

Astrophysical properties of star clusters in the Magellanic Clouds homogeneously estimated by ASteCA

G.I. Perren^{1,3*}, A.E. Piatti^{2,3}, and R.A. Vázquez^{1,3}

¹ Facultad de Ciencias Astronómicas y Geofísicas (UNLP), IALP-CONICET, La Plata, Argentina

² Observatorio Astronómico, Universidad Nacional de Córdoba, Córdoba, Argentina

³ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

Received XXXX XX, 2016; accepted XXXX XX, XXXX

ABSTRACT

Aims. To produce an homogeneous catalog of astrophysical parameters of 239 resolved star clusters located in the Small and Large Magellanic Clouds, observed in the Washington photometric system.

Methods. The cluster sample was processed with the recently introduced Automated Stellar Cluster Analysis (ASteCA) package, which ensures both an automatized and a fully reproducible treatment, together with a statistically based analysis of their fundamental parameters and associated uncertainties. The fundamental parameters determined with this tool for each cluster, via a color-magnitude diagram (CMD) analysis, are: metallicity, age, reddening, distance modulus, and total mass.

Results. We generated an homogeneous catalog of structural and fundamental parameters for the studied cluster sample, and performed a detailed internal error analysis along with a thorough comparison with values taken from twenty-six published articles. We studied the distribution of cluster fundamental parameters in both Clouds, and obtained their age-metallicity relationships.

Conclusions. The ASteCA package can be applied to an unsupervised determination of fundamental cluster parameters; a task of increasing relevance as more data becomes available through upcoming surveys.

Key words. catalogs – galaxies: fundamental parameters – galaxies: star clusters: general – Magellanic Clouds – methods: statistical – techniques: photometric

1. Introduction

The study of a galaxy’s structure, dynamics, star formation history, chemical enrichment history, etc., can be conducted from the analysis of its star clusters. Star clusters in the Magellanic Clouds (MCs) are made up of a varying number of coeval stars which share a chemical composition, are assumed to be located relatively at the same distance from the Sun, and affected by roughly the same amount of reddening. All these factors greatly facilitate the estimation of their fundamental parameters, and thus the properties of their host galaxy. New developments in astrophysical software allow the homogeneous processing of different types of star clusters’ databases. The article series by Kharchenko et al. (see Kharchenko et al. 2005; Schmeja et al. 2014, and references therein) and the integrated photometry based derivation of age and mass for 920 clusters presented in Popescu et al. (2012)¹, are examples of semi-automated and automated packages applied on a large number of clusters.

However, there is no guarantee that by employing a homogeneous method, we will obtain similar parameter values from the same cluster photometric data set across different studies. This is particularly true when the methods require user intervention, which inevitably makes the results subjective to some degree. In Netopil et al. (2015) the open cluster parameters age, reddening, and distance are contrasted throughout seven published databases (metallicity is not inspected as this quantity is usually

assumed rather than adjusted). The authors found that all articles show non-negligible offsets in their fundamental parameter values. This result highlights an important issue: most of the color-magnitude diagram (CMD) isochrone fits are done by-eye, adjusting correlated parameters independently, and often omitting a proper error treatment (see von Hippel et al. 2014, for a more detailed description of this problem). When statistical methods are employed, the used code is seldom publicly shared to allow scrutiny by the community. There is then no objective way to assess the underlying reliability of each set of results. Lacking this basic audit, the decision of which database values to use becomes a matter of preference.

As demonstrated by Hills et al. (2015), assigning precise fundamental parameters for an observed star cluster is not a straightforward task. Using combinations of up to eight filters (*UBVRIJHK_s*) and three stellar evolutionary models, they analyzed the NGC188 open cluster with their Bayesian isochrone fitting technique², and arrived at statistically different results depending on the isochrones and the filters used. NGC188 is a ~4 Gyr cluster with a well defined Main Sequence (MS) mostly unaffected by field star contamination, and with proper motions and radial velocities data available. Therefore, the study of star clusters affected by a significant amount of field star contamination, observed with fewer filters, and without information about its dynamics, should result significantly more complicated. A mismatch between several theoretical evolutionary models, along with an inability to reproduce clusters in the unevolved MS

* e-mail: gabrielperren@gmail.com

¹ Based on their MASSive CLuster Evolution and ANALysis (MASS-CLEAN) package: <http://www.massclean.org/>

² BASE-9: <http://webfac.db.erau.edu/~vonhippt/base9/>

domain, had already been reported in Grocholski & Sarajedini (2003).

The aforementioned difficulties in the analysis of star cluster CMDs will only increase if the study is done by-eye, since: a) the number of possible solutions manually fitted is several orders of magnitude smaller than that handled by a code, b) correlations between parameters are almost entirely disregarded, c) uncertainties can not be assigned through valid mathematical means – and are often not assigned at all –, and d) the final values are necessarily highly subjective. The need is clear for an automated general method with a fully open and extensible code base, that takes as much information into account as possible, and capable of generating reliable results.

In Perren et al. (2015, hereafter Paper I) we presented the Automated STEllar Cluster Analysis (ASteCA³) package, aimed at allowing an accurate and comprehensive study of star clusters. Through a mostly unassisted process the code analyses clusters' positional and photometric data sets to derive their fundamental parameters, and their uncertainties. ASteCA was applied on a group of 20 Milky Way open clusters in Paper I. As shown in that article, the code is able to assign precise parameter values for clusters with low to medium field star contamination, and gives reasonable estimations for heavily contaminated clusters. Every part of this astrophysical package is open and publicly available, and its development is ongoing.

In the present work we apply ASteCA on 239 clusters in the Small and Large Magellanic Clouds (S/LMC), distributed up to $\sim 5^\circ$ and $\sim 8^\circ$ in angular distance from their centers, respectively. The MCs are located close enough to us to allow the study of their resolved star clusters. The large number of cataloged clusters – ~ 4000 are listed in the Bica et al. (2008) catalog – makes them an invaluable resource for investigating the properties of the two most massive galaxies that orbit the Milky Way. The reddening that affects the MCs is relatively small, except for a few regions like 30 Doradus in the LMC, where E_{B-V} can reach values above 0.4 mag (Piatti et al. 2015a). The overall low levels of reddening further simplifies the research of these two galaxies, through the estimation of their clusters' parameters. We make use of photometric data sets in the CT_1 the Washington system (Canterna 1976; Geisler 1996), which is known to be highly sensitive to metal abundance for star clusters older than ~ 1 Gyr (Geisler & Sarajedini 1999). The results obtained here regarding the metal content of the clusters in our sample are thus of relevance for the analysis of the MCs chemical enrichment history.

This is the first study where such a large sample of resolved star clusters is homogeneously analyzed in an automatic way, with all their fundamental parameters statistically estimated rather than fitted by eye or fixed a priori. Having metal content assigned for 100% of our sample is particularly important, especially compared to other star cluster catalogs. The latest version of the well known DAML02 database⁴ (Dias et al. 2002), for example, reports abundances for only 13% out of the 2167 clusters cataloged. Estimations for the total cluster mass is given in few cases, if integrated photometry is provided.

This article is structured as follows. In Sect. 2 we present the star cluster sample used in this work along with numerous stud-

figures/as_RA_DEC.png

Fig. 1. Distribution of our set of analyzed clusters, shown in red, superimposed on to the Bica et al. (2008) database of 3740 star clusters (black dots), for both MCs. The assumed centers for the Clouds are marked with a blue triangle.

ies used to compare and validate our results. Sections 3 and 4 describe the estimation of the fundamental parameters derived with ASteCA, and analyze their uncertainties, respectively. In Sect. 5 a detailed comparison of our results with published values from the literature is performed. Sect. 6 shows the distribution of cluster fundamental parameters in the MCs, and their our catalog, and the age-metallicity relationships (AMRs) for the cluster system. Sect. 7 summarizes our results and concluding remarks.

2. Clusters sample

The data sets used in this work consists of CT_1 Washington magnitudes and their errors for 239 star clusters; 150 of them located in the LMC and the remaining ones in the SMC. In Fig. 1 we show the spatial distribution of the studied clusters (red circles), overplotted on those cataloged by Bica et al. (2008) (black dots).

Table 1 list the sources of the photometric data sets used. These nineteen articles can be merged into a single group, whose results arise from the analysis of the same CT_1 photometry used by ASteCA in the current study. From here on we will refer to this group as the “literature”, and the values presented in each of the articles in Table 1 as “literature values”. Metallicity, age, reddening, and distance modulus (μ_o) values were estimated or assigned for most of the studied clusters, except for the 36 clusters presented in Piatti (2011b) which had only their ages estimated via the δT_1 index (Phelps et al. 1994; Geisler et al. 1997). In most of the listed works the metallicity is fixed to $[Fe/H] = -0.7$ dex, $[Fe/H] = -0.4$ dex, while the distance modulus values are always fixed to $\mu_o = 18.9$ mag and $\mu_o = 18.5$ mag (or very close values), for the S/LMC respectively. Ages reported in the literature were obtained by eye matching, either through the standard isochrone technique or applying the δT_1 index method. Reddenings were estimated in almost all cases interpolating the maps of either Burstein & Heiles (1982), Schlegel et al. (1998),

³ ASteCA is released under a general public license (GPL v3; <https://www.gnu.org/copyleft/gpl.html>) and can be downloaded from its official site: <http://asteca.github.io>

⁴ Latest version: v3.5, 2016 Jan 28; <http://www.wilton.unifei.edu.br/ocdb/>

Table 1. Sources of the CT_1 data sets used in this work. N refers to the number of clusters analyzed per article.

Article	N	Galaxy	Telescope
Geisler et al. (2003)	8	LMC	CTIO 0.9m
Piatti et al. (2003b)	5	LMC	CTIO 0.9m
Piatti et al. (2003a)	6	LMC	CTIO 0.9m
Piatti et al. (2005)	8	SMC	CTIO 0.9m
Piatti et al. (2007a)	4	SMC	CTIO 0.9m
Piatti et al. (2007c)	2	SMC	Danish 1.54m
Piatti et al. (2007b)	2	SMC	Danish 1.54m
Piatti et al. (2008)	6	SMC	Danish 1.54m
Piatti et al. (2009)	5	LMC	CTIO 0.9m / Danish 1.54m
Piatti et al. (2011b)	3	LMC	CTIO 0.9m
Piatti et al. (2011a)	14	SMC	CTIO 1.5m
Piatti (2011a)	9	SMC	Blanco 4m
Piatti (2011b)	36	LMC	Blanco 4m
Piatti (2011c)	11	SMC	Blanco 4m
Piatti (2012)	26	LMC	Blanco 4m
Piatti & Bica (2012)	4	SMC	Blanco 4m
Palma et al. (2013)	23	LMC	Blanco 4m
Maia et al. (2013)	29	SMC	Blanco 4m
Choudhury et al. (2015)	38	LMC	Blanco 4m

Table 2. Sources of Johnson-Kron-Cousins photometric data sets for some clusters in our sample. N refers to the number of clusters in common with our sample per article.

Article	N	Galaxy	Phot
Pietrzynski & Udalski (1999), P99	7	SMC	<i>BVI</i>
Pietrzynski & Udalski (2000), P00	25	LMC	<i>BVI</i>
Hunter et al. (2003), H03	62	S/LMC	<i>UBVR</i>
Rafelski & Zaritsky (2005), R05	24	SMC	<i>UBVI</i>
Chiosi et al. (2006), C06	16	SMC	<i>VI</i>
Glatt et al. (2010), G10	61	S/LMC	<i>UBVI</i>
Popescu et al. (2012), P12	48	LMC	<i>UBVR</i>

or Haschke et al. (2011). Maia et al. (2013) presented total mass estimations for their 29 star clusters sample.

Our cluster sample was also partially studied from Johnson-Kron-Cousins photometry, as listed in Table 2. This group will be referred from here on as the “databases” (DBs), as a way to distinguish them from the “literature” articles mentioned above. In Sect. 5 we show how the parameter values obtained by ASteCA for our sample of clusters, compares to those given in both the literature and the databases.

3. Estimation of star cluster parameters

Fundamental – metallicity, age, distance modulus, reddening, mass – and structural – center coordinates, radius, contamination index, approximate number of members, membership probabilities, true cluster probability – parameters, were obtained either automatically or semi-automatically by the ASteCA package. A detailed description of the functions built within this tool can be found in Paper I, and in the code’s online documentation⁵. The resulting catalog can be accessed via VizieR⁶. We’ve made available online the entire Python codebase developed to analyze the data obtained with ASteCA and generate the figures in this arti-

cle.⁷ In addition, the full output image generated by ASteCA for each processed cluster together with its corresponding membership probability file, can be accessed separately through a public code repository.⁸

3.1. Ranges for fitted fundamental parameters

Before ASteCA processes a cluster photometric data set, the user is required to provide a suitable range for each fundamental parameter. This is accomplished by setting a minimum, a maximum, and a step value in the appropriate input data file. As explained in Sect. 3.4, each combination of the five fundamental parameters represents a unique synthetic cluster CMD, or model. Hence, an interval too broad between the minimum and maximum values for a given parameter, or a very small step – or a combination of both – will result in a large number of possibilities in the resulting fundamental parameters values. The larger the number of parameters values, the larger the number of models the code will process to find the synthetic CMD which best matches the observed cluster CMD. For this reason, the allowed ranges and steps were selected to provide a balance between a reasonably large interval, and a computationally manageable number of total models. Especial care must be taken to avoid defining a range that could bias the results towards a particular region of any fitted parameter. The adopted ranges are listed in Table 3.

Unlike most previous works where the metallicity is a pre-defined fixed value, we do not make any assumptions on the cluster’s metal content. The minimum and maximum $[\text{Fe}/\text{H}]$ values selected are $[-2.2, 0]$ dex; with a step of ~ 0.1 dex. This interval covers completely the usual metallicities reported for clusters in the MCs. The age range is set to 1 Myr–12.6 Gyr with a step of $\log(\text{age/yr}) = 0.05$ dex, encompassing almost the entire allowed range of the CMD service⁹ where the theoretical isochrones were obtained from (see Sect. 3.4).

The maximum allowed value for the reddening of each cluster was determined through the MCEV¹⁰ reddening maps (Haschke et al. 2011), while the minimum value is always zero. These maps contain a large number of observed areas with assigned E_{V-I} values. The TOPCAT¹¹ tool was used to query reddening values within a region as small as possible around the positions of each cluster. For approximately 85% of our sample we found several regions with an associated reddening value, within a box of 0.5 deg centered on the cluster’s position. For the remaining clusters, larger boxes had to be used to find regions with assigned E_{V-I} values. The two most extreme cases are NGC1997 ($\alpha=5^{\text{h}}30^{\text{m}}34^{\text{s}}$, $\delta=-63^{\circ}12'12''$ [J2000.0]) and OHSC28 ($\alpha=5^{\text{h}}55^{\text{m}}35^{\text{s}}$, $\delta=-62^{\circ}20'43''$ [J2000.0]) in the outskirts of the LMC, where boxes of 4 deg and 6 deg respectively were needed to find a region with an assigned reddening value. In both cases, the reddenings given by the Schlafly & Finkbeiner (2011) map for their coordinates are up to two times smaller than the ones found in the MCEV map. We adopted the largest E_{V-I} value of each region, MCEV_{max} , as the upper limit in the reddening range. Three steps are used to ensure that the reddening range is partitioned similarly for all MCEV_{max} values: 0.01

⁷ <https://github.com/Gabriel-p/mc-catalog>

⁸ <https://github.com/Gabriel-p/mc-catalog-figs>

⁹ CMD: <http://stev.oapd.inaf.it/cgi-bin/cmd>

¹⁰ Magellanic Clouds Extinction Values (MCEV): <http://dc.zah.uni-heidelberg.de/mcextinct/q/cone/form>

¹¹ TOPCAT: <http://www.star.bris.ac.uk/~mbt/topcat/>

⁵ ASteCA documentation: <http://asteca.rtfid.org>

⁶ <http://vizier.cesr.cnrs.fr>

Table 3. Fundamental parameters’ ranges used by ASteCA on our set of 239 clusters. The approximate number of values used for each parameter is N . This gives a combined total of $\sim 2.3 \times 10^7$ possible models (or synthetic cluster CMDs), that could be theoretically matched to each studied cluster in our sample.

Parameter	Min	Max	Step	N
[Fe/H]	~ -2.2	0	~ 0.1	23
$\log(\text{age/yr})$	6.	10.1	0.05	82
E_{B-V}	0.0	MCEV_{max}	~ 0.01	~ 12
μ_{SMC}	18.86	19.06	0.02	10
μ_{LMC}	18.4	18.6	0.02	10
Mass (M_{\odot})	10	$[1, 3] \times 10^4$	200	$[50, 150]$

for $\text{MCEV}_{\text{max}} > 0.1$, 0.02 if $0.05 \leq \text{MCEV}_{\text{max}} \leq 0.1$, and 0.005 for $\text{MCEV}_{\text{max}} < 0.05$. The E_{V-I} extinction is converted to E_{B-V} following Tammann et al. (2003): $E_{V-I} = 1.38 E_{B-V}$, with an extinction law of $R_V = 3.1$ applied throughout the analysis.

The distance modulus ranges for the S/LMC Clouds were set to 18.96 ± 0.1 mag and 18.46 ± 0.1 mag respectively; where the mean values are taken from de Grijs & Bono (2015) and de Grijs et al. (2014). Line of sight (LOS) depths for the MCs (front to back, $\pm 1\sigma$) have been reported to span up to 20 kpc in their deepest regions (Subramanian & Subramanian 2009; Nidever et al. 2013; Scowcroft et al. 2015). The 0.1 mag deviations allowed in this work give LOS depths of ~ 5.7 kpc and ~ 4.6 kpc for the S/LMC. This covers more than half of the average LOS depths found in Subramanian & Subramanian (2009) for both MCs.¹² Although this is not enough to cover the entire observed depth ranges for the MCs, it gives the distance moduli a certain liberty to move around their mean values when all parameters are adjusted.

For the total cluster mass, the range was set to a minimum of $10 M_{\odot}$ and a maximum of $10000 M_{\odot}$. This is true for almost all the clusters in our sample, except for 15 visibly massive clusters for which the maximum mass limit was increased to $30000 M_{\odot}$. The step used to divide the mass range was $200 M_{\odot}$, which also sets the minimum possible uncertainty for a cluster’s mass estimate.

Finally, the binary fraction was fixed to a value of 0.5 – considered a reasonable estimate for clusters (von Hippel 2005; Sollima et al. 2010) – to avoid introducing an extra degree of complexity into the fitting process. Secondary masses are randomly drawn from a uniform mass ratio distribution of the form $0.7 \leq q \leq 1$, where $q = M_2/M_1$, and M_1 , M_2 are the primary and secondary masses. This range for the secondary masses was found for example for the LMC cluster NGC1818 in Elson et al. (1998), and represents a value commonly used in analysis regarding the MCs (see Rubele et al. 2011, and references therein).

3.2. Center and radius assignment

In automatic mode, ASteCA uses a Gaussian kernel density estimator (KDE) function to determine the center of the cluster as the point of maximum star density. A radial density profile (RDP) is used to estimate the cluster’s radius, as the point where the RDP reaches the mean star density of the surrounding field.

The function that obtains the center coordinates requires that the density of cluster members make it stand out over the combination of foreground and background field stars, and that no

other over-density is present in the observed frame.¹³ The function for estimating the radius will give good results when the RDP generated is reasonably smooth, and only if the cluster does not occupy the entire observed frame; i.e.: some portion of the surrounding field must be visible. When these conditions are not met for either function, the user needs to run the code on semi-automatic mode. In this mode the center coordinates can be either manually fixed, or found by the code based on an initial set of approximate values. Similarly, in semi-automatic mode the radius must be fixed by the user.

In approximately $\sim 66\%$ of our clusters sample, the center coordinates were obtained via a KDE analysis, based on initial approximate values. Radii values were calculated by ASteCA for $\sim 60\%$ of the clusters, through an RDP analysis of the observed surrounding field. The remaining clusters are those that are highly contaminated, contain very few observed stars, and/or occupy most of the observed frames. This means that an automatic estimation of their centers and/or radii was not possible, and the semi-automatic mode was used.

The contamination index (CI) calculated by the code is a parameter related to the number of foreground/background stars present in the defined cluster region. A value of $CI=0$ means that the cluster is not contaminated by field stars, $CI=0.5$ means that the overdensity is indistinguishable from the surrounding field (i.e.: the average number of field stars expected within the cluster region equals the total number of stars in it), and $CI>0.5$ means that more field stars are expected within the cluster region than cluster stars.¹⁴ For reference, the average CI of the set of clusters with manual radii assignment is $CI \approx 0.9 \pm 0.2$, while for those star clusters whose radii were estimated in automatic mode the average is $CI \approx 0.6 \pm 0.2$.

3.3. Field-star decontamination

A decontamination algorithm (DA) was employed on the CT_1 CMD of each processed cluster. The DA allows cleaning the CMD of field-star contamination, before the isochrone fitting function is used (see Sect. 3.4). The DA processes all stars within the defined cluster region, i.e: the circle centered on the cluster and with radius equal to the cluster radius.

The Bayesian DA presented in Paper I was improved for [this article the present analysis](#); the new DA works in two steps. First, the original Bayesian membership probability (MP) assignment is applied to the CMD of all stars within the cluster region (see Paper I for more details). After that, a cleaning algorithm is used to remove stars of low MPs from the CMD, as shown in Fig. 2. To do this, the full CMD is divided into smaller cells, according to a given binning method. By default ASteCA uses the Bayesian blocks method¹⁵ introduced in Scargle et al. (2013), via the implementation of the astroML package (Vanderplas et al. 2012).¹⁶ This second step is similar to the field-star density based cell-by-cell removal algorithm applied in Bonatto & Bica (2007, B07), which uses a simpler rectangular grid. The main difference, aside from the binning method employed, is that ASteCA can remove

¹³ The single over-density limitation is planned to be lifted in upcoming versions of ASteCA.

¹⁴ See Paper I, Sect. 2.3.2 for a complete mathematical definition of the index.

¹⁵ http://www.astroml.org/examples/algorithms/plot_bayesian_blocks.html

¹⁶ While Bayesian blocks binning is the default setting in ASteCA, several others techniques for CMD star removal are available, as well as five more binning methods.

¹² Average LOS depth, SMC: 4.9 ± 1.2 kpc (bar), 4.23 ± 1.48 kpc (disk); LMC: 4.0 ± 1.4 kpc (bar), 3.44 ± 1.16 kpc (disk).

figures/L62_DA_BF.png

Fig. 2. *Top:* CMD of the cluster region for the SMC-L62 cluster, where n_{memb} is the approximate number of cluster stars based on the star density of the cluster compared to the star density of the field (left), and the CMD of the surrounding field region (right). In both panels N_{accept} is the number of stars that were not rejected due to their large photometric errors. Rejected stars with large errors are shown as green crosses. *Bottom:* cluster region after applying the DA (left); MPs vary according to the colorbar at the top right. Dotted horizontal and vertical lines show the binning used to reject low MP stars cell-by-cell, as obtained via the Bayesian blocks method. Open circles (drawn semi-transparent) represent rejected stars. N_{fit} is the number of stars kept unsubtracted by the cell-by-cell rejection. The best fitted isochrone is overplotted with a green line. The respective generated synthetic CMD is shown in the right panel. N_{synth} is the number of stars in the synthetic cluster CMD, and the dotted lines represent the binning obtained using Knuth’s rule, applied in the synthetic cluster match process.

stars cell-by-cell based on their previously assigned MPs, not randomly as done in B07.

Approximately 70% of our sample was processed with default settings, this is: a Bayesian DA followed by the removal of low MP stars, based on a Bayesian block binning method of the cluster region’s CMD. The remaining clusters were processed with modified settings to allow a proper field-star decontamination. This ~30% of clusters are mainly those with a low number of stars or heavily contaminated by field stars. Changes introduced were, for instance, a different binning method (often a rectangular grid using Scott’s rule, Scott 1979), or skipping the Bayesian MP assignation and only performing a density based cell-by-cell statistical field star removal. In this latter case, the DA works very similarly to the B07 algorithm.

An appropriate field-star decontamination is of the utmost importance, since the cleaned cluster CMDs will be used to estimate the cluster fundamental parameters.

3.4. Isochrone fitting

ASteCA generates synthetic CMDs to match the observed ones and thus derive the cluster’s fundamental parameters. In this work the synthetic CMDs are built using the PARSEC v1.1 the-

oretical isochrones computed by Bressan et al. (2012, B12), and the log-normal initial mass function (IMF) defined by Chabrier (2001).¹⁷ For a given age, metallicity, and total mass, a synthetic CMD is built from a stochastically sampled IMF, and then shifted by reddening and distance modulus.

The Poisson likelihood rate (PLR) developed in Dolphin (2002) is employed to assess the goodness of match between the star cluster’s CMD and a synthetic CMD. As the PLR is a binned statistic, it requires binning the cluster CMD, and all the synthetic CMDs that can be generated according to the fundamental parameter ranges defined in Sect.3.1. This part of the analysis uses Knuth’s rule (Knuth 2006, also implemented via the astroML package) as the default binning method (see bottom right plot in Fig. 2). The inverted logarithmic form of the PLR can be written as

$$LPLR = -\ln PLR = \sum_i m_i - n_i + n_i \ln \frac{n_i}{m_i}, \quad (1)$$

where m_i and n_i are the number of stars in the i th cell of the synthetic and the observed cluster CMD, respectively.¹⁸ In Paper I the total cluster mass parameter could not be estimated, due to the likelihood statistic used (see Paper I, Eq. 11). The LPR defined in Eq. 1 allows us to also consider the mass as a free parameter in the search for the best synthetic CMD. Following the validation performed in Paper I for the metallicity, age, reddening, and distance, we present in Appendix A a validation study for the total mass. We demonstrate that the masses recovered by ASteCA for nearly 800 MASSCLEAN synthetic clusters are in excellent agreement with the masses used to generate them.

The synthetic CMD which best resembles the observed cluster CMD is obtained by minimizing the LPLR. Since we deal with five free fundamental parameters, a 5-dimensional surface of solutions is built from all the possible synthetic CMDs that can be matched to the observed one. ASteCA applies a genetic algorithm (GA) on this surface to derive the cluster fundamental parameters.

The GA uses a set of reasonable default options to perform the matching procedure between the observed and synthetic CMDs. These options had to be modified – mainly extending the depth of the GA search for the best match – for ~30% of the clusters in the sample. This sub-sample is composed of clusters with particularly complicated morphologies, i.e.: very low number of stars, high field star contamination, no MS present in the cluster CMDs, etc.

Once the GA returns the optimal fundamental parameter values, uncertainties are estimated via a standard bootstrap technique. This process takes a significant amount of time to complete, since it involves running the GA several more times on a randomly generated cluster with replacement.¹⁹ We run the bootstrap process ten times per cluster, as it would be prohibitively

¹⁷ For a detailed description on the ASteCA technique for the generation of synthetic CMDs from theoretical isochrones, see: Paper I, Sect. 2.9.1. All the evolutionary tracks from the CMD service (<http://stev.oapd.inaf.it/cgi-bin/cmd>) are currently supported, as well as three other IMFs.

¹⁸ If for any given cell we have $n_i \neq 0$ and $m_i = 0$, a very small number is used instead ($m_i = 1 \times 10^{-10}$) to avoid a mathematical inconsistency with the factor $\ln m_i$.

¹⁹ Generating a new cluster “with replacement”, means randomly picking stars one by one from the original cluster, where any star can be selected more than once. The process stops when the same number of stars as those present in the original cluster have been picked.

costly timewise to run it – as would be ideal – hundreds or even thousands of times.

4. Errors in fitted parameters

It is well known that a parameter given with no error estimation is meaningless from a physical standpoint (Dolphin 2002; Andrae 2010). ~~This fact notwithstanding~~ Despite this, a detailed error treatment is usually ignored in articles that deal with star clusters analysis (Paunzen & Netopil 2006). As explained in Sect. 3.4, ASteCA employs a bootstrap method to assign standard deviations for each fitted parameter in our clusters sample.

Since the code must simultaneously fit a large number of free parameters – five in this case – within a wide range of allowed values, and using only a 2-dimensional space of observed data, i.e., the T_1 versus $(C - T_1)$ CMD²⁰, the uncertainties will expectedly be somewhat large. It is worth noting that unlike manually set errors, these are statistically valid uncertainty estimates obtained via a bootstrap process. This is an important point to make given that the usual by-eye isochrone fitting method not only disregards known correlations among all clusters parameters, it is also fundamentally incapable of producing a valid error analysis (Naylor & Jeffries 2006). Any uncertainty estimate produced by-eye serves only as a mere approximation, which will often be biased towards smaller figures. The average logarithmic age error given in the literature, for example, is 0.16 dex, in contrast with the almost twice as large average value estimated by ASteCA (see below). In Fig. 3 we show the distribution of the standard deviations versus the five fundamental parameters fitted by the code, for the entire sample (SMC and LMC clusters).

The apparent dependence of the metallicity error with decreasing $[\text{Fe}/\text{H}]$ values arises from the fact that ASteCA uses z values, where $[\text{Fe}/\text{H}] = \log(z/z_\odot)$; with $z_\odot = 0.0152$. The error in z is found to be on average $e_z \approx 0.003$ for the whole cluster sample.²¹ This means that, when converting to $e_{[\text{Fe}/\text{H}]}$ using the relation

$$\sigma_{[\text{Fe}/\text{H}]} = \sigma_z / [z \times \ln(10)], \quad (2)$$

the z in the denominator makes $\sigma_{[\text{Fe}/\text{H}]}$ grow as it decreases, while σ_z remains more or less constant. For very small z values (e.g.: $z = 0.0001$), the logarithmic errors can easily surpass $\sigma_{[\text{Fe}/\text{H}]} = 2$ dex. In those cases, as seen in Fig. 3, the error is trimmed to 2 dex which is enough to cover the entire metallicity range.

As can be seen, there are no visible trends in the arrangement of errors for any of the fitted parameters. This is a desirable feature for any statistical method. If the uncertainty of a parameter varied (increase/decrease) with it, it would indicate that ASteCA was introducing biases in the solutions.

Histograms plotted to the right in Fig. 3 show the distribution of errors, and their arithmetic means as a dashed red line. The average $[\text{Fe}/\text{H}]$ and $\log(\text{age}/\text{yr})$ errors for the sample are 0.3 dex and 0.27 dex, respectively. For the SMC $\sigma_{[\text{Fe}/\text{H}]} < 0.2$ dex for ~34% of the clusters. We found similar errors for ~69% of the LMC clusters. Approximately 53% of the combined S/LMC sample show $\sigma_{\log(\text{age}/\text{yr})} < 0.1$ dex. Error estimates for the

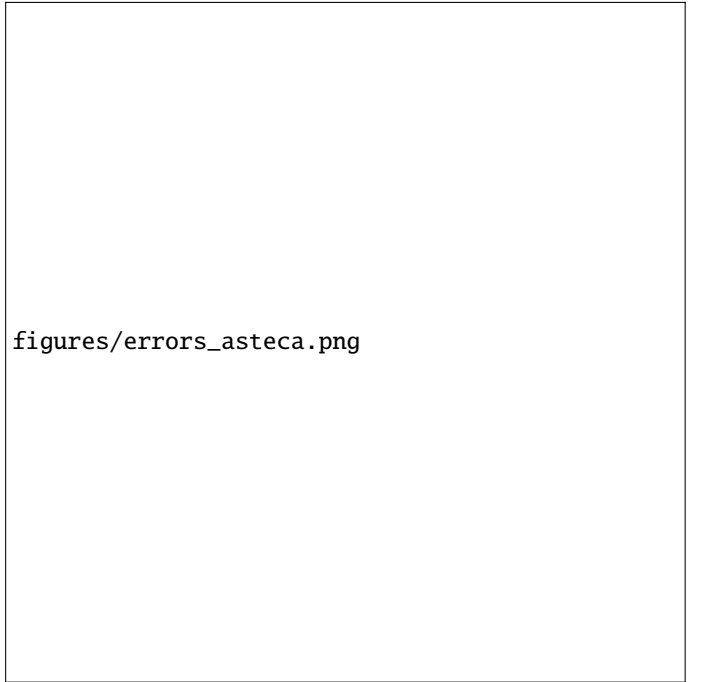


Fig. 3. *Left:* Distribution of errors versus the five parameters fitted by ASteCA. Colors are associated to the CI (see bar in top plot), sizes are proportional to the actual cluster sizes. A small random scatter in the x -axis is added for clarity. *Right:* Error histogram. The mean error value for each parameter is shown in the top right corner, and drawn in the plot with a dashed red line.

remaining parameters are all within acceptable ranges.

Numerically, errors could be lowered applying several different techniques: increase the number of bootstrap runs, increase the number of models evaluated in the GA (“generations”), or reduce the value of the steps in the parameters grid. All of these methods will necessarily extend the time needed to process each cluster, in particular increasing the number of bootstrap runs. Limited computational time available requires a balance between the maximum processing power allocated to the calculations, and the precision one is aiming at. The error values presented here should then be considered a conservative upper limit of the accuracy with which the fundamental parameters of our sample can be obtained by ASteCA.

5. Comparison with published fundamental parameters

We compare in Sect. 5.1 the resulting parameter values obtained by ASteCA, with those taken from the papers that used the same CT_1 data sets (see Table 1), referred as the “literature”. The parameters age, reddening, and mass are also analyzed in Sect 5.2 for a subset of 142 star clusters that could be cross-matched with $UBVRI$ photometry results (see Table 2), referred as the “databases”.

5.1. Literature values

The comparison of ASteCA versus literature values for the metallicity, age, reddening, and distance modulus, is presented in Fig. 4. The left and central panels show the 1:1 identity line for the LMC and the SMC respectively. The right panel diagram

²⁰ We plan to upgrade the code to eventually allow more than just two observed magnitudes, therefore extending the 2-dimensional CMD analysis to an N-dimensional one.

²¹ Approximately 70% and over 80% of the clusters in the S/LMC, respectively, have assigned $e_z = 0.003$ errors.

figures/as_vs_lit_S-LMC.png

Fig. 4. *Left column:* parameters comparison for the LMC. *Center column:* idem for the SMC. *Right column:* BA plot with differences in the sense ASteCA minus literature, for the combined S/LMC sample. For clarity, a small random scatter is added to both axes for the metallicity and distance modulus plots. Mean and standard deviation are shown as a dashed line and a gray band, respectively; their values are displayed in the top left of the plot. Colors following the coding shown in the bar at the right of the figure, for each row. Piatti (2011b) clusters which contain only age information are plotted with $E_{B-V}=0$ color coding.

shows a Bland-Altman (BA) plot for our combined sample of clusters, with the variation in the ~~x-axis~~x-axis proposed by Krouwer (Bland & Altman 1986; Krouwer 2008)²² The BA plot shows differences in the ~~y-axis~~y-axis in the sense $\Delta =$

(ASteCA minus literature), versus values found by the code in the ~~x-axis~~x-axis. Errors in Δ values are calculated combining the errors for both estimates. The mean of the differences, $\bar{\Delta}$, is shown as a dashed line and its standard deviation as a gray band.

²² The BA is also called a “difference” plot or more commonly a “Tukey Mean-Difference” plot. In the original BA plot the default ~~x-axis~~x-axis displays the mean values between the two methods being compared. The Krouwer variation changes the means for the values of one of those methods, called the “reference”. In our case, the reference method is ASteCA so we use its reported values in the ~~x-axis~~x-axis.

An offset is noticeable for the abundance estimates (upper panels of Fig. 4), where ASteCA tends to assign values ~ 0.22 dex larger than those found in the literature. On average, the offset is of ~ 0.27 dex for the SMC and ~ 0.18 dex for the LMC. This effect can be explained by two different processes, in light

of our knowledge that the ASteCA’s best fit CMD matching introduces no biases into the solutions. First, the MC’s star clusters are generally considered to have low metallicities. For example, the $[\text{Fe}/\text{H}]$ values assumed in the 19 literature articles, are exactly -0.7 dex and -0.4 dex for $\sim 60\%$ and $\sim 75\%$ of the S/LMC clusters. The by-eye fit is thus very likely biased towards the assignment of lower abundances. This type of “confirmation bias” in published values has been studied recently by de Grijs et al. (2014), in relation to distance measurements reported for the LMC. Researchers also tend to fit isochrones adjusting it to the lower envelope of a cluster’s sequence; which also contributes to the selection of isochrones of smaller metallicity.²³ Second, ASteCA’s z values are converted to the logarithmic form $[\text{Fe}/\text{H}]$ using a solar metallicity of $z_{\odot}=0.0152$ (Bressan et al. 2012). Values found in the literature, on the other hand, are converted using a solar metal content of $z_{\odot}=0.019$ (Marigo et al. 2008). This means that ASteCA will always give $[\text{Fe}/\text{H}]$ values larger by ~ 0.1 dex, for any fitted isochrone of equivalent z being compared, following

$$\begin{aligned}\Delta[\text{Fe}/\text{H}] &= [\text{Fe}/\text{H}]_{\text{ASteCA}} - [\text{Fe}/\text{H}]_{\text{literature}} \\ &= \log(z/0.0152) - \log(z/0.019) = \log(0.019/0.0152) \\ &\approx 0.0969\text{dex}\end{aligned}\tag{3}$$

These two effects combined can therefore explain the offset found for the abundances assigned to star clusters in the literature versus those estimated by ASteCA. It is important to notice that neither effect is intrinsic to the best likelihood matching method used by the code.

The general dispersion between literature and ASteCA values can be quantized by the standard deviation of the Δ differences in the BA plot. This value is ~ 0.32 dex (top left of BA plot), in close agreement with the mean internal uncertainty found in Sect. 4 for this parameter. Mean metallicity estimates for the MCs using ASteCA values are $[\text{Fe}/\text{H}]_{\text{SMC}} \approx -0.52 \pm 0.44$ dex, and $[\text{Fe}/\text{H}]_{\text{LMC}} \approx -0.26 \pm 0.24$ dex. These averages are similar within their uncertainties, to those obtained using literature values: $[\text{Fe}/\text{H}]_{\text{SMC}} \approx -0.78 \pm 0.23$ dex, and $[\text{Fe}/\text{H}]_{\text{LMC}} \approx -0.42 \pm 0.16$ dex

The comparison between literature and ASteCA ages shows an overall good agreement, with larger differences for some young clusters. Ten clusters with either a shallow photometry, a low number of cluster stars, or heavily contaminated, present an age difference larger than $\Delta \log(\text{age}/\text{yr}) > 0.5$ dex with the values estimated in the literature. These are referred to as “outliers”, and discussed in more depth in Appendix B. Excluding these outliers, the $\Delta \log(\text{age}/\text{yr})$ offset for S/LMC clusters between ASteCA and literature values are ~ 0.02 dex and ~ 0.05 dex, respectively. The mean value for the $\Delta \log(\text{age}/\text{yr})$ parameter including all S/LMC clusters is ~ -0.01 dex, as shown in the BA plot of Fig. 4. This points to an excellent agreement among literature and ASteCA values for the $\log(\text{age}/\text{yr})$. If we exclude the outliers, the mean of the differences increases by a small amount to ~ 0.04 dex. Similarly to what was found for the metallicity, the dispersion between literature and ASteCA values is almost exactly the internal uncertainty found for errors assigned by the code, i.e.: ~ 0.3 dex.

²³ The reason is that increasing an isochrone’s metal content moves it towards redder (greater color) values in a CMD, see for example Bressan et al. (2012), Fig 15.

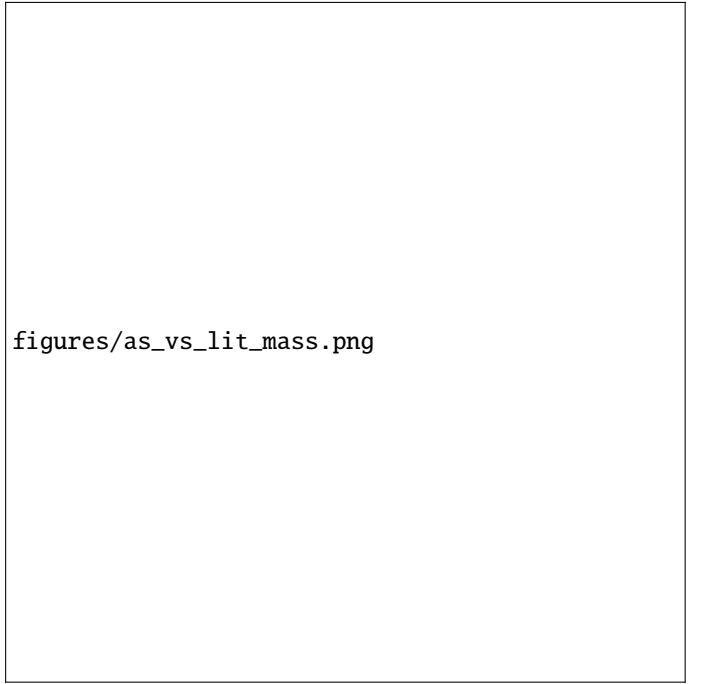


Fig. 5. *Top:* Mass comparison for ASteCA versus literature values. *Bottom:* BA plot with mean and standard deviation of the differences shown as a dashed horizontal line, and a gray band respectively. The mean and standard deviation values are displayed in the top left corner.

The reddening distribution presents maximum E_{B-V} values of ~ 0.15 mag and ~ 0.3 mag for the S/LMC, respectively. The Δ differences are well balanced with a mean of -0.02 and a standard deviation of 0.05 mag, slightly larger than the 0.02 mag average uncertainty found for the errors assigned by the code. We estimate average E_{B-V} values for the S/LMC of 0.03 ± 0.03 mag and 0.05 ± 0.05 mag. Although these are lower estimates – approximately a third – than those used for example in the Hunter et al. (2003) study, our sample does not contain clusters in the regions of the MCs most affected by dust. This explains the much lower mean reddening found in this work.

The distance moduli (μ_0) found by ASteCA in both MCs show a clear displacement from literature values. This is expected, as the distance to clusters in the latter is always assumed to be a fixed constant equivalent to the distance to the center of the corresponding galaxy. The distribution of μ_0 values found by the code covers the entire range allowed in Sect. 3.1. Distances obtained by ASteCA thus vary up to ± 0.1 mag from the default fixed μ_0 values used in the literature. It is worth noting that this variation appears to have no substantial effect on any of the remaining parameters, something that could in principle be expected due to the known correlations between all fundamental parameters (see Paper I, Sect. 3.1.4). This reinforces the idea that using a fixed value for the distance modulus, as done in the literature, is a valid way of reducing the number of free variables at no apparent extra cost.

Since masses are only assigned in Maia et al. (2013, M13) for its sample of 29 SMC clusters, the comparison with ASteCA is presented separately in Fig. 5. The identity plot (top) shows a trend for ASteCA masses to be smaller than those from M13. This is particularly true for two of the clusters with the largest mass estimates – and the largest assigned errors – in M13: H86-

97 ($3300 \pm 1300 M_{\odot}$), and H86-87 ($3100 \pm 1700 M_{\odot}$). The BA plot (bottom) shows that, on average, M13 masses are $\sim 500 \pm 700 M_{\odot}$ larger. To explain these differences, we need to compare the way total cluster masses are obtained by M13 and by ASteCA in this work.

In M13 masses were determined using two methods, both based on estimating a mass function via the T_1 luminosity function (LF) of a cluster. In both cases a field star cleaning process was applied. The first one employs a CMD decontamination procedure (Maia et al. 2010), and the second one cleans the cluster region’s LF by subtracting it a field star LF. The results obtained with these two methods are averaged to generate the final mass values. Although any reasonable field star cleaning algorithm should remove many or most of the foreground/background stars in a cluster region, some field stars are bound to remain. For heavily contaminated clusters this effect will be determinant in shaping the “cleaned” CMD sequence, as cluster stars will be very difficult to disentangle from contaminating field stars. The set of 29 clusters in the M13 sample are indeed heavily affected by field star contamination. This can be seen in Fig. 5, where the color assigned to each cluster corresponds to its CI . The minimum value is $CI \approx 0.55$, which means all cluster regions are expected to contain, on average, more field stars than cluster stars. The presence of a large number of contaminating field stars not only makes the job much harder for the DA, it also necessarily implies that the overall LF will be overestimated, therefore leading – in the case of M13 – to an overestimation of the total mass. In contrast, ASteCA assigns cluster masses taking their values directly from the best match synthetic CMD. Field star contamination will thus have a much lower influence on the code’s mass estimate, limited just to how effective the DA is in cleaning the cluster region. In fact, we find in Appendix A that the code will slightly underestimate the masses by approximately $200 M_{\odot}$, for low mass clusters. These two effects combined explain the $\sim 500 M_{\odot}$ offset found for ASteCA versus literature mass values seen in Fig. 5.

The case of B48 is worth mentioning, as it is the cluster with the largest total mass given in M13 ($3400 \pm 1600 M_{\odot}$). After being cleaned of possible field stars by the DA – see Sect. 3.3 – low mass stars are entirely removed and B48 is left only with its upper sequence ($T_1 < 18.4$ mag). This happens both in the literature and in the analysis done by ASteCA, see left CMD in Fig. 6, and Fig. C8 in M13. The likelihood defined in Eq.1 sees then no statistical benefit in matching the cluster with a SC of a similar age and mass, which will contain a large number of low mass stars. This leads the GA to select as good matches SCs of considerably younger ages ($\log(\text{age}/\text{yr}) < 7.0$ dex) than that assigned in M13 ($\log(\text{age}/\text{yr}) = 7.9 \pm 0.05$ dex), and with much lower mass estimates (see caption of Fig. 6). A good match both in age and in mass could be found by the code, only if the DA was applied with no cell-by-cell removal of low MP stars as shown in the right CMD of Fig. 6. This means that all stars within the cluster region – including field stars – are used in the synthetic CMD matching process, which inevitably questions the reliability of the total mass estimate. Dealing with this statistical effect is not straightforward and will probably require an extra layer of modeling added to the synthetic CMD generation algorithm. As discussed in Appendix B this effect also plays an important role in the significant age differences found between ASteCA and the literature, for a handful of clusters.

figures/B48_DA.png

Fig. 6. *Left:* Best fit isochrone for B48 found by ASteCA when a cell-by-cell removal is applied on its sequence, following the Bayesian MP assignment (removed stars are drawn semi-transparent). The estimated age and total mass are $\log(\text{age}/\text{yr}) = 6.2 \pm 0.6$, and $M = 400 \pm 200 M_{\odot}$. *Right:* Best fit isochrone found when no removal of stars is performed, and the full cluster region is used in the search for the best synthetic cluster match. The estimated age and total mass are now $\log(\text{age}/\text{yr}) = 7.5 \pm 0.3$, and $M = 3000 \pm 900 M_{\odot}$.

5.2. Databases values

In addition to the analysis performed in Sect. 5.1, we compare our results with those taken from seven articles – the “databases” or DBs – where a different photometric system was used; see Table 2. We can further separate these seven DBs into two groups: those where the standard by-eye isochrone fitting method was applied – P99, P00, C06, and G10 – and those where integrated photometry was employed to derive the clusters fundamental parameters – H03, R05, and P12. A total of 142 individual star clusters from our sample could be cross-matched. Where names were not available to perform the cross-match – databases P00, P99, and R05 – we employed a 20 arcsec radius to find matches, based on the equatorial coordinates of the clusters in each DB.

The comparison of ASteCA ages with those from the four isochrone fit DBs is shown in Fig. 7, left and center plot. P99 and P00 analyze SMC and LMC clusters respectively, using Bertelli et al. (1994) isochrones and fixed metallicities for the S/LMC of $z = 0.004, 0.008$. While P99 derives individual reddening estimates based on red clump stars, P00 use reddening values determined for 84 lines-of-sight in the Udalski et al. (1999) LMC Cepheids study. The mean S/LMC extinctions in P99 and P00, for the cross-matched data set, are $E_{B-V} \approx 0.08 \pm 0.02$ mag and $E_{B-V} \approx 0.14 \pm 0.02$ mag²⁴. These are both larger estimates than those found in this work, $E_{B-V} \approx 0.03 \pm 0.03$ mag and $E_{B-V} \approx 0.05 \pm 0.05$ mag for the S/LMC, see Sect. 5.1. Both studies attempt to eliminate field star contamination following the statistical procedure presented in Mateo & Hodge (1986). The

²⁴ Average E_{B-V} extinction for the P database taken from de Grijs & Anders (2006).

figures/cross_match_if.png

Fig. 7. *Left:* age comparison for DBs that used the isochrone fit method, versus ASteCA, where N is the number of clusters cross-matched in each DB. *Center:* “delta” plot, showing the differences between reddening and age, in the sense ASteCA minus DB. Curves represent regions of iso-densities after fitting a 2-dimensional Gaussian Kernel. *Right:* same as previous plot, now showing ASteCA minus literature values for both Clouds.

distance moduli for the S/LMC of 18.65 mag and 18.24 mag employed in these DBs, are approximately ~ 0.25 mag smaller than the canonical distances assumed for each Cloud. This has a direct impact on their obtained ages. In de Grijs & Anders (2006) the authors estimate that had P00 used a value of $\mu_0 = 18.5$ mag instead for the LMC, their ages would be ~ 0.2 – 0.4 dex younger; a similar conclusion is reached by Baumgardt et al. (2013).²⁵ The

same reasoning can be applied to the P99 age estimates. P99 and P00 logarithmic ages are displaced on average from ASteCA values (in the sense ASteCA minus DB) by -0.13 ± 0.6 dex and 0.37 ± 0.5 dex respectively, as seen in Fig. 7 left plot. In the case of P99, the distance modulus correction mentioned above would bring the age values to an overall agreement with those obtained by ASteCA, although with a large scatter around the identity line. P00 age values on the other hand, would end up ~ 0.7 dex below the code’s age estimates after such a correction. Such a large deviation is most likely due to the overestimated reddening values used by P00, as will be shown below.

²⁵ Notice that in Baumgardt et al. (2013) the authors correct the age bias that arises in P00 due to the small distance modulus used, increasing P00 age estimates by 0.2 dex. This is incorrect, ages should have been decreased by that amount.

C06 studied 311 SMC clusters via isochrone fitting applying two methods: visual inspection and a Monte Carlo based χ^2 minimization. The authors also employ a decontamination algorithm to remove contaminating field stars, making this the article that more closely resembles this present work. Distance modulus is assumed to be $\mu_0=18.9$ mag. Reddening and metallicity values of $E_{B-V}=0.08$ mag and $z=0.008$ are used, adjusted when necessary to improve the fit. It is worth noting that the $[\text{Fe}/\text{H}]=-0.4$ dex abundance employed in C06 is closer to the average $[\text{Fe}/\text{H}]=-0.52\pm0.44$ dex value found for the SMC by ASteCA, than the canonical value of $[\text{Fe}/\text{H}]=-0.7$ dex used by default in most works. This is the DB – out of the four isochrone-fit and the three integrated photometry DBs – that best matches ASteCA’s $\log(\text{age}/\text{yr})$ values, with a mean deviation from the identity line of 0.02 ± 0.58 dex.

G10 analyzed over 1500 star clusters with ages <1 Gyr in both Clouds via by-eye isochrone fitting. They assumed distance moduli of (18.9, 18.5) mag, and metallicities of (0.004, 0.008), for the S/LMC respectively. Reddening was adjusted also by-eye on a case-by-case basis. The G10 database presents a systematic bias where smaller logarithmic ages are assigned compared to our values, with an approximate deviation of $\Delta \log(\text{age}/\text{yr})=0.23 \pm 0.46$ (ASteCA minus G10). This is consistent with the results found in Choudhury et al. (2015) (see Fig. 5), and later confirmed in Piatti et al. (2015b,a). G10 does not apply any decontamination method, by which probable field stars are removed from the cluster’s CMD prior to its analysis. Instead, they plot over the cluster region a sample of surrounding field stars 0.1 arcmin away from the cluster’s radius, taken from a 0.5 arcmin concentric annulus. The lack of a proper statistical removal of contaminating foreground/background stars can cause the isochrone fit to be skewed by their presence.

As seen in the center plot of Fig. 7, these four DBs taken as a single group present a clear age-extinction bias, when compared to ASteCA values. This known degeneracy was found in Paper I to be the one with the largest correlation value (see Paper I, Table 3), meaning it is the process most likely to affect isochrone fit studies. The maximum density in this “delta” plot is located around $\Delta E_{B-V} \approx -0.07$ mag and $\Delta \log(\text{age}/\text{yr}) \approx 0.3$ dex. This trend is most obvious for P00 where a rather large average reddening value was employed, compared to the mean value found by ASteCA as stated above. For comparison purposes we show in the right plot of Fig. 7 the same delta plot, this time generated subtracting literature values from ASteCA age estimations. It can be clearly seen that afore mentioned bias is basically non-existent here, pointing to a consistent overall assignation of reddening and ages by the code.

In Appendix C we show the CMDs of each cross-matched cluster for this four DBs (153 in total). The isochrone plotted is the best fit proposed by the corresponding DB, and is compared to the best match found by ASteCA. Those clusters that present the largest age discrepancies between ASteCA and DBs values, are those where the same effect mentioned in Appendix B takes place. A good example of this is SMC-L39, as seen in Figs. C.9 and C.18 for C06 and G10 respectively.

Our age and mass estimates are also compared with three DBs – H03, R05, and P12 – which used integrated photometry to obtain these parameters (see Table 2). Only H03 and P12 obtained total mass values for the clusters in their sample, these are analyzed in Sect. 5.2.1.

figures/cross_match_ip_ages.png

Fig. 8. Age comparison for DBs that used the integrated photometry method, versus ASteCA, where N is the number of clusters cross-matched in each DB.

H03 studied approximately 1000 star clusters in both MCs – 748 in the LMC and 196 in the SMC²⁶ – via $UBVR$ integrated photometry. Ages were assigned based on the Starburst99 model (Leitherer et al. 1999), assuming metallicities, distance moduli, and average E_{B-V} values of (0.004, 0.008), (18.94, 18.48) mag, and (0.09, 0.13) mag, for the S/LMC, respectively. The masses for each cluster were derived through their absolute magnitudes M_V and the mass-luminosity relation, assuming -14.55 mag to be the M_V of a 10 Myr old $10^6 M_\odot$ cluster with $z = 0.008$. This article represents, as far as we are aware, the largest published database of MCs cluster masses to date.

R05 used two models – GALEV (Anders & Fritze-v. Alvensleben 2003) and Starburst99 – combined with three metallicities values – (0.004, 0.008) and (0.001, 0.004, 0.008), used in each model respectively – to obtain ages for 195 SMC clusters. This results in five age estimates for each cluster. Individual reddening values are obtained in the same manner as done in the Harris & Zaritsky (2004) study, assigning reddenings according to fixed age ranges. We averaged all reddening-corrected age values for each matched cluster, and assigned an error equal to the midpoint between the lowest and highest error bound among all reported ages in the article.

P12 uses the same dataset from H03 to analyze 920 LMC clusters through their MASSCLEAN_{colors} and MASSCLEAN_{age} packages (Popescu & Hanson 2010b,a). Metallicities are fixed to $z=0.008$, while the reddening is taken from G10 when available, or fixed to $E_{B-V}=0.13$ mag as done in H03. Ages and masses from duplicated entries in the P12 sample are averaged in our analysis.

As seen in the identity plot of Fig. 8, H03 visibly underestimates ages for younger clusters. In de Grijs & Anders (2006, see Fig. 1) this effect was also registered, which the authors assigned to the photometry conversion done in H03. The aver-

²⁶ To these numbers, 140 and 76 “questionable” (according to H03) S/LMC clusters respectively, can be added to their sample.

age dispersion between H03 and ASteCA logarithmic age values is 0.44 ± 0.56 dex. The same happens for P12 ages, albeit with a smaller mean dispersion in $\log(\text{age}/\text{yr})$ of approximately 0.35 ± 0.44 dex. In P12, their own age estimates are compared with those taken from H03. They find a clear systematic difference with H03 (see P12, Fig. 8), where MASSCLEAN ages are larger than H03 estimates, for clusters with $\log(\text{age}/\text{yr}) < 8$ dex. In our case, most of the clusters cross-matched with P12 are older than 8 dex. Nevertheless the same trend is confirmed, with P12 age values located below the identity line in Fig 8 – i.e.: younger ages compared to ASteCA – but still closer than those from H03. This bias towards smaller age estimates by P12 is consistent with what was found in Choudhury et al. (2015). Contrary to the results found for H03 and P12, the R05 study slightly underestimates ages compared to our values, with a mean $\log(\text{age}/\text{yr})$ dispersion of -0.25 ± 0.63 dex around the identity line. The standard deviation is the largest for the three integrated photometry DBs. In R05 the authors mention the lack of precision in their age measurements, due to the use of integrated colors, and the lack of constraints for the metallicity.

Expectedly, the four isochrone fit studies analyzed previously show a more balanced distribution of ages around the 1:1 relation, in contrast with the DBs that employ integrated photometry. Ages taken from integrated photometry studies are known to be less accurate, and should be taken as a rather coarse approximation to the true values. As can be seen in P12, integrated colors present large scatters for all age values, leading inevitably to degeneracies in the final solutions. The added noise by contaminating field stars is also a key issue, as it is very difficult to remove properly from integrated photometry data. A single overly bright field star can substantially modify the observed cluster’s luminosity, leading to incorrect estimates of its parameters (Baumgardt et al. 2013; Piatti 2014). A detailed analysis of some of the issues encountered by integrated photometry studies, and the accuracy of their results, is presented in Anders et al. (2013).

5.2.1. Integrated photometry masses

Masses are obtained in H03 and P12 via integrated photometry analysis. Baumgardt et al. (2013) also derives masses, but as stated in that article, their results are in good agreement with those from P12, so we do not add this database to our work.

There are 127 clusters in the combined H03 and P12 cross-matched samples. In Fig. 9 we show DBs masses for all cross-matched clusters, versus their logarithmic mass differences defined as:

$$\Delta M_{\log} = \log(M_{\text{ASteCA}}) - \log(M_{\text{DB}}) \quad (4)$$

where *DB* represents either H03 or P12. Crossed-matched clusters are divided into three regions, according to the masses given in either database. Sizes are scaled with each cluster’s radius in parsecs, and colors follow the difference in assigned ages $\Delta \log(\text{age}/\text{yr})$, in the sense ASteCA minus DB values, as given in the colorbar (see rightmost plot). The minimum CI value for clusters in this set is ~ 0.6 , meaning that they comprise a highly contaminated sample of clusters. The horizontal dashed line in each plot is $\overline{\Delta M_{\log}}$, the mean of the combined logarithmic differences for the assigned masses, for both DBs. The gray band shows the one standard deviation region around this mean.

The average logarithmic differences in the $M_{\text{DBs}} \leq 1000 M_{\odot}$ low-mass region – left plot in Fig. 9 – is $\sim 0.3 \pm 0.5$. This

is equivalent to a multiplicative factor of 2 between ASteCA (larger values) and DBs masses. The effect is most visible for masses $< 500 M_{\odot}$, something that had already been noticed in Sect. 5.1. The mean standard mass difference $\Delta M (= M_{\text{ASteCA}} - M_{\text{DBs}})$ is $\sim 40 \pm 1700 M_{\odot}$ in the region $M_{\text{DBs}} < 5000 M_{\odot}$, and $\sim -400 \pm 2500 M_{\odot}$ in the $M_{\text{DBs}} < 10000 M_{\odot}$ region; pointing to a very reasonable scatter around the identity line. Although the standard deviation values are somewhat large, this is nevertheless expected for a set composed of clusters in this mass range. As stated for example in Baumgardt et al. (2013) and P12, low mass clusters – those with $M \lesssim 5000$ or $10000 M_{\odot}$ – will tend to have their estimated integrated photometry masses largely dominated by stochastic processes.

A surprising systematic trend arises when we take a closer look at the mass values assigned by the DBs. It can be evidently noticed that the larger the mass estimated by either DB, the larger the logarithmic difference with ASteCA’s derived value. The left plot in Fig. 9 already shows this trend, but looking at the center and right plots, where DBs mass estimates $> 1000 M_{\odot}$ are shown, the trend becomes unmistakable. In these plots we see clusters with DB masses up to $1 \times 10^5 M_{\odot}$ with ΔM differences of almost $9 \times 10^4 M_{\odot}$. The most discrepant case of SMC cluster NGC419 had to be left out of the right plot in Fig. 9 for clarity, as it is given a mass of $3.9 \times 10^5 M_{\odot}$ by H03 and only $2.8 \times 10^4 M_{\odot}$ by the code. This cluster is one of the nine clusters – five in the LMC and four in the SMC – identified by H03 as “extreme” clusters due to their very low absolute magnitude values (see end of Sect. 4 in H03).

Values for $\overline{\Delta M_{\log}}$ in the medium and large-mass regions shown in the center and right plots of Fig. 9, are $\sim -0.3 \pm 0.4$, and $\sim -0.8 \pm 0.3$ respectively. A mass value estimated by the DBs in either region, is thus on average between 2 and 6 times larger than the same mass calculated by ASteCA. The smaller standard deviations for $\overline{\Delta M_{\log}}$ in both these regions, would seem to imply that this is not just a stochastic effect but rather a systematic one.

Considering that integrated photometry studies assign more credibility to higher mass estimates – as they will tend to be less influenced by stochastic fluctuations – this constitutes indeed an unexpected result. It is also in direct contrast with what we show in Appendix A, where the total mass estimation by ASteCA is validated using MASSCLEAN synthetic clusters. We found that the relative differences in mass – ASteCA minus MASSCLEAN – do not tend to either increase or decrease with the true mass of the cluster. The mass estimations done by ASteCA remain close to the actual cluster masses even for low values. The standard deviation of the average relative differences for the three mass regions explored (as shown in Fig. 9), actually decrease for larger masses, as expected, pointing to a more accurate recovery of the mass as the clusters analyzed grow in size.

In Table 4 we show the five clusters with the largest ASteCA-DBs mass discrepancies, $\Delta M > 2 \times 10^4 M_{\odot}$, ordered locating the ones with the larger DB masses on top. CMD plots for each one of these clusters are presented in Appendix D, along with the best match synthetic clusters generated by the code. For the most extreme cases, both DBs assign total masses that are over an order of magnitude greater than the value found by ASteCA. Ages are largely in good agreement across the two DBs and this work, for these five clusters. As mentioned previously for the SMC cluster NGC419, H03 assigns a mass that is $3.6 \times 10^5 M_{\odot}$ – or equivalently 14 times – larger than the mass obtained by the code. Similarly, P12 gives LMC cluster NGC1917 a mass

figures/cross_match_ip_mass.png

Fig. 9. *Left:* BA mass plot, showing the differences between estimated masses in the H03 and P12 DBs and the code, in the sense ASteCA minus DB; symbols as in Fig. 8. Only DB masses $\leq 5000 M_{\odot}$ are shown here. Colors are assigned according to the difference in $\log(\text{age}/\text{yr})$ estimation of each cluster (colorbar is shown in the right plot), while sizes are proportional to the actual sizes in parsecs. The gray band is the mean $\pm 1\sigma$ for the ΔM values (notice that the axis is scaled by $10^{-4} M_{\odot}$). *Center:* same as previous plot, now showing DB mass values in the range $5000 - 20000 M_{\odot}$. *Right:* same as previous plot, for DB mass values $> 20000 M_{\odot}$.

Table 4. Clusters with large differences ($\Delta M > 20000 M_{\odot}$) in their assigned ASteCA masses, versus the values found in the DBs. Equatorial coordinates are expressed in degrees for the $J2000.0$ epoch. Ages are given as $\log(\text{age}/\text{yr})$.

Cluster	$\alpha(^{\circ})$	$\delta(^{\circ})$	H03 _{age}	P12 _{age}	ASteCA _{age}	H03 _M [M_{\odot}]	P12 _M [M_{\odot}]	ASteCA _M [M_{\odot}]
S-NGC419	17.07917	-72.88417	9.31 \pm 0.12	–	8.95 \pm 0.05	$\sim 3.9 \times 10^5$	–	2.8 \pm 0.3 $\times 10^4$
L-NGC1917	79.7583	-69.001	9.48 \pm 0.09	9.46 \pm 0.08	9.15 \pm 0.08	$\sim 5.9 \times 10^4$	1 \pm 0.05 $\times 10^5$	4 \pm 1 $\times 10^3$
L-NGC1751	73.550	-69.80694	9.48 \pm 0.09	9.06 \pm 0.01	9.1 \pm 0.05	$\sim 9.7 \times 10^4$	6.5 \pm 1 $\times 10^4$	9 \pm 1 $\times 10^3$
S-L27	10.35	-72.89083	9.28 \pm 0.12	–	9.45 \pm 0.06	$\sim 5.5 \times 10^4$	–	1.3 \pm 0.4 $\times 10^4$
L-SL244	76.90417	-68.54194	9.48 \pm 0.09	9.43 \pm 0.01	9.15 \pm 0.09	$\sim 2.9 \times 10^4$	3.5 \pm 0.4 $\times 10^4$	4 \pm 1 $\times 10^3$

25 times larger – $\Delta M = 9.6 \times 10^4 M_\odot$ – than the one derived by ASteCA.

A possible explanation for this large divergence in the calculated masses, is the presence of contaminating field stars. To test this hypothesis we re-processed with ASteCA all clusters in Table 4, this time with no previous decontamination process applied. This means that all stars within the cluster region, including field stars in the same line of sight, will be employed in the obtention of the best synthetic cluster match. The upper limit for the total mass is set to $5 \times 10^5 M_\odot$, to avoid biasing the results by setting a low total mass value. For LMC cluster SL244 the total mass value obtained this way is $3 \times 10^4 M_\odot$, meaning its average DB mass is recovered when no field stars cleaning is performed. The best synthetic cluster match for SMC cluster L27 results in an even higher total mass of $1 \times 10^5 M_\odot$. These two cases clearly highlight the importance of a proper field star decontamination, before the method to derive a cluster's mass is applied.

For the remaining three clusters – S-NGC419, L-NGC1917, and L-NGC1751; the most massive according to H03 and P12 – the masses derived by ASteCA using the contaminated cluster region fall short from the values assigned by the DBs. The closest match is found for LMC clusters NGC1751 and NGC1917, for which a mass of $2 \times 10^4 M_\odot$ is estimated by the code while their average DB mass is $\sim 8 \times 10^4 M_\odot$. In the case of SMC cluster NGC419, the recovered ASteCA mass using the entirety of all the stars in its observed field is $5.5 \times 10^4 M_\odot$, still seven times smaller than the value given by H03. The radius used for NGC419 in our case is larger than the one employed in H03 by more than 20% ($\sim 85''$ versus $\sim 70''$), so we can be sure that this effect is not related to a lack of stars included within the cluster region.

These three clusters share a common extra attribute beyond being the ones with the largest assigned masses in both DBs, and showing the largest discrepancies with ASteCA mass values: they are all identified as clusters presenting the controversial “dual red clump” (dRC) structure (Girardi et al. 2009). This feature was predicted in Girardi et al. (1998) as a grouping of stars with enough mass to ignite helium, while avoiding e^- -degeneracy. The quantitative effect of the dRC structure, on the integrated magnitude of a cluster, was tested on synthetic MASSCLEAN clusters of varying masses. We found that adding a secondary RC composed of about $\sim 15\%$ of the stars present in the main RC (the approximate value found for NGC419 in Girardi et al. 2009), has a very small effect on the synthetic cluster's integrated V magnitude, as well as its (B-V) color.²⁷ For a 1 Gyr synthetic cluster of $10000 M_\odot$, adding stars to form the dRC – located ~ 0.3 mag below the RC, and ~ 0.04 mag towards the bluer region of the (B-V)-vs.-V (B-V) vs. V CMD – results in the integrated V band (and the integrated (B-V) (B-V) color) increasing only a few hundredths of a magnitude. At the same time, just doubling the mass of the synthetic cluster, i.e. $20000 M_\odot$, increases the integrated V band value almost a full magnitude. This difference is large enough to assume that the H03 and P12 models will not mistakenly assign large masses, based on such a small variance in integrated photometry as that produced by a dRC. The excess brightness generated by stars in the dRC region of these three clusters, would thus appear to not be enough to explain the over-estimated total mass values given by H03 and P12 (in particular to NGC419 by H03).

In addition to the presence of a dRC, both NGC419 and NGC1751 show extended or multiple main-sequence turnoffs (MSTO; see: Glatt et al. 2008; Milone et al. 2009;

Rubele et al. 2010, 2011; Girardi et al. 2011) while NGC1917 is known to possess a broadened MSTO (Milone et al. 2009). The origin of this structure is still under debate, as seen in Piatti & Bastian (2016), Milone et al. (2016), and Li et al. (2016). Its influence on the derived masses from integrated photometry studies is not straightforward to assess, nor can its impact be easily discarded.

If we examine the H03 and P12 databases, we find that the aforementioned bias – by which integrated photometric studies increasingly overestimate masses for larger mass clusters – exists even when comparing these studies among themselves. After removing duplicated entries in both DBs, and cross-matching them with a maximum search radius of 20 arcsec, we are left with 670 unique LMC clusters across H03 and P12. Fig. 10 shows BA diagrams for these cross-matched clusters. We plot here the average P12-H03 mass \overline{M}_{DBs} versus their logarithmic differences ΔM_{\log} , as defined in Eq. 4, in the sense P12 minus H03. As done previously, masses are separated into three regimes for clarity. There are only five clusters that show average masses larger than $100000 M_\odot$, and they are all massive globular clusters which are incorrectly assigned a low age and mass by P12.²⁸ For example, the LMC globular cluster NGC1835 is correctly identified by H03 as an old ~ 5 Gyr system, with a total mass of $\sim 1.4 \times 10^6 M_\odot$ (a reasonable value, although a bit over-estimated, according to Dubath et al. 1990). P12 on the other hand classifies this as an extremely young ~ 6.3 Myr cluster with a very low total mass estimate of $1700 M_\odot$.

In Fig. 10 we see that, as the average cluster mass given by these DBs increases, so do their logarithmic differences. The mean value ΔM_{\log} decreases from $\sim 0.8 \pm 0.5$ in the low mass region, to $\sim 0.3 \pm 0.5$ in the medium cluster mass region, to $\sim -0.4 \pm 1$ in the large mass region. Masses go from being over-estimated a factor of ~ 6 by P12 – in relation to H03 – in the low mass region, to being underestimated by a factor of ~ 2.5 in the large mass region. These differences in total mass grow with larger average masses, in a way that is closely related to the difference in the $\log(\text{age/yr})$ values estimated by each DB. Where P12 estimates larger masses than H03 it also assigns larger ages by more than 1.5 dex. Conversely, in the large average mass region, P12 ages can reach differences of up to 3 dex lower than H03. This age-mass positive correlation, by which an older large cluster can be incorrectly identified as a much younger and less massive one or vice-versa, is also noticeable in Fig. 9 albeit to a lesser extent. In Fig. 11 we show how the age-mass correlation affects DBs estimates compared to those taken from ASteCA (left plot), and how this effect is much stronger in P12 and H03 estimates (right plot).

About $\sim 17\%$ of the clusters in H03 and P12, are assigned masses above $5000 M_\odot$ in their respective DBs. A smaller percentage, less than $\sim 8\%$ and $\sim 5\%$, are assigned by those works mass values above $10000 M_\odot$ and $20000 M_\odot$ respectively. Similar proportions are found when inspecting ASteCA's obtained masses. Clusters with relatively large masses represent – as demonstrated by the aforementioned percentages – a small portion of clusters in the DBs. Nonetheless, care should be taken when applying their integrated photometry estimated masses to the study of properties such as the initial cluster mass function (ICMF). Large discrepancies in the mass value assigned for the

²⁷ Where both V and B filters correspond to the Johnson photometric system.

²⁸ These five LMC globular clusters are: NGC1916, NGC1835, NGC1786, NGC1754, and NGC1898. P12 assigns masses below $2000 M_\odot$ in all cases.

figures/H03_P12_mass.png

Fig. 10. *Left:* BA plot for the logarithmic mass differences between P12 and H03 masses, for average mass values below $1000 M_{\odot}$. Clusters are colored according to the difference in their assigned logarithmic ages by each DB – i.e.: $\Delta \log(\text{age}/\text{yr})$ – in the sense P12 minus H03; see colorbar in the right plot. The mean and standard deviation ΔM_{\log} are shown as a dashed black line and a gray region, respectively. *Center:* idem, for average DB masses $1000 < \bar{M}_{DB} < 10000 (M_{\odot})$. *Right:* idem, for average DB masses $\bar{M}_{DB} > 10000 (M_{\odot})$.

most massive clusters, could have a non-negligible impact on the slope of the IMCF.

6. Distribution of fundamental parameters in the analyzed database

We present here a brief summary of the distribution of the values in the parameters space, for the five fundamental parameters obtained with ASteCA, for our set of the 239 Magellanic Clouds clusters. The metallicity and age parameters cover a wide range in both galaxies, and their distribution can be

regarded as a representative randomized sample of the cluster system in the MCs in our catalog.

A method is devised to allow an unbiased analysis of the estimated values in the parameters space, shown in Sect. 6.1. The rest of the fitted parameters – This method is applied on the five fitted parameters, and the results presented in Sect. 6.2. The parameters reddening, distance, and total mass –, do not span a large enough range sufficiently large ranges to account for the values found in the MCs. Their distribution thus can only be though to characterize the state of those star clusters that compose this particular catalog. The metallicity and

figures/age_mass_corr.png

Fig. 11. *TopLeft:* Differences plot for ASteCA minus the combined P12 and H03 DBs. Horizontal and vertical axis show differences in $\log(\text{age}/\text{yr})$, and $\log(M/M_\odot)$, respectively. A 2-dimensional Gaussian kernel density estimate is shown as iso-density black curves. *BottomRight:* idem, for ages and masses of P12 and H03 cross-matched clusters, in the sense P12 minus H03.

age parameters, on the other hand, cover a wide range in both galaxies. Their distribution can be regarded as a representative randomized sample of the cluster system in the MCs. Their relationship is studied in Sect 6.3.

6.1. Method

Histograms are widely used to derive a large number of properties, when a substantial amount of astrophysical data is available. A galaxy’s star formation history (SFH) is a good example of such a property, almost always quantified via a one-dimensional histogram. Their widespread use notwithstanding, the generation of a histogram is affected by well known statistical issues; see Silverman (1986), Simonoff & Udina (1997). Different selected bin widths and anchor positions (point of origin) can make histograms built from the exact same data look utterly dissimilar. In the worst cases, completely spurious sub-structures may appear, leading the analysis towards erroneous conclusions. We bypass these issues by constructing an adaptive (or variable) Gaussian kernel density estimate (KDE) in one and two dimensions, using the standard deviations associated to a given parameter as the bandwidth estimates. The formulas for both KDEs are:

$$KDE_{1D}(x) = \frac{1}{N\sqrt{2\pi}} \sum_{i=1}^N \frac{1}{\sigma_i} e^{-\frac{(x-x_i)^2}{2\sigma_i^2}}, \quad (5)$$

$$KDE_{2D}(x, y) = \frac{1}{2\pi N} \sum_{i=1}^N \frac{1}{\sigma_{xi}\sigma_{yi}} e^{-\frac{1}{2}\left(\frac{(x-x_i)^2}{\sigma_{xi}^2} + \frac{(y-y_i)^2}{\sigma_{yi}^2}\right)}, \quad (6)$$

where N is the number of observed values, x_i is the i -th observed value of parameter x , and σ_{xi} its assigned standard deviation (same for y_i and σ_{yi}). The 1D version of these KDEs

is similar to the “smoothed histogram” used in the Rafelski & Zaritsky (2005) study of SMC clusters. The use of standard deviations as bandwidth estimates means that the contribution to the density map (in 1D or 2D) of parameters derived with large errors, will be smoothed (or “spread out”) over a large portion of the parameter’s domain. Precise parameter values on the other hand – i.e. those with small assigned errors – will contribute to a much more narrow region, as one would expect.

Replacing one and two-dimensional histogram analysis with these KDEs has two immediate benefits: a) it frees us from having to select an arbitrary value for the bandwidth (the most important component of a KDE, equivalent to the bin width of a regular histogram), and b) it naturally incorporates the errors obtained for each studied parameter into its derived probability density function. **Figs. 12 and 13 show 1D and 2D density maps constructed via Eqs. 5 and 6, for two paired parameters, for each of the MCs.** Being probability density functions means that the area under the curve integrates to unity. This makes the distributions for a given parameter comparable for both galaxies, even if the number of observed points – clusters in our case – is not the same (equivalent to a normalized histogram).

6.2. Distribution of parameters within the observed ranges

Figs. 12 and 13 show 1D and 2D density maps constructed via Eqs. 5 and 6, for two paired parameters, for each of the MCs.

A distinct period of cluster formation is visible in the LMC starting around the ~ 5 Gyr mark, which lasted up to ~ 1.3 Gyr ago. A similar, but much less pronounced peak is seen for the SMC, with a clear drop in cluster formation around ~ 2 Gyr. The height difference between the SMC and LMC KDEs is related to the decline in cluster formation. While the LMC sharply drops to almost zero from ~ 1 Gyr to present times, the SMC shows a much softer descent with smaller peaks around ~ 250 Myr and ~ 130 Myr. The well known “age gap” in the LMC between 3–10 Gyrs (Balbinot et al. 2010) is present, visible as a marked drop in the $KDE_{\log(\text{age}/\text{yr})}$ curve at approximately ~ 9.5 dex. No clusters older than this age are found in our work.

The 2D KDE age-metallicity map shows how spread out these values are for clusters in the SMC, compared to those in the LMC which are much more heavily clustered together. Although in this map the abundance of the SMC can be seen to reach substantially lower values than the LMC, the 1D KDE to the right reveals that the $[\text{Fe}/\text{H}]$ parameter peaks between 0 dex and -0.2 dex, for clusters in both Clouds.

The age-mass 2D map shows a clustering around younger ages and smaller masses for the LMC, relative to the SMC. The cluster seen in the bottom right corner of the age-mass SMC map is HW42 ($\alpha=1^h01^m08^s$, $\delta=-74^\circ04'25''$ [J2000.0]), a small cluster (radius < 20 pc) located close to the SMC’s center. Though its position in the map is somewhat anomalous, the 1σ error in its age and mass estimates could move it to $\log(\text{age}/\text{yr}) \approx 9.4$ and $\log(M/M_\odot) \approx 2.6$. This cluster is classified as a possible emissionless association by Bica & Schmitt (1995).

In both **clouds** there is a tendency for the mass and the size of the cluster, to grow with the estimated age, as expected. In the 1D mass KDE, we see that the LMC accumulates most clusters in the low mass regime ($\sim 3000 M_\odot$). This is also true for the SMC which has a larger proportion of large mass clusters, with a distinctive peak around $\sim 30000 M_\odot$.

As seen in Fig. 13 (top), the 1D KDEs of the distance moduli are well behaved and clearly normal in their distribution. A Gaussian fit to these curves results in best fit values of 18.96 ± 0.08 mag for the SMC, and 18.49 ± 0.08 mag for the

figures/as_kde_maps0.png

Fig. 12. One and two-dimensional Gaussian adaptive KDEs for the age, metallicity and mass parameters. Top and right plots are 1D KDEs while the center plots are 2D KDEs. Observed clusters are plotted as red and blue stars for the S/LMC, respectively in the 2D KDEs. Sizes are scaled according to each cluster's radius in parsecs. A small scatter is introduced for clarity.

LMC. The initial ranges given in Table 3, as well as the literature mean distances, are thus properly recovered. Reddening values are much more concentrated in the sample of SMC clusters around $E_{B-V} \approx 0.015$ mag. The clusters in the LMC on the other hand, show that most values are dispersed below $E_{B-V} \approx 0.1$ mag, with a shallower peak located at ~ 0.03 mag.

6.3. Age-metallicity relation

A stellar system's age-metallicity relation (AMR) is an essential tool to learn about its chemical enrichment evolution. This relation is usually presented either as scattered single points in

the age-metallicity space, or as a function created by grouping and averaging metallicity estimates in arbitrary age bins. In Piatti (2010) a method was devised to generate an AMR able to take into account the errors in age values, to produce bins of different sizes. This method has been applied to the obtention of AMRs in Piatti & Geisler (2013), and also adapted to derive star cluster frequency distributions (e.g., Piatti 2013). In this work we generate a new method based on the KDE technique described in Sect. 6, with a number of advantages over previous methods. The mathematical details are shown in Appendix E. Literature AMRs for both MCs, obtained following the same method ex-

figures/as_kde_maps1.png

Fig. 13. Same as Fig. 12 for the $E_{(B-V)}$ and distance modulus parameters.

plained in this section but using age and metallicity values from the articles in Table 1, are shown in Fig E.1 (Appendix E).

The final ASteCA AMRs for the S/LMC can be seen in Fig. 14 as red and blue continuous lines, respectively. Stars show the position of all clusters in our sample for each galaxy, with sizes scaled according to their radii. The shaded regions represent the 1σ standard deviations of the AMR functions. These regions span a $[\text{Fe}/\text{H}]$ width of approximately 0.2 dex for both Clouds, for the entire age range. The blue (top) and red (bottom) vertical segments in the top plot are the bin edges determined for each age interval by Knuth's algorithm, for the LMC and the SMC respectively. As stated in Appendix E, the final AMR functions are mostly unaffected by the chosen binning method. Using Knuth's algorithm results in approximately 12 age intervals of

widths between 0.35 and 1 Gyr, as seen in Fig. 14. If instead we use 100 intervals of ~ 0.1 Gyr width, the only substantial change is that the SMC curve is perturbed in the region $\text{Age} < 800$ Myr where, for the youngest ages, $[\text{Fe}/\text{H}]$ values are raised by ~ 0.1 dex. There are two SMC clusters with very low metal abundance estimates, $[\text{Fe}/\text{H}] < -2$ dex. Having such small metallicity values means that their associated uncertainty will be quite large, as discussed in Sect. 4. This has the effect of spreading their positions in the density map, preventing these low values from affecting the AMR substantially. If these two clusters are excluded from our data, the resulting AMR for the SMC moves upwards in the $[\text{Fe}/\text{H}]$ axis by less than 0.05 dex, for ages below 500 Myr.

Several chemical evolution models and empirically estimated AMRs can be found in the literature for both Magellanic

figures/AMR_asteca.png

Fig. 14. AStEca’s age-metallicity relation for the S/LMC (red/blue solid lines). See text in Sect. 6.3 for more details.

Clouds. Many of the studies performed on the age-metallicity relation of the S/LMC present their results as scattered points in the age vs metallicity plot, rather than fitting a unique AMR function to describe the distribution of those observed values. To allow a straightforward comparison with our AMR functions, we show in Fig. 14 – center and bottom plots – the functions presented in twelve other works.

These studies constitute a representative sample of the different methods and data used in the literature over the past twenty years, for these two galaxies: Pagel & Tautvaisiene (1998, PT98; bursting models), Geha et al. (1998, G98; closed-box model with Holtzman SFH), Harris & Zaritsky (2004, HZ04), Carrera et al. (2008b, C08a; average of four disk frames), Carrera et al. (2008a, C08b; average of thirteen frames), Harris & Zaritsky (2009, HZ09), Noël et al. (2009, N09; 5th degree polynomial fit to the AMRs of their three observed regions), Tsujimoto & Bekki (2009, TB09; 1: no merger model, 2: equal mass merger, 3: one to four merger), Rubele et al. (2012, R12; four tiles average), Cignoni et al. (2013, C13; B: Bologna, C: Cole), Piatti & Geisler (2013, PG13), and Meschin et al. (2014, M14; 0: field LMC0, 1: field LMC1, 2: field LMC2). Details on how these AMRs were constructed will not be given here, as they can be consulted in each reference. All of the above mentioned articles used field stars for the obtention of their AMRs. This is, as far as we are aware, the first work where the AMR function for both galaxies is derived entirely from observed star clusters.

The overall trend of both AMRs coincides with what has already been found in the literature, namely that the metallicity increases rather steadily with younger ages (particularly below 3 Gyr). On average, our AMR estimates are displaced slightly towards more metal rich values, compared to the majority of the literature AMRs. We must keep in mind though, that most of the $[\text{Fe}/\text{H}]$ values in the previous literature are obtained using a solar metallicity of $z_{\odot}=0.019$. Instead in this work we used the more recent value of $z_{\odot}=0.0152$. As was shown in Sect. 5.1, this dif-

ference alone means our $[\text{Fe}/\text{H}]$ estimates will be ~ 0.1 dex above the others.

For the LMC galaxy, Fig. 14 center plot, we see a marked drop in metallicity from ~ -0.45 dex beginning around 3.8 Gyr, and ending 3 Gyrs ago at ~ -0.6 dex. The M14-0 curve seems to reproduce this behavior, but shifted ~ 0.8 Gyr towards younger ages. After the aforementioned drop in the LMC’s AMR, there is a steep climb from 3 Gyr to 2 Gyr reaching almost $[\text{Fe}/\text{H}] \sim -0.3$ dex, and then a sustained but much more shallow increase up to the estimated present day’s metal content of ~ -0.15 dex. Our average metallicity value for present day clusters, coincides reasonably well with the PT98 bursting model. This AMR shows nonetheless a very different rate of increase from 2 Gyr to present times, compared to AStEca’s AMR. The C08a AMR, while lacking finer details, provides a better match for this age range. The AMR functions for the LMC that differ the most from the one obtained with AStEca values, are those taken from HZ09 and G98. These two curves are visibly separated, not only from our AMR, but also from the rest of the group.

The AMR function for the SMC obtained using AStEca’s age and metallicity values, is shown along ten AMRs taken from the published literature in Fig. 14, bottom. The peak around ~ 7.5 Gyr predicted by TB09 in its two merger models (1:1, and 1:4 merger) are not visible in our AMR. Abundances in our AMR remain largely stable around a value of $[\text{Fe}/\text{H}] \approx -0.9$ dex until approximately 3 Gyrs ago, where the rate of growth for the metallicity increases considerably. From that point up to the present day, the average metallicity for clusters in the SMC grows by about 0.4 dex, according to AStEca’s values. The increased rate of growth behavior for ages younger than 3 Gyrs, is only reproduced by the PT8 model, and the HZ04 function. The PT98 model begins diverging from our AMR at ~ 2 Gyr, until a gap of ~ 0.4 dex in metallicity is generated. In contrast, the HZ04 curve remains much closer to our own throughout the entire age range. These two AMRs estimate a present day metallicity very close to the $[\text{Fe}/\text{H}] \approx 0.4$ dex value estimated by AStEca.

Overall, our AMRs can not be explained by any single model or empirical AMR function, and are best reproduced by a combination of several of these. A similar result was found in Piatti & Geisler (2013), although their AMRs – derived from field star population – are significantly different from ours, mainly for the SMC case. It is important to remember that the AMRs estimated using the ages and metallicities derived via AStEca, are averaged over the structure of both Magellanic Clouds. In Fig. 1 we showed that our set of clusters covers a large portion of the surface of these galaxies. If more clusters were available so that the AMRs could be instead estimated by sectors in the S/LMC, it is possible that different results would arise. When clusters in our set are divided by sectors, statistically low numbers are assigned to each – particularly for the SMC – which makes this more detailed study not feasible.

7. Summary and conclusions

We presented an homogeneous catalog for a set of 239 star clusters in the Large and Small Magellanic Clouds, observed with the Washington photometric system. The clusters span a wide range in metallicity, ~~and~~ age, and ~~mass, and~~ are spatially distributed throughout both galaxies. The fundamental parameters metallicity, age, reddening, distance modulus, and total mass were determined using the AStEca package. This tool allows the automated processing of a cluster’s positional and photometric data, resulting in estimates of both its structural and intrinsic/extrinsic properties. As already shown in Paper I, the advan-

tages of using this package include reproducible and objective results, along with a proper handling of the uncertainties involved in the synthetic cluster matching process. This permits the generation of a truly homogeneous catalog of observed clusters, with their most important parameters fully recovered. Our resulting catalog is complete for all the analyzed parameters, including metallicity an mass, two properties often assumed or not obtained at all.

Internal errors show no biases present in our determination of fundamental parameters, as seen in Sect. 4. The analysis of our results in Sect. 5.1, demonstrate that the assigned values for the clusters are in good agreement with published literature which used the same Washington photometry. The metallicity parameter showed to be the most discrepant one, with ASteCA's [Fe/H] values – for both galaxies combined – being on average ~ 0.22 dex larger than those present in the literature. Half of this difference is due to the solar abundance assumed in this work ($z_{\odot}=0.0152$; Bressan et al. 2012). The explanation for the remaining ~ 0.1 dex is the confirmation bias effect, by which most cluster studies will assume the canonical [Fe/H] values rather than derive them through statistically valid means. We also compared our results with those taken from articles that used different photometric systems–, in Sect. 5.2. While the age differences in this case are somewhat larger, they can be mostly explained by effects outside the code. ~~Internal errors show no biases present in our determination of fundamental parameters.–~~

We performed in Sect. 5.2.1 a detailed comparative study of masses obtained through integrated photometry studies, with our own estimates from CMD analysis. The total mass values from the latter studies are shown to be systematically overestimated. This result is opposite to the expectation that larger clusters can have their masses recovered with moderate accuracy, via integrated photometry.

A method for deriving the distribution of any fundamental parameter – or a combination of two of them – is presented in Sect. 6.1. This method takes into account the information contained by the uncertainties, often excluded from the analysis. By relying on Gaussian kernels, it is robust and independent of ad-hoc binning choices. An age-metallicity relation is derived in Sect. 6.3 using the above mentioned method, and the parameter values obtained by ASteCA, for cluster systems in both galaxies. The AMRs generated can not be fully matched by any model or empirical determination found in the recent literature.

We demonstrated that the ASteCA package is able to produce proper estimations of observed star clusters, with their fundamental parameters covering a wide range of values. A necessary statistically valid error analysis can be performed, thanks to its built-in bootstrap error assignment method. The tool is also proven to be capable of operating almost entirely unassisted, on large databases of clusters. This is an increasingly essential feature of any astrophysical analysis tool, given the growing importance of big data and the necessity to conduct research on large astronomical data sets.

Acknowledgements. GIP would like to thank the help and assistance provided throughout the redaction of several portions of this work by: D. Hunter, A. E. Dolphin, M. Rafelski, D. Zaritsky, T. Palma, F. F. S. Maia, B. Popescu, and H. Baumgardt. This research has made use of the VizieR²⁹ catalogue access tool, operated at CDS, Strasbourg, France (Ochsenbein et al. 2000). This research has made use of “Aladin sky atlas”³⁰ developed at CDS, Strasbourg Observatory, France (Bonnarel et al. 2000; Boch & Fernique 2014). This research has made use of NASA’s Astrophysics Data System³¹. This research made use of

the Python language v2.7³² (van Rossum 1995), and the following packages: NumPy³³ (Van Der Walt et al. 2011); SciPy³⁴ (Jones et al. 2001); Astropy³⁵, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013); scikit-learn³⁶ (Pedregosa et al. 2011); matplotlib³⁷ (Hunter et al. 2007). This research made use of the Tool for OPerations on Catalogues And Tables (TOPCAT)³⁸ (Taylor 2005).

References

- Anders, P. & Fritze-v. Alvensleben, U. 2003, A&A, 401, 1063
 Anders, P., Kotulla, R., De Grijs, R., & Wicker, J. 2013, The Astrophysical Journal, 778, 138
 Andrae, R. 2010, ArXiv e-prints [arXiv:1009.2755]
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
 Balbinot, E., Santiago, B. X., Kerber, L. O., Barbuy, B., & Dias, B. M. S. 2010, MNRAS, 404, 1625
 Baumgardt, H., Parmentier, G., Anders, P., & Grebel, E. K. 2013, MNRAS, 430, 676
 Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, A&AS, 106
 Bevington, P. R. & Robinson, D. K. 2003, McGraw-Hill
 Bica, E., Bonatto, C., Dutra, C. M., & Santos, J. F. C. 2008, Monthly Notices of the Royal Astronomical Society, 389, 678
 Bica, E. L. D. & Schmitt, H. R. 1995, ApJS, 101, 41
 Bland, J. M. & Altman, D. G. 1986, Lancet, 1, 307
 Boch, T. & Fernique, P. 2014, in Astronomical Society of the Pacific Conference Series, Vol. 485, Astronomical Data Analysis Software and Systems XXIII, ed. N. Manset & P. Forshay, 277
 Bonatto, C. & Bica, E. 2007, MNRAS, 377, 1301
 Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, AAPS, 143, 33
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
 Burstein, D. & Heiles, C. 1982, AJ, 87, 1165
 Canerna, R. 1976, The Astronomical Journal, 81, 228
 Carrera, R., Gallart, C., Aparicio, A., et al. 2008a, The Astronomical Journal, 136, 1039
 Carrera, R., Gallart, C., Hardy, E., Aparicio, A., & Zinn, R. 2008b, The Astronomical Journal, 135, 836
 Chabrier, G. 2001, ApJ, 554, 1274
 Chiosi, E., Vallenari, A., Held, E. V., Rizzi, L., & Moretti, A. 2006, Astronomy and Astrophysics, 452, 179
 Choudhury, S., Subramaniam, A., & Piatti, A. E. 2015, The Astronomical Journal, 149, 52
 Cignoni, M., Cole, A. A., Tosi, M., et al. 2013, ApJ, 775, 83
 de Grijs, R. & Anders, P. 2006, MNRAS, 366, 295
 de Grijs, R. & Bono, G. 2015, The Astronomical Journal, 149, 179
 de Grijs, R., Wicker, J. E., & Bono, G. 2014, The Astronomical Journal, 147, 122
 Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871
 Dolphin, A. E. 2002, Monthly Notices of the Royal Astronomical Society, 332, 91
 Dubath, P., Meylan, G., Mayor, M., & Magain, P. 1990, A&A, 239, 142
 Elson, R. A. W., Sigurdsson, S., Davies, M., Hurley, J., & Gilmore, G. 1998, MNRAS, 300, 857
 Geha, M. C., Holtzman, J. A., Mould, J. R., et al. 1998, The Astronomical Journal, 115, 1045
 Geisler, D. 1996, AJ, 111, 480
 Geisler, D., Bica, E., Dottori, H., et al. 1997, AJ, 114, 1920
 Geisler, D., Piatti, A. E., Bica, E., & Claria, J. J. 2003, Monthly Notices of the Royal Astronomical Society, 341, 771
 Geisler, D. & Sarajedini, A. 1999, The Astronomical Journal, 117, 308
 Girardi, L., Bertelli, G., Bressan, A., et al. 2002, A&A, 391, 195
 Girardi, L., Eggenberger, P., & Miglio, A. 2011, MNRAS, 412, L103
 Girardi, L., Groenewegen, M. A. T., Weiss, A., & Salaris, M. 1998, MNRAS, 301, 149
 Girardi, L., Rubele, S., & Kerber, L. 2009, MNRAS, 394, L74
 Glatt, K., Grebel, E. K., & Koch, A. 2010, Astronomy and Astrophysics, 517, A50

³² <http://www.python.org/>

³³ <http://www.numpy.org/>

³⁴ <http://www.scipy.org/>

³⁵ <http://www.astropy.org/>

³⁶ <http://scikit-learn.org/>

³⁷ <http://matplotlib.org/>

³⁸ <http://www.starlink.ac.uk/topcat/>

²⁹ <http://vizier.u-strasbg.fr/viz-bin/VizieR>

³⁰ <http://aladin.u-strasbg.fr/>

³¹ <http://www.adsabs.harvard.edu/>

- Glatt, K., Grebel, E. K., Sabbi, E., et al. 2008, *The Astronomical Journal*, 136, 1703
- Grocholski, A. J. & Sarajedini, A. 2003, *MNRAS*, 345, 1015
- Harris, J. & Zaritsky, D. 2004, *AJ*, 127, 1531
- Harris, J. & Zaritsky, D. 2009, *The Astronomical Journal*, 138, 1243
- Haschke, R., Grebel, E. K., & Duffau, S. 2011, *The Astronomical Journal*, 141, 158
- Hills, S., von Hippel, T., Courteau, S., & Geller, A. M. 2015, *The Astronomical Journal*, 149, 94
- Hunter, D. A., Elmegreen, B. G., Dupuy, T. J., & Mortonson, M. 2003, *Astron J*, 126, 1836
- Hunter, J. D. et al. 2007, *Computing in science and engineering*, 9, 90
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open source scientific tools for Python*, [Online; accessed 2016-06-21]
- Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R.-D. 2005, *A&A*, 440, 403
- Knuth, K. H. 2006, *ArXiv Physics e-prints [physics/0605197]*
- Krouwer, J. S. 2008, *Stat Med*, 27, 778
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, 123, 3
- Li, C., Grijs Richard, d., Bastian, N., et al. 2016, *ArXiv e-prints [arXiv:1606.05394]*
- Maia, F. F. S., Corradi, W. J. B., & Santos, Jr., J. F. C. 2010, *MNRAS*, 407, 1875
- Maia, F. F. S., Piatti, A. E., & Santos, J. F. C. 2013, *Monthly Notices of the Royal Astronomical Society*, 437, 2005
- Marigo, P., Girardi, L., Bressan, A., et al. 2008, *A&A*, 482, 883
- Mateo, M. & Hodge, P. 1986, *ApJS*, 60, 893
- Meschin, I., Gallart, C., Aparicio, A., et al. 2014, *MNRAS*, 438, 1067
- Milone, A. P., Bedin, L. R., Piotto, G., & Anderson, J. 2009, *A&A*, 497, 755
- Milone, A. P., Marino, A. F., D'Antona, F., et al. 2016, *MNRAS*, 458, 4368
- Naylor, T. & Jeffries, R. D. 2006, *MNRAS*, 373, 1251
- Netopil, M., Paunzen, E., & Carraro, G. 2015, *Astronomy & Astrophysics*
- Nidever, D. L., Monachesi, A., Bell, E. F., et al. 2013, *The Astrophysical Journal*, 779, 145
- Noël, N. E. D., Aparicio, A., Gallart, C., et al. 2009, *ApJ*, 705, 1260
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, 143, 23
- Pagel, B. E. J. & Tautvaisiene, G. 1998, *MNRAS*, 299, 535
- Palma, T., Clariá, J. J., Geisler, D., Piatti, A. E., & Ahumada, A. V. 2013, *Astronomy & Astrophysics*, 555, A131
- Paunzen, E. & Netopil, M. 2006, *MNRAS*, 371, 1641
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Perren, G. I., Vázquez, R. A., & Piatti, A. E. 2015, *Astronomy & Astrophysics*, 576, A6
- Phelps, R. L., Janes, K. A., & Montgomery, K. A. 1994, *AJ*, 107, 1079
- Piatti, A. E. 2010, *Astronomy and Astrophysics*, 513, L13
- Piatti, A. E. 2011a, *Monthly Notices of the Royal Astronomical Society: Letters*, 416, L89
- Piatti, A. E. 2011b, *Monthly Notices of the Royal Astronomical Society: Letters*, 418, L40
- Piatti, A. E. 2011c, *Monthly Notices of the Royal Astronomical Society: Letters*, 418, L69
- Piatti, A. E. 2012, *Astronomy & Astrophysics*, 540, A58
- Piatti, A. E. 2013, *Monthly Notices of the Royal Astronomical Society*, 437, 1646
- Piatti, A. E. 2014, *MNRAS*, 445, 2302
- Piatti, A. E. & Bastian, N. 2016, *ArXiv e-prints [arXiv:1603.06891]*
- Piatti, A. E. & Bica, E. 2012, *Monthly Notices of the Royal Astronomical Society*, 425, 3085
- Piatti, A. E., Bica, E., Geisler, D., & Claria, J. J. 2003a, *Monthly Notices RAS*, 344, 965
- Piatti, A. E., Clariá, J. J., Bica, E., et al. 2011a, *Monthly Notices of the Royal Astronomical Society*, 417, 1559
- Piatti, A. E., Clariá, J. J., Parisi, M. C., & Ahumada, A. V. 2011b, *Publications of the Astronomical Society of the Pacific*, 123, 519
- Piatti, A. E., de Grijs, R., Ripepi, V., et al. 2015a, *MNRAS*, 454, 839
- Piatti, A. E., de Grijs, R., Rubele, S., et al. 2015b, *MNRAS*, 450, 552
- Piatti, A. E. & Geisler, D. 2013, *The Astronomical Journal*, 145, 17
- Piatti, A. E., Geisler, D., Bica, E., & Claria, J. J. 2003b, *Monthly Notices of the Royal Astronomical Society*, 343, 851
- Piatti, A. E., Geisler, D., Sarajedini, A., & Gallart, C. 2009, *Astronomy and Astrophysics*, 501, 585
- Piatti, A. E., Geisler, D., Sarajedini, A., Gallart, C., & Wischnjewsky, M. 2008, *Monthly Notices of the Royal Astronomical Society*, 389, 429
- Piatti, A. E., Sarajedini, A., Geisler, D., Clark, D., & Seguel, J. 2007a, *Monthly Notices of the Royal Astronomical Society*, 377, 300
- Piatti, A. E., Sarajedini, A., Geisler, D., Gallart, C., & Wischnjewsky, M. 2007b, *Monthly Notices of the Royal Astronomical Society*, 382, 1203
- Piatti, A. E., Sarajedini, A., Geisler, D., Gallart, C., & Wischnjewsky, M. 2007c, *Monthly Notices of the Royal Astronomical Society: Letters*, 381, L84
- Piatti, A. E., Sarajedini, A., Geisler, D., Seguel, J., & Clark, D. 2005, *Monthly Notices of the Royal Astronomical Society*, 358, 1215
- Pietrzynski, G. & Udalski, A. 1999, *AcA*, 49, 157
- Pietrzynski, G. & Udalski, A. 2000, *AcA*, 50, 337
- Popescu, B. & Hanson, M. M. 2010a, *ApJ*, 724, 296
- Popescu, B. & Hanson, M. M. 2010b, *ApJ*, 713, L21
- Popescu, B., Hanson, M. M., & Elmegreen, B. G. 2012, *ApJ*, 751, 122
- Rafelski, M. & Zaritsky, D. 2005, *Astron J*, 129, 2701
- Rubele, S., Girardi, L., Kozhurina-Platais, V., Goudfrooij, P., & Kerber, L. 2011, *MNRAS*, 414, 2204
- Rubele, S., Kerber, L., & Girardi, L. 2010, *MNRAS*, 403, 1156
- Rubele, S., Kerber, L., Girardi, L., et al. 2012, *A&A*, 537, A106
- Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013, *ApJ*, 764, 167
- Schlafly, E. F. & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schmeja, S., Kharchenko, N. V., Piskunov, A. E., et al. 2014, *A&A*, 568, A51
- Scott, D. W. 1979, *Biometrika*, 66, 605
- Scowcroft, V., Freedman, W. L., Madore, B. F., et al. 2015, *ArXiv e-prints [arXiv:1502.06995]*
- Silverman, B. 1986, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Taylor & Francis)
- Simonoff, J. S. & Udina, F. 1997, *Computational Statistics & Data Analysis*, 23, 335
- Sollima, A., Carballo-Bello, J. A., Beccari, G., et al. 2010, *MNRAS*, 401, 577
- Subramanian, S. & Subramanian, A. 2009, *A&A*, 496, 399
- Tammann, G. A., Sandage, A., & Reindl, B. 2003, *A&A*, 404, 423
- Taylor, M. B. 2005, in *Astronomical Society of the Pacific Conference Series*, Vol. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Tsujimoto, T. & Bekki, K. 2009, *The Astrophysical Journal Letters*, 700, L69
- Udalski, A., Soszynski, I., Szymanski, M., et al. 1999, *Acta Astron.*, 49, 223
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science & Engineering*, 13, 22
- van Rossum, G. 1995, *Python tutorial*, Report CS-R9526, pub-CWI, pub-CWI:adr
- Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in *Conference on Intelligent Data Understanding (CIDU)*, 47–54
- von Hippel, T. 2005, *ApJ*, 622, 565
- von Hippel, T., van Dyk, D. A., Stenning, D. C., et al. 2014, in *EAS Publications Series*, Vol. 65, *EAS Publications Series*, 267–287

Appendix A: Total cluster mass validation

The likelihood used in this work, presented in Eq. 1, allows us to set the total cluster mass as a free parameter to be optimized, i.e., treat it the same way we treat the metallicity, age, reddening, and distance parameters. To validate the masses recovered by ASteCA we processed 768 synthetic clusters generated with the MASSCLEAN tool, 384 for each Magellanic Cloud. These synthetic clusters were created to imitate the metallicity, age, and mass range that clusters in both Magellanic Clouds have, while the distance and reddening parameters were fixed. In Table A.1 these values are shown for the five fundamental parameters fitted. The process of generating a synthetic cluster with MASSCLEAN was described in Paper I, Sect. 3. Each MASSCLEAN cluster had its V vs $(B - V)$ CMD (in the $UBRIJHK$ photometric system) analyzed by the code.

In Fig. A.1 we show the results from the best match process applied on the 768 MASSCLEAN synthetic cluster, separated into three mass regions. The **x-axis** displays the true MASSCLEAN mass values. The **y-axis** shows the logarithmic mass differences defined in Eq. 4, in the sense ASteCA minus MASSCLEAN. Colors are associated to the differences in $\log(\text{age}/\text{yr})$ (ASteCA minus MASSCLEAN values), according to the colorbar shown to the bottom right of the right plot. The average age difference for each mass region defined is: -0.3 ± 0.6 dex ($\bar{M} \leq 1000 [M_{\odot}]$), -0.05 ± 0.19 dex ($1000 < \bar{M} \leq 10000 [M_{\odot}]$), -0.01 ± 0.13 dex ($\bar{M} > 10000 [M_{\odot}]$). As expected, cluster with larger masses are assigned more accurate ages by the code. On average, the difference between ASteCA (estimated) minus MASSCLEAN (true) logarithmic ages in the full mass range analyzed is of $\sim -0.1 \pm 0.4$ dex. The gray bands represent the mean and standard deviation for the logarithmic mass differences, ΔM_{\log} , as estimated versus true masses. For each mass region defined above these values are: -0.13 ± 0.24 , -0.01 ± 0.15 , 0.01 ± 0.09 , for the combined set of synthetic clusters belonging to both galaxies. There is no visible trend, other than a larger average difference and dispersion in the estimated mass values, for clusters with low masses ($< 1000 M_{\odot}$). For this set of synthetic clusters – where the total true mass is either $500 M_{\odot}$ or $1000 M_{\odot}$ – the code assigns masses in a range between ~ 200 and $3000 M_{\odot}$. The ΔM_{\log} value obtained for this low mass region means that ASteCA is underestimating these clusters' masses by approximately $\sim 200 M_{\odot}$. This effect is tied to an improper age estimation, where ASteCA incorrectly assigns younger ages to scarcely populated clusters, and compensates the low number of cluster stars by decreasing also the total mass. Such an issue is not unexpected for very low mass clusters.

Table A.2 shows the correlation matrix between the five cluster parameters matched. We see the usual correlations appear (age-metallicity, metallicity-distance, age-reddening, etc.), as found in Table 3 of Paper I. The total mass shows no correlations, except for a very small positive one with the distance modulus. This arises because if the distance is overestimated, low mass stars will be lost in the synthetic CMD due to the maximum magnitude limit of the observed CMD. The likelihood will thus compensate this loss, by increasing the mass in the synthetic CMD matched.

Appendix A.1: Metallicity estimation for different mass values

The metallicity (z) estimated for the entire set is displayed in Fig. A.2, for each mass value used in the generation of the synthetic MASSCLEAN clusters. There are two clear tenden-

cies visible in these plots. First we notice that, as the cluster's mass grows, so does the accuracy of the metallicity estimates. Although the average difference between true and estimated z values remains close to $\bar{\Delta z} \approx 0.001$ ³⁹ for the entire range, its standard deviation drops from ~ 0.01 to 0.004 for the more massive clusters. Almost all of the poorest solutions obtained by ASteCA – those with $|\Delta z| > 0.01$ dex – are associated to low mass scarcely populated clusters. These synthetic clusters can contain as little as 2 true member stars (up to a maximum of 100), with an average of ~ 40 present in their analyzed CMDs. This poor solutions set is composed of 91 synthetic clusters – equivalent to $\sim 12\%$ of the entire processed sample – 58 of which are of low mass ($\leq 1000 M_{\odot}$). The majority of these low mass clusters – 38 out of the 58 – are assigned younger ages by the code, due to an improper field star decontamination process (an expected issue when the number of true members is very low). If we leave out those clusters with a visibly wrong age estimation – i.e., those with $|\Delta \log(\text{age}/\text{yr})| \geq 0.5$ dex – the average difference in z for the entire mass range is found to be $\sim 0.0008 \pm 0.006$ dex; a rather small difference with reasonable dispersion. These subset of clusters with wrong ages estimated by ASteCA represent approximately $\sim 11\%$ of the combined S/LMC MASSCLEAN sample (82 synthetic clusters), and is almost entirely composed of clusters with $M \leq 1000 M_{\odot}$ (73 synthetic clusters).

The second visible tendency is that the code seems to overestimate the metal content for lower metallicities, and overestimate it for the largest ones. A balanced distribution around the $\Delta z = 0$ line is mostly seen for the abundances in the middle portion of the analyzed range. This trend is much more noticeable for lower masses, but can be found for all the mass values shown. This is a statistical artifact that arises due to the necessarily limited metallicity range analyzed by the code. For the synthetic clusters with the lowest metal contents ($z = 0.001$), ASteCA can only assign equal or larger metallicities since negative z values are not possible. Equivalently, for the clusters with the largest abundances ($z = 0.03$) the code can only associate equal or lower metallicities. This is because of the upper limit used in the z range analyzed by ASteCA, which is precisely $z = 0.03$. This “bias” could be avoided for those large metallicity clusters, by increasing the upper limit for the z range in the code. It can not be avoided for the lowest metal abundances.

An external source of errors also needs to be taken into account when analyzing ASteCA's metallicity (and age) estimates, for MASSCLEAN synthetic clusters. This is the intrinsic differences between the Marigo et al. (2008) isochrones – used to generate the synthetic MASSCLEAN clusters – and the PARSEC (Bressan et al. 2012) isochrones – used by the code to find the optimal fundamental parameters. These differences are a source of error in the matching process that is not straightforward to quantify. The two sets of tracks have non-negligible dissimilarities beyond the turn-off points, for most of the age range where they can be produced. This can be seen in Fig. A.3, where isochrones from both sets are compared for five different $\log(\text{age}/\text{yr})$ ages from 7.5 to 9.5 dex. For ages up to 8 dex the PARSEC isochrones present a turn-off point located at lower $\log(L/L_{\odot})$ values, particularly for lower metallicities. This causes a shift in the more evolved portions of the isochrone, displacing the Marigo isochrones towards larger $\log(L/L_{\odot})$ values. Beyond that age this effects reverses, and the PARSEC isochrones are now lifted above the Marigo tracks. Given the

³⁹ This is expected, as $z = 0.001$ is the metallicity step used when processing the synthetic clusters with ASteCA.

Table A.1. Parameters values used to generate the set of 768 MASSCLEAN synthetic clusters.

Parameter	Values	N
z	0.001, 0.004, 0.015, 0.03	4
$\log(\text{age/yr})$	7, 7.2, 7.5, 7.7, 8, 8.2, 8.5, 8.7, 9, 9.2, 9.5, 9.7	12
μ	18.9 (SMC), 18.5 (LMC)	2
E_{B-V}	0.1	1
Mass (M_{\odot})	500, 1000, 5000, 10000, 25000, 50000, 100000, 250000	8

Table A.2. Correlation matrix for parameter deltas, defined for each cluster in the sense AStECa minus MASSCLEAN.

$\Delta param$	Δz	$\Delta \log(\text{age/yr})$	$\Delta \mu$	ΔE_{B-V}	ΔM
Δz	1.	-0.36	0.24	-0.15	0.03
$\Delta \log(\text{age/yr})$	–	1.	-0.15	-0.28	0.01
$\Delta \mu$	–	–	1.	0.05	0.13
ΔE_{B-V}	–	–	–	1.	0.0
ΔM	–	–	–	–	1.

Table B.1. Clusters with large differences in their assigned literature ages versus the values found by the code (“outliers”). Equatorial coordinates are expressed in degrees for the J2000.0 epoch. Ages are given as $\log(\text{age/yr})$ for literature (L) and AStECa (A). The difference between both estimates (L-A) is given in the last column as Δ .

Cluster	$\alpha(^{\circ})$	$\delta(^{\circ})$	L	A	Δ
L-KMHK975	82.49583	-67.87889	8.30	6.70	1.60
L-SL579	83.55417	-67.85639	8.15	7.00	1.15
L-BSDL631	76.64167	-68.42722	8.35	7.50	0.85
L-KMHK979	82.41250	-70.98389	7.90	7.30	0.60
L-H88-316	85.41250	-69.22944	8.25	7.70	0.55
S-L35	12.00417	-73.48611	8.34	6.90	1.44
S-H86-188	15.05833	-72.45833	8.10	6.70	1.40
S-L39	12.32500	-73.37167	8.05	7.00	1.05
S-B134	17.25417	-73.20667	8.15	7.20	0.95
S-K47	15.79583	-72.27361	7.90	7.00	0.90

many known correlations between fundamental parameters (e.g., the age-metallicity, reddening-distance, and distance-metallicity relations), it is not easy to predict how the matching algorithm will resolve such instances.

Appendix B: Outliers

Ten of the analyzed clusters in this work – approximately 4% of the total 239 present in the set – show a difference in age with literature values of $\Delta \log(\text{age/yr}) > 0.5$. Such a large age difference translates into two very dissimilar isochrones fitted to the same observed coeval star sequence, which makes this sub-sample of star clusters stand out from the rest. For these “outliers” no configuration of the DA plus the employed binning methods could be found, that resulted in synthetic CMD matches with age values close to those found in the literature. Five of these clusters belong to the LMC and the remaining five to the SMC, as shown in Table B.1. All the clusters in the outliers sample had smaller ages assigned by the code, compared to their literature values. These differences go from 0.55 dex up to 1.6 dex in the most extreme case of the LMC cluster KMHK975.

In Fig. B.1 the CMDs for these ten clusters are plotted in pairs, two per row. The left CMD in each pair shows the cluster region with the literature isochrone fit shown in red. For nine of the clusters we fit Marigo et al. (2008) isochrones, since this set of theoretical isochrones was used in all but one of the articles

where these clusters were analyzed. The exception is SMC-L35 where we use Girardi et al. (2002) isochrones, as done in the literature. The right CMD shows the same cluster region with the isochrone that generated the best synthetic CMD match found by AStECa, shown in green. The colors of the stars in this CMD correspond to the MPs assigned by the DA, when applied. Semi-transparent stars are those removed by the cell-by-cell density based cleaning algorithm (see Sect.3.3), also when applied. Values for the fundamental parameters in both fits are shown to the bottom left of each CMD.

For almost all of the outliers, the same process can be identified as the main cause responsible for the observed age differences. While the by-eye isochrone fit done in the literature aligns the brighter part of the cluster’s sequence with the turn off point of an older isochrone, AStECa decides instead that this is the top portion of a much younger isochrone with no discernible turn off. The statistical mismatch due to the removal of low mass stars by the DA – discussed in Sect. 5.1 – can also be seen to affect some of the fits here. In particular the SMC clusters SL579 and H86-188 show signs of this effect in the best synthetic CMD match selected by AStECa (see isochrone fits in Fig. B.1, CMD diagrams *b* and *h*).

These age estimates could be improved – in the sense that they could be brought closer to literature values – if a more restrictive age range was used (e.g.: a minimum value of $\log(\text{age/yr})=7.5$ dex instead of 6 dex as used in this work, see Table 3). Lacking external evidence to substantiate this a priori restriction, we choose to keep the values obtained by AStECa, with this section acting as a cautionary note.

Appendix C: Color-magnitude diagrams for the P99, P00, C06, and G10 databases

This appendix presents the CMDs of all clusters cross-matched with our own sample, in the databases P99, P00, C06, and G10 – i.e.: those that used the isochrone fit method in their analysis – in Figs. C.1 to C.18.

Each CMD pair in the figures below shows a given cluster with the fit proposed by the corresponding database to the left (red isochrone), and the one estimated by AStECa to the right (green isochrone). The coloring of the stars in the CMD to the right, correspond to the weights assigned by the decontamination algorithm. Stars drawn semi-transparent were discarded by the

figures/massclean_mass.png

Fig. A.1. Recovered masses by ASteCA for the set of 768 MASSCLEAN synthetic clusters, separated in three mass ranges. Logarithmic mass differences ΔM_{\log} are obtained in the sense ASteCA minus MASSCLEAN (i.e., estimated minus true values), and shown in the ~~y-axis~~y-axis. MASSCLEAN masses in the ~~x-axis~~x-axis are perturbed with a small random scatter to improve visibility.

cell-by-cell cleaning method, and thus not used in the synthetic match process.

figures/massclean_z.png

Fig. A.2. ASteCA metallicity estimates for each mass used to generate the synthetic MASSCLEAN clusters. Colors are associated to the logarithmic age differences, shown in the colorbars to the right. The green dashed horizontal line is the $\Delta[Fe/H]=0$ line, shown as reference.

figures/mar_vs_par_isocho.png

Fig. A.3. Marigo et al. (2008) versus PARSEC (Bressan et al. 2012) isochrones, for different metallicities and ages.

figures/outliers_VS_asteca_0.png

Fig. B.1. CMDs for the outliers set. See description of the plots in the main text of the section.

figures/DB_fit/P99_VS_asteca_0.png

Fig. C.1. CMDs for the P99 database.

figures/DB_fit/P99_VS_asteca_1.png

Fig. C.2. CMDs for the P99 database.

figures/DB_fit/P00_VS_asteca_0.png

Fig. C.3. CMDs for the P00 database.

figures/DB_fit/P00_VS_asteca_1.png

Fig. C.4. CMDs for the P00 database.

figures/DB_fit/P00_VS_asteca_2.png

Fig. C.5. CMDs for the P00 database.

figures/DB_fit/C06_VS_asteca_0.png

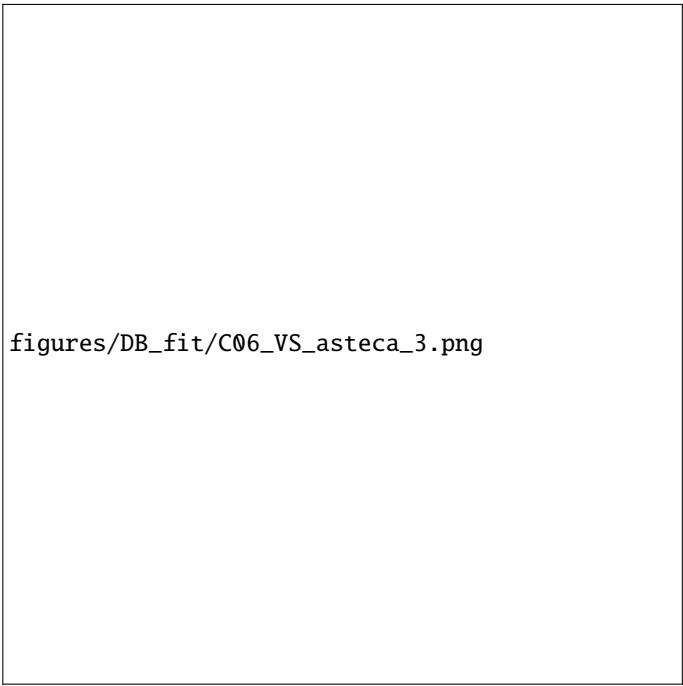
Fig. C.6. CMDs for the C06 database.

figures/DB_fit/C06_VS_asteca_1.png

Fig. C.7. CMDs for the C06 database.

figures/DB_fit/C06_VS_asteca_2.png

Fig. C.8. CMDs for the C06 database.



figures/DB_fit/C06_VS_asteca_3.png

Fig. C.9. CMDs for the C06 database.

figures/DB_fit/G10_VS_asteca_0.png

Fig. C.10. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_1.png

Fig. C.11. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_2.png

Fig. C.12. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_3.png

Fig. C.13. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_4.png

Fig. C.14. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_5.png

Fig. C.15. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_6.png

Fig. C.16. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_7.png

Fig. C.17. CMDs for the G10 database.

figures/DB_fit/G10_VS_asteca_8.png

Fig. C.18. CMDs for the G10 database.

Appendix D: Color-magnitude diagrams for large mass clusters

We present in Fig D.1 CMDs for the five clusters with the largest mass estimates by either the H03 or the P12 databases. These are also the clusters with the largest differences in assigned masses when compared with those derived by ASteCA. See Table 4 for more information about the fundamental parameters assigned to these clusters.

Appendix E: Mathematical description of the age-metallicity relationship

We can take advantage of the KDE technique described in Sect. 6 – used to produce two-parameters density maps – to generate an AMR that is truly representative of the observed data, with some important improvements over previous methods. First, a Gaussian density map has no dependence on the number, size (fixed or variable) or location of bins, as a regular histogram or the aforementioned method would. Second, the errors in the two parameters used to obtain the density map (age and metallicity), are organically included in the function that generates it (as explained in Sect. 6). This means that no ad-hoc procedure is required to incorporate the important information carried by these values, into the final AMR.

The process of creating an AMR function⁴⁰ requires that we associate a unique $[Fe/H]$ to a single age value, for the available age range. After generating the age-metallicity 2D density map, a dense 2D grid is created dividing it into N steps of 0.01 dex width, covering the ranges of both parameters. Every point in this grid is evaluated in the KDE map and its value (w_i) is stored, along with its age-metallicity coordinates (age_i , $[Fe/H]_i$). Each of the N ages defined in the grid is then associated to a single representative $[Fe/H]$ value. This representative metallicity for a given age is obtained as the mean metallicity value, weighted by

the KDE function at that particular age. The formal equation for obtaining it, can be written as

$$\overline{[Fe/H]}_{age_i} = \frac{\sum w_i [Fe/H]_i}{\sum w_i} \quad (E.1)$$

where the summations are performed over each i step for the N defined steps in the metallicity range, $[Fe/H]_i$ is the metallicity value at step i , and w_i is the value of the 2D KDE map for that fixed age and metallicity coordinates. The age_i subindex in Eq. E.1 indicates that this mean metallicity was calculated for a fixed age value, and thus represents a unique point in the AMR. A similar version of this weighted metallicity was employed in Noël et al. (2009, see Eq. 3) to derive AMR estimates for three observed fields. We apply the above formula to all the N ages in the grid defined at the beginning of the process. The standard deviation for each $\overline{[Fe/H]}_{age_i}$ value is calculated through the equation

$$\sigma_{age_i}^2 = \frac{\sum w_i \sum [w_i ([Fe/H]_i - \overline{[Fe/H]}_{age_i})^2]}{(\sum w_i)^2 - \sum w_i^2} \quad (E.2)$$

where again all summations are applied over N , and the descriptions given for the parameters in Eq. E.1 apply. At this point, this method already gives us an AMR function estimate, since every age step is mapped to a unique metallicity. The downsides are that the AMR is noisy due to the very small step of 0.01 dex used, and that the associated errors are quite large. This latter effect arises because the weighted standard deviation, Eq. E.2, will be affected not only by the errors in both measured parameters – through the constructed 2D KDE map – but also by the intrinsic dispersion in the metallicity values found for any given age. To solve this, we calculate the average $[Fe/H]$ for a given age interval, rather than assigning a metallicity value to each age step in the grid. Dividing the age range into intervals requires a decision about the step width, much like when constructing a histogram, bringing back the issue of binning. We have two advantages here: a) we use Knuth’s algorithm (see Sect. 3.4) to obtain the optimal binning for our data, and b) the final AMR function is very robust to changes in the binning method selected, so even the previous choice is not crucial in determining the shape of our AMR. Finally, the $\overline{[Fe/H]}_{age_i}$ values obtained for every age_i within a defined age interval, are averaged. Errors are propagated through the standard formula, disregarding covariant terms (Bevington & Robinson 2003, Eq. 3.14).

In Fig. E.1 we show the AMRs for both Clouds, generated using the metallicity and age values taken from those articles presented in Table. 1, i.e.: the “literature”. Most clusters in these works are assigned fixed metallicities of -0.7 dex and -0.4 dex for the S/LMC, respectively; particularly for estimated ages below 1 Gyr. This explains the average difference of ~ 0.2 dex that can be appreciated, when compared with the AMRs shown in Fig. 14.

⁴⁰ By “AMR function” we mean a curve that spans the entire observed age range, mapping each age value to a single metallicity value.

figures/largemass_VS_asteca_0.png

Fig. D.1. CMDs for the clusters with the largest masses assigned by H03 and/or P12. The best match synthetic cluster is plotted to the right, and the observed cluster region CMD to the left, for each cluster.

figures/AMR_literature.png

Fig. E.1. Age-metallicity relationships for our set of 239 clusters, using $\log(\text{age/yr})$ and $[\text{Fe}/\text{H}]$ values taken from the literature.