Boletim da Sociedade Astronômica Brasileira, **35**, no. 1, 24-28

Characterization of star clusters observed by the VISCACHA Survey using synthetic stellar populations

Bernardo P. L. Ferreira[1], João Francisco C. Santos Jr.[1], Bruno Dias[2], Fransciso Gardin de Carvalho[1], Francisco Ferreira de Souza Maia[3]

[1] DF-ICEx-UFMG e-mail: plfbernardo@gmail.com,jsantos@fisica.ufmg.br,joaofranciscogardin12@gmail.com
[2] IA-FCE-UNAB/Chile e-mail: astro.bdias@gmail.com
[3] IF-UFRJ e-mail: ffsmaia@if.ufrj.br

Abstract. Isochrone fitting is a key aspect of the characterization of a star cluster, allowing for the determination of its age, metallicity, distance and color excess. We present the SIESTA code: a new, computational technique for fitting isochrones based on the statistical comparison of the distribution of cluster stars in the color-magnitude diagram with distributions generated for synthetic populations with different parameters. We apply this technique for four stars clusters from the Small Magellanic Cloud observed by the VISCACHA Survey.

Resumo. O ajuste de isócronas é um aspecto essencial da caracterização de um aglomerado estelar, permitindo a determinação da sua idade, metalicidade, distância e excesso de cor. Nós apresentamos o código SIESTA: uma nova técnica computaciolaç para ajustar isócronas, baseada na comparação da distribuição de estrelas de um aglomerado no diagrama cor magnitude com distribuições geradas para populações sintéticas de diferentes parâmetros. Nós aplicamos esta técnica para quatro aglomerados da Pequena Nuvem de Magalhães observados pelo levantamento VISCACHA.

Keywords. Galaxies: star clusters: general – Galaxies: Magellanic Clouds

1. Introduction

One of the appeals of studying star clusters is that we can obtain a great amount of information about them using only photometric measurements of their stars. By fitting isochrones to the distribution of the stars from a cluster in the Color-Magnitude Diagram (CMD), we can infer its age, metallicity, distance, and color excess (e.g., Dias et al. 2014). After this critical step, other aspects of the star cluster can be analyzed such as its total mass and present-day mass function (e.g., Maia et al. 2014), or its dynamical evolution (e.g., Angelo et al. 2023).

The wealth of information that can be extracted from star clusters makes these objects relevant for better understanding the properties of the galaxies where they are inserted, with examples both in the Milky Way (e.g., Castro-Ginard et al. 2021) and in other galaxies (e.g., Weidner et al. 2004). The VISCACHA Survey[1] (Maia et al. 2019) is an example of the latter, in the context of the Magellanic Clouds. It consists, mainly, of a photometric survey of star clusters in the outskirts of these galaxies and in the Magellanic Bridge, performed using the SAM (SOAR Adaptive Module) instrument, attached to the 4.1m SOAR telescope (SOuthern Astrophysical Research). Its goal, as described in Maia et al. (2019), is to use star clusters to better understand the structure and evolution of the Magellanic System. To achieve that, isochrone fitting is a key aspect of the clusters' analysis.

To perform isochrone fitting, it is common in the literature for authors to use "chi-by-eye" approaches, where the parameter space is explored visually to find a good match between models and observations (e.g., Ferreira et al. 2019; Saroon et al. 2023). Although good fits can be obtained from this type of analysis, it also unavoidably leads to subjective criteria for fine-tuning the best solution and defining uncertainties.

To account for that, several authors started using statistical methods for isochrone fitting, where the parameter space is explored iteratively, and models are mathematically compared with the data to determine the best parameters along with the corresponding uncertainties. Examples of said efforts are the codes ASTECA (Perren et al. 2015), BASE-9 (Robinson et al. 2016), and SIRIUS (Souza et al. 2020).

In this work, we present a new code for performing isochrone fitting, named SIESTA, which stands for Statistical matchIng between rEal and Synthetic STellar populAtions. The SIESTA code is based on the comparison of the observed CMD from a given star cluster with CMDs coming from synthetic populations, in a Bayesian approach, to determine its age, metallicity, distance, color excess, and binary fraction.

We employ this technique to characterize four stellar clusters of the Small Magellanic Cloud (SMC) from the VISCACHA Survey, namely, NGC 152, Lindsay 91, Lindsay 113, and NGC 121. These clusters were chosen for having a wide range of ages, spanning from 1 to 10 Gyr, which makes them a good sample for evaluating the performance of the SIESTA code. They were also homogeneously analyzed by Milone et al. (2023), using data from the Hubble Space Telescope (HST), providing an independent source to compare our results. The four clusters have also been analyzed in previous VISCACHA papers using different techniques. Lindsay 91, and Lindsay 113 were characterized in Oliveira et al. (2023) with the latest version of the SIRIUS code (Souza et al. 2020). NGC 152 was characterized by Dias et al. (2022) using a previous version of SIRIUS. Finally, NGC 121, which is an extensively studied cluster in the SMC, was characterized by Saroon et al. (2023) using visual isochrone fitting.

2. Data

The four clusters were observed by the VISCACHA team in the VI bands using the SAM imager (Tokovinin et al. 2016):

[1] VIsible Soar photometry of star Clusters in tApii and Coxi HuguA

a ground layer, adaptive optics imager in the SOAR telescope, with a 4kx4k CCD detector with 2x2 binning, plate scale of 0.091 $arcsec \cdot pixel^{-1}$ and field of view of $3.1 \times 3.1 arcmin^2$. Due to the adaptive optics performed by SAM, the PSF's full width at half maximum is usually below 0.75". Data reduction and photometry were performed following the methodology presented in Maia et al. (2019). The cluster CMDs were also decontaminated from the contribution of field stars using the method proposed in Maia et al. (2010).

3. The SIESTA code

The SIESTA code characterizes a stellar population by comparing the distribution of observed stars in the CMD with the ones obtained from synthetic populations, generated from a grid of isochrones. The adopted grid comes from the PARSEC-COLIBRIv3.7 database (Bressan 2012; Marigo et al. 2017), the same that was used in Milone et al. (2023), with ages within the range of $6.00 \leq \log Age \leq 10.15$ and metallicities within $-2.19 \leq [M/H] \leq 0.00$, both in steps of $0.01 dex$.

To characterize a given cluster, we must construct its Hess diagram, which is, in summary, a map of the stellar distribution in the CMD. To do that, we use two-dimensional histograms with bin widths chosen to constrain critical features of the CMD, such as the Main Sequence turn-off and/or the Red Clump. Once the binning is defined, we evaluate the 2D histogram of the cluster CMD. To incorporate the photometric uncertainties in the Hess diagrams, we reevaluate this histogram 10,000 times replacing the original data with the stellar magnitudes added with some random Gaussian noise proportional to the observed photometric error. The final Hess diagram is the average of all previous evaluations.

3.1. Generating synthetic populations

The synthetic populations are generated dynamically during the fitting process. The procedures, which we describe below, are illustrated on Figure 1. For a given pair of age and metallicity, an isochrone is selected from the grid and then displaced in the CMD according to a given distance and color excess. The isochrone is then realistically sampled using the initial mass function from Kroupa (2001).

After that, a fraction of the population is converted into non-resolved binary systems, by adding to its flux the contribution of another synthetic star, also sampled from the isochrone. To better emulate the effect of non-resolved systems in the CMD, the secondary star is sampled uniformly, from a mass range of $[qM_1, M_1]$, where M_1 is the mass of the original star and q is a free parameter that must be lesser or equal than 1. By setting a minimum mass for the secondary star in the binary systems, we ensure that their colors and magnitudes will be significantly different than the flux of the original star. This is a common procedure in other methods that rely on synthetic populations for their characterizations (e.g., Kerber et al. 2007; Perren et al. 2015). The main difference is that, in these procedures, the binary fraction is usually kept fixed, while in SIESTA it is fitted together with the age, metallicity, distance, and color excess.

After the binary addition, some of the synthetic stars are randomly removed from the sample, using a completeness function fitted for each cluster. These functions were built empirically, by comparing the number of stars per magnitude in the I band of our observations with the more complete catalogs from Milone et al. (2023), that used HST data. Finally, Gaussian noise is added to the magnitudes of the remaining synthetic stars, according to a photometric error function, which was also fitted for each cluster analyzed, using the relation between the photometric uncertainty and the magnitude obtained from the observations of each star. The noise is truncated at 1 standard deviation, as this better reproduces the spread observed in the data.

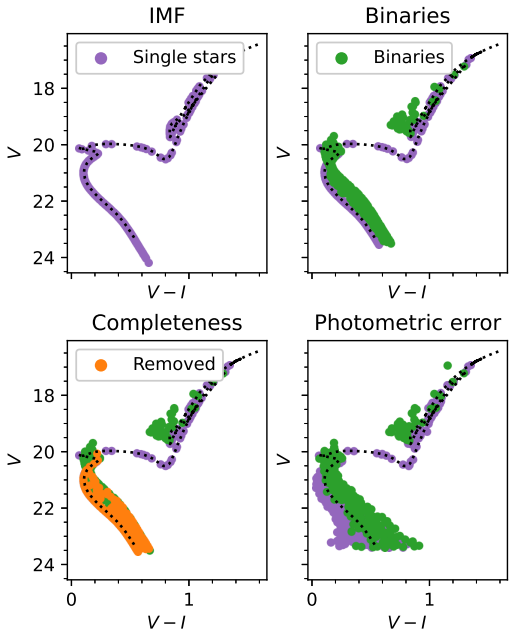

Figure 1. Process for generating a synthetic CMD from an isochrone (age of 1.5 Gyr, metallicity of -0.75, distance of 63 kpc, and color excess of 0.01). The top left plot shows the stars sampled using Kroupa (2001)'s IMF. The top right shows the CMD after the conversion of a fraction of the population into non-resolved binaries. The bottom left plot highlights the stars randomly removed to emulate the photometric incompleteness of the data. Then, the bottom right plot shows the population after the addition of noise in the magnitudes, to emulate the photometric errors. The synthetic CMD has 3000 synthetic stars with a binary fraction of 15%.

The Hess diagram is then constructed for the synthetic population, using the same bins defined for the observed data. Since the process of generating a synthetic population is intrinsically stochastic, we use a large number of synthetic stars (of the order of 10^4) and then renormalize the number of counts to get the average number of objects in a given bin. We also repeat the noise addition in the synthetic data, to emulate the smoothing process applied in the observations.

3.2. Fitting

SIESTA uses a Bayesian approach to find the combination of parameters that creates the synthetic population that is most representative of the observed CMD of a given star cluster. This strategy is grounded on the Bayes Theorem, which, in this context, can be written as:

$$p(params|CMD) \propto p(params) \times p(CMD|params). \quad (1)$$

$p(params|CMD)$ is what we call the *posterior* distribution: the probability of the star cluster we want to analyze having a particular set of physical parameters, conditioned by the observed CMD. Our goal is to find the age, metallicity, distance, color excess, and binary fraction (with a fixed minimum mass fraction q) that maximizes the *posterior*.

$p(params)$ is the *prior* distribution, which represents the probability of the cluster having a given set of parameters, unconditioned by the observed CMD. In specific cases, restrictive *priors* can be chosen based on preliminary knowledge of the cluster. For instance, a normal function could be used as a metallicity *prior* for clusters that were analyzed spectroscopically. More generally, non-informative *priors* (flat functions) can be used within the expected range of the parameters.

Finally, $p(CMD|params)$ is the likelihood function, which represents the probability of a cluster with a given set of parameters displaying the observed CMD. This function can be estimated by evaluating the expected distribution of stars with a certain age, metallicity, distance, color excess, and binary fraction using synthetic populations and then comparing it with the observations using some mathematical formulation.

For SIESTA, we compare the number of stars in each bin of the Hess diagram evaluated for the cluster with the corresponding bin of the Hess diagram of a synthetic population using the likelihood from Tremmel et al. (2013):

$$p(CMD|params) = \Pi_{col,mag} \frac{\Gamma(0.5 + n^{obs}_{col,mag} + n^{synt}_{col,mag})}{\Gamma(1 + n^{obs}_{col,mag})\Gamma(0.5 + n^{synt}_{col,mag})}. \quad (2)$$

The indexes *col* and *mag* identify the specific colors and magnitudes of the centers of each bin, respectively. $n^{obs}_{col,mag}$ refers to the number of stars in a given bin of the observed CMD, and $n^{synt}_{col,mag}$ refers to the corresponding number of stars in the same bin of the synthetic population.

Figure 2 shows an application of this likelihood for the CMD of NGC 152, compared to a synthetic population that clearly does not match the observations. We can see that the likelihood is significantly smaller for the regions where we observe mismatches between the two CMDs, in particular in the Main Sequence turn-off and in the Red Clump.

For sampling the *posterior* distribution, we used the affine invariant Markov Chain Monte Carlo sampler EMCEE from Foreman-Mackey et al. (2013). To obtain final values for age, metallicity, distance, color excess, and binary fraction, we fit, to the corresponding marginalized *posterior* distributions, skewed-normal functions and use the mode as the final value, and the standard deviation as the uncertainty.

4. Results

We proceed to apply the SIESTA code in the characterization of the four clusters NGC 152, Lindsay 113, Lindsay 91, and NGC 121. In the fitting process, we will use non-informative *priors* for age, distance, and color excess, only restricting these quantities to reasonable intervals. For the metallicity, although there are spectroscopic determinations of this quantity for the clusters, we will also use non-informative *priors*, as a way to test SIESTA's ability to properly recover this parameter. Finally, for the binary fraction, we use a minimum mass fraction for the companion star of $q = 0.60$ and a log-normal *prior* distribution, following the empirical observations of Donada et al. (2023).

Figure 3 shows the marginalized *posteriors* obtained for the clusters, together with comparisons between the observed CMDs and the best-fitted synthetic population. For all clusters, we can see that the *posteriors* are "well-behaved" in all parameters, meaning that they appear concentrated in a central value, considered the best-fit for each cluster. The best-fitted synthetic populations also capture the main features observed in the CMDs, constraining the Main Sequence (in particular, the turn-off), the Subgiant and Red Giant branches, and also the position of the Red Clump.

Our results are compiled in Table 1, which also contains values from previous characterizations of the clusters. These results are put in perspective in Figure 4. The metallicities obtained using SIESTA are overall smaller than the values found by Milone et al. (2023). This trend also leads to higher color excess values, when comparing the two works. The systematic metallicity difference is not observed when comparing our results with previous VISCACHA characterizations, although divergences still occur for Lindsay 113 and NGC 121, where they are larger than the corresponding uncertainties. Still, when comparing our results with the spectroscopic determinations, we see that, in all cases, there is a good agreement between them, indicating that the SIESTA code is a reliable tool for recovering the metallicity of star clusters, at least for populations older than $1 Gyr$.

For the ages, there is an overall agreement between results from multiple works. For NGC 152, our results are consistent with the solution from Milone et al. (2023), fitted using the lower part of the cluster's extended Main Sequence turn-off (eMSTO), but is older than the findings of Dias et al. (2022). For Lindsay 113, our solution is older than what's presented in Milone et al. (2023) while matching the results from Oliveira et al. (2023). Finally, for NGC 121, our solution is younger than what's presented in Milone et al. (2023) and Saroon et al. (2023), although differences are within the 2σ limit.

Finally, for the distances, for Lindsay 114, and Lindsay 91, our results are consistent with both works. For NGC 152 they are an intermediary between Dias et al. (2022) and Milone et al. (2023). For Lindsay 113 they are smaller than Milone et al. (2023) and consistent with Oliveira et al. (2023) while the opposite trend is observed for NGC 121, with our results being larger than the findings of Milone et al. (2023), but are consistent with Saroon et al. (2023).

In the case of the binary fractions, most of our results are close to 13%, which is an effect of the *prior* applied in this parameter. For NGC 152, and NGC 121, however, we obtained considerably larger values. For NGC 152, this can be explained by the presence of the cluster eMSTO, which makes its upper main sequence showcase a larger dispersion than what is typically observed. Since the binary fraction tends to spread the stars in the CMD, this leads to the increase in this parameter, even if the physical origin of the eMSTO phenomenon is not typically associated with binarity, but rather with the combination of the presence of fast rotators and internal age variations (e.g., Cordoni et al. 2018). For NGC 121, on the other hand, since the cluster is densely populated, the excess in the binary fraction can be attributed to a large number of stars being superimposed in the image.

5. Conclusions and Perspectives

We have developed the SIESTA code, a new tool for characterizing star clusters by comparing the Hess diagrams of their CMDs with the ones from synthetic populations. The synthetic populations are created using PARSEC+COLIBRI isochrones and a Kroupa IMF. On top of that, we simulate non-resolved binary systems, photometric errors, and photometric completeness. The code uses MCMC sampling in a Bayesian approach to determine the cluster's age, metallicity, distance, color excess, and binary fraction.

We applied the code to four star clusters observed by the VISCACHA Survey, which were previously analyzed by the collaboration, and also by Milone et al. (2023), using HST data. In all cases, we were able to obtain solutions that reproduce the

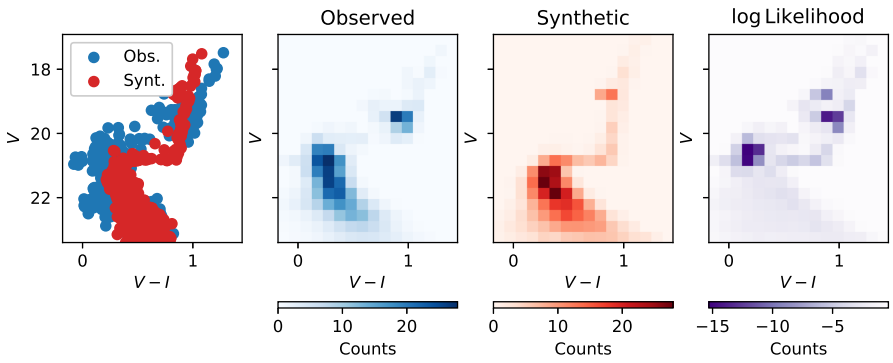

Figure 2. The first panel shows the observed CMD for NGC 152 (blue) and a synthetic population with age of 4 Gyr, metallicity of -1.5, distance of 50 kpc, color excess of 0.05, and binary fraction of 15%. The following two panels show the corresponding Hess Diagrams, and the final one, the logarithm of the likelihood of each bin, calculated using Equation 2.

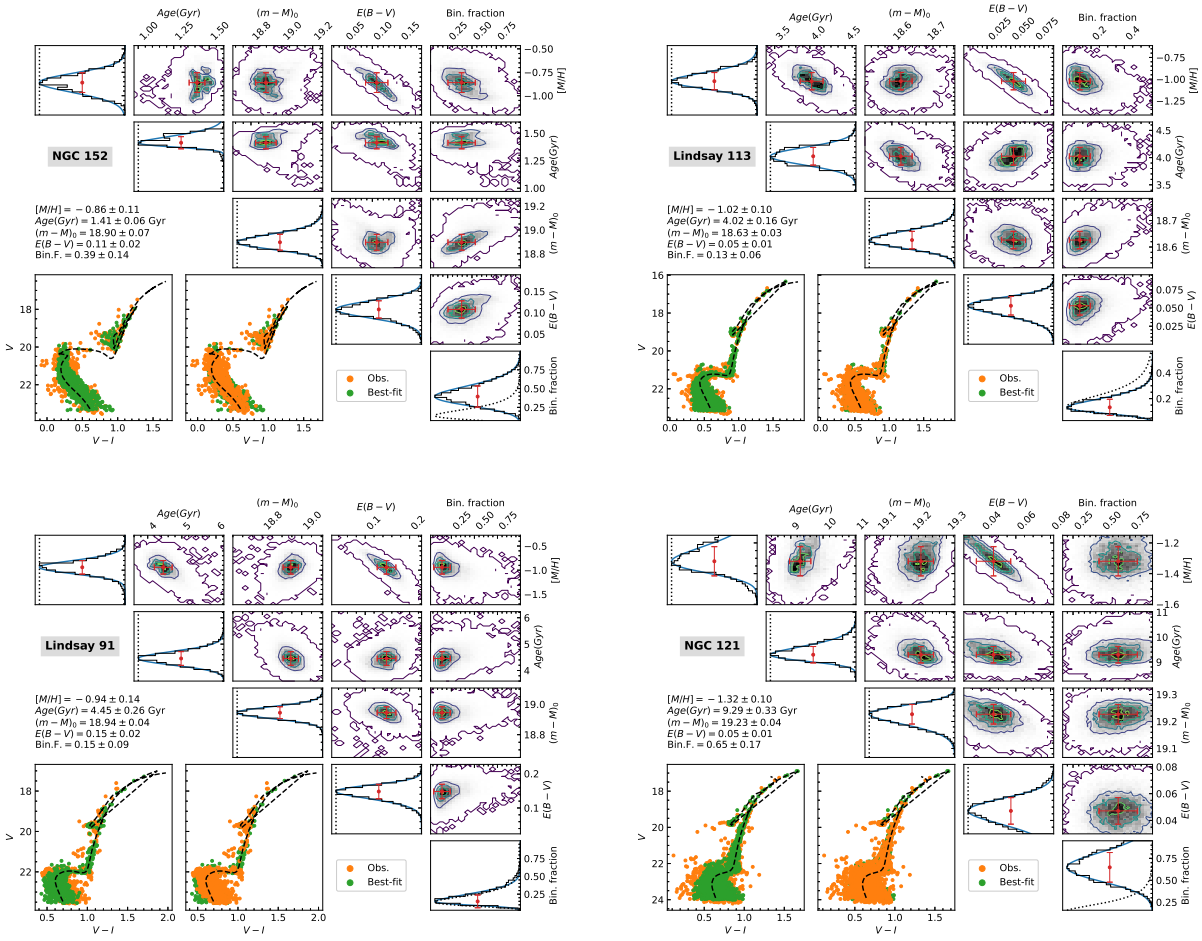

Figure 3. Results of the analysis of the four clusters. For each image, the panels on the top-right side show the corner plots of the marginalized *posterior* distributions. In the main diagonal, we have the 1-dimensional histograms, where the *priors* are shown as dotted lines and the fitted skewed-normal distribution is shown in blue. In the bottom left, the text shows the results of the fits, and the CMDs display comparisons between observations (orange) and the best-fitted synthetic population (green; randomly sampled with the same number of synthetic stars as the analyzed cluster).

main features of the observed CMDs. Our results were also able to systematically recover metallicities consistent with spectroscopic determinations, despite using only non-informative *priors* for this parameter.

Our main goal with the SIESTA code is to use it as a tool for homogeneously characterizing star clusters observed by the VISACACHA Survey, although we recognize that it could be used in other scientific contexts as well. The code could also be expanded, by including more sophisticated phenomena (e.g., multiple populations) or allowing for the use of different grids of isochrones, to evaluate the dependence of the results with the choice of the stellar evolution model. Finally, the synthetic pop-

Table 1. Results from characterizations performed for the four clusters analyzed in this work. For each cluster, values are presented in the following order: first, the fits obtained using SIESTA, as described in the text; second, results from the isochrone fitting performed by Milone et al. (2023) (uncertainties were not evaluated by the authors); third, results from previous isochrone fitting performed by the VISCACHA team using different methods (see text for details); finally, results for the metallicity coming from spectroscopic determinations, the methods used are indicated between parenthesis. For NGC 152 two age values were found by Milone et al. (2023), considering different approaches to the cluster's extended main sequence turn-off (eMSTO). The value outside of the parenthesis corresponds to an isochrone fitted to the lower part of the eMSTO, while the value inside the parenthesis comes from a fit to the upper part of the eMSTO.

Cluster	Source	$[M/H]$	Age (Gyr)	$(m-M)_0$	$E(B-V)$	Bin. F.
NGC 152	This work	-0.86 ± 0.11	1.41 ± 0.06	18.90 ± 0.07	0.11 ± 0.02	0.39 ± 0.14
	Milone et al. (2023)	-0.6	1.45 (1.90)	19.07	0.03	–
	Dias et al. (2022)	$-0.77^{+0.07}_{-0.21}$	$1.27^{+0.04}_{-0.26}$	$18.71^{+0.07}_{-0.06}$	$0.11^{+0.07}_{-0.04}$	–
	Song et al. (2021)	-0.73 ± 0.11 (High-resolution spectroscopy)				
Lindsay 113	This work	-1.02 ± 0.10	4.02 ± 0.16	18.63 ± 0.03	0.05 ± 0.01	0.13 ± 0.06
	Milone et al. (2023)	-0.8	3.60	18.76	0.03	–
	Oliveira et al. (2023)	-0.87 ± 0.08	3.90 ± 0.40	18.69 ± 0.05	0.03 ± 0.03	–
	Parisi et al. (2015)	-1.03 ± 0.04 (Calcium triplet)				
Lindsay 91	This work	-0.94 ± 0.14	4.45 ± 0.26	18.94 ± 0.04	0.15 ± 0.02	0.15 ± 0.09
	Milone et al. (2023)	-0.8	4.20	18.97	0.10	–
	Oliveira et al. (2023)	-0.82 ± 0.08	3.90 ± 0.50	18.86 ± 0.09	0.15 ± 0.03	–
	De Bortoli et al. (2022)	-0.90 ± 0.06 (Calcium triplet)				
NGC 121	This work	-1.32 ± 0.10	9.29 ± 0.33	19.23 ± 0.04	0.05 ± 0.01	0.65 ± 0.17
	Milone et al. (2023)	-1.2	9.70	19.05	0.04	–
	Saroon et al. (2023)	$-1.50^{+0.05}_{-0.15}$	$10.10^{+0.30}_{-0.50}$	$19.16^{+0.06}_{-0.06}$	$0.07^{+0.13}_{-0.03}$	–
	Song et al. (2021)	-1.28 ± 0.06 (High-resolution spectroscopy)				

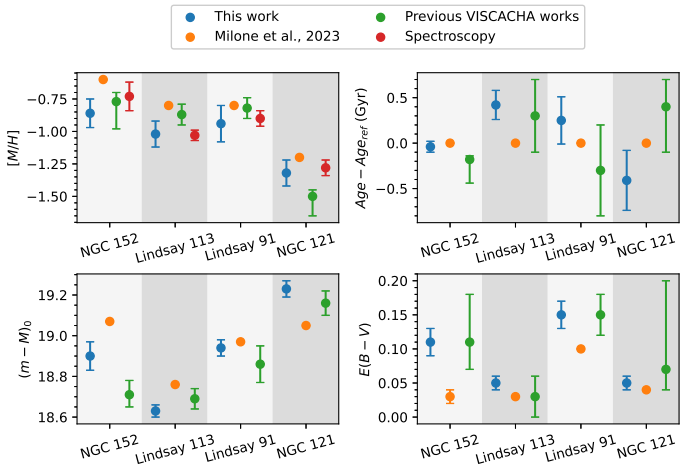

Figure 4. Comparison of the results obtained in this work, and previous characterizations performed for these clusters, listed in Table 1. Each panel refers to a specific parameter. In the age panel (top right), the results of each cluster are shown in relation to the Milone et al. (2023) values.

ulations could be used for inferring masses for the cluster stars, allowing for the estimation of the present-day mass function.

Acknowledgements. BPLF and JFCG acknowledge financial support from CNPq (proc. 141875/2023-2 and 140642/2021-8) and CAPES (Finance Code 0001). BD acknowledges support by ANID-FONDECYT iniciación grant No. 11221366. FFSM acknowledges financial support from CNPq (proc. 404482/2021-0) and from FAPERJ (proc. E-26/201.386/2022 and E-26/211.475/2021). The work is based on observations obtained at the Southern Astrophysical Research (SOAR) telescope, which is a joint project of the Ministério da Ciência, Tecnologia e Inovações do Brasil (MCTI/LNA), the US National Science Foundation's NOIRLab, the University of North Carolina at Chapel Hill (UNC), and Michigan State University (MSU).

References

Angelo, M. S. et al., 2023, MNRAS, 522, 956
Bressan, A., 2012, MNRAS, 427, 127
Castro-Ginardi, A. et al., 2022, A&A, 652, A162
Cordoni, G. et al., 2018, ApJ, 869, 139
Dalessandro, E. et al., 20216 ApJ, 829, 77
De Bortoli, B. J. et al., 2022, A&A, 664, A168
Dias, B. et al., 2014, A&A, 561, A106
Dias, B. et al., 2022, MNRAS, 512, 4334
Donada, J. et al., 2022, arXiv preprint arXiv:2301.11061
Ferreira, F. A. et al., 2019, MNRAS, 483, 5508
Foreman-Mackey, D. et al., 2013, PASP, 125, 306
Kerber, L. et al., 2007, A&A, 462, 139
Kroupa, P., 2001, MNRAS, 322, 231
Maia, F. F. S. et al., 2010, MNRAS, 407, 1875
Maia, F. F. S. et al., 2014, MNRAS, 437, 2005
Maia, F. F. S. et al., 2019, MNRAS, 484, 5702
Marigo, P. et al., 2017, ApJ, 835, 77
Milone, A. P. et al., 2023, A&A, 672, A161
Oliveira, R. A. P. et al., 2023, MNRAS, 524, 2244
Parisi, M. C. et al., 2015, ApJ, 149, 154
Perren, G. I. et al., 2015, A&A, 576, A6
Robinson, E. et al., 2016, ASCL, 1608
Saroon, S. et al., 2023, A&A, 677, A35
Song, Y. et al., 2021, MNRAS, 504, 4160
Souza, S. O. et al., 2020, ApJ, 890, 38
Tokovinin, A. et al., 2016, PASP, 128, 970
Tremmel, M. et al., 2013, ApJ, 766, 19
Weidner, C. et al., 2004, MNRAS, 350, 1503