Draft version July 17, 2023
Typeset using LaTeX default style in AASTeX631

The isochrone fitting problem: a first approach to a generalized statistic fitting

Baeza, Ian[1]

[1] Institute of Astrophysics, Faculty of Exact Sciences, Universidad Andrés Bello, Santiago, Chile

ABSTRACT

The parameter estimation for Stars Clusters through Isochrones has been widely used to make rough estimation even if in several studies there's no statistical formality (at least clearly described within the manuscripts). Along this study we aim to generalized a proper way of estimate those physical properties that commonly are just fitted by eye, and get some errors for these estimations. In specific we study four star clusters, we first determine the most likely candidates through their proper motion and in addition we keep stars with their *color* $\times$ *color* consistent with stellar evolution (in order of get a reliable estimation of the cluster). The we perform an isochrone fitting where first we map the likelyhood of several models and then we run a Markov chain Monte Carlo (MCMC) to retrieve errors of the Age, Metallicity, Global Reddening and Distance.

Keywords: Globular Cluster — Open Cluster — stellar association — Isochrone — MCMC

1. STAR CLUSTERS & ISOCHRONES

1.1. *Open Cluster and Globular Clusters*

Star clusters such as Open Clusters (OPCs, characterized by a young population of stars and found in the surroundings of areas with ongoing star formation Krumholz et al. (2019)) and Globular clusters (GCs, characterized by the high density of stars, an old stellar population II and currently a specific chemical signature Bastian & Lardo (2018)) are groups gravitationally bound stars that born from the same nebula, hence they have the same (very similar) physical properties such Age and chemical composition (To see a more detailed and extended review see and references in Archinal & Hynes (2003); Bastian & Lardo (2018); Chi et al. (2023)). In this sense, star clusters are ultimately described by an Initial Mass Function (IMF, which in simple terms is a function that describes the number of stars with a specific mass that will form given an initial mass (the total mass of a molecular cloud which is the place where stars form), see for complete description of the topic Bastian et al. (2010); Kroupa et al. (2013)) and the subsequent stellar evolution (taking into account all the difficulties that may arise in these processes due to the chaotic and complex nature of the evolution and including the interaction with other stars).

Due these characteristic Star Clusters are powerful tools to several areas in astronomy since the ability to measure and estimate distances ages and metallicity have proven to be a hard task. In this context OPCs can be used as standard candles to measure distance, to estimate a rough estimation of the metallicity enrichment of the location, to trace areas where we have active star formation (not only within our own galaxy but in extragalactic environments) and also are a near tool to study stellar evolution (Krumholz et al. (2019); Kharchenko et al. (2005); Monteiro et al. (2021); Chi et al. (2023)), on the other hand GCs are also reliable estimator of distance, metallicity of the substructures where they reside (such Bulge and Halo of the Milky Way) and are old time capsules that could give us some insights of how was the environment where they form giving us a very specific knowledge about what was the metallicity at certain epoch in the past, so with this information we can reconstruct how was the specific enrichment of the structure where they belong (in specific through this we know that the Bulge has experienced a much more aggressive enrichment than the experienced by the Halo, Bastian & Lardo (2018); Baumgardt & Vasiliev (2021); Vasiliev & Baumgardt (2021)).

1.2. *Stellar evolution Models*

A stellar evolutionary model is a mathematical model that can be used to calculate the evolutionary phases of a star from its genesis through its extinction. The star's mass and chemical composition are utilized as inputs, and the only limits are its luminosity and surface temperature. These models when are used together with the information of a cluster (their IMF) are called isochrones. The meaning in the own name, It's a sample of stellar models compute at the same (iso) age (chrone). In practice, this implies that some theorists utilize a stellar evolution algorithm to compute evolutionary tracks (i.e. a star's attributes through time, or age) for a variety of masses, and then connect all the models of different masses at the same age (e,g Bressan et al. (2012); Nguyen et al. (2022)).

These isochrones are use to determine the physical properties of clusters such Distance. Reddening, Age , and Metallicity (e.g. Chi et al. (2023))

2. DATA AND TARGETS

2.1. *Gaia DR3*

We were working with data from Gaia, a space mission that collects astrometry, photometry, and spectroscopic data from billions of stars in the Milky Way (MW). Gaia Data Release 3 (Gaia DR3) was issued on June 13, 2022, making it the most recent release. The *Gaia DR3* data products combine previously published *Gaia DR3* data products with new data products for the same collection of observations and time. The *Gaia DR3* offers a major improvement over *Gaia DR3* Data Release 2, with parallax and correct motion precision's enhanced (by 30% and by a factor of two, respectively). The DR3 photometry likewise has improved precision, but most importantly, it has considerably better homogeneity across color, magnitude, and celestial position. In comparison to *Gaia EDR3*, *Gaia DR3* provides a larger set of data, including: sources with mean G_RVS-band magnitudes, sources with radial velocity, mean RVS spectra, sources with chemical abundances from RVS spectra, and so on will allow us to compare our metallicities and membership probability estimates with the radial velocity data, for example.(Gaia Collaboration et al. 2022)

2.2. *PARSEC Isochrones*

On the other hand the stellar evolution models that we are going to use are from Bressan et al. (2012); Nguyen et al. (2022), these Isochrones were obtained through the webpage http://stev.oapd.inaf.it/cmd which allow us to select the photometric system (Gaia photometry in this case) and select a sample of certain Ages and metallicities range.

2.3. *Stellar Clusters*

We select 4 star cluster relatively close (compared with the overall sample of GCs and OPCs) since our aim in this study is to perform a reliable fit and retrieve physical properties of these clusters.

- **Messier 4 / NGC 6121 / (M4):** Is the closest GCs to our system has been widely studied SIMBAD[1] has registered at the moment of this manuscript is written 1808 entries referring to this object , we redirect the reader to check http://www.messier.seds.org/m/m004.html in order a brief overview about this cluster. For the initial parameters regarding of this GCs (position, size, proper motion, etc.) we use as reference the values from Baumgardt & Vasiliev (2021); Forbes & Bridges (2010), with this info we retrieve from *Gaia* 1 degree around the position of the cluster corrected by proper motion.
- **47 Tucanae / NGC 104 / (47 Tuc):** Is second brightest GCs in the sky with at least 3825 references registered in SIMBAD[2], as a further read about this cluster you can check http://www.messier.seds.org/xtra/ngc/n0104.html. For the position, size, proper motion, etc. We use the values obtained in Baumgardt & Vasiliev (2021); Forbes & Bridges (2010), with this info we retrieve from *Gaia* 1 degree around the position of the cluster corrected by proper motion.
- **Messier 22 / NGC 6656 / (M 22):** was one of the first GCs to be discovered, accumulating 1342 studies where it's refered according to SIMBAD[3], to have a broader notion about the historical relevance of this cluster you can check http://www.messier.seds.org/m/m022.html. For the main physical parameters of this cluster, we take as reference the values obtained in Baumgardt & Vasiliev (2021); Forbes & Bridges (2010), with this info we retrieve from *Gaia* 1 degree around the position of the cluster corrected by proper motion.
- **Melotte 22 / Pleiades:** is among the nearest OPCs to Earth, it has been widely studied having 3327 references associated in SIMBAD[4], to read more about this cluster we suggest to see http://www.messier.seds.org/m/

[1] SIMBAD M4 link
[2] SIMBAD 47 Tuc link
[3] SIMBAD M 22 link
[4] SIMBAD Melotte 22 link

m045.html which includes a historical overview. Regarding about the initial parameters we use the values in the catalogue provided by Melotte (1915) and the values obtained in van Leeuwen (2009); Brandner et al. (2023), with this info we get from *Gaia* 180 arc-min around the center corrected by proper motion.

3. PROBABILITY OF MEMBERSHIP

We use a classical approach to retrieve the most likely members, using astrometric parameters such as the proper motion we use the general approach of two 2D-Gaussian mixture model (GMM), this model considers that the distribution of proper motions of the stars in a cluster's region can be represented by two (or more) elliptical bivariate Gaussians. (Griggio & Bedin (2022); Xiang et al. (2021); Çakmak & Karataş (2022))

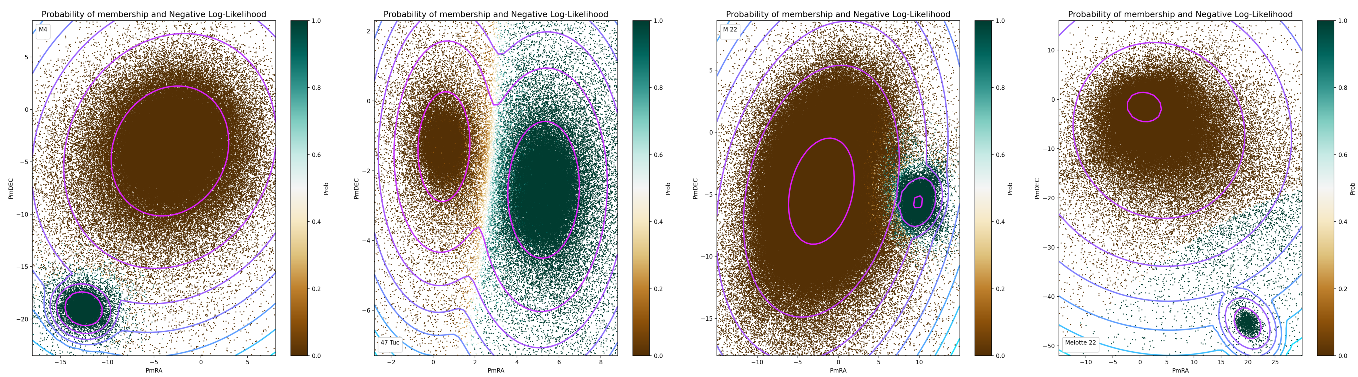

Figure 1. Probabilities achieved for our 4 objects of study through Gaussian Mixture of 2 components (except for Melotte 22 which a third component was introduced 2 of them to describe the field)

This approach in this case was done through the Scikit-Learn Package of python (Pedregosa et al. (2011)) , which has developed functions of clustering to retrieve means and covariance matrices. Using this tool we get the mean motion of the cluster their dispersion and the correlation between the variables. Our results are in good agreement with the literature values from Baumgardt & Vasiliev (2021) for the Globular clusters and in van Leeuwen (2009) for the Melotte 22 cluster. (this values are summarized in a table in the Appendix A).

Due to the limited time the other selection process was done without rigor, it is included for the GCs a selection by position , this means that we selected the center over-density and eliminate the outskirts of the GC. On the other hand for the OPC from our sample we use the over-density in parallax. For the whole sample we include a color x color consistency , this means that the relation between their magnitudes are correlated as we exect from a normal stellar evolution (that it's our aim since we are fitting a Single Stellar Population (SSP)).

4. ISOCHRONE FITTING

The Isochrone fitting was performed in three steps, first with an initial run was performed in order to separate the likely candidate per stellar stage (this will not be the final stages associated it just to perform a interpolation) this must be done in this way since the HR diagram shows a complex pattern that can not be inferred as a function by the color and magnitude, however if we split them in per stellar stage we can make functions that given the color would gave us the mag or that given the mag could gave us the color, with this idea in mind we perform a Gaussian Process (GP) to estimate the trend and the main behaviour of the cluster (this was performed through 3 kernels of Pedregosa et al. (2011), a long RBF to get the main behaviour of the stellar stage, a RationalQuadratic kernel to obtain the small irregularities (like curves or bending the trend), and the WhiteKernel to estimate the intrinsic error, this with the purpose of optimize the fitting (and in this sense we don't have to handle with a very numerous and irregular sample).

The initial fit was done through a minimization of the distance to an isochrone with input values near to the literature (see 2.3 and the references there), in this step we also measure the standard deviation along each axis as well as for the representative data as the original data (it will be used to discard sources that are not well explained through the SSP model selected).

After this was settled the next step is iterate over the isochrones (which were interpolated in order to create a homogeneous distribution and do not bias the measure of distance), and map the likelyhood varying the Age , Metallicity, the reddening (Av, this was assumed as a constant for the cluster and considering proportionality in absorption given in the webpage of PADOVA isochrones) and the distance modulus (Um). Once we map the whole sample we proceed to make a very high resolution association to the real data with the best Isochrone (applying as a rule that sources can not be brighter than point in the isochrone if it is redder and that can not be fainter than a dot in the isochrone if it is bluer). As an example we illustrate these steps for the fitting of Pleiades in fig (3), also the dispersion assigned can be found in B .

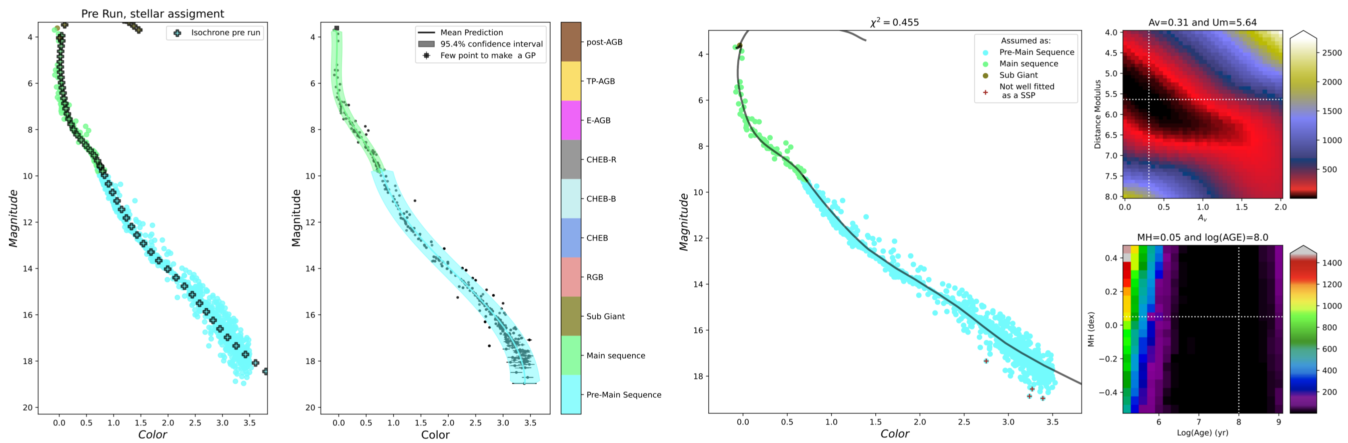

Figure 2. The fitting process, the first panel shows a pre-assignment to determine under which axis we will perform the interpolation (this is done in low resolution since the important is get the trend). The second panel shows the final result of the fitting with the best isochrone found also are showed the likelyhood for MH vs AGE and the specific Av vs Um for this metallicity and age. Regarding to the value it seems obvious that the errors are overestimated we attribute this to the estimation of error along the axis that we are not able to interpolate (since it's not a function, this means that we have more than one point in the dependant variable for some of the independent variable), which were calculated as the dispersion in the sample per length cover by the stage (Percentile 84 - Percentile 16 , divided by 2 and divided by the width covered by the sample in that stage).

This was done by the whole sample of cluster, the results and the code can be found in the GitHub https://github.com/Inebama/Astrostatistics_Iso_Fit.git were we have put the main codes and the figures of output.

5. THE FAILURE STEP

After doing the likelyhood mapping our intentions were clear, we wanted to estimate the errors of our measurements, and since we had the likelyhood the only thing left was interpolate these values (collapse the maps in both axes metallicity and Age) to run a MCMC and in this way estimate the errors, but due to lack of time and the initial steps took us more time than we expected we could not finish this project.

6. ACKNOWLEDGMENTS

I would like to thank to you Professor Giuliano P. for your time and the patient to teach me the useful tools along the course of Astrostatistics

Facilities: Gaia DR3

Software: Astropy (Astropy Collaboration et al. 2013, 2018), Numpy (Harris et al. 2020), Matplotlib (Hunter 2007), Scikit-learn (Pedregosa et al. 2011)

APPENDIX

A. VALUES OBTAINED FROM GMM

Table 1. Mean vector of Motion for each Cluster with their corresponding covariance matrix, obtained with GMM with the package of Scikit-Learn. (In units of mas/yr, and to compute these values we use 2 components one for the field and another for the Cluster except by the Melotte 22 where due the high number of field stars 2 components were required just by the field)

Cluster	Mean [pmRA,pmDec]	Covariance Matrices
M 4	[-12.51880983,-19.02460664]	$\begin{pmatrix} 0.74150974 & -0.06132156 \\ -0.06132156 & 0.49493991 \end{pmatrix}$
47 Tuc	[5.27156186 , -2.54886654]	$\begin{pmatrix} 0.78800102 & 0.03979797 \\ 0.03979797 & 1.03969432 \end{pmatrix}$
M 22	[9.8703678 , -5.63281973]	$\begin{pmatrix} 0.43488632 & 0.07800918 \\ 0.07800918 & 0.35232734 \end{pmatrix}$
Melotte 22	[19.92093319 , -45.44234436]	$\begin{pmatrix} 1.09986888 & -0.37473833 \\ -0.37473833 & 1.71408382 \end{pmatrix}$

B. OBTAINING DISPERSION

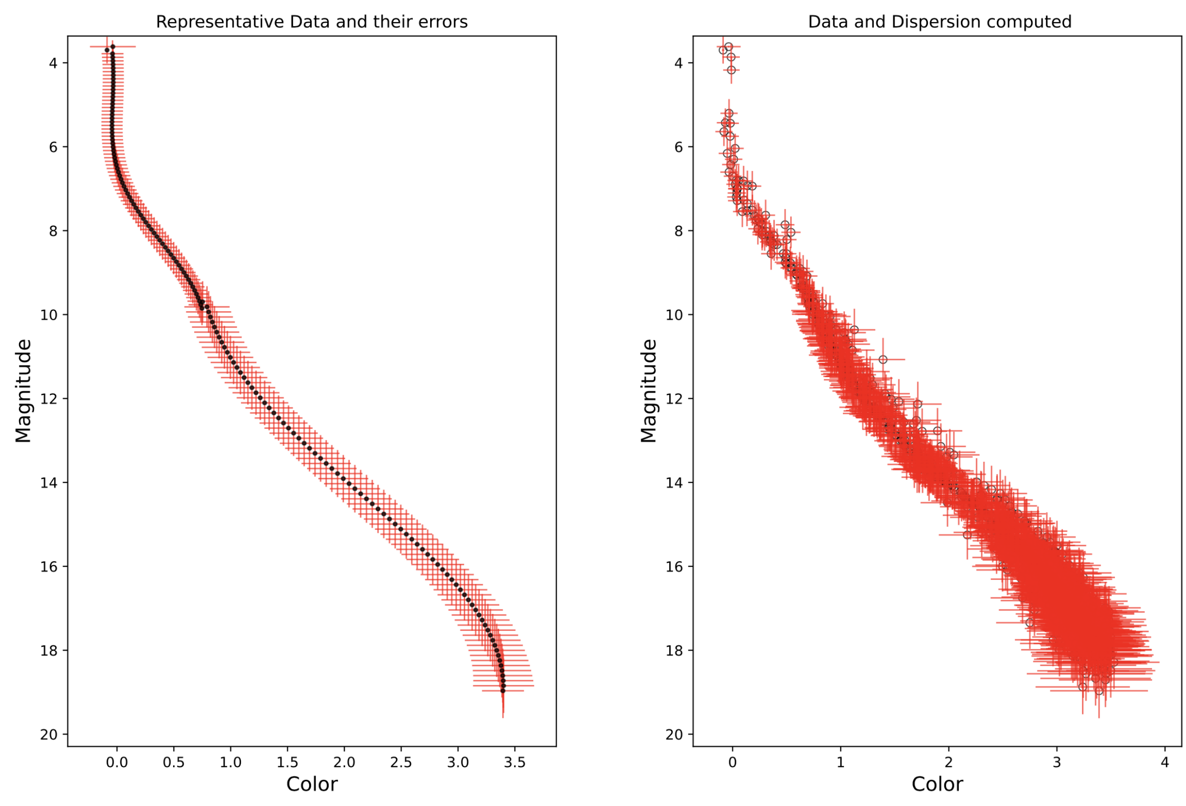

Figure 3. The Dispersion computed for Melotte 22, the representative data set and the original data set based in GP and intrinsic dispersion per initial stellar stage. (Also a handicap was introduced in order to make the stars brighter weight more that the fainter ones)

REFERENCES

Archinal, B. A., & Hynes, S. J. 2003, Star Clusters

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339, doi: 10.1146/annurev-astro-082708-101642

Bastian, N., & Lardo, C. 2018, ARA&A, 56, 83, doi: 10.1146/annurev-astro-081817-051839

Baumgardt, H., & Vasiliev, E. 2021, MNRAS, 505, 5957, doi: 10.1093/mnras/stab1474

Brandner, W., Calissendorff, P., & Kopytova, T. 2023, arXiv e-prints, arXiv:2306.03132, doi: 10.48550/arXiv.2306.03132

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127, doi: 10.1111/j.1365-2966.2012.21948.x

Çakmak, H., & Karataş, Y. 2022, na, 96, 101833, doi: 10.1016/j.newast.2022.101833

Chi, H., Wang, F., Wang, W., Deng, H., & Li, Z. 2023, ApJS, 266, 36, doi: 10.3847/1538-4365/accb50

Forbes, D. A., & Bridges, T. 2010, MNRAS, 404, 1203, doi: 10.1111/j.1365-2966.2010.16373.x

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2022, arXiv e-prints, arXiv:2208.00211. https://arxiv.org/abs/2208.00211

Griggio, M., & Bedin, L. R. 2022, mnras, 511, 4702, doi: 10.1093/mnras/stac391

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R. D. 2005, A&A, 438, 1163, doi: 10.1051/0004-6361:20042523

Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. 2013, in Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations, ed. T. D. Oswalt & G. Gilmore, Vol. 5, 115, doi: 10.1007/978-94-007-5612-0_4

Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, ARA&A, 57, 227, doi: 10.1146/annurev-astro-091918-104430

Melotte, P. J. 1915, MmRAS, 60, 175

Monteiro, H., Barros, D. A., Dias, W. S., & Lépine, J. R. D. 2021, Frontiers in Astronomy and Space Sciences, 8, 62, doi: 10.3389/fspas.2021.656474

Nguyen, C. T., Costa, G., Girardi, L., et al. 2022, A&A, 665, A126, doi: 10.1051/0004-6361/202244166

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

van Leeuwen, F. 2009, A&A, 497, 209, doi: 10.1051/0004-6361/200811382

Vasiliev, E., & Baumgardt, H. 2021, MNRAS, 505, 5978, doi: 10.1093/mnras/stab1475

Xiang, Y., Xi, J., Shao, Z., Wang, M., & Yang, Y. 2021, apss, 366, 17, doi: 10.1007/s10509-021-03923-9