

SAM: Using Python to Fit Isochrones to Clusters

Shannon Bischof^{1, 2,*}

¹*Physics and Engineering Department, Coastal Carolina University.*

²*Applied Physics Major
Engineering Science Major
Applied Mathematics Minor*

(Dated: May 3, 2025)

Color-Magnitude Diagrams (CMD) are graphs of stellar brightness vs. color, which translate to graphs of luminosity vs. temperature. Isochrone model curves can be fitted to a CMD to determine the cluster's distance, age, number of heavy elements (metallicity), and amount of light being blocked by dust (extinction). There are two parts to fitting an isochrone to a cluster of stars. The first is to compare characteristics of the stars (proper motion and distance) to determine which stars are members of the cluster, and the second is to properly fit the isochrone line to the distribution of remaining stars. The process is usually manual, which can be error-prone and time-consuming. Using a Python process limits human error and accelerates the process. The results are output faster than it would be if it were done by hand; however, the results are neither accurate nor precise. This process was tested using a sampling of clusters that had already been measured by Cluster Pro Plus. The results show that this process is not yet accurate.

I. INTRODUCTION:

Color-Magnitude Diagrams (CMDs) are an important tool in astronomy. They plot brightness against color to analyze the properties of star clusters. Isochrones, a curve on a CMD that represents a population of stars of the same age, are essential to determine the properties of star clusters. Traditionally, the process of fitting the CMD to an isochrone is done manually, which is error-prone and time-consuming. The errors can arise from the selection process of the stars to include in the cluster analysis (membership selection) and in the fitting process.

Stellar Analysis Machine (SAM), inspired by Skynet UNC's Astromancer Program [1], aims to automate the traditionally manual isochrone fitting process. Using Astroquery for data retrieval and ASteCA for membership selection and analysis.

SAM converts the distance modulus into distance in kiloparsecs, and the total extinction into reddening extinction. SAM saves the CMDs of each cluster, including filtered versions, isochrone fittings, and synthetic overlays. This study evaluates the accuracy and precision of the SAM in the estimation of stellar parameters, in order to enhance the efficiency of isochrone fitting.

II. METHODS:

Astroquery uses SIMBAD [2] to extract Gaia Data Release 3 (GDR3) [3][4] based on the cluster specified by the user. Astroquery was also used to compile the mass amounts of the GDR3 data from the list shown in VII A. We ensured that the right data were retrieved from GDR3 by comparing the true right ascension and declination of the cluster with that printed by Astroquery on both SIMBAD and GDR3 outputs. ASteCA

[5] uses the GDR3 data to create synthetic clusters and fits isochrones to the CMD. We had ASteCA use a FastMP filter and only kept the stars that had a 75 % chance of being part of the cluster to perform the membership selection. We also had the option to use a Bayesian filter, but that method required a radius to work. FastMP uses proper motion and parallax to estimate membership probabilities. We used the FastMP method to create a filtered version of the cluster because the purpose of this project is to limit user input. We also set the minimum and maximum logarithmic age to 7.0 and 10.1 and the minimum and maximum distance to 7.0 and 17.0. ASteCA required the maximum and minimum to be set so that the graphing mechanic could work. ASteCA was capable of estimating outside that range; however, the graphs would not load if the parameters were estimated to be outside of the range. We then generated the isochrones and created a synthetic cluster that we calibrated using the filtered cluster. SAM retrieved the CMDs with the isochrone and synthetic overlay-ed onto the filtered cluster. This was done so that we could see how the isochrones look compared to the cluster. SAM also saved the credible intervals for the estimations.

We tested the accuracy and precision of SAM. We used data from Cluster Pro Plus (CPP)[1], the data was determined by hand by the Astromancer Program [1]. An assumption that was made was that although there is room for human error, training in the Astromancer Program is sufficient to eliminate large amounts of human error. The Cluster Pro Plus results were also filtered by removing duplicates whose estimated values were far from each other. If we consider the Cluster Pro Plus values to be true, we can judge the accuracy and precision of the SAM values.

We chose four random clusters to show the results in this paper. We applied this method to 68 clusters listed in Appendix VII A which were processed in this manner. CMDs of the clusters are located in the Appendix VII B. To see the code, visit SAM's GitHub: <https://github.com/Perseid-ST/Stellar-Analysis-Machine>

The Cluster Pro Plus data records the distance to a cluster in kilo-parsecs, so SAM converted the distance modulus from

* Contact author: sgbischof@coastal.edu

AStCA to kilo-parsecs using the following equation [6]:

$$D = 10^{\frac{dm-5}{5}} / 1000 \quad (1)$$

Where D is distance in kilo-parsecs and dm is the distance modulus.

Cluster Pro Plus also only accepted the reddening extinction, so SAM also converted the total extinction to reddening extinction using the following equation [5]:

$$E_{BV} = A_V / R_V \quad (2)$$

Where A_V is the total extinction and R_V is the ratio of total to selective extinction. This is around 3.1 for clusters in the Milky Way. The metallicities in Cluster Pro Plus and AStCA have different units as well; however, the units of Cluster Pro Plus metallicity could not be determined.

III. RESULTS:

Below are the CMDs of IC 2944, Platais 9, M45, NGC 2453 and their corresponding credible intervals. In the CMDs, the red line is the isochrone, the blue circles are the GDR3 data points, and the green triangles are the synthetic cluster's data points. We extracted the accepted values from CPP and compared SAM's estimation with the CPP data.

IC 2944 has an accepted reddening of 0.4285 [7], which is a total extinction of 1.33 using Equation 2, and a metallicity of 0.105 [7]. IC 2944 also has an accepted log (age) of 6.6 [7], and an accepted distance of 2.565 kpc [7], which is a distance modulus of approximately 12.04 using Equation 1. As shown in Figure 1(a), SAM has an estimated total extinction of 1.064 ± 0.065 , a distance modulus of 11.151 ± 0.101 , an age of 7.002 ± 0.017 , and a metallicity of 0.015 ± 0.001 . The credible intervals in Figure 1(b) show that the estimations of total extinction (Av), distance modulus (dm), and age (loga) converge, while the metallicity does not. The percentage error for the total extinction is 20 %, the distance modulus is 7.38 %, the age is 6.09 %.

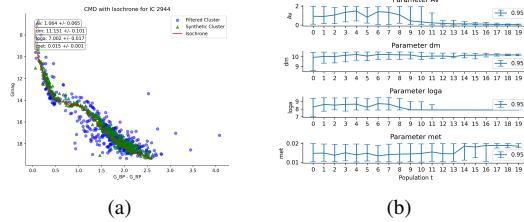


FIG. 1: 1(a) is a CMD of IC 2944 with the synthetic cluster and the isochrone overlay-ed on top of the cluster. 1(b) is the credible intervals of the IC 2944 estimations

Platais 9 has an accepted reddening of 0.05 [7], which is a total extinction of 0.155 using Equation 2, and a metallicity of 0.3 [7]. Platais 9 also has an accepted log(age) of 7.71 [7], and a accepted distance of 0.22 kpc [7], which is a distance modulus of approximately 6.71 using Equation 1. As

shown in Figure 2(a), SAM has an estimated total extinction of 1.047 ± 0.548 , a distance modulus of 12.378 ± 2.733 , an age of 8.521 ± 0.928 , and a metallicity of 0.016 ± 0.003 . The credible intervals in Figure 2(b) show that the estimations of total extinction (Av), distance modulus (dm), age (loga), and metallicity do not converge. The percentage error for the total extinction is 575. 48 %, the distance modulus is 84. 47 %, the age is 10.52 %.

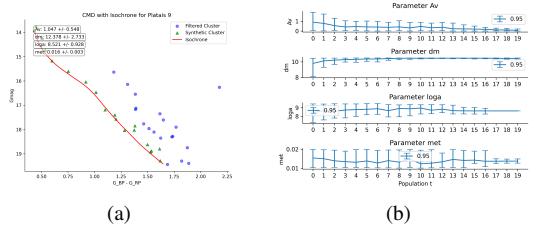


FIG. 2: 2(a) is a CMD of Platais 9 with the synthetic cluster and the isochrone overlay-ed on top of the cluster. 2(b) is the credible intervals of the Platais 9 estimations

Messier 45 has an accepted reddening of 0.01 [7], which is a total extinction of 0.031 using Equation 2, and a metallicity of 0 [7]. Messier 45 also has an accepted log(age) of 8 [7], and a accepted distance of 0.13 kpc [7], which is a distance modulus of approximately 5.57 using Equation 1. As shown in Figure 3(a), SAM has an estimated total extinction of 1.054 ± 0.657 , a distance modulus of 11.93 ± 2.669 , an age of 8.386 ± 0.917 , and a metallicity of 0.015 ± 0.003 . The credible intervals in Figure 3(b) show that the estimations of total extinction (Av), distance modulus (dm), age (loga), and metallicity do not converge. The percentage error for the total extinction is 3300.00 %, the distance modulus is 114.18 %, the age is 4.83 %.

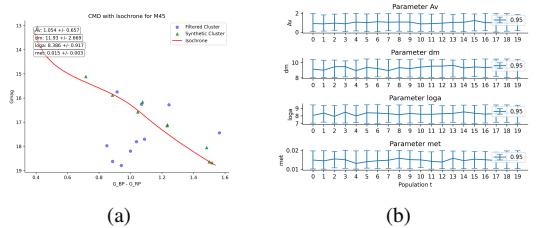


FIG. 3: 3(a) is a CMD of M45 with the synthetic cluster and the isochrone overlay-ed on top of the cluster. 3(b) is the credible intervals of the M45 estimations

NGC 2453 has an accepted reddening of 0.57 [7], which is a total extinction of 1.767 using Equation 2, and a metallicity of -0.42 [7]. NGC 2453 also has an accepted log (age) of 7.50 [7], and an accepted distance of 4.19 kpc [7], which is a distance modulus of approximately 13.11 using Equation 1. As shown in Figure 4(a), SAM has an estimated total extinction of 1.438 ± 0.084 , a distance modulus of 13.191 ± 0.119 , an age of 8.5 ± 0.213 , and a metallicity of 0.013 ± 0.003 . The credible intervals in Figure 4(b) show that the estimations of

total extinction (Av), distance modulus (dm), and age ($\log(a)$) converge, while the metallicity does not. The percentage error for the total extinction is 18.62 %, for the distance modulus is 0.62 %, for the age is 13.33 %.

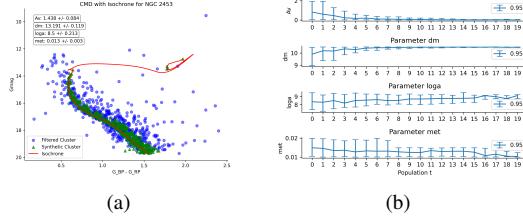


FIG. 4: 4(a) is a CMD of NGC 2453 with the synthetic cluster and the isochrone overlay-ed on top of the cluster. 4(b) is the credible intervals of the NGC 2453 estimations

50 CMDs can be found in Appendix VII B.

IV. DISCUSSION OF RESULTS:

In the following graphs, the results of SAM were compared with those of CPP. In the figures, the blue dashed line shows what the best-fit line would look like if the SAM data were equal to the Cluster Pro Plus data. The green line shows the best-fit line of the data from SAM.

In Figure 5, the SAM produces an estimate of the distance with a best-fit line of $y = 0.172x - 1.285$, however, the data points do not appear to have a linear displacement. The displacement appears to be exponential, which could indicate a flaw in how SAM converts from the distance modulus to kiloparsecs. The low slope indicates that SAM is significantly underestimating the distance. The intercept indicates a consistent error where the estimated distance is less than the CPP distance. The value of R^2 is 0.696, indicating that the SAM data are moderately close to the best-fit line.

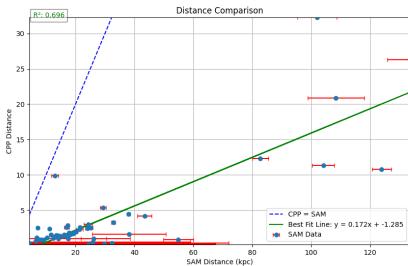


FIG. 5: This figure shows the comparison of distance estimation from SAM with Cluster Pro Plus distance.

In Figure 6, SAM produces an estimate of the $\log(a)$ with a best-fit line of $y = 0.903x + 0.692$. The slope is close to one, which indicates that the age estimation is more likely to be accurate. The intercept indicates a constant error where the estimated age is higher than the CPP age. The value of R^2 is

0.439, indicating that the data do not fit well with the best-fit line.

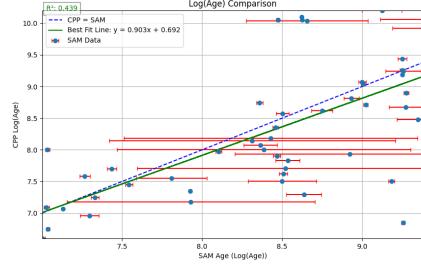


FIG. 6: This figure shows the comparison of $\log(a)$ estimation from SAM with Cluster Pro Plus $\log(a)$.

In Figure 7, SAM produces an estimate of the extinction with a best-fit line of $y = 0.764x + 0.076$. The slope is lower than one, but is closer to one than zero. The intercept indicates that there is a small but constant error where the estimated total extinction is higher than the total extinction from CPP. The value of R^2 is 0.285, indicating that the data do not fit well with the best-fit line.

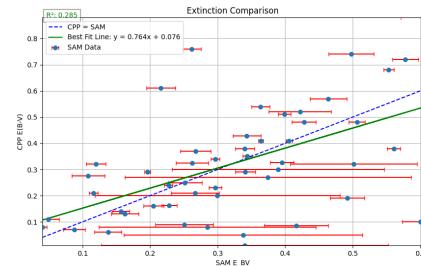


FIG. 7: This figure shows the comparison of extinction estimation from SAM with Cluster Pro Plus extinction.

Because we are uncertain of the units of metallicity in the CPP data, no meaningful comparison can be made.

This shows that this method of stellar analysis is inaccurate and imprecise.

V. CONCLUSION:

In summary, the distance estimation exhibits a significant systematic offset. Although age and extinction estimations appear satisfactory on a larger scale, the precision of individual results is not satisfactory.

Future work will focus on addressing distance error and exploring the underlying causes of image-saving issues. Additionally, we aim to analyze how SAM performs when handling globular clusters, open clusters, and stellar associations separately. In addition, we will look into better isochrone models. Currently, it is unknown whether AStECA could not make the CMDS or whether SAM failed to save the CMD, future work will also solve this problem.

VI. ACKNOWLEDGMENTS:

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

VII. APPENDIX:

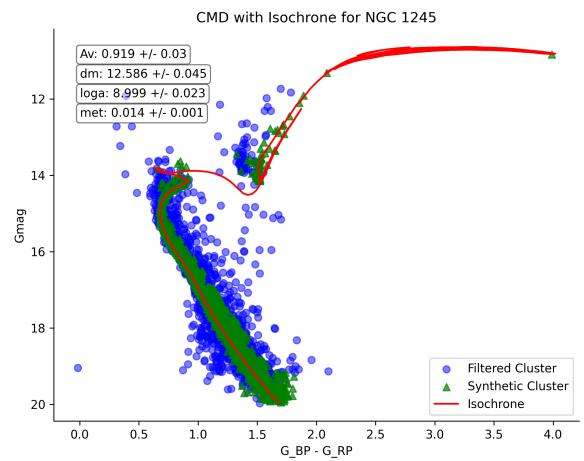
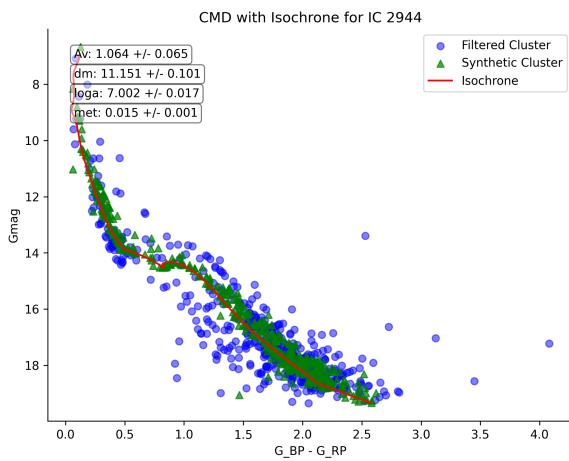
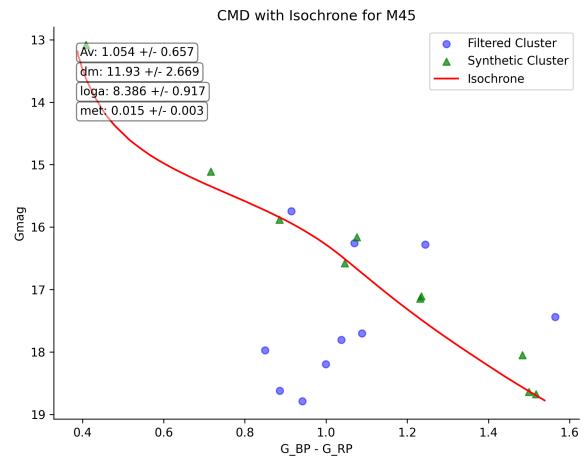
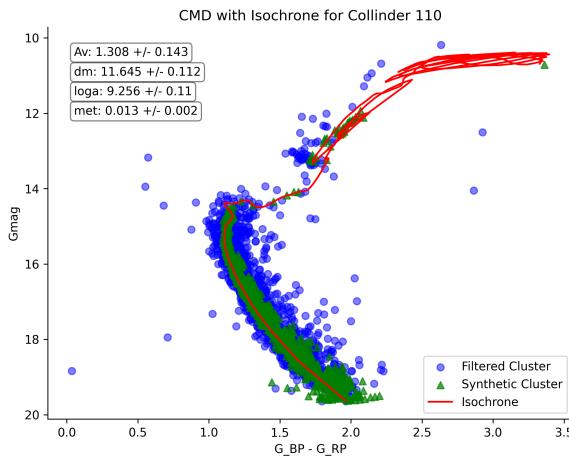
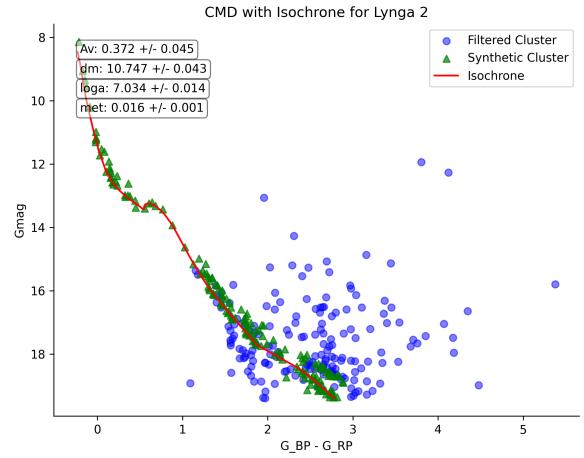
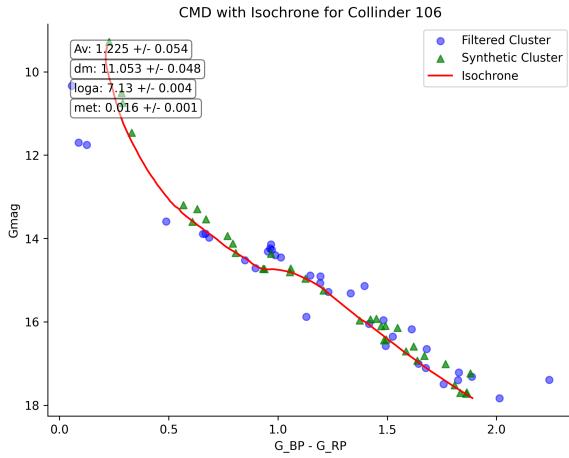
A. List of Clusters Used

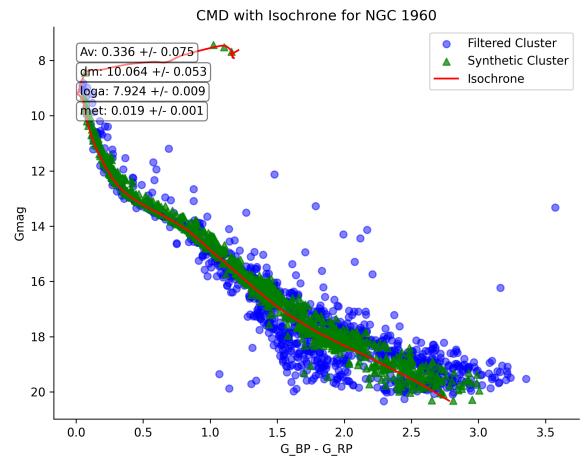
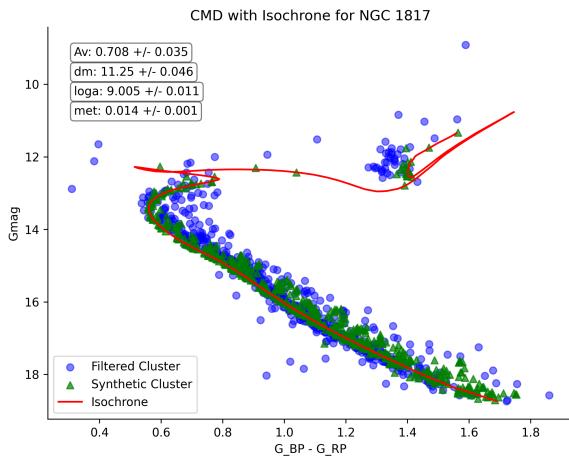
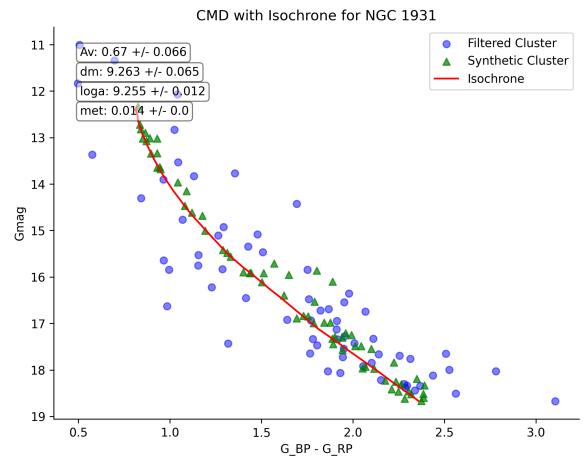
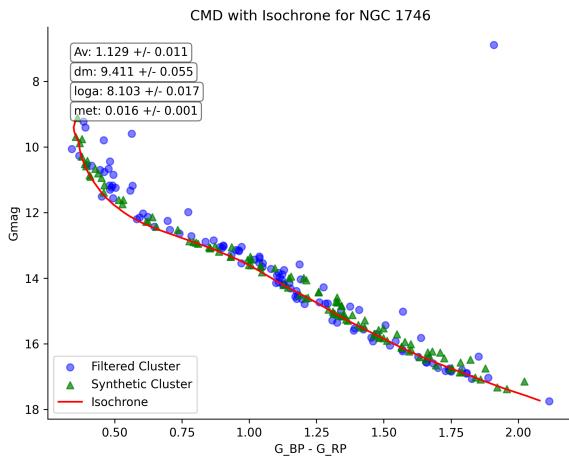
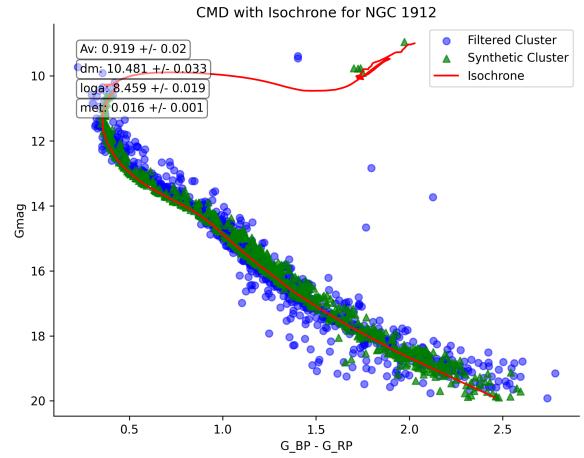
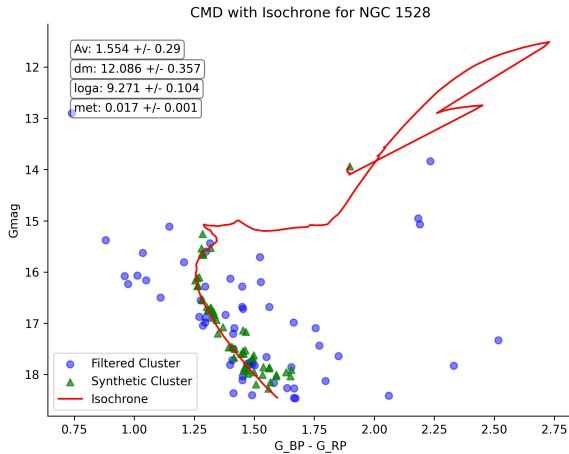
- Collinder 106 • NGC 2516
- IC 2944 • NGC 2546
- IC 4756 • NGC 2632
- Lynga 2 • NGC 2682
- NGC 1817 • NGC 3766
- NGC 1960 • NGC 4755
- NGC 2099 • NGC 5634
- NGC 2158 • NGC 5822
- NGC 2264 • NGC 6139
- NGC 2345 • NGC 6229
- NGC 6025 • NGC 6231
- NGC 6475 • NGC 6235
- Collinder 110 • NGC 6242
- IC 1805 • NGC 6325
- IC 4499 • NGC 6352
- IC 4651 • NGC 6401
- IC 5146 • NGC 6440
- M45 • NGC 6441
- NGC 104 • NGC 6496
- NGC 1245 • NGC 6583
- NGC 1528 • NGC 6717
- NGC 1746 • NGC 6939
- NGC 188 • NGC 7082
- NGC 1912 • NGC 7243
- NGC 1931 • NGC 7492
- NGC 2168 • NGC 7654
- NGC 2269 • NGC 7789
- NGC 2298 • NGC 869
- NGC 2323 • NGC 957
- NGC 2360 • NGC3766
- NGC 2422 • NGC4755
- NGC 2447 • Per OB2
- NGC 2453 • Platais 9
- NGC 2509 • Sco OB4

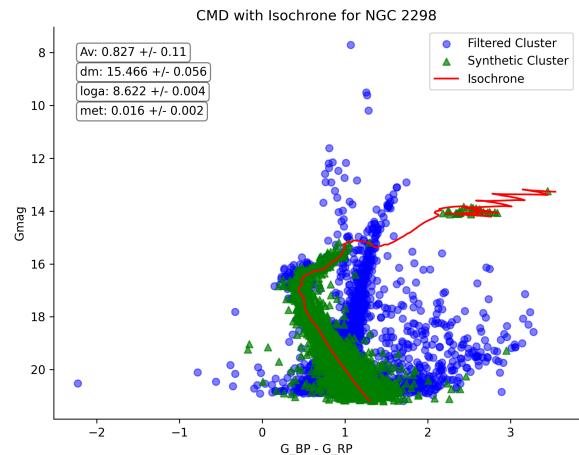
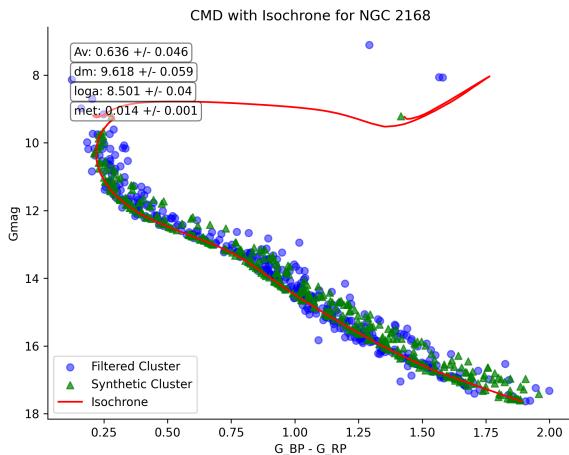
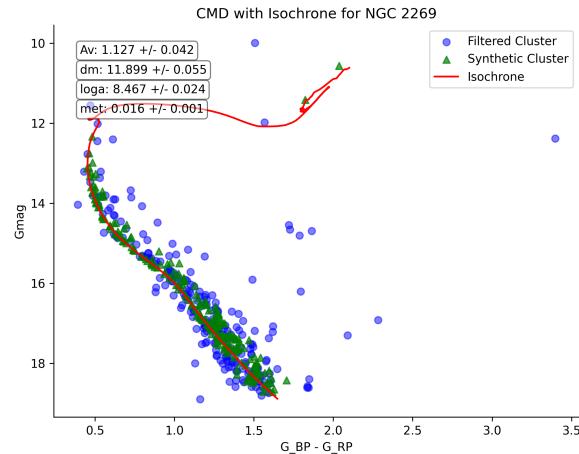
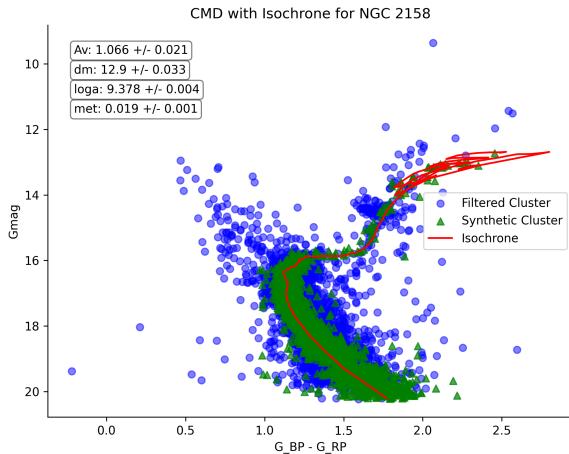
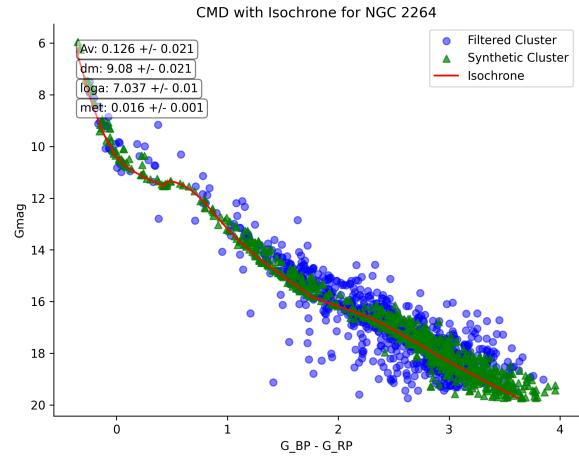
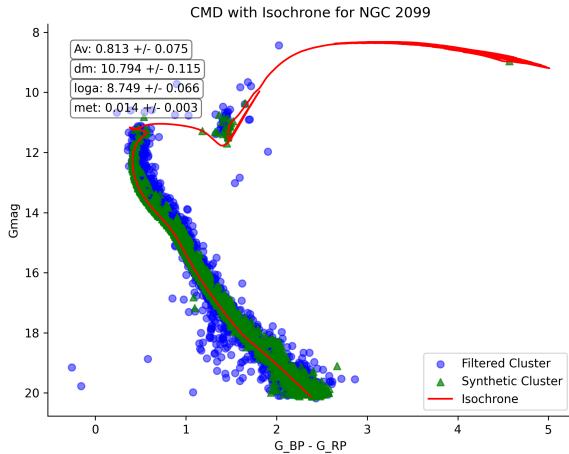
B. CMDs of Some of the Clusters Listed in Appendix A

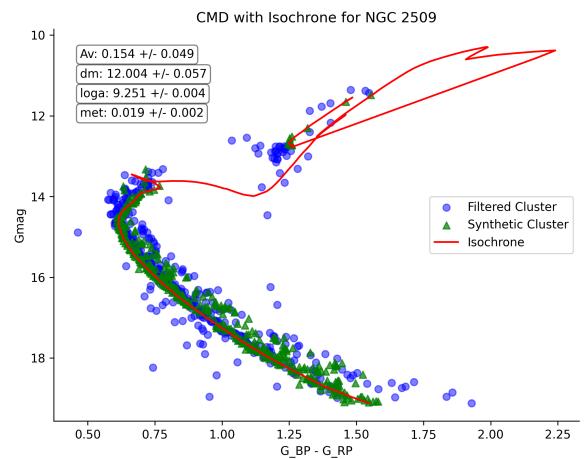
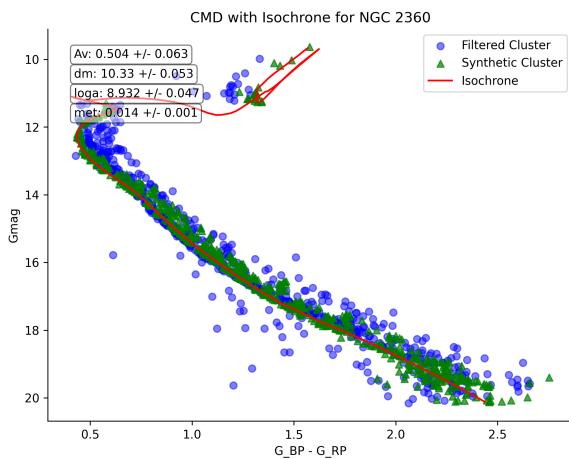
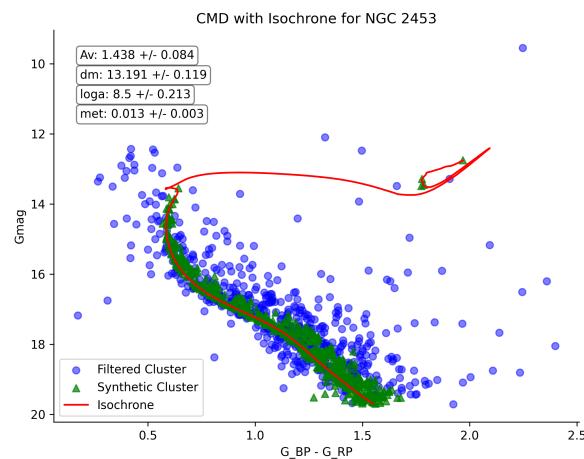
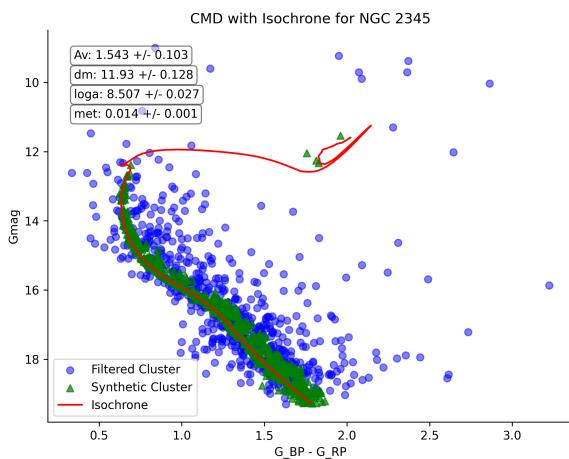
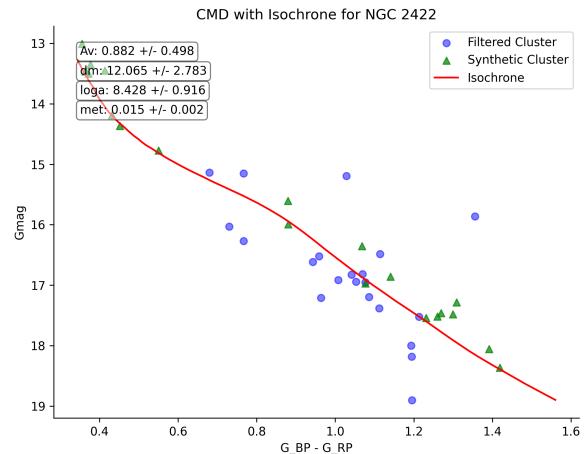
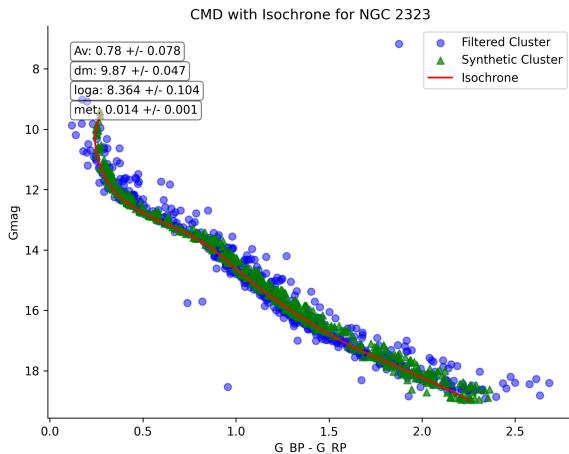
This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France

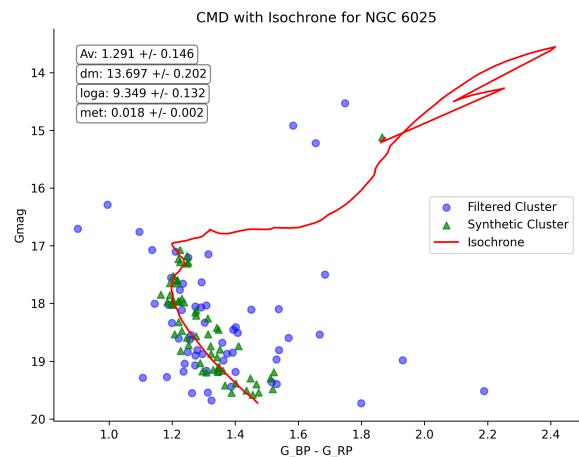
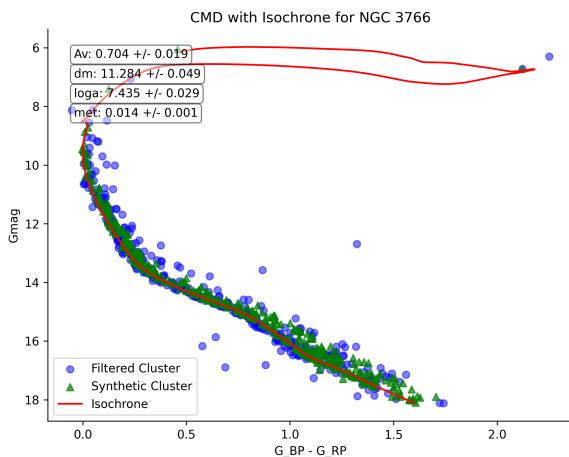
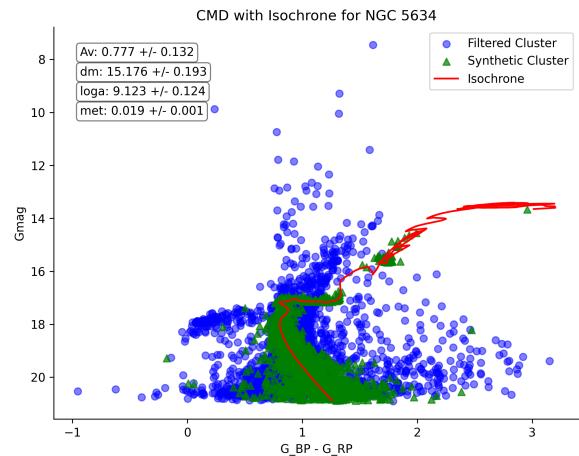
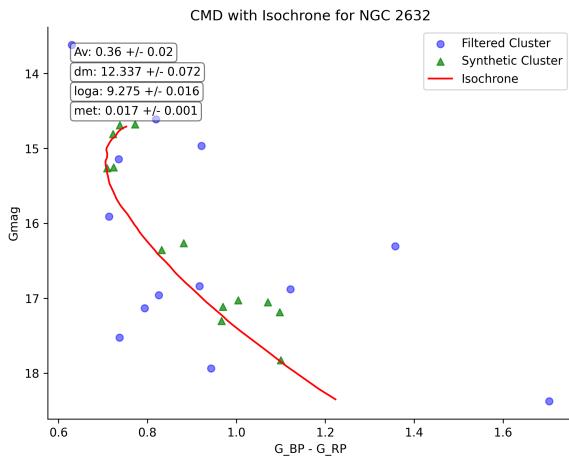
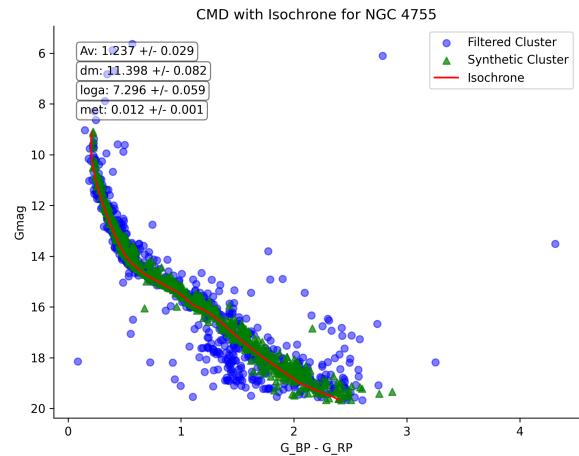
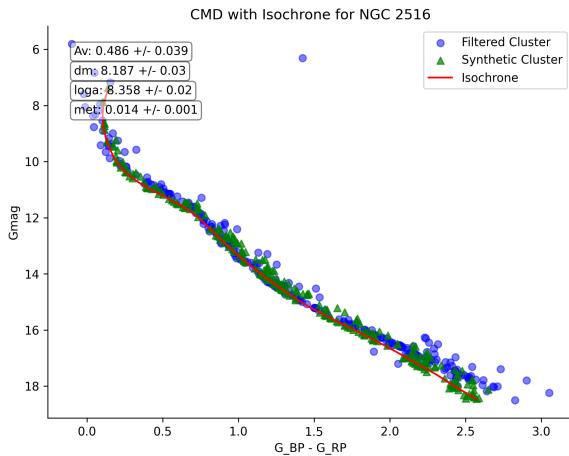
Loop through all filenames in the list

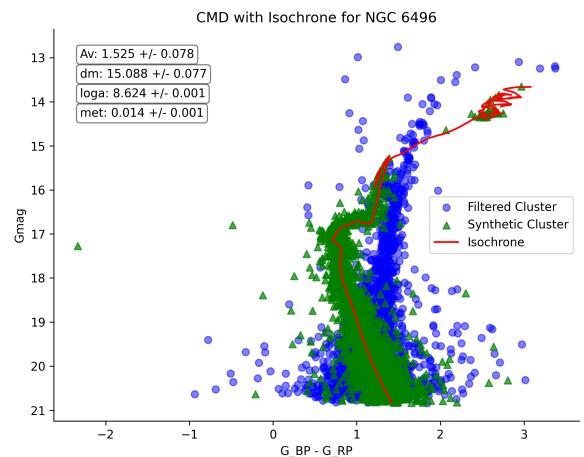
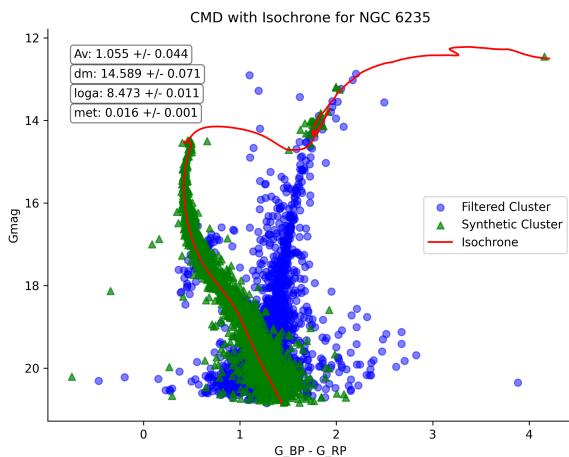
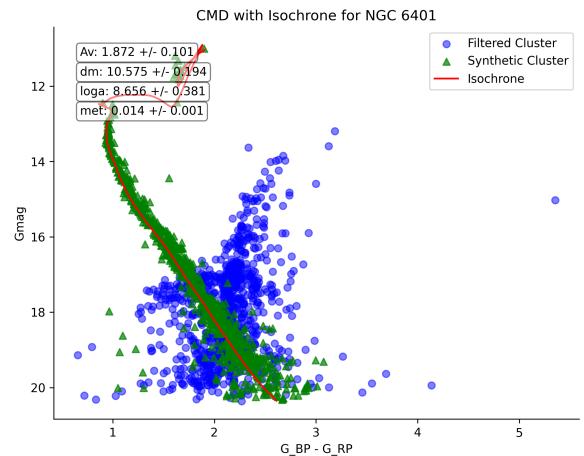
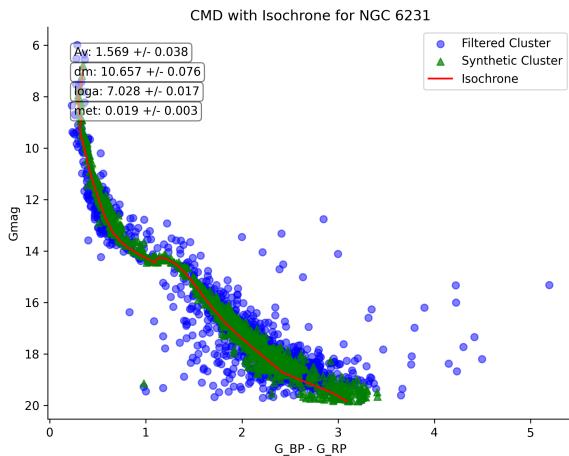
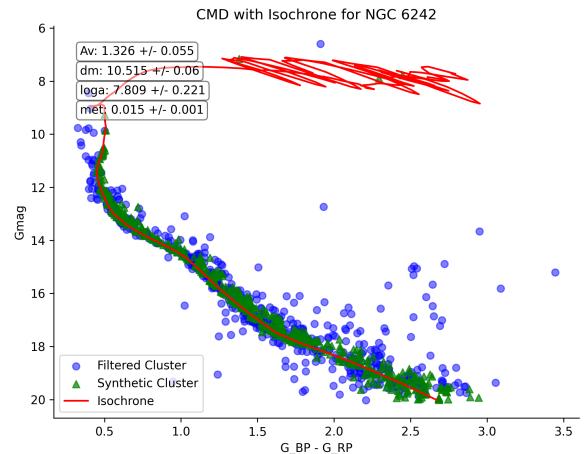
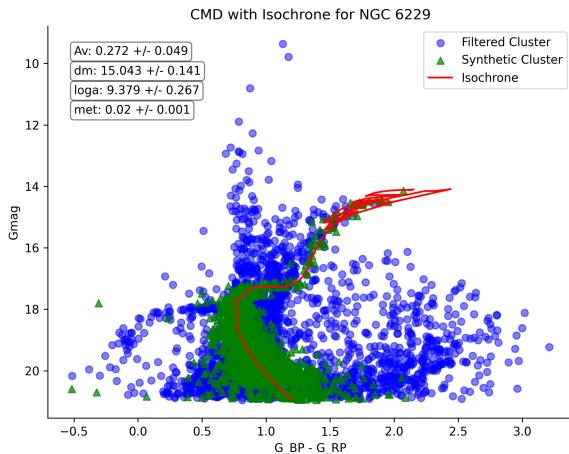


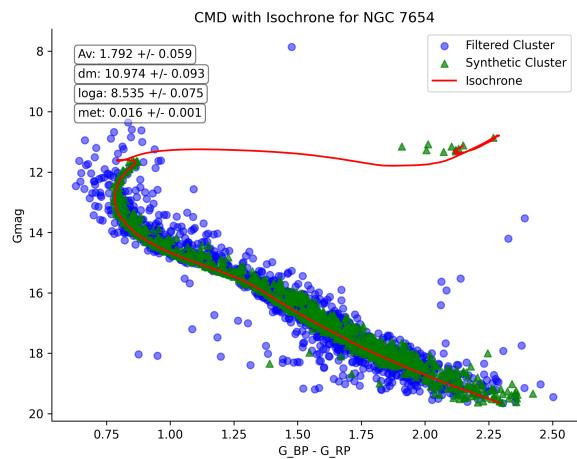
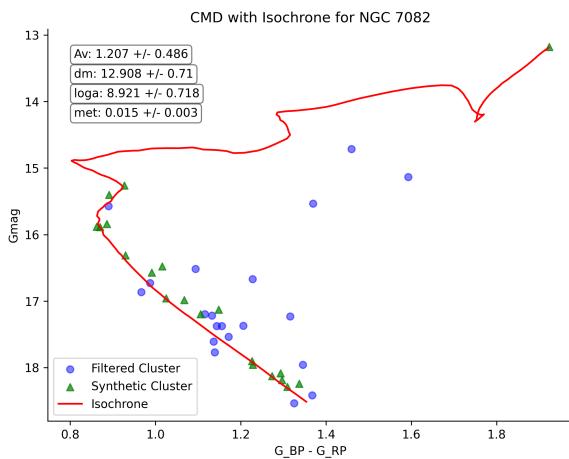
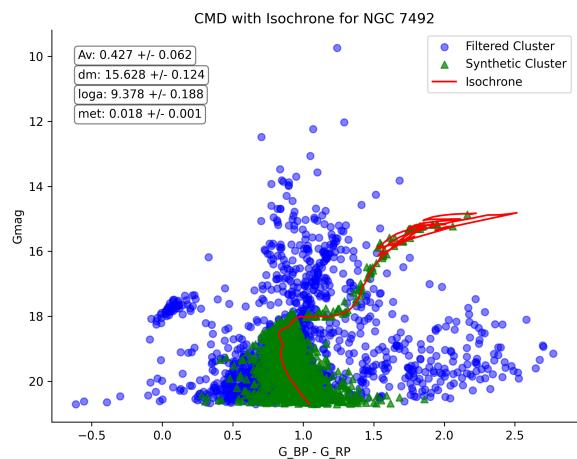
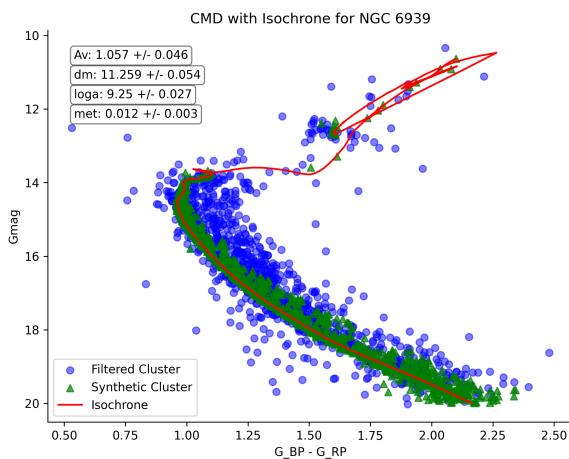
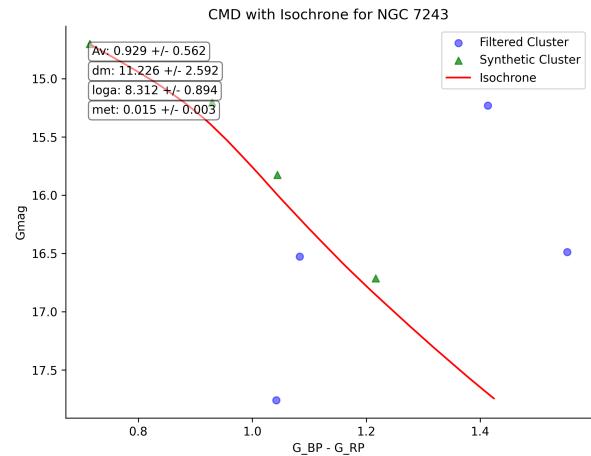
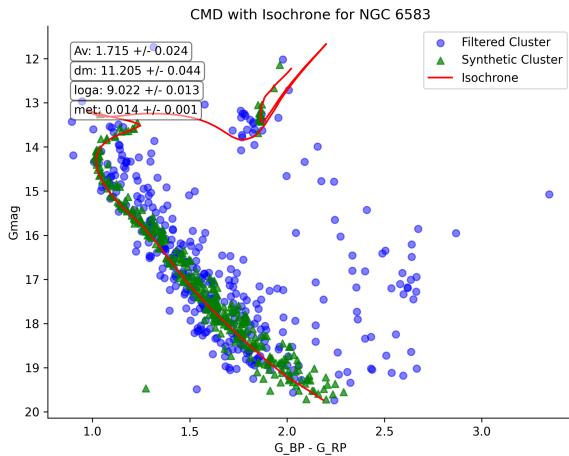


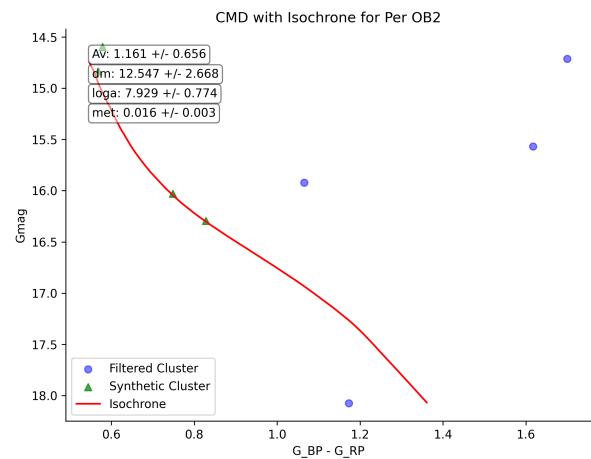
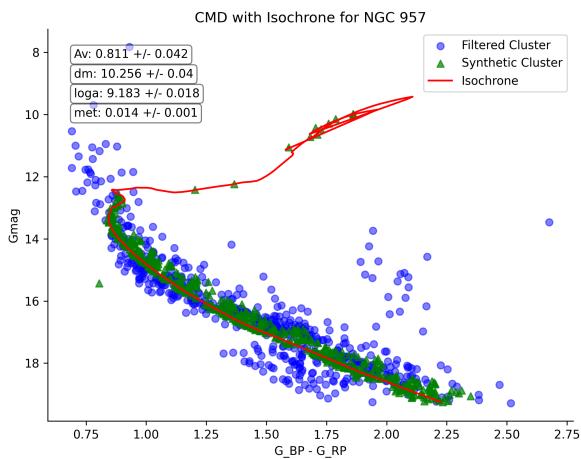
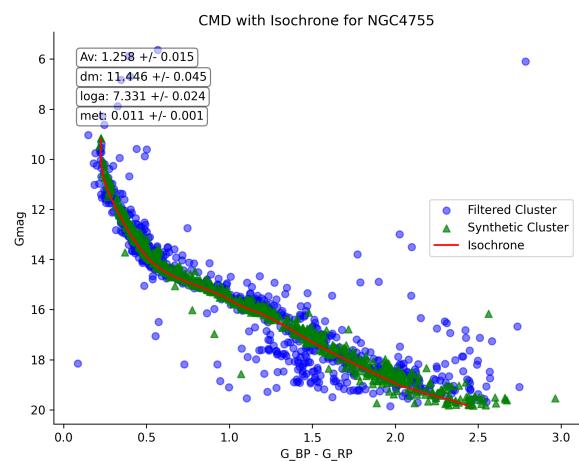
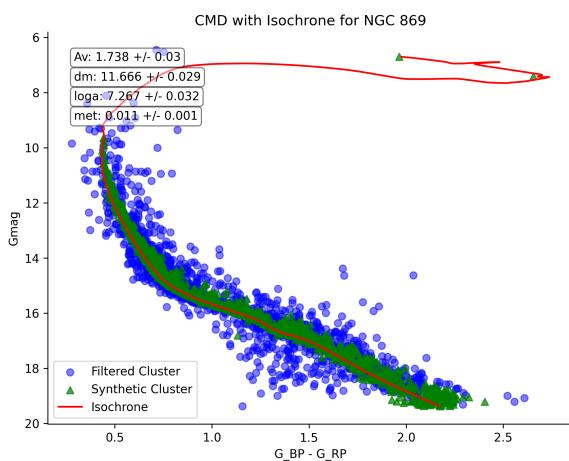
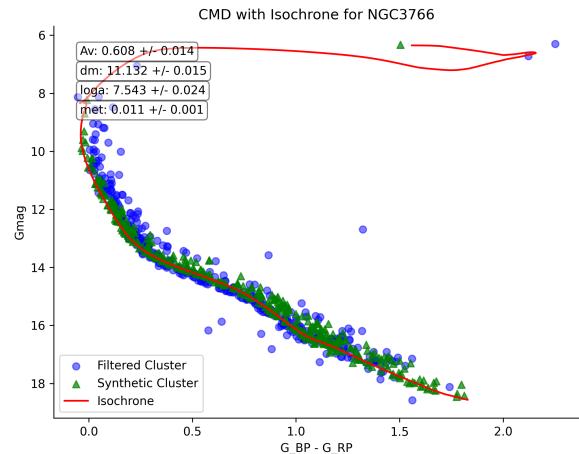
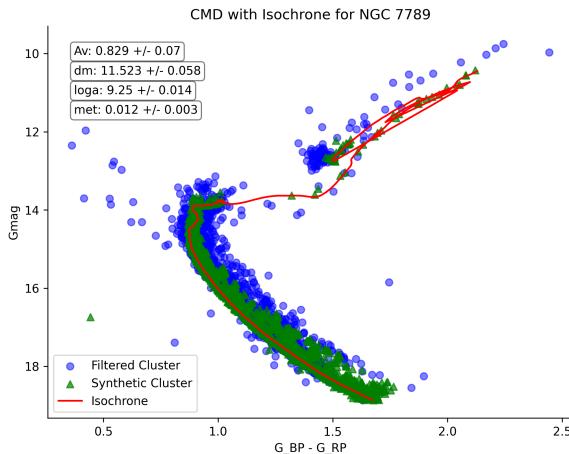


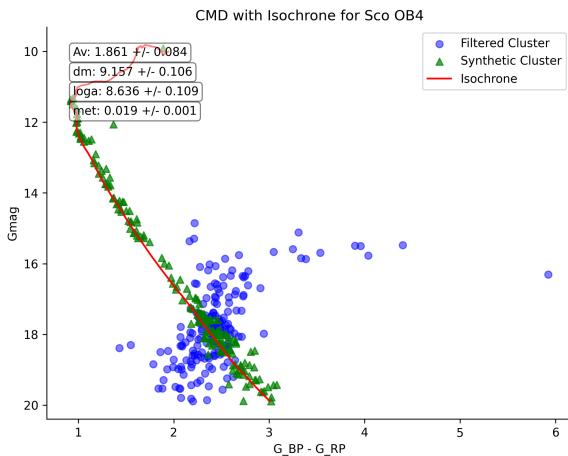
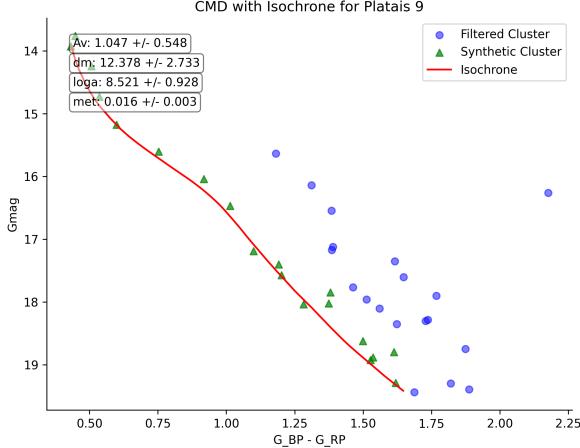












- [1] D. Janzen, R. Fu, D. Reichart, V. Kouprianov, J. Haislip, M. Fitzgerald, L. Selph, J. Torian, and F. James, Skynet cluster pro plus: An interactive web application for stellar astronomy education, *Astronomy Education Journal* **A110**, 1 (2024).
- [2] M. Wenger, F. Ochsenbein, D. Egret, P. Dubois, F. Bonnarel, S. Borde, F. Genova, G. Jasniewicz, S. Laloë, S. Lesteven, and R. Monier, The SIMBAD astronomical database. The CDS reference database for astronomical objects, *The CDS reference database for astronomical objects*, **143**, 9 (2000), arXiv:astro-ph/0002110 [astro-ph].
- [3] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, and et. al., The Gaia mission, *A&A* **595**, A1 (2016), arXiv:1609.04153 [astro-ph.IM].
- [4] Gaia Collaboration, A. Vallenari, A. G. A. Brown, and et. al., Gaia Data Release 3. Summary of the content and survey properties, *A&A* **674**, A1 (2023), arXiv:2208.00211 [astro-ph.GA].
- [5] Perren, G. I., Vázquez, R. A., and Piatti, A. E., Asteca: Automated stellar cluster analysis, *A&A* **576**, A6 (2015).
- [6] B. W. Carroll and D. A. Ostlie, *An Introduction to Modern Astrophysics* (Cambridge University Press, 2017) pp. 57–83.
- [7] This data was retrieved from Cluster Pro Plus [1].