Data Analysis Report on Open Clusters

Owen Johnson [1]

[1] *University College Dublin, UCD*

1. INTRODUCTION

Open clusters have been shown to be instrumental laboratories to explore many forms of astronomy. However, conducting analysis first requires a linear structure before more in-depth details can be explored. Figure 1 gives a brief overview of the sequence the data analysis for open-cluster science should be carried out. With each outlined step following standard practices from similar studies.
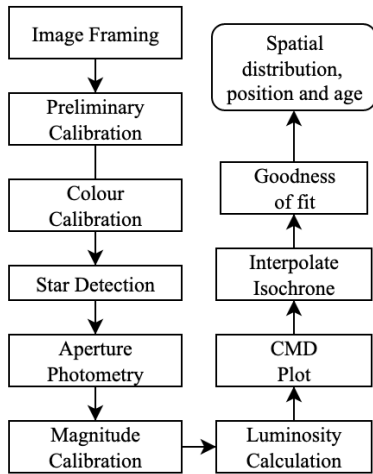

Figure 1. Overview of steps to be carried out during data analysis.

2. PRELIMINARY CALIBRATION

The conventional calibration will be carried out using flat and bias images as with all photometry. However, it is also important to consider the frame size and colour calibration. The use of the telescope at Calor Alto Observatory (CAO) for this project will make use of B, V and R filters, which means that only 50mm of the 100mm detector can be used. This leaves an observable 11' circular unvignetted frame that will need to be cropped before analysis using `numpy.s_`.
This means that about a 10' by 10' area will be suitable for observation after the crop. This will limit the size of the open cluster that can be observed and will need to be taken into consideration when choosing targets.
Color calibration is not as important to consider throughout this study as it aims to be self-contained regarding colour comparison to both experimental and archived data. However, if colour calibration proves to be a necessary step, CAO is equipped with a filter wheel reference sheet along with reference targets. [1]

3. STELLAR DETECTION

Upon first observing open clusters, each member of the clusters must be first identified. After each target has been successfully framed and reduced photometry must be performed to identify each star.
This can be completed using `DAOstarfinder` from the `photutils` python module.[2]
This operates on the DAOFIND algorithm coined by Stetson (1987). When given a threshold value `DAO` will search local density maxima for peaks that surpass this threshold. `DAO` will then fit a 2D Gaussian kernel to find objects of similar shape and size. Furthermore, the applied Gaussian kernel can be used to find the centroid and roundness by marginally fitting a 1D distribution of the Gaussian kernel to the unconvolved data image. The important return of data from the star finder is the x and y centroid positions and roundness.
Once this is complete, it is then possible to continue on the path to magnitude calculation. For a comprehensive discussion of implementation and mathematical framework, see (Howell 2006, Ch. 5.1-5.3) and Stetson (1987)

3.1. *Selecting Detection Parameters*

The two important parameters that require attention is the `threshold` and the `FWHM`.
I. The threshold determines what count number is considered the background and what count number is considered a star.
This value will depend on how noisy the image is so to ensure that they are no false detections the threshold expressed as follow,

$$\texttt{threshold} = 5 \cdot n_{\text{background}} \tag{1}$$

II. The full width at half maximum (FWHM) of the Gaussian distribution defines the resolution of a point source. FWHM is the diameter of the star's image area where intensity has fallen by half its peak value. For

---

accurate photometry at FWHM of at least two pixels is required. Where FWHM is initially taken to be 7 to find a hand full of stars. These stars can be fitted to a 2D Gaussian to find FWHM. The FWHM is updated accordingly, and the process is repeated until the desired result is attained. Where FWHM of distribution is expressed as,

$$\text{FWHM} = 2\sqrt{2\ln(2)} \cdot \sigma \qquad (2)$$

4. INSTRUMENTAL MAGNITUDE CALIBRATION

4.1. *Instrumental Apparant Magnitude Calaculation*

Next is the measurement of the brightness of each detected source. This is completed using `aperture.photometry` from `photoutils`[3]. The basic idea is that the counts measured can be directly related to magnitude. This is done by sampling a specified aperture area around each detected source along with taking the background counts around the source by defining the bounds of the annulus (Bradley et al. 2020).
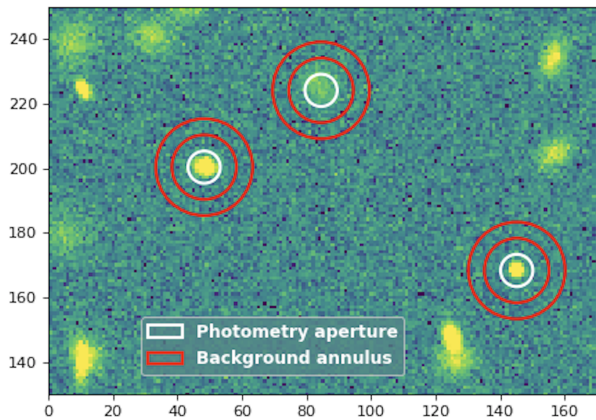

Figure 2. Apertures using `photutils`. White circle is the aperture where the counts are summed and the red annulus is where the background counts are summed. Image curtsey of Bradley et al. (2020)

The number of measured counts from `aperture` photometry is related to magnitude as follows,

$$m_{\text{std}} = -2.5\log_{10} F + \text{ZP} \qquad (3)$$

Where m_{std}, is the calibrated magnitude of a given source in the *standard* system. F, is the background-subtracted counts from a given source, and ZP is the zero point of an image.

4.2. *Insturmental Conversion*

Figure 2 shows how `aperture` handles the aperture and annulus. The aperture value is selected based on what aperture returns the lowest signal to noise ratio (SNR) during calibration. This is done by selecting a star from a frame and sampling other stars around it within a defined radius, i.e. 8 arcmins. This group of stars is then queried to a database (i.e. APASS[4]). To return stars from the selected group with the most accurate astrometric values. The varying aperture values are used to see which returns the lowest SNR. A good guess value, in this case, would be the FWHM that was previously determined. The least noisy aperture will be returned and used for subsequent analysis.

The annulus will be selected based on where a uniform background count can be attained, usually around taken with the inner and outer boundary taken as 5 and 10 pixels extended outside the aperture.
In the case where there is an overlapping of stars, the background count can be taken from a sparsely populated region of the frame. This can be easily implemented using `numpy.s_` and `for` loops in tandem with source detection.

Even though the magnitude of the sources can be calculated, they are only instrumentally apparent magnitudes. The next step is to covert each detected target into the standard photometric system. The purpose in this is to negate any discrepancies between the instrumental system and the standard system. This is done through the calculation of the zero point, ZP.

To find the zero-point, the use of reference stars is needed. A plot can be made between instrumental magnitude against the photometric magnitude to find the zero-point (Budding & Demircan 2007, Ch. 6.1) of the system, see fig. 3. If done correctly, all-instrumental bias will be minimised. This procedure needs to be repeated for each filter used.

When the reference stars have been identified eq. (3) can be re-arranged to calculate the zeropoint,

$$\text{ZP} = m_{\text{std}} + 2.5\log_{10} F \qquad (4)$$

4.3. *Statistical Uncertainty*

Photon's on a CCD obey a Poission distribution therefore before any measurements or analysis are made there is a built in uncertainty. This uncertainty is commonly referred to as signal to noise ratio (SNR). This is sim-
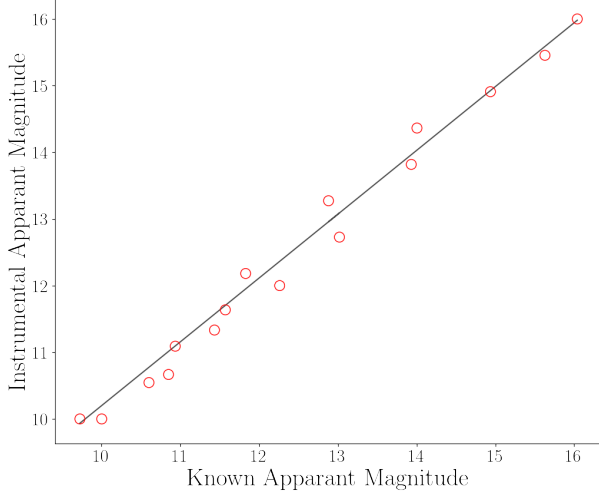
---

[3] https://photutils.readthedocs.io

[4] https://www.aavso.org/apass

Figure 3. Example plot of Magnitude calibration. Using provided data from NGC3286 and `astroquery` to use SIM-BAD magnitude values.

ply the signal from the detected star over the noise in the signal itself (Budding & Demircan 2007, Ch. 5.3). Where it is expressed as follows,

$$n = n_{\text{aperture}} \left(1 + \frac{n_{\text{aperture}}}{n_{\text{annulus}}} \right) \tag{5}$$

$$\text{SNR} = \frac{S_{\text{star}}}{\sqrt{S_{\text{star}} + n\left(S_{\text{bkg}}\right)}} \tag{6}$$

Where n is respective area of pixels and S is the respective photon count. This allows for an error estimation on magnitude to be expressed as follows,

$$\Delta m \sim \frac{1}{\text{SNR}} \tag{7}$$

4.4. *Luminosity and Distance Calculations*

Following ZP being determined, the calculation of distance and luminosity can be determined using the following equations.

$$d = 10^P; \quad P = \frac{5 - m_{\text{std}} - \text{ZP}}{5} \tag{8}$$

Where luminosity can be calculated as follows,

$$L = 4\pi d^2 F \tag{9}$$

An inquiry can also be made into the mass of main-sequence stars present in the stellar population if required by using the following equation.

$$L = L_\odot \left(\frac{m}{m_\odot} \right)^a \quad (3 \lesssim a \lesssim 4) \tag{10}$$
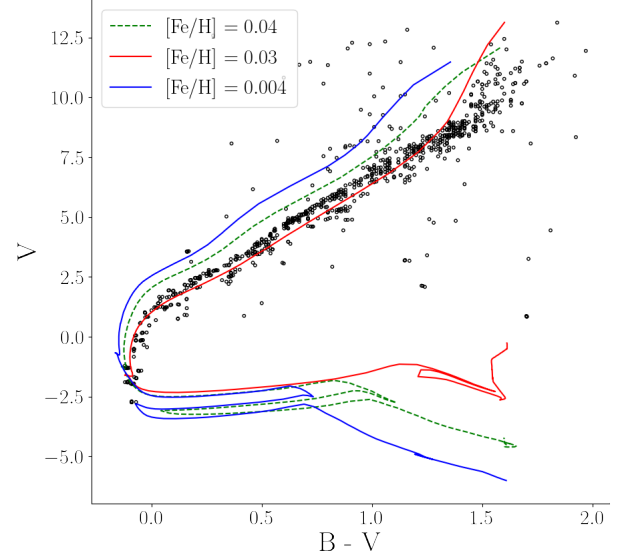

Figure 4. Example CMD plot of M45 using discussed methods using SIMBAD data. MIST isochrones were generated for a range of metallicities between $0.008 < [\text{Fe/H}] < 0.03$. 100 Myr fitted the data closely and parameters are within close range of Vandenberg (1985). All mention packages and databases are used in the creation of this plot as a proof of concept.

5. COLOUR-MAGNITUDE DIAGRAM

Colour magnitude diagrams (CMD) are a variant of Hertzsprung-Russell (HR) diagrams. The HR diagram will summarise the temperatures against magnitudes of the stellar populations, whereas CMDs is magnitude against colour. CMDs are much easier to plot as the CMD does not require the temperature of each star in a cluster. Instead, the ratio of two spectral bands intensity is used. This ratio is directly related to the black body function and related to temperature. The balance is usually expressed as the magnitude difference between two optical spectral bands, i.e. B-V. This can be similarly expanded to flux as it is easier to plot and measure flux in a standard optical spectral band (usually V). Due to this reason, CMDs are used more commonly in cluster observation. See fig. 4 as an example.

Plotting a CMD is relativity easy once proper calibration of magnitudes for the required filters has been performed. It just requires the use of `matplotlib`.

6. THEORTICAL STELLAR ISOCHRONE

Once the CMD has been plotted, it is time to turn to fit a stellar isochrone. This part of the analysis is inferred from much of the information about a stellar population. There are a considerable amount of models

for creating isochrones such as MIST[5] (Choi et al. 2016; Paxton et al. 2018).

`isochrone`[6] can be used in tandem with MIST to quickly generate isochrones with varying parameters. This is done using the `StarModel` object class, allowing tabulated pairs of varying parameters to be quickly downloaded in mass to a binary form using the `grid` extension of `starmodel`. `isochrone` uses grid interpolation based on `pandas` and `scipy` modules to produce multi-indexed isochrone data frames at fast speeds. This allows for isochrone model generation and plotting in a more efficient way than a manual generation through MIST and ultimately will allow for a different range of parameters to be fitted due to ease of use and automation of interpolating process.

Parameter	Inferred from Photometry
RA (J2000)	YES
DEC (J2000)	YES
Galaxy Longitude	YES
Galaxy Latitude	YES
Distance	YES
Distance Modulus (m_{std})	YES
Age	YES
Metallicity	ESTIMATED
Reddening	ESTIMATED

Table 1. Required parameters in calculation of a MIST theortical isochrone. YES - can be directly inferred from observations. ESTIMATED - inferral is possible but will have to be compared with other databases.

6.1. *Reddening*

Reddening is a direct result of propagation through the interstellar medium (ISM), causing light to diffuse. The extent of reddening is inversely proportional to the wavelength of the optical light. Reddening can be expressed as an excess of colour, $E(B-V)$ in a photometric system. With this absorption expressed as follows at a given wavelength,

$$A_v = R_v \, E(B-V) \tag{11}$$

Where A_v is the adsorption value at R_v is the degree of redding at a specific wavelength. Reddening will affect the stars' horizontal position when plotting the CMD as the diffuse through the ISM will decrease detected light.

In correcting reddening, the loss of light will be taken into consideration. V-band adsorption, R_v, has been estimated to be ~ 2.5 towards the Milkey Way's bulge (Nataf et al. 2013). However, this value changes based on the target position. Nevertheless, there are plentiful high-resolution spectroscopic studies to provide reliable estimations.

6.2. *Metallicity Esitmations*

Metallicity proves to be quite useful when analysing open clusters as its it's a strong indicator to what stars are part of the stellar population in the cluster itself. Using discrepencies in magnitude along with comparison with metallicity 'imposter' stars can be idenitifed. However, metallicity primarily uses spectorscopy for accurate estimations, commonly analayesd through comparioson of the Sun using a log scale,

$$\left[\frac{\text{Fe}}{\text{H}}\right] = \log_{10}\left[\frac{\text{Fe/H}}{\text{Fe/H}_\odot}\right] \tag{12}$$

Using this scale, photometric colours can be used to estimate metallicity to a reasonable degree, as shown by Karaali et al. (2011) using U, V and B filters. However, this would require the use of narrowband filters for more significant periods of observation time to attain values equivalent to high-resolution spectroscopic databases such as Tojeiro et al. (2009).

6.3. *Convection Overshooting*

Convection overshooting has been shown to be a necessary consideration for stellar models which has been most notable shown by VandenBerg & Stetson (2004). It is the also the reason why earlier stellar studies assumed that convection cores were enlarged at some given pressure height. When creating a mdoel convection cores should not exceed $r_{\max}$. The maximum size of a convection core is given as follows as,

$$\int^{r_{\max}} (L_{\text{rad}} - L)\frac{1}{T^2}\frac{dT}{dr} \, dr = 0 \tag{13}$$

Successfully correcting for overshooting produced more accurate plots for the later stages in the main sequence as shown in fig. 5. Most theoretical isochrone models readily include convection overshooting as a parameter, thus making for easier implementation onto CMD plots. Inclusion will see better fits in the later main-sequence as shown by fig. 5 and fig. 4

7. 'GOODNESS' OF ISOCHRONE FIT

In fitting the isochrone to observed data also requires a tangible way to show that the isochrone is well fitted. This is important, especially when dealing with minor
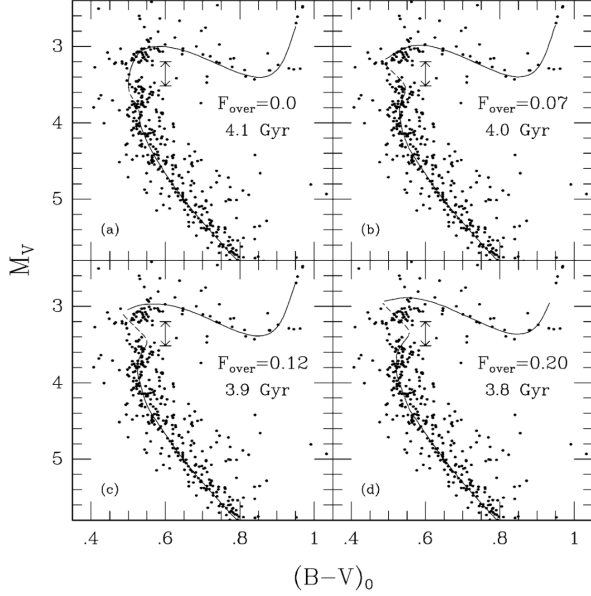
Figure 5. Theortical isochrones fitted with varying values of age and overshooting, F_{over} as shown by Vandenberg (1985)

changes of the estimated parameters in table 1. Two developmental methods have been outlined by Naylor & Jeffries (2006) and Valle et al. (2021), but there is no generally agreed-upon methodology when measuring the 'goodness' of fit. A chi-squared (χ^2) test can be performed using `scipy` and easily implemented. The following expression can be used to perform the test,

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \qquad (14)$$

Where O is the observed value, and E is the expected value. After each interpolation, the isochrone can be compared against the cluster data for each ith point taking E_i to be the closest point on the isochrone. This will be taken as the $(O - E)^2/E$ component to be summed. The code suite created by Naylor & Jeffries (2006) runs quickly and provides similar results as produced in fig. 4 with much more rigour on how well the isochrone fits. However, interpolation of F_{over} parameters does not appear to be implemented as of the most recent stable release.

8. OPEN CLUSTERS HOME IN THE GALAXY

Finally, following the calibration of both magnitude and the theoretical isochrone, an estimation of age, distance, and spatial distribution can be made. Distance and age at this point will be calculated. Along with this any imposters in the stellar population will be removed if spotted as outliers during magnitude or metallicity calibration. From here, the cluster can be classified, and comments can be made on the stellar evolution of the cluster and evolution in both the galactic bulge and arms or any of the previously discussed parameters (Friel 1995).

REFERENCES

Bradley, L., Sipőcz, B., Robitaille, T., et al. 2020, astropy/photutils: 1.0.0, 1.0.0, Zenodo, doi: 10.5281/zenodo.4044744

Budding, E., & Demircan, O. 2007, Introduction to astronomical photometry (Cambridge University Press)

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102, doi: 10.3847/0004-637X/823/2/102

Friel, E. D. 1995, Annual Review of Astronomy and Astrophysics, 33, 381, doi: 10.1146/annurev.aa.33.090195.002121

Howell, S. B. 2006, Handbook of CCD Astronomy, 2nd edn., Cambridge Observing Handbooks for Research Astronomers (Cambridge University Press), doi: 10.1017/CBO9780511807909

Karaali, S., Bilir, S., Ak, S., Yaz, E., & Coşkunoğlu, B. 2011, Publications of the Astronomical Society of Australia, 28, 95–106, doi: 10.1071/AS10026

Nataf, D. M., Gould, A., Fouqué, P., et al. 2013, ApJ, 769, 88, doi: 10.1088/0004-637X/769/2/88

Naylor, T., & Jeffries, R. D. 2006, MNRAS, 373, 1251, doi: 10.1111/j.1365-2966.2006.11099.x

Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, ApJS, 234, 34, doi: 10.3847/1538-4365/aaa5a8

Stetson, P. B. 1987, PASP, 99, 191, doi: 10.1086/131977

Tojeiro, R., Wilkins, S., Heavens, A. F., Panter, B., & Jimenez, R. 2009, The Astrophysical Journal Supplement Series, 185, 1–19, doi: 10.1088/0067-0049/185/1/1

Valle, G., Dell'Omodarme, M., & Tognelli, E. 2021, A&A, 649, A127, doi: 10.1051/0004-6361/202140413

Vandenberg, D. A. 1985, ApJS, 58, 711, doi: 10.1086/191055

VandenBerg, D. A., & Stetson, P. B. 2004, PASP, 116, 997, doi: 10.1086/426340