Finding Properties of Open Clusters Using Hertzsprung-Russell Diagrams

William Fahie
Pembroke College
February 23, 2025

---

1. Abstract

We reduce images of the star clusters NGC0663, NGC1039 and NGC6939 taken by the Philip Wetton Telescope (PWT) in Oxford and merge this with data gathered by the Gaia telescope to refine star membership and eliminate outliers. From this reduced and refined data, we extract the magnitudes of the stars for both B and V filters and plot the results on a Hertzsprung-Russell diagram. We use the PARSEC models [1] to find a best fit for the data, yielding the distances, ages and extinctions of the clusters. For NGC0663 we find the distance to be ~2510 parsecs, the age as ~20×10^6 years and the extinction (colour excess) as ~0.7. For NGC1039 we have ~1050 parsecs, ~13×10^6 years and ~0.3. Finally, for NGC6939 we have ~1910 parsecs, ~2000×10^6 years and ~0.1. The nature of our fitting process makes it difficult to estimate uncertainties for the values, as discussed in section 5, although a very approximate approach is taken at the end of section 4.3.

2. Introduction

The Philip Wetton Telescope (PWT) is a 40cm Schmidt-Cassegrain telescope located within the Oxford Astrophysics Observatory and is largely used for teaching and small-scale research projects. In this case, we examine images it has captured of star clusters in our galaxy. A star cluster is a gravitationally bound group of stars with the same origin and therefore a common age. There exist two types of clusters: open and globular, although we only study open clusters here. Open clusters are smaller than globular clusters, being 3pc – 10pc in diameter and containing up to 1000 stars [2]. They are also weakly bound meaning they disband in a timescale on the order of ~Gyr.

Raw images of stars from telescopes contain artefacts (“spurious signals”) which must be removed first, a procedure known as reduction. In our case, with the PWT, there are three steps we go through to do this: bias subtraction, dark subtraction and flat field division which are outlined in sections 3.1, 3.2 and 3.3 respectively. This is then refined using data collected from the Gaia satellite [3] in section 3.5. Gaia provides highly accurate photometry, distances, and parallaxes for stars in star clusters, so by merging the data for the relevant star clusters with that from the PWT, we can remove outliers which may be field stars i.e. not part of any stellar cluster.

We can define the main sequence for a cluster using a Hertzsprung-Russel (HR) diagram, also known as a colour-magnitude diagram, where the apparent magnitude/brightness (using either a B or V filter, which correspond to the blue and visible parts of the spectrum respectively) is plotted against its colour (B-V). The brightness is a measure of the emitted power of the star, whereas the colour is a measure of its surface temperature.

Although all the stars in each cluster have a common age, they will be of different sizes and so will be at different stages of their evolution, forming a distribution across the HR diagram. The distance, age and extinction (the dimming and reddening of light due to absorption and scattering by interstellar dust and gas) of a cluster will affect the shape of the distribution, and so by using a dataset of stellar evolution models (known as "isochrones") we can find the best fit and obtain these properties of the cluster, all of which can be found in section 4. These properties tell us more about the overall structure, age and composition of our galaxy, and allow us to test our current evolutionary models.

3. Methods

3.1 Bias subtraction

Dark frames are produced by the camera by taking images with the shutter closed. Since no external light is entering the system, any signal present must have been generated by the detector itself and thus needs to be subtracted from the real observations. The PWT generates dark frames for a range of exposure times. Bias frames are the subset of dark frames with zero exposure time so only contain the electronic offset due to the camera electronics. This offset is introduced to ensure that the natural variation of the analogue signal can never introduce a negative value.

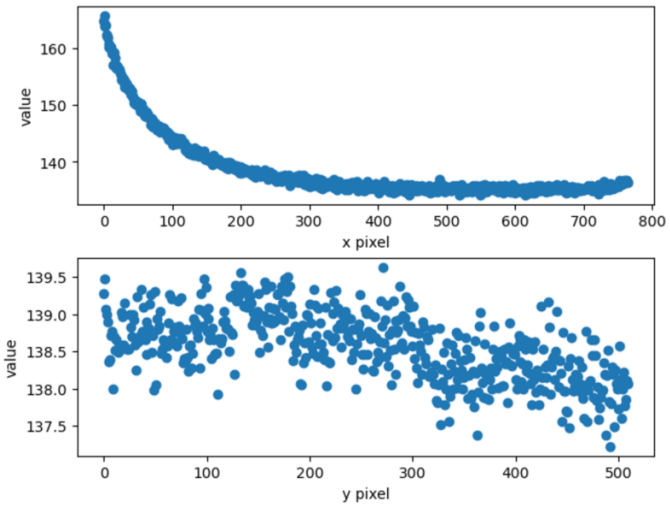

Figure 1: Average bias level along the x-axis (upper) and y-axis (lower) for a single bias frame.

As we can see in Figure 1, the bias level only becomes constant in the x direction after column ~300. We combine the bias frames to form a master bias frame, and within the domain with which the bias level is approximately constant, we find a mean bias of $\beta = 137.0$ and a standard deviation (readout noise) of $\sigma = 2.70$ (see Figure 2).

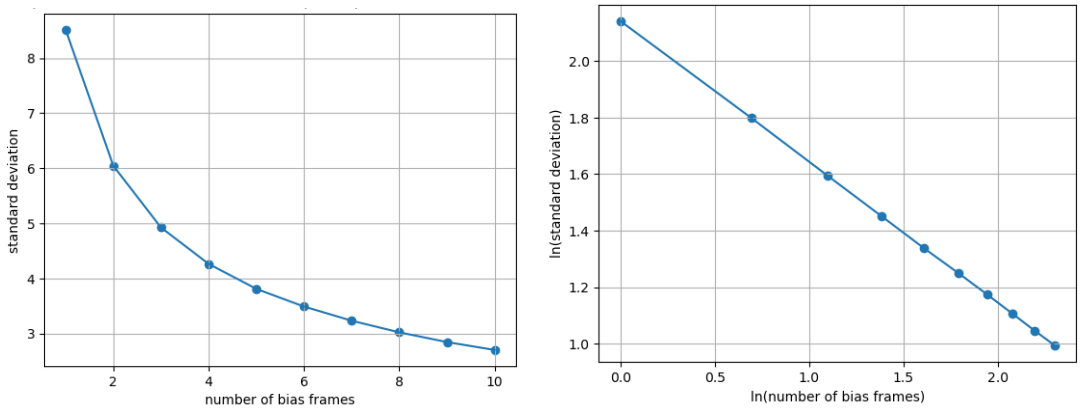

Figure 2: σ vs N (left) and ln(σ) vs ln(N) (right).

We also verify a linear relationship in the right graph of Figure 2. We find $\ln \sigma \propto -\frac{1}{2} \ln N$ i.e. the gradient is very close to -0.5, which is what we expect since $\sigma \propto \frac{1}{\sqrt{N}}$.

3.2 Dark subtraction

We now consider dark frames with a non-zero exposure time. The signals within these dark frames, known as "dark current", are generated by thermal electrons within the detector. In this experiment, we assume temperature is controlled and so the resulting spurious signals due to dark current are a function of time only. The spurious signals follow

$$counts = \beta + \alpha \times t \qquad (1)$$

where α = dark current rate, β = bias, t = exposure time. For each exposure time that the camera took, we find the mean dark level of all the dark frames, which gives the graph shown in Figure 3.

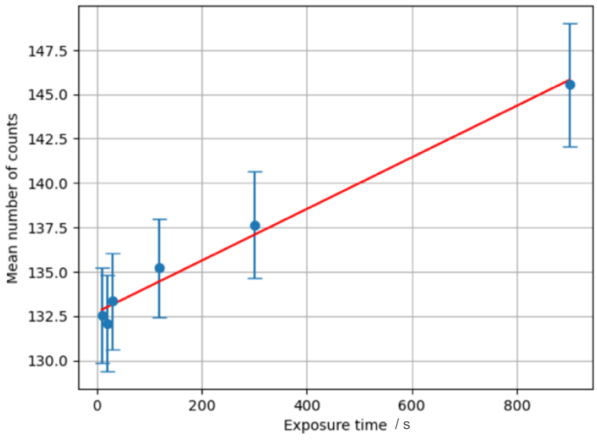

Figure 3: Mean number of counts in the dark frames as a function of exposure time. We find the gradient to be 0.014 ± 0.009 and an intercept of 132.7 ± 2.5.

This gives us a dark current rate (given by the gradient) of α = 0.014 ± 0.009 s^{-1} and a new bias level (given by the intercept) of β = 132.7 ± 2.5. This is 4.3 counts away from our previously calculated bias which is just within the error ranges of the two results. The difference might be due to thermal variations since we treated temperature as a controlled variable.

3.3 Flat field division

Finally, we consider flat field frames. These are generated by capturing a blank part of the sky which has even illumination and hence we can detect any variations in the imaging system. We account for this by dividing our science frames (the actual images of stars we are interested in) by these flat field frames. The PWT telescope captures flat field frames for both blue (B) and green (V, for visible) optical filters, which produce images around 440 nm and 550 nm respectively. As with the bias and dark frames, we average over all the flat field frames to generate master flat frames for both filters.

3.4 Reduction of science frames

We now use our master calibration frames to remove the telescope artefacts from the raw science frames. We follow the formula

$$R' = \frac{R-\beta-\alpha\times t}{F} \qquad (2)$$

where R′ = reduced frame, R = raw frame, F = flat frame and t, β, α have the same meanings as before. Figure 4 shows a comparison between a raw and reduced image.

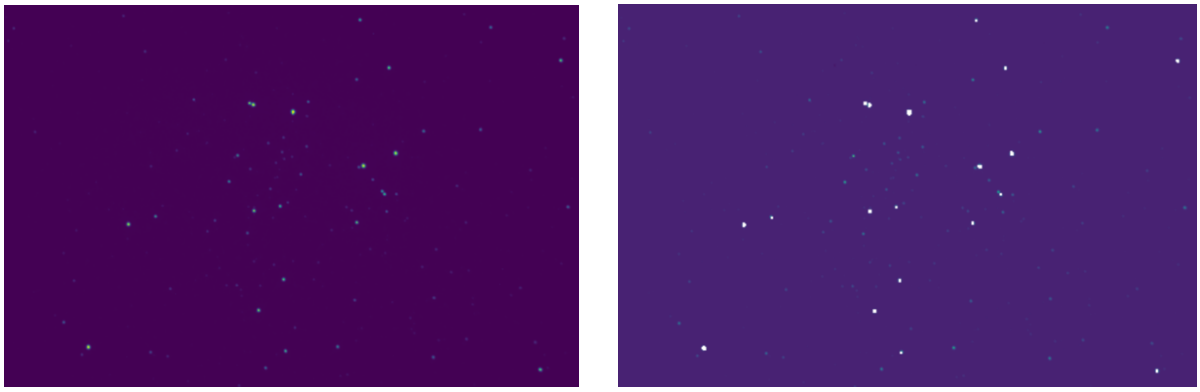

Figure 4: Raw image (left) compared to a reduced image (right) of a single image of NGC0663.

We then combine all the reduced frames of the same filter as seen in Figure 5. The angular size of the patch of the sky covered by these images is 750 arcseconds.

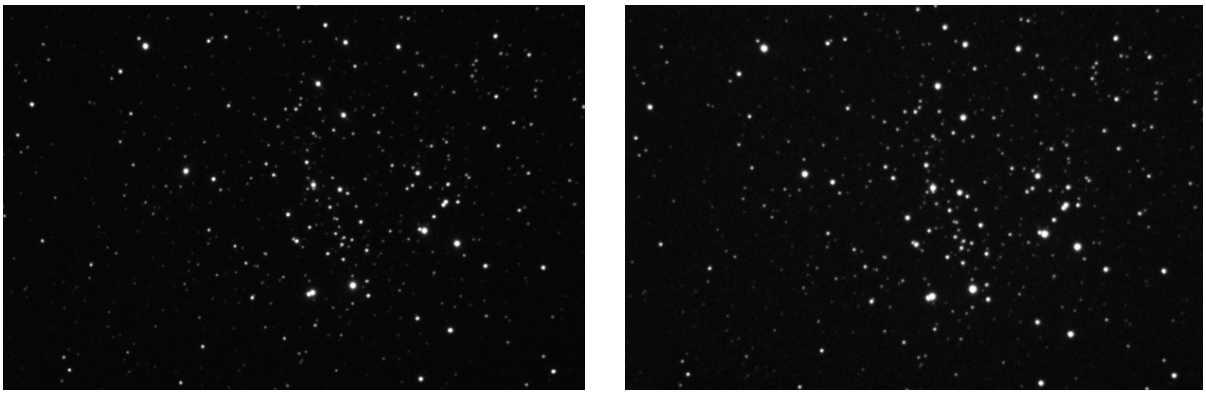

Figure 5: The combined reduced science frames for the V filter (left) and B filter (right) viewed in SAOImage DS9.

We now would like a list of the sky coordinates (RA and Dec – refer to Appendix A for more information) of all the stars in these images so we can measure and analyse their magnitudes. Since there are too many stars to identify by hand, we use the DAOFIND algorithm [4]. The results are shown in Figure 6.

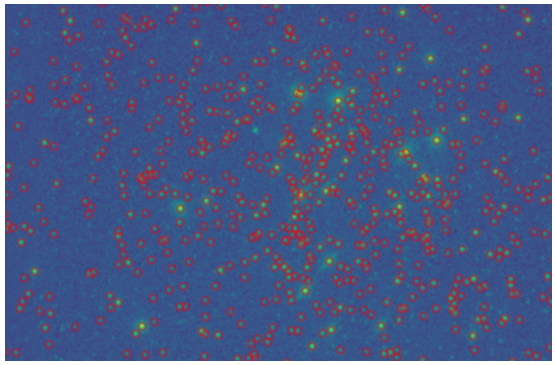

Figure 6: 634 stars identified (circled in red) by the DAOFIND algorithm performed on the reduced image of the NGC0663 cluster (on the V filter in this case). The few uncircled stars were not identified.

3.5 Merging with Gaia

As explained in section 2, not all the stars identified in our images will be members of the star clusters of interest but instead may be field stars. Following the brief description in section 2, the Gaia satellite is a space observatory launched by the European Space Agency in 2013, orbiting around the L2 Lagrange point. It is designed to map the proper motions and parallaxes of over a billion stars in our galaxy (see Appendix A for further information on proper motions and parallaxes).

We use TOPCAT (an interactive graphical viewer and editor for tabular data) to match the coordinates of our identified stars with the Gaia catalogues for NGC0663. This allows us to plot the proper motions of our stars which is shown in Figure 7. We expect all the stars in a given cluster to have similar proper motions and thus appear as a tight bunch, whereas the field stars are spread out more randomly. Within TOPCAT we define a subset of stars (using a simple lasso tool) which are more likely to be members of the cluster. Our new subset contains 535 stars and so we eliminated 99 field stars.

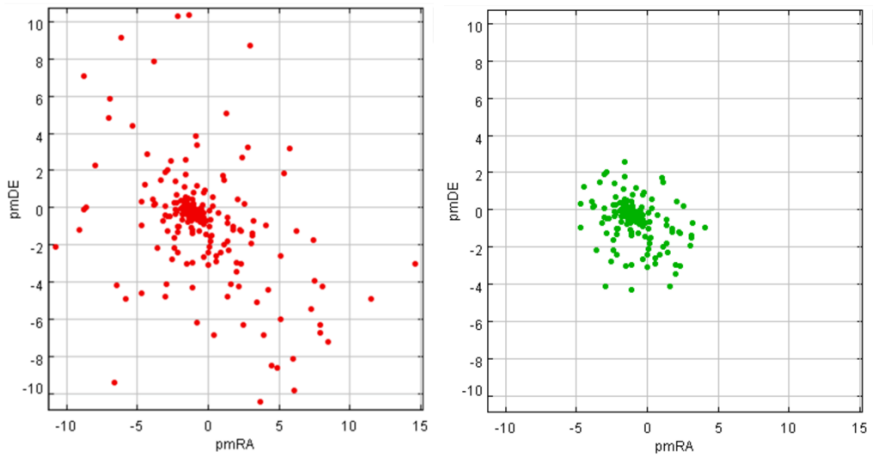

Figure 7: Proper motions of stars captured by the PWT's image of NGC0663. The left (red) shows all the stars, whereas the right (green) shows a refined version where 99 field stars have been omitted.

Having developed an effective reduction process, we repeat all the steps so far for two more star clusters captured by the PWT: NGC1039 and NGC6939. Because these clusters are closer to us, we can use Gaia's parallax data to refine the cluster even further. We plot these as histograms in Figure 8. As with proper motions, we expect the parallaxes of stars in the same cluster to be similar, and so again we define subsets which eliminate likely outliers.

For NGC6939 we identify 763 stars and thus eliminate 192 stars. For NGC1039, it is close enough that we can clearly distinguish the peak centred on zero parallax due to the field stars from the peak due to the cluster. Hence, we can eliminate as many as 807 out of 1039 stars.

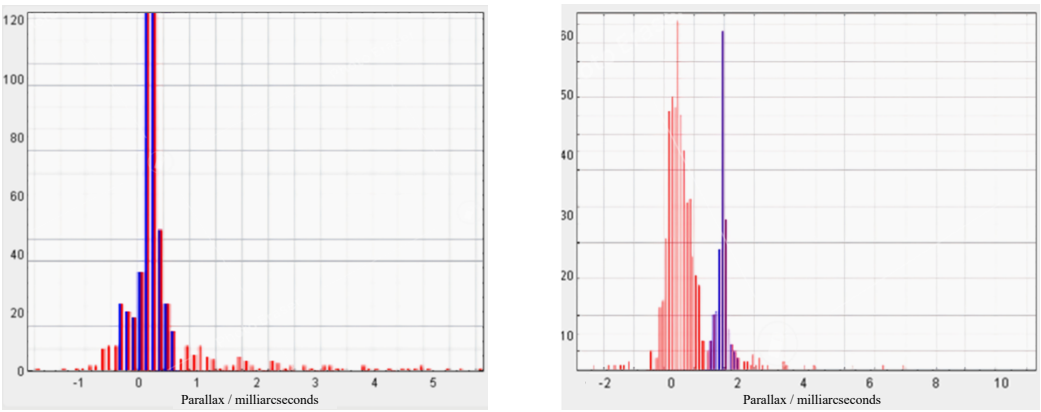

Figure 8: Histograms of the parallaxes for stars in NGC6939 (left) and NGC1039 (right). The x-axis is the parallax measured in milliarcseconds (see Appendix B) and the y-axis is frequency as usual for histograms.).The red bars correspond to field stars we have eliminated whereas the blue bars are our refined subset.

4. Results

4.1 Photometry

Now that we have a refined list of star coordinates, we measure the fluxes of the stars within the reduced images, known as photometry. This is done by measuring the counts (corresponding to photons) within a defined circular aperture. We use QFitsView to fit a 2D elliptical Gaussian so we can measure the Full Width Half Maximum (FWHM) for a handful of stars in the image.

Table 1: FWHM for 4 random stars in the reduced NGC0663 image.

FWHM / px	**Angular size / arcsec**
2.406	3.609
2.482	3.723
2.388	3.582
2.379	3.569

The information in Table 1 shows us that the FWHM is ~2.4 px which, since each pixel is 1.5 arcsec, corresponds to an angular size of ~3.6 arcsec. This is a rough figure used to define the aperture within which we measure the flux, so the exact number and error are not important.

Given that NGC0663 is ~2100 parsecs away from Earth, the angular size of a sun-like star would be ~ 4.4 x 10^{-6} arcsec. Again, the exact number is not important, but it is worth commenting on the order of magnitude. Although the true angular size of stars is much closer to being point-like (~10^{-6} arcsec), the reason the stars in our image are so much larger (~1 arcsec) is because the size we *observe* through a telescope is dominated by the point spread function of the telescope and the atmospheric blurring known as "seeing". These dominating effects apply to all stars which is why they all have roughly the same angular size. We take the optimal radius of the aperture to be 2.5 × FWHM ~6 px.

$$magnitude = zeropoint - 2.5 \times log_{10}(count\ rate) \quad (3)$$

Using the known magnitudes [2] for three stars in the image (1477, 1601, 2615) and measuring their respective count rates within an aperture defined by the derived radius above, we can solve for the zeropoint using Equation 3 (see Appendix B for more on this equation). The results are shown in Tables 2 and 3, and this data gives us a B zeropoint of 19.51 $\pm$ 0.65 and a V zeropoint of 19.75 $\pm$ 0.02.

Table 2: Calculated zeropoints for B for NGC0663.

Star ID#	**B magnitude**	**B counts per sec**	**B filter zeropoint**
1477	10.347	4628.0	19.510
1601	9.798	13885.1	20.154
2615	9.139	7756.2	18.863

Table 3: Calculated zeropoints for V for NGC0663.

Star ID#	**V magnitude**	**V counts per sec**	**V filter zeropoint**
1477	9.590	11255.0	19.718
1601	9.002	31331.2	19.777
2615	8.537	19720.0	19.739

Using Equation 3 with these zeropoints and applying it to our refined lists of stars for the three clusters, we derive the B and V magnitudes for all the stars in our images.

4.3 HR analysis

The colour of a star is quantified by the difference between the B and V magnitude i.e. a smaller B-V means the star is bluer while a bigger B-V means the star is redder. We plot our V magnitudes (y-axis) against the colour B-V (x-axis) for our refined list of stars and fit our evolutionary PARSEC models to them as shown in Figures 9, 10 and 11. These models appear as isochrones which are lines of constant time that represent the expected relationship between colour and brightness for a cluster. We fit these isochrones using sliders shown in Figure 12.

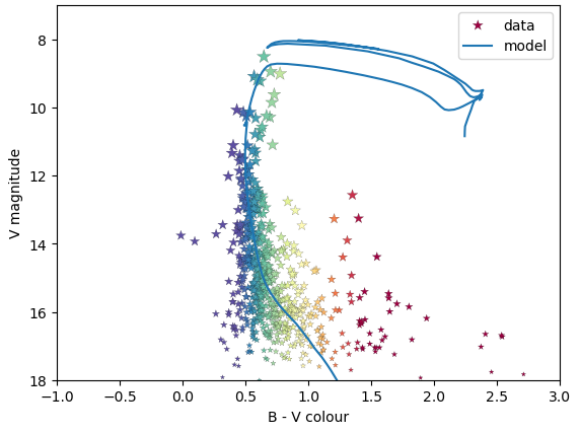

Figure 9: HR diagram for NGC0663.

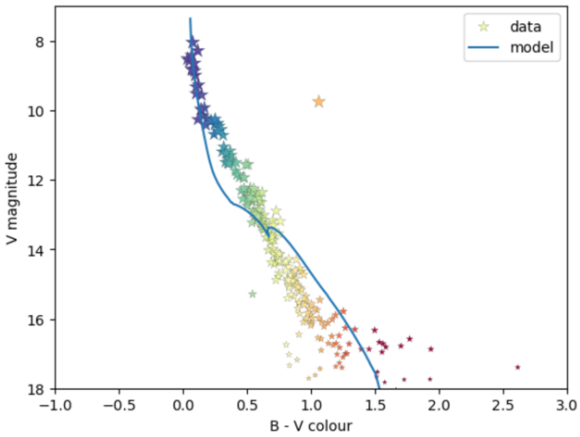

Figure 10: HR diagram for NGC1039.

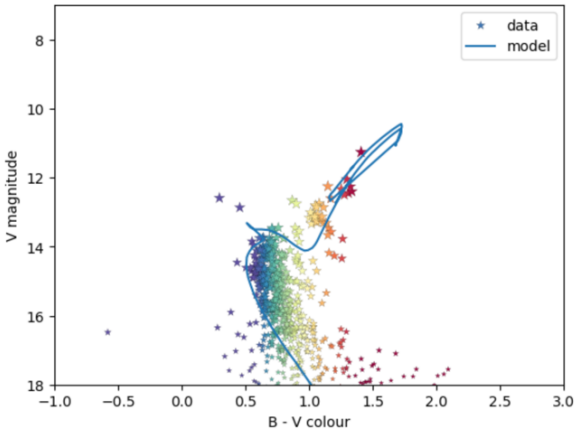

Figure 11: HR diagram for NGC6939.

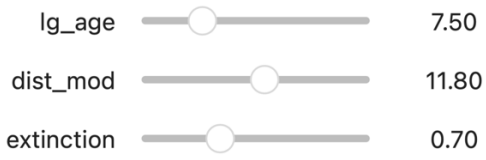

Figure 12: Isochrone fitting sliders.

We find that increasing the distance modulus shifts the model line vertically downwards, which makes sense as a greater distance means the apparent brightness of the star is weaker, hence its observed magnitude is greater (magnitudes are defined inversely with brightness, with the smallest magnitudes/brightest stars at the top – see Appendix B). The distance modulus parameter is given logarithmically as $5\,log_{10}\frac{d}{10}$ (this is the convention for defining distance modulus – again see Appendix B for more on this).

Increasing the extinction shifts the line down and to the right. This also makes sense since extinction is due to interstellar dust which absorbs and scatters light, and this tends to affect blue light (high frequency) more than red. Extinction is quantified by the "colour excess" E(B-V) which is defined in Equation 4, where B-V is a measure of colour as explained earlier.

$$E(B-V) = (B-V)_{observed} - (B-V)_{intrinsic} \quad (4)$$

Finally, adjusting the age affects the turn-off point, which is where the stars leave the main sequence and move towards the red giant branch i.e. towards the top right corner of the HR diagram. More massive stars are brighter and bluer in colour and evolve out of the main sequence faster. This means a young cluster is more likely to have a higher proportion of massive stars which haven't left the main sequence yet, so the turn-off point must be closer to the top left (brighter, bluer). The age parameter is given logarithmically as $\log_{10} t$ (i.e. lg_age=8.0 is 1×10^8 years old).

Table 4: Summary of results obtained from the HR diagrams in Figures 9, 10 and 11.

NGC#	**Distance modulus / log(parsecs)**	**Distance / parsecs**	**Logarithmic age / log(years)**	**Age / 10^6 years**	**Extinction ("colour excess")**
0663	12.0	2510	7.3	20	0.7
1039	10.1	1050	7.1	13	0.3
6939	11.4	1910	9.3	2000	0.1

The results are summarised in Table 4. The most dominant source of error comes from fitting the isochrones to the data since this is done manually and is therefore very subjective. For a very approximate error, we can use the resolution of the parameter sliders which is 0.1 for all of them. For distance and age, which are defined logarithmically, we can use Equation 5 to get a percentage uncertainty of ~20% for all values.

$$\Delta \log_{10} x \approx \frac{\Delta x}{x \ln 10} \rightarrow \frac{\Delta x}{x} \approx \ln 10 \cdot \Delta \log_{10} x \quad (5)$$

5. Conclusion

We have successfully reduced and refined our data and fit models to them to estimate the age, distance and extinction for the star clusters NGC0663, NGC1039 and NGC6939. As discussed at the end of section 4, since the data from Gaia is highly accurate, the most dominant source of error in our results arises from fitting the isochrones to the data, not just due to the resolution of the fitting parameters (which is 0.1 for each one) but simply from manually choosing the best fit. This is difficult to quantify since the best fit is performed by eye and consequently highly subjective, so we accept that our results are approximate values ("ball parks") for the properties.

An extension to this work could be to use an isochrone dataset which allows parameter adjustment with a smaller resolution. One could then use some kind of fitting algorithm, perhaps a least-squares approach which minimises the error (maybe using a machine learning algorithm like backpropagation). Not only would this automate the best fit selection and reduce human bias, but this minimum error would give us a measure of uncertainty in our results which we do not currently have.

Our use of colour-magnitude/HR diagrams was sensible, and a better choice than a magnitude-magnitude diagram (i.e. B against V), as the B-V axis in an HR diagram gives a measure of colour and thus temperature (blue means hotter, and red means cooler) which is very useful to relate to brightness. Merging the PWT's data with Gaia's catalogue was also a highly important step since fitting the data is already the predominant source of uncertainty, and this would have only been more difficult if the distribution included unnecessary field stars. Overall, the results we obtained in Table 4 are consistent with published values for the three clusters NGC0663 [5], NGC1039 [6], NGC 6939 [7], supporting the reliability of the PWT data as well as our analysis and fitting methods. While there is certainly scope for refinement, this experiment successfully demonstrates the methodology for determining star cluster properties.

6. References

[1] A. Bressan et al., 2012, *MNRAS*, 427, 127. PARSEC: Stellar Tracks and Isochrones with the Padova and Trieste Stellar Evolution Code.

[2] Oxford University Physics. [Online]. Available: https://www-teaching.physics.ox.ac.uk/practical_course/scripts/srv/local/rscripts/trunk/Astro/AS35/AS35.pdf. Accessed: 28 April 2025.

[3] European Space Agency (ESA). Gaia Mission Overview. [Online]. Available: https://www.esa.int/Science_Exploration/Space_Science/Gaia. Accessed: 28 April 2025.

[4] P. B. Stetson, 1987, *PASP*, 99, 191. DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry.

[5] A. Wallenquist, 1929, *Uppsala Astronomiska Observatoriums Meddelanden*, no. 42.

[6] W. Dieckvoss, 1954, *Astronomische Nachrichten*, 282, 25.

[7] F. Küstner, 1923, *Veroeffentlichungen der Universitäts-Sternwarte Bonn*, no. 18.

Appendix A – Parallax and proper motion

Proper motion is the apparent motion of a star across the sky i.e. the sideways motion across our view, not the motion towards or away from us. Gaia measures this with two components: pmRA = proper motion in Right Ascension (east-west direction), and pmDE = proper motion in Declination (north-south direction).

We can use proper motion to refine cluster membership for most clusters. Whereas for closer stars, parallax becomes a particularly useful tool. Parallax is the apparent shift of a star's position when observed from two different positions. Gaia gives the apparent shift of the star's position over 6 months, in milliarcseconds. The closer the star, the greater the parallax angle it will experience.

Appendix B – Brightness and magnitude

There is often some confusion in astronomy surrounding magnitude and brightness. In the magnitude system, smaller numbers generally mean brighter stars and vice versa, a convention that dates to Hipparchus (~150 BC) who classified the brightest stars as magnitude 1, and the faintest as magnitude 6. Later astronomers then defined the observed magnitude of light emitted by an object as

$$m_2 - m_1 = -2.5 \log_{10} \frac{F_2}{F_1}$$

where m = observed magnitude and F = flux. The inverse nature is retained through the minus sign. This explains Equation 3 in section 4.1 which we repeat here:

$$magnitude = zeropoint - 2.5 \times log_{10}(count\ rate)$$

The 2.5 factor is conventional and originates from the Pogson scale where a difference in 5 magnitudes corresponds to a factor of 100 in brightness.

$$10^{\frac{5}{x}} = 100 \rightarrow x = 2.5$$

Given that flux is inversely proportional to distance squared, we have

$$m_2 - m_1 = -2.5 \log_{10}(\frac{d_1}{d_2})^2 = 5 \log_{10} \frac{d_2}{d_1}$$

where d = distance. In section 4.3 we give an expression for distance modulus. This is defined as the difference between the observed magnitude m and the absolute magnitude M, since this quantity depends on distance alone. The absolute magnitude is defined as the observed magnitude a star would have if it were exactly 10 parsecs away, hence the distance modulus is given by the following equation.

$$m - M = 5 \log_{10} \frac{d}{10}$$