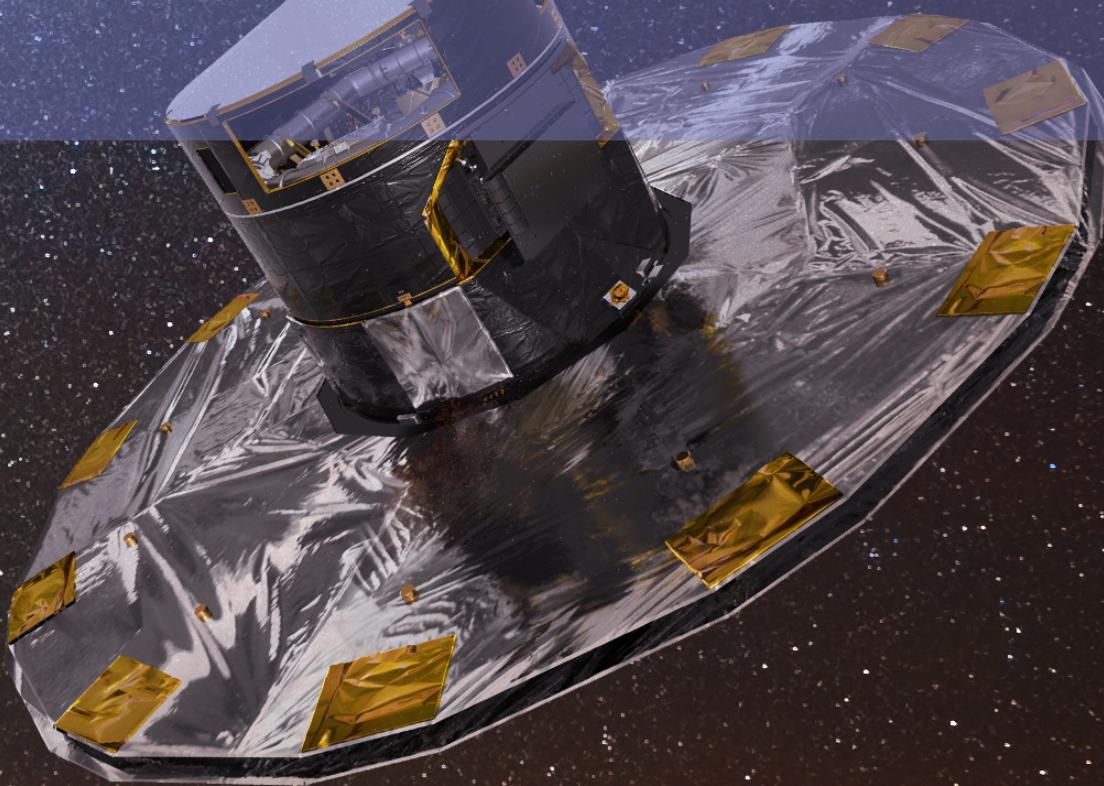


KRITTIKA SUMMER PROJECTS 2023

# Gaia Data Analysis

Aditya Saran, Arnav Jain, Aryan Kumar, Navdha,  
Ravi Kumar, Tamajeet Roychowdhury, Tanmay Patil,  
and Tanishk Mohan





Aditya Saran<sup>1,2</sup>, Arnav Jain<sup>2</sup>, Aryan Kumar<sup>1,2</sup>, Navdha<sup>1,2</sup>, Ravi Kumar<sup>1,2</sup>,  
Tamojeet Roychowdhury<sup>1,2</sup>, Tammay Patil<sup>2</sup>, and Tanishk Mohan<sup>1,2</sup>

<sup>1</sup>Krittika - The Astronomy Club of IIT Bombay, Powai, Mumbai - 400076, India

<sup>2</sup>Indian Institute of Technology Bombay, Mumbai - 400076, India

Copyright © 2023 Krittika IITB

PUBLISHED BY KRITTIKA: THE ASTRONOMY CLUB OF IIT BOMBAY

[GITHUB.COM/KRITTIKAIITB](https://github.com/KrittikaIITB)

First Release, July 2023

# Abstract

Gaia is a space observatory of the European Space Agency (ESA) which was launched in 2013 and expected to operate until 2025. It aims to build the most accurate map of our galaxy so far, by mapping the positions, distances and proper motions of the stars with enormous precision. It currently operates around the Sun-Earth L2 Lagrangian point.

All of Gaia's data is accessible publicly via ADQL interface on Python, using the package `astroquery.Gaia`. This allows us to study the astrometric and certain photometric characteristics of over a billion stars upto magnitudes of 21 in the Gaia catalog using our computers. The Gaia Data Release 3 was released on 13 June 2022.

This project aims to use the data for specific types of stars, and then to use the trends observed to arrive at some interesting qualitative results for those chosen stars.



# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Collecting Data	6
1.2	Coordinate Systems	8
1.2.1	Galactic coordinate system	8
1.2.2	ICRS coordinate system	9
1.3	Photometry	9
1.3.1	Flux	9
1.3.2	Magnitude scale	9
1.3.3	Using filters	10
1.4	Astrometry	10
1.4.1	Parallax	10
1.4.2	Proper Motion	11
1.5	HR Diagrams - the Physics of Stars	11
1.5.1	HR Diagrams	12
1.5.2	Life Cycle of Stars	14
<b>2</b>	<b>The Study of Open Clusters</b>	<b>18</b>
2.1	Introduction	18
2.2	Selecting Cluster Members	19
2.2.1	ADQL Query	19
2.2.2	Beehive cluster	20
2.2.3	M42 Orion	21
2.2.4	Blanco 1 cluster	21
2.2.5	M7	22
2.2.6	M6	23
2.2.7	M34 Spiral Cluster	24

2.2.8	Messier 35 .....	25
2.2.9	M48 .....	26
2.2.10	Trapezium Cluster .....	27
2.2.11	Messier 45 .....	28
2.2.12	Messier 50 .....	29
2.2.13	Messier 38 .....	29
2.2.14	Lambda Orionis .....	30
2.2.15	Alpha Persei Cluster .....	30
<b>2.3</b>	<b>Age, Metallicity and Isochrones</b>	<b>30</b>
<b>2.4</b>	<b>Spectral Classification using HR Diagrams</b>	<b>32</b>
2.4.1	Blanco 1 cluster .....	32
2.4.2	Beehive cluster .....	33
2.4.3	Southern Beehive cluster .....	33
2.4.4	Wishing Well cluster .....	34
2.4.5	Messier 7 .....	35
2.4.6	Messier 35 .....	35
2.4.7	NGC 752 .....	36
2.4.8	IC 4651 .....	37
2.4.9	Messier 67 .....	37
2.4.10	NGC 188 .....	38
2.4.11	M48 .....	39
2.4.12	Trapezium Cluster .....	40
2.4.13	Messier 45 .....	41
2.4.14	Messier 50 .....	42
2.4.15	Messier 38 .....	43
2.4.16	Lambda Orionis .....	44
2.4.17	Alpha Persei Cluster .....	45
<b>2.5</b>	<b>Distribution in Position and Velocity Space</b>	<b>46</b>
2.5.1	Blanco 1 cluster .....	46
2.5.2	Beehive cluster .....	47
2.5.3	Southern Beehive cluster .....	48
2.5.4	Wishing Well Cluster .....	49
2.5.5	Messier 35 .....	50
2.5.6	NGC 752 .....	51
2.5.7	IC 4651 .....	52
2.5.8	Messier 67 .....	53
2.5.9	M48 .....	54
2.5.10	Trapezium Cluster .....	55
2.5.11	NGC 188 .....	55
<b>2.6</b>	<b>Comparing ages of different clusters</b>	<b>56</b>
<b>2.7</b>	<b>Empirical Initial Mass Function</b>	<b>57</b>
<b>3</b>	<b>The Study of Globular Clusters</b>	<b>60</b>
<b>3.1</b>	<b>Need for Improvised Quality Cuts</b>	<b>60</b>
<b>3.2</b>	<b>Unexplored Parts of HR Diagrams - Evolved Stellar Stages</b>	<b>64</b>
3.2.1	Red Clump .....	64
3.2.2	Blue Stragglers .....	64
3.2.3	RR Lyrae Variables .....	65

<b>3.3</b>	<b>Distance Determination with isochrones</b>	<b>65</b>
3.3.1	Using RR Lyrae stars .....	65
3.3.2	Using the Red Clump .....	66
3.3.3	Using the turn-off point .....	68
<b>3.4</b>	<b>Inferring Properties from Isochrones</b>	<b>68</b>
3.4.1	Messier 3 .....	70
3.4.2	Messier 4 .....	70
3.4.3	Messier 5 .....	71
3.4.4	Messier 13 .....	71
3.4.5	Messier 68 .....	72
3.4.6	Messier 79 .....	72
3.4.7	Messier 92 .....	73
3.4.8	Messier 107 .....	73
3.4.9	NGC 6541 .....	74
3.4.10	NGC 104 .....	74
3.4.11	Omega Centauri (NGC 5139) .....	75
3.4.12	Messier 30 .....	75
3.4.13	47 Tucanae .....	76
<b>3.5</b>	<b>Summary</b>	<b>77</b>
<b>4</b>	<b>Analysis of Magellanic Clouds</b>	<b>78</b>
<b>4.1</b>	<b>Large Magellanic Cloud</b>	<b>78</b>
<b>4.2</b>	<b>Small Magellanic Cloud</b>	<b>83</b>
<b>4.3</b>	<b>Magellanic Bridge</b>	<b>88</b>
<b>5</b>	<b>Time-Series Analysis - Auxiliary Data Products</b>	<b>91</b>
<b>5.1</b>	<b>Variable stars</b>	<b>91</b>
5.1.1	Cepheids .....	91
5.1.2	Long Period Variables (LPVs) .....	92
<b>5.2</b>	<b>Analysing Clusters using Spectroscopy data</b>	<b>92</b>
5.2.1	Blanco 1 .....	92
5.2.2	Beehive cluster .....	94
<b>5.3</b>	<b>Analysing Cepheids</b>	<b>97</b>
5.3.1	Individual stars .....	97
5.3.2	Period Luminosity Relationship .....	98
<b>6</b>	<b>Results and Discussion</b>	<b>100</b>
<b>7</b>	<b>References</b>	<b>103</b>



# 1. Introduction

## 1.1 Collecting Data

Gaia will perform its observations from a controlled Lissajous-type orbit around the L2 Lagrange point of the Sun and Earth-Moon system. During its 5-year operational lifetime, the satellite will continuously spin around its axis, with a constant speed of 60 arcsec/s. As a result, over a period of 6 hours, the two astrometric fields of view will scan across all objects located along the great circle ‘perpendicular’ to the spin axis. As a result of the basic angle of  $106.5^\circ$  separating the astrometric fields of view on the sky, objects transit the second field of view with a delay of 106.5 minutes compared to the first field.

Gaia has two telescopes with a fixed angular separation of  $106.5^\circ$ . Both of them rotate at the same angular velocity around an axis perpendicular to line of sight of the two telescopes. Light from both telescopes are focused onto a plane of CCDs using a mirror. The CCD array has 106 individual CCDs, and almost 1 billion pixels.

When light from the first telescope falls onto the CCD array, the first line of CCD is activated and selects the stars from the first telescope to be projected on the main CCD grid. Similarly for the second telescope, the second CCD line is activated.

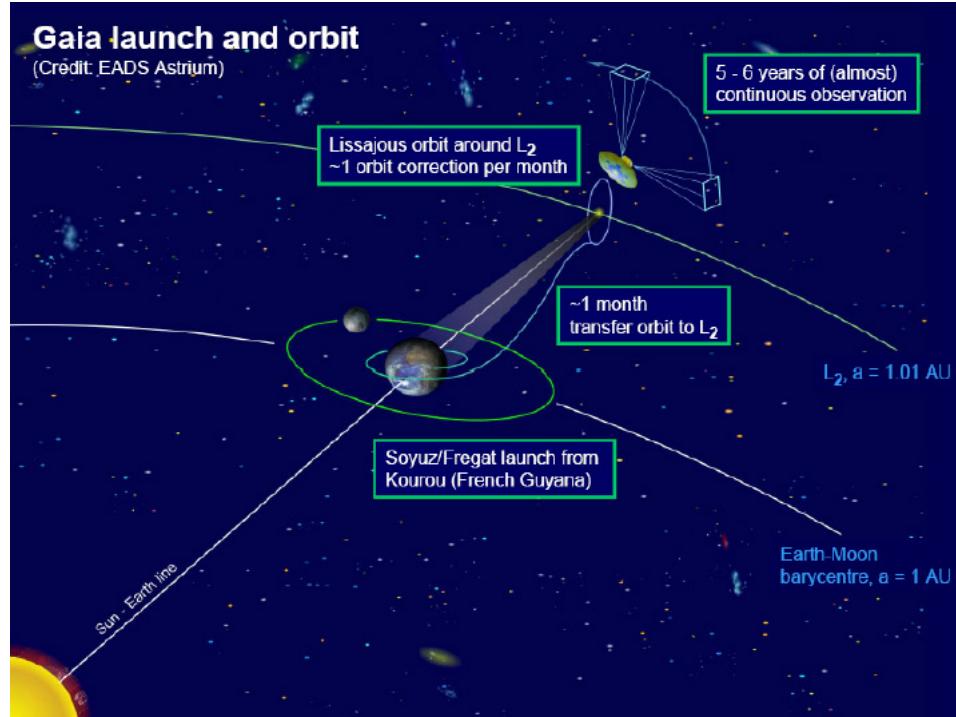


Figure 1.1: Orbit of Gaia around Sun-Earth L2

As the telescope rotates, light from each single star moves across the CCD grid. There are three main measurements made:

1. The first grid of pixels is the *astrometric field*. It tracks the position of the star. As the star moves along each CCD in the CCD grid, the on-board computer finds out its position using the CCD pixels and also finds out its brightness. As the starlight passes through the 9 columns of CCD, each column measures its position and brightness. Multiple measurements help to minimize the error, and also optimizes the amount of data needed to be sent.
2. The next two columns are meant for photometry. Light from the stars is passed through two special prisms, that directs the smaller and larger wavelength components of the star's light onto the blue and red photometers respectively. This provides information about the star's temperature (Wien's law), size and chemical composition (relative abundance of hydrogen, helium etc. using intensity of their spectral lines).
3. Light from the brighter stars is captured on the radial velocity spectrometers, which measure the stars' radial velocity using Doppler effect (using hydrogen spectral lines as the base wavelength).
4. Additionally, as Gaia moves along with Earth around the Sun, it also records each star's parallax, thus allowing us to find the distance to that star.

The collected information from the CCDs is compressed into a data packet to be stored on Gaia's onboard computer, and later transmitted back to ESA's ground stations. This process is repeated for each star as the telescopes rotate. Around 3 million stars are measured every hour. Gaia maps over a billion different stars, up to a limiting magnitude of about 20, and would thus catalogue about 1% of our galaxy's stars. It is also expected to trace numerous asteroids and comets in our Solar System, exoplanets, brown dwarfs, supernovae and quasars - apart from numerous stars from other galaxies (Magellanic Clouds and farther in the Local

Group). Each star is mapped almost 70 times over the course of 5 years before the processed data is released.

## 1.2 Coordinate Systems

### 1.2.1 Galactic coordinate system

The **galactic coordinate system** is a celestial coordinate system in spherical coordinates, with the **Sun** as its **center**, the primary direction aligned with the approximate center of the Milky Way Galaxy, and the fundamental plane parallel to an approximation of the galactic plane but offset to its north. It uses the right-handed convention, meaning that coordinates are positive toward the north and toward the east in the fundamental plane.

#### Galactic longitude:

Longitude ( $\ell$ ) measures the angular distance of an object eastward along the galactic equator from the galactic center. Analogous to terrestrial longitude, galactic longitude is usually measured in degrees( $^{\circ}$ ).

#### Galactic latitude:

Latitude ( $b$ ) measures the angle of an object northward of the galactic equator (or midplane) as viewed from Earth. Analogous to terrestrial latitude, galactic latitude is usually measured in degrees( $^{\circ}$ ).

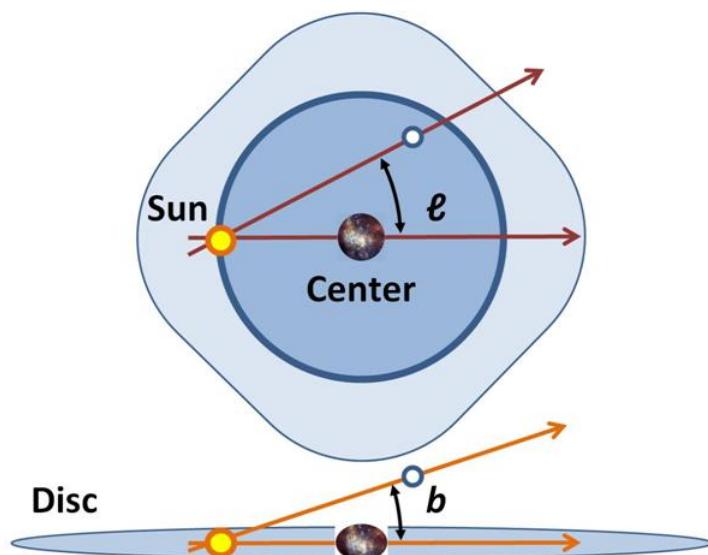


Figure 1.2: Galactic coordinate system

Here, in the from the above figure we can see that the galactic longitude is measured from the line joining the sun and galactic center and it is represented as  $\ell$ .

Similarly, latitude is measured above the galactic equator or midplane and it is represented by  $b$ .

### 1.2.2 ICRS coordinate system

- The **International Celestial Reference System (ICRS)** is the current standard celestial reference system adopted by the **International Astronomical Union (IAU)**.
- Its origin is at the **barycenter** of the **Solar System**.  
**Barycenter** is the center of mass of two or more bodies that orbit one another and is the point about which the bodies orbit.  
Unlike galactic coordinate system (which has sun as origin), ICRS has barycenter of solar system as origin.

## 1.3 Photometry

### 1.3.1 Flux

Luminosity: Luminosity is the energy released by a star in one second. It is generally denoted as  $L$ .

Flux is defined as:-

$$F = \frac{L}{4\pi r^2}$$

where  $r$  is distance of star from earth

### 1.3.2 Magnitude scale

#### Apparent magnitude

The stars' are classified on the basis of their apparent magnitude using the magnitude scale.

Apparent magnitude is defined as:-

$$m = -2.5 \log_{10} \left( \frac{F}{F_0} \right)$$

here,  $F_0$  is the reference flux of a source of our choice. Since the negative log is used in the definition, stars with higher magnitude values are fainter. The brightest object in the sky, the Sun, has an apparent magnitude of -26.74. The brightest star, Sirius, has an apparent magnitude of -1.46. The star Vega was historically assigned the baseline apparent magnitude of 0.00, but this was further refined later with other definitions when Vega was discovered to have slight variations in its luminosity.

#### Absolute magnitude

The apparent magnitude of a star is good enough to tell us about its brightness as seen from Earth, but it does not tell us anything about its intrinsic luminosity i.e. the total energy it radiates, due to the distance factor in the flux. So we use the absolute magnitude, which is the magnitude of a star as seen from a distance of 10 parsecs (1 parsec = 3.26 light years). By comparing the absolute magnitudes of two different stars, we directly have a measure of their luminosities as the distance factor cancels out.

$$M = m + 5 - 5 \log_{10} r$$

where  $r$  is the distance of the star from Earth in parsecs.

### 1.3.3 Using filters

#### Red pass filter

This filter allows higher wavelengths to pass through and blocks the lower wavelengths. The flux is available in the form of red pass apparent magnitude, denoted as  $G_{RP}$ .

#### Blue pass filter

This filter allows lower wavelengths to pass through and blocks the higher wavelengths. The flux is available in the form of blue pass apparent magnitude, denoted as  $G_{BP}$ .

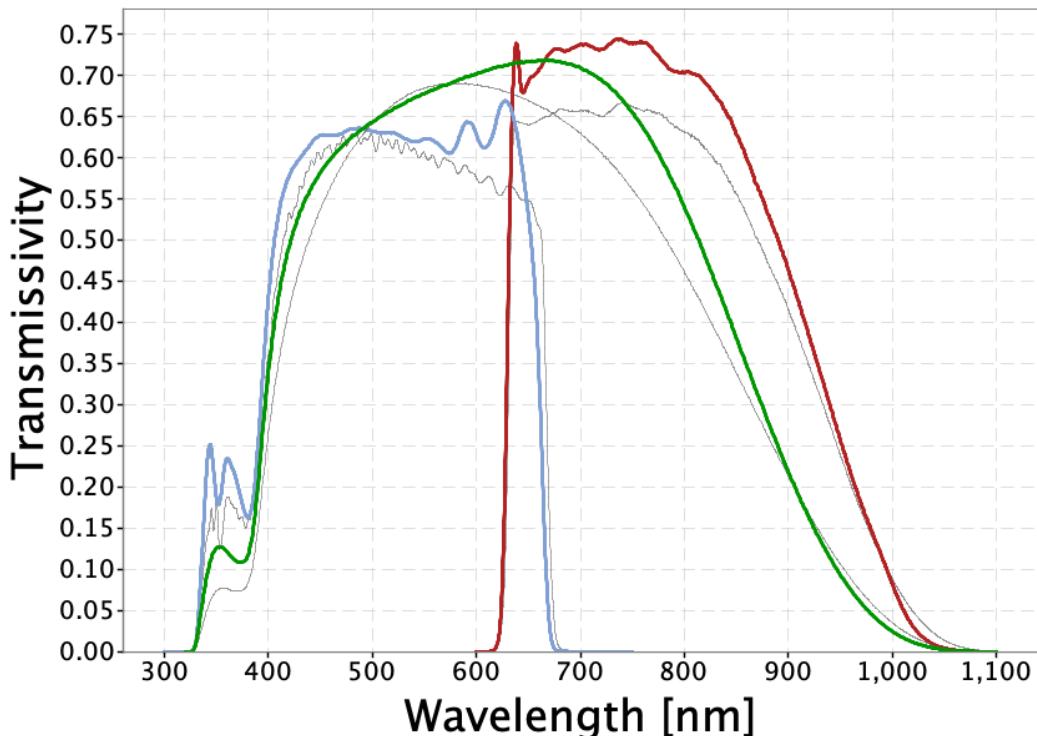


Figure 1.3: Range of Wavelengths admitted by each passband - red is for red pass filter, blue for a blue pass and the green line denotes the entire flux in visible range

## 1.4 Astrometry

### 1.4.1 Parallax

Parallax measures the angular movement of a star w.r.t. a fixed background of the sky. The angular movement is a result of the Earth (and Gaia) revolving around the sun. From different points in the orbit, the line of sight to the star is slightly different resulting in parallax. It has units of angle. The unit of distance parsec is defined as the distance at which a star appears to have a parallax of 1 arcsecond (1/3600 of a degree) as seen from Earth.

Hence, parallax  $p$  of a star is directly related to its distance  $d$  as  $d$  in parsecs =  $\frac{1}{p}$  where parallax is in arcseconds.

### 1.4.2 Proper Motion

A star also has its intrinsic relative velocity w.r.t. Earth-Sun system due to its motion around the galactic center. This results in the star changing its position in the sky by a very small amount each year, known as proper motion. Note that the proper motion only measures the transverse component of the star's velocity, or more precisely, its angular velocity w.r.t. Earth. The transverse component of velocity is not included. Proper motion in Gaia catalogs is measured in milliarcseconds/year.

## 1.5 HR Diagrams - the Physics of Stars

A star can be approximated quite well as a blackbody. For the same, we have the typical radiation curve and Planck's Law as shown below

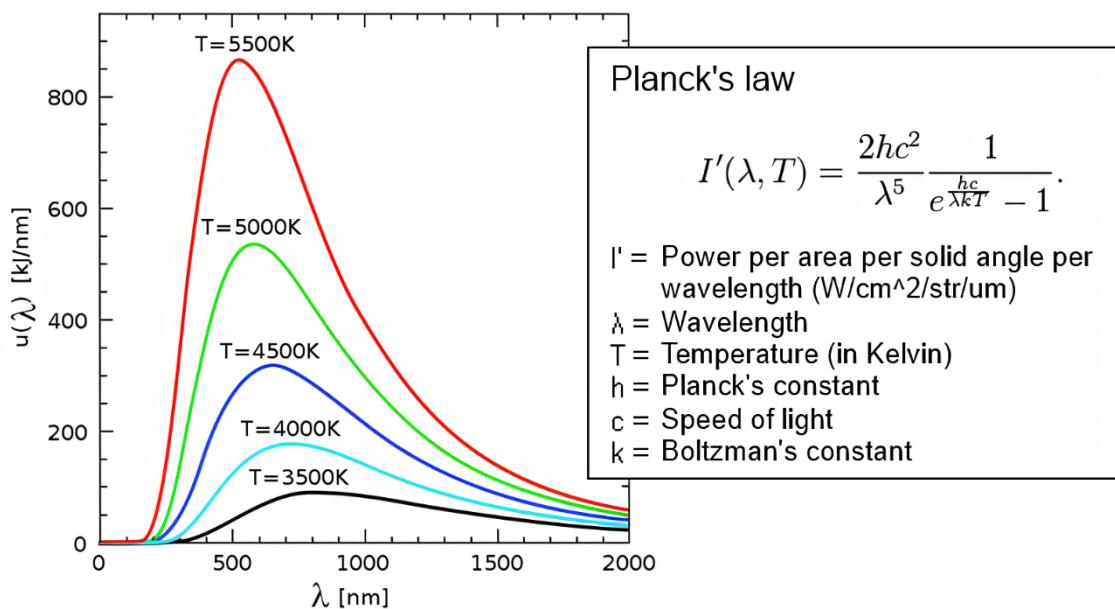


Figure 1.4: Blackbody Radiation Curve

It is easily seen that as temperature increases, the peak shifts to the left. Now consider this with respect to the blue pass and red pass filters. At a lower temperature, there is more flux from stars at higher wavelengths i.e. in the red band. So  $G_{RP}$  is lower in value than  $G_{BP}$  or  $G_{BP} - G_{RP}$  is more positive. The opposite happens for higher temperature stars, i.e.  $G_{BP} - G_{RP}$  is less positive or negative. So if we plot  $G_{BP} - G_{RP}$  on one axis, it gives a good idea of the temperature.

Meanwhile, consider the flux-luminosity relation given by

$$F = \frac{L}{4\pi r^2}$$

and the absolute magnitude

$$M = -2.5 \log_{10} \left( \frac{F}{F_0} \right) = -2.5 \log_{10}(L) + c$$

where  $c$  is a fixed constant. Thus absolute magnitude serves as a measure of luminosity.

Finally, it may be noted that luminosity is related to surface temperature as

$$L = 4\pi\sigma R^2 T^4$$

where  $R$  is the radius and  $T$  the surface temperature of the star.

### 1.5.1 HR Diagrams

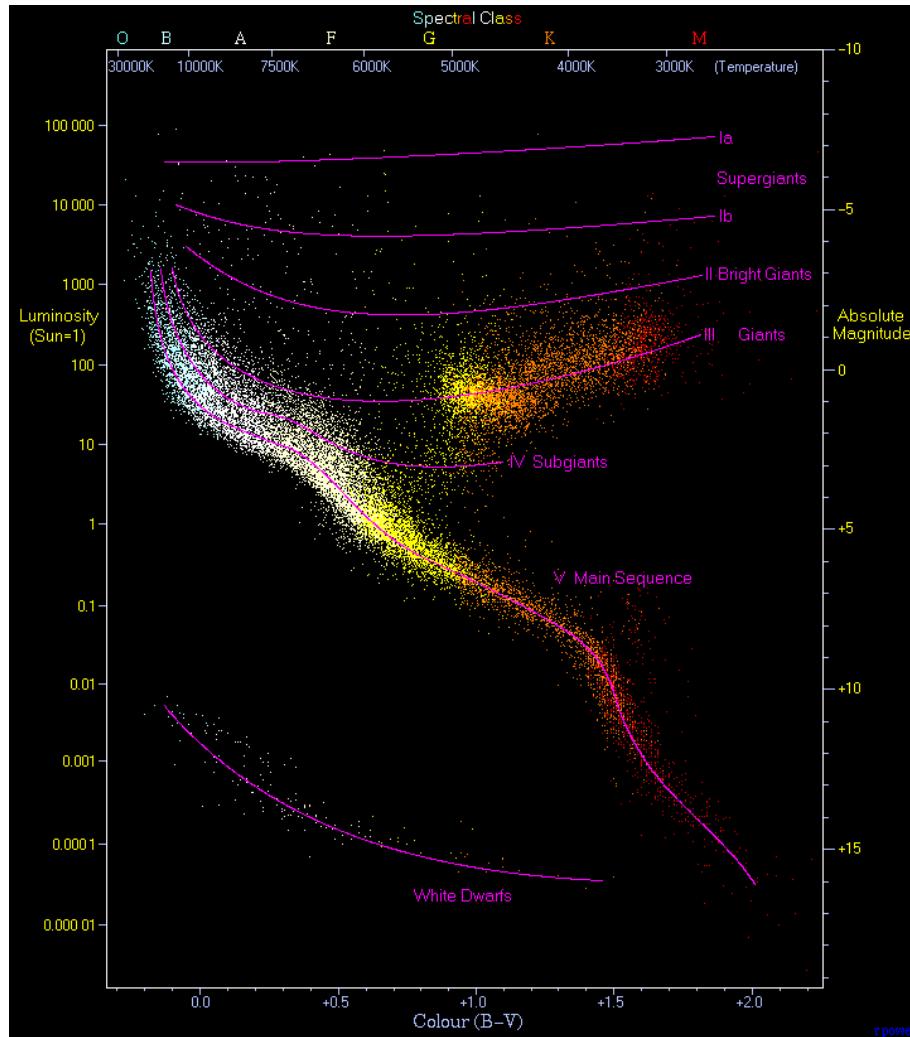


Figure 1.5: Standard HR Diagram

Normally, one would expect there to be no relation between the luminosity (absolute magnitude) and temperature ( $G_{BP} - G_{RP}$ ) due to the radius of a star being a uniformly varying value. However, the internal structure of the star (mainly its mass) dictates that only specific combinations of these are allowed, which was deduced by studying stellar data extensively. These are shown on the Hertzsprung-Russell diagrams (HR diagrams) which look something like above.

Most of the stars in a cluster occupy a region called the main sequence. These are stars that are using hydrogen as their fuel. The main sequence can be divided into classes as follows.

### Spectral Classification

Each star is divided into different spectral types depending on the spectrum produced by the star. Further, each spectral class is divided into 10 subclasses A0, A1, A2, ... A9. The classes and subclasses represent a sequence of temperatures, from hotter (O stars) to cooler (M stars) and from hotter (A0) to cooler (A9).

Spectral Type	Surface Temperature	Distinguishing Features
O	> 25,000K	H; HeI; HeII
B	10,000-25,000K	H; HeI; HeII absent
A	7,500-10,000K	H; CaII; HeI and HeII absent
F	6,000-7,500K	H; metals (CaII, Fe, etc)
G	5,000-6,000K	H; metals; some molecular species
K	3,500-5,000K	metals; some molecular species
M	< 3,500K	metals; molecular species (TiO!)
C	< 3,500K	metals; molecular species (C2!)

### Luminous Classification

Stars are also classified by luminosity class. Luminosity classes are determined from spectral features and photometric measurements, coupled with information regarding the distance to the star and the amount of extinction of the starlight from interstellar material. The luminosity class designation describes the size (gravitational acceleration in the photosphere) of a star from the atmospheric pressure. For larger stars of a given spectral type, the surface gravity decreases relative to what it was on the main sequence, and this decreases the equivalent widths of the absorption lines.

Luminosity Class	Description	Comments
0	Hypergiants	extreme
Ia	Supergiants!	large and luminous
Ib	Supergiants!	less luminous than Ia
II	Bright Giants	
III	Giants	
IV	Sub-Giants	
V	Dwarfs	Main Sequence
sd	Sub-Dwarfs	
D	White Dwarfs	

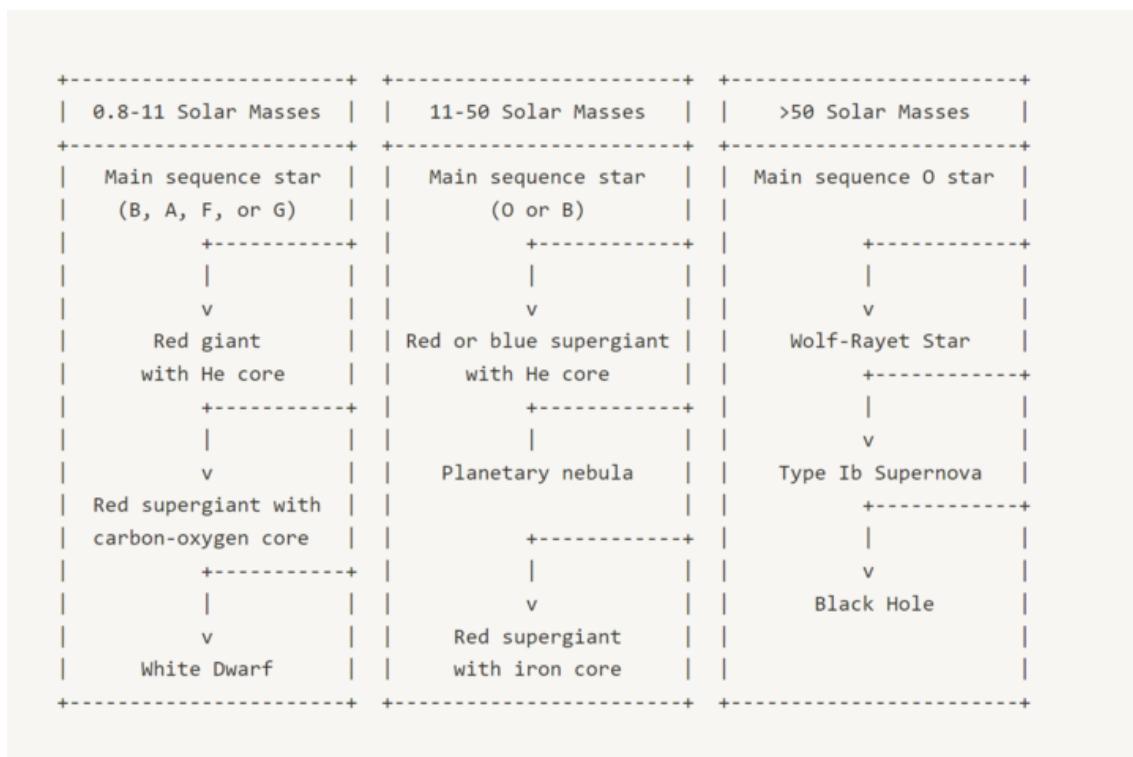
### 1.5.2 Life Cycle of Stars

Stars go through a cycle of birth, life, and death. The life cycle of a star depends on its mass. For example, a star like our Sun will live for about 10 billion years, while a star 20 times its mass will only live for a few million years. During a star's life, it fuses hydrogen into helium, releasing energy in the form of light and heat.

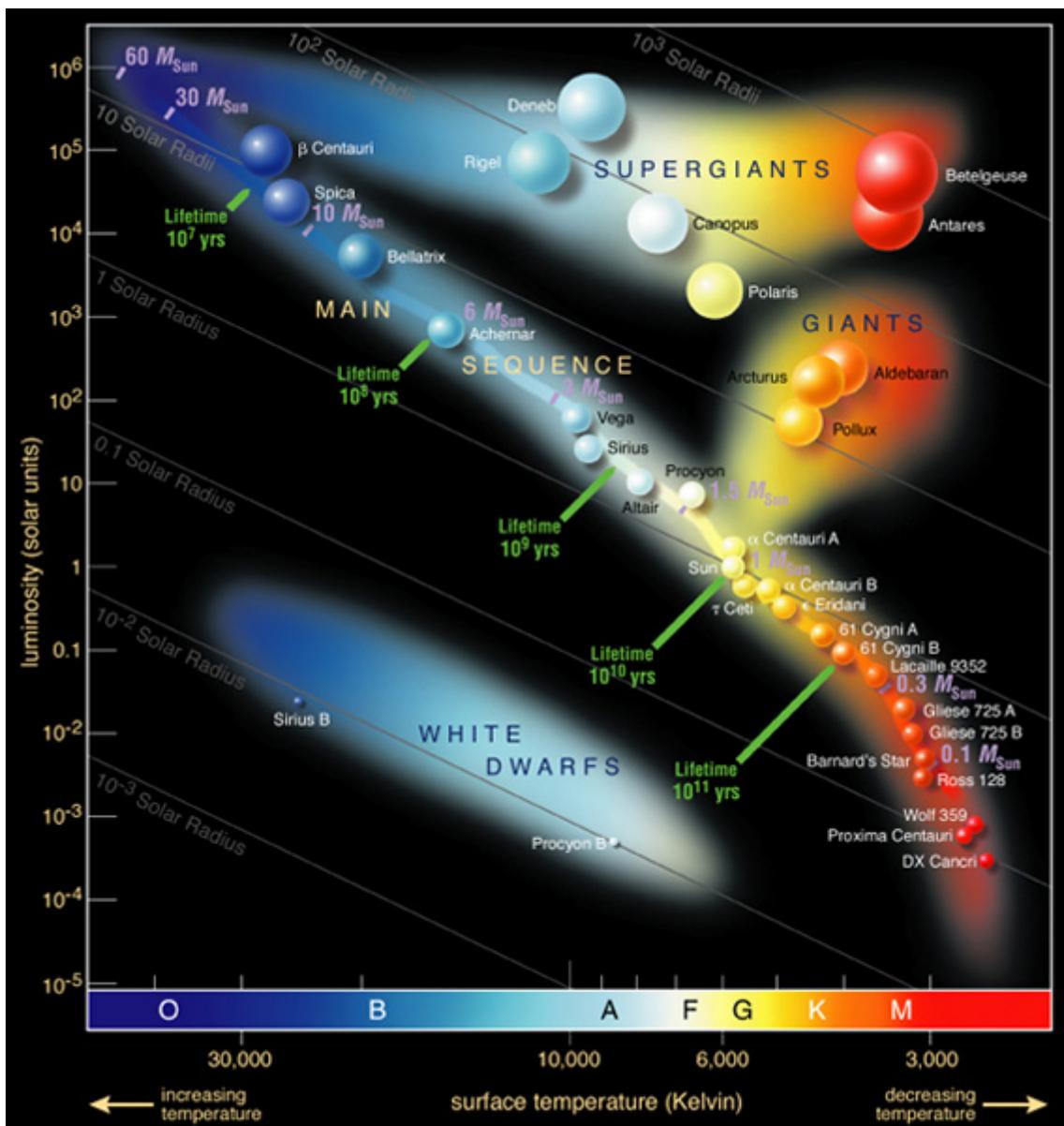
As a star ages and runs out of hydrogen fuel, it may start fusing helium into heavier elements like carbon and oxygen. This causes the star to expand and cool, becoming a red giant. Eventually, the star will shed its outer layers, leaving behind a hot, dense core known as a white dwarf.

If the star is massive enough, it may undergo a violent explosion known as a supernova. During a supernova, the star's core collapses, creating a burst of energy that can briefly outshine an entire galaxy. The explosion also creates heavy elements like gold, silver, and uranium.

Sometimes, the core of a massive star will collapse even further, forming a neutron star. Neutron stars are incredibly dense, with a single teaspoon of neutron star material weighing as much as a mountain. They also have incredibly strong magnetic fields and can spin incredibly fast.



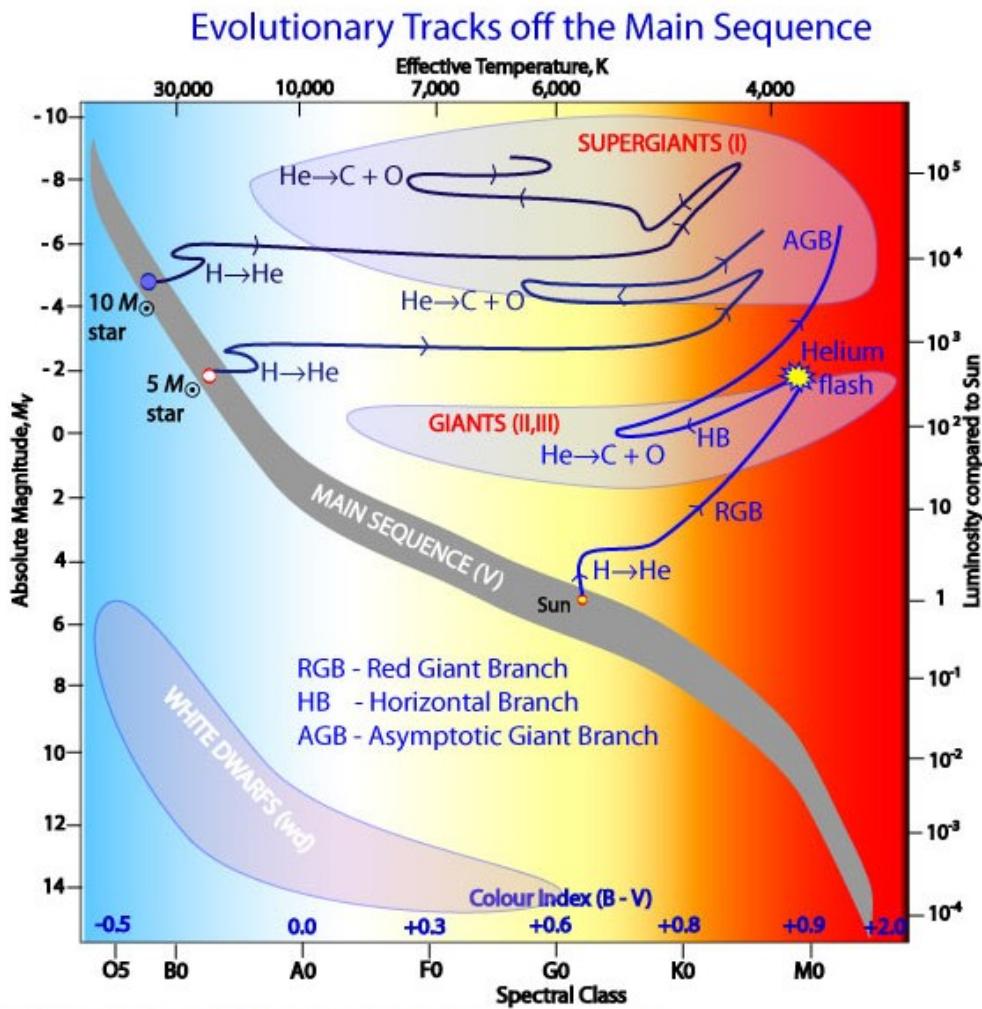
The position of a star on the HR diagram is also dependent on which stage of its life cycle it is in.



### Main Sequence Stars

Main sequence stars are stars that fuse hydrogen into helium in their cores. They are the most common type of star in the universe and are classified by their spectral type and luminosity class. Our Sun is a main sequence star of spectral type G2V and luminosity class V.

Low-mass stars stay on the main sequence, with stars with masses less than 0.5 solar masses unable to burn helium.



Post-main sequence evolutionary tracks for 1, 5 and 10 solar mass stars.

### Red Giant

A red giant is a dying star in the final stages of stellar evolution. Red giant stars are so named because they are much larger than main sequence stars like our Sun.

A medium-sized star, will eventually burn out of hydrogen and start burning helium. As it expands, it will become cooler and thus drift off the main sequence to the right into the Red Giant Branch (RGB).

As it continues to burn helium in the core, the outer layers burn hydrogen and dump the helium in the core. At one point there is so much helium being burnt in the core that it enters a stage called the helium flash. Due to this, its size decreases and temperature increases, moving the star right on the HR diagram into the Horizontal Branch (HB).

Eventually, it runs out of helium and expands to burn heavier elements, moving along the Asymptotic Giant Branch (AGB).

As a red giant star ages, it will cool down and become less luminous, eventually shedding its outer layers and leaving behind a core known as a white dwarf. Some red giants, however, may undergo a supernova explosion instead, creating heavy elements and releasing enormous amounts of energy.

### White Dwarfs

White dwarfs are what is left when stars like our sun have exhausted all of their fuel. They are dense, dim, stellar corpses — the last observable stage of evolution for low- and medium-mass stars. Whilst most massive stars will eventually go supernova, a low or medium-mass star with a mass less than about 8 times the mass of the sun will eventually become a white dwarf. The maximum mass a white dwarf can have is 1.4 solar masses, which is called the Chandrasekhar limit.

These stars are present at the bottom left, as labeled in Figure 1.5.

### Neutron Stars

Neutron stars are incredibly dense, with a single teaspoon of neutron star material weighing as much as a mountain. They are formed when the core of a massive star collapses even further than it would to become a white dwarf, creating a super-dense ball of neutrons. Neutron stars also have incredibly strong magnetic fields and can spin incredibly fast.

### Black Holes

Black holes are formed when the core of a massive star collapses further than it would to become a neutron star. The gravitational pull at the surface of a black hole is so strong that nothing, not even light, can escape it. Anything that gets too close to a black hole will be sucked in and destroyed. Black holes are still not fully understood and remain one of the most mysterious objects in the universe.

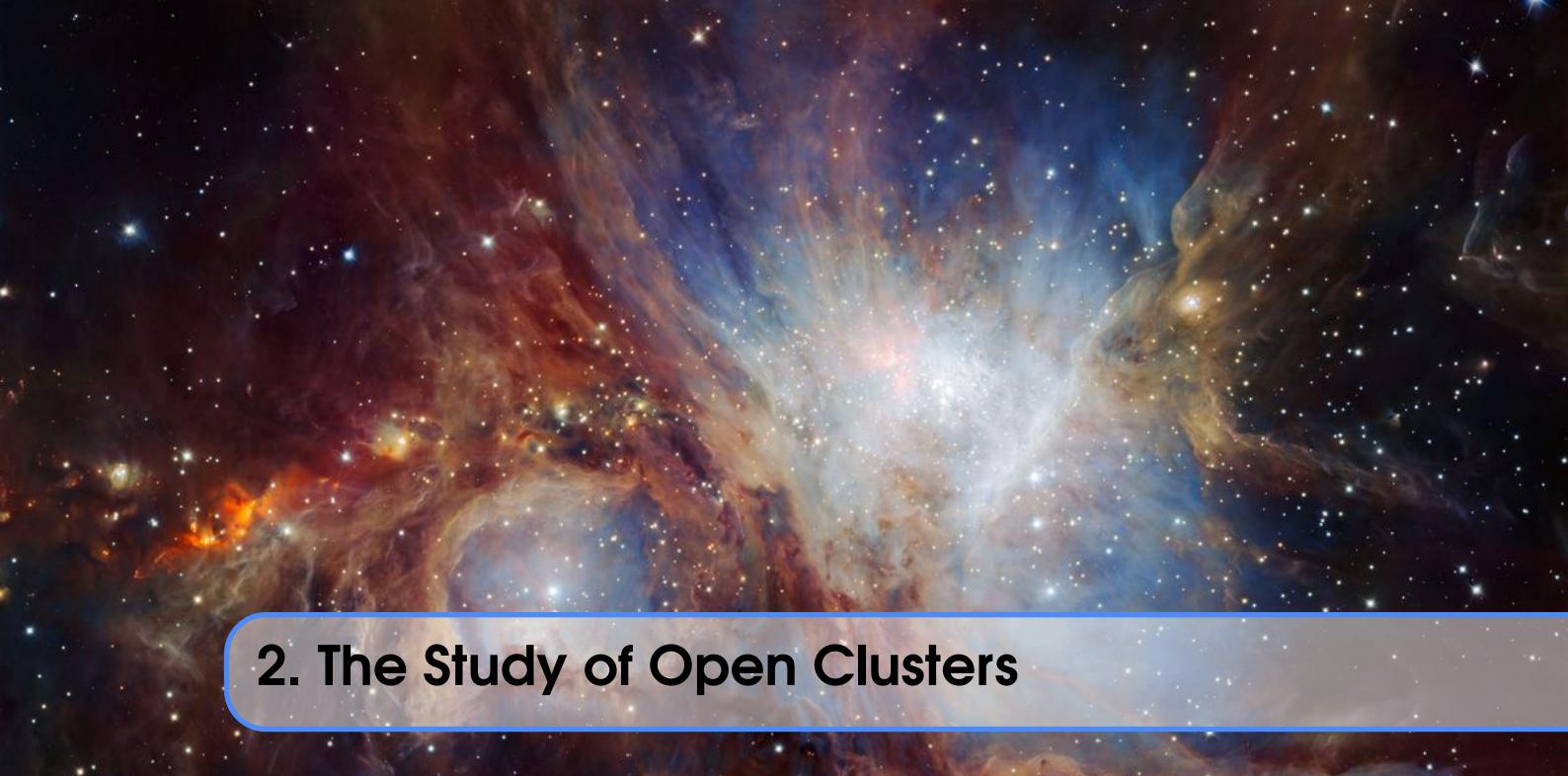
### Metallicity of a Star

The metallicity of a star refers to the abundance of elements heavier than helium in its atmosphere. Stars with high metallicity are thought to have formed from the remnants of previous generations of stars, while stars with low metallicity are thought to be some of the oldest stars in the universe. By studying the metallicity of stars, astronomers can learn more about the early universe and the processes that led to the formation of galaxies and stars.

The formula for metallicity is:

$$(Fe/H) = \log_{10} \left( \frac{Fe}{H} \right) - \log_{10} \left( \frac{Fe}{H}_{\odot} \right)$$

where Fe is the abundance of iron and H is the abundance of hydrogen. The solar value for Fe/H is used as a reference point. (Fe/H) is often used as a measure of metallicity because iron is one of the most abundant heavy elements in the universe. By studying the metallicity of stars, astronomers can learn more about the early universe and the processes that led to the formation of galaxies and stars.



## 2. The Study of Open Clusters

### 2.1 Introduction

There are two main types of star clusters: Open Clusters and Globular clusters. An open cluster refers to a specific type of star cluster composed of tens to a few thousand stars while globular cluster is a group ten thousands to millions of old stars that are bounded very strongly by gravitational forces. Open clusters originate from a common giant molecular cloud and have almost the same age. The Milky Way galaxy alone has yielded the discovery of over 1,100 open clusters, with scientists speculating the existence of many more. These clusters are held together by a loose gravitational attraction among their members but are prone to disruption when they encounter other clusters or gas clouds during their orbits around the Galactic Center.

Open clusters originate from the collapse of giant molecular clouds, which are dense accumulations of gas and dust. These clouds, with masses several thousand times that of the Sun, maintain equilibrium through magnetic fields, turbulence, and rotation. Various factors can disrupt this equilibrium, such as shock waves from supernovae, cloud collisions, or gravitational interactions. The collapse of a molecular cloud triggers star formation within regions of higher density (quantitatively in regions where density is  $10^4$  hydrogen molecules per  $cm^3$ ). The cloud fragments hierarchically into smaller clumps, eventually giving rise to several thousand stars. During this process, protostars remain obscured within the collapsing cloud, visible only in infrared light. The most massive and hottest stars among the newly formed cluster emit intense ultraviolet radiation, ionizing the surrounding gas and creating an H II region. Stellar winds and radiation pressure from these stars disperse the ionized gas at the speed of sound. Over millions of years, the cluster may experience core-collapse supernovae, expelling additional gas from the vicinity. This gas expulsion, combined with other processes, limits the cluster's lifespan to approximately ten million years, halting further star formation.

Gas expulsion has significant effects on open clusters. Only a fraction of the original gas content forms stars, resulting in infant weight loss and significant mortality rates. For example, when Pleiades was formed only one-third of stars remained and the rest became unbounded. The formation of an open cluster depends on the gravitational binding of its stars; otherwise, an unbound stellar association may form. The remaining stars contribute to the Galactic field population.

The following open clusters were selected to be analysed:

1. Blanco 1
2. M44 Beehive Cluster
3. NGC 2516 Southern Beehive Cluster
4. NGC 3532 Wishing Well Cluster
5. M35
6. NGC 751
7. IC 4651
8. Messier 67
9. NGC 188
10. M6 Butterfly Cluster
11. M34 Spiral Cluster
12. M42 Orion
13. M7
14. Messier 45
15. Messier 50
16. NGC 188
17. Messier 38
18. Lambda Orionis
19. Alpha Persei Cluster

## 2.2 Selecting Cluster Members

We considered the top 40,000 brightest stars in the Gaia EDR3 catalog. A common criterion is to choose those that exhibit a good Gaussian fit for distance and possess a prominent peak above the background. This approach ensures that the selected clusters have a well-defined central tendency and distinct characteristics that distinguish them from the surrounding data points. We used the list of open clusters on Wikipedia that have a distance range of less than 2000 pc, as Gaia data on parallax worsens as we increase distance. To make the data better, there was filtering done on the basis of clustering in the pmra/pmdec space as well. Error cuts on pmra and pmdec were not used as they severely reduced the number of stars we got from a query. The selection criteria were:

- Parallax is not NULL
- Error in Parallax < 20
- Order by G band magnitude in descending order

### 2.2.1 ADQL Query

An example of an ADQL query is as follows

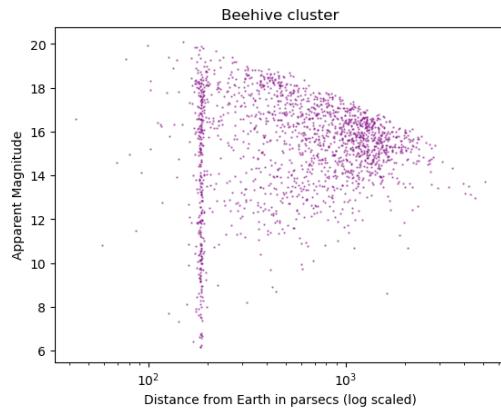
```
query = '''SELECT top 40000 source_id, ra, dec, parallax, phot_g_mean_mag,
phot_bp_mean_mag, phot_rp_mean_mag, pm, pmra, pmdec
from gaiadr3.gaia_source
```

```
where ra between 83.32 and 84.32
AND dec between -5.89 and -4.89
AND parallax_over_error > 10
AND parallax is not NULL
order by phot_g_mean_mag DESC'''
```

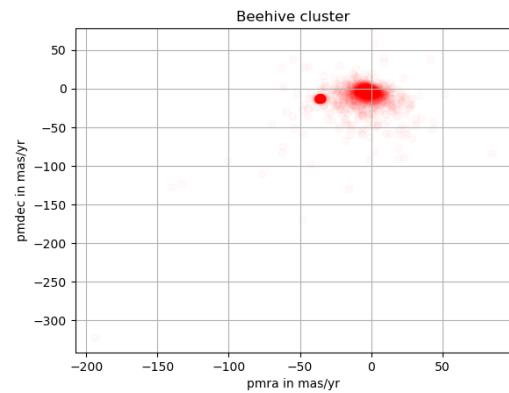
Following is a procedure followed used for all clusters to be analyzed. A few examples have been taken to illustrate the procedure.

### 2.2.2 Beehive cluster

This is the most ideal scenario; for all star clusters, one may not get such a good result. On querying, we checked for any apparent clustering based on a scatter plot between apparent magnitude and logarithmic ally scaled distance from Earth, and then we check for clumping in the pmra/pmdec space.



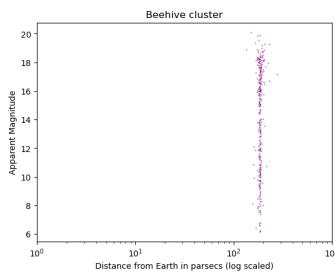
(a) Plot of distance and apparent mag



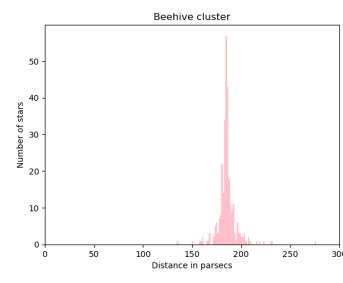
(b) Plot of pmra and pmdec

Figure 2.1: From plot we can infer that the cluster is at some distance larger than 100 parsecs, while the other plot has two clumps; one slightly dense one of background stars at (0,0) and the other very dense one at (-40, -10) which is the cluster

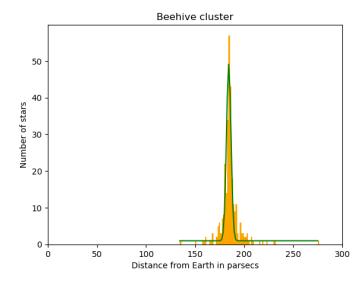
We query again, this time with pmra and pmdec filters. After that, we find the histogram of the distance of stars from Earth and fit a Gaussian distribution to find the mean and standard deviation. We take the clusters position to be at the mean, and member stars to be within 3 times the standard deviation around the mean.



(a) Plot of distance and apparent magnitude, after filtering pmra/pmdc.



(b) Histogram of distance after pmra/pmdc filtering.



(c) Histogram of distance, with a Gaussian curve fitting.

In this manner, we are able to get the cluster members.

### 2.2.3 M42 Orion

We shall again follow the same procedure as for Beehive.

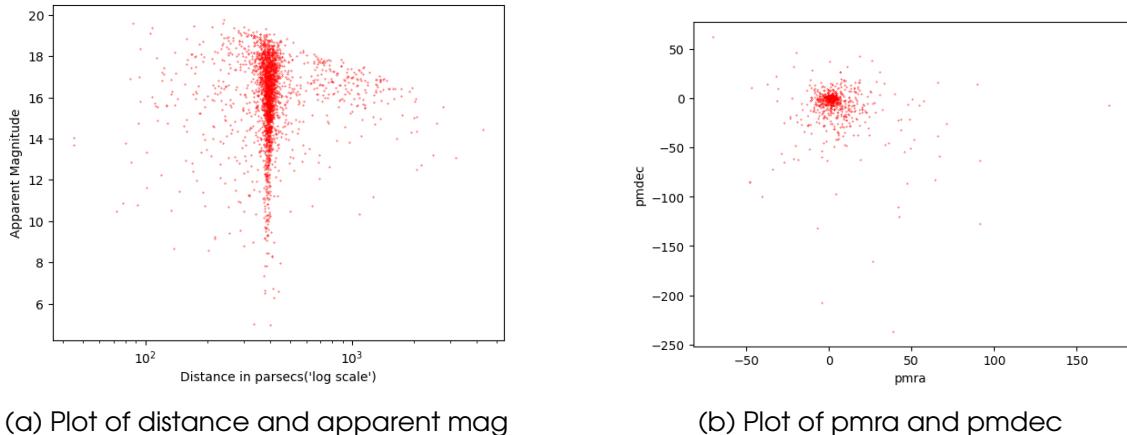


Figure 2.3: Two plots showing the presence of a cluster.

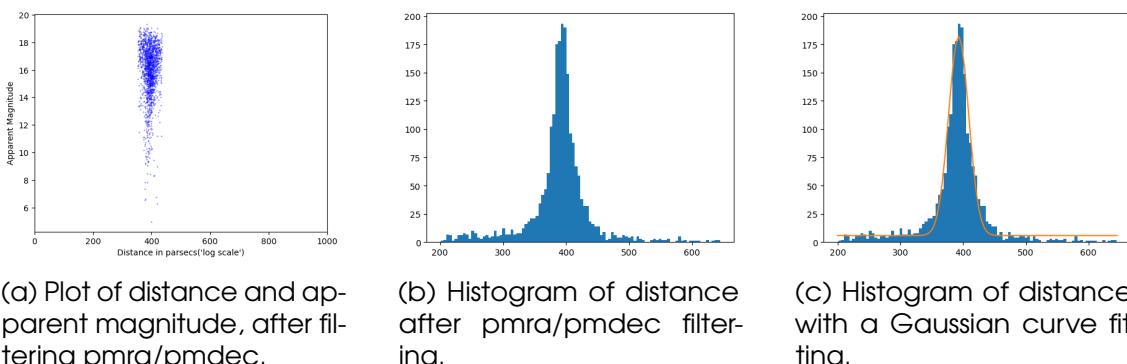
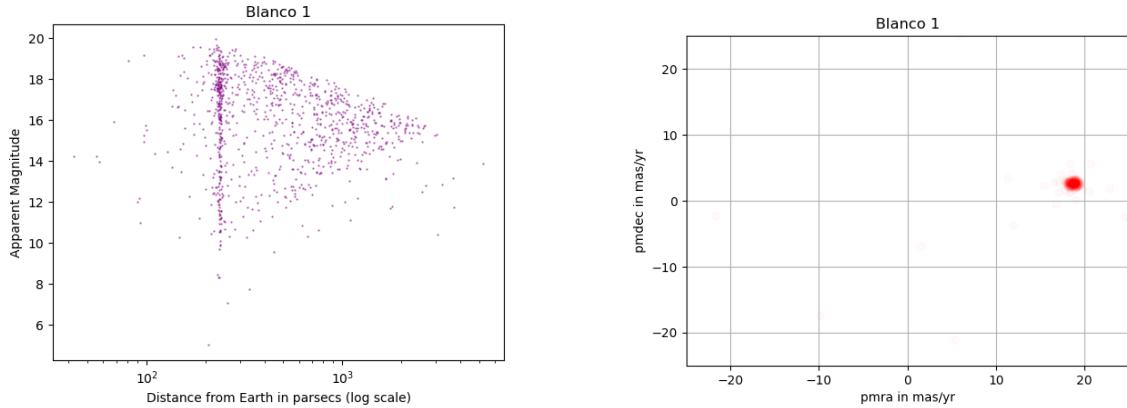


Figure 2.4: The cluster is between 200 and 300 parsecs according to plot a. The data received after pmra/pmdc filtering is very clean, and there are very few stars in the filtered data which are not part of the cluster.

### 2.2.4 Blanco 1 cluster

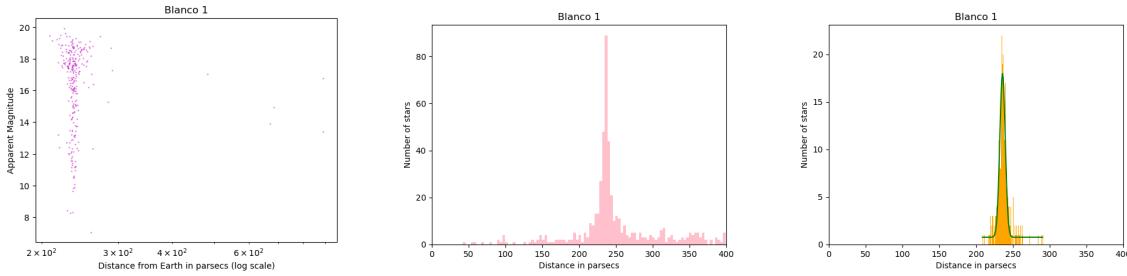
We shall follow the same procedure as for Beehive.



(a) Plot of distance and apparent mag

(b) Plot of pmra and pmdec

Figure 2.5: Two plots showing the presence of a cluster. In plot b we do not get a background star clump as they are lesser and more spread, because of the high transparency they are not visible.



(a) Plot of distance and apparent magnitude, after filtering pmra/pmdc.

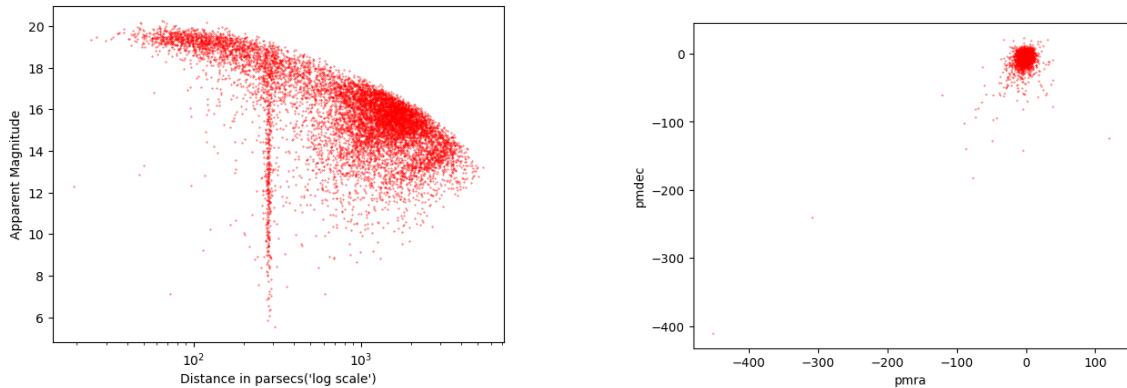
(b) Histogram of distance after pmra/pmdc filtering.

(c) Histogram of distance, with a Gaussian curve fitting.

Figure 2.6: The cluster is between 200 and 300 parsecs according to plot a. The data received after pmra/pmdc filtering is very clean, and there are very few stars in the filtered data which are not part of the cluster.

## 2.2.5 M7

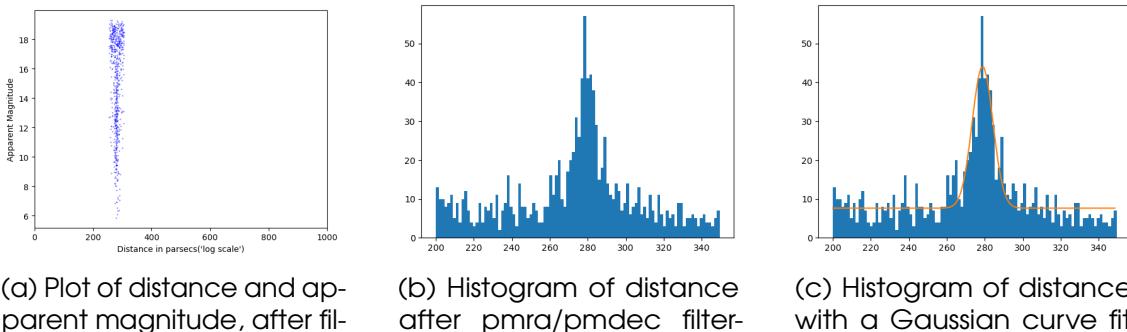
We shall again follow the same procedure as for Beehive.



(a) Plot of distance and apparent mag

(b) Plot of pmra and pmdec

Figure 2.7: Two plots showing the presence of a cluster. In plot b we do not get a background star clump as they are lesser and more spread, because of the high transparency they are not visible.



(a) Plot of distance and apparent magnitude, after filtering pmra/pmdc.

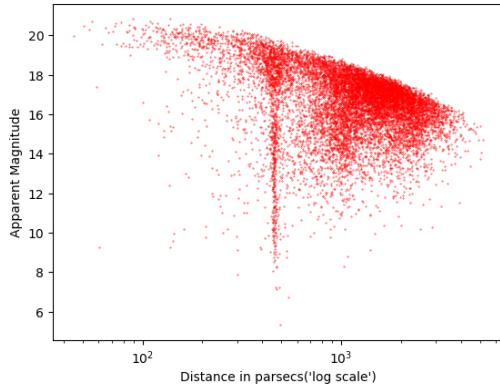
(b) Histogram of distance after pmra/pmdc filtering.

(c) Histogram of distance, with a Gaussian curve fitting.

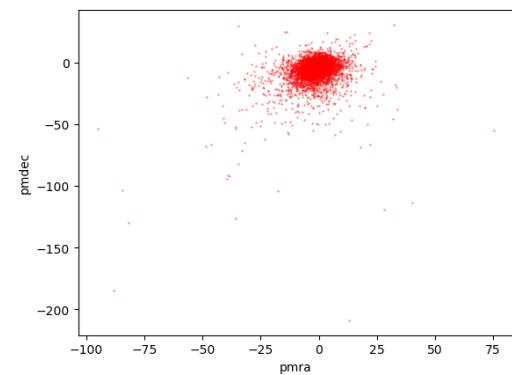
Figure 2.8: The cluster is between 200 and 300 parsecs according to plot a. The data received after pmra/pmdc filtering is very clean, and there are very few stars in the filtered data which are not part of the cluster.

## 2.2.6 M6

We shall again follow the same procedure as for Beehive.

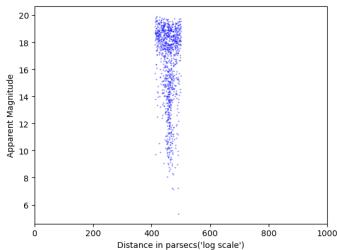


(a) Plot of distance and apparent mag

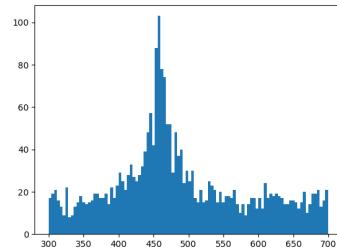


(b) Plot of pmra and pmdec

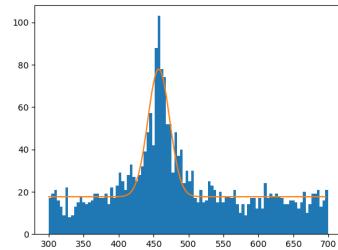
Figure 2.9: Two plots showing the presence of a cluster. In plot b we do not get a background star clump as they are lesser and more spread, because of the high transparency they are not visible.



(a) Plot of distance and apparent magnitude, after filtering pmra/pmdc.



(b) Histogram of distance after pmra/pmdc filtering.

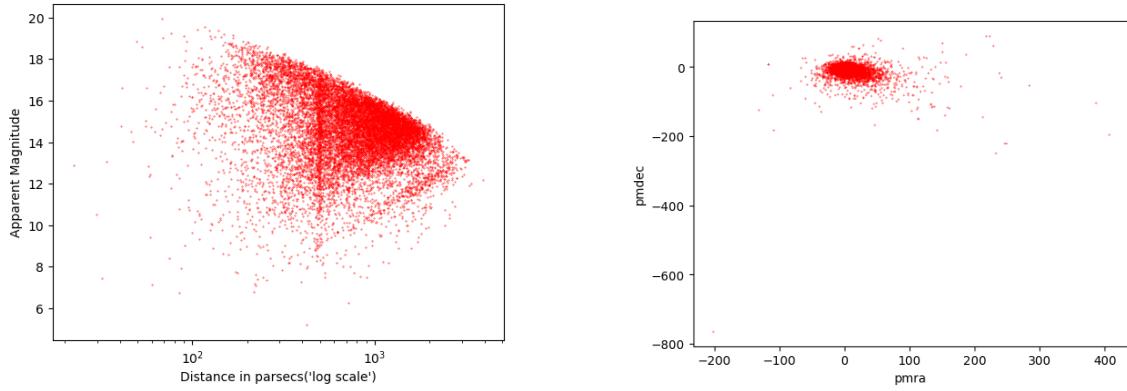


(c) Histogram of distance, with a Gaussian curve fitting.

Figure 2.10: The cluster is between 200 and 300 parsecs according to plot a. The data received after pmra/pmdc filtering is very clean, and there are very few stars in the filtered data which are not part of the cluster.

## 2.2.7 M34 Spiral Cluster

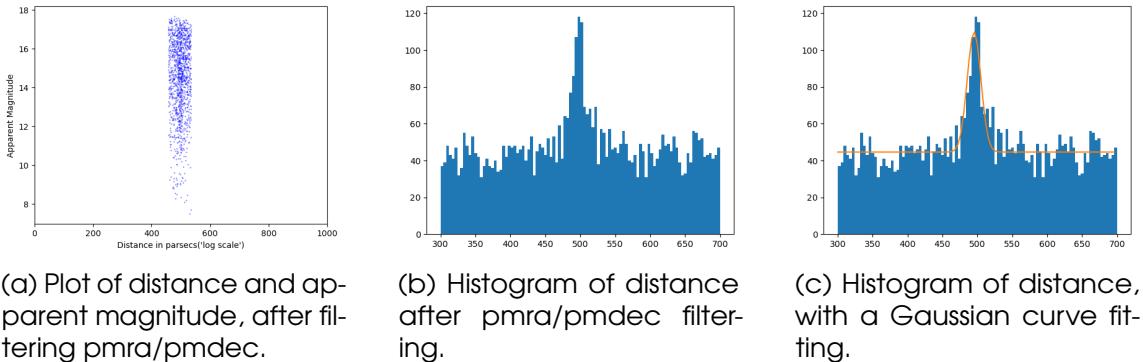
We shall again follow the same procedure as for M6.



(a) Plot of distance and apparent mag

(b) Plot of pmra and pmdec

Figure 2.11: Two plots showing the presence of a cluster.



(a) Plot of distance and apparent magnitude, after filtering pmra/pmdc.

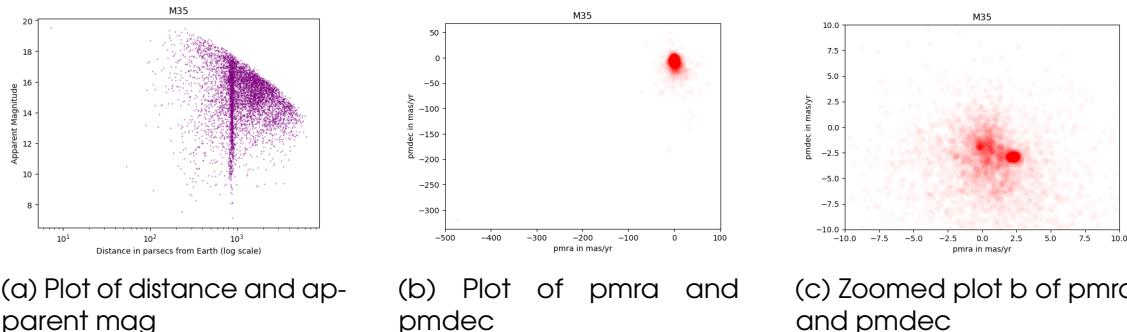
(b) Histogram of distance after pmra/pmdc filtering.

(c) Histogram of distance, with a Gaussian curve fitting.

Figure 2.12: The cluster is between 450 and 550 parsecs according to plot a. The data received after pmra/pmdc filtering is very clean, and there are very few stars in the filtered data which are not part of the cluster.

### 2.2.8 Messier 35

The above cases were ideal. Let us take the case of Messier 35, which is less ideal. There can be worse cases than this one, especially at larger distances which will be apparent when we get to the HR diagrams of NGC 188.



(a) Plot of distance and apparent mag

(b) Plot of pmra and pmdec

(c) Zoomed plot b of pmra and pmdec

Figure 2.13: Two plots showing the presence of a cluster. The plot b does not seem to have a clump which can be the cluster on a trivial observation. On a closer look upon zooming, the cluster clump is immediately visible.

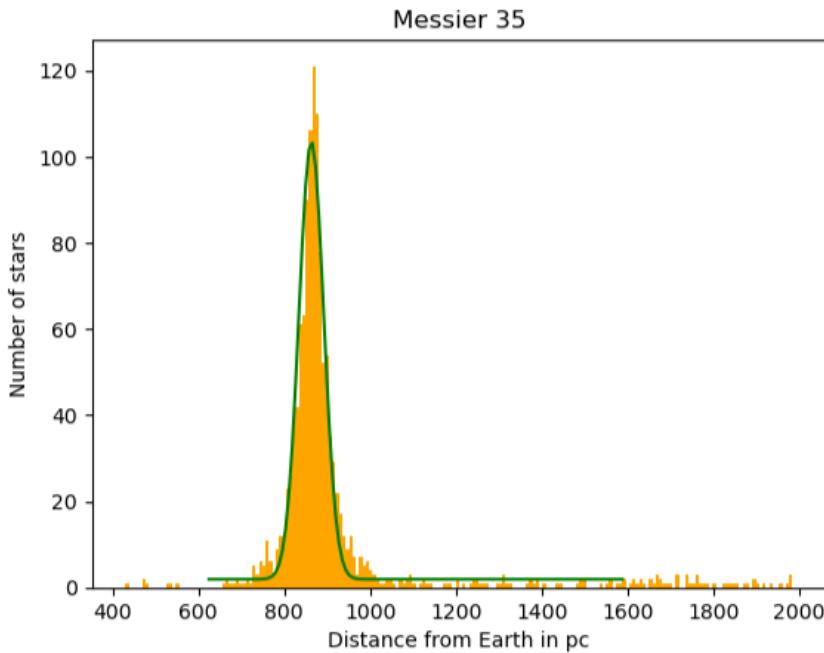
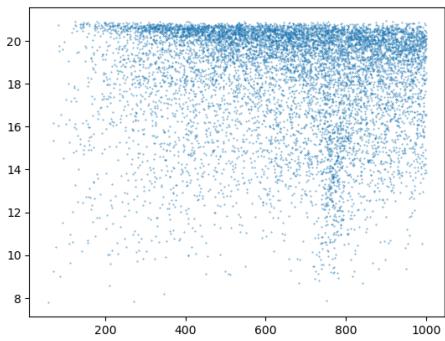


Figure 2.14: Histogram of distance, with a Gaussian curve fitting after pmra/pmdec filtering.

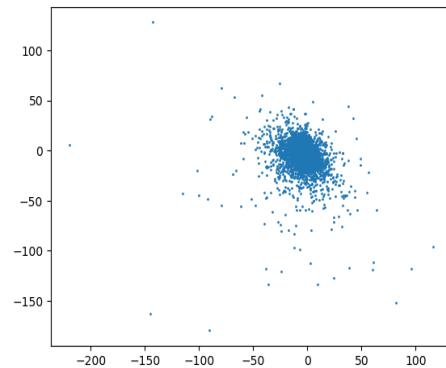
An open cluster cannot have such a big spread; this spread is this big because of more background stars and an increasing error in the parallax of stars in Gaia data as the distance from Earth increases. Decreasing the tolerance of error in parallax leads to too many stars being removed. We still apply the same cuts as before as we cannot differentiate between the background and cluster stars with these basic filters.

### 2.2.9 M48

As mentioned earlier, I have taken the plot of apparent magnitude vs distance and plot of pmra vs pmdec.

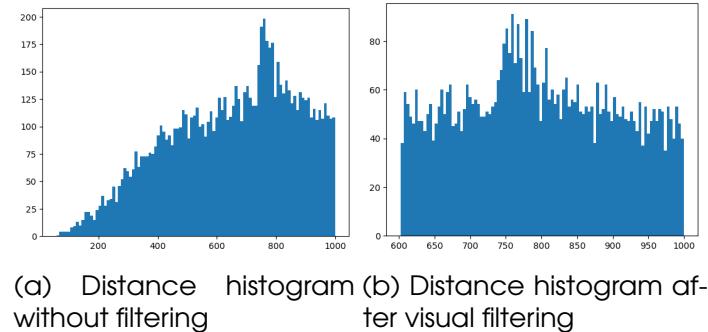


(a) Plot of Apparent magnitude vs Distance



(b) Plot of pmra and pmdec

Now, we will plot distance histogram. A peak comes in the graph as expected at a distance of about 800 parsec. After filtering the data we can see that peak more properly.



Now, we will do Gaussian curve fitting. Below is the graph for that.

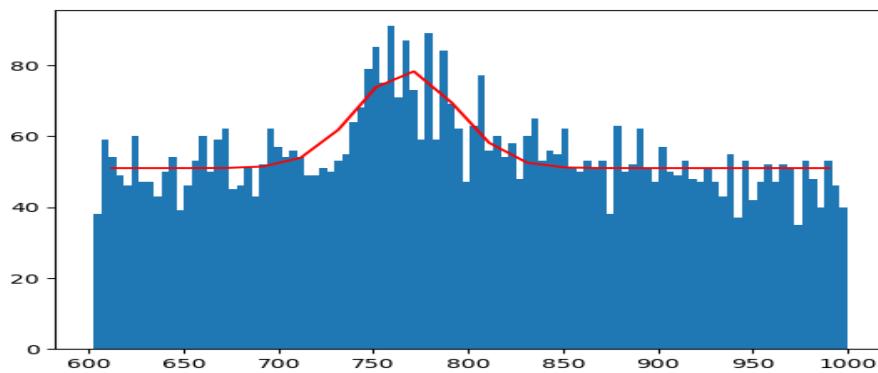
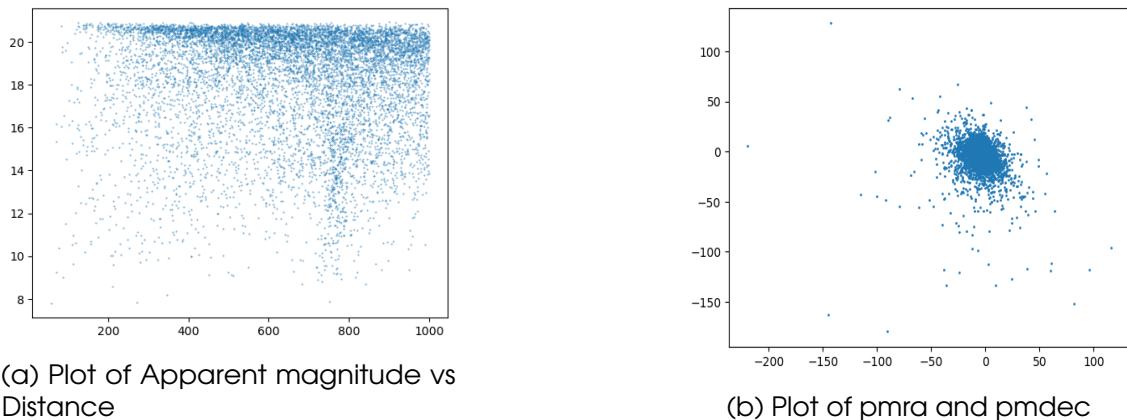


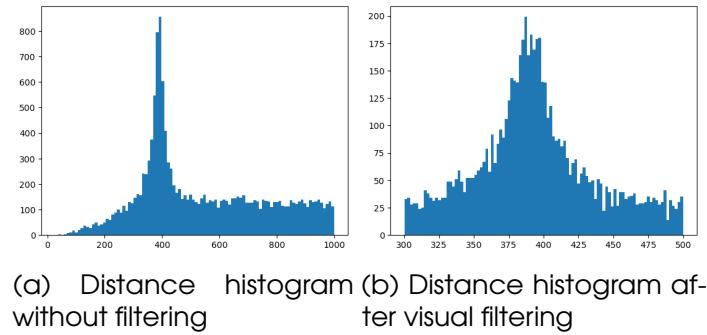
Figure 2.17: Gaussian curve fitting on that distance histogram

### 2.2.10 Trapezium Cluster

I will follow the same procedure as for M48. First we plotted apparent magnitude vs distance where we can see clear evidence for the cluster. Then pmra vs pmdec plot where we can see clear clustering below.



Below is the distance histogram. A peak comes at about 400 parsecs.



Below is the graph after Gaussian curve fitting.

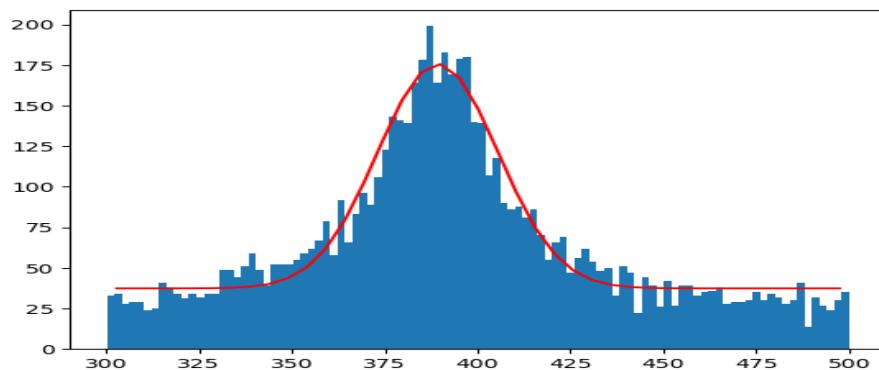


Figure 2.20: Gaussian curve fitting on that distance histogram

### 2.2.11 Messier 45

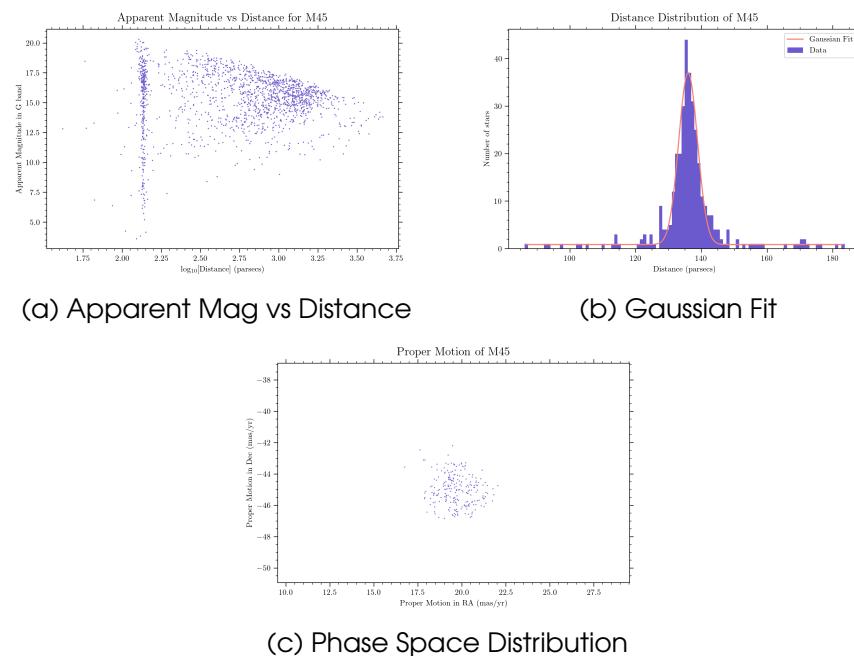


Figure 2.21: M45: Analysis of Distance and Location

### 2.2.12 Messier 50

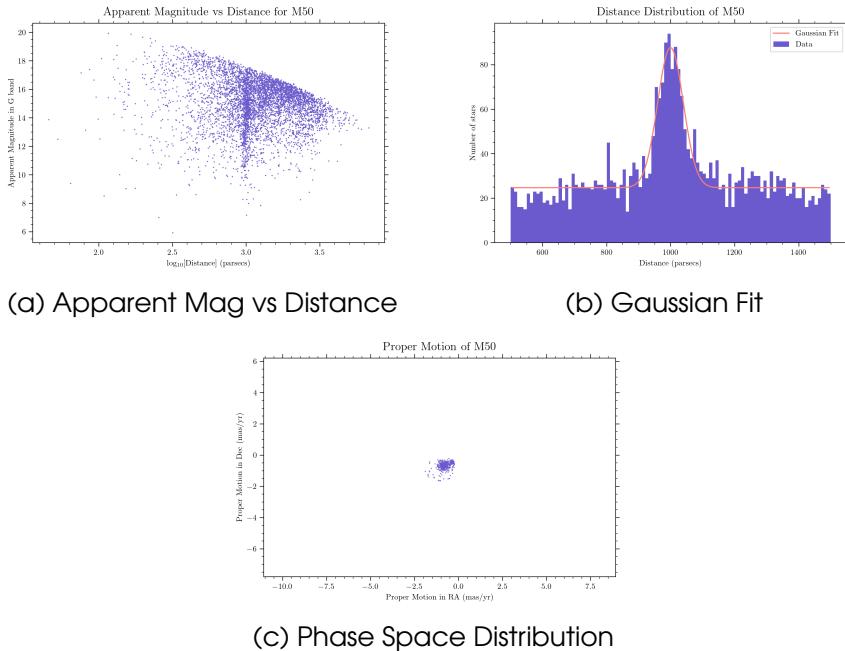


Figure 2.22: M50: Analysis of Distance and Location

### 2.2.13 Messier 38

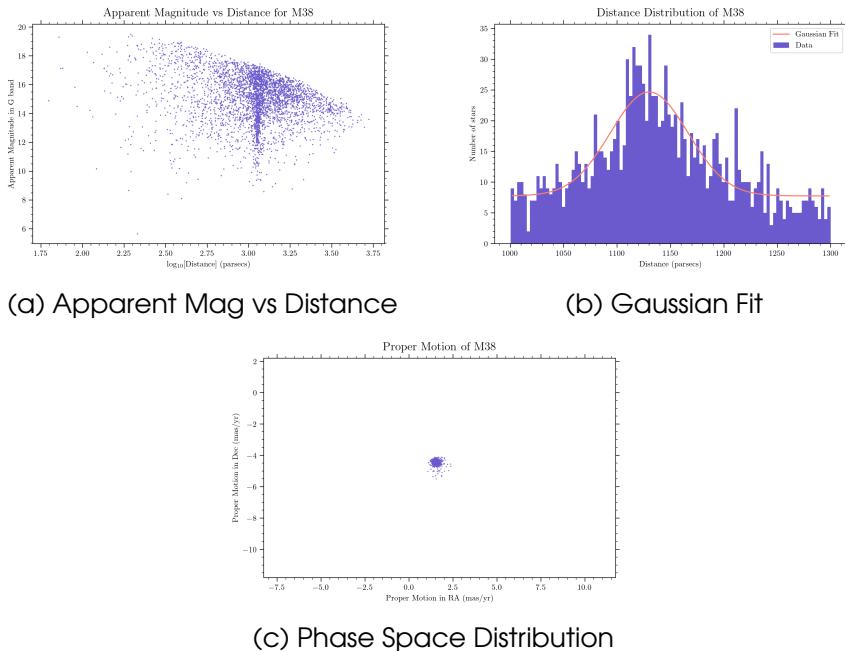


Figure 2.23: M38: Analysis of Distance and Location

### 2.2.14 Lambda Orionis

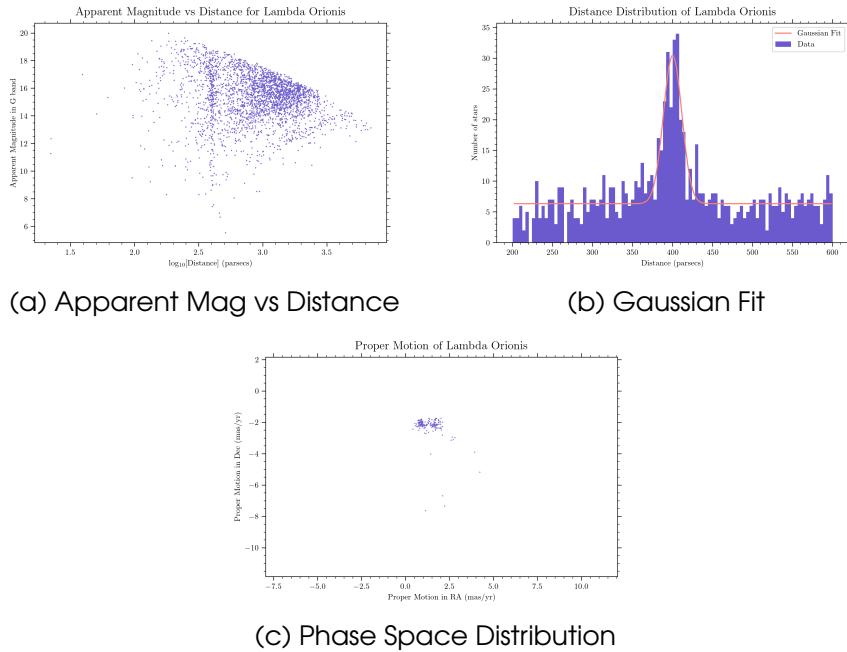


Figure 2.24: Lambda Orionis: Analysis of Distance and Location

### 2.2.15 Alpha Persei Cluster

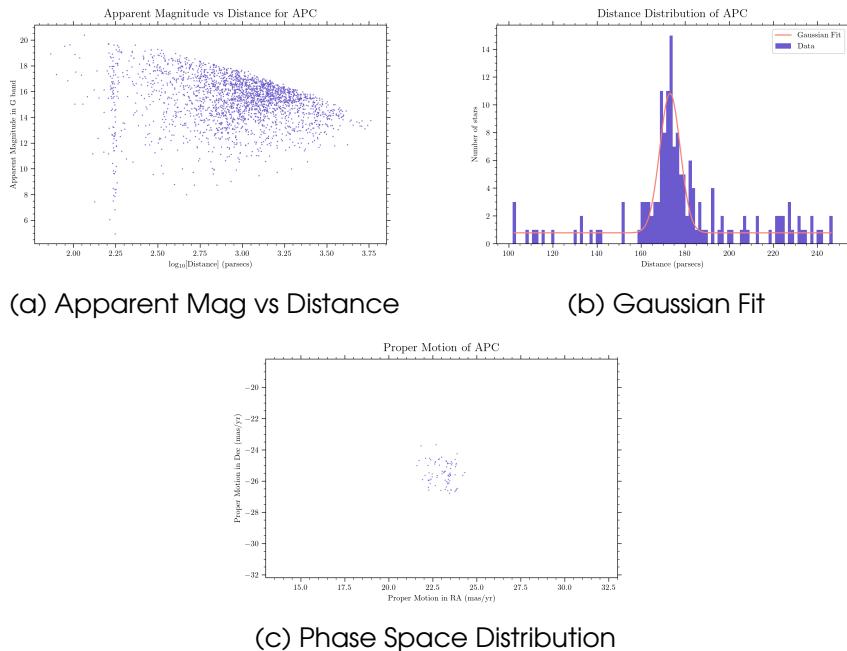
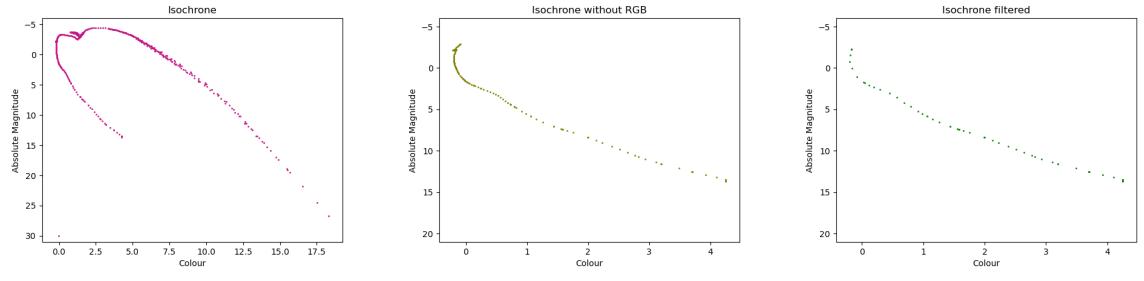


Figure 2.25: Alpha Persei Cluster: Analysis of Distance and Location

## 2.3 Age, Metallicity and Isochrones

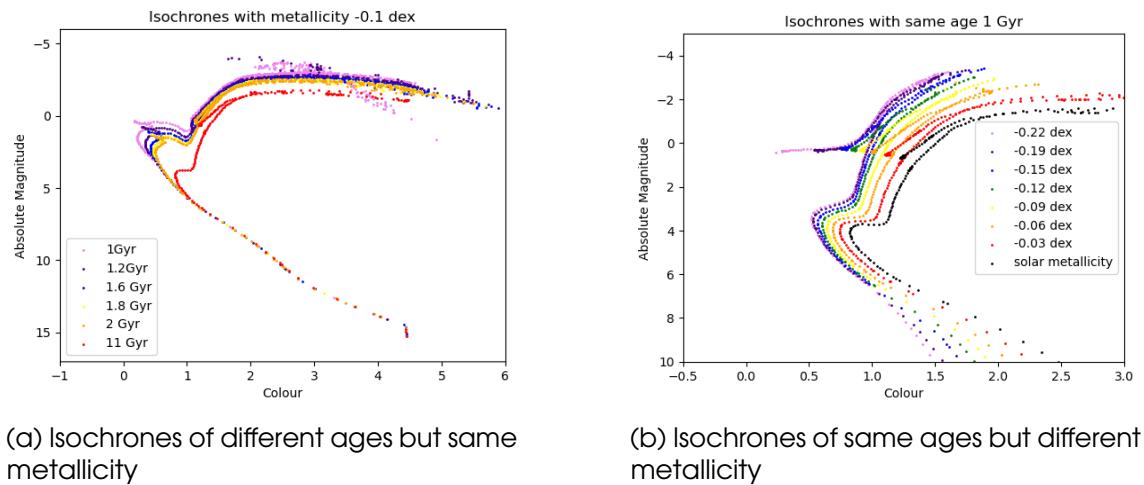
Isochrones are HR diagrams consisting of stars with same age but different mass. An example of the isochrone we will be using is given below.



(a) Plot of a theoretical isochrone.  
(b) Isochrone after removing AGB.  
(c) Isochrone with 51 nearly equally spaced points.

Figure 2.26: Plot (a) has a theoretical isochrone which is 600 million years old and of solar metallicity. It has an extravagant red giant branch, which is generally not there in open clusters, especially at this age. So we filter them out. Giant stars have lost a lot of mass but have higher radii, owing to a lower value of surface gravity (denoted as  $g$ ). We remove all stars with  $\log g < 3$  and get plot (b). After that is the isochrone with 51 roughly equally spaced points, on which we will plot our HR diagrams to give a comparison

Isochrones can tell a lot about a cluster. Since most stars in a cluster are born at a similar time, an isochrone that fits well on them will have the same properties as that of the cluster. This allows us to infer the age and metallicity of a cluster.



(a) Isochrones of different ages but same metallicity  
(b) Isochrones of same ages but different metallicity

Figure 2.27: The red giant branches in these isochrones has been cut to bring the more important distinctions in focus

In plot (a), all isochrones have the same main sequence, and similar, if not identical sharpness of turning points. The way they differ is primarily in the position of the turn off point on the main sequence. Older isochrones, have their turn off point earlier than younger ones; which is expected, as older isochrones will have older stars, and thus lesser stars on the main sequence.

In plot (b), we see that the main sequences are not the same, but slightly shifted. One empirical observation which can also be made is that as metallicity increases, the bend at the turn off point becomes sharper.

## 2.4 Spectral Classification using HR Diagrams

After finding the cluster members, we shall create and examine the HR diagrams of each of them.

Theoretically, the percentage of each spectral class in the main sequence of clusters in the solar neighbourhood is as follows:

1. O class: 0.00003%
2. B class: 0.125%
3. A class: 0.625%
4. F class: 3.03%
5. G class: 7.5%
6. K class: 12%
7. M class: 76%

The HR diagrams we find will have the red giant branch and white dwarfs. To classify, we filter these branches out.

### 2.4.1 Blanco 1 cluster

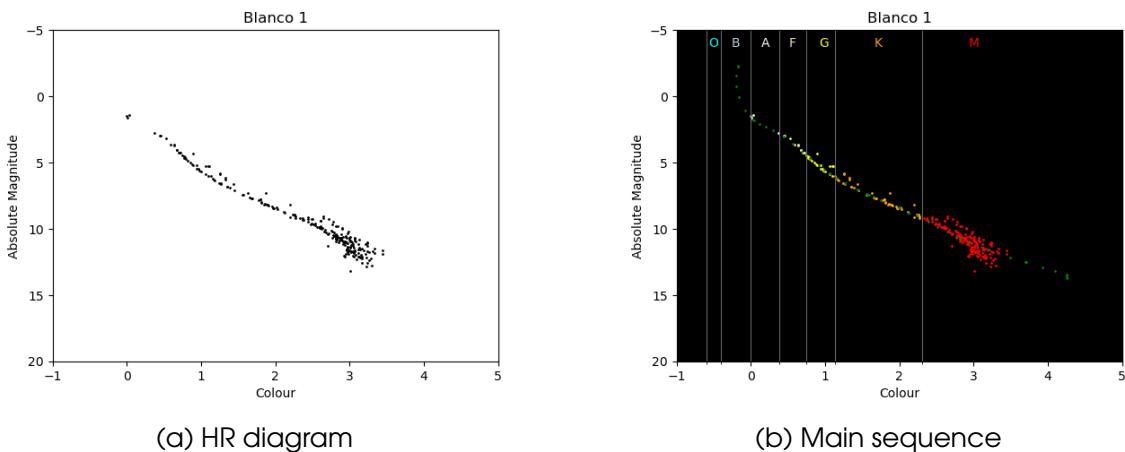


Figure 2.28: The green isochrone in plot b has an age of 100 million years and solar metallicity.

The composition of Blanco 1 observed is:

- O class: 0.0%
- B class: 0.37%
- A class: 1.10%
- F class: 5.49%
- G class: 9.89%
- K class: 21.98%
- M class: 60.81%

As we shall see, this is the closest we will get to the expected values mentioned above.

### 2.4.2 Beehive cluster

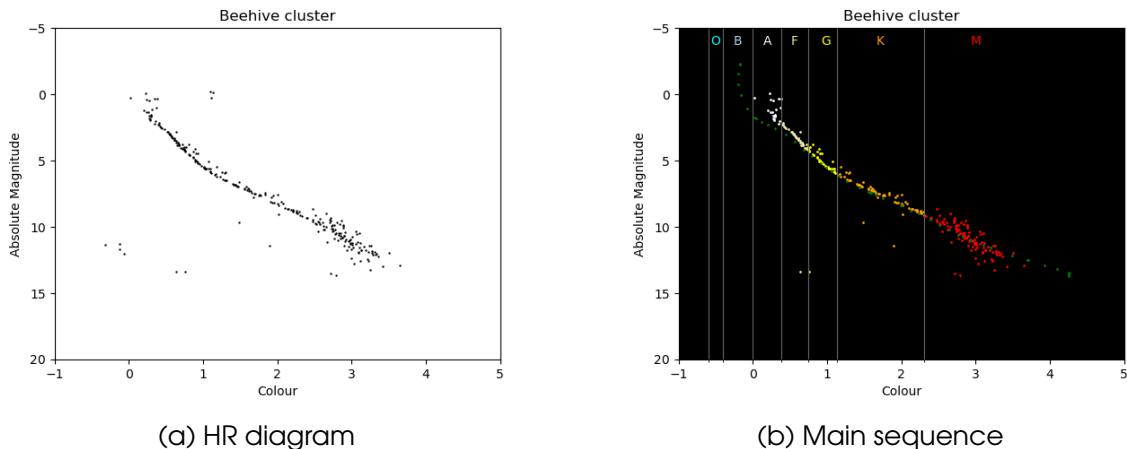


Figure 2.29: The green isochrone in plot b has an age of 100 million years and solar metallicity.

The composition of Beehive cluster observed is:

- O class: 0.0%
- B class: 0.0%
- A class: 8.02%
- F class: 13.92%
- G class: 14.77%
- K class: 24.05%
- M class: 39.24%

### 2.4.3 Southern Beehive cluster

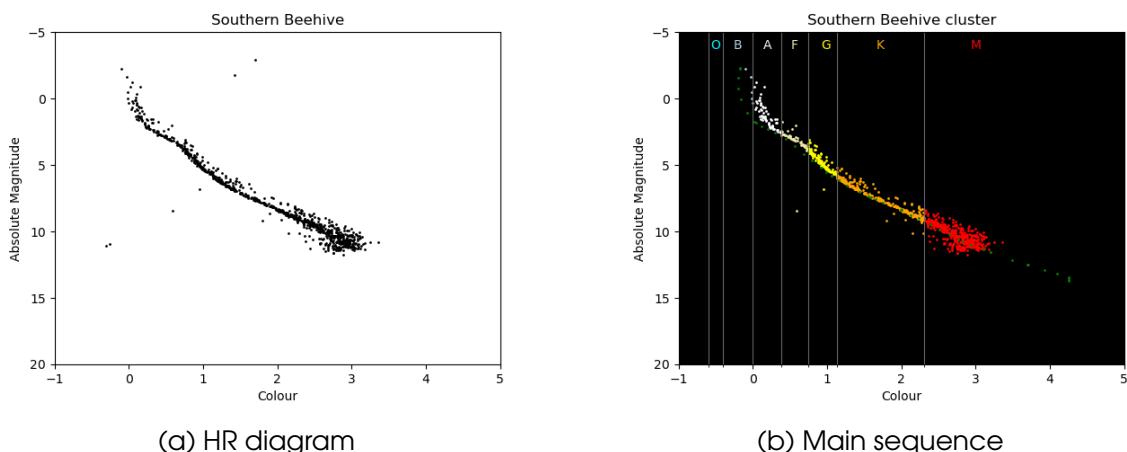


Figure 2.30: The green isochrone in plot b has an age of 100 million years and solar metallicity.

An interesting thing to observe is that there are two main sequences visible above. The composition of Southern Beehive cluster observed is:

- O class: 0.0%
- B class: 0.0%

- A class: 8.02%
- F class: 13.92%
- G class: 14.77%
- K class: 24.05%
- M class: 39.24%

#### 2.4.4 Wishing Well cluster

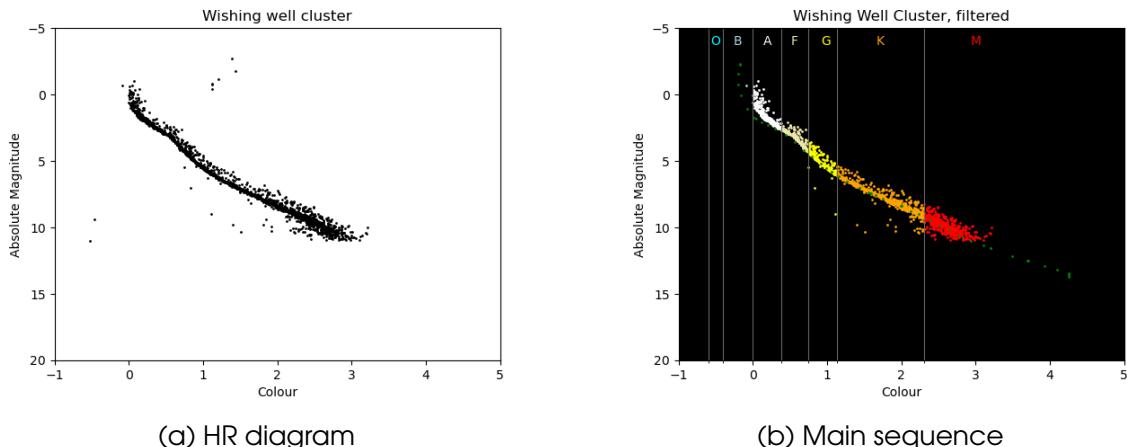
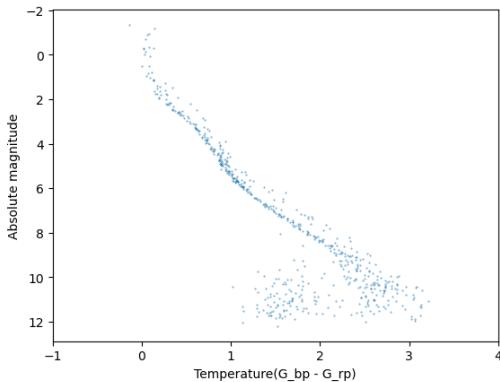


Figure 2.31: The green isochrone in plot b has an age of 100 million years and solar metallicity.

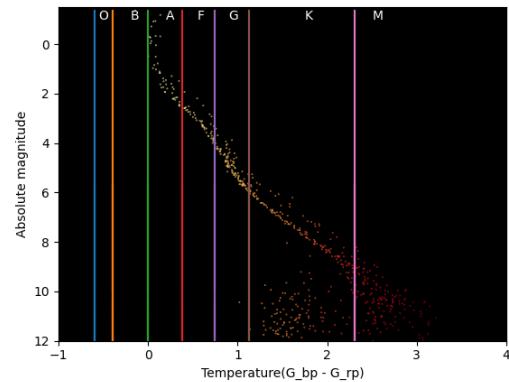
Interestingly, Wishing Well cluster seems to have two main sequences. The composition of Wishing Well cluster observed is:

- O class: 0.0%
- B class: 0.16%
- A class: 11.32%
- F class: 11.27%
- G class: 17.04%
- K class: 35.58%
- M class: 23.50%

### 2.4.5 Messier 7



(a) HR diagram



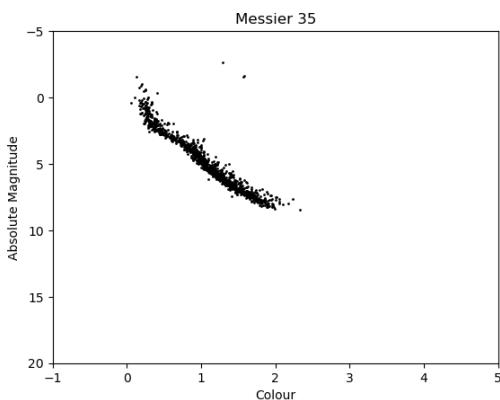
(b) Main sequence

Figure 2.32: The green isochrone in plot b has an age of 100 million years and solar metallicity.

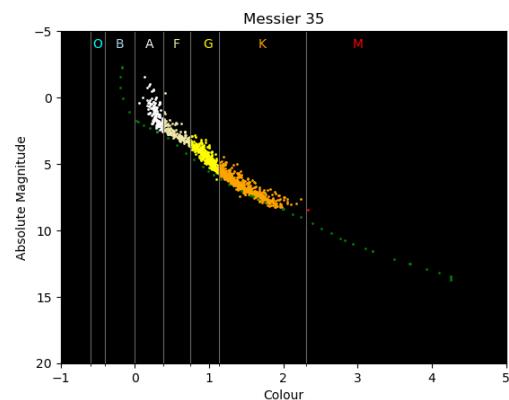
The composition of M7 cluster observed is:

- O class: 0.0%
- B class: 0.0%
- A class: 11.32%
- F class: 12.43%
- G class: 27.18%
- K class: 48.89%
- M class: 0.09%

### 2.4.6 Messier 35



(a) HR diagram



(b) Main sequence

Figure 2.33: The green isochrone in plot b has an age of 100 million years and solar metallicity.

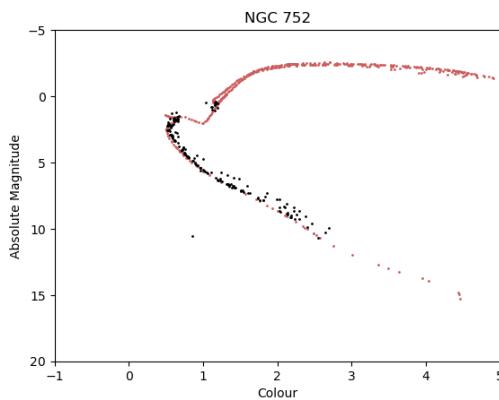
The composition of M35 cluster observed is:

- O class: 0.0%
- B class: 0.0%
- A class: 11.32%

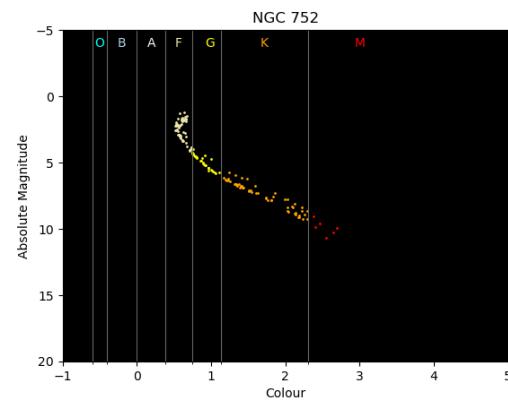
- F class: 12.43%
- G class: 27.18%
- K class: 48.89%
- M class: 0.09%

This decrease in M stars is due to their dimness. At large distances, the apparent magnitude of these stars is very low, and Gaia's lower limit of magnitude is 21. Also, it is difficult to get accurate measurement for these stars and thus get removed in the error cuts.

#### 2.4.7 NGC 752



(a) HR diagram



(b) Main sequence

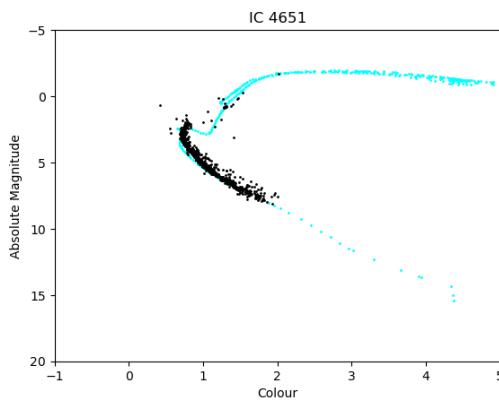
Figure 2.34: The green isochrone in plot (a) has an age of 2 billion years and metallicity of -0.1 dex.

We have an evident turn off point! This is an older cluster, which is why our previous isochrone of age 100 Myr and solar metallicity will not work. There are stars here which have entered the red giant branch.

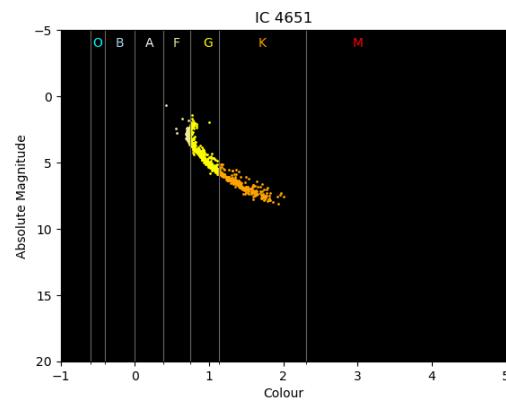
The composition of NGC 188 cluster observed is:

- O class: 0.0%
- B class: 0.0%
- A class: 0.0%
- F class: 36.42%
- G class: 20.53%
- K class: 37.75%
- M class: 3.97%

### 2.4.8 IC 4651



(a) HR diagram



(b) Main sequence

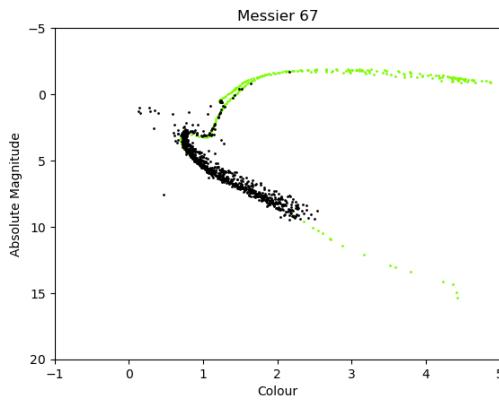
Figure 2.35: The green isochrone in plot b has an age of 3.2 billion years and metallicity of 0.1 dex.

The composition of IC4651 observed is:

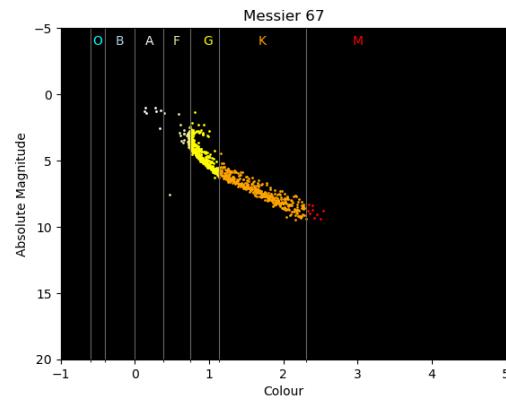
- O class: 0.0%
- B class: 0.0%
- A class: 0.0%
- F class: 15.47%
- G class: 49.87%
- K class: 34.53%
- M class: 0.0%

IC 4651 is older than NGC 752. It also seems to have a higher metallicity.

### 2.4.9 Messier 67



(a) HR diagram



(b) Main sequence

Figure 2.36: The green isochrone in plot (a) has an age of 5 billion years and solar metallicity.

The composition of Messier 67 observed is:

- O class: 0.0%
- B class: 0.0%

- A class: 0.75%
- F class: 8.20%
- G class: 47.60%
- K class: 42.28%
- M class: 0.96%

#### 2.4.10 NGC 188

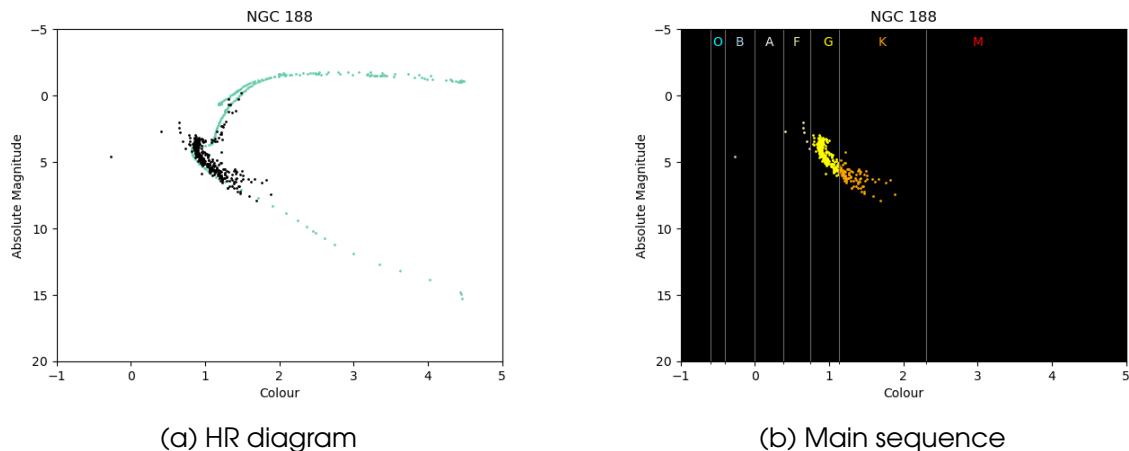


Figure 2.37: The green isochrone in plot (a) has an age of 11 billion years and metallicity of -0.1 dex.

This is the worst open cluster we took the data for. It further solidifies our reasoning that as distance increases, the data worsens as this cluster is around 2kpc away. (The HR diagram in plot (a) has 341 stars only!).

We will explore a way around this problem in the next section for globular clusters, which are even further away.

NGC 188, and Messier 67 are among the oldest open clusters known.

The composition of NGC 188 observed is:

- O class: 0.0%
- B class: 0.38%
- A class: 0.75%
- F class: 1.90%
- G class: 68.25%
- K class: 28.89%
- M class: 0.0%

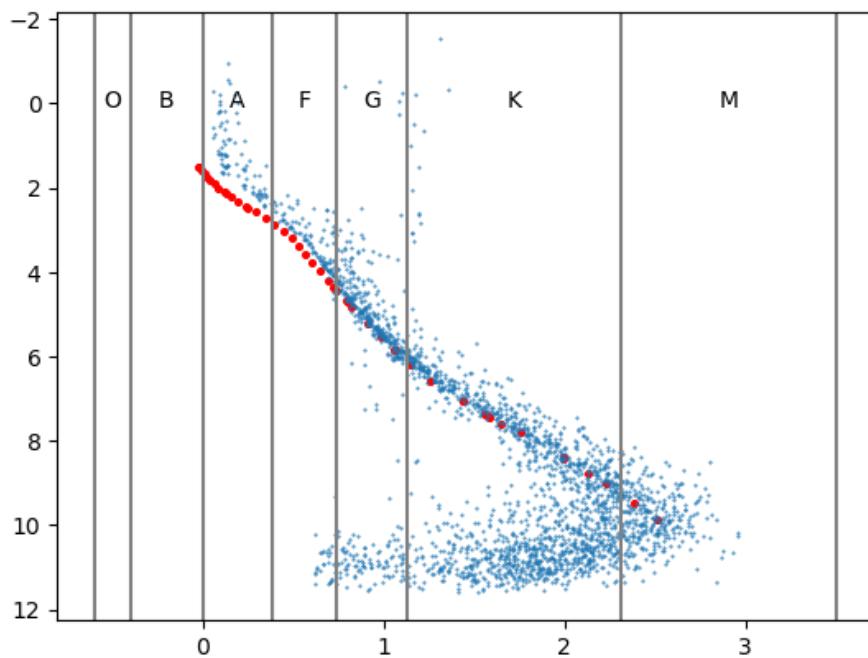
**2.4.11 M48**

Figure 2.38: HR diagram with isochrone fitting

Above isochrone in red has age of 100 million years and 0 metallicity. The composition of M48 as observed is:

- Percent of O type is 0.0
- Percent of B type is 0.0
- Percent of A type is 3.54
- Percent of F type is 5.84
- Percent of G type is 17.36
- Percent of K type is 59.36
- Percent of M type is 12.65

### 2.4.12 Trapezium Cluster

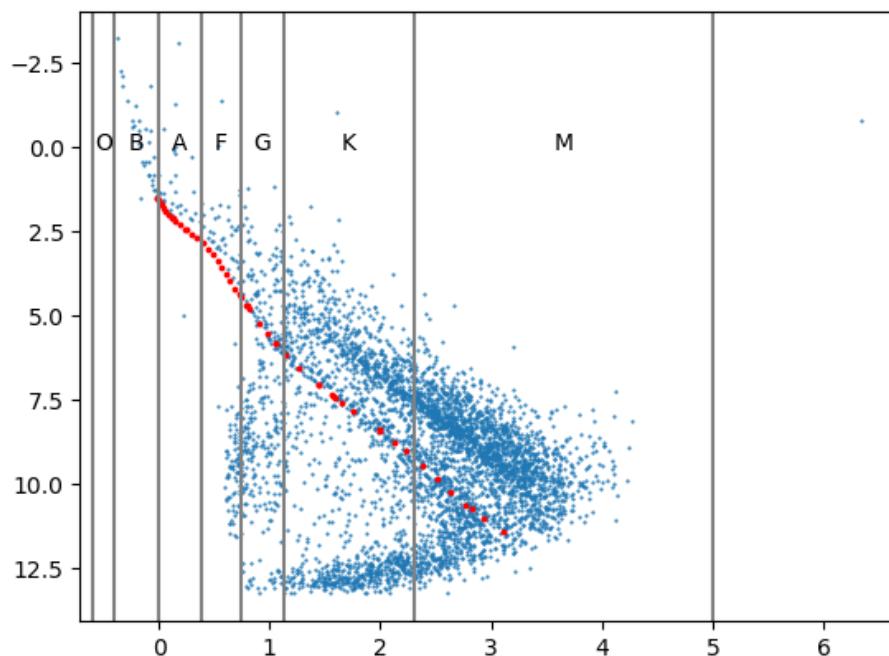


Figure 2.39: HR diagram with isochrone fitting

Above isochrone in red has age of 100 million years and 0 metallicity. The composition of Trapezium Cluster as observed is:

- Percent of O type is 0.0
- Percent of B type is 0.67
- Percent of A type is 0.72
- Percent of F type is 2.54
- Percent of G type is 6.95
- Percent of K type is 32.52
- Percent of M type is 54.56

### 2.4.13 Messier 45

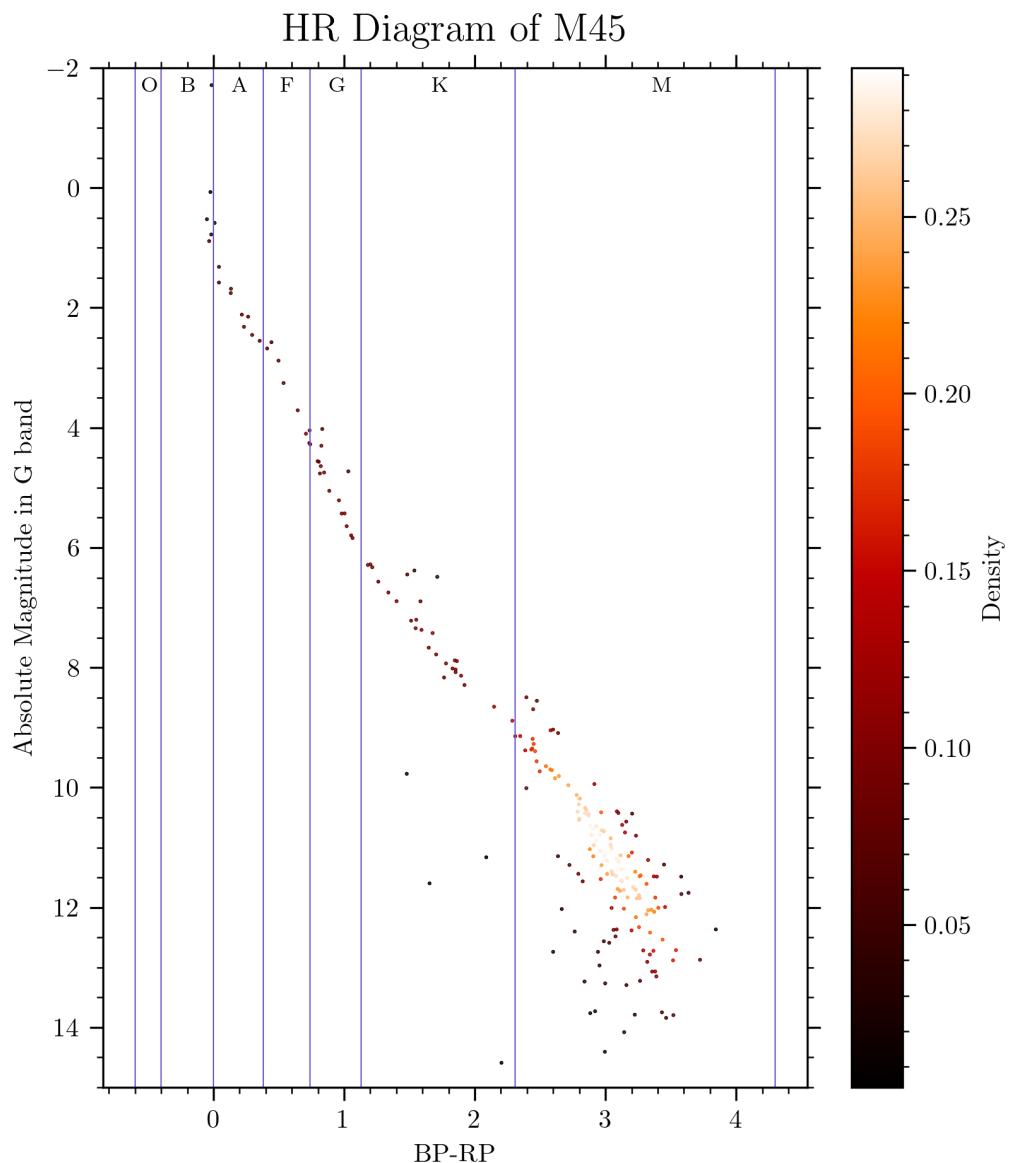


Figure 2.40: M45: HR Diagram

The composition of Messier 45 observed is:

- O class: 0.0%
- B class: 2.22%
- A class: 4.44%
- F class: 3.56%
- G class: 7.11%
- K class: 14.67%
- M class: 68.0%

#### 2.4.14 Messier 50

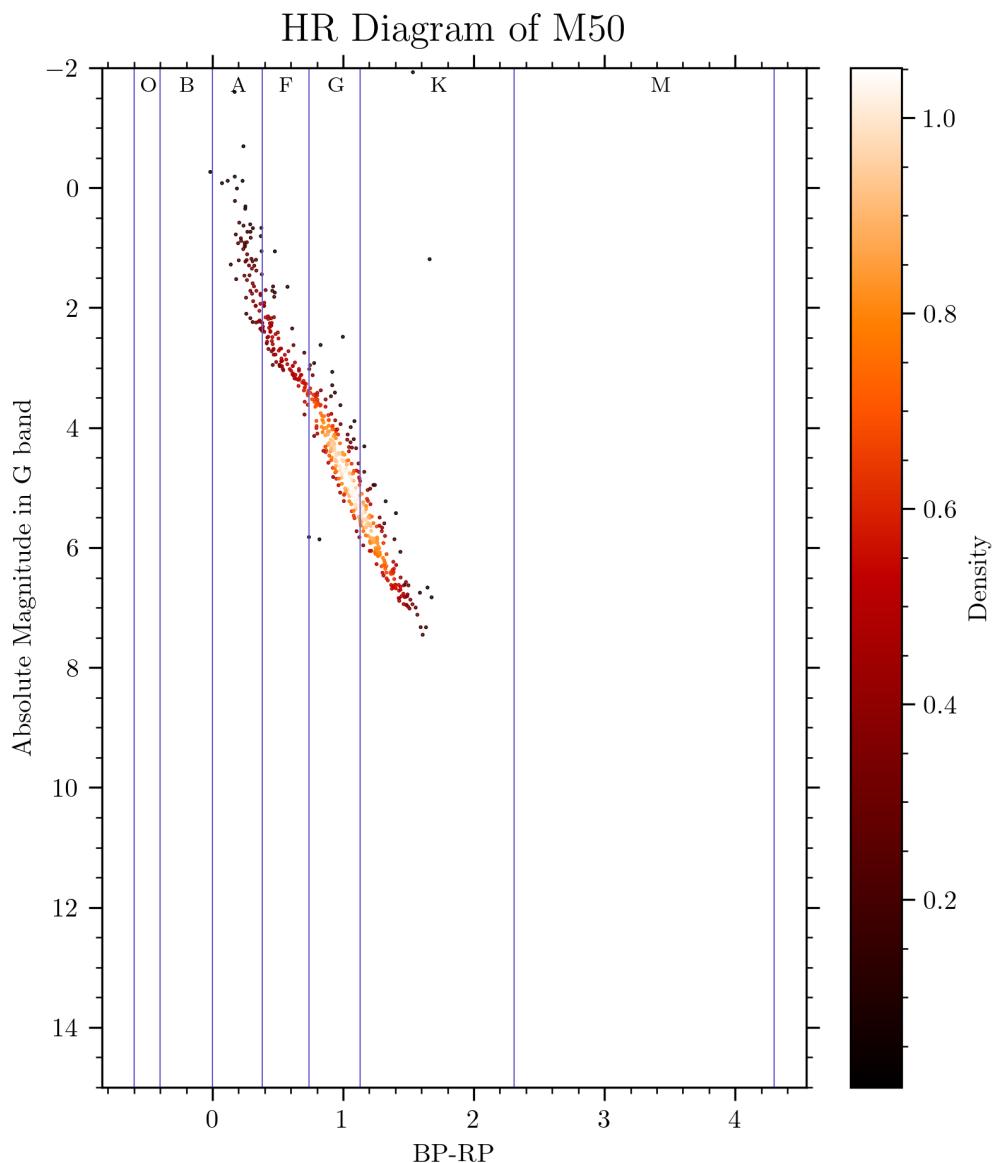


Figure 2.41: M50: HR Diagram

The composition of Messier 45 observed is:

- O class: 0.0%
- B class: 0.18%
- A class: 12.41%
- F class: 18.07%
- G class: 39.78%
- K class: 29.56%
- M class: 0.0%

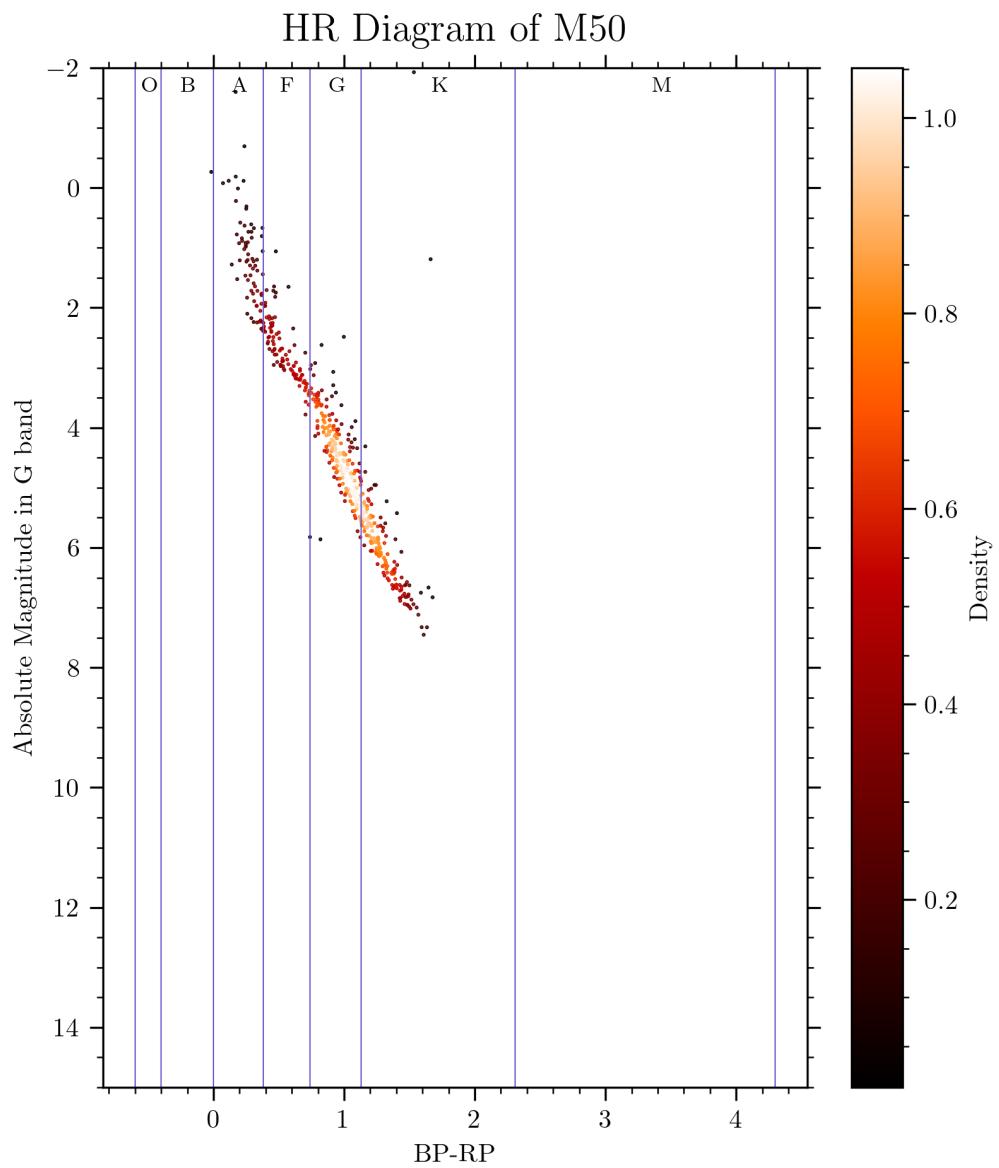
**2.4.15 Messier 38**

Figure 2.42: M38: HR Diagram

The composition of Messier 38 observed is:

- O class: 0.0%
- B class: 0.0%
- A class: 5.69%
- F class: 23.08%
- G class: 34.31%
- K class: 36.92%
- M class: 0.0%

### 2.4.16 Lambda Orionis

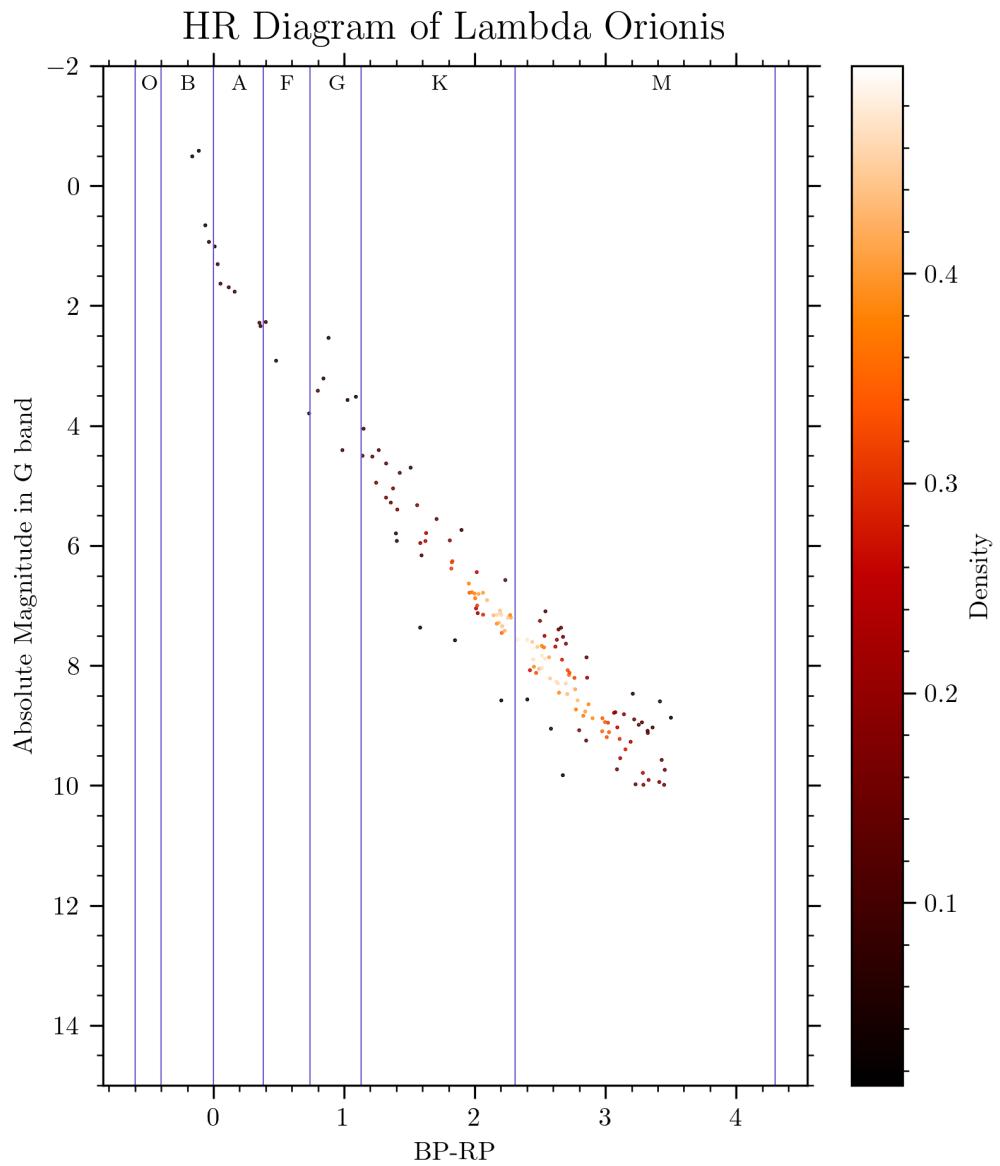


Figure 2.43: Lambda Orionis: HR Diagram

The composition of Lambda Orionis observed is:

- O class: 0.0%
- B class: 2.44%
- A class: 4.27%
- F class: 1.83%
- G class: 3.66%
- K class: 34.15%
- M class: 53.66%

### 2.4.17 Alpha Persei Cluster

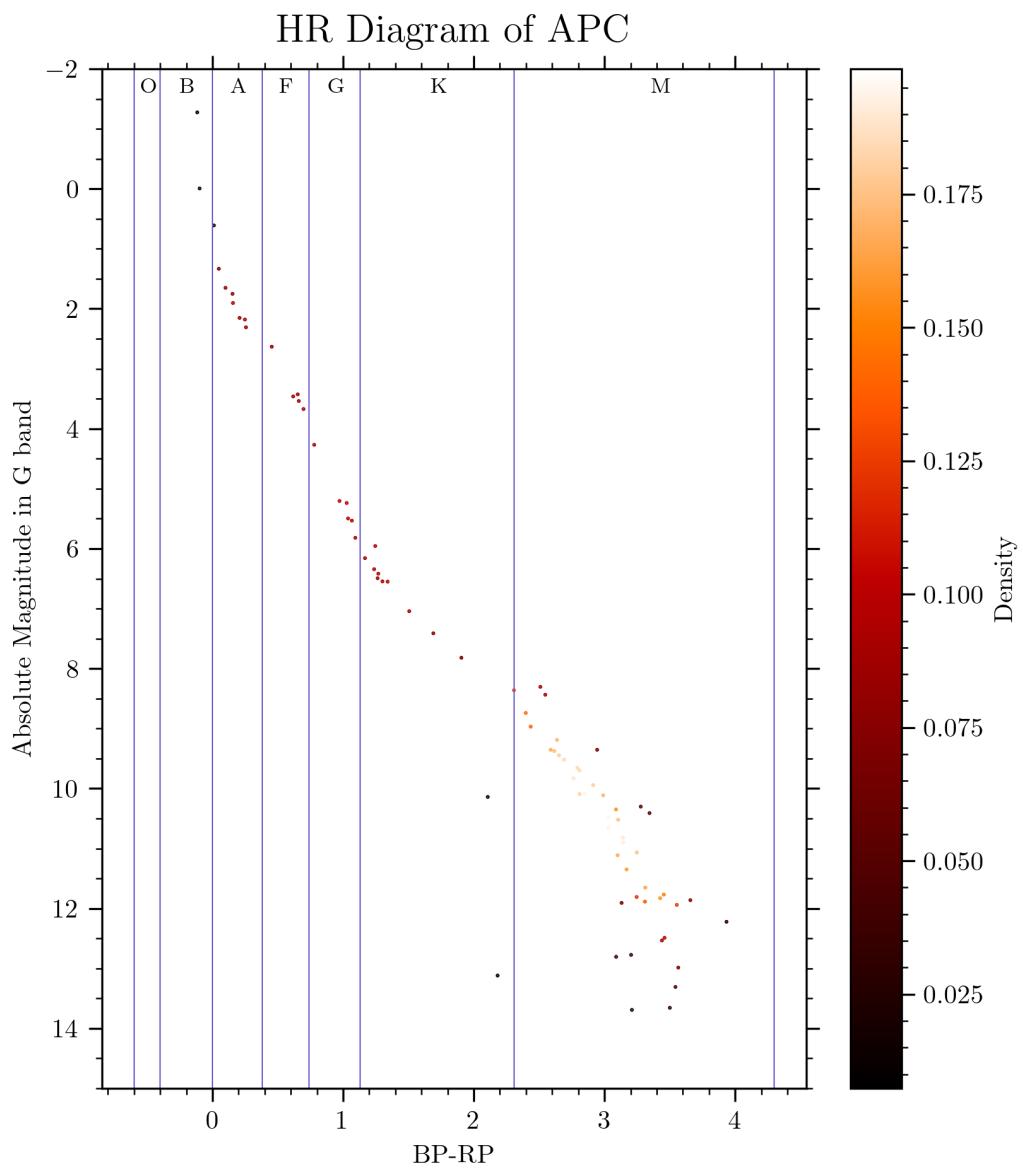


Figure 2.44: Alpha Persei Cluster: HR Diagram

The composition of Alpha Persei Cluster observed is:

- O class: 0.0%
- B class: 2.47%
- A class: 9.88%
- F class: 6.17%
- G class: 7.41%
- K class: 16.05%
- M class: 58.02%

## 2.5 Distribution in Position and Velocity Space

In this section, we shall take the mean position and proper motion to be that of the centre of the cluster and find distance and velocity distributions (not in radial direction with respect to the Earth, since Gaia DR3 does not provide that data for many stars) for all spectral classes. Since no O class star was observed, it shall be omitted from our histograms.

### 2.5.1 Blanco 1 cluster

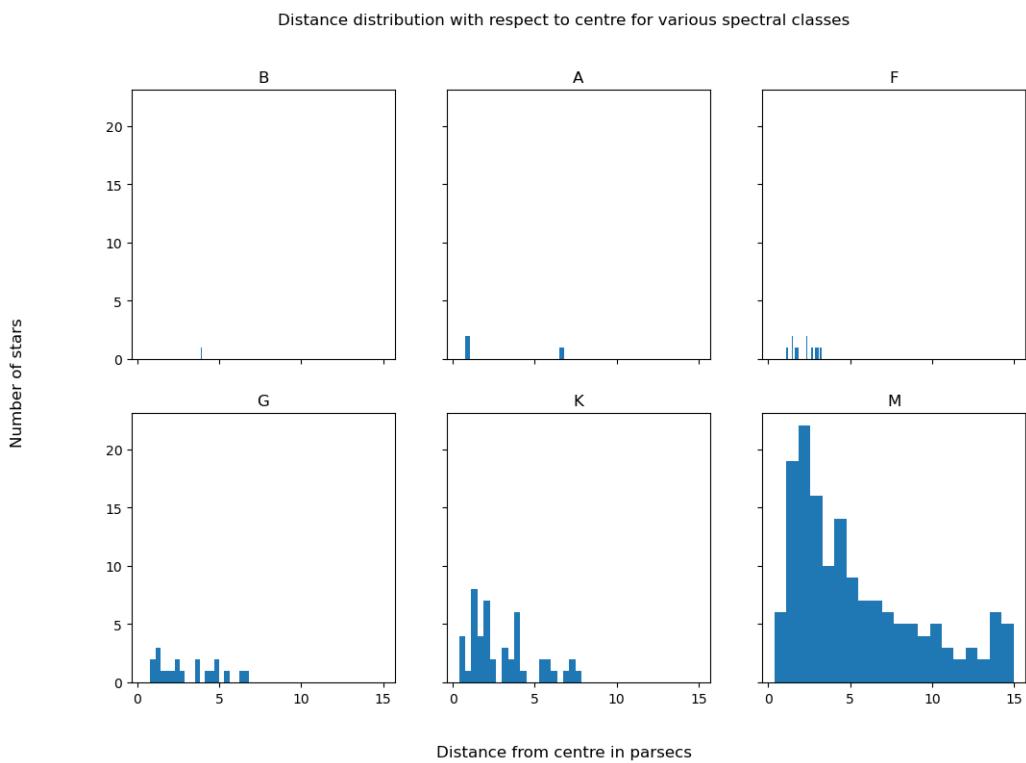


Figure 2.45: Distance distribution for Blanco 1 stars

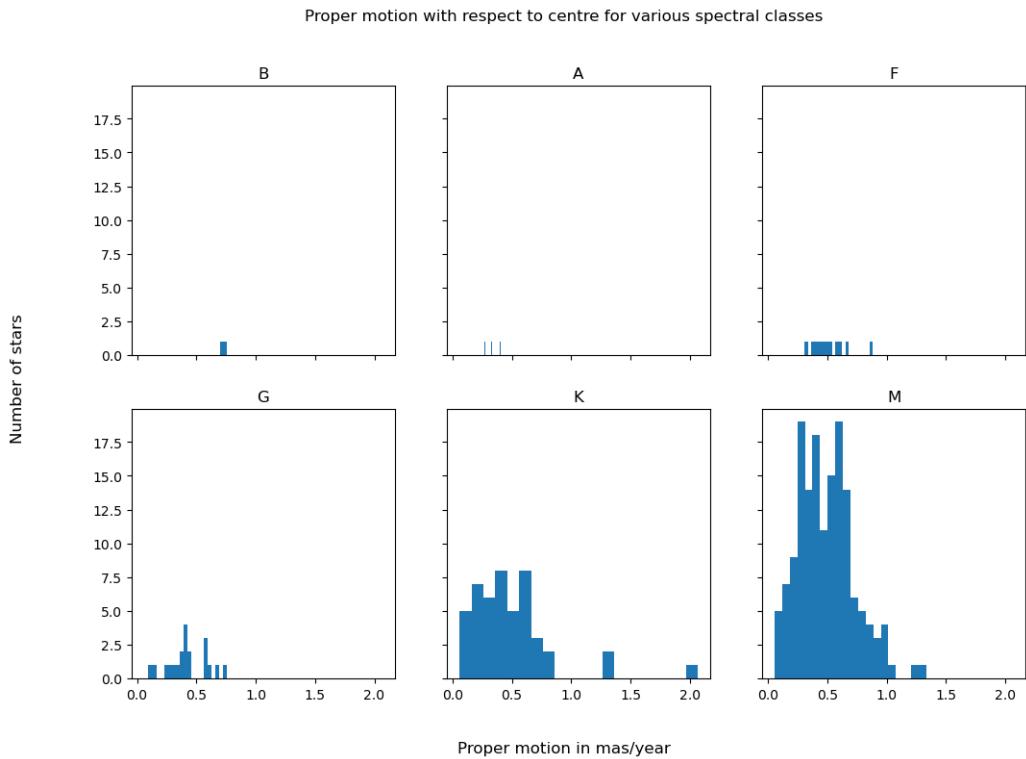


Figure 2.46: Proper motion for Blanco 1 stars

### 2.5.2 Beehive cluster

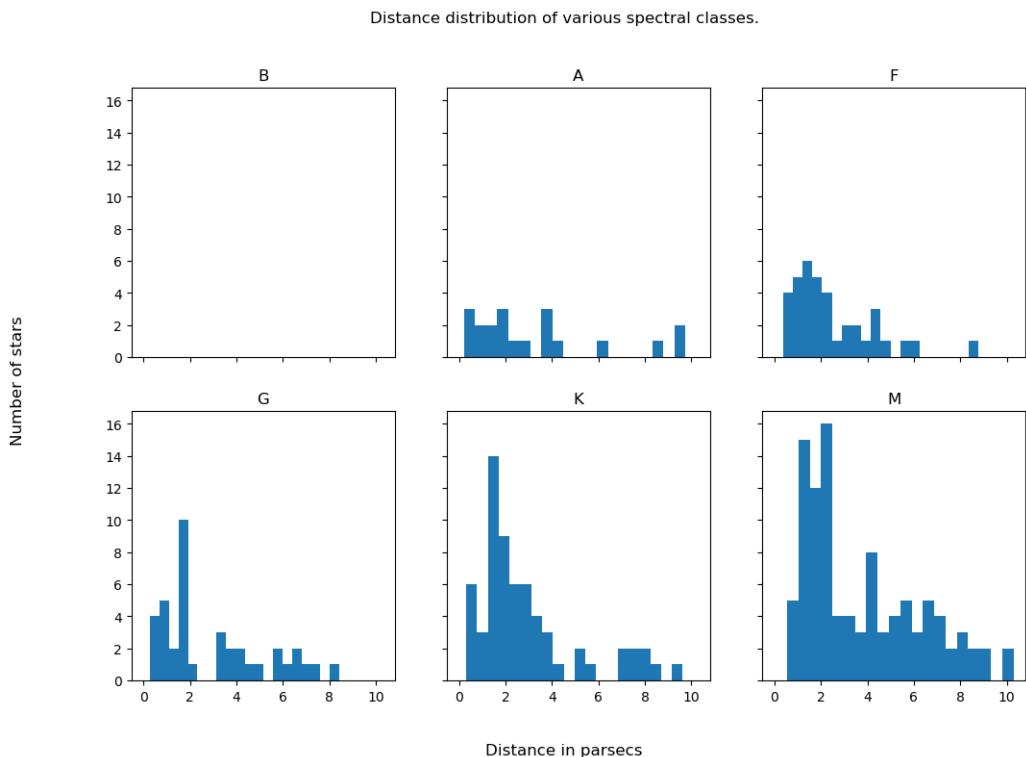


Figure 2.47: Distance distribution for Beehive stars

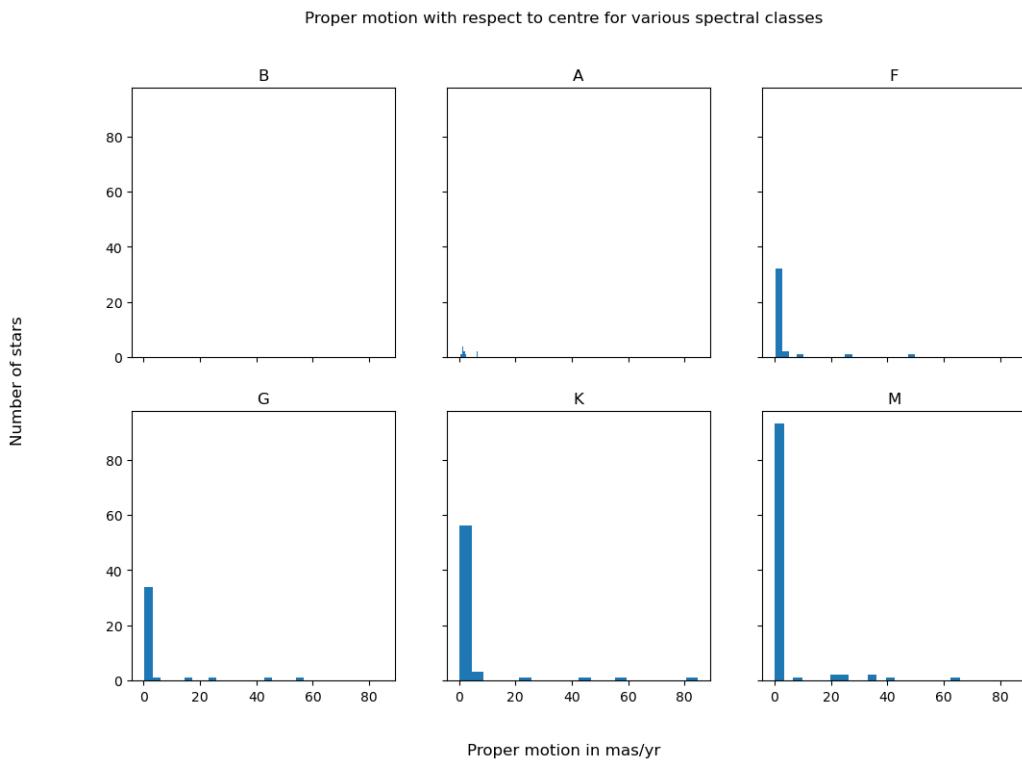


Figure 2.48: Proper motion for Beehive stars

### 2.5.3 Southern Beehive cluster

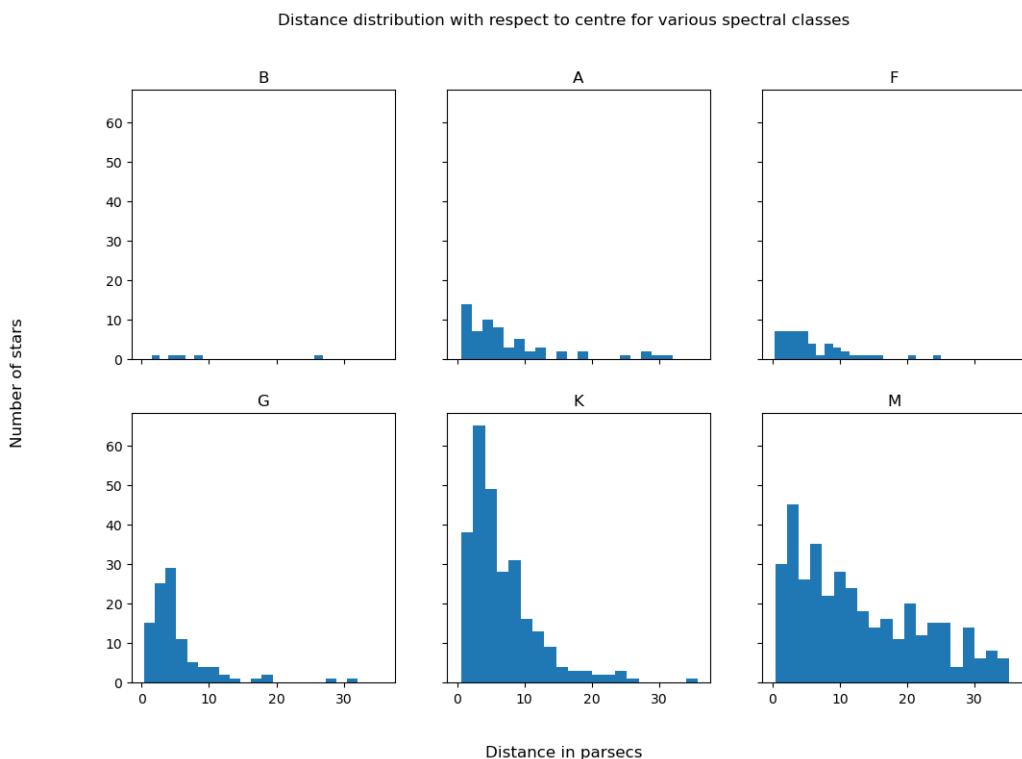


Figure 2.49: Distance distribution for Southern Beehive stars

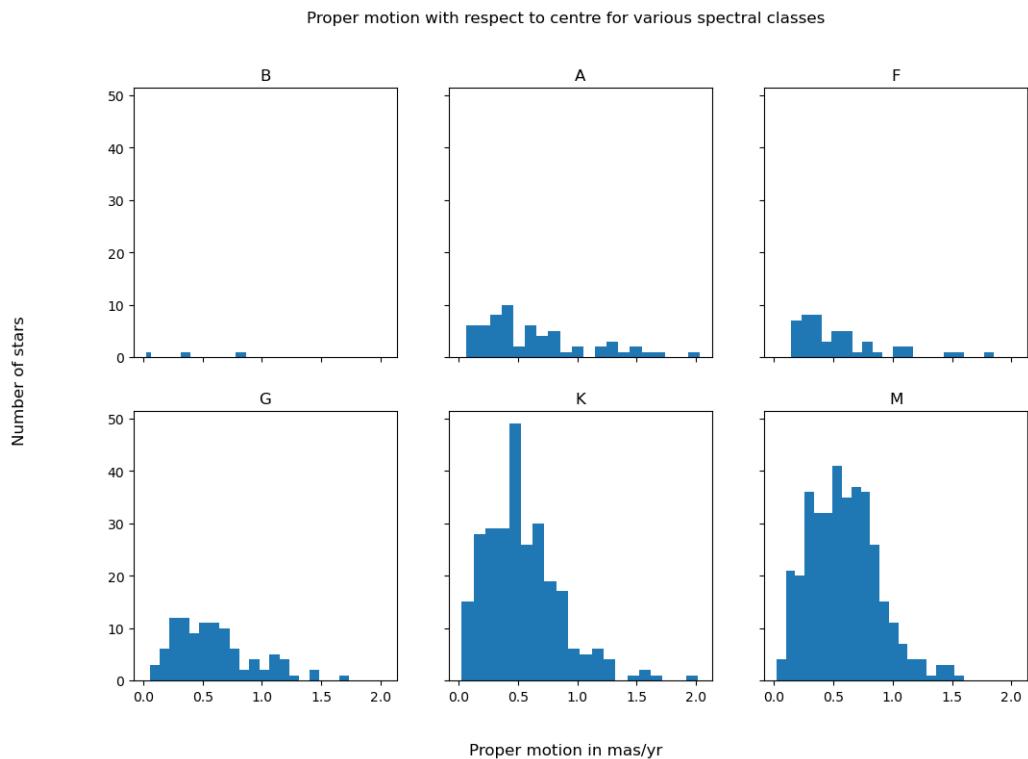


Figure 2.50: Proper motion for Southern Beehive stars

#### 2.5.4 Wishing Well Cluster

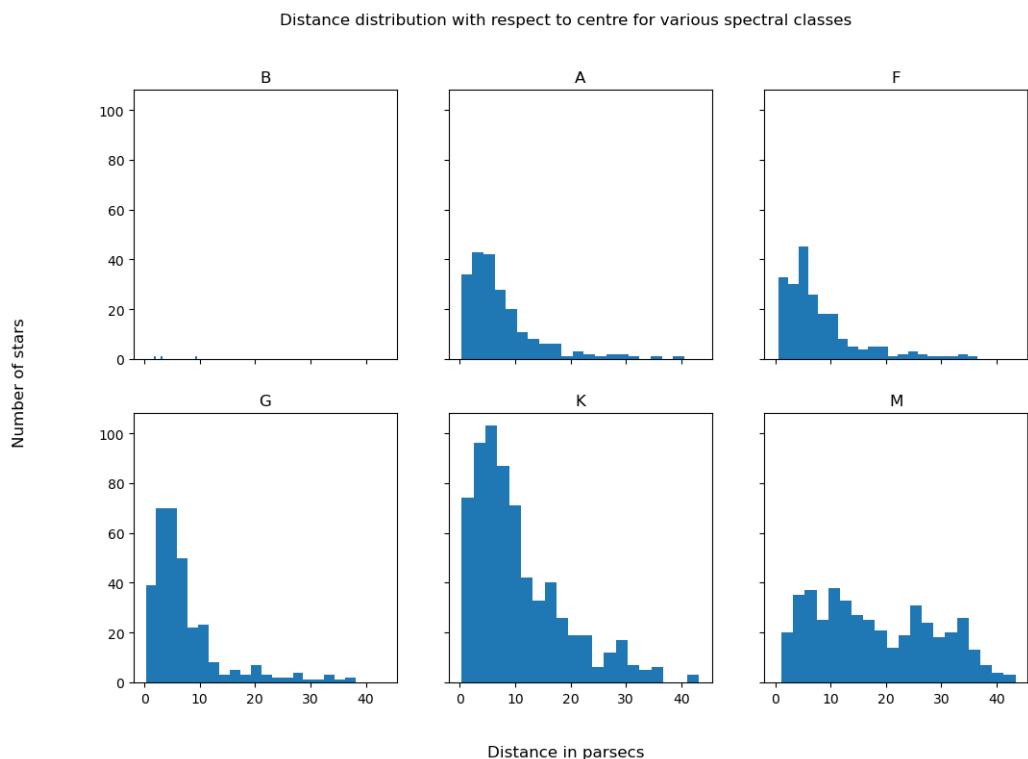


Figure 2.51: Distance distribution for Wishing Well stars

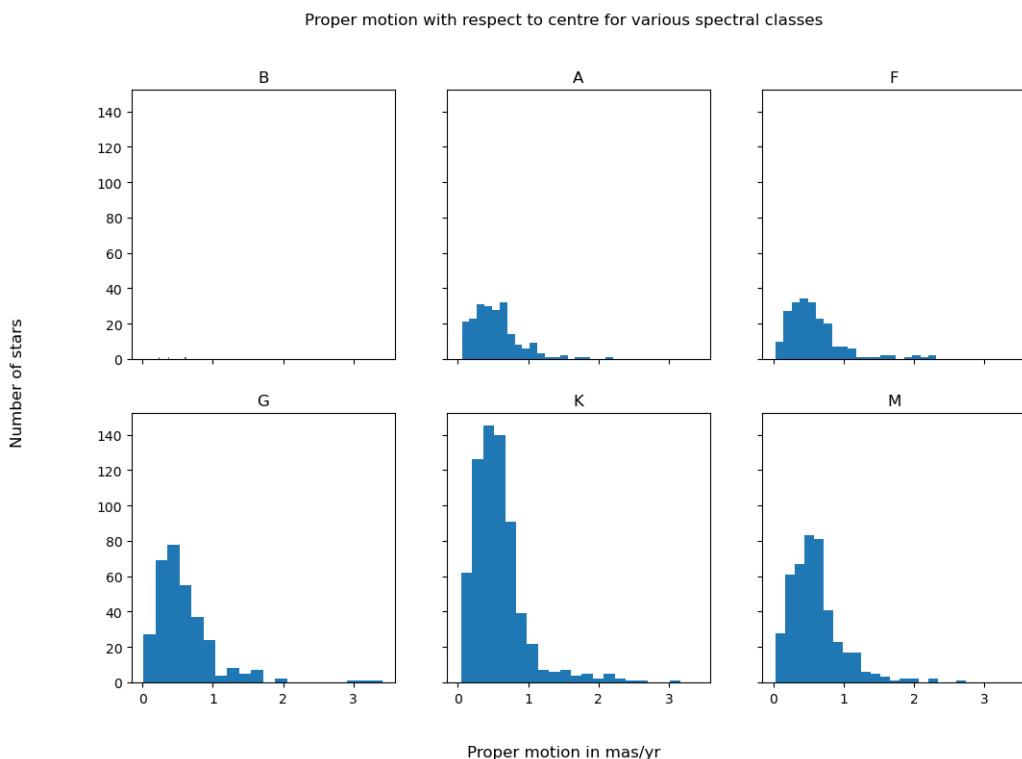


Figure 2.52: Proper motion for Wishing Well stars

### 2.5.5 Messier 35

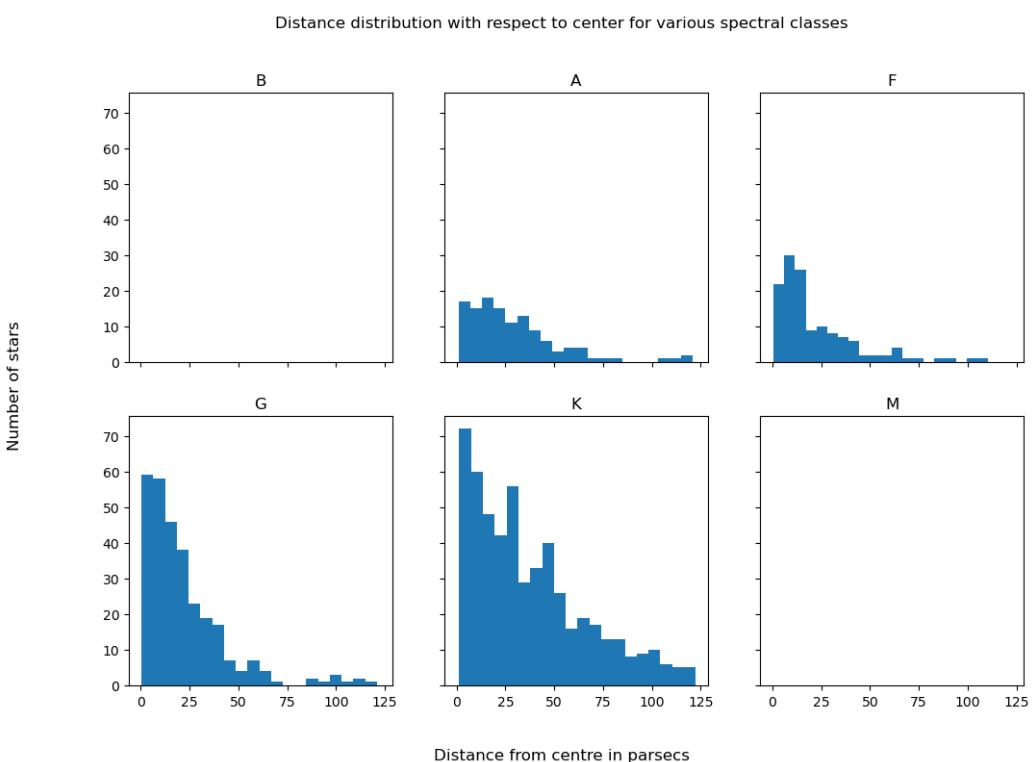


Figure 2.53: Distance distribution for M35 stars

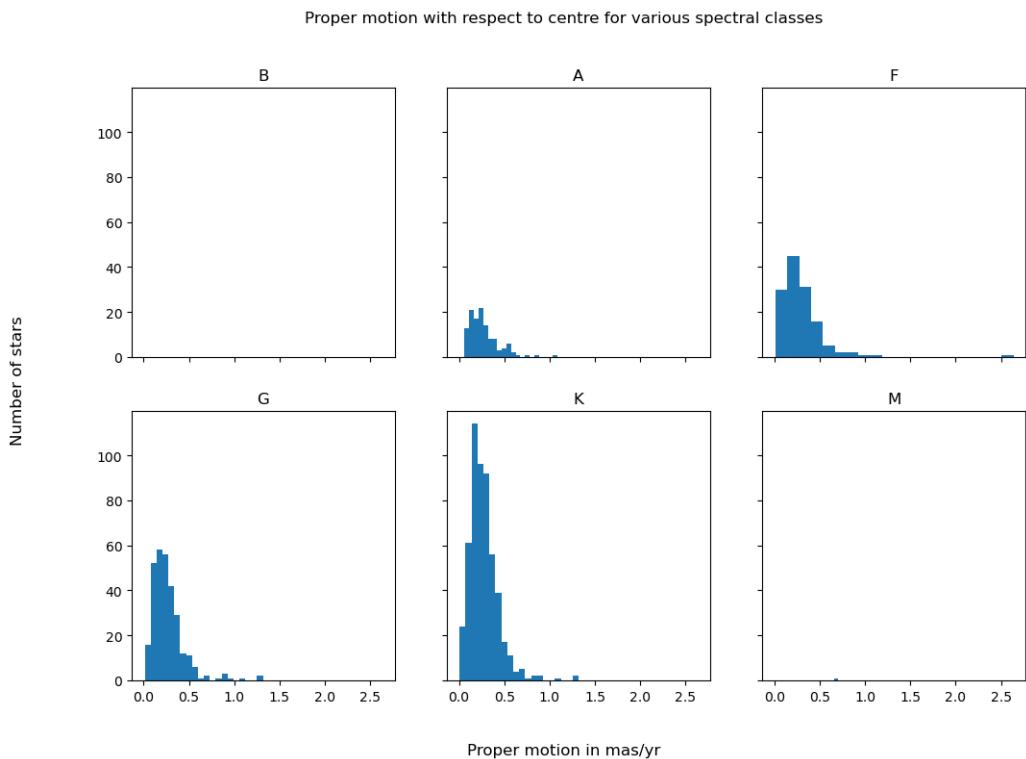


Figure 2.54: Proper motion for M35 stars

## 2.5.6 NGC 752

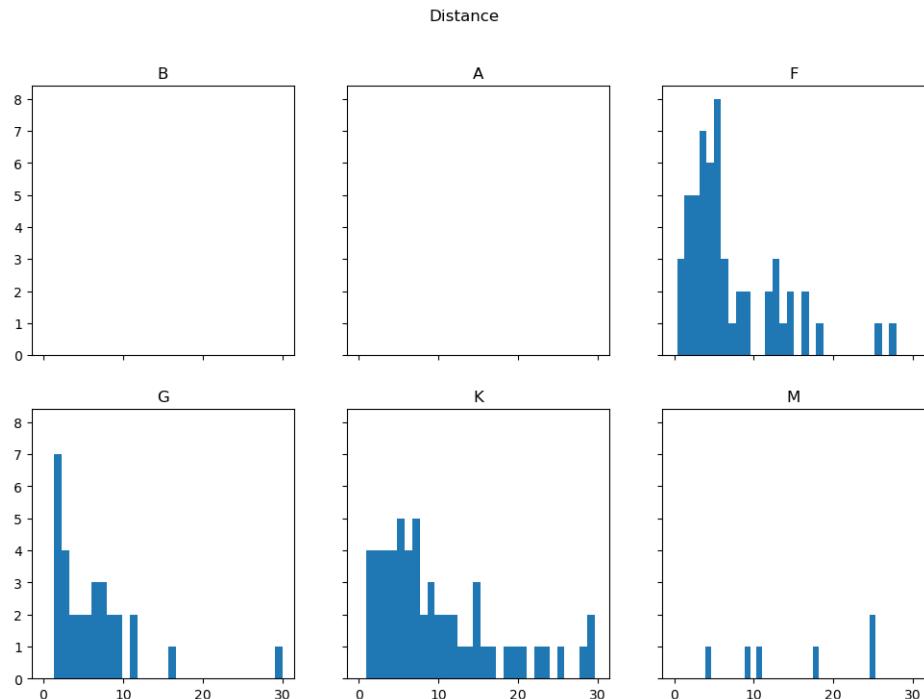


Figure 2.55: Distance distribution for M35 stars

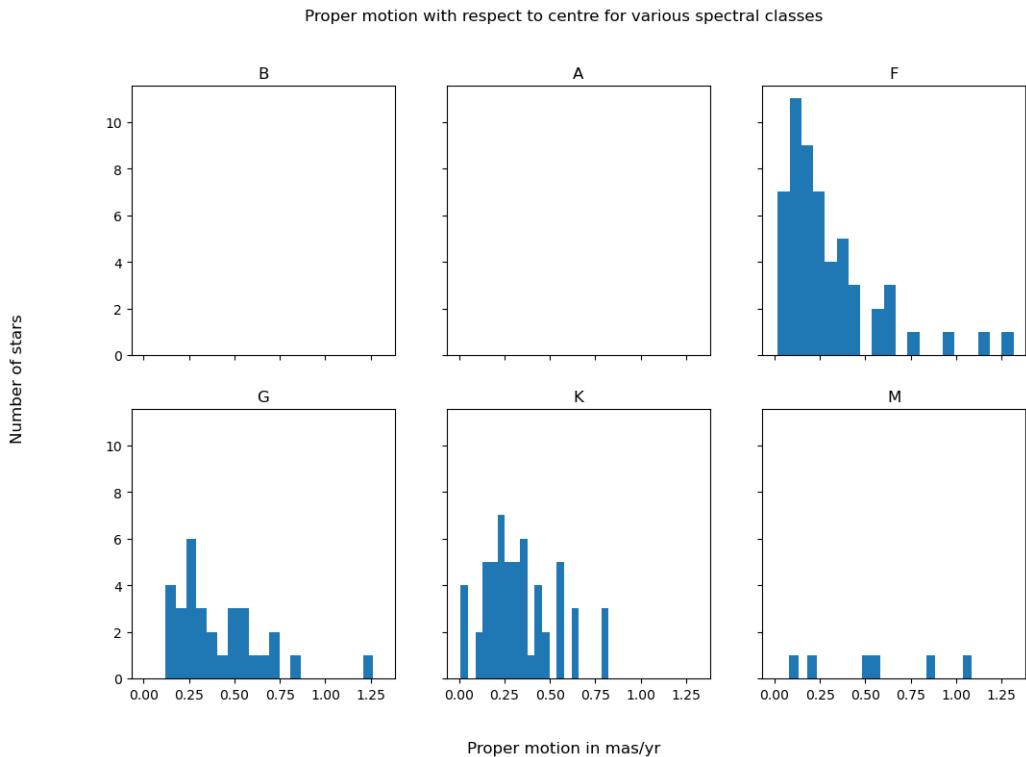


Figure 2.56: Proper motion for M35 stars

### 2.5.7 IC 4651

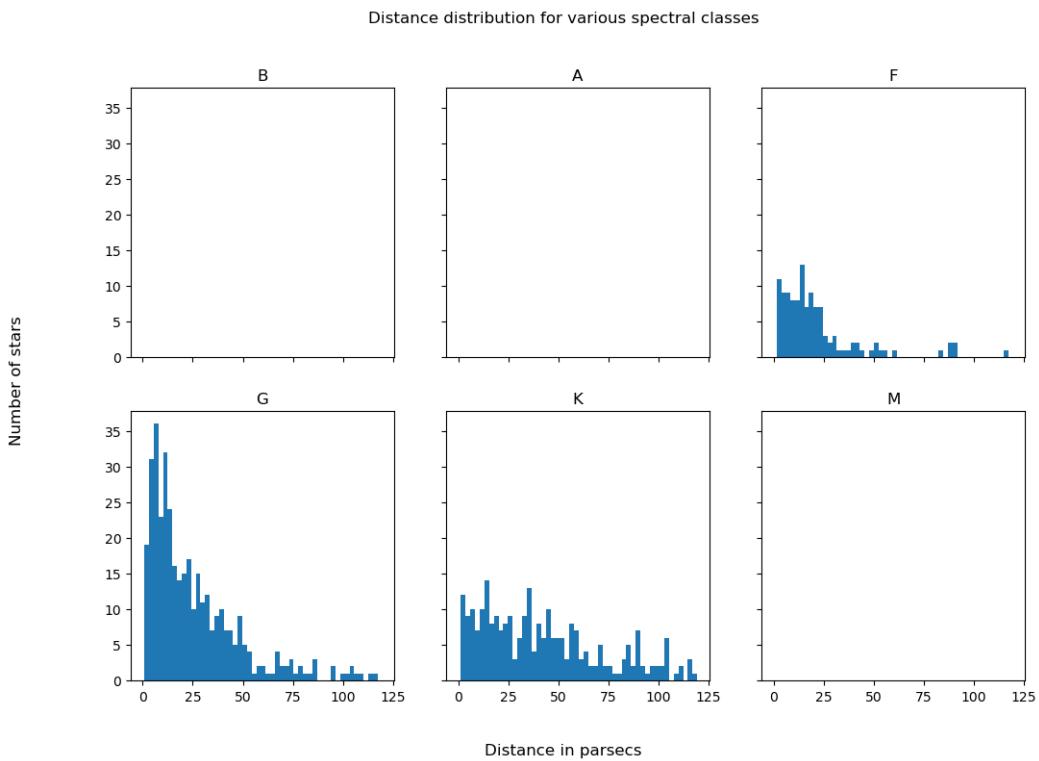


Figure 2.57: Distance distribution for M35 stars

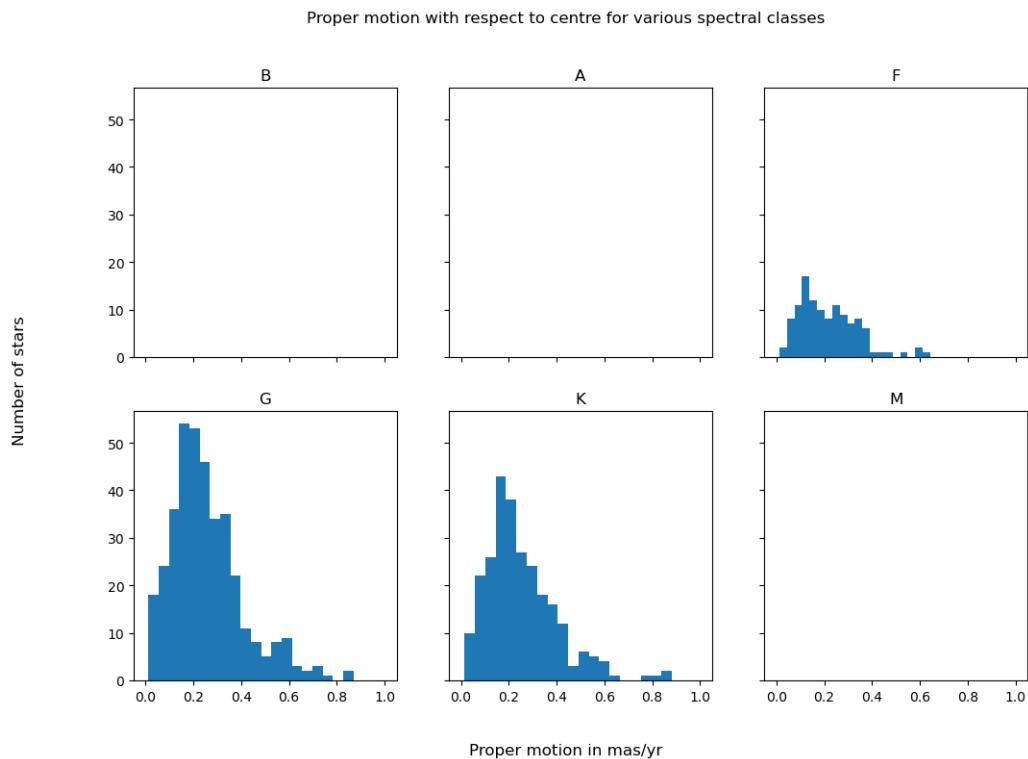


Figure 2.58: Proper motion for M35 stars

### 2.5.8 Messier 67

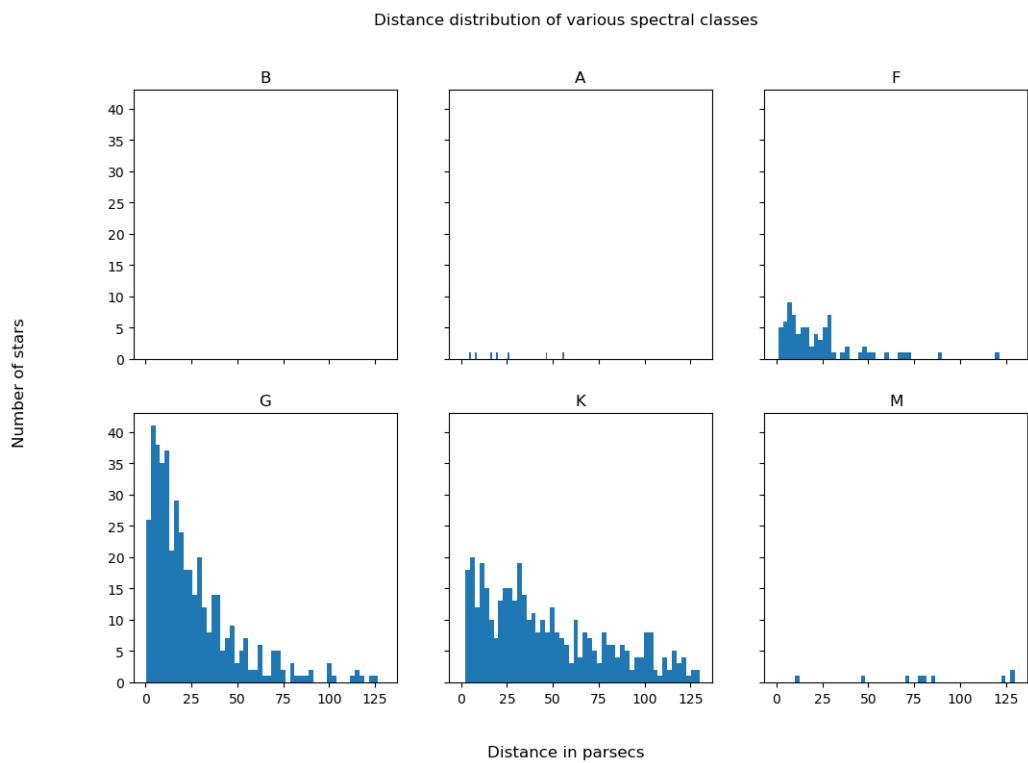


Figure 2.59: Distance distribution for M35 stars

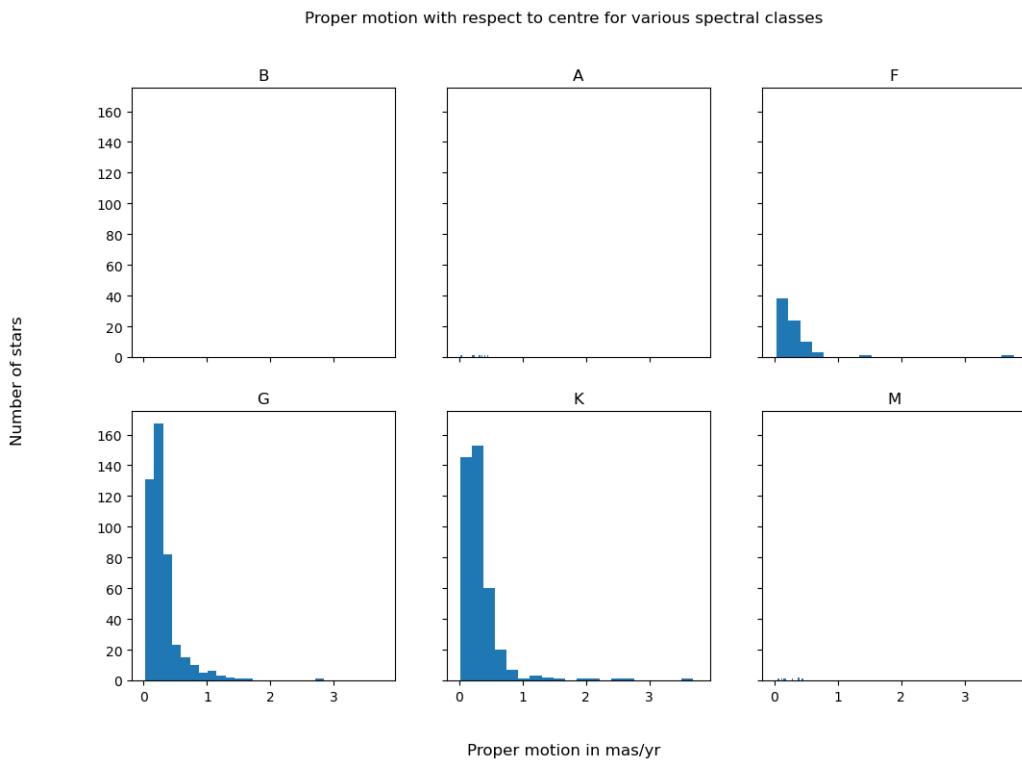
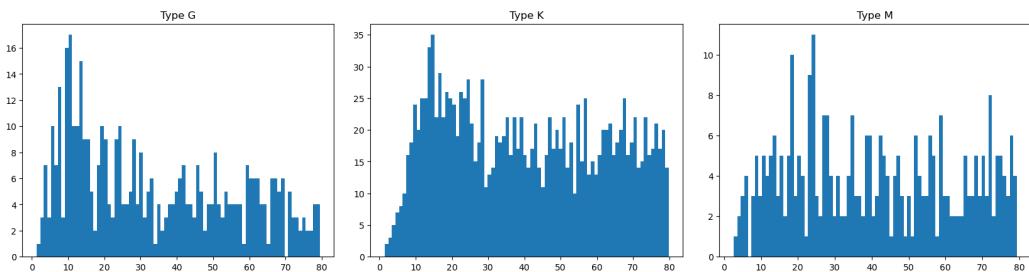
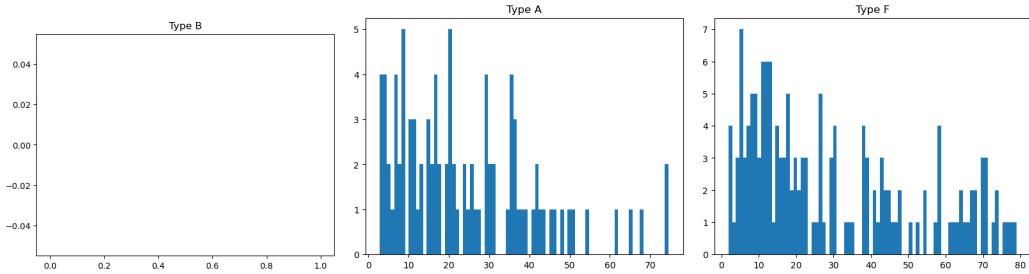
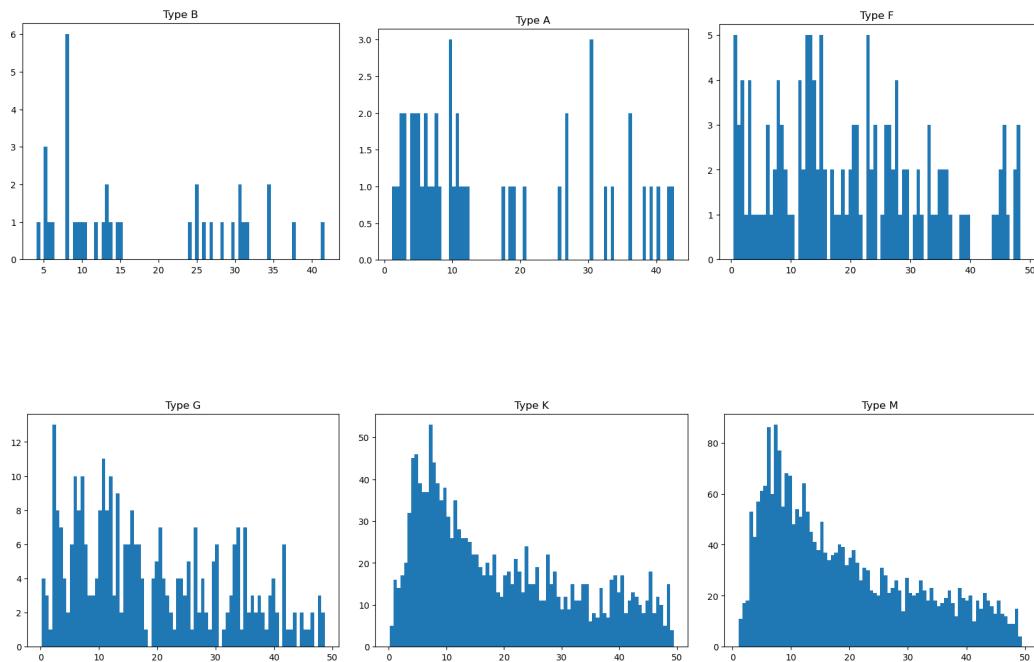


Figure 2.60: Proper motion for M35 stars

### 2.5.9 M48



### 2.5.10 Trapezium Cluster



### 2.5.11 NGC 188

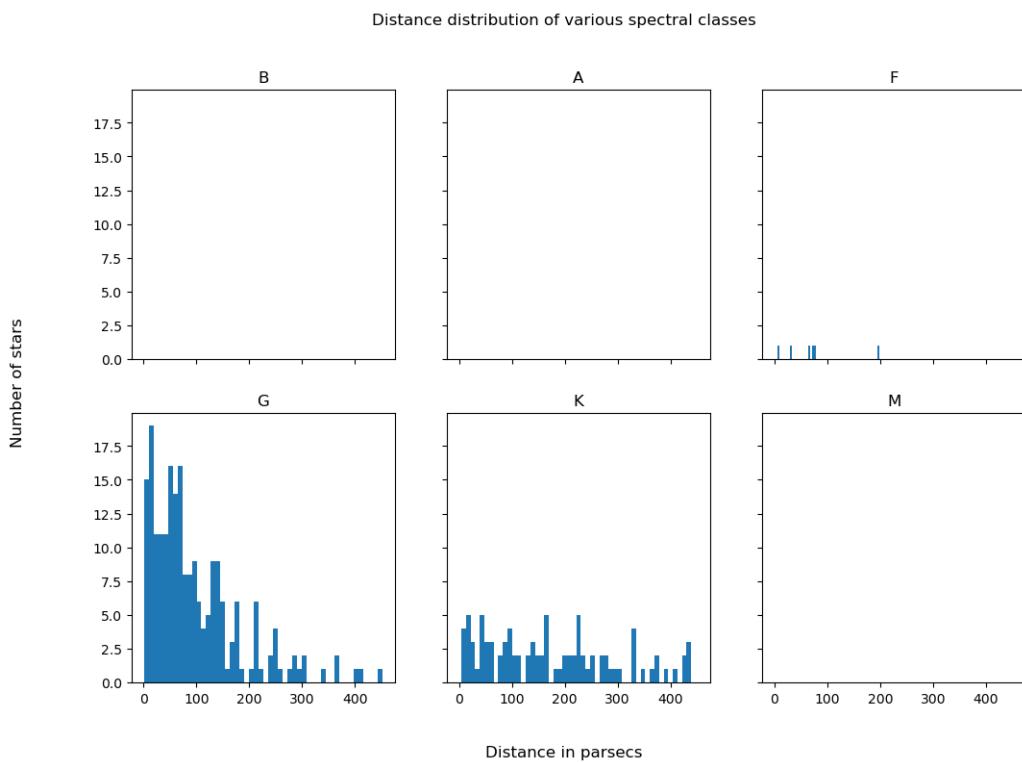


Figure 2.65: Distance distribution for M35 stars

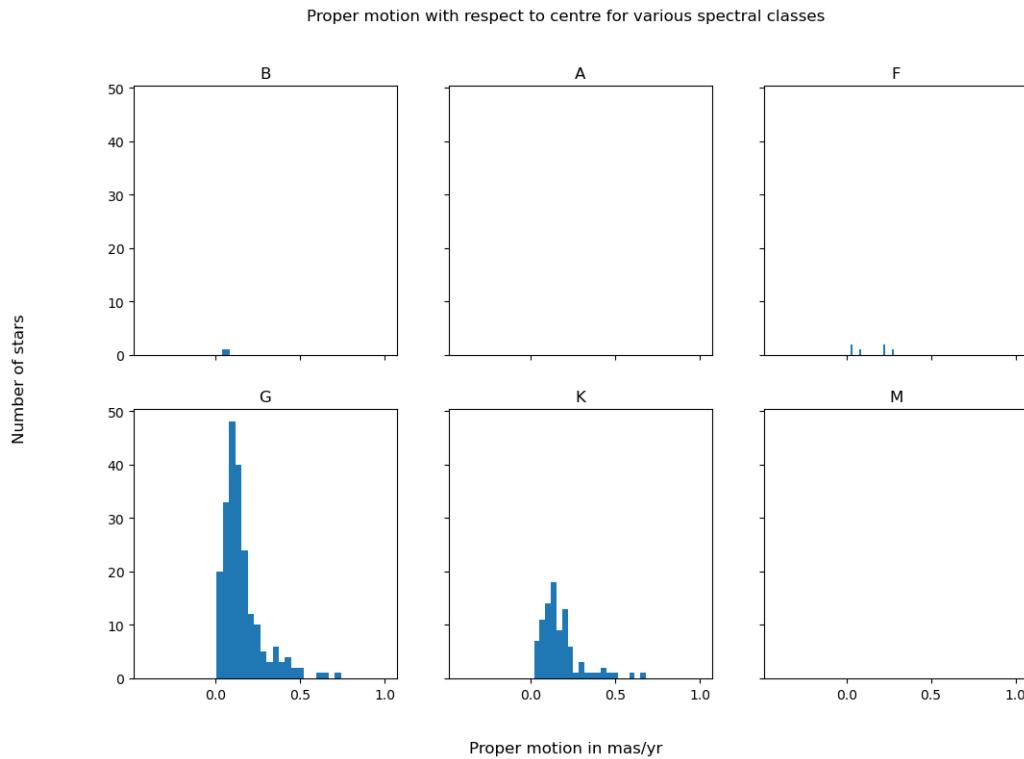


Figure 2.66: Proper motion for M35 stars

A trend is visible among all open clusters; the distance distribution is more spread for M stars, while B stars tend to be more towards the centre of the cluster.

## 2.6 Comparing ages of different clusters

One way to compare ages of clusters, as aforementioned in section 2.3, was by comparing the position of turn off points. There, we verified it by comparing various isochrones of same age and metallicity.

We will now verify it using the HR diagrams we obtained above. Here are 4 clusters chosen to show the difference. All of them have a visible turn-off point to make the comparison easier.

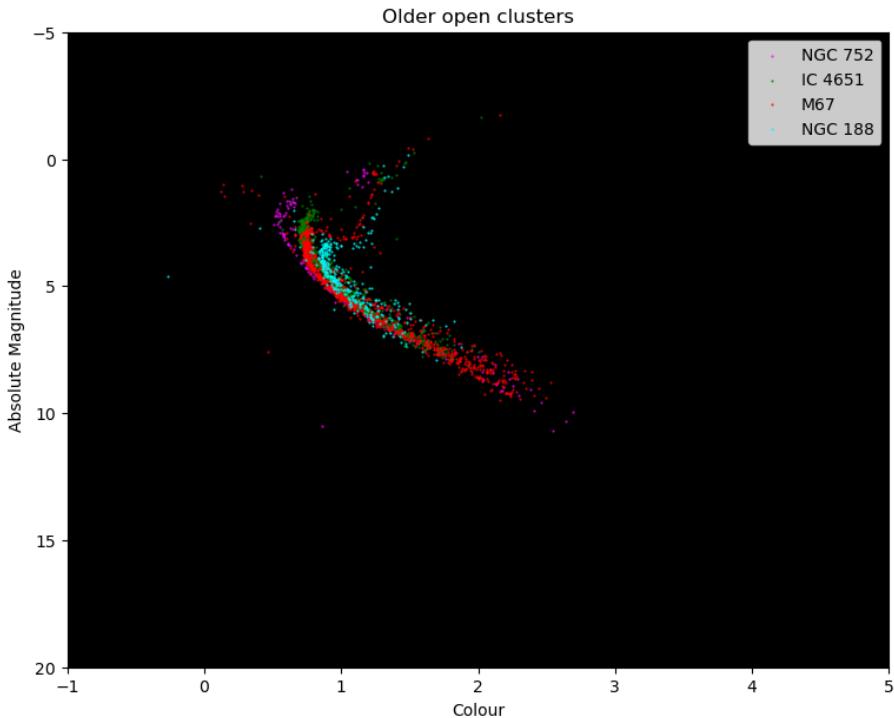


Figure 2.67: Old open clusters.

The main sequences are nearly overlapping, so the different metallicities should not be an issue while comparing. This means that the ages of the clusters are in the following order:

NGC 188 (11 billion) > M67 (5 billion) > IC 4651 (3.2 billion) > NGC 752 (2 billion).

The value inside the brackets are ages of the isochrones on which their HR diagrams were overlaid.

## 2.7 Empirical Initial Mass Function

The Initial Mass Function (IMF) is an empirical function that gives the initial mass distribution of stars for a population of stars. This is different from Present-Day Mass Function (PDMF), which is for present-day masses of stars. The PDMF is highly complicated and we shall not delve into it. As properties of stars are closely related to its mass, the IMF serves as an important tool for astronomers.

There are various different IMFs proposed. We will fit for the Salpeter (1955) IMF.

$$N(m) = Am^{-\gamma}$$

Where  $N(m)$  is the number of stars as a function of initial mass,  $m$  is initial mass,  $A$  is a constant, and  $\gamma$  is the exponentially decaying coefficient.

The PARSEC isochrones which we used in the previous section has a lot of data along with the stars magnitude and colour. It also has data for initial mass of the stars. We will take a cluster, and associate the initial mass of the closest point on the isochrone to it. Then we will plot a histogram and fit an exponentially decaying function on it.

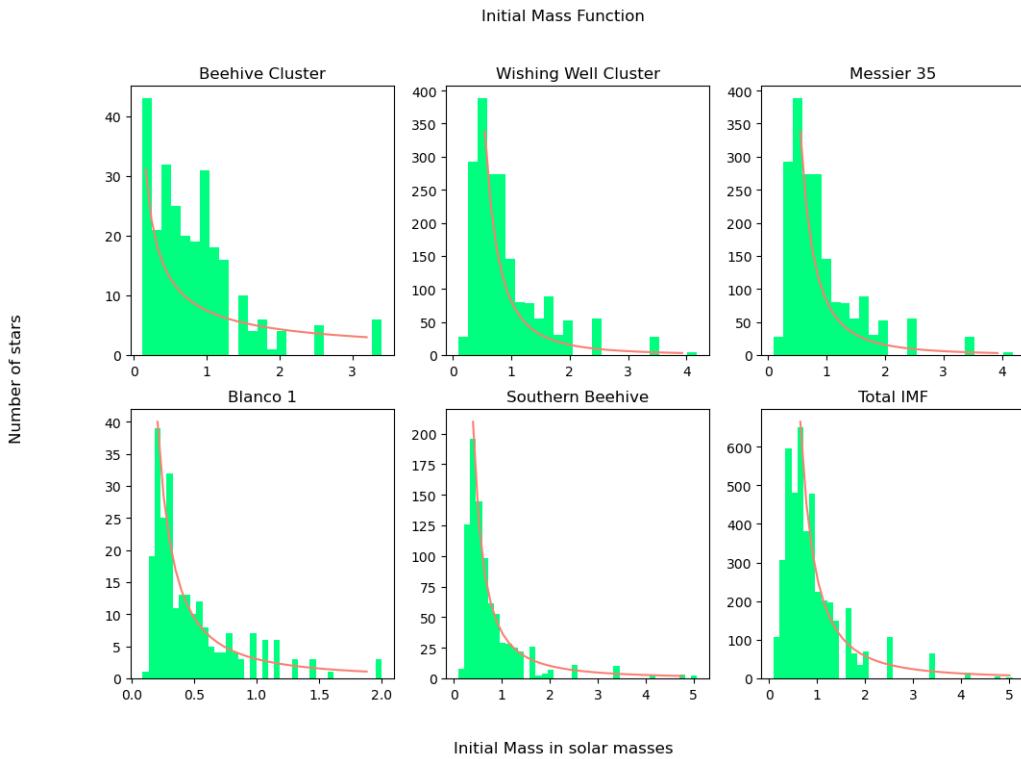


Figure 2.68: Histogram and exponentially decaying function for various open clusters, including their sum in the last plot

Since Gaia data is often not good for the dimmer, lighter stars we fit beginning from the maximum bin to the second last bin (as such stars are rare in clusters, and a bigger sample source is needed). Therefore the last bin is included in the final plot which is the sum of all the clusters used.

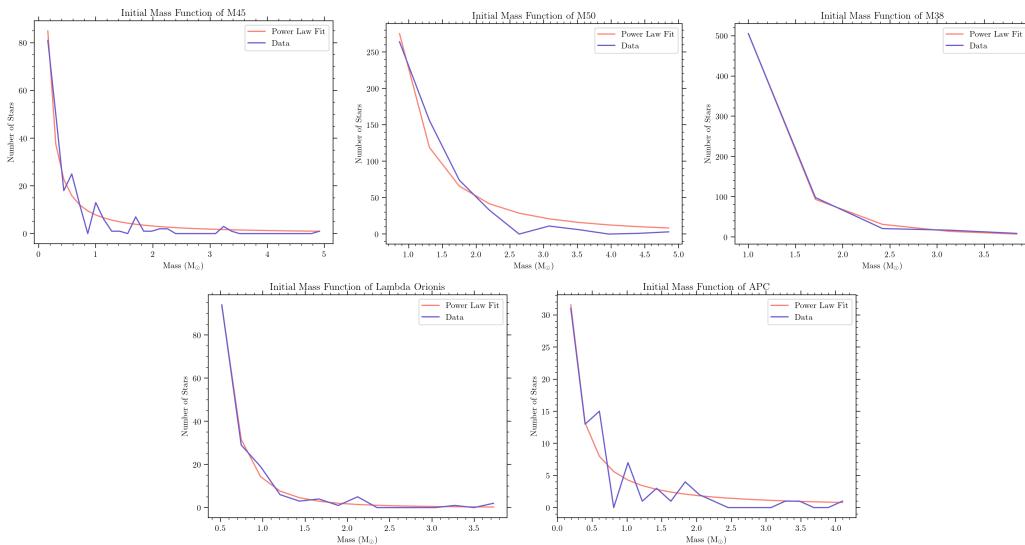
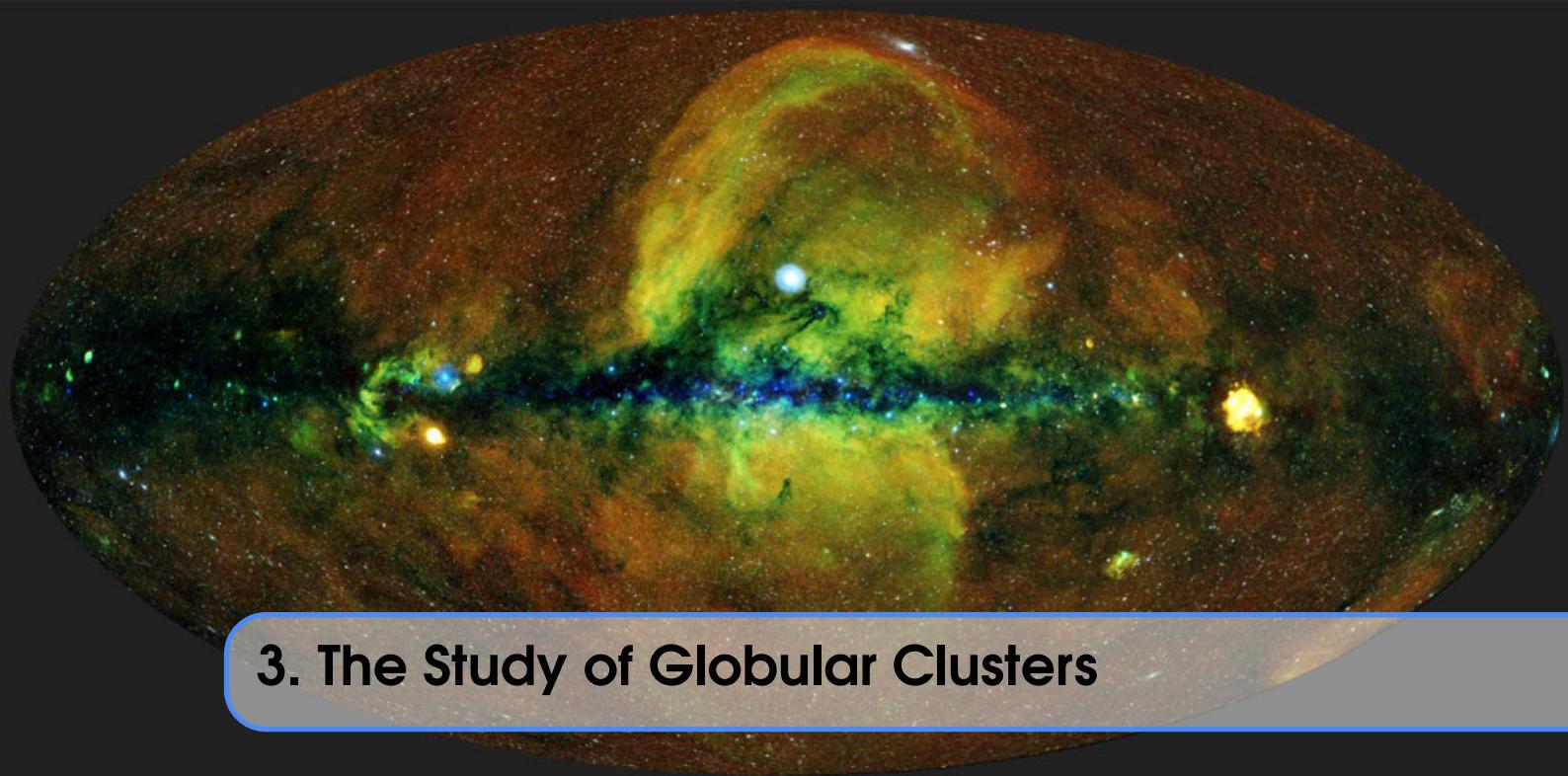


Figure 2.69: Some more Power Law Fits

The constants found were as follows:

Cluster name	Distance (in pc)	Power Law
Beehive	187	$7.45m^{-0.78}$
Wishing Well	405	$81.27m^{-2.39}$
Messier 35	912	$85.11m^{-1.76}$
Blanco 1	253	$3.02m^{-1.65}$
Southern Beehive	346	$37.76m^{-1.88}$
Total of the above	-	$260.70m^{-2.17}$
Messier 45	136.2	$3.87m^{-1.81}$
Messier 50	881	$200.35m^{-3.63}$
Messier 38	1066	$443.6m^{-2.08}$
Lambda Orionis	400	$11.46m^{-3.95}$
Alpha Persei Cluster	175	$5.58m^{-0.99}$
Messier 48	416.6	$12.05m^{-1.69}$
Trapezium Cluster	373.3	$2.68m^{-2.81}$



## 3. The Study of Globular Clusters

Globular clusters are tightly gravitationally bound star clusters which have cluster members of the order of magnitude ten thousand to million. As they are so tightly bound, they are dense, massive, and old, far more so than open clusters. Due to their age, their HR diagrams, as we shall see are very different from open clusters. The most spectacular globular cluster visible is Omega Centauri, was initially thought to be a star. The first globular cluster discovered was Messier 22, by Abraham Ihle.

### 3.1 Need for Improvised Quality Cuts

There are fewer globular clusters as compared to open clusters. As an example, in the Milky Way approximately 1100 open clusters are known, and a far larger amount is predicted to exist. Meanwhile, there are around 150 globular clusters known. So the probability of finding a globular cluster close to Earth decreases. The nearest globular cluster, M4 is itself approximately 1.68kpc away, and most are even further away.

We have already established that Gaia's data on distance worsens the farther you go. In the previous section, we had used parallax error cuts to take only those stars whose data for distances was good. This can no longer be continued. This is because the error cuts remove far too many stars to be of any analytic significance. Instead, we improvise a new way to check for the quality of data of stars.

We use the value called RUWE (Re-normalised Unit Weight Error). Instead of parallax error cuts, we use  $\text{RUWE} < 1.4$  as a standard. The lesser the value of RUWE, the better the data is.

After that, Gaussian fitting was done for ra, dec, pmra and pmdec. Stars within 3 times the standard deviation of the mean for these parameters was taken to be as part of the cluster, and the rest were removed.

The following clusters were analyzed:

1. Messier 4: 21,802 stars
2. Messier 92: 9,441 stars
3. Messier 3: 17,303 stars
4. Hercules Cluster (Messier 13): 19,467 stars
5. NGC 6541: 5,936 stars
6. Messier 68: 5,424 stars
7. Messier 79: 3,733 stars
8. Messier 107: 5,750 stars
9. Messier 5: 3,834 stars
10. Omega Centauri: 10,000 stars
11. NGC 104: 17561 stars
12. 47 Tucanae: 34,482 stars

Since parallax data is so unreliable, we cannot use absolute magnitude to plot our HR diagrams. In a cluster, all stars are nearly at the same distance from Earth. Therefore the conversion from apparent to absolute magnitude is just increasing or decreasing by a constant by this approximation.

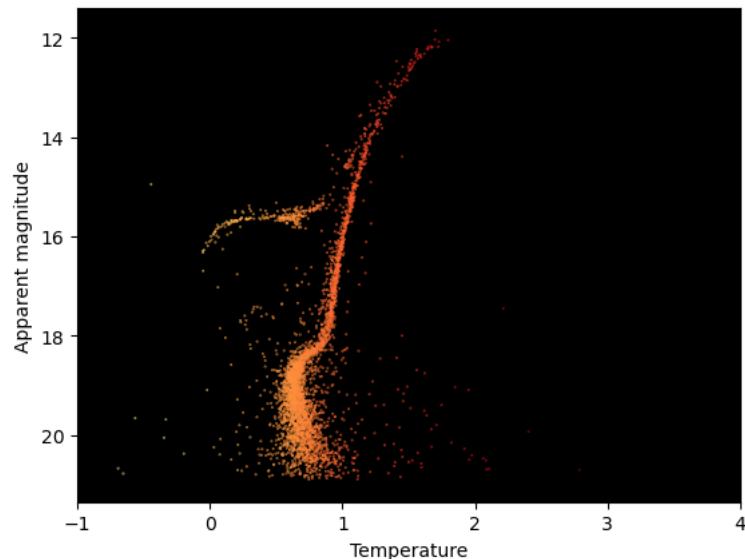


Figure 3.1: HR diagram of M5

The HR diagrams we got are shown below.

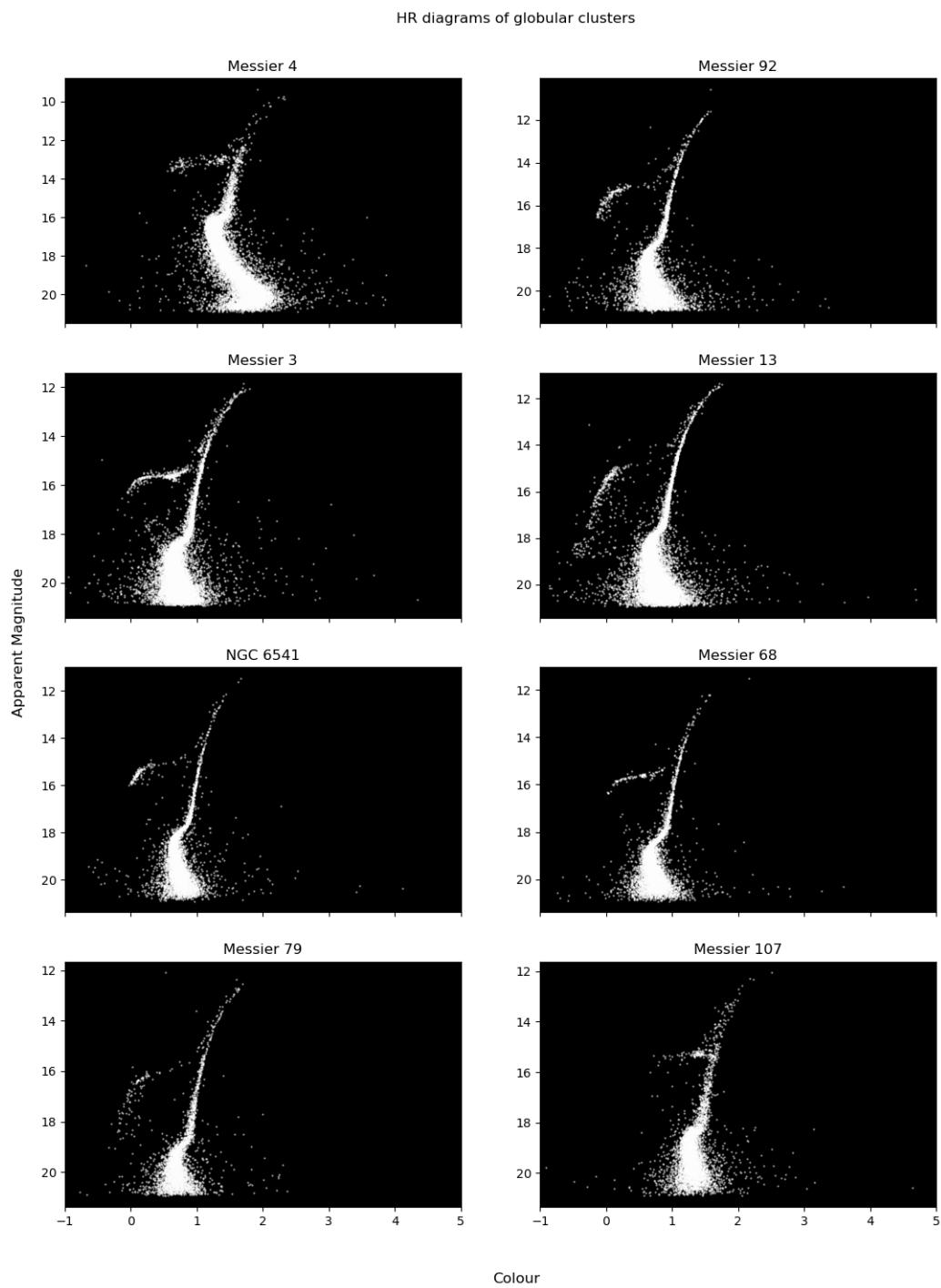


Figure 3.2: HR diagrams for various globular clusters

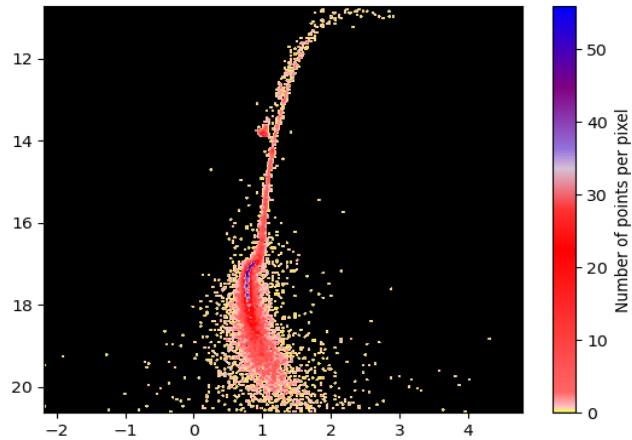


Figure 3.3: HR diagram of NGC 104

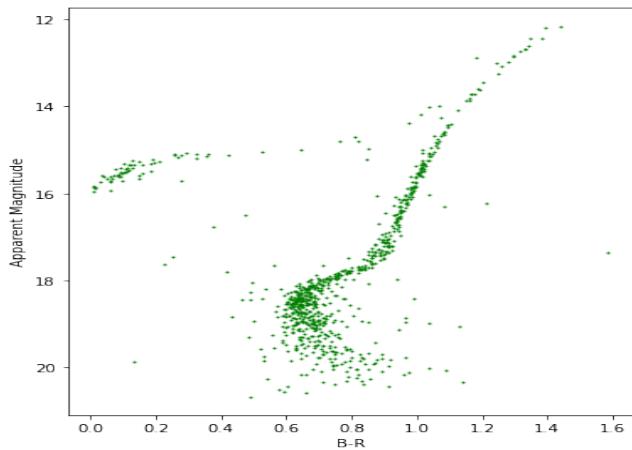


Figure 3.4: HR diagram of Messier 30

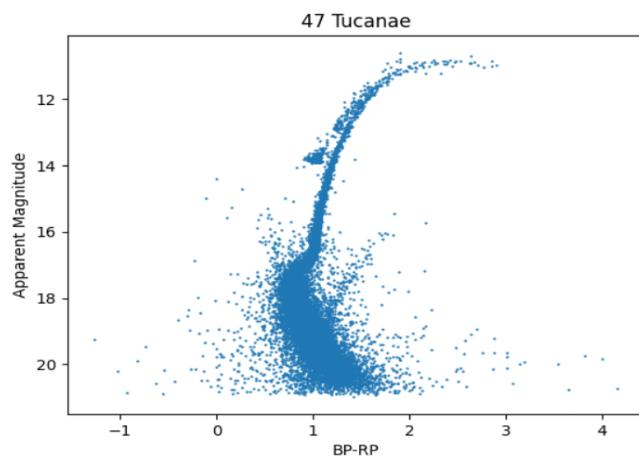


Figure 3.5: HR diagram of 47 Tucanae

One striking feature is that all these diagrams terminate abruptly at around 21, roughly. This is due to Gaia's limiting magnitude of 20.7; that is, Gaia cannot

observe stars with apparent magnitudes more than 20.7. This is why the number of stars we used should not be taken to be a good way to compare the size of a cluster; the cluster may just be too far away for its dimmer stars to be observed by Gaia.

We also observe several new parts in the diagram, which were not present in open clusters. In open clusters we had seen the main sequence and in some cases white dwarfs and red giant branch stars. We will explore these new parts in the next section.

## 3.2 Unexplored Parts of HR Diagrams - Evolved Stellar Stages

The main sequences of the HR diagrams is smaller than the open clusters. This is because a large part of the main sequence is cut off by Gaia's limiting magnitude, and as these clusters are old; so we have a well defined RGB, and we have the following new parts:

1. Well defined Red Giant Branch (RGB): Present in all of these clusters.
2. Horizontal Branch (HB): Present in all of these clusters.
3. Red clump: Easily spotted in M107, present in M3 and M4.
4. Asymptotic Giant Branch (AGB): Present in all of these clusters
5. Blue Stragglers: Present in M13 and M3.
6. RR Lyrae Variables: Present in M92, M3, M68, NGC6541 and M79.

We discussed about RGB, AGB and HB in section 1.5.2, in Red Giants.

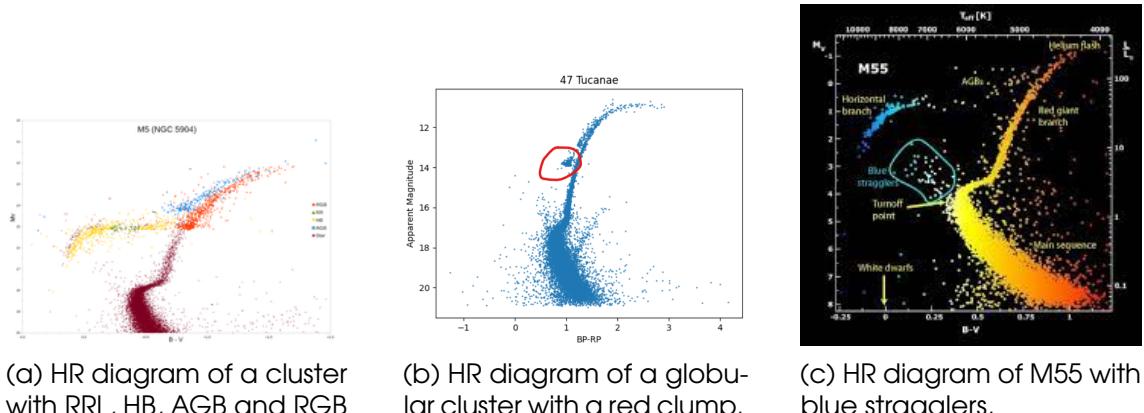
(Insert image with these points marked in the HR diagrams).

### 3.2.1 Red Clump

The red clump is a clustering in the HR diagram. They have an absolute magnitude of +0.5, and are hotter than RGB stars with same magnitude. Stars have a tendency to cluster at the cool end of the HB. This tendency is more in metal-rich clusters, and less so in metal-poor. However, this is not a strict rule as there are various other factors. In certain cases like Omega Centauri, the red clump and RGB stars from different populations may overlap.

### 3.2.2 Blue Stragglers

These are stars which lie beyond the turn-off point and the HB on the HR diagram. Several explanations for their existence exist, two of them being most likely. The first is that they are (or were) binary stars in the process of merging (or have merged). The other is mass transfer between binary stars. There is evidence for both of these hypothesis, and in M3, 47 Tucanae and NGC 6752 both of these mechanisms seem to be operating.



### 3.2.3 RR Lyrae Variables

Variables are stars with variable brightness (magnitude). They are older pulsating class A stars with low metallicity. They have a short pulsation period ranging from 0.05 to 1.2 days, with magnitude variation of 0.3 to 2. These stars occupy a small region on the HB. Their average magnitude is around +0.75.

## 3.3 Distance Determination with isochrones

In this section we will determine the distance of a cluster from the Earth. Gaia parallax data at large distance is unreliable, so we use some other methods, which are as follows.

### 3.3.1 Using RR Lyrae stars

RR Lyrae variables have an absolute magnitude of around +0.75. Assuming all stars in a cluster are nearly at the same distance from the Earth, when converting from apparent to absolute magnitudes, the  $5\log(\frac{d}{10})$  term can be considered as a constant. So the amount we shift the theoretical isochrone down is equal to  $5\log(\frac{d}{10})$ . We tried this with the following isochrones which have RR Lyrae variables.

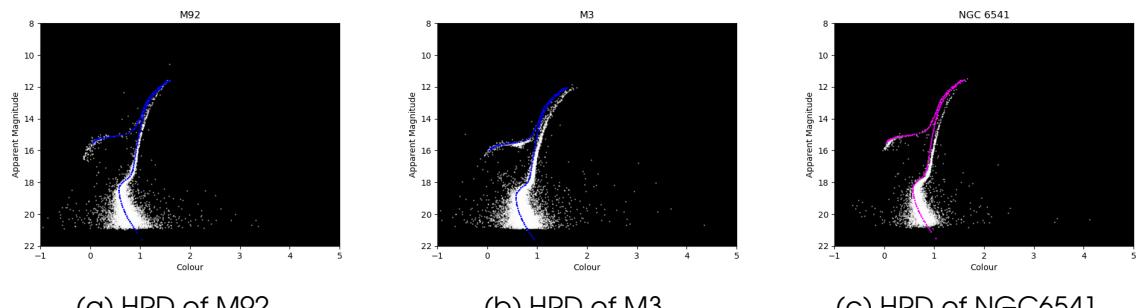


Figure 3.7: The isochrone used here has age of 12.5 Gyr and metallicity of -2.2 dex, chosen for having a well defined HB; since RRL have similar magnitudes, there is no difference across different isochrones. So the same isochrone can be used to fit only the HB.

The isochrone was shifted down by 14.80, 15.24 and 14.75 respectively.

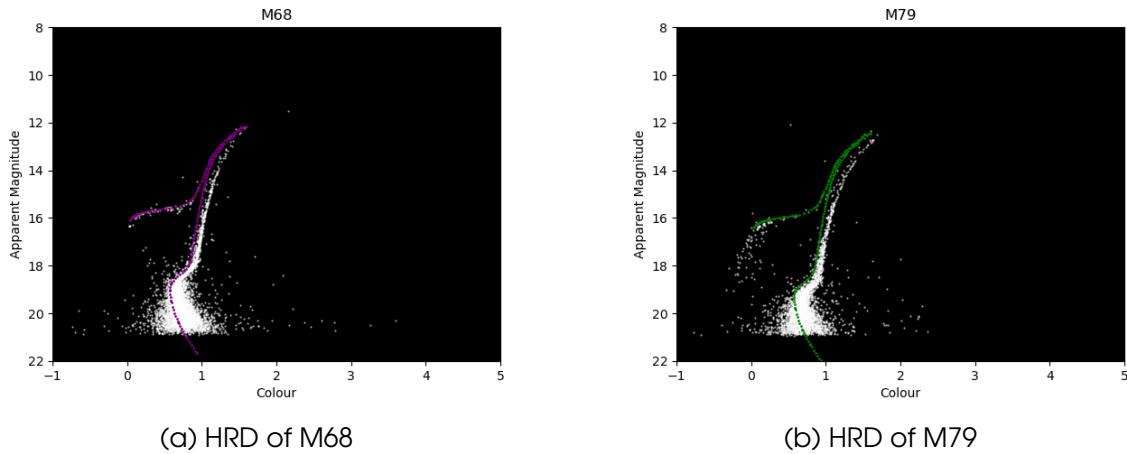


Figure 3.8: The isochrone used here is the same as last, 12.5 Gyr and -2.2 dex.

The isochrones were shifted down by 15.34 and 15.65 respectively.

Taking the amount shifted down to be  $m$ , which is the difference between apparent and absolute magnitude to  $5\log(\frac{d}{10})$  gives us the formula:

$$d = 10^{\frac{m}{5}+1} \text{ parsecs}$$

This gives the following distances for these clusters:

1. M92: 9.12 kpc
2. M3: 11.17 kpc
3. NGC 6541: 8.91 kpc
4. M68: 11.69 kpc
5. M79: 13.49 kpc

### 3.3.2 Using the Red Clump

The red clump stars can also be used in the same fashion as RR Lyrae variables, due to their almost similar absolute magnitudes of +0.5. We analyzed M4, M107 and 47 Tucanae using the red clump.

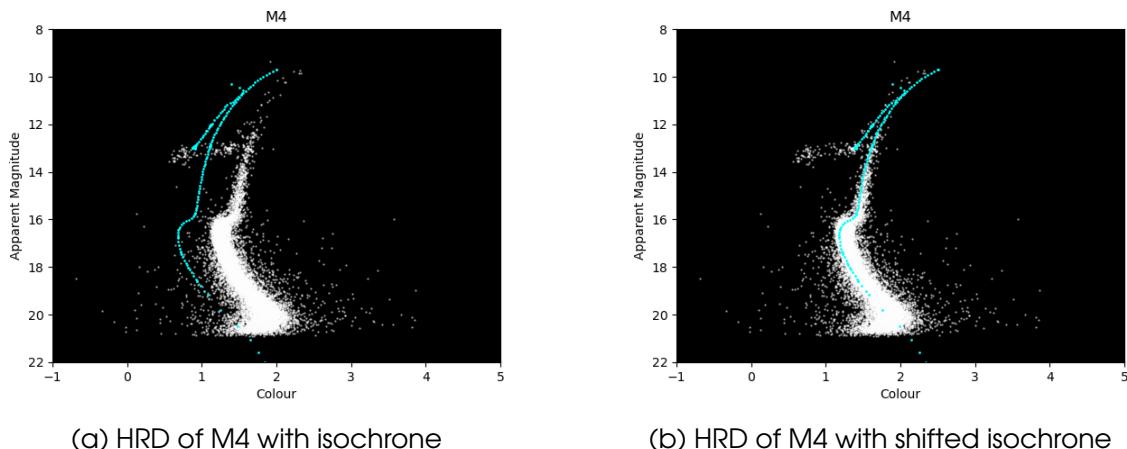


Figure 3.9: The isochrone used here has age of 12.5 Gyr and metallicity of -0.9 dex.

Our theoretical isochrone is to the left of the cluster! While isochrones with different metallicities don't have the same main sequence, the difference is not all

this drastic; this might have happened due to something called **extinction**. This happens due to the presence of dust and gas present in between the object and observer. This scatters the light and makes the object appear more red. It is also called interstellar reddening. The isochrone was shifted by +0.5 to the right and 12.60 down. Why right? Because on reddening, the star will appear brighter in the red passband, and this means that  $G_{RP}$  will decrease. The star will also appear dimmer in the blue passband, so  $G_{BP}$  increases. Therefore  $G_{BP} - G_{RP}$  increases, and the HR diagram shifts to the right.

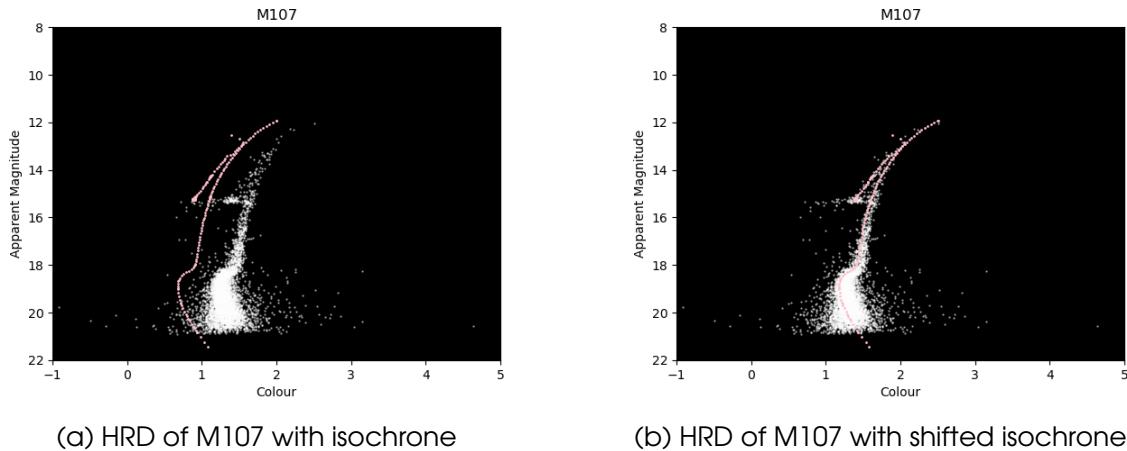


Figure 3.10: The isochrone used here is the same as last, 12.5 Gyr and -0.9 dex.

Here too we have extinction present. We shifted the isochrone to the right by +0.52 and down by 14.80.

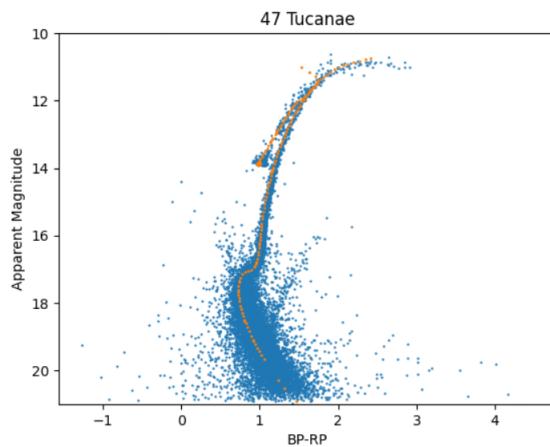


Figure 3.11: 47 Tucanae, Isochrone with 13 Gyr and -0.6 dex used

The above isochrone was shifted down by 13.4 and no shift was done in the right direction.

By using the same formula derived in 3.3.1, we find the distances as:

- M4: 3.31 kpc
- M107: 9.12 kpc
- 47 Tucanae: 4.78 kpc

M4 is one of the closest clusters, which can be seen here; it is just 3.31 kpc away! It is also why we see its main sequence in more detail as compared to other clusters.

### 3.3.3 Using the turn-off point

In the case of M13, we did not have either RR Lyrae variables or red clumps. So we directly found the isochrone which best fits this cluster, and on the basis of the amount we shift it down we used the same formula derived in section 3.3.1.

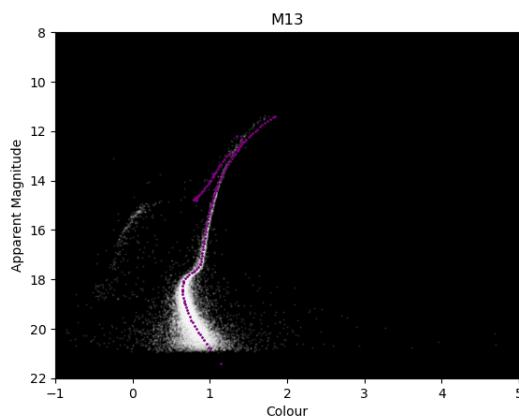


Figure 3.12: HRD of M13. The transparency has been increased to show that the isochrone fits well. The isochrone is aged 12.5 Gyr and has a metallicity of -1.2 dex.

The isochrone was shifted down by 14.40, which means that M13 is 7.59 kpc away.

An interesting point to note was that the distances calculated by us here were off by around 1kpc, in some rare cases around 1.5 or 2 kpc.

## 3.4 Inferring Properties from Isochrones

Along with distance, we can also put a rough estimate on the age and metallicity of a cluster from isochrones. We used PARSEC isochrones, with ages 10 to 13.5 in steps of 0.5 billion years, and metallicities of 0, -0.3, -0.9, -1.2, -1.5, -1.6, -1.9 and -2.2 dex, in total of 72 isochrones and visually found the best fit isochrone, and assigned its age and metallicity to the cluster. As we shall see, in some cases the visual fitting fails to distinguish on which isochrone is the better one. In some cases, zooming at the turnoff point helps, but not always.

The above procedure was done by first comparing isochrones of different metallicities for the same age as shown below:

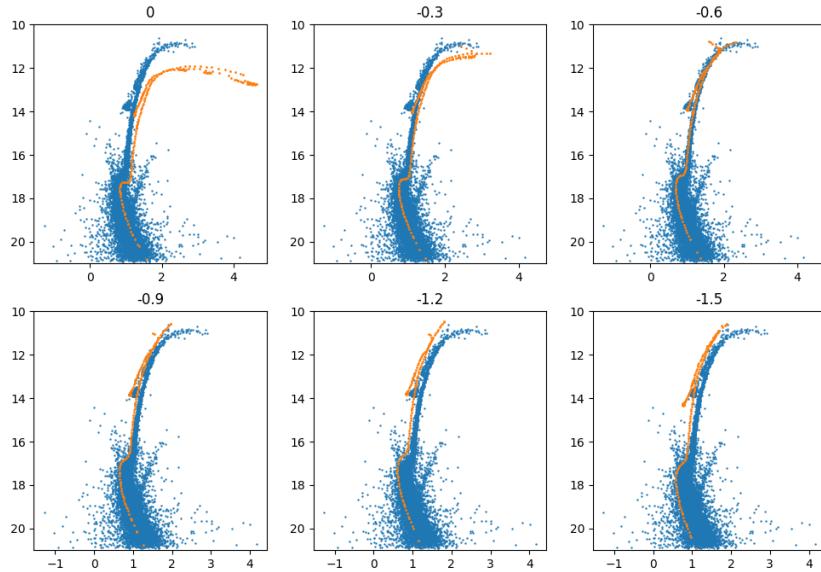


Figure 3.13: For the same age Isochrones of different metallicities are compared.

It is found that for different age groups, ideally the same metallicity gives the best fit. Then for the best metallicity we compare different ages:

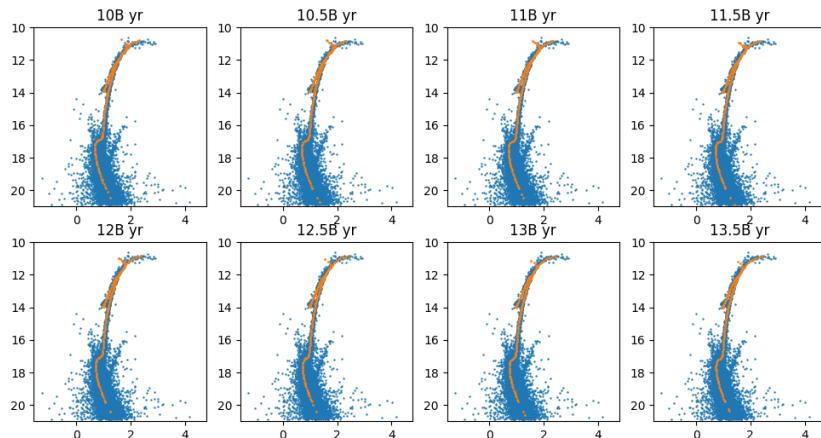


Figure 3.14: For the same metallicity, Isochrones of different ages are compared.

This way we found the best approximation for age and metallicity.

The transparency of all the points in the HR diagram has been increased to show the quality of the fit.

### 3.4.1 Messier 3

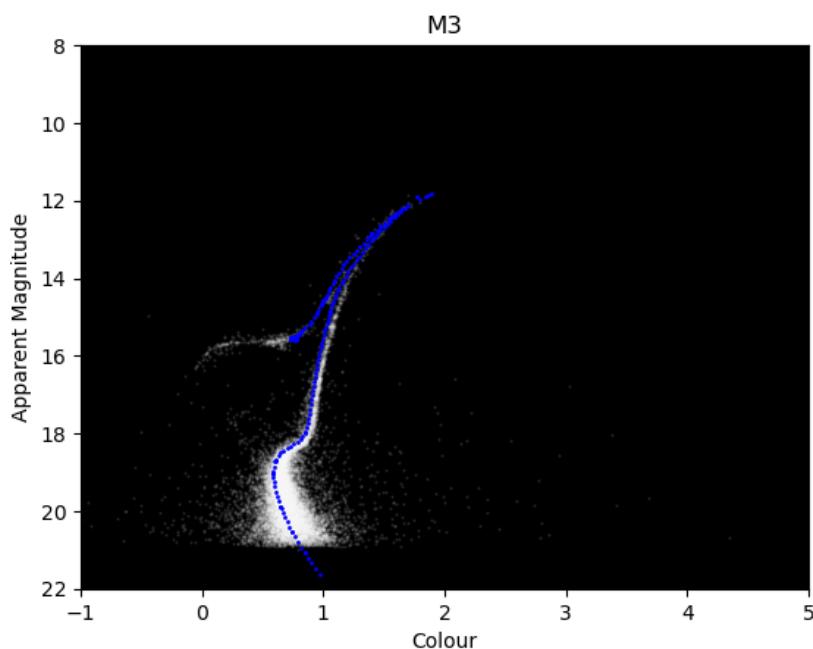


Figure 3.15: HRD of M3. The isochrone is aged 11 Gyr and has a metallicity of -1.5 dex.

### 3.4.2 Messier 4

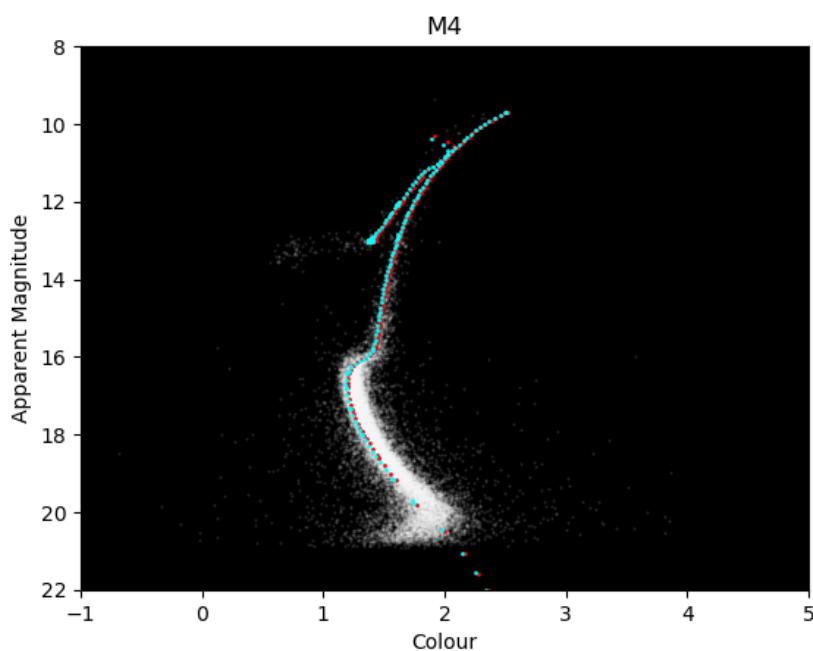


Figure 3.16: HRD of M4. The red isochrone is aged 12.5 Gyr and cyan isochrone is aged 13 Gyr. Both isochrones have metallicity of -0.9 dex.

### 3.4.3 Messier 5

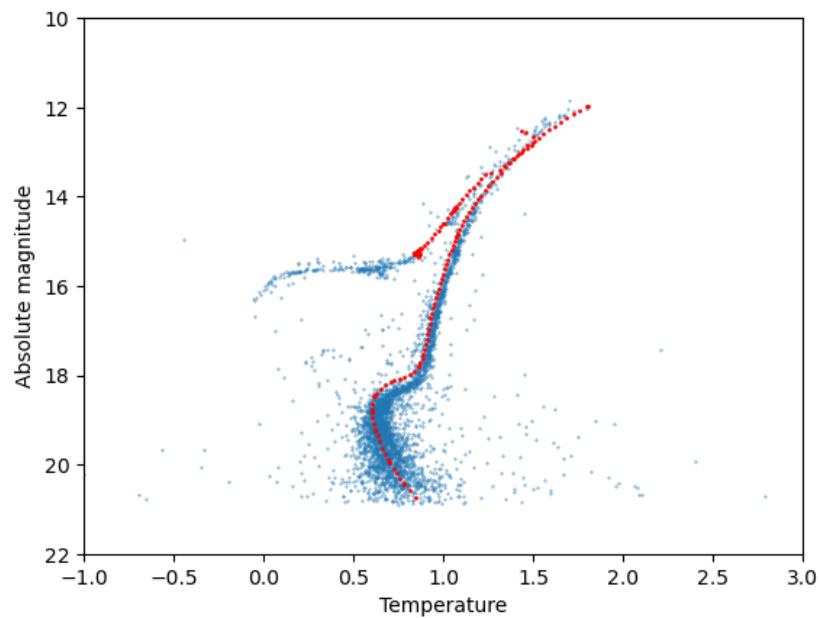


Figure 3.17: HRD of M5. The red isochrone is aged 13 Gyr and cyan isochrone is aged 11 Gyr. Both isochrones have metallicity of -1.2 dex.

### 3.4.4 Messier 13

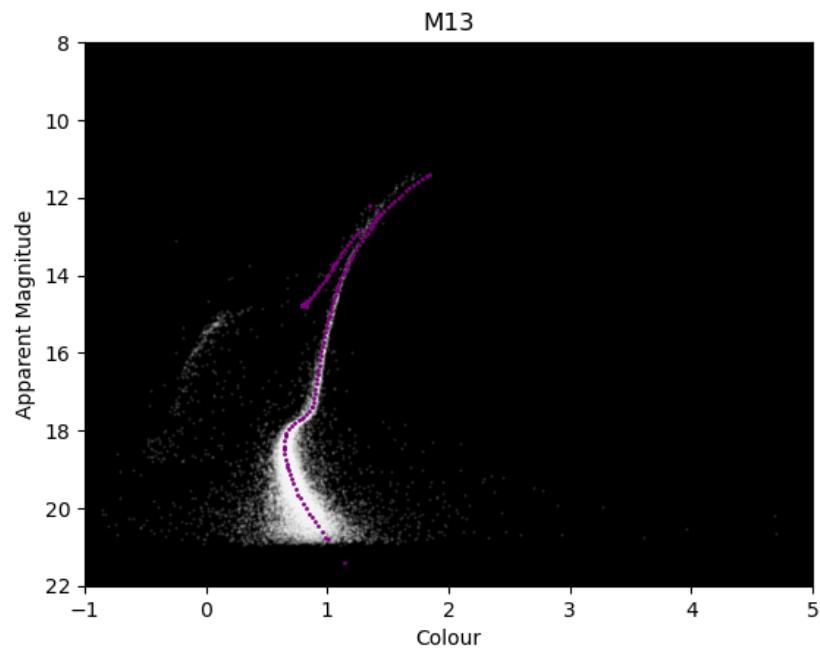


Figure 3.18: HRD of M13. The isochrone is aged 12.5 Gyr and has a metallicity of -1.2 dex.

### 3.4.5 Messier 68

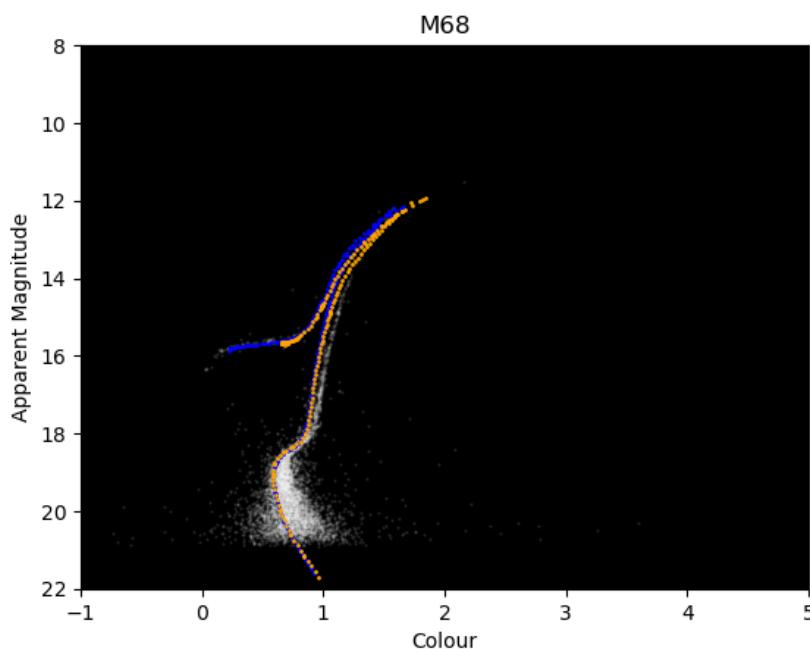


Figure 3.19: HRD of M68. The orange isochrone is aged 11 Gyr and has a metallicity of -1.6 dex, and the blue isochrone is 12.5 Hyr old and has a metallicity of -1.9 dex.

### 3.4.6 Messier 79

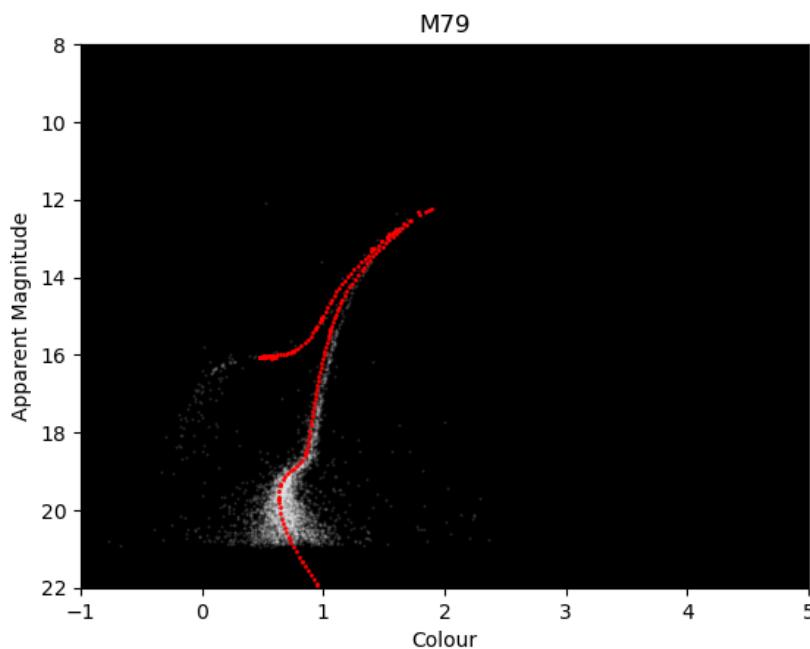


Figure 3.20: HRD of M79. The isochrone is aged 13.5 Gyr and has a metallicity of -1.5 dex.

### 3.4.7 Messier 92

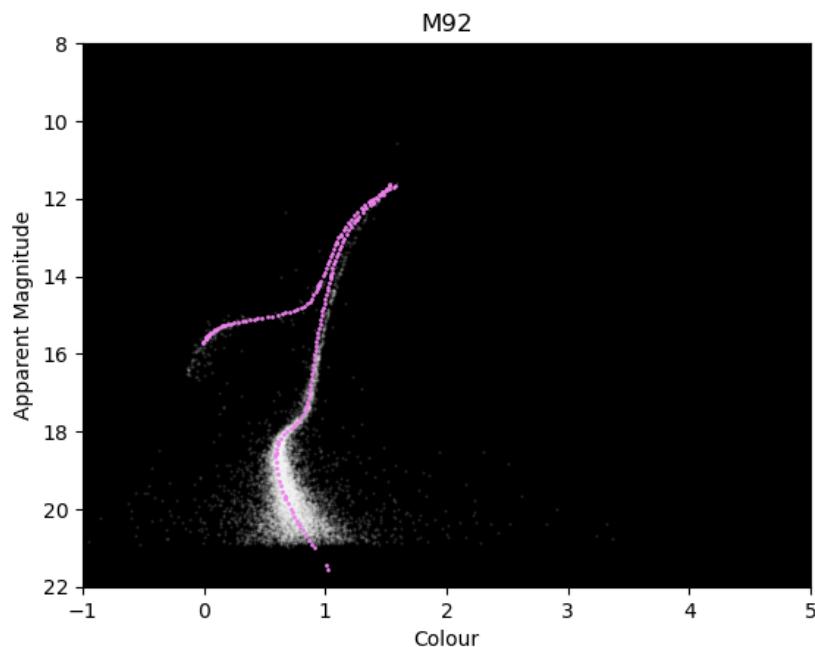


Figure 3.21: HRD of M92. The isochrone is aged 13.5 Gyr and has a metallicity of -2.2 dex.

### 3.4.8 Messier 107

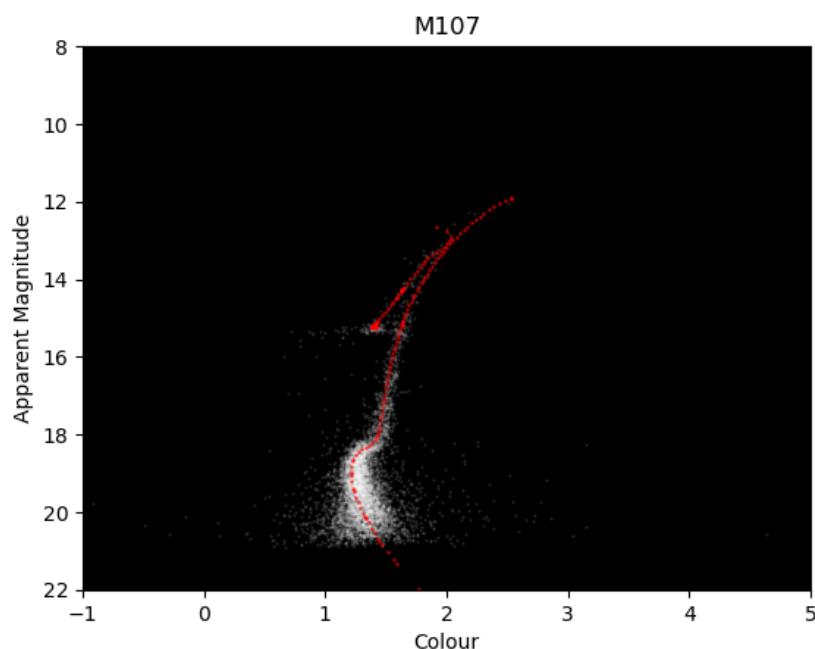


Figure 3.22: HRD of M3. The isochrone is aged 13.5 Gyr and has a metallicity of -0.9 dex.

### 3.4.9 NGC 6541

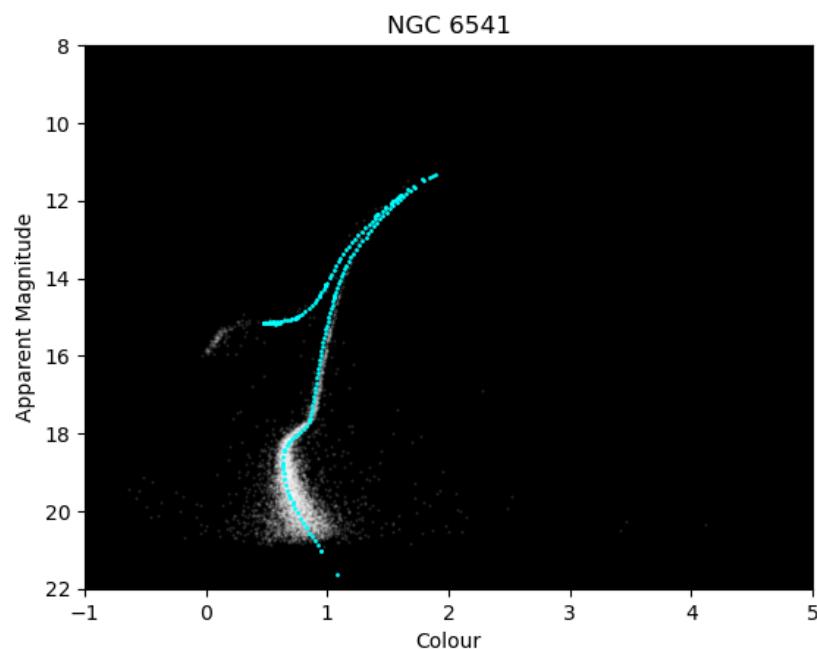


Figure 3.23: HRD of NGC 6541. The isochrone is aged 13.5 Gyr and has a metallicity of -1.5 dex.

### 3.4.10 NGC 104

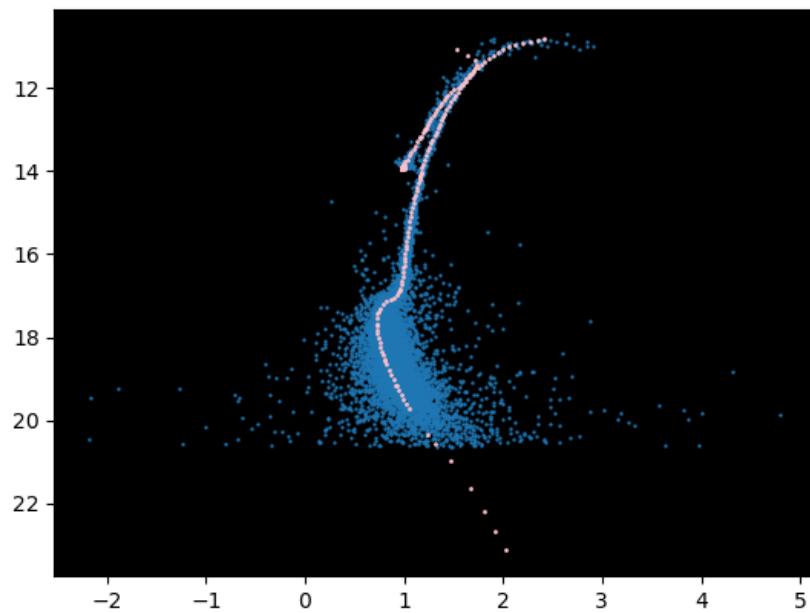


Figure 3.24: HRD of NGC 104. The isochrone is aged 13 Gyr and has a metallicity of -0.6 dex.

### 3.4.11 Omega Centauri (NGC 5139)

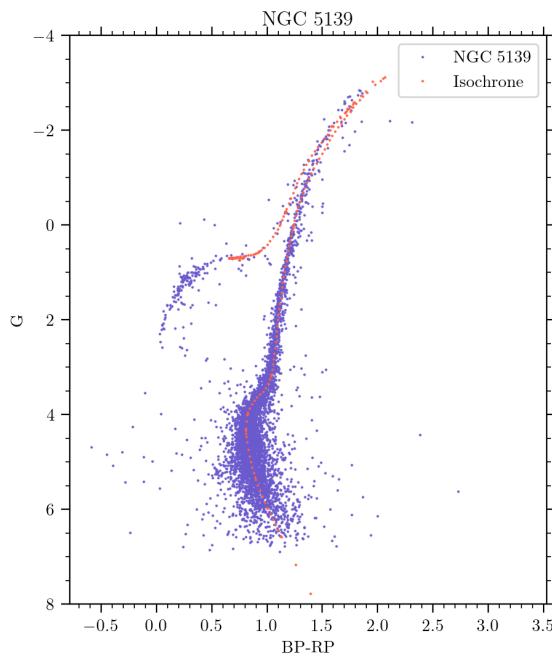


Figure 3.25: HR Diagram of Omega Centauri. The isochrone is aged 13.5Gyr and has a metallicity of -1.5 dex

### 3.4.12 Messier 30

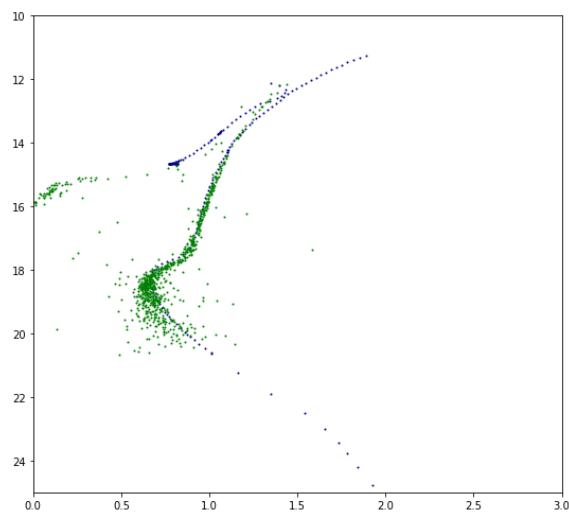


Figure 3.26: HR Diagram of Messier 30. The isochrone is aged 13.5Gyr and has a metallicity of -1.2 dex

### 3.4.13 47 Tucanae

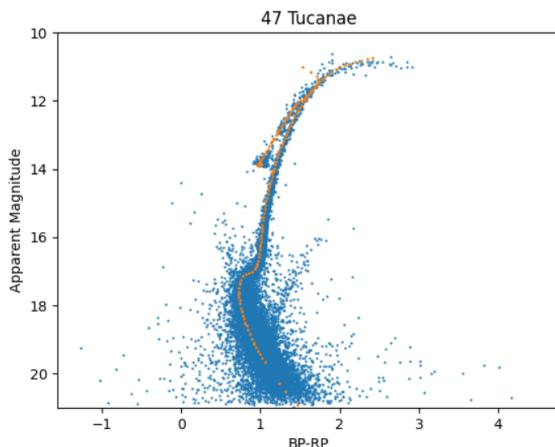


Figure 3.27: HR Diagram of 47 Tucanae. The isochrone is aged 13Gyr and has a metallicity of -0.6 dex

To summarise:

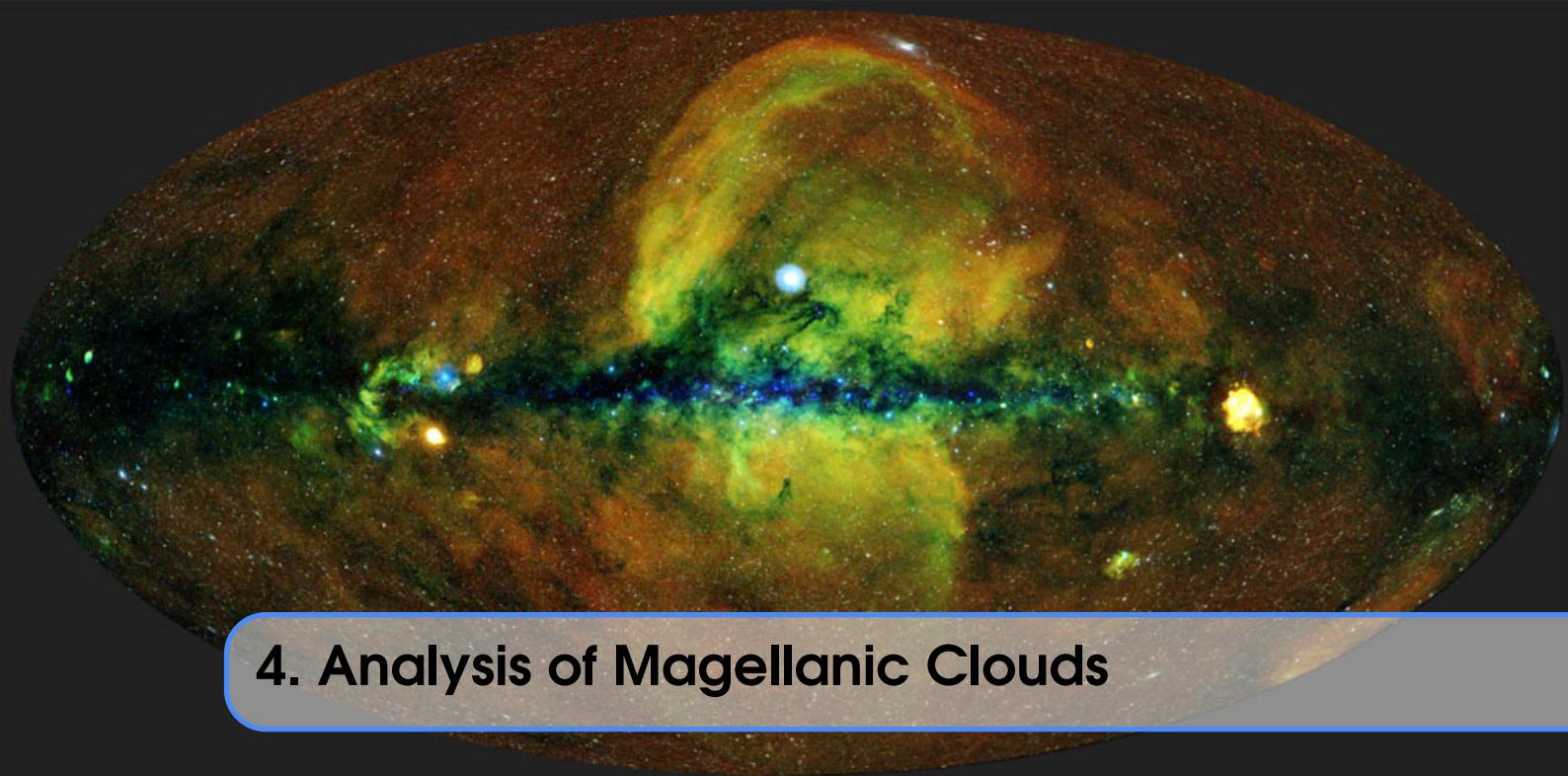
1. M3: 11 billion years, -1.5 dex
2. M4: 12.5/13 billion years, -0.9 dex
3. M13: 12.5 billion years, -1.2 dex
4. M68: 11/12.5 billion years, -1.6/-1.9 dex
5. M79: 13.5 billion years, -1.5 dex
6. M92: 13.5 billion years, -2.2 dex
7. M107: 13.5 billion years, -0.9 dex
8. NGC 6541: 13.5 billion years, -1.5 dex
9. Omega Centauri: 13.5 Billion years, -1.5 dex
10. 47 Tucanae: 13 Billion years, -0.6 dex

This establishes that globular clusters are extremely old. M4, M107 and 47 Tucanae are metal rich, while M92 is one of the most metal poor globular cluster

### 3.5 Summary

To summarise:

Cluster name	Distance (in kpc)	Age (in Gyr)	Metallicity (in dex)	sideways shift (in mag)
M3	11.17	11	-1.5	0
M4	3.31	12.5/13	-0.9	+0.50
M5	20.08	11.5	-1.2	0.0
M13	7.59	12.5	-1.2	0
M68	11.69	11 12.5	-1.6 -1.9	0 0
M79	13.49	13.5	-1.5	0
M92	9.12	13.5	-2.2	0
M107	9.12	13.5	-0.9	+0.52
NGC 6541	8.91	13.5	-1.5	0
NGC 5139	5.24	13.5	-1.5	+0.18
NGC 104	4.45	13	-0.6	0
M30	7.8	13.5	-1.2	0.11
47 Tucanae	4.78	13	-0.6	0



## 4. Analysis of Magellanic Clouds

The Magellanic Clouds are two dwarf galaxies orbiting the Milky Way. They do not have a well-defined shape and thus are irregular galaxies. They are called Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC), based on their sizes. The LMC and SMC have a diameter of 14,000 and 7,000 light-years respectively. To put things into perspective, the Milky Way is about 100,000 light-years across. The large Magellanic cloud is about 50 kpc away, while the SMC is about 63kpc away.

### 4.1 Large Magellanic Cloud

The LMC has RA =  $3^h11^m34^s$  and Dec =  $-73^\circ30'24''$ . On general LMC stars have proper motion less than 5 mas per year. We query for the Large Magellanic Cloud as follows:

To begin we do a query as follows:

```
'''SELECT g.ra, g.dec, g.parallax, g.phot_g_mean_mag, g.phot_bp_mean_mag,
g.phot_rp_mean_mag, g.pmra, g.pmdec, g.pm,
DISTANCE(POINT(80.8942, -69.756),
POINT(g.ra, g.dec)) AS ang_sep
FROM gaiadr3.gaia_source AS g
WHERE 1 = CONTAINS(POINT(80.8942, -69.756), CIRCLE(ra, dec, 5))
AND g.pmra is not null
AND g.pmdec is not null
AND g.parallax BETWEEN 0 and 0.1
AND g.pm < 5
AND g.ruwe < 1.4'''
```

Now we filter the stars by applying cuts on the phase distribution of the stars. Iteratively sigma clipping the stars based on proper motion in RA and Dec.

We get the following plot:

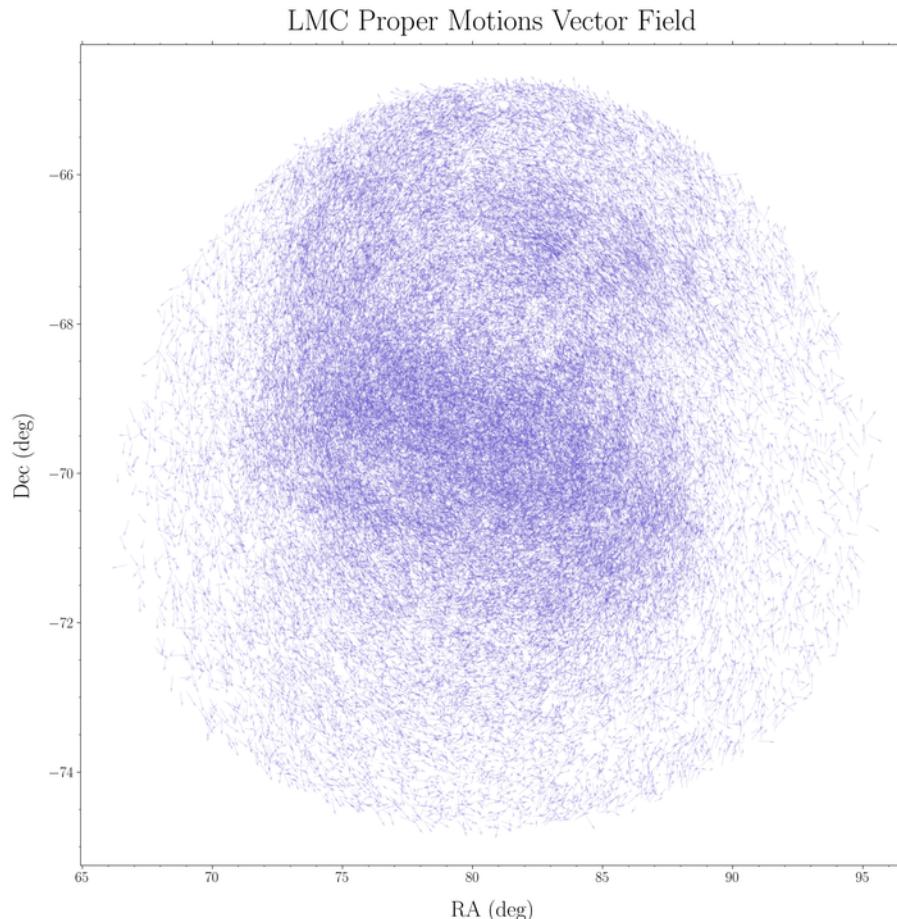


Figure 4.1: Quiver plot of proper motion directions

We can see how the stars seem to revolve around the core of the Large Magellanic Cloud.

Now, we will produce a larger dataset for in-depth analysis of both LMC and SMC.

```
'''select source_id, ra, dec, phot_g_mean_mag, pmra, pmdec, pm, bp_rp, ruwe,
       parallax, parallax_error
  from gaiadr3.gaia_source
 where ra between 65 and 100
   and dec between -76 and -60
   and pmra is not null
   and pmdec is not null
   and bp_rp is not null
   and pm < 5
   and ruwe < 1.4
   and phot_g_mean_mag is not null
 order by parallax_over_error desc'''
```

This query gives us 15,540,606 stars. The scatter plot of these stars is:

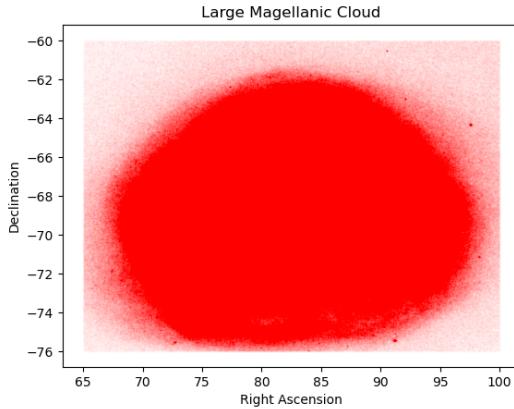


Figure 4.2: RA/Dec scatter plot with size 1 and transparency 0.002

We need some more filtering. So we do 3 sigma filtering on parallax. However, parallax data is not reliable, so we do the same on parallax + error and parallax - error, and then take all stars which lie in any one of them.

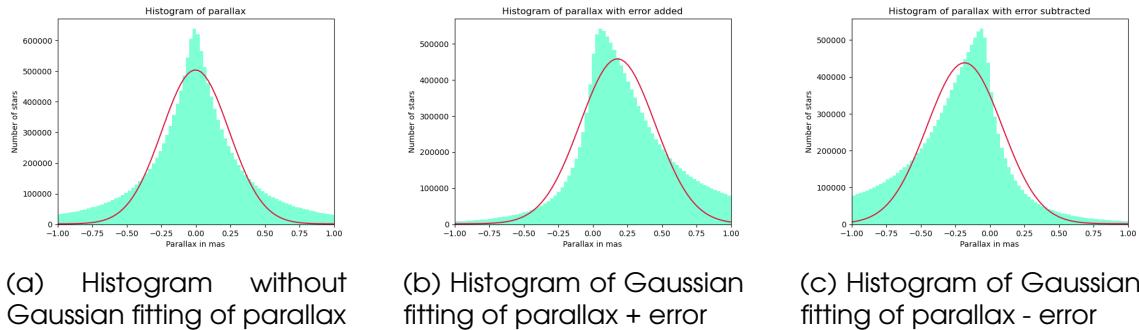


Figure 4.3: We do parallax cuts as mentioned above based on these fits, which gives 13,874,052 stars.

To see the difference, we plot the filtered stars on the unfiltered ones.

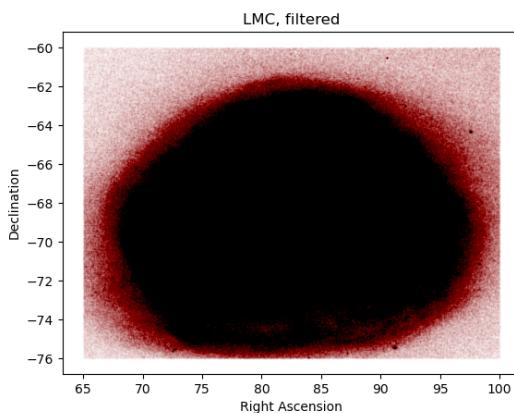


Figure 4.4: RA/Dec scatter plot with size 1 and transparency 0.002. In red is the earlier plot, and in black in the filtered one.

We then divide these stars on the basis of their evolutionary phase. This was done by making polygons on the HR diagram and then taking those stars inside the area of the polygon to be in that phase.

Before that, however we encounter a new region in HR diagrams; Blue Loop (BL). This is a stage where a star changes from a cool star to hotter one and then cooler again. It is called so because it forms a loop towards the blue (the hotter) side of the HR diagram. Some stars undergo this phase multiple times.

The vertices of polygons used were as follows:

- Young 1: (0.18, 16.0), (-0.3, 10.0), (-1.0, 10.0), (-1.0, 16.0), (0.18, 16.0)
- Young 2: (-1.0, 16.0), (0.18, 16.0), (0.34, 18.0), (-1.0, 18.0), (-1.0, 16.0)
- Young 3: (-0.40, 20.5), (-0.6, 19.0), (-0.6, 18.0), (0.34, 18.0), (0.40, 18.9), (0.45, 19.5), (0.70, 20.5), (-0.40, 20.5)
- RGB: (0.80, 20.5), (0.90, 19.5), (1.60, 19.8), (1.60, 19.0), (1.05, 18.41), (1.30, 16.56), (1.60, 15.3), (2.40, 15.97), (1.95, 17.75), (1.85, 19.0), (2.00, 20.5), (0.80, 20.5)
- AGB: (1.6, 15.3), (1.92, 13.9), (3.5, 15.0), (3.5, 16.9), (1.6, 15.3)
- RRL: (0.45, 19.5), (0.40, 18.9), (0.90, 18.9), (0.90, 19.5), (0.45, 19.5)
- BL: (0.90, 18.25), (0.1, 15.00), (-0.30, 10.0), (2.85, 10.0), (1.30, 16.56), (1.05, 18.41), (0.90, 18.25)
- RC: (0.90, 19.5), (0.90, 18.25), (1.60, 19.0), (1.60, 19.8), (0.90, 19.5)

There are parts which are left out of this on purpose; this is because these areas are too mixed, affected by blended stars or contaminated with foreground stars. These areas are exclusive and do not overlap. The HR diagram is as follows:

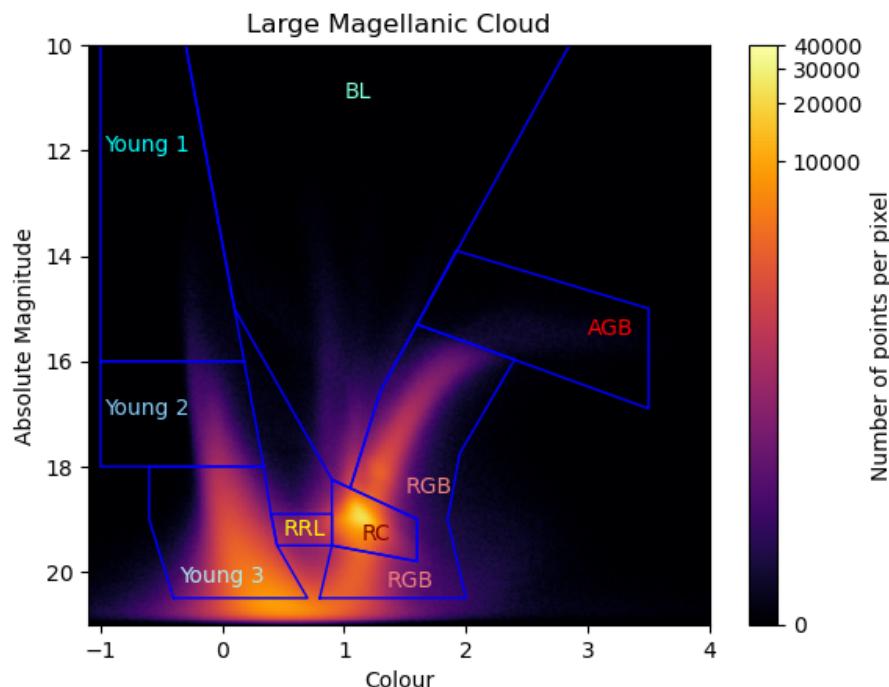


Figure 4.5: HR diagram of LMC. The colourmap has been logarithmically scaled.

On the basis of this segregation we make the scatter plots of all the different classes.

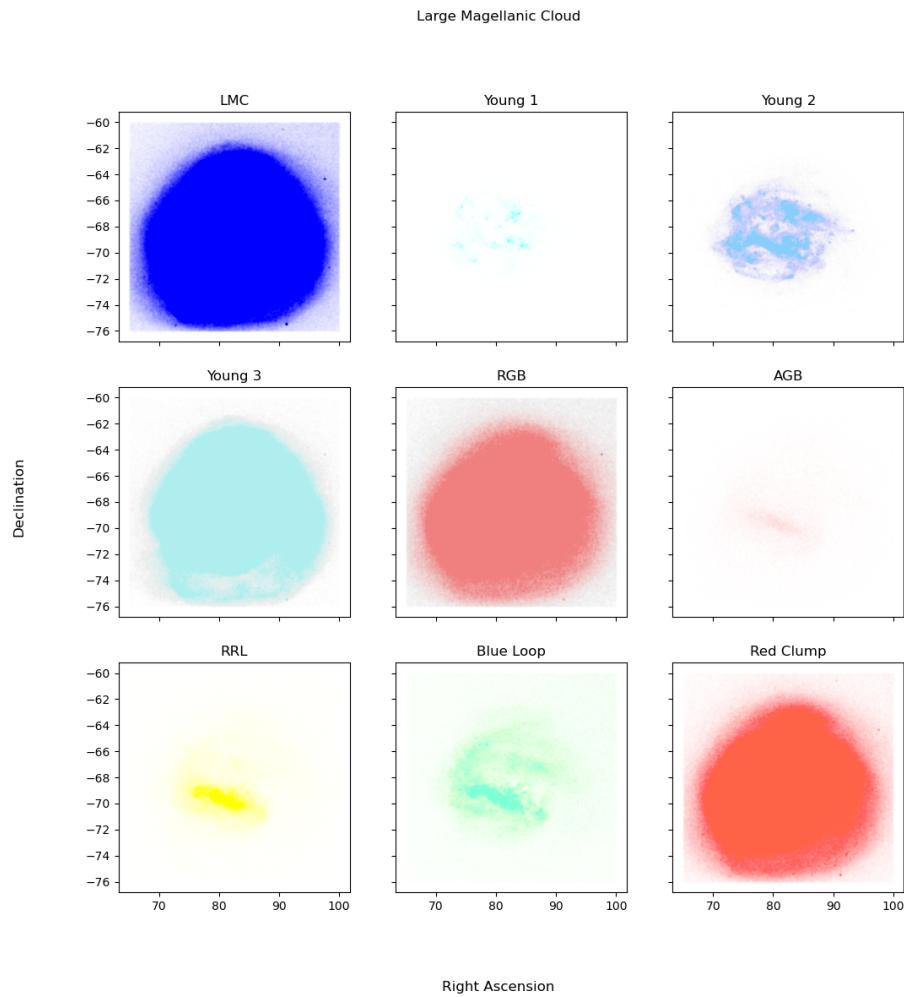


Figure 4.6: The scatter plots of various evolutionary phases of LMC.

Now, since more filtering was done on the basis of HR diagrams, we plot the filtered on unfiltered one again to see the difference. This leaves us with 11,098,117 stars in the LMC.

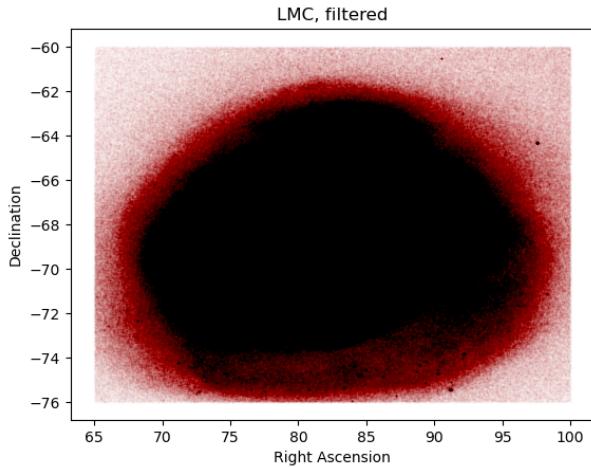


Figure 4.7: RA/Dec scatter plot with size 1 and transparency 0.002. In red is the earlier plot, and in black in the filtered one.

The final plot of LMC is this

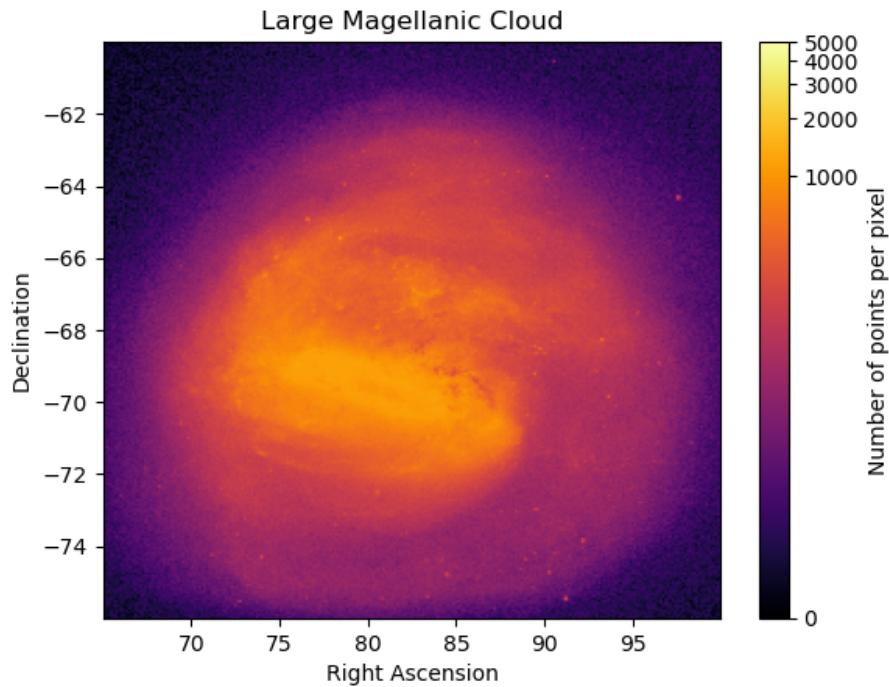


Figure 4.8: RA/Dec scatter plot with a logarithmically scaled colourmap.

## 4.2 Small Magellanic Cloud

The SMC has  $\text{RA} = 00^{\circ}52^m44.8^s$  and  $\text{Dec} = -72^{\circ}49'43''$ . On general SMC stars also have proper motion less than 5 mas per year. We query for the Small Magellanic Cloud as follows:

```
'''select source_id, ra, dec, phot_g_mean_mag, pmra, pmdec, pm, bp_rp,
ruwe, parallax, parallax_error
```

```
from gaiadr3.gaia_source
where
ra between 0 and 30 and
dec between -78 and -66
and pmra is not null
and pmdec is not null
and bp_rp is not null
and ruwe < 1.4
and pm < 5
and phot_g_mean_mag is not null
order by ruwe asc'''
```

This query gives us 2,705,844 stars. This gives us the following ra/dec scatter plot:

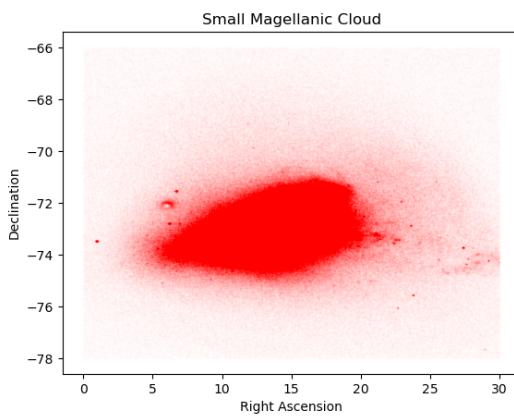


Figure 4.9: RA/Dec scatter plot with size 1 and transparency 0.005

We filter as we did with LMC.

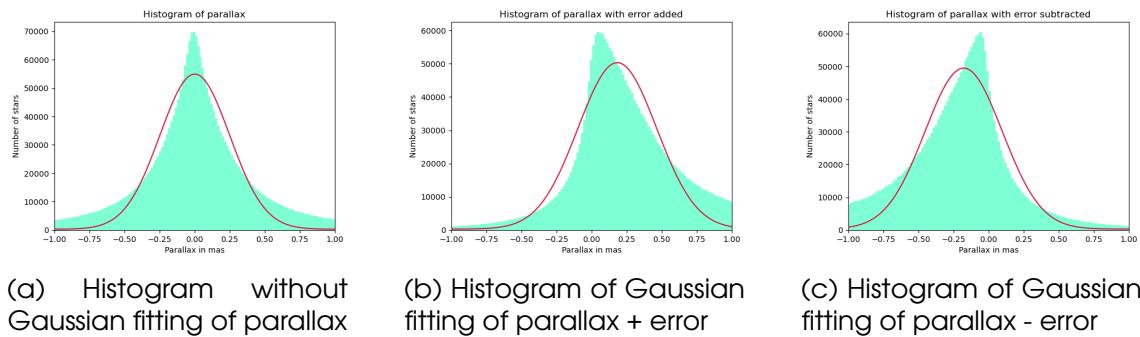


Figure 4.10: We do parallax cuts as mentioned above based on these fits, which gives 2,408,372 stars.

To see the difference, we plot the filtered stars on the unfiltered ones.

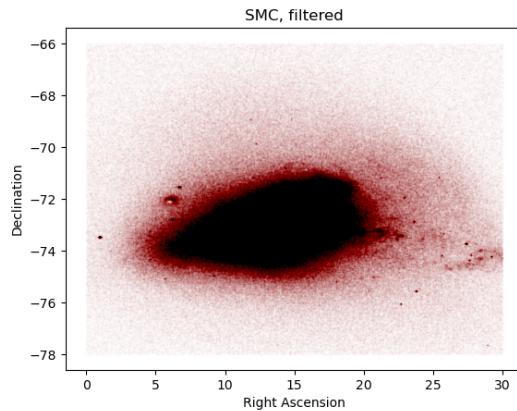


Figure 4.11: RA/Dec scatter plot with size 1 and transparency 0.005. In red is the earlier plot, and in black in the filtered one.

Like LMC, we divide the diagram into polygons again as follows:

- Young 1: (-1.00, 16.50), (-1.00, 10.00), (-0.30, 10.00), (-0.15, 15.25), (0.00, 16.50), (-1.00, 16.50)
- Young 2: (-1.00, 18.50), (-1.00, 16.50), (0.00, 16.50), (0.24, 18.50), (-1.00, 18.50)
- Young 3: (-0.50, 20.50), (-0.65, 20.00), (-0.65, 18.50), (0.24, 18.50), (0.312, 19.10), (0.312, 20.00), (0.50, 20.50), (-0.50, 20.50)
- RGB: (0.65, 20.50), (0.80, 20.00), (0.80, 19.50), (1.60, 19.80), (1.60, 19.60), (1.00, 18.50), (1.50, 15.843), (2.00, 16.00), (1.60, 18.50), (1.60, 20.50), (0.65, 20.50)
- AGB: (1.50, 15.843), (1.75, 14.516), (3.50, 15.00), (3.50, 16.471), (1.50, 15.843)
- RRL: (0.312, 20.00), (0.312, 19.10), (0.80, 19.10), (0.80, 20.00), (0.312, 20.00)
- BL: (0.40, 18.15), (-0.15, 15.25), (-0.3, 10.00), (2.60, 10.00), (1.00, 18.50), (0.80, 18.50), (0.40, 18.15) RC: (0.80, 19.50), (0.80, 18.50), (1.00, 18.50), (1.60, 19.60), (1.60, 19.80), (0.80, 19.50)

The HR diagram is as follows:

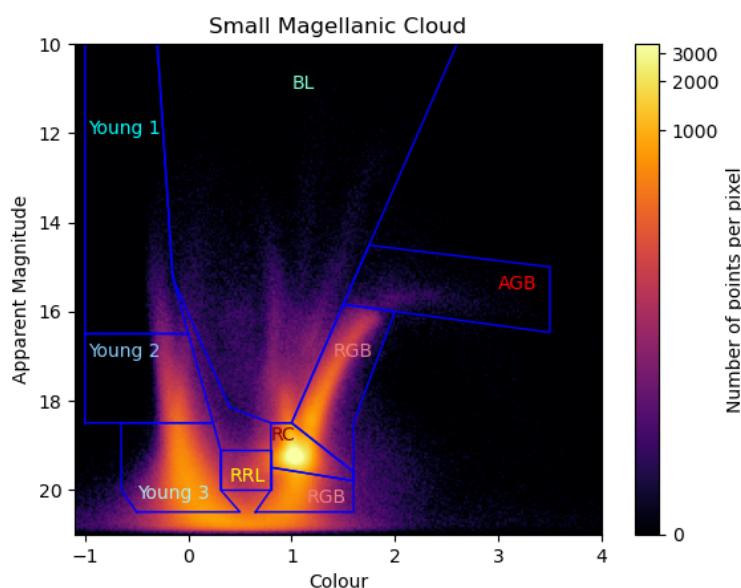


Figure 4.12: HR diagram of SMC. The colourmap has been logarithmically scaled.

On the basis of this segregation we make the scatter plots of all the different classes.

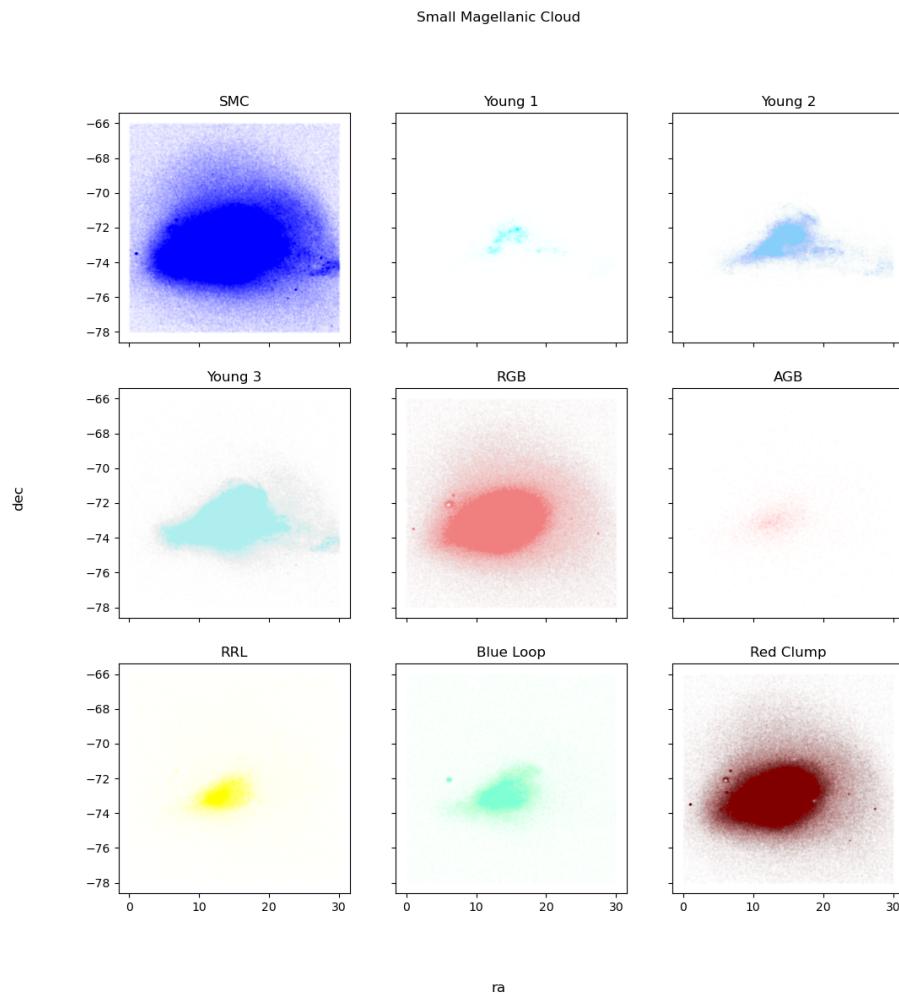


Figure 4.13: The scatter plots of various evolutionary phases of SMC.

Since there was filtering done on the basis of HR diagram too, we plot the filtered on unfiltered scatter again. This leaves us with 2,137,092 stars in the SMC.

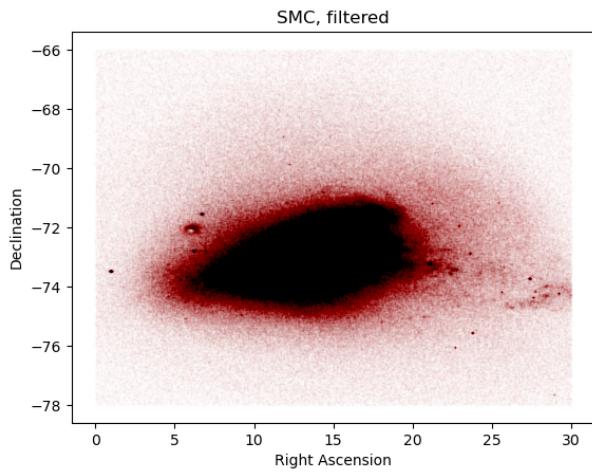


Figure 4.14: RA/Dec scatter plot with size 1 and transparency 0.005. In red is the earlier plot, and in black in the filtered one.

The final plot of SMC is this

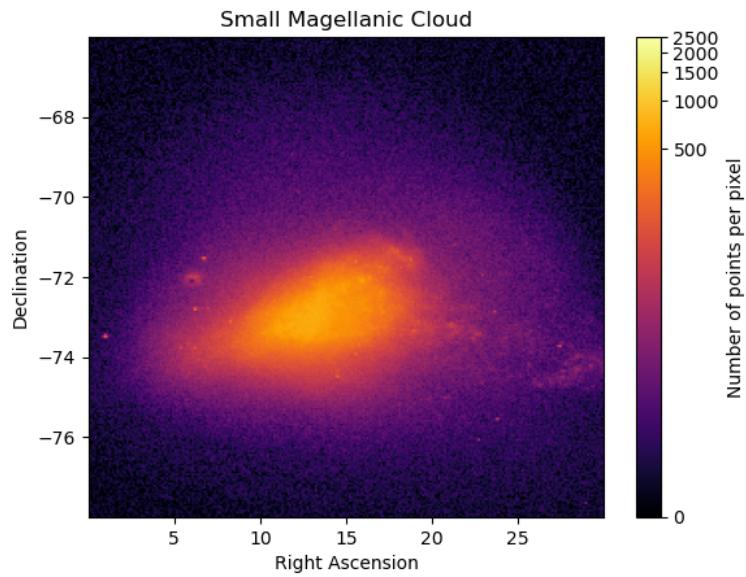


Figure 4.15: RA/Dec scatter plot with a logarithmically scaled colourmap.

As with LMC, we use the same method to plot the proper motions of SMC, just this time with the filtered stars.

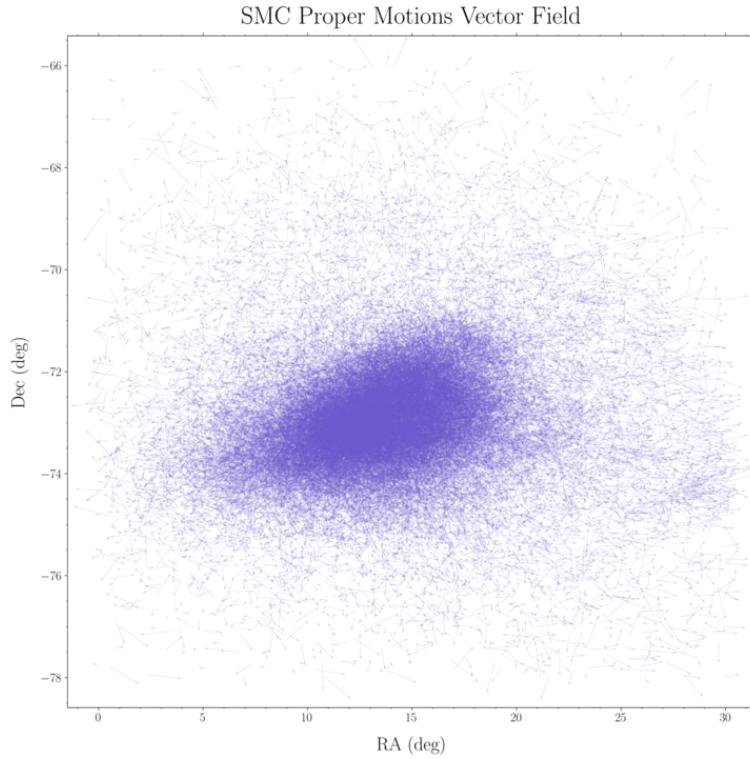


Figure 4.16

We did a different query for LMC to get a smaller sample of stars, as the plot here would not provide much information with such a large number of stars.

### 4.3 Magellanic Bridge

Magellanic Bridge is a stream of hydrogen with some stars inside it, connecting the Large and Small Magellanic Bridge. We query for it as follows:

```
'''select source_id, ra, dec, phot_g_mean_mag, pmra, pmdec, pm, bp_rp, ruwe
from gaiadr3.gaia_source
where
ra between 30 and 70 and
dec between -78 and -62
and pmra is not null
and pmdec is not null
and bp_rp is not null
and ruwe < 1.4
and phot_g_mean_mag is not null
order by ruwe asc'''
```

We get 1,618,591 stars from this query. Since the two galaxies are not at the same distance from earth filtering stars inside bridge is not possible by parallax cuts. So we go directly to HR diagram.

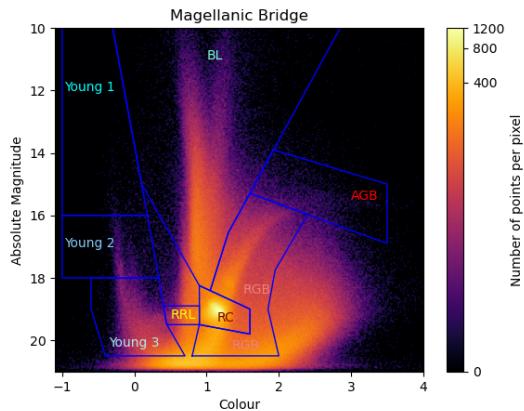


Figure 4.17: HR diagram of MB. The colormap has been logarithmically scaled.

We use the same polygons as LMC. This leaves us with 1,004,461 stars.  
The plot of evolutionary phases is given below:

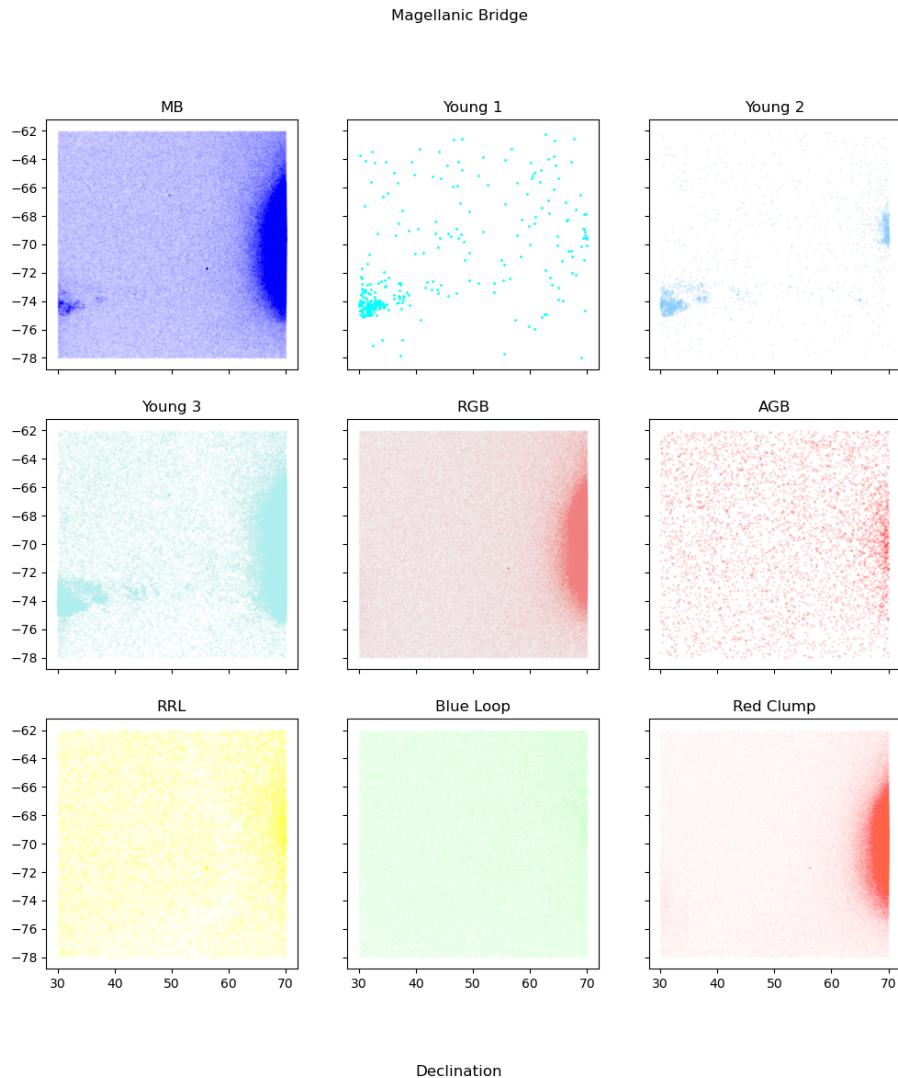


Figure 4.18: The scatter plots of various evolutionary phases of MB.

We divide the stars into young, consisting of young 1, young 2 and young 3 stars, and old which have the rest of the stars and plot the scatter plots again

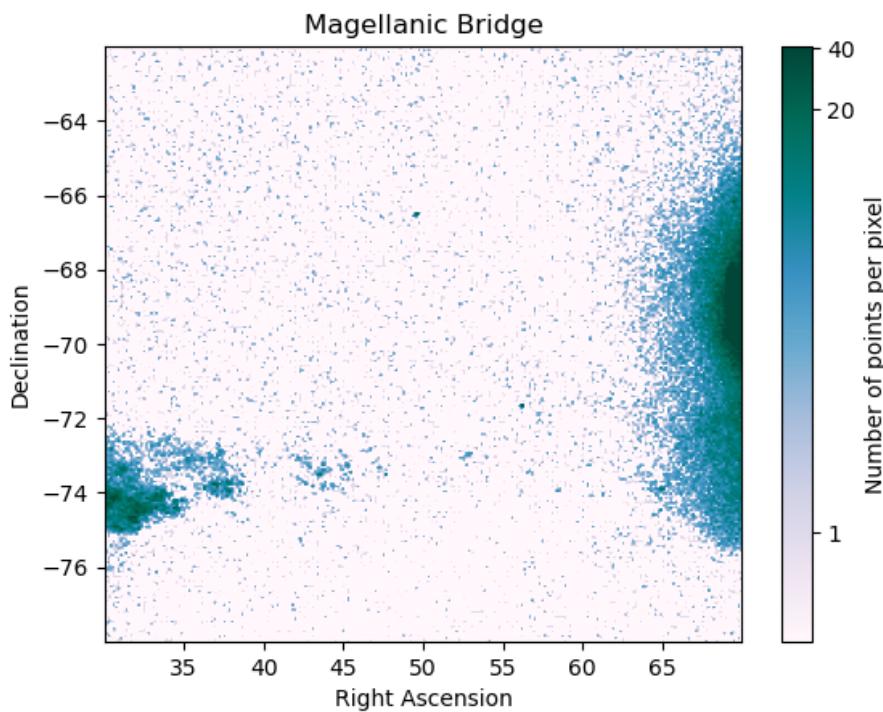


Figure 4.19: Visible bridge in young stars.

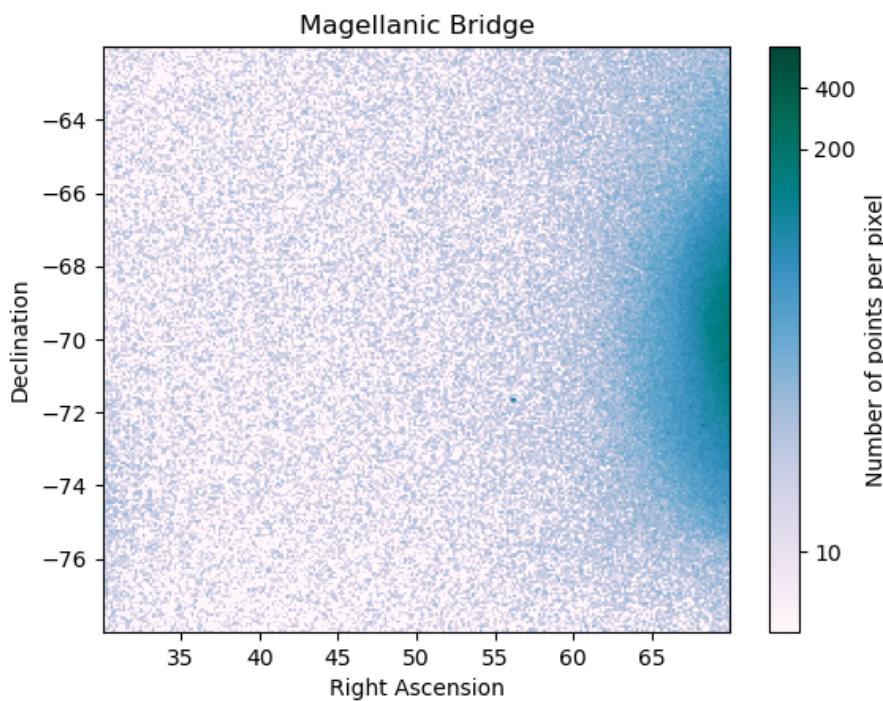


Figure 4.20: The bridge is not visible in old stars.



## 5. Time-Series Analysis - Auxiliary Data Products

In Gaia data release, key parameters are stored in the gaia source table that contains the (mean) astrometric, photometric, and radial-velocity data as well as astrophysical parameters. In addition to this, Gaia has a large amount of non-tabular data such as epoch photometry (Gaia goes over each star multiple times; this has the apparent magnitude and time recorded), mean spectra, (finer for hydrogen line to calculate radial velocity, called radial velocity spectrum or rvs. There is a broader spectra too, but it has a very large least count of 2 nm). These are accessible via the Datalink protocol.

We use the following functions and code given on ESA's guide on how to extract Datalink products.

### 5.1 Variable stars

Variable stars are stars whose brightness varies with time. There are various types of such stars, like Cepheids, RR Lyrae variables, and Long Period Variables. These stars can be studied by plotting apparent magnitude with time, which is called a light curve. We discussed RR Lyrae in the globular cluster section, now we will talk about the other two.

#### 5.1.1 Cepheids

These are stars which expand and contract periodically, and thus their brightness also changes periodically. These stars have very regular periods, ranging from 1 to 70 days and magnitude variation from 0.1 to 2 magnitudes. These are massive stars (~ 8 solar masses), and typically between class F to G/K. Polaris is a multiple star system with Polaris A, B and C. Polaris A is a Cepheid weighing 6 solar masses, while the other two are main sequence stars.

### 5.1.2 Long Period Variables (LPVs)

These are red giants or supergiants with periods of 30 to 1000 days. They have two subclasses; Miras and Semiregulars.

#### Mira

Miras are red giants with a period between 80 to 1000 days. Most medium sized main sequence stars go through this stage as they evolve into the RGB (including the Sun, eventually!). They have amplitude variations of more than 2.5 magnitudes.

#### Semiregular

These are red giants or supergiants with semiregular or irregular periods ranging from 30 to 1000 days. A prominent example is Betelgeuse.

## 5.2 Analysing Clusters using Spectroscopy data

In Gaia Datalink products, we have both the mean spectra and radial velocity spectra. One way to identify different spectral classes is by their spectra, which is distinct. We do this for two clusters, Blanco 1 and Beehive.

### 5.2.1 Blanco 1

There was no spectroscopy data available for class M stars, and no O stars in the cluster. So we skip that class.

#### Class B

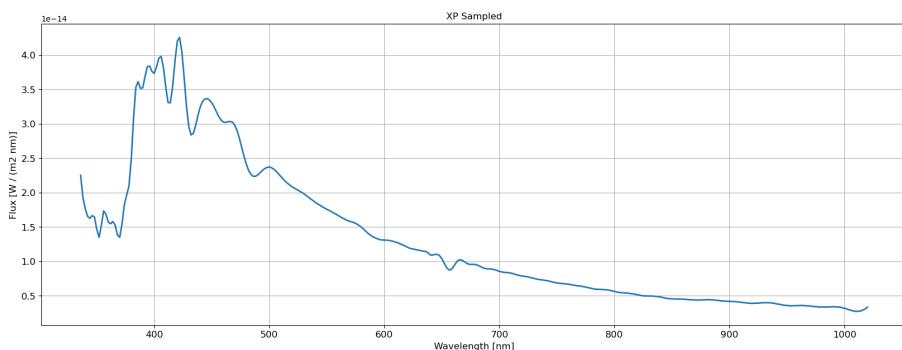


Figure 5.1: Class B spectra of Blanco 1

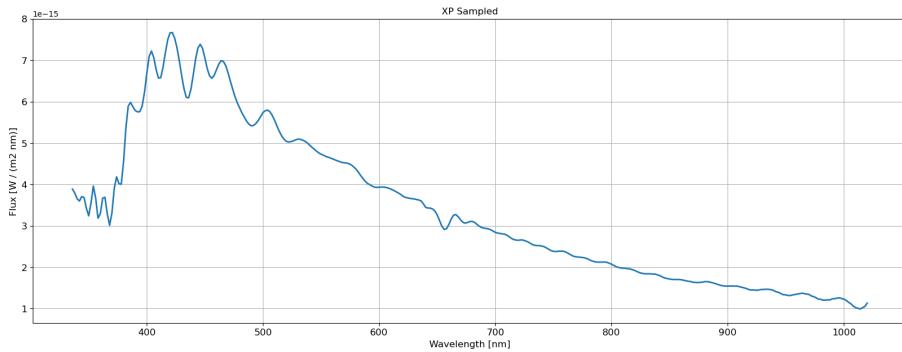
**Class A**

Figure 5.2: Class A spectra of Blanco 1

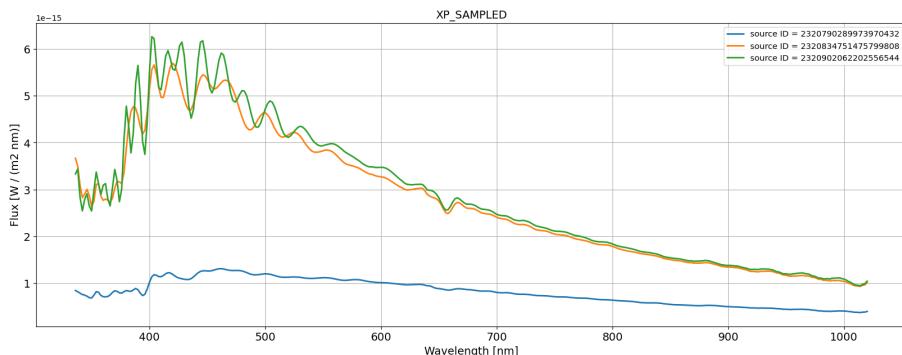
**Class F**

Figure 5.3: Class F spectra of Blanco 1

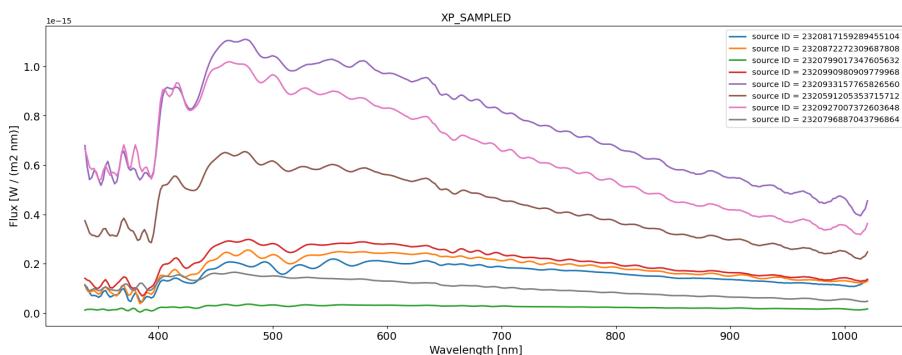
**Class G**

Figure 5.4: Class G spectra of Blanco 1

### Class K

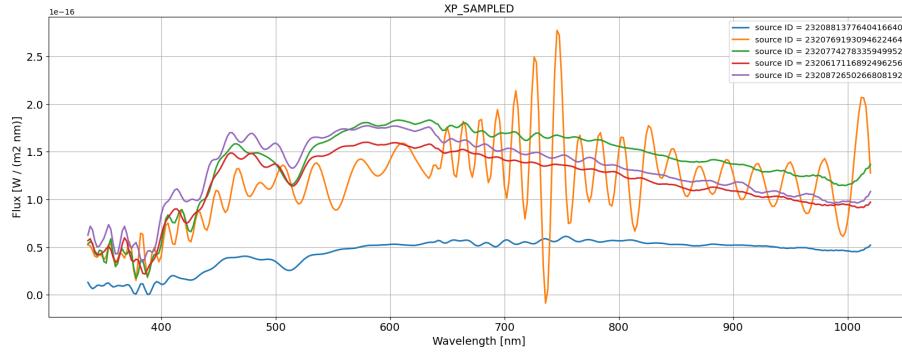


Figure 5.5: Class K spectra of Blanco 1

### 5.2.2 Beehive cluster

We do not have any O or B star, so we skip those classes.

### Class A

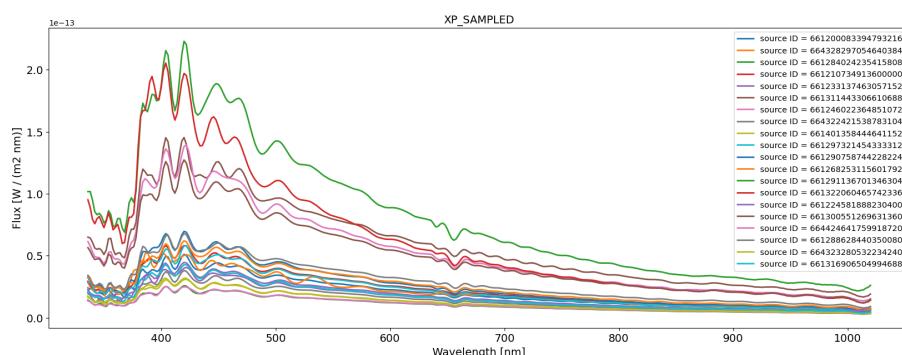


Figure 5.6: Class A spectra of Beehive

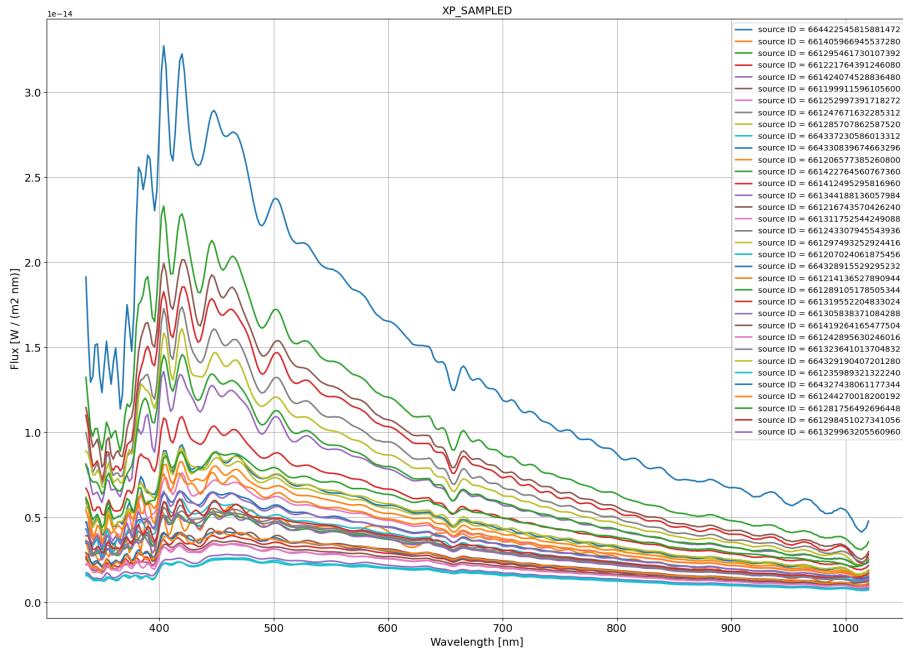
**Class F**

Figure 5.7: Class F spectra of Beehive

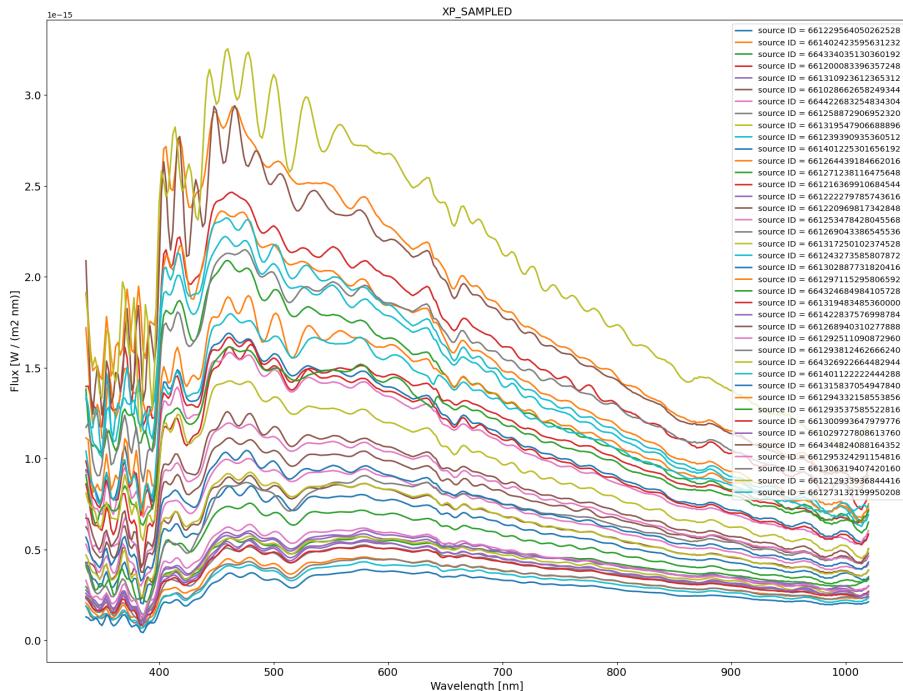
**Class G**

Figure 5.8: Class G spectra of Beehive

### Class K

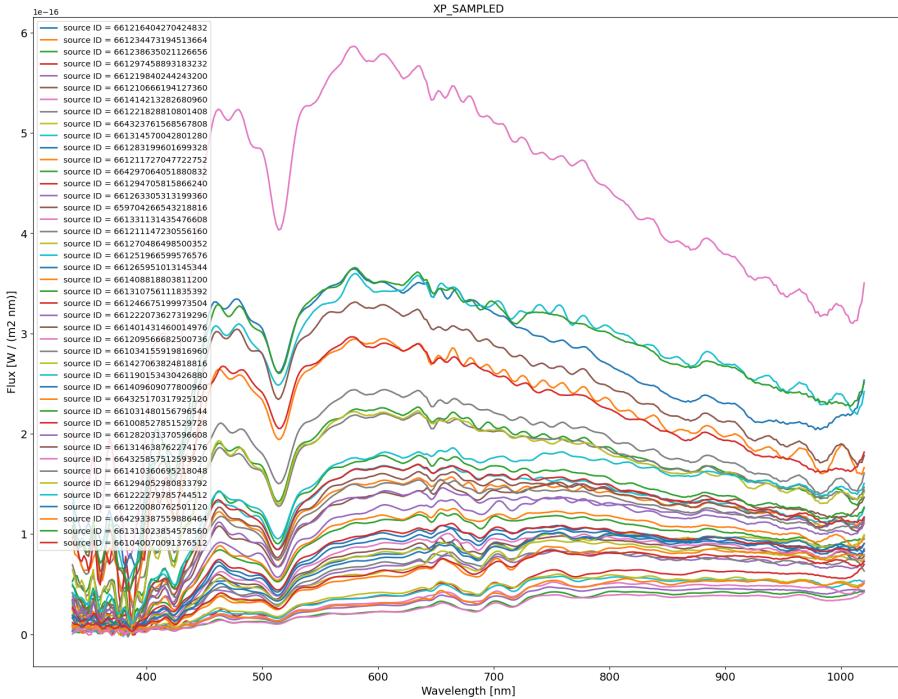


Figure 5.9: Class K spectra of Beehive

### Class M

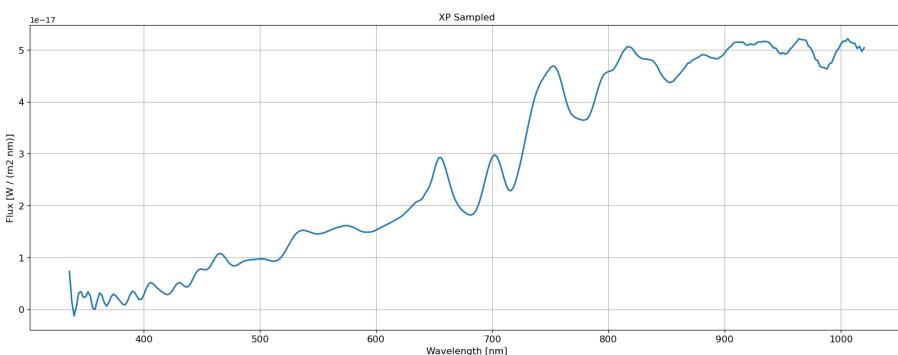


Figure 5.10: Class M spectra of Beehive

### Radial Velocity Spectrum

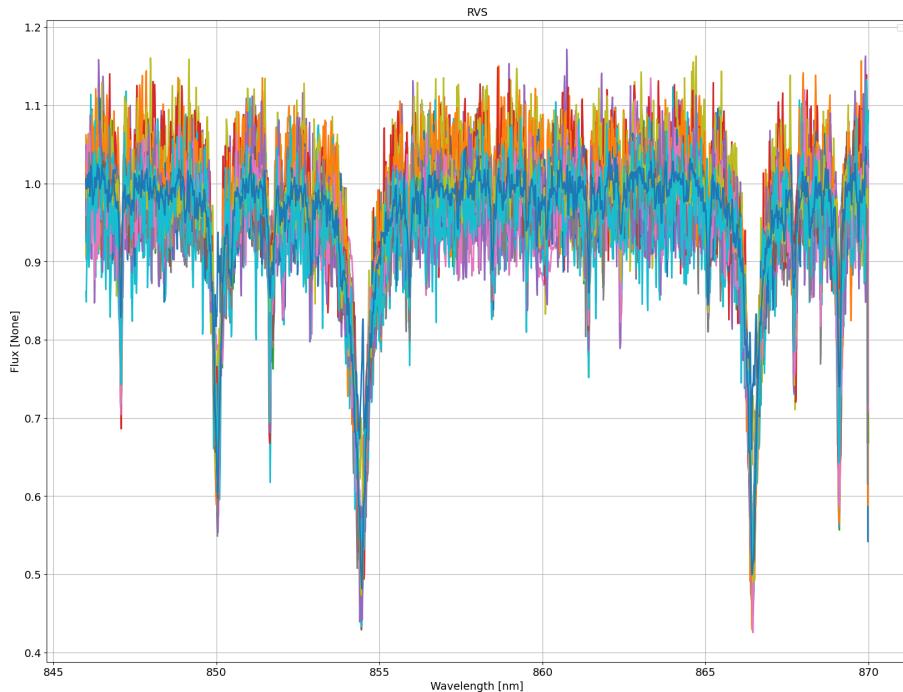


Figure 5.11: Radial Velocity spectra of Beehive

## 5.3 Analysing Cepheids

### 5.3.1 Individual stars

We first look at two Cepheids individually, Eta Aquillae and Delta Cepheid. The following code was used for Delta Cepheid. The source id was taken from Gaia archive.

```
query = f"SELECT source_id, ra, dec, pmra, pmdec, parallax \
FROM gaiadr3.gaia_source \
WHERE source_id = 2200153454733285248"
```

To get epoch photometry, we use the following code:

```
dl_key = 'EPOCH_PHOTOMETRY-Gaia DR3 2200153454733285248.xml'

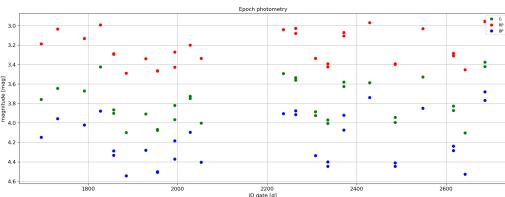
dl_out = extract_dl_ind(datalink, dl_key, figsize=[20,7])
dl_out
```

To get the mean spectra, called xp sampled, we run this:

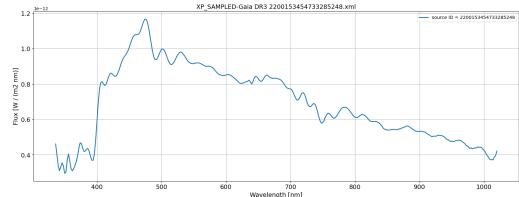
```
dl_key      = 'XP_SAMPLED-Gaia DR3 2200153454733285248.xml'
source_ids = [product.get_field_by_id("source_id").value for product in datalink[dl_key]]
tables     = [product.to_table()                                for product in datalink[dl_key]]

fig        = plt.figure(figsize=[20,7])
source_ids_i = iter(source_ids)
```

```
for inp_table in tables:
    plot_sampled_spec(inp_table, title=dl_key.replace('_COMBINED.xml', ''), legend = f'source_id={source_id}')
plt.show()
```



(a) Epoch photometry of Delta Cepheid.



(b) Mean spectra of Delta Cepheid.

Similarly with Eta Aquillae. This time we only have data for epoch photometry.

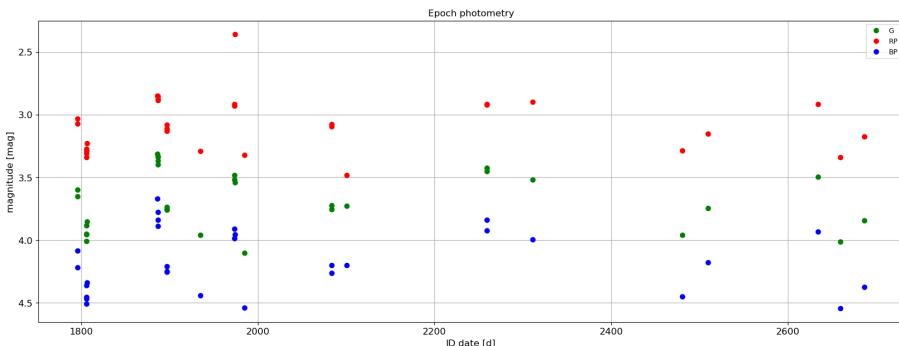


Figure 5.13: Epoch photometry of Eta Aquillae.

### 5.3.2 Period Luminosity Relationship

We try to fit a relationship between the period and the average magnitude of the stars, which is called a period-luminosity relationship. In Cepheids, the plot between logarithm of periods of various stars and their magnitudes is roughly a straight line, which is called the Leavitt Law.

The query for Cepheids is as follows:

```
'''select m.source_id, m.parallax, m.phot_bp_mean_mag, m.phot_rp_mean_mag, v.pf, m.phot_g_m
from gaiadr3.gaia_source as m, gaiadr3.vari_cepheid as v
where m.source_id = v.source_id and
m.parallax is not null and
m.has_epoch_photometry = 'True' and
v.pf > 0 and
m.parallax_over_error > 5 and
m.phot_g_mean_flux_over_error > 5
order by v.pf'''
```

This gives 1450 stars. We then retrieve their epoch photometry.

```
retrieval_type = 'EPOCH_PHOTOMETRY'
data_structure = 'INDIVIDUAL'
data_release = 'Gaia DR3'
```

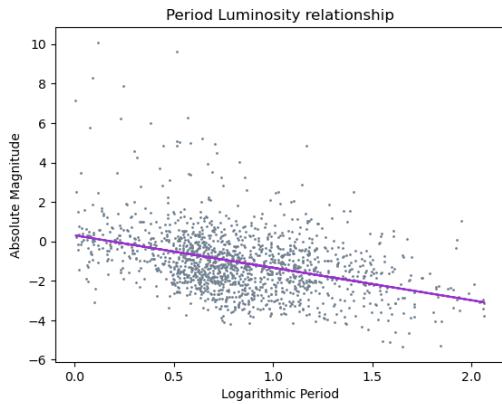
```

datalink = Gaia.load_data(ids=results['source_id'], data_release = data_release, retrieval_
dl_keys  = [inp for inp in datalink.keys()]
dl_keys.sort()

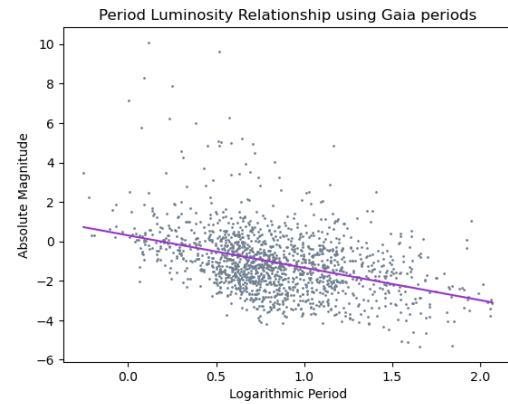
print()
print(f'The following Datalink products have been downloaded:')
for dl_key in dl_keys:
    print(f' * {dl_key}')

```

Then we use Lomb Scargle Periodogram to find their periods ourselves, and compare our results with Gaia periods. The minimum frequency used was 0.008 days, and maximum as 1 day.



(a) Period Luminosity relationship using Lomb Scargle period data.



(b) Period Luminosity relationship using Gaia period data.

In total there were 45 stars that did not agree with Gaia periods within 1% error. This may be due to less data which makes it difficult for Lomb Scargle to get the period.

The equation of line in fig (a) is  $M = -1.67\log(P) + 0.33$ , while the one in fig (b) is  $M = -1.7\log(P) + 0.37$ .



## 6. Results and Discussion

1. In open clusters, we saw that majority of the stars lie on the main sequence. These are divided into various spectral classes. By looking at distance distributions of these clusters, we found lighter, M stars are more spread across the cluster. As we go up the classes, the stars tend to be present towards the center. Also, cluster members have similar proper motions. We fit a power law on initial masses of the younger clusters, and derived the Initial Mass Function given by Salpeter empirically.

---

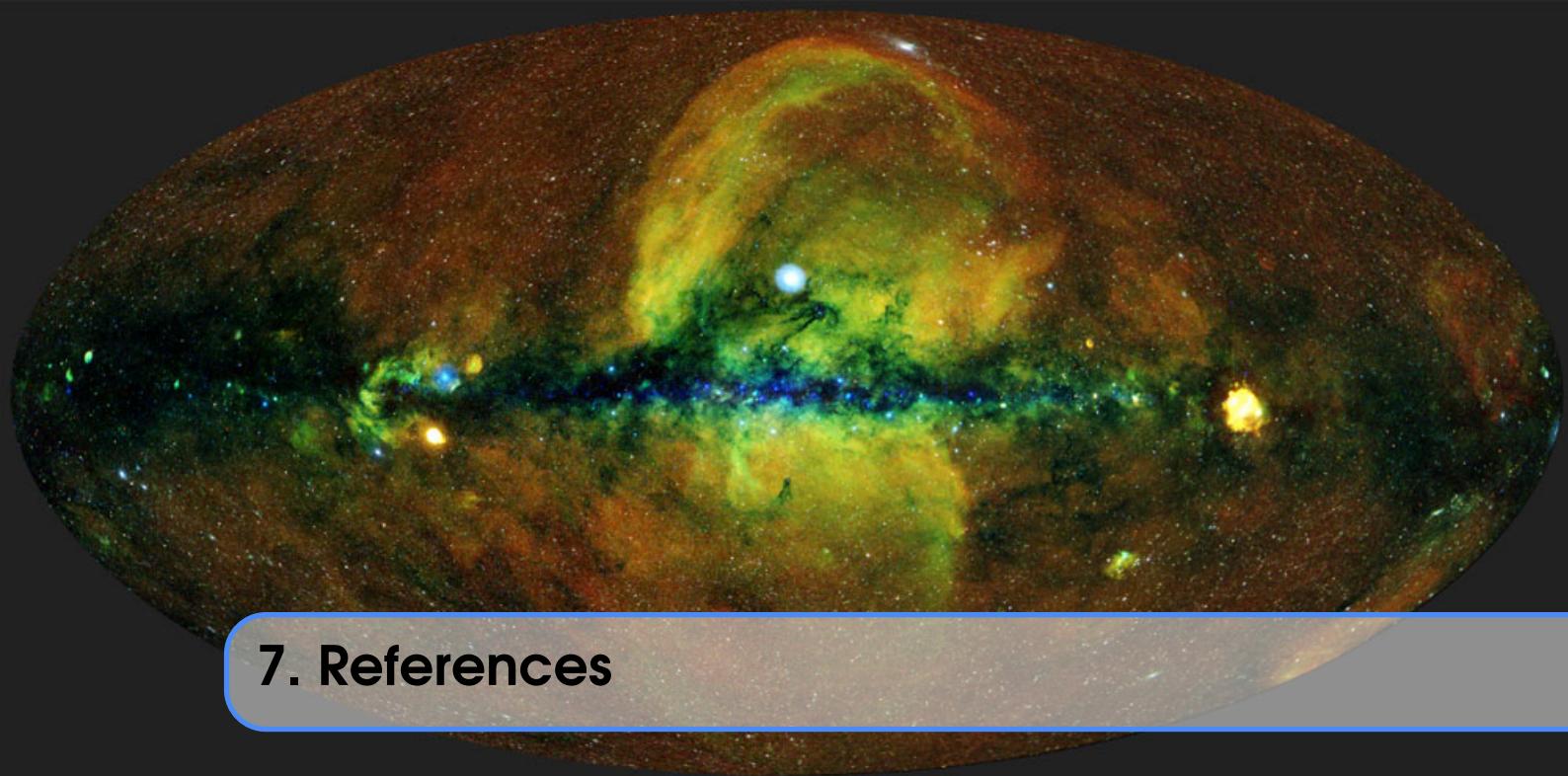
Cluster name	Distance (in pc)	Power Law
Beehive	187	$7.45m^{-0.78}$
Wishing Well	405	$81.27m^{-2.39}$
Messier 35	912	$85.11m^{-1.76}$
Blanco 1	253	$3.02m^{-1.65}$
Southern Beehive	346	$37.76m^{-1.88}$
Total of the above	-	$260.70m^{-2.17}$
Messier 45	136.2	$3.87m^{-1.81}$
Messier 50	881	$200.35m^{-3.63}$
Messier 38	1066	$443.6m^{-2.08}$
Lambda Orionis	400	$11.46m^{-3.95}$
Alpha Persei Cluster	175	$5.58m^{-0.99}$
Messier 48	416.6	$12.05m^{-1.69}$
Trapezium Cluster	373.3	$2.68m^{-2.81}$

2. While analysing globular clusters, we found stars in evolutionary phases not seen before. We found that these clusters, are indeed, much more old, and massive than open clusters. By using fitting various parts of the HR diagram, we found the distance of these clusters from Earth. By visually fitting isochrones we also put an estimate on the age and metallicity of these clusters. The following table summarises this.

Cluster name	Distance (in kpc)	Age (in Gyr)	Metallicity (in dex)	sideways shift (in mag)
M3	11.17	11	-1.5	0
M4	3.31	12.5/13	-0.9	+0.50
M5	20.08	11.5	-1.2	0.0
M13	7.59	12.5	-1.2	0
M68	11.69	11 12.5	-1.6 -1.9	0 0
M79	13.49	13.5	-1.5	0
M92	9.12	13.5	-2.2	0
M107	9.12	13.5	-0.9	+0.52
NGC 6541	8.91	13.5	-1.5	0
NGC 5139	5.24	13.5	-1.5	+0.18
NGC 104	4.45	13	-0.6	0
M30	7.8	13.5	-1.2	0.11
47 Tucanae	4.78	13	-0.6	0

3. We then looked at the two dwarf galaxies, Large and Small Magellanic Clouds. We plotted the average proper motion of the stars inside them, and saw the way in which the speeds increase close to the centre. We then filtered them and found the distribution of stars in various evolutionary stages, and their distribution. Older stars were much more scattered than young ones. The sort-of spiral shape of LMC was seen, while SMC was much more irregular. Also there is a bridge present between the two dwarf galaxies, which is visually apparent when seeing the younger stars.
4. Gaia has a large amount of data, so we used its spectroscopy data for two open clusters, Beehive and Blanco 1. We saw that each spectral class has a distinct spectra. Due to the least count of 2 nm, it is hard to tell which line is at the peak exactly. We also found the radial velocity spectra for Beehive. Then, using epoch photometry data, we first did an exercise on two stars, Delta Cepheid and Eta Aquillae. After that we took 1450 stars and derived the period-luminosity relationship for Cepheids empirically, using Lomb Scargle periodogram. The equation we got was as follows:

$$M = -1.67 \log(P) + 0.33$$



## 7. References

- [Gaia DR3 Data Model](#) - to obtain the table names, column names and the information contained in them
- [Gaia Data Access using Python](#) - gives the procedure and syntax to extract Gaia data
- [Building HR diagrams with Gaia data](#) - the complete HR diagram for all stars in Gaia
- [How to extract the Gaia ancillary data using datalink - Gaia Users - Cosmos](#) - to obtain Datalink products from Gaia archive.
- [Stellar Classification - Wikipedia](#)
- [Post-Main Sequence Stars](#) - for various evolutionary phases
- [Star Clusters](#) - blue stragglers image
- [RR Lyrae Variables - Wikipedia](#) - for RR Lyrae variables image
- Bressan et al. (2012), MNRAS, 427, 127 + Chen et al. (2014, 2015), MNRAS, 444, 2525 + MNRAS, 452, 1068 + Tang et al. (2014), MNRAS, 445, 4287 + Marigo et al. (2017), ApJ, 835, 77 + Pastorelli al. (2019), MNRAS, 485, 5666 + Pastorelli al. (2020), MNRAS, in press - for PARSEC isochrones used
- [Temperature of Stars using Gaia Data](#)
- [Open Clusters - Wikipedia](#)
- [Variable Stars](#) - for types of variable stars.