Application of the Ant Colony Algorithm in the Identification of Globular Clusters

Bachelor's Project Thesis

Jördis Hollander, s2956543, j.hollander.6@student.rug.nl

Supervisors:
Prof. Dr. H. Jaeger
Dr. E. Balbinot

May 23, 2022

Abstract

Globular clusters (GCs) are stellar agglomerates of about $10\,000$ to $100\,000$ stars. They provide an interesting ground to study stellar evolution. The complexity of the Universe makes the precise identification and classification of stellar structures challenging. In this paper, a pipeline for the identification of GCs is developed based on work by M. Mohammadi et al.[*] This pipeline consists of excluding candidate regions through the use of a blob-detection technique. The remaining regions are then processed by the Ant Colony random-walk algorithm. This algorithm investigates a region and provides information on its stellar density in the form of pheromone values. Finally, these results are fed into a *gravity-inspired* clustering algorithm that was developed to interpret the pheromone values to determine potential GCs.

The aim of the research is to determine the accuracy of the pipeline in classifying GCs and investigate possible improvements. This pipeline is then applied on the Gaia DR2 data-set. Different areas consisting of a variety of stellar objects are selected. Some of these areas contain known GCs (Area 1, 2, and 3), while some do not. Each of these areas is split via a rasterization process into evenly distributed rasters. The accuracy of the pipeline is explored by running it on these different rasters and considering firstly if it finds all known GCs and secondly by considering what other stellar structures it classifies as GCs.

For the blob-detection technique a cutoff point for the constant representing the minimal acceptable blob size was identified to be 0.2. Under this cutoff point the blob-detection technique filters away $87.5\,\%$ (813 out of 929) of the candidate rasters and maintains $76.9\,\%$ (20 of the 26) rasters that contain known GCs. For the three areas containing known GCs this is:

➤ Area 1: 7 out of 12 rasters containing GCs

➤ Area 2: 1 out of 1 rasters containing GCs

➤ Area 3: 16 out of 17 rasters containing GCs

The Ant Colony algorithm coupled with the clustering algorithm applied across the same rasters leads to varying results per execution. Across 5 experiments, an average of 51 ± 4 clusters are found. The combined results over all experiments show that the clustering maintains $43.3\,\%$ (13 out of 30) of the known GCs:

➤ Area 1: 7 out of 12 GCs

➤ Area 2: 1 out of 1 GCs

➤ Area 3: 5 out of 17 GCs

The results of the full pipeline identifies 41 clusters of which 27 could be identified as known stellar structures. These clusters are $(31.7\,\%)$ 13 GCs, $(12.2\,\%)$ 5 Open Clusters, $(9.8\,\%)$ 4 Galaxies, $(4.9\,\%)$ 2 Dwarf Galaxies, $(2.4\,\%)$ a Molecular Cloud, $(2.4\,\%)$ an Absorption Nebula, and $(2.4\,\%)$ an Emission Nebula. In addition, it finds $(34.1\,\%)$ 14 clusters that do not correspond to a known stellar structure. For Areas 1, 3, and 4 most of the clusters are found consistently across the experiments. However, in Area 2 the clusters are only found sporadically, with each cluster being found only in at most two experiments.

It is evident from these results that the blob-detection operates as an effective exclusion criteria but with the current constant it does not yet maintain all known GCs. While the pipeline does not identify the majority of the GCs that exist, for those that it is able to identify, it can pinpoint their locations accurately. Further research in tuning the parameters and steering the behavior of the ants is expected to expand the number of GCs identified by the pipeline and to solidify the Ant Colony as a useful tool for exploring the Universe. With further refinement of the process and the synergy of the initial blob filtration and clustering applied on the Ant Colony pheromones, this pipeline can likely be made robust.

---

[*]M. Mohammadi, N. Petkov, K. Bunte, R. Peletier, and F.M. Schleif, "Globular Cluster Detection in the Gaia Survey," *Neurocomputing*, vol. 342, pp. 164–171, 2019.

Acknowledgements

Contents

Chapter 1

Introduction

The speed of light in a vacuum is constant. This fact has enabled us to look into the past and observe how the Universe has evolved over time. From the cosmic soup, to the emergence of the first stars, to the formation of galaxies; the evolution of the Universe has been a perennial source of fascination for mankind. One area of particular interest is the evolution of galaxies. Early on most galaxies were small, but gradually these smaller galaxies merged and amalgamated to form larger galaxies [1]. Over time, as they continued to evolve, they began to manifest a great variety of galactic structures. However, this begins to raise a fundamental question: *How does one classify the stages of evolution for a galaxy or determine where it is in its evolution?*

Hubble's Tuning Fork was the first classification scheme that sought to answer this, based on a galaxy's structure and size [2]. However, this scheme proved to be insufficient in the face of the complexity of the Universe and the variety of possible galactic formations [2]. An alternative approach to this classification scheme is to attempt to determine the age of galaxies based on their composition and not their overall shape. One such method involves determining the ages of various clusters of stars within a galaxy, thereby providing insight into the origin of the stars that constitute it.

Of the various cluster types, the globular cluster (GC) is on average the oldest [3], and thus, the most significant in gauging the age of galaxies. These types of clusters are stellar agglomerates [4] which formed in one of the following two ways [5]:

1. Through the compression of halo gas in the cosmic re-ionization phase early on in the formation of the Universe.

2. In the collapse of enormous molecular clouds triggered by events such as the collision of gas-rich galaxies.

They are typically composed of around $10\,000$ to $100\,000$ stars [6] bound tightly by gravity into a spherical formation. Some GCs are among the oldest objects in the Universe [7] and are, thus, an interesting foundation for studying galactic evolution [8]. These older GCs manifest some specific properties such as low metallicity [9], and through a combination of techniques, such as horizontal branch morphology, analysis of white dwarf cooling sequences, and comparisons using the main-sequence turn-off location [7], [10], may have their age accurately determined, thereby setting the bounds for the age of the galaxy they are contained within. The question remains: *How are these GCs found?*

1.1 Existing Methods of Identifying GCs

Astronomy is primarily an observational science, which, for most of its history, has gleaned information by looking up to the night sky with nothing but the naked eye. It was in this manner that the first GC, Messier 22, was discovered. It was observed by Abraham Ilhe on August 6th, 1665 and as may be seen in Figure 1.1 demonstrates an unusually dense core [11]. This GC became the subject of much research across the 1900s and the statistics it has provided have been a basis for the identification of GCs today. The primary method for detecting new GCs has involved statistical analysis of photometric data across the Universe. Properties such as mean luminosity and metallicity are extracted from known GCs and are used to filter regions based on this spectroscopic information [12], [13]. To date, over 150 GCs have been discovered in our Milky Way, using such techniques [14].

Figure 1.1: Crammed Center of Messier 22 taken by ESA/Hubble [15]

As a brief interlude, the advent of the large-scale tools used in modern astronomy have allowed for the collection of enormous amounts of information within our galaxy and far beyond. Telescopes (refractory, reflector, radiography, spectrography, and x-ray) allow us to extract a variety of information all without leaving the Earth's surface [16]. To collect information without the interference of the Earth's atmosphere we make use of satellites and space observatories that have been launched into orbit around our planet [17]. Occasionally, a space probe will be sent beyond our orbit to collect information from asteroids, planets, or their moons within our solar system [17]. However, for stellar observations, the use of telescopes becomes necessary due to the sheer distances involved. Thus, the techniques used for the identification of GCs were predominantly limited to the evaluation of photometric data and performing statistical analysis.

However, this has changed with the first Gaia data release in September 2016. In this data-set stars are represented by IDs and are coupled with a variety of astrometric, photometric, and spectroscopic readings [18] (see Chapter 2 for details). This allows processing to occur on a per-star basis and allows for a greater variety of techniques to be employed. The state-of-the-art in GC detection across such data-sets is described by Mohammadi et al. in 2018 [8]. In their paper, they make use of 3D kernel density estimations (*KDE*) across subdivisions of the data-set. Using these estimations they then evaluated two different techniques:

- Nearest-neighbors search.
- Kernel based anomaly detection through the training of a support-vector machine on random portions of the sky to search for outliers.

For both techniques, they then use a blob-detection technique based on Difference of Gaussian, as a post-processing method to further filter the regions.

1.2 Objectives

The research in this paper follows from the work of Mohammadi et al. and seeks to automate the discovery of potential GCs using data from stellar catalogs such as the Gaia data-set. The objective is to produce a pipeline using:

1. Blob-detection: based on Difference of Gaussian (*BlobDoG*), as a crude initial filter that reduces the number of candidate regions for further inspection.

2. Ant Colony random-walk algorithm: to compute density information in the form of pheromone values.

3. Clustering: via an algorithm which pools these pheromone values in 3D space gravitationally to determine clusters.

This pipeline is optimized and tested on data from the Gaia DR2 data-set [18], [19]. Different regions consisting of a variety of stellar objects are selected. Some of these regions contain previously found GCs, while some do not. Since we cannot state with absolute certainty the total number of GCs (for any given region, more may yet be found), the accuracy of the pipeline is explored by

running it on these different regions and then evaluating it in two ways. First, we determine if it finds all known GCs, and second we consider what other stellar structures (if any) it classifies as GCs.

A robust classifier for GCs would be a useful addition to the tool-set for exploring the Universe and its utility could be further evaluated with the full publication of the Gaia DR3 data-set slated for release in the first-half of 2022 [20].

It is important to note that due to the scope of this project, the emphasis does not lie in precisely optimizing the parameters associated with the pipeline. The aim is to identify the effectiveness and shortcomings of the individual stages of the pipeline, as well as to identify potential parametric relationships between the different stages.

1.3 Summary

The rest of this paper is structured as follows:

First, we take a closer look at the Gaia Data Release 2 in Chapter 2. We begin by depicting the stellar regions under evaluation. This is followed by a description of various astronomical objects that will be of interest in the analysis. Additionally, the parameters that will be made use of from the data-set are scrutinized.

In the chapter on methodology, Chapter 3, the overview of the pipeline is provided. This is coupled with in-depth descriptions of the three techniques in use, namely, Difference of Gaussian, the Ant Colony algorithm, and the Gravitational Clustering algorithm.

The results are explored in Chapter 4, which discusses the findings for the individual components of the pipeline. Additionally, it includes the statistical results and plots which reveal interesting characteristics of the pipeline. The results for the regions are compared and contrasted to explore the differences.

In Chapter 5, conclusions are drawn about the effectiveness of the pipeline based on the results and discusses how they compare against the objectives that were set forth.

In Chapter 6, the results are interpreted and evaluated to provide an insight into how well the individual components of the pipeline have worked, and attempts to learn from the successes and shortcomings. It also describes worthwhile avenues for further research.

Chapter 2

Data

The stellar data selected for this project is a subset of the Gaia DR2 data-set [19]. This data, collected during the Gaia satellite mission [19], is approximately 500 GB and consists of a variety of parameters collected on a per-star basis. This is in great contrast with typical telescopic data which is primarily raw radiometric imagery and greatly expands the types of processing that may be conveniently applied. The subset of the parameters that are of interest to the pipeline are described in Section 2.1. As a result of hardware constraints, the investigation has been limited to a set of smaller regions from within the Gaia data-set. Four distinct areas were selected, bounded by by right ascension (*RA*) × declination (*Dec*), and represent regions of interest with varying stellar distributions. Table 2.1 sets out the cosmic ranges as well as the number of stars within these areas.

Table 2.1: Areas Under Investigation

Area	RA	Dec	Number of Stars
A1	120° up to 246°	−2° up to 60°	25 486 556
A2	295° up to 308°	15° up to 25°	23 470 239
A3	0° up to 75°	−90° up to −30°	16 781 316
A4	0° up to 45°	30° up to 70°	32 333 936

Rationale for the choice of these four regions:

Area 1 (A1): This region was used in the precursor work of Mohammadi et al. [8]. It is the largest region of the four in terms of their span across RA and Dec. Conversely, it also has the lowest density of stars across the region. This results in large areas that contain very few stars. These *dark* regions are very apparent in Figure 2.1a. It contains several known GCs, including a number in these *dark* regions, which are expected to be identified easily as they stand out from their dark and void stellar surroundings.

Area 2 (A2): The region was chosen because it is a much smaller area with a very high-density of stars across the whole span. Additionally, it contains only one GC. This is useful as a testing ground to see how the pipeline would handle regions featuring a high amount of stars.

Area 3 (A3): This area was also used in the paper by Mohammadi et al. and though its overall density is less than for A2, it features specific regions with an extremely high density of stars. Additionally, it contains 17 GCs, the Magellanic Clouds (the name given to a specific pair of Dwarf galaxies), a supernova remnant, open clusters, and galaxies. The Magellanic Clouds are the two extremely bright, very densely packed regions that may be seen in Figure 2.1c. These two very bright regions also contain some GCs and it would be useful to see how the pipeline handles classifying these clusters with the interference of the Magellanic clouds.

Area 4 (A4): This area has no GCs, but contains the nearby and very bright Andromeda and Triangulum galaxies. These galaxies lie across the range of 30°–50° Dec, and it would be illuminating to see if the pipeline detects them and classifies them as GCs. The remaining range, from 50°–70° contains a large number of nebulae (large regions of very bright loosely-packed stellar gas).

(a) A1

(b) A2

(c) A3

(d) A4
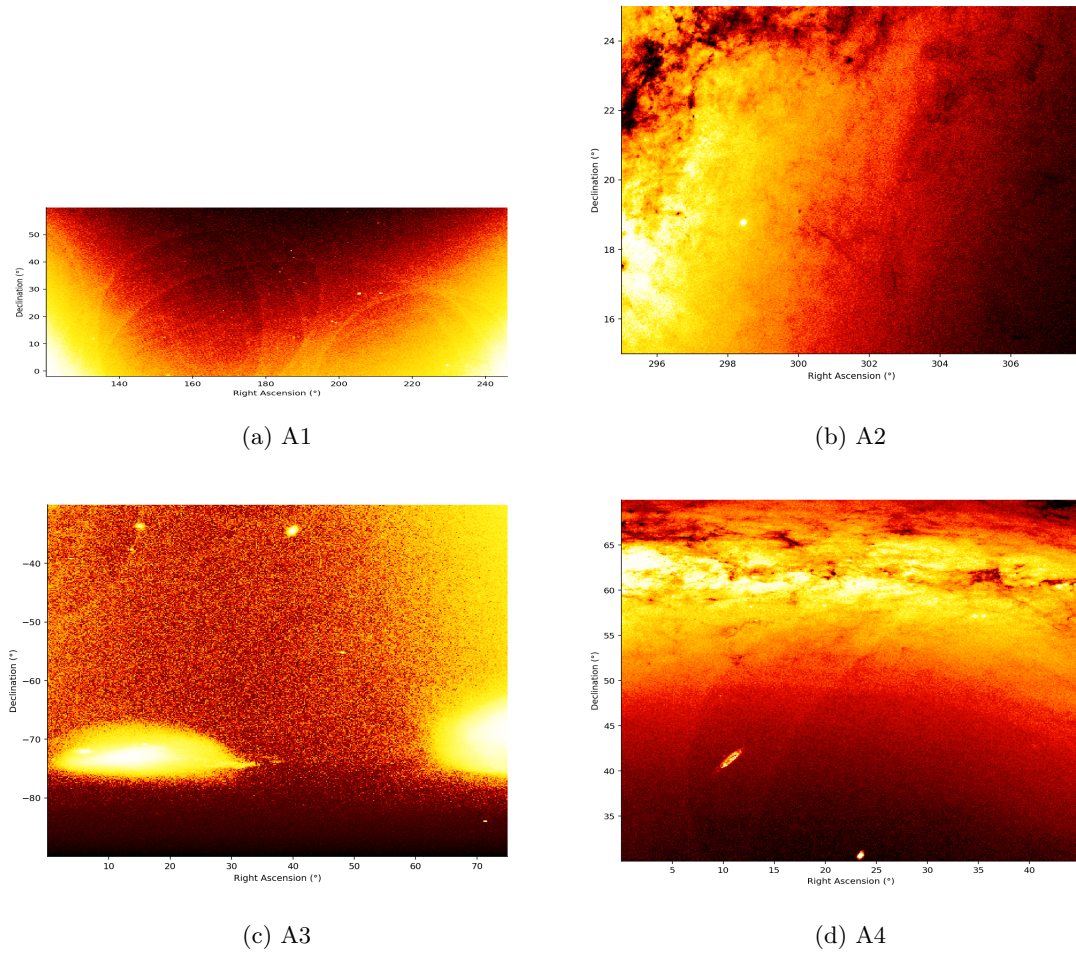
Figure 2.1: Stellar Distribution Heat-maps for the Four Areas

Figure 2.1 depicts heat-maps of the four regions synthesized from the stellar information in the Gaia DR2 data-set. It provides an insight into the population and density of the stars found within the four areas. The brighter areas (yellow–white) contain more stars than the darker areas (red–black). From these heat-maps, the spots of increased stellar density are very evident. However, it is not immediately apparent whether these spots are GCs, Open Clusters (*OC*), galaxies (*Gal*), or some other stellar structure. Figure 2.2, provides an example of some of these stellar structures for Area 1 and further highlights the difficulty in classifying these stellar structures by eye alone.
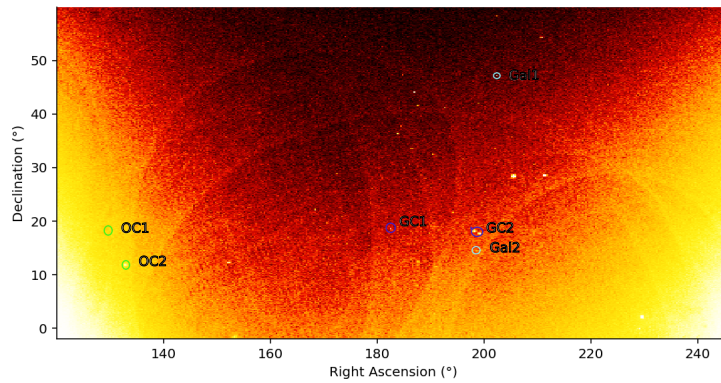
○: Open Clusters	○: Globular Clusters	○: Galaxies
OC1: Messier 44	GC1: NGC4147	Gal1: The Whirlpool Galaxy
OC2: Messier 67	GC2: NGC5024 and NGC5053	Gal2: Malin 1

Figure 2.2: Stellar Structures Present in A1

Area 1 contains more than the three GCs encircled in Figure 2.2. It contains an additional nine GCs (for a total of 12), Area 2 contains just one known GC, and Area 3 contains 17 GCs [21], [22]. See Table 2.2 for information on the RA, the Dec, and the angular diameter (*DIA*) for these globular clusters where this information has been published. Note that the *DIA* is represented in arcminutes (′) which is a measure of angular distance where $1' = \frac{1}{60}°$ and is useful to provide an expectation of the size of a GC.

Table 2.2: Known GCs

GC	RA (°)	Dec (°)	DIA (′)
Area 1			
Palomar 3	151°382 92	00°04′18″0	1′6
Willman 1	162°35	51°03′00″0	7′
Palomar 4	172°319 99	28°58′24″9	1′3
Koposov 1	179°827 09	12°15′36″0	—
NGC 4147	182°526 26	18°32′33″5	4′4
NGC 5024	198°229 45	18°01′05″4	13′
NGC 5053	199°112 88	17°42′00″5	10′
M3	205°548 42	28°22′38″2	18′
NGC 5466	211°363 71	28°32′04″0	9′
Palomar 5	229°021 87	00°06′41″8	8′0
M5	229°639 62	02°04′54″9	21′6
GCI 38	242°752 47	14°57′28″0	2′2
Area 2			
M71	298°44	18°46′45″1	7′2
Area 3			
47 Tucanae	06°023 63	−72°04′52″6	50′
NGC 121	06°701 04	−71°32′08″4	—
NGC 362	15°809 42	−70°50′55″6	14′
NGC 1049	39°968 75	−34°16′08″	—
NGC 1261	48°067 54	−55°12′59″2	6′85
NGC 1466	56°1375	−71°40′17″0	—
Arp Madore 1	58°761 25	−49°36′52″0	—
NGC 1629	67°404 17	−71°50′18″	—
NGC 1651	69°386 25	−70°35′08″	—
NGC 1644	69°415 00	−66°11′49″	—
NGC 1652	69°595 42	−68°40′23″	—
NGC 1841	71°346 25	−83°59′49″	—
NGC 1696	72°125 00	−68°14′35″	—
NGC 1756	73°707 92	−69°14′15″	—
NGC 1786	74°782 91	−67°44′43″	—
NGC 1783	74°786 67	−65°59′07″	—
NGC 1795	74°945 83	−69°48′05″	—

It is these GCs that are used in the evaluation of the results of the pipeline.

2.1 The Parameters

The Gaia DR2 data-set provides up to 88 parameters per star [19]. Of these parameters, there are six that are required for the pipeline:

1. Apparent Magnitude: This is a unitless quantity and is the measure of a star's brightness when observed from Earth (N.B.: A higher apparent magnitude corresponds to a less bright star). This is under the key `phot_g_mean_mag` within the data-set.

2. Right Ascension (RA): This quantity is represented in degrees (°) and when coupled with the declination it provides a position for an astronomical body in the equatorial coordinate system. This is under the key `ra` within the data-set.

3. Declination (Dec): This quantity is represented in degrees (°) and when coupled with the right ascension it provides a position for an astronomical body in the equatorial coordinate system. This is under the key `dec` within the data-set.

4. Parallax: This quantity is measured in milliarcseconds (mas) and is the difference in the apparent position of an object when viewed along two different lines of sight [23]. This is under the key `parallax` within the data-set.

5. Proper Motion of Right Ascension (PM_{RA}): Expressed in $mas\,yr^{-1}$ and is the motion of an astronomical body from the frame of the center of mass of the solar system in right ascension. This is under the key `pmra` within the data-set.

6. Proper Motion of Declination (PM_{Dec}): Expressed in $mas\,yr^{-1}$ and is the motion of an astronomical body from the frame of the center of mass of the solar system in declination. This is under the key `pmdec` within the data-set.

The details underlying the selection of these parameters alongside statistical analysis is provided in the sections that follow.

For the development of the pipeline, it is important to note, that the Gaia DR2 data-set does not include the parallax, PM_{RA}, or PM_{Dec} for *all* stellar objects [24]. Additionally, the Gaia data-set contains some spurious negative values for the parallax [25]. Since these parameters are required for the pipeline, the stars that are missing valid values for these parameters have been filtered out. The details on how many stars are filtered per area and the percentage that remain are given in Table 2.3.

Table 2.3: Filtered Stars

Area	Total Number of Stars	Remaining Stars	Percentage Remaining
A1	25 486 556	17 933 864	70.4%
A2	23 470 239	14 268 513	60.8%
A3	16 781 316	9 961 034	59.4%
A4	32 333 936	22 243 660	68.8%

2.1.1 Apparent Magnitude

The apparent magnitude provides information on the brightness of a star. In isolation this is influenced by the mass of the star and the proximity of the star to the observer. The closer the star is, the brighter it will appear. The more massive the star, the higher its black-body radiation; thus, the more energy it releases in the form of brighter light.

This parameter is used alongside, RA and Dec, to create the 2-dimensional mapping that is fed into the Difference of Gaussian filters to detect blobs. Histograms of the apparent magnitude across the four areas may be seen in Figure 2.3. Note that the frequency is presented on a logarithmic axis and that a larger apparent magnitude corresponds to a less bright star.
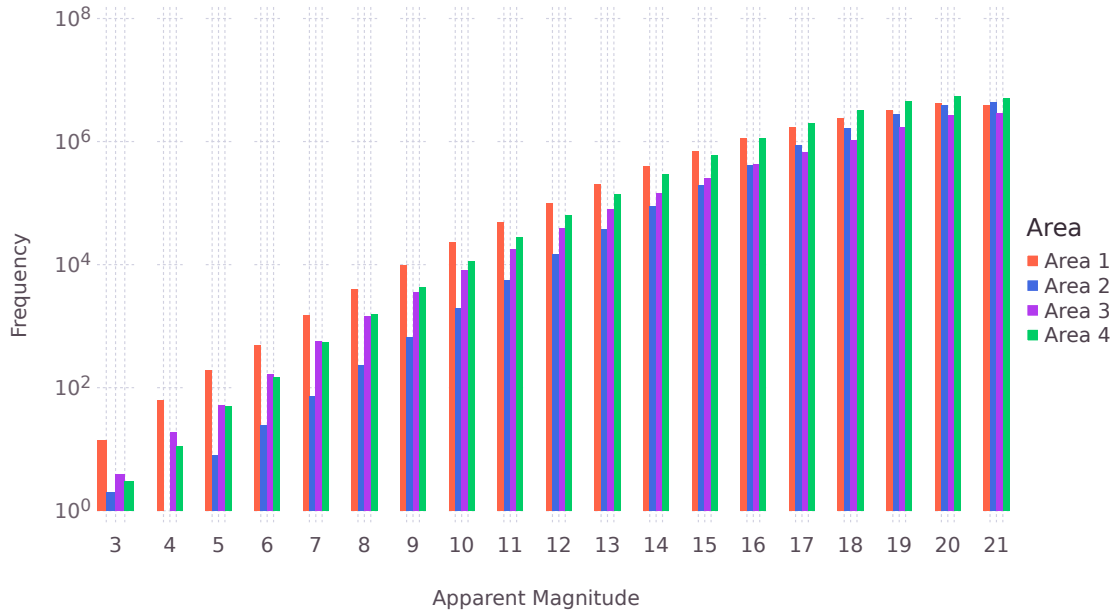
Figure 2.3: Distribution of Apparent Magnitude Across the Four Areas

In Figure 2.3 it is apparent that as the stars grow dimmer, the distribution of stars between the four areas become more uniform. It is also evident that Area 1 contains the most bright stars out of the four regions and that Area 2 contains the least amount of bright stars. However, this does not reveal the full story as each of the four areas contain a differing total number of stars. Thus, it is also necessary to consider the proportion of stars that lie within a specific apparent magnitude against the total number of stars contained within that area. This proportional distribution for each of the four areas may be seen in Figure 2.4.
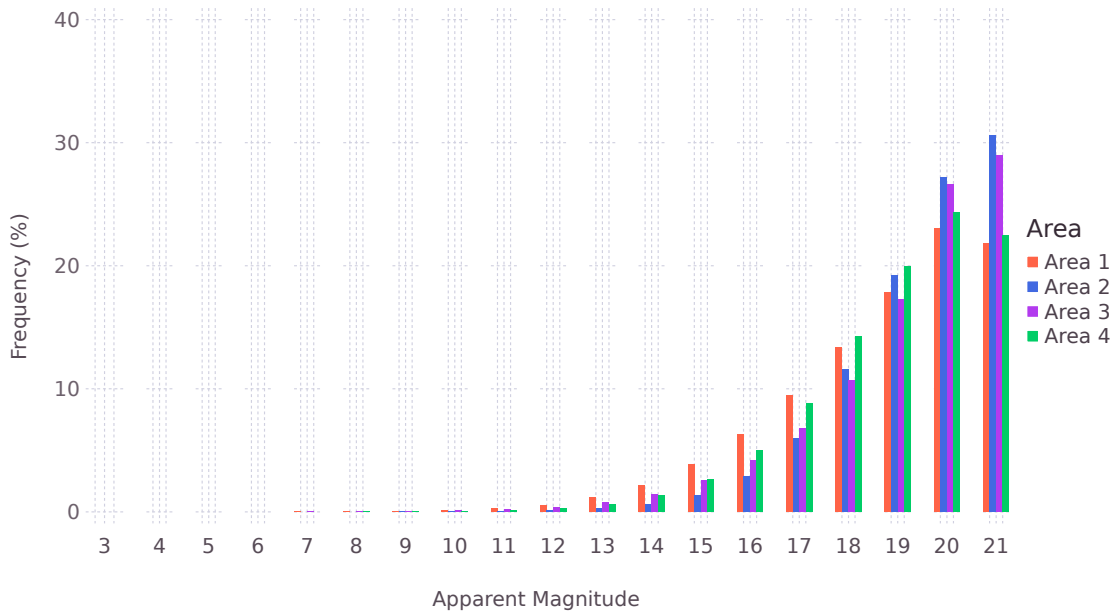
Figure 2.4: Proportional Distribution of Apparent Magnitude Across the Four Areas

Figure 2.4 shows that the areas are predominantly composed of *dim* stars. All four areas have at least 50 % of their stars being of a magnitude of 19 or over.

2.1.2 RA, Dec, and Distance (via Parallax)

The RA and Dec are provided as part of the data-set and are easy to retrieve. Together, these two quantities provide 2-dimensional spherical coordinates for a star. However, two stars that may appear to overlap in terms of their RA and Dec may be many thousands of light-years apart. Thus, to be able to effectively cluster stars it is important to determine the third dimension representing the distance from the Earth. By making use of trigonometry and the angular shifts present in parallax, astronomers were able to determine a method to estimate the distance of an object from the Earth. This distance is measured in parsecs and, as may be seen in Figure 2.5, works by virtue of our orbit around the Sun.
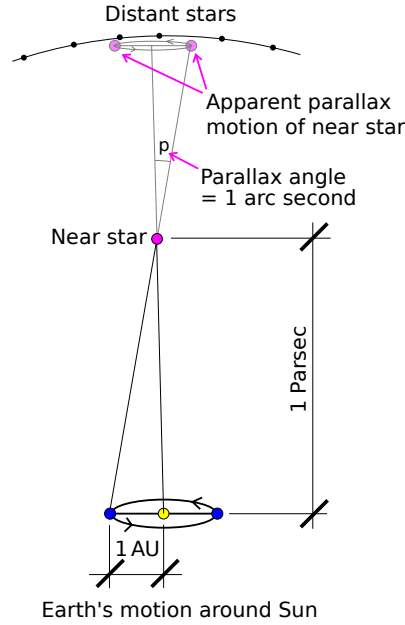
Figure 2.5: Depiction of Parallax Measurement [26]

To determine the distance to some nearby star, we measure the angular shift in the position of a distant star relative to the nearby star at opposite points in our planet's orbit. This angular shift is the parallax and using this we may compute the distance in parsecs:

$$\text{distance} = \frac{1}{\text{parallax}} \qquad (2.1)$$

The distributions of the parallax values across the four areas may be seen in the density plot in Figure 2.6.
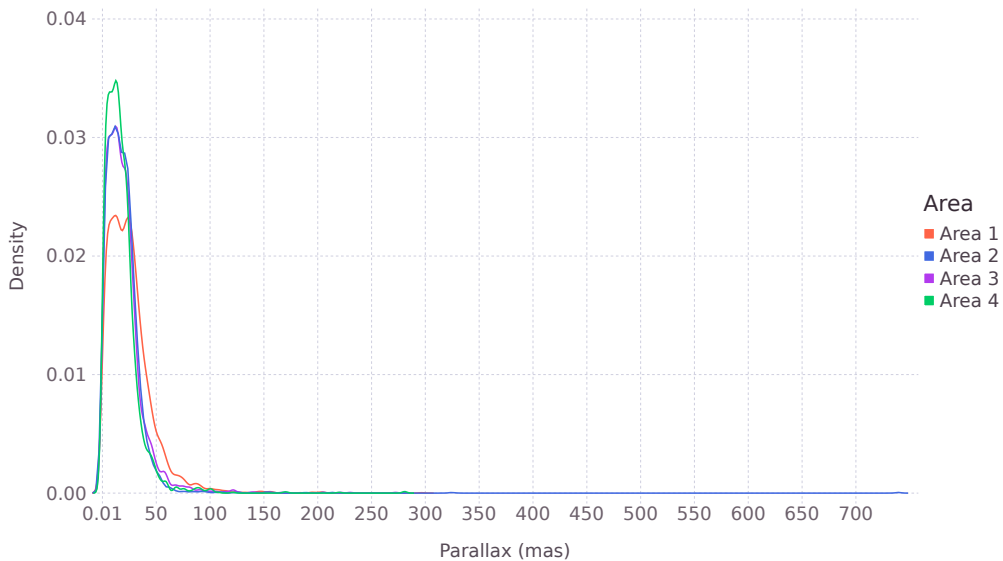
Figure 2.6: Distribution of Parallax Across the Four Areas

It is evident that the parallax values for the stars across the four regions span similar ranges and that they are all predominantly centered within 0.01 mas to 50.0 mas.

2.1.3 Proper Motion

It is also of interest to explore the spread of the data with respect to *proper motion of right ascension*, and *proper motion of declination*. Together they describe the angular shift that a star experiences over time, essentially representing the drift that the star is undergoing. In some cases this drift can be quite significant. Figure 2.7 highlights this by showcasing the apparent shift in the position of Barnard's Star across the years 1991 to 2007.
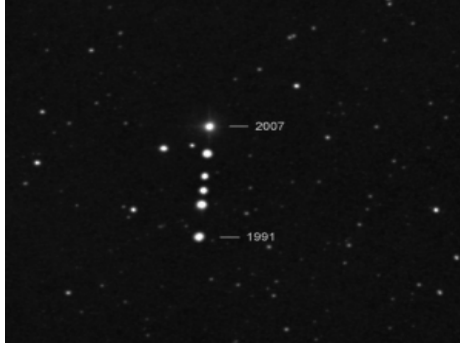
Figure 2.7: Proper Motion of Barnard's Star from 1991–2007 [27]

GCs are characterized by stars that are gravitationally bound to each other and are likely to be experiencing similar drift [4]. Thus, the proper motion values provide a useful metric in identifying potential GCs. The Gaia data-set displays unprecedented accuracy in the proper motion values, with uncertainty in the range of [24]:

- $0.06 \, \text{mas} \, \text{yr}^{-1}$ for apparent magnitudes < 15.

- $0.2 \, \text{mas} \, \text{yr}^{-1}$ for apparent magnitudes of 17.

- $1.2 \, \text{mas} \, \text{yr}^{-1}$ for apparent magnitudes of 20.

Density plots providing information on the distribution of PM_{RA} and PM_{Dec} may be seen in Figure 2.8.
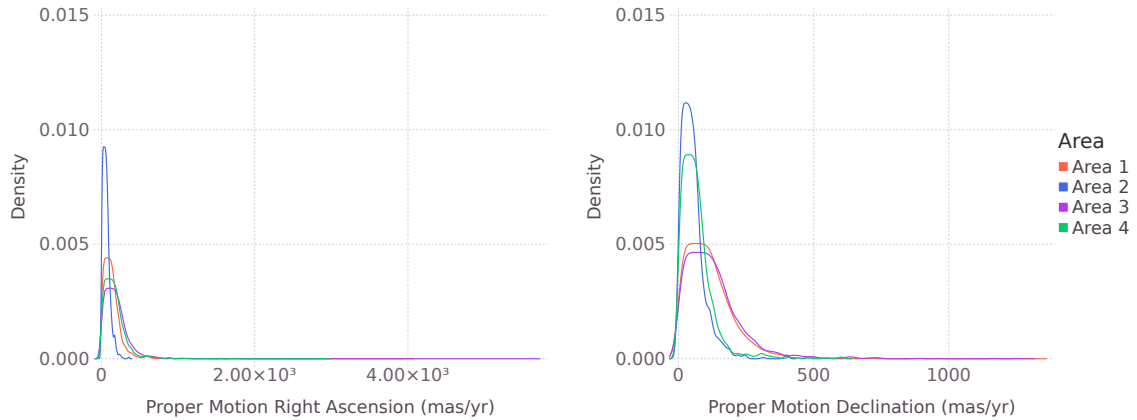
Figure 2.8: Distribution of Proper Motion Across the Four Areas

The plots show a great variety in the motion of the stars. It is evident that the motion of PM_{RA} is larger than PM_{Dec} for all areas. Additionally, the stars within Area 2 are experiencing drift within a tighter bound for both PM_{RA} and PM_{Dec} when compared to the other areas. The remaining three areas share a similar distribution in their stellar motion for RA, but Area 4 seems to have less drift in Dec than either Area 1 or Area 3. The largest movement in the direction of RA is in Area 3 with $5750 \, \text{mas} \, \text{yr}^{-1}$, and the largest movement in the Dec is in Area 1 with $1350 \, \text{mas} \, \text{yr}^{-1}$.

Chapter 3

Methodology

With each requisite parameter being described in Chapter 2, we may now describe the process of identifying GCs. The central pipeline consists of three stages:

1. Initial exclusion via blob-detection based on the Difference of Gaussian algorithm.

2. Pheromone-based density mapping using the Ant Colony random-walk algorithm.

3. Gravitational clustering using the pheromones from the previous stage to pool related stars.*

Note that, as the Ant Colony is an algorithm that incorporates *randomness*, subsequent executions on the same input data will give rise to different results. Since the gravitational clustering is also based directly on the output of the ant colony, these two portions of the pipeline are tested across multiple experiments.

However, before any of these steps may be applied, it is first necessary to rasterize the data, this subdivides it into smaller windows to operate across. This is necessary because the Any Colony algorithm functions better on smaller regions. The pipeline then provides information, per raster, on the clusters contained within that raster. An overview of this process is shown in Figure 3.1.
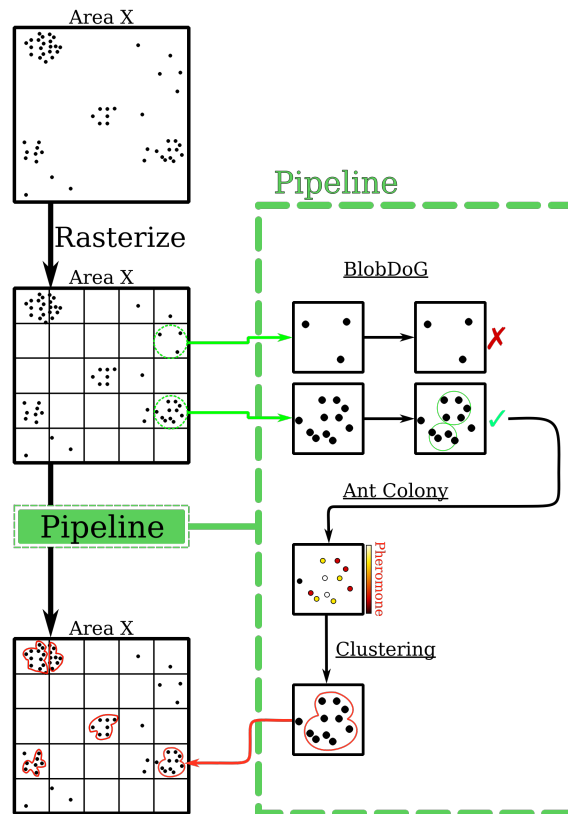
Figure 3.1: Overview of GC Identification

---

*The algorithm used for this stage was developed by the author for this paper.

3.1 Rasterization

Rasterization involves splitting an area into smaller equally sized regions. Theoretically, rasterization could occur across any set of parameters. However, in this instance the rasterization is applied across the equatorial coordinate system, and so the splitting is based on the RA and Dec.

The largest known GC spans a RA × Dec of $1.5° \times 1.5°$ [21] which defines the minimum possible bound of the rasters. In, the work of Mohammadi et al., rasters of $3.0° \times 3.0°$ were used [8]. However, to account for the extraordinary case of two globular clusters of size $1.5° \times 1.5°$ being next to each other, rasters of size $4.0° \times 4.0°$ were chosen. Note, Area 2 is an outlier due to its size and its stellar density. Thus, to explore the impact of the *denseness* of the region on the results, Area 2 is investigated with rasterization in RA × Dec of both $2.0° \times 2.0°$ and $4.0° \times 4.0°$. Figure 3.2, shows an example of rasterization applied across the Small Magellanic Cloud in Area 3.
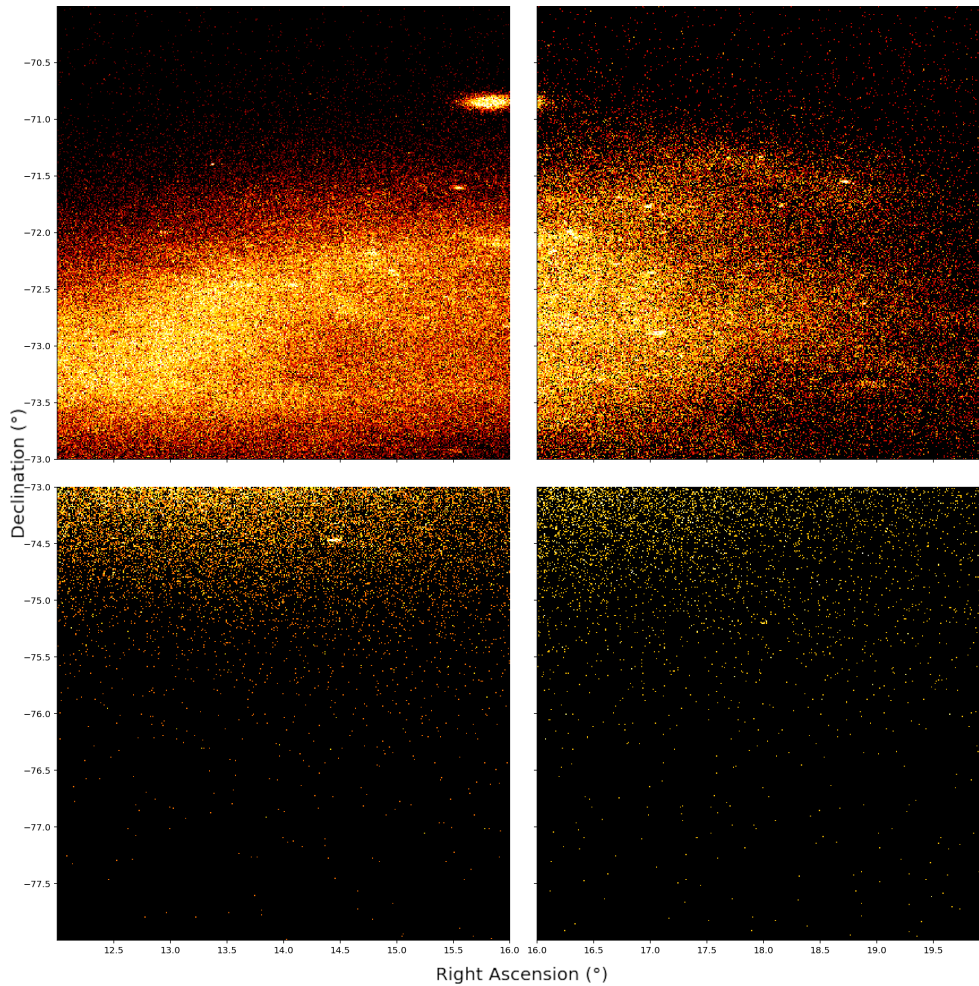
Figure 3.2: Example of Rasterization Across the Magellanic Cloud in Area 3

Rasterizing each area into smaller windows brings several benefits.

1. The Ant Colony algorithm functions better on smaller regions as it is more easily able to explore the state space (see Section 3.3 for more information).

2. The sheer size of the data-set means that a substantial amount of computer memory is required to completely load one of the four areas. The smaller regions generated by the rasterization significantly lowers the operating memory requirements.

3. Smaller regions are more likely to be empty, and get marked as such by the blob detection. This means that the more computationally expensive stages of the pipeline need to process less data.

4. Lastly, since each raster is considered independently across the future stages of the pipeline, they may be operated on in parallel. This reduces the amount of time needed to execute the pipeline.

However, this fixed rasterization scheme raises the issue of splitting a GC apart at the raster boundary. This shortcoming and possible solutions are elaborated on in Chapter 6.

3.2 Raster Exclusion Using BlobDoG

The rasterization results in many smaller regions, upon which, the central pipeline runs. However, the stellar distribution across these rasters is obviously non-uniform. Figure 3.3 displays three different scenarios which dominate the description of the rasters.

(a) Dark Region   (b) Dark Region with Bright Spot   (c) Bright Region
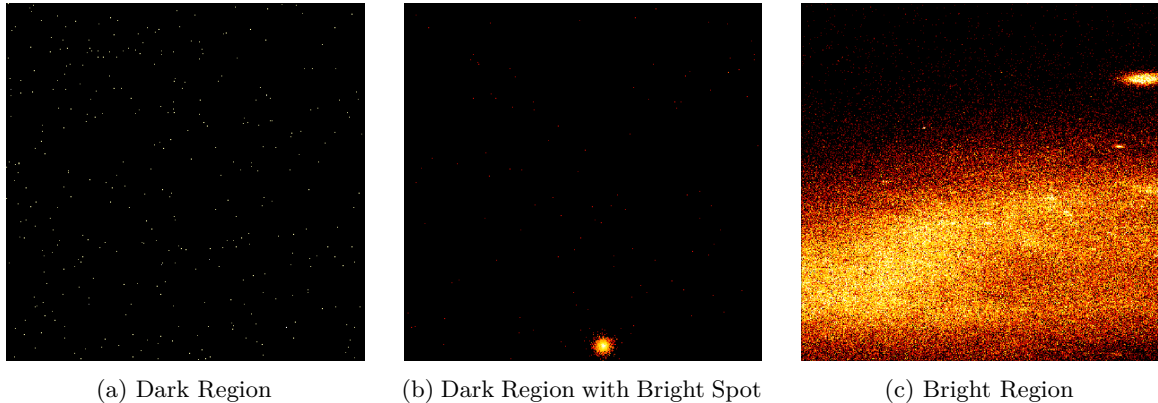
Figure 3.3: Three Types of Rasters

Given the time complexity of the Ant Colony algorithm and the clustering algorithm, it is prudent to reduce the number of rasters to be processed. This is done by filtering out the *dark* regions similar to the region presented in Figure 3.3a. Previous research by Mohammadi et al. made use of a blob-detection technique based on the Difference of Gaussian (BlobDoG) as a post-processing method [8]. However, for this research, the blob detection technique is instead applied as a pre-processing step.

The BlobDoG algorithm is provided a grayscale image and then reports information on the coordinates and size of blobs contained with that image. For a given raster, the RA, Dec, and Magnitude of stars are used to construct its grayscale representation. This results in an image where each pixel of the image has a grayscale luminance corresponding to the magnitude. BlobDoG is then applied on each of these images. Rasters for which BlobDoG reports no blobs correspond to dark regions and are filtered away.

3.2.1 Components of the BlobDoG Algorithm

Gaussian Filters

BlobDoG makes use of Gaussian filters [28]. This technique is fundamental to edge detection and involves the application of a Gaussian kernel on some 2-dimensional grayscale image $I(x, y)$ to result in a *blurred* image [29]. The Gaussian kernel used for this blurring may be seen in Eq (3.1).

$$G(x, y, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{3.1}$$

In Eq (3.1):

- x and y together, represent the coordinates of a specific pixel within the image $I(x, y)$ [30],

- σ represents the standard deviation of the Gaussian distribution, characterizing the width of the Gaussian kernel [30].

This kernel is then applied on some image via the *convolution* operation (represented by $*$), i.e.

$$I_{\text{blurred}}(x, y) = I(x, y) * G(x, y, \sigma)$$

Upon application, each pixel within the image receives a weighted average of the pixels around it. The initial weights for these surrounding pixels are determined by the Gaussian distribution parameterized by σ. The higher the σ value, the higher the weights allocated to (further-away) neighbors, which ultimately results in a stronger blur. Figure 3.4 is an example of the application of this Gaussian blurring on a well-known standard test image of Lenna Forsén [31].

(a) $I(x,y)$ (b) $I(x,y) * G(x,y,\sigma = 2.5)$ (c) $I(x,y) * G(x,y,\sigma = 5.0)$
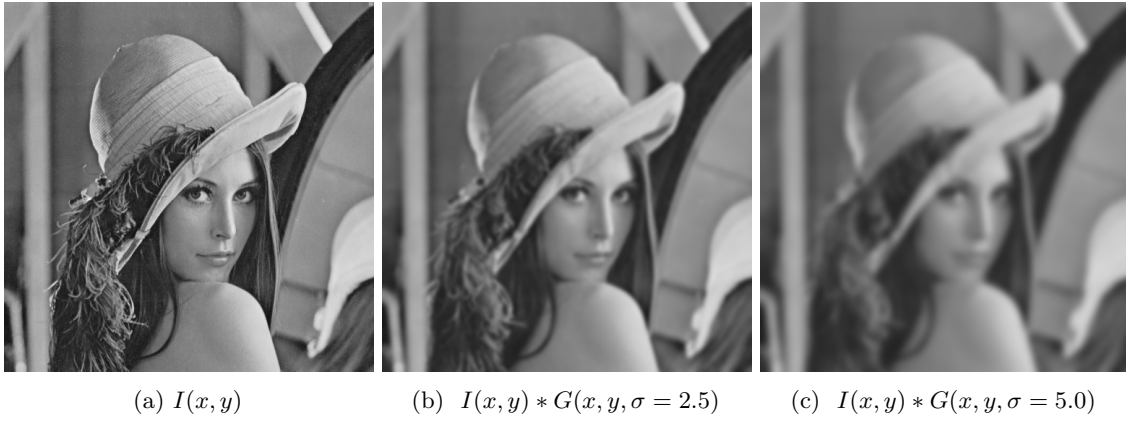
Figure 3.4: Gaussian Blurring applied on $I(x,y) =$ Lenna

Regions within the image that have a sharp contrast have a more pronounced blur, and regions that are more uniform are affected less by the blurring. This results in a blur that preserves boundaries and edges. It is this property that makes Gaussian filters effective in edge detection.

Difference of Gaussian

The technique used for edge detection based on Gaussian filters is the Difference of Gaussian (DoG). As the name implies, DoG is simply the difference between the application of two different Gaussian filters on the same image, $I(x,y)$. The formula for a DoG filter may be seen in Equation (3.2).

$$
\begin{aligned}
DoG(x,y,\sigma) &= I(x,y) * G(x,y,k\sigma) - I(x,y) * G(x,y,\sigma) \\
&= I(x,y) * (G(x,y,k\sigma) - G(x,y,\sigma)) \qquad \text{Convolution distributes over subtraction}
\end{aligned}
\tag{3.2}
$$

The two filters are differentiated by the presence of k, a constant multiplicative factor operating on the σ. As this factor increases, the strength of the blur that the corresponding Gaussian filter manifests, also increases. The difference between these two different applications of the Gaussian filters, then causes [32]:

- the more uniform regions which maintain their similarity to be inhibited,

- the more distinct regions which demonstrate a stronger blur to be excited.

Fundamentally, it may be viewed as a band-pass filter that attenuates the spatial frequencies within the original image [30]. Figure 3.5 demonstrates the application of DoG on the image of Lenna and highlights the utility of DoG for edge detection.

Figure 3.5: Difference between Gaussian Blurring of Figure 3.4b and Figure 3.4c

DoG provides a mechanism to emphasize the edges of the blobs contained within a stellar raster. By identifying these boundaries at different scales the blob detection is able to report on the existence of blobs within a raster.

Blob Detection

The implementation of BlobDoG which was used for this paper comes from the Python library `SciKit` [33]. Their implementation is based on the work of Lowe [28] and uses DoG to evaluate the

frequency response of each pixel within an image at different scale-factors. The values for these scale-factors are proportional to the k present in Eq (3.2). The subsequent applications of DoG at these different scale-factors form a cube of sorts [28], a depiction of which, may be seen in Figure 3.6.

(a) Pixel-by-Pixel Evaluation of BlobDoG Cube   (b) Maximal Frequency Response
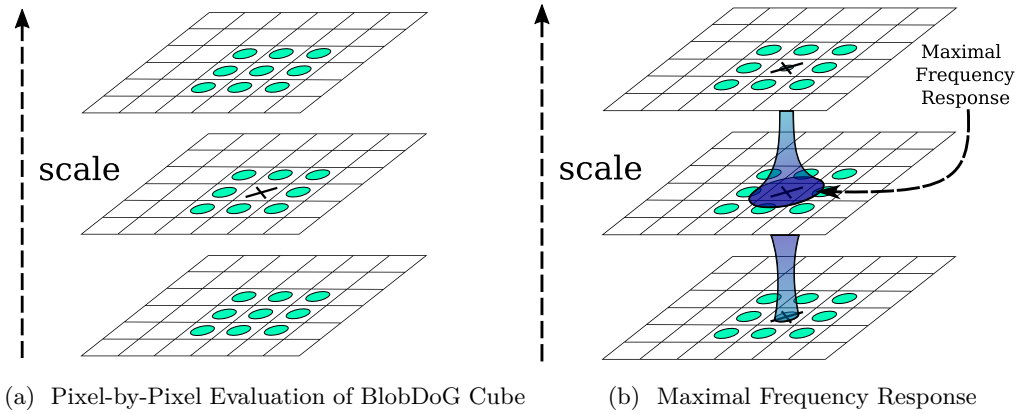
Figure 3.6: Application of BlobDoG

The cross within Figure 3.6a represents the current pixel being evaluated. Its maximal frequency response is determined by comparing its response, against the 26 surrounding neighbors represented by the circles [28]. The scale that results in the overall maxima for a pixel corresponds to the size of the largest cohesive *blob* that contains that pixel. The radius of the resulting blob equates to [33]:

$$radius = \sqrt{2} * scale \tag{3.3}$$

The process in Figure 3.6b is akin to finding the right focus for a pair of binoculars when looking at some specific point. The clarity of the image is analogous to the frequency response. As the frequency response increases, the clarity of the image increases. This continues until the maximal clarity is reached for that point at some specific focus level (the scale). Past this point the frequency response decreases and the clarity of the image begins to go down. This essentially identifies the scale at which each point becomes of interest.

Finally, the implementation uses a threshold to prune away blobs whose maximal frequency is deemed too low, as this corresponds to a blob that is too small [33]. This threshold is represented by $B_{threshold}$ and its value is elaborated on in Section 3.5.

3.2.2   Interpreting the Results of BlobDoG

BlobDoG provides information on the position and scale of any blob that is sufficiently large (based on $B_{threshold}$). If no blobs are detected then that raster may be filtered away. Using Eq (3.3) the radii of any blobs are determined and this is used to plot the results. Figure 3.7 shows the results of BlobDoG when applied on the exemplar rasters that were shown in Figure 3.3.

(a) Dark Region   (b) Dark Region with Bright Spot   (c) Bright Region
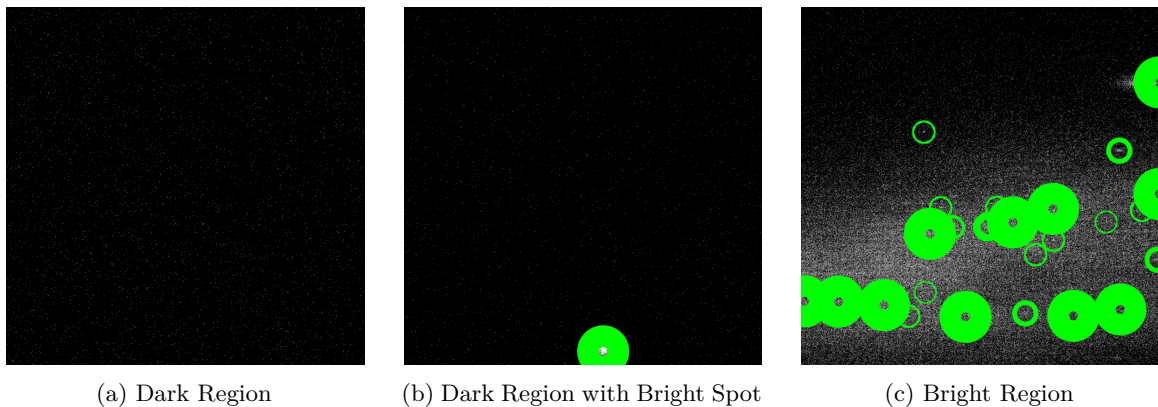
Figure 3.7: DoG Applied to the Three Types of Rasters

Since Figure 3.7a contains no blobs, its corresponding raster will be filtered away. Figure 3.7b and Figure 3.7c contain one or more blobs and their rasters should be processed by the future stages of the pipeline.

3.3 Density Mapping Via the Ant Colony

With the initial filtration handled by BlobDoG, what remains is the task of identifying potential GCs in the rasters that have persisted. It is difficult to precisely encode the properties of GCs such that an exhaustive search algorithm could identify them in a timely fashion. Instead we aim to produce density mappings and cluster the densest regions in a fashion similar to what was done in the previous work by Mohammadi et al. [8].

Since GCs are an agglomerate structure, when identifying them, it is necessary to consider the attributes of individual stars, as well as, the relationship between stars. To this end, it is possible to re-frame the scenario and represent the stars on a graph, with the stars representing the nodes and the relationships between the stars represented as edges. Encoding these edges with weights based on the similarities between the pairs of stars provides a foundation for the use of the *Ant Colony random-walk* algorithm.

This algorithm is based on the swarm behavior that manifests in biological ant colonies [34]. A swarm of ants explores a region by (randomly) discovering the shortest routes to a desirable resource and then leaving *pheromone* trails for other ants to follow. As other ants follow this path and also discover this same resource, the pheromone trail becomes stronger, continuing to attract more and more ants until the resource is consumed. These pheromone values naturally decay over time and allow the ant colony to dynamically optimize their routes based on their environmental context [34].

After the ants have been provided a sufficient amount of time to explore, these pheromone values will result in a random path that approximately describes the network substructure of their environment [35]. The strength of these pheromone trails provide us the density mapping which will be used to cluster candidate GCs (as described in Section 3.4).

3.3.1 Components of the Ant Colony Algorithm

The core algorithm for the Ant Colony may be seen in Alg. 1. Note that the values that were selected for the constants N_{gens}, N_{ants}, and N_{steps} may be found in Section 3.5.

Algorithm 1 Ant Colony [36]

Output: The pheromone vector: $\mathbf{f} = [f_1, f_2, \ldots, f_{N_{\text{stars}}}]$

$\mathbf{f}^0 = [0, 0, \ldots, 0]$ ▷ Length of N_{stars}.
for $t = 1, \ldots, N_{\text{gens}}$ **do**
    $\mathbf{N}^{t-1}_{\text{visitations}} = [0, 0, \ldots, 0]$ ▷ Length of N_{stars}.
    for ant $= 1, \ldots, N_{\text{ants}}$ **do**
        star $\leftarrow$ Random initial position for the ant, based on a uniform random distribution.
        $\mathbf{N}^{t-1}_{\text{visitations}}$ [star] += 1
        for step $= 1, \ldots, N_{\text{steps}}$ **do**
            star $\leftarrow$ Random next position for the ant, weighted by the transition probability: Eq. (3.6) ▷ This uses star (the current position) and $\mathbf{f}^{t-1}$.
            $\mathbf{N}^{t-1}_{\text{visitations}}$ [star] += 1
    Update $\mathbf{f}^{t-1}$ to $\mathbf{f}^t$ using the pheromone update function: Eq. (3.7) ▷ This uses t, $\mathbf{f}^{t-1}$, $\mathbf{N}^{t-1}_{\text{visitations}}$, N_{ants}, and N_{steps}.
return $\mathbf{f}^{N_{\text{gens}}}$

Before any ants can take their first steps, the feature vector $\mathbf{f}$ containing a pheromone value for each star is initialized. The main logic of the algorithm is comprised of a triple nested loop. The outermost loop represents an iteration over a number of generations (N_{gens}). Each subsequent generation is able to make use of the information from the previous generation thereby thoroughly exploring the state space and allowing for more accurate, average pheromone values. This outer loop executes three steps:

1. The initialization of $\mathbf{N}_{\text{visitations}}$ keeps track of the number of visits to each star.

2. The inner loop where a number of ants corresponding to N_{ants} are set loose randomly within the stellar raster.

3. The transitioning of the pheromone values between each experiment using the pheromone update function described by Eq. (3.7).

The inner loop iterates across each ant and executes the random-walk for that ant. First, the ant is spawned at a random starting position. The visitation for that starting position is updated and then the ant proceeds to walk around for a number of steps (corresponding to a total of N_{steps}). This is represented by the innermost loop, where the ant determines its next step in a random fashion

biased by the distribution of pheromones in its neighborhood. This allows it to determine its best path based on the knowledge left behind by previous ants. This step is governed by Eq. (3.6).

Finally, after all the ants have performed their random-walks, the pheromone values for that experiment may be updated. The updating of the pheromone values marks the end of one experiment, after which, the next experiment may begin. From this description, it is evident that the core of the Ant Colony algorithm is fairly simple. However, its strength hinges on the transition probability driving the core decision making process of the ants.

To compute this transition probability it is necessary to compute additional information on the relationship between stars. However, it is not feasible to compute the relationship between all pairs of stars due to the sheer number of stars. Additionally, it is not even useful to compute this data for stars that are too far apart. Thus, from a given star, x_i an ant will only consider a limited amount of neighboring stars ($N_{\text{neighbors}}$). These neighboring stars correspond to the stars that are closest by Euclidean distance. This distance is based on the RA, Dec, and Distance and for two stars x_i and x_j is computed as follows:

$$
\text{Euclidean Distance Between } x_i \text{ and } x_j = \left\{ \left(\text{RA}_{x_i} - \text{RA}_{x_j} \right)^2 + \right.
$$
$$
\left(\text{Dec}_{x_i} - \text{Dec}_{x_j} \right)^2 + \tag{3.4}
$$
$$
\left. \left(\text{Distance}_{x_i} - \text{Distance}_{x_j} \right)^2 \right\}^{\frac{1}{2}}
$$

For the star x_i, the list of its nearest neighbors is represented as $\mathcal{N}_{x_i}$ and this list is assumed to be sorted in ascending order by Euclidean distance from x_i.

3.3.2 Transition Probability

The first step in computing the transition probability is to compute the *weights* between each star and its nearest neighbors. This weight should be proportional to how rewarding it would be for an ant to make a transition from its current star to some neighboring star. For two stars, x_i and x_j, their weight is represented by $w(x_i, x_j)$. Since the main aim is to identify stars with certain similar features that may make them part of the same GC, the weight must constitute the similarity between a pair of stars. Stars within GCs are relatively near to each other and manifest similar motion characteristics. Thus, the parameters used to compute the weights are:

- the RA, Dec, and Distance as these parameters encode the position,

- the PM_{RA} and PM_{Dec} as these parameters encode the motion.

Computing the Weights

The algorithm used to compute the weights is as follows:

Algorithm 2 Computing Weights Between All Stars and Their Neighbors [33]

Initialization: The *stars* where for each star, x_i, the following are accessible:

- The relevant parameters of x_i: RA_{x_i}, Dec_{x_i}, Distance_{x_i}, $\text{PM}_{\text{RA} x_i}$, and $\text{PM}_{\text{Dec} x_i}$
- The ($N_{\text{neighbors}}$) nearest neighbors of x_i: $\mathcal{N}_{x_i}$

Output: A lookup table for the *weights*, w, where the weight from x_i to x_j may be looked-up via $w(x_i, x_j)$.

$$
w = \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_{N_{\text{stars}}} \end{array} \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}
$$
▷ Initialize the lookup table for the weights to 0.

for $x_i \in$ stars **do**

$$
D = \begin{array}{c} \mathcal{N}_{x_i}[1] \\ \mathcal{N}_{x_i}[2] \\ \vdots \\ \mathcal{N}_{x_i}[N_{\text{neighbors}}] \end{array} \begin{pmatrix} \text{RA}_{\mathcal{N}_{x_i}[1]} & \text{Dec}_{\mathcal{N}_{x_i}[1]} & \text{Distance}_{\mathcal{N}_{x_i}[1]} & \text{PM}_{\text{RA}\mathcal{N}_{x_i}[1]} & \text{PM}_{\text{Dec}\mathcal{N}_{x_i}[1]} \\ \text{RA}_{\mathcal{N}_{x_i}[2]} & \text{Dec}_{\mathcal{N}_{x_i}[2]} & \text{Distance}_{\mathcal{N}_{x_i}[2]} & \text{PM}_{\text{RA}\mathcal{N}_{x_i}[2]} & \text{PM}_{\text{Dec}\mathcal{N}_{x_i}[2]} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{RA}_{\mathcal{N}_{x_i}[N_{\text{neighbors}}]} & \text{Dec}_{\mathcal{N}_{x_i}[N_{\text{neighbors}}]} & \text{Distance}_{\mathcal{N}_{x_i}[N_{\text{neighbors}}]} & \text{PM}_{\text{RA}\mathcal{N}_{x_i}[N_{\text{neighbors}}]} & \text{PM}_{\text{Dec}\mathcal{N}_{x_i}[N_{\text{neighbors}}]} \end{pmatrix}
$$
▷ Assign the parameters from each neighbor to the feature matrix, D.

$\overline{D} = D - \mu(D)$ ▷ Center the data by subtracting column-wise means so the mean of each column is 0.

$U, \Sigma, V^{\mathsf{T}} = \text{SVD}(\overline{D})$ ▷ Get the singular values and singular vectors from the Singular Value Decomposition.

$\widehat{D} = \dfrac{\overline{D}}{\|\overline{D}\|}$ ▷ Compute the correlation coefficient for each star.

$\text{weights}_{\mathcal{N}_{x_i}} = |\widehat{D} * V| * \Sigma$ ▷ $\text{weights}_{\mathcal{N}_{x_i}}$ is of length $N_{\text{neighbors}}$ and has one element for each neighbor of x_i.

Insert $\text{weights}_{N_{x_i}}$ into row for x_i in w

return w

The aim in the weight initialization is to compute a single metric capable of uniformly representing the differences between the 5 different parameters across a pair of stars. This problem requires a form of dimensionality reduction which is done using Singular Value Decomposition (SVD) [37]. SVD may be used to generate an orthogonal projection to the original dimensions which minimizes the mean-squared distance of the observed points from the new basis axes. In effect this produces a linear combination of the original parameters which maintains the maximal variance of the constituent dimensions [38].

Computing the Transition Probability

With the weights in place the ants have an initial basis for forming their decisions. However, the ants should also take into account the pheromone trails left by previous ants. By combining the underlying weights with the previous pheromone values the ants may determine a probability for transitioning to the neighboring stars [36].

$$\hat{f}^t(x_j) = \frac{f^t(x_j)}{\sum\limits_{x_k \in \mathcal{N}_{x_i}} f^t(x_k)} \qquad \text{(3.5)}$$

Normalize the pheromone value of x_j (a neighbor of x_i) based on the neighborhood of x_i.

$\hat{f}^t(x_j)$ is the normalized pheromone value of x_j at generation t. It is normalized by evaluating $f^t(x_j)$ (the pheromone value of x_j at generation t) and then dividing it by the sum of the pheromone values of the all the stars in the neighborhood. The transition probability between x_i and x_j is then simply computed by mixing the weight between x_i and x_j and the normalized pheromone value of x_j in the previous generation.

$$P^t(x_i, x_j) = \frac{\left(w(x_i, x_j)\right)^\gamma \left(\hat{f}^{(t-1)}(x_j)\right)^{1-\gamma}}{\sum\limits_{x_k \in \mathcal{N}_{x_i}} \left(w(x_i, x_k)\right)^\gamma \left(\hat{f}^{(t-1)}(x_k)\right)^{1-\gamma}} \qquad \text{(3.6)}$$

Compute the transition probability of going from x_i to x_j.

$P^t(x_i, x_j)$ is the transition probability from x_i to x_j computed for generation t. In the equation, γ is a control parameter that simply determines the relative impact of the weights versus the pheromone values (see Section 3.5 for information on the value selected for γ). The range for γ is $0 \leq \gamma \leq 1$.

- $\gamma = 0$ describes a scenario where the pheromone values are all that are involved in determining the transition probability.

- $\gamma = 1$ describes a scenario where the weights are the only factor contributing to the transition probability.

Finally the result is then normalized through dividing by the summation of the equation applied in the numerator across the whole neighborhood. This brings the transition probability within the range of [0, 1] and ensures that the sum of all the transition probabilities in the neighborhood is 1.

3.3.3 Updating Pheromone Values

With the ants able to compute their transition probabilities, all that remains is to update the pheromone values between generations. The equation to compute the pheromone values for generation t from $t-1$ is as follows [36]:

$$f^{(t)}(x_i) = \underbrace{\frac{\mathbf{N}_{\text{visitations}}^{t-1}[x_i]}{N_{\text{ants}} \times N_{\text{steps}}}}_{\text{Fresh Pheromone Trails}} + \underbrace{(1-\rho)f^{(t-1)}(x_i)}_{\substack{\text{Preexisting Pheromone} \\ \text{Trails}}} \qquad \text{(3.7)}$$

For the fresh pheromone trails, $N_{\text{ants}} \times N_{\text{steps}}$ corresponds to the total number of visitations made by the ants across a single generation. Thus, $\frac{\mathbf{N}_{\text{visitations}}^{t-1}[x_i]}{N_{\text{ants}} \times N_{\text{steps}}}$ is simply the proportion of the visits that were made to x_i of the total number of visitations for that generation. These fresh pheromone trails are then combined with the preexisting pheromone trails. ρ is within the range of $0 \leq \rho \leq 1$ (see Section 3.5 for the value) and controls the evaporation of these preexisting pheromone values [36].

3.3.4 Interpreting the Results of the Ant Colony Algorithm

The results of the algorithm are a pheromone vector $\mathbf{f}$ which contains the pheromone values left by the ants after traversing the raster. Figure 3.8 plots the resulting pheromone heat-maps of the Ant Colony algorithm for the exemplar rasters shown in Figure 3.3.

(a) Dark Region   (b) Dark Region with Bright Spot   (c) Bright Region
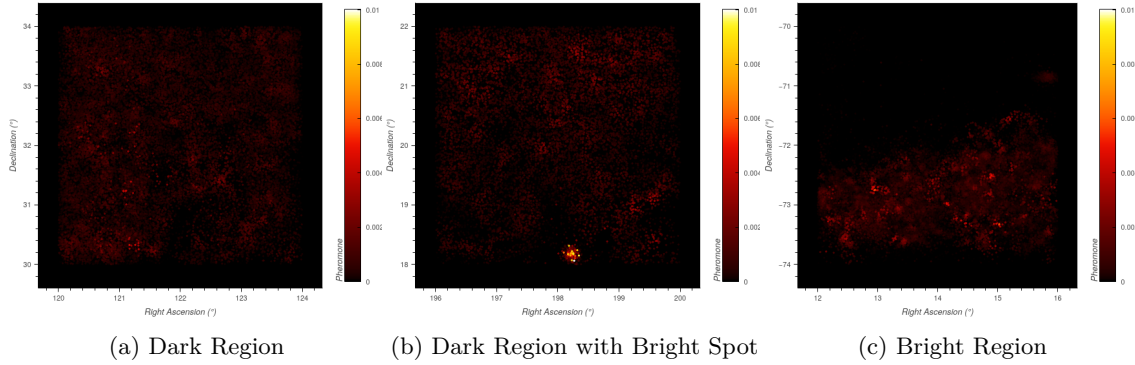
Figure 3.8: Results of the Ant Colony Applied to the Three Types of Rasters

The brighter stars (yellow-white) correspond to stars with higher pheromone values than the darker stars (red-black). The rasters depicted in Figure 3.8a and Figure 3.8c are both relatively uniform with the first being mostly devoid of stars and the second being completely saturated with stars. As a result, it seems that the ants have not highlighted any major region of interest. However, for Figure 3.8b the ants have been able to pinpoint the cluster that is present within the raster. While, this cluster is apparent by eye alone, it is not always so evident from a pheromone heat-map that a cluster has been identified.

As a result, it is necessary to perform further processing to effectively extract potential clusters from the pheromone values. It is expected that the pheromone values within clusters will be high and the values outside of (and between clusters) will be low. This provides a boundary separating a cluster from other stellar structures. This forms the basis of the clustering algorithm.

3.4 Gravitational Clustering Based on Pheromone Mapping

The pheromone values encode the plethora of information related to the RA, Dec, Distance, PM_{RA}, and PM_{Dec} of the stars. They provide a strong heuristic for identifying the boundaries between dense regions of similar stars. The next step is to then perform clustering such that stars with high pheromone values are grouped together. While the raw pheromone values do provide information on the density of the stars, it is quite possible for two stars in different but equally dense regions to have the same pheromone values. Thus, the clustering algorithm must also make use of the position of the star coupled with the pheromone values to determine which cluster a star may belong to.

To explain the clustering algorithm, it is necessary to reformulate the problem at hand. There are a number of points distributed in 3-dimensional space with a (pheromone) parameter attached to them that describes the strength of the attraction they feel to other points. In addition, points that are closer to each other should experience a stronger attractive force than points that are farther apart. This formulation reveals the similarity of this problem to the basic description of gravitational attraction. Newton's equation which governs gravitational attraction is as follows:

$$F_{\text{gravity}} = g \frac{M_{x_i} M_{x_j}}{r^2} \tag{3.8}$$

In Eq. (3.8), g represents the gravitational constant, M_x represents the mass of some object x, and r represents the distance between the *center-of-gravity* (CoG) of x_i and x_j. The clustering algorithm uses this as its foundation but uses the pheromone values in place of the mass to determine the attraction between the stars. Stars which are sufficiently attracted to each other will pool to form a cluster which in turn may continue to grow and attract more stars.

3.4.1 Components of the Clustering Algorithm

In the pheromone-based gravitational clustering, an initial cluster will form out of single star. Stars within the field of attraction of this cluster will be absorbed. The strength of this attraction is based on the pheromone mass of both the cluster and the star and is defined as follows:

$$P_{x_i} = \mathbf{f}[x_i] \qquad \text{\textit{The pheromone mass for a single star.}} \tag{3.9}$$

$$P_C = \sum_{x_k \in C} \mathbf{f}[x_k] \qquad \text{\textit{The pheromone mass for a cluster.}} \tag{3.10}$$

The force of attraction between the cluster (C) and the star (x_i) is then defined as follows:

$$F_{\text{pheromone}} = \frac{P_C P_{x_i}}{r^2} \qquad \begin{array}{l} \text{\textit{The pheromone attraction force between}} \\ \text{\textit{a cluster } C \text{ \textit{and a star} } x_i.} \end{array} \tag{3.11}$$

In Eq. (3.11):

- P_C represents the pheromone mass of the cluster C.

- P_{x_i} represents the pheromone mass of some star x_i.

- r represents the Euclidean distance between the CoG of the cluster C and the star x_i.

This is effective for stars with a pheromone value > 0. However, across a given raster there may be stars that were never visited by the ants. These stars would have a pheromone value $= 0$ and even if they were geometrically contained within a cluster, by Eq. (3.11) they would have an attraction force equal to zero. So, stars with a non-zero pheromone value should be clustered differently than stars with a zero pheromone value. Thus, the basic algorithm for the pheromone clustering, Alg. 3, is composed of these two steps as well as a final filtration step based on a minimum accepted number of stars contained within a GC, $N_{\text{GC}_{\text{min}}}$ (see Section 3.5 for the value that was selected).

Algorithm 3 Pheromone Clustering

Initialization:
- The stars where for each star, x_i the following are available: RA_{x_i}, Dec_{x_i}, and $Distance_{x_i}$.
- The pheromone values for the stars, $\mathbf{f}$ where for some star, x_i its pheromone value may be accessed by $\mathbf{f}[x_i]$.

Output: The set of clusters present across the stars.

$(\text{stars}_{\mathbf{f}=0}, \text{stars}_{\mathbf{f}\neq 0}) \leftarrow$ Partition stars by $\mathbf{f}$ ▷ Stars with a zero pheromone value must be processed separately.

$\text{clusters}_{\text{initial}} \leftarrow$ Cluster the stars with non-zero pheromone values using Alg. 4   ▷ This uses $\text{stars}_{f \neq 0}$.

$\text{clusters}_{\text{all}} \leftarrow$ Cluster the stars with zero pheromone values using Alg. 5   ▷ This uses $\text{clusters}_{\text{initial}}$ and $\text{stars}_{f=0}$.

$\text{clusters}_{\text{filtered}} \leftarrow$ Filter out clusters from $\text{clusters}_{\text{all}}$ that contain fewer than $N_{\text{GC}_{\text{min}}}$ stars

return $\text{clusters}_{\text{filtered}}$

To perform the clustering using the non-zero pheromone values it is necessary to compute the center-of-gravity based on these pheromone values. The equation for this is as follows:

$$\text{CoG}_C = \frac{\sum\limits_{x_i \in C}\left(\text{position}_{x_i} \times P_{x_i}\right)}{|C| \times \sum\limits_{x_i \in C} P_{x_i}} \tag{3.12}$$

The CoG_C is the average position of the stars in C where the contribution of each star is weighted by their pheromone mass. This provides a coordinate in $(\text{RA}, \text{Dec}, \text{Distance})$ which may then be used as part of the non-zero pheromone clustering.

Algorithm 4 Non-Zero Pheromone Clustering

Initialization: The stars with non-zero pheromone values ($\text{stars}_{f \neq 0}$)
Output: The clusters from $\text{stars}_{f \neq 0}$

$\text{clusters} \leftarrow$ Create an empty set to contain the clusters that are found
$\text{stars}_{\text{to process}} = \text{stars}_{f \neq 0}$
while $\text{stars}_{\text{to process}}$ is not empty **do**
    $C \leftarrow$ Create an empty cluster
    $\text{star} \leftarrow$ Pop the star with the greatest pheromone value from $\text{stars}_{\text{to process}}$
    Insert star in C
    repeat
        $\text{cluster}_{\text{changed}} = \texttt{false}$
        for all $x_i \in \text{star}_{\text{to process}}$ **do**
            $\text{CoG}_C \leftarrow$ Compute the center-of-gravity of C using Eq. (3.12)
            $r \leftarrow$ Compute the Euclidean distance between CoG_C and the position of x_i
            $F_{\text{pheromone}} = \dfrac{P_C P_{x_i}}{r^2}$ ▷ Based on Eq. (3.11).
            if $F_{\text{pheromone}} \geq F_{\text{min attraction}}$ **then**
                Remove x_i from $\text{stars}_{\text{to process}}$
                Insert x_i in C
                $\text{cluster}_{\text{changed}} = \texttt{true}$
    until $\text{cluster}_{\text{changed}} == \texttt{false}$
    Insert C in clusters
return clusters

In this algorithm, every star under consideration is added to a list of stars to be processed. While stars remain to be processed, a cluster is generated and populated with the star with the highest pheromone mass of the stars that remain. Then this cluster is grown within the `repeat-until` loop until the cluster ceases to grow. Within this loop all the stars that remain have their position compared with the center-of-gravity of the cluster.

The resulting distance is then used alongside the pheromone mass of the star and the cluster to compute the pheromone attractive force ($F_{\text{pheromone}}$). If this force is sufficiently high the star is absorbed into the cluster, thereby increasing the pheromone mass of the cluster and shifting its center-of-gravity. These clusters are generated and grown repeatedly in this fashion until every star has been assigned to a cluster (even if it is simply a cluster containing that one star).

After this algorithm has run all potential clusters are now found. It is now necessary to incorporate the stars with the zero pheromone values whose positions overlap with the clusters. In essence these stars correspond to objects without a *gravitational* field. Thus, it is necessary to consult the centroid of the clusters instead of their center-of-gravity.

$$\text{centroid}_C = \frac{\sum\limits_{x_i \in C} \text{position}_{x_i}}{|C|} \tag{3.13}$$

The centroid is the average position of the stars in C. With this, the zero pheromone clustering is:

Algorithm 5 Zero Pheromone Clustering

Initialization:
    • The stars with zero pheromone values ($\text{stars}_{f=0}$)
    • The clusters based on the initial non-zero clustering from Alg. 4 ($\text{clusters}_{\text{initial}}$)
Output: The clusters incorporating $\text{stars}_{f=0}$ across $\text{clusters}_{\text{initial}}$

$\text{clusters}_{\text{all}} = \text{clusters}_{\text{initial}}$
for all $x_i \in \text{stars}_{f=0}$ **do**
    $\text{clusters}_{\text{closest}} \leftarrow \text{clusters}_{\text{all}}$ sorted in ascending order by Euclidean distance from centroid of cluster to x_i
    for all $\text{cluster} \in \text{clusters}_{\text{closest}}$ **do**
        if $\begin{aligned} &C_{\text{RA}_{\min}} \leq \text{RA}_{x_i} \leq C_{\text{RA}_{\max}} \wedge \\ &C_{\text{Dec}_{\min}} \leq \text{Dec}_{x_i} \leq C_{\text{Dec}_{\max}} \wedge \\ &C_{\text{Distance}_{\min}} \leq \text{Distance}_{x_i} \leq C_{\text{Distance}_{\max}} \end{aligned}$ **then** ▷ Check if x_i is within the bounds of the cluster.
            Insert x_i into the cluster
            break Out of inner loop
return $\text{clusters}_{\text{all}}$

In the zero pheromone clustering, each star in the set of zero pheromone stars is considered. The preexisting clusters are then searched through in order of closest to farthest based on the Euclidean distance between the centroid of the cluster and the position of the star. The star is then absorbed by the first cluster that has the star within its bounds. If no cluster contains the star it is simply left unclustered.

3.4.2 Interpreting the Results of the Clustering

With all these steps in tow, the results are a set of clusters each of which contain a set of stars. The information of these stars within the clusters may be plotted alongside the rest of the stars to provide a visualization of the clusters contained within a raster. Of the three exemplar rasters, only the dark region with the bright spot shown in Figure 3.3b is identified as containing clusters. A 2-dimensional plot of the pheromone heat-map and the clustering of the raster may be seen in Figure 3.9.
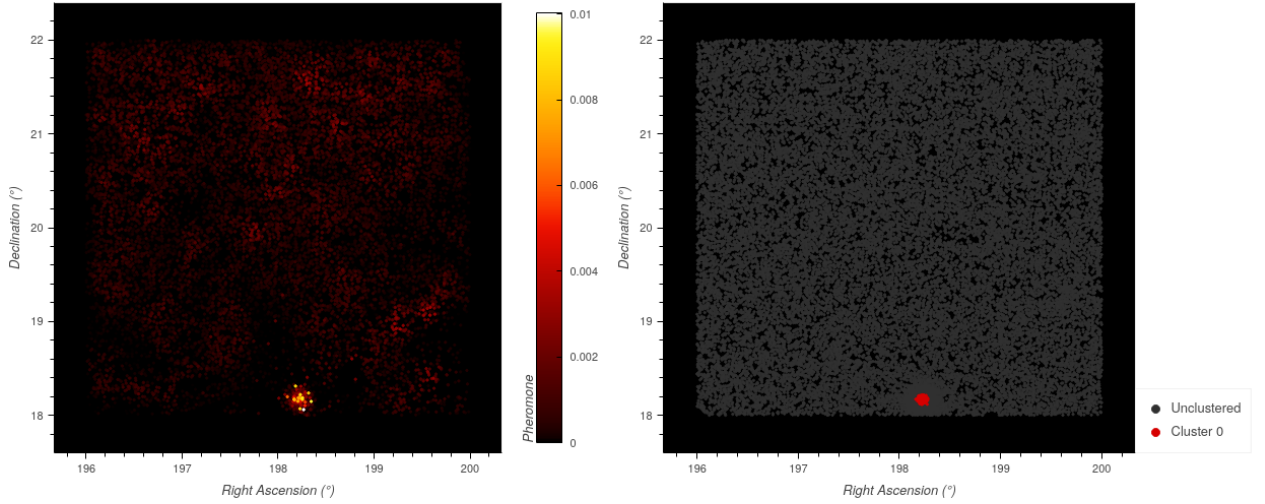
Figure 3.9: Clustering of Dark Region with Bright Spot

Additionally, in this instance only one cluster is identified. However, it is possible for multiple clusters to be involved and for these clusters to overlap in RA and Dec. To provide visualization in such cases 3-dimensional plots have also been used, such as the plot in Figure 3.10 which focuses the region containing the GC.
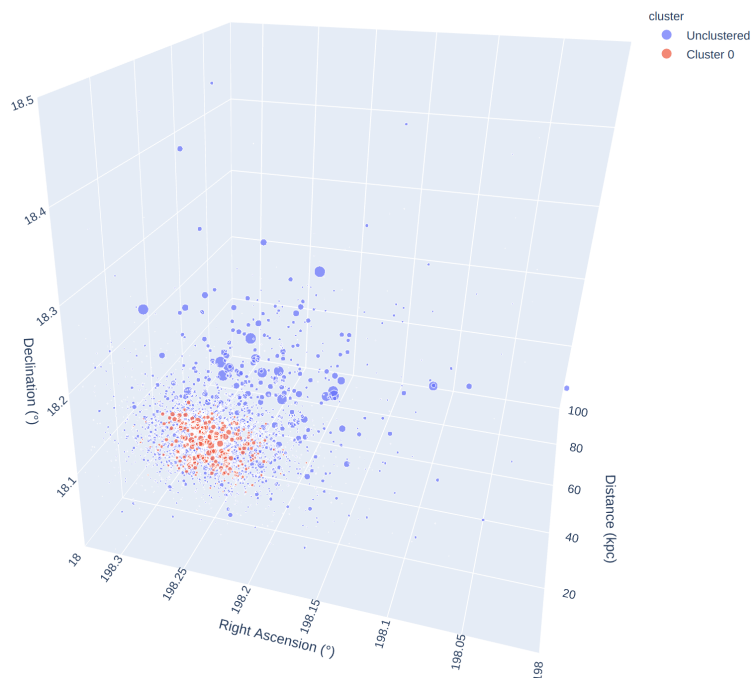

Figure 3.10: 3D Plot of Cluster in Dark Region with Bright Spot

3.5 Constants

There are a variety of constants whose values impact the functioning of the pipeline. They are presented in Table 3.1 with their descriptions, their values, and how those values were determined.

Table 3.1: All Constants

Name	Value	Algorithm	Explanation
$B_{\text{threshold}}$	0.2	BlobDoG	This threshold represents the smallest blob scale that should be considered a valid blob. The value for this constant was determined through experimentation described in Table 4.1.
N_{gens}	5	Ant Colony	Number of generations in the execution of the Ant Colony algorithm. Each generation incorporates the results of the previous generation to uncover network substructure. This value was set based on work by Mohammadi on the Ant Colony algorithm [36]. Future experimentation should be done to optimize this value.
N_{ants}	30	Ant Colony	Number of ants exploring a given raster across a single generation. This value was set based on work by Mohammadi on the Ant Colony algorithm [36]. Future experimentation should be done to optimize this value.
N_{steps}	2000	Ant Colony	Number of steps each ant is allowed to take across a given generation. This value was set based on work by Mohammadi on the Ant Colony algorithm [36]. Future experimentation should be done to optimize this value.
$N_{\text{neighbors}}$	20	Ant Colony	The number of neighbors an ant should consider when transitioning between stars during one step. This is an optimization taking advantage of the fact that the ants' choice of their next star is weighted in part by the distance to that star from the current star. Setting this parameter too low will prevent the ants from spontaneously considering stars further afield. Setting it too high will simply increase the time taken for the execution of the algorithm. This value was determined through trial-and-error and should be experimented on further to identify the lowest bound.
γ	0.9	Ant Colony	This value represents a ratio in the mixture of the pre-existing pheromone values from previous generations and the results from the current generation. At $\gamma = 1$, the ants only incorporate the results of the current generation in determining their next transition. At $\gamma = 0$, the ants only incorporate the results from the previous generation in determining their next transition. This value was set based on work by Mohammadi on the Ant Colony algorithm [36]. Future experimentation should be done to optimize this value.
ρ	0.1	Ant Colony	Across each generation of the Ant Colony algorithm the pheromone values from the previous generation are combined with the results of the current generation. This parameter controls how strong of an influence the previous pheromone values are on the new pheromone values. A value of 0 means that the previous pheromone values do not contribute to the new pheromone values. This value was set based on the work by Mohammadi on the Ant Colony [36] and requires further experimentation.
$F_{\text{min attraction}}$	0.01	Clustering	The minimum force constituting a valid attraction between concentrations of pheromone mass. This value was determined through trial-and-error and should face more thorough experimentation in the future.
$N_{\text{GC min}}$	100	Clustering	Each cluster identified by the clustering contains a set of stars. This value represents the minimum number of stars that should be present in the set for it to be considered a valid cluster and maintained. The smallest stellar cluster consists of 100 stars [39] and thus this was used as the value.

Chapter 4

Results and Findings

This chapter presents the results of the GC detection applied across the areas described in Chapter 2. They consist of a number of tables and plots which correspond to:

1. The three stages of the pipeline:

    i. Raster Exclusion Using BlobDoG (in Section 4.1)

    ii. Ant Colony (in Section 4.2)

    iii. Gravitational Clustering (in Section 4.3)

2. The full pipeline (in Section 4.4)

Note that each stage of the pipeline may depend on the results from the previous stages. Thus, the parametric variation that provides the best results for a given stage is carried forward as the basis for the future stages of the pipeline.

4.1 Raster Exclusion Using BlobDoG

To determine a sufficient value for $B_{\text{threshold}}$, the BlobDoG filter was run across all rasters for all areas while varying the threshold. Table 4.1 shows the number of rasters that remained in each area for the different threshold values.

Table 4.1: Rasters Remaining After the Execution of BlobDoG for Various Values of $B_{\text{threshold}}$

Area	Number of Rasters Remaining (All Rasters & Rasters Containing Known GCs)												Total Number of Rasters	
	$B_{\text{threshold}}$													
	0.1		0.2*		0.3		0.4		0.5		0.6			
	All	GCs	All	GCs	All	GCs	All	GCs	All	GCs	All	GCs	All	GCs
Area 1	512	12	8	7	8	7	7	7	7	7	7	7	512	12
Area 2: $2.0° \times 2.0°$	35	1	28	1	10	1	7	1	3	1	2	1	35	1
Area 2: $4.0° \times 4.0°$	12	1	9	1	8	1	7	1	7	1	5	1	12	1
Area 3	285	12	39	11	29	10	19	9	18	9	16	9	285	12
Area 4	120	0	60	0	10	0	9	0	6	0	4	0	120	0
Total	964	26	144	20	65	19	58	18	41	18	34	18	964	26

When the threshold is set too low one or more blobs are detected in every raster (as is the case with $B_{\text{threshold}} = 0.1$). As this value increases the number of blobs detected in each area is reduced. For these various thresholds $B_{\text{threshold}} = 0.2$ provides the largest reduction in the number of rasters while still maintaining the most of the rasters that contain known GCs. Thus, this threshold is selected as the basis for the future steps of the pipeline. The results of applying BlobDoG with this threshold across each area may be seen in Figure 4.1. This presents the rasters that are kept for this threshold and highlights which GCs are among those kept and among those removed.

---

*This was the threshold that was used for the future stages of the pipeline.

(a) Area 1

(b) Area 2: $2.0° \times 2.0°$

(c) Area 2: $4.0° \times 4.0°$

(d) Area 3

(e) Area 4
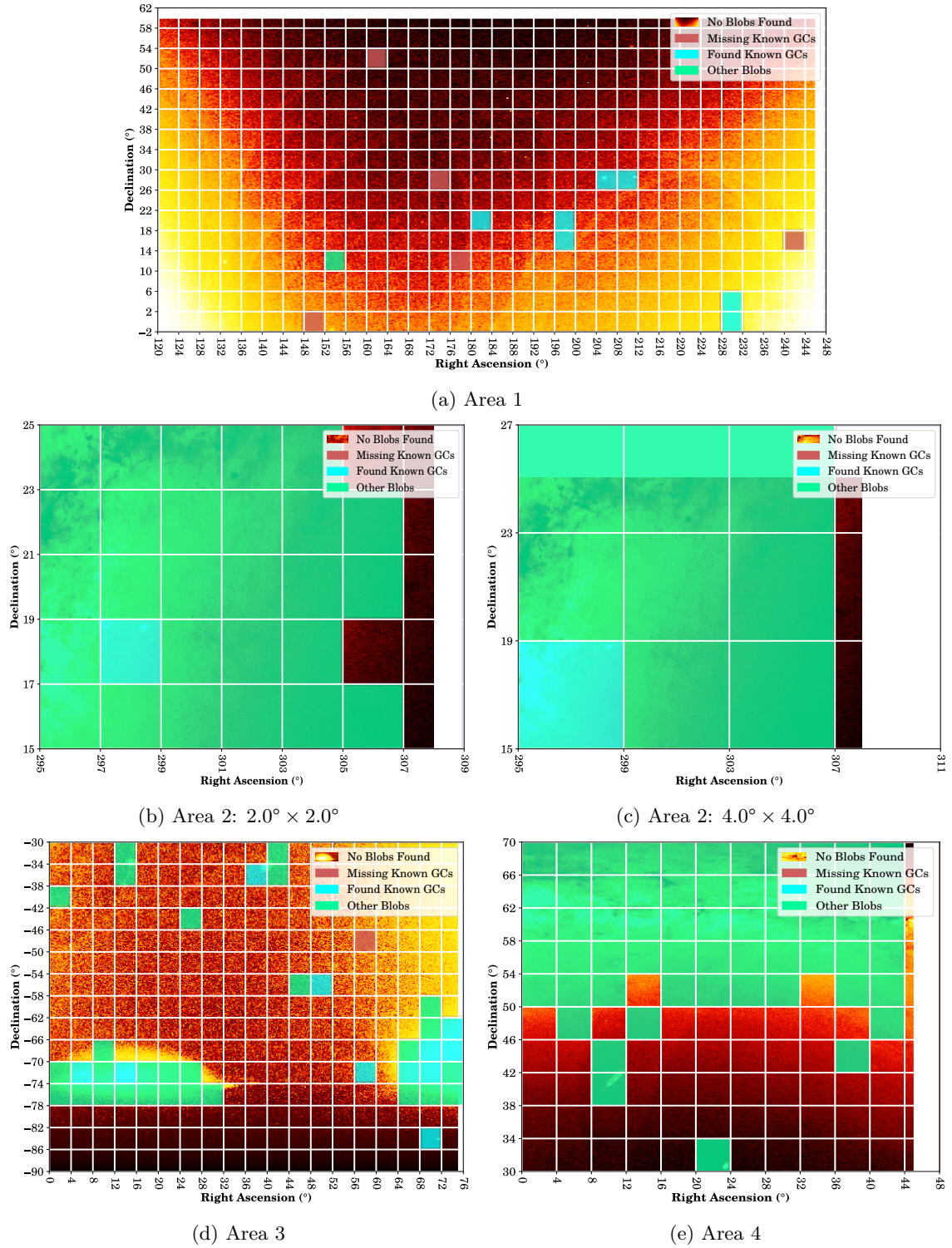
Figure 4.1: BlobDoG Raster Classification with $B_{\text{threshold}} = 0.2$

Each area in this figure is represented as a raster plot where BlobDoG finds *known GCs*, *other blobs* (corresponding to other stellar structures such as galaxies, dwarf galaxies, and open clusters), and *no blobs*. In these plots, rasters containing known GCs that were not maintained by BlobDoG (*missing known GCs*) are also highlighted. To provide a clear depiction of where these blobs are located, the results have been overlaid on the stellar distribution heat-map for each area.

It is evident that rasters in regions with a high number of stars or regions with a large variation in their stellar density have blobs detected by BlobDoG. Figure 4.1b and 4.1c depicting Area 2 and 4.1e depicting Area 4 find blobs mainly at the bright dense regions within those areas. Area 3 represented by Figure 4.1d has most of its blobs detected at the locations of the Magellanic Clouds. Additionally, the dark regions within this area have no blobs detected except in the one raster that contains a known GC. Area 1 finds blobs mainly around the GCs, but does not seem to find any blobs in the brightest regions in the corners of Figure 4.1a.

Table 4.2 lists the GCs present across the areas and specifies whether the raster they are contained within was maintained by BlobDoG.

Table 4.2: Known GCs detected at $B_{\text{threshold}} = 0.2$

Area 1		Area 2		Area 3	
GC	Present After BlobDoG?	GC	Present After BlobDoG?	GC	Present After BlobDoG?
Palomar 3	No	M71	Yes	47 Tucanae	Yes
Willman 1	No			NGC 121	Yes
Palomar 4	No			NGC 362	Yes
Koposov 1	No			NGC 1049	Yes
NGC 4147	Yes			NGC 1261	Yes
NGC 5024	Yes			NGC 1466	Yes
NGC 5053	Yes			Arp Madore 1	No
M3	Yes			NGC 1629	Yes
NGC 5466	Yes			NGC 1651	Yes
Palomar 5	Yes			NGC 1644	Yes
M5	Yes			NGC 1652	Yes
GCI 38	No			NGC 1841	Yes
				NGC 1696	Yes
				NGC 1756	Yes
				NGC 1786	Yes
				NGC 1783	Yes
				NGC 1795	Yes
Proportion Maintained	$^7/_{12}$		$^1/_1$		$^{16}/_{17}$

This table shows that after the raster exclusion by BlobDoG, 7 of 12 known GCs are maintained in Area 1, 1 of 1 known GCs are maintained in Area 2 (for both the $2.0° \times 2.0°$ rasterization and the $4.0° \times 4.0°$ rasterization), and that 16 of 17 known GCs are maintained for Area 3. Thus, BlobDoG maintains 80% of the known GCs.

4.2   Density Mapping Via the Ant Colony

The Ant Colony algorithm is an intermediary step to the final clustering. The primary result is the distribution of pheromone values across the rasters of each area. Since the Ant Colony algorithm makes use of a random-walk the results are non-deterministic and may differ between runs. As a result, to more accurately explore its behavior, 5 experiments were performed.

Table 4.3 shows statistics from the Ant Colony algorithm for each area across each experiment. It displays the mean pheromone values across the set of stars that were visited by at least one ant as well as the mean pheromone values across all of the stars in that area. In addition, it shows the number of stars that were visited by at least one ant across that experiment which is then used to compute the percentage of stars that were visited out of all the stars within an area.

Table 4.3: Pheromone Statistics and Visitations for Each Area Across Each Experiment

Area 1

Experiment	$\mu_{\text{pheromone (visited stars)}}$	$\mu_{\text{pheromone (all stars)}}$	#stars visited	$\%\dfrac{\#\text{stars visited}}{\text{Total} = 17\,933\,864}$
1	4.268×10^{-4}	233.9×10^{-6}	$9\,829\,764$	$54.81\,\%$
2	4.258×10^{-4}	233.9×10^{-6}	$9\,852\,113$	$54.94\,\%$
3	4.272×10^{-4}	233.9×10^{-6}	$9\,821\,846$	$54.77\,\%$
4	4.257×10^{-4}	233.9×10^{-6}	$9\,855\,991$	$54.96\,\%$
5	4.274×10^{-4}	233.9×10^{-6}	$9\,817\,422$	$54.74\,\%$

Area 2: $2.0° \times 2.0°$

Experiment	$\mu_{\text{pheromone (visited stars)}}$	$\mu_{\text{pheromone (all stars)}}$	#stars visited	$\%\dfrac{\#\text{stars visited}}{\text{Total} = 14\,268\,513}$
1	1.731×10^{-4}	20.10×10^{-6}	$1\,656\,651$	$11.61\,\%$
2	1.739×10^{-4}	20.10×10^{-6}	$1\,648\,789$	$11.56\,\%$
3	1.745×10^{-4}	20.10×10^{-6}	$1\,644\,010$	$11.52\,\%$
4	1.704×10^{-4}	20.10×10^{-6}	$1\,682\,903$	$11.79\,\%$
5	1.748×10^{-4}	20.10×10^{-6}	$1\,641\,042$	$11.50\,\%$

Area 2: $4.0° \times 4.0°$

Experiment	$\mu_{\text{pheromone (visited stars)}}$	$\mu_{\text{pheromone (all stars)}}$	#stars visited	$\%\dfrac{\#\text{stars visited}}{\text{Total} = 14\,268\,513}$
1	1.377×10^{-4}	6.892×10^{-6}	$714\,099$	$5.00\,\%$
2	1.304×10^{-4}	6.892×10^{-6}	$754\,141$	$5.29\,\%$
3	1.368×10^{-4}	6.892×10^{-6}	$718\,842$	$5.04\,\%$
4	1.383×10^{-4}	6.892×10^{-6}	$710\,776$	$4.98\,\%$
5	1.304×10^{-4}	6.892×10^{-6}	$754\,194$	$5.29\,\%$

Area 3

Experiment	$\mu_{\text{pheromone (visited stars)}}$	$\mu_{\text{pheromone (all stars)}}$	#stars visited	$\%\dfrac{\#\text{stars visited}}{\text{Total} = 9\,961\,034}$
1	4.721×10^{-4}	234.5×10^{-6}	$4\,946\,519$	$49.66\,\%$
2	4.710×10^{-4}	234.5×10^{-6}	$4\,958\,073$	$49.77\,\%$
3	4.700×10^{-4}	234.5×10^{-6}	$4\,969\,384$	$49.89\,\%$
4	4.729×10^{-4}	234.5×10^{-6}	$4\,938\,536$	$49.58\,\%$
5	4.735×10^{-4}	234.5×10^{-6}	$4\,932\,100$	$49.51\,\%$

Area 4

Experiment	$\mu_{\text{pheromone (visited stars)}}$	$\mu_{\text{pheromone (all stars)}}$	#stars visited	$\%\dfrac{\#\text{stars visited}}{\text{Total} = 22\,243\,660}$
1	2.150×10^{-4}	44.20×10^{-6}	$4\,573\,861$	$20.56\,\%$
2	2.149×10^{-4}	44.20×10^{-6}	$4\,576\,483$	$20.57\,\%$
3	2.141×10^{-4}	44.20×10^{-6}	$4\,592\,141$	$20.64\,\%$
4	2.153×10^{-4}	44.20×10^{-6}	$4\,566\,856$	$20.53\,\%$
5	2.145×10^{-4}	44.20×10^{-6}	$4\,584\,464$	$20.61\,\%$

Area 1 and Area 3 manifest similar statistics even with a significant difference in the total sum of their stars. Conversely, even with the same total number of stars, the ants visited double the amount of stars in Area 2: $2.0° \times 2.0°$ as compared to Area 2: $4.0° \times 4.0°$. Additionally, the ants traversing Area 4 visit a similar number of stars as in Area 3 even though Area 4 has double the quantity of stars. Thus, it is evident that the pheromone results are impacted by more than just

the total number of stars and must also be influenced by the distribution of stars present within the rasters of an area.

To explore the distribution of the pheromone values across each area, Figures 4.2 to 4.6 show the histograms and density plots generated for each experiment.

(a) Pheromone Histogram

(b) Pheromone Density Plot
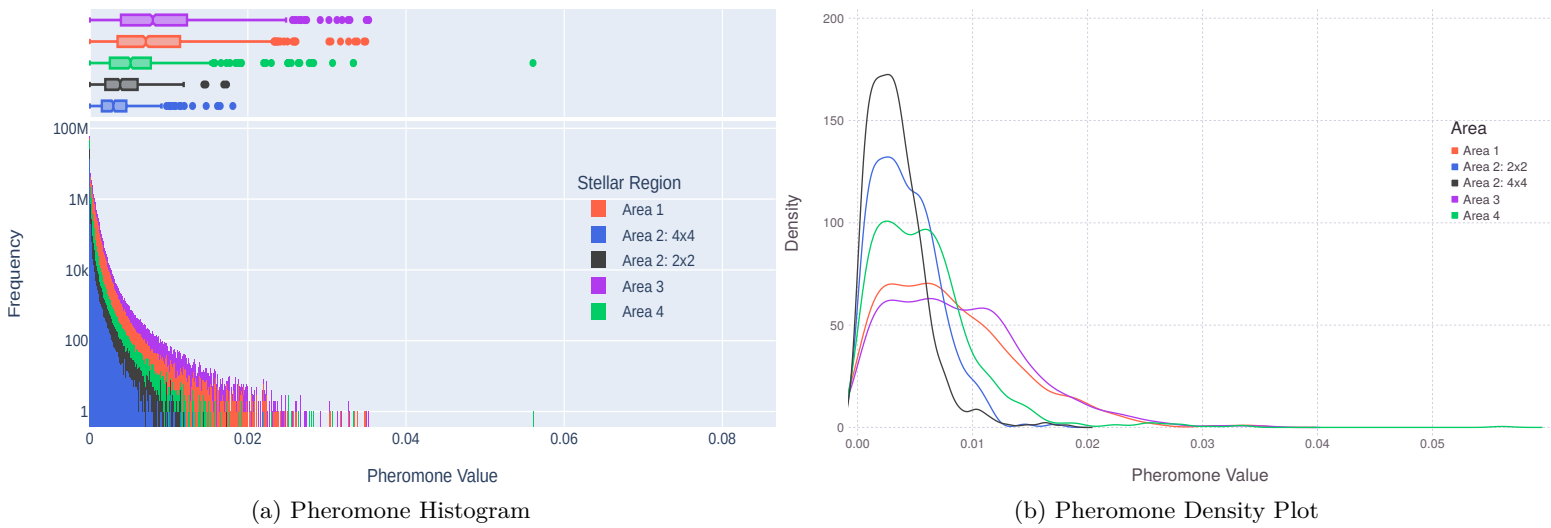
Figure 4.2: Plots of Pheromone Values for Experiment 1

(a) Pheromone Histogram

(b) Pheromone Density Plot
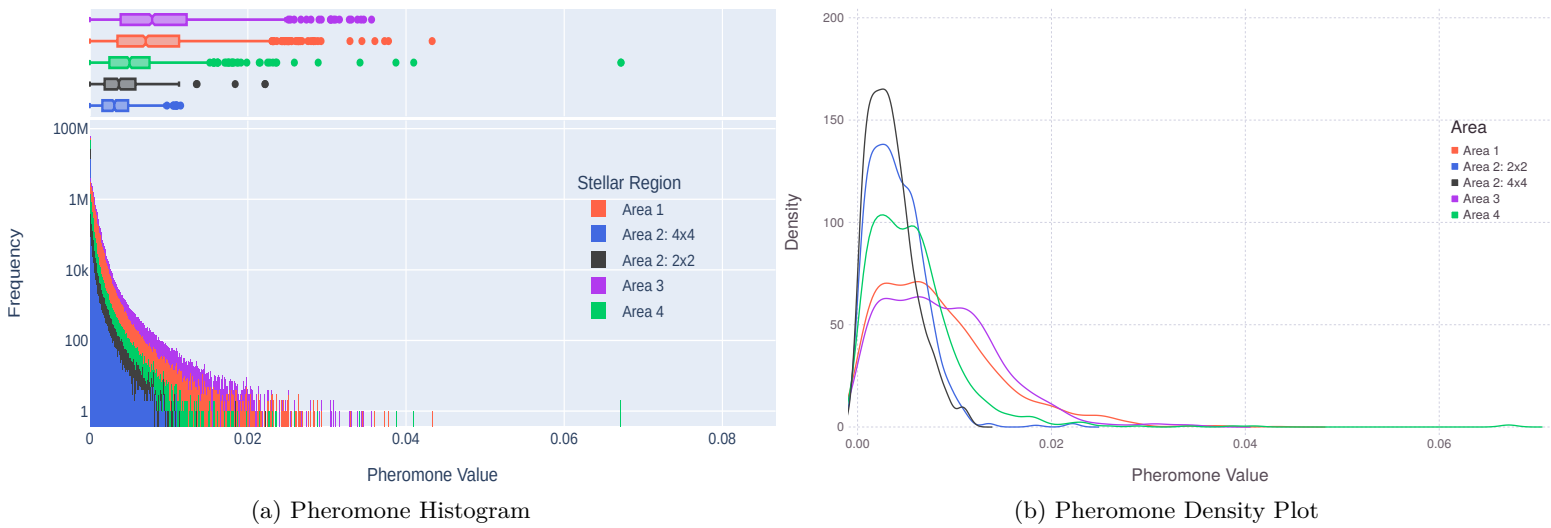
Figure 4.3: Plots of Pheromone Values for Experiment 2

(a) Pheromone Histogram

(b) Pheromone Density Plot
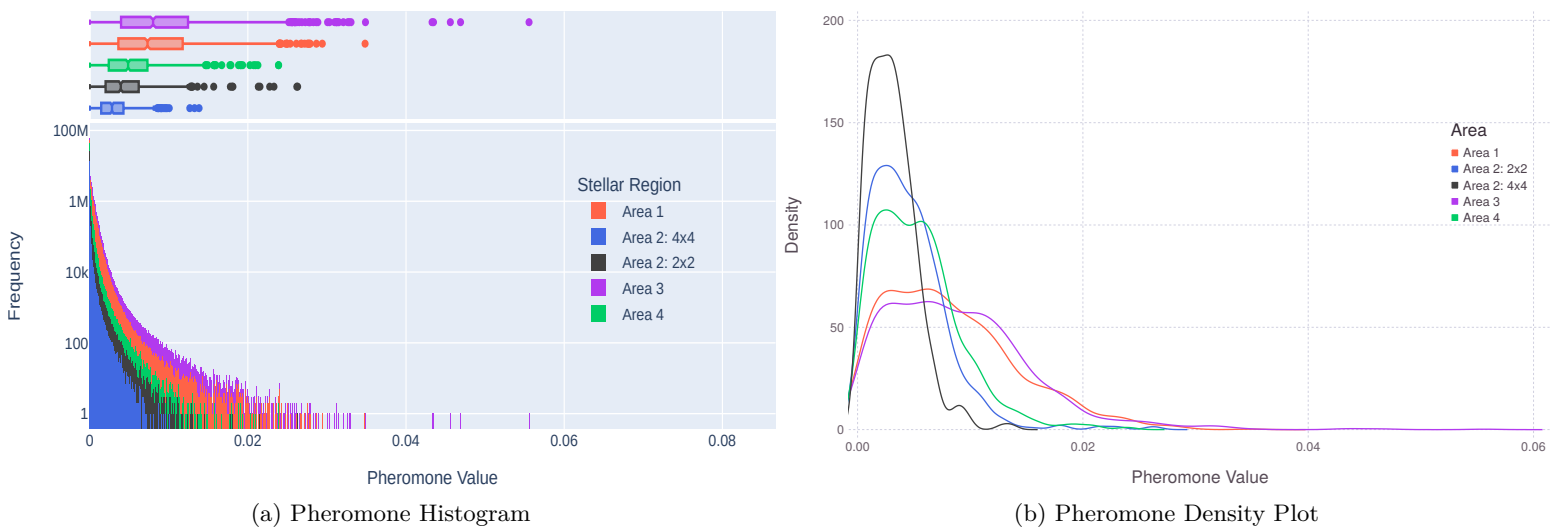
Figure 4.4: Plots of Pheromone Values for Experiment 3

(a) Pheromone Histogram

(b) Pheromone Density Plot
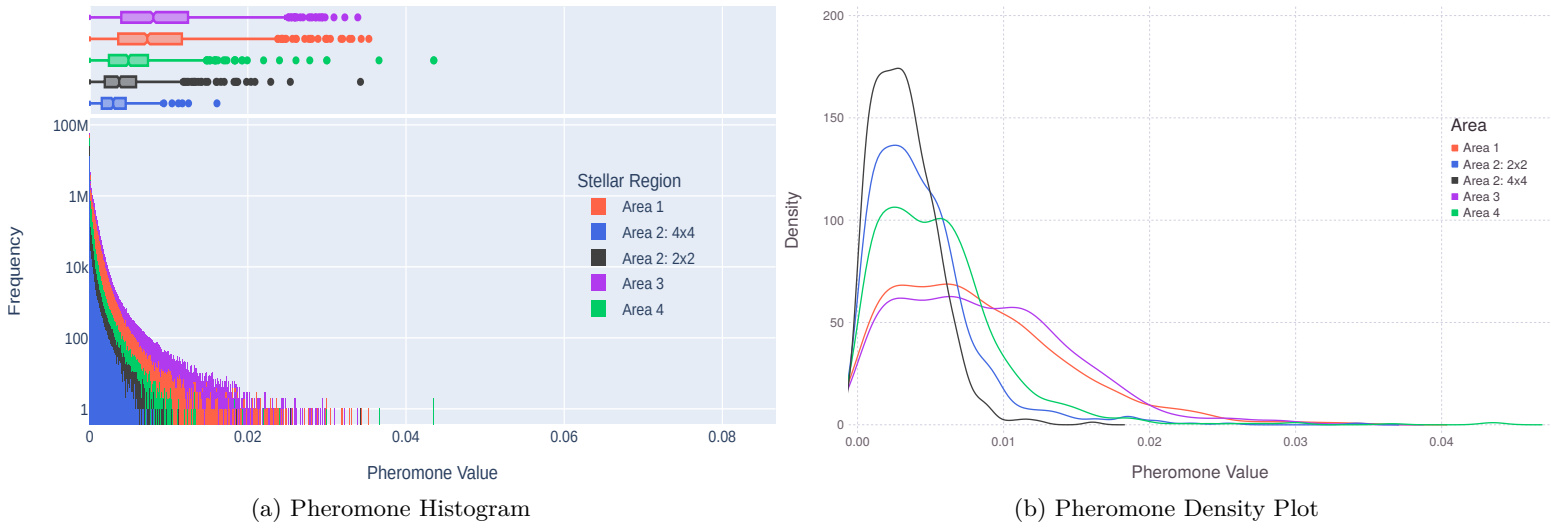
Figure 4.5: Plots of Pheromone Values for Experiment 4

(a) Pheromone Histogram

(b) Pheromone Density Plot
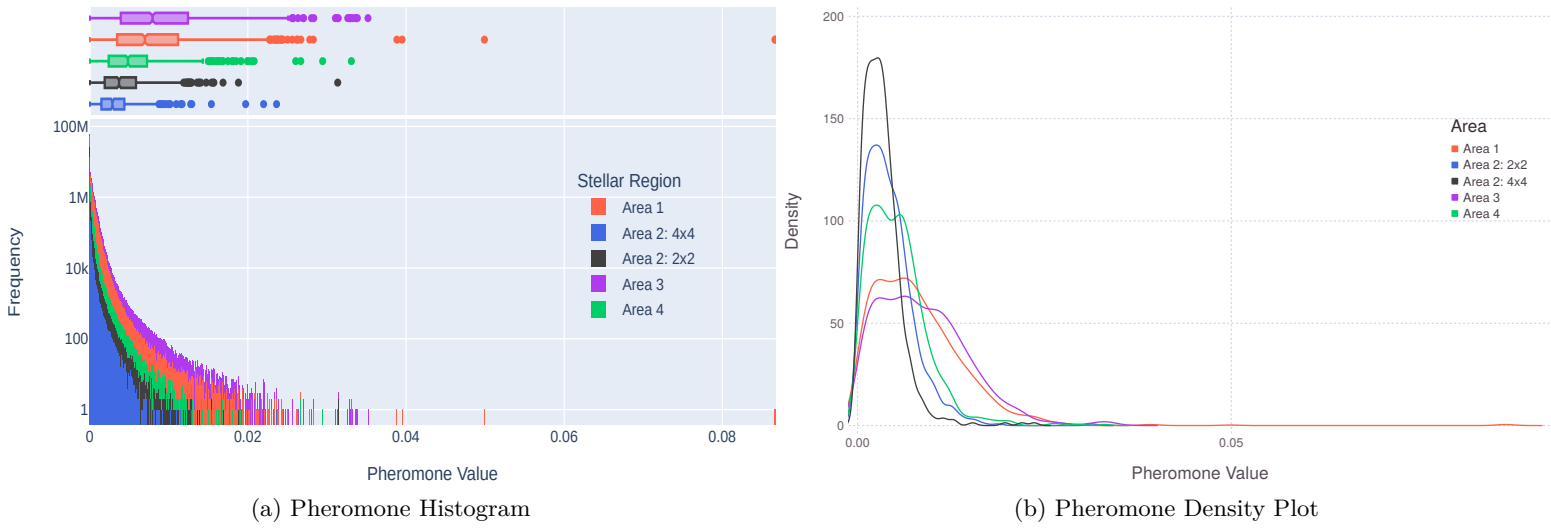
Figure 4.6: Plots of Pheromone Values for Experiment 5

The distribution of pheromone values across each experiment are similar with only minor variations in the outliers present. The bulk of the pheromone values lie between 0.00 and 0.01. Area 1 and Area 3 have similar pheromone distributions despite having a significant difference in both their total amount of stars and the amount of stars visited by at least one ant. Additionally, Area 2: $2.0° \times 2.0°$ has a wider distribution of pheromone values than Area 2: $4.0° \times 4.0°$.

4.3 Gravitational Clustering Based on Pheromone Mapping

The clustering algorithm is deterministic and so the same input will always result in the same output. However, the clustering directly relies on the results from the Ant Colony. Thus, the clustering phase was also tested across five experiments with each experiment using the results of one of the experiments from the Ant Colony phase.

The results of the clustering phase are sets of stars identified across each raster which constitute the clusters identified within that raster. The raw results detailing the statistics on every cluster identified across every experiment may be seen in Appendix 6.3 in Tables C.1 through Table C.4 and Figures B.1 through B.5. These results were aggregated by combining clusters that overlapped in both their RA bounds and their Dec bounds. Raster plots for these combined results may be seen in Figure 4.7.

(a) A1

(b) A2: 2.0° × 2.0°
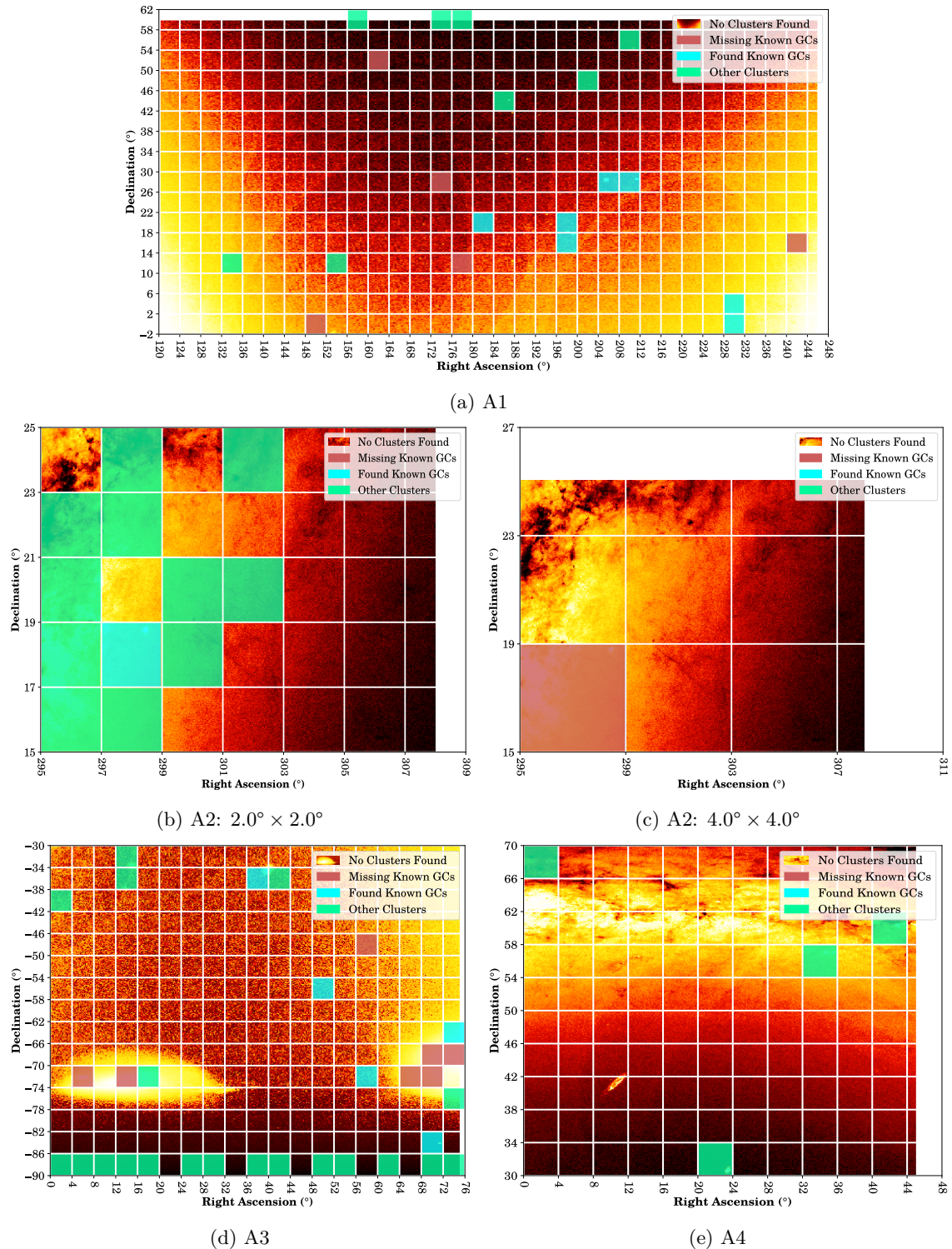

(c) A2: 4.0° × 4.0°

(d) A3

(e) A4

Figure 4.7: Clustering Raster Classification

Similar to the plots for BlobDoG in Figure 4.1, the results for the clustering have been overlaid on the stellar distribution heat-map for each area. However, while the results for BlobDoG are simply a binary decision for each raster, the results for the clustering are a set of clusters contained within each raster. Thus, Figure 4.7 does not represent the full results for the clustering and is instead provided to depict the general scope of where known GCs are *Found* or *Missing*, and where *Other Clusters* are detected. It is evident from the figure that a significant amount of the rasters identified contain GCs. The counts for the total number of clusters alongside how many rasters were involved are shown in Table 4.4.

Table 4.4: Number of Clusters Detected Per Area Across Each Experiment

Area	Experiment									
	1		2		3		4		5	
	# of Clusters	In # of Rasters	# of Clusters	In # of Rasters	# of Clusters	In # of Rasters	# of Clusters	In # of Rasters	# of Clusters	In # of Rasters
Area 1	16	11	19	13	18	12	16	11	15	10
Area 2: 2.0° × 2.0°	3	3	1	1	5	5	5	5	3	4
Area 2: 4.0° × 4.0°	0	0	0	0	0	0	0	0	0	0
Area 3	27	22	28	23	29	24	29	23	28	22
Area 4	2	2	2	2	3	3	3	3	1	1
Total	48	38	50	39	55	44	53	42	47	37

Due to the difference between the number of clusters and the number of involved rasters it is clear that in some cases multiple clusters are found within the same raster. Across the different experiments there is a variety in the number of clusters that were found. However, these differences hover aroud the same ballpark figure, a total of 51 ± 4 clusters. Additionally, across each experiment the arrangement of areas by the number of clusters found remains the same (Area 3 finding the most, followed by Area 1, Area 2: 2.0° × 2.0°, Area 4, and finally Area 2: 4.0° × 4.0°). The results compared against the list of known GCs is as follows:

Table 4.5: Known GCs Detected by Clustering

Area 1			Area 2			Area 3		
GC	Present After Clustering?	Experiments	GC	Present After Clustering?	Experiments	GC	Present After Clustering?	Experiments
Palomar 3	No		M71	Yes	1 4	47 Tucanae	No	
Willman 1	No					NGC 121	No	
Palomar 4	No					NGC 362	No	
Koposov 1	No					NGC 1049	Yes	1 2 3 4 5
NGC 4147	Yes	1 2 3 4 5				NGC 1261	Yes	1 2 3 4 5
NGC 5024	Yes	1 2 3 4 5				NGC 1466	Yes	1 3 4 5
NGC 5053	Yes	1 2 3 4				Arp Madore 1	No	
M3	Yes	1 2 3 4 5				NGC 1629	No	
NGC 5466	Yes	1 2 3 4 5				NGC 1651	No	
Palomar 5	Yes	1 2 3 4 5				NGC 1644	No	
M5	Yes	2				NGC 1652	No	
GCI 38	No					NGC 1841	Yes	1 2 3 4 5
						NGC 1696	No	
						NGC 1756	No	
						NGC 1786	No	
						NGC 1783	Yes	1 2 3 5
						NGC 1795	No	
Proportion Identified	$7/12$			$1/1$			$5/17$	

Most of the GCs that are identified were consistently found across the experiments. The exceptions are the GCs M5 and M71, which were only identified in 1 and 2 experiments respectively. When aggregating the results across all the experiments, the same GCs are discovered in Area 1 and Area 2 as kept by the BlobDoG filter. However, only 5 of the 17 are identified for Area 3 as compared to 16 of the 17 maintained by BlobDoG.

4.4 Full Pipeline

The full pipeline constitutes the initial raster filtration by BlobDoG, followed by the pheromone mapping by the Ant Colony algorithm, which finally culminates in the Gravitational Clustering of the stars based on their pheromone values. This section presents the results of this full pipeline and reframes the results of the previous stages in the context of the full pipeline.

(a) A1

(b) A2: $2.0° \times 2.0°$

(c) A2: $4.0° \times 4.0°$

(d) A3

(e) A4
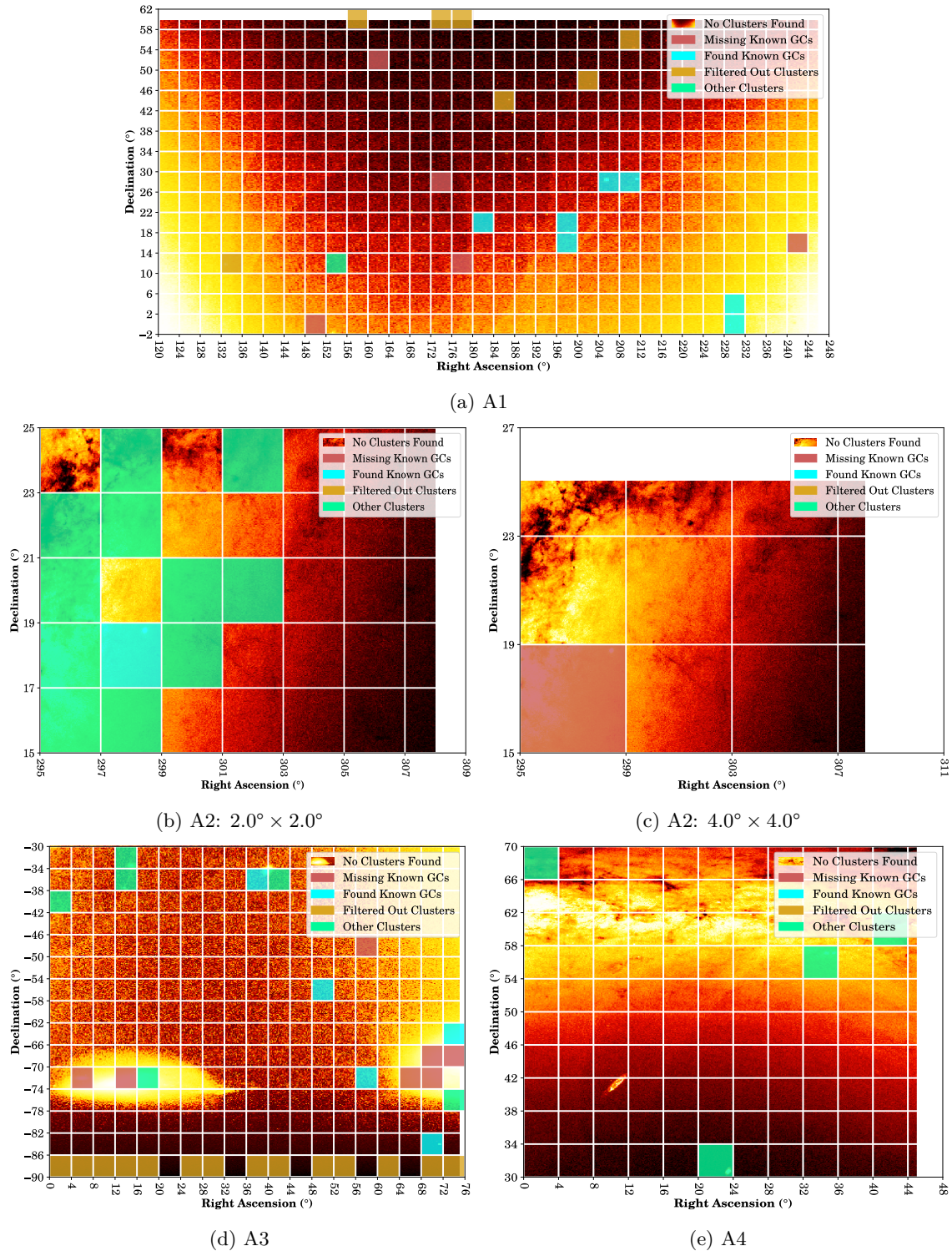
Figure 4.8: Pipeline Raster Classification

Figure 4.8 combines the results of BlobDoG and the clustering. It showcases the rasters that contain *Found Known GCs* or contain *Missing Known GCs* as well as *Other Clusters*. In addition, it shows the rasters that the clustering phase would have identified as containing a stellar structure but which were preemptively filtered away by BlobDoG (*Filtered Out Clusters*).

Table 4.6 describes the clusters that were identified by the clustering but had been filtered away

by BlobDoG. In Table 4.8 the remaining clusters that were identified by the clustering are shown. Additionally, it describes the experiments that the clusters were identified in. Finally, any cluster whose bounds contained an existing stellar structure according to Stellarium [40] has been marked with the name of the structure and its type. Clusters that did not corresponding to an existing stellar structure were given the type of *Nothing*.

Table 4.6: All BlobDoG's Cluster Removals

Type	Name	Experiments
	Area 1	
OC	Golden-Eye Cluster	1 2 3 4 5
Galaxy	NGC 3286 NGC 3288	4
Galaxy	NGC 3770 NGC 3795	2
Galaxy	NGC 3838 PGC 36398 PGC 36585 PGC 36655 PGC 36877	1 2 3 4 5
Galaxy	Box Galaxy	1 2 3 4 5
Galaxy	Whirlpool Galaxy	1 2 3 5
Galaxy	Pinwheel Galaxy	1 2 3 4 5
	Area 3	
Nothing	N.A.	1 2 3 4 5
Galaxy	PGC 993	2 3 4 5
Nothing	N.A.	1 2 3 4 5
Galaxy	PGC 3533	1 2 3 4 5
Galaxy	PGC 216800	1 2 3 4 5
Nothing	N.A.	1 2 3 4 5
Galaxy	PGC 5780	1 2 3 4 5
Nothing	N.A.	1 2 3 4 5
Galactic Cluster	Abell 3037	1 2 3 4
Nothing	N.A.	1 2 3 4 5
Nothing	N.A	1 2 3 4 5
Nothing	N.A.	2 3 4
Nothing	N.A.	1 3 4 5
Nothing	N.A.	1 3 4 5
Nothing	N.A.	1 2 3 4 5
Nothing	N.A.	4 5
Galactic Cluster	Abell 3333	1 2 3 4 5

For Area 1, there were 6 galaxies and 1 OC which was filtered away by BlobDoG. For Area 3, there were 6 galaxies and 11 clusters that were identified as nothing. The average and maximum bounds across the filtered clusters is as follows:

Table 4.7: Statistics on the RA Bounds and the Dec Bounds for the Filtered Clusters

	Average	Max
RA Bounds	01°.12	01°.88
Dec Bounds	00°.97	01°.55

The RA bounds and the Dec bounds reveal that the clusters that BlobDoG have a diameter of approximately 1°.0. This is in contrast to the bounds corresponding with the results of the clusters identified by the clustering that were also maintained by BlobDoG. Additional information on the stellar coordinates of each cluster and their bounds may be found in Table A.1 in Appendix A.

Table 4.8: Clusters Found in All Experiments

Type	Name	Experiments
	Area 1	
Dwarf Galaxy	Regulus Dwarf Galaxy	1 2 3 4 5
GC	NGC 4147	1 2 3 4 5
GC	NGC 5024	1 2 3 4 5
GC	NGC 5053	1 2 3 4
GC	M3	1 2 3 4 5
GC	NGC 5466	1 2 3 4 5
GC	Palomar 5	3
GC	M5	2
	Area 2	
Nothing	N.A.	1
Nothing	N.A.	5
Nothing	N.A.	5
Nothing	N.A.	3
Nothing	N.A.	4
Nothing	N.A.	4
Nothing	N.A.	3
Molecular Cloud	LDN 758	3
OC	NGC 6827	2
Nothing	N.A.	1
Nothing	N.A.	4
Absorption Nebula	LDN 787	5
GC	M71	1 4
Nothing	N.A.	5
Nothing	N.A.	4
Nothing	N.A.	3
Nothing	N.A.	3
Nothing	N.A.	5
	Area 3	
Galaxy	String of Pearls	1 2 3 4 5
Galaxy	Southern Pinwheel Galaxy	1 2 3 4 5
Dwarf Galaxy	Sculptor Dwarf Galaxy	1 2 3 4 5
OC	NGC 419	2
GC	NGC 1049	1 2 3 4 5
Galaxy	Fornax Dwarf Galaxy	1 2 3 4 5
GC	NGC 1261	1 2 3 4 5
GC	NGC 1466	1 3 4 5
GC	NGC 1841	1 2 3 4 5
OC	NGC 1777	1 2 3 4 5
GC	NGC 1783	1 2 3 5
	Area 4	
HII region	NGC 7822	1 2 3 4
Galaxy	Triangulum Galaxy	1 2 3 4 5
OC	Double Cluster	3
OC	NGC 1027	4

This table shows that the full pipeline identifies 13 GCs, 5 OCs, 4 Galaxies, 2 Dwarf Galaxies, a Molecular Cloud *, an Absorption Nebula †, and an Emission Nebula ‡. In addition, it finds 14 clusters that do not correspond to a known stellar structure. For Areas 1, 3, and 4 most of the clusters are found consistently across the experiments. However, in Area 2 the clusters are only found sporadically, being found in at most two experiments.

Table 4.9: Statistics on the RA Bounds and the Dec Bounds for the Identified Clusters

	Average	Max
RA Bounds	$00\overset{\circ}{.}25$	$01\overset{\circ}{.}03$
Dec Bounds	$00\overset{\circ}{.}27$	$01\overset{\circ}{.}07$

The statistics on the bounds the clusters that were identified by the full pipeline are substantially lower than the bounds for those filtered away by BlobDoG. Additional information on the stellar coordinates of each cluster and their bounds may be found in Table A.2 in Appendix A.

---

*A dense region of interstellar gas and dust.

†A very dense region within a molecular cloud whose density obscures the light from stars behind it. These regions are known as stellar nurseries as they have the characteristics permitting the formation of new stars.

‡The formation of a young star from an absorption nebula. The gravitational force of the gas and dust causes the region to heat up to the point of emitting light in the visible spectrum.

Chapter 5

Conclusion

For BlobDoG with a $B_{\text{threshold}} = 0.2$, 7 out of the 12 rasters containing known GCS are maintained in Area 1, the 1 containing a known GC is maintained in Area 2, and 16 out of 17 are maintained in Area 3. It is evident from these results that BlobDoG with this $B_{\text{threshold}}$ does not work perfectly as a pre-processing method, as it filters out known GCs. However, it does keep 80% (24 out of 30) of the known GCs. Furthermore, no raster filtered away by BlobDoG contained any known GCs as identified by the clustering phase that follows. A total of 21 rasters which would have been identified by the clustering as containing a stellar structure were filtered away by BlobDoG, 7 of which are in Area 1 and 14 in Area 3.

From Table 4.6 the stellar structures that are removed by BlobDoG are predominantly galaxies (10 of 24) or do not correspond to any identifiable stellar structure (11 of 24). These clusters that have no known stellar structures, are either a failure of the Ant Colony algorithm or represent potentially undiscovered GC candidates. It is possible that the $B_{\text{threshold}}$ may be too large (especially as it discards known GCs) and thus some of these clusters that were filtered away may be on the cusp for BlobDoG. However, lowering the $B_{\text{threshold}}$ to 0.1 results in no rasters being filtered away. Thus, the ideal $B_{\text{threshold}}$ lies between $0.1 - 0.2$. To summarize BlobDoG operates as an effective exclusion criteria but it is not yet perfect in maintaining all known GCs.

For the Ant Colony algorithm, the amount of stars per raster influence the $\mu_{\text{pheromone}}$ as well as the distribution of stars across that raster. These quantities also influence the percentage of the visited stars in that raster, which can be observed from Table 4.3, when comparing Area 2: $2.0° \times 2.0°$ versus Area 2: $4.0° \times 4.0°$. The $\mu_{\text{pheromone}}$ values, the distribution of the pheromone values, and the stars' visitation percentage in Area 1 and Area 3 are quite similar despite the areas having very different characteristics. Area 1 has the lowest density of stars across a region while Area 3 features regions with an extremely high density of stars. Exploring the behavior of the Ant Colony algorithm with respect to the total number of stars and the density of the stars with the rasters is pivotal to optimizing the clustering phase that follows.

From Table 4.3 and Figures 4.2 to 4.6 it can be observed that the ants behave similarly across each of the experiments. However, it becomes clear from the results of the Clustering that even the small variations in the ants behavior can result in distinct clusterings. This is evident in Table 4.4 where the results of visitation percentage and pheromone distribution were similar but the number of clusters found in each experiment per area were different. Additionally the sporadic results from the clustering of Area 2 reveals that it is necessary to run the algorithm many times to identify the different clusters present and to gauge the certainty with which the ants discover it.

All areas find clusters except for Area 2: $4.0° \times 4.0°$, which highlights the relationship between the functioning of the Ant Colony algorithm and the amount of stars contained within the rasters. Table 4.4 shows that some rasters contain multiple clusters. The known GCs found in Area 1 and Area 2 are the same as found by BlobDoG but Area 3 only finds 5 out of 17 known GCs. These results show promise but highlight the need to fine tune the control parameters of the Ant Colony and the clustering. In Area 1 and Area 3 many clusters were detected in rasters corresponding to the darker regions of the areas which were subsequently filtered by BlobDoG. This shows that the Ant Colony may have a predisposition to identifying clusters more easily in regions with less stars.

Across the full pipeline, the system is able to identify 13 GCs out of the 30 known GCs and identifies a total of 13 rasters out of the 27 rasters containing known GCs. From Table 4.8, 27 out of the 41 clusters that were identified by the pipeline corresponded to a known stellar structure. 13 of these corresponded to GCs and 14 corresponded to some other pre-existing stellar structure.

While the pipeline does not identify the majority of the GCs that exist, for those that it is able to identify, it can pinpoint their locations quite accurately. Further research in tuning the parameters and steering the behavior of the ants is expected to expand the number of GCs identified by the pipeline and to solidify the Ant Colony as a useful tool for exploring the Universe.
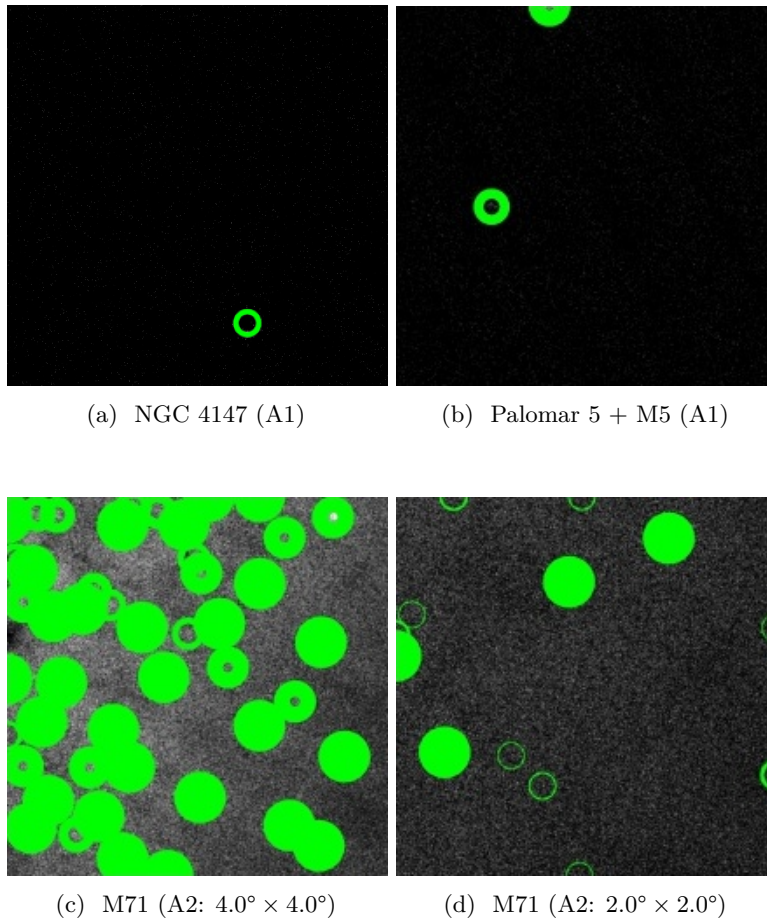
Chapter 6

Evaluation

In this chapter, the results are examined in closer detail to provide an insight into the inner working of the algorithms. The questions raised during the course of the paper are answered, shortcomings and possible improvements are discussed, and the work is contrasted to pre-existing research. This evaluation is performed for each of the three phases of the pipeline:

- Raster Exclusion Using BlobDoG (in Section 6.1)

- Ant Colony (in Section 6.2)

- Gravitational Clustering (in Section 6.3)

6.1 Raster Exclusion Using BlobDoG

What is the behavior displayed by BlobDoG?
To examine the behavior of BlobDoG, results for various rasters across the areas are displayed in Figure 6.1. These rasters span from dark to bright regions and contain GCs, other clusters, and galaxies.

(a) NGC 4147 (A1)   (b) Palomar 5 + M5 (A1)

(c) M71 (A2: 4.0° × 4.0°)   (d) M71 (A2: 2.0° × 2.0°)

(e) Small Magellanic Cloud (A3)   (f) Large Magellanic Cloud (A3)

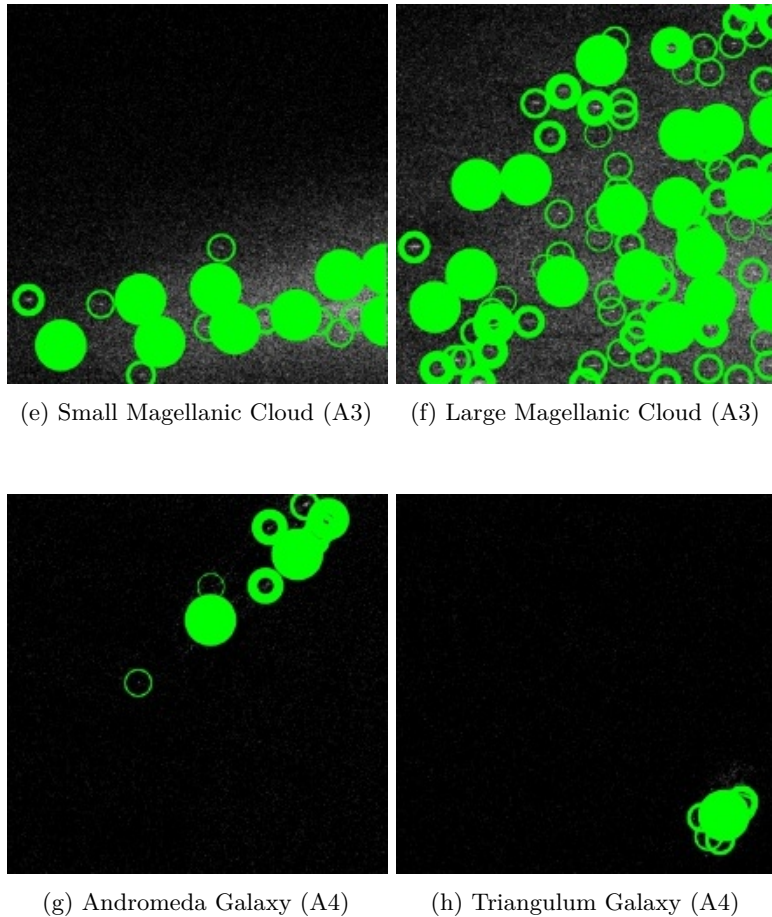(g) Andromeda Galaxy (A4)   (h) Triangulum Galaxy (A4)

Figure 6.1: Interesting Examples of the Blobs Found by BlobDoG

The rasters in Area 1 (Figures 6.1a and 6.1b) show GCs that stand out visibly against their dark stellar background. These blobs are easily found and identified even in the scenario for M5 where the GC is only partially contained within the raster. The rasters in Area 2 (Figures 6.1c and 6.1d) both show the same region containing the GC M71. This GC is highlighted by BlobDoG across both rasterization schemes but at different scales. Additionally, it is evident that the many smaller blobs identified in Figure 6.1d are not present in Figure 6.1c. The rasters in Area 3 (Figures 6.1e and 6.1f) show the bright and busy Magellanic Clouds. No GCs are present in the Small Magellanic Cloud shown in Figure 6.1e. Nevertheless, it is still a busy area with many blobs being identified. The raster containing the Large Magellanic Cloud contains four GCs (NGC 1696, NGC 1756, NGC 1786, and NGC 1795). While these GCs are identified in Figure 6.1f, it is difficult to distinguish them due to the amount of other blobs that are also highlighted. The rasters in Area 4 (Figure 6.1g and 6.1h) contain the Andromeda Galaxy and the Triangulum Galaxy. These galaxies are identified as many smaller blobs rather than a single continuous blob.

It is evident from the plots in Figure 6.1 that BlobDoG is able to identify blobs across a wide range of scenarios. The results in Area 2: $4.0° \times 4.0°$ and the Magellanic Clouds reveal that BlobDoG is more responsive in bright busy areas. This behavior is beneficial as brighter areas containing many stars are more likely to contain clusters. Additionally, the difference in the blob detection between A2: $2.0° \times 2.0°$ and A2: $4.0° \times 4.0°$ reveals that the rasterization scheme directly influences the results of BlobDoG. Thus, more research should be performed on identifying any possible parametric relationship between the $B_{\mathrm{threshold}}$ and the size of the rasters generated by the rasterization scheme.

Why are rasters containing known GCs filtered out at $B_{\mathbf{threshold}} = 0.2$?
BlobDoG fails to identify all known GCs and filters away some rasters containing known GCs. Table 6.1 lists these GCs and provides additional information to aid with identifying a trend underlying these failures.

Table 6.1: Known GCs that are Not Detected with $B_{\text{threshold}} = 0.2$

Area 1	
Koposov 1	A low-luminosity GC with a distance of 48.3 kpc [41].
Palomar 3	A distant GC at 96 kpc with an apparent magnitude of 14.26 [42].
Palomar 4	A distant GC at 109 kpc with an apparent magnitude of 15.65 [21].
GCI 38	A distant GC at 74.7 kpc with an apparent magnitude of 74.7 [21].
Willman 1	An ultra low-luminosity GC with a distance of 38 kpc but with a very low brightness corresponding to an apparent magnitude of 15.2 [43].

Area 3	
Arp Madore 1	A very distant GC at 123.3 kpc [44].

All of these GCs that were not maintained are either very distant, have a low-luminosity, or manifest both of these attributes. It is these factors that lead to BlobDoG failing to identify these clusters. This becomes apparent visually when considering Figure 6.2.

(a) Koposov 1 (A1)   (b) Palomar 3 (A1)   (c) Palomar 4 (A1)

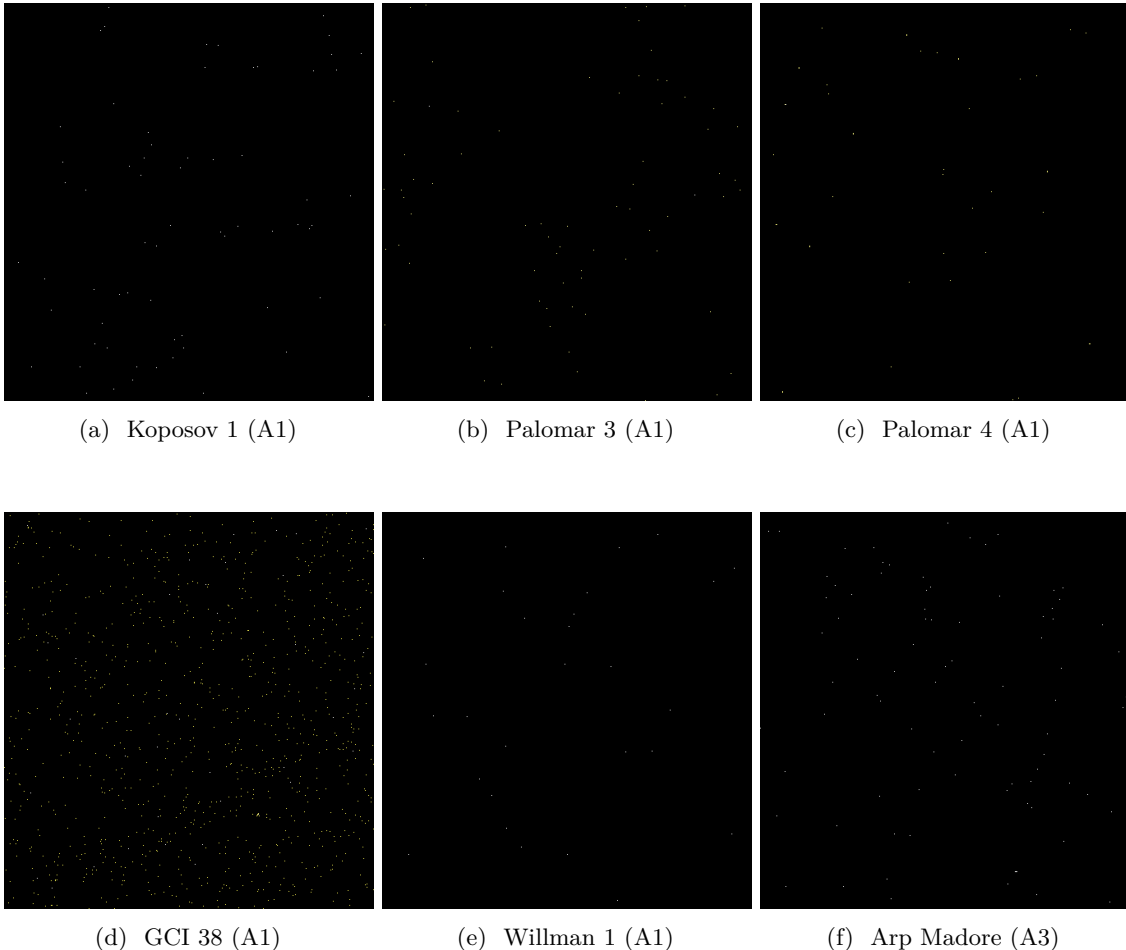(d) GCI 38 (A1)   (e) Willman 1 (A1)   (f) Arp Madore (A3)

Figure 6.2: Known GCs Not Detected with $B_{\text{threshold}} = 0.2$

These GCs are barely visible or correspond with a very small blob (which would then be discarded by the $B_{\text{threshold}} = 0.2$). It might be feasible to resolve this by reducing the $B_{\text{threshold}}$ further. In addition, it may be possible to make use of an alternate rasterization scheme which also rasterizes across the distance. It would then be possible to augment the rasters that are further away to better account for these distant clusters at the cost of accuracy.

What would be the optimal $B_{\mathbf{threshold}}$?

BlobDoG filters away 85 % of the rasters with a $B_{\text{threshold}}$ of 0.2, while maintaining 80 % of the known GCs. In ideal circumstances BlobDoG would filter away many empty rasters while still maintaining every raster containing a GC. Through the testing performed in Section 4.1, it is evident that if an optimal $B_{\text{threshold}}$ were to exist within the rasterization scheme that was tested, it must lie between 0.1 to 0.2. However, it is possible that there is no $B_{\text{threshold}}$ which is able to maintain all known GCs while still filtering away a majority of the rasters. Further experimentation is required to test this.

How could BlobDoG be improved?

It is important to identify the relationship between the $B_{\text{threshold}}$ and the size of the rasters generated by the rasterization scheme. Furthermore, the information provided by BlobDoG could be used to provide better direction to the Ant Colony algorithm. The locations of the blobs that were identified by BlobDoG could be used to provide the initial weights for the stars processed by the Ant Colony. Thus, the initial generation of ants would be predisposed to evaluate the regions containing those blobs.

In addition, the possibility of rasterizing in RA, Dec, as well as distance should be evaluated. By rasterizing across distance as well, it may be possible to enhance the rasters that are further away to better capture distant low-luminosity clusters.

BlobDoG is unable to account for distance in its results. This is in contrast to the clustering algorithm which is able to classify multiple clusters overlapping in RA and Dec. In the work by Mohammadi et al. [8], BlobDoG was used as a post-processing step. If the rasterization is applied across distance, this raises an interesting proposition for the use of BlobDoG in place of the gravitational clustering algorithm. For this final step, BlobDoG would be applied across RA and Dec with the luminosity for each pixel being determined by the pheromone value of the star at that pixel. An example of BlobDoG being applied as a post-processing mechanism may be seen in Figure 6.3.

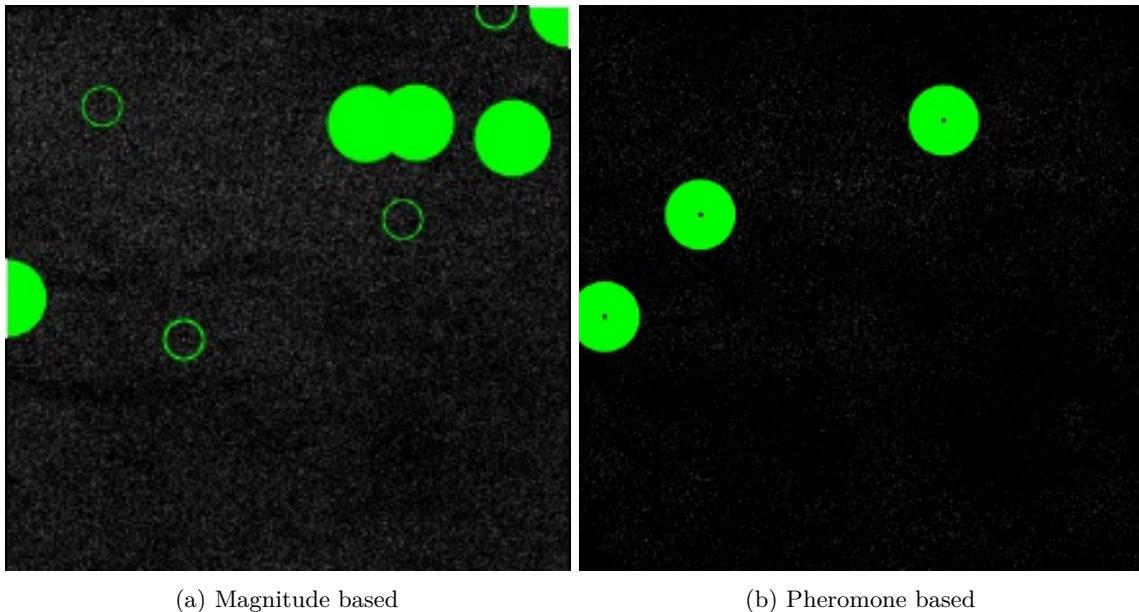(a) Magnitude based   (b) Pheromone based

Figure 6.3: Blobs Identified by BlobDoG In the Same Raster Under Different Parameters

As may be seen in this figure, the results of BlobDoG when applied on the absolute magnitude of the stars within the raster are very different to the results when it is applied to the pheromone values of the stars within the raster. In this instance, BlobDoG is able to precisely identify the pheromone clusters left by the Ant Colony. This highlights its potential as a mechanism to cluster pheromone values. Incorporating this step would result in the pipeline being as follows:

1. Rasterize based on RA, Dec, and Distance.

2. Apply BlobDoG on the rasters based on their apparent magnitude.

3. Generate the pheromone maps using the Ant Colony algorithm.

4. Apply BlobDoG on the pheromone values.

Since BlobDoG operates with a lower time complexity than the clustering algorithm, this reformulation may provide sufficient results while operating substantially faster.

6.2 Density Mapping Via the Ant Colony

What is the behavior displayed by the Ant Colony algorithm?
To examine the behavior of the Ant Colony algorithm, results for various rasters across the areas are displayed in Figure 6.4. The rasters that are examined are the same as those that were used for the evaluation of the behavior of BlobDoG. To compare the distribution of the stars across the rasters against the network substructure discovered by the Ant Colony algorithm, the stellar distribution heat-maps of the rasters (indicated by †) are positioned alongside the pheromone heat-maps (indicated by ‡).

(a) NGC 4147 (A1)†   (b) NGC 4147 (A1)‡   (c) Palomar 5 + M5 (A1)†   (d) Palomar 5 + M5 (A1)‡

(e) M71 (A2: $4.0° \times 4.0°$)† (f) M71 (A2: $4.0° \times 4.0°$)‡ (g) M71 (A2: $2.0° \times 2.0°$)† (h) M71 (A2: $2.0° \times 2.0°$)‡

(i) Small Magellanic Cloud (A3)†   (j) Small Magellanic Cloud (A3)‡   (k) Large Magellanic Cloud (A3)†   (l) Large Magellanic Cloud (A3)‡

(m) Andromeda Galaxy (A4)†   (n) Andromeda Galaxy (A4)‡   (o) Triangulum Galaxy (A4)†   (p) Triangulum Galaxy (A4)‡
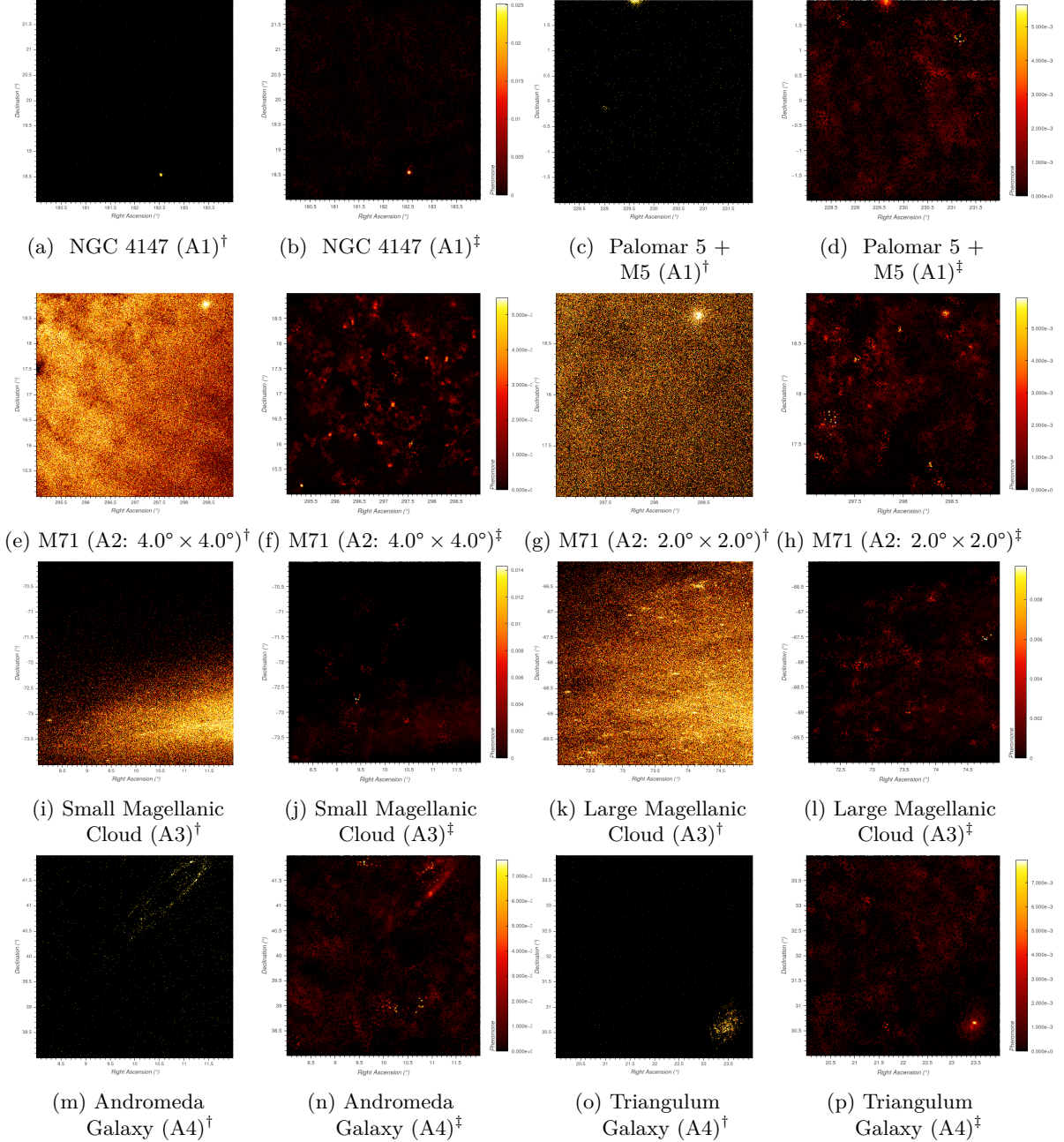
Figure 6.4: Stellar Distribution Heat-maps (†) vs. Pheromone Heat-maps (‡) for Various Rasters

The algorithm functioning as intended should result in pheromone mappings honed in on stellar clusters and other dense regions. From the plots of the rasters in Area 1 and Area 4, it is evident that the algorithm works well at identifying concentrated clusters of stars surrounded by regions of very low concentration. The structure of these rasters is directly observable in the density map uncovered by the Ant Colony. The plots for Area 1 demonstrate the ability of the Ant Colony algorithm to highlight the location of GCs through an increased concentration of pheromone values. Visually, this corresponds to the focused bright spot present in the pheromone heat-map of the rasters. For the plots of Area 4, the pheromone heat-maps show larger bright smudges corresponding to the nearby galaxies contained within those rasters. These pheromone mappings are less intense but still stand

out, clearly formed in the shape of the galaxies at hand. For Area 1 and 4, the algorithm is able to distinctly identify these clusters. However, the results for the other areas are less definitive.

For Area 2: $2.0° \times 2.0°$ it is able to highlight the GC contained within the upper-right corner of the raster, but it does so with many other less relevant regions also being highlighted. This is taken to the extreme with Area 2: $4.0° \times 4.0°$ which does not clearly identify the GC at all. The difference in these outcomes is due to the disparity in the number of stars contained within both rasters. As Area 2: $2.0° \times 2.0°$ contains less stars, the connected structure that this GC demonstrates, represents an increased proportion of the total substructure. Thus, the ants are more likely to gather within the GC to pool their pheromone values than in Area 2: $4.0° \times 4.0°$. Both rasters present a busy starscape and it seems that the ants are being spread too thin to have their pheromone values be reliably identified by the clustering phase that follows.

The plots for Area 3 focus on the rasters containing the Magellanic Clouds. Figure 6.4j is of the Small Magellanic Cloud and has no GCs present, while Figure 6.4l shows the Large Magellanic Cloud which contains the four GCs: NGC 1696, NGC 1756, NGC 1786, and NGC 1795. These GCs were not identified in the results of the final pipeline. The pheromone values across the rasters containing these GCs are distributed erratically by the ants without sufficient concentration. As a result, no cluster gets singled out by the clustering phase. However, the full pipeline does identify many clusters contained within Area 3. These clusters are primarily on the fringes in rasters that do not contain many stars. Many of them are contained within the Dec bound of $-90.0°$ and $-86.0°$. These rasters represent the least busy regions in Area 3.

Table 6.2 lists the number of stars contained within each raster that was examined and presents additional statistics on the pheromone values associated with that raster.

Table 6.2: Visualization of the Star and Pheromone (**f**) Information of Each Raster in Figure 6.4.

Figure	N_{stars}	$\mathbf{f}_{\text{max}}$	$\mathbf{f}_{\text{mean}}$
6.4b NGC 4147 (A1)	26 778	2.513×10^{-2}	3.060×10^{-4}
6.4d Palomar 5 + M5 (A1)	76 687	5.649×10^{-3}	1.069×10^{-4}
6.4f M71 (A2: $4.0° \times 4.0°$)	3 246 820	5.482×10^{-3}	2.524×10^{-6}
6.4h M71 (A2: $2.0° \times 2.0°$)	771 759	5.944×10^{-3}	1.062×10^{-5}
6.4j Small Magellanic Cloud (A3)	301 721	1.430×10^{-2}	2.716×10^{-5}
6.4l Large Magellanic Cloud (A3)	576 494	9.693×10^{-3}	1.421×10^{-5}
6.4n Andromeda Galaxy (A4)	97 875	7.633×10^{-3}	8.372×10^{-5}
6.4p Triangulum Galaxy (A4)	66 763	7.941×10^{-3}	1.227×10^{-4}

What can be observed in this table is that the mean pheromone value ($\mathbf{f}_{\text{mean}}$) from the rasters of Area 2 and Area 3 are much smaller than the values of Area 1 and Area 4. This again emphasizes the relationship of the number of stars and the behavior of the Ant Colony algorithm expressed with the pheromone values.

Ultimately, the Ant Colony algorithm demonstrates effective behavior in specific circumstances (namely rasters with a large variation in stellar density across the raster). However, the algorithm requires further optimization to function as effectively on rasters with a more uniform distribution and those containing many stars.

What are the shortcomings of the Ant Colony algorithm?
There are two major shortcomings in the implementation of the Ant Colony algorithm. Firstly, it is possible that the ants are unable to evaluate their environment sufficiently. This is evident from the inability of the Ant Colony to successfully identify any clusters in Area 2: $4.0° \times 4.0°$. The configuration of the parameters of the Ant Colony play the primary factor in determining the scale at which the Ant Colony is able to successfully operate. It is crucial to identify this relationship between these parameters and the properties of the rasters being evaluated given the central nature of the Ant Colony algorithm to the functioning of the whole pipeline.

Secondly, the Ant Colony algorithm computes the Euclidean Distance between stars in RA, Dec, and Distance. However, this assumes that the basis units of each of these axes are uniform. While this holds between RA and Dec as they are in the same unit, it does not hold with distance. It may be possible to determine some relationship between the RA, Dec, and Distance which operates functionally.

How could the Ant Colony algorithm be improved?

Two main suggestions for improving the Ant Colony algorithm are:

1. Determining the relationship between the configuration constants for the Ant Colony algorithm and the characteristics of the input rasters. The aim would be to identify a method of dynamically setting these constants based on the raster. This information could then also be forwarded alongside the pheromone results to the clustering algorithm.

2. As with the same improvement mentioned for BlobDoG, introducing a rasterization scheme which also cuts across distance. This would make it easier for the Ant Colony algorithm to control the number of stars being processed per raster and thereby improve the uniformity of the behavior of the Ant Colony.

6.3 Gravitational Clustering Based on Pheromone Mapping

What is the behavior displayed by the Gravitational Clustering algorithm?

To explore the behavior of the clustering algorithm, plots are generated to compare the distribution of pheromone values across a raster against the resulting clustering. Figure 6.5 shows such a comparison for the raster containing the GC NGC 4147.
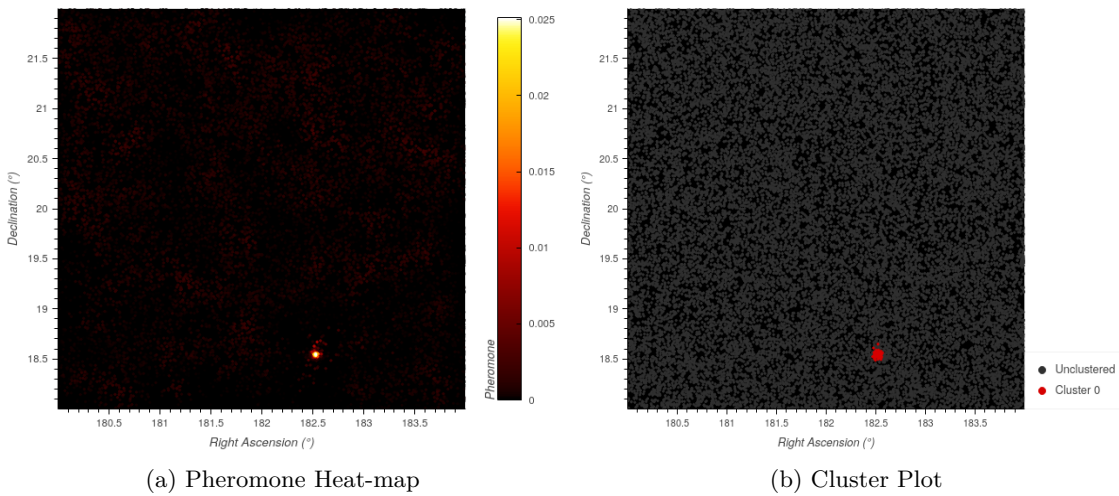

(a) Pheromone Heat-map

(b) Cluster Plot

Figure 6.5: 2D Plots of NGC 4147

The distribution of the pheromone values presented in Figure 6.5a are precisely represented in the clustering presented in Figure 6.5b. These plots are effective at showcasing the clustering across RA and Dec. However, unlike BlobDoG, the Ant Colony algorithm and the Gravitational Clustering algorithm operate by taking into account a third distance dimension. The comparison showing this distance dimension may be seen in Figure 6.6.

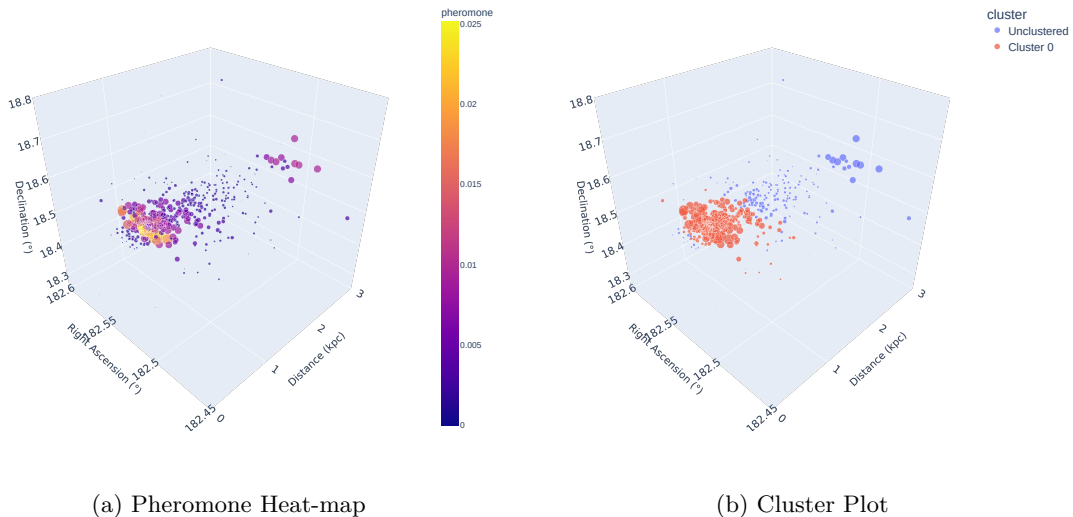(a) Pheromone Heat-map

(b) Cluster Plot

Figure 6.6: 3D Plots of NGC 4147

Figure 6.6 shows 3D scatterplots zoomed-in on the area around the cluster. The size of each point in the scatterplot corresponds to the magnitude of the pheromone value (as does the color in the case of the pheromone heat-map). It is clear from these figures, that the center of the cluster contains the stars with the highest pheromone values. Additionally, a separate group with high pheromone values may be observed within the distance of 2 kpc to 3 kpc. The main grouping evidently contains more than 100 stars whereas this other group contains less than 100 stars and is not classified as a cluster.

It is interesting to consider the scenario where the clustering algorithm does identify multiple clusters within the same raster. This is case for the raster containing the GC NGC 5024 and the 2D plot of the comparison between the pheromones values for this raster and the clustering may be seen in Figure 6.7.

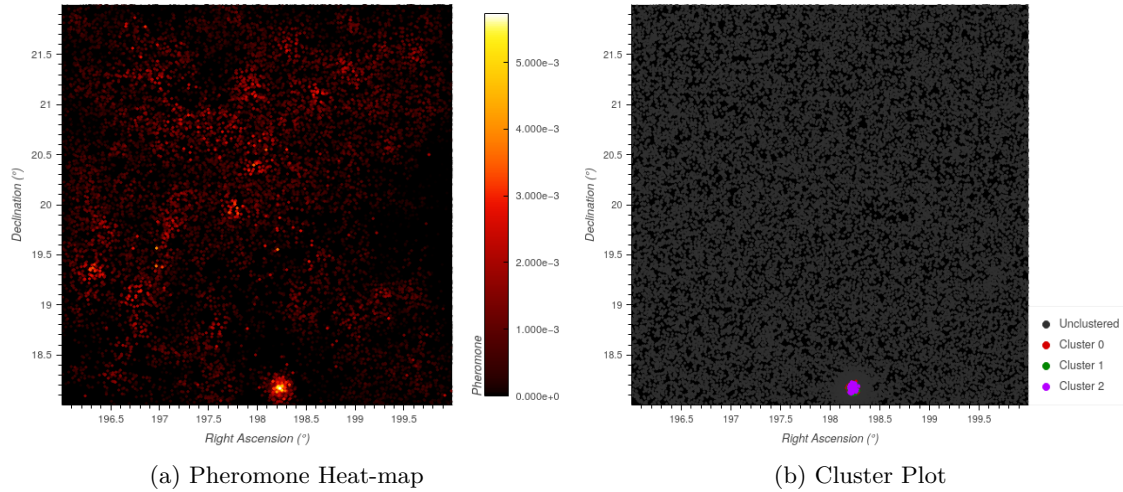(a) Pheromone Heat-map

(b) Cluster Plot

Figure 6.7: 2D Plots of NGC 5024

In Figure 6.7, only the results for *Cluster 2* seem apparent. This is because the clusters overlap in RA and Dec and this is confirmed when considering the plots in Figure 6.8.
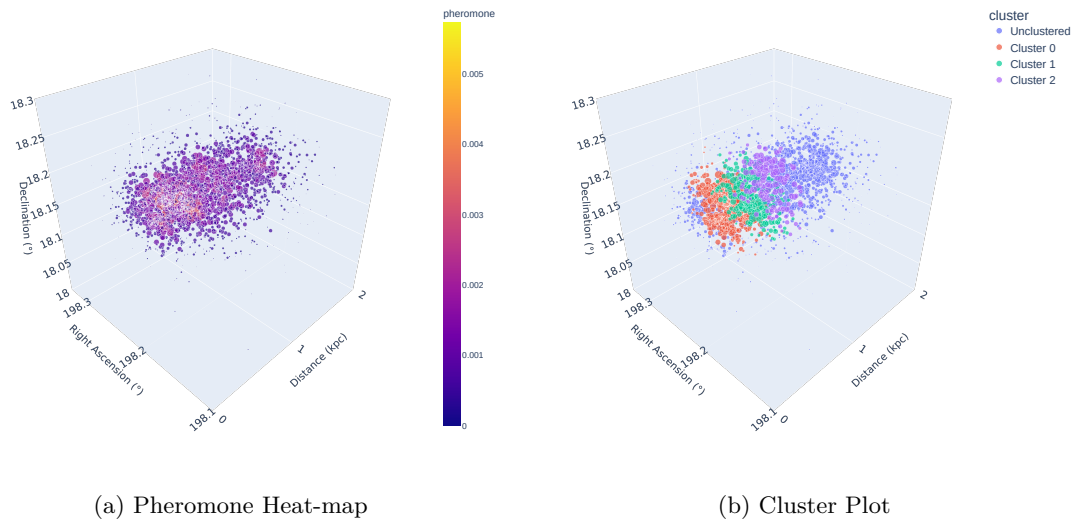

(a) Pheromone Heat-map

(b) Cluster Plot

Figure 6.8: 3D Plots of NGC 5024

The pheromone values seem uniformly distributed over the zoomed-in range displayed in Figure 6.8a. When comparing this region to the region presented in Figure 6.5 there seem to be many more stars present. Additionally, unlike Figure 6.5 it is hard to visually distinguish any sub-clusters across this range. These pheromone values appear as a single agglomerate structure. However, the clustering algorithm does distinguish this group into multiple clusters. It is evident from Figure 6.8b that the clustering algorithm identifies 3 distinct clusters around the center of mass of the pheromone values. These clusters appear as thin slices all in close proximity to each other with

no apparent unclustered regions interlacing them. This indicates an issue in the behavior of the clustering algorithm.

What are the shortcomings of the Gravitational Clustering algorithm?

From Figure 6.8, it becomes apparent that the clustering algorithm functions well in RA and Dec but seems miscalibrated under distance. This is due to the fact that the basis unit for RA, Dec, and distance are treated as the same unit. Thus, a $1°\!.0$ shift in RA or Dec is considered equivalent to a $1.0\,\mathrm{kpc}$ shift in distance. However, in reality this comparison is not physically sensible and it is evident from the results that the relative scaling for these bases are inappropriate. This same issue pervades the Ant Colony as well as it also makes use of Euclidean Distance.

Additionally, the fact that clusters are being split into multiple clusters under distance means that the sizes of the resulting cluster are underestimated. The parameter $N_{\mathrm{GC_{min}}}$ is set to a constant of 100. Thus, it is possible that a cluster whose total number of stars is above 100 has been filtered away as it was split into multiple clusters. This, likely explains why some clusters were identified only in $^4\!/\!_5$ experiments. Instead of making a binary decision, it would be practical to provide a continuous metric describing the validity of a given cluster.

Finally, more experimentation is required to determine an optimal value for $F_{\mathrm{min\ attraction}}$. This evaluation should also consider possible relationships between the properties of the raster and the constants that were used in previous phases of the pipeline.

How could the Gravitational Clustering algorithm be improved?

There are a number of possible improvements that could be considered for the Gravitational Clustering algorithm:

1. More testing must be done to determine any possible parametric relationships between the constants set in the previous phases of the pipeline and the constants used in the clustering phase.

2. The Ant Colony algorithm and the Gravitational Clustering algorithm both make use of Euclidean Distance, which treats the scale of the basis units for RA, Dec, and Distance identically. However, the distance operates with different units to the RA and Dec and this results in miscalibration of the algorithm in considering the distance. This could be temporarily alleviated by special casing the distance metric in the computations or by evaluating another distance metric perhaps under a different geometric scheme (e.g. spherical coordinates).

3. After the initial clusters have been formed, the clustering algorithm makes a binary decision based on the number of stars within the cluster. This decision determines whether or not that cluster should be classified as a valid cluster. This is mainly to remove small clusters that were formed containing very few stars (approaching one star). Instead, it would be appropriate to provide a continuous metric on the certainty of a given cluster being valid. This, step could also evaluate other properties such as metallicity. These metrics could also be used to differentiate between different cluster types.

Appendix A

Extended BlobDoG Tables

Table A.1: Extended BlobDoG Cluster Removals

Type	Name	RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiments
			Area 1			
OC	Golden-Eye Cluster	132°85	11°48′31″41	132°54 − 133°14	11°30′49″94 − 12°09′28″79	1 2 3 4 5
Galaxy	NGC 3286	158°86	58°20′16″71	158°58 − 159°21	58°01′42″82 − 58°39′16″97	4
	NGC 3288					
Galaxy	NGC 3770	174°06	59°12′18″36	173°74 − 174°35	58°52′39″72 − 59°32′26″42	2
	NGC 3795					
Galaxy	NGC 3838	176°51	58°31′52″81	176°10 − 177°04	58°03′39″24 − 59°02′00″30	1 2 3 4 5
	PGC 36398					
	PGC 36585					
	PGC 36655					
	PGC 36877					
Galaxy	Box Galaxy	187°05	44°07′25″99	186°65 − 187°53	43°45′34″72 − 44°26′31″10	1 2 3 4 5
Galaxy	Whirlpool Galaxy	202°46	47°11′30″37	202°08 − 202°75	46°53′49″64 − 47°36′14″96	1 2 3 5
Galaxy	Pinwheel Galaxy	210°79	54°21′33″69	210°55 − 211°07	54°08′42″75 − 54°36′48″15	1 2 3 4 5
			Area 3			
Nothing	N.A.	01°37	−86°17′02″66	00°77 − 01°99	−86°50′12″81 − −86°00′10″72	1 2 3 4 5
Galaxy	PGC 993	07°22	−86°32′25″26	06°51 − 07°91	−87°08′00″46 − −86°00′22″38	2 3 4 5
Nothing	N.A.	08°62	−86°30′45″61	08°10 − 09°19	−87°02′11″48 − −86°04′45″38	1 2 3 4 5
Galaxy	PGC 3533	12°54	−86°30′27″51	12°01 − 13°27	−87°09′21″00 − −86°00′03″34	1 2 3 4 5
Galaxy	PGC 216800	13°58	−87°04′17″60	13°16 − 14°05	−87°30′06″74 − −86°33′10″10	1 2 3 4 5
Nothing	N.A.	17°19	−86°27′27″30	16°46 − 18°12	−87°19′04″59 − −86°00′08″79	1 2 3 4 5
Galaxy	PGC 5780	25°76	−86°27′20″33	25°10 − 26°38	−87°12′58″21 − −86°00′07″08	1 2 3 4 5
Nothing	N.A.	28°42	−86°23′11″26	28°00 − 29°01	−87°02′35″83 − −86°00′56″14	1 2 3 4 5
Galactic Cluster	Abell 3037	36°97	−86°20′57″40	36°50 − 37°48	−86°47′35″26 − −86°00′32″54	1 2 3 4
Nothing	N.A.	41°02	−86°35′40″26	40°22 − 41°82	−87°21′16″69 − −86°00′02″30	1 2 3 4 5
Nothing	N.A	50°74	−86°25′15″67	50°23 − 51°25	−86°54′33″20 − −86°01′02″53	1 2 3 4 5
Nothing	N.A.	53°58	−86°21′59″43	52°98 − 54°19	−86°56′14″36 − −86°00′42″69	2 3 4
Nothing	N.A.	61°90	−86°18′11″33	61°35 − 62°47	−86°45′54″25 − −86°00′05″26	1 3 4 5
Nothing	N.A.	62°74	−86°32′06″73	61°68 − 63°45	−87°10′30″65 − −86°00′22″24	1 3 4 5
Nothing	N.A.	69°40	−86°39′36″30	68°57 − 70°21	−87°33′40″20 − −86°00′27″49	1 2 3 4 5
Nothing	N.A.	71°48	−86°24′13″20	70°93 − 71°97	−86°54′45″20 − −86°01′34″43	4 5
Galactic Cluster	Abell 3333	73°44	−86°37′29″98	72°61 − 74°50	−87°32′24″77 − −86°00′15″80	1 2 3 4 5

Table A.2: Extended Clusters Found in All Experiments

Area 1

Type	Name	RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiments
Dwarf Galaxy	Regulus Dwarf Galaxy	152°12	12°18′02″45	152°00 – 152°25	12°09′58″35 – 12°26′21″46	1 2 3 4 5
GC	NGC 4147	182°53	18°32′29″50	182°46 – 182°67	18°19′57″04 – 18°38′49″73	1 2 3 4 5
GC	NGC 5024	198°23	18°10′02″92	198°15 – 198°31	18°05′47″74 – 18°15′37″24	1 2 3 4 5
GC	NGC 5053	199°11	17°41′59″43	199°03 – 199°20	17°37′41″37 – 17°45′53″90	1 2 3 4
GC	M3	205°55	28°22′26″78	205°44 – 205°65	28°15′48″56 – 28°27′32″13	1 2 3 4 5
GC	NGC 5466	211°36	28°32′00″26	210°98 – 211°47	28°24′58″82 – 28°38′44″69	1 2 3 4 5
GC	Palomar 5	229°61	01°57′47″14	229°51 – 229°72	01°53′35″19 – 01°59′57″80	3
GC	M5	229°64	02°03′26″66	229°59 – 229°71	02°00′09″53 – 02°07′22″18	2

Area 2

Type	Name	RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiments
Nothing	N.A.	295°06	19°09′17″96	295°01 – 295°12	19°05′18″69 – 19°14′36″41	1
Nothing	N.A.	295°29	18°29′05″08	295°21 – 295°41	18°22′57″86 – 18°34′12″04	5
Nothing	N.A.	296°00	17°23′27″66	295°91 – 296°09	17°17′42″90 – 17°29′03″90	5
Nothing	N.A.	296°02	22°43′14″11	295°90 – 296°17	22°37′14″73 – 22°48′57″44	3
Nothing	N.A.	296°05	15°19′35″71	295°94 – 296°13	15°12′51″08 – 15°29′18″70	4
Nothing	N.A.	296°17	22°11′23″46	296°09 – 296°26	22°06′18″71 – 22°15′22″18	4
Nothing	N.A.	296°22	17°31′32″99	296°15 – 296°31	17°25′16″15 – 17°38′16″47	3
Molecular Cloud	LDN 758	296°61	19°08′40″17	296°51 – 296°71	19°04′57″81 – 19°12′23″21	3
OC	NGC 6827	297°60	21°21′28″57	297°50 – 297°68	21°18′05″35 – 21°25′13″22	2
Nothing	N.A.	297°17	23°52′08″87	297°10 – 297°26	23°46′57″79 – 23°56′59″35	1
Nothing	N.A.	297°72	15°37′39″26	297°63 – 297°82	15°28′41″39 – 15°44′36″90	4
Absorption Nebula	LDN 787	298°22	22°12′59″69	298°15 – 298°30	22°05′50″14 – 22°20′26″79	5
GC	M71	298°44	18°46′27″05	298°40 – 298°49	18°42′30″76 – 18°50′02″53	1 4
Nothing	N.A.	299°26	18°34′53″48	299°19 – 299°34	18°28′12″09 – 18°42′43″59	5
Nothing	N.A.	299°71	20°07′38″33	299°65 – 299°77	20°02′01″45 – 20°13′46″18	4
Nothing	N.A.	300°30	18°17′28″09	300°22 – 300°36	18°10′37″66 – 18°25′39″34	3
Nothing	N.A.	301°47	19°52′50″56	301°38 – 301°57	19°48′32″35 – 19°57′00″22	3
Nothing	N.A.	302°15	23°29′50″72	302°01 – 302°24	23°21′47″17 – 23°37′09″98	5

Area 3

Type	Name	RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiments
Galaxy	String of Pearls	03°78	−39°12′55″81	03°55 – 04°00	−39°19′58″48 – −39°06′32″19	1 2 3 4 5
Galaxy	Southern Pinwheel Galaxy	13°73	−37°39′55″00	13°33 – 14°00	−37°53′52″16 – −37°06′18″28	1 2 3 4 5
Dwarf Galaxy	Sculptor Dwarf Galaxy	15°03	−33°43′18″88	14°73 – 15°29	−33°53′59″04 – −33°30′35″11	1 2 3 4 5
OC	NGC 419	17°07	−72°52′56″38	17°01 – 17°12	−72°54′46″05 – −72°50′33″55	2
GC	NGC 1049	39°88	−34°32′06″15	39°59 – 40°00	−34°48′34″74 – −34°15′17″86	1 2 3 4 5
Galaxy	Fornax Dwarf Galaxy	40°08	−34°30′12″32	40°00 – 40°28	−34°43′31″34 – −34°13′02″70	1 2 3 4 5
GC	NGC 1261	48°07	−55°12′59″17	48°00 – 48°14	−55°17′01″45 – −55°10′23″89	1 2 3 4 5
GC	NGC 1466	56°14	−71°40′10″54	56°02 – 56°34	−72°04′39″73 – −71°27′11″21	1 3 4 5
GC	NGC 1841	71°37	−84°00′38″52	70°77 – 71°80	−84°28′49″91 – −83°24′22″53	1 2 3 4 5
OC	NGC 1777	73°95	−74°16′45″35	73°77 – 74°12	−74°28′29″07 – −74°03′19″63	1 2 3 4 5
GC	NGC 1783	74°79	−65°58′40″00	74°71 – 74°89	−65°59′58″83 – −65°49′10″26	1 2 3 5

Area 4

Type	Name	RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiments
HII region	NGC 7822	00°55	67°25′37″17	00°40 – 00°75	67°16′53″23 – 67°34′07″21	1 2 3 4
Galaxy	Triangulum Galaxy	23°47	30°36′49″92	23°31 – 23°61	30°29′08″78 – 30°44′38″01	1 2 3 4 5
OC	Double Cluster	35°52	57°08′26″09	35°43 – 35°63	57°03′30″07 – 57°13′46″71	3
OC	NGC 1027	40°69	61°35′27″79	40°58 – 40°79	61°29′39″61 – 61°41′45″61	4

Appendix B

Raster Plots of the Full Pipeline

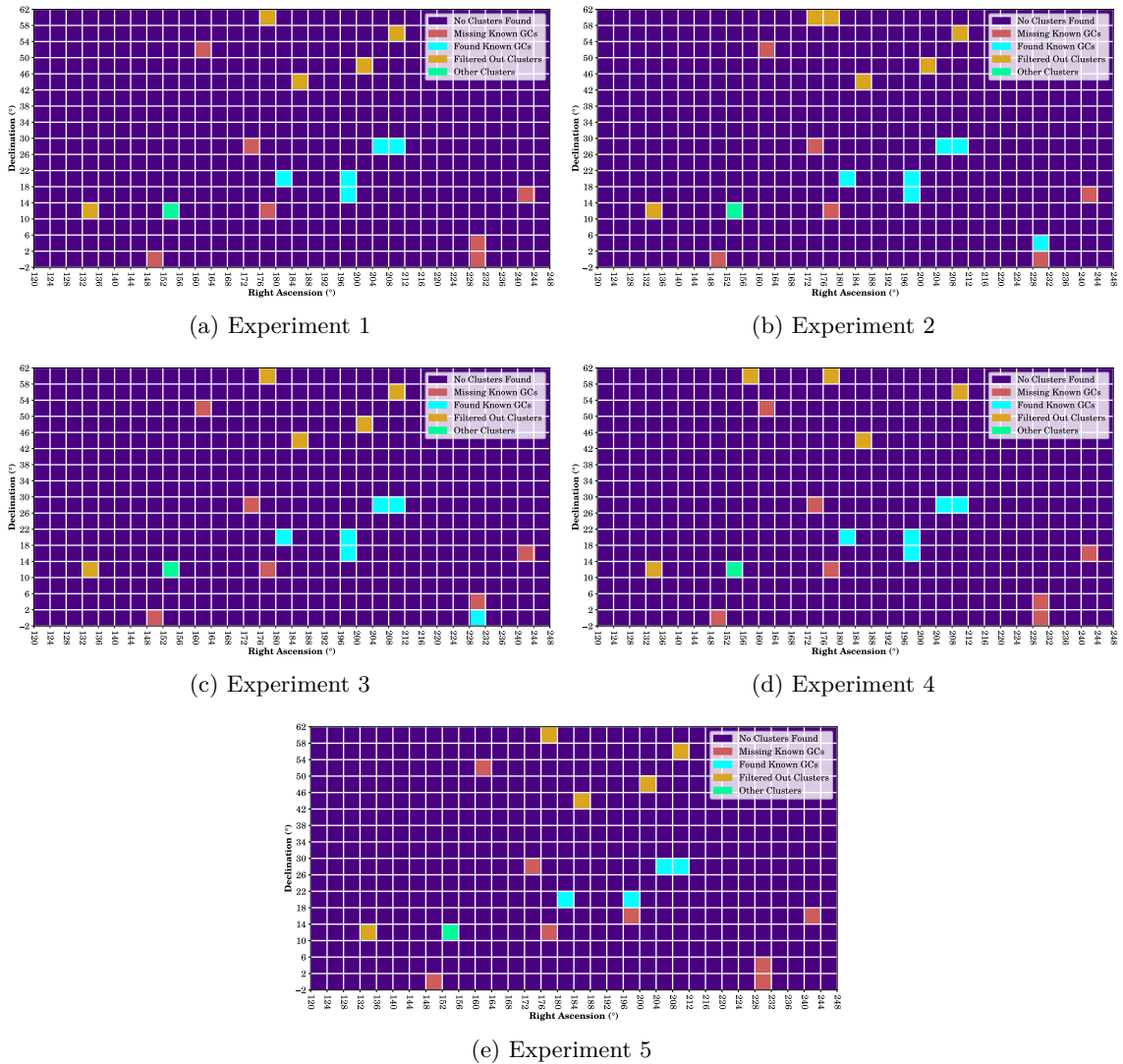Figure B.1: Clustering Results Across A1 for All Experiments

(a) Experiment 1

(b) Experiment 2

(c) Experiment 3

(d) Experiment 4

(e) Experiment 5
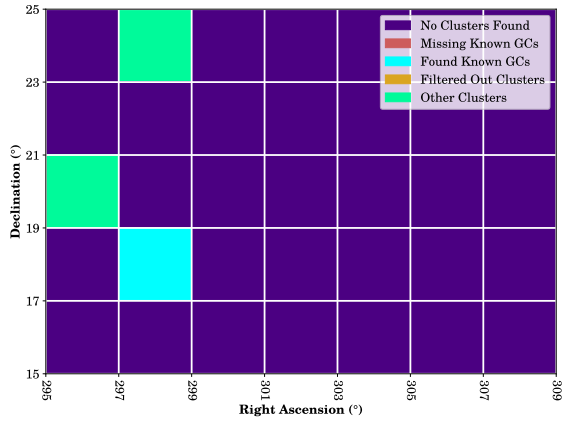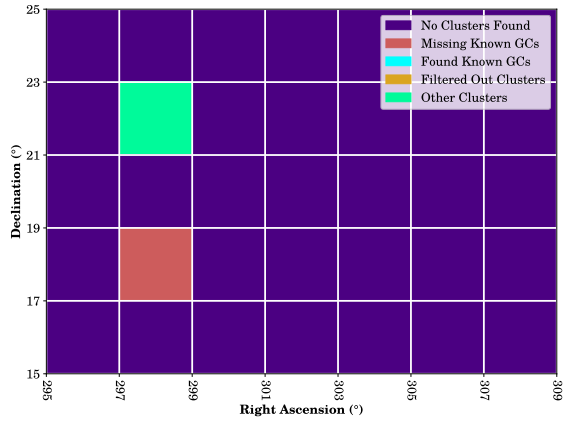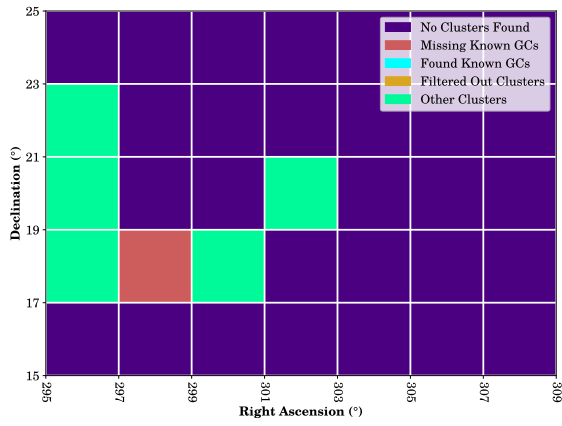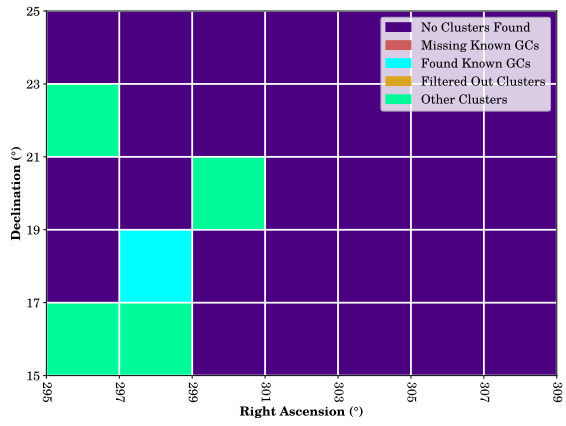
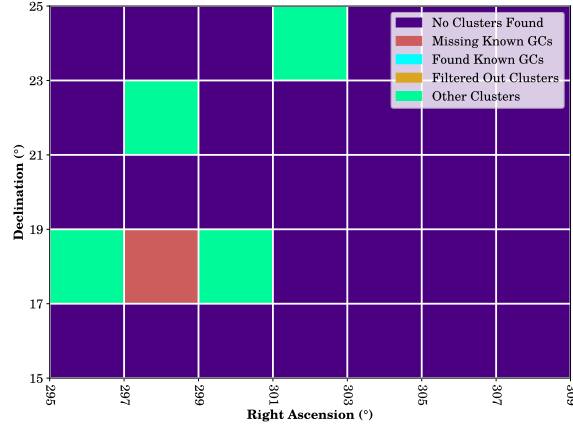Figure B.2: Clustering Results Across A2: $2.0° \times 2.0°$ for All Experiments

(a) Experiment 1

(b) Experiment 2

(c) Experiment 3

(d) Experiment 4
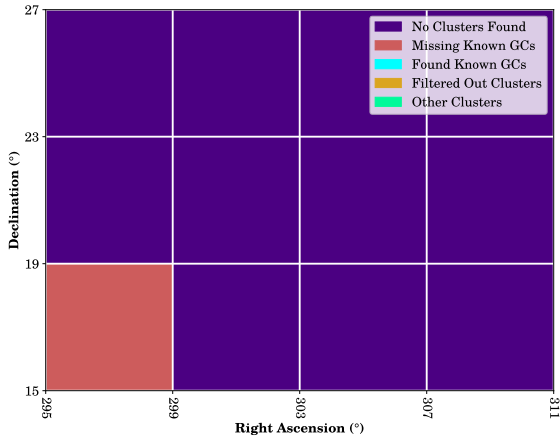
(e) Experiment 5

Figure B.3: Clustering Results Across A2: 4.0° × 4.0° for All Experiments

(a) Experiment 1

(b) Experiment 2

(c) Experiment 3

(d) Experiment 4
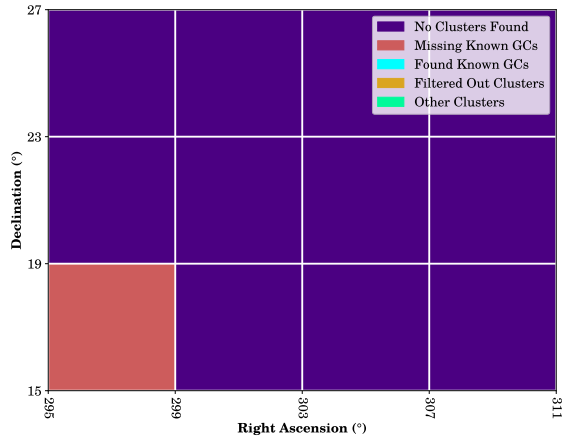
(e) Experiment 5

Figure B.4: Clustering Results Across A3 for All Experiments

(a) Experiment 1

(b) Experiment 2

(c) Experiment 3

(d) Experiment 4

(e) Experiment 5

Figure B.5: Clustering Results Across A4 for All Experiments

(a) Experiment 1

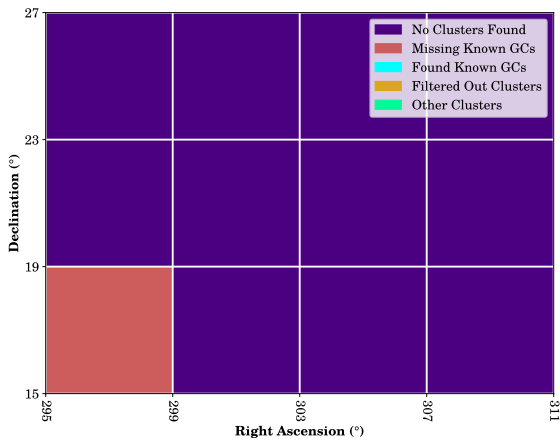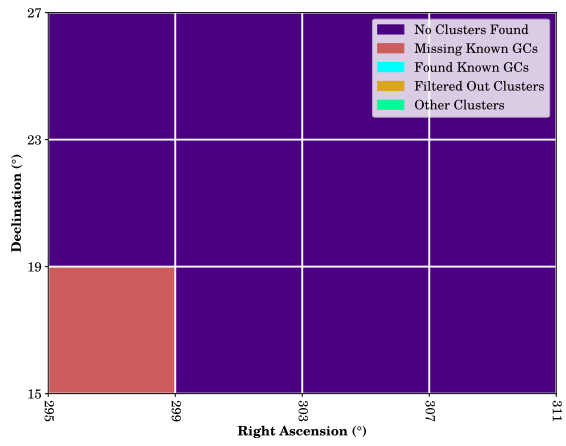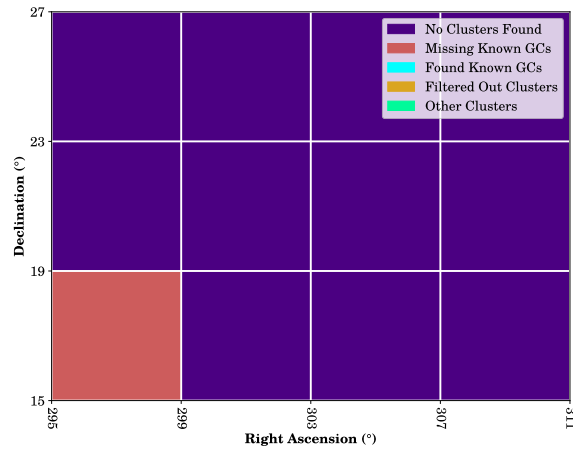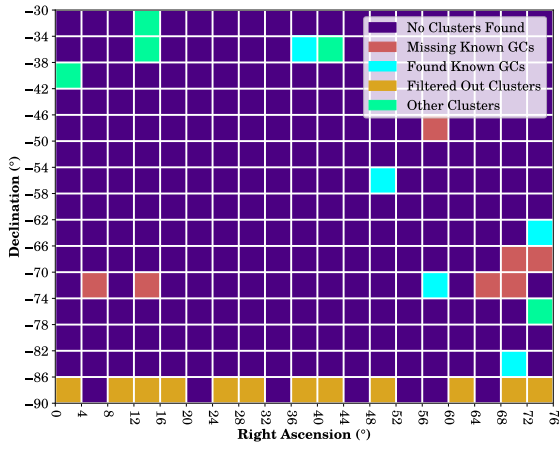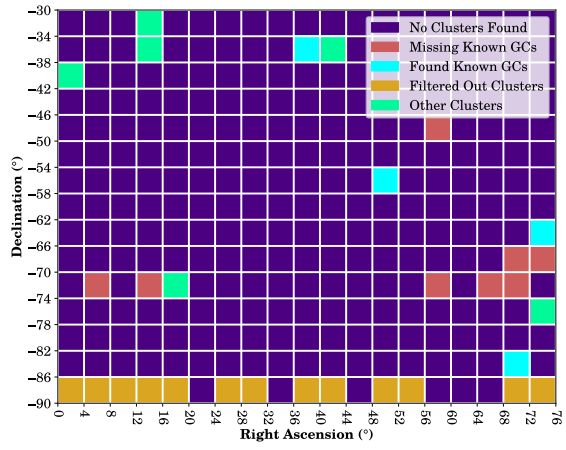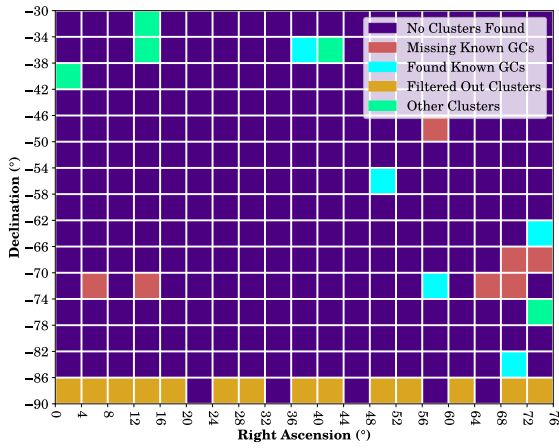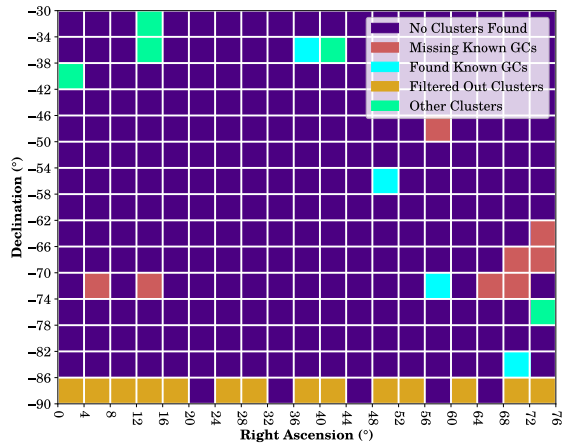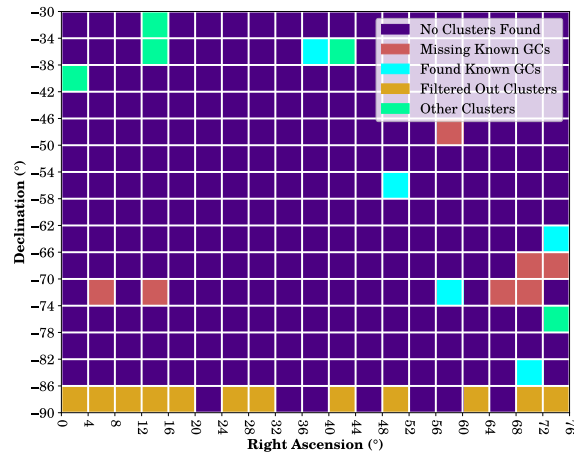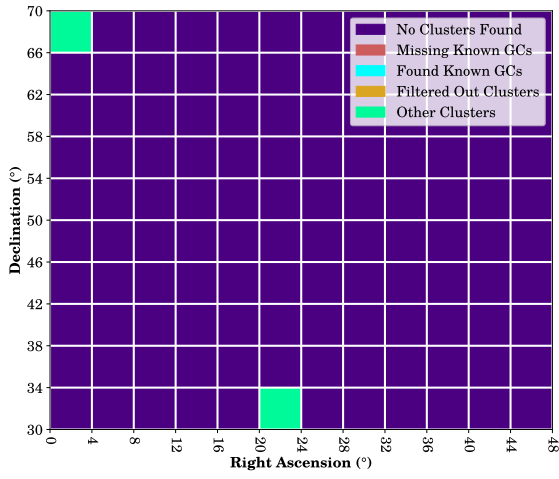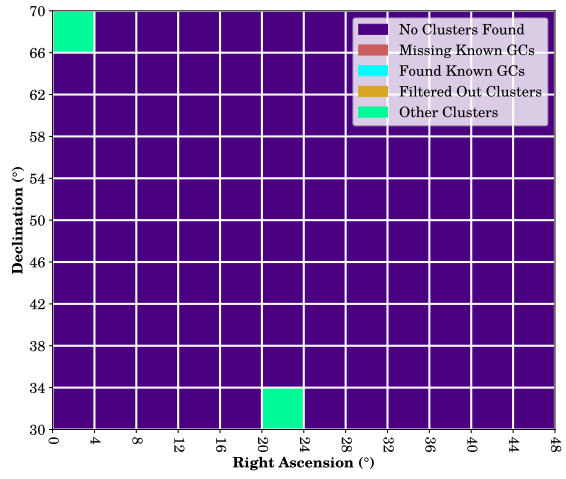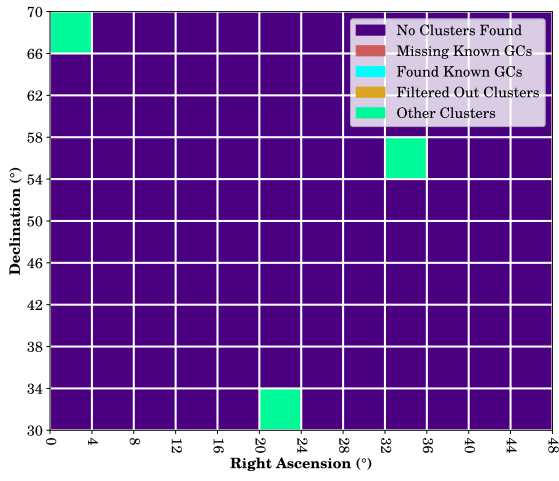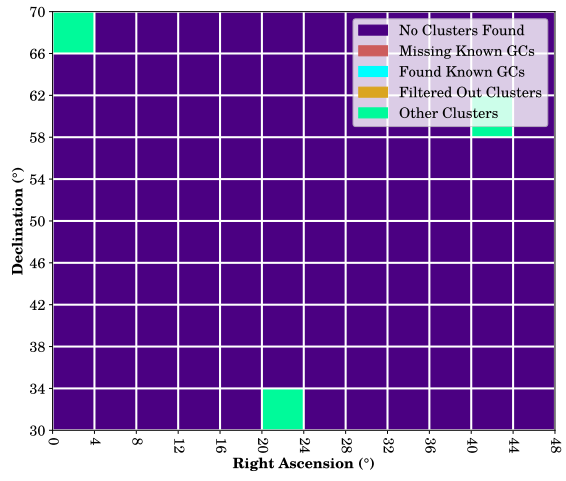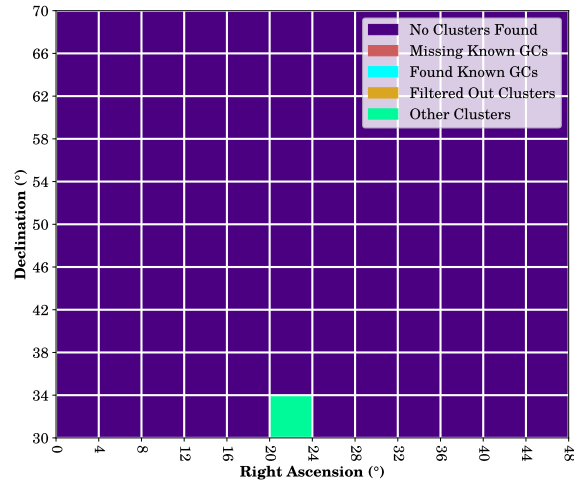(b) Experiment 2

(c) Experiment 3

(d) Experiment 4

(e) Experiment 5

Clusters Found Across All Experiments

Table C.1: Area 1: All Clusters Found Across All Experiments

RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiment
132°85	11°48′04″56	132°64 − 133°09	11°33′07″28 − 12°01′31″53	1
132°85	11°48′06″92	132°61 − 133°10	11°30′49″94 − 12°09′28″79	2
132°85	11°49′08″74	132°64 − 133°10	11°34′28″69 − 12°04′18″96	3
132°85	11°49′06″97	132°64 − 133°10	11°33′07″28 − 12°09′03″91	4
132°85	11°48′09″85	132°54 − 133°14	11°30′56″18 − 12°07′07″47	5
152°12	12°18′06″04	152°00 − 152°25	12°11′00″17 − 12°24′07″58	1
152°12	12°17′56″10	152°01 − 152°24	12°10′07″97 − 12°24′05″15	2
152°12	12°18′06″39	152°00 − 152°25	12°09′58″35 − 12°26′21″46	3
152°12	12°18′00″67	152°00 − 152°25	12°10′07″97 − 12°26′21″46	4
152°12	12°18′03″03	152°00 − 152°25	12°09′58″35 − 12°26′21″46	5
158°86	58°20′16″71	158°58 − 159°21	58°01′42″82 − 58°39′16″97	4
174°06	59°12′18″36	173°74 − 174°35	58°52′39″72 − 59°32′26″42	2
176°50	58°32′04″48	176°18 − 176°94	58°08′47″54 − 58°57′48″63	1
176°50	58°29′44″00	176°18 − 176°89	58°08′15″42 − 58°54′52″21	2
176°53	58°31′39″37	176°10 − 177°04	58°03′39″24 − 59°02′00″30	3
176°50	58°31′59″53	176°18 − 176°89	58°10′05″86 − 58°57′38″37	4
176°51	58°33′56″67	176°18 − 176°95	58°08′47″54 − 59°02′00″30	5
182°53	18°32′25″99	182°46 − 182°67	18°19′57″04 − 18°38′49″73	1
182°53	18°32′31″81	182°47 − 182°57	18°29′10″58 − 18°38′49″73	2
182°53	18°32′31″08	182°47 − 182°59	18°26′52″72 − 18°38′49″73	3
182°53	18°32′28″01	182°47 − 182°57	18°26′52″72 − 18°35′16″31	4
182°53	18°32′30″61	182°47 − 182°57	18°29′10″58 − 18°35′16″31	5
187°04	44°08′16″25	186°77 − 187°20	43°57′53″17 − 44°21′39″41	1
187°06	44°05′33″51	186°65 − 187°53	43°45′34″72 − 44°26′31″10	2
187°04	44°08′29″61	186°77 − 187°20	43°57′59″46 − 44°21′39″41	3
187°04	44°07′14″98	186°72 − 187°47	43°45′34″72 − 44°24′34″75	4
187°06	44°07′35″58	186°84 − 187°47	43°49′33″00 − 44°21′12″38	5
199°11	17°41′53″28	199°06 − 199°18	17°39′09″54 − 17°44′47″28	1
199°11	17°42′02″20	199°03 − 199°20	17°37′41″37 − 17°45′38″77	2
199°11	17°41′56″40	199°06 − 199°18	17°38′08″72 − 17°45′53″90	2
199°11	17°42′04″44	199°06 − 199°16	17°39′16″08 − 17°45′00″74	3
199°11	17°42′00″84	199°06 − 199°18	17°37′49″95 − 17°45′53″90	4
198°23	18°10′11″42	198°16 − 198°29	18°06′30″66 − 18°14′18″80	1
198°23	18°09′47″85	198°17 − 198°29	18°05′58″06 − 18°13′27″93	1
198°23	18°10′19″39	198°18 − 198°29	18°06′22″83 − 18°13′56″37	1
198°23	18°10′13″23	198°17 − 198°28	18°06′46″77 − 18°13′21″44	2
198°24	18°09′36″75	198°19 − 198°28	18°06′54″88 − 18°12′36″07	2
198°23	18°10′10″90	198°18 − 198°28	18°07′01″82 − 18°13′01″76	3
198°23	18°09′48″92	198°19 − 198°28	18°06′51″88 − 18°13′27″93	3
198°23	18°10′05″12	198°15 − 198°31	18°05′47″74 − 18°15′37″24	4

198°23	18°10′12″74	198°16 – 198°28	18°06′46″77 – 18°13′59″04	5
202°47	47°11′20″97	202°22 – 202°72	46°55′18″52 – 47°27′00″17	1
202°49	47°10′49″78	202°22 – 202°72	46°57′07″41 – 47°23′56″32	2
202°44	47°12′06″75	202°08 – 202°75	46°53′49″64 – 47°36′14″96	3
202°45	47°11′43″97	202°19 – 202°72	46°55′18″52 – 47°29′05″63	5
205°55	28°22′16″21	205°48 – 205°63	28°18′32″51 – 28°26′24″14	1
205°54	28°22′12″63	205°50 – 205°60	28°19′48″37 – 28°24′39″45	1
205°55	28°22′17″10	205°49 – 205°62	28°18′34″73 – 28°26′07″30	1
205°57	28°21′39″18	205°50 – 205°63	28°18′26″09 – 28°24′50″05	1
205°55	28°22′15″75	205°48 – 205°63	28°19′05″57 – 28°26′06″42	2
205°54	28°22′24″46	205°48 – 205°62	28°18′46″60 – 28°26′32″55	2
205°55	28°22′52″99	205°49 – 205°63	28°19′29″34 – 28°27′02″14	2
205°55	28°22′13″72	205°48 – 205°61	28°18′34″73 – 28°25′54″94	2
205°55	28°22′55″37	205°45 – 205°63	28°18′23″23 – 28°27′05″68	3
205°55	28°22′28″81	205°49 – 205°62	28°18′57″79 – 28°25′54″94	3
205°55	28°22′14″82	205°48 – 205°62	28°19′05″57 – 28°26′24″14	3
205°54	28°22′24″71	205°49 – 205°60	28°19′10″45 – 28°25′27″66	3
205°53	28°22′56″01	205°44 – 205°63	28°18′07″68 – 28°26′56″59	4
205°55	28°22′17″54	205°48 – 205°63	28°18′26″10 – 28°26′41″83	4
205°55	28°22′53″39	205°47 – 205°63	28°19′00″52 – 28°27′05″68	4
205°55	28°22′33″42	205°48 – 205°62	28°18′34″73 – 28°26′41″54	4
205°55	28°22′10″70	205°48 – 205°62	28°18′48″76 – 28°25′57″14	4
205°55	28°22′47″98	205°46 – 205°65	28°17′59″86 – 28°27′32″13	5
205°55	28°22′22″16	205°46 – 205°65	28°15′48″56 – 28°27′06″97	5
205°56	28°22′42″33	205°48 – 205°63	28°18′29″32 – 28°26′53″07	5
205°54	28°22′03″74	205°49 – 205°61	28°19′28″82 – 28°24′30″13	5
205°55	28°22′46″17	205°48 – 205°63	28°18′11″89 – 28°26′58″28	5
211°36	28°31′59″63	211°26 – 211°45	28°27′03″42 – 28°36′29″84	1
211°36	28°32′04″20	210°98 – 211°47	28°24′58″82 – 28°38′44″69	2
211°36	28°31′48″32	211°30 – 211°42	28°28′07″81 – 28°35′21″12	2
211°36	28°31′50″57	211°29 – 211°45	28°27′41″92 – 28°35′23″77	3
211°37	28°31′57″40	211°29 – 211°44	28°27′40″08 – 28°35′47″21	3
211°37	28°32′03″89	211°30 – 211°43	28°27′22″15 – 28°35′38″87	3
211°36	28°32′02″31	211°29 – 211°45	28°27′22″15 – 28°36′37″66	4
211°36	28°32′05″62	211°31 – 211°42	28°29′13″36 – 28°35′21″12	4
211°36	28°32′04″79	211°31 – 211°44	28°28′31″06 – 28°36′04″44	5
211°37	28°32′05″86	211°30 – 211°43	28°28′14″20 – 28°35′38″87	5
210°79	54°21′26″55	210°57 – 210°99	54°09′13″67 – 54°34′09″66	1
210°78	54°21′28″86	210°55 – 211°03	54°09′13″67 – 54°36′48″15	2
210°79	54°21′46″21	210°57 – 211°07	54°08′42″75 – 54°36′48″15	3
210°79	54°21′36″21	210°57 – 211°03	54°10′56″12 – 54°34′09″66	4
210°79	54°21′30″62	210°55 – 211°03	54°09′13″67 – 54°36′48″15	5
229°61	01°57′47″14	229°51 – 229°72	01°53′35″19 – 01°59′57″80	3
229°64	02°03′26″66	229°59 – 229°71	02°00′09″53 – 02°07′22″18	2

Table C.2: Area 2: All Clusters Found Across All Experiments

RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiment
296°.05	15°19′35″.71	295°.94 – 296°.13	15°12′51″.08 – 15°29′18″.70	4
296°.22	17°31′32″.99	296°.15 – 296°.31	17°25′16″.15 – 17°38′16″.47	3
296°.00	17°23′27″.66	295°.91 – 296°.09	17°17′42″.90 – 17°29′03″.90	5
295°.29	18°29′05″.08	295°.21 – 295°.41	18°22′57″.86 – 18°34′12″.04	5
295°.06	19°09′17″.96	295°.01 – 295°.12	19°05′18″.69 – 19°14′36″.41	1
296°.61	19°08′40″.17	296°.51 – 296°.71	19°04′57″.81 – 19°12′23″.21	3
296°.02	22°43′14″.11	295°.90 – 296°.17	22°37′14″.73 – 22°48′57″.44	3
296°.17	22°11′23″.46	296°.09 – 296°.26	22°06′18″.71 – 22°15′22″.18	4
297°.72	15°37′39″.26	297°.63 – 297°.82	15°28′41″.39 – 15°44′36″.90	4
298°.44	18°46′44″.62	298°.40 – 298°.49	18°43′08″.71 – 18°50′02″.53	1
298°.45	18°46′09″.47	298°.40 – 298°.49	18°42′30″.76 – 18°49′38″.71	4
297°.60	21°21′28″.57	297°.50 – 297°.68	21°18′05″.35 – 21°25′13″.22	2
298°.22	22°12′59″.69	298°.15 – 298°.30	22°05′50″.14 – 22°20′26″.79	5
297°.17	23°52′08″.87	297°.10 – 297°.26	23°46′57″.79 – 23°56′59″.35	1
300°.30	18°17′28″.09	300°.22 – 300°.36	18°10′37″.66 – 18°25′39″.34	3
299°.26	18°34′53″.48	299°.19 – 299°.34	18°28′12″.09 – 18°42′43″.59	5
299°.71	20°07′38″.33	299°.65 – 299°.77	20°02′01″.45 – 20°13′46″.18	4
301°.47	19°52′50″.56	301°.38 – 301°.57	19°48′32″.35 – 19°57′00″.22	3
302°.15	23°29′50″.72	302°.01 – 302°.24	23°21′47″.17 – 23°37′09″.98	5

Table C.3: Area 3: All Clusters Found Across All Experiments

RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiment
03°.78	−39°12′55″.47	03°.58 – 04°.00	−39°19′58″.48 – −39°06′32″.19	1
03°.78	−39°12′55″.22	03°.55 – 04°.00	−39°19′58″.48 – −39°06′32″.19	2
03°.78	−39°12′54″.31	03°.57 – 04°.00	−39°19′58″.48 – −39°07′53″.63	3
03°.78	−39°12′54″.08	03°.58 – 04°.00	−39°19′58″.48 – −39°08′18″.42	4
03°.78	−39°12′59″.98	03°.60 – 03°.99	−39°18′44″.16 – −39°08′43″.04	5
01°.40	−86°15′32″.89	00°.87 – 01°.89	−86°37′55″.48 – −86°00′13″.53	1
01°.36	−86°16′43″.99	00°.87 – 01°.86	−86°46′08″.93 – −86°00′13″.53	2
01°.33	−86°16′09″.03	00°.77 – 01°.86	−86°44′27″.61 – −86°00′13″.53	3
01°.39	−86°19′02″.41	00°.77 – 01°.99	−86°50′12″.81 – −86°00′10″.72	4
01°.36	−86°17′45″.00	00°.77 – 01°.89	−86°46′08″.93 – −86°00′13″.53	5
15°.03	−33°43′14″.41	14°.75 – 15°.28	−33°53′59″.04 – −33°32′20″.92	1
15°.03	−33°43′22″.53	14°.77 – 15°.26	−33°53′59″.04 – −33°32′06″.49	2
15°.03	−33°43′34″.98	14°.77 – 15°.25	−33°53′46″.71 – −33°32′39″.88	3
15°.03	−33°43′19″.87	14°.77 – 15°.27	−33°53′59″.04 – −33°32′20″.92	4
15°.02	−33°43′02″.59	14°.73 – 15°.29	−33°53′54″.33 – −33°30′35″.11	5
13°.72	−37°40′02″.85	13°.35 – 13°.97	−37°53′05″.80 – −37°10′21″.15	1
13°.73	−37°39′44″.18	13°.41 – 14°.00	−37°53′52″.16 – −37°06′46″.06	2
13°.73	−37°39′41″.93	13°.48 – 13°.97	−37°50′22″.87 – −37°10′24″.93	3
13°.73	−37°40′16″.57	13°.38 – 14°.00	−37°50′22″.87 – −37°14′32″.03	4
13°.73	−37°39′49″.49	13°.33 – 13°.97	−37°52′30″.24 – −37°06′18″.28	5
13°.55	−87°04′15″.09	13°.16 – 14°.02	−87°30′06″.74 – −86°38′35″.39	1
12°.60	−86°31′06″.35	12°.01 – 13°.27	−87°09′21″.00 – −86°00′03″.34	1
13°.58	−87°05′13″.56	13°.17 – 14°.02	−87°30′06″.74 – −86°38′35″.39	2
12°.50	−86°30′11″.85	12°.01 – 12°.96	−86°59′10″.79 – −86°04′37″.05	2
13°.59	−87°03′51″.93	13°.16 – 14°.05	−87°30′06″.74 – −86°36′07″.06	3
12°.58	−86°31′24″.77	12°.01 – 13°.21	−87°06′08″.99 – −86°00′03″.34	3
13°.58	−87°03′29″.05	13°.16 – 14°.05	−87°30′06″.74 – −86°33′10″.10	4
12°.45	−86°28′57″.56	12°.01 – 12°.88	−86°58′57″.97 – −86°04′37″.05	4
13°.59	−87°04′38″.39	13°.17 – 14°.02	−87°30′06″.74 – −86°38′35″.39	5
12°.57	−86°30′37″.02	12°.01 – 13°.17	−87°04′26″.50 – −86°00′12″.68	5
17°.07	−72°52′56″.38	17°.01 – 17°.12	−72°54′46″.05 – −72°50′33″.55	2
17°.19	−86°27′08″.05	16°.60 – 17°.77	−86°58′46″.11 – −86°00′14″.19	1
17°.31	−86°32′02″.85	16°.46 – 18°.12	−87°19′04″.59 – −86°00′08″.79	2
17°.19	−86°26′05″.87	16°.60 – 17°.77	−86°58′46″.11 – −86°00′14″.19	3
17°.14	−86°25′36″.56	16°.55 – 17°.73	−87°01′22″.87 – −86°00′08″.79	4

17°11	−86°26′23″20	16°60 − 17°70	−86°58′46″11 −	−86°00′14″19	5
25°77	−86°24′48″89	25°26 − 26°31	−86°56′17″69 −	−86°00′07″08	1
25°77	−86°25′15″14	25°26 − 26°31	−86°56′58″19 −	−86°00′20″07	2
25°75	−86°30′02″36	25°10 − 26°38	−87°12′58″21 −	−86°00′07″08	3
25°76	−86°28′16″53	25°20 − 26°38	−87°05′04″74 −	−86°00′20″07	4
25°77	−86°28′18″73	25°10 − 26°38	−87°09′37″15 −	−86°00′07″08	5
28°43	−86°23′55″69	28°00 − 28°96	−86°56′33″55 −	−86°00′56″14	1
28°44	−86°22′47″53	28°00 − 28°96	−86°56′33″55 −	−86°00′56″14	2
28°48	−86°24′52″50	28°00 − 29°01	−87°02′35″83 −	−86°00′56″14	3
28°37	−86°22′00″38	28°00 − 28°80	−86°51′52″28 −	−86°00′59″93	4
28°37	−86°22′20″20	28°00 − 28°80	−86°51′52″28 −	−86°00′59″93	5
39°88	−34°32′38″24	39°69 − 40°00	−34°45′50″73 −	−34°19′42″98	1
39°88	−34°32′36″01	39°68 − 40°00	−34°44′45″63 −	−34°20′44″78	1
39°88	−34°31′55″80	39°64 − 40°00	−34°45′58″69 −	−34°17′14″90	2
39°86	−34°32′22″78	39°59 − 40°00	−34°48′34″74 −	−34°15′17″86	3
39°88	−34°31′19″18	39°66 − 40°00	−34°44′18″77 −	−34°19′32″50	4
39°89	−34°31′54″73	39°73 − 40°00	−34°42′27″74 −	−34°22′31″23	4
39°87	−34°31′56″34	39°64 − 40°00	−34°46′24″10 −	−34°17′14″90	5
36°95	−86°21′15″31	36°50 − 37°42	−86°47′35″26 −	−86°00′32″54	1
36°96	−86°21′01″91	36°50 − 37°44	−86°45′25″61 −	−86°00′32″54	2
36°99	−86°20′38″53	36°56 − 37°48	−86°45′25″61 −	−86°00′32″54	3
36°97	−86°20′53″84	36°50 − 37°48	−86°47′35″26 −	−86°00′32″54	4
07°24	−86°32′48″43	06°73 − 07°84	−87°03′56″94 −	−86°05′59″01	2
07°21	−86°31′38″72	06°51 − 07°91	−87°08′00″46 −	−86°00′22″38	3
07°21	−86°33′17″47	06°74 − 07°69	−86°58′28″01 −	−86°05′59″01	4
07°21	−86°31′56″40	06°53 − 07°91	−87°06′19″85 −	−86°01′06″38	5
40°09	−34°28′06″22	40°00 − 40°23	−34°39′08″35 −	−34°16′09″70	1
40°09	−34°31′26″30	40°00 − 40°22	−34°40′56″34 −	−34°21′52″31	1
40°07	−34°28′32″63	40°00 − 40°19	−34°37′01″63 −	−34°19′06″33	2
40°07	−34°31′24″25	40°00 − 40°18	−34°38′44″05 −	−34°23′32″68	2
40°07	−34°29′00″02	40°00 − 40°20	−34°37′59″60 −	−34°20′13″11	3
40°09	−34°31′25″81	40°00 − 40°23	−34°43′31″34 −	−34°21′12″65	3
40°09	−34°29′15″88	40°00 − 40°28	−34°43′18″87 −	−34°13′02″70	4
40°06	−34°33′06″33	40°00 − 40°17	−34°39′47″45 −	−34°25′52″01	4
40°07	−34°28′49″68	40°00 − 40°18	−34°37′26″65 −	−34°18′43″10	5
40°09	−34°30′56″11	40°00 − 40°23	−34°40′49″72 −	−34°21′53″81	5
41°03	−86°35′17″99	40°30 − 41°76	−87°14′56″40 −	−86°00′02″30	1
40°99	−86°36′50″35	40°31 − 41°67	−87°14′57″78 −	−86°00′43″39	2
41°02	−86°32′48″88	40°40 − 41°67	−87°09′33″33 −	−86°00′43″39	3
41°00	−86°37′08″04	40°22 − 41°82	−87°21′16″69 −	−86°00′02″30	4
41°03	−86°36′16″02	40°25 − 41°82	−87°18′03″88 −	−86°00′02″30	5
48°07	−55°13′13″34	48°01 − 48°12	−55°15′14″81 −	−55°10′35″09	1
48°07	−55°12′58″62	48°02 − 48°11	−55°14′55″60 −	−55°10′44″95	1
48°07	−55°12′57″70	48°02 − 48°12	−55°15′15″04 −	−55°10′44″95	2
48°07	−55°12′47″25	48°02 − 48°11	−55°14′35″01 −	−55°10′23″89	2
48°07	−55°13′12″45	48°01 − 48°12	−55°15′14″81 −	−55°10′42″13	2
48°07	−55°13′02″05	48°00 − 48°14	−55°16′16″37 −	−55°10′29″24	3
48°07	−55°12′51″10	48°01 − 48°12	−55°15′28″86 −	−55°10′23″89	3
48°07	−55°13′00″97	48°00 − 48°14	−55°17′01″45 −	−55°10′29″24	4
48°07	−55°12′54″68	48°01 − 48°13	−55°15′15″04 −	−55°10′23″89	4
48°07	−55°12′53″61	48°01 − 48°13	−55°15′28″86 −	−55°10′23″89	5
48°07	−55°12′59″11	48°02 − 48°11	−55°15′09″67 −	−55°10′29″24	5
50°75	−86°24′26″26	50°24 − 51°25	−86°50′48″56 −	−86°01′02″53	1
50°73	−86°25′54″13	50°23 − 51°24	−86°50′48″56 −	−86°01′02″53	2
50°73	−86°25′44″94	50°23 − 51°25	−86°54′33″20 −	−86°01′02″53	3
50°77	−86°25′04″71	50°27 − 51°25	−86°50′48″56 −	−86°01′02″53	4
50°75	−86°25′08″31	50°27 − 51°25	−86°50′48″56 −	−86°01′02″53	5
53°58	−86°20′48″87	53°16 − 54°07	−86°47′23″82 −	−86°00′42″69	2
53°59	−86°24′35″88	52°98 − 54°19	−86°56′14″36 −	−86°00′42″69	3
53°58	−86°20′33″55	53°11 − 54°07	−86°47′20″88 −	−86°00′42″69	4
56°14	−71°39′52″64	56°03 − 56°34	−71°59′28″39 −	−71°27′48″18	1
56°14	−71°39′54″60	56°03 − 56°26	−71°55′20″84 −	−71°27′11″21	3
56°15	−71°41′15″45	56°02 − 56°34	−72°04′39″73 −	−71°27′48″18	4

56°14	−71°39′39″48	56°03 − 56°26	−71°48′32″10 − −71°27′48″18	5
61°92	−86°19′49″83	61°39 − 62°47	−86°45′54″25 − −86°00′24″76	1
62°93	−86°27′01″74	62°42 − 63°39	−86°54′10″71 − −86°00′22″24	1
61°90	−86°17′21″94	61°40 − 62°43	−86°41′13″85 − −86°00′24″76	3
62°93	−86°25′46″70	62°42 − 63°45	−86°54′10″71 − −86°00′22″24	3
61°85	−86°18′41″65	61°35 − 62°35	−86°45′54″25 − −86°00′05″26	4
62°22	−86°49′12″30	61°68 − 62°63	−87°10′30″65 − −86°26′43″62	4
61°91	−86°16′51″91	61°40 − 62°35	−86°38′42″34 − −86°00′24″76	5
62°91	−86°26′26″18	62°42 − 63°36	−86°54′10″71 − −86°00′22″24	5
71°37	−84°00′30″40	70°80 − 71°80	−84°28′49″91 − −83°24′22″53	1
71°37	−84°00′35″45	70°80 − 71°80	−84°22′49″86 − −83°29′40″52	2
71°37	−84°01′23″35	70°98 − 71°76	−84°22′41″95 − −83°42′29″79	3
71°36	−84°00′30″16	70°84 − 71°76	−84°22′10″30 − −83°29′40″52	4
71°37	−84°00′13″25	70°77 − 71°80	−84°22′49″86 − −83°26′44″60	5
69°46	−86°38′59″13	68°81 − 70°12	−87°15′42″73 − −86°03′47″98	1
69°36	−86°42′25″19	68°57 − 70°21	−87°33′40″20 − −86°00′27″49	2
69°41	−86°38′42″03	68°71 − 70°06	−87°15′36″88 − −86°03′47″98	3
69°37	−86°38′52″44	68°64 − 70°17	−87°25′08″48 − −86°00′27″49	4
71°49	−86°23′14″97	71°05 − 71°97	−86°48′24″97 − −86°01′34″43	4
71°47	−86°25′11″42	70°93 − 71°97	−86°54′45″20 − −86°01′34″43	5
69°41	−86°39′02″72	68°76 − 70°11	−87°15′42″73 − −86°03′47″98	5
74°79	−65°58′53″76	74°73 − 74°84	−65°59′58″83 − −65°56′56″74	1
74°79	−65°58′52″77	74°72 − 74°85	−65°59′58″83 − −65°55′13″36	2
74°79	−65°58′33″95	74°72 − 74°89	−65°59′55″68 − −65°56′08″71	2
74°79	−65°58′47″84	74°72 − 74°85	−65°59′58″83 − −65°55′01″16	3
74°79	−65°58′34″66	74°72 − 74°88	−65°59′56″04 − −65°54′46″43	3
74°78	−65°58′36″24	74°71 − 74°87	−65°59′58″83 − −65°49′10″26	5
74°79	−65°58′20″78	74°72 − 74°87	−65°59′57″54 − −65°53′46″34	5
73°95	−74°16′54″76	73°84 − 74°09	−74°23′35″77 − −74°08′18″62	1
73°95	−74°16′57″19	73°84 − 74°07	−74°21′33″08 − −74°11′24″40	2
73°95	−74°16′41″94	73°80 − 74°12	−74°26′59″69 − −74°08′18″62	3
73°95	−74°16′32″77	73°78 − 74°11	−74°28′29″07 − −74°03′19″63	4
73°95	−74°16′40″07	73°77 − 74°11	−74°27′26″84 − −74°04′27″48	5
73°30	−86°36′49″34	72°61 − 74°08	−87°20′21″62 − −86°01′36″09	1
73°53	−86°37′55″11	72°68 − 74°40	−87°27′31″75 − −86°00′15″80	2
73°57	−86°36′39″48	72°73 − 74°47	−87°26′18″23 − −86°00′15″80	3
73°34	−86°37′15″79	72°61 − 74°18	−87°25′57″57 − −86°01′36″09	4
73°46	−86°38′50″19	72°61 − 74°50	−87°32′24″77 − −86°00′15″80	5
08°62	−86°31′22″09	08°10 − 09°19	−87°02′11″48 − −86°04′45″38	1
08°62	−86°31′08″60	08°17 − 09°10	−86°56′36″06 − −86°08′31″47	2
08°63	−86°30′15″31	08°17 − 09°10	−86°56′36″06 − −86°05′42″51	3
08°61	−86°30′30″86	08°17 − 09°10	−86°56′36″06 − −86°05′42″51	4
08°63	−86°30′31″20	08°12 − 09°15	−86°56′36″06 − −86°04′45″44	5

Table C.4: Area 4: All Clusters Found Across All Experiments

RA (°)	Dec (°)	RA Bounds (°)	Dec Bounds (°)	Experiment
00°56	67°26′02″32	00°40 − 00°75	67°16′53″23 − 67°34′07″21	1
00°55	67°25′17″38	00°44 − 00°68	67°19′50″40 − 67°30′31″84	2
00°55	67°25′57″82	00°42 − 00°71	67°19′37″98 − 67°34′07″21	3
00°55	67°25′11″16	00°46 − 00°66	67°19′53″51 − 67°29′48″36	4
23°47	30°36′29″71	23°32 − 23°59	30°29′08″78 − 30°43′25″53	1
23°47	30°36′37″36	23°35 − 23°60	30°29′29″25 − 30°43′35″91	2
23°47	30°37′42″05	23°38 − 23°56	30°32′01″53 − 30°43′25″53	3
23°47	30°36′50″79	23°36 − 23°57	30°30′46″37 − 30°43′25″53	4
23°46	30°36′29″71	23°31 − 23°61	30°29′08″78 − 30°44′38″01	5
35°52	57°08′26″09	35°43 − 35°63	57°03′30″07 − 57°13′46″71	3
40°69	61°35′27″79	40°58 − 40°79	61°29′39″61 − 61°41′45″61	4

Bibliography

[1] M. Giavalisco, *Galaxy Evolution.* 2000, p. 2142. DOI: 10.1888/0333750888/2142.

[2] *The Hubble Tuning Fork - Classification of Galaxies*, ESA/Hubble. [Online]. Available: https://www.spacetelescope.org/images/heic9902o/ (visited on 04/08/2020).

[3] R. Jimenez and P. Padoan, "A New Self-consistency Check on the Ages of Globular Clusters," *The Astrophysical Journal*, vol. 463, no. 1, pp. L17–L20, May 1996. DOI: 10.1086/310053. [Online]. Available: https://doi.org/10.1086/310053.

[4] R. Gratton, A. Bragaglia, E. Carretta, V. D'Orazi, S. Lucatello, and A. Sollima, "What is a Globular Cluster? an Observational Perspective.," *The Astronomy and Astrophysics Review.*, vol. 27, no. 1, pp. 1–136, 2019. DOI: doi:10.1007/s00159-019-0119-3. [Online]. Available: https://link.springer.com/article/10.1007/s00159-019-0119-3 (visited on 08/24/2021).

[5] S. van den Bergh, "How Did Globular Clusters Form?" *The Astrophysical Jorunal Letters*, vol. 559, no. 2, pp. L113–L114, Oct. 2001. DOI: 10.1086/323754. arXiv: astro-ph/0108298 [astro-ph].

[6] D. Bamberger, *How do Astronomers Count Stars in a Globular Cluster?* [Online]. Available: https://www.quora.com/How-do-astronomers-count-stars-in-a-globular-cluster?share=1 (visited on 04/06/2020).

[7] B. C. Erin M. O'Malley Christina Gilligan, "Absolute Ages and Distances of 22 GCs Using Monte Carlo Main-sequence Fitting," *The Astrophysical Journal*, vol. 838, no. 2, p. 162, Apr. 2017. DOI: 10.3847/1538-4357/aa6574. [Online]. Available: https://doi.org/10.3847/1538-4357/aa6574.

[8] M. Mohammadi, N. Petkov, K. Bunte, R. Peletier, and F. Schleif, "Globular Cluster Detection in the Gaia Survey," *Neurocomputing*, vol. 342, pp. 164–171, 2019, Advances in artificial neural networks, machine learning and computational intelligence, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2018.10.081. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231219301341.

[9] A. Dotter, A. Sarajedini, and J. Anderson, "Globular Clusters in the Outer Galactic Halo: New Hubble Space Telescope / Advanced Camera for Surveys Imaging of Six Globular Clusters and the Galactic Globular Clusters Age-metalicity Relation," *The Astrophysical Journal*, vol. 738, no. 1, p. 74, Aug. 2011. DOI: 10.1088/0004-637x/738/1/74. [Online]. Available: https://doi.org/10.1088/0004-637x/738/1/74.

[10] D. R. Soderblom, "The Ages of Stars," *Annual Review of Astronomy and Astrophysics*, vol. 48, no. 1, pp. 581–629, 2010. DOI: 10.1146/annurev-astro-081309-130806. [Online]. Available: https://doi.org/10.1146/annurev-astro-081309-130806.

[11] L. Monaco, E. Pancino, F. R. Ferraro, and M. Bellazzini, "Wide-field photometry of the galactic globular cluster m22," *Monthly Notices of the Royal Astronomical Society*, vol. 349, no. 4, pp. 1278–1290, Apr. 2004, ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2004.07599.x. [Online]. Available: http://dx.doi.org/10.1111/j.1365-2966.2004.07599.x.

[12] G. D. Bothun, H. C. Harris, and J. E. Hesser, "Detection of the globular-cluster population around ngc 7814," *Publications of the Astronomical Society of the Pacific*, vol. 104, p. 1220, Dec. 1992. DOI: 10.1086/133111.

[13] M. G. Lee, H. S. Park, and H. S. Hwang, "Detection of a Large-Scale Structure of Intracluster Globular Clusters in the Virgo Cluster," *Science*, vol. 328, no. 5976, pp. 334–336, 2010, ISSN: 0036-8075. DOI: 10.1126/science.1186496. eprint: https://science.sciencemag.org/content/328/5976/334.full.pdf. [Online]. Available: https://science.sciencemag.org/content/328/5976/334.

[14] W. E. Harris. (Dec. 2010). "Catalog of Parameters for Milky Way Globular Clusters: the Database," McMaster University, [Online]. Available: https://physics.mcmaster.ca/~harris/mwgc.dat (visited on 08/25/2021).

[15] (Apr. 10, 2015). "The Crammed Center of Messier 22," ESA/Hubble, [Online]. Available: https://upload.wikimedia.org/wikipedia/commons/7/76/The%5C%5Fcrammed%5C%5Fcentre%5C%5Fof%5C%5FMessier%5C%5F22.jpg (visited on 08/16/2021).

[16] Y.-S. Ting, A. Vaz, and G. Narayan, *How Do We Study the Stars? - Yuan-Sen Ting*, Ted-Ed. [Online]. Available: https://ed.ted.com/lessons/how-do-we-study-the-stars-yuan-sen-ting (visited on 07/16/2021).

[17] *Space Technology*. [Online]. Available: https://clarkscience8.weebly.com/space-technology.html (visited on 07/16/2021).

[18] Gaia Collaboration et al., "Description of the Gaia Mission (Spacecraft, Instruments, Survey and Measurement Principles, and Operations)," *Astronomy and Astrophysics*, vol. 595, Nov. 2016. DOI: 10.1051/0004-6361/201629272. [Online]. Available: https://doi.org/10.1051/0004-6361/201629272.

[19] Gaia Collaboration et al., "Gaia Data Release 2 - Summary of the Contents and Survey Properties," *Astronomy and Astrophysics*, vol. 616, Aug. 2018. DOI: 10.1051/0004-6361/201833051. [Online]. Available: https://doi.org/10.1051/0004-6361/201833051.

[20] (2021). "Gaia Data Release Scenario," European Space Agency, [Online]. Available: https://www.cosmos.esa.int/web/gaia/release (visited on 08/25/2021).

[21] *List of Globular Clusters*. [Online]. Available: https://en.wikipedia.org/wiki/List%5C%5Fof%5C%5Fglobular%5C%5Fclusters (visited on 04/07/2020).

[22] (Aug. 16, 2021). "List of NGC Selected from the VizieR Service," Strasbourg Astronomical Data Center, [Online]. Available: https://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=VII/118/ngc2000 (visited on 08/16/2021).

[23] (Aug. 14, 2021). "Parallax," [Online]. Available: https://en.wikipedia.org/wiki/Parallax (visited on 08/24/2021).

[24] (Apr. 27, 2018). "Gaia Data Release 2 (Gaia DR2)," European Space Agency, [Online]. Available: https://www.cosmos.esa.int/web/gaia/dr2 (visited on 05/27/2020).

[25] X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard, J. de Bruijne, T. Prusti, C. Babusiaux, and H. E. Delgado, "Gaia Data Release 2," *Astronomy & Astrophysics*, vol. 616, A9, Aug. 2018, ISSN: 1432-0746. DOI: 10.1051/0004-6361/201832964. [Online]. Available: http://dx.doi.org/10.1051/0004-6361/201832964.

[26] C. Martin, *Stellarparallax Parsec1*, Sep. 24, 2006. [Online]. Available: https://commons.wikimedia.org/wiki/File:Stellarparallax%5C%5Fparsec1.svg.

[27] Z. Modrono. (Jul. 18, 2017). "Dreaming Peter van de Kamp's Dream," [Online]. Available: https://reddots.space/dreaming-peter-van-de-kamps-dream/ (visited on 08/19/2021).

[28] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–, Nov. 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

[29] *Diffusion Gaussian Kernel*, University of Wisconsin-Madison, 2007. [Online]. Available: http://pages.stat.wisc.edu/~mchung/teaching/MIA/reading/diffusion.gaussian.kernel.pdf (visited on 06/15/2021).

[30] M. A. Michael W. Davidson. (May 17, 2016). "Molecular Expressions Microscopy Primer: Digital Image Processing – Difference of Gaussians Edge Enhancement Algorithm," [Online]. Available: https://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/diffgaussians/index.html (visited on 06/24/2021).

[31] D. Hooker. (Jul. 1973). "Lenna (Test Image)," Playboy Magazine, [Online]. Available: https://upload.wikimedia.org/wikipedia/en/7/7d/Lenna%5C%5F%5C%28test%5C%5Fimage%5C%29.png (visited on 11/11/2021).

[32] D. Marr and E. C. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, pp. 187–217, 1980.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[34] A. Akhtar, "Evolution of Ant Colony Optimization Algorithm – A Brief Literature Review," *arXiv e-prints*, arXiv:1908.08007, arXiv:1908.08007, Aug. 2019. arXiv: 1908.08007 [cs.NE].

[35] F. Xia, J. Liu, H. Nie, Y. Fu, L. Wan, and X. Kong, "Random Walks: A Review of Algorithms and Applications," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. PP, pp. 1–13, Nov. 2019. DOI: 10.1109/TETCI.2019.2952908.

[36] M. Mohammadi, "The Ant Colony," Unpublished, 2020.

[37] S. M. Holland. (Dec. 3, 2019). "Principal Components Analysis (PCA)," [Online]. Available: https://strata.uga.edu/software/pdf/pcaTutorial.pdf (visited on 08/05/2021).

[38] D. L. Ian Butterworth Tom Kwong. (2021). "Principal component analysis," [Online]. Available: https://multivariatestatsjl.readthedocs.io/en/latest/pca.html (visited on 12/19/2020).

[39] E. Dugan. (Jun. 19, 2019). "What is the Difference Between a Globular Star Cluster and an Open Star Cluster?" [Online]. Available: https://astronomy.com/magazine/ask-astro/2019/06/what-is-the-difference-between-a-globular-star-cluster-and-an-open-star-cluster (visited on 08/25/2021).

[40] G. Zotti, S. M. Hoffmann, A. Wolf, F. Chéreau, and G. Chéreau, "The Simulated Sky: Stellarium for Cultural Astronomy Research," *Journal of Skyscape Archaeology*, vol. 6, no. 2, pp. 221–258, Mar. 2021. DOI: 10.1558/jsa.17822. [Online]. Available: https://journal.equinoxpub.com/JSA/article/view/17822.

[41] S. Koposov, J. T. A. de Jong, V. Belokurov, H. Rix, D. B. Zucker, N. W. Evans, G. Gilmore, M. J. Irwin, and E. F. Bell, "The Discovery of Two Extremely Low Luminosity Milky Way Globular Clusters," *The Astrophysical Journal*, vol. 669, no. 1, pp. 337–342, Nov. 2007. DOI: 10.1086/521422. [Online]. Available: https://doi.org/10.1086/521422.

[42] M. E. Sharina, M. V. Ryabova, M. I. Maricheva, and A. S. Gorban, "The Stellar Population and Orbit of the Galactic Globular Cluster Palomar 3," *Astronomy Reports*, vol. 62, no. 11, pp. 733–746, Nov. 2018, ISSN: 1562-6881. DOI: 10.1134/s1063772918110069. [Online]. Available: http://dx.doi.org/10.1134/S1063772918110069.

[43] B. Willman, M. Geha, J. Strader, L. E. Strigari, J. D. Simon, E. Kirby, N. Ho, and A. Warres, "Willman 1 —- A Probable Dwarf Galaxy with an Irregular Kinematic Distrubution," *The Astronomical Journal*, vol. 142, no. 4, p. 128, Sep. 2011, ISSN: 1538-3881. DOI: 10.1088/0004-6256/142/4/128. [Online]. Available: http://dx.doi.org/10.1088/0004-6256/142/4/128.

[44] M. Aaronson, R. A. Schommer, and E. W. Olszewski, "AM-1 : a Very Distant Globular Cluster," *Astrophysical Journal*, vol. 276, pp. 221–228, Jan. 1984. DOI: 10.1086/161605.