

**EXPLORING GALACTIC CHEMICAL EVOLUTION
USING CLUSTER CHEMISTRY
AND VARIABILITY**

by

Vijith Jacob Poovelil

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

Department of Physics and Astronomy

The University of Utah

May 2021

Copyright © Vijith Jacob Poovelil 2021

All Rights Reserved

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

STATEMENT OF DISSERTATION APPROVAL

This dissertation of

Vijith Jacob Poovelil

has been approved by the following supervisory committee members:

Gail Zasowski, Chair May 17, 2021
Date Approved

Anil Seth, Member May 17, 2021
Date Approved

Kyle Dawson, Member May 17, 2021
Date Approved

Andrey Rogachev, Member May 17, 2021
Date Approved

Jeff Phillips, Member May 17, 2021
Date Approved

and by Christoph Boehme, Chair
of

the Department of
Physics and Astronomy

and by David B. Kieda of the Graduate School.

ABSTRACT

A stellar cluster is a gravitationally bound, co-moving group of stars that are believed to have been born together and hence have very similar chemistry and age. These stellar populations are essential tools in our quest to explore the evolution history of the Milky Way and other galaxies.

The chemical homogeneity of stellar clusters contains important clues about interstellar medium mixing efficiency, star formation, and the enrichment history of the Galaxy. I present (i) a new cluster member catalog based only on stellar kinematics, (ii) improved uncertainties for chemical information of the cluster members, and (iii) the dependence of cluster homogeneity on Galactic and cluster properties. I find that cluster homogeneity is uncorrelated with Galactocentric distance, vertical height, age, and metallicity. However, velocity dispersion, which is a proxy for cluster mass, is positively correlated with intrinsic scatter at relatively high levels of significance for Ca and Mg, with possible positive correlation at a low level of significance for Ni, Si, Al, and Fe. The elements that show a correlation with velocity dispersion are those that are predominantly produced by core-collapse supernovae. These findings suggest a quantitative difference between the correlation lengths of elements produced predominantly by different types of supernovae, which would have implications for Galactic chemical evolution models and the feasibility of chemical tagging.

Cluster age is another valuable property for exploring Galaxy evolution and formation. However, determining ages of unresolved clusters is difficult. I perform a pilot study for a project to develop statistical techniques to estimate cluster ages using integrated light from the cluster. I explore the photometric variability of resolved field stars and unresolved background stars within the cluster radius and find non-trivial variability from each of these sources. I also identify variability features in the ensemble light from clusters

corresponding to previously known and unknown potential variable stars. These analyses lay the groundwork for determining cluster ages, particularly for unresolved clusters in other galaxies, using the variability of their ensemble light.

CONTENTS

ABSTRACT	iii
LIST OF TABLES	vii
ACKNOWLEDGMENTS	viii
CHAPTERS	
1. INTRODUCTION	1
1.1 Stellar Clusters	1
1.1.1 Properties	1
1.1.2 Formation and Disruption	3
1.2 Cluster Chemistry and Galaxy Evolution	5
1.2.1 Nucleosynthetic Processes	5
1.2.2 Power of Chemistry	7
1.2.3 Chemical Homogeneity in Clusters	8
1.2.4 Chemical Tagging	9
1.3 Photometric Variability in Clusters	11
1.3.1 Stellar Variability	11
1.3.2 Cluster Age vs Variability	13
References	18
2. OPEN CLUSTER CHEMICAL HOMOGENEITY THROUGHOUT THE MILKY WAY	22
2.1 Introduction	23
2.2 Data	24
2.2.1 Apogee	24
2.2.2 Rederived Abundance Uncertainties	25
2.2.3 Proper Motions	27
2.2.4 Stellar Distances	27
2.2.5 Literature Cluster Parameters	27
2.3 Cluster Membership	27
2.3.1 Method	27
2.3.2 Validation	29
2.3.3 Catalog	30

2.4	Quantifying Cluster Homogeneity	31
2.5	Results	31
2.5.1	Final Selection of Elements and Cluster Members	31
2.5.2	Cluster Distances	32
2.5.3	Abundance Scatter in Clusters	33
2.5.4	Galactic Position, Age, and Metallicity	33
2.5.5	Velocity Dispersion (Cluster Mass)	34
2.6	Summary	36
2.7	APPENDIX A	36
2.8	APPENDIX B	39
2.9	References	40
3.	ASSESSING THE INFORMATION CONTENT OF CLUSTER ENSEMBLE LIGHTCURVES	41
3.1	Introduction	41
3.2	Data and Ensemble Photometry Techniques	42
3.2.1	Cluster Catalogs	42
3.2.2	Time Domain Photometry	43
3.2.3	TESS Lightcurves	44
3.2.4	Data Processing and Detrending	45
3.3	Exploring Cluster Variability Using Ensemble Photometry	49
3.3.1	Field Star Variability	49
3.3.2	Background Variability	50
3.3.3	Identifying Periodic Features in Ensemble Lightcurves	51
3.4	Summary	53
	References	68
4.	SUMMARY AND FUTURE WORK	70
4.1	Cluster Chemical Homogeneity and Chemical Tagging	70
4.2	Cluster Variability and Age	71

LIST OF TABLES

2.1 Sample Table of Cluster Members Selected Using Our Membership Selection in Section 3	30
2.2 Columns from the Table of Catalog Clusters	30
2.3 A Subsample of Columns from Table 2 for the 10 Clusters Used in Section 5	31
2.4 Intrinsic Abundance Scatter (Section 4) and Space Velocity Dispersion (σ_{tot} ; Equation (4)) for the OCs Analyzed in Section 5	33

ACKNOWLEDGMENTS

This work has been made possible by the patient and keen guidance from my supervisor, Prof. Gail Zasowski. When I joined her group as a second year grad student, I had no prior knowledge or training in astrophysics. However, the expert support and supervision from Gail in undertaking multiple projects has helped me complete my PhD within a total time of five years. She has been an excellent teacher and an irreplaceable guide in preparing and training me to be a good researcher.

I would like to thank the rest of my committee members, viz., Anil Seth, Kyle Dawson, Andrey Rogachev, and Jeff Phillips for their continued support and guidance throughout my PhD. They have constantly given me a fresh outlook on my research work and also provided me with essential career guidance whenever necessary.

I would like to thank all my group members including Sten Hasselquist, Nicolas Boardman, Jianhui Lian, Aishwarya Ashok, Benjamin Gibson, Kameron Goold, and Tobin Wainer. The weekly group meetings where we present and discuss our research progress have been incredibly productive.

I would like to thank several individuals in the SDSS collaboration who have continually taken time to help me with their expertise. It has been a pleasure working with the entire group. I would like to thank my collaborators for the TESS project, including Prof. Joshua Pepper at the Lehigh University, John Dinsmore, Avi Patel, and Marios Christodoulou for their contributions and support in one of the research projects I was involved in. The biweekly collaboration meetings and intermittent exchanges have been instrumental in some of the work that I have presented in this thesis.

I have my parents and my brother to thank for all the achievements in my life so far, and for those yet to come. My friends and family have been the pillar of support I could always lean on whenever I needed it the most.

I acknowledge the support during this thesis from the Research Corporation for Science Advancement and from the Heising-Simons Foundation through Scialog awards.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics — Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

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

This work includes data collected by the TESS mission, which are publicly available from the Mikulski Archive for Space Telescopes (MAST). Funding for the TESS mission is

provided by NASA's Science Mission directorate.

CHAPTER 1

INTRODUCTION

The seemingly infinite extent of the night sky has always intrigued the human mind. Ever since we evolved the cranial capacity to do so, we have always wondered how the stars that we see every night came into existence. How are these massive nuclear reactors formed? Why are there different types of stars and what factors determine their initial masses, temperatures, and velocities? How are these hundreds of billions of stars that make up our Galaxy distributed? How have the properties of these populations of stars in our Galaxy evolved over time? How were the shape, size, and brightness of the Galaxy affected by the changes in these stellar populations? How did a snapshot of the Galaxy from Earth look about 4.5 billion years ago?

These are some of the questions that drive Local Group astrophysics. If we can understand how the Milky Way (MW) was born and evolved over time, then we can understand more about the other billions of galaxies in our Universe. Since we are entrenched within it, the MW is the perfect laboratory to precisely study the minute details of how stars and stellar populations contribute to the overall evolution of their parent galaxies.

1.1 Stellar Clusters

1.1.1 Properties

Most stars form in clusters due to the fragmentation of the initial star formation cloud that prevents it from contracting into a single star (Jappsen et al. 2005). These stars that form together may escape in time from their birth clusters as a result of interactions with

stars within the cluster or from gravitational interactions from external sources. Stellar clusters can be divided into two broad types based on several distinguishing properties.

Globular clusters (GCs) are dense collections of stars that are tightly bound by gravity. They are generally old ($\gtrsim 6$ Gyr), massive ($\gtrsim 10^4 M_{\odot}$), and metal-poor, and are mostly found in the galactic halo for spiral galaxies like the MW (Kharchenko et al. 2013; Gratton 2020; Krumholz et al. 2019). They are fairly common with around 150 known GCs in the MW (Harris 1996) and around 500 in M31 (Barmby and Huchra 2001). Almost every large galaxy that has been observed is known to possess a system of GCs (Harris 1991).

The distribution of GC ages within a galaxy are used as indicators of the star formation and assembly history of the galaxy. GC formation is believed to occur over cosmic timescales, in many cases tracing the star formation history of the field population of its host galaxy (Usher et al. 2019; Reina-Campos et al. 2019), and does not require special conditions in the early Universe (Katz and Ricotti 2014). GCs in the MW have median ages of ≈ 10 Gyr with a spread of a few Gyr (Bastian and Lardo 2018). Ages for GCs in the MW have been estimated primarily through photometry by fitting their resolved color-magnitude diagrams (CMDs, Marín-Franch et al. 2009). In the case of extragalactic GCs, for which resolved photometry may not be available, ages have been determined by including integrated spectroscopic information as well, utilising color-metallicity relations (Usher et al. 2012, 2019) and metallicities (Forbes et al. 2015).

The second type of clusters is called open clusters (OCs), and this thesis focuses on projects exploring several properties of OCs, such as chemistry, mass, variability, and age. OCs are groups of stars that are bound by gravity, although weaker than GCs, and are believed to have been born together by the same initial molecular cloud. OCs are generally less massive ($\lesssim 5000 M_{\odot}$) and younger ($\lesssim 6$ Gyr) than GCs (Kharchenko et al. 2013; Krumholz et al. 2019). Despite having higher metallicities in general than GCs, OC metallicities overlap those of the thick-disk population of Galactic GCs, which have metallicities > -0.8 dex (Zinn 1985).

OCs have been observed to trace the metallicity gradient of the Galaxy with the more metal-rich clusters found towards the center (Jacobson et al. 2011; Donor et al. 2018).

The ages and metallicities of OCs can provide clues about their formation and evolution in the Galaxy. For instance, Friel (1995) finds a large scatter in the age and metallicity distribution of OCs at different Galactocentric radii. This observed scatter could be a result of radial migration of individual stars (Minchev and Famaey 2010) or clusters (Bekki 2010) from their birth locations. A more likely explanation for the scatter is that the OCs that have average properties different from those of their environment may have been accreted from infalling dwarf galaxies. Several OCs such as NGC 6791, Berkeley 29, and Saurer 1 have been argued to have arrived in our Galaxy via infall (Carraro and Bensby 2009; Krumholz et al. 2019).

1.1.2 Formation and Disruption

Giant Molecular Clouds (GMCs) contain denser regions, called clumps, whose mass and size are comparable to individual star clusters. More massive GMCs host more massive clumps, which are expected in gas-rich regions within galaxies undergoing vigorous star formation (Escala and Larson 2008). Stellar clusters of a range of masses are formed in these dense clumps within GMCs (Krumholz et al. 2019). Ting et al. (2016) found an upper limit of $3 \times 10^7 M_{\odot}$ for clusters formed in the Galactic alpha-enhanced thick disk. Cluster formation conditions in nearby spiral galaxies are not substantially different from that in the MW.

There are multiple internal and environmental processes that contribute to cluster disruption:

1.1.2.1 Relaxation-Driven Mass Loss

One of the key processes that drive mass loss in clusters is relaxation. When high velocity stars escape the cluster, the Maxwellian distribution is preferentially underpopulated at the high-velocity end. As a result, stars exchange gravitational energy and repopulate the high velocity end. However, since these new high-velocity stars take up larger orbits, they will end up outside the tidal radius and escape the cluster. This underpopulates the high-velocity region again and the effect repeats. Evidence of this is seen in many GCs in the halo, like Pal 5 (Odenkirchen et al. 1997, 2002) and NGC 5466 (Grillmair and Johnson

2006). Similar evidence is seen in OCs, like Berkely 17 (Bhattacharya et al. 2017).

1.1.2.2 Tidal Perturbations

The main encounters of stellar clusters that can cause tidal perturbations are with GMCs (Spitzer 1958). However, most encounters do not disrupt the cluster completely but only result in some mass loss (Prieto and Gnedin 2008; Binney and Tremaine 2008; Athanassoula 2007). Tidal stripping works in conjunction with two body relaxation. A first strip removes the weakly bound stars and then two body relaxation repopulates these high-energy orbits. However, if the tidal shocks are separated by time scales shorter than the relaxation time of the cluster, then their impact is less since there will be no high velocity stars to strip.

OCs can get disrupted if they are in disk-crossing or bulge-crossing orbits since they experience changing tides as they pass through a disk or bulge (Martinez-Medina et al. 2017). Clusters that are far from the plane of the Galaxy make such transits very rarely, while those very close to the plane experience weaker gradient in tidal forces, hence preferentially preserving clusters in these extremes. This could be responsible for the observed distribution of GCs far from and OCs closer to the plane of the MW (Krumholz et al. 2019).

1.1.2.3 Stellar Evolution

Stellar evolution can result in mass loss in the form of core-collapse supernovae (SNe) and stellar winds from Asymptotic Giant Branch (AGB) stars. During the first 40 Myr after a cluster is formed, the major source of mass loss is the gas ejected from Type-II SNe that escape the cluster (Faucher-Giguère and Kaspi 2006). After the SNe cease, the main source of mass loss is stellar winds from AGB stars. This is a gradual process and the actual amount of mass ejected depends on the mass of the cluster. While the ejected gas escapes the cluster, the stellar remnants left behind will remain in the cluster (Kruijssen 2009). As a result of the adiabatic mass loss, the stars settle into a new virial equilibrium, increasing the cluster radius while shrinking the tidal radius. Stars that are outside the tidal radius will then be dispersed from the cluster.

The effect of stellar evolution on cluster disruption depends on the initial concentration of the cluster. If the cluster radius is much smaller than the tidal radius or in other words, the cluster is dense enough, then the effects of mass loss due to stellar evolution will be negligible (Martinez-Medina et al. 2020). In fact, the adiabatic mass loss due to stellar evolution cannot single-handedly disrupt an entire cluster since a new virial equilibrium will constantly be established (Kroupa and Boily 2002; Krumholz et al. 2019). However, the presence of tidal perturbations can drive mass loss beyond the possibility of virial equilibrium and the cluster may disrupt (Baumgardt and Makino 2003; Lamers et al. 2010).

1.2 Cluster Chemistry and Galaxy Evolution

Stars and the interstellar medium (ISM) have a cyclic relation. Stars form from collapsing clouds of gases in the ISM. They return their material into the ISM at the end of their lives and sometimes even during their lifetimes. However, the material they return has a larger fraction of heavier elements, compared to their initial gas, due to various enrichment processes that synthesize larger nuclei elements over the course of the star's lifetime or towards the end of it, e.g., supernovae. The main processes that produce elements other than hydrogen and helium (called ‘metals’ by astronomers) are described in the following section.

1.2.1 Nucleosynthetic Processes

There are several processes that synthesize the plethora of elements that we observe in the universe. The physical and temporal differences between these processes can be used as important clues about the origin and movement of stars in the Galaxy.

- The Big Bang: Quite literally, this is where it all started. The Big Bang produced all of the hydrogen and most of the helium, lithium, and deuterium that we see in the universe today (Alpher et al. 1948).
- Cosmic Ray Spallation: Cosmic rays consist of protons and alpha particles with very

high energies. When these particles collide with interstellar matter, they produce elements such as lithium, beryllium, and boron (Meneguzzi et al. 1971). Spallation is also responsible for producing trace amounts of higher elements; however, the contributions from SNe and neutron processes heavily outweigh this process. As seen from Figure 1.1, beryllium and boron are almost entirely formed from cosmic ray spallation, while lithium has significant contributions from other sources.

- **Hydrostatic Fusion:** While most low mass stars fuse hydrogen into helium in their core for billions of years, the most massive stars quickly run out of hydrogen in its core and starts burning helium into carbon via the triple-alpha process. This sequential hydrostatic burning continues, producing oxygen and silicon in various stages. After this, heavier elements are synthesized in abundances determined by the nuclear quasiequilibrium (Bodansky et al. 1968) that is achieved between alpha particle capture that produces heavier elements and the thermal photon induced photoejection of an alpha particle that produces lighter elements.

These heavier elements produced in massive stars are released into the ISM via the explosive power of SNe. Core collapse SNe are the major contributors of alpha elements like magnesium, silicon, and sulphur since they expel the layers of elements synthesized within massive stars during the hydrostatic burning phase (Hoyle 1954). Type Ia SNe preferentially release an excess of iron-peak elements, as represented in Figure 1.1.

- **Neutron Capture Processes:** Elements above iron are mainly synthesized in neutron rich environments, where the heavy elements already present in the star undergo a succession of neutron captures and eventually β decay into stable isotopes. This is called the r-process or rapid neutron capture process. The r-process requires environments with extremely dense concentrations of neutrons, which can be realized in neutron star mergers.

In the s-process, or the slow neutron capture process, there is enough time between neutron captures for the unstable nucleus to undergo β decay into a stable isotope (Burbidge et al. 1957). Compared to the r-process, the s-process occurs at much lesser neutron densities. The s-process is known to occur mainly in AGB stars, starting from heavy seed nuclei already present in them and slowly producing higher elements, one isotope after the other, over thousands of years.

A third type of neutron capture process, called the i-process or the intermediate process, occurs in certain environments with neutron densities between those required for the r and s processes. The hydrogen-rich envelope of an AGB star injects protons to the intershell region, where both the r-process and s-process form in varying amounts and timescales (Goswami and Goswami 2020).

1.2.2 Power of Chemistry

Chemical information of stellar populations can tell us a lot about the star-formation conditions present during their formation. For example, we can compare the early star-formation rates of two different populations by simply studying their chemistry. Different nucleosynthetic events produce certain specific elements in different amounts. Tracing the abundance of specific elements in the present-day distribution of stars in a galaxy can provide clues to the sequence of enrichment events that occurred prior to their formation. For instance, alpha elements such as magnesium, sulphur, and silicon are released predominantly in core-collapse SNe, such as Type II and Type Ib (Woosley et al. 1973), while Type Ia SNe release substantial quantities of iron-peak elements as well (François et al. 2004). Core-collapse SNe occur soon (≈ 40 Myr) after the formation of a particular population of stars (Woosley and Janka 2005) and Type Ia SNe take over a billion years (Fontaine et al. 2001), which occur when stellar remnants of low to intermediate mass stars (white dwarfs) accrete material and exceed the Chandrasekhar limit. On the other hand, core collapse SNe only occur for high mass stars, which have much shorter lifetimes compared to their low mass counterparts. This means that the distributions of the alpha element abundances and the metallicity of a stellar population will reflect its star formation history.

Figure 1.2 shows an example of alpha element abundance vs metallicity for two distinct populations. The plateau at low metallicity comprises stars that were formed from gas that was predominantly polluted by core-collapse SNe, while stars towards the right of the ‘knee’ were formed once the contributions from Type Ia SNe also started polluting the ISM. The black population is inferred to have a higher star-formation rate since it was able to form more stars that polluted the ISM towards higher metallicities before Type Ia SNe began.

1.2.3 Chemical Homogeneity in Clusters

Since stars within an OC are formed together from the same star forming cloud, they are expected to have very similar chemistry. This is under the assumption that the initial cloud is very well mixed and that the metals are distributed uniformly throughout the cloud. With the advent of large-scale surveys, it is possible to study the spectra of a increasing number of individual stars to analyze their chemical composition. Cluster chemistry is used to explore star formation and the chemical enrichment history of the Galaxy, using techniques such as chemical tagging (§1.2.4). Studying the chemical composition of OCs is an indispensable tool in our exploration of the Galaxy since they are the only form of stellar population that show high levels of chemical homogeneity.

Abundance variations in OCs have been studied sparsely but have been found to be homogeneous in general. The Hyades (De Silva et al. 2006, 2011) and M67 (Liu et al. 2016) are the two of the more widely studied clusters and were found to have abundance variations of less than 0.05 dex. De Silva et al. (2007) demonstrated high chemical homogeneity in 7 elements for 12 Red Giants of the OC Collinder 261. They also found unique chemical signatures in the 3 OCs; viz. Collinder 261, Hyades and HR 1614. Bovy (2016) shows that the scatter of the chemical abundances of stars within an OC are as low as 0.01 dex.

Although most OCs have been observed to be chemically homogeneous, there can be various factors that affect the level of homogeneity. For example, if the metals in the initial cloud are not well-mixed, stars that are formed from it will have considerably different metallicities. Work done by Bland-Hawthorn et al. (2010) suggests that clusters with masses up to $10^4 M_\odot$, but not all of those up to $10^5 M_\odot$, are expected to be homogeneous.

This suggests that ISM mixing efficiency could be dependant on mass of the cluster.

Variations in individual chemical abundances can also arise in stars belonging to different stages of evolution. Even if these stars were born from a well-mixed initial cloud, they can have differences in some elemental abundances due to various processes that take place in stars over the course of their lifetime. For instance, Souto et al. (2019) found significant abundance variations between stars belonging to different evolution stages within the open cluster M67. They propose that this is due to thermal diffusion and dredge-up. Blanco-Cuaresma et al. (2015) also found variations in chemical signatures for stars within the same cluster but belonging to different evolutionary stages. They also found evidence of chemical overlap between different OCs. They stress the importance of obtaining abundances of more elements in future surveys to be able to differentiate between chemically overlapping clusters for the purposes of chemical tagging (§1.2.4).

The chemical overlap between multiple OCs can be overcome by using multiple elemental abundances to study these clusters. For example, Armillotta et al. (2018) found that using multiple metals to trace the homogeneity is crucial to distinguish between two or more parsec scaled clusters, since any single metal tracer may or may not have significant overlap. They also find that regardless of the initial distribution of the metal field in the molecular cloud, stellar abundances are correlated below 1 parsec. This could mean that clusters that have radii larger than 1 pc may only be homogeneous in certain elements depending on the distribution of these metals in the initial molecular cloud.

The large volume of studies focusing on exploring chemical homogeneity in clusters is a testament to its importance in our understanding of the MW and beyond. Metallicity and abundance variations in OCs reflect atomic diffusion in stellar atmospheres, planetary infall, or intrinsic abundance variations that existed at the onset of star formation.

1.2.4 Chemical Tagging

The memory of the motions in space of stars that escape the gravitational field of their natal clusters tend to be lost over durations of time larger than the rotation period of the Galaxy, since they can be affected by time-dependent features such as the transit of the spiral arms in the Galaxy (Krumholz and Ting 2018). However, stars conserve their

chemical abundances almost over their entire lives. This led Freeman and Bland-Hawthorn (2002) to suggest that disrupted clusters can be detected using chemical signatures from sufficiently high-quality abundance data of multiple elements. They posit that reconstructing the clusters in the disk over the age of the Galaxy is possible using chemistry. Chemical tagging refers to identifying groups of stars based on their separation in chemical space.

1.2.4.1 Weak Chemical Tagging

In weak chemical tagging, soft assumptions may be made about certain properties of stellar populations, such as their age or their birth location in the Galaxy. Several conclusions can be made about the properties of stellar populations based on their distinct chemical patterns. For example, stars in the Solar neighborhood have been found to show a clear bimodality in multi-dimensional chemical space (Bensby et al. 2005; Hawkins et al. 2015). One possibility for this observation is that a population of stars migrated to the Solar neighborhood from their birth locations in the metal-rich inner Galaxy. This inference makes soft assumptions about the star formation conditions of stellar populations based on their location in the Galaxy.

1.2.4.2 Strong Chemical Tagging

On the other hand, strong chemical tagging involves using stronger assumptions about the age and birth locations of stars to infer certain properties. Strong chemical tagging spans a range of ideas, including discovering disrupted clusters and finding cluster members and expelled members of intact clusters using only chemical signatures. Various attempts have been made to identify OCs by grouping stars in chemical space. For example, Hogg et al. (2016) found that clusters in 15-dimensional abundance space correlate with clusters in velocity space. However, they report high level of contamination, which suggest that using clustering techniques in abundance space alone may not consistently result in physically co-moving objects. Mitschang et al. (2013) developed a metric to derive an empirical probability function for membership based on the chemical abundances for 35 clusters collected from literature. The authors showed that achieving a high clustering detection efficiency is difficult and that it depends on the level of uniqueness of the co-natal

stars' chemical signatures. Even dissolved stellar clusters can be reconstructed by grouping stars in high-dimensional chemical space against the background (Ting et al. 2012, 2015). It also helps distinguish large scale structures from one another. For example, the Sagittarius dwarf galaxy shows a distinct chemical pattern from the Milky Way in high-dimensional chemical space (Hasselquist et al. 2015).

1.3 Photometric Variability in Clusters

1.3.1 Stellar Variability

Photometric variability refers to the periodic or aperiodic variations in the observed apparent brightness of the star. These variations could arise either due to extrinsic factors such as eclipsing and stellar rotation, or due to intrinsic changes in the physical properties of the stars themselves. These two cases are described below. The classification of different types of variable stars can be seen in Figure 1.3.

1.3.1.1 Extrinsic Variables

In certain cases, the variation in the observed brightness of a star has nothing to do with the physical mechanisms that control its luminosity. An eclipsing binary is a classic example of such a system. In a binary system, every time one star transits the other, there will result in periodic dips in the brightness of the system under observation. The lightcurves for these systems are unique due to the dips of different strengths produced by the primary and secondary eclipses.

Stellar rotation could also cause variability in the apparent brightness of the star. Stars with sizeable starspots appear to dim as the starspots rotate out of view. Magnetic stars have brighter poles that could result in varying brightness as they rotate. Certain stars with ellipsoidal shape can also appear to fluctuate in brightness as the observed shape changes due to rotation.

1.3.1.2 Intrinsic Variables

Variability in the intrinsic brightness of stars is a very common phenomenon. Even our Sun varies in brightness by about 0.1% every 11 years (Fröhlich 2006). Certain types of

stars are known to have significant variability in their intrinsic brightness as a result of physical processes within the star. These so called intrinsic variable stars are crucial to the methodical exploration of our Galaxy and beyond.

Cataclysmic variables are stars whose luminosities increase by a large factor and then drop down to its original state. A classic example that has already been discussed is a SN, which is an explosive outburst that destroys the progenitor star. Pulsating variables are stars whose luminosities vary periodically due to increase in opacity of the stellar atmosphere with increasing temperature. This increased opacity causes heat build up, which then pushes the stellar atmosphere away, cooling it and decreasing the opacity. The result is a cyclic process that causes the stellar atmosphere to repeatedly pulsate (Tao et al. 1998). These stars occupy a narrow, almost vertical region in the Hertzsprung-Russell diagram called the ‘instability strip’, within which they suffer instabilities that cause their luminosities to vary regularly or irregularly. Cepheid variables and RR Lyrae variables are particularly interesting pulsating variables due to their use as distance indicators and their presence in stellar clusters.

Cepheid variables are particularly useful type of variable stars whose intrinsic brightness varies periodically over a range of multiple days (Soszyński et al. 2008). These stars follow a period luminosity relationship, also called the Leavitt law (Leavitt 1908), that links their intrinsic brightness to their pulsation period. This relationship is utilized as a distance indicator to estimate various Galactic scales such as the distance to the Galactic center, the distance of the Sun from the Galactic plane, distances of several globular clusters and external galaxies, and even the Hubble constant (Udalski et al. 1999; Majaess et al. 2009; Freedman and Madore 2010). Since they are massive ($4\text{-}20 M_{\odot}$) young stars (≈ 100 Myr) (Turner 1996), they tend to be present in younger stellar populations as opposed to older GCs.

RR Lyrae variable is another class of variables, whose brightness varies with a period typically less than a day. RR Lyraes are generally older and are used to trace the thick disk and halo populations (Dambis et al. 2013), despite their significant presence at all Galactic latitudes. These stars are very commonly associated with GCs, with over 80% of known

GC variables identified as RR Lyraes (Clement et al. 2001). This property is particularly useful for studies that track cluster properties using information from variable stars.

We have a good understanding of the kind of variable stars present in GCs in the MW. Clement et al. (2001) catalogs all the known variable stars in GCs from previous studies, along with information about the type of variable star and its period of variability. While a wide range of variable stars have been found in these GCs, almost every GC contains several RR Lyrae stars due to their characteristic old age. However, variability studies on OCs are less common and incomplete, mostly comprising of reports of individual detections of variables within the cluster. For example, Anderson et al. (2013) detected Cepheid variables in multiple OCs, and de Marchi et al. (2007) reported on the variable stars found in NGC 6791. Since OCs are relatively younger, RR Lyrae are not commonly found in them.

1.3.2 Cluster Age vs Variability

The fundamental difference in the types of variability sources found in young and old clusters is a vital tool that we can exploit. There are several surveys that study the photometric variability of stars, and many of them target individual clusters as well. With the help of these data, we can look at the periodic and aperiodic variations in the brightness of individual stars or groups of stars in OCs and GCs.

§3 describes the first steps taken towards devising novel techniques to estimate cluster ages using information from the variable stars present in them. Since RR Lyrae have short periods of typically less than a day, the signatures within the light coming from clusters that have an abundance of these stars will be different from those that contain other variable stars, such as Cephieds or eclipsing binaries. The lightcurves of RR Lyrae and Cepheids variables also significantly differ in amplitude and shape.

The ages of clusters are a crucial tool for exploring Galaxy evolution and formation. Our understanding of the Galactic thin and thick disk and the populations of stars within them has been significantly improved by explorations of clusters of different ages (Spina et al. 2021; Salaris et al. 2004). Analyzing the chemistry of stars within clusters of different ages also helps us learn more about the types of nucleosynthetic processes that produce these

elements and the interstellar mixing efficiency at different time scales and locations of the Galaxy (Brown et al. 1991; Krumholz and Ting 2018; Armillotta et al. 2018). Using the ages of clusters within the MW (Kharchenko et al. 2013) and the Small and Large Magellanic Clouds (Bica et al. 2020) to understand the relationship between cluster variability and age, we can estimate ages for unresolved clusters in other galaxies. §3 describes the pilot study performed to develop statistical techniques to determine cluster ages using the light from the cluster as a whole.

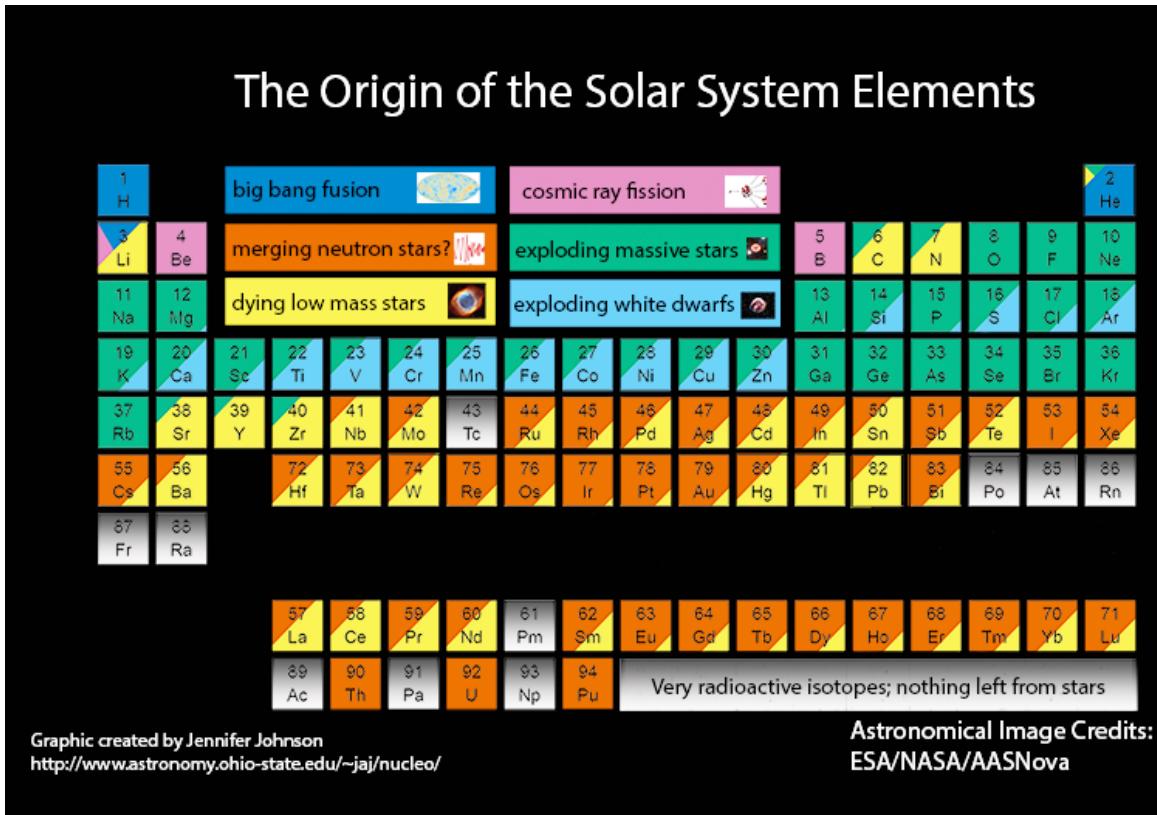


Figure 1.1: Periodic table of elements colored by the different nucleosynthetic processes that release them into the ISM.

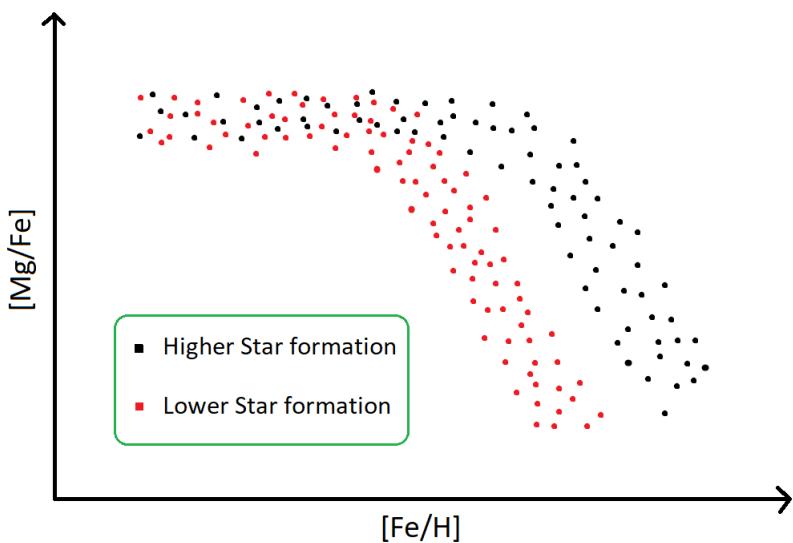


Figure 1.2: $[\text{Mg}/\text{Fe}]$ (alpha-element) vs. $[\text{Fe}/\text{H}]$ (metallicity) for two populations with different star formation rates. The two populations can be distinguished by the position of the knee.

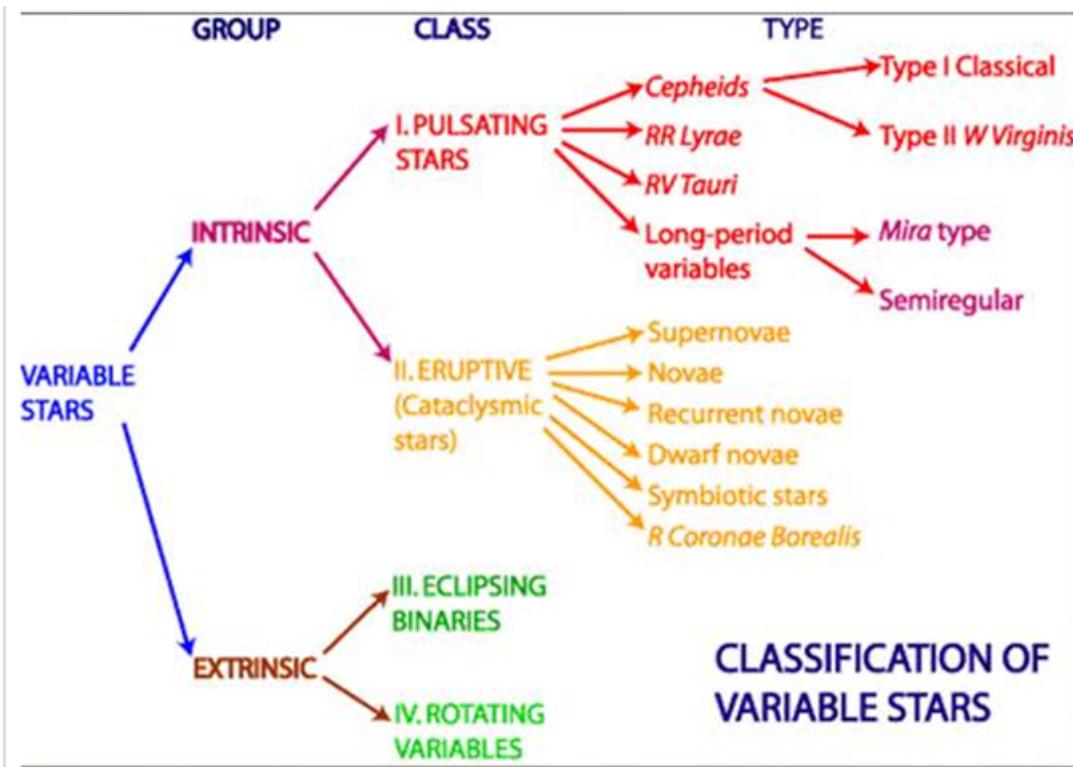


Figure 1.3: Classification of different types of variable stars. Credit: Australia Telescope National Facility (www.atnf.csiro.au)

REFERENCES

- Alpher, R. A., Bethe, H., and Gamow, G. 1948, Physical Review, 73, 803
- Anderson, R. I., Eyer, L., and Mowlavi, N. 2013, MNRAS, 434, 2238
- Armillotta, L., Krumholz, M. R., and Fujimoto, Y. 2018, s, 481, 5000
- Athanassoula, E. 2007, MNRAS, 377, 1569
- Barmby, P., and Huchra, J. P. 2001, AJ, 122, 2458
- Bastian, N., and Lardo, C. 2018, ARA&A, 56, 83
- Baumgardt, H., and Makino, J. 2003, MNRAS, 340, 227
- Bekki, K. 2010, MNRAS, 401, 2753
- Bensby, T., Feltzing, S., Lundström, I., and Ilyin, I. 2005, A&A, 433, 185
- Bhattacharya, S., Mishra, I., Vaidya, K., and Chen, W. P. 2017, ApJ, 847, 138
- Bica, E., Westera, P., Kerber, L. d. O., Dias, B., Maia, F., Santos, J. F. C., J., Barbuy, B., and Oliveira, R. A. P. 2020, VizieR Online Data Catalog, J/AJ/159/82
- Binney, J., and Tremaine, S. 2008, Galactic Dynamics: Second Edition
- Blanco-Cuaresma, S., et al. 2015, A&A, 577, A47
- Bland-Hawthorn, J., Krumholz, M. R., and Freeman, K. 2010, ApJ, 713, 166
- Bodansky, D., Clayton, D. D., and Fowler, W. A. 1968, PhRvL, 20, 161
- Bovy, J. 2016, ApJ, 817, 49
- Brown, J. H., Burkert, A., and Truran, J. W. 1991, ApJ, 376, 115
- Burbidge, E. M., Burbidge, G. R., Fowler, W. A., and Hoyle, F. 1957, Reviews of Modern Physics, 29, 547
- Carraro, G., and Bensby, T. 2009, MNRAS, 397, L106
- Clement, C. M., et al. 2001, AJ, 122, 2587

- Dambis, A. K., Berdnikov, L. N., Kniazev, A. Y., Kravtsov, V. V., Rastorguev, A. S., Sefako, R., and Vozyakova, O. V. 2013, MNRAS, 435, 3206
- de Marchi, F., et al. 2007, A&A, 471, 515
- De Silva, G. M., Freeman, K. C., Asplund, M., Bland -Hawthorn, J., Bessell, M. S., and Collet, R. 2007, AJ, 133, 1161
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., Asplund, M., Williams, M., and Holmberg, J. 2011, MNRAS, 415, 563
- De Silva, G. M., Sneden, C., Paulson, D. B., Asplund, M., Bland-Hawthorn, J., Bessell, M. S., and Freeman, K. C. 2006, AJ, 131, 455
- Donor, J., et al. 2018, AJ, 156, 142
- Escala, A., and Larson, R. B. 2008, ApJL, 685, L31
- Faucher-Giguère, C.-A., and Kaspi, V. M. 2006, ApJ, 643, 332
- Fontaine, G., Brassard, P., and Bergeron, P. 2001, PASP, 113, 409
- Forbes, D. A., Pastorello, N., Romanowsky, A. J., Usher, C., Brodie, J. P., and Strader, J. 2015, MNRAS, 452, 1045
- François, P., Matteucci, F., Cayrel, R., Spite, M., Spite, F., and Chiappini, C. 2004, A&A, 421, 613
- Freedman, W. L., and Madore, B. F. 2010, ARA&A, 48, 673
- Freeman, K., and Bland-Hawthorn, J. 2002, Annual Review of Astronomy and Astrophysics, Vol. 40, 487
- Friel, E. D. 1995, ARA&A, 33, 381
- Fröhlich, C. 2006, Space Science Reviews, 125, 53
- Goswami, P. P., and Goswami, A. 2020, Journal of Astrophysics and Astronomy, 41, 47
- Gratton, R. 2020, in IAU Symposium, Vol. 351, IAU Symposium, ed. A. Bragaglia, M. Davies, A. Sills, and E. Vesperini, 241–250
- Grillmair, C. J., and Johnson, R. 2006, ApJL, 639, L17
- Harris, W. E. 1991, ARA&A, 29, 543
- . 1996, AJ, 112, 1487
- Hasselquist, S., et al. 2015, in American Astronomical Society Meeting Abstracts, Vol. 225, American Astronomical Society Meeting Abstracts #225, 319.05

- Hawkins, K., Jofré, P., Masseron, T., and Gilmore, G. 2015, MNRAS, 453, 758
- Hogg, D. W., et al. 2016, ApJ, 833, 262
- Hoyle, F. 1954, ApJS, 1, 121
- Jacobson, H. R., Pilachowski, C. A., and Friel, E. D. 2011, The Astronomical Journal, 142, 59
- Jappsen, A. K., Klessen, R. S., Larson, R. B., Li, Y., and Mac Low, M. M. 2005, A&A, 435, 611
- Katz, H., and Ricotti, M. 2014, Monthly Notices of the Royal Astronomical Society, 444, 2377
- Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., and Scholz, R.-D. 2013, A&A, 558, A53
- Kroupa, P., and Boily, C. M. 2002, MNRAS, 336, 1188
- Kruijssen, J. M. D. 2009, A&A, 507, 1409
- Krumholz, M. R., McKee, C. F., and Bland -Hawthorn, J. 2019, ARA&A, 57, 227
- Krumholz, M. R., and Ting, Y.-S. 2018, MNRAS, 475, 2236
- Lamers, H. J. G. L. M., Baumgardt, H., and Gieles, M. 2010, MNRAS, 409, 305
- Leavitt, H. S. 1908, Annals of Harvard College Observatory, 60, 87
- Liu, F., Asplund, M., Yong, D., Meléndez, J., Ramírez, I., Karakas, A. I., Carlos, M., and Marino, A. F. 2016, MNRAS, 463, 696
- Majaess, D. J., Turner, D. G., and Lane, D. J. 2009, MNRAS, 398, 263
- Marín-Franch, A., et al. 2009, ApJ, 694, 1498
- Martinez-Medina, L. A., Gieles, M., Gnedin, O. Y., and Li, H. 2020, arXiv e-prints, arXiv:2009.06643
- Martinez-Medina, L. A., Pichardo, B., Peimbert, A., and Moreno, E. 2017, ApJ, 834, 58
- Meneguzzi, M., Audouze, J., and Reeves, H. 1971, A&A, 15, 337
- Minchev, I., and Famaey, B. 2010, ApJ, 722, 112
- Mitschang, A. W., De Silva, G., Sharma, S., and Zucker, D. B. 2013, MNRAS, 428, 2321
- Odenkirchen, M., Brosche, P., Geffert, M., and Tucholke, H. J. 1997, , 2, 477

- Odenkirchen, M., Grebel, E. K., Dehnen, W., Rix, H.-W., and Cudworth, K. M. 2002, AJ, 124, 1497
- Prieto, J. L., and Gnedin, O. Y. 2008, ApJ, 689, 919
- Reina-Campos, M., Kruijssen, J. M. D., Pfeffer, J. L., Bastian, N., and Crain, R. A. 2019, MNRAS, 486, 5838
- Salaris, M., Weiss, A., and Percival, S. M. 2004, A&A, 414, 163
- Soszyński, I., et al. 2008, , 58, 293
- Souto, D., et al. 2019, ApJ, 874, 97
- Spina, L., et al. 2021, MNRAS, 503, 3279
- Spitzer, Lyman, J. 1958, ApJ, 127, 17
- Tao, L., Spiegel, E., and Umurhan, O. M. 1998, in APS Division of Fluid Dynamics Meeting Abstracts, APS Meeting Abstracts, LC.10
- Ting, Y.-S., Conroy, C., and Goodman, A. 2015, ApJ, 807, 104
- Ting, Y.-S., Conroy, C., and Rix, H.-W. 2016, ApJ, 816, 10
- Ting, Y.-S., Freeman, K. C., Kobayashi, C., De Silva, G. M., and Bland-Hawthorn, J. 2012, MNRAS, 421, 1231
- Turner, D. G. 1996, , 90, 82
- Udalski, A., Soszynski, I., Szymanski, M., Kubiak, M., Pietrzynski, G., Wozniak, P., and Zebrun, K. 1999, , 49, 223
- Usher, C., Brodie, J. P., Forbes, D. A., Romanowsky, A. J., Strader, J., Pfeffer, J., and Bastian, N. 2019, MNRAS, 490, 491
- Usher, C., et al. 2012, MNRAS, 426, 1475
- Woosley, S., and Janka, T. 2005, Nature Physics, 1, 147
- Woosley, S. E., Arnett, W. D., and Clayton, D. D. 1973, ApJS, 26, 231
- Zinn, R. 1985, ApJ, 293, 424

CHAPTER 2

OPEN CLUSTER CHEMICAL HOMOGENEITY THROUGHOUT THE MILKY WAY

Reprinted with permission from Vijith Jacob Poovelil, G. Zasowski, S. Hasselquist, A. Seth, John Donor, Rachael L. Beaton, K. Cunha, Peter M. Frinchaboy, D. A. García-Hernández, K. Hawkins, K. M. Kratter, Richard R. Lane, and C. Nitschelm. “Open Cluster Chemical Homogeneity throughout the Milky Way” *ApJ*, vol. 903, 55, 2020.



Open Cluster Chemical Homogeneity throughout the Milky Way

Vijith Jacob Poovellil¹ , G. Zasowski¹ , S. Hasselquist^{1,12} , A. Seth¹ , John Donor², Rachael L. Beaton³, K. Cunha^{4,5} , Peter M. Frinchaboy² , D. A. García-Hernández^{6,7} , K. Hawkins⁸ , K. M. Kratter^{4,9} , Richard R. Lane¹⁰, and C. Nitschelm¹¹

¹ Department of Physics & Astronomy, University of Utah, Salt Lake City, UT, 84112, USA

² Department of Physics & Astronomy, Texas Christian University, Fort Worth, TX, 76129, USA

³ Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA

⁴ Steward Observatory, University of Arizona, Tucson, AZ, 85721, USA

⁵ Observatório Nacional, 20921-400 Rio de Janeiro, RJ, Brazil

⁶ Instituto de Astrofísica de Canarias (IAC), E-38205 La Laguna, Tenerife, Spain

⁷ Universidad de La Laguna (ULL), Departamento de Astrofísica, E-38206 La Laguna, Tenerife, Spain

⁸ Department of Astronomy, The University of Texas at Austin, Austin, TX, 78712, USA

⁹ Department of Astronomy, University of Arizona, Tucson, AZ, 85721, USA

¹⁰ Instituto de Astronomía y Ciencias Planetarias, Universidad de Atacama, Copiapó 485, Copiapó, Chile

¹¹ Centro de Astronomía (CITEVA), Universidad de Antofagasta, Avenida Angamos 601, Antofagasta 1270300, Chile

Received 2020 May 22; revised 2020 September 4; accepted 2020 September 14; published 2020 November 2

Abstract

The chemical homogeneity of surviving stellar clusters contains important clues about interstellar medium (ISM) mixing efficiency, star formation, and the enrichment history of the Galaxy. Existing measurements in a handful of open clusters suggest homogeneity in several elements at the 0.03 dex level. Here we present (i) a new cluster member catalog based only on APOGEE radial velocities and Gaia-DR2 proper motions, (ii) improved abundance uncertainties for APOGEE cluster members, and (iii) the dependence of cluster homogeneity on Galactic and cluster properties, using abundances of eight elements from the APOGEE survey for 10 high-quality clusters. We find that cluster homogeneity is uncorrelated with Galactocentric distance, $|Z|$, age, and metallicity. However, velocity dispersion, which is a proxy for cluster mass, is positively correlated with intrinsic scatter at relatively high levels of significance for $[\text{Ca}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$. We also see a possible positive correlation at a low level of significance for $[\text{Ni}/\text{Fe}]$, $[\text{Si}/\text{Fe}]$, $[\text{Al}/\text{Fe}]$, and $[\text{Fe}/\text{H}]$, while $[\text{Cr}/\text{Fe}]$ and $[\text{Mn}/\text{Fe}]$ are uncorrelated. The elements that show a correlation with velocity dispersion are those that are predominantly produced by core-collapse supernovae (CCSNe). However, the small sample size and relatively low correlation significance highlight the need for follow-up studies. If borne out by future studies, these findings would suggest a quantitative difference between the correlation lengths of elements produced predominantly by Type Ia SNe versus CCSNe, which would have implications for Galactic chemical evolution models and the feasibility of chemical tagging.

Unified Astronomy Thesaurus concepts: Chemical enrichment (225); Chemical abundances (224); Star clusters (1567); Milky Way evolution (1052)

Supporting material: figure set, machine-readable tables

1. Introduction

The chemical composition of stars that we see today is a consequence of a sequence of past enrichment events that polluted the interstellar medium (ISM). Hence, by studying stellar chemistry, we can learn how these events contributed to enrich the ISM and improve our understanding about the evolution of the Galaxy over time. Open clusters (OC) are particularly interesting objects as they consist of stars that were born together from the same initial molecular cloud and hence are believed to be chemically homogeneous. Studying the chemistry of these objects can help us trace the ISM pollution rates of different nucleosynthetic processes and ISM mixing efficiency in different locations in the Galaxy. One must also rely on the assumption of chemical homogeneity of OCs to identify common birth sites using only the chemical signatures of stars, also called chemical tagging (Freeman & Bland-Hawthorn 2002). Measuring the level of homogeneity of OCs and understanding the factors that can affect it are crucial for the feasibility of chemical tagging.

Cluster chemical homogeneity has been studied in many globular clusters (GCs) and a few OCs. GCs are observed to have inhomogeneities and anticorrelations in most of the light elements (e.g., Carretta et al. 2010; Milone et al. 2018; Mészáros et al. 2020). Chromosome maps show light-element abundance scatter in GCs with masses down to 10^4 M_\odot (Saracino et al. 2019). Heavy-element abundance variations are also observed, but only in a small number of massive GCs (e.g., Gratton 2020).

The chemistry of OCs, on the other hand, has not been as widely explored, and questions remain regarding the level—or even presence of—*intrinsic* chemical scatter. The Hyades is a well-studied cluster that has been argued to be chemically homogeneous (De Silva et al. 2006, 2011), although other recent work on the same cluster identified abundance variations of around 0.02 dex (Liu et al. 2016b). M67, another well-studied OC, was found to potentially be inhomogeneous in certain elements from the analysis of two solar twins in the cluster by Liu et al. (2016a). However, Bovy (2016) showed that the scatter of several elemental abundances relative to hydrogen within M67, NGC 6819, and NGC 2420 is as low as 0.03 dex, using APOGEE spectra. De Silva et al. (2007)

¹² NSF Astronomy and Astrophysics Postdoctoral Fellow.

demonstrated chemical homogeneity to the 0.05 dex level in seven elements using 12 red giants in the OC Collinder 261, and Bertran de Lis et al. (2016) calculated a scatter of [O/Fe] of $\lesssim 0.01$ dex in several clusters.

In addition to chemical homogeneity within a single cluster, which is the focus of this paper, the efficacy of chemical tagging also depends on the degree of chemical overlap between different clusters. Intriguingly, some studies have identified chemically indistinguishable pairs of clusters (e.g., NGC 2458 and NGC 2420; Garcia-Dias et al. 2019), a high degree of overlap in chemical signatures between clusters (e.g., Blanco-Cuaresma et al. 2015), and pairs of chemically indistinguishable stars inside two distinct clusters (Ness et al. 2018).

Despite the tension in the literature findings, OCs are frequently assumed to be chemically homogeneous in cluster studies or for chemical tagging applications. However, there are theoretically and observationally motivated mechanisms that could cause OCs to be inhomogeneous to some level in certain elements. Dredge-up and atomic diffusion can drive abundance inhomogeneities between cluster members in different stellar evolutionary stages. For instance, Souto et al. (2019) found evidence of atomic diffusion in the open cluster M67, resulting in abundance differences of up to 0.1 dex between stars in different evolutionary states. Blanco-Cuaresma et al. (2015) also found variations in chemical signatures for stars belonging to different evolutionary stages within the same cluster. Models of planetary engulfment suggest this process could be responsible for some of the observed elemental scatter in OCs, like the Pleiades (Spina et al. 2018).

Aside from these mechanisms that can change the observable chemical composition of stars during their lifetime, inhomogeneities could also arise from intrinsic scatter present in the giant molecular cloud before star formation began or significant pollution over the many megayear timescale of the cluster's formation (Krumholz et al. 2019). The correlation length of elements in the initial cloud depends on the series of various enrichment events that produced them and on the efficiency of mixing of the ISM in that region. Using hydrodynamical simulations, Armillotta et al. (2018) found that chemical abundances in OCs should be correlated up to 1 pc, irrespective of the initial correlation lengths of the elements. Thus, given that a typical OC has an effective radius of around 4–5 pc (Kharchenko et al. 2013), we may see inhomogeneities in cluster abundances depending on the initial distribution of the elements before star formation began.

Bland-Hawthorn et al. (2010) suggest that all clusters with masses up to $10^4 M_{\odot}$, and a significant fraction of those up to $10^5 M_{\odot}$, are expected to be homogeneous. However, large star-forming clouds that form more-massive clusters may be subject to pollution from massive stars that become supernovae before star formation is complete. For example, the Sun is suggested to have formed in a cluster with a high-mass star that became a supernova while the Sun was still a protostar (Looney et al. 2006).

The many possible mechanisms described above that induce abundance scatter within OCs imply that the level of scatter may depend on properties like the nucleosynthetic groups of elements, range of evolutionary state, cluster size, cluster mass, etc. Any systematic difference between the levels of homogeneity for different metals (e.g., between alpha elements and

iron-peak elements) may indicate how different enrichment events can affect ISM mixing efficiency.

Although a few well-studied OCs have been shown to be chemically homogeneous within the observational uncertainties of large-scale surveys, there has not been a survey that has looked at a large number of clusters and systematically studied how their chemical homogeneity depends on various Galactic and cluster properties. So then it is necessary to look at these properties to seek out subtle patterns or behavior that can distinguish among mechanisms. This paper describes the first such systematic study of OC chemical homogeneity as a function of Galactic and cluster parameters such as Galactocentric distance (R_{GC}), vertical height ($|Z|$), age, and mass.

The paper is organized as follows: Section 2 describes the data that we have used from APOGEE (Section 2.1), Gaia (Section 2.3), distance catalogs (Section 2.4), and cluster catalogs (Section 2.5). We also describe the procedure we use to derive improved uncertainties for APOGEE abundances (Section 2.2). Section 3 explains our kinematics-based cluster membership selection (Section 3.1), the validation of our cluster members (Section 3.2), and the final catalog (Section 3.3). In Section 4, we describe how we quantify cluster chemical homogeneity, and Section 5 contains the analysis of cluster chemical homogeneity versus cluster and Galactic properties. Section 6 summarizes the main results from the paper.

2. Data

2.1. Apogee

We adopted stellar parameters, chemical abundances, and radial velocities (RVs) from the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017). APOGEE, one of the component surveys of the Sloan Digital Sky Survey IV (SDSS-IV; Blanton et al. 2017), is a high-resolution, near-infrared spectroscopic survey of $\sim 500,000$ stars across the Milky Way (Zasowski et al. 2013b, 2017). Observations are taken with two custom-built, 300 fiber spectrographs (Wilson et al. 2019), one at the 2.5 m Sloan Telescope at the Apache Point Observatory (Gunn et al. 2006) and one at the 2.5 m du Pont Telescope at Las Campanas Observatory (Bowen & Vaughan 1973). We use data from the 16th SDSS data release (DR16; Ahumada et al. 2019).

The pipelines that reduce the data and derive RVs are described in Nidever et al. (2015), and the APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP) is detailed in García Pérez et al. (2016). The DR16 ASPCAP values were derived by optimizing the comparison of the APOGEE spectra with synthetic spectra computed with Turbospectrum (Plez 2012) and MARCS model atmospheres (Gustafsson et al. 2008). The optimization is carried out using the FERRE code (Allende Prieto et al. 2006). The description of the APOGEE data products and abundance reliability for DR14 are presented in Holtzman et al. (2018) and Jönsson et al. (2018), while those for DR16 are described in Jönsson et al. (2020).

Because our cluster membership determination (Section 3) is based solely on kinematic properties, initially we only require reliable kinematical measurements: RVs (discussed here) and proper motions (PMs; Section 2.3). Additional cuts on data quality used for the analysis in Section 5 are described in Section 5.1. To ensure reliable kinematic cluster member

selection, we restrict our sample to have APOGEE RV uncertainties (VERR) of $<0.1 \text{ km s}^{-1}$. We also remove a small number of stars with implausibly large velocities by requiring $|\text{VHELIO_AVG}| < 5000 \text{ km s}^{-1}$.

To remove potential binaries from our sample, which may inflate the characteristic velocity signatures inferred for our clusters, we remove stars with visit-to-visit RV variations (VSCATTER) $>1 \text{ km s}^{-1}$ (Badenes et al. 2018; Price-Whelan et al. 2020) and stars found in the binary catalog of Price-Whelan et al. (2020). Given that the time baseline for the majority of APOGEE sources is less than a year, APOGEE is sensitive to detect binaries with periods less than a few years and separated by distances less than a few astronomical units (Price-Whelan et al. 2020). We note that because the VSCATTER limit that we impose is more sensitive to massive binaries, which tend to be more centrally concentrated, and that binary properties are correlated with metallicity (e.g., Badenes et al. 2018; Moe et al. 2019), it is conceivable that these binary rejection cuts could induce some metallicity-dependent spatial sampling patterns. We confirmed that this is not the case in our sample because this limit removes a tiny fraction of stars ($<5\%$), whose spatial distributions are indistinguishable from those that pass these cuts.

2.2. Rederived Abundance Uncertainties

2.2.1. Motivation and Outline

Earlier studies have highlighted the possibility that the uncertainties for some of the APOGEE abundances in earlier data releases are overestimated (e.g., Ness et al. 2018), and our own preliminary analyses of the abundance dispersions in our clusters supported this assessment. Prior to DR16, uncertainties on the abundances in the APOGEE data releases were determined by examining the spread of abundances in well-sampled clusters (Holtzman et al. 2018), which required assuming that the cluster had no intrinsic spread. This assumption makes the uncertainties unsuitable for the study of chemical homogeneity.

To address this, we recalculate the random abundance uncertainties for the stars in our analysis using an improved method.¹³ We worked closely with the APOGEE team to subsequently adapt our approach into the DR16 uncertainty determination. The approach used in DR16 (Jönsson et al. 2020) relies on a parametric fit to calculate the uncertainties for all the APOGEE stars. In this paper, however, we adopt a nonparametric approach because we find that a simple analytic function cannot adequately capture the relationships between uncertainty and signal-to-noise ratio (S/N), T_{eff} , and [M/H].

In brief, we use the differences in $[\text{X}/\text{Fe}]$ values derived by ASPCAP for independent spectra of the same star to compute a relationship between the abundance uncertainties and S/N, T_{eff} , and [M/H] (Section 2.2.2). Then, we use this relationship to compute the uncertainties for our cluster members (Section 2.2.3).

¹³ We emphasize that this empirical procedure for determining the uncertainties in the abundances accounts only for the random component of the uncertainties, which is essential for deriving the intrinsic abundance scatter in a cluster. Any systematic components, such as stellar parameter-dependent abundance variations due to departures from LTE or hydrostatic equilibrium, or systematics in the atomic data or instrumental distortions, are not captured by this procedure.

2.2.2. Calculating Uncertainties Using Multiple Visits

In general, the ASPCAP pipeline is run on the stacked spectrum of each star, which comprises all visits to that star. For a subset of APOGEE stars, however, ASPCAP is run on the individual visit-level spectra, providing multiple independent sets of stellar parameter and abundance measurements for single stars.¹⁴ We use these sets to estimate the random uncertainty of the ASPCAP measurements as a function of S/N, T_{eff} , and [M/H]. The variations between ASPCAP values for spectra of the same star, at the same S/N, provide a more realistic representation of the random measurement uncertainties than the ones derived from cluster dispersions in earlier APOGEE DRs.

We define a sample of stars, hereafter called the Uncertainty Training (UT) sample, with ASPCAP solutions derived from two or more visit spectra with similar S/Ns ($\Delta(\text{S/N})/(\text{S/N}) \leq 20\%$), resulting in similar temperature ($\Delta T_{\text{eff}} \leq 100 \text{ K}$) and metallicity ($\Delta [\text{M}/\text{H}] \leq 0.07 \text{ dex}$) values. These similarity criteria are imposed to ensure that differences in $[\text{X}/\text{Fe}]$ are not due to different global spectral fits. We restrict our analysis to giant stars (using the ASPCAP_CLASS column and a limit of $\log g < 3$).

The final UT sample of 8729 stars is then divided into five bins of S/N: 50–70, 70–100, 100–130, 130–200, and >200 . These bins are chosen to provide finer sampling at lower S/N, where the effect of the S/N on output parameters is larger; above $\text{S/N} \approx 150$, there is a weak relationship with uncertainties. We explored dividing the sample into finer bins in S/N. However, there were no significant changes in the final derived uncertainties, and there was an increased risk of undersampling each bin so that the estimated uncertainty would not capture the effect of the varying range of stellar parameters.

We adopt a Voronoi binning scheme in $T_{\text{eff}}\text{--}[M/H]$, within these fixed S/N ranges, to ensure both reliable measurements of the uncertainties in less populated regions of the parameter space and high-resolution measurements where possible. The Python package vorbin (Cappellari & Copin 2003) is used to group the UT sample into 2D bins of T_{eff} and [M/H], targeting at least 30 stars per bin. The final bins are populated with between 9 and 69 stars per bin, with an average of around 33.

The differences between pairs of visit-level $[\text{X}/\text{Fe}]$ values for individual stars can be used to compute the standard deviation of the distribution from which the pairs were originally drawn. We assume this distribution to represent the intrinsic ASPCAP random uncertainty. The quantities are related by

$$e_{[\text{X}/\text{Fe}],k} = \frac{\sqrt{\pi}}{2} \text{median}(|[\text{X}/\text{Fe}]_i - [\text{X}/\text{Fe}]_j|), \quad (1)$$

where $e_{[\text{X}/\text{Fe}],k}$ is the abundance uncertainty associated with the k th bin of T_{eff} and [M/H] at a given S/N, and $[\text{X}/\text{Fe}]_i$ and $[\text{X}/\text{Fe}]_j$ refer to abundance measurements derived from two independent visit spectra of the same star. The top row of Figure 1 shows the distribution of UT stars in [M/H] and T_{eff} , in bins of S/N, colored by the $e_{[\text{Mg}/\text{Fe}],k}$ values. This demonstrates the complex pattern of $e_{[\text{Mg}/\text{Fe}],k}$ derived in this way and the difficulty of describing such behavior with simple analytical expressions. Similar figures for all elements we

¹⁴ These measurements are contained in the “allCal” file as part of APOGEE’s data releases.

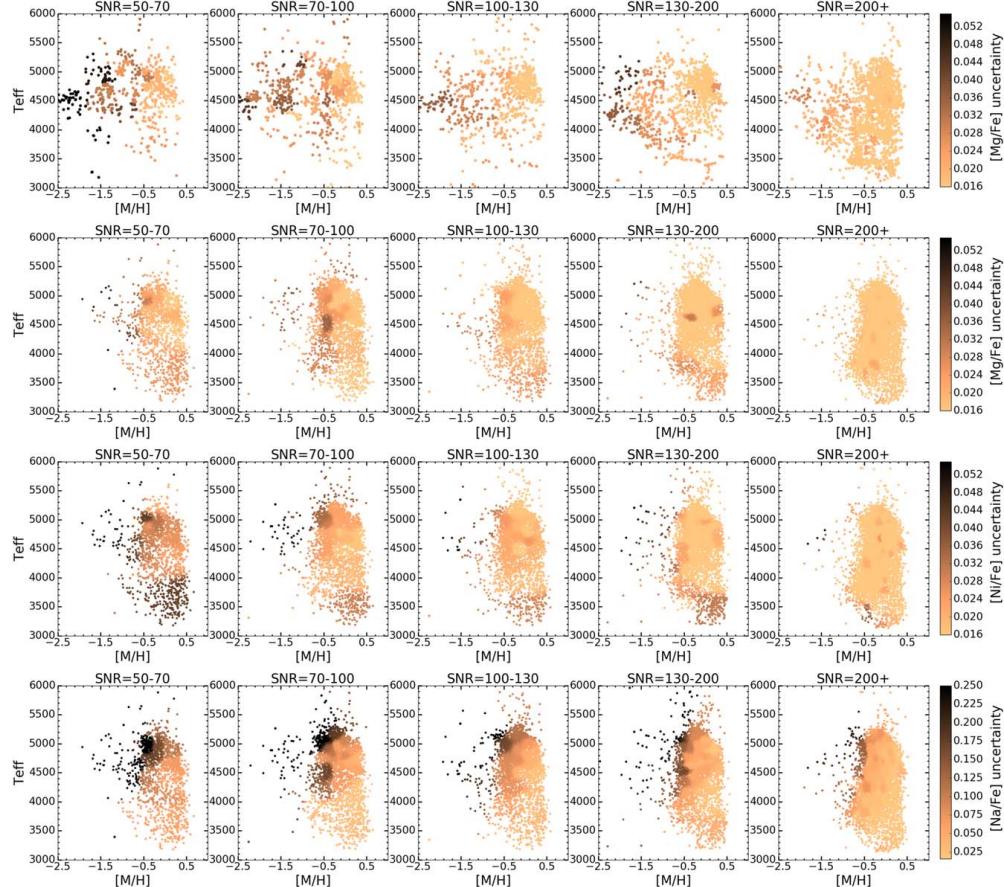


Figure 1. First row: the Uncertainty Training (UT) sample, divided by S/N and Voronoi-binned by $T_{\text{eff}} - [\text{M}/\text{H}]$, as described in the text (Section 2.2). The stars in each bin are colored by the computed $e_{[\text{X}/\text{Fe}],k}$ for that bin; [Mg/Fe] is shown here as an example. Note that many adjacent bins have nearly identical e values and are indistinguishable in this representation. Second row: weighted uncertainties for [Mg/Fe] of the cluster member sample (Section 3). The uncertainties assigned to the cluster members trace the pattern seen in first row. Third row: weighted uncertainties for [Ni/Fe] of the cluster member sample. Fourth row: weighted uncertainties for [Na/Fe] of the cluster member sample. Note that the range in the color bar has been increased.

analyze in Section 5 can be found in Appendix A (Figures A1 and A2).

We looked for potential dependencies of $e_{[\text{X}/\text{Fe}],k}$ on [X/Fe] itself—e.g., that stars with enhanced [Mg/Fe] may have different uncertainties than stars with solar [Mg/Fe] at the same T_{eff} , [M/H], and S/N—by deriving $e_{[\text{X}/\text{Fe}],k}$ in the same manner as above, but separately for stars with high and low [Mg/Fe]. We found no significant differences in the $e_{[\text{X}/\text{Fe}],k}(T_{\text{eff}}, [\text{M}/\text{H}], \text{S/N})$ patterns. We also found no difference in the results discussed in Section 5 when using uncertainties derived from the solar [Mg/Fe] part of the UT sample compared to when using the full sample.

2.2.3. Assigning Uncertainties to Stars

Given the computed array of $e_{[\text{X}/\text{Fe}],k}$ for each bin of T_{eff} , [M/H], and S/N, we can then sort any other star into a bin and assign it $e_{[\text{X}/\text{Fe}],k}$ values. We perform this sorting by training a

Gaussian naive Bayes classifier algorithm (Pedregosa et al. 2011) on the UT sample and then applying the trained classifier to our cluster member sample (Section 3.3). This results in a bin membership probability for each star; we assign each star an uncertainty summed from all of the Voronoi bins and weighted by each bin’s probability for that star:

$$e_{[\text{X}/\text{Fe}]} = \sum_k p_k e_{[\text{X}/\text{Fe}],k}, \quad (2)$$

where k is the bin index, p_k is the probability of bin k for this star, and $e_{[\text{X}/\text{Fe}]}$ is the final uncertainty for the star. Adopting this weighted average uncertainty ensures that all stars falling within a range of stellar parameters will not be assigned identical values for their abundance uncertainties due solely to the binning scheme. This approach also smooths the transitions between the bin edges. Ultimately, the uncertainties assigned to a star with stellar parameters near a bin edge are not driven by

the distribution of the bins but by the uncertainties estimated from the UT sample that fall within a neighboring area of the star in [M/H] and T_{eff} .

The bottom three rows of Figure 1 show examples of these rederived uncertainties for [Mg/Fe], [Ni/Fe], and [Na/Fe] (note the differences in color scaling) along [M/H] and T_{eff} , in bins of S/N. All show the expected improvement in precision with higher S/N. Other generic patterns are also clearly visible—for example, the increase in uncertainty at low metallicities and/or high temperatures, where lines become weaker, and at very low temperatures, where lines become increasingly blended. Mg and Ni have lower uncertainties compared to Na, which is expected due to the difficulty in measuring Na lines in APOGEE. Each element also has its own unique patterns, reflecting the range of difficulty in measuring lines of different elements in different parts of stellar parameter space.

The DR16 uncertainties are not systematically higher or lower than the uncertainties derived here for stars in our cluster sample (Section 3), and we find qualitatively similar results for the analysis in Section 5 if we use either set of values.

2.3. Proper Motions

For the cluster membership selection in Section 3, we use PMs from DR2 of the Gaia mission (Gaia Collaboration et al. 2018). We require that the errors in the PM measurements be smaller than 2.0 mas yr⁻¹ and the renormalized unit weight error (RUWE) be less than 1.4 (Ziegler et al. 2020). In addition, cluster-specific limits are imposed on the spatial distribution, magnitude, and color of the stars, as described in Section 3.1.3.

2.4. Stellar Distances

For our analysis in Section 5, we use spectrophotometric distances calculated using the method described in Rojas-Arriagada et al. (2017, RA17). We also use the StarHorse (Queiroz et al. 2018, 2019) and astroNN (Leung & Bovy 2019) distances to compare with our RA17 estimates. The results of this paper are not affected by the choice of the distance catalog used to determine cluster distances. Further discussion can be found in Section 5.

2.5. Literature Cluster Parameters

We adopt the Milky Way Star Clusters catalog (Kharchenko et al. 2013, hereafter K13) as the base catalog for our membership search Section 3. We use K13 cluster center coordinates and angular radii to define the search limits, and we consider the cataloged distances, ages, and metallicities (in conjunction with APOGEE-derived values) when assessing our membership selection. From the total sample of 3208 clusters in K13, we only consider the 366 clusters that have six or more APOGEE stars (Section 2.1) within their cluster radii (Section 3.1.1). We narrow our cluster sample to the 10 most-populated high-quality clusters for our analysis in Section 5.

3. Cluster Membership

The first step of our open cluster analysis is to identify cluster members in the APOGEE sample. Numerous methods have been demonstrated in the literature, typically adopting some combination of RVs, PMs, metallicities, and position in the CMD (e.g., Frinchaboy & Majewski 2008; Mészáros et al. 2013;

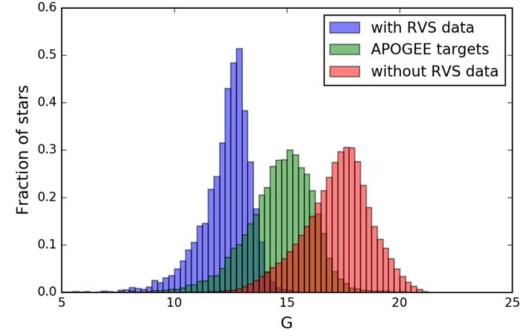


Figure 2. Gaia-DR2 G -band magnitude of stars with (blue) and without (red) Gaia/RVS data (normalized separately), compared to the APOGEE stars (green) belonging to our clusters. Because Gaia-DR2 only has RVs for stars with $G \lesssim 13$, and most of our clusters have RGB stars fainter than this, we do not use Gaia/RVS for the RVs in our selection.

Donor et al. 2018). As we are interested in the chemical homogeneity of the clusters, we design our membership selection around kinematical information only: RVs from APOGEE (Section 2.1) and PMs from Gaia-DR2 (Section 2.3). This is similar in spirit to the approach taken by Cantat-Gaudin et al. (2018), hereafter C18, who use Gaia-DR2 information only.

Figure 2 shows the distribution of the Gaia-DR2 G -band magnitude of stars with (blue) and without (red) Gaia/RVS data, compared to the APOGEE stars (green) in the vicinity of several of our cluster candidates. This figure highlights why APOGEE RVs are necessary for the objects in our sample; due to a combination of distance and extinction, most of our stars are too faint to have Gaia-DR2 RVS radial velocities.

3.1. Method

Figures 3–5 demonstrate the procedure described below for a well-studied open cluster (NGC 6819), a poorly studied cluster (FSR 0494), and a K13 cluster not recovered by our membership method (ASCC 116), respectively.

3.1.1. Cluster Coordinates

For each cluster, we start our membership search with APOGEE and Gaia-DR2 stars within twice the cluster radius, $2R_{\text{cluster}}$, using central coordinates and “total” cluster radii (their $r_2 = R_{\text{cluster}}$) from K13 (Section 2.5). The “central stars” (within $1R_{\text{cluster}}$) define the cluster’s kinematical signature, and the “annulus stars” (between $1.5–2R_{\text{cluster}}$) define the background distribution (e.g., Figures 3(a)–(b)). We only consider the 366 K13 clusters that have six or more stars within $1R_{\text{cluster}}$ in the APOGEE catalog that meet the quality criteria above.

3.1.2. Radial Velocities

We search for RV peaks associated with comoving stars in each cluster (e.g., Figure 3(c)) location by subtracting a kernel density estimate (KDE) of the annulus stellar RV distribution (shown in green) from that of the central stars (shown in blue). The residual (shown in red) is then fit with a Gaussian to determine the central RV ($\langle RV \rangle$) and the width (σ_{RV}) of the dominant peak. We also measure the ratio of the Gaussian amplitude (A_{RV}) to the standard deviation of the residuals (σ_{resid}) more than $3\sigma_{RV}$ away from the Gaussian center; this

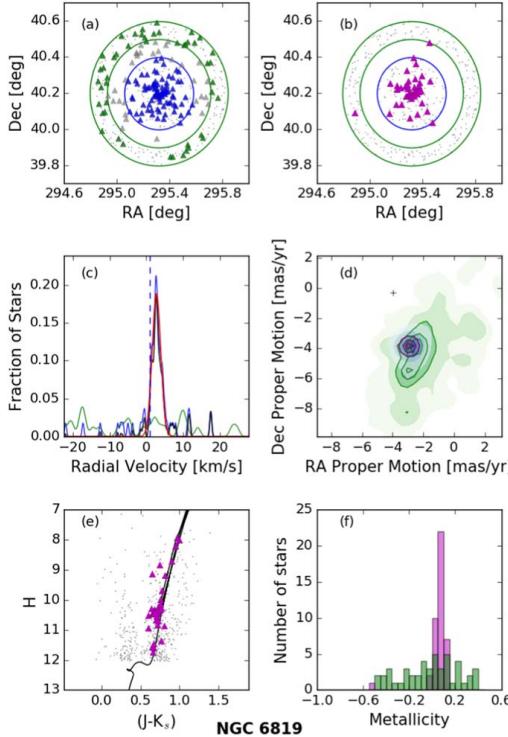


Figure 3. Proof-of-concept membership selection for NGC 6819: APOGEE stars within the annulus and central regions are shown in green and blue, respectively, while final cluster members are shown in purple. (a) and (b) Stellar distribution in R.A. and decl. APOGEE and Gaia-DR2 stars are shown as triangles and points, respectively. (c) and (d) Distribution of RVs and PMs. Fits for the subtracted distributions in RV and PM are shown in red. Diagnostic plots for final cluster members: (e) color-magnitude diagram along with a Padova isochrone corresponding to the cluster. (f) Metallicity distribution of cluster stars as compared to annulus stars. See text for details.

metric quantifies the strength of the RV signal. Visual inspection demonstrates that for our clusters, an $A_{\text{RV}}/\sigma_{\text{resid}} > 9.5$ corresponds to a cluster in RV space. Smaller values tend to be dominated by noisy residuals driven by a low number of annulus stars.

3.1.3. Proper Motions

We obtain PM information for all stars within $2R_{\text{cluster}}$ using the Gaia TAP+ query from the astroquery package in Python and keep stars that pass the quality cuts mentioned in Section 2.3.

Next, we ensure that the Gaia-DR2 data sample has the same color-magnitude range as the APOGEE stars by matching to the “apogeeObject” files used in the APOGEE targeting pipeline, which contains the 2MASS (Skrutskie et al. 2006), Spitzer-IRAC GLIMPSE (Benjamin et al. 2005; Churchwell et al. 2009), and AllWISE (Wright et al. 2010; Cutri et al. 2013) photometry used to calculate extinction (Majewski et al. 2011; Zasowski et al. 2013b, 2017). We then restrict the Gaia-DR2 stars to the same $(J - K_s)_0$ and H limits

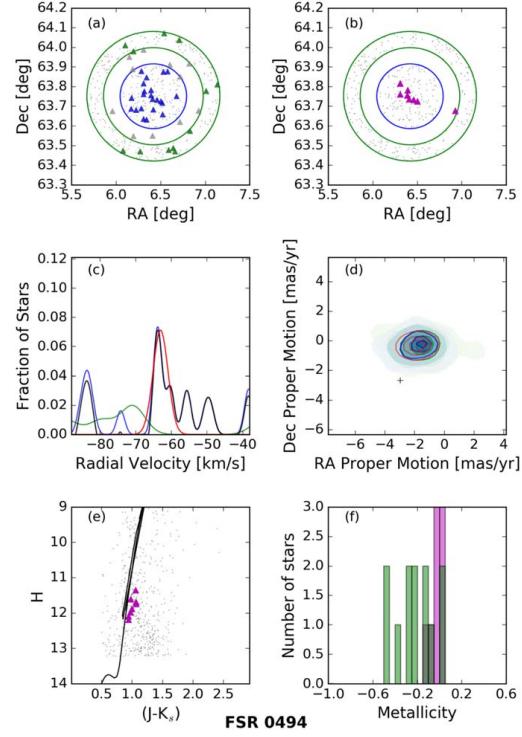


Figure 4. Same as Figure 3, but for FSR 0494, a lesser studied OC.

sampled by the APOGEE stars in the vicinity of that cluster (generally $(J - K_s)_0 \geq 0.5$ and $7 \leq H < 12.2$).

We use this cut to ensure that the PM distribution obtained from Gaia-DR2 stars is an accurate representation of the APOGEE stars that we are considering for membership. However, we compare the membership with and without using this cut for all 10 clusters studied in Section 5. Although the PM distribution is altered slightly, the final cluster members determined are the same for these clusters irrespective of the color-magnitude cut.

In a small fraction of cases (6%), the apogeeObject files do not span the full background annulus region, but we have confirmed that the distributions of $\mu_{\text{R.A.}}$ and $\mu_{\text{decl.}}$ do not change across the small angular scales of our clusters, so we consider even these partial annuli to be representative of the background distributions.

As with the RVs (Section 3.1.2), we compare KDEs of the central and annular distributions to identify any signal of the cluster, this time with 2D KDEs ($\mu_{\text{R.A.}} \times \mu_{\text{decl.}}$, shown in Figure 3(d)). Because the annular PM distribution is much less noisy than in the case of RVs, we model the entire central PM distribution (shown in blue) as the sum of a 2D Gaussian and a scaled copy of the annular PM distribution (shown in green). The best-fit Gaussian (shown in red) center ((PM)) and 2D dispersion ($\sigma_{\text{PM}}(\text{R.A., decl.})$) are taken as the PM distribution of the comoving cluster stars.

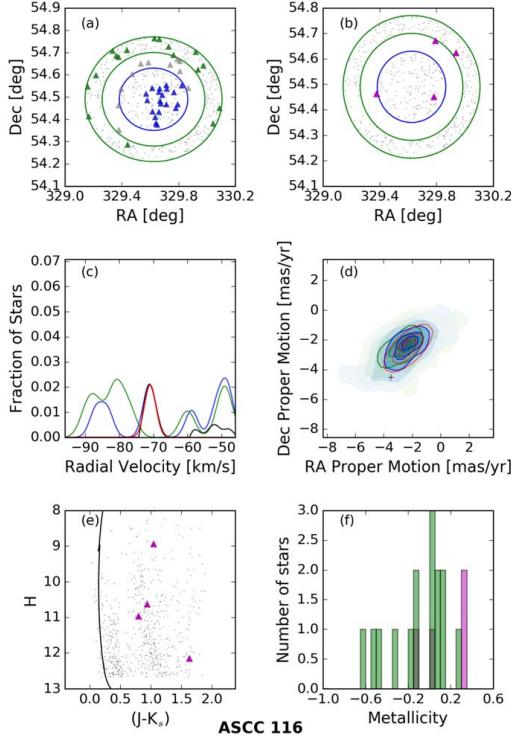


Figure 5. Same as Figure 3, but for ASCC 116, shown here as an example of a cluster where the diagnostic plots do not confirm the presence of an OC.

3.1.4. Computing Membership Probabilities

We first compute cluster membership probabilities, based on RVs and PMs, for each star within $2 R_{\text{cluster}}$ of a cluster. These probabilities are the values of Gaussian distributions with the means and standard deviations derived from the RV and PM fitting in Sections 3.1.2–3.1.3, scaled to have a maximum value of unity. We consider as likely members stars that fall within a 3σ window (shown in purple in Figures 3–5) on the combined RV–PM probability. We give equal weighting to each kinematic dimension while calculating the combined probability, because weighting each dimension on how distinct it is from the background did not yield any changes in the final selected cluster members. We choose a selection window of 2σ on the combined probability for the analysis in Section 5 because we observed a few outliers in the metallicity distribution (e.g., Figure 3(f)) that were removed when we used a stricter cut.

Examples of this entire procedure and its results are demonstrated in Figures 3–5. The color coding for all panels is as follows: locations and kinematical information for the central APOGEE stars are plotted in blue, for the annulus APOGEE stars in green, for the final members in purple, and for the Gaia-DR2 stars used for the PM background in gray points. The top row (panels (a) and (b)) show the on-sky distribution of stars, with the inner R_{cluster} in a blue circle and the outer annulus enclosed in green circles at 1.5 and $2 R_{\text{cluster}}$.

Panel (a) shows the stars used in characterizing the cluster (Sections 3.1.2–3.1.3), and panel (b) shows the final cluster members (Section 3.1.4).

The middle row (panels (c) and (d)) show the RV and PM distributions of the central and annulus stars, with the fitted residual peaks in red. The bottom row (panels (e) and (f)) are not used for membership selection and are only examined when setting reliability flags for entire clusters (Section 3.2).

Figure 3 shows the recovery of the well-studied cluster NGC 6819 (e.g., Hole et al. 2009; Platais et al. 2013; Yang et al. 2013; Wu et al. 2014; Lee-Brown et al. 2015), Figure 4 shows the recovery of the less-studied cluster FSR 0494 (Froebrich et al. 2007; Zasowski et al. 2013a; Donor et al. 2018), and Figure 5 shows the nonrecovery of the cluster cataloged as ASCC 116 (e.g., Kharichenko et al. 2005; Cantat-Gaudin et al. 2018).

3.2. Validation

In the membership selection examples in Figures 3–5, the left side of the bottom row (panel (e)) shows the $(J - K_s, H)$ color–magnitude diagram of the cluster members and background stars, along with a shifted PARSEC isochrone (Bressan et al. 2012; Marigo et al. 2017) corresponding to the cluster’s distance, metallicity, and extinction, either known from K13 or approximated from the cluster members themselves. The right side of the bottom row (panel (f)) shows the APOGEE metallicity distributions of the annulus stars (green) and the kinematically selected member stars (purple). We use these two pieces of data when setting reliability flags for clusters in the final catalog (Section 3.3). Although we do not use metallicity in determining cluster members, we use it to flag clusters that do not have a clear and distinct metallicity distribution function (MDF), compared to the background stars, described below.

We classify clusters into “GOOD,” “WARN,” “INSUFFICIENT_DATA,” and “UNRECOVERED” categories based on our confidence in the recovery of genuine cluster members:

1. **GOOD:** clusters that have distinct kinematics ($A_{\text{RV}}/\sigma_{\text{resid}} > 9.5$; Section 3.1.2) and metallicity dispersion less than 0.13 dex.
2. **WARN:** clusters that have distinct kinematics ($A_{\text{RV}}/\sigma_{\text{resid}} > 9.5$; Section 3.1.2) and metallicity dispersion greater than 0.13 dex.
3. **INSUFFICIENT_DATA:** clusters that have distinct kinematics ($A_{\text{RV}}/\sigma_{\text{resid}} > 9.5$; Section 3.1.2) but fewer than five cluster members, making it difficult to interpret the diagnostic CMD and MDF distributions (e.g., Figures 5(e)–(f)).
4. **UNRECOVERED:** Clusters that do not have distinct kinematics ($A_{\text{RV}}/\sigma_{\text{resid}} < 9.5$; Section 3.1.2).

We observed that the cluster metallicity distributions are generally either very tight or indistinguishable from the background. By visual inspection of all the clusters, this bifurcation is captured using a metallicity dispersion cut of 0.13 dex, with clusters having a higher metallicity dispersion classified using the “WARN” flag.

Clusters that have too few stars to be validated using the diagnostic plots are included in the “INSUFFICIENT_DATA” classification. Figure 5 shows an example of such an object, where our membership method fails to confirm a comoving object at the location of a K13 cluster. Such cases are expected, because APOGEE uses specific sets of color and magnitude

Table 1
Sample Table of Cluster Members Selected Using Our Membership Selection in Section 3

APOGEE	Cluster	RAdeg deg	DEdeg deg	N σ_{RV}	N σ_{PM}	Dist arcminutes
2M00000068+5710233	NGC 7789	0.0029	57.1732	37.79	24.49	33.99
2M00001199+6114138	NGC 7790	0.0500	61.2372	12.24	3.93	13.36
2M00001328+5725563	NGC 7789	0.0554	57.4323	8.70	5.09	47.91
2M00002012+5612368	NGC 7789	0.0839	56.2102	33.14	2.58	39.47
2M00002853+6119307	NGC 7790	0.1189	61.3252	5.35	1.33	16.94

(This table is available in its entirety in machine-readable form.)

limits and does not target the entire sky homogeneously. We include Figure 5 to highlight the robustness of this method and validation for false positives. For the analysis in Section 5, we only use clusters that have a “GOOD” validation flag.

3.3. Catalog

We generate two catalogs based on the kinematic membership selection above. One contains all of the stars within $2R_{\text{cluster}}$ that meet the membership criteria described in Section 3.1.4 for the GOOD, INSUFFICIENT_DATA, and WARN clusters (Section 3.2). This includes, for each star, the APOGEE ID and stellar coordinates, the name of and distance (in arcminutes) from the center of the cluster to which it belongs, and the number of sigmas from the center of the membership probability distribution (Section 3.1.4) in both RV and PM dimensions. A sample of this table is shown in Table 1, which is published in its entirety in machine-readable format.

The second catalog contains the properties of the clusters themselves, outlined in Table 2. For each GOOD, INSUFFICIENT_DATA, and WARN cluster, we give the central coordinates, radius, and age from K13, along with the average distances and metallicities of the member stars, a suite of kinematic fitting parameters and metrics, and other metadata. Table 2 is published in its entirety in machine-readable format. A subsample of Table 2 containing important kinematic and chemical information for 10 ten OCs used in Section 5 is shown in Table 3. Of the 366 K13 clusters with 6 or more APOGEE stars within $1R_{\text{cluster}}$, 34 are included with the GOOD flag, 38 have WARN, 11 have INSUFFICIENT_DATA, and 283 are flagged as UNRECOVERED.

Figure 6 summarizes several properties of the GOOD clusters from our catalog. Figure 6(a) shows the clusters’ Galactic $R_{\text{GC}}\text{-}Z$ distribution, and Figure 6(b) shows the distribution of their mean [M/H] and log (age) values. In Figure 6(c), we plot the distribution of mean [M/H] and [Mg/Fe] over a background of APOGEE stars with similar Galactic radius and height ($R_{\text{GC}} = 5\text{--}15$ kpc and $|Z| < 2$ kpc), selected using the same quality criteria described in Sections 2.1 and 5. Figure 6(d) shows a histogram of the number of cluster members identified, with the cutoff of nine members used in Section 5 indicated with a red dashed line.

We find member sample sizes similar to those of Donor et al. (2020), who also studied APOGEE DR16, for the clusters in common. We calculate a metallicity gradient (using [M/H]) of -0.096 ± 0.016 dex kpc $^{-1}$ for the sample, spanning $R_{\text{GC}} = 7\text{--}12$ kpc and $|Z_{\text{GC}}| < 1$ kpc. This value is within the uncertainties of, but slightly steeper than, previous calculations of the metallicity gradient (e.g., -0.079 ± 0.005 and -0.085 ± 0.019 dex kpc $^{-1}$; Jacobson et al. 2011; Donor et al. 2018).

Table 2
Columns from the Table of Catalog Clusters

Column	Description
Cluster	Name of cluster
RAdeg	Central R.A. ^a [degrees]
DEdeg	Central decl. ^a [degrees]
Radius	Adopted cluster radius ^a [arcminutes]
Dist	Median RA17 spectrophotometric distance of cluster members [kiloparsecs] (Section 2.4)
e_Dist	Dispersion of spectrophotometric distance [kiloparsecs]
logAge	Log (age) of cluster ^a [dex]
e_logAge	Uncertainty in log(age) of cluster ^a [dex]
M/H	Mean [M/H] of cluster members [dex] (Section 5.4)
e_M/H	Standard deviation of [M/H] of cluster members [dex]
RV	Mean of best-fit Gaussian to RVs [km/s] (Section 3.1.2)
e_RV	Standard deviation of best-fit Gaussian to RVs [km/s] (Section 3.1.2)
ARV	Amplitude of best-fit Gaussian to RVs (Section 3.1.2)
pmRA	Mean μ_{α} of best-fit Gaussian to $\mu_{\alpha} \times \mu_{\delta}$ [mas/year] (Section 3.1.3)
pmDE	Mean μ_{δ} of best-fit Gaussian to $\mu_{\alpha} \times \mu_{\delta}$ [mas/year] (Section 3.1.3)
e_pmRA	Standard deviation in RA of best-fit Gaussian to $\mu_{\alpha} \times \mu_{\delta}$ [mas/year] (Section 3.1.3)
e_pmDE	Standard deviation in DEC of best-fit Gaussian to $\mu_{\alpha} \times \mu_{\delta}$ [mas/year] (Section 3.1.3)
Apm	Amplitude of best-fit Gaussian to $\mu_{\alpha} \times \mu_{\delta}$ (Section 3.1.3)
Rpm	Rotation angle [radians] (Section 3.1.3)
Spm	Scale of the annular PM distribution (Section 3.1.3)
ARV/sigma	Amplitude over residual for RV (Section 3.1.2)
Nmem	Number of selected cluster members
FLAG	Validation flag for the cluster (Section 3.2)
Mg/Fe	Mean [Mg/Fe] of cluster members [dex]
e_Mg/Fe	Standard deviation of [Mg/Fe] of cluster members [dex]
Z	Vertical distance from the Milky Way disk [kpc] (Section 5.4)
Rgc	Galactocentric distance of cluster [kpc] (Section 5.4)
e_Rgc	Uncertainty in R_{GC} [kpc]
3Dvel	3D velocity dispersion [kpc] (Section 5.5)
e_3Dvel	Uncertainty in 3D velocity dispersion

Note.

^a From K13.

(This table is available in its entirety in machine-readable form.)

We also compared our membership with the C18 membership for our GOOD clusters. Considering stars with APOGEE observations, C18 has about 5% more members for each cluster than we do, but these stars typically have RVs inconsistent with the peak of the cluster. For a few clusters, we find additional members (about 4% of the total) that are not present in C18. These stars do not have measured Gaia-DR2 parallaxes, and we

Table 3
A Subsample of Columns from Table 2 for the 10 Clusters Used in Section 5

Cluster	Dist ± e_Dist kpc	RV ± e_RV km s ⁻¹	pmRA ± e_pmRA mas yr ⁻¹	pmDE ± e_pmDE mas yr ⁻¹	Nmem	M/H ± e_M/H dex	Mg/Fe ± e_Mg/Fe dex
NGC 1245	3.19 ± 0.19	-29.18 ± 0.79	0.55 ± 0.57	-1.67 ± 0.49	26	-0.080 ± 0.025	-0.028 ± 0.024
NGC 188	1.85 ± 0.18	-41.96 ± 0.33	-2.32 ± 0.54	-0.94 ± 0.52	29	0.100 ± 0.029	0.033 ± 0.026
NGC 2204	3.69 ± 1.19	92.09 ± 1.01	-0.54 ± 0.55	1.96 ± 0.51	27	-0.282 ± 0.096	0.014 ± 0.049
NGC 2420	2.19 ± 0.43	74.22 ± 0.93	-1.15 ± 0.50	-2.16 ± 0.59	18	-0.201 ± 0.067	0.004 ± 0.027
NGC 2682	0.75 ± 0.13	34.05 ± 0.66	-10.98 ± 0.55	-2.95 ± 0.56	381	-0.007 ± 0.058	0.004 ± 0.034
NGC 6705	0.93 ± 0.61	35.51 ± 1.65	-2.31 ± 1.29	-5.05 ± 0.79	15	0.172 ± 0.056	-0.058 ± 0.036
NGC 6791	4.65 ± 0.95	-47.05 ± 1.39	-0.42 ± 0.52	-2.27 ± 0.51	59	0.346 ± 0.050	0.099 ± 0.034
NGC 6819	2.37 ± 0.61	2.74 ± 1.18	-2.93 ± 0.54	-3.88 ± 0.57	48	0.057 ± 0.099	-0.008 ± 0.022
NGC 7789	1.97 ± 0.76	-54.77 ± 1.12	-0.91 ± 0.51	-1.95 ± 0.53	25	-0.018 ± 0.085	-0.017 ± 0.033
VDB 131	2.39 ± 0.53	-31.89 ± 1.68	-6.00 ± 0.67	0.19 ± 0.72	13	0.152 ± 0.071	-0.038 ± 0.033

Note. Shown are the mean cluster heliocentric distances, metallicities, [Mg/Fe] abundances, and the means of the Gaussian fits for each kinematic dimension, along with their standard deviations. “VDB 131” is short for VDBERGH-HAGEN 131.

believe that this is the reason they have been excluded from C18. However, these additional members we find do have measured RA17, StarHorse, and astroNN distances (Section 2.4) that are generally similar to the distances of the members common to both membership catalogs. We repeated the analysis in Section 5 using only the common members and found similar results and interpretations.

4. Quantifying Cluster Homogeneity

To study the chemical homogeneity of the kinematically identified cluster members in Section 3, we need a robust homogeneity metric that takes into account the members’ nonuniform abundance uncertainties (Section 2.2). Previous studies of intrinsic abundance scatter adopted a variety of metrics, such as the rms of the abundances (De Silva et al. 2007) and a χ^2 -like measurement of the distance between pairs of stars in an N -dimensional chemical space (Ness et al. 2018).

We adopt a Maximum Likelihood Estimator (MLE) approach to determining the intrinsic abundance scatter of a group of stars, similar to the MLE in Kovalev et al. (2019). This choice is based on the speed and simplicity of the method, combined with its consistency with other tested metrics (see below).

Given a distribution of abundances [X/Fe] with their corresponding uncertainties, we can estimate the likelihood that these values were drawn from a Gaussian distribution centered at $\mu_{[X/Fe]}$ with a standard deviation of $\sigma_{[X/Fe]}$ using

$$L \equiv \prod_{i=1}^n \frac{1}{\sqrt{2\pi}(\sigma_{[X/Fe]}^2 + e_i^2)^{1/2}} \exp\left(\frac{-(x_i - \mu_{[X/Fe]})^2}{2(\sigma_{[X/Fe]}^2 + e_i^2)}\right), \quad (3)$$

where x_i is the chemical abundance of a particular element for a cluster member and e_i is the corresponding abundance uncertainty. By finding where the maximum of this function lies on the $\mu_{[X/Fe]} - \sigma_{[X/Fe]}$ plane shown in Figure 7, we can estimate the parameters of the Gaussian distribution from which these data points are drawn. Here, we are most interested in the value of $\sigma_{[X/Fe]}$, because it represents the intrinsic scatter of the abundances within the cluster. We estimate the asymmetric uncertainty in the value of $\sigma_{[X/Fe]}$ using the distribution of the likelihood function along the $\sigma_{[X/Fe]}$ axis. We

take the first and third quartile ranges of this distribution as the lower and upper uncertainty limits on $\sigma_{[X/Fe]}$.

We verified that this method can recover an input $\sigma_{[X/Fe]}$ value from mock abundances that have been perturbed by uncertainties assigned from stars in several of our clusters for a given element. During this test, we noticed the existence of a bias for the MLE estimator with respect to the number of stars in each cluster—specifically, clusters with fewer stars ($N < 15$) were systematically estimated to have lower scatter than the actual value. We resolved this issue by fitting this bias with an exponential function and scaling the derived MLE scatter based on the number of members in each cluster.

We looked at the distribution of individual stellar APOGEE RVs versus abundances in clusters to ensure that the calculated value of $\sigma_{[X/Fe]}$ was not being driven by outlier stars in each dimension. We ensured that there was no evident trend between the intrinsic scatter and dispersion of T_{eff} or $\log g$ for the clusters that we study. We verified for multiple clusters that the elemental abundance distributions followed a Gaussian distribution because this is an assumption intrinsic to the MLE method. We also studied how the $\sigma_{[X/Fe]}$ changed for stars that belonged to different evolutionary stages within the same cluster.

As a consistency check, we compared the MLE-based $\sigma_{[X/Fe]}$ to that estimated from other metrics. One other metric we considered compares the cumulative distribution of pairwise distances in N -dimensional chemical space of simulated abundances with real cluster data. This metric is more computationally expensive than the MLE method but produces results that are entirely consistent.

5. Results

5.1. Final Selection of Elements and Cluster Members

We select elements for our analysis from the full set available in APOGEE using a variety of criteria. Some elements are known to have issues with accurate abundance determinations with ASPCAP, at least in certain ranges of the stellar parameter relevant to our stars (e.g., S, K, Na, and Ti; Hawkins et al. 2016), and we discard these. We also remove C, N, and O from further analysis because the abundances of these elements are affected by different stages of dredge-up over the course of the evolution of the star.

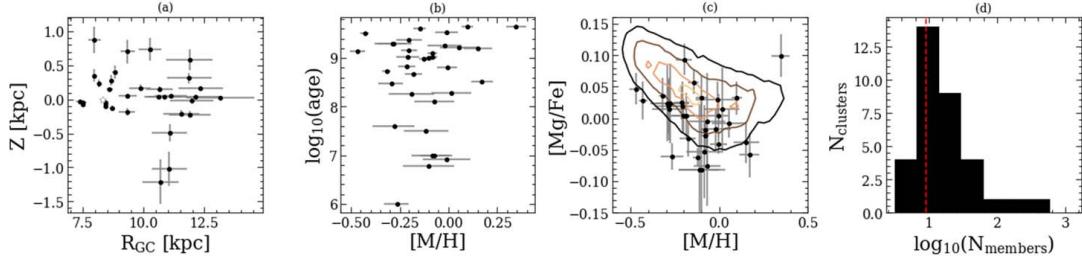


Figure 6. Summary of cluster catalog (Section 3.3). Panel (a) shows the distribution in R_{GC} and Z of clusters flagged as GOOD, using heliocentric distance estimates from the APOGEE member stars. Panel (b) shows the distribution of GOOD clusters in $[\text{M}/\text{H}]$ (from the APOGEE member stars) and $\log(\text{age})$ (from K13). Panel (c) shows the distribution of GOOD clusters in $[\text{M}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$ (against a background of MW stars). Panel (d) shows the histogram of the number of cluster members in the GOOD clusters. For the analysis in Section 5, we only use clusters with at least nine members, shown by the vertical dashed line.

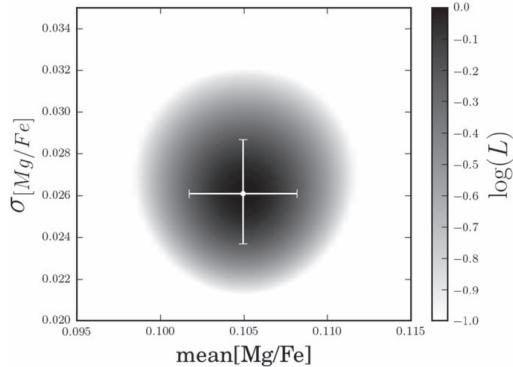


Figure 7. Example of the $\mu_{[\text{X}/\text{Fe}]} - \sigma_{[\text{X}/\text{Fe}]}$ plane of the likelihood function (Equation (3)) used to determine the intrinsic scatter ($\sigma_{[\text{X}/\text{Fe}]}$) for $[\text{Mg}/\text{Fe}]$ in NGC 6791.

We use the rederived uncertainties (Section 2.2) to calculate and compare the chemical homogeneity of OCs to that of groups of field stars selected to match the clusters in spatial extent and stellar parameters. A few elements (e.g., P, V, and Ce) for which field star samples show a lower scatter in abundance are removed from further analysis, because field stars chosen in this way are expected to have higher scatter in abundances than OCs. Based on these quality cuts, we only use $[\text{Mg}/\text{Fe}]$, $[\text{Al}/\text{Fe}]$, $[\text{Si}/\text{Fe}]$, $[\text{Ca}/\text{Fe}]$, $[\text{Cr}/\text{Fe}]$, $[\text{Mn}/\text{Fe}]$, $[\text{Ni}/\text{Fe}]$, and $[\text{Fe}/\text{H}]$ for the analyses in this section.

The intrinsic chemical scatter within clusters ($\sigma_{[\text{X}/\text{Fe}]}$) used in this section are derived from GOOD clusters (Section 3.2) with at least nine members (Section 3.1.4) that are within $1 R_{\text{cluster}}$ from the cluster center and that meet the following APOGEE bitmask¹⁵ criteria:

1. BRIGHT_NEIGHBOR and VERY_BRIGHT_NEIGHBOR==0 (STARFLAG bits 2 and 3),
2. SUSPECT_BROAD_LINES==0 (STARFLAG bit 17),
3. METALS_BAD==0 (ASPCAPFLAG bit 19),
4. ALPHAFE_BAD==0 (ASPCAPFLAG bit 20), and
5. STAR_BAD==0 (ASPCAPFLAG bit 23).

¹⁵ <https://www.sdss.org/dr16/algorithms/bitmasks/>

We further restrict our sample to giant stars (using the ASPCAP_CLASS designation and a limit of $\log g < 3$) with $S/N > 50$. This $\log g$ limit is implemented to remove stars whose abundances could potentially be affected by atomic diffusion (Souto et al. 2019; Semenova et al. 2020). Finally, we remove stars that lie in ranges of T_{eff} , $[\text{M}/\text{H}]$, and S/N in which the distribution of the observed visit-to-visit abundance variations (Section 2.2.2) is non-Gaussian. We find that in a small number ($\sim 3\%$) of the stellar parameter bins used to derive abundance uncertainties, a significant fraction of the stellar visit pairs result in abundance differences > 0.5 dex. Cluster members that fall in these bins are removed from measurements of the abundance scatter of that particular element, because the computed abundance uncertainty may not reflect the true deviation from the correct answer.

These limits result in 10 GOOD OCs with sufficient members in all elements for further analysis. We explored several combinations of these limits (e.g., the minimum number of stars required in each OC, the minimum S/N condition), all for which the final results and interpretations remain the same as described below. The membership plots for these selected clusters are included in Appendix B (Figure B.1). Note that most of the outliers in the MDFs of the clusters (panel f, shown in purple) fail to pass the quality cuts mentioned above and so are not included in the analysis.

5.2. Cluster Distances

We use stellar distances to compute the median cluster heliocentric distances, which are used to calculate the Galactocentric distance (R_{GC}), height above the midplane ($|Z|$) (Section 5.4), and space velocity dispersion (Section 5.5). We find extremely good agreement in these median distances using four distance catalogs: StarHorse, astroNN, RA17 (Section 2.4), and distances calculated using the Gaia-DR2 parallax of the cluster members.

The two exceptions are VDBERGH-HAGEN 131 and NGC 6705, where the four catalogs give median distances that vary by a factor of ~ 2 . VDBERGH-HAGEN 131 stands out as being the most heavily reddened ($E(J - K_s) \sim 0.6$) and the most differentially reddened ($\sigma_{E(J - K_s)} \sim 0.12$), as evident in its CMD in Appendix B. We observed that the CMD for VDBERGH-HAGEN 131 dereddened with the RA17 reddening estimates produces a tighter locus than with StarHorse and astroNN. Additionally, for both VDBERGH-HAGEN 131 and NGC 6705, the RA17 distances for the members have a

Table 4
Intrinsic Abundance Scatter (Section 4) and Space Velocity Dispersion (σ_{tot} ; Equation (4)) for the OCs Analyzed in Section 5

Cluster	$\sigma_{[\text{Fe}/\text{H}]}$ dex	$\sigma_{[\text{Mg}/\text{Fe}]}$ dex	$\sigma_{[\text{Al}/\text{Fe}]}$ dex	$\sigma_{[\text{Si}/\text{Fe}]}$ dex	$\sigma_{[\text{Ca}/\text{Fe}]}$ dex	$\sigma_{[\text{Cr}/\text{Fe}]}$ dex	$\sigma_{[\text{Mn}/\text{Fe}]}$ dex	$\sigma_{[\text{Ni}/\text{Fe}]}$ dex	σ_{tot} km s^{-1}
NGC 1245	0.0211 \pm 0.003	0.0246 \pm 0.003	0.0233 \pm 0.005	0.0154 \pm 0.003	0.0252 \pm 0.004	0.0582 \pm 0.009	0.0000 \pm 0.004	0.0127 \pm 0.002	5.41 \pm 1.32
NGC 188	0.0219 \pm 0.004	0.0143 \pm 0.004	0.0341 \pm 0.010	0.0000 \pm 0.002	0.0148 \pm 0.004	0.0135 \pm 0.009	0.0240 \pm 0.006	0.0074 \pm 0.003	3.10 \pm 0.77
NGC 2204	0.0422 \pm 0.007	0.0414 \pm 0.007	0.0511 \pm 0.011	0.0443 \pm 0.008	0.0224 \pm 0.005	0.0000 \pm 0.009	0.0134 \pm 0.007	0.0068 \pm 0.004	6.49 \pm 1.58
NGC 2420	0.0314 \pm 0.006	0.0160 \pm 0.008	0.0210 \pm 0.007	0.0000 \pm 0.003	0.0202 \pm 0.005	0.0430 \pm 0.011	0.0157 \pm 0.005	0.0000 \pm 0.003	4.18 \pm 0.99
NGC 2682	0.0263 \pm 0.004	0.0148 \pm 0.003	0.0132 \pm 0.005	0.0153 \pm 0.003	0.0000 \pm 0.003	0.0728 \pm 0.012	0.0134 \pm 0.004	0.0028 \pm 0.002	1.68 \pm 0.35
NGC 6705	0.0359 \pm 0.008	0.0124 \pm 0.004	0.0275 \pm 0.008	0.0112 \pm 0.003	0.0161 \pm 0.004	0.0326 \pm 0.009	0.0096 \pm 0.005	0.0072 \pm 0.004	3.50 \pm 0.68
NGC 6791	0.0491 \pm 0.004	0.0268 \pm 0.003	0.0709 \pm 0.007	0.0205 \pm 0.003	0.0263 \pm 0.003	0.0693 \pm 0.008	0.1146 \pm 0.010	0.0285 \pm 0.003	7.59 \pm 1.83
NGC 6819	0.0343 \pm 0.003	0.0081 \pm 0.002	0.0329 \pm 0.004	0.0211 \pm 0.002	0.0174 \pm 0.002	0.0246 \pm 0.005	0.0252 \pm 0.003	0.0100 \pm 0.002	4.24 \pm 0.98
NGC 7789	0.0318 \pm 0.004	0.0043 \pm 0.003	0.0300 \pm 0.006	0.0102 \pm 0.003	0.0128 \pm 0.003	0.0000 \pm 0.007	0.0245 \pm 0.004	0.0000 \pm 0.003	3.23 \pm 0.71
VDB 131	0.0771 \pm 0.014	0.0205 \pm 0.007	0.0576 \pm 0.002	0.0123 \pm 0.004	0.0183 \pm 0.006	0.0357 \pm 0.013	0.0331 \pm 0.010	0.0161 \pm 0.007	5.81 \pm 1.33

Note. “VDB 131” is short for VDBERGH-HAGEN 131.

slightly smaller dispersion, compared to the StarHorse, astroNN, and Gaia-DR2 parallax-based distance values.

So, we adopt the RA17 distances for these two clusters and for consistency for all of the clusters. We emphasize that the results described below are independent of the catalog used to calculate the distance.

5.3. Abundance Scatter in Clusters

We calculate the abundance scatter in 10 OCs for 8 elements (Mg, Al, Si, Ca, Fe, Si, Mn, and Ni) using the method discussed in Section 4. We measure a nonzero intrinsic scatter ($\sigma_{[\text{X}/\text{Fe}]}$) in most cases. From Table 4, we see that all clusters except NGC 2204, NGC 6791, and VDBERGH-HAGEN 131 have $\sigma_{[\text{Fe}/\text{H}]}$ very close to previously determined limits for the scatter in $[\text{Fe}/\text{H}]$ (a range of 0.02–0.04 dex; De Silva et al. 2007; Bovy 2016; Kovalev et al. 2019). Two of these three have $\sigma_{[\text{Fe}/\text{H}]}$ less than 0.05 dex, with the exception of VDBERGH-HAGEN 131, which is a lesser studied cluster with no previous abundance determinations or abundance scatter studies performed.

VDBERGH-HAGEN 131 also exceeds the limit (0.03 dex) predicted by Bovy (2016) for $\sigma_{[\text{Al}/\text{Fe}]}$. Although chemical abundances have been determined for some red giants in NGC 2204 (e.g., Jacobson et al. 2011; Carlberg et al. 2016), there have been no studies focused on its chemical homogeneity. We find $\sigma_{[\text{Mg}/\text{Fe}]}$, $\sigma_{[\text{Al}/\text{Fe}]}$, and $\sigma_{[\text{Si}/\text{Fe}]}$ in NGC 2204 to be higher than average literature limits (\sim 0.03 dex) for other OCs.

However, the most interesting case we observe is NGC 6791, a high-metallicity OC whose chemistry has been well studied (e.g., Cunha et al. 2015). We measure a value of $\sigma_{[\text{Mn}/\text{Fe}]}$ for NGC 6791 that is very high compared to the $\sigma_{[\text{Mn}/\text{Fe}]}$ values for the rest of our clusters. The $\sigma_{[\text{X}/\text{Fe}]}$ values for the other elements in NGC 6791 fall within the limits quoted by Bovy (2016), except for $\sigma_{[\text{Al}/\text{Fe}]}$ (limit \sim 0.03 dex). Donor et al. (2020) also report a particularly high uncertainty of 0.13 dex in their mean $[\text{Mn}/\text{Fe}]$ for this cluster, where their uncertainty is defined as the 1σ scatter in cluster $[\text{Mn}/\text{Fe}]$ abundances in APOGEE. We have verified that this atypically high measurement of $\sigma_{[\text{Mn}/\text{Fe}]}$ is not a result of nonmembers with discrepant $[\text{M}/\text{H}]$ measurements that may have been selected as members (e.g., $\sigma_{[\text{Fe}/\text{H}]} < 0.05$ for this cluster, which is highly unlikely if contamination were large). We have verified that Mn lines for NGC 6791 members can be reliably measured over a range of T_{eff} at high $[\text{M}/\text{H}]$. We also find no systematic increase in

random uncertainties at higher $[\text{M}/\text{H}]$ nor any systematic shift in $[\text{Mn}/\text{Fe}]$ abundances with T_{eff} .

We compared the abundance scatter between elements that are observed to have a high abundance variation in GCs and those that do not. Of the elements that are included in our study, Mg, Al, and in few cases Si are those that have confirmed observations of significant abundance scatter and anticorrelations in GCs (Gratton 2020). As described above, the abundance scatter in Mg, Al, and Si for NGC 2204 stands out above the literature limits for OCs. However, we do not observe a selectively higher abundance scatter in these elements for any of our other OCs.

5.4. Galactic Position, Age, and Metallicity

We find that cluster abundance scatter is uncorrelated with Galactocentric distance and vertical height from the plane of the Milky Way for all the elements we consider. Figures 8(a)–(b) shows examples of the trend of cluster scatter in $[\text{Mg}/\text{Fe}]$ with respect to Galactocentric distance and vertical height, respectively.

For ages, we use values from the K13 catalog that have reported uncertainties in their age measurements (7 out of the 10 clusters). We find that cluster scatter is uncorrelated with cluster age. An example plot is shown for $[\text{Mg}/\text{Fe}]$ in Figure 8(c). We calculate the mean metallicity ($[\text{M}/\text{H}]$) of each cluster, using its APOGEE members, and find that metallicity is uncorrelated with cluster scatter. An example plot is shown for $[\text{Mg}/\text{Fe}]$ in Figure 8(d).

In Figure 8(b), although we see a relatively higher correlation coefficient compared to the rest of the subplots, we do not believe that this shows the presence of a significant correlation because this trend is not evident in any other element ($[\text{X}/\text{Fe}]$ or $[\text{Fe}/\text{H}]$) that we consider. We also looked for correlations between chemical scatter and these Galactic/cluster properties in selected subgroups, such as thin and thick disk clusters, but did not find anything of significance.

We examined whether cluster scatter was correlated with physical cluster size, which we calculated using the angular cluster radius from K13 and the median stellar distance, and found no relationship. However, we note that we consider these size values to be highly uncertain, as they depend on the choice of angular radius definition and in at least some cases, clearly do not match the kinematically clumped stars at that location.

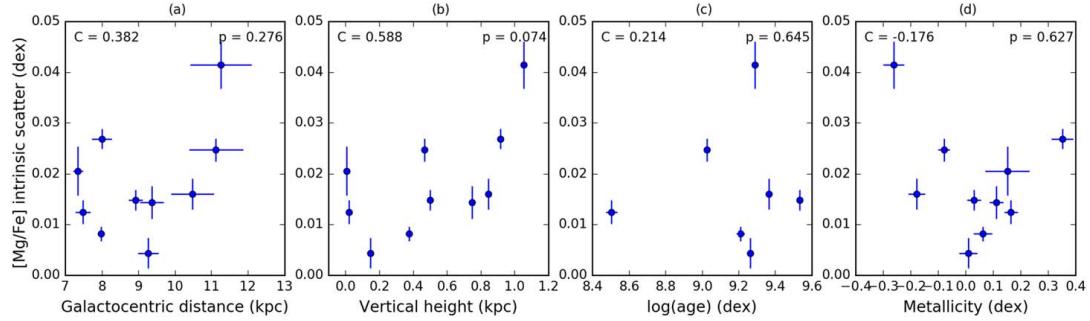


Figure 8. Dependence of cluster [Mg/Fe] homogeneity on Galactocentric distance, vertical height, log (age), and metallicity of the cluster. The Spearman correlation coefficient (C) and corresponding p value (p) are shown for each property.

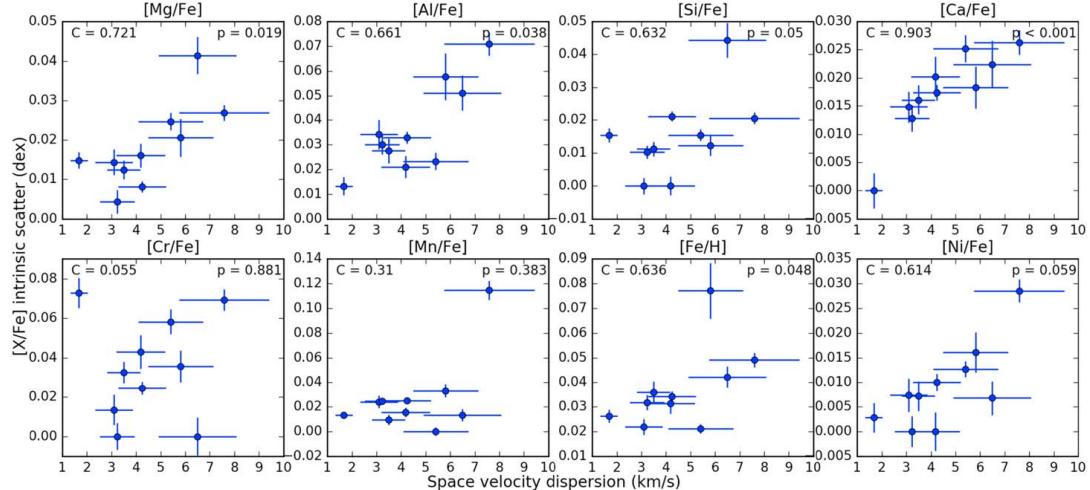


Figure 9. Dependence of the cluster [X/Fe] homogeneity on space velocity dispersion. The Spearman correlation coefficients (C) and corresponding p values (p) are shown for each element.

5.5. Velocity Dispersion (Cluster Mass)

5.5.1. Correlation with Velocity Dispersion

We calculate the 3D velocity dispersion of a cluster, a proxy for cluster mass, from its RV and PM dispersions and heliocentric distance using the following equation:

$$\sigma_{\text{tot}} = \sqrt{\sigma_{RV}^2 + (\sigma_{\mu_a}^2 + \sigma_{\mu_b}^2)d_{\text{helio}}^2}, \quad (4)$$

where σ_{tot} is the space velocity dispersion, corrected by the uncertainties as described in Section 4; σ_{RV} , σ_{μ_a} , and σ_{μ_b} are the dispersions in the cluster for each kinematic dimension; and d_{helio} is the heliocentric distance, assigned as the median of the stellar member distances from RA17 (Section 2.4). We observe a strong correlation between the calculated RV and PM velocity dispersions, which ensures that σ_{tot} is not being driven by any one dimension alone.

We find that the cluster chemical scatter is positively correlated with the space velocity dispersion of the cluster at relatively high levels of significance ($p < 0.019$) for [Mg/Fe]

and [Ca/Fe]; possibly positively correlated at a low level of significance ($0.038 < p < 0.059$) for [Ni/Fe], [Si/Fe], [Al/Fe], and [Fe/H]; and uncorrelated ($p > 0.38$) for [Cr/Fe] and [Mn/Fe]. Figure 9 shows the intrinsic scatter in [Fe/H] and the rest of the abundances as a function of space velocity dispersion, along with the associated Spearman correlation coefficients (C) and p values.

To understand why only certain elements show this trend between intrinsic scatter and σ_{tot} , we look for natural ways to group elements based on their properties. For example, we notice that this trend is not exclusive to the α elements that we study. Although intrinsic scatter is positively correlated with σ_{tot} in Mg, Ca, and Si (albeit at low significance), we observe a similar trend in an odd-Z element (Al) and an iron-peak element (Ni) at lower significance. In Section 5.3 we discuss how the abundance scatters we observe behave for elements that show abundance variations and anticorrelations in GCs (e.g., Mg, Al, and Si). Again, this trend is not restricted to these three elements. However, we find that if we group elements based on their dominant nucleosynthetic process of production, we see a

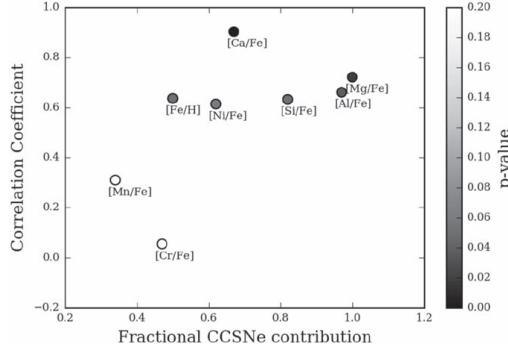


Figure 10. Fractional contribution from CCSNe vs. the Spearman correlation coefficient of space velocity dispersion with respect to the intrinsic $[X/\text{Fe}]$ scatter.

distinction between elements that are produced predominantly by core-collapse supernovae (CCSNe) versus Type Ia SNe.

In order to visualize the differences in the strengths of correlation in groups of elements and to explore the roles that different enrichment events may have played in them, we use the empirically determined fractional contribution of CCSNe (f_{cc}) for each element from Weinberg et al. (2019). Figure 10 shows the Spearman correlation coefficients from Figure 9 against f_{cc} , with the points colored by the p value of their correlation. Here, f_{cc} represents the fraction of each element contributed by CCSNe at a given metallicity $[\text{Mg}/\text{H}]$, assuming that these elements are produced exclusively by Type Ia SNe and CCSNe. We calculate f_{cc} from Equation (11) in Weinberg et al. (2019), using the median $[\text{Mg}/\text{H}]$ value for each of our clusters.

From Figure 10, in addition to metallicity (traced by $[\text{Fe}/\text{H}]$), the elements that show a correlation between intrinsic abundance scatter and cluster velocity dispersion (with $C > 0.6$) at both higher (with $p < 0.019$) and lower ($0.038 < p < 0.059$) levels of significance are the ones that are produced mostly by CCSNe.

5.5.2. Caveats of the Significance of the Correlations

While these correlations between abundance scatter and space velocity dispersion are interesting, we note that the statistical analysis is done using only 10 clusters and the elements that we list as correlated have varying levels of significance (p values). Here we explore the caveats associated with these correlations. One effect of the low sample size is seen when we randomly remove any one cluster from the analysis using a jackknife resampling. For certain elements, the correlation becomes insignificant if we remove a specific cluster (e.g., removing NGC 2204 in the case of Si or removing NGC 6791 in the case of Ni) from the analysis. However, we found that no one cluster is responsible for systematically reducing the significance across the set of all the elements. Because we arrive at this cluster sample by preferentially selecting clusters with high quality membership and reliable abundance uncertainties (Section 5.1), we have no *a priori* reason to drop any particular cluster from any particular elemental trend.

We explored the effect of using a different correlation metric (Kendall's tau) on the strength and significance of the correlations seen in Figure 9. We do not use the Pearson metric in this comparison because it is susceptible to outliers and assumes

linear relationships. The Spearman and Kendall correlation metrics are highly correlated themselves, though their magnitude is not equal, and they are more robust to outliers as they use the ranks of the variables rather than the actual values. We show the Spearman metric in Figures 9 and 10 because it is more commonly used, although we note that Kendall may be more accurate for small sample sizes. The significance of the correlations for Al, Si, Fe, and Ni depends somewhat on the choice of the metric, with Kendall's tau associated with p values up to 0.07, which reinforces our classification of these elements as possible correlations. However, Mg and Ca scatters are significantly correlated with velocity dispersion irrespective of the metric used, while Mn and Cr remain uncorrelated.

5.5.3. Potential Implications

As discussed in Section 5.5.1, it is interesting that the elements that show some level of correlation with space velocity dispersion (Figure 9) are those that are predominantly produced by CCSNe. However, due to the small sample size, both in clusters and elements, and other caveats discussed in Section 5.5.2, we cannot definitively conclude that these correlations are significant. However, in case these findings are validated by future larger studies, here we explore the potential implications of this result.

Figure 10 suggests that the nucleosynthetic processes that are responsible for the production of elements in the ISM may have an observable effect on the final abundance scatter within the cluster. It also hints at the existence of a difference in the ejection radii between the pollution mechanisms of Type Ia SNe and CCSNe.

Why does the chemical scatter within a cluster depend on the mass of the cluster? Equally interesting, why is this correlation present in certain elements and not in others? The dependence of the intrinsic abundance scatter on cluster mass can be understood by looking at what we know about cluster formation processes. Clusters are formed from giant molecular clouds in filamentary structures (Kounkel & Covey 2019), forming strings of star-forming gas. So, gas accreted to form more-massive clusters will not be limited to a sphere surrounding the final cluster, but rather spans a larger range in distance. Fujii (2015) explored the possibility that young massive clusters may be formed by hierarchical merging of subclusters or smaller open clusters.

Although the OCs we are using in our analysis are not as massive as the young massive clusters discussed in Fujii (2015), this mechanism hints that more-massive clusters that we study may have been formed by accreting gas over larger ranges in distance in the initial cloud. This would result in the more-massive clusters having a larger scatter in metallicity and abundances of certain elements depending on their correlation lengths and mixing efficiency in the initial clouds before star formation began. Krumholz et al. (2019) also suggest that massive clusters are formed over extended formation times rather than a single freefall timescale. This could again increase the chances of the star-forming cloud being polluted by exploding high-mass stars that have already formed.

How are variations in the correlation lengths of different metals in an initial cloud reflected in the chemical homogeneity of the final clusters formed from the cloud? Armillotta et al. (2018) studied this using hydrodynamical simulations, and they observed that if the initial field of the metal is correlated over smaller distances ($< 6 \text{ pc}$), turbulent mixing will efficiently

smooth out these inhomogeneities. Metal fields that are correlated over much larger distances (>40 pc) will also yield homogeneous stellar chemistry as the variations span a range larger than the typical cloud size. However, metals that are correlated on intermediate lengths in the ISM (6–40 pc) can have a higher scatter in the stellar abundances of their final clusters. So, with this reasoning, elements that have correlation lengths within the intermediate range in the initial cloud may be expected to have a higher abundance scatter in the final stellar members for massive clusters.

Furthermore, because elements belonging to different nucleosynthetic groups have been shown by Krumholz & Ting (2018) to have different correlation lengths in the initial cloud, we may observe this trend only in certain elements or nucleosynthetic groups. They propose that there should be no significant differences between the correlation lengths of Type Ia SNe and CCSNe since both types of explosions have comparable energy budgets. However, Figure 10 suggests the presence of a quantitative difference between the correlation lengths of these two mechanisms that pollute the ISM, and that this difference may manifest itself in the abundance scatter of nucleosynthetic element groups in the final stellar populations.

6. Summary

We have identified cluster members for a large number of open clusters in the Kharchenko et al. (2013) catalog using only their kinematic information: radial velocities from APOGEE and PMs from Gaia-DR2. We provide a catalog of cluster properties and members for 83 clusters with a range of detection qualities (Section 3.2). This cluster membership catalog will be useful for anyone interested in studying cluster chemistry. In addition, we derived new uncertainties for the APOGEE elemental abundances, as a function of stellar parameter and S/N, for the cluster members.

We also studied the dependence of cluster chemical homogeneity on various Galactic and cluster properties. As seen from Figure 9, Mg and Ca show a strong, relatively significant correlation between cluster chemical scatter and velocity dispersion, while Ni, Si, Al, and Fe may also exhibit a possible positive correlation, albeit at low significance. It is interesting that these elements are those that are predominantly produced by CCSNe. However, we urge caution in these findings due to the small sample size and p values close to 0.05. Nevertheless, if true, these findings suggest a quantitative difference between the correlation lengths of the metals dispersed into the ISM as a result of Type Ia SNe and CCSNe, under the assumption that scatter is set by mixing processes. The existence of an intrinsic difference in the distance to which the elements are expelled by these two SN explosions would affect our understanding of the pollution rates and mixing efficiency in the ISM. For a definite determination, not only is the exploration of more elements required, but also a larger sample of open clusters. If validated by future larger studies, this result should be included in existing and future Galactic chemical evolution models and simulations.

These results also have potential implications for chemical tagging, which first assumes that OCs are intrinsically chemically homogeneous and then attempts to determine birth siblings, cluster members, or conatal objects using only the chemical signatures of the stars. We find that the abundance scatter in most elements for our clusters are within the limits previously found (e.g., Bovy 2016). However, if future studies with a larger OC

sample and more elements find similar empirical dependencies of the cluster homogeneity on velocity dispersion, these results should be considered in future work using chemical tagging. For example, the most-massive OCs could either be altogether avoided in chemical tagging studies or be studied with caution for elements that are predominantly produced by CCSNe.

We thank the anonymous referee for thoughtful comments that improved the clarity of the paper. V.P., G.Z., K.H., and K.M.K. are grateful for support from the Research Corporation for Science Advancement through a ScialogR award. S.H. is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-1801940. J.D. and P.M.F. acknowledge support for this research from the National Science Foundation (AST-1311835 & AST-1715662). D.A.G. H. acknowledges support from the State Research Agency (AEI) of the Spanish Ministry of Science, Innovation and Universities (MCIU) and the European Regional Development Fund (FEDER) under grant AYA2017-88254-P. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the US Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Software: astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), astroquery (Ginsburg et al. 2019), FERRE (Allende Prieto et al. 2006), functools (Hendricks 2015), matplotlib (Hunter 2007), numpy (Oliphant 2006; Van Der Walt et al. 2011), pandas (Pandas Development Team 2020), scikit-learn (Pedregosa et al. 2011), scipy (Virtanen et al. 2020), Turbospectrum (Plez 2012), vorbin (Cappellari & Copin 2003).

Appendix A

In Figures A1 and A2 we show the uncertainty training sample plots for all the elements we use in our analysis in Section 5.

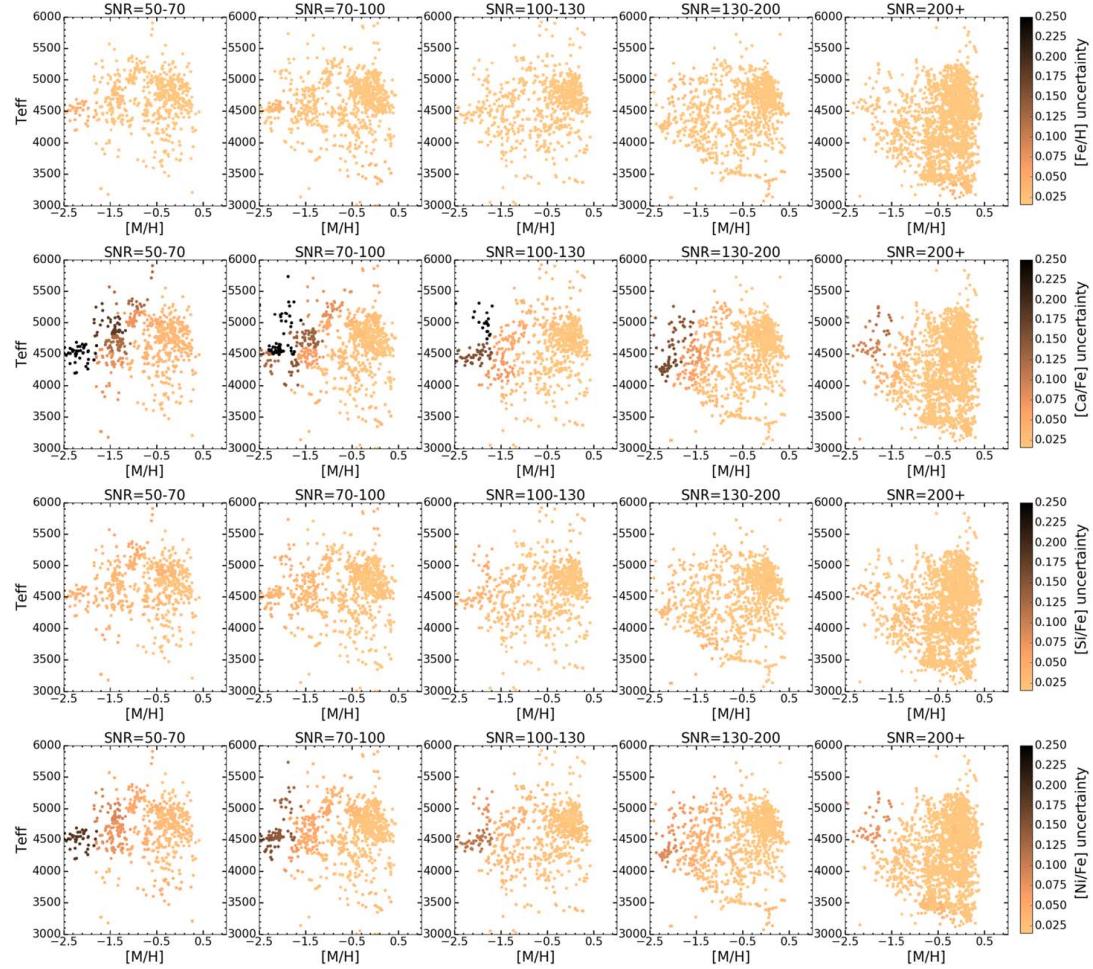


Figure A1. The Uncertainty Training (UT) sample for $[Fe/H]$, $[Ca/Fe]$, $[Si/Fe]$, and $[Ni/Fe]$, similar to the first row of Figure 1.

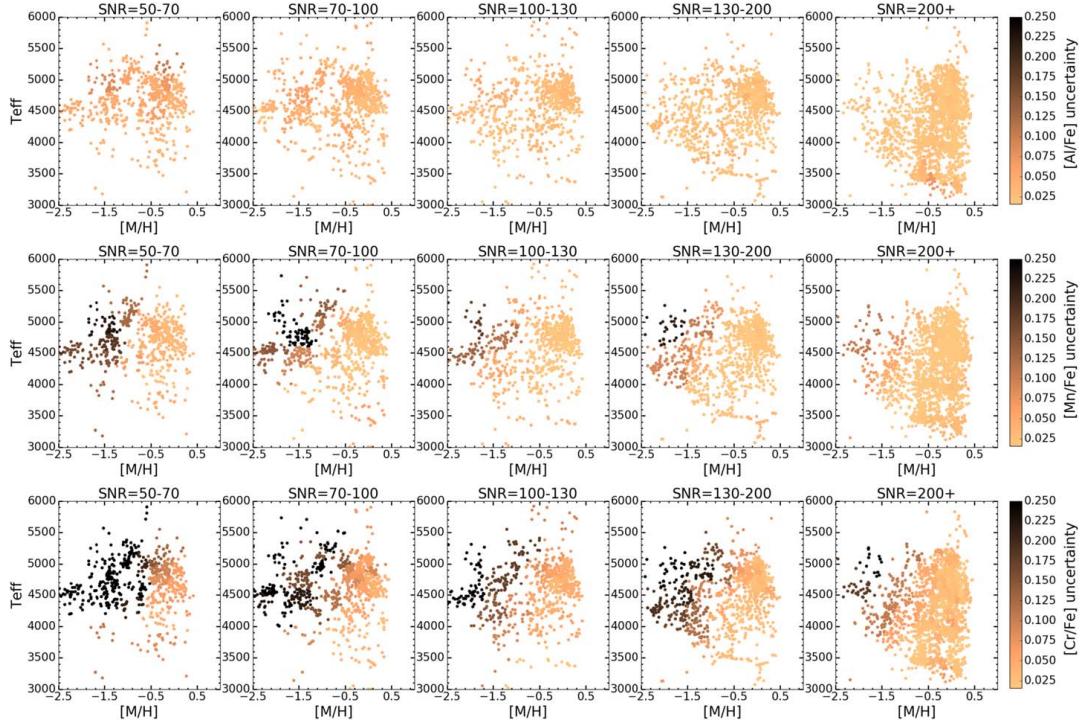


Figure A2. The Uncertainty Training (UT) sample for [Al/Fe], [Mn/Fe], and [Cr/Fe], similar to the first row of Figure 1.

Appendix B

Here we show the membership plots for all of the clusters that we use in the final analysis in Section 5. Figure B.1 is shown here as an example.

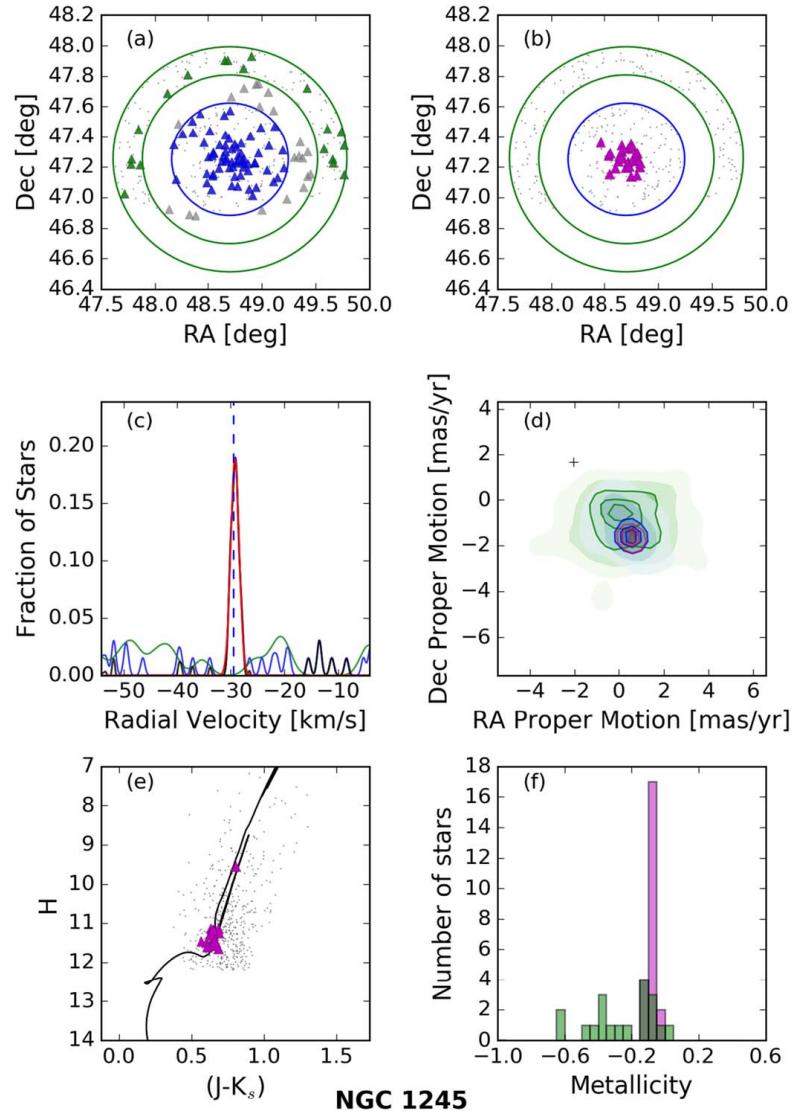


Figure B1. Membership plots for all of the clusters that we use in the final analysis in Section 5. The panels are the same as Figure 3. (The complete figure set (10 images) is available.)

ORCID iDs

- Vijith Jacob Poovelil <https://orcid.org/0000-0002-9831-3501>
 G. Zasowski <https://orcid.org/0000-0001-6761-9359>
 S. Hasselquist <https://orcid.org/0000-0001-5388-0994>
 A. Seth <https://orcid.org/0000-0003-0248-5470>
 K. Cunha <https://orcid.org/0000-0001-6476-0576>
 Peter M. Frinchaboy <https://orcid.org/0000-0002-0740-8346>
 D. A. García-Hernández <https://orcid.org/0000-0002-1693-2721>
 K. Hawkins <https://orcid.org/0000-0002-1423-2174>
 K. M. Kratter <https://orcid.org/0000-0001-5253-1338>
 C. Nitschelm <https://orcid.org/0000-0003-4752-4365>

References

- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, *ApJS*, 249, 3
 Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, *ApJ*, 636, 804
 Armillotta, L., Krumholz, M. R., & Fujimoto, Y. 2018, *MNRAS*, 481, 5000
 AstroPy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
 Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, *ApJ*, 854, 147
 Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2005, *ApJL*, 630, L149
 Bertran de Lis, S., Allende Prieto, C., Majewski, S. R., et al. 2016, *A&A*, 590, A74
 Blanco-Cuaresma, S., Soubiran, C., Heiter, U., et al. 2015, *A&A*, 577, A47
 Bland-Hawthorn, J., Krumholz, M. R., & Freeman, K. 2010, *ApJ*, 713, 166
 Blanton, M. R., Bershadsky, M. A., Abolfathi, B., et al. 2017, *AJ*, 154, 28
 Bovy, J. 2016, *ApJ*, 817, 49
 Bowen, I. S., & Vaughan, A. H., Jr. 1973, *ApOpt*, 12, 1430
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
 Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *A&A*, 618, A93
 Cappellari, M., & Copin, Y. 2003, *MNRAS*, 342, 345
 Carlberg, J. K., Cunha, K., & Smith, V. V. 2016, *ApJ*, 827, 129
 Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2010, *A&A*, 516, A55
 Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, 121, 213
 Cunha, K., Smith, V. V., Johnson, J. A., et al. 2015, *ApJL*, 798, L41
 Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, 1
 De Silva, G. M., Freeman, K. C., Asplund, M., et al. 2007, *AJ*, 133, 1161
 De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2011, *MNRAS*, 415, 563
 De Silva, G. M., Sneden, C., Paulson, D. B., et al. 2006, *AJ*, 131, 455
 Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2018, *AJ*, 156, 142
 Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2020, *AJ*, 159, 199
 Freeman, K., & Bland-Hawthorn, J. 2002, *ARA&A*, 40, 487
 Frinchaboy, P. M., & Majewski, S. R. 2008, *AJ*, 136, 118
 Froebrich, D., Scholz, A., & Raftery, C. L. 2007, *MNRAS*, 374, 399
 Fujii, M. S. 2015, *PASJ*, 67, 59
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, 616, A1
 García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, 151, 144
 García-Díaz, R., Allende Prieto, C., Sánchez Almeida, J., & Alonso Palicio, P. 2019, *A&A*, 629, A34
 Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, *AJ*, 157, 98
 Gratton, R. 2020, in IAU Symp. 351, Star Clusters: From the Milky Way to the Early Universe, ed. A. Bragaglia et al. (Cambridge: Cambridge Univ. Press), 241
 Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, *AJ*, 131, 2332
 Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, *A&A*, 486, 951
 Hawkins, K., Masseron, T., Jofré, P., et al. 2016, *A&A*, 594, A43
 Hendricks, P. 2015, Funtools: Functional Programming in R., <https://CRAN.R-project.org/package=funtools>
 Hole, K. T., Geller, A. M., Mathieu, R. D., et al. 2009, *AJ*, 138, 159
 Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, *AJ*, 156, 125
 Hunter, J. D. 2007, *CSE*, 9, 90
 Jacobson, H. R., Friel, E. D., & Pilachowski, C. A. 2011, *AJ*, 141, 58
 Jacobson, H. R., Pilachowski, C. A., & Friel, E. D. 2011, *AJ*, 142, 59
 Jönsson, H., Allende Prieto, C., Holtzman, J. A., et al. 2018, *AJ*, 156, 126
 Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, *AJ*, 160, 120
 Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R.-D. 2005, *A&A*, 438, 1163
 Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, *A&A*, 558, A53
 Kounkel, M., & Covey, K. 2019, *AJ*, 158, 122
 Kovalev, M., Bergemann, M., Ting, Y.-S., & Rix, H.-W. 2019, *A&A*, 628, A54
 Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, 57, 227
 Krumholz, M. R., & Ting, Y.-S. 2018, *MNRAS*, 475, 2236
 Lee-Brown, D. B., Anthony-Twarog, B. J., Deliyannis, C. P., Rich, E., & Twarog, B. A. 2015, *AJ*, 149, 121
 Leung, H. W., & Bovy, J. 2019, *MNRAS*, 489, 2079
 Liu, F., Asplund, M., Yong, D., et al. 2016a, *MNRAS*, 463, 696
 Liu, F., Yong, D., Asplund, M., Ramírez, I., & Meléndez, J. 2016b, *MNRAS*, 457, 3934
 Looney, L. W., Tobin, J. J., & Fields, B. D. 2006, *ApJ*, 652, 1755
 Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
 Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, 739, 25
 Marigo, P., Girardi, L., Bressan, A., et al. 2017, *ApJ*, 835, 77
 Mézáros, S., Holtzman, J., García Pérez, A. E., et al. 2013, *AJ*, 146, 133
 Mézáros, S., Masseron, T., García-Hernández, D. A., et al. 2020, *MNRAS*, 492, 1641
 Milone, A. P., Marino, A. F., Renzini, A., et al. 2018, *MNRAS*, 481, 5098
 Moe, M., Kratter, K. M., & Badenes, C. 2019, *ApJ*, 875, 61
 Ness, M., Rix, H. W., Hogg, D. W., et al. 2018, *ApJ*, 853, 198
 Nidever, D. L., Holtzman, J. A., Allende Prieto, C., et al. 2015, *AJ*, 150, 173
 Oliphant, T. E. 2006, A Guide to NumPy, Vol. 1 (USA: Trelgol Publishing)
 Pandas Development Team 2020, andas-dev/pandas: Pandas, latest, T., Zenodo, doi:10.5281/zenodo.3509134
 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825
 Platais, I., Gosnell, N. M., Meibom, S., et al. 2013, *AJ*, 146, 43
 Plez, B. 2012, Turbospectrum: Code for Spectral Synthesis, Astrophysics Source Code Library, ascl:1205.004
 Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2020, *ApJ*, 895, 2
 Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, 156, 123
 Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, *A&A*, 638, 76
 Queiroz, A. B. A., Anders, F., Santiago, B. X., et al. 2018, *MNRAS*, 476, 2556
 Rojas-Arrigada, A., Recio-Blanco, A., de Laverny, P., et al. 2017, *A&A*, 601, A140
 Saracino, S., Bastian, N., Kozhurina-Platais, V., et al. 2019, *MNRAS*, 489, L97
 Semenov, E., Bergemann, M., Deal, M., et al. 2020, arXiv:2007.09153
 Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
 Souto, D., Allende Prieto, C., Cunha, K., et al. 2019, *ApJ*, 874, 97
 Spina, L., Meléndez, J., Casey, A. R., Karakas, A. I., & Tucci-Maia, M. 2018, *ApJ*, 863, 179
 Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, 13, 22
 Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, 17, 261
 Weinberg, D. H., Holtzman, J. A., Hasselquist, S., et al. 2019, *ApJ*, 874, 102
 Wilson, J. C., Hearty, F. R., Skrutskie, M. F., et al. 2019, *PASP*, 131, 055001
 Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
 Wu, T., Li, Y., & Hekker, S. 2014, *ApJ*, 786, 10
 Yang, S.-C., Sarajedini, A., Deliyannis, C. P., et al. 2013, *ApJ*, 762, 3
 Zasowski, G., Beaton, R. L., Hamm, K. K., et al. 2013a, *AJ*, 146, 64
 Zasowski, G., Cohen, R. E., Chojnowski, S. D., et al. 2017, *AJ*, 154, 198
 Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013b, *AJ*, 146, 81
 Ziegler, C., Tokovinin, A., Briceño, C., et al. 2020, *AJ*, 159, 19

CHAPTER 3

ASSESSING THE INFORMATION CONTENT OF CLUSTER ENSEMBLE LIGHTCURVES

3.1 Introduction

The presence of variable stars in clusters can be valuable in determining certain properties of the clusters. The ensemble light from a cluster will change as the population ages since the types of variable stars that dominate the variability of the ensemble light will also change. For instance, Cepheid variables are found primarily in populations younger than 100 Myr, typically OCs, while there is a significantly higher fraction of RR Lyrae stars in older GCs (Clement et al. 2001, §1.3). There are significant differences in the light from these two types of variable stars, including period, amplitude, and the shape of the lightcurve (LC). The differences between the fundamental types of variable stars that populate different clusters is a vital piece of information that we can utilise to determine cluster ages. With the help of ensemble photometry, we can quantify these differences for a plethora of clusters of different ages and metallicities, predominantly in the Milky Way (MW, Kharchenko et al. 2013), and the Small and Large Magellanic clouds (SMC and LMC, respectively, Bica et al. 2020).

This chapter describes the first steps of a project to develop statistical techniques to estimate cluster ages using ensemble photometry. The variability signals from clusters can be captured and quantified using different metrics of the time-series ensemble photometric data of these clusters. By characterizing variability for clusters spanning a range of age and metallicity, we seek to estimate cluster ages using just their ensemble LCs. This technique

will be particularly useful to estimate ages of unresolved stellar populations and clusters in galaxies for which we have photometric data.

In this chapter, I describe a pilot study to understand where the age information resides in a cluster’s ensemble lightcurve. I explore how to optimally process the ensemble LCs, cleaning and extracting the actual signal from the cluster members, after accounting for the contamination, background, and systematics (§3.2.4). In §3.3.1 and §3.3.2, I explore the intrinsic variability of the possible contaminants in the ensemble LC, viz., field stars and unresolved background stars in the cluster. In the same section, I study the effect of systematics on the ensemble LC by using the Lomb-Scargle (LS) periodogram to identify the features corresponding to actual variable stars and those that may be caused by remnant systematics (§3.3.3.1). This is complemented with a detailed pixel-by-pixel analysis for clusters that do not have member variability information (§3.3.3.2).

3.2 Data and Ensemble Photometry Techniques

3.2.1 Cluster Catalogs

For this project, I use several variability and membership catalogs for both open clusters (OCs) and globular clusters (GCs). The Clement et al. (2001) catalog is a compilation of all known variability sources in GCs, along with information about the type of variable star and its period of variability. There is no equivalent global catalog that documents all known variables stars in OCs, and hence I describe the individual cluster variability studies in the relevant cluster sections.

I adopt the Milky Way Star Clusters catalog (Kharchenko et al. 2013, hereafter K13) as the base catalog for cluster information such as cluster center coordinates and angular radii to define TESS and *Gaia* search limits. In addition to individual studies that measure distances, ages, and metallicities of clusters (e.g., Donor et al. 2018; Bergbusch et al. 1991; Alonso-Santiago et al. 2017; Lennon et al. 1990), the catalogued values from K13 are also used whenever available. The Bica et al. (2020) catalog contains age and metallicity information of clusters in the Magellanic Clouds.

Figure 3.1(a) shows the distribution of the clusters on the age-metallicity plane for which we have estimates of age and metallicity from literature. The other subplots in Figure 3.1

show clusters that also have (b) accessible TESS data, (c) information from literature about the variability sources present in them, (d) both TESS data and variability information. A similar age-metallicity distribution for the clusters in the Magellanic Clouds is shown in Figure 3.2.

3.2.2 Time Domain Photometry

Ensemble photometry refers to the combined lightcurve for a group of objects, e.g., a stellar cluster. This technique is essential for analyzing signals from unresolved clusters and galaxies, where accessing LCs for each individual star is not possible. It is important to understand what information is contained in the ensemble LC of a cluster and how we can identify the important bits of information, while discarding the noise. The light from an ensemble LC contains the following components:

- Cluster member flux (signal): The ensemble light acquired from the cluster will contain information from both variable and non-variable cluster members within the aperture selected. This signal is what we are primarily interested in for our analysis.
- Field star flux (contamination): In addition to cluster members, there will also unavoidably be resolved field stars within the vicinity of the cluster that will contribute to the ensemble light. While this may contaminate the ensemble light from the cluster, it could also lead to discoveries of previously unknown variable stars within the field of the cluster.
- Unresolved star flux (background): While resolved stars will contribute the largest flux to the ensemble light, the faint background surrounding the cluster core and resolved field stars contains non-zero flux from unresolved stars. The variations in the flux from these background stars may depend on the location of the cluster field. It is important to understand the photometric variations of the unresolved stars that make up the background flux from a cluster field.
- Instrument noise (systematics): Each telescope and its associated instrumentation will produce systematic noise that depends on several factors like the target location

and observation time. There are several methods to account for this noise and clean the signal to study the components of the light that we are actually interested in. One such method for removing systematic noise from TESS LCs is described in §3.2.4.

3.2.3 TESS Lightcurves

Several surveys and missions look at time domain photometry for stars and planets, e.g., ULTRACAM (Dhillon et al. 2007), WASP (Pollacco et al. 2006), NGTS (Wheatley et al. 2018), HiPERCAM (Dhillon et al. 2016), (Udalski et al. 1992), and *Kepler* (Koch et al. 2010). In addition to these, the Transiting Exoplanet Survey Satellite (TESS) has revolutionized exoplanet search and has discovered multiple planets of different sizes around nearby stars (Huang et al. 2018; Dragomir et al. 2019). TESS plans to catalog thousands of planet candidates and considerably increase the current number of known exoplanets. However, despite its mission focus, TESS data are not only useful for detecting exoplanets, but also to study stellar variability. The considerably long baseline of one month to one year of observations for each star, depending mainly on its ecliptic latitude (Ricker et al. 2014), enables us to track a large range of variability source types, such as RR Lyrae, Cepheids, and eclipsing binaries.

Ricker et al. (2015) explain the specifications of the TESS mission, including details about the spacecraft and the instrumentation. The spacecraft is currently in a highly elliptical 13.7-day orbit around the Earth. The instrumentation consists of four wide-field optical charge-coupled device cameras. Flux measurements of target stars are recorded every 2 min, and full frame images are recorded every 30 min. Each sector is observed continuously for two spacecraft orbits (27.4 days), with occasional momentum dumps needed due to the ≈ 1.5 Nm of angular momentum build-up induced by solar radiation pressure. At the perigee, observations are interrupted for around 16 h to point TESS’s antenna toward Earth, downlink data, and resume observing. This process produces systematics in the observed flux towards the extremes of each transit that need to be corrected using techniques like the cotrending basis vectors provided by TESS for each CCD or linear regression techniques using flux pixels outside the aperture (Lightkurve Collaboration et al. 2018), described in §3.2.4.

The most important systematic error is due to random pointing variations (“spacecraft jitter”) where the motion of the star image on the CCD will introduce changes in the measured brightness due to the nonuniform quantum efficiency of the CCD pixels (Ricker et al. 2015). Another common source of systematic errors is the saturation of pixels at magnitudes of approximately $I_C = 7.5$. However, TESS is still capable of high precision photometry with a bright limit of $I_C \approx 4$ or maybe even brighter, if the photometric aperture is large enough to contain all of the excess charge. Public data releases from TESS occur every 4 months, supporting and encouraging community-wide efforts to study new planets and variable stars¹.

3.2.4 Data Processing and Detrending

The major part of the analyses described in this chapter involves the use of TESS data. Here I describe the steps followed to locate, download, process, and clean ensemble LCs of clusters. This process is made easier by the Lightkurve package (Lightkurve Collaboration et al. 2018) in Python, which has functions and routines for each of these steps. Lightkurve searches for TESS data products using the name of the cluster and downloads a cutout of the full frame image. Next, I define the cluster aperture as a circle with the radius in pixels of the cluster from K13; each pixel in TESS covers 21 arcsecs in the sky. The light from the cluster aperture will be the primary raw signal of the cluster. The ‘background’ aperture is defined as the pixels that lie outside the cluster aperture and within the TESS cutout and have flux less than a designated threshold. The light from the background aperture is used to correct for systematics in the light from the cluster aperture.

There are several methods to process the cluster ensemble lightcurves, cleaning and extracting the intrinsic signal from the cluster members. This process, called detrending, involves accounting for systematics arising from spacecraft motion, pointing variations, and data downlink. For detrending ensemble LCs, I utilise a built in functionality in Lightkurve that uses background pixels within the field of the cluster to remove the systematics in the cluster ensemble LC. This technique uses linear regression to subtract the features

¹Link to data releases: <https://archive.stsci.edu/missions-and-data/tess/data-products>

that are common to the cluster and background apertures. However, this tool contains several parameters, such as the number of PCA vectors used for the linear regression, the background pixel flux limit, and aperture radius, that can be tuned to obtain the optimal corrected LC. To figure out a combination of parameters that adequately correct the ensemble LCs of different clusters, I perform a series of analyses using ensemble light from multiple clusters (§3.2.4.1, §3.2.4.2) and mock data generated using known variable stars present in clusters (§3.2.4.3).

3.2.4.1 Background flux threshold

One of the parameters that can be tuned in the detrending process is the flux threshold above which pixels will be excluded from the background used for the systematics correction. These high flux pixels should be avoided since they most likely correspond to resolved field stars. The background pixels should ideally represent only the systematics due to the spacecraft motion and pointing variations and should contain minimal contamination from field stars.

An ideal value for the flux threshold is determined by visualizing the distribution of background pixels for different values of background threshold and a histogram of the flux values for the pixels within the TESS cutout. Figure 3.3 show the pixels that are selected as the background aperture for different values of the threshold for NGC 6637. The histograms of the flux from the TESS cutout for these clusters are shown in Figure 3.4, with the background thresholds indicated with dashed lines. From this figure, one can see that the 85th percentile captures the bulk of the background flux and avoids the tail end of the flux distribution. On increasing the percentile, the background aperture captures more of the light from the pixels on the periphery of the resolved field stars seen in Figure 3.3. The 85th percentile flux threshold was observed to consistently avoid including flux from these resolved bright field stars for multiple clusters, hence improving our background aperture selection.

3.2.4.2 Background spatial variations

Next I study the spatial variations of the background pixels in the detrending process. For this I divide the background into four quadrants and perform the detrending using pixels within each quadrant. These quadrants are shown in the left plot of Figure 3.5 for NGC 6637. I download and process LCs from the pixels in each of these four background quadrants. The right plot in Figure 3.5 shows the ratio of the background LC from the third quadrant with LCs from each of the four quadrants. It is evident from this figure that there are some differences between the features in the LC of background pixels taken from different parts of the TESS cutout.

These differences could be a result of significantly reducing the number of pixels included in the background aperture as we use individual quadrants in the TESS cutout. If this is the case, there may be differences in the background spatial variations between clusters towards the Galactic center and the halo. Since clusters towards the Galactic center are expected to have a higher density of field stars, they would have fewer pixels included in the background aperture, potentially increasing the spatial variations in the background. This analysis suggests that one should be aware of the spatial variations in the background pixels of a cluster and that it is important to characterize these variations with respect to the location of the cluster.

3.2.4.3 Number of PCA vectors

The goal behind this analysis is to identify an optimal range for the number of PCA vectors to be used in the detrending process. Too few would not sufficiently account for the complex systematics that are signature to typical TESS LCs; too many could remove interesting features in the LC that may belong to a variable star in the cluster.

To achieve this goal, I generate mock LCs for clusters by downloading and combining individual LCs for all known variable stars in a cluster and adding background noise from background pixels within the field of the cluster. I then use the detrending method described in §3.2.4 on the mock lightcurves to test if the additional noise is removed and the original mock LC is restored. I use different values for the number of PCA vectors, and compare the resulting detrended LCs. Figure 3.6 shows the original mock LC, the

LC with the additional background noise, and the LCs detrended using different number of PCA vectors, along with the corresponding periodograms for each case. It is clear from the figure that a higher number of PCA vectors is not necessary for the detrending process. The detrended LCs using fewer PCA vectors look very similar to the original mock LC, sufficiently accounting for and removing the background noise. One can also notice how the detrended LC becomes flatter with higher number of PCA vectors. This is also evident in the periodograms, with those for the higher number of vectors seeming to remove certain features from the original periodogram. Excessively correcting the LC could potentially remove some of the intrinsic signal coming from variable stars within the cluster. This analysis suggests that restricting the number of PCA vectors used in the detrending process to fewer than ten is safer.

I also verify that this method is robust to different backgrounds by repeating the analysis above with background noise from the field of a different cluster. The detrending method is found to effectively subtract the background, and the final corrected LC is found to be very similar to the original mock LC. In summary, the detrending tool will efficiently correct for the systematics in the ensemble LC with a lower number of PCA vectors (fewer than ten) and is robust to different types of backgrounds.

3.3 Exploring Cluster Variability Using Ensemble Photometry

In this section I describe several analyses of cluster variability using ensemble TESS photometry and discuss their results. The techniques used to download, process, and clean ensemble LCs from TESS data for individual clusters are described in §3.2.4. To quantify the strength of periodic features in ensemble LCs, I use the Lomb-Scargle periodogram, which calculates the power at a range of frequencies using least-square sinusoidal fits to the LC. I have also explored other techniques to quantify the variability in ensemble LCs such as the damped random walk model, the root mean square of the flux, and the Welch-Stetson statistic (Welch and Stetson 1993). The efficiency and reliability of these metrics with respect to their reflection of age information from clusters have to be further explored and understood.

3.3.1 Field Star Variability

Here I explore the photometric variability of field stars located outside the aperture of the cluster (§3.2.4) but within the field of the cluster cutout from TESS. It is important to understand the variability of field stars that are not members of the cluster since an aperture of any radius that we define will unavoidably contain a certain fraction of field stars. Since we plan to use this ensemble photometry method on unresolved clusters, we have to characterize field star variability using resolved clusters in our Galaxy.

Here I consider NGC 6637, a GC in the Galactic bulge, and NGC 3201, a GC in the outer disk. I selected these two clusters due to their distinct field star populations and concentrations expected from the densely populated Galactic bulge compared to the outer disk of the Galaxy. Here I only consider objects outside the cluster radius specified in K13, with flux greater than a specified threshold. These conditions isolate the high flux objects outside the cluster range, which are most likely the field stars. Figure 3.7 shows the ensemble field star LC for NGC 6637 from the selected aperture and the corresponding LS periodogram. From the ensemble LC it is clear that there is a periodic signal at around 3 days, and this is confirmed by the periodogram as well. In case of NGC 3201, the ensemble LC in Figure 3.8 shows a range of variability in different periods and the periodogram also

highlights this.

This analysis shows that field stars within the range of a cluster may contain some variables that could contaminate the ensemble signal that we are hoping to obtain from the cluster. One way to minimize the contamination from field stars is to consider only smaller radii when using ensemble photometry. This however would have to be balanced with the desired completeness for the analysis since a larger fraction of genuine cluster members would also be excluded for smaller radii. A better approach is to statistically characterize field star variability throughout the Galaxy as a function of their ages and metallicities. This understanding can help us account for the variability of the background stellar population of a cluster based on its location in the Galaxy.

3.3.2 Background Variability

Here I explore the photometric variability of the background pixels within the cluster radius. These pixels will contain faint light from unresolved stars that make up the background of the cluster field. It is important to understand the behaviour of these background unresolved stars since the light from them is a component that contributes to the ensemble light from a cluster aperture. Similar to the case of field stars in §3.3.1, when we eventually extend the ensemble LC method to unresolved clusters outside the MW, we will encounter contamination from the background stars within the aperture that we define for the cluster. Hence, here I attempt to understand the type and strength of variability in the pixels that make up the background of the cluster field. For this I only consider pixels within an aperture of 1-1.5 times the ‘cluster radius’ given in K13, with flux less than a threshold determined from §3.2.4.1. This way I am able to isolate only the low-flux pixels outside the cluster radius, avoiding the resolved field stars in the area.

Figure 3.9 shows the ensemble LC for NGC 6637 and the corresponding LS periodogram from the background pixels chosen using the aperture described above. From the periodogram, one can see that there is some clear variability at certain periods. However, it is worth noting that the range of variations of the flux in Figure 3.9 is much smaller than that for the field stars in Figures 3.7 and 3.8. The variations in the flux of unresolved background stars were consistently found to be small compared to that of resolved field

and cluster stars in several clusters.

The contribution from the background unresolved stars could be minimized by applying a lower flux threshold on the aperture that we consider for our ensemble cluster photometry. This would selectively remove the low-flux pixels from the analysis and only consider the resolved stars or groups of stars within the cluster radius. However, this technique would not be effective against unresolved clusters that we are eventually planning to analyze with the ensemble photometry method. Hence, a statistical characterization of the photometric variability of background stellar population as a function of age and metallicity is important to ensure minimal contamination in our ensemble photometry method.

3.3.3 Identifying Periodic Features in Ensemble Lightcurves

The ensemble LC for a cluster will contain multiple time varying features that could be caused by different sources within the cluster aperture. The variable stars or sources corresponding to these features can be tracked by identifying the pixels that show strong signatures of variability. LS periodograms can help us identify these high-variability pixels for each frequency.

These features can be seen in Figure 3.10, where I compare the periodograms of the ensemble LCs of old, metal-poor GCs (top row) with those of young, metal-rich OCs (bottom row). Each dominant feature in the periodogram for a cluster has been labelled by the variability sources that most likely produced it. These sources have been determined by studying individual pixels within the cluster aperture that contribute to the ensemble LS periodogram. This tool of visualizing the pixels that contribute to the periodic behaviour in the ensemble LC is valuable in validating the ensemble LC analysis. It helps in confirming that the features seen in the ensemble LC of a cluster are indeed produced by previously determined variable stars that fall within the pixels that contribute to these features. For this, in §3.3.3.1, I consider clusters for which we have comprehensive knowledge of the variables present in them, which are mostly GCs. For clusters that don't have many known variable stars in their vicinity, I look for pixels that contribute to the features observed in the periodogram of the the ensemble LC. This is a strong tool to help detect variable stars that were previously unknown by tracking the stars that fall within these particular pixels.

I explore this method for several OCs described in §3.3.3.2.

3.3.3.1 Ensemble photometry with GCs

I test the accuracy and reliability of using ensemble photometry to determine variables within a cluster by verifying that the features observed in the LS periodogram of the ensemble LC are indeed generated by individual variables that we know are present in a cluster. For this, I use GCs since the Clement et al. (2001) catalog contains information of all known variables stars present in several GCs and references to the studies that identified them.

The top two rows of Figure 3.11 shows the ensemble LC and the corresponding periodogram for NGC 6522. The dominant peak in the periodogram suggest that there may be variability source(s) present within the selected aperture whose brightness varies at that particular frequency. On searching the Clement et al. (2001) catalog, I find a few known variable stars within the aperture of the cluster shown in the bottom row of Figure 3.11 with their frequency of variation and apparent magnitude. One can see that there are two known variables that fall in the same frequency range as the strongest peak in the LS periodogram. These two very bright stars are W Virginis pulsating variables and may likely be contributing to the dominant feature in the ensemble LC periodogram.

Similarly for NGC 5272, another GC, I was able to identify the variables responsible for producing a strong feature in the ensemble periodogram. From Figure 3.12 one can recognize that a group of RR Lyrae stars may likely be responsible for the feature in the periodogram at $f \approx 2.0 \text{ days}^{-1}$. This technique is used on multiple clusters to identify the variable stars that may be contributing to the annotated peaks in the periodograms for individual clusters in Figure 3.10. This analysis improves our understanding of how the information from variables of different frequencies and brightness present in the cluster are captured by the periodogram of the ensemble LC.

3.3.3.2 Ensemble photometry with OCs

Here I identify potentially new variable stars in OCs by tracking the individual pixels that contribute to the strong periodic features in the combined light. Variability studies of

OCs are generally sparse and incomprehensive, mostly comprising of reports of individual detections of variables in the cluster (e.g., Anderson et al. 2013; de Marchi et al. 2007). I have devised a technique that helps in detecting previously unknown variables present in the cluster by analyzing LCs from individual pixels and comparing them with the ensemble LC for the cluster. One example of this technique used for NGC 2422 is shown in the second column in Figure 3.13, which colors each pixel by the power of the periodogram corresponding to the individual LC from that pixel in the frequency range shown in the third column. The light from the variable stars that fall within the radius of the pixel should be responsible for the particular dominant feature in the periodogram. For the OCs that have little or no information about previously known variable stars, this technique can be used to identify potential variable stars.

To determine the sources of observed variability at distinct frequencies, a cone search is done on DR2 of the *Gaia* mission (Gaia Collaboration et al. 2018). For example, in the fourth column in Figure 3.13, the orange points represent the *Gaia* stars that fall within the pixels with the strongest contribution to the dominant feature in the ensemble periodogram. Some of these stars shown in orange may be variables contributing to the dominant feature shown in the periodogram in Figure 3.13. The star with $G \approx 12$ is the most likely candidate due to its high brightness. This periodic variability is evident from the LC of the individual strongest pixels shown in Figure 3.14. The LC phase-folded at $f \approx 0.72 \text{ days}^{-1}$, and the corresponding LS periodogram, are also included to highlight that the variations have the same frequency as the strongest peak in the periodogram in Figure 3.13. Using this technique, I have identified several potential variable stars within the range of OCs and have annotated Figure 3.10 with this information.

3.4 Summary

The main goal of the project described in this chapter is to develop techniques to estimate cluster properties using ensemble photometry. To understand the components of the ensemble light that is obtained from the TESS survey, I explore several methods to optimize the data processing and detrending techniques applied to the LCs (§3.2.4). I

identify optimal values for parameters such as the flux threshold to select pixels that make up the background aperture, and the number of PCA vectors to be used for the detrending process.

With the help of the optimized detrending tool, I use the ensemble photometry method described in §3.2.4 to explore the variability of resolved field stars (§3.3.1) and unresolved background stars (§3.3.2) within the cluster range. I find that there is non-trivial variability in the ensemble light from both the resolved and unresolved stars that make up the field of the cluster. The effects of these sources of contamination should be considered while analyzing the final ensemble LCs of clusters. The variability of resolved field stars and unresolved background stars should be statistically characterized as a function of age and metallicity to minimize contamination from these sources in the ensemble photometry method.

The ensemble LC method is also used to acquire and analyze the variability signatures of cluster members from known GCs and OCs (§3.3.3). I analyze the ensemble LCs of GCs with previously known variable stars to validate the method and confirm that the information from individual variability sources are being captured by the ensemble LCs as well. I identify previously unknown potential variable stars in OCs by comparing the periodic features in the ensemble LC with those of individual pixels that contribute to the variability. The stars that fall within the pixels that show strong periodic features should contain the variability sources corresponding to the periodic features in the ensemble LC.

The analyses performed in this chapter lays the groundwork for using techniques to quantify the variability in ensemble LCs and extract information of the physical properties of clusters. Further analyses have to be done to achieve this goal. For example, we have to explore different types of metrics and their efficiency in quantifying the photometric variability. Furthermore, we should explore combinations of these metrics to characterize photometric variability for a large number of clusters for which we have accurate age and metallicity determination (refer §3.2.1). Once we have a trained model, we can estimate properties like the age and metallicity of clusters using only ensemble photometry. This technique is particularly useful for unresolved clusters in nearby galaxies. The scope and

potential impact of this methodology high, and we have performed the pilot study that will act as support and groundwork for future explorations.

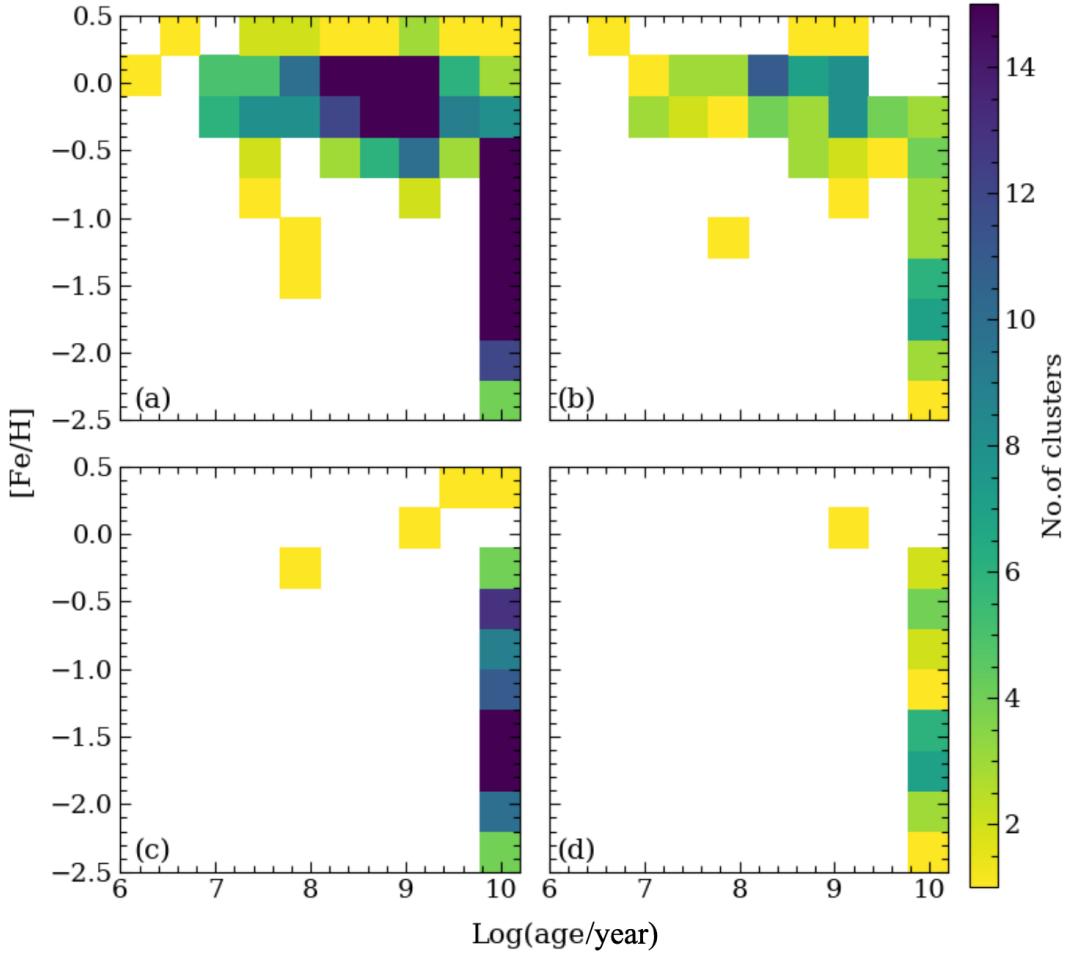


Figure 3.1: Shows the age-metallicity distribution of clusters for which we have good age and metallicity estimates from literature. Individual subplots show clusters with (a) no additional constraints, (b) accessible TESS data, (c) information from literature about the variability sources present in them, (d) both TESS data and variability information.

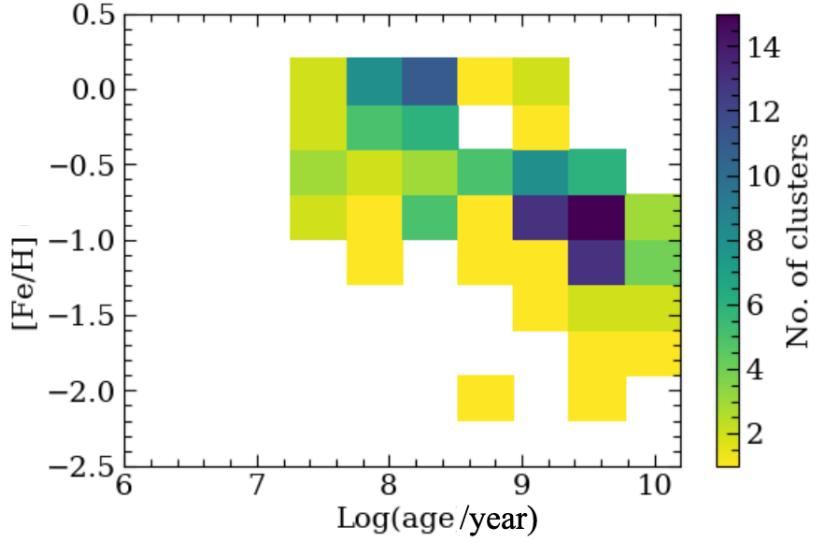


Figure 3.2: The age-metallicity distribution of clusters in the Magellanic Clouds for which we have age and metallicity estimates (Bica et al. 2020).

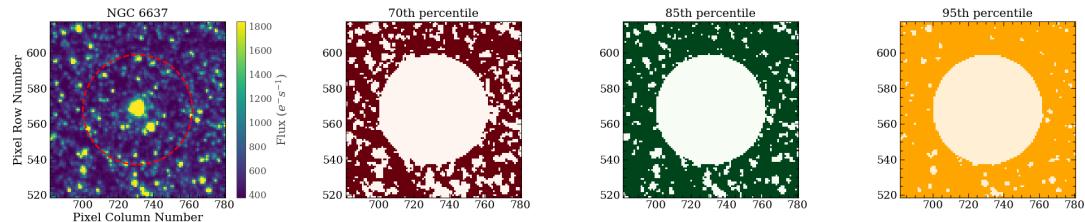


Figure 3.3: From left: (a) shows flux map and cluster aperture for NGC 6637; (b), (c), and (d) show pixels included in the background aperture when the background flux threshold is set to the 70th, 85th, and 90th percentiles of the flux distribution (refer §3.2.4.1)

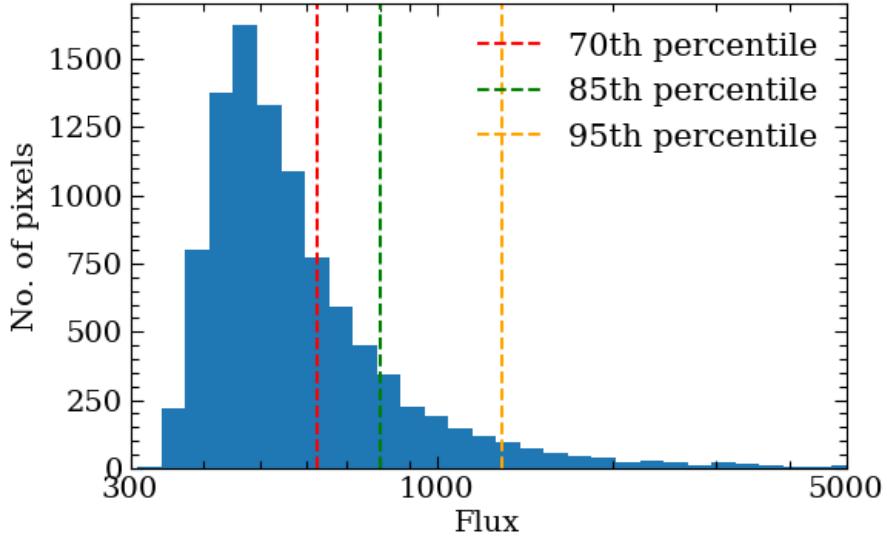


Figure 3.4: Histogram of flux in the TESS cutout for NGC 6637 with different background thresholds indicated using dashed lines colored by the apertures used in Figure 3.3 (refer §3.2.4.1).

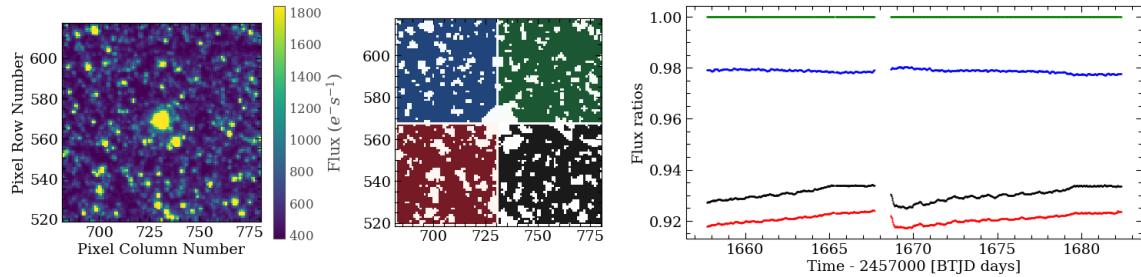


Figure 3.5: Distribution of background pixels within each quadrant for NGC 6637 (refer §3.2.4.2). The colors in the left plot correspond to each quadrant and the right plot shows ratio of the TESS LCs from pixels within the each quadrant with the third quadrant for NGC 6637. The center plot shows the flux map for the cluster cutout.

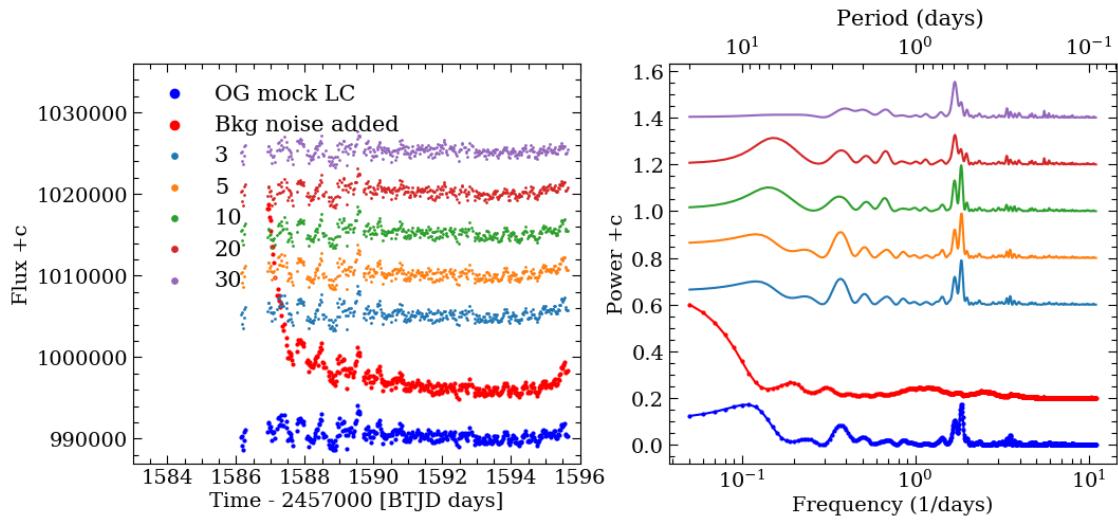


Figure 3.6: LCs from the mock LC analysis (§3.2.4.3) detrended using different numbers of PCA vectors for NGC 3201.

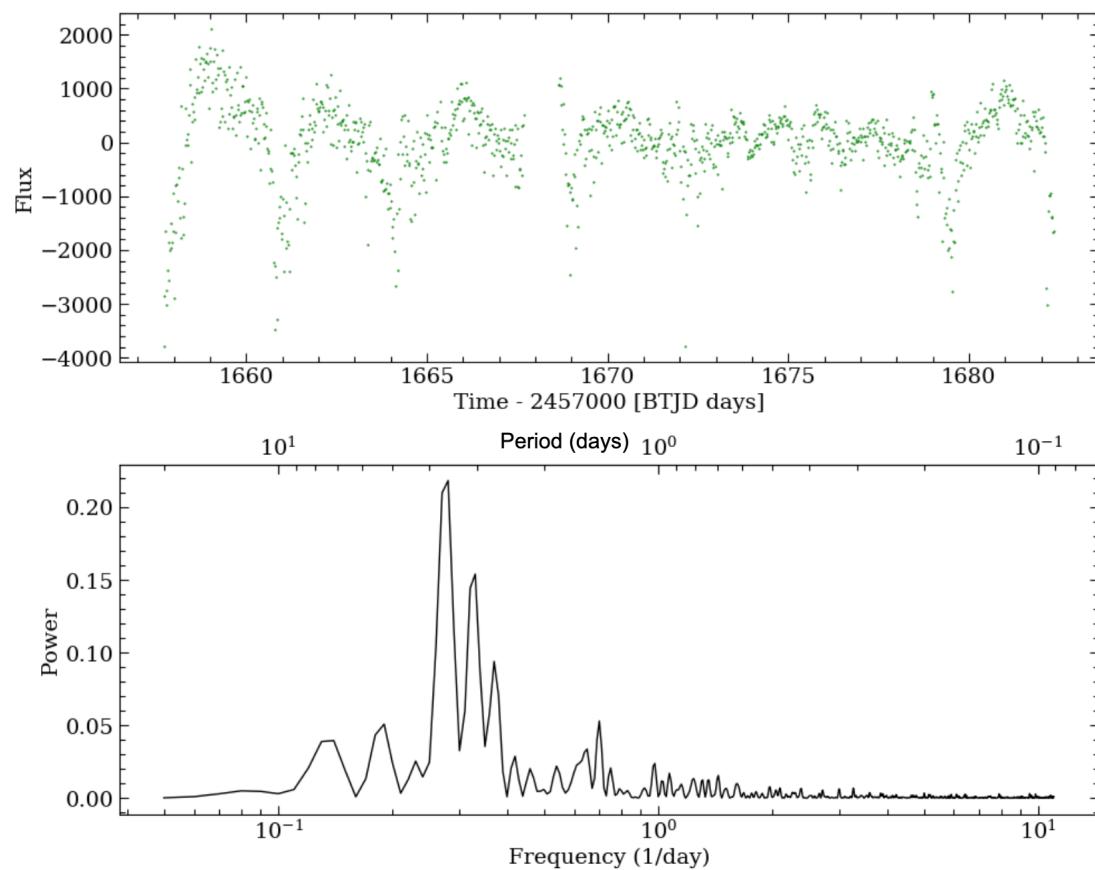


Figure 3.7: The ensemble LC for field stars in NGC 6637 along with the corresponding periodogram.

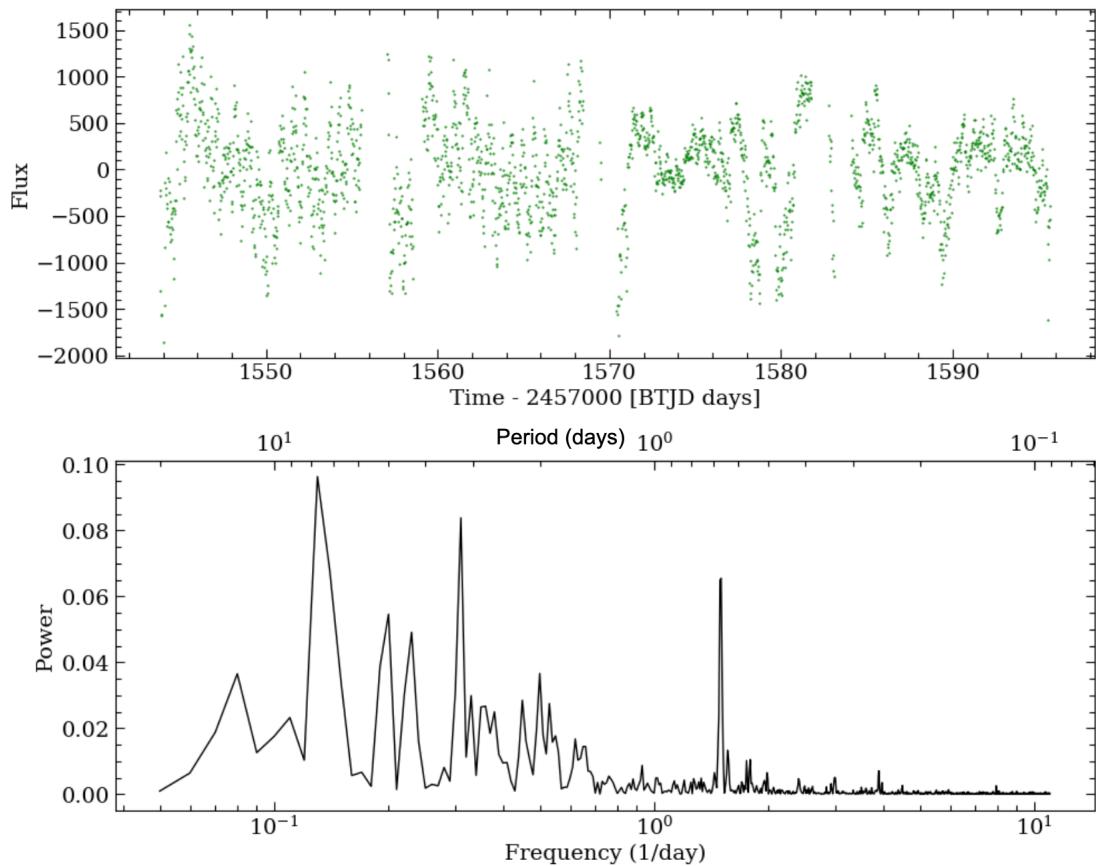


Figure 3.8: The ensemble LC for field stars in NGC 3201 along with the corresponding periodogram.

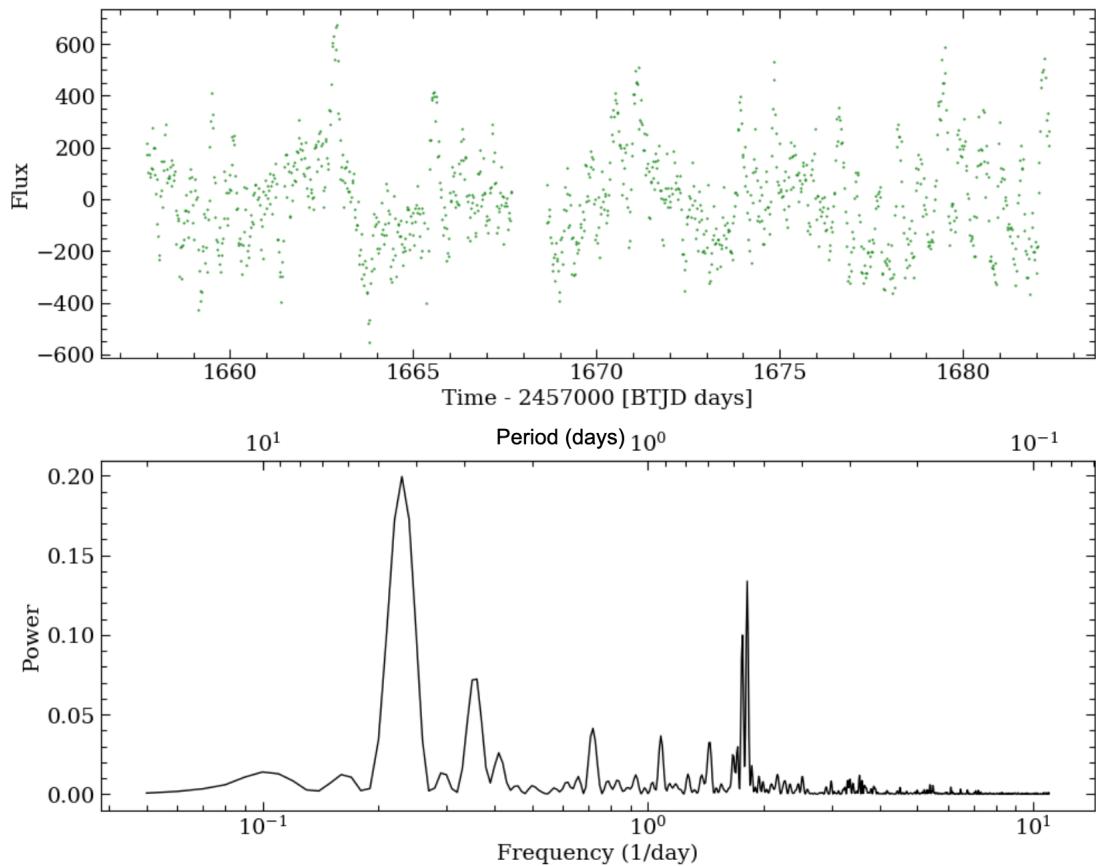


Figure 3.9: The ensemble LC for unresolved background stars in NGC 6637 along with the corresponding periodogram.

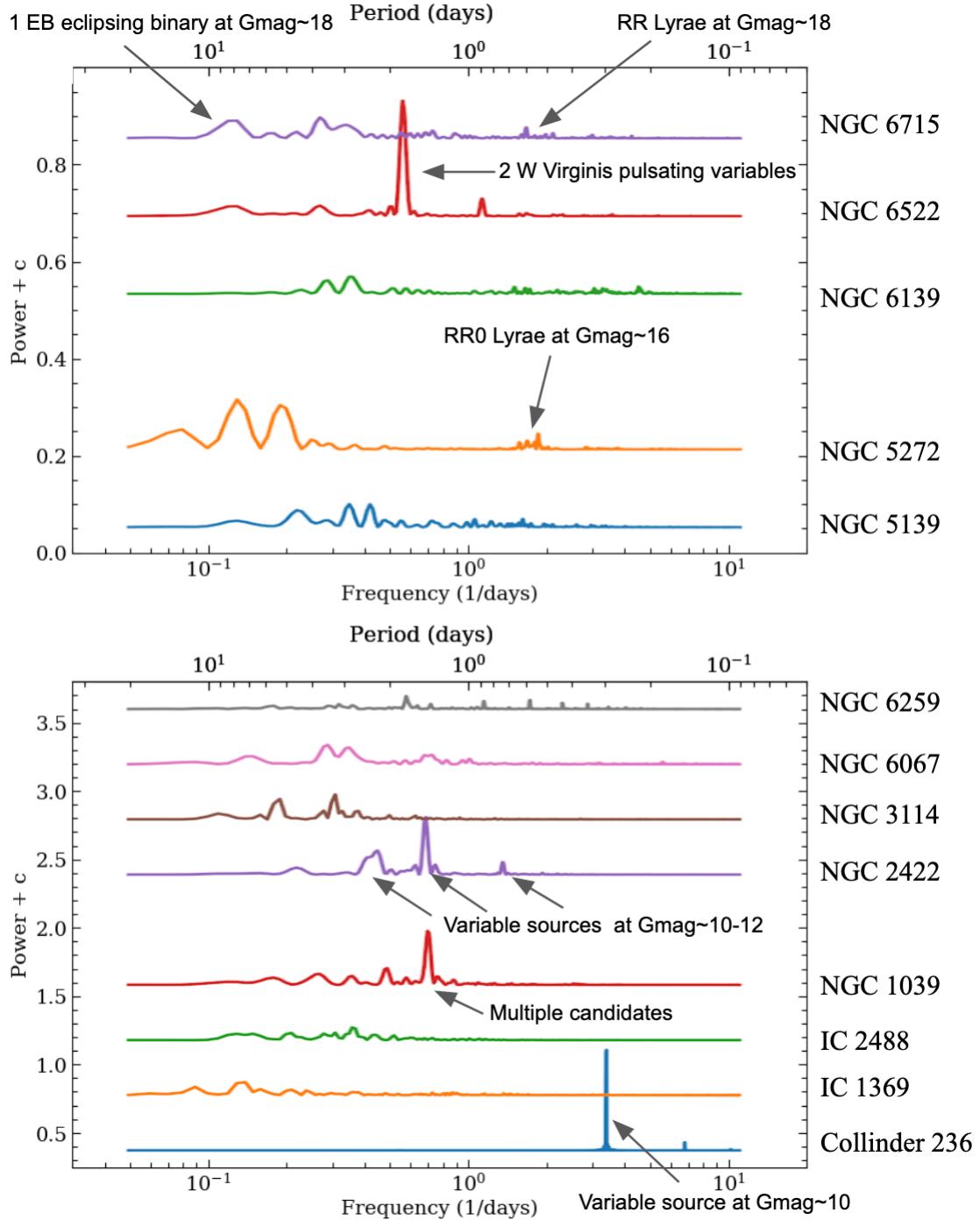


Figure 3.10: Periodograms of ensemble LCs from old, metal-poor GCs (top row) and young, metal-rich OCs (bottom row) with each dominant feature labelled by the stars potentially contributing to them. The individual periodograms are also labelled by the clusters they correspond to.

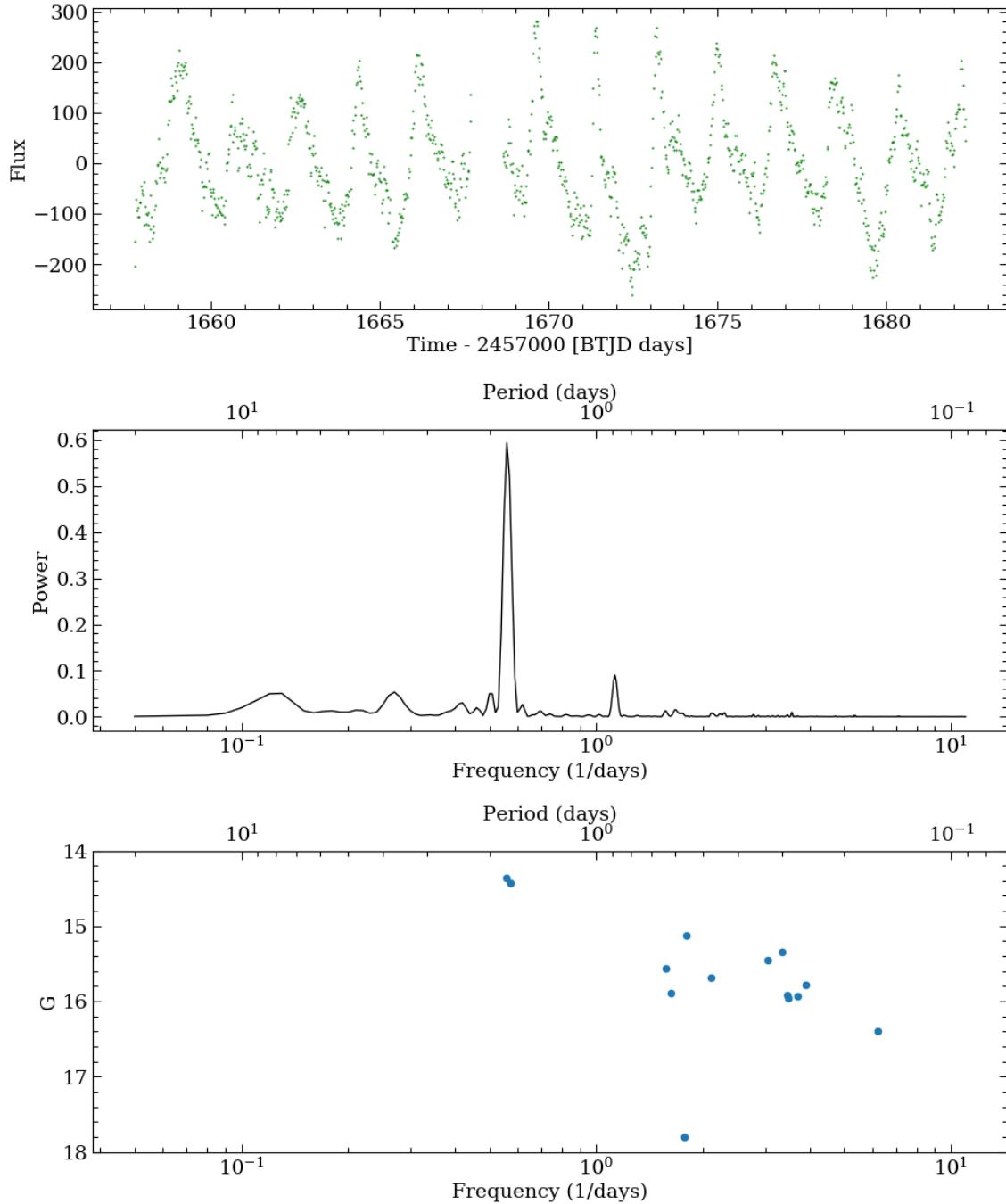


Figure 3.11: The ensemble lightcurve, periodogram, and known variable stars in NGC 6522 are shown. Notice that the clear variable signature in the lightcurve, which is also confirmed by the strong peak in the periodogram, corresponds to two already known variables.

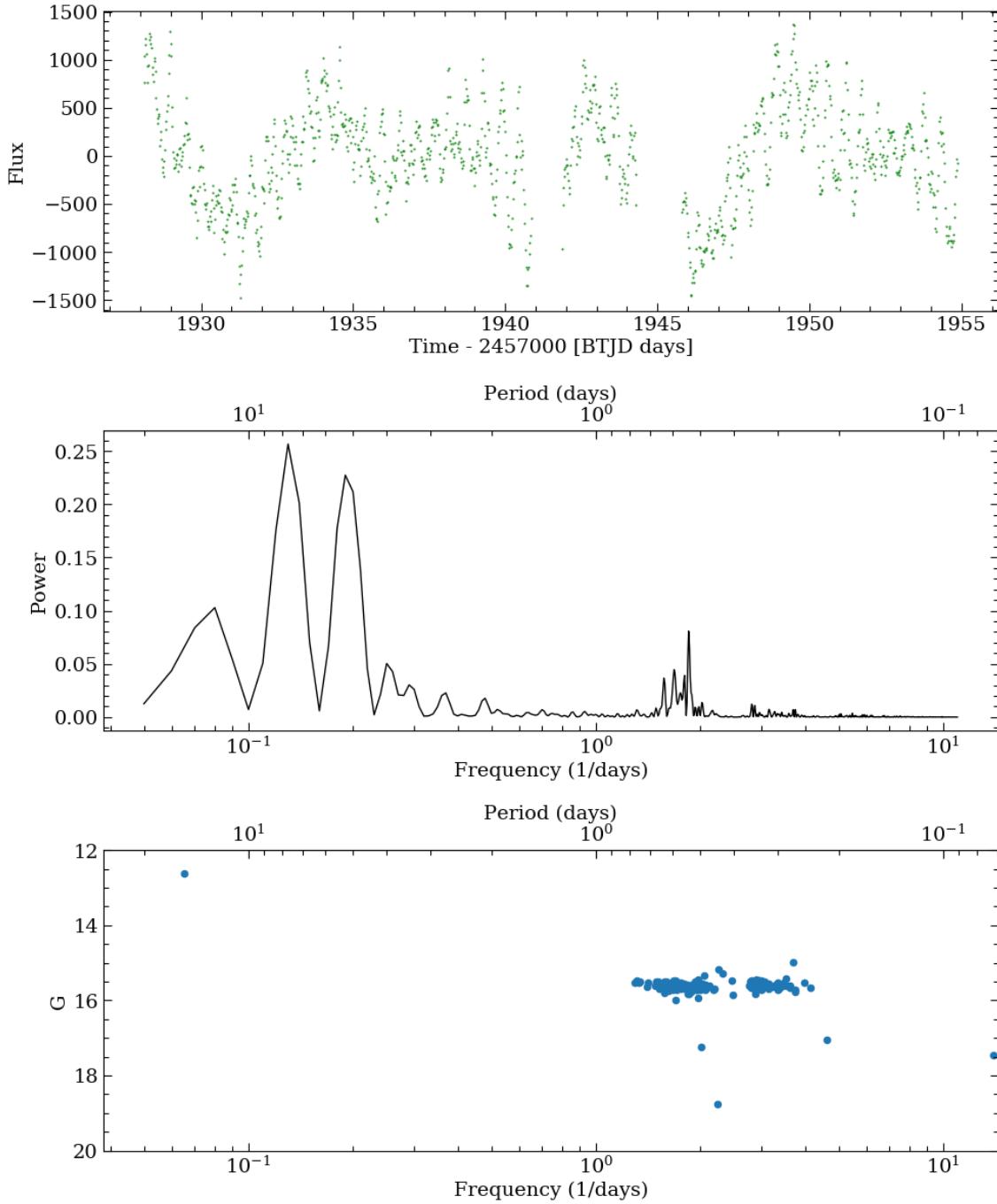


Figure 3.12: The ensemble lightcurve, periodogram, and known variable stars in NGC 5272 are shown. Notice that the variability feature at $\text{freq} \approx 2.0 \text{ days}^{-1}$ in the periodogram corresponds to group of already known variables.

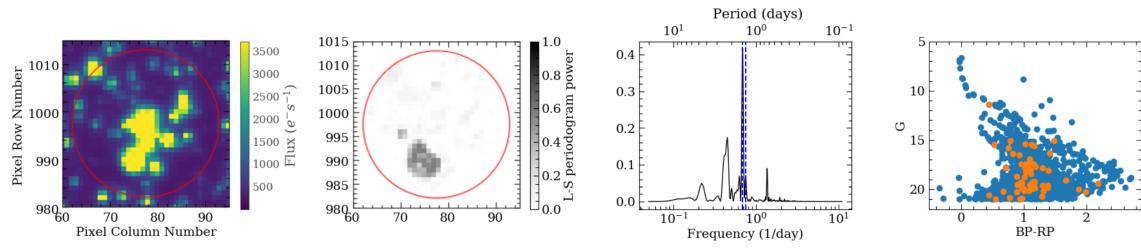


Figure 3.13: From left: (a) Flux map for NGC 2422 with the cluster aperture shown. (b) Pixels in the aperture colored by the power of the periodogram corresponding to the individual LC from that pixel for a specific frequency range. (c) Shows the Lomb-Scargle periodogram for the ensemble light from NGC 2422 with the frequency range used for the selection in (b). (d) Gaia stars within the aperture are shown in blue and those within the strongest pixels in (b) are shown in orange.

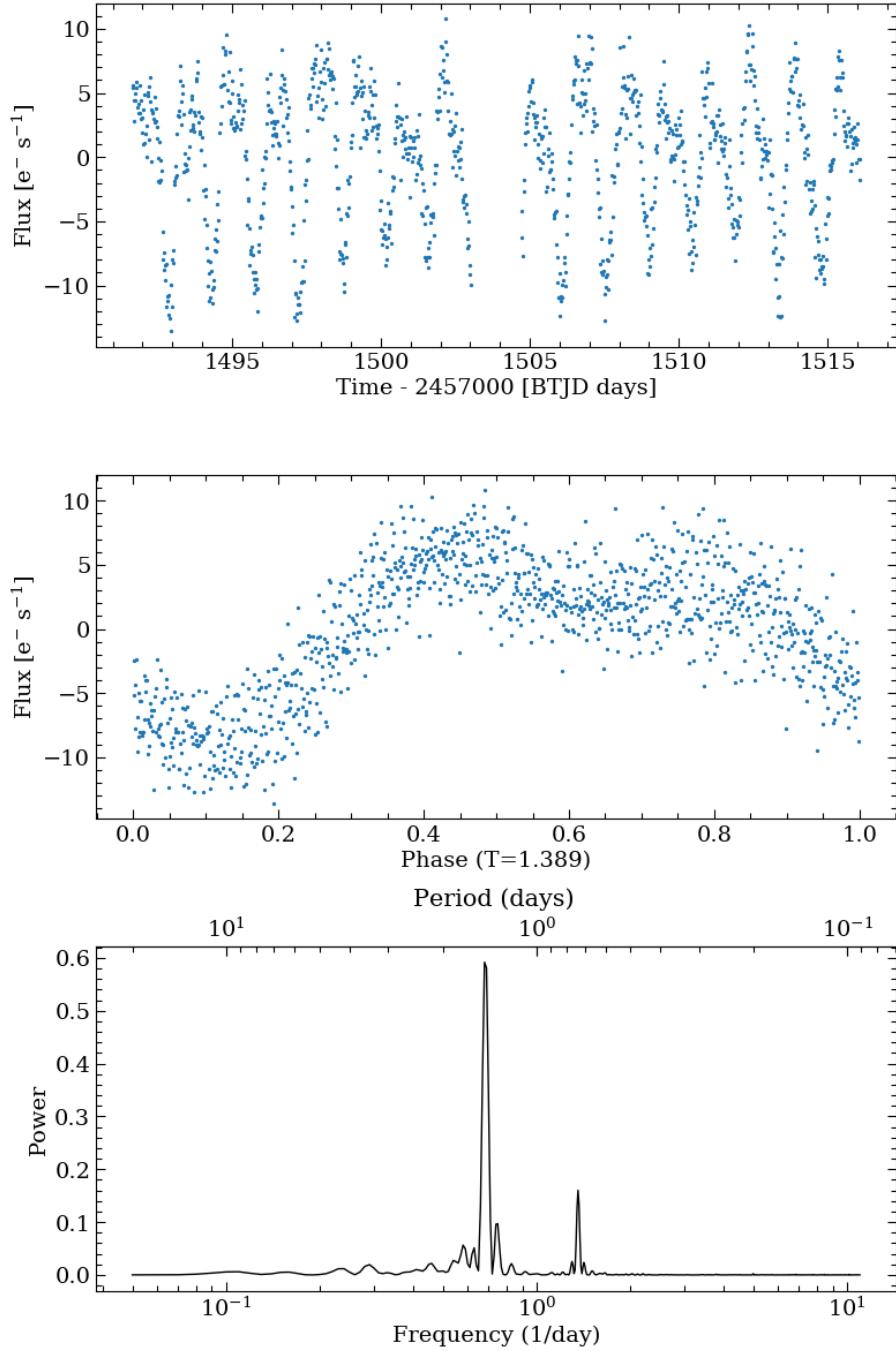


Figure 3.14: The detrended LC for the pixels with the strongest contribution to the feature observed in the ensemble periodogram at $f \approx 0.72$ (1/day) in Figure 3.13.

REFERENCES

- Alonso-Santiago, J., Negueruela, I., Marco, A., Tabernero, H. M., González-Fernández, C., and Castro, N. 2017, MNRAS, 469, 1330
- Anderson, R. I., Eyer, L., and Mowlavi, N. 2013, MNRAS, 434, 2238
- Bergbusch, P. A., Vandenberg, D. A., and Infante, L. 1991, AJ, 101, 2102
- Bica, E., Westera, P., Kerber, L. d. O., Dias, B., Maia, F., Santos, J. F. C., J., Barbuy, B., and Oliveira, R. A. P. 2020, VizieR Online Data Catalog, J/AJ/159/82
- Clement, C. M., et al. 2001, AJ, 122, 2587
- de Marchi, F., et al. 2007, A&A, 471, 515
- Dhillon, V. S., et al. 2007, MNRAS, 378, 825
- Dhillon, V. S., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, and H. Takami, 99080Y
- Donor, J., et al. 2018, AJ, 156, 142
- Dragomir, D., et al. 2019, ApJL, 875, L7
- Gaia Collaboration et al. 2018, A&A, 616, A1
- Huang, C. X., et al. 2018, ApJL, 868, L39
- Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., and Scholz, R.-D. 2013, A&A, 558, A53
- Koch, D. G., et al. 2010, ApJL, 713, L79
- Lennon, D. J., Dufton, P. L., Fitzsimmons, A., Gehren, T., and Nissen, P. E. 1990, A&A, 240, 349
- Lightkurve Collaboration et al. 2018, Lightkurve: Kepler and TESS time series analysis in Python, Astrophysics Source Code Library
- Pollacco, D. L., et al. 2006, PASP, 118, 1407

Ricker, G. R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, ed. J. Oschmann, Jacobus M., M. Clampin, G. G. Fazio, and H. A. MacEwen, 914320

Ricker, G. R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M., and Mateo, M. 1992, , 42, 253

Welch, D. L., and Stetson, P. B. 1993, AJ, 105, 1813

Wheatley, P. J., et al. 2018, MNRAS, 475, 4476

CHAPTER 4

SUMMARY AND FUTURE WORK

4.1 Cluster Chemical Homogeneity and Chemical Tagging

In §2, I devise a novel method to determine stellar cluster members using only their kinematic information. The cluster members identified for a large number of clusters in the K13 catalog using this technique is available for download as a machine readable table along with the final published paper. In §2, I also explore the dependence of cluster chemical homogeneity on various Galactic and cluster properties. I find that Mg and Ca show a strong, relatively significant correlation between cluster chemical scatter and velocity dispersion, while Ni, Si, Al, and Fe may also exhibit a possible positive correlation, albeit at low significance. Despite the small sample size, if true, these findings suggest a quantitative difference between the correlation lengths of the metals dispersed into the ISM as a result of Type Ia SNe and CCSNe. The existence of an intrinsic difference in the distance to which the elements are expelled by these two SNe explosions would affect our understanding of the pollution rates and mixing efficiency in the ISM. This study provides reason and motivation to explore this dependence on more elements and a larger sample of open clusters. If validated by future larger studies, this result should be included in existing and future Galactic chemical evolution models and simulations.

The results from the cluster chemical homogeneity project also have potential implications for chemical tagging, which assumes that OCs are intrinsically chemically homogeneous. Chemical tagging attempts to determine birth siblings, cluster members, or co-natal objects using only the chemical signatures of the stars. I find that the abundance scatter in most elements for our clusters are within the limits previously found. However,

a larger OC sample and more elements should be used in future studies to explore the empirical dependencies of cluster homogeneity on velocity dispersion. Such efforts will have considerable impact on the feasibility of chemical tagging. For example, the most massive OCs could either be altogether avoided in chemical tagging studies, or be studied with caution for elements that are predominantly produced by CCSNe.

4.2 Cluster Variability and Age

In §3, I explore the photometric variability of resolved field stars and unresolved background stars within the cluster range. The effects of the non-trivial variability observed from each of these sources should be considered while analyzing the final ensemble LCs of clusters. In future work, the contamination from field stars can be reduced by limiting the radius for the cluster aperture while considering the ensemble light from it. Similarly, the variations from the unresolved background stars can be minimized by imposing a lower flux threshold in the cluster aperture.

The ensemble LC method is also used to acquire and analyze the variability signatures of cluster members from known GCs and OCs. I analyze the ensemble LCs of GCs with previously known variable stars to validate the method and confirm that the information from individual variability sources are being captured by the ensemble LCs as well. I identify previously unknown potential variable stars in OCs by comparing the periodic features in the ensemble LC with those of individual pixels that contribute to the variability.

The analyses performed in §3 lays the groundwork for using statistical techniques to quantify the variability in ensemble LCs and extract information of the physical properties of clusters. Further analyses have to be done in the future to achieve this goal. For example, we have to explore different types of metrics and their efficiency in quantifying the photometric variability. Furthermore, we should use combinations of these metrics to characterize photometric variability for a sample of clusters which we have accurate age and metallicity determinations. Then we can estimate properties like the age and metallicity of clusters using only ensemble photometry. This technique is particularly useful to determine properties of unresolved clusters in other galaxies. Once fully developed, the scope and impact of this technique is huge, and this project describes the pilot study that will act as

support and groundwork for future explorations.