THE CLUSTERING OF YOUNG STELLAR CLUSTERS IN NEARBY GALAXIES

A Dissertation Presented

by

KATHRYN A. GRASHA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2018

Astronomy

THE CLUSTERING OF YOUNG STELLAR CLUSTERS IN NEARBY GALAXIES

A Dissertation Presented

by

KATHRYN A. GRASHA

Approved as to style and content on 04/09/2018 by:

Daniela Calzetti, Chair

Mauro Giavalisco, Member

Stella Offner, Member

Krista Gile, Member

S. Schneider, Department Chair
Astronomy

DEDICATION

To my better half.

EPIGRAPH

Don't compare your life to others.

There is no comparison between the Sun and the Moon.

They shine when it's their time.

ACKNOWLEDGMENTS

First and foremost, I am eternally grateful for my husband, Andrew, for his unending support and encouragement to live a fulfilling life. He experienced all of the ups and downs of my research and his patience and willingness to proof read countless papers and aid in the identification of indecipherable blobs of light known as star clusters describes dedication that not many would care to endure. I am indebted to my undergraduate adviser, Prof. Jeremy Darling. He undertook the enormous burden and responsibility to start me down the path of research in astronomy. His unwavering reassurance and support prepared me to succeed in the challenging path known as graduate school. I also thank my family, my parents Rosella and Matthew, for their support in my effort to pursue my chosen field of astronomy, and my siblings Paul, Sharon, Rebecca, and Mary, for letting me have all the brains in the family; you all know how I do not like to share.

My most sincere gratitude goes to my dissertation adviser, Prof. Daniela Calzetti, for the patient guidance, encouragement, enthusiasm, and advice she has provided throughout my time as her student. She gave an immeasurable amount of her time, as a researcher and a mentor, to help me develop the grit and resilience to become a successful, independent scientist. She believed in me when I did not believe in myself and I have been extraordinarily lucky to have a supervisor who cares so much about my work and myself.

I thank my dissertation committee for their guidance, support, and patience: Prof. Daniela Calzetti, Prof. Mauro Giavalisco, Prof. Krista Gile, and Prof. Stella Offner. I also acknowledge the people who have helped me during the course of writing the results presented in this dissertation. In particular, Dr. Andrew Battisti, Timothy

Conklin, David Schenck, and Brandt Gaches for comments, discussion, coding help, and never ending moral support that immensely enhanced and developed my research. I thank the department staff for literally keeping the department together and the astronomy faculty at UMass as a whole, for creating an exceptional community in which to be a graduate student.

I am fortunate to have been part of the fantastic LEGUS collaboration. This dissertation would not be possible without the shear amount of work from all the members of the LEGUS team. I want to especially acknowledge all the star cluster identifiers who put in an enormous amount of seemingly thankless effort, time, and eye strain to help create the star cluster catalogs that are core to this research. My work and accomplishments would not exist without the dedication from all of them. Dr. Angela Adamo has also been instrumental in the development of my work and never hesitated to reach out, no matter how tough the questions asked by a referee.

I am exceedingly grateful for my graduate advisers, Prof. Daniela Calzetti and Prof. Todd Tripp, for their dedication to ensure my financial support during my graduate studies. I also thank the Massachusetts Space Grant Consortium, the American Astronomical Society, and the Mary Dailey Irvine Fund for the numerous travel grants that allowed me to present this dissertation work at conferences around this beautiful globe.

This work is based on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, under NASA Contract NAS 5–26555. These observations are associated with Program 13364 (LEGUS). Support for Program 13364 was provided by NASA through a grant from the Space Telescope Science Institute.

Finally, I want to thank all the amazing and inspirational women that I had the privilege to learn from over the course of my undergraduate and graduate career,

specifically Prof. Daniela Calzetti, Prof. Erica Ellingson, Prof. Alexandra Pope, Prof. Stella Offner, Alyssa Sokol, Dr. Bethany Wilcox, Dr. Allison Kirkpatrick, Lauren Kahre, Rhonda Holton, Karen Williamson, and Laura Archibald Kelley. More generally, I want to acknowledge all the women in STEM who battle everyday their own self and a culture that judges their worth, their ability, their effort, and their achievements based on their gender identity. Never doubt yourself. You are capable.

ABSTRACT

THE CLUSTERING OF YOUNG STELLAR CLUSTERS IN NEARBY GALAXIES

MAY 2018

KATHRYN A. GRASHA

B.A., UNIVERSITY OF COLORADO BOULDER

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Daniela Calzetti

Star clusters form the basic building blocks of galaxies. Star clusters span a wide range of ages, from a few million years to billions of years, making them exceptional tracers of the star formation histories of their host galaxies. Star formation is the process by which galaxies build up their stellar populations and their visible mass and occurs in a continuous, hierarchical fashion across a large dynamical range, from individual stars up to kiloparsec-scale correlated ensembles of stellar aggregates. It is the formation, evolution, and eventual destruction of these large hierarchical star-forming complexes that provide an essential role in understanding the mechanism, process, and dynamical evolution of star formation on sub-galactic scales.

In the first part of this dissertation, we use a sample of eight local galaxies and find that young star clusters are reliable tracers for the hierarchical patterns of star formation. We characterize the correlation using a two-point statistic, which allows us to analyze global correlations. We find that star formation is coherent over scales of

a few hundred parsec up to a few kpc depending on the galaxy, and in all cases, these hierarchies are short lived and unbound, dissolving in a few tens to a hundred Myr. The recovered correlations between the spatial separations and ages of star clusters contained within these unbound hierarchical structures are consistent with theoretical expectations of arising from a turbulence-driven interstellar medium (ISM). More importantly, we found evidence that the maximum size of correlated star formation is driven by galactic shear. This work represents the first large-scale study of its kind outside of the Local Group to characterize turbulence as the physical driver of correlated star formation and marks a turning point in the effort to link local star forming structures to those that are common at high redshift.

In the second part of this dissertation, we combine our star cluster catalogs with exquisite molecular gas observations to connect the detailed stellar population information to the natal gas from which it formed. In a study of two galaxies, we find that the timescale for star clusters to lose association with their natal clouds is of order a few Myr, a result that is independent of the spiral structure of the host galaxy. These works represent the first such large-scale cross-correlation between star clusters and molecular gas outside the Local Group and provide vital theoretical guidance to connect gas to stellar components.

We also introduce initial work that employs the use of machine learning as a process to identify star clusters, a quicker and more homogeneous method to implement than the traditional visual classification technique employed for most of the stellar cluster catalogs.

This dissertation is based on three papers that have appeared in the peer-reviewed literature with two more papers submitted to peer-reviewed literature[1].

[1]These papers are Grasha et al. (2015, 2017a,b, 2018a,b). One other paper unrelated to the topic of this dissertation has also been written during the course of my graduate career (Grasha et al., 2013).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

INTRODUCTION

Virtually all the light originating from outside Earth is created via a nuclear fusion process in the center cores of stars, nearly perfect spheres of hot plasma in hydrostatic equilibrium, a complex interplay between the outward push of photon energy and inward gravitational energy. The light from our nearest star, the Sun, is responsible for the creation of the energy that sustains life on Earth. The collective light from nearby stars create recognizable patterns in the night time sky that have captured the imagination and creativity of individuals and cultures for many millennia.

The Milky Way[2] galaxy is composed of billion stars in a variety of masses, which determines the size, luminosity, and lifetime of a star. It was not until the 1920s that observations revealed that the Milky Way is just one of many galaxies throughout the universe (Hubble, 1925), solving the so-called 'Great Debate' (Shapley & Curtis, 1921) on the extragalactic nature of 'spiral nebula' outside the Milky Way. It is stars that form the basic building blocks of the visible matter in all these galaxies.

Understanding the nature of how the vast diversity of galaxies form and evolve is a major outstanding topic in modern astronomy. Star formation is ultimately responsible for giving rise to the variety of observed stellar population ages of these galaxies and is a signpost of their formation history. Due to the scale-free, hierarchical nature of star formation, large-scale structures within galaxies such as spiral arms, circumnuclear rings, etc., are built upon the accumulation of smaller, denser components of star clusters and individual stars (Elmegreen & Elmegreen, 1983). The highly

[2]The term Milky Way is a translation from Latin *via lactea*, from the Greek phrase *galaxías kýklos* — milky circle.

ordered star formation process at the individual stellar scale will inevitably impact the global morphology of a galaxy, and a vast amount of effort has been invested in the theoretical understanding of the underlying physics (Hopkins, 2013a; Hopkins et al., 2013; Krumholz, 2014), but an adequate and comprehensive connection over a wide range of scales has yet to be observationally achieved to help guide theories. A detailed understanding requires characterizing the physical dependencies of star formation within the changing environment of individual galaxies across galaxy-scales. The results presented in this dissertation represent ongoing efforts to connect the scales of star formation, with a particular emphasis on the role star clusters play in tracing the star formation patterns within galaxies and to quantify how the clustering of star formation evolves both in space and in time.

CHAPTER 1

OVERVIEW OF YOUNG STAR CLUSTERS AS TRACERS OF RECENT STAR FORMATION

This chapter introduces the background material relevant to the research presented in this dissertation and includes: an overview of star clusters and their properties, methods for quantifying the process of star formation and its patterns, and the molecular reservoirs from which star clusters originate.

1.1 Introduction to Star Clusters

The majority of stars do not form in isolation but instead are believed to form in coeval groupings called star clusters (Lada & Lada, 2003; Portegies Zwart et al., 2010). Star clusters come in all shapes and sizes and are composed of a few hundred to tens of millions of individual stars, with typical sizes of one to tens of parsec (pc) and total stellar masses of $100 - 10^7$ $M_\odot$ (Portegies Zwart et al., 2010). These stellar systems, from $\sim$0.1 – 1000 stars/pc^3, form at the same time within their parent cloud of gas and dust[1]. Only a small fraction, order $\sim$5%, of these dust-embedded stellar groupings will remain gravitationally bound and emerge as long-lived star clusters (e.g., Goddard et al., 2010; Adamo et al., 2011; Johnson et al., 2016). Long-lived star clusters are fundamental laboratories for studying stellar evolution, star formation history of a galaxy, and the shape of the stellar initial mass function (IMF; e.g., Salpeter, 1955; Kroupa, 2002; Bastian et al., 2010), a method to parameterize

[1]Recent observations, however, have demonstrated that star clusters are not composed of single stellar populations and provide evidence of multiple generations of star formation (see, e.g., Piotto et al., 2005; Villanova et al., 2007; Peacock et al., 2013).

the relative number of stars as a function of their mass at the time of their initial formation.

This dissertation focuses on young ($\lesssim$100 Myr) star clusters as tracers of recent star formation. As young stars are typically associated with the interstellar medium (ISM), the vast majority of star formation, and thus the clusters themselves, will be located within the disk of a galaxy. Due to the dust extinction present in the Galactic plane, which hinders our ability to detect stellar systems outside the small volume of the Milky Way near our Sun, studies of young star clusters are primarily reserved for external galaxies[2] to mitigate this effect. This is in contrast to globular[3] clusters, old and massive stellar clusters that primarily reside in the halo of galaxies and do not trace recent star formation. Figure 1.1 shows the spatial distributions of known Milky Way star clusters, demonstrating the distinct spatial location of observed young clusters, also referred to as "open" clusters because they are gravitationally unbound, versus globular clusters. Figure 1.2 shows R136, a very young and massive star cluster in the Large Magellanic Cloud (LMC; distance 50 kpc), as well as a zoom in of M 83, a nearby galaxy (distance 4.5 Mpc), where the star clusters are clearly seen emerging from their natal clouds within the spiral arms of the galaxy.

1.2 Overview of Stellar Population and Galaxy Properties

There are numerous physical properties that can be measured in stellar populations and galaxies. Parameters listed here are relevant to the research in this dissertation.

---

[2]Notable examples of Milky Way studies of star clusters are Evans et al. (2009); Gutermuth et al. (2009); Lada et al. (2012), among others. These studies employ near-infrared wavelengths to identify the star clusters where extinction is less severe.

[3]Derived from the Latin word *globulus* — a small sphere.

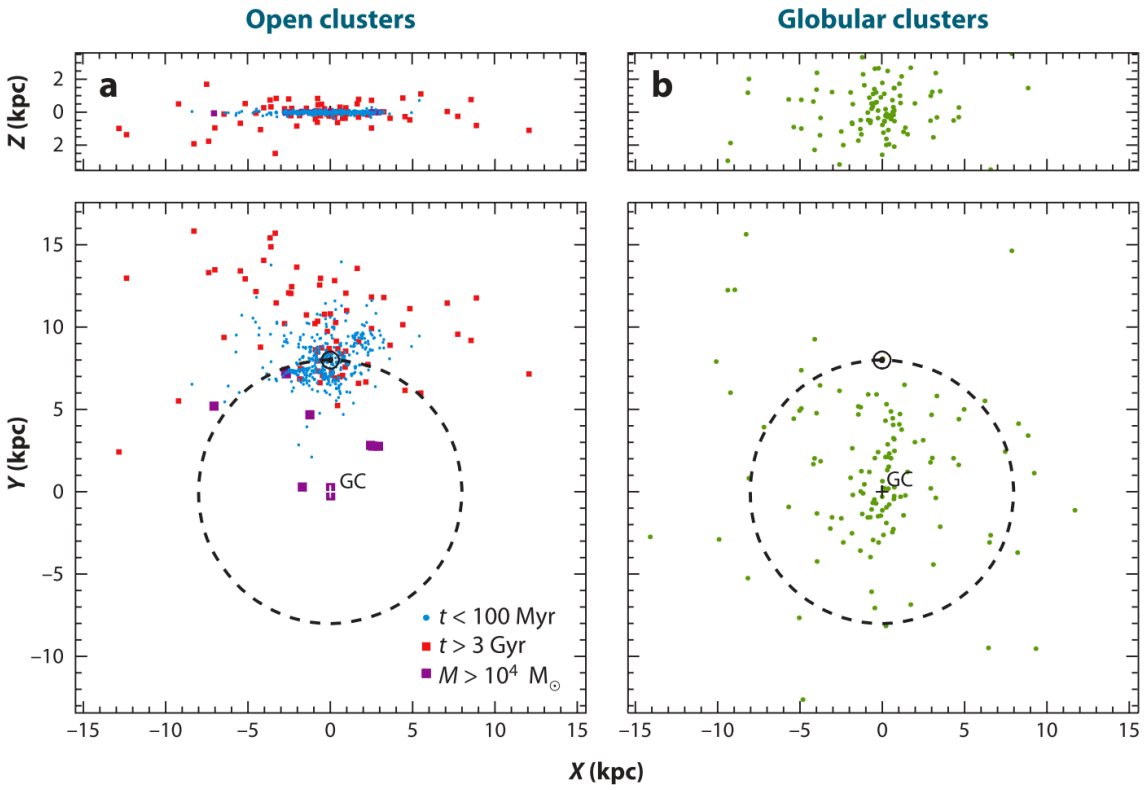

Figure 1.1. Distribution of star clusters in the Milky Way. The upper panels show an edge-on view of the Galactic plane whereas the lower panels show the face-on view. *Left*: The spatial distribution of young (<100 Myr; blue dots) open (unbound) star clusters are preferentially located in or near the Galactic plane compared to the distribution of old (>3 Gyr; red squares) clusters. The excess of points near the Sun is a selection effect. The center of the Galaxy is labeled "GC". *Right* : globular (bound) clusters (green dots) are typically located in the bulge and halo. Image adapted from Portegies Zwart et al. (2010).

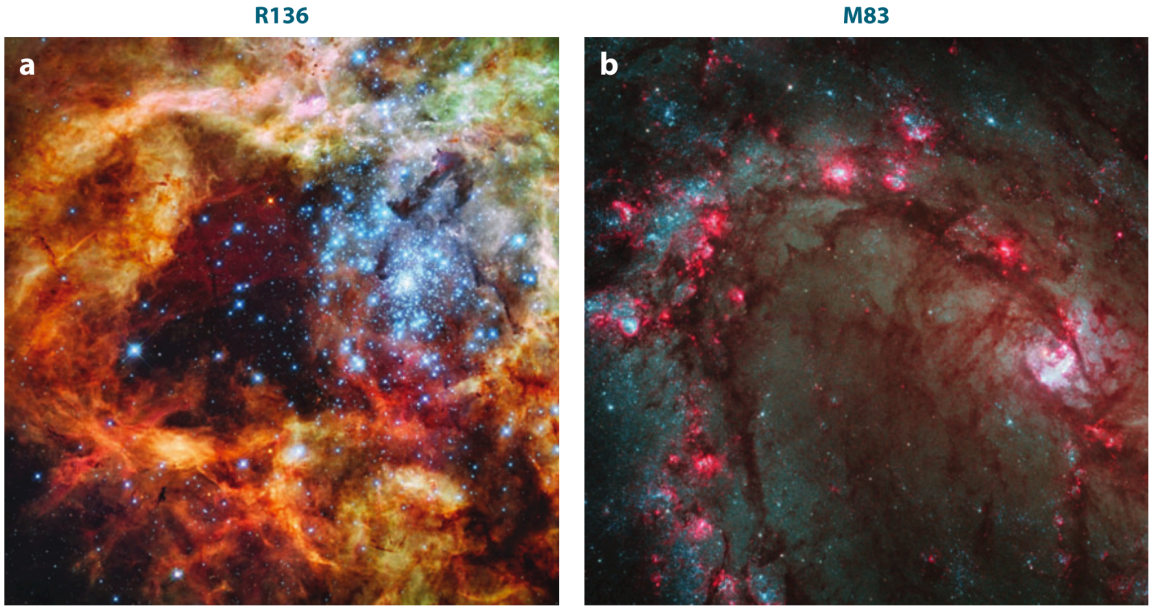

Figure 1.2. Examples of resolved young star clusters. (a) A 50x50 pc^2 region around the 10^5 $M_\odot$ star cluster R136 in the 30 Doradus region of the Large Magellanic Cloud (blue objects; distance of 50 kpc). (b) Young star clusters forming in the spiral arms of M 83 (blue objects; distance of 4.5 Mpc). Young stars emit a large quantity of high energy radiation, which ionizes the surrounding molecular medium (called H II regions, denoting the hydrogen is ionized), giving rise to the distinctive pink regions that surround young clusters. Image adapted from Portegies Zwart et al. (2010).

- **Luminosity $\mathbf{L}_\lambda$**. Quantifies the total amount of energy per unit time per unit wavelength emitted by a source at a specific wavelength λ. As astronomical objects are located at a distance D from Earth, the observed flux (i.e., apparent brightness) is the intrinsic luminosity decreased by a factor of $4\pi D^2$.

- **Magnitude**. Integrated fluxes from astronomical sources are typically quoted in magnitudes, a logarithmic measure of the brightness of an object in a defined transmission curve[4]. The integration over the reference intensity in a given transmission curve is then related to a magnitude m via:

$$m = -2.5 \log \left(\frac{\int_0^\infty I(\lambda) T(\lambda) d\lambda}{\int_0^\infty I_0(\lambda) T(\lambda) d\lambda} \right) + m_0, \tag{1.1}$$

  where $I(\lambda)$ is the intensity received as a function of wavelength, $I_0(\lambda)$ the reference intensity that depends on the adopted photometric system and transmission curves, $T(\lambda)$ is the transmission efficiency for the filter as a function of wavelength, and m_0 is the reference magnitude, the magnitude for a flux through the filter equal to the reference flux. The magnitude scale is designed such that brighter objects have smaller numbers; an increase in magnitude by 2.5 means that an object is a factor of 10 fainter. The factor of -2.5 is present for historical reasons[5].

  When a magnitude is calculated using its luminosity distance (i.e., the apparent brightness of the source as viewed from Earth), the magnitude is called an apparent magnitude, m. Apparent magnitudes are used to quantify the brightness of sources at specific wavelengths and the measurement of apparent magnitudes

[4] Also called filters, passbands, or just bands.

[5] The magnitude system was invented in the 2nd century BCE by Hipparchus and is based on the fact that our naked eyes detect light with a logarithmic scale of intensity. Nowadays, it mostly serves to confuse people. Astronomers have been migrating towards more physical systems of measuring the intensity or luminosity of a source.

is known as photometry. The absolute magnitude M is defined as the apparent magnitude if the object were viewed at a distance of 10 parsec (32.6 light years). Placing all objects at a standard distance allows for a direct comparison between the absolute luminosities of the sources.

- **Color**. The difference in the magnitude measured between two wavelengths (i.e., passbands) is referred to as the color of the object. Younger sources have correspondingly bluer colors, resulting from the majority of the emission arising at bluer (shorter) wavelengths, compared to older, redder sources. If the passband at a shorter wavelength has a corresponding lower magnitude (a brighter object with more flux) than the magnitude measured for a longer wavelength passband, the color will be referred to as "blue".

- **Star formation rate (SFR)**. Quantifies the rate that the gas mass is converted into stars within a galaxy per unit time. There are numerous ways that this quantity can be estimated, most commonly with direct tracers such as stellar light, or indirect tracers, such as inferring the reprocessed light from interstellar dust by young stars. The conversion from light to SFR implies an assumption on the mass distribution of the stars, both detected and undetected components.

- **Stellar population age**. Refers to the luminosity-weighted average age of a stellar population. This age measurement is typically estimated from broad-band photometry (Section 1.3), though it can also be measured from specific spectral features (e.g., Hα emission of photospheric absorption lines) that serve as indicators of stars of a certain spectral type. Accurate measurements of the age of stellar populations is vital to this work. For galaxies, the mean stellar population age is closely related to the star formation history (SFH), which describes the time-dependent evolution of the SFR.

- **Dust Extinction and Color Excess E(B–V)**. Dust within a galaxy is efficient at attenuating light, with the strongest impact occurring in the UV (bluer wavelengths) and the attenuation efficiency decreasing towards longer (redder) wavelengths (Draine, 2003). For this reason, dust attenuation will tend to make spectral energy distributions (SED) appear artificially redder relative to the intrinsic SED of their stellar population.

  The effects of dust can be modeled as an extinction curve, which applies the modification to the spectrum of a galaxy as a function of wavelength. The color excess of the nebular gas emission, E(B–V), is used as a measurement of the amount of dust extinction. It corresponds to the change in color between the B and V passbands that arises as a result of dust, with values of E(B–V) = 0 corresponding to dust-free systems.

  To calculate the change in observed flux from the intrinsic flux, the following equation is usually used:

$$F_o(\lambda) = F_i(\lambda) 10^{-0.4 \text{ E(B-V) } k(\lambda)}, \tag{1.2}$$

  where $F_o(\lambda)$ is the observed (reddened) flux, $F_i(\lambda)$ is the intrinsic (unreddened) flux, E(B–V) is the color excess, and $k(\lambda)$ is the relative change in absorption due to dust as a function of wavelength.

1.3 Deriving Star Cluster Properties with Spectral Energy Distribution Fitting

Star formation occurs in the heart of molecular clouds, collapsing from gravitational fragmentation and forming hundreds to millions of stars that are gravitationally bound (McKee & Ostriker, 2007). Under the assumption that the formation of the stars occur at the same time, all stars within a cluster are treated as a single stel-

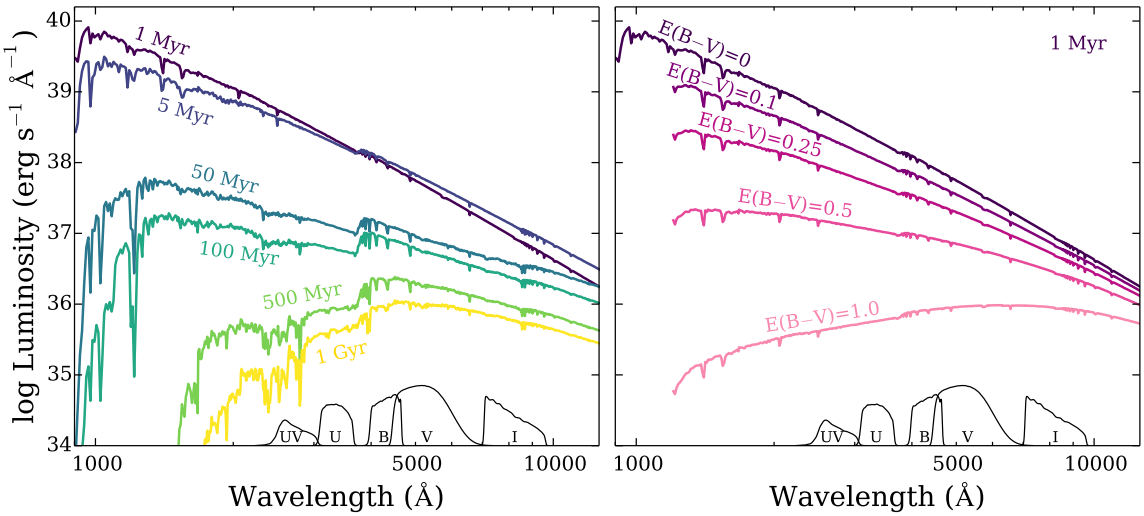

Figure 1.3. Spectral energy distributions between 1 Myr and 1 Gyr of various extinction. *Left*: The different colors show different SEDs for dust-free single stellar population (SSP) from Starburst99 models (Leitherer et al., 1999) without nebular emission. The strong dependency of the UV emission ($\lambda \lesssim 3000$ Å) as tracers of the massive, short lived stars that identify young stellar populations is evident from the steep decrease in the luminosity of these shorter wavelengths with increasing age. The HST transmission curves are shown with labels on the bottom axis. *Right*: The effect of dust on the SED of a 1 Myr SSP model assuming a starburst attenuation curve (Calzetti et al., 2000) with color excess E(B–V) values of 0 to 1. The degeneracy in SED shape between increasing age and dust (e.g., compare 50 Myr curve on left to E(B-V)=0.5 curve on right) necessitates the inclusion of the NUV and U bands to provide leverage in distinguishing these effects.

lar population (SSP). A SSP is the combination of all the light from the stars in the population and their initial mass distribution is parameterized by a chosen IMF. The resulting SED of the SSP will evolve with time as the stars themselves evolve (Figure 1.3).

The primary source of emission at UV (100 to 3000 Å) through near-infrared (NIR; 7000 Å to 1 mm) wavelengths arises from the stellar populations in a galaxy. Massive, short-lived stars produce a copious amount of ionizing photons and are the primary source of emission at UV wavelengths, and thus signify recent $\lesssim$100 Myr star formation activity. In comparison, less massive and longer lived stars dominate at

IR wavelengths and are reliable indicators of the total stellar mass (Kennicutt, 1998; Kennicutt & Evans, 2012).

The HST NUV and U bands, illustrated in Figure 1.3, are both blueward of the Balmer break at 4000 Å, a sensitive indicator of the age for young (<1 Gyr) stellar populations as the strength of the Balmer absorption lines is strongly age dependent (Bruzual, 1983). The NUV and U wavelengths are essential in order to help constrain the observed degeneracy between age and extinction (Anders et al., 2004) that causes a young, dusty stellar population (or galaxy) to mimic an older, dust-free population, as shown in Figure 1.3 (e.g., Witt et al., 1992).

To derive the properties (age, mass, and extinction E(B–V)) for a set of photometric data points, a grid of models is constructed and then convolved with the same broad band filters as the observations. It is imperative to account for the missing UV flux and correct for dust effects (Eq. 1.2) in order to accurately derive properties from SED fitting. A best-match is then found between the convolved model data and observations through minimizing the summed squares of the residuals to the model. Figure 1.4 shows an example of SED fitting performed for a galaxy.

1.4 The Hierarchical Patterns of Star Formation

The formation of a star is a seemingly minor, but important, event that occurs during the lifetime of a galaxy. The events of star formation never occur in isolation, but instead are a component in the complex network of the star formation process (Elmegreen, 2011). Star clusters form in patterns and groupings with each other in what we call hierarchical star formation; this "social" formation process leads to the assembly of structures composed of young stellar components in varying degrees of sizes and gravitational boundness, with the individual components serving as part of the larger grouping, connected in both space and time (see Elmegreen, 2010). This observed hierarchy in star formation has important implications in the evolution of

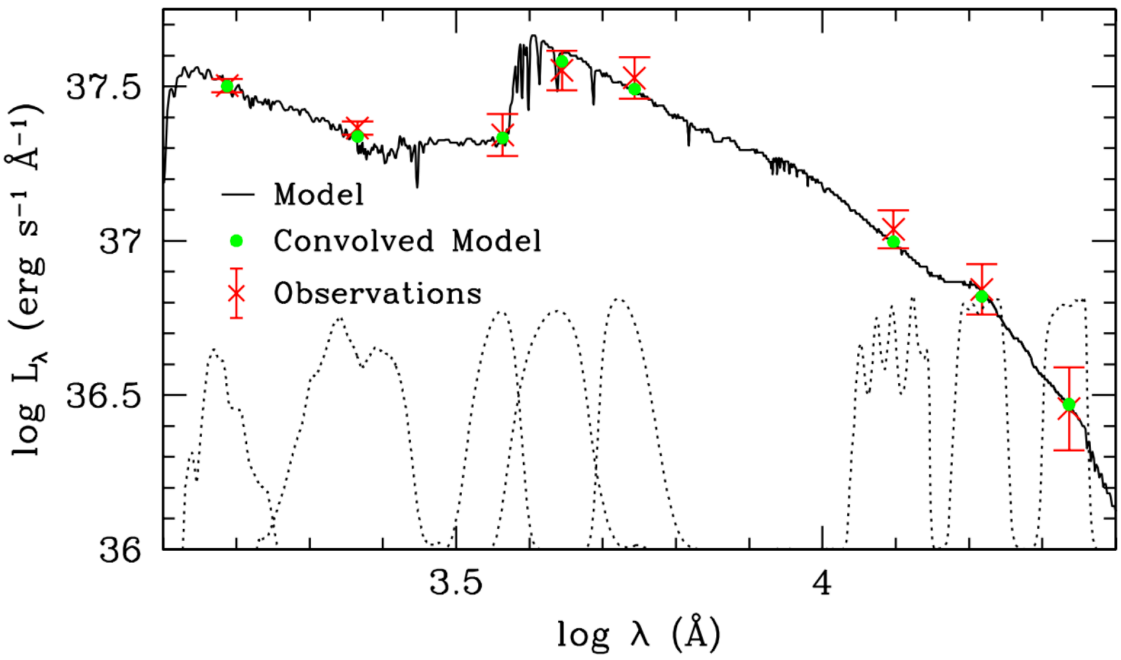

Figure 1.4. Best-fit Starburst99 SED for a galaxy. The wavelength coverage is from the FUV to the Ks bands. The black spectrum represents the synthetic Starburst99 spectrum, the red points represent the galaxy observations with accompanying 1σ errors, the eight dotted lines represent the filter transmission curves for the FUV, NUV, U, B, V, J, H, and Ks bands, and the solid green circles represent the convolved flux values of the Starburst99 curve with each transmission curve. Matching stellar models to the observed SED allows for estimation of physical properties of the galaxy. Image adapted from Grasha et al. (2013).

galaxy structure, the buildup of a galaxy's mass, and may have a role in determining the IMF itself (Larson, 1982).

The observed hierarchy is a result of stellar or gas kinematics with velocity dispersions that scale with a fractional power of the size of the region, which relates to the duration of star formation within a region of a given size (Figure 1.5). This is an expected result from fragmenting molecular clouds, and the resulting star clusters should reflect the same spatial distribution as inherited from their natal clouds. A schematic of how cluster formation and the formation of larger structures arises from a hierarchical star formation process is shown in Figure 1.6. The dispersion of unbound, large-scale hierarchical groupings of stellar aggregates occurs in a gradual fashion. Characterizing the parameters that drive the dissolution is the main scientific driver for this work and is presented in Chapter 2 and Chapter 3. In Chapter 4, we present a study aiming at characterizing the physical driver of the observed hierarchies.

1.5 The Molecular Interstellar Medium

The cores of giant molecular clouds (GMCs), extended and short-lived structures formed by molecular and neutral gas and dust (McKee & Ostriker, 2007), are the seeds of star formation. GMCs are constantly evolving objects and show a strong coupling with their environment (Krumholz et al., 2014; Klessen & Glover, 2016), revealing the important implications this has on the resulting star formation. At large scales, GMCs are spatially organized in complexes with filamentary and highly structured morphology (Figure 1.7), and are usually located in regions of high density, such as spiral arms (e.g., Grabelsky et al., 1987; Wilson et al., 2005; Goodman et al., 2014; Zucker et al., 2015) with the densest regions always clustered with other regions. As GMCs represent dense peaks within the distribution of gas in the ISM, the efficiency of star formation is enhanced within these clouds (Elmegreen, 2008),

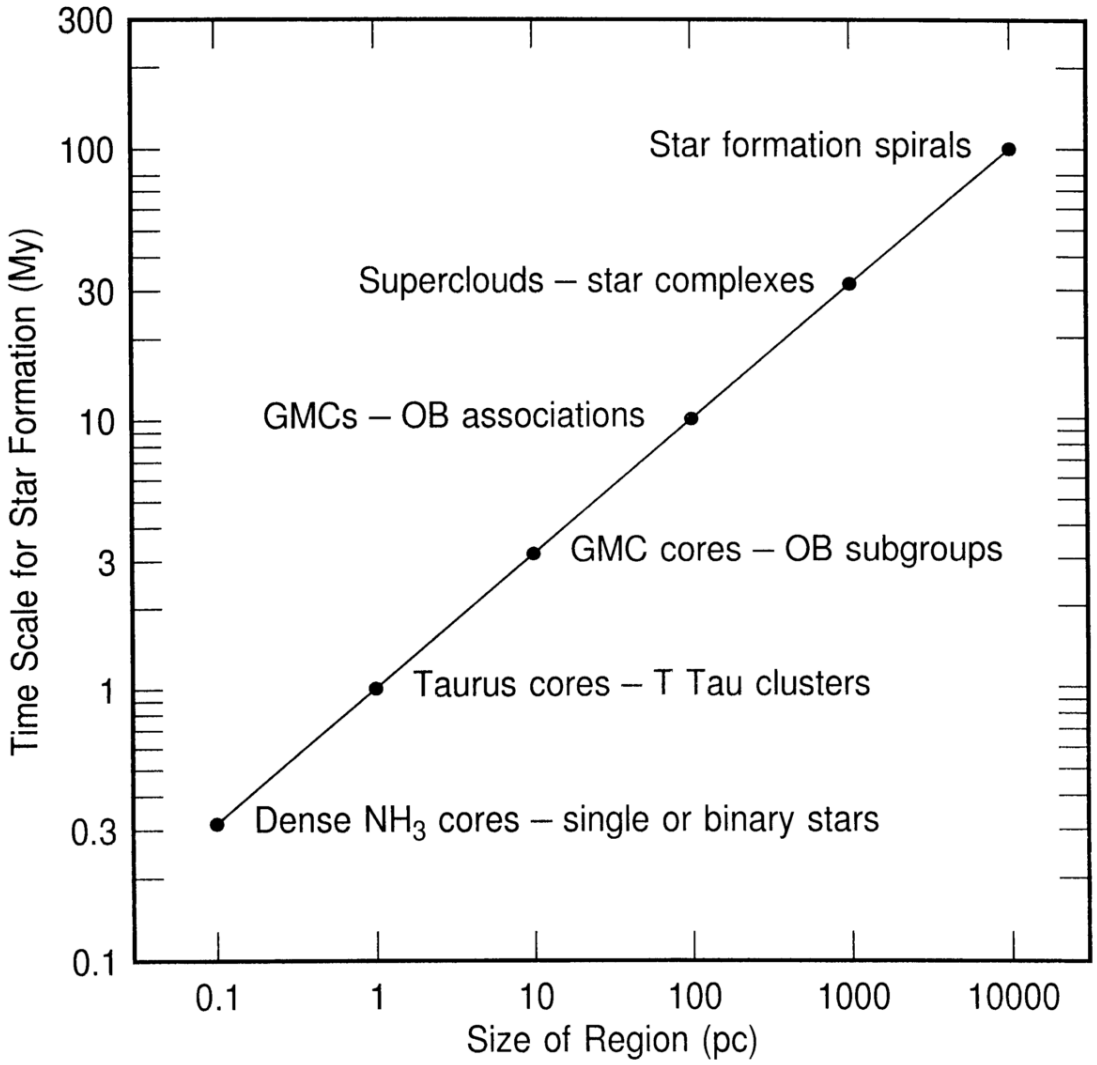

Figure 1.5. Size dependence of the duration of star formation in various regions. The image shows the relation expected from turbulence, with the timescale for star formation scaling as the square root of the size of the star-forming region ($t \propto \sqrt{R}$). Image adapted from Efremov & Elmegreen (1998).

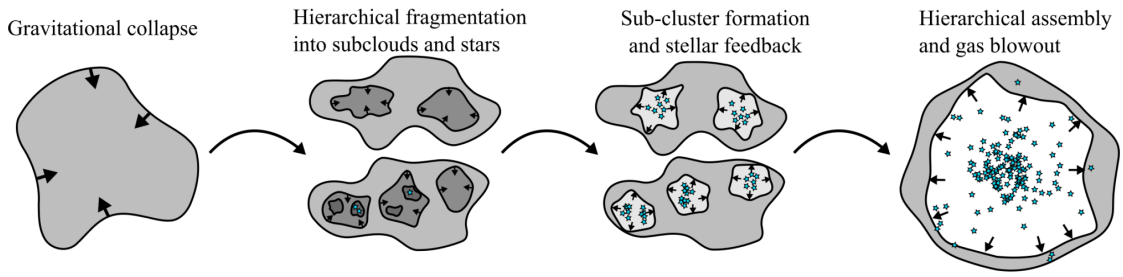

Figure 1.6. Model of cluster formation from hierarchical star formation. *Far left*: An unstable molecular cloud undergoes gravitational collapse. *Center left*: The gravitational instability causes hierarchical fragmentation, producing a hierarchy of subclouds that eventually fragment into individual stars. *Center right*: Stars that fragmented out of the same sub-clouds form in sub-clusters. Feedback from massive stars starts to evacuate gas locally. *Far right*: The sub-clusters merge hierarchically into a single cluster as stellar feedback blows out any remaining gas. Fragmenting molecular clouds result in the products of star formation – stars and star clusters – that are hierarchical (i.e., are correlated with each other in time and in space). Image adapted from Grudić et al. (2017).

giving rise to ensembles of star formation products that appear to be directly associated with GMCs. The efficiency of star formation is always highest in the densest region, and thus, the majority of star formation occurs within GMCs and these, as well as star clusters, will usually trace the spiral arms in the disk of spiral galaxies (see Figure 1.2). However, the overall efficiency of star formation process is low, with empirical measurements for Milky Way and extragalactic clouds being of order a few percent (e.g., Leroy et al., 2008; Murray, 2011; Usero et al., 2015).

Since the seminal work of Wilson et al. (1970), carbon monoxide (CO) emission is used as the primary tracer of molecular gas. It has a crucial role in the investigation of the physical foundation of star formation, tied to the process of galaxy evolution via the amount, distribution, and kinematics of the molecular gas[6]. Physical cloud

[6]Ideally, we would study molecular gas by studying the line emission from molecular hydrogen H_2. Unfortunately, the temperature within these clouds is too low to excite even the lowest accessible rotational transition of H_2, forcing us to rely on information provided by other, indirect, observational tracers. CO is the second most abundant molecular species within typical molecular clouds, and thus is a widely used tracer despite not being a perfect tracer of H_2 (Tielens & Hollenbach, 1985).

parameters, such as the mean density and turbulent Mach number, link star formation to molecular clouds, affecting the ability of a gas cloud to fragment and collapse to form stars and star clusters (Krumholz & McKee, 2005; Hennebelle& Chabrier, 2013). Thus star clusters present an approach to examine the link between star formation and gas physical conditions.

Gravity plays an indisputable role in shaping the structure of molecular clouds. Turbulence is also thought to play a fundamental in governing and shaping the overall properties of molecular clouds, gas dynamics, and the ISM as well (Mac Low & Klessen, 2004; McKee & Ostriker, 2007). Turbulence occurs when the inertial forces in a fluid are much greater than its viscosity, resulting in large-scale instabilities that cascade energy down to smaller spatial scales. Turbulent motions within molecular clouds are often greater than the sound speed c_s[7] in the medium, resulting in supersonic shocks that create regions of enhanced density. Supersonic turbulence within the ISM can result in density fluctuations over a wide range of spatial scales, and if the density enhancements are significant enough, they may be susceptible to gravitational collapse. However, unlike gravity, turbulent motions may not always be conducive to forming structures, acting to disrupt the density structure of the gas and suppress star formation. This complex interaction between turbulence and gravity in the ISM is fundamental in regulating its density structure, which in turn controls the star-forming potential of molecular clouds, and thus, plays a vital role in the overall efficiency and structure of star formation (Padoan & Nordlund, 2002, 2012; Krumholz & McKee, 2005; Federrath, 2013; Federrath & Klessen, 2013).

[7]The sound speed c_s is the speed that pressure disturbances travel through a medium. It depends on the pressure P and mass density ρ, defined as $c_s \equiv \partial P/\partial\rho$. For isothermal gas in a neutral medium, $c_s = \sqrt{k_b T/\mu} \approx 1$ km s^{-1}, where μ is the mean molecular weight and k_b is the Boltzmann constant which relates the average kinetic energy of particles in a gas with the temperature of the gas.

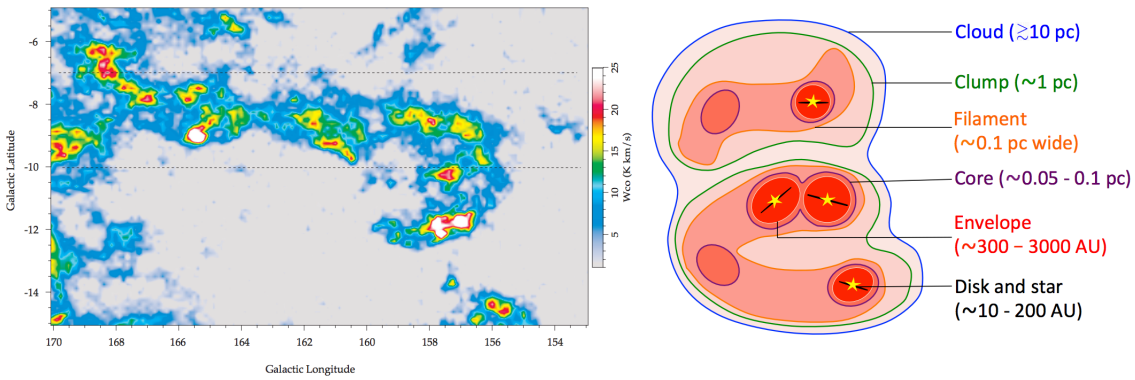

Figure 1.7. Molecular gas map of the integrated intensity of ^{12}CO emission toward the California cloud. The color scales show the velocity-integrated intensity of the CO(1-0) transition line, with star formation efficiency highest in the densest regions (red/white in color; left image). Image adapted from Lada et al. (2009). *Right*: A cartoon display of a molecular cloud showing hierarchical structures inside the cloud. The figure shows the cloud, clumps, filaments, cores, and envelopes, with typical size scales listed. The image is not drawn to scale. The scale-free nature of star clusters is believed to be inherited from the scale-free nature of the density clustering within the original cloud, with the clustering at the core-scale manifesting itself to create kpc-scale star formation structures. Image adapted from Pokhrel et al. (2018).

The time that a molecular cloud remains a coherent entity, capable of undergoing star formation, is a key factor in accurate descriptions of star formation and galaxy evolution. The cloud lifetime, how long a cloud can form stars, and how long a cloud remains associated with its stars, are still open questions related to star formation (Heyer & Dame, 2015). Whether or not the stellar correlations are inherited from the originating molecular clouds and the timescale for association between molecular clouds and star clusters are investigated in Chapter 5 and Chapter 6.

1.6 The Need for Space-Based Observations

The first known globular star cluster, M22, was discovered in 1665 by Abraham Ihle, a German amateur astronomer. However, it was not known that star clusters were composed of individual stars until Charles Messier observed M4 in 1764 with sufficient resolution to identify the individual stellar components (Boyd, 2008). Since

then, significant progress has been achieved in characterizing young star clusters and the process of star formation, also thanks to the launch of the Hubble Space Telescope (HST) in 1990. The combination of HSTs relatively large imaging field of view (2.7" x 2.7"), superb angular resolution (0.04"/pixel), high throughput at blue and ultraviolet (UV) wavelengths, and exquisite photometric stability is unrivaled by any other current telescope. HST observations provide the UV observations to accurately break the age/dust degeneracy (Figure 1.3) as well as sufficient resolution to distinguish star clusters from individual stars. HST has unarguably ushered in a new era in the field of star cluster research by enabling statistically large and reliable studies of young star clusters in nearby galaxies.

1.7 The Legacy ExtraGalactic Ultraviolet Survey

All of the work in this dissertation is drawn from the Legacy ExtraGalactic Ultraviolet Survey (LEGUS; Calzetti et al., 2015) a Cycle 21 HST program that imaged 50 nearby ($\sim$3.5–16 Mpc) galaxies with the UV and optical (UVIS) channel on the Wide Field Camera 3 (WFC3). Each pointing is covered by five broad band filters: NUV (F275W), U (F336W), B (F438W), V (F555W), and I (F814W), with the WFC3/UVIS. If already observed, we leveraged the existing observations taken with the Advanced Camera for Surveys (ACS) using the Wide Field Channel (WFC) as to not duplicate observations.

The set of five broad band filters (Figure 1.3) for the LEGUS observations is dictated by three necessities:

(1) Separate bright stars from faint star clusters.

(2) Derive accurate (time steps of $\delta\tau \lesssim 10$ Myr) star formation histories (SFHs) of stars.

(3) Obtain extinction-free ages and masses for clusters with age accuracy of 0.2 dex at intermediate ages, and mass accuracy of 0.3 dex.

The aim of the LEGUS survey is to investigate the relation between star formation and its galactic environment in nearby galaxies, over the range from individual star systems to galaxy-scale (a few kpc) structures. These data allow the study and investigation of the evolution of star formation within galaxies, both spatially and temporally, and is helping to paint a more accurate picture for theories of galaxy evolution, the physical underpinning of the gas-star formation relation, and the nature of star formation at high-redshift[8]. The work in this dissertation represents attempts in the effort to link local star-forming structures to the remarkably different and clumpier structures that are observed to be common in galaxies at high-redshift (e.g., Conselice et al., 2004; Elmegreen et al., 2007; Förster Schreiber et al., 2011; Guo et al., 2012, 2018; Soto et al., 2017).

[8]Redshift z is a unitless quantity that corresponds to the shifting of a spectral features due to the expansion of the universe. Redshift is measured from the difference in the observed and rest-frame wavelength, $z = (\lambda_{\rm obs} - \lambda_{\rm rest})/\lambda_{\rm rest}$. Redshift relates non-linearly to the age of the universe, where $z = 0$ describes the present day universe. All galaxies in this dissertation are considered "local" and are assumed to be at $z \sim 0$.

CHAPTER 2

THE SPATIAL DISTRIBUTION OF THE YOUNG STELLAR CLUSTERS IN THE STAR FORMING GALAXY NGC 628

This chapter[1] focuses on characterizing the function form of the correlation function and analyzing the errors. This will enable star clusters as a tracer for the hierarchy of star formation.

2.1 Abstract

We present a study of the spatial distribution of the stellar cluster populations in the star-forming galaxy NGC 628. Using Hubble Space Telescope broadband WFC3/UVIS UV and optical images from the Treasury Program LEGUS (Legacy ExtraGalactic UV Survey), we have identified 1392 potential young ($\lesssim$100 Myr) stellar clusters within the galaxy using a combination of visual inspection and automatic selection. We investigate the clustering of these young stellar clusters and quantify the strength and change of clustering strength with scale using the two-point correlation function. We also investigate how image boundary conditions and dust lanes affect the observed clustering. The distribution of the clusters is well fit by a broken power law with negative exponent α. We recover a weighted mean index of $\sim$–0.8 for all spatial scales below the break at 3."3 (158 pc at a distance of 9.9 Mpc) and an index of $\sim$–0.18 above 158 pc for the accumulation of all cluster types. The strength of

[1]These results are based on research published by Grasha et al. (2015), and is reproduced by permission of the AAS.

the clustering increases with decreasing age and clusters older than 40 Myr lose their clustered structure very rapidly and tend to be randomly distributed in this galaxy, whereas the mass of the star cluster has little effect on the clustering strength. This is consistent with results from other studies that the morphological hierarchy in stellar clustering resembles the same hierarchy as the turbulent interstellar medium.

2.2 Introduction

Star formation leads to the creation of stellar clusters (Lada & Lada, 2003) and most, if not all, stars form in some type of clustered structure. Observations of local star forming regions have shown that clustering is a common feature, resulting from the fractal properties of the interstellar medium (ISM) under the effects of turbulence (Elmegreen & Efremov, 1997; Elmegreen et al., 2014). As a result, gravitationally bound clusters occupy the smallest and densest regions of the hierarchy of giant molecular cloud (GMC) complexes forming on the larger scales ($\sim$1 kpc).

As clusters can be observed to greater distances than individual stars, young stellar clusters provide an excellent means to investigate the connection between the continuous distribution of star formation on small scales to galactic formation at large scales. While studies on the hierarchical clustering of stellar populations have become more sophisticated over time (e.g., see the early papers of Payne-Gaposchkin, 1974; Efremov, 1995), whether or not the stars and the clusters are mapping the same type of hierarchy (i.e., the dense peaks are randomly distributed within each hierarchy or they are a biased representation of each hierarchy) is still a question that needs to be answered. The answer will help understand the role of cluster formation and evolution that is complementary to studies of the cluster formation efficiency. The spatial distribution of newly formed stars is also important as it provides insight not only on the processes of star formation, but also on the evolution and environmental dependencies within stellar clusters and their migration from their

clustered complexes. For instance, we expect that the observed hierarchical clustering disappear with age (Elmegreen et al., 2006; Elmegreen, 2010): the densest regions with the shortest mixing timescales lose their substructures first, whereas the larger, unbound regions will lose their substructure over time owing to random initial motions and tidal forces. The migration timescale for which stars and clusters 'abandon' their clustered natal structure is not well constrained for most systems, but such knowledge is vital toward understanding how star formation evolves in both space and time.

One of the most powerful ways to probe the clustering distribution of galactic components is with the two-point correlation function (Peebles, 1980; Zhang et al., 2001; Odekon, 2006, 2008), a statistical tool to quantify the excess probability of finding one object within a specified distance of another object against that of a random, unclustered distribution. Applying correlation functions to study the clustering of clusters will provide insight to the (1) physical process of cluster formation; (2) the extent that formation of clusters is hierarchical; and (3) whether or not the clustered distribution of young stellar clusters reflects the fractal structure of the interstellar gas (Efremov & Elmegreen, 1998; Elmegreen & Efremov, 1996; Elmegreen & Elmegreen, 2001; Bastian et al., 2007).

Few studies have been done on the clustering distribution of stellar clusters thus far, showing the need for a systematic approach to the problem across a wide variety of galaxies and environments. In one of the first studies on the hierarchy occurring within stellar cluster ensembles, Zhang et al. (2001) found that a power-law correlation function well described the clustering hierarchy up to 1 kpc for the star clusters within the Antennae galaxies. Scheepmaker et al. (2009), in a study of the stellar clusters in M 51, found that clusters showed an age-dependency in their degree of clustering, where the youngest clusters were more clustered compared to older clusters.

Investigations of clustering behavior across time and scale lengths on stellar systems have been performed in a few local galaxies in recent years as well. In a study

of the stellar samples of M 31, Gouliermis et al. (2014) found that the youngest (<25 Myr) stars were more clustered than older stars (<300 Myr) and that the observed clustering changed within different regions of the galaxy. Studies of the stellar structures in both the Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC) have shown that stars are born highly fractal and evolve toward a uniform distribution within the crossing time of the galaxy ($\sim$ 175 and 80 Myr for the LMC and SMC, respectively; Gieles et al., 2008; Bastian et al., 2009; Bonatto & Bica, 2010). Within NGC 1313, Pellerin et al. (2007) found that infant mortality — the process by which most clusters do not survive the death of the most massive stars in the cluster, forcing the remaining unbound clusters to dissolve over a short period of time ($\lesssim$ 10 Myr: Lada & Lada, 2003; Fall et al., 2005; Bastian & Goodwin, 2006; Chandar et al., 2010) — is an efficient process in disrupting, and therefore, destroying stellar clusters by showing that B-stars can be found outside of star clusters. These unbound stars in the diffuse field can account for a large portion of the UV light from a galaxy (Meurer et al., 1995; Tremonti et al., 2001; Hoopes et al., 2001; Crocker et al., 2015) and the disruption and dissolution of clusters is crucial in populating the diffuse stellar field. The evolution of the clustering of star clusters with time will thus depend on the combination of two factors: how fast the cluster migrates within a galaxy and how long the typical star cluster survives.

If the cluster dispersion timescale has a dependency with galactic environment, we would expect to see their age distribution change over different ambient environments within a galaxy. Indeed, in a study by Sánchez-Gil et al. (2011), the young stellar populations within NGC 628 exhibited age gradients across the spiral arms, with the youngest star formation regions concentrated along the center of the spirals, and an age gradient from the inner to the outer parts of the galaxy. Radial stellar migration can further flatten the observed age gradient, implying steeper age gradients than those measured. Additionally, Bastian et al. (2011a) have found strong dependency

of clusters age with galaxy position, where the age distribution varies as a function of galactocentric distance, becoming shallower toward the outskirts of the galaxy. Theoretical work (Elmegreen & Hunter, 2010; Kruijssen et al., 2011) suggests that the age distribution of clusters do vary with environment and that it can influence the dissolution timescale, with clusters disrupting faster when the gas surface density $\Sigma_{\rm gas}$ is high and living longer when $\Sigma_{\rm gas}$ is low. On the other hand, if cluster dispersion is independent of environment, the age distribution of all clusters should be the same, regardless of position within a galaxy. Studies of both LMC and SMC have shown that there is a flat age distribution of clusters, concluding that cluster disruption does not significantly affect the age distribution and that there is little dependence on environment for the first ∼200 Myr (Gieles et al., 2007; Baumgardt et al., 2013). Early results of the Panchromatic Hubble Andromeda Treasury (PHAT; Dalcanton et al., 2012) survey show that the clusters of M 31 show no evidence for cluster dissolution at early times (30–100 Myr). Using the first PHAT clusters catalog of 601 clusters, Fouesneau et al. (2014) confirms that cluster disruption has little to no effect prior to the timescale of 100 Myr in Andromeda. The spatial correlation between clusters will allow us to study the extent of evolution in the clusters and whether or not the formation of clusters is hierarchical (Efremov & Elmegreen, 1998; Bastian et al., 2007).

Ambient environment is crucial for how clusters form, and quite possibly, it is even more critical to investigate if stellar clusters trace star formation. Studies on the clustering of stars have revealed important information on the nature of the star formation process itself. The recent results of Gouliermis et al. (2015a) show an environmental dependence in the clustering distribution of stars in the nearby spiral M 31, where stars in the outer regions of the galaxy experienced less disruption and larger amounts of clustering compared to stars located in the inner regions. This agrees with the work of Silva-Villa et al. (2014), which showed an environmental

dependence on the cluster population within M 83, where the clusters in the outer region showed less disruption and flatter age distributions compared to clusters in the inner region. Work by Gouliermis et al. (2015b) for NGC 6503 shows that younger stars follow a hierarchical distribution whereas older stars display a homogeneous, less clustered distribution. By comparing the clustering results within this work to the clustered results of stars, we will be able to infer if and how star clusters also trace star formation.

The nearby spiral galaxy in this study, NGC 628, was observed as part of the Legacy ExtraGalactic UV Survey[2] (LEGUS; Calzetti et al., 2015), a Cycle 21 Hubble Space Telescope (HST) program which has imaged 50 nearby (∼3.5–12 Mpc) galaxies in five UV and optical bands (NUV,U,B,V,I) with the UV/Visible (UVIS) channel on the Wide Field Camera 3 (WFC3). The aim of LEGUS is to investigate the relation between star formation and its galactic environment in nearby galaxies, over scales ranging from individual star systems to kpc-sized structures. NGC 628 is a face on grand design spiral galaxy with a large number of star clusters to provide a statistically powerful test for changes in clustering strength with scale. The occurrence of hierarchical structures within NGC 628 was already investigated by Elmegreen et al. (2006) using ACS data in B, V, I, and Hα bands. These authors examined the distributions of size and luminosity of star-forming regions, finding that both of these were well described with a power law, indicative of a hierarchical structure of stellar components within the galaxy.

The work herein primarily focuses on describing the two-point correlation function as a tool to study the clustering properties of the young stellar clusters. For this goal, we use the galaxy NGC 628 as an example. We hope to address the following points once the clustering properties of a larger number of LEGUS galaxies have

[2]https://legus.stsci.edu/

been studied: (1) whether stars clusters are clustered and how strongly; (2) how the clustered distribution of stars compare to that of star clusters; (3) how the clustering depends on age and environment and the dissolution timescale of the clusters; and (4) whether or not clusters can be used to trace the structure of star formation.

The galaxy selection and cluster identification process is described in Section 2.3. The methodology of two-point correlation function is introduced in Section 2.5. In Section 2.6, we describe the results and analysis and how we use the correlation function to draw conclusions about the properties of our star clusters. We discuss our results concerning hierarchy of the stellar clusters in Section 2.7. Finally, we summarize the findings of this chapter in Section 2.8.

2.3 NGC 628

We study the face-on grand design spiral galaxy NGC 628 located at a distance of 9.9 Mpc (Olivares et al., 2010) with no apparent bulge. NGC 628 is the largest galaxy of the M 74 galaxy group and has a global SFR(UV) of approximately 3.7 $M_{\odot}$ yr^{-1} (Lee et al., 2009). The stellar disk appears largely undisturbed in studies of the gas kinematics with optical and UV imaging (Herbert-Fort et al., 2010) and has a disk thickness of 0.25 kpc (Peng, 1988). We select this galaxy primarily due to the low-inclination angle (25 degrees), relatively large angular size (10'.5 x 9'.5), clearly defined spiral arms, and high number of clusters available for analysis.

The images of NGC 628 for our analysis were obtained in five broad band filters, NUV (F275W), U (F336W), B (F438W), V (F555W) and I (F814W) in October 2013 with the UV-optical channel (UVIS). Observations of NGC 628 consist of two pointings, one towards the center of the galaxy, NGC628c, and one towards the galaxy's eastern region, NGC628e, adjacent to the first pointing. All five filters in both pointings are aligned and referenced to the UVIS/F336W filter. Both pointings are combined into a single mosaic for analysis, shown in Figure 2.1. The images have

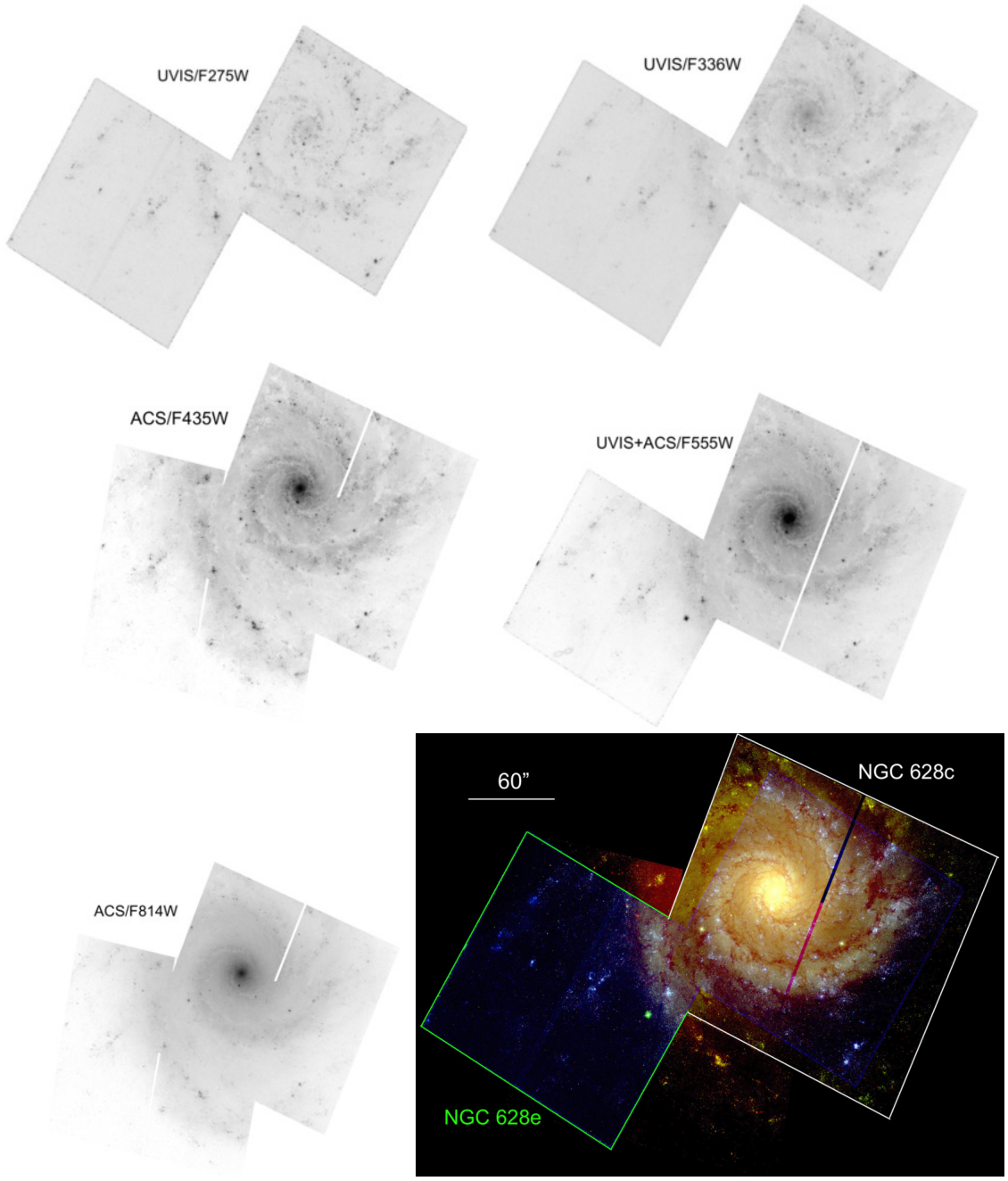

Figure 2.1. Mosaic images of the five broad band filters of NGC 628. All images are rotated with North up. The bottom right shows the RGB color composite mosaic UVIS/F275W and UVIS/F336W (blue), UVIS/F438W and UVIS/F555W (green), and UVIS/F814W (red). The white line outlines the central pointing, NGC 628c, and the green line outlines the east pointing, NGC 628e. The white horizontal bar in the upper left denotes the length scale of 60”, 3.3 kpc at the distance of NGC 628.

a scale of 0.04 arcsec per pixel, corresponding to a pixel resolution scale of 1.9 pc at a distance of 9.9 Mpc. A general description of the standard data reduction of the LEGUS datasets is available in Calzetti et al. (2015).

2.4 Cluster Selection and Identification

Stellar clusters within NGC 628 are identified by first undergoing an automated process using SExtractor (Bertin & Arnouts, 1996) that avoids background sources and single, bright stars. Each catalog includes sources which satisfy the two following conditions: (1) the V band concentration index (CI) must be greater than the stellar CI peak value (CI=1.4 and 1.3 for the inner and outer pointing, respectively, in order to make the selection criteria more uniform between the varying pixel scales of the cameras); and (2) the source should be detected in two contiguous bands with a signal-to-noise greater than 3.

As stars appear unresolved even at the highest HST resolution power, their CI will vary little and their CI distribution will be highly picked around an average value typical of a stellar PSF. On the other hand, clusters are partially resolved and their sizes can vary, therefore, they generally have CI larger than that of stellar values. Using CI distributions of the extracted cluster candidates, we selected the smallest CI value that would allow us to remove the bulk of the stellar interlopers from our catalog. Since the resolution power is also dependent of the HST camera used, the values used for the inner and outer frame of NGC 628 are different. Simulations of CI as function of effective radius of the clusters show us that the chosen limits corresponds to $R_{\rm eff} = 1$ pc at the distance of NGC 628.

In order to secure reliable measurements from SED fitting, each source is required to have a 3σ detection in at least four of the five photometric bands, necessary to adequately break the age-extinction degeneracy. The physical properties of each cluster (ages, extinctions, masses) have been derived using deterministic stellar population

models (Yggdrasil; Zackrisson et al., 2011) and a χ^2 fitting approach as described in Calzetti et al. (2015), which includes uncertainties estimates (see Adamo et al., 2010). In this paper we use standard catalogues which contains cluster physical properties derived with deterministic models with solar metallicity for both stars and gas, an average covering factor of 50%, and a Milky Way extinction curve (Cardelli et al., 1989) with a foreground E(B–V)= 0.06. Photometry is performed with a circular aperture of 4 pixels in radius, with the background measured within an annulus of 7 pixels in inner radius and 1 pixel in width. Single stellar populations are used to determine the ages of our clusters; for clusters that are better described with multiple stellar populations, our reported age is recovering the mean age of the cluster.

The uncertainties we derive in age, mass, and extinction are about 0.1 dex. Within the LEGUS collaboration we are also undertaking cluster analysis based on stochastically sampled stellar libraries and Bayesian approaches (Krumholz et al., 2015a). A comparison between the standard approach based on Yggdrasil deterministic models (Zackrisson et al., 2011) and the analysis performed with SLUG models (da Silva et al., 2012) shows that the average derived cluster properties do agree down to clusters masses of 5000 $M_{\odot}$. Below this mass range we observe the largest uncertainties in derived ages, masses, and extinctions because, as already widely discussed in the literature (e.g., Cerviño & Luridiana, 2004), deterministic models are not able to correctly interpret the large scatter in color caused by the stochastic sampling of the cluster IMF. However, in this analysis, we are mainly focusing on a morphological division of the cluster catalogues, based on visual inspection. We will also explore the change in clustering using two age ranges (i.e., 1 to 40 Myr and larger than 40 Myr) and two mass ranges (below and above log(M) = 3.6 $M_{\odot}$). A realistic error to associate to the value of 40 Myr is the average uncertainty of 0.1 dex observed in deterministic cluster properties; this corresponds to $40(+20/-10)$ Myr. The low mass range will

be affected by stochastic effects and indeed we will discuss this point in Section 2.6.3. More details will be given in a forthcoming paper (Adamo et al., 2017).

Each object within the cluster catalog that has an absolute magnitude brighter than −6 mag (well within the 90% completeness limits) in the V band is then visually inspected and assigned one of four categories: (1) a symmetric, centrally concentrated cluster; (2) a concentrated cluster with some degree of asymmetry; (3) a multiple peaked system; or (4) a spurious detection such as a foreground/background source, single bright star, bad pixel, or a source that lies too close to the edge of the chip. Class 4 objects are excluded from the final cluster catalog. Figure 2.2 shows what a typical cluster looks like for each classification. Classifications 1, 2, 3 are considered to be genuine star clusters or associations. The classification of each object is compiled from visual inspection from at least four independent members of the LEGUS team. The final cluster catalog is compiled by comparing all the results from each individual, after checking for consistency.

The distribution of masses and ages of the clusters is shown in Figure 2.3, with the three classes identified by different colors. In the same figure, the selection limit of $M_{F555W} = -6.0$ mag (apparent magnitude of 23.98 mag) for the clusters is shown by a continuous line, which follows the characteristic age-mass correlation for ages $> 10^7$ yr. General properties of the three cluster classes that will be detailed below include that class 3 clusters tend to be on average younger and less massive than Class 1 and Class 2 clusters.

Cluster candidates that meet the requirements of a minimum cutoff of CI > 1.4 (CI > 1.3 for the clusters in NGC 628e) and detection in at least four filters but are fainter than an absolute V-band magnitude of −6 are not visually classified; however, these unclassified sources are still added to the cluster catalogs, labeled as Class 0 cluster candidates. We are greatly interested in the faint cluster candidates in order to increase our sample, especially at the low-mass end (Krumholz et al., 2015a).

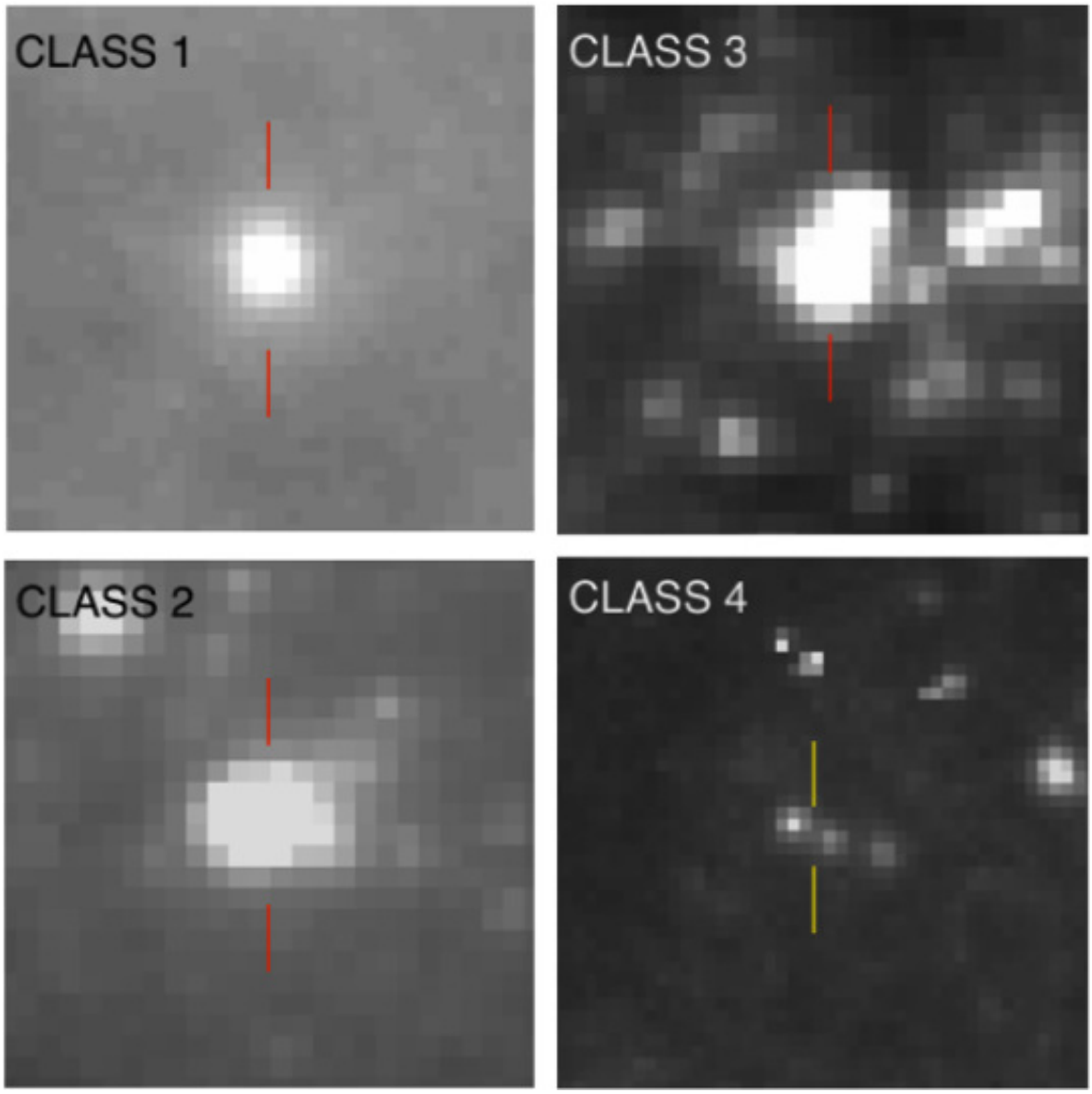

Figure 2.2. LEGUS classification system for the star clusters. An example is shown for Class 1 (symmetrical), 2 (asymmetrical), 3 (multiple peak), and 4 (star or spurious object) in the F555W filter.

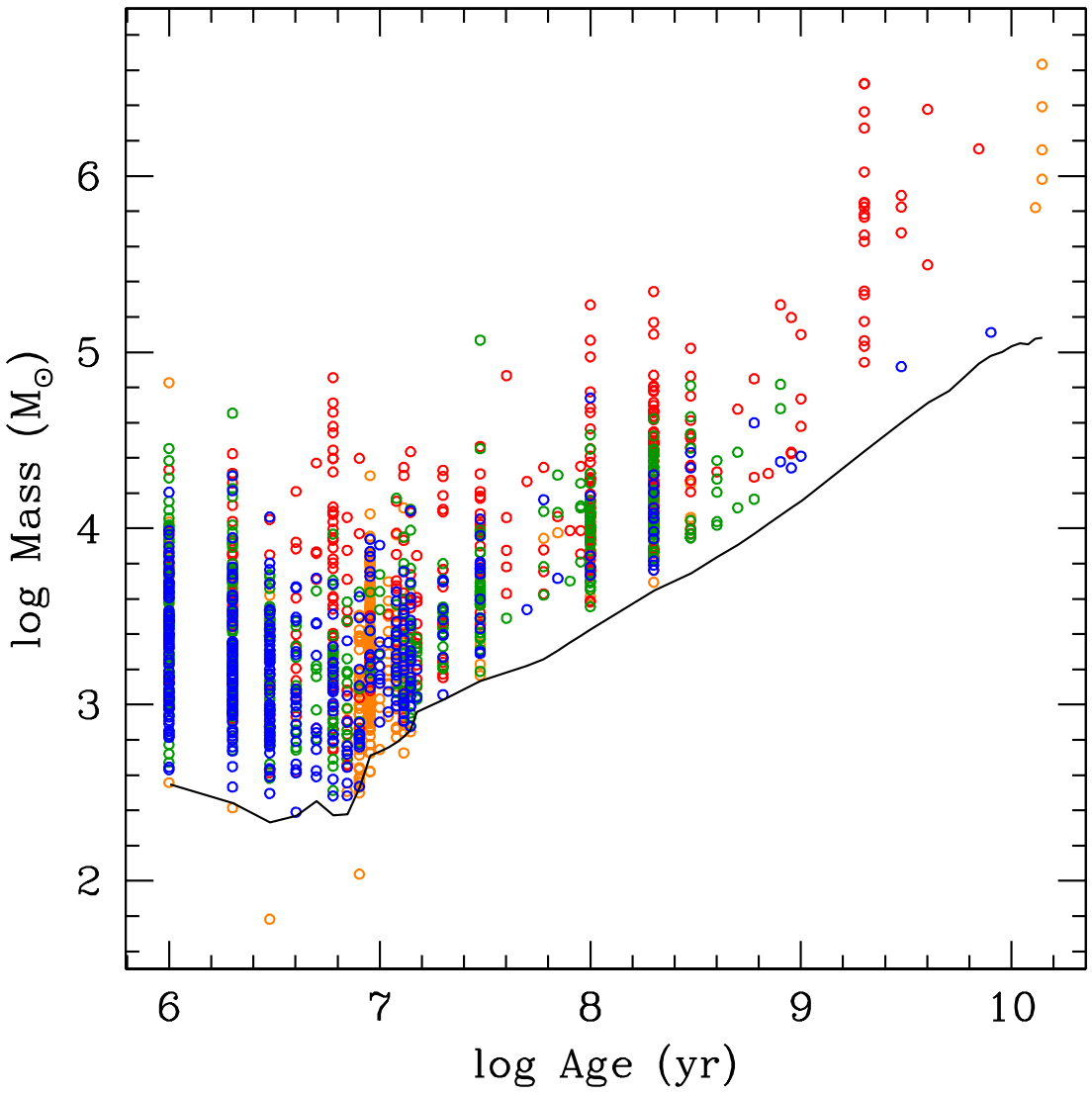

Figure 2.3. Age-mass diagram for star clusters in NGC 628. The colors represent our different cluster classifications: Class 1 (symmetrical, red circles), 2 (asymmetrical, green circles), 3 (multiple peak, blue circles), and 0 (selected as fainter than $M_{F555W} = -6.0$ mag, orange circles). The solid line is the 90% completeness limit.

Because these are not visually inspected sources, we acknowledge that there will be contamination of spurious objects in this class. We impose additional constraints on the Class 0 (not visually inspected) objects by analyzing those with CI $\leq$ 1.6 mag separately from those with CI $>$ 1.6 mag. This additional criterion of selecting the broadest of the Class 0 sources will increase our chance of identifying genuine clusters within the classification (see Section 2.6.2 for a more detailed discussion). For all Class 0 sources, we place a limit on the broadest CI of 4 to exclude background galaxies that may be included in the catalog. We have identified 345 Class 0 cluster candidates. For the 128 Class 0 candidates with CI $>$ 1.6 mag, we performed a visual check to quantify the accuracy of these objects being clusters. The results can be found in Appendix A, where we find that the contamination rate of non-clusters in the Class 0 is as high as 57%.

NGC 628c(e) has a total of 1019 (245) star clusters classified as Class 1, 2, or 3 and 345 (39) Class 0 cluster candidates with more than four filters available. We combine both pointings for our analysis, resulting in a total of 1264 genuine (Class 1, 2, and 3) and 384 Class 0 cluster candidates. There are 17 clusters in the overlapping region of the east and central pointings that are present in both catalogs; we have removed the duplicate clusters within the NGC 628e catalog. The breakdown of all the classifications in age, mass, CI, and color excess E(B–V) are shown in Figure 2.4. Figure 2.5 shows the breakdown of only Class 0 objects with a CI cut at 1.6 mag, where we expect the sources with the broadest CI indices to exclude all non-cluster objects. As expected, the clusters that are classified as multiple peak sources (Class 3) have the largest CI value on average. Class 1 (centrally concentrated) clusters have larger masses and older ages when compared to the rest of the clusters in the galaxy. The stellar clusters (Class 1, 2, 3) show small extinction, with an average color excess of E(B–V) $\sim$ 0.15. On the other hand, Class 0 candidates exhibit moderate

extinction, with an average of E(B–V) = 0.46, which remains roughly the same for all Class 0 clusters, independent of CI.

2.5 The Two-Point Correlation Function

Correlation functions provide statistical means by which to measure and provide detailed quantification on the clustering distributions of galaxy constituents. One of the most commonly used quantitative measure of clustering structure is the three-dimensional two-point correlation function, $\xi(r)$, which measures the magnitude of clustering as a function of scale size.

Originally defined by Peebles (1980) as $\mathrm{d}P = \bar{n}^2[1+\xi(r)]\ \mathrm{d}V_1\mathrm{d}V_2$ for cosmological applications, the two-point correlation function $\xi(r)$ is defined as a measure of the probability of finding a neighboring object, above what is expected for an unclustered random Poisson distribution, in a shell element with volume element $\mathrm{d}V$ at a distance r from any object with mean number density of the galactic structure, $\bar{n}$. The two-point correlation function, $\xi(r)$, is usually fit as a power law, $\xi(r) = (r/r_0)^{-\gamma}$, where r_0 is the characteristic scale-length of the clustering, defined as the scale at which $\xi(r) = 1$. The correlation function has been observed to have the same slope γ across galaxy and cluster systems (e.g., Peebles, 1975; Postman et al., 1998; Daddi et al., 2000; Brodwin et al., 2008); the uniformity of the measured indices seem to suggest a common underlying dynamics on all scales.

The spatial distribution of galactic components can also be measured in two-dimensions as projected onto the plane of the sky. In our study, we measure the two-dimensional projected angular correlation function $\omega(\theta)$, defined as the probability above Poisson of finding two star clusters with an angular separation θ as $\mathrm{d}P = N^2[1+\omega(\theta)]\ \mathrm{d}\Omega_1\mathrm{d}\Omega_2$, where N is the surface density of clusters per steradian with two infinitesimal elements of solid angle $\mathrm{d}\Omega_1$ and $\mathrm{d}\Omega_2$, separated by angle θ.

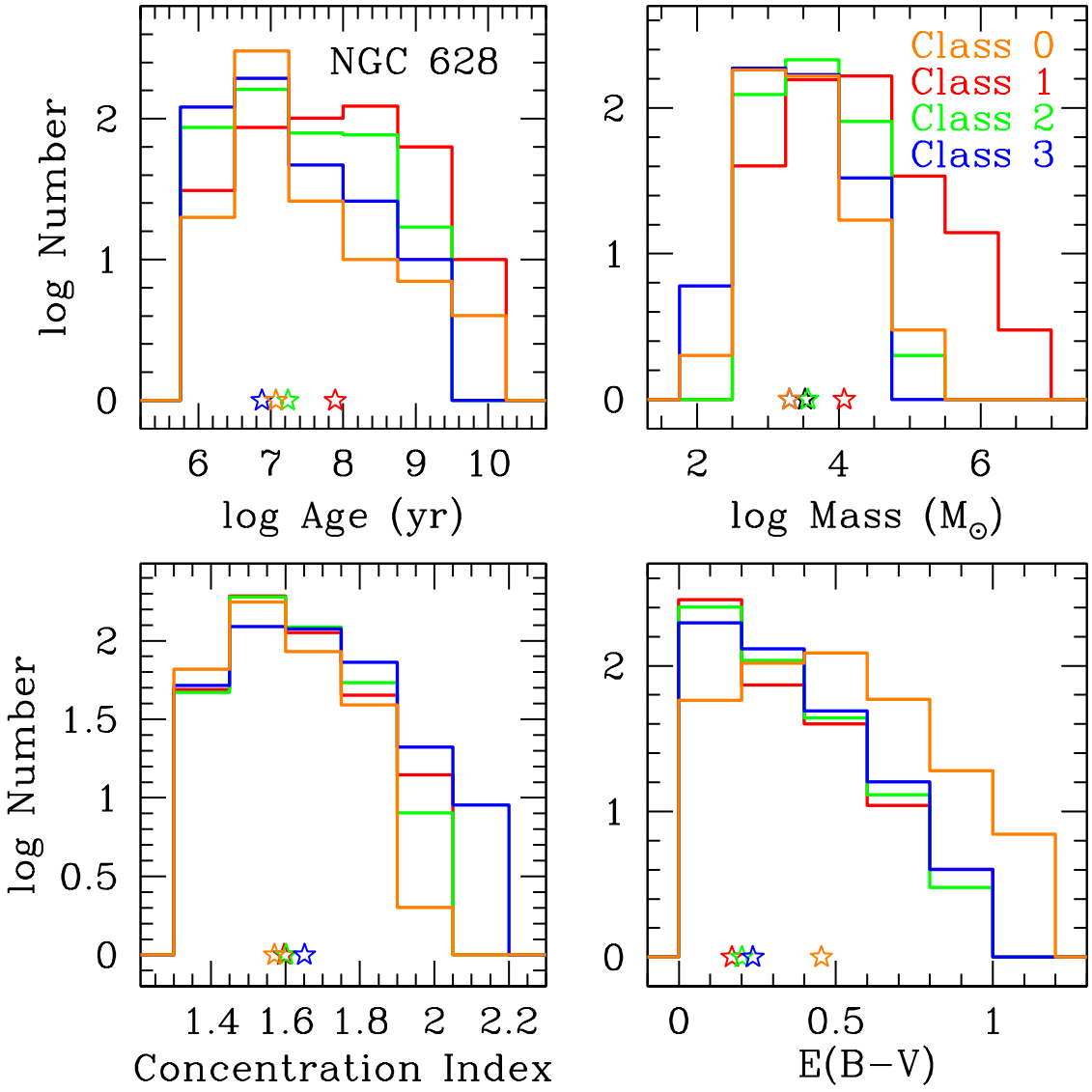

Figure 2.4. Histograms of age, mass, CI, and E(B–V) for the star clusters in NGC 628. The colors represent the cluster classification as defined in Section 2.4: Class 0 sources are orange, Class 1 are red, Class 2 are green, and Class 3 are blue. The distribution of E(B–V) for Class 0 sources is notably different from the genuine star clusters. The open star symbols at the bottom of each plot show the median value for each cluster classification.

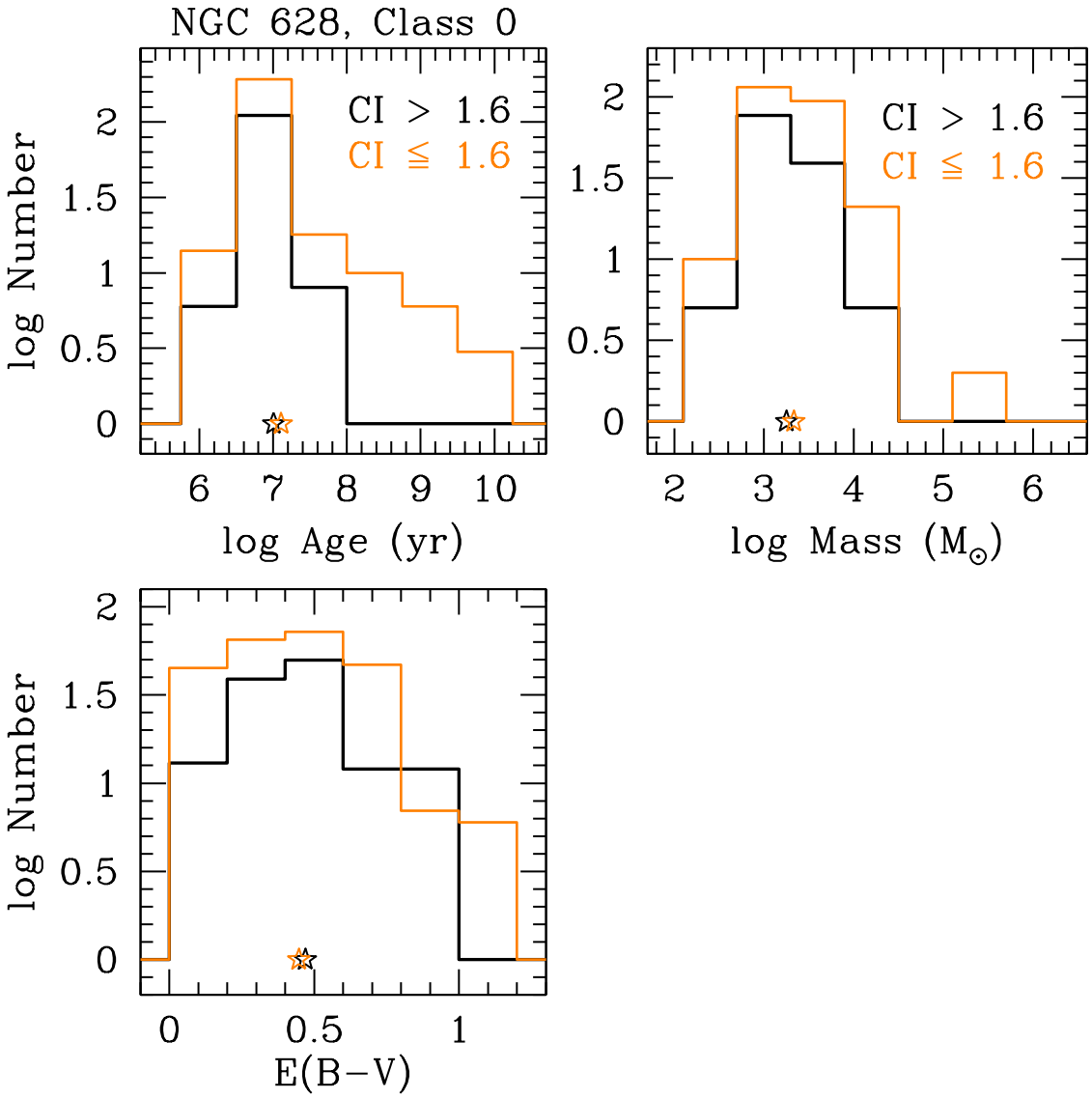

Figure 2.5. Histograms of age, mass, CI, and E(B–V) for the class 0 star cluster candidates in NGC 628. Class 0 clusters candidates with a CI $>$ 1.6 mag are in black and the clusters with CI $\leq$ 1.6 mag are shown in orange. The open star symbols at the bottom of each plot show the median value for each cluster classification. While the average value of mass or E(B–V) does not change between the sources with broader profiles compared to those with narrower profiles, there is an age difference in that CI$\leq$1.6 have a tail at larger ages.

To measure $\omega(\theta)$, pairs of stellar clusters are counted as a function of separation, compared by what is expected for an unclustered distribution. A random catalog (in x, y position) of sources must be produced, populating the same sky coverage and geometry (e.g., edges, masks) as the data with randomly distributed points. We define masks as areas that exclude all data, such as the ACS chip gap, or a reduction in the observed surface area of available data with respect to the global average, such as dust lanes or interarm regions of the galaxy. We account for masks in the random catalog that are present in the real data. The ratio of pairs of clusters observed in the data relative to pairs of points in the random catalog is then used to estimate $\omega(\theta)$. In this study, we will implement the Landy & Szalay (1993, LS) estimator, which cross-correlates the data and random catalog to minimize edge and mask effects, written as,

$$\omega_{\rm LS}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \tag{2.1}$$

where DD is the number of data-data pairs, DR is the number of cross-correlated data-random pairs, and RR is the number of random-random pairs with the same mean density and sampling geometry with separation between θ and $\theta + \delta\theta$. The pairs are computed as,

$$\begin{aligned} DD(\theta) &= \frac{P_{DD}(\theta)}{N(N-1)} \\ DR(\theta) &= \frac{P_{DR}(\theta)}{NN_R} \\ RR(\theta) &= \frac{P_{RR}(\theta)}{N_R(N_R-1)}, \end{aligned} \tag{2.2}$$

where N and N_R are the total number of data and random points in the survey volume, respectively and $P_{DD}(\theta)$, $P_{DR}(\theta)$, and $P_{RR}(\theta)$ represents the total pair counts in each separation $\theta \pm \delta\theta$ bin for the data-data, data-random, and random-random pairs, respectively. The size of the bin is determined by the sample size; we calculate our correlation function with eight spatial bins, spaced logarithmically between 0.16" (4 pixels) and 200" (5000 pixels), corresponding to spatial scales of 4.8 pc to 9.6 kpc,

the largest scale we can investigate with our data. Bin numbers were selected as a compromise between resolution and total number of clusters available to be sampled within each bin. When computed as above, a random distribution of an unclustered population will result in a flat correlation $\omega(\theta) = 0$, while a clustered distribution will have $\omega(\theta) > 0$. For hierarchical structuring, a general trend of decreasing $1 + \omega(\theta)$ with radius is expected (Gomez et al., 1993; Larson, 1995). This correlation function has also been successfully used for characterizing the clustering behavior of both stars and stellar clusters (e.g., Gomez et al., 1993; Zhang et al., 2001; Scheepmaker et al., 2009; Gouliermis et al., 2014, 2015b) in several nearby galaxies. Although the formulation in Eq 2.1 is optimized for taking into account edge effects, we still attempt to reproduce as closely as possible the geometry of the galaxy region sampled. As NGC 628 is nearly face-on, we do not need to take into account deprojecting the data before computing the correlation function.

It is also important that the random catalog must be large enough to not introduce Poisson error in the estimator. This is checked by ensuring that the RR pair counts in the smallest bin are high enough such that Poisson errors are negligible in the total error budget. As a result, the number of random points need to be greater than the data points. We achieve this for a random catalog that is 100 times greater than the size of the cluster sample. The uncertainties in the measurements of $\omega(\theta)$ in each radius bin are estimated as a Poisson error.

2.5.1 Application of the Clustering Estimator

Applying the two-point correlation function to our star clusters helps us to identify common age structures and derive the correlation length as a function of age and location. Figure 2.6 shows the correlation function for both pointings of NGC 628 as well as the mosaic. It is interesting to note that the Class 1 (centrally concentrated) clusters have a nearly flat relationship across all spatial distributions, while

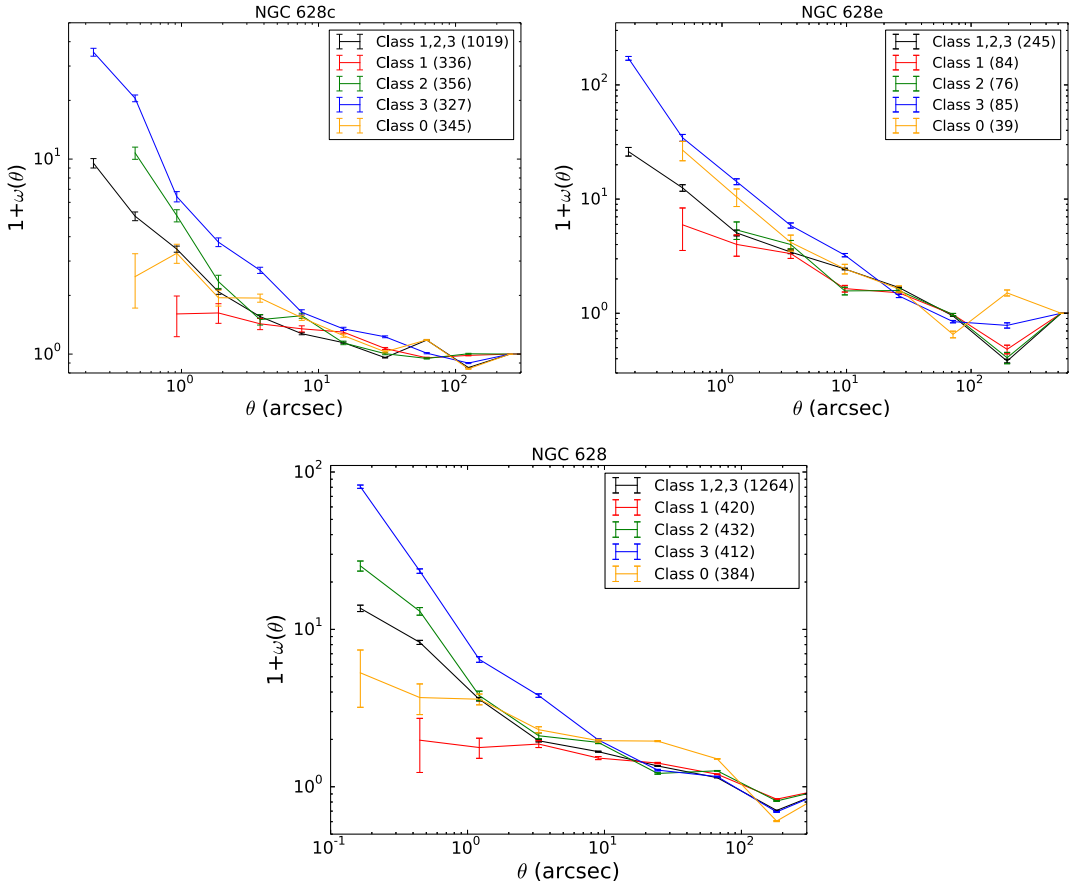

Figure 2.6. Two-point correlation function $1+\omega(\theta)$ for the clusters in NGC 628. The central pointing, NGC 628c (top left), east pointing, NGC 628e (top right), and the mosaic, NGC 628 (bottom) are all displayed a function of angular distance (arcsec). The colors represent the classification of each cluster, as defined in Section 2.4. The numbers in parentheses show the number of clusters in each classification.

the strongest clustering is seen in Class 3 (multiple peak) sources. The correlation for Class 0, which potentially may include also spurious sources, is flatter than that for our most structured cluster but still has a distribution that is more fractal (i.e., a steeper exponent) than class 1. We will discuss in Section 2.6.1 and Section 2.6.2 the differences between the correlations for the separate classes of clusters, and how we will attempt to single out the Class 0 clusters without visual inspection.

As mentioned above, the LS estimator is minimally sensitive to the random sample size and is effective at handling both edge and masking effects, as found in Landy &

Szalay (1993) and Kerscher et al. (2000). Using the central pointing, NGC 628c, to test how changing the geometry and masks within the random data sample is going to affect the results, we found that a random catalog of the same observing geometry as the real data with a random sample size of 10,000 objects already converges to the same correlation function values as a random catalog with 500,000 sources. As a sanity check, we create random fields with 1200 and 70,000 objects and compute the correlation function between the two catalogs and find that the computed $\omega(\theta)$ is zero across all angular scales.

We also verify that the LS correlation function is only minimally sensitive to the presence of masks within the data. After populating the random catalog for NGC 628c, we first remove the small chip gap (see Figure 2.7a) of several hundred pixels that runs across the length of the detector from the random catalog. In order to test how missing data alters the correlation function, we simulated dust lanes in the random catalog that coincide with the spiral arms in the galaxy where there is a dearth of observed star clusters. Adding anywhere between a single dust lane (about 6% of the total area) up to four dust lanes (13% of the total data area) have a negligible effect on the resulting correlation function, shown in Figure 2.7b. However, the clusters are already under sampled in the interarm regions and their removal along with the random objects in the same region, as expected, does not affect the results. Removing the random sources from these under sampled regions results in an expected decrease of the observed clustering at small scales as it removes the data-random cross correlation count pairs for the clusters that reside next to these regions but does not have other major effects.

To test the effect of masks in regions that are not under sampled, we take a random catalog with the same geometry as covered by the real data and remove regions from the random catalog that cover 10 or 20% of the total area within the galaxy and recompute the correlation function. We keep all the data and just remove

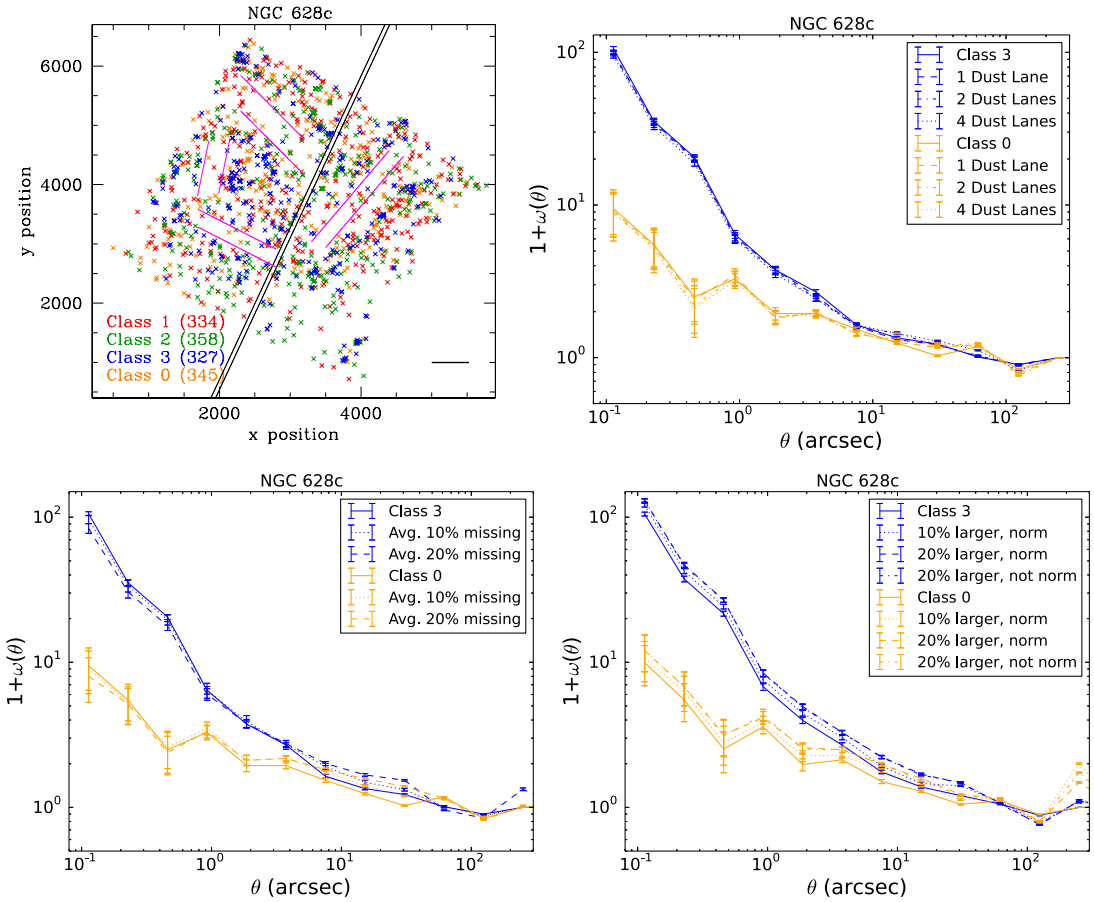

Figure 2.7. Testing the effect of geometry, random catalog size on the resulting correlation function for NGC 628c. Plot A (top left): Class 1 objects are shown in red, Class 2 are shown in green, Class 3 are shown in blue, and Class 0 are shown in orange. The numbers in parentheses show the number of clusters in each classification. The solid black line in the bottom right represents the spatial scale of 1 kpc at the distance of the galaxy. The two solid black lines denote the ACS chip gap and the magenta lines show the regions where we removed objects from the random catalog in order to simulate masking of dust lanes. Plot B (top right): Two-point correlation function $1 + \omega(\theta)$ when data are removed to simulate dust lanes. Removing an increasing amount of data does negligibly decrease the value of $\omega(\theta)$ at small radii. Plot C (bottom left): Two-point correlation function $1 + \omega(\theta)$ where we mask regions from the random catalog that corresponds to either 10 or 20% of the total area covered by the data. Shown is the correlation result from averaging the output from 10 realizations, each with random regions excluded from the galaxy. Plot D (bottom right): Two-point correlation function $1 + \omega(\theta)$ where we alter the coverage of the random catalog area to be 10% and 20% larger than the real data. In both cases, the number of sources in the random catalog is scaled accordingly to account for the increase in area coverage in order to maintain a constant surface density. When the random catalog is larger than the area covered by the data, the correlation strength is artificially increased.

the region from the random field. We perform this test 10 times, moving the region of missing data around the galaxy, and take the average resulting correlation function, shown in Figure 2.7c. For areas that are outside of the dust lanes (i.e., we are not selecting regions where the clusters are particularly under sampled), we observe a slight decrease in the clustering at small spatial scales and a transition toward increased clustering at larger spatial scales. As we are removing regions that are populated more than the dust lane regions, this can be best understood as an extreme example of removing data in regions that are not under sampled. The increase of clustering strength at the largest spatial scales is similar to what is observed when we artificially increase the random catalog size. These differences are not particularly significant from the statistical point of view.

We finally test the case of a random catalog with area that is 10% or 20% larger than that of the actual data (Figure 2.7d). For this test, we normalize the number of objects within the random catalog with the size of the random field geometry so that each field has the same surface density of random sources. In this case, we observe an increase in the strength of the clustering (higher values for the two-point correlation function). Likewise, a random catalog that has a smaller area than the coverage of the real data will result in a decrease in clustering across all scales. The effect of matching area and shape becomes especially important for galaxies that do not span the entire width of the chip and have to be described with ellipses for a subset of the chip coverage. While not applicable to NGC 628, it does affect other LEGUS galaxies in the survey.

2.6 Results and Analysis

2.6.1 Quantifying the Correlation Strength

The projected angular two-point correlation function (Eq. 2.1) is usually well described with a power-law,

$$\omega(\theta) = A_\omega \theta^\alpha, \tag{2.3}$$

where the slope α measures the strength of the clustering and the amplitude A measures the correlation length of the clustering; we use both to determine if the clustering is consistent with being scale-free. In a fully hierarchical model, the clusters will be correlated with other clusters from other groupings on larger scales, together forming a group higher up in the hierarchy of star formation. This would result in a smooth decline of $1 + \omega(\theta)$ in both radius and age. Figure 2.8 shows the two-point correlation function after combining both pointings in NGC 628. Figure 2.8 also shows each cluster class in a separate panel, and fit each correlation function with a broken power law (except for Class 0 and 1 which are better fit with a single power law). We compute the fit and determine the slopes using the Levenberg-Marquardt non-linear least square minimization fit, where the functional form is given as,

$$\log[1 + \omega(\theta)] = \begin{cases} A_1 + \alpha_1 \log(\theta) & : \log(\theta) < \beta \\ A_2 + (\alpha_1 - \alpha_2)\beta + \alpha_2 \log(\theta) & : \log(\theta) > \beta, \end{cases} \tag{2.4}$$

and where the breakpoint β is the logarithm of the position of the separation break along the x-axis (spatial scale), A_1 and A_2 are the clustering amplitudes before and after the break, and α_1 and α_2 are the slopes to the power law before and after the breakpoint, respectively. Slopes, amplitudes, and break-points are free parameters in the fit. The fitted parameter results for NGC 628 are listed in Table 2.1.

For a hierarchical (scale-free or self-similar) distribution, the total number of clusters N within an aperture increase with radius r as $N \propto r^{\alpha+2}$, where α is related to the (two-dimensional or projected) correlation fractal dimension D2 $= \alpha + 2$ (e.g., see Falgarone et al., 1991). The distribution of star formation and interstellar gas over a large range of environments is observed to show a (three-dimensional) fractal dimension of D3 ~ 2.3 (e.g., Elmegreen & Efremov, 1996; Elmegreen & Elmegreen, 2001), in agreement with the predicted fractal dimension for the density structure of

a turbulent ISM (Federrath et al., 2009). Projection of this three-dimensional fractal on a plane will result in a two-dimensional fractal dimension D2 = D3 – 1 if the perimeter–area dimension of a projected 3D structure is the same as the perimeter–area dimension of a slice of the structure (Elmegreen & Scalo, 2004). The two-point correlation function results are qualitatively identical to a power spectra analysis. However, as shown in simulations by Gouliermis et al. (2014), the D2/D3 conversion is not trivial.

As can be seen in the power law fits to NGC 628 in Figure 2.8, the power law index recovered ranges from $\alpha = -0.14$ to -1.51, covering spatial scales from 5 pc to nearly 5 kpc. If we take the weighted mean of the measured indices before the break point at 3".3, we recover an average slope of $\alpha \sim -0.8$, in agreement with what is expected from a hierarchical distribution of gas. After the break point, we observe a dramatic decrease in the fractal distribution, informing us that the spatial distribution of the clusters (excluding Class 1 and 0 sources) is systematically less clustered at larger spatial length-scales. Results of a study of the hierarchical structure of star formation within a subset of 12 LEGUS galaxies by Elmegreen et al. (2014), suggests that hierarchically structured star-forming regions are common unit structures, which can be several hundred parsecs, but also that observed self-similarity of young stellar structures down to parsec scales is indicative that individual star clusters form stellar groupings on larger scales.

Given the 2'.7 FOV of WFC3/UVIS, the largest physical size that can be probed for a single pointing in NGC 628, is 7.8 kpc. A pixel scale corresponds to a physical size of 1.90 pc at the distance of 9.9 Mpc, with a typical star cluster around 5–10 pc in diameter. Each of the classes is best-fit with a broken power law at 3".3, corresponding to a spatial scale length of 158 pc at a distance of 9.9 Mpc, except for Class 1 clusters, best-fit with a single power law. The break in the power law corresponds to the observed line-of-sight thickness of the galactic disk (see Section 2.6.1.2 and

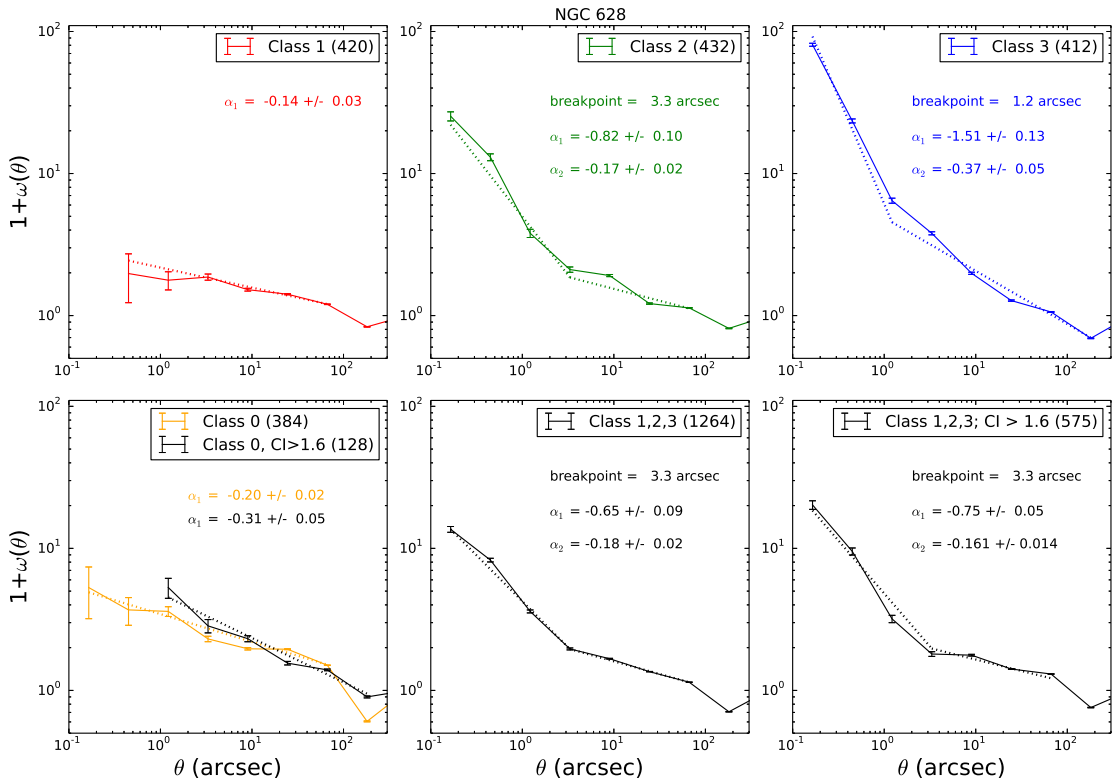

Figure 2.8. Two-point correlation function $1 + \omega(\theta)$ for the clusters of NGC 628 a separated into the individual classes. The fitted broken power laws, given as α_1 and α_2 along with uncertainties to the power law slope, are listed in each panel. The bottom left shows all the Class 0 sources (yellow) and only those limited to CI > 1.6 (black). We also report the power law fit to the correlation function for Class 1, 2, and 3 clusters (bottom panel, middle) and all of the clusters with CI > 1.6 (bottom right plot).

Table 2.1. Two-Point Correlation Function Power Law Parameters of NGC 628

Class	Number	A_1	α_1	β (arcsec)	A_2	α_2
Class 1	420	2.24(3)	−0.14(3)	...	...	...
Class 2	432	5.31(3)	−0.82(10)	3.3	2.76(2)	−0.17(2)
Class 3	412	8.17(7)	−1.51(13)	1.2	3.4(2)	−0.37(5)
Class 0	384	3.44(5)	−0.20(2)	...	...	...
Class 0; CI> 1.6	128	4.8(1.2)	−0.31(5)	...	...	...
Class 1, 2, 3	1264	3.91(19)	−0.65(9)	3.3	2.29(2)	−0.18(2)
Class 1, 2, 3; CI> 1.6	575	3.9(3)	−0.75(5)	3.3	2.26(2)	−0.161(14)

Notes. Columns list the (1) Classification of stellar clusters – Class 1: symmetrical clusters; Class 2: asymmetrical clusters; Class 3: multiple peak clusters; and Class 0: non-visually identified clusters, (2) Number in each classification, (3) Amplitude A_1 of the angular correlation function before the breakpoint, (4) Slope α_1 of the angular correlation function after the breakpoint, (5) Location of the break point β. Cluster classifications that are best-fit with a single power law do not have a breakpoint, (6) Amplitude A_2 of the angular correlation function past the breakpoint, and (7) Slope α_2 of the angular correlation function after the breakpoint. Numbers in parentheses indicate uncertainties in the final digit(s) of listed quantities, when available.

Section 2.7 for discussion). Class 3 clusters exhibit the break below the rest of the clusters at 1".2 in addition to exhibiting a much steeper slope across all scale lengths. The exact location of the breakpoint naturally depends on the number of bins used to calculate the correlation function (See Section 2.4). Past the 3".3 break, all the classes show a correlation function with a power law index of $\alpha \sim -0.2$, indicating that any occurring clustering has considerably weakened by this length scale, lasting until a few kpc. The clustering observed increases below this breakpoint, albeit to varying degrees between the cluster classifications. These results indicate that star formation is clumpy and the stellar clusters form highly clustered distributions for spatial separations of r less than the break point. For spatial scales above the breakpoint, the clustering observed becomes more homogeneous and less clustered.

In order to address how clusters behave versus associations, we compare Class 1+2 (bound) clusters to just Class 3 clusters in Figure 2.9. We believe that Class

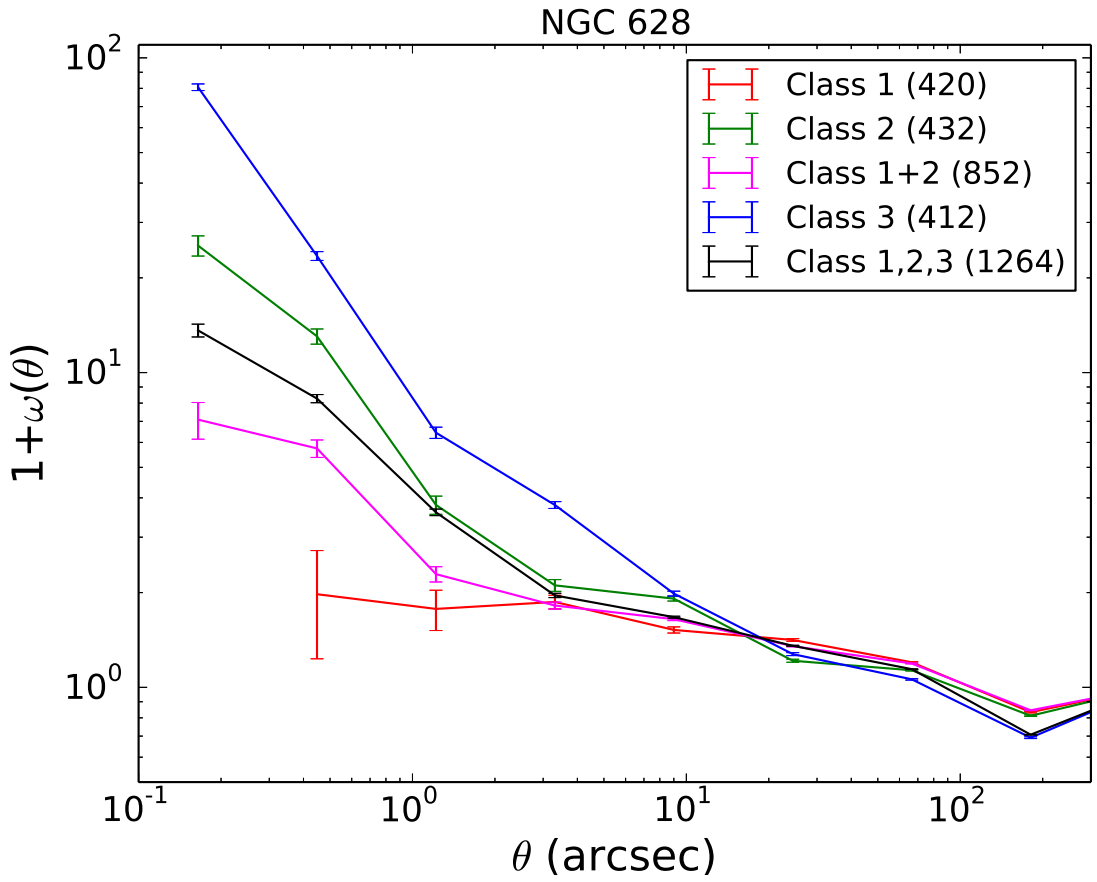

Figure 2.9. Two-point correlation function $1+\omega(\theta)$ for the bound cluster classifications compared to the global behavior in NGC 628. Class 1 and 2 (bound) are shown in magenta and compared against the combination of all the Class 1, 2, and 3 (black) clusters.

1 and 2 are gravitationally bound clusters in different states of relaxation whereas Class 3 (multiple peak) clusters quite possibly be unbound associations. We see that the clustering behavior of the Class 1+2 is different than what is observed for just the Class 3 clusters alone. These results indicate that bound clusters (class 1 and 2) have a correlation behavior that is weaker, corresponding to a different spatial distribution when compared to what we observe for Class 3 and the compilation of Class 1, 2, and 3 clusters combined. This strengthens our results that as (bound) clusters relax, they become less clustered and behave differently than what is observed for associations.

2.6.1.1 Break in the Power Law

Excluding Class 0 and 1 sources, distributions are described with a broken power law with the break point occurring at 158 pc averaged between all the genuine clusters (class 1, 2, and 3), though Class 3 clusters lose their self-correlation at much shorter scale lengths – 58 pc. In fact, we may expect a smaller characteristic scale for the younger clusters and a larger one for older clusters, if clusters and stars represent the same hierarchy (Efremov, 1995). Class 1 clusters are the least self-correlated of all the cluster morphologies with a nearly flat exponent, implying that the distribution of these clusters is nearly uniform across all scales. The flattening of the exponent found after the break point for classes 2 and 3 is comparable to the uniform distribution of Class 1, implying the loss of self-correlation distribution for these types of clusters after the break point at the larger scale. For the cumulative correlation function of class 1, 2, and 3, we still see strong self-correlation with a break point below 158 pc, suggesting that the Class 2 and 3 clusters are dominating the small-scale correlation.

A similar break in power spectra within other galaxies has been observed and interpreted as due to the line of sight thickness of galactic disks (Lazarian & Pogosyan, 2000; Elmegreen et al., 2001; Odekon, 2008; Block et al., 2010) when the line-of-sight depth is smaller than the transverse size. A study of the brightest stars in the LMC, M 31, and M 33 by Odekon (2008) identified a transition in the correlation function to a higher correlation dimension (weaker clustering), with the transition marking the large-scale regime where disk geometry and dynamics set the scale for structure. This transition from the small-scale three-dimensional turbulence to large-scale two-dimensional disturbances in the disk, observable with a transition to a more shallow power law at the transition, has been detected for the HI gas in the LMC with a transition occurring at 180 pc by Padoan et al. (2001); Elmegreen et al. (2001) and 290 pc by Kim & Park (2007). We may very well be seeing the analogous transition within our young stellar populations.

2.6.1.2 Dependency of Results on Bin Numbers

The choice of annuli width used to calculate the correlation function in Eq. 2.1 will naturally influence the results and the exact location of the breakpoint of the correlation behavior. In order to ensure adequate sampling of clusters in each bin across the scale length that we are investigating, our bin numbers were selected such that the smallest angular bin contained a minimum of two clusters, as the smallest separation size is where our data are the sparsest. Figure 2.10 shows how changing the bin size from five bins to 20 bins moves the location of the breakpoint, averaged over Class 1, 2, and 3 clusters, between 2 to 14 arseconds (96 pc to 672 pc, respectively). Decreasing the bin numbers aids to smooth out occurring variations at both small and large angular separation. We can see that the break point only starts to disappear at the coarsest sampling of 5 bins, highlighting that galaxies with a large number of clusters available are needed to provide statistically sound results. In order to see the breakpoint at the disk thickness, we need to be able to resolve the equivalent scale.

2.6.2 Age and Concentration Index

The correlation function of Class 0 objects tends to be very shallow, confirming that Class 0 objects contain a compilation of every possible object available. In order to attempt to identify only the star clusters within Class 0 objects, we take very conservative cuts in CI and then further divide into age bins. The youngest objects that are genuine star clusters should exhibit an increase the strength of the power-law in the correlation function. We see this in the correlation function for Class 3 clusters, which on average are younger and more clustered compared to Class 1 clusters. In order to test if the correlation does change with cluster age, we divide our clusters by age, starting at 10 Myr, and recompute the correlation function. Figure 2.11 shows the correlation function for all cluster classifications as well as only the clusters within each classification that are younger than 40 Myr, which is the age where we

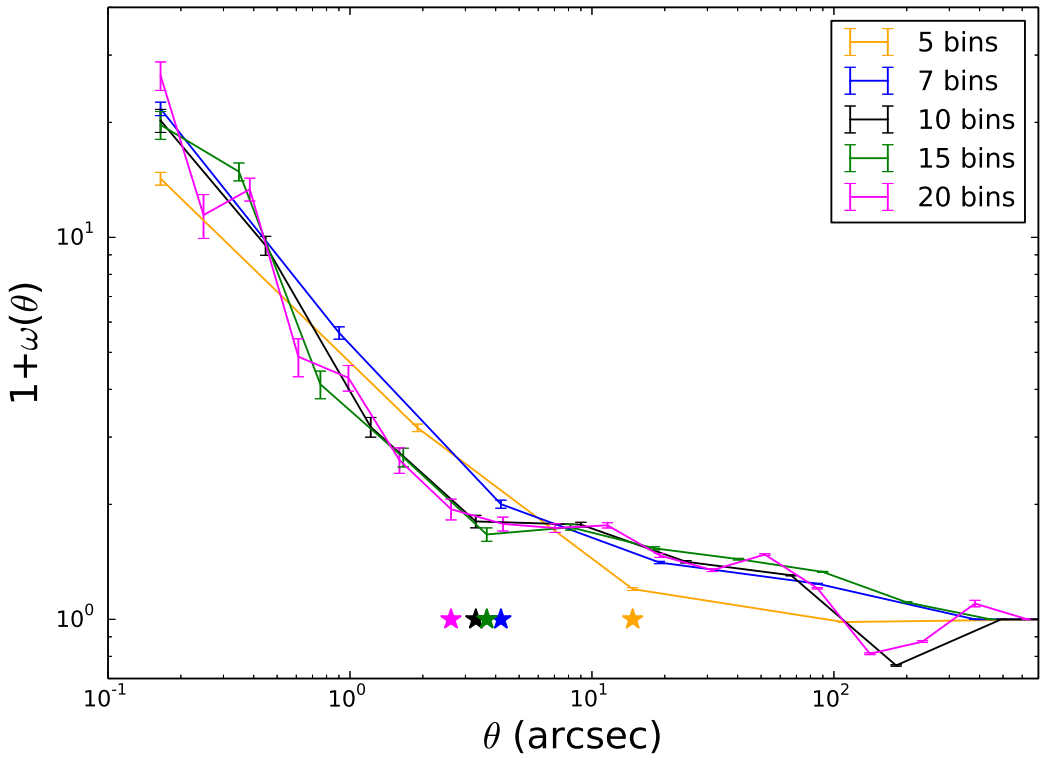

Figure 2.10. Correlation function computed using different widths of annuli. We vary the spatial scale from 5 to 20 bins to inspect the impact of the size of annuli on the breakpoint. The star symbols show the location of the breakpoint for each result.

see the most noticeable transition from clustered to non-clustered. Excluding Class 0 and 1 clusters, the correlation function does increase when we only consider the youngest clusters in each class. Separating the broadest sources (CI $>$ 1.6 mag) for the combination of the Class 1, 2, and 3 sources from their narrower counterparts (far right, bottom panel of Figure 2.11) only marginally increases the clustering strength. The shallowness of the slope for the Class 1 sources is not entirely surprising as these are the oldest and most massive clusters in the galaxy, where we possibly are viewing a subset of clusters that have dispersed and have lost their clustered, non-homogeneous distribution. Class 3 sources, which are on average the youngest in the galaxy, are systematically the most clustered class observed.

As can be seen in our analysis, at 20 My we see a transition at which the clustering structure disappears and the clusters display a more homogeneous, non-clustered distribution, in place by ages of 40 Myr. However, the clustering strength of Class 3 clusters are only minimally effected with an increase in age. All star clusters regardless of their classification tend to be randomly distributed above this age, where random motions and shear effects can explain the observed data trends. The flatness of the correlation function of the age distribution for Class 1 clusters as compared to Class 3 clusters can be evidence of cluster dissolution as we see a dramatic decrease in the total strength of the clustering with increasing age.

We also show the Class 0 sources in Figure 2.11 divided into concentration index bins above and below CI of 1.6 mag, where the broader sources (CI $>$ 1.6 mag) do show evidence of increased clustering. We do not show the Class 0 sources for ages older than 40 Myr as there is a dearth of available data. When we combine the youngest and broadest sources, limiting our selection to our very conservative cut of CI $\geq$ 1.6 mag and divided into extremely young ($\leq$ 10 Myr) and young ($\leq$ 40 Myr) age bins, we do not see a difference in the strength of the clustering. However, out of

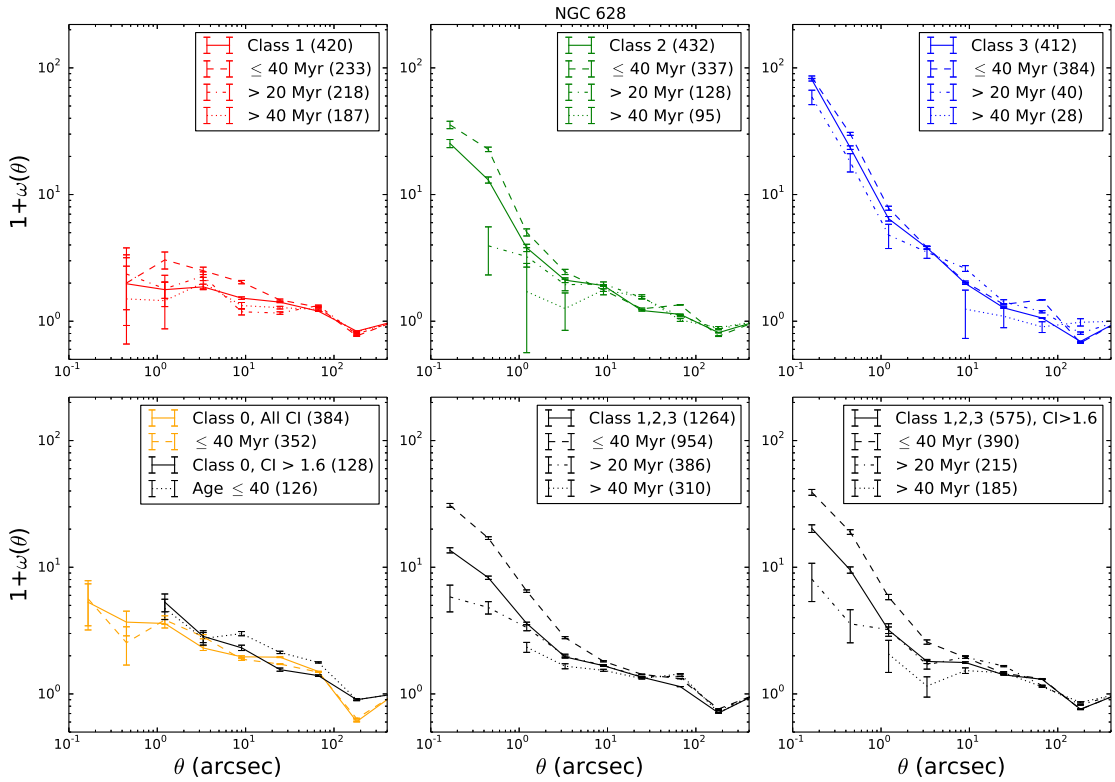

Figure 2.11. Two-point correlation function $1+\omega(\theta)$ in NGC 628 divided by cluster types and by cluster age. The strength of clustering increases when we only consider the youngest clusters within each classification. For Class 0 sources, we also show how the CI (compact vs broad) values influence the clustering, where the broad Class 0 sources (CI $>$ 1.6 mag) appear to have a slight increase in the clustering strength. Diving the genuine clusters (i.e., Class 1, 2, and 3) by CI value does not change the amount of clustering.

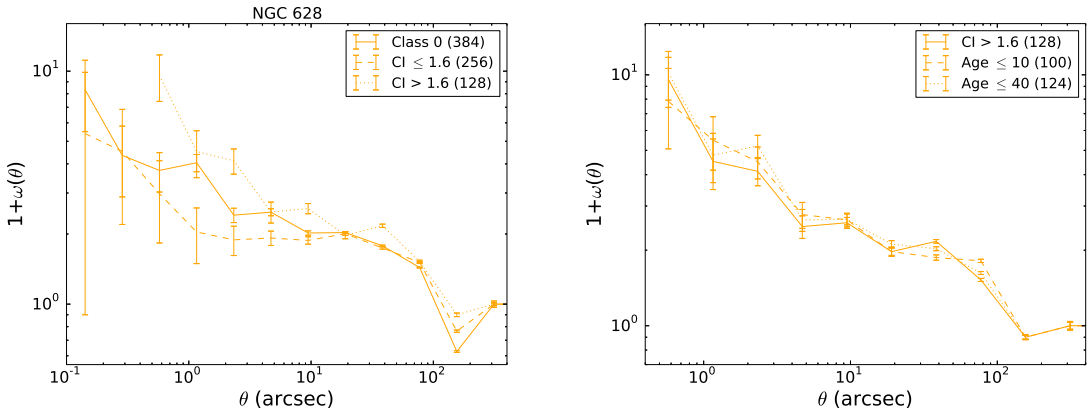

Figure 2.12. Correlation function for Class 0 sources, divided into bins by their CI value. We classify compact clusters as those with a CI mag less than 1.6 and broad clusters as those with a CI> 1.6 (top). The compact clusters show less evidence of clustering with increasing spatial scale for where we have data, which is to be expected if the broadest objects are clusters. Bottom: The two-point correlation function $1 + \omega(\theta)$ for Class 0 clusters that are above a CI cutoff of 1.6, divided into bins of ages below 10 and 40 Myr. While genuine clusters should be identified as the younger clusters of the broader CI objects, showing more evidence of clustering at smaller spatial scales, the lack of older clusters in the CI > 1.6 bin makes the correlation function nearly identical between all the ages.

all the CI ≥ 1.6 mag sources (128), 98% have ages of ≤ 40 Myr, which indicates that sources with high values of CIs are sources with younger ages.

As we saw in the power law fits of Section 2.6.1, Class 0 sources are described with a fairly shallow power law of index $\alpha = -0.20$, which is not as steep as expected if the objects are part of a larger hierarchical distribution. However, the clustering does increase when we limit our selection to class 0 clusters with CI > 1.6, as can be seen in Figure 2.12. If we repeat the power law fits to the CI > 1.6 clusters, we find the power law index increases to $\alpha = -0.31$, up from $\alpha = -0.20$. This is still not as steep as the slope observed for Class 2 or 3 clusters, however, it compares to slopes measured by other star cluster studies in the range of -0.4 to -0.8 (Scheepmaker et al., 2009) and -0.7 to -1 (Zhang et al., 2001).

The age where we start to see a decrease in the clustering of stellar clusters, 40 Myr, is significantly younger than the age at which the clustering structure of

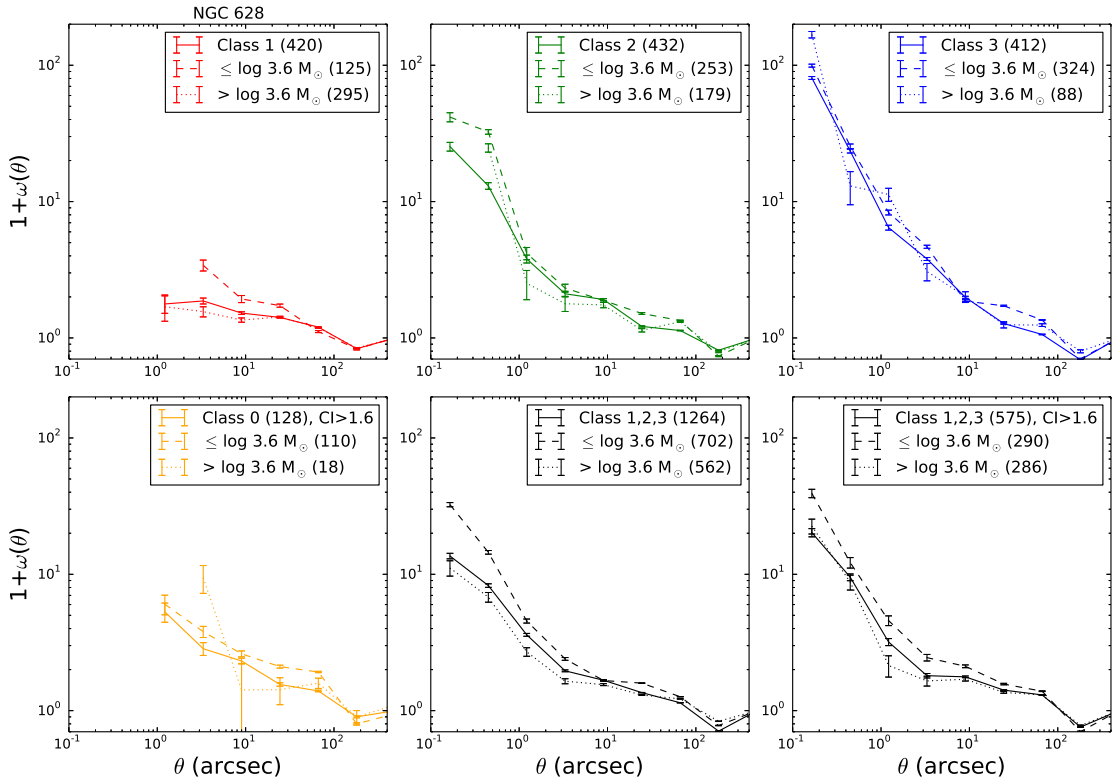

Figure 2.13. Two-point correlation function $1+\omega(\theta)$ for all the cluster classifications in NGC 628 divided into mass bins. Clusters with masses below $\log(M/M_{\odot}) = 3.6$ are shown in the dashed line and above are represented by the dotted line. For the Class 0 sources, the sources with CI > 1.6 mag are shown in orange, where we see a large increase in the clustering for the massive sources.

stars starts to dissipate (300 Myr, for eg., Bastian et al., 2011b). The difference in age can be explained if clusters lose their substructure through a combination of both dissolution and the migration. While stars maintain their individuality over their lifetime, clusters can dissolve or evaporate over a much shorter timescale, implying that a cluster may lose its identity while the stars that compose it do not.

2.6.3 Mass

We investigate the role that the mass of stellar clusters play in the hierarchy, formation, and dispersion of stellar structures.

Figure 2.13 shows how the clustering changes when we divide the clusters in the sample into bins of high and low mass, where we choose the value of $\log(M/M_{\odot}) = 3.6$ to have adequate numbers of clusters sampling each bin across all cluster classifications. For Classes 1 and 2, we notice slightly stronger correlations for the lower mass clusters, where the correlation strength starts to increase for the lower mass clusters at the smallest spatial scales. The strength of the clustering for different masses is not significantly different for Class 3 clusters. Combining all the genuine clusters (bottom middle and right panels in Figure 2.13) accentuates the increased clustering strength for lower mass clusters. We do recognize that clusters with log < 3.6 $M_{\odot}$ may be biased by stochastic IMF sampling, resulting in unreliable mass measurements at the low-mass end. This makes it difficult to draw specific conclusions on how the mass affects clustering and is further compounded by the fact that we lack clusters populating the lowest spatial scales on the correlation function. For the Class 0 sources that are limited to CI > 1.6, a large increase in clustering is observed for the large mass bin of $\log(M/M_{\odot}) > 3.6$. If we consider all the class 0 sources regardless of CI value, this same jump is not seen and there is no significant difference between the low and high-mass sources. This increase in clustering for the more massive clusters is not seen for any of the other cluster types within NGC 628.

2.7 Discussion

The hierarchical morphology in stellar ensembles has been investigated by Elmegreen et al. (2014) with 12 LEGUS galaxies, including NGC 628. The power laws of star forming regions suggest that hierarchically structured regions are represented with a common unit length of a few hundred parsecs. The observed power law of hierarchical structure of star forming complexes observed in galaxies is consistent with the model that star formation is regulated by turbulence where gas compression form successively smaller clouds within larger clouds (see Vázquez-Semadeni et al., 2009;

Elmegreen, 2010; Kritsuk et al., 2013; Elmegreen et al., 2014). The hierarchical nature of star formation will likewise drive the observed hierarchy of young stars, and therefore, the distributions of young stellar clusters. The same processes will likely form a secondary correlation for age, making larger regions older in proportion to the turbulent crossing time (Efremov & Elmegreen, 1998). These results indicate that the young stellar clusters that we observe should also be imprinted with the hierarchical structure of the natal cloud structure from which they are born. The influence of the self-similar nature of star formation is demonstrated on the distributions of resolved massive young stars on galactic scales and the power laws in their parameters correlations, as shown for NGC 6822 (Gouliermis et al., 2010) as well as with the much deeper LEGUS data for NGC 6503 (Gouliermis et al., 2015b). Our results are consistent with the study of the luminosity function of the young stellar populations in NGC 628 by Adamo et al. (2017), found to be well-described by a power law distribution with index close to -2, suggesting that clusters form in a turbulence-driven fractal ISM.

Excluding Class 0 and 1 sources, our correlation functions for NGC 628 are best described with two power laws connected at a break point corresponding to 158 pc, though this break is much shorter for Class 3 clusters. The thickness of the galactic disk of NGC 628 is around 0.25 kpc (Peng, 1988; Ma et al., 1998), a factor of two different from our measured break point in the power spectrum. Previous results of power spectra in galactic observations have shown that breaks in the power law are related to the line-of-sight thickness of galactic disks. Observations of the LMC disk with IR emission by Block et al. (2010) show a two component power spectrum with a break at 100–200 pc, occurring at a depth that is comparable to the disk line-of-sight thickness. The same break in the power spectra for the LMC at 100 pc is also seen with HI observations (Elmegreen et al., 2001), interpreted as the line-of-sight thickness. If the line-of-sight thickness is responsible for the observed break in power

spectra, dwarf galaxies could lack the break as the line-of-sight depth is comparable to the transverse length (Westerlund, 1997; Roychowdhury et al., 2010). Increasing the scale height results in an increase of chance alignment along the line of sight, serving to decrease both the clustering scale length and magnitude of clustering.

Interestingly enough, the youngest population (≤ 10 Myr) of star clusters observed by Scheepmaker et al. (2009) were observed to have the shallowest slope ($\alpha = -0.4$), where their oldest star clusters ($30 \leq \text{Age(Myr)} < 400$) have the steepest slope of $\alpha = -0.8$. Our youngest clusters on average (Class 3) display the most prominent clustering behavior whereas the oldest clusters on average (Class 1) systematically show minimal clustering with a very shallow slope. Despite being fairly flat, the clustering results of our young clusters within Class 0 which have not been visually classified and could still include non-clusters, appear to be consistent with other observations, even if they fall short of the measured slopes of $\alpha \sim -0.8$ by Zhang et al. (2001). While these early results seem to suggest that the correlation function of young clusters may be similar between different galaxies (i.e., it hints toward a universal fractal dimension of hierarchical star formation), more studies of different galaxies are necessary to determine the validity of this statement.

As shown in Figure 2.11, the strength of the clustering decreases with both increasing spatial scale and increasing age, excluding our multiple peak clusters (Class 3), though the increase in the clustering is only marginal for the symmetric clusters (Class 1). At ages younger than 20 Myr, we see a transition from the stellar clusters displaying a highly clustered distribution toward a flat, non-clustered distribution for ages greater than 40 Myr. This is consistent with star clusters born within a highly clustered structures and the clustering dissipating in a little as a few Myr after the formation from random motions of the clusters and shear effects.

We see a slight increase in clustering with decreasing mass (Figure 2.13), predominately only affecting the smallest spatial scales. This is not an expected result as in

a hierarchy/turbulence model universe, everything is self similar so there should be no mass dependency on the clustering results. We know that there is an age effect on the clustering correlation and to understand if there is a bias present that makes our selection of low-mass clusters younger; Figure 2.14 shows clusters younger than 40 Myr divided into high and low mass bins. When only considering our youngest clusters, the dependency of the mass of a cluster on the clustering strength disappears and we can see that the correlation is nearly identical for clusters above and below our mass bin of 3900 $M_{\odot}$. Thus the mass dependence seen in Figure 2.13 is probably an age effect (i.e., low mass clusters tend to be younger and more correlated than high mass clusters). We lack number statistics in NGC 628 to be able to investigate how mass influences the clustering behavior for clusters older than 40 Myr and we defer it to future studies.

We see that the opposite behavior occurs for the broadest Class 0 candidate cluster sources as there is a steep rise in the clustering for massive clusters (Figure 2.13 and Figure 2.14). However, as we are dealing with small number statistics (there are only two objects within the first two bins of the CI $>$ 1.6 and mass $>$ 3900 $M_{\odot}$ bin) the clustering strength increase is accompanied with large error bars, making any conclusion unreliable at best. With greater cluster numbers, we will be able to statistically investigate if the properties of the star clusters change as a function of environment and if this is reflected in a change of the observed clustering properties.

2.8 Summary and Conclusion

In this paper, we present a technique to investigate the spatial clustering of the young stellar clusters in the LEGUS galaxy NGC 628 with UV and optical data taken with WFC3/UVIS from HST. The inclusion of NUV observations within this study provides reliable measurements of both the age and masses of the clusters, giving us

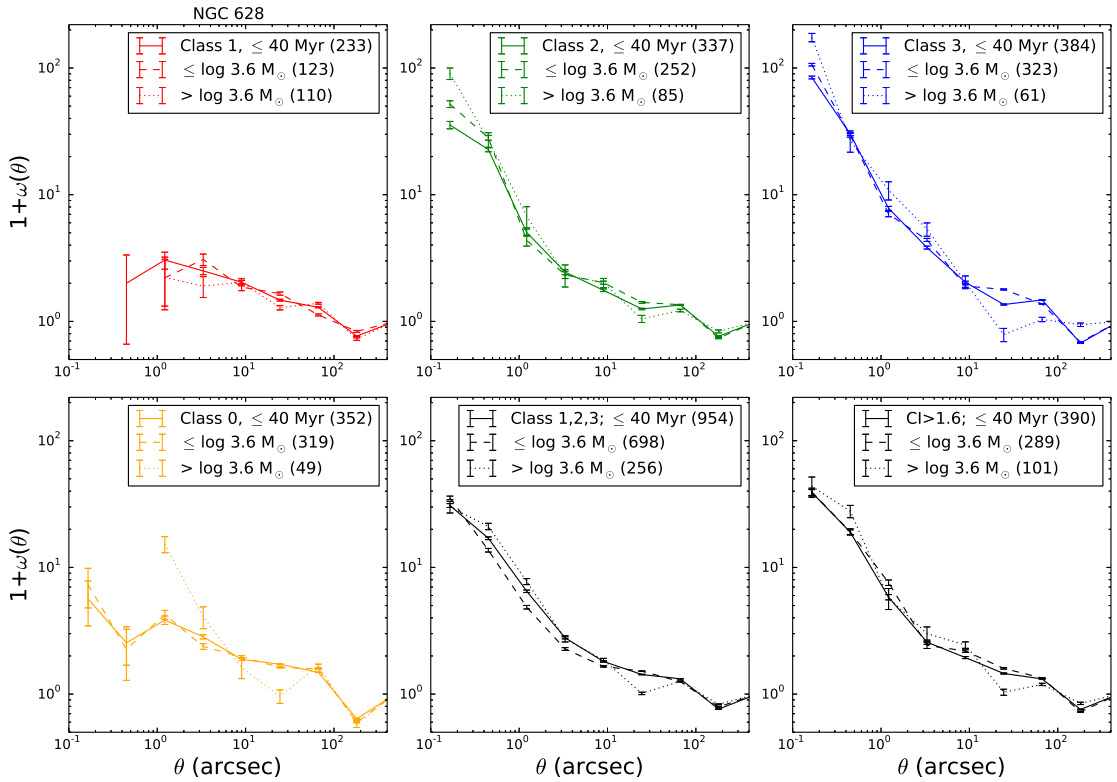

Figure 2.14. Two-point correlation function $1+\omega(\theta)$ for all young (¡40 Myr) clusters in NGC 628 divided into mass bins. Clusters with masses below (dashed line) and above (dotted line) the mass of $\log(M/M_{\odot}) = 3.6$.

an unprecedented, high-angular resolution view of the clustering in a way that has not been possible before.

Through visual inspection, we have identified 1264 stellar clusters within the galaxy that have an absolute V-band magnitude brighter than -6, CI > 1.4, and detection in at least four filters. For the fainter (V-band magnitude fainter than -6) clusters that did not undergo visual inspection (a total of 384 objects) that still have detection in at least four filters, we take cuts in both concentration index and age to separate out possible cluster candidates from spurious, non-cluster candidates. Using this method, we identify 128 possible clusters of out a total of 384 faint sources, increasing the number of possible stellar clusters available for analysis to 1392.

We implement the two-point correlation function to describe and study the amount of clustering occurring within all our cluster candidates. We find that the clusters of NGC 628 have correlation functions that can be well described with a power law slope consistent with what is expected if the clustering is part of a hierarchy and that the observed clustering strength increases within younger clusters. Our results reveal the internal hierarchical morphology present in the spatial distribution of the stellar clusters. We find that the general observed hierarchy of the young stellar clusters decreases monotonically across the dynamical range of the galactic disk from 7 – 10 kpc.

Our clusters with centrally concentrated light profiles have the highest average mass and oldest ages and show the flattest power law slope of $\alpha \sim -0.14$ across all spatial scales. Our multiple peaked clusters have the steepest slope of $\alpha \sim -1.51$ and correspond to the clusters with the lowest average masses and youngest ages in the galaxy. The asymmetrical clusters display a slope of $\alpha \sim -0.82$. The weighted average of the measured power law slope for the combination of all the genuine clusters is $\alpha \sim -0.8$. These results from the correlation function reveal that our cluster population is part of a self-similar distribution with an average two-dimensional fractal

distribution D2 = 1.2. This fractal distribution is consistent with the hierarchical morphology of star forming structures formed from fractal gas that has undergone turbulent fragmentation.

Both the asymmetrical and multiple peaked sources show a break in their power law, at 3".3 and 1".2 respectively, corresponding to a spatial scale length of 158 pc and 58 pc; the average slope at length-scales beyond the breakpoint is $\alpha \sim -0.24$. These results suggest that the stellar clusters do form in clustered structures at the small spatial scales, while the clustering is erased and the spatial distribution is observed to be more homogeneous at larger spatial scales beyond the break at 3".3. Conversely, the symmetrical clusters have a single shallow slope of $\alpha \sim -0.14$, indicating nearly uniform distribution across all scales. The measured slope for the faint non-visually classified sources have a power law slope of $\alpha \sim -0.20$; when we exclude sources with measured CI values less than 1.6, the slope steepens to $\alpha \sim -0.31$, which is still below what we find for the average of the genuine clusters in the survey. The asymmetrical and multiple peak clusters dominate the observed clustering structure.

In addition to the clustering strength decreasing with increasing spatial scale, we also find that the clustering strength decreases with increasing age, consistent with the fractal distribution being erased as the clusters age. We find a dramatic decrease in the clustering at a timescale of 40 Myr. We find that dividing up clusters by mass has a minor influence on the clustering behavior, however, the low-mass bin (< 4000 $M_\odot$) is subject to stochastic IMF effects and as such, these masses are highly uncertain. The mass dependency on the clustering behavior vanishes when we only consider the young clusters below 40 Myr.

Future investigations of a larger sample of LEGUS galaxies will provide a better understanding of whether there is a common correlation length observed in the stellar clusters across galaxies and if the clustering depends on the ambient galactic environment. The clustering nature of the young stellar clusters will not only provide a

window to investigate whether or not star formation occurs in hierarchical patterns in both space and time, it will also inform us on the clustering properties of stars at the time of their formation, providing important information on the nature of the star formation process itself. If stars and clusters map the same type of hierarchy, this would allow the use of observations on the clustering distribution of stellar clusters to trace and study star formation, providing a way to understand the cluster formation efficiency in relation to the star formation efficiency.

CHAPTER 3

THE HIERARCHICAL DISTRIBUTION OF THE YOUNG STELLAR CLUSTERS IN SIX LOCAL STAR FORMING GALAXIES

This chapter[1] focuses on expanding the results as laid out in Chapter 2 to a sample of six galaxies. This will enable us to find a dependence of the clustering of the young star clusters on galaxy environment.

3.1 Abstract

We present a study of the hierarchical clustering of the young stellar clusters in six local (3–15 Mpc) star-forming galaxies using Hubble Space Telescope broadband WFC3/UVIS UV and optical images from the Treasury Program LEGUS (Legacy ExtraGalactic UV Survey). We identified 3685 likely clusters and associations, each visually classified by their morphology, and we use the angular two-point correlation function to study the clustering of these stellar systems. We find that the spatial distribution of the young clusters and associations are clustered with respect to each other, forming large, unbound hierarchical star-forming complexes that are in general very young. The strength of the clustering decreases with increasing age of the star clusters and stellar associations, becoming more homogeneously distributed after ∼40–60 Myr and on scales larger than a few hundred parsecs. In all galaxies, the associations exhibit a global behavior that is distinct and more strongly correlated from

[1]These results are based on research published by Grasha et al. (2017a), and is reproduced by permission of the AAS.

compact clusters. Thus, populations of clusters are more evolved than associations in terms of their spatial distribution, traveling significantly from their birth site within a few tens of Myr, whereas associations show evidence of disruption occurring very quickly after their formation. The clustering of the stellar systems resembles that of a turbulent interstellar medium that drives the star formation process, correlating the components in unbound star-forming complexes in a hierarchical manner, dispersing shortly after formation, suggestive of a single, continuous mode of star formation across all galaxies.

3.2 Introduction

Star clusters are gravitationally bound stellar structures, with radii between 0.5 to several parsecs and masses between 10^3 and 10^7 $M_\odot$ (Portegies Zwart et al., 2010). Because most, if not all, stars form in some type of stellar aggregate (Lada & Lada, 2003), stellar clusters are a direct product of the star formation process within galaxies. Compounded by the fact that young stellar clusters are intrinsically brighter than single stars, star clusters become important tracers of the recent star formation history in galaxies beyond which individual stars cannot be detected.

Within the hierarchical model, star formation occurs within structures that have smoothly varying densities and sizes that range from pc to kpc scales, with denser regions nested within larger, less dense areas (e.g., Elmegreen et al., 2006; Bastian et al., 2007). Bound star clusters form at these peak densities within the hierarchy. Most structures within the hierarchy are themselves gravitationally unbound and the stellar components are expected to inherit their clustered substructure from the molecular clouds from which they are born (Scalo, 1985). Recent analyses of 12 local galaxies (Elmegreen et al., 2014) found that the clustering of star formation remains scale-free, up to the largest scales observable, for both starburst galaxies and more quiescent star-forming galaxies. This result is consistent within the framework

where the self-similar structure of the interstellar medium (ISM), regulated by turbulence, is believed to be the primary driver for the hierarchical nature of star formation (Elmegreen & Efremov, 1996; Elmegreen et al., 2014). Thus, extensive star-forming regions of several hundred parsecs or larger are expected to represent common structures, related in both space and time in a hierarchical manner that determines the structure and morphology of all galaxies.

The evolution and erasure of the unbound hierarchical structures has been the focus of investigation in recent years, where observations of local galaxies support an age-dependent clustering of the stellar components (e.g., Pellerin et al., 2007; Bastian et al., 2009; Scheepmaker et al., 2009; Gieles et al., 2011; Pellerin et al., 2012; Baumgardt et al., 2013; Gouliermis et al., 2014, 2015b; Grasha et al., 2015), and the clustering becomes progressively weaker for older populations. Hierarchical clustering is expected to dissipate with age (Elmegreen et al., 2006; Elmegreen, 2010) as the densest regions with the shortest mixing timescales lose their substructures first, whereas the larger, unbound regions will lose their substructure over longer periods of time owing to tidal forces and random velocities (Bate et al., 1998), dispersing over time to form the field population. Characterizing the clustered nature of star formation provides insight into how star formation is organized across a galaxy, by correlating local environmental conditions at sub-galactic scales – such as feedback and turbulence – to the global properties – such as dynamics and morphology – of entire galaxies and constrain the migration timescale for which stars and clusters abandon their natal structure. This will in-turn provide a vital connection between the inherently different processes of clustered star formation seen within local galaxies and the large kpc-scale star-forming structures that appear to be common at high-redshift (Immeli et al., 2004; Elmegreen et al., 2009; Förster Schreiber et al., 2011; Guo et al., 2012).

In this paper, we study the young stellar cluster populations of six galaxies as part of the Legacy ExtraGalactic UV Survey[2] (LEGUS; Calzetti et al., 2015), a Cycle 21 Hubble Space Telescope (HST) program with images of 50 nearby (∼3.5–15 Mpc) galaxies in five UV and optical bands (NUV,U,B,V,I) with the UV/Visible (UVIS) channel on the Wide Field Camera 3 (WFC3) and re-using archival ACS images when appropriate. The aim of LEGUS is to investigate the relation between star formation and its galactic environment in nearby galaxies, over scales ranging from individual star systems to kpc-sized structures. These data will help to establish a more accurate picture of galaxy formation and the physical underpinning of the gas-star formation relation. The relatively nearby location of these galaxies provides us with the high-angular resolution needed to acquire large numbers of star clusters to perform statistically accurate tests for changes in clustering strength across a representative range of galactic environments. Investigations of a few galaxies from the LEGUS project have already observationally demonstrated the relatively young (∼40–60 Myr) dispersal timescales of star-forming structures (Gouliermis et al., 2015b; Grasha et al., 2015).

This work builds on our previous paper (Grasha et al., 2015) on a study of the nearby star-forming galaxy NGC 628 using the two-point correlation function as a tool to quantify the clustering properties of the young stellar clusters, finding that the youngest clusters are spatially clustered within unbound, star-forming complexes that disperse with time. In this work, we expand our sample to investigate the clustering distribution of the stellar clusters within a larger sample of galaxies and a wider range of galactic environments. We will use the correlation function to identify common age structures, the extent that the distribution of clusters is hierarchical, on which timescale it disperses, and the dependencies of global properties (galaxy type) has on

[2]https://legus.stsci.edu/

the clustering results, if any. This will in turn inform on the nature of local, resolved star formation.

The galaxy selection is described in Section 3.3 and the cluster identification process is described in Section 3.4. The methodology of two-point correlation function is introduced in Section 3.5. In Section 3.6, we describe the results and analysis and how we use the correlation function to draw conclusions about the properties of our star clusters. We discuss our results concerning hierarchy of the stellar clusters in Section 3.7. Finally, we summarize the findings of this study in Section 3.8.

3.3 Sample Selection

In this paper, we select six local (<13 Mpc) galaxies, ranging from dwarf to grand design spirals, with visually identified stellar cluster catalogs available (see Section 3.4), from the LEGUS survey. The galaxies and their general properties are listed in Table 3.1 and shown in Figure 3.1 along with the clusters. All galaxies were observed in five broad band filters: NUV, U, B, V, and I; the list of filters used can be found in Calzetti et al. (2015). Both NGC 628 and NGC 7793 have two pointings, combined into a single mosaic for analysis; the remaining galaxies have one pointing.

3.3.1 NGC 7793

NGC 7793 is a spiral galaxy in the Sculptor group at a distance of 3.44 Mpc classified as morphological type SAd. With an angular size of 9'.3 x 6'.3, both a west and east pointing were observed to cover a significant portion of the galaxy. A study of the resolved stars has found that the radial profile exhibits a break at 5.1 kpc; beyond the disk break, the younger populations exhibit a steeper profile, indicative of high levels of stellar radial migration (Radburn-Smith et al., 2012).

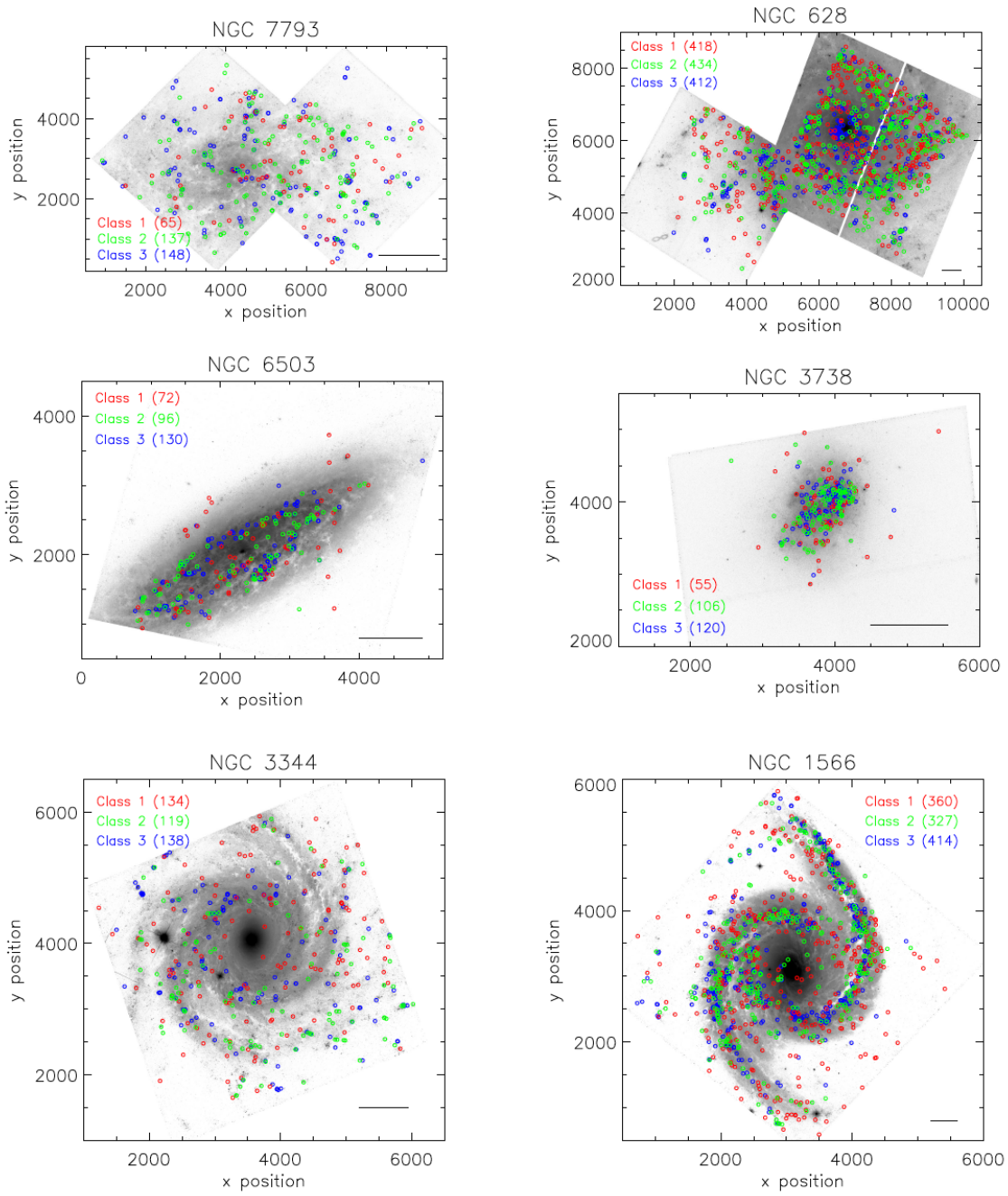

Figure 3.1. V-band images of six LEGUS galaxies overlaid with the positions of star cluster and association candidates. The cluster color corresponds to the morphological classification as described in Section 3.4. The numbers in parenthesis represent the total number of clusters within the classification. The solid black line in the bottom right of each figure represents the spatial scale of 1 kpc at the distance of that galaxy.

Table 3.1. Properties of Six LEGUS Galaxies Sample

Name	Morph.	T	Inclin. (deg.)	Dist. (Mpc)	SFR(UV) ($\frac{M_\odot}{yr}$)	M_* ($M_\odot$)	Σ_{SFR} ($\frac{M_\odot}{yr\ kpc^2}$)	R_{25} (arcm.)	Scale (pc/")	CI_{cut} (mag)
NGC 7793	SAd	7.4(0.6)	47.4	3.44	0.52	3.2×10^9	0.00907	4.67	16.678	1.3(e)/1.4(w)
NGC 3738	Im	9.8(0.7)	40.5	4.90	0.07	2.4×10^8	0.01187	1.26	23.756	1.4
NGC 6503	SAcd	5.8(0.5)	70.2	5.27	0.32	1.9×10^9	0.00620	3.54	25.550	1.25
NGC 3344	SABbc	4.0(0.3)	23.7	7.0	0.86	5.0×10^9	0.00558	3.54	33.937	1.35
NGC 628	SAc	5.2(0.5)	25.2	9.9	3.67	1.1×10^{10}	0.00444	5.23	47.956	1.4(c)/1.3(e)
NGC 1566	SABbc	4.0(0.2)	37.3	13.2	5.67	2.7×10^{10}	0.01285	4.16	64.995	1.35

Notes. Columns list the (1) galaxy name (ordered in increasing distance); (2) morphological type as listed in NED, the NASA Extragalactic Database; (3) RC3 morphological T–type as listed in Hyperleda (http://leda.univ-lyon1.fr); (4) inclination, in degrees; (5) redshift-independent distance in Mpc, (6) star formation rate ($M_\odot$ yr^{-1}), calculated from the GALEX far-UV, corrected for dust attenuation as described in Lee et al. (2009); (7) stellar masses ($M_\odot$) obtained from the extinction-corrected B-band luminosity and color information as described in Bothwell et al. (2009); (8) star formation rate surface density Σ_{SFR}; (9) optical radius of the galaxy R_{25} from de Vaucouleurs et al. (1991) in arcmin; (10) scale in pc $arcsec^{-1}$; and (11) concentration index (CI) cutoff between stars and star clusters (see Section 3.4). Both NGC 7793 and NGC 628 were observed with two pointings and the different CI cutoff for each pointing [central (c), east (e), or west (w)] is given.

3.3.2 NGC 3738

NGC 3738 is an irregular dwarf galaxy that makes up part of the Messier 81 group, located at 4.9 Mpc. NGC 3738 is a very small system with an apparent size of 2'.6 x 2'.2 and is the only Irregular galaxy in this study.

3.3.3 NGC 6503

NGC 6503 is a ring dwarf spiral galaxy of morphological type SAcd located at a distance of 5.27 Mpc. The region of intense star formation in the galaxy is sufficiently compact to be fully covered with a single pointing and is observed to be organized into a ring (Knapen et al., 2006). Work by Gouliermis et al. (2015b) shows that younger stars are organized in a hierarchical distribution whereas older stars display a homogeneous, less clustered distribution with a structure dispersion timescale of ~ 60 Myr.

3.3.4 NGC 3344

NGC 3344 is an isolated barred spiral galaxy with a morphological classification of SABbc, located at a distance of 7 Mpc. With apparent dimensions of 7'.1 x 6'.5, we have a single pointing of NGC 3344 that covers most of the inner region of the galaxy. There is a presence of ring-like morphological features at 1 kpc and 7 kpc, with a small bar present within the inner ring (Verdes-Montenegro et al., 2000).

3.3.5 NGC 628

NGC 628 is a morphological type SAc face-on grand design spiral galaxy located at a distance of 9.9 Mpc with no apparent bulge. The largest galaxy of the M 74 galaxy group with an angular size of 10'.5 x 9'.5, a central and east pointing were obtained. The clustering of the young stellar clusters in this galaxy was investigated in an earlier paper by Grasha et al. (2015), where an age dependency for the clustering strength was found with a randomization time scale of $\sim$ 40 Myr, after which the clusters displayed a flatter, more homogenous distribution.

3.3.6 NGC 1566

The most distant galaxy in our sample and the brightest member of the Dorado group, NGC 1566 is a face-on spiral galaxy with an intermediate-strength bar type classified as a SABbc. NGC 1566 is shown to host a low-luminosity AGN (Combes et al., 2014) with a star forming ring at 1.7 kpc (Smajić et al., 2015). We acknowledge that the distance of NGC 1566 is uncertain in the literature; in this work, we adopt the value of 13.2 Mpc, as reported in Calzetti et al. (2015).

3.4 Cluster Identification, Selection, and Characterization

A general description of the standard data reduction of the LEGUS datasets is available in Calzetti et al. (2015). A detailed description of the cluster selection, identification, photometry and SED fitting for the LEGUS galaxies is presented in

Adamo et al. (2017). Here we summarize the aspects of that paper that are relevant for the current analysis.

Stellar clusters within each galaxy are identified first through an automated process using SExtractor (Bertin & Arnouts, 1996) from the white-light image produced with the five standard LEGUS bands (see Calzetti et al. (2015) for the method used to produce white-light images). The configurations of SExtractor are optimized to detect a minimum of a 4σ source in at least 5 contiguous pixels within a region 30x30 pixels in size. These initial catalogs are visually inspected to see if any obvious clusters are excluded, improving on the initial SExtractor parameters.

The photometry within each band, for all sources, is corrected for foreground Galactic extinction (Schlafly & Finkbeiner, 2011). With the exception of NGC 7793, photometry is performed with a circular aperture of 4 pixels (3.8 to 10 parsecs) in radius, with the background measured within an annulus of 7 pixels (7–18 parsecs) in inner radius and 1 pixel in width. NGC 7793 uses an aperture radius of 5 pixels (3.3 parsec). Each automatic catalog includes sources which satisfy the two following conditions: (1) the V band concentration index (CI; difference in magnitudes measured with an aperture of radius 1 pixel compared to the magnitude within 3 pixels) must be greater than the stellar CI peak value; and (2) the source must be detected in at least two contiguous filters (the reference V band and either B or I band) with a photometric error $\sigma_\lambda \leq 0.35$ mag. The first condition above is intended to minimize stellar contamination from the automatic generated cluster catalog. The second condition corresponds to cluster candidate with a signal-to-noise greater than 3; this is imposed to obtain reliable constraints on the derived cluster properties (age, mass, and extinction). The median errors on the photometry of the star clusters in NGC 628 are in the range 0.04–0.08 mag.

As stars are unresolved even at the highest HST resolution power, their CI will vary little and their distribution will be highly peaked around an average value typical

of a stellar PSF within the same galaxy. Additionally, stellar clusters are partially resolved and their sizes can vary, therefore, on average, they will have larger CI values than stellar values. Using CI distributions of the extracted cluster candidates, we select a CI value that separates cluster candidates from the bulk of the stellar interlopers within each system. As the resolution power is dependent on both the HST camera used and the distance of the galaxy, the values used for the CI cutoff will vary from galaxy to galaxy and for different HST cameras. Table 3.1 lists the cutoff CI values between stars and clusters for each frame.

The LEGUS automatic cluster catalog is further down-selected based on the availability of SED fitting and visual classification outcome. In order to secure reliable measurements from SED fitting, each source is required to have a detection in at least four of the five photometric bands, necessary to adequately break the age-extinction degeneracy. The physical properties of each cluster (age, extinction, and mass) are derived using deterministic stellar population models (Yggdrasil; Zackrisson et al., 2011) and a χ^2 fitting approach which includes uncertainty estimates (see Adamo et al., 2010). The uncertainties of the age and mass estimates have errors of $\sim$0.1 dex. To analyze the cluster populations, we use models with solar metallicity for both stars and gas, an average covering factor of 50%, and a starburst attenuation curve. The Appendix B details how the adopted attenuation curve affects the derived ages for the star clusters, however, the adoption of different curves only minimally impacts the clustering results.

Within each galaxy, we visually inspect the subsample of the automatic cluster candidates that have an absolute magnitude brighter than -6 mag in the V band (excluding NGC 1566, where the detection limit is -8 mag). The magnitude limit is introduced according to the detection limits of the LEGUS sample, which enables selecting down to $\sim$1000 $M_\odot$, 6 Myr old clusters with color excess E(B–V)= 0.25 (Calzetti et al., 2015). The magnitude cut is imposed on the aperture-corrected

F555W magnitudes for all the LEGUS galaxies, with the exception of NGC 1566, which we applied an absolute V magnitude of -8, required to reach the same detection limits compared to the rest of the sample. At the distance of NGC 6503, an absolute magnitude limit of $M_V = -6$ mag corresponds to a visual apparent magnitude of 22.6 mag. At the distance of NGC 1566, the equivalent visual apparent magnitude at $M_V = -8$ mag is also 22.6 mag.

The visual inspection is performed in order to minimize the contamination within our final cluster catalogs, necessary to produce robust star cluster catalogs. Each inspected object from the automatic cluster catalog is assigned one of four classifications: (1) a symmetric, centrally concentrated cluster, usually displaying a homogeneous color; (2) a concentrated cluster with elongated density profiles and less symmetric light distribution, usually displaying a homogeneous color; (3) a less-compact, multiple peaked system that is blue in color, on top of diffuse light; or (4) a spurious detection such as a foreground/background source, single bright star, bad pixel, or a source that lies too close to the edge of the chip. Class 4 objects are contaminants and excluded from the final cluster catalog. Table 3.2 lists the median UV and optical colors for each cluster class as well as the compactness, measured by the median CI value, for each galaxy. In general, class 3 associations exhibit bluer colors in all bands across all galaxies and are marginally more extended compared to class 1 and 2 clusters.

The visual classification of each object is performed by at least three independent members of the LEGUS team and the final *visually classified* cluster catalog is compiled by comparing all the results from each individual. Figure 3.2 shows the fractional distribution of cluster types (classes 1, 2, and 3) within each galaxy. The observed increase in the fraction of class 1 clusters, and, to a minor extent, of both class 1 and 2 relative to the total, for increasing distance may be an effect of decreasing spatial resolution at larger galaxy distance as clusters are not fully resolved and our

Table 3.2. Color and Compactness of Cluster Classes

Class	$m_{UV} - m_U$ (mag)	$m_U - m_B$ (mag)	$m_V - m_R$ (mag)	CI (mag)
		NGC 7793		
Class 1	−0.29	−0.64	0.53	1.57(0.18)
Class 2	−0.35	−1.12	0.57	1.57(0.19)
Class 3	−0.46	−1.29	0.28	1.62(0.21)
		NGC 3738		
Class 1	−0.14	−0.22	0.57	1.72(0.21)
Class 2	−0.52	−0.62	0.47	1.67(0.18)
Class 3	−0.57	−0.86	0.56	1.76(0.11)
		NGC 6503		
Class 1	−0.36	−0.48	0.67	1.50(0.17)
Class 2	−0.53	−0.67	0.51	1.53(0.17)
Class 3	−0.56	−0.99	0.57	1.56(0.23)
		NGC 3344		
Class 1	−0.21	−0.58	0.68	1.50(0.12)
Class 2	−0.38	−1.28	0.40	1.51(0.13)
Class 3	−0.50	−1.46	0.13	1.60(0.16)
		NGC 628		
Class 1	−0.039	−0.38	0.71	1.57(0.14)
Class 2	−0.20	−0.85	0.59	1.57(0.14)
Class 3	−0.30	−1.18	0.52	1.62(0.18)
		NGC 1566		
Class 1	0.065	−0.38	0.64	1.48(0.11)
Class 2	−0.21	−1.02	0.53	1.50(0.14)
Class 3	−0.29	−1.28	0.42	1.57(0.17)

Notes. Columns list the (1) Classification of stellar clusters; (2) median UV color $m_{\rm F275W} - m_{\rm F336W}$; (3) median optical color $m_{\rm F336W} - m_{\rm F435W}$; (4) median optical color $m_{\rm F555W} - m_{\rm F814W}$; and (5) median concentration index. Numbers in parentheses indicate uncertainties in the final digit(s) of listed quantities, when available.

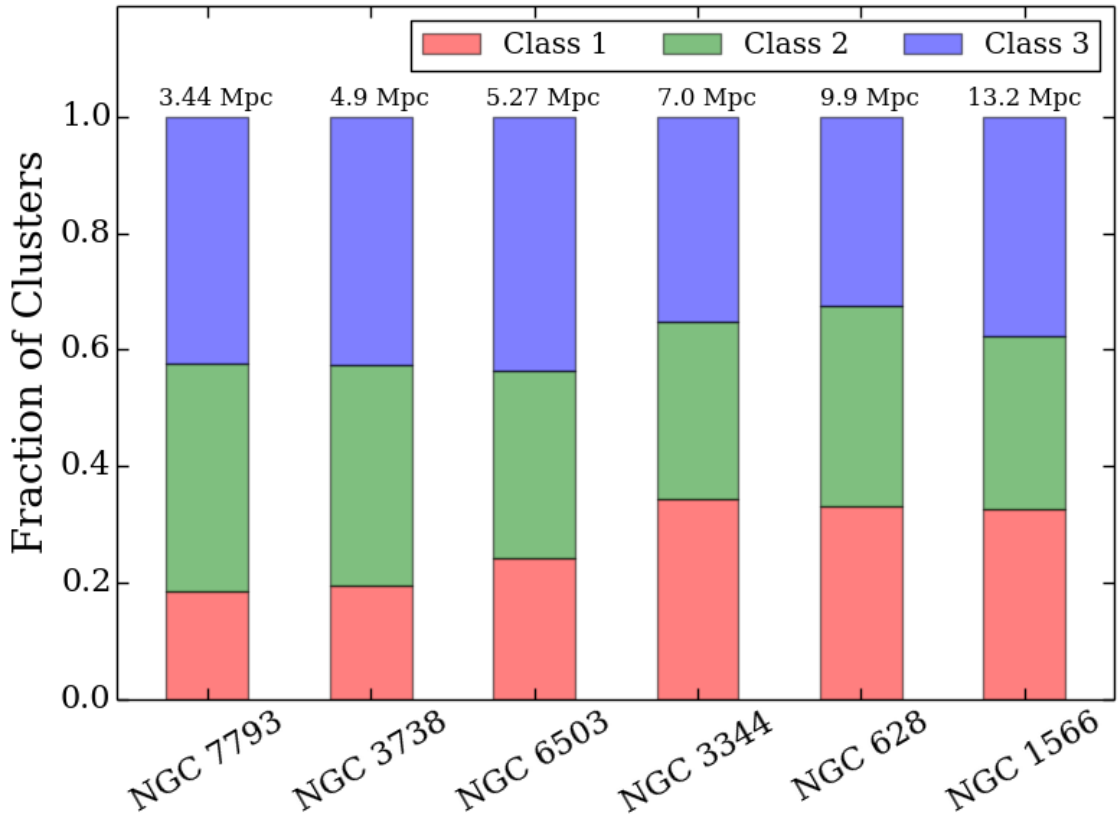

Figure 3.2. Fractional distribution of cluster types within each galaxy. The color represents the cluster classification and the galaxies are ordered in increasing distance, listed on the top of each bin.

ability to recover their morphology is dependent on both the distance of the galaxy and the severity of crowding where the individual clusters are located. However, this is a small effect, especially when class 1 and 2 sources are considered together and compared with the frequency of class 3 sources.

The main difference between the classifications is linked to the morphology of the systems. Classifications 1 and 2 are considered to be compact star clusters candidates, and for brevity, we will refer to these as star clusters throughout this paper (Adamo et al., 2017). It is very challenging to constrain the dynamical state of young stellar clusters; we are not able to determine whether or not the star clusters are actually gravitationally bound. Class 3 objects may constitute a different type of

source compared to class 1 and 2 clusters, where their multi-peak and asymmetrical morphology suggest that they are likely stellar associations, dissolving on short time scales due to being gravitationally unbound (Gieles & Portegies Zwart, 2011); we refer to class 3 sources as 'associations' throughout this work. Due to the nature in the difference of the physical properties (see Section 3.4.1 and Section 3.4.2) of class 1 and 2 clusters versus the multi-peak associations, throughout our analysis we quite often examine the (possibly bound) class 1 and 2 clusters together and compare their clustering results separately to that of the class 3 associations.

Detailed descriptions of the completeness tests, applied to NGC 628 as a test source, and how the impact of the assumptions and the selection criteria affect the LEGUS cluster catalogs can be found in Adamo et al. (2017). In brief, at the distance of NGC 628 (9.9 Mpc) and a CI cutoff of 1.4 mag for the central pointing, we are unable to resolve very compact clusters with a size of $R_{\rm eff} = 1$ pc and smaller. We are not worried about missing clusters in our analysis as the size distribution of clusters within NGC 628 peaks at values around 3 pc (Ryon et al., 2017), well above our completeness limit of 1 pc. The magnitude limits corresponding to 90% recovery range from 22.6 to 24.3 mag, depending on the band. Additionally, clusters older than $\sim$300 Myr are affected by incompleteness due to evolutionary fading, however, only the very youngest clusters $\lesssim$ 100 Myr have the strongest influence on the clustering results, and thus, we are not heavily concerned with incompleteness affecting the results and analysis. We investigate selection effects due to stochastic sampling of the IMF and the impact on the derived cluster ages in Section 3.5.1 and the impact on the clustering results and analysis when applying mass cuts to the sample.

3.4.1 Age Distribution

Figure 3.3 shows the distribution of the stellar cluster ages for each galaxy broken down by cluster classification along with the median age. Class 1 clusters (symmet-

rical) have the oldest median age and the median age of the population decreases for class 2 (asymmetrical) clusters and class 3 (multiple peak) associations, respectively. A notable feature is the flat or decreasing distribution toward larger ages, especially visible in the class 3 associations, consistent with cluster disruption on timescales of a few tens of Myr or less; cluster fading can compound this effect as well. The class 3 population does behave distinctly differently compared to class 1 and 2 clusters: in addition to a decrease in their numbers at ages older than a few 10 Myr, their clustering properties in NGC 628 behave quite differently than class 1 and 2 clusters (Grasha et al., 2015), and their disruption rate is significantly higher compared to class 1 and 2 clusters (Adamo et al., 2017). Thus, we maintain that the class 3 associations represent a distinct type of system and appear to be born with lower densities than class 1 and 2 clusters and are not the remnants of dispersed/disrupted class 1 or 2 clusters.

3.4.2 Mass Distribution

Figure 3.4 shows the distribution of the cluster masses within each galaxy, divided by cluster classification. The clusters display mass trends across all six galaxies that are dependent on their classification, similar to what is observed for the age distributions (Section 3.4.1): class 1 clusters, on average, are more massive compared to the rest of the clusters and the class 3 associations are the least massive. The three morphological categories of cluster candidates, once we take into account their different age and mass properties, also reinforces the physical differences between the classifications, as discussed in Section 3.4.1. Class 1 clusters are older and more massive stellar clusters, indicative of evolved systems that are relaxed and have likely already survived disruption. Class 3 associations, on the other hand, are clusters that are much younger and less massive compared to bound stellar clusters.

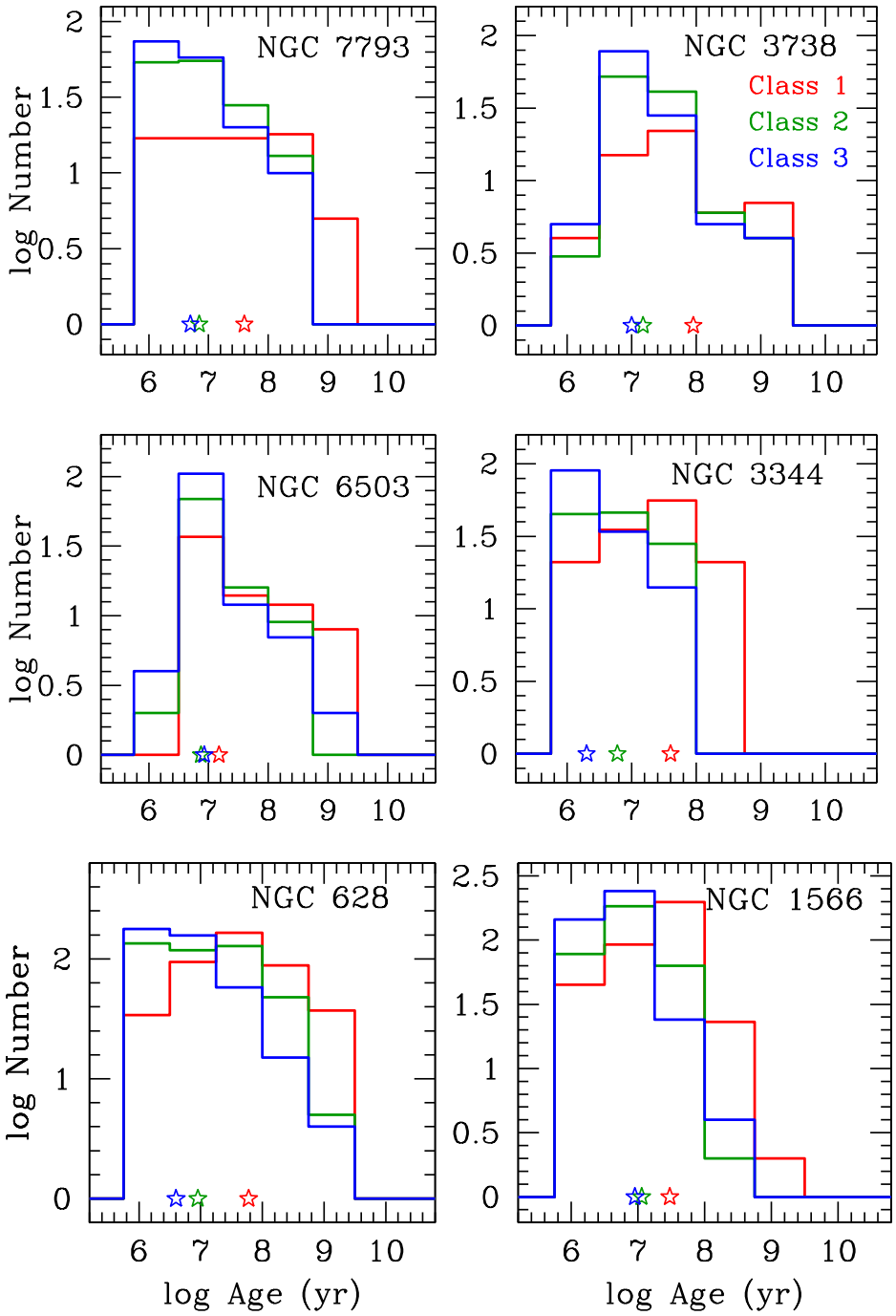

Figure 3.3. Distribution of the star cluster ages for each cluster class in all the galaxies. The clusters are separated by class type: Class 1 (red; symmetrical), 2 (green; asymmetrical), and 3 (blue; multiple peak). The open star symbol shows the median mass of each distribution. Class 1 clusters are always older whereas class 3 associations are always younger.

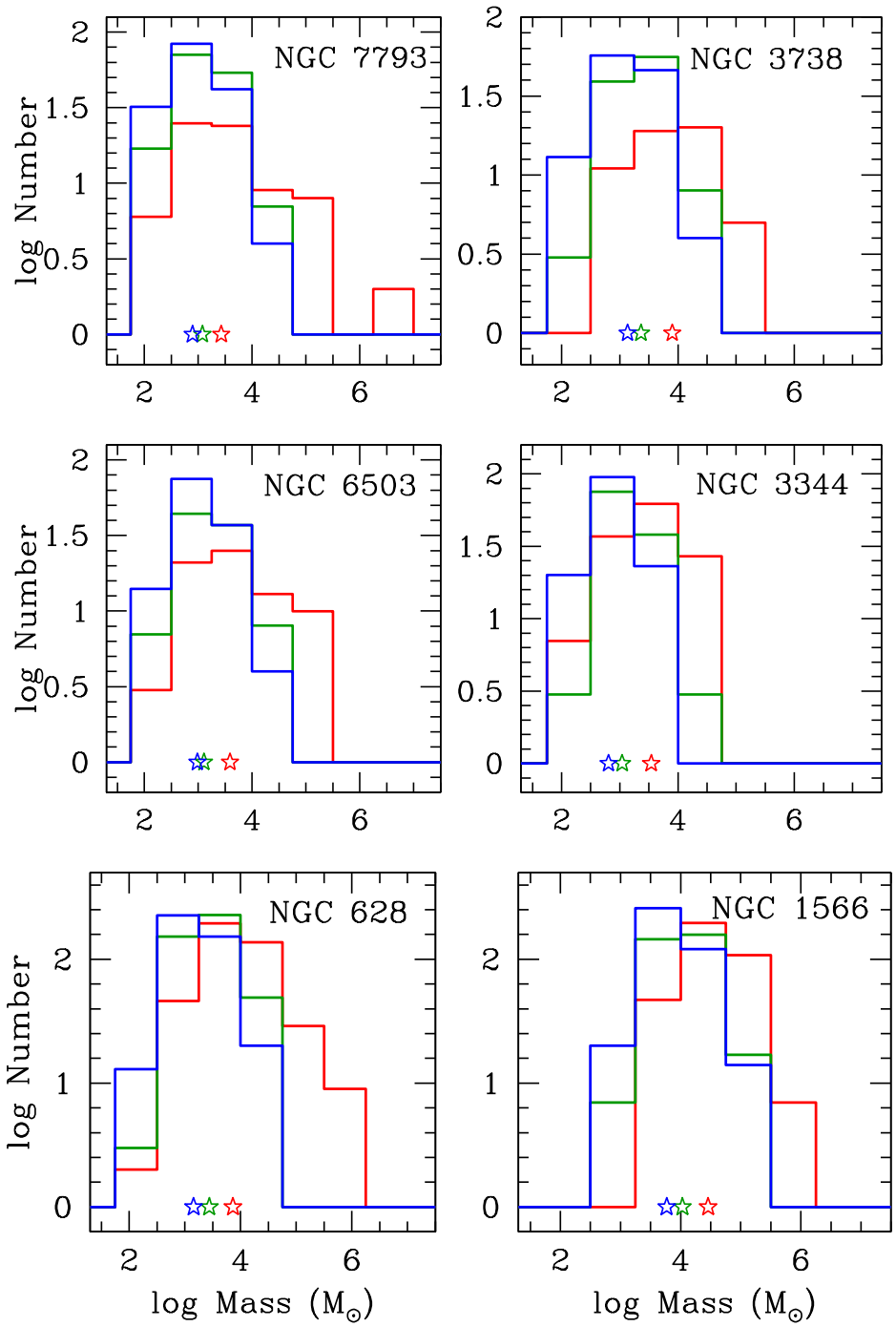

Figure 3.4. Distribution of the star cluster masses for each cluster class in all the galaxies. The clusters are separated by class type: Class 1 (red; symmetrical), 2 (green; asymmetrical), and 3 (blue; multiple peak). The open star symbol shows the median mass of each distribution. Class 1 clusters are always more massive whereas class 3 associations are always less massive.

3.4.3 Deprojection of the Galactic Disk

In order to assess the spatial distribution of the young stellar clusters free from the effect of projection of the galactic disk. We deproject the stellar cluster positions from the plane of the sky to the plane of the galaxy using the simple assumption that each galaxy can be described with an axisymmetric flat rotating disc.

The positions of star clusters are corrected for the inclination i of the galaxy in a two step process. First the intrinsic x, y detector coordinates in the plane of the galaxy are rotated as

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta, \end{aligned} \tag{3.1}$$

where x', y' are the rotated coordinate axes and the position angle θ is determined by the orientation of the observed field of view. The rotated pixel coordinates are then deprojected for the line of sight inclination angle as x' and $y'/\cos i$. We have found that for galaxies with an inclination angle below 40°, correction for projection has a minimal impact on the relative position of the clusters, and therefore, does not influence the clustering results. Figure 3.5 shows the spatial distributions of the stellar clusters, where the positions of the clusters in three galaxies with $i > 40°$ have been deprojected, NGC 7793, NGC 3738, and NGC 6503. For all calculations in this paper, we use the deprojected positions of the star clusters for the above three galaxies. Due to the low inclination of NGC 628, NGC 3344, and NGC 1566, the star cluster positions are not corrected for projection effects and we use the cluster positions as shown in Figure 3.1.

3.5 The Two-Point Correlation Function

We implement the angular two-point correlation function, $\omega(\theta)$, as projected onto the plane of the sky, to measure the magnitude of clustering as a function of scale size for the young stellar clusters in six nearby galaxies (Table 3.1). The two-dimensional

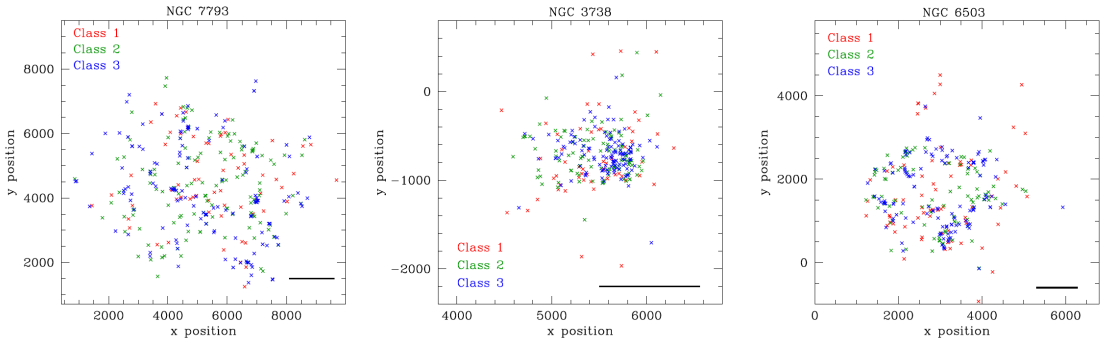

Figure 3.5. Deprojected pixel positions for highly inclined galaxies. Clusters are color coded by the classification: class 1 clusters are shown in red, class 2 clusters are shown in green, and class 3 associations are shown in blue. NGC 628, NGC 3344, and NGC 1566 are not corrected for inclination as the deprojection is a small effect for these galaxies. The solid black line in the bottom right of each figure represents the spatial scale of 1 kpc at the distance of that galaxy.

correlation function $1+\omega(\theta)$ is defined as the probability above Poisson of finding two star clusters with an angular separation θ as $\mathrm{d}P = N^2[1+\omega(\theta)]\ \mathrm{d}\Omega_1\mathrm{d}\Omega_2$, where N is the surface density of clusters per steradian with two infinitesimal elements of solid angle $\mathrm{d}\Omega_1$ and $\mathrm{d}\Omega_2$, separated by angle θ (Peebles, 1980). For truly random Poisson distribution, the two-point correlation function will be flat across all scales, such that $1+\omega(\theta) = 1$ and a clustered stellar sample will have $1+\omega(\theta) > 1$ at small values of θ, declining with increasing scales toward that of a flat, non-clustered distribution. The correlation function of a fractal (self-similar) distribution is described with a single power-law as $1+\omega(\theta) = (r/r_0)^{-\gamma}$, where r_0 is the characteristic scale-length of the clustering and γ describes the hierarchical ordering (see Section 3.6.1).

To measure $\omega(\theta)$, pairs of stellar clusters are counted as a function of their separation (deprojected if necessary), compared to what is expected for an unclustered distribution. The unclustered distribution (in x, y position) of sources must populate the same sky coverage and geometry (e.g., edges, masks) as the observational data. We define masks as areas that exclude all data, such as the ACS chip gap or areas with dust lanes, where there is a reduction in the observed number of clusters with

respect to the global average. We reproduce, as closely as possible, the geometry of the galaxy region sampled for each random catalog. The ratio of pairs of clusters observed in the data relative to pairs of points in the random catalog is then used to estimate the correlation function $\omega(\theta)$. In this study, we implement the Landy-Szalay estimator (LS; Landy & Szalay, 1993), as described in Chapter 2. The size of the bin is determined by the sample size, selected as a compromise between resolution and total number of clusters available to be sampled within each bin. The formulation in Eq 2.1 is optimized to mitigate edge effects and systematic errors on the computed correlation function, necessary to alleviate unrealistic calculations of $\omega(\theta)$ at large spatial scales due to the deviation from a random distribution which can arise from a limited field of view and edge effects. In Grasha et al. (2015), we investigate the influence of the random catalog size and geometry on the outcome of the two-point correlation function.

Figure 3.6 shows the correlation functions for all six galaxies, calculated for the entire sample of star clusters and divided by their cluster classification. As expected for star-forming structures formed hierarchically, there is a systematic decrease in the clustering with increasing spatial scales across all galaxies. Additionally, the clustering strength of the class 1+2 (bound) clusters across all the systems is much weaker than what is observed for the class 3 associations alone. Thus, the bound clusters – for all ages – have a more homogeneous and relaxed distribution throughout their galaxy at a given spatial scale compared to the (highly clustered) distribution of associations. This can result from either different formation/evolutionary paths between bound and unbound clusters and/or an age dependency of the clustering strength, as found in Grasha et al. (2015) and investigated further in Section 3.6.2 and Section 3.6.4 for these systems.

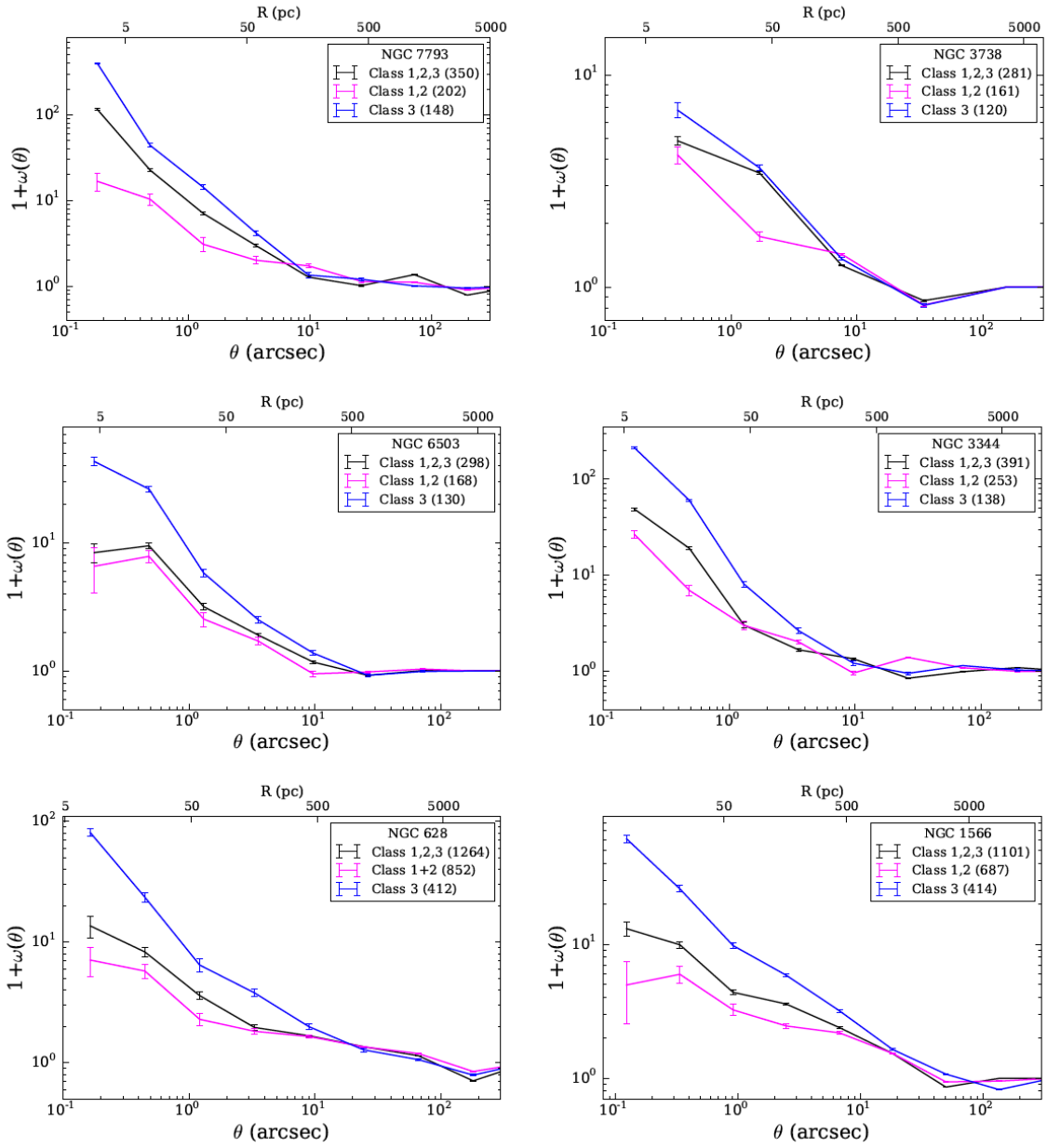

Figure 3.6. Two-point correlation function of the clusters for the six galaxy study. The plows are displayed as a function of angular distance (arcsec) and the top axis showing the corresponding spatial scale at the distance of the galaxy. The colors represent the classification of each cluster, as defined in Section 3.4: class 1+2 clusters are shown in magenta, class 3 associations are shown in blue, and all clusters are shown in black. The numbers in parentheses show the number of clusters in each classification. The 1σ uncertainties are the sample standard deviation of bootstrap resamples within each bin.

3.5.1 Selection Effects

At the distance of NGC 628 (9.9 Mpc), the LEGUS cluster catalogs are complete down to 5000 $M_{\odot}$ (Adamo et al., 2017). As outlined in Section 3.4, we do not consider stochastic sampling of the IMF when deriving the mass and age of the clusters, which becomes increasingly important for clusters with masses below $\sim$ 5000 $M_{\odot}$. However, as found in simulations by Krumholz et al. (2015a), the properties of the cluster population as a whole are relatively similar between both conventional deterministic and stochastic fitting procedures. A comparison of the cluster ages show that there is no difference in the results for ages that are derived with stochastic versus non-stochastic models (Grasha et al., 2017b). As the clustering results is most heavily influenced by the youngest clusters (<100 Myr; see Section 3.6.2 as well as Grasha et al., 2015), which may lie below the mass limit of 5000 $M_{\odot}$ (Adamo et al., 2017), it is important to examine how mass-cuts influence the clustering analysis to ensure that our results are not biased against the detection limitations.

Figure 3.7 shows the clustering correlation for the young (< 100 Myr) clusters in NGC 628 with mass cuts applied at both 3000 and 5000 $M_{\odot}$. For both cases, the clustering strength does not change when applying a mass cut and the correlation functions are in agreement with each other and with the entire cluster population. The reason there is no change to the observed correlation function for high/low mass clusters is easily understood if we consider that low-mass clusters ought to be populated throughout the galaxy in a similar manner as the entire cluster population. Consequently, incompleteness effects will not manifest themselves in a way that missing clusters (in this case, low-mass clusters) are preferentially located within one region of a galaxy, affecting the global clustering distribution. Missing low-mass clusters evenly throughout a galaxy only acts to reduce the total clusters available for analysis, and if anything, serves to artificially reduce the observed global clustering.

Due to this negligible effect, we do not perform a mass cut on any galaxy and have no need to worry about selection effects affecting the clustering results.

3.6 Results and Analysis

3.6.1 Quantifying the Correlation Strength

In a two-dimension (projected) self-similar distribution, the total number of clusters N increases with radius r as $N \propto r^{D_2}$, where D_2 is the two-dimensional fractal dimension (Mandelbrot, 1982). Fractal distributions exhibit a power-law dependency of the correlation function $1 + \omega(\theta)$ with increasing radius of the form $1 + \omega(\theta) \propto r^{\alpha}$ (Calzetti et al., 1989; Larson, 1995) and the number of clusters for every radial aperture will increase as $N \propto r^{\alpha} \times r^2 \propto r^{\alpha+2}$. Thus, we can then relate the power-law slope α from the correlation function $1+\omega(\theta)$ to determine the two-dimensional fractal dimension as $D_2 = \alpha + 2$. A flat, non-clustered distribution of $\alpha = 0$ will result in a fractal geometric dimension of $D_2 = 2$; a derived steep (negative) slope will be indicative of a clustered distribution, resulting in a fractal dimension less than 2. We determine the slope α from the correlation function $1+\omega(\theta)$ at small scales in log-log space through a Levenberg-Marquardt non-linear least square minimization fit that is the same as in Chapter 2, Eq. 2.3.

Often, the correlation function is best described with a double power law as opposed to a single slope (e.g., Larson, 1995; Gouliermis et al., 2014), determined through minimizing χ^2. For those cases, we determine two slopes using the same two-component power low as 2.4. As is in Chapter 2, both slopes and the breakpoint are free parameters in the fit.

For a hierarchical model, the distribution of star formation and interstellar gas, over a large range of environments, is shown to exhibit a projected, two-dimensional fractal dimension of $D_2 \sim 1.2$–1.6 (e.g., Beech, 1987; Falgarone et al., 1991; Elmegreen & Falgarone, 1996; Sánchez et al., 2005; Elmegreen et al., 2006; Sánchez & Alfaro,

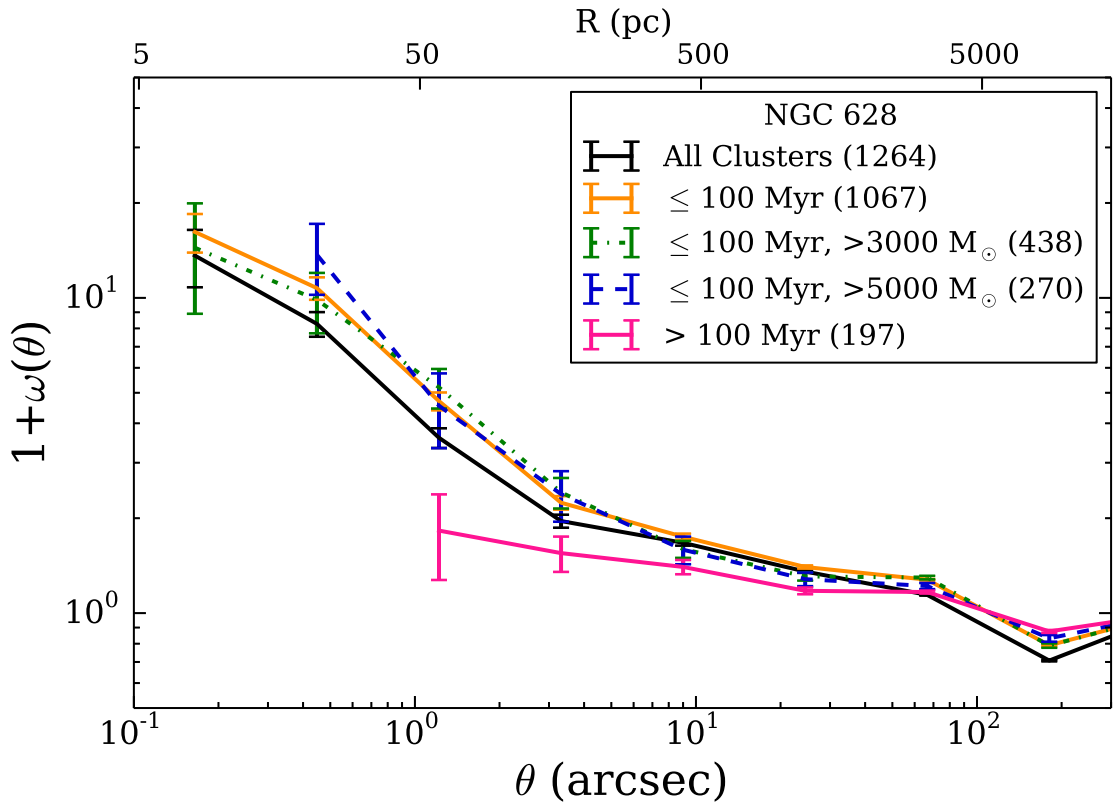

Figure 3.7. Two-point correlation function $1 + \omega(\theta)$ of NGC 628 showing the effect of a mass cutoff. The colors represent different age and mass ranges and the numbers in parentheses show the number of clusters in each age/mass bin. For all clusters with ages less than 100 Myr, the correlation function does not change for different mass cuts. Thus, we are not worried about selection effects affecting the clustering results and analysis. However, there is a pronounced decrease in the clustering strength for clusters older than 100 Myr; this effect of age on the clustering strength is discussed in Section 3.6.2.

2008). Smaller values of the fractal number D_2 correspond to higher fractal dimensions (i.e., systematically more clustering) and steeper slopes for the correlation function. These values for the 2D fractal dimension are in agreement with the predicted fractal dimension for the three-dimensional, D_3, density structure of the ISM (Federrath et al., 2009) in a turbulent medium. Often, the connection between the three-dimensional fractal dimension D_3 and the projected two-dimensional fractal D_2 is assumed to be $D_2 = D_3 - 1$ (Mandelbrot, 1982). However, the conversion between D_2 and D_3 is not straightforward (Gouliermis et al., 2014).

Table 3.3 lists the distribution of power law slopes and amplitudes for all the clusters in each galaxy. All galaxies are best-fit with a single power law, excluding NGC 628 and NGC 3344, which are best-represented by a double power law. We fit over the dynamical range up to the scale where the correlation function becomes flat (R_0, the correlation length) as values of $1 + \omega(\theta) = 1$ corresponds to a random (non-clustered) distribution. Similar breaks in the power-law have been observed in both the correlation function of young stars in M 31 (Gouliermis et al., 2015a) and in the spectral correlation function of the Large Magellanic Cloud (LMC; Padoan et al., 2001), attributed to the disk scale height and corresponding to the transition between turbulent motions at small scales to disk dynamics at large scales.

3.6.2 Age Effects

The general distribution of stellar clusters across all the galaxies displays a fairly consistent picture: class 3 associations are very strongly clustered at the smallest spatial scales (Figure 3.6) and in general, have a younger median age distribution in comparison to the class 1 and 2 clusters (Figure 2.4). In order to investigate an age-dependency to the star-forming structures and the timescale over which the clustered substructures disperse, we divide the clusters by age to recover where the difference in clustering is maximal. We then recompute the correlation functions

Table 3.3. Power-Law Parameters for the Six Galaxies

Class	Number	A_1	α_1	β (arcsec)	A_2	α_2
			NGC 7793			
Class 1,2,3	350	14.0(6)	−1.16(5)		...	
			NGC 3738			
Class 1,2,3	281	3.27(3)	−0.394(12)		...	
			NGC 6503			
Class 1,2,3	298	3.97(19)	−0.46(6)		...	
			NGC 3344			
Class 1,2,3	391	4.7(3)	−1.68(13)	1.3	3.37(13)	−0.42(3)
			NGC 628			
Class 1,2,3	1264	4.81(5)	−0.75(4)	3.3	2.39(5)	−0.16(2)
			NGC 1566			
Class 1,2,3	1101	6.06(6)	−0.492(13)		...	

Notes. Power law fits for separations measured in angle (θ). Columns list the (1) Classification of stellar clusters; (2) Number in each classification; (3) Amplitude A_1 of the angular correlation function before the breakpoint; (4) Slope α_1 of the angular correlation function before the breakpoint; (5) Location of the break point β. Cluster classifications that are best-fit with a single power law do not have a breakpoint; (6) Amplitude A_2 of the angular correlation function past the breakpoint; and (7) Slope α_2 of the angular correlation function past the breakpoint. Numbers in parentheses indicate uncertainties in the final digit(s) of listed quantities, when available.

in these different age bins, determined individually within each galaxy, to explore how the age influences both the size and the clustering strength of the star-forming complexes within each galaxy.

Figure 3.8 shows how the clustering strength depends on the age of the clusters. For all galaxies and cluster classifications, there is a clear dichotomy between younger and older clusters, with clustering decreasing in strength for clusters older than 20–60 Myr. This timescale for the transition from a clustered distribution of the clusters at the youngest ages to a more smooth, homogeneous distribution occurs very rapidly, and in agreement with similar studies where the observed clustered distribution is slowly lost as the clusters/stars age. (e.g., Scheepmaker et al., 2009; Sánchez & Alfaro, 2009; Gouliermis et al., 2015b).

While small numbers for most of the galaxies is a limiting factor in our analysis, NGC 3738 shows the most striking influence of age on the clustering strength, where there is a dramatic decrease in the observed clustering for all the clusters within the irregular galaxy. NGC 628 and NGC 1566 also exhibit a substantial decrease in clustering with increasing cluster ages. NGC 7793, NGC 6503, and NGC 3344 only exhibit a slight change in the clustering with age, and within each of the cases, there are a lack of any older ($\gtrsim$50 Myr) clusters at the smallest spatial scales. Thus, the amount of clustering occurring at small spatial scales is very uncertain for the older clusters in most of the galaxies.

3.6.3 The Effect of Global Galactic Properties on the Correlation Length R_0

To compare the clustering results between the galaxies, we convert the length scales in arseconds to a spatial scale (parsec). We see in Figure 3.9 that the strength of the clustering for a fixed physical scale within each galaxy reflects the difference in structure of the galaxy, where bigger/brighter galaxies are associated with a larger

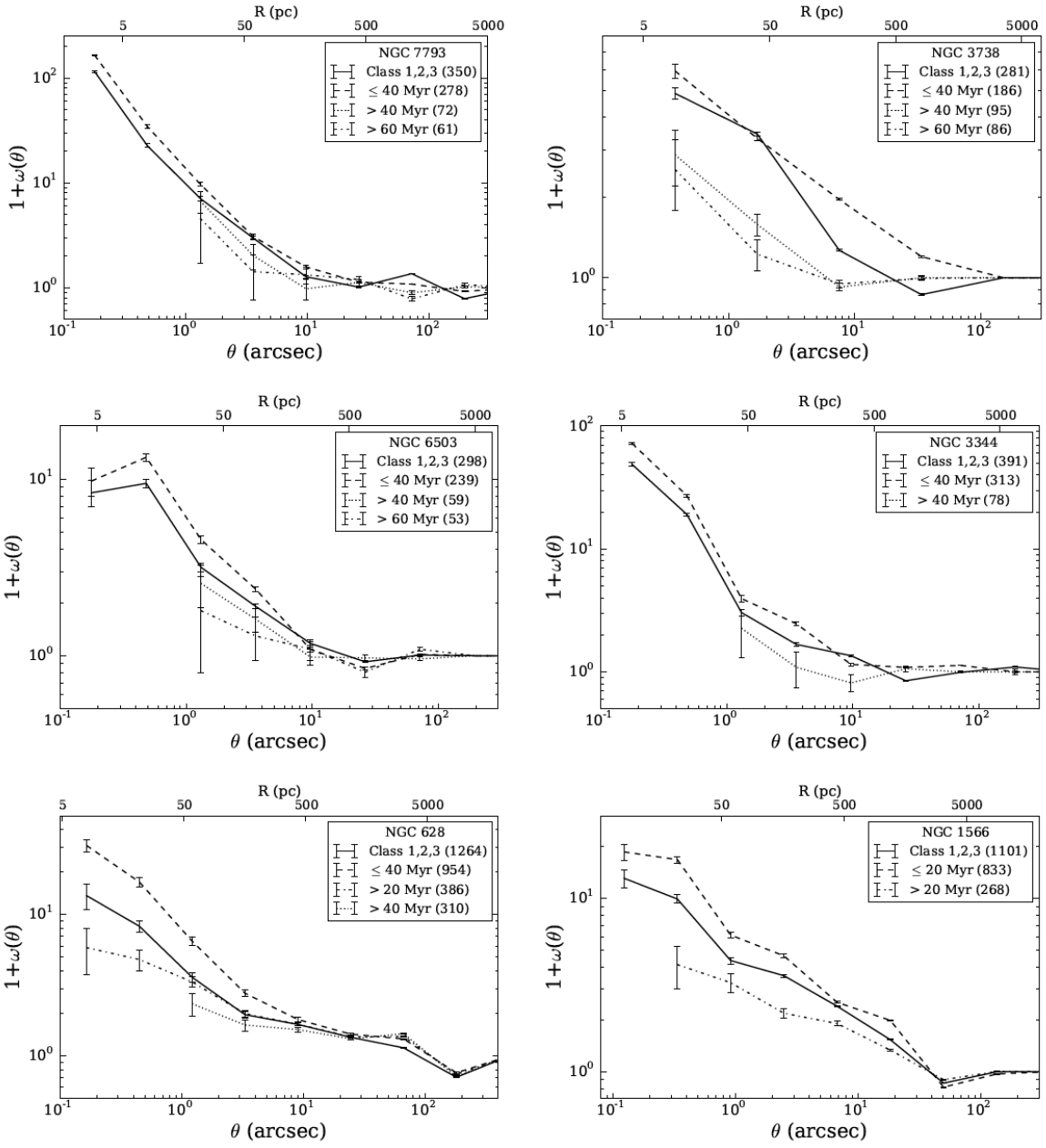

Figure 3.8. Two-point correlation function $1 + \omega(\theta)$ for all six galaxies divided by ages below/above specific age divisions determined by requiring that the clustering among the young star clusters is maximized. The strength of clustering increases when we consider the youngest clusters within each classification and affects the smallest spatial scales first within all galaxies.

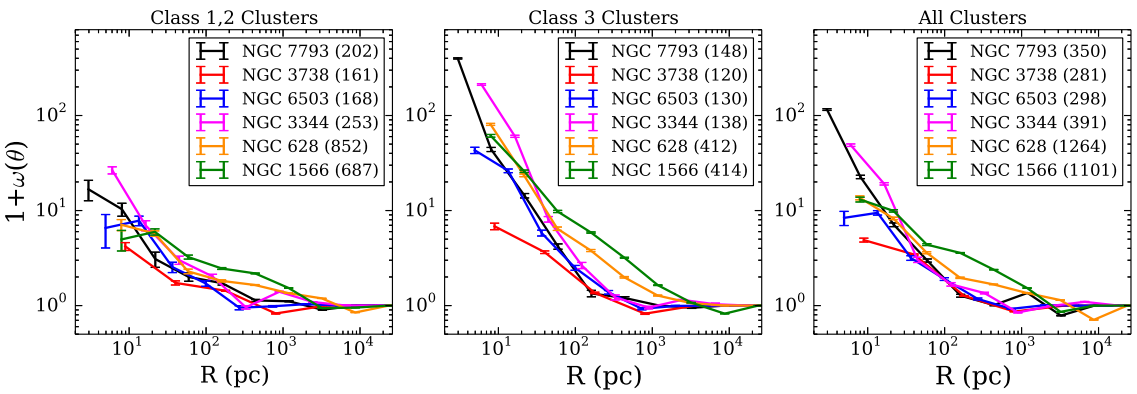

Figure 3.9. Combined two-point correlation functions $1 + \omega(\theta)$ as a function of physical scale. The plots are divided by cluster class: the class 1 and 2 clusters (left), the class 3 associations (middle), and all clusters (right). The numbers in parentheses show the number of clusters in each classification.

amplitude for a fixed distance between cluster pairs. The larger clustering amplitude of NGC 1566 compared to the rest of the sample is quite possibly observational and not physical, and much of the discrepancy may be due to adopting a distance that is further than assumed as the distance to the galaxy is quite uncertain.

To investigate the individual galactic properties on the clustering results, we compare the correlation length R_0 to different physical properties of each galaxy for all cluster classifications. The correlation length R_0 is defined where the correlation function flattens ($\omega = 0$), i.e. the distributions of star clusters become uncorrelated. The correlation length R_0 can describe the size of the typical star-forming complexes within a galaxy where the youngest star clusters were born and still reside. Thus, in addition to R_0 having an age-dependency, the correlation length may also depend on global galactic properties or any processes that influence the global star formation within these clustered star-forming structures.

Figure 3.10 shows the correlation length of each galaxy compared to different physical properties: morphological T–type, SFR, star formation rate surface density $\Sigma_{\rm SFR}$, and the stellar mass. In general, larger and brighter galaxies exhibit larger correlation lengths than smaller galaxies. This implies that the distribution of star-

forming regions within larger galaxies are hierarchical over longer spatial scales, both in the strength of the observed clustering and the extent of the complexes ability to survive over larger scales compared to what is observed in smaller galaxies. Galaxies with higher SFRs and lower morphological T numbers also have a general trend for larger correlation lengths. This indicates that the young stellar clusters in flocculent and dwarf/irregular galaxies with lower SFRs are distributed more homogeneously at the same spatial scales as observed in spiral galaxies, and thus, these systems globally have smaller star-forming complexes. There is no trend on the SFR surface density $\Sigma_{\rm SFR}$ of the correlation length, which is surprising as recent observations suggest that the SFR efficiency scales with the SFR surface density (Cook et al., 2016). While these observed trends are weak, limited by the small sample sizes, future studies with the full LEGUS sample of galaxies will help improve these trends and investigate why galaxies with larger SFRs correlate with large correlation lengths but no such trend is observed with the $\Sigma_{\rm SFR}$.

3.6.4 Combined Age Results

In order to compare the results and analyze how all clusters behave globally across all galaxies, we combine each cluster class across all galaxies by summing the weighted number of pairs as a function of physical separation, as shown in Figure 3.11. The improved statistics from the increase in clusters per class permits to better investigate common age effects among the galaxies. We fit a power law to each cluster class to the total sample of clusters and to the clusters above/below an age of 40 Myr, where that choice in age comes from the typical median age where each galaxy shows maximal clustering (see Figure 3.8). Table 3.4 shows the power law fits for the weighted average correlation for each class. Up to a breakpoint of 112 parsecs, the entire cluster population *averaged over all six galaxies* has a recovered power law slope of $\alpha = -0.83$, very similar to the fractal dimension of both stars and local

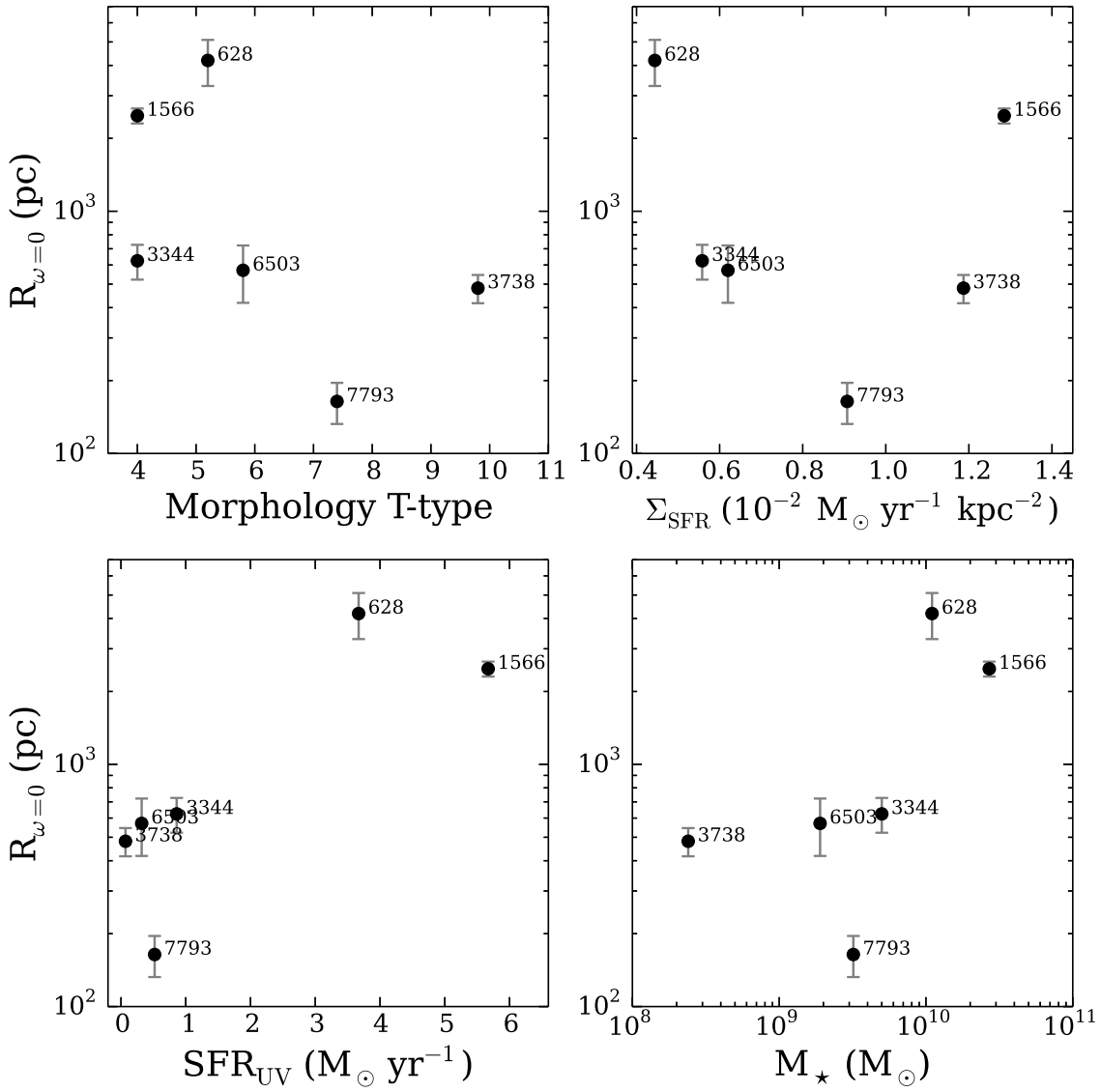

Figure 3.10. Correlation length $R_{\omega=0}$ as a function of various galactic parameters. $R_{\omega=0}$ is where the correlation function (Figure 3.9) becomes flat ($\omega = 0$), displayed as a function of the morphological type, star formation rate surface density $\Sigma_{\rm SFR}$, the SFR, and the stellar mass of each galaxy. The galaxy names are shown with their numbers only (without the leading NGC). There are general trends for larger correlation lengths in galaxies that are larger/brighter, have stronger spirals, and bigger SFRs. A larger sample of galaxies is necessary to more accurately describe how trends of galactic physical properties influence the clustering results. No trend is seen between the correlation length R_0 and $\Sigma_{\rm SFR}$.

interstellar clouds (see, e.g., Falgarone et al., 1991; Elmegreen et al., 2006; Sánchez & Alfaro, 2008), inherit their hierarchy at birth and preserve the natal gas pattern for a few tens of Myr. The global slope for all clusters is very similar to the slope recovered for the class 2 clusters of $\alpha = -0.85$, whereas the combined 1+2 clusters have a shallower slope of $\alpha = -0.57$ and class 1 clusters have a very shallow, single slope of $\alpha = -0.18$ across all spatial scales, despite the steep upturn at the smallest spatial scales, which arises from small numbers within that bin (see Figure 3.11 for the scatter). Associations, on the other hand, exhibit a much steeper slope with $\alpha = -1.12$.

As can also be seen in Figure 3.11, every classification is observed to display a broken power law, which is not necessarily true for each individual galaxy (Figure 3.6). The reason behind this is that we are increasing statistics by combining clusters from numerous galaxies, which enables us to recover more subtle effects. This is best described in Figure 3.9, where, for example, the correlation function for NGC 6503 and NGC 3738 becomes flat at a smaller spatial scale than what is observed for NGC 628, likely a difference that arises from the global galactic properties (see Figure 3.10). As a result, when the correlation function for all the galaxies are averaged together, there is a general break around 100-200 pc for all cluster types, where the slope beyond the break for all cluster types is nearly uniform ($\alpha \sim -0.2$) and only varies at the smallest spatial scales before the breakpoint.

Figure 3.12 shows how the ages of clusters influence the global weighted average correlation function across all six galactic environments. We divide the clusters into their class and further separate into age bins of older/younger than 40 Myr and compute the correlation function for each trial, taking the weighted average across all galaxies. As can be seen, for each class of cluster, the strength of the clustering disappears for clusters of older ages. While there is large scatter between all the

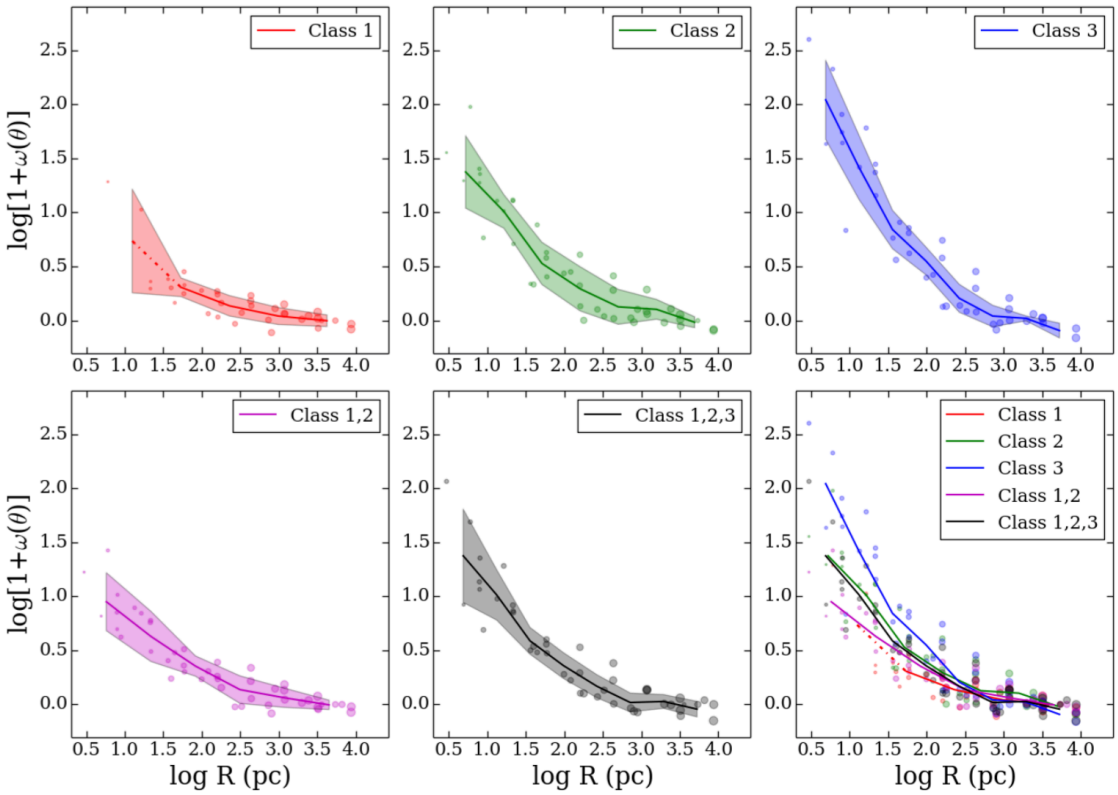

Figure 3.11. Weighted two-point correlation functions $1 + \omega(\theta)$ for all six galaxies as a function of physical scale (parsec) separated by cluster class. The solid line shows the weighted average of each class, weighted by number of pairs in each data point. The size of the data points represents the number of pairs for each point and the shaded region represents the 1σ scatter. We represent power-law fit for the class 1 clusters at short separations with a dashed line as we are limited by very small numbers within that bin for those clusters.

Table 3.4. Power-Law Parameters for Weighted Average

Class	A_1	α_1	β (arcsec)	A_2	α_2
All Ages					
Class 1	62(2)	−0.86(2)	59	3.2(3)	−0.14(3)
Class 2	63(3)	−0.73(7)	112	4.8(3)	−0.18(3)
Class 3	463(40)	−1.13(7)	112	6.6(4)	−0.23(07)
Class 1,2	25(4)	−0.58(4)	93	3.6(2)	−0.15(2)
Class 1,2,3	69(5)	−0.77(7)	112	4.0(3)	−0.17(4)
Age $\leq$ 40 Myr					
Class 1	52(10)	−0.66(14)	128	6.4(3)	−0.23(3)
Class 2	119(15)	−0.83(3)	112	8.1(5)	−0.26(2)
Class 3	565(18)	−1.15(13)	112	7.3(4)	−0.232(16)
Class 1,2	39(2)	−0.60(3)	186	3.51(13)	−0.139(19)
Class 1,2,3	115(6)	−0.85(3)	112	5.6(3)	−0.21(2)
Age > 40 Myr					
Class 1	2.45(14)	−0.13(2)		...	
Class 2	3.98(10)	−0.160(18)		...	
Class 3	3.3(3)	−0.161(14)		...	
Class 1,2	2.90(13)	−0.130(18)		...	
Class 1,2,3	2.61(15)	−0.164(6)		...	

Notes. Columns list the (1) Classification of stellar clusters; (2) Amplitude A_1 of the angular correlation function before the breakpoint; (3) Slope α_1 of the angular correlation function after the breakpoint; (4) Location of the break point β in parsecs; (5) Amplitude A_2 of the angular correlation function past the breakpoint; and (6) Slope α_2 of the angular correlation function after the breakpoint. All values are derived by using as independent coordinate R in parsecs. Numbers in parentheses indicate uncertainties in the final digit(s) of listed quantities, when available.

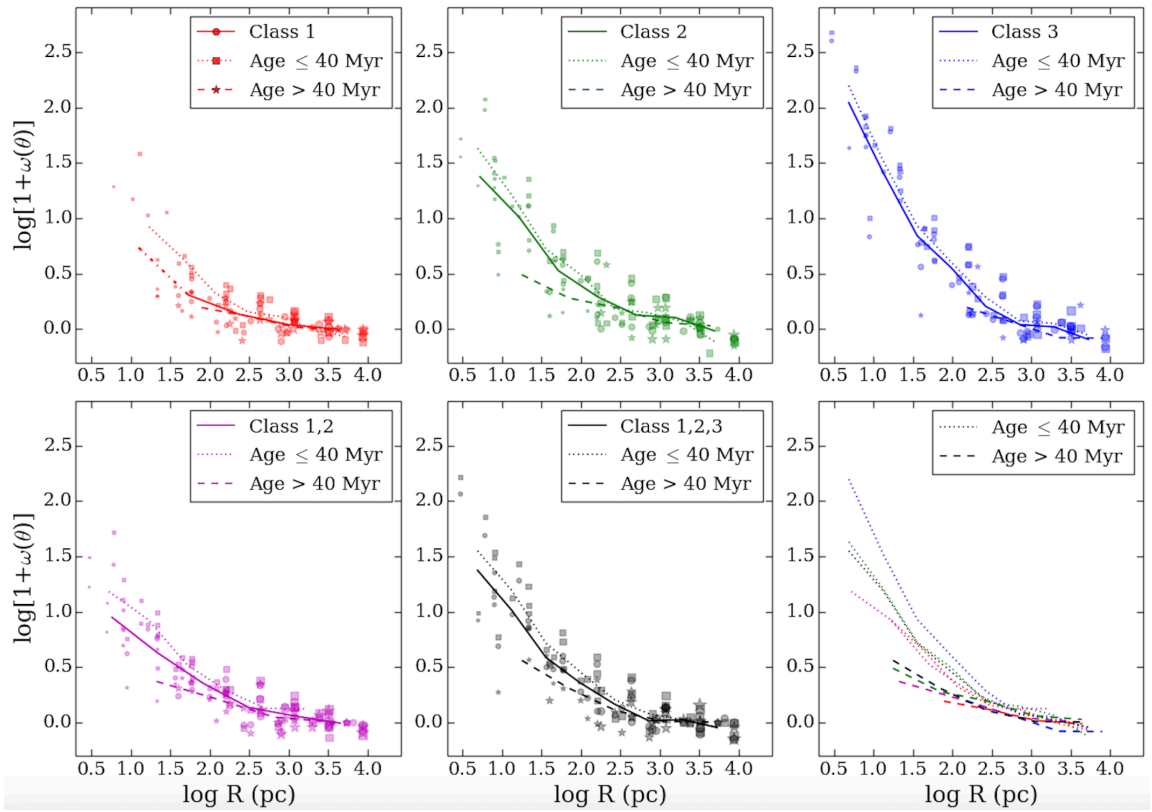

Figure 3.12. Weighted two-point correlation functions $1 + \omega(\theta)$ divided by age for each cluster class. The solid line (and circle symbols) show the weighted average of each class, weighted by number of pairs in each data point, the dotted line (and square symbols) shows the weighted average for clusters with ages less than 40 Myr and the dashed line (and star symbols) represents clusters older than 40 Myr. The size of the data points represents the number of pairs for each point. We represent the class 1 clusters at short separations with a dashed line as we are limited by very small numbers within that bin for those clusters. The last panel shows all the trends for clusters older/younger than 40 Myr.

galaxies, by ages of 40 Myr, the clustering strength has already decreased across all systems.

What is significant is that the behavior of the class 1 and 2 clusters older than 40 Myr show an appreciable flattened distribution out to small spatial scales ($\sim$10 pc), indicating that the clusters that survive past 40 Myr become randomized at all spatial scales. The class 3 associations, on the other hand, show an absence of older ($\gtrsim$40 Myr) clusters within $\lesssim$100 pc of each other, suggesting early dissolution. We conclude that the observed decrease in clustering strength is an effect of the cluster classes marking an age sequence in the cluster classes. More importantly, this common behavior is observed in six different galaxies characterized by quite different environments, as defined by their morphologies, which ranges from Sbc to Irregular, indicating that the clusters randomize on fairly short timescales.

3.6.5 Binary Clusters

The hierarchical structure of star formation results in the highest efficiency of star formation occurring within dense, crowded environments at the smallest scales. This has the possibility in resulting in binary or groupings of multiple star clusters at small scale lengths that are capable of interacting with other. The existence of binary star clusters have been observed in the LMC (Bhatia & Hadzidimitriou, 1988; Dieball et al., 2002) and the MW (Subramaniam et al., 1995; de la Fuente Marcos & de la Fuente Marcos, 2009).

Most cluster pairs are observed to have projected separations less than 20 pc, relatively young ages, and show small age differences between the pairs, indicative that the pair was formed at birth. Thus, binary star clusters are believed to be short–lived, with lifetimes of 10–100 Myr (Bhatia, 1990; de la Fuente Marcos & de la Fuente Marcos, 2010; Priyatikanto et al., 2016). Within the Magellanic Clouds, between $\sim$10–20% of clusters are potentially part of a binary/multiple system (see

Dieball et al., 2002, and references therein). The MW shows a lower prevalence of binary star clusters, however, this could be an observational effect, resulting from the relatively short lifetime of binary star clusters compounded by the paucity of young star clusters observed within the MW (Portegies Zwart et al., 2010).

Assuming a maximum cluster pair separation of 20 pc is required for a binary system to be bound and interacting, this separation is smaller than the closest cluster pairs we observe at ages older than 40 Myr (Figure 3.12, dashed lines in the bottom right plot). We conclude that none of our clusters reside in bound binaries or groupings at these separations past this timescale.

When we consider clusters with ages less than 40 Myr, about 6.5% of our total cluster sample (241 clusters out of 3685 total across all six galaxies) includes pairs closer than 20 pc, and could be binary clusters. Recent results of young star clusters in the same LEGUS galaxies show that pairs of clusters with small separations are more likely to be coeval and exhibit similar ages (Grasha et al., 2017b). The same findings of similar ages with small separations for pairs of clusters is also observed in the MW (Efremov & Elmegreen, 1998) and the LMC (de la Fuente Marcos & de la Fuente Marcos, 2009). Similar ages for small separations is an expected result from turbulence driving the hierarchical structure of the ISM (Elmegreen & Efremov, 1996). These complementary studies indicate that clusters favor being born close together at roughly the same time within groupings of multiple star clusters, resulting in the observed strong clustering at the youngest ages and the smallest separations. The clustered complexes disperse with age, effectively decreasing the observed clustering and removing the cluster pairs at the shortest separations (Figure 3.12). However, we are unable to dynamically constrain if close cluster pairs are gravitationally bound to each other, or if they will simply disperse with age.

3.7 Discussion

In this work we have identified stellar clusters and associations to investigate, their clustering properties with respect to the rest of the stellar clusters in each galaxy, and the timescale over which the clustered substructure that these clusters reside in dissipates. For a hierarchical star formation model, the turbulent-driven ISM induces scale-free star formation across all spatial scales, resulting in the clustering of star clusters and creating substructure on pc-scale scale lengths, dispersing and distributing over wider areas as they age (Portegies Zwart et al., 2010). Amid this general framework (e.g., Zhang et al., 2001; Gouliermis et al., 2015b), there are still many unanswered questions relating the properties of star clusters to the process of star formation, including (1) the timescale for the survival of the substructure; (2) the extent that the environment influences cluster formation, evolution, and disruption, and hence, these hierarchical star-forming complexes; and (3) how gravitationally bound stellar systems behave compared to loosely bound stellar systems (associations), as dense cloud systems and associations may differ as early as birth (see, e.g., Wright et al., 2014).

The two-point correlation function provides a means to investigate and quantify the stellar aggregate distribution within extended star-forming structures and accurately determine the randomization timescale for the components to become uncorrelated relative to each other. We find that the young clusters are clustered with respect to each other and that the distribution randomizes with increasing age. Despite spanning different environments, with morphological types from Sbc to irregular, similar dispersion timescales of a few tens of Myr are observed in all six galaxies.

The visual identification technique we implement allows us to distinguish between class 1 and 2 clusters, which are potentially gravitationally bound stellar systems, and class 3 associations, providing us the ability to investigate how these two different types of stellar aggregates evolve over time. The young age distribution and the

morphology, characterized by multi-peaks and asymmetries in the light distribution, indicates that these clusters are most likely stellar associations, unbound stellar systems that evaporate and disperse on short time scales (tens of Myr; Gieles & Portegies Zwart, 2011). This is additionally supported by Adamo et al. (2017) for star clusters in NGC 628, finding that class 3 associations tend to be systematically younger and to have a steeper mass function than class 1 and 2 clusters. We find that associations behave differently in their clustering properties compared to the class 1 and 2 compact clusters across all galaxies. Class 1 systems have relatively weak clustering and in general are the most massive cluster type and exhibit older ages. At the other end of the spectrum, class 3 associations display a very young age distribution and exhibit very strong clustering within each galactic system. In addition, class 3 associations display a steep decrease in their numbers with increasing age and exhibit a median age that is always below 10 Myr (Figure 2.4). The short timescales of survivability further reinforce the idea that these stellar systems are not likely gravitationally bound systems and they disrupt soon after formation, and thus, are still located near their birth sites, in agreement with the observed increase of clustering for class 3 sources. For a cluster of mass 10^4 $M_\odot$ and a radius of 5 pc, the crossing timescale within that cluster is a few Myr. Hence, for class 1 and 2 clusters with an average population age $\gtrsim$ 10 Myr, the crossing time is shorter than the age, and we can consider these clusters as potentially gravitationally bound systems.

Recent work on the number densities of the young stellar clusters in NGC 628 by Adamo et al. (2017) also shows a rapid decline in class 3 associations, disappearing on time scales of $\sim$50 Myr, comparable to those of hierarchically structured star-forming regions that we find in this study. Furthermore, class 1 clusters appear to already be less clustered at birth compared to the class 3 associations (Wright et al., 2014; Adamo et al., 2017), as seen in Figure 3.12: when we only consider clusters less than 40 Myr, class 1 clusters still display a flatter clustered distribution

compared to what is observed for class 3 associations. Thus, class 1 clusters do not trace the multi-scale hierarchy as directly as the associations, which may reflect the distribution of the densest peak (the putative birth sites of the bound clusters) within the ISM hierarchical structure. The increased clustering observed with young clusters – of all morphologies, though the clustering strength is strongest for the associations – supports the idea that all recently born stars and star clusters are formed within hierarchical star-forming complexes and their distribution reflects the fractal patterns of their parent molecular clouds. The relatively quick dispersal time of the associations could provide a significant contribution to the stellar field population of their host galaxies (see, e.g., Maíz-Apellániz, 2001).

Studies to constrain the properties of young star clusters, connecting them with their surrounding environment, are reasonably straightforward and achievable with extragalactic studies. It is more difficult to conduct studies of young star clusters within the disk of the MW galaxy, as the observations face complications such as foreground/background confusion, distance uncertainties, and severe line-of-sight reddening, which makes it arduous to detect young stellar systems. Despite these challenges in detecting and interpreting the observations of stellar clusters in the MW galaxy, the young stellar clusters within 2 kpc of our Solar System exhibit evidence of formation within hierarchical structures from the observed correlation of their age and separations, possibly driven by a turbulent ISM (de la Fuente Marcos & de la Fuente Marcos, 2009). The same observational signature is also present in the age and positions of the young star clusters in the LMC (Efremov & Elmegreen, 1998) and other local galaxies (Grasha et al., 2017b). Additionally, local star-forming environments in the MW display distributions that are consistent with predictions of hierarchically structured star-formation (Bressert et al., 2010) that is also observed in extragalactic studies, arising from the molecular cloud hierarchical structure (Elmegreen, 2008; Elmegreen et al., 2014).

This study finds that the youngest clusters (<50 Myr) primarily trace the spiral arms while older clusters are more randomly distributed throughout the galaxy. We conclude that star-forming complexes are relatively young, predominately located within the spiral arms of their galaxies, and are among the largest scales of the star formation hierarchy. The total amount of global star formation captured within clustered structures versus the amount of star formation that is 'unclustered' (i.e., the stellar field population) has important consequences on the mode (i.e., one universal mode or multiple, distinct modes) of star formation. We will investigate and address these issues in a forthcoming paper.

If star clusters have a velocity dispersion of $\sim$1 km s^{-1} (Larson, 1981) typical of their birth clouds, the clusters are capable of traveling 1 parsec per Myr. Shear effects, spiral motions, and colliding GMCs can compound the initial random motion, capable of boosting the velocity up to 10 km s^{-1} (see, e.g., Whitmore et al., 2005; Gustafsson et al., 2016). Stellar systems with a total random motion of 10 km s^{-1} can travel 100 pc within 10 Myr, making a 50 Myr timescale for randomization of star clusters within star-forming substructures entirely reasonable, resulting in a correlation length of a couple hundred of parsecs over that time period consistent with our findings.

The amplitude in the correlation function reflects the difference in structure between the galaxies, the amplitude scales with the magnitude of the host-galaxy: for a fixed scale, the faintest galaxy, NGC 3738, has the smallest value of $\omega + 1$, and the brightest galaxy, NGC 1566, has the largest value (Figure 3.9). The different slopes and clustering strengths exhibited by the different classes and ages of clusters imply different fractal dimensions and hierarchies for different galaxies, suggesting that each galaxy has its own influence in the formation and survival of star-forming structures. Despite the significant variations in the power-law slope – a proxy for the fractal dimension – among the galaxies, a global average over all galaxies with $\alpha = -0.83$ is consistent with the stellar clusters forming from fractal gas in a hierarchical star

formation framework. While we have found that the local environment does influence the distribution of the stellar systems – larger galaxies with higher SFRs have larger clustering correlation lengths – the exact influence of the local environment on observed differences in the global clustering results and on the nature of the 20–60 Myr randomization timescale – requires a larger sample of galaxies that span a broad range of physical parameters.

3.8 Summary and Conclusion

We present a study of the clustering of the young stellar clusters in six nearby galaxies, drawn from the LEGUS sample of 50 local star-forming galaxies with UV and optical data taken with WFC3/UVIS from HST, with the addition of archival optical ACS images. Taking advantage of both the high-angular resolution observations and the reliable measurements of ages and masses, we identify and visually classify the brightest star clusters (luminosities in the V-band brighter than -6 mag). All stellar clusters are divided into two broad categories of morphology: compact cluster candidates – according to our classification scheme, we consider these to be bound cluster candidates – and multi-peak systems, likely expanding or loosely bound stellar associations. In total, we identify 2323 compact clusters and 1362 associations, resulting in a total of 3685 stellar clusters in six galaxies of varying distance from 3.44 Mpc to 13.2 Mpc. Multiband cluster photometry is combined using deterministic single stellar population models to derive an age, mass, and color excess E(B–V) for every star cluster with errors of 0.1 dex or less for all derived measurements.

We implement the angular two-point correlation function to quantify the clustering of star clusters within each galaxy, finding that the correlation functions are well described with a single power-law slope, and in some cases, a double power-law. The general observed correlation between young stellar clusters decreases monotonically as a function of separation, as expected in hierarchical structures. We also find that

the clustering strength is stronger for the younger clusters compared to the older ones at the same separation lengths, which tend to be weakly or non-clustered within each galaxy and across all cluster types. Generally, the mass of the star clusters has little effect on the clustering strength; it is primarily governed by age. The near-flat slopes measured for the older clusters corroborate the effectiveness of the two-point correlation function to differentiate between a randomly distributed population from that of a clustered, fractal distribution.

We conclude that the observed decrease in clustering strength is the result of the clusters taking on more uniform positions throughout their galaxy, erasing the observed substructure with time. The timescale for the clustering erasure for the hierarchical clustering begins around 20–60 Myr, and the characteristic size of the clustered structures is about 100–300 pc across different galaxies. Thus, these identified star-forming complexes are very young.

The morphological classifications – compact clusters and stellar associations – not only provide robust cluster catalogs free of stellar contaminants, but also yield insight into the properties of the stellar systems. While both the stellar clusters and associations inherit the imprints of the hierarchical structure of the ISM at birth, their evolutionary paths differ from each other after their formation. The compact stellar clusters survive for longer time periods, with median ages of 9–40 Myr (and their ages are likely to become older, but we are limited by selection effects, see Adamo et al., 2017), whereas associations display very young median ages between 2–10 Myr. The difference in the physical properties of the bound stellar systems compared to the associations has an important impact on the clustering properties of the two stellar systems: associations are very strongly clustered in all six galaxies and in comparison, the (older) compact stellar clusters are less strongly correlated in their spatial distributions. We conclude that clusters display a more homogeneous distribution as a result of their birth sites being more homogeneously distributed.

Finally, we take the weighted average correlation of all the star clusters across all six galaxies as a function of spatial scale to investigate the average global clustering observed across the different environments. We fit a power-law to the weighted average correlation function of all the clusters in all the galaxies and measure a slope of -0.77, expected for a star formation process dominated by turbulence. The age-dependency to the clustering of star clusters found for individual galaxies becomes notably compelling when averaged across all galaxies: for all clusters with ages less than 40 Myr, we find that the clustering slope is flat and uniform, with values between -0.13 and -0.16, regardless of cluster type. This indicates that star clusters rapidly disperse from their hierarchically organized star-forming regions regardless of the galaxy type or size of the region. There are a large range of slopes for each individual galaxy, especially observed in the distribution of the youngest (age $\lesssim$40 Myr) clusters, indicative that there is a range of fractal hierarchies between galaxies and that the local environment can influence the initial formation of star clusters within the clustered structures. Indeed, initial results do suggest that larger correlation lengths and stronger clustering is exhibited in galaxies that are larger and have higher SFRs. Identifying the exact physical nature for the randomization timescale – and how the local and/or global environment, such as the difference in timescales for spirals versus ellipticals or the inner versus outer regions of a galaxy, influences the results – is an important goal for future work when a larger sample of galaxies becomes available.

CHAPTER 4

HIERARCHICAL STAR FORMATION IN TURBULENT MEDIA: EVIDENCE FROM YOUNG STAR CLUSTERS

This chapter[1] characterizes the correlation between the age and separation. This will enable us to test if the spatial distributions observed are driven by turbulence.

4.1 Abstract

We present an analysis of the positions and ages of young star clusters in eight local galaxies to investigate the connection between the age difference and separation of cluster pairs. We find that star clusters do not form uniformly but instead are distributed so that the age difference increases with the cluster pair separation to the 0.25–0.6 power, and that the maximum size over which star formation is physically correlated ranges from ∼200 pc to ∼1 kpc. The observed trends between age difference and separation suggest that cluster formation is hierarchical both in space and time: clusters that are close to each other are more similar in age than clusters born further apart. The temporal correlations between stellar aggregates have slopes that are consistent with predictions of turbulence acting as the primary driver of star formation. The velocity associated with the maximum size is proportional to the galaxys shear, suggesting that the galactic environment influences the maximum size of the star-forming structures.

[1]These results are based on research published by Grasha et al. (2017b), and is reproduced by permission of the AAS.

4.2 Introduction

Star formation is believed to be hierarchical in space and time as a result of turbulence and self-gravity (Scalo, 1985; Li et al., 2005; Elmegreen, 2010; Kritsuk et al., 2013) within the interstellar medium (ISM). The largest scales of the hierarchical structure are star-forming disks and within them are the accumulation of smaller components in the hierarchy: unbound cluster and stellar complexes, star clusters, and individual stars. Stars mostly form together in some sort of ensemble of clusters and associations (Lada & Lada, 2003) and it is currently thought that they rarely form in isolation (see, however, Lamb et al., 2016). Furthermore, star clusters are clustered with respect to each other (Efremov, 1978) in large complexes, imprinted with the fractal structure of the Giant Molecular Clouds (GMCs) from which they are born (Scalo, 1985; Elmegreen & Falgarone, 1996; Sánchez et al., 2010), slowly dispersing as the clusters age. Characterizing the correlation behavior of star formation across a range of galaxy properties provides a validation of this picture and a crucial understanding of how star formation progresses in both space and time across galactic scales.

Within the framework of star formation models that are regulated by turbulence, gas compression will break larger clouds into successively smaller ones, giving rise to the observed hierarchical structure (Elmegreen, 1999; Vázquez-Semadeni et al., 2009; Hopkins, 2013a). Such turbulent fragmentation processes, in addition to creating a hierarchy in the distribution of star clusters' properties (e.g., mass, size), will also create correlations with star cluster ages; in a structure where clusters have formed out of the same GMC, clusters that form close together will have closer ages compared to clusters that are further apart. Coeval cluster pairs have been observed both within the Large Magellanic Cloud (LMC; Bhatia & Hadzidimitriou, 1988; Dieball et al., 2002) and the Milky Way (MW; De Silva et al., 2015). Thus, within this picture, larger structures display older ages. This is interpreted as the duration of

star formation proceeding faster in smaller regions than in larger ones (Efremov & Elmegreen, 1998), in proportion to the turbulent crossing time.

Turbulence-driven star formation predicts that the age of star-forming structures will increase in proportion to the square root of the size (Elmegreen & Efremov, 1996). If the age-separation effects were driven simply by the stellar drift of the structure, a linear relation between the structures would be expected (Blaauw, 1952; Zwicky, 1953). A diffusion-driven expansion would produce a squared relation between age and size, in agreement with expectations from a random walk: the total distance traveled by a random scattering process is related to the square root of the number of random steps taken, where the number of steps is proportional to the time. These are testable predictions, when the ages of the structures can be determined with sufficient accuracy.

The age-separation distribution of star clusters was originally investigated within the LMC by Efremov & Elmegreen (1998), finding that the average age difference between pairs of star clusters increases with their separation as $\Delta t(\text{ Myr}) \sim 3.3 \times S(\text{pc})^{0.35\pm0.05}$ up to 780 pc. In a similar venue, de la Fuente Marcos & de la Fuente Marcos (2009) found local MW cluster pairs to also exhibit age differences with their separation with a 0.40 ± 0.08 power once the effects of incompleteness and cluster dissolution are taken into account. These relationships are similar to the size-linewidth relation of GMCs that show that the crossing time in a GMC increases as the square root of the size of the star forming region (Larson, 1981; Elmegreen & Efremov, 1996). Numerical simulations investigating the propagation of star formation in a turbulent medium by Nomura & Kamaya (2001) derive an age-separation relation to the 0.5 power for star clumps with separations >50 pc, comparable with the relation seen in both Efremov & Elmegreen (1998) and de la Fuente Marcos & de la Fuente Marcos (2009). The hierarchy in the star cluster structures is expected to have an upper limit in size; beyond this, separate regions form independently from one an-

other (Elmegreen et al., 2014), where turbulence can no longer regulate the cluster positions. Thus, the cluster relation should flatten at the maximum size of a galaxy's star-forming region, resulting in a turnover where the age difference between cluster pairs becomes a random function of the separation.

In this chapter, we analyze the average age difference between pairs of star clusters in eight local galaxies to investigate their hierarchical distribution, generalizing previous results beyond the MW and LMC. Our galaxy and cluster sample selection is described in Section 4.3. The age difference versus separation for cluster pairs is presented in Section 4.4. We discuss our results in Section 4.5 and summarize the findings in Section 4.6.

4.3 Sample Selection

The eight galaxies analyzed here were observed as part of the HST Treasury program Legacy ExtraGalactic UV Survey[2] (LEGUS; Calzetti et al., 2015), a Cycle 21 survey of 50 nearby star-forming galaxies observed in NUV, U, B, V, and I with WFC3/UVIS. The galaxy images have all been drizzled to a common scale resolution of the native WFC3 pixel size (0.0396 arcsec per pixel). The eight galaxies of this study (seven spirals and one irregular dwarf) are those for which star cluster catalogs have been produced by the LEGUS team at the time of this writing. General descriptions of the LEGUS survey, observations, and data products are available in Calzetti et al. (2015).

4.3.1 Star Cluster Identification and Selection

A detailed description of the cluster selection, identification, photometry, and SED fitting is presented in Adamo et al. (2017). We summarize here briefly the aspects that are important for the current paper.

[2]https://archive.stsci.edu/prepds/legus/dataproducts-public.html

Catalogs of candidate stellar clusters are first identified using SExtractor (Bertin & Arnouts, 1996) on a white-light image generated by using the five available photometric bands. This automated process produces a list of clusters with their position in pixel coordinates, number of filters the source was detected in, and the concentration index (CI; difference in magnitudes measured with an aperture of radius 1 pixel and an aperture of radius 3 pixels) of the source. The candidate clusters must satisfy two criteria in this automated process: (1) the CI in the V-band must exceed the stellar CI peak value; and (2) photometric error of $\sigma_\lambda \leq 0.35$ mag in at least two bands (the reference V band and either B or I band). Corrections for foreground Galactic extinction (Schlafly & Finkbeiner, 2011) are applied to the photometry. The concentration index cutoff between stars and star clusters are determined independently for each galaxy and for each pointing.

Before photometry is performed, the automated cluster candidates are down-selected through excluding any sources that are not detected with a 3σ detection in at least four of the five available photometric bands. The cluster candidates then undergo SED fitting to measure their physical properties (age, color excess E(B–V), and mass) using deterministic stellar population models (Yggdrasil; Zackrisson et al., 2011). The conditions placed on the photometry result in uncertainties of the ages and masses of $\sim$0.1 dex. The cluster SED fitting implement Padova isochrones, solar metallicity, and a range of extinction/attenuation curves. For this analysis, we select the SED fits performed with the starburst attenuation curve (Calzetti et al., 2000), which provides reasonable fits for both spiral and dwarf galaxies.

The sources in the automated catalogs include sources that are not star cluster candidates, such as background galaxies, foreground stars, multiple star pairs in crowded regions, and bad pixels and/or edge effects. To remove these contaminants, we visually inspect each cluster candidate that has an absolute magnitude brighter than -6 mag in the V-band. Each source is assigned one of four possible classifica-

tions: (1) class 1 clusters are compact and symmetrical, displaying a homogeneous color and a FWHM that is more extended than the stars within the galaxy, (2) class 2 clusters are compact and elongated, displaying elliptical light profiles; (3) class 3 clusters are non-compact and reveal multiple peaks on top of diffuse emission; and (4) class 4 sources are contaminants within the catalog; we remove these interlopers from the final *visually inspected* cluster catalog. The final cluster classifications are based on inspection by three separate individuals within the LEGUS team, excluding NGC 1566, where 368 of the clusters (33% of the total) were identified with only one human classification.

The clusters in this study comprise clusters with class 1, 2, and 3, which we further down-select to include only clusters with ages less than 300 Myr to ensure that we are not strongly affected by incompleteness due to evolutionary fading (see Section 4.4.4) or cluster dissolution. Table 4.1 lists the galaxies and number of young cluster candidates ($<$ 300 Myr) within each galaxy.

We exclude the clusters within the strong star-forming region in the north-east corner of NGC 5194 to avoid introducing bias from the structure, bringing the total clusters with ages <300 Myr from 1171 to 821. A detailed investigation on the star clusters of NGC 5194 will be performed in a forthcoming paper (Grasha et al., 2018a; Messa et al., 2018).

4.3.2 Incompleteness and Selection Effects

Completeness tests are performed and discussed in Adamo et al. (2017) on NGC 628 at a distance of 9.9 Mpc; we briefly summarize those tests here. Completeness tests on the cluster catalogs show that the catalogs are complete down to the CI cutoff of each galaxy, as listed in Table 4.1, corresponding to clusters with $R_{\rm eff} \sim 1$ pc at 10 Mpc. With the exception of NGC 1566, NGC 628 is the most distant galaxy in our list (Table 4.1), thus we expect the other catalogs to be complete to a smaller ef-

Table 4.1. Galaxy Properties for the Δt–R Study

Galaxy	Morph	Dist. (Mpc)	R_{25} (kpc)	Morph.	SFR_{UV} ($M_\odot$ yr^{-1})	$N_{clusters}$	CI_{cut} (mag)	Scale (pc/px)	Incl. (deg.)	P.A. (deg.)
NGC 7793	SAd	3.44	4.65(0.11)	7.4(0.6)	0.52	343	1.3(e)/1.4(w)	0.66	47.4	90
NGC 1313	SBd	4.39	5.8(0.3)	7.0(0.4)	1.15	673	1.4	0.83	40.7	39
NGC 3738	Im	4.90	1.78(0.08)	9.8(0.7)	0.07	254	1.4	0.94	40.5	141
NGC 6503	SAcd	5.27	5.44(0.13)	5.8(0.5)	0.32	283	1.25	1.1	70.2	120
NGC 3344	SABbc	7.0	7.23(0.17)	4.0(0.3)	0.86	388	1.35	1.3	23.7	140
NGC 5194	SAbc	7.66	12.5(0.3)	4.0(0.3)	6.88	821	1.35	1.5	51.9	163
NGC 628	SAc	9.9	13.1(0.4)	5.2(0.5)	3.67	1205	1.4(c)/1.3(e)	1.9	25.2	25
NGC 1566	SABbc	13.2	15.9(0.8)	4.0(0.2)	5.67	1099	1.35	2.5	37.3	32

Notes. Columns list the: (1) galaxy name, ordered by increasing distance; (2) morphological type as listed in NED, the NASA Extragalactic Database; (3) distance from Calzetti et al. (2015); (4) optical radius of the galaxy R_{25} from NED; (5) RC3 morphological T–type as listed in Hyperleda (http://leda.univ-lyon1.fr); (6) Star Formation Rate (SFR), calculated from the GALEX far-UV, corrected for dust attenuation as described in Lee et al. (2009); (7) number of star clusters with ages < 300 Myr; (8) concentration index (CI) cutoff between stars and star clusters (see Section 4.3). NGC 7793 and NGC 628 have different CI cutoffs for each pointing [central (c), east (e), or west (w)], labeled separately; (9) pixel resolution in parsec $pixel^{-1}$; (10) inclination, in degrees, from Calzetti et al. (2015); and (11) position angle (P.A.), measured in degrees (0 to 180) from the North to the East. A P.A.= 0 corresponds to a galaxy with the longest axis oriented along the North-South direction. Numbers in parentheses indicate 1σ uncertainties in the final digit(s) of listed quantities, when available.

fective radius. A CI cut is necessary in order to remove stellar contaminants from the catalog. Size distributions of star clusters have been shown to display a log-normal distribution that typically peaks around 3 parsecs across galaxies (Ryon et al., 2017), and hence, we do not expect any bias within the catalogs as the typical cluster radius is well above the detection limit of 1 parsec. As a result, the distance of the galaxy is not expected to impact the cluster recovery fraction. Crowding effects between the inner and outer regions of a galaxy is also negligible on the cluster catalogs.

In addition to using deterministic models to deriving the physical properties of the stellar clusters, the cluster properties are also derived using a Bayesian analysis method to stochastically-sample cluster evolutionary models, as performed by Krumholz et al. (2015a) through implementation of the Stochastically Lighting Up Galaxies (SLUG; da Silva et al., 2012; Krumholz et al., 2015b) code. Stochastic effects of the IMF become progressively more important for deriving accurate ages and masses for clusters with masses below $\sim$5000 $M_\odot$. Krumholz et al. (2015a) finds in

both NGC 628 and NGC 7793 that the global properties of the cluster populations are relatively similar between the conventional deterministic and stochastic fitting procedures. Section 4.4.6 shows how our main results are nearly the same when the cluster properties (age and mass) are derived with stochastic models. As the biases between the two methods (deterministic and stochastic) of deriving the cluster properties averages out for the mean of the entire population (Krumholz et al., 2015a), our results are fairly insensitive to the assumption of incomplete sampling of the IMF at the low-mass cluster range with deterministic fitting.

The LEGUS cluster catalogs are fundamentally luminosity limited, as seen in the rising top envelope of the age-mass diagram of clusters (Grasha et al., 2015; Adamo et al., 2017), biasing the catalogs towards younger clusters. However, as shown in Adamo et al. (2017) for NGC 628 and NGC 5194, a cut in absolute V magnitude of -6 mag is more conservative than the depth of the V-band image, i.e., -6 mag is more luminous than the detection limit. In some parts of this paper (e.g., Section 4.4.2) we apply mass cuts to the clusters, as a mass-limited sample will prevent bias in the age distribution caused by a luminosity-limited sample of young clusters. At 10 Mpc, an age limit of 200 Myr yields a complete cluster sample down to masses of 5000 $M_{\odot}$ (Adamo et al., 2017), i.e., close to the mass limit where stochastic sampling of the IMF begins to become important. Our absolute V magnitude limit is determined by the detection limits of the LEGUS sample, which aims at selecting down to $\sim$1000 $M_{\odot}$, 6 Myr old clusters with color excess E(B–V)= 0.25 (Calzetti et al., 2015). The only exception to the above is NGC 1566, located at a greater distance than NGC 628. For NGC 1566, a higher absolute V magnitude cut of -8 mag needs to be applied, resulting in a smaller age range for which our catalogs are complete to 5000 $M_{\odot}$.

In Section 4.4.2 below, we test for selection effects by repeating our analysis after making mass cuts at both 3000 $M_{\odot}$ and 5000 $M_{\odot}$, which yield complete samples up to ages of 100 Myr and 200 Myr, respectively. If completeness affects or drives the

maximum possible age of a star-forming region, we would expect to see that in the results because the age limit changes as a function of the mass cut (see Figure 3 of Grasha et al., 2015), and thus, we would expect a correlation between the age to which the sample is complete and the maximum age in the observed $\Delta t - R$ correlation. In Section 4.4.2, we find that the effect of mass cuts on the results is negligible and well within the uncertainties of the data. Therefore, all analyses throughout this paper include all clusters < 300 Myr and we do not apply a mass cut in order to strengthen available statistics, as our results are not affected by completeness issues, either in age or mass.

4.3.3 Deprojection of the Galactic Disk

It is essential to properly account for and deproject the positions of the star clusters to accurately use their spatial distributions free from the effect of projection. We deproject the pixel coordinate positions of all the stellar clusters in each galaxy by assuming that each galaxy can be well-described with an axisymmetric flat rotating disc. The deprojection of the cluster positions is performed in a two step process.

The x and y pixel coordinates are first rotated by the position angle α, determined by the orientation of the observed field of view for each galaxy. The rotated coordinates x' and y' are determined as $x' = x\cos\alpha + y\sin\alpha$ and $y' = -x\sin\alpha + y\cos\alpha$. The rotated coordinates are then deprojected by the inclination angle i of the galaxy as $x_{\rm deproject} = x'$ and $y_{\rm deproject} = y'/\cos i$. We use the deprojected coordinate positions $x_{\rm deproject}$ and $y_{\rm deproject}$, converted to a physical scale within each galaxy, for all calculations involving the positions of the clusters using the spatial resolution of a pixel at the distance of each galaxy, as listed in Table 4.1.

4.4 Analysis and Results

4.4.1 $\Delta t - R$ Relation

Following the work of Efremov & Elmegreen (1998), we consider all clusters younger than 300 Myr within each galaxy, and compute the absolute age difference Δt between each pair as a function of their deprojected (see Section 4.3.3) physical separation (R) for nine equally spaced separation bins in log space. We allow pair separations up to 8000 parsecs in each galaxy, performing a least-squares fit to the data following a double power-law as,

$$\Delta t(\text{ Myr}) = \begin{cases} A_1 \times R(\text{pc})^{\alpha} & R \leq R_{\max} \\ A_2 \times R^0 & R > R_{\max}. \end{cases} \tag{4.1}$$

with α the slope before the breakpoint at a maximum separation $R_{\max}$ where the age–separation trends flatten and A_1 and A_2 the intercept before and after the breakpoint, respectively. We force the slope at scale lengths larger than $R_{\max}$ to zero and the location of $R_{\max}$ is a free parameter in the fit, determined where chi-squared is minimized. We take into account and propagate the uncertainties of individual cluster ages for each cluster pair for the Δt calculation when making the mean age-difference at each bin. The number of bins are chosen as a compromise between available statistics, where we require a minimum of 10 data points in each bin, and resolution, though on average, the smallest bin for each galaxy has over 20 pairs. The influence of the choice of bin size is investigated further in Section 4.4.5.

Assuming that cluster pairs are members of the same region, the observed turnover is a measure of a galaxy's maximum star-forming region size, and the age differences at the same separation can be used as proxies for the duration or lifetime of star formation within the region. Table 4.2 lists the power-law fit parameters for each galaxy.

Table 4.2. Age Difference and Spatial Separation Results

Galaxy	Intercept (A_1)	Slope (α)	Max Age (Myr)	$R_{\rm max}$ (pc)	Velocity (km s^{-1})	$v_{\rm S}$ (km s^{-1})
NGC 7793	4.0(0.3)	0.47(0.06)	48(19)	203(30)	4.0(1.1)	2.0(1,2)
NGC 1313	16(1)	0.26(0.08)	85(26)	585(183)	6.1(1.4)	2.4(3)
NGC 3738	9.0(0.5)	0.24(0.04)	45(9)	869(152)	15(2)	5.0(4)
NGC 6503	1.7(0.3)	0.6(0.2)	62(21)	275(104)	5.7(2.5)	1.9(5)
NGC 3344	1.0(0.2)	0.6(0.2)	41(14)	338(131)	8.5(3.1)	3.7(6)
NGC 5194	6.8(1.0)	0.36(0.07)	83(20)	947(231)	13(3)	5.5(7)
NGC 628	7.4(0.6)	0.33(0.07)	66(18)	788(179)	13(4)	6.9(7,8)
NGC 1566	2.3(0.2)	0.41(0.14)	30(9)	508(179)	20(7)	10(9)

Notes. Power-law fits for the equally spaced separation bins in log space for the galaxies in Figure 4.1. Columns list the: (1) galaxy name, ordered by increasing distance; (2) intercept A_1 from Eq. 4.1; (3) slope α from Eq. 4.1; (4) A_2, the average age difference between cluster pairs at the turnover; (5) size $R_{\rm max}$ of the star forming region at the turnover; (6) traveling velocity (size/average age) of the region at the turnover; and (7) average shear velocity between cluster pairs at $R_{\rm max}$ and references for the rotation curves used to derive the shear. Numbers in parentheses indicate 1σ uncertainties in the final digit(s) of listed quantities, when available.
References for rotation curves: (1) – Carignan & Puche (1990); (2) – Dicaire et al. (2008); (3) – Ryder et al. (1995); (4) – Oh et al. (2015); (5) – Greisen et al. (2009); (6) – Verdes-Montenegro et al. (2000); (7) – Daigle et al. (2006); (8) – Combes & Becquaert (1997); (9) – Agüero et al. (2004).

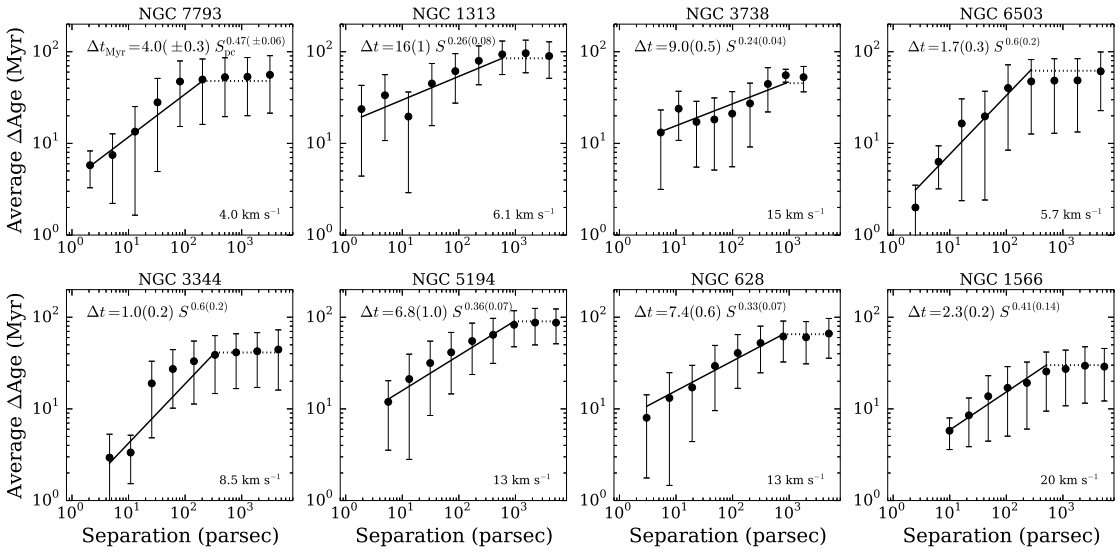

Figure 4.1. Age difference between cluster pairs as a function of separation between the cluster pairs. The black data points and error bars show the average age and 1σ spread for each cluster pair in each separation bin. The bins are logarithmically-spaced, which provide optimal sampling (see Section 4.4.5). The average age difference between cluster pairs increases systematically with their separation, indicating that the duration of star formation is longer for larger regions and that younger star-forming structures are less extended than older regions. A linear fit to the $\Delta t - R$ relation is shown in the solid black line and the dotted line shows where the relationship is flat beyond the breakpoint $R_{\rm max}$. The lower right hand corner of each panel lists the average velocity at the breakpoint in the double power law.

Figure 4.1 shows the average age-separation as a function of increasing separation between cluster pairs in addition to the power-law fit to each galaxy (eq. 4.1). The maximum separation of correlated cluster pairs corresponds to sizes ∼200–1000 pc and average age separations of ∼20–100 Myr.

From Figure 4.1, we calculate a velocity from the ratio of the size of the turnover $R_{\rm max}$ to the average age difference at that size. This velocity is the average speed at which turbulence moves through the star-forming region. The velocity is listed in the lower right hand corner of each panel in Figure 4.1.

4.4.2 Impact of Mass Cuts on the $\Delta t - R$ Relation

If completeness in age and/or mass drives the observed separation of cluster pairs, we would expect to observe a correlation between the limiting age we are sensitive to at a given mass cut and the maximum age at the turn over point. We repeat the analysis in Section 4.4.1 after applying different mass cuts for the galaxies with enough statistics to do a proper study: NGC 1313, NGC 5194, NGC 628, and NGC 1566 to see if we see a trend for an increase in the age accompanying increases in the mass cut. Figure 4.2 shows negligible increase in the maximum age of the $\Delta t - R$ relation, about 10% or less, which is well within the errors. We do note that NGC 628 does show a higher increase in $R_{\rm max}$ at the largest mass cut, about 25%. However, this is still within the error bars, and, as shown in Section 4.4.7, the $\Delta t - R$ correlation for NGC 628 disappears when we randomize the clusters positions and ages, demonstrating that the correlation is not an effect of selection biases.

Additionally, for all the galaxies in Table 4.2, the maximum age is always shorter than the completeness age limit of 200 Myr at the mass cut-off of 5000 $M_{\odot}$. We conclude that the luminosity-limited nature of our catalog, and therefore, potential biases induced by this on the ages, are not driving the observed $\Delta t - R$ correlation.

4.4.3 Shear and Global Galactic Properties

To better understand the impact of galactic shear on the turbulent velocities, we remove the effect of shear in all galaxies at the turnover separation $R_{\rm max}$ between all clusters in all galaxies. First, we calculate the angular velocity as $\Omega = V_{\rm rot}/R_{\rm g}$, where $V_{\rm rot}$ is the rotational velocity at the galactocentric position $R_{\rm g}$ of each cluster. Then we take the average galactocentric distance $R_{\rm g}$ multiplied by the difference in angular rates, $\Delta\Omega$, for each pair of clusters within the $R_{\rm max}$ separation bin. This is the relative velocity from shear for each pair, $V_{\rm S}$, which we subtract from the turbulent velocity to get a velocity that is corrected for the effect of shear. The relative velocity from

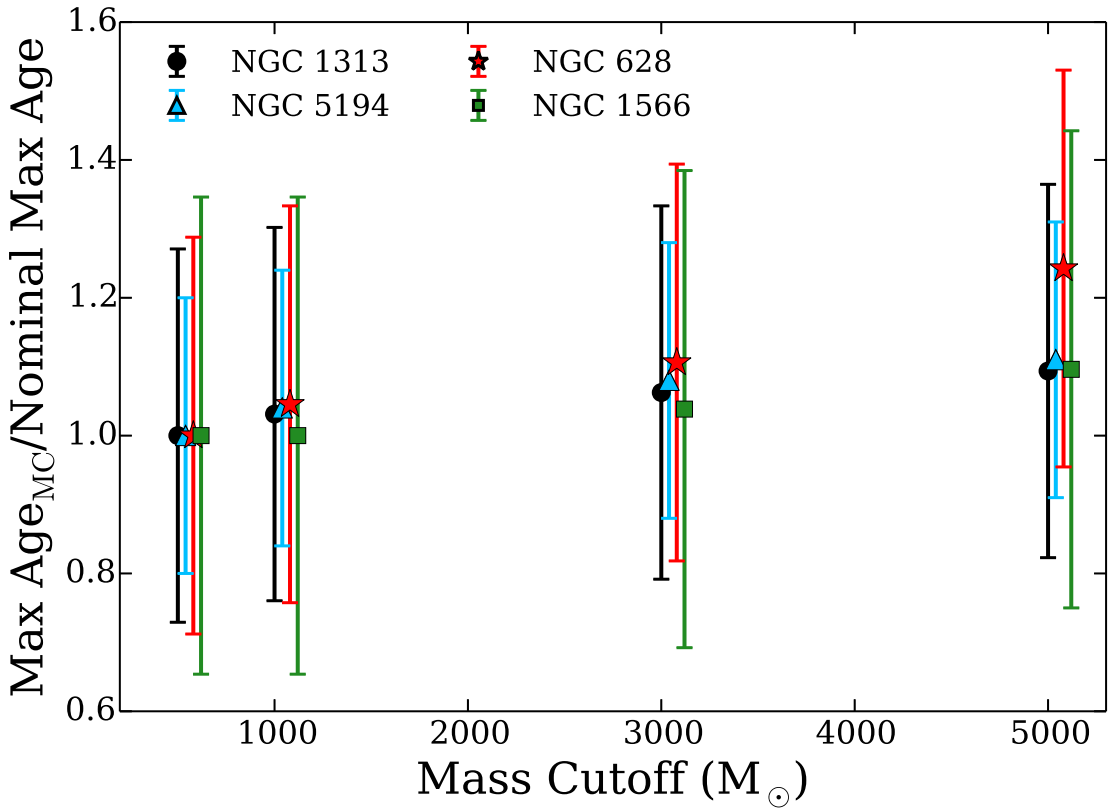

Figure 4.2. Maximum age for each galaxy at different mass cutoff limits normalized by the nominal maximum age. These results are derived for a minimum mass limit of 500 $M_\odot$ (Table 4.2) as a function of the cutoff mass at 1000, 3000, and 5000 $M_\odot$. If the $\Delta t - R$ relation is a result of the luminosity limit of the sample we would expect an increase in the maximum age as we increase the mass cutoff. There is a slight upward trend, but the values are well within the errors, implying that the catalog completeness is not responsible for driving the $\Delta t - R$ relation. The points corresponding to different galaxies are shifted slightly along the x-axis as to not lie on top of each other.

shear for each galaxy is listed in Table 4.2 along with the references from which the rotation curves are taken. V_S quantifies the average difference in azimuthal velocities on the scale of the largest separation and informs on how much shear influences the relative velocity at the edge of a star-forming region. We find that in general the contribution of shear to the measured velocity between cluster pairs is at most only slightly greater than the 1σ error of the velocity.

The velocity from the $\Delta t - R$ relation, when normalized to the velocity component produced by shear (Figure 4.3), is independent of the turnover size $R_{\rm max}$. Thus the maximum speed associated with the largest structure of star formation in a galaxy is linked directly to the velocity difference from shear within the same structure. These results indicate that while turbulence is quite possibly the dominant process defining the $\Delta t - R$ relation, there are dependencies on the environment of the host galaxy, which affect the measured maximum sizes, age-differences, and velocities.

Figure 4.4 shows the trend of the velocity at $R_{\rm max}$ for each galaxy with the dust-corrected star formation rate (SFR), R_{25} optical radius, and morphological Hubble T–type. We find that larger spirals (correlation coefficient $r = 0.98$) with higher SFRs ($r = 0.86$) and smaller T–type morphologies ($r = 0.83$) exhibit larger $R_{\rm max}$ velocities, although the trends with SFR and T-type are weak. This is not unexpected as larger galaxies are known to have larger star-forming complexes (Elmegreen et al., 1996). The irregular galaxy, NGC 3738, does not follow the trend that we observe for spirals. Because the cluster catalogs are complete to clusters more extended than $R_{eff} \sim$ 1 parsec (Adamo et al., 2017) at a distance of 10 Mpc, and we expect this to apply to all galaxies excluding NGC 1566 (13 Mpc), we do not expect any significant bias in our cluster analysis due to the distance of the galaxies. The apparent correlation between an increase in the velocity (or maximum size $R_{\rm max}$) with increasing galaxy distance (Table 4.2) is an observational bias in our sample as the smallest galaxies in our sample are at the closest distances (Figure 4.5).

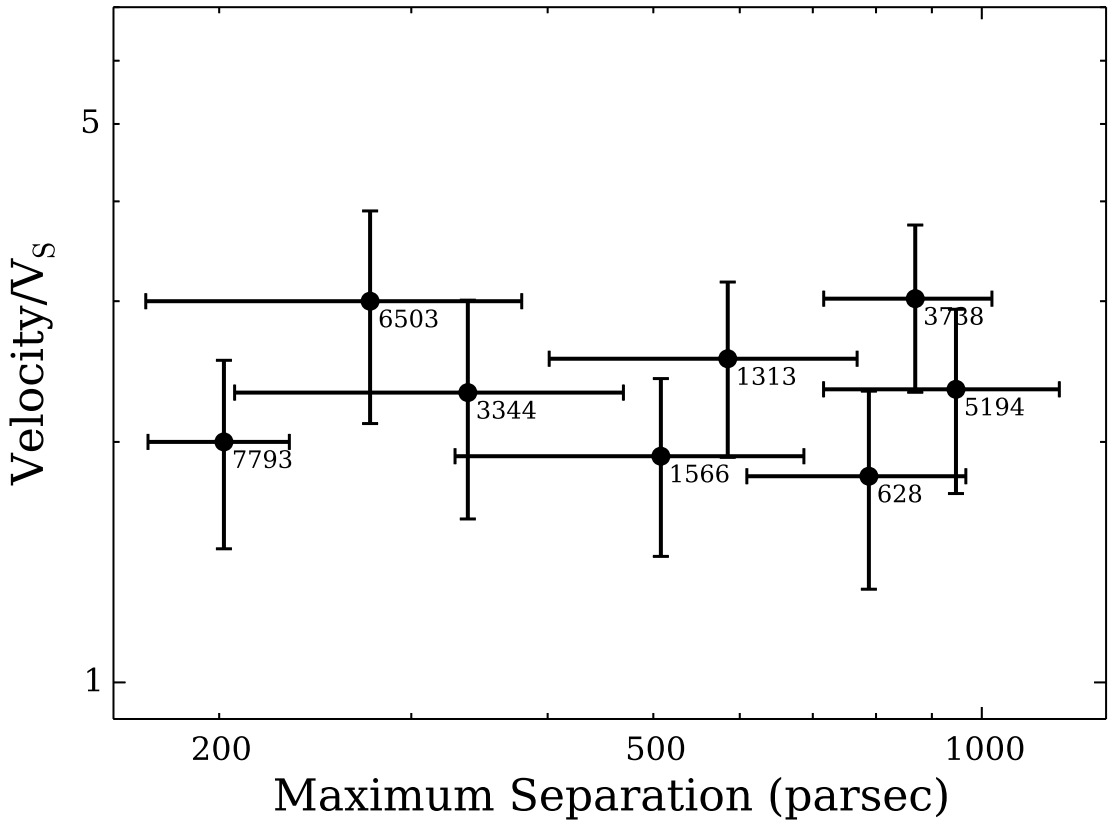

Figure 4.3. Velocity at $R_{\rm max}$ divided by the velocity difference due to shear $V_{\rm S}$ as a function of the turnover size between cluster pairs $R_{\rm max}$. The ratio $Velocity/V_{\rm S}$ is independent of $R_{\rm max}$, implying that shear is responsible for determining the maximum size of star-forming regions. Error bars display the standard 1σ error in the mean.

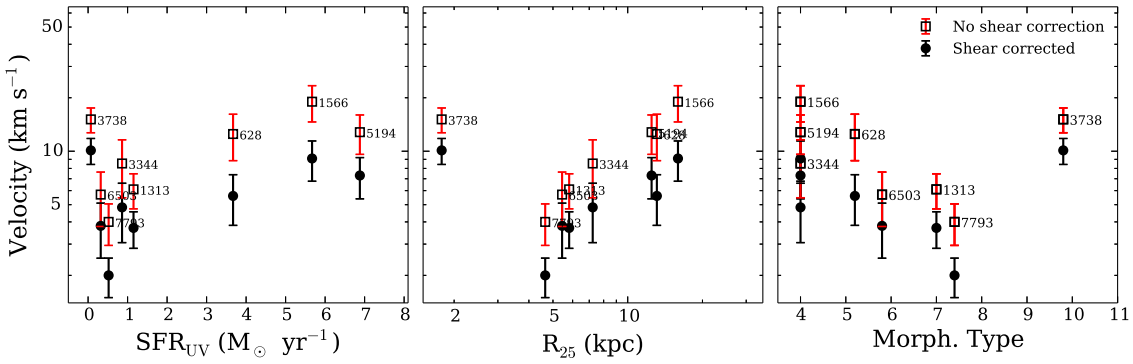

Figure 4.4. Velocity at $R_{\max}$ as a function is several galactic parameters. Shown are the UV SFR (left), R_{25} optical radius (center), and morphology T–type (right). Open squares with red error bars have not been corrected for shear while black solid circles with error bars have been corrected for shear. Larger spiral galaxies and higher SFRs correlate with larger velocities; the Irregular galaxy NGC 3738 does not follow the spirals' trend.

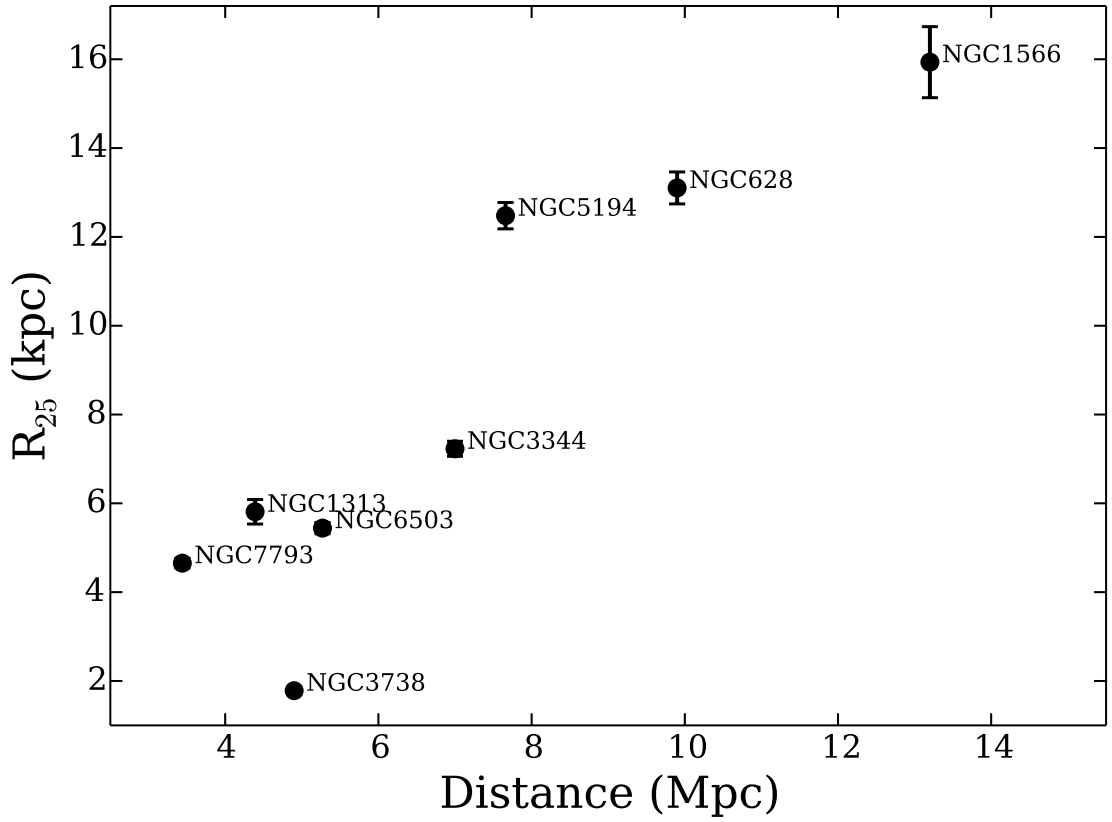

Figure 4.5. R_{25} radius as a function of the distance for each galaxy. The largest galaxies, and therefore, the largest traveling velocities calculated from the $\Delta t - R$ relation, are found at the largest distance within our survey.

4.4.4 Cluster Evolution

Clusters are disrupted rapidly due to evaporation, merging, and tidal field interactions (for a review, see Portegies Zwart et al., 2010), removing up to 90% of clusters within each age dex. At older ages, cluster samples start to suffer from incompleteness due to evolutionary fading and rising mass completeness limits at ages older than a few hundred Myr (e.g., Fouesneau et al., 2014; Adamo et al., 2017). Incompleteness due to embedded clusters, confusion with associations, and infant mortality is important in the age range 1–10 Myr (Gieles et al., 2011). We can test the impact of cluster evolution on our results by inspecting the $\Delta t - R$ relation for clusters with ages above ∼10 Myr and for ages up to 100 Myr compared to 300 Myr. We compute the $\Delta t - R$ relation in this manner for four of our galaxies with the most populous cluster catalogs in our sample with the best number statistics available: NGC 1313, NGC 5194, NGC 628, and NGC 1566.

Based on the considerations above, we expect that clusters with ages of 10–100 Myr are those for which the effect of both cluster disruption, fading, and dynamical mass loss will have minimal impact on the $\Delta t - R$ relation. Figure 4.6 shows this along with the $\Delta t - R$ relation for the age range 1–300 Myr and 1–100 Myr. Table 4.3 lists the power-law fits of the four galaxies available for all the different age ranges using logarithmic bins like in Figure 4.1. The best-fit values (e.g., $R_{\rm max}$) do change between the two binning methods of Figure 4.1 and Figure 4.6, but the values are well within the error. The scatter is significant in the age range of 10–100 Myr because of small number statistics, and the correlation should be represented with a single power-law. However, we elect to keep the double power-law fit in order to enable easy comparison of this age bin with the other two.

Evolutionary effects on the clusters already appear to have impacted the age–separation results by 300 Myr, flattening the slopes and increasing the characteristic star formation timescale compared to clusters that are only 100 Myr old in NGC 628,

NGC 5194, and NGC 1313. On the other hand, NGC 1566 shows a different trend where the measured slope decreases when we lower the age limit from 300 to 100 Myr. However, the slope of NGC 1566 marginally increases when we further decrease the age limit to the range 10–50 Myr, the only system were we have enough statistics to test the relation in the range for clusters in this smaller age range. The ability to perform tests that pinpoint the age where evolutionary effects become important depend on the size of the available catalog. Not accounting for the effect of cluster evolution may act to decrease the underlying slope.

An inspection of the clusters in the range 10-100 Myr compared to 1–100 Myr in Figure 4.6 – effectively removing clusters that are subject to violent gas expulsion (e.g., Baumgardt et al., 2013) – show that the $\Delta t - R$ relation becomes shallower as clusters younger than 10 Myr are removed. The relation steepens marginally within NGC 1566, but as the total age range is narrow, the scatter is considerable and we do not consider it a significant effect.

The clusters in the range 10–100 Myr for both NGC 1313 and NGC 5194 show an extremely shallow relation compared to what is observed for the rest of the systems. We are limited by small number statistics for these two systems for this age range, having only 228 and 253 clusters, respectively, compared to 447 and 545 clusters for NGC 628 and NGC 1566, respectively. Thus, the scatter is significant and part of the observed shallowing of the slope is likely due to small number statistics for the shortest spatial scales with the youngest age differences.

In general, the slope steepens when we reduce the upper age limit. de la Fuente Marcos & de la Fuente Marcos (2009) also found that the MW cluster pairs exhibit a steepening slope after correcting for cluster dissolution and incompleteness. Our catalogs are most sensitive to clusters with young ages: between 45%–66% of the star clusters in each galaxy have ages $\lesssim$10 Myr. The effect of removing clusters with ages <10 Myr impacts the age–separation relation in each galaxy differently, but in

all cases, it is apparent that the youngest, most recent star forming regions drive the observed $\Delta t - R$ relation at the smallest scales. While the initial star cluster formation may be driven by turbulence, after the star clusters have aged just a few tens of Myrs, the imprint of turbulence on their age–separation relation appears to have diminished.

The recovered slope of NGC 628 exhibits a significant flattening for clusters older than 100 Myr which is not observed for the other galaxies. This agrees with the findings of Adamo et al. (2017) that cluster disruption is not important for ages between 10 and 200 Myr in NGC 628, hence the dramatic change in the slope from 100 to 300 Myr. Both NGC 5194 (Gieles et al., 2005) and NGC 1566 (Hollyhead et al., 2016) also show slow dissolution rates of ~ 100 Myr for their cluster populations, unlike NGC 1313, which exhibits a fast cluster disruption timescale of 25 Myr (Pellerin et al., 2007).

Figure 4.7 shows the ratio of the traveling velocity to the velocity difference due to shear $V_{\rm S}$ as a function of the turnover size for the different age ranges of Figure 4.6. The four galaxies with enough clusters to perform an analysis of the dependence of the $\Delta t - R$ relation with age are also the largest galaxies with the largest derived $R_{\rm max}$ value. While there is a general trend for a higher ratio of $V/V_{\rm S}$ at the turnover point when tracing younger star-forming regions within a given galaxy (i.e., a lower contribution of shear within the youngest, smallest complexes), selecting the same age range for all galaxies results in a flat trend between $V/V_{\rm S}$ and $R_{\rm max}$. As long as we compare the results using the same age range for the clusters, the ratio $V/V_{\rm S}$ appears to be constant across the selected cluster age range.

The difference in the observed slopes or maximum age is not an artifact of galaxy distance because, comparing Table 4.1 and Table 4.2, there is no correlation between maximum correlation age and distance. Shear inside the galaxy may be an important

factor in determining the duration of star-formation and size of regions, as shown in Section 4.4.3 and Figure 4.3 and Figure 4.7.

4.4.5 Binning Method and the Dependence of $R_{\rm max}$ on Age

In this section, we investigate the choice of bin size and method of binning on the $\Delta t - R$ results, specifically the sensitivity of the location of the turnover point $R_{\rm max}$ to the binning selection.

The total spread of possible age differences between cluster pairs is significant, especially at the largest separations. While the upper envelope of age differences only increases with increasing separation, there will always be cluster pairs with $\Delta t = 0$ values at all separations. As performed in both Efremov & Elmegreen (1998) and de la Fuente Marcos & de la Fuente Marcos (2009), we bin the data, finding that the best representation of the underlying distribution is when the binning method is performed in equal logarithmic-spacing, as shown in Figure 4.1. However, in order to fully understand how our choice of binning methods may potentially impact the location of $R_{\rm max}$, which drives the resulting velocity at that location, we also compute the $\Delta t - R$ correlation for regular intervals of separations of 20 parsecs for our selected age range of 1–300 Myr.

Figure 4.8 shows the results: as expected, when we increase the number of bins, the scatter between individual bins increases. Additionally, choosing the bins such that they are for regular intervals of separation greatly limits the resolution at small separations due to sparse numbers of clusters. The derived values of $R_{\rm max}$ remain consistent relative to those listed in Table 4.2. The $\Delta t - R$ relations in Figure 4.6 for different age ranges can be used to investigate the dependency of the size of a star-forming region, $R_{\rm max}$.

For all galaxies, when the maximum age of the clusters is lowered and we only consider clusters younger than 100 Myr for the $\Delta t - R$ relation, we find that the size of

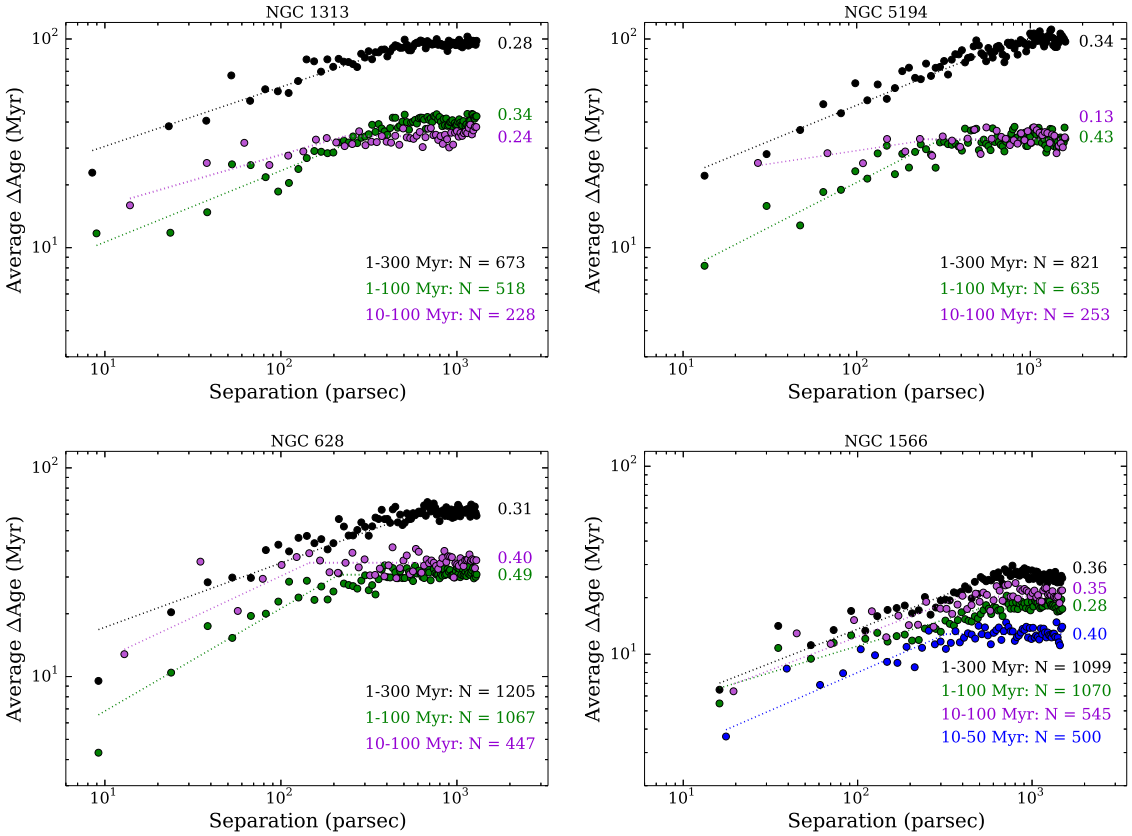

Figure 4.6. Three different age ranges to test for the effect cluster evolution on the $\Delta t - R$ relation for an aging cluster population. We do this by computing the age difference between cluster pairs as a function of separation in logarithmic scale between the cluster pairs of four galaxies, divided into: 1–300 Myr (black), 1–100 Myr (green), and 10-100 Myr (purple). NGC 1566 contains enough clusters to also investigate the trend at ages 10–50 Myr (blue). The range 10-100 Myr is selected to avoid the effects of violent gas disruption (ages <10 Myr) and cluster dissolution and fading (ages $\gtrsim$ 100 Myr). A fit to the $\Delta t - R$ relation is performed and the numbers to the right of each age range show the value of the slope α, with the full power-law fits listed in Table 4.3. The scatter present in the 10-100 Myr range is significant and should be represented with a single power-law; however we perform the fits in this age using a double power law like in the other cases, to enable direct comparisons among the three age ranges considered. Older clusters result in larger, more extended star-forming sizes, increasing the size of $R_{\rm max}$. Cluster dissolution and fading also has the greatest impact for clusters older than a couple hundred Myr, increasing the average separation between pairs and flattening the slope in the $\Delta t - R$ correlation as clusters age.

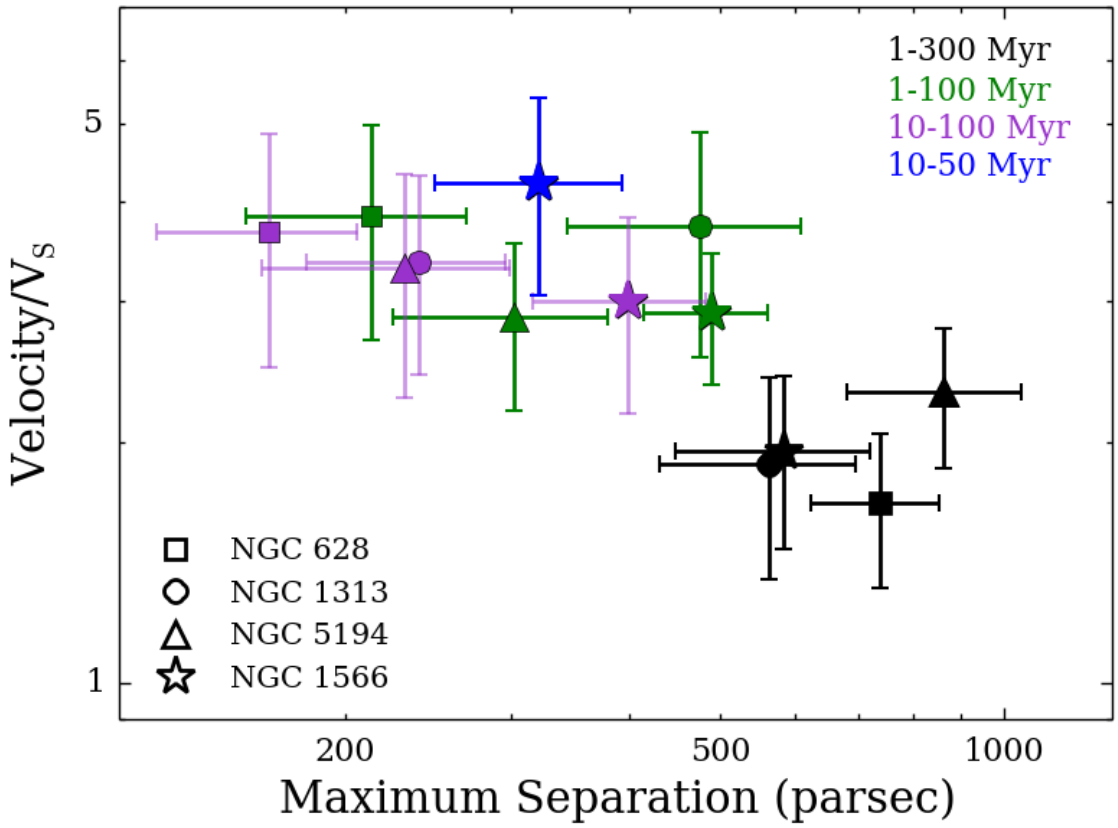

Figure 4.7. Velocity at $R_{\rm max}$ divided by the velocity difference due to shear $V_{\rm S}$ calculated at the different age ranges from Figure 4.6. The black symbols show the $V/V_{\rm S}$ ratio for clusters in the age range 1–300 Myr, green show the range range 1–100 Myr, purple show the age range 10–100 Myr, and blue show the age range 10–50 Myr (only for NGC 1566). Individual galaxies are represented by different symbols: NGC 628 (squares), NGC 1313 (circles), NGC 5194 (triangles), and NGC 1566 (stars). The range 10–100 Myr would be better represented by a single power-law (Figure 4.6); however, we elect to fit that age range with a double power-law to enable direct comparison with the other two age bins. We highlight this choice for the 10–100 Myr age bins by marking the error bars for this age bin with lighter lines. For a given galaxy, younger (i.e., smaller) star-forming regions have a total traveling velocity which is a few times larger than the shear, while approaching the $V_{\rm S}$ value as the maximum region size increases. Within each age range, however, $V/V_{\rm S}$ remains constant from galaxy to galaxy. Error bars display the standard 1σ error in the mean.

Table 4.3. Age Difference and Spatial Separation Results for Varying Age Ranges

Galaxy	A_1 (Myr)	Slope (α)	Max Age (Myr)	$R_{\rm max}$ (pc)	V/V$_{\rm S}$
			10–50 Myr		
N1313			...		
N5194			...		
N628			...		
N1566	1.3(0.3)	0.40(0.09)	13(3)	326(72)	4.2(1.2)
			10–100 Myr[1]		
N1313	9(2)	0.24(0.06)	34(8)	239(57)	3.4(0.9)
N5194	16(4)	0.13(0.03)	33(9)	258(68)	3.3(1.0)
N628	5.0(1.3)	0.40(0.11)	35(10)	141(40)	3.7(1.2)
N1566	2.4(0.5)	0.35(0.07)	20(4)	408(93)	3.0(0.8)
			1–100 Myr		
N1313	4.8(1.3)	0.34(0.07)	40(11)	486(132)	3.7 (1.2)
N5194	2.8(0.7)	0.43(0.08)	33(6)	294(77)	2.9(0.7)
N628	2.2(0.6)	0.49(0.13)	31(8)	211(56)	3.8(1.1)
N1566	3.0(0.4)	0.28(0.04)	18(2)	493(73)	2.9(0.5)
			1–300 Myr		
N1313	16(3)	0.28(0.07)	96(23)	560(132)	1.9(0.5)
N5194	10.0(1.3)	0.34(4)	99(13)	842(180)	2.3(0.5)
N628	7.8(1.2)	0.31(0.05)	64(10)	736(114)	1.7(0.4)
N1566	2.1(0.4)	0.40(0.07)	26(5)	590(136)	1.9(0.5)

Notes. Power-law fits for the logarithmic separation bins for the galaxies in Figure 4.6. Columns list the: (1) galaxy name, ordered by increasing distance; (2) intercept A_1; (3) slope α; (4) A_2, the average age difference between cluster pairs at the turnover, $R_{\rm max}$; (5) size $R_{\rm max}$ of the star forming region at the turnover for the age range from Eq. 4.1; and (6) the ratio of the traveling velocity to the velocity different from shear (Figure 4.7). Each value is listed for four different age ranges: 10–50 Myr, 10–100 Myr, 1–100 Myr, and 1–300 Myr. Only NGC 1566 was investigated for the range 10–50 Myr. Numbers in parentheses indicate 1σ uncertainties in the final digit(s) of listed quantities, when available.

[1]The clusters in the 10–100 Myr range of Figure 4.6 are best-described with a nearly flat, single power-law with slope with no breakpoint. However, in order to compare the ratio of the velocity from shear in Figure 4.7, we keep the double power-law fit.

the star-forming region $R_{\rm max}$ decreases (Figure 4.6 and Table 4.3). This is expected as the youngest clusters will trace the regions of the most recent star formation, occurring at the smallest size scales. Thus, increasing the age of the star-forming region increases the extent of the structure that the star clusters are tracing, and as expected, increases the maximum age of the region as well. Only half our sample have large enough cluster catalogs to investigate the age-dependency on the $\Delta t - R$ relation. The changing of the $\Delta t - R$ relation with different age ranges provides important insight on the effect of cluster dissolution and fading on the results, as discussed in the previous section.

4.4.6 Sensitivity of the $\Delta t - R$ Relation on Stochastically Derived Properties

As mentioned in Section 4.3.2, the cluster properties for two of the LEGUS galaxies, NGC 7793 and NGC 628, have been derived through the implementation of SLUG (Krumholz et al., 2015a). SLUG provides stochastically derived posterior probability distribution functions (PDFs) for ages, masses, and extinctions of the clusters, assuming different priors for both the cluster mass function and dissolution rate. We utilize the full posterior PDFs of the cluster candidates in NGC 7793 and NGC 628 to investigate the implications on the $\Delta t - R$ relation of deriving the ages stochastically compared to traditionally deterministic SED modeling. We ensure that the underlying physical assumptions (e.g., metallicity, dust attenuation models) are consistent between the two codes.

Figure 4.8 shows the $\Delta t - R$ relation for NGC 7793 and NGC 628 where the ages are determined both from deterministic techniques and with the stochastic modeling code SLUG (Krumholz et al., 2015a). We take the stochastic age of each cluster as the peak of the marginal posterior probabilities.

We find that the global trends from ages determined stochastically are similar to what is observed for cluster properties derived from traditional SED fitting techniques. There is a slight decrease in the average age separations for clusters with properties derived stochastically within NGC 7793 and no difference is observed within NGC 628. Any difference is well within the uncertainties on the age differences as determined from deterministic models. We conclude that there is no concern of any bias introduced in our results that is caused by the traditional deterministic fitting procedures.

4.4.7 Randomization Tests

4.4.7.1 Shuffling the Ages

We perform two tests to assess whether the correlation found between the age and separation of cluster pairs is truly statistically significant or is an effect of our sample's luminosity-limited selection. In the first test, using the 1205 star clusters in NGC 628, we keep the real cluster positions and shuffle the ages randomly and recompute the $\Delta t - R$ relation, using the same bin choices as Figure 4.6 for the 1–300 Myr clusters. Figure 4.9 shows 50,000 iterations of randomly shuffling the ages and keeping the separations the same. We see that the act of shuffling the ages causes the correlation to disappear, implying that the observed $\Delta t - R$ relation is statistically significant and is not a random effect.

For each iteration, we also determine the slope, where we only fit cluster pairs up to 500 pc to remove any uncertainty in the location of $R_{\rm max}$. In the 50,000 trials, there was never a random trial with a slope in agreement within 1σ of the slope obtained for the real data of NGC 628 (0.33 ± 0.07). This provides further evidence of a physical effect where clusters tend to form coevally, while distant pairs tend to form at different times; the effect is not the result of chance alignment or biases in the survey selection.

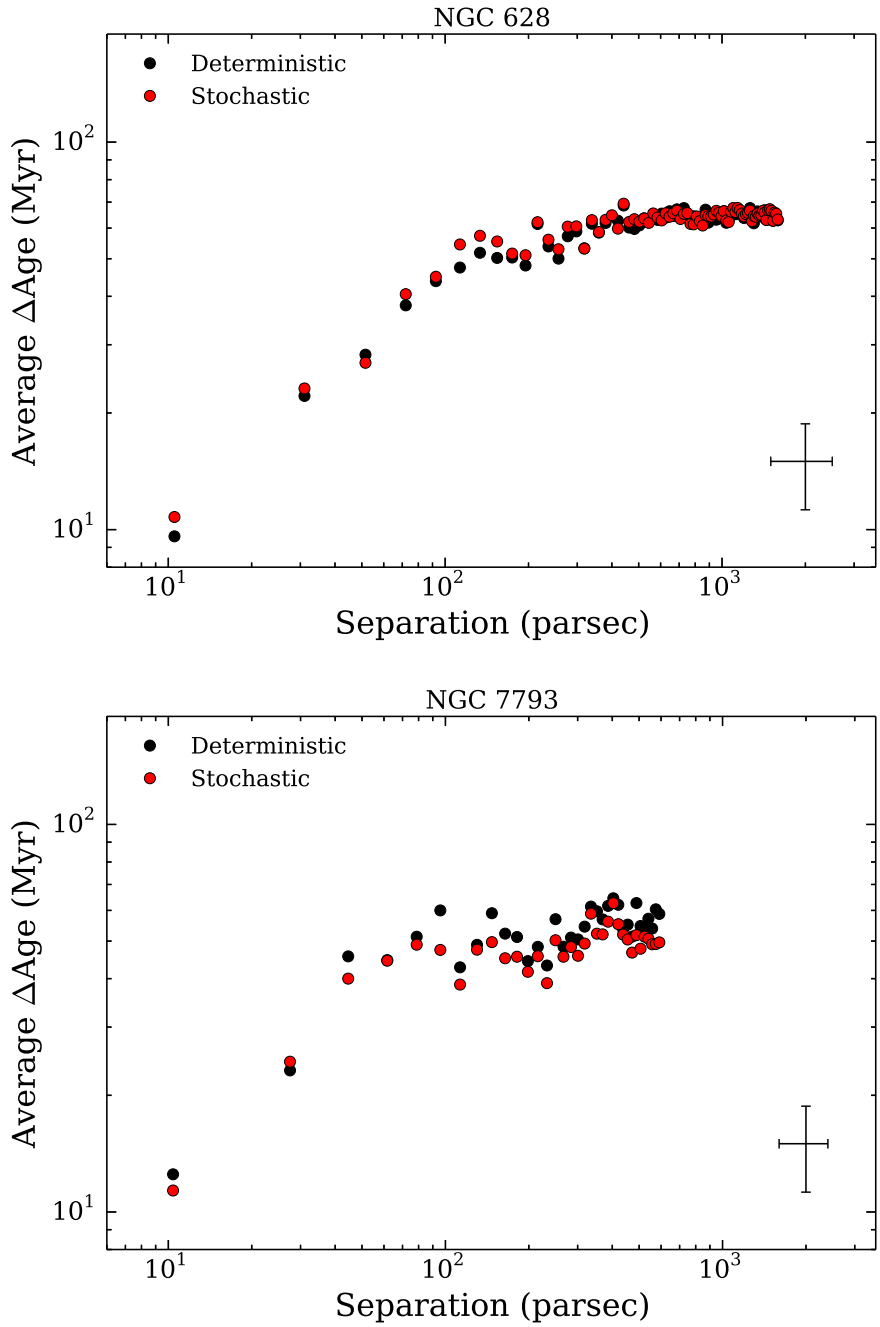

Figure 4.8. Testing the effect of stochastic versus deterministic models on the $\Delta t - R$ relation. The age difference between cluster pairs as a function of separation of the cluster-rich galaxy NGC 628 (top) and the cluster-poor galaxy, NGC 7793 (bottom). The average age difference Δt is shown as a function of their physical separation R for linear intervals of separation of 20 parsecs. The black points show the results if the ages are determined with deterministic models and the red points show the results for the ages derived with stochastic modeling (Section 4.4.6). The bottom right shows the average error for each point. The difference between the $\Delta t - R$ relation for the stochastic models are well within the scatter of the deterministic models.

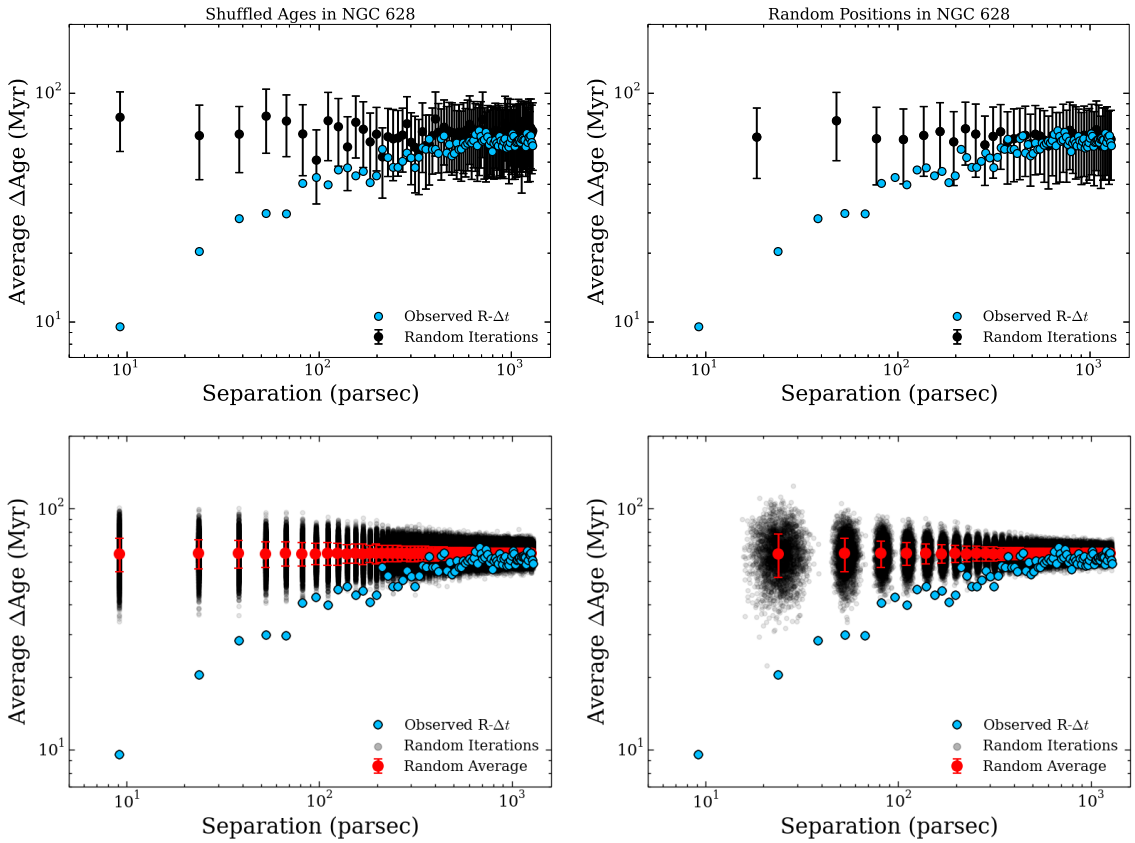

Figure 4.9. $\Delta t - R$ trend for NGC 628 with randomized positions and shuffled ages. The plots show the age difference between cluster pairs as a function of separation in logarithmic scale between the cluster pairs of NGC 628. Blue points show the trend for the real data of NGC 628. *Top left*: One example of the $\Delta t - R$ relation where we randomly shuffle the ages among the real cluster positions. *Top Right*: One example with randomizing the cluster positions within NGC 628 (see Section 4.4.7.2). *Bottom Left*: The gray points show 50,000 iterations where cluster ages are shuffled randomly among the real cluster positions, and the red points show the average of the shuffled points with the 1σ error in the mean. *Bottom Right*: The gray points show 15,000 iterations when cluster positions are randomized within the galaxy, and the red points show the average of the randomized points with the 1σ error in the mean. There are a lack of close cluster pairs with randomized positions as real cluster positions are highly clustered, especially at young ages. The correlation disappears for randomized ages and positions across all pair separation lengths, suggesting that the $\Delta t - R$ trend found in Figure 4.1 and shown in blue for the clusters is statistically significant and determined by a real physical effect.

4.4.7.2 Randomizing the Positions

The second test we perform to check whether the age distribution is driving the observed behavior is assigning a random spatial position within the footprint of NGC 628 to each of the fitted cluster ages. Figure 4.9 shows the $\Delta t - R$ relation results for 15,000 iterations of randomizing the positions of the star clusters. This test of randomizing the positions is more involved than simply shuffling the ages as we have to ensure that enough cluster pairs are within the smallest separation bin in Figure 4.9. Because we use the actual cluster positions when we randomize the ages in Section 4.4.7.1, we did not have this problem in that first test.

The observational bins in NGC 628 contains 58 cluster pairs within the smallest separation bin centered at 10 pc, and 86 cluster pairs with separations in the 20 pc bin. We place the requirement of a minimum of 25 pairs in the least populated bin, though the results do not change if we lower the threshold. In an initial run of 50,000 iterations, using the same bin sizes as Figure 4.6, only two trials populated the minimum of 25 cluster pairs in the 10 pc bin, far short from the 58 cluster pairs that are observed in the true cluster positions of NGC 628 in the same bin. We already infer from this that the positions of the clusters are not random and the $\Delta t - R$ relation cannot be the result of a statistical effect and the probability of a chance alignment is nearly negligible. Indeed, the clusters of NGC 628 do show a very clustered distribution, especially at young ages, compared to randomized populations (Grasha et al., 2015).

To ease computational time and deal with the dearth of close cluster pairs in the randomized positions, we decrease the number of bins by half, increasing the average separation of the smallest bin centered at 10 pc to around 20 pc. We still require the minimum of 25 clusters in the smallest bin. On average, only 5% of the iterations have enough close clusters such that 25 points populated the smallest bin at 20 pc, i.e., in order to obtain 15,000 successful runs with enough clusters populating the

right panels of Figure 4.9, we performed over 300,000 trials. This is a significant result as well, even after requiring a minimum of 25 points with separations around 20 pc, still substantially less than what is observed in real clusters of NGC 628 with 86 cluster pairs at these separations, the cluster positions are not driven by randomized processes.

With similar findings with the shuffled age test, we find that the correlation disappears when the positions are randomized, suggesting that the $\Delta t - R$ correlation is not a statistical effect. This is further reinforced by the lack of multiple cluster pairs at short distances in the randomized data compared to the presence of a significant number of close, young clusters pairs observed between real clusters. For 15,000 successful trials of randomizing the positions, the probability of a random trial having a slope in agreement within 1σ of the value obtained for the real $\Delta t - R$ correlation of NGC 628 (0.33 ± 0.07) is 0.02% (3 out of 15,000 trials). The cluster positions, combined with their ages, are indeed imprinted with information on the physical mechanism that drives cluster formation and the $\Delta t - R$ correlation is statistically robust.

4.4.7.3 Unbinned Results

As briefly discussion in Section 4.4.5, the scatter in the intrinsic $\Delta t - R$ across all separation lengths is considerable and the clusters will gather in rows along the y-axis owing to the discrete ages resulting from SED fitting. Figure 4.10 shows the unbinned distribution for all the cluster pairs in NGC 628. We compare it to the one of the 15,000 trials where the clusters are assigned randomized positions (top right panel of Figure 4.9). The cluster ages are the same in both cases, only the positions are changing between the two panels of Figure 4.10. There is a noticeable increase in younger cluster pairs at small separations of less than a few hundred parsecs in the real data that is not present in the randomized trial, which becomes more apparent

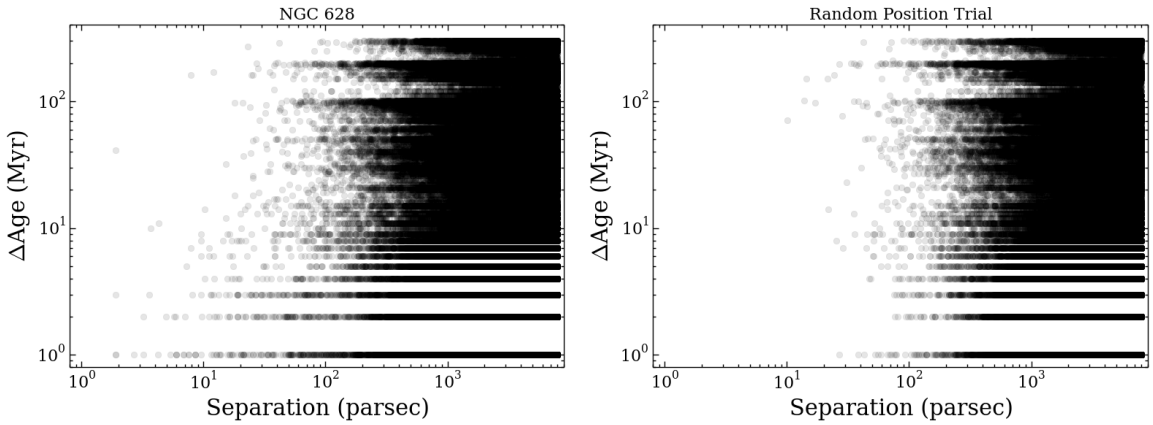

Figure 4.10. Unbinned $\Delta t - R$ relation for NGC 628. The age difference between cluster pairs as a function of separation between the cluster pairs without any binning for NGC 628 and the random cluster position trial in Figure 4.9. There is an excess of clusters at smaller separations for smaller age differences in the actual data that is not recovered in the simulations.

when we bin across spatial separations in Figure 4.9. In the real data, the majority of clusters with separations less than 50 pc are generally less than ∼30 Myr old, while in the random trial, there is no trend in age for clusters pair separations less than 50 pc. Therefore, despite the scatter present, the unbinned data support the inference from the binned data, that the $\Delta t - R$ relation is statistically robust, further supporting our results from the previous section using both randomized ages and randomized positions.

4.5 Discussion

Larson (1981) was the first to establish that turbulence is responsible for the observed velocity dispersion and size correlation of GMCs, where the line widths increase as a power of their radius: $\sigma \propto R^{\beta}$ with $\beta = 0.38$. This relation indicates that turbulence is faster in larger regions. The slope for the size–velocity dispersion relationship is now believed to be slightly steeper: $\beta \propto 0.5$ (e.g., Solomon et al., 1987; Rosolowsky et al., 2008; Rice et al., 2016). Our results show that the duration of star

formation within star-forming regions in eight galaxies is $\Delta t \propto R^{[0.25,0.6]}$, suggesting that the $\Delta t - R$ cluster relation is consistent with resulting from turbulence. For a hierarchical ISM arising from turbulence, the velocity difference Δv between points separated in space by Δx varies as $\Delta v \propto \Delta x^{\beta}$ and the crossing time Δt over that distance varies as $\Delta t \propto \Delta x / \Delta v \propto \Delta x^{1-\beta}$. Thus, our slope α equals $1-\beta$ for slope β in the size-linewidth relation of interstellar turbulence. For a turbulent model, $\beta \sim 0.5$ (Rice et al., 2016), results in $\alpha \sim 0.5$. This value of α is consistent with observations in the LMC (Elmegreen & Efremov, 1996; Efremov & Elmegreen, 1998), the MW (de la Fuente Marcos & de la Fuente Marcos, 2009), and with our sample. Furthermore, correctly accounting for the effects of cluster evolution brings the observed slopes closer to the expected value of 0.5 from turbulence (see Section 4.4.4 and de la Fuente Marcos & de la Fuente Marcos, 2009).

The change from a power-law distribution to a flat distribution in the $\Delta t - R$ cluster relation may describe the transition from a scale-free turbulent motion at small-scales to uniform large-scale galactic dynamics, occurring around 0.5–1 kpc (Elmegreen et al., 1996; Sánchez et al., 2010; Dutta et al., 2013). It is likely related to the maximum size of a coherent star-forming region, which may be the turbulent Jeans length, expected from self-gravity on the largest scale (Elmegreen & Elmegreen, 1983), or the length given by galaxy rotation (Escala & Larson, 2008). The separation sizes are also similar to the maximum correlated size of cluster complexes as probed using the two-point correlation function in the same galaxies (Grasha et al., 2015, 2017a).

While the age and separations of the youngest clusters can be well-described with a turbulent model, a better understanding on the effects of cluster evolution and survival can help improve our understanding of turbulent-driven star formation and unravel whether the observed deviations of the slopes from the theoretical 0.5 value are the results of other evolutionary or environmental effects. The maximum size of

a turbulent cell appears to be limited by shear: we find a tight relation between the maximum turbulent velocity and the average velocity difference due to shear at the size of the largest correlated star-forming region (Figure 4.3).

The combination of both the galaxy size and the shear determines the average age of star-forming regions in the $\Delta t - R$ correlation, varying by a factor of five across the sample. Larger galaxies exhibit larger star-forming complexes, impacting the size of $R_{\rm max}$, and within each complex, the age difference of a sub-region scales with roughly the square root of the size of the sub-region (Figure 4.1). Shear limits the average ages of the regions by limiting the ratio of the maximum size to the average age. The duration of star formation within a region is proportional to a few crossing times and the recovered velocities are similar to the turbulence motions of the ISM (~ 10 km s^{-1}; Heiles & Troland, 2003).

Clusters formed together in groups that exhibit similar ages is an expected outcome of hierarchical star formation model (Bhatia & Hadzidimitriou, 1988; Dieball et al., 2002; De Silva et al., 2015). The relation between time and distance we find implies that young clusters born together in close pairs (or groupings) do indeed show similar ages, providing further support that star formation, as traced by these young stellar clusters, is organized in a hierarchical manner (e.g., Hopkins, 2013a). The $\Delta t - R$ relation is statistically robust and significant, the correlation disappearing when both the ages and positions are randomized within each galaxy (Section 4.4.7). Star formation is not a random process: both the positions and ages are consistent with being driven by turbulence and it is unlikely to have a random process give rise to the large number of close-age cluster pairs observed in local galaxies.

4.6 Summary and Conclusion

In this work, we investigate the relation between age difference and the separation of young (<300 Myr) cluster pairs in eight local galaxies to investigate whether a

correlation exists between the duration of star formation and the size of the star-forming region. Our main results can be summarized as follows:

1. Clusters that are born closer to each other exhibit smaller age differences compared to clusters that are born further apart. This time-distance correlation among young clusters implies that on average younger star-forming structures are less extended than older regions.

2. The average age difference between pairs of clusters increases with their separation as the 0.25 to 0.6 power. The duration of star formation and the size of the star forming regions are consistent with expectations from turbulence: larger regions can form stars over longer timescales as the duration of star formation scales with the square root of the size over longer timescales as an effect of turbulence cascading down to smaller size regions. The slopes we recover tend to be closer to 0.5 if only clusters younger than $\sim$100 Myr are considered, indicating that the hierarchical distribution, inherited by a turbulent ISM, dissipates with time.

3. The power-law relation reaches a maximum size where the age difference between pairs becomes constant. The maximum size increases with galaxy size, and the ratio of the maximum size to the average age difference – values that are similar to the turbulent speed of the ISM – increases with galaxy shear. These results imply that star formation proceeds hierarchically in giant star complexes with a duration on small scales that is proportional to the local turbulent crossing time. The complexes are then dispersed by shear.

4. For the spiral galaxies in our sample, the maximum velocity marginally correlates with the galaxy's size; weaker correlations are found with the galaxy's SFR and morphology. The dwarf galaxy in our sample does not fit the trend of the spirals.

At small scales, turbulence appears to be a primary candidate driver of the spatial and temporal scales for star formation processes although other potential drivers, like self-gravity (Li et al., 2005) will need to be further investigated. Future studies and improved simulations over an expanded range of galactic parameters that also consider cluster feedback into the ISM will help provide the necessary insight to the secondary processes that impact the overall organization of star formation across a wide range of scales.

CHAPTER 5

CLASSIFYING STAR CLUSTERS WITH MACHINE LEARNING TECHNIQUES WITH LEGUS: CORRELATING YOUNG STAR CLUSTERS WITH MOLECULAR CLOUDS IN M 51

This chapter[1] presents the machine learning (ML) identification process to produce the star cluster catalogs and correlates the star cluster catalogs with that of GMCs within the grand design spiral NGC 5194 in order to quantify the association timescale of star formation with molecular gas.

5.1 Abstract

We present a pilot test using machine learning (ML) to generate homogeneous, self–consistent, and repeatable catalogs of star clusters. The ML technique is applied to NGC 5194, a nearby galaxy that is part of the Legacy Extragalactic UV Survey (LEGUS), and compared against clusters that are visually confirmed. Using an existing Bagged Decision Tree algorithm, we achieve a good agreement between the clusters with both identification methods: the visual identification yields about 75% agreement between different human classifiers, while the overall accuracy of the ML classification achieves 67% agreement with the visual classifications. Outside the footprint where visual classification is available, compact and centrally concentrated clusters are readily identified, while asymmentrical systems with multiple peaks (e.g., stellar associations) are generally missed. Post-facto visual inspection of the ML-only

[1]These results are based on research published by Grasha et al. (2018a), and is reproduced by permission of the AAS.

classification still finds ∼70% recovery rate for centrally concentrated clusters. We use these clusters for our investigation of NGC 5194. We compare the distribution of star clusters relative to molecular gas in the inner 9 kpc, where CO measurements are available, finding that younger star clusters are spatially closer to molecular clouds than older star clusters. We also find that a significant fraction of young star clusters are still associated with a molecular cloud (i.e., located within the footprint), with median ages of 4 Myr whereas unassociated star clusters are significantly older at 40 Myr. Lastly, we find that younger star clusters are preferentially located within the stellar spiral arms and are hierarchically distributed within star-forming regions of ∼1 kpc in size that disperse after 50–100 Myr; the fractal dimensions of the clusters are steeper than those of the GMC hierarchy. Given the distance to their nearest GMC, this timescale for disassociation corresponds to traveling velocities for the star cluster of 10 km s^{-1}. We find evidence for a galactocentric-dependence on the coherence of star formation, where clusters located near the inner region of the galaxy reside in significantly smaller star-forming complexes and display a more homogeneous distribution compared to star clusters further from the center, suggesting a correlation between the survival of a cluster complex and its environment.

5.2 Introduction

A central tenet of our understanding of star formation is that it is correlated in such a way that stars form in stellar aggregates (e.g., Lada & Lada, 2003; Portegies Zwart et al., 2010) that we call 'clustered star formation'. In such a distribution, individual components of a population are more likely to appear near other members than if they were randomly distributed. Observations have directly measured the correlation in the spatial distribution between young stars, stellar clusters, and associations (e.g., Gomez et al., 1993; Zhang et al., 2001; Oey et al., 2004; Scheepmaker et al., 2009; Gouliermis et al., 2014, 2015b, 2017; Grasha et al., 2015, 2017a,b; Sun

et al., 2017a,b), providing overwhelming evidence for the coherence of star formation over multiple scales and within an extensive variety of galactic environments.

This hierarchical distribution of star formation is understood as a consequence of dense gas concentrated primarily within giant molecular clouds (GMCs) that undergo fragmentation (Carlberg & Pudritz, 1990; McLaughlin & Pudritz, 1996), under the influence of both gravitational collapse (de Vega et al., 1996) and turbulence (e.g., Elmegreen & Scalo, 2004; Federrath et al., 2009; Girichidis et al., 2012; Hopkins et al., 2013; Guszejnov & Hopkins, 2016) and other feedback processes (e.g., Krumholz, 2014) that act to suppress star formation. Such hierarchical fractal structure in the interstellar medium (ISM) results in a power-law distribution of the ISM components from which we can measure the fractal dimension, or the strength of clustering (e.g., Elmegreen & Falgarone, 1996; Roman-Duval et al., 2010). Within this framework, young star clusters should inherit their clustered distribution from the ISM structure from which they are born; observations corroborate the clustering of molecular gas (e.g., Elmegreen & Falgarone, 1996; Sánchez et al., 2010; Miville-Deschênes et al., 2010) as well as protostellar cores (Johnstone et al., 2000, 2001; Enoch et al., 2006; Stanke et al., 2006; Young et al., 2006) and young stellar objects (Bressert et al., 2010). The hierarchical distribution is slowly lost with time, either due to random motions that remove the fractal imprint, becoming more uniformly distributed with age, or due to the merging of sub-clusters into larger clusters (e.g., Gieles et al., 2008; Bastian et al., 2009; Davidge et al., 2011).

Star clusters, identifiable in galaxies up to distances of ∼100 Mpc, are born at the densest regions of the hierarchy (Elmegreen, 2008) and provide a sensitive and direct observational signature of the star formation process. Despite this, it is not well understood to what extent the galactic environment, both locally and globally, influences the evolution and survival of star clusters or the complexes from which they are born. These outstanding issues need to be properly addressed in order

to accurately characterize the connection between star formation occurring at the scales of individual stellar aggregates, with scaling laws that are uncorrelated with the molecular clouds below ∼100 pc (e.g., Onodera et al., 2010; Schruba et al., 2010), and the global scaling relation between star formation and gas reservoirs of entire galaxies (Kennicutt, 1998; Kennicutt & Evans, 2012).

Within the local universe, star clusters are more extended than point sources (individual stars) in Hubble Space Telescope (HST) imaging, although the ability to discriminate between stars and star clusters decreases with increasing distance, higher stellar background/crowding, and lower resolution at different bands. Compounded with the diversity of star cluster properties (radii of 1–10 pc and masses of $100 - 10^7\ M_\odot$; Portegies Zwart et al., 2010), it is difficult to optimize a set of parameters for a 'one size fits all' automated identification approach. Because of this, the most common procedure used to secure catalogs of star clusters is through visual inspection. The difficulty in identification is further compounded by the absence of any discernible features; this is opposite to galaxies, that often have substructures, like spiral arms, that make crowdsourcing a viable option with, for example, the Galaxy Zoo project (Lintott et al., 2008, 2011). The subjective judgment from visual identification often produces inhomogeneous results and inconsistent cluster classifications between different galaxies and different classifiers as there is no obvious definition of a star cluster, especially when the clusters are unresolved. This has resulted in conflicting results within the literature, where separate teams produce different cluster catalogs using varying criteria for what constitutes a star cluster from the same images (see, e.g., Adamo et al., 2015; Chandar et al., 2015). Furthermore, as astronomical data sets become increasingly larger and substantially more complex to process and analyze, implementing automated classification techniques becomes a promising and necessary approach. Manual classification is time-consuming, requiring a significant

amount of time in order to recreate catalogs as our understanding of the physical properties of star clusters evolves (Portegies Zwart et al., 2010).

In this paper, we present a pilot experiment of off-the-shelf ML algorithms applied to the galaxy NGC 5194, in order to test the effectiveness of these algorithms in identifying young stellar clusters in nearby galaxies. ML approaches are both faster and more efficient than the traditional approach of visual identification. A common 'supervised learning' ML algorithm implements the use of decision trees to perform feature recognition (Breiman, 2001; Kotsiantis, 2007), a parameterization approach where the classification is a two step process: (1) a training or learning step, where a classification model is constructed based on data with known labels (i.e., telling the ML algorithm what is and what is not a cluster), searching for similarities in the source data with the same labels; and (2) a classification step, where the training model is used to predict the class label for unlabeled data. Supervised learning algorithms require a pre-existing catalog of images, where each image contains a single, centered source with a pre-determined label, and thus, still requires human intervention for creating the training set catalogs. The accuracy of the results depends on the size and accuracy of the labels in the training set. Much like human identification, distinguishing sets of objects, such as stars (point-like sources) from star clusters (more extended than a point source or made of multiple, connected point sources) becomes increasingly difficult as the sources become fainter, due to either a loss of spatial resolution and/or a decrease in the signal to noise, or location in crowded environments, making it harder to distinguish the object in question from numerous nearby neighbors or contaminants. Thus, ML algorithms are expected to suffer the same limitations as human classifiers, but have the advantage of higher speed and efficiency. For instance, ML algorithms applied to astronomical data has been used to measure the morphological parameters and features of galaxies (e.g., Storri-Lombardi

et al., 1992; Ball et al., 2004; Banerji et al., 2010; Beaumont et al., 2014; Dieleman et al., 2015; Huertas-Company et al., 2015; Xu & Offner, 2017).

This paper is part of a broader goal of creating high-quality star cluster catalogs at exquisite resolution to provide a more comprehensive understanding of cluster formation, evolution, and dissolution. M 51 (NGC 5194 and NGC 5195) is part of the Legacy ExtraGalactic UV Survey[2] (LEGUS, HST GO–13364; Calzetti et al., 2015), a Cycle 21 HST treasury program of 50 nearby (∼3.5–16 Mpc) galaxies in five UV and optical bands (NUV, U, B, V, and I) with the goal of investigating the connection between local star forming regions – as traced with young stellar clusters – and global star formation within the nearby universe. Results with the LEGUS datasets so far include support for a hierarchical star formation process (Elmegreen et al., 2014; Gouliermis et al., 2017; Grasha et al., 2015, 2017a) with the age and size distribution of the hierarchies driven by turbulence (Gouliermis et al., 2015b; Grasha et al., 2017b), investigation of the effect on the evolution of galaxies from the radiative and mechanical feedback of star clusters (Calzetti et al., 2015b), tests for variations in the initial mass function of star clusters (Krumholz et al., 2015a; Ashworth et al., 2017), and generated extinction maps using the stellar catalogs to investigate variations in the dust–to–gas ratio with metallicity (Kahre et al., 2018). We apply this new, automatically produced catalog to a cross-correlation analysis comparing star clusters to the GMC catalog from the PAWS project; this is the first such large-scale cross-correlation outside the Local Group.

In the pre-LEGUS epoch, only a handful of statistically robust cluster catalogs existed for local galaxies (see Adamo et al., 2017, and references therein). The LEGUS dataset is increasing the number of young stellar clusters in the nearby universe by ten-fold, where the addition of the UV and U filters below the Balmer break are

[2]https://archive.stsci.edu/prepds/legus/

critical for the derivation of accurate physical properties of each star cluster (e.g., Anders et al., 2004). Consistency in the cluster candidate classification between galaxies is crucial in order to provide a common ground on which to investigate and compare results to current theories of star formation. While most of the LEGUS cluster catalogs have been produced with the standard visual inspection technique, this paper explores the possibility of expanding the number and depth of the catalogs via ML.

M 51 is an excellent benchmark to investigate the clustering nature of star clusters as a function of galactocentric distance in addition to the connection of the young stellar clusters with maps of molecular clouds; both of these investigations are presented in this paper. The former has been the main topic of two previous papers (Grasha et al., 2015, 2017a), but not yet performed in such a cluster-rich galaxy to explore trends at sub-galactic scales. The paper is organized as follows: The galaxy selection and reduction process are described in Section 5.3. The cluster selection and identification process are described in Section 5.4. The ML methodology and procedure are outlined in Section 5.5. The results are described in Section 5.6, where Section 5.6.1 briefly introduces the molecular gas data and the two-point correlation function is described in Section 5.6.2. We discuss our results concerning hierarchy of the stellar clusters and the connection to the gas reservoirs in Section 5.7. Finally, we summarize the findings of this study in Section 5.8.

5.3 Sample Selection and Data Reduction

In this paper, we have selected M 51 (NGC 5194 with its smaller interacting companion, NGC 5195, collectively known as the Whirlpool galaxy) from the LEGUS survey as an optimal target to experiment with ML identification as a complement to the visual classification of star clusters.

We select NGC 5194 due to its large size (angular size of ∼11' × 7' and stellar mass $M_\star = 2.4 \times 10^{10}$ $M_\odot$) and a star formation rate SFR(UV) = 3 $M_\odot$ yr^{-1}, providing a large and robust number of young clusters. NGC 5194 has a relatively low inclination (22°; Colombo et al., 2014b), mitigating projection effects, a distance of 7.66 Mpc (Tonry et al., 2001), sufficiently nearby to mitigate confusion between stellar systems and individual stars.

The LEGUS dataset of NGC 5194 provides 5 pointings in the NUV (F275W) and U (F336W) broad-bands, observed with WFC3/UVIS. The new UV/U data consist of 4 pointings from LEGUS, covering the arms and outskirts of the galaxy with a deeper exposure covering the central nucleus of the galaxy (GO–13340; PI: S.Van Dyk). Archival B (F435W), V(F555W), and I (F814W) images with ACS/WFC (GO–10452; PI: S. Beckwith) cover the entire NGC 5194+NGC 5195 system with 6 pointings, re-reduced using the same pipeline with the UV and U WFC3/UVIS images from the LEGUS project. The UVIS data only cover a portion of the galaxy (see Figure 5.1 as well as Figure 1 of Messa et al., 2018), and as a result, our clusters are only located within the footprint of the UVIS data as these blue wavelengths are necessary in order to break the age/dust degeneracy and accurately derive the ages of the young stellar clusters.

Reduced science frames are drizzled to a common scale resolution, corresponding to the native WFC3 pixel size (0.03962 arcsec/pixel). The frames have all been aligned and rotated with North up. Detailed descriptions of the standard data reduction of the LEGUS datasets are available in Calzetti et al. (2015).

5.4 Creating the Visually-Identified Star Cluster Catalogs

A detailed description of the cluster selection, identification, photometry, and SED fitting procedures for the LEGUS galaxies is presented in Adamo et al. (2017). The production of the cluster catalog and completeness tests applied to NGC 5194 are

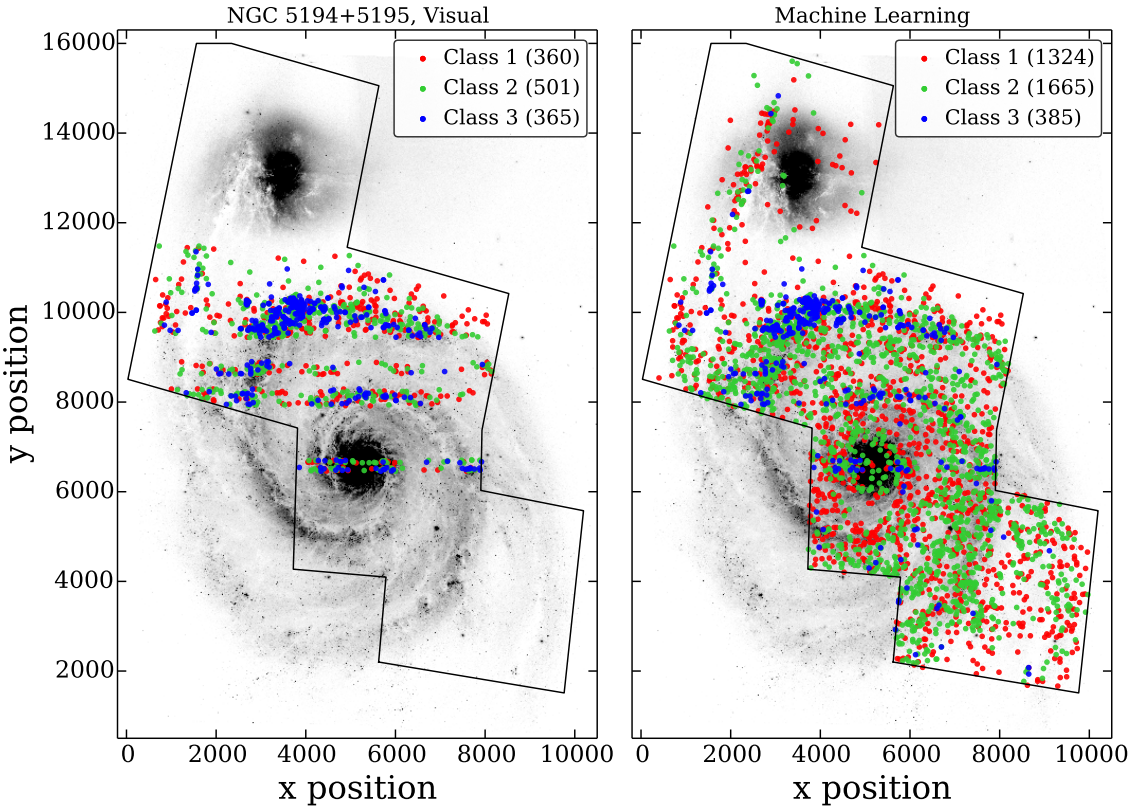

Figure 5.1. NGC 5194 showing the location of the visually identified class 1, 2, 3 clusters (left; Section 5.4.2) and ML-identified clusters (right; Section 5.5) over the V-band ACS image. The black outline shows the UVIS footprint. Cluster classifications are based off of morphology and our morphological cluster classification is a good approximation to also a dynamical classification, with class 1 and 2 being gravitational-bound sources whereas class 3 clusters are stellar associations (Grasha et al., 2015; Adamo et al., 2017; Ryon et al., 2017). Clusters were only visually identified in four select regions (the northern spiral arm, inner arm regions, and the galaxy center) and these were used as the training set for the rest of the galaxy. The ML process is able to reproduce the original class 3 associations from the sub-set of visual classifications but vastly under predicts the total occurrence of class 3 associations in the global sample.

detailed in Messa et al. (2018) and the completeness limit in the five broad bands and across the various environmental sub-regions are further detailed in Messa et al. (2018b). We summarize here briefly the aspects of Adamo et al. and Messa et al. that are important for the current Chapter.

Procuring cluster catalogs is a multi-step process: extracting the cluster candidates (Section 5.4.1) through an automatic catalog and then performing visual inspection of this catalog to remove contaminants (Section 5.4.2). Completing these steps results in our final robust visually identified cluster catalog.

5.4.1 Star Cluster Selection

The *automated catalog* of star cluster candidates is extracted from the V-band image with source extractor (SExtractor; Bertin & Arnouts, 1996). For NGC 5194, the SExtractor input parameters are optimized to extract sources with at least a 10σ detection in a minimum of 10 contiguous pixels. This automatic procedure returns the positions of candidate clusters within the image and the concentration index (CI; magnitude difference of each source within an aperture of 1 pixel compared to 3 pixels) of each source. The CI is related to the size of stellar systems (Ryon et al., 2017) and can be used to differentiate between individual stars and stellar clusters; stars, in general, have narrower light profiles, and therefore, smaller CI values compared to star clusters. The CI value that separates stars from star clusters within each system and image is determined through an iterative inspection of the CI distribution from the output of the SExtractor parameters. The CI reference value used to distinguish between unresolved sources (stars) and resolved sources (candidate clusters) within NGC 5194 is 1.35 mag. At the distance of NGC 5194, the pixel resolution is 1.47 parsec/pixel.

All candidate clusters must satisfy two criteria in the above automated procedure: (1) the CI in the V-band must exceed the stellar CI peak value; and (2) be detected

in at least two contiguous bands (the reference V band and either B or I band). Standard aperture photometry is performed for each cluster candidate using a fixed science aperture of 4 pixels in radius with a local sky annulus at 7 pixels (1 pixel wide) in all five filters. Aperture corrections to account for missing flux are based on isolated clusters (see Messa et al., 2018) and calculated by subtracting the standard photometry in the fixed science aperture from the total photometry inside a 20 pixel radius with a 1 pixel sky annulus. Corrections for foreground Galactic extinction (Schlafly & Finkbeiner, 2011) are applied to the photometry.

Following this step, all cluster candidates detected in at least four bands with photometric error $\sigma \leq 0.3$ undergo spectral energy distribution (SED) fitting procedures to procure the age, mass, and color excess E(B–V) of each source. The SED fitting analysis is performed with deterministic Yggdrasil single stellar population (SSP) models (Zackrisson et al., 2011). The Yggdrasil spectral synthesis code is based on Starburst99 (Leitherer et al., 1999) stellar population spectra with the photoionized nebula predicted by Cloudy (Ferland et al., 1998, 2013). All cluster catalogs for the LEGUS galaxies use a Kroupa (2001) universal IMF (but see Ashworth et al. (2017) for a generalization to a variable IMF). The cluster properties in this paper are derived using Padova isochrones that include TP–AGB stars (Vázquez & Leitherer, 2005) and the starburst attenuation curve (Calzetti et al., 2000). The χ^2 fitting procedure is described in Adamo et al. (2010) and the error analysis is described in Adamo et al. (2012).

5.4.2 Visual Inspection & Star Cluster Classification

After the first step — extraction of the clusters from an automatic catalog and SED fitting — all clusters with an absolute magnitude brighter than −6 mag in the V-band undergo visual inspection by a minimum of three independent classifiers within the LEGUS team to produce the final *visual catalog*. The magnitude limit is

defined by the detection limits of the LEGUS sample, which enables selecting down to a ∼1000 $M_{\odot}$, 6 Myr old cluster with E(B–V) = 0.25 (Calzetti et al., 2015). A total of 10925 cluster candidates brighter than our magnitude cut off pass all the criteria in the extraction process (Section 5.4.1) and undergo the visual classification procedure, although only 2487 cluster candidates actually underwent the visual classification process, selected within four bands across the UVIS footprint (Figure 5.1). The regions were selected to cover the northern spiral arm, inner arm regions, and the very center of the galaxy.

The visual inspection step is necessary and currently the only way to remove non-cluster interlopers within the automatically extracted catalog to ensure the robustness of the final cluster catalog. While creating more conservative cuts in the automatic extraction parameters such as raising the CI value (i.e., selecting only the most extended sources) may exclude stellar contaminants, it is difficult to remove bad pixels, MW stars, or background galaxies as those types of interlopers are not standardized in size or color, and indeed CI cuts alone are inadequate at procuring reliable star cluster catalogs (Grasha et al., 2015). As a result of the vast diversity of star clusters, there is no unique combination of criteria that will lead to the detection of all bona-fide clusters while at the same time not introducing false detections. This is detailed in Appendix C and necessitates the (albeit subjective) visual selection process that goes into the creation of most star cluster catalogs.

The visual classification is performed using a custom interactive tool that utilizes DS9 with V-band and red-green-blue (RGB) composite images. Based on the morphology and color, each sources gets classified under one of four classes (see Adamo et al., 2017):

> **Class 1**: Symmetric and centrally compact star clusters. Usually uniform in color.

Class 2: Compact, asymmetric star clusters with some degree of elongation. Usually uniform in color.

Class 3: Clusters that show multiple-peaked profiles on top of an underlying diffuse emission.

Class 4: Non-clusters, including but not limited to foreground stars, asterisms, background galaxies, saturated pixels, etc.

The initial visual classification procedure of 2487 cluster candidates in the four regions across the UVIS (see Figure 5.1) was performed by members of the LEGUS team, with a yield of 1226 (49%) considered class 1, 2, or 3 clusters. Each source is inspected by at least three classifiers, and we find that the classification is often subject to human interpretation. The resulting agreement between classifiers is around 70%–75%. For large disagreements (more than one class discrepancy between the classifiers), additional classifiers are involved, until a consistent classification for the candidate is achieved. Due to the time required to make the final robust catalog of the initial star clusters, only about 25% (2487 out of 10925) of the cluster candidates are visually inspected.

NGC 5194 has been the subject of numerous studies and multiple catalogs have been produced for the young star cluster populations at varying resolutions and coverage of the galaxy disk (e.g., Bastian et al., 2005; Lee et al., 2005; Gieles et al., 2005; Scheepmaker et al., 2007; Chandar et al., 2016). The predominant disparities in most catalogs results from morphological differences as opposed to the extraction process for clusters in overlapping regions between prior studies. A comparison of our cluster catalogs derived using the standard LEGUS process relative to previous catalogs is detailed in Messa et al. (2018) and will be further addressed in Chandar et al. (in prep).

5.5 Machine Learning

The application of ML algorithms to identify star clusters promises to be an optimal and fast solution to attain homogeneous and internally-consistent catalogs of star clusters without the need for human intervention. Using the LEGUS four-class classification, we run our pilot test using ML to automate the classification procedure and the clusters that have been visually classified as a training set.

5.5.1 Machine Learning Method

Our ML classification employs existing algorithms that implement feature learning through execution of decision trees on raw pixel values from the reduced and aligned mosaics of NGC 5194; all cluster classification with ML is performed using only color and flux as training features. Decision trees fall under the category of 'ensemble learning' algorithms: a supervised ML procedure in which a number of constraints is implemented at each step before leading to a decision and assigning a final classification. For our star clusters, the ML algorithm is first trained on the sub-sample of pre-classified clusters from human-identification before application to the unclassified data without a human classification. We refer the reader to previous work on the detailed methodology of decision trees by, e.g., Breiman (1984) and Freund (1995); Freund & Schapire (1999). The ML procedure as applied to the star clusters in this study is outlined in the following steps.

- **Creating the image patches.** We extract 15×15 pixel image patches (22×22 parsec) centered around the coordinates of all 10,925 cluster candidates in the automated catalog (Section 5.4.1) from the full mosaic of NGC 5194 in three colors (F336W, F55W, and F814W). The classification uses raw pixel values as features, resulting in a collection of 1×675 dimensional feature vectors from each image (a 15×15 vector per color). The ML algorithm does not implement any use of color features, i.e., our features are a 1D vector. A full utilization

of the 2D structure inherent to image data and the role of color in the ML classification is a goal for future development.

To augment the dataset, each image patch is repeated with a rotation of 90, 180, and 270 degrees. The rotation serves to both increase the size of the training samples and to ensure that the machine learns to recognize objects independently of orientation.

- **Defining the training set.** The 2487 sources that have visual classifications (Section 5.4.2) are each represented by a single feature vector; the canonical task of ML is to separate clusters (i.e., a set of feature vectors) from non-clusters (another set of feature vectors), where the cluster feature vectors are further classified into one of three classifications and similarities among objects of the same class are identified and 'trained' from the vector features. Commonalities encompass the color (e.g., class 3 associations are typically bluer in color than class 1 or 2 clusters), shape (e.g., class 1 clusters are nearly always symmetric in shape whereas class 2 clusters have some degree of elongation) and number of peaks, among other quantities. The success of the ML performance depends on both the fidelity and number of training examples. Inconsistencies in the classifications from the visual procedure naturally introduce inaccuracies in the ability for ML to identify commonalities between the features of each class.

  The training set is further augmented through inclusion of the LEGUS visual cluster catalogs from NGC 628, NGC 3344, NGC 6503, and NGC 7793, adding ∼3300 more visually-identified sources (Grasha et al., 2017a). Table 5.1 lists all the galaxies, their distances, and their visually identified clusters that are in our training sample. Each source from the additional cluster catalogs are also augmented with rotations of 90, 180, and 270 degrees. In total, the visually-identified sources of NGC 5194 make up 38% of the all training dataset we

use. The inclusion of the other visual cluster catalogs serve to help even out the distribution of cluster classes as 50% of the sources in NGC 5194 are non-cluster contaminants (Figure 5.3). The ability to accurately transfer the classification model of clusters across galaxies will likely only be effective when each galaxy is at approximately the same distance (3.5–10 Mpc) and the star clusters are visually comparable (i.e., we do not include semi-resolved clusters within diffuse, nearby dwarf systems in the training set of unresolved clusters as found in NGC 5194.

- **Implementing the decision trees.** We implement a Bagged Decision Tree algorithm[3], a structured classifier that makes step-by-step choices based on the value of the feature vectors. Decision trees operate by continually splitting the data into two subsets through binary decisions until all star clusters in the subset are of the same class. For example, starting with the first classification, all input clusters are separated into two possible outcomes, belonging to class 1 or not, repeating for all four classifications. We then carry out an ensemble of 100 bagged trees[4] before assigning the final class to a cluster, i.e. we train separate trees on subsamples of the entire dataset, arriving at the classification for each individual object by the majority vote over the hundred iterations in the ensemble, selecting a random subset of predictors to use at each decision split. The proportion of trees agreeing on the most popular vote corresponds to the confidence of that ensemble classifier. The confidence in the classification increases with the number of trees used; the machine algorithm always converges to an internally accurate final classification after 100 trees. The final four-class

[3]We use the TreeBagger function from the Statistics and Machine Learning Toolbox in Matlab.

[4]The ensemble number 100 was chosen as a convenient end-point for the algorithm for this test, although other tests show that error minimization was achieved already after 50 or 60 bagged trees. See Appendix D and Figure D.1.

classification assigned to each individual object is the classification of the most confident of the four binary classifiers.

- **Determining accuracy.** We exclude a randomly selected 10% of the training dataset out of the training process for evaluation purposes, repeating this step 10 times, to increase statistics. The results from the trained data are then compared to the excluded clusters for computing classification accuracy and confusion matrices from the known visual classifications. This is done for each decision tree that is computed. The accuracy (also known as precision) is the fraction of relevant instances (i.e., true positives) among the retrieved 10% instances. Our final reported accuracy is the error obtained with a 10-fold cross-validation. Cross validation combines the average measures of a fit to derive a more accurate estimate of the prediction performance of a model.

  When using the visually identified sources in NGC 5194 alone as training sets, we achieve a 4-class accuracy of 62%. Increasing the training sample to include the LEGUS spirals listed above increases the accuracy to 67% (Figure 5.2). The fact that the accuracy increases with the additional training samples demonstrates the ML's ability to transfer the trained models to classify objects in different galaxies at varying distances and/or environments.

5.5.2 Comparison of Visual-Identified Clusters to Machine Learning Clusters

Figure 5.3 shows the number of clusters between the four-classifications for the two identification methods: the ML algorithm run on the entire ensemble of cluster candidates and the visual inspection performed on about 25% of the total ensemble. The percentage of ML class 1 or 2 clusters remain roughly the same between the two subsamples, indicating that the ML is able to correctly reclassify and identify new class 1 or 2 clusters. A post-ML visual inspection performed by one of the LEGUS

Table 5.1. Galaxies and Star Clusters in the Training Sample

Galaxy	Distance (Mpc)	N_1	N_2	N_3	N_4	N_{total}
NGC 5194	7.7	360	501	365	1261	2487
NGC 628	9.9	418	434	412	658	1922
NGC 3344	7.0	134	119	138	75	466
NGC 6503	5.3	72	96	130	168	466
NGC 7793	3.4	65	137	148	97	447
Total		1049	1206	1193	2259	5788

Notes. Columns list the galaxy, its distance in Mpc, and number of visual clusters with classifications 1, 2, 3, and 4 used as the training sample in the ML algorithm.

human classifiers on one of the regions without previous visual clusters confirms that the recovery rate for class 1 and 2 clusters is around 70%. In other words, the human classifier agrees with the ML classification in 70% of the cases (see details in Section 5.5.4). The dearth of class 3 associations indicates that the ML method is unable to recognize new class 3 associations. We do not observe an excess of class 1 and 2 clusters in the region outside the regions with visually classified clusters, leading us to conclude that the ML considers almost all class 3 associations as class 4 contaminants. This results in a larger percentage of class 4 non-clusters in the ML catalog compared to the visual catalog under the assumption that the total sample is similar to the human-classified sample in its distribution of classes. This informs us that the ML procedure can identify single peak, compact clusters (i.e., class 1 or 2) whereas it is unable to distinguish between multiple peak systems that are clusters from the chance alignment of unrelated stars. The cause for this is further investigated in Section 5.5.3.

In Appendix E, we compare the mass and age function of the visual cluster catalog to the ML cluster catalog calculated according to the method of Messa et al. (2018).

We find that both the age and mass functions are consistent between the two catalogs within their uncertainties. Thus, using the ML procedure to classify star clusters compared to the traditional visual approach does not introduce biases in any part of the cluster parameter space.

5.5.2.1 Contamination of Stars in the Cluster Catalog

Figure 5.4 shows the distribution in color-color space that separates single stars from star clusters, using the method as demonstrated by Whitmore et al. (2010). In our catalogs, the color-color space occupied by single stars and clusters is well-separated and the contamination rate of individual stars in the ML cluster catalog is low, $\sim$2%. This demonstrates that our visual classification procedure, as well as the ML procedure, is successful in differentiating between star clusters and contaminants. However, color-color cuts alone do not suffice to differentiate star clusters from stars (Whitmore et al., 2010). Size also is not sufficient in differentiating between stars and star clusters (Whitmore et al., 2010; Grasha et al., 2015). The inability to use color and size cuts to cull stars from star cluster catalogs necessitates the visual identification step in creating robust and reliable star cluster catalogs. We demonstrate this in Appendix C.

5.5.2.2 Comparing Properties of the Visual versus ML Clusters

Figure 5.5 shows the distribution of ages and masses for each classification. There are differences in the properties of the different cluster classes: class 1 clusters are the oldest and most massive out of the three labels in the classification, where the clusters decrease in both age and mass for class 2 clusters and 3 associations, respectively. The difference in the observed age and mass properties of the LEGUS classification has been demonstrated in other LEGUS galaxies (see Adamo et al., 2017; Grasha et al., 2017a), showing the validity in the ML classifications compared to the visually identified clusters. For all the analysis throughout the paper, we typically group class 1 and

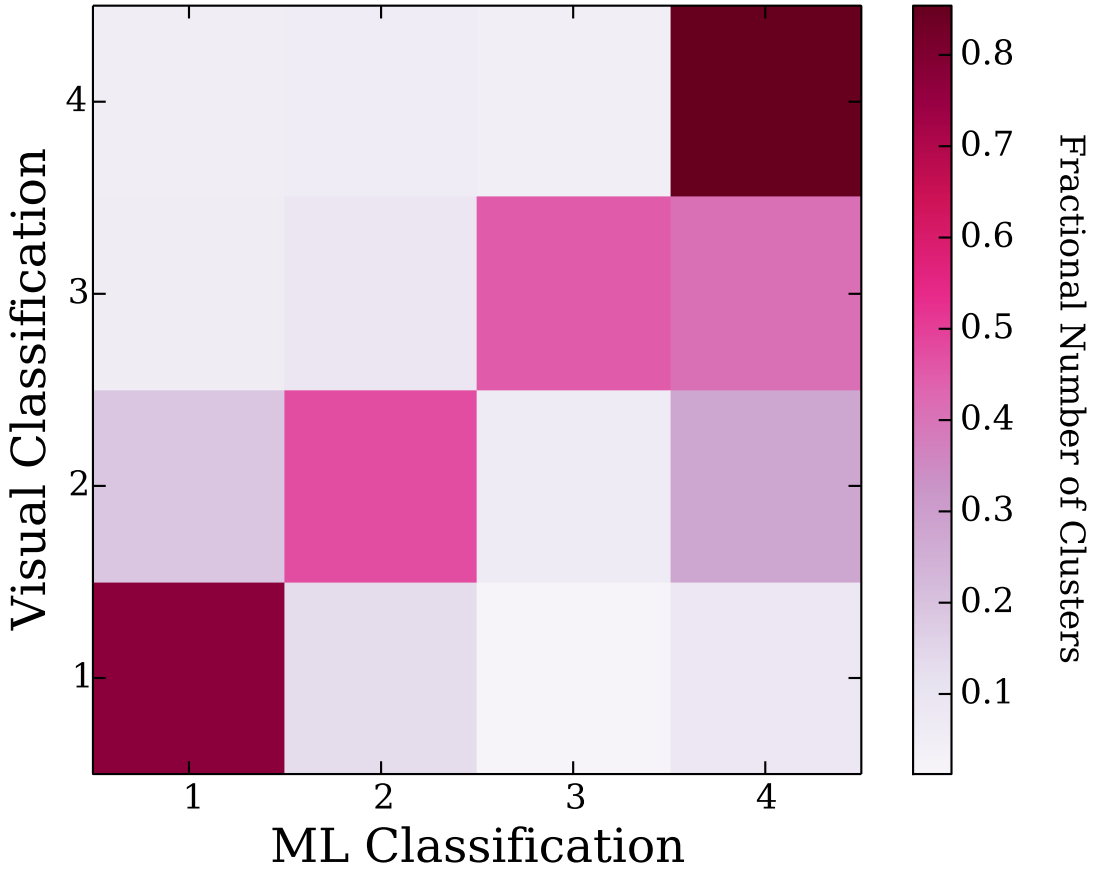

Figure 5.2. Confusion matrix for the cluster class agreement between the visual and the ML classes, built from all LEGUS spiral galaxies for which we have visual classifications. Each color patch shows the fractional overlap between the visual and ML catalog; a perfect agreement would have a 1 to 1 correspondence along the diagonal. The cross validation results give an overall accuracy for the ML technique to identify and reclassify the four classes of 67%. The confusion of class 3 associations and class 4 contaminants is easily seen.

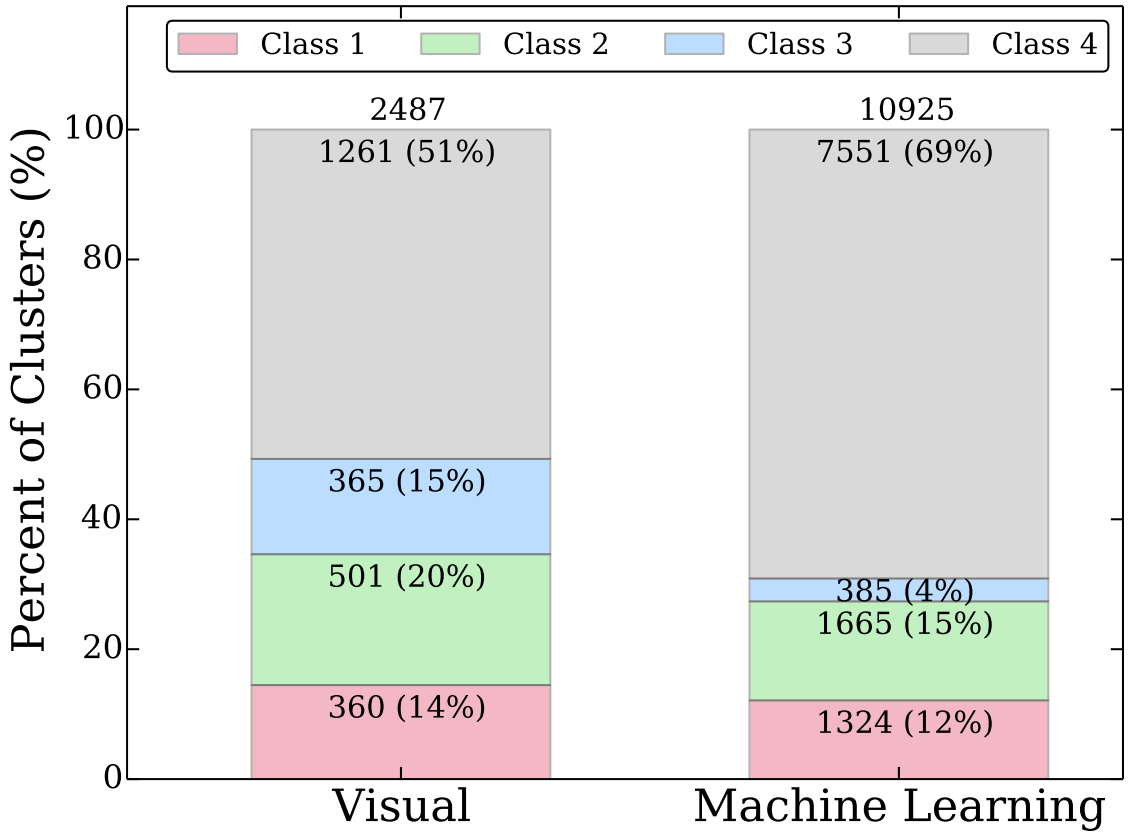

Figure 5.3. Histograms of star clusters identified by humans and by our ML method. We list the number of clusters of each classification from the visual identification procedure (left column) and all the ML classified clusters (right column) along with the relative percentage of each class. The relative number of class 1 (red) and 2 (green) sources increases (roughly) proportionally between the two identification methods but there is a severe dearth of class 3 (blue) clusters identified outside of human-classified clusters, which instead are categorized into the class 4 non-clusters (gray). In total, the number of star clusters in the ML catalog is 2.7 times larger than the visual identified sample.

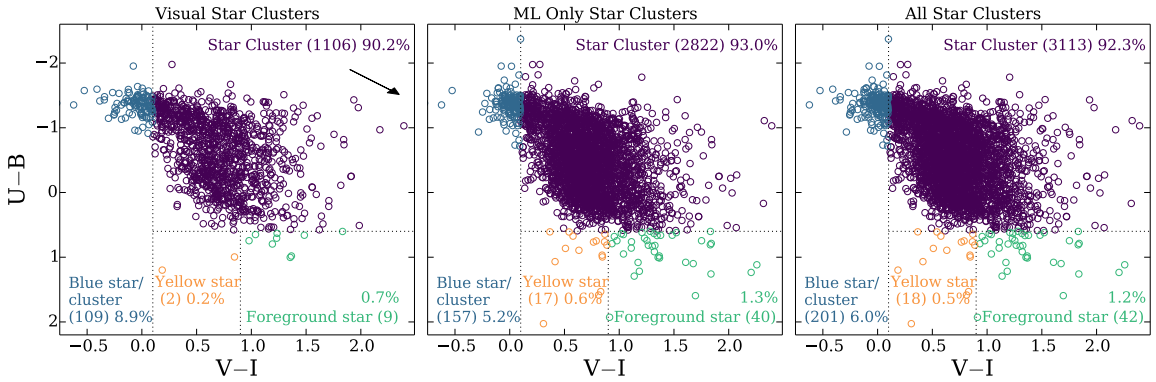

Figure 5.4. U–B versus V–I colors for the total ML cluster catalog of NGC 5194. Shown is the visual catalog (left), ML-only clusters (middle), and the total cluster catalog of NGC 5194 (right). The clusters are color coded by the different regions that deliminate in color space stars versus star clusters from Whitmore et al. (2010). The numbers in parenthesis list the total number of sources within each region. We also list the fraction of each subregion compared to the total star cluster sample in each panel. The ML procedure does an exceptional job at identifying star clusters from individual stars and the contamination is only ∼2 % for the star clusters, comparable to the contamination in the visually identified star cluster catalog. The arrow marks reddening corresponding to $A_V = 1$ mag (E(B–V) = 0.25 assuming a starburst attenuation curve).

2 together into a single group and compare them against the properties of the class 3 associations. This division is physically motivated, as class 1 and 2 clusters appear to be genuinely bound star clusters and are distinct in their properties, compared to class 3 systems, which we consider associations (see Adamo et al., 2017; Ryon et al., 2017). The division is also motivated as class 3 associations are underrepresented compared to the total number of clusters recovered by the ML procedure.

As shown in Figure 5.5: the properties of the ML clusters that do not have a visual classification (purple line) are in general slightly older and more massive than the clusters that are visually identified (pink line), but consistent with each other at the $\sim 2\sigma$ level. This suggests that the star clusters identified by humans are younger, and due to the luminosity-limited nature of the catalogs, younger clusters have a statistically higher chance of being less massive. This is also potentially impacted by the fact that the vast majority of the visual clusters are identified in the northern arm. Work by Messa et al. (2018b) in NGC 5194 demonstrates that the environment increases not only the physical properties of the clusters but also their survivability. This issue can be avoidable in future work by ensuring that the training sample spans the full extent of the galaxy to mitigate any environmental biases.

5.5.3 Limitations and Failure Examples

Visual classification utilizes multiple parameters to classify a star cluster: RGB colors, surface contours, radial profiles, and surface plots (for visual representations, see Adamo et al., 2017). Radial profiles and surface plots aid in separating individual stars from clusters; in general, even the most compact clusters are more extended than individual stars. RGB images help to separate background galaxies, which tend to be redder in color compared to those displayed by clusters. Color also helps discriminate class 3 associations from asterisms composed of individual stars, often

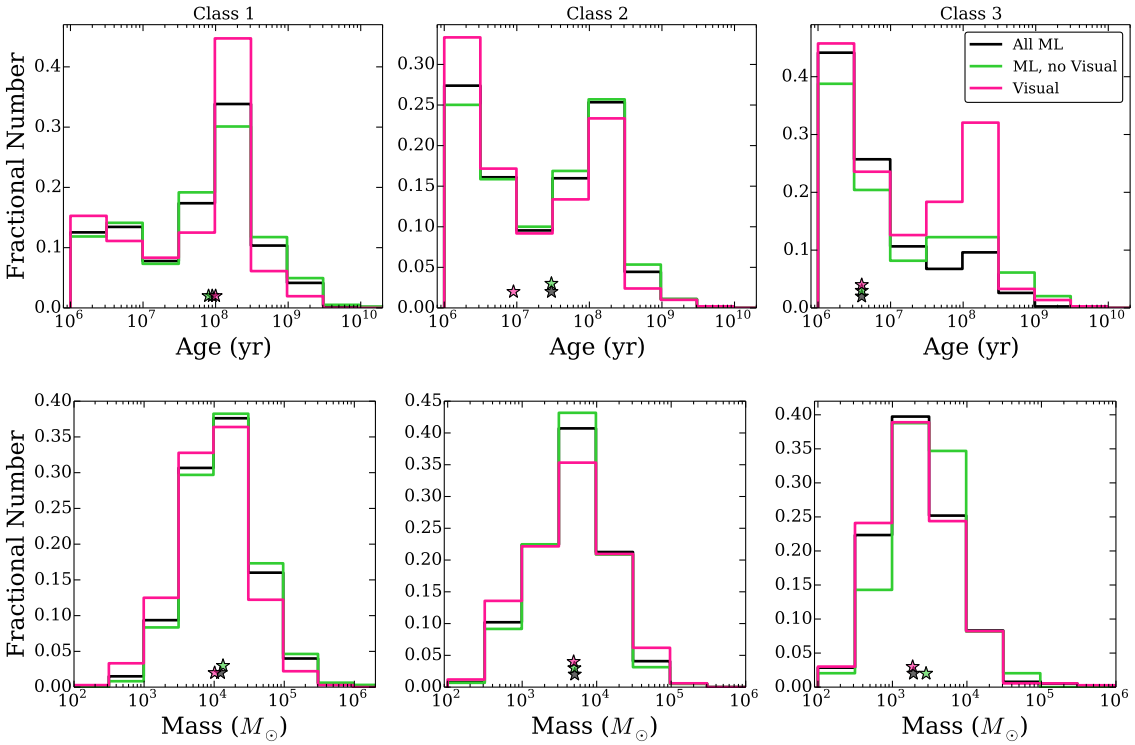

Figure 5.5. Distribution of cluster ages (top row) and masses (bottom row) for each class type (1, 2, 3) for the visual clusters (pink), total ML catalog (black), and ML clusters without a visual classification (purple). The star indicates the median of each distribution. There is a decreasing age and mass distribution for the class 1, 2, and 3 clusters, respectively. Additionally, there is a general trend for the clusters with a visual classification (pinks) to be both younger and less massive than the general population of clusters as found with the ML process that do not have a visual classification (purple), however, the distributions of visual and ML cluster properties are consistent with each other. Symbols that overlap are given an vertical offset for clarity.

distinct in color. Surface contours help in discriminating circular class 1 clusters from asymmetric class 2 clusters.

In our current experiment, the ML algorithm only implements color features (i.e., flux values) and does not directly utilize radial profiles or surface contours/plots. The lack of additional parameters beyond color to train the ML algorithm partly plays a role in the dearth of recovered class 3 sources in the automated catalog. As a whole, class 1 and 2 clusters are fairly uniform and have minimal variation compared to the general category of class 3 sources, which encompass a diversity of compactness and number of peaks. The ML procedure appears to suffer from confusion between the non-uniformity of class 3 and that of class 4 non-clusters, further compounded with noisy labels whereas the commonality in the class 1 and 2 clusters are more easily defined.

Figure 5.6 displays RGB patches for a small subset of clusters that are assigned the same class from both the visual and ML process. Figure 5.7 displays the same except for a subset of failures where sources were visually identified as a cluster (class 1, 2, or 3) but labeled as a non-cluster (class 4) by the ML. In a large number of cases, nearby single stars, which are easily identifiable by eye as not part of the cluster in the center of each image patch, act to confuse the ML algorithm, classifying it as a multiple peak (or non-cluster) when in reality the stars are simply a chance alignment, which is currently unrecognized by the ML. Uneven background is a changeable parameter in the interactive visual identification tool, but constant in all cluster images that are evaluated by the ML algorithm. This is another problematic issue as reliable identification becomes especially arduous in regions with high background due to loss of signal to noise but which can be fixed by improvements in the ML algorithm.

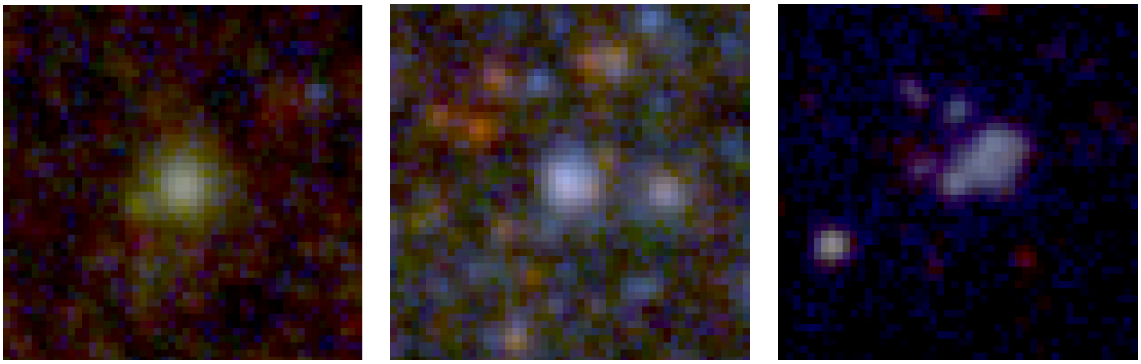

Figure 5.6. RGB images (50 by 50 pixels; 74 parsec in length) for clusters in NGC 5194 with an agreement between the visual classification and the ML classification for class 1 (left), 2 (middle), and 3 (right) clusters. The contrast levels are the same for all the images, which remains fixed for the ML classification, as opposed to the visual classification, where the user has the option to change the contrast levels.

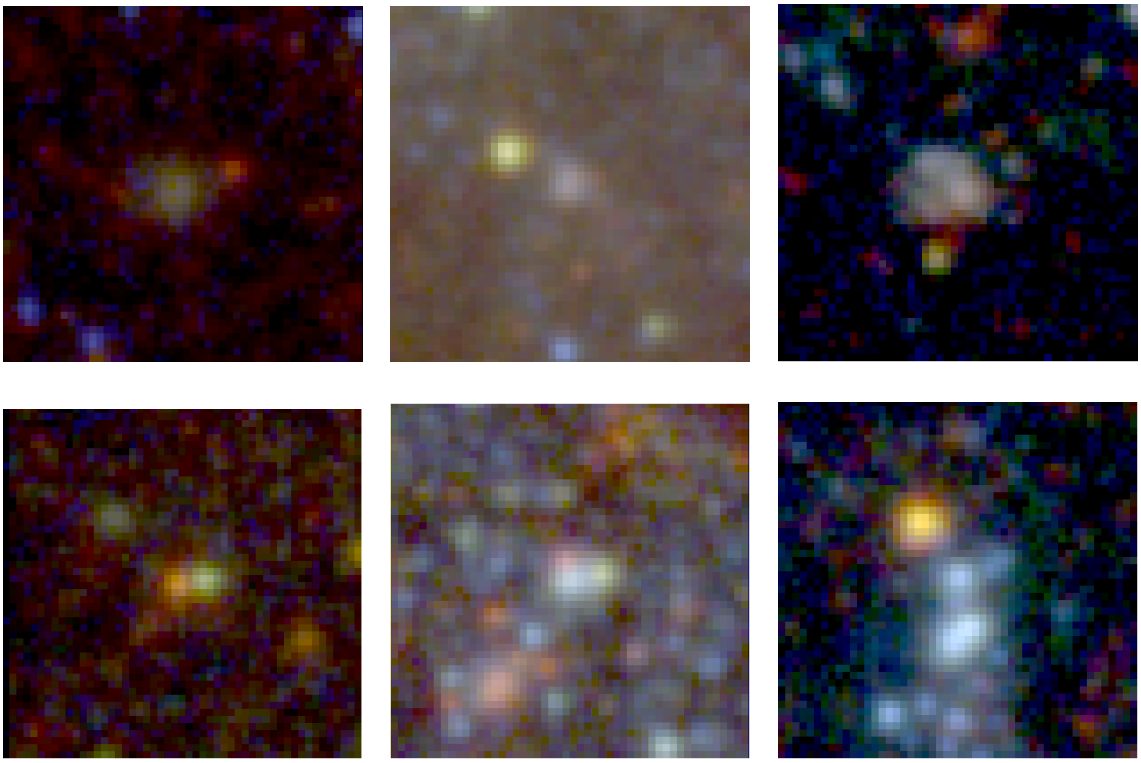

Figure 5.7. Same as Figure 5.6, but for clusters with a disagreement between the visual classification and the ML classification. Top row: Visual class 1 (left), 2 (middle), and 3 (right) clusters that were classified as non-clusters (class 4) with the ML. Bottom row: Class 4 non-clusters from the visual identification that were classified as class 1 (left), 2 (middle), or 3 (right) by the ML.

5.5.4 Spot Checks of the ML Catalog

An independent analysis is performed on 337 star clusters (14% of the training sample) within the training regions to compare the ML procedure classifications to the visual classifications. The aim of this spot check is to better understand if the ML is producing star clusters that align with the visual classification procedure. The results of this manual check recover fractions of clusters in the visual training set that reflect the agreement between experts (∼70–75%; Section 5.5.2). The agreement rate for the checks performed on the ML classifications are 3–5% lower than what is recovered for the manual checks of the visual classifications. This level of agreement is what we expect given the accuracy of our ML procedure (Figure 5.2) and demonstrate that the ML procedure is recovering clusters that are consistent with the LEGUS visual procedure based off of the morphology of the star clusters.

5.6 Results and Analysis

5.6.1 Correlating the Young Star Clusters to Molecular Gas

To connect the young star clusters to their environment, we compare the cluster locations to the location of molecular gas. We use the GMC catalog (Colombo et al., 2014a) from the Plateau de Bure Interferometer Arcsecond Whirlpool Survey (PAWS; Schinnerer et al., 2013; Pety et al., 2013). PAWS mapped the molecular gas in the central 9 kpc of NGC 5194 in its ^{12}CO(1–0) line emission at a cloud-scale resolution of ∼40 pc. The GMCs identified by Colombo et al. (2014) account for about half (54%) of the flux in the PAWS map. The remainder lies in structures that are blended, continuous, diffuse, low S/N or otherwise intractable to their cloud-finding approach. By construction, the GMCs represent significant peaks in the CO emission, and assuming that these correspond to the cluster-forming structures seems reasonable. Figure 5.8 shows the inner region with the star clusters and GMCs that lie within the footprint

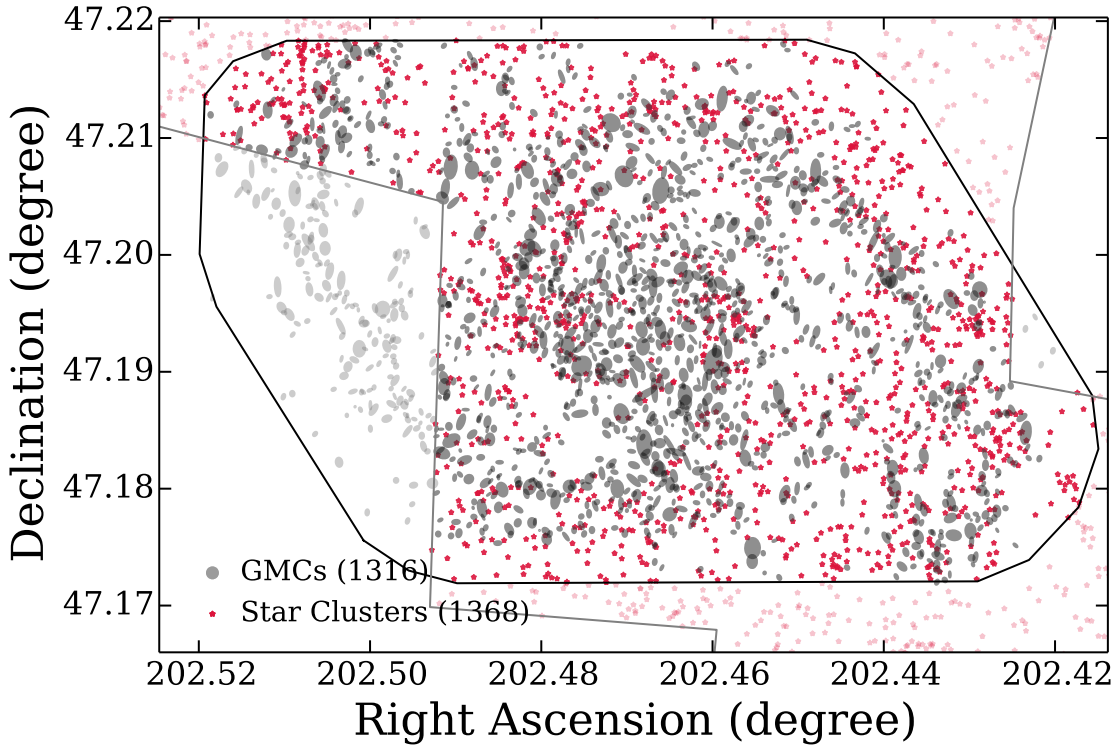

Figure 5.8. Location and sizes of the GMCs and star clusters in the central region of NGC 5194. The black line shows the outline of the PAWS coverage for the GMCs and the gray line shows the UVIS coverage for the star clusters. The numbers listed are the star clusters and GMCs within each footprint. Star clusters and GMCs that are not located within the PAWS and UVIS footprint are shown in light pink and light gray, respectively, and excluded from all star cluster–GMC comparisons.

coverage of both UVIS and PAWS, for a total of 1368 class 1, 2, and 3 star clusters (1268 class 1 or 2) and 1316 GMCs.

Previous observations show that stars and star clusters quickly become unassociated with the GMCs from where they are born, either due to drift or from blowing out the gas. The timescale for most clusters to become exposed and lose association with their molecular gas is as early as 2 Myr within M 83 (Hollyhead et al., 2015) to ∼5 Myr from studies of the Antennae galaxy (Whitmore et al., 2014). We perform multiple tests to correlate the location of star clusters with the molecular clouds to investigate if star clusters become less associated with molecular clouds with age.

5.6.1.1 Distance of Star Clusters from Molecular Clouds

In our first test, we take the star clusters within the PAWS coverage and calculate the shortest distance to the center of their nearest GMC (Figure 5.9). For the entire sample, the median of the shortest distance for the closest star cluster–GMC pair is 87 ± 2 parsec. For comparison, the median diameter of GMCs in this galaxy is 72 pc. The 1σ errors are bootstrap estimates based on 10,000 samples. We divide the star cluster sample into half, corresponding to a galactocentric distance of 2.7 kpc, to investigate if distance from the galaxy center or age of the clusters is the primary driver in the change of distance between pairs of star clusters and GMCs. Star clusters at galactocentric distance less than 2.7 kpc are substantially closer to their nearest GMC at 66 ± 2 pc whereas the shortest distance between all cluster–GMC pair doubles to 132 ± 6 pc at galactocentric distances greater than 2.7 kpc. We further separate the two radial bins into two age bins, older and younger than 10 Myr. We find that for galactocentric distances less than 2.7 kpc, the younger clusters are separated from their nearest GMCs by 59 ± 2 pc, and the older clusters are separated from their GMCs by 74 ± 4 pc. Similarly, at galactocentric distances larger than 2.7 kpc, the younger and older clusters are separated from their nearest GMCs by 118 ± 9 pc and 143 ± 7 pc, respectively.

When only considering age, the star clusters younger and older than 10 Myr exhibit median distances of 71 ± 3 parsec and 107 ± 4 parsec, respectively, which are different separation at a 7–σ level significance. The median radius of the GMC population in NGC 5194 is 36 parsec, thus star clusters that are younger than 10 Myr lie at distances twice the size of the cloud's radius whereas clusters older than 10 Myr are found at median distances three times the size of the molecular cloud.

While on average younger star clusters are closer in proximity to their nearest GMC, the galactocentric distance is a stronger predictor of the trend for increasing distance between clusters and GMCs. Both the aggregate midplane pressure

(Elmegreen, 1989) and the thermal gas pressure (Wolfire et al., 2003) are expected to fall with radius in disk galaxies. Both the surface density and velocity dispersion of the gas are observed to decrease with radius in NGC 5194 (e.g., Schuster et al., 2007; Tamburro et al., 2009), as does the mean volume density of the gas (Chen et al., 2015; Bigiel et al., 2016). In this case, one explanation for our observation would be that in the lower density and environment at larger galactocentric radii, there simply are less clusters and molecular clouds in total, leading to a lower chance of young clusters that are associated with their parent cloud. Star clusters and their unbound complexes do respond to the environment that they live in; we find that clusters at distances in the outer region of the galaxy reside in complexes significantly larger than clusters closer to the galaxy center, an effect of the increased shear, midplane pressure, and turbulence present. We investigate this further in Section 5.6.2.3. These results do not inform us if a star cluster is actually associated with the closest GMC; we investigate this more in the next section.

5.6.1.2 Star Clusters Associated with GMCs

In our next test, we examine the properties of the young star clusters that are still associated with a molecular cloud. We define the association of a star cluster and a GMC as cases where the position of the star cluster falls within the footprint of the GMC. Likewise, a cluster becomes unassociated (i.e., "detached") from its parental GMC when it has drifted away from its natal environment. In addition to detaching via drift, clusters are capable of photo-evaporating and blowing away via stellar feedback/winds the ambient molecular gas; these cases reveal themselves as holes in the GMC but are indeed still associated with the ambient molecular gas (see, e.g., Hony et al., 2015). We treat the GMCs as ellipses and even if there are clusters that have evacuated their surrounding material, blowing out holes in the adjacent gas, these cases would still be considered associated, since they fall within the footprint

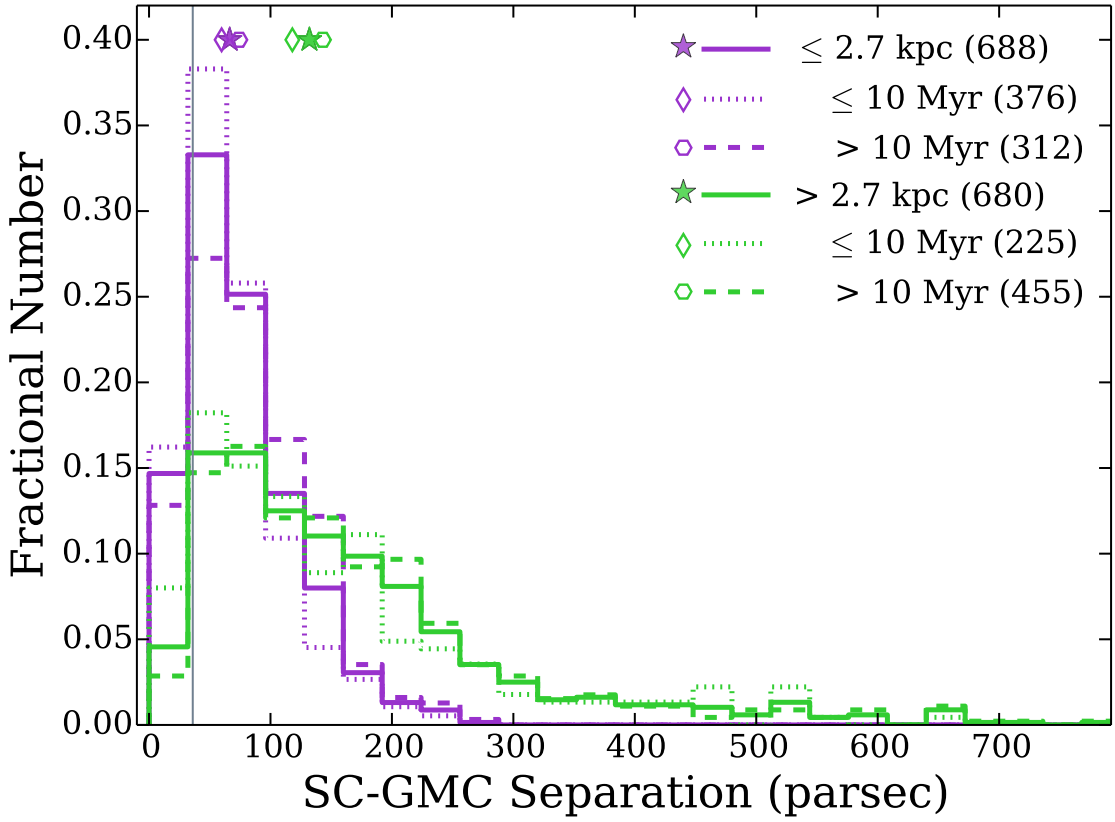

Figure 5.9. Fractional distribution of the shortest distance for each star cluster to the closest GMC broken into two radial bins and age bins. The symbols show the median value for each distribution. Star clusters less than a galactocentric radius of 2.7 kpc (purple) are 66 ± 2 pc from the nearest GMC, with ages less than 10 Myr (dotted line) showing slightly shorter distances of 59 ± 2 pc compared to ages greater than 10 Myr (dashed line) at 74 ± 4 pc. Star clusters at galactocentric distances greater than 2.7 kpc (green) are 132 ± 6 pc from the nearest GMC, with distance dropping for the youngest ($\leq$10 Myr; dotted line) to 118 ± 9 pc and increasing to 143 ± 7 pc for the older clusters (>10 Myr; dashed line). On average, younger star clusters are closer in proximity to a GMC than older star clusters although the distance from the center of the galaxy has a bigger impact, with the average SC–GMC distance increasing with increasing galactocentric distance. The 1σ errors are bootstrap estimates based on 10,000 samples. The vertical line shows the median GMC radius of 36 pc.

of the GMC ellipse. Figure 5.10 shows an example of star clusters that are located within the 1 $R_{\rm GMC}$ ellipse footprint of a GMC in a zoom in of the galaxy. We repeat the analysis for star clusters within an annulus between 1 and 2 $R_{\rm GMC}$, between 2 and 3 $R_{\rm GMC}$, and those beyond 3 $R_{\rm GMC}$ distance.

We expect that if star clusters do disperse from their GMCs with time, clusters that lie further than 3 $R_{\rm GMC}$ will be older than those within 1 $R_{\rm GMC}$. When counting the star clusters that lie within a GMC, we only allow each cluster to be assigned to a single GMC. In situations where multiple clouds lie on top of each other and a star cluster can potentially be paired up with more than one cloud, the star cluster is assigned to the most massive GMC. If we instead assign a star cluster to the closest GMC regardless of its mass, the results remain robust as only 3% percent of the star clusters lie on top of more than one cloud, and in many situations, the closest cloud is still often the most massive as well (see further discussion in Section 5.6.1.3). We perform the search in order of increasing distance from any GMC such that all star clusters that reside within 1 $R_{\rm GMC}$ are assigned to their associated cloud before moving onto star clusters that reside in the annuli between 1 and 2 $R_{\rm GMC}$ and 2 and 3 $R_{\rm GMC}$, respectively.

We find a total of 143 star clusters that lie within the footprint of 118 GMCs ($\leq 1\ R_{\rm GMC}$). An additional 356 clusters are between 1 and 2 radii of a GMC center and 227 clusters are between 2 and 3 radii of a GMC, leaving 642 star clusters unassociated with any cloud ($\geq 3R_{\rm GMC}$). Figure 5.11 shows the distribution of the cluster age and masses. The median age of all clusters in NGC 5194 is 30^{+6}_{-6} Myr, significantly older than clusters located within one radius of a GMC with median ages of 4^{+1}_{-2} Myr. Of the clusters still within a GMC radius, nearly 40% have ages less than 3 Myr. The median age rises slightly for clusters between one and two radii of a GMC with a median age of 6^{+3}_{-1} Myr, and clusters that are between 2 and 3 radii from a GMC have median ages of 20^{+10}_{-7} Myr. Star clusters that are unassociated with any GMC

have median ages of 40^{+20}_{-6} Myr. The 1σ uncertainties are bootstrap estimates based on 10,000 samples; the age trend is significant at the 6–σ level.

More massive star clusters are only marginally more likely to be found within a GMC: the median cluster mass is 10450 ± 1090, 9480 ± 840, 7180 ± 800, and 7680 ± 860 $M_\odot$ for clusters within 1 $R_{\rm GMC}$, 2 $R_{\rm GMC}$, 3 $R_{\rm GMC}$, and unassociated, respectively, although the difference is not significant given the scatter. Table 5.2 lists the properties of the star clusters and, in addition, is divided by the region – center, spiral arm, and interarm – of the galaxy as defined by the PAWS dataset (Colombo et al., 2014a). We are limited by incompleteness within the very center of the galaxy due to the high background and we do not detect sources older than $\sim$10 Myr due to the blending of older sources with similar color to the background light. The cluster catalogs of Chandar et al. (2017) also show a dearth of clusters in the inner $\sim$1 kpc of the galactic center. This appears to not affect the clusters in the central GMC region that are within 1 $R_{\rm GMC}$ where the completeness age is typically older than the age of clusters that are still associated with a GMC ($\sim$4 Myr), though it begins to impact the results for clusters with distances greater than 2 $R_{\rm GMC}$ as those systems are typically older than the completeness age within the center.

We find average E(B–V) values of 0.21, 0.24, 0.24, and 0.25 for clusters within 1 $R_{\rm GMC}$, 2 $R_{\rm GMC}$, 3 $R_{\rm GMC}$, and unassociated, respectively. For a starburst attenuation curve, these correspond to A_V's of 0.83, 0.89, 0.96, and 1.0, respectively. The scatter in E(B–V) for a given age range is significant and it does not correlate strongly with age (Messa et al., 2018). Clusters closer to a GMC do not appear to be any more embedded than those further from a GMC, as indicated by their E(B–V) values. The difference in age between star clusters inside/outside GMCs therefore cannot be explained by significantly higher extinction affecting the star clusters within the spiral arms or GMCs. If anything, we would expect the youngest clusters to be embedded within the clouds, and therefore not to be recovered by our inspections that rely on

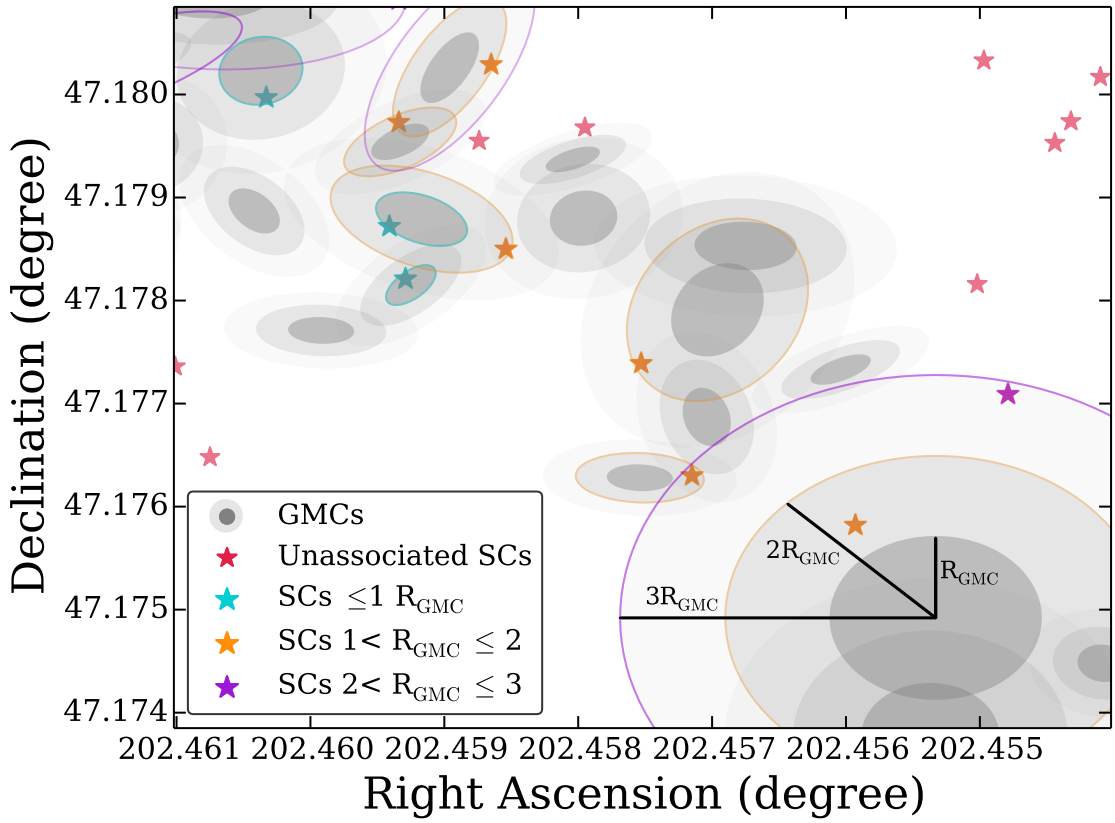

Figure 5.10. Diagram showing how we define associated star clusters. Associated star clusters located inside 1 $R_{\rm GMC}$ (turquoise), 1–2 $R_{\rm GMC}$ (orange), and 2–3 $R_{\rm GMC}$ (purple) while unassociated star clusters are shown in pink. We consider star clusters to be associated with a GMC if they lie within the footprint of the molecular cloud. GMCs that have a corresponding cluster are outlined in the same color as the radial location of the cluster.

detection of UV/blue emission from the clusters. This consideration further reinforces the significance of our result, that the youngest clusters are recovered in proximity of GMCs.

5.6.1.3 Are Massive GMCs more likely to Host Multiple Star Clusters?

We also examine if more massive GMCs are statistically more likely to host multiple star clusters. Figure 5.12 shows the distribution of the mass of GMCs as a function of number of star clusters located within its footprint, for both star clusters

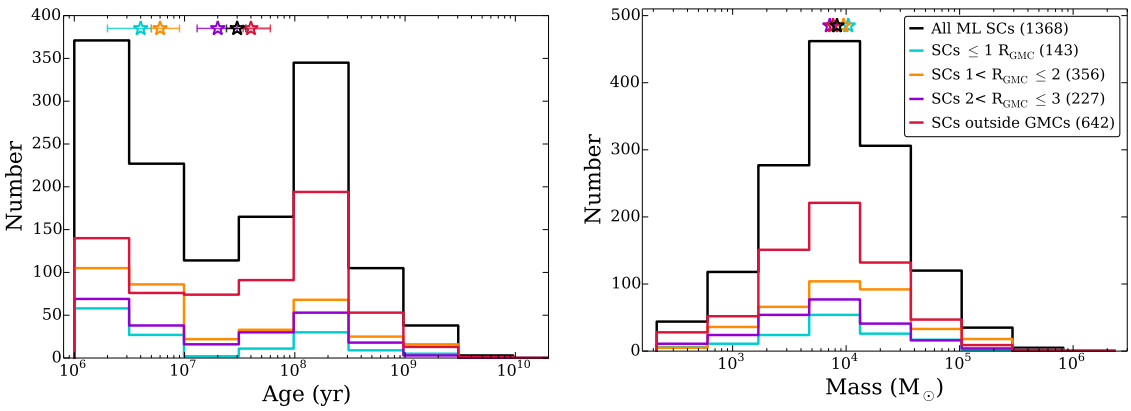

Figure 5.11. Histogram of star cluster ages as a function of distance from their nearest GMC. Left: Distribution of the ages of star clusters for the entire sample (black), star clusters located $\leq$1 $R_{\rm GMC}$ (turquoise), star clusters within 1–2 $R_{\rm GMC}$ (orange), 2–3 $R_{\rm GMC}$ (purple), and star clusters unassociated with a GMC (>3 $R_{\rm GMC}$; red). The stars show the median age of each distribution (30^{+6}_{-6} Myr, 4^{+1}_{-2} Myr, 6^{+3}_{-1} Myr, 20^{+10}_{-7} Myr, and 40^{+20}_{-6} Myr, respectively) and the 1σ errors on the median are bootstrap estimates based on 10,000 samples. Star clusters located within a GMC are generally much younger than star clusters that are further away from a GMC and the age progressively increases with increasing distance from the GMC. Right: Distribution for the masses of the star clusters. The stars show the median mass of each distribution (8300 ± 750 $M_\odot$, 10450 ± 1090 $M_\odot$, 9480 ± 840 $M_\odot$, 7180 ± 800 $M_\odot$, and 7680 ± 860 $M_\odot$). The mass distribution of star clusters located within a GMC (turquoise) is statistically the same as the mass distribution of star clusters that are unassociated with a GMC (red). The 1σ errors from bootstrap estimates based on 10,000 samples are not shown in the plot as the uncertainties are $\sim$10% of the median value.

Table 5.2. Associated Cluster Properties of NGC 5194

Region	N_{GMC}	N_{SC}	Median Age (Myr)	Median Mass ($M_\odot$)
Within $1R_{GMC}$	118	143	4^{+1}_{-2}	10450(1090)
Center	35	40	4^{+1}_{-1}	10360(1510)
Spiral Arm	53	63	4^{+2}_{-2}	12640(2480)
Inner Arm	30	40	3^{+2}_{-1}	7050(2280)
Within $2R_{GMC}$	250	356	6^{+3}_{-1}	9480(840)
Center	62	96	$4^{+0.5}_{-1}$	12250(2020)
Spiral Arm	98	140	15^{+8}_{-5}	10700(1560)
Inner Arm	90	120	7^{+2}_{-4}	7570(1140)
Within $3R_{GMC}$	175	227	20^{+10}_{-7}	7180(800)
Center	53	43	$3^{+0.5}_{-0.5}$	6690(1310)
Spiral Arm	29	88	35^{+12}_{-9}	7290(1610)
Inner Arm	93	96	30^{+10}_{-6}	7120(960)
Unassociated	924	642	40^{+20}_{-6}	7680(860)
Center	239	26	3^{+3}_{-2}	8470(2760)
Spiral Arm	383	124	50^{+13}_{-15}	11430(1440)
Inner Arm	302	492	40^{+6}_{-8}	6940(430)
Total	1316	1368	30^{+6}_{-6}	8300(750)
Center	335	205	$3^{+1}_{-0.5}$	9750(1140)
Spiral Arm	540	415	40^{+8}_{-7}	10480(870)
Inner Arm	441	748	$40\ ^{+5}_{-5}$	6990(720)

Notes. Properties of star clusters depending on their association with a GMC or location within different regions of the galaxy as defined in the PAWS dataset (Colombo et al., 2014a). Columns list the: (1) location of the star cluster; (2) number of GMCs; (3) number of star clusters; (4) median age of the star clusters; and (5) median mass of the star clusters. The 1σ uncertainties are bootstrap estimates based on 10,000 samples. Numbers in parentheses indicate uncertainties in the final digit(s) of listed quantities, when available.

found within 1 $R_{\rm GMC}$ and those found within 2 $R_{\rm GMC}$. There is a marginal statistically insignificant trend for more massive GMCs to host multiple star clusters. As the number of associated star clusters increases, the total spread in the mass of the host GMC does decrease.

Star clusters that lie on top of multiple GMCs can potentially impact this result as we always assign a cluster to the most massive GMC in such an event. However, there are only four star clusters (3%) within one radius of a GMC that have the possibility of being paired up to more than one cloud. For all four cases, these star clusters were the only clusters associated with either GMC. Changing the associated GMC for these clusters away from the most massive GMC only moves four data points downward and both the median and the quartiles remain unchanged (Figure 5.12a). For the star clusters within 2 radii of the center of a cloud, 52 (15%) lie ontop of two GMCs and three clusters (0.8%) lie on top of three clouds. We randomly assign these clusters to different clouds and recompute the values in Figure 5.12b and we find that the effect is minimal and that the median value always falls within the first and third quartiles. Additionally, we find no evidence for more massive GMCs to be hosts of more massive star clusters.

5.6.2 The Two-Point Correlation Function

We implement the angular two-point correlation function $\omega(\theta)$ to measure the magnitude of clustering as a function of projected distance between the star clusters. A detailed description of the formalism and methodology of the two-point correlation function as applied to star clusters within other LEGUS galaxies can be found in both Grasha et al. (2015) and Grasha et al. (2017a). The correlation function provides a way to identify common age structures to derive the randomization timescale for when the coherence of star-forming hierarchies becomes uncorrelated (Section 5.6.2.1) as well as the correlation length — the size scale for the star-forming regions — as a

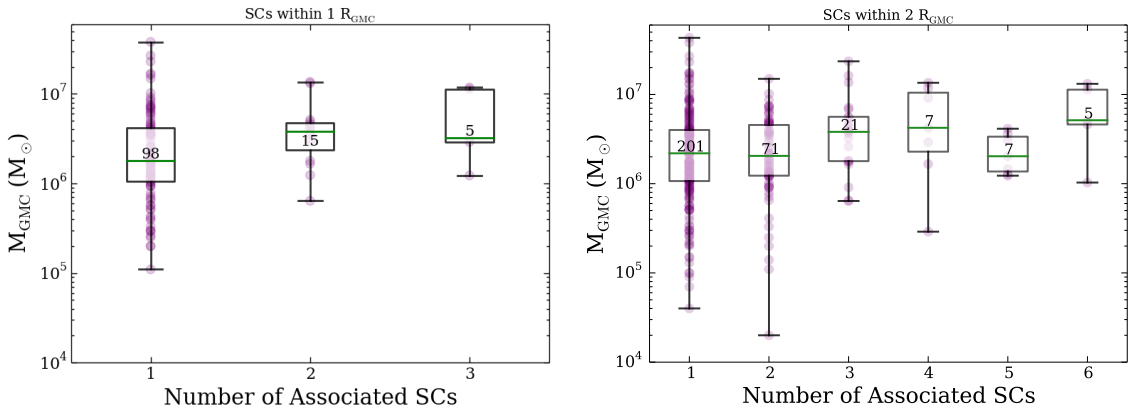

Figure 5.12. Boxplots of the GMC masses for the number of associated star clusters. The purple points show the underlying distribution for individual GMCs, with the total number listed in the middle of each box, and the green line shows the median of the GMC masses. The box encloses 50% of the distribution (the first and third quartile) and the whiskers extend to the minimum and maximum. The left figure shows the distribution for the 118 GMCs with 143 star clusters within one radius of the cloud and the right figure shows the distribution of all star clusters within two radii of each cloud.

function of age and location within a galaxy (Section 5.6.2.3). Here we list the details necessary for the application to the star clusters and the GMCs within NGC 5194.

The correlation function (Peebles, 1980) is defined as a measure of the probability of finding a neighboring object, above what is expected for an unclustered random Poisson distribution. In this study, we implement the angular two-dimensional correlation function $\omega(\theta)$, as projected onto the plane of the sky, as the probability above Poisson of finding two star clusters within an angular separation θ as $\mathrm{d}P = N^2[1 + \omega(\theta)]\ \mathrm{d}\Omega_1\mathrm{d}\Omega_2$, where N is the surface density of clusters per steradian with two infinitesimal elements of solid angle $\mathrm{d}\Omega_1$ and $\mathrm{d}\Omega_2$, separated by angle θ.

The original work of Peebles (1980), along with the application to stellar populations by Gomez et al. (1993) and Zhang et al. (2001), fits the power law distribution to $\omega(\theta)$. Following the convention of Scheepmaker et al. (2009) and Gouliermis et al. (2015b), we fit the power law distribution to $1 + \omega(\theta)$, similar as well to the work of

Odekon (2008). For values of $\omega(\theta) >> 1$, the power law fit will be the same regardless for the two methods. We calculate $\omega(\theta)$ the same way as introduced in Chapter 2.

The physical motivation for applying the power law to $1 + \omega(\theta)$ is that in a fully hierarchical (fractal) model, the distribution will have a smooth power law decline of $1 + \omega(\theta) \propto \omega^{\alpha}$ (Calzetti et al., 1989) with increasing separation between pairs until they reach the correlation length of the hierarchy, where the clustering becomes consistent with being random ($1 + \omega(\theta) = 1$). Inside such a distribution, the number of star clusters inside an annulus increases with the radius r as $N = r^{\alpha} \times r^2 \propto r^{\alpha+2}$, relating the slope α measured from $1 + \omega(\theta)$ directly to the two-dimension fractal number of the hierarchy as $\mathrm{D2} = \alpha + 2$, the fractal dimension for a distribution of objects that lie in the plane of a galaxy. We might expect the slope of $1 + \omega(\theta)$ to change on scales smaller than the thickness of the disk but we do not have enough clusters to see the effect. Interstellar gas has a hierarchical morphology structure with a typical fractal dimension of $D2 \sim 1.5$ (Elmegreen et al., 2006).

Figure 5.13 shows the two-point correlation function for the visually identified clusters, the subset of the ML clusters that is located within the visual regions, and the entire ML cluster catalog excluding the 132 clusters in NGC 5195. The random catalog for the visual clusters follows the same regions where the visual star clusters reside whereas the entire ML catalog follows the UVIS footprint and excludes the region of NGC 5195. We find a smooth and steady decline with increasing radius, well described with a power law. This behavior is expected for a hierarchical distribution and has been observed in the distribution of star clusters and stars within other galaxies (Gomez et al., 1993; Zhang et al., 2001; Odekon, 2006, 2008; Scheepmaker et al., 2009; Gouliermis et al., 2014, 2015b).

Similar to previous studies we have performed on the clustering of the star clusters in LEGUS galaxies (Grasha et al., 2015, 2017a), there is a significant difference in the distribution of each class of clusters. The class 1 clusters of NGC 5194 reveal a nearly-

flat distribution, indicating that these clusters have a more homogeneous and uniform distribution and are less likely to reside within larger unbound star-forming structures compared to class 2 and 3 clusters. Class 3 associations, for a given spatial scale, have a much stronger clustering strength compared to the class 1 or 2 clusters. In Grasha et al. (2017a), we detailed how the morphological classification quite possibly reveals implications on the dynamical state of the clusters; class 1 and 2 have ages and properties that suggest they are gravitationally bound clusters whereas the properties of class 3 associations indicate that they are short-lived systems, possibly dispersing soon after formation, if not already unbound at the time of their formation. The observed differences in the spatial correlations between possible clusters and unbound associations is further supported by the difference in the luminosity functions of the two types of stellar systems, indicating that the classifications are not an evolutionary sequence but delineate distinct physical systems (Adamo et al., 2017).

When we compare the visual and the ML catalog across the entire galactic system, we see that for class 1 and 2 clusters, the amplitude and slope are slightly decreased. This may simply be a result of the ML clusters having slightly older ages compared to the visual clusters (Figure 5.5); there is a strong age-dependency to the clustering, with the strength decreasing with age, investigated further in Section 5.6.2.1. The correlation function for the class 3 associations for the entire ML catalog is noticeably different compared to the distribution of the class 3 associations in the visual catalog. The reason for the extreme amplitude is a result of the ML procedure being unable to recognize and reclassify new class 3 associations outside of the original visual sample; only 42 new class 3 associations (12% more) were identified even though the visual regions encompass only 28% of the entire area of the UVIS coverage for NGC 5194. As a result, the strength of the signal observed is not physically meaningful but rather the result of artificial clustering resulting from correlating the clusters located in a small region of the galaxy with a random catalog that encompasses a much larger

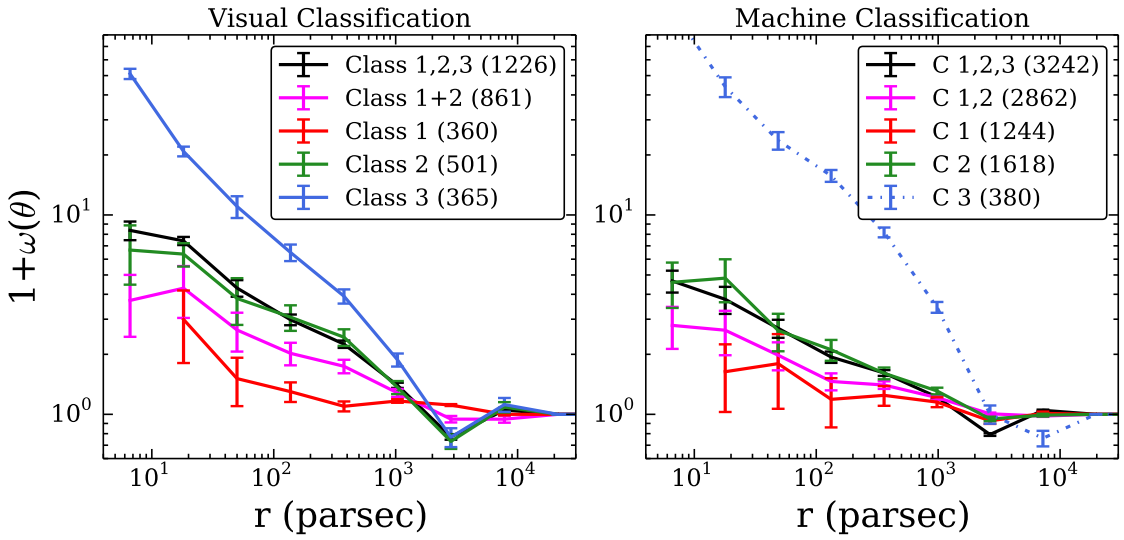

Figure 5.13. Two-point correlation function $1 + \omega(\theta)$ for the visually identified clusters (left) and the entire ML-classified clusters (right) as a function of physical scale for NGC 5194. The numbers in parentheses show the number of clusters in each classification. The dearth of class 3 associations in the ML catalog outside the visual regions (Figure 5.1) give rise to the artificial clustering occurring within those sources, represented with a dashed line. The 132 clusters within NGC 5195 are excluded from the clustering analysis.

region. The lack of new class 3 associations, which tend to be younger, may help explain the shallowing of the slope for the entire ensemble of clusters compared to the visual sample. Table 5.3 lists the power law fits to the distributions. We only fit the power law over the dynamical range where the correlation function is above 1, where we define the correlation length. We additionally only list the power law functions for combination of the class 1, 2, and 3 correlations as that distribution is statistically the same as the class 1 and 2 distribution due to the small number of class 3 associations.

5.6.2.1 Age Effects

The star clusters in NGC 5194 are preferentially associated with the spiral arms of the galaxy (Figure 5.1). As shown in Messa et al. (2018), the concentration within the

spiral arms is particularly noticeable for the very young clusters (<10 Myr), though the clustering persists for clusters up to ages of ∼100 Myr before diminishing. Thus, the hierarchical distribution slowly dissipates with time as the result of star clusters randomizing throughout the galaxy, creating a decline in $1+\omega(\theta)$ with increasing age in addition to length scale.

Figure 5.14 shows the correlation functions for different age ranges. We see a significant decrease in the slope and the amplitude of the clustering with an increase in the age of the clusters. The youngest clusters (<10 Myr) exhibit a slope of -0.50 ± 0.03, significantly different from the global measurement of -0.32 ± 0.02. 50–100 Myr after their formation, the recovered slope is nearly flat at -0.11; the star clusters that are able to travel from their star-forming complexes take on distributions that are consistent with being randomized. Thus, clusters become less correlated with each other and the coherence of star formation as traced with star clusters subsides after 50–100 Myr; this randomization timescale is consistent with what we find in other local galaxies (Grasha et al., 2017a). While the randomization timescale is as short as 50 Myr, structures with lower fractal dimensions still survive for longer periods, as e.g., is found for stars in M 31, where structure survives for at least 300 Myr (Gouliermis et al., 2015a), in accordance to what we found in NGC 6503 (Gouliermis et al., 2015b).

The lack of any signal in the class 1 clusters across the entire galaxy is easily understood in comparing the distribution and ages of the clusters in Figure 5.17. The center, inter arm, and arms are clearly visible in the distribution of the class 2 clusters whereas there is no structure present in the distribution of the class 1 clusters. The class 2 clusters are additionally significantly younger than that of class 1 clusters. The trend of younger — in this case, class 2 — clusters still residing within their star-forming hierarchies predominantly within the spiral arms is expected as clusters form preferentially within the spiral arms.

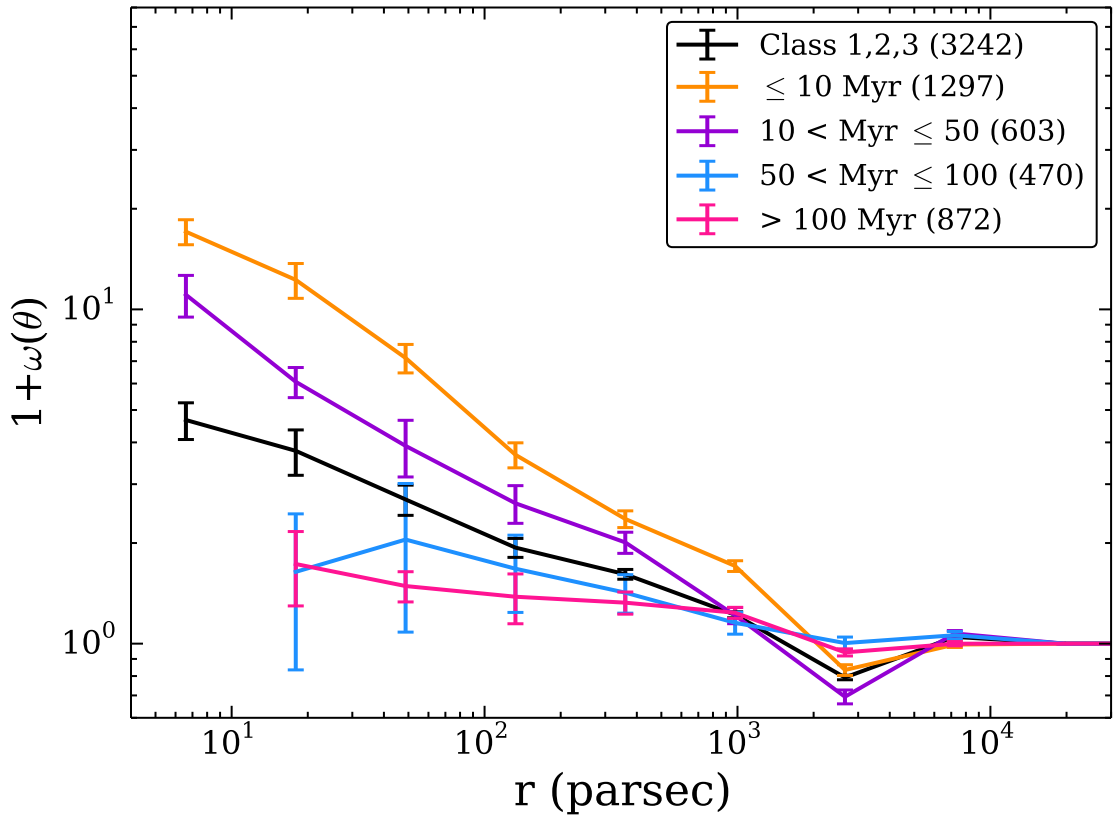

Figure 5.14. Age dependency of the two-point correlation function $1 + \omega(\theta)$ for the clusters as a function of the spatial scale. The numbers in parentheses show the number of clusters for each age range. There is a strong age dependency to the clustering, becoming consistent with a randomized distribution after 100 Myr. Additionally, there is a lack of clusters older than 50 Myr at small separation lengths compared to what is seen for younger clusters, further supporting that older clusters are less correlated with each other.

Table 5.3. Power Law Parameters of NGC 5194

Class	N_{SC}	A_ω	α (")	R_ω (pc)
		Visual		
Class 1,2,3	1226	25.2(1.5)	−0.43(0.02)	1940(150)
		ML[1]		
Class 1,2,3	3242	10.6(0.8)	−0.32(0.02)	1440(150)
≤10 Myr	1297	48(4)	−0.50(0.03)	2200(220)
10 < Myr ≤ 50	603	25(3)	−0.44(0.04)	1330(190)
50 < Myr ≤ 100	470	2.7(0.3)	−0.11(0.03)	8460(2640)
>100 Myr	872	1.93(0.16)	−0.06(0.02)	26900(11800)
≤ 4 kpc	1396	5.8(0.8)	−0.32(0.08)	250(70)
> 4 kpc	1846	14.9(0.9)	−0.34(0.02)	3070(270)
		Star Clusters in PAWS field		
Class 1,2,3	1368	5.6(0.5)	−0.27(0.04)	551(96)
≤10 Myr	601	36(7)	−0.55(0.10)	633(170)
		GMCs		
GMCs	1507	2.3(0.2)	−0.09(0.03)	5040(1400)

Notes. Columns list the (1) Classification of stellar clusters – Class 1: symmetrical clusters; Class 2: asymmetrical clusters; and Class 3: multiple peak clusters; (2) Number of star clusters in each classification; (3) Amplitude A of the angular correlation function; (4) Slope α of the angular correlation function; and (5) Correlation length R_ω when the function becomes consistent with a random distribution $(1 + \omega(\theta) = 1)$. Numbers in parentheses indicate uncertainties in the final digit(s) of the listed quantities, when available.

[1] Excludes the star clusters in NGC 5195

5.6.2.2 Comparing the Clustering of Star Clusters to Molecular Clouds

Under the assumption that the young star clusters inherit their initial distribution from their natal clouds, we expect the initial stellar configurations to reflect the highly structured nature of the ISM gas. We again use the PAWS GMC catalog (Colombo et al., 2014a) to calculate the clustering present in the distribution of the molecular clouds within the inner 9 kpc. Figure 5.15 shows the resulting correlation functions, for all the molecular clouds and the subset of clouds only located within the UVIS field, though the distribution between the two subsets of GMCs is nearly identical. This allows us to compare the distribution of the latter to the star clusters that are also located at the intersection of the PAWS and UVIS fields.

The largest difference between the distribution of the star clusters and the GMCs in Figure 5.15 compared to the global sample in Figure 5.13 is the correlation length where the function becomes consistent with a random distribution. The star clusters within the inner 9 kpc have correlation lengths ($1+\omega(\theta) = 1$) of a few hundred parsec, significantly smaller than what is observed for the entire sample (Figure 5.15). Beyond this, the correlation signal is consistent with being random. The GMCs show a fairly large correlation length of $\sim$5000 pc, but part of this is due to the shear shallowness present in their distribution.

The measured slope for the GMCs (Table 5.3) from the correlation function is -0.10 ± 0.03, significantly shallower than the subset of star clusters located within the PAWS region at -0.27 ± 0.04, where we only measure the slope for the GMCs at distances greater than the median radius of 36 pc. The difference in the measured slopes between star clusters and GMCs additionally increases when we only consider the youngest ages, where the hierarchy becomes substantially more clustered for clusters with ages less than 10 Myr old (Figure 5.15). The clustering of the GMCs is exceptionally flat compared to that of the star clusters, although it is consistent with that of star clusters older than 100 Myr. There is an excess in the clustering of the

GMCs around a few hundred pc, arising due to substructure present in the galaxy. While we expect the hierarchy of the youngest star clusters to mirror that of the GMCs, we find that this is not the case and that the slopes greatly deviate for the youngest clusters.

The same trend for an observed shallowing of the GMC distribution compared to the young star clusters is also observed within the flocculent galaxy NGC 7793 (Grasha et al., 2018b). The results from these two galaxies illustrate that the excess in the distribution of the star clusters must arise from an inherently more clustered distribution compared to that of GMCs, indicating GMCs may produce more than one star cluster, where the production of star clusters is sequential as not simultaneous.

To investigate this further, we compare the distribution of the youngest and most massive star clusters to the most massive GMCs (Figure 5.16). In attempt to match the mass limit between that of star clusters and GMCs, for a given cut in the star cluster mass, we assume a star formation efficiency (SFE) of 1% and use that to estimate the resulting stellar mass for a given GMC mass limit. SFE is expected to be of order a few percent in nearby galaxies (Leroy et al., 2008; Usero et al., 2015). There is a general increase in the slope of the GMC distribution with increasing mass, signifying that when we include all the molecular clouds in the correlation function, there is a washing out of the clustered signal as observed for stellar products. The most massive molecular clouds ($> 5 \times 10^6$ $M_\odot$) start to show a significantly more clustered distribution, albeit with a slightly different distribution than that of the star clusters, however, we can only do the comparison for GMCs with a masscut at $> 3 \times 10^6$ $M_\odot$ as we have inadequate young, massive star clusters above that value.

The increased clustering for increasing GMCs mass implies that our optically identified star clusters arise from a specific subset of molecular clouds, partly explaining the difference in the spatial distribution in Figure 5.15. We conclude that it is the most massive clouds that are more likely to produce star clusters, given the compar-

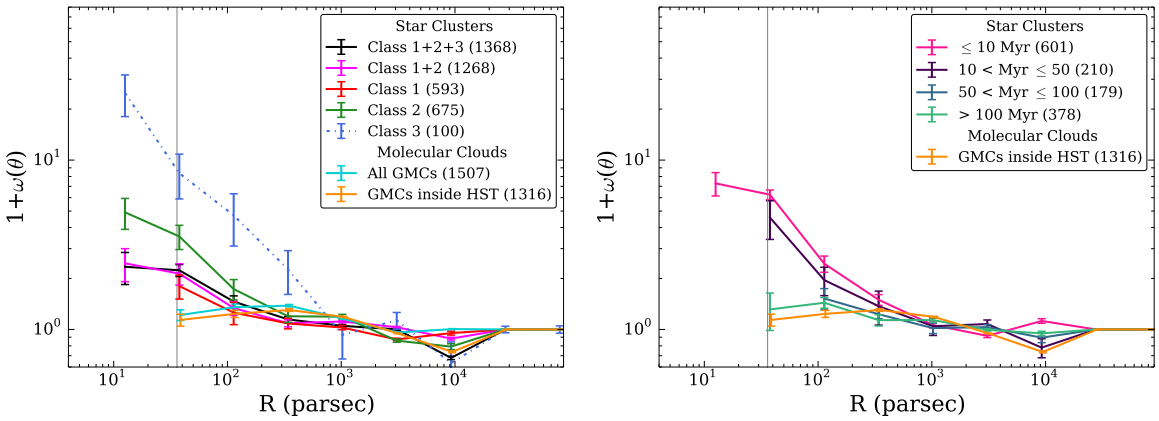

Figure 5.15. Two-point correlation function $1+\omega(\theta)$ for the GMCs and the clusters that are located within both the UVIS and PAWS footprints. The numbers in parentheses show the number of clusters in each classification and the vertical dashed lines show the median radius of the GMCs at 36 pc. The GMCs have minimal structure, reaching a flat distribution accompanied by a slight excess at ∼800 pc. We perform the fits such that the slope for the GMC is measured for all scales greater than ∼40 pc. The star clusters, compared to the global sample in Figure 3.6, show distributions that exhibit slopes suppressed and show much shorter correlation lengths. The youngest (<10 Myr) star clusters do not mirror the observed slope of the GMCs and the clustering is quickly lost around 50 Myr. The dearth of class 3 clusters give rise to the artificial clustering and the excess of class 2 over class 1 is an age effect (see Figure 5.17).

ison in their spatial distributions, although an improvement can be made in further studies with an increase in resolution of the CO gas to examine the relations at smaller spatial scales below our current resolution.

5.6.2.3 Radial Trends

To investigate a potential radial dependence of the star cluster hierarchies, we divide the star clusters of NGC 5194 into two galactocentric bins, separated by the corotation radius of ∼4 kpc (Querejeta et al., 2016), and computing the correlation function for the clusters in these two radial bins. Figure 5.18 shows that the clustering does indeed depend on the position of the clusters within the stellar disk. For a given spatial scale, the clustering amplitude is significantly stronger for clusters

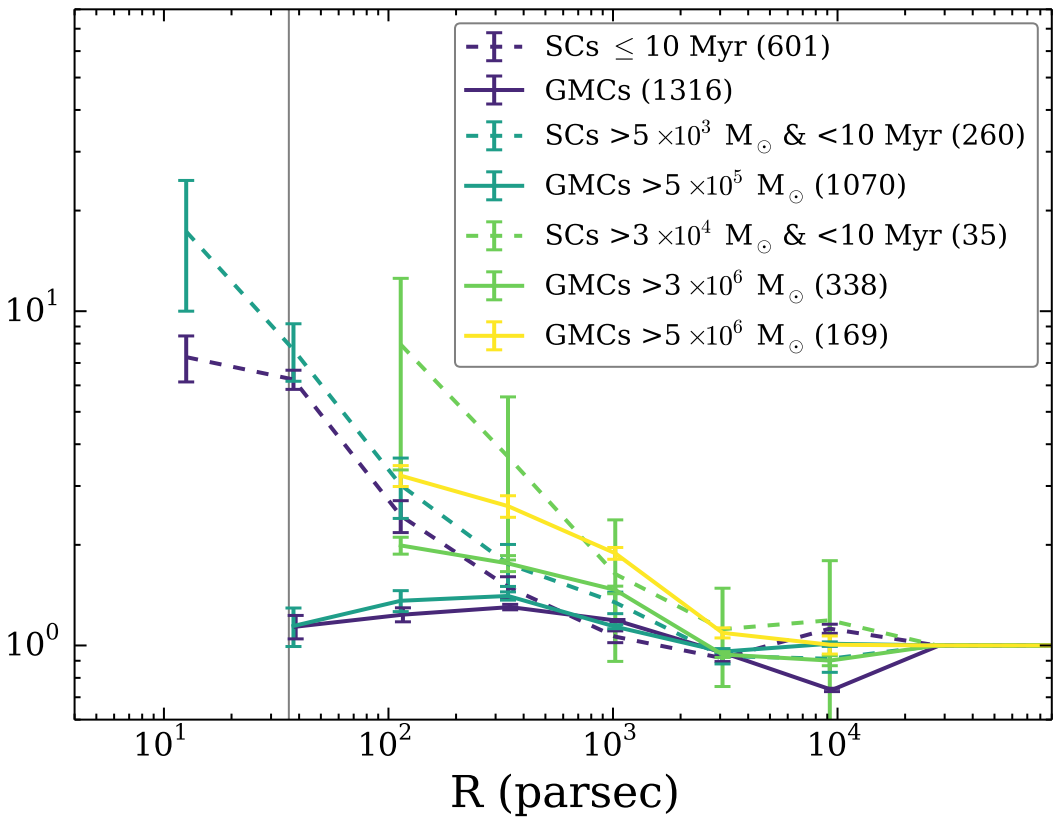

Figure 5.16. Two-point correlation function $1+\omega(\theta)$ with mass cuts for the GMCs. We compare the distribution of the most massive and youngest star clusters to the most massive GMCs on the far right panel. We show star clusters (dashed purple line) with mass cuts at 5000 $M_\odot$ (dashed blue line) and 3×10^4 $M_\odot$ (dashed green line). We also show the equivalent GMC distribution (solid lines) assuming a 1% SFE. The clustering in the distribution of the GMCs increase with mass, but at a given mass cutoff, the slopes are different from that observed for the youngest star clusters. The observed slope for the GMC distribution stars to increase for masses greater than 5×10^6 $M_\odot$ (yellow line), but there are not enough clusters to compare this distribution to assuming a SFE of a few percent.

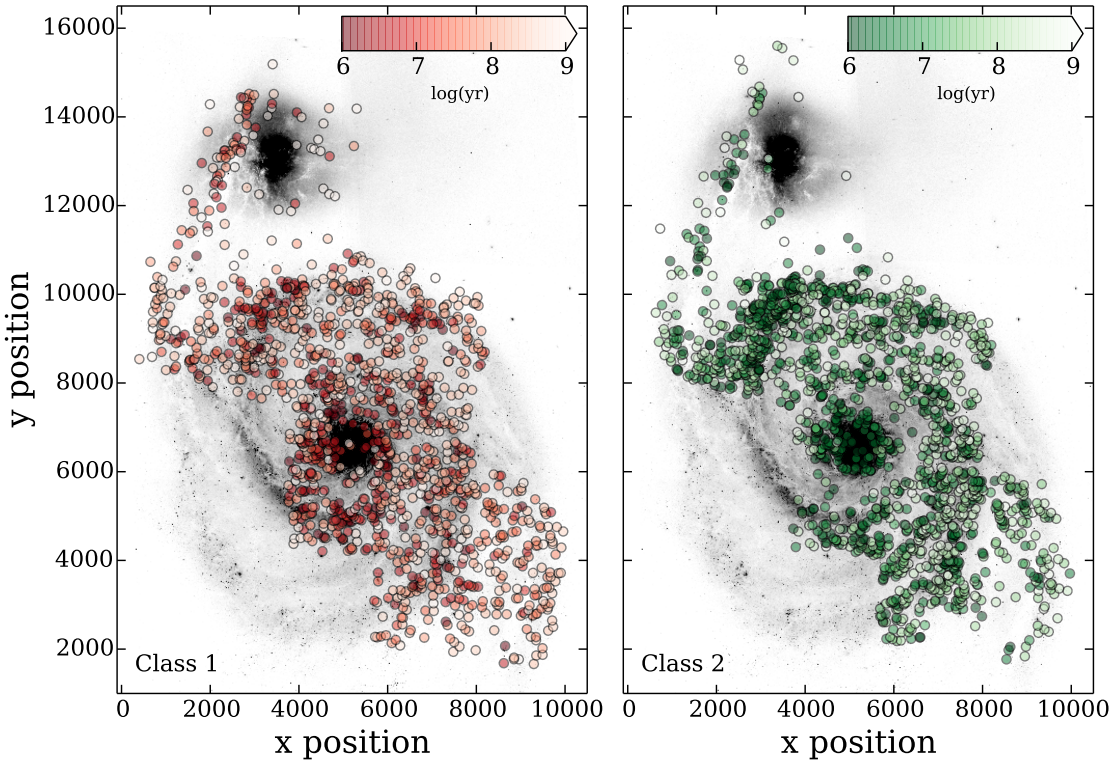

Figure 5.17. Class 1 (left) and 2 (right) clusters shown over the V-band image of NGC 5194. The star clusters are colored according to their ages with dark colors corresponding to younger ages and all clusters older than 1 Gyr are shown as white. The class 1 sources are much more uniformly distributed while the class 2 clusters are predominantly tracing the spiral arms and the center region. The typical ages of class 1 clusters are much older than the ages of the class 2 clusters (Figure 5.5). We do not show the class 3 associations. Clusters are mostly concentrated along the spiral arms, which is particularly noticeable for the youngest clusters of both classes.

located at distances greater than the co-rotation radius than for clusters within the co-rotation radius. More notably, the correlation length is also significantly larger, around ∼2 kpc, for clusters outside the co-rotation radius whereas the size of the structures for the clusters within the co-rotation radius is ∼200 pc. One of the biggest drivers of the observed difference is the lack of strong stellar spiral features within the inner region of the galaxy, resulting in significantly smaller star-forming complexes (Section 5.6.2.2 and Table 5.3).

This suggests longer lived star-forming regions reside at larger galactocentric radii. This is likely a result of lower shear within the outer regions of the galaxy, allowing clusters to live longer than in regions near the center of the galaxy. The effect of the local ambient environment in also driving the maximum correlation size is further discussed in Section 5.7.

5.6.2.4 Comparison to Previous Work

The clustering of the star clusters in NGC 5194 using the two-point correlation function was performed in a study by Scheepmaker et al. (2009) using star clusters derived with WFPC2 U and B, V, and I bands with ACS. Their final luminosity-limited cluster catalog contains 1580 clusters that range from 2.5 Myr to 500 Myr and is not visually inspected for contaminants. Their clustering shows a smooth decline with scale size and age, exhibiting recovered slopes of -0.4, consistent with projected fractal dimension of 1.2 in the young clusters ($\lesssim$30 Myr) with the slope becoming more shallow as the clusters age. We recover a slope of -0.3 for the whole sample, rising to -0.5 for the youngest clusters, consistent within the errors. The correlation lengths between the two studies are similar, becoming flat around a few kiloparsecs.

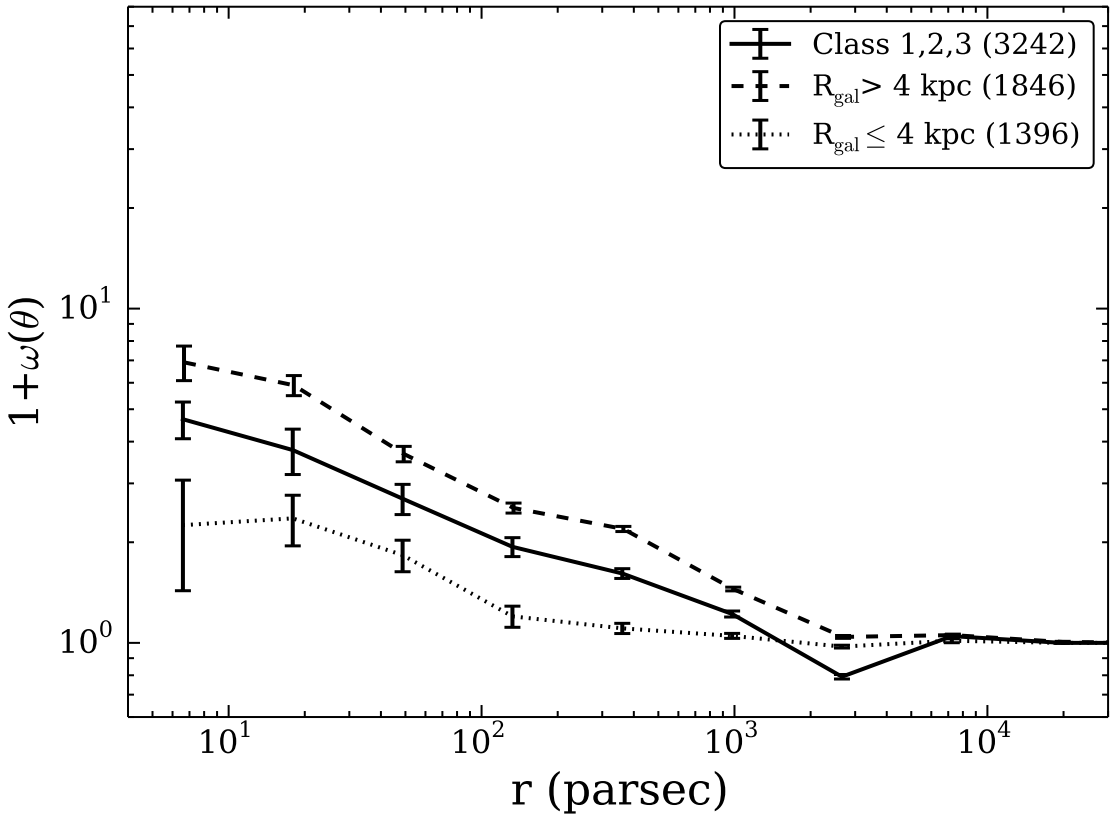

Figure 5.18. Two-point correlation function $1 + \omega(\theta)$ for star clusters at galactocentric radii greater/less than 4 kpc. The numbers in parentheses show the number of clusters in each classification. Star clusters further from the center of the galaxy reside in larger hierarchies, exhibiting greater, but shallower, correlation lengths than star clusters near the center.

5.7 Discussion

The physical mechanisms that give rise to the shape of the hierarchical structures is believed to be a direct consequence of turbulence in the ISM operating self-similarly at a range of scales (Elmegreen et al., 2000; Mac Low & Klessen, 2004; Hopkins, 2013a; Guszejnov et al., 2017), and responsible for triggering the fragmentation of large clouds into smaller fragments. We find that GMCs that reside within a spiral arms are more likely to host a star cluster (45%; Table 5.2) compared to those in the interarm (25%). This suggests that the majority of recent star formation is more likely to occur within GMCs located in the denser regions of stellar spiral arms and the association with star formation lasts for about one crossing time.

The correlation for star cluster complexes within the central 9 kpc exhibit correlation lengths of a few hundred parsec. By ∼50–100 Myr, the clustering signal has dispersed and is statistically the same as a random distribution, requiring a cluster velocity of only ∼1 km s^{-1} to achieve randomization within this timescale. The clustering scale length in the inner 4 kpc is significantly shorter than the clustering scale length of 2 kpc for clusters located beyond 4 kpc (Figure 5.18).

The shorter correlation sizes near the center are likely a result of the ambient environment that can influence the dissolution of individual clusters and unbound complexes through shear and pressure. The clustering becomes consistent with random at much smaller sizes in the inner region of the galaxy. This is consistent with the observed radial trend of clusters in the outer regions of NGC 5194 experiencing significantly less disruption (e.g., Silva-Villa et al., 2014; Messa et al., 2018b). The decrease of shear, turbulence, and ISM mid-plane pressure with increasing distance from the galaxy center also increases the survivability of bound clusters, and hence, the lifetimes and sizes of the complexes in which they reside (e.g., Kruijssen, 2012; Johnson et al., 2016). Indeed, simulations do show that star and star cluster formation depends on the shear content in pre-collapse GMCs and that presence of shear in disk

galaxies impedes the formation of very massive clusters, preferentially forming systems of smaller clusters and structures (Weidner et al., 2010). We would expect that if clusters cannot form, neither can clustering persist. While there is an age dependency on the amplitude of the clustering, younger star-forming regions do not necessarily exhibit shorter correlation lengths than older clusters (Figure 5.14). A recent study in NGC 2336 found that UV-bright star-forming knots also show evidence for older star-forming regions at larger radii compared to the central part of the galaxy, quite possibly arising from the lower tidal shear present at larger galactocentric distances (Rahna et al., 2018).

Through an analysis of average age between pairs of star clusters in Grasha et al. (2017b), we find that the global average of young (<300 Myr) star-forming regions in NGC 5194 is ∼950 pc in size with average ages of 83 Myr, similar to the size scale recovered with the two-point correlation function in this work. From the parameters derived for a typical star-forming region of NGC 5194, the velocity for a single crossing time is 13 km s^{-1}. This is consistent with the estimated velocity of 9.5 km s^{-1} we find that is required for star clusters to disassociate from their nearest molecular cloud given the distance and their current age from the closest GMC (Section 5.6.1.2). A study by Whitmore et al. (2014) in the Antennae galaxy shows that the timescale needed for GMC complexes with a radius of 200 pc to diffuse in a crossing time is ∼10 Myr. The results from Whitmore et al. are comparable to the values we derived in this work, with proportionally larger complex sizes and diffusion timescales within NGC 5194.

The typical crossing timescale for an arm GMCs from PAWS, with a radius of 40 parsec and a velocity dispersion of 8 km s^{-1}, results in a crossing timescale of 5 Myr (Schinnerer et al., 2017). This is comparable to the 4–6 Myr timescale we derived for when star clusters are still associated with their progenitor clouds. This suggests that the internal velocity, and thus, the crossing time scale of molecular

clouds may determine how long a star cluster remains associated. The natural time unit for a GMC is the free-fall time and describes the timescale for which the cloud experiences no other forces outside of its own gravitational collapse. The free-fall timescale only depends on the density ρ of the molecular gas, $t_{ff} = \sqrt{3\pi/32G\rho}$ and places a lower limit on a given cloud lifetime (Heyer & Dame, 2015). The lifetime of a typical GMC is expected to be a few free-fall timescales; observations show GMC lifetimes of ∼20–30 Myr (Meidt et al., 2015) within NGC 5194. The short timescales for clusters to become visible from their GMCs compared with the longer time-scales for which the clouds are destroyed is also expected from simulations (e.g., Dale et al., 2015). The 4–6 Myr timescale for the disassociation of star clusters from molecular gas will thus arise from both the relative time at which the young star clusters emerge due to cluster motions as well as a dependency on the timescale for star clusters to erode cavities within their molecular reservoirs from feedback effects (i.e., ionizing winds). Star clusters are small in comparison to the angular extent of the GMCs, and GMC 'destruction' remains a local process that does not encompass the GMC as a whole.

Molecular clouds are observed to evolve over time and exhibit different levels of star formation activity, such as in the LMC, where not all GMCs show evidence of star formation and more evolved and older clouds are more likely to be associated with optical stellar clusters Kawamura et al. (2009). The same study derives a dis-assocation timescale between star clusters and GMCs of $\sim 7 - 10$ Myr, however, no significant difference is observed in the properties (size or line width) between the varying evolutionary stages of the GMCs. Within this proposed picture for the evolutionary sequence of GMCs, most active star formation will lead to the dispersal of the cloud. However, the cloud population of NGC 5194 appears to undergo regrowth and transformation rather than a complete dispersal after star formation (Meidt et al., 2015), which suggests that the observed spatial separation between older star clusters

and GMCs cannot arise solely from GMC dissolution and must require a component of motion that acts to separate star clusters from their natal clouds.

5.8 Summary and Conclusion

The LEGUS project (Calzetti et al., 2015) is producing an extensive suite of high-quality, homogeneous, and accurate properties of clusters in nearby galaxies, which are crucial in order to address the role of star clusters in the star formation process in a consistent manner across galaxies. In this work, we present the methods used to test an off-the-shelf ML procedure to identify and create catalogs of star clusters for LEGUS, which we have adapted to galaxy NGC 5194. The goal is to decrease the time and increase the efficiency required for identification and classification of star clusters in nearby galaxies, currently performed via time-intensive visual inspection.

Our main results can be summarized as follows:

1. We use the LEGUS 4-class cluster classification of the visually identified star clusters as a training set to teach the ML algorithm to distinguish between different classes of clusters based on their features. These learned decision features are then applied to the remaining clusters that have no classifications. The ML techniques identify an additional 2148 clusters from a training set of 1226 star clusters that have been visually identified, thus bringing the total to 3374. Cross validation of the ML algorithm on the visually inspected catalog indicates that the ML has an overall (4-class) accuracy of 67%. For comparison, human classifiers agree among each other about 70%-75% of the time. This suggests that an off-the-shelf ML algorithm is already effective at providing classifications that are close to the accuracy rate of humans. However, outside this overlap region, the success of the ML algorithm decreases for certain classes of clusters, depending on the morphology of the cluster. Single peak, extended clusters are reliably identified at similar rates as observed in the visual training

sample whereas associations, which typically exhibit multi-peaked structure and color gradients, are not identified at the same rate and are interpreted as non-cluster contaminants. Thus, further developments of ML algorithms will need to investigate methods to bring the accuracy to higher levels for all classes, but specifically for the multi-peak clusters/associations.

2. The locations of the youngest star clusters are highly correlated with the molecular gas (Figure 5.11) from the GMC catalog from PAWS data. Star clusters that are associated (i.e., located within the footprint of a GMC) are significantly younger – median ages of 4 Myr – than star clusters that are unassociated with molecular clouds, which have median ages of 40 Myr. For the star clusters located within three radii of a GMC, we derive an ISM velocity of $\sim$10 km s^{-1} required for them to become dissociated from their natal clouds. We find no significant evidence that more massive GMCs are more likely to host multiple star clusters (Figure 5.12). The average distance between a star cluster and the nearest molecular cloud increases with galactocentric distance, corroborating a picture in which the separation of clusters from their birth locations is due to random motions (Figure 5.9). Furthermore, clusters that are at least as old as the lifetime of molecular clouds ($\sim$20 Myr) or older are typically located at distances of at least three radii from their nearest GMC. Thus, the timescale for cluster disassociation ($\sim$4-6 Myr) is mainly driven by the random motions that move the clusters away from the GMCs, and only partially impacted by GMC dissolution effects.

3. We implement the two-point correlation function to quantify the time-scale for the survival of star-forming hierarchical structures and to determine the sizes of the clustered regions. We find that the amount of clustering of star clusters decreases with increasing spatial scale (Figure 3.6), consistent with star clus-

ter formation occurring within hierarchical star-forming complexes of ∼1 kpc in size. The power law slope of the distribution, depending on the method of cluster identification, ranges from −0.3 to −0.4, where the clusters identified with the ML process exhibit a shallower distribution. The clustering depends strongly on the age of the clusters (Figure 5.14), indicative that star-forming hierarchies disperse with time, with randomization timescales of ∼50–100 Myr. The size and the strength of the clustering also depend on the galactocentric radius; larger clustered star-forming regions are preferentially located further from the galaxy center (Figure 5.18). This environmental dependence of the clustering of the star clusters is consistent with clusters near the center experiencing increased disruption. This is in agreement with the strong age-dependency of the survival of the cluster complexes with complexes near the inner region of the galaxy being smaller and showing distributions that are close to random. The correlation function for the GMCs shows a power law slope of −0.10, very shallow compared to the star clusters, consistent with being nearly randomized. This may suggest that the structure of GMCs is more easily randomized compared to that of star clusters.

CHAPTER 6

CONNECTING YOUNG STAR CLUSTERS TO CO MOLECULAR GAS IN NGC 7793 WITH ALMA–LEGUS

This chapter[1] presents initial ALMA–LEGUS observations of NGC 7793 that correlates the star clusters at sub-GMC resolutions of ∼15 pc.

6.1 Abstract

We present an investigation of the relationship between cloud properties and the associated stellar clusters in the nearby flocculent galaxy NGC 7793. We combine the star cluster catalog from the UV/optical HST LEGUS (Legacy ExtraGalactic UV Survey) program with the ∼15 parsec resolution of CO(2–1) observations from ALMA of the inner $\sim 2 \times 3$ kpc of NGC 7793. We find a strong spatial correlation between young star clusters and giant molecular clouds (GMCs) such that all clusters still associated with a molecular cloud (i.e., located within the physical footprint) are younger than 11 Myr and display a median age of 2 Myr. The age distribution increases gradually as the star cluster–GMC distance increases, with star clusters that are spatially unassociated with molecular gas exhibiting a median age of 7 Myr. Thus, star clusters are able to emerge from their natal clouds long before the timescale required for clouds to disperse. To investigate if the hierarchy observed in the stellar components is inherited from the GMCs, we quantify the amount of clustering in the spatial distributions of the components and find that the star clusters within the

[1]These results are based on research published by Grasha et al. (2018b), and is reproduced by permission of the AAS.

ALMA coverage exhibit a fractal dimension slope of -0.61 ± 0.03, twice as steep as the observed molecular cloud hierarchy with slope of -0.28 ± 0.04. This is among the first observational studies to examine the temporal association between GMCs and star clusters, and while we find enhanced spatial correlations between clouds and young stellar clusters on scales $\lesssim$40 pc, there is a lack of mirroring between the observed hierarchies of star formation and GMCs with the star clusters exhibiting a significantly more fractal distribution. We find, however, that the spatial distributions are consistent with each other when comparing the most massive GMCs to the youngest and most massive star clusters when assuming a star formation efficiency of a few percent. This suggests that it is important to match the mass limit star clusters and GMCs to properly compare how clusters and GMCs respond to clustering. The analysis of the association between GMCs and star clusters provides reliable estimates of their temporal association and the evolutionary timescales to better our understanding the gas and star formation cycle across galaxies.

6.2 Introduction

Star formation is a hierarchical, scale-free process in both space and time (Efremov & Elmegreen, 1998) spanning from individual stars to entire star-forming galaxies, and is a consequence of giant molecular clouds (GMCs) converting their molecular mass into newly formed stars via fragmentation. The observed hierarchical nature of stars and star clusters (e.g., Gomez et al., 1993; Zhang et al., 2001; Odekon, 2008; Gouliermis et al., 2014; Sun et al., 2017a,b), inherited from their progenitor clouds/cores (e.g, Elmegreen & Falgarone, 1996; Johnstone et al., 2000, 2001; Sánchez et al., 2010), is believed to be imposed by the hierarchical nature of turbulence (Elmegreen & Efremov, 1996, 1997; Padoan & Nordlund, 2002; Hopkins, 2013a,b) throughout the interstellar medium (ISM).

The Legacy ExtraGalactic Ultraviolet Survey[2] (LEGUS, HST GO–13364; Calzetti et al., 2015), a Hubble Space Telescope (HST) Treasury program of 50 local ($\lesssim$15 Mpc) galaxies observed in the UV and optical regimes have allowed for unprecedented investigations into the star formation hierarchies across a diverse population of galaxies, improving our understanding of the processes of star formation in the local universe (Elmegreen et al., 2014; Gouliermis et al., 2015b, 2017; Grasha et al., 2015, 2017a,b). Extending these studies of the characteristic size of star-forming structures to higher redshifts, where the vast majority of stars seen today were formed (Madau & Dickinson, 2014), is the only way to differentiate if the observed clumpiness arises from in-situ structure formation from instabilities within the galactic disk (e.g., Dekel et al., 2009; Bournaud et al., 2014) and/or as a result of galaxy interactions (e.g., Di Matteo et al., 2008) through exploring the relation between clump sizes and galactic properties (e.g., Wisnioski et al., 2012).

The dynamical structure of a galaxy is capable of impacting the local star formation process (e.g., Renaud et al., 2013), and consequentially, the overall organization and survival of star-forming structures. As such, high-resolution studies of local galaxies are ideal to study the complex interplay between gas and the subsequent impact on star formation. Early results from the Plateau de Bure Interferometer Arcsecond Whirlpool Survey (PAWS; Schinnerer et al., 2013; Pety et al., 2013) reveal that the spiral structure of M 51 impacts the gas density of the local environment, which alters the properties of the GMCs, and thus, star-formation. This affects the overall organization and structure of the ISM at scales of individual GMCs (Hughes et al., 2013), which will inevitability impact the products of star formation (Messa et al., 2018b).

[2]https://archive.stsci.edu/prepds/legus/

The work in this paper focuses on the flocculent galaxy NGC 7793, one of the closest galaxies within the LEGUS. A primary science goal of LEGUS is aimed to further our understanding of the connection between localized sites of star formation and the global star formation process of entire galaxies. LEGUS has vastly increased the number of high-quality catalogs of young star clusters over a broad range of galactic environments, improving our understanding of the parameters responsible for star cluster formation and evolution in a homogeneous and consistent manner (Kim et al. in prep).

The goal of this paper is to investigate the connection between molecular gas and associated stellar populations, a topic previously examined in another LEGUS galaxy, the interacting spiral system NGC 5194 (Grasha et al., 2018a) with molecular CO data from PAWS (Colombo et al., 2014a). This paper complements the prior study with recent CO observations for NGC 7793 (Bittle et al., 2018) with the Atacama Large Millimeter/submillimeter Array (ALMA) and serves as an excellent comparison to connect the properties of star clusters and molecular gas to the spiral structure in two different galactic systems. The lifetimes of the star-forming complexes were previously investigated for NGC 7793 in Grasha et al. (2017a,b), where we find relatively small size scales of $\sim$200 parsec for the coherence of star formation, consistent with expectations for flocculent galaxies. NGC 7793 is additionally a cluster-poor system, which makes it an interesting system to study the properties of star clusters in the context of their connection to the spiral structure and the timescale for star formation to remain associated with molecular clouds. These results are quite relevant to the longstanding issue of molecular cloud lifetimes, both before and during star formation, and its relation to the environment (e.g., Kawamura et al., 2009; Miura et al., 2012; Whitmore et al., 2014; Heyer & Dame, 2015).

This Chapter is organized as follows. The galaxy selection and reduction process as well as the acquisition and reduction of the ALMA data are described in Sec-

tion 6.3. The cluster selection and identification process is described in Section 6.4. The correlation of the star clusters to the molecular gas is described in Section 6.5 and we quantify the hierarchy of the star clusters and the GMCs in Section 6.5.4. Finally, we summarize the findings of this study in Section 6.6.

6.3 NGC 7793

6.3.1 The HST UV/Optical Observations

In this paper, we select NGC 7793 from the LEGUS survey and observe the molecular gas with ALMA (Section 6.3.2) to study the impact of spiral arm structures on the properties of both star clusters and the molecular gas reservoir. NGC 7793 is a nearby (3.44 ± 0.15 Mpc assuming a distance modulus of 27.68 ± 0.09; Pietrzyński et al., 2010) SAd flocculent galaxy in the Sculptor group, with a relatively low star formation rate (SFR(UV) ~ 0.52 $M_\odot$ yr^{-1}; Calzetti et al., 2015), and a central oxygen abundance of 12 + log (O/H) $= 8.50 \pm 0.02$ and an oxygen abundance gradient of -0.0662 ± -0.0104 dex/kpc (Pilyugin et al., 2014).

The HST observations of NGC 7793 cover two pointings, the western and eastern part of the galaxy. The eastern pointing is observed in all five bands (F275W, F336W, F438W, F555W, F814W) with the WFC3. The west pointing is observed in only three bands (F275W, F336W, F438W) with the remaining two bands (F555W, F814W) taken from archival ACS observations (GO–9774; P.I. S.S. Larsen). All archival ACS images are re-reduced using the same pipeline as the UV and U images with WFC3/UVIS from the LEGUS project. Reduced science frames are drizzled to a common scale resolution, to match the native WFC3 pixel size (0.0396 arcsec/pixel), corresponding to a resolution element of 0.66 parsec/pixel. The frames have all been aligned and rotated with North up. We refer the reader to the detailed descriptions of the standard reduction of the LEGUS data sets in Calzetti et al. (2015).

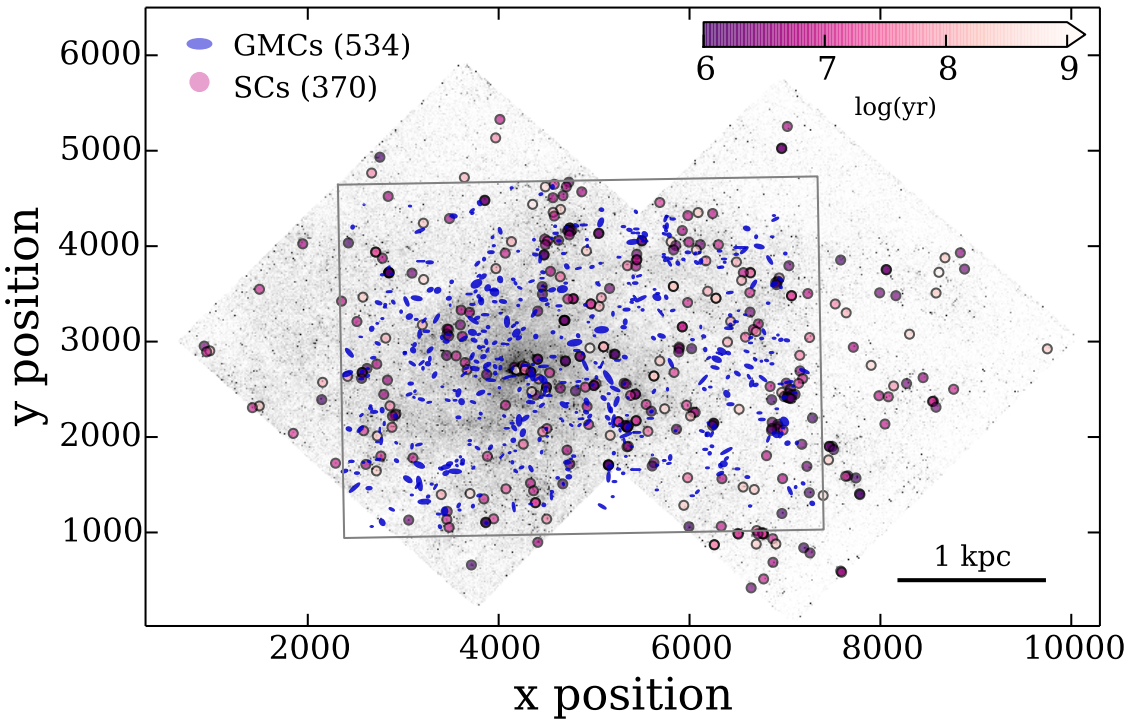

Figure 6.1. Positional location and sizes of the GMCs (blue ellipses) and star clusters (circles) in NGC 7793 over the UVIS/F438W image. The star clusters are colored according to their ages, where the youngest are darker pink and the oldest are white. The gray rectangle shows the outline of the ALMA coverage. Star clusters not located within the ALMA coverage are excluded from all star cluster–GMC comparisons.

Figure 6.1 shows the gray-scale V-band image of both the east and west UVIS pointings overlaid with the positions of the GMCs and ALMA coverage (Section 6.3.2) and the star clusters (Section 6.4).

6.3.2 The ALMA CO Observations

A detailed quantification of the data acquisition, reduction, and creation of the GMC catalog, along with the properties of the molecular clouds (mass function, power spectrum, etc.) and how they depend on their location within NGC 7793, will be

described further in Bittle et al. (2018). We list here the details necessary for this study.

We observed the CO (2–1) transition with ALMA band 6 (230.36366 GHz) in the inner 180"×114" (3 × 2 kpc) of NGC 7793 (ALMA programs 2015.1.00782 & 2016.1.00674; PI: K.E. Johnson). The total integration time is 3 hours with the 12-m array with 149 pointings at 72.6 seconds of integration per pointing with a mosaic spacing of 12.9" between pointings. Our angular resolution of 0.85" allows us to resolve sub-GMC sizes (8 pc scales) and we are sensitive to emission on scales as large as 11" (∼175 pc), which allow us to recover the largest molecular complexes in the galaxy. We have sufficient sensitivity to detect molecular clouds with masses of $\sim 10^4$ $M_\odot$, matching the associated mass scale of the stellar clusters with $\gtrsim 10^4$ $M_\odot$. We also achieve a velocity resolution of 1.2 km s^{-1}, sufficient to resolve individual clouds with expected velocity dispersions of order 2–3 km s^{-1}. The final rms noise of the observations is ∼4 mJy/beam.

6.3.3 Creating the GMC Catalog

We create the GMC catalog from the CO(2-1) position-position-velocity data with the CPROPS segmentation algorithm for spectral line emission (Rosolowsky & Leroy, 2006). After identifying a typical cube rms noise, local maxima are identified above 4-σ. Emission above 2-σ that lie in at least two continuous velocity channels around an identified peak is then assigned to that peak. For a local maximum within a kernel range of another local maximum, we define a shared contour level. If the peaks both lie above 3.5-σ of the isophote, they are determined to be unique entities and the emission is separated appropriately. From these finalized identifications of clouds, we are able to measure cloud properties such as size (R_{GMC}), line width (σ_v), and luminosity (L_{CO}) for each identified cloud.

Table 6.1. Global Properties of the GMCs in NGC 7793

	M	R	L_{CO}	σ_v
	($10^4 M_\odot$)	(pc)	(10^3 K km s^{-1} pc^2)	(km s^{-1})
All	$3.2^{+3.7}_{-1.5}$	$13.4^{+6.5}_{-3.9}$	$5.2^{+6.0}_{-2.4}$	$2.6^{+0.7}_{-0.6}$
≤ 1 kpc	$3.4^{+4.6}_{-1.6}$	$13.4^{+6.2}_{-4.4}$	$5.5^{+7.4}_{-2.6}$	$2.6^{+0.7}_{-0.6}$
> 1 kpc	$2.9^{+2.6}_{-1.3}$	$13.5^{+7.0}_{-3.5}$	$4.8^{+4.1}_{-2.2}$	$2.7^{+0.7}_{-0.5}$

Notes. Columns list (1) Environment, either all, ≤ 1 kpc, or >1 kpc; (2) M, the virial mass in $M_\odot$; (3) R, the deconvolved effective radius in pc; (4) L_{CO}, the integrated CO luminosity in K km s^{-1} pc^2; and (5) σ_v, the deconvolved velocity dispersion in km s^{-1}. We report the median, lower quartile (Q25) and upper quartile (Q75) of the distributions.

By construction, the GMCs represent significant peaks in the CO emission and we assume that these correspond to the cluster-forming structures. Using the position angle, radius, and semimajor axis, we represent the GMCs as ellipses in Figure 6.1. Table 6.1 lists the general properties of the GMCs in NGC 7793 for the entire population as well as those that lie at distances less/greater than galactocentric distances of 1 kpc (Figure 5.9).

6.4 Creating the Visually-Identified Star Cluster Catalogs

A detailed description of the cluster selection, identification, photometry, spectral energy distribution (SED) fitting procedures, and completion limits for the LEGUS galaxies is presented in Adamo et al. (2017). We summarize here briefly the aspects that are important for the current paper.

6.4.1 Star Cluster Selection

The LEGUS process of producing cluster catalogs is a multi-step process, with an initial automated extraction of the cluster candidates (Section 6.4.1), followed by a visual identification of a sub-sample of the brightest clusters to remove contaminants

(Section 6.4.2) in order to create the final science-ready, visually identified cluster catalog.

The initial *automated catalog* of star cluster candidates is extracted from a white-light image generated by using the available photometric bands (see Calzetti et al., 2015) with source extractor (SExtractor; Bertin & Arnouts, 1996). The SExtractor parameters are optimized to extract sources with at least a 3σ detection in a minimum of 5 contiguous pixels. This automatic procedure returns the positions of candidate clusters within the image and the concentration index (CI; magnitude difference of each source within an aperture of 1 pixel compared to 3 pixels). The CI is related to the size of stellar systems (Ryon et al., 2017) and can be used to differentiate between individual stars and stellar clusters. Star clusters, in general, have more extended light profiles, and therefore, larger CI values compared to individual stars. The CI reference value used to distinguish between unresolved sources (stars) and resolved sources (candidate clusters) within NGC 7793 is 1.3 mag; we disregard any sources with a CI value less than this reference value.

Standard photometry is performed on the cluster candidates using a science aperture radius of 5 pixels and a sky annulus at 7 pixels with a width of 1 pixel. Average aperture corrections are estimated using a cluster control sample (see Adamo et al., 2017), estimated as the difference between the magnitude of the source within a 20 pixel radius with a 1 pixel sky annulus minus the magnitude of the source at the science aperture. These average aperture corrections in each filter are applied to the standard photometry of all the clusters. Corrections for foreground Galactic extinction (Schlafly & Finkbeiner, 2011) are also applied to the photometry. All positions are corrected for an inclination of 47.4 degrees.

All cluster candidates detected in at least four of the five bands with photometric error $\sigma \leq 0.3$ undergo spectral energy distribution (SED) fitting procedures and error propagation as described in Adamo et al. (2010, 2012) to extract the age, mass, and

color excess E(B–V) of each source. The SED fitting analysis is performed with deterministic Yggdrasil single simple stellar population (SSP) models (Zackrisson et al., 2011). The Yggdrasil spectral synthesis code is based on the Starburst99 (Leitherer et al., 1999) stellar population synthesis models with the photoionized nebular emission predicted by Cloudy (Ferland et al., 1998, 2013). All cluster catalogs for the LEGUS galaxies implement a Kroupa (2001) universal initial mass function (IMF), but see Ashworth et al. (2017) for variable IMF cases. For NGC 7793, the cluster catalog SED is derived using the Padova-AGB SSP stellar isochrone tracks (Vázquez & Leitherer, 2005) and a starburst attenuation curve (Calzetti et al., 2000) with the assumption that stars and gas undergo the same amount of reddening. For wavelengths longer than $\sim$ 3000 Å, attenuation and extinction dust models are similar in shape and normalization, and thus, the adopted dust model does not greatly impact the results (Grasha et al., 2017a).

The star cluster properties of NGC 7793 are also derived using a Bayesian analysis method with stochastically sampled cluster evolutionary models (Krumholz et al., 2015a) using the Stochastically Lighting Up Galaxies (SLUG; da Silva et al., 2012; Krumholz et al., 2015b) code. SLUG returns the full posterior probability distribution function (PDF) of the physical properties of each cluster rather than a single best fit. This approach does not assume a fully sampled IMF, which becomes especially relevant for the accurate derivation of cluster properties in cluster masses below $\sim 10^{3.5}$ $M_{\odot}$ (Cerviño & Luridiana, 2004), which is near the completeness limit of our catalogs for cluster ages up to 200 Myr at distances of $\sim$10 Mpc (Adamo et al., 2017). NGC 7793 is located at 3.44 ± 0.15 Mpc, and as a result, we expect the cluster catalog to be complete to significantly less massive clusters. We do find that the relatively nearby distance of NGC 7793 results in a catalog with cluster properties that are predominately younger and less massive compared to the more distant LEGUS galaxies (see Grasha et al., 2017a). Indeed, the cluster population shows a median

age of 7 Myr and a median mass of only 1080 $M_{\odot}$. We do not impose a mass cut to account for a stochastically sampled IMF and/or completeness issues as only 63 clusters in our sample are more massive than 5000 $M_{\odot}$. We examine the effects that the method of deriving the ages has on our results and find that the global properties of the entire ensemble of clusters are consistent between the two methods and the science results appear to be relatively insensitive whether the properties are derived deterministically or stochastically. We do perform checks by repeating calculations with placing mass cuts and find that our results are not biased with the inclusion of low-mass clusters.

6.4.2 Visual Inspection and Star Cluster Classification

After the extraction of the clusters and the SED fitting procedure, all clusters with an absolute magnitude brighter than −6 mag in the V-band undergo visual inspection by a minimum of three independent classifiers within the LEGUS team to procure the final *visual catalog*. This −6 magnitude limit is defined by the detection limits of the LEGUS sample and enables selection down to a ∼1000 $M_{\odot}$, 6 Myr old cluster with E(B–V) = 0.25 (Calzetti et al., 2015). Only 484 cluster candidates are brighter than the magnitude cut off to undergo the visual classification procedure.

The visual classification is performed within ds9 on a V-band and three-color composite image in addition to the surface contours, radial profiles, and surface plots of each source (see Adamo et al., 2017). The LEGUS cluster classification is based on the morphology and color of each source and is necessary in order to exclude non-cluster contaminants within the automatically extracted catalog to ensure the robustness of the final cluster catalog. Each sources gets classified under one of four classes: (1) symmetric and centrally compact star clusters, usually uniform in color; (2) compact, asymmetric star clusters with some degree of elongation, usually uniform in color; (3) multiple-peaked profiles that show an underlying diffuse emission, color

gradients are common; and (4) non-clusters, including but not limited to foreground stars, asterisms, background galaxies, saturated/bad pixels, etc. The final cluster catalog for NGC 7793 contains 370 clusters of class 1, 2, and 3.

In general, class 1 and 2 clusters are older and are potentially gravitationally bound systems with ages greater than their crossing time (Ryon et al., 2017) whereas class 3 objects are in general much younger, and due to their multi-peak nature, we refer to these systems as compact associations. Initial work in other LEGUS galaxies show that the morphological classification may contain information about the dynamical state of the clusters as well (Adamo et al., 2017; Grasha et al., 2017a). The class 3 compact associations have spatial distributions that are significantly more structured compared to the distribution of class 1 and 2 clusters as a result of the younger age distribution of class 3 associations. In this work, we compare the total star cluster sample to that of the GMCs and we refer the reader to Grasha et al. (2017a) for a more detailed discussion on the impact of the classification, thus age-dependency, on the dissolution of the large-scale clustering observed for the star clusters. NGC 7793 is one of the closest galaxies in the LEGUS sample, and as such, we have sufficient resolution to resolve down to very compact or loose associations that may otherwise be missed in more distant systems. 42% of the clusters in the NGC 7793 catalog are class 3 associations, among the highest rate out of all the LEGUS cluster catalogs (see Figure 2 of Grasha et al., 2017a). Thus we are confident that we are efficient at identifying what may be considered young and loose stellar aggregates within NGC 7793.

6.5 Results and Analysis

6.5.1 Separation of Young Star Clusters and GMCs

To connect the young star clusters to their environment, we compare the projected spatial locations of the clusters to the projected location and sizes of peaks in the

molecular gas. Previous observations show that stars and star clusters quickly become unassociated with the GMCs from where they are born, either due to drift or from blowing out the gas. The timescale for most clusters to become exposed is as early as 2 Myr within M 83 (Hollyhead et al., 2015), ∼5 Myr from a study of the Antennae galaxy (Whitmore et al., 2014), and within M 33, star exposed clusters show a peak around 5 Myr with an embedded phase that lasts only 4 Myr (Corbelli et al., 2017). Within Milky Way molecular clouds, the lifetimes of various star formation phases can be exceptionally short, $\lesssim$1 Myr (Battersby et al., 2017). In this study, our primary science goal is to determine the timescales for the association of GMCs with star formation and how quickly star clusters move away from their natal origins. If a star cluster and a GMC overlap in projection, we consider that pair an association.

6.5.1.1 Shortest Distance between Clusters and GMCs

We take all the star clusters within the ALMA coverage, reducing the total clusters in the catalog from 370 to 293, and measure the shortest distance to the center of the nearest GMC. Figure 4.8 shows the distribution of the shortest star cluster – GMC distance. For the entire sample, the median of the distances is 53±5 pc. The youngest star clusters (ages less than 10 Myr) have a median star cluster–GMC distance that drops to 41 ± 4 pc and the remaining star clusters that are older than 10 Myr have a median cluster–GMC pair distance of 66 ± 5 pc. Younger star clusters are closer in proximity to a GMC than older star clusters despite the small numbers of star clusters and GMCs in NGC 7793.

In Figure 6.2 we also show the expected distribution expected if the same number of the star clusters are randomized within the ALMA field of view. There is a clear excess in the distribution of the youngest clusters with ages $\lesssim$10 Myr exhibiting strong spatial correlations with the GMCs. The older star clusters (>10 Myr) show much

weaker correlations, with distributions that are almost consistent what is expected from a random cluster population.

The galactocentric distance r_{GC} of the closest star cluster/GMC pairs has a bigger influence on the age difference and we investigate this affect by dividing the sample in half at 1 kpc. Clusters located closer than $r_{GC} < 1$ kpc have a median SC–GMC separation of only 39 ± 4 pc, rising to 64 ± 7 pc for clusters beyond 1 kpc (Figure 6.3). In general, we do expect more things to be located in the central kpc of a galaxy, so it may not be surprising that statistically separations are closer within the inner region of the galaxy. However, this does not explain the trend for an increasing star cluster – GMC separation with increasing age, also demonstrated in NGC 5194 (Grasha et al., 2018a). We further quantify this in Section 6.5.2.

6.5.2 Properties of Star Clusters Associated with GMCs

Under the clustered star formation model, with the proper multi-wavelength observational data set, we can quantify the timescale of emergence, and thus the velocity required for star clusters to disassociate from their natal molecular clouds. We investigate the disassociation timescale by tracking how the distribution in the age of the cluster populations changes as a function of the distance to the center of every GMC. As star clusters do disperse from their GMCs with time (Grasha et al., 2018a), isolated clusters will exhibit significantly older ages compared to clusters that are located within, or even in close proximity to, a GMC. We do not allow double counting and each cluster can be assigned to only one GMC. In situations where multiple clouds lie on top of each other, the star cluster is assigned to the closest GMC, and in most cases, that is always the most massive one as well.

Within NGC 7793, we find 13 star clusters (4%) lie within the extent of 12 GMCs (2%). 31 clusters are located at distances just outside their nearest GMC ($1 \times R_{\rm GMC}$), but less than 2 radii of a GMC center. 26 star clusters located at distances greater

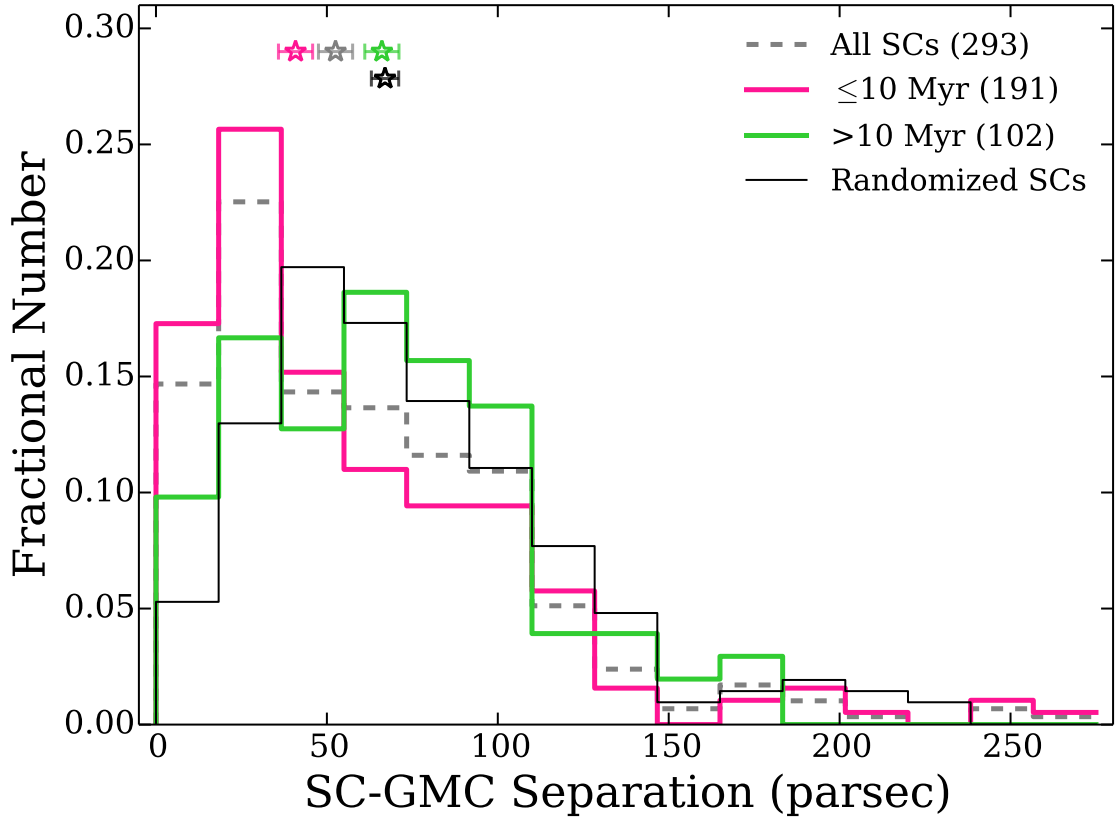

Figure 6.2. Fractional distribution of the shortest distance for each star cluster to the closest GMC for NGC 7793. The clusters are divided into two age bins for ages. The star symbols show the median value for each distribution along with the 1σ uncertainties from bootstrap estimates based on 10,000 samples. The global average (dashed gray) for the shortest distance between each star cluster to the closest GMC is 53 ± 5 parsec. Star clusters with ages less than 10 Myr (pink) show shorter distances of 41 ± 4 pc versus 38 ± 4 pc compared to ages older than 10 Myr (green) at 66 ± 5 pc. Younger star clusters are closer in proximity to a GMC than older star clusters and the difference is significant given the scatter and there is a clear excess observed for the youngest (<10 Myr) star clusters given the expected distribution if the star clusters were distributed randomly (thin black line, median value of 67 ± 4 pc).

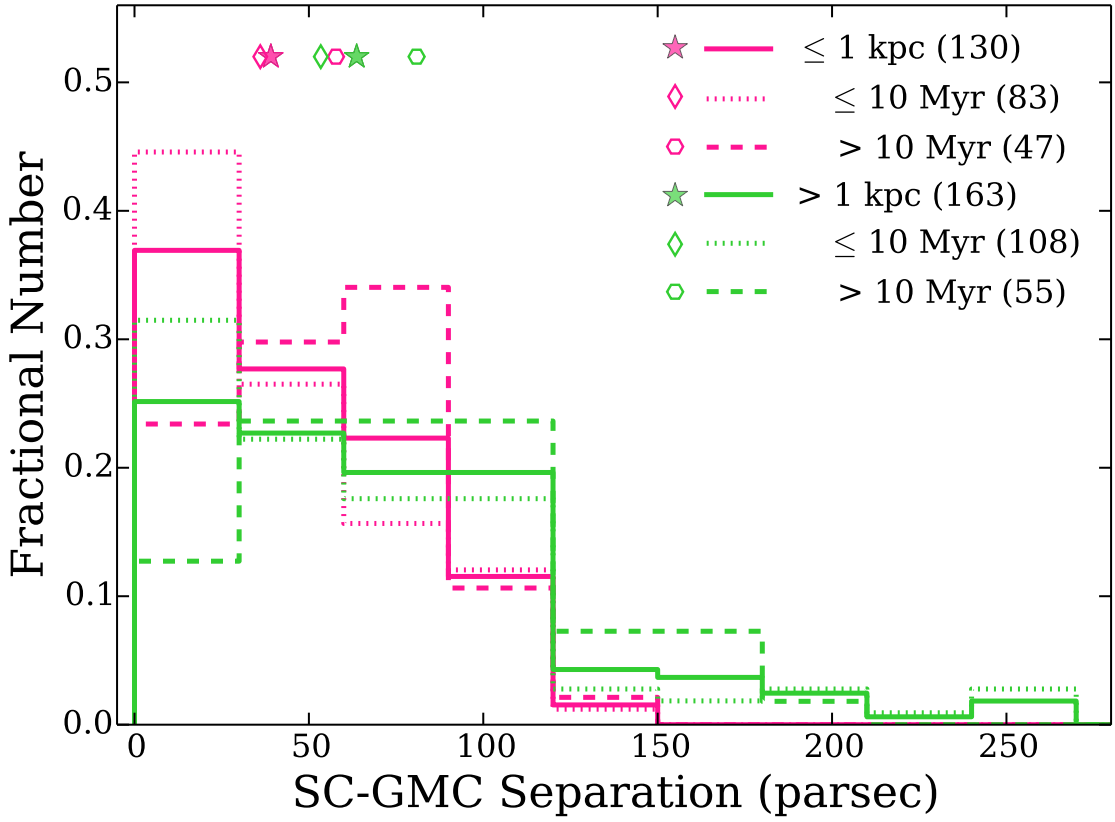

Figure 6.3. Fractional distribution of the shortest distance for each star cluster to the closest GMC broken into galactocentric radial bins and age bins. The symbols show the median value for each distribution. Star clusters less than a galactocentric radius of 1 kpc (pink) are on average 39 ± 4 pc from the nearest GMC with ages less than 10 Myr (dotted pink line) showing slightly shorter mean distances of 36 ± 4 pc compared to ages greater than 10 Myr (dashed pink line) at 58 ± 10 pc. Star clusters at galactocentric distances greater than 1 kpc (green) are on average 64 ± 7 pc from the nearest GMC, with mean distance decreasing for the youngest ($\leq$10 Myr; dotted green line) to 53 ± 9 pc and increasing to 81 ± 8 pc for the older clusters (>10 Myr; dashed green line). On average, younger star clusters are closer in proximity to a GMC than older star clusters although the distance from the center of the galaxy has a larger impact with the average SC–GMC distance increasing significantly with increasing galactocentric distance. Errors are bootstrap estimates from 10,000 samples and are not shown on the plot.

than 2 but less than 3 radii away from their nearest GMC. The remaining 224 star clusters (76%) are unassociated with any cloud (i.e., located at distances greater than $3 \times R_{\rm GMC}$ from the center of their nearest GMC). We select $3 \times R_{\rm GMC}$ as the definition of an "unassociated" as that is the distance where the star clusters show median ages that are above that of the global stellar population.

Figure 6.4 shows the distribution of the cluster ages for those with and without associations to GMCs. The median age of all clusters is 6 Myr, which is 3 times older than clusters located in proximity to GMCs at both <1 $R_{\rm GMC}$ and 1–2 $R_{\rm GMC}$. Clusters that are between 2 and 3 radii from a GMC have median ages of 3 Myr. Star clusters that are unassociated with any GMC are on average older than all clusters and those associated with GMCs. The small range in age sampled by the cluster population makes the distribution uncertain for increasing distance from a GMC, but the clear excess in the youngest clusters (Figure 6.4) at the shortest distances (Figure 6.3) spatial correlation of clouds and young stellar clusters that is lost rapidly. The young ($\lesssim$ 5 Myr) star clusters in this study can have age uncertainties that are a considerable fraction of the actual assigned value. It is difficult to assess a true difference between 2 Myr and 3 Myr old clusters given our age uncertainties, but the fact that the median age changes dramatically from clusters closest to a GMC of 2–3 Myr to 7 Myr for those greater than $3R_{\rm GMC}$ is a reliable measurement. We estimate the 1σ uncertainties on the median age estimates from bootstrap measurements based on 10,000 samples. The relatively low scatter given the small age range corroborates the significance in the observed age trend (significant at the 3.4–σ level due to very small cluster numbers).

Table 6.2 lists the properties of the star clusters given their distance from the nearest GMC. The trend for younger clusters to be in close proximity to molecular clouds is expected as these exposed stellar systems have started to evacuate their surrounding natal material, but have not lived long enough to either traveled far

enough to have erased the imprint of their birth location or to completely clear away the molecular material in their immediate vicinity (e.g., Corbelli et al., 2017). The ages we recover here are younger than typical GMC dissolution timescales of ∼10–30 Myr from both observations and theoretical expectations (e.g., Engargiola et al., 2003; Murray, 2011; Dobbs & Pringle, 2013; Heyer & Dame, 2015; Meidt et al., 2015; Grudić et al., 2018; Jeffreson & Kruijssen, 2018).

We find that distribution of E(B–V) values for a given star cluster are unaffected by their proximity to a GMC. The global median for all star clusters has an A_V of 0.49 assuming a starburst attenuation curve. Star clusters that are still associated with a GMC exhibit median A_V's of 0.45 whereas star clusters unassociated with GMCs ($> 3R_{\rm GMC}$) actually show slightly larger values of 0.51. The scatter in E(B–V) for a given age range is considerable and correlates poorly with age (Bastian et al., 2005; Adamo et al., 2010; Messa et al., 2018). The difference in age between star clusters inside/outside GMCs therefore cannot be explained by significantly higher extinction affecting the star clusters within the spiral arms or GMCs as the star clusters are already exposed, as shown by the resulting A_V values, indicating the significance of this result.

We take all 69 star clusters located at distances less than 3 $R_{\rm GMC}$ and find that the average velocity required for a star cluster to travel to its present location given its age, assuming it started from the center of its nearest cloud, is 6.2 ± 0.9 km s^{-1}. This velocity is larger than the velocity dispersion in individual GMCs (Table 6.1). These direct measurements of the age and velocity for the dissociation of clusters from their natal clouds are in agreement with expected ISM velocities and timescales for the dissolution of clusters from their natal hierarchical complexes between 40–60 Myr within NGC 7793 (see Section 6.5.4 and Grasha et al., 2017a).

The young star clusters will naturally erode the GMCs in which they are embedded, creating growing cavities caused by the winds and ionizing radiation of their

Table 6.2. Associated Cluster Properties in NGC 7793

Region	N_{SC}	N_{GMC}	Age (Myr)	$Mass_{SC}$ ($M_\odot$)	E(B–V) (mag)
$\leq 1R_{GMC}$	13	12	2(1)	540(280)	0.45(0.15)
1– 2 R_{GMC}	31	22	2(1)	1100(370)	0.53(0.17)
2–3 R_{GMC}	25	20	3(2)	690(490)	0.32(0.15)
Unassociated	224	492	7(1)	1150(260)	0.51(0.04)
Total	293	534	6(1)	1040(212)	0.49(0.04)

Notes. Properties of star clusters depending on their association with a GMC. Columns list the: (1) radial distance of the star clusters from a GMC; (2) number of star clusters; (3) number of GMCs; (4) median age of the star clusters; (5) median mass of the star clusters; and (6) median E(B–V) values. The 1σ uncertainties are bootstrap estimates based on 10,000 samples.

massive stars (see, e.g., Lada, 1987). However, these are effects visible in the vicinity of the clusters while the remainder of the GMCs is not affected. The clusters are small in comparison to the angular extent of the GMCs, and thus the destruction of the GMC by the young star clusters remains a local process that does not encompass the GMC as a whole given the timescales we measure and the typical lifetimes of a cloud.

6.5.3 Comparison between NGC 7793 and NGC 5194

In Grasha et al. (2018a) we performed a similar study to constrain the timescales of the association of GMCs and star clusters in NGC 5194 (7.7 Mpc) by combining the LEGUS cluster catalog ($\sim$ 1300 star clusters) with the GMC catalog ($\sim$1300 GMCs) at 40 pc resolution (Colombo et al., 2014a) in the inner 6×9 kpc of the galaxy with the Plateau de Bure Interferometer Arcsecond Whirlpool Survey (PAWS; Schinnerer et al., 2013; Pety et al., 2013) dataset. NGC 5194 is a grand design spiral galaxy that is currently interacting with its smaller companion, NGC 5195.

The median age of the star clusters in NGC 5194 is 30 Myr, five times older than the median population in NGC 7793. The increase in the total number of clusters

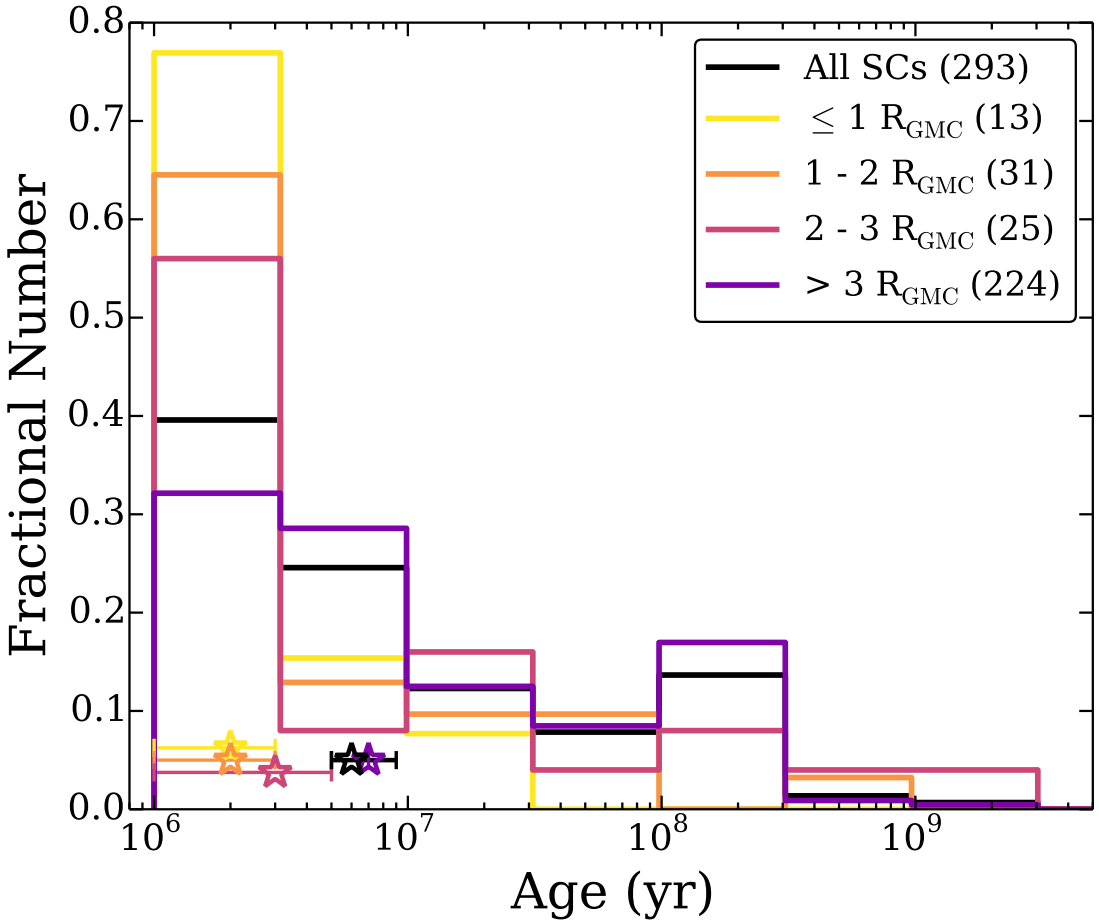

Figure 6.4. Distribution of the ages of star clusters from a GMC for NGC 7793. Displayed is the trend for the entire sample (black), star clusters located $\leq$1 $R_{\rm GMC}$ (yellow), within 1–2 $R_{\rm GMC}$ (orange), 2–3 $R_{\rm GMC}$ (pink), and star clusters unassociated with a GMC (>3 $R_{\rm GMC}$; purple). The stars show the median age of each distribution (6 ± 1 Myr, 2 ± 1 Myr, 2 ± 1 Myr, 3 ± 2 Myr, and 7 ± 1 Myr, respectively; offset in the y-direction) and the 1σ uncertainties are bootstrap estimates based on 10,000 samples. Star clusters located within a GMC are generally much younger than star clusters not within close proximity to a GMC and the age increases with increasing distance from the GMC.

by 4.7 times allows for a clearer trend and a stronger constraint on the disassociation timescale albeit in a vastly different galactic environment. In NGC 5194, the clusters located $\leq$1 $R_{\rm GMC}$ exhibit median ages of $\simeq$4 Myr, rising to $\simeq$6 Myr and $\simeq$15 Myr when located in annuli between 1–2 and 2–3 $R_{\rm GMC}$, respectively. Star clusters unassociated with any GMC (>3 $R_{\rm GMC}$) exhibit ages 10 Myr older than the global median.

The difference in the association timescale recovered between the two systems can partly arise from resolution effects: at about half the distance, the spatial resolution of NGC 7793 is twice that of NGC 5194, with a UV/optical resolution element of 16 pc/arcsec compared to 36 pc/arcsec in NGC 5194. The resolution for the GMC catalog corresponds to physical scales of $\sim$ 20 pc compared to $\sim$40 pc within NGC 5194. This allows the catalogs of NGC 7793 to be complete toward the more common lower mass clusters and GMCs down to masses of $\sim 10^4 M_\odot$ whereas the PAWS survey was limited to only the most massive GMCs $\gtrsim 2.5{\times}10^5 M_\odot$.

The difference in the association timescale between star clusters and GMCs, 2 Myr in NGC 7793 versus 4 Myr in NGC 5194, and unassociation cluster ages, 3 versus 15 Myr, can also arise partly from the inherently different cluster, GMC, and ISM properties between the two galaxies. If we reconsider the age of clusters still associated with GMCs in the two galaxies as percentages of the global median age, by the time the star clusters have traveled a minimum of $2R_{\rm GMC}$, they are half the median age of the total population in their respective systems. In addition to the association timescale that is twice as large, the velocity necessary to disassociate star clusters from their birth places is $\simeq$9.5 km s^{-1}, nearly 50% faster than what we observe in the star cluster – GMC pairs in NGC 7793 at 6.2 $\pm$ 0.9 km s^{-1}. Thus the difference in the timescales may simply be a result of a slower shear allowing the star clusters to remain associated with their host GMCs for longer periods of time.

The origin of the grand design spiral arms in NGC 5194 is the direct result of an interaction with its smaller companion, NGC 5195. For this system, the GMCs

are primarily in the arms and there is almost no age trend as a function of distance across the arm. This lack of an age trend across a spiral arm is seen observationally (Shabani et al., submitted) as well as in numerical simulations (Dobbs et al., 2017). In comparison, NGC 7793 is a flocculent spiral arm system, and thus star formation it is not as well organized as what is observed in systems with stronger and more well-defined arms. This potentially will also work to impact the differences observed in the time scales and distances between GMCs and star clusters between these two systems.

Within the SMC, Mizuno et al. (2001) found a positive correlation between young HII emission objects that trace star formation timescales of $\lesssim$6 Myr and CO clouds, where $\sim$35% of their sample of young emission objects are found to be associated with molecular clouds. Older, emissionless stellar objects (ages $\sim 6-100$ Myr) did not show a strong correlation, with only $\sim 14\%$ of this older sample showing any type of spatial association with a molecular cloud. This suggests that CO clouds quickly dissipate after the formation of their star clusters and any correlations for older clusters can be the result of chance alignment. This temporal timescale for the spatial association of young star clusters with with CO clouds is in agreement with results in the LMC with disassociation timescales of $\lesssim$ 10 Myr (Fukui et al., 1999; Yamaguchi et al., 2001). Within the LMC, Kawamura et al. (2009) finds a disassocation timescale between clusters and GMC of $\sim 7-10$ Myr and there is no significant observed difference in the size or line width between the varying evolutionary stages of GMCs. We derive timescale of $2-3$ Myr within NGC 7793 is smaller than those measured in either NGC 5194 or the LMC and SMC, but in all cases, the estimated lifetime for the GMCs is at least a factor of two greater than the estimated star cluster temporal association.

6.5.4 The Two-Point Correlation Function

We implement the angular two-point correlation function $\omega(\theta)$ to measure the magnitude of clustering as a function of projected distance between the star clusters above that of the global background. A detailed description of the formalism and methodology of the two-point correlation function as applied to star clusters within other LEGUS galaxies can be found in both Grasha et al. (2015) and Grasha et al. (2017a). The correlation function provides a method to identify the sizes of star-forming regions as well as common age structures to derive the randomization timescale for the coherence of star-forming hierarchies. Here we list the details necessary for the application to the GMCs within NGC 7793. We have previously computed the correlation function for the star clusters in NGC 7793 and will compare those results to the distribution we find for the GMCs within the ALMA coverage.

The angular correlation function $\omega(\theta)$ is the probability of finding a neighboring object within an angular separation θ above what is expected for a random Poisson distribution $\mathrm{d}P = N^2[1+\omega(\theta)]\ \mathrm{d}\Omega_1\mathrm{d}\Omega_2$, where N is the surface density of clusters per steradian with two infinitesimal elements of solid angle $\mathrm{d}\Omega_1$ and $\mathrm{d}\Omega_2$ (Peebles, 1980). A clustered distribution has an excess of pair counts at small separations, resulting in a declining power law distribution of $1 + \omega(\theta)$ toward larger length scales whereas a random distribution of an unclustered population results in a flat correlation at all length scales.

We estimate the correlation function using the Landy & Szalay (1993) estimator by counting pairs of star clusters (or GMCs) in increasing annuli and comparing those cluster counts to expectations from unclustered distributions. We supplement the GMC data with a catalog of sources that are randomly populated with the same sky coverage as the ALMA observations. The correlation function of stellar components has been demonstrated to be well-described with a power law $1 + \omega(\theta) = A_\omega\theta^\alpha$ following the convention of Calzetti et al. (1989), where the slope α measures the

strength of the clustering and the amplitude A_ω measures the correlation length of the clustering. Interstellar gas exhibits a hierarchical morphology structure with a typical fractal dimension of $D2 = \alpha + 2 \simeq 1.5$ (Elmegreen et al., 2006), allowing the projected correlation function to serve as a direct comparison of the distribution of individual GMCs and star clusters to the expected hierarchical distribution of the ISM.

Figure 6.5 shows the two-point correlation function for the GMCs and the subset of clusters located within the ALMA coverage. The random catalog follows the outline of the ALMA coverage in Figure 6.1. The star clusters exhibit a smooth and steady decline with increasing radius, well described with a power law which is nearly identical to the global star cluster distribution (see Grasha et al., 2017a), expected for a distribution that is hierarchical. We perform single power law fits in log-log space to both distributions over scale lengths $\lesssim$100 pc when $1 + \omega(\theta) > 1$. We recover a slope of -0.61 ± 0.03 for the star clusters whereas the GMCs exhibit a considerably shallower slope of -0.28 ± 0.04.

The distribution of the GMCs is significantly flatter than the star cluster distribution, a result also seen in NGC 5194 (Grasha et al., 2018a). The close association in time between GMCs and star formation suggests that the hierarchy of star clusters should be reflected by the GMCs (see, e.g., Dobbs et al., 2014), however, mirroring the exact spatial distribution would indicate that each GMC can result in the creation of only one star cluster. The observed excess in the clustering of the star clusters compared to the distribution of the GMCs in NGC 7793 and NGC 5194 strongly suggests that the natal distribution of star clusters must be more structured than that of the GMCs. This could arise from requiring a GMC to produce more than one star cluster, though the production of star clusters does not need to be simultaneous but could be sequential, and/or indicates that not all individual GMCs form a star cluster. Molecular clouds do evolve over time and exhibit different levels of star

formation activity (e.g., Yamaguchi et al., 2001; Kawamura et al., 2009; Ochsendorf et al., 2017). In particular, Kawamura et al. (2009) found that not all GMCs show evidence of star formation and more evolved clouds appear to be associated with optical stellar clusters, and thus, we may be sampling star formation regions in different evolutionary states where exposed star clusters reside in regions where the molecular gas reservoir has already disrupted.

Not all GMCs will form a star cluster; the inclusion of the entire GMC population in the calculation for Figure 6.5 may quite possibly be erasing the clustered distribution of the GMCs that are most likely forming the star clusters. To more fairly compare the distributions of the star clusters and GMCs, we limit the star clusters to those more massive than 1000 $M_{\odot}$ and match the mass limit between the star clusters and GMCs by assuming a star formation efficiency (SFE) of 1% and 3%. This allows us to exclude low mass GMCs that are most likely not actively undergoing star formation.

There is a general increase in the slope of the GMC distribution with increasing mass although we are unable to compare the distribution of the massive GMCs below scales of $\sim$40 pc due to their small numbers. The increased clustering for increasing the GMC mass implies that our optically identified star clusters arise from a specific subset of molecular clouds. We conclude that it is the most massive clouds that are more likely to produce star clusters.

Despite the similar spatial distributions between the most massive GMCs to the youngest and most massive star clusters, the clustering observed in the GMCs is still weaker than that observed for the star clusters. Star clusters originate in groupings within clustered GMCs and a stronger excess of close neighbors increases when considering the youngest clusters of the hierarchy (Grasha et al., 2015, 2017a). The different lifetimes of star clusters and GMCs may impact the dispersal timescale from the structures, and hence, the survival of the large scale hierarchies. The cluster age

distribution for NGC 7793 suggests the presence of disruption (Mora et al., 2009; Silva-Villa & Larsen, 2011). Thus the cluster dissolution timescale will have some degree of impact on the clustering results as the estimated randomization timescale for the dispersion of the hierarchies is fairly short, ∼40–60 Myr (Grasha et al., 2017a). On the other hand, the lifetimes of GMCs, upwards of tens of Myr depending on the system, are significantly shorter than the derived lifetimes of star-forming complexes. As a result, the hierarchical structure present in the spatial distribution of the GMCs disappears even quicker due to their rapid destruction.

A steeper star cluster distribution compared to the GMC distribution could be the result of the formation of a star cluster triggering the additional star formation of other clusters within a cloud (Elmegreen & Lada, 1977). Elmegreen (2018) shows that the power law in the two-point correlation function for clusters is a result from the accumulation of different generations of star formation. This would imply that the shallow slope observed for the GMC clustering can be the result of a single generation of star formation. This may be further exasperated by not being sensitive to the dense peaks of CO-dark molecular gas, making it difficult to detect the structures of the dense ISM where the vast majority of H_2 may be located and where stars are actively forming (Grenier et al., 2005; Wolfire et al., 2010; Glover & Smith, 2016).

The two-point correlation function is a global statistic and does not allow an analysis of the individual complexes, which is instead particularly suited to analyze variations in the complex clump size between a large sample of galaxies. While we have heavily implemented the use of young star clusters, other tracers of star formation such as Hα emission, HI, and CO will be particularly insightful to explore the variations in the time and size scales across galaxies with varying physical conditions.

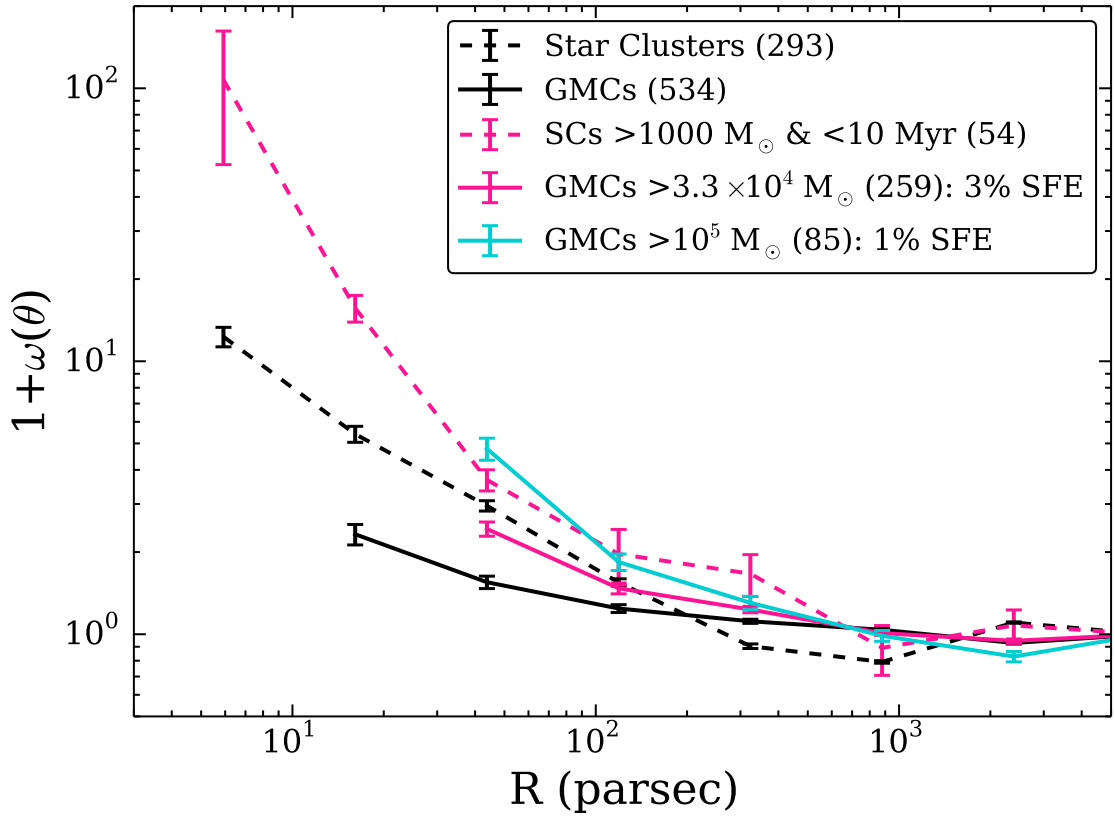

Figure 6.5. Two-point correlation function $1 + \omega(\theta)$ for NGC 7793 comparing the star clusters to the GMCs. The star clusters are shown in dotted lines and the GMCs are solid lines. The black line shows the global average for both the star clusters and the GMCs that are within the ALMA coverage (gray box of Figure 6.1). The numbers in parentheses list the number of objects. The power law relation is expected for a hierarchical distribution and is consistent with the distribution of star clusters and stars within other galaxies. The observed clustering is very flat for the GMCs given the strong age-dependency to the clustering. To compare how GMCs and star clusters respond to clustering, we show the distribution of the most massive and youngest star clusters to that of the most massive GMCs. We show the star clusters with a mass cut at 1000 $M_\odot$ (dashed pink line) and the equivalent GMC distribution with mass cuts assuming a 3% SFE (solid pink line) and a 1% SFE (solid teal line). The clustering in the distribution of the GMCs increase with mass, but at a given mass cutoff, becoming similar to the distribution of the youngest and most massive star clusters.

6.6 Summary and Conclusion

The impact of spiral structure and feedback from stellar populations on molecular clouds has broad-ranging implications for star formation in the local universe. The high-quality homogeneous star cluster catalogs from the HST LEGUS project (Calzetti et al., 2015) with reliable age measurements down to a few Myr, combined with exquisite molecular CO data (Bittle et al., 2018), are both essential in addressing the relationship of star clusters with the properties of molecular gas in a consistent and precise manner. In this work, we present a study on the connection between the star clusters in the flocculent LEGUS galaxy NGC 7793 with high-resolution molecular gas from ALMA at ∼20 pc, resolving down to individual GMCs. We compare the locations of star clusters to that of GMCs to investigate the age at which star clusters remain associated with their molecular gas, complementing a similar study we performed in the grand design spiral system NGC 5194 (Grasha et al., 2018a) to investigate the impact of spiral structure on the organization of star formation.

Our main results are summarized as follows.

1. We combine the LEGUS star cluster catalog with ∼20 pc resolution observations of CO(2–1) with ALMA of the inner ∼2×4 kpc of NGC 7793 to study over what timescales the products of star formation are still associated with molecular gas.

2. The youngest star clusters are predominately located in close vicinity to a GMC, showing an enhanced spatial correlation between the molecular clouds and young stellar clusters ($\lesssim$10 Myr) on scales $\lesssim$40 pc.

3. Star clusters that are still associated with a GMC (i.e., located within the footprint of a GMC) exhibit median ages of 2 Myr compared to the global median age of the clusters at 6 Myr. Star clusters that reside at distances $> 3R_{\rm GMC}$ to the nearest GMC are considered unassociated with any molecular cloud and exhibit median ages of 7 Myr. These age estimates approximate

the erosion of molecular gas from the young star clusters, helping to constrain their dissolution time scales in environments with recent massive star formation. Thus, in NGC 7793, star clusters disassociate from their natal clouds is short timescales of 2–3 Myr.

4. We implement the two-point correlation function to compare the hierarchical distribution of the GMCs to that of the star clusters. The star clusters are well described with a decreasing power law with increasing spatial scale, with a slope of -0.61 ± 0.03. The GMCs show a significantly flatter distribution with a power law slope of -0.28 ± 0.04 over the same scale lengths, signifying that star clusters are born more structured compared to the distribution of GMCs or are simply able to retain their structure for longer timescales. This suggests that not all GMCs form a star cluster and those that do most likely form more than one star cluster. The shorter lifetimes of GMCs compared to the stellar hierarchy, which persists for ∼40–60 Myr before dissolving, may also compound enhanced clustering present in the star cluster hierarchy.

5. The clustering observed for the star clusters and GMCs become similar to each other when comparing the most massive GMCs to the youngest and most massive star clusters. This shows the importance of matching the mass limit for star clusters and GMCs through assuming a SFE of a few percent. This allows for a better identification for the subset of GMCs that are most likely responsible for the formation of the current population of star clusters in order to compare the clustering present in the distribution of the star clusters and GMCs.

In the future, we aim to broaden this method to a larger range of galactic environments to determine the exact nature of the local environment on the disassociation timescale between star clusters and GMCs. We also seek for further improvements in deriving the disassociation timescales at different GMC evolutionary stages, ex-

panding the exposed star formation stage we currently implement with UV/optical clusters to also include the embedded star formation phase with IR-selected clusters. The exact differences in the structure or other properties of the GMCs as a function of the age of the associated star clusters is also a further avenue to explore.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Summary of this Dissertation

The focus of this dissertation has been on developing methods that improve our ability to trace the hierarchy of star formation, with a focus on using star clusters to derive the size and timescales of the unbound, large-scale hierarchies. The scientific novelty of this thesis is the use of star cluster catalogs from the LEGUS survey. These catalogs provide a well-characterized census of the star cluster population in the local universe, obtained in a high-resolution and homogeneous fashion across a diverse range of systems and environments. The LEGUS cluster catalogs increased the total number of available clusters in extragalactic galaxies by a factor of 10, allowing for the exploration of a parameter space in galaxy properties that have not been possible before.

In Chapter 2, we showed a proof of concept of using the projected two-point correlation function to characterize the observed correlations between stellar populations of NGC 628. The two-point correlation statistic performs a global average of the enhancement in clustering observed in the star clusters above that of the randomly distributed background. We find that the younger the star clusters, the more concentrated and clumped in groups their spatial distribution is, a natural expectation for star formation that proceeds hierarchically, showing fractal distributions that are consistent with observations of stellar distributions and hierarchy (Gouliermis et al., 2014, 2015a,b). This demonstrates that star clusters are indeed reliable tracers of the

scale-free nature of star formation and using correlation functions is a reliable method to constrain the global star formation patterns within a galaxy.

Performing the above analysis in a statistical and global manner, as opposed to inspecting individual clusters and star-forming complexes, is ideal for large samples of galaxies that cover a range of masses and galaxy types, and is essential in order to quantify galaxy-to-galaxy variations in a way that reduces noise levels that would otherwise make it difficult. Additionally, a statistical analysis of the *star cluster population as a whole* is one way to reduce the stochastic and highly variable estimates that come with inspecting properties of individual star clusters, such as stochastic IMF sampling and uncertainties in SED fitting from limited available bandpasses.

In Chapter 3, we applied the two-point correlation method to six galaxies in order to better understand the environmental dependence on the structure of star formation in a diverse range of galaxies. The stellar populations in these systems are well characterized with power laws, with a wide range of correlation lengths, from 200 pc up to 2000 pc. Despite the broad range observed for the correlation length of star formation, the dispersal timescales observed across all galaxies cover a much more limited range, about a factor 2 variation, from ∼40–100 Myr. Past the distance scale marked by the correlation length, the components are observed to have a random, homogeneous distribution. This indicates that there is a common physical mechanism that disperses star-forming structures across galaxies that may be independent of the exact environment and is determined by a common, underlying physical mechanism.

In Chapter 4, we constrain the physical mechanism that gives rise to the observed power law correlations between cluster ages and their separations. We measure the age difference between cluster pairs as a function of their separation, finding a power law with indices consistent with theoretical expectations from turbulence as Age $\propto \sqrt{\text{Separation}}$. In all systems, the star clusters show a transition from this expected power law distribution to that of a flat distribution. This break point is represents

the transition from 3D hierarchical motions to 2D bulk galactic dynamical motions and is interpreted as the maximum size of star-forming region. Shear of the host galaxy appears responsible for determining this maximum size over which correlated star forming regions can remain coherent although limits due to scale heights of the galaxies can still play a role.

In Chapter 5, we correlate the star clusters and the molecular gas from IRAM observations within NGC 5194 to constrain the timescale of association between star formation and the natal molecular reservoir. This represents the first study of its kind outside the Local Group to relate star clusters with high-resolution CO observations at resolutions of roughly the scale size of GMCs in an extraordinarily cluster-rich galaxy. We show that there is a dramatic difference in the age of star clusters that are still in association with a molecular cloud, showing ages of $\sim$5 Myr compared to typical ages of $\sim$50 Myr for clusters unassociated with a GMC. This galaxy uses a cluster catalog that is primarily identified with a newly developed machine learning process, compared to the traditional visual identification approach. The machine learning approach shows to be a promising avenue for future studies as it is unwieldy to classify massive samples by visual inspection.

In our concluding Chapter 6, we combine ALMA CO observations of NGC 7793 to correlate the star clusters to sub-GMC molecular gas at a resolution of $\sim$15 pc. This galaxy complements the study of Chapter 5 as NGC 7793 is a flocculent system whereas NGC 5194 is a grand design spiral galaxy, allowing for an excellent comparison on the influence of spiral dynamics on star formation hierarchies and the possible impact on the timescale of the association between star formation and molecular gas. NGC 7793 shows a slight decrease in the age of the star clusters that are still associated with their natal GMC, but more importantly, the two-point correlation function reveals a more strongly clustered distribution in the star clusters than what is observed for the GMCs. Such findings indicate that GMCs must give birth to multiple

star clusters in order to create a more clustered spatial distribution. When we match the mass limits of the GMCs and star clusters by assuming a star formation efficiency of a few percent, the spatial clustering present of the most massive and youngest star clusters start to reflect that of the most massive GMCs. This highlights the importance of comparing star clusters to the GMCs that are more likely resulting in active star formation.

The most important implications derived from this dissertation work that have added to the knowledge of the field of astrophysics are listed below.

1. The rapid dispersal of star-forming complexes as traced with star clusters corroborates a single, clustered mode of star formation, with the stars and star clusters dispersing between a few tens to one hundred Myr to form the galactic field population.

2. The relation describing the two-point correlation function remains a power law with increasing stellar population ages. This implies that the components are distributed in a scale-free manner and that the dispersal of these large-scale structures occurs via a scale-free process as well (i.e., the dispersal process is independent of scale). Numerical simulations (Elmegreen, 2018) rule out random stellar motions as the dispersal mechanism, as that would cause a dearth at the smallest scales first and erase the observed power law. A more detailed investigation into this is an area for future endeavors (Section 7.2.3).

3. Shear appears to be the mechanism determining the largest scale for the coherence of star formation. This dominates over the effect of local environment in regulating the size and survival of star-forming structures and instead reveals the importance influence of the global (large-scale) environment.

4. Star clusters show evidence of disassociating from their natal molecular material rapidly after their formation, ∼4–6 Myr. This migration timescale of a

few Myr is consistent with the clouds' crossing time with traveling velocities of 5–10 km s^{-1}.

5. Star clusters show clustered distributions that are stronger than that of GMCs, however, the spatial clustering present in the most massive GMCs is very similar to that of the youngest and most massive star clusters.

7.2 Future Work

There are several promising future research directions that build on the framework developed over the course of this dissertation. Future directions will take full advantage of the homogeneous and high-quality nature of the LEGUS datasets and catalogs as probes of the star formation process in local galaxies.

7.2.1 Environmental Influence on Cluster and Star Formation Structural Parameters

The LEGUS cluster catalogs represent a major step forward in the quantity and diversity of galaxy environments over which to investigate and characterize the stellar populations and their connection to the star formation process. One future project is to extend the studies of star clusters and their hierarchical distribution to dwarf galaxies. These systems represent a significantly different physical environments than the spiral galaxies that are primarily focused on in this work. A substantial portion of the LEGUS sample is dwarf galaxies, so a large and robust sample of clusters is attainable for such a project. This would provide an exceptional test of the environmental influence on the self-similar nature of the patterns of star formation in lower density systems.

7.2.2 How Turbulence and Dynamics Shape the Clustered Behavior

Some of the largest improvements to be made in future endeavors will be on the theoretical aspect of star formation. Current observations are unable to discriminate among different models for the underlying physics that is ultimately responsible for how star formation proceeds. Star formation theories that are regulated via supersonic turbulence (Mac Low & Klessen, 2004; Krumholz & McKee, 2005) create a cascading series of density fluctuations, where long-lasting high density fluctuations gravitationally collapse and form the observed hierarchical fragmentation and structure. On the other hand, in star formation theories that are regulated via self-gravity, the structured hierarchy can be explained due to gravity rather than turbulence (Li et al., 2005; Ballesteros-Paredes et al., 2011). In reality, it could a combination of multiple mechanisms, gravitational instability, turbulence, and stellar feedback, with different mechanisms dominating at different scales.

Stronger observational constraints can help clarify current theoretical expectations. Current observational uncertainties are substantial, limited in almost every case by small number statistics, but the observed power law for the $\Delta R - t$ relation (Chapter 4 and Figure 4.1) show indices that deviate from what is expected if arising solely from turbulence (however, the observations are consistent with the errors). This indicates that gravitational instabilities may also play a role in driving the age and size correlations between newly formed star clusters (Li et al., 2005). The correlations may also be different in different galactic environments. While the work in this dissertation only included young star clusters (i.e., UV emission) as tracers of recent star formation, future research can exploit other indicators of star formation as well, such as Hα ionized gas emission. Multi-wavelength studies with the inclusion of IR to constrain the youngest, embedded star clusters, and sub-mm and radio bands to constrain the gas reservoirs at scales similar to those of individual clusters, will help break these degeneracies to accurately model and constrain the physical mechanism

of star formation. Such data also allow for more detailed exploration of the interplay between star clusters and their immediate environmental surroundings.

7.2.3 The Physical Mechanism that Drives the Scale-free Dispersal

Another area that needs to be set on a stronger foundation is the characterization of the dispersal mechanism of these hierarchical structures. The two-point correlations of star clusters display distributions that are characterized by power laws, with a strong age-dependency to the steepness of the relation, decreasing for older cluster ages and becoming nearly flat at $\sim$100 Myr, at which point the excess in clustering has disappeared. Because the distribution remains consistent with a power law with stellar population ageing implies that the dispersal mechanism of hierarchical complexes is a scale-free process as well, regardless of if the hierarchies are 200 pc or 2 kpc.

This cannot result solely from random motions as that would preferentially remove only the smallest scales. The work in this dissertation and in other studies (e.g., Gieles et al., 2008; Bastian et al., 2009; Gouliermis et al., 2015b) instead corroborate a steady decrease in the power law with stellar population age, indicating that all scales must be erased simultaneously. Simulations show that the inherent nature of continuous star formation may give rise to the age-dependent scale-free dispersal (Elmegreen, 2018), caused by the continual buildup of hierarchical stellar structures and the overlap of distinct regions with a mixing of younger stars forming out of a different hierarchy of interstellar gas compared to that of older stellar populations. Implementing observational tests of this will move us forward in understanding the physical evolution of star-forming structures and patterns.

7.2.4 Redshift Evolution of Star Formation: Implications in the Era of JWST

With the launch of the James Webb Space Telescope (*JWST*) observatory, we will start to observe the very first galaxies in the early universe. Star clusters provide an indispensable component in regulating the star formation cycle and in governing the evolution of galaxies, two of the primary science themes of the *JWST* mission. The spectroscopic and photometric capabilities of *JWST* will allow for the characterization of the properties of young stellar systems and their dependence on their immediate surroundings with resolution and sensitivity that is unrivaled by any other current observing facility. *JWST* will be able to resolve star-forming complexes at high-redshift, determine the factors that drive the largest galactic scales on which star formation is coherent, and discover how these scales may evolve with redshift, allowing a framework to compare the physics and properties of high-redshift star-forming clumps and construct a complete picture of star formation over all of cosmic time.

7.3 Closing Remarks

Like all great scientific endeavors, sometimes the questions we set out to answer are more complex than we originally envisioned. With the answer to each question solved, a new and intriguing problem always arises to take its place. The work I have begun here is not complete and I take pride in contributing toward the greater sum of humanity's knowledge, no matter how small my niche may seem.

APPENDIX A

CLUSTER IDENTIFICATION WITHIN CLASS 0

In order to ensure that our method selecting the broad (CI $>$ 1.6) sources within the Class 0 candidates accurately identifies genuine star clusters, we visually checked the 128 Class 0 sources with CI $>$ 1.6, finding that 73 are true star clusters (57%). While this is a high contamination rate and does affect our cluster results, placing a limit for sources at CI $>$ 1.6 does improve the blind selection of clusters and allows us to identify a correlation at a much higher degree of confidence than without the selection (dotted orange versus black in Figure A.1). Additionally, small numbers limits our results of NGC 628 and we will only fully be able to investigate more significant results when we employ the galaxies within a larger selection of the LEGUS sample.

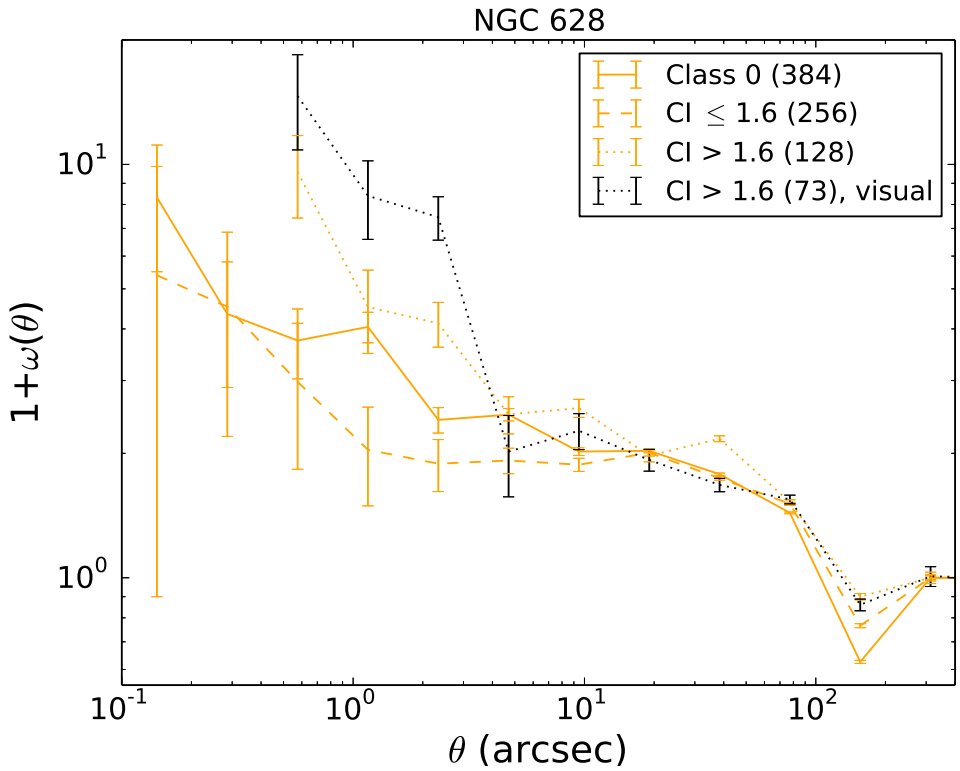

Figure A.1. Correlation function for the non-visually identified Class 0 sources, divided according to CI values. The black line shows the Class 0 sources that have been visually checked to verify that they are genuine clusters.

APPENDIX B

THE EFFECT OF DIFFERENT DUST MODELS ON THE CLUSTERING RESULTS

We use the photometric catalogs to derive the ages of our star clusters using a SED fitting algorithm, which we will touch upon below, but are available in detail in Adamo et al. (2017). The clusters in this paper for their SED fits assume a single stellar population (SSP) using two stellar libraries to create two sets of SSP models, Padova AGB and Geneva tracks without rotation, both available in Starburst99 (Leitherer et al., 1999). We assume the initial mass function (IMF) is fully sampled and adopt a Kroupa IMF, with stellar masses between 0.1 and 100 $M_{\odot}$. The models are reddened prior to SED fitting the photometry and the grid includes increasing internal reddening of E(B–V) = [0.0, 1.5] with steps of 0.01 mag. The fit incorporates three different extinction and/or attenuation laws: (1) the Milky Way extinction law from Cardelli et al. (1989); (2) the starburst attenuation law by Calzetti et al. (2000), assuming the stars and gas suffer the same reddening; and (3) the differential starburst attenuation law where we assume the gas emission suffers higher extinction than the stars.

In total, there are 12 star cluster catalogs available for each galaxy, produced with deterministic models and fitting procedures: a combination of the two photometric approaches for aperture correction (average based and CI based; see Adamo et al., 2017, for a detailed analysis and treatment on how these two approaches affect the derived properties), two stellar libraries (Geneva and Padova AGB), and three extinction/attenuation curves (Milky Way, Starburst, and the differential Starburst). For a few galaxies, there are also catalogs based on using a Bayesian analysis

method together with stochastically sampled cluster evolutionary models, presented by Krumholz et al. (2015a) using the Stochastically Lighting Up Galaxies (SLUG; da Silva et al., 2012) code. We do not incorporate the stochastic models in our results and refer the reader to Krumholz et al. (2015a) for a detailed analysis on how the deterministic approach affects the derived cluster properties compared to the stochastic approach.

Of the 12 deterministic models available, our reference cluster catalog uses the Padova stellar evolutionary models assuming solar metallicity and the starburst attenuation law. As different flavors of stellar libraries and assumptions in the dust geometry used to build cluster evolutionary tracks are bound to influence the derived cluster properties, we check here how our results depend on the assumptions made to derive the physical properties from the cluster photometry. Because our primary science goal is determining the timescale for which the hierarchical cluster structures randomize and dissipate, happening on scales of a few tens of Myr, we examine the effect that the different models will have on the derived age of our stellar clusters.

The two-point correlation functions of Figure 3.6, Figure 3.9, and Figure 3.11 are independent of derived cluster properties as they only take into account the location of each cluster within the host galaxy. The cluster properties only start to influence the results when we make bins in age, looking at the clustering as a function of both spatial scale and age, as we do in Figure 3.8 and Figure 3.12. Figure B.1 shows the age of the clusters within each galaxy, derived using our reference starburst attenuation model, as a function of age of the clusters determined using both the differential starburst and Milky Way extinction law. We also show the line that delineates where we make our age bin for clusters older/younger than 40 Myr for Figure 3.8 and Figure 3.12. Ideally, we want to avoid models where the clusters appear in the upper left hand or lower right hand part of each plot, as these clusters will move between the old/young age bin depending on which dust model is assumed. The starburst and

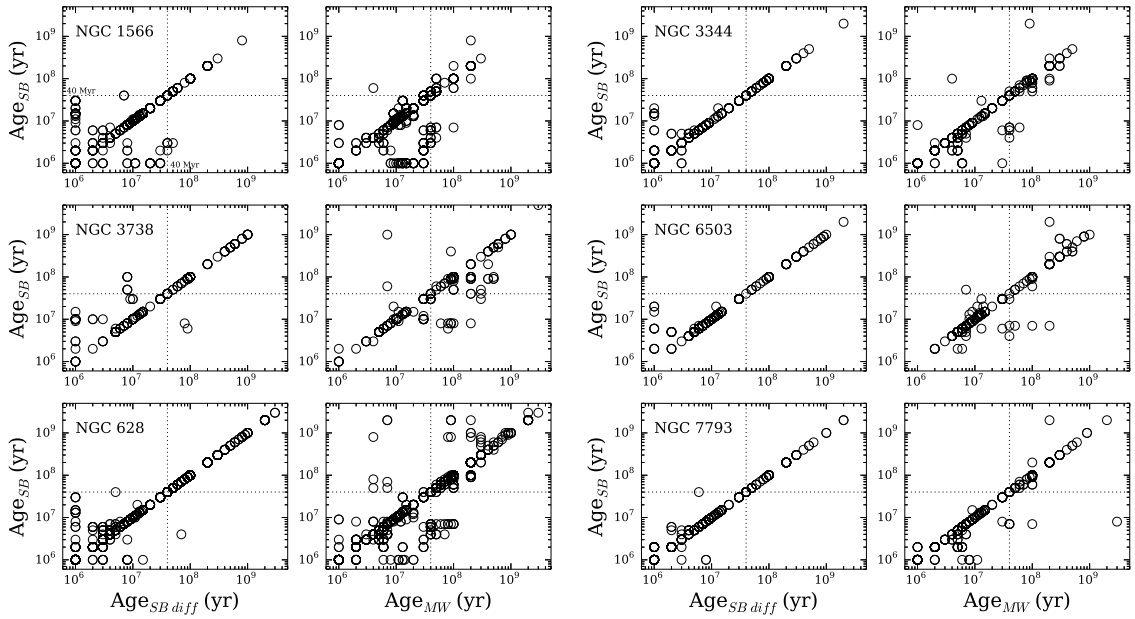

Figure B.1. Ages of the star clusters as determined with different attenuation curves. The y-axis shows the age of the star clusters, as determined using a starburst (SB) attenuation curve, as a function of the age as determined with a differential starburst attenuation curve (SB diff; left column) and a Milky Way extinction curve (MW; right column), for all six galaxies in the sample. The horizontal and dotted lines delineate an age cutoff of 40 Myr. The one-to-one correlation in the ages deviates most at the youngest ages (primarily below $\sim$ 10 Myr), which is expected, as the shortest wavelengths are where the different dust models have the greatest deviation from each another.

differential starburst laws show fairly consistent one-to-one relations in the recovered ages, where the greatest deviations are at the youngest ages, primarily confined to ages below $\sim$10 Myr. The youngest ages will have the biggest flux contribution from the shortest wavelengths (i.e., UV) and subject to the larger effects from attenuation, responsible for the increase in age spreads at younger ages. The starburst and Milky Way derived-ages deviate more relative to each other, and across a larger age range, especially within NGC 628 and NGC 3738.

To better understand the exact influence of the different models on the derived properties, we recompute the two-point correlation function analysis for NGC 628 using the Milky Way extinction curve and NGC 3738 using the differential starburst

curve attenuation curve, and compare them to the results in Figure 3.8 using the starburst model curve. A large number of sources moving between the age-separation at 40 Myr will have a substantial effect on the clustering results.

Figure B.2 shows the impact that different ages have on the two-point correlation function for NGC 3738 when calculated using the starburst attenuation curve (Figure 3.6; our reference model), compared to the differential starburst curve (as NGC 3738 has about a quarter of the metallicity of the Milky Way galaxy, we do not consider the Milky Way extinction curve). The differential starburst law gives a total of four clusters with ages younger than those recovered with the starburst attenuation curve. The class 1 clusters show the greatest difference in the results, a result of class 1 having the least number of total clusters compared to the total number of class 2 and 3 clusters.

Figure B.3 shows the two-point correlation function for NGC 628 estimated using the Milky Way extinction curve compared to our reference starburst model. There are a total of nine clusters that jump between the age bin of lower/higher than 40 Myr when derived with the Milky Way curve. While there is a large spread in the ages (seen in Figure B.1 for the Milky Way versus starburst curve, the large number of clusters means the results are fairly insensitive to any particular model. The only difference is a minor increase in the slope at the lowest spatial bin for the youngest clusters in class 1.

While the derived ages determined with different models from the SED fits do influence the overall properties of the star clusters, the adopted dust curve does not make a statistical impact on our results nor does it prohibit the comparison of results if the data set is analyzed using two different sets of models.

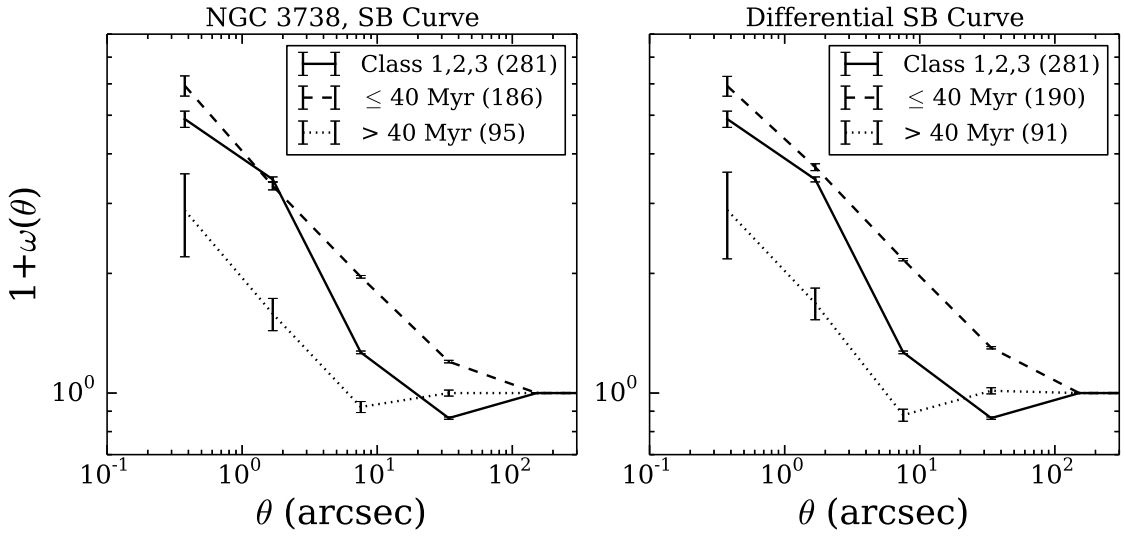

Figure B.2. A demonstration on how the different ages impact the two-point correlation function for NGC 3738. We show the starburst attenuation law (left; same as Figure 3.8), which is our reference, compared to the differential starburst attenuation law (right) for the three classifications of clusters.

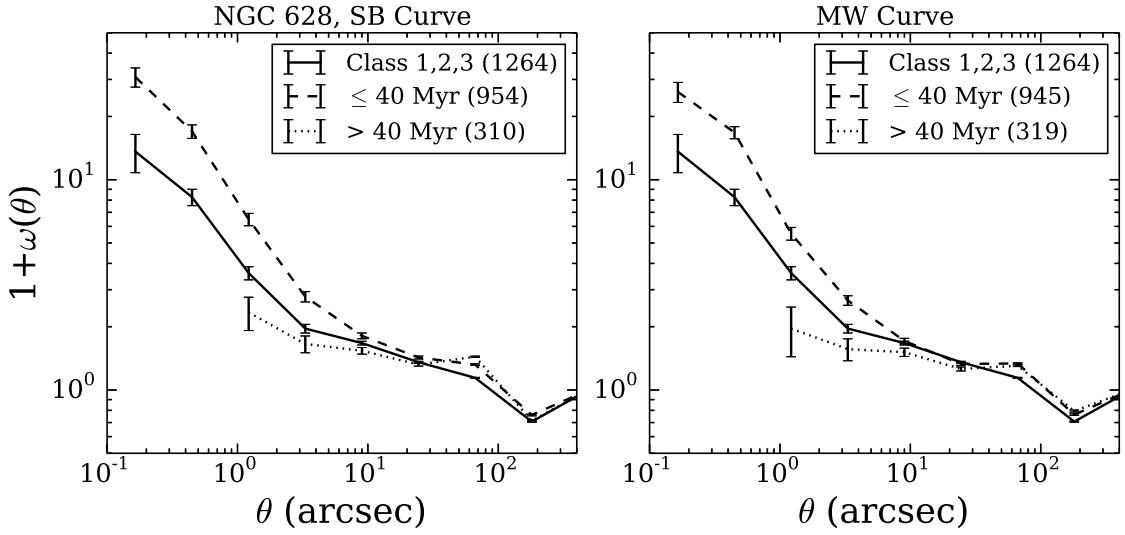

Figure B.3. A demonstration on how the different ages impact the two-point correlation function for NGC 628. The reference starburst attenuation curve (left, same as Figure 3.8) is compared against a Milky Way extinction curve (right) for all the classifications of clusters. The correlation function for the clusters whose properties are determined with a starburst attenuation law versus a Milky Way extinction law does not change.

APPENDIX C

THE REQUIREMENT OF VISUAL IDENTIFICATION TO DISTINGUISH INDIVIDUAL STARS FROM STAR CLUSTERS

C.1 Color Alone is not Sufficient to Differentiate between Stars and Star Clusters

One of the primary challenges in observations of unresolved star clusters is the difficulty to distinguish faint, young star clusters from those of individual stars (Figure C.1). The fact that star clusters are composed of stars, the light in a single cluster may be dominated by a single red or yellow supergiant, especially in low-mass clusters (Silva-Villa & Larsen, 2010), introducing scatter into the typical color space occupied by star clusters (Whitmore et al., 2010). Discreteness in sampling of the IMF can also strongly affect integrated properties (e.g., colors, magnitudes, ages, masses Cerviño & Luridiana, 2006; Maíz Apellániz, 2009) and may contribute to the observed color scatter, especially within unresolved, low-mass (10^4 $M_\odot$) systems (Silva-Villa & Larsen, 2011). Photometric uncertainties also make it difficult to distinguish stars from star clusters for systems that are fainter than $M_v \sim -7$ (Whitmore et al., 1999, 2010). In addition to confusion with individual stars, background galaxies also act as interlopers in cluster color space and these objects are only removable from automated catalogs via visual inspection (Larsen, 2004; Annibali et al., 2011).

It is expected though that most star clusters, even globular clusters, will exhibit colors bluer than $V - I$ of 1.5 (Forbes et al., 1997; Larsen, 2001; Santiago, 2008) although reddening due to dust can act to push clusters past this limit. Even if such

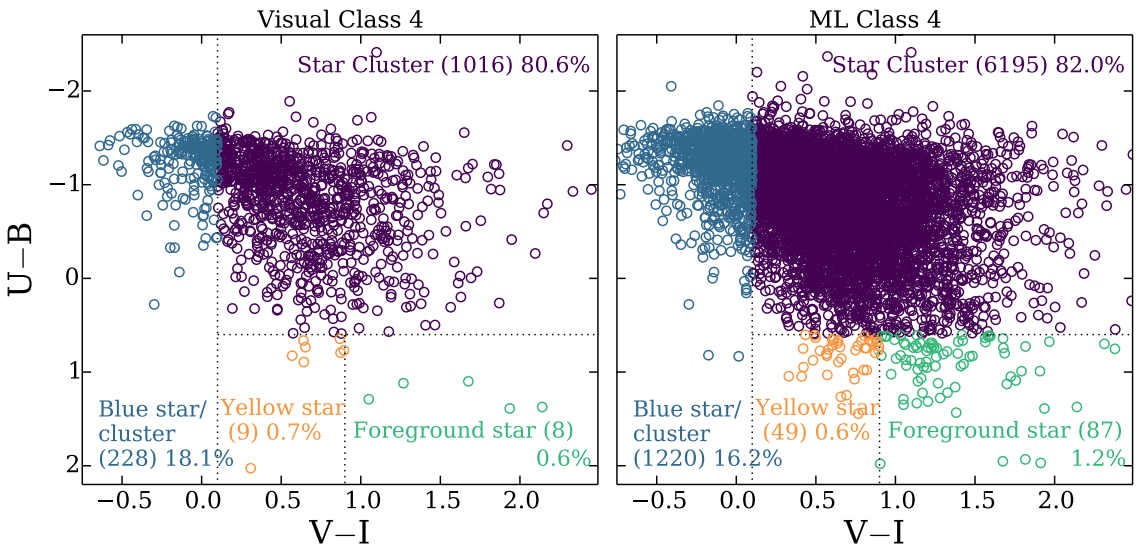

Figure C.1. U–B versus V–I colors for the class 4 cluster contaminants. The left shows the clusters from the visual identification and the ML-only clusters are on the right. The regions are the same as shown in Figure 5.4 from Whitmore et al. (2010). The numbers in parenthesis list the total number of sources within each region and also show the fraction of each subregion compared to the total sample in each panel. We keep the axis range the same as Figure 5.4 to ease comparison with the actual star cluster sample. Despite the vast majority of the non-cluster contaminates falling in the star cluster color-color subspace, the sources with visual identification are not clusters. The sources with ML-only identification show the same fraction within each subregion. It is possible that in some clusters, the light from a single supergiant star will dominate the light profile, creating confusion in the observed colors. The high rate of contaminants showing the same color as star clusters is why color alone is insufficient at distinguishing between star clusters and stars (or other contaminants, such as background galaxies), necessitating the need for visual identification.

a cut in color space is performed on our catalog, this will only exclude ∼2% of the objects from the total catalog, revealing the difficult in identifying star clusters via color alone.

C.2 Size Alone is not Sufficient to Differentiate between Stars and Star Clusters

The creation of any star cluster catalog almost always undergoes an initial CI cut to help in culling individual stars from the automated catalog (e.g., Bastian et al., 2005; Scheepmaker et al., 2007; Mora et al., 2009; Silva-Villa & Larsen, 2010; Whitmore et al., 2010; Bastian et al., 2012; Chandar et al., 2014; Adamo et al., 2017). It is still common, however, to have foreground stars or background galaxies that can be larger than the CI cutoff, requiring visual inspection to remove these from the catalog.

In Grasha et al. (2015), we sought to identify cuts in CI values to reliably extract star clusters while at the same time excluding contaminants without the need for visual identification in NGC 628 (distance of 10 Mpc). We find that there is too much overlap in the sizes of clusters with contaminants, most commonly foreground stars, unassociated stellar asterisms, and background galaxies. The overlap present in the sizes between clusters and contaminants is why size alone is inadequate as a parameter to identify clusters is demonstrated in Figure C.2).

Due to the vast diversity of star clusters, there is no unique combination of criteria that will lead to the detection of all bona-fide clusters in a galaxy while at the same time not introducing false detections. The inability to easily separation stars from star clusters based off their color or size, the two most obvious parameters to investigate, necessitates the additional process of visual inspection on top of an automated procedure. Even with sufficient resolution to identify individual stars in each star cluster, the cluster catalogs of M 31 still underwent the visual identification process for their selection (Johnson et al., 2012, 2015). The visual inspection of cluster candidates does introduce a level of subjectivity in the final product, and as a result, the clusters catalogs of the same galaxy can vary from different authors (see, e.g., Adamo & Bastian, 2015, for a summary). Typically, the vast majority of the dif-

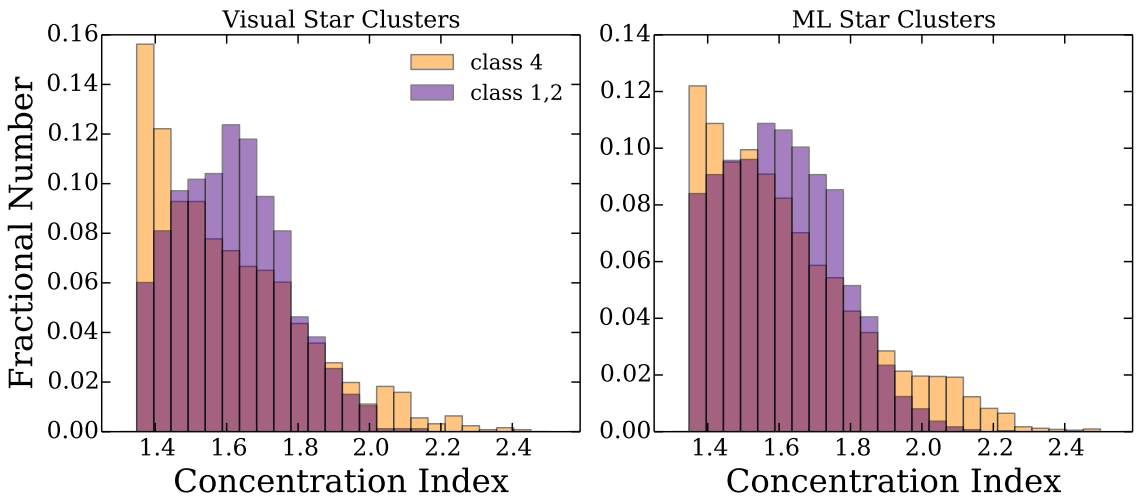

Figure C.2. Fractional histogram of the concentration index of the star cluster catalogs. Both class 1+2 star clusters (purple) and class 4 non-cluster contaminants (orange) for sources with visual identification (left) and ML identification (right) are shown. While CI cuts are a reliable way to separate individual stars from star clusters at very local distances, many contaminants overlap and show similar size distributions. Therefore, size alone is not sufficient in order to to differentiate between stars and clusters. Many clusters are compact enough such that they remain indistinguishable from stars even at the resolution provided by HST imaging.

ferences that arise between different catalogs are mostly due to source classification and the morphological definition of a cluster. This is why the LEGUS classification delineates between clusters based on their morphology in order to ease comparison with previous studies and as well as future studies.

APPENDIX D

ACCURACY IN THE PERFORMANCE OF THE ML ALGORITHM

Tree bagging (Bootstrap Aggregation) is a simple 'ensemble learning' ML algorithm, ideal for when a goal is to reduce the variance of a decision tree. Several subsets of data from training samples are chosen randomly with replacement and each collection of subset data is used to train their decision trees. This method creates an ensemble of different models and the average of all the predictions from different trees are used is more robust than a single decision tree.

Typically 50–150 decision trees are implemented before assigning a class to a candidate where the classifier error decreases with the number of grown trees. We implement 100 trees in our method, a compromise between computational time and algorithm performance (see Figure D.1).

The only hyperparameter we vary is the number of decision trees in the TreeBagger function in Matlab. We use the default parameters for the rest of the hyperparameter inputs: minimum number of observations per tree leaf is 1, sample with replacement, Gini's diversity index for the split decision, and always implement a 10-fold cross-validation. By default, TreeBagger bags classification trees.

The performance of a classifier can be visualized and analyzed with the receiver operating characteristic (ROC) curve which compares the fraction of true positives versus that of true negatives. The closer the points in a given curve to the top left corner, the better the performance of the classifier and signifies a better predictive performance. The optimal point is the value for which the point on the ROC curve

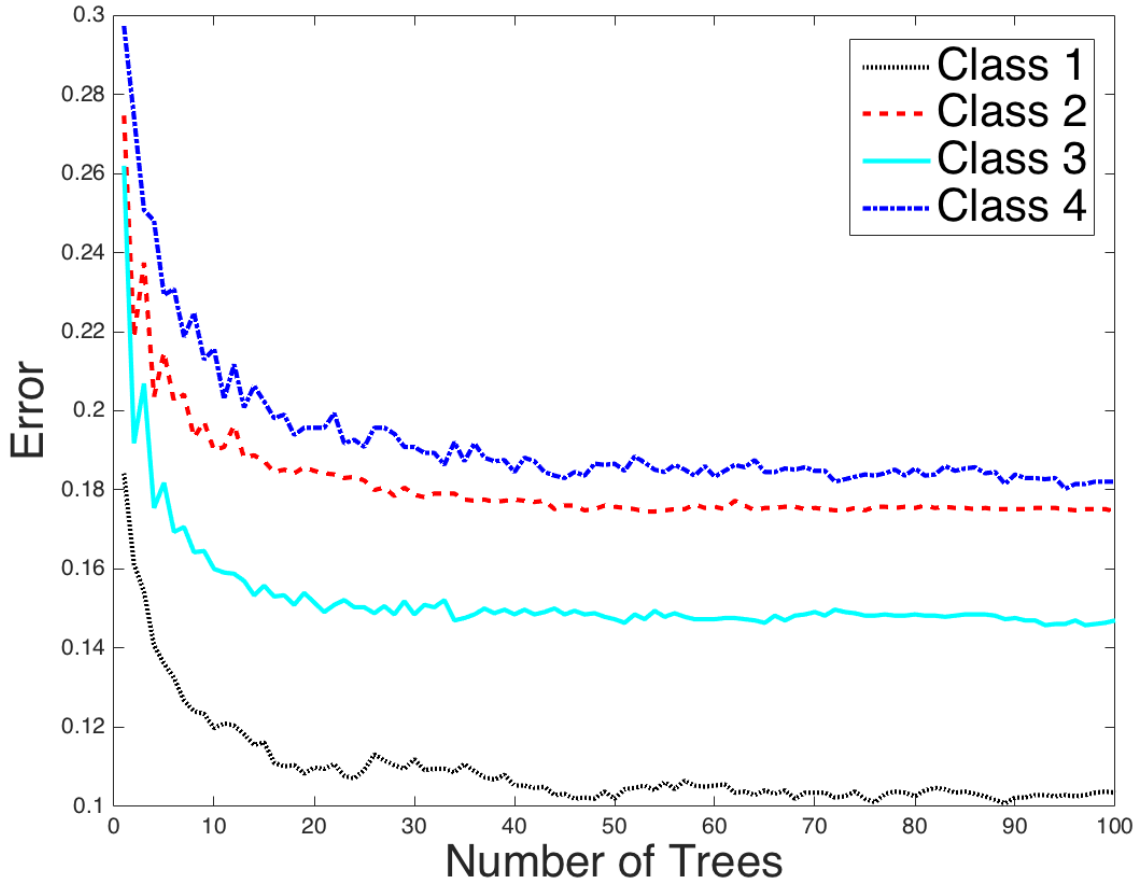

Figure D.1. Cross-validated fractional error in the performance of the bagging classifier as a function of the number of bagged trees. Colors show different cluster classifications.

has the minimum distance to the top left corner. Figure D.2 shows the ROC curves for each class and the ROC curves predict each class versus the rest (e.g., class 1 vs class 2, 3, and 4). The error and 99% confidence thresholds for each cluster classification in the final catalog are 0.103 ± 0.016 for class 1, 0.175 ± 0.017 for class 2, 0.147 ± 0.018 for class 3, and 0.182 ± 0.011 for class 4.

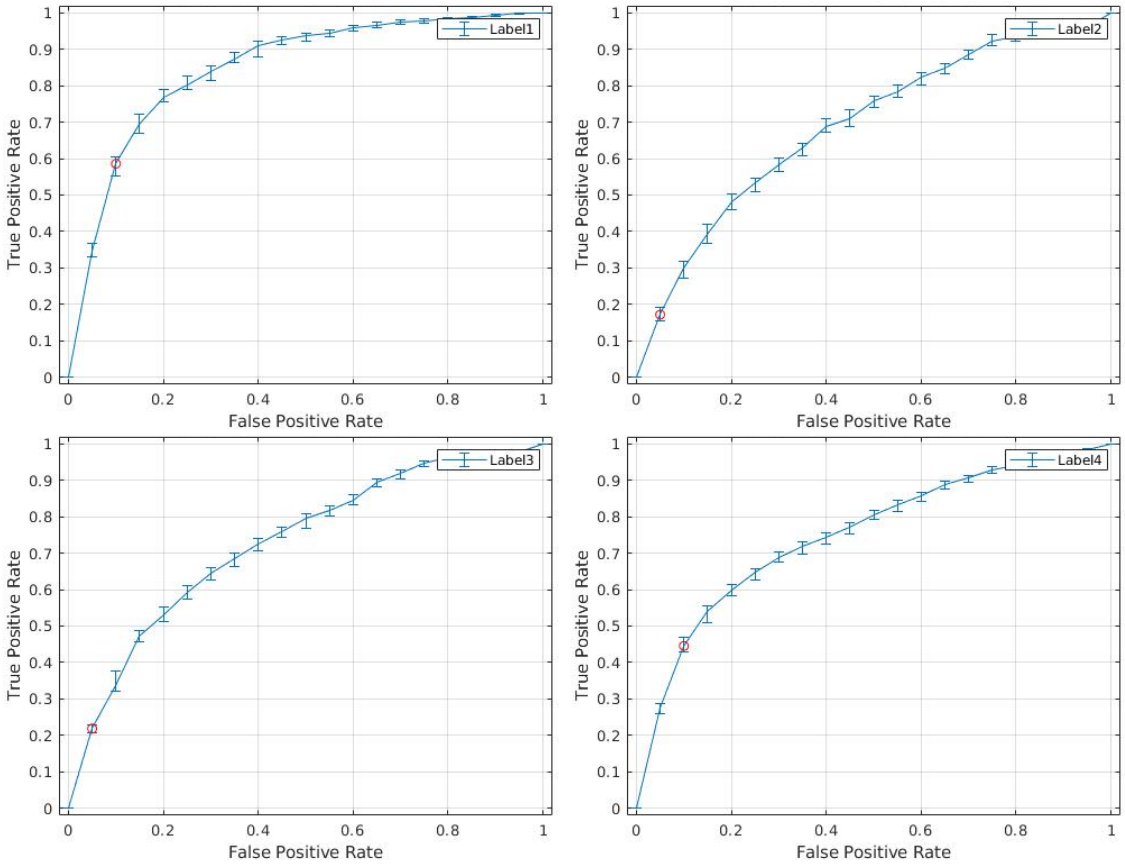

Figure D.2. ROC curves of the four cluster classifications. Shown is the true positive rate as a function of the false positive rate. The performance for class 1 clusters can easily be seen as the best out of all the classifications. The ML performance for class 1 has an accuracy of .897, class 2 accuracy of 0.825, class 3 accuracy of 0.853, and class 4 accuracy of 0.818. The red circle on each curve is the optimal ROC point.

APPENDIX E

MASS AND AGE FUNCTIONS OF VISUAL VERSUS ML STAR CLUSTERS

As the luminosity is an observed quantity, it can be quantified without any assumption of stellar models and or SED fits. The luminosity function of young star clusters is generally described by a power law function $dN/dL \propto L^{-\alpha}$, with an almost universal index of $\alpha \sim 2$ as observed in local spiral galaxies (e.g., Larsen, 2002; de Grijs et al., 2003). Using the same method of Messa et al. (2018), we have computed the luminosity function of the visual star cluster catalog and the ML star cluster catalog on the V-band filter, finding a power law fit with slope 1.96 $\pm$ 0.04 for the visual cluster catalog and 1.92 $\pm$ 0.03 for the ML cluster catalog. These are both consistent within the errors and demonstrates that the ML process does not penalize or mis-classify clusters with specific luminosities.

In order to explore the properties derived from SED fits to ensure we are not inaccurately classifying and introducing biases in the ML catalog, we calculate the mass function using the mass-complete sample (clusters with masses above 5000 $M_\odot$ and ages less than 200 Myr). The fit performed both with a single power law and truncated mass function which serves to test the effect of random sampling from the mass function as such an effect can produce a truncation effect (see da Silva et al., 2012). For the single power law for the mass function, we recover slopes of -2.12 ± 0.05 for the visual cluster catalog and -2.18 ± 0.02 for the ML cluster catalog. These slopes are consistent with the canonical -2 power law commonly seen in star cluster studies (e.g., Battinelli et al., 1994; de Grijs et al., 2003)

When fitted with a Schechter function with a truncation mass, we recover a slope of -1.85 ± 0.05 with a truncation mass at $1.42 \pm 0.21 \times 10^5$ $M_\odot$ for the visual catalog. The ML catalog is best fit with a slope of -1.88 ± 0.03 and a truncation mass of $1.39 \pm 0.14 \times 10^5$ $M_\odot$. We show the mass functions from both methods in Figure E.1 along with simulated mass functions. The simulation mass functions are obtained via a bootstrapping technique from 1000 Monte Carlo trials as described inMessa et al. (2018). We show the median expectation, the 50 %, and the 90 % limits of the simulated function results.

As the mass functions of ML and visual classified populations are consistent with each other within the uncertainties, we conclude the ML procedure is not introducing any bias in the mass parameter space of the clusters.

Our last check to ensure we are not misclassifying or missing certain populations of the clusters with the ML procedure is to compare the age functions of the catalogs. We use the same method as implemented in Messa et al. (2018) by dividing the sample in age bins of 0.5 and normalizing the number of sources in each bin by the age range spanned by each bin (Figure E.2). The points are fitted with a simple power law $dN/dt \propto t^{-\alpha}$ up to log(age/yr) = 8.5, where incompleteness starts to affect the results. We exclude sources in the fit of the age function that are younger than 10 Myr as they potentially contaminate our sample due to the rapid decline in the number of clusters surviving past ∼10 Myr (see Lada & Lada, 2003).

We recover a single power law fit to the age function for both classification method, with a slope of -0.35 ± 0.17 for the visual catalog and -0.44 ± 0.05 for the ML catalog. For both methods, the differences are within the uncertainties.

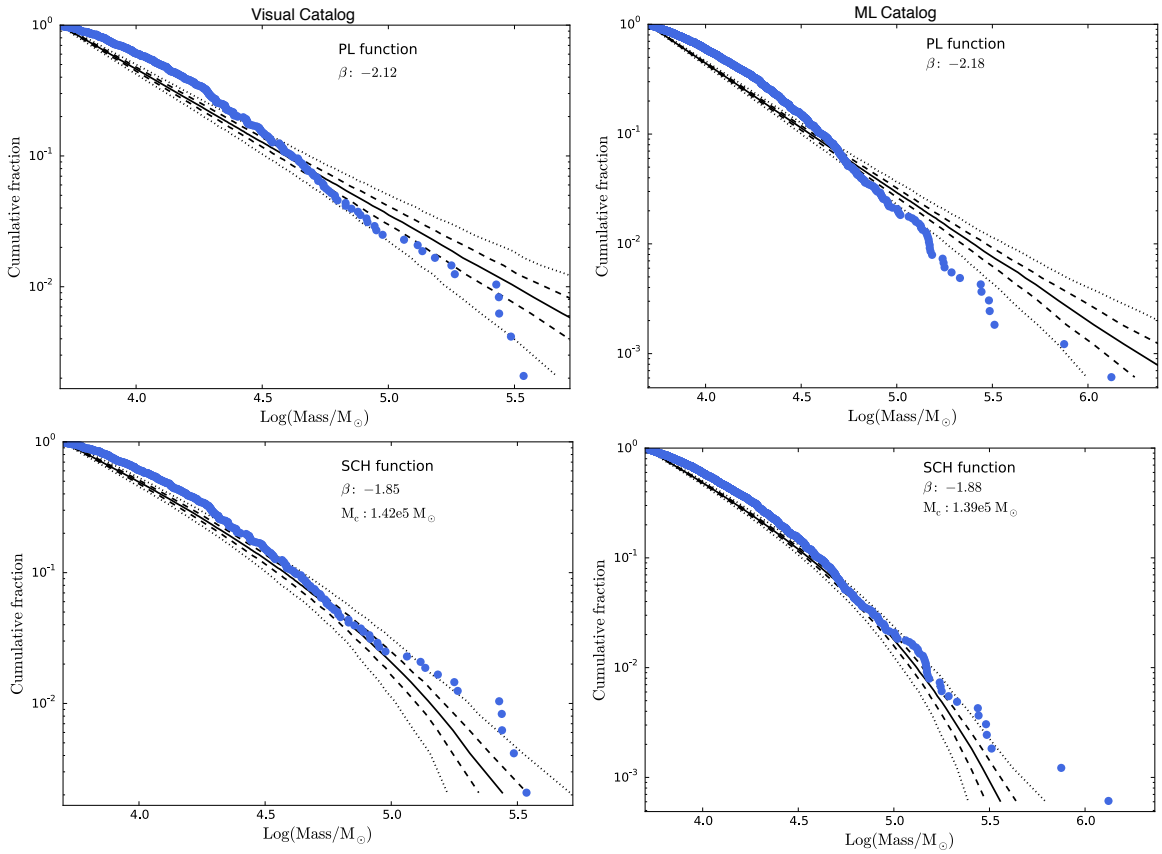

Figure E.1. The mass functions for the star clusters in NGC 5194. Show are star clusters (blue) in the visual catalog (left column) and the ML catalog (right column). Monte Carlo simulations are performed 1000 times and are shown in the black lines. The median mass distributions (solid lines) and the limits within 50 percent (dashed) and 90 percent (dotted) of the simulations are plotted. The single power law fits and simulations are shown on the top row and the Schechter mass function are on the bottom. The differences observed between the two catalogs are consistent within the uncertainties.

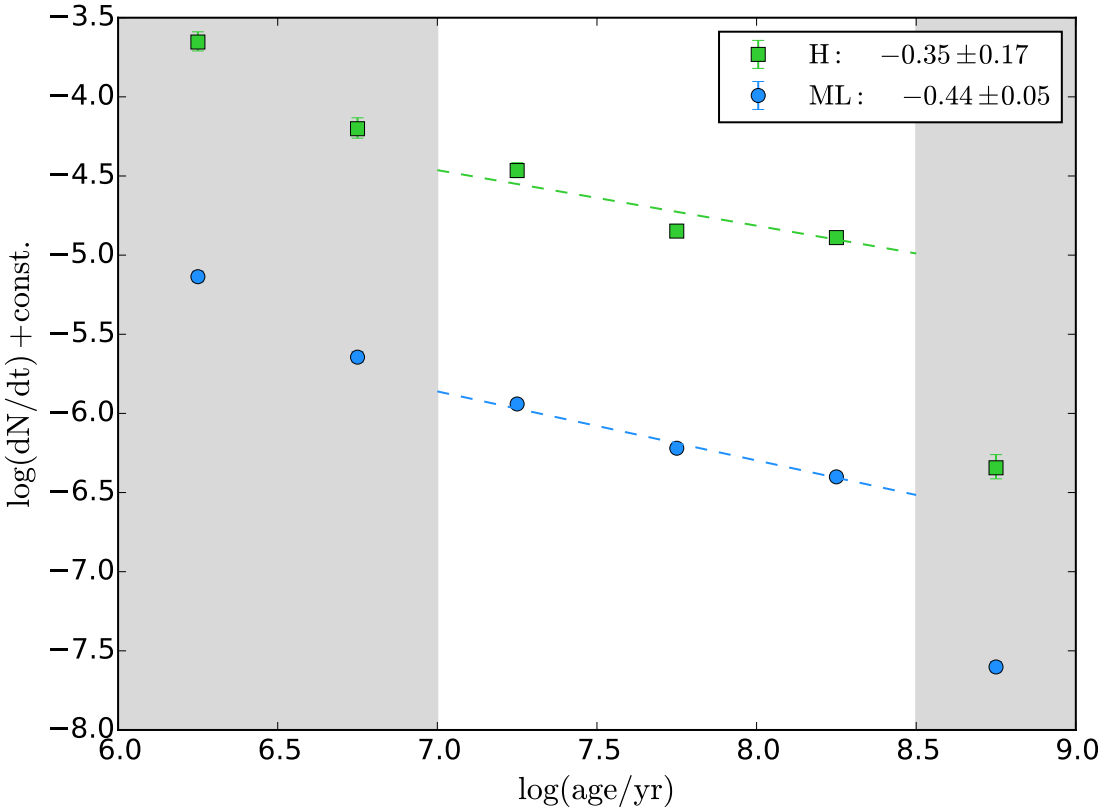

Figure E.2. Age function of the star clusters in NGC 5194. Shown are the visually identified star clusters (H: green) and the ML star cluster (blue). Dashed lines represent the best-fitting curve for the bins in the range log(age) = 7–8.5 yr. The grey-shaded areas mark the part of the functions excluded from the analysis due to incompleteness (old ages) and possible contamination by unbound sources (young ages). The differences observed between the two catalogs are within the uncertainties.

BIBLIOGRAPHY

Adamo, A., Östlin, G., Zackrisson, E., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 407, 870

Adamo, A., Östlin, G., Zackrisson, E. 2011, *Monthly Notices of the Royal Astronomical Society*, 417, 1904

Adamo, A., Smith, L.J., Gallagher, J.S., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 1185

Adamo, A., Kruijssen, J.M.D., Bastian, N., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 246

Adamo, A. & Bastian, N. 2015, arXiv:1511.08212

Adamo, A., Ryon, J.E., Messa, M., et al. 2017, *The Astrophysical Journal*, 841, 131

Agüero, E.L., Díaz, R.J., & Bajaja, E. 2004, *Astronomy & Astrophysics*, 414, 453

Anders, P., Bissantz, N., Fritze-v. Alvensleben, U., de Grijs, R. 2004, *Monthly Notices of the Royal Astronomical Society*, 347, 196

Annibali, F., Tosi, M., Aloisi, A., & van der Marel, R.P. 2011, *Astronomical Journal*, 142, 129

Ashworth, G., Fumagalli, M., Krumholz, M.R., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 469, 2464

Ball, N.M., Loveday, J., Fukugita, M., et al. 2004 *Monthly Notices of the Royal Astronomical Society*, 348, 1038

Ballesteros-Paredes, J., Hartmann, L W., Vázquez-Semadeni, E., Heitsch, F., & Zamora-Avilés, M.A. 2011, *Monthly Notices of the Royal Astronomical Society*, 411, 65

Banerji, M., Lahav, O., Lintott, C.J., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 406, 342

Bastian, N., Gieles, M., Lamers, H.J.G.L.M., Scheepmaker, R.A., & de Grijs, R. 2005, *Astronomy & Astrophysics*, 431, 905

Bastian, N. & Goodwin, S.P. 2006, *Monthly Notices of the Royal Astronomical Society*, 369, L9

Bastian, N., Ercolano, B., Gieles, M., et al. 2007, *Monthly Notices of the Royal Astronomical Society*, 379, 1302

Bastian, N., Gieles, M., Ercolano, B., & Gutermuth, R. 2009, *Monthly Notices of the Royal Astronomical Society*, 392, 868

Bastian, N., Covey, K.R., & Meyer, M.R. 2012, *Annual Review of Astronomy and Astrophysics*, 48, 339

Bastian, N., Adamo, A., Gieles, M., et al. 2011a, *Monthly Notices of the Royal Astronomical Society*, 417, 6

Bastian, N., Weisz, D.R., Skillman, E.D., et al. 2011b, *Monthly Notices of the Royal Astronomical Society*, 412, 1539

Bastian, N., Adamo, A., Gieles, M., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 419, 2606

Bate, M.R., Clarke, C.J., & McCaughrean, M.J. 1998, *Monthly Notices of the Royal Astronomical Society*, 297, 1163

Battersby, C., Bally, J., Svoboda, B. 2017, *The Astrophysical Journal*, 835, 263

Battinelli, P., Brandimarti, A., & Capuzzo-Dolcetta, R. 1994, *Astronomy & Astrophysics Supplement*, 104, 379

Baumgardt, H., Parmentier, G., Anders, P., & Grebel, E.K., 2013, *Monthly Notices of the Royal Astronomical Society*, 430, 676

Beaumont, C.N., Goodman, A.A., Kendrew, S., Williams, J.P., & Simpson, R. 2014, *The Astrophysical Journal Supplement Series*, 214, 3

Beech, M. 1987, *Astrophysics and Space Science*, 133, 193

Bertin, E. & Arnouts, S. 1996, *Astronomy & Astrophysics Supplement*, 117, 393

Bhatia, R.K. & Hadzidimitriou, D. 1988, *Monthly Notices of the Royal Astronomical Society*, 230, 215

Bhatia, R.K. 1990, *Publications of the Astronomical Society of Japan*, 42, 757

Bigiel, F., Leroy, Adam K., Jiménez-Donaire, Maria J., et al. 2016, *The Astrophysical Journal*, 822, L26

Bittle, L., et al. 2018, in prep

Blaauw, A. 1952, *Bulletin of the Astronomical Institutes of the Netherlands*, 11, 414

Block, D.L., Puerari, I., Elmegreen, B.G., & Bournaud, F. 2010, *The Astrophysical Journal Letters*, 718, L1

Bonatto, C. & Bica, E. 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 996

Bothwell, M.S., Kennicutt, R.C., & Lee, J.C. 2009, *Monthly Notices of the Royal Astronomical Society*, 400, 154

Bournaud, F., Perret, V., Renaud, F., et al. 2014, *The Astrophysical Journal*, 780, 57

Boyd, R.N. 2008, "An introduction to nuclear astrophysics". University of Chicago Press. p. 376. ISBN 0-226-06971-0.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. 1984, Classification and Regression Trees. Belmont, CA: Wadsworth International Group

Breiman, L. 2001, *Machine Learning*, 45, 5

Bressert, E., Bastian, N., Gutermuth, R., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 409, L54

Brodwin, M., Dey, A., Brown, M.J.I., et al. 2008, *The Astrophysical Journal*, 687, L65

Bruzual, A.G. 1983, *The Astrophysical Journal*, 273, 105

Calzetti, D., Giavalisco, M., & Ruffini, R. 1989, *Astronomy & Astrophysics*, 226, 1

Calzetti, D., Armus, L., Bohlin, R.C., et al. 2000, *The Astrophysical Journal*, 533, 682

Calzetti, D., Chandar, R., Lee, J.C., et al. 2010, *The Astrophysical Journal*, 719, L158

Calzetti, D., Lee, J.C., Sabbi, E., et al. 2015, *Astronomical Journal*, 149, 51

Calzetti, D., Johnson, K.E., Adamo, A., et al. 2015b, *The Astrophysical Journal*, 811, 75

Cardelli, J.A., Clayton, G.C., & Mathis, J.S. 1989, *The Astrophysical Journal*, 345, 245

Carignan, C. & Puche, D. 1990, *Astronomical Journal*, 100, 394

Carlberg, R.G. & Pudritz, R.E. 1990, *Monthly Notices of the Royal Astronomical Society*, 247, 353

Cerviño, M. & Luridiana, V. 2004, *Astronomy & Astrophysics*, 413, 145

Cerviño, M. & Luridiana, V. 2006, *Astronomy & Astrophysics*, 451, 475

Chandar, R., Fall, S.M., & Whitmore, B.C. 2010, *The Astrophysical Journal*, 711, 1263

Chandar, R., Whitmore, B.C., Calzetti, D., & OConnell, R. 2014, *The Astrophysical Journal*, 787, 17

Chandar, R., Fall, S.M., & Whitmore, B.C. 2015, *The Astrophysical Journal*, 810, 1

Chandar, R., Whitmore, B.C., Dinino, D., et al. 2016, *The Astrophysical Journal*, 824, 71

Chandar, R., Chien, L.-H., Meidt, S., et al. 2017, *The Astrophysical Journal*, 845, 78

Chen, H., Gao, Y., Braine, J., & Gu, Q. 2015, *The Astrophysical Journal*, 810, 140

Colombo, D., Hughes, A., Schinnerer, E., et al. 2014a, *The Astrophysical Journal*, 784, 3

Colombo, D., Meidt, S.E., Schinnerer, E., et al. 2014b, *The Astrophysical Journal*, 784, 4

Combes, F. & Becquaert, J.-F. 1997, *Astronomy & Astrophysics*326, 554

Combes, F., Garcia-Burillo, S., Casasola, V., et al. 2014, A&A, 565, 97

Conselice, C.J., Grogin, N.A., Jogee, S., et al. 2004, *The Astrophysical Journal*, 600, L139

Cook, D.O., Dale, D.A., Lee, J.C., Thilker, D., Calzetti, D., & Kennicutt, R.C. 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 3766

Corbelli, E., Braine, J., Bandiera, R., et al. 2017, *Astronomy & Astrophysics*, 601, 146

Crocker, A.F., Rupali, C, Calzetti, D., et al. 2015, *The Astrophysical Journal*, 808, 76

da Silva, R.L., Fumagalli, M., & Krumholz, M. 2012, *The Astrophysical Journal*, 745, 145

Daddi, E., Cimatti, A., Pozzetti, L., Hoekstra, H., Rottgering, H.J.A., Renzini, A., Zamorani, G., & Mannucci, F. 2000, *Astronomy & Astrophysics*, 361, 535

Daigle, O., Carignan, C., Amram, P., et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 367, 469

Dalcanton, J.J. & Stilp, A.M. 2010, *The Astrophysical Journal*, 721, 547

Dalcanton, J. J., et al. 2012, *The Astrophysical Journal Supplement Series*, 200, 18

Dale, J.E., Ercolano, B., & Bonnell, I.A. 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 987

Davidge, T.J., Puzia, T.H., Mcconnachie, A.W. 2011, *The Astrophysical Journal Letters*, 728, L23

de Grijs R., Anders P., Bastian N., Lynds R., Lamers H.J.G.L.M., & ONeil E.J. 2003, *Monthly Notices of the Royal Astronomical Society*, 343, 1285

de la Fuente Marcos, R. & de la Fuente Marcos, C. 2009, *The Astrophysical Journal*, 700, 436

de la Fuente Marcos, R. & de la Fuente Marcos, C. 2010, *The Astrophysical Journal*, 719, 104

De Silva, G.M., Carraro, G., D'Orazi, V., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 453, 106

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H.G., Jr., et al. 1991, Third Reference Catalogue of Bright Galaxies (ver 3.9; New York: Springer)

de Vega, H.J., Sánchez, N., & Combes, F. 1996, *Nature*, 383, 56

Dekel, A., Sari, R., & Ceverino, D. 2009, *The Astrophysical Journal*, 703, 785

Dicaire, I., Carignan, C., Amram, P., et al. 2008, *Astronomical Journal*, 135, 2038

Dieball, A., Müllerm, H., Grebel, E.K., 2002, *Astronomy & Astrophysics*, 391, 547

Dieleman, S., Willett, K.W., & Dambre, J. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1441

Di Matteo, P., Bournaud, F., Martig, M., et al. 2008, *Astronomy & Astrophysics*, 492, 31

Dobbs, C.L. & Pringle, J.E. 2013, *Monthly Notices of the Royal Astronomical Society*, 432, 653

Dobbs, C.L., Pringle, J.E., & Naylor, T. 2014, *Monthly Notices of the Royal Astronomical Society*, 437, L31

Dobbs, C.L., Adamo, A., Few, C., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 3580

Draine, B.T. 2003, *Annual Review of Astronomy and Astrophysics*, 41, 241

Dutta, P., Begum, A., Bharadwaj, S., Chengalur, J.N., 2013, *New Astronomy*, 19, 89

Elemgreen, B.G. & Lada, C.J. 1977, *The Astrophysical Journal*, 214, 725

Efremov, Y.N. 1978, *Soviet Astronomy Letters*, 4, 66

Efremov, Y.N. 1995, *Astronomical Journal*, 110, 2757

Elmegreen, B.G. 1989, *The Astrophysical Journal*, 338, 178

Efremov, Y.N. & Elmegreen, B.G. 1998, *Monthly Notices of the Royal Astronomical Society*, 299, 588

Elmegreen, B.G. & Elmegreen, D.M. 1983, *Monthly Notices of the Royal Astronomical Society*, 203, 31

Elmegreen, B.G & Falgarone, E. 1996, *The Astrophysical Journal*, 471, 816

Elmegreen, B.G. & Efremov, Y.N. 1996, *The Astrophysical Journal*, 466, 802

Elmegreen, B.G., Elmegree, D.M., Salzer, J.J., Mann, H. 1996, *The Astrophysical Journal*, 467, 579

Elmegreen, B.G. & Efremov, Y.N. 1997, *The Astrophysical Journal*, 480, 235

Elmegreen, B.G. 1999, *The Astrophysical Journal*, 527, 266

Elmegreen, B.G., Efremov, Y., Pudritz, R., & Zinnecker, H. 2000, in Mannings V., Boss A.P., Russell S.S., eds, Protostars & Planets IV. Univ. Arizona Press, Tucson, p. 179

Elmegreen, B.G. & Elmegreen, D.M. 2001, *Astronomical Journal*, 121, 1507

Elmegreen, B.G., Kim, S., & Staveley-Smith, L. 2001, *The Astrophysical Journal*, 548, 749

Elmegreen, B.G. & Scalo, J.M. 2004, *Annual Review of Astronomy and Astrophysics*, 42, 211

Elmegreen, B.G., Elmegreen, D.M., Chandar, R., et al. 2006, *The Astrophysical Journal*, 644, 879

Elmegreen, D.M., Elmegreen, B.G., Ravindranath, S., & Coe, D.A. 2007, *The Astrophysical Journal*, 658, 763

Elmegreen B.G. 2008, *The Astrophysical Journal*, 672, 1006

Elmegreen, D.M., Elmegreen B.G., Marcus, M.T., et al. 2009, *The Astrophysical Journal*, 701, 306

Elmegreen, B. G. 2010, Proc. Int. Astronomical Union, IAU Symp. 266, Star Clusters: Basic Galactic Building Blocks Throughout Time and Space 3 (Cambridge: Cambridge Univ. Press)

Elmegreen, B.G. & Hunter, D.A. 2010, *The Astrophysical Journal*, 712, 604

Elmegreen, B.G. 2011, EAS Publications Series, 51, 31

Elmegreen, D.M., Elmegreen, B.G., Adamo, A., et al. 2014, *The Astrophysical Journal Letters*, 787, L15

Elmegreen, B.G. 2018, *The Astrophysical Journal*, 853, 88

Engargiola, G., Plambeck, R.L., Rosolowsky, E., & Blitz, L. 2003, *The Astrophysical Journal Supplement Series*, 149, 343

Enoch, M.L. Young, K.E., Glenn, J., et al. 2006, *The Astrophysical Journal*, 638, 293

Escala, A. & Larson, R.B. 2008, *The Astrophysical Journal*, 685, 31

Evans, N.J., Dunham, M.M., Jørgensen, J.K., et al. 2009, *The Astrophysical Journal Supplement Series*, 181, 321

Falgarone, E., Phillips, T.G., & Walker, C.K. 1991, *The Astrophysical Journal*, 378, 186

Fall, S.M., Chandar, R., & Whitmore, B.C. 2005, *The Astrophysical Journal*, 631, L133

Federrath, C., Klessen, R.S., & Schmidt, W. 2009, *The Astrophysical Journal*, 692, 364

Federrath, C. 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 1245

Federrath, C. & Klessen, R.S. 2013, *The Astrophysical Journal*, 763, 51

Ferland, G.J., Korista, K.T., Verner, D.A., Ferguson, J.W., Kingdon, J.B., & Verner, E.M. 1998, *Publications of the Astronomical Society of the Pacific*, 110, 761

Ferland, G.J., Porter, R.L., van Hoof, P.A.M., et al. 2013, *Revista Mexicana de Astronomía y Astrofísica*, 49, 137

Freund, Y. 1995, *Information and Computation*, 121, 256

Freund, Y. & Schapire, R.E. 1999, *Journal of Japanese Society for Artificial Intelligence*, 14, 771

Forbes, D.A., Brodie, J.P., & Huchra, J. 1997, *Astronomical Journal*, 113, 887

Förster Schreiber, N.M., Shapley, A.E., Genzel, R., et al. 2011, *The Astrophysical Journal*, 739, 45

Fouesneau, M., Johnson, L.C., Weisz, D.R., et al. 2014, *The Astrophysical Journal*, 786, 117

Fukui, Y., Mizuno, N., Yamaguchi, R., Mizuno, A., Onishi, T., et al. 1999, *Publications of the Astronomical Society of Japan*, 51, 745

Gieles, M., Bastian, N., Lamers, H.J.G.L.M., & Mout, J.N. 2005, *Astronomy & Astrophysics*, 441, 949

Gieles, M., Larmers, H.J.G.L.M., & Portegies Zwart, S.F. 2007, *The Astrophysical Journal*, 668, 268

Gieles, M., Bastian, N., & Ercolano, B. 2008, *Monthly Notices of the Royal Astronomical Society*, 391L, 93

Gieles, M., Heggie, D. C., & Zhao, H. 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 2509

Gieles, M. & Portegies Zwart, S.F. 2011, *Monthly Notices of the Royal Astronomical Society*, 410, L6

Girichidis, P., Federrath, C., Allison, R., Banerjee, R., & Klessen, R.S. 2012, *Monthly Notices of the Royal Astronomical Society*, 420, 3264

Glover, S.C.O. & Smith, R.J. 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 3011

Goddard, Q.E., Bastian, N., Kennicutt, R.C. 2010, *Monthly Notices of the Royal Astronomical Society*, 405, 857

Gomez, M., Hartmann, L., Kenyon, S.J., & Hewett, R. 1993, *Astronomical Journal*, 105, 1927

Goodman, A.A., Alves, J., Beaumont, C.N., et al. 2014, *The Astrophysical Journal*, 797, 53

Gouliermis, D.A., et al. 2010, *The Astrophysical Journal*, 725, 1717

Gouliermis, D.A., Hony, S., & Klessen, R.S. 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 3775

Gouliermis, D.A., Beerman, L.C., Bianchi, L., et al. 2015a, in "Lessons from the Local Group", eds. K.C. Freeman et al.(New York: Springer), p. 289

Gouliermis, D.A., Thilker, D., Elmegreen, B.G., et al. 2015b, *Monthly Notices of the Royal Astronomical Society*, 452, 3508

Gouliermis, D.A., Elmegreen, B.G., Elmegreen, D.M., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 468, 509

Grabelsky, D.A., Cohen, R.S., Bronfman, L., Thaddeus, P., & May, J. 1987, *The Astrophysical Journal*, 315, 122

Grasha, K., Calzetti, D., Andrews, J.E., Lee, J.C., & Dale, D.A. 2013, *The Astrophysical Journal*, 773, 174

Grasha, K., Calzetti, D., Adamo, A., et al. 2015, *The Astrophysical Journal*, 815, 93

Grasha, K., Calzetti, D., Adamo, A., et al. 2017a, *The Astrophysical Journal*, 840, 113

Grasha, K., Elmegreen, B.G., Calzetti, D., et al. 2017b, *The Astrophysical Journal*, 842, 25

Grasha, K., Calzetti, D., Fedorenko, K., et al. 2018a, *The Astrophysical Journal*, submitted

Grasha, K., Calzetti, D., Bittle, L., et al. 2018b, *The Astrophysical Journal*, submitted

Greisen, E.W., Spekkens, K., & van Moorsel, G.A. 2009, *Astronomical Journal*, 137, 4718

Grenier, I.A., Casandjian, J.-M., Terrier, R. 2005, Science, 307, 1292

Grudić, M.Y., Guszejnov, D., Hopkins, P.F., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, arxiv: 1708.09065

Grudić, M.Y., Hopkins, P.F., Faucher-Giguère, C.-A., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, 475, 3511

Guo, Y., Giavalisco, M., Ferguson, H.C., et al. 2012, *The Astrophysical Journal*, 757, 120

Guo, Y., Rafelski, M., Bell, E.F., et al. 2018, *The Astrophysical Journal*, 853, 108

Gustafsson, B., Church, R.P., Davies, M.B., & Rickman, H. 2016, *Astronomy & Astrophysics*, 593, 85

Guszejnov, D. & Hopkins, P.F. 2016, *Monthly Notices of the Royal Astronomical Society*, 459, 9

Guszejnov, D., Hopkins, P.F., & Krumholz, M.R. 2017, *Monthly Notices of the Royal Astronomical Society*, 468, 4093

Gutermuth, R.A., Megeath, S.T., Myers, P.C., Allen, L.E., Pipher, J. L., & Fazio, G.G. 2009, *The Astrophysical Journal Supplement Series*, 184, 18

Heiles, C. & Troland, T.H. 2003, *The Astrophysical Journal*, 586, 1067

Hennebelle, P. & Chabrier, G. 2013, *The Astrophysical Journal*, 770, 150

Herbert-Fort, S., Zaritsky, D. Christlein, D., & Kannappan, S.J. 2010, *The Astrophysical Journal*, 715, 902

Heyer, M. & Dame, T.M. 2015, *Annual Review of Astronomy and Astrophysics*, 53, 583

Hollyhead, K., Bastian, N., Adamo, A., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 1106

Hollyhead, K., Adamo, A.; Bastian, N., Gieles, M., & Ryon, J.E. 2016, *Monthly Notices of the Royal Astronomical Society*, 460, 2087

Hony, S., Gouliermis, D.A., Galliano, F., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 448, 1847

Hoopes, C.G., Walterbos, R.A.M., & Bothun, G.D. 2001, *The Astrophysical Journal Letters*, 631, L133

Hopkins, P.F. 2013a, *Monthly Notices of the Royal Astronomical Society*, 428, 1950

Hopkins, P.F. 2013b, *Monthly Notices of the Royal Astronomical Society*, 430, 1653

Hopkins, P.F., Narayanan, D., & Murray, N. 2013, *Monthly Notices of the Royal Astronomical Society*, 432, 2647

Hubble, E.P. 1925, *The Astrophysical Journal*, 62, 409

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, *The Astrophysical Journal Supplement Series*, 221, 8

Hughes, A., Meidt, S., Schinnerer, E., et al. 2013, *The Astrophysical Journal*, 779, 44

Immeli, A., Samland, M., Westera, P., & Gerhard, O. 2004, *The Astrophysical Journal*, 611, 20

Infante, L. & Pritchet, C.J. 1995, *The Astrophysical Journal*, 438, 565

Jeffreson, S.M.R. & Kruijssen, J.M.D. 2018, *Monthly Notices of the Royal Astronomical Society*, 476, 3688

Johnson, L.C., Seth, A.C., Dalcanton, J.J., et al. 2012, *The Astrophysical Journal*, 752, 95

Johnson, L.C., Seth, A.C., Dalcanton, J.J., et al. 2015, *The Astrophysical Journal*, 802, 127

Johnson, L.C., Seth, A.C., Dalcanton, J.J., et al. 2016, *The Astrophysical Journal*, 827, 33

Johnstone, D., Wilson, C.D., Moriarty-Schieven, G., Joncas, G., Smith, G., Gregersen, E., & Fich, M. 2000, *The Astrophysical Journal*, 454, 327

Johnstone, D., Fich, M., Mitchell, G.F., & Moriarty-Schieven, G. 2001, *The Astrophysical Journal*, 559, 307

Kahre, L., Walterbos, R.A., Kim, H., et al. 2018, *The Astrophysical Journal*, 855, 133

Kawamura, A., Mizuno, Y., Minamidani, T., et al. 2009, *The Astrophysical Journal Supplement Series*, 184, 1

Kennicutt, R.C. 1998, *Annual Review of Astronomy and Astrophysics*, 36, 189

Kennicutt, R.C. & Evans, N.J. 2012, *Annual Review of Astronomy and Astrophysics*, 50, 531

Kerscher, M., Szapudi, I., & Szalay, A.S. 2000, *The Astrophysical Journal Letters*, 535, L13

Kim, S. & Park, C. 2007, *The Astrophysical Journal*, 663, 244

Klessen, R.S. & Glover, S.C.O. 2016, Star Formation in Galaxy Evolution: Connecting Numerical Models to Reality, Volume 43. ISBN 978-3-662-47889-9 (Springer-Verlag Berlin Heidelberg)

Knapen, J.H., Mazzuca, L.M., Böker, T., Shlosman, I., Colina, L., Combes, F., & Axon, D.J. 2006, *Astronomy & Astrophysics*, 448, 489

Kotsiantis, S.B. 2007, *Informatica*, 31, 249

Kritsuk, A.G., Lee, C.T., & Norman, M.L. 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 3247

Kroupa, P. 2002, *Science*, 295, 82

Kroupa, P. 2001, *Monthly Notices of the Royal Astronomical Society*, 322, 231

Kruijssen, J.M.D., Pelupessy, F.I., & Lamers, H.J.G.L.M. 2011, *Monthly Notices of the Royal Astronomical Society*, 414, 1339

Kruijssen, J.M.D. 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 3008

Krumholz, M.R. & McKee, C.F. 2005, *The Astrophysical Journal*, 630, 250

Krumholz, M.R. 2014, *Physics Reports*, 539, 49

Krumholz, M.R. Bate, M.R., Arce, H.G., et al. 2014, Protostars and Planets VI, Henrik Beuther, Ralf S. Klessen, Cornelis P. Dullemond, and Thomas Henning (eds.), University of Arizona Press, Tucson, 914 pp., p.243-266

Krumholz, M.R., Adamo, A., Fumagalli, M., et al. 2015a, *The Astrophysical Journal*, 812, 147

Krumholz, M.R., Fumagalli, M., da Silva, R.L., Rendahl, T., & Parra, J. 2015b, *Monthly Notices of the Royal Astronomical Society*, 452, 1447

Lada, C.J. 1987, IAUS, 115, 1

Lada, C.J. & Lada, E.A. 2003, *Annual Review of Astronomy and Astrophysics*, 41, 57

Lada, C.J., Lombardi, M., & Alves, J.F. 2009, *The Astrophysical Journal*, 703, 52

Lada, C.J., Lombardi, M., & Alves, J.F. 2012, *The Astrophysical Journal Letters*, 745, 190

Lamb, J.B., Oey, M.S., Segura-Cox, D.M., et al. 2016, *The Astrophysical Journal*, 817, 113

Landy, S.D. & Szalay, A.S. 1993, *The Astrophysical Journal*, 412, 64

Larson, R.B. 1981, *Monthly Notices of the Royal Astronomical Society*, 194, 809

Larson, R.B. 1982, *Monthly Notices of the Royal Astronomical Society*, 200, 159

Larson, R.B. 1995, *Monthly Notices of the Royal Astronomical Society*, 272, 213

Larsen, S.S., Forbes, D.A., & Brodie, J.P. 2001, *Monthly Notices of the Royal Astronomical Society*, 327, 1116

Larsen S.S. 2002, *Astronomical Journal*, 124, 1393

Larsen, S.S. 2004, *Astronomy & Astrophysics*, 416, 537

Lazarian, A. & Pogosyan, D. 2000, *The Astrophysical Journal*, 537, 720

Lee, M.G., Chandar, R., Whitmore, B.C. 2005, *Astronomical Journal*, 130, 2128

Lee, J.C., Gil de Paz, A., Tremonti, C., et al. 2009, *The Astrophysical Journal*, 706, 599

Leitherer, C., Schaerer, D., Goldader, J.D., et al. 1999, *The Astrophysical Journal Supplement Series*, 123, 3

Leroy, A.K., Walter, F., Brinks, E., et al. 2008, *Astronomical Journal*, 136, 2782

Li, Y., Mac Low, M.-M., & Klessen, R.S. 2005, *The Astrophysical Journal*, 628, 823

Limber, D.N. 1954, *The Astrophysical Journal*, 119, 655

Lintott, C., Schawinski, K., Slosar, A., et al. 2008, *Monthly Notices of the Royal Astronomical Society*, 389, 1179

Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 410, 166

Ma, J., Peng, Q.-H., & Gu, Q.-S. 1998, *Astronomy & Astrophysics Supplement*, 130, 449

Mac Low, M.-M. & Klessen, R.S. 2004, Rev. Mod. Phys., 76, 125

Madau, P. & Dickinson, M. 2014, *Annual Review of Astronomy and Astrophysics*, 52, 415

Maíz-Apellániz, J. 2001, *The Astrophysical Journal*, 563, 151

Maíz Apellániz, J. 2009, *The Astrophysical Journal*, 699, 1938

Mandelbrot, B.B. 1982, The Fractal Geometry of Nature (New York: Freeman)

McKee, C.F. & Ostriker, E.C. 2007, *Annual Review of Astronomy and Astrophysics*, 45, 565

McLaughlin, D.E. & Pudritz, R.E. 1996, *The Astrophysical Journal*, 457, 578

Meidt, S.E., Schinnerer, E., García-Burillo, S., et al. 2015, *The Astrophysical Journal*, 804, 78

Messa, M., Adamo, A., Östlin, G., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, 473, 996

Messa, M., Adamo, A., Calzetti, D., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, submitted

Meurer, G.R., Heckman, T.M., Leitherer, C., Kinney, A., Robert, C., & Garnett, D.R. 1995, *Astronomical Journal*, 110, 2665

Miura, R.E., Kohno, K., Tosaki, T., Espada, D., Hwang, N., et al. 2012, *The Astrophysical Journal Supplement Series*, 761, 37

Miville-Deschênes, M.-A., Martin, P.G., Abergel, A., et al. 2010, *Astronomy & Astrophysics*, 518, L104

Mizuno, N., Tubio, M., Yamaguchi, R., Onishi, T., & Fukui, Y. 2001, *Publications of the Astronomical Society of Japan*, 53, L45

Mora, M.D., Larsen, S.S., Kissler-Patig, M., Brodie, J.P., & Richtler, T. 2009, *Astronomy & Astrophysics*, 501, 949

Murray, N. 2011, *The Astrophysical Journal*, 729, 133

Nomura, H. & Kamaya, H. 2001, *Astronomical Journal*, 121, 1024

Ochsendorf, B.B., Meixner, M., Roman-Duval, J., Rahman, M., & Evans II, N.J. 2017, *The Astrophysical Journal*, 841, 109

Odekon, M.C. 2006, *Astronomical Journal*, 132, 1834

Odekon, M.C. 2008, *The Astrophysical Journal*, 681, 1248

Oey, M.S., King, N.L., & Parker, J.W. 2004, *Astronomical Journal*, 127, 1632

Oh, S.-E., Hunter, D.E., Brinks, E., et al. 2015, *Astronomical Journal*, 149, 180

Olivares, E.F., Hamuy, M., Pignata, G., et al. 2010, *The Astrophysical Journal*, 715, 833

Onodera, S., Kuno, N., Tosaki, T., et al. 2010, *The Astrophysical Journal*, 722, 127

Padoan, P., Kim, S., Goodman, A., & Staveley-Smith, L. 2001, *The Astrophysical Journal Letters*, 555, L33

Padoan, P. & Nordlund, Å. 2002, *The Astrophysical Journal*, 576, 870

Padoan, P., Nordlund, Å, 2011, *The Astrophysical Journal*, 730, 40

Payne-Gaposchkin, C.H. 1974, *Smithsonian Contribution Astrophysics*, 16

Peacock, M.B., Zepf, S.E., & Finzell, T. 2013, *The Astrophysical Journal*, 769,126

Peebles, P.J.E. 1975, *The Astrophysical Journal*, 196, 647

Peebles, P.J.E. 1980, 'The Large-Scale Structure of the Universe', (Princeton, N.J.: Princeton University Press)

Pellerin, A., Meyer, M., Harris, J., & Calzetti, D. 2007, *The Astrophysical Journal Letters*, 658, L87

Pellerin, A., Meyer, M., Calzetti, D. & Harris, J. 2012, *Astronomical Journal*, 144, 182

Peng, Q-H. 1988, *Astronomy & Astrophysics*, 206, 18

Pety, J., Schinnerer, E., Leroy, A.K., et al. 2013, *The Astrophysical Journal*, 779, 43

Pietrzyński, G., Gieren, W., Hamuy, M., et al. 2010, *Astronomical Journal*, 140, 1475

Pilyugin, L.S., Grebel, E.K., & Kniazev, A.Y. 2014, *Astronomical Journal*, 147, 131

Piotto, G., Villanova, S., Bedin, L.R., et al. 2005, *The Astrophysical Journal*, 621, 777

Pokhrel, R., Myers, P.C., Dunham, M.M., et al. 2018, *The Astrophysical Journal*, 853, 5

Portegies Zwart, S.F., McMillan, S.L.W., & Gieles, M. 2010, *Annual Review of Astronomy and Astrophysics*, 48, 431

Priyatikanto, R., Kouwenhoven, M.B.N., Arifyanto, M.I., Wulandari, H.R.T., & Siregar, S. 2016, *Monthly Notices of the Royal Astronomical Society*, 457, 1339

Postman, M., Lauer, T.R., Szapudi, I., & Oegerle, W. 1998, *The Astrophysical Journal*, 506, 33

Querejeta, M., Meidt, S.E., Schinnerer, E., et al. 2016, *Astronomy & Astrophysics*, 588, 33

Radburn-Smith, D.J., Roškar, R. Debattista, V.P., et al. 2012, *The Astrophysical Journal*, 753, 138

Rahna, P.T., Das, M., Murthy, J., Gudennavar, S.B., & Bubbly, S.G. 2018, arxiv: 1803.00506

Renaud, F., Bournaud, F., Emsellem, E., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 1836

Rice, T.S., Goodman, A.A., Edwin, A.B., et al. 2016, *The Astrophysical Journal*, 822, 52

Roman-Duval, J., Jackson, J.M., Heyer, M., Rathborne, J., & Simon, R. 2010, *The Astrophysical Journal*, 723, 492

Rosolowsky, E. & Leroy, A.K. 2006, *Publications of the Astronomical Society of the Pacific*, 118, 590

Rosolowsky, E.W., Pineda, J.E., Kauffmann, J., & Goodman, A.A. 2008, *The Astrophysical Journal*, 679, 1338

Roychowdhury, S., Chengalur, J.N., Begum, A., & Karachentsev, I.D. 2010, *Monthly Notices of the Royal Astronomical Society*, 404, L60

Ryder, S.D., Staveley-Smith, L., Malin, D., & Walsh, W. 1995, *Astronomical Journal*, 109, 1592

Ryon, J.E., Gallagher, J.S. III, Smith, L.J., et al. 2017, *The Astrophysical Journal*, 841, 92

Salpeter, E.E. 1955, *The Astrophysical Journal*, 121, 161

Sánchez, N., Alfaro, E. J., & Pérez, E. 2005, *The Astrophysical Journal*, 625, 849

Sánchez, N. & Alfaro, E.J. 2008, *The Astrophysical Journal Supplement Series*, 178, 1

Sánchez, N. & Alfaro, E.J. 2009, *The Astrophysical Journal*, 696, 2086

Sánchez, N., Añez, N., Alfaro, E., Odekon, M.C. 2010, *The Astrophysical Journal*, 720, 541

Sánchez-Gil, M.C., Jones, D.H., Pérez, E., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 753

Santiago, B.X. 2008, *Astronomy & Astrophysics*, 492, 23

Scalo, J.M. 1985 in Protostars and Planets II, ed. D.C. Black & M.S. Mathews (Tucson: Univ. of Arizona Press), 201

Scheepmaker, R.A., Haas, M.R., Gieles, M., Bastian, N., Larsen, S.S., & Lamers, H.J.G.L.M. 2007, *Astronomy & Astrophysics*, 469, 925

Scheepmaker, R.A., Lamers, H.J.G.L.M., Anders, P., & Larsen, S.S. 2009, *Astronomy & Astrophysics*, 494, 81

Schinnerer, E., Meidt, S.E, Pety, J., et al. 2013, *The Astrophysical Journal*, 779, 42

Schinnerer, E., Meidt, S.E, Colombo, D., et al. 2017, *The Astrophysical Journal*, 836, 62

Schlafly, E.F. & Finkbeiner, D.P. 2011, *The Astrophysical Journal*, 737, 103

Schruba, A., Leroy, A.K., Walter, F., Sandstrom, K, & Rosolowsky, E. 2010, *The Astrophysical Journal*, 722, 1699

Schuster, K.F., Kramer, C., Hitschfeld, M., Garcia-Burillo, S., & Mookerjea, B. 2007, *Astronomy & Astrophysics*, 461, 143

Shapley, H. & Curtis, H.D. 1921, Bulletin of the National Research Council, 2, 171

Silva-Villa, E. & Larsen, S.S. 2010, *Astronomy & Astrophysics*, 516, 10

Silva-Villa, E. & Larsen, S.S. 2011, *Astronomy & Astrophysics*, 529, 25

Silva-Villa, E., Adamo, A., Bastian, N., Fouesneau, M., & Zackrisson, E. 2014, *Monthly Notices of the Royal Astronomical Society*, 440, L116

Smajić, S., Moser, L., Eckart, A., et al. 2015, *Astronomy & Astrophysics*, 583, 104

Solomon, P.M., Rivolo, A.R., Barrett, J., & Yahil, A. 1987, *The Astrophysical Journal*, 319, 730

Soto, E., de Mello, D.F., Rafelski, M., et al. 2017, *The Astrophysical Journal*, 837, 6

Stanke, T., Smith, M.D., Gredel, R., & Khanzadyan, T. 2006, *Astronomy & Astrophysics*, 447, 609

Storri-Lombardi, M.C., Lahav, O., Sodre, L, & Storri-Lombardi, L.J. 1992, *Monthly Notices of the Royal Astronomical Society*, 259, 8

Subramaniam, A., Gorti, U., Sagar, R., Bhatt, H. 1995, *Astronomy & Astrophysics*, 302, 86

Sun, N.-C., de Grijs, R., Subramanian, S., et al. 2017a, *The Astrophysical Journal*, 835, 171

Sun, N.-C., de Grijs, R., Subramanian, S., et al. 2017b, *The Astrophysical Journal*, 849, 149

Tamburro, D., Rix, H.-W., Leroy, A.K., Mac Low, M.-M., Walter, F., Kennicutt, R.C., Brinks, E., & de Blok, W.J.G. 2009, *Astronomical Journal*, 137, 4424

Tielens, A.G.G.M. & Hollenbach, D. 1985, *The Astrophysical Journal*, 291, 722

Tremonti, C.A., Calzetti, D., Leitherer, C. & Heckman, T.M. 2001, *The Astrophysical Journal*, 555, 322

Tonry, J.L., Dressler, A., Blakeslee, J.P., et al. 2001, *The Astrophysical Journal*, 546, 681

Usero, A., Leroy, A.K., Walter, F., et al. 2015, *Astronomical Journal*, 150, 115

Vázquez, G.A. & Leitherer, C. 2005, *The Astrophysical Journal*, 621, 695

Vázquez-Semadeni, E., Gómez, G.C., Jappsen, A.-K., Ballesteros-Paredes, J., & Klessen, R.S. 2009, *The Astrophysical Journal*, 707, 1023

Verdes-Montenegro, L., Bosma, A., & Athanassoula, E. 2000, *Astronomy & Astrophysics*, 356, 827

Villanova, S., Piotto, G., King, I.R., et al. 2007, *The Astrophysical Journal*, 663, 296

Weidner, C., Bonnell, I.A., & Zinnecker, H. 2010, *The Astrophysical Journal*, 724, 1503

Westerlund, B.E. 1997, The Magellanic Clouds, Cambridge Astrophysics Series Vol. 29 (Cambridge: Cambridge Univ. Press)

Whitmore, B.C., Zhang, Q., Leitherer, C., Fall, S.M., Schweizer, F., & Miller, B.W. 1999, *Astronomical Journal*, 118, 1551

Whitmore, B.C, Gilmore, D., Leitherer, C., et al. 2005, *Astronomical Journal*, 130, 2104

Whitmore, B.C., Chandar, R., Schweizer, F., et al. 2010, *Astronomical Journal*, 140, 75

Whitmore, B.C., Brogan, C., Chandar, R., et al. 2014, *The Astrophysical Journal*, 795, 156

Wilson, R.W., Jefferts, K.B., & Penzias, A.A. 1970, *The Astrophysical Journal Letters*, 161, L43

Wilson, B.A., Dame, T.M., Masheder, M.R.W., & Thaddeus, P. 2005, *Astronomy & Astrophysics*, 430, 523

Wisnioski, E., Glazebrook, K., Blake, C., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 422, 3339

Witt, A.N., Thronson, H.A.J., & Capuano, J.M.J. 1992, *The Astrophysical Journal*, 393, 611

Wolfire, M.G., McKee, C.F., Hollenbach, D., Tielens, A.G.G.M. 2003, *The Astrophysical Journal*, 587, 278

Wolfire, M.G., Hollenbach, D. & McKee, C.F. 2010, *The Astrophysical Journal*, 716, 1191

Wright, N.J., Parker, R.J., Goodwin, S.P., & Drake, J.J. 2014, *Monthly Notices of the Royal Astronomical Society*, 438, 639

Xu, D. & Offner, S.S.R. 2017, *The Astrophysical Journal*, 851, 149

Yamaguchi, R., Mizuno, N., Mizuno, A., et al. 2001, *Publications of the Astronomical Society of Japan*, 53, 985

Young, K.E., Enoch, M.L., Evans, N.J., et al. 2006, *The Astrophysical Journal*, 644, 326

Zackrisson, E., Rydberg, C.-E., Schaerer, D., Östlin, G., & Tuli, M. 2011, *The Astrophysical Journal*, 740, 13

Zhang, Q., Fall, S.M., & Whitmore, B.C. 2001, *The Astrophysical Journal*, 561, 727

Zucker, C., Battersby, C., & Goodman, A. 2015, *The Astrophysical Journal*, 815, 23

Zwicky, F. 1953, *Publications of the Astronomical Society of the Pacific*, 65, 205