OPTICAL GALAXY CLUSTER DETECTION ACROSS A WIDE REDSHIFT RANGE

by

Jiangang Hao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2009

Doctoral Committee:

Professor Timothy A. Mckay, Chair
Professor Ratindranath Akhoury
Professor August Evrard
Professor Gregory Tarle
Assistant Professor Sally Oey

To My Mom

ACKNOWLEDGMENTS

The current work is a result of several years efforts, trial & error and back-and-forth improvements. It would not be possible without the help and support of my advisor, colleagues, friends and family.

I am especially grateful to my advisor, Tim Mckay. He is an amazing person, with intriguing ideas and keen perceptions. I especially benefit from his blending of strong curiosity in Science and down-to-earth attitude toward research. I am really lucky for the right timing to get involved in his interesting and important project on galaxy cluster detection. It is really a joyful experience to work with him. Besides his great help on my research, I also learned a lot from his way of mentoring students, organizing groups and developing research projects, which are all very important for my future career.

I am greatly indebted to Ben Koester. For a long time, even now, he was my major source of information for clusters, astronomy and IDL techniques. Without his generous help, my research would be twice as hard. His previous work on cluster finding sets such a high standard that I spent two years trying to find a way to go further. Thanks, Ben.

I am very lucky to be in the office next to Gus Evrard, our leading theorist in this field. I really enjoy many discussions with him and his ways of telling me what is possible, practical and what is not. Many of his ideas are an integral part of this project.

I am also grateful to my friend and officemate Brian Nord, with whom I have gone to many conferences and workshops together over the past 5 years. We share the same office and I benefited a lot from numerous conversations with him. From him, I also learned a lot about American culture, people and life.

I thank my friend Jing Shao and enjoy our many fruitful discussions. He clarified many of my confusions about the microscopic aspects of our Universe. My friend Guin-Dar Lin, with whom I have shared an apartment in Ann Arbor, gave me a lot of help on getting my Linux computer to work. I enjoy very much our discussions about some interesting physics in our daily life. My friend Bingqiang Wang at Shanghai Supercomputing Center also gave me much advice on computing related issues and brought my attention to many fancy software packages.

I gave my sincere thanks to Dragan Huterer and Carlos Cunha for their clarifications on the theoretical parts of cluster cosmology. I thank Eli Rykoff and Erin Sheldon for their help on taming IDL.

I am also thankful to many other professors and colleagues. They provided all kinds of help for me over the past 5 years. They are: Fred Adams, Ratindranath Akhoury, Jim Annis, Matt Becker, Michael Busha, Anbo Chen, Aaron Fenyes, David Gerdes, David Johnston, Eva-Marie Proszkow, Elena Rasia, Eduardo Rozo, Joel Smoller, Michael Schubnel,, Greg Tarle, Risa Wechsler, Kimmy Wu and York Peng Yao, and I thank you all for your help.

I am greatly indebted to many people from the statistics community, who gave me many good suggestions and comments. Especially, Jon Kettenring, Kerby Shedden and Hyunsook Lee gave me a lot of help on various topics in statistics, from which I have learned a lot.

Last but not least, I want to thank my wife, Jing. Without her support and understanding, I could not go this far. Her love and encouragement are so important for me to pass through the difficult times in my research and life.

PREFACE

"There is a theory which states that if anybody ever discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened"

Douglas Adams

Where do we come from? What are we? Where are we going? These questions have perplexed human beings ever since we started intellectual reasoning. However, most attempts to answer these questions in history were simply based on logic, not on evidence. If we look these questions from the perspective of modern physics, we immediately realize that the above questions should be traced back to the following questions. Where does the Universe come from? What is the Universe made of? Where is the Universe going?

Over the past century, revolutionary progress in physical science enabled us to discuss the Universe in terms of empirical evidence rather than pure speculation. Over the past decade, progress in technology has pushed Cosmology into the precision era. The composition, evolution and fate of the Universe has became one of the major

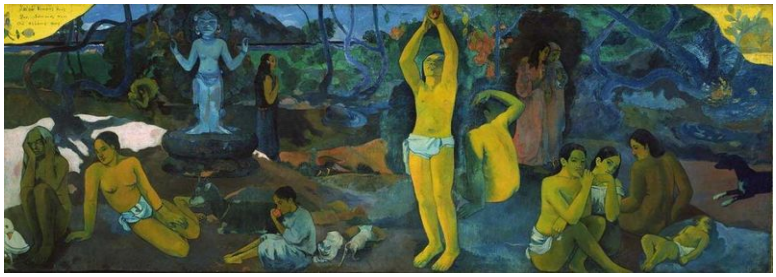


Figure 1. *Where Do We Come From? What Are We? Where Are We Going?* Painting by Paul Gauguin, (1897-1898). Courtesy of Museum of Fine Arts, Boston. Adapted from Wikipedia

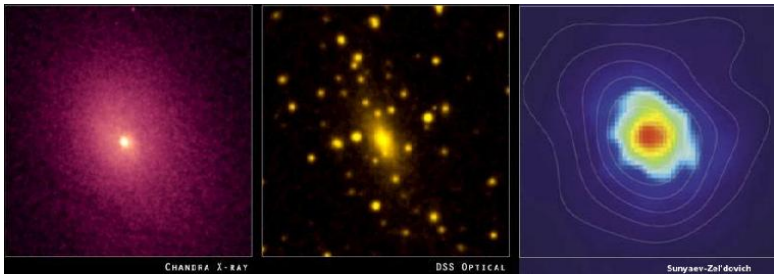


Figure 2. Galaxy cluster show itself in multi-wavelength of electromagnetic emission. From left to right, each corresponding to X-ray, optical and microwave band.

topics for serious scientists.

Among the four known fundamental interactions, only electromagnetism and gravity can carry interactions over a long range. In our Universe, most objects are electrically neutral but massive. Therefore the dominant interaction on large scales in the Universe is gravity. However, the "charge" for gravitational interaction is Mass and the strength of gravity is lot weaker ($\sim 10^{36}$) than their electromagnetic counterpart, making them VERY weakly coupled to our instruments. As a result, only when the gravitational effects are manifested by electromagnetic phenomena, we can reliably detect them to high precision with our current technology. Therefore, we need something that is gravitationally significant so that cosmological information can be effectively encoded in, while also manifesting strong electromagnetic features so that we can reliably detect it.

Under the above criterion, galaxy clusters stand out immediately. They are so massive that they get the strongest imprints during the evolution of the Universe. On the other hand, they also show strong features in various wavelengths of the electromagnetic emission (e.g. see Figure 2), leading to their reliable detection. This unique characteristic make galaxy clusters a very important probe for cosmology.

To probe the Universe with galaxy clusters, the first step is to build a large catalog for clustered galaxies. This is the major topic of this work!

CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER 1

Introduction

"The most incomprehensible thing about the world is that it is comprehensible"

Albert Einstein

1.1 Current Status of Cosmology

The most exciting discovery in Physics and Astronomy over the past decade is the accelerating expansion of our Universe initially revealed by Supernovae experiments (Perlmutter et al., 1999; Riess et al., 1998). Despite initial skepticism, this discovery has been confirmed by other independent experiments based on the cosmic microwave background (Spergel et al., 2003, 2007), the galaxy power spectrum (Tegmark et al., 2004) and baryon acoustic oscillations (Eisenstein et al., 2005). This discovery challenges many of our established notions about the evolution and composition of the Universe. It requires dramatic changes: either gravity changes from attractive to repulsive on very large scales or the cosmic composition needs to be changed to include substance that can offset gravitational attraction and produce repulsive interactions at large scales.

Given the fact that gravity has been well tested on solar system scales and its attractive nature is consistent with galaxy cluster scale observations, it is difficult to imagine that gravity turns repulsive at larger scales. This requires gravity to become zero at a certain scale when it turns from attractive to repulsive. Therefore, most people prefer to imagine a substance that has negative pressure and can drive the acceleration of the cosmological expansion. This mysterious substance is now called

Dark Energy, and observations suggest that it makes up of $\sim 74\%$ of the total energy density of the Universe.

What is dark energy? This question is one of the major topics of current Physics and Astronomy. A plethora of theoretical models for dark energy have been proposed and most are consistent with current observations. Therefore, tighter constraints on the energy density and evolution of dark energy from new experiments are essential for narrowing down the list of candidates.

There are several different ways to constrain the energy density and evolution of dark energy, each of which leads to degeneracy in certain directions of parameter space. Therefore, a combination of complementary methods will help to reduce the degeneracies and give tighter constraints on each parameter. The existence of dark energy will affect both the geometry of the Universe and large scale structure formation. Among the possible experiments, some constrain dark energy mainly via geometry (e.g. CMB, Supernovae); while some others provide constraints from both structure growth and geometry (e.g. BAO, Galaxy Clusters). In Figure 1.1, we show the constraints on dark energy from various experiments. In this work, we will mainly consider the constraints from galaxy clusters.

1.2 Structure Formation and Galaxy Clusters

The large scale structures in our Universe are grown from initial fluctuations of quantum fields during inflation. The fluctuations are amplified under gravity and evolve to form large scale structures. Luminous matter[1] does not play a major role in this process because its total mass is only about one fifth of another unknown substance: Dark Matter. Figure 1.2 shows a summary of the cosmic energy budget for different components. Compared to dark energy, we know a little bit more about dark matter. Dark matter is very massive and interacts mainly via gravity. For more details about dark matter, refer to the review paper (Bertone et al., 2005). A major difference between dark matter and dark energy is that dark energy produces effectively repul-

[1]Strictly speaking, should be the Standard Model particles

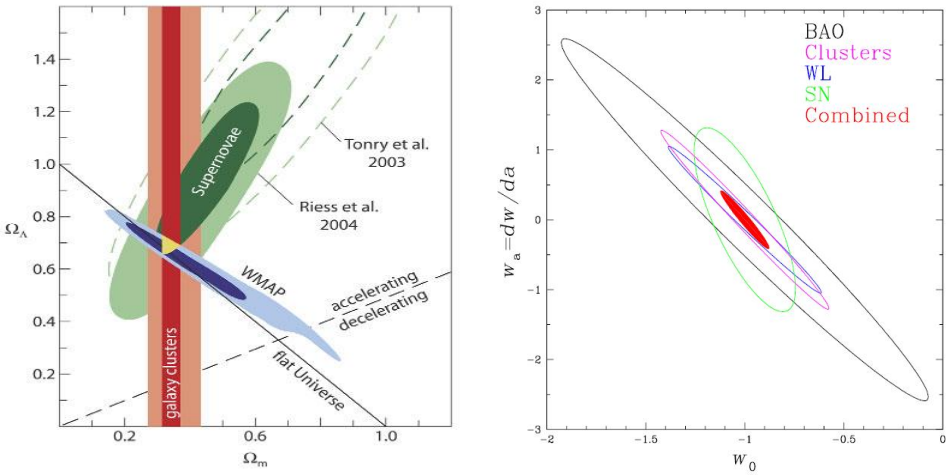


Figure 1.1. The constraints on the dark energy from various observations. The panel on the left assumes dark energy is a vacuum energy and the right panel assumes dark energy is a dynamically evolving component with equation of state parameterized as $w(a) = w_0 + \frac{dw}{da}(1-a)$. The panel on the left is adopted from ESO press release (ESO, 2004) and the right panel is adopted from the Dark Energy Survey science proposal (Frieman et al, 2006)

sive interaction, while dark matter exerts ordinary gravity as described by general relativity.

Therefore, the major driving force for structure formation is from dark matter. Big fluctuations of the dark matter density field collapse under gravity and form bound systems, called dark matter halos. Luminous matter then cools into the dark matter halos to form galaxies and stars. The smaller halos carried with galaxies merge together to form bigger halos whose luminous counterparts form galaxy clusters. The

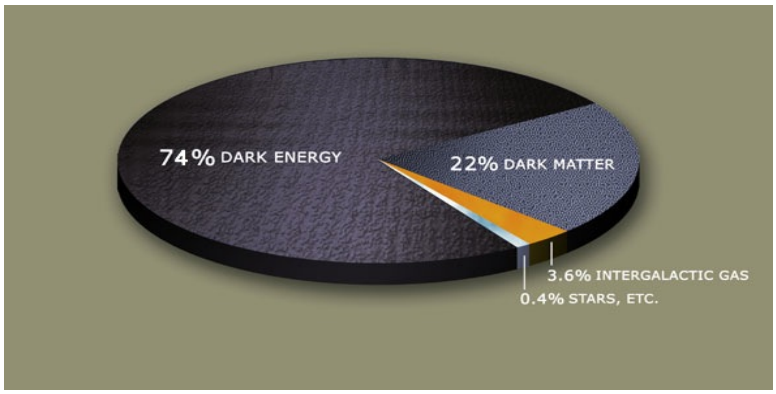


Figure 1.2. Current constraints on the energy densities of different constituents in our Universe. The image is reproduced from (Wikipedia, 2009)

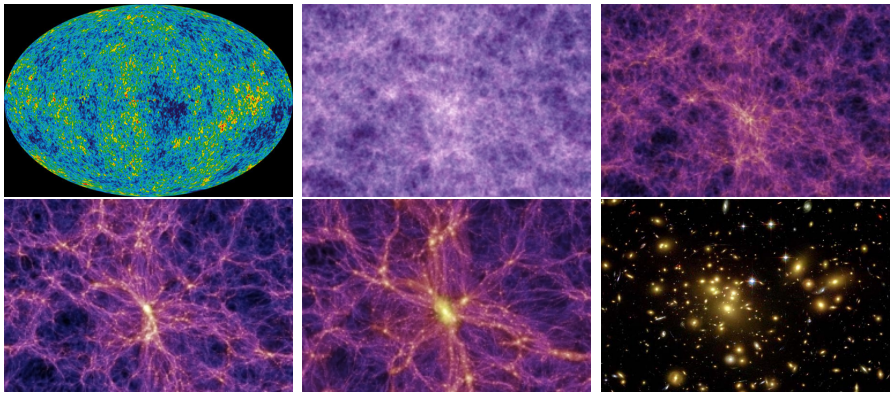


Figure 1.3. The formation of large scale structures and galaxy clusters from the initial fluctuations in primordial plasma. From top left to bottom right are, CMB map(from WMAP website), Millennium Simulation of structure formation at redshift 18.3, 5.7, 1.4 and 0.0 (from VIRGO CONSORTIUM website) and galaxy cluster (from Spacedaily.com)

pictures in Figure 1.3 show the stages of structure formation. Large scale N-body simulations (Evrard et al., 2002; Springel et al., 2005) can reliably predict the formation of halos of various mass scales given a specific set of cosmological parameters. Because these predictions are robust, it is possible to constrain the cosmological parameters from measurement of halo abundance and the spatial distribution. However, dark matter halos neither emit nor absorb electromagnetic radiation, making their detection very hard. Therefore, we should look for something that can be a good tracer of dark matter halos at large mass. Galaxy clusters are reliable probes of dark matter halos.

1.3 Galaxy Clusters as Cosmological Probes

In this section, we will briefly outline the theoretical connection between galaxy cluster and cosmological parameters and the challenge.

1.3.1 The Hard Reality

As we pointed out in the previous section, dark matter plays the major role in large scale structure formation. N-body simulation can predict the abundance and distribution of dark matter halos under a set of cosmological parameters (Sheth & Tormen, 1999). However, dark matter halos do not have electromagnetic emissions and can-

not be detected directly. We need to seek for something that can be easily detected and is a good tracer of dark matter halos. Galaxy clusters are appropriate tracers of massive dark matter halos.

Observationally, we can detect galaxy clusters from three major wavelengths of their electromagnetic emission: microwave, optical and X-rays. In the microwave band, cluster detection is facilitated by the Sunyaev-Zeldovich effect (SZE) (Sunyaev & Zeldovich, 1970; Carlstrom et al., 2002). The CMB photons are scattered by the high energy electrons in the intracluster medium (ICM) of galaxy clusters, leaving imprints of galaxy clusters on the CMB map. One very attractive advantage of detecting clusters using SZE is its redshift independence, allowing a "uniform" detection across a wide redshift range. For more details about SZE cluster detection, see Carlstrom et al. (2002, and references therein). Most recently, four clusters detected from SZE have been reported by the South Pole Telescope team (Staniszewski et al., 2008, SPT), which is good evidence of the feasibility of SZE cluster detection.

However, SZE does not differentiate clusters along the line of sight. The observed SZE signal is an integrated effect of all clusters along the same line of sight. Therefore, the "clusters" detected from SZE are two dimensional projections of the actual clusters. To resolve clusters along the line of sight, optical follow up is needed to get the redshifts for the SZE detected clusters.

In the X-ray band, galaxy clusters also show strong emission. The hot gas trapped in the deep potential well interacts violently, leading to strong X-ray emission. Since the X-ray emission is closely related to the hot gas trapped in the potential well of the cluster, their luminosity is negligible outside of the cluster, making the detection of galaxy clusters in the X-ray band less prone to contamination due to projection.

However, detection of clusters in the X-ray band cannot be done through ground based experiments due to strong atmospheric absorption. Since the 1990s, several satellite based X-ray observatories have been deployed to detect clusters in the X-ray band. The ROSAT all-sky survey (Voges et al., 1999, and references therein) is the first full sky X-ray survey and gives rise to several X-ray cluster catalogs: NORAS (Böhringer et al., 2000), REFLEX (Böhringer et al., 2004) and 400 $\deg^2$ (Burenin

et al., 2007). More recently, XMM Newton (Jansen et al., 2001) and the Chandra satellite (Vikhlinin et al., 2006) were launched to study the detailed X-ray properties of selected galaxy clusters. Similar to detecting clusters using SZE, X-ray cluster detection cannot provide the redshift of the cluster. Additional follow up in the optical band is necessary to assign redshift to the detected clusters.

In the optical band, galaxy clusters show strong features. Compared to the aforementioned cluster detection in the microwave band and the X-ray band, optical cluster detection is less expensive and can detect the redshifts of the clusters. The disadvantage lies in that the contamination due to projection in optical band is much more than that in the other two band because optical emission is not exclusive to clustered galaxies. Especially, when we look for clusters across a wide redshift range, the contamination due to projection will significantly degrade cluster detection as well as measurements of their properties. We will leave the details to later sections.

If we want to constrain cosmological parameters using galaxy clusters, we essentially need to reconstruct the distribution of halos and their masses from clusters. However, reconstructing halo distribution from galaxy clusters is not an easy and clean job. Current N-body simulations cannot address this issues because the physics of how galaxies accumulate to halos is not yet well settled. The following two major challenges emerge when we reconstruct the halo distributions from clusters.

The first challenge is the bias between clusters and halos. This is partially a issue of definition of clusters and halos. Galaxy clusters (baryonic matter) will trace the dark matter halos through gravitational interaction, but not necessarily as a one-to-one mapping. One can imagine that one halo can associate with more than one clusters. In a cluster detection process, different cluster finding algorithms working at different wavelengths have different "definitions" about the observable features of clusters. Therefore, clusters detected in different wavelengths using different algorithms can be very different. For example, clusters detected in the X-ray and microwave band essentially map out the over-dense intracluster medium while clusters detected optically map the over-density of the galaxy distribution. For certain fully relaxed systems, we can expect some concordance among them. But in general,

there may be a big offset among them.

If we impose the one to one matching of clusters and halos, we need to percolate close pairs of clusters to form bigger ones. In this way, we hide the bias problem by increasing the scatter of mass observable relation. If we want to obtain tighter cosmological constraints, we need to model this bias so that we can decrease the scatter.

The second challenge comes from the masses of the clusters or the underlying halos. We need to reliably determine the statistical behavior of masses in order to constrain cosmological parameters. However, the masses are not directly observable. We need to choose certain mass proxies that can be detected directly. Though the final calibration between mass proxies and masses requires independent measurements of masses, we can still get some rough estimates about the mass scatter from various mass proxies based on simulations.

For clusters detected in the X-ray band, the X-ray luminosity and temperature are correlated to the mass of the underlying halos. The mass scatter [2] from X-ray luminosity L_X is $\sim$40% for fixed L_X (Stanek et al., 2006) while the mass scatter achieved from the X-ray temperature (T_X) is $\sim$15% at fixed T_X (Vikhlinin et al., 2006). For clusters detected from SZE, the Compton Y parameter integrated within a radius enclosing a contour of 500 times the critical density correlates strongly with the cluster mass, leading to a scatter of $\sim$10%-15% (Rudd, 2007). For optically selected clusters, the masses for individual clusters are not easily attainable. However, by stacking clusters of similar optical richness, the masses can be obtained through weak lensing analysis (Johnston et al., 2007; Sheldon et al., 2007a) and velocity dispersion (Becker et al., 2007). The corresponding scatter in mass for a given richness bin are $\sim$13% and $\sim$60% for weak lensing and velocity dispersion analysis respectively. For a review on different mass proxies, refer to Voit (2005).

Note that since large mass scatter will degrade our cosmological constraints, we need to pin down the scatter as much as possible. Clearly, detecting clusters at only one wavelength will not be very effective. Therefore, multi-wavelength synergy is

[2]The scatter in mass quoted here is the scatter in log mass, i.e. $\sigma_{\ln M}$.

needed for narrowing down the mass-observable scatters. Suppose we solve all the above challenges to a satisfactory level. How do we relate the halos to cosmological parameters? In the following subsections, I will outline the theoretical connections among halo abundance, clustering and cosmological parameters.

1.3.2 Theoretical Connection I: Number Counts

N-body simulations based on a given set of cosmological parameters can predict the comoving number density of dark matter halos above certain mass thresholds. This means that if we can measure/count the comoving number density of dark matter halos above a certain mass, we can reverse the process to constrain cosmological parameters. To calculate the comoving density, we need two pieces of information: i) The direct number counts of halos above a certain mass; ii) The comoving volume corresponding to the halos. The first part is a result of the structure growth in a given cosmology while the second part results from the geometry of that cosmology.

Comoving volume:

The differential co-moving volume in the Friedman-Robertson-Walker Universe can be calculated through the following equation:

$$dV_c = \frac{c(1+z)^2 D_A(z)^2}{H_0 E(z)} d\Omega dz \tag{1.1}$$

where D_A is the angular diameter distance

$$D_A(z) = \frac{c}{H_0(1+z)} \int_0^z \frac{dz'}{E[z']} \tag{1.2}$$

For the special case of a FRW universe with dark energy equation of state can be written as $w(z) = p(z)/\rho(z)$, the $E(z)$ is defined as:

$$E(z) = \frac{H(z)}{H_0} = \left[\Omega_{k,0}(1+z)^2 + \Omega_{R,0}(1+z)^4 + \Omega_{M,0}(1+z)^3 + \Omega_{DE} \exp\{\int_0^z \frac{-3[1+w(z')]dz'}{1+z'}\}\right]^{1/2} \tag{1.3}$$

Comoving number density:

The comoving number density of dark matter halos in a given cosmology can be

obtained from N-body simulations. The general form of the differential comoving number density is:

$$dn = f(M, z)dM \tag{1.4}$$

Where $f(M, z)$ is the so called mass function. Based on N-body simulation, an empirical fitting formula for $f(M, z)$ is obtained by Jenkins et al. (2001):

$$f(M, z) = \frac{dn}{d\ln M} = 0.3\frac{\rho_M}{M}\frac{d\ln\sigma^{-1}}{d\ln M}\exp[-|\ln\sigma^{-1} + 0.64|^{3.82}] \tag{1.5}$$

where σ is the variance of the mass density perturbation field smoothed using a window function.

$$\sigma^2(M, z) = \frac{G^2(z)}{2\pi^2}\int_0^\infty k^2 P(k)W^2(k, M)dk \tag{1.6}$$

and $G(z)$ is the growth factor of linear perturbation. It is the solution of the the 2nd order differential equation:

$$G'' + \left[\frac{5}{2} - \frac{3}{2}w(z)\Omega_{DE}(z)\right]G'' + \frac{3}{2}[1 - w(z)]\Omega_{DE}(z) = 0 \tag{1.7}$$

where ' denotes the derivative with respect to $\ln a$ and a is the scale factor. The initial conditions are: $G(0) = 1$ and $G'(0) = 0$ (Hu, 2005). In a ΛCDM universe, the growth function is in the following simple form:

$$G(z) = \frac{5}{2}E(z)\int_z^\infty \frac{1 + z'}{E(z')^3}dz' \tag{1.8}$$

$W(kR)$ is the Fourier transform of the window function

$$W(k, M) = \frac{3[\sin(kR) - kR\cos(kR)]}{kR} \tag{1.9}$$

where

$$M = \frac{4\pi R^3}{3}\rho_M = \frac{4\pi R^3}{3}\Omega_{M,0}\rho_{c,0} = \frac{H_0^3\Omega_{M,0}R^3}{2G_N} \tag{1.10}$$

and $P(k)$ is the power spectrum of the matter field. It can be related to the initial power spectrum P_{ini} as (Eisenstein & Hu, 1998)

$$P(k) = P_{ini}T^2(k)\left[\frac{G(z)}{G(0)}\right]^2. \tag{1.11}$$

$T(k)$ is the linear transfer function, which can be fitted as

$$T(k(q)) = \frac{L(q)}{L(q) + C(q)q^2} \tag{1.12}$$

where

$$L(q) = \ln(e + 1.84q) \tag{1.13}$$

$$C(q) = 1.44 + \frac{325}{1 + 60.5q^{1.11}} \tag{1.14}$$

and q scales to k as

$$q = \frac{k}{\Gamma_{eff}}(T_{CMB}/2.7K)^2 \tag{1.15}$$

with

$$\Gamma_{eff} = \Omega_M h\left[1 - 0.38\ln(431\Omega_M h^2)\frac{\Omega_b}{\Omega_M} + 0.38\ln(22.3\Omega_M h^2)(\Omega_b/\Omega_M)^2\right] \tag{1.16}$$

Number Counts:

With the above formulae, if we can measure the numbers of halos per solid angle and redshift slice, we can connect them to the cosmological parameters immediately by:

$$\frac{dN(z)}{dzd\Omega} = \frac{cd_A^2(1+z)^2}{H(z)}\int_{M_{lim}(z)}^{\infty} d\ln M f(M,z) \tag{1.17}$$

Where M_{lim} is the mass threshold. So far, we have demonstrated how the number counts of dark matter halos relate to the underlying cosmological parameters. How-

ever, the number count only contains information about the first moment of the halo distribution. The second moment, i.e. the spatial clustering, can relate to cosmological parameters in a different manner.

1.3.3 Theoretical Connection II: Clustering

Halo clustering can be described by the two point correlation $\xi_{hh}(r)$ or its Fourier counterpart power spectrum $P_{hh}(k)$. The $P(k)$ in Eq.(1.11) is the power spectrum of the matter field, not the halo distribution. To first order, the power spectrum of the halo distribution relates to the matter power spectrum by multiplication of a bias via the following equation (Majumdar & Mohr, 2004)

$$P_{hh}(k,z) = b_{eff}^2(z)P(k,z) \tag{1.18}$$

where b_{eff} is the bias defined by

$$b_{eff}(z) = \frac{\int dM b(M,z) f(M,z)}{\int dM f(M,z)} \tag{1.19}$$

with the bias given by (Sheth & Tormen, 1999)

$$b(M,z) = 1 + \frac{a\delta_c^2/\sigma^2 - 1}{\delta_c} + \frac{2p}{\delta_c[1 + (a\delta_c^2/\sigma^2)^p]} \tag{1.20}$$

where $a = 0.75$, $p = 0.3$ and $\delta_c = 1.69$. On the other hand, we can measure the power spectrum of the halos by inversion of the point correlation function or by direct measurements. Then, we can go on to constrain the cosmological parameters.

1.3.4 Counts in Cell and Self-Calibration

The equations outlined in the previous sections are basic recipes that we can use to relate observational data and model parameters. However, measuring the power spectrum from the Fourier inverse of the two point correlation is often numerically noisy. Fortunately, we do not have to measure the power spectrum to compare it with theory because the variance of the number counts are related to the power spectrum (Hu & Kravtsov, 2003)

$$< n_i n_j > -\bar{n}^2 = \bar{n}^2 b^2 \int \frac{d^3k}{(2\pi)^3} W_i(k) W_j(k) P(k) \tag{1.21}$$

where n_i is the number of halos in cell i and $\bar{n}$ is the mean number counts of halos in each cell. Clearly, measuring the variance of the number counts is a lot more tractable than measuring the power spectrum directly. Note that the mean number counts $\bar{n}$ can be calculated from Eq.1.17. For more details about how to execute these calculations in realistic way, refer to Cunha (2008). In this way, we can integrate the number counts and clustering together to compare with the counts and variance of the counts of halos in each cell. This is in fact the practical way to relate data and model parameters.

However, it is worth noting that such a scheme will lead to some loss of information because of the cell division. Without prior knowledge of the optimal cell division scheme, some useful information is lost. Therefore, although the counts in cell scheme facilitates the comparison between data and theory, its application requires further investigation into the optimal cell division for a given survey. If we try the direct method instead of counts in cell, we will suffer from the noisy measurements of the power spectrum. There is a trade off in terms of errors using different methods. But a comparison should be good for better cosmological constraints.

From the above picture, one can see that the mean number counts and their variance depend on cluster mass in different ways. If we combine the two pieces of information together, we can self-calibrate the mass-observable relations, leading to tighter constraints on the cosmological parameters (Majumdar & Mohr, 2004; Hu, 2003; Lima & Hu, 2004). Figure 1.4 shows the improvement of cosmological constraints using self-calibration.

1.3.5 Cluster Cosmology in Action

Given all the complexities and uncertainties, some initial work on cosmological constraints using optically selected clusters has been carried out based on the maxBCG cluster catalog (Rozo et al., 2007b,a, 2009). The constraints on σ_8 and Ω_M are shown in Figure 1.5.

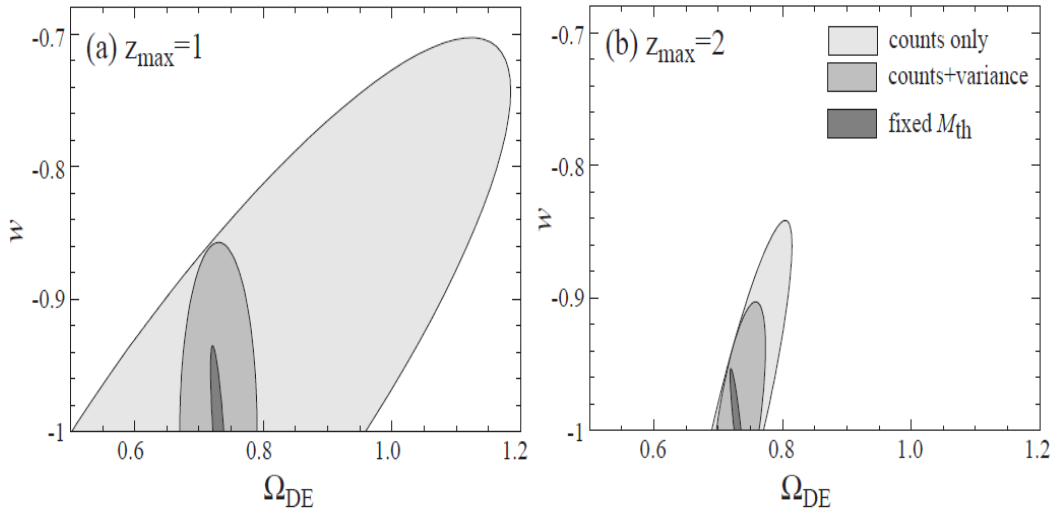


Figure 1.4. The cosmological constraints when using self-calibration. (a) is for a survey with maximum redshift $z_{max} = 1.0$ and (b) is for $z_{max} = 2.0$. Fixed M_{th} means we have perfect knowledge about the mass of the clusters. The plots are reproduced from (Lima & Hu, 2004).

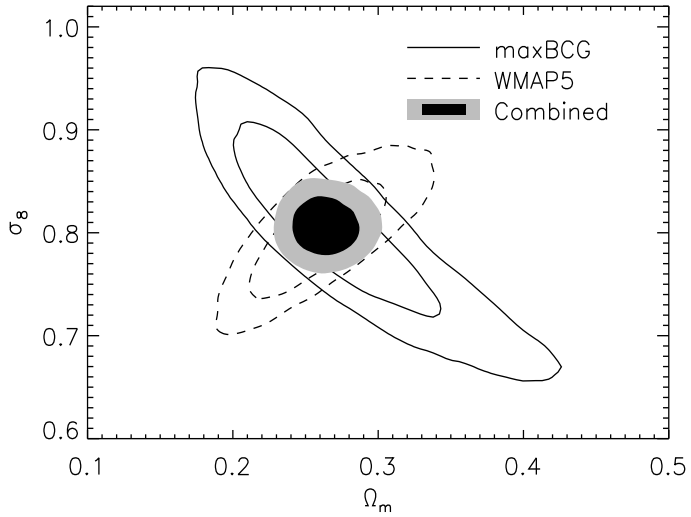


Figure 1.5. Constraints on σ_8 and Ω_M from maxBCG and WMAP5 by assuming a spatially flat ΛCDM cosmology. Contours show the 68% and 95% confidence regions for maxBCG (solid), WMAP5 (dashed), and the combined results (filled ellipses). The thin axis of the maxBCG-only ellipse corresponds to $\sigma_8(\Omega_m/0.25)^{0.41} = 0.832 \pm 0.033$. The joint constraints are $\sigma_8 = 0.807 \pm 0.020$ and $\Omega_m = 0.265 \pm 0.016$ (one-sigma errors)(Rozo et al., 2009).

No doubt, with more clusters and better estimated masses based on deeper and multi-wavelength data, we can further improve the above constraints on cosmological parameters. Particularly, clusters provide a an approach to constrain cosmological parameters, which is complementary to SNeIa, BAO and CMB methods. Building a large galaxy cluster catalog is an important step towards this goal. In the next section, we will describe the major challenges of cluster detection in the optical band.

1.4 Optical Galaxy Cluster Detection

In order to use galaxy clusters for precision cosmology, a large cluster catalog with high purity, completeness and well calibrated masses is indispensable. Galaxy clusters can be detected at different wavelengths, such as the x-ray band, optical band and microwave band, each yielding a unique mass-observable relation. Looking for galaxy clusters in optical data enjoys high signal to noise, a large volume of available data, less expensive experiment setups and fairly accurate redshift information. In this section, we will briefly review the various optical cluster detection algorithms and their pros and cons.

1.4.1 Generic Data Clustering Analysis

Before we delve into optical cluster detection, we will first look at the generic data clustering analysis. Data clustering is a very generic problem in data mining, machine learning and pattern recognition. It plays an important role in artificial intelligence. Generally speaking, there are two basic types of clustering analysis methods: partitional and hierarchical. In partitional clustering analysis, objects are partitioned into non-overlapping groups and each object belongs to only one group. The number of the groups are determined by minimizing cost functions that reflect the clustering structures. In hierarchical analysis, objects are partitioned into nested groups based on some linkage cuts in a progressive way and then organized as a hierarchical tree. In Fig. 1.6, we show the two types of clustering analysis for some artificial data. It is clear that both methods require a metric to allow distance between data points to be calculated. Many methods also require that the data points belong to one of

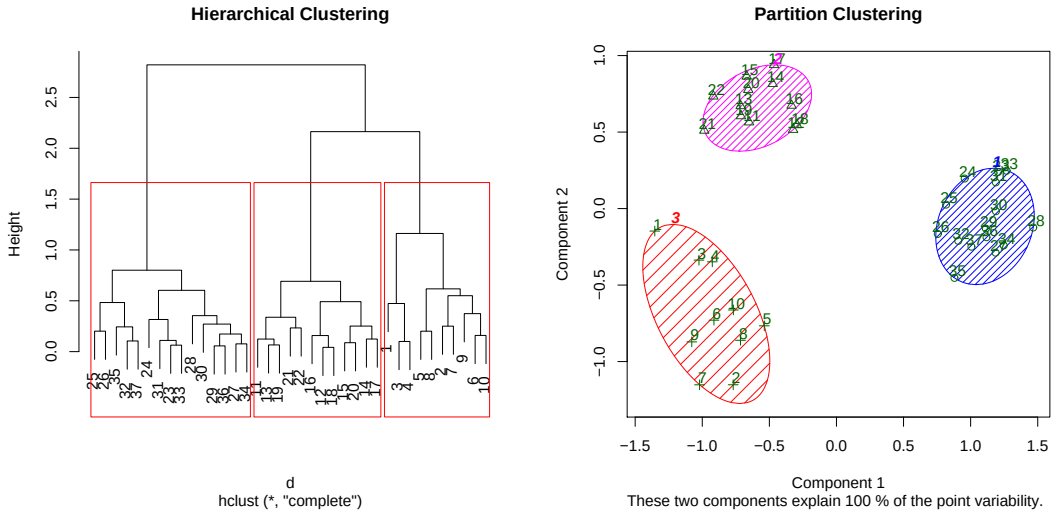


Figure 1.6. The hierarchical (left) and partitional (right) clustering analyses. The data are generated from $x \sim N(1, 0.3^2), N(0, 0.2^2), N(-1, 0.2^2)$ and $y \sim N(1, 0.3^2), N(0, 0.2^2), N(1.5, 0.2^2)$. For the hierarchical clustering, we the clusters are divided based on the complete linkage, i.e. the maximum distance between elements of each cluster. For the partitional clustering, we use kmean algorithm. In both case, we use the Euclidean distance.

.

the clusters. In galaxy clustering analysis, we do not have precise distance measures. Also, only a small fraction of all galaxies are clustered and most of them are not clustered. These features make galaxy cluster detection a lot more complicated.

1.4.2 Galaxy Cluster Detection and De-projection

The fundamental challenge for optical galaxy cluster selection is that we are trying to detect three dimensional clusters with precise information only in two dimensions (RA/DEC). The huge uncertainties in galaxy positions along the line of sight lead to projections, which deteriorate our richness estimates for all clusters and confuse cluster detection as we go to lower richness systems. In Figure 1.7, we show a somewhat exaggerated situation to illustrate the effects of projection on cluster detection. We use color to represent the third dimension information about the data points. If we have different kinds of information on the third dimension, i.e. different colors, we will have very different clustering results. In real optical cluster detection, though we can get some information about the 3D clustering based on the clustering in the projected RA/DEC plane, we need additional information to accurately recover 3D

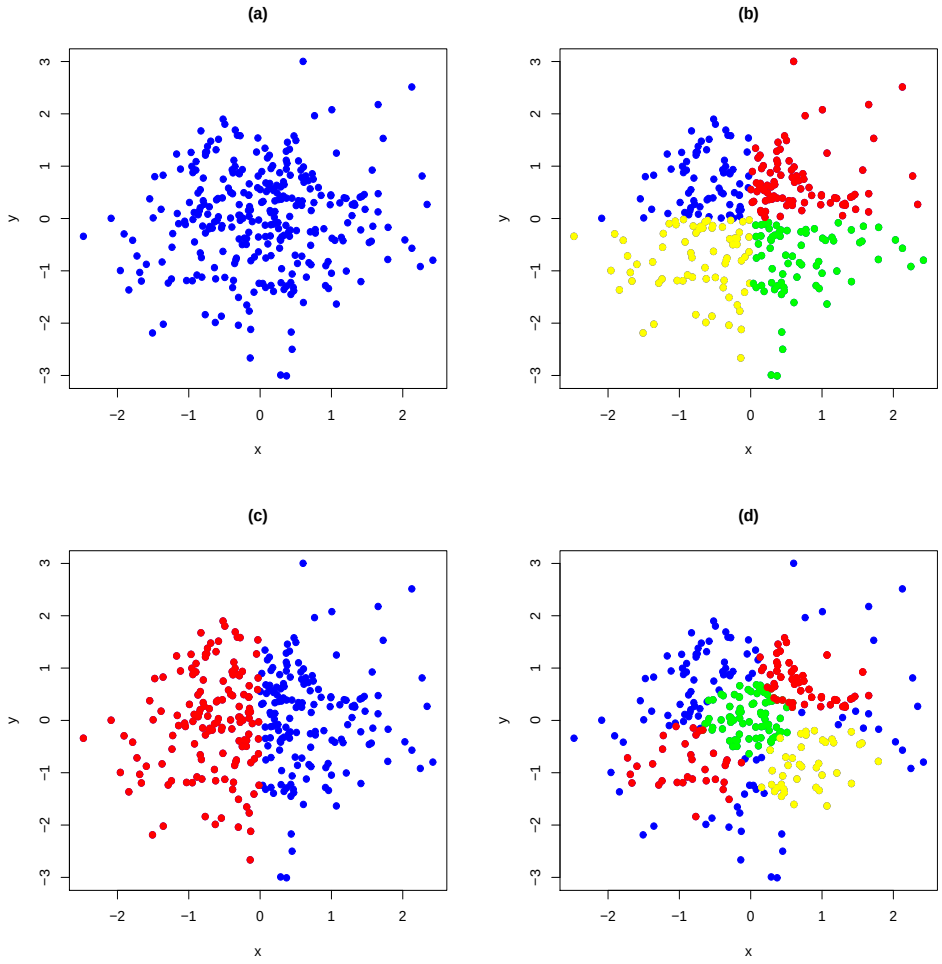


Figure 1.7. This plot is an exaggerated example of the effect of projection on clustering analyses. Each dot has 3 dimensional information (x, y, color). If we change the color information, any clustering algorithm will identify different objects.

clustering. Therefore, every optical cluster finding algorithm needs to effectively remove the projected field galaxies before calculating the over-density in the RA/DEC plane.

Historically, the ability to locate the positions of galaxies along the line of sight is limited by the technology available to that specific age. Abell (Abell, 1957) built his cluster catalog by visually examining black and white photographs of the sky. The clusters he identified are projected over-densities in the RA/DEC plane with certain magnitude cuts. In this case, the magnitude cuts play the role of de-projecting the field galaxies along the line of sight. However, due to the broad luminosity function

of galaxies, magnitude can only roughly tell how far away the galaxy is. It becomes less effective at resolving galaxy positions along the line of sight at higher redshift. Before the advent of multi-color and wide-field digital imaging, magnitude was the only way to de-project galaxies along the line of sight. This severely limited optical cluster detection in the past.

Over the past 30 years, various algorithms for optical galaxy cluster detection have been developed based on photometric data (Huchra & Geller, 1982; Davis et al., 1985; Shectman, 1985; Efstathiou et al., 1988; Couch et al., 1991; Lidman & Peterson, 1996; Postman et al., 1996; Kepner et al., 1999; Annis et al., 1999; Gladders & Yee, 2000, 2005; Gal et al., 2000, 2003; Kim et al., 2002; Goto et al., 2002; Ramella et al., 2002; Lopes et al., 2004; Botzler et al., 2004; Berlind et al., 2006; Koester et al., 2007b; Li & Yee, 2008). For a thorough recent review, please see Gal (2006). Though these methods differ in many detailed respects, we can classify them according to the de-projection methods they employ. In Table. 1.4.2, we list the major cluster finding algorithms and the de-projection methods used.

The de-projection method used in a specific algorithm is affected by the properties of the data available when the algorithm was developed. For early times, only single band data were available and therefore the major de-projection methods were all essentially magnitude based. Some of them use magnitude directly while others use the photometric redshift (photoz) calculated from the magnitude. Since both magnitude and the photoz derived from it are poor indicators of the positions of galaxies along the line of sight, these de-projection methods don't work well for the non-local Universe. Though they are quite effective for detecting massive clusters, they cannot maintain good purity and completeness for clusters with lower/intermediate richness across a wide redshift range. Moreover, the projection also creates large scatters in the richness-mass relation derived from these methods.

With the advent of modern multi-band digital imaging technology, large scale CCD imaging surveys greatly alleviate the projection effects that plagued optical galaxy cluster detection. In a precise multi-band sky survey, we have magnitude information from more than one band, allowing better reconstruction of the galaxy

Algorithm	Type of data applied	De-projection method
Percolation[a]	Single band/Simulation	Magnitude/photoz
Smoothing Kernels[b]	Single band	Magnitude
Adaptive Kernel[c]	Single band	Magnitude
Matched Filter [d]	Single band	Magnitude
Hybrid and Adaptive Matched Filter[e]	Single band/multi-band	Magnitude/photoz
Voronoi Tessellation[f]	Multi-band	Colors
Cut-and-Enhance [g]	Multi-band	Colors
Modified Friends of Friends[h]	Multi-band	Photoz
C4[i]	Multi-band	All Colors
Percolation with Spectroscopic redshift[j]	Multi-Band	Spectroscopic Redshift
Cluster Red Sequence[k]	Multi-band	Red sequence
MaxBCG [l]	Multi-band	Red sequence
GMBCG	Multi-band	Red sequence

Table 1.1. Summary of optical cluster finding algorithms

[a] Huchra & Geller (1982); Davis et al. (1985); Efstathiou et al. (1988); Ramella et al. (2002)

[b] Shectman (1985)

[c] Gal et al. (2000, 2003)

[d] Postman et al. (1996)

[e] Kepner et al. (1999); Kim et al. (2002); Dong et al. (2008)

[f] Kim et al. (2002); Lopes et al. (2004)

[g] Goto et al. (2002)

[h] Li & Yee (2008)

[i] Miller et al. (2005)

[j] Berlind et al. (2006)

[k] Gladders & Yee (2000, 2005)

[l] Annis et al. (1999); Koester et al. (2007a,b)

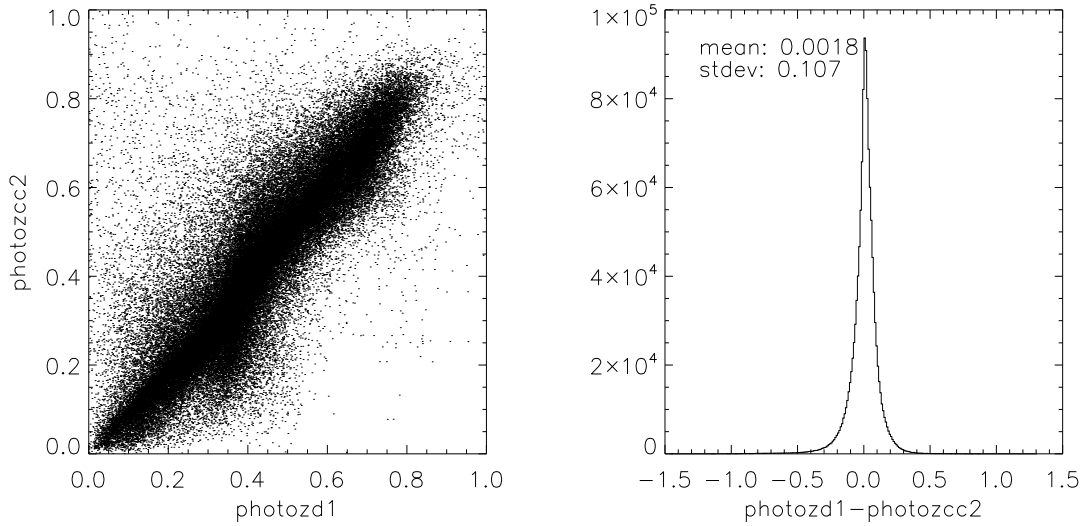


Figure 1.8. We show the scatter of the two photoz estimators (photozcc2 and photozd1) from the well tested neural network algorithms for the SDSS DR6 data (Oyaizu et al., 2007). Though the algorithm is already well tuned, the two photozs still lead to a photoz difference of 0.1.

spectra. Even the crude Spectral Energy Distribution (SED) information provided by colors provides very effective information for locating galaxies along the line of sight.

The existence of 4000 Å break in the spectra of red elliptical galaxies makes the color that contains the break strongly correlate with redshift. As a result, the clustered galaxies are also tightly clustered in the corresponding color space, forming the so called red sequence. They have narrow color scatter (~ 0.05) and a slightly tilted color magnitude relation, which has been measured precisely in Hao et al. (2009). We leave a more detailed introduction about red sequence to Chapter 2.

There are basically two ways to de-project galaxies using the multi-color data: (1) using multi-color data to obtain photometric redshifts or (2) using the color clustering of red sequence galaxies directly. The first approach is very straightforward. There are well tested Neural Network algorithms to assign photozs to galaxies based on the multi-color data and these lead to errors ~ 0.03 for bright galaxies (Oyaizu et al., 2007). However, the photozs are not perfect. In particular, the error distribution for photozs are often non-Gaussian, and limitations on training sets lead to bias in photoz estimation for fainter and bluer galaxies.

In Figure 1.8, we compare photozs based on two estimators from (Oyaizu et al., 2007). The photozd1 is obtained by training only on the magnitudes while photozcc2 is obtained by training only on colors. The two photozs are in agreement by 0.1. Even if the photoz precision for all galaxies was as good as 0.03, it is still not sufficient for cluster member selection, because the member galaxies' equivalent redshifts difference should be about ± 0.003 assuming a velocity dispersion of 900 km/sec. Therefore, though it is a lot better than the magnitude based de-projection used in the early days, using the photoz for de-projection has important limits, especially when training sets remain limited.

As an alternative, we might stay closer to the data and look for clustering directly in color space. Red sequence galaxies in clusters display a scatter in color ~ 0.05. This is comparable to the measurement errors on the colors from SDSS data. Therefore, much of the information available to de-project field galaxies is already there in color space. At this point, we need to address a perplexing question before we proceed. Why do colors do better than photozs that are essentially derived from colors? In particular, photozs are obtained by using multi-color/magnitudes while the ridgeline color is only one color. This would suggest that photozs should do better than red sequence colors.

To resolve this puzzle, we need to clarify two points. First, it is not always true that more information leads to better results. When you have more information, some of it is useful while other parts may be useless or even harmful. For the galaxy photoz case, most of the leverage for obtaining photometric redshifts comes from the 4000 Å break as revealed in the galaxies' colors. Other colors provide little information while adding noise. Therefore, when using all color/magnitude, you essentially need to supply additional information about which color is more informative than the others. In practice, that means appropriate weights should be assigned to each color/magnitude before combining them to get photozs. Second, there is a fundamental limit on the precision of photozs from the all the machine learning algorithms. They cannot achieve better results than the scatter existing in the training set. Especially, when the training set is biased, the results can be significantly skewed for those ob-

jects that are not well represented by the training set. For example, the faint blue galaxies are not well represented in most spectroscopic training sets, which will make their photozs a lot worse than those for bright red galaxies.

For red sequence colors, we have two additional pieces of information. One is spatial clustering, because we normally look for red sequence around a cluster. Another is that we can separate red sequence galaxies from the blue galaxies in advance of the clustering analyses. For this combination of reasons, de-projection using red sequence colors out-performs de-projection using photozs.

The major advantage of using red sequence color for de-projection is that we can push the cluster detection to lower richness limits a lot better than we can do using photozs. For very big clusters, one can find them with any means. But for lower richness systems, appropriate de-projection is crucial for detection and richness measurement. For cosmological constraints, clusters with a wide mass and redshift range will provide substantially more leverage on cosmological parameters.

CHAPTER 2

E/S0 Ridgeline Measurements Using Error Corrected Gaussian Mixture Model

2.1 Overview of Red Sequence Galaxies

Modern multi-band digital imaging technology allows precise characterization of galaxy color. The red-sequence is ubiquitous in the galaxy population (Hogg et al., 2004), and the close mapping between apparent color and redshift for these galaxies provides a powerful tool for reducing the effects of projection.

The predominantly red, bright, passively evolving red sequence, or "E/S0 ridgeline" (Visvanathan & Sandage, 1977; Annis et al., 1999) found in the cores of clusters of varied richness up to at least $z \sim 1.4$ (Bower et al., 1992; Smail et al., 1998; van Dokkum et al., 1998; Barrientos, 1999; Blakeslee et al., 2003; Mullis et al., 2005; Eisenhardt et al., 2005; De Lucia et al., 2007) provide an efficient means for cluster detection, and have become an integral part of modern cluster cosmology. The red sequence itself is ubiquitous in the galaxy population (Renzini, 2006, e.g.), and in clusters red sequence galaxies dominate the bright end of the cluster luminosity function (Sandage et al., 1985; Barger et al., 1998). These galaxies are extremely tightly clustered in color space, containing old populations of stars whose observed color varies smoothly with redshift (e.g. Gladders & Yee, 2000). The pervasiveness of this phenomenon in clusters minimizes projection effects, enabling efficient optical cluster detection and providing accurate photometric redshifts (Gladders & Yee, 2000; Koester et al., 2007b). Simple counting of photometrically identified cluster red sequence galaxies (Koester et al., 2007a, e.g) has also been shown to be an effective proxy for cluster mass (Becker et al., 2007; Sheldon et al., 2007b; Johnston

et al., 2007), with more sophisticated applications yielding improvements in richness as a cluster mass proxy (Rozo et al., 2008a). In the era of precision cosmology, the extent to which the red sequence can be exploited for cluster cosmology depends how accurately its characteristics can be measured at a given redshift.

In addition to its relevance to cluster cosmology, the red sequence plays an important role in constraining the complex physical processes that drive galaxy formation and evolution at all scales. At the field scale this includes measurements of the red galaxy luminosity function (Wake et al., 2006; Faber et al., 2007), the clustering of red galaxies in various environments (Zehavi et al., 2005; Coil et al., 2008), and color-magnitude relations of spectroscopically (Cool et al., 2006) and morphologically identified early-type galaxies. The high density environments of clusters of galaxies are dominated by red sequence galaxies. They dominate in the cores of rich clusters to at least $z \sim 1$ forms the basis for various monolithic collapse scenarios (e.g. Bower et al., 1992; Blakeslee et al., 2003; Mei et al., 2009). Faber et al. (2007) summarize some of these results to fill out a picture of galaxy formation that includes a mechanism for the formation of the red sequence.

In color-magnitude space, the red sequence is typically characterized by slope, zero point, and scatter. Various models posit that in the rest frame, the scatter in the red sequence is driven primarily by age effects, its slope is a manifestation of the mass-metallicity relation, and the zero point is set by combination of age and mass-metallicity differences (e.g. Bernardi et al., 2005; De Lucia et al., 2007; Faber et al., 2007)

Studies of the cluster red sequence have been undertaken in many cases by simply measuring the photometric color-magnitude relation (e.g. López-Cruz et al., 2004; De Lucia et al., 2007), supplemented with HST morphological information (e.g Gladders et al., 1998) and sometimes with spectroscopy. Extra morphological and spectroscopic data allow precise separation of E and S0-types from the rest of the galaxy population, as well as refined identification of cluster members (Blakeslee et al., 2003; Mei et al., 2009). The situation also benefits significantly from precise color measurements afforded by deep, CCD-based imaging (e.g van Dokkum et al., 1998). In the litera-

ture, the red sequence has been measured with various levels of scrutiny in dozens of individual clusters.

In the past several years, researchers have turned to the considerable resources of the Sloan Digital Sky Survey (SDSS) and similar wide field surveys to probe the red sequence of field galaxies and the red sequence of ellitptical galaxies in various environments (Hogg et al., 2004; Bernardi et al., 2005, 2006; Cool et al., 2006). These studies have included both spectroscopically and morphologically identified red galaxies at $z \sim 0.1$, and that aim to constrain galaxy evolution scenarios to the cosmologically-relevant luminous red galaxy (LRG) samples extending to $z \sim 0.6$ (e.g. Cool et al., 2006).

The maxBCG cluster catalog (Koester et al., 2007a) is the largest optical galaxy cluster catalog based on the photometric data from Sloan Digital Sky Survey. The clusters are identified using the maxBCG algorithm, which is a variant of a matched filter algorithm with the inclusion of color filters based on the red sequence galaxy (Koester et al., 2007b). The clusters in the maxBCG catalog range from redshift 0.1 to 0.3 in an approximately volume limited way. With the maxBCG cluster catalog, we are positioned to use the SDSS to make among the most statistically robust photometric measurements of the cluster red sequence, using nearly 14,000 clusters between $0.1 \leq z \leq 0.3$. In this chapter we focus on the slope and scatter of the red sequence. While the maxBCG sample affords exquisite precision, we show clearly the systematic effects photometric errors can have on the measurement of the underlying slope and scatter of the red sequence, and introduce a method for properly handling these effects. This method, based on an Error-Corrected Gaussian Mixture Model (ECGMM, see section 2.3), reliably recovers the properties of the ridgeline by taking measurement errors into account. After presenting the method, we describe its application to the measurement of maxBCG clusters. Of particular relevance to cluster cosmology are the observed mean, scatter, and slope of the E/S0 ridgeline for all maxBCG clusters. These results are presented, along with a discussion of observed trends with redshift.

2.2 Intrinsic Properties of the Red Sequence Ridgeline

The red sequence ridgeline in galaxy clusters shows that cluster galaxies condense in color space in addition to real space. The old stellar populations which dominate emission from early-type galaxies give rise to remarkably similar galaxy colors. The close mapping between galaxy color and redshift for these galaxies is primarily a result of the restframe 4000 Å break in their spectra, and thus the most informative color for cluster finding at a given redshift depends on where the 4000 Å break resides. For the SDSS filter sets, the 4000 Å break will be well within the g band as long as the redshift is below 0.35. So, the the most informative SDSS color for the clusters in the maxBCG catalog is $g - r$.

In the vicinity of a detected cluster, there are both cluster member galaxies and projected field galaxies. Red sequence galaxies form a part of the member population, whose colors are clustered tightly and can be approximated with a Gaussian distribution with narrow width. On the other hand, the field galaxy and blue member galaxy colors are not tightly clustered and could be approximated by a Gaussian distribution with a broader width[1]. The problem of separating the ridgeline from the field can be specified as following: What are the two Gaussian components (one for the ridgeline and one for the field) that represent the color distribution in the vicinity of a galaxy cluster? If this double Gaussian is an adequate model for describing the overall color distribution, the one dimensional Gausian Mixture Model (GMM) is well suited to the problem.

In the traditional applications of GMM, measurement errors are not considered. In our case, there are non-negligible measurement errors associated with the galaxy colors. We are interested in measuring the intrinsic color scatter of cluster members, absent contamination by the increasing measurement errors of faint galaxies. Without accounting for these errors, one can expect observed color scatter to increase as the

[1]There are complicated situations where the distribution in color space is not simply unimodal or bimodal, for example when two clusters are seen in projection. For maxBCG clusters, the redshift range is only from 0.1 to 0.3. Since an 0.4 L* magnitude cut is also applied, the chance of two or more overlapped clusters is low. Therefore, a unimodal or bimodal distribution in color space is a good approximation.

measurement errors become larger. While the intrinsic color scatter may increase as redshift increases (because the 4000 Å line break is shifting toward r band and making the $g-r$ color less discriminative.), measurement errors may make us overestimate the increase in intrinsic scatter with redshift. To avoid this problem, we include measurement error into our mixture model likelihood function. We will call this an Error-Corrected Gaussian Mixture Model (ECGMM) and derive the corresponding Expectation Maximization (EM) recursive relation in the following section.

2.3 Error Corrected Gaussian Mixture Model

2.3.1 The Expectation Maximization Algorithm

In this section we present the details of the ECGMM used in this thesis for ridgeline studies. In what follows, we describe how to fit a multicomponent Gaussian mixture model to a one dimensional distribution of data with both intrinsic scatter and measurement error. Our method is a extension of the traditional expectation maximization method for GMM (Dempster et al., 1977).

Assume the data are to be modeled by a mixture of N Gaussians fit to the distribution of M data points. Subscript i cycles through N and j cycles through M. We use μ_i, σ_i and w_i to denote the location, width and weight of each Gaussian component. y_j and δ_j denote the data points and their measurement errors. For brevity, we denote the parameters (μ_i, σ_i and w_i) collectively by θ. The likelihood of the parameters given the data and measurement errors is then:

$$\mathcal{L}(\theta|y) = \prod_{j=1}^{M} \left[\sum_{i=1}^{N} \frac{w_i}{\sqrt{2\pi(\sigma_i^2+\delta_j^2)}} \exp\left(-\frac{(y_j-\mu_i)^2}{2(\sigma_i^2+\delta_j^2)} \right) \right] \tag{2.1}$$

The optimal parameters θ can be estimated by maximizing the above likelihood function. The Expectation Maximization algorithm provides an efficient way to get the maximum likelihood estimators in such a setting. To utilize this, we need to introduce a hidden variable, z_j, which tells which Gaussian component the data point y_j is sampled from. In our case, unlike the standard EM prescription, we have non-negligible measurement errors present. In the following, we will show the derivation

Variable notation	Meaning
y_1,..., y_j, ..., y_m:	Colors of BCGs and member galaxies.
z_1,..., z_j, ..., z_m:	Hidden variables that tell which Gaussian component the y_j is sampled from.
δ_1,..., δ_j,..., δ_m:	Measurement errors for every y_j.
μ_1, ..., μ_i, ..., μ_n:	Mean of each Gaussian component.
σ_1, ..., σ_i, ..., σ_n:	Width of each Gaussian component.
w_1, ..., w_i, ..., w_n:	Weights of corresponding Gaussian components.

Table 2.1. The notations used in our derivation of ECGMM algorithm

of the iteration relation of the parameters that can maximize the likelihood 2.1.

We introduce the notations as shown in Table.2.1. The parameter t represents the t^{th} iteration. The likelihood of the parameters given the data after convolving with the measurement errors is given by Eq.2.1. The optimal parameters can be obtained by maximizing the above likelihood. However, if we introduce hidden variables, z, that tell us which Gaussian component the y_j is sampled from, then the maximization process is significantly simplified. The corresponding pdf of data given z and θ is

$$p(y|z_j = i, \theta^{(t)}) = \prod_{j=1}^{M} p(y_j|z_j = i, \theta_i^{(t)}) = \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi(\sigma_i^{(t)2} + \delta_j^2)}} \exp\left[-\frac{(y_j - \mu_i^{(t)})^2}{2(\sigma_i^{(t)2} + \delta_j^2)} \right] \tag{2.2}$$

The weight of each Gaussian Component in the mixture is given by $w_i = p(z_j = i|\theta)$. The estimation of hidden variable can be related to Eq.2.2 by the Bayes formula as follows:

$$p(z_j = i|y_j, \theta^{(t)}) = \frac{p(z_j = i, y_j|\theta^{(t)})}{p(y_j|\theta^{(t)})} = \frac{p(y_j|z_j = i, \theta^{(t)})p(z_j = i|\theta^{(t)})}{\sum_{i=1}^{N} p(y_j|z_j = i, \theta^{(t)})p(z_j = i|\theta^{(t)})} \tag{2.3}$$

The EM algorithm iteratively update the parameters θ by maximizing the expected log likelihood

$$Q(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{M} p(z_j = i|y_j, \theta^{(t)})\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_i^2+\delta_j^2) - \frac{(y_j-\mu_i)^2}{2(\sigma_i^2+\delta_j^2)} + \ln p(z_j = i|\theta^{(t)})\right] \tag{2.4}$$

under the constraint $\sum_{i=1}^{N} p(z_j = i|\theta^{(t)}) = 1$. Using the Lagrange Multiplier approach, we redefine

$$\tilde{Q}(\theta) = Q(\theta) - \lambda\left[\sum_{i=1}^{N} p(z_j = i|\theta^{(t)}) - 1\right] \tag{2.5}$$

with λ as the multiplier.

$$\frac{\partial\tilde{Q}(\theta)}{\partial\mu_i} = \sum_{j=1}^{M}\left[p(z_j = i|y_j, \theta^{(t)})\left(\frac{y_j-\mu_i}{\sigma_i^2+\delta_j^2}\right)\right] = 0 \tag{2.6}$$

From Eq.2.6, we can arrive at the following recursive relation for μ:

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^{M} y_j p(z_j = i|y_j, \theta_i^{(t)})/(1+\delta_j^2/\sigma_i^{(t)2})}{\sum_{j=1}^{M} p(z_j = i|y_j, \theta_i^{(t)})/(1+\delta_j^2/\sigma_i^{(t)2})} \tag{2.7}$$

Similarly, we have

$$\frac{\partial\tilde{Q}(\theta)}{\partial\sigma_i} = \sum_{j=1}^{M} p(z_j = i|y_j, \theta^{(t)})\left[\frac{\sigma_i^2(1+\delta_j^2/\sigma_i^2) - (y_j-\mu_i)^2}{\sigma_i^4(1+\delta_j^2/\sigma_i^2)^2}\right] = 0 \tag{2.8}$$

Note that since σ_i and δ_j are entagled within the summation, there is not a simple analytic solution for σ_i. However, since the algorithms are iterative in nature and the major contribution for the update of σ_i is from $(y_j-\mu_i)^2$, we can approximate σ_i in δ_j^2/σ_i^2 with its value in t^{th} iteration. Then we can solve for the $(t+1)^{th}$ iteration relation for σ_i as:

$$\sigma_i^{(t+1)} = \left[\frac{\sum_{j=1}^{M}(y_j-\mu_i)^2 p(z_j = i|y_j, \theta_i^{(t)})/(1+\delta_j^2/\sigma_i^{(t)2})}{\sum_{j=1}^{M} p(z_j = i|y_j, \theta_i^{(t)})/(1+\delta_j^2/\sigma_i^{(t)2})}\right]^{1/2} \tag{2.9}$$

Our numerical test shows that such an approximation works fine in practice. For $w_i = p(z_j = i|\theta)$, we have

$$\frac{\partial \tilde{Q}(\theta)}{\partial w_i} = \sum_{j=1}^{M} p(z_j = i|y_j, \theta^{(t)})/w_i - \lambda = 0 \tag{2.10}$$

which leads to

$$w_i = p(z_j = i|\theta) = \frac{1}{\lambda} \sum_{j=1}^{M} p(z_j = i|y_j, \theta^{(t)}) \tag{2.11}$$

Using the condition $\sum w_i = 1$, we have $\lambda = M$. Substitute λ back to Eq. 2.11, we arrive at:

$$w_i^{(t+1)} = \frac{1}{M} \sum_{j=1}^{M} p(z_j = i|y_j, \theta_i^{(t)}) \tag{2.12}$$

In the above iteration relations Eq.2.7, Eq.2.9 and Eq.2.12, t and $t+1$ denote the round of iterations. When we ignore the measurement errors δ_j, the above recursive relation reduces to the standard EM recursive relation for Gaussian Mixture Model. The above relations are easily generalized to the multivariate case by simply replacing the data with data matrix, the mean with a mean vector and the variance with a covariance matrix.

It is likely that we can also improve the fit by adding more Gaussian components, although this is clearly not good in the sense of parsimony. So, we need to somehow decide on the number of Gaussian components by trading off quality of fit against the number of introduced free parameters. To accomplish this, we use the Bayesian Information Criterion (BIC) (Schwarz, 1978; Connolly et al., 2000) to determine how many mixtures we should use. The BIC is defined as:

$$BIC = -2 \log \mathcal{L}_{max} + k \log(M) \tag{2.13}$$

Where k is the number of free parameters. For mixture models with different number of mixtures, we compare corresponding BICs, and select the model with the smallest BIC.

2.3.2 Monte Carlo Test of The ECGMM

Before we delve into real data, we first conduct some Monte Carlo tests to see whether the ECGMM approach can reliably identify the cluster and background Gaussian components. These tests are used to determine whether this method can reliably recover the true parameters input in the simulation, and to see whether the extracted parameters are generally unbiased for varying levels of measurement error.

For this purpose, we generate two Gaussian random data sets, one representing cluster member colors, denoted CL, and the other representing the field galaxies/blue galaxies' colors, denoted as BG. The CL set is generated from $N(0.5, 0.04^2)$ and the BG set is generated from $N(0, 0.3^2)$. To represent clusters with different richness, we allow the normalization (also denoted as N_{gals} in the plots) of CL data set to vary as 10, 15, 20, 25, 30, 40, 50, 60 and 70 while keeping the normalization of the BG set as 30. These numbers are chosen to make the simulation as close to the real data as possible. Then we combine CL and BG to create mock data sets that mimic the colors of both cluster members and background galaxies in a field. It is worth noting that these mock colors are error free so far. Next, we will add some noise to them to mimic the measurements errors. To do this, we first generate random numbers from a uniform distribution in the range of [0, 0.1], which play the role of δ_j in Eq.2.1. Then, we generate from $N(0, \delta_j^2)$ and add them to the noise free data set to produce a noise added mock color data set. In Figure 2.1, we plot the results from the ECGMM fitting. The results show that for clusters with $N_{gals} \geq 10$, the method gives very reliable estimates for the locations (μ) and widths (σ) of the Gaussian components.

Next we test for possible bias in the estimators. For each cluster richness N_{gals}, we replicate the data as well as errors 200 times and then apply our methods to each to obtain estimates for the parameters. In each case, we calculate the bias of parameters θ (the σ and μ in our case) defined as $E(\hat{\theta}) - \theta$. In Figure 2.2, we plot the results from both GMM and ECGMM for comparison. Clearly, the introduction of error correction(as shown in the bottom two panels) is essential for removal of the bias of the width resulting from measurement error(as shown in the top two panels).

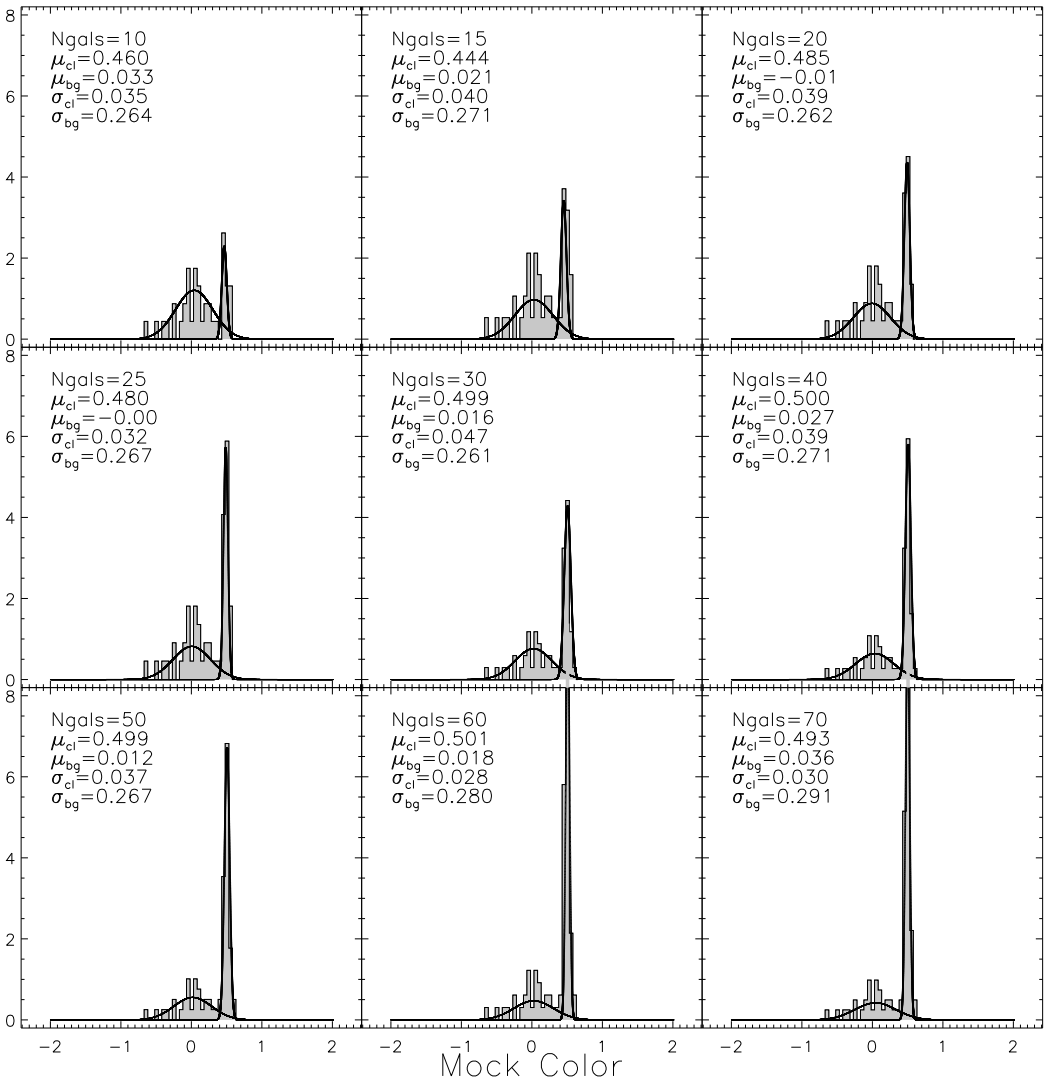


Figure 2.1. The ECGMM fitting to the mock color's distribution. μ and σ denote the locations and widths of the corresponding Gaussian components. The true μ are 0 and 0.5 for BG and CL sets respectively. The true σ are 0.3 and 0.04 for BG and CL sets respectively.

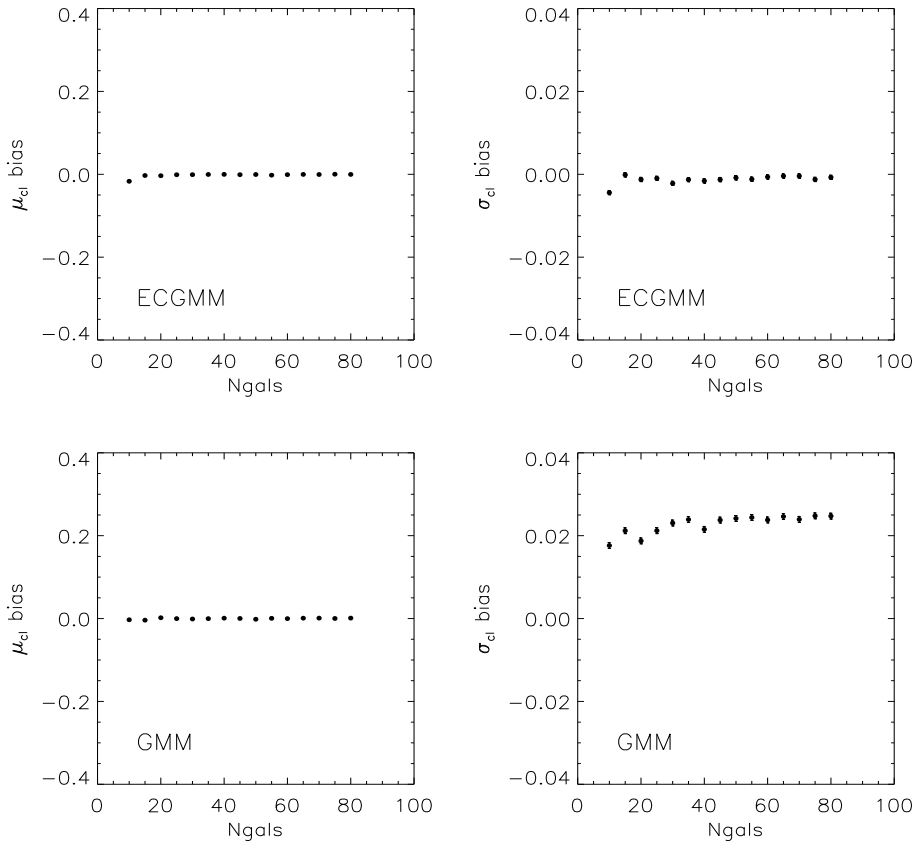


Figure 2.2. Monte Carlo test of the bias, $(E(\hat{\theta}) - \theta)$, of the GMM estimators for the location and width using GMM (bottom two panels) and ECGMM (top two panels) as a function of richness for the cluster component of the mock clusters (see text).

2.3.3 Bootstrapping to Increase The Robustness of ECGMM

Though the ECGMM is generally quite stable for identifying the parameters of the Gaussian Components, it can fail occasionally due to a very inappropriate choice of initial parameters or some very large measurement errors for certain galaxies. To make our measurement more robust, we introduced a bootstrap-like scheme. Suppose we have N data points. We randomly pick one of the data points and record it. We then repeat this process N times and get N recorded data points. These N points form one resampling set of the original data set. Now, we apply the ECGMM to this new data sample and measure the corresponding parameters. After this, we start a second round, getting another resampling set with N data points in it and measure the parameters using ECGMM again. We repeat this process X times and obtain X estimates of each parameter. We throw away those outlier estimates (estimates beyond the upper and lower inner fences[2]) for each parameter and use the mean of good estimates as the value of each parameter. Using this scheme, our resulting parameter estimates are much more robust, at a cost of a tolerable increase in computation time. In this application, we took X to be 50.

2.4 Precision Measurements of E/S0 Ridgeline for MaxBCG Clusters

2.4.1 Evolution of Ridgeline and Its Width

We apply the above prescriptions of ECGMM to the maxBCG catalog (Koester et al., 2007a), measuring the red sequence $g-r$ ridgeline. The procedures are as follows: for each cluster in maxBCG catalog, we choose a scaled aperture R_{200}^{lens} to ensure we are considering equivalent regions of clusters of varied masses and therefore varied richness. R_{200}^{lens} is the critical radius, interior to which the mean mass density of the cluster is 200 times of the critical energy density of the Universe. Based on the weak lensing analysis (Johnston et al., 2007; Hansen et al., 2007), the scaling

[2]In statistics, the lower inner fence is defined by $Q_1 - 1.5IQR$ and higher inner fence is $Q_3 + 1.5IQR$, where Q_1 and Q_3 are the first and third quartiles respectively. The IQR is the interquartile range, defined as $Q_3 - Q_1$.

relation between R_{200}^{lens} and the original maxBCG richness N_{200} is given by $R_{200}^{lens} = 0.182(N_{200})^{0.42}$.

Next, we identify all SDSS galaxies inside this aperture range, fainter than the BCG, and brighter than 0.4 L* at the redshift of the cluster. Then, we apply the ECGMM procedure to the $g-r$ colors and corresponding measurement errors of these galaxies. One of the resulting two Gaussian components from the ECGMM will represent the cluster red sequence color distribution while the other represents the background/blue galaxy color distribution. To determine which Gaussian Component belongs to the cluster, we calculate the likelihood of the BCG's $g-r$ color for each Gaussian Component. The component for which the BCG has a higher likelihood is assigned as the cluster component and the other is declared background.

For comparison, we measure the red-sequence location and width using both ordinary GMM and ECGMM. The top panel of Fig. 2.3 shows the evolution of the average g-r ridgeline location and width measured using ordinary GMM. We observe the well-known trend in the average ridgeline zeropoint. In addition, there is strong apparent evolution in the average ridgeline width, which becomes nearly 100% larger by $z = 0.3$. However, from the lower two panels which are measured using ECGMM, one can see very clearly the power of ECGMM in constraining the intrinsic width of the ridgeline without contamination from measurement error. The results show that the mean observed $g-r$ ridgeline location retains the same linear dependence on redshift while the mean width of the ridgeline shows a weak dependence on redshift, with the $g-r$ scatter $\sigma(z=0.1) = 0.044 \pm 0.001$ and $\sigma(z=0.3) = 0.057 \pm 0.002$ or a broadening by $\sim 20\%$ from $z = 0.1$ to $z = 0.3$. Unsurprisingly, the strong dependence of the scatter on redshift from the GMM is mostly due to the increased measurement errors for cluster members at higher redshift.

2.4.2 Ridgeline Tilt From Galaxy Clusters

It has been pointed out that the color-magnitude relation (CMR) of cluster member galaxies has a negative slope (e.g Kodama & Arimoto, 1997; Gladders et al., 1998), so that fainter member galaxies are generally bluer. The evolution of these CMR slopes

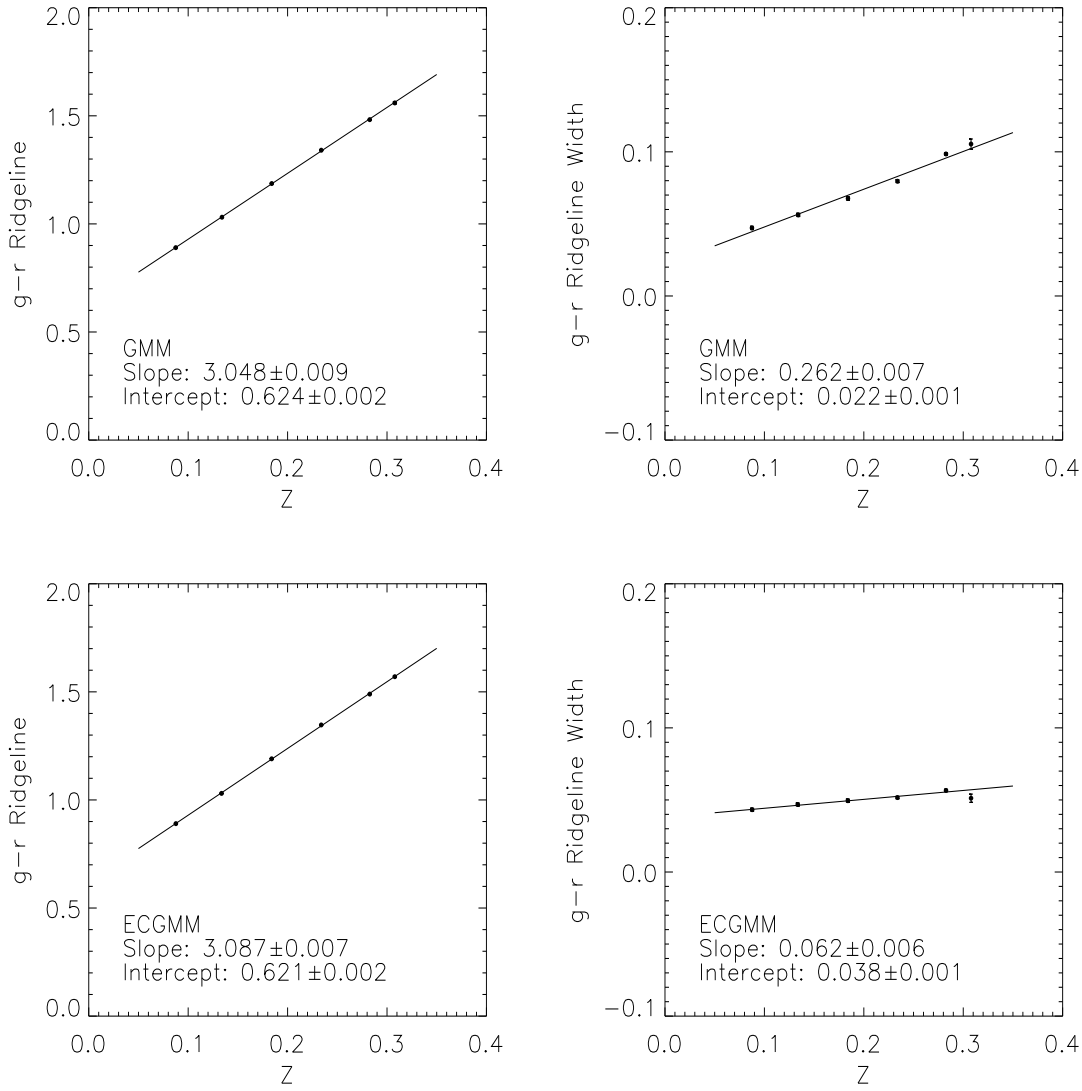


Figure 2.3. Tracking the $(g - r)$ red sequence zeropoint and width as a function of redshift, measured using ordinary GMM (upper panels) and ECGMM (lower panels) respectively. After error correction, the broadening of the observed red-sequence width with redshift is greatly suppressed, revealing the effect of photometric errors on the observed broadening.

with respect to redshift and richness has been difficult to address, largely due to the lack of a sufficiently large cluster catalog with well measured photometry for all its galaxies. The maxBCG catalog provides about 14,000 galaxy clusters, extending over $0.1 < z < 0.3$, which enables us to measure the slope of the CMR for clusters with good statistics across a range in both richness and redshift.

Measurement of the slope of the CMR typically proceeds by identification of the cluster red-sequence, followed by some iterative process of outlier removal, and a determination of cluster "member" galaxies which are then used to measure the slope and zeropoint of the CMR. We apply the method described in previous sections to measure the color distribution of individual clusters and to select the members for every cluster by requiring the color difference between each member galaxy and ridgeline color to be less than $\pm 2\sigma$ (σ is the convolved ridgeline width, given by the best-fit ECGMM, and measurement errors of individual member galaxy's color). We choose 2σ because this is roughly where the background component's likelihood dominates over the cluster component's likelihood. Based on this identification of membership driven by the ECGMM, we fit the members in the CMR of every cluster with a straight line and call its slope as the slope of ridgeline in what follows.

The distribution of ridgeline slopes for maxBCG clusters are shown in Figure 2.4 in bins of $\Delta z = 0.03$. Despite the substantial scatter in slope among individual clusters, we can see from Figure 2.4 and Figure 2.5 that the mean slope of the red sequence ridgelines for clusters deviates from zero for $0.1 < z < 0.3$. For any bin, the error on the mean places the measurement many σ from zero.

In Figure 2.5, it is apparent that the observed trend of the mean ridgeline slope with redshift is statistically significant (about 15 σ different from 0): the slope becomes steeper by a factor of 2.5 by $z = 0.3$. In Figure 2.6, we plot the evolution of the slopes vs richness in each redshift slice, which shows that the dependencies of ridgeline slope with respect to richness is weak, as observed elsewhere (e.g. Hogg et al., 2004). Clearly, the observed slope of the red-sequence is not associated with cluster richness, and is unsurprisingly a strong function of redshift.

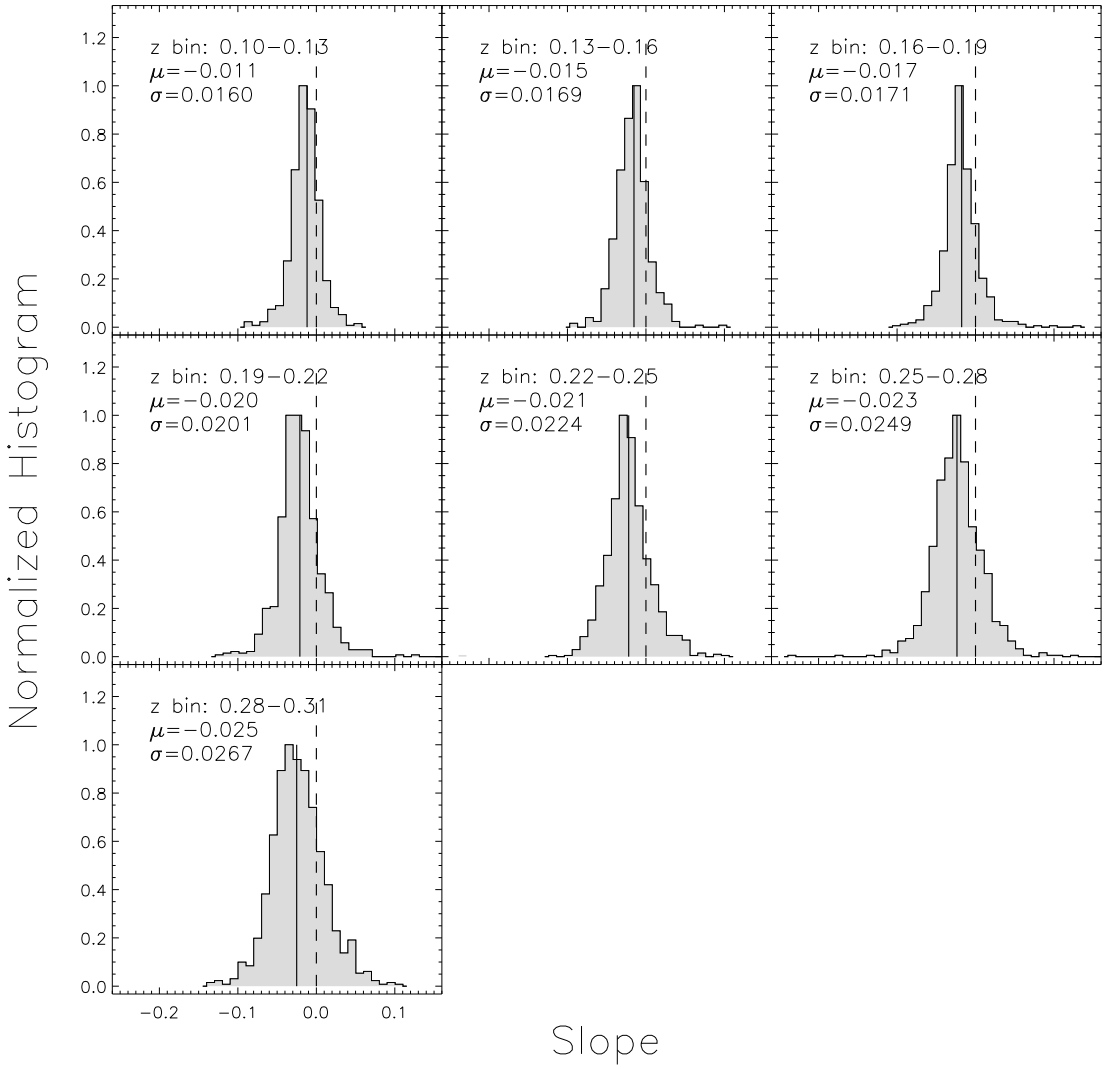


Figure 2.4. The distributions of measured ridgeline slopes for clusters in steps of 0.03 in redshift. μ and σ denote the mean and width of the distribution The dashed line corresponds to zero.

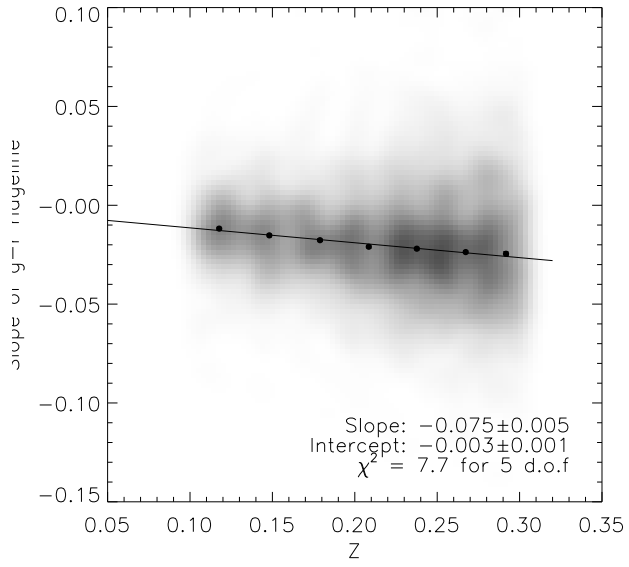


Figure 2.5. Tracking the observed red-sequence slope vs redshift. The gray clouds represent the slope measurements from individual clusters. The black solid circles and error bars (too small to be seen on the plot) are the weighted mean and the standard deviation to the weighted mean for each redshift bin ($\Delta = 0.03$).

2.4.3 Ridgeline Tilt from Spectroscopic Data

The above measurement is based only on a photometric determination of red sequence galaxies. The level to which projection plays into this selection is as yet unknown. The true red-sequence galaxy population in some physical volume, either in a cluster or in the field, is contaminated by dusty foreground galaxies which can be rejected via spectroscopy, and by the peculiar velocities of the galaxies themselves.

To address the possibility of foreground contamination, it is interesting to see if the above results are preserved in a spectroscopic sample of galaxies. To achieve this goal, we use galaxies with spectra from DR6 of the SDSS Value Added Galaxy Catalog (VAGC) (Blanton et al., 2005) for a comparison. Due to the selection effects of the spectroscopic data, we will choose only the galaxies with redshift from 0.1 to 0.20 and brighter than 0.4 L* magnitude at their respective redshifts.

By extension from the photometric sample and from previous work, we know that the slope does not vary with environment, so the field sample represented by our spectroscopy should be a fair representation of the expected slope in clusters.

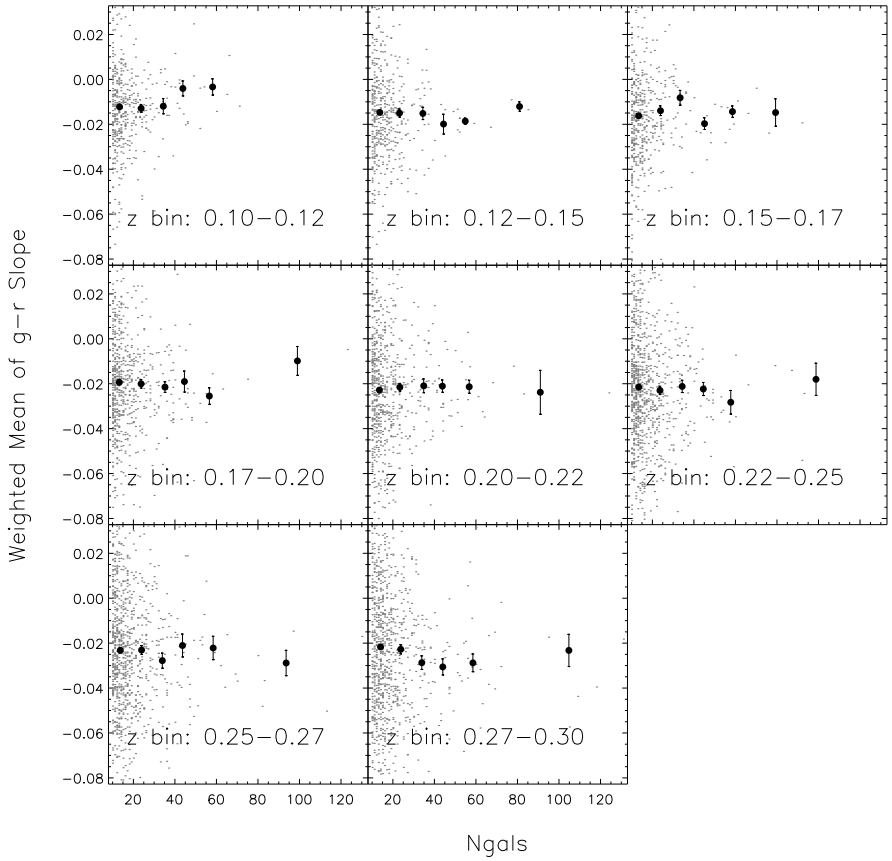


Figure 2.6. The evolution of mean ridgeline slope vs richness at different redshift slices. The richness bins brackets are chosen as N_{200} =[10,20,30,40,60,80,161]. The light dark points are from individual clusters. The black solid dots and error bars are weighted mean and standard deviation of the slopes in each N_{gals} bin for every redshift slice. Given the low statistics at rich bins, we did not see strong trends of the slope evolution w.r.t richness.

Our procedures are as follows: we first bin the galaxies into bins of size $\Delta z = 0.003$, which corresponds to velocity slices of 900km/sec. The color distribution of the galaxies in each bin shows clear bimodality (to panel of Fig. 2.7). Then, we separate the red sequence galaxies in each bin using ECGMM. The red sequence galaxies correspond to the Gaussian component with bigger $g - r$ value and we choose $\pm 2\sigma$ from the peak location as red galaxy samples for each redshift slice, in a fashion similar to the one we used for cluster galaxies. Then, we fit the CMR of galaxies' $g-r$ colors and i-band magnitude with a line in every bin, recording their corresponding mean

and slope. In the bottom panel Figure 2.7, we choose 6 redshift bins ($\Delta z = 0.003$) to illustrate the red/blue galaxy separation and the ridgeline slope fitting in each bin. Finally, we fit the variation of slope with redshift with a line to look for a trend. The results are shown in the left panel of Figure 2.8. As a comparison to the cluster sample, we also plot the mean variation of the ridgeline slope for clusters in the same redshift range [0.1, 0.2] in the right panel of Figure 2.8. When the redshift range is changed, the slope of the fitted line for the cluster sample becomes steeper as compared to Figure 2.5. The reason lies in that the linear fit to the trend is only the first order approximation. For our purpose here, we just need to require that the cluster sample and spectroscopic sample have the same redshift range so as to compare them fairly.

A comparison shows that the slope from spectroscopic and cluster samples differ by $\simeq 20\%$ on the $g - r$ vs. i slope at $z \simeq 0.1$: -0.013 vs. -0.010 and by about the same amount at $z = 0.2$. The trend with redshift is similar in both samples, but the slopes are generally steeper in the spectroscopic sample.

2.4.4 Possible Reasons for The Evolution of Ridgeline Slope

Based on the measurement results in previous sections, we have the following observations: i) The mean slope of CMR is negative and it become more and more negative as redshift increases. This is in agreement with the results in Gladders et al. (1998). ii) The slopes are almost independent of cluster richness. Even the non-clustered spectroscopic data show similar slopes and redshift dependence.

What can we learn from this? The origin of negative slope in the CMR has been considered mainly to be a result of the initial mass-metallicity enrichment difference among elliptical galaxies of different masses (Gladders et al., 1998). The ages of galaxies are also thought to be partially responsible for the ridgeline tilt, but play a minor role compared to metallicity.

As for the redshift evolution of the slopes, we need to make it clear that the measured slopes are sensitively dependent on the galaxy populations considered. The colors of galaxies in a narrow redshift slice (~ 0.003) show bimodality. The ridgeline

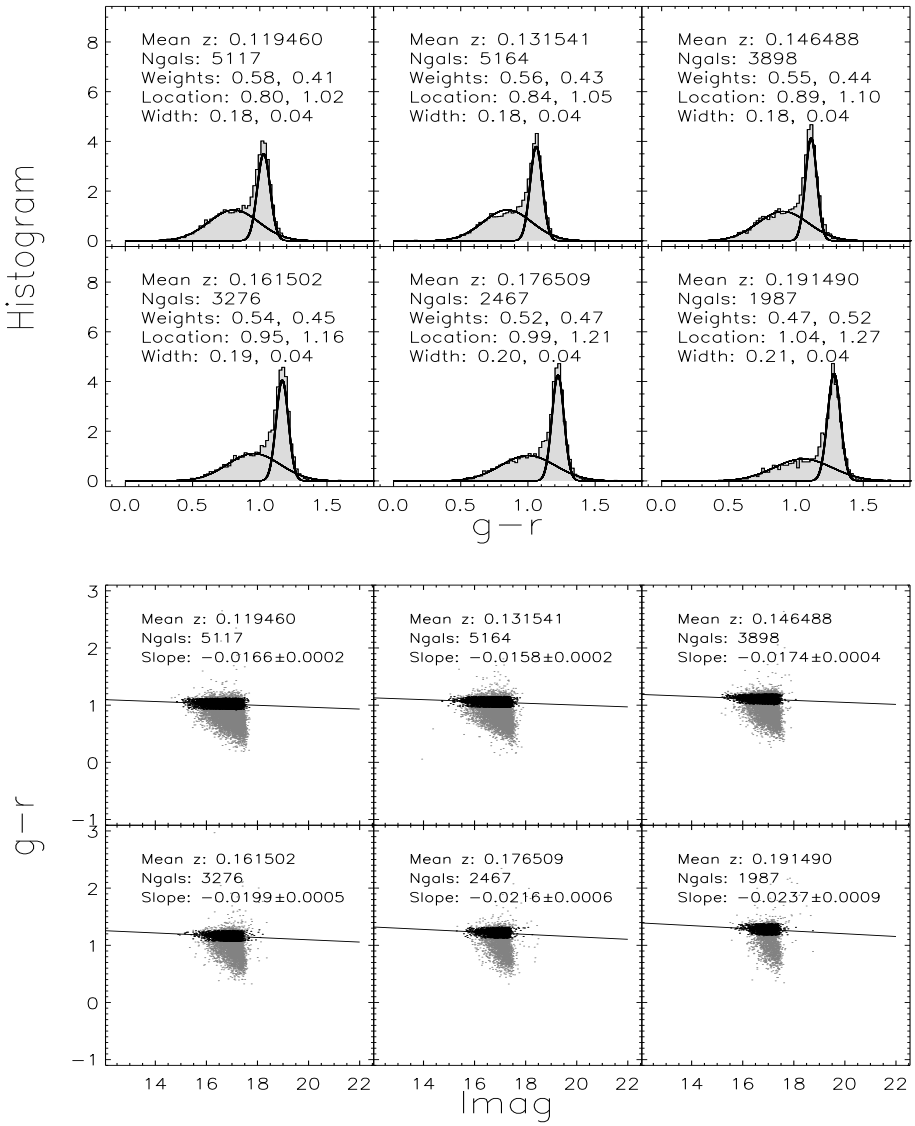


Figure 2.7. Evaluating the ECGMM-derived red sequence slopes in SDSS spectroscopy of field galaxies. The normalized color histograms (top panel) for $\Delta z = 0.003$ slices in spectroscopic redshift clearly show the presence of the red and blue components in the field galaxy distribution. ECGMM is used to separate the two components, the redder of which is to measure the CMR (bottom panel).

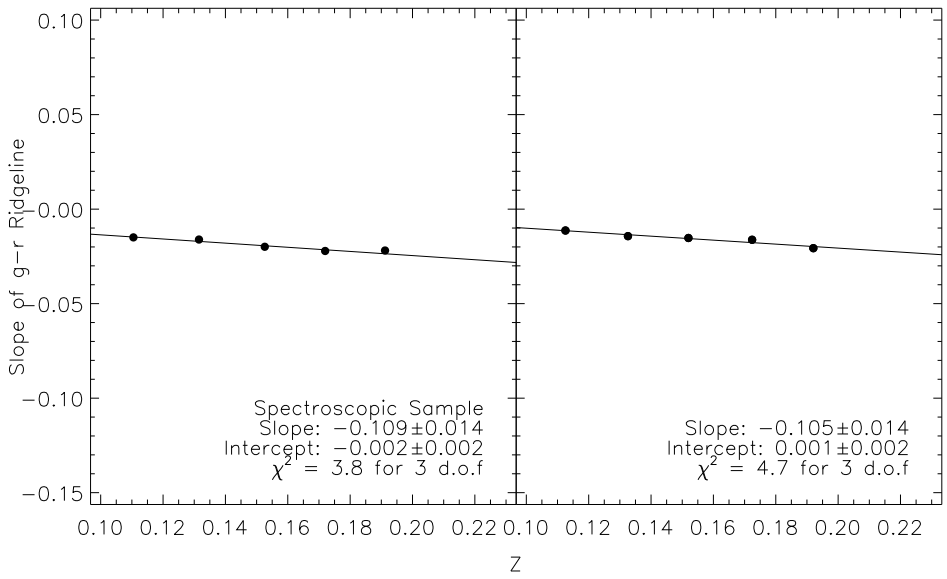


Figure 2.8. The comparison of the evolution of the slopes of CMR for the spectroscopic sample and the cluster sample. Since the spectroscopic sample is biased due to selection effects at $z \geq 0.2$, we just plot both samples in the redshift range from 0.1 to 0.2.

slope is basically meaningful only for red sequence galaxies. Then, the issue is how to isolate (or select) the red sequence galaxies. Generally, the red sequence galaxies are separated in terms of the color bimodality, i.e. the red component of the bimodal distribution. However, the red galaxies selected in terms of the color cannot guarantee the same population of galaxies in terms of other physical properties, such as metallicity, age and SFR, across the whole redshift range. This is likely a major reason for the change in slopes as redshift increases.

To show this quantitatively, we use a star formation rate (SFR) cut in addition to the color cut to select the red galaxies on spectroscopic data. The SFR measurement we used Brinchmann et al. (2004),is cross matched with the spectroscopic data we used in section 2.4.3. As we increase our cuts on SFR to include only older and deader galaxies, the ridgeline slope evolution with redshift become milder and milder. The resulting plot is Figure 2.9.

2.4.5 Discussion

To this point, the measurements have been presented in the observed frame and stand alone. Photometric cluster detection and the quantities derived (e.g. richness)

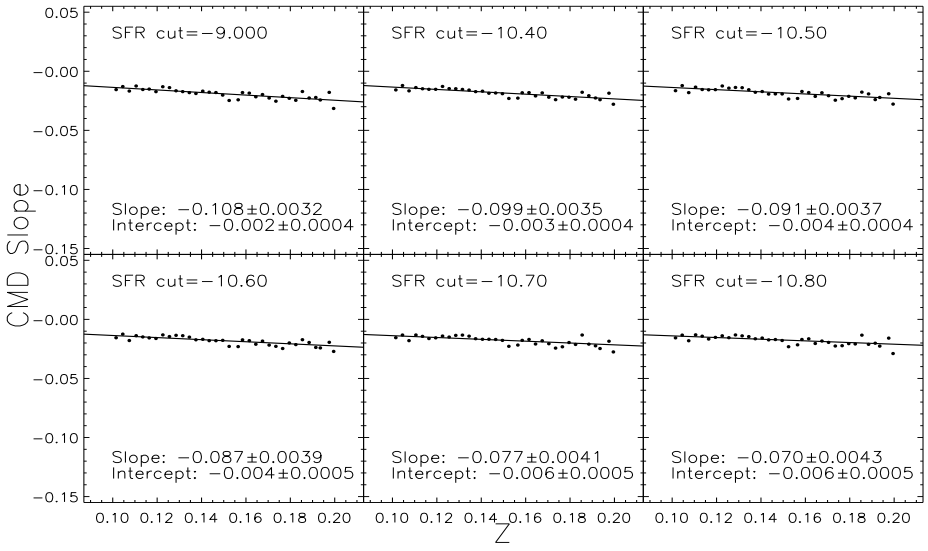


Figure 2.9. The evolution of ridgeline slope for different star formation rate (SFR) cuts. The evolution becomes milder as we cut off those galaxies with higher SFR. Due to the limitation of the available data, we cannot cut more to further reduce the slope, but we expect the slope can go to zero if we remove those high SFR galaxies completely and consistently across redshift.

operate in the frame of the observer, and predictions from galaxy formation models and mock galaxy catalogs can be evaluated in light of these precision measurements. They have particular applicability to calibration of optical cluster detection efforts, especially to those that rely on the properties of the red sequence. The methodologies developed herein allow the "bootstrapping" of optical algorithms: basic cluster finders locate the clusters, and precision measurements (such as these) of said clusters lead to refinements in those algorithms. For illustrative purposes, we list the relevant observational considerations to be made in understanding the context of these measurements with respect to previous work in the literature, and then highlight a few of our more interesting results.

In general, there are five places where the comparison to previous work must be treated with caution, which can be summarized as follows: 1) redshifting of the galaxy spectra through the bandpasses under consideration, which imparts trends in the observed colors, 2) selection effects imposed by the color selection (e.g. Franzetti et al., 2007), 3) aperture effects, i.e. the aperture used to measure the color in different banspasses (e.g. Scodeggio, 2001; Blakeslee et al., 2006), 4) projection effects. 5)

actual evolution in the red sequence.

At some level, any of the aforementioned issues may play into our results: (i) at $z \simeq 0.1$, the CMR of photometrically-selected galaxies is noticeably shallower than previous spectroscopic measurements of the color magnitude relation (Hogg et al., 2004; Cool et al., 2006), (ii) the slopes are almost independent of cluster richness; (iii) the photometric error-corrected scatter of the red-sequence broadens mildy with redshift; (iv) the observed mean slope of the CMR is negative and it becomes more negative as redshift increases.

Naively, we expect that our measurement of the slope of the red-sequence, $-0.013\pm$ 0.003 mags mag^{-1} at $z = 0.1$ corresponds to the SDSS spectroscopic analysis of Hogg et al. (2004), for which the slope is -0.022 mags mag^{-1} in $^{0.1}(g-r)$. In addition to the fact that the Hogg et al. (2004) measurements are corrected to the $z = 0.1$ rest-frame, one possible difference comes from our definition of the red sequence: Hogg et al. (2004) use a 2σ clipping algorithm to define the red sequence and to iteratively reject outliers. While they split the sample by Sersic index, sigma-clipping may be more permissive of objects near the "blue cloud" to be included in the red-sequence, while the method presented in this thesis automatically accounts for the presence of these objects. Our slope measurements at a given redshift may also be biased shallow, as the initial 2 σ cut derived from the ECGMM fit does not account for the slope in the red-sequence itself, i.e. the cut is applied in the same way regardless of magnitude. Ideally, an iterative procedure would be employed to determine the best-fit line for each cluster and the 2σ cut would be applied as a function of magnitude. Unfortunately, the small number statistics for low richness clusters do not permit this to be implemented in a robust fashion.

Insofar as richness and local density are similar indicators of environment, the second observation (ii) that the slope is almost independent of environment is in basic agreement with Hogg et al. (2004), who use SDSS spectroscopy at $z \sim 0.1$ to compare galaxies with high ($n \geq 2$) Sersic indicies in different environments characterized by their local density.

After the photometric error correction is performed by ECGMM, a trend in the

scatter with redshift remains (iii), such that the scatter increases with increasing redshift. At high redshift $z \simeq 1$, the color-magnitude relation has been measured in a handful of clusters (Mei et al., 2009; Koester et al., 2009; Santos et al., 2009, e.g.) with the general conclusion that the restframe scatter in the CMR does not evolve with redshift. More locally, the SDSS Luminous Red Galaxy (LRG) sample has been used to measure various redshifted frames of bright ($L \gtrsim 2.2L_*$) red galaxies (Cool et al., 2006). Cool et al. (2006) find the intrinsic rest-frame scatter ${}^{0.16}(g-r) = 35.4 \pm 3.7$ and ${}^{0.37}(g-r) = 43.5 \pm 6.2$ mmags mag^{-1}, consistent with no evolution. However, our own *observed* frame measurements reveal an increase in the scatter of $\simeq 20\%$ over a similar time period. The ultimate explanation for this discrepancy is likely found in the observed CMR steepening with increasing redshift. Because we can not robustly subtract off the CMR for individual clusters, the measured width of the color distribution of the CMR will be broadened by the increasing tilt of the CMR.

Result (iv) is in qualitative agreement with the results in Gladders et al. (1998) who find a similar trend in the slope for a sample of 44 Abell clusters at $z \leq 0.15$ and 6 clusters at $0.2 \leq z \leq 0.75$, the largest previous study of its kind. In their study of the scatter of the CMR in LRGs, Cool et al. (2006) report no significant trend with redshift in the rest-frame slope of LRGs over $0.16 < z < 0.37$ in either the cluster or the field, but caution that the sample is not-well suited to measuring the slope. The observed factor of 2.5 increase in the magnitude in our measurement of the slope is likely due to a combination of the lack of k-corrections and selection effects (e.g. Franzetti et al., 2007) derived from color cuts that may preferentially include a larger and larger fraction of galaxies with siginificant star-formation at increasing redshifts.

A further contribution to the inflated slope may come from the choice of the color aperture. van Dokkum et al. (1998) and Scodeggio (2001) note the importance of the use of adaptive apertures, which place the color measurements of large and small galaxies on the same footing. This point motivates our choice of `MODEL_MAGS` from the SDSS, which are derived from the best-fit convolution of the local PSF with a deVaucoleurs model in the r-band. This same best-fit model is then used to compute the flux in both the g and r-bands.

2.5 Summary

In this chapter, we've presented the ECGMM, a new purely photometric method which characterizes the red sequence ridgeline in large statistical cluster samples. This provides precise measures of the mean variation of the red sequence ridgeline location and width with respect to redshift, properly corrected for photometric errors. The measured slopes, scatters, and zeropoints are directly applicable to improved cluster finding efforts and to characterization of known galaxy clusters.

Applying the method to maxBCG clusters approximately recovers known properties of the red sequence, namely its slope and the variation of the slope with redshift, and the insensitivity of the slope to environment. It also preliminarily suggests that the width of the red-sequence increases with redshift, and that the slope of the red-sequence grows substantially by $z \simeq 0.3$, but we caution that these observed trends may be attributable to a host of observational effects that we have made no attempt to correct. Color selection effects, the lack of k-corrections, and the details of the measurement of the individual cluster CMRs require proper attention before applying these results to models of galaxy formation.

CHAPTER 3

GMBCG Algorithm for Optical Cluster Detection

3.1 Overview

As pointed out in the previous section, de-projecting the field galaxies is crucial for efficient optical cluster detection. Red sequence galaxies in clusters are clustered tightly in color space in addition to the ra/dec plane, providing a robust way to de-project field galaxies. Therefore, if we can identify red sequence clustering in color space and then combine it with clustering in the ra/dec plane, we will be able to single out clusters with minimum contamination from projection.

Galaxies' color distribution around a cluster can be well approximated by a mixture of two Gaussian distributions. The redder and narrower Gaussian distribution corresponds to the color distribution of the cluster's red sequence members, while the bluer and wider one corresponds to the background galaxies' color distribution. If there is no cluster, then the color space will be a single Gaussian distribution with a wider width because there is no red sequence component. Therefore, we can fit the color distribution with mixtures of Gaussians, and use certain criterion (we use Bayesian Information Criterion in this project) to determine how many mixtures give the best fit. This can tell us whether there is red sequence clustering in color space. One complication in our case is that errors in the measurement of colors are not negligible and proper modeling of them is essential for proper fitting of the mixture models. The traditional Gaussian Mixture Model does not consider the measurement errors and we therefore developed an error corrected Gaussian Mixture Model to include them (Hao et al., 2009). It has been applied to measure the ridgeline properties of maxBCG clusters, leading to precise and unbiased information about the evolution

Table 3.1. The ridgeline color in different redshift ranges for SDSS filters

Ridgeline Color:	g-r	r-i	i-z
Redshift Range:	$0.0 \sim 0.35$	$0.35 \sim 0.70$	$0.70 \sim 1.0$

of the E/S0 ridgeline and its width.

As long as we single out the red sequence galaxies, we will reduce the problem to a clustering analysis on the ra/dec plane. One can then use either parametric (such as convolving a model kernel) or non-parametric (such as Voronoi Tessellation) methods to analyze the clustering. The key part is separating the red sequence galaxies. When we apply such a scheme to data spanning a wide redshift range, there are four other complications to consider.

The first is that as redshift increases, the E/S0 ridgeline color changes accordingly. This is mainly a result of the 4000 Å break shifting across the color filters. Because of this effect, the most informative color (ridgeline color) will vary as redshift increases. For the set of SDSS filters, the relation between ridgeline color and redshift is shown in Table. 3.1

Beyond $z \sim 1.0$, one needs infrared color information, such as Y, J or K. Therefore, when detecting clusters in data spanning a wide redshift range, it is necessary to determine which ridgeline color we should work on for given galaxies. To make the decision, we use photometric redshift (photoz). As we noted before, the uncertainty of photoz is about $0.02 \sim 0.03$, which is quite large for selecting cluster members, but will suffice to determine which ridgeline color should be used for each galaxy. At the border regions, the two adjacent ridgeline colors can yield similar results. Therefore, we can determine the ridgeline color based on photoz of the galaxy. This requires that we need to get the photoz for every galaxy beforehand using other methods.

A second complication of cluster finding across a wide redshift range arises from the increased chance of overlapping clusters, one at low redshift and another at relatively high redshift. Such an overlap will complicate the distribution in color space, turning it from bimodal to tri-modal or even more. To reduce such possibilities, we use a broad photoz window (such as $\pm$ 0.25 in photoz) to clip on the target galaxy/-

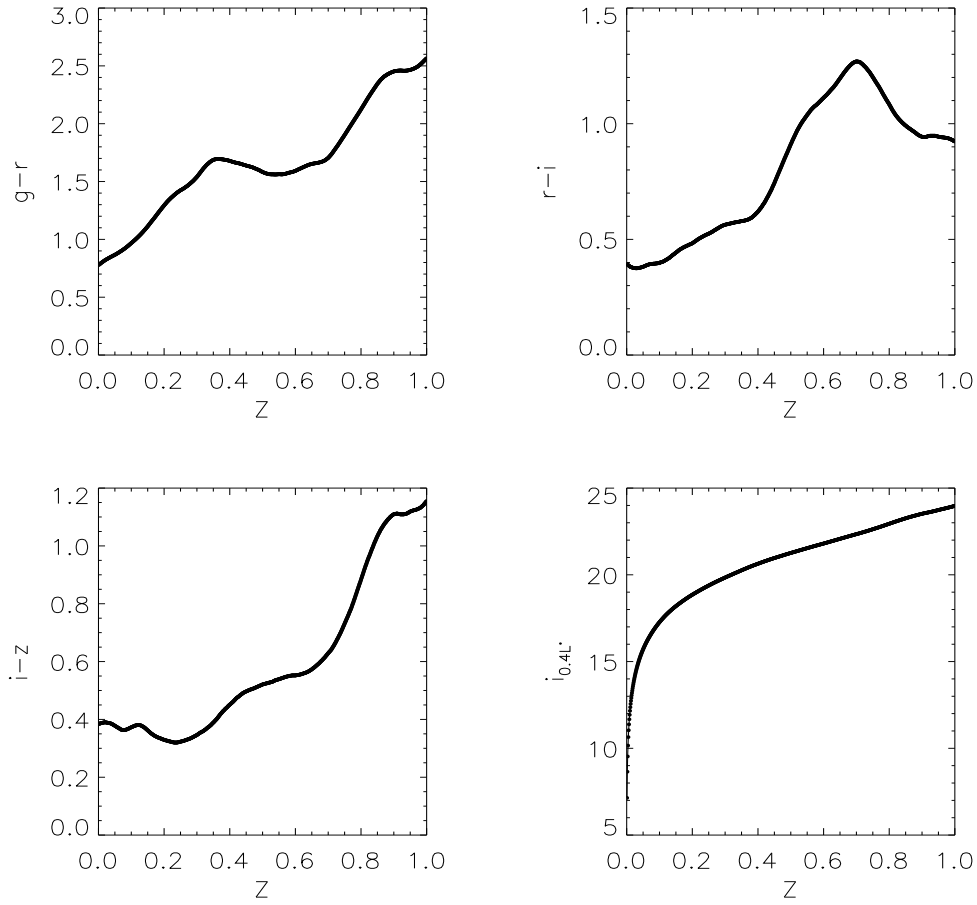


Figure 3.1. The top two and bottom left panels are the color evolution based on a color model of the red sequence galaxies (Koester et al., 2007a). The bottom right panel is the I band apparent magnitude corresponding to $0.4L^*$ at different redshifts.

cluster. The available photoz precision is adequate for this purpose too. In addition to photoz clips, we also apply magnitude cuts and require the galaxies brighter than 0.4 L^*, where L^* is the characteristic magnitude in the Schechter Luminosity function. For our application, the $0.4L^*$ as a function of redshift is shown in the lower right panel of Figure 3.1, which is adopted from (Koester et al., 2007a). These two measures can do very well for removing the foreground and background clusters and thus simplify the color space structure around the target galaxies.

The third complication concerns richness measurement of the clusters when the ridgeline color is changed. The red sequence galaxies selected from different color band have different degrees of contamination from the background. This is the fundamental

limit of all color based red sequence selection. This has a relatively minor effect on our cluster detection. As long as we have the cluster catalog available, we will need to further calibrate the richness measured from different color bands using other means, such as gravitational lensing.

The last but not least complication is the increase of measurement error as redshift increase. Generally, measurement errors depend on the brightness of the galaxy. As redshift increases, the fraction of faint galaxies increases and so do the average measurement errors. As a result, our cluster detection will inevitably be redshift dependent. As we will show in what follows, our detection likelihood function will reflect this tendency. In the following subsections, we will show how to quantify the likelihood functions to be used in cluster detection.

3.2 Brightest Cluster Galaxy as Cluster Centers

In our algorithm, the center of the cluster is assumed to be the Brightest Cluster Galaxy (BCG). This is a very useful and reasonable assumption, with good physics, algorithmic and computational motivations. The major physics motivation for focusing on the BCG as a cluster center is that the gas in a cluster is dragged to the bottom of the gravitational potential well, making the central galaxy a lot brighter. The BCG is widely thought to be co-located with the most bound particles typically identified with the cluster center in simulation. Therefore, this is a pretty fair assumption. The assumption is also very useful algorithmically. The BCG can serve as a "noise damper" during the cluster finding process. Without the BCG assumption, the projected galaxy density field would be very sensitive to foreground and background galaxies, which we cannot remove completely by color cuts. A few incorrectly de-projected galaxies can severely affect the projected density centers in the RA/DEC plane and therefore affect cluster detection. For all these reasons, searching for a BCG as the center can boost the efficiency of cluster detection.

3.3 E/S0 Ridgeline Selection

As a first step, we show how to determine the appropriate ridgeline color ($g-r$, $r-i$, etc) for a given candidate BCG. Suppose we have photometric redshifts for every galaxy, denoted as z_p with an uncertainty denoted by σ_p. The true but unknown redshift of the galaxy is denoted by z (unknown). If we assume the z_p errors are normally distributed (Abazajian & Sloan Digital Sky Survey, 2008), we can write down the probability that the true redshift is within the redshift range z_{min} and z_{max} as:

$$P(z_{min} \leq z \leq z_{max}) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \int_{z_{min}}^{z_{max}} dz \exp\left[-\frac{(z-z_p)^2}{2\sigma_p^2} \right] \tag{3.1}$$

The advantage of making a cut in probability space is that it allows us to take uncertainties into account. One can put a threshold probability requirement for galaxies being selected as a given ridgeline. As a result, some galaxies will not be assigned to any ridgeline if they have very big uncertainties. If you take the philosophy that a galaxy should belong to either ridgeline color, then it will be the same as a direct cut on photoz. For Sloan filter system, the ridgeline and corresponding redshift range is specified in Table.3.1.

3.4 BCG Candidates Pre-selection

In principle, every galaxy is a potential BCG and should be tested. However, this will be computationally expensive and noise prone. Since we roughly know what a BCG should look like in terms of its color and redshift, we do a very coarse pre-screening to remove galaxies that are obvious not BCGs from our search lists. No doubt, there is a risk of removing some true BCGs, but we can reduce the chance of this by imposing a very broad cuts. To implement this pre-selection, for a given test galaxy, we introduce the following BCG color likelihood, which quantifies the closeness of a given galaxy's color to the predicted ridgeline color at its redshift.

Based on the color model derived from the SDSS red galaxies with passive evolution to high redshift(Koester et al., 2007a), we can get a simple relation between

ridgeline color and redshift. We denote the model predicted ridgeline color at a given the redshift z as $\bar{c}(z)$. The $\bar{c}(z)$ is shown in Figure 3.1. Then, the likelihood that a given galaxy with known z_p and color c in the corresponding ridgeline is given by

$$L^{BCG}_{candidate} = \frac{1}{2\pi\sigma_p\sqrt{\sigma_c^2 + w_c^2}} \times \int_{-\infty}^{\infty} dz \exp\left[-\frac{(z - z_p)^2}{2\sigma_p^2} - \frac{(c - \bar{c}(z))^2}{2(\sigma_c^2 + w_c^2)} \right] \qquad (3.2)$$

where σ_c is the measurement uncertainty of the galaxy's corresponding color and w_c is the corresponding ridgeline width at that color. We then impose a threshold on this likelihood to throw away non-BCGs. To avoid excluding potential real BCGs, we keep the threshold very low, i.e. 0.1. However, there is still danger for false rejection of potential BCGs due to some very incorrect photozs. To make our cut insensitive to the photozs, we do not impose the cut directly on the likelihood Eq.3.2. Instead, we select potential BCGs based on this likelihood, examining their distributions in color-color space, and make cuts in color space. When we apply cuts in color space, we make our cuts a lot bigger to cover the region where our selected BCGs reside. By such a scheme, we can avoid removing those potential BCGs with catastrophic photozs.

3.5 Red Sequence Member Galaxy Selection

The sizes of clusters are varied. Therefore, using a variable aperture size to measure the properties of clusters is desirable. For a candidate cluster, we should try a series of different aperture radii, and select the one that gives best S/N. However, this can be computationally expensive. As a substitute, we take a two-step approach to achieve the above purpose. First, we measure the richness of the cluster by using a fixed aperture size. Second, we scale the radius based on our measured fixed aperture richness and remeasure everything using the scaled aperture size. The key part of this procedure is we need to have appropriate scaling relations so as to improve the Signal/Noise. In the following, we will show how we implement this.

3.5.1 Fixed Aperture Membership and Richness

For a candidate BCG, we identify cluster members in the following way. We draw a 0.5 Mpc circle around the candidate BCG and get all the galaxies fainter than it but brighter than the 0.4L* magnitude at the candidate BCG's redshift. Since we already know which color ridgeline we should use for this candidate BCG, we use a two-component Gaussian mixture to fit the distribution of the colors of all the galaxies selected above (See Chapter 2). To remove possible overlap of two or more clusters along the line of sight, we apply a photoz clip to consider only galaxies within a photoz difference of ± 0.25. The precision of photoz is adequate for this purpose. Also, a Bayesian Information Criterion (BIC) is introduced to determine how many Gaussian components are appropriate (Hao et al., 2009).

To determine which Gaussian component the candidate BCG belongs to, we compare its corresponding likelihoods of belonging to each of the two Gaussian components (one corresponds to the color distribution of red sequence galaxies and another corresponds to the color distribution of the field and blue galaxies) and select the most likely Gaussian component as the candidate BCG's red sequence galaxy distribution. Its mean will correspond to the location of the ridgeline and its standard deviation will be the width of the ridgeline. All the galaxies whose colors are within $\pm 2\sigma$ of the ridgeline location will be tagged as members. The number of member galaxies is denoted as $N_{gals}^{0.5Mpc}$. The reason we apply the cut $\pm 2\sigma$ is that this is normally the place where the background likelihood dominates over cluster likelihood. The justification of doing this is that the member galaxies of the cluster should also cluster in color space. The two component Gaussian Mixture Model can reliably pick up the right peak in the color space as verified by our Monte Carlo simulation test (Hao et al., 2009).

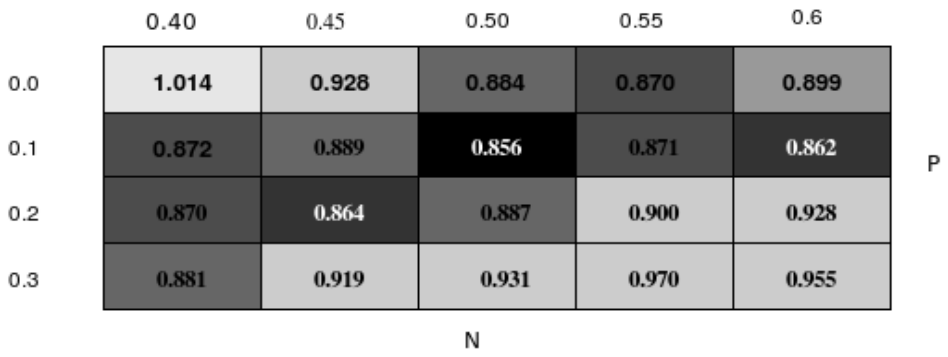


Figure 3.2. Comparing different scaling relations. The number in each box is the $\sigma_{\ln L_X}$ defined in (Rykoff et al., 2008; Rozo et al., 2008b). Based on the result, we choose $N = 0.5$ and $P = 0.1$.

3.5.2 Scaled Aperture Size and Richness

For selecting the appropriate aperture, we assume the scaling relation to be a power law, as motivated by Hansen et al. (2007).

$$R_{scale} = N(N_{gals}^{0.5Mpc})^P \tag{3.3}$$

where N and P are the normalization and power respectively. The crucial part is that we need to fix N and P in some way so that the resulting R_{scale} is best for the cluster in terms of S/N. The criteria we used here is the scatter in $\ln L_X$ at fixed richness estimated in a similar way for the maxBCG clusters (Rykoff et al., 2008; Rozo et al., 2008b). The smaller the scatter, the better the estimated richness and therefore the better the aperture. The performance of a grid N, P is shown in Figure 3.2. The best scaling relation is at $N = 0.5$ and $P = 0.1$, which yield $\sigma_{\ln L_X} = 0.86 \pm 0.02$ for the top 2000 clusters. Compared with $\sigma_{\ln L_X|N_{200}} = 0.96 \pm 0.03$ based on maxBCG scaled richness N_{200}, the current scaling is a big improvement.

Once we have the scaled aperture, we repeat the procedures for the fixed apertures, substituting the corresponding scaled aperture for 0.5 Mpc. The corresponding richness is denoted as N_{gal}^{scaled}.

3.6 Cluster Likelihood

Next we need to quantify how "likely" a galaxy is to be a BCG and measure its strength of clustering. We introduce two likelihoods to quantify these. It is worth noting that they are called likelihoods because they quantify the uncertainty and are normalized to unity. Mathematically, such a measure is equivalent to likelihood. They are not likelihoods from the frequentist's view and do not correspond to any underlying random process. The first likelihood is based on how close a candidate BCG's color is to the corresponding ridgeline color selected as described in 3.5.2. We introduce the following *cluster BCG likelihood* to quantify this.

$$L_{cluster}^{BCG} = \frac{1}{\sqrt{2\pi(\sigma_{gm}^2 + \sigma_c^2)}} \exp\left[-\frac{(c - c_{gm})^2}{2(\sigma_{gm}^2 + \sigma_c^2)} \right] \tag{3.4}$$

where σ_{gm} is the width of the Gaussian component corresponding to the cluster, and c_{gm} is the cluster's ridgeline color from the Gaussian component's peak location. c is the color of the candidate BCG and σ_c is the corresponding measurement error. We also require $|c - c_{gm}| \leq 2\sqrt{\sigma_{gm}^2 + \sigma_c^2}$ for the candidate BCG to be considered as a BCG.

We quantify the strength of clustering in the projected ra/dec plane by convolving the selected members with a radial kernel/profile. This is essentially a radially weighed number count of the members. Here, we choose the projected NFW profile (Bartelmann, 1996; Navarro et al., 1997; Koester et al., 2007b) as the radial kernel. It is worth noting that the type of kernel used is not as important as its scale, which has been revealed by statistical kernel density analyses (Silverman, 1986; Scott, 1992). Therefore, the specific kernel we use won't significantly bias the detection of clusters that deviate from the kernel shape. In this work, we will stick to the NFW radial kernel. We introduce the *clustering strength likelihood* as

$$L_{cluster}^{strength} = \sum_{k=1}^{N_g} \Sigma(x_k) \tag{3.5}$$

where N_g is the total number of member galaxies and

$$\Sigma(x) = \frac{2\rho_s r_s}{x^2 - 1} f(x), \tag{3.6}$$

$r_s = r_{200}/c$ is the the scale radius, ρ_s is the projected critical density, $x = r/r_s$ and

$$f(x) = \begin{cases} 1 - \frac{2}{\sqrt{x^2-1}} \tan^{-1}\sqrt{\frac{x-1}{x+1}} & x > 1 \\ 1 - \frac{2}{\sqrt{1-x^2}} \tanh^{-1}\sqrt{\frac{1-x}{x+1}} & x < 1 \\ 0 & x = 1 \\ 0 & x > 20. \end{cases} \tag{3.7}$$

The profile is truncated at $r = 100h^{-1}$ kpc to avoid divergence of $f(x)$ and we choose $r_s = 150$ kpc in our implementation. For details, refer to (Koester et al., 2007b).

Now, we have two likelihoods; the cluster BCG likelihood $L^{BCG}_{cluster}$ Eq.(3.4) and clustering strength likelihood $L^{strength}_{cluster}$ Eq.(3.5), which capture the color and spatial information of a cluster respectively. Now, we combine them together as a measure of how strong the clustering is and how likely the galaxy is a BCG.

$$L^{tot}_{cluster} = L^{BCG}_{cluster} \times L^{strength}_{cluster} \tag{3.8}$$

This likelihood is essentially a measure of the convolved density map modulated by the closeness of candidate BCG's color to the ridgeline color. We will use it as our major criterion for selecting clusters.

3.6.1 Luminosity Weighted Radial Density Likelihood

In addition to the likelihood we introduced in the previous section, we also measured another likelihood, the luminosity weighted radial likelihood $L^{strength}_{Lum}$. This likelihood is measured in a similar way as $L^{strength}_{cluster}$ except we attach a luminosity weight (W_{lum}) to each galaxy. The luminosity weight is determined by the ratio of each galaxy's i-band magnitude to the i-band magnitude corresponding to 0.4L* at the candidate cluster BCG's redshift.

$$L_{Lum}^{strength} = \sum_{k=1}^{N_g} \Sigma(x_k) \times W_{lum}(k) \tag{3.9}$$

The advantage of introducing such a measure is that its ratio to the non-luminosity weighted $L_{cluster}^{strength}$ is a good indicator of whether the candidate BCG is a star or a galaxy. It is a double check for the star/galaxy separation of the input catalog.

3.6.2 Implementation of the Algorithm

With all the quantities calculated from the above definitions, the implementation of the cluster selection is straightforward. There are basically four steps:

1. Preselecting BCG candidates.

2. For every galaxy in the candidate list, evaluate the total likelihood $L_{cluster}^{tot}$. During this process, the member galaxies are searched in the full galaxy catalog, not the candidate list.

3. Rank the candidate BCGs by total likelihood and remove those candidates from this list which are identified as members of another candidate BCG with higher total likelihood. In Figure 3.4, we show the distribution of likelihoods around a cluster. The BCGs with lower likelihood will be merged with the one with highest likelihood.

4. Repeat the above process and eventually obtain a list of BCGs and their cluster members. Based on the richness measured in 0.5 Mpc, we calculate a scaling R_{scaled} for every BCG. Then, repeat process 1) – 2) by changing the searching aperture to R_{scaled} from 0.5 Mpc. We then arrive at a final list of BCG members and BCGs with scaled richness N_{gals}^{scale}.

The procedures are summarized as a flowchart in Figure 3.3

3.6.3 Post Percolation Procedure

The above process is essentially a process of detecting the peaks of the smoothed density field with the height of the peaks measured by $L_{cluster}^{tot}$. This quantity is a

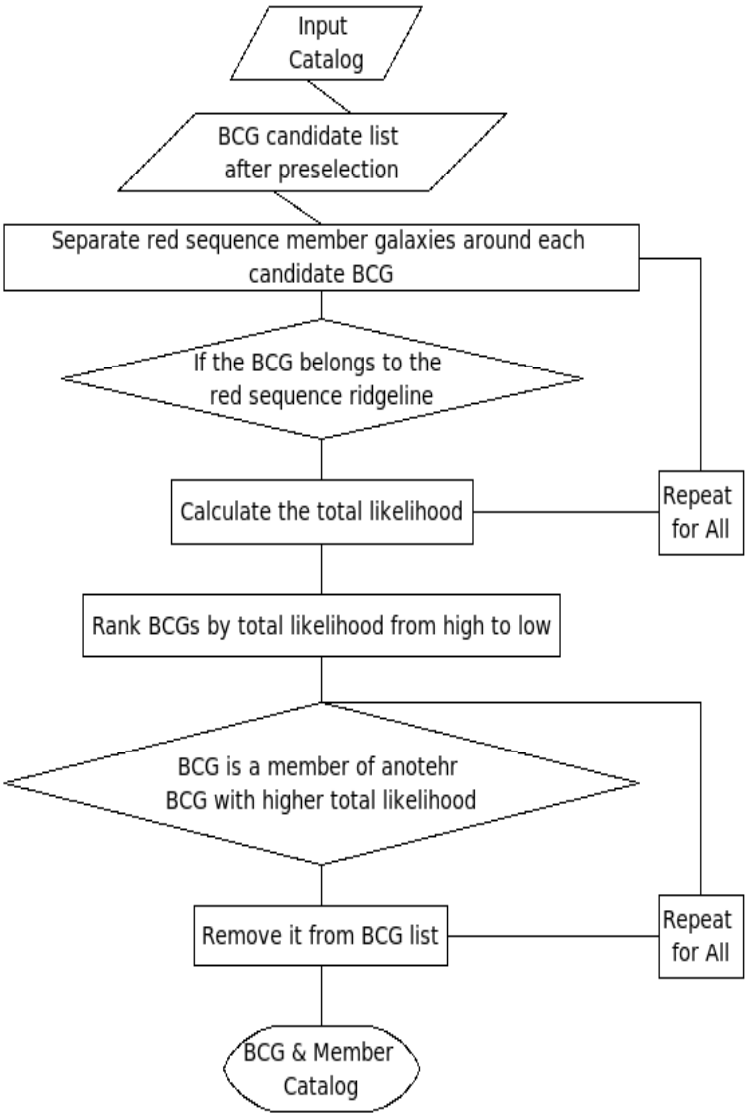


Figure 3.3. Flowchart for the implementation of the GMBCG algorithm

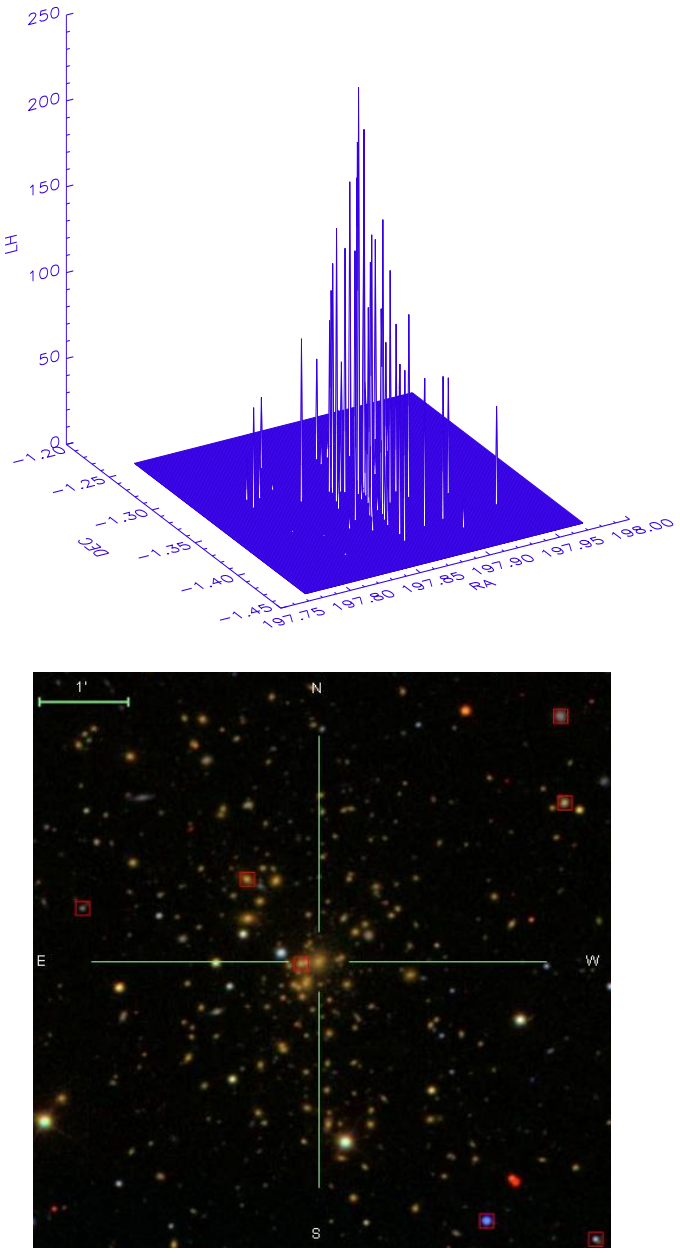


Figure 3.4. Upper panel shows candidate BCGs with lower likelihood will be merged into the cluster whose BCG has higher likelihood. Lower panel show the actual cluster (Abell 1689). The highest peak in the upper panel correspond to the brightest galaxy in the field.

combination of the peak height of density field $L_{cluster}^{strength}$ and $L_{cluster}^{BCG}$ which quantifies the closeness of the BCG's color to the corresponding ridgeline color. From the $L_{cluster}^{tot}$, we can get two pieces of information: how likely the target galaxy is a Brightest Cluster Galaxy and how strongly the galaxies are clustered around it. These two are not completely independent.

In our cluster finding process, the center of the cluster is assumed to be the brightest cluster galaxy. Therefore, it is possible that several peaks (quantified by $L_{cluster}^{tot}$) are identified in the field of a brightest cluster galaxy. We then need to blend several peaks and merge them into the same clusters using some criteria. We call this process post percolation. The major motivation for not directly blending the peaks during the cluster finding process is that we want to have some flexibility on how to merge the peaks. In many cases, the sub-peaks are indicators of potential sub-structures of clusters.

To be specific, we merge the peaks in the following way. For an identified peak (on a potential BCG), we identify a cylindrical region in the ra/dec plane and redshift space. The radius of the cylinder is specified by the radius of the cluster and the height is specified by its photoz ± 0.05. If another fainter candidate BCG falls within this cylinder, that fainter BCG peak will be merged into the brighter BCG peak in a similar way as shown in Figure 3.4, but with likelihood replaced by the I band magnitude. This is only one way to do the percolation, and may not be the optimal one. In reality, one needs to test this against some known clusters and see if the percolation scheme is over/under blending (Koester et al., 2007a).

3.7 Comparison with MaxBCG Algorithm

It is interesting to collect the major differences between the GMBCG and maxBCG algorithms (Koester et al., 2007b). maxBCG is a matched filter based algorithm with an additional filter from the red sequence colors. Using this algorithm, a large optical cluster catalog has been created (Koester et al., 2007a), which has high purity and completeness base on tests on both a Monte Carlo catalog and a N-body mock catalog.

The difference between GMBCG and maxBCG can be summarized as three major aspects:

1. Generally speaking, maxBCG is a generalized matched filter algorithm with the inclusion of a color filter in addition to radial and magnitude filters. It varies the filter at a grid of testing redshifts. The redshift at which the model filter maximizes the match with data is selected as the redshift of the cluster. GMBCG is not a matched filter like algorithm and it does not maximize the match for any filter. It uses a statistically well motivated mixture model to identify red sequence galaxies. The radial NFW kernel serves as a smoothing kernel rather than a model filter. Therefore, GMBCG will not bias against clusters that do not follow the assumed model filter.

2. maxBCG assumes a ridgeline model and fixed background for all clusters while GMBCG does not assume any model priori. It uses the well established mixture model to determine the ridgeline and background cluster by cluster. The advantage is that it automatically adjusts the cluster and background parameters across a wide redshift range.

3. In the maxBCG algorithm, the photozs of the clusters are estimated as a part of the execution of the algorithm. But in GMBCG, photozs are obtained from other methods such as neural networks, nearest neighbor polynomial, etc.

From the above comparisons, it should be clear that GMBCG is more easily extended to high redshift and less biased against atypical clusters.

CHAPTER 4

GMBCG Catalog For SDSS DR7

In this chapter, we apply the GMBCG algorithm to the latest data release of the Sloan Digital Sky Survey (data release 7, DR7 hereafter), and construct an optical cluster catalog of more than 53,000 rich clusters across the redshift range $0.1 \sim 0.5$. We also identify some especially rich high redshift clusters beyond redshift 0.5. We cross match the GMBCG clusters to X-ray clusters and maxBCG clusters, and test the completeness and purity of the catalog against a Monte Carlo catalog based on the DR7 data. In the following, I will introduce the details of how such a large catalog was assembled using the GMBCG algorithm.

4.1 SDSS DR7 Data

The Sloan Digital Sky Survey (SDSS) (York et al., 2000) is a multi-color digitized CCD imaging and spectroscopic sky survey, utilizing a dedicated 2.5-meter telescope at Apache Point Observatory, New Mexico. It has recently completed mapping over one quarter of the whole sky up to the medium redshift range in *ugriz* filters. DR7 is a mark of the completion of the original goals of the SDSS and the end of the phase known as SDSS-II (Abazajian & Sloan Digital Sky Survey, 2008). It includes a total imaging area of 11663 square degrees with 357 million unique objects identified.

In this paper, we will mainly detect clusters on the so called Legacy Survey area, which "provided a uniform, well-calibrated map in ugriz of more than 7,500 square degrees of the North Galactic Cap, and three stripes in the South Galactic Cap totaling 740 square degrees" (Abazajian & Sloan Digital Sky Survey, 2008). In Figure 4.1 we show the coverage map.

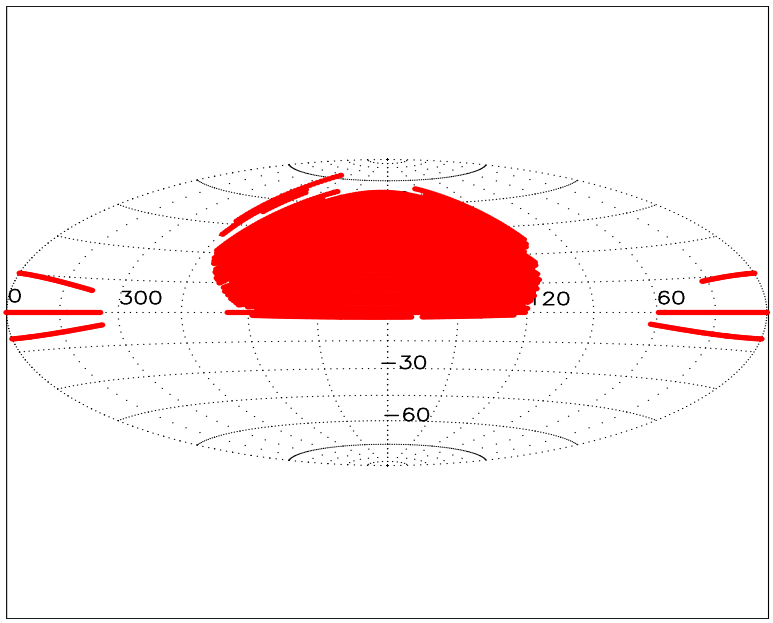


Figure 4.1. The photometric imaging coverage of the SDSS legacy survey. The figure is from the SDSS official website (www.sdss.org).

After running our GMBCG algorithm on the data, we assembled a large and approximately volume-limited cluster catalog with more than 53,000 clusters above a certain richness threshold and across a redshift range from 0.1 to 0.5 (main catalog). In addition, we identified some big clusters at redshifts above 0.5 (high redshift catalog), which are almost at the detection limit of SDSS data.

4.2 Input Catalog

To prepare an input galaxy catalog, we download the galaxy catalog from the Photo-Primary view with type set to 3 (galaxy) and i-band magnitude less than 21.0 from the Casjob database (http://casjobs.sdss.org/CasJobs/). Meanwhile, we also download the photoz table and cross match the objects to the galaxy catalog to attach photozs to each galaxy. In DR7, the photozs in the photoz table are calculated based on a nearest neighbor polynomial algorithm (Abazajian & Sloan Digital Sky Survey, 2008). Since we cross matched the photoz table to get photozs, the selection criteria used in the photoz table are also applied to our final input catalog.

In addition to the above selection requirements, we also throw away those galaxies

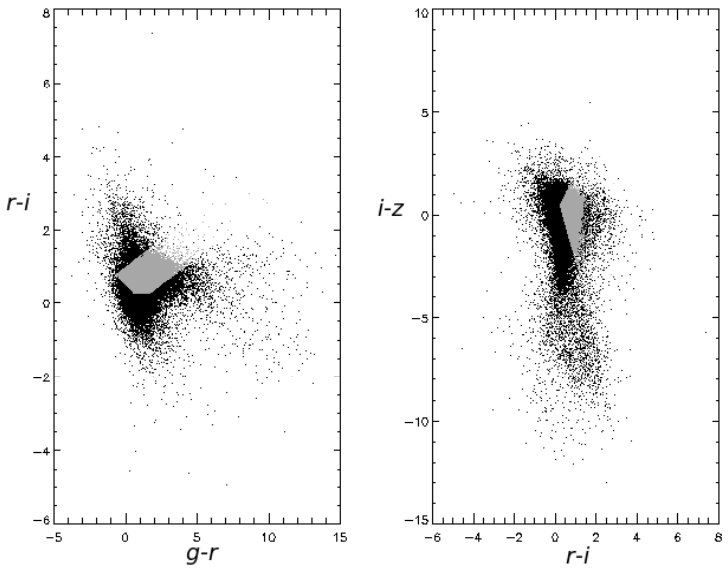


Figure 4.2. The BCG preselection in color - color space for the SDSS DR7 data. The gray parts are our preselected BCGs in color-color space.

with very bad measurements (photometric errors greater than 20 percent). We then made some cuts in color-color space to select the search list for candidate BCGs as described in section 3.4. The corresponding cuts in color-color space are shown in Figure 4.2. We apply very tolerant cuts which keep 70% of the total galaxies in our candidate BCG search list, eliminating only those with quite atypical colors.

After the above procedures, we prepare an input catalog for our cluster finder. It is worth noting that we did not apply any star/galaxy separation procedures other than the ones generated by the standard DR7 pipeline. This is a relatively tolerant selection that may be contaminated by occasional bright stars that are not well separated from galaxies. However, as we will show in what follows, we measured another quantity, "weighted_nfw_lh". By comparing this with the standard "nfw_lh" we can reliably eliminate any bright stars which make it into in our resulting catalog.

Tag Name in Catalog	Definition
OBJID	Unique ID of each galaxy in SDSS DR7
RA	Right Ascention
DEC	Declination
PHOTOZ	Photometric redshift from the photoz table in DR7
PHOTOZ_ERR	Errors of photoz
SPZ	Spectroscopic redshift
GMR	$g-r$ color[a]
GMR_ERR	Error of $g-r$ color
RMI	$r-i$ color
RMI_ERR	Error of $r-i$ color
MODEL_COUNTS	Model magnitude[b]
MODEL_COUNTS_ERR	Error of model magnitude
LIM_I	I band magnitude corresponding $0.4L^*$ at photoz
BCGLH	Cluster BCG likelihood, $L_{cluster}^{BCG}$
NFW_LH	Clustering strength likelihood, $L_{cluster}^{strength}$
WEIGHTED_NFW_LH	Luminosity weighted radial density likelihood, $L_{Lum}^{strength}$
LH	Total cluster likelihood, $L_{cluster}^{tot}$
NGALS	Number of member galaxies inside 0.5 Mpc circle from BCG
GM_SCALEDR	The scaled aperture R_{scale}
GM_SCALED_NGALS	Number of member galaxies inside GM_SCALEDR from BCG
GM_GMR	Location of the Gaussian component corresponding to cluster red sequence in $g-r$ color
GM_GMR_WDH	Width of the Gaussian component corresponding to cluster red sequence in $g-r$ color
GM_RMI	Location of the Gaussian component corresponding to cluster red sequence in $r-i$ color
GM_RMI_WDH	Width of the Gaussian component corresponding to cluster red sequence in $r-i$ color
GM_NN	Number of Gaussian Mixtures
GM_SCALED_NGALS_UNIF	Rescaled richness, see section 4.3.2
NFW_LH_UNIF	Rescaled NFW_LH, see section 4.3.2

Table 4.1. The tags in the cluster catalog

[a]All colors are calculated using model magnitude

[b]For details, see http://www.sdss.org/DR7/algorithms/photometry.html

4.3 Cluster Catalog

4.3.1 Catalog Facts

We apply the GMBCG algorithm to the input catalog and generate a full catalog of galaxy clusters for the SDSS DR7. We search clusters from redshift 0.05 to 0.60, but only include in the main final catalog the redshift range $0.1 \sim 0.5$ to keep the quality of the catalog high. In Table 4.3.1, we list the tags in the final cluster catalog and their corresponding definitions. There are about 53,000 clusters with the rescaled $L_{cluster}^{strength}$ (see next section) greater than 8.5 after applying the post percolation procedure. The redshift distribution of the clusters are shown in Figure 4.3 and the richness distribution is shown in Figure 4.4. Example clusters at different redshift are shown in Fig. 4.3.1.

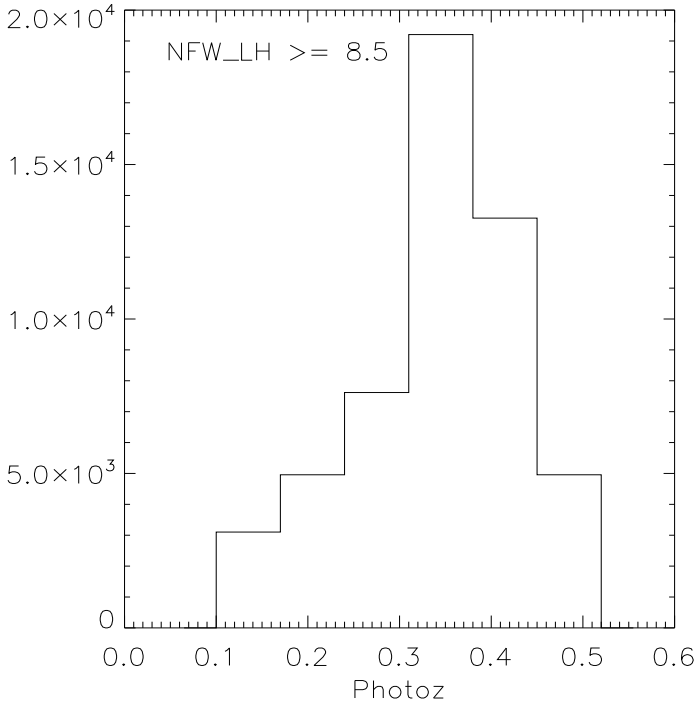


Figure 4.3. Redshift distribution of GMBCG clusters with rescaled NFW_LH greater or equal to 8.5

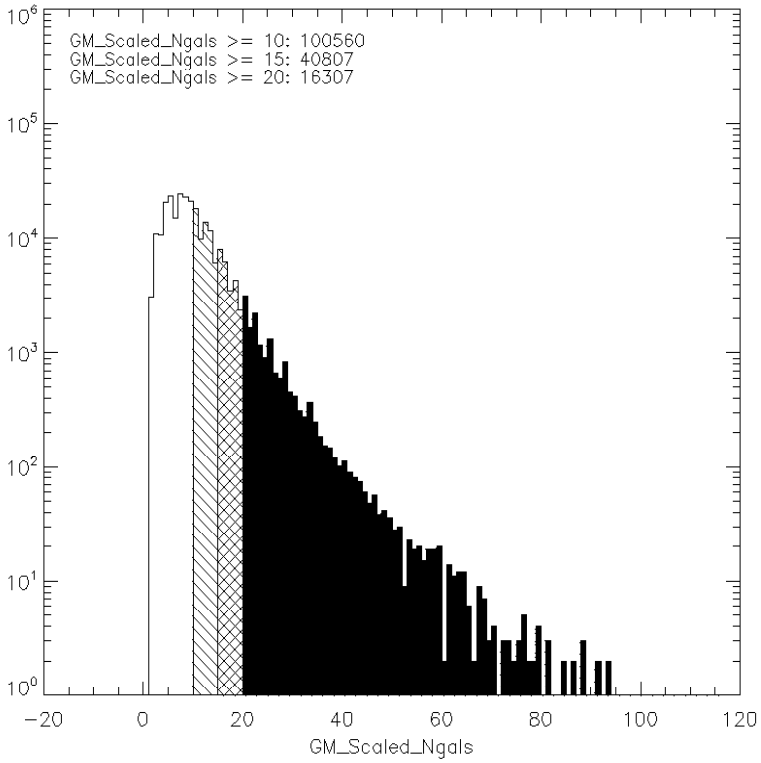
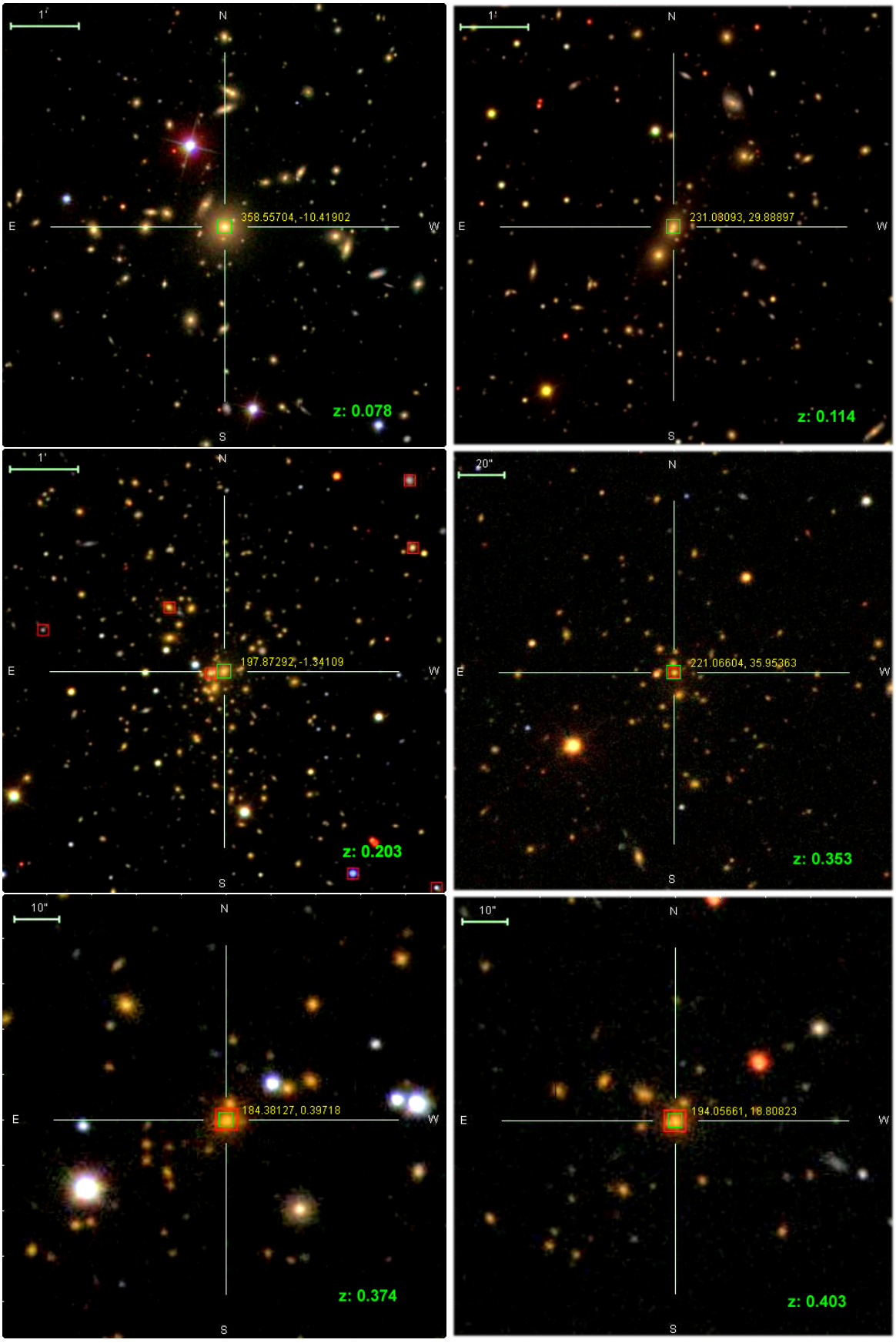
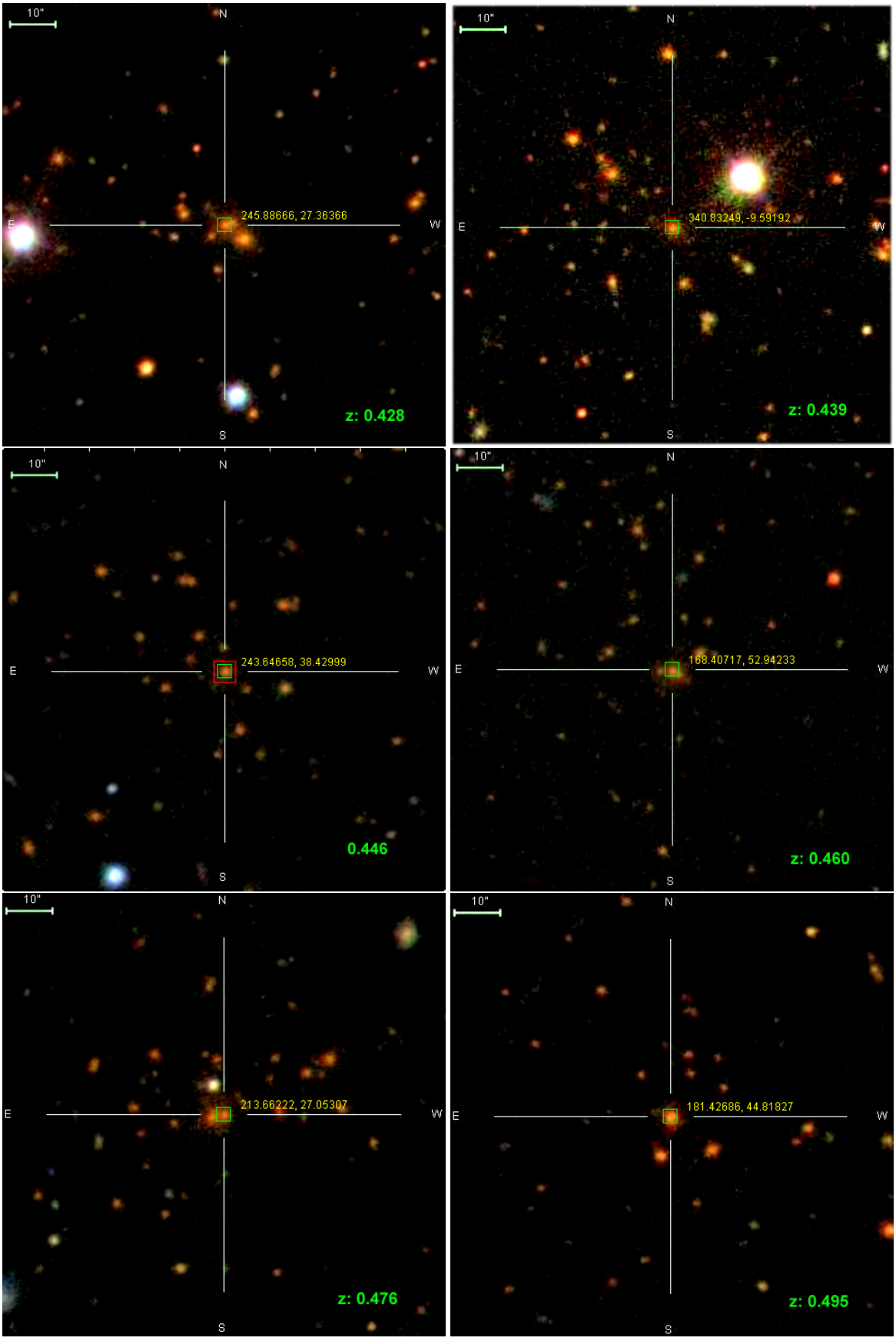


Figure 4.4. Richness distribution of GMBCG clusters. The richness is a rescaled richness (see next section).


1'
N
E
W
S
358.55704, -10.41902
z: 0.078
1'
N
E
W
S
231.08093, 29.88897
z: 0.114
1'
N
E
W
S
197.87292, -1.34109
z: 0.203
20"
N
E
W
S
221.06604, 35.95363
z: 0.353
10"
N
E
W
S
184.38127, 0.39718
z: 0.374
10"
N
E
W
S
194.05661, 18.80823
z: 0.403


10"
N
E
W
S
245.88666, 27.36366
z: 0.428
10"
N
E
W
S
340.83249, -9.59192
z: 0.439
10"
N
E
W
S
243.64658, 38.42999
0.446
10"
N
E
W
S
168.40717, 52.94233
z: 0.460
10"
N
E
W
S
213.66222, 27.05307
z: 0.476
10"
N
E
W
S
181.42686, 44.81827
z: 0.495

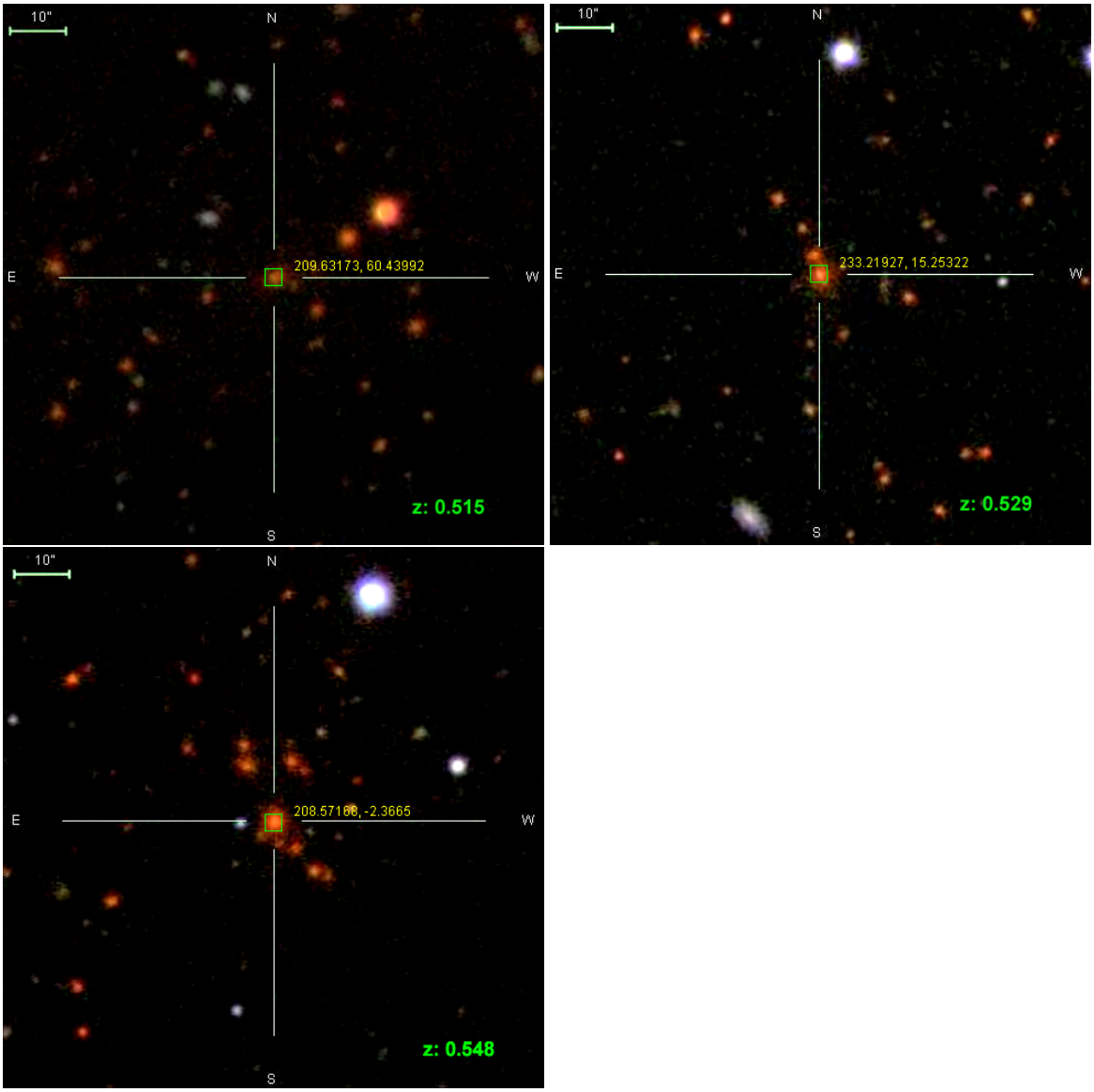


Figure 4.5. The above images are some of our detected clusters from SDSS DR7. The green letters on the images indicate the BCGs' spectroscopic redshift.

4.3.2 Richness Re-scaling

In the redshift range $0.1 \sim 0.5$, only the $g-r$ and $r-i$ ridgeline colors are used, and the switch between them is determined by the *photoz* of the target galaxy. Since we measure the richness by counting the number of galaxies falling within 2σ of the ridgeline, the resulting richness from $g-r$ and $r-i$ are not directly comparable. In part this is due to a changing degree of background contamination as the ridgeline moves through color space (see Figure 4.8). Generally, the richness measured from $r-i$ is higher than that measured from $g-r$. To make the richness more consistent across the whole redshift range, we rescale the richness measured from $r-i$ color. Clearly, mass is the only true parameter with which we should relate the two different richnesses. Therefore, a complete resolution of this problem requires to map out the mass-richness relation for both richnesses and then bring them to the same ground. However, for the moment, we will use a simple first order approach. That is, we require the statistical distribution of richness measured from two different colors to be the same. The scaling relation that matches the two distributions is clearly non-linear. The procedure we use is to match the richness at different percentile bins of the two distributions and re-scale them linearly in each bin. Then, we fit a polynomial to the scaling relation across all the bins and get a "continuous" scaling relation. The richness from $r-i$ color will be re-scaled by this continuous scaling relation and the results are shown in Figure 4.6 and Figure 4.7. Since the scaling relation is monotonously increasing, the scaled richness will not alter the cluster ranking based on the original richness in the $r-i$ ridgeline region.

4.3.3 Bimodality in Color Space

As we have shown in previous sections, the color distribution around a cluster normally shows bimodality. But there are situations where the cluster is so big that its members completely dominate the field within the aperture we impose. In this case, the color distribution may be unimodal. In our implementation of the GMBCG algorithm, we also consider this situation as a potential cluster as long as the width of the dominant unimodal distribution is narrow enough.

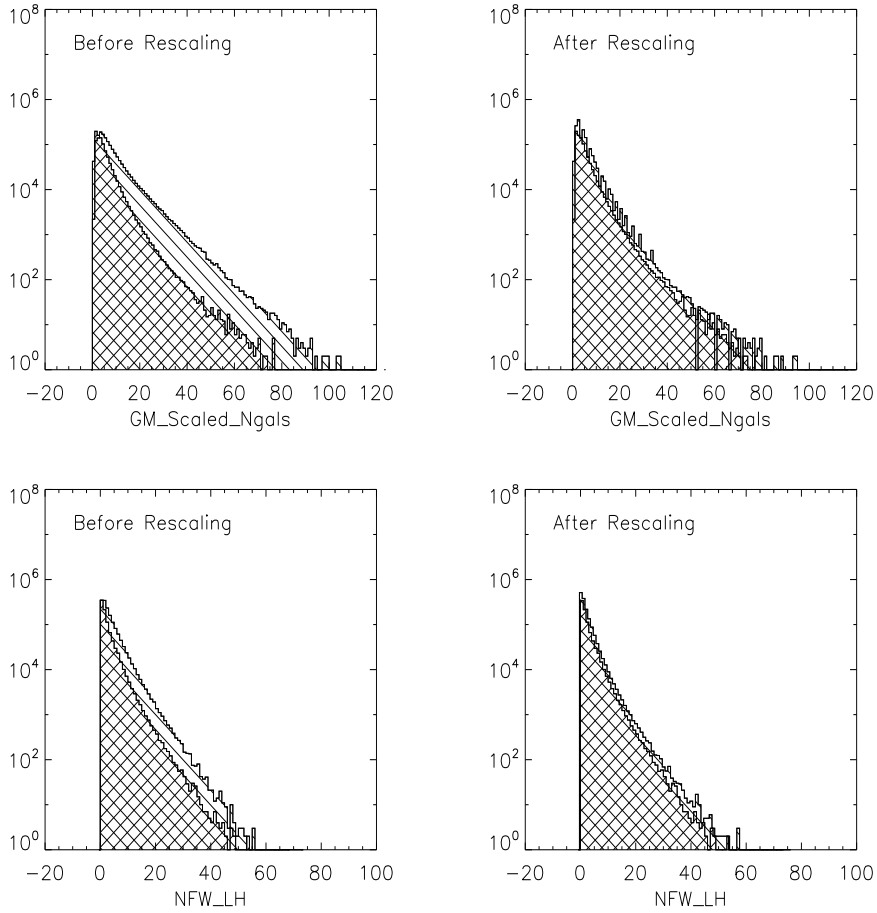


Figure 4.6. The re-scaling of the richness and nfw_lh measured based on $g-r$ color (histogram with filled line at 45°) and $r-i$ color (histogram with filled line at $-45°$). Corresponding nonlinear re-scaling relations are multiplied to the richness and nfw_lh based on $r-i$ color. This procedure roughly matches the two distribution.

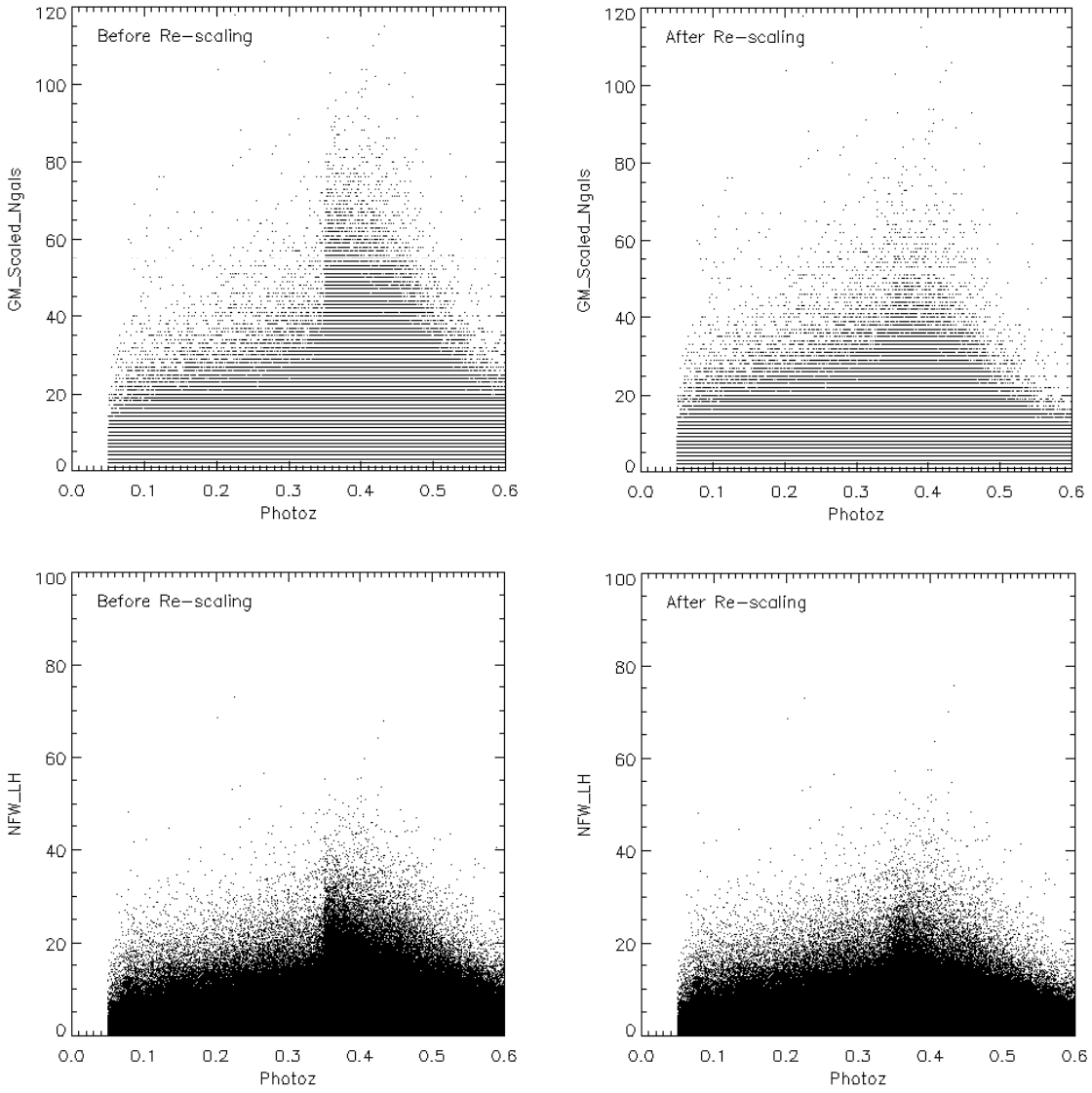


Figure 4.7. The scatter plot of photoz vs. richness and nfw_lh before and after re-scaling. Clearly, after re-scaling, the bump around 0.35 due to the change of ridgeline become significantly damped.

In the case of a bimodal color distribution, the separation between the two Gaussian components will vary as redshift changes, leading to different degrees of overlap. This overlap of the two Gaussian components partially tells us the fraction of projected galaxies when we impose the color cuts on the red sequence galaxies. Therefore, the richness for the clusters should be appropriately weighted to account for the projection. In Figure 4.8, we show the color distribution of clusters at different redshifts. From the plot, the 2σ cut we imposed for selecting red sequence members is almost where the likelihood of red sequence galaxy becomes equal to that of background/blue galaxies, which substantiates our previous choice.

4.4 Evaluating the Catalog

Any cluster finding algorithm can be evaluated by two simple criteria: completeness and purity. Completeness quantifies whether the cluster finder can find all true clusters, while purity quantifies whether the clusters found by the cluster finder are real clusters. However, calculating the completeness and purity requires that we know in advance what is a true cluster. This issue can only be completely resolved when we have a high resolution simulation catalog that can properly reflect the galaxies' colors as well as their interaction with dark matter halos. However, this is a very difficult task, complicated by various factors such as limitations of the resolution of simulation, unknown behaviors of galaxies at high redshift, unknown evolution of all types of galaxies and distribution, etc.

In this section, we introduced a simple and robust Monte Carlo Mock Catalog to test our cluster finder. The result can, at least, tell us the purity and completeness of our cluster catalog with respect to the model clusters we put in.

4.4.1 Monte Carlo Mock Catalog

There are four steps to create the Monte Carlo catalog:

1. *The base catalog preparation*: we pull out 25 stripes of galaxy catalog (about 70%) from DR7 of the Sloan Digital Sky Survey and remove all the known cluster galaxies based on our cluster catalog. We shuffle the remaining galaxies

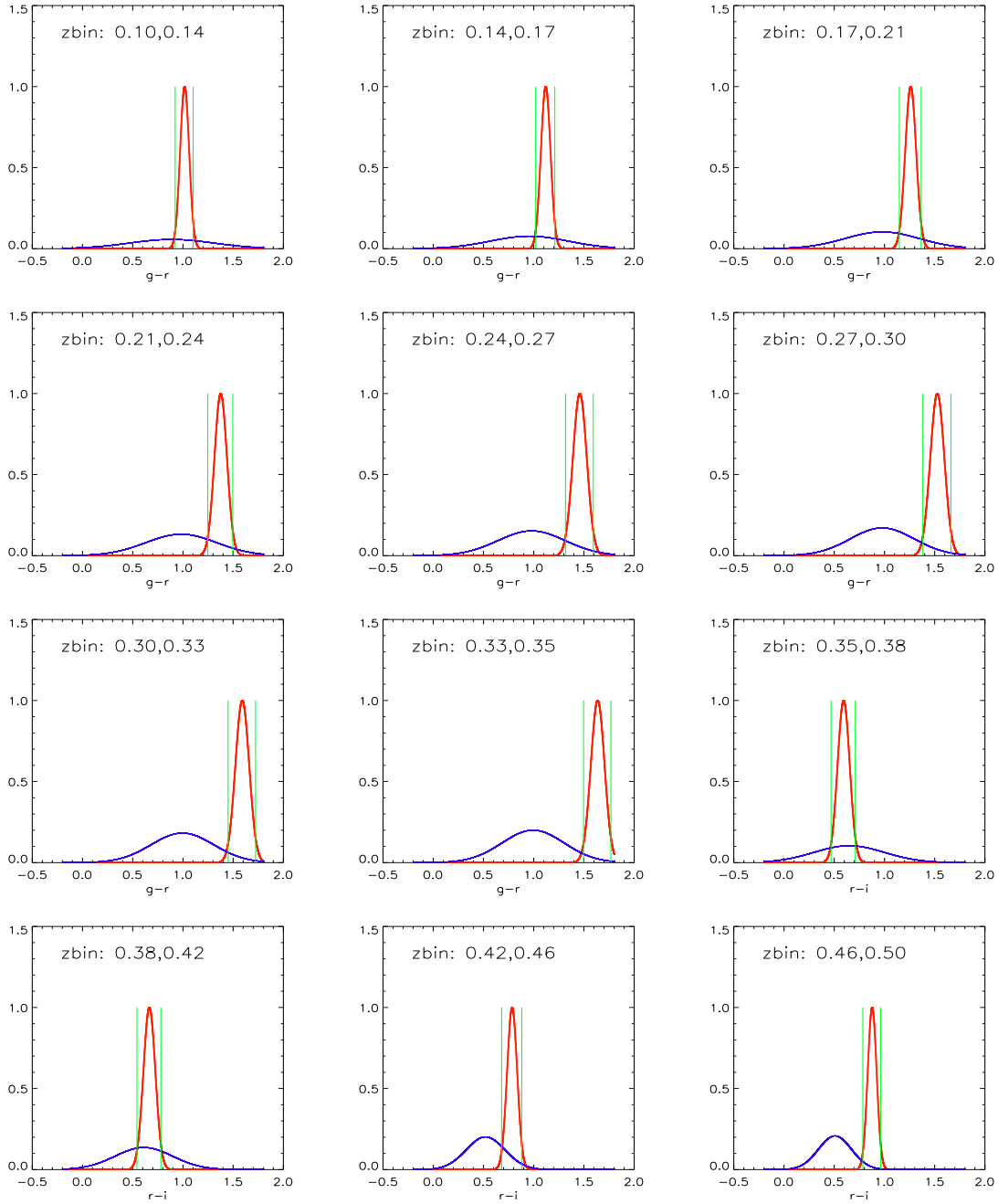


Figure 4.8. The bimodal distribution of red sequence galaxy colors and background/blue galaxies. The results are based on the average results in each redshift bin as indicated in the plots. The green vertical lines are the 2 σ clip of the red sequence peak.

by using random ra and dec to replace their true ra and dec, while keeping their colors and other properties unchanged.

2. *Model cluster selection*: We pick up 51 big clusters whose redshift ranges from 0.1 to 0.5 from our cluster catalog and confirm them by visually checking their SDSS images. These clusters are very big and we are quite sure they are real clusters. Each cluster has a BCG and about 30-100 member galaxies.

3. *Model mock clusters resampling*: Pick up a BCG randomly from the 51 model clusters and then select a fixed number of member galaxies from the corresponding model cluster's members. The fixed number is randomly chosen from [10, 15, 20, 25, 30, 35, 40, 45, 50]. In this way, we can generate a resampled model cluster of a given richness.

4. *Putting resampled model clusters into base catalog:* For every stripe of the base catalog, we select 500 resampled model clusters and put them into the base catalog so that their corresponding BCGs replace 500 randomly chosen galaxies in the base catalog. Then, we will have a Monte Carlo catalog that are based on the real photometry of the SDSS DR7 data.

The above procedures are summarized in the flowchart in Figure 4.9. Through this procedure, we can produce very realistic mock clusters in a realistic setting. However, since we sample from the big clusters to generate small clusters, the brightness of the smaller clusters we generated might be biased and appear brighter than they should be. Will this affect our results a lot? The answer is no. Because we assume the BCG as cluster center and as long as this property is reserved in the mock clusters, we will be safe. The increased brightness of the smaller mock clusters will only affect the S/N of the color of each galaxy and this will not have a big impact since we have already modeled the measurement errors using the Error Corrected Gaussian Mixture Model. So, this artifact in the mock catalog will not affect our objective of testing the completeness and purity of our GMBCG catalog.

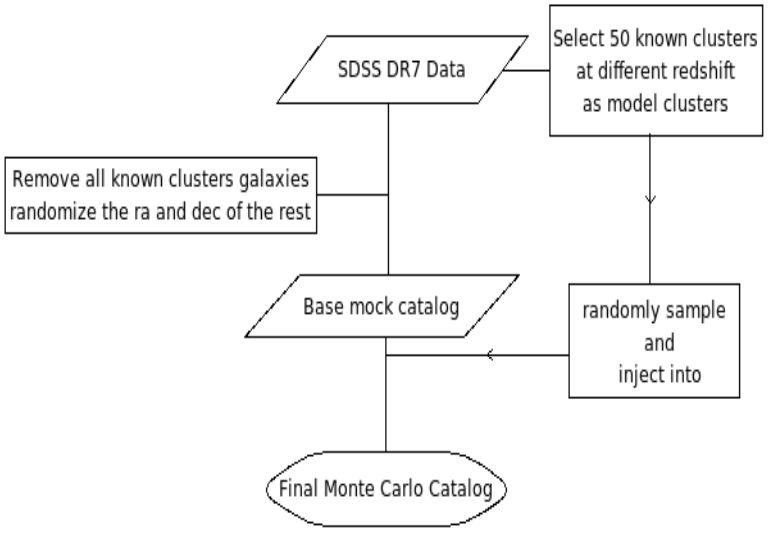


Figure 4.9. The flowchart of generating the mock catalog

Another unrealistic artifact is that we randomize the field galaxies. In the real world, their distribution is not so random. However, their correlation is very weak compared to the correlation around clusters. Therefore, treating it as a Poisson random distribution should not have a big impact on our cluster detection, but will affect the richness estimate. In this section, we only consider if we can detect the clusters we put there and not the richness correlation. So, for this purpose, the Monte Carlo catalog should suffice.

4.4.2 Completeness and Purity

To test the completeness and purity of our cluster finder, we run it on the Monte Carlo catalog created above. Then, we cross match the detected clusters and the model clusters using a simple cylinder matching, i.e. searching in a cylinder of 0.5 Mpc in radius and ± 0.02 in redshift (this is a very small cylinder). At a given redshift bin and above a given N_{gal}, if we denote the number of model clusters that are matched to the found clusters by $N_{model}^{match}(z, N_{gal})$, the total number of model clusters by $N_{model}(z, N_{gal})$, the number of found clusters that are matched to model clusters by $N_{found}^{match}(z, N_{gal}^{scaled})$ and the total number of found clusters by $N_{found}(z, N_{gal}^{scaled})$, the

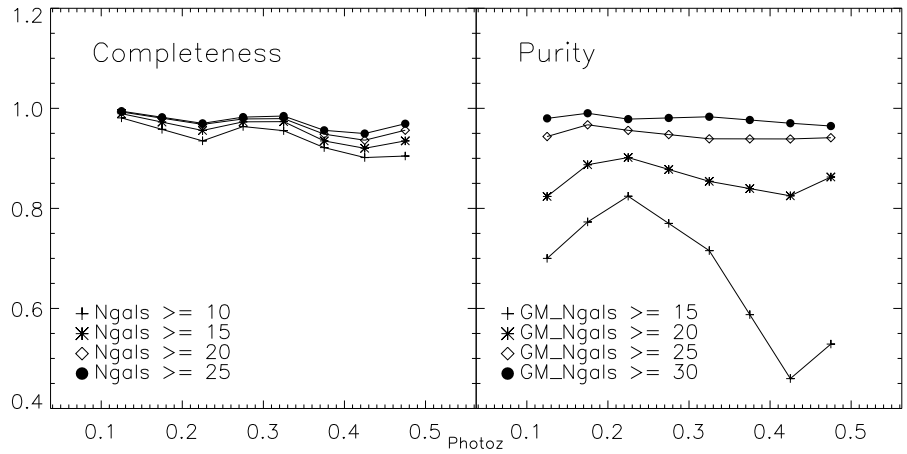


Figure 4.10. The completeness and purity of the GMBCG catalog based on the Monte Carlo Catalog. In the completeness plot, N_{gal} is the number of member galaxies of our input model clusters. In the purity plot, N_{gal}^{scaled} is the number of member galaxies measured by the cluster finder. Since there are residual red galaxies in the catalog before we put clusters in, the measured cluster richness (N_{gal}^{scaled}) is generally higher than our input richness

completeness and purity can then be defined as

$$completeness = \frac{N_{model}^{match}(z, N_{gal})}{N_{model}(z, N_{gal})} \tag{4.1}$$

$$purity = \frac{N_{found}^{match}(z, N_{gal}^{scaled})}{N_{found}(z, N_{gal}^{scaled})} \tag{4.2}$$

The results of the completeness and purity are plotted in Figure 4.10. The plot show that the GMBCG algorithm can yield a highly complete and pure cluster catalog.

4.5 Cross-Matching of GMBCG to ROSAT X-ray Clusters

Optical identification of peaks in the galaxy distribution represents only one of many methods used to find clusters. Other observables employed in cluster detection include thermal emission of x-rays from the hot intracluster medium, weak-lensing distortion of background sources, and the Sunyaev-Zeldovich effect on the cosmic microwave background. Each method has certain advantages and disadvantages along with a unique proxy for the mass of a cluster, which can be used to probe cosmological

constraints. Therefore, it is important that our cluster finding algorithm be able to detect those clusters found by alternative means. X-ray cluster catalogs are the most appealing candidate for exploring this question. Numerous x-ray catalogs exist with large sky coverage overlapping the DR7 survey area. Follow up optical examination is frequently performed on these catalogs to confirm their identity as clusters and to obtain accurate redshifts.

Just as when matching to optically identified catalogs, complications do arise. It is not always the case that the BCG lies exactly on the X-ray peak. There exists significant scatter in the x-ray luminosity-richness relation (Rykoff et al., 2008). Furthermore, the DR7 catalog contains clusters down to a richness threshold much lower then current x-ray catalogs can detect. The main goal of this exercise is to test the extent to which our algorithm is able to identify the brightest x-ray clusters.

We compare the DR7 catalog to three x-ray identified cluster catalogs: NORAS (Böhringer et al., 2000), REFLEX (Böhringer et al., 2004) and 400 deg^2 (Burenin et al., 2007). NORAS and REFLEX are composed of clusters identified from extended sources on the ROSAT all-sky survey x-ray maps. Together they cover the northern and southern galactic caps and are flux limited at 3×10^{-12} ergs s^{-1}cm^{-2} in the 0.1 - 2.4 KeV energy band. The 400 deg^2 catalog is composed of serendipitous clusters found in the high galactic latitude ROSAT pointings. It is flux limited at 1.4 s^{-1}cm^{-2} in the 0.5 - 2.0 KeV energy band. Sources from all three catalogs have been confirmed as clusters through follow up optical identification. Combining these catalogs yields 229 unique clusters in the survey area spanned by DR7.

A cylindrical search is performed on the combined x-ray catalogs in order to determine if these clusters were found by the GMBCG algorithm. We consider two clusters a match if they have a physical separation in the projected plane $sep < 2.0$ Mpc and a redshift difference $|z_{xray} - z_{photo}| < 0.05$. By this criteria, 222 out of 229 X-ray clusters are matched with at least one GMBCG cluster. As pointed out in Koester et al. (2007a), when the matching separation is greater than 1 Mpc, the matches are often chance matchs. Therefore, only matches at separations less than 0.5 Mpc are reliable matches. On the other hand, the GMBCG clusters are at different

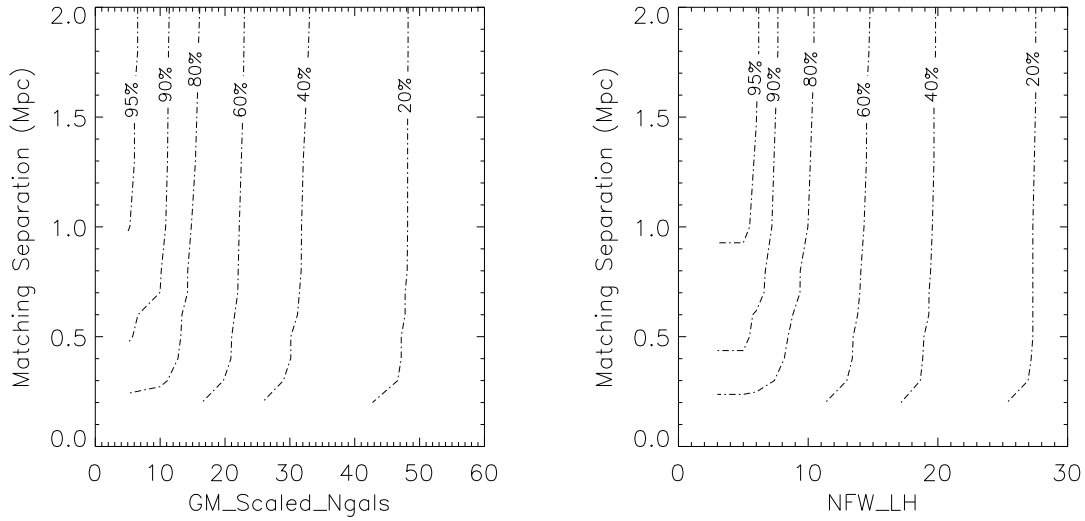


Figure 4.11. Contour of the matching ratio to ROSAT clusters for different separation and richness/nfw_lh. The percentages are calculated by the ratio of the matched clusters with matching separation less than the point on the contour while richness/nfw_lh greater than the point on the contour.

confidence levels if they have different N_{gal}^{scaled} or clustering strength ($L_{cluster}^{strength}$). So, the general notion is that the bigger the cluster and the shorter the separation, the higher probability of a true match. The rule of thumb for good clusters is $N_{gal}^{scaled} \geq 10$ or ($L_{cluster}^{strength} \geq 6$. In Figure 4.11 we show the contour plots of the matching ratio in terms of separation and cluster richness/clustering strength. The results show that we can reliably recover 90% of the X-ray clusters (separation is less than 0.5 Mpc and richness is greater than 10.)

In Figure 4.12, we show two high L_x ROSAT clusters with and without optical clusters matched. The one without a match is clearly not a cluster. The other non-matched ROSAT clusters are in similar situation, i.e. they are likely to be incorrectly identified as as X-ray clusters. In Figure 4.13, we listed 6 such clusters.

4.6 Cross-Matching to MaxBCG Clusters

As a further test on the completeness of the GMBCG Catalog, we make a comparison to the maxBCG catalog (Koester et al., 2007a). The maxBCG catalog consists of 13,827 clusters in the redshift range $0.1 < z < 0.3$ with a lower threshold on richness

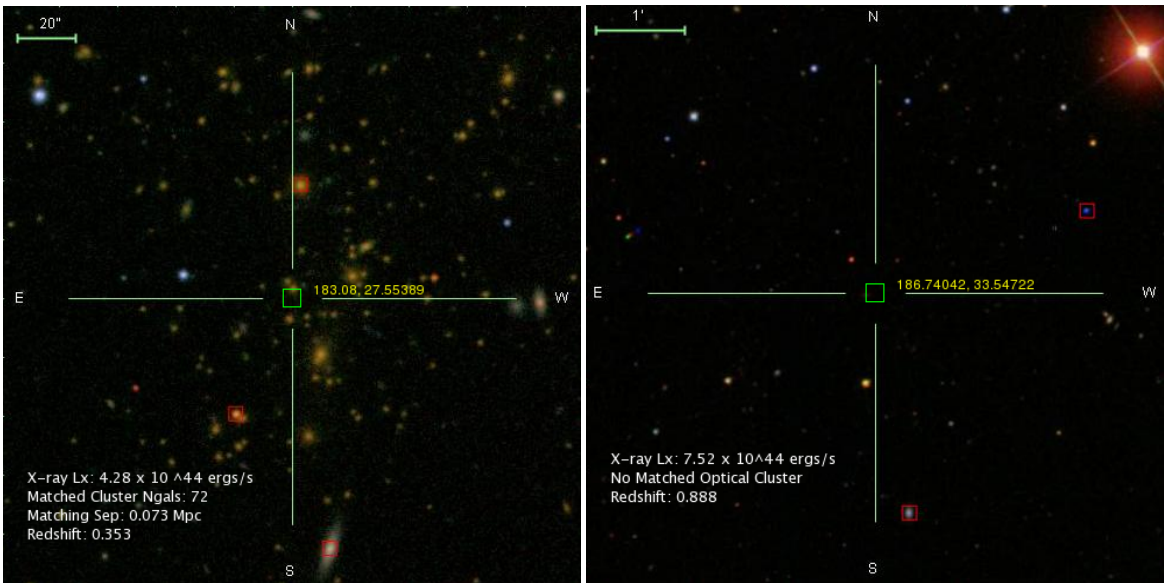


Figure 4.12. The left panel shows a high L_X ROSAT that has an optical cluster match. The right panel show another high L_X ROSAT that does not have an optical cluster match

set at $N_{200} = 10$. It is derived from DR5 of the Sloan Digital Sky Survey and covers a slightly smaller area than the new GMBCG catalog.

Several complications arise in the process of performing cluster-to-cluster matches between catalogs, namely redshift uncertainties, centering differences between the two algorithms, and scatter in the richness measurements. Although many similarities exist between the maxBCG and GMBCG algorithms, it is not always the case that they choose the same central galaxy for a given cluster. When matching clusters a careful cut must then be made in the two-dimensional physical separation in order to allow for this centering difference, while at the same time minimizing unreal matches due to random projection. Uncertainty in the photometric redshifts can yield a similar problem along the line of sight because a cut in $\Delta z = |z_{maxBCG} - z_{GMBCG}|$ must be made that accommodates these errors. Finally, the richness measurements themselves have large scatter, i.e. clusters that appear in one catalog may be given a richness value below threshold in the other and be unavailable for match. In what follows we briefly examine each of these problems in considering the best matching scheme to implement in order to quantify the agreement between the GMBCG catalog and the maxBCG catalog.

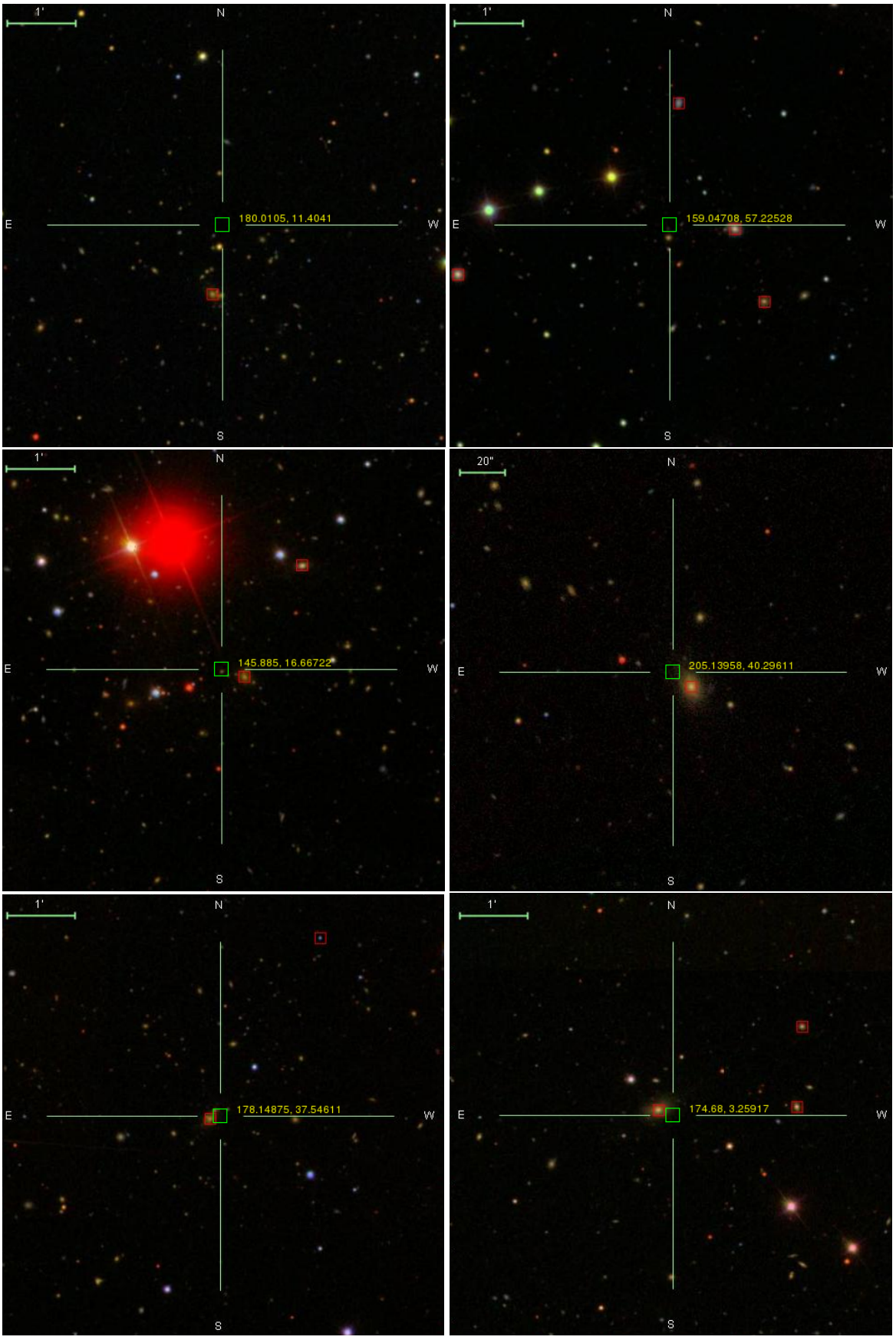


Figure 4.13. Six ROSAT clusters that do not have any matched optical clusters.

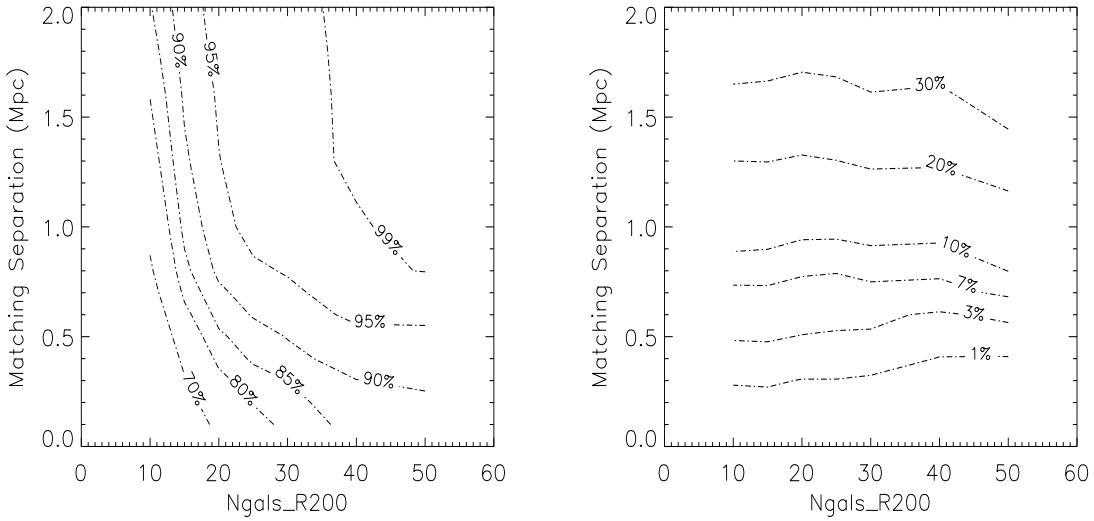


Figure 4.14. Left panel is the contour of matching fraction of the maxBCG clusters to the GMBCG clusters at different richness (Ngals_R200) and different separations. In the right panel, random points are used instead of the GMBCG clusters.

The uncertainty in redshift estimates for maxBCG clusters is $\sigma_z \sim 0.015$ (Koester et al., 2007a). In the GMBCG catalog, the uncertainty of the photozs at redshift below 0.3 is ~ 0.02 (Abazajian & Sloan Digital Sky Survey, 2008). Therefore, a redshift difference of ~ 0.05 between the two catalogs is an appropriate selection window for the matching. As for the radial separation, given the fact that the maxBCG clusters are blended within a separation of $R200 \sim 1.0 - 2.0$ Mpc (Koester et al., 2007b), a radial separation of ~ 2.0 Mpc is appropriate for our matching search. Similar as matching to X-ray clusters, the smaller the matching separation, the higher the probability of real matches. But it is a little more complicated in the current case because the lower the richness of the maxBCG clusters, the less likely they are true clusters (for the massive X-ray clusters, we are more certain they are real clusters). Therefore, we make the following contour plots about the matching ratio with respect to both the separation and the richness of maxBCG clusters.

Such a matching yields that 11,754 out of 13,823 ($\sim 85\%$) clusters in maxBCG catalog have a match in the blended GMBCG catalog where a threshold of $L_{cluster}^{strength} \geq 5$ is applied. If we loosen this threshold and take into account the intrinsic scatter in the richness relations of the two catalogs, 13,789 out of 13,823 ($\sim 99.8\%$) maxBCG

clusters match to GMBCG clusters. Those non-matched clusters are all at low richness Moreover, 8,979 clusters of the 13,789 matched clusters ($\sim 65.1\%$) have identical BCGs identified from the GMBCG catalog. In the left panel of Figure 4.14, we show the matching fraction of the GMBCG clusters to maxBCG clusters at different richness (maxBCG clusters) and separation. As a comparison, we make a similar plot by matching random points to maxBCG clusters.

CHAPTER 5

Statistical Methods and Precision Cosmology

"If your experiment needs statistics, you ought to have done a better experiment."

Lord Ernest Rutherford

One of the major goals in Science is to build models to describe large amounts of observations/data with fewer parameters. The way we can verify if our model is reasonable is to see if the model can reproduce the data in a statistically non-distinguishable way. Therefore, precision cosmology requires not only precision measurements, but also precision data analysis methods.

Astrophysics experiments are different from other laboratory experiments in that we cannot get what we wish, we can only get what the Universe provides. In addition to this disadvantage, we also suffer from the fact that most measurements are a combination of intrinsic scatter and measurement errors, which are in general hard to separate without some assumptions. Therefore, appropriate statistical methods are especially important for astronomical data analysis. In this chapter, I am going to describe some of my work related to statistical methods in data analysis. This works is not yet systematic nor complete, but it nonetheless leads to some very interesting applications.

5.1 Robust Fitting

The widely used scheme for fitting astronomical data is the maximum likelihood estimation (MLE) scheme, in which the "best" estimates of model parameters should maximize the likelihood of the parameters given the data, i.e. $L(\boldsymbol{\theta}|\text{data})$. To get

correct parameter estimates, one needs to correctly model the data and defines an appropriate $L(\boldsymbol{\theta}|\text{data})$ to be maximized. In practice, most analyses assume Gaussianity of the residual distribution and therefore take the likelihood $L(\boldsymbol{\theta}|\text{data})$ to be Gaussian. Maximizing such a likelihood function is mathematically equivalent to the well known least square estimation method (LSE, or weighted least square estimation, WLSE). Despite its wide application, one must bear in his mind the disadvantage of LSE – it is highly sensitive to outliers in the data. Statistically, the definition of an outlier is tricky and model dependent. That is, if you fit the data with model A, then the outliers may not be outliers if you fit the data with model B. In physics, models have been generally fixed by physics considerations and what we are concerned with is parameter estimation. Some outliers in data may alter the constraints on certain parameters, which will lead to misinterpretation of the physics if the parameters are in critical regions. There are two ways to get out of this trouble. First, scrutinize the data carefully and remove the outliers. With luck, you can sucessfully remove outliers instead of data points corresponding to interesting physics. Second, use more robust and outlier insensitive methods to fit the data.

In this section, I will introduce a robust analysis based on SNIa distance modulus and redshift data, though in principle the method is applicable to any other astronomical data analysis without increase of computation if using the Markov Chain Monte Carlo (MCMC) method. There are two major reasons for choosing SNIa data for this analysis: the physics model for the Supernovae distance modulus and redshift relation is straightforward and well understood and the SNIa data are delicate with many uncertainties entering into the calibrations and seeding the potential of outliers.

Our main purpose in this section is not to lay down precise new constraints on the cosmological parameters. Instead, we will focus on the demonstration of the differences in parameter estimates from LSE and robust methods. Therefore, we do not impose any prior on Hubble parameter from other experiments.

The sensitivity of the LSE method to outliers mainly stems from the fact that its merit function is of second order in the residuals. To be specific, we demonstrate this using the SNIa distance modulus ($\{\mu_i\}$), redshift ($\{z_i\}$) and measurement error

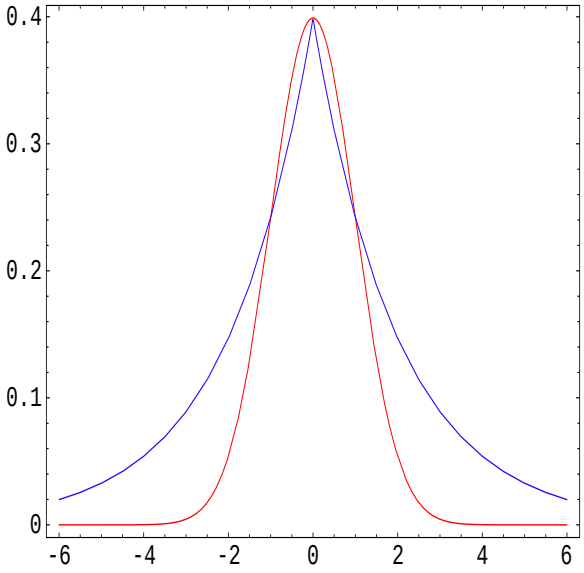


Figure 5.1. Gaussian distribution (red) and two-sided exponential distribution (blue).

$(\{\sigma_i\})$

$$\chi^2 = \sum_i \left[\frac{\mu_i - \mu(\boldsymbol{\theta}; z_i)}{\sigma_i} \right]^2 \tag{5.1}$$

where $\mu(\boldsymbol{\theta}; z_i)$ is the distance modulus predicted from a given cosmological model with cosmological parameters $\boldsymbol{\theta}$ at redshift z_i. This corresponds to a likelihood of parameters given the data

$$L(\boldsymbol{\theta}|\{z_i\}, \{\mu_i\}, \{\sigma_i\}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[- \frac{(\mu_i - \mu(\boldsymbol{\theta}; z_i))^2}{2\sigma_i^2} \right] \tag{5.2}$$

Thus, a natural consideration of robust analysis that is less sensitive to outliers is to consider the fist order of the residuals as the merit function (Press et al., 2002), i.e. with the likelihood function

$$L(\boldsymbol{\theta}|\{z_i\}, \{\mu_i\}, \{\sigma_i\}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[- \frac{|\mu_i - \mu(\boldsymbol{\theta}; z_i)|}{2\sigma_i} \right] \tag{5.3}$$

In Fig.(5.1), we show the two different types of distributions. When there are outliers, the double side exponential distribution is a better representation of the residual distribution and more robust.

Comparing Eq.(5.2) and Eq.(5.3), one can observe that maximizing the latter by methods involving derivatives (such as the popular Levenberg-Marquardt method) will not be valid any longer due to the discontinuity of the derivative for Eq.(5.3) at the origin. For models with fewer parameters, one can directly use a laborious brute-force method to grid the parameter space uniformly, evaluating the likelihood function at all the grid points and select the maximum after marginalizing over other nuisance parameters. This requires tremendous computation power as the number of parameters increases and the grid size decreases, rendering the analysis less feasible. Markov Chain Monte Carlo (MCMC) (Liu, 2002) methods provide a powerful solution to this class of problems. Instead of searching the full grid of the parameter space, MCMC will sample the posterior distributions of each parameter by "jumping" in parameter space with appropriate acceptance/rejection rules.

Next, we fit the data to a cosmological model with dynamic dark energy, allowing the equation of state of dark energy and Hubble constant to be free parameters in addition to Ω_M. We also assume a flat Universe prior and constraints to the parameters to be $w \geq -1$ and $\Omega_M \geq 0$. The Supernovae data we used is from the gold sample of Riess et al. (2004). The fitting function is

$$\mu(z_i;\Omega_M,w,h) = 5\log\left[\frac{3000(1+z)}{h}\int_0^z dz[\Omega_M(1+z)^3+(1-\Omega_M)(1+z)^{3(1+w)}]^{-1/2}\right]+25 \tag{5.4}$$

The results are presented in Figure 5.2 and 5.3. In doing the fitting, we use a Markov Chain with 100,000 iterations after the initial burn in. The acceptance ratio is 0.247 for both LSE method and the robust method. Though the difference between the parameters from LSE and the robust method are not statistically significant, the latter one has less tight constraints than the former one. These constraints are more realistic because the error distributions are more tolerant to data points in the tails. In the current case with Supernovae data, there is not a very big difference. But for future combined data analyses, such as WMAP+SN+WL+CL, the difference may be significant.

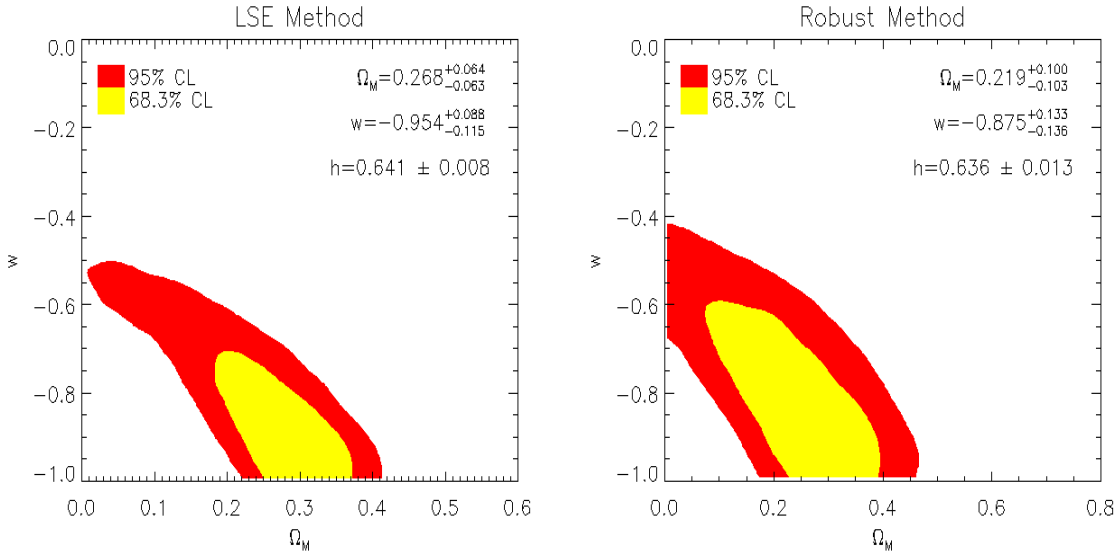


Figure 5.2. (a) Contour plot of w-Ω_M by LSE method; (b) Contour plot of w-Ω_M by robust method. Though in this case the difference is not statistically significant, the robust method gives less tight constraints, which is more realistic.

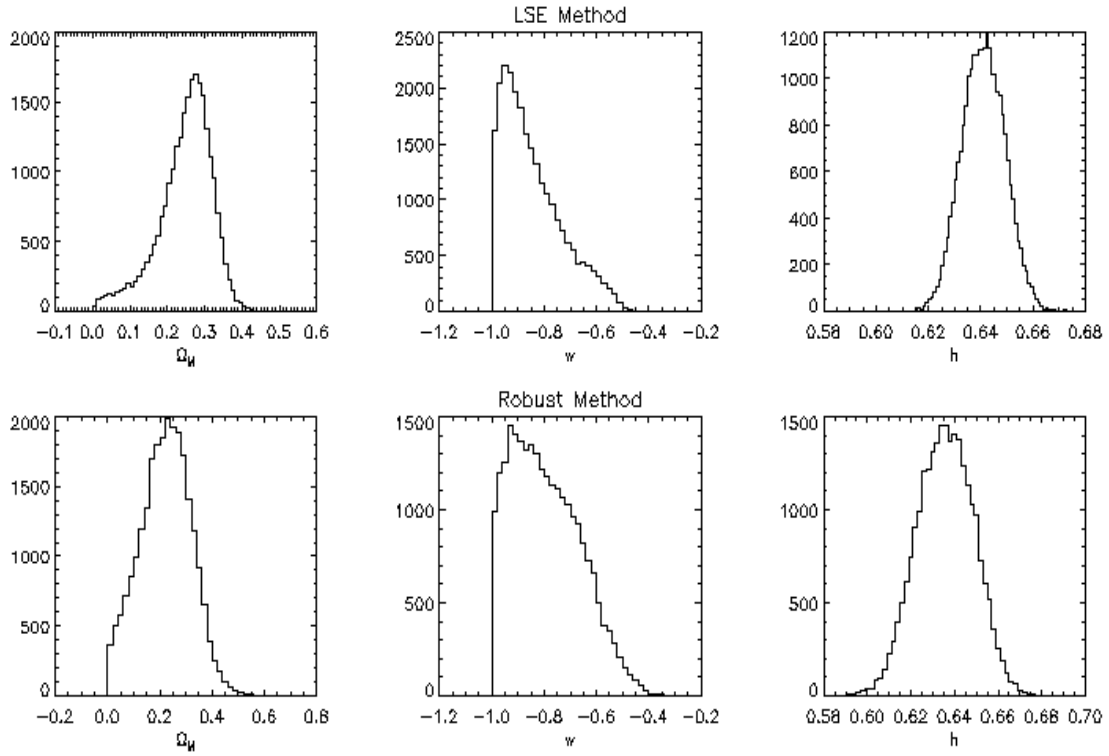


Figure 5.3. Posterior distribution of every parameters. The corresponding estimates of the parameters are shown in Fig.(5.2). The distribution of w and Ω_M are affected by our prior constraints $w \geq -1$ and $\Omega_M \geq 0$.

5.2 Figure of Merit for Dark Energy Experiments[1]

When we talk about constraints on dark energy, a very challenging task is to determine what to constrain. Generally, we can describe dark energy by two characteristics: the energy density (Ω_{DE}) and the equation of state ($w(z) = p(z)/\rho(z)$) and therefore the constraints should be on Ω_{DE} and $w(z)$.

However, constraining $w(z)$ is a tricky issue. Generally speaking, since we do not know the form of $w(z)$ *a priori*, there is no reason to impose a fixed parametric form for $w(z)$ and a non-parametric $w(z)$ is more favorable. We should allow some freedom on the form of $w(z)$ and let the data determine which form is preferred. In this direction, Huterer & Starkman (2003) considered $w(z)$ as piece-wise function of different redshift bins $w_i(z_i)$ and using SNIa data to constrain the value of w_i in each bin. Further work along this direction was done in (Saini, 2003; Wang & Mukherjee, 2004; Huterer & Cooray, 2005; Wang & Tegmark, 2005; Shapiro & Turner, 2006; Stephan-Otto, 2006; Krauss et al., 2007) and most recently in (Albrecht et al., 2009). However, as pointed out in (Genovese et al., 2008), such a scheme leads to big variance on each of the estimated $w(z_i)$. As an improvement, a new scheme is proposed to fit the non-parametric $w(z)$(Genovese et al., 2008) with a series of B-splines extending to dimension k and determining the appropriate dimension k by using the Bayesian Information Criterion (BIC) based on the SNIa data.

On the other hand, parametric $w(z)$ promises efficient fitting and smaller variance on the resulting parameters on $w(z)$. But its full power is based on the assumption that the parametric form of $w(z)$ is very close to the "true" form, which we actually do not know *a priori*.

A very important question here is how do we compare the constraints on $w(z)$ from different experiments. In the Dark Energy Task Force (DETF) report, one figure of merit was suggested based on a specific parameterization of $w(z) = w_0 + w_a z/(1+z)$. "The DETF figure of merit is the reciprocal of the area of the error ellipse enclosing the 95% confidence limit in the $w_0 - w_a$ plane. Larger figures of merit indicate greater

[1]Tables and plots in this section are mostly reproduced from Liu et al. (2008)

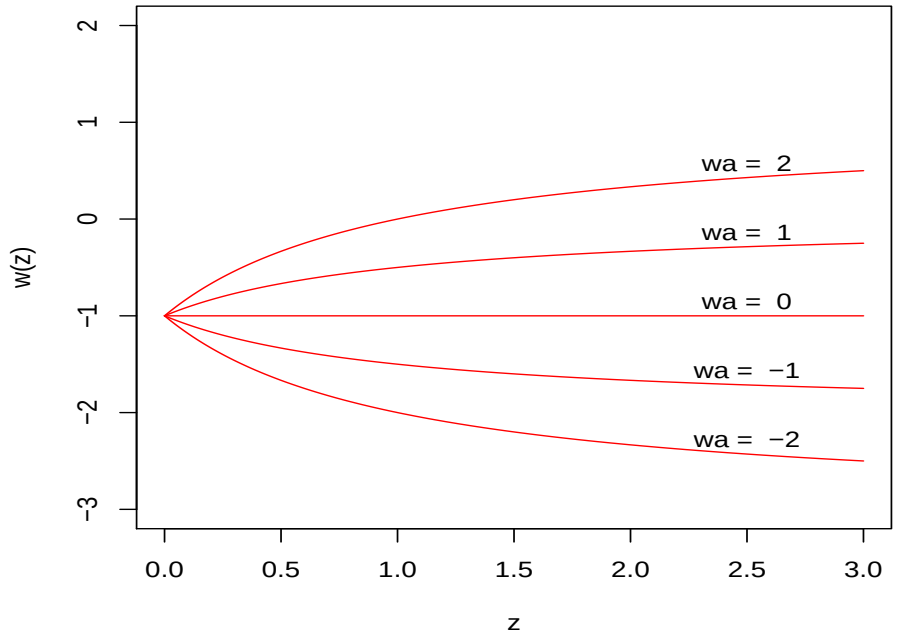


Figure 5.4. The evolution of the parametrization $w(z) = w_0 + w_a z/(1+z)$. Since w_0 just sets the asymptotic value, we fixed it as -1 in the plot.

accuracy" (Albrecht et al., 2006). One can immediately notice that this figure of merit is heavily dependent on the choice of the form of the parametrization. For the given $w(z) = w_0 + w_a z/(1+z)$, no matter how you vary the parameters w_0 and w_a, it just can represent a specific family of the evolution of w. In Figure 5.4, we show the evolution of $w(z)$ for various w_a.

If the real evolution of $w(z)$ does not follow the family of evolutions depicted in Figure 5.4, the DETF figure of merit will not truly reflect the reality of the constraints. Therefore, a new figure of merit that is not dependent on a specific parametrization is more desirable. In Liu et al. (2008), we proposed to use the area of the 1 σ error band around w(z) as a new figure of merit. It is defined as

$$A_W = 2\int_0^{z_{max}} \{Var[w(z)]\}^{1/2} dz \tag{5.5}$$

where $Var[w(z)]$ is the variance of $w(z)$. This is essentially the integrated error

Name	Definition	Reference
Linear	$w(z) = w_0 + w_a z$	Huterer & Turner (2001)
		Weller & Albrecht (2002)
UIS	$w(z) = w_0 + w_a z$ $(z < 1)$	Upadhye et al. (2005)
	$w(z) = w_0 + w_a$ $(z \geq 1)$	
CPL	$w(z) = w_0 + w_a \frac{z}{1+z}$	Chevallier & Polarski (2001)
		Linder (2003)
Family I	$w(z) = w_0 + w_a (\frac{z}{1+z})^n$	Liu et al. (2008)
Family II	$w(z) = w_0 + w_a \frac{z}{(1+z)^n}$	Liu et al. (2008)

Table 5.1. Some major 2-parameter parameterizations of equation of state

across a redshift range. Since our main goal is to know how well $w(z)$ is constrained, we should look at the integrated error rather than the error at a given pivot point. Clearly, the area A_W quantifies how well we can constrain the $w(z)$ across the whole redshift range from 0 to z_{max}. It is the overall information we have about $w(z)$. No matter how you parameterize $w(z)$, A_W can be calculated accordingly and compared. The smaller the A_W, the better the overall constraints on $w(z)$.

There are many different parametric forms for $w(z)$ which have been proposed. In Table 5.2, we listed several two-parameter parameterizations. We will focus on the two-parameter parameterizations in this work because they provide a starting point for parameterizing the redshift dependence of the equation of state $w(z)$. Clearly, as more parameters are introduced, we can achieve a better fit to the data. Comparing models with different numbers of parameters normally involves some arbitrariness on the choice of penalization for larger number of parameters. For example, one can choose either Bayesian Information Criterion (BIC) or Akaike Information Criterion. In this work, we do not compare models with different number of parameters. We will instead focus on two-parameter parameterizations.

As an example to show the variance of w(z) for a specific parametrization, let's consider the linear parametrization $w(z) = w_0 + w_a z$. The variance is $Var[w(z)] = Var(w_0) + z^2 Var(w_a) + 2z Cov(w_0, w_a)$. Clearly, no matter how small the constraints you place on w_0 and w_a, the variance $Var[w(z)]$ increases rapidly as redshift increases, meaning that we have increased uncertainties about $w(z)$ as redshift increases. However, the CPL parametrization will not diverge so rapidly as redshift increases. That

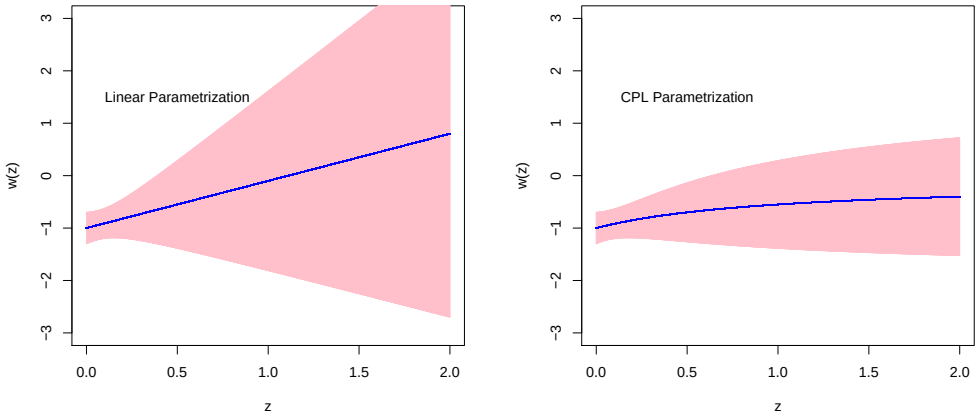


Figure 5.5. The 1 σ error bands for linear parametrization (left) and CPL parametrization (right). In the plot, we fix $w_0 = -1 \pm 0.3$, $w_a = 0.9 \pm 1.8$ and $Cov(w_0, w_a) = -0.2$. The same parameters will lead to different information about $w(z)$ for different parameterizations.

is to say, suppose we have the same constraints on w_0 and w_a for both cases, the CPL parametrization essentially gives us more information (say, the Fisher Information) about the $w(z)$ across a wide redshift range. in Figure 5.5, we show the 1 σ band for both parameterizations with the same constraints on w_0 and w_a.

Therefore, the CPL parametrization is better than the linear parametrization in this sense. But is the CPL parametrization the best? To answer this question, we consider two families of nested parameterizations that both have the CPL parametrization as a special case. We then look at their corresponding A_W in addition to their minimum χ^2 from the fit. The point here is that if they have similar minimum χ^2, the one that has less A_W is more favored.

Among the many observations that can help constrain the shape of $w(z)$, SNIa data provide the most straightforward constraints. Therefore, in this work, we will study the effects of different parameterizations on our understanding of the evolution of dark energy based on SNIa data.

In Table 5.2, we tabulated the constraints on w_0, w_a and the corresponding A_W for different parameterizations based on the SNIa data. In Fig. 5.6, 5.7, 5.8, we plot the corresponding 1 σ band for different parameterizations. In Fig. 5.9, we plot

Model	$\chi^2_{\rm min}$	w_0	w_a	A_W
Linear	195.409	-1.12628±0.281052	0.0811196±1.18901	2.92336
UIS	195.412	-1.1192±0.27732	0.0485532±1.17151	2.21778
CPL	195.411	-1.12456±0.331918	0.0961458±1.89159	1.88532
Family I				
n=2	195.413	-1.11369±0.21235	0.135963±4.91012	3.1127
n=3	195.402	-1.12407±0.181828	1.52726±14.2855	5.13255
n=4	195.314	-1.15242±0.164779	13.4211±34.8666	7.01332
Family-II				
n=2	195.409	-1.13475±0.412811	0.203332±3.09753	1.13799
n=3	195.399	-1.17258±0.546306	0.631377±5.28164	0.641504
n=4	195.356	-1.29367±0.787516	2.28402±9.61853	0.960275

Table 5.2. The minima of χ^2 and areas of the $w(z)$ band for different models using 192 SNIa data (Davis et al., 2007; Wood-Vasey et al., 2007; Riess et al., 2007). For the nested family I and II, we omit the $n = 1$ case because they both reduce to the CPL parametrization as $n = 1$

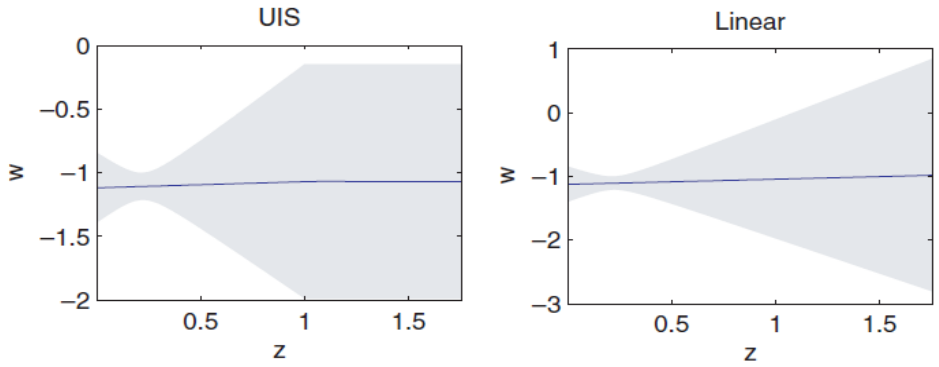


Figure 5.6. Evolution of $w(z)$ under the UIS (left) and linear (right) parameterizations

$\chi^2_{\rm min} - A_W$ for different parameterizations. Based on this plot, we can see that family II with $n = 4$ or 3 performs better than the CPL parametrization.

We must point out that this work is still an initial test for the method. The A_W are closely related to the testing power when we perform the hypothesis test to compare $w(z)$ for theoretical models and experiments. More systematic analyses by incorporating more experimental data is underway.

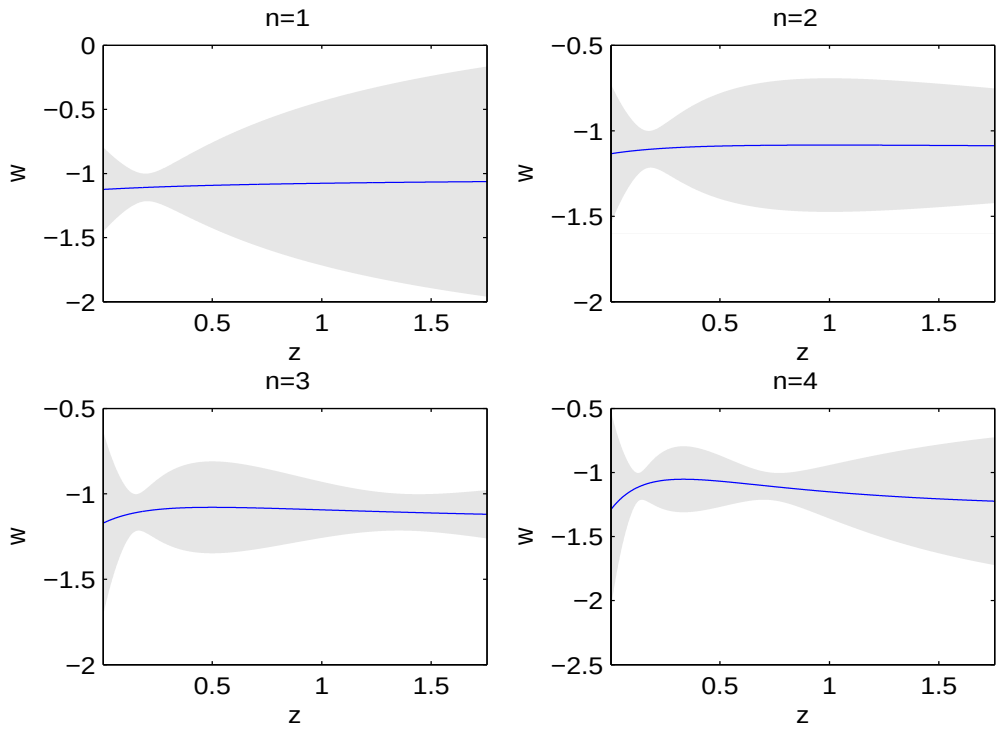


Figure 5.7. Evolution of $w(z)$ under the family I parameterizations of different n.

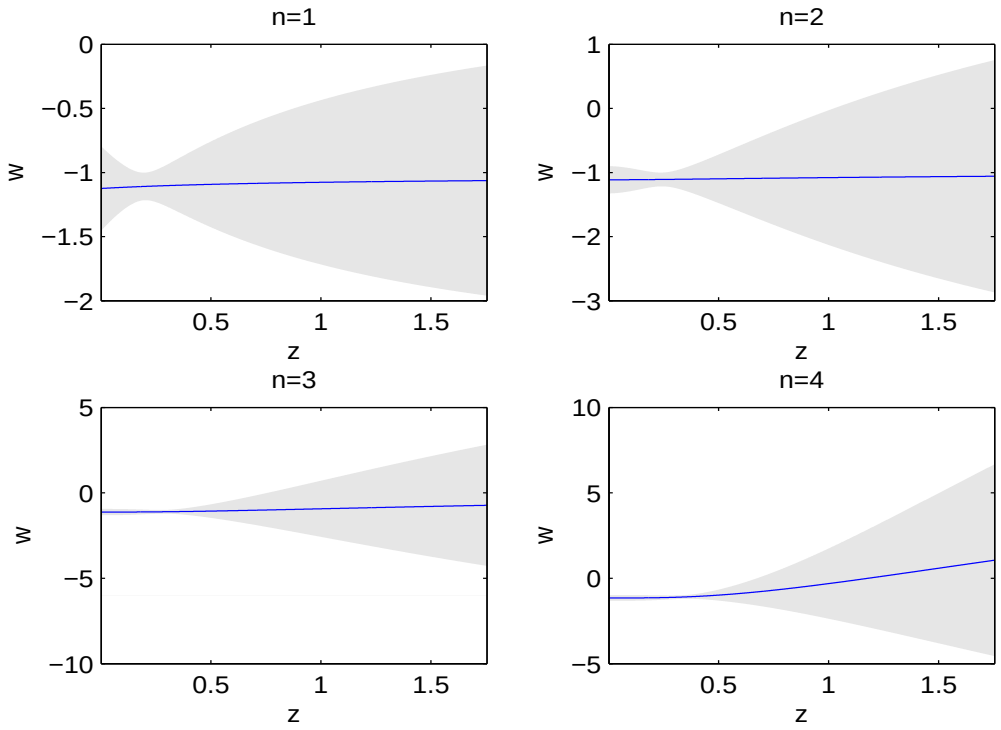


Figure 5.8. Evolution of $w(z)$ under the family II parameterizations of different n.

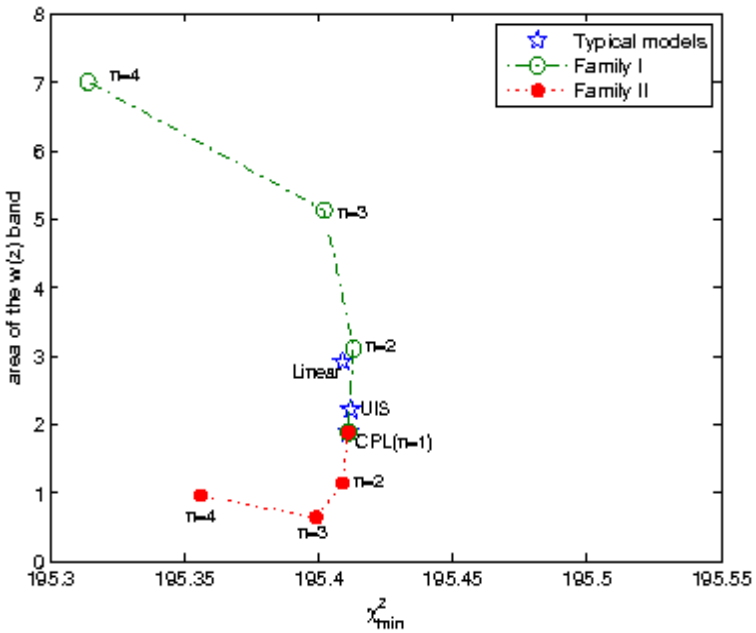


Figure 5.9. The phase diagram in the $\chi^2_{\rm min} - A_W$ plane for parametrizations listed in Table.5.2 based on SNIa data fit. Based on this plot, we can see that family II with n=3 or 4 outperform the CPL parametrization

CHAPTER 6

The Outlook

6.1 Conclusions

In this thesis, I developed a new optical cluster detection algorithm, GMBCG, which detects galaxy clusters by taking advantage of red sequence galaxies. The algorithm is easily extendible to high redshift by switching to the appropriate color ridgeline as redshift increases. The most important feature of the algorithm is that it does not require prior knowledge of very specific filters in color space and real space. Instead, it detects color clustering using an Error Corrected Gaussian Mixture Model (ECGMM) and then convolves the selected red sequence galaxies with a radial kernel. In the process, the algorithm does not *match* to any model filters. Therefore, it is less biased compared with matched filter algorithms.

Another major difference between the current algorithm and matched filter algorithms is the photometric redshift. In GMBCG, the photometric redshifts are taken as independently determined inputs. GMBCG itself does not optimize any matching to filters with respect to photometric redshifts. Based on the detected red sequence, the algorithm can produce a photometric redshift as a byproduct.

The detection limit in GMBCG depends mainly on the proper functioning of the ECGMM. Based on our Monte Carlo test, the ECGMM can work reliably when there are more than 10 red sequence galaxies. By using the resampling technique (as in section 2.3.3), we can make the ECGMM fitting significantly more robust.

By applying the algorithm, we have assembled a catalog of more than 53,000 rich clusters based on the DR7 of SDSS. The main sample of clusters spans a wide redshift range from 0.1 to 0.5 in an approximately volume limited way. This new catalog will

enable further studies of the evolution of clusters with sufficient statistics.

6.2 Follow-ups

The most immediate analysis based on the GMBCG clusters will be a weak gravitational lensing analysis for stacked clusters. Finding clusters is important, but knowing their masses is even more important for cluster cosmology. A good cluster finding algorithm should yield not only the list of clusters, but also some observable measurements (such as richness) that can be consistently mapped to masses across a wide redshift range.

Recently, we have used weak lensing analysis to estimate the masses of clusters in the maxBCG catalog (Sheldon et al., 2007c,a; Johnston et al., 2007). We will soon extend this analysis to the GMBCG clusters, which have different, presumbly better richness measurements. Meanwhile, a number of clusters in the new GMBCG catalog show strong lensing signatures, arcs, which will allow us to estimate their masses using strong lensing techniques. A cross comparison between the results from different lensing techniques will provide additional information on the clusters' masses. Most important, an empirical halo occupancy distribution (HOD) and its evolution could be constructed by analyzing the cluster masses via gravitational lensing.

Two additional techniques for estimating masses of maxBCG clusters, the velocity dispersion (Becker et al., 2007) and X-ray luminosity (Rykoff et al., 2008), will also be extended to the new GMBCG clusters. These analysis are complementary to the weak lensing analysis and can help to get tighter constraints on the mass richness scaling relation. Since GMBCG clusters span redshifts 0.1 to 0.5, they will allow us to measure the evolution of cluster properties across a broader redshift range than was possible with maxBCG.

GMBCG clusters, drawn from a much larger volume than maxBCG, will be especially useful for cosmological constraints. With more clusters and a wider redshift range, we expect to lay down considerable tighter constraints on σ_8 and Ω_M.

Finally, an important application of the GMBCG algorithm is to mock surveys. Mock catalog can help to understand the performance of the cluster finder at high

redshift and the interplay between clusters and dark matter halos. This is an indispensable step for cluster cosmology.

6.3 Dark Energy Survey and the Future

As I pointed out in the introduction, the major motivation for our cluster detection project is for cluster cosmology. The Dark Energy Survey (DES) (The Dark Energy Survey Collaboration, 2005) is the next generation multi-color digital sky survey aimed at providing tighter constraints on dark energy evolution. The GMBCG algorithm is especially designed for this survey. Numerous tests of the algorithm against the DES simulation are undergoing. All our experience with the algorithm on SDSS data will prepare us for the deeper data from DES.

On a similar time scale as DES, cluster detection efforts in the microwave band are underway. The South Pole Telescope project (Staniszewski et al., 2008, SPT) is designed to detect galaxy clusters via the Sunyaev-Zeldovich effect in the microwave band. It covers 4000 $\deg^2$ of the southern sky, which are also covered by DES in the optical band. Therefore, SPT and DES are highly complimentary for cluster detection and their mass calibration. The planned mission, eROSITA (http://www.mpe.mpg.de/projects.html #erosita), is an X-ray all-sky survey in the medium energy range up to 10 keV. It will provide rich data for detecting clusters in the X-ray band. In the very near future, the combined analysis of clusters detected from these multi-wavelength data will significantly improve the purity, completeness and mass calibration of galaxy clusters. No doubt, cluster cosmology is becoming a hot research topic and tighter constraints on the evolution of dark energy will be imposed.

Besides dark energy, a major task in physics and astronomy is to map out the observable part of our Universe and catalog all the celestial objects we can detect in the next 10 to 20 years. We are now in a unique historical era. We are building the basic atlas of the Universe, just as explorers did for the Earth 500 years ago. Galaxy clusters are the largest gravitationally bound systems in our Universe. Besides cosmology, building a cluster catalog, on its own, is an important contribution to this

ambitious task in cosmic mapping. The completion of the GMBCG cluster catalog for the SDSS DR7 is one significant step towards cluster cartography in the near future.

APPENDICES

APPENDIX A

C++ class for implementation of ECGMM

```
/*-------------------------------------------------------------------------

  NAME:
    ecGMMClass.h

  PURPOSE:

    This C++ class implement the EM algorithm for the one dimensional
    Error Corrected Gaussian Mixture Model as specified in
    Jiangang Hao et al, 2008

  PLATFORM:

    This code is tested on Redhat Linux and Cygwin with gcc version 3.4.4
    in 2008.

  NOTE: since different people will use the mixture model very differently, I did
        not provide a simple function, instead, i put the most basic brick: the
        ecGMMClass that you can customize very freely by following the example.

  DEPENDENCIES:

    The class itself does not require other libraries than the standard C++
    libraries. If you want to use random number to test it, I recommend you
    use the GNU gsl random number generators. In the ecGMM.cpp file, the random
    generator is used. So, make sure you have gsl properly installed.

  USAGE:

    This class produce an object, in which you can specify the data as
    well as the EM iterations. The input part including:

   W : your initial guess of the weights of Gaussian components
   Mu: your initial guess of the locations of Gaussian components
   Sigma: your initial guess of the width of Gaussian components
   x: the data array
   xErr: the measurement errors corresponding to x
   N: the number of mixtures.

  REVISION HISTORY:

    Created September-2008: Jiangang Hao, University of Michigan
    Copyright (C) 2008 Jiangang Hao, Dept.of Phys., University of Michigan
                       jghao@umich.edu  or jianganghao@gmail.com

    This program is free software; you can redistribute it and/or modify
    it under the terms of the GNU General Public License as published by
    the Free Software Foundation; either version 2 of the License, or
    (at your option) any later version.

    This program is distributed in the hope that it will be useful,
```

```
    but WITHOUT ANY WARRANTY; without even the implied warranty of
    MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
    GNU General Public License for more details.

    You should have received a copy of the GNU General Public License
    along with this program; if not, write to the Free Software
    Foundation, Inc., 51 Franklin St, Fifth Floor, Boston, MA  02110-1301  USA
-------------------------------------------------------------------------------*/

#include<iostream>
#include<fstream>
#include<cmath>
#include<vector>
using namespace std;

class ecGMM {
  public:

  vector<double> x, xErr, W, Mu, Sigma;
  double GaussErrFun(double y, double yErr, double mean, double sd);
  double pSum(int j);
  double pIJ(int i, int j);
  void W_update();
  void Mu_update();
  void Sigma_update();
  void update();            //replace W,Mu and Sigma with the new one
  double lhood();           // Calculate the likelihood
  double BIC();             // Calculate the BIC of the maximum likelihood
  int N;                    // number of mixture
  int M;                    // number of data points
  double ep;                // precision goal

  ecGMM(int N){
       for(int k=0; k<N; k++){
          Wt.push_back(0.);
          Mut.push_back(0.);
          Sigmat.push_back(0.);
       }}                   //constructor to initialize Wt, Mut, Sigmat

  private:
  vector<double> Wt, Mut, Sigmat; // store the update
};

//--------------------definition of member functions--------------------//

double ecGMM::GaussErrFun(double y, double yErr, double mean, double
sd) {
  double PI=3.14159265;
  double res=0.;
  res = exp(- pow((y-mean),2.0)/(yErr*yErr+sd*sd)/2.0)/sqrt(2.0*PI*(yErr*yErr+sd*sd));
  return res;
}

//--------define \sum_k=0^N p(y_j|z_j=k,\theta^{(t)})---------------//

double ecGMM::pSum(int j) {
  double res = 0.;
  for(int k = 0; k < N; k++)
    {
      res = res + GaussErrFun(x[j],xErr[j],Mu[k],Sigma[k])*W[k];
    }
  return res;
}

//--------define p(z_j=i|y_j,\theta^{(t)})-----------------------------//

double ecGMM::pIJ(int i, int j) {
```

```
  double res;
  res = GaussErrFun(x[j],xErr[j],Mu[i],Sigma[i])*W[i]/pSum(j);
  return res;
}

//--------Update Weight and store it in Wt-----------------------------//

void ecGMM::W_update() {
  double res;
  for(int i = 0; i < N; i++)
    {
      res = 0.;
      for(int j = 0; j < M; j++)
    {
        res = res + pIJ(i,j);
    }
      Wt[i] = res/double(M);
    }
}

//--------Update Mu and store it in Mut--------------------------------//

void ecGMM::Mu_update() {
  double resu, resd;
  for(int i = 0; i < N; i++)
    {
      resu = 0.;
      resd = 0.;

      for(int j = 0; j < M; j++)
    {
      resu = resu + x[j]*pIJ(i,j)*pow(Sigma[i],2)/(pow(Sigma[i],2)+pow(xErr[j],2));
      resd = resd + pIJ(i,j)*pow(Sigma[i],2)/(pow(Sigma[i],2)+pow(xErr[j],2));
    }
      Mut[i] = resu/resd;
     }
 }

//--------Update Sigma and store it in Sigmat--------------------------//

void ecGMM::Sigma_update() {
   for(int i = 0; i < N; i++)
    {
     double resu = 0.;
     double resd = 0.;
      for(int j = 0; j < M; j++)
    {
      resu = resu + pow((x[j] - Mu[i]),2)*pIJ(i,j)*pow(Sigma[i],2)
                                    /(pow(Sigma[i],2)+pow(xErr[j],2));
      resd = resd + pIJ(i,j)*pow(Sigma[i],2)/(pow(Sigma[i],2)+pow(xErr[j],2));
    }
      Sigmat[i] = sqrt(resu/resd);
    }
}

//--------update all parameters ---------------------------------------//

void ecGMM::update() {
  for(int i=0; i<N; i++)
    {
      W[i] = Wt[i];
      Mu[i] = Mut[i];
      Sigma[i] = Sigmat[i];
    }
}

//------calculate the likelihood-------------------------------------//
```

```
double ccGMM::lhood() {
  double res1, res2;

  res1=1.;
  for(int j=0;j<M;j++)
    {
      res2=0.;
      for(int i=0; i<N; i++)
    {
      res2 = res2 + W[i]*GaussErrFun(x[j],xErr[j],Mu[i],Sigma[i]);
    }
      res1=res1*res2;
    }
  return(res1);
}

//-----return the BIC-------------------------------------------------//

double ccGMM::BIC() {
  double res;
  res=-2.0*log(lhood())+double(3.*N-1.)*log(double(M));
  return(res);
}
```

APPENDIX B

Parameter fitting with MCMC

In this appendix, we show briefly how the MCMC methods (Metroplis Hasting algorithm) work. For more details, refer to (Liu, 2002). For simplicity, we consider a model with parameters θ. Given the parameters at t^{th} iteration as $\theta^{(\mathbf{t})}$, we make the variations as

$$\theta' = \theta^{(\mathbf{t})} + \epsilon \tag{B.1}$$

where ϵ is a set of random variables generated and re-scaled. Then, we draw $U \sim \texttt{Uniform}[0, 1]$. If $U \leq r(\theta^{(\mathbf{t})}, \theta')$, then we make the following updates

$$\theta^{(\mathbf{t+1})} = \theta' \tag{B.2}$$

On the other hand, if $U > r(\theta^{(t)}, \theta')$, we make the update as

$$\theta^{(\mathbf{t+1})} = \theta^{(\mathbf{t})} \tag{B.3}$$

The $r(\theta^{(t)}, \theta')$ is defined as

$$r(\theta^{(t)}, \theta') = \min[1, \frac{\pi(\theta')}{\pi(\theta^{(t)})}] \tag{B.4}$$

where $\pi(\theta)$ is the posterior probability distribution of the parameters given the data. After sufficient long sampling, we will recover the corresponding posterior probability distributions of θ, from which we will get all information about that parameters, i.e. their means, variances and etc.

BIBLIOGRAPHY

Abazajian, K., & Sloan Digital Sky Survey, f. t. 2008, ArXiv e-prints

Abell, G. O. 1957, The distribution of rich clusters of galaxies. A catalogue of 2712 rich clusters found on the National Geographic Society Palomar Observatory Sky Survey (Chicago: Univiersity of Chicago Press, 1957)

Albrecht, A., et al. 2009, ArXiv e-prints

—. 2006, ArXiv Astrophysics e-prints

Annis, J., et al. 1999, in Bulletin of the American Astronomical Society, Vol. 31, Bulletin of the American Astronomical Society, 1391–+

Barger, A. J., et al. 1998, ApJ, 501, 522

Barrientos, L. F. 1999, PhD thesis, AA(UNIVERSITY OF TORONTO (CANADA))

Bartelmann, M. 1996, A&A, 313, 697

Becker, M. R., et al. 2007, ApJ, 669, 905

Berlind, A. A., et al. 2006, ApJS, 167, 1

Bernardi, M., Nichol, R. C., Sheth, R. K., Miller, C. J., & Brinkmann, J. 2006, AJ, 131, 1288

Bernardi, M., Sheth, R. K., Nichol, R. C., Schneider, D. P., & Brinkmann, J. 2005, AJ, 129, 61

Bertone, G., Hooper, D., & Silk, J. 2005, Physics Reports, 405, 279

Blakeslee, J. P., et al. 2003, ApJ, 596, L143

—. 2006, ApJ, 644, 30

Blanton, M. R., et al. 2005, AJ, 129, 2562

Böhringer, H., et al. 2004, A&A, 425, 367

—. 2000, ApJS, 129, 435

Botzler, C. S., Snigula, J., Bender, R., & Hopp, U. 2004, MNRAS, 349, 425

Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, MNRAS, 254, 601

Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, MNRAS, 351, 1151

Burenin, R. A., Vikhlinin, A., Hornstrup, A., Ebeling, H., Quintana, H., & Mescheryakov, A. 2007, ApJS, 172, 561

Carlstrom, J. E., Holder, G. P., & Reese, E. D. 2002, Ann. Rev. Astron. Astrophys., 40, 643

Chevallier, M., & Polarski, D. 2001, International Journal of Modern Physics D, 10, 213

Coil, A. L., et al. 2008, ApJ, 672, 153

Connolly, A. J., Genovese, C., Moore, A. W., Nichol, R. C., Schneider, J., & Wasserman, L. 2000, ArXiv Astrophysics e-prints

Cool, R. J., Eisenstein, D. J., Johnston, D., Scranton, R., Brinkmann, J., Schneider, D. P., & Zehavi, I. 2006, AJ, 131, 736

Couch, W. J., Ellis, R. S., MacLaren, I., & Malin, D. F. 1991, MNRAS, 249, 606

Cunha, C. 2008, ArXiv e-prints

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

Davis, T. M., et al. 2007, ApJ, 666, 716

De Lucia, G., et al. 2007, MNRAS, 374, 809

Dempster, A., Laird, N., & Rubin, D. 1977, Journal of the Royal Statistical Society, Series B, 39, 1

Dong, F., Pierpaoli, E., Gunn, J. E., & Wechsler, R. H. 2008, ApJ, 676, 868

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, MNRAS, 235, 715

Eisenhardt, P. R., et al. 2005, in Bulletin of the American Astronomical Society, Vol. 37, Bulletin of the American Astronomical Society, 1344–+

Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605

Eisenstein, D. J., et al. 2005, ApJ, 633, 560

ESO. 2004, Cosmological Constraints, [Online; accessed 2-April-2009]

Evrard, A. E., et al. 2002, ApJ, 573, 7

Faber, S. M., et al. 2007, ApJ, 665, 265

Franzetti, P., et al. 2007, A&A, 465, 711

Frieman et al. 2006, Dark Energy Survey Science Program, [Online; accessed 2-April-2009]

Gal, R. R. 2006, ArXiv Astrophysics e-prints

Gal, R. R., de Carvalho, R. R., Lopes, P. A. A., Djorgovski, S. G., Brunner, R. J., Mahabal, A., & Odewahn, S. C. 2003, AJ, 125, 2064

Gal, R. R., de Carvalho, R. R., Odewahn, S. C., Djorgovski, S. G., & Margoniner, V. E. 2000, AJ, 119, 12

Genovese, C. R., Freeman, P., Wasserman, L., Nichol, R. C., & Miller, C. 2008, ArXiv e-prints, 805

Gladders, M. D., Lopez-Cruz, O., Yee, H. K. C., & Kodama, T. 1998, ApJ, 501, 571

Gladders, M. D., & Yee, H. K. C. 2000, AJ, 120, 2148

—. 2005, ApJS, 157, 1

Goto, T., et al. 2002, AJ, 123, 1807

Hansen, S. M., Sheldon, E. S., Wechsler, R. H., & Koester, B. P. 2007, ArXiv e-prints, 710

Hao, J., Mckay, T. A., & others. 2009, in preparation

Hogg, D. W., et al. 2004, ApJ, 601, L29

Hu, W. 2003, Physical Reviews D, 67, 081304

Hu, W. 2005, Lecture notes, Website, `http://background.uchicago.edu/~whu/Courses/Ast321_05`

Hu, W., & Kravtsov, A. V. 2003, ApJ, 584, 702

Huchra, J. P., & Geller, M. J. 1982, ApJ, 257, 423

Huterer, D., & Cooray, A. 2005, Physical Reviews D, 71, 023506

Huterer, D., & Starkman, G. 2003, Physical Review Letters, 90, 031301

Huterer, D., & Turner, M. S. 2001, Physical Reviews D, 64, 123527

Jansen, F., et al. 2001, A&A, 365, L1

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372

Johnston, D. E., et al. 2007, ArXiv e-prints, 709

Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R., & Xu, G. 1999, ApJ, 517, 78

Kim, R. S. J., et al. 2002, AJ, 123, 20

Kodama, T., & Arimoto, N. 1997, A&A, 320, 41

Koester, B. P., et al. 2009, ArXiv e-prints

—. 2007a, ApJ, 660, 239

—. 2007b, ApJ, 660, 221

Krauss, L. M., Jones-Smith, K., & Huterer, D. 2007, New Journal of Physics, 9, 141

Li, I. H., & Yee, H. K. C. 2008, AJ, 135, 809

Lidman, C. E., & Peterson, B. A. 1996, AJ, 112, 2454

Lima, M., & Hu, W. 2004, Physical Reviews D, 70, 043504

Linder, E. V. 2003, Physical Review Letters, 90, 091301

Liu, D.-J., Li, X.-Z., Hao, J., & Jin, X.-H. 2008, MNRAS, 714

Liu, J. S. 2002, Monte Carlo Strategies in Scientific Computing (Springer)

Lopes, P. A. A., de Carvalho, R. R., Gal, R. R., Djorgovski, S. G., Odewahn, S. C., Mahabal, A. A., & Brunner, R. J. 2004, AJ, 128, 1017

López-Cruz, O., Barkhouse, W. A., & Yee, H. K. C. 2004, ApJ, 614, 679

Majumdar, S., & Mohr, J. J. 2004, ApJ, 613, 41

Mei, S., et al. 2009, ApJ, 690, 42

Miller, C. J., et al. 2005, AJ, 130, 968

Mullis, C. R., Rosati, P., Lamer, G., Böhringer, H., Schwope, A., Schuecker, P., & Fassbender, R. 2005, ApJ, 623, L85

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Oyaizu, H., Lima, M., Cunha, C. E., Lin, H., Frieman, J., & Sheldon, E. S. 2007, ArXiv e-prints, 708

Perlmutter, S., et al. 1999, Astrophys. J., 517, 565

Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P., & Christensen, J. A. 1996, AJ, 111, 615

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2002, Numerical Recipes in C++: The Art of Scientific Computing (Cambridge University Press)

Ramella, M., Geller, M. J., Pisani, A., & da Costa, L. N. 2002, AJ, 123, 2976

Renzini, A. 2006, Annual Reviews in Astronomy and Astrophysics, 44, 141

Riess, A. G., et al. 2007, ApJ, 659, 98

Riess, A. G., et al. 1998, Astron. J., 116, 1009

—. 2004, Astrophys. J., 607, 665

Rozo, E., et al. 2008a, ArXiv e-prints

—. 2008b, ArXiv e-prints

Rozo, E., Wechsler, R. H., Koester, B. P., Evrard, A. E., & McKay, T. A. 2007a, ArXiv Astrophysics e-prints

Rozo, E., et al. 2007b, ArXiv Astrophysics e-prints

—. 2009, ArXiv e-prints

Rudd, D. H. 2007, PhD thesis, The University of Chicago

Rykoff, E. S., et al. 2008, ApJ, 675, 1106

Saini, T. D. 2003, MNRAS, 344, 129

Sandage, A., Binggeli, B., & Tammann, G. A. 1985, AJ, 90, 1759

Santos, J. S., et al. 2009, ArXiv e-prints

Schwarz, G. 1978, Annals of Statistics, 6, 461

Scodeggio, M. 2001, AJ, 121, 2413

Scott, D. 1992, John Wiley

Shapiro, C., & Turner, M. S. 2006, ApJ, 649, 563

Shectman, S. A. 1985, ApJS, 57, 77

Sheldon, E. S., et al. 2007a, ArXiv e-prints, 709

—. 2007b, ArXiv e-prints, 709

—. 2007c, ArXiv e-prints, 709

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Silverman, B. W. 1986, Chapman & Hall

Smail, I., Edge, A. C., Ellis, R. S., & Blandford, R. D. 1998, MNRAS, 293, 124

Spergel, D. N., et al. 2003, Astrophys. J. Suppl., 148, 175

—. 2007, Astrophys. J. Suppl., 170, 377

Springel, V., et al. 2005, Nature, 435, 629

Stanek, R., Evrard, A. E., Bohringer, H. B., Schuecker, P., & Nord, B. 2006, Astrophys. J., 648, 956

Staniszewski, Z., et al. 2008, ArXiv e-prints

Stephan-Otto, C. 2006, Physical Reviews D, 74, 023507

Sunyaev, R. A., & Zeldovich, Y. B. 1970, Astrophys. Space Sci., 7, 3

Tegmark, M., et al. 2004, Physical Reviews D, 69, 103501

The Dark Energy Survey Collaboration. 2005, ArXiv Astrophysics e-prints

Upadhye, A., Ishak, M., & Steinhardt, P. J. 2005, Physical Reviews D, 72, 063501

van Dokkum, P. G., Franx, M., Kelson, D. D., Illingworth, G. D., Fisher, D., & Fabricant, D. 1998, ApJ, 500, 714

Vikhlinin, A., et al. 2006, Astrophys. J., 640, 691

Visvanathan, N., & Sandage, A. 1977, ApJ, 216, 214

Voges, W., et al. 1999, A&A, 349, 389

Voit, G. M. 2005, Rev. Mod. Phys., 77, 207

Wake, D. A., et al. 2006, MNRAS, 372, 537

Wang, Y., & Mukherjee, P. 2004, ApJ, 606, 654

Wang, Y., & Tegmark, M. 2005, Physical Reviews D, 71, 103513

Weller, J., & Albrecht, A. 2002, Physical Reviews D, 65, 103512

Wikipedia. 2009, Dark energy — Wikipedia, The Free Encyclopedia, [Online; accessed 2-April-2009]

Wood-Vasey, W. M., et al. 2007, ApJ, 666, 694

York, D. G., et al. 2000, AJ, 120, 1579

Zehavi, I., et al. 2005, ApJ, 630, 1