# A Search for Nearby Dwarf Galaxies

by

Chengyu Xi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics

Waterloo, Ontario, Canada, 2014

### Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Dwarf galaxies can be used to trace invisible dark matter substructure. Studies of the properties of dwarf galaxy populations are crucial in solving the so-called "Missing Satellite Problem", and hold important implications for galaxy formation. We use data from the COSMOS survey to study the statistical properties of nearby satellite galaxies, including their clustering, spatial distribution and abundance as a function of the luminosity relative to the central galaxies. The COSMOS data have 30-band photometric redshifts of very high quality, which greatly help our measurements. We detect an over-density of galaxies at projected separation $< 600$ kpc, corresponding to $7.9 \pm 0.5$ satellites per central galaxy. We also measure the relative abundance of satellites as a function of the magnitude difference between satellites and central galaxies, the so-called "relative luminosity function". After correcting for the variation of the magnitude limits for different primaries, we find a detection of $17.7 \pm 5.1$ satellites per central galaxy at $\Delta m_g \equiv g_{sat}^+ - g_{main}^+ \leq 10$. This result is two magnitudes fainter than most of the recent studies. This result is of two magnitudes less deep than the result of Speller & Taylor (2014), but is much more complete at the faint end, as we include the compact and low surface brightness galaxies that are excluded in their work. Additionally, we investigate satellite alignment with the orientations of the central galaxies, searching for the so-called "Holmberg effect". We split our primary sample into several sub-samples, including blue/red, bright/faint, and early/late SED types. The blue and red and early and late sub-samples have roughly similar g-band luminosities. For early type and red primaries, we find some evidence for alignment between satellite positions and the major axes of the primary galaxies. We also study the dependence of the amplitude of the clustering signal on the primary properties. We find that early-type primaries have a stronger clustering signal with their satellite populations, a larger population per central galaxy and possibly a more extended radial distribution of satellites than the late-type primaries on average. This may indicate differences in the properties or in the baryonic evolution of the host halos of the two types of primaries.

## Acknowledgements

I would like to thank Dr. James Taylor for his help through out the work, making this thesis possible.

I would like to thank the members of the COSMOS collaboration for their work on the COSMOS catalogues, and for providing support and access to both public and private catalogues.

I would also like to thank Iris Yan Chan and Xiaoyu Wu for their support with writing.

# Dedication

I want to dedicate this thesis to my sister, to whom I always feel guilty for ruining her childhood, but also grateful for her company and support all the time.

# Table of Contents

# List of Tables

# List of Figures

xiii

# Chapter 1

# Introduction

## 1.1   Cold Dark Matter Structure Formation

Dark matter (DM) comprises over 80% of the total matter density of the universe (or 26.5% of the total energy, while baryons are only 4.9%, most recent results of the Planck collaboration (Planck Collaboration et al., 2013)), and it thus has a dominant effect in the current standard picture of structure formation. The Cold Dark Matter (CDM) model and numerical simulations predict that structures in the universe are formed in a hierarchical manner, where the initial density fluctuations present at early times grow through gravitational instability and eventually collapse and virialize into dense, relaxed systems called "halos". Halos then continue to form larger structures through accretion and mergers (Blumenthal et al., 1984).

Dark matter halos are the fundamental non-linear units of the cosmic structure and also provide the environments where galaxies (e.g. the Milky Way) are formed (White & Rees, 1978). This cosmic structure consisting of dark matter halos causes the matter in the universe to cluster, rather than being randomly distributed. Dark matter particles and sub-structures are spatially associated inside individual host halos (also called "one-halo term" clustering), which themselves are associated in large-scale structures (also called "two-halo term" clustering) (Navarro, Frenk & White, 1996; Cooray & Sheth, 2002). The nature of this clustering can be found through the observation of the visible galaxies, which are predicted to form inside individual halos and thus can be used to trace the dark matter structures that are themselves invisible. The galaxy clustering at large scales, or the two-halo term clustering, has been well measured using observations of bright galaxies (e.g. the Luminous Red Galaxy (LRG) catalogues from the Sloan Digital Sky Survey (SDSS) (York

et al., 2000)) and has been successfully matched with the prediction from the CDM model. On the other hand, the one-halo term clustering at small scales, which involves observing generally fainter satellite galaxies, is more difficult to measure in practice and is also more theoretically complex, due to the effects from baryonic processes that are not well known. Chapter 3 will focus on the one-halo term clustering measurement.

High resolution numerical simulations have predicted various properties of dark matter halos. First of all, halos are defined as relaxed, virialized systems. The virial mass of dark matter halos is the total mass inside the virial radius (Frenk & White, 2012):

$$M_{vir} = (4\pi r_{vir}^3/3)\Delta_{vir,c} \; \rho_c \qquad (1.1)$$

where $M_{vir}$ is the virial mass of the halo, $r_{vir}$ is the virial radius of the halo, $\rho_c$ is the critical density of the universe and $\Delta_{vir,c} \approx 200$ in a universe with mean density equal to the critical density, according to linear theory. When we use $\Delta_{vir,c} = 200$, the virial radius is equivalent to the $R_{200}$, which is defined as the radius within which the average density of the halo is equal to 200 times of the mean density of the universe. Similarly, the $M_{200}$ is defined as the mass inside $R_{200}$ of the halo. Inside individual dark matter halos, the mass density of halos with different masses can be fit with a universal density profile as a function of radius from the halo center, which was proposed by Navarro, Frenk & White (1996), also known as the NFW profile:

$$\rho_{NFW}(r) = \frac{\rho_s}{\frac{r}{R_s}(1 + \frac{r}{R_s})^2} \qquad (1.2)$$

where $R_s$ is the scale radius (the radius at which the logarithmic slope of the density profile is -2) and $\rho_s$ is the characteristic mass of the halo. This density profile provides us with an initial guess at the spatial distribution of satellite galaxies: the halo mass densities of the region with radius close to $R_s$ are $\propto r^{-2}$ (for MW-like host halos, $R_s$ is typically 1/10 of the virial radius, which is around 30 kpc), so the corresponding projected density will go as $\propto r_p^{-1}$ (where $r_p$ is the projected separation). Assuming the substructures of the halos follow the same distribution, the abundance of substructures of a given halo in a given projected radial bin $2\pi r dr$ will be approximately a constant. This rough estimate is a good match to our measurements of the actual satellite spatial distribution. This result will be presented in chapter 4.

Additionally, numerical simulations also show that individual dark matter halos usually do not have spherical shapes, but are ellipsoidal or triaxial (Jing & Suto, 2002). It is interesting but remains unclear whether the shapes of the dark matter halos are related to the morphologies of the galaxies we observed within them (e.g. galactic disks of spiral

galaxies). If we assume the shapes of the dark matter halos and the galaxies are associated through some mechanism like the exchange of angular momentum, we can expect that the distribution of the substructures of the halos may have certain alignment with central galaxies sitting in the centers of the host halos. Furthermore, this alignment pattern may be observable by tracing the dark matter substructures with the dwarf galaxies that are located inside the virial radius of the host halos of the central galaxies. This will be discussed in detail in chapter 4.

## 1.2   Galaxy Formation in Dark Matter Halos

In theory, galaxies are considered to form inside dark matter halos, and this is also supported by abundant observational evidence. Some of the main pieces of the observational evidence include: 1) In 1939, the measurement of rotation curve (stellar rotation velocity as a function of radius) for the Andromeda Nebula (Babcock, 1939) indicated that much of its mass is distributed at a larger radius than where its stars are found. 2) Kinetic studies of spiral galaxies show their stellar mass is generally insufficient to stabilize their galaxy disks, and spherical dark matter halos are needed (Ostriker & Peebles, 1973). 3) Gravitational lensing can directly provide evidence for the spatial association of the visible galaxies with invisible structures of much larger mass.

The hierarchical structure formation model along with high resolution numerical simulations also predict that progenitor halos can often survive the process of merging into larger systems, giving a population of substructures, or subhalos. The baryonic materials inside the progenitor halos may also manage to cool and form stars before the mergers take place (Yang et al., 2006). This gives a population of satellite galaxies around the central galaxies, which are formed in the centers of the larger systems. The existence of those satellite galaxies can be found through the observation of the dwarf galaxies that are close to our Milky Way (MW). These dwarf galaxies typically have relatively high velocity dispersion ($\sigma_{dwarf} > 10$ km s$^{-1}$) with only very low stellar mass. This indicates that they are bound within local, over-dense dark matter structures. Furthermore, there are more than 25 identified dwarf galaxies that are located within 300 kpc from the MW (McConnachie, 2012), which is about the virial radius of the MW (Moore et al., 1999; Klypin, Zhao & Somerville, 2002), indicating the host dark matter structures of these dwarf galaxies cannot be independent dark matter halos but must be subhalos of the MW, and that these dwarf galaxies themselves must be the satellites of the MW.

However, when comparing to dwarf galaxies to the high resolution CDM based simulations, the predicted subhalos that are able to host the dwarf galaxies inside the galactic

halos like those of the MW and M31 outnumber the dwarf galaxies we observe dramatically (Klypin et al., 1999; Moore et al., 1999; Diemand, Kuhlen & Madau, 2007; Springel et al., 2008; Kravtsov, Gnedin & Klypin, 2004). This is known as the missing satellite problem (MSP). Although more faint satellites around the MW and M31 have been discovered during the past two decades or so (Grebel, Dolphin & Guhathakurta, 2000; van den Bergh, 2000b; Zucker et al., 2004; Willman et al., 2005; Zucker et al., 2006, 2007; Irwin et al., 2007; Martin et al., 2006; Liu et al., 2008; Martin, de Jong & Rix, 2008; Simon & Geha, 2007; Watkins et al., 2009; Belokurov et al., 2008, 2009, 2010), the number of observed satellites is still much smaller than the prediction. This problem may indicate some mechanism shuts down galaxy formation at small scales (Taylor & Babul, 2004; Taylor, Silk & Babul, 2004), or may challenge the CDM model and favour other dark matter models such as the Warm Dark Matter (WDM) model (Moore et al., 2000; Spergel & Steinhardt, 2000; Yoshida et al., 2000; Bode, Ostriker & Turok, 2001; Craig & Davis, 2001; Zavala et al., 2009; Lovell et al., 2011). Another possible explanation for this discrepancy may be the incompleteness of the observed satellite populations of the MW and M31 (Simon & Geha, 2007; Walsh, Willman & Jerjen, 2009; Willman et al., 2004; Koposov & Belokurov, 2008). Nevertheless, the MSP is of great importance and contains valuable implications for structure formation and galaxy formation.

## 1.3   The Local Group

The Local Group (LG) is the group of galaxies consisting of the MW and nearby galaxies, including the Andromeda galaxy (M31) and the satellite galaxies of the MW and M31. M31 is the nearest bright, spiral galaxy to the MW and is about 780 kpc from us (Ribas et al., 2005; McConnachie, 2012). It has a similar (though probably slightly higher) luminosity and mass as the MW (Karachentsev & Kashibadze, 2006). Both host halos of the MW and the M31 are estimated to have virial radii of $\sim$ 300 kpc (Klypin, Zhao & Somerville, 2002), so their halos do not quite overlap with each other.

The LG is the best measured galaxy system we have. It has a large population of nearby dwarf galaxies, with over 50 confirmed satellite members and more being discovered in the recent years. McConnachie (2012) reviews all the 100 nearby (within 3 Mpc from the Sun) dwarf galaxies that were known by 2012, of which 76 galaxies are very likely to be members of the LG. Among these LG members, there are 27 galaxies that are within 300 kpc (the suggested virial radius of the MW halo), which are all likely bound satellite galaxies of the MW, with one exception of Leo I (the current velocity measurements of Leo I suggest it may be unbound with the MW halo unless its tangential velocity components

are overestimated ([Zaritsky et al., 1989](); [Mateo, Olszewski & Walker, 2008]()). ). In addition, there are three more galaxies (Leo T, NGC 6822 and Phoenix) that have distances larger than 300 kpc, but are likely to be dynamically bound with the MW ([McConnachie, 2012]()), which makes them the potential satellites of the MW. For the M31 sub-group, there are 21 galaxies within 300 kpc (suggested virial radius of the M31) from the center of M31, with two of them (Andromeda XII and XIV) possibly unbound given their relatively large tangential velocities, assuming the halo mass of M31 to be $10^{12} M_\odot$ ([McConnachie, 2012]()). Additionally, the Pegasus dwarf galaxy (DIG) and IC 1613 are also the potential satellites of M31, as they are arguably dynamically associated, just as Leo T, NGC 6822 and Phoenix are likely associated the MW. The associations of the remaining LG members (excluding the MW and M31) with any dynamical subgroup are ambiguous.

The dwarf galaxies of the LG have been measured within a very broad luminosity range. The satellites of the MW and M31 are distributed in the V-band absolute magnitude range of $-17 \leqslant M_V \leqslant -6$, and the least luminous members (candidates, e.g. Segue I, Segue II, Willman I, Bootes II) of the LG have V-band absolute magnitude down to $M_V \sim 1 - 3$ ([McConnachie, 2012]()). The spatial distribution of the satellites of the MW appears to be anisotropic. The 11 innermost satellites are distributed almost perpendicular to the MW disk ([Lynden-Bell, 1982](); [Kroupa, Theis & Boily, 2005]()). However, there may be more undiscovered satellites that are close to the plane of the MW, due to the extinction of the galactic disk. This possibility of unknown satellites make it more complicated to draw a firm conclusion on the real isotropy of the MW satellites. We will determine the alignments of the nearby dwarf galaxies outside the LG with their central galaxies, and discuss the implications in section [4.2]()

## 1.4   Outside the Local Group

The Local Group is, and probably will always be, our best-observed system. Given the high quality of the LG data, they play an essential role in studying problems like the missing satellite problem. However, before drawing any conclusion from the LG studies to explain galaxy evolution in the broader universe, we need to know whether the LG and its halos are typical enough to represent other similar systems, which remains an open question ([van den Bergh, 2000a](); [Weisz et al., 2011](); [Lovell et al., 2011]()). Currently it is still very difficult for numerical simulations and semi-analytic studies to fully reproduce the MW satellite population, especially to reproduce the two extremely bright satellites (LMC and SMC) of the MW ([Benson et al., 2002](); [Koposov et al., 2009](); [Okamoto et al., 2010](); [Liu et al., 2011]()), which are also suggested to be observationally rare in MW-like systems ([Robotham et al.,

2012).

To answer the question of how typical the LG is, we obviously need the observations of dwarf galaxies in other systems outside the LG that are similar to the MW and M31. In this way, we will be able to study the satellite properties such as the abundance and spatial distribution in a broader cosmological context. A better understanding of the statistical behaviour (e.g. average properties and statistical scatter) of these systems will help to determine whether the LG represents a typical case or a statistically rare one. In addition, studying the dwarf galaxies outside the LG allows us to understand satellite behaviour in various cases beyond the MW-like systems. We will be able to study the dependence of the satellite populations on the properties of the central galaxies and the local environment, which has useful implications to galaxy formation. We will analyse of the dependence of satellite behaviour on the central galaxy properties in chapters 3 to 5.

This thesis is organized as follows. In Chapter 2, we introduce the COSMOS survey, the catalogue we use and our galaxy selection criteria. In Chapter 3, we present a measurement of clustering between our bright primary sample and faint secondary sample, using the two-point correlation function. In Chapter 4, we present an analysis of the spatial distribution of the satellites, with a calculation of their radial density distribution and their angular alignment with the orientation of their central galaxies. In Chapter 5, we present an analysis of satellite abundance as a function of luminosity relative to the central galaxies. In Chapter 6, we discuss some of the possible systematic uncertainties in our measurement. In Chapter 7, we briefly present a preliminary analysis of satellite populations using the CFHT Stripe 82 survey. Finally, we summarize our results in Chapter 8.

# Chapter 2

# Data and Reduction

## 2.1  The COSMOS Survey

The Cosmic Evolution Survey (COSMOS) is a project that is designed to probe various astronomical topics over a large redshift range that extends up to $z \sim 6$, covering 2 $\deg^2$ sky area. The survey includes multi-wavelength imaging and spectroscopy from X-ray to radio wavelength, including the Hubble Space Telescope (HST) imaging (Scoville et al., 2007). The wide wavelength coverage and multiple bands give the potential for accurate photometric redshift measurement (Ilbert et al., 2009).

## 2.2  The COSMOS Photo-z Catalogue

The clustering (satellite) study intensively relies on a reliable measurement of redshift. Spectroscopic redshifts provide robust and highly accurate estimates, but they are extremely expensive and inefficient for faint objects, which are, however, very important for satellite study. On the other hand, the photometric redshifts (photo-zs), which are based on the observed color in a few broad-band filters (e.g. 5 in the case of the SDSS survey), are a cheap, efficient method to collect redshift information both for bright and faint galaxies. However, they usually have large uncertainties and do not meet the accuracy requirements for studying small-scale clustering.

The COSMOS survey has produced a photometric redshift catalogue using a very large number of bands, which provides a compromise between these two techniques. Is has 30-

7

bands in total, including some intermediate or narrow bands, compared to the five broad bands using by SDSS.

## 2.2.1  Photo-z Derivation

The photometric redshift in this catalogue is based on a $\chi^2$ template-fitting procedure, using the *LePhare* code [1] (S.Arnouts & O.Ilbert) (Ilbert et al., 2009). The full galaxy SED template library includes 7 SED templates of elliptical galaxies and 12 templates of spiral galaxies and 11 SEDs of starburst (SB) galaxies, which are shown in Fig. 2.1. For each template, the code scales the SED to different redshifts, with a grid spacing of $\delta z = 0.01$. Then it compares the predicted SED to the observed flux from the 30 filters. The merit function $\chi^2(z, T, A)$ is defined as (Ilbert et al., 2009):

$$\chi^2 = \sum_{f=1}^{N_f} \left( \frac{F_{obs}^f - A \times F_{pred}^f(z, T) \; 10^{-0.4 s_f}}{\sigma_{obs}^f} \right)^2 \tag{2.1}$$

where $N_f = 30$, refers to the number of filter; $F_{obs}^f$ refers to the observed flux at filter $f$; $F_{pred}^f(z, T)$ refers to the predicted flux at filter $f$, using template $T$ and fitting with redshift $z$; $A$ is the normalization; $s_f$ is the zero-point of the filter $f$; $\sigma_{obs}^f$ refers to the uncertainty of the observed flux of the filter $f$. The photo-z is estimated from the minimization of the function $\chi^2(z, T)$. The redshift probability distribution function (PDFz) is also derived from the $\chi^2$ function:

$$P(z) \propto exp \left( -\frac{\chi^2(z) - \chi^2_{min}}{2} \right). \tag{2.2}$$

The $1\sigma$ error, based on the definition of PDFz, is estimated by:

$$\chi^2(z_{max}, z_{min}) = \chi^2_{min} + 1. \tag{2.3}$$

Throughout this thesis, the error of the photo-z is defined as $(z_{max} - z_{min})/2$.

Some examples of demonstrating the precision of this $\chi^2$ template-fitting method are shown in Fig. 2.2, where the predicted flux of the best fit $F(z, T)$ is compared to some of the filters.

By comparing the photo-z redshift $z_p$ to corresponding spectroscopic redshift $z_s$ (e.g. from zCOSMOS (Lilly et al., 2007)), the quality of the photo-z can be evaluated with

---

[1] www.oamp.fr/people/arnouts/LE_PHARE.html
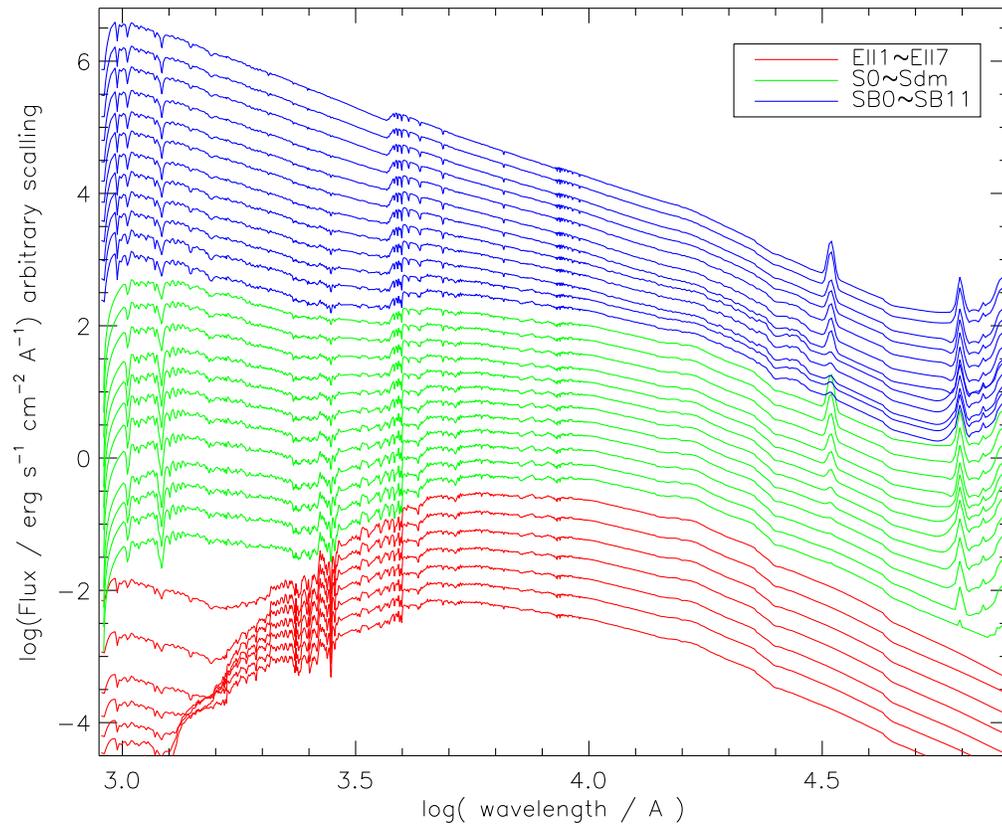
Figure 2.1: The 31 templates that the COSMOS photo-z catalogue used.
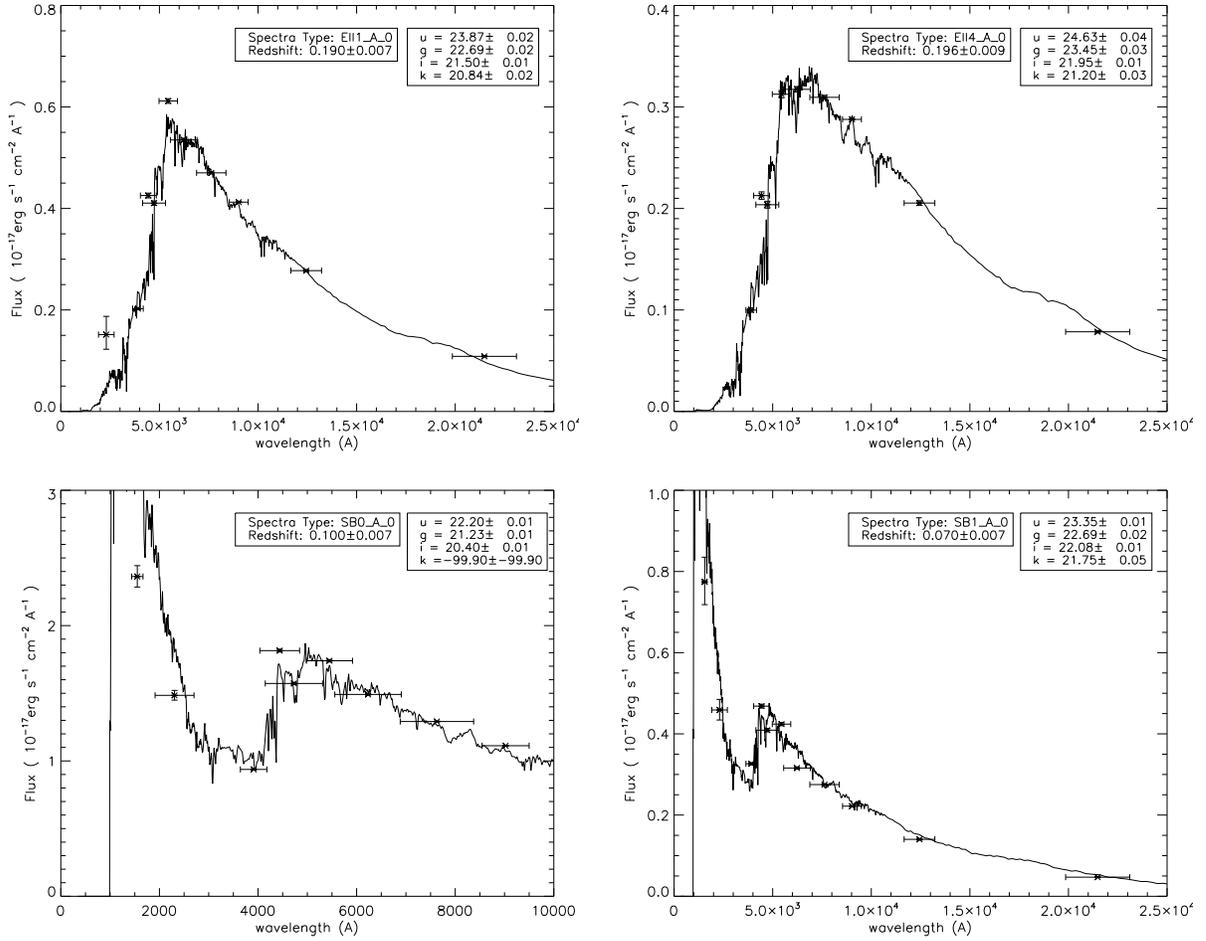
Figure 2.2: Four examples of how the spectral templates fit the photometry in multiple filters to estimate the redshift.

parameter $\sigma_{\Delta z} = |z_p - z_s|/(1\sigma \; error)$. The quality shows two feature (Ilbert et al., 2009): The photo-z has significantly higher uncertainty at high redshift ($z \geq 1.25$); The accuracy of the photo-z is high for bright objects ($\sigma_{\Delta z}/(1+z_s) = 0.007$ at $i^+ < 22.5$ (Ilbert et al., 2009)), slightly worse for median magnitude ($\sigma_{\Delta z}/(1+z_s) = 0.011$ at $i^+ \sim 23.1$, and is significantly degraded for very faint objects ($\sigma_{\Delta z}/(1+z_s) = 0.053$ with a catastrophic failure rate of 20% for objects with $24 < i^+ < 25$).

## 2.3 The Spectroscopic Catalogue

In addition to the photo-z catalogue, we also use redshifts from the internal spectroscopic catalogue of the COSMOS collaboration (updated on October 9th, 2012), for the objects that have these available. The redshifts of this catalogue are compiled from various sources, including the NASA extragalactic database (NED) and the z-COSMOS survey (Lilly et al., 2007). The z-COSMOS is a large spectroscopic redshift survey covering the COSMOS field, using Visible Multi-Object Spectrograph (VIMOS) on the 8-meter Very Large Telescope (VLT). The survey contains two parts: zCOSMOS-bright and zCOSMOS-deep. The zCOSMOS-bright contains about 20,000 bright galaxies with a I-band criteria of $I_{AB} < 22.5$ and with redshift range $0.1 < z < 1.2$, covering the whole COSMOS Advanced Camera for Survey (ACS) field; While the zCOSMOS-deep only covers the 1 deg$^2$ central area of the COSMOS ACS field but with higher redshift $1.4 < z < 3.0$, and contains about 10,000 galaxies.

The spectroscopic catalogue allows independent test for the accuracy of the 30-band photometry method and directly provides highly accurate redshift estimates for a part of the sample we used in this study.

## 2.4 The COSMOS Shape Catalogue

Compared to ground-based observation, the space telescopes can have better resolution as they are free from the effects of atmospheric seeing. The best seeing for ground-based observation is typically 0.5-1.0", while the full width half-maximum (FWHM) of the point-spread function (PSF) of the Advanced Camera for Surveys (ACS) and the Wide Field Camera (WFC) on the Hubble Space Telescope (HST) is 0.12" (Leauthaud et al., 2007). The high resolution can provide a huge advantage of resolving close pairs, which is very helpful for our purposes.
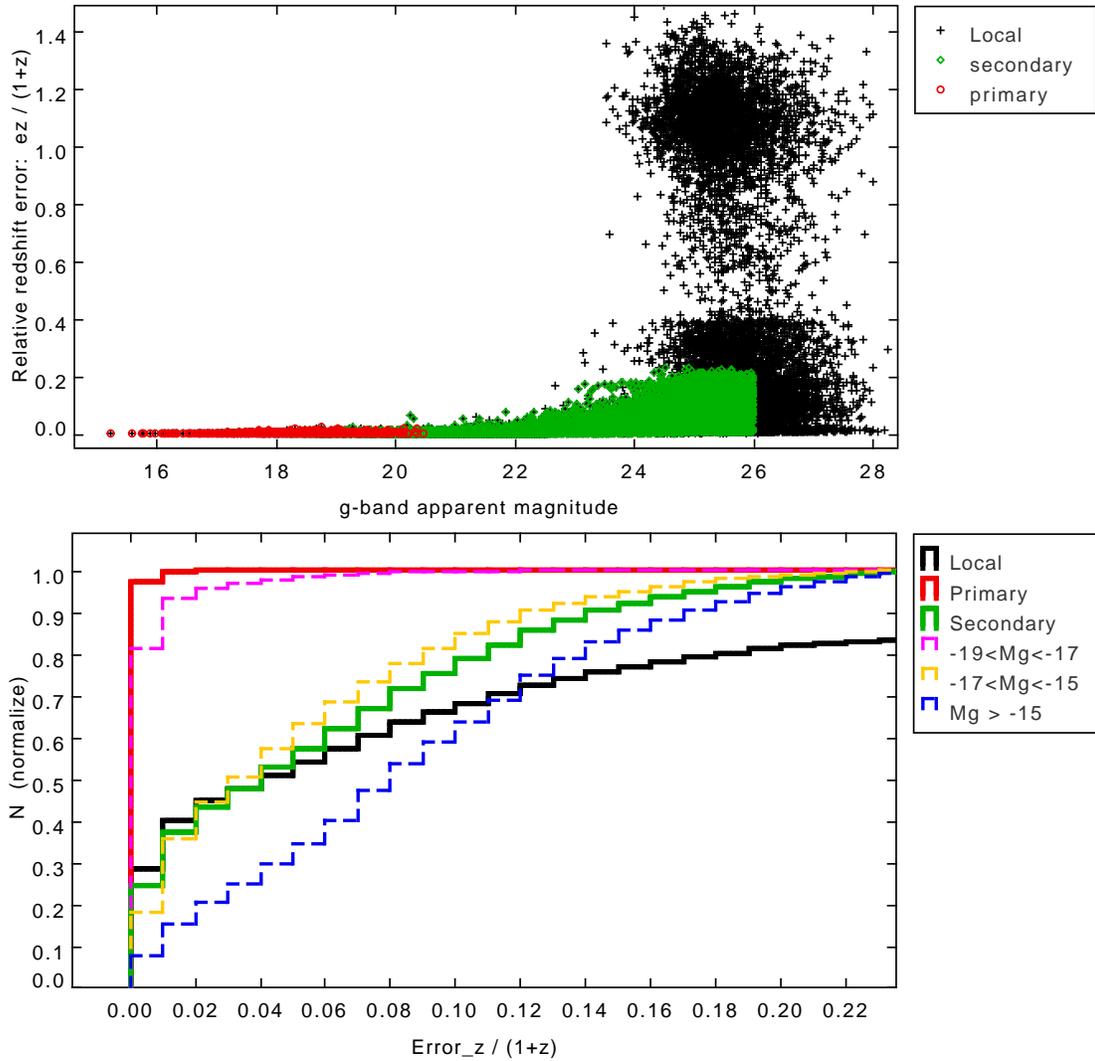
Figure 2.3: Photo-z quality of the COSMOS shape catalogue (after the photo-zs of some bright galaxies are corrected using their spectroscopic redshift). The top plot shows the distribution of photo-z error as a function of the g-band apparent magnitude; the bottom plot shows the normalize cumulative counts as a function of photo-z error. The local sample (black) is cut at $z < 0.3$, while the selections of the primary (red) and the secondary sample (green) are described in section 2.6 and section 2.7. Note the secondary sample is also separated into three magnitude range bins(pink, yellow, blue) in the bottom plot.

The COSMOS shape catalogue was constructed by Leauthaud et al. (2007), with the primary goal of minimizing the effects of contaminations by stars, cosmic rays, diffraction spikes and other artifacts, providing clean PSF-corrected shape measurement, together with the high quality COSMOS photometric redshift. They first extracted objects from the ACS/WFC imaging with a "Hot-Cold" method employed by Rix et al. (2004), which runs SExtractor[2] twice, once with a configuration optimized for detecting only bright objects ("cold" step) and once with a configuration optimized for detecting faint objects ("hot" step). They then constructed the base catalogue by combining the two samples produced separately from the two steps. This two-step method helped to improve the detection of close pairs and improved the completeness of the very luminous objects and the very faint objects (Leauthaud et al., 2007). Additionally, a carefully defined (generally based on automatic algorithm and made by hands in a few cases when the algorithm fails), polygonal-shaped mask is created around each bright object during the "cold" step, removing a certain number of spurious detections. Afterwards, instead of using the automatic classification, they classified the galaxies/stars on the peak-surface-brightness-to-magnitude (both parameters are generated from the SExtractor) plane, which was tested to be a more robust method (Leauthaud et al., 2007). Finally, the catalogue was carefully cleaned again by masking out other contaminated regions (including reflection ghost, asteroids and satellite trails) and removing astronomical sources such as HII regions around bright galaxies and stellar clusters. In the end, the completed catalogue contains $1.2 \times 10^6$ objects ($3.9 \times 10^5$ galaxies) with a limiting magnitude of F814W = 26.5 after all those cuts (Leauthaud et al., 2007).

The COSMOS photo-z catalogue provides the 30-band redshift information for 89% of the objects selected from the ACS/WFC imaging while the remaining 11% of the objects are mainly small objects that cannot be covered by the ground-based imaging and thus have no valid photo-zs. In addition, there are larger areas are masked out for the ground-based imaging and the objects inside these areas do not have good photo-zs, either. Overall, there are about 73% objects of the shape catalogue with accurate photometric redshifts, including $2.8 \times 10^5$ galaxies (Leauthaud et al., 2007). Fig. 2.3 shows the quality of the photo-z of the local galaxies of the shape catalogue. Note that a part of the redshifts (mainly bright galaxies) are revised using the COSMOS spectroscopic redshifts. Fig. 2.4 shows the photometric redshift versus g-band absolute magnitude distribution at low redshift for the COSMOS shape catalogue, as well for our selected primary and secondary samples.

---

[2]SExtractor is a program that builds a catalogue of objects from an astronomical image, also see Bertin & Arnouts (1996)
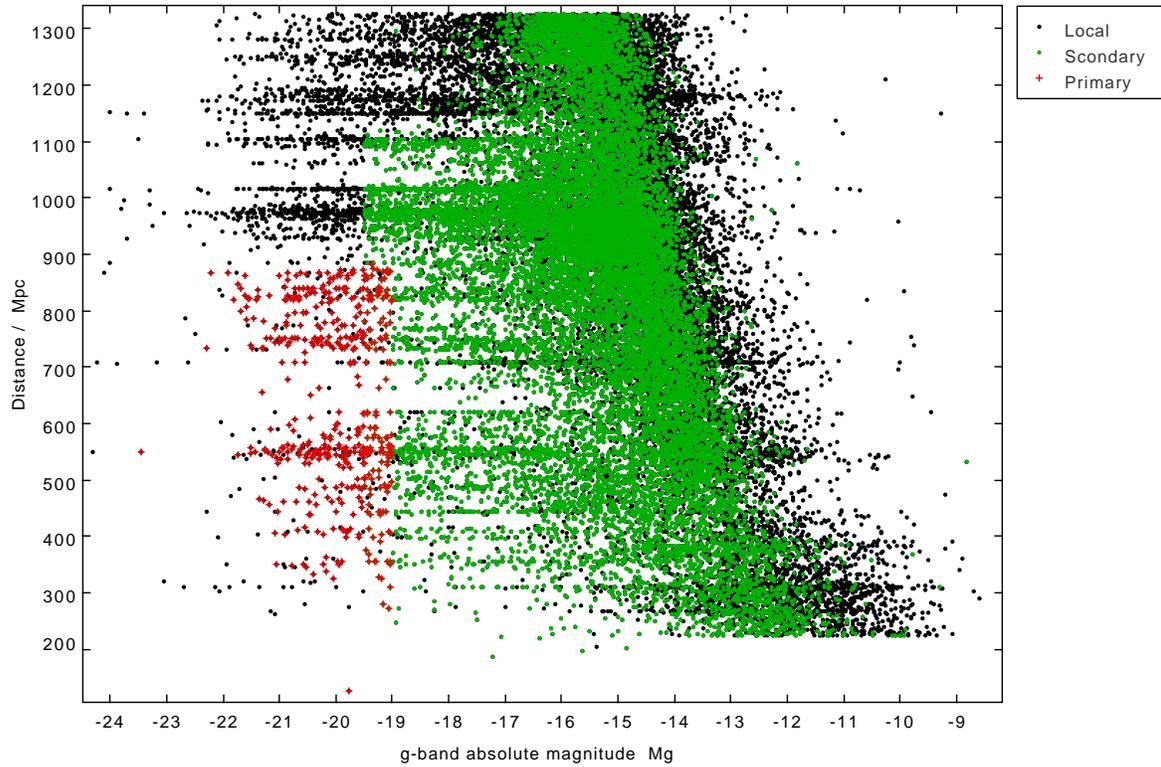
Figure 2.4: The Distance along the light-of-sight versus g-band absolute magnitude of the shape catalogue (after using the spectroscopic catalogue for redshift correction) at low redshift. Our primary samples are marked with red crosses, the secondary samples are marked with green dots.

## 2.5 Masking the Field

Halos and scattered light from other stars, bad pixels and other problems may cause flaws in certain images on an image. Galaxy detection and photometry measurement may be unreliable in these regions. A common strategy to reduce catalogue contamination is to mask these regions out, as was done with the COSMOS shape catalogue (Fig. 2.6, left panel). The masks are designed to fit the shape of the Point Spread Function (PSF) of the stars and other artifacts (holes, spikes, ect), so that it prevent us from using the low quality data and not to lose too much effective information at the same time. However, it also brings more complicated boundary effects for any further statistical studies. It is important to determine the shapes of the masked regions and to apply this consistently for the clustering analysis.

In the COSMOS shape catalogue, the parameter "Mask-s" indicates that whether the object should be masked or not. Besides deciding whether to mask out data point, we need to produce a sky description of the masking to ensure that we can produce a random sample that has the exactly same spatial distribution with the data. In order to determine the shape the masked regions, we divided the ACS field into $N \times N$ grids and counts how many effective data points are inside each grid cell, whereas the grid cells containing "no objects" (lower number of objects than certain threshold) will be defined as the masked regions. Fig. 2.5 shows a histogram of the counts per grid in four resolution choices of making the grids. Although we want high resolution to map out mask boundaries, excess resolution runs the risk of over-estimating the masked areas. This is because the low average count per grid, as a consequence of the high resolution (e.g. the case of $400 \times 400$ resolution in Fig. 2.5), can introduce a high Poisson noise, which will then require a higher threshold of counting to deal with. However, a higher threshold means a sacrifice of more effective data points.

As the comprise between mapping out the boundaries of masked regions in a high resolution, as well as keeping relatively complete data, we chose a resolution of $300 \times 300$ and a threshold of two points per grid and defined all the grids with point number equal or below the threshold to be the masked regions. The masking that we defined will be later used to generate a random sample that matches with the secondary sample. Additionally, we also excluded all the data points of the secondary sample in the masked regions to guarantee that the random sample and the secondary sample will have almost the same 2-dimensional distribution on the sky.
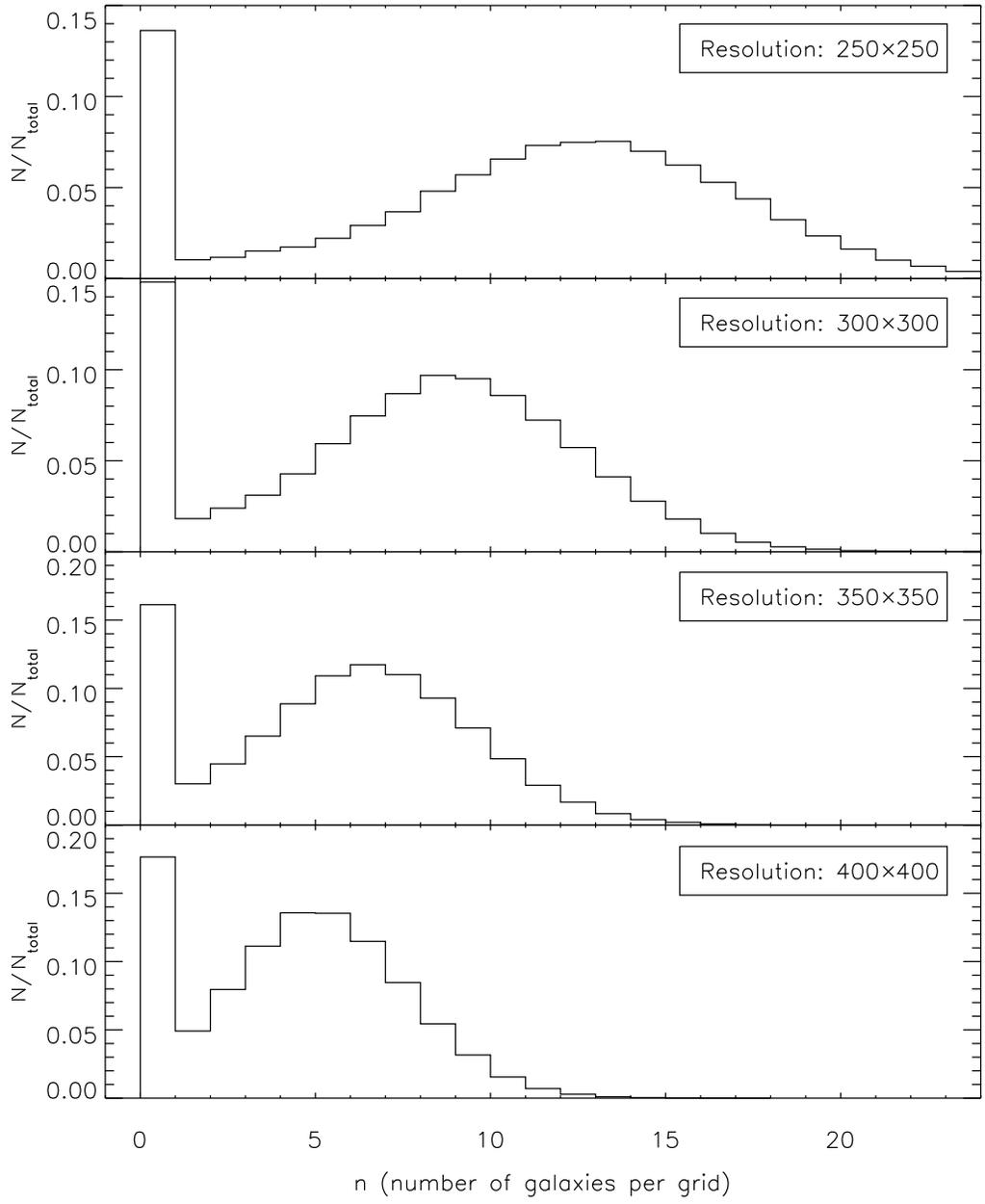
Figure 2.5: Distribution of object counts per grid cell for four different grid resolutions.
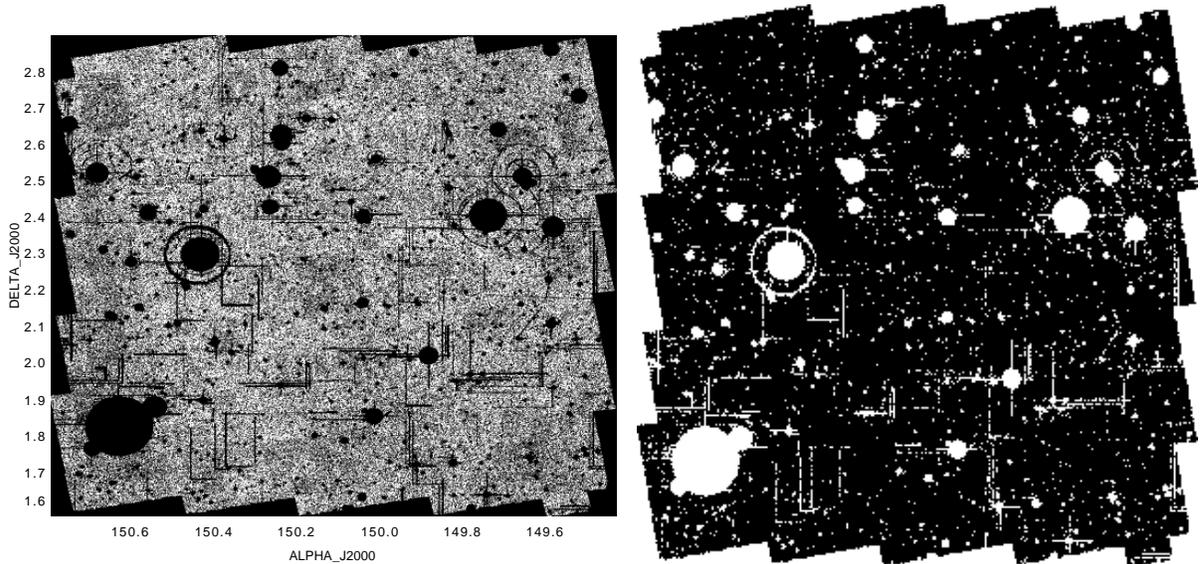
Figure 2.6: A comparison of sky coverage between the masked data COSMOS shape catalogue and the masking matrix we used in our study. The left plot shows the sky distribution of masked the COSMOS shape catalogue, while the right plot shows the masking matrix we get, based on the method described in the masking section. Notes the RA increases from right to left for both panels.

## 2.6 The Primary Sample

The detailed construction steps of our primary sample are described as below, along with the number of remaining galaxies after each step shown in the brackets:

1. We use the COSMOS shape catalogue as the base catalogue (737,932 galaxies in total);

2. We apply a redshift cut $0 < ZPHOT$ (photo-z) $< 0.3$ to select the relatively local galaxies (34,274 galaxies left);

3. We revise the redshift, using the spectroscopic catalogue: we matched these pri-

17

maries to the low redshift galaxies from the spectroscopic catalogue, and corrected the redshift of the galaxies that are also available in the spectroscopic data (34,274);

4. We apply a further redshift cut: $0 < mZPHOT$ (photo-z with spectroscopic redshift correction) $\leq 0.2$ (15,858);

5. We apply a magnitude cut of the g-band absolute Magnitude $-50.0 < M_g < -19.0$ (623);

6. We exclude the data in the masked regions, selecting data with parameter "Mask_s" == 1; (502)

7. We apply an apparent magnitude cut: We found a part our primary selection are abnormally faint ($g^+ \sim 21.5 - 25.0$) in g-band magnitude, indicating errors in the absolute magnitude estimates. We exclude them by cutting the sample with: $g^+ < 21.0$. (486)

8. We add supplementary primary samples from the SDSS: We searched the SDSS DR9 photometric and spectroscopic database by searching for galaxies that have magnitude of $g^+ < 18.0$ within the ACS field. We found 5 additional galaxies that meet the criteria above but not are included in the our primary list. And three of them also meet our absolute magnitude criteria: $M_g < -19.0$. We added these three galaxies into our primary sample. (489)

9. We apply an additional "Kick-out" cut (see section 2.8 for further explanation): Despite the bright magnitude cut we applied, the sample we selected may still contain galaxies that are not truly primaries but satellites of other brighter galaxies. We define the galaxies that have one or more brighter companions ($\Delta_{M_g} > 1.0$, $\Delta_V < 300km/s + 1\sigma_V$, $\Delta_{r_p} < 600$kpc) to be potential satellites. We cut out these galaxies (96) from the primary and build a "Kick-out" primary list. (393 galaxies in the "Kick-out" primary sample.)

After the reduction, there are 489 bright galaxies left as our primary sample (without the additional "Kick-out" cut). Fig. 2.7 show the redshifts of the primaries versus their g-band absolute magnitude, where the green dots are the primaries whose redshifts have been revised, using the spectroscopic catalogue, the blue dots are the three supplementary galaxies obtained from the SDSS. The top left and top right panels in Fig. 2.8 show the g-band absolute magnitude and apparent magnitude distribution of the primary sample.
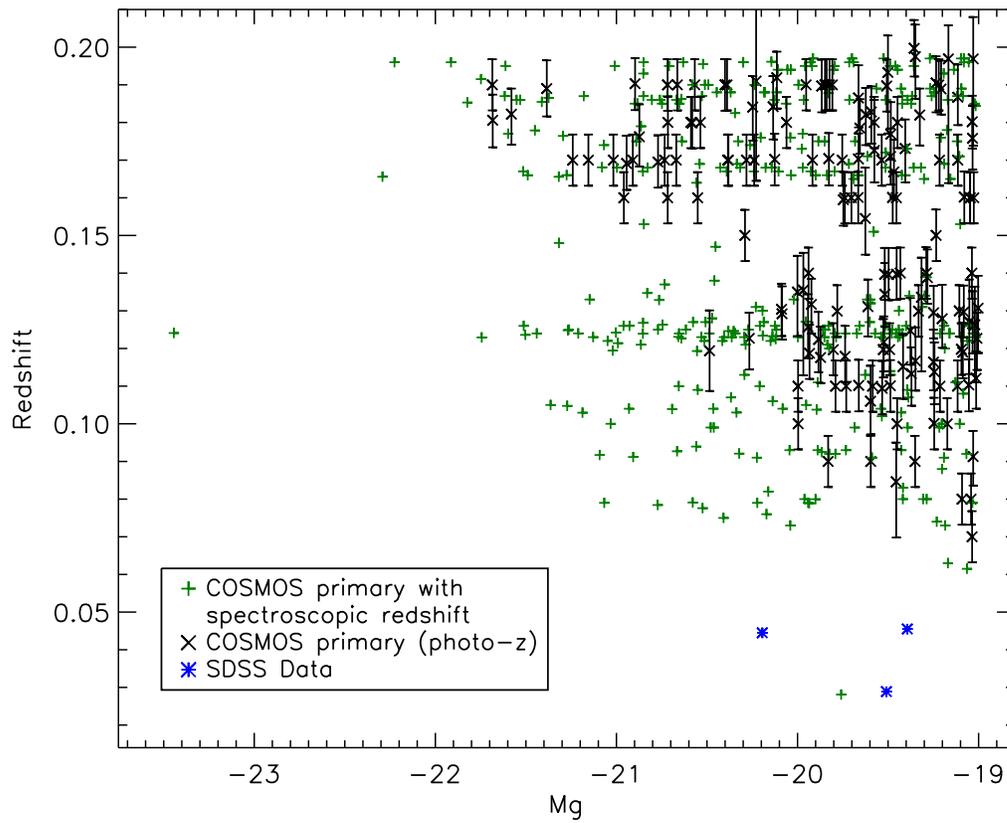
Figure 2.7: The g-band absolute magnitude–redshift distribution of the Primary sample

## 2.7 Cleaning the Secondary Sample

We used the shape catalogue as base catalogue and picked the galaxy samples within redshift range of 0.0 to 0.25 ($z_{low68} \leq 0.25$ in this case, using $z_{low68}$ instead of z and a slightly higher up limit to ensure that the secondary sample includes all the possible faint galaxies that may cluster with the primary galaxies) and with the absolute magnitude in g-band fainter than -19.5 as the secondary sample. Then we matched the sample to the spectroscopic catalogue and replace the redshift of the galaxies that were also available in the spectroscopic catalogue, reducing the error in the redshift to zero.

The detailed construction steps of our secondary sample are described as below, along with the number of remaining galaxies after each step shown in the brackets:

1. The sample was initially built based on the COSMOS shape catalogue (737,932);

2. We applied a redshift cut: $0 < ZPHOT$(photo-z) $< 0.3$ (34,274);

3. We revised the redshift using the spectroscopic catalogue: we matched these galaxies to the low redshift galaxies ($z < 0.6$) from the spectroscopic catalogue, and corrected the redshift of the galaxies that are also available in the spectroscopic data (34,274);

4. We applied a further redshift cut of $0 < z_{low68} < 0.25$ (31,135), where $z_{low68}$ the modified redshift lower limit. Note that we use a lower limit and a relatively larger range than the primary to include relatively complete secondary sample in case that some faint galaxies have poor photometric redshift.

5. We applied an absolute magnitude cut: $-19.5 < M_g < -0.1$; (29,947)

6. Data masking: using mask parameter "mask_s", we selected objects in the unmasked regions; (22,638)

7. We applied a further apparent magnitude cut using g-band magnitude: $MAG\_AUTO < 26.0$. (20,7033)

8. We excluded the galaxies with large redshift errors: $\sigma_z = (z_{high68} - z_{low68})/2 < 0.25$. (16,702)
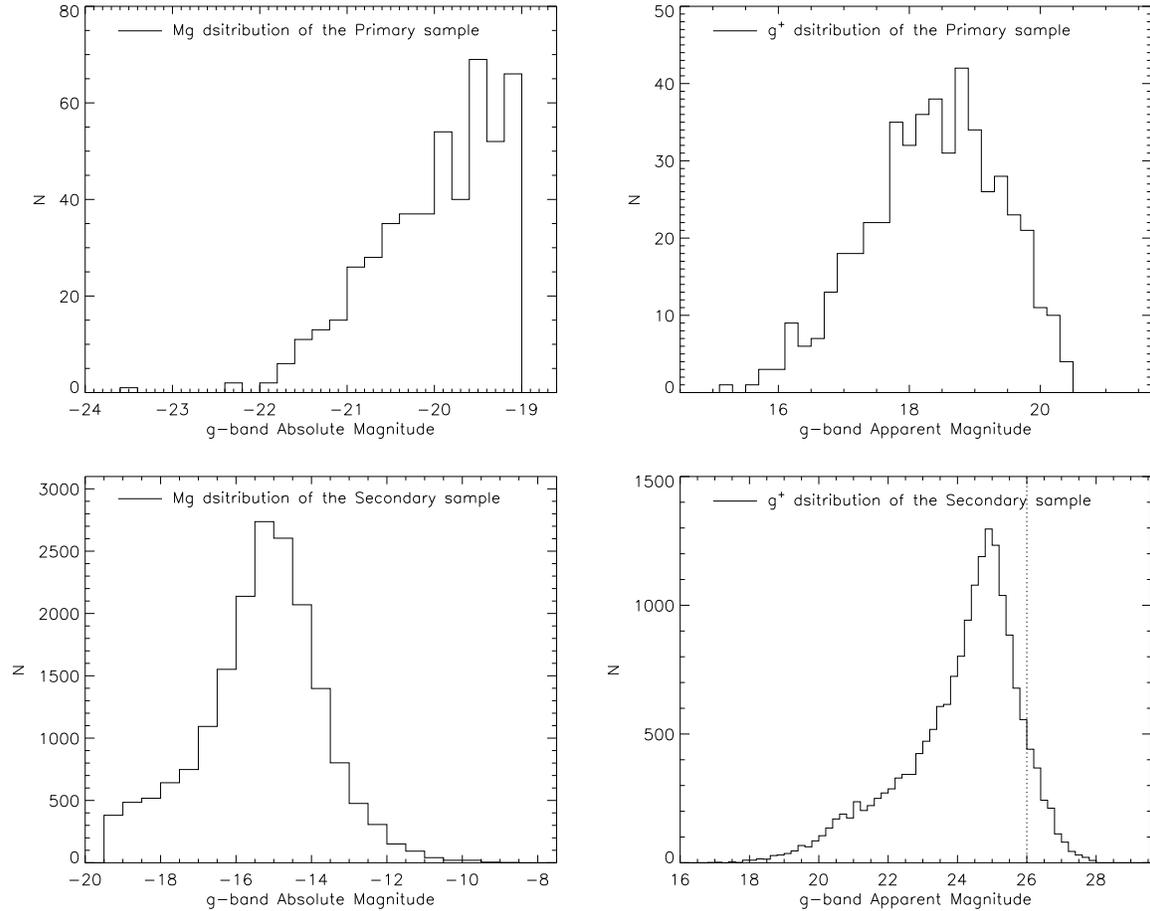
Figure 2.8: The first plot shows the distribution of absolute g-band magnitude of the primary sample; the second one shows the distribution of the apparent magnitude of the primary sample; the third one shows the distribution of the absolute magnitude of the secondary sample; the last one shows the distribution of apparent magnitude of the secondary sample with the $g = 26.0$ cut we applied to the sample.

## 2.8 Properties and Further Sub-samples

We looked at the detailed magnitude distribution of our primary and secondary sample, shown in Fig. 2.8.

Previous studies usually focus on the primary galaxies that are relatively isolated from other bright galaxies (Speller & Taylor, 2014; Nierenberg et al., 2012; Wang & White, 2012; Wang et al., 2014; Strigari & Wechsler, 2012), which probably correspond to individual dark matter halo systems as opposed to more massive groups . Focusing the isolated primaries has an advantage of being able to easily avoid satellite over-counting but it also dramatically reduces the sample size. Specifically, the isolation cut is a particular unfavored choice for the COSMOS in term of sacrificing the sample size. Since the COSMOS field only contains few extremely bright sources, we have choose a relatively faint luminosity limit to achieve a reasonable number size of our primary sample. As a result, the surface density of our primaries is relatively high, which makes it less likely for the primaries to be isolated with each other (see Section 6.4 for more detailed discussion).

Instead of an isolation cut, we applied a "Kick-out" technique to clean our primary sample. One of the main motivations of the isolation cut and many other primary selection criteria is to select the galaxies that are the most likely to be the central galaxies of their systems, because otherwise (e.g. mistaking a satellite as the central galaxies), we would be looking at the unimportant objects that are representative for the systems. The goal of the "Kick-out" method is to find those candidates that are not truly the primaries of their systems and kick them out of our primary list. Specifically, we checked the regions within a projected distance of $r_p < 600$ kpc and velocity difference of $|\Delta V| < 300 km/s + \sigma_{\Delta V}$ around each primary (where $\sigma_{\Delta V}$ is the uncertainty of the relative velocity between two primaries). We chose the distance 600 kpc to match the range within which we searched for satellites, also see section. 4.1.3. We chose 300 km s$^{-1}$ as a consertive relative velocity limit, as CDM simulations predict that typical orbital velocities of subhalos in MW-like host halos are around or below 200 km s$^{-1}$ (Springel et al., 2008). If we find another galaxy that is one or more magnitudes brighter, the fainter member of the pair is then identified as a secondary and is kicked out of the primary list. By applying this technique, we can ensure that all the galaxies in our primary list are genuinely the dominant galaxies in their immediate neighbourhood. By doing the "Kick-out" cut, we reduced the number of the primary sample from 489 to 392 galaxies. The size of the remaining "Kick-out" primary sample is still fairly large, which allows us to split it into sub-samples for further analysis.

Table 2.1: Morphological classification from NED for the three extra primary galaxies from SDSS.

| Object name (from NED) | Ra(degree) | Dec(degree) | Type | Reference |
|---|---|---|---|---|
| 2MASX J10005519+0223437 | 150.23005 | 2.395523 | S1 | Véron-Cetty & Véron (2006) |
| 2MASX J10000814+0245542 | 150.03379 | 2.765140 | Sb | Nair & Abraham (2010) |
| DSS J100236.57+014836.0 | 150.65241 | 1.810012 | Sc | N/A |

"N/A" means Not Available.

## The Bright and Faint Sub-samples

We split our primary sample into another two sub-samples using a luminosity criterion based on the g-band absolute magnitude: We defined the galaxies with $M_g < -20.5$ as the Bright sub-sample, which has 120 galaxies; correspondingly, we also define the galaxies with $M_g > -20.5$ as the Faint sub-sample, which includes 272 galaxies. Note that the full "Kick-out" primary sample has a mean absolute magnitude in g-band of $\langle M_g \rangle \approx -20.1$, the Bright primary sub-sample has $\langle M_g \rangle \approx -21.0$ and the Faint primary sub-sample has $\langle M_g \rangle \approx -19.7$

## The Elliptical/S0 and Spiral Sub-samples

We have the SED type information from the catalogue (Column "MODD"), which allows us to cut the primary into elliptical and spiral galaxies. For the three SDSS-based galaxies, we searched their images (shown in Fig. 2.9) and estimated their morphologies by eye. We then searched the NASA/IPAC Extragalactic Database (NED) for further morphological identification. The search results are shown in Tab. 2.1 . Note that the third galaxy only has the SDSS system classification, without any reference available. Nevertheless, the system classification of that galaxy (Sc) agrees with our manual morphology estimate.

In the end, we have 62 galaxies in the Elliptical/S0 sub-sample; and we have 330 galaxies in the Spiral sub-sample. Fig. 2.10 shows their g-band absolute magnitude versus their color. Note that the mean absolute magnitude in g-band of the Elliptical/S0 sub-sample is $\langle M_g \rangle \approx -20.2$, slightly brighter than the Spirial sub-sample, which has $\langle M_g \rangle \approx -20.1$.
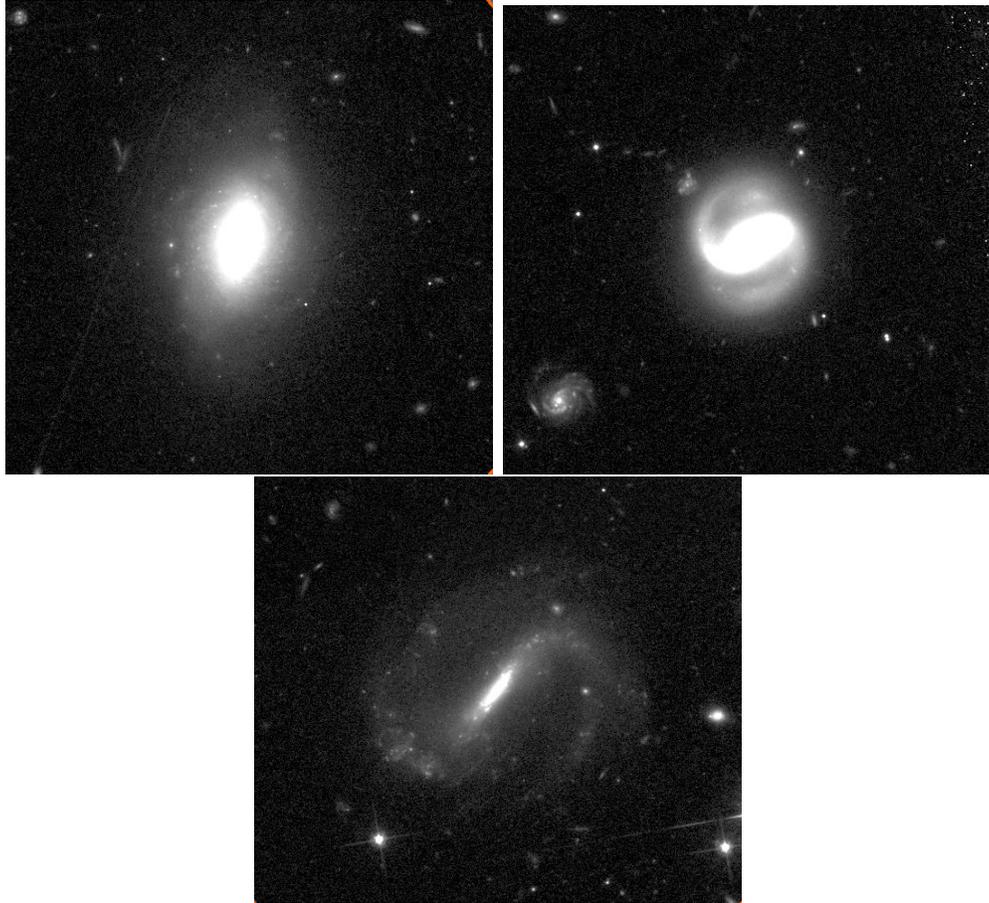
Figure 2.9: The images of the three SDSS-based galaxies. The images are obtained using the COSMOS skywalker (http:www.mpia-hd.mpg.deCOSMOSskywalker#). The morphological types of the three galaxies are identified as (from left to right, top to bottom): S1, Sb, Sc.

**The Red and Blue Sub-samples**

We also looked into the color index $B - V$ distribution of our primary, which is shown in Fig. 2.10. Based on the two peaks shown in the figure, we divided our primary sample into a Red ($B - V \leq 0.48$) sub-sample and a Blue ($B - V > 0.48$) sub-sample. The color of the three SDSS-based galaxies use another color index criterion, as we lack the B-band photometry: $g - r > 0.4$ to be Red, $g - r \leq 0.4$ to be Blue. Eventually, we have 197 galaxies in the Red sub-sample and 195 galaxies in the Blue sub-sample. Note that the Red sub-sample has a mean absolute magnitude in g-band of $\langle M_g \rangle \approx -20.3$, whereas the Blue sub-sample has $\langle M_g \rangle \approx -19.9$.
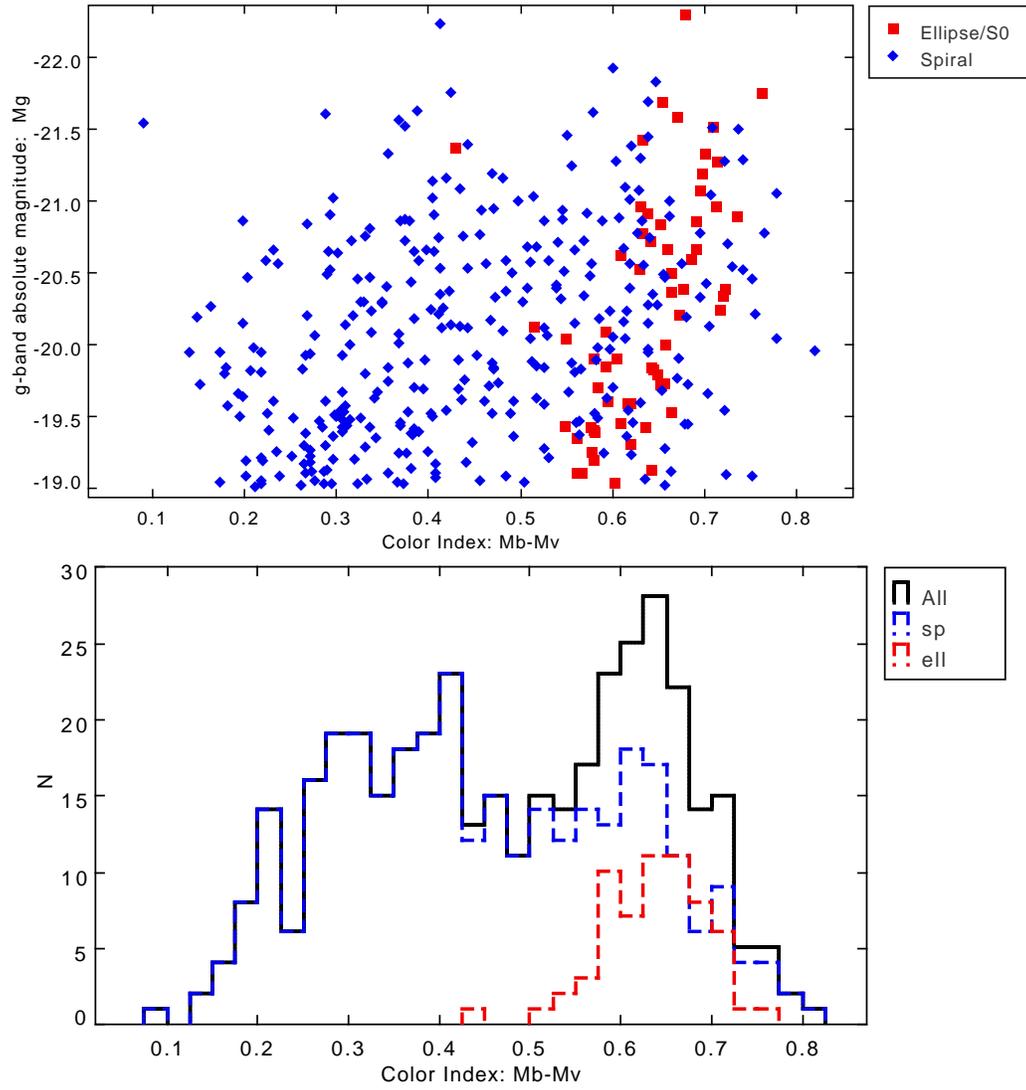
Figure 2.10: The color $(B - V)$— magnitude distribution the kick-out primary sample, Elliptical/S0 sample and Spiral sample.

# Chapter 3

# Clustering Analysis

## 3.1  Distance Calculation

There are many methods to determine the distance of an astronomical object. Among those methods, using the Hubble Flow to determine distance from its redshift is the most commonly used method. The redshift of an object is mainly contributed from its Hubble flow velocity, which is tightly related to the expansion rate and therefore varies at different epochs. If two objects are very close to one-another along the line of sight, we are seeing them at roughly the same epoch, and the expansion rate is constant. The separation along the line of sight of the two objects in the comoving space $\delta D_C$ is (Hogg, 1999):

$$\delta D_C = \frac{c}{H_0 E(z)} \, \delta z \qquad (3.1)$$

where $H_0$ is the Hubble constant, $\delta z$ refers to the redshift difference between the two objects, and $E(z)$ is a function of redshift that is defined as (Hogg, 1999):

$$E(z) \equiv \sqrt{\Omega_M (1+z)^3 + \Omega_K (1+z)^2 + \Omega_\Lambda} \qquad (3.2)$$

where $\Omega_M, \Omega_K, \Omega_\Lambda$ are the density parameters for mass, curvature, cosmological constant. Note that we did not include the radiation component as its contribution at low-redshift universe is negligible.

Thus the comoving distance along the line of sight of an object at redshift $z$ can be given by the integration (Hogg, 1999):

$$D_C = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}. \qquad (3.3)$$

27

The transverse comoving distance $D_M$ depends on the curvature and can be related to the comoving distance along the line of sight by $D_C$ (Hogg, 1999):

$$D_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_k}} sinh[\sqrt{\Omega_K} D_C/D_H] & \text{for } \Omega_K > 0 \\ D_C & \text{for } \Omega_K = 0 \\ D_H \frac{1}{\sqrt{\Omega_k}} sin[\sqrt{\Omega_K} D_C/D_H] & \text{for } \Omega_K < 0 \end{cases} \quad (3.4)$$

where $D_H = \frac{c}{H_0}$ is the Hubble distance. Based on observation, the curvature is, if not zero, negligible. Therefore, $D_M = D_C$ is a good approximation.

Another useful concept is the Angular Diameter Distance $D_A$, which is defined as:

$$D_A = \frac{x}{\theta} \quad (3.5)$$

where $x$ is the actual size of an object (or actual transverse separation of two objects), and $\theta$ is its observed angular size (the observed angular separation). The angular diameter distance can be relate to the comoving distance by (Hogg, 1999):

$$D_A = \frac{D_M}{1+z} = \frac{D_C}{1+z} \quad (3.6)$$

Here I have also listed several cosmology models for the distance calculation (Hinshaw et al., 2013; Planck Collaboration et al., 2013), shown in Table.3.1

Table 3.1: Cosmological parameters, based on the results of WMAP9 and Planck experiments.

| Parameter | WMAP-9 | Planck | | Planck+WP | |
|---|---|---|---|---|---|
| | | Best fit | 68% limits | Best fit | 68% limits |
| $\Omega_b h^2$ | $0.02264 \pm 0.00050$ | 0.022068 | $0.02207 \pm 0.00033$ | 0.022032 | $0.02205 \pm 0.00028$ |
| $\Omega_c h^2$ | $0.1138 \pm 0.0045$ | 0.12029 | $0.1196 \pm 0.0031$ | 0.12038 | $0.1199 \pm 0.0027$ |
| $\Omega_\Lambda$ | $0.721 \pm 0.025$ | 0.6825 | $0.686 \pm 0.020$ | 0.6817 | $0.685^{+0.018}_{-0.016}$ |
| $H_0$ | $70.0 \pm 2.2$ | 67.11 | $67.4 \pm 1.4$ | 67.04 | $67.3 \pm 1.2$ |

$H_0$ is in unit of $km\ s^{-1} Mpc^{-1}$

Throughout this work, I use the $Planck + WP$ cosmology, which specifically includes:

$$H_0 = 100h\ km\ s^{-1} Mpc^{-1} = 67.0\ km\ s^{-1} Mpc^{-1} \quad (3.7)$$
$$\Omega_m = (\Omega_b h^2 + \Omega_c h^2)/h^2 = 0.32 \quad (3.8)$$
$$\Omega_\Lambda = 0.68 \quad (3.9)$$
$$\Omega_K = 1 - \Omega_m - \Omega_\Lambda = 0 \quad (3.10)$$

## 3.2 The Clustering Calculation

### 3.2.1 The Two-point Correlation Function

The Two-point Correlation Function (2PCF) is one of the most common and robust statistical methods for characterising the clustering of galaxies. It is widely used for large-scale-structure studies, tracing the massive galaxies in the universe. We find it can be also particularly useful at small scales, tracing the abundance of faint galaxies and measuring their clustering with respect to their host galaxies.

The application of the 2PCF in Astronomy started with the angular correlation functions, when redshift catalogues were rare. These early studies (Davis & Geller, 1976; Dressler, 1980) determined the dependence of galaxy correlation on galaxy morphology. The angular correlation is defined as the joint probability $\delta P$ of finding two galaxies at an angular separation of $\delta\theta$, with respect to a random distribution (Peebles, 1980).

$$\delta P(\theta) = N^2[1 + w(\theta)]\delta\Omega_1\delta\Omega_2 \tag{3.11}$$

where $N$ is the mean surface density of the data, and $\Omega_1, \Omega_2$ are two small solid angles on the sky. If $w(\theta)$ is always zero, the sample distribution is equivalent to a random one.

As more redshift surveys were being conducted, it became practice to calculate the three-dimension correlation function, with count pairs based on distance separation instead of angular separation:

$$\delta P(s) = N^2[1 + \xi(s)]\delta V_1\delta V_2 \tag{3.12}$$

where the distance $s$ is defined as (with $c = 1$):

$$s \equiv [z_1^2 + z_2^2 - 2z_1z_2cos(\theta_{12})]^{1/2}/H_0 \ . \tag{3.13}$$

At small separation scales, it approximates to:

$$s \approx (\pi^2 + r_p^2)^{1/2} \tag{3.14}$$

where $\pi$ refers to the radial separation and $r_p$ refers to the transverse separation (Davis & Peebles, 1983), which are defined as:

$$\pi = (z_1 - z_2)/H_0 \tag{3.15}$$
$$r_p \equiv (z_1 + z_2)/H_0 \times tan(\theta_{12}/2)$$

However, the peculiar velocity of the galaxies significantly affects the radial separation measurement and thus introduces systematic uncertainty for the 3-D 2PCF measurement, which is known as the redshift distortion. As a result, it recently became popular to measure the clustering in the Anisotropic Correlation Function (ACF) $\xi(r_p, \pi)$, which splits the 2PCF measurement into two directions—radial and transverse. While this is similar to the angular correlation function, the measurement that is perpendicular to the line of sight does not suffer the uncertainty from the redshift distortion. By integrating the ACF along the line of sight ($\pi$ direction), we can get the Projected Correlation Function (PCF) $w(r_p)$, which utilizes the redshift information while minimizing the effects from redshift distortion.

$$w(r_p) = 2 \int_0^\infty \xi[(r_p^2 + \pi^2)^{1/2}]d\pi \tag{3.16}$$

The technique of PCF is very useful for our project, not only because it reduces the uncertainties due to peculiar velocities, but also because it minimizes the effect from the relatively low quality of redshifts of the faint galaxies from the COSMOS catalogue. As discussed previously, COSMOS provides very accurate photometric redshifts overall, but the accuracy of the photo-zs drops as the galaxies become fainter. This means we should not fully rely on the photo-zs of the secondary sample to estimate $\pi$ and the correlation function.

For our purpose here, we calculated the 2PCF in a slightly different way. First the projected distance is determined solely from the redshifts of the primary galaxies instead of the average of two galaxies, as the former is more accurate:

$$r_p \equiv dA|_{z=z_p} \times tan(\theta_{ps}) \tag{3.17}$$

where $dA$ here refers to angular-diameter distance that corresponds to the redshift of the primary galaxies; $\theta_{ps}$ refers to the angular separation between primary and secondary galaxies. We use angular-diameter distance to get the projected distance as it gives a more direct physical interpretation. Secondly, we only integrate a finite range along the line of sight to get the PCF:

$$w(r_p) = \int_{\pi|z=z_p-z_d}^{\pi|z=z_p+z_d} \xi[(r_p^2 + \pi^2)^{1/2}]d\pi \tag{3.18}$$

where $z_p$ is the redshift of the primary, $z_d$ is a redshift difference cut defined as:

$$z_d = 0.0022 + \sigma z_s \tag{3.19}$$

where $\sigma z_s$ is the redshift uncertainty of the secondary sample, while 0.0022 is a manual cut, which corresponds to the redshift that may be caused by the peculiar velocity as large

as $2 \times 300$ km s$^{-1}$ plus the largest separation scale we care about ($\sim 1 Mpc$). We did not extend the integral to infinity, as we know from large scale structure that:

$$\xi(r_p, \pi|_\infty) \longrightarrow 0 \tag{3.20}$$

At very large scales, pairs do not contribute much to the correlation signal and the remaining signal comes from the "two-halo" term (Cooray & Sheth, 2002). We are focusing on the scale where we can possibly find satellites and thus the larger scales are not our target but simply will increase the noise of our measurement.

## 3.2.2 Estimators

There are several estimators for calculating the correlation function:

$$\xi_n = \frac{DD}{RR} - 1 \ , \tag{3.21}$$

$$\xi_{DP} = \frac{DD}{DR} - 1 \ , \tag{3.22}$$

$$\xi_{Hew} = \frac{DD - DR}{RR} \ , \tag{3.23}$$

$$\xi_{Ham} = \frac{DD \cdot RR}{DR^2} - 1 \ , \tag{3.24}$$

$$\xi_{LS} = \frac{DD - 2DR + RR}{RR} \ . \tag{3.25}$$

The top one is the natural estimator, which is based on the definition of the correlation function (Peebles, 1980; Peebles & Hauser, 1974). The following estimators are, from top to bottom, contributed by Davis & Peebles (1983, $\xi_{DP}$), Hewett (1982, $\xi_{Hew}$) , Hamilton (1993, $\xi_{Ham}$), and (Landy & Szalay, 1993, $\xi_{LS}$). The $DD$ refers to the pair counting of galaxies within the sample, $RR$ refers to the expected pair counting for a random distribution that has the same sampling geometry and mean density (or proper normalization), $DR$ refers to the cross-pair counting between the sample and random distribution with suitable normalization.

Ideally, if the artificial random distribution perfectly matches the observation data, sharing the exact same sampling spacing and mean density everywhere, the uncertainties of all the estimators above originate from the Poisson error of bin counts (Peebles, 1980; Landy & Szalay, 1993). In practice, the major challenge of estimating the correlation function is that the observed samples are often extremely complicated. The complexity

involves the boundaries and masking of the field, the variance of the actual sample density due to the different observation conditions at different areas or large-scale cosmic variance, which makes it extremely difficult to reproduce the sampling properties of the observation data (Landy & Szalay, 1993). Inevitably, the discrepancy will introduce the calculation biases and extra variances from the "real" correlation function. The different estimators mentioned above were historically proposed to overcome and minimize these biases and variances.

Among the numerous estimators of the correlation function, the Landy & Szalay estimator is usually recognized as the most popular estimator. In theory, the LS estimator is able to minimize the error down to the Poisson level (Landy & Szalay, 1993). Its power has also been tested by simulations (Kerscher, Szapudi & Szalay, 2000), which shows that the Landy & Szalay estimator generally has the best performance for most applications, with the Hamilton estimator next to it. We also tested the different estimators by using them to calculate the correlation between two randomly generated samples that have slightly different spatial distributions. The Landy & Szalay estimator did beat the other estimators, though the difference from the Hamilton estimator is almost negligible.

As a conclusion, the Landy & Szalay estimator is chosen for our 2PCF calculations throughout our clustering study.

### 3.2.3 Cross-correlation Between Bright and Faint Galaxies

The 2PCF essentially characterizes the joint spatial distribution relation (clustering) between two groups of objects. The samples can be either a single group, in which is the result is the so-called "auto correlation function", or between two groups of objects that may have very different properties. In our case, we are particularly interested in the clustering of faint galaxies (the secondary sample) around bright galaxies (the primary sample), as they can be the potential candidates for the satellite systems.

The projected cross-correlation function between bright and faint galaxies gives $\xi_{BC}(r)$ the estimated joint probability of finding a faint galaxy in a infinitely small area $\delta A_F$, at a projected distance $r$ to one bright galaxy, which is found in another infinitely small area $\delta A_B$:

$$\delta P(r) = n_B n_C [1 + \xi_{BC}(r)] \delta A_B \delta A_F \tag{3.26}$$

where the $n_B$ and $n_F$ are the mean two-dimensional density of the bright galaxies and faint galaxies.

We use the popular Landy-Szalay estimator in the redshift space as it shows highest signal-to-noise ratio in our calculation.

$$\xi_{BF}(r) = \frac{D_B D_F - D_B R_F - D_F R_B}{R_B R_F} + 1 \qquad (3.27)$$

where $D_B D_F, D_B R_F, D_F R_B, R_B R_F$ are the normalized pair numbers at the projected redshift separation s, between four sets of data—$D_B$: the Data of Bright galaxies; $D_F$: the Data of Faint galaxies; $R_B$: a random sample distributed in the same field as $D_B$; $R_F$: a random sample distributed in the same field as $D_F$. In the case where the bright and faint galaxies share the same field, it is not necessary to produce random samples individually for the bright and faint samples and use one shared random sample, as long as the size of the random sample is significantly larger than any of the original data.

## 3.2.4   The Random Sample and Spherical Pair Counting

Building a three-dimensional random sample requires the number density of the random sample to scale with the effective volumes of the observed sample at different redshifts. A common way to meet this requirement is using a "wash out" method, which gives the random sample the same redshift distribution as the observed sample, while randomizing the RA and DEC distribution of the random sample. However, this method does not work for surveys of small sky coverage, in which case massive galaxy clusters and other large structures can lead to density peaks in the redshift distribution and artificially imprint the random sample with the same over-dense feature. As the COSMOS covers a region of only $2 deg^2$, it is not appropriate to use the "wash out" method.

Therefore, rather than building a three-dimensional random sample, we chose to build a two-dimensional random sample and use it with a "spherical pair counting" method. In this method, we construct a redshift layer for each primary, by cutting the part of the secondary sample that are far away from the primary along the line of sight, using a redshift criterion:

$$|z_s - z_p| < 0.0022 + 2\sigma_{z_s} \ , \qquad (3.28)$$

where $z_s$ refers to the photometric redshift of the secondary sample, $z_p$ refers to the redshift of the primary, $\sigma_{z_s}$ refers to the error of the redshift of the secondary sample. Note that we include the secondary sample within redshift difference of 0.0022, which corresponds to the redshift that may be caused by the peculiar velocity as large as $2 \times 300$ km s$^{-1}$ plus the largest separation scale we care about ($\sim 1 Mpc$). We then project all the secondary sample within the redshift layer and the two-dimensional random sample onto the spherical

surface where the primary lies, and count the pairs using the projected separation on this surface. Note that the number of secondary sample within the redshift layer is different for each primary, which requires an individual normalization for each primary:

$$[D_pD_s(p_i)] = \frac{D_pD_s(p_i)}{N_s(p_i)} \tag{3.29}$$

$$[D_pR(p_i)] = \frac{D_pR(p_i)}{N_R} \tag{3.30}$$

$$[D_sR(p_i)] = \frac{D_sR(p_i)}{N_s(p_i)N_R} \tag{3.31}$$

$$[RR(p_i)] = \frac{RR(p_i)}{N_RN_R} \tag{3.32}$$

where $[D_pD_s(p_i)], [D_pR(p_i)], [D_sR(p_i)]$ and $[RR(p_i)]$ are the normalized pair counts for the primary $p_i$; $D_pD_s(p_i), D_pR(p_i), D_sR(p_i)$ and $RR(p_i)$ are the pair counts for the primary $p_i$ before normalizing; $N_s$ and $N_R$ are the number of the secondary sample within the redshift layer of the primary $p_i$ and the number of random sample. Finally, the normalized count for each primary will be summed up to give the final count.

Applying the redshift cut that we mentioned above is very important for the clustering calculation; in this way we can largely remove the foreground and background, depending on the quality of the redshift for each primary. The COSMOS photo-zs are quite accurate for the relatively bright objects while the quality degrades for fainter objects. We test the significance of the improvement that the redshift cut can bring to the clustering measurement, by comparing the projected cross-correlation measurements with and without the redshift cut. The test is done using the same primary sample (full primary sample), together with a bright secondary sub-sample and a faint secondary sub-sample. From the test results that are shown in Fig. 3.1, we can see that this redshift cut can greatly improve the clustering measurement, especially for bright secondary sample, for which we have relatively accurate photo-zs.

### 3.2.5   Error Estimate

The error of the 2PCF calculation is estimated by using the so-called "bootstrap resampling technique" (Barrow, Bhavsar & Sonoda, 1984). The idea behind this technique is to randomly generate a set of re-samples from the original sample, and rerun the calculation with the re-samples. The averaged variance between the results of those recalculations is used as the error of the original calculation.
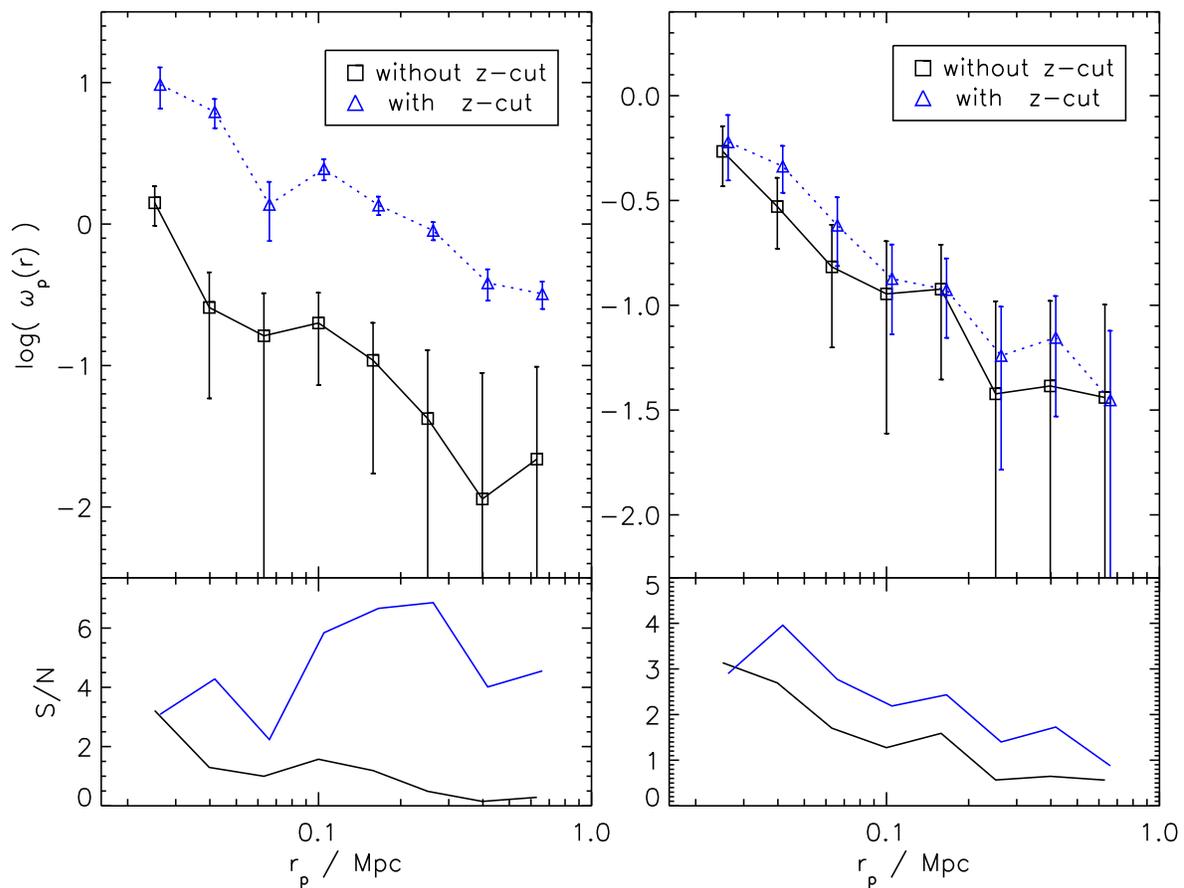
Figure 3.1: The power of applying redshift cut to the 2PCF calculation. Left side is the correlation between the full primary sample and the secondary sample within $-19 < M_g < -17$, the right side is the correlation between the full primary sample and the secondary sample within $-17 < M_g < -15$. Both plots show the contrast between using and not using the redshift cuts

35

In practice, we produce 100 randomly generated bootstrap arrays for our primary sample before the calculation, of which each element indicates how many times a primary should be used in a bootstrap sample. Specifically, each element has an value of $1 - N_p$, whereas the sum of the array should be equal to $N_p$, where $N_p$ is the number size of the primary sample. Thus each array defines a bootstrap sample. In the next step, we apply the bootstrap arrays and weighting parameters to the counting process and produced 100 sets of differently weighted counting results through one single calculation process. We then use the randomly weighted counting results to generate 100 different 2PCF results using the same estimator. Finally, we use the $rms$ of those results as the error estimate for our 2PCF calculation.

Our error estimate was also tested by comparing it to the results of several artificially half-cut sub-samples and by applying our code to the correlation calculation between two randomly generated samples. In the former case, the error fitted the variance among the calculation results of those sub-samples. In the latter case, the code returned a clustering signal within about $1\sigma$ of zero in all cases.

## 3.2.6  Results and Discussion

We tested several combinations of primary and secondary sample for the projected two-point cross-correlation function measurement. We discussed the dependence the measurements on several properties of the primary and secondary samples below.

### Dependence on the Secondary Luminosity

We divided our secondary sample into three sub-samples with the g-band absolute magnitude and use them to calculate the correlation with the full primary sample. The three sub-samples are:

1. Secondary sample with g-band absolute magnitude $-19.0 < M_g < -17.0$; (2,395 galaxies)

2. Secondary sample with g-band absolute magnitude $-17.0 < M_g < -15.0$; (8,271 galaxies)

3. Secondary sample with g-band absolute magnitude $M_g > -15.0$ and g-band apparent magnitude $g < 26.0$. (9,014 galaxies)

The result is shown in Fig. 3.2. From the plots, we can see the correlation functions have a strong dependence on the magnitude of the secondary sample. The brighter secondary sample has much strong correlation signal than the faint samples, as well as a better signal-to-noise ratio.

## Dependence on the Primary luminosity

We split our primary into two sub-samples by the g-band absolute magnitude: one relatively bright sub-sample of 120 galaxies with $M_g \leq -20.5$ and one relatively faint sample with 272 galaxies with $M_g > -20.5$ .

The result is shown in Fig. 3.3. From the plots we can see the correlation strongly depends on the luminosity of the primary sample. The correlation functions of the full primary sample, the relatively bright sample and the relatively faint sample share similar slope. The bright sample has a significantly higher signal detection than the faint sample and the overall average. The bright sample also has almost the same signal-to-noise ratio as the full primary sample, despite the much smaller sample size.

## Dependence on the Primary SED Type

The shape catalogue has a column called "MODD", which lists the spectral template that best fits the photometry of each galaxy. We can use this information to classify and divided our primary sample into two sub-samples: 63 Elliptical/S0 galaxies and 392 Spiral galaxies.

The result is shown in Fig. 3.4, which shows that the SED type of the primries has some effects on the correlation measurement. The elliptical galaxies have relatively high signal detection and good signal-to-noise ratio, considering the extremely small sample size.

This result may be related to the dependence of the clustering signal on the primary luminosity, as the elliptical galaxies have a higher luminosity on average; it may also be a result of that fact that ellipitical primaries are more likely to locate in groups and clusters; or it may indicate some essitential difference of the galaxy formation between these two types of galaxies.

## Dependence on the Primary Colour

We looked into the colour of our primary sample (Fig. 2.10), and split our primary sample into two sub-samples with the criterion of $B - V \sim 0.48$. We got a red type sample with 197 galaxies and a blue type sample with 195 galaxies.
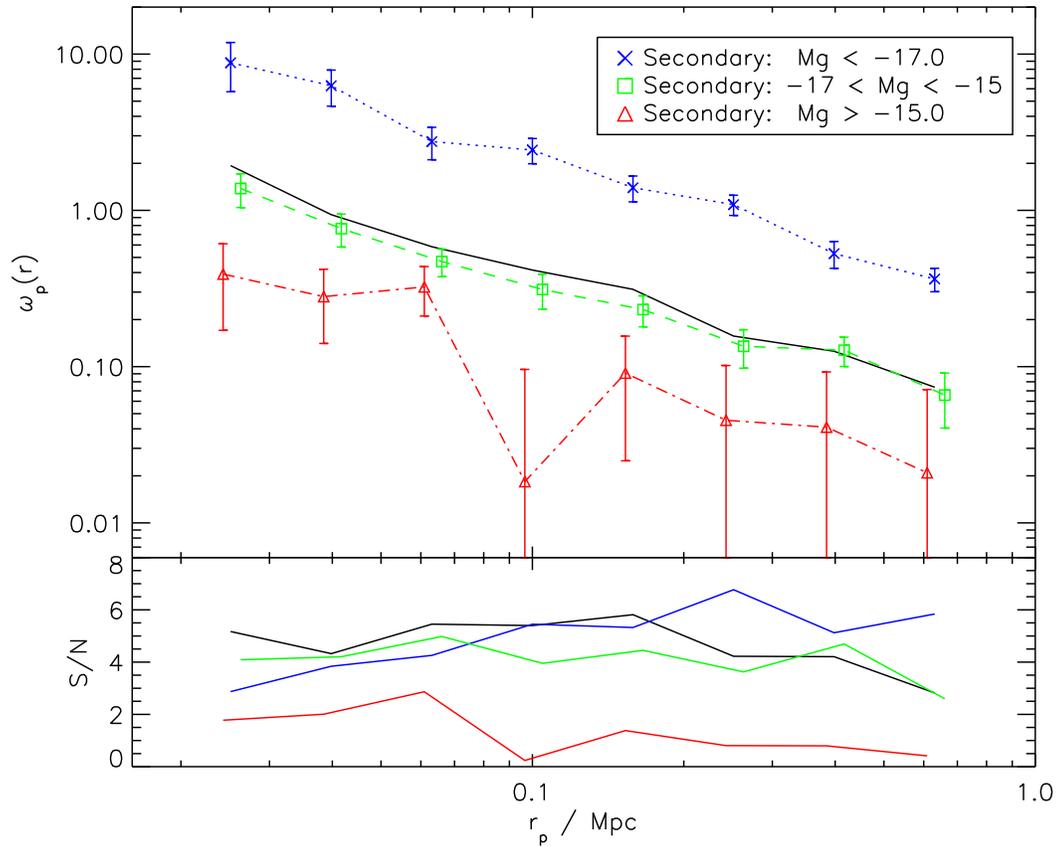
Figure 3.2: The effect of secondary brightness. The blue, green and red lines show the projected cross-correlation function between the full primary sample and three secondary sub-samples that cut by brightness, with a comparison to the measurement using the full secondary sample (black line).
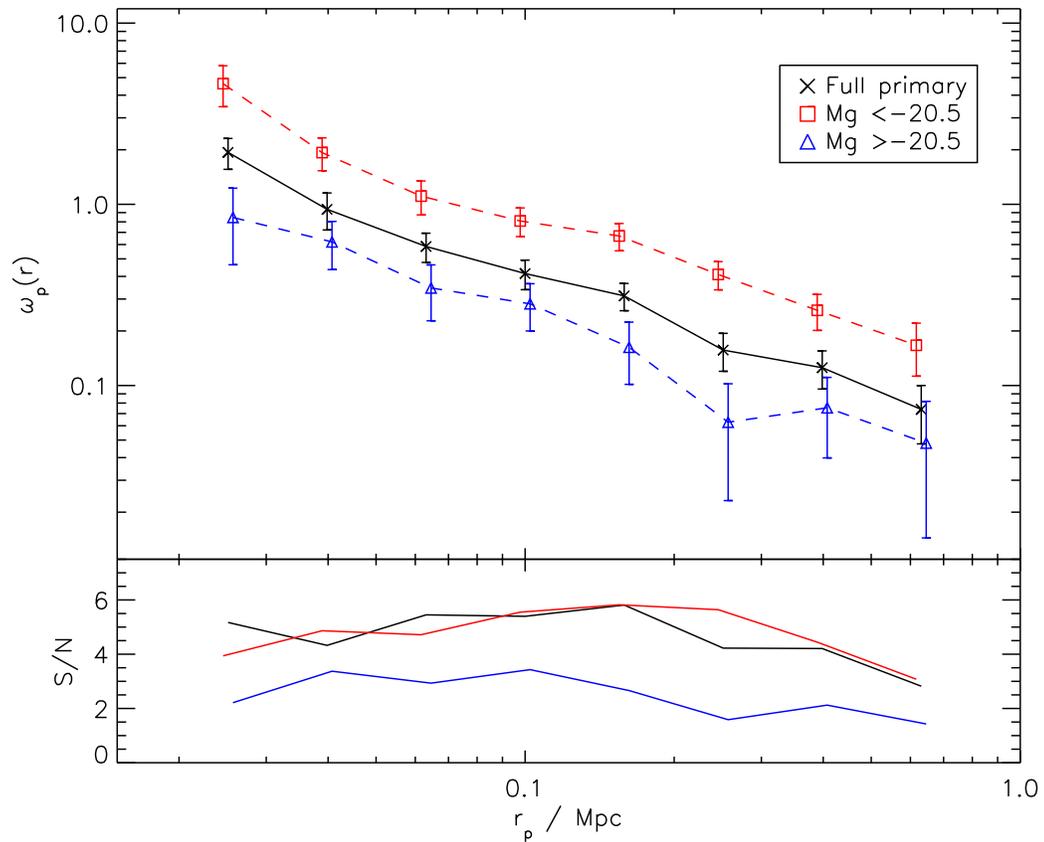
Figure 3.3: The effect of primary brightness. The figure shows the projected cross-correlation function for two primary sub-samples (Bright $M_g < -20.5$ and Faint $-20.5 < M_g < -19.0$ and a secondary sub-sample $M_g < -15.0$
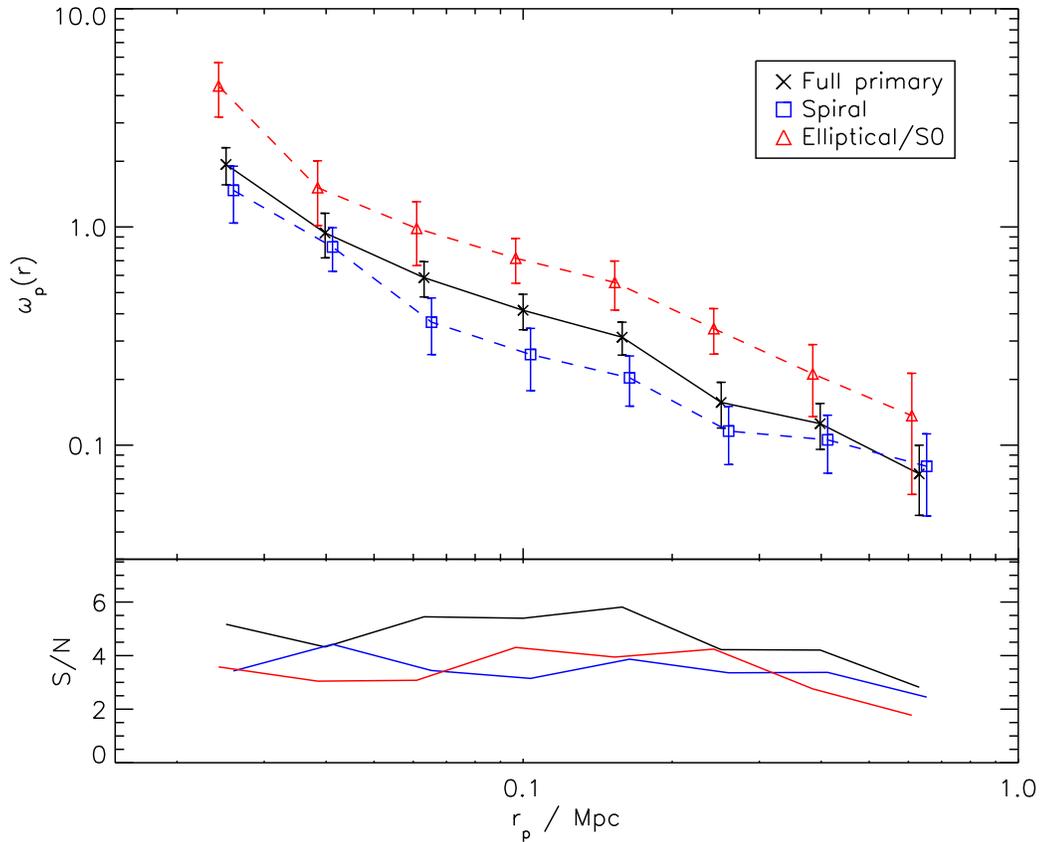
Figure 3.4: The effects of primary SED type. The figure shows he cross-projected correlation function for two primary sub-samples (Ellipticals and Spirals) and a secondary sub-sample $M_g < -15.0$

When we split the primaries into the Red and Blue samples discussed previously, the result is shown in Fig. 3.5. The plot shows that the clustering depends on the colour of the primary, although the dependence is not that strong as on the primary luminosity or the primary SED type. Overall, the red primary sub-sample shows higher correlation detection and better signal-to-noise ratio than the blue primary sub-sample.
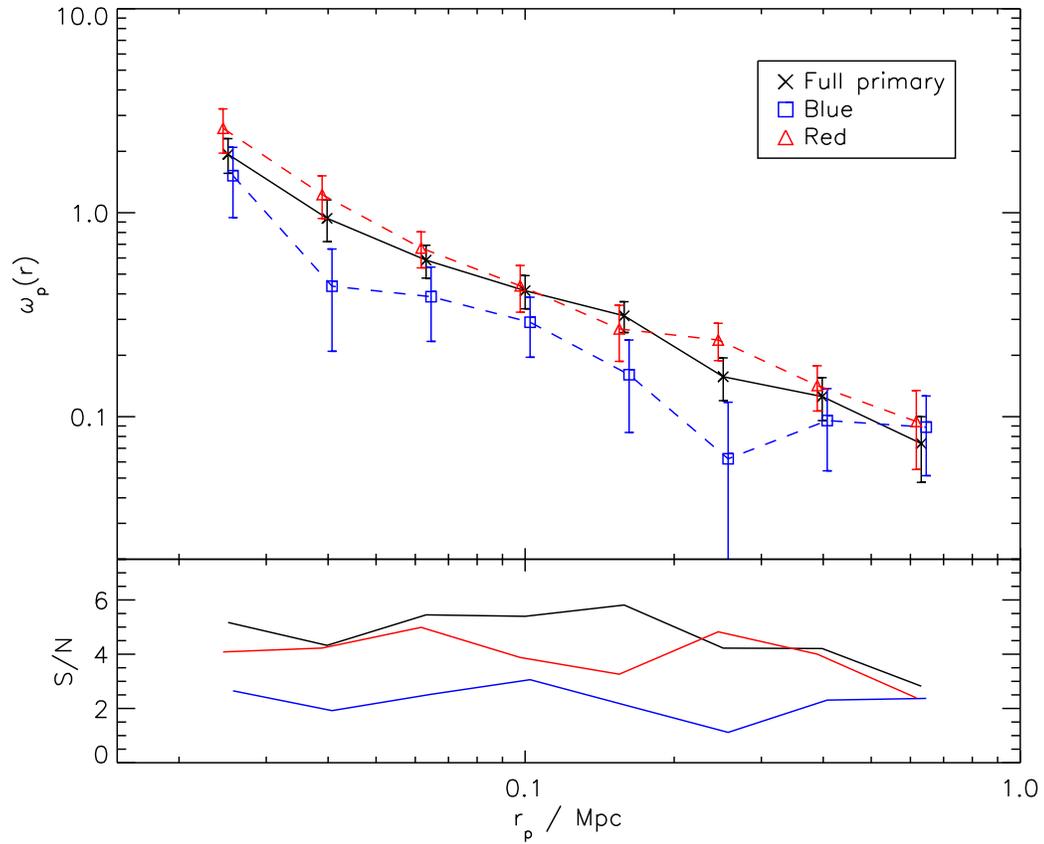
Figure 3.5: The effect of primary colour on the clustering signal. The figure shows the cross-projected correlation function for two primary sub-samples (Red and Blue) and a secondary sub-sample $M_g < -15.0$

# Chapter 4

# Satellite Spatial Distribution

This chapter summarizes work on two aspects of the spatial distribution of the satellites — the radial distribution and the angular distribution (galaxy alignment). For the radial distribution, we compare two different methods of background subtraction and conclude that the locally averaging method works better with the COSMOS data. We then discuss and determine the proper distance range that we should use for satellite counting. Finally, we look into the dependence of the radial distribution on the properties of the primary. For the angular distribution, we study the alignment of the position angle of the satellites with the position angle of the projected major axis of the primaries.

## 4.1   Radial Profile

Given some projected distance range from the central galaxies, the galaxies inside the range consist of the actual satellites and background galaxies. Since the satellites are relatively faint and thus usually have uncertain redshift information, it is difficult to directly distinguish the satellites from the background[1]. The key to achieve good measurements of the satellite abundance is to subtract the background population statistically.

Here we apply two methods to subtract the background and calculate the satellite radial density profile. The essential difference between these two methods is whether the background is defined as a global average over the whole primary sample, or whether it is determined for each galaxy individually.

---

[1]By "background", we mean both background galaxies and foreground galaxies and we do not distinguish them in this thesis, because neither of them are real satellites, and both need to be subtracted from the satellite counts.

### 4.1.1 Radial Profile using a Global Average Background

It has been a common practice to use the global mean of the secondary sample to estimate the background (Wang et al., 2011; Wang & White, 2012; Wang et al., 2014; Jiang, Jing & Li, 2012). It has been argued that the distribution of background galaxies should be uncorrelated, and thus in effect randomly distributed, with respect to the central galaxies (Jiang, Jing & Li, 2012). We can use a random sample that contains a large number of points to fill in the volume that has exactly the same masked areas and boundaries as the observations. The distribution of this random sample should represent accurately the uncorrelated background, once we scale it down to the size of the observed secondary sample. We can then use the mean density of the random sample averaged over the effective observation volume to estimate the background.

However, with our clustering measurement, we can clearly see the correlation signal is non-zero even at very large projected distances (close to 1 Mpc), where there should be no genuine satellites. This implies that the local background of faint galaxies is slightly higher than the global mean density. The higher local background is actually an expected phenomenon, contributed by the large-scale correlation that is also known as the "two-halo term" correlation (Cooray & Sheth, 2002). If we can measure the extra amount contributed by the two-halo term, in principle it should be able to give better estimate of the real local background.

From our discussion of the 2PCF calculation in the previous chapter, we expect the correlation signal to be strong at small separations, and drops with projected distance. Specifically, from the linear-scale plot of the correlation function (Fig. 4.1), we can tell the correlation signal flattens out when the projected distance reaches $r_p > 700$ kpc, at which scale the satellite contribution is negligible and the large scale structure dominates (Speller & Taylor, 2014; Wang & White, 2012). If we assume the signal of the two-halo term is approximately flat within several virial radii from the central galaxies and the total correlation signal only comes from the one-halo term and the two-halo term, we can use the correlation signal at large projected distance as the mean two-halo term correlation, and then subtract it from the total correlation signal to get the one-halo term signal.

We define the averaged background correlation signal as:

$$\hat{\omega}_{bg} = \frac{\int_{r_2}^{r_3} \omega(r_p) dr_p}{r_3 - r_2}. \tag{4.1}$$

If we assume the background does not vary with radius around the primary, which is a proper approximation at relatively small scales ($r_p < 1.0 Mpc$), we can subtract the
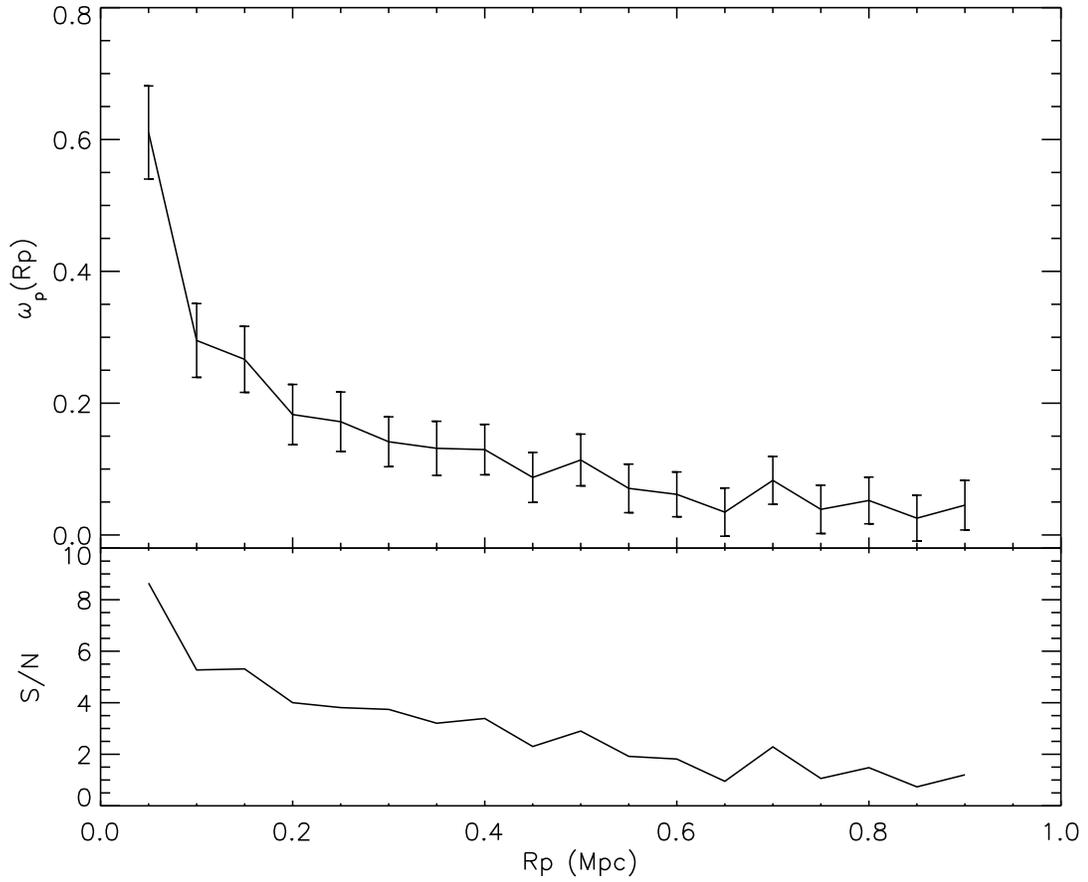
Figure 4.1: Correlation calculation with the full primary sample with the full secondary sample, plotted with linear scales. Top: The projected two-point correlation function, calculated using the full primary sample and the full secondary sample, and plotted with linear scales. Bottom: Signal-to-noise ratio for the same.

contribution from the background and estimate the satellite number as a function of the radius:

$$n(r_p) = (\omega(r_p) - \hat{\omega}_{bg}) D_p R(r_p) \frac{N_p N_s}{N_p N_R} \tag{4.2}$$

where the $D_p R(r_p)$ is the non-normalized primary-random pair profile, and the $\frac{N_p N_s}{N_p N_R}$ normalizes the profile to the size of the primary-secondary pair sample.

The error of this estimator mainly comes from the error of the correlation $\omega(r_p)$, while the error from the $D_p R(r_p)$ is negligible as the size of the random sample is big enough to minimize the Poisson uncertainty.

$$\sigma_n = \sqrt{\sigma_\omega^2 + \sigma_{\omega_{bg}}^2} D_p R(r_p) \frac{N_p N_s}{N_p N_R} \tag{4.3}$$

A result of using the projected 2PCF with the full primary and the full secondary sample is shown in Fig. 4.2.

## 4.1.2 Radial Profile using Local Background

First of all, we can remove a significant amount of background by making use of the accurate photo-zs available of our dataset. Around each primary, we exclude any galaxy that has difference of line-of-sight velocity larger than $2 \times 300$ km s$^{-1}$ by $2\sigma$ confidence ($|\Delta V| < 300 km/s + 2\sigma_V$). This cuts out most of the background and foreground galaxies and constructs a layer around each primary that consists of the galaxies that are at roughly the same redshift. We then assume satellites are only distributed in the regions that are close to the central galaxies, such that:

$$n_{sat}(r) = 0 \quad (\text{when } r > r_2) \tag{4.4}$$

Then we assume of density of background is a constant and we use the galaxy counts at the outer region that is relatively far away from the central galaxies to estimate the density of the background:

$$n_{bg} = n_{tot}|_{r>r_2} = \frac{N_i^{tot}|_{r>r_2}}{A|_{r_2}^{r_3}} \tag{4.5}$$

where $N_i^{tot}$ and $n_{tot}$ refer to the total counts and the density of the total counts of all the galaxies within the radial bin around the primary $i$; $A|_{r_2}^{r_3}$ refers to the effective area within projected distance range of $r_2 - r_3$.
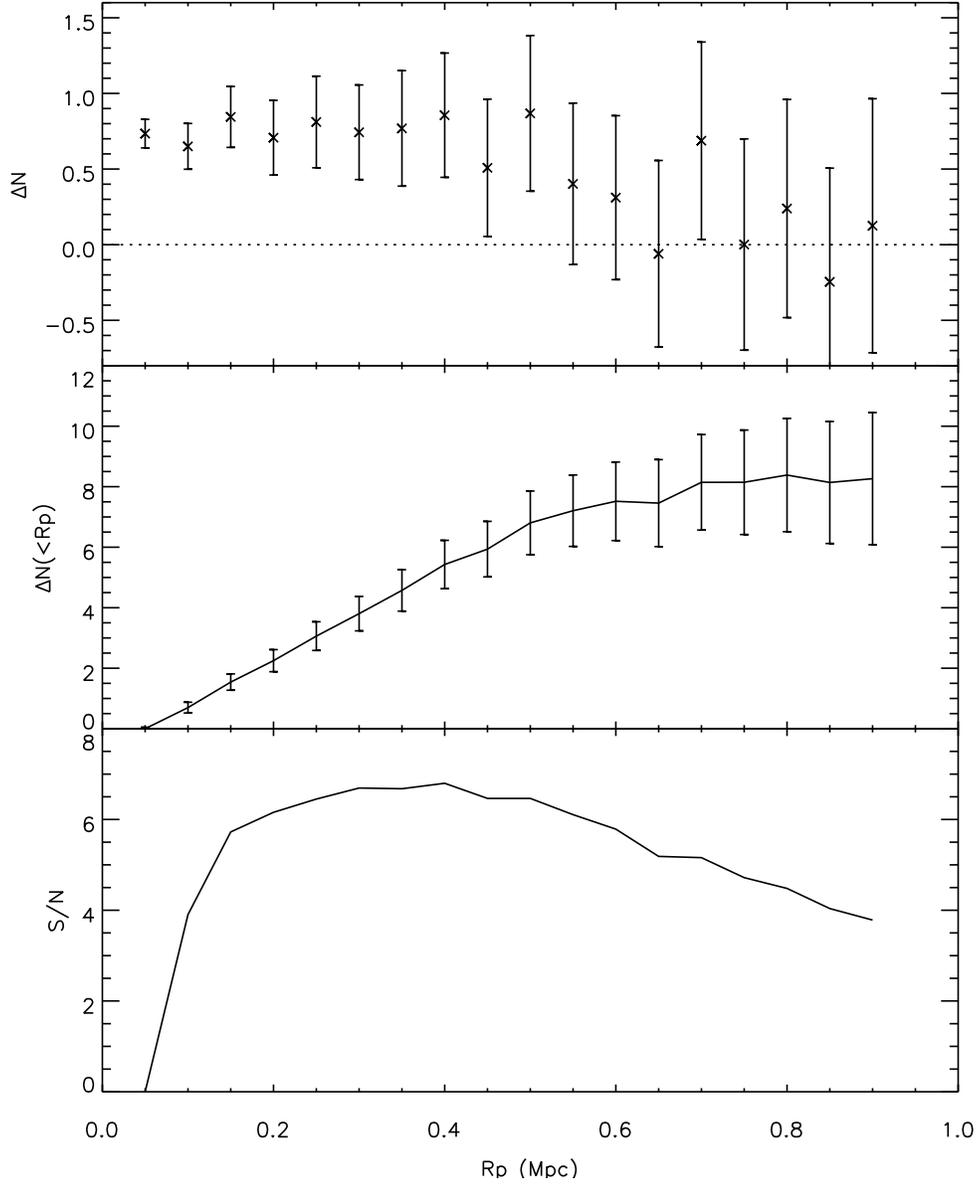
Figure 4.2: The satellite radial density profile, derived from the correlation function. The result here is converted from the projected 2PCF calculated using the full primary and secondary samples. The top panel shows the satellite profile in radial bins of 50 kpc, ranging from 25 kpc to 925 kpc. The second panel shows the cumulative counts, excluding the first bin. The bottom panel shows the signal-to-noise ratio of the cumulative counts.

For Primary galaxy $i$, the satellite number within distance range $r_1 - r_2$ can be given by subtracting the background within the same range:

$$
\begin{aligned}
N_i^{sat}|_{r_1}^{r_2} &= (N_i^{tot} - N_i^{bg})|_{r_1}^{r_2} \\
&= N_i^{tot}|_{r_1}^{r_2} - A|_{r_1}^{r_2} \, n_{bg} \\
&= N_i^{tot}|_{r_1}^{r_2} - \frac{A|_{r_1}^{r_2}}{A|_{r_2}^{r_3}} N_i^{bg}|_{r_2}^{r_3}
\end{aligned}
\tag{4.6}
$$

Due to the masking and other boundary effects, the areas around the primaries are usually not complete, which leads to the miscounting for satellites. To correct this effect, we need to know the area completeness of each radial bin for each primary. We define:

$$
\gamma_i(r, r + dr) = \frac{A'_i(r, r + dr)}{A_i(r, r + dr)}
\tag{4.7}
$$

where $A'_i(r, r + dr)$ refers to the area of the radii range $(r, r + dr)$ without any masking or boundary conditions. The corrected satellite number can then be expressed as:

$$
N_i^{sat'}|_{r_1}^{r_2} = \int_{r_1}^{r_2} \gamma_i(r, r + dr) \, n_i^{sat}(r, r + dr) dr
\tag{4.8}
$$

To get all the areas $(A, A')$ around each primary, we randomly generate a sample of 300,000 objects to fill an area that covers the whole ACS field. The counts of the random sample can provide the complete area without any masking and boundary effects for each primary. Then we apply the masking we obtained for the COSMOS shape catalogue (see Section 2.5) to this random sample. The counts of the random sample can provide an estimate of the effective area observed around each primary. Technically, we conduct the counting process in radial bins of 50 kpc. We use the radial range of $750 \sim 1000$ kpc as outer range for background estimate, and the radial range of $50 \sim 600$ kpc as the satellite counting range. The detailed reason for choosing these ranges are described in Section 4.1.3.

We applied this method to produce the satellite radial density profile plot (Fig. 4.3) with the same primary and secondary samples that are used in producing Fig. 4.2. These include the 489 galaxies from the full primary sample and 17,143 galaxies from the full secondary sample. The error bars are estimated from Poisson uncertainty of the secondary counts.

Comparing Fig. 4.2 and Fig. 4.3, we can see the two methods produce very similar results with the same signal level and profile shape, while the result from local background

48
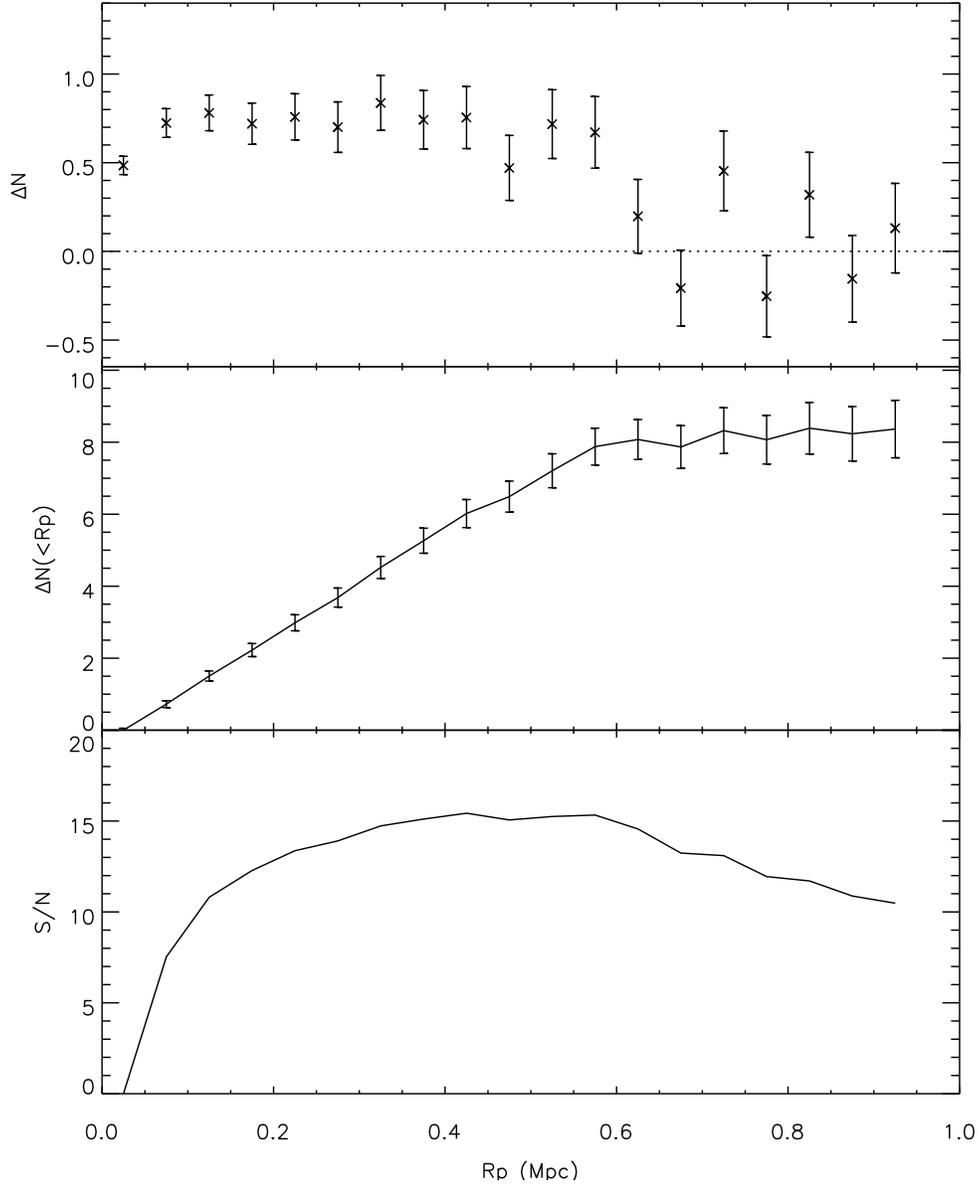
Figure 4.3: The satellite radial number density profile, using a local background subtraction. The top panel shows the satellite profile in radial bins of 50 kpc, ranges from 0 kpc to 950 kpc. The second panel shows the cumulative counts, which excludes the first bin to avoid possible contaminations (see Section 4.1.3 for detail of distance range choice). The bottom panel shows the signal-to-noise ratio of the cumulative counts.

49

subtraction has better signal-to-noise ratio. The reason the globally estimated background has lower signal-to-noise, may be because of cosmic variance between the background regions behind individual primaries. This variance can be the result of the essential variance of the halo mass, the variance of the two-halo term correlation (cosmic variance), and perhaps more importantly, the difference of effective observation volumes between individual primaries. Our galaxy samples are distributed in the redshift range of $0.02 < z < 0.2$, which has 5 magnitude difference in distance modulus between the lowest redshift and highest redshift. Therefore, individual primaries can potentially have very different effective observation volume and different background counts because of that. Overall, we concluded that using the local background subtraction is a better method for our purposes, while using an average background subtraction provides verification of these results, albeit at slightly lower signal-to-noise.

### 4.1.3   Distance Dependence

The typical physical scale of galaxies is on the order of 10 kpc, while the typical scale of galaxy clusters on the order of 1 Mpc. Our primary sample has an average g-band absolute magnitude of $\bar{M}_g = -20.1$, which is fainter than the MW. We expect the virial radius of our primaries should be less than 500 kpc except for a few of the most massive systems, and thus this is the radial range within which we expect that most satellites will be found. This is also confirmed by Fig. 4.2 and Fig. 4.3, where the satellite number that is contributed by the outskirts ($r > 0.6$ Mpc) is much less significant than the inner region. Therefore, we set the upper limit of distance of satellite counting to 600 kpc.

In addition, we should exclude pairs within the innermost radial bins for two reasons: First of all, there are faint detections as a result of the image splitting of the bright extended galaxies. The catalogues we use are initially selected by SExtractor (Leauthaud et al., 2007; Ilbert et al., 2009; Lilly et al., 2007; Capak et al., 2007), which is an image analysis program (Bertin & Arnouts, 1996). A single bright extend source can sometimes be identified as multiple ones when the source has an irregular shape or when parts of it, like galactic arms, appear to be separated from the source. Despite of the excellent work of PSF modelling and masking by Leauthaud et al. (2007), a part of the secondary samples in the shape catalogue that are very close to the primary can still actually be contributed by the multiple detections of the central galaxies (some examples can be seen in Fig. 4.6). The second concern is the optical contamination. Areas close to bright source can be potentially contaminated by bright globular clusters, HII regions that are associated with the central galaxies (Speller & Taylor, 2014) and the halos and scattered light of the central bright source that are not completely masked out. The contamination can make it difficult

to resolve the images of close pairs in some cases. In general, it can add extra systematic errors to the photometry measurement (Wang & White, 2012) (also see Fig. 4.7), and further cause extra uncertainty for the photo-z measurement, as the contaminations come from sources that may have different redshifts.

For these reasons above, we decided not to use the close pairs. To establish where to place the cut on separation, we examined the images of 30 close galaxy pairs and evaluated the detection by analysing the images. We selected 500 close pairs, which all have a very faint companion ($g^+ > 23.0$). Then we searched the images of 32 bright-faint galaxy pairs within the separation range of $r_p < 22$ kpc and 24 pairs within the range of $r_p \sim 22 - 60$ kpc. We evaluated the resolvability of those pairs in the images and classified them into three categories —"clearly separated", "questionable" and "ambiguous". We show some examples of images of questionable pair detection in Fig. 4.5 and some really bad detections ("ambiguous") in Fig. 4.6, in contrast to the images of the clearly separated pairs shown in Fig. 4.4. Finally we plotted our manual classification of those pairs with the distribution of their projected physical separations and the photo-z error of the faint companions, which is shown in Fig. 4.7.

From the plot, we can see that the questionable cases are mostly distributed within 32 kpc from the central galaxy. The unresolvable detections are distributed within a separation of 20 kpc. In addition, some faint galaxies within the small separation have significantly larger errors in their photometry. This implies that at small physical separation, the risk of having fake faint galaxy detection does exist. Fortunately, the risk is significantly lower when the separation is larger than about 30 kpc. We concluded that a lower separation limit of 30 kpc for pair counting can remove most of the problematic and questionable pair detections and we decided to set the limit to be 50 kpc, as a relatively safe choice. Finally, the projected separation range for pair counting is set to 50 – 600 kpc.

## 4.1.4  Dependence on Central Galaxy Properties

In this section, we will test the satellite radial distribution's dependence on the primary properties by dividing the primary into different sub-samples and comparing the corresponding results. Specifically, we will test the effects from the primary luminosity, the primary morphology (SED type) and the colour of the primary sample.
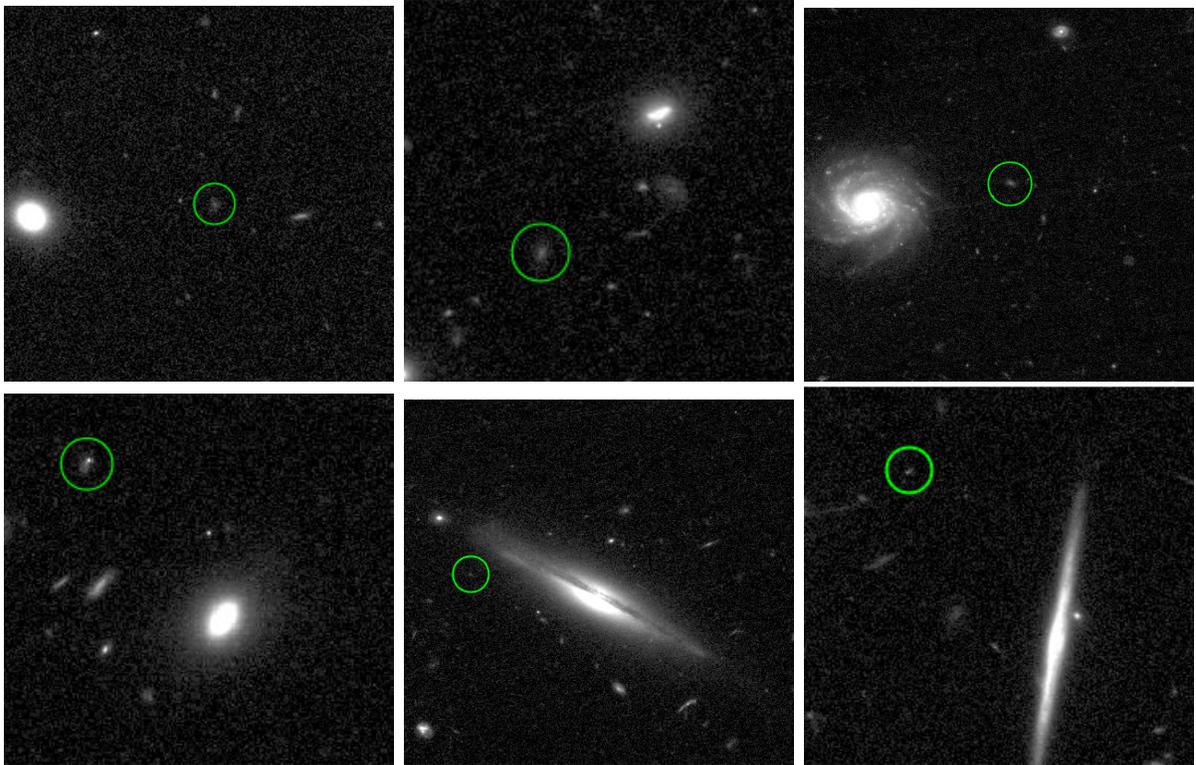
Figure 4.4: Examples of clearly separated bright-faint galaxy pairs. The faint galaxies of these pairs are circled in green. The images are obtained with image search tool "COSMOS SkyWalker" (http://www.mpia-hd.mpg.de/COSMOS/skywalker/#), using the coordinate information from the COSMOS shape catalogue.
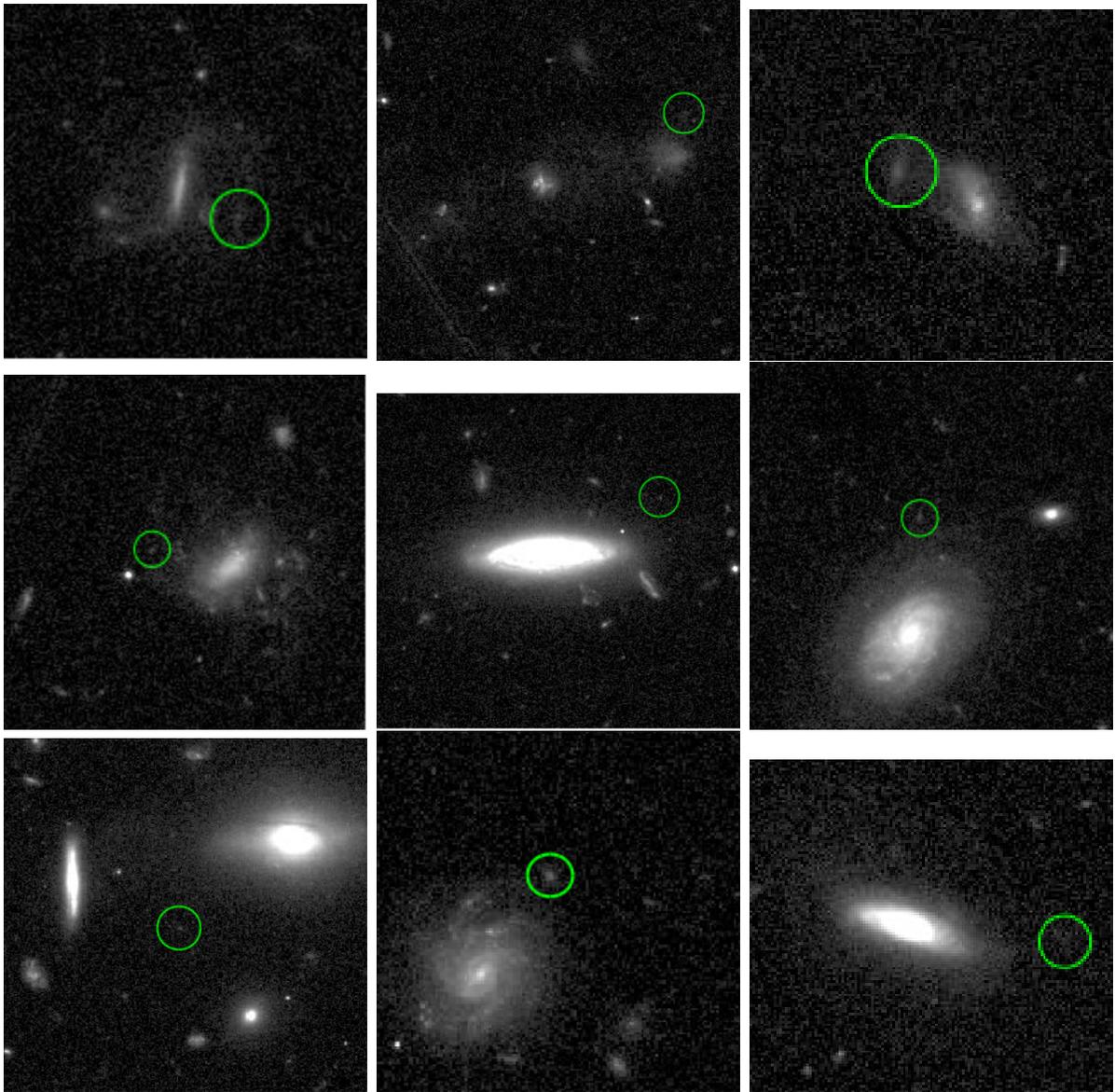
Figure 4.5: Examples of bright-faint galaxy pairs that the detection of the faint companions are questionable. The faint galaxies of these pairs are circled in green. The images are obtained with image search tool "COSMOS SkyWalker" (http://www.mpia-hd.mpg.de/COSMOS/skywalker/#), using the coordinate information from the COSMOS-shape catalogue.
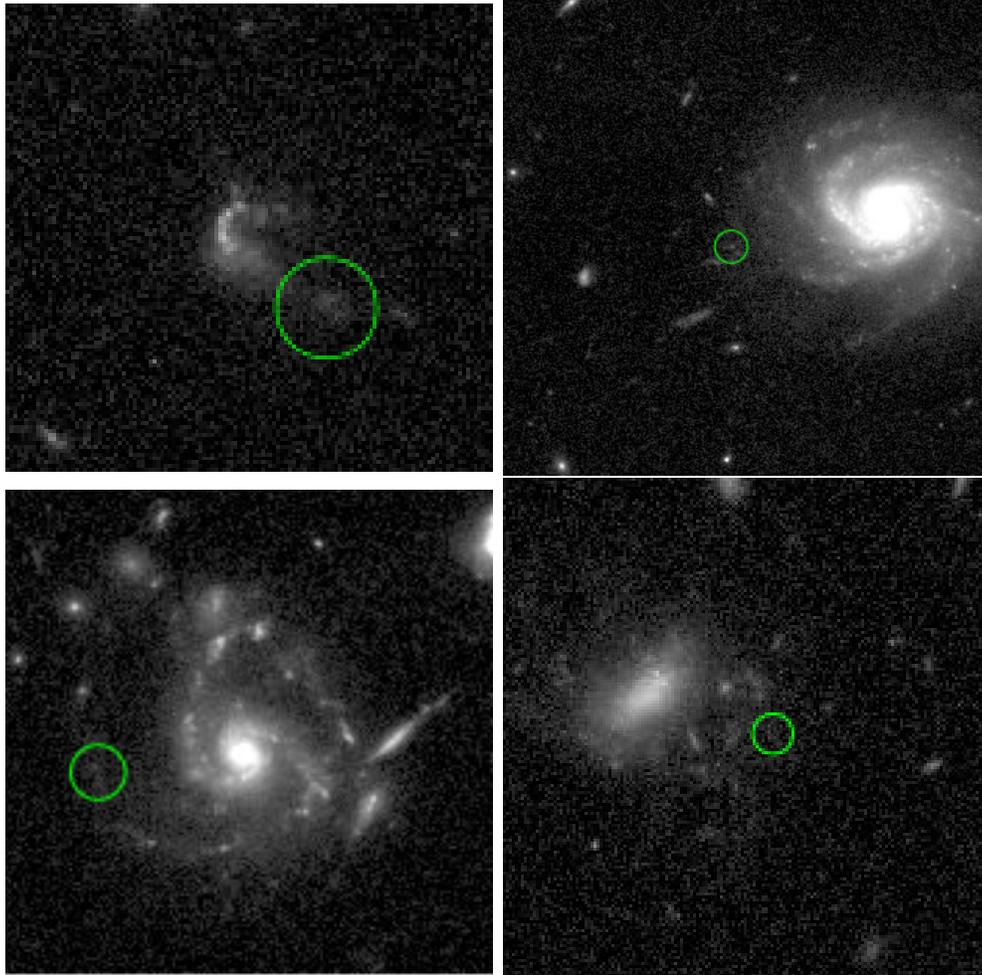
Figure 4.6: Examples of bright-faint galaxy pairs that can be barely resolved as separated objects. The faint galaxies of these pairs are circled in green. The images are obtained with image search tool "COSMOS SkyWalker" (http://www.mpia-hd.mpg.de/COSMOS/skywalker/#), using the coordinate information from the COSMOS shape catalogue.
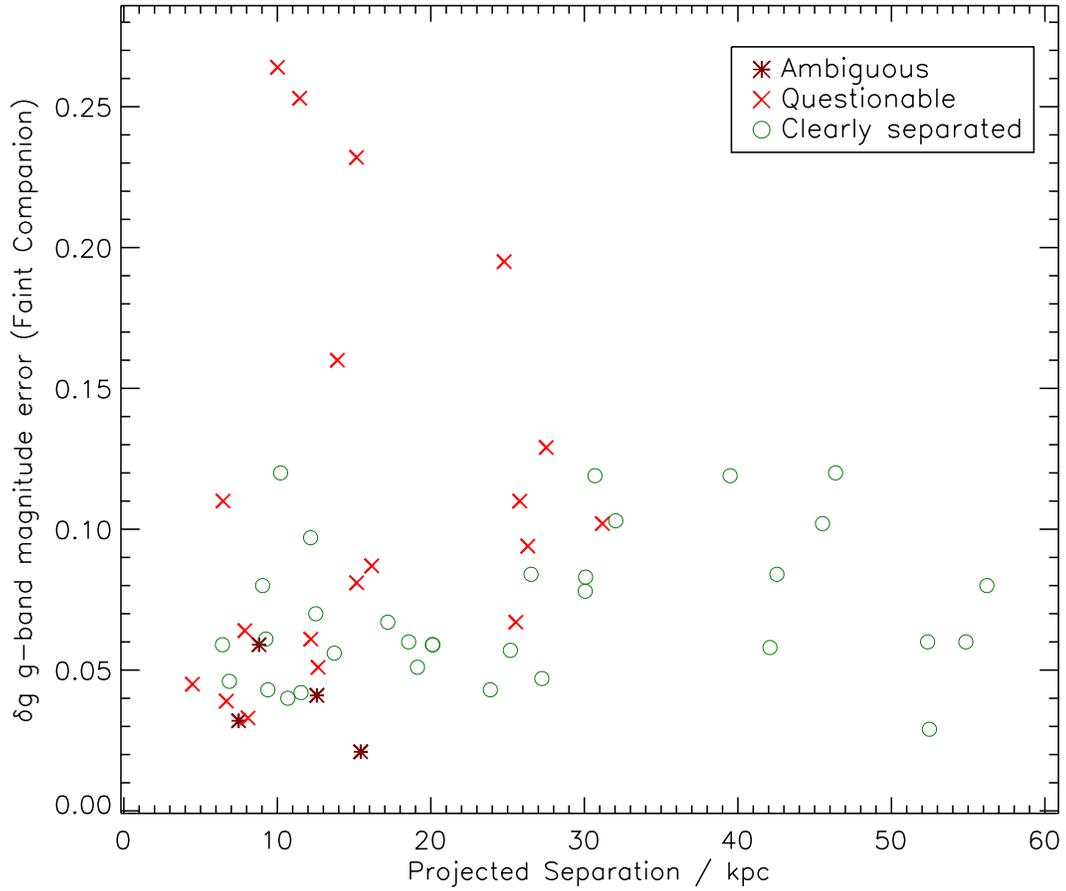
Figure 4.7: The results of bright-faint galaxy pair image checking. The color corresponds to our evaluation of how well the pairs are resolved as separted objects. There are four red points for ambiguous pairs; 19 blue points for questionable pair detections and 33 green points for clearly separated pairs. And the plot shows how the resolvability is distributed with projected separation and the error of g-band magnitude.

## Dependence on Primary Luminosity

We use the g-band magnitude $M_g = -20.5$ as the dividing line and split the kick-out primary sample into two sub-samples — a bright sub-sample with 120 galaxies and a faint sub-sample with 272 galaxies. We then calculated the excess number counts $\Delta N$ for the two sub-samples separately, using the full secondary sample for both calculations. The results are shown in Fig. 4.8, along with the result of the full kick-out primary sample for comparison.

We find that the bright primaries have more satellites than faint primaries overall. This general trend is expected if we assume that bright galaxies are located in more massive halos on average. Another interesting result is that the number profile of faint primaries seems drop more quickly with radius. As we can see in the plot, the faint primaries seem have significantly less satellites per bin at the range of $250 \sim 500$ kpc than the inner regions. In contrast, the bright primaries have a quite flat radial profile throughout the range of $0 \sim 500$ kpc. This arguably implies satellites of faint primaries are less widely distributed than satellites of bright primaries. The more extended distribution of the bright primaries may indicate that they are located in halos with a larger virial radius (more massive) than the faint primaries.

## Dependence on Primary SED Type

To see the dependence of the primary SED type, we used the catalogue parameter "MODD", which indicates the best fit spectrum template for the 30-band photometry, to separate the primaries into early and late sub-samples, as in section 2.8. We used the primaries with $1 \leq MODD \leq 9$, which corresponds to galaxy type E1 to S0, as our "Elliptical/S0" sub-sample, including 62 galaxies, and use the primaries with $MODD > 9$, which corresponds to galaxy type S1 to SB11 (star burst), as the "Spiral" sub-sample, including 230 galaxies. For both SED type sub-samples, we used the full secondary sample to calculate their satellite radial density profile and compared to the profile of the full sample, shown in Fig. 4.9 .

From the plot, we can see that the satellite radial distribution has a strong dependence on the primary SED type. The Elliptical/S0 type primaries have much higher satellite radial number density than the Spiral type primaries. The satellite distribution of the Elliptical/S0 type primaries also extends to larger radii (signal-to-noise ratio peaks at $\sim 600$ kpc) than the Spiral type primaries (signal-to-noise ratio peaks at $\sim 200$ kpc). As in the previous figure, the more extended distribution of the elliptical primaries may
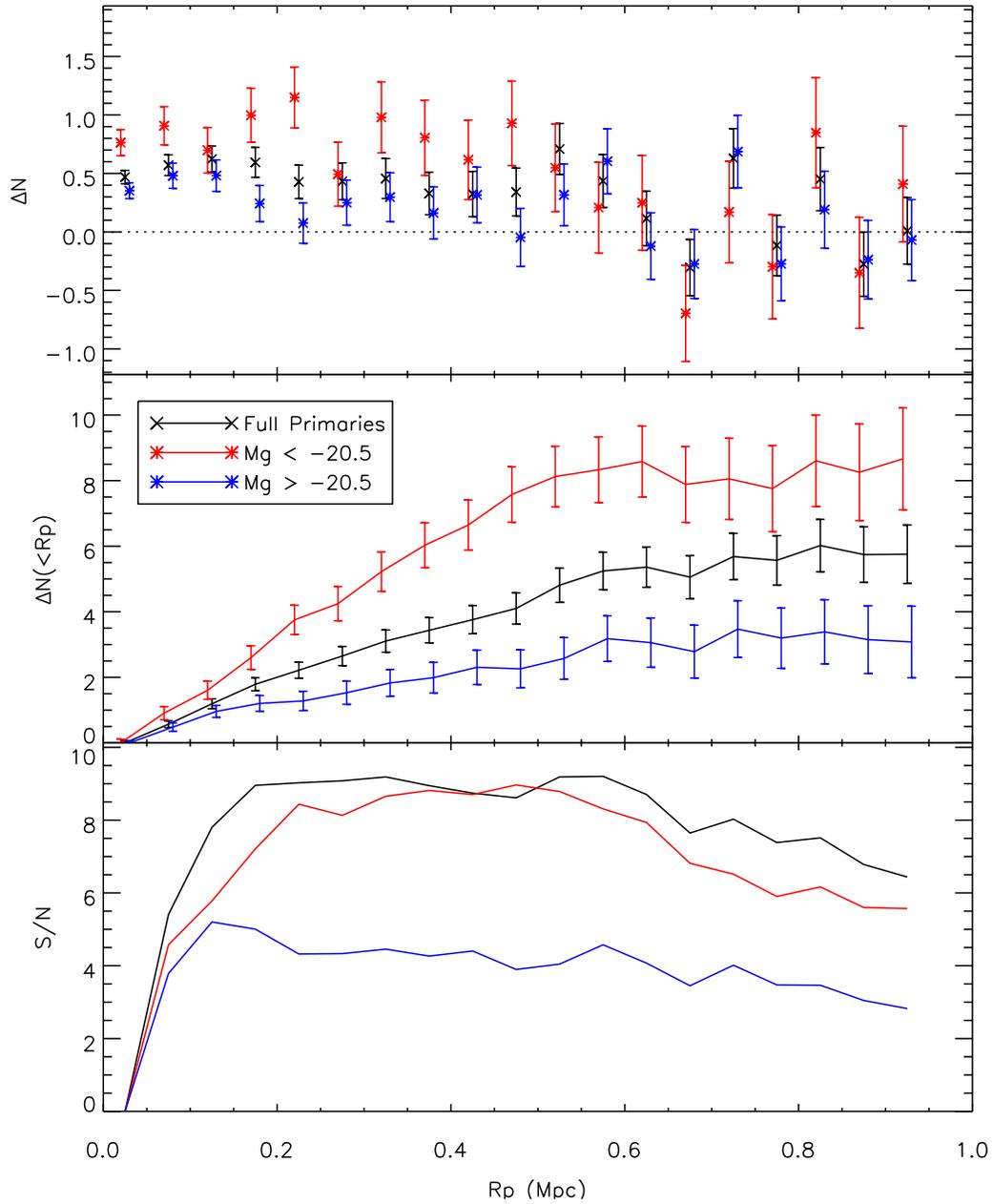
Figure 4.8: The satellite radial distribution of faint (blue) and bright (red) primary galaxies, with a comparison of the full "Kick-out" primaries (black).
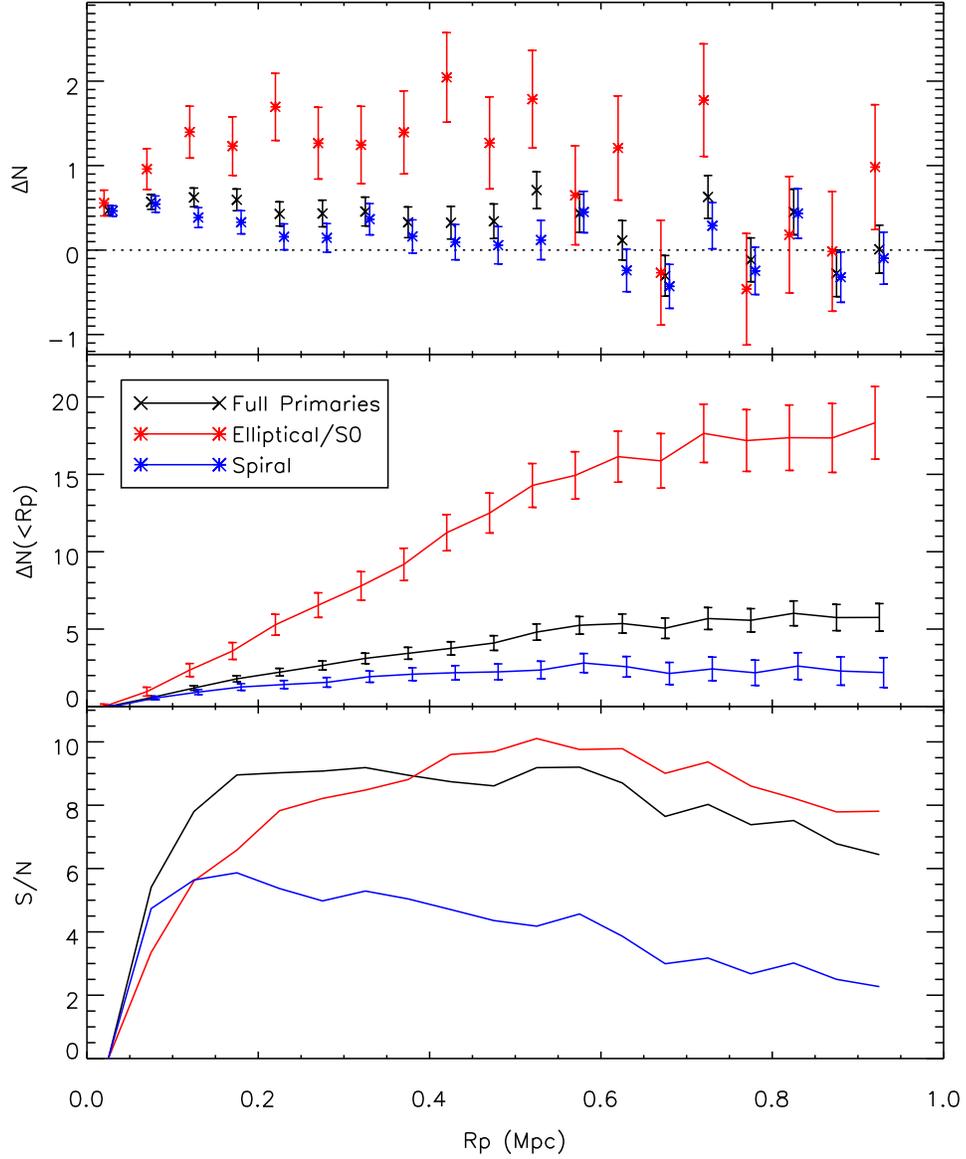
Figure 4.9: The radial distribution of satellites around Elliptical/S0 (red) type and Spiral type (blue) primary galaxies, with a comparison of the full "Kick-out" primaries (black).

indicate that they are located in halos with larger virial radius (more massive) or more likely to be located in more massive halo groups than the spiral primaries.

**Dependence on Primary Colour**

We used a color criterion to divide the primary sample into a Red ($B - V \leq 0.48$) sub-sample and a Blue ($B - V > 0.48$) sub-sample. As we do not have B-band information for the SDSS-based galaxies, we used a slightly different color criterion for them, taking $g - r > 0.4$ to be Red, and $g - r \leq 0.4$ to be Blue. After the cuts, we got 197 galaxies in the Red sub-sample and 195 galaxies in the Blue sub-sample. We then calculated their satellite radial density profiles with the full secondary sample, and compared the results to the satellite radial density profile using the full "Kick-out" primary sample, as shown in Fig. 4.10.

From the plot, we can see that the satellite radial distribution also has a very strong dependence on the primary colour. The Red type primaries have much higher satellite radial number density than the Blue type primaries. The satellite distribution of the Red type primaries also extends to larger radii (signal-to-noise ratio peaks at $\sim 450$ kpc) than the Blue type primaries (signal-to-noise ratio peaks at $\sim 150$ kpc). Furthermore, the more extended distribution of the red primaries may indicate that they are located in halos with larger virial radius (more massive) or more likely to be located in more massive halo groups than the blue primaries.

## 4.2   Galaxy Alignment

The numerical simulations based on the CDM model predict the shapes of DM halos are usually not spherical, but are more likely ellipsoidal or triaxial (Jing & Suto, 2002). Some studies also show that the subhalos have anisotropic distributions that are slightly aligned with the major axis of the host halos (Knebe et al., 2004; Zentner et al., 2005), which is probably due to angular correlations of satellite infall directions with large scale filaments (Tormen, Bouchet & White, 1997; Knebe et al., 2004; Zentner et al., 2005). Assuming satellite galaxies can trace the angular distribution of subhalos and the orientations of the central galaxies are also aligned with their host halos, we might then expect to observe some anisotropy in the distribution.
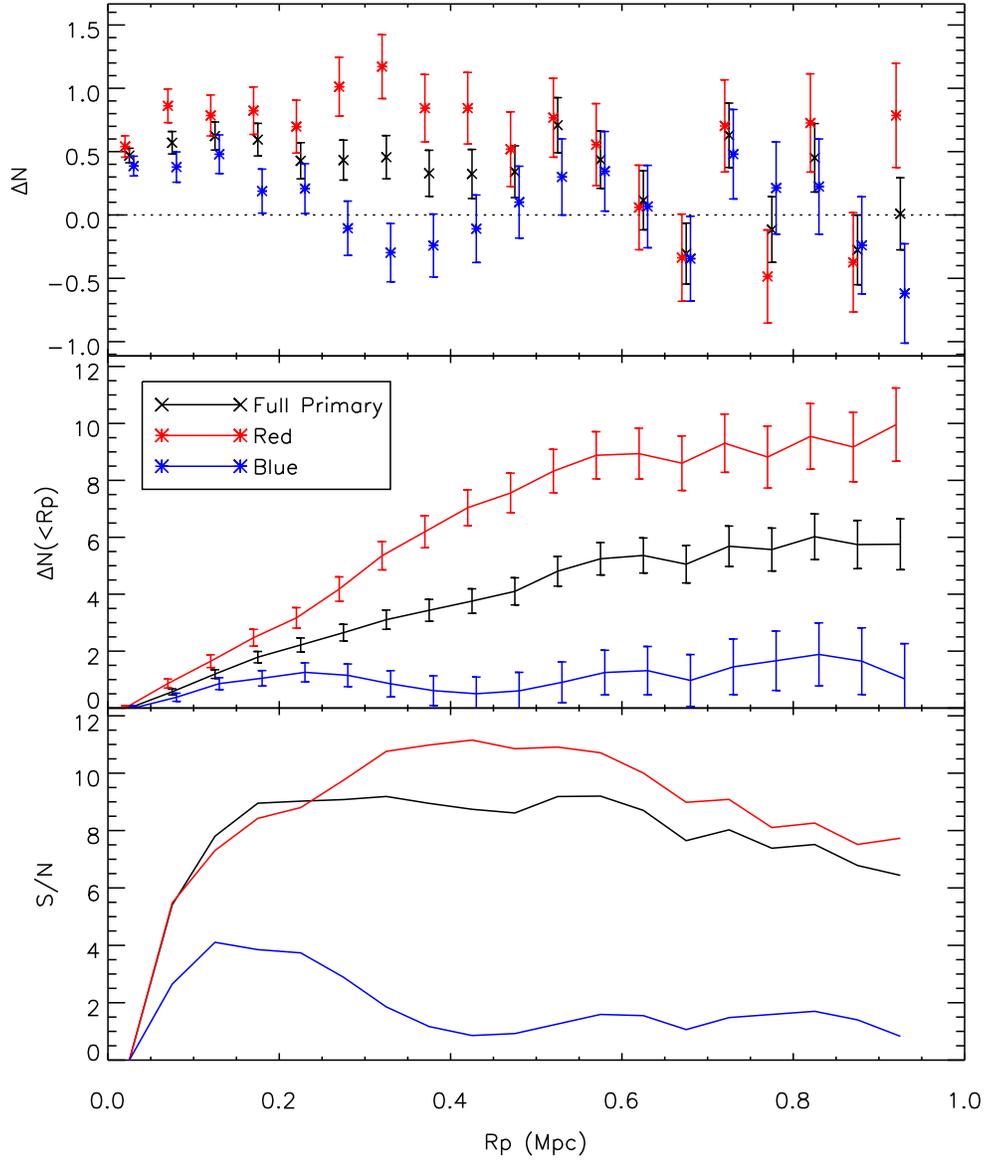
Figure 4.10: The radial distribution of satellites around Red type (red) primary galaxies and Blue type (blue) primary galaxies, with a comparison of the full "Kick-out" primaries (black).

60
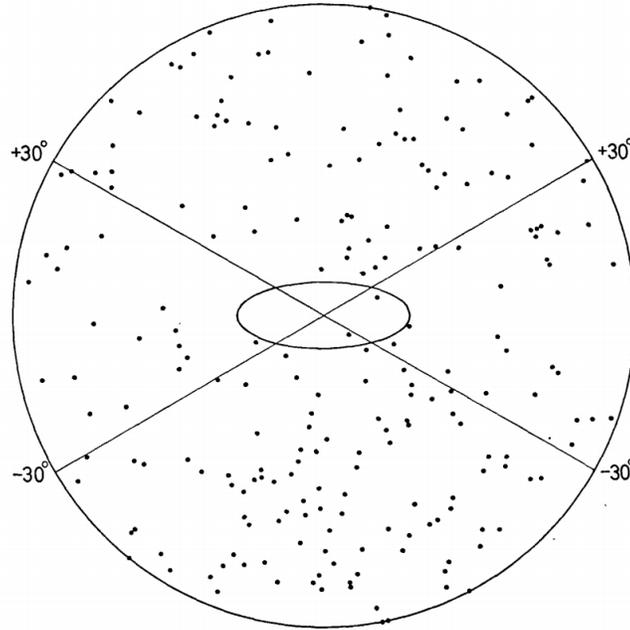
Figure 4.11: The galaxy alignment found by Holmberg. The plot shows the combined spatial distribution of 218 galaxies around 58 Sa central galaxies (Holmberg, 1969).

## 4.2.1 The Holmberg Effect

Holmberg (1969) investigated 174 groups of galaxies and studied the distribution of their position angles and separations with respect to their central galaxies. He found that for spiral primaries, the companion galaxies preferentially align with the minor axis of their central galaxies (Holmberg, 1969), as shown in Fig. 4.11. However, this galactic alignment effect, which is known as "Holmberg effect", was not confirmed by many subsequent studies over the next decade (Hawley & Peebles, 1975; Sharp, Lin & White, 1979; MacGillivray et al., 1982). Note that the original detection of Holmberg was restricted to a small range of projected separation: $r_p < 50kpc$ (Yang et al., 2006; Holmberg, 1969).

Later on, Zaritsky et al. (1997) attempted to measure the Holmberg effect within a much larger projected distance range. They did not find any significant detection of alignment for the distance range of $r_p = 0 \sim 200$ kpc, but they did manage to detect a preferred satellite-central alignment with the minor axis of the central galaxies for a larger range of separation of $300kpc \leqslant r_p \leqslant 500kpc$ (Zaritsky et al., 1997), which are already larger than the typical virial radii of the host halos of those central galaxies (Yang et al., 2006).

The alignment at large distance scale was later confirmed by Sales & Lamdas (2004), using the two-degree Field Galaxy Redshift Survey (2dFGRS) data, despite the fact that they were using a relatively small line-of-sight velocity constraints ($|\Delta v| < 160$ km s$^{-1}$) to identify satellites (Sales & Lambas, 2004). The MW satellites may also show a Holmberg effect — the 11 innermost satellites are distributed almost perpendicular to the MW disk (Lynden-Bell, 1982; Kroupa, Theis & Boily, 2005), although extinction complicates this conclusion.

More recently, however, the opposite results was found by several groups (Brainerd, 2005; Yang et al., 2006; Azzaro et al., 2007; Wang et al., 2009; Zhang et al., 2013). These studies used SDSS data and analysed several types of primaries and satellites. They all found that the satellites preferred to align with the major axis of the central galaxies, with complicated dependence with the primary and secondary properties.

Given the conflicting conclusions of these studies, we wished to test the isotropy of the satellite-central distribution independently with our sample. In addition, the COSMOS data has the great advantage of high quality photometric redshifts for the satellites. Note that the poor quality of Sloan photometry forced the previous studies using SDSS to consider a very large range of line-of-sight velocity difference to include all the potential satellites. The high resolution imaging in COSMOS is also very helpful to resolve the potential satellites within the innermost radial bins.

## 4.2.2   Methodology

**Position Angle Measurement**

We define the orientation of the primary galaxies by using the position angle $\varphi_a$ of their major axis that is provide by the COSMOS shape catalogue, which is measured from the North (the increasing Dec direction), and going East (the increasing Ra direction). Using the sample position angle system, the position angle of the galaxy-galaxy line $\varphi_{GG}$ for the galaxies around our primary sample is calculated based on their coordinate information:

$$\varphi_{GG} = tan^{-1}\left(\frac{(\alpha_S - \alpha_P)cos(\delta_P)}{\delta_S - \delta_P}\right) \tag{4.9}$$

where $\alpha_P$ and $\alpha_s$ are the right ascensions (RA) of the primary and secondary galaxies, and $\delta_P$ and $\delta_S$ are the declinations (DEC) of the primary and secondary galaxies.

We define the relative position angle $\theta$ as:

$$\theta = \begin{cases} |\varphi_{GG} - \varphi_a| & \text{When } |\varphi_{GG} - \varphi_a| < 90° \\ 180° - |\varphi_{GG} - \varphi_a| & \text{When } |\varphi_{GG} - \varphi_a| > 90° \end{cases} \tag{4.10}$$

where $\theta = 0°$ implies the secondary sample perfectly aligns with the major axis; while $\theta = 90°$ implies the secondary sample perfectly aligns with the minor axis.

## Binning Strategy

Previous studies (Yang et al., 2006; Zhang et al., 2013) showed that satellite-central alignment has a dependence on the projected radius from the central. So rather than simply binning the satellite counts in position angle, we choose to bin the pair counts $n(\theta, r_p)$ in both position angle $\theta$ and projected separation $r_p$ from the central galaxies. The radial binning is also helpful for the area incompleteness correction (see section 6.1). Although the average satellite contribution per radial bin becomes insignificant at large projected separation ($r_p > 600$ kpc) from the central galaxies, the radial distribution may be slightly different for different position angles. Therefore, we choose to count the satellites in a larger radial range, extending to $r_p \leq 950$ kpc, than what we did in the previous section.

## Background Subtraction

We use essentially the same method for estimating the background as the locally-averaged background we used for the last section. We use the number of pairs at the projected separation range of 750 kpc $< r_p <$ 1000 kpc as the satellite background counts.

Assuming the background is isotropic for each primary, then we have:

$$n_i^{bg}(\theta, r_p)dr_pd\theta = \frac{A_i(r_p)dr_p}{A_i^{out}}N_i^{out} \cdot \frac{d\theta}{360°}, \tag{4.11}$$

where $n_i^{bg}$ refers to the background of the two-dimensional number density distribution around the primary $i$; $A_i(r_p)dr_p$ and $A_i^{out}$ are the area of a inner radial bin and the area of the outer region (750 – 1000 kpc as defined above.) around the primary $i$. Note that the effective area are estimated from counting the random sample around each primary, as we did in the previous section.

Now we have the background for each position angle-projected distance bin. By subtracting the background, we then can have a 2-D distribution of observed satellite number density with the two parameters:

$$n_i^{sat}(\theta, r_p)dr_p d\theta = n_i(\theta, r_p)dr_p d\theta - \frac{A_i(r_p)dr_p}{A_i^{out}}N_i^{out} \cdot \frac{d\theta}{360°} . \qquad (4.12)$$

where $n_i$ is the total observed pair number count around the primary $i$ and $n_i^{sat}$ is the excess pair number count after background subtraction (the observed number of satellites). We can fix the area incompleteness effect with the same method we used in the previous section (also see section 6.1 for detailed explanation and discussion):

$$n_i^{sat'}(\theta, r_p)dr_p d\theta = \gamma_i(r_p, r_p + dr_p)n_i^{sat}(\theta, r_p)dr_p d\theta \qquad (4.13)$$

where $\gamma_i(r_p, r_p + dr_p)$ is the reciprocal of the area incompleteness at the radial bin of $r_p \sim r_p + dr_p$ for the primary $i$, which is defined as:

$$\gamma_i(r, r + dr) = \frac{A'_i(r, r + dr)}{A_i(r, r + dr)} \qquad (4.14)$$

where $A'_i(r, r + dr)$ refers to the area of the radii range $(r, r + dr)$ without any masking or boundary conditions.

The $n_i^{sat'}(\theta, r_p)$ describes the two-dimensional number density of the satellite as a function of two parameters, which itself is already very interesting result. When we integrate the 2-D distribution over certain radial ranges, we can the angular satellite number density distribution over those radial ranges for individual primaries:

$$n_i^{sat'}(\theta)|_{r_1}^{r_2} = \int_{r_1}^{r_2} n_i^{sat'}(\theta, r_p)dr_p . \qquad (4.15)$$

We then can average the distribution over all the primaries:

$$N'_{sat}(\theta)|_{r_1}^{r_2} = \frac{1}{N_p}\sum_{i=1}^{N_p} N'_{sat}(\theta)|_{r_1}^{r_2} \qquad (4.16)$$

When we are looking the angular satellite number density distribution and trying to find whether there is a Holmberg effect, we care more about the relative distribution between each angular bin than the individual density values. Thus we define the normalized angular satellite probability distribution function as:

$$f_{sat}(\theta) = \frac{N'_{sat}(\theta)}{\langle N'_{sat}(\theta)\rangle} \qquad (4.17)$$

where $\langle N'_{sat}(\theta)\rangle$ is the averaged angular satellite number density over the position angles.

### 4.2.3  Results

**Visualizing the Projected Satellite Density Field**

We included all the pairs within a projected radii of $r_p \leqslant 1 Mpc$ and velocity difference of $|\Delta V| < 300 km/s + 1\sigma_V$ from our primaries. We then combined all the systems by shifting to coordinates centred on the central galaxies, and rotated so the major axes of central galaxies are aligned. Then we plotted all the pairs with their position angles and scaled projected separations from the central galaxies, which gives us a density map of all the close pairs around our primary sample. Subtracting the background density field that we generated previously, the density field left describes the projected spatial distribution of the satellites, which can be easily visualized.

Examining the derived density maps around several sub-samples of our primary sample, we noticed that the satellites of all types of primary samples seem to have fairly isotropic distribution at small projected radial separations ($r_p < 350$ kpc). At large separations ($r_p > 450$ kpc), the satellites, however, seem to show some anisotropic distribution patterns, where we found that the early-type primaries, for instance, seem more likely to align with the major axis of the central galaxy disks, showing an anti-Holmberg effect. These different alignment patterns between the inner regions and outer regions inspired us to measure the alignment at different radial ranges.

**Position Angle Distribution at Inner and Outer Radii for Different Primary Samples**

We measured and integrated the two-dimensional satellite number density and the normalized angular satellite probability distribution over three radial ranges: full radii (50 – 800 kpc), inner radii (50 – 400 kpc) and outer radii (450 – 800 kpc). We omitted the outermost radial bins (800 kpc $\leqslant r_p \leqslant$ 950 kpc) for the consideration of signal-to-noise level. We also investigated the dependence of the galactic alignment on the primary properties (e.g. primary luminosity, galaxy type and color), by using the corresponding sub-samples to calculate the results separately. The sub-samples involved include the full base primary sample before applying "Kick-out"(489 galaxies) , the "Kick-out" primary sample (392 galaxies), the bright primary sub-sample (120 galaxies), the faint primary sub-sample (272 galaxies), the early-type (Elliptical/S0) primary sub-sample (62 galaxies), the late-type (Spiral) primary sub-sample (329 galaxies), the red primary sub-sample (197 galaxies) and the blue primary sub-sample (195 galaxies), whose detailed selections are explained in section 2.8.

The results over the full radial range are shown in Fig. 4.12. Integrating over all radii, the satellite distributions of all subsamples of primaries appear to have some anisotropic patterns, which are however hard to confirm given the signal-to-noise ratio. The results over the inner radial range are shown in Fig. 4.13, from which we can find that the satellites around most types of primaries are fairly isotropic. The pair detections in each bin are mostly consistent with average within $1\sigma$ uncertainty, except for one $1.5\sigma$ detection for the blue primaries. This detection is still not significant enough to confirm any anisotropic distribution. The results over the outer radial range are shown in Fig. 4.14, from which we can see that quite a few types of primaries seem to have anisotropic satellite distributions, including the full base primary sample, the "Kick-out" primary sample, the faint primary sample, the early-type primary sample and the red primary sample. All those primary samples show the same alignment pattern, where their satellites appear to have a preferred alignment with the major axis of the central galaxies. The early-type primary sample and the red primary sample show the strongest detection. The early-type primary sample has a $2\sigma$ detection and a $1.5\sigma$ detection; whereas the red primary sample has $3\sigma$ detection. In addition to the detections in individual bins, the detections of these primary sample over all angular bins show consistent trend. At the same time, no significant anisotropic satellite distribution patterns are found for the late-type and blue primary samples.

## 4.2.4   Discussion

We can draw the following conclusions from the plots:

The position angle distributions of the full radial range 50 – 800 kpc, for all types of primaries do not show any significant anisotropy. The faint primary, early-type primary and red primary samples show possible indications of the anti-Holmberg effect, but the detections at the angular bin is only marginal.

When the angular distributions are measured in two projected distance ranges, the inner regions (50 – 400 kpc) and outer regions (450 – 800 kpc), interesting patterns occur. The satellite distributions in the outer regions show several possible preferred alignment detections while the satellite distributions in the inner regions are fairly isotropic. This is expected as the anisotropy of subhalo distributions are mainly due to large scale filaments (Tormen, Bouchet & White, 1997; Knebe et al., 2004; Zentner et al., 2005), which affect the subhalo distribution of outer regions more significantly than the inner regions. If we assume our detections in the outer regions are real anti-Holmberg detections, the detections first may indicate the orientations of the central galaxies have preferred alignments with the orientation of the host halos. Secondly, the detections may show some essential difference
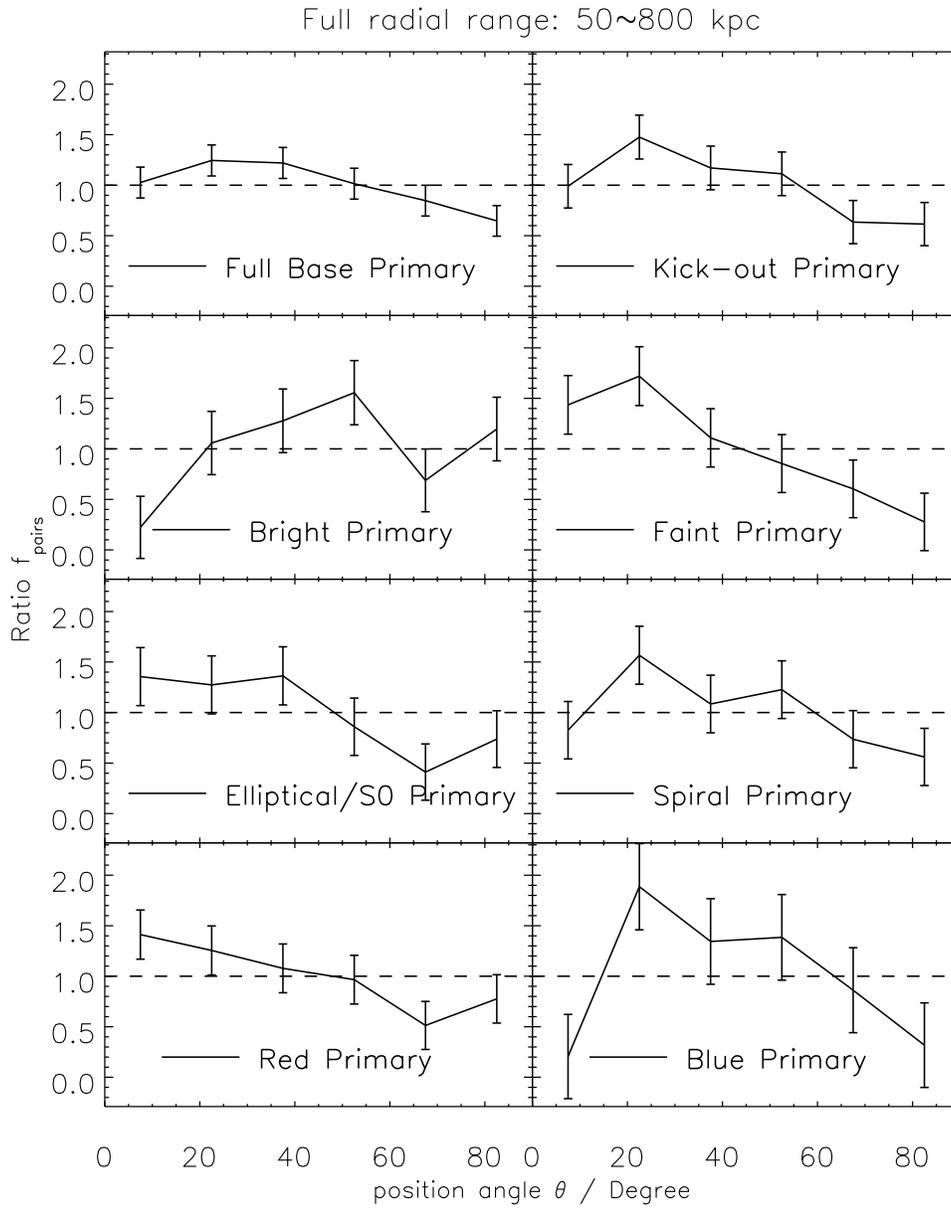
Figure 4.12: The normalized angular satellite number density distribution averaged within $50 - 800$ kpc. See section 2.8 for detailed explanations of each primary sub-samples used in this plot.
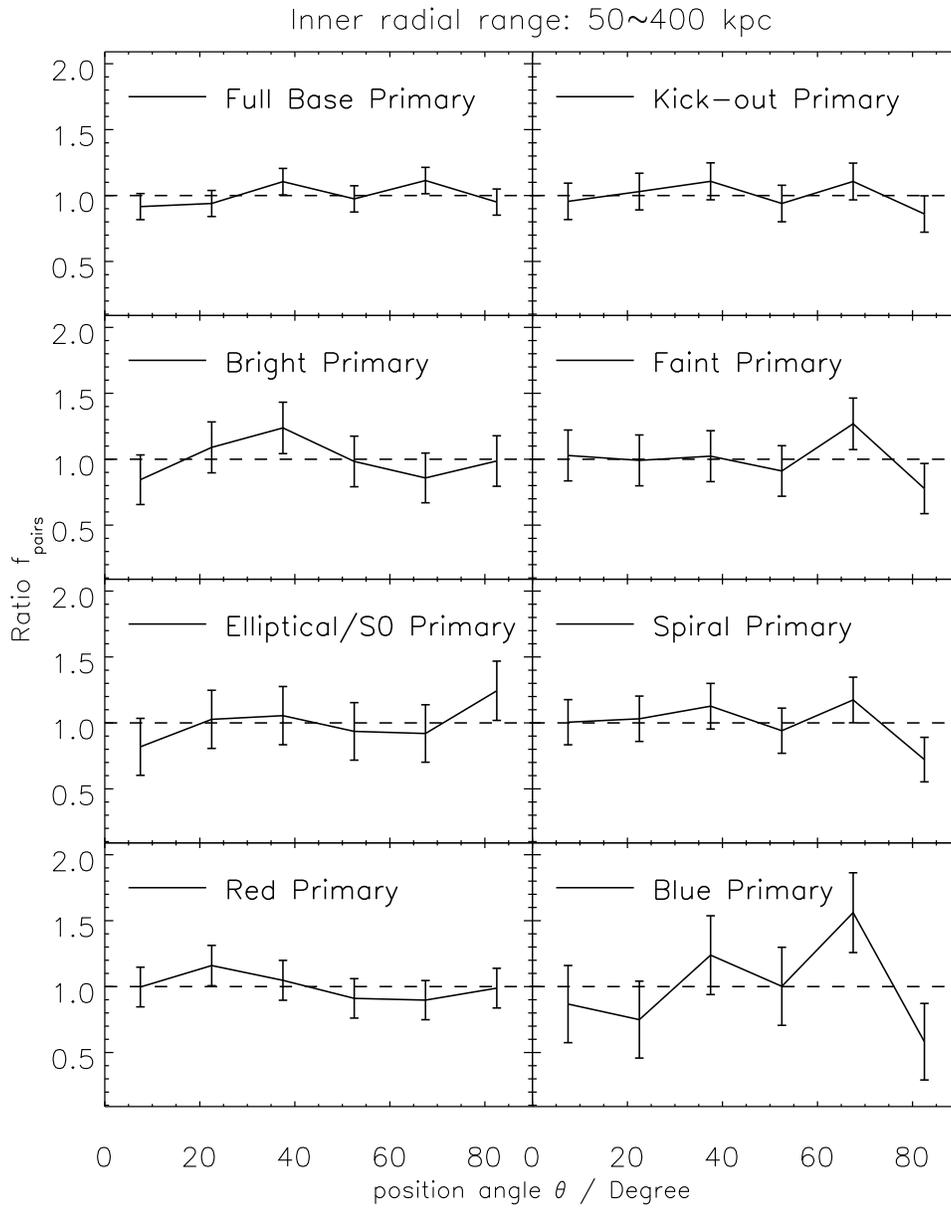
Figure 4.13: The normalized angular satellite number density distribution averaged within 50–400 kpc. See section 2.8 for detailed explanations of each primary sub-samples used in this plot.
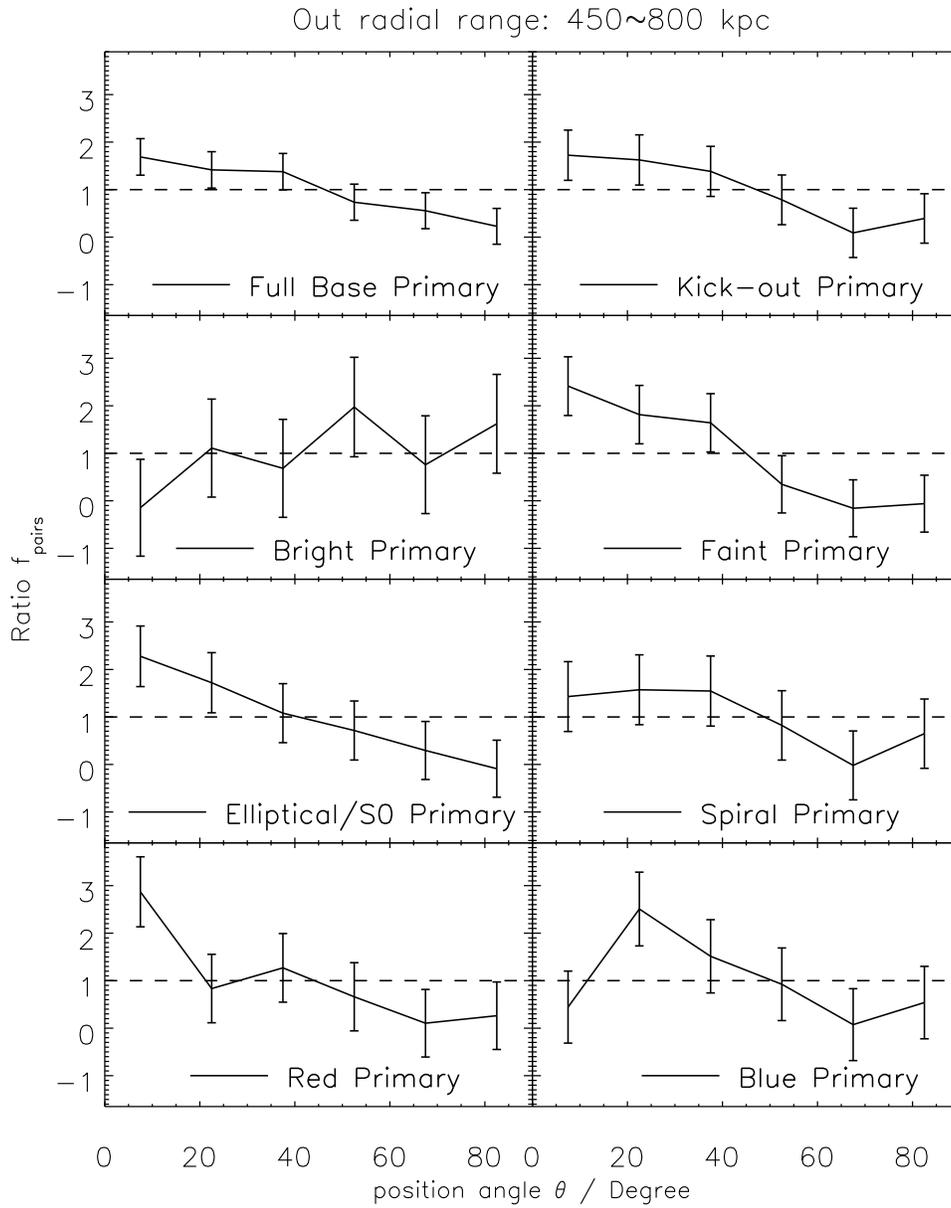
Figure 4.14: The normalized angular satellite number density distribution averaged within 450 − 800 kpc. See section 2.8 for detailed explanations of each primary sub-samples used in this plot.

of satellite alignments between the early-type and late-type primary galaxies, which may hold possible clues for galaxy formation.

# Chapter 5

# The Relative Luminosity Function

The Relative Luminosity Function (RLF) $N(\Delta m)$ describes the mean number of satellites per primary that are at the given magnitude range $\Delta m$ relative to the primary, where we define $\Delta m = m_{sat} - m_{main}$ . This function has been studied previously by many other groups (Liu et al., 2011; Guo et al., 2011b,a, 2012, 2013; Nierenberg et al., 2011, 2012; Strigari & Wechsler, 2012; Speller & Taylor, 2014), and provides a point of comparison between the satellite abundance to the predicted abundance of DM substructures. The dark matter simulations predict that dark matter halos should host a constant number of subhalos within a given dark matter mass fraction of the host halo (Kravtsov, 2010; Nierenberg et al., 2012). The scale invariance relates to the measurable RLF, given the assumption that the satellite and central galaxies trace the subhalos and host halos. The RLF is expected to be similar to the relative mass function but different, due to the non-linearity of luminosity-mass relationship and stellar-mass-to-halo-mass relationship.

In this chapter, we construct our relative luminosity function measurements using the COSMOS datasets. We explain our method of calculating the RLF in section 5.1; In section 5.2, we then present our results of RLF measurements on the satellite population around our primary sample and the primary sub-samples; and we discuss the dependence of the RLF on the properties and compare our measurements to the previous studies.

## 5.1   Methodology

To build the RLF, we need to know the distribution of satellites as a function of their luminosity relative to their central primary. At the same time, we also need to know

their radial distribution so that we are able to fix the area incompleteness caused by the masking and other boundary conditions. Our method is to count the galaxy pair number with a binning in two parameters, including the binning of the relative magnitude difference between the primary and the secondary sample, and the binning of the projected separation from the central primaries. We use a bin-size of 1 magnitude ($\Delta m_{k+1} - \Delta m_k = 1.0$) for the relative luminosity binning, and a bin-size of 50 kpc for the radial binning ($r_{p_{j+1}} - r_{p_j} = 50$ kpc). Thus, for each primary, the number of secondary-primary pairs insider the bin $\Delta m_k$ and $r_{p_j}$ can be written as:

$$n_i(r_{p_j}, \Delta m_k) = N_i^{pair}(r_{p_{j-1}} < r_p < r_{p_j}, \Delta m_{k-1} < \Delta m < \Delta m_k) \tag{5.1}$$

For the $\Delta m$, we use the difference of the g-band apparent magnitude between the secondary and the primary. Furthermore, We use the pair counts at the same relative luminosity bin at the outer radial range of $750 kpc \leqslant r_p \leqslant 1000 kpc$ as the background, written as $n_i^{outer}$. We then estimate the effective area for the each radial bin and for the outer region by counting a randomly generated sample of 300,000 objects around each primary $A_i^{outer}$. Note that we also count the random-primary pairs before any masking or boundary condition is made to the random sample $A'^{outer}_i$, which gives an estimate of the "real" area for each primary and will be used later to fix the area incompleteness.

We also count the random-primary pairs for each radial bin and estimate the background for each radial bin by scaling the secondary-primary counts of the outer region with the ratio of the random counts:

$$n_i^{bg}(r_{p_j}, \Delta m_k) = \frac{A_i(r_{p_j})}{A_i^{out}} n_i^{outer}(\Delta m_k) \tag{5.2}$$

For each bin, we then subtract the background from the secondary-primary pairs and get the satellite number in that bin.

$$n_i^{sat}(r_{p_j}, \Delta m_k) = n_i(r_{p_j}, \Delta m_k) - n_i^{bg}(r_{p_j}, \Delta m_k) \tag{5.3}$$

Finally we correct the area incompleteness effect by multiplying the satellite number of each radial bin with the reciprocal of the area completeness of the corresponding bin:

$$n_i^{sat'}(r_{p_j}, \Delta m_k) = n_i^{sat}(r_{p_j}, \Delta m_k) \times \frac{A'_i(r_{p_j})}{A_i(r_{p_j})} \tag{5.4}$$

where $n_i^{sat'}(r_{p_j}, \Delta m_k)$ refers to the derived satellite number of the bin, without any masking or boundary conditions for the primary "$i$"; $A'_i(r_{p_j})$ and $A_i(r_{p_j})$ refer to the random-primary pair counts of the corresponding radial range of the primary "$i$", whose ratio gives an estimate the reciprocal of the area completeness of the radial bin.

We use Poisson statistics to estimate the uncertainty of satellite number, which mainly comes from the counting of secondary-primary pairs:

$$\sigma_{n_i(r_{p_j}, \Delta m_k)} = \sqrt{n_i(r_{p_j}, \Delta m_k)}. \tag{5.5}$$

The Poisson uncertainty of the random-primary counting is omitted here as the large number size of the random sample makes it totally negligible in the end. The error of the satellite number in each bin can be expressed as:

$$\sigma_i^{sat'}(r_{p_j}, \Delta m_k) = \frac{A'_i(r_{p_j})}{A_i(r_{p_j})} \sqrt{n_i(r_{p_j}, \Delta m_k) + \left(\frac{A_i(r_{p_j})}{A_i^{out}}\right)^2 n_i^{outer}(\Delta m_k)}. \tag{5.6}$$

The remaining work is to combine the individual bins and sum them up to the whole RLF. First, we sum the satellite numbers (that is the excess secondary counts relative to the background) of all the relative magnitude-radial bins over the radius range of $50 - 600$ kpc.

$$dN_i(\Delta m_k) = \sum_{r_{p_j}=50kpc}^{600kpc} n_i^{sat'}(r_{p_j}, \Delta m_k). \tag{5.7}$$

The error can be expressed as:

$$\sigma_i^2(\Delta m_k) = \sum_{r_{p_j}=50kpc}^{600kpc} \left[\sigma_i^{sat'}(r_{p_j}, \Delta m_k)\right]^2. \tag{5.8}$$

The next step is to average the satellite number per relative magnitude bin over the primaries, which is actually is not that straightforward because of the magnitude depth of the survey. When the relative magnitude gets big enough, there would be no corresponding secondary galaxies for some of the primary galaxies, due to the magnitude limit of the secondary sample. These primary galaxies should not be included during the averaging for that relative magnitude bin as they are actually not contributing anything. So the averaged satellite number for each relative magnitude bin should be:

$$d\bar{N}(\Delta m_k) = \frac{\sum_{i=1}^{N_p} \sigma_i^2(\Delta m_k)\, \delta(m_i^p, \Delta m_k)}{\sum_{i=1}^{N_p} \delta(m_i^p, \Delta m_k)}, \tag{5.9}$$

where $N_p$ refers to the total number of the primary sample, $m_i^p$ refers to the magnitude of the primary $i$, and $\delta(m_i^p, \Delta m_k)$ is the criterion for whether the primary should be taken

73

into account, which can be defined as:

$$\delta(m_i^p, \Delta m_k) = \begin{cases} 1 & \text{if } m_i^p < m_{limit} - \Delta m_k, \\ 0 & \text{if } m_i^p > m_{limit} - \Delta m_k, \end{cases} \tag{5.10}$$

where $m_{limit}$ is the magnitude limit of the secondary sample.

Similarly, the overall error of the satellite number at the relative magnitude bin $\Delta m_k$ can be expressed as:

$$\bar{\sigma}^2(\Delta m_k) = \frac{\sum_{i=1}^{N_p} dN_i(\Delta m_k)\, \delta(m_i^p, \Delta m_k)}{\sum_{i=1}^{N_p} \delta(m_i^p, \Delta m_k)}, \tag{5.11}$$

At this point, the equation of $d\bar{N}(\Delta m_k)$ already gives the differential format of the RLF in principle. However, it would be inconsistent to consider the RLF for $\Delta m < 1.0$, since by construction, the "Kick-out" primary sample[1] excludes pairs with relative magnitude difference of less than one. Thus we decide to omit the detection of the first relative magnitude bin, which corresponds to $\Delta m < 1.0$. This would affect the cumulative result by about 0.1 satellite per primary.

## 5.2 Results

### 5.2.1 Overall Results of the "Kick-out" Primary Sample

We first calculated the RLF for our "Kick-out" primary sample with the full secondary sample. We detected an excess pair signal from companion down to $\Delta m = 10$, which is two magnitudes fainter than most of the previous studies, while brighter than the measurement by Speller & Taylor (2014) by two magnitudes. Correcting for the magnitude limits of COSMOS, we detected a total excess of objects within separation $< 600$ kpc, corresponding to about 17.7 ± 5.1 satellites per central galaxy at $\Delta m \leq 10$ and 5.5 ±0.5 satellites per central galaxy at $\Delta m \leq 7$, where the detection has the highest signal-to-noise ratio.

Fig. 5.1 compares the RLF measurement using the "Kick-out" primary sample (392 galaxies) to the measurement using the full base primary sample (489 galaxies) and the RLF of the MW satellites. From the plot, we can see our detection using the "Kick-out" primary sample is similar to the MW satellites overall, while it is lower than the MW

---

[1]Definition of the "Kick-out" sample, see section 2.6
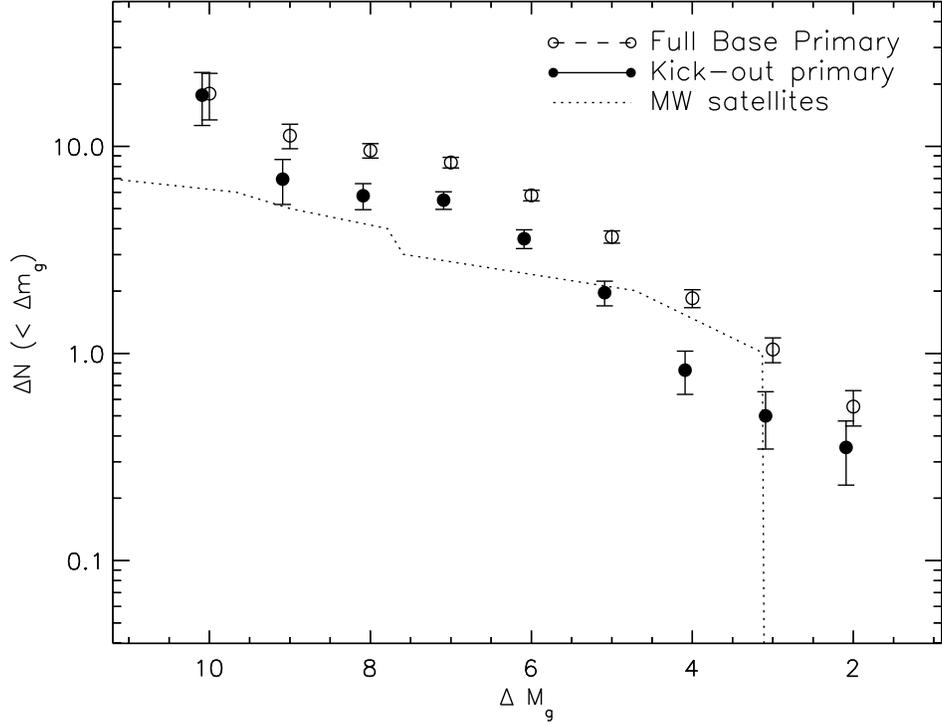
Figure 5.1: The RLF of using the "Kick-out" primary sample and the RLF of using the base primary sample without kicking out any potential satellites.

satellite sample at $3 < \Delta m < 5$ and higher than the MW satellite sample at $5 < \Delta m < 8$. Specifically, at $\Delta m = 3$, we have a satellite detection of $0.50 \pm 0.15$ galaxies per central galaxy, which may suggest the LMC is a statistically rare case ($> 3\sigma$).

Additionally, we found that the RLF of the "non-Kick-out" primary sample has much higher number of satellites at a given relative luminosity. The difference is probably caused by the fact that the "non-Kick-out" primary sample actually contains some satellite galaxies. The detections around those galaxies are actually contributed by their host galaxies, whose luminosities are much higher. This means we shifted the RLF signals that should correspond to a larger relative magnitude to the lower relative magnitude end, producing misleading results.
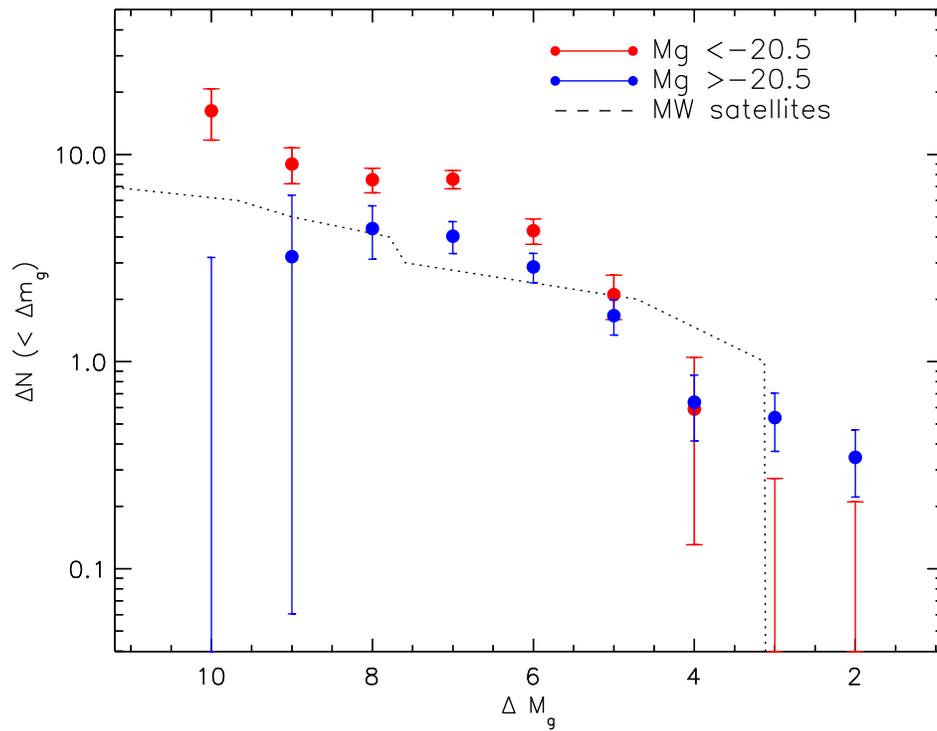
Figure 5.2: The RLF of using the Bright primary sample that has 120 galaxies with $M_g < -20.5$ and the RLF of using the Faint primary sample that has 272 galaxies with $M_g > -20.5$.
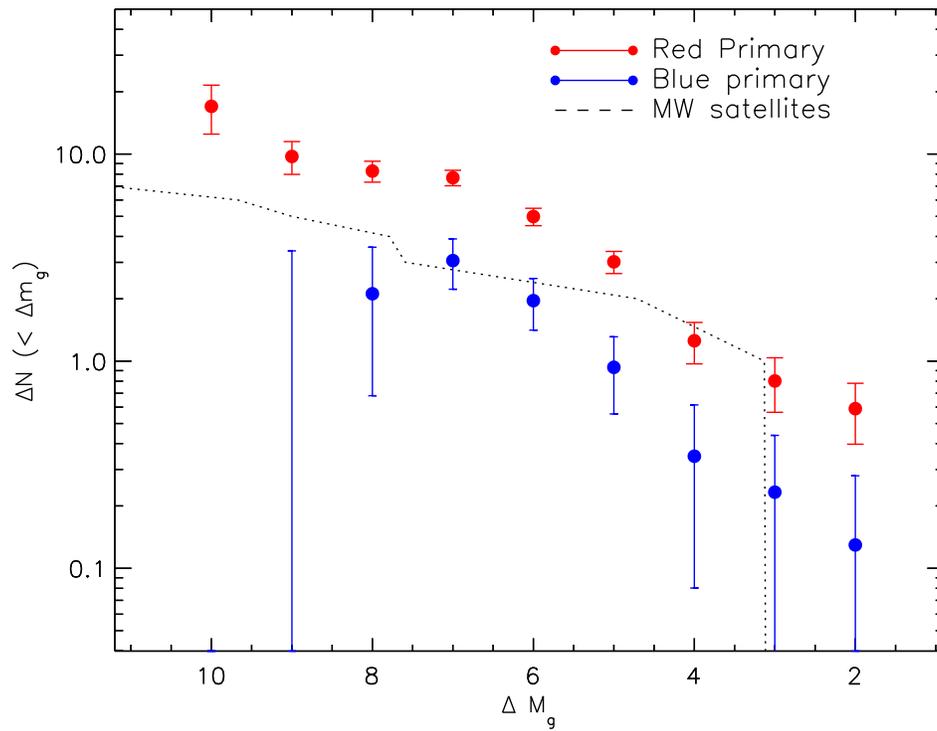
Figure 5.3: The RLF of using the Red primary sample that has 197 galaxies with $B - V < 0.48$ and the RLF of using the Blue primary sample that has 195 galaxies with $B - V > 0.48$.
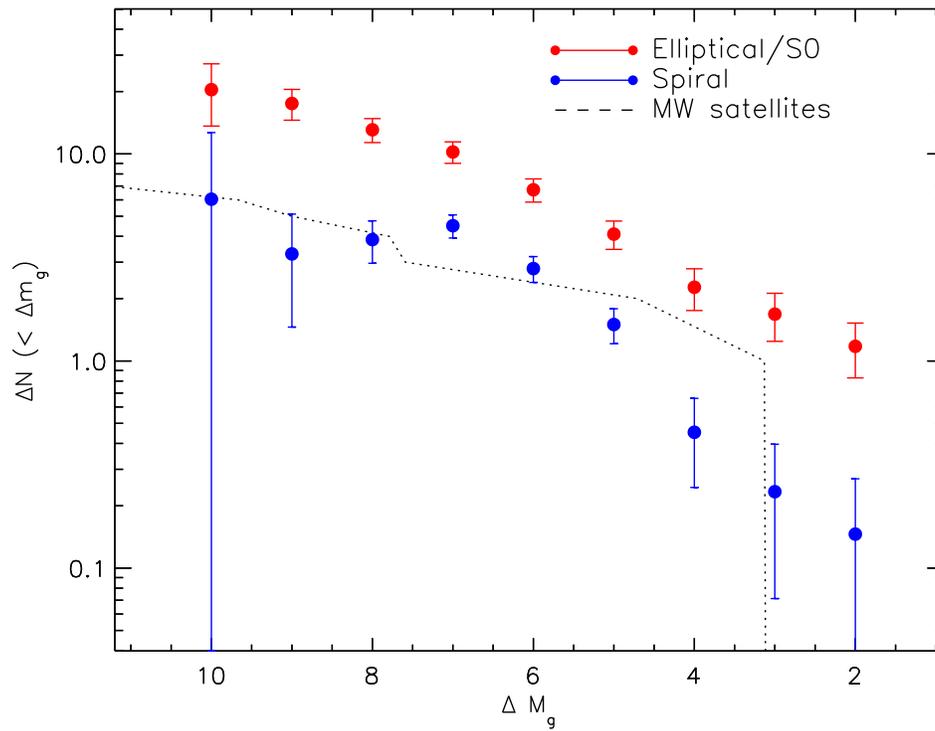
Figure 5.4: The RLF of using the Elliptical/S0 primary sample that has 62 galaxies with the spectrum template type "MODD" $\leq 8$ and the RLF of using the Spiral primary sample that has 330 galaxies with the spectrum template type "MODD" $> 8$.

## 5.2.2 Dependence on Primary Properties

Just as we did in the previous chapters, we also split the "Kick-out" primary sample into 6 sub-samples based on the galaxy properties. Then we calculated their RLF with the same secondary sample (full secondary sample) separately and compared the results to show the dependence of the RLF on the properties of the central galaxies.

The RLFs of the Bright primaries ($M_g < -20.5$, 120 galaxies) and the Faint primaries ($M_g > -20.5$, 272 galaxies) are shown in Fig. 5.2. From the plots, we can see that the Red primaries have a significantly higher (three times more) average satellite number over the relative magnitude range than the Blue primaries. This is actually quite a surprising result. From Fig. 2.10, we know that there is not much luminosity difference between the Red primaries and Blue primaries. Therefore, this huge satellite number difference that cannot even be found in between the Bright and Faint primaries may suggest that the satellite abundance has a strong dependence on the color of the central galaxies.

The comparison of the RLFs of the Red primaries ($B - V < 0.48$, 197 galaxies) and the Blue primaries ($B - V > 0.48$, 198 galaxies) is shown in Fig. 5.3. From the plot, the bright primaries seem to have significantly larger number of satellites at the large relative magnitude end while having fewer satellites at the small relative magnitude end than the faint primaries. These differences are very likely to be artificial, however. As our secondary samples have a brighter limit of g-band absolute magnitude $M_g > -19.0$, there is a magnitude gap of 1.5 magnitude between the Bright primary and the secondary sample, which will lower the satellite number of the Bright primary sample at the small relative magnitude end. On the other hand, our secondary also has a fainter limit of g-band apparent magnitude of $g^+ < 26.0$. The satellite counting will become incomplete when the relative magnitude become higher than a certain point. This effect will obviously happen earlier (lower relative magnitude) for the Faint primary sample, which may be responsible for the difference of the RLF at the large end of the relative magnitude scale.

The RLFs of the Elliptical/S0 primaries ("MODD" $\leq 8$, 63 galaxies) and the Spiral primaries ("MODD" $> 8$, 272 galaxies) are shown in Fig. 5.4. Similarly to the result of the Red primaries compared to the Blue primaries, the Elliptical/S0 primaries have a significantly higher satellite detection than the Spiral primaries, over the full relative magnitude range. We can conclude that the RLF also has a very strong dependence on the galaxy types of the central galaxies.

Overall, we find that the RLF does not show a very significant dependence on the primary luminosity, except for the differences at the low $\Delta m$ and the high $\Delta m$ ends, which are likely to be artificial as previously discussed. This luminosity invariance is very

similar to the scale invariance of dark matter substructures predicted by the simulations. In contrast, we find the RLF has a very strong dependence on the primary galaxy type. Early-type (elliptical, red) primaries have a significantly larger satellite population than the late-type (spiral, blue) primaries. Possible explanations involve the systematic underestimation of the photometry of the bright early-type galaxies, different luminosity-to-stellar-mass relationships, different stellar-mass-to-halo-mass relationships, or different environments between early and late type galaxies (e.g. Early-type primaries may be more likely to be found in groups, and exist in more massive host halos than late-type primaries.) In section 6.2, we test the possible effects of the potential bias of the photometry measurement for the bright early-type galaxies to the RLF measurement. Based on the results of these tests, we conclude that the systematics of the photometry can cause overestimate of the RLF of the early-type galaxies by only a small amount, but are not significant enough to explain the difference of the RLFs between the early-type primaries and the late-type primaries. If the difference is a result of different luminosity-stellar-mass relations, it gives a possible clue to the missing satellite problem. Star formation, for instance, probably cuts off in low mass subhalos, as a result of photo-ionization and feedback from SNe and AGN. The different environments of the subhalos around different types of central galaxies may suppress star formation in different ways. If the difference comes from some essential differences between the two types of galaxies, either in their stellar-mass-to-halo-mass relations or in their environments, it gives very important implications for galaxy formation (e.g. maybe more baryonic materials are striped away from the subhalos around the early-type central galaxies, as a result of different subhalo in-fall history, interactions with the disks of the central galaxies, or different feedback from AGN and SNe), and provides a test for the semi-analytical models that attempt to explain the missing satellite problem.

### 5.2.3   Comparison to Previous Studies

As mentioned in the beginning, the relative luminosity function has been measured by many previous studies (e.g. Liu et al. (2011); Guo et al. (2011b,a, 2012, 2013); Nierenberg et al. (2011, 2012); Strigari & Wechsler (2012); Speller & Taylor (2014) - see Nierenberg et al. (2012) and Speller & Taylor (2014) for a better summary). Fig. 5.5 compares our RLF measurement using the "Kick-out" primary sample to the measurement of the "low mass, low redshift" samples from Nierenberg et al. (2012, hereafter N12) (blue open triangles and red open squares represent early and late types of primary, offset slightly for clarity), the measurement and upper limits from Strigari & Wechsler (2012, hereafter SW11) (purple crosses and arrows indicate the measurement and the upper limits respectively) and to the

measurement from Speller & Taylor (2014, hereafter Speller14) (green diamonds). Note that the previous studies used r-band photometry while we are using g-band, which may effect the $\Delta m$ slightly. We investigated the averaged value of the magnitude difference of $\langle r - g \rangle$, which is about 0.3 for our primary sample, and is about 0.6 for the secondary sample. This means that those r-band-based results should be shifted slightly to the low $\Delta m$ end for a better comparison. The data points of previous studies in Fig. 5.5 are offset by $0.15 \sim 0.25$ for this reason and to add clarity.

From the plot, we can see all previous results are in good agreement for $\Delta m \sim 1 - 6$, in the case our measurement is slightly higher and has much better signal-to-noise ratios. This discrepancy becomes significantly smaller when the isolation is applied to our primary sample (see section 6.4 and Fig. 6.4). So we can conclude that the discrepancy is mainly due to the fact that we include the galaxies in groups and clusters in our "Kick-out" primary sample, which arguably implies that the primaries in groups and clusters are sitting in more massive host halos and have larger populations of satellites. Additionally, at $\Delta m \sim 7 - 8$, our measurement is lower than the measurement of N12 early types and SW11, which is not a surprise as SW only gives the estimate of upper limit, and the early-type primaries are expected to have a larger satellite population as we have shown in the previous section. Finally, we can see at $\Delta m \sim 6 - 10$, our measurement is higher than the measurement of Speller14 (even when isolation cut is applied, see Fig. 6.4). This is because that Speller & Taylor (2014) uses a size-magnitude cut to reduce the background, which increases the incompleteness of the faint secondary sample and thus lowers the measurement at the high $\Delta m$ end. The COSMOS has a better survey depth than the SDSS (down to $g^+ \sim 26$) and very high quality photo-zs, which allow us to measure the satellite abundance at the faint end with a much higher completeness. Also note that Speller & Taylor (2014) uses satellite counting range of $50 - 400$ kpc while we are using $50 - 600$ kpc. This may allow our result to include slightly more satellites for the very bright primaries by definition.
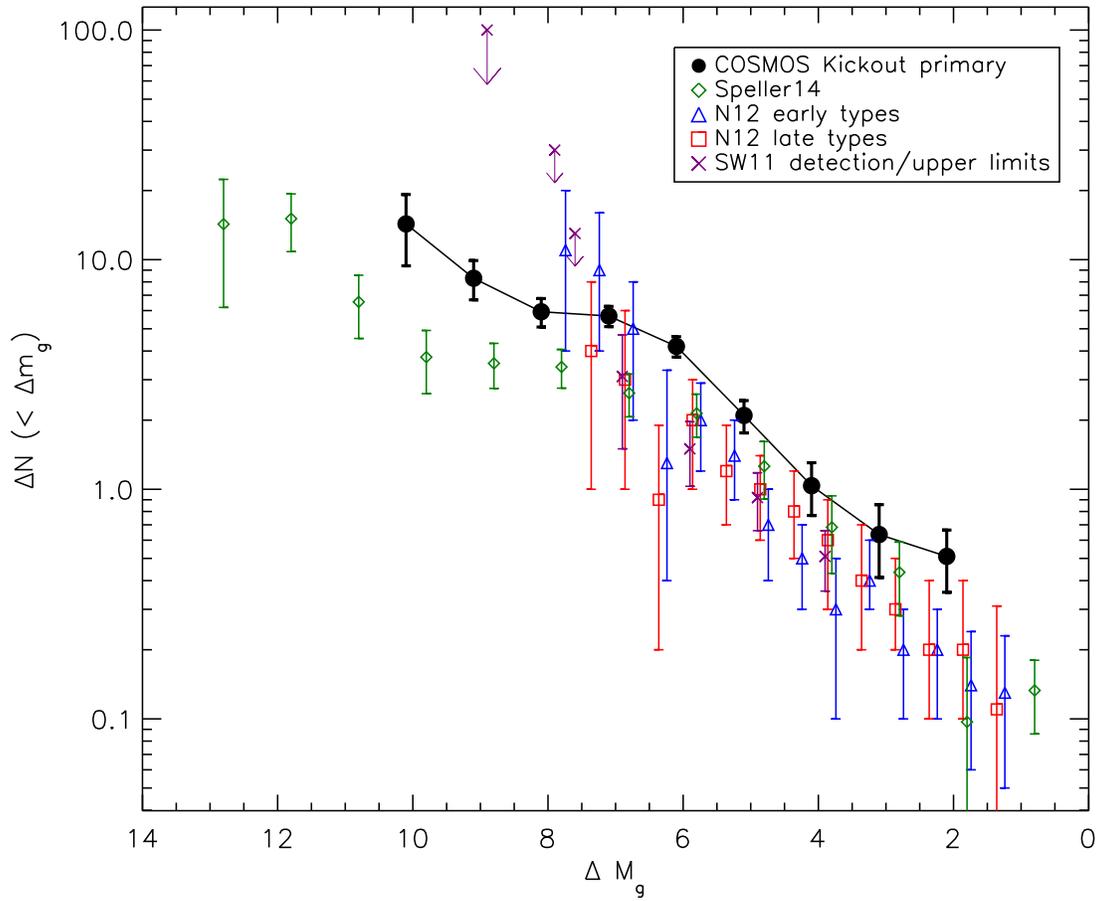
Figure 5.5: The RLF for the kick-out primary sample, compared to previous measurements.

# Chapter 6

# Systematics and Correction

## 6.1 Area Incompleteness Correction

Due to masking and our inner radial cut, the areas around the primaries are not complete. Therefore, the satellites that are located in those regions are not included in our abundance estimate.

To correct this affect, we assumed the masking and other boundary conditions are statistically isotropic, based on which we measured the effective area around each primary using random sample counts. We first filled the field that extends slightly over the ACS field with a sufficiently large random sample ($5 \times 10^6$ data points), without applying any boundary conditions. Afterwards, we counted the random points around individual primaries in radial bins of 50 kpc, and compared it to the same counts, but only after applying to the random sample the masking and boundaries corresponding to those in the observations. The ratio of the two random sample counts indicates the ratio of the areas that are missing in each radial bin for each primary. Accordingly, we corrected the area incompleteness by multiplying the satellite number each bin with the ratio of the full random counts to the counts after masking.

The key assumption for this area correction is that the masked area should be statistically no different from any other areas that are located in the same radial bin. This assumption is fairly reasonable but we have no good way to estimate the effect of this assumption being incorrect. If the masking was correlated with the large scale filaments, for instance, or if it was aligned with the principal axes of bright galaxies, our correction would introduce extra bias. Overall, however, this correction seems unlikely to introduce significant extra bias into our results.

## 6.2 Photometry of the Bright Early Type Galaxies

Recent studies pointed out that there is a trend in the traditional flux measurement to underestimate the total flux of bright early type galaxies, as a result of over-subtracting the background (He et al., 2013; Bernardi et al., 2013). This implies possible bias when we select the elliptical primaries for our study.

This effect may slightly affect our RLF measurement as it implies that we slightly underestimated the brightness of our early-type primaries and therefore also underestimated the value of $\Delta m$ for the RLF measurement. Equivalently speaking, the RLF signal from some primaries have a larger value of $\Delta m$ and the plot is expected to be shifted slightly to the larger magnitude difference end. We can test the significance of this systematic through a process of decreasing the magnitude of our blue primaries (making them brighter) by a small random amount and observing how it changes the result. He et al. (2013) remeasured the photometry for 2,949 bright ($M_r < -22.5$) SDSS early-type galaxies with improved algorithms and they claimed an average difference of about $0.16 \sim 0.26$ r-band magnitude measurement, depending on which algorithm is applied. According to their result, we added a Gaussian random distributed magnitude difference $\Delta m = -0.2 \pm 0.2$ to the magnitude of our elliptical/S0 type primary galaxies and then reran our RLF calculation for 10 times. We averaged the results from the 10 recalculations as the shifted RLF measurement, which is shown in Fig. 6.1, along with the original result for comparison.

The result of adding the extra magnitude does slightly affect the measurement. The RLF after the magnitude changes is lowered by a small amount, or equivalently speaking, is shifted to the high relative luminosity end by a small amount. This shows that the systematics of the photometry measurement of the bright early type galaxies are potentially important for the RLF measurement, but is not significant enough to explain the difference of the detection between the elliptical and spiral galaxies.

## 6.3 Photometric Errors in Areas Close to Bright Galaxies

The photometry measurement of faint objects can be affected by the halos and refractions from nearby bright stars and galaxies, which brings it extra uncertainty and further affects our relative luminosity function measurement. As shown in Fig. 4.7, the error of g-band magnitude measurement of the COSMOS shape catalogue is significantly higher when a faint secondary gets close ($< 30$ kpc) to a bright primary. However, as we only consider
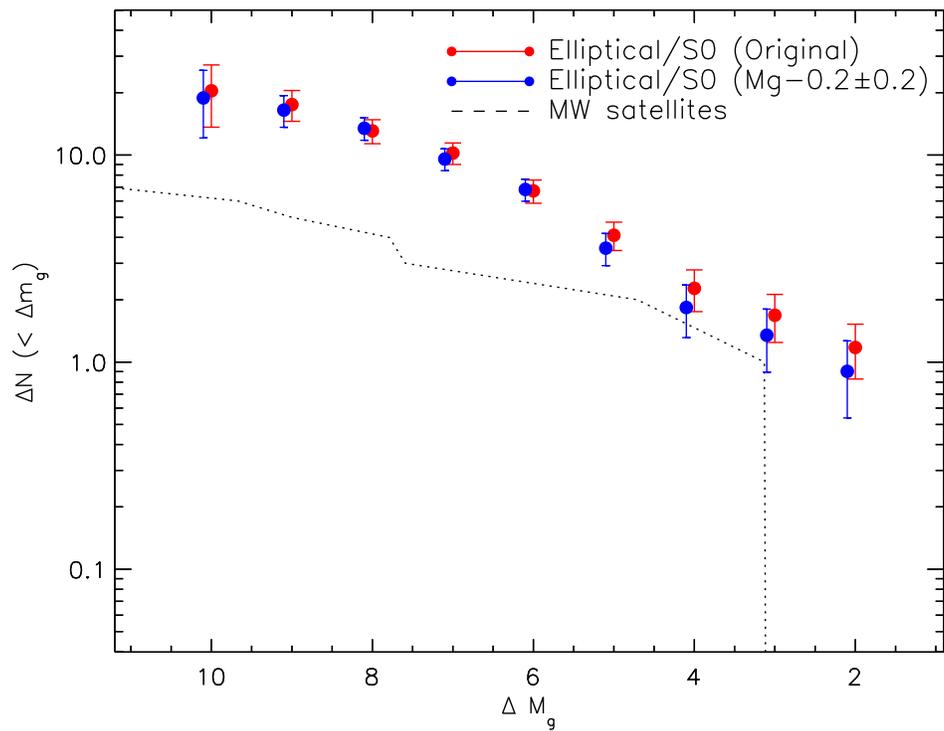
Figure 6.1: The RLF of Elliptical/S0 type primaries after adding extra Gaussian random luminosity of $\Delta M_g \sim 0.2 \pm 0.2$ (blue dots), with the original measurement included for comparison (Red dots).

the pairs with projected distance $r_p > 50$ kpc, the extra uncertainty to the photometry measurements from bright companions should be fairly small. We can expect the g-band magnitude error to be smaller than 0.15 on average, according to Fig. 4.7. In principle, we can test this effect to our RLF measurement by adding extra Gaussian random error of $\sigma_g = 0.2$ to the g-band magnitude values of our secondary sample. However, we can easily expect this effect would be comparable or smaller to the effects caused by the photometry bias of the bright early type galaxies we tested in the previous section, which is almost negligible.

## 6.4   Isolation Effects

### 6.4.1   Isolation Criteria on the Primary

The COSMOS survey covers only a small fraction of the sky ($2\ deg^2$). It contains a very limited number of nearby, very bright galaxies (only 5 galaxies with $M_g < -22.0$). Therefore, we have to choose a relatively faint magnitude cut ($M_g < -19.0$) to obtain a primary sample with reasonable size. At the same time, however, it also means the observation volume will be so crowded that it is hard to find enough primaries that are isolated from the other primaries. We tested cutting the base primary sample with various sets of isolation criteria. The relation between the number of remaining primaries and the isolation criteria is shown in Fig. 6.2. From the plot we can see, even with the criteria of cutting galaxy pairs with $|\Delta V| < 300$ km s$^{-1}$ and $r_p < 1.0$ Mpc, there would be only 105 primaries left after the isolation, which is a much smaller number size than simply using the "kick-out" technique.

Additionally, we also test another two conditional isolation criteria. One of them is similar to the method used by Guo et al. (2011a), whose essential idea is to exclude the primaries that have companions that have similar or higher luminosity. Specifically, we search for companions around each primary with the criteria of $r_p < r_{isl}$ and $|\Delta V| < 2 \times \Delta V_{isl}$, and once we find a primary has companions with $g_c - g_p < \Delta m_{isl}$, we exclude that primary. We test isolation criteria with $\Delta V_{isl} = 300$ km s$^{-1}$ and $\Delta m_{isl} = 1.0$, and plot the number of remaining isolated primary sample as a function of the isolation radius, which is shown with the blue dots in Fig. 6.2. We can see that this isolation criterion saves more primaries at given isolation radius compared to the non-conditional isolation criteria (with $r_{isl} = 1.0$ Mpc, we will have 152 primaries), as a result of ignoring the companions that are much fainter than the central galaxies.

The other isolation criterion we test is based on the "Kick-out" method (see section 2.8 for the detailed description for the "Kick-out" method). We first use half of the isolation radius as the "kick-out" radius and exclude the galaxies that are not truly primaries out from the primary list first. Afterwards, we apply the usual isolation cut to the remaining primaries, where we exclude any primary that has a companion from the current primary list and produce an isolated primary sample. Overall, the essential idea of this criterion is to isolate the primaries from other "real" primaries, ignoring the galaxies that are not truly the primaries of any system. The size of isolated primary sample produced with this method is also plotted as a function of the isolation radius, which is shown in Fig. 6.2 (red dots). We can see that this method provides the largest population of primaries with a given isolation radius. Specifically, we will have 224 primaries with $r_{isl} = 1.0$ Mpc, $\Delta V_{isl} = 300$ km s$^{-1}$ and 204 primaries with $r_{isl} = 1.2$ Mpc and the same $\Delta V_{isl}$.

In the following test, we use the isolated primary sample (204 galaxies) produced with the method that is based on the "Kick-out" method, as this method provides a relatively larger population of primaries than other methods. We are also using $r_{isl} = 1.2$ Mpc, $\Delta V_{isl} = 300$ km s$^{-1}$, to be consistent with the previous "Kick-out" radius of 0.6 Mpc. As the size of the isolated primary sample is relatively small, we do not split it further into more sub-samples and directly compare the results of using this isolated sample to the results of using the "Kick-out" primary sample.

## 6.4.2   Primary Isolation Effects on the 2PCF Measurement

We test the effect of primary isolation on the clustering measurement. Fig. 6.3 shows the comparison between the result of using the full "Kick-out" primary sample and the result of using the isolated primary for the projected cross-correlation function calculation with a same secondary sub-sample selected by $-19.0 < M_g < -15.0$. From the plot, we can see that the two detections of clustering signal have similar amplitude, while the clustering signal of the isolated primary sample seems to have a steeper slope. This may imply that the isolated primaries have a more concentrated mass distribution and they may arguably be located in more concentrated host halos. Additionally, the signal-to-noise ratio of the detection using the isolated primary sample is significantly lower (about half) than the detection of using the full "Kick-out" primary sample, which is expected as the size of the isolated primary sample is much smaller.

87
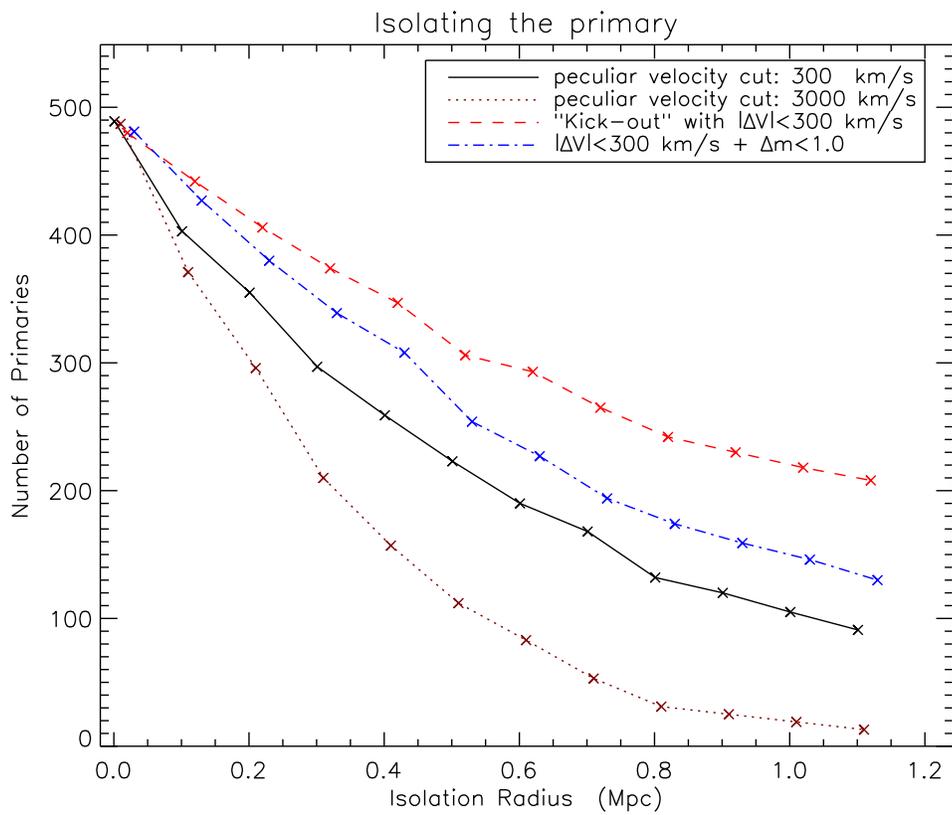
Figure 6.2: The number of primary sample with different isolation criteria.
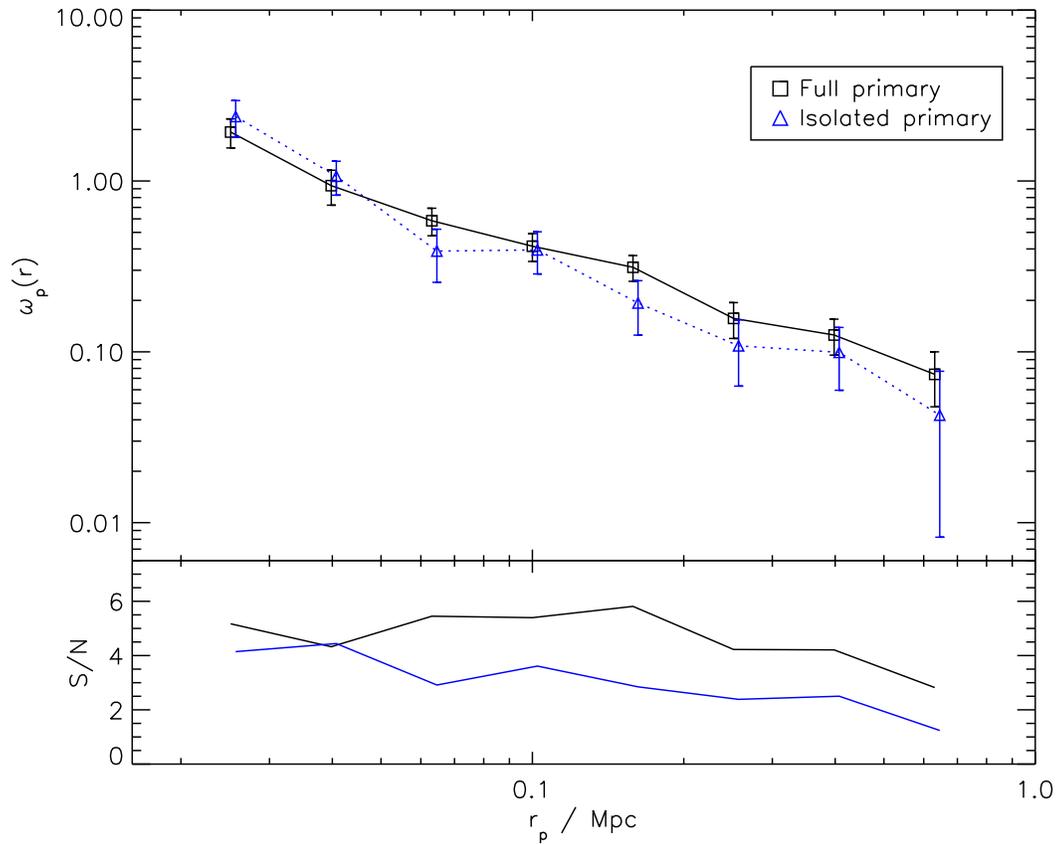
Figure 6.3: The isolation effect. The cross-projected correlation function between the full primary samples and a secondary sub-sample $-19.0 < M_g < -15.0$, and the correlation between the isolated primary sample and the same secondary sub-sample

### 6.4.3 Primary Isolation Effects on the Relative Luminosity Function

Here we also test the isolation effects on the RLF by recalculating the RLF with the isolated primary sample produced by the "Kick-out"-based method, which uses isolation radius of 1.2 Mpc, velocity difference cut of $|\Delta V| < 2 \times 300$ km s$^{-1}$ and contains 204 primaries. The result of this recalculation should give a better comparison to the previous studies on the RLF that are also using isolated primary samples, as shown in Fig. 6.4. In the plot, the measurements of the "low mass, low redshift" samples from Nierenberg et al. (2012, hereafter N12) are shown in blue open triangles and red open squares represent early and late types of primary; the measurement and upper limits from Strigari & Wechsler (2012, hereafter SW11) are shown in purple crosses and arrows, whereas the crosses and arrows indicate the measurement and the upper limits respectively; the measurement from Speller & Taylor (2014, hereafter Speller14) is shown in green diamonds. Note that the previous studies are conducted using r-band photometry while we are using g-band, which may effect the $\Delta m$ slightly. Specifically, the averaged value of the magnitude difference of $\langle r - g \rangle$ is about 0.3 for our primary sample and about 0.6 for the secondary sample. This means that those r-band-based results should be shifted slightly to the low $\Delta m$ end for a better comparison. The data points of previous studies in Fig. 6.4 are offset by $0.15 \sim 0.25$ for this reason and for clarity.

Compared to Fig. 5.5, the result of using the isolated primary sample shows an overall slightly lower RLF detection and a significantly lower signal-to-noise ratio. The lower signal-to-noise ratio is simply due to smaller sample size and the lower detection arguably indicates that isolated primaries are more likely to be located in less massive host halos and have smaller satellite populations than the primaries in groups and clusters. Additionally, the result of using the isolated primary sample has a better agreement with the previous studies at $\Delta m \sim 3 - 7$ but still slightly higher than the other results, which is arguably contributed by the better completeness of our measurement. Our measurement also appears to have a higher detection at $\Delta m \sim 2$, which is possibly due to our isolation criterion allowing relatively bright companions for each primary or may be just a statistical fluctuation. Finally, our result after applying the isolation cut is still higher than Speller & Taylor (2014) at $\Delta m \sim 7 - 10$, which is probably due to the fact that our secondary sample is more complete at the faint end, indicating the photo-zs of the COSMOS is a more powerful tool to reduce the background than the magnitude-size cut used in Speller & Taylor (2014).
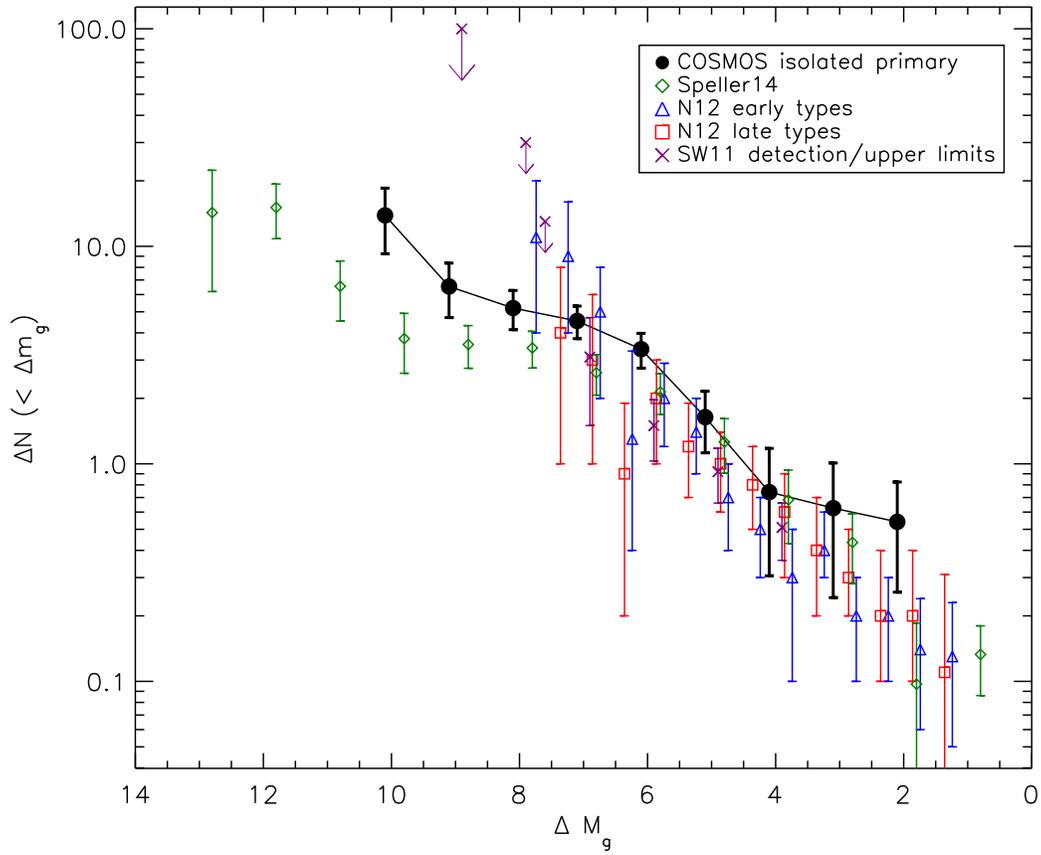
Figure 6.4: RLF using "kick-out"-isolated primary, along with four results from previous studies on the RLF for comparison.

## 6.5 Photometric Depth

From the bottom right panel of Fig. 2.8, we can see our secondary sample is relatively complete at $g^+ < 25.0$ and the number counts per magnitude bin drop dramatically after that. This incompleteness of the secondary sample at the faint end also leads to the incompleteness of our RLF measurement when the relative luminosity is larger than a certain point. In order to estimate at what relative luminosity the completeness of our RLF measurement starts to drop, we can artificially cut a part of our faint secondary sample and test how it affects the result.

As our secondary was originally cut at $g^+ < 26.0$, we applied three brighter test cuts of $g^+ < 25.0$, $g^+ < 24.5$ and $g^+ < 24.0$ as the luminosity faint limits of the secondary sample. We then used the corresponding secondary samples to recalculate the RLF with the "Kick-out" primary sample. The results of those recalculations are shown in Fig. 6.5. From the plot, we can see the measurements using the secondary samples with luminosity cuts of $g^+ < 25.0$, $g^+ < 24.5$ and $g^+ < 26.0$ basically agree with the other measurements, except for the detections at $\Delta m_g \sim 10$, where the detection using the original secondary sample with $g^+ < 26.0$ has a higher amplitude than the other two. Additionally, we found that the measurement using the secondary with luminosity cut within $g^+ < 24.5$ well matches with the rest at the relative luminosity range of $\Delta m_g \lesssim 7$, although it is lower than the other detections at the higher relative luminosity end. This implies that our RLF measurements are relatively complete down to $\Delta m_g \sim 7$ even when we cut out the secondaries that have $g^+ > 24$. Note that the secondary sample of $24.0 < g^+ < 25.0$ is relatively complete. Therefore, we can conclude that our RLF measurement using the "Kick-out" primary sample and the full secondary sample with $g^+ < 26.0$ is probably complete down to $\Delta m_g \sim 8$ .

## 6.6 Surface Brightness Limit

As much as a survey is limited by the magnitude depth, it is also, perhaps more importantly, limited by the surface brightness of the targets. This means we lose the sources that are intrinsically faint or have extended shapes that cause low surface brightness as a result. Losing low surface brightness objects leads to an underestimate of the satellite abundance, and may also introduce some extra bias. The relatively low density of the very extended sources may cause them to be more strongly affected by feedback processes (e.g. star formation and AGN activity). Or perhaps it is the other way around, and the low surface brightness of these objects are already the consequence of the environment. Nevertheless,
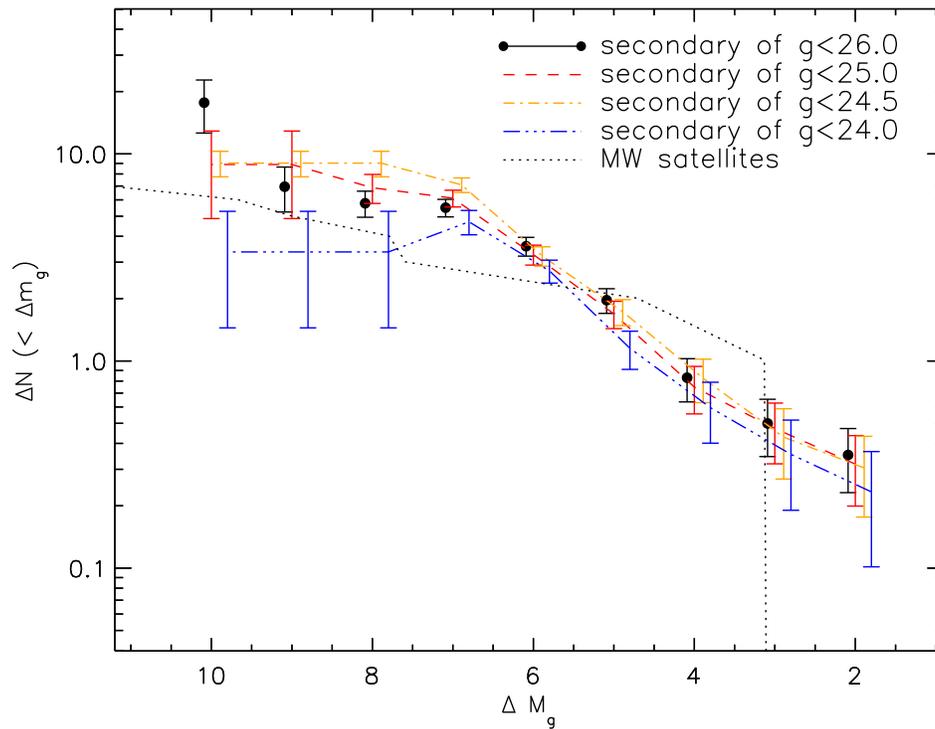
Figure 6.5: The RLFs of using the "Kick-out" primary sample and the secondary samples with the luminosity cuts of $g^+ < 25.0$, $g^+ < 24.5$ and $g^+ < 24.$, with a comparison to the result of using the secondary sample with the original luminosity cut of $g^+ < 26.0$.

these low surface brightness objects may have a different way of interacting with their primaries and therefore have a different clustering and spatial distribution with respect to the host galaxies.

## 6.7 Lensing Magnification

The lensing effect, which is caused by the gravitational potential of a massive object deflecting the light rays that are passing close by, can magnify the apparent luminosity of background sources. When we are looking at lose pairs of bright and faint galaxies, this lensing effect artificially amplify the luminosity of the faint galaxies. It will not affect our results on the clustering and spatial distribution of satellites, but it may cause bias for our RLF measurement, which depends on accurate relative luminosity information. We would like to determine how serious the lensing effect can be in our RLF measurement.

The amount of surface brightness magnification caused by a circular symmetric lens can be given by (Bartelmann & Schneider, 2001):

$$\mu = \frac{1}{|(1-\kappa)^2 - \gamma^2|} \tag{6.1}$$

where $\kappa$ is the isotropic convergence, which quantifies the ability of the lens to cause the light rays passing close by to converge; while $\gamma$ is the anisotropic tangential shear, which describes the lens' ability to distort the images behind the lens. Considering the sample isotropic case for rough estimate, then the magnification is dominated by the $\kappa$, which can be defined as (Johnson et al., 2014):

$$\kappa = \frac{\Sigma}{\Sigma_{crit}} = \frac{4\pi G}{c^2} \frac{D_L D_S}{D_{LS}} \Sigma \tag{6.2}$$

where $\Sigma$ is the surface mass density of the lens, and $\Sigma_{crit}$ is the so-called critical surface mass density;$D_L, D_S$ and $D_{LS}$ refer to the distance from the observer to the lens and source, and the distance between the lens and source. For our case, we are interested in close pairs with $|\Delta V| < 300 km/s + 2\sigma_V$ with the most remote objects sit at $z \sim 0.2$. When the both galaxies of a pair have the same redshift, obviously no lensing effects will occur. The worst case happens where the two galaxies of a pair have the largest separation along the line of sight. As we cut the sample with redshift error smaller than 0.25: $\sigma_z < 0.25$, the most extreme case would be when the redshift of the primary galaxy has the highest value ($z_p = 0.2$) and the secondary is separated from the primary by the maximum value

of the redshift error $z_s = 0.45$, which will give us (refer to section 3.1 for the calculation of luminosity distance):

$$\frac{D_L D_S}{D_{LS}} \approx 1700 \ Mpc \tag{6.3}$$

We define the surface mass density to be approximately:

$$\Sigma \approx \frac{M}{\pi R^2} \tag{6.4}$$

where the smallest radius we are concerned is about 50 kpc, where we cut the inner region pairs. The typical halo mass for our primary sample is around $10^{12} M_\odot$, while the mass inside the projected radius of 50 kpc should be much smaller (about $2 \times 10^{11} M_\odot$ if we assume a spherical NFW density profile and a virial radius of 300 kpc). We can take $10^{12} M_\odot$ as a conservatively high mass estimate. Therefore, the highest magnification effect we can get is about:

$$\kappa_{max} \leq \frac{4G}{c^2} \ 1700 Mpc \ \frac{10^{12} M_\odot}{(50 kpc)^2} \approx 0.13 \tag{6.5}$$

This corresponds to a difference of magnitude of $2.5 \ log(1/(1 - \kappa_{max})^2) \approx 0.3$.

Note that this conservative estimation of lensing magnification only applies to those secondaries that have the largest redshift errors and lie at the smallest projected separations, but far behind the most massive primaries along the line of sight, which are a very small part of our secondary sample. This means the lensing effect on the real satellites, which are close to the primaries along the line of sight, is negligible. On the other hand, the lensing magnification slightly brighten some very faint background galaxies that may be not be detected otherwise, which may cause a small over-density of faint background galaxies around the primaries and thus artificially increase satellite abundance by a small amount at the faint end. Overall, the effect from lensing magnification is small and probably negligible, but it may worth further analysis and correction for studies at higher redshift, where the lensing effect will be more significant.

# Chapter 7

# CS82

## 7.1 Introduction to the Survey

The Canada-France-Hawaii Telescope Stripe 82 Survey (CS82) is an i-band follow-up survey of the SDSS. It covers the SDSS equatorial Stripe 82 region, which originally covers more than 200 deg$^2$ with a high density of spectroscopic redshifts (it has over 100,000 redshift measurements so far and the ongoing surveys such as SDSS-III BOSS and WiggleZ are adding more new spectra (Shan et al., 2014)). The CS82 survey contains a total of 173 tiles (165 tiles CFHTStripe-82 and 8 CFHTLS Wide tiles). Each CS82 tile was obtained in four dithered observations with a total exposure time of $4 \times 410s$, each resulting in a $5\sigma$ limiting magnitude in about 2 arcsec diameter aperture of about $i_{AB} = 24.0$ (Li et al., 2014) (Shan et al., 2014). Each stripe also covers about 1 deg$^2$ and thus the CS82 has final sky coverage of 173 deg$^2$, which drops to an effective sky coverage of $\sim 124$ deg$^2$ (Shan et al., 2014). The stripes are mostly distributed between $-40.0 < RA < 45.0$, $-1.0 < DEC < +1.0$, as shown in Fig. 7.1.

The CS82 survey has a fairly good i-band depth with excellent seeing conditions. The i-band depth reaches $i_{AB} \sim 24.0$ for point sources and $i_{AB} \sim 23.5$ for galaxies, while the typical seeing is between 0.4$''$ and 0.8$''$, with a median of 0.59$''$ (Leauthaud et al., 2012; Li et al., 2014).

Figure 7.1: The distribution of CS82 field, with the names of each stripe. This figure is taken from a talk by Leauthaud et al. (2012).

## 7.2 A Preliminary Look at the Data

Before working with the final reduced data set currently under production for the survey, we have used public-release data[1] to estimate the potential satellite clustering signal in CS82. We first looked at one of the CS82 stripes that is named "S86p3m", which is located at $3.45 \leq RA \leq 4.30$ and $-0.99 \leq DEC \leq -0.02$ (it can be found at the bottom left corner of Fig. 7.1). The catalogue of "S86p3m" provided by the Megapipe contains 152,890 objects, which drops to 104,044 objects after we applied a galaxy-star cut to exclude most of the point sources (stars) and other image artifacts. The catalogue does not provide redshift information, which is however required for the primary galaxies for the subsequent analysis. We built the primary sample with the following steps:

1. we searched NED to obtain the redshift information for the bright galaxies in the "S86p3m" field. With the criteria of $z \leq 0.15$ and "Object Type"="Galaxy", the NED returned 108 galaxies with reliable spectroscopic redshifts;

---

[1]Download through the Megapipe, see Megapipe: http://www2.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/megapipe/

2. We calculated their g-band absolute magnitude $M_g$ with their g-band apparent magnitude and their redshifts. We then applied a g-band absolute magnitude cut of $M_g < 19.5$, which left us 74 galaxies;

3. We applied an isolation criterion: we excluded all the close bright galaxy pairs that have small velocity difference of $|\Delta V| < 300$ km s$^{-1}$ and a small projected separation of $r_p < 500$ kpc, which left us 39 isolated primary galaxies.

The remaining 39 galaxies will be used as the primary sample and the reduced catalogue from the field "S86p3m" (104,044 galaxies) will be used as the secondary sample for our subsequent analysis.

## 7.3 Preliminary Results

We calculated the RLF using the NED-based primary sample, along with the secondary sample based on the public data of the field "S86p3m". The method is very similar to the method described in section 5.1, but we did not apply any cuts on the background as we do not have any redshift information for the secondary sample. Instead, we simply calculate the projected separation between the secondary galaxies and a given primary, assuming all the secondaries are at the same redshift as that primary. We then used the projected distance range of 50 – 400 kpc for pair counting and used the projected distance range of 600 – 800 kpc for the background estimation. The result is shown in Fig. 7.2. Overall, it appears similar to our previous results using the data from the COSMOS survey and has a surprisingly good signal-to-noise ratio of 2.3 at $\Delta m = 9$.

The preliminary result of the RLF using the data from a single field ("s86p3m") may be showing the great potential of the CS82 survey. If the data of the other 172 fields provide measurements of similar quality, the signal-to-noise ratio of the RLF for the whole entire survey can reach more than 20, which would allow us to divide the sample into quite a few sub-samples and study the effects of the galaxy properties in an extremely detailed way. However, this expectation is over-optimistic as this measurement using a single field is very likely to a lucky statistical incidence; we also tested the RLF calculation with the data of another field "S86p5m", whose detection only has a signal-to-noise ratio much less than 1 at $\Delta m = 8$ and does not reach the depth of $\Delta m \sim 9$. In addition, without any redshift information for the secondary sample, it will inevitably introduce more systematics and statistical uncertainties. For instance, we found that the RLF detection is sensitive to the pair counting range. Once the counting range is extended to projected distance of 50 – 500 kpc, we found that the signal-to-noise ratio will drop to 1.2 at $\Delta m = 9$.
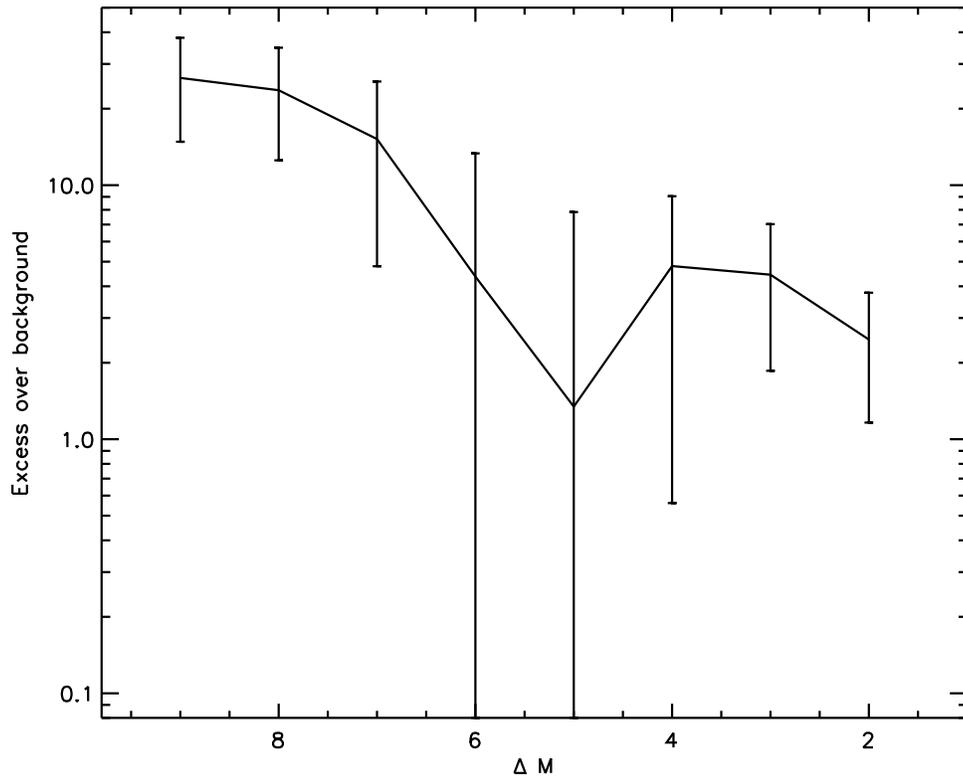
Figure 7.2: The RLF measurement of using the stripe "S86p3m" from the CS82 public data and the NED-based isolated primary sample.

All things considered, the CS82 is still likely to be of great potential for measuring satellite populations, and future work can be done to improve the measurements. For instance, the final reduced catalogues with better completeness than the public data will surely improve the analysis. We can also use the redshift information from other sources for a part of our secondary sample. Furthermore, we can try other methods to pre-reduce the background (e.g. the magnitude-size cut used by Speller & Taylor (2014)).

# Chapter 8

# Conclusion

We have used data from the COSMOS survey to search for dwarf satellites around nearby bright galaxies (z < 0.2). We have detected these satellite populations statistically based on their clustering around bright galaxies. We have also studied both the radial and angular distributions, as well as the relative luminosity function of the satellites, as a function of central properties. Our conclusions are summarized as follows:

- We measured the projected two-point cross-correlation function (2PCF) between the bright primary galaxies and the faint secondary sample. We found the high quality photo-zs of the COSMOS data can greatly improve the clustering measurement. We also found the correlations are strongly dependent on both the primary luminosity and secondary luminosity. The brighter primary and secondary galaxies show stronger correlations than the faint ones. The secondary sample clusters more strongly around Red primary galaxies than around Blue primaries.

- We measured the radial distribution of satellite galaxies around the bright primary galaxies, using the "Kick-out" primary sample and six primary sub-samples based on that sample, along with the full secondary sample. We detected an excess over the background of 7.9 ± 0.5 galaxies per central galaxy within 600 kpc projected radius from the central galaxies, using the full "Kick-out" primary sample. Secondly we found that the radial distributions of satellites around all the sub-samples of primaries are relatively flat within the inner radial bins and drop to zero in the outer regions, following a pattern similar to that predicted for an NFW density profile. This implies the satellites may, at a certain level, trace the dark matter substructures inside individual host halos. Third, we found that the radial satellite number density

has a strong dependence on the primary luminosity. Brighter galaxies have a higher number of satellites per radial bin and their satellite distributions extend to larger radius than the faint primaries. This may imply that the brighter primaries are more likely to be located in halos of larger virial radii (thus more massive) than the faint primaries, or may imply that the bright primaries are more likely to be associated with massive groups rather than isolated halos. Lastly, we found that the radial satellite number density has a strong dependence on the morphological (SED) type of the primary. Early-type primaries have a much larger population of satellites on average than the late-type primaries, and their satellite distribution also extends to larger radii. There are several possible explanations for this difference. It could be that early-type primaries are more likely to be located in more massive individual halos or are associated with massive halo groups. It could also imply different luminosity-mass relations for early-type and late-type bright galaxies, or perhaps it implies different baryonic physics in the two types of galaxies. For further investigation of these possibilities, we may need to repeat the measurement with a larger primary sample and focus on the isolated primaries. It would also be interesting to study the dependence of radial satellite number density on the stellar mass and see whether the discrepancy between the two types of galaxies will go away when the primary samples are selected in a restricted stellar mass range.

- We studied the alignment of the satellite-central position angle with the orientation of the central galaxies, trying to detect the so-called Holmberg effect. First of all, we found that the alignment has a dependence on the projected radius from the central galaxy. When the alignment measurements were conducted separately for the inner regions and the outer regions, we found that a part of the systems showed convincing alignment patterns. This already arguably shows that the orientation of the central galaxies are related to the asymmetry of the host halo in certain ways. Specifically, in the inner regions around all types of primary galaxies, the distribution of the satellites were found to have no preferred alignment with the major axis of the central galaxies, while in the outer regions, the satellites of some types of primary samples were found to have a preferred alignment with the major axis, including the satellites of the faint primary sample, the early-type primary sample and the red primary sample. In the case of late-type galaxies, no significant anisotropy was found in the satellite distribution at any radius. These alignment patterns may contain information about the formation history of different types of galaxies.

- We determined the Relative Luminosity Function of the satellites for 392 "Kick-out" primary galaxies, along with the full secondary sample. First of all, we detected

$17.7 \pm 5.1$ satellites per primary down to $\Delta m = 10$ and detected $5.5 \pm 0.5$ satellites per primary at $\Delta m = 7$, where we obtained the best signal-to-noise ratio. Second, we compared our measurement of the MW satellite. We found that overall, our measured satellite abundance exceeds that of the MW satellites (at $\Delta m \sim 5 - 10$), which may be because our primary sample includes elliptical galaxies and galaxies in massive groups. We also found that our measured satellite abundance is significantly lower than that of the MW satellites at the relatively low luminosity range of $\Delta m \sim 3 - 5$, which may suggest the LMC-MW combination is a statistically rare case. Third, we also looked into the dependence of the RLF on the properties of the primary galaxies. The early-type galaxies were found to have a significantly larger population of satellites than the late-type primaries over the whole magnitude range. We discussed the possible systematics in our measurement and concluded those systematics are not significant enough to explain the discrepancy between the satellite abundances of the different types of primaries. The discrepancy may indicate different host halo properties, different luminosity-mass relations and possible different baryonic effects between the two types of primaries. For further investigation of these possibilities, we may need to repeat the measurement with a larger primary sample and focus on the isolated primaries. This will help to constrain the host halo properties. In addition, we can use stellar mass information (available for the COSMOS data) to test the possible luminosity-mass effects.

- It would also be interesting to study the difference between the satellite populations of different properties (e.g. color, surface brightness, etc.). Satellites with different infall history and subject to different environment effects may show different properties as a consequence. For instance, a lot current semi-analytic studies try to solve the MSP by adding strong feedback. However, this causes an over-prediction of the population of red satellites as a result (Guo et al., 2011b; Liu et al., 2011; Wang et al., 2014). In addition, assuming feedback has strong effects on the star formation in subhalos, we can thus expect the red satellites, which are more likely under the effects of the feedback, would have different angular alignments and radial profiles from the blue satellites. Overall, studies of the effect of the secondary properties on the satellite-primary clustering, satellite spatial distribution and satellite abundance can provide more useful implications for galaxy formation and the structure formation.

Finally, we introduced preliminary work on satellite populations in the CS82 survey. We showed that the CS82 survey can provide a primary sample of much larger size, which probably will allow us to measure the satellite abundance with much higher signal-to-noise ratio. Furthermore, the larger size of the primary sample in CS82 may also allow

us to study dependence of satellite abundance on the primary galaxy properties in a more detailed way.

# References

Azzaro M., Patiri S. G., Prada F., Zentner A. R., 2007, MNRAS, 376, L43

Babcock H. W., 1939, Lick Observatory Bulletin, 19, 41

Barrow J. D., Bhavsar S. P., Sonoda D. H., 1984, MNRAS, 210, 19P

Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291

Belokurov V. et al., 2008, ApJ, 686, L83

Belokurov V. et al., 2010, ApJ, 712, L103

Belokurov V. et al., 2009, MNRAS, 397, 1748

Benson A. J., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2002, MNRAS, 333, 177

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40

Brainerd T. G., 2005, ApJ, 628, L101

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

Capak P. et al., 2007, ApJS, 172, 99

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Craig M. W., Davis M., 2001, New Astronomy, 6, 425

Croton D. J. et al., 2006, MNRAS, 365, 11

Davis M., Geller M. J., 1976, ApJ, 208, 13

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

Diemand J., Kuhlen M., Madau P., 2007, ApJ, 667, 859

Dressler A., 1980, ApJ, 236, 351

Frenk C. S., White S. D. M., 2012, Annalen der Physik, 524, 507

Grebel E. K., Dolphin A. E., Guhathakurta P., 2000, in Astronomische Gesellschaft Meeting Abstracts, Vol. 17, Astronomische Gesellschaft Meeting Abstracts, Schielicke R. E., ed., p. 61

Guo Q., Cole S., Eke V., Frenk C., 2011a, MNRAS, 417, 370

Guo Q., Cole S., Eke V., Frenk C., 2012, MNRAS, 427, 428

Guo Q., Cole S., Eke V., Frenk C., Helly J., 2013, MNRAS, 434, 1838

Guo Q. et al., 2011b, MNRAS, 413, 101

Hamilton A. J. S., 1993, ApJ, 417, 19

Hawley D. L., Peebles P. J. E., 1975, AJ, 80, 477

He Y. Q., Xia X. Y., Hao C. N., Jing Y. P., Mao S., Li C., 2013, ApJ, 773, 37

Hewett P. C., 1982, MNRAS, 201, 867

Hinshaw G. et al., 2013, The Astrophysical Journal Supplement Series, 208, 19

Hogg D. W., 1999, ArXiv Astrophysics e-prints

Holmberg E., 1969, Arkiv for Astronomi, 5, 305

Ilbert O. et al., 2009, ApJ, 690, 1236

Irwin M. J. et al., 2007, ApJ, 656, L13

106

Jiang C. Y., Jing Y. P., Li C., 2012, ApJ, 760, 16

Jing Y. P., Suto Y., 2002, ApJ, 574, 538

Johnson T. L., Sharon K., Bayliss M. B., Gladders M. D., Coe D., Ebeling H., 2014, ArXiv e-prints

Karachentsev I. D., Kashibadze O. G., 2006, Astrophysics, 49, 3

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kerscher M., Szapudi I., Szalay A. S., 2000, ApJ, 535, L13

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

Klypin A., Zhao H., Somerville R. S., 2002, ApJ, 573, 597

Knebe A., Gill S. P. D., Gibson B. K., Lewis G. F., Ibata R. A., Dopita M. A., 2004, ApJ, 603, 7

Koposov S., Belokurov V., 2008, Observational Constraints on the "Missing Satellite" Problem from SDSS, Koribalski B. S., Jerjen H., eds., p. 195

Koposov S. E., Yoo J., Rix H.-W., Weinberg D. H., Macciò A. V., Escudé J. M., 2009, ApJ, 696, 2179

Kravtsov A., 2010, Advances in Astronomy, 2010

Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482

Kroupa P., Theis C., Boily C. M., 2005, A&A, 431, 517

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Leauthaud A. et al., 2007, ApJS, 172, 219

Leauthaud A. et al., 2012, in American Astronomical Society Meeting Abstracts, Vol. 219, American Astronomical Society Meeting Abstracts #219, p. 324.06

Li C., Gadotti D. A., Mao S., Kauffmann G., 2009, MNRAS, 397, 726

Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, MNRAS, 368, 21

Li R. et al., 2014, MNRAS, 438, 2864

Lilly S. J. et al., 2007, ApJS, 172, 70

Liu C., Hu J., Newberg H., Zhao Y., 2008, A&A, 477, 139

Liu L., Gerke B. F., Wechsler R. H., Behroozi P. S., Busha M. T., 2011, ApJ, 733, 62

Lovell M. R., Eke V. R., Frenk C. S., Jenkins A., 2011, MNRAS, 413, 3013

Lynden-Bell D., 1982, in Astrophysical Cosmology Proceedings, Brueck H. A., Coyne G. V., Longair M. S., eds., pp. 85–108

MacGillivray H. T., Dodd R. J., McNally B. V., Corwin, Jr. H. G., 1982, MNRAS, 198, 605

Martin N. F., de Jong J. T. A., Rix H.-W., 2008, ApJ, 684, 1075

Martin N. F., Ibata R. A., Irwin M. J., Chapman S., Lewis G. F., Ferguson A. M. N., Tanvir N., McConnachie A. W., 2006, MNRAS, 371, 1983

Mateo M., Olszewski E. W., Walker M. G., 2008, ApJ, 675, 201

McConnachie A. W., 2012, AJ, 144, 4

Moore B., Gelato S., Jenkins A., Pearce F. R., Quilis V., 2000, ApJ, 535, L21

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19

Nair P. B., Abraham R. G., 2010, ApJS, 186, 427

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Nierenberg A. M., Auger M. W., Treu T., Marshall P. J., Fassnacht C. D., 2011, ApJ, 731, 44

Nierenberg A. M., Auger M. W., Treu T., Marshall P. J., Fassnacht C. D., Busha M. T., 2012, ApJ, 752, 99

Okamoto T., Frenk C. S., Jenkins A., Theuns T., 2010, MNRAS, 406, 208

Ostriker J. P., Peebles P. J. E., 1973, ApJ, 186, 467

Peebles P. J. E., 1980, The large-scale structure of the universe

Peebles P. J. E., Hauser M. G., 1974, ApJS, 28, 19

Planck Collaboration et al., 2013, ArXiv e-prints

Ribas I., Jordi C., Vilardell F., Fitzpatrick E. L., Hilditch R. W., Guinan E. F., 2005, ApJ, 635, L37

Rix H.-W. et al., 2004, ApJS, 152, 163

Robotham A. S. G. et al., 2012, MNRAS, 424, 1448

Sales L., Lambas D. G., 2004, MNRAS, 348, 1236

Scoville N. et al., 2007, ApJS, 172, 1

Shan H. Y. et al., 2014, MNRAS, 442, 2534

Sharp N. A., Lin D. N. C., White S. D. M., 1979, MNRAS, 187, 287

Simon J. D., Geha M., 2007, ApJ, 670, 313

Speller R., Taylor J. E., 2014, ApJ, 788, 188

Spergel D. N., Steinhardt P. J., 2000, Physical Review Letters, 84, 3760

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Springel V. et al., 2008, MNRAS, 391, 1685

Strigari L. E., Wechsler R. H., 2012, ApJ, 749, 75

Taylor J. E., Babul A., 2004, MNRAS, 348, 811

Taylor J. E., Silk J., Babul A., 2004, in IAU Symposium, Vol. 220, Dark Matter in Galaxies, Ryder S., Pisano D., Walker M., Freeman K., eds., p. 91

The Fermi-LAT Collaboration et al., 2013, ArXiv e-prints

Tolstoy E., Hill V., Tosi M., 2009, ARA&A, 47, 371

Tormen G., Bouchet F. R., White S. D. M., 1997, MNRAS, 286, 865

van den Bergh S., 2000a, The Galaxies of the Local Group. Cambridge

van den Bergh S., 2000b, PASP, 112, 529

Véron-Cetty M.-P., Véron P., 2006, A&A, 455, 773

Walsh S. M., Willman B., Jerjen H., 2009, AJ, 137, 450

Wang W., Jing Y. P., Li C., Okumura T., Han J., 2011, ApJ, 734, 88

Wang W., Sales L. V., Henriques B. M. B., White S. D. M., 2014, MNRAS, 442, 1363

Wang W., White S. D. M., 2012, MNRAS, 424, 2574

Wang Y., Park C., Yang X., Choi Y.-Y., Chen X., 2009, ApJ, 703, 951

Watkins L. L. et al., 2009, MNRAS, 398, 1757

Weisz D. R. et al., 2011, ApJ, 743, 8

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Willman B. et al., 2005, ApJ, 626, L85

Willman B., Governato F., Dalcanton J. J., Reed D., Quinn T., 2004, MNRAS, 353, 639

Wong A. W. C., Taylor J. E., 2012, ApJ, 757, 102

Yang X., van den Bosch F. C., Mo H. J., Mao S., Kang X., Weinmann S. M., Guo Y., Jing Y. P., 2006, MNRAS, 369, 1293

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

Yoshida N., Springel V., White S. D. M., Tormen G., 2000, ApJ, 544, L87

Zaritsky D., Olszewski E. W., Schommer R. A., Peterson R. C., Aaronson M., 1989, ApJ, 345, 759

Zaritsky D., Smith R., Frenk C. S., White S. D. M., 1997, ApJ, 478, L53

Zavala J., Jing Y. P., Faltenbacher A., Yepes G., Hoffman Y., Gottlöber S., Catinella B., 2009, ApJ, 700, 1779

Zentner A. R., Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2005, ApJ, 629, 219

Zhang Y., Yang X., Wang H., Wang L., Mo H. J., van den Bosch F. C., 2013, ApJ, 779, 160

Zucker D. B. et al., 2006, ApJ, 643, L103

Zucker D. B. et al., 2004, ApJ, 612, L121

Zucker D. B. et al., 2007, ApJ, 659, L21