The Galaxy–Halo Connection Across Cosmic History

Shogo Ishikawa

Department of Astronomical Science
School of Physical Sciences
SOKENDAI (The Graduate University for Advanced Studies)

ACKNOWLEDGMENTS

I am deeply grateful for my supervisor, Professor Nobunari Kashikawa, for his continuous supports of my research and the Ph.D. life in NAOJ. He devoted his precious time to discuss and improve my work, and gave me useful comments and inspirations whenever I got into difficulties. I have learnt a lot of things from him, and all of his teachings, remarks, and the attitude and the passion for research are my valuable assets in my life.

I would like to show my deep appreciation for Professor Tadayuki Kodama and Dr. Takashi Hamana, who are also my supervisors in SOKENDAI/NAOJ. They supported my research and gave useful comments from various aspects to improve my study and papers.

I am grateful for my colleagues and the collaborators of this research, Dr. Jun Toshikawa, Mr. Yoshifumi Ishizaki, Mr. Masafusa Onoue, Mr. Hisakazu Uchiyama, Dr. Masayuki Tanaka, Dr. Yuu Niino, and Dr. Kohei Ichikawa. I also thank people in Optical and Infrared Astronomy Division in NAOJ, people in Department of Astronomical Science in SOKENDAI, Subaru Telescope, and my friends.

Finally, I would like to show my deep and sincere gratitude to my parents, Noboru Ishikawa and Yuko Ishikawa, my sister, Mayu Moriyama (Ishikawa), and my grandparents, Tetsuzo Suzuki and Sumi Suzuki, for their understanding, helps, supports, and continuous encouragements to me.

ABSTRACT

In this thesis, I discuss the relationship between observable galaxies and their underlying invisible host dark haloes over the global cosmic time at $z = 0-5$ based upon the precision clustering analyses. At each redshift, I collect a large number of galaxy samples enough to obtain high-quality two-point auto angular correlation function (ACF), which is a useful estimator to measure the strength of galaxy clustering quantitatively. Furthermore, the high accurate ACFs enable to carry out more precise clustering analyses using a "halo occupation distribution (HOD)" formalism, which characterizes the occupation number and the distribution of galaxies within the virialized dark haloes as a function of the dark halo mass.

At low-redshift ($z < 1.4$) Universe, I collect numerous galaxy samples using the data of the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP) Wide-layer. By using the photometric images with the extremely wide survey area taken by the Subaru/Hyper Suprime-Cam, I collect a large number of massive galaxies ($M_\star > 10^{11} M_\odot$) up to $z = 1.4$ as well as the less massive galaxies down to the limiting magnitude of $i < 25.5$ mag. Deep optical multi-wavelength images enable to apply a spectral energy distribution (SED) fitting technique to derive the stellar mass and the photometric redshift of each galaxy sample with high accuracies. I divide galaxy samples into subsamples according to their photometric redshift and the stellar mass in order to investigate the dependence of the clustering properties upon the physical properties by computing the ACFs of each subsample. Thanks to the numerous galaxy samples, derived ACFs show excellent S/N ratios, and the HOD model is able to reproduce the observed ACFs over all magnitude ranges. The mean halo masses of low-z dark haloes increase monotonically with increasing the stellar-mass limits and the cosmic time, indicating the evidence that massive haloes host massive galaxies and the halo growth via halo mergers. Our observed stellar-to-halo mass ratios (SHMRs) are in good agreement with the theoretical model of Behroozi et al. (2013b), although the slope at high-mass end shows a deviation from the model prediction.

The epoch at $z \sim 2$, known as the "cosmic noon", is important era to study formation and evolution of galaxy; however, it has been very difficult to construct a large number of galaxy samples, which is essential to carry out the precision clustering analyses. By combining our original data with publicly available data, I apply BzK/gzK selection to the wide-field imaging data over 5.2 deg^2, which is the largest-ever survey area of the BzK/gzK study. I select 41,112 sgzK galaxies down to $K_{\rm AB} < 23.0$. The ACFs based upon the largest sgzK sample with high quality enable to perform the HOD analysis. The mean halo mass and the HOD mass parameters are found to increase monotonically with increasing K-band magnitude, suggesting that more luminous sgzK galaxies reside in more massive dark haloes. Following the dark halo mass evolution using the extended Press–Schechter formalism and the number evolution of satellite galaxies in a dark halo, I find that faint Lyman break galaxies at $z \sim 4$ could evolve into the faintest sgzK galaxies ($22.0 \leq K_{\rm AB} \leq 23.0$) at $z \sim 2$ and into the Milky-Way-like galaxies or elliptical galaxies in the local Universe, whereas the most luminous sgzK galaxies ($18.0 \leq K_{\rm AB} \leq 21.0$) could evolve into the most massive systems in the local Universe. In addition, the SHMRs of the sgzK galaxies are found to be consistent with the theoretical prediction, as well as the previous observational results at $z \sim 2$.

At high-redshift ($z > 3$) Universe, I carry out the HOD analysis using the dropout galaxy samples at $z \sim 3$, 4, and 5 using the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) Deep Field. Deep- and wide-field images of the CFHTLS Deep Survey enable to obtain sufficiently accurate ACFs to apply the HOD analysis. The mean dark halo masses, $\langle M_{\rm h} \rangle = 10^{11.7-12.8} h^{-1} M_\odot$, is found to

increase with the stellar-mass limit. The threshold dark halo mass to possess a central galaxy within dark haloes, $M_{\rm min}$, show almost identical behavior to their stellar mass, compared to low-z with the same stellar-mass limit, whereas the threshold dark halo mass to possess one satellite galaxy, M_1, show systematically higher values at $z = 3 - 5$ than those of low-z over the entire stellar-mass range. The satellite fractions, which represent the percentages of the satellite galaxy with respect to the total samples, are found to be significantly small compared to $z \sim 2$, indicating the drastically increase of satellite galaxies from $z = 3 - 2$. Along with the high M_1 values and the low satellite fractions of high-z galaxies, satellite galaxies form inefficiently within dark haloes at high-z. Assuming the main-sequence of star-forming galaxies, I computed the SHMRs, which is found to agree with those derived using the SED fitting method. The observed SHMRs are highly consistent with the theoretical predictions based on the abundance-matching method within 1σ confidence intervals. The pivot halo mass, $M_{\rm h}^{\rm pivot}$, which is the mass at the most efficient star-formation in galaxies, is, for the first time, found to increase with cosmic time at $z > 3$, and the SHMRs at $M_{\rm h}^{\rm pivot}$ show little evolution, indicating that mass growth rates of stellar components and dark haloes are comparable at $3 < z < 5$.

From the precision clustering/HOD analyses, I firstly achieve to trace the redshift evolution of $M_{\rm h}^{\rm pivot}$ across cosmic history by observations. The evolution of $M_{\rm h}^{\rm pivot}$ is follows the theoretical prediction of Behroozi et al. (2013b): $M_{\rm h}^{\rm pivot}$ increases with decreasing the redshift at $z = 3 - 5$ and has a peak at $z \sim 2$, which is known as the highest star-formation era, and then decrease the star-formation efficiency with increasing the cosmic time at $z = 2 - 0$. This result of the redshift evolution of star-formation efficiencies is consistent with the theoretical models (e.g., Behroozi et al. 2013a,b; Moster et al. 2013) as well as the observational results (e.g., Hopkins & Beacom 2006). Moreover, I find that $M_{\rm h}^{\rm pivot}$ to be almost unchanged around $\log(M_{\rm h}^{\rm pivot}/M_\odot) = 12.1 \pm 0.2$ over cosmic time at $0 < z < 5$ and conclude that galaxy formation is ubiquitously most efficient near a halo mass of $\langle M_{\rm h} \rangle \sim 10^{12} M_\odot$ over cosmic time.

LIST OF ABBREVIATIONS

ABBREVIATION	DESCRIPTION
ACF	Two-Point Angular Auto Correlation Function
BAO	Baryon Acoustic Oscillation
CDM	Cold Dark Matter
CFHTLS	Canada–France–Hawaii Telescope Legacy Survey
CLF	Conditional Luminosity Function
HOD	Halo Occupation Distribution
HSC	Hyper Suprime-Cam
LBG	Lyman break galaxy
NIR	Near Infra-Red
SED	Spectral Energy Distribution
SHMR	Stellar-to-Halo Mass Ratio
SMOKA	Subaru–Mitaka–Okayama–Kiso Archive
SPH	Smoothed Particle Hydrodynamics
WIRDS	WIRCam Deep Survey

CONTENTS

ACKNOWLEDGMENTS i

ABSTRACT ii

LIST OF ABBREVIATIONS iv

1 INTRODUCTION 1

1.1 Dark Matter and Structure Formation . . . 1

1.1.1 Large-scale structure of the Universe . . . 1

1.1.2 Dark side of the Universe . . . 2

1.2 Galaxy–Halo Connection . . . 5

1.2.1 Galaxy clustering . . . 6

1.2.2 Halo occupation distribution model . . . 7

1.2.3 Interpretation of observed galaxy clustering using the HOD model . . . 7

1.2.4 Beyond the "classical" HOD model . . . 10

1.2.5 Stellar-to-halo mass ratio . . . 11

1.3 This Thesis . . . 15

2 A THEORETICAL FRAMEWORK OF GALAXY CLUSTERING ANALYSIS 17

2.1 Galaxy Clustering Statistics . . . 17

2.1.1 Two-point angular auto correlation function . . . 17

2.1.2 Error estimation of the ACF . . . 20

2.1.3 Real-space galaxy clustering . . . 21

2.1.4 Dark halo mass estimation from large-scale galaxy clustering . . . 23

2.2 Analytical Dark Halo Models . . . 25

2.2.1 Halo mass function . . . 28

2.2.2 Density profile of the dark halo . . . 29

2.2.3 Large-scale halo bias . . . 31

2.3 The Halo Occupation Distribution Formalism . . . 33

2.3.1 Overview of the HOD model . . . 33

2.3.2 Galaxy distribution within the dark halo . . . 34

2.3.3 Galaxy power spectrum and correlation function . . . 41
2.4 HOD Analysis . . . 42

3 GALAXY–HALO CONNECTION IN LOW-REDSHIFT UNIVERSE 44
3.1 Overview . . . 44
3.1.1 Clustering analyses of spectroscopically observed galaxies . . . 44
3.1.2 SED fitting technique and photometric redshift . . . 44
3.1.3 Clustering analyses using photometric redshifts . . . 46
3.1.4 Motivation of this study . . . 47
3.2 Details of Data and Samples . . . 48
3.2.1 Data description . . . 48
3.2.2 Sample selection . . . 48
3.3 Clustering Analysis . . . 57
3.3.1 Angular correlation functions of HSC galaxy samples . . . 57
3.3.2 The HOD analysis . . . 57
3.4 Results . . . 62
3.4.1 Mean halo mass . . . 62
3.4.2 Satellite fraction . . . 62
3.4.3 Stellar-to-halo mass ratio . . . 64

4 GALAXY–HALO CONNECTION IN MID-REDSHIFT UNIVERSE 70
4.1 Overview . . . 70
4.1.1 Characteristics of mid-redshift . . . 70
4.1.2 Clustering analysis in mid-redshift Universe . . . 70
4.1.3 Motivation of this study . . . 71
4.2 Photometric Data and Sample Selection . . . 72
4.2.1 Photometric data . . . 72
4.2.2 K-selected catalogue . . . 73
4.2.3 Band corrections . . . 74
4.2.4 gzK selection method . . . 78
4.3 Clustering Analysis . . . 83
4.3.1 Angular correlation function . . . 83

4.3.2 Subsamples . . . 83
4.3.3 Redshift distributions and completeness . . . 86
4.4 Clustering Properties of sgzK Galaxies . . . 88
4.4.1 ACFs of cumulatively resampled sgzK galaxies . . . 88
4.4.2 Clustering in real space . . . 90
4.4.3 Differential luminosity subsample . . . 92
4.4.4 Dark halo mass estimation by the large-scale clustering of sgzK galaxies . . . 94
4.5 HOD Analysis . . . 96
4.5.1 Procedure of the HOD analysis . . . 96
4.5.2 Results of the HOD analysis . . . 97
4.5.3 Dark halo masses from the HOD model . . . 100
4.6 Discussion . . . 104
4.6.1 The luminosity dependence of the HOD parameters . . . 104
4.6.2 The evolution of sgzKs by tracing the halo mass evolution . . . 106
4.6.3 Galaxy evolution by tracing the number of satellite galaxies . . . 109
4.6.4 Stellar-to-halo mass ratio . . . 111

5 GALAXY–HALO CONNECTION IN HIGH-REDSHIFT UNIVERSE 113
5.1 Overview . . . 113
5.1.1 Clustering analysis in high-redshift Universe . . . 113
5.1.2 Motivation of this study . . . 114
5.2 Data and Sample Selection . . . 115
5.2.1 Optical data . . . 115
5.2.2 NIR data . . . 117
5.2.3 Photometry and sample selection . . . 118
5.3 Stellar Mass Estimation . . . 118
5.3.1 SED Fitting . . . 122
5.3.2 Main-sequence of star-forming galaxies . . . 124
5.3.3 Consistency of the stellar mass estimation between the SED fitting and the MS relation . . . 125
5.4 Clustering Analysis . . . 125

5.5 HOD Parameters . . . 133
5.5.1 HOD mass parameters . . . 133
5.5.2 Mean halo masses . . . 136
5.5.3 Satellite fractions . . . 140
5.6 Stellar-to-Halo Mass Ratios of High-Redshift Galaxies . . . 141
5.6.1 SHMRs from main sequence vs. SED fitting . . . 141
5.6.2 Comparison with literature . . . 141
5.6.3 Evolution of the SHMRs and the pivot halo masses . . . 145

6 DISCUSSION 149
6.1 Reanalysis of sgzK Galaxies . . . 149
6.1.1 Clustering and HOD analysis . . . 149
6.1.2 Stellar-to-halo mass ratio . . . 152
6.2 Redshift Evolution of Satellite Fraction . . . 153
6.3 Redshift Evolution of Stellar-to-Halo Mass Ratio . . . 154

7 CONCLUSIONS AND FUTURE PROSPECTS 159
7.1 Summary and Conclusions . . . 159
7.2 Future Prospects . . . 162

REFERENCES 164

1. INTRODUCTION

1.1. Dark Matter and Structure Formation

1.1.1. Large-scale structure of the Universe

After "the Great Debate" about the scale of the Universe between Harlow Shapley and Heber Curtis, spiral nebulae have been recognized as extragalactic objects like our Milky Way Galaxy. Before that, some studies have reported that gaseous emission lines of nebulae except for the Andromeda Nebula show red-shifted spectra (e.g., Slipher 1917), indicating that most of the extragalaxies go away from our galaxy. Georges Lemaître interpreted the redshift of extragalaxies is originated from the expansion of the Universe (Lemaître 1927) based upon the general relativity theory, which is constructed by Albert Einstein (Einstein 1911, 1916), and Edwin Hubble discovered a tight correlation between the recession velocities of galaxies (equivalent to galaxy redshifts) and their radial distances, which is known as "Hubble's law" (Hubble 1929). Above the historical studies have established the picture of "Galaxy Universe": the Universe is composed of numerous galaxies.

Using the redshift of galaxy as a quantity of galaxy distance, a distribution map of galaxies can be drawn. A first attempt to draw the 3D galaxy distribution map was the Center for Astrophysics (CfA) Redshift Survey, which carried out the spectroscopic galaxy observations to measure the redshifts of the large number of galaxy at Harvard-Smithsonian Center for Astrophysics (Huchra & Geller 1982; Huchra et al. 1983). The CfA survey, which is a pioneering redshift survey, gave great insights into knowledge of structure of the Universe and opened the door for the idea of the hierarchical structure model of the Universe in units of galaxies; however, the number of spectroscopically observed galaxy was about $2,000$ and the observable redshift was limited for the local Universe ($z \sim 0.05$). By advancing the capabilities and efficiency of the observational instruments, many ambitious redshift surveys were implemented to achieve more precise descriptions of the galaxy 3D map by increasing the number of galaxy and limitation of redshift for the purpose of revealing the large-scale structure of the Universe.

The second CfA Redshift Survey (CfA2 Survey) started in 1984 and collected $\sim 20,000$ local bright galaxies in the northern sky. de Lapparent et al. (1986) firstly showed the map of galaxies by slicing the survey area of CfA2 Survey with narrow declination range (Δdecl ~ 6 degrees with ΔR.A. ~ 135 degrees) and Geller & Huchra (1989) discovered a supercluster that consists of a lot of galaxy clusters (Figure 1). That amazingly vast galactic filament was termed "the Great Wall" (the scale of the Great Wall is estimated beyond 500 million light years) and it was also found that the Great Wall is surrounded by a wide area with extremely low galaxy density, known as a "cosmic void".

The Sloan Digital Sky Survey (SDSS; York et al. 2000) is the largest redshift survey that is carried out both the multi-wavelength photometric observation as well as the spectroscopically observation using a wide-angle optical telescope mounted at Apache Point Observatory. This survey covered $\sim 12,000$ deg^2, which corresponds to $\sim 1/4$ of the total sky, by five optical photometric images (u-, g-, r-, i-, and z-band; Fukugita et al. 1996) and selected about 375 million objects. The targets of the spectroscopic observation were selected based upon above the photometric observation and redshifts of about one million galaxies and $100,000$ quasars were confirmed. Gott et al. (2005) produced the map of the Universe from the scale of the Kuiper-Belt objects in our solar system to $z \sim 1.69$ using the SDSS data and discovered an "SDSS Great Wall" of galaxies with length of ~ 1.37 billion light

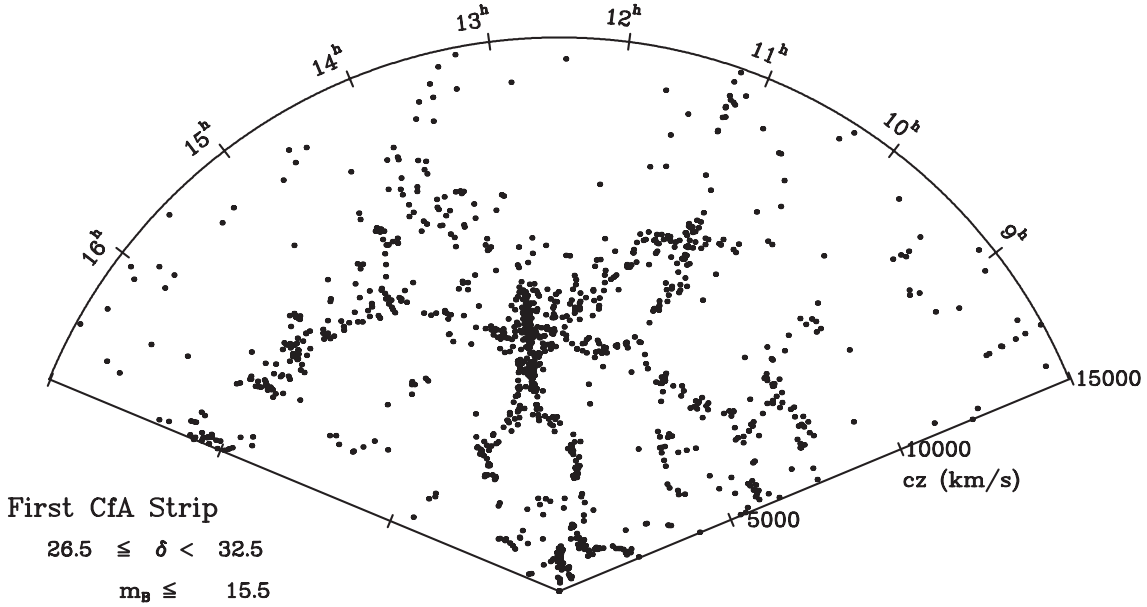

Figure 1.— The galaxy distribution map drawn by the second CfA Redshift Survey (CfA2 Survey). In this figure, about 1,100 spectroscopically observed galaxies are plotted. The filamentary structure of galaxies can be seen. It is noted that the long-drawn galaxy distribution around the center of the figure is not real structure but the effect of a Finger-of-God. This figure is originally presented in Geller & Huchra (1989) and credit of this figure is the Smithsonian Astrophysical Observatory.

years ($\sim$ 423 Mpc), which is the largest structure in the Universe.

1.1.2. Dark side of the Universe

The 3D map of the Universe drawn by the extensive redshift surveys revealed there are a lot of structures such as filaments, voids, sheets, and nods, and the hierarchical structure of the present-day Universe (galaxies, galaxy groups, galaxy clusters, and galaxy superclusters) is formed by combined those structures complexly. However, it is known that the total amount of baryonic mass is quite small to form the large-scale structure of the Universe which can be seen in present-day Universe within Hubble time (which is an approximately age of the Universe calculated by an inverse Hubble constant, $1/H_0 \sim 14.4$ Gyr) by numerical simulations, known as the "missing mass problem". To explain the shortage of the total amount of mass of the Universe and the anisotropy of the cosmic microwave background radiation (CMB), some cosmological models have been proposed to compensate the missing masses of the visible matters.

One of the most leading hypothesis to describe the picture of the Universe is a "ΛCDM" (e.g., Turner et al. 1984; Davis et al. 1985) paradigm, which is constructed based upon the idea of the "cold dark matter" model (CDM; Peebles 1982; Bond et al. 1982; Blumenthal et al. 1984). Dark matter is a hypothetical matter to fill a gap of the lack of the total mass of the Universe. The original idea of dark matter has been proposed by Fritz Zwicky to explain the mass gap between the galaxy cluster and field spiral galaxies (Zwicky 1933, 1937). He investigated the mass-to-luminosity ratio of the Coma galaxy

cluster and local Kapteyn stellar systems and reported that the mass-to-luminosity ratio of Coma cluster is $M/L \sim 500$, whereas each Kapteyn system shows $M/L \sim 3$, indicating that galaxy clusters are bounded by more massive gravitational system compared to field galaxies, and galaxy clusters are not the accumulated objects of galaxies. Another evidence for the existence of dark matter is a rotation curve of spiral galaxies. Rubin & Ford (1970) measured rotational velocities of the arms of Andromeda galaxy as a function of the radius from the galaxy center, and reported that radial velocities do not follow the Keplerian motion, showing the constant rotational velocity, irrespective of its radius. Many observational studies support the flat rotational velocity for spiral galaxies (e.g., Kent 1987; Sofue & Rubin 2001), and infer that invisible matter is extended to the outer regions of spiral galaxies to keep the rotational velocity constant. In addition, there are a lot of observational evidence for supporting the existence of dark matter such as the X-ray observation (e.g., Clowe et al. 2006; Akerib et al. 2014) and the weak gravitational lensing effect (e.g., Massey et al. 2007; Miyazaki et al. 2015).

Above problems (the missing mass problem and the rotation velocity problem) and observational results can be naturally explained by introducing the invisible matter in structure formation model. Some dark matter candidates have been proposed from the astrophysical aspect (massive astrophysical compact halo objects; MACHOs) and the aspect from the elementary particle theory (weakly interacting massive particles; WIMPs). MACHO is a baryonic candidate of dark matter that is too faint to detect by observations such as black holes, neutron stars, white and brown dwarfs, and planets. However, recent observations show that Big Bang nucleosynthesis model cannot produce the total amount of baryon enough for dark matter (Dar 1995) and the total mass of MACHO is much smaller than the expected mass of dark matter (Tisserand et al. 2007); thus, it is not expected that MACHO hardly contributes the total mass of dark matter. WIMPs is an elementary particle candidate of dark matter and it is assumed that they do not show the electromagnetic interaction, only show the gravitational interaction. Due to the achievement to approximately reproduce the large-scale structure of the Universe seen today by the cosmological simulations with WIMPs, it has been regarded that the possible candidate for dark matter is likely to be WIMPs, although it is still unclear which elementary particle is WIMPs.

By introducing dark matter and investigating the evolution of the distribution across the cosmic time by the cosmological N-body simulations, it is found that dark matter gradually forms clumpy structures by congregating together by their gravity. Thanks to the high-quality capability of recent calculating equipments, one can trace the evolutionary history of dark matter distribution from $z \sim 130$ to $z = 0$ and very small dark matter clumps by the latest cosmological dark matter N-body simulations, e.g., the Millennium simulation (Springel et al. 2005), the Bolshoi simulation (Klypin et al. 2011), and the MultiDark simulation (Prada et al. 2012), due to the high particle mass resolution as well as the spatial resolution. Figure 2 is an example of the cosmological N-body simulation with dark matter. One can find the clumpy and the filamentary structure in the simulation boxes, which are similar structure seen in the 3D map of the galaxies.

In summary, both the observational problems such as the missing mass problem and the constant rotation curve of the spiral galaxies as well as the theoretical problems to reproduce the large-scale structure of the Universe by the cosmological simulations are resolved simultaneously by introducing the assumed matter, termed dark matter. Considering the results of the cosmological simulations and the observational constraints, the candidate of dark matter is preferable to be massive elementary

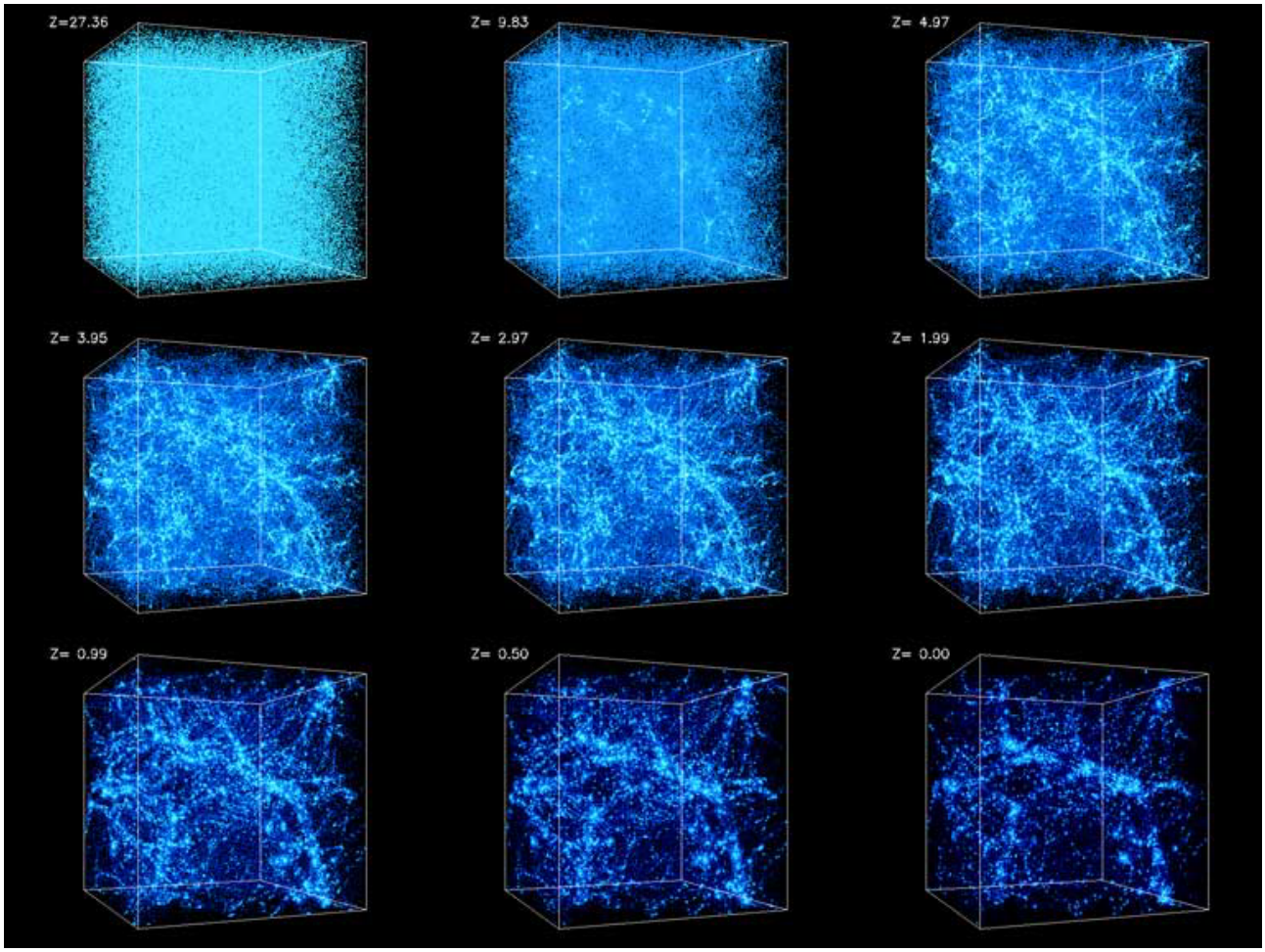

Figure 2.— Redshift evolution of dark matter distribution from $z = 30$ (top left) to $z = 0$ (bottom right) by the cosmological N-body simulation. The box size is 43 Mpc. One can find the clumpy and the filamentary structure in the simulation boxes. Credit of this figure is Kavli Institute for Cosmological Physics (KICP) at the University of Chicago.

particles with no electromagnetic interactions (CDM).

1.2. Galaxy–Halo Connection

According to the paradigm of the ΛCDM, all of galaxies form and evolve within the clump of dark matter, termed a "dark halo". Therefore, the structure formation in the Universe is largely governed by dark matter. Understanding of the formation scenario of the large-scale structure of the Universe boils down to the question of 1) how dark matter and dark haloes, which are harbors of galaxies, have been evolved, and 2) how galaxies have formed and evolved within dark haloes. The dark matter distribution is well studied by the cosmological N-body simulations precisely (see Section 1.1.2); however, the detailed scenario of galaxies is still an outstanding question. One can largely gain insights into the structure formation model if observed galaxies can be connected to the underlying dark matter (Figure 3). Nevertheless, it is quite difficult to link galaxies to their host dark haloes because dark matter cannot be observed directly by the telescope.

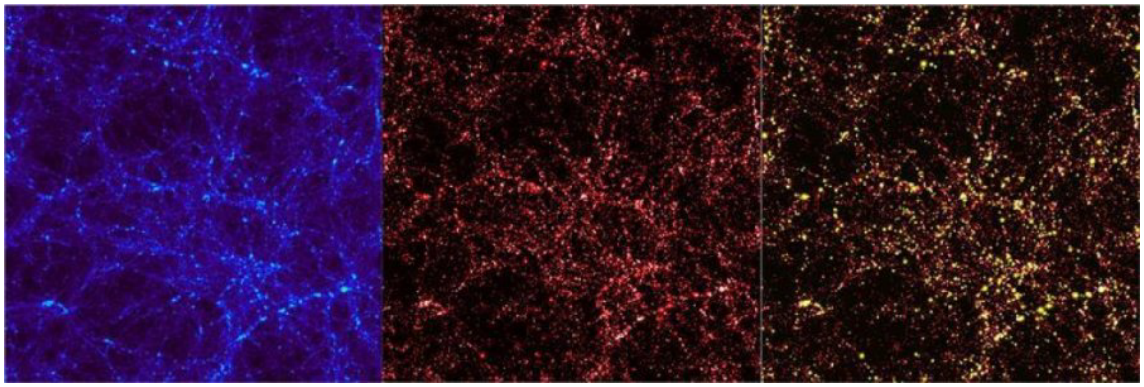

Figure 3.— The image of the distribution of dark matter and galaxies obtained by the numerical simulation. Left panel shows the results of the cosmological N-body simulation at $z \sim 2$ and the central panel is a simplified view of the network of dark matter. The right panel displays the sites of highly star-forming regions shown in yellow. The credit of this figure is The Virgo Consortium/Alexandre Amblard/ESA. Refer to Amblard et al. (2011) for more details.

There are various technique for the connection between galaxies and their host dark haloes, i.e., a galaxy clustering technique (e.g., Totsuji & Kihara 1969; Peebles 1980; Eisenstein et al. 2005; Zehavi et al. 2011), an abundance-matching method (e.g., Colín et al. 1999; Kravtsov & Klypin 1999; Vale & Ostriker 2004; Conroy et al. 2006), and a weak-lensing technique (e.g., Tyson et al. 1990; Kaiser 1998; Bacon et al. 2000; Wittman et al. 2000). However, the galaxy clustering is the almost only method to relate high-redshift observed galaxies to dark haloes. In this section, I will give a detailed prescription how observed galaxies can be linked to their host dark haloes along the description of the galaxy clustering.

1.2.1. Galaxy clustering

One of the most critical methods to connect observed galaxies to the underlying invisible dark matter is to use the galaxy clustering. Galaxy clustering represents how galaxies bound together by their gravity. Galaxies within massive dark haloes show high amplitude of the galaxy clustering because massive dark haloes strongly correlate with each other, whereas galaxies within less massive dark haloes show the weak clustering amplitude. Thus, one can unveil the properties of invisible dark haloes by investigating the strength of amplitudes of the galaxy clustering.

The pioneering study of the galaxy clustering is Totsuji & Kihara (1969), who firstly applied a two-point angular correlation function (ACF) for galaxy samples obtained by Shane & Wirtanen (1967) to estimate the galaxy clustering statistically. The ACF, which is practically described $\omega(\theta)$, represents how galaxies distribute over random/uniform distribution with a separation angle of θ. A probability to find both a galaxy at an element of solid angle $\delta\Omega_1$ and another one at $\delta\Omega_2$ with a separation angle of θ simultaneously, $\delta P(\theta)$, can be written using the ACF as follows:

$$\delta P(\theta) = \bar{N}\,[1 + \omega(\theta)]\,\delta\Omega_1\delta\Omega_2, \tag{1}$$

where $\bar{N}$ is an averaged surface density of galaxy on the celestial sphere. By defining the two-point real-space auto correlation function as $\xi(r)$, which represents the excess of galaxy distribution over the random distribution with distance r in real-space, δP can be easily extended into real-space as,

$$\delta P(r) = \bar{n}\,[1 + \xi(r)]\,\delta V_1\delta V_2, \tag{2}$$

where $\bar{n}$ is an averaged number density of galaxy and δV is a comoving volume element. The ACF is quite useful quantity to estimate the galaxy clustering strength statistically; however, the drawback of the ACF is that one have to collect the sufficient number of galaxy samples to obtain ACFs with high S/N ratio.

Many observational studies attempt to reveal the relationship between various galaxies and host haloes via the clustering analysis. Zehavi et al. (2005, 2011) carried out precise clustering analyses using the large number of local galaxies obtained by the SDSS survey and discussed the correlation between galaxy properties and strengths of the galaxy clustering. Using the selection methods for the specific galaxy populations or the photometric redshift via the SED fitting technique (see Section 3.1.2), galaxy clustering analyses have been applied for galaxies at $z > 1$ (e.g., Adelberger et al. 2005; Kashikawa et al. 2006; McCracken et al. 2010; Wake et al. 2011; Coupon et al. 2012; Ishikawa et al. 2015). Furthermore, Barone-Nugent et al. (2014) computed the ACF at $z \sim 7.2$ and derived the bias parameter (refer to Section 2.1.4) using the data of the Hubble eXtreme Deep Field (XDF; Illingworth et al. 2013) and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011).

By dividing galaxy samples into subsamples according to their baryonic properties such as stellar masses, luminosities, and magnitudes, one can discuss the dependence of the clustering properties on baryonic properties of galaxies. One can also reveal the relationship between the dark halo masses and the properties of galaxies; however, it is necessary to assume some analytical formulae and the approximated picture of a one-to-one correspondence between galaxies and dark haloes. Thus, more precise model is required to interpret observed galaxy clustering in terms of the realistic physical processes.

1.2.2. Halo occupation distribution model

A "halo occupation distribution" (HOD, e.g., Ma & Fry 2000; Seljak 2000; Berlind & Weinberg 2002; Berlind et al. 2003; van den Bosch et al. 2003a) is a powerful theoretical approach that relates the galaxy distribution to the dark matter distribution. The HOD formalism describes the galaxy distribution within the host dark halo by characterizing them from the perspective of the probability distribution $P(N|M)$ that a halo of virial mass M contains N galaxies with specific physical properties, such as color or galaxy type. One of the advantages of the HOD formalism is that the HOD parameters have explicit physical meanings; this enables us to easily interpret the relationship between the galaxies and the host haloes.

The largest difference between the HOD framework and other galaxy–halo correspondence methods, such as the abundance-matching method, is that the HOD model is based on a more realistic halo model and does not assume a one-to-one correspondence between galaxies and host haloes. It is known that different types of galaxy pairs contribute to the ACF between large-angular scales and small-angular scales. At the small-angular scale, galaxy clustering is contributed to mainly by galaxy pairs that are located in the same dark halo, referred to as the "1-halo term", whereas galaxy clustering at the large-angular scale is attributed to galaxy pairs that reside in different dark haloes, referred to as the "2-halo term". These two components result in different power-law slopes of the ACF at the two angular scales; these characteristics are well described using the HOD formalism. In Figure 4, I present the conceptual diagram to represent what type of the galaxy pair is dominant for the HOD formalism at each angular scale.

The HOD model predicts a power spectrum of galaxies by assuming some analytical formulae for characterizing the dark haloes and the occupation manner of galaxies within dark haloes. The occupation of galaxies can be described by formulating the number of galaxies within dark haloes, $N_{\rm tot}(M_{\rm h})$, as a function of dark halo mass $M_{\rm h}$. Total number of galaxies is composed of two types of galaxies: a central galaxy and satellite galaxies. One can describe the number of galaxies within dark haloes by considering the above galaxy types as $N_{\rm tot}(M_{\rm h}) = N_{\rm cen}(M_{\rm h}) + N_{\rm sat}(M_{\rm h})$, where $N_{\rm cen}(M_{\rm h})$ and $N_{\rm sat}(M_{\rm h})$ represent the number of the central and the satellite galaxies within dark haloes as a function of halo mass $M_{\rm h}$. It is noted that $N_{\rm cen}(M_{\rm h})$ satisfies the condition of $N_{\rm cen}(M_{\rm h}) \in [0, 1]$ because dark haloes cannot possess multiple central galaxies due to the definition of the central galaxy. Once the galaxy power spectrum from the HOD model is computed, one can derive a correlation function of galaxies in real space by the Fourier transformation using a so-called "Wiener-Khintchine relation". The ACF can be derived by projecting the real-space two-point correlation function using a "Limber's approximation" (Limber 1953). I plotted the galaxy power spectrum, real-space correlation function, and the projected ACF calculated by the HOD formalism in Figure 5. Detailed descriptions of the theoretical framework of the HOD formalism are given in Section 2.3.

1.2.3. Interpretation of observed galaxy clustering using the HOD model

Recently, many studies use the HOD model to interpret the observed galaxy clustering signals in high-redshift Universe as well as local/low-redshift Universe. Zehavi et al. (2005) carried out the HOD analysis to investigate the dependence of the galaxy properties such as galaxy color or its luminosity on the galaxy clustering using the numerous local galaxy samples ($200,000$ galaxies over $2,500$ deg^2

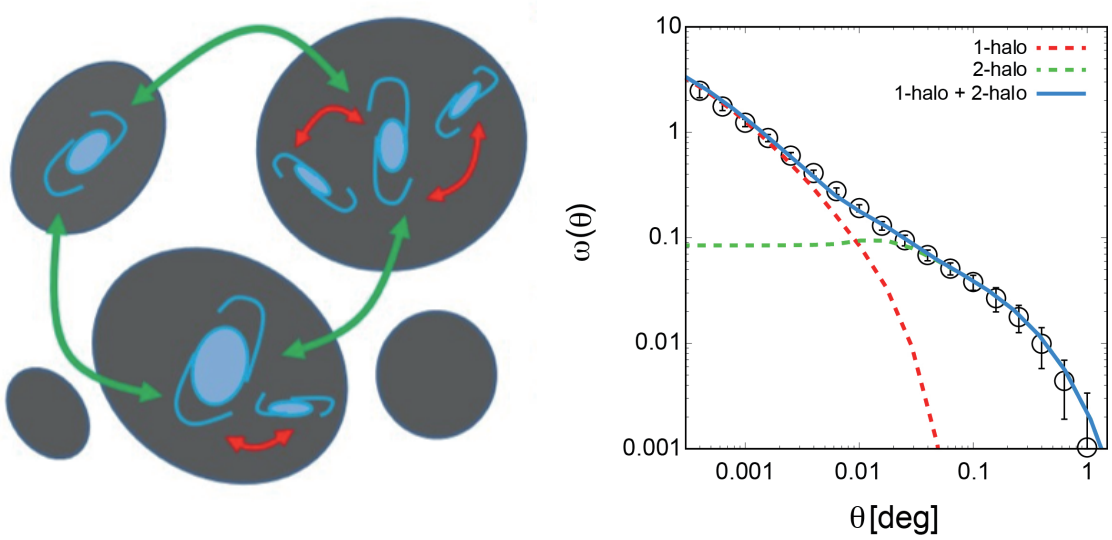

Figure 4.— A conceptual diagram to explain the HOD formalism. Left figure is an illustration to represent how galaxies correlate with each other and right figure is the observed ACF and the best-fit HOD fitting. In the HOD model, massive dark haloes can possess one central galaxy and multiple satellite galaxies within them, whereas less massive dark haloes contain only one or no central galaxy. Within massive dark haloes, strong clustering of the central–satellite galaxy clustering and the satellite–satellite galaxy clustering can be observed (red arrows in left figure); these clustering signals appear at small-angular scales in ACFs, termed the "1-halo term" (dotted red line in right figure). On the other hand, at large-angular scale, only the galaxy clustering between central galaxies within different dark haloes (green arrows in left figure) appears to the ACF, termed the "2-halo term" (dotted green line in right figure). The observed ACF can be represented by the sum of each component (solid blue line in right figure), the 1-halo term and the 2-halo term. It is noted that the observed ACF of right figure is measured by the galaxies obtained by the HSC SSP survey at $0.80 < z_{\rm phot} < 1.10$ with stellar-mass limit of $\log\left(M_\star/M_\odot\right) > 9.8$.

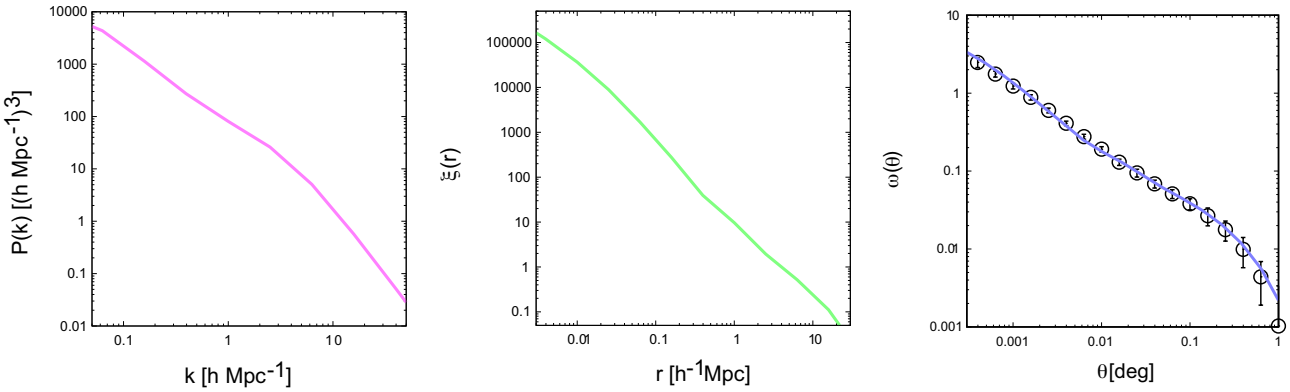

Figure 5.— The galaxy power spectrum, $P(k)$, (left panel), the real-space two-point correlation function, $\xi(r)$, (middle panel), and the ACF, $\omega(\theta)$, (right panel) predicted by the HOD formalism. First, the HOD model compute the power spectrum of galaxies from the assumption of some analytical formulae and the galaxy occupation within dark haloes. Power spectrum can be transformed from Fourier space into real space using the Wiener-Khintchine relation, and the transformed function os the real-space two-point correlation function. The ACF is derived by projecting the real-space two-point correlation function using the Limber's approximation. One can compare observed galaxy ACFs with the ACF calculated by the HOD model and investigate the properties of host dark haloes of galaxy samples by constraining the occupation of galaxies within dark haloes. The power spectrum, the real-space two-point correlation function, and the ACF presented in this figure are derived from the best-fit HOD model for galaxies obtained by the HSC SSP survey at $0.80 < z_{\rm phot} < 1.10$ with stellar-mass limit of $\log\left(M_\star/M_\odot\right) > 9.8$.

at $z < 0.22$) obtained by the second data release of SDSS survey, and Zehavi et al. (2011) extended their analyses by increasing the redshift range and the number of galaxy samples (700,000 galaxies over 8,000 deg^2 at $z < 0.25$) using the completed data of SDSS redshift survey (see Section 3.1.1).

As is the case with the local Universe, low-z Universe has been well investigated the relation between galaxies and their host haloes because of the easiness in collecting a large number of galaxy samples (e.g., Matsuoka et al. 2011; Wake et al. 2011; Coupon et al. 2012; Skibba et al. 2014; Coupon et al. 2015; Rodríguez-Torres et al. 2016); however, $z \sim 2$ Universe has not been well studies due to the difficulty of the sample selection. McCracken et al. (2010) collected the star-forming galaxies (sBzK galaxies) and passively evolving galaxies (pBzK galaxies) at $z \sim 2$ using the "BzK selection technique" (Daddi et al. 2004) in the COSMOS field, and Béthermin et al. (2014) firstly applied the HOD analysis on the sBzK galaxies using the photometric catalogue of McCracken et al. (2010). Moreover, Ishikawa et al. (2016) carried out precise clustering analysis and HOD analysis on star-forming galaxies at $z \sim 2$, which is approximately twice sample number compared to McCracken et al. (2010) selected by Ishikawa et al. (2015).

It has been investigated in higher-redshift Universe ($z > 3$) compared to $z \sim 2$ because the selection method of the high-redshift galaxy has already been established in the mid-1990s (e.g., Steidel et al. 1996) and they can be selected only using the optical images. Harikane et al. (2016) carried out clustering and HOD analyses of Lyman break galaxies at $z = 4 - 7$ by combining data of Hyper Suprime-Cam and publicly available data of Hubble Space Telescope. They investigated the redshift evolution of SHMRs but the accuracy of the clustering analysis is not good due to the small number of galaxy samples. Ishikawa et al. (2017) collected a large number of LBG samples enough to carry out precise clustering analysis in high-z Universe and confirmed the validity of the model prediction of the stellar-to-halo mass ratio presented by Behroozi et al. (2013b).

1.2.4. Beyond the "classical" HOD model

The HOD formalism has been developed to interpret galaxy clustering with relating the underlying dark matter and succeeded to reveal the relation between galaxies and host haloes; however, there is still room for improvement. For example, the HOD model does not assume that central galaxy should be brighter or heavier than satellite galaxies within the same dark haloes. Moreover, there is no condition that massive haloes should possess brighter or heavier galaxies compared to less massive haloes, provided that galaxies satisfy the threshold of the stellar mass (or the flux) for its sample selection. However, it seems oddly by considering the observational results; the central galaxy of galaxy clusters is the brightest galaxies among the member galaxy, termed a "brightest cluster galaxy (BCG)", and the stellar masses of central galaxies monotonically increase with increasing their host haloes according to stellar mass–halo mass relations in the local Universe (e.g., Leauthaud et al. 2012; Kravtsov et al. 2014; Coupon et al. 2015).

To fill a gap between the "classical" HOD formalism as a toy model and realistic physical conditions, some improved HOD models have been proposed. The early attempt was to consider the constraint of the observed luminosity functions or the stellar-mass functions in the HOD model, known as a "conditional luminosity function model" (CLF; Yang et al. 2003; van den Bosch et al. 2003a) and a "conditional mass function model" (CMF; Moster et al. 2010). The CLF model considers the average number of galaxies within the dark haloes with mass $M_{\rm h}$, $\Phi(L|M_{\rm h})$, with luminosities within the range

of $L \pm L/2$, whereas the CMF model considers the average number of galaxies within the dark haloes with mass $M_{\rm h}$, $\Phi(M_\star|M_{\rm h})$, with stellar masses within the range of $M_\star \pm M_\star/2$. Therefore, one can say that the CFL and the CMF model is the differential form of the classical HOD model with respect to the luminosity or the stellar mass. By constraining the number of galaxies according to the luminosity or the stellar mass of galaxies, it seems to be resolved the issue that HOD model distributes galaxies regardless their baryonic properties. Nevertheless, both the luminosity functions and the stellar-mass functions are suffered from relatively large scatters and not observed correctly especially in high-z Universe. Leauthaud et al. (2012) carried out the HOD analysis for the galaxy samples at $z \lesssim 1$ in the COSMOS field using an extended HOD model that is introduced the constrain of the parameterized form of the stellar-to-halo mass ratio (Behroozi et al. 2010) constructed by Leauthaud et al. (2011). Moreover, Zu & Mandelbaum (2015) proposed another post HOD model to compensate the stellar-mass incompleteness of galaxy samples by introducing the observed stellar-mass function in the formulated HOD model, named an "`iHOD` model", where `i` denotes "incompleteness".

The classical HOD model assumes that galaxy clustering is primarily governed only by the halo mass, albeit other physical properties of dark haloes. However, recent studies have reported that galaxy clustering strengths also depend on the age of host haloes, known as a "halo assembly bias" (e.g., Gao & White 2007; Reed et al. 2007; Miyatake et al. 2016). Hearin et al. (2016) developed an improved HOD formalism, called a "decorated HOD model", which is taken into account the effect of the halo assembly bias in their HOD model. Zentner et al. (2016) reanalyzed the galaxy clustering of the SDSS galaxies, which are originally analyzed by Zehavi et al. (2011), using the decorated HOD model and concluded that the halo assembly bias cannot be ruled out its possibility for the SDSS galaxies. Another big feature of the decorated HOD model is that this model uses the results of the cosmological N-body simulations for information of dark haloes instead of employing the analytical formulae of dark haloes; one assigns observed galaxies to dark haloes following the occupation manner of the HOD formalism in the similar manner as the abundance-matching method and the ACF from the decorated HOD model can be calculated by measuring the clustering of galaxies within the simulation box. However, we should keep in mind that detailed properties of the halo assembly bias is under discussion. Miyatake et al. (2016) found a strong correlation between concentrations of cluster member galaxies and the density fluctuations of dark matter using the local SDSS clusters; however, it is still unclear whether the above correlation can be applied for galaxy-size haloes and the more distant clusters. Tinker et al. (2016) investigate the correlation between halo ages and galaxy ages to identify the halo assembly bias using a galaxy group catalogue (Blanton et al. 2005) and concluded that massive galaxies show correlation between a large-scale environment and a large-scale density fluctuation, whereas the correlation between them cannot be seen for less massive galaxies ($M_\star \lesssim 10^{12} M_\odot$). Further elucidation of the halo assembly bias and improvement of the HOD model with the assembly bias may require future observations.

1.2.5. Stellar-to-halo mass ratio

Another method that can be used to relate the dark halo to the distribution of baryons in the dark halo is termed the "stellar-to-halo mass ratio" (SHMR), which is defined as the ratio between the dark halo mass and the galaxy stellar mass, represents the conversion efficiency from baryonic matter into stellar components within dark haloes. The SHMR has been the subject of many theoretical studies aiming to reveal or constrain the properties of galaxies and the conversion efficiency from

baryons into galaxies in dark haloes (e.g., Behroozi et al. 2010; Moster et al. 2010; Yang et al. 2012; Behroozi et al. 2013a,b). On the $M_{\rm h}$ versus $M_\star/M_{\rm h}$ diagram, the SHMR has a peak at $M_{\rm h} \sim 10^{12} M_\odot$, which corresponds to the most efficient mass, termed a "pivot halo mass" ($M_{\rm h}^{\rm pivot}$; Leauthaud et al. 2012), to convert from baryonic matter into stars, which is consistent with theoretical prediction (e.g., Blumenthal et al. 1984). However, the SHMR at the $M_{\rm h}^{\rm pivot}$ is much smaller than the universal baryon fraction ($\Omega_{\rm b}/\Omega_{\rm m} \sim 0.167$) (e.g., Behroozi et al. 2010; Leauthaud et al. 2012; Yang et al. 2012), indicating that a part of baryon in dark haloes are not converted into stars. This characteristic dark halo mass is determined by the equilibrium of the suppression mechanism that star formation becomes equal between the shallow gravitational potential wells for less-massive dark haloes and active galactic nuclei (AGN) feedback, and high virial temperature of massive dark halo.

Yang et al. (2012) presented a functional form of the SHMR up to $z \sim 4$ using the observational stellar-mass functions and the two-point correlation functions via the CLF model. Moster et al. (2013) carried out numerical simulations to constrain the SHMR at $0 < z < 4$ with observational constrains of stellar-mass functions. They found that the most efficient halo mass shows mild redshift evolution; $\log(M_{\rm h}^{\rm pivot}/M_\odot) \sim 11.8$ at $z = 0$ and $\log(M_{\rm h}^{\rm pivot}/M_\odot) \sim 12.5$ at $z \sim 4$, indicating that star formation in high-redshift preliminary takes place only in massive dark haloes. Behroozi et al. (2013b) predicted that the redshift evolution of the SHRM from $z = 8$ to $z = 0$ using numerical simulations. Figure 6 is the results of their SHMRs obtained by the numerical simulation with observational constrains of stellar-mass functions, specific star-formation rates, and cosmic star-formation rates at each redshift. They concluded that star formation is the most efficient at $M_{\rm h} \sim 10^{12} M_\odot$ and its evolution is quite small at $0 < z < 5$; however, there was a trend that the pivot halo mass became larger with increasing redshift from $z = 0$ to $z = 3$, and the conversion efficiency for the massive dark halo was the highest when $2 < z < 3$. Besides above theoretical approach to the SHMRs, many studies tackle to reveal the details and the redshift evolution of SHMRs using the semi-analytical model (e.g., De Lucia & Blaizot 2007; Guo et al. 2011; Birrer et al. 2014; Somerville et al. 2015).

In addition to the theoretical works, observational studies have been revealed the SHMR, almost of which are implemented by clustering analyses. Leauthaud et al. (2012) carried out HOD analyses for $z < 1$ galaxies in the COSMOS field; they employed an improved HOD model that included the constraints of the SHMR of central/satellite galaxies, enabling them to investigate the SHMR of central and satellite galaxies separately. Leauthaud et al. (2012) showed that the $M_{\rm h}^{\rm pivot}$ mass varies with the redshift, which is evidence of mass downsizing at $0 < z < 1$, and the SHMR of satellite galaxies becomes dominant for massive dark halo ($M_{\rm h} > 10^{13} M_\odot$; Figure 7). Coupon et al. (2015) also investigated the SHMRs of galaxies at $z \sim 0.8$ by adopting the HOD+SHMR model of Leauthaud et al. (2012) in the CFHTLenS (Heymans et al. 2012) and the VIPERS fields (Guzzo et al. 2014; Garilli et al. 2014). They showed an excess of the SHMR for massive dark haloes compared to the results of Behroozi et al. (2013b), indicating that the total amount of stellar components within massive haloes is dominant from the satellite galaxies.

Recent observational studies also reveal the SHMRs even at $z > 1$ Universe by overcoming the difficulties to estimate galaxy stellar masses as well as dark halo masses. McCracken et al. (2015) obtained galaxies at $0 < z < 3$ by the SED fitting technique in the UltraVISTA field (McCracken et al. 2012). They calculated the SHMR up to $z < 2.5$; however, SHMRs at $z > 1$ are not estimated correctly due to the large scatters of the photometric redshifts. Ishikawa et al. (2016) carried out the HOD analysis on star-forming galaxies at $z \sim 2$ and revealed the SHMR, showing that the decreasing

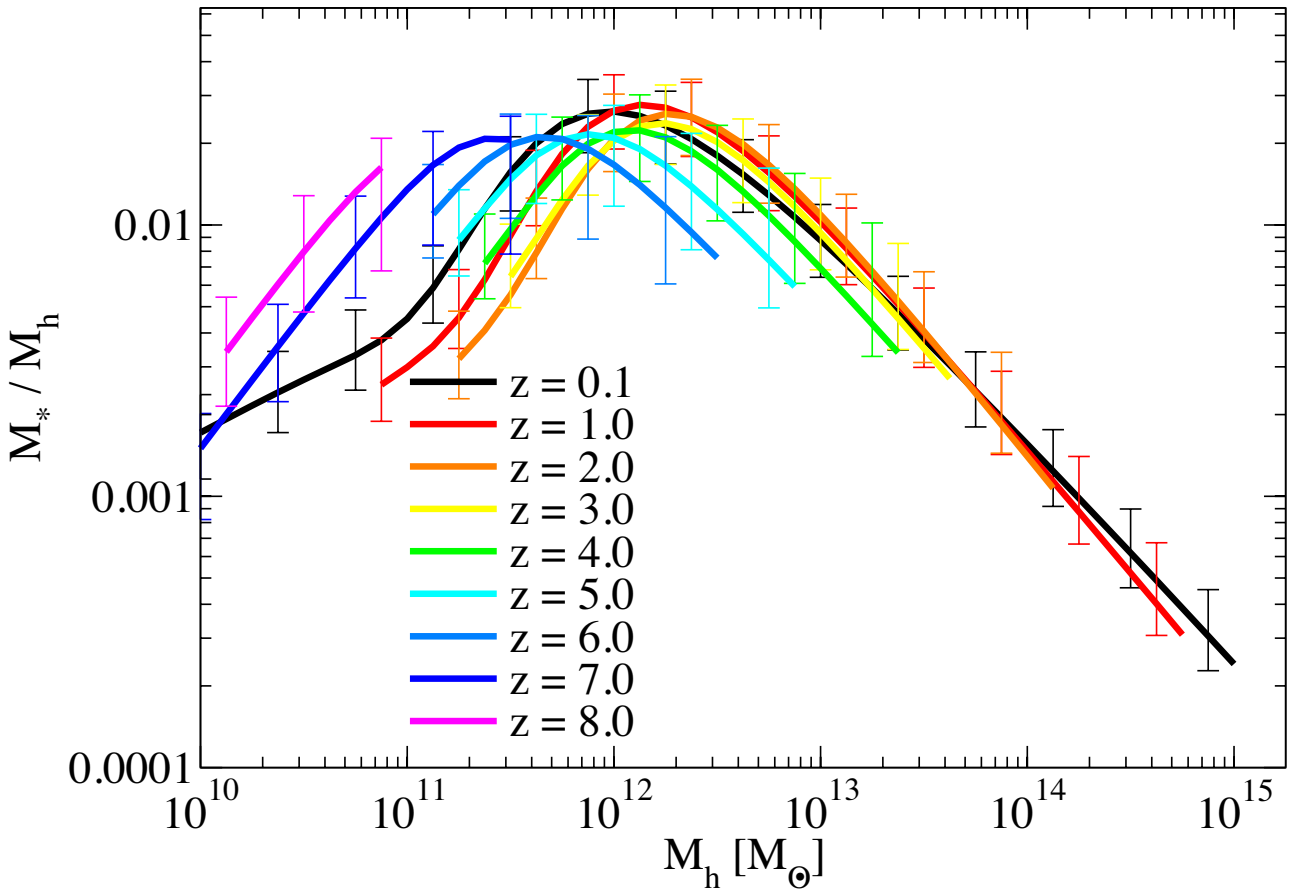

Figure 6.— The redshift evolution of the SHMRs calculated by numerical simulations (Behroozi et al. 2013b). In this simulation, they only considered the contribution of central galaxies; thus, these SHMRs are the relation of the central galaxies. The SHMRs are calculated using observational constraints of galaxy stellar-mass functions, specific star-formation rates, and cosmic star-formation rates from $z = 0$ to $z = 8$ at each redshift. This figure shows that the most efficient halo mass ($M_{\rm h}^{\rm pivot}$) is $M_{\rm h}^{\rm pivot} \sim 10^{12} M_{\odot}$ regardless its redshift and the values of the SHMRs at $M_{\rm h}^{\rm pivot}$ do not significantly change at $z = 0 - 8$. This figure is originally presented by Behroozi et al. (2013b).

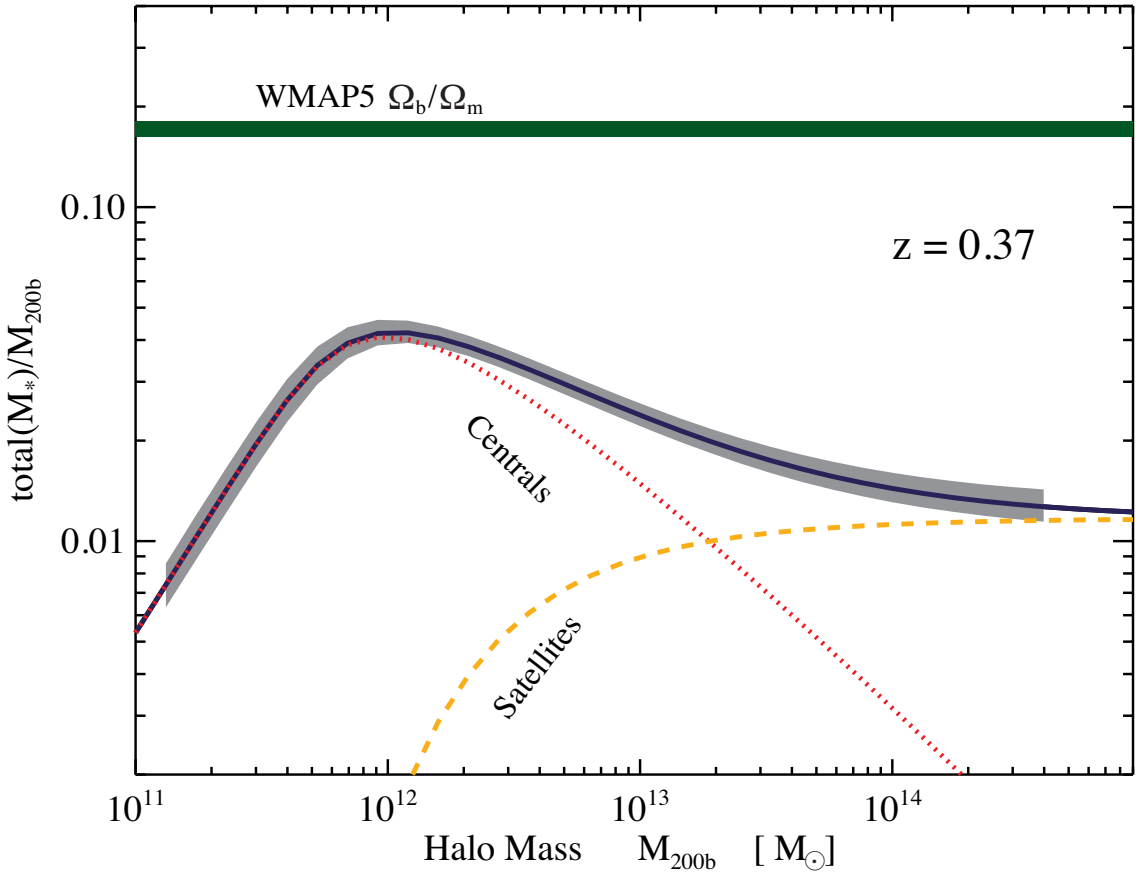

Figure 7.— The results of the SHMRs of galaxies at $z \sim 0.37$ in the COSMOS field derived by Leauthaud et al. (2012). They used the improved HOD model including the constrains of the SHMR, and investigate the relationship between galaxies and dark haloes by a joint analysis of the spatial galaxy clustering, the galaxy–galaxy lensing, and the stellar-mass function developed by Leauthaud et al. (2011). The total SHMR is attributed by central galaxies for less massive dark haloes and consistent with the results of Behroozi et al. (2013b), whereas satellite galaxies are dominant for massive dark haloes ($M_{\rm h} \gtrsim 10^{13} M_\odot$). This figure is originally presented by Leauthaud et al. (2012).

trend of the SHMR of massive galaxies can be seen even at high-z. Harikane et al. (2016) calculated the SHMRs of quite less massive ($M_\star \sim 10^{7-9} M_\odot$) Lyman break galaxies (LBGs) at $z = 4 - 7$ using the archive data of the Hubble Space Telescope and the Subaru Telescope/Hyper Suprime-Cam. They found that the redshift evolution of their SHMRs was consistent with the theoretical predictions via the abundance-matching technique (Behroozi et al. 2013b), and baryon conversion efficiency at $z \sim 4$ increased monotonically up to $\langle M_{\rm h} \rangle \sim 10^{12} M_\odot$. Moreover, Ishikawa et al. (2017) carried out HOD analyses on LBGs at $z \sim 3$, 4, and 5 using the publicly available data of the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Gwyn 2012) Deep Fields and achieved to calculate the $M_{\rm h}^{\rm pivot}$ at $z > 1$ firstly. They showed increasing trend of $M_{\rm h}^{\rm pivot}$ with increasing cosmic time at $z > 3$, and the SHMRs at $M_{\rm h}^{\rm pivot}$ show little evolution, indicating that mass growth rates of stellar components and dark haloes are comparable at $3 < z < 5$.

1.3. This Thesis

The ultimate goal of this study is to reveal how the large-scale structure of the Universe formed and evolved. To resolve this issue, it is necessary to unveil how galaxies formed and evolved within underlying dark haloes. Since the evolutionary history of dark matter, which are bearers of the structure formation, is well studied by the cosmological N-body simulation, one can gain insights into the structure formation scenario by linking galaxy evolution to the dark halo evolution. However, understanding of the scenario of galaxy formation and evolution is essential to complex physical mechanisms with baryons and it is difficult to research the evolutionary relationship among galaxy populations by tracing the evolution of their baryon characteristics.

Another approach to study the evolutionary history of galaxies is to trace the evolution of host halo masses of various galaxy populations. One can estimate the evolutionary model of galaxy to determine the properties of galaxies, including the dark halo mass, the stellar mass, the star-formation rate, and the morphology over cosmic time, and to compare these quantities with theoretical models. The dark halo mass of galaxies is a particularly important parameter to trace the mass assembly history of galaxies because, according to the ΛCDM model, dark haloes grow monotonically with cosmic time by merging, irrespective of the baryonic processes. However, the measurement of the mass of the dark halo is no straightforward. One of the most effective methods to determine the mass of dark haloes is to use the galaxy clustering.

Dark halo masses tell us the evolutionary connection among galaxy populations; however, this method just gives a possibility for the galaxy evolution since baryonic evolutions are not considered. Therefore, other properties of galaxies should be taken into account to establish the valid evolutionary history of galaxies. It is difficult to obtain baryonic properties of galaxy, such as the stellar mass, age, star-formation rate, and the amount of dust extinction; however, one can derive those quantities by carrying out the SED fitting technique, provided that multiple color photometric data are available. In addition, one can derive the distribution pattern of galaxies by applying the HOD formalism. The evolutionary history can be constructed validly by investigating the evolution of satellite fractions and SHMRs, as well as the evolution of dark halo masses among various galaxy populations at various redshifts.

In this thesis, I discuss the relationship between galaxies and their host dark haloes at $z = 0.3 - 5.4$

based upon the results of precision clustering analyses. Thanks to quite wide survey areas, I succeed to collect the large number of galaxy samples enough to obtain the high-quality ACF at each redshift. Furthermore, our high accurate ACFs enable us to carry out more precise clustering analyses using the HOD formalism. In addition, I reveal the baryonic properties of galaxies using multi-wavelength data, especially photometric redshifts and galaxy stellar masses, by applying the SED fitting technique and investigate the properties of dark haloes by the properties of baryons. By combining both aspects of galaxies of baryons and dark haloes, I target to unveil the evolutionary history of galaxies from $z \sim 5.4$ to $z = 0.3$ by tracing the evolution of baryonic properties such as satellite fractions and SHMRs, as well as the evolution of host dark halo masses.

This thesis is organized as follows. In Section 2, I present a theoretical framework of the galaxy clustering and the HOD formalism. The results of clustering and HOD analyses are given in Section 3, 4, and 5: in Section 3, I show the results of clustering using/HOD analyses at $0.3 < z < 1.4$ using the data of the Hyper Suprime-Cam Subaru Strategic Project (HSC SSP) survey, the relations between star-forming galaxies at $1.4 < z < 2.5$ (sgzK galaxies) and their host haloes by accurate clustering analyses based on the wide-field survey is presented in Section 4, and the results of the precision clustering/HOD analyses and the details and the evolution of SHMRs of LBGs at $2.5 < z < 5.4$ is given in Section 5. In Section 6, I give the organized discussion across cosmic history using the above results. Conclusion of this thesis and future prospects are presented in Section 7. Adopted cosmological parameters and the typical physical scales under those parameters are shown in the overview of each section.

2. A THEORETICAL FRAMEWORK OF GALAXY CLUSTERING ANALYSIS

2.1. Galaxy Clustering Statistics

2.1.1. Two-point angular auto correlation function

A Two-point angular auto correlation function (ACF) is an extensively used physical quantity to estimate the strength of the galaxy clustering statistically. The first attempt to apply the correlation function to measure the strength of galaxy clustering is Totsuji & Kihara (1969), and ACFs are widely accepted as a useful tool for clustering analysis after introduction of the famous textbook (Peebles 1980).

The very early studies of galaxy clustering (e.g., Davis et al. 1978; Phillipps et al. 1978) simply count the separation angles of galaxies because the number of galaxies that are confirmed their redshifts are quite small. Those early studies revealed that the ACF follows the power law as

$$\omega(\theta) = A_\omega \theta^{1-\gamma}, \tag{3}$$

where A_ω is an amplitude of the ACF and γ is a power-law slope. The simple counting method is, however, no longer used for almost all today's clustering analyses, including my study of this thesis, because of dealing enormous galaxy samples. Therefore, various ACF estimators have been proposed to evaluate the ACFs by dealing the galaxy distribution statistically. The most commonly used ACF estimator is proposed by Landy & Szalay (1993) as,

$$\omega_{\rm LS}(\theta) = \frac{\rm DD - 2DR + RR}{\rm RR}, \tag{4}$$

where DD, DR, and RR are the normalized pair counts of galaxy–galaxy, galaxy–random, and random–random samples with separation angle of $\theta \pm \delta\theta$, respectively. To estimate the excess of galaxy distribution over the homogeneous distribution, it is necessary to to generate random samples over the survey region. The random distribution is required to trace the geometry of galaxy distribution accurately, including the masked regions and edges of the photometric images. The number of random samples is typically more than ten times as many as the number of galaxy samples to reduce the Poisson error.

It is known that the variance of the estimator of Landy & Szalay (1993) is small that is comparable to Poisson error (e.g., Kerscher et al. 2000; Vargas-Magaña et al. 2013). Besides the estimator of Landy & Szalay, various ACF estimators have been proposed (e.g., Hewett 1982; Davis & Peebles 1983; Rivolo 1986; Hamilton 1993). Definitions of some of those ACF estimators, ($\omega_{\rm DP}$, $\omega_{\rm Ham}$, $\omega_{\rm Nat}$ represent the estimator of Davis & Peebles, Hamilton, and a natural ACF definition, respectively), are as follows:

$$\omega_{\rm DP} = \frac{\rm DD - DR}{\rm DR}, \tag{5}$$

$$\omega_{\rm Ham} = \frac{\rm DD \times RR - DR^2}{\rm DR^2}, \tag{6}$$

and

$$\omega_{\rm Nat} = \frac{\rm DD - RR}{\rm RR}, \tag{7}$$

respectively.

Pons-Bordería et al. (1999) compared the behaviors of some ACF estimators using a so-called "Cox point process", which is a process to compute ACFs using samples that are knows the "true" correlation function. They assigned galaxies to the CDM simulation box, whose box size was 80 h^{-1}Mpc, and compared the Cox ACF with ACF estimators, repeating this procedure 10 times. The referenced ACF estimators were proposed by Peebles & Hauser (1974), Davis & Peebles (1983), Rivolo (1986), Hamilton (1993), Landy & Szalay (1993), and Stoyan & Stoyan (1994). Pons-Bordería et al. (1999) concluded that, at a large-angular scale, Hamilton (1993) and Landy & Szalay (1993) show apparently small variances compared to others but values of ACF do not significantly change between ACF estimators, and Stoyan & Stoyan (1994) can estimate ACFs with small errors at a small-angular scale. However, the main conclusion of Pons-Bordería et al. (1999) was that there is no supreme estimator and each estimator has its advantages and disadvantages, thus, one should choose an appropriate ACF estimator for their purpose.

Kerscher et al. (2000) also investigated the difference of the ACF estimators by comparing them with the "true" correlation function of Virgo Hubble volume simulation cluster catalogue via direct counting. The result was almost identical to the result of Pons-Bordería et al. (1999); deviations of Hamilton (1993) and Landy & Szalay (1993) are smaller than others and those ACFs are comparable at a large-angular scale, whereas there is no significantly difference among estimators at a small-angular scale. They also investigated that the impact of the random size on the ACF, and concluded that the estimator of Landy & Szalay (1993) is preferable to Hamilton (1993) because of the small effect of the number of random samples.

The comparison of the ACFs calculated by some ACF estimators using our galaxy samples are shown in Figure 8. The ACFs are calculated by the estimators of Landy & Szalay (1993), Davis & Peebles (1983), Hamilton (1993), and the natural definition, respectively. Galaxy samples of these ACFs satisfy the conditions of $0.55 < z_{\rm phot} < 0.80$ and $\log(M_\star/M_\odot) > 11.2$, which are selected by the HSC SSP Wide-layer photometric catalogue of Mizuki at the region 3. The shaded region of each panel represents the 1σ Poisson error computed by the equation (10). As discussed in Pons-Bordería et al. (1999) and Kerscher et al. (2000), ACFs at small-angular scales are not significantly different among ACF estimators. On the other hand, at large-angular scales, ACFs of Davis & Peebles (1983) and the natural definition overestimate compared to those of Landy & Szalay (1993) and Hamilton (1993). The difference between the ACF derived by Landy & Szalay (1993) and Hamilton (1993) is quite small over all angular scales. The only one different point between these estimators is the most smallest angular scale; the ACF of Hamilton (1993) has a characteristic increase that cannot be explained by the halo model, i.e., the HOD model. We should keep in mind that one cannot determine the appropriate ACF estimator from this comparison because the "true" ACF of these galaxy samples is still unknown.

In this thesis, I employ the Landy & Szalay (1993) ACF estimator to quantify the strengths of galaxy clusterings because of following reasons. First, Landy & Szalay estimator can evaluate the ACFs with small variances at all angular scales. Especially, deviations at large-angular scales have large impact on the estimation accuracy of dark halo masses. Dark halo masses are calculated by the amplitude of the ACF. ACFs are typically fitted at $0.01 < \theta < 1$ (in units of the degree) to estimate the clustering strength and large errors in this angular scale lead to large uncertainties of dark halo masses. In addition, $M_{\rm min}$, which is an HOD mass parameter to scale the threshold dark halo mass

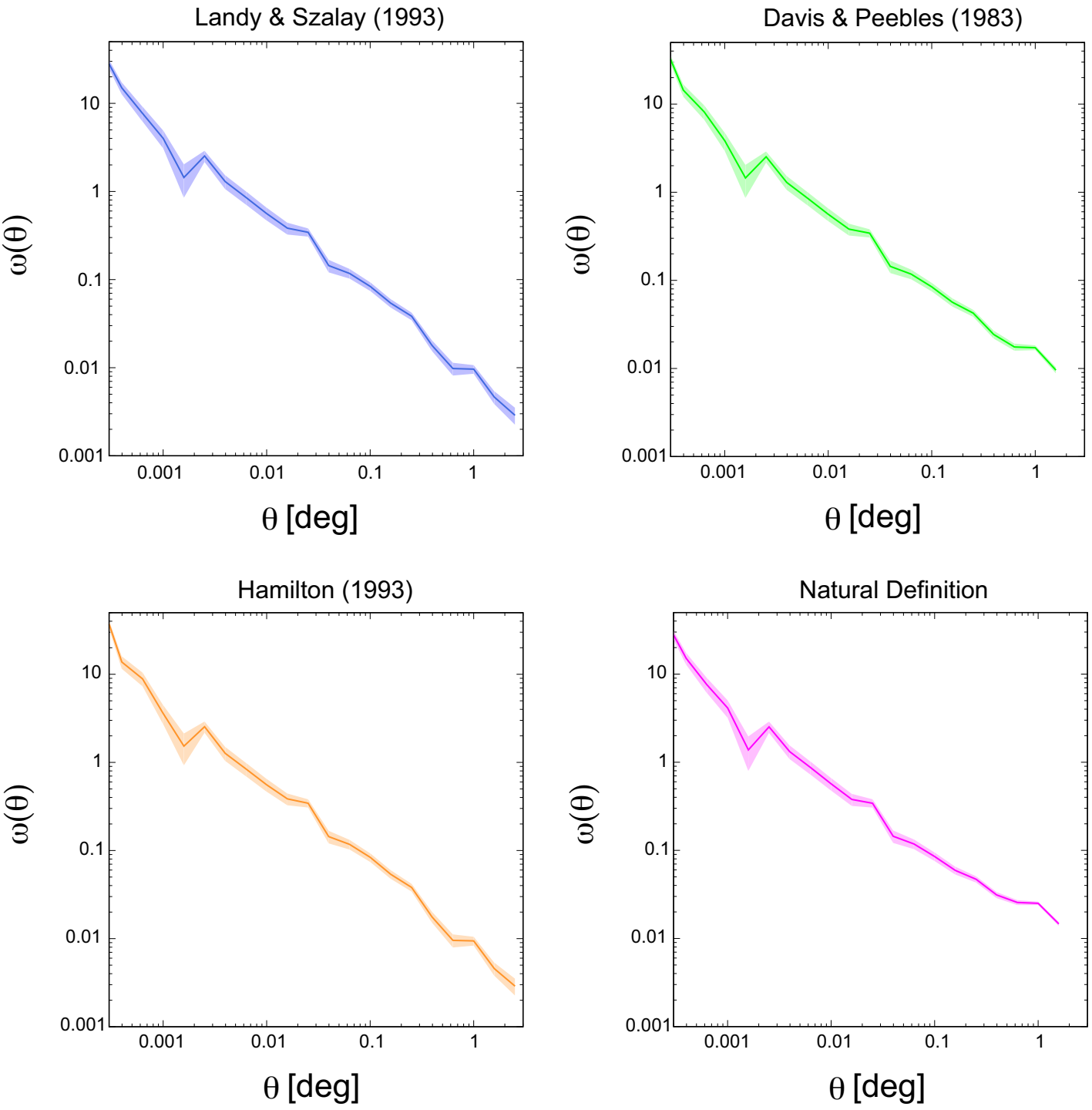

Figure 8.— Comparison of the ACFs calculated by the estimator of Landy & Szalay (1993, blue), Davis & Peebles (1983, green), Hamilton (1993, orange), and the natural definition (red), respectively. Solid lines are the values of ACF at each angular scales and shaded regions represent the 1σ Poisson errors calculated by the equation (10). All of these ACFs are computed by the galaxy samples satisfying $0.55 < z_{\rm phot} < 0.80$ and $\log(M_\star/M_\odot) > 11.2$ obtained by data of Hyper Suprime-Cam SSP Wide-layer at the region 3.

to possess the central galaxy, is sensitive to the ACFs at large-angular scales. Dark halo masses from HOD analyses and SHMRs will be suffered large uncertainties if ACFs are noisy at large-angular scales. The other reason is that most of previous clustering studies employ the estimator of Landy & Szalay (1993). Although Hamilton (1993) can evaluate identical ACFs to Landy & Szalay (1993), comparison should be done using the same estimator. From here, all of the observed ACF will be calculated through the definition of Landy & Szalay (1993).

It has been reported that observed ACFs at large-angular scales are underestimated due to the limitation of the survey field, known as the "integral constraint" (IC; Groth & Peebles 1977). The IC is determined the area of the survey field and its geometry, thus, one can estimate the IC using the random–random pair counting as

$$\mathrm{IC} = \frac{\Sigma\theta^{1-\gamma}\mathrm{RR}(\theta)}{\Sigma\mathrm{RR}(\theta)}, \tag{8}$$

assuming a power-law form (equation 3) for the intrinsic ACF (Roche et al. 1999). ACFs can be corrected their biased large-angular bins using the IC as

$$\omega(\theta) = \omega_{\mathrm{meas}}(\theta) \times \frac{\theta^{1-\gamma}}{\theta^{1-\gamma} - \mathrm{IC}}, \tag{9}$$

where $\omega_{\mathrm{meas}}(\theta)$ is a measured ACF.

2.1.2. Error estimation of the ACF

The ACF is suffered from statistical errors, especially for the case that the number of galaxy samples is quite small. The simplest method to evaluate the error of the ACF is the Poisson error, which can be calculated as,

$$\delta\omega(\theta) = \sqrt{\frac{1+\omega(\theta)}{DD(\theta)}}. \tag{10}$$

As above definition, the Poisson error is biased if the number of galaxy pairs, $DD(\theta)$, is small, i.e., clustering signals at small-angular scales are significantly suffered from large Poisson errors.

The Poisson error has mainly two problematic points: the Poisson error is not considered 1) the field-to-field variance and 2) the correlation between different angular bins. Recent extensive surveys, e.g., Subaru Hyper Suprime-Cam survey, the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016), and the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), cover more than $1,000$ deg^2 and clustering signals differ at the part of survey field due to the factor of nature such as cosmic variance, as well as the artificial factor such as the variance of the observational depth. Moreover, unlike the power spectrum in Fourier space, galaxy clustering signals in real-space measurements are correlated with each other; thus, one should account for the covariance of the ACF measurement.

To address these issues, some error estimation methods have been proposed to evaluate the clustering uncertainties by resampling the observed galaxy samples. Efron (1979) developed a "bootstrap resampling method" and the bootstrap method is firstly applied to the estimation of the uncertainties of the galaxy clustering by Barrow et al. (1984). The procedure of the bootstrap method is as follows: one randomly resamples the galaxy samples allowing for redundancy as many as the number of original galaxy samples, and calculate the ACF at each step, repeating this step for many times. The error of

the ACF can be evaluated by the root mean squares of each angular bin. This method can estimate the correlation between other angular bin by calculating the covariance matrix (see below); however, the field-to-field variance is not considered. Additionally, the countable galaxy pair is reduced because the separation angle of duplicative galaxy pair is zero; this method resamples by just excluding some galaxy samples and estimates the variance by the subset of galaxy samples that is selected randomly with small sample size.

Another estimation method is a "jackknife resampling method", basically concepts of which was developed by Quenouille (1949, 1956) and Tukey (1958). Shao (1986) and Shao & Wu (1989) constructed the "delete-one jackknife method", which divide survey field into sub-field and calculate the ACF by deleting one sub-field, repeating this procedure as times as the number of sub-regions. The jackknife resampling method can consider the field-to-field variance as well as the angular bin-to-bin correlations through the covariance matrix.

The covariance matrix, C_{ij}, is a $N \times N$ square matrix that represent the covariance between the i-th and j-th angular bin for the angular correlation function that consists of N-th angular bins, $\omega(\theta) = \{\omega(\theta_1), \omega(\theta_2), \cdots, \omega(\theta_N)\}$. The covariance matrix through each resampling method can be evaluated as follows:

$$C_{ij}^{\rm BS} = \frac{1}{N}\sum_{k=1}^{N}\left(\omega_k\left(\theta_i\right) - \bar{\omega}\left(\theta_i\right)\right) \times \left(\omega_k\left(\theta_j\right) - \bar{\omega}\left(\theta_j\right)\right), \tag{11}$$

and

$$C_{ij}^{\rm JK} = \frac{N-1}{N}\sum_{k=1}^{N}\left(\omega_k\left(\theta_i\right) - \bar{\omega}\left(\theta_i\right)\right) \times \left(\omega_k\left(\theta_j\right) - \bar{\omega}\left(\theta_j\right)\right), \tag{12}$$

where $C_{ij}^{\rm BS}$ and $C_{ij}^{\rm JK}$ are the (i, j) element of the covariance matrix derived by the bootstrap resampling method and jackknife resampling method, respectively. $\bar{\omega}\left(\theta_i\right)$ is a mean ACF of i-th angular bin of N times realization. The factors of each covariance matrix represents the correction factor that converts from the resampled sub-volume into the original volume, i.e., for each step, the bootstrap method resamples galaxy samples as the same volume as the original samples by allowing the duplicated selection, whereas the jackknife method decreases the resampled volume from N to $N-1$ by omitting one of divided subsample (cf., Norberg et al. 2009). According to the above definitions, the covariance matrix is a symmetric matrix, $C_{ij} = C_{ji}$, and the diagonal elements approximately correspond to the dispersions of each angular bin, i.e., $\delta\omega(\theta_i) \propto \sqrt{C_{ii}}$.

In this thesis, I adopt the jackknife resampling method to estimate the errors and the covariance matrices of observed ACFs. The number of sub-regions will be determined for each galaxy subset by considering the total survey field. Excessively small sub-region compared to the entire field causes little variation especially for the estimation at large-angular scale, whereas too much large sub-region cannot capture the field-to-field variance. The covariance matrix will be evaluated using the equation (12) throughout this thesis.

2.1.3. Real-space galaxy clustering

The ACF is a correlation function projected onto the celestial sphere and integrated along the line of sight. One can evaluate a real-space galaxy correlation function that contains more precise clustering information if radial distances or redshifts of galaxy samples are available. Real-space two-

point correlation function as a function of separation radius, $\xi(r)$, can be calculated via the same definitions of ACFs, provided that random samples are generated in the three-dimensional survey volume. The amplitude of the ACF is diluted along the line of sight; however, the clustering strength of the real-space correlation function reflects the intrinsic strength of sample galaxies because it is also considered the galaxy clustering along the direction of depth.

Observed real-space two-point correlation functions are also obtained in a power-law form as a function of separation radius as,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}. \tag{13}$$

The characterizing parameters of $\xi(r)$ are as follows: a power-law slope, γ, and a distance where $\xi(r)$ becomes unity, r_0, called a correlation length. Unity of the galaxy correlation function means that, in that scale, galaxies correlate with each other in a state of the density perturbation that is a boundary between the linear regime and the non-linear regime.

The correlation length is an important physical parameter to characterize the galaxies and their host haloes, and verify the non-linear perturbation theory; however, there is a difficulty in deriving the correlation length of various galaxy populations because of the necessity of redshift information. Limber (1953) succeeded to associate the amplitude of the ACF with the correlation length using the redshift distribution of galaxy samples, known as the Limber equation. The formulation of Limber equation improved by Peebles (1980) is described as,

$$A_\omega(1\mathrm{rad}) = r_0^\gamma \sqrt{\pi} \frac{\Gamma(\gamma - 1/2)}{\Gamma(\gamma/2)} \frac{\displaystyle\int_0^\infty dz\, F(z) D_\mathrm{A}^{1-\gamma}(z) \left(\frac{dN}{dz}\right)^2 g(z)}{\displaystyle\left(\int_0^\infty dz \frac{dN}{dz}\right)^2}. \tag{14}$$

The term of $A_\omega(1\mathrm{rad})$ corresponds to the amplitude of the value of the ACF at $\theta = 1$ radian, Γ is the Gamma function as $\Gamma(x) = \int_0^\infty dt\, t^{x-1} e^t$, $D_\mathrm{A}(z)$ is the angular diameter distance to redshift z defined under the assumption of $\Omega_k = 0$ as,

$$D_\mathrm{A}(z) = \frac{1}{1+z} \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m (1+z')^3 + \Omega_\Lambda}}, \tag{15}$$

where c is a speed of light and H_0 is a Hubble constant. $\frac{dN}{dz}$ is a redshift distribution of the galaxy sample and $g(z)$ is a function depends upon the cosmological parameters given as,

$$g(z) = \frac{H_0}{c} (1+z)^2 \sqrt{1 + \Omega_m z + \Omega_\Lambda [(1+z)^{-2} - 1]}. \tag{16}$$

$F(z)$ is a function describing the redshift evolution of the two-point correlation function $\xi(r)$, i.e.,

$$F(z) = \frac{(1+z)}{(1+z_c)^{-(3+\epsilon)}}, \tag{17}$$

where z_c is the center of the redshift distribution, and ϵ is assumed as $\epsilon = -1.2$ (see also Ouchi et al. 2004b). It should be noted that the Limber equation is an approximate expression to convert an amplitude of the projected correlation function into the real-space. Therefore, the accuracy of the result will become worse by increasing the redshift range (cf., Simon 2007).

Real-space two-point correlation function can describe more precise clustering characteristics of galaxies compared to the projected two-point angular correlation function; however, one should keep in mind that real-space correlation functions are suffered from the effect of the "redshift-space distortion (RSD)". RSD is caused by adopting the redshift as a radial distance and its origin is large peculiar velocities of galaxies in strong fields, such as the center of the galaxy cluster, compared to the Hubble flow. In local Universe ($z \ll 1$), the redshift and the radial distance of galaxies can be approximated as,

$$cz \sim H_0 d + v_{pecl}, \tag{18}$$

where d is a radial distance and v_{pecl} is a peculiar velocity of galaxy, respectively. In this expression, the redshift can be disturbed by the large peculiar velocity and galaxies will be elongated along the line of sight in the redshift space compared to the real-space distribution. This difference of the galaxy distribution between the redshift space and the real space around the strong gravity field is known as the effect of "Finger of God" (FoG, Jackson 1972), which is termed by that the galaxy distribution can be seen as pointing toward observers. The FoG effect only appears in a small scale, non-linear perturbation field such as the center of the galaxy cluster. Additionally, RSD effect also can be seen in a large scale, approximately linear perturbation field. By considering the comprehensive gravity field, high-density gravity fields attract galaxies, whereas low-density gravity fields like the cosmic voids seem to shun galaxies. Thus, the shape of large-scale galaxy distribution in redshift space seems to be collapsed along the tangential direction of line of sight. This RSD effect in large scale is called "Kaiser effect" (Kaiser 1987). To avoid these RSD effects in real-space correlation function, many studies integrate their real-space correlation functions along the line of sight and obtain projected two-point correlation functions even if redshifts of galaxies are determined with high accuracies (e.g., de la Torre et al. 2007; Zehavi et al. 2011; Durkalec et al. 2015).

As seen above, it is difficult to derive and treat the real-space correlation functions; however, the real-space correlation function is important physical quantity for the structure formation theory. Correlation functions represent the density fluctuation in real space and it can be related to a power spectrum, $P(k)$, which is the density fluctuation in Fourier space, known as "Wiener-Khintchine relation" as follows:

$$P(k) = 4\pi \int dx\, x^2 \frac{\sin kx}{kx} \xi(x), \tag{19}$$

and

$$\xi(x) = \frac{1}{2\pi^2} \int dk\, k^2 \frac{\sin kx}{kx} P(k). \tag{20}$$

These relations are useful to convert galaxy power spectra into correlation functions in the HOD analysis (see Section 2.3.3).

2.1.4. Dark halo mass estimation from large-scale galaxy clustering

Using the analytical formulae of the structure formation theory and assuming halo models, one can evaluate dark halo masses from galaxy correlation functions. The following expressions and analytical models are mainly developed by Mo & White (1996), Mo et al. (1996), Sheth & Tormen (2002), and Mo & White (2002), which have been shown to provide a connection between the bias of the dark haloes and the masses of the dark haloes largely based upon the ellipsoidal halo collapse model (Sheth

et al. 2001), provided that the bias of the dark haloes and the redshift distribution of galaxies are given.

The effective bias parameter, $b_{\rm eff}$, which relates the distribution between galaxies and underlying dark matter, is defined as,

$$b_{\rm eff} = \sqrt{\frac{\xi_{\rm g}(8h^{-1}{\rm Mpc})}{\xi_{\rm DM}(8h^{-1}{\rm Mpc})}}, \tag{21}$$

where $\xi_{\rm g}(8h^{-1}{\rm Mpc})$ and $\xi_{\rm DM}(8h^{-1}{\rm Mpc})$ are the real-space correlation functions of galaxies and dark matter with a scale of $8h^{-1}$Mpc. $\xi_{\rm g}(8h^{-1}{\rm Mpc})$ can be determined using equation (13) as,

$$\xi_{\rm g}(8h^{-1}{\rm Mpc}) = \left(\frac{8h^{-1}{\rm Mpc}}{r_0}\right)^{-\gamma}. \tag{22}$$

The correlation length can be estimated by the Limber equation (equation 14). Analytical formulae of the correlation function (or power spectrum) of dark matter have been proposed by the results of the cosmological N-body simulation (e.g., Hamana et al. 2002; Smith et al. 2003).

The bias parameter of the dark halo itself can be calculated by the analytical formulae given by Sheth et al. (2001) and Mo & White (2002). Here, I give a summary of these models. The linear growth factor $D(z)$ is defined as,

$$D(z) = \frac{g(z)}{g(0)(1+z)}, \tag{23}$$

where $g(z)$ depends upon the cosmological parameters, i.e.,

$$g(z) \approx \frac{5}{2}\Omega_m(z) \times \left\{\Omega_m^{4/7}(z) - \Omega_\Lambda(z) + \left[1 + \frac{\Omega_m(z)}{2}\right]\left[1 + \frac{\Omega_\Lambda(z)}{70}\right]\right\}^{-1}. \tag{24}$$

It is noted that the linear growth factor satisfies $D(0) = 1$. Here, $\Omega_m(z)$ and $\Omega_\Lambda(z)$ are the cosmological parameters at a given redshift, and they satisfy,

$$\Omega_m(z) = \frac{\Omega_{m,0} \times (1+z)^3}{E^2(z)}, \tag{25}$$

and

$$\Omega_\Lambda(z) = \frac{\Omega_{\Lambda,0}}{E^2(z)}, \tag{26}$$

where $E(z)$ is given by,

$$E(z) = \left[\Omega_{\Lambda,0} + (1 - \Omega_{m,0} - \Omega_{\Lambda,0})\,(1+z)^2 + \Omega_{m,0}\,(1+z)^3\right]^{1/2}. \tag{27}$$

The bias parameter of the dark halo with a mass of M, $b_{\rm h}(M)$ is given by,

$$b_{\rm h}(M) = 1 + \frac{1}{\delta_{\rm c}}\left[\nu^2 + b\nu^{2(1-c)} - \frac{\nu^2/\sqrt{a}}{\nu^{2c} + b(1-c)(1-c/2)}\right], \tag{28}$$

where ν is defined by,

$$\nu = \frac{\sqrt{a}\delta_{\rm c}}{D(z)\sigma(M)}, \tag{29}$$

and a, b, c, and $\delta_{\rm c}$ are assumed to be $a = 0.707$, $b = 0.5$, $c = 0.6$, and $\delta_{\rm c} = 1.69$, respectively (Sheth et al. 2001, see also Section 2.2.3 of this thesis). The term $\sigma(M)$ represents the root mean square of the density field fluctuation with a mass scale M, i.e.,

$$\sigma^2(M) = \frac{1}{2\pi^2}\int_0^\infty dk\, k^2 P_m(k)\tilde{W}^2(kR), \tag{30}$$

and the redshift dependence is included via the linear growth factor, $D(z)$. The term $P_m(k)$ is the matter power spectrum at $z = 0$ given by,

$$P_m(k) = P_m(k,t)|_{t=0} = A_0 D^2(z=t) k^{n_{\rm s}} T^2(k,t)\big|_{t=0}\,, \tag{31}$$

where A_0 is a normalization parameter determined by the value of σ_8, $D(z = t)$ is the linear growth factor at a redshift that corresponds to cosmic time t, and $T(k, t)$ is a transfer function, which determines the shape of the initial power spectrum. In this study, I employ fitting formulae of the transfer function proposed by Eisenstein & Hu (1998). The term $\tilde{W}(x)$ is a top-hat window function, i.e.,

$$\tilde{W}(x) = \frac{3\,(\sin x - x\cos x)}{x^3}, \tag{32}$$

and R is a Lagrangian radius of the dark halo, which gives the dark halo mass dependence, i.e.,

$$R(M) = \left(\frac{3M}{4\pi\bar{\rho}_0}\right)^{1/3}, \tag{33}$$

where $\bar{\rho_0}$ represents the current mean density of the Universe. Behaviors of $\sigma(M)$ and $P_m(k, z)$ are shown in Figure 9 and Figure 10.

One can estimate the mean dark halo mass, $\langle M_{\rm h}\rangle$, and the minimum dark halo mass, $M_{\rm min}$, from a comparison between the effective bias parameter and the bias parameter of the dark halo (see also Hayashi et al. 2007). The mean dark halo mass is defined when the bias of the dark halo with a mass of $\langle M_{\rm h}\rangle$ is equal to the effective bias, as

$$b_{\rm eff} = b_{\rm h}(\langle M_{\rm h}\rangle), \tag{34}$$

whereas the minimum dark halo mass, when the effective bias parameter from equation (21) is equal to another expression of the effective bias, is defined as,

$$b_{\rm eff} = \frac{\displaystyle\int_{M_{\rm min}}^\infty dM_{\rm h}\, b_{\rm h}(M_{\rm h}) n(M_{\rm h})}{\displaystyle\int_{M_{\rm min}}^\infty dM_{\rm h}\, n(M_{\rm h})}, \tag{35}$$

where $n(M_{\rm h})$ is the halo mass function.

2.2. Analytical Dark Halo Models

In this section, I define some analytical halo models to represent the properties of dark haloes before introducing the HOD model. The HOD formalism is established under ansatz that all of galaxies are formed within dark haloes and the following halo models.

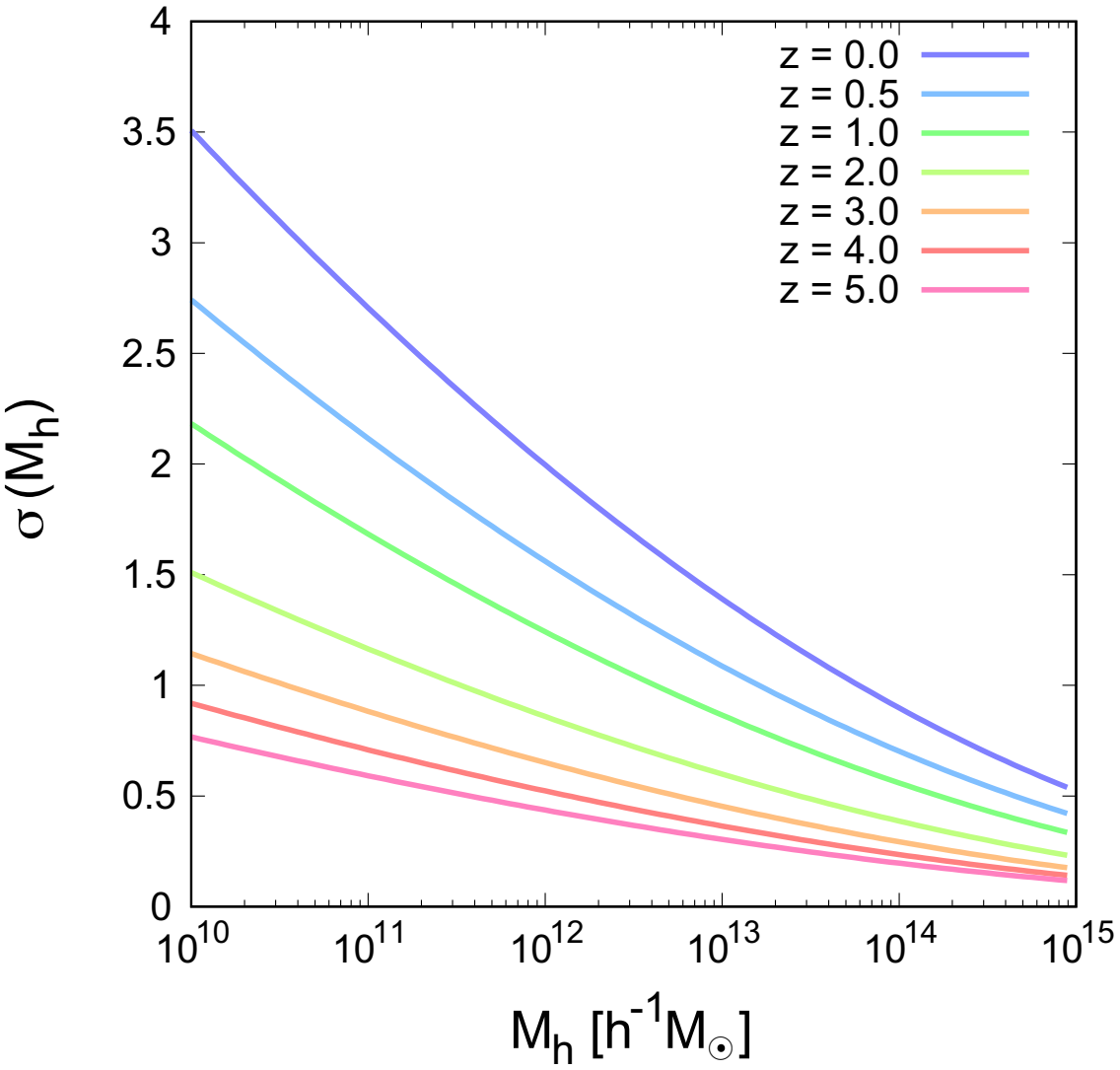

Figure 9.— The redshift dependence of the root mean square of the density field fluctuation as a function of the dark halo mass, $\sigma(M_{\rm h})$. $\sigma(M_{\rm h})$ is calculated by the equation (30). Cosmological parameters are assumed the Planck cosmologies.

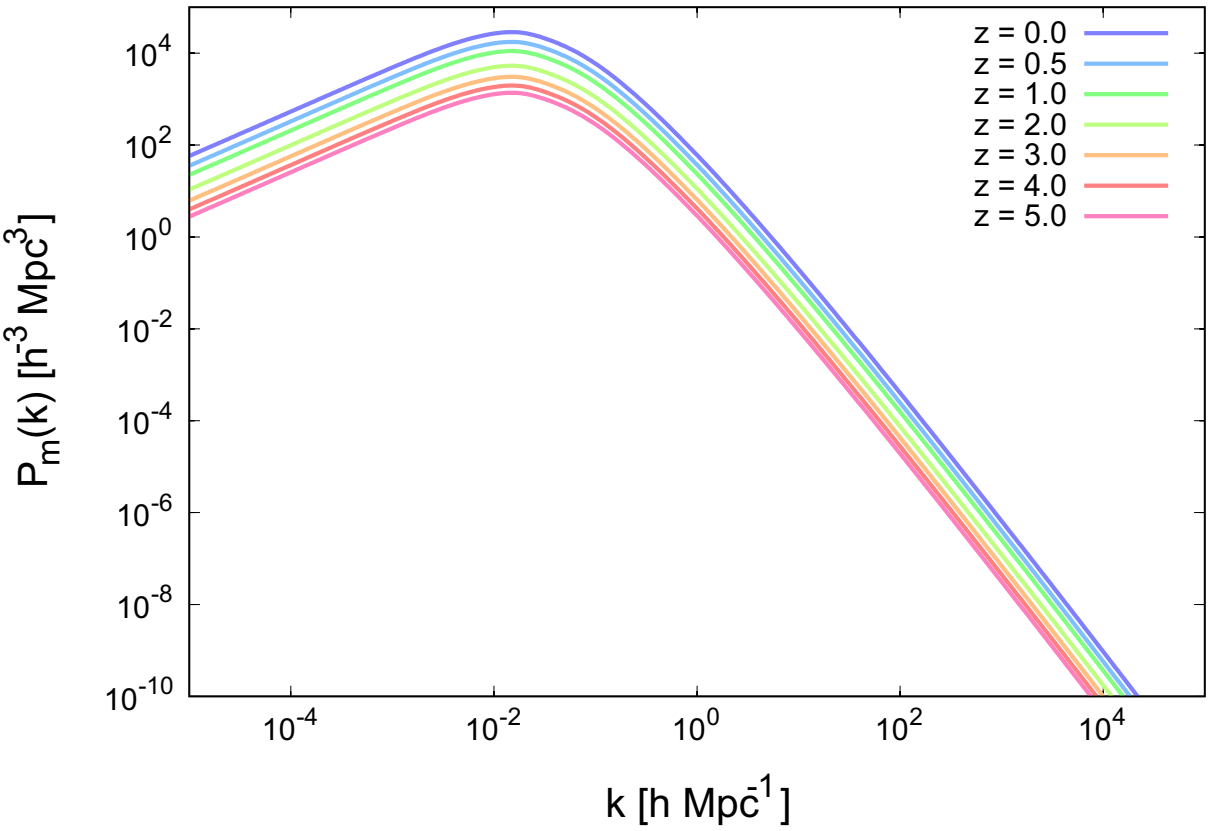

Figure 10.— The redshift dependence of matter power spectrum, $P_m(k)$, assuming the transfer function of Eisenstein & Hu (1998). $P_m(k)$ is calculated by the equation (31). Cosmological parameters are assumed the Planck cosmologies.

2.2.1. Halo mass function

In describing the dark halo by analytical formulae, the halo mass function has been well studied. The halo mass function represents the number density of dark haloes as a function of the dark halo mass at a given redshift. The first successful work of describing the halo mass function is a formalism of Press & Schechter (1974), which combines the linear perturbation theory that can be treated easily with the non-linear regime to estimate the development of the structures of the Universe precisely. Press & Schechter (1974) formalism assumes that the initial density fluctuation follows a Gaussian distribution. Dark matter around the peak of the density fluctuation collapses with spherical symmetry and forms a dark halo if the peak exceeds the critical density fluctuation, δ_c.

The probability distribution that forms a dark halo at a given redshift , $P(\delta > \delta_c|z)$, can be describes as,

$$P\left(\delta > \delta_c|z\right) = \int_{\delta_c}^{\infty} d\delta_{M_\mathrm{h}} P\left(\delta_{M_\mathrm{h}}|z\right), \tag{36}$$

where $P(\delta_{M_\mathrm{h}}|z)$ is a probability distribution of the density fluctuation with $\delta \in [\delta_{M_\mathrm{h}}; \delta_{M_\mathrm{h}} + d\delta_{M_\mathrm{h}}]$ at an arbitrary redshift as,

$$P\left(\delta_{M_\mathrm{h}}|z\right) = \frac{1}{\sqrt{2\pi\sigma^2(M_\mathrm{h})}} \exp\left(-\frac{\delta^2_{M_\mathrm{h}}}{2\sigma^2(M_\mathrm{h})}\right) d\delta_{M_\mathrm{h}}. \tag{37}$$

It is noted that $\sigma(M_\mathrm{h})$ can be calculated by the equation (30). The $P(\delta_{M_\mathrm{h}}|z)$ keeps Gaussianity because of the assumption of the initial Gaussian distribution. The Press & Schechter formalism evaluates the number density of dark haloes by multiplying $P(\delta > \delta_c|z)$ by a factor of 2 because of the reason known as a "cloud-in-cloud problem". The reason of multiplying the factor of 2 is as follows: in the probability distribution of $P(\delta > \delta_c|z)$, the region with negative density fluctuation cannot collapse and form dark haloes; however, negative regions can form dark haloes by taken into the collapsed objects that originate from the regions of positive density fluctuations. Therefore, one should take the negative density fluctuation regions into consideration of the formation of dark haloes and multiply a factor of 2 for regarding the probability distribution of $P(\delta > \delta_c|z)$ as the probability distribution of dark haloes with a halo mass of M_h.

From the above discussion, the fraction of dark haloes with mass of more than M_h, $f(M > M_\mathrm{h}\,|\,z)$, can be written as,

$$f(M > M_\mathrm{h}\,|\,z) = 2\int_{\delta_c}^{\infty} d\delta_{M_\mathrm{h}} P\left(\delta_{M_\mathrm{h}}|z\right) = \frac{2}{\sqrt{\pi}} \int_{\delta_c/\sqrt{2}\sigma(M_\mathrm{h})}^{\infty} dx \exp\left(-x^2\right). \tag{38}$$

The number density of the dark halo at a given redshift can be derived by differentiating $f(M > M_\mathrm{h}\,|\,z)$ with respect to the halo mass and multiplying the mean number density of dark halo, $\bar{\rho_0}/M_\mathrm{h}$, as,

$$\begin{aligned} n(M_\mathrm{h})dM_\mathrm{h} &= \sqrt{\frac{2}{\pi}} \frac{\bar{\rho_0}}{M^2_\mathrm{h}} \frac{\delta_c}{\sigma(M_\mathrm{h})} \left|\frac{\partial \ln \sigma(M_\mathrm{h})}{\partial \ln M_\mathrm{h}}\right| \exp\left(-\frac{\delta^2_c}{2\sigma^2(M_\mathrm{h})}\right) \\ &= \sqrt{\frac{2}{\pi}} \frac{\bar{\rho_0}}{M^2_\mathrm{h}} \nu \left|\frac{\partial \ln \sigma(M_\mathrm{h})}{\partial \ln M_\mathrm{h}}\right| \exp\left(-\frac{\nu^2}{2}\right), \end{aligned} \tag{39}$$

where ν is defined as $\nu \equiv \delta_c/\sigma(M_\mathrm{h})$.

The Press–Schechter formalism can relatively well describe the number density of the dark halo and has been accepted its usefulness; however, it is known that the halo mass function based upon the

Press & Schechter formalism cannot represent the results of the cosmological N-body simulations: the halo mass function of Press & Schechter (1974) overestimates (underestimates) the number density of the dark halo below (beyond) the "knee" of the halo mass function compared to the N-body simulation. To address these discrepancies, some "extended" Press–Schechter formalisms have been proposed (e.g., Sheth & Tormen 1999; Jenkins et al. 2001; Tinker et al. 2008; Behroozi et al. 2013b).

Sheth & Tormen (1999) improved the Press–Schechter formalism by introducing the ellipsoidal collapse model and proposed the functional form of their halo mass function as,

$$n(M_{\rm h})dM_{\rm h} = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}_0}{M_{\rm h}^2}A\left(1+\left(a\nu^2\right)^{-p}\right)\sqrt{a\nu^2}\left|\frac{\partial \ln\sigma(M_{\rm h})}{\partial \ln M_{\rm h}}\right|\exp\left(-\frac{a\nu^2}{2}\right), \tag{40}$$

where A is a normalization parameter that is determined by the following condition:

$$\int n(M_{\rm h})\frac{M_{\rm h}}{\bar{\rho}_0}dM_{\rm h} = 1. \tag{41}$$

In the formalism of Sheth et al. (2001), the parameters of a and p are assumed $a = 1/\sqrt{2}$ and $p = 0.3$, and the normalization parameter is $A \approx 0.3222$, respectively. The halo mass function of Sheth & Tormen (1999) is well consistent with the results of the N-body simulations (refer to Figure 2 of Springel et al. 2005). The redshift evolution of the Sheth–Tormen halo mass function is presented in Figure 11.

In this thesis, I assume the halo mass function of Sheth & Tormen at any redshift.

2.2.2. Density profile of the dark halo

Following the formulation of the gravitational collapse model of Sheth & Tormen (1999), dark haloes are formed and identified around the local peaks of the density field. The center of the dark halo should be the most concentrated part and the density decrease with increasing the distance from its center.

Navarro et al. (1997) presented a radial density profile of dark haloes, $\rho(r|M_{\rm h})$, by the cosmological N-body simulation known as the "NFW profile" as follows:

$$\rho(r|M_{\rm h}) = \frac{\rho_{\rm s}}{(r/r_{\rm s})(1+r/r_{\rm s})^2}, \tag{42}$$

where $r_{\rm s}$ is a scale radius and $\rho_{\rm s}$ is a density with radius of $r_{\rm s}$. Total halo mass can be written using the virial radius, $r_{\rm vir}$ as

$$M = \int_0^{r_{\rm vir}} dr 4\pi r^2 \rho(r). \tag{43}$$

Assuming the NFW profile, above total mass is as follows (cf., Cooray & Sheth 2002; Takada & Jain 2003):

$$M = \frac{4\pi\rho_{\rm s}r_{\rm vir}^3}{c^3}\left[\ln(1+c) - \frac{c}{1+c}\right], \tag{44}$$

where c is a concentration parameter defined as $c = r_{\rm vir}/r_{\rm s}$. I adopt the following relation of the concentration parameter (Bullock et al. 2001):

$$c\,(M,z) = \frac{c_0}{1+z}\left(\frac{M}{M_*\,(z)}\right)^{\beta}, \tag{45}$$

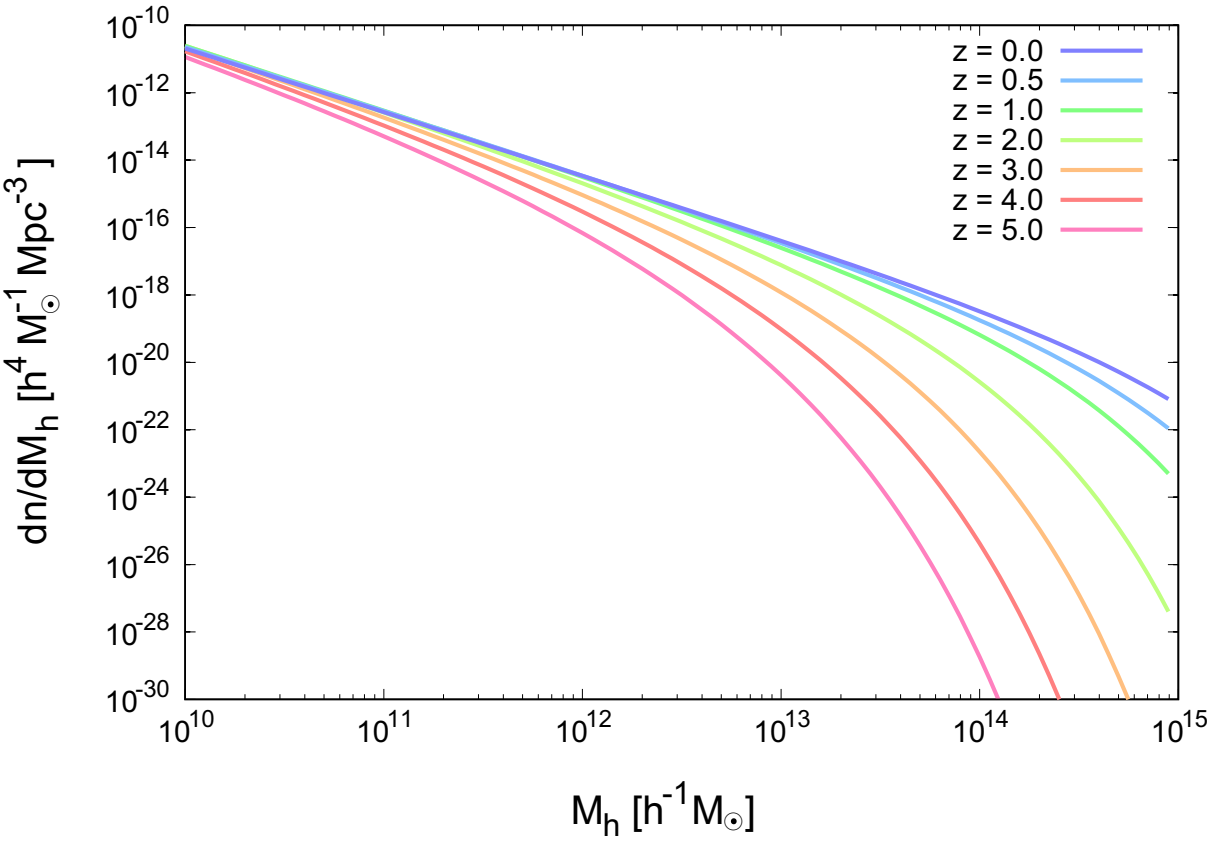

Figure 11.— The redshift evolution of the halo mass function proposed by Sheth & Tormen (1999) calculated by the equation (40). The parameters of a and p in the equation (40) are adopted as $a = 1/\sqrt{2}$ and $p = 0.3$, respectively. Cosmological parameters are assumed the Planck cosmologies.

where $M_*(z)$ is a characteristic mass scale defined as

$$\nu\left(M_*, z=0\right) = \frac{\delta_c(z=0)}{D(z=0)\sigma(M_*)} = 1. \tag{46}$$

I choose the parameter set of c_0 and β as $(c_0, \beta) = (11, -0.13)$ that is the same values of Zehavi et al. (2011) and Coupon et al. (2012).

The HOD model computes the galaxy clustering from the galaxy power spectrum in Fourier space; thus, it is required the Fourier-transformed expression of the NFW profile. Following the Cooray & Sheth (2002), the normalized density profile in Fourier space, $u(k|M_{\rm h})$, is

$$u(k|M_{\rm h}) = \frac{\int dr \rho(r|M_{\rm h}) e^{-ikr}}{\int dr \rho(r|M_{\rm h})}. \tag{47}$$

Under the assumption of the spherically symmetric density profile, it can be written as

$$u(k|M_{\rm h}) = \int_0^{r_{\rm vir}} dr 4\pi r^2 \frac{\sin kr}{kr} \frac{\rho\left(r|M_{\rm h}\right)}{M_{\rm h}}. \tag{48}$$

For the NFW profile, it can be written as follows:

$$\begin{aligned} u(k|M_{\rm h}) &= \frac{4\pi\rho_{\rm s} r_{\rm vir}^3}{c^3 M_{\rm h}} \Bigg\{ \sin\left(kr_{\rm vir}/c\right) \Big[{\rm Si}\left(\left[1+1/c\right] kr_{\rm vir}\right) - {\rm Si}\left(kr_{\rm vir}/c\right) \Big] - \frac{kr_{\rm vir}}{\left(1+1/c\right) kr_{\rm vir}} \\ &+ \cos\left(kr_{\rm vir}/c\right) \Big[{\rm Ci}\left(\left[1+1/c\right] kr_{\rm vir}\right) - {\rm Ci}\left(kr_{\rm vir}/c\right) \Big] \Bigg\}. \end{aligned} \tag{49}$$

The expressions of ${\rm Si}(x)$ and ${\rm Ci}(x)$ are the sine integral and the cosine integral defined as

$${\rm Si}(x) = \int_0^x dt \frac{\sin\left(t\right)}{t} \tag{50}$$

and

$${\rm Ci}(x) = -\int_x^\infty dt \frac{\cos\left(t\right)}{t}, \tag{51}$$

respectively. I present the halo mass dependence of the NFW profile in Fourier space in Figure 12.

2.2.3. Large-scale halo bias

As I mentioned in Section 2.2.1, dark haloes are formed by the spherically (or ellipsoidal) collapse at the peak of the density fluctuation, known as the "peak-background split" (e.g., Cole & Kaiser 1989; Mo & White 1996). Thus, dark haloes are the biased tracers of the invisible underlying dark matter distribution and the biased galaxy clustering reflects the bias of their host haloes. Clustering of dark haloes is largely governed by the halo bias; the large-scale halo bias is one of the most important parameters to determine the galaxy-galaxy clustering with large separation angle, i.e., the 2-halo term of the HOD formalism (see Section 2.3.3).

The large-scale halo bias, $b_{\rm h}(M_{\rm h})$, represents the difference of the fluctuation between dark haloes and underlying dark matter, i.e.,

$$\frac{\delta n_{\rm DM}(M_{\rm h})}{n_{\rm DM}(M_{\rm h})} = \left[1 + b_{\rm h}(M_{\rm h})\right] \delta, \tag{52}$$

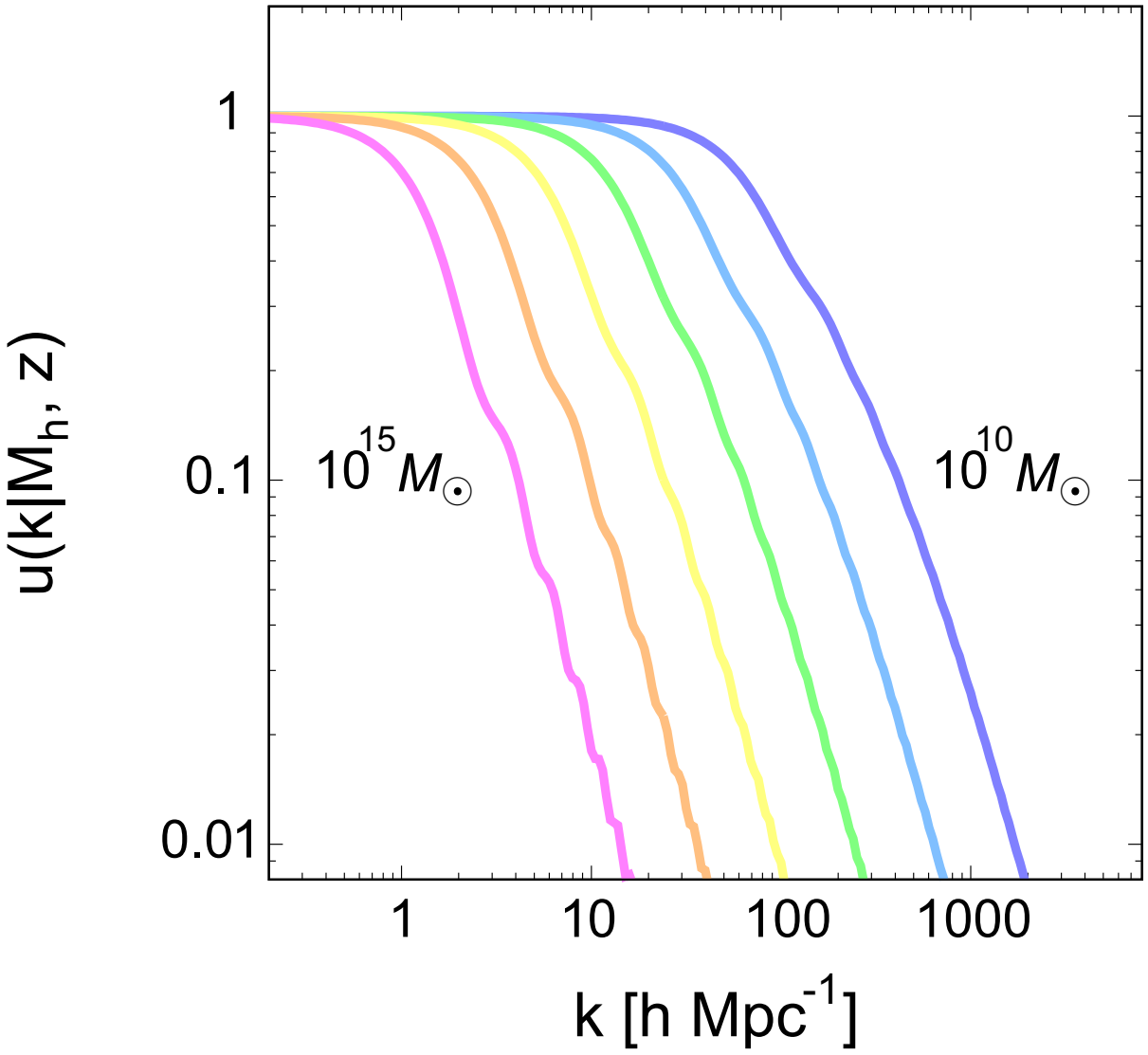

Figure 12.— The halo mass dependence of the NFW profile of dark haloes in Fourier space as a function of the wavenumber. Profiles of dark hales from less massive ones ($M_{\rm h} = 10^{10} M_{\odot}$; blue) to massive ones ($M_{\rm h} = 10^{15} M_{\odot}$; red) are shown.

where $n_{\rm DM}(M_{\rm h})$ is a number density of dark haloes with mass $M_{\rm h}$ and δ is a density fluctuation of underlying dark matter. The halo bias parameter has been drawn by assuming the halo mass function of the Press–Schechter expression (Cole & Kaiser 1989; Mo & White 1996) as

$$b_{\rm h}(M_{\rm h}) = 1 + \frac{\nu^2 - c}{\delta_{\rm c}}, \tag{53}$$

where ν is defined as $\nu = \delta_{\rm c}/\sigma(M_{\rm h})$ and $\delta_{\rm c}$ is a critical density to form dark haloes by collapse. Sheth et al. (2001) improved this formulation by introducing the halo mass function of Sheth & Tormen (1999) for equation (28), although Tinker et al. (2005) updated the parameter set as $(a, b, c) = (0.707, 0.35, 0.80)$ based upon the results of their numerical simulation. Additionally, Tinker et al. (2005) introduced the scale-dependent term for the halo bias:

$$b_{\rm h}^2(M_{\rm h}, r) = b_{\rm h}^2(M_{\rm h}) \frac{[1 + 1.17\xi_m(r)]^{1.49}}{[1 + 0.69\xi_m(r)]^{2.09}}, \tag{54}$$

where $\xi_m(r)$ is a matter correlation function in real space. Throughout this thesis, I adopt the modified scale-dependent halo bias proposed by Tinker et al. (2012):

$$b_{\rm h}^2(M_{\rm h}, r) = \begin{cases} b_{\rm h}^2(M_{\rm h}) \dfrac{[1 + 1.17\xi_m(r)]^{1.49}}{[1 + 0.69\xi_m(r)]^{2.09}} & \text{if} \quad r \geq 2R_{\rm halo} \\ \\ b_{\rm h}^2(M_{\rm h}) \dfrac{[1 + 1.17\xi_m(2R_{\rm halo})]^{1.49}}{[1 + 0.69\xi_m(2R_{\rm halo})]^{2.09}} & \text{if} \quad r < 2R_{\rm halo}, \end{cases} \tag{55}$$

where $R_{\rm halo}$ represent the radius of the focusing halo. This halo bias model is the precisely described expression of equation (54) for the overlapped dark haloes along the line of sight of which the center of one dark halo is outside of another dark halo, i.e., $r \geq R_{1,\rm halo} + R_{2,\rm halo}$. On the other hand, the dark halo pair whose center of one dark halo is within another dark halo, i.e., $r < R_{1,\rm halo} + R_{2,\rm halo}$, satisfies $b_{\rm h}(M_{\rm h}, z) = 0$ from the halo exclusion model (Zheng 2004).

2.3. The Halo Occupation Distribution Formalism

2.3.1. Overview of the HOD model

In this Section, I will present the briefly introduction of a halo occupation distribution (HOD; e.g., Seljak 2000; Ma & Fry 2000; Berlind & Weinberg 2002; Cooray & Sheth 2002; Berlind et al. 2003; van den Bosch et al. 2003a; Kravtsov et al. 2004; Hamana et al. 2004) model. The HOD model has been developed to connect observable galaxies to the background dark matter distribution. The HOD formalism describes the galaxy distribution within the host dark halo by characterizing them from the perspective of the probability distribution, $P(N|M)$, that a halo of virial mass M contains N galaxies with specific physical properties, such as color, stellar mass, or galaxy population. One of the advantages of the HOD formalism is that the HOD parameters have explicit physical meanings; this enables us to easily interpret the relationship between the galaxies and the host haloes (see Section 2.3.3).

To reveal the structure formation of the Universe, it is important to determine the amplitudes the dark matter density fluctuation at various physical scales via power spectrum, $P(k)$, which describes

the amplitude of fluctuation in Fourier space as a function of the wavenumber. Seljak (2000) developed the analytical model of the dark matter power spectra at a scale of the non-linear regime as well as a scale of the linear perturbation based upon the formalism of Press & Schechter (1974), which describes the number density of the virialized dark haloes at an arbitrary cosmic time. Additionally, Seljak (2000) extended their dark matter power spectrum to the galaxy power spectrum by introducing a new parameter for describing the mean number of galaxies within the vilialized dark halo. Berlind & Weinberg (2002) investigated the validity of the predicted galaxy clustering based upon the model galaxy power spectrum (e.g., Seljak 2000; Ma & Fry 2000) by comparing the results of the cosmological N-body simulation, and organized their theoretical framework of galaxy distribution and clustering from the aspect of the fluctuation of dark matter and termed it "halo occupation distribution" model. After the establishment of the HOD formalism, van den Bosch et al. (2003a) and Hamana et al. (2004) applied the HOD model to the observed galaxy correlation function and derived the properties of host dark haloes of their galaxy samples.

Details of the HOD analyses and those results at each redshift are summarized in the first part of Section 3, 4, and 5.

2.3.2. Galaxy distribution within the dark halo

One of the biggest characteristics of the HOD formalism is to formulate the number of galaxies within dark haloes as a function of the halo mass. In a very early halo occupation model that is constructed by Berlind & Weinberg (2002) and Hamana et al. (2004), the number of total occupied galaxies within a dark halo, i.e., the halo occupation function, $N_{\rm tot}^{\rm Hamana}(M_{\rm h})$, is assumed to follow the power law as,

$$N_{\rm tot}^{\rm Hamana}(M_{\rm h}) = \begin{cases} 0 & (M_{\rm h} < M_{\rm min}) \\ \left(\dfrac{M_{\rm h}}{M_1}\right)^{\alpha} & (M_{\rm h} \geq M_{\rm min})\,. \end{cases} \tag{56}$$

This functional form contains three free parameters: $M_{\rm min}$, M_1, and α. One can interpret the physical meanings of these HOD free parameters; $M_{\rm min}$ is a threshold dark halo mass to be able to possess a galaxy stochastically, M_1 is a typical dark halo mass to be able to possess one galaxy, and α is a galaxy formation efficiency. We note that galaxies are not distinguished central galaxy from satellite galaxies in this early halo occupation function. The behavior of this functional form is shown in Figure 13.

This simple, very early halo occupation function is widely used and gives clues to connect simple physical quantities such as dark halo masses to complex baryonic properties such as galaxy distributions within dark haloes (e.g., Ouchi et al. 2005; Hamana et al. 2006; Kovač et al. 2007; Orsi et al. 2008; Quadri et al. 2008). However, recent studies hardly adopt this halo occupation model because of the lack of the possibility of the existence of central galaxies within less massive dark haloes for this model. Therefore, some more realistic improved halo occupation models have been proposed to well describe the galaxy distribution within dark haloes.

Kravtsov et al. (2004) and Zehavi et al. (2005) have developed a halo occupation model that separates the contribution of the central galaxy and the satellite galaxy. The occupation function of

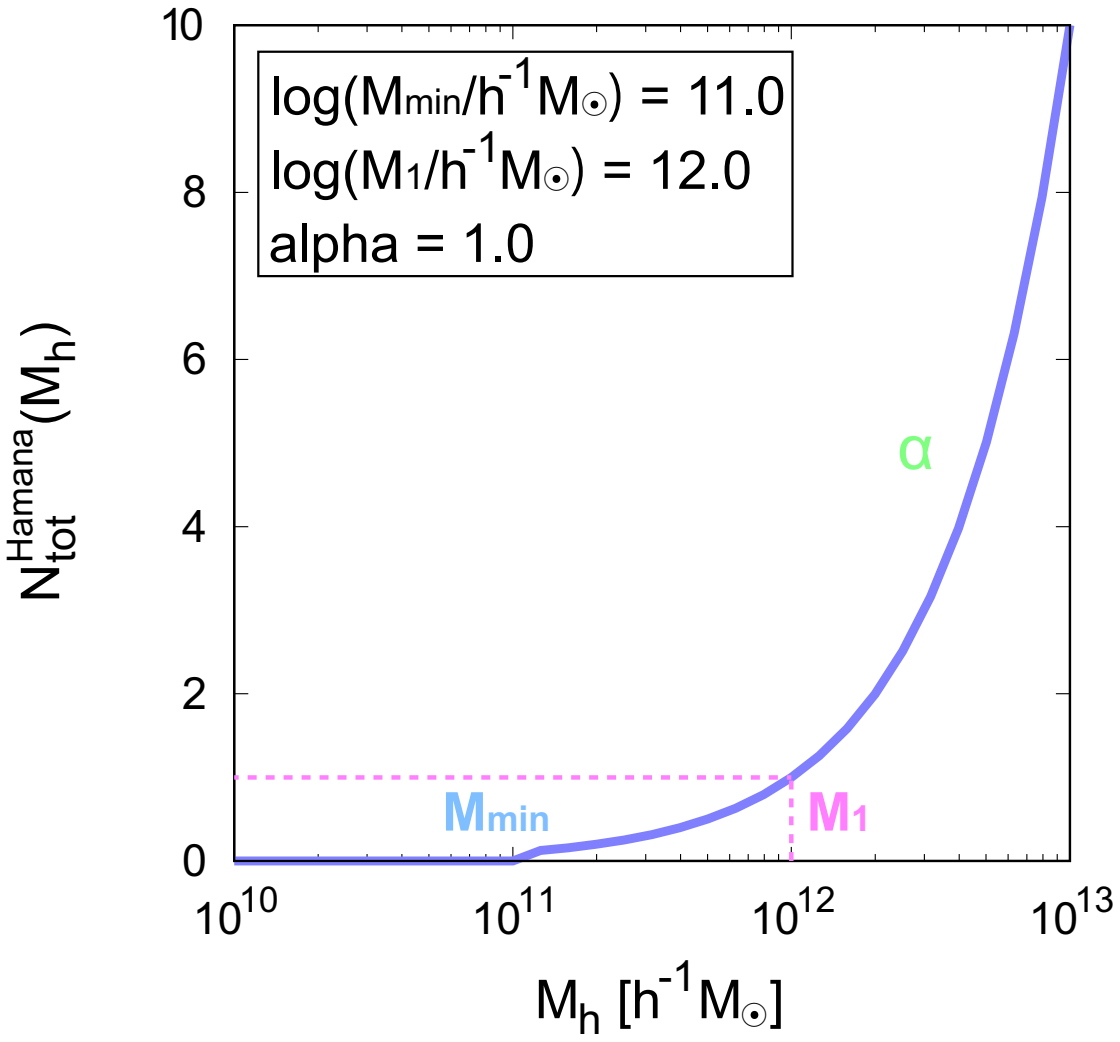

Figure 13.— The relationship between the number of occupied galaxies within the dark halo and the mass of the dark halo. This relation is calculated by the halo occupation function of the equation (56) that is proposed by Berlind & Weinberg (2002) and Hamana et al. (2004). The HOD parameters are assumed as $\log(M_{\rm min}/h^{-1}M_\odot) = 11.0$, $\log(M_1/h^{-1}M_\odot) = 12.0$, and $\alpha = 1.0$. The blue solid line is the number of total galaxies including both the central and the satellite galaxy because this model does not distinguish those galaxy types.

the central galaxy, $N_{\rm cen}^{\rm Zehavi}(M_{\rm h})$, and the satellite galaxy, $N_{\rm sat}^{\rm Zehavi}(M_{\rm h})$, are as follows:

$$N_{\rm cen}^{\rm Zehavi}(M_{\rm h}) = \begin{cases} 0 & (M_{\rm h} < M_{\rm min}) \\ \\ 1 & (M_{\rm h} \geq M_{\rm min})\,, \end{cases} \tag{57}$$

and

$$N_{\rm sat}^{\rm Zehavi}(M_{\rm h}) = \left(\frac{M_{\rm h}}{M_1}\right)^{\alpha}. \tag{58}$$

The total number of occupied galaxies, $N_{\rm tot}^{\rm Zehavi}(M_{\rm h})$ is given by

$$N_{\rm tot}^{\rm Zehavi}(M_{\rm h}) = N_{\rm cen}^{\rm Zehavi}(M_{\rm h}) \left[1 + N_{\rm sat}^{\rm Zehavi}(M_{\rm h})\right]. \tag{59}$$

Their occupation model of the central galaxy is adopted a Heaviside step function because the number of central galaxy cannot exceed the unity. The HOD free parameters of this occupation model is the same as those of equation (56); however, the physical meanings among them are slightly different. In this model, $M_{\rm min}$ is a threshold dark halo mass to possess a central galaxy, M_1 is also a threshold dark halo mass to possess a satellite galaxy (i.e., one central galaxy and one satellite galaxy are contained within the dark halo with mass of M_1, in total), and α is a formation efficiency of the satellite galaxies. The shape of this occupation function is presented in Figure 14.

Zheng et al. (2005) also proposed the halo occupation function that distinguishes the central galaxies from the satellite galaxies like the formalism of Zehavi et al. (2005). The characteristic of the formalism of Zheng et al. (2005) is introducing a smoothed cutoff parameter for the central galaxy occupation, which originates from the results of the semi-analytical model and the smoothed particle hydrodynamics (SPH) simulation (refer to Figure 3 of Zheng et al. 2005). The functional form of the central galaxy occupation, $N_{\rm cen}^{\rm Zheng}(M_{\rm h})$, developed by Zheng et al. (2005) is as follows:

$$N_{\rm cen}^{\rm Zheng}(M_{\rm h}) = \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log(M_{\rm h}) - \log(M_{\rm min})}{\sigma_{\log M}}\right)\right]. \tag{60}$$

The notation of ${\rm erf}(x)$ represents the error function that satisfies ${\rm erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x dt \exp\left(-t^2\right)$. The satellite galaxy occupation, $N_{\rm sat}^{\rm Zheng}(M_{\rm h})$, can be expressed as

$$N_{\rm sat}^{\rm Zheng}(M_{\rm h}) = \left(\frac{M_{\rm h} - M_0}{M_1}\right)^{\alpha}, \tag{61}$$

provided that $M_{\rm h}$ satisfies $M_{\rm h} \geq M_0$. The total number of occupied galaxies, $N_{\rm tot}^{\rm Zheng}(M_{\rm h})$, can be described as

$$N_{\rm tot}^{\rm Zheng}(M_{\rm h}) = N_{\rm cen}^{\rm Zheng}(M_{\rm h}) \left[1 + N_{\rm sat}^{\rm Zheng}(M_{\rm h})\right]. \tag{62}$$

The occupation function of Zheng et al. (2005) is tuned by five HOD free parameters: $M_{\rm min}$, M_1, M_0, $\sigma_{\log M}$, and α. The physical meanings of these parameters can be interpreted as follows: $M_{\rm min}$ is a threshold dark halo mass to possess a central galaxy with a 0.5 possibility, M_1 is an approximately typical dark halo mass to possess one satellite galaxy (the reason of "approximately" is due to the subtraction of M_0), M_0 is a threshold dark halo mass to be able to possess a satellite galaxy, $\sigma_{\log M}$ is a smoothed cutoff parameter that controls the width of the scatter between the halo mass and the

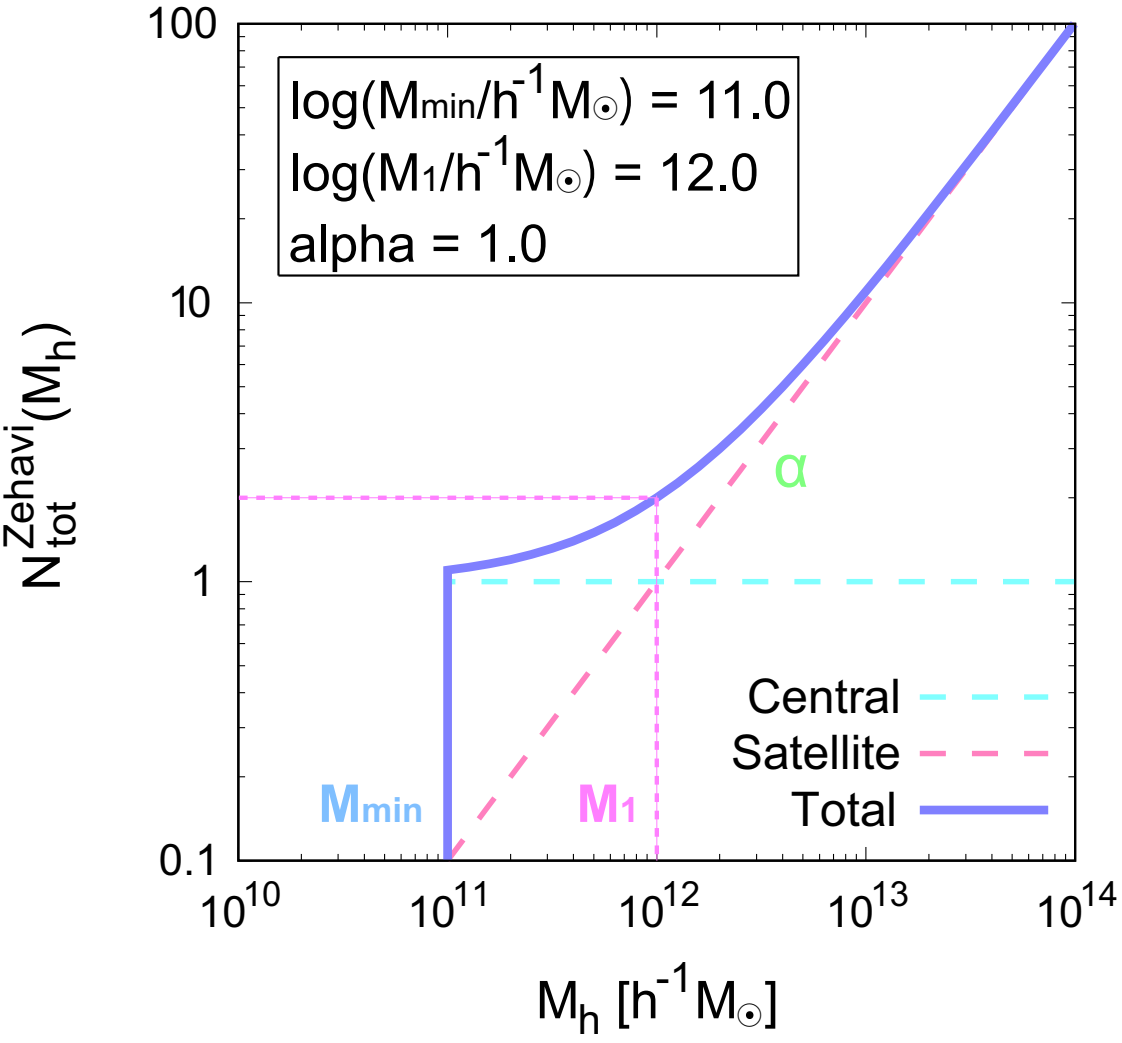

Figure 14.— The relationship between the number of occupied galaxies within the dark halo and the mass of the dark halo. The dashed cyan line is the occupation function of the central galaxy, $N_{\rm cen}^{\rm Zehavi}(M_{\rm h})$, whose shape is the Heaviside step function, the dashed red line is the occupation function of the satellite galaxy, $N_{\rm sat}^{\rm Zehavi}(M_{\rm h})$, and the solid blue line represents the total occupied galaxies, $N_{\rm tot}^{\rm Zehavi}(M_{\rm h})$. This relation is calculated by the halo occupation function of the equation (59) that is proposed by Kravtsov et al. (2004) and Zehavi et al. (2005). The HOD parameters are assumed as $\log(M_{\rm min}/h^{-1}M_\odot) = 11.0$, $\log(M_1/h^{-1}M_\odot) = 12.0$, and $\alpha = 1.0$.

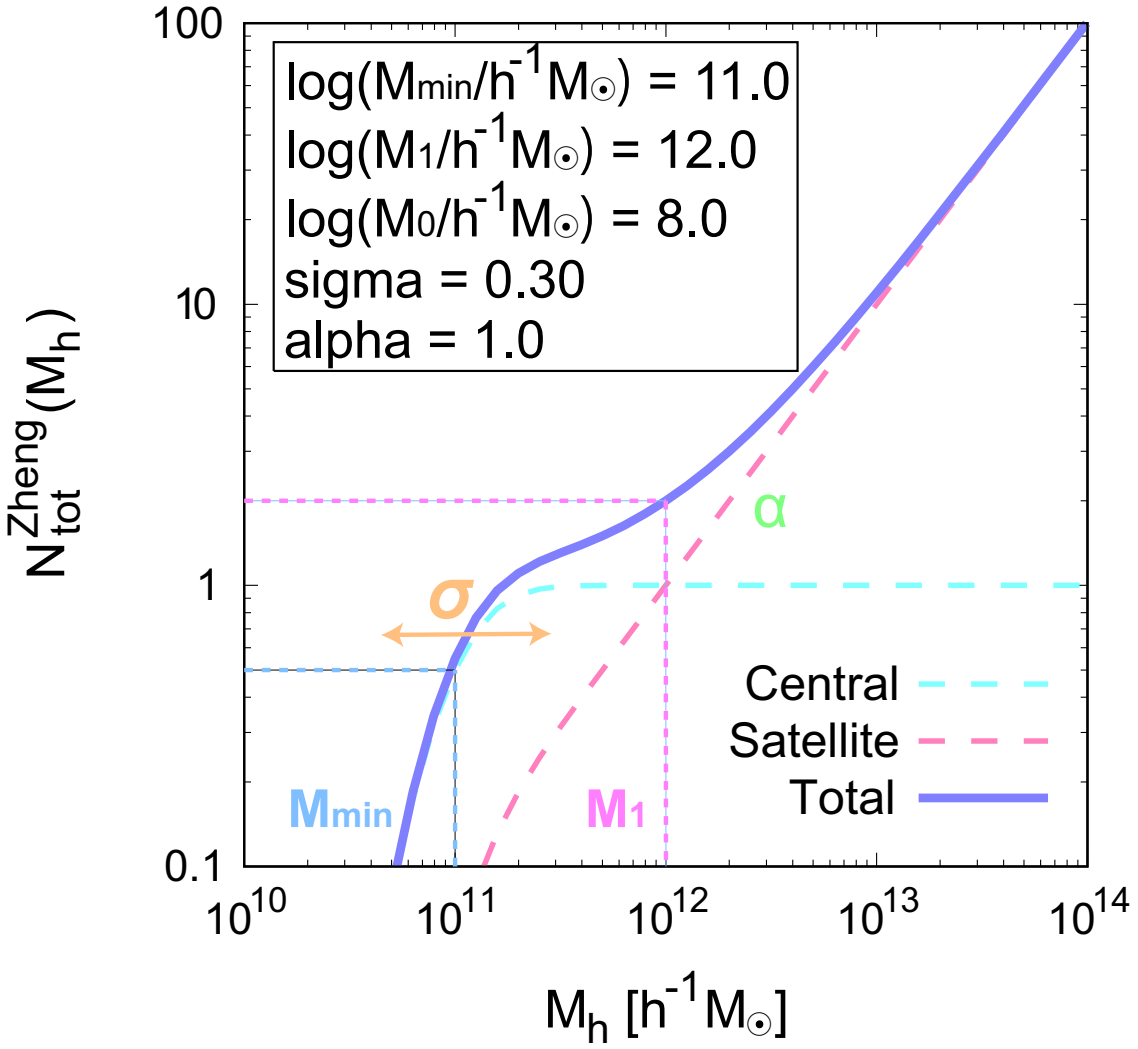

Figure 15.— The relationship between the number of occupied galaxies within the dark halo and the mass of the dark halo. The dashed cyan line is the occupation function of the central galaxy, $N_{\rm cen}^{\rm Zheng}(M_{\rm h})$, whose shape is the error function, the dashed red line is the occupation function of the satellite galaxy, $N_{\rm sat}^{\rm Zheng}(M_{\rm h})$, and the solid blue line represents the total occupied galaxies, $N_{\rm tot}^{\rm Zheng}(M_{\rm h})$. This relation is calculated by the halo occupation function of the equation (62) that is proposed by Zheng et al. (2005). The HOD parameters are assumed as $\log(M_{\rm min}/h^{-1}M_\odot) = 11.0$, $\log(M_1/h^{-1}M_\odot) = 12.0$, $\log(M_0/h^{-1}M_\odot) = 8.0$, $\sigma_{\log M} = 0.30$, and $\alpha = 1.0$.

galaxy baryonic properties (e.g., stellar masses, magnitudes, and luminosities), and α is a formation efficiency of the satellite galaxies. A schematic view of this occupation model is given in Figure 15.

Recently, Geach et al. (2012) presented a new occupation function to describe the complicated distribution of Hα emitters (HAE) predicted by the GALFORM (Cole et al. 2000) simulation (Bower et al. 2006; Font et al. 2008; Lagos et al. 2011). To represent the composite occupation of the Gaussian-like distribution and the step function with smoothed cutoff of the central galaxy, and the occupation of the Poisson distribution of the satellite galaxies at high-mass end, Geach et al. (2012) proposed the occupation functions of the central galaxy, $N_{\rm cen}^{\rm Geach}(M_{\rm h})$, and the satellite galaxy, $N_{\rm sat}^{\rm Geach}(M_{\rm h})$, as follows:

$$N_{\rm cen}^{\rm Geach}(M_{\rm h}) = F_c^B\left(1-F_c^A\right)\exp\left[-\frac{\log\left(M_{\rm h}/M_c\right)^2}{2\sigma_{\log M}^2}\right] + F_c^A\left[1+{\rm erf}\left(\frac{\log\left(M_{\rm h}/M_c\right)}{\sigma_{\log M}}\right)\right], \quad (63)$$

and

$$N_{\rm sat}^{\rm Geach}(M_{\rm h}) = F_s\left[1+{\rm erf}\left(\frac{\log\left(M_{\rm h}/M_{\rm min}\right)}{\delta_{\log M}}\right)\right]\left(\frac{M_{\rm h}}{M_{\rm min}}\right)^{\alpha}. \quad (64)$$

The total number of galaxies of this formalism, $N_{\rm tot}^{\rm Geach}(M_{\rm h})$, is given by

$$N_{\rm tot}^{\rm Geach}(M_{\rm h}) = N_{\rm cen}^{\rm Geach}(M_{\rm h}) + N_{\rm sat}^{\rm Geach}(M_{\rm h}). \quad (65)$$

The characteristic of this formalism is that both the central and the satellite galaxy occupation consists of two components: at the low-mass end, the central (satellite) galaxy occupies the haloes as the Gaussian-like (step-like) distribution, and the step-function (Poisson) distribution at the high-mass end. These complicated functional forms contain eight HOD free parameters: F_c^A, F_c^B, M_c, $\sigma_{\log M}$, F_s, $M_{\rm min}$, $\delta_{\log M}$, and α. F_c^A and F_c^B are the normalization parameters that control the amplitudes of the Gaussian-like distribution and the step-like function of the central galaxy, respectively. M_c is a cutoff halo mass of the central galaxy and $\sigma_{\log M}$ determine the width of the Gaussian and the smoothed cutoff. F_s tunes the amplitude of the satellite galaxy occupation, $M_{\rm min}$ is a typical dark halo mass to possess one satellite galaxy, $\delta_{\log M}$ is a width of the smoothed cutoff of the satellite galaxy occupation at the low-mass end, and α is a satellite formation efficiency parameter. This occupation model is drawn in Figure 16.

In addition, Harikane et al. (2016) introduced the duty cycle, which normalizes the amplitude of the total galaxy occupation, based upon the formalism of Zheng et al. (2005) to analyze the LBGs at $z = 4-7$ as

$$N_{\rm tot}^{\rm Harikane}(M_{\rm h}) = DC \times \left\{N_{\rm cen}^{\rm Zheng}(M_{\rm h})\left[1+N_{\rm sat}^{\rm Zheng}(M_{\rm h})\right]\right\}, \quad (66)$$

where DC is a duty-cycle parameter with $DC \in [0,1]$.

These above occupation models have been mainly developed to represent general galaxy distributions within dark haloes, and the galaxy occupation models for peculiar populations such as the submillimeter galaxies (SMGs) and AGNs are also differently constructed (e.g., Richardson et al. 2013; Skibba et al. 2015; Magliocchetti et al. 2016); thus, one should select an adequate occupation model for their analysis.

All of the HOD analyses implemented in this thesis employ the halo occupation function proposed by Zheng et al. (2005) because this occupation function is regarded as a "standard" halo occupation model and our galaxy samples are not particular galaxy populations. Hereafter, I describe the $N_{\rm cen}^{\rm Zheng}$,

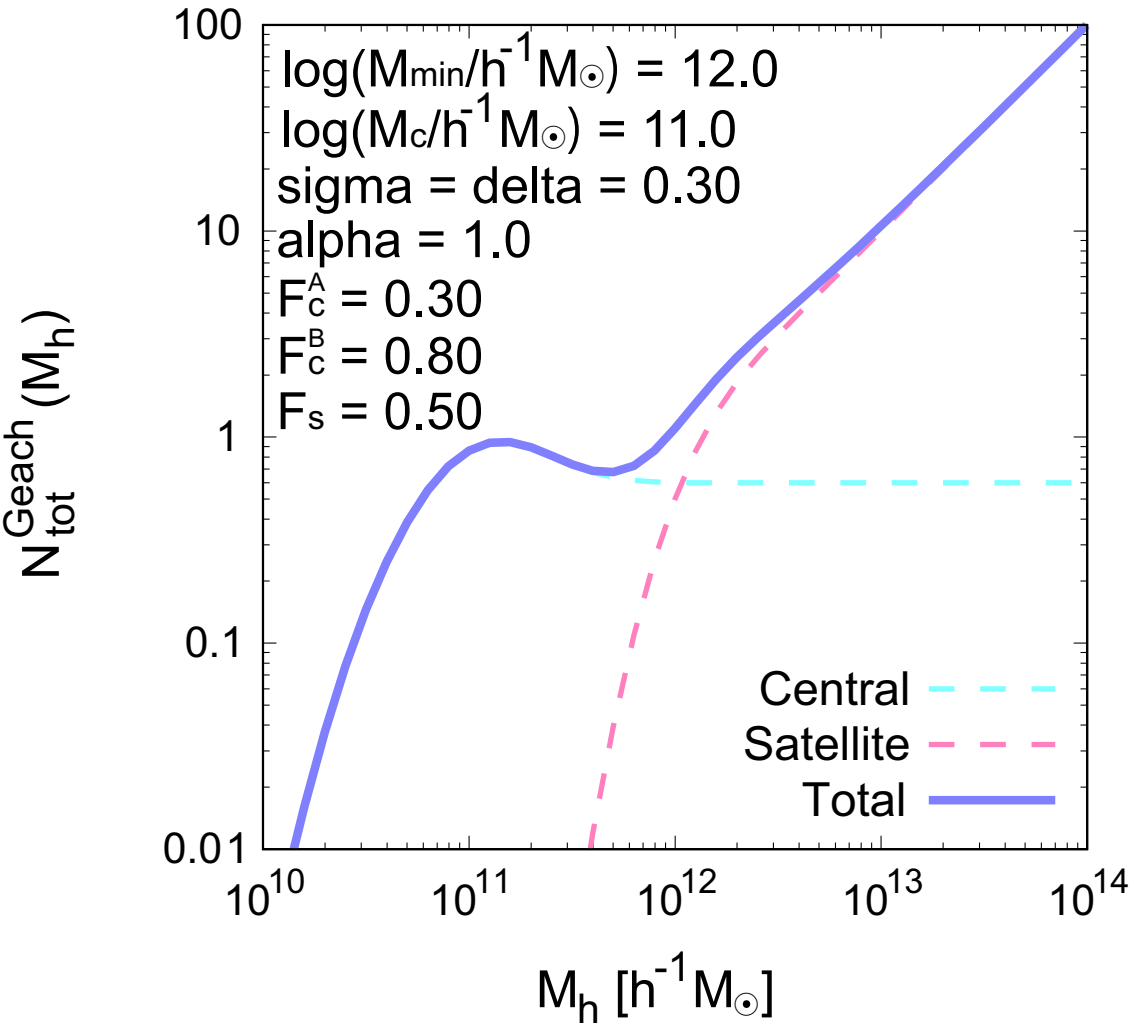

Figure 16.— The relationship between the number of occupied galaxies within the dark halo and the mass of the dark halo. The dashed cyan line is the occupation function of the central galaxy, $N_{\rm cen}^{\rm Geach}(M_{\rm h})$, the dashed red line is the occupation function of the satellite galaxy, $N_{\rm sat}^{\rm Geach}(M_{\rm h})$, and the solid blue line represents the total occupied galaxies, $N_{\rm tot}^{\rm Geach}(M_{\rm h})$. This relation is calculated by the halo occupation function of the equation (65) that is proposed by Geach et al. (2012). The HOD parameters are assumed as $\log(M_{\rm min}/h^{-1}M_{\odot}) = 12.0$, $\log(M_c/h^{-1}M_{\odot}) = 11.0$, $\sigma_{\log M} = 0.30$, $\delta_{\log M} = 0.30$, $\alpha = 1.0$, $F_c^A = 0.30$, $F_c^B = 0.80$, and $F_s = 0.50$.

$N_{\rm sat}^{\rm Zheng}$, and $N_{\rm tot}^{\rm Zheng}$ as $N_{\rm cen}$, $N_{\rm sat}$, and $N_{\rm tot}$, respectively, for the sake of simplicity. However, we should keep in mind that the halo occupation function of Zheng et al. (2005) is only confirmed its validity for the galaxies at $z \leq 1$. There is no creditable support for justification of this model at high-z Universe.

2.3.3. Galaxy power spectrum and correlation function

The basic idea of the HOD model is to predict the correlation function of galaxies based upon the halo bias model. The galaxy correlation function is, however, a formalism in real space, and it is difficult to compute the correlation function directly because the density fluctuation is not independent on the scales. The simplest strategy to compute galaxy correlation function is to convert the galaxy power-spectrum in Fourier space into galaxy correlation functions using the Wiener-Khintchine relation (equation 20) because galaxy power-spectrum is independent from the scale in the linear regime.

The galaxy correlation functions are composed of two distinct terms: the "1-halo term" and the "2-halo term" (cf., Ma & Fry 2000; Berlind & Weinberg 2002). The 1-halo term is dominated by the non-linear, small scale galaxy clustering, i.e., galaxy pairs within the same dark halo, whereas the 2-halo term represent the linear, large scale galaxy clustering that originates from the central galaxy pairs within different dark haloes. The transition scale between the 1-halo term and the 2-halo term typically corresponds to the virial radius of dark haloes. In the HOD model, one computes the galaxy power spectra that correspond to the 1-halo term, $P_g^{\rm 1h}(k)$, and the 2-halo term, $P_g^{\rm 2h}(k)$, and, then summing them to evaluate the galaxy power spectrum of the entire scale as

$$P_g(k) = P_g^{\rm 1h}(k) + P_g^{\rm 2h}(k). \tag{67}$$

The 1-halo term, which is the correlation between galaxy pairs within the same dark halo, furthermore can be divided into two distinct terms: the correlation 1) between the central galaxy and the satellite galaxies and 2) between the satellite galaxies within the same dark haloes. The central–satellite correlation term, $P_g^{\rm cs}(k)$, and the satellite–satellite correlation term, $P_g^{\rm ss}(k)$, can be calculated as

$$P_k^{\rm cs}(k,z) = \frac{2}{n_g^2}\int_{M_{\rm low}}^{M_{\rm high}} dM_{\rm h}\langle N_{\rm cen}N_{\rm sat}\rangle(M_{\rm h})\frac{dn(z)}{dM_{\rm h}}\left|u(k,M_{\rm h},z)\right|, \tag{68}$$

and

$$P_k^{\rm ss}(k,z) = \frac{1}{n_g^2}\int_{M_{\rm low}}^{M_{\rm high}} dM_{\rm h}\langle N_{\rm sat}(N_{\rm sat}-1)\rangle(M_{\rm h})\frac{dn(z)}{dM_{\rm h}}\left|u(k,M_{\rm h},z)\right|^2, \tag{69}$$

where $\frac{dn(z)}{dM_{\rm h}}$ is the halo mass function of Sheth & Tormen (1999) at redshift z, $u(k,M_{\rm h},z)$ is the NFW profile in Fourier space, and $\langle N_{\rm cen}N_{\rm sat}\rangle(M_{\rm h})$ and $\langle N_{\rm sat}(N_{\rm sat}-1)\rangle(M_{\rm h})$ are the expectation numbers of central–satellite and satellite–satellite galaxy pairs, respectively. n_g is the number density of galaxies calculated by

$$n_g = \int_{M_{\rm low}}^{M_{\rm high}} dM_{\rm h}\frac{dn(z)}{dM_{\rm h}}N_{\rm tot}(M_{\rm h}). \tag{70}$$

Using the results of Bullock et al. (2002) and Zheng et al. (2005), one can approximate the expectation numbers of galaxy paris as:

$$\langle N_{\rm cen}N_{\rm sat}\rangle(M_{\rm h}) = N_{\rm cen}(M_{\rm h})N_{\rm sat}(M_{\rm h}), \tag{71}$$

and

$$\langle N_{\rm sat}(N_{\rm sat}-1)\rangle(M_{\rm h}) = N_{\rm sat}(M_{\rm h})^2, \tag{72}$$

respectively. Throughout this thesis, the upper and lower limit of the integration with respect to the halo mass are set $M_{\rm high} = 10^{16}h^{-1}M_\odot$ and $M_{\rm low} = 10^3 h^{-1}M_\odot$, respectively. The total power spectrum of the 1-halo term is given by

$$P_g^{\rm 1h}(k,z) = P_k^{\rm cs}(k,z) + P_k^{\rm ss}(k,z). \tag{73}$$

The power spectrum of the 2-halo term can be calculated as

$$P_g^{\rm 2h}(k,z) = P_m(k,z)\left[\frac{1}{n_g}\int_{M_{\rm low}}^{M_{\rm high}} dM_{\rm h}\frac{dn(z)}{dM_{\rm h}}N_{\rm tot}(M_{\rm h})b_{\rm h}(M_{\rm h})\left|u(k,M_{\rm h},z)\right|^2\right]^2, \tag{74}$$

where $P_m(k,z)$ is the matter power spectrum and $b_{\rm h}(M_{\rm h})$ is the large-scale halo bias proposed by Tinker et al. (2010).

The real-space galaxy correlation function, $\xi(r)$, can be computed by the Fourier transformation:

$$\xi(r) = \frac{1}{2\pi}\int_0^\infty dk k^2 P_g(k)\frac{\sin kr}{kr}. \tag{75}$$

The projected angular auto correlation function, $\omega(\theta)$, can be calculated using the derivative limber equation (Limber 1953; Phillipps et al. 1978) as

$$\omega(\theta) = \frac{\displaystyle\int_{z_{\rm min}}^{z_{\rm max}}\left(\frac{dN}{dz}\right)^2\int_{k_{\rm min}}^{k_{\rm max}} dk\,dz\frac{k}{2\pi}P_g\left(k,z\right)J_0\left(kD_c(z)\theta\right)}{\displaystyle\left(\int_{z_{\rm min}}^{z_{\rm max}} dz\frac{dN}{dz}\right)^2}, \tag{76}$$

where $\frac{dN}{dz}$ is the normalized selection function of the galaxy sample and $J_0(x)$ is the zero-th order Bessel function. $D_c(z)$ is the radial comoving distance defines as

$$D_c(z) = \frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3+\Omega_\Lambda}}, \tag{77}$$

where c is a speed of light.

2.4. HOD Analysis

In this section, I present the procedure of the HOD analysis. The HOD analysis is, in short, to compare the observed ACF with the predicted ACF, which is computed by the HOD model to constrain the HOD parameters. Besides the shape of the ACF, the constraint of the number density of galaxy is also imposed to represent the observed galaxy distribution. As mentioned in Section 2.1.2, one should account for the covariance in the clustering measurements. In general, the correlations between the different scales are dealt with considering the covariance matrix.

The constraint of the HOD parameters is implemented through the χ^2 fitting method:

$$\chi^2 = \sum_{i,j}\left[\omega_{\rm obs}\left(\theta_i\right)-\omega_{\rm HOD}\left(\theta_i\right)\right]\left(C^{-1}\right)_{ij}\left[\omega_{\rm obs}\left(\theta_j\right)-\omega_{\rm HOD}\left(\theta_j\right)\right] + \frac{[n_g^{\rm obs}-n_g^{\rm HOD}]^2}{\sigma_{n_g}^2}, \tag{78}$$

where $(C^{-1})_{ij}$ is an (i,j) element of the inverse covariance matrix, $n_g^{\rm obs}$ and $n_g^{\rm HOD}$ are the number densities of galaxies from observation and the HOD model, and $\sigma^2_{n_g}$ is the statistical 1σ error of $n_g^{\rm obs}$. We note that both Poisson error and cosmic variance are considered in assessing $\sigma^2_{n_g}$ (Trenti & Stiavelli 2008). $n_g^{\rm obs}$ is calculated by

$$n_g^{\rm obs} = \frac{N_{gals}}{\Omega \int_z dz \frac{dV}{dz}}, \tag{79}$$

where N_{gals} is the total number of galaxies, Ω is the solid angle of the observation field, and $\frac{dV}{dz}$ is a comoving volume element. $n_g^{\rm HOD}$ can be calculated using the best-fit halo occupation function as

$$n_g^{\rm HOD} = \int dM_{\rm h} M_{\rm h} n(M_{\rm h}) N_{tot}(M_{\rm h}). \tag{80}$$

The mean halo mass, $\langle M_{\rm h} \rangle$, and the satellite fraction, $f_{\rm sat}$, can be estimated from the best-fit HOD parameters, which are derived by above procedure. The mean halo mass and the satellite fraction are defined as

$$\langle M_{\rm h} \rangle = \frac{1}{n_g} \int dM_{\rm h} M_{\rm h} n(M_{\rm h}) N_{\rm tot}(M_{\rm h}) \tag{81}$$

and

$$f_{\rm sat} = 1 - \frac{1}{n_g} \int dM_{\rm h}(M_{\rm h}) N_{\rm cen}(M_{\rm h}). \tag{82}$$

It is noted that n_g is calculated by equation (80).

3. GALAXY–HALO CONNECTION IN LOW-REDSHIFT UNIVERSE

3.1. Overview

3.1.1. Clustering analyses of spectroscopically observed galaxies

The relationship between galaxies at $z < 1.4$ and their host haloes have been well investigated. The biggest work of the low-z Universe is implemented by Zehavi et al. (2011), who carried out accurate clustering analyses using a large number of $z < 0.25$ galaxy sample obtained by Sloan Digital Sky Survey (SDSS; York et al. 2000) Seventh Data Release ($\sim 700,000$ galaxies over $8,000$ deg^2 total survey field). Radial distances as well as angular positions of each galaxy by the spectroscopic observation of SDSS enabled them to carry out real-space galaxy clustering analyses to derive the high-accuracy clustering information, and then real-space correlation functions were integrated along the line-of-sight to exclude the effect of the redshift-space distortions.

Zehavi et al. (2011) divided the numerous local SDSS galaxies into volume-limited subsamples by their k-corrected absolute r-band magnitudes and investigated the properties of host dark haloes by the clustering measurement and the HOD analysis. They found that the clustering strengths of galaxies and large-scale galaxy biases increase slowly at a luminosity of $L < L_\star$, where $L_\star$ corresponds to $M_r = -20.5$ magnitude for their analysis, and increase rapidly at a $L \gtrsim L_\star$, which is consistent with previous HOD studies using the data of large galaxy redshift surveys (e.g., Norberg et al. 2001; Zehavi et al. 2005), and the formation efficiency of satellite galaxies within massive dark haloes is nearly independent from the stellar masses of satellite galaxies, showing the same trend with HOD studies at $z < 1$ (e.g., Wake et al. 2011). Furthermore, SDSS galaxy samples were also divided into star-forming galaxies and passively evolving galaxies according to their k-corrected $(g - r)$ colors to investigate the dependence of clustering and host halo properties on galaxy populations as well as galaxy luminosities. The dependence of the clustering amplitude on galaxy luminosity was clearly different between galaxy populations; the clustering amplitude of the star-forming galaxies evolves weakly for faint galaxies ($L \lesssim 0.4L_\star$) but steadily for bright galaxies, whereas passively evolving galaxies show little evolution of the clustering amplitude, and Zehavi et al. (2011) interpreted this different characteristic as the difference of the fraction of the central and satellite galaxies.

3.1.2. SED fitting technique and photometric redshift

As seen in the previous section, Zehavi et al. (2011) succeeded in revealing the relationship between local galaxies with various baryonic characteristics (absolute magnitude, galaxy color, and galaxy population) and their host haloes; however, the redshift range of their clustering analysis was quite limited ($z < 0.25$) because they confined their galaxy samples to spectroscopically observed galaxies. It is essential to collect a large number of galaxies to achieve the clustering analysis with high S/N ratio, and one can obtain a lot of galaxies by photometric observations instead of spectroscopic observations.

A spectral energy distribution (SED) fitting technique, which is introduced firstly by Baum (1962), is a critical tool to estimate the redshift (photometric redshift) using photometric data. One of the

advantages of the SED fitting technique is that one can obtain the photometric redshift (as well as baryonic properties in some cases) of a large number of galaxies at once, which is difficult to be achieved by the spectroscopic observations. Moreover, the SED fitting technique can evaluate the photometric redshift for faint galaxies even for falling below the sensitivity limit of the spectroscopic observation. Therefore, we can derive a lot of galaxies at $z_{\rm phot} < 1.4$, which is a typical accuracy limit of photometric redshift in case of being limited by using optical images, by applying the SED fitting technique for the wide and deep photometric images obtained by extensive surveys such as Cosmic Evolution Survey (COSMOS; Scoville et al. 2007), Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Gwyn 2012), and Subaru Hyper Suprime-Cam Survey (Miyazaki et al. 2013).

The wide variety of models and codes of the SED fitting method have been proposed such as a machine-learning method, a template fitting method, and assuming Bayesian physical priors. The machine-learning fitting method, for example, ANNz (Collister & Lahav 2004), PhotoRApToR (Brescia & Cavuoti 2014), employs an algorithm to sophisticate their fitting procedures using neural network or random tree. This method is based upon the empirical fitting technique, which estimates the photometric redshift by constraining the color/magnitude–redshift relation through linear or polynomial fitting functions (cf., Connolly et al. 1995; Brunner et al. 1997). The advantage of the machine-learning method is to be able to treat large-sized databases that approach or exceed terabyte scale (e.g., Ball et al. 2008), whereas the disadvantage is that one has to prepare the training set to learn and sophisticate its algorithm. In general, luminosities of the training samples do not fall below the limitation of the spectroscopic observation; thus, the adoptable objects of this fitting method are limited for relatively bright samples.

In contrast, the template fitting method, LePhare (Arnouts et al. 1999; Ilbert et al. 2006), HyperZ (Bolzonella et al. 2000), Z-PEG (Le Borgne & Rocca-Volmerange 2002; Le Borgne et al. 2004), and EAZY (Brammer et al. 2008), simply compares the observed galaxy fluxes with SED templates that are generated by the spectral synthesis models (e.g., Bruzual & Charlot 2003) or observed galaxy SEDs (e.g., Coleman et al. 1980). The best-fit SED model and the photometric redshift are evaluated through the χ^2 fitting method. The biggest benefit to use this technique is that one can obtain physical properties of galaxies such as the stellar mass, the age, the star-formation rate, and the total amount of dust extinction, as well as their photometric redshifts because the shape of SED templates of synthesis models are determined and generated based upon the galaxy physical quantities. This fitting method is, however, largely affected by prepared SED models (i.e., emission lines, initial mass functions, dust extinction models, absorption models of the intergalactic medium, and star-formation histories) and physical parameter ranges. To assess the plausible physical parameters that follow the results of observations as well as photometric redshift, it has been developed the template fitting technique with Bayesian physical priors, which is firstly proposed and demonstrated by Benítez (2000), and examples of the SED fitting codes with Bayesian physical a priori probability distributions are GalMC (Acquaviva et al. 2011), iSEDfit (Moustakas et al. 2013), SEABASs (Rovilos et al. 2014), and Mizuki (Tanaka 2015).

As seen above, each SED fitting method has characteristics and suitable/unsuitable scientific usages. The clustering analysis, which is a subject of this thesis, is suitable for the template fitting method because it is in our interest for the relation between faint galaxies and their host haloes as well as those of bright galaxies, and for the dependence of clustering properties on the baryonic properties.

3.1.3. Clustering analyses using photometric redshifts

In this section, I summarize the results of some important clustering analyses using galaxy samples collected by their photometric redshifts.

Coupon et al. (2012) carried out clustering analyses using data of CFHTLS Wide field, which covers 133 deg^2. Depth of the photometric data of the CFHTLS Wide are not particularly deep (mean limiting magnitudes that correspond to the 50% completeness for point sources are $u^* \sim 25.3$, $g' \sim 25.5$, $r' \sim 24.8$, $i' \sim 24.5$, and $z' \sim 23.6$ magnitudes, respectively); however, wide survey field enabled them to collect relatively bright galaxies ($M_g - 5\log h < -21.8$) up to $z \sim 1.2$. They performed the SED fitting for their galaxy samples down to $i' < 22.5$ using LePhare code (Arnouts et al. 1999; Ilbert et al. 2006) and achieved the high-accuracy photometric redshift estimation ($\sigma_{|z_{\rm phot}-z_{\rm spec}|}/(1+z_{\rm spec})) = 0.037$ for galaxies at $0.2 < z_{\rm phot} < 1.2$, compared to spectroscopic redshifts), though the stellar mass estimation suffered from large uncertainties due to the shallowness of the optical images or the lack of the NIR images.

They divided their galaxy samples into "blue cloud" galaxies and "red sequence" galaxies according to the galaxy type of the best-fit SED model ("blue" galaxies correspond to the late-type galaxies, whereas "red" galaxies are the early-type galaxies), and those galaxies are furthermore divided into redshift bins by information of photometric redshift. The ACFs are computed for each population/redshift bin with high S/N ratio up to large-separation angular scale due to the large number of galaxy samples, and they found that bright, redder galaxy samples show more strongly clustering compared to faint, bluer galaxy samples. HOD analyses were applied for their ACFs and revealed that, at a fixed luminosity threshold, red galaxies reside in more massive haloes than blue galaxies, and galaxies are tend to be hosted more massive haloes at low redshift ($z \sim 0$) compared to higher-z Universe ($z > 0.6$) by tracing the redshift evolution of the number density of galaxy and the fraction of HOD mass parameters, $M_1/M_{\rm min}$. Galaxy stellar masses were estimated by the correlations between the stellar mass and the B-band luminosity calibrated by the results of the COSMOS survey (Ilbert et al. 2009), but blue galaxies were not able to be evaluated their stellar masses due to the large scatter in the stellar mass-to-luminosity relation. Using the stellar masses from the above relation and halo masses from HOD analyses, stellar-to-halo mass ratios (SHMRs) were calculated and the evolution of the pivot halo mass, $M_{\rm h}^{\rm pivot}$, was investigated. They showed that $M_{\rm h}^{\rm pivot}$ of the total sample increases with increasing redshift, whereas $M_{\rm h}^{\rm pivot}$ of the red galaxy sample show little redshift evolution, and interpreted this trend as the difference of the star-forming activities; the lack of the star-formation of the red galaxies makes $M_{\rm h}^{\rm pivot}$ constant.

Leauthaud et al. (2012) investigated the relationship between galaxies at $0.2 < z_{\rm phot} < 1.0$ and their dark haloes in the COSMOS field by the jointly analysis of galaxy-galaxy weak lensing, galaxy abundance, and galaxy spatial clustering using a theoretical framework developed by Leauthaud et al. (2011). Using deep and wide wavelength range of photometric data of the COSMOS survey, Leauthaud et al. (2012) succeeded in collecting fainter galaxy samples compared to Coupon et al. (2012) with high accurate photometric redshift. Galaxy stellar masses were also estimated by the SED fitting technique using the Bayesian code developed by Bundy et al. (2006) with uncertainty of $0.1 \sim 0.2$ dex. They revealed that $M_{\rm h}^{\rm pivot}$ increase with increasing redshift from $z_{\rm phot} \sim 0.2$ to $z_{\rm phot} \sim 1.0$, showing the trend of the galaxy downsizing (e.g., Cowie et al. 1996; Juneau et al. 2005). However, the values of the SHMRs at s$M_{\rm h}^{\rm pivot}$ are not significantly changed as $(M_{\rm h}/M_\star)^{\rm pivot} \sim 27$, indicating that the quenching of star-formation may depend on the value of the SHMR, not simply on the halo mass.

McCracken et al. (2015) presented the results of the clustering analysis using the percent-level precision photometric redshifts derived by combining the data of the COSMOS survey with the UltraVISTA survey (McCracken et al. 2012). The UltraVISTA survey is a very deep NIR observation, which has been carrying out the wide-field NIR camera, VIRCAM, mounted on the VISTA telescope (Emerson & Sutherland 2010). The limiting magnitudes of J-, H-, and $K_{\rm s}$-band reach ~ 24 magnitudes (5σ, $2''$ diameter aperture in the AB magnitude) and the survey area is ~ 1.5 deg^2, where is a subset of the COSMOS field. McCracken et al. (2015) obtained galaxy samples down to $K_{\rm s} < 24.0$ up to $z_{\rm phot} < 2.5$ derived by the SED fitting procedure using the LePhare code, and the accuracy of the photometric redshift is achieved less than 1% contamination of catastrophic redshift error for $z_{\rm phot} < 1.5$ samples. They divided their galaxy samples into subsamples by the photometric redshift and stellar mass, and carried out HOD analysis to investigate the dependence of clustering properties $M_{\rm h}^{\rm pivot}$ are evaluated up to $z \sim 2$ via the abundance-matching technique in addition to HOD analysis, and compared their results with those in literature (Coupon et al. 2012; Leauthaud et al. 2012; Hudson et al. 2015; Martinez-Manso et al. 2015). They concluded that $M_{\rm h}^{\rm pivot}$ does not significantly evolve from $z = 0$ to $z \sim 2$ as a consequence of the little evolution of halo mass functions and stellar mass functions up to $z \sim 2$.

3.1.4. Motivation of this study

Recent wide-field photo-z galaxy surveys (e.g., Coupon et al. 2012, 2015; van Uitert et al. 2016) collected total, unbiased galaxy samples up to $z \sim 1$ and discussed the properties of dark haloes that harbour the low-z galaxy samples using the HOD formalism or the galaxy–galaxy lensing technique (refer to Section 3.1.3). Those studies precisely revealed the SHMRs even for satellite galaxies and also gave the inference of the dependence on their environments; however, their analyses were confined to low-z massive galaxies, i.e., $\log(M_\star/M_\odot) \gtrsim 10.0$ at $z \lesssim 1$, due to the shallowness of photometric images. On the other hand, deep-field photo-z surveys (e.g., McCracken et al. 2015) have revealed the relationship between galaxies and their host dark haloes of less massive galaxies ($M_\star \sim 10^9 M_\odot$) as well as massive galaxies up to $z \sim 3$. Nevertheless, the accuracy of the clustering analysis, especially of galaxies with $M_\odot \gtrsim 10^{10} M_\odot$, was not enough because of the small survey area (1.5 deg^2 for McCracken et al. 2015).

To investigate the precise relationship between low-z galaxies and their host dark haloes, I carry out precision clustering and HOD analyses using the extensive dataset obtained by the Hyper Suprime-Cam Subaru Strategic Project (HSC SSP) survey (Miyazaki et al. 2013). The HSC SSP survey is a large-scale optical-image survey using the new imaging instrument mounted on the prime focus of the Subaru Telescope, Hyper Suprime-Cam (Miyazaki et al. 2012, details of the HSC SSP survey is presented in Section 3.2). Using the excellent dataset of the HSC SSP survey, I progress the study to fill a gap between previous clustering studies in low-z with extremely high S/N ratios. By taking advantages of the wide-field imaging capability and the moderately deep imaging of the HSC SSP survey, I can collect a lot of less massive galaxies near the mass limit ($M_\star \sim 10^{9.6} M_\odot$) as well as massive galaxies ($M_\star \sim 10^{11} M_\odot$).

In addition, the HSC SSP survey provides the photometric-redshift catalogue; the photometric-redshift catalogue contains the objects with physical quantities such as the photometric redshift, the star-formation rate, the galaxy stellar mass, the age, and the amount of dust extinction. Using the

HSC photo-z catalogue, I can compute precise ACFs with dividing the galaxy samples into subsamples according to their baryonic natures and derive the properties of host dark haloes via the HOD analysis.

Throughout this section, I employ the flat lambda cosmology ($\Omega_{\rm m} = 0.27, \Omega_\Lambda = 0.73$, and $\Omega_{\rm b} = 0.045$), the Hubble constant as $h = {\rm H}_0/100{\rm kms}^{-1}{\rm Mpc}^{-1} = 0.7$, the matter fluctuation amplitude at $8h^{-1}$Mpc as $\sigma_8 = 0.8$, and spectral index of the primordial power spectrum as $n_{\rm s} = 1$, which are based upon the results of the *WMAP* seven-year data (Komatsu et al. 2011) to compare with the results of previous works. With these assumed cosmological parameters, the ages of the Universe at $z = 0.5$ and $z = 1.0$ are ~ 6.14 and $4.22\ h^{-1}$Gyr, and physical angular scales of 1 arcsec correspond to 4.32 and $5.71\ h^{-1}$kpc, respectively.

3.2. Details of Data and Samples

3.2.1. Data description

To construct galaxy samples, I used the data of the HSC SSP survey (Miyazaki et al. 2013), which is the extensive optical imaging survey operated by Subaru Telescope. The HSC (Miyazaki et al. 2012) is the latest imaging instrument mounted on the prime focus of Subaru Telescope, which has 104 science CCDs and covers 90 arcmin diameter (~ 1.5 deg^2) of the sky at a pixel scale of 0.168 arcsec pixel^{-1}. The HSC SSP survey started its program in March 2014 and is designed to implement observation ~ 300 nights over $5-6$ years. The quality of the photometric images is quite high: the median seeing in the i-band is ~ 0.6 arcsec.

The HSC SSP survey consists of three distinct surveys: the Wide-layer, the Deep-layer, and the Ultra Deep-layer. The Wide-layer covers extremely wide area ($\sim 1,400$ deg^2) by five optical bands (g-, r-, i-, z-, and y-band) with moderate depth ($r_{{\rm lim},5\sigma} \sim 26.1$ mag, where $r_{{\rm lim},5\sigma}$ is a 5σ limiting magnitude with $2''$ aperture size), whereas the Ultra Deep-layer is a highly deep-imaging survey ($r_{{\rm lim},5\sigma} \sim 27.7$ mag) by five optical broad bands and four narrow bands with a small survey area (~ 3.5 deg^2). Transmission functions of the broad-band filters of the HSC are presented in Figure 17. In this thesis, I used the data of the Wide-layer because the HSC SSP survey has not yet been completed and the Deep-layer and the Ultra Deep-layer do not reach their limiting magnitudes.

3.2.2. Sample selection

Galaxies are derived from the photometric-redshift catalogue using an SED fitting code with Baysian physical priors, Mizuki (Tanaka 2015), based on the HSC SSP data released in 2015. Details of the Mizuki can be found in Section 5.3.1. I set magnitude thresholds to obtain galaxies from entire survey area homogeneously as follows: $g < 26.0$, $r < 25.6$, $i < 25.4$, $z < 24.6$, and $y < 23.9$. Galaxies with large reduced χ^2 in the SED fitting, i.e., $\chi^2_{\rm SEDfit}/{\rm dof} \geq 5.0$, are excluded. The effective survey area of the data released in 2015 is ~ 162.8 deg^2 after masking the edges of the images and around the saturated objects.

I divided galaxy samples into subsamples according to their photometric-redshift. Thanks to the large number of galaxy samples, galaxies can be resampled into four redshift bins: $0.30 < z_{\rm phot} < 0.55$, $0.55 < z_{\rm phot} < 0.80$, $0.80 < z_{\rm phot} < 1.10$, and $1.10 < z_{\rm phot} < 1.40$. The lowest photometric redshift is

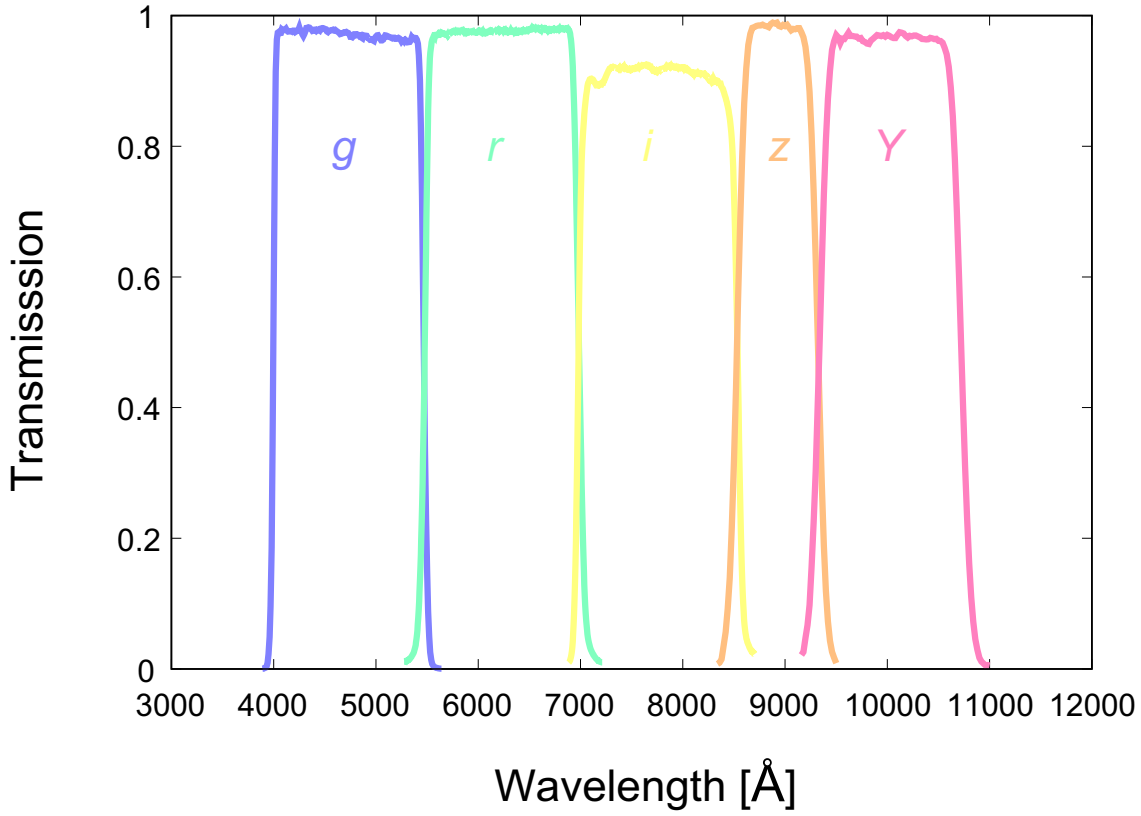

Figure 17.— Transmission functions of the HSC broad-band filters. The g- (blue), r- (green), i- (yellow), z- (orange), and y-band filter (red) are presented.

chosen as $z_{\rm phot} = 0.30$ due to the small sample size of the nearby galaxies and the accuracy of the SED fitting for those galaxies is not high, whereas the highest photometric redshift is set to $z_{\rm phot} = 1.40$ because NIR observations are essential for collecting the galaxies at $z > 1.4$. To avoid contaminations from other redshift bins, I only use galaxies satisfying the following condition: the upper (lower) limit of the 1σ confidence interval of the photometric redshift, $z_{\rm phot,upper}$ ($z_{\rm phot,lower}$), is smaller (larger) than the largest (smallest) redshift of the redshift bin, i.e.:

$$z_{\rm smallest} \leq z_{\rm phot,lower} \leq z_{\rm phot} \leq z_{\rm phot,upper} \leq z_{\rm largest}, \tag{83}$$

where $z_{\rm smallest}$ ($z_{\rm largest}$) represents the smallest (largest) redshift of that redshift bin. The redshift distributions and the stellar-mass distributions of the HSC galaxy samples are shown in Figures 18 and 19.

To determine stellar-mass limits of each redshift bin, I compare stellar-mass functions (SMFs) of the HSC galaxy samples with SMFs in the COSMOS field. I utilize the public photometric-redshift catalogue (Muzzin et al. 2013) using the data of the COSMOS survey (Scoville et al. 2007) and the UltraVISTA survey (McCracken et al. 2012) to compute the SMFs in the COSMOS field of each redshift bin. The SMF, $\Phi(M_\star)$, is simply calculated as:

$$\Phi(M_\star) = \frac{1}{\Delta M_\star} \sum_{i=1}^{N} \frac{1}{\int_{z_{\rm min}}^{z_{\rm max}} dz \frac{dV}{dz}}, \tag{84}$$

where $\Delta M_\star$ is the size of the stellar-mass bin, N is the number of galaxy samples in the stellar-mass bin, $z_{\rm min}$ ($z_{\rm max}$) is the lower (upper) limit of redshift, and $\frac{dV}{dz}$ is a comoving volume element. The stellar-mass limits, $M_{\star,\rm limit}$, of each redshift bin are determined as the 70% stellar-mass completeness of the COSMOS SMFs; $\log(M_{\star,\rm limit}/M_\star) = 9.4$, 9.6, 9.8, and 10.0 for the redshift bin of $0.30 < z_{\rm phot} < 0.55$, $0.55 < z_{\rm phot} < 0.80$, $0.80 < z_{\rm phot} < 1.10$, and $1.10 < z_{\rm phot} < 1.40$, respectively. The number of the galaxy samples of each redshift bin and stellar-mass bin are presented in Table $1 - 4$.

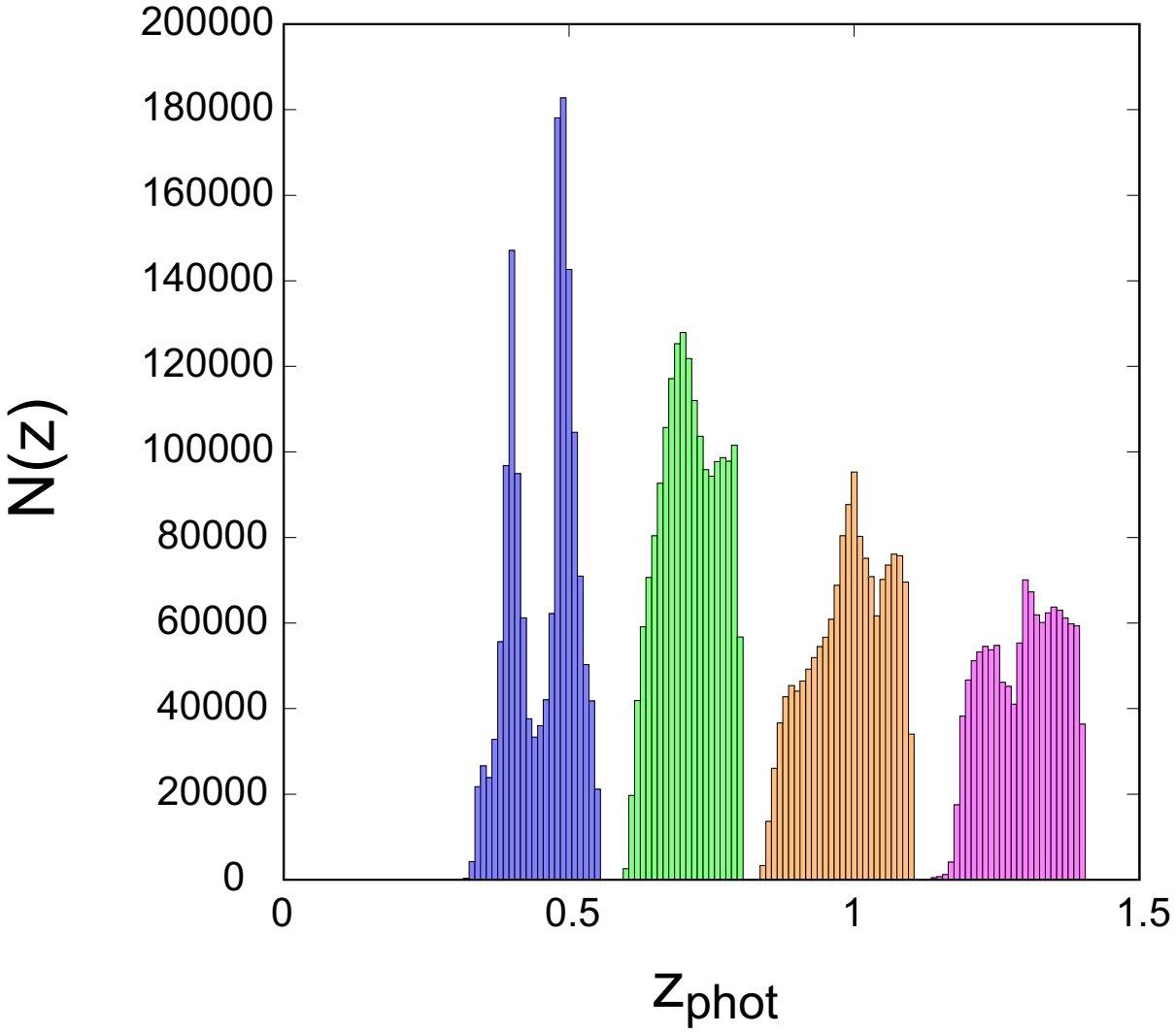

Figure 18.— The redshift distributions of galaxies obtained by the HSC SSP survey Wide-layer. Photometric redshifts are computed by an SED fitting code with Bayesian physical priors, MIZUKI. Galaxies are dividid into four redshift bins according to their photometric redshift: $0.30 < z_{\rm phot} < 0.55$ (blue), $0.55 < z_{\rm phot} < 0.80$ (green), $0.80 < z_{\rm phot} < 1.10$ (orange), and $1.10 < z_{\rm phot} < 1.40$ (red). Conditions of the galaxy selection are presented in Section 3.2.2.

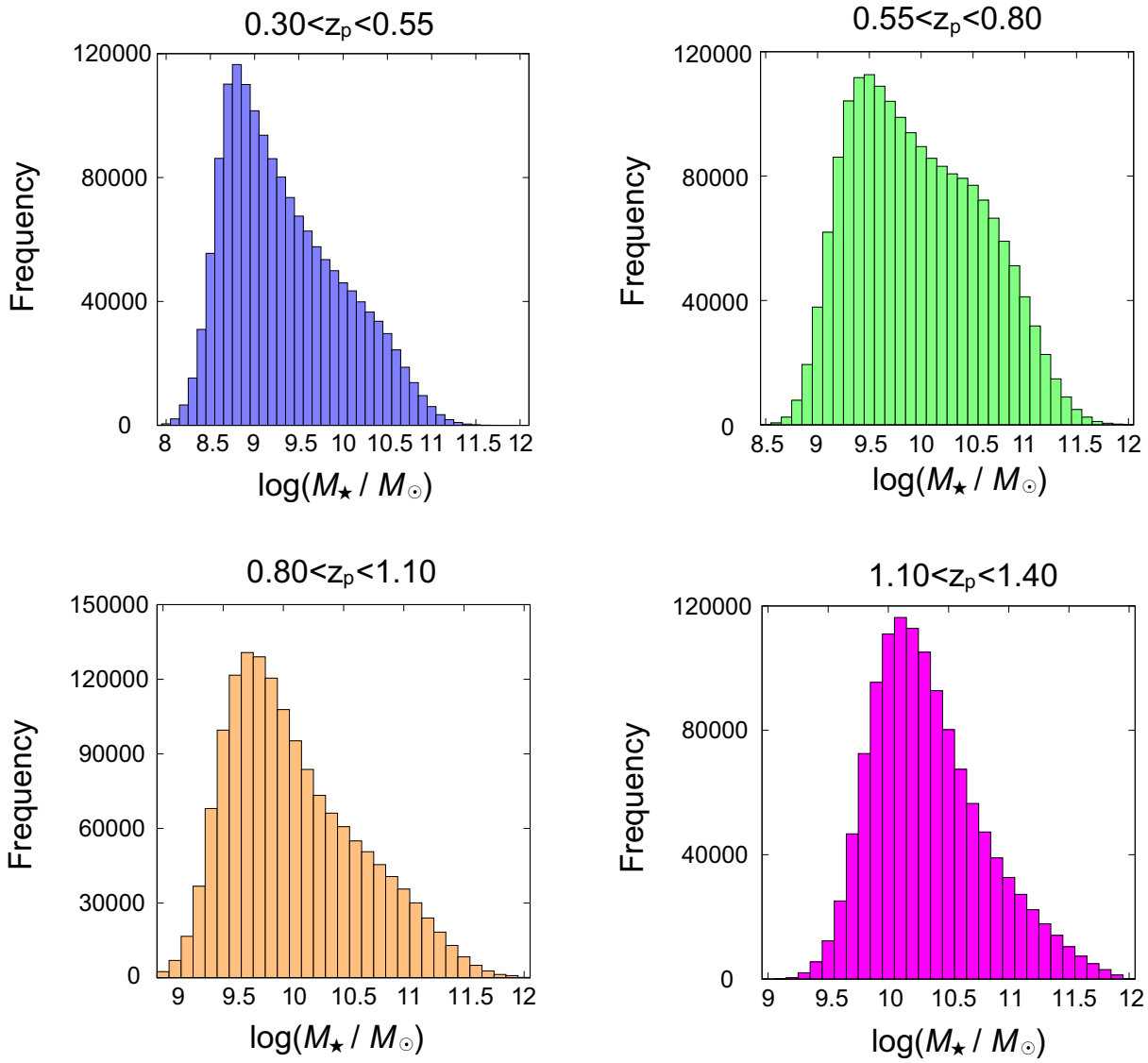

Figure 19.— Distributions of stellar masses of each redshift bin. Stellar masses of each galaxy samples are evaluated through the SED fitting technique.

Table 1: The number of the galaxy samples and the best-fit HOD parameters of each subsample limited by the stellar mass at $0.30 < z_{\rm phot} < 0.55$

$\log(M_{\star,\rm limit}/M_\odot)$	N	$\log(M_{\rm min}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\sigma_{\log M}$	α	$\log(M_{\rm h}/h^{-1}M_\odot)$	$f_{\rm sat}$	χ^2/dof
9.4	459,192	$11.24^{+0.02}_{-0.02}$	$12.74^{+0.06}_{-0.06}$	$0.07^{+0.04}_{-0.02}$	$1.12^{+0.06}_{-0.05}$	12.98 ± 0.04	0.19 ± 0.01	3.04
9.6	361,046	$11.35^{+0.02}_{-0.02}$	$12.78^{+0.07}_{-0.06}$	$0.20^{+0.05}_{-0.06}$	$1.12^{+0.07}_{-0.05}$	13.01 ± 0.04	0.20 ± 0.01	3.08
9.8	277,145	$11.46^{+0.03}_{-0.02}$	$12.80^{+0.07}_{-0.05}$	$0.29^{+0.05}_{-0.07}$	$1.14^{+0.06}_{-0.05}$	13.05 ± 0.04	0.23 ± 0.01	2.94
10.0	205,207	$11.58^{+0.03}_{-0.03}$	$13.00^{+0.08}_{-0.08}$	$0.28^{+0.05}_{-0.06}$	$1.22^{+0.05}_{-0.07}$	13.09 ± 0.05	0.21 ± 0.01	2.87
10.2	142,898	$11.72^{+0.03}_{-0.04}$	$13.02^{+0.10}_{-0.08}$	$0.27^{+0.08}_{-0.08}$	$1.15^{+0.07}_{-0.09}$	13.12 ± 0.05	0.19 ± 0.01	3.12
10.4	89,542	$11.91^{+0.04}_{-0.03}$	$13.17^{+0.09}_{-0.06}$	$0.42^{+0.10}_{-0.11}$	$1.28^{+0.08}_{-0.07}$	13.14 ± 0.05	0.18 ± 0.01	3.62
10.6	47,265	$12.17^{+0.07}_{-0.07}$	$13.37^{+0.10}_{-0.10}$	$0.52^{+0.12}_{-0.14}$	$1.14^{+0.10}_{-0.13}$	13.22 ± 0.06	0.17 ± 0.01	2.99
10.8	20,289	$12.67^{+0.08}_{-0.07}$	$13.97^{+0.14}_{-0.16}$	$0.45^{+0.15}_{-0.16}$	$1.05^{+0.13}_{-0.13}$	13.25 ± 0.05	0.11 ± 0.01	1.57
11.0	6,720	$13.10^{+0.11}_{-0.12}$	$14.21^{+0.14}_{-0.20}$	$0.75^{+0.18}_{-0.16}$	$1.48^{+0.06}_{-0.08}$	13.33 ± 0.10	0.05 ± 0.01	2.96
11.2	1,731	$13.67^{+0.09}_{-0.10}$	$14.89^{+0.22}_{-0.26}$	$0.56^{+0.23}_{-0.24}$	$1.07^{+0.20}_{-0.25}$	13.74 ± 0.11	0.04 ± 0.01	0.76

Table 2: The number of the galaxy samples and the best-fit HOD parameters of each subsample limited by the stellar mass at $0.55 < z_{\rm phot} < 0.80$

$\log(M_{\star,\rm limit}/M_\odot)$	N	$\log(M_{\rm min}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\sigma_{\log M}$	α	$\log(M_{\rm h}/h^{-1}M_\odot)$	$f_{\rm sat}$	χ^2/dof
9.6	889,036	$11.41^{+0.01}_{-0.01}$	$12.55^{+0.02}_{-0.01}$	$0.40^{+0.06}_{-0.05}$	$1.09^{+0.07}_{-0.08}$	12.72 ± 0.04	0.28 ± 0.01	1.37
9.8	735,865	$11.52^{+0.01}_{-0.02}$	$12.68^{+0.02}_{-0.02}$	$0.44^{+0.07}_{-0.07}$	$0.91^{+0.08}_{-0.06}$	12.78 ± 0.03	0.24 ± 0.01	1.42
10.0	599,589	$11.62^{+0.02}_{-0.02}$	$12.84^{+0.04}_{-0.05}$	$0.28^{+0.08}_{-0.09}$	$1.13^{+0.06}_{-0.08}$	12.86 ± 0.03	0.23 ± 0.01	1.82
10.2	474,774	$11.74^{+0.03}_{-0.02}$	$12.96^{+0.04}_{-0.06}$	$0.36^{+0.10}_{-0.11}$	$1.13^{+0.04}_{-0.06}$	12.92 ± 0.04	0.21 ± 0.01	2.28
10.4	357,678	$11.90^{+0.04}_{-0.05}$	$13.15^{+0.05}_{-0.07}$	$0.35^{+0.12}_{-0.14}$	$1.22^{+0.05}_{-0.08}$	12.98 ± 0.04	0.18 ± 0.02	2.17
10.6	246,762	$12.12^{+0.06}_{-0.05}$	$13.31^{+0.04}_{-0.04}$	$0.47^{+0.12}_{-0.15}$	$1.20^{+0.04}_{-0.06}$	13.02 ± 0.05	0.16 ± 0.01	2.91
10.8	150,766	$12.36^{+0.06}_{-0.06}$	$13.65^{+0.06}_{-0.07}$	$0.42^{+0.11}_{-0.14}$	$1.03^{+0.08}_{-0.09}$	13.07 ± 0.05	0.12 ± 0.02	3.35
11.0	77,342	$12.74^{+0.09}_{-0.08}$	$13.96^{+0.11}_{-0.11}$	$0.57^{+0.16}_{-0.19}$	$1.20^{+0.09}_{-0.11}$	13.14 ± 0.06	0.08 ± 0.02	3.89
11.2	31,556	$13.12^{+0.11}_{-0.10}$	$14.29^{+0.10}_{-0.11}$	$0.65^{+0.11}_{-0.15}$	$1.15^{+0.11}_{-0.12}$	13.28 ± 0.07	0.06 ± 0.02	3.77
11.4	9,523	$13.50^{+0.13}_{-0.11}$	$14.72^{+0.14}_{-0.15}$	$0.38^{+0.16}_{-0.18}$	$0.92^{+0.12}_{-0.14}$	13.70 ± 0.06	0.05 ± 0.03	3.95

Table 3: The number of the galaxy samples and the best-fit HOD parameters of each subsample limited by the stellar mass at $0.80 < z_{\rm phot} < 1.10$

$\log(M_{\star,\rm limit}/M_\odot)$	N	$\log(M_{\rm min}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\sigma_{\log M}$	α	$\log(M_{\rm h}/h^{-1}M_\odot)$	$f_{\rm sat}$	χ^2/dof
9.8	743,925	$11.48^{+0.01}_{-0.01}$	$12.86^{+0.01}_{-0.01}$	$0.11^{+0.01}_{-0.01}$	$1.19^{+0.01}_{-0.02}$	12.72 ± 0.02	0.17 ± 0.01	0.97
10.0	562,992	$11.56^{+0.02}_{-0.02}$	$12.90^{+0.06}_{-0.06}$	$0.07^{+0.04}_{-0.05}$	$1.24^{+0.05}_{-0.06}$	12.79 ± 0.04	0.18 ± 0.02	0.87
10.2	420,352	$11.69^{+0.01}_{-0.01}$	$12.93^{+0.04}_{-0.06}$	$0.16^{+0.05}_{-0.4}$	$1.18^{+0.07}_{-0.09}$	12.82 ± 0.03	0.19 ± 0.01	0.96
10.4	310,325	$11.83^{+0.03}_{-0.03}$	$12.98^{+0.04}_{-0.06}$	$0.33^{+0.09}_{-0.10}$	$1.08^{+0.09}_{-0.10}$	12.85 ± 0.03	0.20 ± 0.01	0.64
10.6	221,021	$12.01^{+0.05}_{-0.04}$	$13.16^{+0.04}_{-0.04}$	$0.27^{+0.11}_{-0.11}$	$1.09^{+0.10}_{-0.11}$	12.87 ± 0.04	0.19 ± 0.01	0.83
10.8	147,477	$12.31^{+0.07}_{-0.07}$	$13.57^{+0.08}_{-0.07}$	$0.36^{+0.18}_{-0.20}$	$1.12^{+0.11}_{-0.18}$	12.92 ± 0.05	0.12 ± 0.02	0.91
11.0	88,454	$12.59^{+0.07}_{-0.07}$	$13.86^{+0.12}_{-0.11}$	$0.30^{+0.20}_{-0.19}$	$1.12^{+0.15}_{-0.13}$	13.03 ± 0.05	0.10 ± 0.02	1.53
11.2	44,937	$12.91^{+0.07}_{-0.07}$	$14.07^{+0.20}_{-0.19}$	$0.72^{+0.18}_{-0.24}$	$1.16^{+0.20}_{-0.21}$	13.10 ± 0.08	0.05 ± 0.01	0.90
11.4	17,824	$13.20^{+0.09}_{-0.11}$	$14.52^{+0.24}_{-0.28}$	$0.61^{+0.23}_{-0.26}$	$1.14^{+0.25}_{-0.26}$	13.25 ± 0.06	0.04 ± 0.02	0.77

Table 4: The number of the galaxy samples and the best-fit HOD parameters of each subsample limited by the stellar mass at $1.10 < z_{\rm phot} < 1.40$

$\log(M_{\star,\rm limit}/M_\odot)$	N	$\log(M_{\rm min}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\sigma_{\log M}$	α	$\log(M_{\rm h}/h^{-1}M_\odot)$	$f_{\rm sat}$	χ^2/dof
10.0	686,638	$11.63^{+0.01}_{-0.01}$	$12.97^{+0.02}_{-0.03}$	$0.14^{+0.03}_{-0.04}$	$1.25^{+0.03}_{-0.05}$	12.51 ± 0.03	0.13 ± 0.01	2.87
10.2	508,118	$11.73^{+0.02}_{-0.03}$	$13.11^{+0.03}_{-0.03}$	$0.32^{+0.05}_{-0.06}$	$1.12^{+0.05}_{-0.06}$	12.59 ± 0.02	0.11 ± 0.01	2.21
10.4	350,273	$11.84^{+0.02}_{-0.02}$	$13.22^{+0.04}_{-0.05}$	$0.07^{+0.04}_{-0.04}$	$1.23^{+0.05}_{-0.08}$	12.63 ± 0.02	0.11 ± 0.04	2.28
10.6	230,681	$11.99^{+0.01}_{-0.01}$	$13.32^{+0.04}_{-0.05}$	$0.27^{+0.06}_{-0.06}$	$1.23^{+0.07}_{-0.08}$	12.67 ± 0.03	0.10 ± 0.01	2.04
10.8	147,436	$12.23^{+0.02}_{-0.03}$	$13.59^{+0.07}_{-0.08}$	$0.31^{+0.10}_{-0.14}$	$1.14^{+0.12}_{-0.13}$	12.76 ± 0.05	0.08 ± 0.01	2.95
11.0	90,174	$12.56^{+0.04}_{-0.04}$	$13.63^{+0.10}_{-0.11}$	$0.66^{+0.09}_{-0.12}$	$1.11^{+0.12}_{-0.14}$	12.79 ± 0.08	0.07 ± 0.02	3.51
11.2	50,743	$12.86^{+0.05}_{-0.05}$	$14.10^{+0.12}_{-0.14}$	$0.77^{+0.06}_{-0.10}$	$1.15^{+0.18}_{-0.16}$	12.88 ± 0.08	0.03 ± 0.01	3.65
11.4	24,859	$13.14^{+0.07}_{-0.07}$	$14.34^{+0.14}_{-0.16}$	$0.69^{+0.07}_{-0.11}$	$1.15^{+0.20}_{-0.22}$	13.04 ± 0.07	0.02 ± 0.01	3.97

3.3. Clustering Analysis

3.3.1. Angular correlation functions of HSC galaxy samples

The ACFs of HSC galaxy samples are computed using the estimator of Landy & Szalay (1993). In calculating the ACFs using the Landy & Szalay estimator, one should generate random samples over galaxy distributions. In oder to reduce the Poisson noise, the number of random sample should be 100 times larger than the number of galaxy sample; therefore, computational costs become bigger with increasing the number of galaxy samples. To reduce the computational cost of the ACF calculation, I employ the Burnes-Hut tree algorithm (Barnes & Hut 1986) and succeed in decreasing the cost from the order of $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$, where N represents the number of random samples.

The data of the HSC Wide-layer released in 2015 is composed of seven distinct regions. I compute ACFs at each region and then sum those ACFs with weights to derive the ACF of the entire survey field. i.e.,

$$\omega(\theta) = \sum_{i=1}^{7} \frac{\alpha_i \omega_i(\theta)}{\alpha_i(\theta)}, \tag{85}$$

where i represents the number of the region and $\alpha_i(\theta)$ is a weight at the angular scale of θ of the ACF of the i-th region. $\alpha_i(\theta)$ can be calculated using the Poisson error, $\sigma_i(\theta)$, as $\alpha_i(\theta) = 1/\sigma_i^2(\theta)$. Thanks to the extremely wide survey area, ACFs are derived with high quality and precision even at the large-angular scale ($\theta \sim 1$ deg), and the integral constraint is found to be a quite small, IC ~ 0.2, with assuming the power-law slope of $\gamma = 1.8$.

The error estimation is performed by the "delete-one" jackknife resampling method (e.g., Shao 1986; Norberg et al. 2009). I divide the HSC Wide-layer survey field into 80 sub-fields and compute ACFs by excluding one sub-field as is the same manner for deriving the "true" ACF. Each sub-field has ~ 2 deg^2 that is enough area to see the field-to-field variance as well as the variance of each angular bin. The error of ACFs are derived by the root mean squares of each angular bin through the jackknife resampling technique and the covariance matrices are also evaluated using the equation (12).

3.3.2. The HOD analysis

High-quality ACFs of the HSC galaxy samples enable to carry out the HOD analysis to investigate precise properties of dark haloes. I use the halo occupation model proposed by Zheng et al. (2005) (equation 60 − 62) with the halo exclusion model, the halo mass function is adopted of Sheth & Tormen (1999), the radial density profile of dark haloes is used the NFW profile (Navarro et al. 1997), and the large-scale halo bias model is the analytical formulate of Tinker et al. (2010). The redshift distributions are used the photometric redshift of each stellar-mass bin.

The results of the ACFs and the HOD fittings are presented in Figure 20 − Figure 23 and values of the best-fit HOD parameters are given in Table 1 − Table 4. All of the stellar-mass limited observed ACFs are well represented by the HOD model at all of the angular scales.

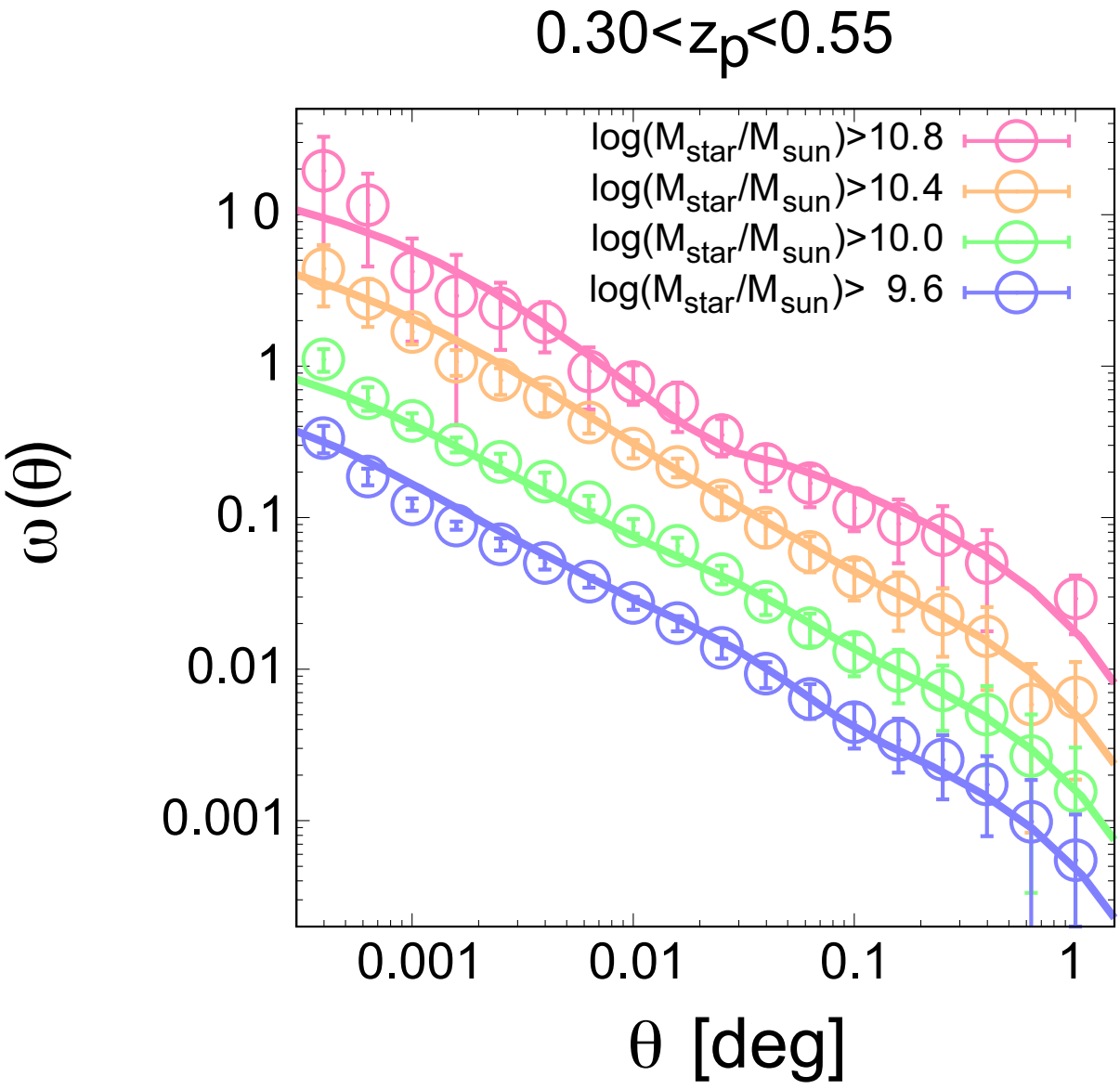

Figure 20.— ACFs of the galaxy samples at $0.30 < z_{\rm phot} < 0.55$ obtained from the HSC SSP photometric-redshift catalogue (circles with error bars) and the best-fit HOD fittings of each stellar-mass bin (solid lines). To clarify the results, amplitudes of the observed ACFs and the HOD fittings are scaled arbitrarily.

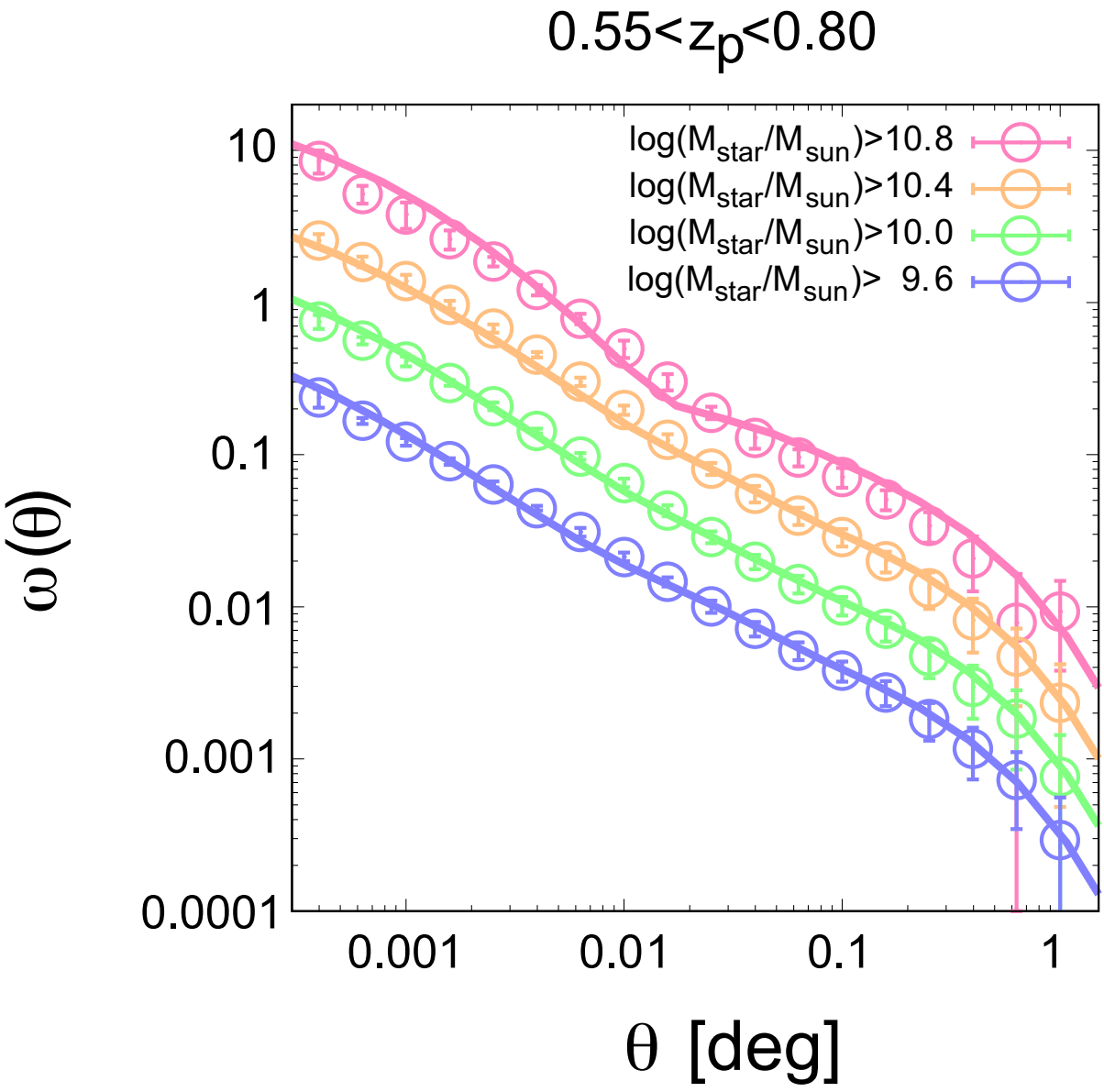

Figure 21.— Same as Figure 20, but for galaxies at $0.55 < z_{\rm phot} < 0.80$.

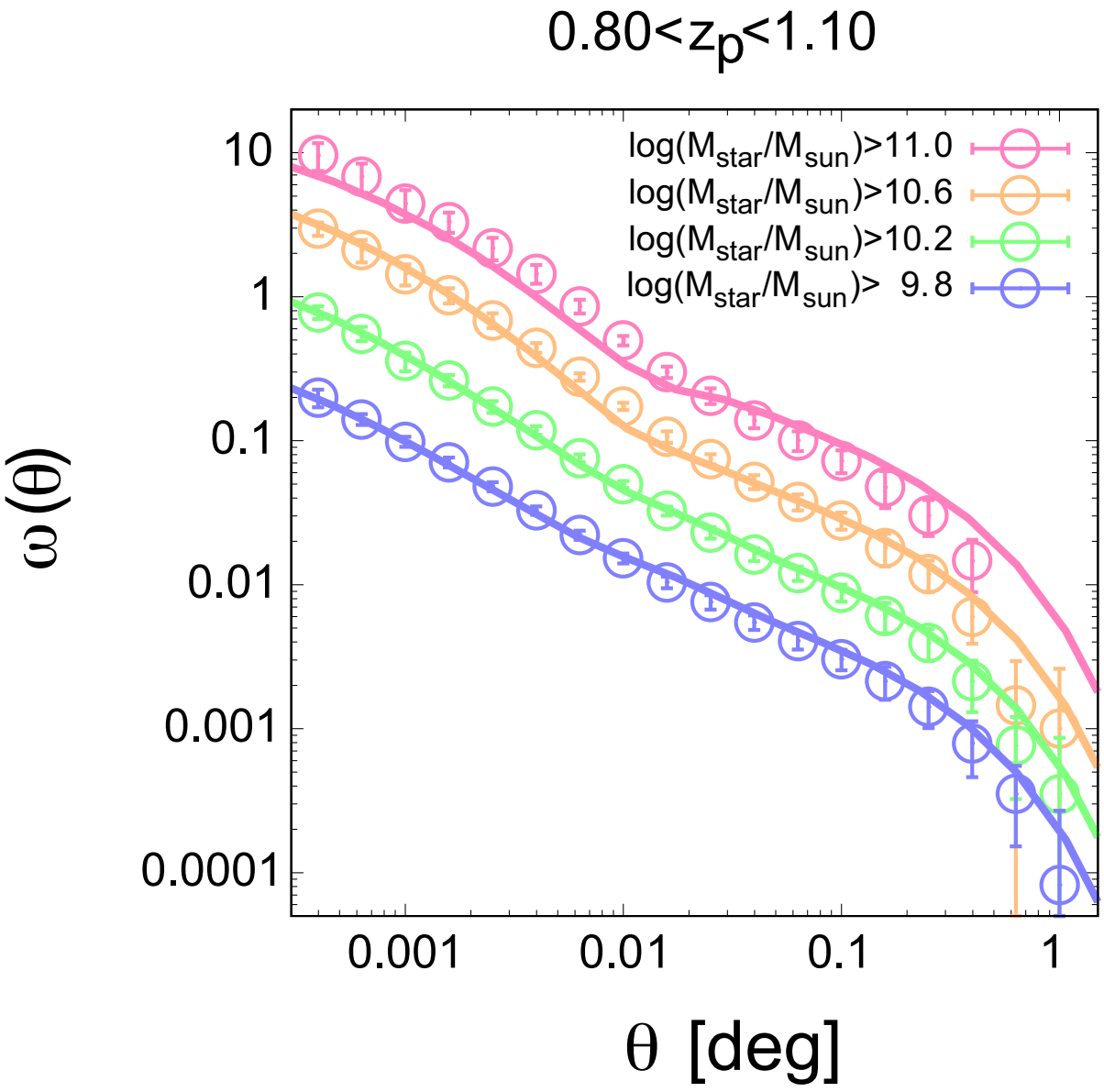

Figure 22.— Same as Figure 20, but for galaxies at $0.80 < z_{\rm phot} < 1.10$.

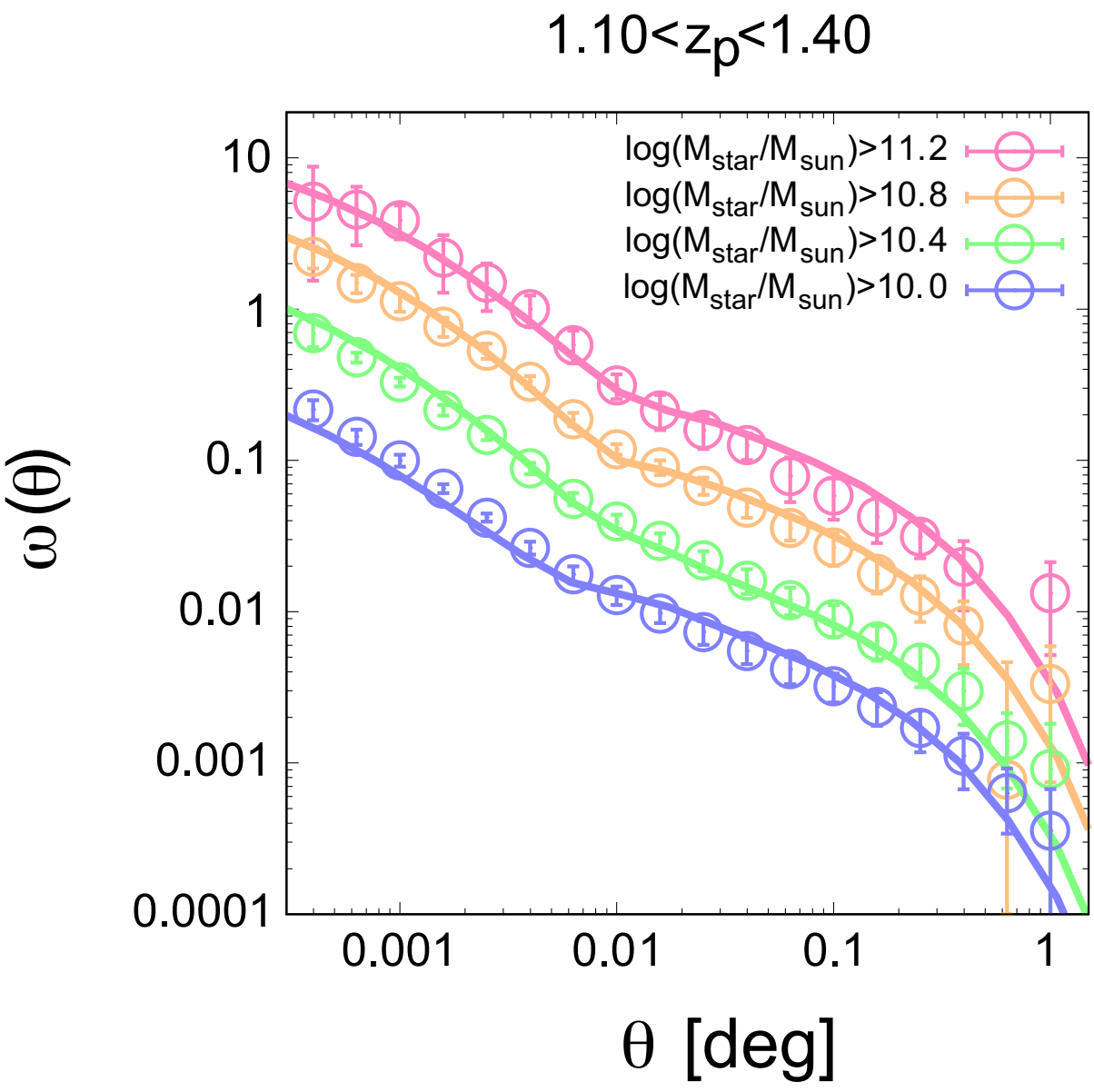

Figure 23.— Same as Figure 20, but for galaxies at $1.10 < z_{\rm phot} < 1.40$.

3.4. Results

Here, I present the results of the HOD analysis on the HSC galaxies. I exclusively discuss the redshift evolution and the stellar-mass dependence of the mean halo mass, the satellite fraction, and the SHMR of low-redshift galaxies.

3.4.1. Mean halo mass

I calculate mean halo masses of our photo-z galaxy samples using the equation (81) and the results are shown in Figure 24. As definition, the mean halo mass represents the mass of host haloes weighted by the number of galaxies within dark haloes. For comparison, I also plot the results of Wake et al. (2011), who implemented the HOD analysis on galaxies selected by the photometric redshifts using the data of the NEWFIRM Medium Band Survey (NMBS; van Dokkum et al. 2009).

The mean halo masses of less-massive galaxies gradually increase with increasing the limit of the stellar mass, whereas mean halo masses increase steeply for massive galaxies, indicating that most of the massive galaxies are central galaxies of massive dark haloes and a certain amount of less massive galaxies are satellite galaxies within massive haloes. This characteristic is also seen in the results of HOD analyses using the luminosity-limited galaxy samples (e.g., Coupon et al. 2012). In addition, at all redshift ranges, mean halo masses increase with decreasing the redshift by fixing the stellar masses of galaxy samples. This increase of the mean halo mass reflects that the halo mergers and halo growths are ongoing over cosmic time.

The results of Wake et al. (2011) is consistent with our mean halo masses of galaxies at $0.80 < z_{\rm phot} < 1.10$. The redshift range of Wake et al. (2011) is slightly different; however, most of their galaxy samples are in the redshift range of $0.80 < z_{\rm phot} < 1.10$ according to their redshift distribution. The errors of Wake et al. (2011) are relatively small compared to our results. This is largely due to the fixed HOD parameters; Wake et al. (2011) fixed as $\sigma_{\log M} = 0.15$ and $\alpha = 1.0$ in their HOD analyses.

3.4.2. Satellite fraction

Satellite fractions, which represent how many satellite galaxies exist in each stellar-mass limited bin, are presented in Figure 25. Definition of the satellite fraction is given in equation (82). I also show the results of satellite fractions of photometric-redshift selected galaxies of Wake et al. (2011) and McCracken et al. (2015).

The satellite fractions show small decrease or flatly evolution from less massive to intermediate stellar masses ($\log(M_{\star,\rm limit}/M_{\odot}) \sim 10.6$) and appear steeply cutoff for massive galaxies, which is the same trend of the local galaxies (e.g., Z. Zheng et al. 2007; Zehavi et al. 2011) and the results of simulations (e.g., Artale et al. 2016). One possible explanation of slightly evolution is that haloes merged onto other haloes at $z > 1$ are sufficiently destructed and taken harbored galaxies within merging haloes into satellite galaxies of merged haloes though cosmic time. Low satellite fractions of HSC galaxies at $1.10 < z_{\rm phot} < 1.40$ even for less massive galaxies agree with the above scenario: galaxies within merging haloes are rarely introduced as satellite galaxies due to the short time-scale after merging events. Satellite fractions of massive galaxies, on the other hand, show small and com-

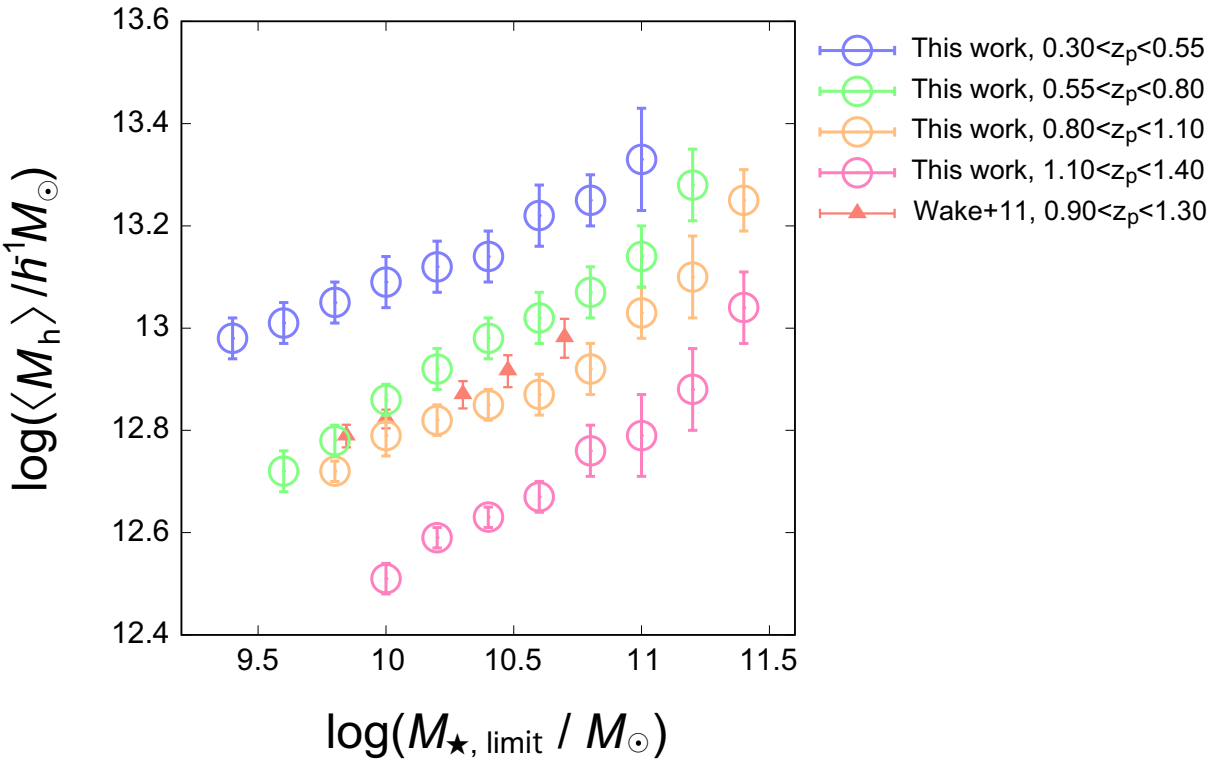

Figure 24.— Mean halo masses of HSC photo-z galaxies as a function of the stellar-mass threshold. For comparison, I also plot the results of photo-z selected galaxies at $0.9 < z_{\rm phot} < 1.3$ by Wake et al. (2011).

parable values across all of the redshift ranges, suggesting that massive galaxies tend to be central galaxies of moderately massive haloes than be satellite galaxies within massive dark haloes. However, further interpretation of behaviors of satellite fractions requires precise analyses, i.e., satellite fractions of star-forming/passive populations and fractions of star-forming galaxies help to interpret the characteristics of total galaxy samples.

By comparing with literature, our satellite fractions show good agreement with results of McCracken et al. (2015), especially for low-z bins, although satellite fractions at $1.1 < z_{\rm phot} < 1.5$ of McCracken et al. (2015) show systematically higher values than our results at $1.10 < z_{\rm phot} < 1.40$. This may be caused by the low accuracy of the photometric redshift estimation for their high-z galaxy samples; the SHMRs of high-z bins cannot be fitted by a double power-law functional form and authors implied the effect of unreliable photo-z estimation (see Figure 10 in McCracken et al. 2015). Satellite fractions of Wake et al. (2011), on the other hand, show higher values compared to our results as well as McCracken et al. (2015) over entire stellar-mass ranges. This could be due to the fixed HOD parameters in analyses of Wake et al. (2011). I also compare our results with satellite fractions of Skibba et al. (2015), who investigated the galaxy spatial clustering in the PRIsm MUlti-object Survey (PRIMUS; Coil et al. 2011; Cool et al. 2013) and the DEEP2 (Davis et al. 2003) surveys by analyzing them using both the HOD-based mock galaxy catalogues using the Bolshoi dark matter simulation (Klypin et al. 2011) with a halo-finding algorithm, "ROCKSTAR" (Behroozi et al. 2013c), and the analytical HOD model. Decreasing trend with increasing the stellar-mass thresholds and redshift is consistent with our results.

3.4.3. Stellar-to-halo mass ratio

Using the stellar masses of galaxies evaluated by the SED fitting technique and the dark halo masses derived by the HOD analysis, I can compute the SHMRs of the HSC galaxies. To compare with literature, stellar-mass limits and $M_{\rm min}$ are adopted for $M_\star$ and $M_{\rm h}$, i.e., SHMRs are computed by $M_{\star,{\rm limit}}/M_{\rm min}$.

I present the results of SHMRs of galaxies at each redshift bin in Figure 26 − Figure 29. For comparison, I also plot predictions by numerical simulations performed by Behroozi et al. (2013b) and Moster et al. (2013), results of the semi-analytical model by Somerville et al. (2015), and the observational result of Coupon et al. (2015). It is noted that SHMRs of Moster et al. (2013) is plotted at the median redshift of each redshift bin.

Our SHRMs are well consistent with model predictions by Behroozi et al. (2013b) within 1σ confidence levels, especially for less massive haloes of $M_{\rm h} < M_{\rm h}^{\rm pivot}$. This is due to our highly accurate HOD analysis based on the large number of galaxy samples.

The largest characteristic of our SHMRs are that massive ends of our SHMRs except for $0.30 < z_{\rm phot} < 0.55$ show excess compared to relations of Behroozi et al. (2013b), albeit within 1σ confidence intervals. The origin of these excesses is unclear. For $0.80 < z_{\rm phot} < 1.10$ and $1.10 < z_{\rm phot} < 1.40$ bins, our results also can be described by the results of $z \sim 0.8$ by Coupon et al. (2015), who investigated the SHMR by dividing the contribution from central galaxies and satellite galaxies. The consistency with Coupon et al. (2015) may imply that our SHMRs at massive ends also contain the SHMRs of satellite galaxies; however, fractions of satellite galaxies of massive stellar-mass bins are quite small. According

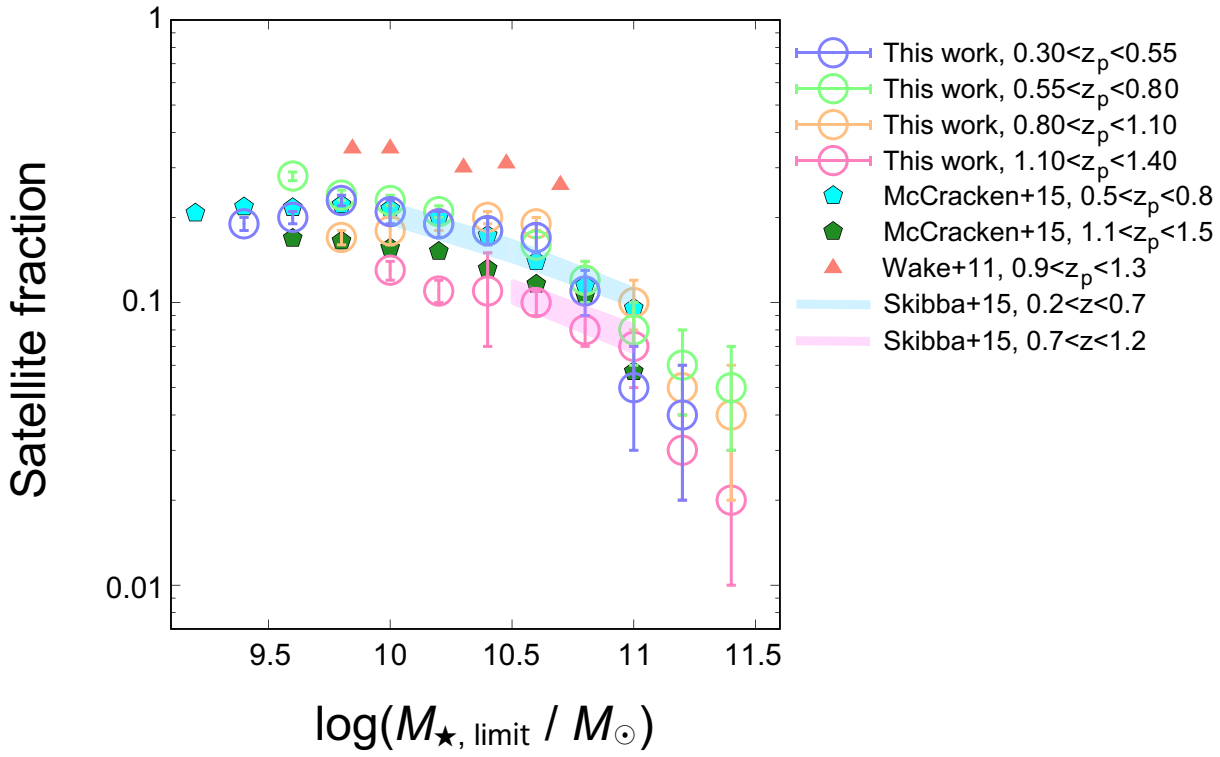

Figure 25.— Satellite fractions of HSC photo-z galaxies as a function of the stellar-mass threshold. For comparison, I also plot the results of satellite fractions of photometric-redshift selected galaxy samples of Wake et al. (2011) and McCracken et al. (2015), and using mock galaxy catalogues by Skibba et al. (2015).

to the results of HOD analyses, satellite fractions, $f_{\rm sat}$, are $f_{\rm sat} \lesssim 0.1$ for $\log(M_{\star,{\rm limit}}/M_\odot) > 11.0$ bins; it is not affirmable to dominate SHMRs of satellite galaxies with small fractions compared to central galaxies. Another possibility is just a statistical fluctuation due to the relatively small sample sizes. The S/N ratios of ACFs and the fitting results of massive bins are not good compared to less massive bins, i.e., values of reduced χ^2 fittings are $\chi^2/{\rm dof} \sim 4$ for massive bins at $0.55 < z_{\rm phot} < 0.80$ and $1.10 < z_{\rm phot} < 1.40$ galaxies. Nevertheless, the numbers of galaxy samples are more than $10,000$ and it seems to be enough sample sizes to carry out accurate HOD fittings, unless most of massive galaxy samples are contaminated by fake objects. To clarify the reason of these excesses at massive ends, I will check the SHMRs of the HSC galaxies using the latest photo-z catalogue that will be updated the number of sample galaxies and the accuracies of $z_{\rm phot}$ in the future.

For two low-z bins, our results are also reproduced by the results of Moster et al. (2013), who calculated the relations by the abundance-matching method, whereas, for two high-z bins, SHMRs of Moster et al. (2013) are underestimated compared to our observational SHMRs at less massive haloes. This may be caused by the observational constraints of Moster et al. (2013); stellar-mass functions to constrain in derivation of SHMRs of Moster et al. (2013) are not deep (Pérez-González et al. 2008; Santini et al. 2012). On the other hand, Behroozi et al. (2013b) used multiple stellar-mass functions computed by various types of surveys, i.e., broad band surveys (UV−IR, FIR; X. Zheng et al. 2007; Rujopakarn et al. 2010), narrow-band surveys (Hα; Ly et al. 2011), and radio surveys (1.4 GHz; Smolčić et al. 2009), to consider all of types of galaxy populations. Moreover, Behroozi et al. (2013b) also took into account the cosmic star-formation rates and the specific star-formation rates, whereas Moster et al. (2013) only considered the observational constraint for stellar-mass functions.

The SHMRs of Somerville et al. (2015), who derived the relations by the semi-analytical model based on Gnedin & Kravtsov (2010), are not conflict to our results around $M_{\rm h}^{\rm pivot}$. For less-massive end, our SHMRs of all of the redshift bins show smaller values than those of Somerville et al. (2015); this is due to the overproduction of less-massive galaxies in their simulations. Overproductions of low-mass galaxies also make $M_{\rm h}^{\rm pivot}$ small compared to our observational results as well as the theoretical predictions by other studies.

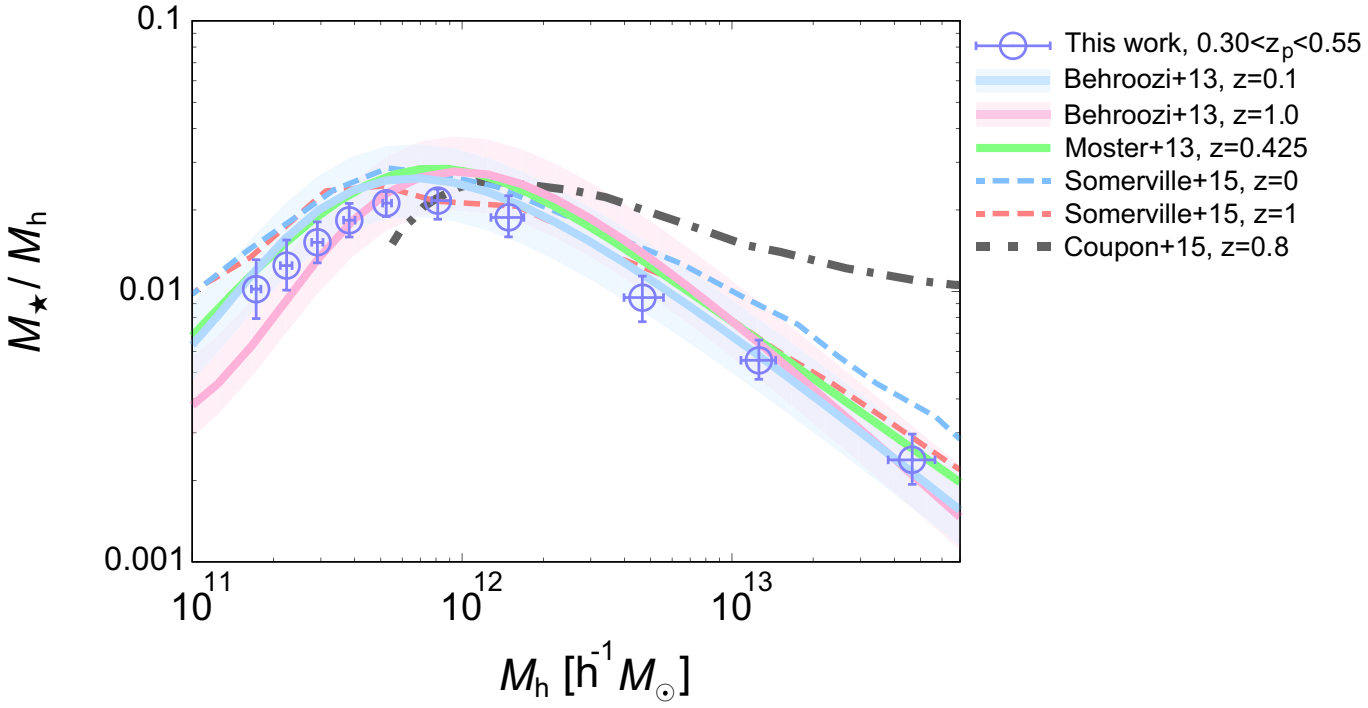

Figure 26.— Comparison of our observed SHMRs of galaxies at $0.30 < z_{\rm phot} < 0.55$ (median redshift: $\bar{z}_{\rm phot} = 0.425$) with literature. I compare our results with predictions by numerical simulations (solid blue line: Behroozi et al. 2013b at $z = 0.1$, solid red line: Behroozi et al. 2013b at $z = 1.0$, solid green line: Moster et al. 2013 at the median redshift), results of the semi-analytical model (dashed blue line: Somerville et al. 2015 at $z = 0$, dashed red line: Somerville et al. 2015 at $z = 1$), and the observational result (chain black line: Coupon et al. 2015 at $z = 0.8$). Stellar mass errors of our results are adopted for the root mean squares of the stellar mass of each stellar-mass bin, whereas the halo mass errors are the errors of $M_{\rm min}$ obtained by the HOD fittings. Shaded regions of Behroozi et al. (2013b) by each color represent the 1σ confidence intervals of each redshift.

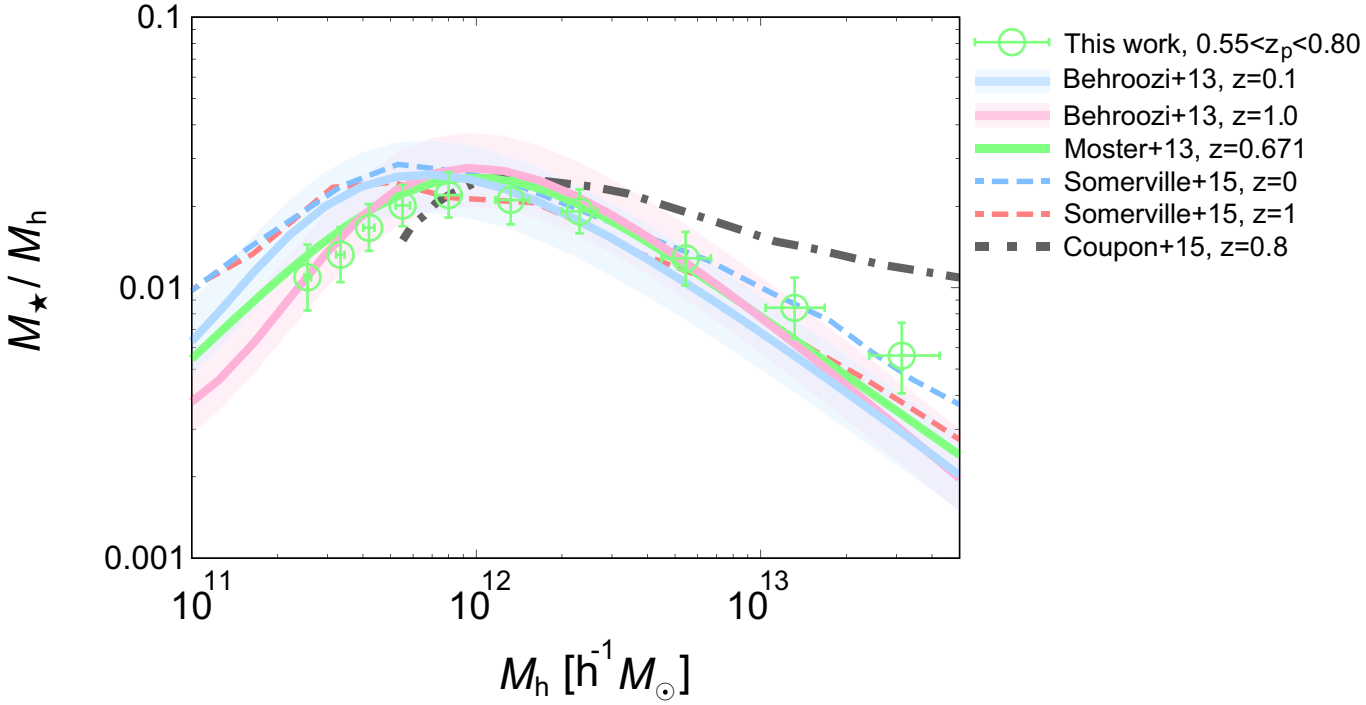

Figure 27.— Same as Figure 26, but for galaxies at $0.55 < z_{\rm phot} < 0.80$ (median redshift: $\bar{z}_{\rm phot} = 0.671$).

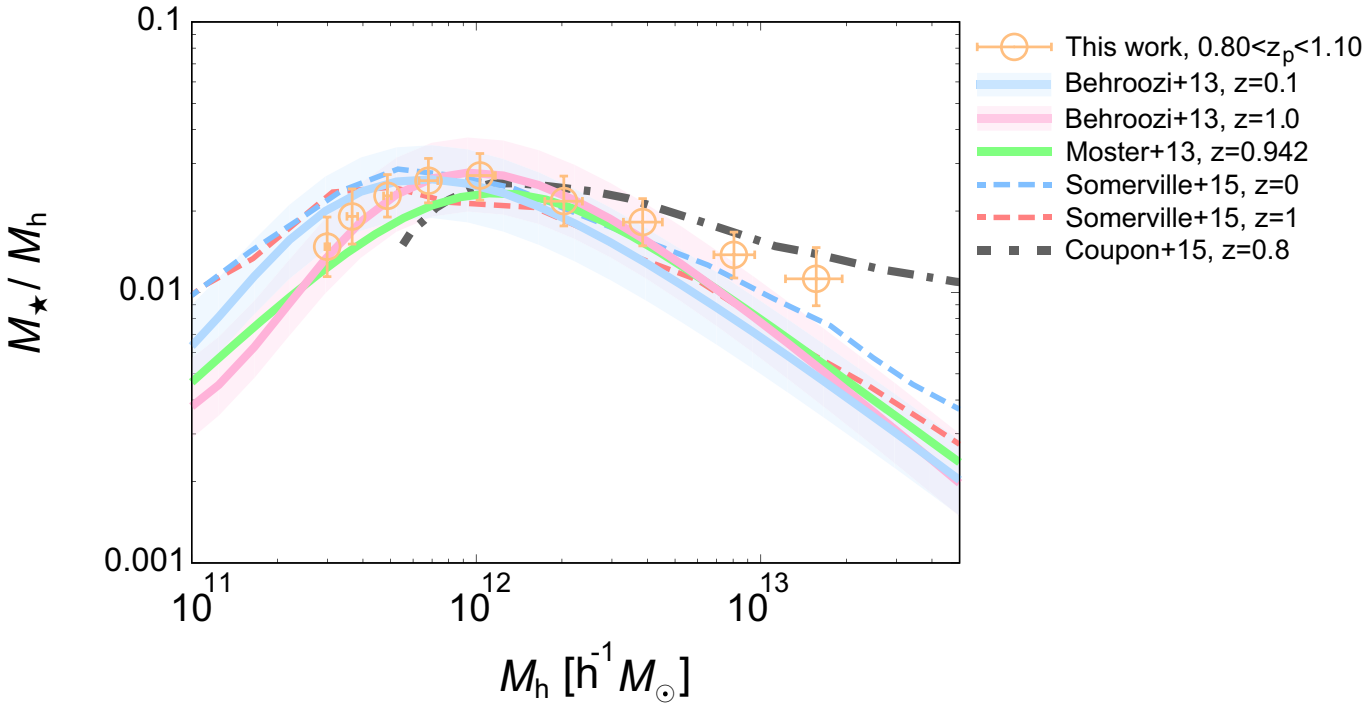

Figure 28.— Same as Figure 26, but for galaxies at $0.80 < z_{\rm phot} < 1.10$ (median redshift: $\bar{z}_{\rm phot} = 0.942$).

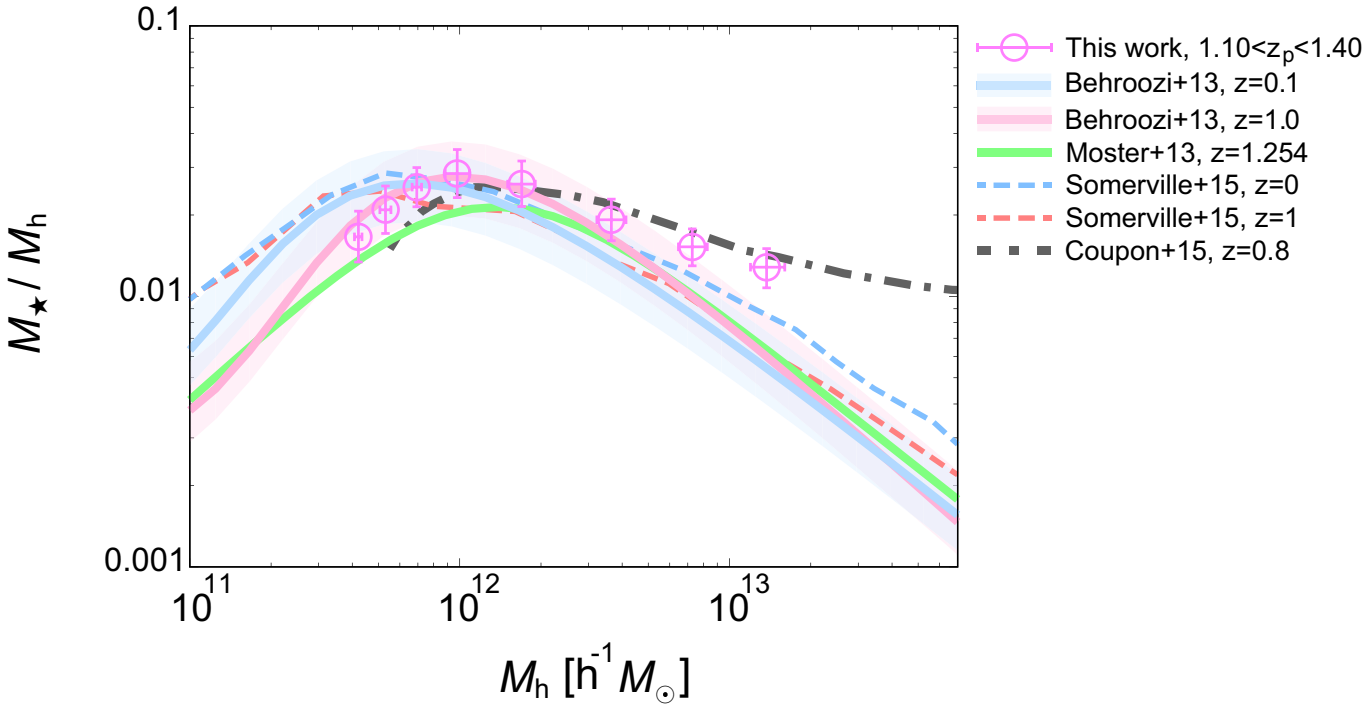

Figure 29.— Same as Figure 26, but for galaxies at $1.10 < z_{\rm phot} < 1.40$ (median redshift: $\bar{z}_{\rm phot} = 1.254$).

4. GALAXY–HALO CONNECTION IN MID-REDSHIFT UNIVERSE

4.1. Overview

4.1.1. Characteristics of mid-redshift

Galaxy observations have shown that $z \sim 2$ has many important characteristics. For example, star formation in galaxies peaked at $1 < z < 2$ (e.g., Hopkins & Beacom 2006; Burgarella et al. 2013), galaxy mass rapidly assembled at $1 < z < 3$ (e.g., Arnouts et al. 2007; Pérez-González et al. 2008; Santini et al. 2009; Ilbert et al. 2010), and the number density of QSOs peaked at $z \sim 2$ (e.g., Palanque-Delabrouille et al. 2013). Therefore, $z \sim 2$ is an important era to obtain clues to the formation and evolutionary history of galaxies by comparing the physical properties of other galaxy populations and tracing the mass evolution of dark haloes. To determine the accurate dark halo mass, it is required to collect a large number of galaxy samples to obtain the ACFs with high S/N ratio and implement a precision clustering analysis and the HOD analysis, if possible. However, $z \sim 2$ is also known the era in which it is difficult to select galaxies; thus, constructing a large galaxy sample and carrying out the accurate clustering analysis is still challenging.

The distinctive spectral features of galaxies at $z \sim 2$ are red-shifted from the optical wavelengths to the NIR wavelengths. Wide-field NIR observation using ground-based telescopes is complicated because of the limited physical size of the NIR detectors and (with the exception of some atmospheric windows) poor sky transparency, as well as strong OH airglow emission lines and thermal emission. For these reasons, wide-field galaxy survey data are lacking at $z \sim 2$ (this commonly referred to as the "redshift desert").

A powerful method to select unbiased data with high completeness $z \sim 2$ galaxies has been proposed, namely the "BzK selection" technique (Daddi et al. 2004). The BzK selection method requires only B-, z-, and K-band photometric data, and enables us to select both star-forming BzKs (hereafter sBzKs) and passively evolving BzKs (hereafter pBzKs) simultaneously in a single color–color diagram. However, a precise clustering analysis of BzK galaxies requires a large number of galaxy samples and has not yet been performed due to the difficulties in obtaining wide-field K-band imaging data.

4.1.2. Clustering analysis in mid-redshift Universe

A number of studies have attempted to reveal the clustering properties of BzK galaxies. Kong et al. (2006) constructed BzK galaxy samples over 1200 arcmin^2. They conducted the clustering analysis of sBzK galaxies and pBzK galaxies and concluded that pBzK galaxies are more strongly clustered than sBzK galaxies, suggesting that the morphology–density relation seen in the local Universe must already have been in place at $z \sim 2$.

Hayashi et al. (2007) carried out a deep BzK galaxy survey in the Subaru Deep Field to investigate the properties of faint sBzK galaxies and reported that the clustering strength of sBzK galaxies depends upon the K-band luminosity. They measured dark halo masses of sBzK galaxies by following

the procedure shown in the Section 2.1.4 and revealed that the minimum dark halo mass to possess the galaxy, $M_{\rm min}$ (the definition of this $M_{\rm min}$ is different from the $M_{\rm min}$ in the HOD parameter, albeit the same notation), and the mean halo mass, $\langle M_{\rm h} \rangle$, follow the relation of $\langle M_{\rm h} \rangle \approx 3 M_{\rm min}$.

McCracken et al. (2010) succeeded in constructing a large number of BzK galaxy samples in the COSMOS field (Scoville et al. 2007) over 1.9 deg^2 down to $K < 23.0$ and carried out a precise clustering analysis. Wide survey area and multi-wavelength data from the COSMOS survey enabled them to determine the comoving correlation length for both sBzK galaxies and pBzK galaxies, respectively.

More recently, Bielby et al. (2014) reported the clustering properties of $z \sim 2$ galaxies by splitting the galaxy samples into star-forming galaxies and passive galaxies based upon the photometric redshift and galaxy color. They revealed that passive galaxies are more strongly clustered than star-forming galaxies; however, very massive star-forming galaxies ($M_\star \sim 10^{11} M_\odot$) exhibited comparable clustering strength to passive galaxies. Their clustering analysis of star-forming galaxies at $z \sim 2$ was, however, confined to massive galaxies because of a shortage of galaxy samples at $z \sim 2$.

Béthermin et al. (2014) succeeded in calculating the dark halo mass of sBzK galaxies over wide mass range ($10^{10.4} M_\odot < M_\star < 10^{11.4} M_\odot$) using both galaxy clustering and HOD analyses in the COSMOS field. The dark halo masses determined using both methods were in good agreement and were consistent with the predictions of semi-analytical models. They firstly calculated the dark halo mass at $z \sim 2$ by the HOD analysis; however, the clustering analysis was carried out using the ACFs with bad S/N ratio, especially at the large-angular scale ($\theta \gtrsim 0.1^\circ$), which made it difficult to determine the dark halo mass accurately.

4.1.3. Motivation of this study

The most important issue to be resolved at this epoch is to determine the accurate dark halo mass. Some previous studies attempted to estimate the dark halo mass by the HOD analysis as well as the clustering analysis; however, the evaluated dark halo masses were suffered from large uncertainties due to the small number of galaxy samples. Béthermin et al. (2014) achieved to construct the largest galaxy sample at $z \sim 2$ at that time (25,683 sBzK galaxies) and carried out the HOD analysis, although the derived dark halo masses contain large errors, i.e., more that 0.2 dex for the smallest uncertainties.

It is also necessary to extend the survey area to obtain ACFs with high S/N ratio, which is essential to measure the precise correlation lengths as well as the dark halo masses with small errors. Moreover, small survey field may be affected by the cosmic variance. The largest survey area of the previous BzK studies conducted in the COSMOS field over 1.9 deg^2 (McCracken et al. 2010; Béthermin et al. 2014). However, it has been reported that the galaxy clustering patterns in the COSMOS field are reported to be different from other fields, especially at $z \sim 2$ (cf., McCracken et al. 2010; Sato et al. 2014). McCracken et al. (2015) found that the correlation length at $1.1 < z < 1.5$ are smaller than other redshift ranges in the COSMOS field and concluded that the peculiar characteristics of clustering at $z \sim 2$ is caused by the cosmic variance (cf., Clowes et al. 2013). Sato et al. (2014) computed the ACFs of the star-forming galaxies and the passive galaxies by combining the optical data of the CFHTLS Deep Fields with the NIR data of the WIRCam Deep Survey (WIRDS; Bielby et al. 2012) and compared their ACFs with those of sBzK galaxies (Kong et al. 2006; Hayashi et al. 2007;

Blanc et al. 2008; Hartley et al. 2008; McCracken et al. 2010), and found that the ACF of McCracken et al. (2010) showed clearly lower amplitude compared to other studies. Therefore, dark halo masses estimated by Béthermin et al. (2014) are thought to be also underestimated because they used the same dataset with McCracken et al. (2010).

To resolve the above problematic issues, I carried out the clustering analysis of star-forming galaxies at $z \sim 2$. By combining our original data with data from publicly available data, I collected g-, $z_{\rm B}/z$-, and K-band imaging data over 5.2 deg^2, which represents the largest area of BzK/gzK survey. Thanks to the wide-field survey area, I achieved to construct the largest-ever star-forming galaxy sample at $z \sim 2$ (sgzK galaxies) by applying the gzK selection method. A large number of sgzK galaxy samples enable to obtain high-quality ACFs over all angular scales enough to implement the HOD analysis to reveal the properties of host dark haloes of sgzK galaxies. Using the $(z_{\rm B} - K)$ color and the K-band magnitude, I evaluate the stellar mass of each sgzK galaxy, which enables to compute the SHMRs of the sgzK galaxies.

Throughout this section, I assume the flat lambda cosmology ($\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$), the Hubble constant as $h = {\rm H}_0/100\,{\rm km\,s^{-1}\,Mpc^{-1}} = 0.7$, and the matter fluctuation amplitude at $8h^{-1}$Mpc as $\sigma_8 = 0.8$. With these cosmological parameters, the age of the Universe at $z = 2$ is $\sim 2.26h^{-1}$Gyr and 1 arcsec corresponds to $5.86h^{-1}$kpc in the comoving scale. All of the results and discussions in this section are presented in Ishikawa et al. (2015) and Ishikawa et al. (2016).

4.2. Photometric Data and Sample Selection

4.2.1. Photometric data

I collected wide-field imaging data using both our own data and publicly available archive data. Photometric data of the K-band were obtained from the archives of the United Kingdom Infra-Red Telescope (UKIRT) Deep Sky Survey (UKIDSS; Lawrence et al. 2007) in the Deep Extragalactic Survey (DXS). The UKIDSS data were acquired using the Wide Field Camera (WFCAM; Casali et al. 2007), which is composed of four $2,048 \times 2,048$ pixel detectors, with a separation of $12.'8$ between detectors. Each detector covers $13.'65 \times 13.'65$ of the sky at a pixel scale of $0.''187$ pixel^{-1}. The central wavelength and the full-width at half-maximum (FWHM) of the K-band of the WFCAM are $\lambda_{\rm c} = 2.2\mu$m and FWHM $= 0.34\mu$m, respectively.

The $z_{\rm B}$-band photometric data are obtained during the course of another observational program (Kashikawa et al. 2015) using the Subaru Telescope/Suprime-Cam (Miyazaki et al. 2002) from June 22 to June 24, 2009. Suprime-Cam is a wide-field imaging instrument covering $34' \times 27'$ of the sky at a pixel scale of $0.''22$ pixel^{-1}. The $z_{\rm B}$-filter ($\lambda_{\rm c} = 8,842$Å, FWHM $= 689$Å) is a custom-made filter that divides the SDSS z-band filter at $9,500$Å and their bluer one (Shimasaku et al. 2005). One of our observation field "VIMOS4" is centered on $(22^{\rm h}20^{\rm m}00^{\rm s}, +00°42'00''; {\rm J2000.0})$ and covers $\sim$3 deg^2. Here, 3 deg^2 of the VIMOS/DXS field was covered by 13 Suprime-Cam field-of-views (FoVs). It is adopted a common dithering circle pattern comprising a full cycle of dithering consisting of 5 pointings. The total integration time was $1,800$ seconds per FoV. The sky condition was very good, with a seeing size of $0.''6$. Photometric calibration was carried out using the spectroscopic standard stars GD153 and Feige110. The data were reduced using the pipeline software package SDFRED

(Yagi et al. 2002; Ouchi et al. 2004a). The package includes bias subtraction, flat fielding, a correction for image distortion due to the prime focus, PSF matching, sky subtraction, and mosaicking.

The z-band photometric data are also retrieved from the SMOKA data archive server (Baba et al. 2002) to extend the survey field to 5.2 deg^2. These data satisfy the conditions that they connect continuously to our $z_{\rm B}$-band field, and the limiting magnitude is comparable to our $z_{\rm B}$-band data. The z-band data were also obtained using the Subaru Telescope/Suprime-Cam.

I retrieved g-band images, which cover the entire field where K and $z_{\rm B}/z$ images are available, using the data of the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Gwyn 2012) Wide Survey, instead of using B-band. The CFHTLS data were acquired using the CFHT MegaCam, which covers $57.'6 \times 56.'4$ of the sky at a pixel scale of $0.''187$ pixel^{-1}, where the central wavelength and FWHM of the g-band of MegaCam are $\lambda_{\rm c} = 4,870$Å and FWHM $= 1,450$Å, respectively.

The $(g, z_{\rm B}, K)$ bands of imaging data are collected over 3.2 deg^2 and (g, z, K) bands over 2.0 deg^2. The survey field of this study covers 5.2 deg^2, comprising the above two regions with different imaging datasets.

Fang et al. (2012) showed that the star-forming and the passive galaxies at $z \sim 2$ are also able to be selected using the $(g - z)$ and the $(z - K)$ colors, which is similar to BzK color selection. They constructed sgzK and pgzK samples by modifying the original BzK color-selection criteria and validated their criteria using the stellar population synthesis model (Bruzual & Charlot 2003). I applied a similar gzK selection process to our data, although color correction from our $(g - z_{\rm B})$ data to the VLT-$(B - z)$ data is required.

4.2.2. K-selected catalogue

The g- and $z_{\rm B}/z$-band images were matched to the geometry of the K-band image using the IRAF[1] task `geomatch`. Object detection and photometry were carried out using SExtractor, version 2.8.6 (Bertin & Arnouts 1996). Following object detection in the K-band, I used SExtractor in "double-image mode" to measure the object fluxes in the g- and $z_{\rm B}/z$-bands with $2''$ aperture photometry. The size of the unit image was $\sim 4,170 \times 4,170$ pixels, which corresponds to $\sim 27'.8 \times 27'.8$. Masking were carefully conducted on the low-S/N regions, such as near the edges of the images and close to the saturated objects, to remove objects with an uncertain flux. The number of masks was 690 over the field, and the effective survey area was 4.8 deg^2. In total, 339,581 objects were detected in our field after masking.

The limiting magnitudes were measured with a 3σ, $2''$ aperture in AB magnitude. I used the IRAF task `limitmag` in SDFRED to determine the limiting magnitudes and applied averaged data as limiting magnitudes in each band. Table 5 lists a summary of the limiting magnitudes and the areas covered by each band. These limiting magnitudes were almost identical (δ mag ~ 0.08) over the entire field.

[1] IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

Table 5: Limiting magnitudes and covering areas of each band

	limiting magnitude (3σ, $2''\phi$ in AB)	area (deg^2)
g	26.08	5.2
z_{B}	25.55	3.2
z	25.36	2.0
K	23.31	5.2

4.2.3. Band corrections

I used the BzK selection method to select $z \sim 2$ galaxy samples; however, the filter set of our data differed slightly from the VLT-BzK filter set (Daddi et al. 2004). Figure 30 shows a comparison of the transmission of our $gz_{\mathrm{B}}K$-filters with that of VLT-BzK filters. When adapting our $gz_{\mathrm{B}}K$ data to the VLT-BzK data, careful correction for differences in the filters is essential to adequately select $z \sim 2$ galaxies and to compare the results.

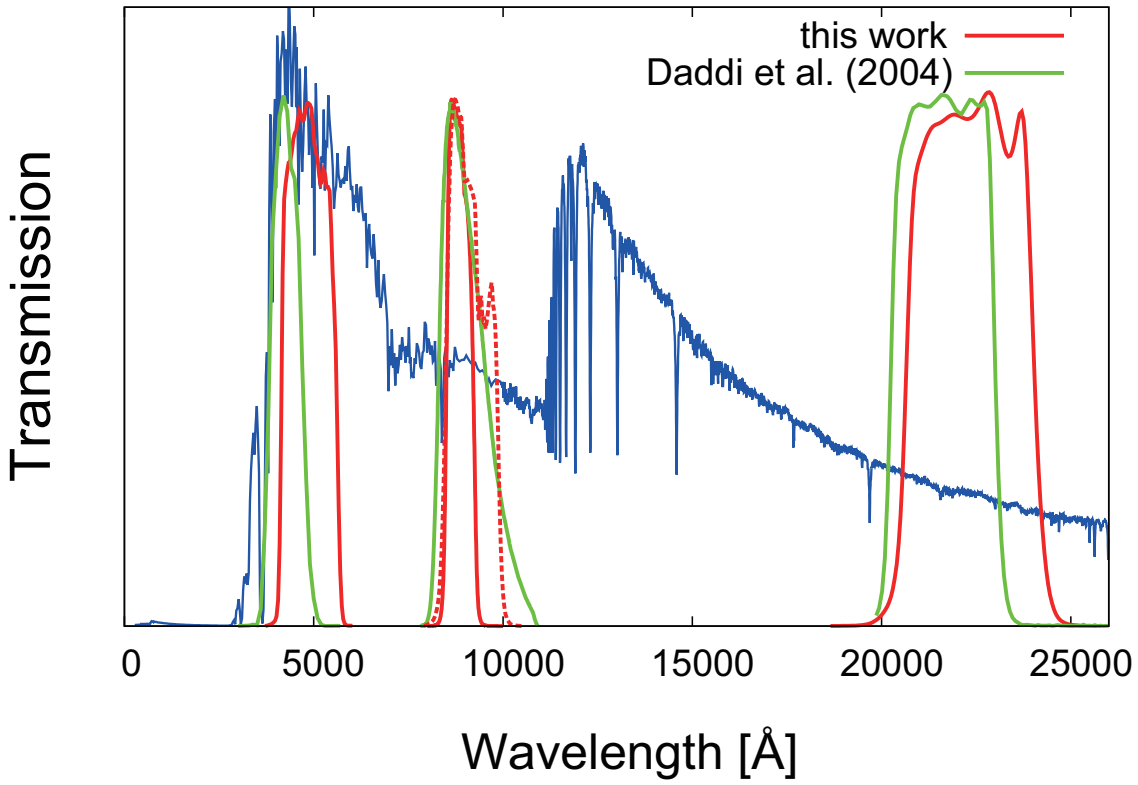

Figure 30.— Differences of the filter transmission profiles between our $gz_{\rm B}K$-filters (red) and the VLT-BzK filters (green) used in Daddi et al. (2004). Our filters are composed of CFHT/MegaCam g-band, Subaru/Suprime-Cam $z_{\rm B}/z$-band, and UKIRT/WFCAM K-band. Suprime-Cam z-band is presented by the dashed line. The blue line represents the typical SED of sgzK galaxies.

Daddi et al. (2004) used the K_s-band (Skrutskie et al. 2006), whereas I used K-band imaging data from the UKIDSS archive. I used the following conversion equation (Carpenter 2001) between these filters:

$$K_s = K + 0.002 + 0.026 \times (J - K). \tag{86}$$

J-band imaging data were also retrieved from the UKIDSS archive, as both K-band photometric data and J-band magnitudes were available. To simplify the notation, I hereafter describe the K_s-band magnitudes converted from K-band magnitudes using equation (86) as "K". Following conversion of the K-band magnitudes, our K-selected catalogue is limited down to $K = 23.0$, which corresponds to a 70% completeness cut (see Section 4.3.3). All galaxies were extracted from our catalogue using the BzK diagram (see Section 4.2.4), and sorted them in magnitude bins from $K = 16.0$ to $K = 23.0$ with $\delta K = \pm 0.25$.

Figure 31 shows a comparison of the number counts of our K-selected galaxies with previously reported studies. Our K-selected catalogue is consistent with previous works.

The BzK selection technique selects $z \sim 2$ galaxies in the color–color diagram of $(B - z)$ versus $(z - K)$. I attempted to carry out color conversion from $(g - z_B)$, $(g - z)$, and $(z_B - K)$ to $(B - z)$ and $(z - K)$ following K-band magnitude conversion. First, to correct for differences between the z_B-band and the z-band, I simultaneously measured the z_B- and the z-band fluxes for the stars in the ~ 0.22 deg^2 region, where photometric data are available for both, and derived the following conversion equation:

$$(z - K) = 0.940 \times (z_B - K) + 0.00330, \tag{87}$$

using a least-squares regression method.

In the same way, the following expressions were determined for conversion:

$$(g - z) = 0.900 \times (B - z)_{Subaru} - 0.120, \tag{88}$$

and

$$(g - z_B) = 0.961 \times (g - z) + 0.115, \tag{89}$$

respectively. I employed the public photometric catalogue of the COSMOS survey, in which g-, B-, and z-band photometries are available, to derive the equation (88).

McCracken et al. (2010) introduced empirical relations to translate $(B - z)$ color derived by the Subaru/Suprime-Cam data to $(B - z)$ color derived by the VLT for objects with $(B - z)_{Subaru} < 2.5$,

$$(B - z)_{VLT} = 1.0833 \times (B - z)_{Subaru} + 0.053, \tag{90}$$

and with $(B - z)_{Subaru} \geq 2.5$,

$$(B - z)_{VLT} = (B - z)_{Subaru} + 0.27. \tag{91}$$

Using these empirical relations and equation (88), I obtain the relations that convert from the Subaru-$(g - z)$ color to the VLT-$(B - z)$ color for objects with $(g - z) < 2.13$,

$$(B - z)_{VLT} = 1.20 \times (g - z) + 0.198, \tag{92}$$

and with $(g - z) \geq 2.13$,

$$(B - z)_{VLT} = 1.11 \times (g - z) + 0.404. \tag{93}$$

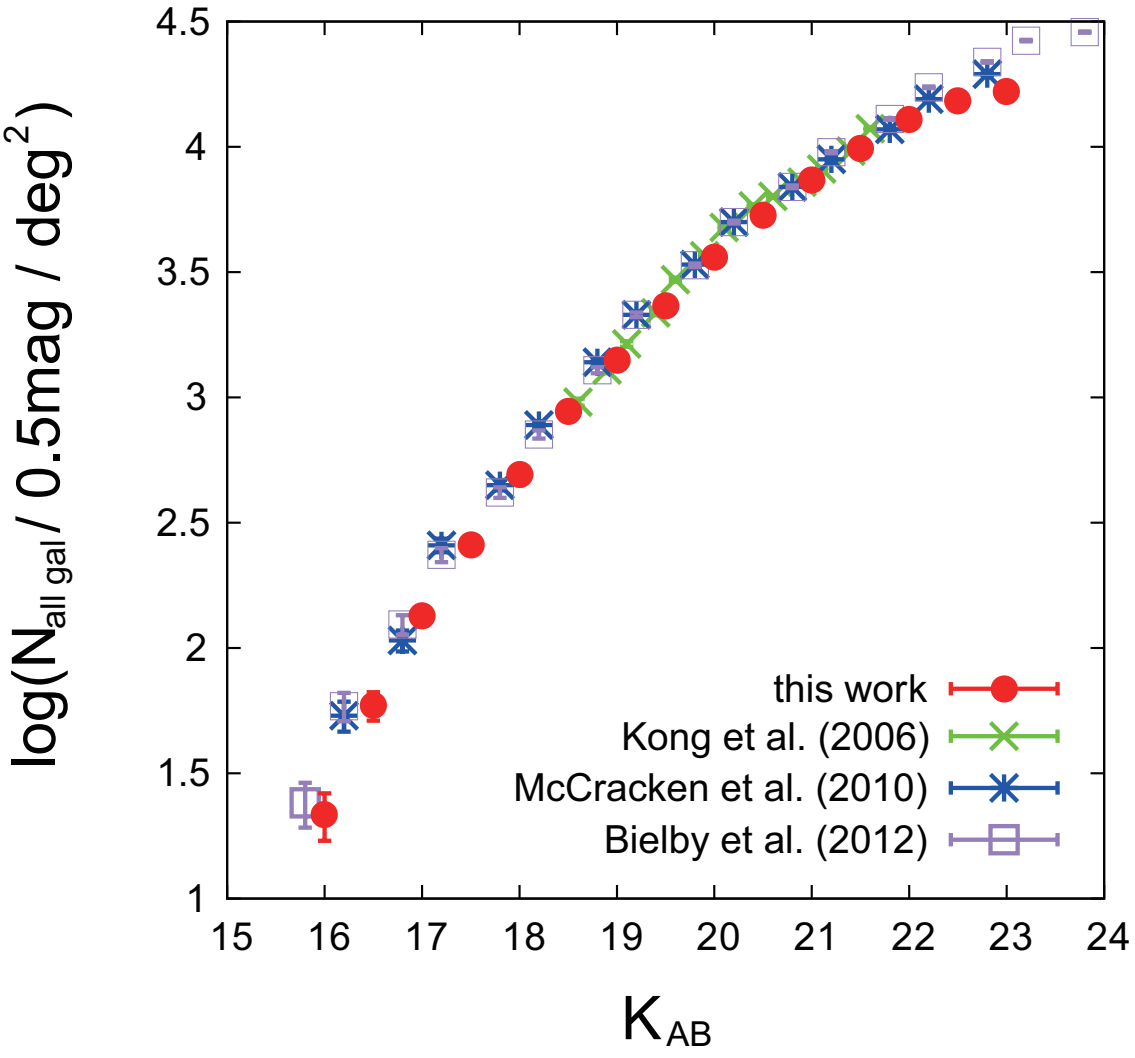

Figure 31.— Number count of the K-band selected galaxies after the correction of the detection completeness in our VIMOS4 field (red circles). Green, blue, and purple symbols represent the results of Kong et al. (2006), McCracken et al. (2010), and Bielby et al. (2012), respectively.

To convert from the Subaru-$(g - z_{\rm B})$ color for objects with $(g - z_{\rm B}) < 2.16$,

$$(B - z)_{VLT} = 1.25 \times (g - z_{\rm B}) + 0.0535, \tag{94}$$

and with $(g - z_{\rm B}) \geq 2.16$,

$$(B - z)_{VLT} = 1.16 \times (g - z_{\rm B}) + 0.271. \tag{95}$$

These color convention equations confirmed their validities using simulated galaxy SED models. The various SED models of galaxies were generated by varying physical properties such as the star-formation history, the star-formation rate, the age, and the dust extinction using the stellar population synthesis models (Bruzual & Charlot 2003). The dust extinction law was assumed to follow Calzetti et al. (2000). The $(g - z_{\rm B})$ and $(z_{\rm B} - K)$ colors were determined from those SEDs, and translated them to the $(B - z)$ and $(z - K)$ colors using the relations presented above, and then traced the redshift evolution on the color–color diagram. I found that the star-forming galaxies and the passively evolving galaxies at $1.4 \lesssim z \lesssim 2.5$ with our band corrections met the original selection criteria proposed by Daddi et al. (2004).

4.2.4. gzK selection method

I constructed the $z \sim 2$ star-forming galaxy samples by applying the BzK selection method. Our BzK selection method is termed as the "gzK selection" method, and selected star-forming galaxy samples are termed as sgzK galaxies because I use g-band photometric data instead of B-band data.

The criterion used to select the sBzK galaxies proposed by Daddi et al. (2004) is given by

$$(z - K) \geq (B - z) - 0.2, \tag{96}$$

and that for pBzK is given by

$$(z - K) < (B - z) - 0.2 \cap (z - K) > 2.5. \tag{97}$$

These original criteria is applied after converting $(g, z_{\rm B}, z, K_{\rm s})$ photometric data to (B, z, K) photometric data using equations (87) and (92)–(95).

Figure 32 shows a color–color diagram for all the objects down to $K = 23.0$ in our VIMOS4 field, where the blue and red dots represent the sgzK galaxies and the pgzK galaxies, respectively. The number of selected sgzK galaxies was N(sgzKs) = 41,112, and the number of pgzKs was N(pgzKs) = 1,313. Our g-band photometry was too shallow to detect all of the pgzK galaxies, which had very red $(g - z)$ color (at least $g - z > 2.7$); therefore, I consider only the sgzKs galaxies. It should be noted that the g-band limiting magnitudes of non-detected objects in the g-band were replaced by the 2σ limiting magnitude in the g-band, whereas non-detected objects in the $z_{\rm B}/z$-band were removed from our catalogue because their locations in the color–color diagram were uncertain.

Figure 33 shows a comparison of the number count between the sgzK and the sBzK galaxies from our data and from previous studies. After correcting the completeness (see Section 34), our data were found to be consistent with almost all previous studies. Table 6 lists a summary of the number count of our sgzK galaxies. Our results were in agreement with almost all previous studies and were in particularly good agreement with the data reported by McCracken et al. (2010). Our study and

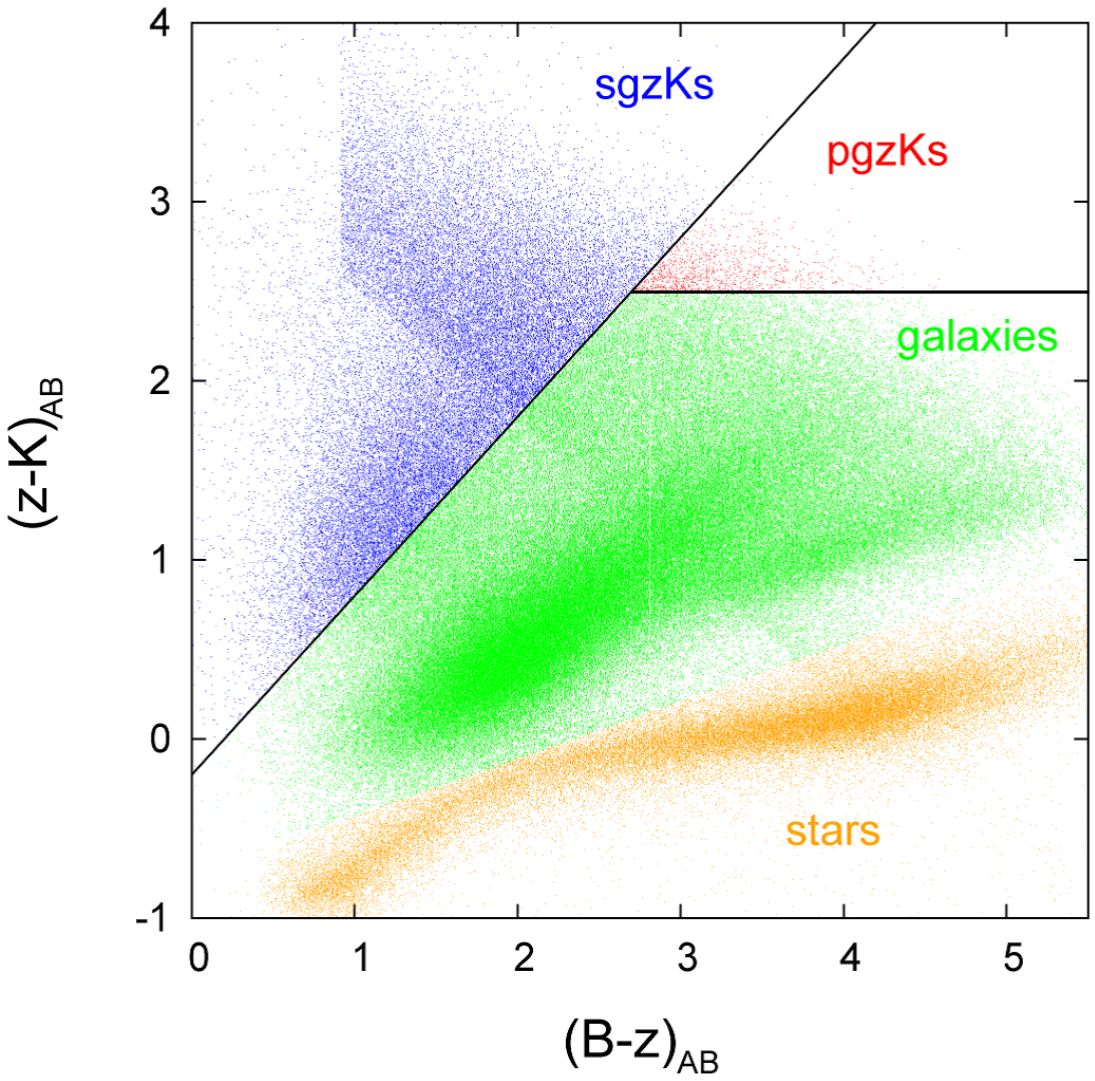

Figure 32.— A color–color diagram of our VIMOS4 field. The blue, red, green and orange dots represent sgzK galaxies, pgzK galaxies, other galaxies and stars, respectively.

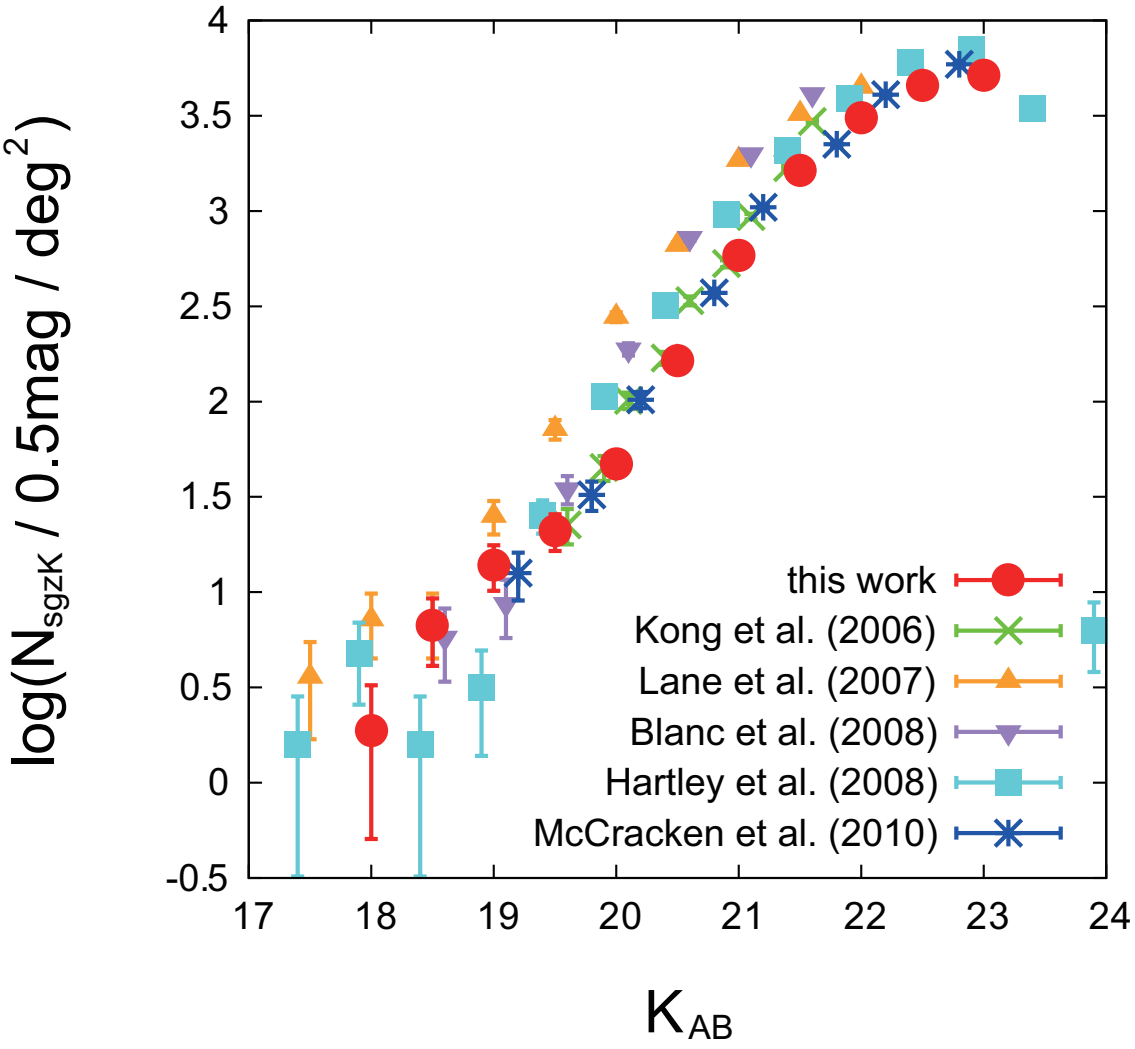

Figure 33.— Comparison of the number counts between our sgzK galaxies (red) after completeness correction and the sBzK galaxies of the previous studies. Green crosses, orange uptriangles, purple downtriangles, cyan squares and blue stars correspond to Kong et al. (2006), Lane et al. (2007), Blanc et al. (2008), Hartley et al. (2008) and McCracken et al. (2010), respectively.

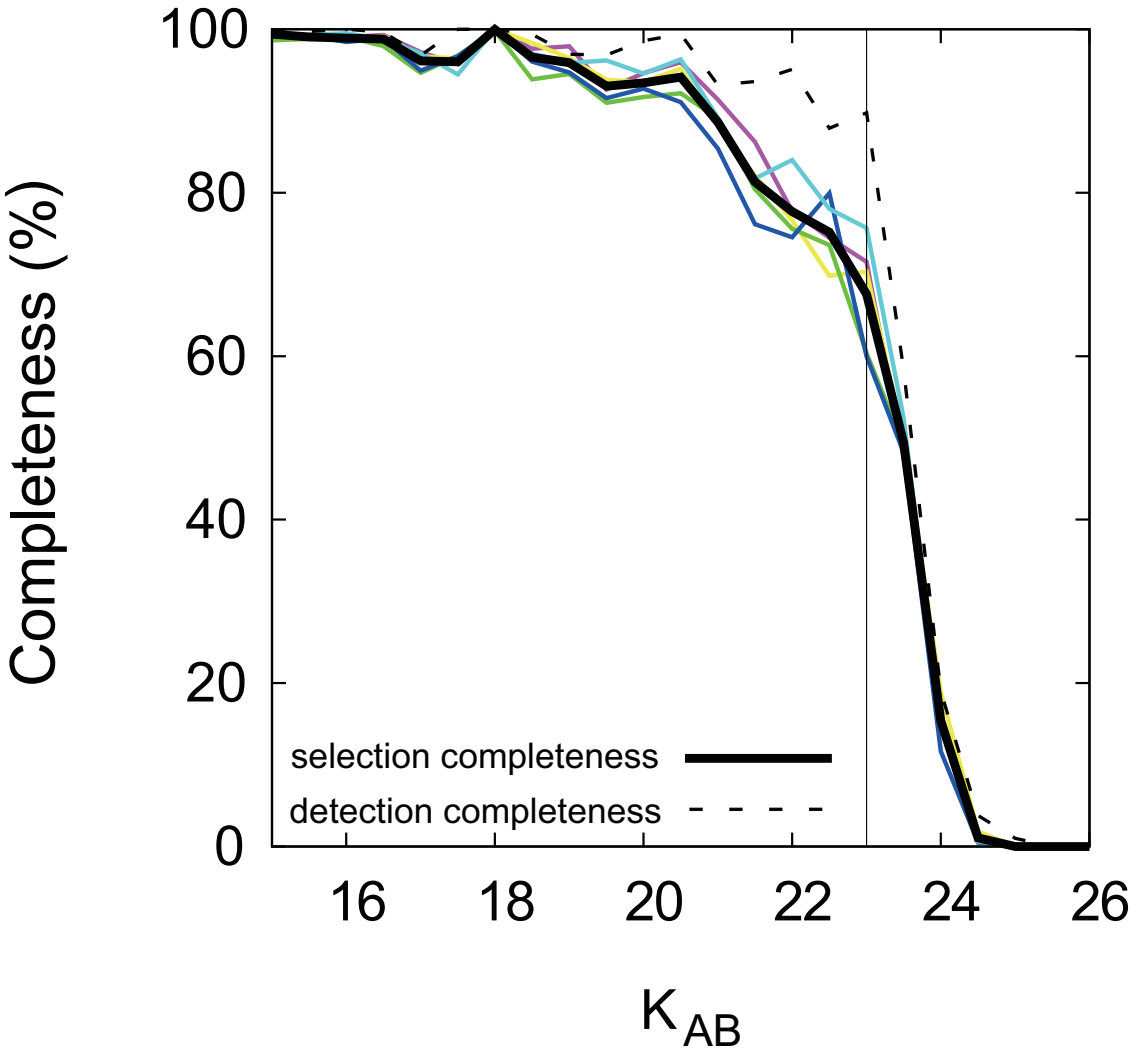

Figure 34.— Completeness of our sample. The colored thin lines represent the results in different regions and the black solid thick line is the average of the results in 5 regions. The dotted line is the detection completeness, which represents how inputted sgzK galaxies were detected as objects. I applied $18.0 \leq K \leq 23.0$ magnitude cut to our sample, corresponding to approximately 70% completeness cut.

Table 6: The sgzK number count of each magnitude bin after completeness correction.

$K_{\rm AB}$	$\rm N_{sgzK}$	$\log(\rm N_{sgzK})$ deg^{-2}
18.0	9	0.273
18.5	32	0.825
19.0	67	1.143
19.5	101	1.323
20.0	225	1.673
20.5	790	2.217
21.0	2,801	2.766
21.5	7,844	3.213
22.0	14,789	3.489
22.5	21,875	3.659
23.0	24,704	3.712

McCracken et al. (2010) are particularly in good agreement each other, while other previous studies have slightly higher number counts than our study. This might be caused by their small survey area.

However, the bright end of our sgzK number count showed an excess compared with these data reported by Blanc et al. (2008) and Hartley et al. (2008). Our very wide survey field enabled us to detect quite rare bright sgzKs at $K \lesssim 19$.

In our color–color diagram, there are some anomalous features. First, in the sgzK region, one can see the sharp vertical line at $(B-z, z-K) \sim (1.0, 2.6)$ to $\sim (1.0, 3.3)$, which originated from the non-detected objects in the g-band. The z/$z_{\rm B}$-band magnitudes of the objects on this line are so close to the limiting magnitude that the $(B-z)$ color is nearly constant. It is noted that the $(B-z)$ color was converted from original $(g-z)$ and $(g-z_{\rm B})$ colors by equation (92) and (94). In addition, a coherent line is seen from $(B-z, z-K) \sim (1.0, 2.6)$ to $\sim (4.0, 0.0)$ in the diagram. This line is caused by the limiting magnitude in the K-band, making the constant $(B-K)$ color.

As shown in Figure 33, the number count of our sgzK sample is consistent with the previous results, suggesting that these anomalous features do not largely affect the sgzK selection; however, g-band faint pgzK galaxies, which show stronger correlation, especially at the small angular scale, than sgzK galaxies (e.g., McCracken et al. 2010; Zehavi et al. 2011), could come into the sgzK sample. To estimate the contamination of g-band faint pgzK galaxies in our sgzK sample, I use the K-selected photometric redshift (photo-z) catalogue in the COSMOS field (Muzzin et al. 2013), which is approximately 1.0 magnitude deeper than our sample. The sgzK and pgzK galaxies were extracted from the COSMOS photo-z catalogue by applying our filter correction to that catalogue and the same color selection criteria. Then, I evaluated how many pgzK galaxies are incorrectly selected as sgzK galaxies by replacing g-band magnitudes by the 2σ limiting magnitude when the objects are fainter than the 2σ limiting magnitude. The contamination fractions of each limiting magnitude were found to be $f^{\rm pgzK}_{22<K<23} = 8.8\%$, $f^{\rm pgzK}_{21<K<22} = 8.8\%$, and $f^{\rm pgzK}_{18<K<21} = 6.0\%$, respectively. Therefore, I concluded that the contamination from pgzK galaxies to our sgzK sample is not substantial.

4.3. Clustering Analysis

4.3.1. Angular correlation function

I computed the ACF of the sgzK galaxies using the estimator of Landy & Szalay (1993). The angular bin was set from $\log(\theta) = -3.6$ to $\log(\theta) = 0.2$, with a bin size $\delta\log(\theta) = 0.2$ to determine the ACF for sgzKs galaxies. $300,000$ random samples were generated homogeneously over the survey field in order to reduce the Poisson error. The random samples were distributed to avoid the masked region, which we used when obtaining sgzK galaxy samples. Errors and the covariance matrices, C_{ij}, of our sgzK galaxy samples were calculated by the jackknife resampling method. I divided our survey region into $N = 64$ and calculated the ACF by removing one sub-region, repeating this process 64 times.

To confirm the validity of the homogeneity, I check the ACF of stars in our sample. The distribution of stars can be regarded as random distribution; therefore, their ACF is expected to be almost zero at all angular scale. Figure 35 shows the ACF of stars. The amplitude of each angular bin is approximately zero, indicating that our sample are almost selected homogeneously.

4.3.2. Subsamples

I divided our sample into subsamples according to the K-band luminosity to investigate the luminosity dependence of the clustering properties. Two subsample sets were created: a cumulative luminosity resampling subset and a differential luminosity resampling subset. The former was for use in comparing the results with previous studies, in which cumulative luminosity subsamples were created due to the small limited of sBzK data. Our large sgzK sample were divided into five subsamples, i.e., $18.0 \leq K \leq 21.0$, $18.0 \leq K \leq 21.5$, $18.0 \leq K \leq 22.0$, $18.0 \leq K \leq 22.5$, and $18.0 \leq K \leq 23.0$. The sample number of each subsample are summarized in Table 7.

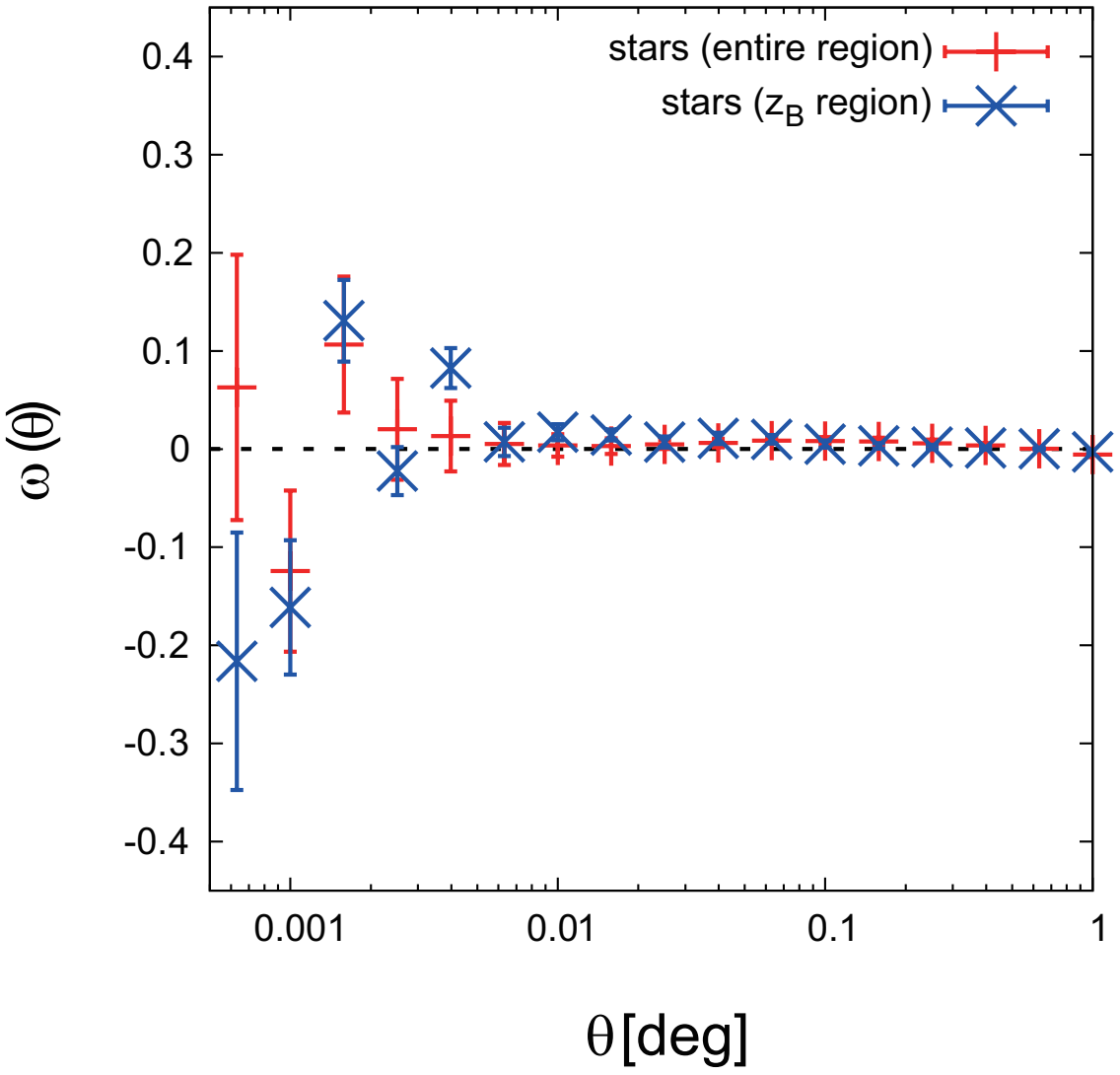

Figure 35.— The ACFs of stars (red lines, entire survey region; blue crosses, z_B-band covered region) in our survey field satisfying $18.0 \leq K \leq 23.0$.

Table 7: Detailed data of cumulative/differential subsamples

	N(sgzK)	$z_{\rm eff}$	$A_\omega(1^\circ) \times 10^{-3}$	r_0 $[h^{-1}$ Mpc]	$b_{\rm eff}$	$M_{\rm min}$ $[h^{-1}M_\odot]$	$M_{\rm h}$ $[h^{-1}M_\odot]$	$M_\star$ $[h^{-1}M_\odot]$
$18.0 \leq K \leq 21.0$	2,086	1.43	18.48 ± 4.57	10.52 ± 1.30	4.16 ± 0.46	$(1.71^{+0.76}_{-0.60}) \times 10^{13}$	$(3.26^{+1.23}_{-1.02}) \times 10^{13}$	$(1.61^{+1.83}_{-0.85}) \times 10^{11}$
$18.0 \leq K \leq 21.5$	6,243	1.61	7.23 ± 1.40	5.61 ± 0.57	2.36 ± 0.21	$(1.84^{+0.97}_{-0.63}) \times 10^{12}$	$(4.39^{+1.82}_{-1.45}) \times 10^{12}$	$(1.24^{+1.00}_{-0.55}) \times 10^{11}$
$18.0 \leq K \leq 22.0$	15,247	1.68	4.36 ± 0.27	4.67 ± 0.17	2.00 ± 0.07	$(8.45^{+1.75}_{-1.50}) \times 10^{11}$	$(2.14^{+0.27}_{-0.31}) \times 10^{12}$	$(8.64^{+7.00}_{-3.87}) \times 10^{10}$
$18.0 \leq K \leq 22.5$	29,158	1.73	3.83 ± 0.14	4.47 ± 0.08	1.92 ± 0.03	$(6.78^{+0.54}_{-0.50}) \times 10^{11}$	$(1.78^{+0.14}_{-0.13}) \times 10^{12}$	$(6.13^{+5.56}_{-2.92}) \times 10^{10}$
$18.0 \leq K \leq 23.0$	41,112	1.76	3.33 ± 0.09	4.12 ± 0.07	1.79 ± 0.03	$(4.60^{+0.38}_{-0.38}) \times 10^{11}$	$(1.23^{+0.10}_{-0.10}) \times 10^{12}$	$(4.95^{+4.80}_{-2.44}) \times 10^{10}$
$18.0 \leq K \leq 21.0$	2,086	1.43	18.48 ± 4.57	10.52 ± 1.30	4.16 ± 0.46	$(1.71^{+0.76}_{-0.60}) \times 10^{13}$	$(3.26^{+1.23}_{-1.02}) \times 10^{13}$	$(1.61^{+1.83}_{-0.85}) \times 10^{11}$
$21.0 < K \leq 22.0$	13,163	1.65	4.81 ± 0.25	4.88 ± 0.26	2.08 ± 0.06	$(1.07^{+0.15}_{-0.17}) \times 10^{12}$	$(2.55^{+0.34}_{-0.32}) \times 10^{12}$	$(6.02^{+3.91}_{-2.37}) \times 10^{10}$
$22.0 < K \leq 23.0$	25,865	1.73	3.32 ± 0.09	4.18 ± 0.07	1.81 ± 0.03	$(4.93^{+0.44}_{-0.43}) \times 10^{11}$	$(1.32^{+0.09}_{-0.12}) \times 10^{12}$	$(2.80^{+1.77}_{-1.08}) \times 10^{10}$

With the cumulative subsample set, however, it was difficult to describe the clustering properties of sgzKs by their K-band luminosity because cumulative resampling dilutes the differences among luminosity subsamples. Taking advantage of the large number of samples, we also divided our sample differentially according to the luminosity to assess the K-band luminosity dependence of sgzK clustering. I divided our sgzK samples into three subsamples based upon their K-band luminosity, i.e., $18.0 \leq K \leq 21.0$, $21.0 < K \leq 22.0$, and $22.0 < K \leq 23.0$, which allowed us to investigate how clustering and the physical properties of sgzK galaxies depend on the K-band luminosity.

4.3.3. Redshift distributions and completeness

I estimated the redshift distribution of our gzK-selected sample by applying the gzK selection method (including the filter corrections) to the K-selected COSMOS photo-z catalogue (Muzzin et al. 2013). It is assumed that the redshift distributions estimated from the COSMOS field (~ 1.9 deg^2) were the same as our large (~ 5.2 deg^2) field and that the cosmic variance was negligible.

I used `z_m2`, which usually provides the most feasible $z_{\rm phot}$, as described in the EAZY manual (Brammer et al. 2008) in the COSMOS photo-z catalogue. I extracted galaxies with a K-band magnitude in the range $18.0 \leq K \leq 23.0$ at $0 < z_{\rm phot} < 3$ and constructed a galaxy catalogue containing $79,284$ galaxies. It should be noted that I excluded the galaxies with a large $z_{\rm phot}$ error, i.e., ($|\texttt{zp} - \texttt{z\_m2}| > 0.5$), where `zp` is the best-fit redshift determined using SED fitting. The number of excluded galaxies was 318.

The $(g - z_{\rm B})$ and $(z_{\rm B} - K)$ colors were calculated using the best-fit SEDs and by estimating $z_{\rm phot}$ for each galaxy, and the g- and $z_{\rm B}$-band magnitudes were derived according to the K-band magnitude from the COSMOS photo-z catalogue. Artificial galaxies with assigned $(g, z_{\rm B}, K)$ band photometries were then randomly distributed as point sources on the g-, $z_{\rm B}$-, and K-band images. A total of $6,000$ distributed galaxies were randomly selected from the galaxy catalogue allowing for redundancy in each limiting magnitude. Source detection and sgzK color selection were carried out as with our data, and I obtained the redshift distributions for each luminosity subsample by repeating these processes 100 times and averaging the results of all steps.

Figure 36 and 37 show the redshift distributions that satisfy the limiting magnitudes of our subsamples. The redshift range that satisfies the sgzK criterion is $1.4 \lesssim z \lesssim 2.5$, which indicates the validity of our colour corrections. Our redshift distributions are almost consistent with those of previous studies (e.g., McCracken et al. 2010; Fang et al. 2012), though these are slightly shifted to lower z. This could be caused by the differences in the filters and the slight shallowness in the K-band of our sample. It should be noted that the brightest subsample ($18.0 \leq K \leq 21.0$) may be contaminated from low-z galaxies. I corrected for the effects of this contamination in the ACF for this brightest subsample by multiplying $1/(1 - f_c)^2$, where f_c is the contamination fraction ($f_c \sim 0.2$), assuming that the contaminating sources are not clustered each other. It should be noted, however, that this assumption is strictly incorrect for contaminated low-z galaxies, whose clustering amplitudes are difficult to be estimated. Other subsamples exhibited less contamination with low/high-z galaxies ($f_c < 0.1$), and we did not apply a correction factor to these data.

The redshift distribution that I estimated using the above procedure shows the completeness of our sample. Figure 34 shows our sample completeness; we obtained completeness of 70% with a

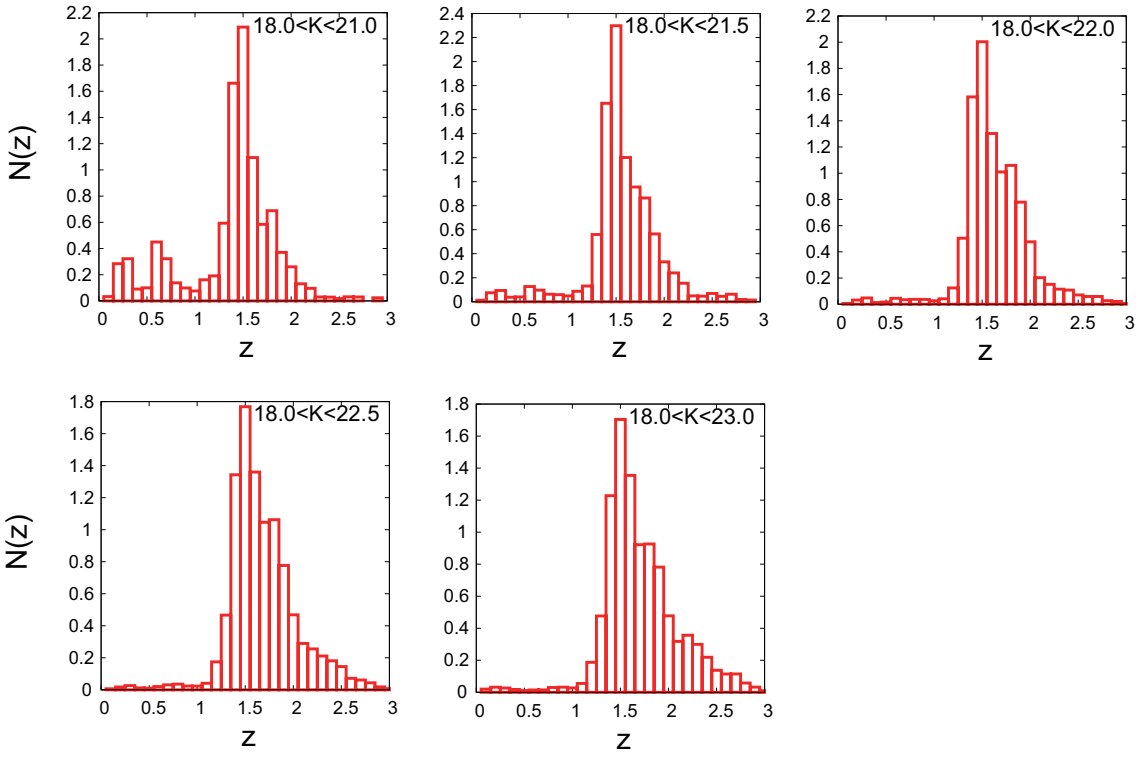

Figure 36.— Redshift distributions of each cumulative subsample. The mean number of redshift bin with $\delta z = 0.1$ width is normalized to one. I extracted the galaxies at $0 < z < 3$ from the COSMOS K-selected catalogue (Muzzin et al. 2013) and applied gzK selection as we did on our own galaxy sample.

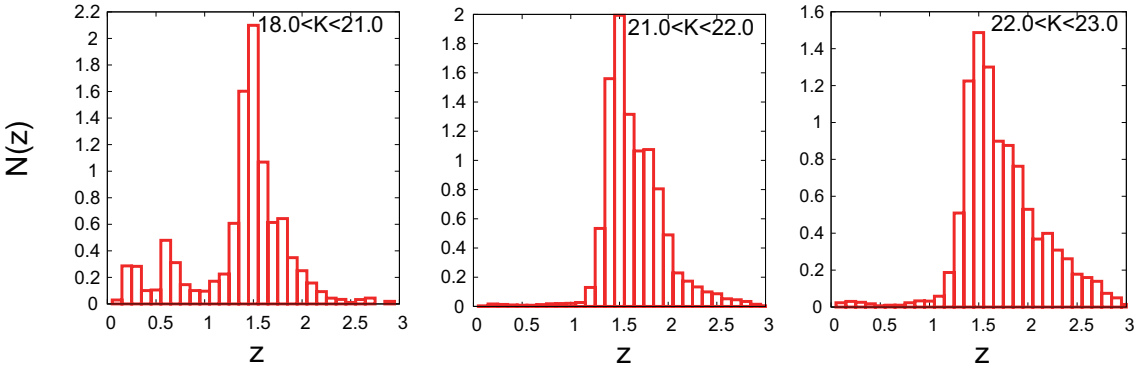

Figure 37.— Redshift distributions of each differential subsample. The method to derive these redshift distributions is the same as Figure 36.

limiting magnitude of $K_{\rm AB} = 23.0$.

4.4. Clustering Properties of sgzK Galaxies

4.4.1. ACFs of cumulatively resampled sgzK galaxies

The upper panel of Figure 38 shows the ACFs of each limiting-magnitude subsample. Bright sgzK galaxies were more strongly clustered than faint sgzK galaxies, which is consistent with previous studies (e.g., Kong et al. 2006; Hayashi et al. 2007; McCracken et al. 2010). The bin sizes of bright sgzK subsamples ($18.0 \leq K \leq 21.0$, $18.0 \leq K \leq 21.5$) were doubled ($\delta\log(\theta) = 0.4$) to gain the S/N ratio. Because of the accurate ACFs due to the large number of sgzK samples, our ACFs show an apparent excess from the power law at small angular scale (i.e., $\theta \lesssim 0.01^\circ$). These characteristics are attributed to the so-called "1-halo term", which comes from galaxy clustering in the same dark halo. $\theta \sim 0.01^\circ$ corresponds to $0.21\,h^{-1}$Mpc at $z \sim 2$ at a physical scale and is comparable to the virial radius ($\sim r_{200}$) of a dark halo at $z = 2$ with a mass of $\sim 10^{13} M_\odot$, where the virial radius is $r_{200} \sim 0.12\,h^{-1}$Mpc. The virial radius of the dark halo was estimated using

$$r_{200} = \left[\frac{GM_{200}}{100H^2(z)}\right]^{1/3}, \tag{98}$$

where G is the gravitational constant, M_{200} is the virial mass of the dark halo (Carroll et al. 1992; Ferguson et al. 2004), and $H(z)$ is the Hubble parameter with a redshift z as given by

$$H(z) = H_0[\Omega_m(1+z)^3 + (1-\Omega_m-\Omega_\Lambda)(1+z)^2 + \Omega_\Lambda]^{1/2}. \tag{99}$$

However, the ACFs at large angular scale (i.e., $\theta \gtrsim 0.01^\circ$) were well approximated by a power law with an index of $\gamma = 1.8$, referred to as the "2-halo term", which originates from galaxy clustering in different dark haloes. These two components can be accurately described by the HOD model, which predicts the galaxy distribution in the dark halo as a function of the halo mass. The results of the HOD analysis are presented in Section 4.5.2.

I also show the model prediction of the dark matter ACF at $z \sim 2$ using the selection function of sgzKs with a limiting magnitude of $18.0 \leq K \leq 23.0$. This dark matter ACF was computed using the analytical formulae proposed by Hamana et al. (2002).

The lower panel of Figure 38 shows the bias parameters of the sgzK galaxies, which are defined as

$$b(\theta) = \sqrt{\frac{\omega_{\rm sgzK}(\theta)}{\omega_{\rm DM}(\theta)}}, \tag{100}$$

where $\omega_{\rm sgzK}(\theta)$ and $\omega_{\rm DM}(\theta)$ are the amplitudes of the ACFs of sgzKs and dark matter at the angular scale θ, respectively. The values of our bias parameter were in the range $3 < b < 5$ at the large scale, which is consistent with the results of Blanc et al. (2008). The bias parameters were larger at small angular scales than at large angular scales, indicating the excess from the power law.

By applying the least χ^2 fitting of the single power law (equation 3) with fixed $\gamma = 1.8$, I obtained the best-fit values of the amplitude of the ACF at 1°. I fitted the power law to the data for large

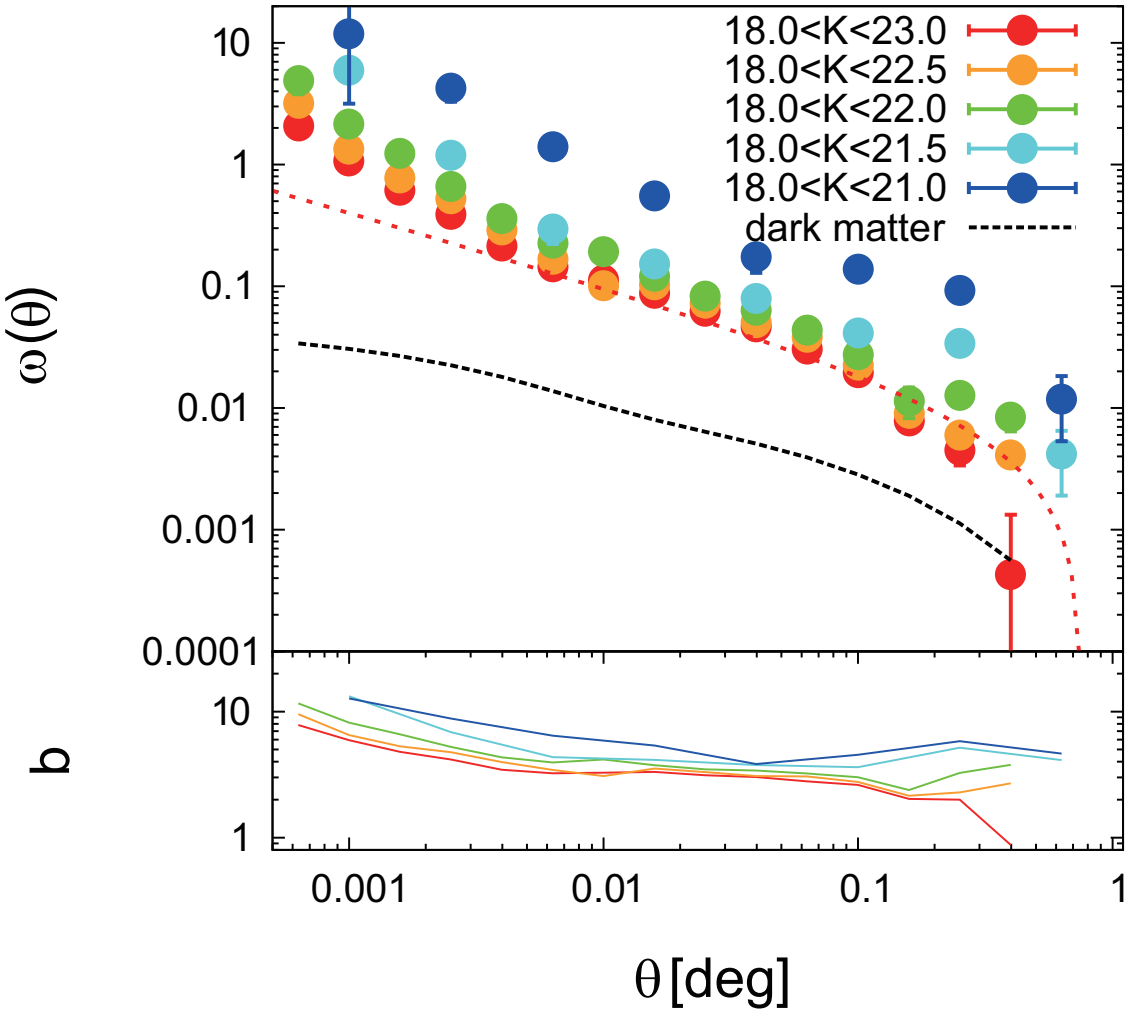

Figure 38.— Upper panel: ACFs of the cumulatively resampled sgzK galaxies. The limiting magnitudes of each sgzK subsample are $18.0 \leq K \leq 23.0$, $18.0 \leq K \leq 22.5$, $18.0 \leq K \leq 22.0$, $18.0 \leq K \leq 21.5$, and $18.0 \leq K \leq 21.0$ (red, orange, green, cyan, and blue circles). I confirmed the fact that more bright sgzK galaxies show the strongly clustering, reported by Hayashi et al. (2007). The red-dotted line represents the result of single power-law fit to the ACF of the total sample at large angular scales to show the excess from a power law at small angular scales. The dotted line represents the model prediction of the dark matter ACF computed using the nonlinear power spectrum. Lower panel: The bias parameters of sgzKs. Bias parameter is defined as $b(\theta) = \sqrt{\omega_{\rm sgzK}(\theta)/\omega_{\rm DM}(\theta)}$.

angular scale ($\theta \gtrsim 0.01^\circ$) where the 1-halo term is negligible. Table 7 lists the resulting amplitudes of the ACF; one can see that the clustering amplitude is dependent on the K-band luminosity.

Our measurement of the amplitudes at $18.0 \leq K \leq 23.0$ was $A_\omega(1^\circ) = (3.33 \pm 0.09) \times 10^{-3}$, which is larger than the value of $(1.79 \pm 0.17) \times 10^{-3}$ reported by Hartley et al. (2008), as well as $(1.27 \pm 0.23) \times 10^{-3}$ reported by McCracken et al. (2010). For $18.0 \leq K \leq 22.0$, we found $A_\omega(1^\circ) = (3.83 \pm 0.14) \times 10^{-3}$, which is larger than the value of $(3.14 \pm 1.12) \times 10^{-3}$ reported by Blanc et al. (2008) and $(2.12 \pm 0.65) \times 10^{-3}$ by McCracken et al. (2010). This deviation may be due to the fact that the survey area of the previous studies was too small to provide a high-quality signal at the large scale ($\theta > 0.1^\circ$). The ACFs reported in most previous studies were truncated at $\theta < 0.1^\circ$ or declined due to the effects of integral constraints. With our results, however, which are based on wide-field data, I calculated the ACF over a wide angular scale of $0.01^\circ < \theta < 0.5^\circ$, which enabled us to more accurately determine the amplitude, especially for the 2-halo term of the ACF. The amplitude of the ACF can be calculated from the large-scale galaxy clustering; the angular range was approximately $0.01^\circ \lesssim \theta \lesssim 0.1^\circ$, which makes it difficult to calculate the amplitude accurately if intermediate to large-scale clustering is not well determined.

In addition, Sato et al. (2014) pointed out that the clustering amplitude reported by McCracken et al. (2010) was weaker than those of the other studies. Sato et al. (2014) also presented the correlation functions, which is inconsistent with the result of McCracken et al. (2010), at the COSMOS field, though the origin of this discrepancy was unclear.

The results of our brightest three bins were in good agreement with those of Kong et al. (2006) although the error bars for their data were relatively large. Blanc et al. (2008) attributed the large amplitude reported by Kong et al. (2006) to the effects of cosmic variance; however, our results, where the survey field was more than 10 times larger than that of Kong et al. (2006), are less affected by cosmic variance. For this reason, I can calculate the amplitude of the ACF of K-bright sgzK galaxies more accurately than Kong et al. (2006) were able to.

4.4.2. Clustering in real space

The correlation amplitude of the ACF, A_ω, can be transformed into the three-dimensional correlation length by assuming a redshift distribution, where the correlation length corresponds to the three-dimensional clustering strength. Detailed procedure to evaluate the real-space correlation length is given in Section 2.1.3. The redshift distributions I used are described in Section 4.3.3.

The correlation lengths are listed in Table 7, and Figure 39 shows a comparison of our results with those of previous studies. I found that brighter sgzKs have larger correlation lengths, which indicates that brighter galaxies reside in more massive haloes and exhibit stronger clustering. It should be noted that because of the large sample size, our results are characterized by smaller error bars than those of previous studies. Our results also show excellent agreement with previous studies over all magnitudes, with the exception of Hartley et al. (2008). This discrepancy between the result of Hartley et al. (2008) and rest of the studies I compare might be caused by the inaccuracy of their sample selection (see also McCracken et al. 2010; Sato et al. 2014). As discussed in section 4.4.1, our ACFs have higher amplitudes than the previous studies, whereas our correlation lengths exhibit good agreement with the previous results. This might be because of the slight difference of the redshift distribution (i.e., our

sample has a redshift distribution shifted to lower-z than the previous studies). We have confirmed that the correlation length becomes smaller when the redshift distribution shifts to low-z.

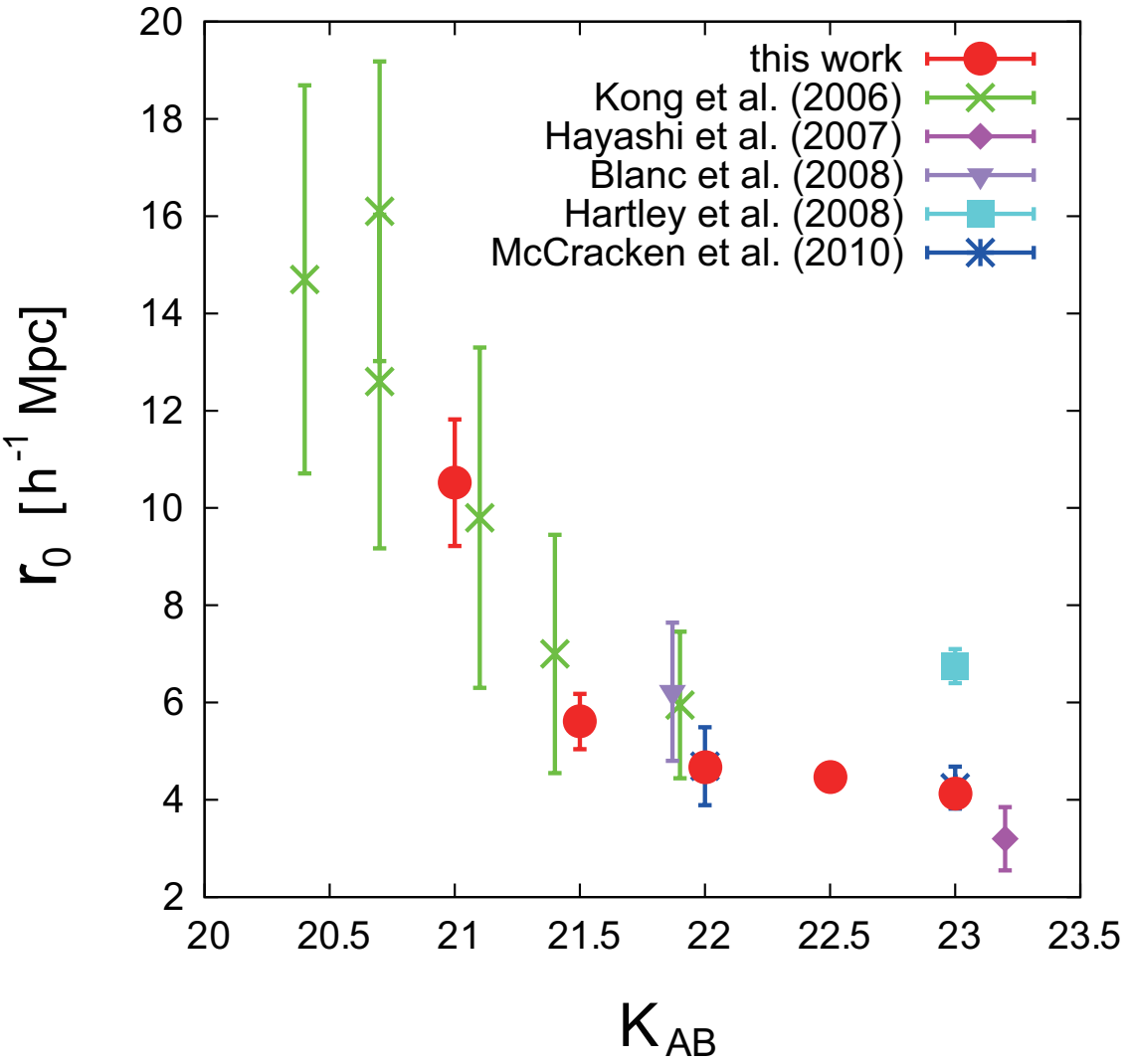

Figure 39.— Comparison of our correlation lengths of sgzK galaxies with previous studies. Our results (red filled circles) are consistent with previous studies over the whole magnitude range and measured with small error bars. All correlation lengths are in units of h^{-1}Mpc, where $h = 0.7$.

The correlation length for sgzK galaxies with $18.0 \leq K \leq 21.0$ were determined using the contamination-corrected ACF. Our correlation length of the brightest subsample is relatively large compared with the result of Kong et al. (2006), which was not corrected for contamination; however, as shown in Figure 36, the brightest subsample was more contaminated by the low/high-z galaxies and should be corrected accordingly. The difference in the correlation lengths can be explained by the lack of correction for contamination in the data reported by Kong et al. (2006).

4.4.3. Differential luminosity subsample

In this section, I show the results of clustering analysis on subsamples with different luminosities. Figure 40 shows the ACFs of sgzK galaxies with different luminosities. The bin size of the brightest sgzK subsample ($18.0 \leq K \leq 21.0$) was increased to ($\delta\log(\theta) = 0.4$) to increase the S/N ratio. The error of each data point was larger than that for the cumulative subsamples due to the smaller sample size; however, these ACFs clearly show dependence on the K-band magnitudes and have the apparent excesses at small angular scale, as was the case for the cumulative subsamples.

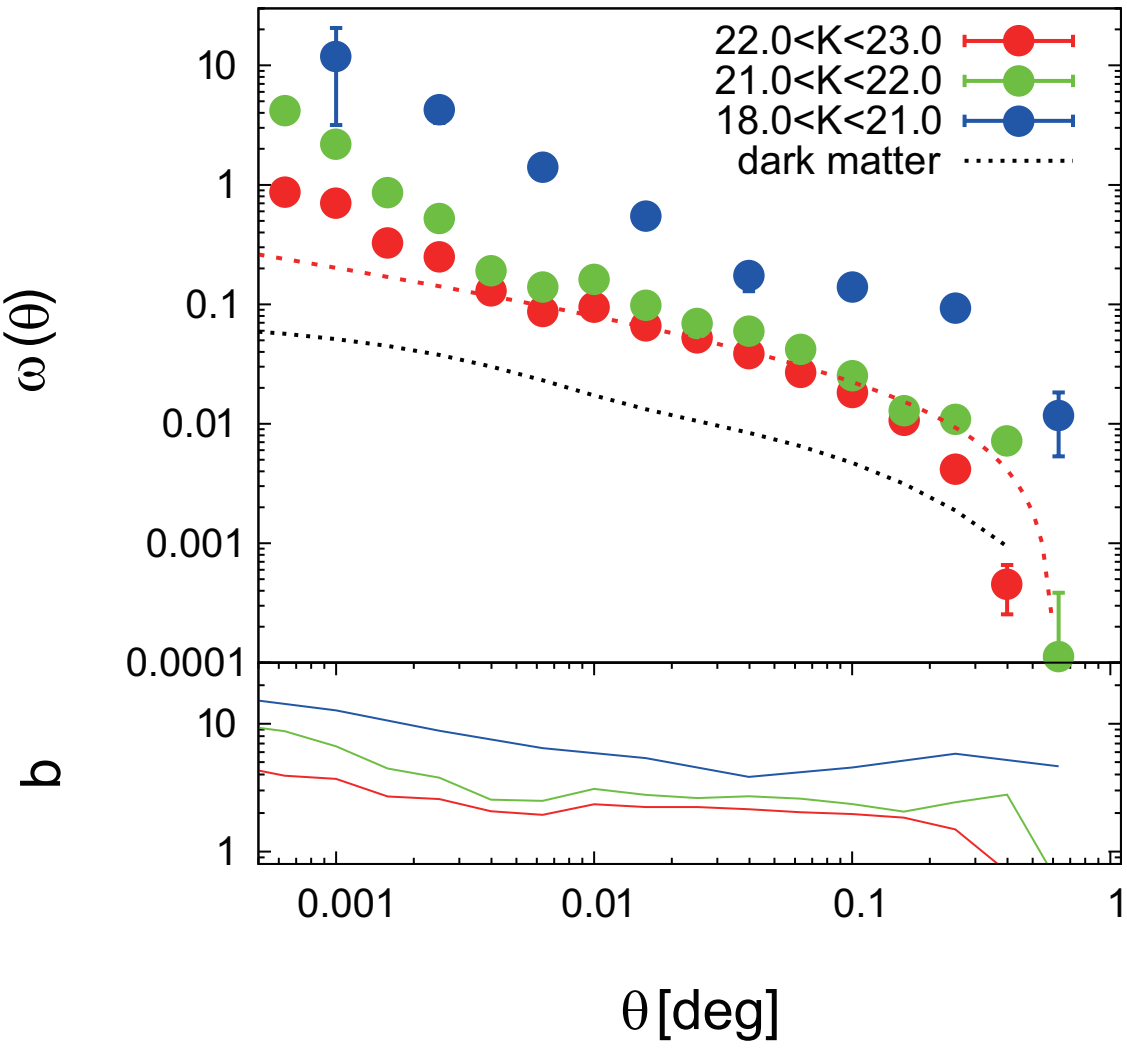

Figure 40.— Upper panel: ACFs of the differentially resampled sgzK galaxies. The limiting magnitudes of each sgzK subsample are $22.0 < K \leq 23.0$, $21.0 < K \leq 22.0$, and $18.0 \leq K \leq 21.0$ (red, green and blue circles). Lower panel: The bias parameters of sgzKs.

In the same manner as with the cumulative resampling, I fitted our sgzK subsamples using a power law to determine the amplitudes of the ACFs and the correlation lengths. These data are listed in Table 7.

Bielby et al. (2014) have showen that the correlation length of a star-forming galaxy depends

upon the stellar mass, whereas the correlation length of a passive galaxy does not (or is only weakly dependent of the stellar mass), in combination with the results of Coil et al. (2008), Bielby et al. (2010), and McCracken et al. (2010). Figure 41 shows the relationship between the correlation length and the stellar mass for both this study and previous studies (Bielby et al. 2014; Béthermin et al. 2014). We note that the stellar masses of our samples are estimated by their $(z-K)$ colors and K-band magnitudes (see Section 4.4.4), whereas Béthermin et al. (2014) and Bielby et al. (2014) estimated by the SED fitting. The dependence of our correlation lengths upon the stellar mass is consistent with previous studies.

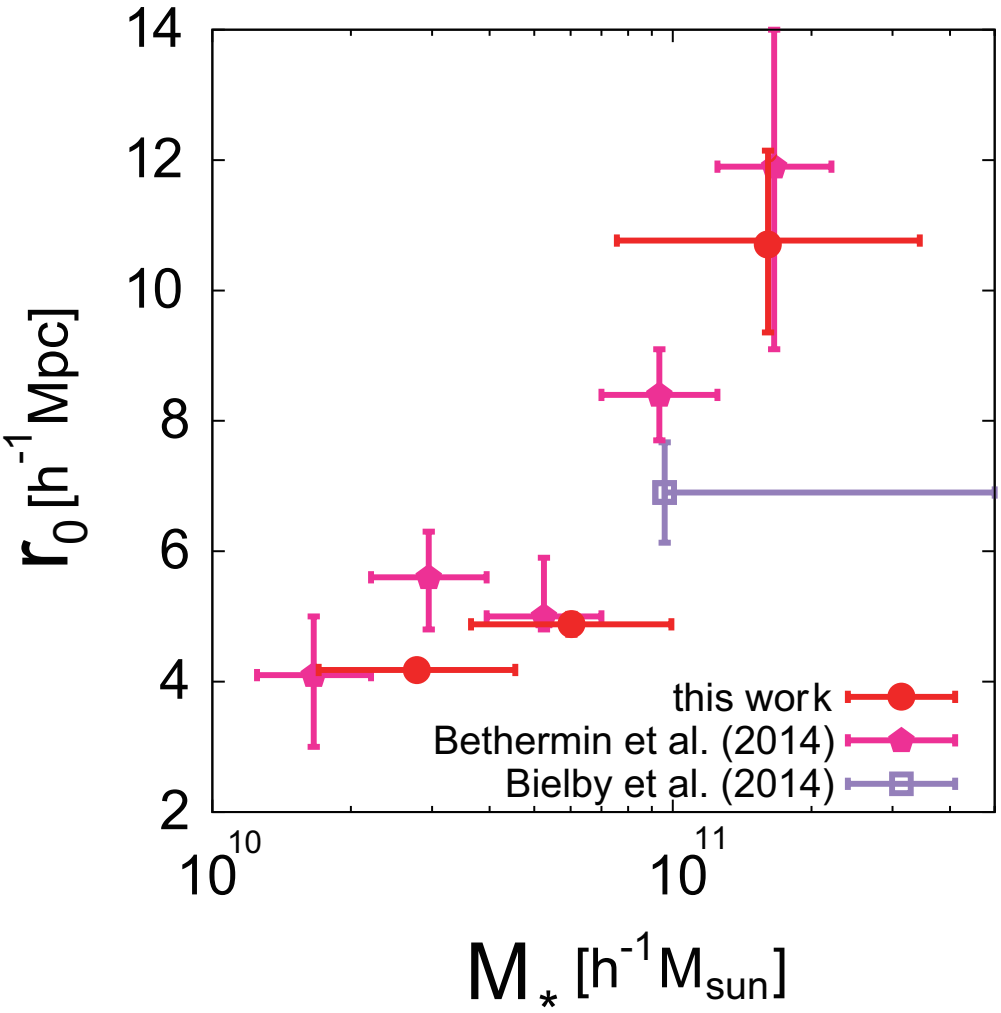

Figure 41.— Dependence of the correlation length upon the stellar mass of sgzK/sBzK galaxies. Our measurements (red circles) show good agreement with Béthermin et al. (2014) and Bielby et al. (2014) (magenta pentagons and purple squares, respectively). All correlation lengths are in units of h^{-1}Mpc, where $h = 0.7$.

4.4.4. Dark halo mass estimation by the large-scale clustering of sgzK galaxies

I calculated the dark halo mass residing in our sgzK samples using the results of our accurate clustering analysis. The procedure to estimate the dark halo mass from the amplitude of the ACF is presented in Section 2.1.4. Matter power spectrum used in this section is generated by the publicly available code CAMB (Lewis et al. 2000; Challinor & Lewis 2011), which is based on the code to calculate the linear cosmic microwave background (CMB) anisotropy spectra, CMBFAST (Seljak & Zaldarriaga 1996; Zaldarriaga & Seljak 2000).

Table 7 lists the dark halo masses. Our method to calculate the dark halo mass is based upon the wide survey area and can thus be expected to be accurate, as the results depend strongly on the clustering signals at large angular scales. Our measurements satisfy $M_{\rm h} \approx 3M_{\rm min}$ over almost all limiting magnitudes, which is consistent with the results of Hayashi et al. (2007), who reported that the mean dark halo mass is mainly determined by the less massive haloes, which are more numerous than the more massive ones. It should be noted that this measurement method of using large-scale clustering assumes that each dark halo must contain a single galaxy. This assumption may be erroneous, however, as massive haloes have been reported to contain multiple galaxies (e.g., galaxy groups/clusters in the local Universe), whereas less massive dark haloes may contain no galaxy at all. The further analysis using the HOD model may be beneficial to provide a more detailed description of the structure model.

Bielby et al. (2014) and Béthermin et al. (2014) determined the dark halo mass by this method, though they divided their sample by the stellar mass. The $(z-K)$ colour and K-band magnitude allowed us to estimate the galaxy stellar mass by using the galaxy model of Koyama et al. (2013, and the references therein). The conversion equation from the $(z-K)$ and K-band magnitudes of sgzK galaxies to stellar mass is given by

$$\log\left(M_\star/10^{11}M_\odot\right) = -0.4\times(K-21.90) + (0.086 - 1.28\times\exp\left(-0.921\times(z-K)\right)). \qquad (101)$$

It is note that this was derived by assuming the Salpeter IMF (Salpeter 1955) and that the formation redshift is $z_{\rm f}=5$. The scatter in the stellar mass of each galaxy was ≈ 0.3 dex. The stellar mass of each sgzK were estimated using this equation and determined the average stellar mass, together with the standard deviation of each subsample. The relations between $M_\star$ and $(z-K)$ color with fixing K-band magnitude are shown in Figure 42.

It is found that the minimum mass of the dark halo that resides in sgzK galaxies satisfying $18.0 \leq K \leq 23.0$ is $M_{\rm min} = (4.60\pm0.38)\times10^{11}h^{-1}M_\odot$ and the mean halo mass is $M_{\rm h} = (1.23\pm0.10)\times10^{12}h^{-1}M_\odot$, which are approximately three times more massive than Hayashi et al. (2007) reported for $K<23.2$. In addition, Blanc et al. (2008) reported a minimum dark halo mass down to $K \lesssim 22.0$ of $M_{\rm min}\approx 3\times10^{12}h^{-1}M_\odot$, which is also more massive than our estimation, i.e., $M_{\rm min} = (8.45^{+1.75}_{-1.50})\times10^{11}h^{-1}M_\odot$. These inconsistencies may be caused by the shallowness of their K-band photometry data and the small sample size. I also compared our results with those of Bielby et al. (2014) and Béthermin et al. (2014) by calculating the stellar mass of each subsample. Our results were in good agreement in terms of the mean halo mass reported by Bielby et al. (2014) and Béthermin et al. (2014) over the entire stellar mass range. However, Bielby et al. (2014) only estimated the dark halo mass of the massive galaxies, $\mathrm{M}_\star \sim 10^{11}h^{-1}M_\odot$, and the error bars of the dark halo mass given by Béthermin et al. (2014) were relatively large. Here, however, the dark halo masses of $z\sim2$ galaxies were determined for a wide range of K-band luminosities (and stellar masses), and we are able to

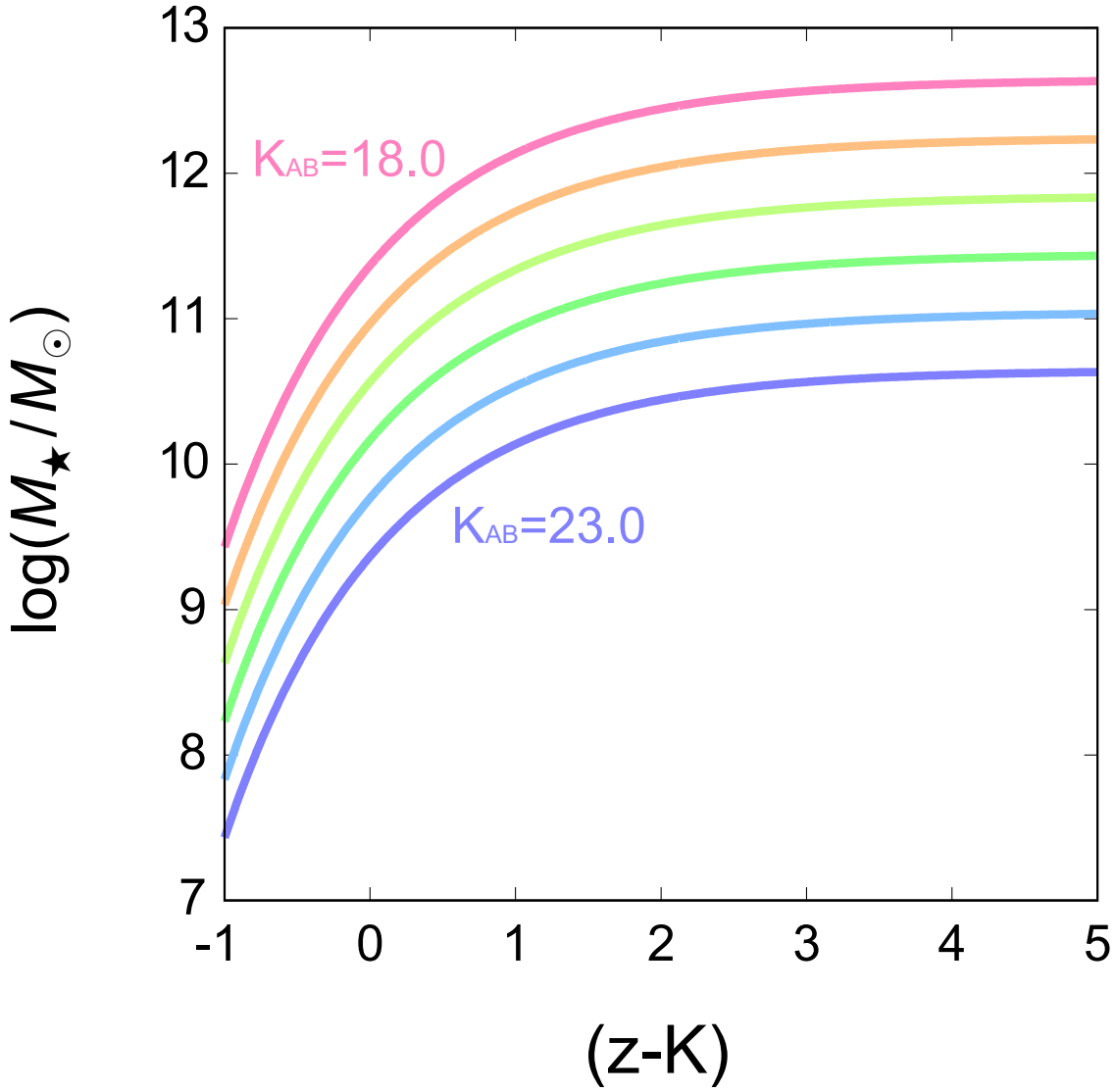

Figure 42.— Stellar masses of galaxies estimated by the equation (101) as a function of $(z-K)$ color. K-band magnitudes are fixed from $K = 23.0$ mag (blue) to $K = 18.0$ mag (red) with varying $\delta K = 1.0$ mag.

provide relatively small error bars.

4.5. HOD Analysis

4.5.1. Procedure of the HOD analysis

To evaluate the more detailed properties of host dark haloes, the HOD analysis was applied to our high-quality ACFs of sgzK galaxies. The halo occupation function of Zheng et al. (2005) was employed to represent the distribution of sgzK galaxies within the dark halo. Béthermin et al. (2014) also applied the HOD analysis with the same occupation model and succeeded to represent the observed ACF by the HOD model, albeit the S/N ratio of observed ACFs were not good.

To compute the ACF from the HOD formalism, it is assumed that the halo mass function proposed by Sheth & Tormen (1999), the large-scale halo bias of Tinker et al. (2005), and the profile of dark haloes as NFW profile (Navarro et al. 1997). The redshift distributions of each subsample were adopted the results of the COSMOS photo-z catalogue (Section 4.3.3).

As mentioned in Section 4.3.1, covariant matrices were evaluated by the jackknife resampling method by dividing our survey fiend into 64 subfields. The correlation coefficients, which is defined as $r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$, for each subsample are calculated and shown in Figure 43.

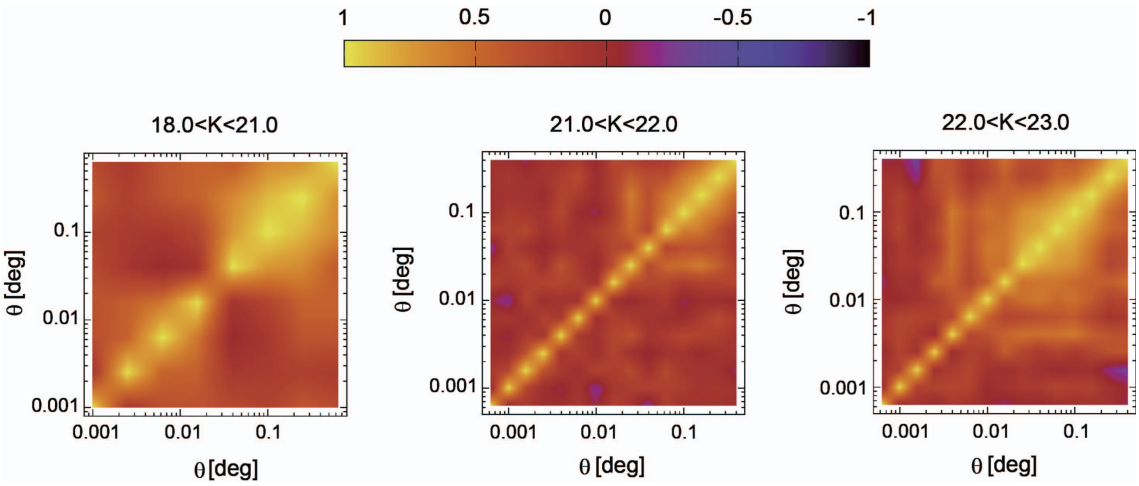

Figure 43.— Correlation coefficient of the sgzK galaxy sample. Left-hand, middle, and right-hand panels show the correlation coefficient results of $18.0 \leq K \leq 21.0$, $21.0 < K \leq 22.0$, and $22.0 < K \leq 23.0$ subsamples, respectively. These results were derived using the jackknife resampling method.

I fitted observed ACFs by varying the five free HOD parameters ($M_{\rm min}$, M_1, M_0, $\sigma_{\log M}$, and α), using the publicly available code, "CosmoPMC", to investigate the best-fit HOD parameters in parameter space (Wraith et al. 2009; Kilbinger et al. 2010). First, the best-fit HOD parameters of the overall sgzKs ($18.0 \leq K \leq 23.0$) were estimated by varying the five parameters over a wide range. Then, differentially resampled ACFs were used to investigate the dependence of the HOD parameters on K-band luminosity. The HOD analyses were carried out in two ways: (1) by varying the five HOD parameters and (2) by only varying $M_{\rm min}$, M_1, and M_0 with fixed $\sigma_{\log M} = 0.254$ and $\alpha = 1.19$

corresponding to the best-fit values for the total sample. In addition to the mean halo mass, $\langle M_{\rm h} \rangle$, the expectation number of sgzK galaxies within the dark halo, $\langle N_g \rangle$, is also calculated by following the description of Hamana et al. (2004) as

$$\langle N_g \rangle = \frac{\int_{M_{\rm min}}^{\infty} dM\, n(M) N(M)}{\int_{M_{\rm min}}^{\infty} dM\, n(M)}, \tag{102}$$

where the low-mass cutoff of the integration was set to $M_{\rm min}$ to exclude the contribution of low-mass dark haloes with no galaxies.

4.5.2. Results of the HOD analysis

Figure 44 shows the best-fit HOD models to the observed ACFs of the sgzK. It is clear that our HOD model can describe the observed ACFs of sgzKs. The best-fit HOD parameters with 1σ error are listed in Table 8. I show the results of the best-fit HOD parameters for the case varying all HOD parameters, as well as the case with fixed $\sigma_{\log M}$ and α. Fixed $\sigma_{\log M}$ and α are derived from the HOD analysis of the total sgzK (satisfying $18.0 \leq K \leq 23.0$) samples. Some studies report the best-fit values of $\sigma_{\log M}$ and α of various galaxy populations, although most of the HOD analysis studies used a fixed $\sigma_{\log M}$ and/or α.

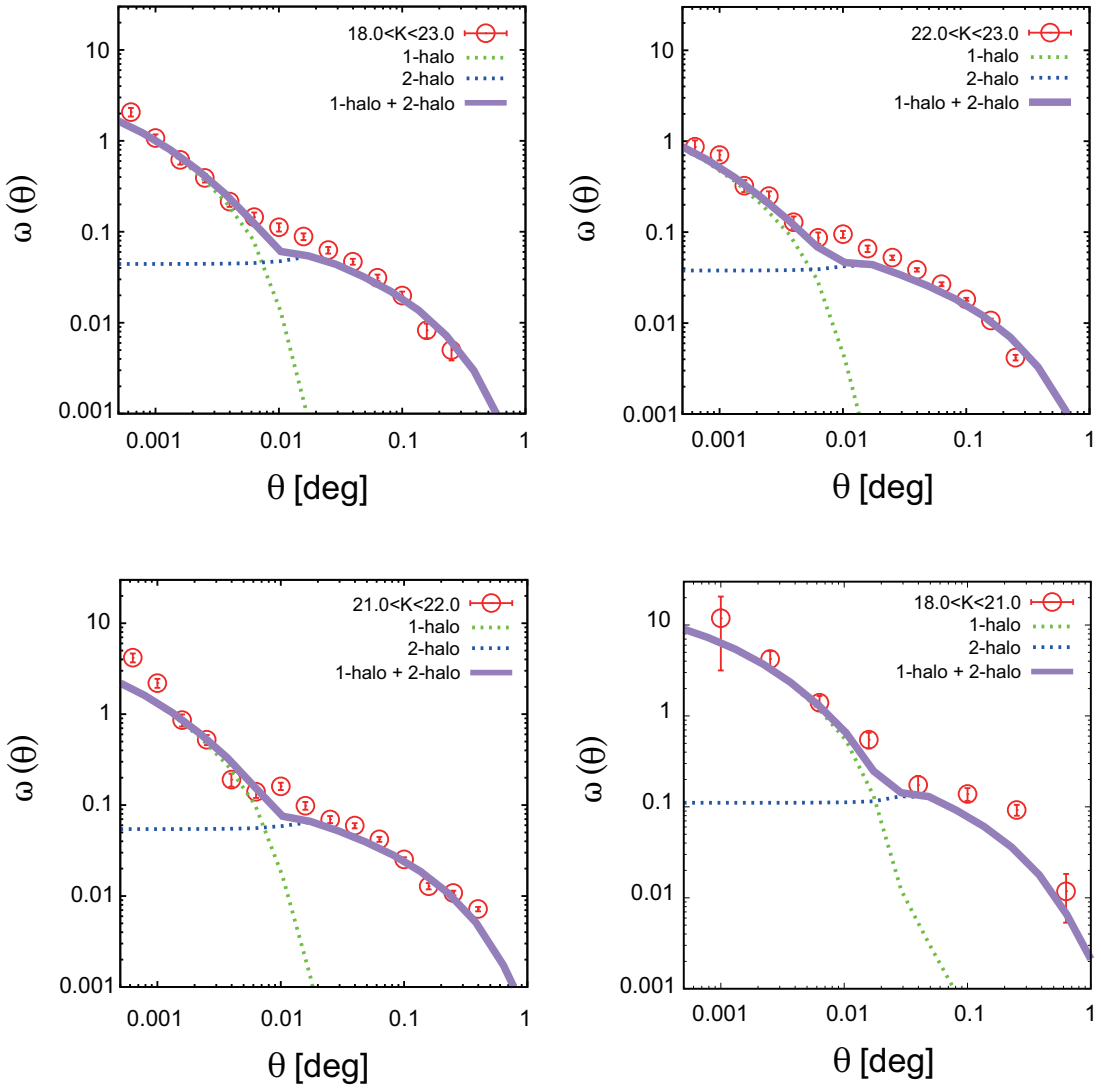

Figure 44.— A comparison between the observed ACF (red circles) and the best-fit HOD ACF. The dashed green (blue) curve represents the ACF of 1-halo (2-halo) term, and the solid purple curve shows a best-fit HOD ACF composed of the sum of the 1-halo term and the 2-halo term. We note that the full sample result (top-left) was derived by varying the five HOD parameters (see text for more details); the and results for the differential magnitude-cut subsamples shown in the other panels were derived by fixing $\sigma_{\log M}$ and α.

Table 8: The best-fit HOD parameters.

magnitude threshold	M_1 $[h^{-1}M_\odot]$	$M_{\rm min}$ $[h^{-1}M_\odot]$	M_0 [a] $[h^{-1}M_\odot]$	$\sigma_{\log M}$	α	$\langle M_{\rm h}\rangle$ $[h^{-1}M_\odot]$	$\langle N_g\rangle$	$\frac{\chi^2}{\rm dof}$
$22.0 < K \leq 23.0$	$(2.62^{+0.35}_{-0.25}) \times 10^{14}$	$(2.88^{+0.08}_{-0.08}) \times 10^{12}$	$(4.10^{+11.8}_{-2.89}) \times 10^{10}$	$0.404^{+0.095}_{-0.132}$	$0.961^{+0.125}_{-0.075}$	$(5.00^{+1.13}_{-0.85}) \times 10^{12}$	$0.788^{+0.080}_{-0.051}$	1.57
$21.0 < K \leq 22.0$	$(1.53^{+0.97}_{-0.31}) \times 10^{14}$	$(3.37^{+0.11}_{-0.10}) \times 10^{12}$	$(1.80^{+6.27}_{-1.40}) \times 10^{11}$	$0.292^{+0.243}_{-0.122}$	$0.983^{+0.141}_{-0.150}$	$(6.61^{+1.14}_{-2.33}) \times 10^{12}$	$0.855^{+0.125}_{-0.127}$	2.71
$18.0 \leq K \leq 21.0$	$(1.75^{+0.60}_{-1.34}) \times 10^{16}$	$(1.15^{+0.76}_{-0.43}) \times 10^{14}$	4.45×10^9	$0.477^{+0.165}_{-0.531}$	$0.993^{+0.262}_{-0.155}$	$(5.89^{+8.26}_{-3.13}) \times 10^{13}$	$0.638^{+0.084}_{-0.056}$	2.46
$22.0 < K \leq 23.0$	$(6.92^{+0.81}_{-0.65}) \times 10^{13}$	$(1.98^{+0.05}_{-0.06}) \times 10^{12}$	$(3.92^{+13.5}_{-2.93}) \times 10^{10}$	0.254 (fixed)	1.19 (fixed)	$(4.55^{+0.42}_{-0.60}) \times 10^{12}$	$0.877^{+0.071}_{-0.064}$	3.02
$21.0 < K \leq 22.0$	$(8.39^{+3.11}_{-1.76}) \times 10^{13}$	$(3.25^{+0.11}_{-0.10}) \times 10^{12}$	$(1.91^{+11.9}_{-1.65}) \times 10^{11}$	0.254 (fixed)	1.19 (fixed)	$(6.91^{+0.77}_{-1.05}) \times 10^{12}$	$0.883^{+0.095}_{-0.084}$	2.64
$18.0 \leq K \leq 21.0$	$(3.55^{+6.13}_{-2.23}) \times 10^{15}$	$(5.22^{+2.24}_{-1.42}) \times 10^{13}$	4.03×10^9	0.254 (fixed)	1.19 (fixed)	$(6.12^{+1.92}_{-1.37}) \times 10^{13}$	$0.759^{+0.028}_{-0.022}$	2.52

[a] The number of the brightest sgzK was too small to put constraints on the 1σ limit for M_0.

Most studies support our fixed $\sigma_{\log M}$ and α (e.g., Zehavi et al. 2005; Coupon et al. 2012). Martinez-Manso et al. (2015) carried out HOD analysis on *Spitzer*-selected galaxies at $z = 1.5$ by fixing $\sigma_{\log M} = 0.2$ and $\alpha = 1.0$, which are consistent with our fixed $\sigma_{\log M}$ and α. In the local Universe, Zehavi et al. (2011) also showed $\sigma_{\log M} \sim 0.2$ and $\alpha \sim 1$ for their faint and intermediate galaxy sample ($M_r^{max} < -20.5$), although the result of their full sample, including the brightest galaxies ($M_r^{max} < -22.0$), was $\sigma_{\log M} \sim 0.7$. From here, I adopt the results of the HOD analysis in which $\sigma_{\log M}$ and α are fixed.

At ~ 0.01 deg, there are systematic spikes in our ACFs, whose characteristics cannot be explained by the HOD model. The origin of this discrepancy is unclear. One possible explanation is that the halo model cannot describe the observed ACF at the scale of the virial radius. Several previous studies, especially at $z \sim 2$, also showed the small excess of the observed ACF from the best-fit HOD at this scale (e.g., Wake et al. 2011; Martinez-Manso et al. 2015; McCracken et al. 2015); however, further discussion is beyond the scope of this study.

The errors of the mean halo mass, $\langle M_{\rm h} \rangle$, and the expectation number of galaxies, $\langle N_g \rangle$, were estimated from the minimum/maximum values of $\langle M_{\rm h} \rangle$ and $\langle N_g \rangle$ by varying the free parameter set. Figure 45 shows χ^2 confidence contour maps of the HOD parameters derived from the least χ^2 method on the M_1–$M_{\rm min}$ parameter planes fixing M_0, $\sigma_{\log M}$, and α of those best-fit values. As shown by the data listed in Table 8, the mass quantities M_1, $M_{\rm min}$, M_0, and $\langle M_{\rm h} \rangle$ almost increased as the magnitude thresholds became brighter. This is evidence that brighter sgzK galaxies reside in more massive dark haloes. Figure 46 shows the halo occupation functions of each luminosity subsample as a function of the dark halo mass. I show both halo occupation functions in the case of varying five HOD parameters (dashed lines) and fixing $\sigma_{\log M}$ and α (solid lines).

The expectation number of galaxies within a dark halo, $\langle N_g \rangle$, was similar between each subsample. It follows that the ratio of the number of K-faint sgzKs in less-massive dark haloes to the number of K-bright sgzK galaxies in massive dark haloes was close to unity. Nevertheless, we should keep in mind that $\langle N_g \rangle$ is strongly affected by α, which represents the satellite galaxy formation efficiency in dark haloes, and which was fixed in this analysis. A possible dependence of the galaxy limiting magnitude on α should be investigated using a larger galaxy sample as part of a future extensive survey, such as the Hyper Suprime-Cam (HSC) survey.

4.5.3. Dark halo masses from the HOD model

In Section 4.4.4, dark halo masses were determined corresponding to sgzKs using large-scale galaxy clustering (Mo & White 1996), which assumes a one-to-one correspondence. HOD analysis enables us to derive a more accurate dark halo mass based on a more realistic HOD halo model.

Figure 47 shows a comparison of the dark halo masses calculated using the HOD analysis and the ACFs. The dark halo masses calculated using the HOD analysis were several times larger than those calculated using the ACFs. It is noted that the definitions of the mean halo masses with these two methods differ slightly. The mean halo mass from the HOD analysis is weighted by the number of galaxies in a dark halo, which tends to make $M_{\rm h}$ larger than the estimate from the ACF; that is because HOD formalism does not assume a one-to-one correspondence, as also mentioned in Geach et al. (2012).

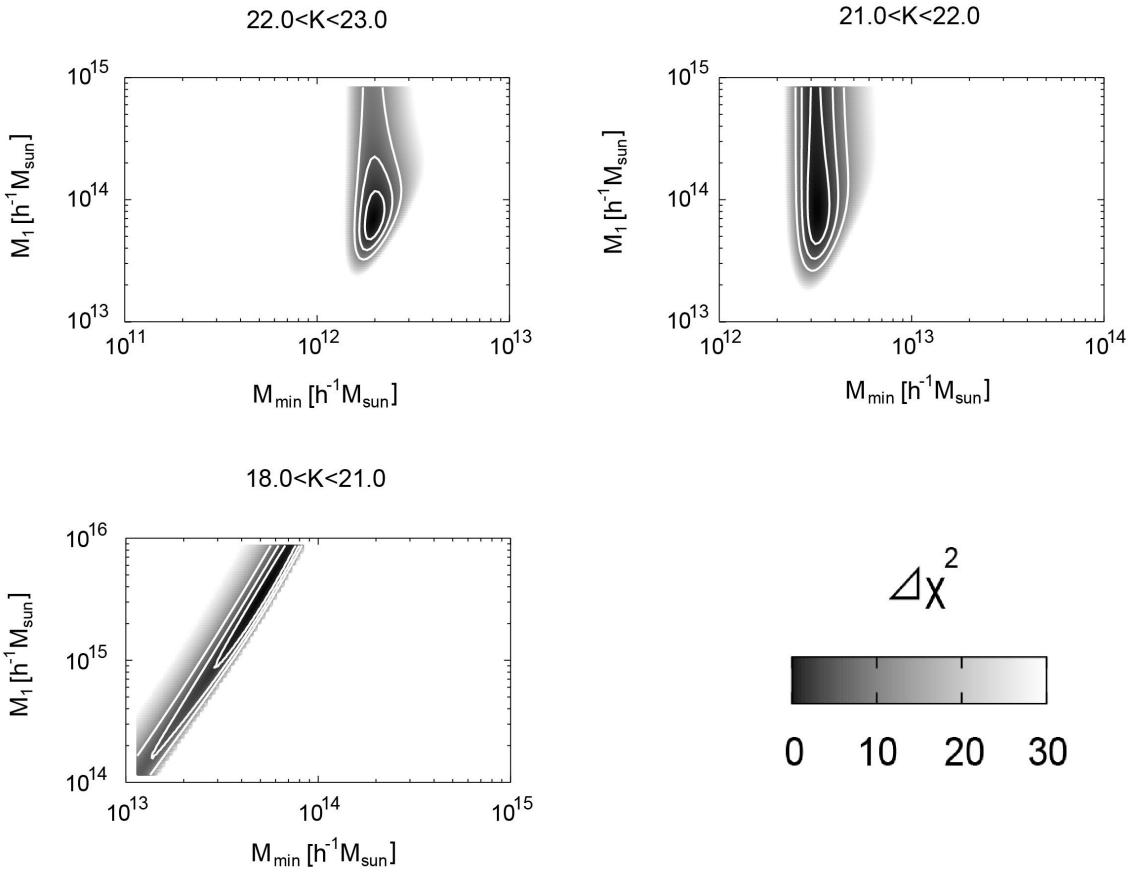

Figure 45.— Confidence contour maps of the HOD parameters derived from the least χ^2 method on the M_1 – $M_{\rm min}$ parameter planes, fixing M_0, $\sigma_{\log M}$, and α of the best-fit values. The grey-scale indicates the difference in χ^2 from the value of the best-fit parameters. The contour lines represent 1σ, 2σ, and 3σ confidence levels.

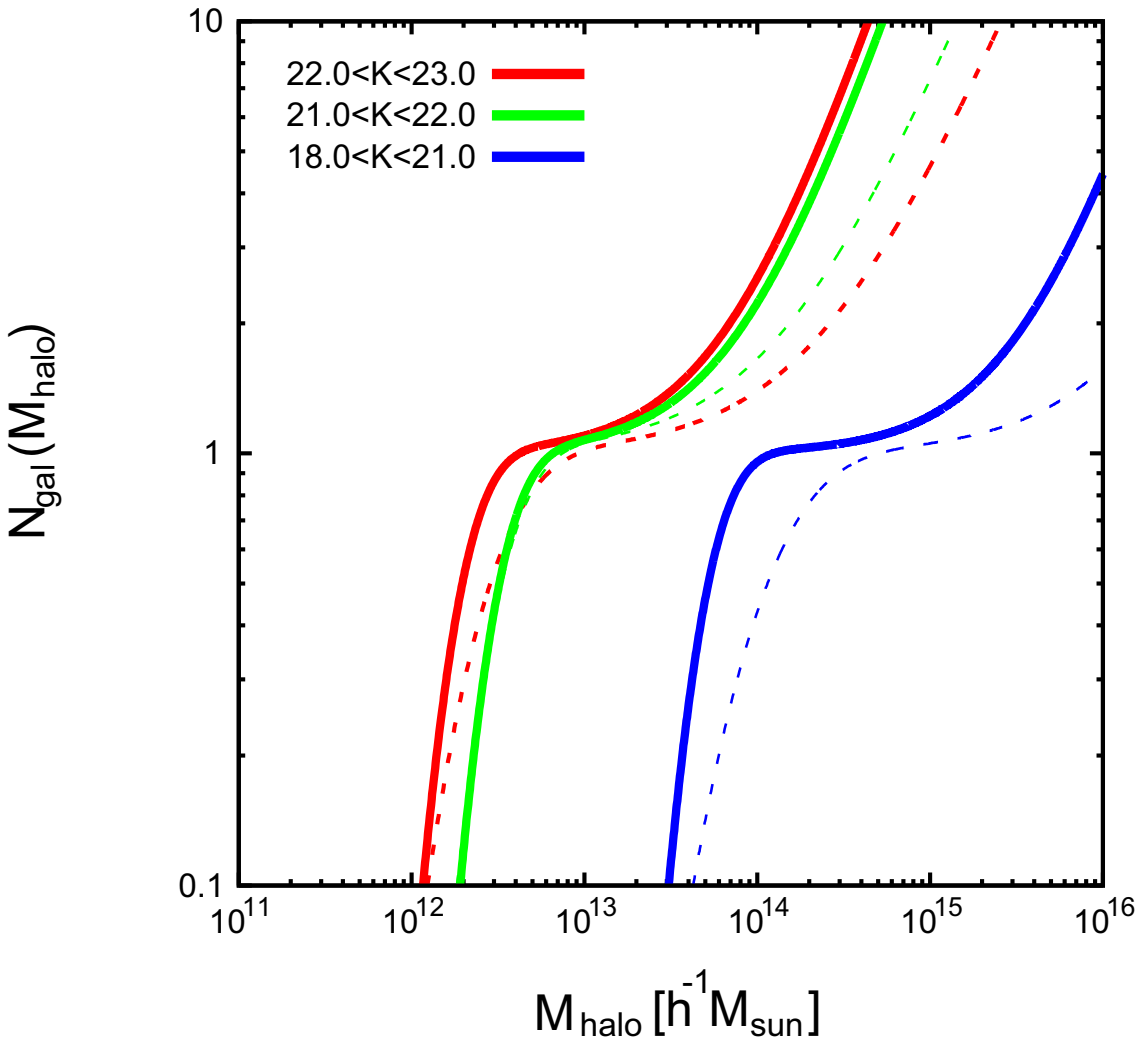

Figure 46.— The halo occupation functions of the best-fit HOD models of each sgzK subsample. I show both results of varying all free parameters (dashed lines) and fixing $\sigma_{\log M}$ and α (solid lines).

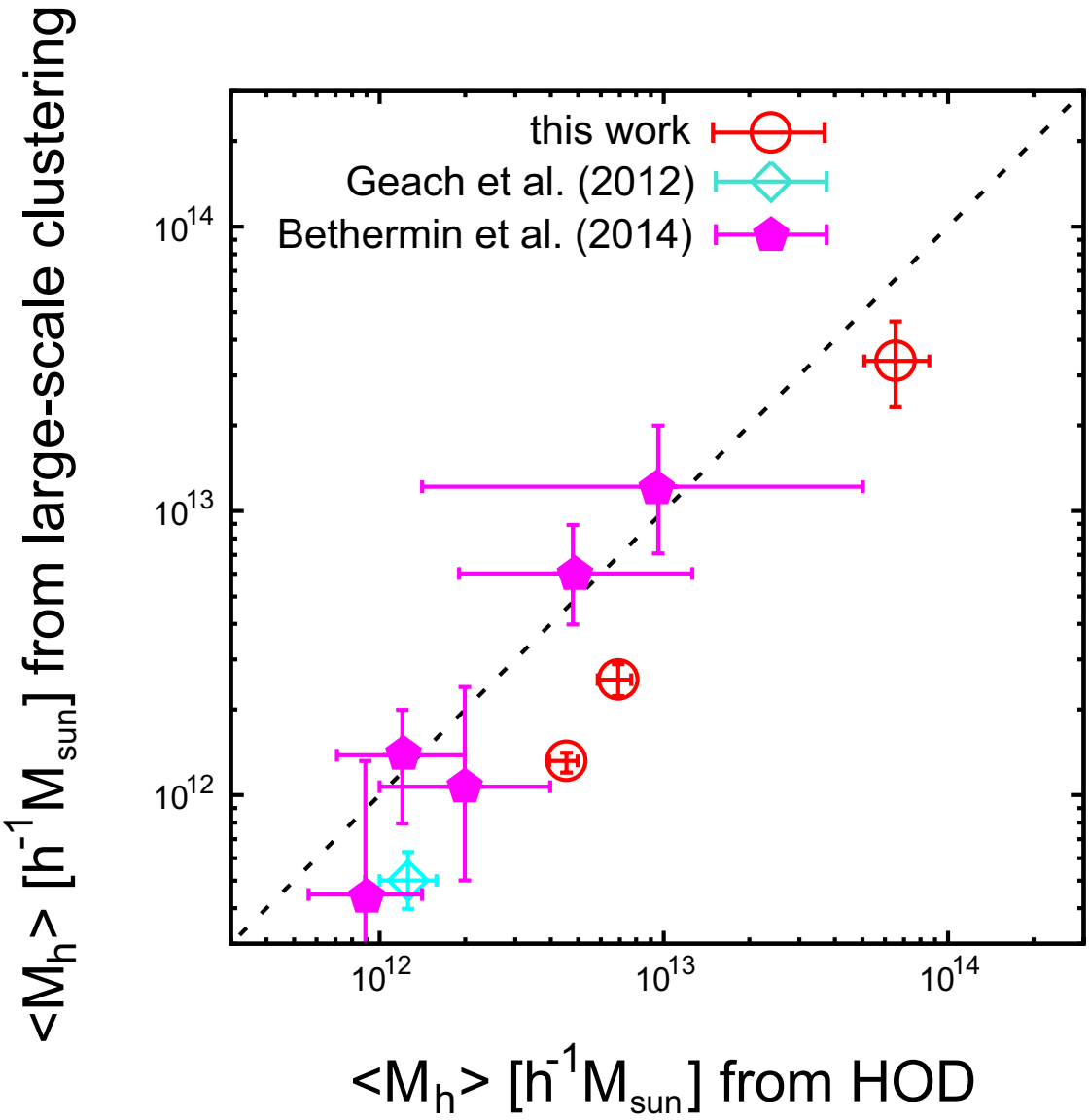

Figure 47.— A comparison of the dark halo masses calculated from the HOD model and large-scale galaxy clustering. The dotted-black line represents the one-to-one relationship between halo masses derived by both methods. Note that the results reported by Béthermin et al. (2014) were the dark halo masses of stellar-mass-limited subsamples.

I also compared our sgzK dark halo masses with those of the previous studies, as shown in Figure 47. Geach et al. (2012) investigated the mass of Hα emitters (HAEs) at $z = 2.23$, and Béthermin et al. (2014) estimated the mass of sBzKs at $z \sim 2$ in the COSMOS field. Our results are in good agreement with these data. Béthermin et al. (2014) reported little difference between data obtained using the two methods. However, this may be caused by large error bars in their HOD analysis.

4.6. Discussion

4.6.1. The luminosity dependence of the HOD parameters

As shown in Section 4.5.2, the HOD parameters exhibit a dependence on the K-band luminosity, which has been reported previously (e.g., Zehavi et al. 2011; Martinez-Manso et al. 2015). This is caused by the fact that more-luminous galaxies generally reside in more-massive dark haloes.

Figure 48 shows the relationship between the threshold absolute K-band magnitudes of the sgzK galaxies and the HOD parameters M_1 and $M_{\rm min}$. Zehavi et al. (2005) showed that the halo mass of galaxies with luminosity $L < L_*$ was weakly dependent on the luminosity of the galaxy, whereas the halo mass of galaxies with a luminosity $L > L_*$ increased sharply with increasing luminosity. This is because a large fraction of baryons within a massive dark halo tend to form satellite galaxies that are below the luminosity threshold, and then accrete to the central galaxy. Zehavi et al. (2011) also found that the clustering amplitude of galaxies with luminosity $L < L_*$ increased slowly, whereas the clustering amplitude increased sharply for $L \gtrsim L_*$, showing that M_1 and $M_{\rm min}$ had a similar dependence on the luminosity. Moreover, Zheng et al. (2009), using the same HOD model with varying $\sigma_{\log M}$ and α, reported a similar relation for luminous red galaxies at $z \sim 0.3$, showing that the dark halo mass of the low-mass halo is proportional to the luminosity of the central galaxy ($L \propto M_{\rm h}$), whereas high-mass dark haloes followed the relation $L \propto M_{\rm h}^{0.5}$.

As shown in Figure 48, our data show different slopes for the bright sgzK galaxies and the faint sgzK galaxies, although the slope of the power law was somewhat uncertain because we have only three data points, which makes it impossible to determine the turning magnitude; this trend is also seen in the local Universe (e.g., Zheng et al. 2009; Zehavi et al. 2011). Our results for high-mass sgzK galaxies exhibit a similar behavior to the high-mass low-z galaxies (Zehavi et al. 2005, 2011; Zheng et al. 2009), indicating that baryons are less likely to accrete into the central galaxies at $z \sim 2$, which is also seen for low-z galaxies. Zheng et al. (2009) inferred that baryons are likely to be consumed to form faint satellite galaxies rather than accrete into bright central galaxies. Our result of a massive dark halo indicates that the trend may be true, even in the $z \sim 2$ Universe.

$M_{\rm min}$, representing the threshold dark halo mass of the central galaxy formation, are larger at $z \sim 2$ than $z = 0$. This suggests that galaxies at higher-z can be formed only in very massive dark haloes, whereas less-massive dark haloes can form galaxies at lower-z, which is known as downsizing (e.g., Fontanot et al. 2009). On the other hand, M_1, which represents the threshold dark halo mass of the satellite galaxy formation, are also larger at $z \sim 2$ than $z = 0$. Our result implies that satellite galaxies are less likely to be formed in the high-z Universe than in the local Universe; also, the HOD mass parameters show the same luminosity dependence as that for the local Universe. It should be noted that the galaxy populations of this work and Zehavi et al. (2011) are not completely identical;

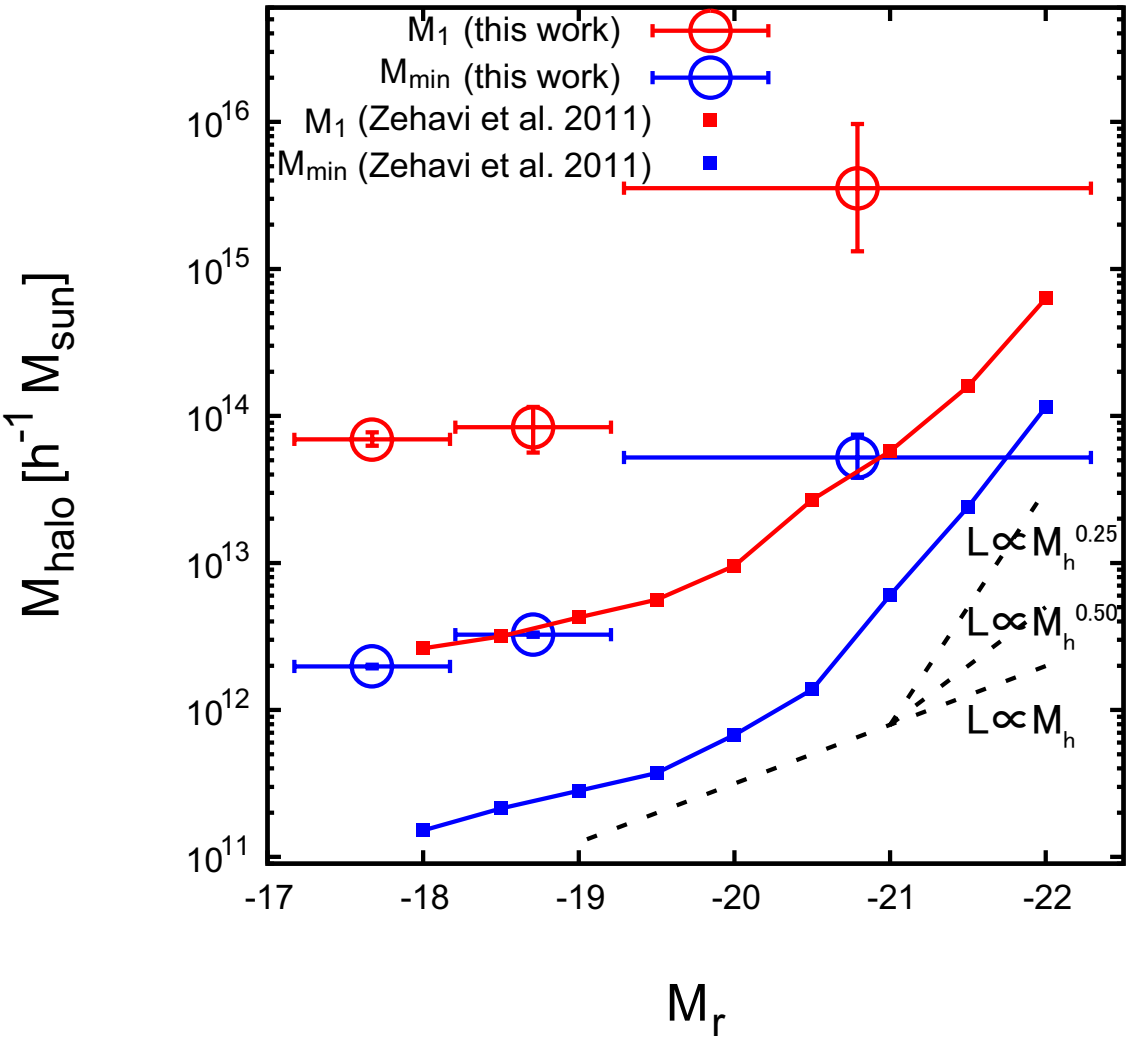

Figure 48.— The relationship between the threshold absolute magnitudes of sgzK galaxies and the HOD parameters for the characteristic halo mass whereby one dark halo possesses one satellite galaxy, M_1; and the threshold mass that can possess a galaxy within a dark halo, $M_{\rm min}$. The red points show M_1, the blue points show $M_{\rm min}$, and the square symbols with lines are the results from Zehavi et al. (2011), derived using SDSS galaxy samples. The dotted lines show various relationships between the luminosity of galaxies and the halo mass (i.e., $L \propto M_{\rm h}$, $M_{\rm h}^{0.5}$, and $M_{\rm h}^{0.25}$, from bottom to top).

Zehavi et al. (2011) used volume-limited SDSS galaxy samples ($z \leq 0.2$), whereas our sample is confined to the star-forming galaxy at $z \sim 2$.

4.6.2. The evolution of sgzKs by tracing the halo mass evolution

The mass assembly history can be traced from the mass evolution of dark haloes because the halo mass increases monotonically as a function of cosmic time due to merging. The halo mass evolution can be evaluated from the extended Press–Schechter model (EPS model; Press & Schechter 1974; Bond et al. 1991; Bower 1991), which describes the number density of objects in the Universe by modeling the halo merger, which is essential for the ΛCDM structure formation. A more detailed description of the EPS model can be found in Hamana et al. (2006, and the references therein).

First, I trace the dark halo mass evolution of Lyman break galaxies (LBGs) at $z \sim 4$ derived by Hamana et al. (2006) to investigate the relationship between sgzK galaxies and the higher-z galaxy population. Figure 49 shows a comparison of the expected dark halo masses between LBGs and sgzKs. It should be noted that the halo occupation models that are adopted by this work and Hamana et al. (2006) are different: the equation (56) was employed for the analysis of Hamana et al. (2006). All assumptions (e.g., cosmological parameters and halo mass function) of Hamana et al. (2006), except for σ_8 that was assumed to be $\sigma_8 = 0.9$ in Hamana et al. (2006), are the same as those we adopted. LBG samples of Hamana et al. (2006) were taken from the Subaru/XMM-Newton Deep Survey (Ouchi et al. 2005). The mass evolutions of LBGs from $z = 4$ to $z \sim 2$ are shown in Figure 49, where the hatched regions represent the 68% confidence intervals of the redshift evolution of the mean dark halo masses of LBGs. The dark halo mass of our bright sgzK galaxies ($21.0 < K \leq 22.0$) was equivalent to the evolved dark halo mass of bright LBGs ($i' < 26.0$) at $3 < z < 4$, whereas the dark halo mass of our faint sgzKs ($22.0 < K \leq 23.0$) was equivalent to the mass of the evolved dark halo mass of faint LBGs ($i' < 27.0$), although there was a slight difference between the mass of the bright sgzK galaxies and the faint sgzK galaxies. This result may be expected, because galaxies with a higher SFR at larger redshifts produce a considerably greater number of stars and evolve into galaxies with larger stellar mass at smaller redshifts. However, I should consider that these two galaxy populations were selected using different selection criteria. It should be noted that the dark halo masses of sgzK galaxies with magnitude $21.0 < K \leq 22.0$ and $22.0 < K \leq 23.0$ satisfy the evolved dark halo masses of both bright and faint LBGs within the range of 1σ confidence intervals, calculated using the EPS formalism. Thus, I cannot clearly determine which LBG is an ancestor of the bright/faint sgzK galaxies; however, the probability distribution functions (PDFs) of mass assembly of LBGs show that the descendants of bright (faint) LBGs at $z \sim 4$ are likely to be bright (faint) sgzK galaxies (see right panel of Figure 49). It is necessary to consider the evolutionary link of baryonic properties (e.g., stellar mass and galaxy age), in addition to the evolution of dark halo mass, to validate the evolutionary history of galaxies.

Figure 49 shows that there is a trend whereby the mean halo mass was smaller for LBGs and sgzK galaxies with fainter limiting magnitudes. This is caused by different physical mechanisms for each galaxy population. For LBGs, the detection band was the i'-band, which corresponds to ~ 1500Å in the rest frame. LBGs are known to be actively star-forming, and the UV radiation is mainly emitted by young, massive stars; i.e., more-luminous LBGs at UV wavelengths have higher SFRs, which are known to be strongly correlated with $M_\star$ ("main-sequence" of the SFR–stellar mass plane). Hathi et al. (2013) reported that LBGs up to $z \sim 5$ follow the same correlation between the SFR and the stellar

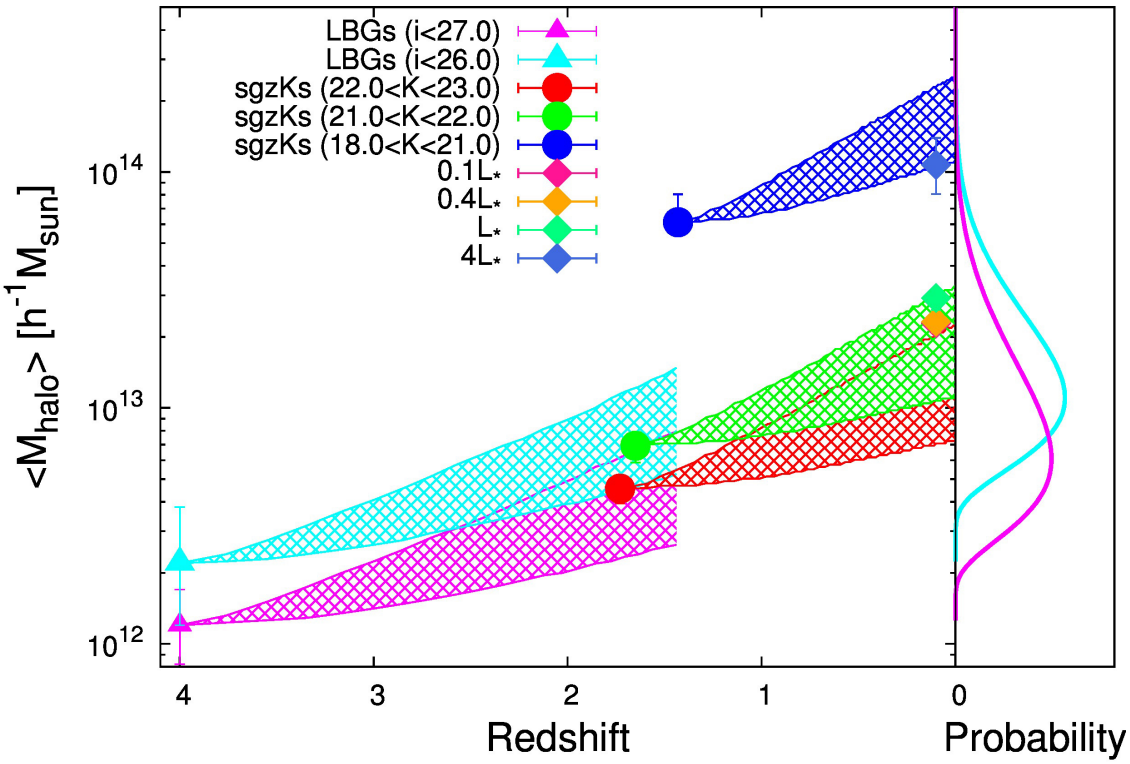

Figure 49.— Left panel: The mass evolutions of each galaxy population. The masses of the LBGs satisfying $i' < 27.0$ (magenta) and $i' < 26.0$ (cyan) at $z = 4$ are those reported by Hamana et al. (2006); the dark halo mass evolutions were calculated using the EPS formalism with a 68% confidence interval as shown by the hatched regions. Our sgzK galaxies calculated using the HOD formalism are also shown, with the dark halo mass evolutions from $z \sim 2$ to $z \sim 0$. In addition, I show the mean dark halo masses of SDSS galaxies reported by Zehavi et al. (2011), which were estimated using their reported HOD parameters. Right panel: The probability distribution functions (PDFs) of mass assembly of the LBGs. The PDF shows the probability of mass evolution of a dark halo that possesses faint LBG (magenta) and bright LBG (cyan) at $z = 4$ to $z \sim 2$.

mass, which is consistent with the results of the cosmological hydrodynamical simulation reported by Finlator et al. (2006). Therefore, the star-formation activity of LBGs was more active with a massive dark halo at $z \sim 4$, and the stellar mass of the galaxies were much higher than UV-faint LBGs that reside in less-massive dark haloes.

It is noted that sgzK galaxies were selected according to the K-band luminosity, which is contributed to mainly by old, less-massive stars, and corresponds to the stellar mass directly. A strong correlation has been reported between the K-band luminosity of sBzKs and the stellar mass (e.g., Lin et al. 2012; Béthermin et al. 2014). However, the correlation between the dark halo mass of sBzKs/sgzKs and the SFRs remains unclear. Béthermin et al. (2014) argued that the dark halo mass of sBzKs increased up to $200 <$ SFR $M_{\odot}\mathrm{yr}^{-1}$ and exhibited a plateau for higher SFRs; however, such flattening of the SFR–halo mass relation is significant only at 1.7σ; and this should be confirmed via a more precise analysis in the future.

The expected number of LBGs in a dark halo reported by Hamana et al. (2004) was $\langle N_g \rangle \sim 0.4$; however, it is found that the expected number of sgzK to be $\langle N_g \rangle \sim 0.8$. This discrepancy may result from differences in the selection criteria. The number of satellite galaxies selected by stellar mass at lower-z was more than that selected by SFR at higher-z. In addition, it was not possible to identify the ancestors of the brightest sgzK galaxies at $z \sim 4$. This is because the counterparts of the brightest sgzK galaxies at $z \sim 4$ are thought to be brighter LBGs than those LBGs satisfying $i' < 26.0$; therefore, those bright LBGs are rare objects in the distant Universe. Further extensive surveys in the future, i.e., HSC survey, are expected to reveal a sufficient number of such bright LBG candidates to determine the dark halo mass more accurately.

I traced the expected evolutionary history of our sgzK galaxies to the local Universe. Figure 49 shows the mass evolutions of sgzK galaxies from $z \sim 2$ to $z \sim 0$ (shown by the hatched regions). The mean dark halo masses of local galaxies with various luminosities were calculated by Zehavi et al. (2011), and these data were applied to the HOD analysis of local galaxies ($z \leq 0.2$) from the SDSS DR7 to investigate the dependence of the clustering strength on the physical properties of galaxies (e.g., luminosity and color).

From the dark halo mass evolution estimated by the EPS formalism, the faintest sgzK galaxies ($22.0 < K \leq 23.0$) at $z \sim 2$ ($\langle M_{\mathrm{h}} \rangle \sim 4 \times 10^{12} h^{-1} M_{\odot}$) evolve to $\langle M_{\mathrm{h}} \rangle \sim (7 - 20) \times 10^{12} h^{-1} M_{\odot}$, which corresponds to the dark halo mass of the local late-type galaxies, such as the Milky-Way, or typical early-type galaxies. Intermediate luminosity sgzK galaxies ($21.0 < K \leq 22.0$) at $z \sim 2$ with a mass of $\langle M_{\mathrm{h}} \rangle \sim 7 \times 10^{12} h^{-1} M_{\odot}$ evolve to $M_{\mathrm{h}} \gtrsim 10^{13} h^{-1} M_{\odot}$, which corresponds to the dark halo mass of massive elliptical galaxies or galaxy groups in the local Universe. The dark halo mass of the brightest sgzK galaxies ($18.0 \leq K \leq 21.0$) at $z \sim 2$, with a mass of $\langle M_{\mathrm{h}} \rangle \sim 1 \times 10^{14} h^{-1} M_{\odot}$, evolve to $\langle M_{\mathrm{h}} \rangle \sim (2 - 4) \times 10^{14} h^{-1} M_{\odot}$, which corresponds to the dark halo mass of the most massive systems in the local Universe, such as rich clusters of galaxies.

Béthermin et al. (2014) investigated galaxy evolution by tracing the dark halo mass from $z \sim 2$ to $z \sim 0$ using the halo growth model reported by Fakhouri et al. (2010). They inferred that sBzK galaxies with a mass of $\langle M_{\mathrm{h}} \rangle \sim 3 \times 10^{11} M_{\odot}$ grew to the most massive field galaxies with a mass of $\langle M_{\mathrm{h}} \rangle \sim 3 \times 10^{12} M_{\odot}$, a halo with a mass of $\langle M_{\mathrm{h}} \rangle \sim 3 \times 10^{12} M_{\odot}$ grew to form galaxy groups, and the sBzK/pBzK galaxies with a mass of $\langle M_{\mathrm{h}} \rangle \sim 3 \times 10^{13} M_{\odot}$ evolved into galaxy clusters in the local Universe. This differs slightly from our analysis; I find that the halo mass evolution was approximately

~ 0.5 dex from $z \sim 2$ to $z \sim 0$, whereas the halo mass evolution reported by Béthermin et al. (2014) was approximately one order of magnitude higher. This difference in the halo growth rate led to differences in the corresponding progenitors of the $z \sim 2$ galaxies.

In summary, our results suggest that faint LBGs at $z \sim 4$ could evolve into faint sgzK galaxies ($22.0 < K \leq 23.0$) at $z \sim 2$ and to the Milky-Way-like galaxies or elliptical galaxies in the local Universe, whereas bright LBGs at $z \sim 4$ could evolve into intermediate luminosity sgzK galaxies ($21.0 < K \leq 22.0$) and into the most-massive elliptical galaxies or central galaxies of the galaxy groups in the local Universe, and the most-luminous sgzK galaxies ($18.0 \leq K \leq 21.0$) could evolve into the most massive systems in the local Universe; i.e., the central galaxies of the galaxy clusters. However, these evolutionary scenarios were determined by only considering the evolution of the dark halo mass. It may not be straightforward to relate different galaxy populations at different redshifts, as they may follow different star-formation histories, which are traced by other physical properties such as SFR, stellar mass, and age. In the following section, I discuss the evolution of galaxy populations, which provides another point of view from which to analyze our galaxy evolution model.

4.6.3. Galaxy evolution by tracing the number of satellite galaxies

One advantage of the HOD formalism is that it enables us to describe the number of galaxies in a dark halo for a given dark halo mass. Because of this characteristic of the HOD formalism, I was able to trace the number evolution of satellite galaxies within a dark halo, which is not possible if I assume one-to-one galaxy–halo correspondence. In this section, I discuss the number evolution of satellite galaxies in a dark halo using our sgzKs, assuming that galaxies evolve so as to follow the dark matter halo evolution predicted using the EPS formalism (see Section 4.6.1).

First, I consider the evolution of dark halo with M_1'; this defines the mass of the dark halo occupied by only one galaxy at $z \sim 2$, satisfying $N(M_1') = 1$ for each luminosity subsample, to investigate the number evolution of the satellite galaxy within the dark halo. I follow the evolution of the dark halo with mass of M_1' based on the EPS formalism from $z \sim 2$ to $z = 0$; the dark halo mass at $z = 0$ was determined as the mass in which the PDF of the dark halo mass evolution reaches a peak; its error is in the range of 1σ confidence interval of the PDF. The number of the satellite galaxies in the dark halo at $z = 0$ was calculated by the halo occupation function (equation 62) at an evolved dark halo mass, with the best-fit HOD parameters of Zehavi et al. (2011). It is noted that I took into account both errors of the evolved dark halo mass and HOD parameters of Zehavi et al. (2011) to evaluate the number of satellite galaxies.

Figure 50 shows the expected number of satellite galaxies at $z = 0$ for a dark halo, which corresponds to a central galaxy with no satellite galaxies at $z \sim 2$. In this figure, the horizontal axis represents the absolute rest r-band magnitudes, which corresponds to the K-band limiting magnitude of the sgzK galaxies at $z \sim 2$, and the vertical axis is the number of satellite galaxies that are expected to be formed in the dark haloes at $z = 0$. The color difference indicates a different absolute magnitude of satellite galaxies at $z = 0$. More-luminous galaxies tend to have more satellite galaxies at $z = 0$, and a large fraction of these satellite galaxies are faint. In dark haloes that contain the faintest central sgzK galaxies with $M_{\rm r} \gtrsim -19$ at $z \sim 2$, approximately three satellite galaxies with luminosity equal to that of central sgzK galaxies at $z \sim 2$ are expected to form. However, brighter satellite galaxies are unlikely to form in the dark haloes of these faint sgzK galaxies. In other words, dark haloes with

the faintest sgzK evolve into dark haloes that contain $\lesssim 10$ faint galaxies and no bright galaxies at $z = 0$. In the dark haloes that contained the brightest central sgzK galaxies $M_{\rm r} \gtrsim -22$ at $z \sim 2$, approximately 100 satellite galaxies are expected to be formed by $z = 0$. This is consistent with the results described in Section 4.6.2; i.e., local galaxy groups/clusters that contain 100 or more member galaxies. Few of the brightest satellite galaxies had similar luminosity to those of sgzK galaxies; i.e., sgzK galaxies that reside in haloes at $z \sim 2$ would become a central galaxy of a galaxy cluster in the local Universe, especially the brightest cluster galaxies (BCGs).

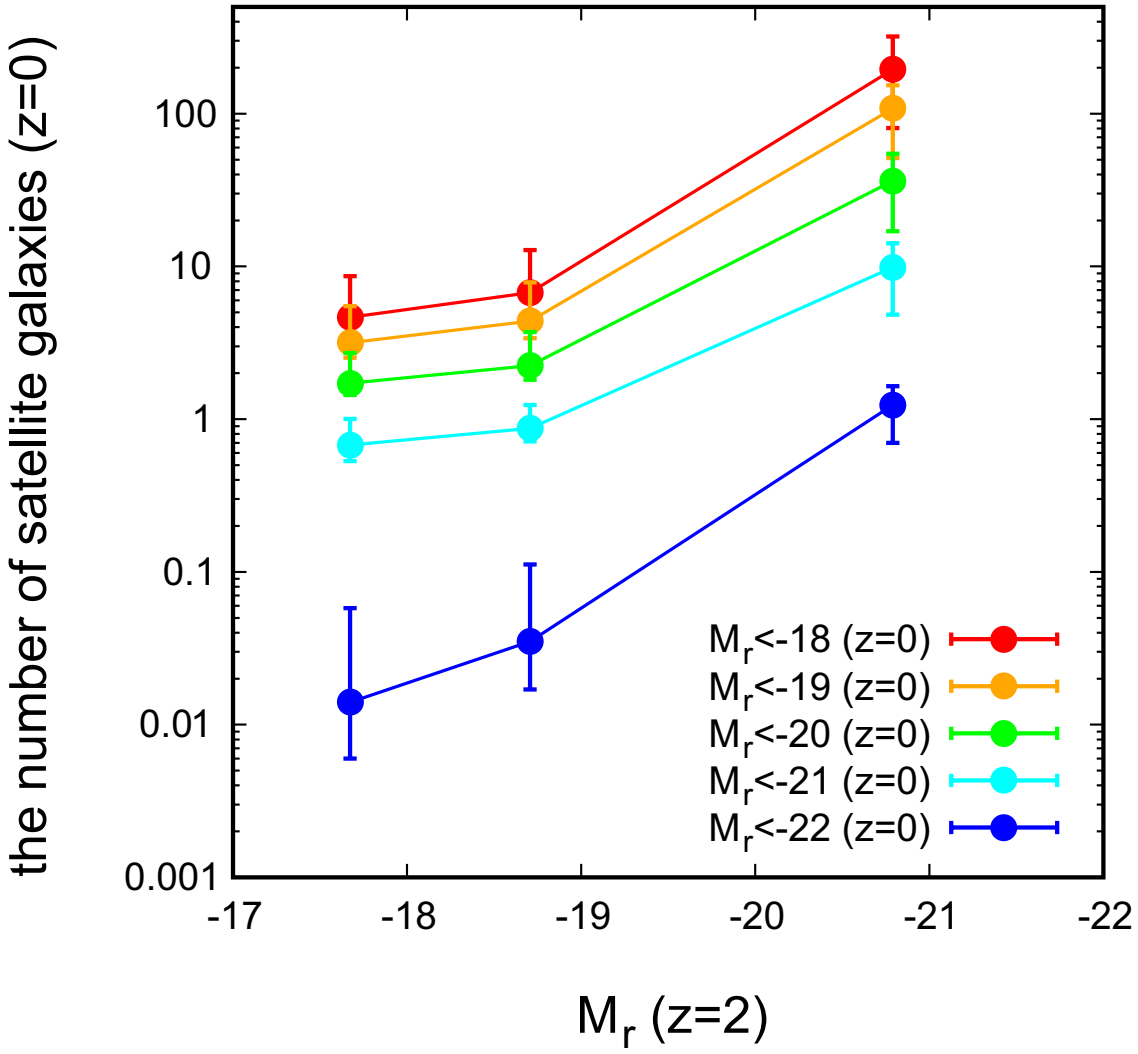

Figure 50.— The evolution of the number of satellite galaxies in the dark halo. The horizontal axis is the magnitude threshold of the absolute magnitude of sgzK galaxies at $z \sim 2$, and the vertical axis is the number of satellite galaxies in the dark halo at $z \sim 0$, which contains only one sgzK satisfying each limiting magnitude. The different colors of the data points correspond to differences in the magnitude threshold of the satellite galaxies.

Based upon the discussion of the number evolution of satellite galaxies, the faintest sgzK galaxies appear to evolve into the Milky-Way-like galaxies, whereas the brightest sgzK galaxies evolve into the central galaxies of galaxy clusters in the local Universe, especially BCGs. These results are consistent

with the implications of the discussion of galaxy evolution from considerations of the evolution of the dark halo mass.

4.6.4. Stellar-to-halo mass ratio

Figure 51 shows a comparison between the results of our SHMRs, the theoretical prediction of Behroozi et al. (2013b), and the observed SHMRs reported by Foucaud et al. (2010), Geach et al. (2012) and McCracken et al. (2015). The stellar mass was determined from the K-band magnitudes and $(z - K)$ colors of each of the sgzK galaxies (Section 4.4.4), and $\langle M_{\rm h} \rangle$ was used as the dark halo masses. It should be noted that the SHMR, in principle, should be calculated by the ratio between the dark halo mass and all stellar components within the dark halo (e.g., Kravtsov et al. 2014); however, Behroozi et al. (2013b) calculated the SHMRs via an abundance matching method. Our data, as well as those of Geach et al. (2012) and McCracken et al. (2015), were calculated using an HOD analysis. Our data points are almost in good agreement with the theoretical model, as well as the other observations. The SHMRs of our sgzKs exhibit almost identical values at $z \sim 0$ within the 1σ confidence level, consistent with the theoretical predictions for SHMRs with $M_{\rm h} \gtrsim 10^{12} h^{-1} M_{\odot}$, and do not significantly change between $z \sim 0$ and $z \sim 2$.

Our results are almost consistent with the prediction of the model reported by Behroozi et al. (2013b); however, there was a slight difference in the trend, especially for the SHMR of the least massive haloes; the SHMR calculated from our data was smaller than the prediction of the model at $z \sim 0$ as well as at $z \sim 2$. This discrepancy may be due to differences in the galaxy populations. Our data were based upon sgzK galaxies, whereas Behroozi et al. (2013b) did not take into account galaxy types when assigning galaxies to dark haloes via abundance matching. Tinker et al. (2013) also investigated this relationship over the redshift range of $0.22 < z < 1.00$, distinguishing the differences in galaxy populations as star-forming galaxy samples, passive galaxy samples, and all galaxy samples. They inferred that, especially for galaxies with a halo mass of $M_{\rm h} \gtrsim 10^{12}\ M_{\odot}$, the stellar mass of a central galaxy depends on the population of the galaxy samples with a fixed dark halo mass due to differences in the growth rates of galaxies. With our samples, however, we consider only star-forming galaxies, whereby the central galaxy has a smaller stellar mass than the other two galaxy populations (Tinker et al. 2013). This may lead to a smaller SHMR at $M_{\rm h} \sim 3 \times 10^{12} h^{-1} M_{\odot}$ than that reported by Behroozi et al. (2013b). The smaller SHMR of our data compared with that of McCracken et al. (2015) that sampled all galaxies, including passive galaxies, may also be attributed to differences in the galaxy populations. The HAEs (Geach et al. 2012), which are in the early phase of star-forming activity, also show significantly lower SHMR than other galaxies.

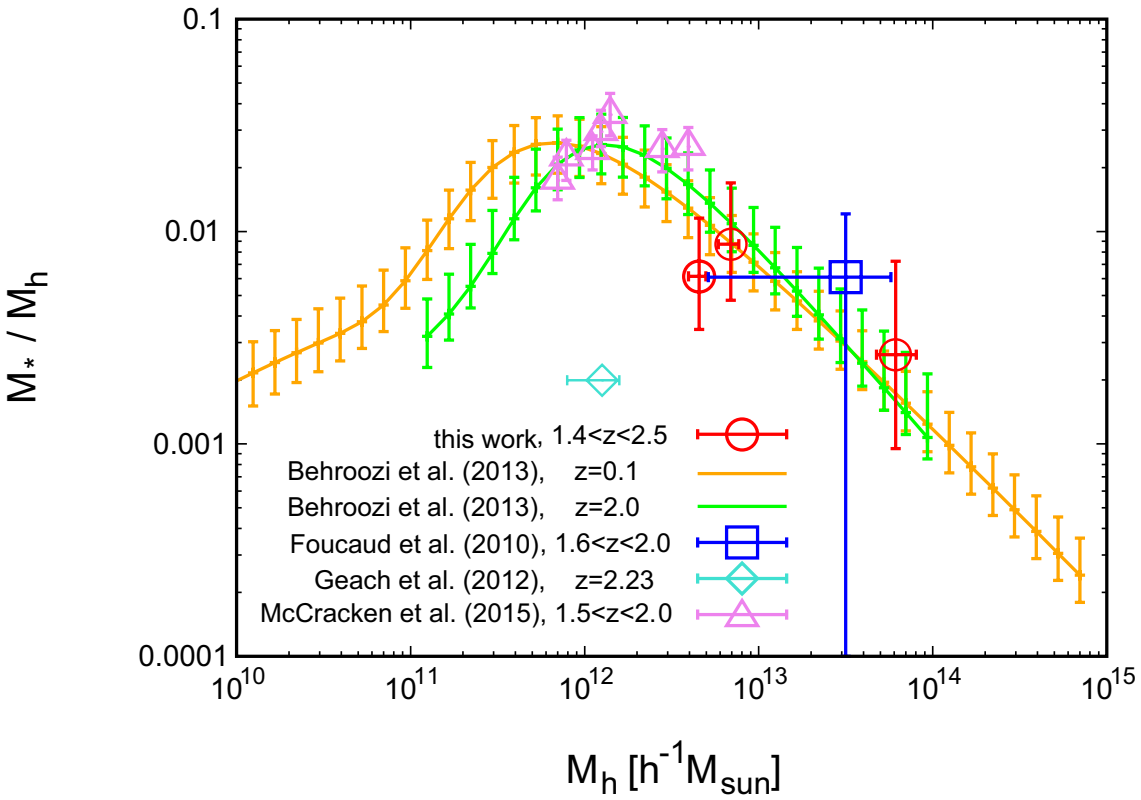

Figure 51.— A comparison of SHMRs between our results for sgzKs (red) and the theoretical predictions reported by Behroozi et al. (2013b). The SHMR reported by Foucaud et al. (2010) (blue), Geach et al. (2012) (cyan), and McCracken et al. (2015) (purple) are also shown. The calculation method used to obtain our dark halo masses was HOD analysis, and the stellar masses were found using the $(z - K)$ color and K-band luminosities of the sgzK galaxies. Geach et al. (2012) and McCracken et al. (2015) also calculated dark halo masses using the HOD analysis, whereas the method used by Foucaud et al. (2010) was the large-scale galaxy clustering. The stellar mass of all of those studies was calculated by the SED fitting. The solid curves with error bars show results of the theoretical model reported by Behroozi et al. (2013b).

5. GALAXY–HALO CONNECTION IN HIGH-REDSHIFT UNIVERSE

5.1. Overview

5.1.1. Clustering analysis in high-redshift Universe

Clustering properties of $z > 3$ galaxies have been relatively investigated because, unlike the case of $z \sim 2$, one can select high-z galaxies only by the optical photometric images. The selection method of high-z galaxies is based upon the idea known as "dropout technique" (e.g., Steidel et al. 1996). The continuum flux from galaxies shorter than the 912 Å Lyman limit will be strongly absorbed by the neutral hydrogen gases, i.e., the intergalactic medium (IGM), and strong break can be observed at the wavelength corresponds to Lyman limit in the rest frame, $912 \times (1+z)$ Å in their SEDs; thus, galaxies "dropout" from the photometric images whose effective wavelength is shorter than redshifted Lyman limit. By capturing the break and the UV continuum by three photometric bands, high-redshift galaxies can be selected. The selected high-z galaxies using the dropout technique are termed the "dropout galaxies" of the "Lyman break galaxies (LBGs)".

The biggest advantage of this method is that the Lyman break moves into the optical wavelength for $z > 3$ galaxies: the $(u-g)$ color can detect the Lyman break and becomes red, while the $(g-r)$ color capture the UV continuum and shows blue for $z \sim 3$ galaxies. Therefore, $z \sim 3$ galaxies can be easily detected and selected only by three optical photometric images with u-, g-, and r-bands using ground-base telescope. In this case, the selected dropout galaxies are termed "u-dropout galaxies" because the u-band photometric image corresponds to the dropout band. By shifting the dropout band to longer wavelength, the dropout technique can be applied up to $z \sim 5$ or 6, although deep observation of long-wavelength end of the optical band is difficult. We note that one can detect higher-z galaxies through the dropout technique by using the NIR images.

The selection method for high-redshift galaxies has been established; however, deep and wide-field observation is required to capture the Lyman break clearly and collect sufficient number of samples enough to carry out the clustering analysis. The ACFs of the dropout galaxies at $z = 3 \sim 5$ have been well studied from the early 2000s (e.g., Giavalisco & Dickinson 2001; Wechsler et al. 2001; Ouchi et al. 2001; Porciani & Giavalisco 2002; Arnouts et al. 2002; Adelberger et al. 2003; Foucaud et al. 2003). Those studies mainly computed the correlation lengths, the galaxy bias parameters, and the roughly dark halo masses by the power-law fitting of the ACFs, and discussed the evolutionary link between the dropout galaxies and the log-z galaxy populations; however, their discussions were largely limited by the relatively large uncertainties due to the bad S/N ratios of their ACFs. Moreover, the angular scales of the ACFs of those studies were quite small because of the small survey field, and the amplitudes of those ACFs were suffered from large errors, which also lead large uncertainties for the deduced quantities such as the correlation length and the dark halo mass.

Ouchi et al. (2005) collected a large number of dropout galaxies at $z = 4$ and derived ACFs by dividing i-band magnitudes with high S/N ratios. They succeeded to obtain apparently excess at the small-angular scale, the 1-halo term, and carried out the HOD analysis using the formalism of Hamana et al. (2004) to derive the detailed physical properties of the dropout galaxies. The mean dark halo masses were calculated $\langle M_{\rm h} \rangle \sim 10^{11-12} M_\odot$ and the expectation number of the dropout galaxies at

$z = 4$ decreased with increasing the rest-frame UV luminosities.

Hamana et al. (2006) also carried out the HOD analysis of the dropout galaxies at $z = 4$, which were selected by Ouchi et al. (2005), and investigated the evolutionary history between the dropout galaxies and the old passively evolving galaxies (OPEGs) at $z \sim 1$ by tracing the evolution of the dark halo mass. The dark halo mass evolution was predicted based upon the extended Press–Schechter formalism, which was developed by Bond et al. (1991), Bower (1991), and Lacey & Cole (1993). They revealed that the OPEGs are likely to be the progenitors of the dropout galaxies

Hildebrandt et al. (2009) collected a large number of dropout galaxies at $z = 3$, 4, and 5 using the deep and wide-field optical photometric images obtained by the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Gwyn 2012). The ACFs divided by the rest-frame UV luminosities with high S/N ratios were obtained in the CFHTLS Deep fields, and the UV-luminosity dependence and the redshift evolution of the clustering properties were investigated. They also applied the HOD analysis for their dropout samples using the formalism of Hamana et al. (2004) and estimated the mean halo masses and the expectation number of galaxies of dropout galaxies. The mean halo masses were calculated $\langle M_{\rm h} \rangle \sim 10^{12.0-12.7} h^{-1} M_{\odot}$, which were almost consistent with the result of Ouchi et al. (2005) with the same luminosity threshold, and showed little redshift evolution at fixed rest-frame UV luminosity.

Recently, Harikane et al. (2016) carried out the HOD analysis on the dropout galaxies at $z = 4-7$ using the data of publicly available data of Hubble Space Telescope (*HST*) and the early-released data of the Subaru/Hyper Suprime-Cam SSP survey (Miyazaki et al. 2013). They obtained the ACFs of the high-redshift dropout galaxies by dividing the galaxy stellar mass, which is evaluated by the main-sequence of the star-forming galaxies (Shibuya et al. 2015) and performed the HOD analysis assuming the halo occupation function with the duty cycles of the dropout galaxies (see Section 2.3.2). The extremely deep photometric data of the HST enabled them to select very less massive galaxies ($M_{\star} \sim 10^7 M_{\odot}$ even at $z \sim 4$) and investigated the evolution of the SHMRs at the low-mass end. They identified the evolution of the SHMRs at $z = 4-7$ and showed that the observed SHMRs at the low-mass end were consistent with the model prediction of Behroozi et al. (2013b). The decreasing trend of their low-mass SHMRs were also consistent with the theoretical results of the semi-analytical models.

5.1.2. Motivation of this study

One of the most important problematic points of this epoch is the accuracies of the ACFs due to the difficulties of collecting the large number of samples. The results of Harikane et al. (2016) are suffered from the large uncertainties because of the bad S/N ratios of their ACFs. To investigate the accurate physical properties of dropout galaxies at high-redshift Universe by carrying out the precise HOD analyses, derivation of the high-quality ACFs is an essential issue.

Some previous studies carried out the HOD analysis for high-z galaxies and derived the physical properties of the dropout galaxies. Hildebrandt et al. (2009) applied the HOD analysis for the ACFs with high S/N ratio derived from the dropout galaxies up to $z \sim 5$; however, they used the formalism developed by Hamana et al. (2004), which is based upon the very-early occupation model. Moreover, Hildebrandt et al. (2009) used the early-released data of the CFHTLS, which is completed the survey

project in 2012 with updating the depth and the quality of the photometric data. Recently, deep and wide-field NIR observation overapping a part of the CFHTLS Deep fields has been carried out by the CFHT Wide-field InfraRed Camera (WIRCam; Puget et al. 2004), which is mounted at the prime focus of the CFHT. Using the NIR photometric data as well as the optical photometric data, one can derive the well accurate photometric redshift and the galaxy stellar mass by the SED fitting technique.

In addition, SHMRs of the dropout galaxies help us to constrain the formation and evolution at high-redshift Universe. Harikane et al. (2016) derived the SHMRs at $z = 4-7$ and traced their redshift evolution, albeit in the limitation for the less massive galaxies ($M_\star \lesssim 10^9 M_\odot$). The star-formation history and the galaxy-formation efficiency can be discussed if the SHMRs of the dropout galaxies will be computed over the wide dark halo mass range and the peak of the SHMRs, known as the pivot halo masses (cf., Leauthaud et al. 2012) will be captured.

To address these issues, I carry out the precise clustering analyses of dropout galaxies at $z \sim 3$, 4, and 5 using the final data release of the CFHTLS. Deep- and wide-field images of the CFHTLS Deep Survey enable us to obtain sufficient accurate ACFs to apply the HOD analysis. By combining the optical data from CFHTLS with the NIR data from the WIRCam Deep Survey (WIRDS; Bielby et al. 2012), the SED fitting technique can be applied to evaluate the accurate photometric redshift of the dropout galaxies as well as their stellar mass. The stellar masses are also calculated by assuming the main-sequence of the star-forming galaxies, and compute the SHMRs by both stellar-mass estimation methods to check the consistency of the SHMRs. Furthermore, the mean halo masses and the satellite fractions are also calculated with high accuracies, and discuss the evolutionary history of galaxies at high-redshift Universe.

Throughout this section, I employ the flat lambda cosmology ($\Omega_{\rm m} = 0.27, \Omega_\Lambda = 0.73$, and $\Omega_{\rm b} = 0.045$), the Hubble constant as $h = {\rm H}_0/100{\rm kms}^{-1}{\rm Mpc}^{-1} = 0.7$, the matter fluctuation amplitude at $8h^{-1}$Mpc as $\sigma_8 = 0.8$, and spectral index of the primordial power spectrum as $n_{\rm s} = 1$, which are based upon the results of the *WMAP* seven-year data (Komatsu et al. 2011) to compare with the results of previous works. With these assumed cosmological parameters, the ages of the Universe at $z \sim 3$, 4, and 5 are ~ 1.56, 1.12, and 0.852 h^{-1}Gyr, and physical angular scales of 1 arcsec correspond to 5.56, 5.03, and 4.55 h^{-1}kpc, respectively. All of the results and discussions are presented in Ishikawa et al. (2017, in prep.).

5.2. Data and Sample Selection

5.2.1. Optical data

To select the enough number of high-z dropout galaxies to obtain high-quality ACFs, I utilized the publicly available data of the final data release of the CFHTLS. CFHTLS is an optical multi-wavelength (u^*-, g-, r-, i-, and z-band) survey that is carried out by the CFHT MegaCam (Boulade et al. 2003) start from mid-2003. The transmission functions of the optical bands of the MegaCam are shown in Figure 52.

The CFHTLS performed two distinct surveys: the CFHTLS Deep Survey and the CFHTLS Wide Survey. The CFHTLS Wide Survey covers ~ 155 deg^2 which is sufficient field size to evaluate the

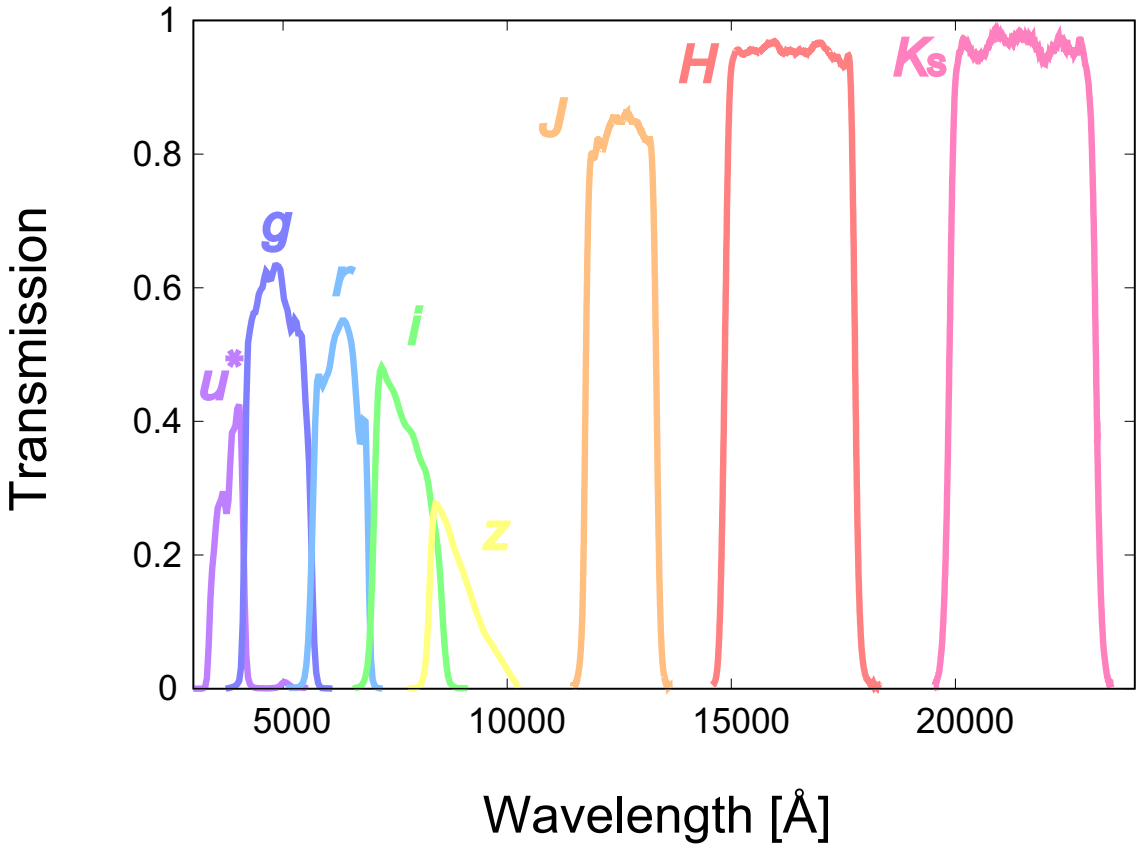

Figure 52.— The transmission functions of the optical photometric bands of the MegaCam (u^*-, g-, r-, i-, and z-band) and the NIR photometric bands of the WIRCam (J-, H-, and K_s-band).

high-quality ACFs; however, the depth of the photometric images are shallow to select the enough number of high-redshift galaxies to implement precise clustering analysis. To achieve the high accuracy enough to implement the HOD analysis, I used the data obtained by the CFHTLS Deep Survey. The CFHTLS Deep Fields consists of four distinct fields, D1, D2, D3, and D4 field, and each of which covers ~ 1 deg^2. The D2 field is a subset of the COSMOS field (Scoville et al. 2007), in which galaxy clustering patterns are reported to be different from other fields, especially at $z \sim 2$ (cf. McCracken et al. 2010; Sato et al. 2014) due to the cosmic variance (Clowes et al. 2013; McCracken et al. 2015). We can eliminate the effects of the cosmic variance in angular correlation functions by summing up the results of four distinct fields.

The photometric image of the i-band of the Deep Survey is approximately 3 magnitudes deeper than the data of the CFHTLS Wide Survey (cf., Coupon et al. 2012; Toshikawa et al. 2016). Limiting magnitudes are measured with a 3σ 2 arcsec aperture in AB magnitude using the IRAF task `geomatch` in SDFRED. The mean 3σ limiting magnitudes of each optical band are $u^* = 28.07$, $g = 28.27$, $r = 27.76$, $i = 27.31$, and $z = 26.38$, respectively, with little field-to-field variation. Details of the photometric images at each field are summarized in Table 9. The effective survey area of the CFHTLS Deep Field was 3.38 deg^2 after masking the regions around saturated objects and frame edges.

Table 9: Survey areas and the limiting magnitudes (3σ, $2''$ aperture) at each field of the CFHTLS and WIRDS

Field	Effective Area [a]	u^*	g	r	i	z	J	H	K_s
	[deg^2]	[mag]	[mag]	[mag]	[mag]	[mag]	[mag]	[mag]	[mag]
D1	0.851	28.12	28.32	27.77	27.30	26.39	24.07	24.60	24.63
D2	0.806	28.07	28.19	27.70	27.30	26.45	24.07	24.55	24.38
D3	0.878	28.14	28.38	27.91	27.48	26.43	25.03	24.87	24.66
D4	0.843	27.96	28.19	27.67	27.17	26.26	25.02	24.59	24.46

[a] The effective areas of the WIRDS are smaller than these values.

5.2.2. NIR data

Part of the survey field of the CFHTLS Deep Survey is overlapped by the WIRCam Deep Survey (WIRDS; Bielby et al. 2012), a wide and deep NIR survey that used a Wide-Field Infra-Red Camera (WIRCam; Puget et al. 2004) mounted at the prime focus of the CFHT. The dropout galaxy samples can be constructed only by the optical images; however, the SED fitting technique can be applied by combining the data of the NIR. The WIRDS provides deep, high-quality J-, H-, and K_s-band photometric images over 1.94 deg^2, where deep optical images of the CFHTLS are available. The transmission functions of these NIR passbands are also plotted in Figure 52. I used the second data release of the WIRDS and all of the data were reduced at CFHT and TERAPIX, as is the case for the CFHTLS.

The limiting magnitudes of these NIR images are measured in the same manner as the optical data. The mean 3σ limiting magnitudes of each NIR band are $J = 24.76$, $H = 24.65$, and $K_s = 24.53$. The limiting magnitudes of each field is summarized in Table 9. A more detailed description of WIRDS

data can be found in Bielby et al. (2012).

5.2.3. Photometry and sample selection

Object detection and the sample selection of the dropout galaxies were performed in the same manner as Toshikawa et al. (2016). In this section, I present a brief outline of the photometry and sample selection method. See also Toshikawa et al. (2016) for more details.

Objects were detected by SExtractor version 2.8.6 (Bertin & Arnouts 1996) in i-band images; other optical and NIR magnitudes were measured using "double image" mode. Fixed aperture (1.4 arcsec, which was approximately twice as much as the typical seeing size of i-band images) photometry was applied to measure the fluxes of objects. The object catalogue was limited down to $i \lesssim 27.3$, which corresponds to a 3σ limiting magnitude in the i-band.

Galaxies at $z \sim 3$, 4, and 5 (corresponding to u-, g-, and r-dropout galaxies, respectively) were selected following the color criterion developed by van der Burg et al. (2010) and Toshikawa et al. (2012) as

$$\begin{aligned} u-\text{dropouts}: \quad & \Big(1.0 < (u^* - g)\Big) \cap \Big(-1.0 < (g-r) < 1.2\Big) \\ \cap \quad & \Big(1.5 \times (g-r) < (u^* - g) - 0.75\Big), \end{aligned} \tag{103}$$

$$\begin{aligned} g-\text{dropouts}: \quad & \Big(1.0 < (g-r)\Big) \cap \Big(-1.0 < (r-i) < 1.0\Big) \\ \cap \quad & \Big(1.5 \times (r-i) < (g-r) - 0.80\Big) \cap (u^* > m_{u^*,2\sigma}), \end{aligned} \tag{104}$$

and

$$\begin{aligned} r-\text{dropouts}: \quad & \Big(1.2 < (r-i)\Big) \cap \Big(-1.0 < (i-z) < 0.7\Big) \\ \cap \quad & \Big(1.5 \times (i-z) < (r-i) - 1.00\Big) \cap (g > m_{g,2\sigma}). \end{aligned} \tag{105}$$

It is noted that $m_{u^*,2\sigma}$ and $m_{g,2\sigma}$ represent the 2σ limiting magnitudes of the u^*- and g-bands. The total number of u-, g-, and r-dropout galaxies selected by the above criterion for the entire CFHTLS Deep Field were 63,563, 47,760, and 9,477, respectively. The effective survey area of the CFHTLS Deep Fields was 3.38 deg^2, after masking the regions around saturated objects and frame edges. Sky distributions of each dropout galaxy sample are presented in Figure 53, Figure 54, and Figure 55. Toshikawa et al. (2016) also selected the dropout galaxies at $z \sim 6$ (i-dropout galaxies); however, the clustering analysis cannot be carried out for the i-dropout galaxies in this work due to the small sample size. Please refer to Toshikawa et al. (2016) for the validity and the sample completeness of the selection methods.

5.3. Stellar Mass Estimation

In this section, I estimate the stellar mass of the dropout galaxy samples in two ways: using 1) an SED fitting technique, and 2) a main-sequence of star-forming galaxies.

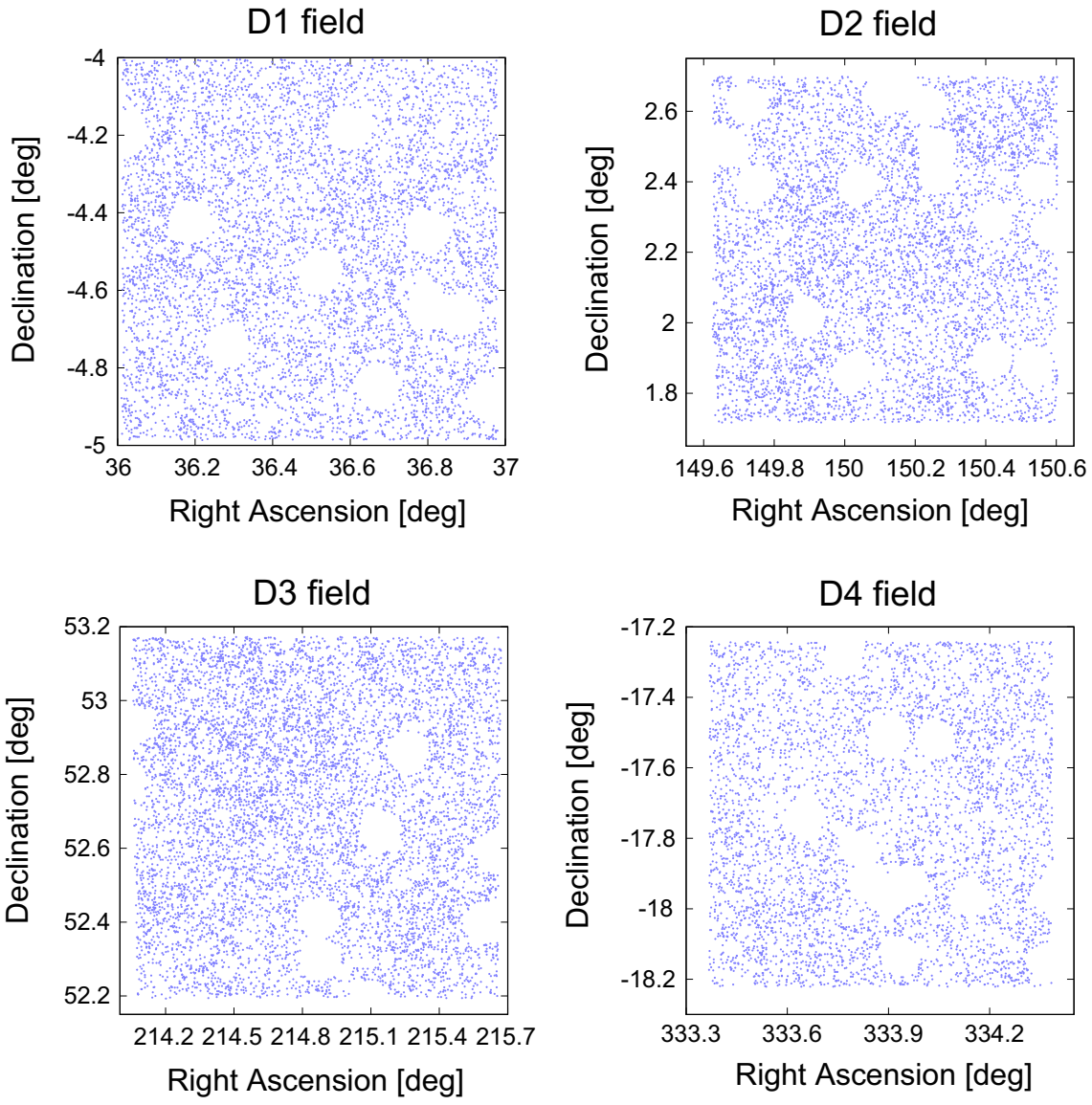

Figure 53.— The sky distribution of the u-dropout galaxies in each CFHTLS Deep field. Each panel represents the distribution in the D1 (left top), D2 (right top), D3 (left bottom), and D4 field (right bottom), respectively. Dropout galaxy samples presented in this figure are limited as $\log M_\star/h^{-1}M_\odot = 10.0$, which is evaluated by assuming the main-sequence of star-forming galaxies (see Section 5.3.2).

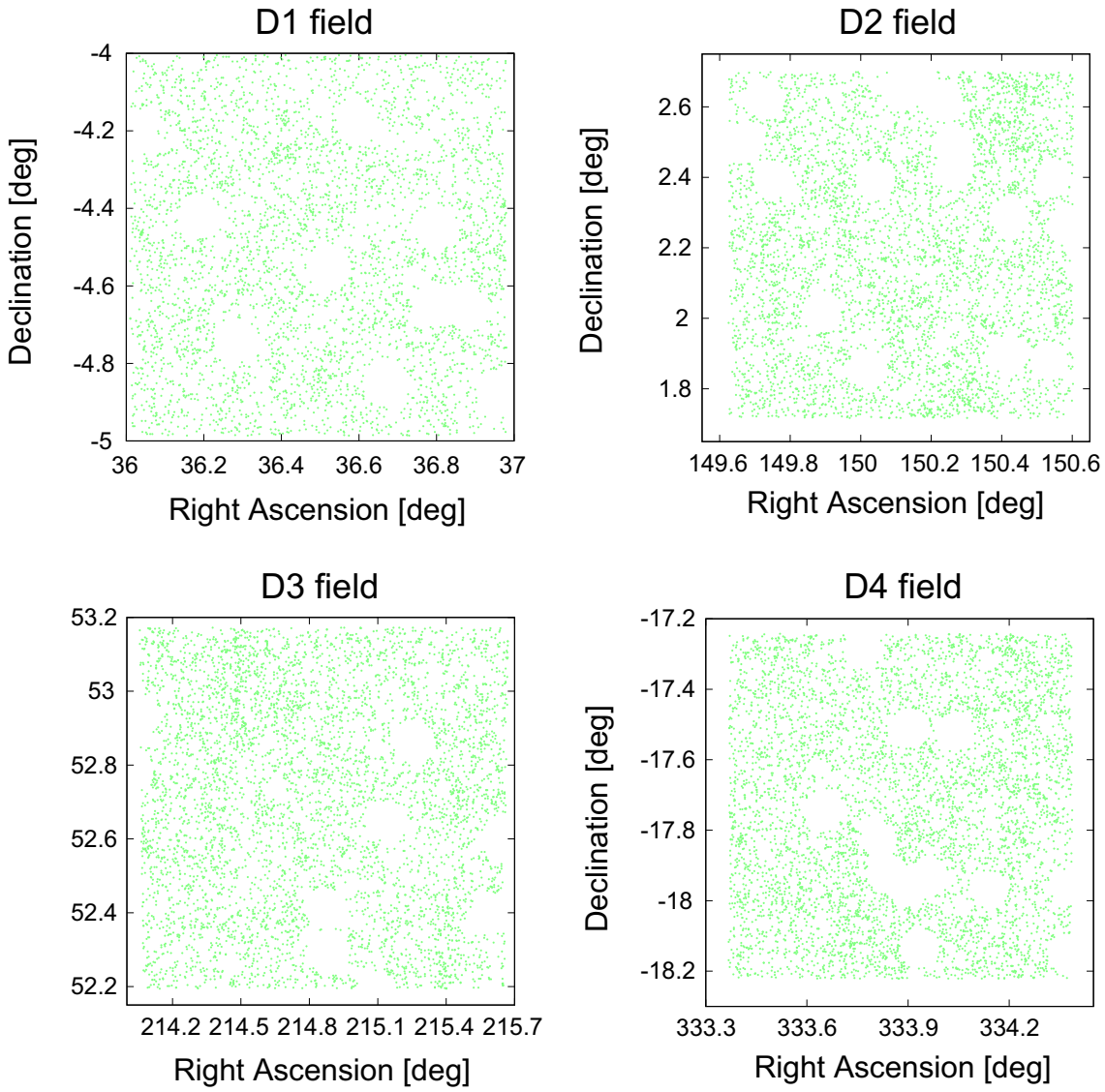

Figure 54.— Same as Figure53, albeit plotting the g-dropout galaxies.

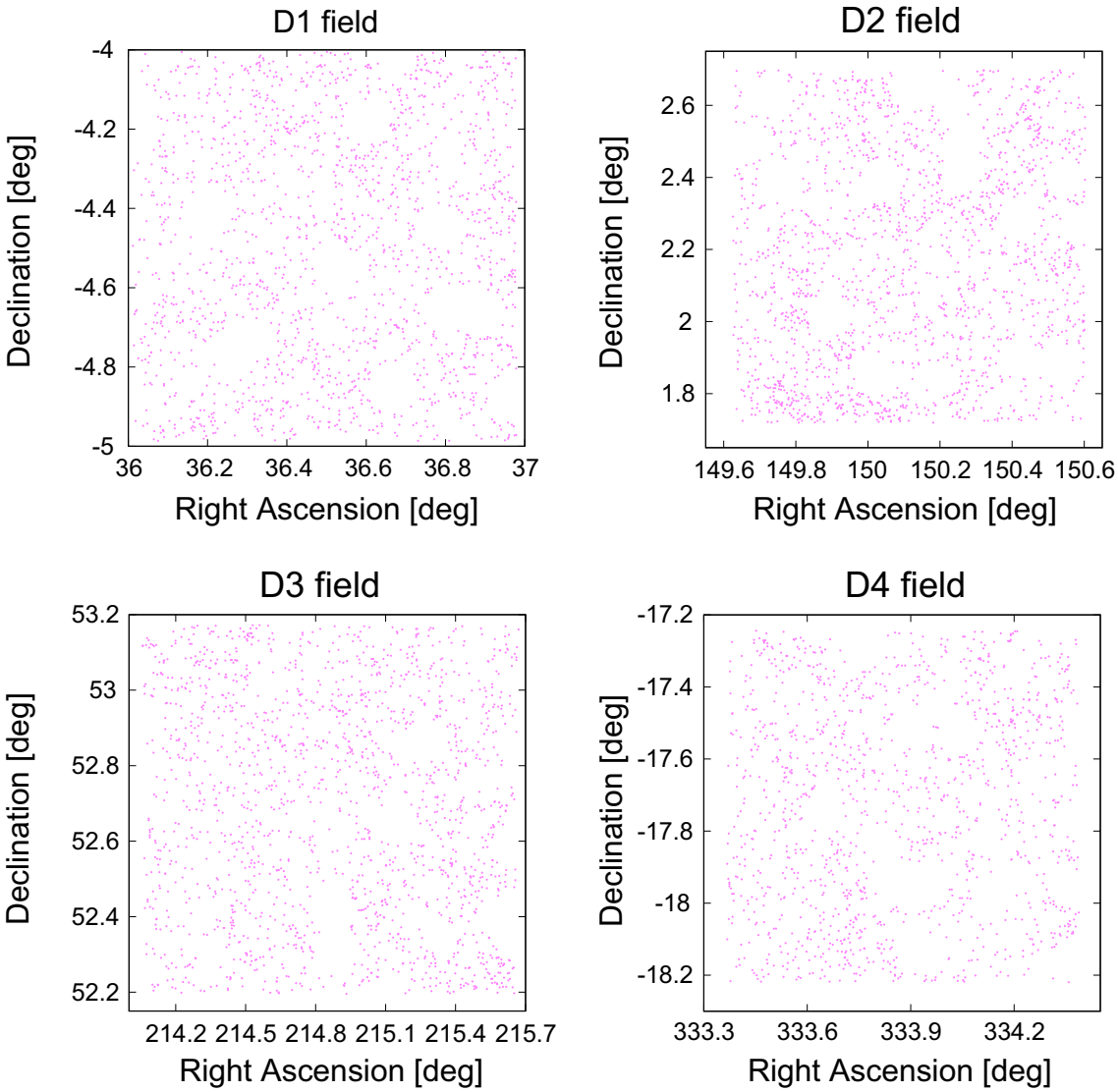

Figure 55.— Same as Figure53, albeit plotting the r-dropout galaxies.

5.3.1. SED Fitting

By combining the photometric images of CFHTLS and WIRDS, five optical data and three NIR data were available and we can apply an SED fitting technique to derive the photometric redshift of each dropout sample. I used an SED fitting code with Bayesian physical priors, MIZUKI (Tanaka 2015). The SED fitting is only applied to galaxies detected in NIR images, to ensure independence of their stellar-mass estimation from those of the main sequence of star-forming galaxies.

Galaxy SED templates are generated by the spectral synthesis model of Bruzual & Charlot (2003). An exponential-decay model with varying declination time-scale, τ, is assumed for the star-formation history of the galaxy template. The SED templates are only considered for solar metallicity abundance. I confirm that the stellar masses and photometric redshifts are not significantly changed when including SED templates with sub-solar abundance. It is also assumed that the initial mass function (IMF) is a Chabrier IMF (Chabrier 2003), the dust attenuation follows the Calzetti curve with varying the optical depth, τ_V (Calzetti et al. 2000), and the IGM attenuation follows the relation of Madau (1995). Nebular emission lines were added to the SED templates of Bruzual & Charlot (2003) with the intensity ratios of Inoue (2011) and other dust extinction law proposed by Calzetti (1997).

Photometric redshifts are evaluated through the likelihood:

$$\mathcal{L} \propto \exp\left(-\chi^2_{\rm SEDfit}/2\right), \tag{106}$$

where the $\chi^2_{\rm SEDfit}$ can be computed as

$$\chi^2_{\rm SEDfit} = \sum_i \frac{\left(f_{i,\rm obs} - \alpha f_{i,\rm model}\right)^2}{\sigma^2_{i,\rm obs}}. \tag{107}$$

$f_{i,\rm obs}$ and $f_{i,\rm model}$ are the observed and the model SED fluxes of the i-th filter, and $\sigma_{i,\rm obs}$ is the uncertainty of the i-th observed flux. α is a normalization parameter that controls the amplitude of the model SED. Physical priors are multiplied by the likelihood to obtain posteriors. It should be noted that the results of the SED fitting (e.g., the redshift distribution and the stellar-mass distribution) do not significantly change by putting off the physical priors. I evaluate the photometric redshifts of $17,341$ u-dropout galaxies and $13,298$ g-dropout galaxies, respectively.

A part of our dropout sample have been implemented spectroscopic observations by Toshikawa et al. (2016); we compare the photometric redshift, $z_{\rm phot}$, with the spectroscopic redshift, $z_{\rm spec}$, to check its accuracy. Figure 56 is a $z_{\rm spec}$ verses $z_{\rm phot}$ diagram of u- and g-dropout galaxies. The numbers of spectroscopic samples are 42 (u-dropout galaxies) and 83 (g-dropout galaxies), respectively. The photometric redshifts show good agreement with the spectroscopic redshifts; it is improved a reliability of the results of our SED fitting.

Figure 57 is the photometric-redshift distributions of u-, g-, and r-dropout galaxies. The means and standard deviations of these redshift distributions are $z_p = 3.11 \pm 0.32$, $z_p = 3.62 \pm 0.28$, and $z_p = 4.67 \pm 0.32$, respectively. These distributions are approximately consistent with the results of Toshikawa et al. (2016), who calculated the redshift distributions of u-, g-, r-, and i-dropout galaxies using a mock LBG catalogue generated by Bruzual & Charlot (2003) SED models. Our photometric-redshift distributions also agree with the results with Hildebrandt et al. (2009), who implemented the SED fitting of the u-, g-, and r-dropout galaxies at CFHT Deep fields only using the

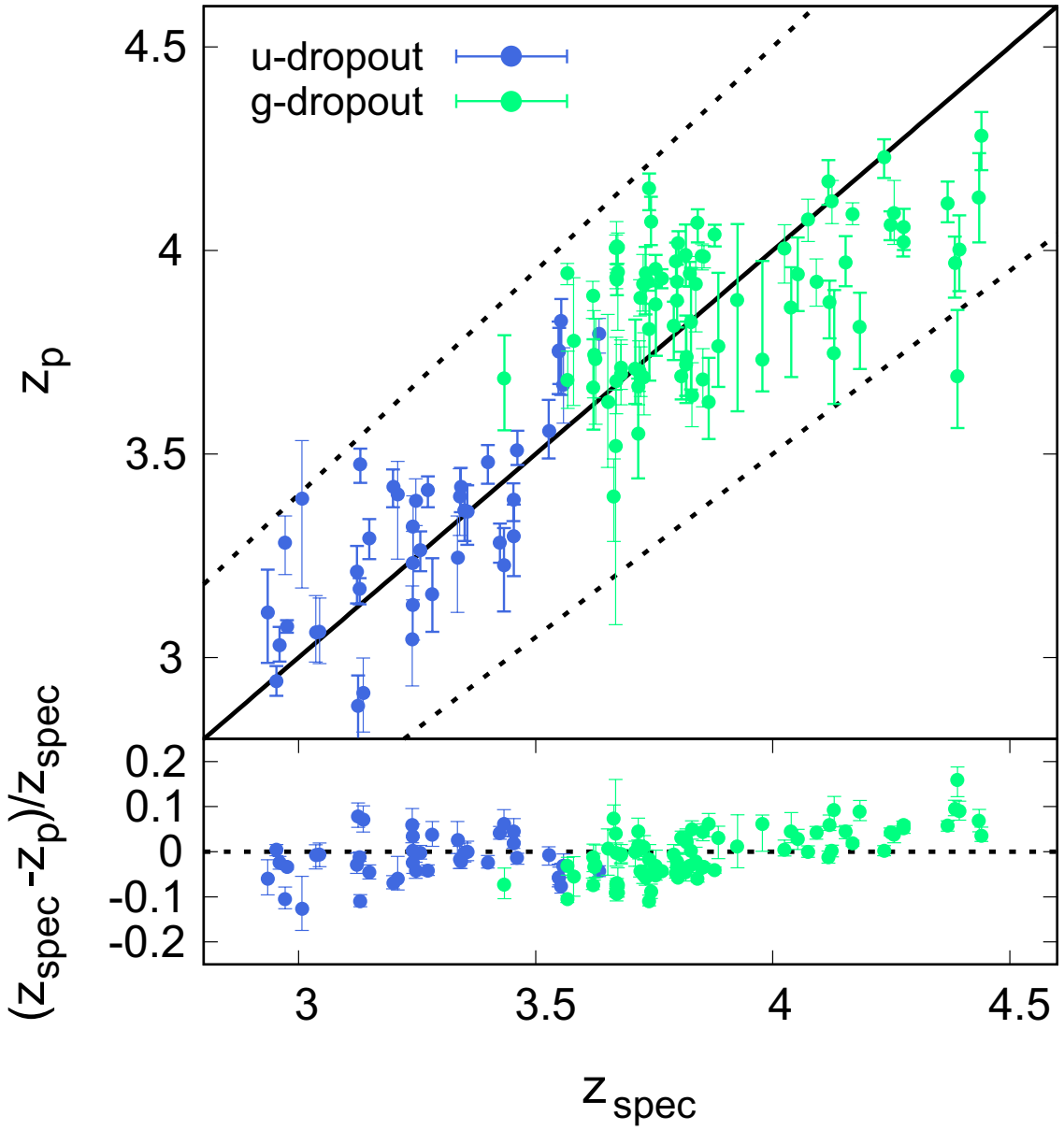

Figure 56.— Top panel shows a comparison between the photometric redshift and the spectroscopic redshift. Blue and green points indicate the u- and g-dropout galaxies. The solid black line is a one-to-one correspondence and the dotted black lines represent $|z_{\rm spec} - z_{\rm phot}|/(1 + z_{\rm spec}) = \pm 0.10$. Bottom panel shows the relative error between the photometric redshift and the spectroscopic redshift as a function of the spectroscopic redshift.

optical photometric data. Hildebrandt et al. (2009) used two SED fitting codes, Bayesian Photometric Redshifts (BPZ; Benítez 2000) and HyperZ (Bolzonella et al. 2000), and carried out the simulations using the color catalogue based upon the templates of Bruzual A. & Charlot (1993) and Maraston et al. (2006); the means and deviations of the peak redshift calculated by above four distinct methods were $z_p = 3.28 \pm 0.15$, $z_p = 3.87 \pm 0.32$, and $z_p = 4.74 \pm 0.14$ for the u-, g-, and r-dropout galaxies, respectively.

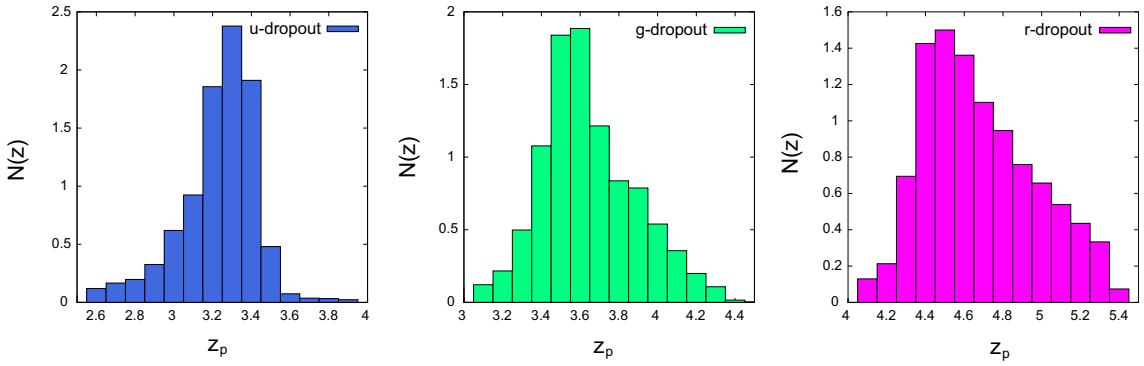

Figure 57.— The normalized photometric-redshift distributions of u- (left), g- (center), and r-dropout galaxies (right) with stellar masses of $\log(M_\star/M_\odot) \geq 10.0$. Photometric redshifts and stellar masses were evaluated using the spectral energy distribution (SED) fitting code with physical priors, Mizuki (Tanaka 2015). The means and standard deviations of these redshift distributions are $z_p = 3.11 \pm 0.32$, $z_p = 3.62 \pm 0.28$, and $z_p = 4.67 \pm 0.32$, respectively.

5.3.2. Main-sequence of star-forming galaxies

The SED fitting method allows estimation of the stellar mass; however, the total sample number decreased due to the limited survey area in which NIR data were available. Therefore, we also utilize a "main sequence" of star-forming galaxies (MS; e.g., Daddi et al. 2007; Rodighiero et al. 2011; Koyama et al. 2014) to evaluate galaxy stellar masses, instead of the SED fitting results, to use all of the dropout galaxy samples obtained in the entire CFHTLS field and achieve high S/N clustering analyses.

The MS relation is a tight correlation between galaxy stellar masses ($M_\star$) and star-formation rates (SFRs) for star-forming galaxies. The star-formation rates of our dropout galaxies are converted from their rest-frame 1450Å luminosities using the relation proposed by Kennicutt (1998). The power-law slope of the rest-frame UV continuum, β, of each dropout galaxy is measured by $(r - i)$ or $(i - z)$ colors and 1450Å luminosities were determined by extrapolating from r-, i-, or z-band magnitudes with assuming β. Dust extinction of UV flux is corrected by assuming the dust extinction low developed by Calzetti et al. (2000).

I adopt a simple linear correlation for MS as

$$\mathrm{SFR}\,(z, M_\star) = A\,(z) \times \frac{M_\star}{10^{11} M_\odot} M_\odot \mathrm{yr}^{-1}, \tag{108}$$

which is the same relation that Tanaka (2015) put as a prior for their SED fitting technique. $A(z)$ is

a redshift-dependence term defined as

$$A(z) = \begin{cases} 10 \times (1+z)^{2.1} & (z < 2) \\ 19 \times (1+z)^{1.5} & (z \geq 2). \end{cases} \tag{109}$$

Observational results (e.g., Magdis et al. 2010; Salmon et al. 2015; Álvarez-Márquez et al. 2016) and smoothed particle hydrodynamics simulations (e.g., Katsianis et al. 2015) support that dropout galaxies at $z = 3$, 4, and 5 follow our assumed MS relation. I compare stellar-mass functions of each dropout galaxy with the results of Santini et al. (2012) and Song et al. (2016), and the lowest stellar-mass limit is determined as the mass of which the observed stellar-mass functions reach $\sim 70\%$ completeness.

5.3.3. Consistency of the stellar mass estimation between the SED fitting and the MS relation

I compute stellar masses of the dropout galaxy samples with two independent estimates: a main sequence of galaxies and an SED fitting technique. The MS relation is a convenient way to assess stellar masses of star-forming galaxies from their UV luminosities; however, derived stellar-mass could suffer from non-negligible uncertainties due to the relatively large scatter. On the other hand, the SED fitting technique is frequently used to give more reliable estimates of stellar mass, although broad wavelength coverage of the data set is required.

The Balmer/4000Å break is an essential spectral feature to obtain accurate stellar masses in the SED fitting technique. It should be noted that the Balmer break can be traced by WIRCam data for u- and g-dropout galaxies, but not for r-dropout galaxies.

Figure 58 is a comparison of stellar masses estimated using the SED fitting technique and the MS relation for u-dropout galaxies in the D1 field (3,623 galaxies). These two estimations show nearly a one-to-one correspondence, albeit with relatively large scatter. $\pm 0.2'$ dex scatter in equation (108), whereas the small scatter of the SED fitting technique for the massive galaxies (red cross at the right of Figure 58) originates from their apparent Balmer/4000Å break. The same consistency can also be obtained for g-dropout galaxies. I assume that these two estimates are consistent with respect to the other, with minimal significant difference. Hereafter, we will use the MS relation that allows stellar mass estimation down to the faint magnitudes for the entire CFHTLS fields, even without WIRCam data, to estimate stellar mass in the following analyses. I also assume that the consistency between these two estimates was valid for r-dropout galaxies. The effects of these two stellar-mass estimation on the SHMR results are discussed in Section 5.6.1.

5.4. Clustering Analysis

The large number of dropout galaxy samples enables to carry out a precision clustering analysis to derive the high-quality ACFs enough to implement the HOD analysis to investigate the relationship between dropout galaxies and their host dark haloes. I divide dropout galaxy samples into subsamples by their stellar mass, which is evaluated by the MS relation. The lowest stellar mass limits are $M_{\star,\mathrm{limit}} = 10^{9.4} M_\odot$ for the u- and g-dropout galaxies and $M_{\star,\mathrm{limit}} = 10^{9.8} M_\odot$ for r-dropout galaxies,

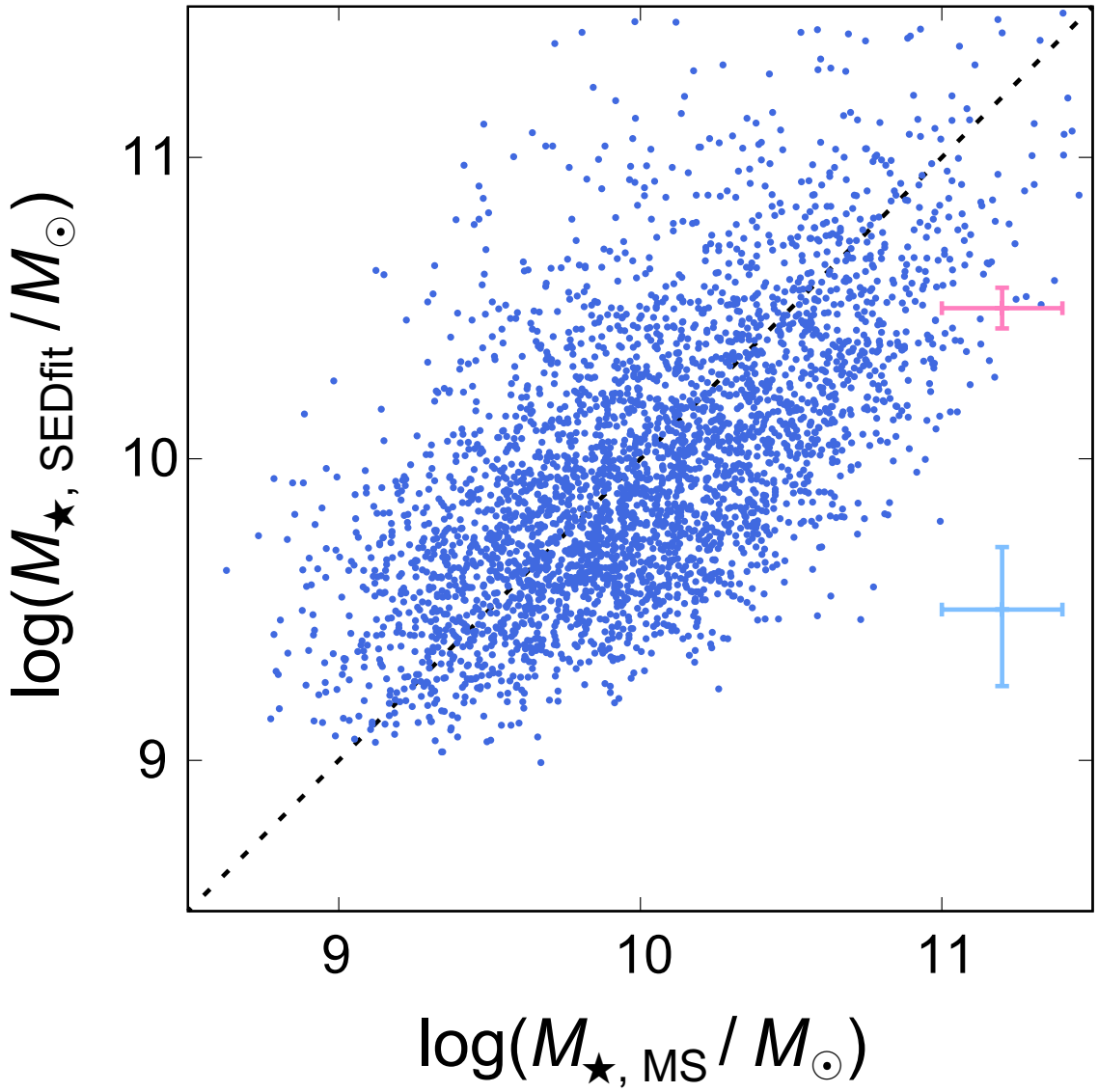

Figure 58.— Comparison of stellar masses of u-dropout galaxies in the D1 field derived by the SED fitting technique and assuming the main-sequence of star-forming galaxies. The black dotted line represents the relation of the one-to-one correspondence. Crosses shown in the figure at the bottom right represent the typical errors of $\log(M_{\star,\mathrm{SEDfit}}/M_{\odot}) < 10$ (blue) and $\log(M_{\star,\mathrm{SEDfit}}/M_{\odot}) \geq 10$ (red) samples.

which are determined by the 70% completeness limit derived by the comparison with the stellar mass functions from literature (Santini et al. 2012; Song et al. 2016). The subsamples of each dropout galaxy are generated with 0.2 dex increments of the stellar-mass limit. The number of dropout galaxy samples of each subsample are summarized in Table 10. ACFs of some subsamples of each dropout galaxy are shown in Figure 59 − Figure 61.

I correct the effect contaminations from low-z galaxies, which decrease the measured clustering amplitudes. I assume that contamination sources have homogeneous distribution; contamination effects are corrected by multiplying clustering amplitudes by $1/(1-f_{\rm c})^2$, where $f_{\rm c}$ is the contamination fraction calculated by the photometric-redshift distributions. It should be noted that the effect of contamination correction is much smaller than the statistical jackknife error of each angular bin.

Table 10: The number of the dropout galaxy samples and the best-fit HOD parameters of each subsample limited by the stellar mass

	$\log(M_{\star,\mathrm{limit}}/M_\odot)$	N	$\log(M_{\mathrm{min}}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\log(M_{\mathrm{h}}/h^{-1}M_\odot)$	f_{sat}
u-dropout	9.4	59,233	$11.38^{+0.03}_{-0.02}$	$13.17^{+0.06}_{-0.06}$	$11.78^{+0.02}_{-0.02}$	0.036 ± 0.002
	9.6	50,159	$11.49^{+0.02}_{-0.02}$	$13.34^{+0.05}_{-0.05}$	$11.86^{+0.01}_{-0.01}$	0.031 ± 0.002
	9.8	36,890	$11.60^{+0.02}_{-0.02}$	$13.48^{+0.06}_{-0.05}$	$11.94^{+0.01}_{-0.01}$	0.027 ± 0.002
	10.0	23,666	$11.70^{+0.02}_{-0.02}$	$13.54^{+0.06}_{-0.06}$	$12.01^{+0.01}_{-0.02}$	0.028 ± 0.002
	10.2	13,081	$11.83^{+0.02}_{-0.02}$	$13.68^{+0.07}_{-0.06}$	$12.11^{+0.01}_{-0.01}$	0.026 ± 0.002
	10.4	5,901	$12.06^{+0.02}_{-0.02}$	$14.04^{+0.11}_{-0.10}$	$12.28^{+0.02}_{-0.02}$	0.017 ± 0.002
	10.6	2,230	$12.17^{+0.02}_{-0.02}$	$14.28^{+0.15}_{-0.10}$	$12.34^{+0.01}_{-0.01}$	0.015 ± 0.004
	10.8	664	$12.47^{+0.03}_{-0.03}$	$14.43^{+0.36}_{-0.25}$	$12.55^{+0.02}_{-0.02}$	0.013 ± 0.005
	11.0	201	$12.76^{+0.03}_{-0.03}$	$14.80^{+0.14}_{-0.24}$	$12.76^{+0.02}_{-0.02}$	0.009 ± 0.004
g-dropout	9.4	41,373	$11.32^{+0.03}_{-0.02}$	$13.05^{+0.06}_{-0.05}$	$11.64^{+0.02}_{-0.02}$	0.036 ± 0.002
	9.6	37,537	$11.39^{+0.02}_{-0.02}$	$13.08^{+0.05}_{-0.04}$	$11.70^{+0.01}_{-0.01}$	0.038 ± 0.002
	9.8	28,741	$11.51^{+0.02}_{-0.01}$	$13.29^{+0.06}_{-0.05}$	$11.79^{+0.01}_{-0.01}$	0.029 ± 0.002
	10.0	23,666	$11.63^{+0.02}_{-0.01}$	$13.36^{+0.05}_{-0.05}$	$11.88^{+0.01}_{-0.02}$	0.031 ± 0.002
	10.2	18,333	$11.78^{+0.02}_{-0.02}$	$13.46^{+0.06}_{-0.06}$	$11.99^{+0.01}_{-0.01}$	0.032 ± 0.002
	10.4	4,443	$11.94^{+0.03}_{-0.02}$	$13.67^{+0.10}_{-0.08}$	$12.11^{+0.02}_{-0.02}$	0.027 ± 0.003
	10.6	1,674	$12.12^{+0.02}_{-0.02}$	$13.86^{+0.15}_{-0.10}$	$12.32^{+0.03}_{-0.03}$	0.005 ± 0.002
	10.8	541	$12.42^{+0.03}_{-0.03}$	$14.38^{+0.39}_{-0.26}$	$12.49^{+0.02}_{-0.02}$	0.005 ± 0.001
	11.0	145	$12.71^{+0.02}_{-0.03}$	$14.80^{+0.14}_{-0.24}$	$12.75^{+0.02}_{-0.02}$	0.004 ± 0.004
r-dropout	9.8	6,707	$11.45^{+0.06}_{-0.06}$	$13.07^{+0.02}_{-0.02}$	$11.65^{+0.05}_{-0.05}$	0.036 ± 0.007
	10.0	6,271	$11.57^{+0.01}_{-0.01}$	$13.37^{+0.08}_{-0.11}$	$11.74^{+0.02}_{-0.02}$	0.020 ± 0.002
	10.2	4,925	$11.74^{+0.03}_{-0.03}$	$13.62^{+0.10}_{-0.11}$	$11.85^{+0.02}_{-0.02}$	0.013 ± 0.002
	10.4	2,806	$12.01^{+0.02}_{-0.02}$	$13.95^{+0.10}_{-0.09}$	$12.08^{+0.02}_{-0.02}$	0.013 ± 0.002
	10.6	1,181	$12.16^{+0.05}_{-0.04}$	$14.25^{+0.21}_{-0.33}$	$12.19^{+0.03}_{-0.04}$	0.009 ± 0.003

The halo occupation is assumed the formalism proposed by Zheng et al. (2005), whose HOD free parameters are $M_{\rm min}$, M_1, M_0, $\sigma_{\log M}$, and α. To achieve better constraint on the HOD mass parameters, especially for the $M_{\rm min}$ and the M_1, I fixed two HOD parameters as $\sigma_{\log M} = 0.30$ and $\alpha = 1.0$. $M_{\rm min}$, M_1, and M_0 are varied at the ranges of $M_{\rm min} \in [10; 14]$, $M_1 \in [11; 15]$, and $M_{\rm min} \in [8; 13]$. These parameter ranges are estimated to be sufficient by the previous HOD studies of other redshift galaxies because there is no high-z HOD studies that adopt the same HOD model as ours and are comparable to our stellar-mass ranges.

In predicting ACFs from the HOD model, it is assumed the halo mass function (HMF) proposed by Sheth & Tormen (1999), the large-scale halo bias factor of Tinker et al. (2010), and the density profile of the dark halo as a NFW profile (Navarro et al. 1997). Halo mass–concentration relation is employed the model of Takada & Jain (2003). Photometric redshifts of each dropout galaxy subsample are employed as their redshift distributions. I use the analytical formula of HMF of Sheth & Tormen (1999) because it is easy to treat the fitting function and compare our results with literature; however, recent cosmological N-body simulations suggest that the HMF of Sheth & Tormen (1999) overpredicts the massive end at high redshift ($z \gtrsim 2$). Ishiyama et al. (2015) calculated the redshift evolution of the HMF using the result of the ν^2GC (New Numerical Galaxy Catalog) simulation (Makiya et al. 2016) and found that their best-fitted HMF at high-mass end ($\langle M_{\rm h} \rangle \gtrsim 10^{13} M_\odot$) at $z > 3$ is lower than Sheth & Tormen (1999). However, the difference is small within their 1σ errors even at least $z = 7$. Tinker et al. (2008) and its improved model tuned to high-z galaxies by Behroozi et al. (2013b) showed that, at high redshift, massive end of the HMF of Sheth & Tormen (1999) overpredicts compared to the results of the N-body simulations. However, the difference of the HMF at $z = 3$ between Sheth & Tormen (1999) and Behroozi et al. (2013b) is negligible for less-massive haloes ($\langle M_{\rm h} \rangle \sim 10^{12} M_\odot$), whereas the apparent excess ($\sim 70\%$ excess of the Sheth & Tormen HMF) can be seen for massive haloes ($\langle M_{\rm h} \rangle \sim 10^{14} M_\odot$), which is a typical halo mass range of M_1 (see Table 10). However, both HMFs are within the amplitudes difference at 1σ error range of M_1; thus, we conclude that the difference of the HMF does not make a large impact on the final results. I used publicly available code, "CosmoPMC", to constrain the HOD parameters and evaluate the best-fit ACFs (Wraith et al. 2009; Kilbinger et al. 2010). The HOD parameters are estimated by the MCMC simulation. The covariance matrix is estimated by the jackknife resampling method. I divided our survey field into 64 sub-fields and evaluated the ACF by excluding one sub-field. This procedure was repeated 64 times.

The results of the HOD analyses are plotted over the observed ACFs in Figure 10. The excesses for ACFs from a single power law at small-angular scales are well described by the HOD model. The best-fit HOD parameters and deduced parameters are presented in Table 10.

Our HOD fittings slightly deviate from the observed ACFs at the 2-halo term, i.e., $0.01 \lesssim \theta \lesssim 0.1$ for u- and g-dropout galaxies. This could be due to the fixing HOD parameters, especially $\sigma_{\log M}$ that controls the occupation of the central galaxy near the dark halo mass of $M_{\rm min}$. I check whether the HOD fitting will be improved by varying all of the HOD free parameters for u-dropout samples and the results are shown in Figure 62 and Table 11. It is noted that the two most massive subsamples, i.e., $\log(M_\star/M_\odot) \geq 10.8$ and $\log(M_\star/M_\odot) \geq 11.0$ bins, have not enough S/N ratios to fit the HOD model with five free parameters. By varying $\sigma_{\log M}$ and α, fittings of the 2-halo terms seem to be improved; however, the ACF of the 1-halo term at small scales and the transition scales between the 1-halo term and the 2-halo term cannot be reproduced. In addition, values of the χ^2 are not significantly changed by varying $\sigma_{\log M}$ and α due to the decrease of the degree-of-freedom. Moreover,

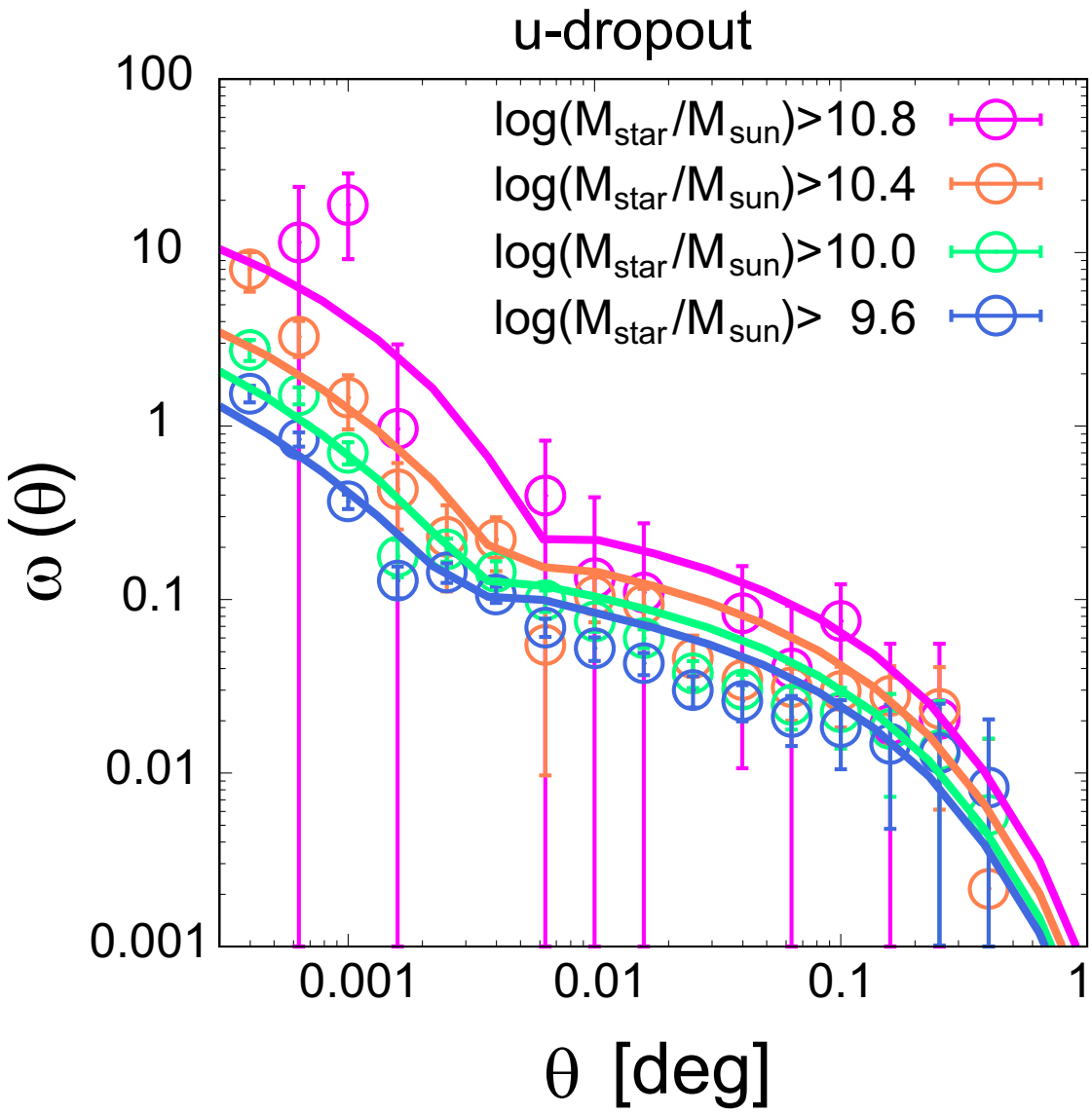

Figure 59.— Observed angular correlation functions (ACFs) of u-dropout galaxies as a function of separation angular scales in degree. The dropout galaxy samples are divided by their stellar masses evaluated by assuming the main sequence of star-forming galaxies. Error bars of each ACF are evaluated by the jackknife resampling technique. Solid lines are the best-fit ACFs of each subsample calculated by the halo occupation distribution (HOD) model. Correlations between angular bins are taken into consideration via covariance matrixes computed by jackknife resampling.

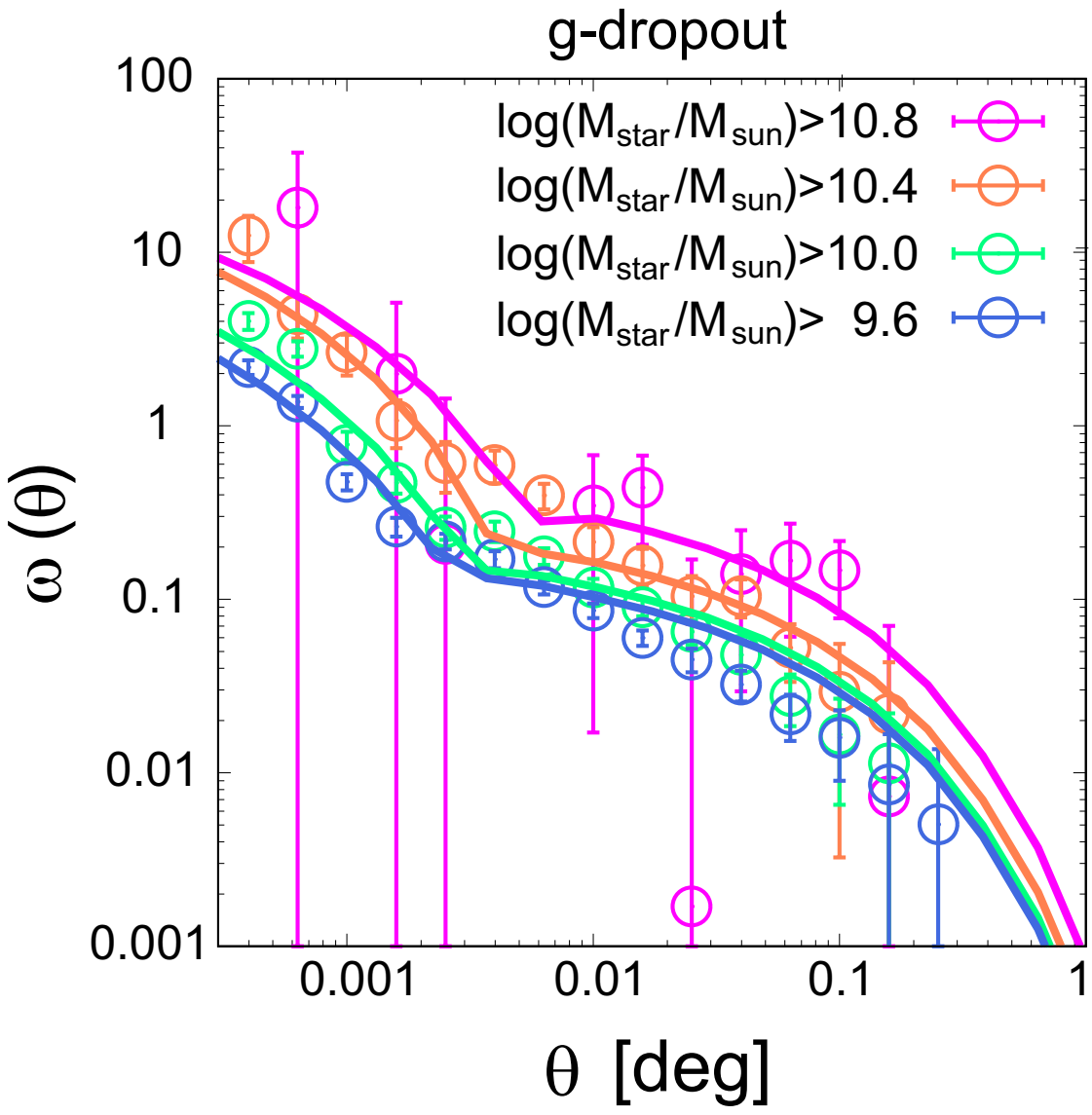

Figure 60.— Same as Figure 59, but for *g*-dropout galaxies..

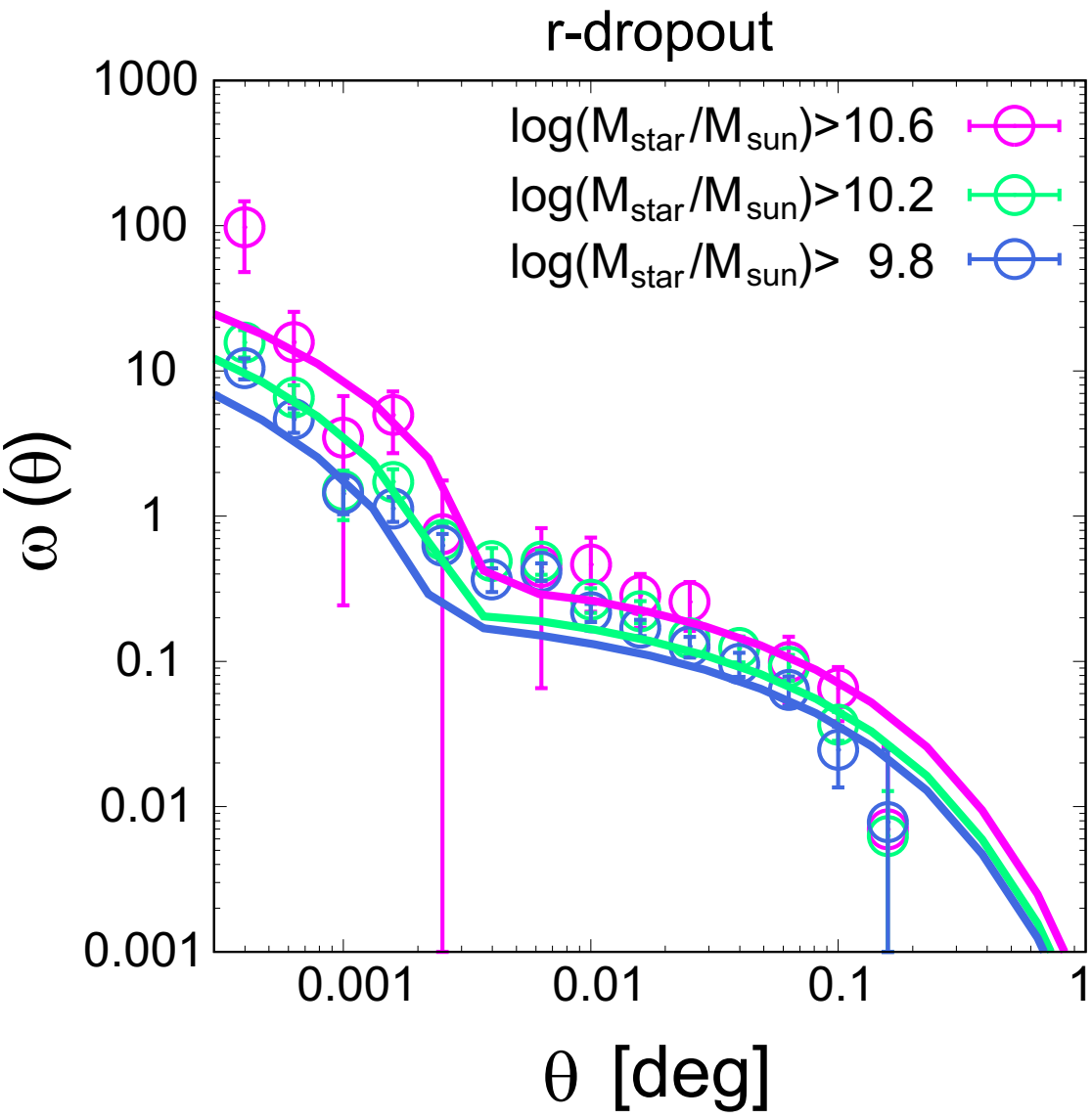

Figure 61.— Same as Figure 59, but for r-dropout galaxies..

all of the fitting results show very small $\sigma_{\log M}$, i.e., $\sigma_{\log M} < 0.1$, indicating that the occupation of central galaxies follows almost the step function. This could imply that the halo occupation function at $z > 3$ is not the same as that of local Universe. The validity of the halo occupation function in the form of equation (60) and (61) should be verified for high-z galaxies; however, the discussion is beyond the scope of this thesis. In this study, I adopt the results of the HOD fitting by fixing the parameters of $\sigma_{\log M} = 0.30$ and $\alpha = 1.0$ to confine the following discussion to $M_{\rm min}$ and M_1.

Table 11: The best-fit HOD parameters with 1σ errors of u-dropout galaxies limited by the stellar mass by varying all of the HOD free parameters

$\log(M_{\star,\rm limit}/M_\odot)$	$\log(M_{\rm min}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\sigma_{\log M}$	α	χ^2/dof
9.4	$11.32^{+0.03}_{-0.04}$	$13.23^{+0.23}_{-0.30}$	$0.08111^{+0.01988}_{-0.04580}$	$0.8907^{+0.2407}_{-0.1186}$	4.96
9.6	$11.50^{+0.04}_{-0.03}$	$13.70^{+0.46}_{-0.33}$	$0.04049^{+0.01503}_{-0.01866}$	$0.9791^{+0.2134}_{-0.1930}$	6.38
9.8	$11.55^{+0.05}_{-0.03}$	$13.71^{+0.35}_{-0.34}$	$0.05911^{+0.02037}_{-0.03020}$	$0.9260^{+0.2410}_{-0.1534}$	5.92
10.0	$11.62^{+0.02}_{-0.04}$	$13.47^{+0.38}_{-0.29}$	$0.04425^{+0.01418}_{-0.02438}$	$0.9737^{+0.2278}_{-0.1883}$	5.79
10.2	$11.81^{+0.01}_{-0.02}$	$13.71^{+0.59}_{-0.20}$	$0.02057^{+0.00897}_{-0.00650}$	$1.076^{+0.154}_{-0.250}$	5.36
10.4	$11.89^{+0.01}_{-0.01}$	$13.90^{+0.43}_{-0.32}$	$0.02673^{+0.01209}_{-0.01104}$	$0.9768^{+0.2236}_{-0.1942}$	5.23
10.6	$12.14^{+0.01}_{-0.01}$	$14.13^{+0.28}_{-0.33}$	$0.01740^{+0.00833}_{-0.00473}$	$1.075^{+0.151}_{-0.153}$	5.12

5.5. HOD Parameters

In this section, I will compare the HOD parameters with results of previous HOD studies of high-z galaxies ($3 \lesssim z \lesssim 5$) as well as lower-z galaxies ($z \lesssim 1$), and investigate the dependence of these parameters on the stellar mass and redshift. The units of all of the mass parameters will be converted from $h^{-1}M_\odot$ into $M_\odot$ by assuming $h = 0.7$ when I compare them with literature.

5.5.1. HOD mass parameters

One of the advantages of adopting the HOD formalism is that each HOD parameter has an explicit physical meaning; $M_{\rm min}$ and $M_1 + M_0$ are the characteristic halo masses to possess one central and one satellite galaxy within the dark haloes, respectively. In Figure 63, I show the best-fit values of $M_{\rm min}$ and M_1 of each subsample as a function of the threshold galaxy stellar mass. $M_{\rm min}$ and M_1 show approximately linear increases with the stellar-mass limit. I also plot $M_{\rm min}$ and M_1 at $0.5 < z < 0.8$ and $1.1 < z < 1.5$ by McCracken et al. (2015) based upon HOD analysis for all galaxies collected by photometric redshifts, although $M_\star$ of McCracken et al. (2015) represents the median stellar mass. By comparing the results at low-z, the increasing trends of $M_{\rm min}$, i.e., the threshold halo mass to possess a central galaxy, are not significantly different at high-z, indicating that the formation efficiencies of the central galaxies are almost the same, at least $z = 5 - 0$.

However, the behavior of M_1, which is the threshold halo mass to host a satellite galaxy, is clearly different between $z = 3 - 5$ and $z = 0 - 1$. First, M_1 increases linearly with stellar mass from the less massive end to intermediate stellar mass ($M_\star \sim 10^{10.6} M_\odot$), and evolves drastically toward the massive

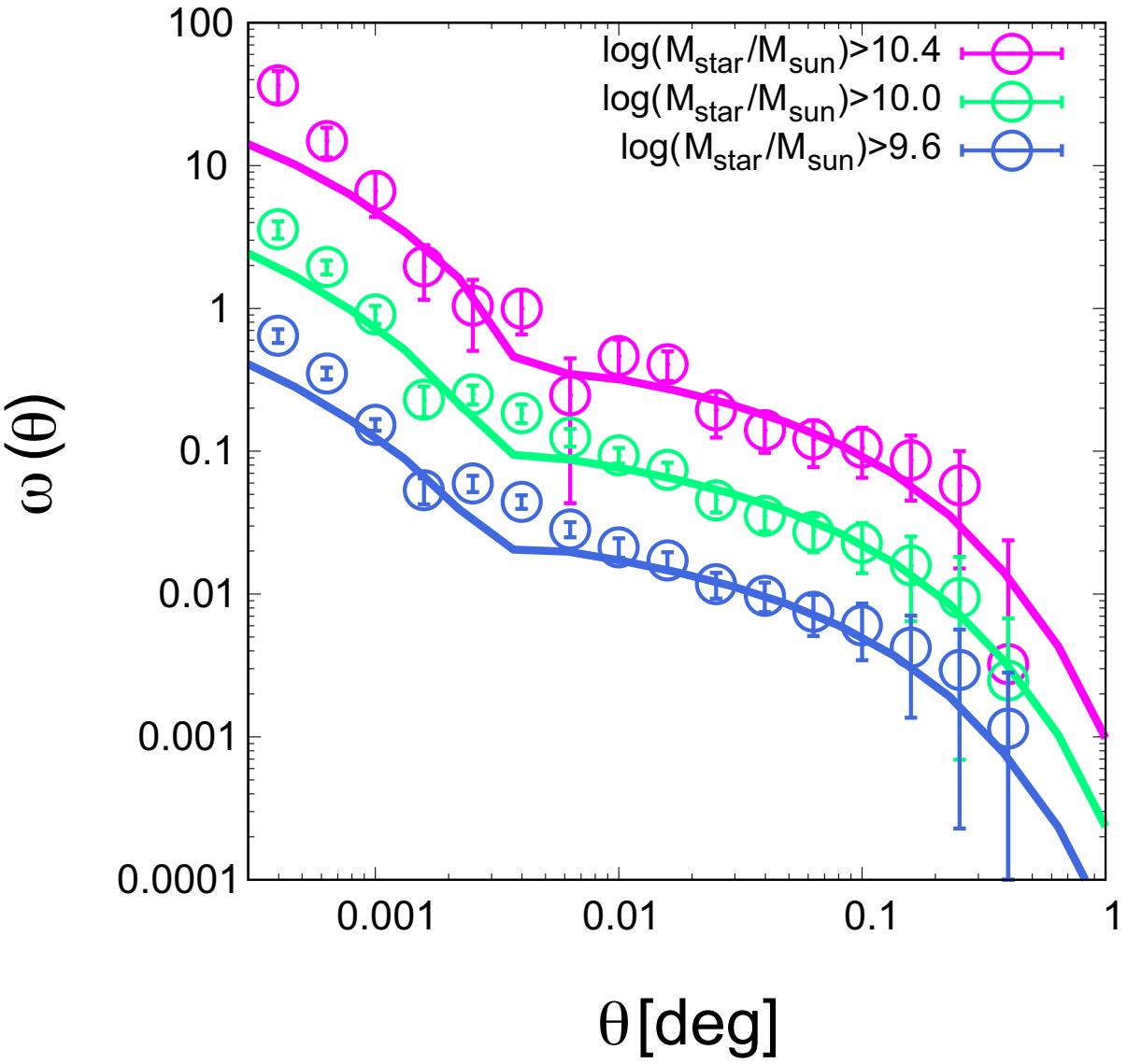

Figure 62.— Results of the HOD fitting of u-dropout galaxies by varying all of the HOD free parameters, $M_{\rm min}$, M_1, M_0, $\sigma_{\log M}$, and α. To clarify the results, amplitudes of the observed ACFs and the HOD fittings are scaled arbitrarily.

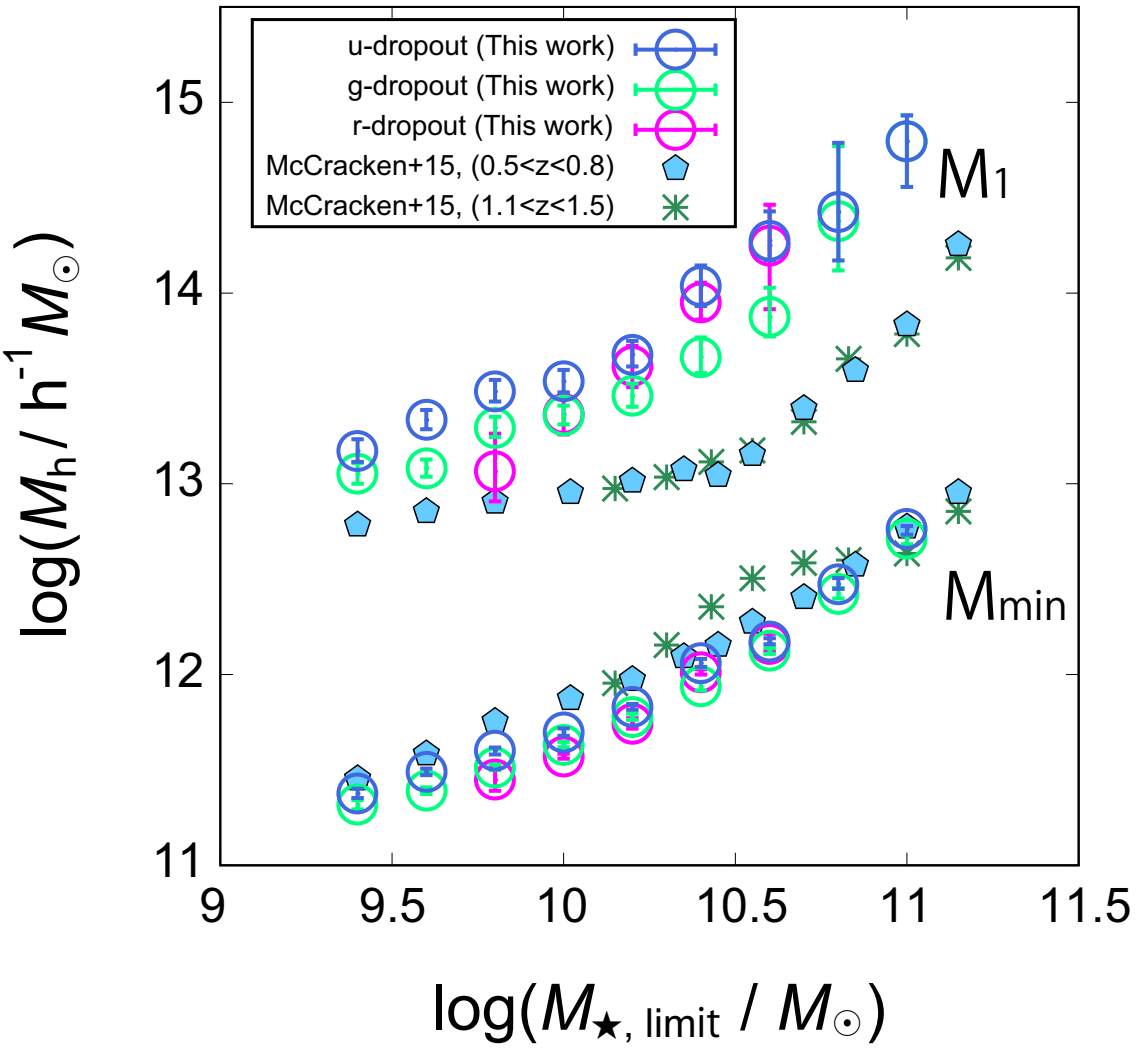

Figure 63.— Best-fit HOD parameters from HOD fitting as a function of the stellar-mass limit. The HOD parameters $M_{\rm min}$ and M_1 represent characteristic halo masses to possess one central galaxy and a satellite galaxy within dark haloes, respectively. The results of McCracken et al. (2015) at $0.5 < z < 0.8$ and $1.1 < z < 1.5$ are also shown.

stellar-mass end at low-z Universe. This trend is the same as the results of high-z; however, the turning stellar mass apparently shifts toward the lower stellar mass compared with low-z. The rapid increase of M_1 at high stellar mass indicates that massive satellite galaxies are formed less efficiently than less massive satellite galaxies. Zheng et al. (2009) also found a marked increase in M_1 and concluded that this is due to the short time-scale to destroy merged subhaloes because major mergers onto massive haloes have taken place until recently (Zentner et al. 2005). The lower turnover stellar mass of M_1 at high-z Universe suggests that dynamical frictions for accreted haloes are not yet sufficient even for less massive haloes.

In addition, M_1 are apparently excess for $z = 3-5$ compared with $z = 0-1$, inferring that satellite galaxies only form within massive dark haloes in high-z Universe. One possible explanation is that dark haloes at high-z Universe have higher virial temperature than those in local Universe, because high-z dark haloes have not been completely virialized. This prevents intra-halo gases from cooling and consequently star formation efficiency decreases. Another possibility is the effect of feedback from central galaxies. Our galaxy samples were confined to star-forming galaxies and galactic winds, if it is more frequent or powerful at high-z, from central galaxies may hamper the formation of satellite galaxies.

5.5.2. Mean halo masses

The mean halo mass, $\langle M_{\rm h} \rangle$, and the satellite fraction, $f_{\rm sat}$, can be estimated from the best-fit HOD parameters for each subsample. The mean halo mass and the satellite fraction are defined as equation (81) and equation (82), respectively.

The mean halo masses as a function of the stellar-mass threshold of each dropout galaxy are shown in Figure 64 − Figure 66. The large number of our dropout galaxy sample enables us to measure the masses of dark haloes with small error bars, even for very massive haloes ($M_{\rm h} \sim 10^{13} M_{\odot}$) that could not be estimated by previous studies, though the fixed $\sigma_{\log M}$ and α artificially reduced the errors. At a fixed stellar-mass limit, the mean halo masses decrease slightly with increasing redshift.

I compare the mean halo masses measured by the previous HOD studies for high-z Universe; Hamana et al. (2004), Lee et al. (2006), Hildebrandt et al. (2009), Bian et al. (2013), and Harikane et al. (2016). Our results are almost consistent with these studies within 1σ confidence level; however, the halo masses of $z \sim 5$ dropout galaxies calculated by Hamana et al. (2004) and Lee et al. (2006) were lower than those of Hildebrandt et al. (2009) and our results. It should be noted that other studies adopted the different HOD model, as well as cosmological parameters, from the ones in the present study. Hamana et al. (2004) proposed and used an early halo occupation model that does not distinguish the central galaxy from satellite galaxies, and Lee et al. (2006), Hildebrandt et al. (2009), and Bian et al. (2013) also employed the same occupation model.

It should also be taken into consideration that cosmological parameters are not unified between these studies. All of the referenced works used simple flat ΛCDM cosmology ($\Omega_{\rm m} = 0.3$, $\Omega_{\Lambda} = 0.7$), whereas I adopted WMAP seven-year cosmologies ($\Omega_{\rm m} = 0.27$, $\Omega_{\Lambda} = 0.73$). The largest impact on the halo mass estimation is the value of σ_8. This study and Harikane et al. (2016) adopted $\sigma_8 = 0.8$, but other referenced studies used $\sigma_8 = 0.9$; thus, our results may underestimate the halo mass compared with studies using $\sigma_8 = 0.9$.

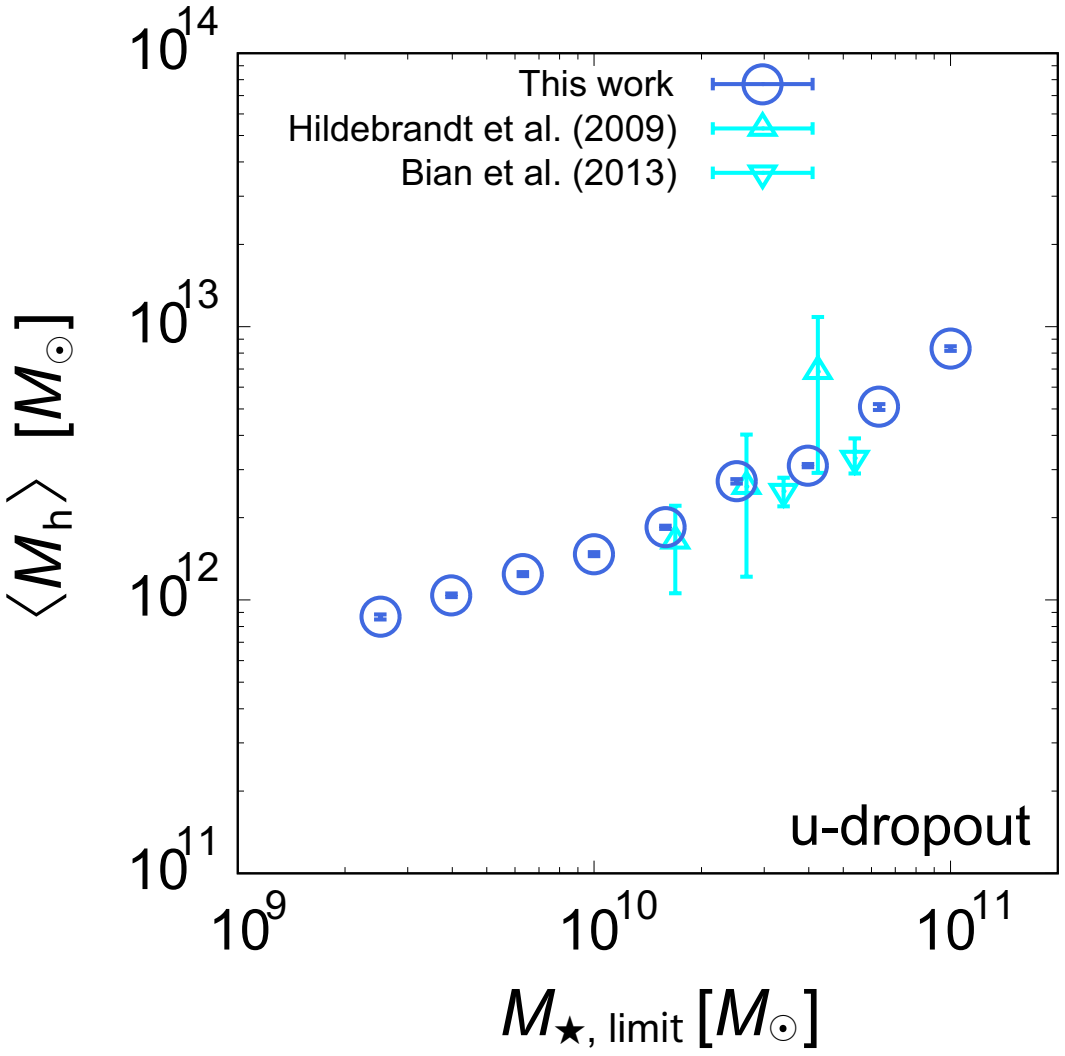

Figure 64.— Mean halo masses of u-dropout galaxies as a function of the stellar-mass limit. For comparison, we also show the mean halo masses of previous HOD studies. Note that the stellar-mass limits from literature are converted from their UV-magnitude limits assuming the UV slope of $\beta = -1.5$. All masses are in units of $M_\odot$, assuming that $h = 0.7$.

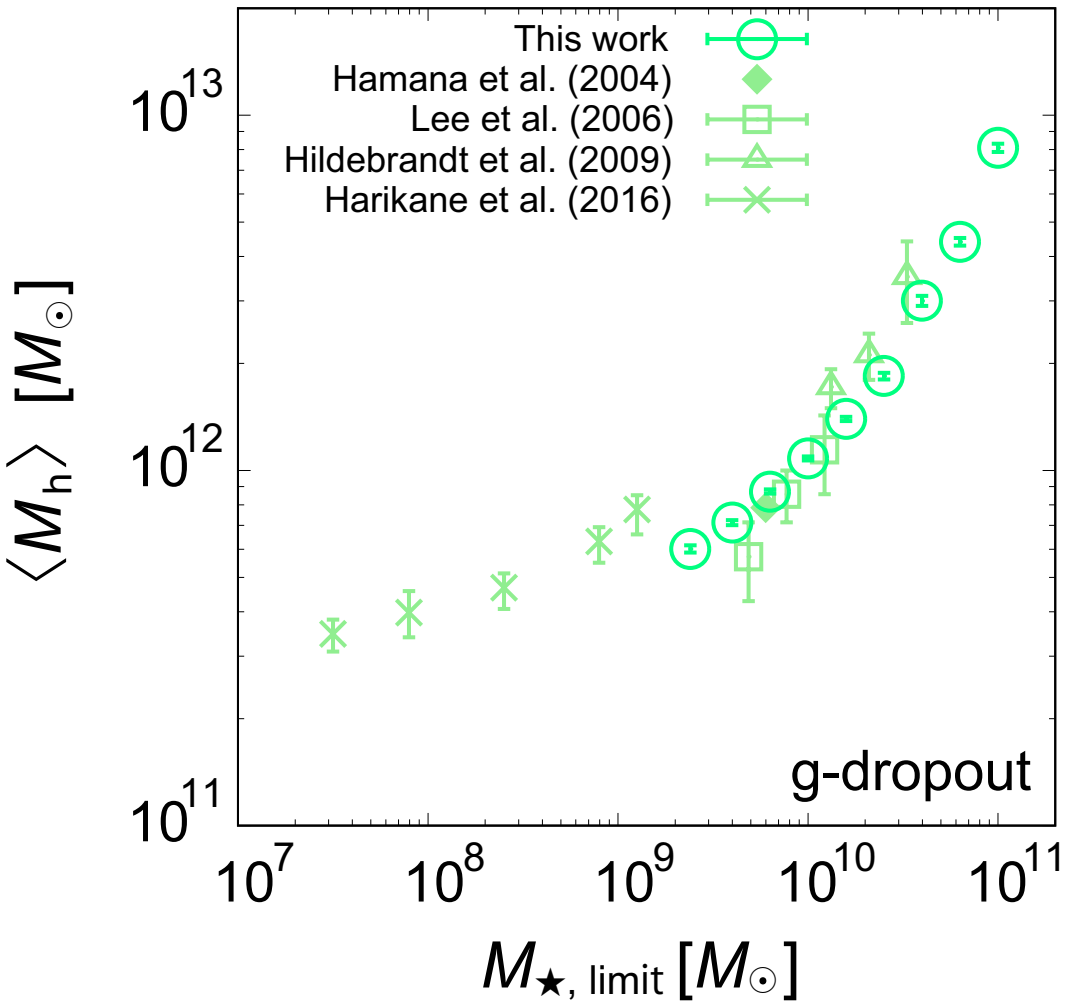

Figure 65.— Same as Figure 64, but for *g*-dropout galaxies.

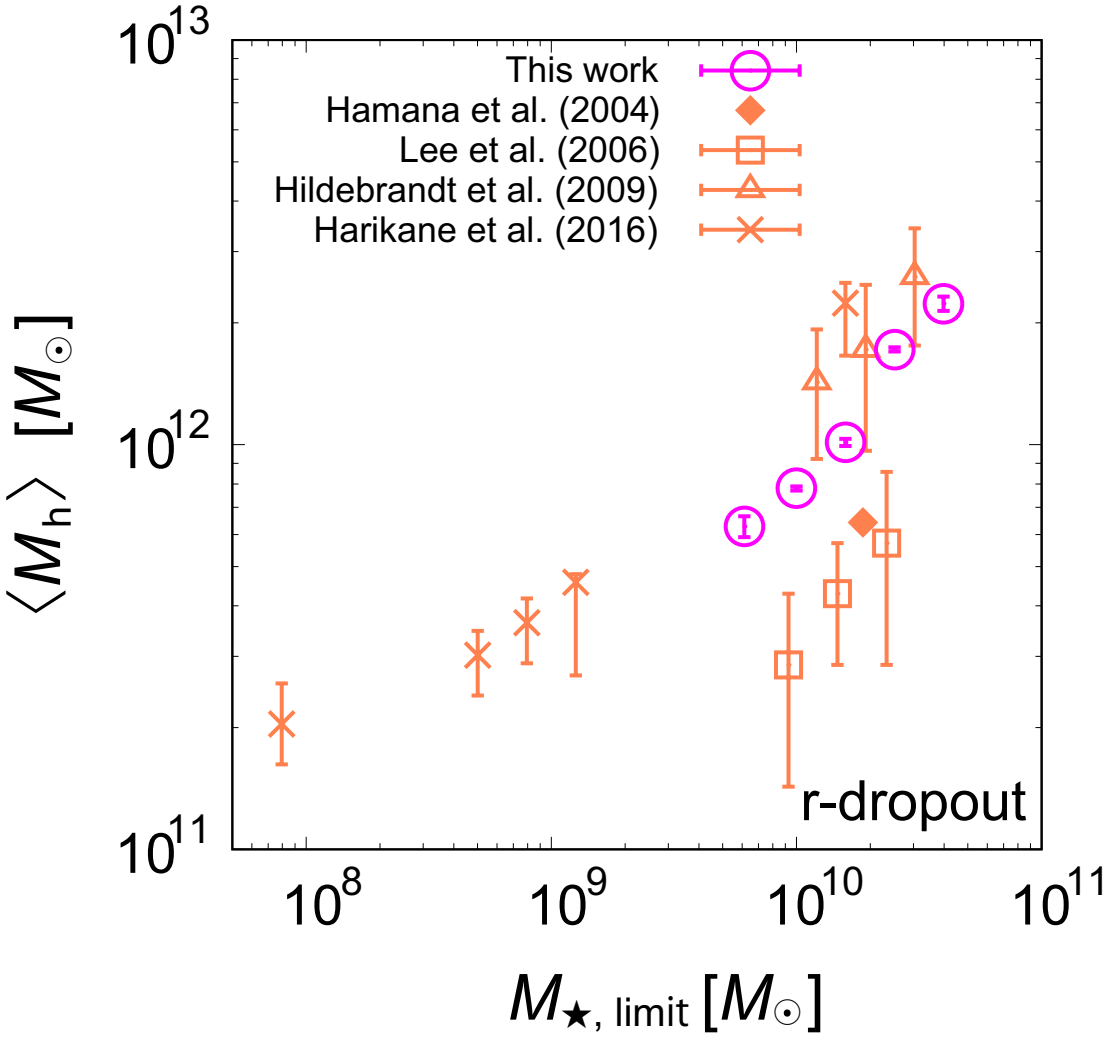

Figure 66.— Same as Figure 64, but for r-dropout galaxies.

5.5.3. Satellite fractions

In Figure 67, we show the satellite fractions as a function of the stellar-mass threshold of each dropout galaxy. As expected, the satellite fractions decrease gradually with increasing galaxy stellar mass and then seem to drop sharply by increasing the stellar-mass limit at each redshift. The turning stellar mass of the satellite fraction corresponds approximately to the turnover stellar-mass limit of M_1.

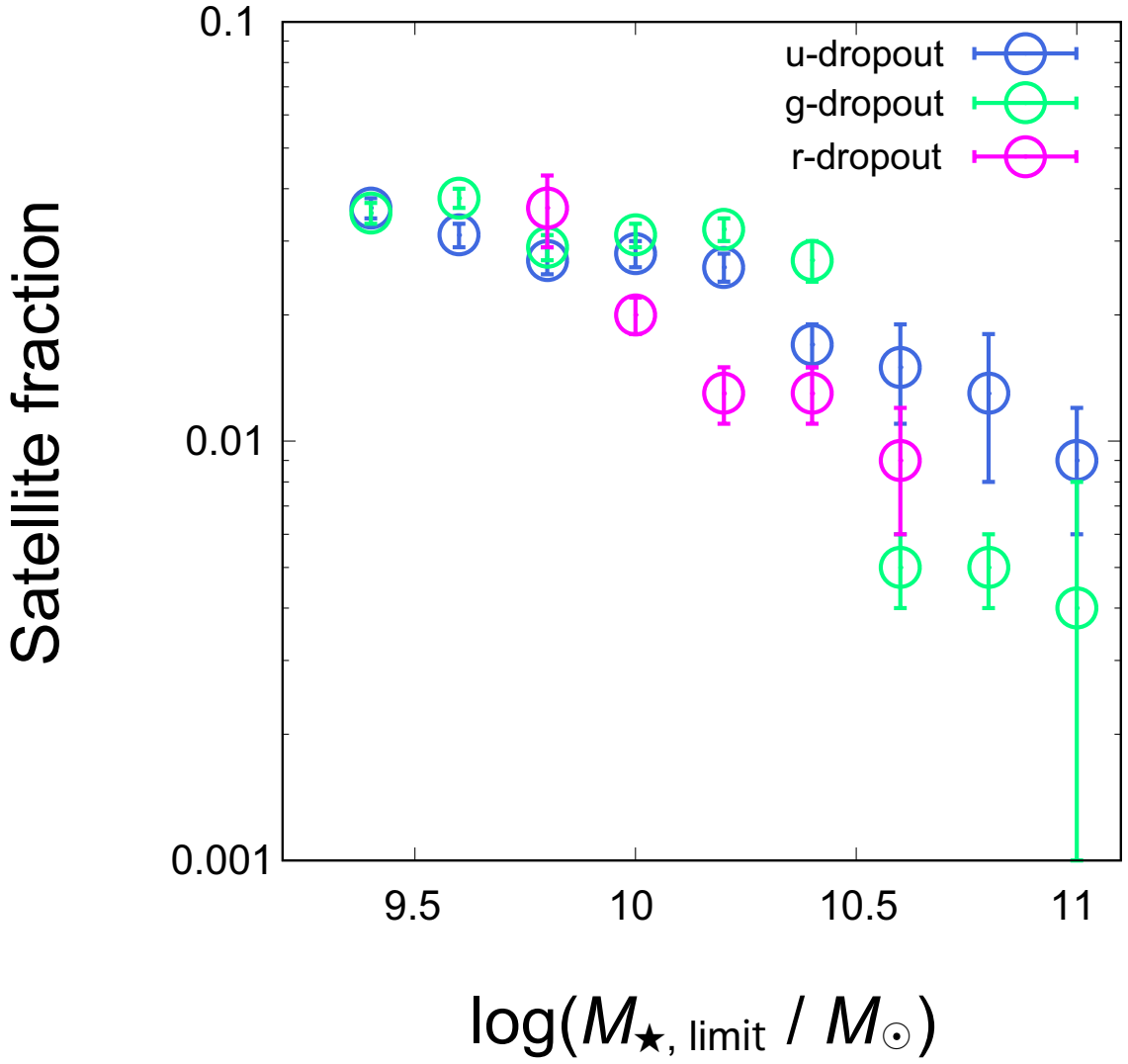

Figure 67.— Evolution of satellite fractions of u-, g-, and r-dropout galaxies as a function of stellar-mass limit.

Our results show that the satellite fractions are quite low, even for less massive galaxies ($f_{\rm sat} \sim 0.04$ for $M_\star = 10^{9.4} M_\odot$); this is completely different from the results of the $0 < z < 2$ Universe (e.g., Wake et al. 2011; Zehavi et al. 2011; Coupon et al. 2012; Martinez-Manso et al. 2015; McCracken et al. 2015). Wake et al. (2011) investigated the satellite fraction of galaxies at $1 < z < 2$ collected by the SED fitting method and showed that the satellite fraction decreases with increasing the stellar mass; however, the values of $f_{\rm sat}$ have a large gap compared to our results ($f_{\rm sat} = 0.21$ for galaxy

samples with $M_\star = 3.0 \times 10^{10} M_\odot$ at $\bar{z} = 1.9$). Martinez-Manso et al. (2015) carried out HOD analyses for the *Spitzer*/IRAC detected galaxies at $z \sim 1.5$ using the *Spitzer* South Pole Telescope Deep-Field Survey (SSDF; Ashby et al. 2013) and showed that the satellite fraction of their faintest magnitude-limit samples ($[4.5]_{\rm limit} = 18.6$ mag with a median stellar mass of $M_\star = 10^{9.87} M_\odot$) was $f_{\rm sat} = 0.23$. They compared the satellite fractions of Zehavi et al. (2011), who implemented HOD analyses using numerous local SDSS galaxies, and inferred that the satellite fraction grows slowly from $f_{\rm sat} \sim 0.2$ at $z \sim 2$ to $f_{\rm sat} \sim 0.3$ at $z \sim 0$.

Low-z studies collected all galaxies, whereas our high-z studies can only select star-forming galaxies with high star-formation rates. Thus, we may have missed quiescent high-z galaxies with different baryonic natures. Our results indicate that there are few satellite galaxies even for less massive galaxies at high-z; the satellite fraction showed little evolution at $z = 3 - 5$ and significant evolution from $z \sim 3$ to $z \sim 2$. Accurate satellite fractions in high-z Universe require extensive and complete high-z galaxy survey in the future.

5.6. Stellar-to-Halo Mass Ratios of High-Redshift Galaxies

In this section, I compare SHMRs of our dropout galaxy samples with previous theoretical and observational studies. First, I check the consistency of the SHMRs calculated by different stellar-mass estimation methods, by the MS relation and the SED fitting technique, and then compared our results with literature and discuss the redshift evolution. For the comparison, I use the threshold stellar masses of subsamples as $M_\star$ and $M_{\rm min}$ as $M_{\rm h}$ to compute the SHMRs of dropout galaxies.

5.6.1. SHMRs from main sequence vs. SED fitting

In Figure 68 − Figure 70, I compare the SHMRs calculated by the MS relation and the SED fitting method of u- and g-dropout galaxies. The SED fitting, which cannot be applied reliably to faint samples, i.e., less massive galaxies, has a higher lowest-mass bin than the MS relation, and also shows a relatively large scatter than those from the MS relations due to the small survey volume where NIR data are available. The results indicate that the stellar-mass estimated by SED fitting and the MS relation are not significantly different on the measurement of SHMRs. They are consistent within 1σ error over almost the entire mass range. It is noted that our SED fitting includes the prior of the MS relation (equation 108), which is the same relation used to estimate the stellar masses from the main sequence. Hereafter, I use the SHMRs derived from the MS relations, which can be applied for wide stellar-mass ranges with high S/N ratios based upon wide-field data. I also assume that the consistency is valid for r-dropout galaxies, the stellar masses of which cannot be estimated from SED fitting.

5.6.2. Comparison with literature

In Figure 68, our observed SHMRs are compared with the results of Moster et al. (2013), a semi-analytical model by Somerville et al. (2015) (GK model; Gnedin & Kravtsov 2010), and Behroozi et

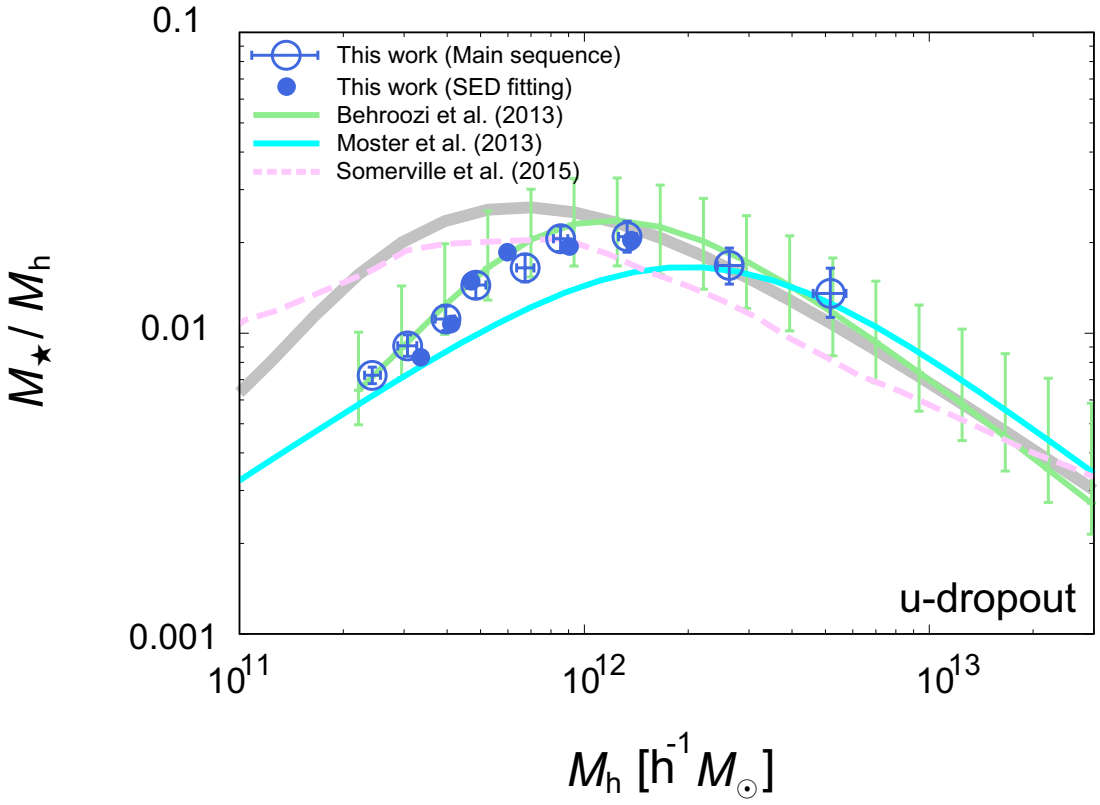

Figure 68.— Comparison of our SHMRs of u-dropout galaxies with literature. The open circles with error bars are our SHMRs computed from the stellar masses by assuming the relation of the main sequence of star-forming galaxies, whereas filled circles are evaluated via an SED fitting technique. Solid lines are model predictions from the abundance-matching method (green with error bars: Behroozi et al. 2013b at $z = 3$, grey: Behroozi et al. 2013b at $z = 0.1$, and cyan: Moster et al. 2013) and the result of the semi-analytical model by Somerville et al. (2015).

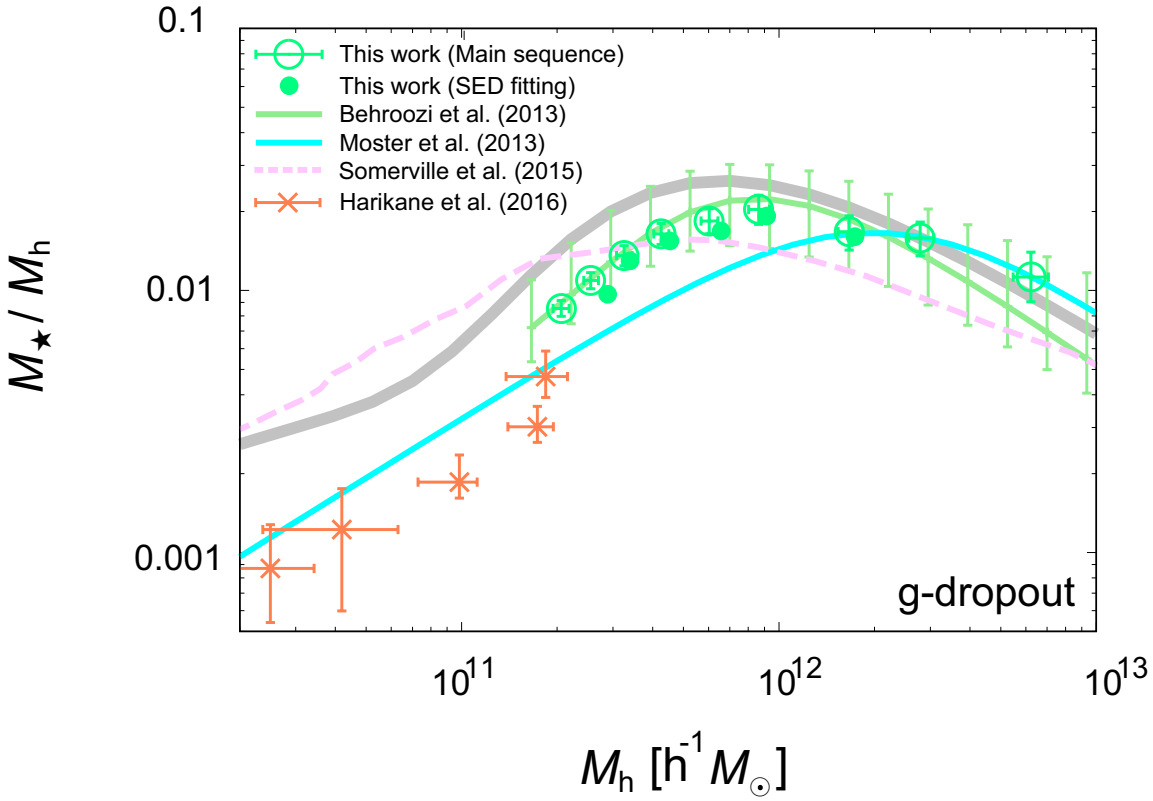

Figure 69.— Same as Figure 68, but for g-dropout galaxies.

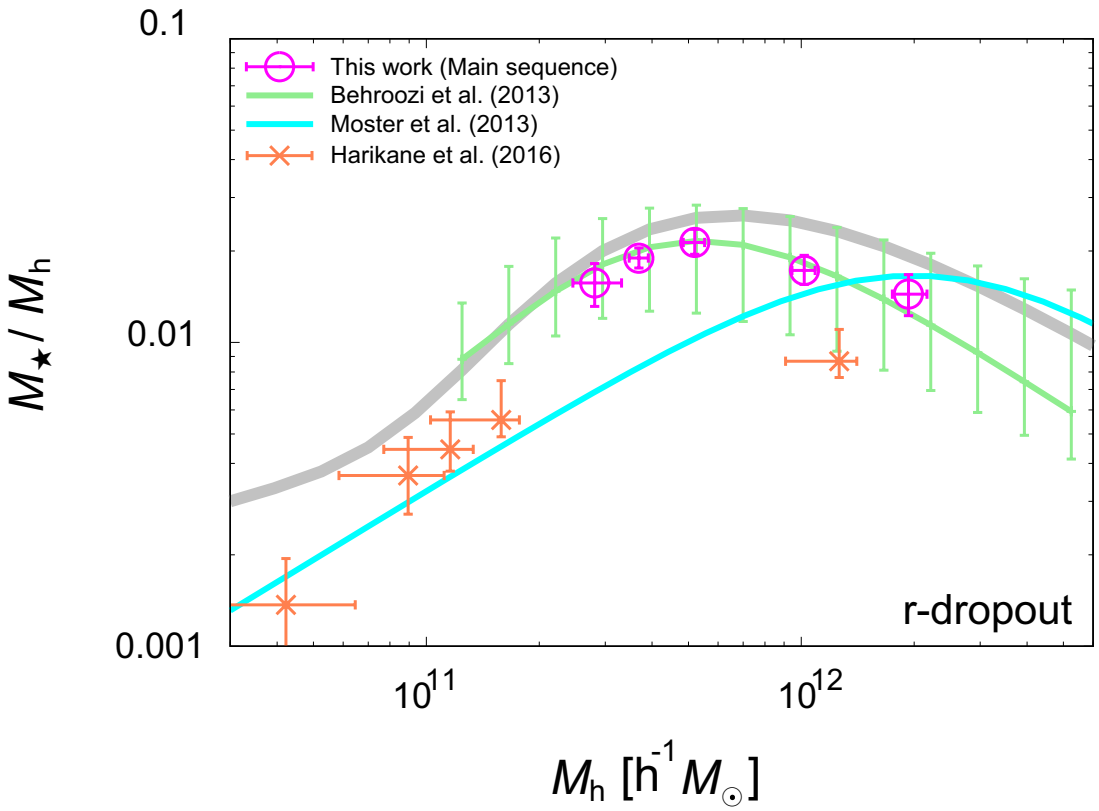

Figure 70.— Same as Figure 68, but for r-dropout galaxies.

al. (2013b) calculated by numerical simulations and observational data from Harikane et al. (2016). The assignment method of central galaxies with host dark haloes of Moster et al. (2013) and Behroozi et al. (2013b) are the abundance-matching method and Harikane et al. (2016) is the HOD formalism. Moster et al. (2013) and Behroozi et al. (2013b) presented their own functional forms of SHMRs as a function of the halo mass and those relations take into account redshift evolutions by being calibrated with the stellar-mass functions at each redshift. I plot the predicted SHMRs of each author at $z = 3.0$, 4.0, and 5.0 in Figure 68 − Figure 70.

Our results show an excellent agreement with the SHMRs of Behroozi et al. (2013b) in all redshift bins, indicating that there is little difference between the HOD technique and the abundance-matching method as a galaxy–halo connection method, at least for the high-redshift Universe. In addition to the constraints on the stellar-mass functions, Behroozi et al. (2013b) also introduced observational constraints, such as cosmic SFR and specific SFRs (sSFRs), which are essential parameters to characterize high-z galaxies. For $z = 3$ and 4, Moster et al. (2013) well reproduces our observational results only for the massive end but underestimates from the less massive end to intermediate halo mass ($\langle M_{\rm h}\rangle < 10^{12.0}h^{-1}M_{\odot}$ for $M_{\star,\rm limit} \sim 10^{10.4}M_{\odot}$). This discrepancy may be caused by that Moster et al. (2013) could not constrain the SHMRs for less massive galaxies by stellar-mass functions, whereas Behroozi et al. (2013b) could reproduce it up to $M_{\star} \sim 10^{9.4}M_{\odot}$, which is comparable to our stellar-mass limit. The functional form of Moster et al. (2013) cannot describe our observational results at $z = 5$ because their adopted stellar-mass functions (Pérez-González et al. 2008; Santini et al. 2012) do not reach $z = 5$. The $M_{\rm h}^{\rm pivot}$ of Somerville et al. (2015), on the other hand, show lower halo masses compared to observation. This may be due to the overproduction of less massive galaxies in their model. It is noted that our SHMRs are inconsistent with the theoretical prediction of Yang et al. (2012), who calculated the SHMR based upon the conditional luminosity function (CLF; Yang et al. 2003; van den Bosch et al. 2003b).

Harikane et al. (2016) observed the SHMRs for very less massive dropout galaxies at $z = 4$ and 5. Their results are consistent with our results, though their estimations show slightly lower SHMR than ours. This may be due to differences in the employed cosmological parameters, or because they introduced a fixed duty-cycle parameter of star-formation in their halo occupation model (equation 66). The UV-faint dropout galaxies with low star-forming activities may be missed in their dropout galaxy selection criterion or appear to be "quenched" by their star-forming phase. By introducing the fixed duty cycle, the expected number of the dropout galaxies (or total amount of stellar masses) for fixed halo masses change, possible affecting the values of SHMR and $M_{\rm min}$.

5.6.3. Evolution of the SHMRs and the pivot halo masses

Due to the availability of deep and wide field imaging data of the CFHTLS, our sample can, for the first time, trace the evolution of the pivot halo mass at $3 < z < 5$. We fit our SHMRs using the parameterized form proposed by Behroozi et al. (2010) as follows:

$$\log(M_{\star}(M)) = \log(\epsilon M_1) + f\left(\log\left(\frac{M}{M_1}\right)\right) - f(0), \tag{110}$$

where $f(x)$ satisfies the following relationship:

$$f(x) = -\log(10^{\alpha x} + 1) + \delta\frac{(\log(1 + \exp(x)))^{\gamma}}{1 + \exp(10^{-1})}. \tag{111}$$

In this parameterized function, M_1 and ϵ represent the pivot halo mass and the SHMR at the pivot halo mass, respectively. First, we investigated the best-fit parameters for high-quality u-dropout galaxies, and then derived the best-fit M_1 and ϵ for g- and r-dropout galaxies with the fixed same α, δ, and γ, which control the power-law slope at the less massive and massive ends as u-dropout galaxies. The $M_{\rm h}^{\rm pivot}$ and SHMRs at $M_{\rm h}^{\rm pivot}$ of each dropout galaxy are summarized in Table 12.

Table 12: Best-fit values of the pivot Halo mass and the SHMRs at pivot halo mass

	$\log(M_{\rm h}^{\rm pivot}/h^{-1}M_\odot)$	$M_\star/M_{\rm h}^{\rm pivot}(\times10^{-2})$
u-dropout	12.08 ± 0.029	1.922
g-dropout	11.94 ± 0.033	2.041
r-dropout	11.74 ± 0.070	2.023

Figure 71 shows the SHMRs and best-fit functions of each redshift. At a fixed halo mass, our SHMRs show negative evolution for galaxies within less massive dark haloes ($\langle M_{\rm h}\rangle \lesssim 10^{12.0}h^{-1}M_\odot$) and positive evolution for massive haloes with decreasing redshift. The arrows above the SHMRs indicate the best-fit $M_{\rm h}^{\rm pivot}$ of each dropout sample. First of all, $M_{\rm h}^{\rm pivot}$ at $z = 3-5$ are measured to be $\sim 10^{12}h^{-1}M_\odot$, which is comparable to local Universe. The $M_{\rm h}^{\rm pivot}$ slightly increases but the SHMRs at $M_{\rm h}^{\rm pivot}$ are not significantly changed at $z = 3-5$, which is consistent with the observational results of LBGs at $z = 4-7$ (Stefanon et al. 2016). The efficiency of the star formation changes very little from $z = 5$ to $z = 0$, which is consistent with time-independent star-formation efficiency model (Behroozi et al. 2013b). Fakhouri et al. (2010) reported the mean mass accretion rates as $\langle \dot{M}_{\rm halo}\rangle_{\rm mean} \propto (\langle M_{\rm h}\rangle/10^{12}M_\odot)\times(1+z)^{2.5}$ at $z \gtrsim 1$ based upon the results of the cosmological N-body simulations (Springel et al. 2005; Boylan-Kolchin et al. 2009). Dropout galaxies show active star-forming activities that is comparable to the above rapid growth of dark halo masses in high-redshift Universe. Combined with the low satellite fractions, these highly star-forming activities may not be governed by galaxy mergers.

Here, $M_{\rm h}^{\rm pivot}$ shows the small positive evolution with cosmic time, which is opposite to previous lower-z studies at $z = 0-1$ (e.g., Coupon et al. 2012; Leauthaud et al. 2012). In Figure 72, I compare the $M_{\rm h}^{\rm pivot}$ with previous HOD studies and a model prediction of the pivot halo mass evolution by Behroozi et al. (2010) calculated by equation (110) and (111). I show observationally, for the first time, the increasing trend of $M_{\rm h}^{\rm pivot}$ with cosmic time at $z > 3$; this $M_{\rm h}^{\rm pivot}$ evolution is well described by the model prediction using the abundance-matching technique. $M_{\rm h}^{\rm pivot}$ evolves from $M_{\rm h}^{\rm pivot} = 10^{11.74}h^{-1}M_\odot$ at $z \sim 5$ to $M_{\rm h}^{\rm pivot} = 10^{12.08}h^{-1}M_\odot$ at $z \sim 3$, which can be described as $\log(M_{\rm h}^{\rm pivot}/M_\odot) \propto -0.22\times(1+z)$, and little evolution of SHMRs at $M_{\rm h}^{\rm pivot}$ indicates that stellar-mass growth is comparable to halo mass growth in high-z Universe. This idea is in agreement with the observational results of the rapid stellar-mass evolution from $z > 3$, when intense halo assembly is expected, toward the era of the cosmic noon (e.g., Hopkins & Beacom 2006; Suzuki et al. 2015). In contrast, at $z < 2$, previous HOD studies show an increasing trend of $M_{\rm h}^{\rm pivot}$ with increasing redshift. This trend is evidence for galaxy downsizing, i.e., galaxy star-formation is taking place within less massive haloes with cosmic time from $z \sim 2$.

More interestingly, $M_{\rm h}^{\rm pivot}$, which is the halo mass with the most efficient star-formation activity, is almost unchanged around $\log(M_{\rm h}^{\rm pivot}/M_\odot) = 12.1 \pm 0.2$ by comparing with the local values as well

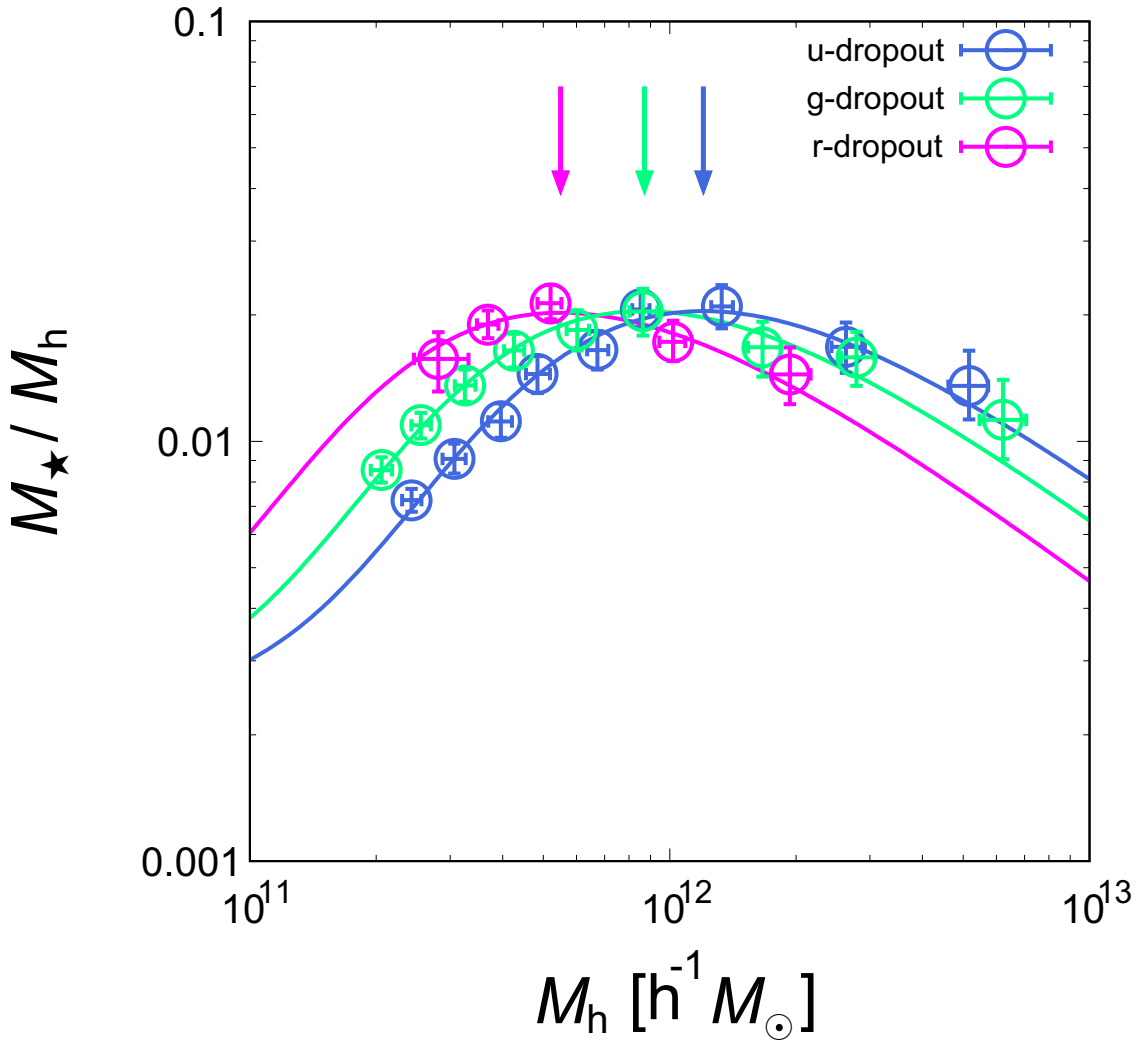

Figure 71.— SHMRs of u- (blue), g- (green), and r-dropout galaxies (red). Solid lines are the best-fit functions of SHMRs fitted by the parameterized function proposed by Behroozi et al. (2010). Arrows above the SHMRs indicate the pivot halo masses of each dropout galaxy.

as the model prediction (Behroozi et al. 2013b) over cosmic time at $0 < z < 5$. This suggests that the peak in the star-formation efficiency and its normalization are almost time independent, while environment of $10^{12} M_{\odot}$ haloes has significantly evolves since $z = 5$. This has been theoretically predicted by considering of the shock heating, the supernova feedback, the gas cooling, and the gas accretion (e.g., Rees & Ostriker 1977; White & Rees 1978; Dekel & Birnboim 2006). Our study shows an observational evidence that the pivot halo mass around $10^{12} M_{\odot}$ represents the characteristic dark halo mass for galaxy formation.

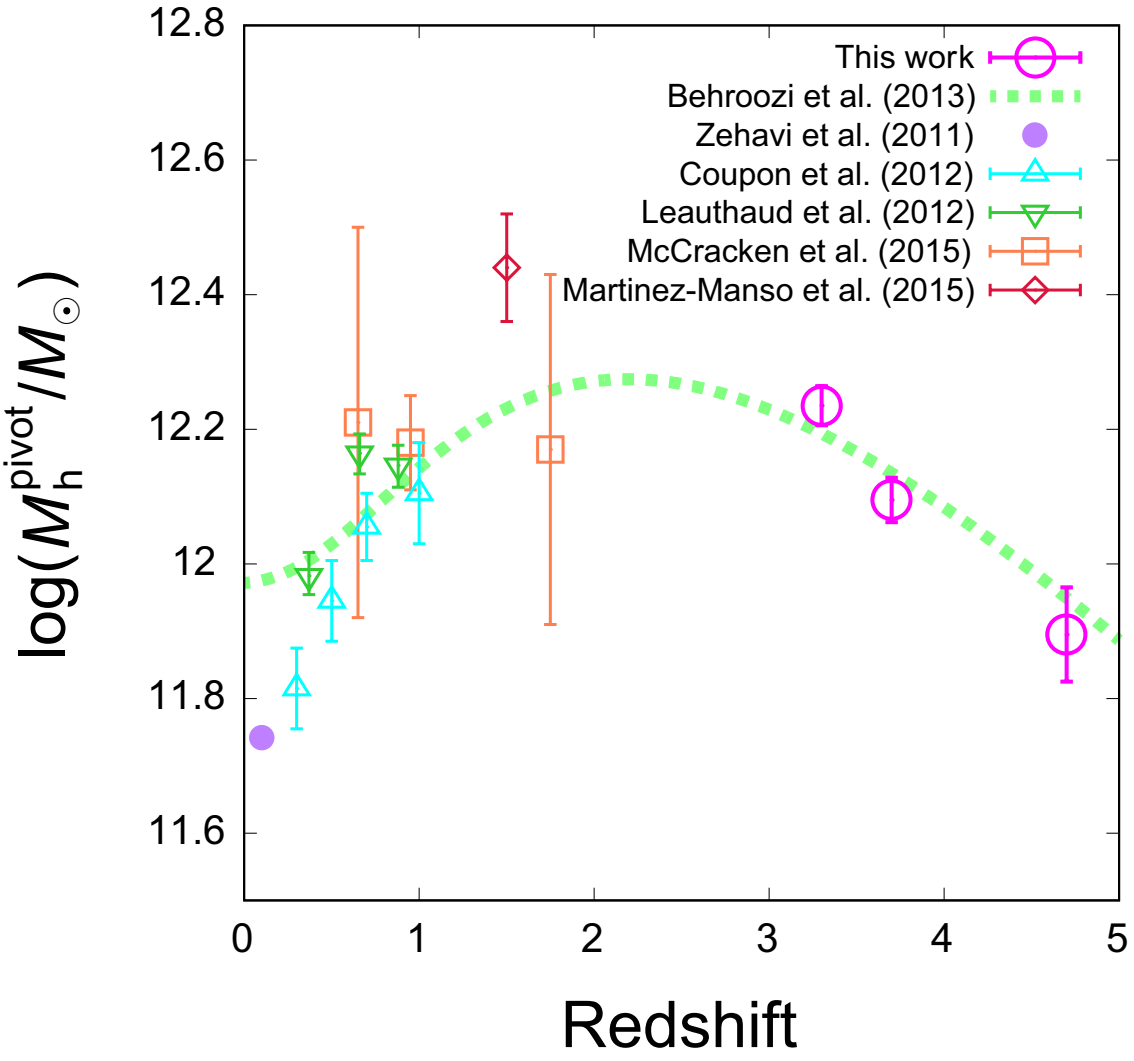

Figure 72.— Pivot halo masses calculated by HOD analyses as a function of redshift. The dashed green line indicates model prediction of the pivot halo mass evolution calculated by the equation (110) and (111) (Behroozi et al. 2010). Pivot halo mass is given in units of $M_{\odot}$ on logarithmic scale by assuming $h = 0.7$.

6. DISCUSSION

In this section, I discuss the relationship between galaxies and their host dark haloes across cosmic history. First, I reanalyze sgzK galaxies by adopting the *WMAP* seven-year cosmologies and dividing sgzK galaxies into subsamples according to their stellar masses to achieve fairly comparison with results of other redshift ranges.

6.1. Reanalysis of sgzK Galaxies

6.1.1. Clustering and HOD analysis

I presented the detailed clustering properties and the relation between sgzK galaxies and their host dark haloes in Section 4. The analysis was performed by assuming the simple flat ΛCDM cosmologies ($\Omega_{\rm m} = 0.30$ and $\Omega_\Lambda = 0.70$) and sgzK galaxies were binned according to their K-band magnitude; however, analyses for low-z galaxies (Section 3) and high-z galaxies (Section 5) were carried out by assuming the *WMAP* seven-year cosmologies ($\Omega_{\rm m} = 0.27$ and $\Omega_\Lambda = 0.73$) and galaxies were binned by their stellar masses. Stellar masses of each sgzK galaxies have already been estimated by their $(z-K)$ colors and the K-band magnitudes using the equation (101). Thus, I reanalyze the sgzK galaxies by adjusting the condition with analyses for low/high-z galaxies.

The stellar mass distribution of sgzK galaxies is shown in Figure 73. The stellar-mass limit is determined by comparing the stellar-mass function of our sgzK galaxies with that derived using the COSMOS photometric redshift catalogue (Muzzin et al. 2013). I set the lower stellar mass as $\log(M_{\star,\rm limit}/M_\odot) = 10.4$, which satisfies 70% selection completeness. It is noted that pgzK galaxies cannot be analyzed due to the low selection completeness.

The sgzK galaxies are divided into subsamples according to their stellar masses with a bin size of 0.2 dex. The number of sample galaxies are listed in Table 13. Clustering analyses are carried out in the same manner as the original analysis: I use the ACF estimator of Landy & Szalay (1993), the integral constraint is corrected, and the errors of ACFs are evaluated by the "delete-one" jackknife resampling method with dividing survey fiend into 64 sub-fields. Derived ACFs are presented in Figure 74.

The HOD analysis is also implemented. I assume the halo mass function of Sheth & Tormen (1999), the large-scale halo bias of Tinker et al. (2010), the halo radial density profile of Navarro et al. (1997), and the halo occupation function of Zheng et al. (2005) with the halo exclusion model. Redshift distributions are evaluated using the COSMOS photometric redshift catalogue (Muzzin et al. 2013). In this analysis, I vary all of the HOD free parameters to obtain fine fittings. The results of the HOD fitting are plotted over the observed ACFs in Figure 73, and the best-fit parameters and the deduced parameters, i.e., mean halo masses and satellite fractions, are given in Table 13.

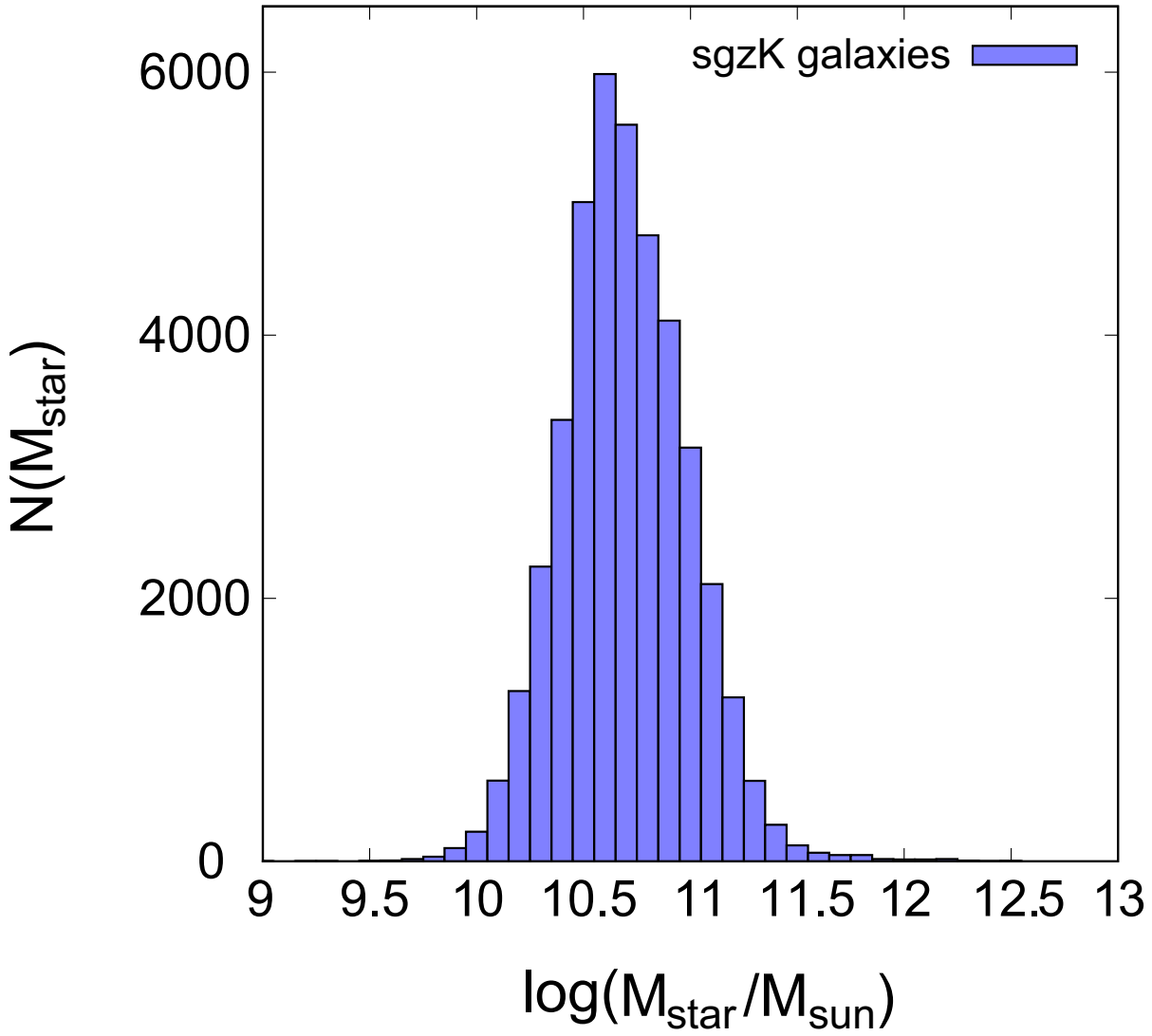

Figure 73.— The stellar mass distribution of sgzK galaxies which satisfy $K \leq 23.0$. Stellar masses are evaluated by the $(z - K)$ color and the K-band magnitude of each sgzK galaxy.

Table 13: The number of the sgzK galaxy samples and the best-fit HOD parameters of each subsample limited by the stellar mass

$\log(M_{\star,\mathrm{limit}}/M_\odot)$	N	$\log(M_{\mathrm{min}}/h^{-1}M_\odot)$	$\log(M_1/h^{-1}M_\odot)$	$\sigma_{\log M}$	α	$\log(M_{\mathrm{h}}/h^{-1}M_\odot)$	f_{sat}	χ^2/dof
10.4	35,026	$12.08^{+0.02}_{-0.03}$	$13.31^{+0.03}_{-0.02}$	$0.16^{+0.06}_{-0.04}$	$1.14^{+0.02}_{-0.03}$	12.60 ± 0.01	0.09 ± 0.01	1.47
10.6	25,124	$12.31^{+0.03}_{-0.03}$	$13.35^{+0.03}_{-0.04}$	$0.43^{+0.06}_{-0.06}$	$1.26^{+0.03}_{-0.05}$	12.65 ± 0.02	0.08 ± 0.01	1.78
10.8	14,098	$12.68^{+0.05}_{-0.06}$	$13.60^{+0.05}_{-0.04}$	$0.72^{+0.06}_{-0.06}$	$1.11^{+0.08}_{-0.10}$	12.67 ± 0.02	0.07 ± 0.01	2.23
11.0	6,030	$13.10^{+0.07}_{-0.07}$	$14.08^{+0.10}_{-0.08}$	$0.78^{+0.08}_{-0.09}$	$1.12^{+0.09}_{-0.12}$	12.79 ± 0.06	0.03 ± 0.01	2.58
11.2	1,771	$13.60^{+0.09}_{-0.10}$	$14.48^{+0.13}_{-0.12}$	$0.92^{+0.08}_{-0.10}$	$1.18^{+0.11}_{-0.14}$	12.87 ± 0.07	$0.01^{+0.01}_{-0.00}$	3.11

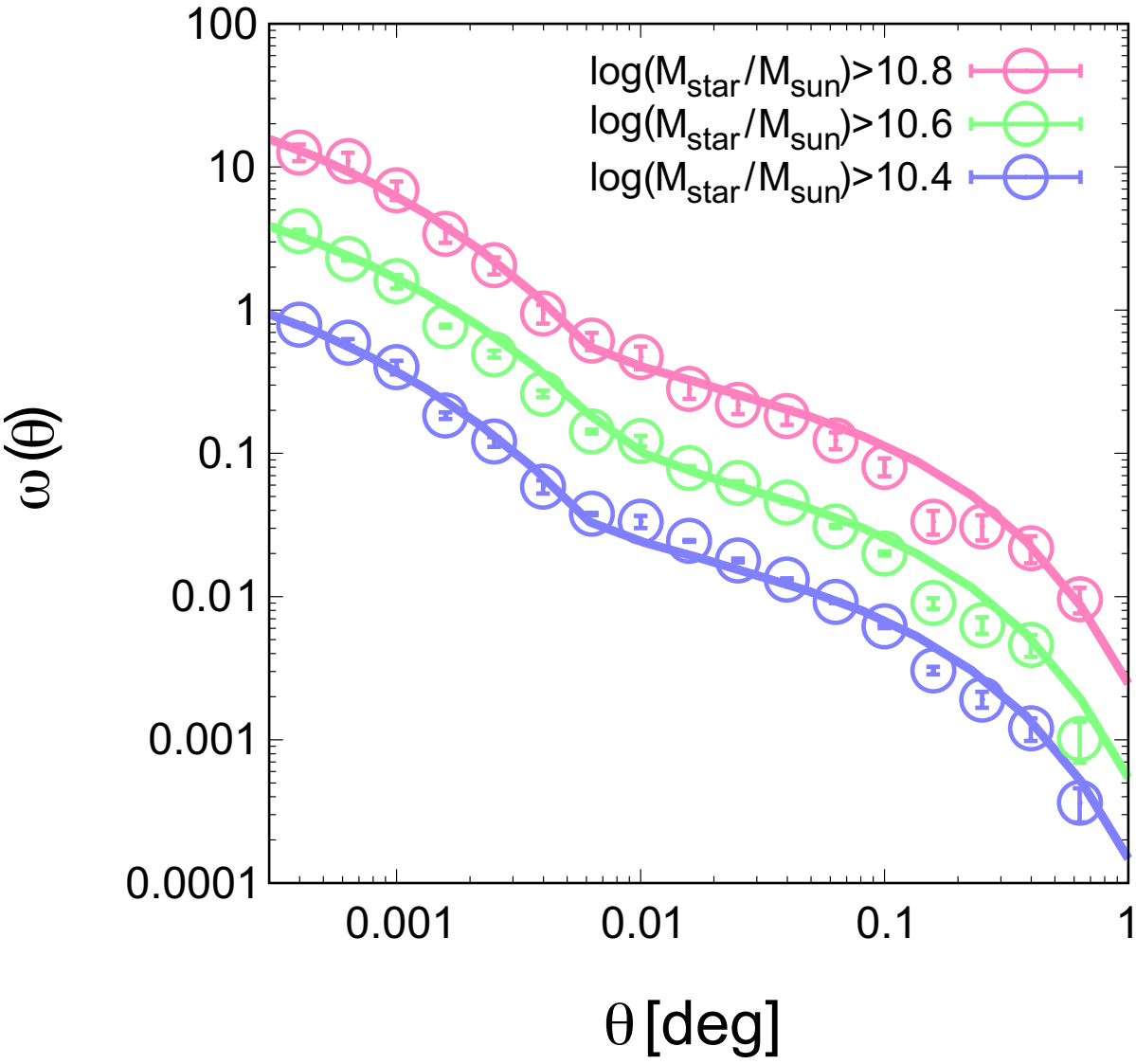

Figure 74.— ACFs and the best-fit HOD fittings of sgzK galaxies as a function of separation angle scales in degree. Errors are estimated by the jackknife resampling method. Solid lines are the best-fit ACFs calculated by the HOD formalist of each stellar-mass bin. The y-axix is scaled arbitrarily to clarify the results.

6.1.2. Stellar-to-halo mass ratio

I recalculate the SHMRs of sgzK galaxies and the results are shown in Figure 75. For comparison, model predictions calculated by the numerical simulation of Behroozi et al. (2013b) and Moster et al. (2013), and by the semi-analytical model of Somerville et al. (2015) are also plotted. The covering halo-mass range differs from Figure 51 due to the difference of the definition of the halo mass; I used mean halo masses, $\langle M_{\rm h} \rangle$, in Figure 51, whereas characteristic halo masses, $M_{\rm min}$, are adopted in

Figure 75. As is the case in Figure 51, the SHMRs of less massive bins show smaller values compared to the results of Behroozi et al. (2013b); however, intermediate to massive bins are well consistent with the results of the abundance-matching method (Behroozi et al. 2013b; Moster et al. 2013) as well as the semi-analytical model (Somerville et al. 2015).

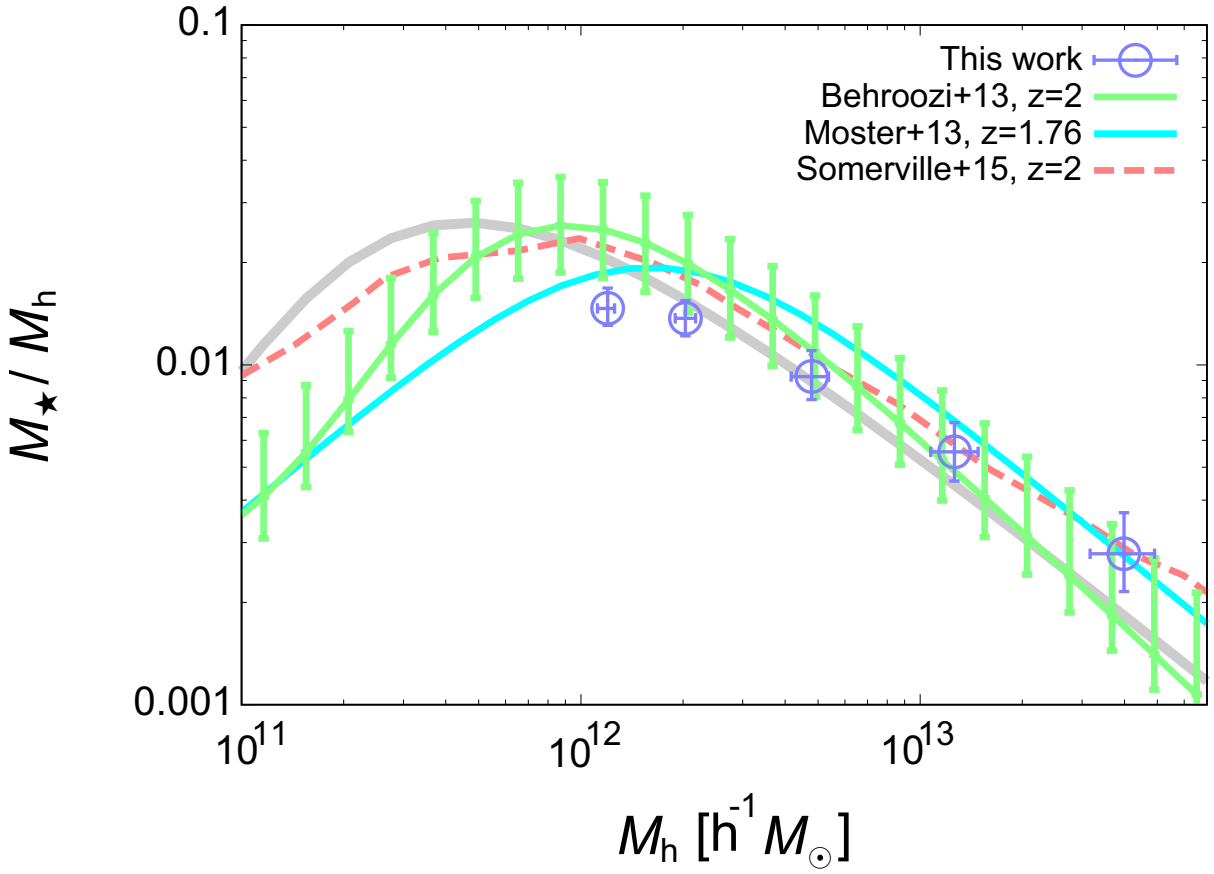

Figure 75.— The recalculated SHMRs of the sgzK galaxies (blue circles with error bars). For comparison, I also plot the model predictions by the numerical simulations of Behroozi et al. (2013b, solid green line) and Moster et al. (2013, solid cyan line), and by the semi-analytical model of Somerville et al. (2015, dashed orange line), respectively. The solid grey line is the SHMR at $z = 0.1$ presented by Behroozi et al. (2013b).

6.2. Redshift Evolution of Satellite Fraction

By the accurate HOD analysis across cosmic time, redshift evolution of the satellite fraction can be traced. I present satellite fractions of each galaxy sample as a function of the stellar-mass threshold in Figure 76. One should keep in mind that the selection criterion of our galaxy samples are not unified; I select total galaxies for low-z samples ($0.30 < z_{\rm phot} < 1.40$) according to their i-band magnitudes, and star-forming galaxies for mid-z samples ($1.4 \lesssim z \lesssim 2.5$) by their stellar masses magnitudes and for high-z samples ($z \gtrsim 2.5$) by their star-formation rates. At all redshifts, satellite fractions are plateau for less massive galaxies and gradually decrease with increasing the stellar-mass threshold

from intermediate stellar-mass threshold. This characteristic is also seen in previous HOD studies (e.g., Z. Zheng et al. 2007; Zehavi et al. 2011; Wake et al. 2011; Coupon et al. 2012; McCracken et al. 2015), inferring that massive galaxies are likely to be central galaxies of intermediate-massive dark haloes rather than to be satellite galaxies within massive dark haloes, irrespective of these redshifts.

As discussed in Section 3.4.2, our satellite fractions of low-z samples are almost consistent with the results of photo-z samples of McCracken et al. (2015) and show small evolution from $z_{\rm phot} \sim 1.40$ to $z_{\rm phot} \sim 0.30$. Coupon et al. (2012) also reported the small evolution of satellite fractions at $z = 0-1$ and found a consistency with the evolutionary trend of satellite fractions estimated by the galaxy–galaxy lensing technique (Mandelbaum et al. 2006).

At $z \sim 2$, satellite fractions of sgzK galaxies are comparable with those at $1.5 < z_{\rm phot} < 2.0$ by McCracken et al. (2015). According to the redshift distribution of total sgzK galaxies (bottom-right panel of Figure 36), most of the sgzK galaxies are selected at $z < 2$; thus, it is plausible that satellite fractions have comparable values between sgzK galaxies and galaxies confined to be $1.5 < z_{\rm phot} < 2.0$.

From the comparison among redshifts, satellite fractions show a large gap between $z \lesssim 2$ and $z > 3$ and those at $2.0 < z_{\rm phot} < 2.5$ from McCracken et al. (2015) locate between them, implying that the fraction of satellite galaxies drastically evolve at $z = 3 - 2$ and $z = 2 - 1$. Jose et al. (2013), who investigated the clustering of LBGs at $z = 3-5$ using the semi-analytical model, reported that $\sim 40\%$ of dark haloes heavier than $\sim 10^{12} M_\odot$ contain central LBGs at $z = 3 - 5$. Fakhouri et al. (2010) investigated halo merger rates using the results of the Millennium Simulation and found that dark haloes with mass $M_{\rm h} > 10^{12} M_\odot$ at $z = 2$ are expected to experience the halo merger from $z = 3$ to $z = 2$, and the halo merger rate from $z = 2$ to $z = 1$ is still high for dark haloes with $M_{\rm h} \sim 10^{12-13} M_\odot$ at $z = 1$. Therefore, the evolution of the satellite fraction at $z = 3 - 2$ and $z = 2 - 1$ seems to be largely supported by the halo mergers.

However, it is necessary for more precise clustering analyses to discuss the evolution of satellite fraction in high-z Universe since our satellite fractions of high-z galaxies are suffered from relatively large errors. Detailed evolutionary scenario of the satellite fractions will be presented when clustering and HOD analyses on the sgzK galaxies and LBGs using the HSC SSP data will be completed. Moreover, galaxy population should be unified to understand the accurate evolutionary history of satellite galaxies; in other words, satellite fractions of total galaxies, i.e., including passively evolving galaxies and galaxies with low star-formation rates, in high-z Universe are required. It will be achieved by extensive surveys using next-generation telescopes such as WFIRST.

6.3. Redshift Evolution of Stellar-to-Halo Mass Ratio

I show the redshift evolution of our observatinal SHMRs in Figure 77. To determine the slopes and the $M_{\rm h}^{\rm pivot}$ of SHMRs at each redshift, I fit observational SHMRs by the functional form (equation 110 and 111) of the SHMR proposed by Behroozi et al. (2010). Behroozi et al. (2013b) presented redshift-dependent forms of each free parameters with 1σ limits:

$$\log(\epsilon) = -1.777^{+0.133}_{-0.146} + \left(-0.006^{+0.113}_{-0.361}\,(a-1) + \left(-0.000^{+0.003}_{-0.104}\right) z\right)\nu - 0.119^{+0.061}_{-0.012}\,(a-1)\,, \quad (112)$$

$$\log(M_1) = 11.514^{+0.053}_{-0.009} + \left(-1.793^{+0.315}_{-0.330}\,(a-1) + \left(-0.251^{+0.012}_{-0.125}\right) z\right)\nu, \quad (113)$$

$$\alpha = -1.412^{+0.020}_{-0.105} + \left(0.731^{+0.344}_{-0.296}\,(a-1)\right)\nu, \quad (114)$$

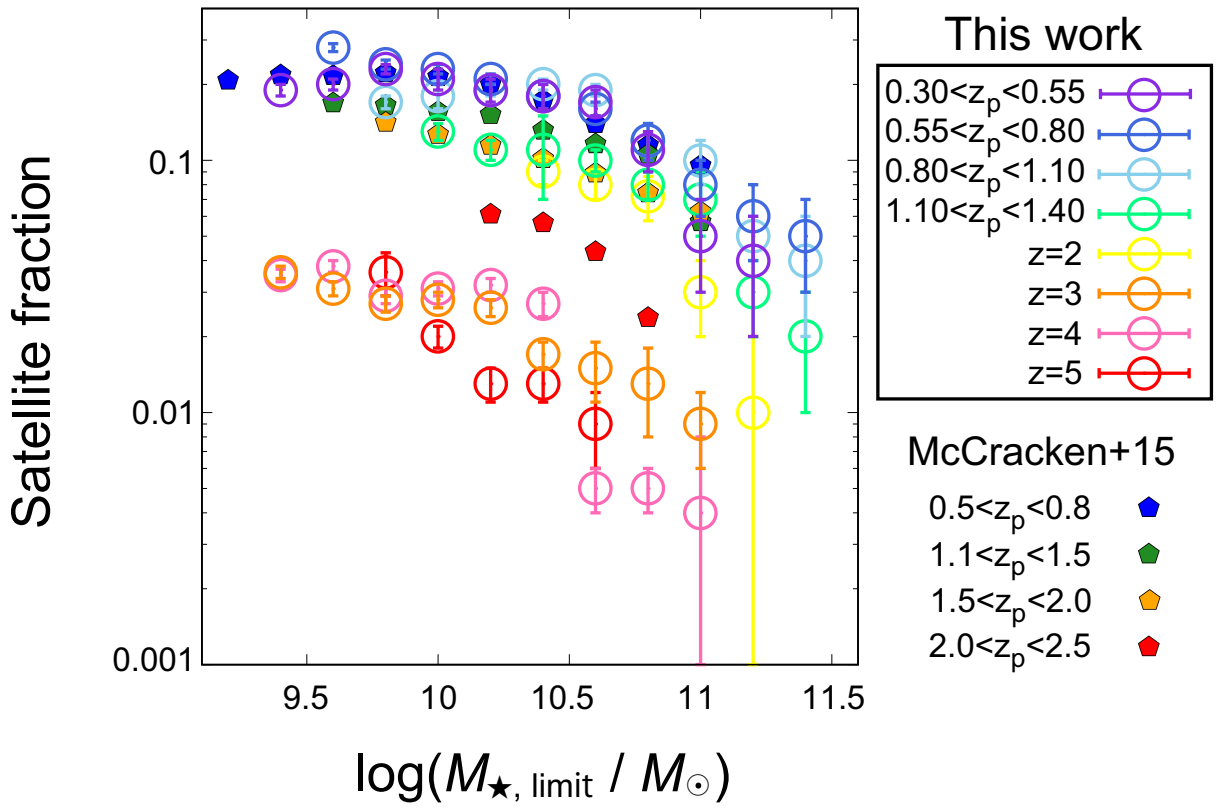

Figure 76.— Redshift evolution of the satellite fraction as a function of samples stellar-mass threshold. I select total galaxy samples by photometric redshifts ($0.30 < z_{\rm phot} < 1.40$), sgzK galaxies ($1.4 \lesssim z \lesssim 2.5$), and the Lyman break galaxies ($z \gtrsim 3$), respectively. Satellite fractions derived by McCracken et al. (2015) are also plotted for comparison.

$$\delta = 3.508^{+0.087}_{-0.369} + \left(2.608^{+2.446}_{-1.261}\,(a-1) + \left(-0.043^{+0.958}_{-0.071}\right) z\right) \nu, \tag{115}$$

and

$$\gamma = 0.316^{+0.076}_{-0.012} + \left(1.319^{+0.584}_{-0.505}\,(a-1) + \left(+0.279^{+0.256}_{-0.081}\right) z\right) \nu, \tag{116}$$

where a is a scale factor, i.e., $a = 1/(1+z)$, and ν satisfies $\nu = \exp\left(-4a^2\right)$. I constrain these SHMR parameters at each redshift by the MCMC simulation: the number of steps is $10,000$ times, the burn-in phase is set $1,000$ times, and the initial parameters are values of $\epsilon(z)$, $M_1(z)$, $\alpha(z)$, $\delta(z)$, and $\gamma(z)$ at each redshift. It should be noted that SHMRs at $z \sim 2$ (sgzK galaxies) and $z \sim 5$ (r-dropout galaxies) are fitted by fixing the slope of the faint end, α, as their initial states due to the small number of observational data points.

The best-fit SHMR relations of each redshift are shown in Figure 78. I also plot the redshift evolution of $M_{\rm h}^{\rm pivot}$ and the comparison with literature in Figure 79. The redshift evolution of our observational $M_{\rm h}^{\rm pivot}$ shows good agreement with the theoretical prediction by Behroozi et al. (2013b) at all redshifts. $M_{\rm h}^{\rm pivot}$ shows apparent redshift evolution: $M_{\rm h}^{\rm pivot}$ is the lowest halo mass at $z \sim 5$ and shifts toward the higher halo mass with increasing cosmic time up to $z = 2-3$, and then decrease with decreasing redshift from $z = 2-3$ to $z \sim 0$. The highest value of $M_{\rm h}^{\rm pivot}$ is at $z = 2-3$, which is known the era with the highest star-formation efficiency in cosmic history (e.g., Hopkins & Beacom 2006). The large error of $M_{\rm h}^{\rm pivot}$ at $z \sim 2$ is originated from the lack of the SHMRs in less-massive end: our $M_{\rm h}^{\rm pivot}$ at $z \sim 2$ is evaluated only by fitting the massive end of the SHMRs. Our measurements are consistent with values of literature at all redshifts with small errors except for the results of Martinez-Manso et al. (2015). Martinez-Manso et al. (2015) calculated the SHMR using the *Spitzer*/IRAC detected galaxies and the difference of $M_{\rm h}^{\rm pivot}$ may be due to the difference of galaxy populations.

The amplitude of SHMRs does not significantly change over cosmic time. The result at $z \sim 2$ is smaller than other redshifts; however, it is unreliable since $M_{\rm h}^{\rm pivot}$ cannot be captured by our observation. More precise SHMRs at $z \sim 2$, i.e., the low-mass end of sgzK galaxies and by separating galaxies into sgzK galaxies and pgzK galaxies, should be calculated using the data of the HSC SSP Deep/Ultradeep layers. On the other hand, two redshift bins in low-z Universe ($0.80 < z_{\rm phot} < 1.10$ and $1.10 < z_{\rm phot} < 1.40$) show higher values compared to other redshifts. These bins also show the deviation of massive-end slopes from the model prediction of Behroozi et al. (2013b); I should check whether these deviations (amplitudes of the SHMR and massive-end slopes) are real using the latest photometric-redshift catalogues. The redshift evolution of both slopes at less-massive and massive ends also should be discussed in future.

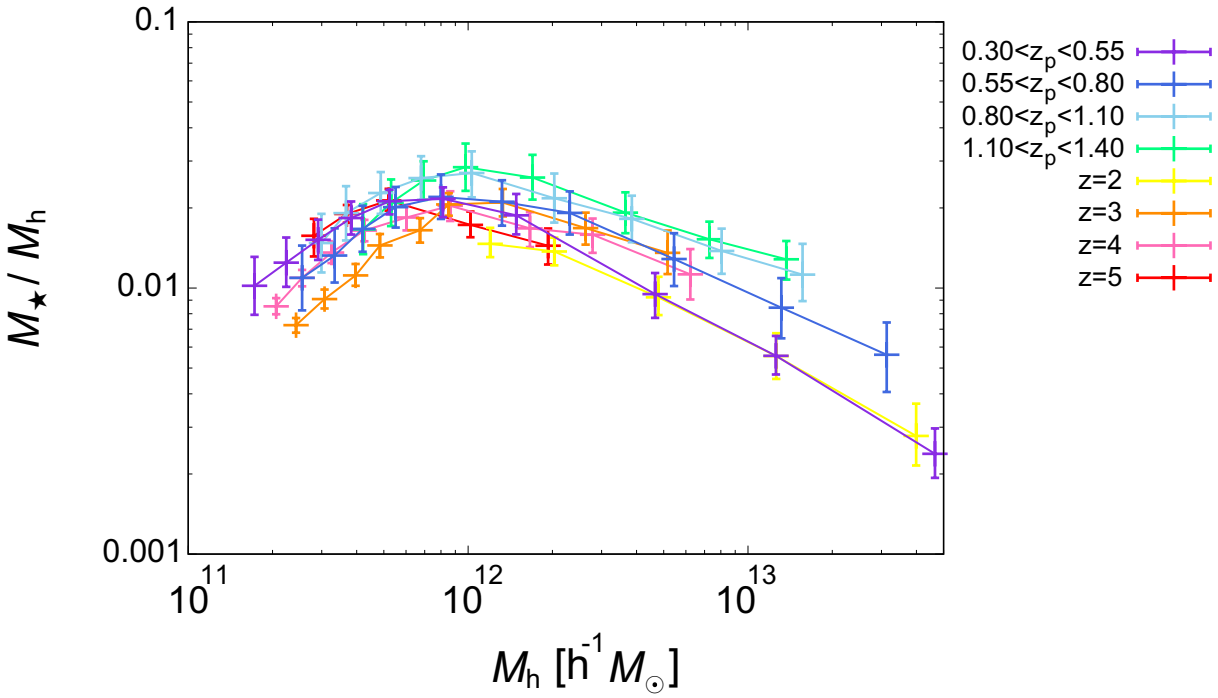

Figure 77.— Redshift evolution of the observational SHMRs with 1σ errors. Galaxy samples are selected total galaxies according to their i-band magnitudes ($0.30 < z_{\rm phot} < 1.40$) and star-forming galaxies according to their stellar masses ($z \sim 2$) and the star-formation rates ($z \gtrsim 3$), respectively.

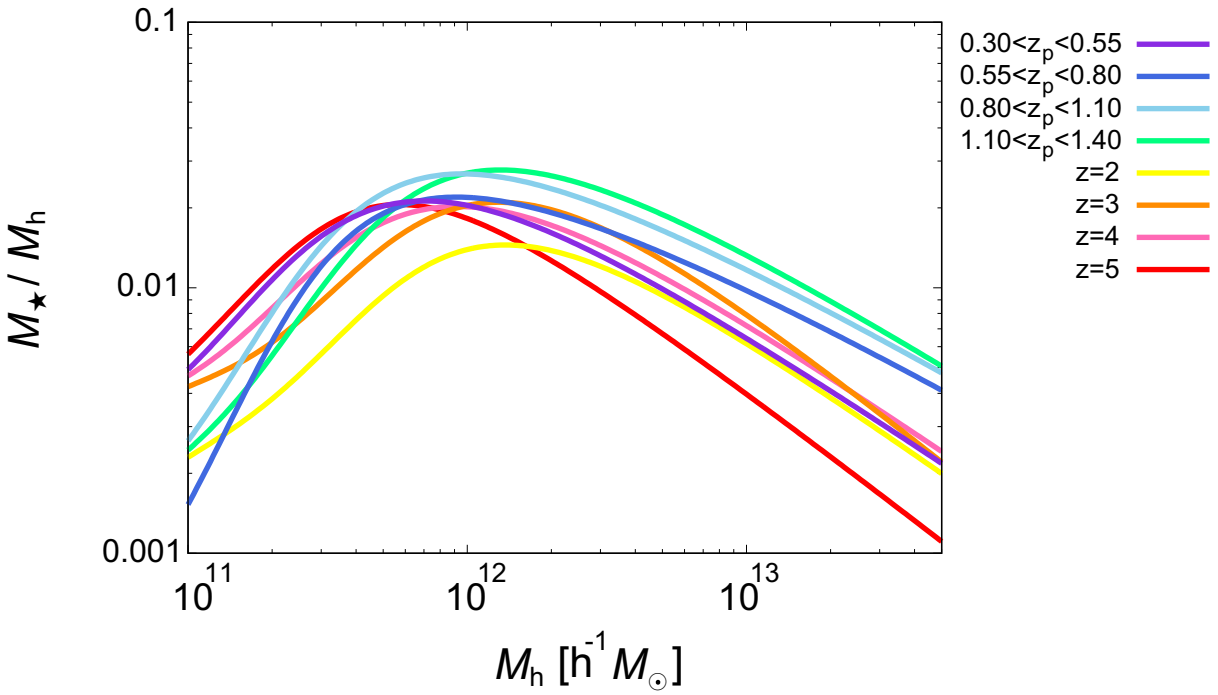

Figure 78.— Redshift evolution of the best-fit SHMRs fitted by the functional form (equation 110 and 111) presented by Behroozi et al. (2010, 2013b).

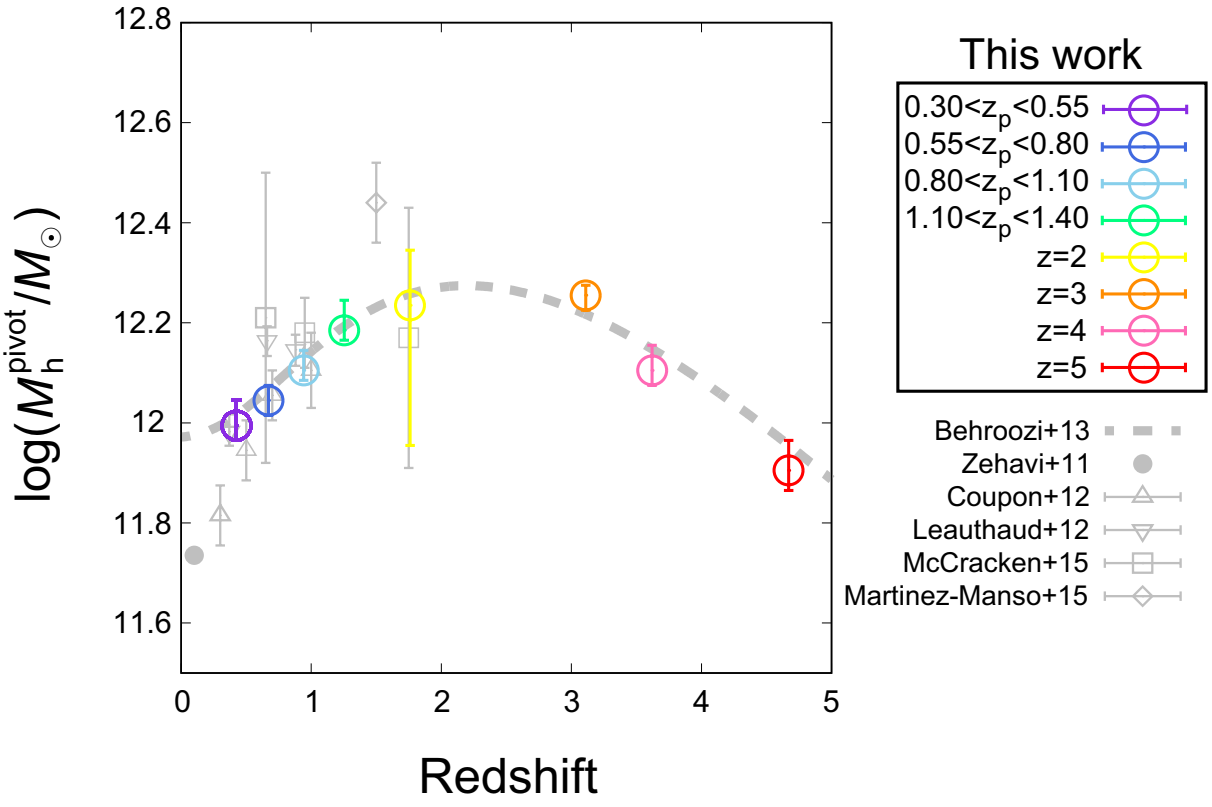

Figure 79.— Evolution of $M_{\rm h}^{\rm pivot}$ as a function of redshift. For comparison, I also plot the observational results of Zehavi et al. (2011), Coupon et al. (2012), Leauthaud et al. (2012), McCracken et al. (2015), and Martinez-Manso et al. (2015), and the theoretical prediction by Behroozi et al. (2013b). Redshift evolution of $M_{\rm h}^{\rm pivot}$ of Behroozi et al. (2013b) is calculated via the equation (110) and (111). $M_{\rm h}^{\rm pivot}$ are given in units of $M_\odot$ on logarithmic scale by assuming $h = 0.7$.

7. CONCLUSIONS AND FUTURE PROSPECTS

7.1. Summary and Conclusions

In this thesis, I present the results of precision clustering analyses to link observed galaxies to their underlying invisible host dark haloes across the global cosmic history from $z = 5.4$ to $z = 0.3$. Thanks to the wide survey fields, I achieve to construct the largest-ever galaxy samples at each redshift and obtain high-quality two-point angular auto-correlation functions (ACFs) with quite high S/N ratios. To connect galaxy samples to their host haloes, I apply the halo occupation distribution (HOD) formalism to observed ACFs and derive the properties of dark haloes such as the mean halo masses and the satellite fractions. In addition to evaluate the properties of dark haloes, I also investigate the baryonic properties of galaxies, i.e., galaxy stellar masses, star-formation rates, and photometric redshifts, to discuss galaxy evolutionary history by linking the galaxy properties of baryons as well as dark matter. Evaluation of both galaxy stellar masses and dark halo masses enables to compute the stellar-to-halo mass ratios (SHMRs), which describes the star-formation efficiency within dark haloes as a function of dark halo mass. This thesis firstly attempts to trace the evolution of SHMRs over cosmic time by observations and derive the evolutionary history of the pivot halo mass, $M_{\rm h}^{\rm pivot}$, which represents the most efficient halo mass at each redshift.

The major conclusions of this thesis are summarized as follows.

In low-z Universe ($0.3 < z < 1.4$):

1. I precisely derive ACFs of low-z galaxies using the photometric-redshift catalogue (Tanaka 2015) of HSC SSP data. Numerous galaxy samples enable to obtain ACFs with small sizes of redshift and stellar-mass bins, which are essential to trace the redshift evolution and investigate the stellar-mass dependence of clustering properties. Clustering strengths increase with increasing the stellar-mass limits and highly accurate ACFs are well represented by the HOD model. HOD analyses reveal that massive haloes host massive stellar-mass galaxies, which are consistent with previous HOD studies of low-z galaxies (e.g., Zehavi et al. 2011; Wake et al. 2011; Coupon et al. 2012; McCracken et al. 2015). By fixing the stellar-mass limit, mean halo masses increase with decreasing redshift, showing evidence of mass evolution of dark haloes by halo merging.

2. Satellite fractions show small redshift evolution from $z \sim 1.4$ to $z \sim 0.3$, which is partly due to the evolution of halo mass functions with cosmic time. Our satellite fractions are almost consistent with the observational results of McCracken et al. (2015) as well as the results of Skibba et al. (2015).

3. I calculate SHMRs for galaxies at $0.3 < z < 1.4$ with wide stellar-mass range. Our SHMRs are almost consistent with the model prediction of Behroozi et al. (2013b) within 1σ confidence levels, especially for less-massive haloes of $\langle M_{\rm h} \rangle < M_{\rm h}^{\rm pivot}$. This is due to our high-accuracy HOD analysis based on the large number of galaxy samples selected by the HSC SSP survey. Besides the model prediction presented by Behroozi et al. (2013b), our SHMRs are not conflict to other theoretical models such as the abundance-matching method by Moster et al. (2013) and the semi-analytical model by Somerville et al. (2015), as well as the observational result by Coupon et al. (2015). Nevertheless, our SHMRs show apparent excess at the massive ends except for the SHMR at $0.30 < z < 0.55$. The origin of these excess is still unclear; however,

one possible explanation is the overprediction of stellar masses in the SED fitting procedure. I also succeed in determining $M_{\rm h}^{\rm pivot}$ at $0.3 < z < 1.4$ and $M_{\rm h}^{\rm pivot}$ shows increasing trend with increasing redshift.

In mid-z Universe ($1.4 < z < 2.5$):

1. Based on wide-field observation, I collect 41,112 sgzK galaxies, which is approximately twice the number of sBzK galaxies reported by McCracken et al. (2010). The large number of sgzK galaxies make it possible to resample both into cumulative and differential luminosity subsamples to investigate the dependence of sgzK clustering properties on the K-band magnitude.

2. Correlation lengths of sgzKs are in good agreement with the results of previous studies, except for Hartley et al. (2008). Error bars of our measurements are quite small because of the large galaxy samples. I also find the correlation between the K-band luminosities and the correlation lengths, implying that luminous sgzKs reside in massive dark haloes.

3. Observed ACFs of sgzKs are well described by the halo model based on the HOD formalism. The HOD mass parameters, $M_{\rm min}$, M_1, M_0, and $\langle M_{\rm h} \rangle$, almost all increased monotonically as a function of the magnitude threshold, suggesting that more-luminous sgzK galaxies reside in more-massive dark haloes. I find that $M_{\rm min}$ and M_1 for faint sgzK galaxies are approximately proportional to the luminosity, whereas bright sgzK galaxies follow $L \propto M_{\rm h}^{0.5}$, as is the case for galaxies in the local Universe. This is because baryons in massive dark haloes tend to form faint satellite galaxies, which are below the luminosity threshold, rather than accrete central galaxies; this confirms the same trend at $z \sim 2$. Our measurements of $M_{\rm min}$ are larger than in the local Universe (Zehavi et al. 2011) over all magnitudes, implying that galaxies are formed only in massive dark haloes at $z \sim 2$, and that less-massive dark haloes form more slowly, known as downsizing. Additionally, M_1 is also larger than that of (Zehavi et al. 2011), suggesting that satellite galaxies are less likely to be formed at $z \sim 2$ than in the local Universe.

4. I investigate the relationship between sgzK galaxies and other galaxy populations at low/high-z Universe by tracing the halo mass evolution using the extended Press–Schechter (EPS) formalism. Faint LBGs ($i' < 27.0$) at $z \sim 4$ are expected to evolve into faint sgzK galaxies ($22.0 < K < 23.0$), whereas bright LBGs ($i' < 26.0$) are expected to evolve into intermediate-luminous sgzK galaxies ($21.0 < K < 22.0$). The brightest sgzK galaxies at $z \sim 2$ appear to have evolved from more luminous LBGs. Furthermore, by comparing with the results of HOD analyses in low-z, our analyses suggest that faint sgzK galaxies could evolve into Milky-Way-like galaxies or elliptical galaxies, intermediate luminosity sgzK galaxies could evolve into massive elliptical galaxies or the central galaxies of galaxy group-size haloes, and the brightest sgzK galaxies could evolve into the central galaxies of galaxy clusters; i.e., BCGs in the local Universe. These evolutionary scenario of sgzK galaxies is also supported by the analysis of the number of satellite galaxies.

In high-z Universe ($2.5 < z < 5.4$):

1. The clustering amplitude of each dropout galaxy depends on the galaxy stellar mass. I reveal that more-massive LBGs in stellar mass show stronger clustering amplitude, which is consistent

with the results of the previous LBG clustering study (Hildebrandt et al. 2009) and the results of our low-z study.

2. Characteristic halo masses possessing both central and satellite galaxies ($M_{\rm min}$ and M_1) show approximately linear growth at $z = 3-5$, and $M_{\rm min}$ follows the same increasing trend with stellar masses at low-z, whereas M_1 at $z = 3-5$ shows higher values than $z = 0.5-1.5$ over the entire stellar-mass range, suggesting that satellite galaxies are formed inefficiently within dark haloes at high-z. This implies that the high virial temperature and/or the strong feedback effect from the central galaxy in high-z dark haloes may prevent effective satellite galaxy formation. In addition, the M_1 increases more sharply than at low-z. This discrepancy of M_1 dependence on the stellar mass may be due to the short timescale to accrete galaxies for being satellite galaxies.

3. The mean halo masses calculated by the best-fit HOD model are $\langle M_{\rm h}\rangle = 10^{11.7-12.8}h^{-1}M_{\odot}$, which are consistent with previous HOD studies. The satellite fractions of dropout galaxies are less than 0.04 for our stellar-mass limit, indicating that satellite galaxies rarely form in high-redshift dark haloes. By comparing the satellite fractions at $z = 0-2$, it is expected a drastic evolution of the number of satellite galaxies from $z \sim 3$ to $z \sim 2$, despite the difference of selection criterion among galaxy populations (e.g., the galaxy stellar mass and the SFR).

4. The SHMRs are computed using two independent stellar mass estimation methods: the SED fitting technique and by assuming the main sequence of star-forming galaxies. I confirm that these two estimations agree with each other.

5. The observed SHMRs at $z = 3-5$ are in good agreement with the model prediction of Behroozi et al. (2013b). I derive observationally, for the first time, the pivot halo mass at $3 < z < 5$, which shows a slight increasing trend with cosmic time at $z > 3$. In contrast, the values of SHMRs for pivot halo masses show little evolution. This suggests that the mass growth rates of stellar components and dark haloes are comparable at $3 < z < 5$.

6. I find $M_{\rm h}^{\rm pivot}$ to be almost unchanged around $\log(M_{\rm h}^{\rm pivot}/M_{\odot}) = 12.1 \pm 0.2$ over cosmic time at $0 < z < 5$, suggesting that the peak in star-formation efficiency and its normalization are almost time-independent, while the environment of $10^{12}M_{\odot}$ haloes significantly evolves after $z = 5$. This has been theoretically predicted by considering of the shock heating, supernova feedback, gas cooling, and gas accretion. The pivot halo mass at $\langle M_{\rm h}\rangle \sim 10^{12}M_{\odot}$ can be, in a manner of speaking, the characteristic dark halo mass for galaxy formation.

Using the above results of clustering analyses at each epoch of the Universe, I discuss the evolutionary history of the relationship between galaxies and their host dark haloes across cosmic history. By comparing the satellite fraction of each redshift, I find the drastic evolution of the satellite fraction from $z \sim 3$ to $z \sim 2$ and from $z \sim 2$ to $z \sim 1$. At $z > 3$, the fraction of satellite galaxies are quite small even for considering less massive satellite galaxies and evolve approximately twice as much as high-z at $z \sim 2$, and the satellite fraction further increase at $z \sim 1$. This may be due to the increase of the massive halo abundance, the frequent halo merger, and highly star-formation activities at these epochs.

I achieve to trace the evolution of $M_{\rm h}^{\rm pivot}$ by observations and confirm that the evolutionary scenario of $M_{\rm h}^{\rm pivot}$ predicted by the numerical simulation (Behroozi et al. 2013b) is valid up to $z \sim 5$. At high redshift, star formations proceed within less massive haloes and the mass of the most efficient

haloes increase with increasing cosmic time by evolving the dark haloes, and the halo mass with the largest star-formation efficiency has a peak $z \sim 2$. After the era of the cosmic noon, galaxies within the most efficient haloes at $z > 2$ progress in quenching and the efficiently star-forming haloes shift toward the less massive haloes with decreasing redshift. Overall, $M_{\rm h}^{\rm pivot}$ and its normalization are found to be almost unchanged during $0 < z < 5$. My study throughout this thesis provides observational evidence that galaxy formation is ubiquitously most efficient near a halo mass of $\langle M_{\rm h} \rangle \sim 10^{12} M_{\odot}$ over cosmic time, which has been theoretically predicted by consideration of the shock heating, the supernova feedback, the gas cooling, and the gas accretion (e.g., Rees & Ostriker 1977; White & Rees 1978; Dekel & Birnboim 2006). Meanwhile, redshift evolution of the massive/less-massive slopes and the amplitude of SHMRs should be investigated in future using the latest photometric-redshift catalogues of the HSC SSP survey, whose accuracy of the estimation of the stellar masses and the photometric redshifts is much higher than the current catalogue that I used in this thesis.

7.2. Future Prospects

First of all, I will precisely investigate the evolution of the satellite fractions and the SHMRs using the updated photometric-redshift catalogue of the HSC SSP survey. By expanding the survey field and improving the SED fitting procedure, I will be able to carry out clustering analysis with higher accuracies than the results of this thesis. Increasing of massive galaxy samples enables to check whether the deviation of the SHMR slope at the massive end compared to Behroozi et al. (2013b) is real. In addition to the analysis of the HSC Wide layer, I will carry out clustering analyses using the data of the HSC Deep layer to extend our analysis up to $z \sim 3$ utilizing the extremely high-accuracy photometric redshifts based on deep optical and NIR data.

I also use data of the HSC Deep/Ultradeep layer to implement extremely deep gzK galaxy survey. sgzK galaxies selected in this thesis is down to $K < 23.0$ and $M_{\rm h}^{\rm pivot}$ at $z \sim 2$ only cannot be measured in the redshift range of this thesis due to the shallowness of the K-band magnitude. Moreover, deep photometric data of g-band allows to select pgzK galaxies, which cannot be investigated their clusterings in this thesis. By revealing the SHMRs of sgzK galaxies at low-mass end and the pgzK galaxies, I can discuss the star-formation within dark haloes even at the era of the "redshift desert" precisely. The gap of the satellite fraction between $z \sim 3$ and $z \sim 2$ also can be filled, which makes a better understanding of galaxy evolution at that epoch.

Furthermore, I plan to investigate the environmental dependence of the SHMRs. van Uitert et al. (2016) investigated the relation between the SHMRs and the local environment using the GAMA group catalogue and concluded that no significant difference can be seen between the different environment; however, their results are largely suffered from statistical errors due to the small sample size. Utilizing the wide survey area of the HSC Wide layer, more than 100 galaxy cluster candidates have already been identified at $z \sim 4$ (J., Toshikawa et al., in prep.) and the number of galaxy clusters and those redshift will drastically increase when the HSC SSP survey will be completed. I will perform clustering/HOD analyses by dividing galaxy samples according to their local densities and check whether the environmental dependence and its redshift evolution can be observed. This will be an indirect verification of the existence of the halo assembly bias.

A research to connect quasars to their host haloes is also in progress (S., Ishikawa & K., Ichikawa,

in prep.). I have selected quasar samples hosted by the star-forming galaxies at $z \sim 2$ (sBzK galaxies) using the COSMOS photometric-redshift catalogue (Muzzin et al. 2013) by matching the data of *WISE* All-Sky Data (Wright et al. 2010). ACFs of quasars hosted by the sBzK galaxies have already calculated; however, we cannot apply the same halo occupation model to the quasar clustering in interpreting the quasar occupation since "satellite quasars" hardly exists in the Universe. I will construct the halo occupation statistics for quasars and investigate the properties of quasar such as the SHMRs and the duty cycles of the quasars, as well as the properties of quasar host haloes. This research can be extended to the clustering analyses confining to Type I quasars using the next-generation spectrograph which will be mounted on the Prime Focus of the Subaru Telescope, PFS.

The halo occupation model should be verified. Ishikawa et al. (2017) pointed out the possibility that the standard HOD model cannot be applied for high-redshift galaxies; thus, I would like to check whether the current HOD model is valid for high-z galaxy samples and reconstruct the theoretical framework to link galaxies to the underlying dark matter, i.e., more sophisticated post HOD formalism.

In oder to gain insights into the structure formation scenario, it is necessary to put observational constraints for dark energy, as well as to investigate the co-evolution of dark matter and galaxies. I would like to constrain the properties of dark energy by detecting and tracing the redshift evolution of signals of the baryon acoustic oscillation (BAO) using the HSC SSP data. BAO provides the $100h^{-1}$Mpc scales at each redshift, which is a characteristic baryonic scale imprinted in the galaxy distribution; therefore, one can trace the expansion history of the Universe by tracing the acoustic peak that reflects the scale of the $100h^{-1}$Mpc at each redshift. The acoustic peak of the BAO has been detected in low-z Universe (i.e., $z \lesssim 0.4$) by the spectroscopic observations (e.g., Eisenstein et al. 2005; Anderson et al. 2014). Kazin et al. (2014) succeeded in detecting the acoustic peak in $z = 0.7$ Universe by the photometric survey (WiggleZ Dark Energy Survey; Drinkwater et al. 2010); however, both the scale of the acoustic peak and its redshift were uncertain because they selected galaxy samples using galaxy colors and depth of imaging data were shallow. I plan to detect the peaks of BAO up to $z = 1.4$ using the HSC SSP data. Deep imaging data of the HSC SSP survey provide accurate photometric redshifts and wide-field survey fields enable to calculate the auto correlations at large-angular scale, which is essential to capture the signal of the BAO. The final goal of my BAO study is to gain observational constraints for dark energy by investigating the expansion history of the Universe, and test the ΛCDM cosmology with highly precision by constraining the value of ω, which is a constant of the equation of state.

REFERENCES

Acquaviva, V., Gawiser, E., & Guaita, L. 2011, ApJ, 737, 47

Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45

Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697

Akerib, D. S., Araújo, H. M., Bai, X., et al. 2014, Physical Review Letters, 112, 091303

Álvarez-Márquez, J., Burgarella, D., Heinis, S., et al. 2016, A&A, 587, A122

Amblard, A., Cooray, A., Serra, P., et al. 2011, Nature, 470, 510

Anderson, L., Aubourg, É., Bailey, S., et al. 2014, MNRAS, 441, 24

Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540

Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, MNRAS, 329, 355

Arnouts, S., Walcher, C. J., Le Fèvre, O., et al. 2007, A&A, 476, 137

Artale, M. C., Pedrosa, S. E., Trayford, J. W., et al. 2016, arXiv:1611.05064

Ashby, M. L. N., Stanford, S. A., Brodwin, M., et al. 2013, ApJS, 209, 22

Baba, H., Yasuda, N., Ichikawa, S.-I., et al. 2002, Astronomical Data Analysis Software and Systems XI, 281, 298

Bacon, D. J., Refregier, A. R., & Ellis, R. S. 2000, MNRAS, 318, 625

Ball, N. M., Loveday, J., & Brunner, R. J. 2008, MNRAS, 383, 907

Barnes, J., & Hut, P. 1986, Nature, 324, 446

Barone-Nugent, R. L., Trenti, M., Wyithe, J. S. B., et al. 2014, ApJ, 793, 17

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

Baum, W. A. 1962, Problems of Extra-Galactic Research, 15, 390

Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 762, L31

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109

Benítez, N. 2000, ApJ, 536, 571

Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587

Berlind, A. A., Weinberg, D. H., Benson, A. J., et al. 2003, ApJ, 593, 1

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

Béthermin, M., Kilbinger, M., Daddi, E., et al. 2014, A&A, 567, A103

Bian, F., Fan, X., Jiang, L., et al. 2013, ApJ, 774, 28

Bielby, R. M., Finoguenov, A., Tanaka, M., et al. 2010, A&A, 523, A66

Bielby, R., Hudelot, P., McCracken, H. J., et al. 2012, A&A, 545, A23

Bielby, R. M., Gonzalez-Perez, V., McCracken, H. J., et al. 2014, A&A, 568, AA24

Birrer, S., Lilly, S., Amara, A., Paranjape, A., & Refregier, A. 2014, ApJ, 793, 12

Blanc, G. A., Lira, P., Barrientos, L. F., et al. 2008, ApJ, 681, 1099

Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, AJ, 129, 2562

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

Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476

Bond, J. R., Szalay, A. S., & Turner, M. S. 1982, Physical Review Letters, 48, 1636

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Boulade, O., Charlot, X., Abbon, P., et al. 2003, Proc. SPIE, 4841, 72

Bower, R. G. 1991, MNRAS, 248, 332

Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150

Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503-1513

Brescia, M., & Cavuoti, S. 2014, Astrophysics Source Code Library, ascl:1408.022

Brunner, R. J., Connolly, A. J., Szalay, A. S., & Bershady, M. A. 1997, ApJ, 482, L21

Bruzual A., G., & Charlot, S. 1993, ApJ, 405, 538

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559

Bullock, J. S., Wechsler, R. H., & Somerville, R. S. 2002, MNRAS, 329, 246

Bundy, K., Ellis, R. S., Conselice, C. J., et al. 2006, ApJ, 651, 120

Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, A&A, 554, A70

Calzetti, D. 1997, American Institute of Physics Conference Series, 408, 403

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Carpenter, J. M. 2001, AJ, 121, 2851

Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499

Casali, M., Adamson, A., Alves de Oliveira, C., et al. 2007, A&A, 467, 777

Chabrier, G. 2003, PASP, 115, 763

Challinor, A., & Lewis, A. 2011, Phys. Rev. D, 84, 043516

Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJ, 648, L109

Clowes, R. G., Harris, K. A., Raghunathan, S., et al. 2013, MNRAS, 429, 2910

Coil, A. L., Newman, J. A., Croton, D., et al. 2008, ApJ, 672, 153

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8

Cole, S., & Kaiser, N. 1989, MNRAS, 237, 1127

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393

Colín, P., Klypin, A. A., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 523, 32

Collister, A. A., & Lahav, O. 2004, PASP, 116, 345

Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, AJ, 110, 2655

Cool, R. J., Moustakas, J., Blanton, M. R., et al. 2013, ApJ, 767, 118

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

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201

Coupon, J., Kilbinger, M., McCracken, H. J., et al. 2012, A&A, 542, A5

Coupon, J., Arnouts, S., van Waerbeke, L., et al. 2015, MNRAS, 449, 1352

Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839

Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156

Davis, M., Geller, M. J., & Huchra, J. 1978, ApJ, 221, 1

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

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

Davis, M., Faber, S. M., Newman, J., et al. 2003, Proc. SPIE, 4834, 161

Dar, A. 1995, ApJ, 449, 550

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2

de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, ApJ, 302, L1

de la Torre, S., Le Fèvre, O., Arnouts, S., et al. 2007, A&A, 475, 443

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429

Durkalec, A., Le Fèvre, O., Pollo, A., et al. 2014, arXiv:1411.5688

Durkalec, A., Le Fèvre, O., Pollo, A., et al. 2015, A&A, 583, A128

Efron B. 1979, Ann. Statist., 7, 1

Einstein, A. 1911, Annalen der Physik, 340, 898

Einstein, A. 1916, Annalen der Physik, 354, 769

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

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560

Emerson, J., & Sutherland, W. 2010, The Messenger, 139, 2

Fang, G., Kong, X., Chen, Y., & Lin, X. 2012, ApJ, 751, 109

Fakhouri, O., Ma, C.-P., & Boylan-Kolchin, M. 2010, MNRAS, 406, 2267

Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, ApJ, 600, L107

Finlator, K., Davé, R., Papovich, C., & Hernquist, L. 2006, ApJ, 639, 672

Font, A. S., Bower, R. G., McCarthy, I. G., et al. 2008, MNRAS, 389, 1619

Foucaud, S., McCracken, H. J., Le Fèvre, O., et al. 2003, A&A, 409, 835

Foucaud, S., Conselice, C. J., Hartley, W. G., et al. 2010, MNRAS, 406, 147

Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., Santini, P. 2009, MNRAS, 397, 1776

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Gao, L., & White, S. D. M. 2007, MNRAS, 377, L5

Garilli, B., Guzzo, L., Scodeggio, M., et al. 2014, A&A, 562, A23

Geach, J. E., Sobral, D., Hickox, R. C., et al. 2012, MNRAS, 426, 679

Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

Giavalisco, M., & Dickinson, M. 2001, ApJ, 550, 177

Gnedin, N. Y., & Kravtsov, A. V. 2010, ApJ, 714, 287

Gott, J. R., III, Jurić, M., Schlegel, D., et al. 2005, ApJ, 624, 463

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Groth, E. J., & Peebles, P. J. E. 1977, ApJ, 217, 385

Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101

Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, A&A, 566, A108

Gwyn, S. D. J. 2012, AJ, 143, 38

Hamana, T., Yoshida, N., & Suto, Y. 2002, ApJ, 568, 455

Hamana, T., Ouchi, M., Shimasaku, K., Kayo, I., & Suto, Y. 2004, MNRAS, 347, 813

Hamana, T., Yamada, T., Ouchi, M., Iwata, I., & Kodama, T. 2006, MNRAS, 369, 1929

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

Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, ApJ, 821, 123

Hartley, W. G., Lane, K. P., Almaini, O., et al. 2008, MNRAS, 391, 1301

Hathi, N. P., Cohen, S. H., Ryan, R. E., Jr., et al. 2013, ApJ, 765, 88

Hayashi, M., Shimasaku, K., Motohara, K., et al. 2007, ApJ, 660, 72

Hearin, A. P., Zentner, A. R., van den Bosch, F. C., Campbell, D., & Tollerud, E. 2016, MNRAS, 460, 2552

Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146

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

Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, A&A, 498, 725

Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

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

Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, ApJS, 52, 89

Hudson, M. J., Gillis, B. R., Coupon, J., et al. 2015, MNRAS, 447, 298

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841

Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236

Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, ApJ, 709, 644

Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6

Inoue, A. K. 2011, MNRAS, 415, 2920

Ishikawa, S., Kashikawa, N., Toshikawa, J., & Onoue, M. 2015, MNRAS, 454, 205

Ishikawa, S., Kashikawa, N., Hamana, T., Toshikawa, J., & Onoue, M. 2016, MNRAS, 458, 747

Ishikawa, S., Kashikawa, N., Toshikawa, J., et al. 2017, ApJ, 841, 8

Ishiyama, T., Enoki, M., Kobayashi, M. A. R., et al. 2015, PASJ, 67, 61

Jackson, J. C. 1972, MNRAS, 156, 1P

Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, MNRAS, 321, 372

Jose, C., Subramanian, K., Srianand, R., & Samui, S. 2013, MNRAS, 429, 2333

Juneau, S., Glazebrook, K., Crampton, D., et al. 2005, ApJ, 619, L135

Kaiser, N. 1987, MNRAS, 227, 1

Kaiser, N. 1998, ApJ, 498, 26

Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, ApJ, 637, 631

Kashikawa, N., Ishizaki, Y., Willott, C. J., et al. 2015, ApJ, 798, 28

Katsianis, A., Tescari, E., & Wyithe, J. S. B. 2015, MNRAS, 448, 3001

Kazin, E. A., Koda, J., Blake, C., et al. 2014, MNRAS, 441, 3524

Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189

Kent, S. M. 1987, AJ, 93, 816

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

Kilbinger, M., Wraith, D., Robert, C. P., et al. 2010, MNRAS, 405, 2381

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102

Kodama, T., Arimoto, N., Barger, A. J., & Arag'on-Salamanca, A. 1998, A&A, 334, 99

Kodama, T., Bell, E. F., & Bower, R. G. 1999, MNRAS, 302, 152

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Kong, X., Daddi, E., Arimoto, N., et al. 2006, ApJ, 638, 72

Kovač, K., Somerville, R. S., Rhoads, J. E., Malhotra, S., & Wang, J. 2007, ApJ, 668, 15

Koyama, Y., Smail, I., Kurk, J., et al. 2013, MNRAS, 434, 423

Koyama, Y., Kodama, T., Tadaki, K.-i., et al. 2014, ApJ, 789, 18

Kravtsov, A. V., & Klypin, A. A. 1999, ApJ, 520, 437

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35

Kravtsov, A., Vikhlinin, A., & Meshscheryakov, A. 2014, arXiv:1401.7329

Lacey, C., & Cole, S. 1993, MNRAS, 262, 627

Lagos, C. D. P., Lacey, C. G., Baugh, C. M., Bower, R. G., & Benson, A. J. 2011, MNRAS, 416, 1566

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

Lane, K. P., Almaini, O., Foucaud, S., et al. 2007, MNRAS, 379, L25

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Leauthaud, A., Tinker, J., Behroozi, P. S., Busha, M. T., & Wechsler, R. H. 2011, ApJ, 738, 45

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159

Le Borgne, D., & Rocca-Volmerange, B. 2002, A&A, 386, 446

Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881

Lee, K.-S., Giavalisco, M., Gnedin, O. Y., et al. 2006, ApJ, 642, 63

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Limber, D. N. 1953, ApJ, 117, 134

Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, ApJ, 756, 71

Ly, C., Lee, J. C., Dale, D. A., et al. 2011, ApJ, 726, 109

Ma, C.-P., & Fry, J. N. 2000, ApJ, 543, 503

Madau, P. 1995, ApJ, 441, 18

Magdis, G. E., Rigopoulou, D., Huang, J.-S., & Fazio, G. G. 2010, MNRAS, 401, 1521

Magliocchetti, M., Popesso, P., Brusa, M., et al. 2016, arXiv:1606.08286

Makiya, R., Enoki, M., Ishiyama, T., et al. 2016, PASJ, 68, 25

Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715

Maraston, C., Daddi, E., Renzini, A., et al. 2006, ApJ, 652, 85

Martinez-Manso, J., Gonzalez, A. H., Ashby, M. L. N., et al. 2015, MNRAS, 446, 169

Massey, R., Rhodes, J., Ellis, R., et al. 2007, Nature, 445, 286

Matsuoka, Y., Masaki, S., Kawara, K., & Sugiyama, N. 2011, MNRAS, 410, 548

McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202

McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156

McCracken, H. J., Wolk, M., Colombi, S., et al. 2015, MNRAS, 449, 901

Miyatake, H., More, S., Takada, M., et al. 2016, Physical Review Letters, 116, 041301

Miyazaki, S., Komiyama, Y., Sekiguchi, M., et al. 2002, PASJ, 54, 833

Miyazaki, S., Komiyama, Y., Nakaya, H., et al. 2012, Proc. SPIE, 8446, 84460Z

Miyazaki, S., & HSC Collaboration 2013, HSC Legacy Survey Proposal

Miyazaki, S., Oguri, M., Hamana, T., et al. 2015, ApJ, 807, 22

Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347

Mo, H. J., Jing, Y. P., & White, S. D. M. 1996, MNRAS, 282, 1096

Mo, H. J., & White, S. D. M. 2002, MNRAS, 336, 112

Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903

Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, ApJ, 767, 50

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJS, 206, 8

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

Norberg, P., Baugh, C. M., Hawkins, E., et al. 2001, MNRAS, 328, 64

Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19

Orsi, A., Lacey, C. G., Baugh, C. M., & Infante, L. 2008, MNRAS, 391, 1589

Ouchi, M., Shimasaku, K., Okamura, S., et al. 2001, ApJ, 558, L83

Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004a, ApJ, 611, 660

Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, ApJ, 611, 685

Ouchi, M., Hamana, T., Shimasaku, K., et al. 2005, ApJ, 635, L117

Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2013, A& A, 551, A29

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

Peebles, P. J. E. 1980, The large-scale structure of the universe, Princeton University Press, 1980. 435 p.,

Peebles, P. J. E. 1982, ApJ, 263, L1

Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, ApJ, 675, 234-261

Phillipps, S., Fong, R., Fall, R. S. E. S. M., & MacGillivray, H. T. 1978, MNRAS, 182, 673

Pons-Bordería, M.-J., Martínez, V. J., Stoyan, D., Stoyan, H., & Saar, E. 1999, ApJ, 523, 480

Porciani, C., & Giavalisco, M. 2002, ApJ, 565, 24

Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2012, MNRAS, 423, 3018

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Puget, P., Stadler, E., Doyon, R., et al. 2004, Proc. SPIE, 5492, 978

Quadri, R. F., Williams, R. J., Lee, K.-S., et al. 2008, ApJ, 685, L1

Quenouille, G., 1949 Ann. Math. Statist., 3, 355

Quenouille, G., 1956, Biometrika, 43, 353

Reed, D. S., Governato, F., Quinn, T., Stadel, J., & Lake, G. 2007, MNRAS, 378, 777

Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541

Richardson, J., Chatterjee, S., Zheng, Z., Myers, A. D., & Hickox, R. 2013, ApJ, 774, 143

Rivolo, A. R. 1986, ApJ, 301, 70

Roche, N., Eales, S. A., Hippelein, H., & Willott, C. J. 1999, MNRAS, 306, 538

Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJ, 739, L40

Rodríguez-Torres, S. A., Chuang, C.-H., Prada, F., et al. 2016, MNRAS, 460, 1173

Rovilos, E., Georgantopoulos, I., Akylas, A., et al. 2014, MNRAS, 438, 494

Rubin, V. C., & Ford, W. K., Jr. 1970, ApJ, 159, 379

Rujopakarn, W., Eisenstein, D. J., Rieke, G. H., et al. 2010, ApJ, 718, 1171

Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183

Salpeter, E. E. 1955, ApJ, 121, 161

Santini, P., Fontana, A., Grazian, A., et al. 2009, A&A, 504, 751

Santini, P., Fontana, A., Grazian, A., et al. 2012, A&A, 538, A33

Sato, T., Sawicki, M., & Arcila-Osejo, L. 2014, MNRAS, 443, 2661

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

Seljak, U., & Zaldarriaga, M. 1996, ApJ, 469, 437

Seljak, U. 2000, MNRAS, 318, 203

Shane, C. D., & Wirtanen, C. A. 1967, Publ. Lick. Obs., 22, Part 1

Shao, J., & Wu, C. J. 1986, Ann. Math. Statist., 4, 1322

Shao, J., & Wu, C. J. 1989, Ann. Math. Statist., 17, 1176

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

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Sheth, R. K., & Tormen, G. 2002, MNRAS, 329, 61

Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15

Shimasaku, K., Ouchi, M., Furusawa, H., et al. 2005, PASJ, 57, 447

Simon, P. 2007, A&A, 473, 711

Skibba, R. A., Smith, M. S. M., Coil, A. L., et al. 2014, ApJ, 784, 128

Skibba, R. A., Coil, A. L., Mendez, A. J., et al. 2015, ApJ, 807, 152

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Slipher, V. M. 1917, Proceedings of the American Philosophical Society, 56, 403

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311

Smolčić, V., Schinnerer, E., Zamorani, G., et al. 2009, ApJ, 690, 610

Sofue, Y., & Rubin, V. 2001, ARA&A, 39, 137

Somerville, R. S., Popping, G., & Trager, S. C. 2015, MNRAS, 453, 4337

Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, ApJ, 825, 5

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2016, arXiv:1611.09354

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17

Stoyan, D., & Stoyan. H. 1994, Fractals, Random Shapes and Point Fields (New York: Wiley)

Suzuki, T. L., Kodama, T., Tadaki, K.-i., et al. 2015, ApJ, 806, 208

Takada, M., & Jain, B. 2003, MNRAS, 344, 857

Tanaka, M. 2015, ApJ, 801, 20

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709-728

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Tinker, J. L., Sheldon, E. S., Wechsler, R. H., et al. 2012, ApJ, 745, 16

Tinker, J. L., Leauthaud, A., Bundy, K., et al. 2013, ApJ, 778, 93

Tinker, J., Wetzel, A., Conroy, C., & Mao, Y.-Y. 2016, arXiv:1609.03388

Tisserand, P., Le Guillou, L., Afonso, C., et al. 2007, A&A, 469, 387

Trenti, M., & Stiavelli, M. 2008, ApJ, 676, 767-780

Toshikawa, J., Kashikawa, N., Ota, K., et al. 2012, ApJ, 750, 137

Toshikawa, J., Kashikawa, N., Overzier, R., et al. 2016, ApJ, 826, 114

Totsuji, H., & Kihara, T. 1969, PASJ, 21, 221

Tukey J. W., 1958 Ann. Math. Statist., 29, 614

Turner, M. S., Steigman, G., & Krauss, L. M. 1984, Physical Review Letters, 52, 2090

Tyson, J. A., Valdes, F., & Wenk, R. A. 1990, ApJ, 349, L1

Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189

van den Bosch, F. C., Yang, X., & Mo, H. J. 2003, MNRAS, 340, 771

van den Bosch, F. C., Mo, H. J., & Yang, X. 2003, MNRAS, 345, 923

van der Burg, R. F. J., Hildebrandt, H., & Erben, T. 2010, A&A, 523, A74

van Dokkum, P. G., Labbé, I., Marchesini, D., et al. 2009, PASP, 121, 2

van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, MNRAS, 459, 3251

Vargas-Magaña, M., Bautista, J. E., Hamilton, J.-C., et al. 2013, A&A, 554, A131

Wake, D. A., Whitaker, K. E., Labbé, I., et al. 2011, ApJ, 728, 46

Wechsler, R. H., Somerville, R. S., Bullock, J. S., et al. 2001, ApJ, 554, 85

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

Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I., & Bernstein, G. 2000, Nature, 405, 143

Wraith, D., Kilbinger, M., Benabed, K., et al. 2009, Phys. Rev. D, 80, 023507

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868-1881

Yagi, M., Kashikawa, N., Sekiguchi, M., et al. 2002, AJ, 123, 66

Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057

Yang, X., Mo, H. J., van den Bosch, F. C., Zhang, Y., & Han, J. 2012, ApJ, 752, 41

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

Zaldarriaga, M., & Seljak, U. 2000, ApJS, 129, 431

Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59

Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H. 2005, ApJ, 624, 505

Zentner, A. R., Hearin, A., van den Bosch, F. C., Lange, J. U., & Villarreal, A. 2016, arXiv:1606.07817

Zheng, Z. 2004, ApJ, 610, 61

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791

Zheng, X. Z., Bell, E. F., Papovich, C., et al. 2007, ApJ, 661, L41

Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760

Zheng, Z., Zehavi, I., Eisenstein, D. J., Weinberg, D. H., & Jing, Y. P. 2009, ApJ, 707, 554

Zu, Y., & Mandelbaum, R. 2015, MNRAS, 454, 1161

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Zwicky, F. 1937, ApJ, 86, 217