UNIVERSITÉ D'AIX-MARSEILLE
École Doctorale Sciences de la Matière
Faculté des Sciences
LABORATOIRE d'ASTROPHYSIQUE DE MARSEILLE

THÈSE DE DOCTORAT

Discipline : Astrophysique
Spécialité : Cosmologie

Anna DURKALEC

Properties and evolution of galaxy clustering at $2 < z < 5$ based on the VIMOS Ultra Deep Survey.

Soutenue publiquement le 11 decembre 2014
au Laboratoire d'Astrophysique de Marseille
devant le jury composé de:

Prof. Veronique Buat	Présidente du jury
Prof. Luigi Guzzo	Rapporteur
Prof. Alain Blanchard	Rapporteur
Dr. Sylvain de la Torre	Examinateur
Prof. Agnieszka Pollo	Examinatrice
Prof. Olivier le Fèvre	Directeur de thèse

Acknowledgments

Foremost, I would like to express my deepest thanks to my supervisor Olivier Le Fèvre, for his constant involvement in my research, his availability, encouragement and most of all for his enthusiasm, which was a key element to motivate me throughout my PhD.

This thesis would not be possible without the immense knowledge (and willingness to share it) of my friends and collaborators. In particular I would like to express my gratitude to Sylvain de la Torre, Agnieszka Pollo, Ummi Abbas, Olivier Ilbert, Laurence Tresse, Brian Lemaux, Paolo Cassata, Janin Porr and Lidia Tasca for all helpful discussions and good advises

Huge 'Dziekuje!' goes to Agnieszka Pollo who, quite literally, kicked me out to start this PhD. Thank you for everything! Mostly for your time, countless discussions, your patient corrections, and for your willingness to help, wherever and whenever.

Also I would like thank Martin Kilbinger for assistance with implementing my HOD module into CosmoPMC code. HOD fitting performed in this work would not be possible without him.

I warmly acknowledge funding from the European Research Council through the Advanced Grant 2010-AdG-268107-EARLY

Finally I would like to thank all my friends from LAM. My officiates in the first place - Romain and Bin-Wei. We spend most of the last 3 years together in one office and I need to say that I couldn't wish for the greater (and weirder) company. Also Ola, Kasia, Lu-

cia, Raffaela, Danka, Debora, Ana, Anna, Javier, Samuel, Mario, Firas, Bruno, Thibaud, Giovanni, Bastien, Pol and many others (whose not mentioning I will probably regret) - thank you guys for all 'outlam' activities.

Wilekie dzieki Moim Rodzicom, którym dedykuje te prace. Bez ich wsparcia, zainteresowania i miłości nie osiagnełabym niczego.

Rodzicom. Dzieki nim wciaz spelniaja sie marzenia.

Contents

Introduction **2**

1 Galaxy formation and evolution in the large scale structure **5**
1.1 Foundations of the Big Bang cosmology 6
1.2 Parametrization of the universe 8
1.2.1 ΛCDM hierarchical model 9
1.3 Theoretical predictions 11
1.3.1 Evolution of dark matter haloes 13
1.3.2 Predicted evolution of the two-point correlation function 16
1.4 Observations of the Universe large-scale structure 18

2 Data sample. VIMOS Ultra Deep Survey **21**
2.1 Spectroscopic target selection 22
2.2 Redshift measurements and reliability flags 23
2.3 Absolute magnitude and stellar mass 24
2.4 Advantages of the VUDS survey 24

3 Quantifying the large-scale distribution of galaxies **29**
3.1 Correlation function measurements 30
3.1.1 General correlation function estimators 31
3.1.2 Redshift-space and projected real-space correlation function 32
3.2 Halo Occupation Distribution (HOD) model 33
3.3 The VUDS mock catalogues 39
3.4 Optimization of correlation function measurements 40

3.4.1 Real-space correlation function from multiple independent fields . . 41
3.4.2 Correlation function measurements details 43
3.5 HOD model fitting tests 53
3.6 Galaxy bias 56

4 Clustering measurements of the high redshift galaxy population $2 < z < 5$ 59
4.1 The sample selection for clustering analysis 61
4.2 Impact of the reliability of redshift measurements 63
4.3 Cosmic variance between VUDS fields 65
4.4 Results in the framework of two approximations 67
4.4.1 Power-law approximation 67
4.4.2 HOD modelling 69
4.5 Comparison with other studies 71
4.6 Evolution of the host dark matter halo 75
4.7 Relation between dark and luminous matter 76
4.8 Summary 79

5 Luminosity and stellar mass dependence of galaxy clustering at $z \sim 3$ 83
5.1 Sample selection for the analysis 87
5.2 Correlation function results 87
5.3 Linear large-scale galaxy bias 88
5.4 Summary 88

6 The stellar mass $M_\star$ – halo mass M_h relation at z∼3 89
6.1 Data sample selection 91
6.2 The star formation efficiency 93
6.3 Summary 95

7 Conclusion 97

Bibliography 100

A Publications 109

B HOD model code documentation 179

List of Figures 188

List of Tables 191

Abstract

This thesis focuses on the study of the properties and evolution of galaxy clustering for galaxies in the redshift range $2.0 < z < 5.0$ from the VIMOS Ultra Deep Survey (VUDS), which is the largest spectroscopic galaxy survey at $z > 2$. Thanks to the size and high reliability of the VUDS galaxy sample I was able to measure the spatial distribution of a general galaxy population at redshift $z \sim 3$ for the first time with a high accuracy. The measurements have been performed using a set of tools, which I developed during my study. I quantified the galaxy clustering by estimating and modelling the projected (real-space) two-point correlation function, $w_p(r_p)$, for a general population of $3,022$ galaxies with robust spectroscopic redshifts in two independent fields (VVDS-02h and COSMOS) covering in total 0.8 deg^2. I also extended the clustering measurements to the luminosity and stellar mass-selected sub-samples in order to study the dependence of galaxy clustering on these two physical parameters. My results show that the clustering strength of the general galaxy population does not change significantly from redshift $z \sim 3.5$ to $z \sim 2.5$, but in both redshift ranges strongly depends on luminosity and stellar mass, with more luminous and more massive galaxies being more clustered than less luminous (massive) ones. Using the halo occupation distribution (HOD) formalism I measured an average host halo mass at redshift $z \sim 3$ significantly lower than the observed average halo masses at low redshift. Additionally, in agreement with the current scenario of the hierarchical formation of the large-scale structure, within our sample I observed an increase of the average host halo masses from $z \sim 3.5$ to $z \sim 2.5$. I concluded that the observed star-forming population of galaxies at $z \sim 3$ might have evolved into the massive and bright ($M_r < -21.5$) galaxy population at redshift $z = 0$. Additionally, I interpret clustering measurements in terms of a linear large-scale galaxy bias, quantifying how *biased* are the VUDS galaxies with respect to the underlying dark matter density field and how

this bias depends on their properties. I find that the linear large-scale galaxy bias is significantly higher than the bias of intermediate and low redshift galaxies. Finally, taking the advantage of the HOD modelling of clustering measurements and the SED fitting over the large set of multi-wavelength VUDS data I computed the stellar-to-halo mass ratio (SHMR) and the integrated star formation efficiency (ISFE) to study the efficiency of star formation and stellar mass assembly. I find that the integrated star formation efficiency is quite high at $\sim 16\%$ for the average galaxies at $z \sim 3$.

Introduction

Cosmology is the science of big questions. One may say, questions as big as the size of the Universe itself. What is the origin of the Universe? How the highly concentrated large scale structure observed at $z = 0$ has formed and evolved through time? What is the difference between the underlying dark matter density distribution and the observed luminous structures? What is the relation between these two components, and how this relation changed through time? What processes drive and regulate star formation and mass growth in galaxies along cosmic time? Finally, how the current paradigm of galaxy formation and evolution stands in the light of the observations?

Modern observations of the local Universe (e.g. Sloan Digital Sky Survey, SDSS) provide data for detailed studies of the shape and properties of the $z = 0$ large-scale structure, in order to answer some of these questions, Intermediate redshift ranges also have been recently extensively examined. Therefore, a natural next step is to extend these studies on early epochs of galaxy formation and evolution. In this thesis I present an unprecedented study of the galaxy clustering performed at $z > 2$, thanks to the data available from the biggest spectroscopic galaxy survey in the redshift range $2 < z < 6$ - the VIMOS Ultra Deep Survey (VUDS). In my work I quantified the galaxy clustering by estimating and modelling the projected (real-space) two-point correlation function, $w_p(r_p)$. I also extended the clustering measurements to the luminosity and stellar mass-selected sub-samples in order to study the dependence of galaxy clustering on these two physical parameters. Finally, taking the advantage of the HOD modelling of clustering measurements and the SED fitting over the large set of multi-wavelength VUDS data, I computed the stellar-to-halo mass ratio (SHMR) to obtain the insight into the efficiency of star formation and stellar mass assembly at redshift $z \sim 3$.

My work can be considered as a precursor of next generation studies coming with the

advent of the new powerful facilities like e.g. the James Webb Space Telescope (JWST) and Extremely Large Telescopes (ELTs).

This thesis is organized into seven chapters which present different aspects of my work. The first chapter describes the general framework and current cosmology 'state of art', i.e the base of my study. It includes a summary of paradigms, theories, assumptions, as well as observable facts, which together constitute our understanding of the origins and evolution of the Universe. I concentrate on the aspects important to my work, i.e on the current paradigms of the evolution of the large-scale structure. In the second chapter I introduce my data source - the VIMOS Ultra Deep Survey (VUDS) - the biggest spectroscopic redshift survey at $z > 2$. I present the most important features of this survey and discuss its uniqueness in terms of galaxy clustering measurements. The third chapter gives the detailed description of tools and methods, which have been used in this work to quantify galaxy clustering. I present the most important features of the galaxy correlation function measurements, such as: the correlation functions estimators, corrections for observational limitations, and the parametrisations used to quantify the results. Additionally, I show the results of various tests, which have been performed before the final measurements from the VUDS data. The fourth chapter contains the description of my study of the properties of galaxy clustering at $z \sim 3$. The results are discussed in the framework of the halo occupation distribution model (HOD). The dark matter halo mass and large-scale galaxy bias evolution is traced, by connecting the result of my study, through various models, with similar measurements at lower redshift. Chapter five summarizes measurements of the luminosity and stellar mass dependency of the galaxy clustering at two redshift ranges $z \sim 2.5$ and $z \sim 3.5$. Finally, in sixth chapter, I study the efficiency of the star formation and stellar mass assembly in galaxies, through measurements of the stellar-to-halo mass ratio (SHMR). At the end I conclude and discuss my results in terms of the current scenario of structure formation and evolution. Additionally, the papers accepted or submitted for publication in refereed journals related to my thesis work are presented in Appendix A:

- Paper 1: *The VIMOS Ultra-Deep Survey:* $\sim 10,000$ *galaxies with spectroscopic redshifts to study galaxy assembly at early epochs* $2 < z \simeq 6$, O. Le Fèvre, L.A.M Tasca, P. Cassata, B. Garilli,... A. Durkalec et al. (2014) - in press.
- Paper 2: *The evolution of clustering length, large-scale bias and host halo mass at* $2<z<5$ *in the VIMOS Ultra Deep Survey (VUDS)*, A. Durkalec, O. Le Fèvre, S. de la Torre, A. Pollo et al. (2014) - submitted.
- Paper 3: *The stellar-to-halo mass ratio using galaxy clustering in VUDS: a high star formation efficiency at* $z \simeq 3$, A. Durkalec, O. Le Fèvre, S. de la Torre, A. Pollo et al. (2014) - submitted.

CHAPTER 1

Galaxy formation and evolution in the large scale structure

Cosmology by its definition is the study of the origin, evolution and eventual fate of the universe. However, there is a lot of unknowns about the beginning of the universe, only uncertain theories about its future, and some huge questions about what happens in the middle. One of the most important of these questions is how the complex structure of galaxies observed in the local universe formed and evolved through time.

Thanks to the studies of the oldest observed radiation - Cosmic Microwave Background Radiation (CMB)(Penzias & Wilson 1965) we can partially answer for the first part of this question i.e. from what and how the structure formed in the first place? The CMB is the furthest back in time we can explore using light. It comes from very early stages of the evolution of the universe when it was only $380,000$ years old and shows that, back then, matter was almost uniformly spread in space. Except of the presence of a very small, random inhomogenities (see Fig. 1.1). The observed temperature fluctuations, which are essentially the footprints of these inhomogenities, reach only $\Delta T/T \sim 10^{-5}$. Although, we still can not fully explain the origins of these observed mass distribution fluctuations, in our current paradigm of galaxy formation, they are believed to be the beginning of complex and strongly concentrated large-scale structure observed in the local universe. The question remains: how did it happen? How these small inhomogenities evolved?

The current general understanding of the structure growth follows the logical pattern, and is based mostly on gravity. First of all the primordial CMB fluctuations are believed to grow under the influence of gravity. Matter was streaming into the potential wells, creating consequently bigger and denser areas, up to the point when the gas was dense enough for the galaxy to form. The process of galaxy formation, of course, is complex

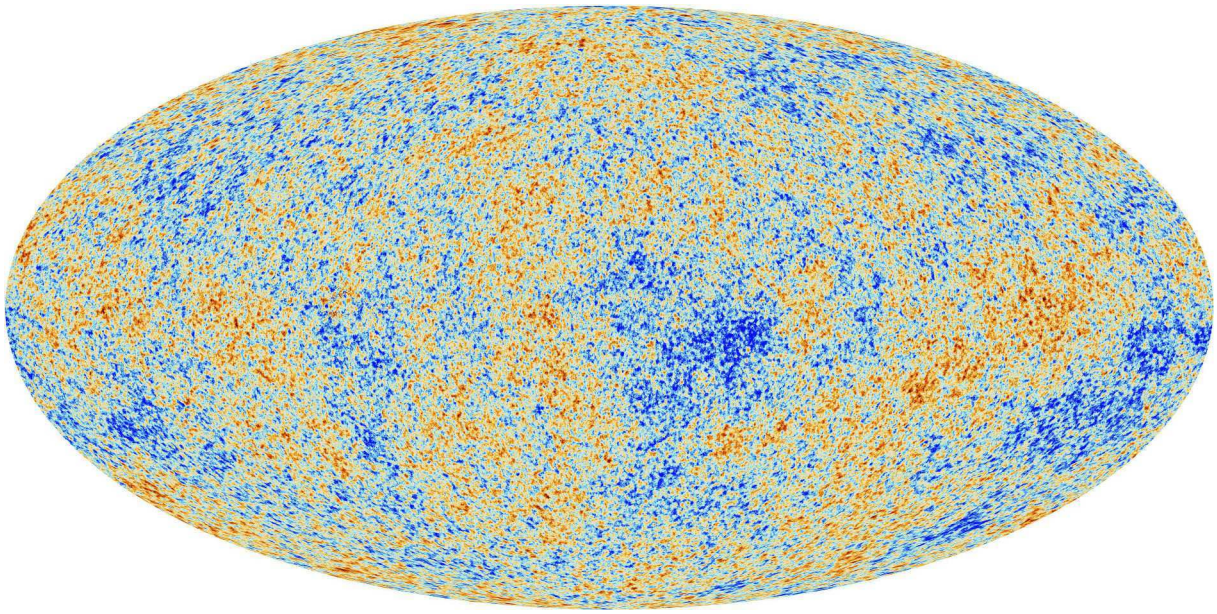


Figure 1.1: Cosmic Microwave Background Radiation (CMB) as seen by ESA's Planck satellite. Credit: ESA and the Planck Collaboration.
Picture from `http://sci.esa.int/planck/`.

and complicated and it includes sophisticated physical processes concerning gas, dust, stars, black holes, all interacting with each other (Bromm et al. 2009). After, or even during, this process the galaxies and dark matter haloes, which hosts them, interact with each other in various ways (i.e. mergers, gas accretion) and grow 'hierarchically' creating clusters, superclusters and bigger formations, causing constant changes of the whole structure through time.

By my work I am making an attempt to study these changes. Through my study, I take a closer look on the shape, properties and evolution of the Large Scale Structure at early stages of galaxy formation, by measuring galaxy clustering at $2 < z < 5$ and comparing the results with the similar ones obtained for the local and moderate Universe.

In this chapter I present the general framework of the current cosmology 'state of art'. I describe the theories, paradigms and assumptions which create the ground base of my study.

1.1 Foundations of the Big Bang cosmology

There have been many different concepts of the Universe since ancient times. True extragalactic astronomy, however, started in the eighteenth century, when astronomers discovered, for the first time, the fuzzy no-point objects. They called them *nebulae*. At this

time, the nature of these objects was an enigma, as they were considered to be either objects which come from our Milky Way or as the extragalactic individual universes. In 1925 the problem was solved by Edwin Hubble who demonstrated, by estimating the distance to these objects, that some of the nebulae are indeed extragalactic bodies, and that they are comparable in size and luminosity to our galaxy. Few years later he also showed, that the distance and radial velocity of these galaxies are linearly related (Hubble 1929), which led him to the conclusion that the Universe is expanding.

Earlier, in 1916, when Albert Einstein created the theory of general relativity, he realized that his theory is useful to construct models, which can describe the universe as a whole. In that time commonly thought that the Universe is static. Solutions of his field equations indicated however that the Universe is expanding. Convinced that this expansion was not possible he therefore introduced a cosmological constant Λ into the equations, to force static solutions.

In 1922 Alexander Friedman presented both static and expanding Universe solutions - called Friedman models. However, his work hasn't been noticed until Gorges Lemaitre independently rediscovered the same solutions in 1927.

Nowadays observations indicate that the Universe not only expands, but also accelerates. This means that at very early epochs the Universe has to be denser and hotter, therefore its initial state had to be characterized by an extreme temperature and density: *The Big Bang*.

These last statements are essentially the ground base of the current *standard cosmological model* - the model which offers comprehensive explanations of various observations at different stages of the evolution of the Universe. This model well explain the existence and structure of the Cosmic Microwave Background Radiation (CMB), the observed shape of the large scale structure and the distribution of galaxies, the accelerating expansion of the universe, the abundance of hydrogen and others.

The standard cosmological model is based upon the hypothesis that:

- Our observational location is not unusual in any case, meaning that the properties of the observed universe are very similar at every point and there are no preferential directions in space
- The Universe is homogeneous and isotropic (at sufficiently large scale) - Cosmological principle.
- Einstein's theory of general relativity describe the dynamic of expansion and evolution of the universe.

Under these assumptions, according to the standard cosmological model, the Universe was born nearly 14 billion years ago, and it is evolving ever since. The evolution of the

Universe can be described in terms of cosmological models, by using physically motivated cosmological parameters

1.2 Parametrization of the universe

There is general agreement that a standard model which describes the universe in the excellent way exists. However we are still far from the agreement whatever it was already created or not. The model which currently is able to explain the various sets of observations and cosmological probes is called ΛCDM (Lambda Cold Dark Matter, see section 1.2.1), which is a parametrization of the Big Bang cosmological model mentioned in the previous section.

In general all existing modern models, which try to describe the evolution of the Universe, are essentially theories based on general relativity. This means, that the geometry of space-time, within these models, is determined by the matter/energy content of the universe through the Einstein field equation:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.1}$$

Here $R_{\mu\nu}$ is the Ricci tensor, describing the local curvature of the space-time, R is the curvature scalar, $g_{\mu\nu}$ is the metric, $T_{\mu\nu}$ is the energy-momentum tensor of the matter content of the Universe, and Λ is the cosmological constant.

Under the assumption of homogeneity and isotropy (cosmological principle) on large scales the $g_{\mu\nu}$ is given by the Friedman-Robertson-Walker (FRW) metric (Eq. 1.2), which allows the Ricci tensor $R_{\mu\nu}$ and curvature scalar R to be expressed in terms of the scale factor $a(t)$ and the curvature signature k:

$$ds^2 = -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2 \theta d\phi^2 \right) \right] \tag{1.2}$$

where $k = +1{,}0$ or -1 if the spatial hypersurfaces have positive, zero or negative curvature respectively (corresponding to the 'open', 'flat' and 'close' Universe). Four space-times coordinates t, r, ϕ, and θ represents time and three polar coordinates respectively. The evolution of the scalar factor $a(t)$ - i.e the expansion of the Universe, is determined by a Friedman equation:

$$H^2(a) = H_0^2 \left[\Omega_\Lambda + \Omega_m a^{-3} + \Omega_r a^{-4} - (\Omega_{tot} - 1)\, a^{-2} \right], \tag{1.3}$$

where a is conventionally taken as to be $a = 0$ at the present day, so that the redshift z is given by $a = 1/(1 + z)$. The Hubble parameter $H \equiv \dot{a}/a$, gives the instantaneous rate of

expansion of the Universe, and H_0 is its present value [1]. The different Ω parameters give the density of each of the component, at the present day - radiation, matter, curvature and cosmological constant, normalized to the critical density:

$$\Omega_i \equiv \frac{\rho_i}{\rho_{crit}} = \frac{8\pi G\rho_i}{3H_0^2} \tag{1.4}$$

where ρ_{crit} is the critical density for which the spatial geometry is flat [2]. Here we account for contributions from radiation Ω_r (negligible in the matter-dominated epoch, present), matter Ω_m, a cosmological constant Ω_Λ, and a curvature term Ω_k. The curvature however is determined by the total energy density of the other components:

$$\Omega_{tot} = \Omega_r + \Omega_m + \Omega_\Lambda = 1 + \Omega_k \tag{1.5}$$

Thus, the complete state of the homogeneous Universe can be described by giving the values of all the density parameters Ω and the Hubble constant h (the present-day Hubble parameter being written $H_0 = 100h$ km s^{-1} Mpc^{-1}).

1.2.1 ΛCDM hierarchical model

Among all cosmological theories, which try to explain the formation and evolution of cosmic structure, the ΛCDM model is in the best agreement with a broad variety of astrophysical observations. It assumes that the Universe is flat and contain three components:

- Dark Energy - described from a historical reasons by the cosmological constant Λ. Represents the 68.3% of the total content of the Universe, and is used to explain the observed accelerating expansion of space against the attractive effects of gravity.
- Cold dark matter (CDM) - Representing 26.8% of the total content of the Universe. Its presence was proved by indirect observations, but their nature is unknown. However it is described as being cold (i.e. the velocity of dark matter is far less than the speed of light at the epoch of radiation-matter equality); non-baryonic (i.e. consisting of matter other than protons and neutrons); dissipationless (i.e. cannot cool by radiating photons); and collisionless (i.e. the dark matter particles interact with each other and other particles only through gravity and possibly the weak force).
- Visible matter - Representing the remaining 4.9% of the Universe content. Consists a baryonic matter such as protons, neutrons and electrons. Essentially the matter of which planets, stars and galaxies are made of.

[1] Here we adopt the common notation $H_0 = 100h$ km s^{-1} Mpc^{-1}

[2] Currently the critical density of the Universe have value $\rho_{crit} = 2.775 \cdot 10^{11} h^2 M_\odot \text{Mpc}^{-3}$

Table 1.1: Current values of the cosmological parameters of the ΛCDM model and the 68% confidence limits for constrained parameters. Column 3 gives the results for the *Planck* temperature power spectrum data alone, while column 4 combine the *Planck* temperature data with *Planck* lensing (both form Planck Collaboration et al. 2014). The results shown in the last column are based on combining WMAP5 with data on the baryon acoustic oscillation (BAO) feature in the distribution of galaxies and on the redshift-distance relation of Type 1a supernovae (SN).

Parameter	Notation	*Planck*	*Planck*+lensing	WMAP+BAO+SN
Hubble constant [a]	H_0	67.4 ± 1.4	67.9 ± 1.5	70.1 ± 1.3
Scalar amplitude	A_s	$(2.23 \pm 0.16) \times 10^{-9}$	$(2.19^{+0.12}_{-0.14}) \times 10^{-9}$	$(2.457^{+0.092}_{-0.093}) \times 10^{-9}$
Scalar spectral index	n_s	0.9616 ± 0.0094	0.9635 ± 0.0094	$0.960^{+0.014}_{-0.013}$
Dark energy density	$\Omega_\Lambda h^2$	0.686 ± 0.020	0.693 ± 0.019	0.721 ± 0.015
Cold dark matter density	$\Omega_c h^2$	0.1196 ± 0.0031	0.1186 ± 0.0031	0.1143 ± 0.0034
Baryon matter density	$\Omega_b h^2$	0.02207 ± 0.00033	0.02217 ± 00.00033	0.02265 ± 0.0059
Reoinization optical depth	τ	0.097 ± 0.038	0.089 ± 0.032	0.084 ± 0.016

[a] In [km s^{-1} Mpc^{-1}]

In its simplest form the ΛCDM model is based on six parameters: scalar amplitude A_s, scalar spectral index n_s, CDM (Cold Dark Matter) density Ω_c, baryon density Ω_b, Hubble parameter h, and reionization optical depth τ. Different values of these parameters produce different structures distribution in the Universe. Great effort has been put in order to find the values of the cosmological parameters and the most robust estimations of cosmological parameters have been preformed using CMB observations.

Since the physics behind the observable quantities of the CMB is well understand, the observational results can be used to put stringent constrains on cosmology. Recently, great progress has been made in our observational assessment of the CMB, starting with the preliminary detection of anisotropy on large angular scales by COBE, through WMAP measurements of acoustic peaks, damping tail and polarization, up to insanely accurate *Planck* measurements. Thanks to these observations all model parameters of the ΛCDM cosmology have already been constrained with great accuracy, as presented in Table 1.1.

1.3 Hierarchical evolution of structures - theoretical predictions

According to the current paradigm the formation of the structure started from small scale matter over-densities. We can observe the footprint of these primordial imhomogenities in Cosmic Microwave Background Radiation (CMB), but their origin is still an enigma. One of the possible explanations is provided by the *inflation* theory. According to this theory the observed initial perturbations were formed during rapidly exponential expansion of the Universe at its early stages of evolution - Inflation. After this epoch the small-scale perturbations started to grow with time - over-dense regions pull matter from under-dense regions. When the over-density of perturbation reached $\delta\rho/\rho \sim 1$ it separated from the general Universe expansion and started to collapse under its own gravity. As a result dark matter segregated from baryonic matter. First one formed into dark matter haloes, while the second one started to cool. The dark matter haloes continue to grow in mass (and size) either by accreting material from their neighbourhood or by merging with other haloes. The history of dark halo mergers can be illustrated in the form of the 'merger tree'. Figure 1.2 shows the example of how a single, massive halo, which we can observe at the present epoch ($z = 0$) was created from small massive halos. The merger process occurs in many different types, in general the two most important types can be defined as:

- Major mergers - dark matter halos of similar masses with central galaxies merge producing a new central galaxy. Usually this process is accompanied by strong star

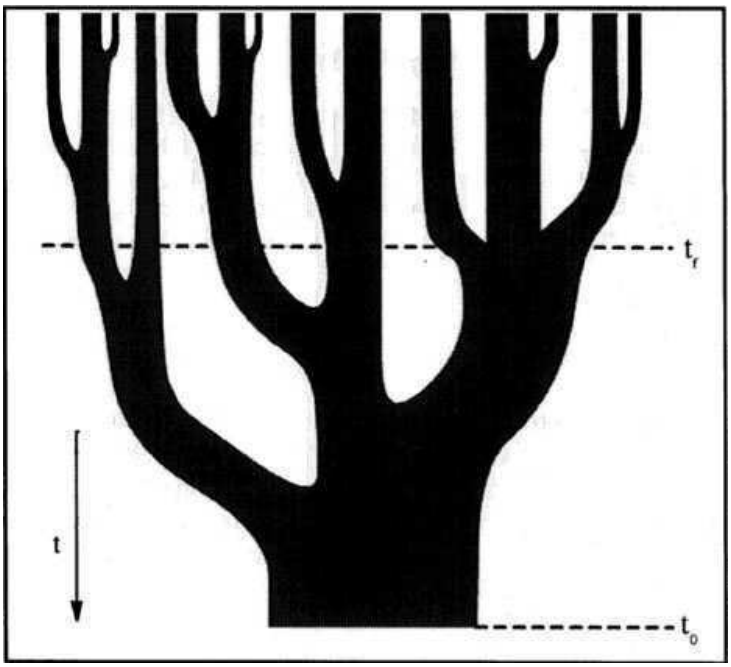


Figure 1.2: A schematic representation of a 'merger tree' depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_j are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

formation or AGN activity.

- Minor mergers - dark matter halos of significantly different masses merge. Dark matter components merge and consolidate into one massive dark matter halo, while galaxies inside it form a system with the massive galaxy near the centre of the new dark matter halo and the satellite galaxy (or galaxies).

Tracing the evolution of baryonic matter is much more complex. If gas cooling processes are effective, matter accumulate in the central part of dark matter halo, creating a protogalaxy. During the collapse gas could fragment into smaller parts, which may evolve into stars. Unfortunately the main part of this process is still not fully understand. For example we are not able to predict what part of the initial mass of protogalaxy is about to be transformed into stars, and how much time it will take. What's more we cannot assume that all baryonic matter is cooling and forming stars. So-called feedback processes prevent gas from cooling or reheat after it has become cold. This might be for instance supernovae explosions, which are strong enough to blow out gas from the galaxy, or the presence of an AGN (Active Galactic Nuclei) in the neighbourhood of the galaxy. Even if we know that these processes play an important role in galaxy formation, we still don't know which one of them dominate and how exactly they occur.

There were made attempts to describe large-scale structure evolution theoretically by using gravitational instability calculations (Press & Schechter1974). This formalism allows to calculate many properties of the population of dark matter haloes, such as their mass function, the mass distribution of their progenitors, their merger rate and their clustering properties. In the following sections we present briefly attempts to describe formation and evolution of dark matter haloes and galaxies.

1.3.1 Evolution of dark matter haloes

Newtonian theory of small perturbations

We can consider the evolution of the initial cosmological perturbations by using of Newtonian perturbation theory in the linear regime. This approximation applies to the structures with sizes much smaller than the horizon size, so that causality can be considered instantaneous, and with the density contrast relative to the background much smaller than unity.

In this framework the time evolution of a fluid of the density ρ and velocity u under the influence of a gravitational field with potential Φ is given by the continuity equation (which describes mass conservation), Euler's equations (the equations of motion), and the

Poisson equation (describing the gravitational field):

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) &= 0 && \text{(continuity)} \\ \frac{\partial u}{\partial t} + (u \cdot \nabla)\, u &= -\frac{1}{\rho}\nabla p - \nabla \Phi && \text{(Euler)} \\ \nabla^2 \Phi &= 4\pi G \rho && \text{(Poisson)} \end{aligned} \tag{1.6}$$

These equations can be rewritten in the comoving frame by using the position x in the comoving coordinate, the peculiar velocity v, density fluctuations(perturbations) $\delta(t,x) \equiv \rho(t,x)/\bar{\rho}(t,x)$ and the gravitational potential $\phi(t,x)$. Also under the assumption of homogeneous universe (cosmological principle) dominated by the collisionless dark matter, the non-linear effects are negligible and the above equations can be linearized. Then the density fluctuations, transformed to the k space [3] is given by:

$$\ddot{\delta_k} + 2\frac{\dot{a}}{a}\dot{\delta}_k - 4\pi G\bar{\rho}\delta_k = 0 \tag{1.7}$$

here $\dot{a}/a = H(t)$ is the Hubble parameter. The above equation can be solved with respect to the time t, by defining $\delta_k(t) = D(t)\delta_k(t_0 = 0)$, where D is the so-called *growth factor*. Equation 1.7 is the second-order differential equation. Hence, there are two independent solutions of this equation: a decaying mode $\delta_k(t) = D_-(t)\delta_k(0)$ and a growing mode $\delta_k(t) = D_+(t)\delta_k(0)$, which describe monotonic decrease and increase of the density fluctuations with time t, respectively. The former mode becomes negligibly small as the Universe expands, and thus one is usually interested in the growing mode alone. The growth factor depends on the cosmological parameters and for ΛCDM cosmology it can be described as:

$$D_+(z) \simeq 2.5\Omega_m(1+z)^{-1}\left[\Omega_m^{4/7} - \Omega_\Lambda + \left(1+\frac{\Omega_m}{2}\right)\left(1+\frac{\Omega_\Lambda}{70}\right)\right]^{-1} \tag{1.8}$$

The linear growth rate of density fluctuation $D(z)$ cosmological parameter dependency is presented in Figure 1.3.

Dark Matter power spectrum

As described in the previous section in principle one can specify the cosmic density field by specifying the density perturbation $\delta(t,x)$ at every point in space (or equivalently, to specify δ_k for all k). Therefore, the large-scale structure of the Universe can be described by the evolution of the dark matter density $\delta(x)$ field seeded in the early epochs of the

[3] By using $\delta_k(t) = \frac{1}{V}\int \delta(t,x)\exp(ik\cdot x)dx$, where the V is the volume of the large box on which perturbation field is assumed periodic.

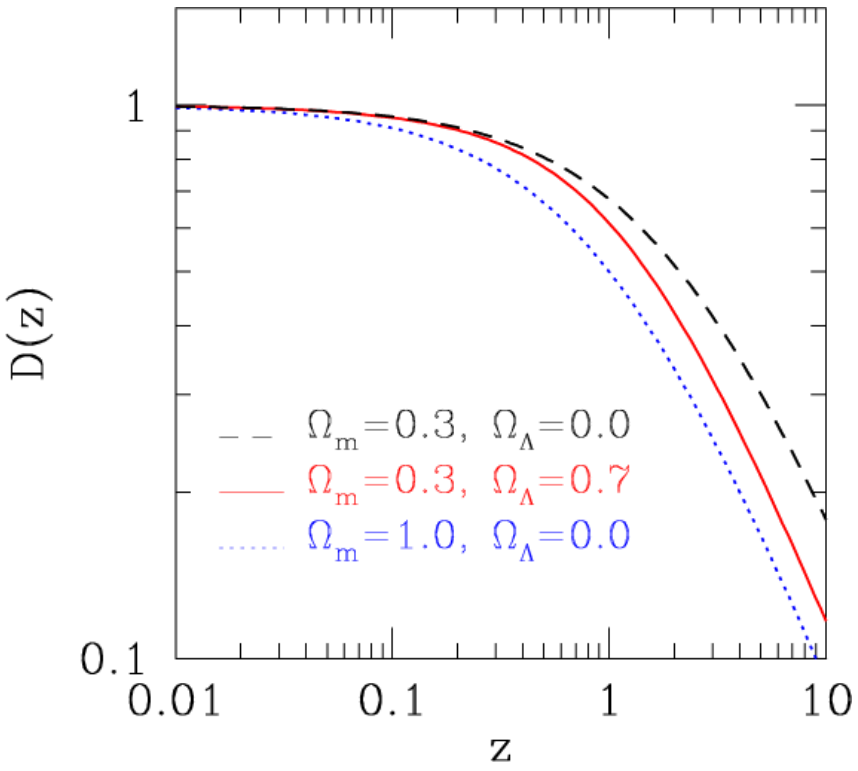


Figure 1.3: Linear growth rate of density fluctuations for different cosmologies.

Universe. It is useful however to describe this field in the form of the power spectrum of dark matter, which records information of the strength of fluctuations on different scales. The power spectrum at any redshift can be written as:

$$P(k, z) = P_p(k) \times T^2(k) \times D_+^2(z) \tag{1.9}$$

where P_p is the initial power spectrum, $T(k)$ is the transfer function that describes the interactions between the different components of the Universe, and $D_+(z)$ is the linear growth factor describing the change in the power spectrum independent of scale.

In the absence of a complete theory for the origin of the density perturbations, the initial (untransferred) power spectrum is commonly assumed to be a power law:

$$P_p(k) \propto k^n, \tag{1.10}$$

where n is called the spectral index. This approximation is generally in agreement with the power spectra shape predicted inflation models.

To completely specify the power spectrum $P(k)$ one also have to specify its amplitude. Again because we do not yet have a refined theory for the origin of the cosmological perturbations, this amplitude is not predicted a priori but rather has to be fixed by the

observations. Even for inflation models, where we can make some predictions for the shape of the initial power spectrum, the current theory has virtually no predictive power regarding the amplitude.

For a power spectrum with a given shape, the amplitude is fixed when we know the value of $P(k)$ at any k, or the value of any statistic that depends only on $P(k)$. One historical prescription for normalizing a theoretical power spectrum involves the variance of the galaxy distribution when sampled with randomly placed spheres of radii R. The predicted variance of the density field is related to the power spectrum by:

$$\sigma^2(R) = \frac{1}{2\pi} \int P(k)\ W^2(kR)\ k^2\ dk \tag{1.11}$$

where

$$W(kR) = \frac{3}{(kR)^3}\ [\sin(kR) - kR\cos(kR)] \tag{1.12}$$

is the Fourier transform of the spherical top-hat window function. The value of power spectrum normalization $\sigma(R)$ derived from the temperature fluctuations in CMB by *Planck* (Planck Collaboration et al. 2014) at $R = 8h^{-1}$Mpc for the ΛCDM cosmological model is equal $\sigma_8 = 0.83$. Thus, one can in principle normalize the theoretical power spectrum by requiring $\sigma(R) = 0.83$ at $R = 8h^{-1}$Mpc.

The linear matter power spectrum $P(k)$ extrapolated to the current times $z = 0$ is presented in Figure 1.4.

1.3.2 Predicted evolution of the two-point correlation function

The two-point correlation function (fully described in chapter 3), main tool used in this thesis to quantify large-scale galaxy distribution, is the Fourier transform of the power spectrum $P(k)$.

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k)\ \exp(ik \cdot r) d^3k = \frac{1}{2\pi} \int_0^\infty P(k)\ \frac{\sin kr}{kr} k^2 dk \tag{1.13}$$

Because of that connection, from the theoretical point of view, the two-point correlation function is a simple way to characterize the matter distribution in the Universe. Since the density field of the Universe evolves over time (see Sec. 1.3), the correlation function evolves as well - $\xi = \xi(r, t)$. In the framework of the Newtonian linear perturbation theory (and in the matter-dominated era of the history of the Universe) in the co-moving coordinates the density contrast can be decomposed into time-dependent and space-dependent factors: $\delta(\vec{r}, t) = D(t)\delta_0(\vec{r})$, where $\delta_0(\vec{r})$ is the present day value of the density contrast at a given location, and $D(t)$ is refer as the growth factor (its value depends on the parameters of the assumed cosmological model). In practice, it means that the spatial

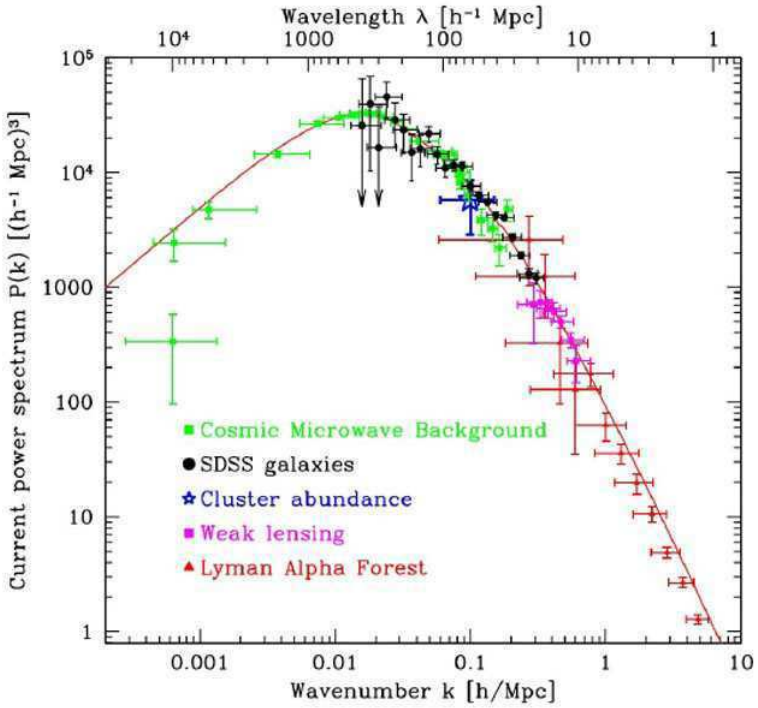


Figure 1.4: Linear matter power spectrum $P(k)$ as a function of wavenumber k from various measurements of cosmological structure. The best fit ΛCDM is presented as a solid line. Figure from Tegmark et al. (2004).

shape of the density fluctuations in co-moving coordinates does not change, and only their amplitude increases. Consequently, only the amplitude of the correlation function should change as well: $\xi(r,t) = D^2(t)\xi(r,t_0)$, and the increase of its amplitude over time depends on the cosmological parameters (Peebles 1980, Schneider 2006).

Thus, the measurement of the evolution of the amplitude of a correlation function of a given population of galaxies provides - in principle - an easy way to test the paradigm of the gravitational perturbations as the origin of the large scale structure of the Universe, and a way to estimate cosmological parameters. However, the reality is not that simple. One of the main reasons are the limitations of the linear theory: dark matter haloes and, consequently, galaxies formed in them, are not unbiased tracers of the underlying matter density field (Press & Schechter 1974).

A second important reason is galaxy evolution: comparing exactly the same population of galaxies at different redshifts is still a very difficult task. Because of these two reasons, the simulations of the evolution of the dark matter density field and dark haloes forming in it, possibly complemented with the semi-analytic models of galaxy formation became a useful tool to compare the evolution of the correlation function with the theoretical

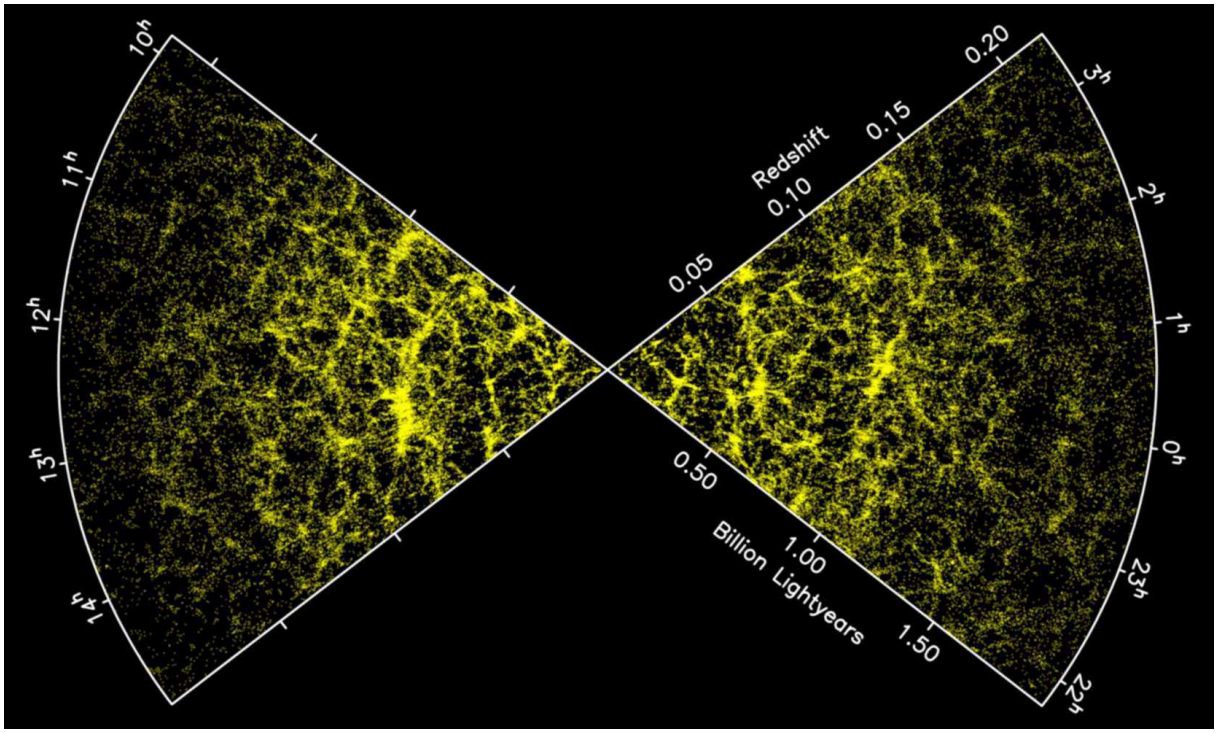


Figure 1.5: The spatial distribution of $\sim 80\,000$ galaxies in the 2dFGRS. Adapted from `http://www.roe.ac.uk/roe/support/pr/pressreleases/2dfslice_black.gif`.

predictions.

1.4 Observations of the Universe large-scale structure

Obviously all theoretical predictions described in previous sections would be meaningless without observational confirmations. Therefore the large galaxy surveys remain the fundamental cosmological tools, which have the biggest contribution in investigating the formation and evolution of large scale structure.

These surveys revealed that the distribution of galaxies is not random, but shows a variety of structures. Some galaxies are located in high-density clusters containing from few tens up to several hundreds of galaxies. The majority of all galaxies, however, are distributed in low-density filamentary or sheet-like structures, which surround large voids (with diameters up to $\sim$ 100Mpc). The example of galaxy structure distribution observed in the 2dF Galaxy Redshift Survey (2dFGRS, Magliocchetti & Porciani 2003) is presented in Figure 1.5.

Since each galaxy is associated with the host dark matter halo, one might naively ex-

pect that the galaxy distribution reflects the large-scale mass distribution in the Universe. On the other hand, if the process of galaxy formation is highly stochastic, or galaxies form only in special, preferred environments, the relation between the galaxy distribution and the matter distribution may be far from straightforward. Therefore, detailed studies of the spatial distribution of galaxies, in principle can convey information regarding both the overall matter distribution, which is strongly cosmology dependent, and the physics of galaxy formation.

Such studies, usually performed using the galaxy correlation function, have been produced for most large galaxy surveys, such as the Sloan Digital Sky Survey (SDSS, Connolly et al. 2002, Zehavi et al. 2004), 2dF Galaxy Redshift Survey (2dFGRS, Magliocchetti & Porciani 2003), VIMOS-VLT Sky Survey (VVDS, Le Fèvre et al. 2005, Pollo et al. 2006), VIMOS Public Extragalactic Survey (VIPERS, Marulli et al. 2013), and DEEP2 (Coil et al. 2006).

Thanks to this remarkable effort it is now very well established that the strength of the galaxy clustering for the general population of galaxies is only mildly evolving from intermediate redshifts $z \sim 1$ to $z \sim 0$. It has been found that galaxy clustering depends on a variety of galaxy properties like the luminosity, morphology, colour and spectral type of galaxies: luminous galaxies tend to be more clustered than faint ones and red galaxies with old stellar population are found to be more clustered than young blue ones (Norberg et al. 2002, Zehavi et al. 2004, 2012, Pollo et al. 2006, de la Torre et al. 2007, Coil et al. 2008, Quadri et al. 2008, Meneux et al. 2006, 2008, 2009, Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012).

At redshifts higher than $z \sim 2$ the situation is less clear. Various difficulties, mainly in collecting statistically significant and representative samples, need to be overcome. While a number of attempts to measure galaxy clustering have been made at early epochs, interpreting results is not straightforward, mainly because of uncertainties connected to redshift determination, small volumes covered, and different galaxy populations selected from a range of methods. Most of the measurements at high redshifts ($z > 1$) are produced using photometric surveys targeting specific classes of galaxies or applying specific observation techniques, like the BzK method (Kong et al. 2006, Lin et al. 2012) or Lyman-break galaxy (LBG) selection (Foucaud et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Savoy et al. 2011, Bielby et al. 2013). In general, within a given sample and based on angular correlation function measurements, some evidence has been presented at $z > 1$ showing that clustering also seems to depend on luminosity, stellar mass or colour as observed on lower redshift, with bright galaxies clustering more strongly than faint ones (Savoy et al. 2011) and passive galaxies clustering more strongly than star-forming galaxies at a given stellar mass (Lin et al. 2012). However the relation

of galaxy samples used in these analyses to the general population of galaxies is not well established, which makes the study of evolution of the galaxy clustering at $z > 2$ difficult. Moreover these galaxies can not be easily connected with the galaxy populations at lower redshifts, which make it difficult to conduct a consistent study of galaxy clustering evolution from high redshift to $z = 0$.

CHAPTER 2

Data sample. VIMOS Ultra Deep Survey

This thesis is based on the biggest (up to the date) and the most reliable data sample available at high redshift $z > 2$ - The VIMOS Ultra Deep Survey (VUDS, Le Fevre et al. 2014). It is a spectroscopic survey of $\sim 10,000$ galaxies performed with the VIMOS multi-object spectograph at the European Southern Observatory Very Large Telescope (Le Fèvre et al. 2003). Spectroscopic target selection for this survey is based on a *inclusive* combination of photometric redshifts and color properties. It covers a total area of 1 deg^2 in three independent fields, reducing the effect of cosmic variance, an important property for galaxy clustering measurements. The fields location and the cover area for each of these fields are summarized in Table 2.1. With ~ 6000 galaxies with reliable spectroscopic redshifts in $2 < z < 6$ this survey is the largest at these redshifts and offers the opportunity for unprecedented studies of star forming galaxy population and its distribution in large scale structures during early phases of galaxy formation and evolution.

The features of the VUDS galaxy survey, such as survey strategy, target selection, data processing and redshift measurements have been fully described in Le Fevre et al. (2014) (included at the end of this thesis). In this chapter I present the features of the VUDS survey, which are relevant in terms of galaxy clustering measurements. I also designate these aspects of the VUDS survey, which makes my study unique and important in terms of understanding the structure formation and evolution.

Table 2.1: The VUDS fields.

Field	α_{2000}	δ_{2000}	Area $[deg^2]$	Effective area $[deg^2]$	Depth
COSMOS	$10h00m01.0s$	+0.2 deg 12'40"	0.50	0.49	$i_{AB} \simeq 25$
VVDS-0.2h	$0.2h26m00.0s$	-0.4 deg 30'00"	0.31	0.25	$i_{AB} \simeq 25$
ECDFS	$0.3h32m28.0s$	-27 deg 48'30"	0.14	0.06	$i_{AB} \simeq 25$

2.1 Spectroscopic target selection

The VUDS Survey was designed to study the galaxy population beyond redshift $z \sim 2$ and up to the redshits possible in the wavelength range up to $\lambda = 9350\mathring{A}$ accessible with the VIMOS spectrograph on the VLT, corresponding to the Lyα at redshift up to $z \sim 6.6$. One of the unique features of the VUDS survey, in comparison to other observations performed at high redshift, is the target selection method, which assure that the selected galaxies are not limited to the specific type/class of galaxies. The target selection is primarily based on photometric redshifts and their probability distribution function (PDF), complemented by color-color selections and analysis of the SED shape when not already in the photo-z selection. Additionally, when the geometry of the observed slit-masks (see Chapter 3) allows it after placing the previous priority targets, a purely magnitude-selected random sample of galaxies with $23 \leq i_{ab} \leq 25$ has been added to the target list. This can be summarized as follow:

1. The priority spectroscopic targets must verify $z_{phot} + 1\sigma \geq 2.4$, and $i_{AB} \leq 25$. As degeneracies are known to occur, the sources for which either the first or the second peak in the photometric redshift probability distribution function satisfy this criterion, have been selected.

2. Sources which were not selected from the primary z_{phot} criterion (1) but satisfy one of the ugr, gri or riz LBG color-color selection criteria have been added to the target list.

3. The full SED over all filters available in ugrizYJHK was used to identify galaxies with the break in the continuum and not identified by any of the z_{phot} or LBG criteria, extending the *dropout* technique. These targets were allowed to have $K_{AB} < 24$ and

hence not necessary be brighter than $i_{AB} = 25$.

4. Finally, when space on the slit mask was available after the above selection, a random sample of galaxies with $23 < i_{AB} < 25$ has been added to the sample.

As presented in Table 2.2 $\sim 80\%$ of galaxies have been selected based on the 1st or 2nd peak of the photometric redshift probability distribution function (1). The remaining $\sim 20\%$ are about equally split between LBG color-color selection (2)(3) and the purely magnitude selected galaxies (4).

The second key feature of the VUDS Survey is the extended wavelength coverage of the VIMOS spectroscopic observations covering $3650 < \lambda < 9350\mathring{A}$, which minimizes any instrument-induced redshift desert and strongly reduces the degeneracies in redshift measurements. Finally the last key point is the integration time of $\simeq$ 14h per target, which allows reaching S/N on the continuum at $8500\mathring{A}$ of $S/N = 5$ for $i_{Ab} = 25$, and $S/N = 5$ for an emission line with a flux $F = 1.5 \times 10^{-18}$ erg/s/cm$^2/\mathring{A}$.

2.2 Redshift measurements and reliability flags

Redshift measurements were carried in the way similar to the one developed and introduced for the VVDS survey (Le Fèvre et al. 2005, Le Fèvre et al. 2013),the zCOSMOS survey (Lilly et al. 2007) and the VIPERS survey (Guzzo et al. 2014). The core engine for redshift measurement is cross-correlation of the observed spectrum with reference templates using the EZ redshift measurement code (Garilli et al. 2010). At the end of the process, each redshift measurement is assigned with a reliability flag which expresses the reliability of the measurement in the following way:

- Flag 0: No redshift could be assigned
- Flag 1: $\sim 50\%$ probability to be correct
- Flag 2: $\sim 70 - 80\%$ probability to be correct
- Flag 3: $95 - 100\%$ probability to be correct
- Flag 4: 100% probability to be correct
- Flag 9: $\sim 80\%$ probability to be correct; spectrum has a single emission line

Compromising between the number of available galaxies and the reliability of their redshift we choose for this study galaxies assigned with flag $2 - 4$ and 9. The redshift measurement for these galaxies is very reliable, with probability to be correct from 70% to 100%. We test, however, the possible influence of the fraction of galaxies with wrongly measured redshift on the correlation function measurements. The procedure and results of this tests are presented in Chapter 4.2.

Table 2.2: VUDS spectroscopic target selection criteria.

Criterion	Redshift	Limiting magnitude	Fraction of targets
(1) Photometric redshift, 1st PDF peak	$z_{phot} + 1\sigma \geq 2.4$	$22.5 \leq i_{AB} \leq 25$	58.3%
(1) Photometric redshift, 2nd PDF peak	$z_{phot} \geq 2.4$	$22.5 \leq i_{AB} \leq 25$	19.3%
(2) LBG ugr	$2.7 < z < 3.5$	$i_{AB} \leq 27$	1.9%
(2) LBG gri	$3.5 < z < 4.5$	$i_{AB} \leq 27$	1.9%
(2) LBG riz	$4.5 < z < 5.5$	$i_{AB} \leq 27$	3.6%
(3) Break SED-based	-	$K_{AB} \leq 24$	3.5%
(4) Magnitude-selected	-	$23 \leq i_{AB} \leq 25$	11.5%

2.3 Absolute magnitude and stellar mass

The VUDS sample covers a large parameter space both in the observed and the physical properties of the galaxy sample. This sample benefits from an extended multi-wavelength dataset (see Le Fevre et al. 2014). The multi-wavelength photometry was used to compute absolute magnitudes and stellar masses from SED fitting using the 'Le Phare' code (Ilbert et al. 2006), as described in details by Ilbert et al. (2013). For the VUDS survey measurements have been used 1) spectroscopic redshifts (when were available); 2) the template set presented by Bruzual & Charlot (2003) with 3 metallicities ($Z = 0.004, 0.008, 0.02$) and the Calzetti et al. (2000) extinction law; 3) two delayed star formation history models (SFH) with peaks at 1 and 3 Gyr and the exponentially declining SFHs, with nine possible τ values in the range $0.1 < \tau < 30$ Gyr.

The distributions in absolute NUV ($2300\mathring{A}$ rest) magnitude and stellar mass of the sample are presented in Figure 2.1. At redshift $z \sim 2 - 5$ the NUV absolute magnitude ranges from NUV $= -19.5$ to NUV $= -23$, while the stellar mass from $10^9 M_\odot$ to $10^{11} M_\odot$.

2.4 Advantages of the VUDS survey for galaxy clustering studies

All kinds of experimental studies strongly depend on the quality and reliability of data which are used for them. This is especially true for studies of the large-scale galaxy

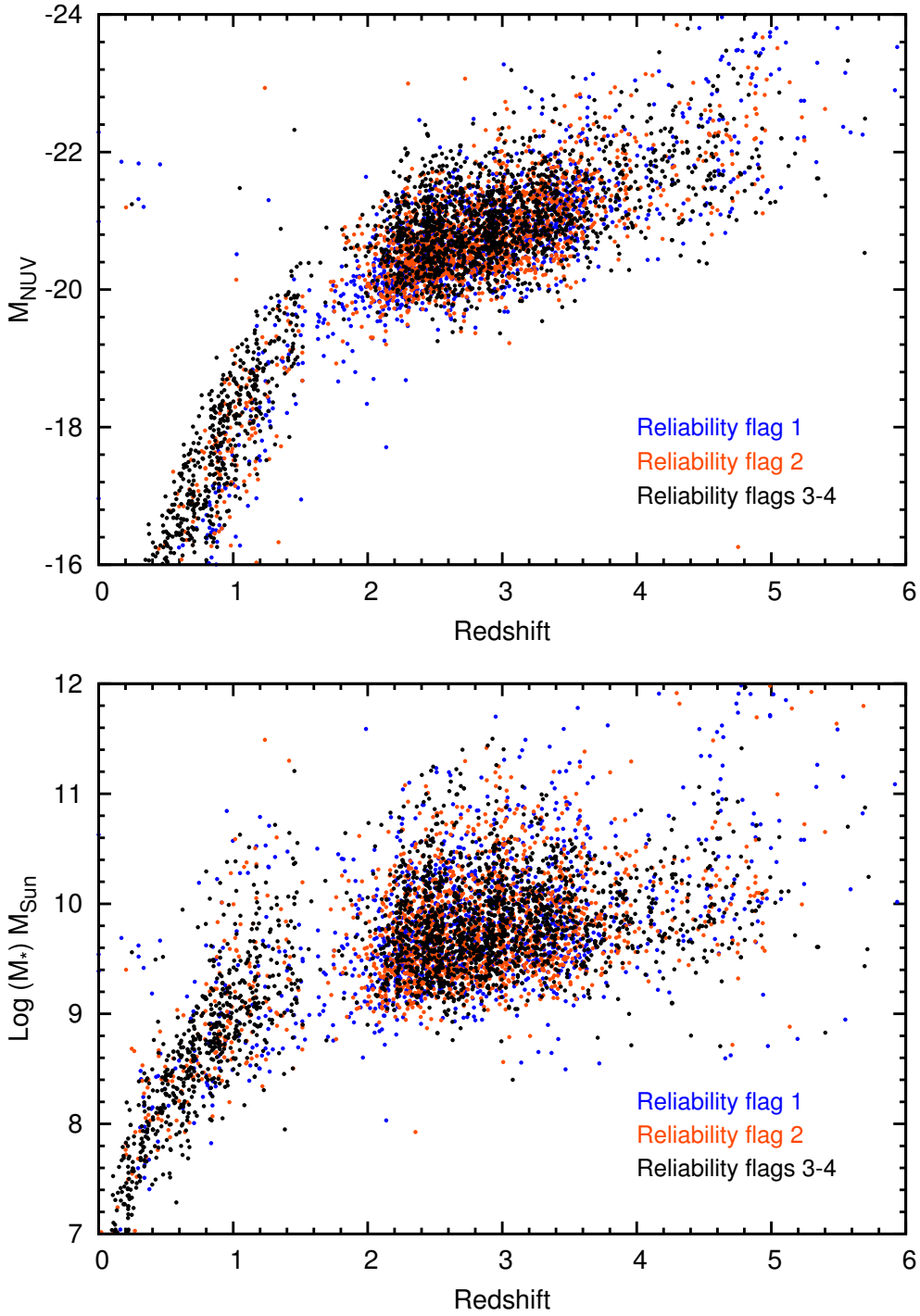


Figure 2.1: The distribution of absolute NUV magnitude (*top*) and stellar mass (*bottom*). Stellar masses are derieved from template SED fitting at the spectroscopic redshift (see text). In both figures black points indicate galaxies with reliability flag 3-4, orange flag 2, and blue flag 1.

distribution at high redshift. For these studies, in addition to quality, the richness of the available sample is also very important, because of the specific nature of quantifying methods which are used, in particular correlation function, which is a statistical tool. In this work we investigate the galaxy clustering at early epochs of galaxy formation, making use of the VUDS galaxy sample. Until now clustering studies at high redshift were very limited, due to the limitations of the available data samples, related to the reliability of their redshift measurements, covered areas, depth, target selection methods. Below, we present the uniqueness of the VUDS data in comparison to the previous galaxy surveys at redshift $z > 2$ in terms of applicability for clustering measurements.

The are several characteristics of the galaxy sample, which are crucial to create a perfect base for galaxy clustering studies. Below, we list the most important of these features:

1. **The number of galaxies, along with the survey volume and the size of the observed area**. The more galaxies we have in the sample, the larger volume they occupy and the larger area on the sky is covered by the observations, the better measurement we can expect. The main classical method used to study galaxy clustering - the correlation function - is a purely statistical tool. Then, it is important that our sample is large and representative in a statistical sense. More objects occupying larger volumes result in a more representative and robust measurement, with small error values, which allows for the detailed interpretation.

2. **High reliability of redshift measurement - spectroscopic redshifts**. A good quality measurement of a large-scale galaxy distribution at given redshift (especially $z > 2$) requires robust redshift measurements for observed galaxies. Otherwise, the correlation function may be distorted. The low accuracy of redshift measurements results (a) in contamination of a given redshift bin by galaxies located at other redshifts and (b) in a larger uncertainty of estimated galaxy positions and, consequently, lower precision in galaxy pair counts at different spatial scales. These effects introduce a "blur" of the large scale structures and result - at minimum - in the decrease of the measured clustering length with respect to the real one. This problems cannot be fully compensated using large data samples.

3. **The selection method used to specify targets for observation**. This characteristic of the survey is important in terms of interpretation of the results. Galaxy clustering is proven to be highly dependent on galaxy properties. Thus, the correlation function measurements based on the sample of very specific galaxies, i.e of certain type or colour (targeted in a certain way), are not representative. These studies of clustering properties of selected types of galaxies are important and valuable.

However, for the more robust studies, which allow for comparisons with theoretical predictions, the galaxy sample needs to represent the general population of galaxies, in order to be easily compared and related to galaxy populations at lower redshifts.

Previously published galaxy samples available at high redshift, for which clustering measurements were performed, mostly satisfy the first condition i.e. the number of objects and the depth of these surveys would assure statistically sound correlation measurement (e.g. McCracken et al. 2010 presented a large sample of 27,757 star-forming galaxies with average photometric redshift $z \sim 2$). However, these are mostly photometric galaxy samples, and even though recently they have reached an impressive accuracy in the redshift measurements, they still suffer from *catastrophic failures* - which happen when the true redshift is mis-estimated by a significant amount. The level of these *catastrophic failures*, even if low at a few percent level (Ilbert et al. 2013), can still result in large unknowns because of the shape of the $N(z)$ of flux limited samples. In particular, an error of 1% at the peak, $z \simeq 1$, of the $N(z)$ of a $i_{AB} = 25$ sample could spread galaxies with wrong photometric redshift to higher redshifts, e.g. to $z \sim 3$ where the projected galaxy density is less than 10 times lower than the $N(z)$ at peak, and can then represent several tens of percent of uncertainty. One recent example is the difficulty to distinguish $z \sim 5$ very massive objects from lower redshifts $z \sim 2$ galaxies of lower mass (Wiklind et al. 2008, Caputi et al. 2012).

Thus, the best galaxy samples for the clustering measurements are those with spectroscopically confirmed redshifts, so that they satisfy our second condition. Such surveys can assure a) that the number of falsely measured objects added or removed from the galaxy sample due to the wrong redshift measurements is reduced, and b) that the "blurr" of the large scale structure occurring due to low accuracy of position measurements is minimized. Fulfilling these conditions guarantees the robust correlation function estimation.

Until now, however, published spectroscopic galaxy surveys at redshift $z > 2$ were either severely limited in number of galaxies or they targeted certain types of galaxies, e.g. extremely massive red objects or sources selected using a Lyman-break technique which selects a star-forming galaxies (LBG e.g. Adelberger et al. 2005; LAE e.g. Ouchi et al. 2004). This leads to the third, pointed by us, feature of the 'perfect' galaxy sample for clustering measurements. The LBG samples reach ~ 2000 redshifts at $z \sim 3$, only ~ 150 at $z \sim 4$, and a few tens of galaxies beyond that (e.g. Steidel et al. 2003, Vanzella et al. 2009, Bielby et al. 2013). Samples of LAE galaxies selected with narrow band techniques and spectroscopically confirmed reach a few hundred objects beyond $z = 3$ (Ouchi et al. 2008, Kashikawa et al. 2011). The only spectroscopic and purely magnitude limited survey extisting presently, with some galaxies observed at high redshift is the VVDS (VIMOS VLT Deep Survey). It contains 1000 galaxies at $z > 2$ (Le Fevre et al.

2013).

In Figure 2.2 the distribution of VUDS galaxies with respect to other high redshift surveys is presented. The large number of spectroscopically confirmed redshifts in the range $2 < z < 5$ makes the VUDS Survey unique, especially for the type of studies presented in this thesis, i.e galaxy clustering analysis. In contrast to other available galaxy samples available at high redshift the VUDS satisfies all three conditions listed in the beginning of this section. Being a spectroscopic redshift survey the VUDS allows to provide a very reliable redshift measurement (2). It contains $\sim 10,000$ objects covering $1deg^2$ - large number of galaxies and relatively big area (1). Additionally, the VUDS selecting method do not target only galaxies of specific type (3), since the target selection is based mainly on photometric redshift (see Sec. 2.1), but rather a general, representatively sampled, population.

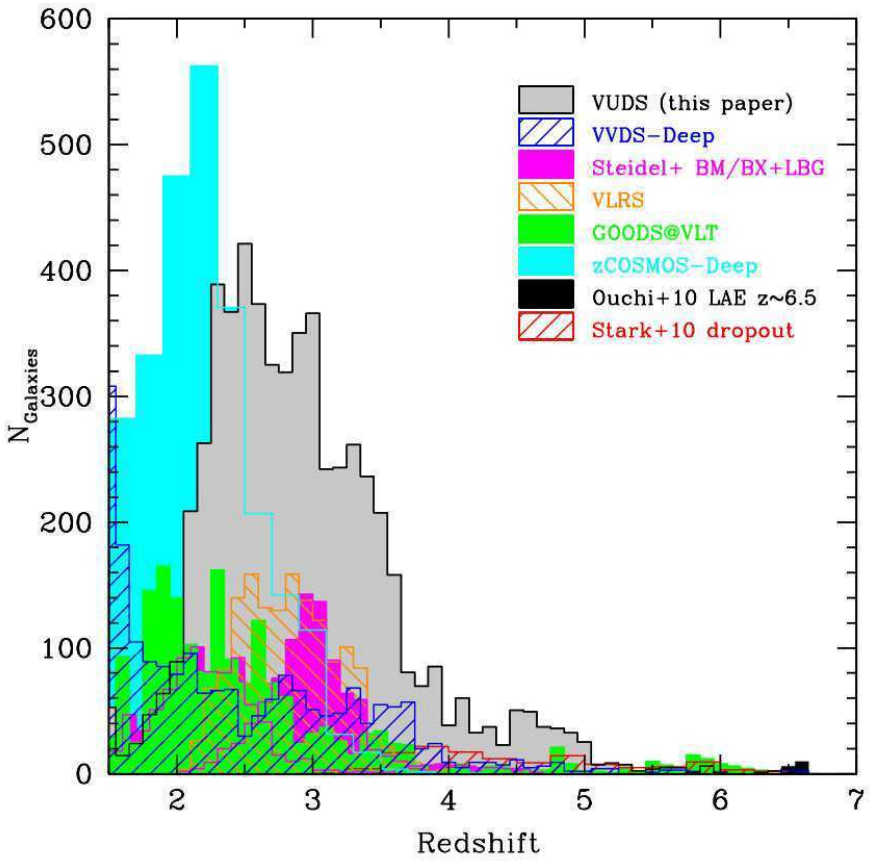


Figure 2.2: Comparison of the redshift distribution of the VUDS survey (presented in grey) with other published spectroscopic surveys at $z > 2$. Figure adapted from Le Fevre et al. (2014).

CHAPTER 3

Quantifying the large-scale distribution of galaxies

Since the very first galaxy surveys have revealed the complex structure of the Universe, the mapping of its evolution has became an important part of cosmology. The investigations revealed that the Universe is composed of very dense regions like clusters and filaments, but also almost empty void regions. Additionally, the underlying dark matter structure and its evolution is shown to follow the visible baryonic matter, though the luminous matter is *biased* in relation to the dark matter distribution. It is still unclear of how exactly this baryonic - dark matter relation looks like, especially in the early stages of galaxy formation, and how it changes through time.

The correlation function is a commonly used tool to describe how galaxies are clustered as a function of scale, and allows to put constrains on the evolution of galaxies (Kauffmann et al. 1999, Zehavi et al. 2011). It is based on the very simple idea of measuring the probability of finding two galaxies at a given separation (Peebles 1980).

The galaxy correlation function can be interpreted by introducing two kinds of approximation. The first, traditional and the most extensively used is based on a power-law function $\xi(r) = (r/r_0)^{-\gamma}$ with two free parameters r_0 - correlation length, which describe how strongly galaxies are clustered and the slope γ (Davis & Peebles 1983). The second, more recent and detailed approximation is based on halo occupation models (Zehavi et al. 2004). In this framework the correlation function is build from two components (terms), which have their influence on different scales. The *One-halo* term dominates on small scales ($< 1-2$ h^{-1}Mpc) and describes clustering of galaxies which reside within one dark matter halo. While the *two-halo* term fully describes large scale (> 5 h^{-1}Mpc) galaxy clustering between different halos. The transition scales are modelled by the interplay of

these two terms.

In the following chapter we present, in details, the most important features of the galaxy correlation function measurements, such as: the correlation functions estimators, corrections for observational limitations, and the parametrisations used to quantify the results.

3.1 Correlation function measurements

The correlation function describes the codependency (correlation) of two variables separated in space or time. In astronomy, it is therefore, the most basic statistical tool, used to measure and investigate the dependencies of separated galaxies, i.e galaxy clustering.

In its simplest, two-point form, the correlation function of two random variables $X(s)$ and $X(t)$ located in two different points in space, s and t, can be defined as:

$$C(s,t) = corr(X(s), X(t)). \tag{3.1}$$

Here, *corr* is the 'Pearson's correlation coefficient', also called simply 'the correlation coefficient'. It is a measure of the strength and direction of the linear relationship between two variables, defined as the covariance of the variables divided by the product of their standard deviations.

This definition can be naturally extended to the higher orders, if required. Then the n-point correlation function can be defined as:

$$C(s_1, s_2, \cdots, s_n) = corr\left(X(s_1), X(s_2), \cdots, X(s_n)\right) \tag{3.2}$$

assuming that every random variable $X(s_j)$ has only one component.

However, it is the simple two-point correlation function $\xi(r)$, tracing the amplitude of galaxy clustering as a function of scale, which is the most commonly used quantitative measure of the large scale structure. It is defined as the probability, above random, of finding a pair of two galaxies (in general objects) separated by a given distance r (Peebles 1980) and it can be presented as:

$$dP = \left[1 + \xi(r)\right] \rho dV_1 dV_2, \tag{3.3}$$

where dP is the probability of finding two independent galaxies in two infinity small [1] volumes dV_1 and dV_2 separated by r, while galaxy number density is ρ.

When galaxies are not clustered, hence they are randomly distributed, then $\xi(r) = 0$ for every r. Also when galaxies are anty-correlated then $-1 < \xi(r) < 0$.

[1] The probability of finding two galaxies within the same volume is negligible

3.1.1 General correlation function estimators

Although the idea behind the correlation function is relatively simple, its definition, mentioned earlier, cannot be straightforwardly applied to compute the correlation function from the real data samples. This is caused by the limitations of galaxy surveys themselves. In principle, to retrieve the true correlation function one would need an unlimited survey which covers the whole sky, and includes all the galaxies. Naturally, in practice such surveys are unreachable. Thus, correlation function estimators must be defined to minimize negative effects related to the limited number of objects and limited areas covered by available surveys.

Some estimators, to measure correlation function $\xi(r)$, have been proposed through the history. One of the earliest, on which the following ones were based, was presented by Davis & Peebles (1983). According to them, the correlation function $\xi(r)$, at separation r, can be computed as the ratio between the number of observed galaxy pairs, with respect to the number of pairs between galaxies in data sample and objects in the 'random catalogue'.

$$\xi_{DP}(r) = \frac{n_R}{n_G}\frac{GG(r)}{GR(r)} - 1, \tag{3.4}$$

where, $GG(r)$ is the number of distinct galaxy-galaxy pairs at separation r in the data catalogue, $GR(r)$ is the number of pairs between data and random catalogue, and n_G and n_R are the mean number density of galaxies in the data and random catalogues.

This definition evolved later to the similar estimator, but suffering from smaller statistical errors, proposed by Hamilton (1993).

$$\xi_H(r) = \frac{GG(r)RR(r)}{(GR(r)^2)} - 1 \tag{3.5}$$

where $RR(r)$ is the number of pairs of objects at separation r within 'random catalogue'.

Presently, however, the estimator most commonly used, thanks to its well established capability to minimize the above mentioned observational limitations of the data, is the Landy & Szalay (1993) estimator:

$$\xi_{LS}(r) = \left(\frac{n_R}{n_G}\right)^2 \frac{GG(r)}{RR(r)} - \left(\frac{n_R}{n_G}\right)\frac{2GR(r)}{RR(r)} + 1, \tag{3.6}$$

which is the one used in this study.

All these estimators, highly depend on the catalogue of randomly distributed objects, which must satisfy several conditions (Bernardeau et al. 2002):

1. First of all the 'random catalogue' has to contain much more objects than data catalogue which will be used to compute the correlation function ($N_R \gg N_G$).

Too few random objects introduce additional distortions into $\xi(r)$ measurements, especially on the small scales ($r_p < 3h^{-1}$Mpc).

2. Secondly, the 'random catalogue' has to reflect almost ideally the geometrical and redshift properties of the corresponding galaxy sample - it should cover the same sky area (volume) and the same redshift range.

For each estimation of the correlation function in this study, the creation of the 'random catalogue' has been treated carefully. Details about the numerous tests performed to measure the influence of the geometrical shape and the number of generated objects in the 'random catalog', on the correlation function can be found in Section 3.4.2. All of them allowed to reduce significantly the negative effects related to the survey limitations.

3.1.2 Redshift-space and projected real-space correlation function

Apart from the survey limitations, mentioned above, there is also another problem one has to face while measuring the real space correlation function - the peculiar velocities of galaxies on the redshift-space plane.

On small spatial scales ($r \lesssim 1h^{-1}$Mpc), within groups and clusters, galaxies have large random motions relative to each other, due to the gravitational field of the structures. This leads to the situation, when even though galaxies are at the same or similar distance r form the observer, their measured redshift z differs significantly, due to the peculiar motions. In other words, the galaxy spatial distribution in redshift-space (i.e. when redshift is interpreted as a cosmological redshift and only used to compute distance) is distorted with respect to the real-space galaxy distribution (i.e. being a function of physical distance). The corresponding distortions are known as the 'coherent infall' and "fingers of God" (Kaiser 1987).

This situation complicates the measurements of the two-point correlation function $\xi(r)$ in real-space, because, as the result of peculiar velocities, the redshift-space separations differ from true physical comoving separations between two galaxies. Therefore various distortions are introduced to the $\xi(r)$ measurement.

A commonly applied method to eliminate contamination from peculiar velocities is based on the fact, that they affect only the redshift and do not have any influence on the position on the sky. One can eliminate redshift distortions by dividing redshift-space separations into two components - parallel π and perpendicular r_p to the line of sight, defining the redshift-space correlation function $\xi(r_p, \pi)$.

This is achieved in the following way. For two objects at given redshifts z_1 and z_2 with observed radial velocities v_1 and v_2 we can define (Fisher et al. 1994) the line of sight vector $\overrightarrow{l} \equiv (v_1 + v_2)/2$ and the redshift difference vector $\overrightarrow{s} = v_1 - v_2$ and also:

$$r_p \equiv \left(\overrightarrow{s} \cdot \overrightarrow{s} - \pi^2\right)^{1/2} \tag{3.7}$$

$$\pi = \frac{\overrightarrow{s} \cdot \overrightarrow{l}}{|\overrightarrow{l}|} \tag{3.8}$$

The redshift-space correlation $\xi(r_p, \pi)$ function can be then estimated by using e.g. the Landy & Szalay estimator (Eq.3.6).

Finally, integrating $\xi(r_p, \pi)$ along the line of sight gives us a projected correlation function $w_p(r_p)$, which is a counterpart of the real-space correlation function, free from the redshift-space distortions (Davis & Peebles 1983):

$$w_p(r_p) = 2\int_0^\infty \xi(r_p, \pi)d\pi = 2\int_0^\infty \xi_0\Big[(r_p^2 + y^2)^{\frac{1}{2}}\Big]dy \tag{3.9}$$

Here, y is the real-space separation along the line of sight, and ξ_0 stands for the real space correlation function computed for $r = \sqrt{r_p^2 + y^2}$. In practice, the upper integral limit π_{max} has to be finite in order to avoid adding noise to the result. After performing a number of tests, we decided to use $\pi_{max} = 20\ h^{-1}$Mpc.

In most cases, especially in the local Universe, $\xi(r)$ is well described by a power law function $\xi(r) = (r/r_0)^{-\gamma}$, where r_0 and γ are correlation length and slope, respectively. With this parametrization, the integral in equation 3.9 can be computed analytically and $w_p(r_p)$ can be expressed as:

$$w_p(r_p) = r_p\Big(\frac{r_0}{r_p}\Big)^\gamma \frac{\Gamma(\frac{1}{2})\Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})}, \tag{3.10}$$

where Γ is the Euler's Gamma Function.

A more detailed description of the shape of the real space correlation function can be done e.g. in the framework of the Halo Occupation Distribution (HOD) Models, which we discuss in Section 3.2.

3.2 Halo Occupation Distribution (HOD) model

In practice $\xi(r)$ is usually well fitted by the power law. However, there is no strict theoretical reason to force the galaxy correlation function to assume the power law shape. And indeed, recent measurements show deviations from the power law shape of galaxy

correlation functions. The strongest ones are observed on small scales, especially for the most luminous galaxies (Coil et al. 2006, Pollo et al. 2006, Zehavi et al. 2011), while probably the most famous ones are baryonic acoustic oscillations (BAOs) at large scales (Eisenstein et al. 2005, Percival et al. 2010).

The small scale behaviour of the correlation function of galaxies can be well interpreted in the framework of the halo occupation distribution (HOD) models, which are based on the relations between the distribution of dark matter and galaxies. The HOD models describe bias in terms of the probability $P(N|M)$ that a dark matter halo of a mass M contains N galaxies of a given type. Recently this approach has been very successfully used to model the shape of the two-point correlation function.

In this work we also apply a halo occupation model, which closely follows the current commonly used description, so that the results can be easily compared to the results taken from the literature. Below we describe the most important features and components of the HOD model we use, while for the general overview of the HOD philosophy we encourage the reader to look into the review paper by Cooray & Sheth (2002). A very thorougful description is also given by Coupon et al. (2012).

In the HOD framework the correlation function can be split into two components. The *one-halo* component $\xi^{1h}(r)$ dominates on scales smaller than the size of one dark matter halo (usually $< 1 - 2$ h^{-1}Mpc), while the *two-halo* component $\xi^{2h}(r)$ dominates on larger scales. The *one-halo* term arises from the pairs of galaxies located within the same halo, while the *two-halo* term is built by the pairs of galaxies hosted by different halos. Consequently, the correlation function can be written as:

$$\xi(r) = 1 + \xi^{1h}(r) + \xi^{2h}(r). \tag{3.11}$$

The properties of the first term, i.e. $\xi^{1h}(r)$ can be described with the use of the halo occupation model which tries to follow the distribution of galaxies located in one dark matter halo. The second term, $\xi^{2h}(r)$ depends on the statistical properties of the large scale dark matter density field and the distribution of dark matter haloes with respect to this density field. Detailed properties of these two terms are described in the following sections.

Number of galaxies occupying the dark matter halo

The most important part of the halo occupation model is the parametrized halo occupation function $\langle N_g|M\rangle$ i.e. the number of galaxies per halo. We parametrized it in the way used by Zehavi et al. (2005b) and motivated by Kravtsov et al. (2004). In particular, we

express the halo occupation function, as a sum of a central galaxy and satellite galaxies.

$$\langle N_g|M\rangle = \langle N_c|M\rangle + \langle N_s|M\rangle \tag{3.12}$$

The mean occupation function for central galaxies $\langle N_c|M\rangle$ is represented by a step function, while the mean halo occupation function for satellite galaxies $\langle N_s|M\rangle$ is approximated by a Poisson distribution with the mean being a power-law:

$$\begin{aligned} \langle N_g|M\rangle &= 1 + \left(\frac{M}{M_1}\right)^{\alpha} \quad \text{for} \quad M > M_{min} \\ &= 0 \qquad\qquad\qquad\quad \text{otherwise,} \end{aligned} \tag{3.13}$$

where M_{min} is the minimum mass needed for a halo to host one central galaxy, and M_1 is the mass of a halo having on average one satellite galaxy, while α is the power law slope of the satellite mean occupation function.

Our purpose was to obtain the best fit HOD with as high accuracy as possible. As one can notice the HOD model used in this work is simplified in comparison with the ones widely used at lower redshift ranges, e.g. by Zehavi et al. (2011) for fitting the SDSS data. However, based on the available statistic of the VUDS sample, the three parameter model appears to be the best solution in order to retrieve the robust measurements of M_{min} and M_1. A similar approach was taken by Abbas et al. (2010) in the case of the measurements from the similarly numerous VVDS surveys at $z \sim 1$ Universe (Abbas et al. 2010). Since the details of the interplay between galaxy evolution and large scale structure development at epochs as early as the ones probed by the VUDS are still poorly understood, it is reasonable to keep the description simplified and do not disperse our interpretation introducing too many free variables. The main results presented in this study essentially should not differ also when more complex HOD models (f.ex the five parameter model proposed by Zheng et al. 2005) are applied.

Small scales: *One-halo term* ξ_{1h}

The *one-halo* term is considered for the small separations scales $r < 1 - 2h^{-1}$Mpc. It describes the part of the correlation functions dominated by the pairs of galaxies located within the same dark matter halo, when the total number of $\langle N_g|M\rangle$ galaxies are placed in this halo. Since the halo occupation function is expressed by the sum of central and satellite galaxies it is also common practice to comprise *one-halo* term into the central-satellite $\xi_{cs}(r)$ and satellite-satellite $\xi_{ss}(r)$ terms.

$$1 + \xi_{1h}(r) = 1 + \xi_{cs}(r) + 1 + \xi_{ss}(r) \tag{3.14}$$

with the two components given by:

$$1 + \xi_{cs} = \int_0^\infty n(M) \ \frac{\langle N_c|M\rangle\langle N_s|M\rangle}{n_g^2} \ \frac{\rho(r|M)}{M} \ dM \tag{3.15}$$

$$1 + \xi_{ss} = \int_0^\infty n(M) \ \frac{\langle N_s|M\rangle^2}{2n_g^2} \ \frac{\lambda(r|M)}{M^2} \ dM \tag{3.16}$$

The number density of galaxies n_g within this framework is represented as:

$$n_g = \int_0^\infty n(M) \ \langle N_g|M\rangle \ dM \tag{3.17}$$

Naturally, in order to present the complete galaxy-dark matter relation, the properties of the dark matter halo itself need to be modelled additionally to the galaxy distribution. From this point of view the most important properties of the dark matter halo are (1) the halo mass function $n(M, z)$, for which we adopted the fitting formula proposed by Tinker et al. (2008) and (2) the halo density profile $\rho(r|M)$ for which we assumed the form described by Navarro, Frenk, & White (1997) (its auto-convolution $\lambda(r|M)$ was taken from Sheth et al. 2001).

1. The mass function, which represents the halo number density per unit of mass, can be parametrized as (Tinker et al. 2008):

$$n(M) = f(\sigma) \ \frac{\overline{\rho_m}}{M} \ \frac{d\ln\sigma^{-1}}{dM} \tag{3.18}$$

The function $f(\sigma)$ is expected to be universal to the changes in redshift and cosmology. It is parameterized as:

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right]\exp(-c/\sigma^2) \tag{3.19}$$

for $\Delta = 200$ and redshift $z = 0$ parameters are set as $A_0 = 0.186$, $a_0 = 1.47$, $b_0 = 2.57$, $c_0 = 1.19$, and $\alpha = 0.01$. Based on the measurements of the mass function of dark matter halos in a large set of collisionless cosmological simulations of flat ΛCDM cosmology, preformed by Tinker et al. (2010), the first three parameters are expected to evolve with redshift and are allowed to vary as a power law of $(1 + z)$.

$$\begin{aligned} A(z) &= A_0(1+z)^{-0.14} \\ a(z) &= a_0(1+z)^{-0.06} \\ b(z) &= b_0(1+z)^{-\alpha} \end{aligned} \tag{3.20}$$

where the zero subscript indicates the value set for $z = 0$.

2. To describe the dark matter halo density profile the NFW formalism has been used (Navarro et al. 1997) in the form:

$$\rho_h(r|M) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2} \tag{3.21}$$

where scale radius r_s and corresponding density at this radius ρ_s are given by:

$$\begin{aligned} r_s^3 &= \frac{3M}{4\pi\overline{\rho_m}\Delta c^3} \\ \rho_s &= \frac{\overline{\rho_m}\Delta c^3}{3\Omega_m(z)[\ln(1+c) - c/(1+c)]} \end{aligned} \tag{3.22}$$

In these equations:

- $\overline{\rho}_m = \overline{\rho}_0(1+z)^3$ is the mean density of the Universe at the epoch of analysis.
- $\overline{\rho}_0 = 2.775 \times 10^{11}(\Omega_{m,0}h^2)[M_\odot Mpc^{-3}]$ is the mean density of matter at present day.
- $\delta_c = 1.68647$ is the initial average overdensity of the region which will collapse at given redshift.
- $\Delta = 200$ is the critical overdensity for halo virialisation.
- $\sigma(M)$ is the variance in the initial density fluctuation field when smoothed with a tophat filter of scale $R = (3M/4\pi\overline{\rho}_m)^{1/3}$
- c is the concentration parameter, for which we assume (Bullock et al. 2001):

$$c(M,z) = \frac{11}{1+z}\left[\frac{M}{M_\star}\right]^{-0.13} \tag{3.23}$$

Here $M_\star$ is the characteristic mass scale for which $\sigma(M_\star) = \delta_c(z=0) = 1.68647$.

Large scales: *Two-halo term* ξ_{2h}

The *two-halo* term $\xi_{2h}(r)$ is dominated by the pairs of galaxies located in separated halos, which implies that it fully describes the correlation function at large scales ($r > 5h^{-1}$Mpc). However, while modelling the *two halo* term it is necessary to remember about the scales on which *one-halo* transits into *two-halo* ($2 < r < 5h^{-1}$Mpc), where ξ_{1h} drops rapidly with r while ξ_{2h} rises. It is important to ensure that *two-halo* pairs indeed come from separated independent halos. This is regulated by the halo exclusion (Zheng 2004). Since the size of the dark matter halo can be related to its mass, the halo mass range allowed for a given separation r, can be limited to the maximum mass $M_{lim}(r)$ for which halos

with central galaxies separated by r do not overlap. With this implementation of halo exclusion the *two-halo* term can be derived in Fourier space by taking:

$$P_{2h}(k,r) = \left[\int_0^{M_{lim}(r)} n(M)\ b_h(M,r)\ \frac{\langle N_g|M\rangle}{n'_g(r)}\ u(k|M)\ dM\right]^2\ P_m(k) \tag{3.24}$$

where for the halo mass function $n(M)$ we adopt the formula proposed by Tinker et al. (2008) as described in the previous section (3.18); $u(k|M)$ is the Fourier transform of the dark matter halo profile $\rho_h(r|M)$ (NFW profile, Navarro et al. 1997, Eq. 3.21); $P_m(k)$ is the matter power spectrum in the nonlinear form given by Smith et al. (2003); the restricted number density n'_g is given by:

$$n'_g = \int_0^{M_{lim}(r)} n(M)\langle N_g|M\rangle dM \tag{3.25}$$

To compute the upper integration limit $M_{lim}(r)$ the method proposed by Tinker et al. (2005) has been followed. The restricted number density given by Eq. 3.25 have been matched with the result from the following expression:

$$n'^2_g = \int_0^{\infty} n(M_1)\langle N|M_1\rangle dM_1 \int_0^{\infty} n(M_2)\langle N_g|M_2\rangle P(x) dM_2 \tag{3.26}$$

Here $x = r/(R_{vir1} + R_{vir2})$ represents the ratio between the halo separation and the sum of virial radii of two halos, so $P(x)$ essentially represents the probability of two halos to not overlap. Tinker et al. (2005) found that this probability is well approximated by $P(y) = (3y^2 - 2y^3)$ in the range $0 \leq y \leq 1$, and it takes values of $P(y) = 0$ and $P(y) = 1$ for $y < 0$ and $y > 1$ respectively, if $y = (x - 0.8)/0.29$.

The *two-halo* term depends strongly on the estimation of the large-scale dark matter halo bias $b_h(M, z, r)$, for which the formula proposed by Tinker et al. (2010) has been used:

$$b(\nu) = 1 - A\frac{\nu^a}{\nu^a + \delta_c^a} + B\nu^b + C\nu^c \tag{3.27}$$

here $\nu = \delta_c/\sigma(M)$ and fitting parameters obtained for $\Delta = 200$ as $A = 1.0$, $a = 0.13$, $B = 0.183$, $b = 1.5$, $C = 0.27$, and $c = 2.4$. According to Tinker et al. (2005) at separations $r \leqslant 3h^{-1}$Mpc the finite size of the halos must be taken into account, while computing halo bias. They proposed parametrizing this scale dependence by the amplitude of the nonlinear matter correlation function in following way, which has been adopted for our study:

$$b_h^2(\nu, r) = b_h^2(\nu)\ \frac{[1 + 1.17\ \xi_m(r)]^{1.49}}{[1 + 0.69\ \xi_m(r)]^{2.09}} \tag{3.28}$$

where the $b_h(\nu)$ is the large scale bias given by Eq. 3.27 and the matter correlation function ξ_m is the fourier transformed matter power spectrum P_m.

Finally the *two-halo* term P_{2h} can be transformed to the real-space correlation function by taking:

$$\xi'_{2h}(r) = \frac{1}{2\pi^2} \int_0^\infty P_{2h}(k,r)\ k^2\ \frac{\sin kr}{kr}\ dk \tag{3.29}$$

where $\xi'_{2h}(r)$ denotes that we have calculated the *two-halo* term for a restricted range of the halo masses, limited by $M_{lim}(r)$. This could be converted to the $\xi_{2h}(r)$ for the entire halo (and therefore galaxy) population by:

$$1 + \xi_{2h} = \left(\frac{n'_g(r)}{n_g}\right)^2 (1 + \xi'_{2h}) \tag{3.30}$$

Deduced parameters

By using received best-fitting parameters it is possible to obtain quantities describing halo and galaxy properties, weighted by the number of galaxies, like the average host halo mass M_h

$$\langle M_h \rangle(z) = \int M\ n(M,z)\ \frac{\langle N_g|M\rangle}{n_g(z)}\ dM, \tag{3.31}$$

the mean galaxy bias b_g

$$b_g(z) = \int b_h(M)\ n(M,z)\ \frac{N(M)}{n_g(z)}\ dM, \tag{3.32}$$

the fraction of central galaxies per halo is

$$f_c(z) = \int n(M,z)\ \frac{N_c(M)}{n_g(z)}\ dM \tag{3.33}$$

consequently the satellite fraction:

$$f_s(z) = 1 - f_c(z), \tag{3.34}$$

In this equations n_g is the number density of galaxies given by Eq. 3.17, $n(M,z)$ is the dark matter mass function is given by Eq. 3.18, $b_h(M,z)$ is the large-scale halo bias described by Eq. 3.27, and $\langle N_g|M\rangle$ stands for the halo occupation function (Eq.3.14).

3.3 The VUDS mock catalogues

To estimate uncertainties of the correlation function measurements performed in this study and to test robustness of our clustering measurements (see Sec. 3.4.2) we make use of a large number of mock galaxy samples, which are designed to mirror the VUDS sample in the range $2 < z < 5$ in possibly the most realistic way.

The 66 independent mock samples have been created based on a method which includes the halo occupation distribution (HOD) and the Stellar-to-Halo Mass Relation (SHMR). This method is the same as developed and introduced by de la Torre et al. (2013), and we refer the reader to the details presented in this paper. We provide below a brief summary of its most important features.

In the first step the mock catalogues containing stellar masses have been created from the MultiDark N-body simulation (Prada et al. 2012) and Pinocchio halo lightcones (Monaco et al. 2002). We followed the stellar mass to halo mass ratio (SMHR) approach based on the assumption of a monotonic relation between halo/subhalo masses and the stellar masses of the galaxies associated with them. We first populated the haloes in lightcones with subhaloes. For this we randomly distributed subhaloes around each distinct halo following a NFW profile (Navarro et al. 1997), so that their number density satisfies the subhalo mass function proposed by Giocoli et al. (2010). Then we assigned a galaxy to each halo and subhalo, with a stellar mass given by the SHMR of Moster et al. (2013).

The procedure which has been followed in the next step of mock catalogues creation is similar to the one used in the VVDS, zCOSMOS and VIPERS surveys also based on VIMOS observations as the VUDS survey (Meneux et al. 2006, Iovino et al. 2010, de la Torre et al. 2011b, 2013). In order to obtain fully realistic VUDS mock catalogues we add the detailed survey selection criteria (see Chapter 2 and Le Fevre et al. 2014). This procedure in the end produces mock parent photometric galaxy catalogues. We then apply the slit-positioning algorithm (SSPOC) with the same settings as used for the VUDS survey (Bottini et al. 2005). This allows us to reproduce the VUDS footprint on the sky for a proper simulation of the small-scale angular incompleteness in the mocks.

As a result of this procedure, 66 realistic mock galaxy catalogues have been produced. They contain the detailed survey completeness function and observational biases of the VUDS survey.

3.4 Optimization of correlation function measurements based on VUDS survey

To perform ideal galaxy clustering evolution study one would need to estimate the correlation function measurements for data from the galaxy survey which would cover the whole sky and include all galaxies. Naturally we don't dispose such a survey yet. The influence of the limited area and relatively small number of galaxies can be minimized by using the proper correlation function $\xi(r)$ estimator (see Section. 3.1.1 for discussion). There are also other, more sophisticated, limitations related to certain selection methods

and instruments used to perform the galaxy survey, which must be addressed in order to perform a reliable study, because all of them have been found to distort the correlation function measurements up to some level. In case of the VUDS survey, which has been used in this work, these sophisticated correlation function distortions are mostly related to the shape of the VIMOS multispectrograph, which carried the observations.

In the following sections we present the correction scheme developed in order to maximize robustness of the correlation function measurements performed in this study.

3.4.1 Real-space correlation function from multiple independent fields

In this study the dataset comes from the observations performed in two independent VUDS fields. Therefore it was necessary to find the most efficient and robust way to compute combined correlation functions. There are two possible methods to estimate the real-space correlation function $w_p(r_p)$ from multiple fields.

In the first method $\xi(r_p, \pi)$ is measured for each field separately (e.g. by using Eq. 3.6). Then, the final $\xi^{final}(r_p, \pi)$ for both fields is obtained as the average weighted by the number of galaxies within each area:

$$\xi^{final}(r_p, \pi) = \frac{1}{N_{gal}^{tot}} \sum_{i=0}^{n_{field}} N_{gal,i} \xi_i(r_p, \pi), \tag{3.35}$$

where N_{gal}^{tot} is the total number of galaxies within the whole sample. Such a method allows for a direct estimation of cosmic variance between the fields, but - because of a relatively small number of objects in each r bin in each computation for every single field - it results in a higher noise, especially on the small scales.

In the second method, the final $\xi(r_p, \pi)$ is computed for all fields simultaneously, and the differences in size and numbers between the fields are accounted for by an appropriate weighting scheme. In particular, each pair is weighted by the number of galaxies per unit volume for the given field:

$$\xi^{final}(r_p, \pi) = \frac{\sum_{i=0}^{n_{field}} w_i \cdot \Big(GG_i(r_p, \pi) - 2GR_i(r_p, \pi) + RR_i(r_p, \pi) \Big)}{\sum_{i=0}^{n_{field}} w_i \cdot RR_i(r_p, \pi)}, \tag{3.36}$$

where $w_i = (N_{gal,i}/V_i)^2$.

In the case of the VUDS data, both methods give nearly identical final results of $w_p(r_p)$. As expected, the first method results in a slightly noisier final correlation function

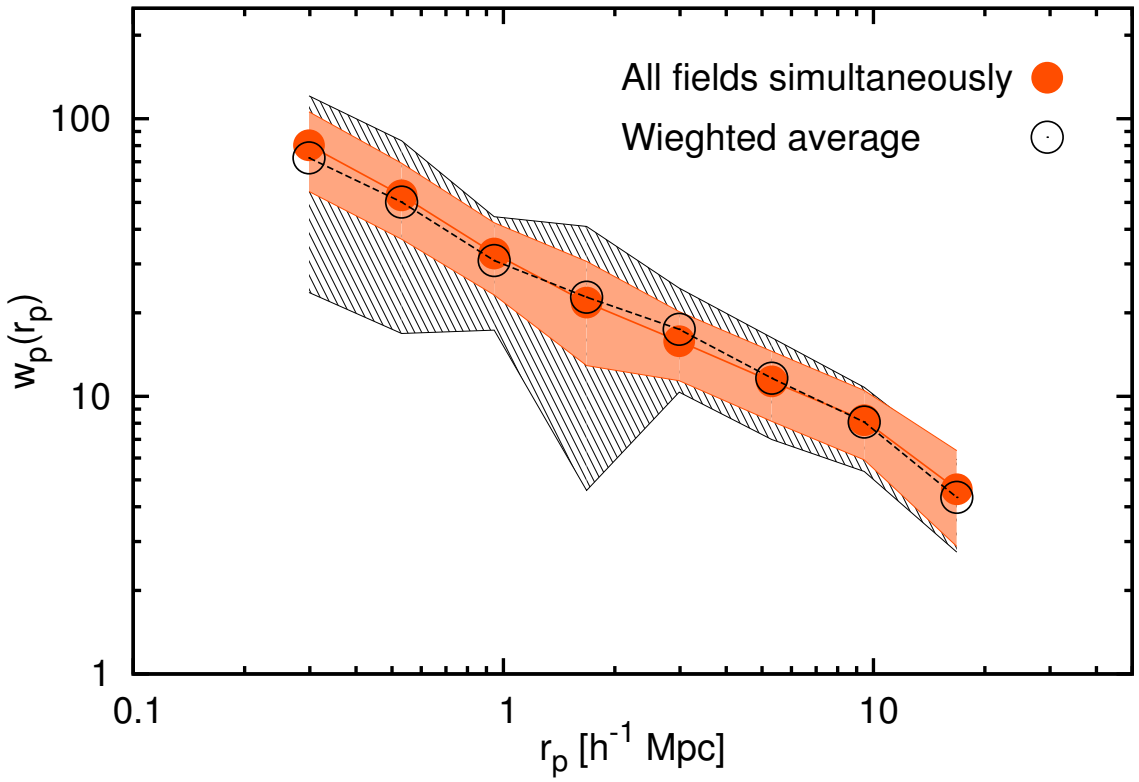


Figure 3.1: The real-space correlation function $w_p(r_p)$ measured for data sample with redshift $2 < z < 5$, spread into two independent VUDS fields. The open black circles and the black shaded area represent $w_p(r_p)$ measurement performed by using weighted average method (i.e. Eq. 3.35) and associated uncertainties, respectively. The orange colour indicates the results for the combined $w_p(r_p)$, but using the second listed method (i.e. using Eq. 3.36).

measurement, and, consequently, larger uncertainties, especially on small scales. The representative example is presented on Figure 3.1.

For most of the computations presented in this work we decided to use the second method, i.e. the composite correlation function given by Eq. 3.36. The first method, however has been used to demonstrate the effect of the cosmic variance between two VUDS fields.

3.4.2 Correlation function measurements details

Correction scheme

In real galaxy surveys, there exists a number of observational biases which may influence the clustering measurements. These biases are related to the construction of the survey. Both the parent photometric catalogue, from which target galaxies are selected, and the final spectroscopic catalogue may be the source of these biases. Below, the most important survey features that cause the biases are listed. Next, we describe the corrections applied to minimize their effects. Since the VUDS survey is performed on the same multi-spectrograph VIMOS as the VVDS, zCOSMOS and VIPERS surveys, the correction scheme adopted here is based on the same methodology as the ones used for the other VIMOS-based surveys in the past. In particular,the main corrections used here have been proposed and fully described by Pollo et al. (2005) and de la Torre et al. (2011a).

The main observational biases in the VUDS data can be listed as follows.

1. In the VUDS survey some of the areas are excised from the observations, due to the VIMOS lay-out (see Fig. 3.2). The field of view of this instrument consists of four quadrants separated by $2'$ gaps. No galaxies are observed in these gaps between quadrants, and that influences the pair counts.

2. Obviously, not all galaxies from the photometric target candidate sample can be spectroscopically observed. Each galaxy spectrum occupies a certain area - a slit in a form of an elongated rectangle of the width of $\sim 1''$ (as presented on Figure 3.4)- on the CCD, which does not permit too close galaxies to be targetted simultaneously. It means that the galaxies from the parent photometric sample must be chosen is specific way to effectively allocate as many as possible 'spectral slits' for each observation. In the case of the VIMOS observations, the slit allocation is performed automatically by the Super-SPOC code (Bottini et al. 2005). As a result, the spectroscopically observed sample is not a random representation of the parent photometric sample, and the introduced bias is relatively complex, especially on the small scales.

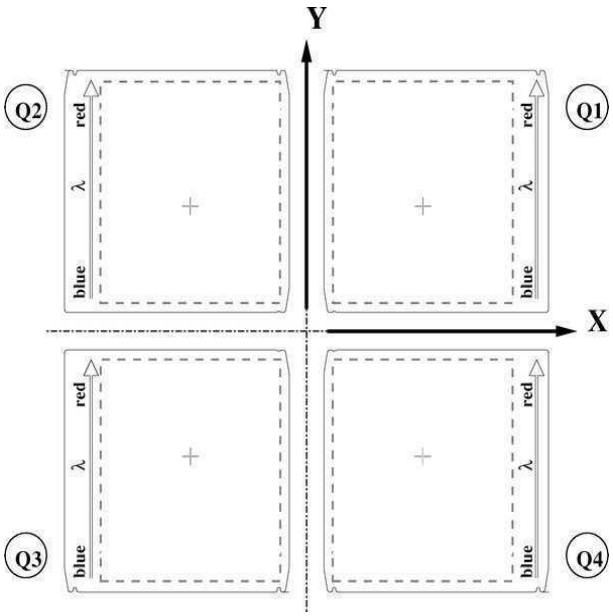


Figure 3.2: The lay-out of VIMOS multiobject spectograph at the European Southen Observatory Very Large Telescope. Picture comes from Bottini et al. (2005).

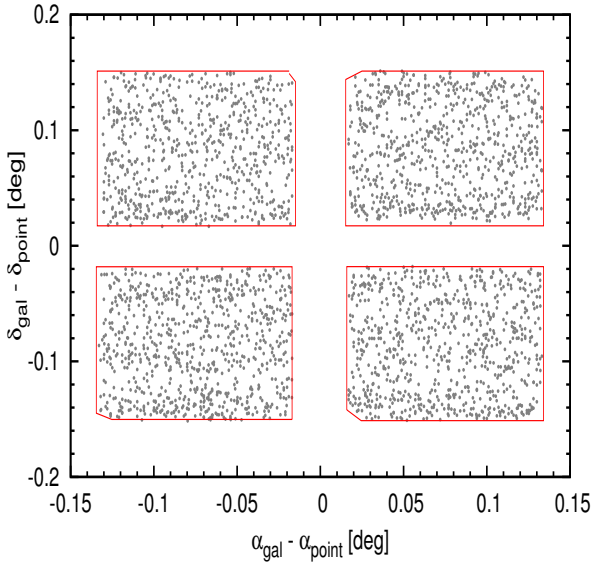


Figure 3.3: Distribution of the VUDS galaxies in the rest frame angular coordinates of their pointings. The red lines indicates the borders applied to each quadrant separately.

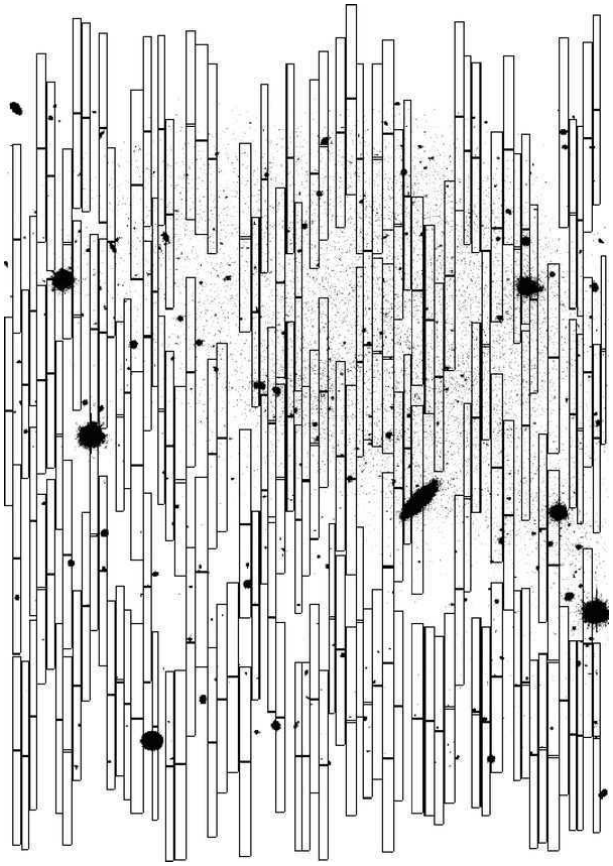


Figure 3.4: Example of allocation of the spectra slits on one of the quadrants of the VIMOS multi-spectrograph, which has been used for VUDS survey observations. Some of the galaxies are excluded form observations due to the slit allocation method. Picture comes from Bottini et al. (2005).

These biases, as well and a few other, less important effects have been described in detail by Pollo et al. (2005).

However, these biases can be minimized by a combination of corrections, which were presented by Pollo et al. (2005) and de la Torre et al. (2011a), and which we found the best working for the VUDS data:

1. **Random sample construction**. The first, most basic part of the correction scheme is the appropriate construction of the random catalogue which needs to be geometrically identical (with the exception of the small scale non-linear biasses introduced by the SSPOC) with the spectroscopic sample. Generating random objects in this catalogue, we take into account the shape of the single pointings and quadrants, which are related to the shape of VIMOS spectrograph, as presented in Figure 3.3. Additionally, we exclude the regions removed from the parent photometric sample (e.g. due to the presence of a bright star), by applying the same photometric mask to the random sample.

   The number of objects generated in the 'random catalogue', also plays a crucial role. In practice it should be significantly more numerous than the real data sample ($N_R >> N_G$). Too few random objects introduce additional distortions into $\xi(r)$ measurements, especially on small scales ($r_p < 3h^{-1}$Mpc). A set of tests have been performed to measure the strength of this effect for different N_R and in result the optimal number of objects necessary to minimize effect, have been retrieved. For the two random catalogues $N_R = 10,000$ and $N_R = 100,000$ the correlation function $w_p(r_p)$ has been estimated 5 times for the same data sample (i.e. the general population of VUDS galaxies in the redshift range $2 < z < 5$). Then, the best-fit power-law parameters r_0 and γ have been obtained for each of these measurements.

   As expected and presented in Figure 3.5, the variation of the correlation function from the average is significant when the random catalogue have $N_R = 10,000$ (especially on small scales), but negligible when $N_R = 100,000$. In effect the correlation length r_0 varied by 6% in the first case, and only by 0.5% when the more numerous random catalog was used. Hence, for the measurements presented in this work we decided to use $N_R = 100,000$, for which additional noise was found to be negligible.

   These first-order corrections reduce most of the negative effects on the correlation function.

2. **A global correction**. In order to account for the missing pairs (due to the VIMOS limitations and the SSPOC strategy) we assign a "global" weight to each galaxy-

galaxy pair. Assuming that the parent catalog is free from angular incompleteness, we define a weighting function $f(\theta)$ as a ratio between the mean number of pairs in the parent photometric catalog and the meain number of pairs in the spectroscopic sample, as a function of angular separation (de la Torre et al. 2011a):

$$f(\theta) = \frac{1 + w_{par}(\theta)}{1 + w_{spec}(\theta)}, \tag{3.37}$$

where $w_{par}(\theta)$ and $w_{spec}(\theta)$ are the angular correlation functions of the parent photometric and spectroscopic samples, respectively. Then, each pair from the spectroscopic sample separated by the angular distance θ is weighted by this ratio $f(\theta)$.

3. **Small scale corrections**. As it comes out, even this strategy does not fully account for the small-scale angular effects of the SSPOC target selection strategy. In order to account additionally for these local small-scale biases we additionally use the local weighting scheme which is also using the parent photometric catalog as a reference. In this case, we count how many galaxies around the targeted galaxy we are missing due to the limited space for spectroscopic slits. Each targeted galaxy pair is then weighted proportionally to its representativenes of the surrounding density field. Similarly to Pollo et al. (2005) we found that the optimal size of the weighting area is a circle of radius $\sim 40''$.

The correcting scheme including all these ingredients was confirmed to be optimal after a series of tests performed using all 66 VUDS mock catalogues (see Sec. 3.3). In figure 3.6 we present the results of the correlation function measurements for the mock catalogues in our three redshift ranges. All the points correspond to the means of the results from 66 VUDS mock catalogues, while the errors are computed as the standard deviations between them. The three upper panels show the comparison between (1) the 'true' (computed for the whole mock parent sample) correlation function and (2) the 'observed' correlation function computed only with the most obvious correction number (1), i.e. the appropriate geometrically cut random sample. Lower panels present the comparison of the same 'true' correlation function and (3) the 'observed' correlation function when the full correcting scheme, described above, was applied.

It is clearly seen that the introduced corrections significantly improve the retrieved signal from 'observed' sample at all the spatial scales, and in all three considered redshift bins. However, it can be noticed that on small scales ($r_p < 3h^{-1}$Mpc), there is still some signal missing, even after introducing the full optimal correcting scheme (see lower panels of Figure 3.6). For these separations, our measurement is on average underestimated by $10\% \pm 3\%$ with respect to the 'true' value of $w_p(r_p)$ computed from the parent sample. At

separations larger than $r_p \sim 3h^{-1}$Mpc the value of $w_p(r_p)$ is also slightly underestimated, but only by about $1.6\% \pm 0.9\%$.

It is a significant improvement in comparison to the average $28\% \pm 5\%$ bias which is observed for the measurements without the correcting scheme. However it still influence the value of the correlation length r_0, which in particular may be underestimated, with respect to the 'true' underlying values. In order to check the strength of this underestimation, a power-law function has been fitted to the 'true' (computed for the parent mock) correlation function as well as to the 'observed' (the mock catalog after SSPOC selection), with the full correcting scheme, both measured as the average from 66 VUDS mock catalogs. The retrieved 'true' correlation length, has been found to be on average higher, by $\Delta r_0 = 0.42 \pm 0.34$, with respect to the 'observed' r_0. Similarly, the 'observed' slope was underestimated by $\Delta\gamma = 0.06 \pm 0.04$, which is almost negligible but not entirely.

These values have been found significant enough for the remaining systematic underestimation of the $w_p(r_p)$ to be taken into account during the correlation function measurements. To correct the correlation function for the observed underestimation, we compute

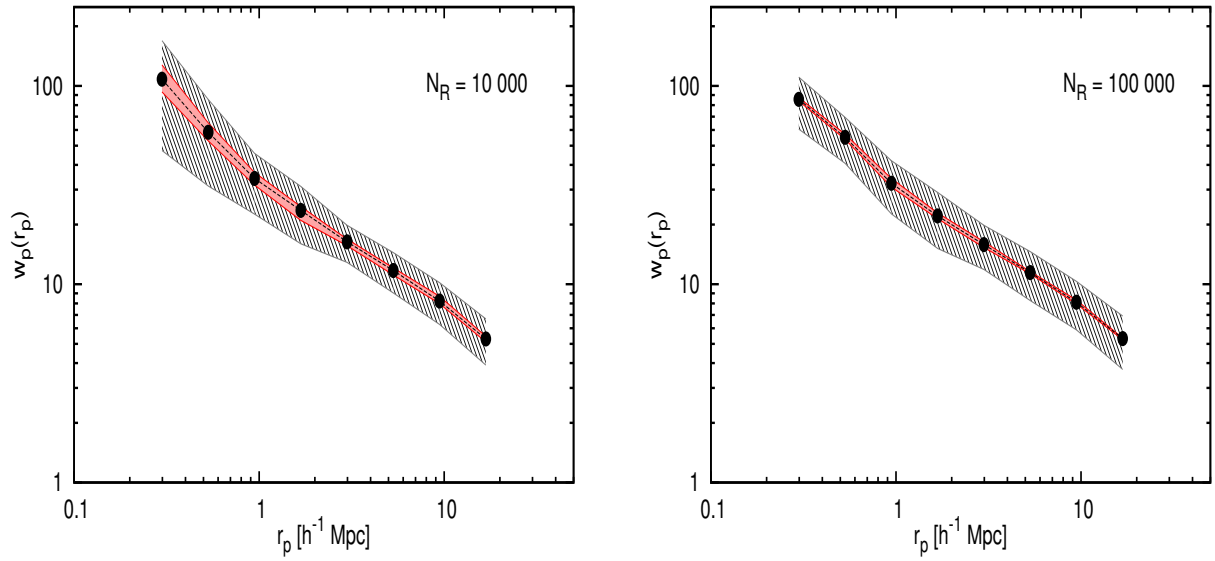


Figure 3.5: The dispersion of projected correlation function $w_p(r_p)$ observed when the random catalogue contains $N_R = 10,000$ (*left panel*) and $N_R = 100,000$ (*right panel*) generated objects. Black filled circles represents the average form 5 independent measurements performed on the same dataset. Red filled areas show the maximum deviations from the average, while the black shaded area the associated average uncertainties of the measurement.

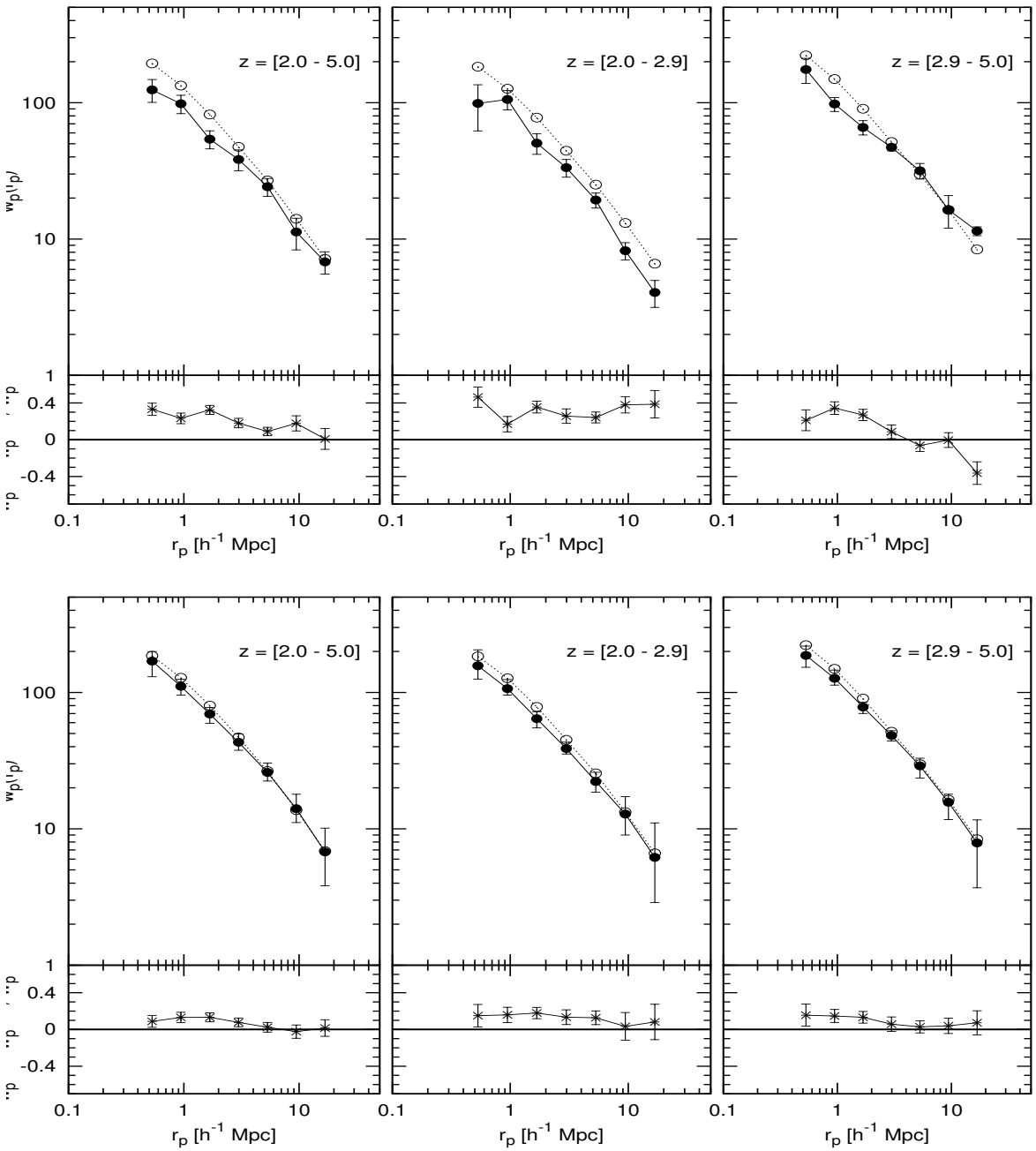


Figure 3.6: Projected two point correlation function $w_p(r_p)$ measured for 66 VUDS mock catalogs in three redshift ranges. The points correspond to the mean measurement of all 66 mock catalogs, while the errors are computed as their standard deviations. The 'true' $w_p(r_p)$ computed for the whole parent sample (open circles) is compared to that measured from the 'observed' sample (filled circles). *Upper panel:* Without corrections (2) and (3). *Lower panel:* After our full correction scheme was applied.

the ratio w_p^{par}/w_p^{obs} between the 'true' $w_p^{par}(r_p)$ and 'observed' $w_p^{obs}(r_p)$ correlation function (based on the VUDS mock) on every scale r_p for each redshift range separately. The correlation function measurements are then multiplied by this ratio for each separation r_p, resulting in the final correlation function measurement $w_p(r_p)$ on which the power-law and HOD fitting have been performed.

All results presented in this work include the correlation function correction described above.

Correlation function error estimate methods

Estimating the errors in two-point correlation function is the subject of the wide discussion in the literature since the very first measurements have been introduced (see Hamilton 1993, Fisher et al. 1994, Bernardeau et al. 2002). It has been shown that the simple Poisson method significantly underestimates the error bars and it requires statistical corrections. One of these corrections, proposed by Kaiser (1986), requires multiplying Poissonian errors by factor $1 + 4\pi n J_3$, with n being the number density of objects and $J_3 = \int^{r_j} r^2 \xi(r) dr$. However even with this corrections the errors still are underestimated (Fisher et al. 1994).

Recently the most extensively used methods are three following ones. In the first method the whole data set is split into smaller sub-samples for each the correlations are estimated independently. Then the uncertainties are computed as the rms values.

Another method is called "bootstrap resampling" (Barrow et al. 1984). It is based on the idea of perturbing the data set, by randomly creating a large number of comparable pseudo data-sets, which differ only slightly form the original sample. The certain number of objects is randomly selected form the data sample (objects are allowed to be drawn multiply times) creating subsamples for which the correlation is computed. The procedure is repeated N times, giving as the result the variation along the 'real' result.

The last method is based on sets of independent mock surveys - simulated catalogues based on large N-body simulations coupled with physical definitions of galaxies. The correlations computed for each independent mock catalogue gives the variation of the result at each scale.

After performing set of tests we decided to use the combined method which take advantage form both bootstrap resampling and mock catalogues, similar to the method proposed by Pollo et al. (2005). The details of this method are described below.

Correlation function shape fitting

The fitting procedure to estimate the power-law parameters r_0 an γ from the projected correlation function $w_p(r_p)$ follows Fisher et al. (1994) (see also Guzzo et al. 1997 and Pollo et al. 2005). For both bootstrap representations and mocks the associated covariance matrix $\mathbf{C}$, between the values of w_p on ith and kth scale can be computed as:

$$\mathbf{C}_{ik} = \Big\langle \left(w_p^j(r_i) - \langle w_p^j(r_i)\rangle_j\right)\left(w_p^j(r_k) - \langle w_p^j(r_k)\rangle_j\right) \Big\rangle_j \tag{3.38}$$

where "$\langle\rangle$" indicates an average over all bootstrap or mock realizations, the $w_p^j(r_k)$ is the value of w_p computed at $r_p = r_i$ in the cone j, where $1 < j < N_{mock}$ for VUDS mocks and $1 < j < N_{boot}$ for bootstrap data.

The covariance matrix reconstructed from $N_{mock} = 66$ VUDS mock catalogues could not be directly applied to the observed data, because it caused the fit to be often unstable. Diagonal elements of matrix reached the realistic values but the off-diagonal non-zero elements differed significantly from those pertaining to the data sample. For this reason we computed our error bars using the realistic scatter of the VUDS mock catalogues, but off-diagonal elements are computed applying the bootstrap method. This means that for each redshift, luminosity and mass ranges we measured the correlation function $w_p(r_p)$ form 1) VUDS data; 2) $N_{mock} = 66$ VUDS mock catalogs; 3) $N_{boot} = 100$ bootstrap resamplings of the VUDS data.

Power-law fitting

It is well established that the values of w_p are not independent at different separations, hence it is not possible to use a simple χ^2 minimization to find the best-fit parameters. However, the covariance matrix $\mathbf{C}$ is symmetric and real, so when it is not singular it can be inverted. If we call $\mathbf{H} = \mathbf{C}^{-1}$, we can fit w_p by minimizing a generalized χ^2, which is defined as

$$\chi^2 = \sum_{i,j}^{N_D} \left(w_p^{obs}(r_i) - w_p^{mod}(r_i)\right) \mathbf{H}_{ij} \left(w_p^{obs}(r_j) - w_p^{mod}(r_j)\right) \tag{3.39}$$

HOD fitting

To explore the HOD parameters space we implemented our HOD model to the publicly available CosmoPMC[2] code, which is using "Population Monte Carlo" (PMC) technique to sample likelihood space (Wraith et al. 2009, Kilbinger et al. 2011). PMC is the adaptive importance-sampling technique (Cappé & Moulines 2007) which allows efficient

[2] `http://www2.iap.fr/users/kilbinge/CosmoPMC/`

Table 3.1: Tests of HOD fitting procedure at low redshift on Luminosity-threshold Samples. Bold characters indicate parameters obtained from fitting the HOD model created for this study, while values in parenthesis represent corresponding parameters estimated by Zehavi et al. (2011).

M_r^{max}	z_{eff}	$\log M_{min}$	$\sigma_{\log M}$	$\log M_0$	$\log M_1$	α
-22.0	0.13	**14.07±0.09**	**0.63±0.06**	**13.09±0.56**	**14.95±0.11**	**1.34±0.15**
		(14.06±0.06)	(0.71±0.07)	(13.72±0.53)	(14.80±0.08)	(1.35±0.49)
-21.5	0.11	**13.38±0.13**	**0.67±0.03**	**13.27±0.16**	**14.42±0.08**	**0.94±0.04**
		(13.38±0.07)	(0.69±0.08)	(13.35±0.21)	(14.20±0.07)	(1.09±0.17)
-21.0	0.09	**12.85±0.07**	**0.57±0.04**	**12.90±0.11**	**14.00±0.03**	**1.12±0.07**
		(12.78±0.10)	(0.68±0.15)	(12.71±0.26)	(13.76±0.05)	(1.15±0.06)
-20.5	0.08	**12.13±0.05**	**0.22±0.06**	**11.73±0.09**	**13.45±0.03**	**1.15±0.09**
		(12.14±0.03)	(0.17±0.15)	(11.62±0.72)	(13.43±0.04)	(1.15±0.03)
-20.0	0.06	**11.82±0.03**	**0.10±0.07**	**12.95±0.09**	**12.96±0.05**	**0.99±0.04**
		(11.83±0.03)	(0.25±0.11)	(12.35±0.24)	(12.98±0.07)	(1.00±0.05)
-19.5	0.05	**11.56±0.02**	**0.18±0.08**	**12.39±0.10**	**12.82±0.03**	**1.03±0.06**
		(11.57±0.04)	(0.17±0.13)	(12.23±0.17)	(12.75±0.07)	(0.99±0.04)
-19.0	0.04	**11.45±0.07**	**0.29±0.15**	**9.54±0.09**	**12.89±0.04**	**1.19±0.10**
		(11.45±0.04)	(0.19±0.13)	(9.77±1.41)	(12.63±0.04)	(1.02±0.02)
-18.5	0.04	**11.24±0.14**	**0.14±0.10**	**10.38±0.60**	**12.34±0.02**	**1.13±0.04**
		(11.33±0.07)	(0.26±0.21)	(8.99±1.33)	(12.50±0.04)	(1.02±0.03)
-18.0	0.03	**11.00±0.09**	**0.70±0.20**	**10.33±1.04**	**12.19±0.07**	**1.10±0.06**
		(11.18±0.04)	(0.19±0.17)	(9.81±0.62)	(12.42±0.05)	(1.04±0.04)

sampling of the parameter space for the large number of samples. For each galaxy sample we fit the real-space correlation function $w_p(r_p)$ and the number density of galaxies n_g, by summing both contributions in log-likelihood obtained by:

$$\chi^2 = \sum_{i,j}^{N_D} \left(w_p^{obs}(r_i) - w_p^{mod}(r_i)\right) \mathbf{C}_{ij}^{-1} \left(w_p^{obs}(r_j) - w_p^{mod}(r_j)\right) + \frac{\left(n_g^{obs} - n_g^{mod}\right)^2}{\sigma_{n_g}^2} \qquad (3.40)$$

where n_g^{mod} is given by Eq. 3.17 at the mean redshift of the sample. The data covariance matrix is approximate by taking Eq. 3.38. The error on the galaxy number density σ_{n_g} contains Poisson noise and cosmic variance.

3.5 HOD model fitting tests

The HOD model code created for this study along with the fitting procedure has been tested before introducing the measurements presented in the following chapters. In particular we performed the HOD fitting of the correlation function measurements based on the low redshift ($z \sim 0$) Sloan Digital Sky Survey (SDSS) data sample, performed by Zehavi et al. (2011) (1). In addition to that we implemented the similar test for $w_p(r_p)$ estimated at intermediate ($z \sim 1$) redshift for the VIMOS VLT Deep Survey data, by Abbas et al. (2010) (2). In the following section we present the representative results from these tests.

1. In order to perform the first test we modified the halo occupation function $\langle N_g|M\rangle$ to one used by Zehavi et al. (2011) for which the mean occupation function is characterized by a smoothed step function for the central galaxy and a power-law number of satellites increasing with halo mass. It is represented by:

$$\langle N_g|M\rangle = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log M - \log M_{min}}{\sigma_{\log M}}\right)\right] \times \left[1 + \left(\frac{M - M_0}{M_1}\right)^{\alpha}\right] \qquad (3.41)$$

   Here erf is the error function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$ and the five free parameters are: the minimum halo mass M_{min} for which 50% of halos hosts galaxy, the width $\sigma_{\log M}$, cutoff mass scale M_0, halo mass M_1 for which halo host one central and one satellite galaxy of a given type, and slope α.

   All these parameters were allowed to vary within the ranges motivated by the previous studies: $10 < \log M_{min} < 15$, $10 < \log M_1 < 15$, $9 < \log M_0 < 14$, $0.1 < \sigma_{\log M} < 0.8$, $0.6 < \alpha < 2$. The real-space correlation function $w_p(r_p)$ measured by Zehavi et al. (2011) for the luminosity-threshold subsamples, have been fitted using PMC adaptive importance-sampling technique as described in Sec. 3.4.2.

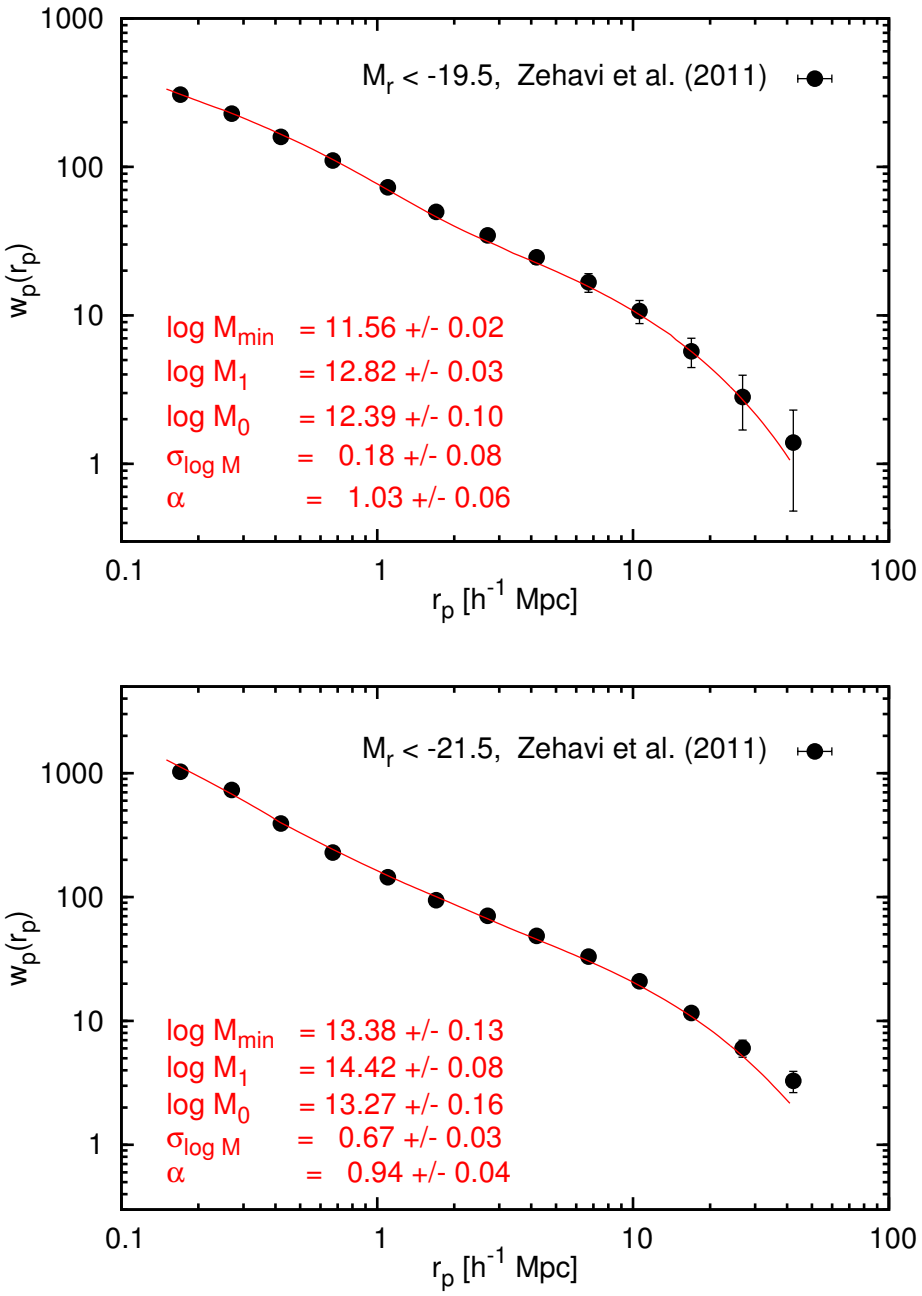


Figure 3.7: HOD fit tests preformed on the correlation function measurements for two selected luminosity threshold subsamples $M_r < -19.5$ (*upper panel*) and $M_r < -21.5$ (*lower panel*). The symbols and errorbars indicate the measurements of projected correlation function $w_p(r_p)$ form the SDSS Survey performed by Zehavi et al. (2011). The solid lines present the best-fit halo model by using method developed for this study.

Table 3.2: Tests of HOD fitting procedure on intermediate redshift range on Luminosity-threshold Samples. Bold characters indicate parameters obtained from fitting the HOD model created for this study, while values in parenthesis represent the corresponding parameters presented by Abbas et al. (2010).

M^{tresh}	z_{mean}	$\log M_{min}$	$\log M_1$	α
< -18.0	0.44	**11.15±0.29** (11.05±0.87)	**12.80±0.77** (12.73±1.00)	**0.88±0.18** (0.99±0.10)
	0.66	**11.28±0.26** (11.21±0.18)	**12.94±0.38** (13.03±0.11)	**1.10±0.30** (1.27±0.08)
< -18.5	0.53	**11.47±0.21** (11.37±0.16)	**12.71±0.21** (12.93±0.06)	**0.91±0.13** (1.11±0.12)
	0.76	**11.50±0.29** (11.63±0.19)	**13.15±0.50** (13.31±0.18)	**1.35±0.21** (1.31±0.09)
< -19.0	0.59	**11.62±0.35** (11.88±0.08)	**13.22±0.62** (13.57±0.1)	**0.94±0.21** (1.33±0.16)
	0.88	**11.76±0.29** (11.76±0.08)	**13.07±0.47** (13.39±0.08)	**1.12±0.16** (1.19±0.24)
< -19.5	0.67	**12.22±0.25** (12.06±0.12	**13.70±0.52** (13.72±0.10)	**1.18±0.19** (1.43±0.27)
	0.99	**11.80±0.71** (11.92±0.16)	**13.40±0.43** (13.43±0.22)	**0.97±0.19** (1.22±0.25)
< -20.0	0.67	**12.11±0.23** (12.06±0.07)	**14.09±0.37** (13.78±0.04)	**1.40±0.15** (1.66±0.18)
	1.05	**12.18±0.28** (12.35±0.11)	**13.87±0.78** (13.80±0.16)	**1.38±0.14** (1.42±0.41)

Best-fit HOD parameters for each subsample are in the excellent agreement with these presented by Zehavi et al. (2011) as shown in Table 3.1. Additionally the two representative examples of the fit (for the $M_r < -19.5$ and $M_r < -21.5$ subsamples) are presented in Figure 3.7

2. Another set of tests has been performed on the correlation function measurements presented by Abbas et al. (2010) for the VVDS-Deep data. This time the halo occupation function $\langle N_g|M\rangle$ have been parametrized as described in section 3.2 i.e with three free parameters: M_{min} - the minimum mass needed for a halo to host one central galaxy, M_1 - the mass of a halo having on average one satellite galaxy, and α - the power law slope of the satellite mean occupation function. The parameters have been allowed to vary within the same ranges as in the previous test. The best-fit parameters obtained by us in comparison to these presented by Abbas et al. (2010) are listed in Table 3.2.

Taking into account that the best-fit parameters measured in both tests are in excellent agreement with the ones presented both in Zehavi et al. (2011) and Abbas et al. (2010), we conclude that the model and the fitting method have been constructed well and the following measurements are robust.

3.6 Galaxy bias

According to the current cosmological paradigm of structure formation, galaxies form and evolve inside dark matter halos (White & Rees 1978). In other words, there exists a connection between dark matter distribution and galaxy structures - in the dense dark matter regions galaxies should be more clustered. The baryonic matter distribution, however, is biased with respect to the dark matter density field. The strength of this effect is referred to as *galaxy bias*.

As the first approximation in this work we used a linear model of the galaxy bias. It assumes a linear relation between galaxy $\sigma_{R,g}(z)$ and mass $\sigma_{R,m}$ rms at a given redshift:

$$\sigma_{R,g}(z) = b\sigma_{R,m} \tag{3.42}$$

where b is the galaxy bias. We also assume that b is independent of scale R, which is true especially for large scales. Usually the adopted value of $R = 8h^{-1}$Mpc. Locally, at $z = 0$, the mass fluctuations reach the value $\sigma_{8,m}(z = 0) = 0.83$ (Planck Collaboration et al. 2014). In our model the redshift evolution of this quantity is described as:

$$\sigma_{8,m}(z) = \sigma_{8,m}(z = 0)D(z), \tag{3.43}$$

where

$$D(z) = \frac{g(z)}{g(0)(1+z)}, \tag{3.44}$$

and $g(z)$ is the normalized *growth factor*, which describes how fast the local perturbations grow with the scale factor. Its value depends on the assumed cosmological model, i.e. cosmological parameters. After Carroll et al. (1992) we write:

$$g(z) = 2.5\Omega_m \left[\Omega_m^{4/7} - \Omega_\Lambda + \left(1 + \frac{\Omega_m}{2}\right)\left(1 + \frac{\Omega_\Lambda}{70}\right)\right]^{-1}, \tag{3.45}$$

where Ω_m and Ω_Λ are the matter and dark energy density parameters at a given redshift.

The galaxy rms at a given scale R $\sigma_{8,g}$ can be retrieved by using the correlation function parameters r_0 and γ:

$$\sigma_{8,g} = \sqrt{C_\gamma \left(\frac{r_0}{8h^{-1}\mathrm{Mpc}}\right)^\gamma}, \tag{3.46}$$

with

$$C_\gamma = \frac{72}{2^\gamma(3-\gamma)(4-\gamma)(6-\gamma)}. \tag{3.47}$$

In this study, we refer to the galaxy large-scale bias computed by this method as b_L^{PL}. This is also used to compute the values based on measurements of the correlation function from other surveys when the b_L estimations were not presented by the authors; in these cases we use their power-law best fit parameters r_0 and γ (e.g. from Le Fèvre et al. 2005). Another estimate of linear large-scale galaxy bias can be obtain from the HOD model. We compute the large-scale galaxy bias b_L^{HOD} using the fit of the halo occupation function (Eq. 3.32).

CHAPTER 4

Clustering measurements of the high redshift galaxy population $2 < z < 5$

As I already pointed out at the beginning of this thesis (see Chapter 1) there are still a lot of unknowns considering the formation and evolution of the large scale structure of the universe. The answer to these questions (in the framework of the assumed cosmological model) can be provided by the studies of the galaxy clustering at different epochs. These kind of studies are mostly carried out by using the galaxy correlation function $\xi(r)$ (or its projected version $w_p(r_p)$) along with its two approximations: power-law and the halo occupation models (HOD).

As listed in the Section 1.4 a remarkable number of clustering studies have been performed up do date at low ($z \sim 0$) and intermediate ($z \sim 1$) redshift ranges (e.g Norberg et al. 2002, Zehavi et al. 2004, 2012, Pollo et al. 2006, de la Torre et al. 2007, Coil et al. 2008, Quadri et al. 2008, Meneux et al. 2006, 2008, 2009, Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012). However, the overall picture of structure evolution is far from being complete. There is still need for comprehensive studies at higher redshift ranges $z > 2$, to probe the early stages of structure assembly.

We are making an attempt to improve the current view of the clustering of galaxies at $z > 2$. We investigate how galaxy clustering evolved from the early phases of dark matter and galaxy assembly to present times. In this chapter we present a clustering analysis of the VIMOS Ultra Deep Survey (VUDS), the largest spectroscopic survey covering the redshift range $2 < z < 6$ in a continuous way to date (Le Fèvre et al. 2014), using a sample of ~ 3000 galaxies with confirmed spectroscopic redshifts in the range $2 < z < 5$. The

Table 4.1: The number of galaxies in VUDS fields for each reliability flag in the redshift range $2 < z < 5$, as used in this study.

VUDS field	reliability flag $zflag$				Area (deg^2)
	2	3	4	9	
COSMOS	939	607	322	72	0.50
VVDS-02h	471	384	207	20	0.31
Total	1410	991	529	92	0.81

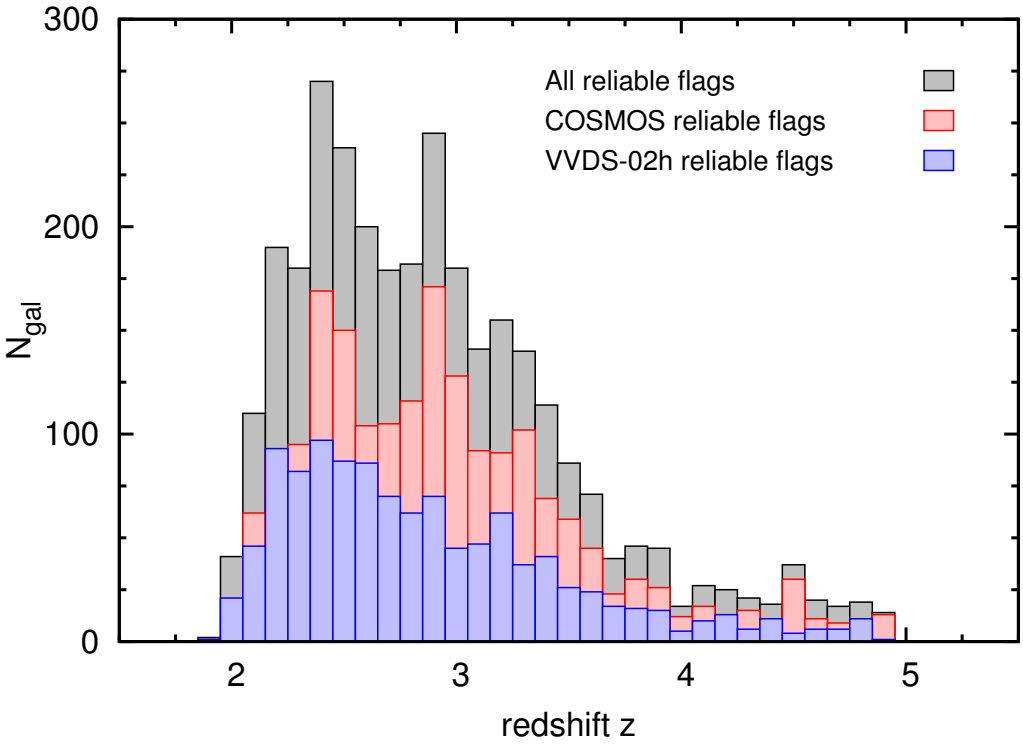


Figure 4.1: The redshift distribution of the used VUDS galaxy sample for redshift range $2.0 < z < 5.0$. The filled gray histogram represents the total sample of reliable galaxies (with $zflag = 2 - 4, 9$, $\sim 80\%$ reliability), while the red and blue histograms represent the contribution from COSMOS and VVDS2h fields respectively.

sample covers a total area of 0.81 deg^2 observed in two independent fields, COSMOS and VVDS-02h, with a mean redshift value of $z \sim 3$. We measure the redshift-space $\xi(r, \pi)$ and projected real-space $w_p(r_p)$ correlation function. We use the power-law approximation to measure the evolution of the correlation strength up to a median redshift $z \sim 3.5$ for the general star-forming galaxy population. We use the halo occupation distribution (HOD) model to put constrains on the properties of the dark matter halos hosting star-forming galaxies at redshifts $z \sim 3$. Methods used to estimate the correlation function, correcting schemes, the correlation function uncertainties and fitting procedures are discussed in Chapter 3. In this chapter we present and discuss the main results of our study. This work has been presented in the paper draft attached in Appendix A.

4.1 The sample selection for clustering analysis

In this study we are using data form the VUDS galaxy survey (see Chapter 2). We choose only these objects with assigned reliability flag $zflag = 2, 3, 4, 9$ in the redshift range $2 < z < 5$ (see Sec. 4.2 for discussion about flag selection). All used objects were observed in two independent fields, COSMOS and VVDS-02h, covering a total area 0.81 deg^2, as data for the third VUDS field in the ECDFS is still too sparse to be used for clustering analysis. The covered area corresponds to a volume $\sim 3 \times 10^7 \mathrm{Mpc}^3$ sampled in the redshift range $2 < z < 5$. In total our sample consists of 3022 galaxies with reliable spectroscopic redshifts in the range $2 < z < 5$. The sample is summarized in Table 4.1.

The spatial distribution of the spectroscopic galaxy sample in each of these fields is presented in Figure 4.2, while Figure 4.1 presents their redshift distributions. The mean redshift value for the whole sample is $z \sim 2.96$. We use 3 different redshift bins to study our sample. In a first sample we include the whole star-forming galaxy population observed by VUDS in the range $2 < z < 5$ with a mean redshift $z \sim 2.96$. We also define a *low* redshift sample with a mean redshift $\bar{z} = 2.5$ in the range z=[2,2.9] and a *high* redshift sample with $\bar{z} = 3.45$ in z=[2.9,5], which roughly represent two equal cosmic time bins. The general properties of these samples including median luminosity M_U^{median} and stellar mass $\log M_*^{median}$ are listed in Table 4.2.

Because of the $i_{AB} = 25$ limit of the VUDS selection, the galaxies in the high redshift sample are ~ 0.5 mag brighter and $\sim 0.2\ M_\odot$ more massive than those in the low redshift sub-sample, which is the expected effect taking into account the VUDS selection strategy. The impact of these differences on the clustering measurements of our three samples are fully discussed in Section 4.4.1.

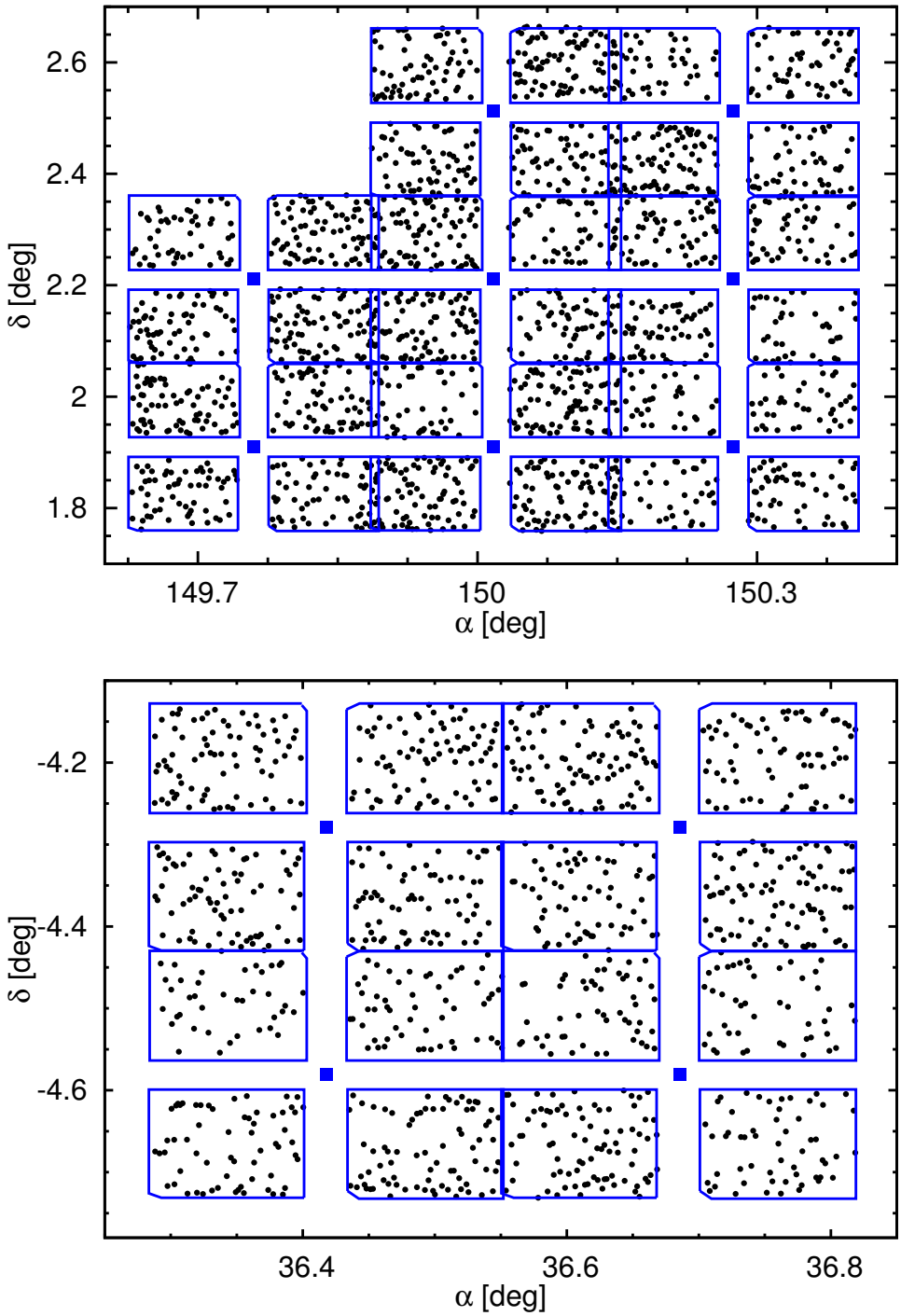


Figure 4.2: The distribution of galaxies with spectroscopic redshifts $2 < z < 5$ in two independent VUDS fields: COSMOS on the *upper* panel and VVDS-02h on *lower* panel. The blue squares indicates VIMOS pointing centers.

4.2 Impact of the reliability of redshift measurements on the correlation function accuracy

As discussed in previous chapters, the precision and reliability of the galaxy correlation function measurement strongly depends on the reliability and accuracy of the measurement of their radial distance, i.e. on the quality of the redshift measurement in the survey. This section is dedicated to describe how reliable is galaxy clustering study performed on the base of the VUDS data. For my study I selected the VUDS galaxies with quality flag $2, 3, 4$ and 9 (see Chapter 2 for the description of the quality flags). In this way I chose only galaxies with the most reliable redshift measurements, which in turn should result in a robust estimation of the correlation function. In particular, the 'confidence level' of the whole selected sample can be estimated using the numbers of galaxies with different flags and their corresponding reliability estimates:

$$P = (N_{f4} + 0.95N_{f3} + 0.75N_{f2} + 0.5N_{f1} + 0.8N_{f9})/N_g, \tag{4.1}$$

where N_{fi} is the number of galaxies with a given i flag and N_g is the total number of galaxies within the sample. From this formula, for our selected VUDS sample the total 'confidence level' is as high as $\sim 86\%$, which is unprecedented at these redshifts and at this depth $i_{AB} = 25$.

However, a question appears: How the remaining $\sim 14\%$ inaccuracy may possibly affect my correlation function measurement? To clarify this, I performed a set of tests, in order to check how a part of galaxy sample with possibly wrongly measured redshifts, would influence the final correlation function measurements. To perform these tests, I used 10 independent VUDS mock catalogues (see Sec. 3.3).

The presence of galaxies with wrongly measured redshift in the data sample can distort the correlation function $w_p(r_p)$ at each r_p. In particular, we expect a decrease in the measured correlation length r_0. In the mock catalogues, we can simulate these distortions by changing the spatial distribution of a corresponding fraction of mock galaxies in the mock catalogues. For my simulations I assume that galaxies with reliability flag 3 and 4 do not introduce any significant distortions to the correlation function measurements, thanks to the $95 - 100\%$ confidence level. I also assume that the possible influence of flag 9 galaxies is negligible, due to their low number in the sample (see table 4.1). Thus, the largest fraction of galaxies with possibly wrongly measured redshift is expected among flag 2 galaxies. Hence, I focused my attention on these galaxies.

I measured how strongly the correlation function would be distorted if 15%, 25%, 30% or 50% (highly unlikely case) of flag 2 galaxies would have wrongly measured redshifts. Within the VUDS galaxy sample selected for this study, flag 2 galaxies repre-

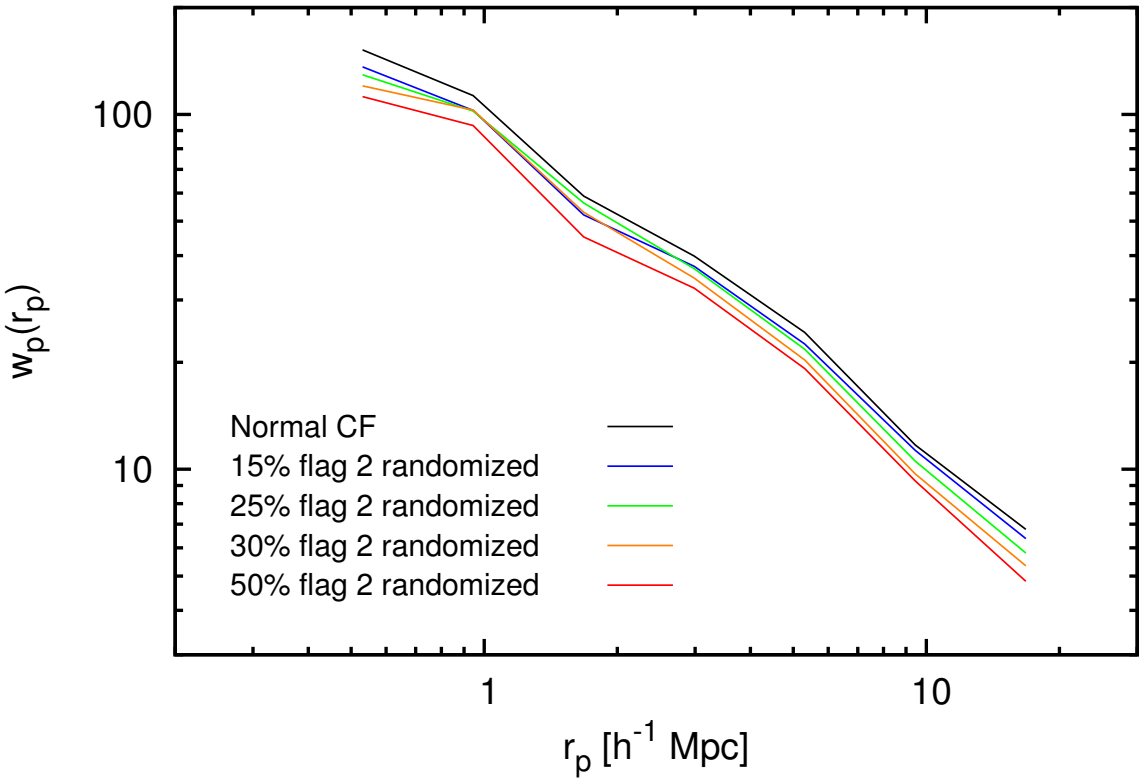


Figure 4.3: Influence of the possible incorrect redshift measurements on the correlation function. Solid lines represent the projected two point correlation function $w_p(r_p)$. Colours indicate the different levels of sample randomization (as labeled). The black solid line represents undistorted correlation function

sent 48% of the total sample (see Tab. 4.1). Therefore, in the mock catalogues I randomly selected a corresponding fraction of objects to play the role of flag 2 galaxies (i.e., $N_{f2}^{mock} = 0.48 N_{tot}^{mock}$). Then, in order to mimic their possibly wrong redshift measurements, I randomized the spatial coordinates (α and δ) of 15%, 25%, 30% and 50% of these objects, to obtain differently distorted samples. Finally, I computed the projected correlation function $w_p(r_p)$ and fitted it with the two parameter power-law (see Section 3.1 and 3.4.2).

The procedure has been repeated for 10 independent mock catalogues and the final estimation of $w_p(r_p)$ was computed as the average. The results, i.e. $w_p(r_p)$ computed for each of these four scenarios are presented in Figure 4.3. As expected, the amplitudes of the estimated projected correlation functions for the samples with wrongly measured redshifts are underestimated, with respect to the undistorted one (black line), and the underestimation increases with the increasing percentage of mis-estimated redshifts. However, the changes are not large except of the unrealistic 50% case. The correlation length for the fraction of 30% randomized galaxies differ form the 'real' r_0 only by $\Delta r_0 = 0.13 \pm 0.04$.

Based on these simple tests we conclude that the correlation function measurements performed in this study are robust, in spite of a possible contamination from wrong redshift measurements. Assuming the worst possible case, i.e. the lower limit for flag 2 galaxies 'confidence level' being 70%, the resultant clustering strength r_0 will be underestimated only by $\Delta r_0 = 0.13$, which is the $\sim 3.3\%$ of the measured amplitude. Therefore, I conclude that the upper error bar of the measured $r0$ should be rescaled by a corresponding value, to fully account for the uncertainties introduced by possible wrongly measured redshifts in the data.

4.3 Cosmic variance between VUDS fields

The galaxy sample used in this study has been observed in two separate fields on the sky: COSMOS and VVDS-02h. Therefore, the first task I faced was to find a estimation method of the correlation function which would guarantee the minimization of the cosmic variance between two fields on the correlation function and hence maximize the robustness of the final result. The details and final outcome of this procedure are described in Chapter 3. However, we can also profit from the fact that we have measurements in two separate fields: we can use these data to test how the correlation function varies between the two different VUDS fields, in order to check whether our correlation function measurement in not dominated by one of them, and to make an estimate of a cosmic variance between them.

To perform this measurement, I computed the projected two-point correlation function

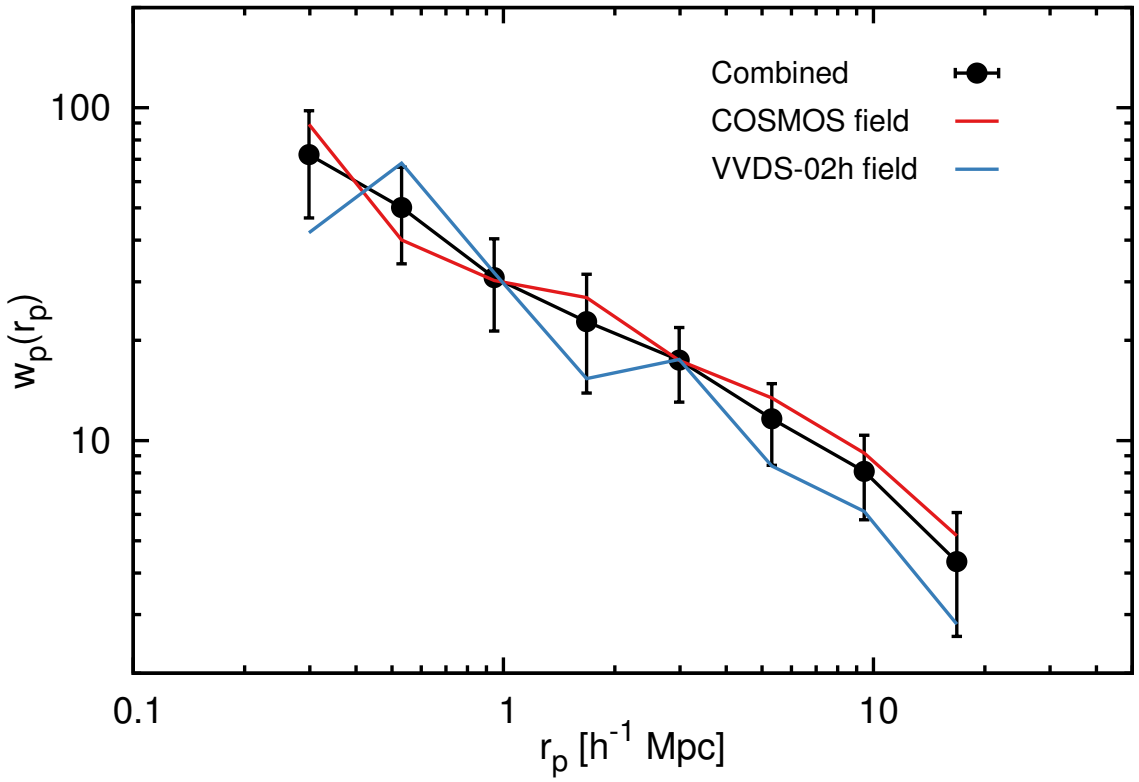


Figure 4.4: Cosmic variance in two VUDS fields: COSMOS and VVDS-02h. The red and blue solid lines indicate the two-point correlation function $w_p(r_p)$ measured in two independent fields: COSMOS and VVDS-02h, respectively. The black points indicate the correlation function composite of these two in the $2 < z < 5$.

in the redshift range $2 < z < 5$, for each of two VUDS fields separately. The results of these measurements are presented in Figure 4.4. Not surprisingly, the two computed $w_p(r_p)$ differ due to the cosmic variance. Galaxies observed in the VVDS-02h field are less clustered than in the COSMOS field. However, given the error bars of the measurements, these differences are not significant.

I conclude that the correlation function measurements for both fields are consistent, taking int account the error bars, and for both fields the signal is retrieved on all scales $0.3 < r_p < 16h^{-1}$Mpc. As a result, the composite function is not dominated by the signal from any of the individual fields at any separation r_p and can be regarded as a representative correlation function for both of them.

4.4 Results in the framework of two approximations

4.4.1 Power-law approximation

We fit a power-law to $w_p(r_p)$, with two free parameters r_0 and γ, to quantify the clustering strength. All the points in the range 0.3 h^{-1}Mpc $< r_p <$ 16 h^{-1}Mpc are used (see Sec. 3.4.2 for the fitting method details). The measured best-fit parameters are listed in Table 4.2 and the power-law fit of the correlation function for each redshift sub-sample is shown in Figure 4.5 together with the (r_0,γ) error contours.

Within our sample we find $r_0 \sim 3.95^{+0.48}_{-0.54}\ h^{-1}$Mpc in z=[2.0,2.9] and $r_0 \sim 4.35^{+0.60}_{-0.76}\ h^{-1}$Mpc in z=[2.9,5]. The slight increase of r_0 with redshift is probably due to the small luminosity differences between samples, although the difference is marginally significant given measurement errors. This observed increasing trend is however in good agreement with the theoretical predictions coupled with cosmological simulations (Katz et al. 1999, Weinberg et al. 2004, Zheng & Weinberg 2007, Conroy et al. 2006). It can be explained by the combination of two main effects:

- The increase of the correlation length r_0 of matter with cosmic time which reflects the actual growth of large scale structures with time.
- However, galaxies are not *unbiased* tracers of the dark matter field. They form in the dark matter haloes which themselves trace the underlying density field of dark matter in a biased way, and the way they form and evolve inside these haloes introduces additional bias. In particular, the strongest over-densities of matter are the first to grow and accumulate matter. As a result, at the higher redshifts, galaxies reside in haloes more biased with respect to the dark matter density field than their lower z counterparts. As a result, they are expected to have a correlation length r_0

Table 4.2: Properties of general population VUDS samples in the redshift range $2.0 < z < 5.0$.

z range	z_{mean}	M_U^{median}	$\log M_*^{median}$	N_{gal}	r_0	γ	b_L^{PL}
$[2.0-5.0]$	2.95	-21.55	9.76	3022	$3.97^{+0.36}_{-0.38}$	$1.70^{+0.09}_{-0.09}$	$2.68^{+0.22}_{-0.21}$
$[2.0-2.9]$	2.50	-21.31	9.66	1556	$3.95^{+0.48}_{-0.54}$	$1.81^{+0.02}_{-0.06}$	$2.39^{+0.32}_{-0.29}$
$[2.9-5.0]$	3.47	-21.81	9.86	1466	$4.35^{+0.60}_{-0.76}$	$1.60^{+0.12}_{-0.13}$	$3.26^{+0.47}_{-0.37}$

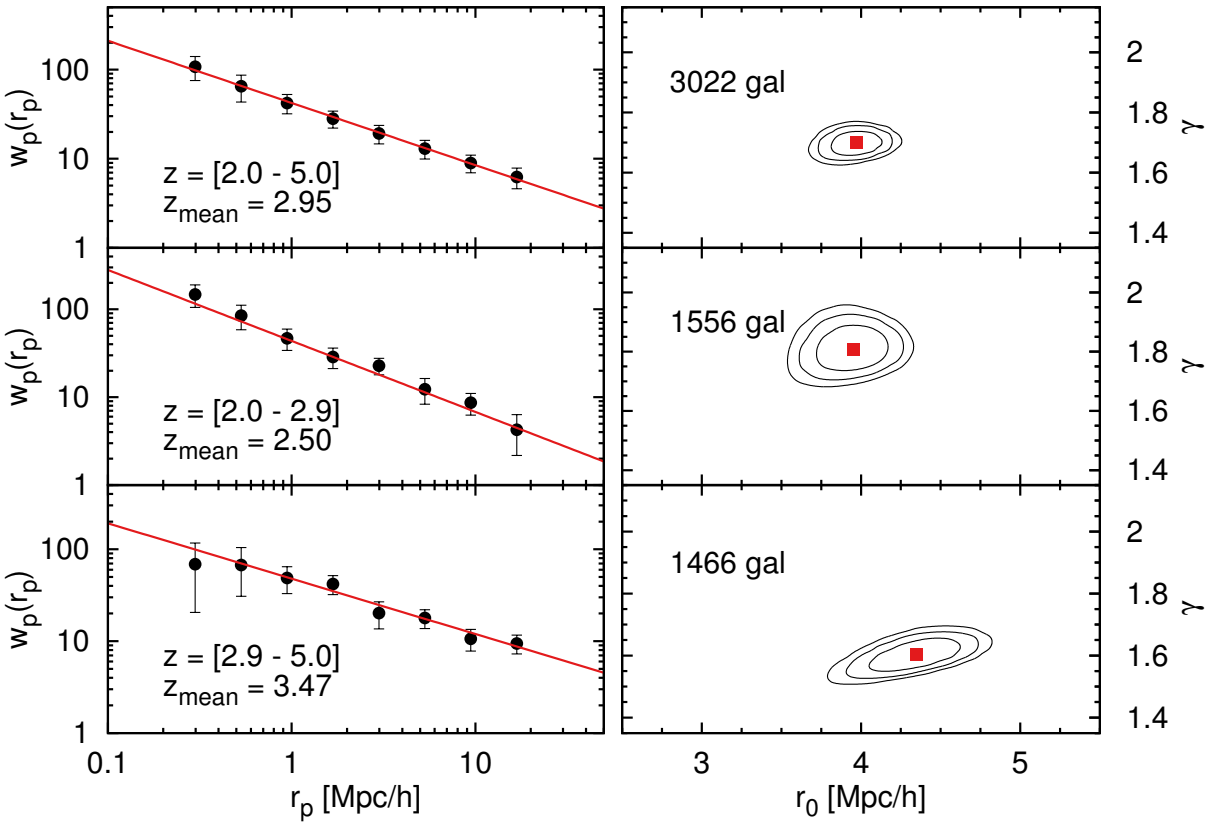


Figure 4.5: *Left panel*: The projected two-point correlation function $w_p(r_p)$. The symbols and errorbars denote measurements of the composite correlation function from the VUDS survey. *Right panel*: The associated best-fit power-law parameters r_0 and γ along with error contours for the general galaxy population separated in three redshift ranges.

higher than the corresponding correlation length for dark matter in comparison to later epochs.

These two effects combine in such a way that the expected evolution of the correlation length $r0$ with time is actually rather weak, and its detailed value depends on the cosmological model (Katz et al. 1999, Weinberg et al. 2004). Our measurements are consistent with this theoretical picture.

Additionally, the measurements are complicated by two other effects: namely, observational biases and galaxy evolution. In particular, the increase of r_0 with z may be caused by the luminosity difference between two samples at $z \sim 2.5$ and $z \sim 3.5$. Due to the magnitude selection $i_{AB} \leq 25$ of the VUDS sample we observe systematically more luminous galaxies at higher z. As a result, our high redshift sample contains on average more luminous galaxies, with median $M_U = -21.81$, while for the lower redshift sample the median $M_U = -21.31$. The luminosity dependence of galaxy clustering is a well established fact. It has been observed both at early and late cosmic epochs (e.g. Zehavi et al. 2011, Coupon et al. 2012, Abbas et al. 2010, Pollo et al. 2006). Luminous galaxies, at all z investigated so far, were found to be more strongly clustered than the fainter ones. This trend is usually related to the fact that they are located in more massive dark matter haloes and, consequently, stronger fluctuations of the dark matter density field. If this luminosity dependency of galaxy clustering is also true for the redshift range investigated in this work, the observed increase of r_0 with z may be caused by the fact that our higher redshift sample is on average more luminous that the lower one. The differences in clustering observed for the luminosity cut sub-samples are discussed in Chapter 5.

As we mentioned before, yet another factor to be taken into account is the evolution of galaxies themselves, in particular a global UV luminosity evolution of $0.3-0.5$ magnitude observed between $z \sim 3.5$ and $z \sim 2.5$ (Reddy & Steidel 2009).

The estimation of the detailed impact of all these four effects (structure growth, halo bias, selection efects, galaxy luminosity evolution) for the general population is not a simple task. It requires careful measurements and HOD modelling in volume-selected samples at different redshifts.

4.4.2 HOD modelling

The best-fit HOD model defined in section 3.2 is fit to each of three redshift sub-samples as presented in the left panels of Figure 4.6. The HOD model provides a good fit to the real-space correlation function $w_p(r_p)$ in all the redshift ranges, particularly in z=[2,2.9] where it is an excellent representation of the data. For the highest redshift bin z=[2.9,5] the fit is not as good. This is mainly due to the more noisy measurement of the correlation

Table 4.3: The HOD and and HOD-based parameters for the redshift sub-samples

z *range*	$\log M_{min}$	$\log M_1$	α	$\log\langle M_h\rangle$	b_g
$[2.0-5.0]$	$11.04^{+0.33}_{-0.36}$	$12.11^{+0.51}_{-0.45}$	$1.29^{+0.12}_{-0.11}$	$11.75^{+0.23}_{-0.28}$	$2.82^{+0.27}_{-0.16}$
$[2.0-2.9]$	$11.12^{+0.23}_{-0.35}$	$12.09^{+0.32}_{-0.35}$	$1.30^{+0.08}_{-0.10}$	$12.01^{+0.16}_{-0.18}$	$2.55^{+0.16}_{-0.21}$
$[2.9-5.0]$	$11.18^{+0.56}_{-0.70}$	$12.55^{+0.85}_{-0.88}$	$0.73^{+0.23}_{-0.30}$	$11.61^{+0.46}_{-0.40}$	$3.48^{+0.31}_{-0.26}$

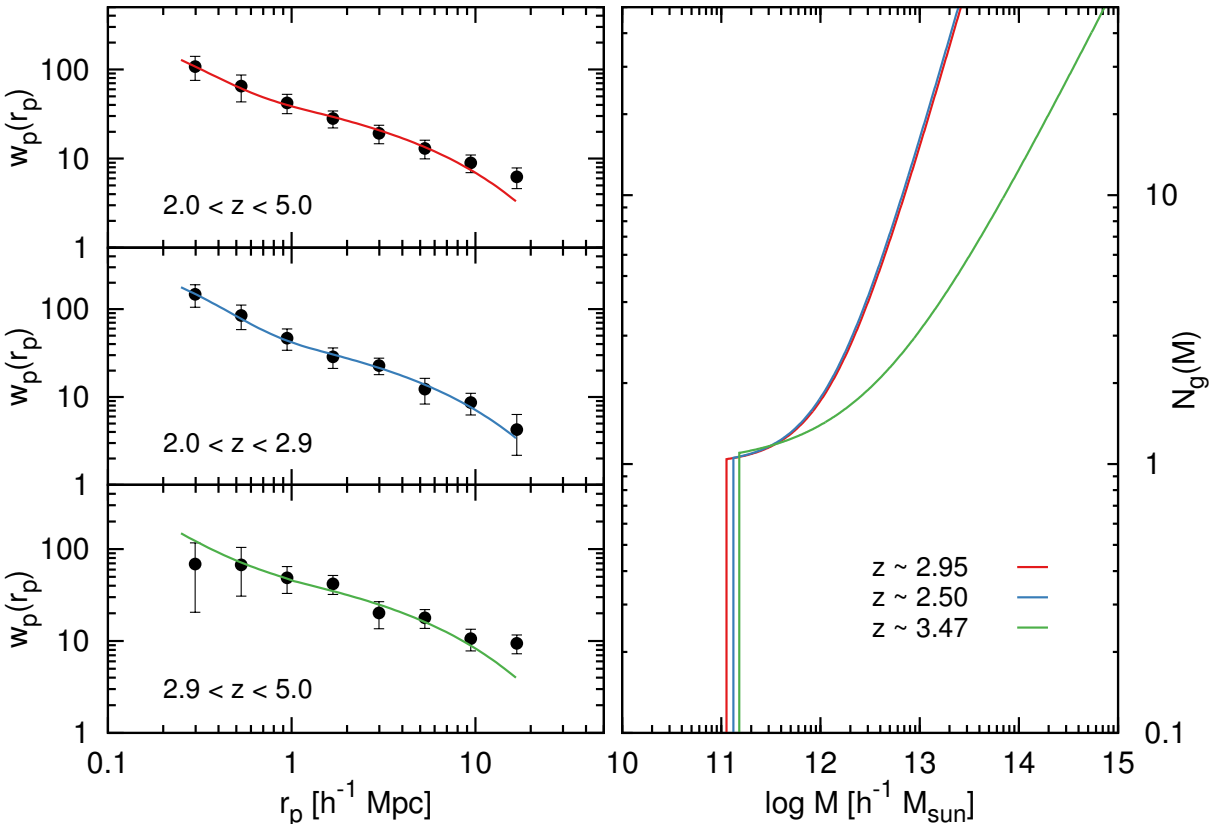


Figure 4.6: *Left panel:* The correlation function for three redshift sub-samples as in Figure 4.5. The solid lines present the best-fit halo occupation model. *Right panel:* The evolution of the halo occupation for the three redshift ranges.

function in this redshift range. Nevertheless the data is of high enough quality to allow for a meaningful HOD fit.

The three HOD parameters, the minimum halo mass M_{min}, the satellite occupation halo mass M_1, and the high halo mass slope α (see sec. 3.2), are inferred from the full covariance matrix (see section 3.4.2) and are listed in Table 4.3. The parameters are allowed to vary within a large range set from previous observations at high and low redshift: $10 < \log M_{min} < 14$, $10 < \log M_1 < 15$, $0.6 < \alpha < 2.0$. The right panel of Figure 4.6 presents the halo occupation function obtained from the HOD fit within three redshift ranges. We do not observe any significant difference in halo masses between the two redshift ranges probed for the general VUDS galaxy population. The minimum halo mass, for which a halo hosts at least one central galaxy M_{min}, has comparable value for the $z \sim 2.5$ and $z \sim 3.5$ samples: we find $M_{min} = 10^{11.12\pm0.33}$ at $z \sim 2.5$ and $M_{min} = 10^{11.18\pm0.56}$ at $z \sim 3.5$.

The satellite occupation mass M_1 is noticeably higher in the higher redshift bin compared to the lower redshift sample, although the errors are quite large. These uncertainties are related to the weak one-halo term signal in the correlation function for the higher redshift measurement.

The high halo mass slope α of the satellite mean occupation function is around $\sim$ 1.3 for z=[2,2.9], significantly steeper than $\alpha = 0.73$ found in z=[2.9,5]. Although the uncertainties on this parameter are large, we may speculate that the number of satellites was somewhat higher in the higher redshift bin, indicating an increase in galaxy merging. The galaxy merger rate derived from pair counts in VUDS indicates a high merging rate at $z \sim 3$ (Tasca et al. 2014a), which might be the signal observed in the HOD. However, this result from clustering will need to be confirmed from larger samples.

The average host halo mass $< M_h|g >$, large-scale galaxy bias b_L^{HOD} and associated 1σ errors, are reported in Table 4.3. The redshift evolution of these parameters is presented in Figure 4.8 - for the average host halo mass, and in Figure 4.9 - for the large-scale galaxy bias.

4.5 Comparison with other studies

The correlation power-law parameters $r_0 = 3.97 \pm 0.48$ and $\gamma = 1.70 \pm 0.09$ at $z \sim 3$ are comparable with the clustering parameters for the star forming galaxies at intermediate ($z \sim 1$) redshift (e.g. Le Fèvre et al. 2005, Pollo et al. 2006), but the r_0, observed at $z \sim 3$ is lower than found for galaxies in the local Universe (e.g. Zehavi et al. 2011, Norberg et al. 2002). This has to be placed in the global perspective, since the galaxies are expected to be more clustered with time.

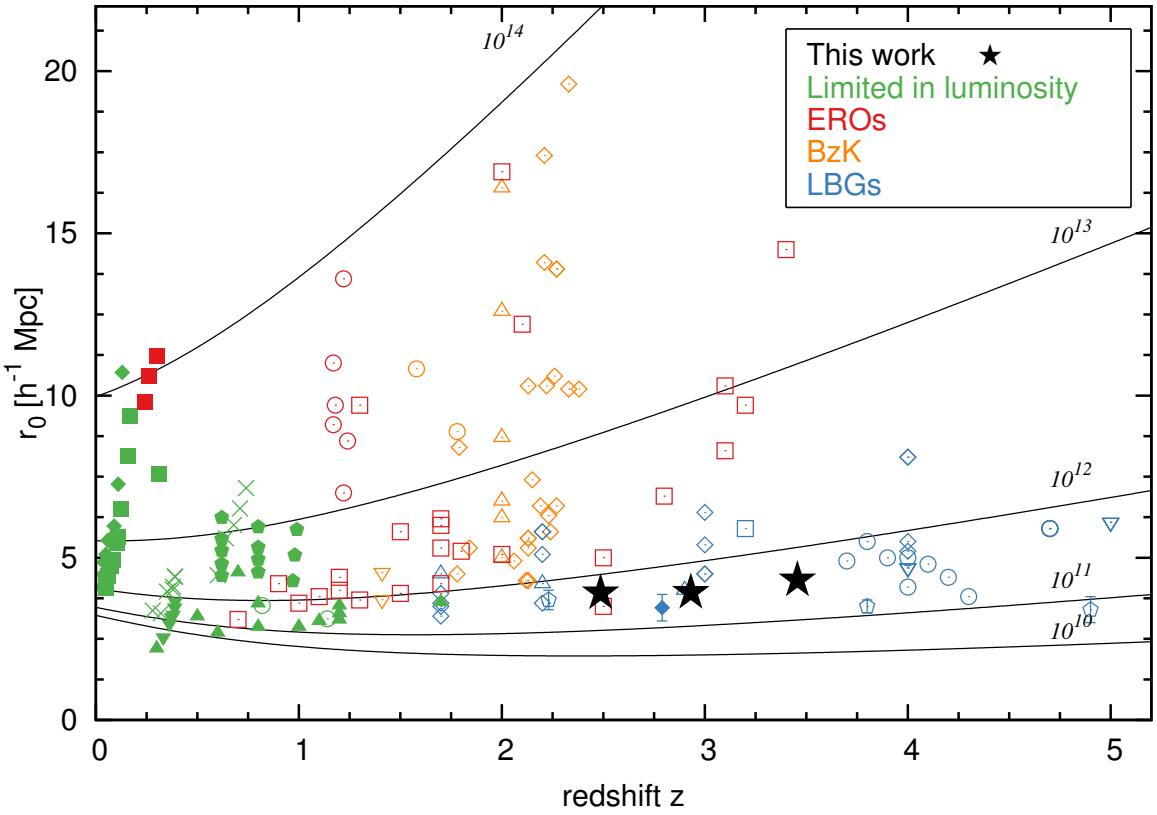


Figure 4.7: Correlation length r_0 as a function of redshift from various surveys of different types of objects. Black stars indicate the measurements obtained from this work in three redshift ranges. Different colours indicate different galaxy populations targeted by using various techniques, as indicated in the upper right corner. Open symbols indicate measurements based on photometric data, while filled symbols are for measurements from spectroscopic data. *Blue:* LBG galaxies (Foucaud et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Savoy et al. 2011, Bielby et al. 2013, Barone-Nugent et al. 2014). *Orange:* BzK galaxies (Blanc et al. 2008, Hartley et al. 2010, McCracken et al. 2010, Lin et al. 2012). *Green:* Galaxy samples from surveys limited in luminosity (Norberg et al. 2002, Coil et al. 2006, Le Fèvre et al. 2005, Pollo et al. 2006, Zehavi et al. 2011, Marulli et al. 2013, Skibba et al. 2009). *Red:* EROs or massive red galaxies (Daddi et al. 2003, Zehavi et al. 2005a, Brown et al. 2008). The five solid curves show the correlation length of dark halos with different minimum masses, as labelled.

A direct comparison of the galaxy clustering strength at different epochs is a difficult task for several reasons. Different measurements are based on various galaxy samples, each with a unique selection function. Since galaxies cluster differently depending on their properties (like luminosity, colours or stellar masses), and in practice every catalogue contains a galaxy population somewhat different, the differences in clustering measurements may primarily be attributed to different galaxy sample properties as much as to evolutionary effects. The inhomogeneities between galaxy samples are especially pronounced for the high redshift samples. At $z > 2$ correlation functions have mostly been measured from surveys targeting specific classes of galaxies, e.g. extremely massive red objects or sources selected using a Lyman-break technique selecting star-forming galaxies. In addition, until recently most clustering measurements at $z > 2$ were produced from the angular correlation function computed on photometric samples with the knowledge of the redshift being only approximate and based on color selection (BzK e.g. Daddi et al. 2003, McCracken et al. 2010; LBG e.g. Adelberger et al. 2005; LAE e.g. Ouchi et al. 2004). Compared to spectroscopic samples, the lower accuracy and precision in redshift measurements in these samples and contamination in a given redshift bin from galaxies at other redshifts introduces a blurring of large scale structures which is difficult to overcome even when using large samples.

Our r_0 estimations are in excellent agreement with other available results at similar high redshifts (e.g. Daddi et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Hartley et al. 2010, Savoy et al. 2011), as presented in Figure 4.7. While r_0 in the VUDS sample is on the lower side of previous clustering measurements at these redshifts, most other samples have specific selection functions from which is it expected that they have a higher clustering bias than our sample, particularly for the redder or more luminous samples. Our measurements can be compared to the measurements by Bielby et al. (2013), based on a large sample of 2135 spectroscopically confirmed LBGs in $\sim 10\,000$ arcmin^2 at $z = 2.8$, the largest sample with a spectroscopically-based measurement from a large volume at these redshift prior to our measurements. They find a clustering length $r_0 = 3.46 \pm 0.41$ at $z = 2.79$ and a slope $\gamma = 1.5 - 1.6$, flatter than for local galaxies, in good agreement with our measurements given the associated errors. Savoy et al. (2011) measured a correlation length $r_0 = 3.5 \pm 0.5$ from ~ 1400 UV-selected galaxies with photometric redshifts $z \sim 3$ and $23 < R < 27$ in the Keck Deep Fields, and Geach et al. (2012) found $r_0 = 3.7 \pm 0.3$ from the clustering of Hα emitters (HAEs) at $z = 2.23$ from HiZELS, both results being very similar to ours at comparable redshifts. We conclude that our measurements are in general agreement with the most recent and least biased measurements available in the literature. Our results are among the most reliable today given the VUDS survey area and the number of galaxies used in our study.

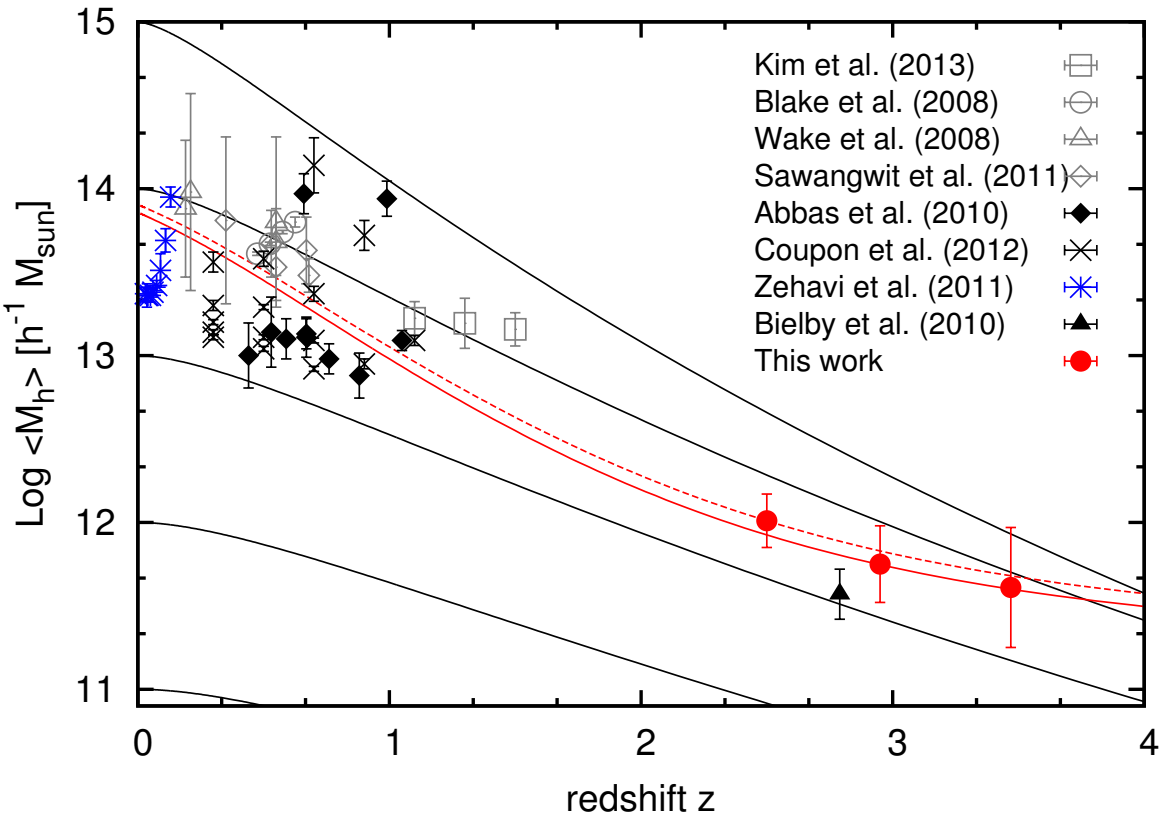


Figure 4.8: The evolution of the number-weighted average host halo mass given by Eq. 3.31 for the three redshift ranges analysed in this study. The red filled circles indicate mass estimations from VUDS. Black and grey symbols represent the results of previous work based on spectroscopic and photometric surveys respectively. The solid black lines indicate how a host halo of a given mass M_0 at $z = 0$ evolves with redshift, according to the model given by van den Bosch (2002). The solid red line represents the halo mass evolution derived using Eq. 4.3, with the HOD parameters obtained from the best-fit HOD model at a redshift $z \sim 3$. The dashed red line is using the HOD best-fit parameters for $z \sim 2.5$. VUDS galaxies with a typical $L_\star$ luminosity are likely to evolve into galaxies with a luminosity $>L_\star$ today.

4.6 Evolution of the host dark matter halo

We draw in Figure 4.7 the model predictions for the evolution of r_0 of dark matter halos following the prescription given by Mo & White (2002) (which is based on the Press & Schechter (1974) formalism). According to this model, dark matter halos of an average mass $\langle M_h \rangle \sim 10^{11.5} M_\odot$ have the same r_0 as observed VUDS galaxies at $z \sim 3$. If we take the rough assumption that at $z \sim 3$ haloes are occupied mostly by only one galaxy (central), this mass would be the average mass of a halo hosting the typical VUDS galaxy. However, this result is only a first approximation, since it is based on simplified assumptions of the Press & Schechter (1974) theory and needs to be refined.

The halo occupation model (HOD) is expected to give more accurate predictions, since it accounts for the actual number and distribution of galaxies occupying haloes. We obtain the number weighted average mass $\langle M_h \rangle$ of haloes hosting VUDS galaxies using the best-fit HOD parameters (Eq. 3.31), in three redshift ranges. We find that the halo mass decreases with redshift (see Figure 4.8), evolving from $\log\langle M_h \rangle = 11.61^{+0.46}_{-0.40} h^{-1} M_\odot$ at $z \sim 3.5$ to $\log\langle M_h \rangle = 12.01^{+0.16}_{-0.18} h^{-1} M_\odot$ at $z \sim 2.5$. These estimations are in reasonable agreement with the recent observations by Bielby et al. (2013) who found an average mass of $\log\langle M_h \rangle \sim 11.57 \pm 0.15 M_\odot$ for haloes hosting LBG galaxies at a redshift $z \sim 2.79$.

It can also be seen from Fig. 4.8 that our estimated host halo masses are much lower than the ones typically observed in the local Universe. At $z \sim 3$ the typical host halo masses are located in the range between $\langle M_h \rangle \sim 10^{11} - 10^{12} M_\odot$ while locally these masses usually reach the values between $10^{13} - 10^{14} M_\odot$. These results indicate that at some stage dark matter halos must have experienced a rapid accretion phase, as expected in the framework of the hierarchical mass assembly.

A universal formula for the mass growth of cold dark matter halos has been derived by van den Bosch (2002). The history of a halo with a given mass M_0 at $z = 0$ can be traced back in time using a simple formula:

$$\log\langle \Psi\left(M_0, z\right)\rangle = -0.301 \left[\frac{\log\left(1+z\right)}{\log\left(1+z_f\right)}\right]^{\nu}, \tag{4.2}$$

where z_f and ν are free fitting parameters that depend on halo mass and cosmological parameters. To obtain these fitting parameters we follow the analytical formula from van den Bosch (2002). We then trace the evolution of the galaxy population sampled by VUDS from $z \sim 3$ to the present epoch to predict the mass of halos hosting the present day descendants of the VUDS galaxies. We find that in this model the typical VUDS halo with a mass $\langle M_H \rangle \sim 10^{11.75} M_\odot$ at $z \sim 3$ should evolve into a halo with a mass $\langle M_H \rangle \sim 10^{13.5} M_\odot$ at $z = 0$. In the local SDSS galaxy sample Zehavi et al. (2011) found that halos with these masses are typically occupied by star forming galaxies with

luminosity $M_r < -20.5$. According to the van den Bosch (2002) model the star forming galaxies at $z \sim 3$ in VUDS with a typical characteristic luminosity $L_\star$ would likely evolve into galaxies equivalent to or brighter than $L_\star$ at the present day.

The above comparison assumes that each halo is occupied by only one galaxy. This is not expected to generally be the case, and the above picture, while broadly correct might need to be adjusted. In order to trace the evolution of dark matter haloes and the hosted galaxy population in a more realistic way we use both the halo mass growth model $\Psi(M_0, z)$ and the halo occupation function $\langle N_g|M\rangle$ at redshift $z = 3$. The average halo mass M_h as a function of redshift z is measured taking (see Eq. 3.31 for comparison):

$$\langle M_h\rangle(z) = \int dM \Psi^{-1}(M, z) n(M, z) \frac{\langle N_g|M\rangle}{n_g(z)}, \tag{4.3}$$

here the $\Psi^{-1}(M, z)$ is the inverse mass growth function proposed by van den Bosch (2002) (Eq. 4.2), $n(M, z)$ is the dark matter mass function, and n_g the galaxy number density. This allows us to trace the history of the typical dark matter halo hosting the average VUDS galaxy from a redshift $z = 3$ to the present day, as presented in Figure 4.8. We find that the typical halo of mass $\langle M_h\rangle = 10^{11.75}$ at $z \sim 3$ should evolve into a halo of mass $\langle M_h\rangle = 10^{13.9}$ at redshift $z = 0$. Comparing again our results from this improved model with Zehavi et al. (2011), halos with these masses are typically occupied by star forming galaxies with $M_r < -21.5$, above the characteristic luminosity $M_\star$ at $z \sim 0$. This means that the general galaxy population observed at redshift $z = 3$ evolves into the brightest and most massive galaxies, which occupy the most massive halos observed in the local Universe, as is expected in the hierarchical mass growth paradigm.

4.7 Relation between dark and luminous matter at high redshift

The hierarchical model of galaxy formation predicts an increase of the galaxy bias with redshift Mo & White (1996). In this section we discuss the relation between the distribution of galaxies and the underlying dark matter density field at different redshifts, by comparing the galaxy bias values measured from VUDS to the bias of galaxy populations at different redshifts from the literature.

Figure 4.9 shows the linear large-scale galaxy bias b_L, computed using methods based on power-law b_L^{PL} (see Sec. 3.6) and HOD b_L^{HOD} approximations (see Sec. 3.2). The results obtained from both methods are comparable within errors, and small differences are likely due to the less accurate power-law approximation. The general VUDS galaxy population shows a bias decreasing with decreasing redshift. According to our measurements, from

$z \sim 3.5$ to $z \sim 2.5$ the linear bias decreased by $\Delta b_L \sim 0.9$. Both our values at $z = 2.5$ and $z = 3.5$ are significantly higher than measurements at later epochs reported in the literature. For comparison, we plot the linear large–scale bias inferred for the intermediate ($z = [0.5 - 1.5]$) redshift range by Marinoni et al. (2005), based on VVDS-Deep survey, for $z > 2$ by Geach et al. (2012) based on observations of $H\alpha$ emitters, and by Barone-Nugent et al. (2014) for $z \sim 3.8$ $LBGs$, along with the values we computed using the best-fit power-law parameters r_0 and γ from Le Fèvre et al. (2005), Adelberger et al. (2005) and Bielby et al. (2013) (the method is described in section 3.6). The values presented in Table 4.3 are similar to the bias measurements recently observed for $LBGs$ at redshift $z > 2$. Bielby et al. (2013) found $b_{gal} = 2.33 \pm 0.17$ from their LBG-selected spectroscopic sample, and at an even higher redshift $z \sim 3.8$ Barone-Nugent et al. (2014) estimated $b_L = 3.0 \pm 0.2$ for similarly selected galaxies.

The visible decrease of the galaxy bias with cosmic time can be explained in terms of the hierarchical scenario of structure formation. In this framework galaxies formed at early epochs in the most dense regions (highest peaks of the density field) which are the most biased with respect to the underlying average mass density field. As the mass density field evolved with time, galaxy formation is then expected to systematically move into less dense, hence less biased, regions. Additionally, it is likely that the gas in the most dense regions became too hot for collapse, and thus star formation could not take place in the strongest over-densities any more (e.g. Blanton et al. 1999). Therefore, the observed decrease of galaxy bias with cosmic time may be explained as galaxy formation is moving to less dense areas as cosmic time increases in a downsizing trend (De Lucia et al. 2006).

In Figure 4.9 we also draw the predicted evolution of the large-scale galaxy bias using different prescriptions. We consider three different theoretical descriptions of the biasing functions, based on different ideas of how evolution may proceed: the *passive*, the *merging*, and the *star forming* biasing models (see e.g. Marinoni et al. 2005). In the first model the number of galaxies (halo occupation) is conserved as a function of time and the bias evolves assuming the form presented in Eq. 3.32. The second model takes into account galaxy merging, as proposed by Mo & White (1996), who gave analytical prescriptions for computing the bias of halos using the Press & Schechter formalism. If we assume that galaxies can be identified with dark matter halos, an approximate expression for the biasing of all halos of mass $> M$ existing at redshift z (but which collapsed at redshift greater than the observation redshift, see discussion in Matarrese et al. 1997) is given by:

$$b(M, z) = 1 + \frac{1}{\delta_c} \left(\frac{\delta_c^2}{\sigma^2(M, z)} - 1 \right) \tag{4.4}$$

where $\delta_c \simeq 1.69$ is the linear overdensity of a sphere which collapses in an Einstein-de

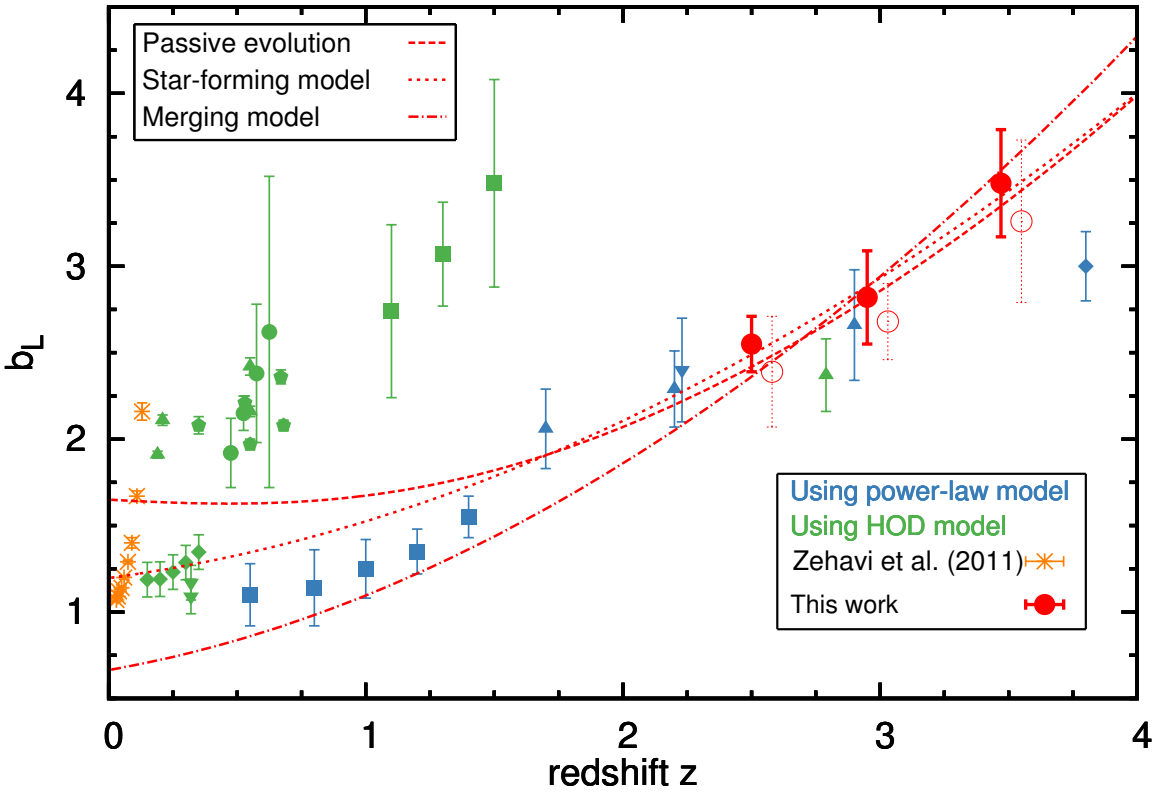


Figure 4.9: Large-scale linear galaxy bias retrieved by using two methods. Red filled circles indicate measurements of b_L^{HOD}, while the open red circles represent b_L^{PL}) in three redshift ranges. Green and blue point represents the results from previous works, retrieved by using HOD (Blake et al. 2008, Wake et al. 2008, Ross & Brunner 2009, Sawangwit et al. 2011, Kim et al. 2014) and power-law frameworks (Marinoni et al. 2005, Adelberger et al. 2005, Geach et al. 2012, Barone-Nugent et al. 2014, Bielby et al. 2013) We show 3 different models for the evolution of bias: a *passive* evolution model (red dashed line) a *merging* model (red dot-dashed line), a *star-forming* model (dotted line), as described in the text.

Sitter Universe and $\sigma(M, z)$ is the linear rms fluctuations on scales corresponding to mass M at the redshift of observation. The *star-forming* model assumes that the distribution of galaxies with luminosity $> L$ is well traced by halos with mass $> M$, and predicts the biasing of objects that just collapsed at the redshift of observation (e.g., Blanton et al. 2000). In this model,

$$b(M, z) = 1 + \frac{\delta_c}{\sigma^2(M, z)} \tag{4.5}$$

represents the biasing of galaxies that formed in a narrow time interval around redshift z (i.e., galaxies which experienced recent star formation at redshift z).

From the halo mass analysis presented in Section 4.6 we infer that VUDS galaxies would evolve into galaxies with a luminosity $M_r < -21.5$ in the local universe (Section 4.6), for which the bias is $b_L = 1.7$. The star forming model predicts that the VUDS galaxy population would evolve into a weakly biased class of galaxies, corresponding to faint low mass galaxies. The merging model seem to have an evolution too steep compared to what is observed as it predicts an evolution to a bias significantly lower than populations of galaxies with the lowest clustering at z=0 (Zehavi et al. 2011). The model that would make the VUDS galaxies with $b_L \sim 2.8$ at z=3 evolve to $b_L = 1.7$ at z$\sim$0 is the passive model. However, it is likely that more sophisticated models would need to be developed to reproduce the joint evolution of DM halo mass and galaxy bias.

4.8 Summary

In this chapter I described a comprehensive study of general galaxy population clustering properties at redshift range $2 < z < 5$. I computed projected two-point correlation function $w_p(r_p)$ based on a spectroscopically confirmed data sample, containing in total 3022 galaxies selected form VUDS survey. The results have been interpreted in the framework of power-law and HOD model approximations.

The correlation function measurements and its power-law fitting (see Chapter 3) have been performed by using the code originally developed by Luigi Guzzo and Agnieszka Pollo (Pollo et al. 2005) and adopted into C language by Sylvain de la Torre, to whom we give all credits. I adapted the code to the VUDS survey characteristics, in particular setting-up the tilling geometry of the 56 different VIMOS quadrants used in this study. The $w_p(r_p)$ modelling in the HOD framework (see Sec. 3.2) have been performed using the code develop for this study. Full code documentation can be found in Appendix B. Finally, in order to perform the HOD fitting, I implemented the coded HOD model to the publicly available CosmoPMC [1] code, which adopt the "Population Monte Carlo" (PMC)

[1] `http://www2.iap.fr/users/kilbinge/CosmoPMC/`

technique to sample likelihood space (Wraith et al. 2009, Kilbinger et al. 2011). All methods and codes have been extensively tested before performing the final measurements from the VUDS data in order to check reliability, limitations and correct errors (for example see Section 3.5). Thus I am assured that all the tools I used are robust.

Clustering studies presented in this chapter are among the first ones performed using spectroscopically confirmed galaxies at high redshifts ($2 < z < 5$) and they, therefore, have a higher accuracy and lower systematics than previous studies. The strength of the clustering of the star-forming galaxies within the VUDS survey mildly increases with redshift, which can be observed through the increase of the correlation length, from $r_0 = 3.95 \pm 0.48$ at $z \sim 2.5$ to $r_0 = 4.35 \pm 0.60$ at $z = 3.5$. The slope of the correlation function is found to be $\gamma = 1.81^{+0.02}_{-0.06}$ and $1.60^{+0.12}_{-0.13}$ at $z = 2.5$ and 3.5, respectively, similar or slightly flatter than the canonical $\gamma = 1.8$ observed at low redshifts. These measurements are in agreement with the most recent and least biased measurements available in the literature at similar redshifts. Additionally, the power-law parameters I derive from the correlation function measurements are comparable to the clustering parameters for star forming galaxies at intermediate redshifts, but the clustering length is lower than for galaxies in the local Universe.

Through the use of the HOD modelling of two-point correlation function, I put constrains on the properties of the underlying dark matter haloes hosting the observed galaxies and trace their evolution with time. I found that the average halo mass obtained from the HOD model fit reaches $\langle M_h \rangle = 10^{11.75} h^{-1} M_\odot$ at redshift $z = 3$, which is lower than observed locally. By assuming the mass growth model from van den Bosch (2002) these halos should evolve into halos with a mass $\langle M_h \rangle (z = 0) = 10^{13.9} h^{-1} M_\odot$ at $z = 0$. As halos with such a mass are occupied by star forming galaxies with $M_r < -21.5$ in the local universe, I infer that the star-forming galaxy population observed at a redshift $z \sim 3$ evolved into the massive and bright galaxy population observed today, as predicted in the hierarchical mass growth paradigm.

Finally, I have been able to answer the questions about the relation between dark and luminous matter at high redshift, by measuring the large-scale galaxy bias. I observed that the linear large-scale galaxy bias b_L computed by using best-fit correlation parameters, increases from $b_L^{PL} = 2.68 \pm 0.22$ at $z = 2.5$ to $b_L^{PL} = 3.26 \pm 0.47$ at $z = 3.5$ and on average is much higher than the bias values observed at low and intermediate redshift ranges. I confirmed these measurements by obtaining the galaxy large-scale bias using the HOD model. I found the results to be comparable to the ones obtained from power-law parameters, reaching a value $b_L^{HOD} = 2.82 \pm 0.27$ at redshift $z \sim 3$. These bias values are comparable with measurements obtained at similar redshifts, but are much higher than bias values observed at low and intermediate redshift ranges, where b_L reaches values in

the range $1-2$. I discuss simple models to link the z$\sim$3 galaxies in our sample to lower redshift populations, and find that a combination of passive halo evolution maintaining halo parameters constant and a star-forming evolution would be needed to reproduce the bias evolution. However, as we do observe evolution in the halo parameters, it seems that these models need to be refined to properly reproduce the bias evolution.

The VUDS data provide a robust reference for the clustering of galaxies at an epoch before the peak in star formation rate. I infer that the star-forming population of galaxies at $z \sim 3$ should evolve into the massive and bright ($M_r < -21.5$) galaxy population which typically occupies haloes of mass $\langle M_h \rangle = 10^{13.9}$ at redshift $z = 0$. More complete modelling seems necessary to relate high redshift to low redshift galaxy populations, possibly taking into account non-linear effects, to better explain the galaxy clustering properties at early epochs in the life of the universe.

All results presented in this chapter can be regarded as the precursor to many other studies that can be carried out with large samples at high redshift, including e.g. more detailed analyses with galaxy samples of different colour, stellar masses and luminosities.

CHAPTER 5

Luminosity and stellar mass dependence of galaxy clustering at $z \sim 3$

Galaxy clustering highly dependents on the various properties of galaxies. It has been shown in many studies, performed both at $z < 2$, that the luminous and massive galaxies tend to be more strongly clustered than their less luminous and less massive counterparts. Additionally, it has been found that the clustering strength varies as a function of morphology, colour and spectral type. Galaxies with bulge dominated morphologies, red colours, or spectral types indicating old stellar population also exhibits stronger clustering and a preference for dense environments (Norberg et al. 2002, Zehavi et al. 2004, 2012, Pollo et al. 2006, de la Torre et al. 2007, Coil et al. 2008, Quadri et al. 2008, Meneux et al. 2006, 2008, 2009, Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012). Naturally, important is to test if the same dependencies apply for the high redshift ranges $z > 2$. This task, however, has been difficult due to the limitations of the available data samples (as discussed in Sec. 2.4).

In this paragraph I take an advantage of the large VUDS data sample and compute the projected two-point correlation function $w_p(r_p)$ for the galaxy samples limited in luminosity and stellar mass in two redshift ranges. This allowed me to put constrains on the clustering dependency of these two galaxy properties. Below I present the results of correlation function measurements and power-law fitting, as well as the measurements of the large-scale galaxy bias. Also, I discuss the results in terms of the current scenario of mass growth.

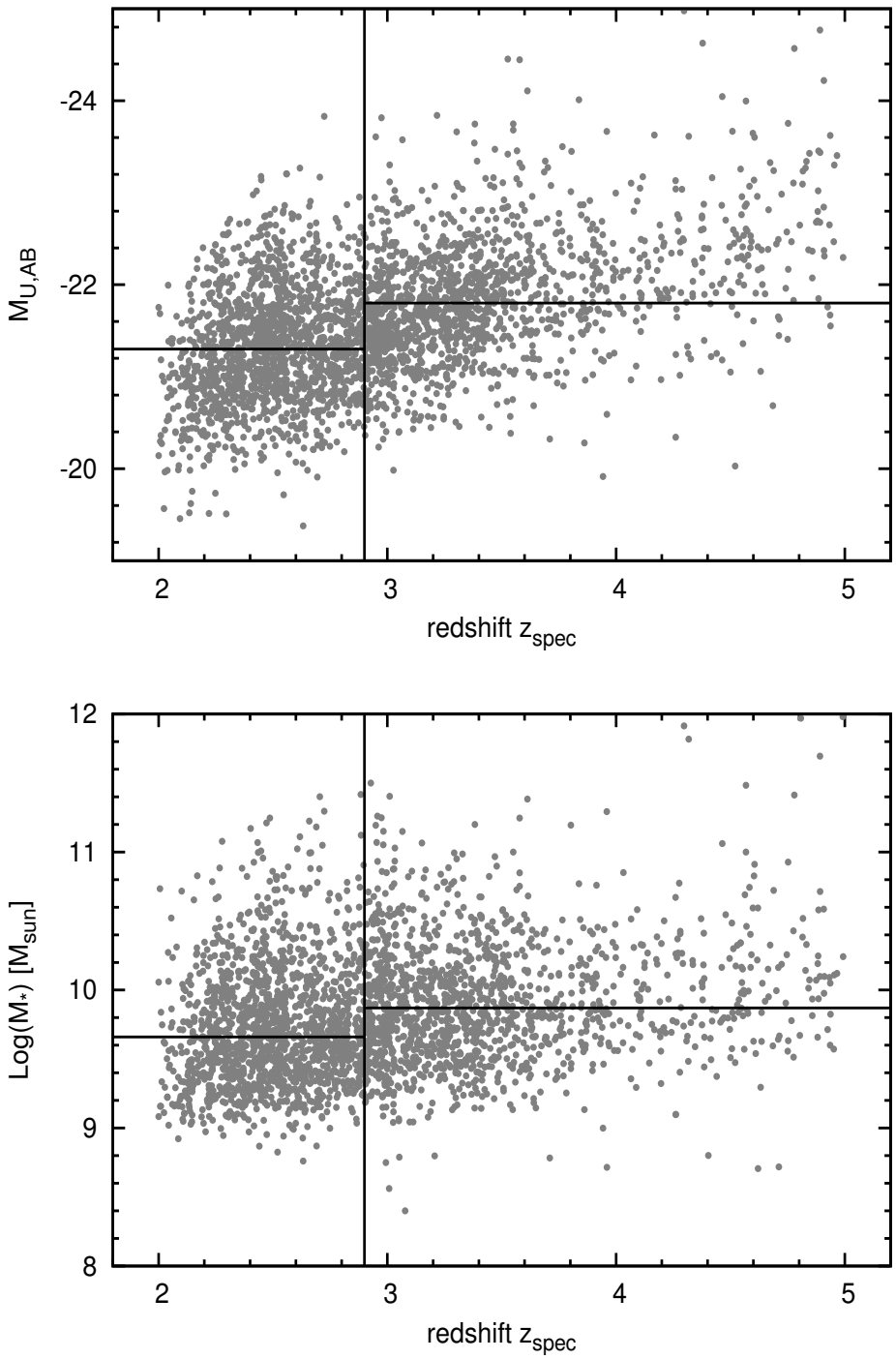


Figure 5.1: The distribution of absolute U-band magnitudes (top) and stellar mass (bottom) in used galaxy sample. Solid vertical lines indicates the boundaries of our redshift slices, while the horizontal lines show the cuts of luminosity and stellar mass sub-samples.

Table 5.1: Properties of luminosity limited VUDS sub-samples in the redshift range $2.0 < z < 5.0$

M_U *range*	M_U^{median}	z range	z_{mean}	N_{gal}	r_0	γ	b_L
> -21.3	-20.84	$[2.0 - 2.9]$	2.46	765	$4.23^{+0.36}_{-0.40}$	$1.80^{+0.09}_{-0.09}$	$2.51^{+0.24}_{-0.21}$
< -21.3	-21.80	$[2.0 - 2.9]$	2.54	767	$5.11^{+0.74}_{-0.84}$	$1.71^{+0.13}_{-0.10}$	$2.99^{+0.46}_{-0.42}$
> -21.8	-21.34	$[2.9 - 5.0]$	3.30	712	$4.39^{+0.54}_{-0.64}$	$1.65^{+0.13}_{-0.12}$	$3.16^{+0.40}_{-0.34}$
< -21.8	-22.39	$[2.9 - 5.0]$	3.62	713	$5.41^{+0.74}_{-1.04}$	$1.80^{+0.19}_{-0.20}$	$4.15^{+0.90}_{-0.68}$

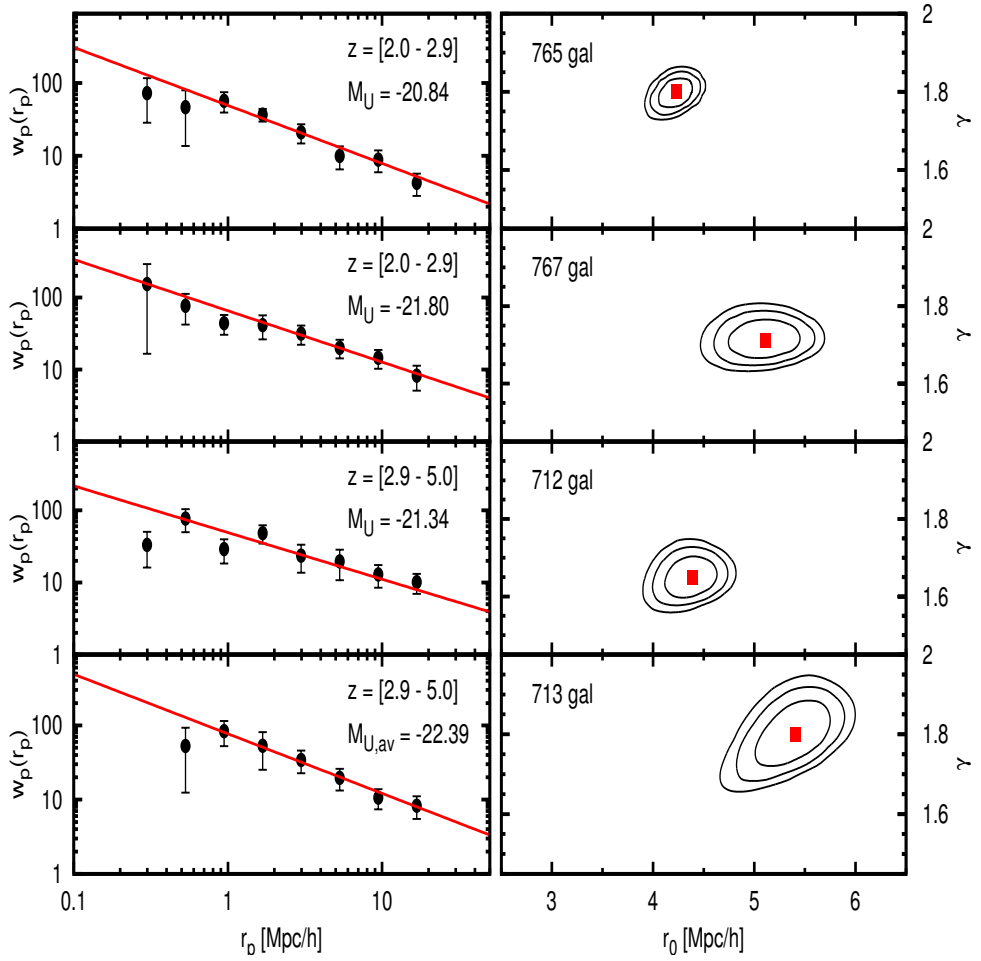


Figure 5.2: The projected two-point correlation function $w_p(r_p)$ and the associated best-fit parameters r_0 and γ along with error contours for the magnitude-threshold sub-samples in two redshift ranges: $z = [2.0 - 2.9]$ - first two rows, and $z = [2.9 - 5.0]$ - last two rows.

Table 5.2: Properties of stellar mass limited VUDS sub samples in the redshift range $2.0 < z < 5.0$

$\log M_*$ *range*	$\log M_*^{median}$	z range	z_{mean}	N_{gal}	r_0	γ	b_L
< 9.66	9.41	$[2.0 - 2.9]$	2.49	766	$3.93^{+0.36}_{-0.38}$	$1.84^{+0.12}_{-0.11}$	$2.37^{+0.23}_{-0.22}$
> 9.66	10.06	$[2.0 - 2.9]$	2.51	769	$4.67^{+0.48}_{-0.48}$	$1.63^{+0.10}_{-0.09}$	$2.73^{+0.25}_{-0.25}$
< 9.87	9.58	$[2.9 - 5.0]$	3.42	734	$4.23^{+0.70}_{-0.94}$	$1.68^{+0.14}_{-0.15}$	$3.16^{+0.61}_{-0.46}$
> 9.87	10.27	$[2.9 - 5.0]$	3.52	694	$4.77^{+0.94}_{-1.24}$	$1.62^{+0.13}_{-0.14}$	$3.55^{+0.79}_{-0.60}$

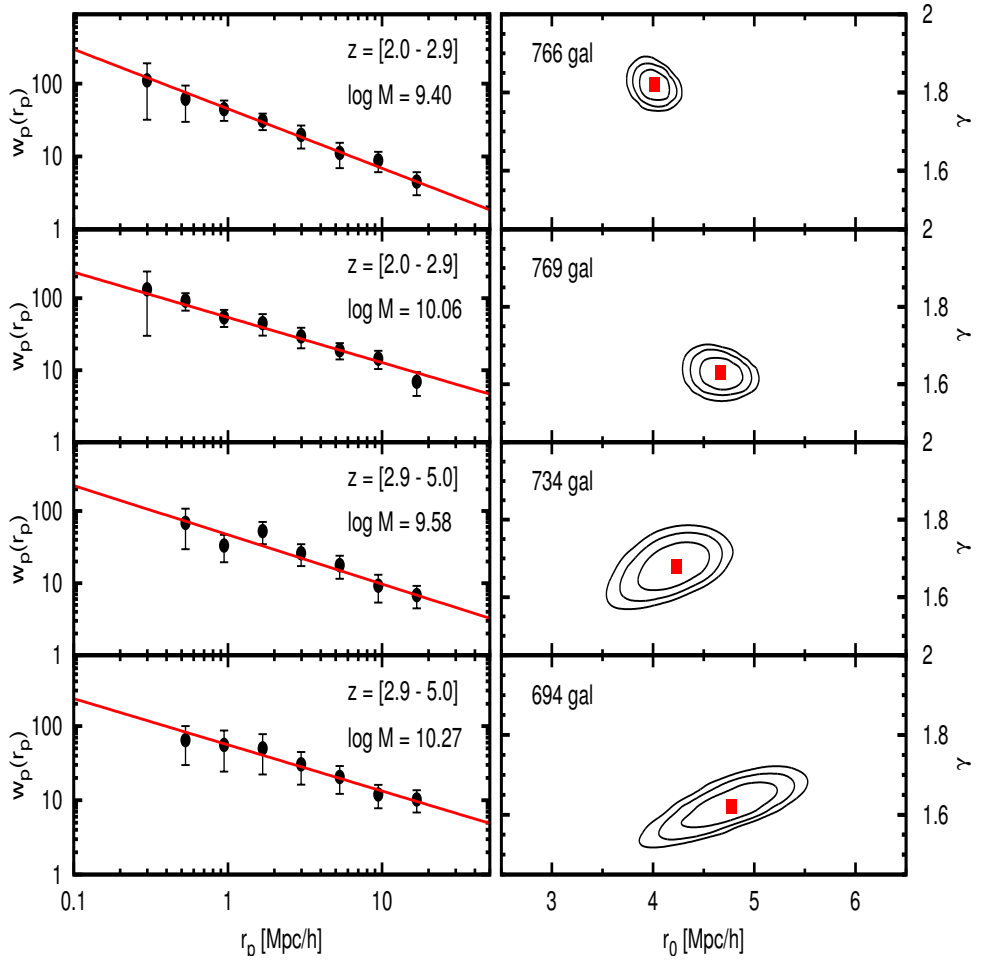


Figure 5.3: The projected two-point correlation function $w_p(r_p)$ and the associated best-fit parameters r_0 and γ along with error contours for the stellar mass-threshold sub-samples in two redshift ranges: $z = [2.0 - 2.9]$ - first two rows, and $z = [2.9 - 5.0]$ - last two rows.

5.1 Sample selection for the analysis

To test for the dependency of clustering on luminosity and stellar mass in the VUDS galaxy sample I created two sub-samples limited by absolute U-band magnitude and stellar mass, within each of two redshift ranges: $2.0 < z < 2.9$ and $2.9 < z < 5.0$, as listed in Tables 5.1 and 5.2. From the $i_{AB} \leq 25$ limit of the VUDS Survey (see Chapter 2 for details) I can define volume-limited samples with $M_U \lesssim -20.5$ for the lower-redshift sample and $M_U \lesssim -21.0$ for the higher-redshift one. The boundaries of the luminosity and stellar mass sub-samples are therefore dictated by the compromise between the need to have volume complete samples and the need to have a statistically significant number of galaxies as shown in Figure 5.1.

5.2 Correlation function results

To quantify the luminosity and stellar mass dependence of galaxy clustering the power-law function $\xi(r) = (r/r_0)^{-\gamma}$ has been fitted to the correlation function measured in each of the stellar mass and luminosity sub-samples. The best-fit parameters are reported in Table 5.1 - for sub-samples limited in luminosity and in Table 5.2 - for sub-samples limited in stellar mass. Figure 5.2 and 5.3 shows the final result of power-law fit along with the associated error contours.

A strong luminosity and stellar mass dependence of galaxy clustering can be observed for the two redshift bins at $z \sim 2.5$ and $z \sim 3.5$, with more luminous or more massive galaxies clustering more strongly than less luminous or less massive galaxies. The correlation length in the $z \sim 2.5$ redshift bin rises from $r_0 = 4.2h^{-1}$Mpc to $r_0 = 5.1h^{-1}$Mpc with the U-band luminosity brighter by one magnitude from $M_U = -20.8$. The same trend is observed in the $z \sim 3.5$ sample, with the correlation length rising even faster with luminosity, with r_0 increasing by $\sim 1h^{-1}Mpc$ when M_U is getting one magnitude brighter from $M_U = -21.3$ to $M_U = -22.3$. A similar behavior is observed for sub-samples limited by stellar mass, with the correlation length increasing by $\sim 0.7h^{-1}$Mpc in the $z \sim 2.5$ sample for masses increasing from an average $\log(M_\star) = 9.4$ to 10.1, and by a $\sim 0.5h^{-1}$Mpc at $z \sim 3.5$ for masses increasing from $\log(M_\star) = 9.6$ to 10.3. For these luminosity and stellar mass limited galaxy samples the correlation slope γ varies between 1.6 and 1.8 without any specific trend.

5.3 Linear large-scale galaxy bias

Based on the clustering parameters r_0 and γ, I compute the large-scale galaxy bias b_L^{PL} (see Sec. 3.6). Similarly to what is known from the lower-z Universe (Pollo et al. 2006, Coil et al. 2006, Meneux et al. 2009, Zehavi et al. 2011, Marulli et al. 2013) we observe a clear segregation with respect to luminosity and stellar mass, with less luminous/massive galaxies being less biased than luminous/massive ones (as reported in Tables 5.1 and 5.2). The values change by $\Delta b_L^{PL} = 0.55$ at $z \sim 2.5$ and by $\Delta b_L^{PL} = 0.99$ at $z \sim 3.5$, for the sub-samples cut in luminosity, and by $\Delta b_L^{PL} = 0.36 - 0.39$ for stellar mass sub-samples, respectively.

These results are in good agreement with the hierarchical model of galaxy formation. Also in the framework of this model we can explain the luminosity dependence of the relative bias which we found. Our results suggest that the brightest (most massive) galaxies correspond to the highest fluctuations of the total mass density field and they are highly biased. On the other hand, fainter and less massive galaxies form and occupy less dense regions, and, as a result, they are less biased with respect to the underlying dark matter field

5.4 Summary

In this chapter I present the results of the study of the luminosity an stellar mass dependence of galaxy clustering. I computed projected two-point correlation function $w_p(r_p)$ for sample of ~ 3000 galaxies separated into luminosity and stellar mass sub-samples in two redshift ranges: $z = [2.0, 2.9]$ and $z = [2.9, 5.0]$ with median $z = 2.50$ and $z \sim 3.47$ respectively. Similarly to the observations performed for lower redshift samples we found that the luminous and more massive galaxies cluster more strongly then less luminous and massive ones in both observed redshift ranges. e.g the r_0 increases with luminosity by $\Delta r_0 = 0.48h^{-1}$Mpc for the low redshift sample and by $\Delta r_0 = 0.99h^{-1}$Mpc for the high redshift. Using the best-fit parameters we computed large-scale galaxy bias. I found that the more luminous/massive galaxies are more strongly biased with respect to the underlying dark matter mass distribution than less luminous/massive, which can be very well explained in the framework of the hierarchical mass growth paradigm. Galaxies at z 3, 10Gyr back in the life of the Universe, were already segregated in luminosity and mass. It is therefore likely that the dominant process leading to this segregation is the evolution of the DM halos rather than galaxy physics.

CHAPTER 6

The stellar mass $M_\star$ – halo mass M_h relation at z∼3

In this thesis I study the evolution of the large-scale structure of the universe, in particular the evolution of luminous-dark matter relation. In the previous chapter I focused on the clustering properties of galaxies. I have shown that the dark matter haloes grow and change from $z \sim 2.5$ to $z \sim 3.5$. These changes naturally effects the galaxies reside inside these dark matter haloes. In this chapter I study the effectiveness of processes, which regulate star formation and mass growth in galaxies along cosmic time.

Galaxies are thought to form via dissipative collapse in the deep potential wells of dark matter haloes (e.g. White & Rees 1978, Fall & Efstathiou 1980). In this paradigm, cooling processes bring baryons in the high density peaks of the halos, where the conditions for gas fragmentation triggering star formation (Bromm et al. 2009). Current models connecting star formation and stellar mass evolution on the one hand and the formation histories of DM halos on the other hand are relying on approximations in the physics used and are in need to be further informed by observational data to reduce the uncertainties in the modelling process (e.g. Conroy & Wechsler 2009).

The efficiency of the baryons assembly into stars is an important ingredient to understand galaxy formation but remains poorly constrained observationally. In recent years it has been proposed to derive this efficiency from the comparison of DM halo masses with galaxy stellar masses. With the measurement of the characteristic mass of dark matter halos M_h now available from observational data and of stellar mass $M_\star$ derived from the analysis of the spectral energy distribution of galaxies, coupled to the knowledge of the cosmological density of baryons and DM, one can infer the conversion rate from baryons to stellar mass.

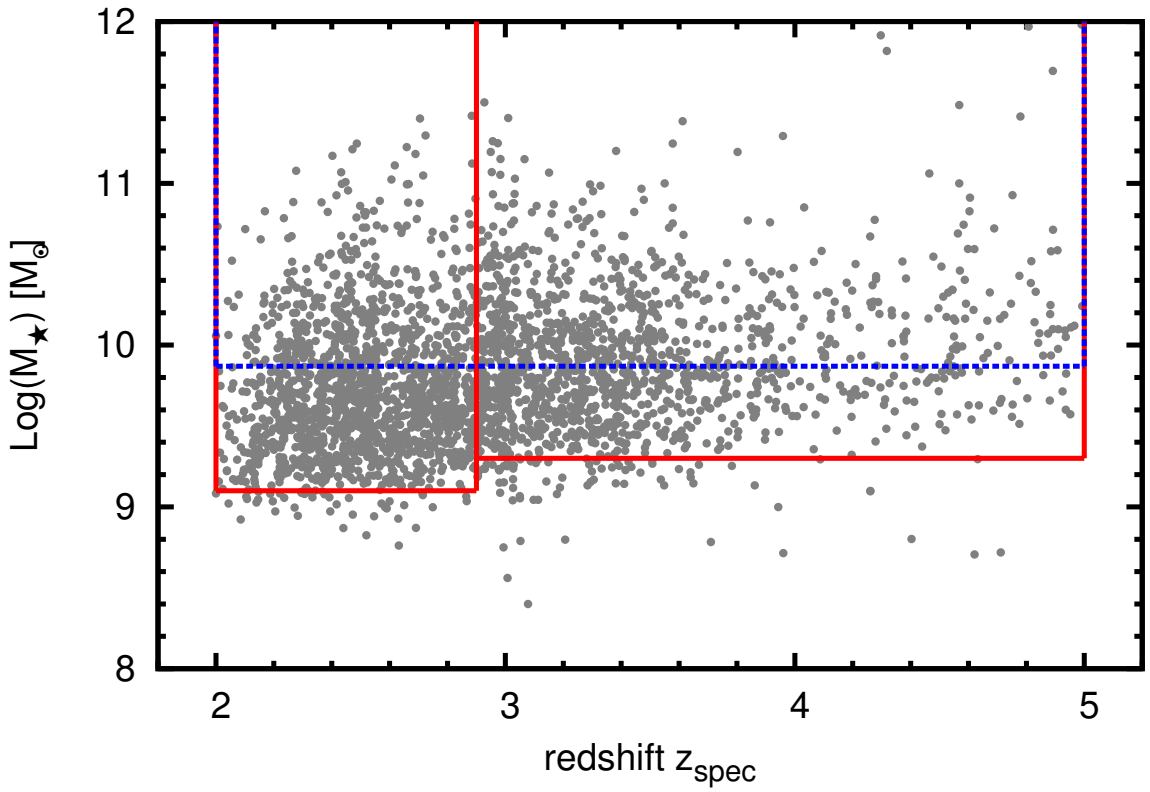


Figure 6.1: The stellar mass distribution of selected VUDS sample. Vertical red solid lines indicates the boundaries of our redshift cuts, while the horizontal lines indicates the cuts in stellar mass. The dashed blue line indictate mass cut at $M_{\star}$=7.4×10^9 $M_{\odot}$ applied to define a high-mass sample (see text).

Two methods have been used so far to make this link between $M_\star$ and M_h. Halo occupation models (e.g. Zehavi et al. 2005b, Leauthaud et al. 2012) and the abundance matching (Conroy et al. 2009, Moster et al. 2013). The efficiency with which the galaxies converted baryons into stars is encoded in how $M_\star$ depends on M_h as a function of redshift, providing a benchmark against which galaxy evolution models can be tested. Using observed stellar mass functions, abundance matching models have led to the derivation of the Stellar Mass to Halo Mass ratio (SHMR), SHMR=$M_\star/M_h$ (Behroozi et al. 2010). Behroozi et al. (2010) find that the integrated star formation efficiency (ISFE) at a given halo mass peaks at 10-20% of available baryons for all redshifts from 0 to 4.

The SHMR is characterized by a peak at around $M_h = 10^{12} M_\odot$, with a lower ratio at lower and higher mass (Behroozi et al. 2013, Behroozi & Silk 2014). The lower efficiency at masses below this peak may indicate that supernova feedback might be sufficient to remove gas from the galaxy as the halo gravitational potential is lower (e.g. Silk 2003, Bertone et al. 2005, Béthermin et al. 2013). At higher masses, the decrease in star formation efficiency might be produced when cold stream are replaced by isotropic cooling (e.g. Dekel & Birnboim 2006, Faucher-Giguère et al. 2011) or some high energy feedback process like that produced by AGNs.

In this chapter I report on the first measurements of the SHMR from galaxies at $2 < z < 5$. We use the halo mass M_h derived from HOD modelling of the 2-point correlation function $w_p(r_p)$, and stellar masses derived from SED fitting (see Chapter 2 for details) to derive the SHMR for several galaxy samples, and compare it to SHMR models.

6.1 Data sample selection

As for the previous studies we use 3022 galaxies with reliable spectroscopic redshifts from VUDS survey (see Chapter 2). The galaxies in this sample have a stellar mass in the range $9 < log(M_*) < 11 M_\odot$ as presented in Figure 6.1. The whole sample has been divided into two redshift ranges: $2 < z < 2.9$ with $\log M_\star^{tresh} = 9.1 M_\odot$ and $2.9 < z < 5.0$ for which $\log M_\star^{tresh} = 9.3 M_\odot$, where $M_\star^{tresh}$ is the lower mass boundary of the sample resulting from the survey selection function. Additionally in order to estimate the SMHR for more massive galaxies we define a sub-sample from galaxies in the redshift range $2 < z < 5$ and with $\log M_\star > 9.87 M_\odot$. This mass limit is the practical limit for which we can measure a galaxy correlation function signal accurately enough at each observed scale $0.1 < r_p < 10$ h^{-1}Mpc for the HOD fit to converge.

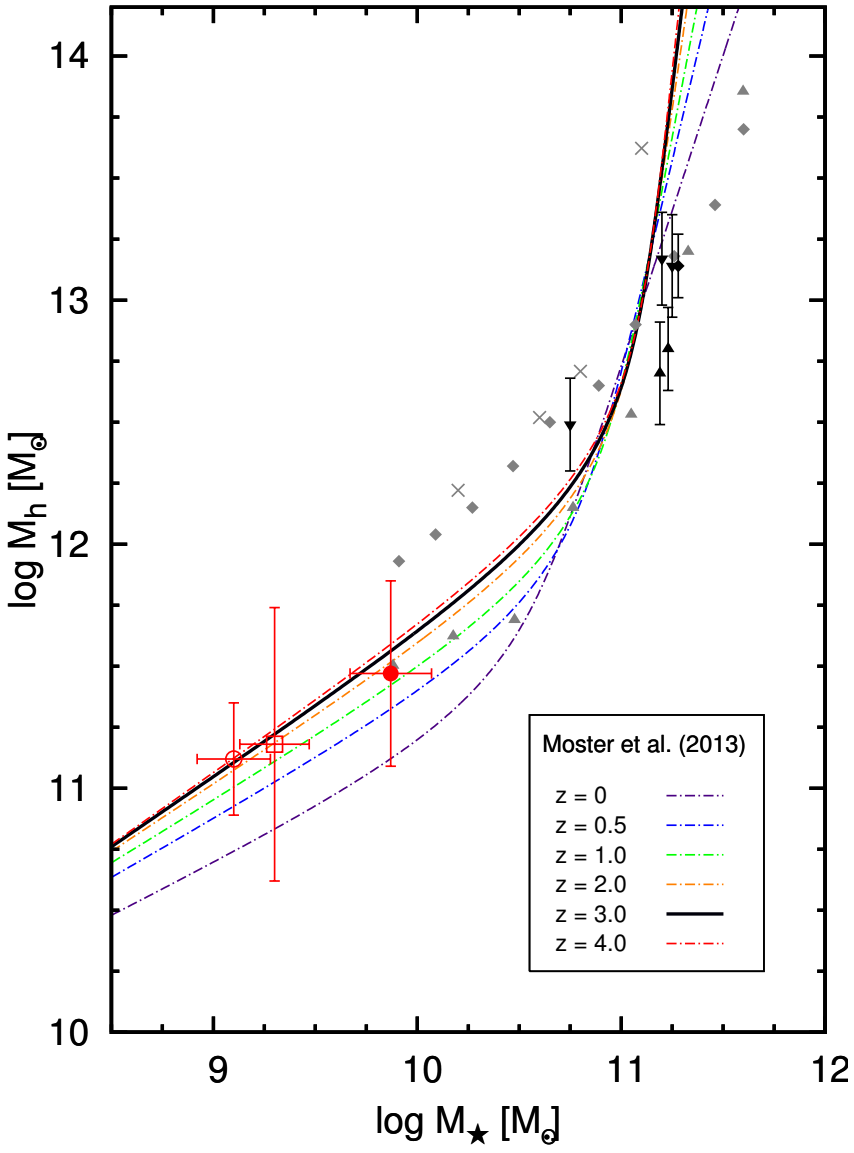


Figure 6.2: The relation between the stellar mass $M_\star$ and the halo mass M_h in the VUDS survey for different stellar mass and redshift cuts (red symbols). $M_\star$ is derived from SED fitting of the multi-wavelength photometric data using the known spectroscopic redshift; the error bar in $M_\star$ indicates the expected uncertainty of the SED fitting method. M_h is obtained from HOD modelling of the two-point correlation function in different redshift and mass ranges. The VUDS data is compared to measurements at low and intermediate redshift form satellite kinematics (SK, Conroy et al. 2007, More et al. 2011, weak lensing (WK, Mandelbaum et al. 2006), galaxy clustering (GC, Foucaud et al. 2010). The lines represents the model predictions derived from abundance matching at various redshift (Moster et al. 2013)

Table 6.1: The stellar mass to halo mass ratio (SMHR), and the integrated star formation efficiency (ISFE) in the VUDS survey

Redshift range	z_{mean}	Stellar mass range	$\log M_\star^{tresh}$	$\log M_h^{min}$	SMHR$\times 10^2$	ISFE $^a\times 10^2$
$[2.0-2.9]$	2.50	$[9.10-11.40]$	$9.10^{+0.15}_{-0.16}$	$11.12^{+0.33}_{-0.36}$	$0.95^{+0.50}_{-0.35}$	$6.16^{+3.23}_{-2.26}$
$[2.9-5.0]$	3.47	$[9.30-11.40]$	$9.30^{+0.17}_{-0.19}$	$11.18^{+0.56}_{-0.70}$	$1.32^{+0.98}_{-0.57}$	$8.52^{+6.32}_{-3.68}$
$[2.0-5.0]^b$	3.00	$[9.87-11.40]$	$9.87^{+0.13}_{-0.15}$	$11.47^{+0.38}_{-0.43}$	$2.51^{+1.23}_{-0.89}$	$16.19^{+7.94}_{-5.74}$

^{a}defined as SMHR$\times\Omega_m/\Omega_b$
^{b}high mass sample

6.2 The star formation efficiency

Our results are presented in Table 6.1 and in the upper panel of Figure 6.2. We find that for the low redshift sample $z\sim 2.5$ the stellar mass for halos of mass $\log M_h = 11.12\pm 0.33$ is $\log M_\star = 9.1$, while at $z\sim 3.5$ the halo mass reaches $\log M_h = 11.18\pm 0.56$ for a stellar mass $\log M_\star = 9.3$.

From these measurements we find that $\log(M_\star/M_h)$ is ranging from -2.02 ± 0.2 for the low mass sample and up to -1.6 ± 0.2 for the most massive sample, at a redshift $z\sim 3$. As shown in Figure 6.2 these results are compared to various measurements at low and intermediate redshift $z<1$, obtained using different methods (Mandelbaum et al. 2006, Conroy et al. 2007, Foucaud et al. 2010, More et al. 2011, Hartley et al. 2013, Moster et al. 2013). Our measurements are in excellent agreement with the model derived from abundance matching at a redshift $z=3$ (Moster et al. 2013).

Our SHMR measurements are the first to be performed at z$\sim$3 from a clustering and HOD analysis, as made possible from the large VUDS spectroscopic redshift survey. We find that the SHMR is 1% to 2.5% for galaxies with intermediate stellar masses (at z$\sim$3) ranging from $\sim 10^9$ M$_\odot$ to $\sim 7\times 10^9$ M$_\odot$.

Following Conroy & Wechsler (2009) we compute the integrated star formation efficiency (ISFE) η=M$_\star$/M$_h$/f_b with f_b the universal baryon fraction $f_b=\Omega_b/\Omega_m=0.155$ (Planck Collaboration et al. 2014). Results are reported in Table 6.1. We find that the ISFE changes from 6.2% to 16.2% for galaxies with a stellar mass from M$_\star$=1.3×10^9 $M_\odot$ to M$_\star$=7.4×10^9 $M_\odot$. The IFSE therefore increases with stellar mass.

The star formation efficiency of $\sim$16% in a halo with M$_h$=3×10^{11} M$_\odot$ is quite close to a maximum of $\sim$20% occurring at 10^{12} M$_\odot$ in SMHM models Behroozi et al. (2010). The star formation efficiency of $\sim$16% in a halo with M$_h$=3×10^{11} M$_\odot$ is quite close to a maximum of $\sim$20% occurring at 10^{12} M$_\odot$ in SMHM models Behroozi et al. (2010) (see also Moster et al. 2013). We use a simple mass growth model to derive the time scale

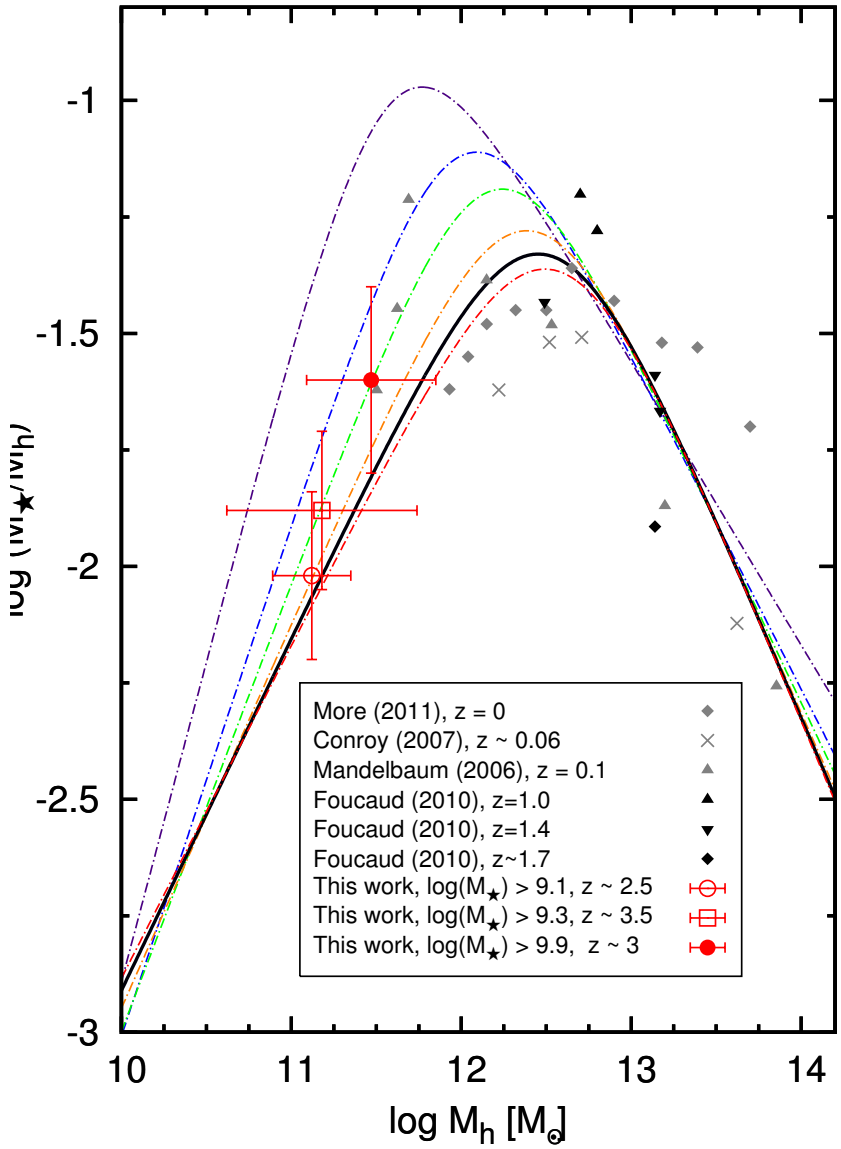


Figure 6.3: The halo mass M_h over stellar mass $M_\star$ ratio vs. halo mass at $z = 3$ in the VUDS survey. The colour scheme is the same as for the Figure 6.2.

for which our most massive galaxy sample would reach the maximum predicted in the SMHM relation from Behroozi et al. (2010). In this model the mass growth of DM haloes is described by a mean accretion rate $\langle \dot{M} \rangle_{mean}$ as proposed by Fakhouri et al. (2010), while galaxies grow in $M_\star$ via star formation using the median SFR for our sample from Tasca et al. (2014a), as well as through mergers with a constant accretion in stars of $\sim 1 M_\odot/yr$ (Tasca et al. 2014b). Because the halo accretion rate and SFR are dependent on redshift and halo/stellar mass, their values are computed every $\delta t = 5$ Myr to account for the changing mass and redshift.

We find that the SMHM relation would reach a maximum $\log(M_\star/M_h) \simeq -1.25$ about 350Myr after the observed epoch (i.e. at z$\sim$2.6) and that at this time halo and stellar masses will be $M_h = 10^{11.6} M_\odot$ and $M_\star = 2\times10^{10} M_\odot$ respectively. According to the model proposed by Moster et al. (2013) the SMHM relation is expected to reverse its trend after reaching the maximum point, but the slope of the function remains the same (see Figure 6.1). Therefore, assuming that the dark matter haloes continue to grow with time, the stellar mass accretion must be stopped or the dark matter accretion would need to increase with an accretion rate $\langle \dot{M} \rangle$ 10 times larger. There are no indications (e.g. in the N-body simulations), which would support the second scenario. On the other hand the stellar mass $M_\star = 2 \times 10^{10}$ computed at the maximum of SMHM relation is comparable to the 'quenching mass' as discussed in Peng et al. (2010). This result leads to the conclusion that, on average, the massive galaxy population with $M_\star^{min} \sim 0.5 - 1 \times 10^{10}$ $M_\odot$ observed at $z \sim 3$ will experience star formation quenching within a few hundred million years.

The SHMR is a simple yet efficient tool to probe star formation efficiency at the epoch of rapid stellar mass assembly provided one obtains robust measurements on both $M_\star$ and M_h, as is now possible with VUDS, complementing more indirect estimates using e.g. abundance matching. A more extensive exploration of the efficiency of star formation over a larger range of halo masses is becoming possible with new surveys coming on-line, and it would be interesting to probe higher masses than done in this paper to evaluate the halo mass for which the star formation efficiency is the highest. Extending these measurement further in redshift from clustering analysis will require the power of new facilities like Sumire-PFS, JWST or ELTs.

6.3 Summary

In order to measure efficiency of the star forming and stellar mass assembly in galaxies, we used the ratio between dark matter halo masses M_h and stellar masses $M_\star$ (SHMR). In our study we used $\sim$ 3000 spectroscopically confirmed galaxies from VIMOS Ultra Deep Survey covering redhift range $2 < z < 5$. Our estimation of SHMR is the first

performed at such high redshift. We measured dark matter host halo masses by using HOD modelling of the correlation function, and multi-wavelength SED fitting of the data to derive stellar masses. We find that the SMHR ranges from 1% to 2.5% for galaxies with $M_\star$=1.3×10^9 $M_\odot$ to $M_\star$=7.4×10^9 $M_\odot$ in dark matter haloes with M_h=1.3×10^{11} $M_\odot$ to M_h=3×10^{11} $M_\odot$. Additionally, we compute the integrated star formation efficiency (ISFE) η=$M_\star/M_h/f_b$ with f_b the universal baryon fraction. We find that the star formation efficiency is moderate $6-9\%$ for the lower mass galaxies, while it is as high as 16% for galaxies with median stellar mass of the sample $5-7 \times 10^9$ $M_\odot$. Low star forming efficiency indicate that star formation is suppressed by some supernovae or AGN feedback. On the other side the observed high ISFE for the average galaxy population at $z \sim 3$ indicates that typical galaxies are efficiently building-up their stellar masses at the key epoch in the mass assembly process.

CHAPTER 7

Conclusion

During my thesis work I studied the properties and evolution of the large scale structure of the Universe at high redshift $z > 2$ corresponding to a cosmic time when the age of the Universe was less than 3 Gyr. This unprecedented study was made possible thanks to the depth and reliability of the spectroscopic galaxy sample from the VIMOS Ultra Deep Survey (see Chapter 2 and also Le Fevre et al. 2014 in the Appendix A). In my work I was trying to bring us closer to the answers of some fundamental questions of modern cosmology. How did the large scale structure evolved? How did the relation between luminous and dark matter change through cosmic time and how this relation depends on the galaxy properties? What processes drive and regulate star formation and mass growth in galaxies along cosmic time?

I used galaxy clustering to quantify the physical properties of large scale structures at $2 < z < 5$. I measured the two-point correlation function - a commonly used statistical tool to describe how galaxies are clustered as a function of scale. I interpreted these measurements using two kinds of approximations. The first one is based on a classical description of the correlation function using a power-law function $\xi(r) = (r/r_0)^{-\gamma}$ with two free parameters r_0 - the correlation length, which describes how strongly galaxies are clustered and the slope γ (see Section 3.1 and also Davis & Peebles 1983). The second, more detailed approximation is based on a halo occupation model (HOD, see Section 3.2 and also Zehavi et al. 2004), which I adapted to the data available at high redshifts.

In order to perform these measurements, I needed to prepare the necessary tools. Therefore, at the first stage of my work I developed the code to model the HOD (see App. B), which then I put into extensive tests in order to check its reliability, limitations and

correct possible errors (see Sec. 3.5). Next, I developed the method to compute composite correlation function from the data coming from different fields (see Sec. 3.4.1), since the VUDS observations I used have been performed in two individual areas (COSMOS and VVDS-02h). To estimate uncertainties in the correlation function measurements performed in this study and to test the robustness of my clustering measurements I made use of a large number of mock galaxy samples, which were designed to mirror the properties of the VUDS sample in the range $2 < z < 5$ in the possibly most realistic way (see Sec. 3.3 for details). After a set of tests on mock catalogues I came out with the correction scheme which minimizes the effects related to the survey observational biases, hence maximized the robustness of the correlation function estimation (see Sec. 3.4.2). Assured that all my tools and methods were robust I performed the measurements on the VUDS data sample.

First, I took a closer look on the shape, properties and evolution of the Large Scale Structure at early stages of galaxy formation, by measuring galaxy clustering of the general galaxy population at $2 < z < 5$ and comparing the results with similar measurements obtained for the local and mid-redshift Universe (see Chapter 4). I found my results to be in a general agreement with the most recent and least biased measurements available in the literature at similar redshifts (see Sec. 4.5). I noted that my measurements are among a few (well, maybe one or two other worldwide) able to report measurements of the correlation function from accurate spectroscopic redshifts, at these high redshifts. I observed that the clustering strength r_0 of the star-forming galaxies within the VUDS survey remains almost stable or even midly increases from $z \sim 2.5$ to $z \sim 3.5$. This observed trend, however surprising at the first glance, can be easily interpreted in the framework of the theoretical predictions coupled with cosmological simulations. It can be explained by the combination of three main effects 1) the actual growth of large scale structure, made mainly of dark matter, with cosmic time, 2) the fact that galaxies are biased tracers of the dark matter field and in the hierarchical scenario of the structure formation the brightests galaxies are more probable to form first in the strongest density fluctuations, and 3) the fact that in the magnitude-limited survey like VUDS we observe intrinsically more luminous (hence, likely belonging to more massive dark mater halos) galaxies at higher redshift (see discussion in Sec. 4.4.1). Next, in the framework of the halo occupation distribution (HOD) formalism I estimated an average host halo mass at redshift z $\sim$ 3 to be $\langle M_h \rangle = 10^{11.75} h^{-1} \mathrm{M}_{\odot}$. This value is significantly lower than the average halo masses observed at low redshift, which is in agreement with the hierarchical theory of mass growth - the structure grows in mass and size through time, and the largest structures start to grow the first. I checked if it is possible to trace the history of the host dark matter haloes, and through that find into which present galaxy population the

VUDS sample would evolve. Using the halo mass growth model proposed by van den Bosch (2002) I found that the average halo observed at redshift $z \sim 3$ should evolve into halos with a mass $\langle M_h \rangle (z = 0) = 10^{13.9} h^{-1} \mathrm{M}_{\odot}$ at $z = 0$. As halos with such a mass are occupied by star forming galaxies with $M_r < -21.5$ in the local universe, I infer that the star-forming galaxy population observed at a redshift z ~ 3 might have evolved into the massive and bright galaxy population observed today (see Sec. 4.6 for discussion). However, it is likely that more complex models are needed to understand the detailed evolution of these galaxies and their host halos. Using my results, I studied also the evolution of the luminous-dark matter relation through the measurements of the large-scale galaxy bias (see Sec. 4.7). I observed that the bias values increase with redshift and all are much higher than the bias values observed at low and intermediate redshift ranges. This behaviour in general is in a perfect agreement with the predictions of our current paradigm of structure formation and evolution based on the hierarchical model of structure formation. The first galaxies are expect to form in the highest peaks of the mass distribution, hence these galaxies are the most biased ones. Then, as the mass density field evolves with time, galaxy formation systematically moves into less dense, hence less biased, regions. The detailed tracing of the changes of galaxy bias from high to low redshift, however, appeared to be a difficult task. I found that simple models are not able to reproduce the evolution of large-scale galaxy bias and that more complex models, like those involving a combination of passive halo evolution maintaining halo parameters constant and a star-forming evolution, would be needed to fully describe the bias evolution (see Sec. 4.7 for discussion).

I then I extended my clustering measurements to luminosity and stellar mass-selected sub-samples in order to study the dependence of galaxy clustering on these two physical parameters (see Chapter 5 for results). As already observed at low redshifts, I found that also at $z > 2$ the galaxy clustering strongly depends on galaxy properties. I observed that more luminous and more massive galaxies are more strongly clustered than their less massive and less luminous counterparts. This is telling us that the processes differentiating the clustering for different luminosities and masses have already been effective by z$\sim$3. Additionally, this trend seems to be stronger at high redshift than at lower z, but this conclusion is uncertain, since all the sub-samples used in my work have a mean luminosity and stellar mass above the characteristic (at given redshift) values of $L_{\star}$ and $M_{\star}$.

Finally, in the last part of my thesis work, I studied the efficiency of star formation and stellar mass assembly in galaxies at redshift $z \sim 3$ (see Chapter 6). I computed the stellar-to-halo mass ratio (SHMR) and the integrated star formation efficiency (ISFE), by taking the advantage of the HOD modelling of clustering measurements and the SED fitting over the large set of multi-wavelength VUDS data. These kind of measurements are

the first to be performed at redshift $z \sim 3$ and I found them to be in excellent agreement with e.g. the SHMR models proposed by Moster et al. (2013). According to my results, the massive galaxies in the VUDS survey have a star formation efficiency close to the maximum possible efficiency, at $\sim 16\%$, in comparison to the maximum $\sim 20\%$ efficiency occurring in SHMR models. On the other hand, I measure that the low mass galaxies have a star formation efficiency at a level $6 - 9\%$, which suggests that the star formation is suppressed by some quenching mechanisms, like produced by feedback effects.

Most of my measurements are among the first ones ever performed with a high accuracy at redshift $z > 2$. As such, they provide a valuable benchmark for the interpretation of the co-evolution of galaxies and Large Scale Structure at early epochs of galaxy formation (from the times when the Universe was only 1.5Gyr old) and put constrains on the efficiency of the processes which drive the star formation and mass assembly in galaxies at that time. All of this information can be used among others as an input to improve galaxy formation models (like semi-analytical models) and simulations (like the latest hydro-dynamical simulations), which are still uncertain at high redshifts and need to be informed by improved observational constraints. My measurements and results can also be considered as a precursor of next generation studies coming with the advent of the new powerful facilities like e.g. the James Webb Space Telescope (JWST) and Extremely Large Telescopes (ELTs).

At the very beginning of this thesis I quoted some unanswered, major cosmological questions. Of course my work cannot possibly give the final answer for these questions. I hope, however, that this study puts some new perspective and slightly improves our understanding of the Universe as it was at the early stages of its evolution. I like to think that this thesis can serve as the basis for new follow-up studies, taking us even closer to definitive answers.

Bibliography

Abbas, U., de la Torre, S., Le Fèvre, O., et al. 2010, MNRAS, 406, 1306

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

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

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

Behroozi, P. S. & Silk, J. 2014, ArXiv e-prints

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

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Phys. Rep., 367, 1

Bertone, S., Stoehr, F., & White, S. D. M. 2005, MNRAS, 359, 1201

Béthermin, M., Wang, L., Doré, O., et al. 2013, A&A, 557, A66

Bielby, R., Hill, M. D., Shanks, T., et al. 2013, MNRAS, 430, 425

Blake, C., Collister, A., & Lahav, O. 2008, MNRAS, 385, 1257

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

Blanton, M., Cen, R., Ostriker, J. P., & Strauss, M. A. 1999, ApJ, 522, 590

Blanton, M., Cen, R., Ostriker, J. P., Strauss, M. A., & Tegmark, M. 2000, ApJ, 531, 1

Bottini, D., Garilli, B., Maccagni, D., et al. 2005, PASP, 117, 996

Bromm, V., Yoshida, N., Hernquist, L., & McKee, C. F. 2009, Nature, 459, 49

Brown, M. J. I., Zheng, Z., White, M., et al. 2008, ApJ, 682, 937

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

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

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

Cappé, O. & Moulines, E. 2007, ArXiv e-prints

Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, ApJ, 750, L20

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

Coil, A. L., Newman, J. A., Cooper, M. C., et al. 2006, ApJ, 644, 671

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

Connolly, A. J., Scranton, R., Johnston, D., et al. 2002, ApJ, 579, 42

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., Prada, F., Newman, J. A., et al. 2007, ApJ, 654, 153

Conroy, C. & Wechsler, R. H. 2009, ApJ, 696, 620

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

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

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

Daddi, E., Röttgering, H. J. A., Labbé, I., et al. 2003, ApJ, 588, 50

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

de la Torre, S., Guzzo, L., Peacock, J. A., et al. 2013, A&A, 557, A54

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

de la Torre, S., Le Fèvre, O., Porciani, C., et al. 2011a, MNRAS, 412, 825

de la Torre, S., Meneux, B., De Lucia, G., et al. 2011b, A&A, 525, A125

De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, MNRAS, 366, 499

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

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

Fakhouri, O., Ma, C.-P., & Boylan-Kolchin, M. 2010, MNRAS, 406, 2267

Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189

Faucher-Giguère, C.-A., Kereš, D., & Ma, C.-P. 2011, MNRAS, 417, 2982

Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1994, MNRAS, 267, 927

Foucaud, S., Conselice, C. J., Hartley, W. G., et al. 2010, MNRAS, 406, 147

Foucaud, S., McCracken, H. J., Le Fèvre, O., et al. 2003, A&A, 409, 835

Garilli, B., Fumana, M., Franzetti, P., et al. 2010, PASP, 122, 827

Geach, J. E., Sobral, D., Hickox, R. C., et al. 2012, MNRAS, 426, 679

Giocoli, C., Tormen, G., Sheth, R. K., & van den Bosch, F. C. 2010, MNRAS, 404, 502

Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, A&A, 566, A108

Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37

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

Hartley, W. G., Almaini, O., Cirasuolo, M., et al. 2010, MNRAS, 407, 1212

Hartley, W. G., Almaini, O., Mortlock, A., et al. 2013, MNRAS, 431, 3045

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841

Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55

Iovino, A., Cucciati, O., Scodeggio, M., et al. 2010, A&A, 509, A40

Kaiser, N. 1986, MNRAS, 219, 785

Kaiser, N. 1987, MNRAS, 227, 1

Kashikawa, N., Shimasaku, K., Matsuda, Y., et al. 2011, ApJ, 734, 119

Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, ApJ, 637, 631

Katz, N., Hernquist, L., & Weinberg, D. H. 1999, ApJ, 523, 463

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188

Kilbinger, M., Benabed, K., Cappe, O., et al. 2011, ArXiv e-prints

Kim, J.-W., Edge, A. C., Wake, D. A., et al. 2014, MNRAS, 438, 825

Kong, X., Daddi, E., Arimoto, N., et al. 2006, ApJ, 638, 72

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35

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

Le Fevre, O., Cassata, P., Cucciati, O., et al. 2013, ArXiv e-prints

Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14

Le Fèvre, O., Guzzo, L., Meneux, B., et al. 2005, A&A, 439, 877

Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1670–1681

Le Fevre, O., Tasca, L. A. M., Cassata, P., et al. 2014, ArXiv e-prints

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159

Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70

Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, ApJ, 756, 71

Magliocchetti, M. & Porciani, C. 2003, MNRAS, 346, 186

Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715

Marinoni, C., Le Fèvre, O., Meneux, B., et al. 2005, A&A, 442, 801

Marulli, F., Bolzonella, M., Branchini, E., et al. 2013, A&A, 557, A17

Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, MNRAS, 286, 115

McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202

Meneux, B., Guzzo, L., de la Torre, S., et al. 2009, A&A, 505, 463

Meneux, B., Guzzo, L., Garilli, B., et al. 2008, A&A, 478, 299

Meneux, B., Le Fèvre, O., Guzzo, L., et al. 2006, A&A, 452, 387

Mo, H. J. & White, S. D. M. 1996, MNRAS, 282, 347

Mo, H. J. & White, S. D. M. 2002, MNRAS, 336, 112

Monaco, P., Theuns, T., & Taffoni, G. 2002, MNRAS, 331, 587

More, S., van den Bosch, F. C., Cacciato, M., et al. 2011, MNRAS, 410, 210

Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121

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

Norberg, P., Baugh, C. M., Hawkins, E., et al. 2002, MNRAS, 332, 827

Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301

Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, ApJ, 611, 685

Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)

Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Percival, W. J., Reid, B. A., Eisenstein, D. J., et al. 2010, MNRAS, 401, 2148

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16

Pollo, A., Guzzo, L., Le Fèvre, O., et al. 2006, A&A, 451, 409

Pollo, A., Meneux, B., Guzzo, L., et al. 2005, A&A, 439, 887

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

Quadri, R. F., Williams, R. J., Lee, K.-S., et al. 2008, ApJ, 685, L1

Reddy, N. A. & Steidel, C. C. 2009, ApJ, 692, 778

Ross, A. J. & Brunner, R. J. 2009, MNRAS, 399, 878

Savoy, J., Sawicki, M., Thompson, D., & Sato, T. 2011, ApJ, 737, 92

Sawangwit, U., Shanks, T., Abdalla, F. B., et al. 2011, MNRAS, 416, 3033

Schneider, P. 2006, Extragalactic Astronomy and Cosmology

Sheth, R. K., Hui, L., Diaferio, A., & Scoccimarro, R. 2001, MNRAS, 325, 1288

Silk, J. 2003, MNRAS, 343, 249

Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, MNRAS, 399, 966

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311

Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, ApJ, 592, 728

Tasca, L. A. M., Le Fèvre, O., & et al. 2014a, ArXiv e-prints

Tasca, L. A. M., Le Fèvre, O., López-Sanjuan, C., et al. 2014b, A&A, 565, A10

Tegmark, M., Strauss, M. A., Blanton, M. R., et al. 2004, Phys. Rev. D, 69, 103501

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41

van den Bosch, F. C. 2002, MNRAS, 331, 98

Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, ApJ, 695, 1163

Wake, D. A., Sheth, R. K., Nichol, R. C., et al. 2008, MNRAS, 387, 1045

Weinberg, D. H., Davé, R., Katz, N., & Hernquist, L. 2004, ApJ, 601, 1

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

Wiklind, T., Dickinson, M., Ferguson, H. C., et al. 2008, ApJ, 676, 781

Wraith, D., Kilbinger, M., Benabed, K., et al. 2009, Phys. Rev. D, 80, 023507

Zehavi, I., Eisenstein, D. J., Nichol, R. C., et al. 2005a, ApJ, 621, 22

Zehavi, I., Weinberg, D. H., Zheng, Z., et al. 2004, ApJ, 608, 16

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005b, ApJ, 630, 1

Zheng, Z. 2004, ApJ, 610, 61

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791

Zheng, Z. & Weinberg, D. H. 2007, ApJ, 659, 1

APPENDIX A

Publications

In this appendix we present the scientific publications already published or submitted, by the end of the PhD. Papers presented here are summarized in chapter 2 (Paper 1), chapter 4 (Paper 2) and chapter 6 (Paper 3) of this thesis.

- Collaboration paper:
  - Paper 1: **The VIMOS Ultra-Deep Survey: $\sim 10,000$ galaxies with spectroscopic redshifts to study galaxy assembly at early epochs** $2 < z \simeq 6$, O. Le Fèvre, L.A.M Tasca, P. Cassata, B. Garilli et al. (2014) - in press.
- First author:
  - Paper 2: **The evolution of clustering length, large-scale bias and host halo mass at 2<z<5 in the VIMOS Ultra Deep Survey (VUDS)**, A. Durkalec, O. Le Fèvre, S. de la Torre, A. Pollo et al. (2014) - submitted.
  - Paper 3: **The stellar-to-halo mass ratio using galaxy clustering in VUDS: a high star formation efficiency at z≃3**, A. Durkalec, O. Le Fèvre, S. de la Torre, A. Pollo et al. (2014) - submitted

Astronomy & Astrophysics manuscript no. vuds˙survey
March 18, 2014
© ESO 2014

arXiv:1403.3938v1 [astro-ph.CO] 16 Mar 2014

The VIMOS Ultra-Deep Survey: ∼10 000 galaxies with spectroscopic redshifts to study galaxy assembly at early epochs $2 < z \simeq 6$ ⋆

O. Le Fèvre[1], L.A.M. Tasca[1], P. Cassata[1], B. Garilli[3], V. Le Brun[1], D. Maccagni[3], L. Pentericci[4], R. Thomas[1], E. Vanzella[2], G. Zamorani[2], E. Zucca[2], R. Amorin[4], S. Bardelli[2], P. Capak[12], L. Cassarà[3], M. Castellano[4], A. Cimatti[5], J.G. Cuby[1], O. Cucciati[5,2], S. de la Torre[1], A. Durkalec[1], A. Fontana[4], M. Giavalisco[13], A. Grazian[4], N. P. Hathi[1], O. Ilbert[1], B. C. Lemaux[1], C. Moreau[1], S. Paltani[9], B. Ribeiro[1], M. Salvato[14], D. Schaerer[10,8], M. Scodeggio[3], V. Sommariva[5,4], M. Talia[5], Y. Taniguchi[15], L. Tresse[1], D. Vergani[6,2], P.W. Wang[1], S. Charlot[7], T. Contini[8], S. Fotopoulou[9], C. López-Sanjuan[11], Y. Mellier[7], and N. Scoville[12]

[1] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France
[2] INAF–Osservatorio Astronomico di Bologna, via Ranzani,1, I-40127, Bologna, Italy
[3] INAF–IASF, via Bassini 15, I-20133, Milano, Italy
[4] INAF–Osservatorio Astronomico di Roma, via di Frascati 33, I-00040, Monte Porzio Catone, Italy
[5] University of Bologna, Department of Physics and Astronomy (DIFA), V.le Berti Pichat, 6/2 - 40127, Bologna
[6] INAF–IASF Bologna, via Gobetti 101, I–40129, Bologna, Italy
[7] Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, 75014 Paris, France
[8] Institut de Recherche en Astrophysique et Planétologie - IRAP, CNRS, Universit de Toulouse, UPS-OMP, 14, avenue E. Belin, F31400 Toulouse, France
[9] Department of Astronomy, University of Geneva ch. d'cogia 16, CH-1290 Versoix
[10] Geneva Observatory, University of Geneva, ch. des Maillettes 51, CH-1290 Versoix, Switzerland
[11] Centro de Estudios de Física del Cosmos de Aragón, Teruel, Spain
[12] Department of Astronomy, California Institute of Technology, 1200 E. California Blvd., MC 249–17, Pasadena, CA 91125, USA
[13] Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA
[14] Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, D-85741, Garching bei München, Germany
[15] Research Center for Space and Cosmic Evolution, Ehime University, Bunkyo-cho 2-5, Matsuyama 790-8577, Japan

e-mail: olivier.lefevre@lam.fr

Received...; accepted...

ABSTRACT

We present the VIMOS Ultra Deep Survey (VUDS), a spectroscopic redshift survey of ∼10 000 very faint galaxies to study the major phase of galaxy assembly $2 < z \simeq 6$. The survey covers 1 deg^2 in 3 separate fields: COSMOS, ECDFS and VVDS-02h, with targets selection based on an *inclusive* combination of photometric redshifts and color properties. Spectra covering $3650 < \lambda < 9350$Å are obtained with VIMOS on the ESO-VLT with integration times of 14h. Here we present the survey strategy, the target selection, the data processing, as well as the redshift measurement process, emphasizing the specific methods adapted to this high redshift range. The spectra quality and redshift reliability are discussed, and we derive a completeness in redshift measurement of 91%, or 74% for the most reliable measurements, down to $i_{AB} = 25$, and measurements are performed all the way down to $i_{AB} = 27$. The redshift distribution of the main sample peaks at $z = 3 - 4$ and extends over a large redshift range mainly in $2 < z < 6$. At $3 < z < 5$, the galaxies cover a large range of luminosities $-23 < M_{NUV} < -20.5$, stellar mass $10^9 M_\odot < M_* < 10^{11} M_\odot$, and star formation rates 1 $M_\odot$/yr$< SFR < 10^3 M_\odot$/yr. We discuss the spectral properties of galaxies using individual as well as stacked spectra. The comparison between spectroscopic and photometric redshifts as well as color selection demonstrate the effectiveness of our selection scheme. With ∼ 6000 galaxies with reliable spectroscopic redshifts in $2 < z < 6$ expected when complete, this survey is the largest at these redshifts and offers the opportunity for unprecedented studies of the star-forming galaxy population and its distribution in large scale structures during the major phase of galaxy assembly.

Key words. Galaxies: evolution – Galaxies: formation – Galaxies: high redshift – Cosmology: observations – Cosmology: large-scale structure of Universe – Astronomical Databases: surveys

1. Introduction

The study of the first billion years of galaxy evolution is one of the key frontiers of modern cosmology. The current theoretical paradigm rests on the hierarchical build-up of dark matter halos in a ΛCDM cosmology (see e.g. Mo, van den Bosch & White 2010). Galaxies formed in these deep potential wells, are expected to transform primordial gas into stars from the initial reservoir and to be fed from new accreted gas. As dark matter halos merge, galaxies in them are also expected to merge, these events deeply transforming the dynamics and overall star, gas, and dark matter content of the merging galaxies into the newly formed one. Along with these processes super-massive black

⋆ Based on data obtained with the European Southern Observatory Very Large Telescope, Paranal, Chile, under Large Program 185.A-0791.

holes are expected to form at the bottom of the potential wells. Complex processes are invoked to regulate the growth of galaxies, including supernovae or AGN feedback possibly quenching star formation, or the role of different environments impacting the way galaxies are nurtured. All these processes are combined in increasingly sophisticated numerical models producing galaxy simulations in representative volumes (e.g. Springel et al. 2008) coupled to semi-analytic description of galaxy evolution (e.g. Guo et al. 2011). These simulations face the challenge to reproduce both the internal physics in complex galaxy systems and the general volume-averaged properties of large galaxy populations as a function of cosmic time, and need better constraints from observations to be thoroughly tested.

Impressive progress has been made on the observational front over the past two decades in an attempt to test and detail a galaxy formation and evolution scenario from robust measurements. A key element driving observational progress is the need to cover all major phases of galaxy evolution from the early formation and galaxy assembly to today, a formidable endeavor. Deep galaxy surveys have florished to conduct this exploration. The latest few billion years have been extensively mapped by surveys like the 2dFGRS (Colless et al. 2001) and then the various stages of the SDSS (Abazajian et al. 2009), setting a firm observational reference for status of galaxies after more than 13 billion years of evolution. At larger redshifts deep surveys are providing a complex picture with strong evolutionary features like the build-up along cosmic time of stellar mass in galaxies of different types or the star formation history. Different types of imaging and spectroscopic surveys are playing a complementary role, the deepest studies being performed in photometry and augmented with photometric redshifts (e.g. Ilbert et al. 2006), and with spectroscopic surveys bringing accurate redshifts and spectro-photometry, spectral features properties, as well as internal velocity information. The contribution of space observations in combination with ground-based surveys has been key to provide morphological information (Rix et al. 2004, Koekemoer et al. 2007, Koekemoer et al. 2011) and access to photometric bands invisible from the ground either in the far UV or in the mid to far infrared. Despite this progress, the exploration of early phases of galaxy formation and evolution is still largely incomplete. We do not know which objects ignited first at the end of the dark ages, when and how the universe was reionised, when and how the first massive galaxies formed, and the importance of any physical connection between galaxy and black-hole formation and growth. Star-forming galaxies are providing key information to understand how galaxies grow with time, in particular enabling to measure fundamental quantities such as the cosmic star formation history (e.g. Lilly et al. 1996, Madau et al. 1996, Tresse et al. 2007, Bouwens et al. 2009, Cucciati et al. 2012, Madau & Dickinson 2014), and the history of stellar mass assembly (e.g. Arnouts et al. 2007, Ilbert et al. 2013).

At redshifts $z \sim 1$ the pioneering CFRS survey (Lilly et al. 1995, Le Fèvre et al. 1995) was followed by more extensive galaxy redshift surveys covering larger volumes like the DEEP2 (Davis et al. 2003), the VVDS (Le Fèvre et al. 2005a, Le Fèvre et al. 2013), zCOSMOS (Lilly et al., 2007), now reaching the 100 000 redshift mark at $z \sim 1$ with VIPERS (Guzzo et al. 2014). These surveys have brought a wealth of quantitative and accurate measurements now reaching large enough areas of a few tens of square degrees and volumes of $\sim 5 \times 10^7 h^{-3}\mathrm{Mpc}^3$ so that the most fundamental statistical quantities describing the galaxy population like the luminosity function (LF), the mass function (MF) or the correlation function (CF) are becoming very accurate and less affected by the cosmic variance related to the small fields of earlier studies. At higher redshifts ($z >\sim 2$), the rapid progress has been driven by the effectiveness in selecting high redshift galaxies and, most importantly, by the impressive gains in sensitivity and efficiency provided by high multiplex multi-slit spectrographs like LRIS (Oke et al. 1995) and DEIMOS (Faber et al. 2003) on the Keck telescope, FORS (Appenzeller et al. 1998) and VIMOS (Le Fèvre et al. 2003) on the VLT. The effectiveness of the Lyman-break galaxies (LBG) selection has provided the capability to find large numbers of galaxies at $z > 2.5$, and is continuing to be the single-most used technique to select galaxies at the highest possible redshifts (e.g. Steidel et al. 2003, Bouwens et al. 2009, Ellis et al. 2013). It is supplemented by narrow band imaging techniques isolating Lyman-α emitters (LAE; Taniguchi et al. 2005, Shimasaku et al. 2006; Ouchi et al. 2008), highly efficiency when reaching sufficiently deep to dig deeper into the LAE luminosity function at increasingly high redshifts. In addition to these pre-selection techniques, deep purely magnitude selected surveys were conducted in order to probe a large population mostly free of pre-selection biases. The largest to date probing redshifts $z > 1.5$ is the $i-band$ magnitude selected VVDS survey covering up to $z \sim 6.5$ with > 1000 galaxies with redshifts $z \geq 2$ (Le Fèvre et al. 2005a, 2013). Other magnitude selected surveys have attempted using redder bands to alleviate selecting only galaxies with strong rest-frame UV continuum. The K20 survey used K-band magnitude selection down to $K = 20$ to identify extremely red objects (EROs) at $z \sim 1.5 - 2$ which turned out to be dust obscured star forming galaxie or old passive early-type galaxies (Cimatti et al. 2002). In a following work the GMASS survey selected objects on the basis of Spitzer NIR photometry with $m_{4.5\mu m} < 23.0$ (AB) coupled to photometric redshift $z_{phot} > 1.4$ to identify a few hundred galaxies with $1.5 < z < 3$, including 13 red passive galaxies (Cimatti et al. 2008). However, and despite these attempts, the approach using pure magnitude selection is costly in observing time when going much beyond redshift $z \sim 2$ or so.

Performing a complete galaxy census is a basic *Astronomy* input necessary for any subsequent *astrophysical* analysis. While at redshifts ~ 1 this census is mostly complete down to stellar mass $10^8\mathrm{M}_\odot$, it is not yet the case at redshifts $z > 1$ for several reasons. First, the color selection schemes applied to photometric samples to extract the high redshift populations, while efficient to identify galaxies, are affected by significant incompleteness, loosing some fraction of the population at the selected redshift, or by contamination from galaxies at other redshifts. While the latter can be dealt with by obtaining spectroscopic redshifts, the former remains a serious difficulty especially at faint magnitude and at the highest redshifts. Unfortunately the level of incompleteness strongly depends on the photometric filters used for imaging, the depth of the observations, as well as the image quality, which requires a case by case study involving source simulations complicated by the need to make apriori hypotheses on the properties of galaxies one is trying to establish. It was realised that color-color selection like the LBG technique at $z > 2.5$ or like the BzK working at $z \simeq 2$ (Daddi et al. 2004) would miss a part of the general galaxy population in their selection process (Le Fèvre et al. 2005b), therefore making the galaxy census incomplete (Le Fèvre et al. 2014). The consequences of this maybe far-reaching, as incompleteness in counts leads to underestimates in luminosity density, star formation rates, as well as mass density, just to cite these few. An important aspect of ongoing and future studies is to revisit galaxy counts as a function of redshift making sure that no significant population is missing and that no significant bias is introduced in deriving astrophysical quantities.

A key element is then the availability of large samples of galaxies with a well defined and well controlled selection function. Spectroscopic redshift surveys play a key role as they provide samples with confirmed redshifts. Photometric redshift surveys are widely used and have now reached an impressive accuracy. However the level of 'catastrophic failures' when photometric redshifts disagree with their training set of spectroscopic redshifts, even if low at a few percent (Ilbert et al. 2013), could still produce large unknowns because of the shape of the N(z) of flux limited samples. An error of 1% at the peak, $z \simeq 1$, of the N(z) of a $i_{AB} = 25$ sample could spread galaxies with wrong photometric redshifts to higher redshifts, e.g. at $z \sim 3$ where the projected galaxy density is less than 10 times the N(z) at peak, which could then represent several tens of percent of uncertainty. One recent example is the difficulty to distinguish $z \sim 5$ very massive objects from lower redshifts $z \sim 2$ galaxies of lower mass (Wiklind et al. 2008, Caputi et al. 2012). Obtaining a spectroscopic redshift therefore remains a fundamental measurement.

Unfortunately, the total number of galaxies spectroscopically confirmed today by all these surveys at $z > 2$ is still limited. Published LBG samples reach ∼2000 redshifts at $z \sim 3$, only ~ 150 at $z \sim 4$, and a few tens of galaxies beyond that (e.g. Steidel et al., 2003, Vanzella et al., 2009, Bielby et al. 2013). Samples of LAE galaxies selected with narrow band techniques and confirmed in spectroscopy reach a few hundred objects beyond $z = 3$ (Ouchi et al. 2008, Kashikawa, 2011). The VVDS has assembled ∼35 000 galaxies with spectroscopic redshifts down to $i_{AB} = 24.75$, but the high redshift tail at $z > 2$ contains about 1000 galaxies (Le Fèvre et al. 2014). In front of the difficulty of obtaining large samples of spectroscopicaly confirmed galaxies, many surveys use samples defined solely on the basis of photometric color selection techniques like LBG or other simple color cuts, relying on completeness and contamination estimates difficult to control. To overcome the uncertainties linked to small existing spectroscopic samples, and to understand the biases and limitations of photometry-based studies in their ability to provide a complete census of the galaxy population, extremely deep spectroscopic surveys over large volumes are needed.

Here we present VUDS, the VIMOS Ultra Deep Survey of ∼10 000 galaxies specifically designed to study the early phases of galaxy evolution $2 < z < 6+$. The VUDS sample contains an unprecedented number of galaxies with secure spectroscopic redshifts at this epoch, obtained in three different fields: COSMOS, ECDFS and VVDS-02h. The survey design including the target selection is based mainly on photometric redshifts, as presented in Section 2. The VIMOS multi-slit spectroscopic observations, the data reduction and redshift measurement scheme are discussed in Section 3. Properties of the VUDS sample are presented in Section 4, including the redshift distribution of the sample, the distribution of intrinsic properties like stellar mass and star formation rate, and the average spectra properties based on high signal-to-noise stacked spectra. After comparing to other surveys in Section 5, we summarize our results in Section 6 and conclude on the usefulness of the VUDS to study the early phases of galaxy assembly with unprecedented accuracy.

All magnitudes are given in the AB system unless specified, and we use a Cosmology with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $h = 0.7$.

2. Survey design

2.1. Science drivers

When the VUDS survey observing time proposal was approved at the end of 2009, spectroscopic redshift surveys were still of limited scope at the epoch of galaxy assembly significantly beyond redshifts $z \sim 2 - 3$. The VUDS survey therefore aimed to address several core science goals at the epoch $2 < z < 6+$, including: (i) the early history of the global star formation rate, (ii) the build up of the mass function for different galaxy populations, and the contribution to mass assembly and star formation of merging, feedback and accretion, (iii) the study of very young galaxies in their early stages of assembly, including the earliest stellar populations like Pop-III. (iv) the identification of proto-structures and the effects of environment at early stages of galaxy evolution, (v) the study of the clustering of galaxies to infer the mass growth of underlying dark matter halos,

These broad science goals still remain hot science topics today and are the main focus of VUDS.

2.2. Survey strategy overview

The VUDS survey is designed to study the galaxy population beyond redshift $z \simeq 2$ and up to the highest redshifts possible in the wavelength range up to $\lambda = 9350$Å accessible with the VIMOS spectrograph on the VLT (see below), corresponding to Lyα at a redshift up to $z \simeq 6.6$. A total of one square degree is observed in three well separated fields, the COSMOS field, the extended Chandra Deep Field South (ECDFS) and the VVDS-02h field, in order to mitigate cosmic variance effects. This program was awarded Large Program status with 640h of VLT-VIMOS observing time.

A key feature of the survey is the target selection. It is primarily based on photometric redshifts and their probability distribution function (PDF), complemented by color-color selections and analysis of the SED shape when not already in the photo-z selection. When the geometry of the observed slit-masks allows it after placing the previous priority targets, a purely magnitude-selected random sample of galaxies with $23 \leq i_{AB} \leq 25$ has been added to the target list. This is further described in Section 2.5.

The second key feature is the extended wavelength coverage of the VIMOS spectroscopic observations covering $3650 \leq \lambda \leq 9350$Å, which minimizes any instrument-induced redshift desert and strongly reduces the degeneracies in redshift measurements (Le Fèvre et al. 2014).

The last key point is the integration time of $\simeq$ 14h per target, which allows reaching a S/N on the continuum at 8500Å of $S/N = 5$ for $i_{AB} = 25$, and $S/N = 5$ for an emission line with a flux $F = 1.5 \times 10^{-18}$ erg/s/cm^2/Å.

2.3. Survey fields

VUDS covers three of the most observed extragalactic fields: the COSMOS field, the ECDFS and the VVDS-02h (also known as the CFHTLS-D1/XMM-LSS field). A total of 0.5 square degrees are observed in the COSMOS field, 0.31 square degrees in the VVDS-02h field, and 0.14 square degrees in the ECDFS. The fields location and covered area for each are summarized in Table 1, and identified in Figures 1 to 3.

2.4. Multi-wavelength data and photometric catalogues

By design the three survey fields cumulate a wealth of deep multi-wavelength data and spectroscopy. This is an important part of VUDS as multi-wavelength coverage of the spectral energy distribution (SED) is essential in combination with accurate spectroscopic redshifts, in particular to derive key physical quantities like absolute magnitudes, stellar masses, star formation rates or ages. We describe for each field the most relevant existing data.

The COSMOS field (Scoville et al. 2007) is extensively observed at all wavelengths. Its location was setup close to one of the original VVDS fields, recentered to a region with less galactic extinction when better extinction maps became available. The field was observed with HST-ACS in the F814W filter (Koekemoer et al. 2007) providing high spatial resolution imaging down to a depth $AB(F814W) = 27.2$ (5σ). Extensive imaging observations were obtained in g, r, i, and z bands from the Subaru SuprimeCam (Taniguchi et al. 2007), as well as from the CFHT Megacam from the CFHT-Legacy Survey including the u-band, and photometry is available from 12 medium bands (references available in Ilbert et al. 2009). The UltraVista survey is acquiring very deep near-infrared imaging in the Y, J, H and K bands using the VIRCAM camera on the VISTA telescope (McCracken et al. 2012). The UltraVista DR2 release reaches a depth $K_{AB} = 24.8$ at 5σ (in 2 arcsec apertures), and it is planned to increase the depth down to $K_{AB} = 25.5$ in half the area by the end of the survey. The VUDS VIMOS pointings are placed in a way to optimize the overlap with the deeper UltraVista strips (Figure 1). Following on the initial Spitzer observations (Sanders et al. 2007), deeper Spitzer-warm observations are obtained by the Splash program, reaching $AB = 25.4$ (5σ at 3.6μm) (Capak et al. in prep.). The COSMOS field has deep Herschel imaging reaching 8 mJy at 250 μm (Oliver et al. 2012). The CANDELS survey (Grogin et al. 2011) in this field includes WFC3 imaging in the F125W (J) and F160W (H) filters down to $H_{AB} = 27.1$ (5σ); the CANDELS area is almost entirely covered by the VUDS pointings (Figure 1). Extensive spectroscopy was obtained in the COSMOS field prior to VUDS. The zCOSMOS survey (Lilly et al. 2007) contains 20 000 galaxies with $0 < z < 1.2$ selected with $i_{AB} \leq 22.5$ and ∼ 6000 galaxies selected with $g_{AB} \simeq 25 - 25.5$ with $1.4 < z < 3$ and a median redshift $z = 2.15$ (Lilly et al. in preparation). Deep spectroscopy from Keck has identified several hundred galaxies in the same redshift range as VUDS (e.g. Capak et al. 2011). VUDS targeting avoided to reobserve those galaxies with existing reliable redshifts, when known. A wealth of observations at other wavelengths are available in this field, we refer the reader to the COSMOS web site (`http://cosmos.astro.caltech.edu/`) for the detailed list and properties.

The ECDFS field is the focus of a number of deep multi-wavelength surveys. Built around the field with deep 1Ms Chandra observations (Giacconi et al. 2002), the field was extended and is covered with deep UBVRI imaging down to $R_{AB} = 25.3$ (5σ, Cardamone et al. 2010 and references therein). Several spectroscopic surveys have been conducted (Le Fèvre et al. 2004, Cardamone et al. 2010, Cooper et al. 2012), with more than 5000 galaxies with redshifts $0 < z < 2$, but the number of spectroscopically confirmed galaxies at $z > 2$ remains small. The central part of this field cumulates a number of deep HST imaging with the HDF-South (Williams et al. 2000), GEMS (Rix et al. 2004), and ERS (Windhorst et al. 2011) surveys, and more recently with the CANDELS survey (Grogin et al. 2012) adding WFC3 near-IR imaging reaching as deep as $H_{AB} = 27.3 - 27.6$. The SERVS Spitzer-warm obtained 3.6 and 4.5 μm imaging data down to $AB = 23.1$ (Mauduit et al. 2012).

The VVDS-02h field was orginally defined as the 'deep field' of the VVDS survey. It was observed in BVRI at CFHT (Le Fèvre et al. 2004), followed by deeper u', g, r, i observations as the D1 deep field in the CFHTLS survey reaching $i_{AB} = 25.44$ at 50% completeness in the latest DR7 (Cuillandre et al. 2012). Deep infrared imaging has been obtained with the WIRCAM at CFHT in YJHK bands down to $Ks_{AB} = 24.8$ also at 50% completness (Bielby et al. 2012). Extensive multi-slit spectroscopy has been obtained with VIMOS as the 'Deep' and 'Ultra-Deep' surveys of the VVDS, with magnitude-selected samples down to $I_{AB} = 24$ and $i_{AB} = 24.75$ respectively. A total of 11 139 objects have spectroscopic redshift measurements from the VVDS, including 774 with $z > 2$. This field was observed in all Spitzer bands as part of the SWIRE survey (Lonsdale et al. 2003), reaching a magnitude in the 3.6 microns band of $AB = 21.5$. More recently the SERVS survey obtained deeper data with Spitzer in the 3.6 μm and 4.5 μm bands down to $AB = 23.1$ at 3.6 μm (Mauduit et al. 2012). It is one of the fields of the HERMES survey (Oliver et al. 2012), matched to CFHTLS and VVDS data (Lemaux et al. 2014). A number of other deep data are available including X-ray (Pierre et al. 2004) and radio (Bondi et al. 2003) observations.

In each field photometric catalogues including at least from u to 4.5 μm bands have been assembled, matching different imaging sources, and extracting photometry in 2 arcsecond apertures in dual mode using SExtractor (Bertin & Arnouts 1996).

2.5. Spectroscopic target selection

As shown from the VVDS, the redshift distribution of faint galaxies, e.g. down to $i_{AB} \simeq 25$ is peaked at $z \sim 1.3 - 1.5$, with a rapidly decreasing high redshift tail (Le Fèvre et al. 2014). Magnitude selection is then inefficient in picking-up $z > 2$ galaxies among the far more numerous foreground. This prompted the use of the Lyman-break selection technique, as pioneered by Steidel and collaborators (Steidel et al., 1996). This technique makes simple use of the color properties of galaxies as a function of redshift, identifying a locus in color-color space associated to specific redshift ranges.

The development of the photometric redshift technique and its demonstrated success in terms of the redshift measurement accuracy with typical errors of less 5% on $\Delta z/(1+z)$ and a low rate of catastrophic errors when compared to spectroscopic redshifts (e.g. Ilbert et al. 2010, Cardamone et al. 2010) has opened new perspectives in pre-selecting galaxies. In essence, photometric redshift techniques are a generalisation of color selection techniques such as LBG or BzK to the complete spectral energy distribution (SED) which must verify a color distribution accross wavelengths compatible with reference galaxy templates. We elected to use the best possible photometric redshifts as our main selection criterion. Photometric redshifts have been computed with the code Le Phare (Arnouts et al. 1999, Ilbert et al. 2006) using the multi-band photometric catalogues covering from the u-band to the Spitzer 4.5μm band. The techniques are described in Ilbert et al. (2009), and more recently in Ilbert et al. (2013). The observed photometric data are correlated against 33 templates covering a range from early to late type galaxies of different ages, star formation histories and metallicities, leaving the E(B-V) extinction as a free parameter. Emission lines with a flux empirically computed from the UV continuum flux are added to the templates. The 'best' redshift is assigned from

the median of the marginalized redshift probability distribution function (PDF).

To select the spectroscopic targets in VUDS we followed an *inclusive* rather than exclusive strategy, adding samples pre-selected from several different criteria, as described below:

- The spectroscopic targets must verify $z_{phot} + 1\sigma \geq 2.4$, and $i_{AB} \leq 25$. As degeneracies are known to occur, we selected sources for which either the first or the second peak in the photometric redshift probability distribution function satisfy this criterion.
- Sources which were not selected from the primary z_{phot} criterion but satisfy one of the ugr, gri or riz LBG color-color selection criteria have been added to the target list.
- The full SED over all filters available in ugrizYJHK is used to identify galaxies with a break in the continuum and not identified by any of the z_{phot} or LBG criteria, extending the *dropout* technique. These targest are allowed to have $K_{AB} \leq 24$ and hence may not necessarily be brighter than $i_{AB} = 25$.
- Finally, when space on the slit mask was still available after the above selections, a random sample of galaxies with $23 \leq i_{AB} \leq 25$ has been selected.

For the last half of the survey observations, priority for slit placement was given to the targets with $z_{estimate} \geq 4$ to increase the number of objects at these redshifts in the final sample.

In adding these different selection criteria we aim at maximizing the pre-selection of objects above $z = 2.4$. Given the dispersion of photometric redshift errors, we expect the redshift distribution of VUDS sources to start rising at $z \simeq 2$. While the selection strategy described above could lead to an increase in the contamination from objects at lower redshifts, this contamination remains quite small as described below. For each of the targeted samples, slits are placed at random using the `vmmps` slit mask design software (Bottini et al. 2005), providing a fair and representative sample of the general population.

These various selection criteria are summarized in Table 2. A large fraction ∼ 75% of the sample has been selected on the basis of the first peak of photometric redshifts, ∼ 15% on the second peak of photometric redshifts, and the rest ∼ 10% about equally split between the LBG color-color selection and purely magnitude selected samples.

3. Observations, data processing, and redshift measurement

3.1. VIMOS on the VLT

The VIsible Multi-Object Spectrograph (VIMOS) is installed on the European Southern Observatory Very Large Telescope unit 3 Melipal. VIMOS is a wide field imaging multi-slit spectrograph (Le Fèvre et al. 2003), offering broad band imaging capabilities, as well as multi-slit spectroscopy. VIMOS is a high performance Multi-Object Spectrograph (MOS) with 4 parallel channels, each a complete imaging-spectrograph with a field of view 8×7 arcmin^2, or a total field of 224 arcmin^2. The key features of VIMOS are a high multiplex (number of slits) and the excellent sky subtraction accuracy reaching $\sigma_{sky\ residual} \simeq 0.1\%$ of the sky signal (Le Fèvre et al. 2013).

For VUDS we use the low resolution multi-slit mode of VIMOS; with the 4 channels, this offers the largest multiplex for multi-slit spectroscopy. Following Scodeggio et al. (2009) we have optimized the slit length allowing lengths as small as 6 arcseconds, maximizing the number of observed slits.

3.2. VIMOS observations

We observe a total of 16 VIMOS pointings, at the coordinates of the pointings identified in Table 3. Fifteen pointings are observed with both the LRBLUE grism covering $3650 \leq \lambda \leq 6800$Å and the LRRED grism covering $5500 \leq \lambda \leq 9350$Å, leading to a full wavelength coverage of $3650 \leq \lambda \leq 9350$Å. With slits one arcsecond wide, these grisms provide a spectral resolution $R = 230$ quite uniform over the wavelength range. At this resolution, each of the four 2048×4096 pixels detectors can accommodate 3–4 full length spectra along the dispersion direction, and given the projected space density of VUDS targets we therefore observe on average $N_{slits} \simeq 600$ individual slits simultaneously. Each slit may contain not only the VUDS pre-selected target but also serendipitous objects falling in the slit by chance.

One pointing on the ECDFS (#3) has been observed with the MR grating. This setup covers $5000 \leq \lambda \leq 9500$Å with a spectral resolution $R = 580$. With this resolution about 2 full length spectra can be placed along the dispersion direction, and a total of ∼ 220 objects have been observed in this pointing. The targets for this pointing have been optimized towards the highest redshifts $z > 4.5$; the improved resolution compared to the LR grisms making it somewhat easier, in principle, to identify emission lines like Lyα between the OH sky emission features (but possibly at the expense of sensitivity on continuum measurements).

Most VUDS observations were obtained after the CCD detectors were upgraded in summer 2010. At that time, the original blue–sensitive thinned E2V CCDs were changed to red–optimized thick E2V CCDs in 2010 (Hammersley et al. 2010). The global VIMOS sensitivity at 9000Å increased by ∼ ×2, making it comparable to the FORS2 sensitivity in the red with a field of view 4.8× larger, and significantly reducing the fringing above 8000Åthanks to the thicker substrate.

To reach a total integration of 14h, each of the LRBLUE or LRRED grism observations consist on average of 13 observing blocks (OBs) executed at the telescope. Each OB includes three spectroscopic exposures of 1250 to 1350 seconds obtained dithering $-0.75, 0, +0.75$ arcseconds along the slit. The OBs specify the observing conditions that must be met, including a seeing better than 1 arcsecond, sky transparency set to 'clear', airmass less than 1.5, dark time with constraints on the moon phase and distance from the field (lunar illumination 0.3 to 0.5, and distance to the observed field of more than 60-90 degrees). Arc lamp and flat field calibrations are obtained after each set of 3 OBs, and flux calibration on standard stars are performed in the standard ESO procedure. When ready, OBs are sent to the VLT-VIMOS service observing queue via the P2PP tool and get executed when the atmospheric and moon conditions are met.

3.3. Spectroscopic data processing, redshift measurement, and reliability flags

The spectroscopic data processing followed the same general principles as defined for the VVDS (Le Fèvre et al. 2005a, 2013). We summarize this process below and emphasize the specific data processing steps that we follow for VUDS.

The general outline of the VUDS data processing is now a standard for multi-slit spectroscopy. We use the VIPGI environment to process the spectra (Scodeggio et al. 2005). First the 40 individual 2D spectrograms coming from the 13 OBs for one of the LRBLUE or LRRED observations are extracted finding the location of the slit projection on the detector using the expecta-

tion from the slit mask design. Sky subtraction is performed with a low order spline fit along the slit for each wavelength sampled. The sky subtracted 2D spectrograms are combined with sigma clipping to produce a single stacked 2D spectrogram calibrated in wavelength and flux. The 2D spectrogram is collapsed along the dispersion direction to produce a slit profile in which objects are identified. The spectral trace of the target and other detected objects in a given slit are linked to the astrometric frame to identify the corresponding target in the parent photometric catalogue. This processing is performed separately for each of the LRBLUE and LRRED observations. The 2D and 1D fully calibrated spectra are then cross matched per observed slit, and joined to form spectra with full $3650 \leq \lambda \leq 9350$Å wavelength coverage. It is to be noted that, in the wavelength range of overlap $5500 \leq \lambda \leq 6800$Å, the end spectra are the average of the LRBLUE and LRRED, cumulating about 28h of observation. At the end of this process 1D sky-corrected stacked spectra are extracted. They are fully calibrated both in wavelength and in flux using spectrophotometric standard stars. In this process additional objects can be detected in the slit and have subsequently their 1D spectra extracted. These objects are called "secondary detections", and a photometric counterpart is searched for in the photometric catalogue. If an object is found within less than one arcsecond from the spectral trace, the secondary spectrum identification is assigned the identifier of the object in the photometric catalogue. If no object is found in the photometric catalogue, the object is extracted and a new entry is produced in the catalogue with coordinates at the location corresponding to the trace of the object. It might happen that a single emission line is identified upon visual examination of the 2D spectrogram of a given slit, but not extracted by the automated procedure: these objects are flagged and then manually extracted. This procedure is particularly important for objects with single emission lines and no or very little detected continuum, which often turn out to be Lyα emitters at high redshifts (e.g. Cassata et al. 2011).

Upon comparison between the photometric magnitudes and the magnitudes derived from the calibrated spectra, we realized that the $u-$band part of the spectra (and to a lesser extent the $g-$band) was lacking photons at the $\sim 40\%$ ($\sim 15\%$) level. We proceeded to add three well-defined corrections to the spectra: (1) atmospheric absorption, (2) atmospheric refraction, and (3) Galactic extinction. The atmosphere absorbs photons depending on the airmass along the light path; this is corrected using the prescription defined for the Paranal observatory in Patat et al. (2011). In addition atmospheric refraction acts as a small prism before entering the telescope, spreading the incoming light into a spectrum with length depending on the airmass and paralactic angle, the angle of the slit to the zenithal angle. With slit-masks placed on sky objects using an $r-$band filter prior to spectroscopic observations, and slits one arcsecond wide, this may introduce a significant loss of uv-blue photons falling out of the slit. The Galactic extinction on distant sources also produces a chromatic correction, which has been applied using the E(B-V) maps of Schlegel et al. (1998). In this process we also proceed to produce spectra with continuum flux which are calibrated on the $i-$band photometric flux, therefore correcting for any slit losses occuring in that band if the object extension is larger than the slit width. Adding these corrections we are able to correct the uv-blue spectroscopic flux in such a way that there is excellent agreement between the spectroscopic flux and the photometric flux measurements at better than the 5% level in all wavelengths $3650 \leq \lambda \leq 9350$Å. This is of particular importance when fitting the spectra to reference templates in order to derive internal galaxy properties (Thomas et al., in prep.). One of the last observations obtained on this program includes a number of repeated observations on a sub-sample of VUDS galaxies. When processed it will allow to estimate redshift (velocity) measurements accuracy, as well as to have an independent check of the reliability of each of the redshift reliability flags (see below). Based on the previous VVDS, zCOSMOS and VIPERS surveys with the same VIMOS instrument, we expect the redshift accuracy to be in the range $dz/(1+z) = 0.0005 - 0.0007$, or an absolute velocity accuracy $150 - 200$ km/s (Le Fèvre et al. 2013). Relative velocities in the same slit (along the slit profile) can be measured to a better precision using e.g. accurate spectral line fitting.

The 2D and 1D spectra are then available for spectroscopic redshift measurements using the EZ environment (Garilli et al. 2010). The core algorithm to find a redshift is the cross-correlation with templates, confronted to a separate estimate of an emission line redshift when applicable. A key element for the cross-correlation engine to deliver a robust measurement is the availability of reference templates covering a large range of galaxy and star types, as well as a large range of rest wavelengths. This last point has to be carefully dealt with when measuring the highest redshift galaxies, as it is necessary that the templates go far enough in the UV, bluer than the Lyman-912Å limit, for the wavelength overlap between the observed galaxy and the redshifted template to be large enough to provide a robust correlation signal. We have used templates built over the years from VIMOS observations for the VVDS (Le Fèvre et al. 2005a, 2013) and the zCOSMOS survey (Lilly et al. 2007). As is relevant for $z > 2$ we have used templates with and without Lyman$-\alpha$ emission.

The redshift measurements are first obtained from an automated run with EZ. This serves as a basis for a visual examination of each spectrum, with an iteration on the redshift measurement using EZ in manual mode, if necessary. We find that at these high redshifts more than half of the spectra need manual intervention to properly measure a redshift. This is mainly due to residual defects in the spectra like sky features residuals after sky subtraction or second order spectra superimposition. As the manual intervention remains an important feature in this process, we have implemented a method to minimize measurement biases linked to one single person. One spectrum is measured by two team members separately, and these measurements are then confronted to produce a single measurement agreed on by the two measurers. With this scheme, we have implemented the same redshift reliability estimator as developed for the CFRS (Le Fèvre et al. 1995) and refined for the VVDS (Le Fèvre et al. 2005, 2013), zCOSMOS (Lilly et al. 2007) or the VIPERS (Guzzo et al. 2013) surveys. The reliability of a redshift measurement is expressed with a flag giving the range of probability for a redshift to be right. The reliability flag may take the following values:

- 0: No redshift could be assigned (the redshifts are then set to 9.9999).
- 1: 50–75% probability to be correct
- 2: 75–85% probability to be correct
- 3: 95–100% probability to be correct
- 4: 100% probability to be correct
- 9: spectrum with a single emission line. The redshift given is the most probable given the observed continuum, it has a $\sim 80\%$ probability to be correct.

The probabilities associated to these reliability flags are remarkably stable because of the process involving several independant people, smoothing out individual biases (Le Fèvre et al. 2013).

VUDS enters a redshift domain which has never been probed by spectroscopic redshift surveys on this scale. The expertise of the VUDS team members grew steadily as more and more of the data was being processed. Upon examination of the first measurements we realised that redshifts were probably wrongly assigned for a small but sizeable (∼ 10%) fraction of the objects. Several standard cases for erroneous measurements were identified, including the possible confusion between early M stars (M0-M3) and $z \sim 5$ absorption-line only galaxies (going both ways), the assignment of Lyα-1215Å instead of OII-3727Å (or vice-versa), the setting of a continuum break to the Balmer-D4000 break rather than the Lyα break (or vice-versa). In addition, the reliability flag as defined above was sometimes either too cautious when a measurer found a spectrum he/she was not yet familiar with, or too optimistic. In view of this, we opted to create a 'Tiger Team' (OLF, PC, EV, BG, DM, VLB, OLF, LP, LTa) in charge to conduct an additional redshift check, provided with a set of well defined reference cases and their treatment. This redshift check was done again by two independent people going through all the measured spectra, separately identifying which ones needed to have their redshift and/or reliability flag modified, and agreeing on the modifications. The check of all pointings was done by four pairs of 2 people, each couple proposing a list of modifications to be examined and agreed upon by the other 'Tiger team' members. We compare in Figure 4 the *old redshifts* with the *Tiger team redshifts* which summarizes this process. At the end of this process, about 10% of the objects had either a redshift or a flag change. While this does not garantee that there are no more obvious 'catastrophic failures' in spectroscopic redshift meassurements, this process ensures a homogeneous treatment of all spectra and the reduction of the main degeneracies present in measuring the redshifts of high redshift galaxies. Obviously, for the fainter objects where the information content of the spectra is not sufficient to solve a possible degeneracy, the reliability flag is assigned to the 'flag 1' category.

One key element of the selection function of the VUDS sample is the target sampling rate (TSR) defined as the ratio of the observed galaxies (all reliability flags) to the underlying parent photometric populations from which the spectroscopic targets have been selected. We find a global TSR of ∼ 30% for the VUDS survey, similar for the three observed fields. With respect to the total population with $i_{AB} \leq 25$, the parent sample of galaxies satisfying the VUDS selection criteria represents 10%, and hence the observed VUDS sample represents 3.3% of all galaxies with $i_{AB} \leq 25$. The most reliable redshifts including flag 2, 3, 4 and 9 account for 70.2% of the sample, flags 1 represent 21.4% and 8.4% could not be measured (flag 0). The spectroscopic success rate (SSR) as a function of magnitude is shown in Figure 5. Down to $i_{AB} = 25$ 91% of targets have a redshift measurement (flags 1, 2, 3, 4, 9), and 74.3% have a reliable measurement (flags 2, 3, 4, 9). This fraction is decreasing to 58% in the last 0.25 magnitude bin before $i_{AB} = 25$. The spectroscopic success rate then goes down with magnitude with a reliable redshift obtained for $\simeq$ 30% of the galaxies targeted at $i_{AB} = 26$. A complete description of the survey selection function including the analysis of the spectroscopic success rate at different redshifts will be provided elsewhere (Tasca et al. in prep.).

The experience gained in this process is invaluable for future massive high redshift surveys. As in Le Fèvre et al. (2013), we emphasize that redshift measurement *at these high redshifts* is a complex process which deserves dedicated and expert care beyond a simplistic 'good' vs. 'bad' redshift measurement scheme to fully exploit the information content of faint object spectroscopy at the instrumental limit. This is further discussed in Section 3.5 in view of the aposteriori comparison between spectroscopic and photometric redshifts.

3.4. VUDS spectra

We present sample spectra over the redshift range of the survey in Figures 6 to Figures 9. The signal-to-noise per sampling element (∼ 7Å) of the spectra at 1500Å rest-wavelength has a mean of $S/N = 4.5$, and $\sigma_{S/N} = 2.1$, the S/N per spectral resolution element being ∼ 2× higher. This gives access to a range of spectral features and properties for each individual galaxy. The main spectral lines identified in individual spectra are Lyman−α1215Å (in emission or in absorption), OIλ1303, CIIλ1334, the SiIV-OIV doublet at λ1394-1403Å, SiIIλ1527, CIVλ1549, FeIIλ1608, HeIIλ1640, AlIIλ1671, AlIIIλ1856, CIII]λ1909 (see the list in Table 4). Below Lyman−α, and depending on the IGM absorption, Lyman−β, Lyman−γ and the Lyman limit at 912Å can be identified. The average spectral properties in different redshift ranges are described in section 4.3.

An overview of the VUDS galaxy population over $2 < z <\sim 6$ is shown in Figure 10. This figure is built from all spectra with reliability flags 3 and 4 in $2 < z < 4$ and all spectra with flags 2, 3 and 4 for $z > 4$ assembled in one single image, one spectrum per image line. The display is quite striking as the eye is able to follow up some of the weakest spectral features up to the highest redshifts. The Lyα line is readily visible in absorption or emission all along the redshift range, with the blue wing of the broad damped Lyα clearly visible in all spectra. Going to the higher redshifts one can note the Lyβ, Lyγ, and Lyman-break (912Å).

3.5. Comparison between spectroscopic and photometric redshifts

With VUDS targets selected in large part from their photometric redshifts (Table 2), we make here a comparison between the photometric redshifts z_{phot} computed from the multi-wavelength data set and used as an input to the target selection, and the spectroscopic redshifts z_{spec}. We compare z_{phot} and z_{spec} for the VUDS sample with reliable flags 3 and 4 in Figure 11. As close to 100% of the z_{spec}(flags 3,4) are the "truth", we can directly test the accuracy and degeneracies of the z_{phot}. We present in Figure 15 the distribution of $\delta z = (z_{spec} - z_{phot})/(1 + z_{spec})$ for flags 3+4 and 2+9 for $z < 1.5$ and z> 2 separations arbitrarily chosen to distinguish a low redshift regime where the z_{phot} computation heavily relies on rest-frame visible domain features like the D4000 spectral break, and a high redshift domain where the computation rests on UV rest-frame features like the Lyα break produced by the intervening IGM and the Lyman-continuum limit. The width of the distribution for flags 3+4 below redshift $z = 1.5$ is an excellent $\sigma(\delta z/(1 + z)) = 0.02$, particularly in the COSMOS and ECDFS fields which benefit from medium band photometry. Above redshift $z = 2$ it is about double which is still excellent but signals the increasing difficulty to assign accurate photometric redshifts from broad band photometry only. It is immediately visible in Figure 11 that there is a secondary tight relation with a low z_{phot} corresponding to a high z_{spec}, which is easily explained by the degeneracy between the D4000 continuum break and the Lyman−α break when com-

puting z_{phot}, and our use of the secondary peak of the z_{phot} PDF to select spectroscopic targets. We draw in Figure 11 the relation $z_{phot} = (1215/4000) \times (1 + z_{spec}) - 1$ and ±15% around this mean. A number of galaxies are well within these limits, representing 9% of the flag 3 and 4 sample at $z > 2$. In all, 95.2% of our flag 3+4 sample at $z_{spec} > 2$ verify either $z_{spec} = z_{phot}$ or $z_{phot} = (1215/4000) \times (1 + z_{spec}) - 1$ within 15%, meaning that our selection function is particularly effective in picking–up galaxies at these redshifts, with low 'catastrophic failure' rate, as further discused in Tasca et al. (in prep.).

We can use the distribution of (z_{spec}, z_{phot}) for flags 3 and 4 to evaluate the level of agreement between z_{phot} and z_{spec} for other spectroscopic reliability flags 2, 1 and 9. The comparison between z_{phot} and z_{spec} for reliability flag 2, 1 and 9 are presented in Figure 12, 13 and 14, respectively. These plots are qualitatively similar to Figure 11, with the 1:1 and the 4000:1215 relationships well populated. For flags 2+9 and $z > 2$ we find $\sigma(\delta z) = 0.044$ hence not much different from the flags 3+4 distribution, and with about 75% to 80% of the objects in this category within 15% of the 1:1 and 1215:4000 relations, which fully supports that these redshifts have a high level of reliability. As noted above, the secondary peak in these distributions for $z > 2$ is produced by the degeneracy betweeen the D4000 and 1215Å continuum breaks. This peak is more pronounced for flag 2+9, with about 23% of the objects in this category within 15% of the 1215:4000 relation, meaning that the objects with these reliability flags are more prone to this degeneracy. A possible reason is that the magnitudes of flag 2+9 objects are on average fainter than the flag 3+4 counterparts. Defining the catastrophic failure rate in the selection of $z > 2.4$ VUDS galaxies as the fraction of galaxies which are outside either $\mid z_{spec} - z_{phot} \mid \leq 0.15 \times (1 + z_{spec})$ or $z_{phot} - (1215/4000) \times (1 + z_{spec}) - 1 \leq 0.15$, it is 20% for flag 2, 37% for flag 1 and 24% for flag 9. Taking into account the intrinsic catastrophic failure rate of 5% for photometric redshifts as observed for flag 3+4 this gives a qualitative estimate on the reliability level of the different spectroscopy flags.

Based on the excellent match between spectroscopic and photometric redshifts obtained for flag 3 and 4, we have added a decimal point to the reliability flag as defined in Section 3.3 translating the level of agreement between the spectroscopic redshift and the photometric redshift for each galaxy. This decimal point may take five different values from 1 for a poor agreement, to 5 for an excellent agreement; more specifically: *.1 means that the spectroscopic and photometric redshifts have a difference $dz = \mid z_{spec} - z_{phot} \mid / (1 + z_{spec}) \geq 0.5$, *.2 is for $0.3 \leq dz < 0.5$, *.3 for $0.2 \leq dz < 0.3$, *.4 for $0.1 \leq dz < 0.2$, and *.5 for $dz < 0.1$. Adding a *.5 decimal therefore further increases the reliability of the spectroscopic redshift measurement, while a *.1 decimal would rather lower it. However, for the higher spectroscopic reliability flag 3 and 4, a low photometric decimal *.1 or .2 would rather indicate that the photometric redshift is likely to be incorrect. This scheme allows to define a sample depending on the level of robustness required by a particular analysis.

This analysis brings two general comments for studies based on photometric redshifts. First, if VUDS targets had been selected based only on the primary peak of the z_{phot} PDF, ∼ 17.5% of the sample with $z \geq 2$ would have been lost compared to our selection. Second, for studies based only on a z_{phot} sample, which necessarily assign a redshift using the primary peak of the redshift PDF, 14.2% of high redshift galaxies would be missed because they would be wrongly placed at low redshift instead of $z > 2.3$. This average value shows a variation with redshift, which will be discussed in a forthcoming paper.

4. General properties of the VUDS sample

In this section we report on the main properties of the VUDS sample, giving an overview of the parameter space probed by the survey. As of this writing, the VUDS sample contains 6 250 objects with a measured redshift, including 6 003 galaxies, 20 AGNs and 227 stars, and no redshift measurement could be obtained for 750 objects. These numbers will increase by 15-20% when data processing will be completed. The projected sky distribution of the VUDS sample follows the layout of the VIMOS pointings, as identified in Figures 1 to 3. With 8 pointings on the COSMOS field, ∼ 4150 objects previously without measurements have been observed and ∼ 3700 have now spectroscopic redshift measurements (these numbers are being revised following the last data processing), covering a total area of 1800 arcmin^2. Over 1125 arcmin^2 in the VVDS-02h field the data for ∼ 2300 targets have been processed and ∼ 2100 objects have now spectroscopic redshifts, and we expect 20% more when data processing will be completed. In the ECDFS, one pointing has been processed so far with ∼ 550 objects observed and ∼ 500 with a redshift measurement, and with 2 more pointings to process (one with LR and one with MR grisms) we expect this number to increase by about 60%. We use the current sample to discus general properties of the sample below.

4.1. Sample properties

The efficiency of the survey target selection can be estimated from the redshift distribution of the observed sources.

The redshift distribution of the VUDS sample is shown in Figure 16. The N(z) is bimodal, with a high redshift component from $z \sim 2$ to $z \sim 6.5$, and a low redshift component mainly at $z < 1.5$. The sample above $z = 2$ is the sample of interest for the main science goals of VUDS; it is representing about 80% of the total sample, with currently more than 4500 objects with redshift measured with $z \geq 2$, and, extrapolating for the remaining data to be processed the VUDS sample, it will contain ∼ 6000 objects with $2 < z < 6.5$ in the end. VUDS is today the largest sample of galaxies with spectroscopic redshifts in any of the redshift bins $2 < z < 3$, $3 < z < 4$, $4 < z < 4.7$ or $4.7 < z < 5.3$, as further discussed in Section 5.

The sample below $z \simeq 2$ is made of several sub-samples: (i) galaxies for which the second peak of the z_{phot} PDF is at $z > 2.4$ but which turned out to be at the lower redshift indicated by the first peak, (ii) galaxies for which the selection criteria for high redshift failed and which are rather at low redshifts, (iii) galaxies with $i_{AB} \leq 25$ which have been used as mask fillers and for which the N(z) is expected to peak at $z \sim 1.5$ (Le Fèvre et al. 2014), and (iv) some serendipitous sources falling in the slits by chance. The mean redshift for the sample above $z = 2$ is $\bar{z} = 3.0$. About 10% of the sample is above $z = 4$, and the high redshift tail goes up beyond redshift $z = 6$, the highest reliable redshift so far being $z = 6.5363$. The current redshift distribution for $z \geq 2$ in the COSMOS, ECDFS and VVDS-02h fields is shown in Figure 17 to Figure 19. The redshift bin of $dz = 0.01$ enables to show the strong clustering present at all redshifts probed by the survey. Some of the densest peaks are remarkable examples of clustering in the early universe, as discussed in Cucciati et al. (submitted) and Lemaux et al. (submitted).

As our sample is z_{phot} selected, it is interesting to check were the VUDS galaxies are distributed in several standard color-color diagrams. Following other studies (e.g. Kurk et al. 2013, Le Fèvre et al. 2014), this *a posteriori* analysis gives those studies using color-color selection an indication of both the ef-

ficiency of the selection and the contamination by galaxies at other redshifts than the redshift of interest in the color space selected. We examine here the *gzK* diagram used to select galaxies with $1.4 < z < 2.5$, and the *ugr* diagram for galaxies with $2.5 < z < 3.5$. Other color-color diagrams like the *gri* or *riz* for $3.5 < z < 4.5$ and $4.5 < z < 5.5$, will be investigated in a forthcoming paper.

The BzK selection is based on the identification of the Balmer and D4000 break crossing the z band for $1.4 < z < 2.5$ (Daddi et al. 2004). We present the *gzK* diagram in Figure 20, with the color-color area adjusted to take into account the different wavelength coverage of the *g* filter compared to the B filter (Bielby et al. 2012, Le Fèvre et al. 2014). The *gzK* criteria are efficient to select 92% of the VUDS galaxies with $K \leq 24$. However, we note that the level of contamination of a *gzK* selected sample down to $K_{AB} = 24$ would be quite high as 58% of galaxies in the selection area of the *gzK* diagram selecting the $1.4 < z < 2.5$ redshift range would be outside this range at $z < 1.4$ or $z > 2.5$ (right panel of Figure 20); at $K_{AB} = 22$ we find this contamination to be lower, at the ∼ 30% level. This trend is similar to that found by Le Fèvre et al. (2014), who identified a magnitude-dependent contamination level. As the VUDS data is selected with $z_{phot} > 2.4$ the contamination level is further enhanced compared to the pure $i-band$ magnitude selection of the VVDS.

Going to higher redshifts, we show the distribution of VUDS galaxies with $2.5 < z < 3.5$ in the *ugr* color-color diagram in Figure 21. Here again a large fraction ∼ 80% of galaxies appear in the expected locus of the color-color diagram. The contamination by galaxies at $z < 2.5$ or $z > 3.5$ is quite high at about 40%. We note that we did not attempt to optimize the redshift range on the basis of the exact shape of the photometric bands used for the CFHTLS photometry used in this plot, but we rather elected to show the distribution of a large population.

This a posteriori color-color analysis shows that the VUDS sample is behaving as generaly expected for galaxies at these redshifts. The photometric redshift selection allows to identify galaxies beyond the classical color-color locus of the *gzK* and *ugr* diagrams. We also point out the strong contamination present in color-color selected samples from galaxies at other redshifts outside the targeted redshift range.

4.2. Absolute magnitudes, stellar masses and star formation rates

The distribution of apparent magnitudes with redshift is shown in Figure 22. In all ∼ 90% of the distribution is within $23 \leq i_{AB} \leq 25$, while ∼ 10% have $25 \leq i_{AB} \leq 27$.

The VUDS sample covers a large range of galaxy physical properties. Using the VUDS spectroscopic redshifts we perform SED fitting on the multi-wavelength photometry using the code Le Phare (Ilbert et al. 2006), as described in Ilbert et al. (2013). We summarize here a few key points, but we refer to the recent description of the fitting process in Ilbert et al. (2013) for a detailed account and associated limitations. Galaxy luminosities are transformed into stellar mass using the best fit synthetic template, from a list of templates built from Bruzual & Charlot (2003) stellar population synthesis models with 3 metallicities ($Z = 0.004$, $Z = 0.008$, and solar $Z = 0.02$), and exponentially declining and delayed SFR with 9 different τ values from 0.1 to 30 Gyr. We use a Calzetti (2000) extinction law, and emission lines are added to the synthetic spectra as described in Ilbert et al. (2009). The output of this fitting process includes among other parameters: absolute magnitudes integrated into standard bands, stellar masses, star formation rates, and extinction. The distributions in absolute NUV (2300Å rest) magnitude, stellar masses, and star formation rate of the sample are presented in Figure 23. At redshifts $z \sim 3-4$ the NUV absolute magnitude ranges from $NUV = -20.5$ to $NUV = -23$, the stellar mass from $10^9 M_\odot$ to $10^{11} M_\odot$, and the star formation rate from below $1 M_\odot$/yr up to several hundred $M_\odot$/yr.

The VUDS survey therefore covers a large parameters space both in the observed properties and the physical properties of the sample galaxies.

4.3. Average spectral properties: stacked spectra

Stacked spectra have been produced from previous surveys, like Shapley et al. (2003) for $z \sim 3$, or for B, V and R dropout galaxies (Vanzella et al. 2009). The large statistics and wavelength coverage of VUDS offers the opportunity to produce composite spectra over a large redshift coverage. The average spectral properties of galaxies over the redshift range $2 \leq z_{spec} \leq 6.5$ are derived from stacking VUDS spectra in different redshift bins. For each redshift bin, the average spectra are produced using the `odcombine` task in IRAF, averaging spectra after scaling to the same median continuum value, i.e. luminosity weighted, and weighting spectra to their mean flux in the same rest-frame wavelength range. Average spectra using sigma clipping, removing at each wavelength pixel those with a value $1.5-3$ times the 1σ value, have been compared to the straight weighted average described above, and very little difference have been observed when the number of spectra is large (> 50). For smaller samples, sigma clipping helps improving the S/N of the stacks by removing e.g. the left-over signatures of the sky subtraction process.

The average spectra of all galaxies for several redshift bins are shown in Figures 24 to 28, together with the average spectra of those with and without Ly-α emission. We discuss some of the key features below.

We find that the fraction of galaxies with any trace of Lyα emission is strongly changing with redshift with $30.6 \pm 1.8\%$ in $2 \leq z \leq 3$, $38.3 \pm 2.9\%$ in $3 \leq z \leq 4$, $61.6 \pm 7.0\%$ in $4 \leq z \leq 4.7$, and $66.6 \pm 14.2\%$ in $4.7 \leq z \leq 5.3$. The detailed properties of the Lyα emitting fraction and implications are examined in details in the accompaning paper by Cassata et al. (submitted).

Besides Lyα a number of spectral features are noteworthy, as indicated in Table 4. The main absorption features redder than Lyα include the SiIIλ1260, OIλ1303, CIIλ1334, SiIVλ1394/1403, SiIIλ1527, CIVλ1549 lines. In emission, weak CIII]λ1909 is quite common, and one can identify S-shape absorption-emission features for e.g. SiIIλ1260, OIλ1303 or CIVλ1549, indicative of strong outflows. The HeIIλ1640 emission is ubiquitous, as seen on all spectra where this line is in the wavelength range (Figures 24 to 26). This line, when several hundreds of km/s in width, indicates the presence of strong winds around Wolf-Rayet stars. If narrow it may indicate some other processes like the presence of a population of low-metallicity stars with properties akin to Population III (Cassata et al. 2013).

Below the Lyα line, the main features identified in galaxies at these redshifts are SiIIλ1192, and then the Lyβ and Lyγ lines, followed by the Lyman-limit producing a continuum break at 912Å. It is interesting to note that the flux below 912Å in our stacked spectra is not zero as would be expected if the lyman continuum escape fraction was 0%. We observe a significant detection of flux below the Lyman-limit in the stacked spectra of galaxies at $3 < z < 4$, $4 < z < 4.7$ and $4.7 < z < 5.3$. The ra-

tio of continuum flux density in the [1400,1500]Å range over the flux in [800,900]Å range below the 912Å Lyman-limit is $f_{1500}/f_{900} = 32 \pm 3$, 39 ± 5, and 33 ± 10, in these three redshift bins respectively. Although these measurements are based on a much larger sample covering a larger redshift range, these values are comparable to the observed value of 58 ± 18 reported by Shapley et al. (2006) at $z \sim 3$. While it is tempting to interpret this in terms of the Lyman-continuum escape fraction, we note that at the faint magnitudes we are observing, the flux observed below 912Å in stacked spectra could also be coming in part from objects more or less contaminated by lower redshift interlopers producing observed flux below that of the Lyman-limit at the rest-frame of the distant source (e.g. Vanzella et al. 2010). The *observed* non-zero flux below 912Å has consequences on the selection of high redshift galaxies based on the Lyman-break technique, as expected colors of high redshift galaxies are affected by apriori hypotheses, e.g. on the f_{1500}/f_{900} ratio. This and the corrected Lyman escape fraction will be analyzed in forthcoming papers.

Our stacked spectra beautifully show the evolution of the 'stair-case' pattern of IGM absorption as a function of redshift produced by continuum blanketing from the Lyman series of galaxies along the line of sight in the volume probed (Madau 1995). The comparison of the observed mean and distribution of IGM absorption properties as a function of redshift in VUDS to the models of Madau (1995) and Meiksin (2010) will be discussed in Thomas et al. (in preparation).

A detailed analysis of the spectral properties of stacked spectra and individual galaxies in the VUDS sample will be presented in forthcoming papers.

5. Comparison with other spectroscopic surveys at high redshifts

Large spectroscopic surveys at $z > 2$ published in the literature are understandably relatively limited compared to surveys at $z < 1.5$ because of the faintness of the sources and requirements on deep imaging to perform pre-selection.

We compile a list, likely not exhaustive, of the main surveys in the redshift range $2 < z < 7$ in Table 5, and we show the corresponding redshift distribution of these surveys in Figure 29. The largest numbers of galaxies can be found in the range $2 < z < 3$ with the LBG-selected surveys of Steidel et al. (1999, 2003, 2004), the zCOSMOS-Deep survey (Lilly et al. 2007, Lilly et al. in prep.), the VLRS (Bielby et al. 2013), and the VVDS-Deep/Ultra-Deep (Le Fèvre et al. 2013). In the redshift range $2.5 < z < 3.5$, these surveys accumulate ∼ 2700 galaxies with spectroscopic redshifts. The VUDS survey contributes ∼ 2800 spectroscopic redshifts, therefore as much as all these other surveys combined, but in a single well-controled survey.

Beyond redshift $z \sim 3.5$ both the number of spectroscopic surveys and the number of galaxies with measured spectroscopic redshifts dramatically fall down. Spectroscopic campains on the GOODS-South area with VIMOS and FORS2 on the VLT have produced about 114 galaxies with redshifts $z \sim 3-6$. The Keck-DEIMOS surveys of Stark et al. 2010 contribute more than 300 galaxies in the range $3 < z < 6.5$. VUDS brings more than 800 new galaxies with spectroscopic redshifts in the range $3.5 < z < 5$, or about twice the number of all other surveys combined as can be seen in Figure 29.

At the highest redshifts $z > 5$, we enter the game of single redshift confirmations following-up on galaxies identified as *dropout*, i.e. with a sharp change in flux between two adjacent photometric bands (e.g. Stark et al. 2010, Curtis-Lake et al. 2012), or following narrow-band Lyman Alpha Emitter candidates (e.g. Ouchi et al. 2010). These different surveys are bringing several tens of spectroscopic identifications beyond $z = 5$. VUDS contributes about 60 new sources with spectroscopic redshifts $z > 5$. This sample is being consolidated, particularly at $z > 6$, and will be the subject of future studies.

An important element to keep in perspective is the area which has been sampled by a survey. As discussed in Moster et al. (2011), the fluctuations in number density of objects observed in deep surveys result from the cosmic variance in the distribution of galaxies in large scale structures. Surveys with one square degree at $z \sim 3$ will be subject to an uncertainty of about 10% from cosmic variance, in addition to the uncertainties related to the number of objects in a survey. For fields of the GOODS size (∼ 150arcmin^2), the cosmic variance at $z \sim 3$ is expected to be in excess of 50% for galaxies with stellar masses $M_* > 10^{10.5}$M$_\odot$. We compare in Figure 30 the number of spectra per square degree vs. the area for different surveys, a high redshift version of that presented in Baldry et al. (2010) and updated by Le Fèvre et al. (2013). While at $z \sim 3$, several surveys have surveyed about 1 square degree (Steidel et al. 2003, Lilly et al. 2007, Bieby et al. 2013, or Le Fèvre et al. 2013), the highest redshift surveys have surveyed only about 0.1 deg^2 (e.g. Vanzella et al. 2009, Stark et al. 2010). In this context, VUDS has selected galaxies at any redshift $z > 2$ in three different fields for a total of ∼ 1deg^2.

6. Summary

The VIMOS Ultra Deep Survey (VUDS) is a deep spectroscopic redshift survey aiming to study the early phases of galaxy assembly at $2 < z < 6.5$ from a sample of ∼ 10 000 galaxies observed with the VIMOS multi-slit spectrograph at the ESO-VLT. The survey target selection is based on photometric redshifts derived from extensive multi-wavelength data, combined to color and color-color selection as well as magnitude-selection. Most of the sample is limited down to $i_{AB} = 25$, but galaxies are observed as faint as $i_{AB} = 27$.

The combination of a wide wavelength coverage from 3650Å to 9350Å, and exposure times of ≃ 14h, lead to a spectroscopic success rate in redshift measurement of about 91% (74% for spectroscopic reliability flags 2 to 9) down to $i_{AB} = 25$. The comparison of photometric redshifts to the VUDS spectroscopic redshifts shows that the VUDS strategy minimizes the loss of galaxy populations compared to more restrictive selection criteria.

We report on the general properties of the sample based on ∼ 80% of the data which has already been processed. The redshift distribution of the current sample at $z \geq 2$ peaks at a mean $z = 3$, and extends beyond $z = 6$. A secondary sample at $z < 2$ is the result of the selection function and provides interesting very low intrinsic luminosity galaxies. The average spectral properties of galaxies with $z > 2$ are discussed based on high S/N stacks of VUDS spectra in several increasing redshift bins. Galaxies with and without Lyα in emission are found at any redshift, but it is found that the fraction of galaxies with Lyα in emission increases with redshift. This is quantified in the accompaning paper by Cassata et al. (submitted). Using stacked spectra we find that there is observed flux below the 912Å Lyman limit in all the redshift ranges explored, the origin of this is being investigated and will be the subject of future papers.

Following an early measurement of the merger rate at $z \sim 3$ (Tasca et al. 2013), several papers presenting results from the

VUDS are submitted together with this paper (Cucciati et al. submitted, Cassata et al. submitted, Lemaux et al. submitted, Amorin et al. submitted). A number of other analyses are in progress.

VUDS is the first deep spectroscopic survey covering such a large redshift range with such a large sample of galaxies with confirmed spectroscopic redshifts. It is ideally suited for detailed studies of the galaxy population at early times $2 < z < 6$. When the full data set will be completed, it is foreseen to make future VUDS data releases publicly available.

Acknowledgements. We thank ESO staff for their continuous support for the VUDS survey, particularly the Paranal staff conducting the observations and Marina Rejkuba and the ESO user support group in Garching. This work is supported by funding from the European Research Council Advanced Grant ERC-2010-AdG-268107-EARLY and by INAF Grants PRIN 2010, PRIN 2012 and PICS 2013. AC, OC, MT and VS acknowledge the grant MIUR PRIN 2010–2011. DM gratefully acknowledges LAM hospitality during the initial phases of the project. This work is based on data products made available at the CESAM data center, Laboratoire d'Astrophysique de Marseille. This work partly uses observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

References

Abazajian et al., 2009, ApJS, 182, 543
Arnouts, S., et al., 2007, A&A, 476, 137
Appenzeller, I., et al. 1998, The Messenger, 94, 1
Baldry, I., et al., 2010, MNRAS, 404, 86
Bertin, E., Arnouts, S., 1996, A&AS, 117, 393
Bielby, R., et al., 2012, A&A, 545, 23
Bielby, R., et al., 2013, MNRAS, 430, 425
Bondi, M., et al., 2003, A&A, 403, 857
Bouwens, R. et al., 2009, ApJ, 705, 936
Bruzual, G., & Charlot, S., 2003, MNRAS, 344, 1000
Calzetti, D., et al., 2000, ApJ, 533, 682
Capak, P., et al., 2011, Nature, 470, 233
Caputi, K., et al., 2012, ApJ, 750, 20
Cardamone, C., et al., 2010, ApJS, 189, 270
Cassata, P., Le Fèvre, O., et al., 2011, A&A, 525, 143
Cassata, P., Le Fèvre, O., et al., 2013, A&A, 556, 68
Cimatti, A., et al., 2002, A&A, 381, 68
Cimatti, A., et al., 2008, A&A, 482, 21
Colless, M., et al., 2001, MNRAS, 328, 1039
Cooper, M.C., et al., 2012, MNRAS, 425, 2116
Cucciati, O., et al., 2012, A&A, 539, 31
Cuillandre, J.C., et al., 2012, Proc. SPIE 8448, Observatory Operations: Strategies, Processes, and Systems IV, 84480
Curtis-Lake, E., et al., 2012, MNRAS, 422, 1425
Daddi, E., et al., 2004, ApJ, 617, 746
Davis, M., et al., 2003, SPIE, 4834, 161
Ellis, R.S., et al. 2013, ApJ, 763, L7
Faber, S. M., et al. 2003, Proc. SPIE, 4841, 1657
Garilli, B., et al., 2010, PASP, 122, 827
Giacconi, R., et al., 2002, ApJS, 139, 369
Grogin, N., et al., 2011, ApJS, 197, 35
Guo, Q., et al., 2011, MNRAS, 413, 101
Guzzo, L., et al., 2014, arXiv:1303.2623
Hammersley, P. et al., 2010, The Messenger, 142, 8
Ilbert, O., et al., 2006, A&A, 457, 841
Ilbert, O., et al., 2009, ApJ, 690, 1236
Ilbert, O., et al., 2013, A&A, 556, 55
Kashikawa, N., et al., 2011, ApJ, 734, 119
Koekemoer, A., et al., 2007, ApJS, 172, 196
Koekemoer, A., et al., 2011, ApJS, 197, 36
Le Fèvre, O., Crampton, D., Lilly, S.J., Hammer, F., Tresse, L., 1995, ApJ, 455, 60
Le Fèvre, O., et al., 2003, SPIE, 4841, 1670
Le Fèvre, O., et al., 2004, A&A, 417, 839
Le Fèvre, O., et al., 2005a, A&A, 439, 845
Le Fèvre, O., et al., 2005b, Nature, 437, 519
Le Fèvre, O., et al., 2013, A&A, 559, 14
Le Fèvre, O., et al., 2014, A&A, arXiv:1307.6518
Lemaux, B., et al., 2014, A&A, in press, arXiv:1311.5228
Lilly, S.J., Le Fèvre, O., Crampton, D., Hammer, F., 1995a, ApJ, 455, 50
Lilly, S. J., Le Fèvre, O., Hammer, F., Crampton, D., 1996, ApJ, 460, 1
Lilly, S.J., Le Fèvre, O., et al., 2007, ApJS, 172, 70
Lonsdale, C.C., et al., 2003, PASP, 115, 897
Madau, P., 1995, ApJ, 441, 18
Madau, P., Ferguson, H.C., Dickinson, M., Giavalisco, M., Steidel, C.C., Fruchter, A., 1996, MNRAS, 283, 1388
Madau, P., & Dickinson, P., 2014, ARAA, in press (arXiv:1403.0007)
Mauduit et al., 2012, PASP, 124, 714
McCracken, H.J., et al., 2012, A&A,
Mo, H., van den Bosch, F., White, S., Galaxy Formation and Evolution, Cambridge University Press, 2010
Moster, B.P., Somerville, R.S., Newman, J.A., Rix, H-W., 2011, ApJ, 731, 113
Oke, J.B., 1995, PASP, 107, 3750
Oliver, S., et al., 2012, MNRAS, 424, 1614
Ouchi, M., et al., 2008, ApJS, 176, 301
Patat, N., et al., 2011, A&A, 527, 91
Pierre, M., et al., 2004, JCAP, 09, 011
Popesso, et al., 2009, A&A, 494, 443
Rix, H.W., et al., 2004, 2004, ApJS, 152, 163
Sanders, D., et al., 2007, ApJS, 172, 86
Schlegel, D., et al., 1998, ApJ, 500, 525
Scodeggio, M., et al., 2005, PASP, 117, 1284
Scodeggio, M., et al., 2009, The Messenger, 135, 13
Scoville, N., et al., 2007, ApJS, 172, 1
Shapley, A., et al., 2006, ApJ, 651, 688
Shimasaku, K., 2006, PASJ, 5, 313
Springel, V. et al., 2008, MNRAS, 391, 1685
Stark, D., et al., 2010, MNRAS, 408, 1628
Steidel, C.C., Giavalisco, M., Pettini, M., Dickinson, M., Adelberger, K.L., 1996, ApJ, 462, 17
Steidel, C.C., Adelberger, K.L., Giavalisco, M., Dickinson, M., Pettini, M., 1999, ApJ, 519, 1
Steidel, C.C., et al., 2003, ApJ, 592, 728
Steidel, C.C., et al., 2004, ApJ, 604, 534
Taniguchi, Y., et al., 2005, PASJ, 57, 165
Taniguchi, Y., et al., 2007, ApJS, 172, 9
Tasca, L., et al., 2014, A&A, in press, arXiv:1303.4400
Tresse, L., et al., 2007, A&A, 472, 403
Vanzella, E., et al., 2008, A&A, 478, 83
Vanzella, E., et al., 2009, ApJ, 695, 1163
Vanzella, E., et al., 2010, MNRAS, 404, 1672
Wiklind, T. et al., 2008, ApJ, 676, 781
Williams, R., et al., 2000, AJ, 120, 2735
Windhorst, R., et al., 2011, ApJS, 193, 27

Table 1. VUDS fields

Field	α_{2000}	δ_{2000}	b	l	Area	*Depth*
COSMOS	$10h00m04.0s$	+02 deg 12′40″	42.1	236.8	1800 arcmin^2	$i_{AB} \simeq 25$
ECDFS	$03h32m28.0s$	−27 deg 48′30″	−54.0	223.5	675 arcmin^2	$i_{AB} \simeq 25$
VVDS-02h	$02h26m00.0s$	−04 deg 30′00″	−57.5	172.0	1125 arcmin^2	$i_{AB} \simeq 25$

Table 2. VUDS spectroscopic target selection criteria

Criterion	Value	Limiting magnitude	Fraction of targets
Photometric redshift, 1st PDF peak	$z_{phot} + 1\sigma \geq 2.4$	$22.5 \leq i_{AB} \leq 25$	58.3%
Photometric redshift, 2nd PDF peak	$z_{phot} \geq 2.4$	$22.5 \leq i_{AB} \leq 25$	19.3%
LBG ugr	$2.7 < z < 3.5$	$i_{AB} \leq 27$	1.9%
LBG gri	$3.5 < z < 4.5$	$i_{AB} \leq 27$	1.9%
LBG riz	$4.5 < z < 5.5$	$i_{AB} \leq 27$	3.6%
Break SED-based	–	$K_{AB} \leq 24$	3.5%
Magnitude-selected	–	$23 \leq i_{AB} \leq 25$	11.5%

Table 3. VUDS: observed VIMOS fields

Pointing	α_{2000}	δ_{2000}	Grism
COSMOS-P01	$09h59m02.39s$	+01 deg 54′35.9″	LRBLUE & LRRED
COSMOS-P02	$10h00m04.08s$	+01 deg 54′35.9″	LRBLUE & LRRED
COSMOS-P03	$10h01m05.76s$	+01 deg 54′35.9″	LRBLUE & LRRED
COSMOS-P04	$09h59m02.39s$	+02 deg 12′41.4″	LRBLUE & LRRED
COSMOS-P05	$10h00m04.08s$	+02 deg 12′41.4″	LRBLUE & LRRED
COSMOS-P06	$10h01m05.76s$	+02 deg 12′41.4″	LRBLUE & LRRED
COSMOS-P07	$10h00m04.08s$	+02 deg 30′46.7″	LRBLUE & LRRED
COSMOS-P08	$10h01m05.76s$	+02 deg 30′46.7″	LRBLUE & LRRED
ECDFS-P01	$03h32m25.99s$	−27 deg 41′59.9″	LRBLUE & LRRED
ECDFS-P02	$03h32m34.00s$	−27 deg 53′59.9″	LRBLUE & LRRED
ECDFS-P03[a]	$03h32m15.00s$	−27 deg 49′59.9″	MR
VVDS02-P01	$02h26m44.51s$	−04 deg 16′42.8″	LRBLUE & LRRED
VVDS02-P02	$02h25m40.34s$	−04 deg 16′42.8″	LRBLUE & LRRED
VVDS02-P03	$02h26m44.51s$	−04 deg 34′50.3″	LRBLUE & LRRED
VVDS02-P04	$02h25m40.34s$	−04 deg 34′50.3″	LRBLUE & LRRED
VVDS02-P05	$02h24m36.14s$	−04 deg 44′57.9″	LRBLUE & LRRED

[a] The ECDFS-P03 has been observed with a VIMOS PA=70deg on the sky.

Table 4. Main spectral features in VUDS spectra

Spectral line	λ_{rest} (Å)	Line type
Lyman−limit	912.0	Continuum break
Lyman−γ	972.0	HI absorption
Lyman−β	1025.2	HI absorption
SiIIλ1192	1192.0	ISM, blend 1190+1193
Lyman−α	1215.7	HI emission & absorption
SiIIλ1260	1260.4	ISM
OI+SiII-1303	1303.2	ISM, blend
CIIλ1334	1334.5	ISM
SiIVλ1394	1393.8	ISM
SiIVλ1403	1402.8	ISM
SiIIλ1527	1526.7	ISM
CIVλ1549	1549.1	ISM, blend 1548.2+1550.8
FeIIλ1608	1608.5	ISM
HeIIλ1640	1640.0	Nebular
AlIIλ1671	1670.8	ISM
FeIIλ1855	1854.7	ISM
FeIIλ1863	1862.8	ISM
CIII]λ1909	1908.7	Nebular, blend 1907+1909
FeIIλ2344	2343.5	ISM
FeIIλ2371	2370.5	ISM
FeIIλ2402	2402.6	ISM
FeIIλ2594	2593.7	ISM
MgIIλ2796	2796	ISM

Table 5. Comparison of the VUDS survey with other spectroscopic redshift surveys at $z > 2$ in the literature, by order of increasing mean redshift

Survey	*Area* deg^2	*Depth* i_{AB} *eq.*	N_{obj} *in Survey*	N_{obj} *at* $z \geq 2$	z_{range}	z_{mean}	*Selection*	*Reference*
VVDS-Deep	0.74	24.00	11 601	634	0 – 5	0.92	$17.5 \leq I_{AB} \leq 24.0$	Le Fèvre et al. 2013
VVDS-UDeep	0.14	24.75	941	341	0 – 4.5	1.38	$23.0 \leq i_{AB} \leq 24.75$	Le Fèvre et al. 2014
Steidel-z2	0.48	$R = 25.5$	851	588	1.4 – 2.5	2.0	$BM - BX$	Steidel et al. 2004
zCosmos-Deep	1	23.75	∼ 7 500	∼ 4 100	1.5 – 3	2.1	$B_{AB} \leq 25$+color	Lilly et al. 2007, Lilly et al. in prep.
VLRS	1.62	24.7	2 135	2 135	2 – 3.5	2.8	$23 < R < 25$	Bielby et al. 2013
LBG-z3	0.38	24.8	1 000		2.7 – 3.5	3.2	$R_{AB} < 25.5$+color	Steidel et al. 2003
GOODS	0.09	$z = 26$	887	114	3	3.5	$(i_{775} - z_{850}) > 0.6, z_{850} < 26$	Vanzella et al. 2009*andref.therein*
VVDS-LAE	0.74	–	217	217	0 – 6.7	3.5	$23.0 \leq i_{AB} \leq 24.75$	Cassata et al. 2011
LBG-z4	0.38	25.0	300	300	3.5 – 4.5	4.0	$I_{AB} < 25$color	Steidel et al. 1999
Dropout-z456	0.09	–	310	310	3 – 6.5	4.5	Dropout, Lyα break	Stark et al. 2010
LAE-z6	1	∼ 27	16	16	∼ 6.5	6.5	LAE narrow band	Ouchi et al. 2010
VUDS	1	25 – 27	∼ 10 000[a]	∼ 6 000[a]	2 – 6.7	3.7	$i_{AB} < 25$+photo-z	This paper

[a] Adjusted to take into account new data not yet processed

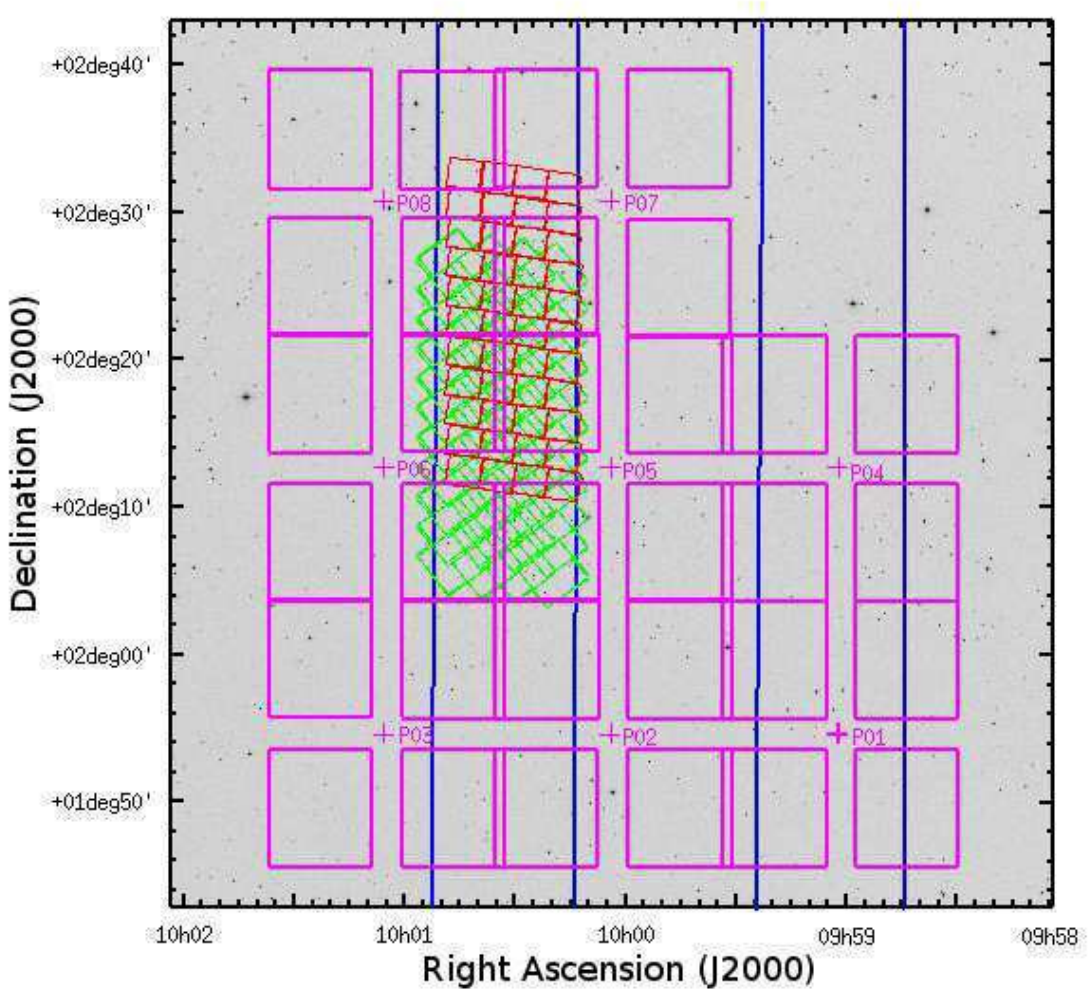


Fig. 1. Layout of the observed 8 VUDS VIMOS pointings in the COSMOS field. The center of each VIMOS pointing is identified by a cross with the pointing number (see Table 3), while the imprint of the 4 quadrants is in magenta. The positions of the UltraVista Deep stripes overlapping with the VUDS area are identified by the blue regions. The CANDELS ACS-F814W (in green) and WFC3-F160W (in red) areas are indicated. The size of the image is 1×1 deg^2.

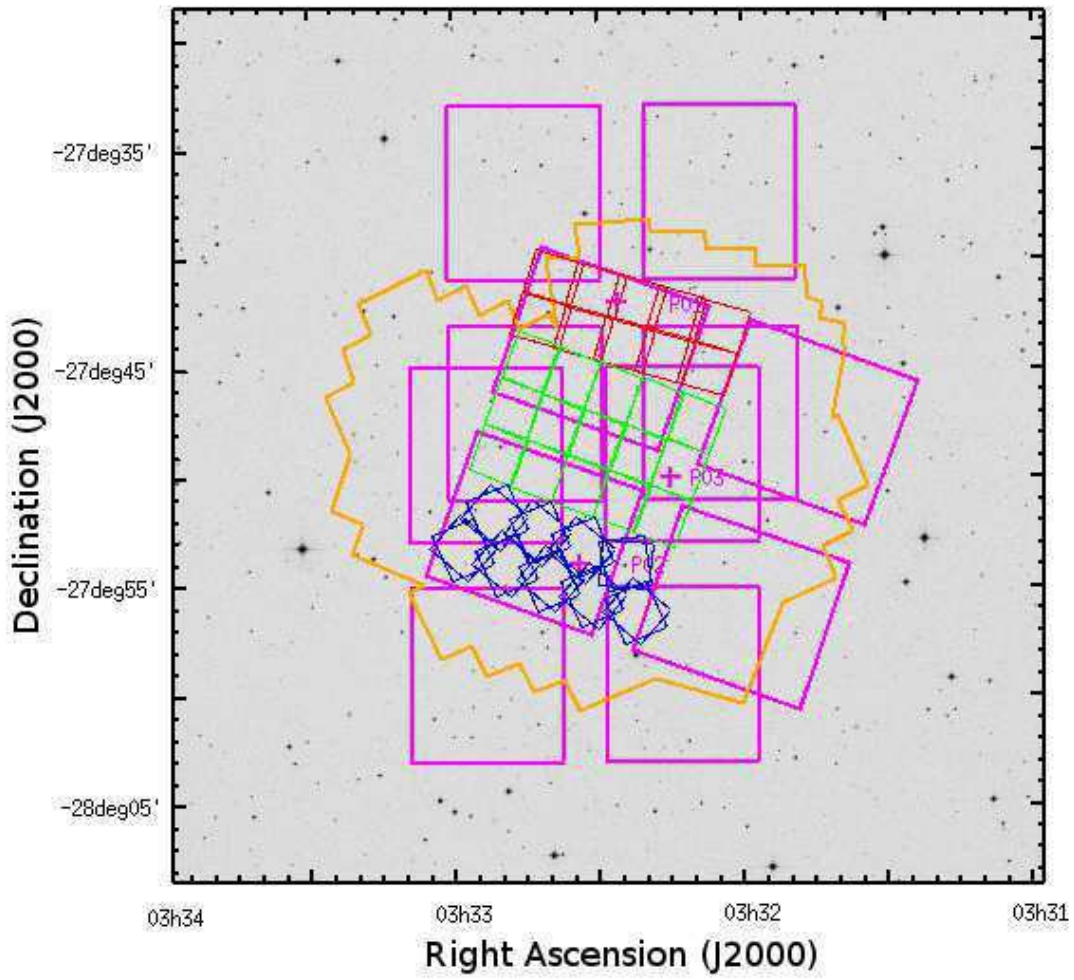


Fig. 2. Layout of the observed 3 VUDS VIMOS pointings in the ECDFS field. The center of each VIMOS pointing is identified by a cross with the pointing number (see Table 3), while the imprint of the 4 quadrants is in magenta. The positions of the CANDELS WFC3 deep area is indicated in green, and the CANDELS wide area in blue. The red region indicates the WFC3 coverage of the ERS. The outline of the existing ACS-F814W imaging is identified in orange. The size of the image is 0.67×0.67 deg^2.

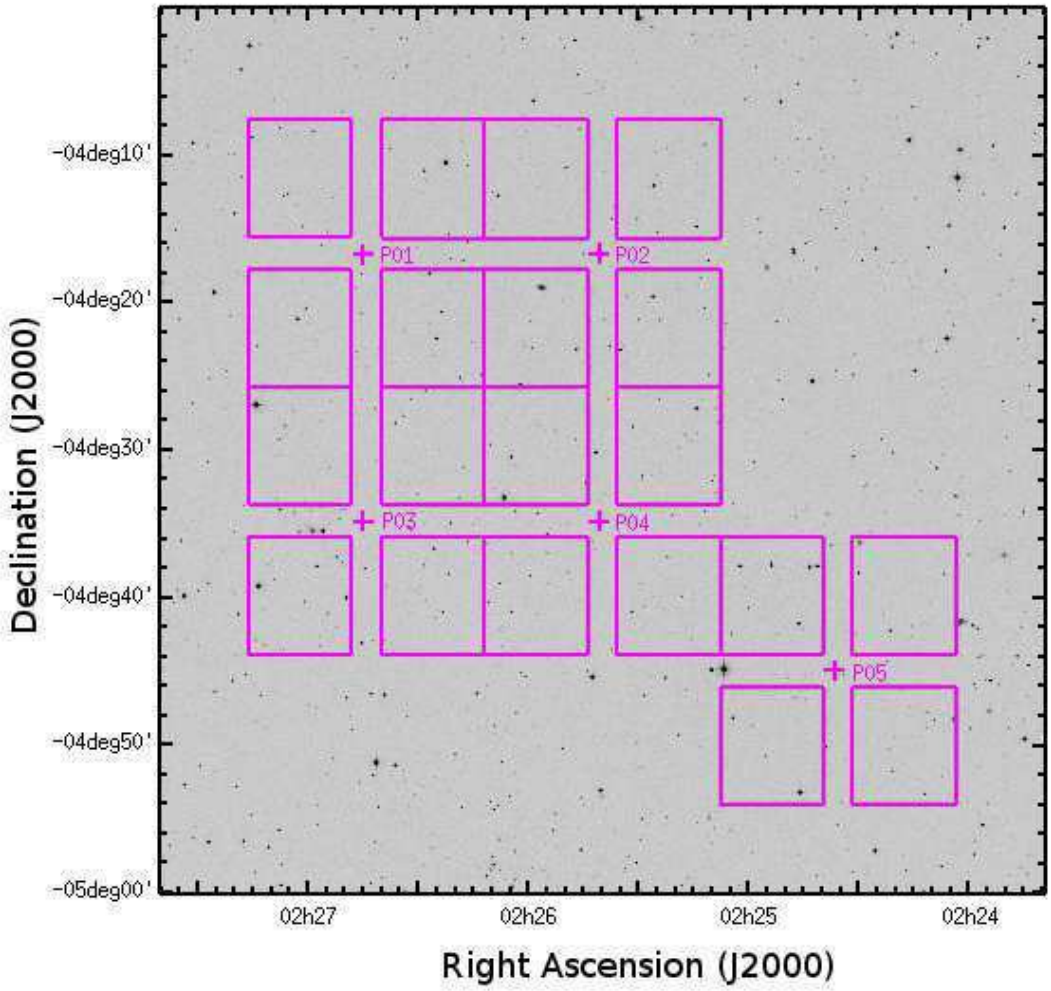


Fig. 3. Layout of the observed 5 VUDS VIMOS pointings in the VVDS-02h field. The center of each VIMOS pointing is identified by a cross with the pointing number (see Table 3), while the imprint of the 4 quadrants is in magenta. The size of the image is 1×1 deg^2. All of the VIMOS pointings are covered by CFHTLS visible photometry and WIRDS near infrared photometry.

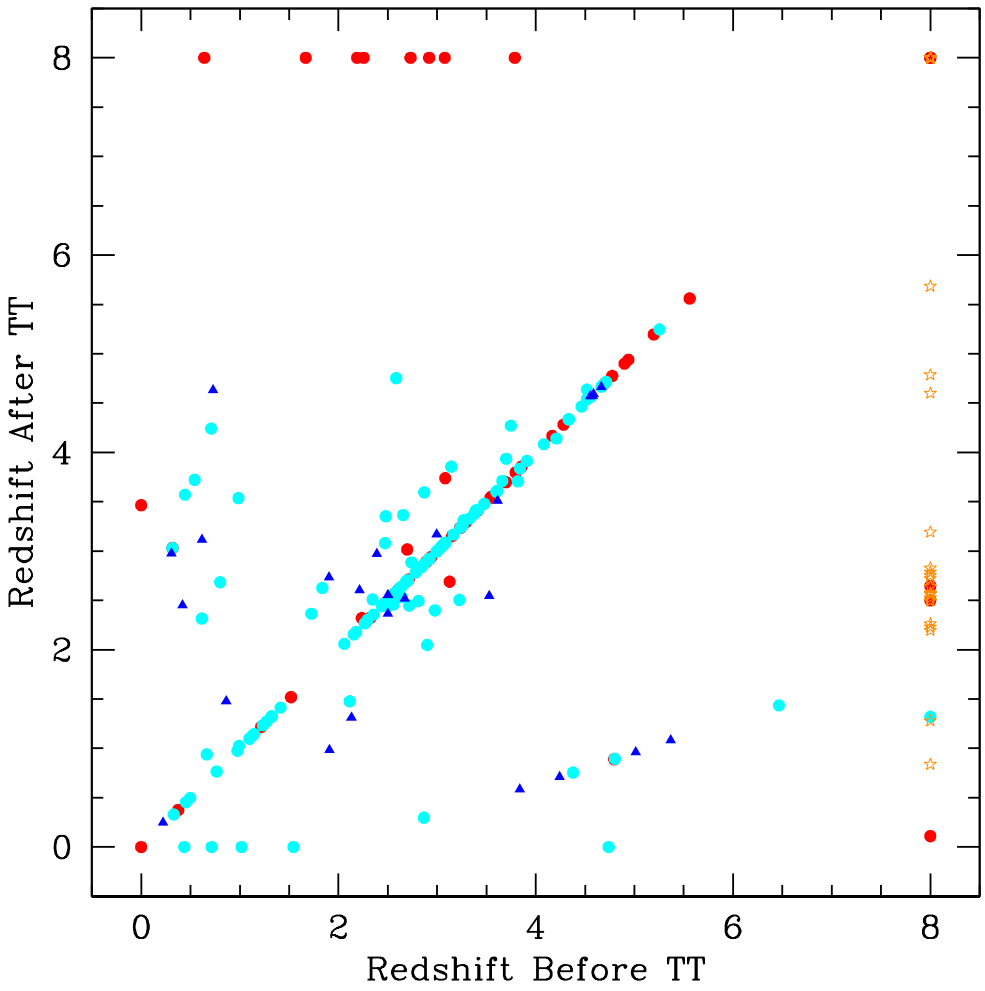


Fig. 4. Comparison between VUDS spectroscopic redshifts before and after the *Tiger Team* (TT) check of redshifts measured after a first measurement pass (see text). About 10% of all objects had a redshift change (6%) or a flag change (4%). Red circles represent objects for which the reliability flag was downgraded after the Tiger Team work; filled cyan circles are those objects with upgraded flags; and blue triangles are objects keeping the same flag. Objects which had undetermined redshifts (flag=0) before the TT work are placed at $z_{before} = 8$ and represented with orange starred symbols. Objects which have undetermined redshifts after the TT work (either keeping their original undetermined status or the TT work decided to downgrade them) are the red circles at $z_{after} = 8$.

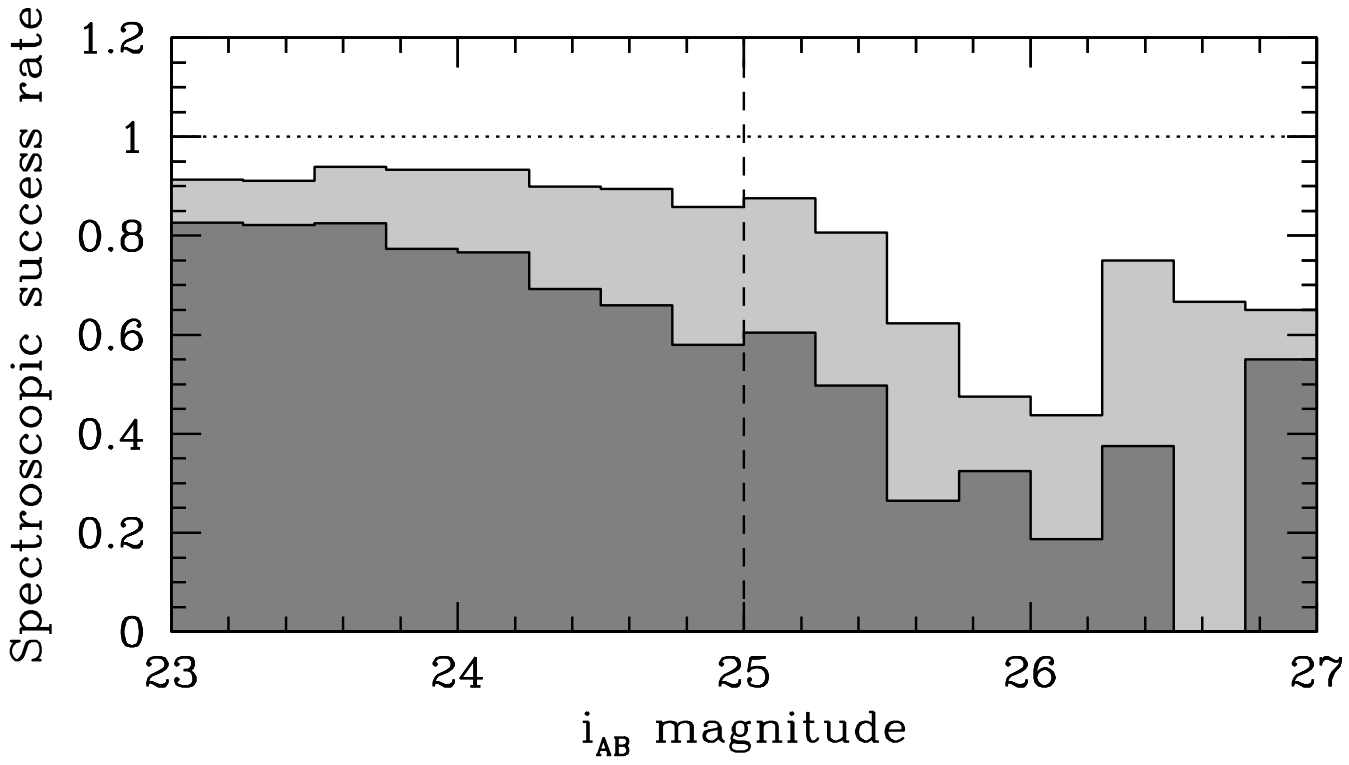


Fig. 5. Spectroscopic success rate versus i_{AB} magnitude, for all objects with a redshift measurement (light grey) and all objects with a > 75% reliable redshift measurement (flags 2, 3, 4, and 9; dark grey).

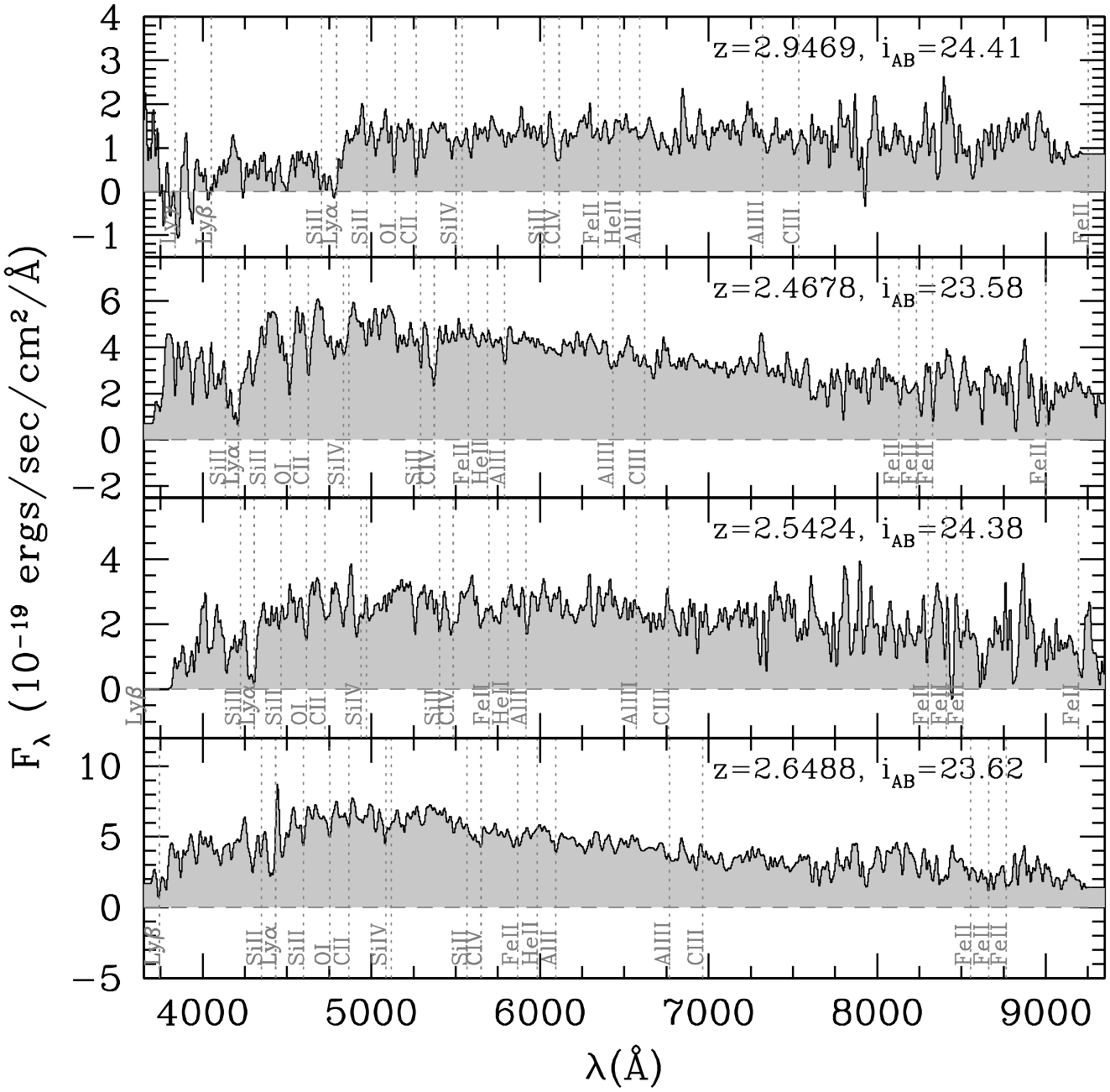


Fig. 6. Sample VUDS spectra in the range $2 < z < 3$.

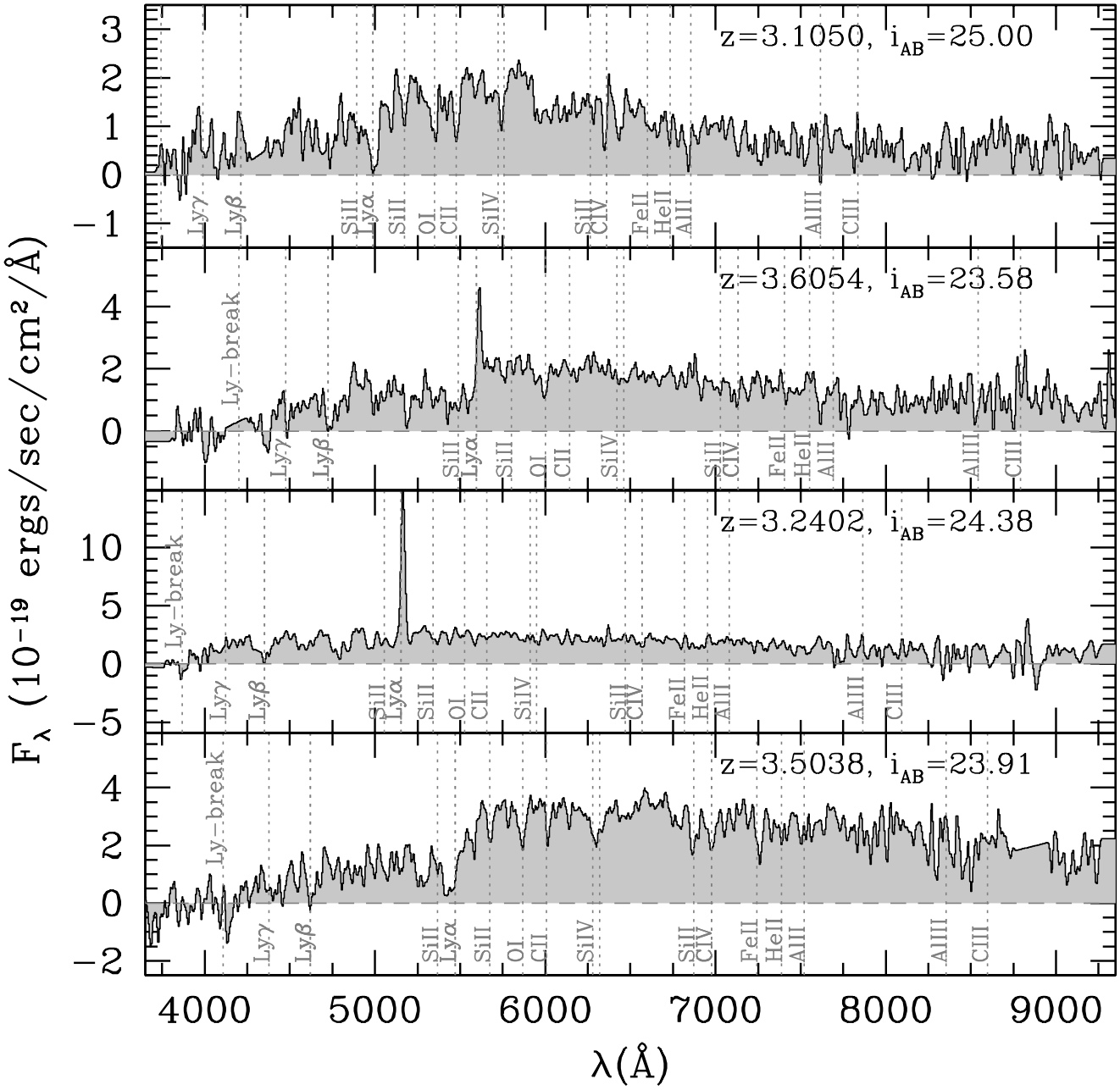


Fig. 7. Sample VUDS spectra in the range $3 < z < 4$.

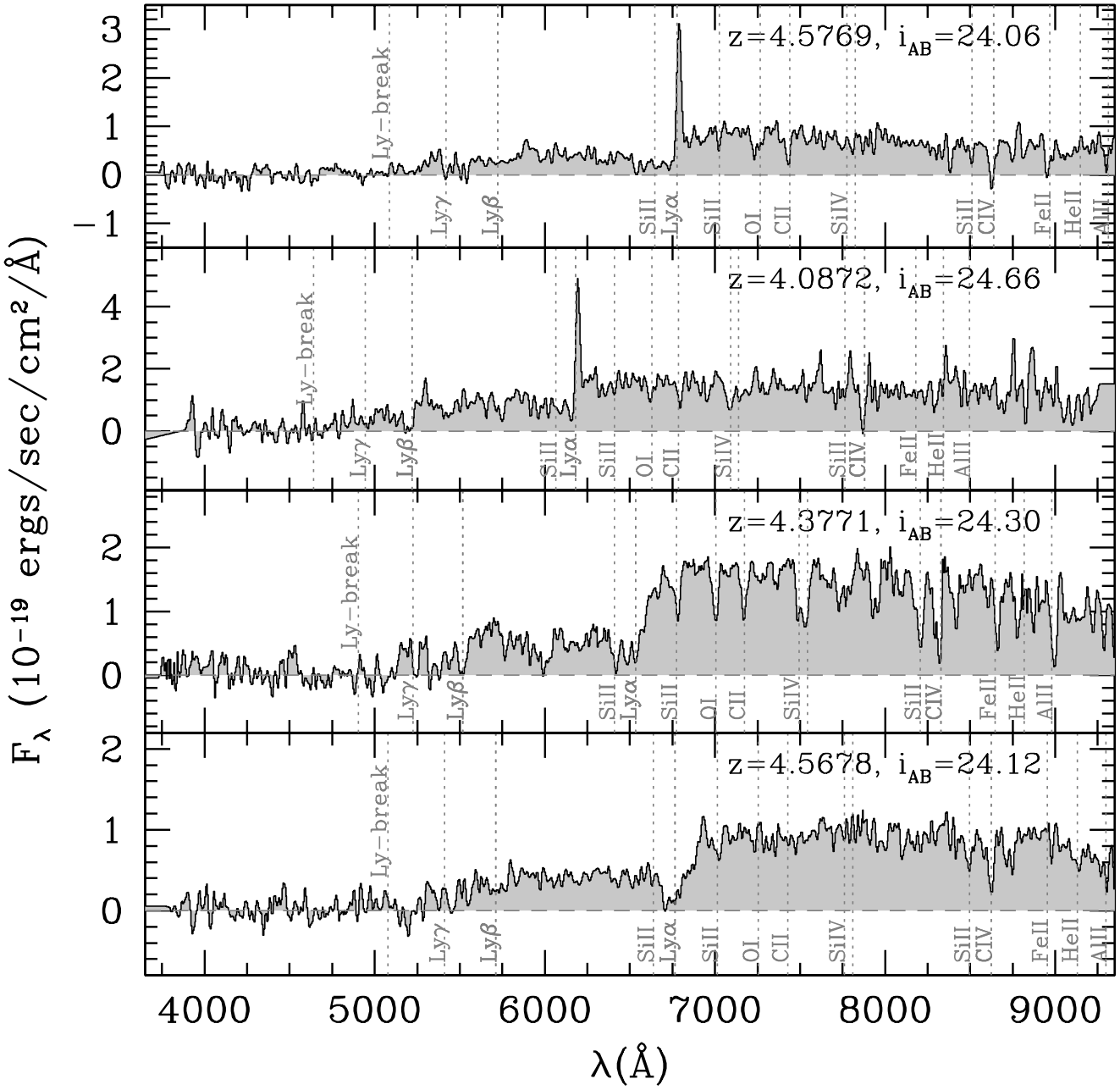


Fig. 8. Sample VUDS spectra in the range $4 < z < 5$.

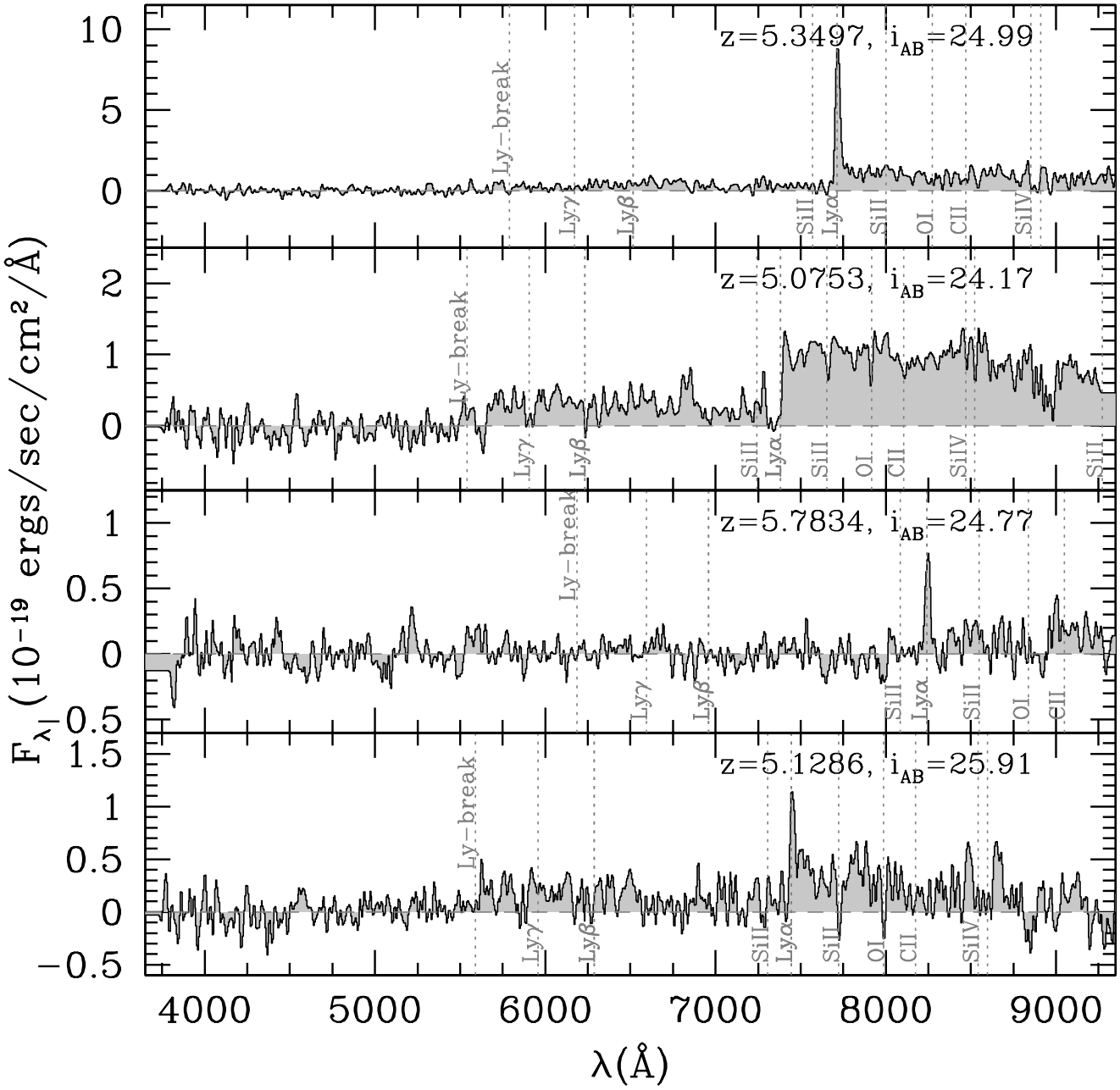


Fig. 9. Sample VUDS spectra in the range $5 < z < 6$.

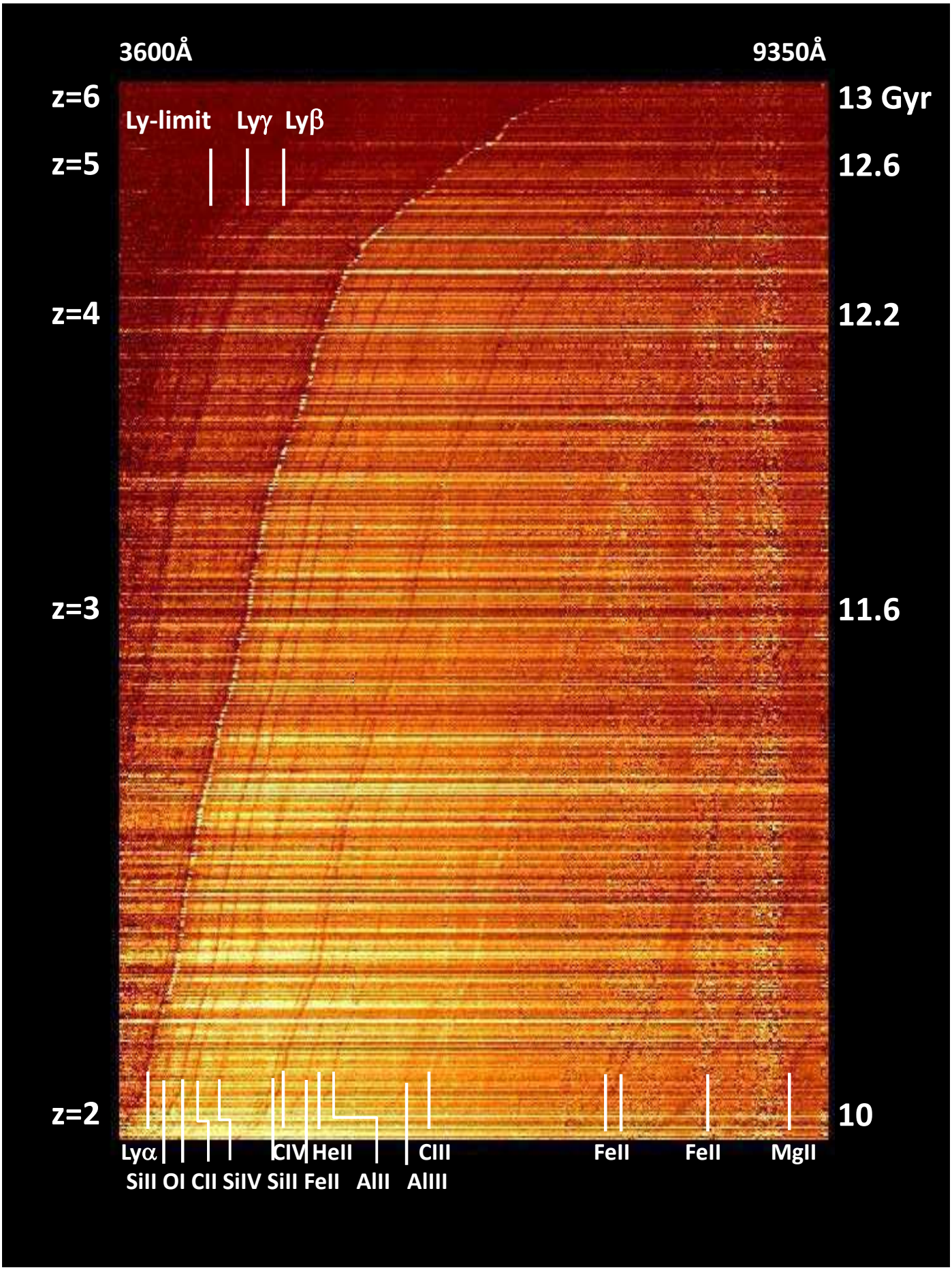


Fig. 10. Complete overview of the spectra obtained in the VUDS survey, with the wavelength going from 3650 to 9350Å on the X-axis. The redshift is increasing (in a non linear way) as indicated along the left Y-axis, with the corresponding look-back time indicate along the right Y-axis. The image is build with the spectra of all $2 < z < 4$ VUDS galaxies with flags 3 and 4, and all spectra with flags 2, 3 and 4 for $z > 4$, ordered one per image line by increasing redshift. All the main emission and absorption lines can be readily identified on this image, even faint ones, thanks to the increased contrast produced by the continuous display of spectra. The main spectral lines are identified on the top left (below Lyα) and at the bottom of the plot (above Lyα), as listed in Table 4. The vertical bands appearing at fixed wavelength in the red correspond to increased residual noise after subtraction of the strong atmospheric OH-bands.

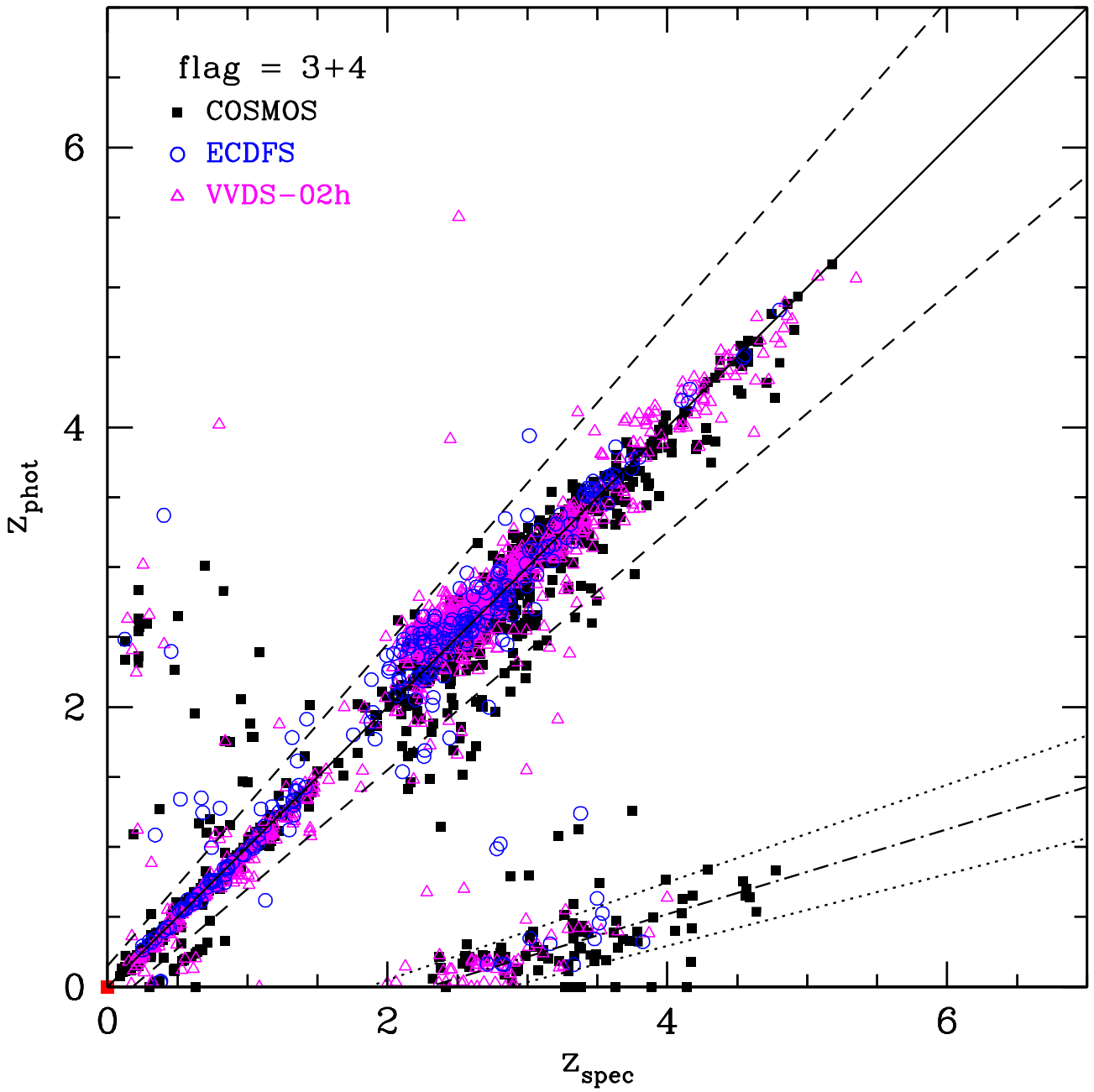


Fig. 11. Comparison between VUDS spectroscopic redshifts with spectroscopic flag 3 and 4 (95-100% reliable) and the SED-derived photometric redshifts using the Le Phare code. Galaxies in the three different fields are identified as filled squares (COSMOS field), open triangles (VVDS-02h), and open circles (ECDFS). The 1:1 equality relation is drawn as a continuous line with 15% errors expressed as $0.15 \times (1 + z)$ drawn as dashed lines. The known degeneracy of photometric redshifts between a Balmer–4000Å break and a Lyα–1215Å break is identified by the dot-dash line with 15% errors as the dotted lines.

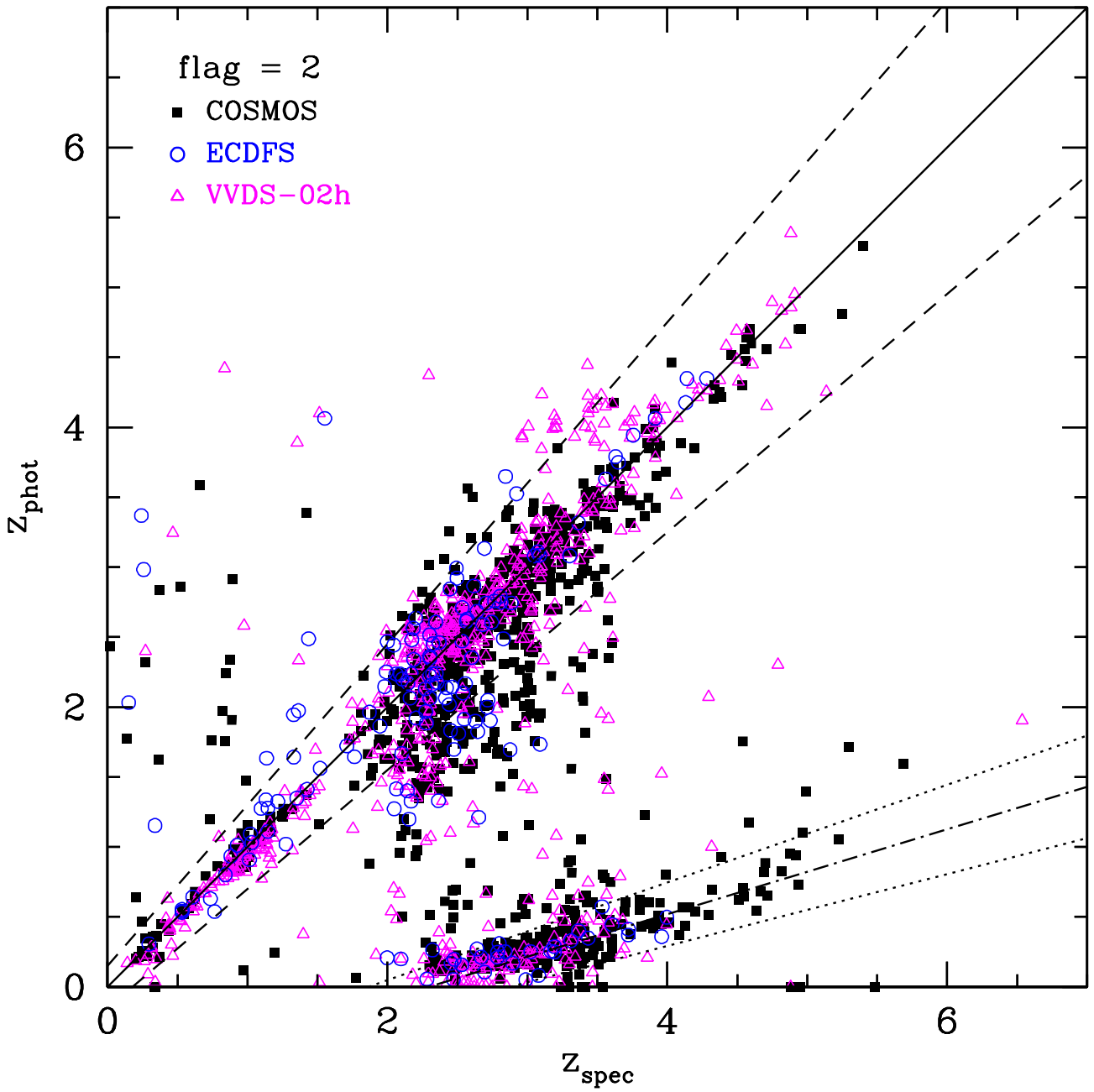


Fig. 12. Same as Figure 11 but for spectroscopic flags 2 (≃ 85% reliable).

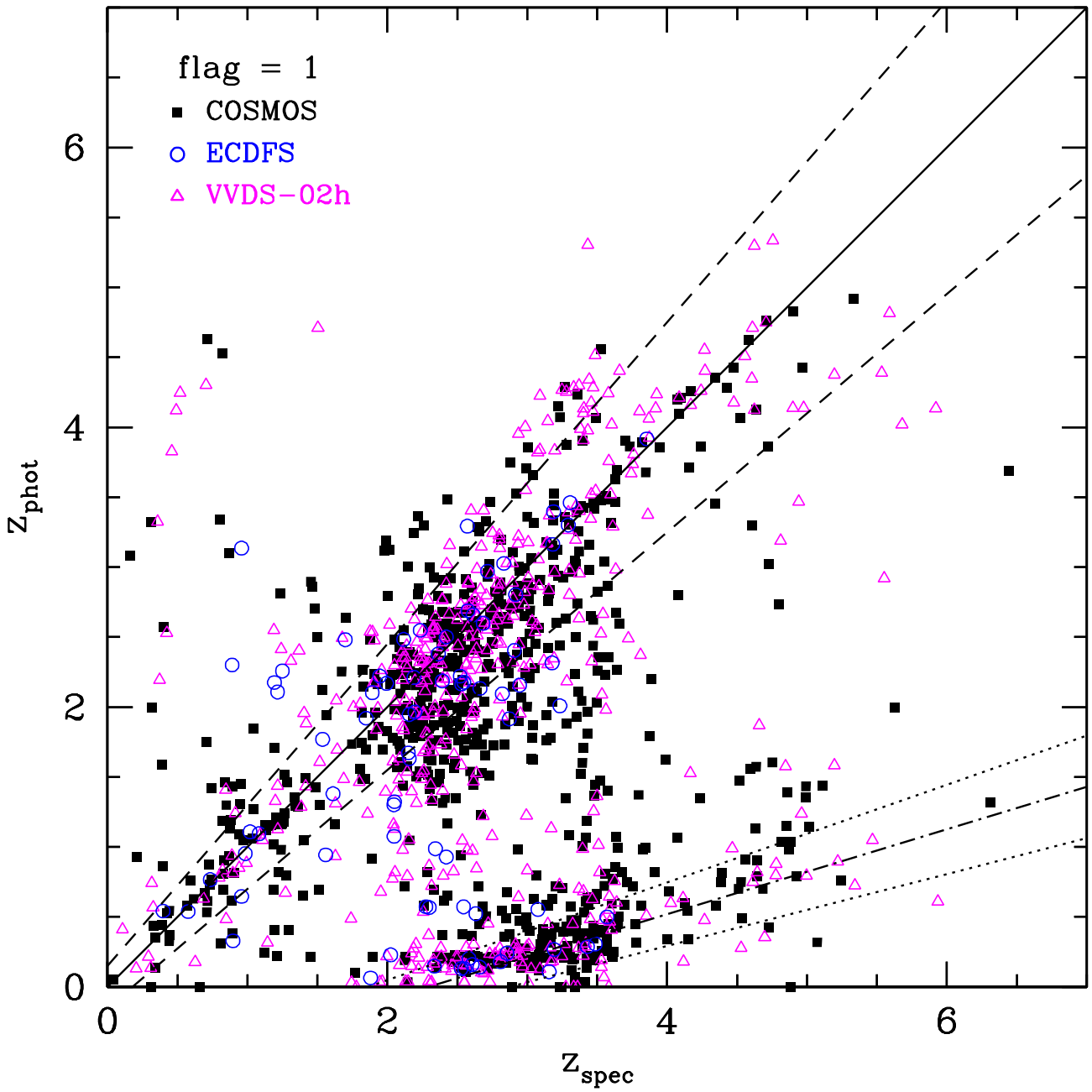


Fig. 13. Same as Figure 11 but for spectroscopic flags 1 (≃ 50 – 70% reliable).

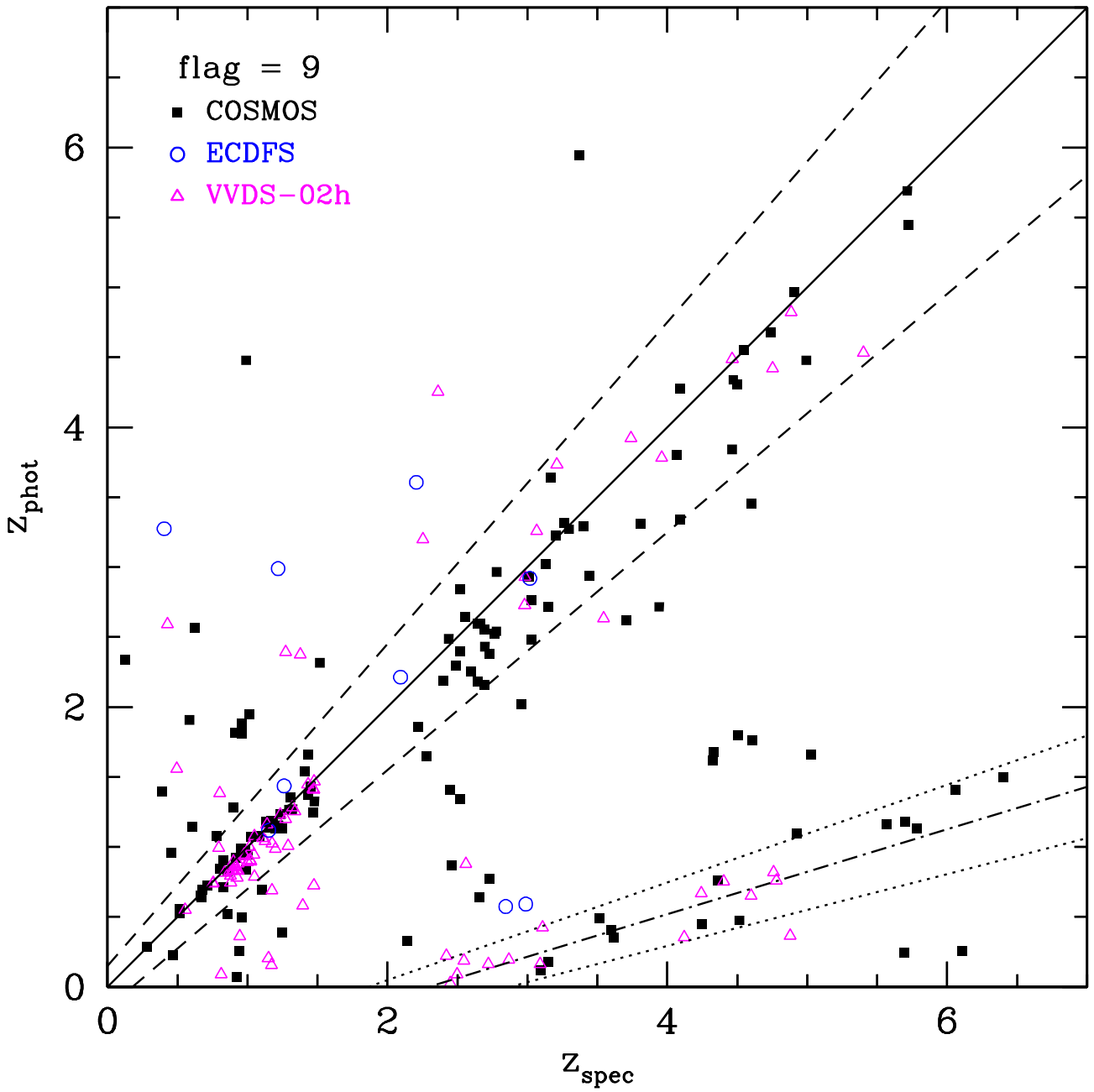


Fig. 14. Same as Figure 11 but for spectroscopic flags 9 (≃ 80% reliable).

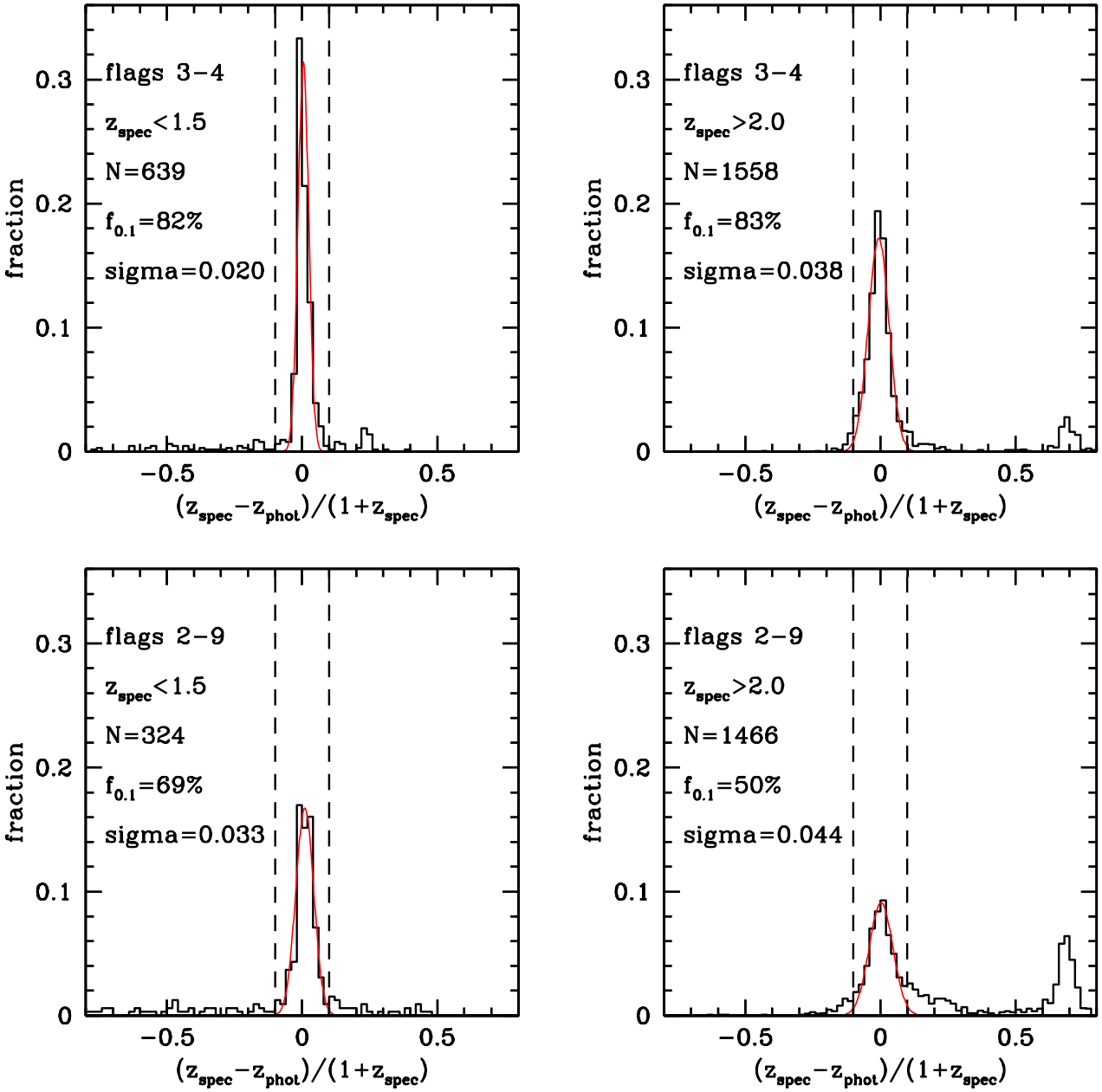


Fig. 15. Distribution of the difference between photometric redshifts and VUDS spectroscopic redshifts for flags 3+4 (top panels) and flags 2+9 (bottom panels), for redshifts $z < 1.5$ (left panels) and redshifts $z > 2$ (right panels). The number of galaxies is indicated in each panel together with the σ of the distribution, and the fraction of galaxies verifying $\delta z = (z_{spec} - z_{phot})/(1 + z_{spec}) \leq 0.1$. The distribution peaked at 0 is for the main sample, while a secondary peak appearing beyond $\delta z = 0.5$ is produced by the degeneracy between the D4000 and the 1215Å continuum breaks (see text).

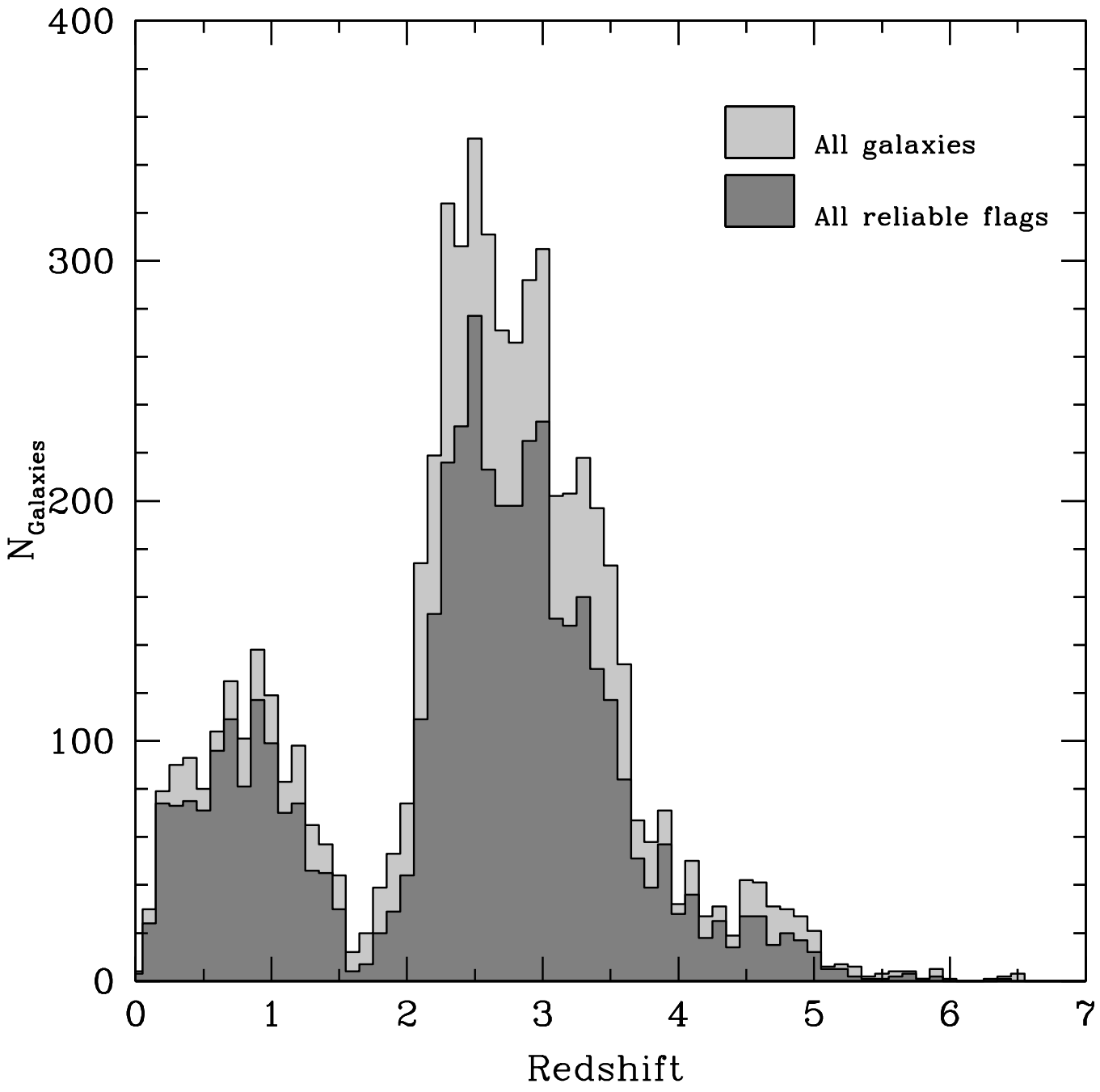


Fig. 16. The current redshift distribution from 6057 galaxies already processed in the VUDS survey, for all objects with a redshift measurement (light grey) and all objects with a > 80% reliable redshift measurement (flags 2, 3, 4, and 9; dark grey).

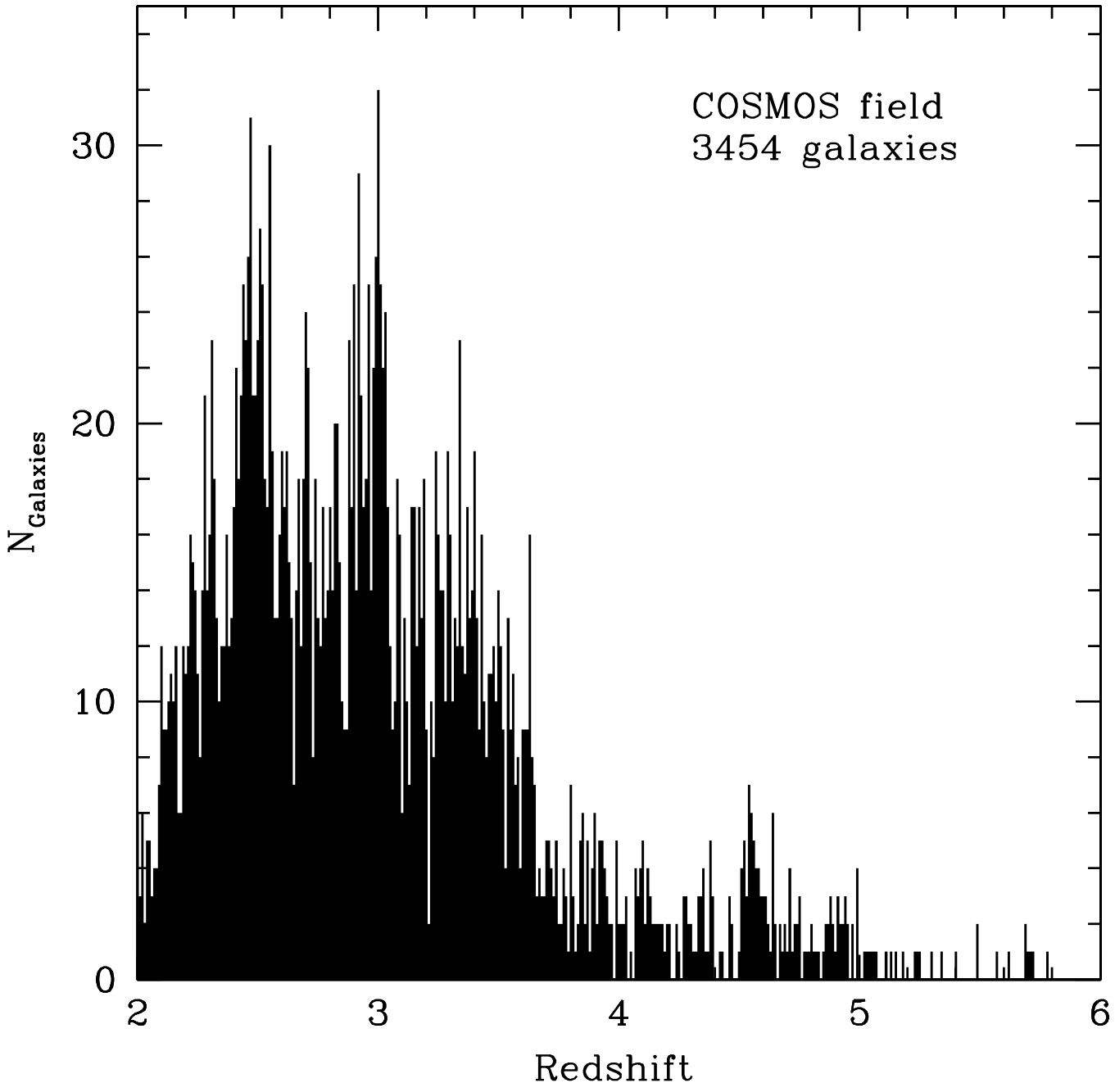


Fig. 17. The current redshift distribution from 3495 galaxies already identified in the COSMOS field by the VUDS survey (all objects with a redshift measurement are used).

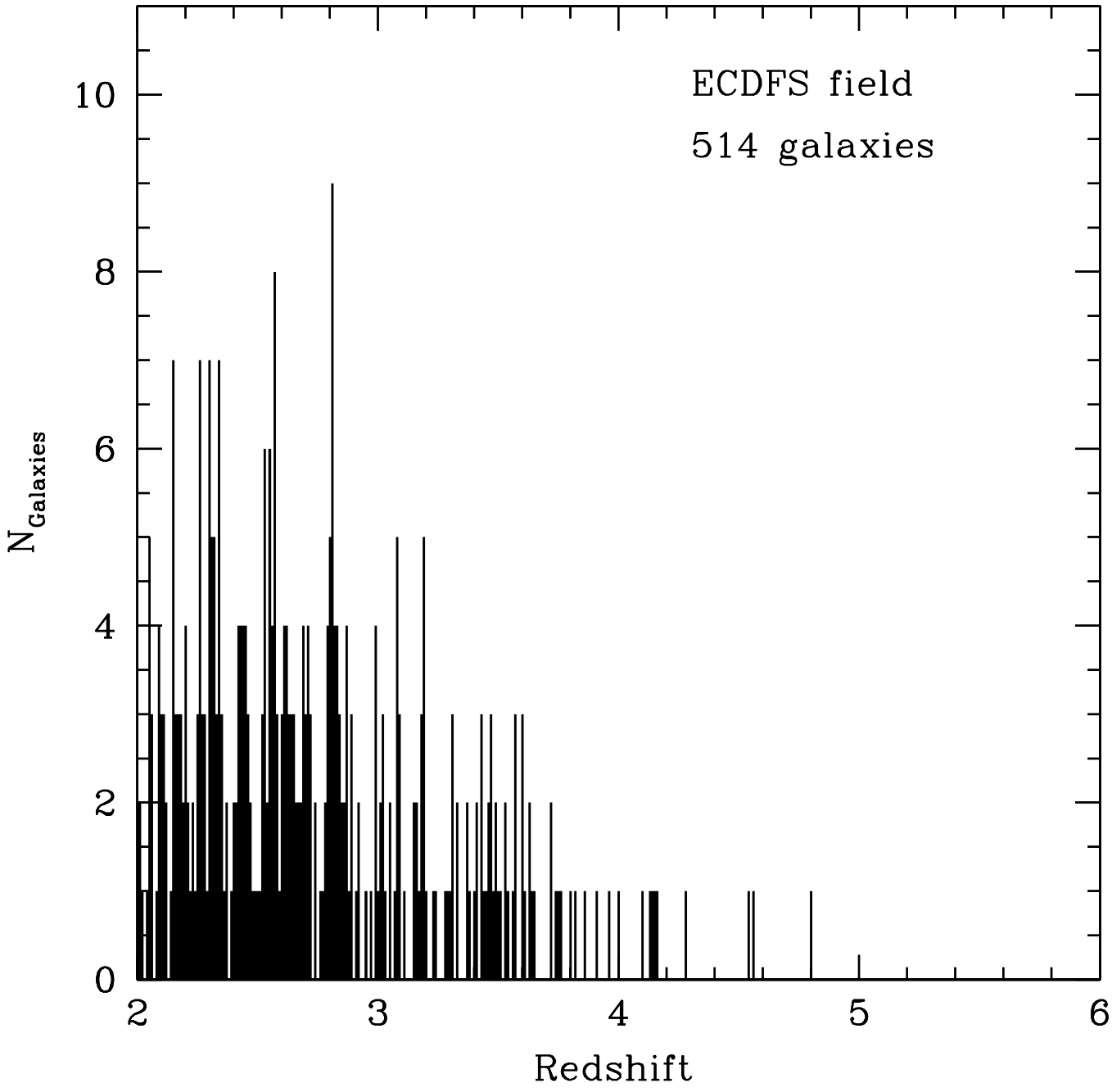


Fig. 18. The current redshift distribution from 514 galaxies already identified in the ECDFS field by the VUDS survey (all objects with a redshift measurement are used).

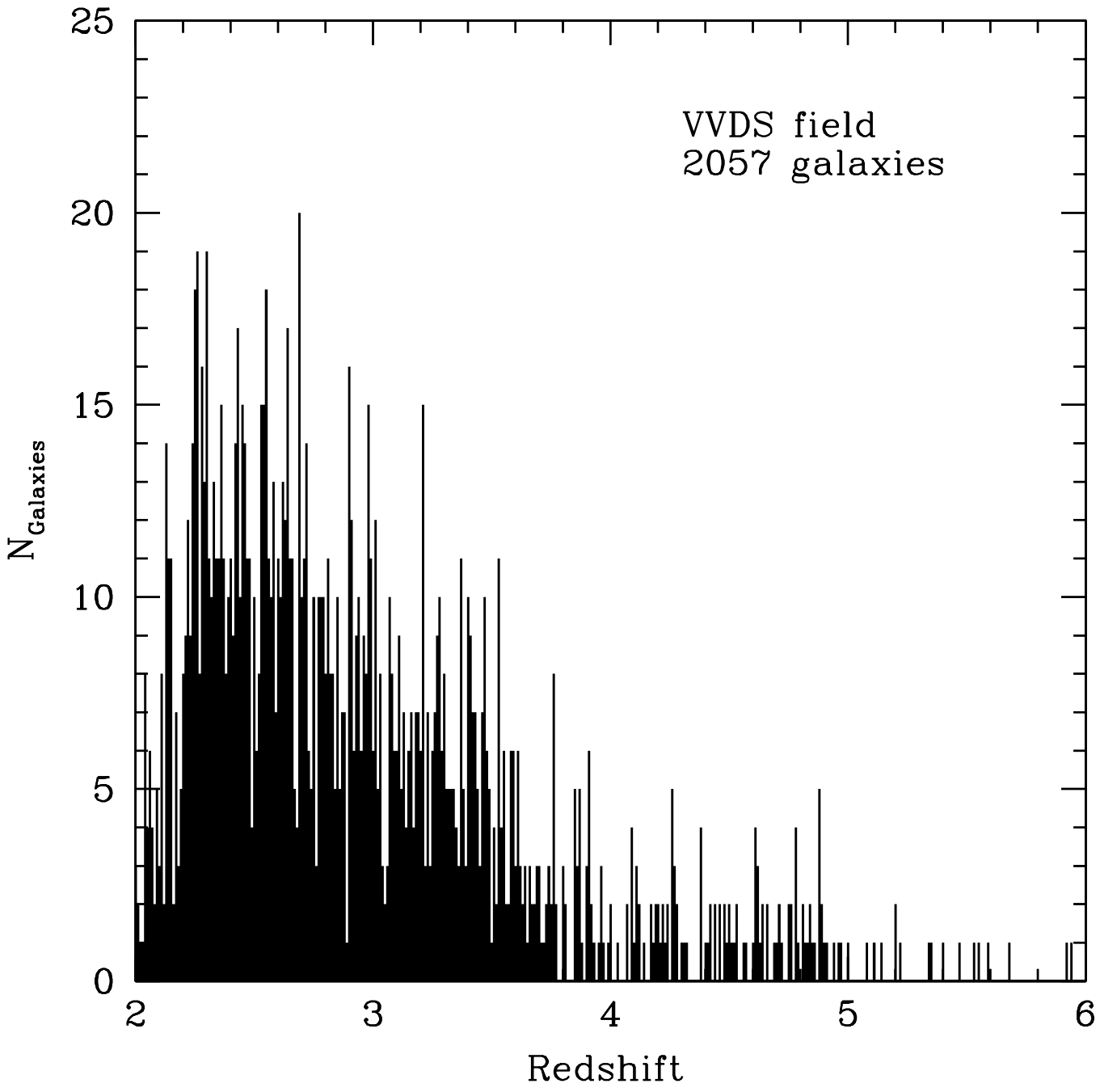


Fig. 19. The current redshift distribution from 2059 galaxies already identified in the VVDS-02h field by the VUDS survey (all objects with a redshift measurement are used).

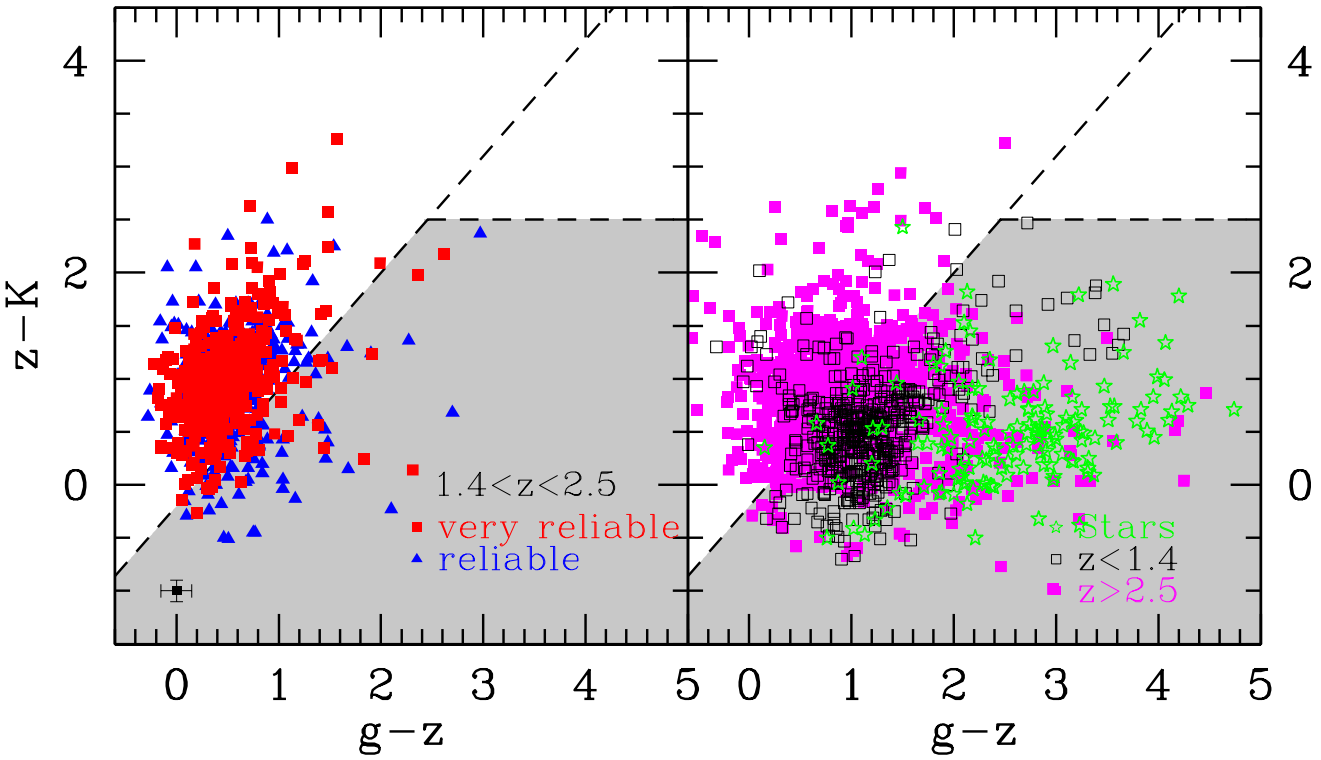


Fig. 20. (left panel) (g-z) vs. (z-K) color-color diagram for VUDS galaxies with $1.4 < z < 2.5$ (flag 3+4: red squares, flag 2: blue triangles). (right panel) same for galaxies either with $z < 1.4$ (open squares) or $z > 2.5$ (magenta filled squares). Star are represented by green starred symbols. The *gzK* selection would have selected 91% of the VUDS galaxies with $1.4 < z < 2.5$, but with a high level of contamination as 58% of galaxies in the selection area of the *gzK* are outside the redshift range $1.4 < z < 2.5$, coming mostly from galaxies at $z > 2.5$.

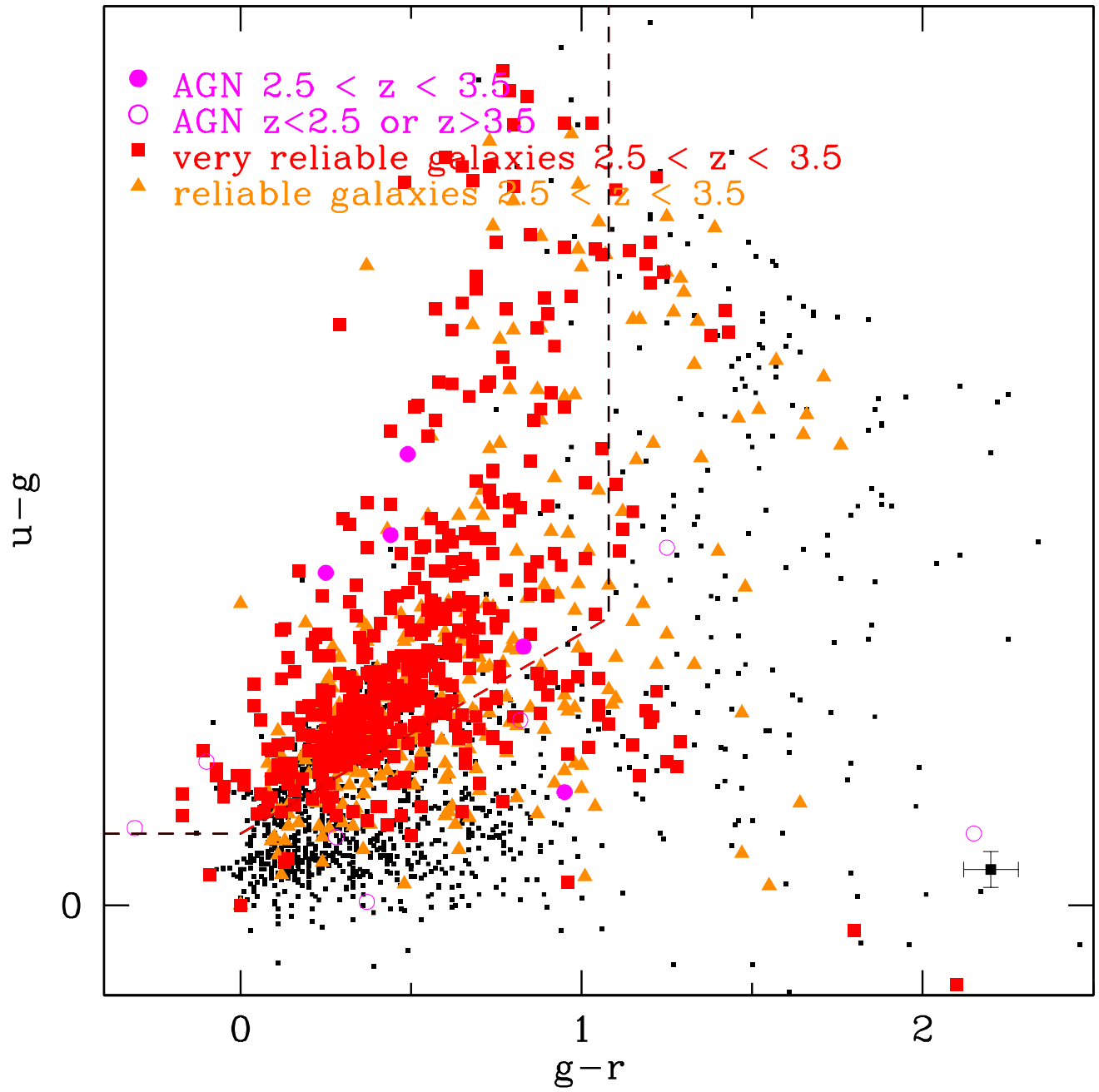


Fig. 21. (left panel) (u-g) vs. (g-r) color-color diagram for VUDS galaxies with $2.5 < z < 3.5$ (flag 3+4: red squares, flag 2: blue triangles). (right panel) same for galaxies either with $z < 2.5$ (open squares) or $z > 3.5$ (magenta filled squares). The *ugr* selection would have selected 80% of the VUDS galaxies with $2.5 < z < 3.5$, but with a high level of contamination as 40% of galaxies in the selection area of the *ugr* are outside the redshift range $2.5 < z < 3.5$.

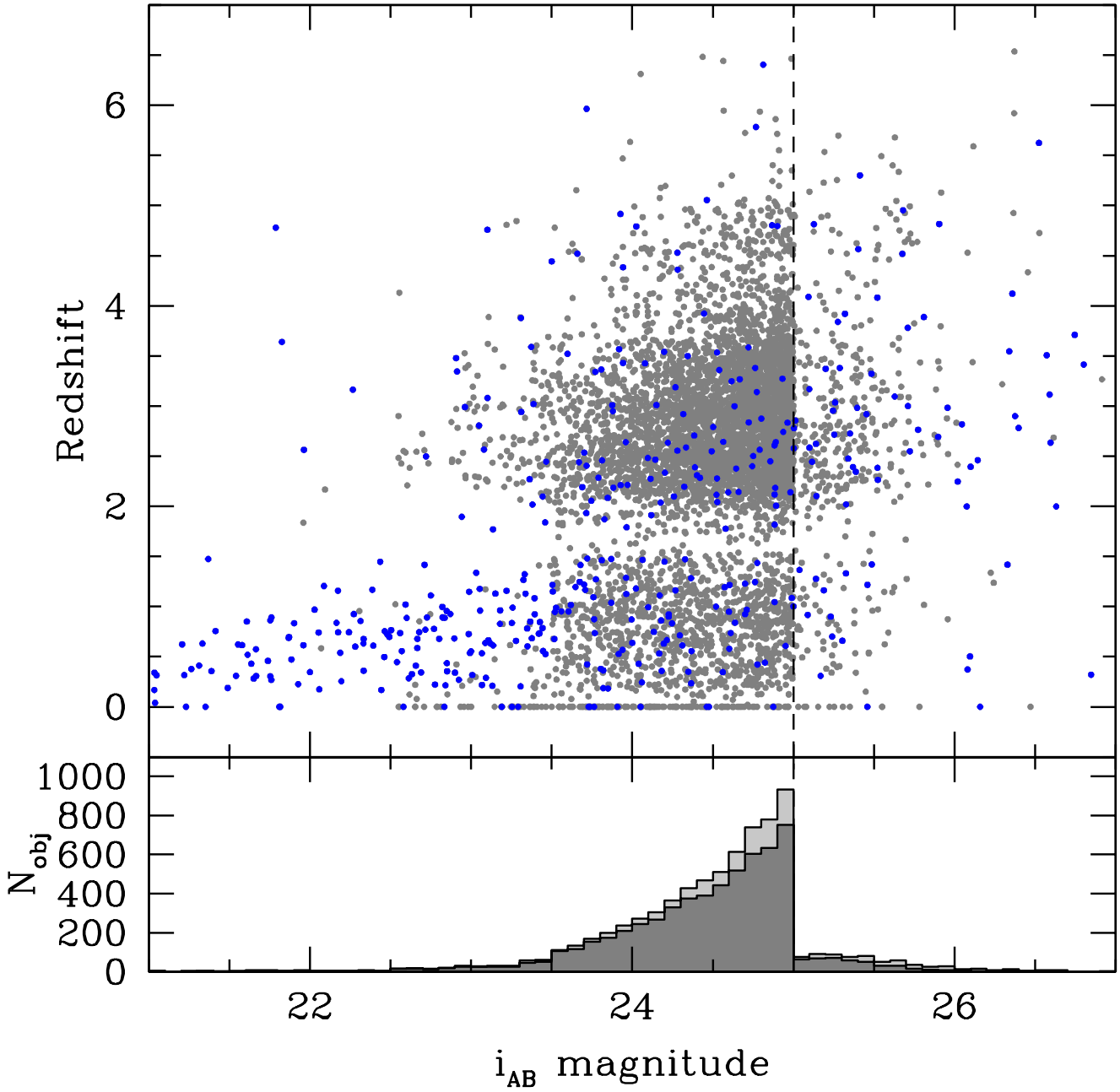


Fig. 22. (top panel) i_{AB} magnitude - redshift distribution for the full VUDS sample (grey points) and for serendipitous objects in the slits (blue points). (bottom panel) Distribution of i_{AB} magnitudes in the VUDS survey for all objects observed (light grey), and for those with a redshift measurement (any non 0 flag, dark grey). The $i_{AB} = 25$ imposed on the sample selected by photometric redshifts is indicated. The fainter objects are pre-selected from one of the other color or SED criteria.

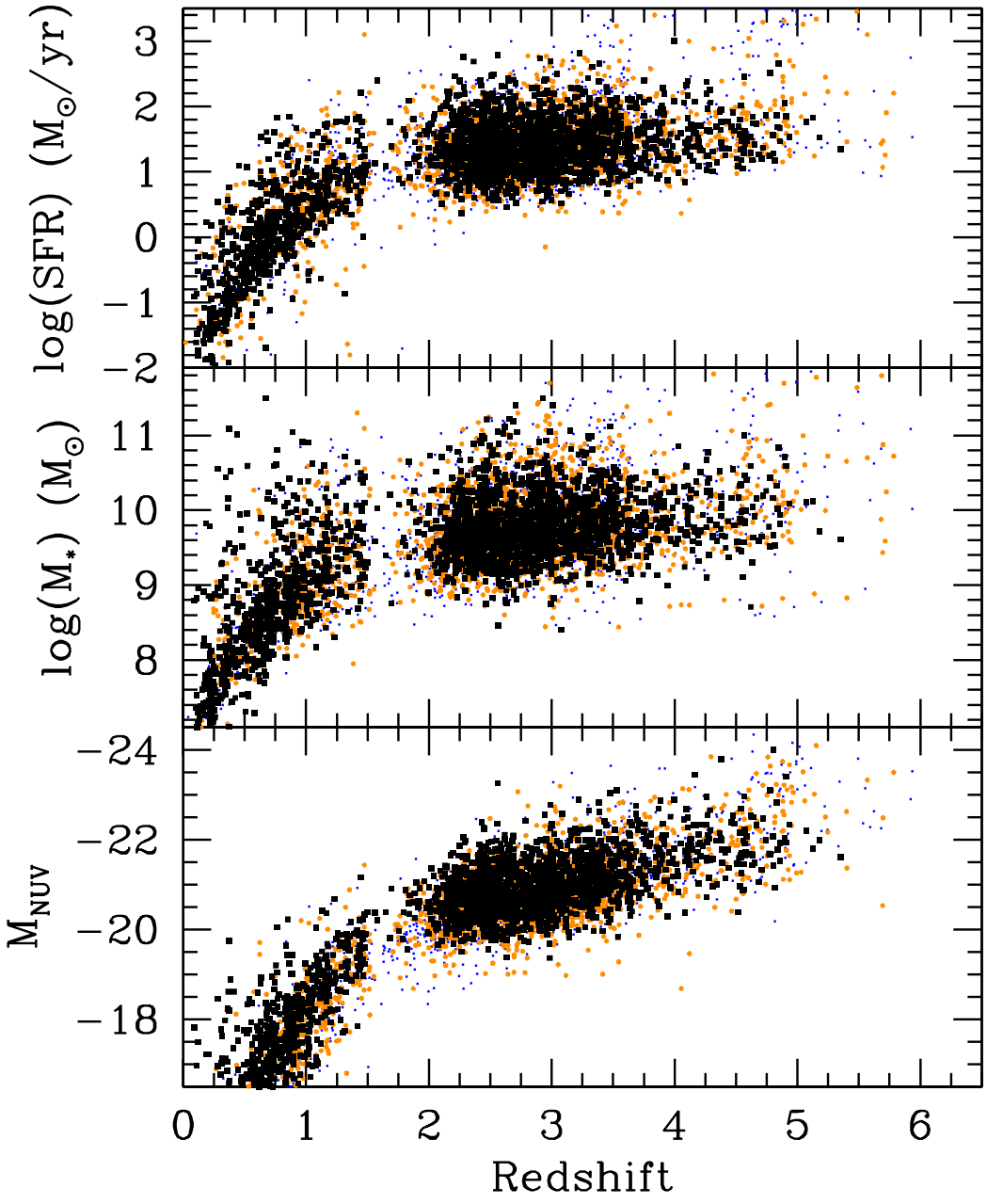


Fig. 23. The distribution of absolute U-band magnitudes (bottom), stellar masses (center), and star formation rate (top). Stellar masses and star formation rates are derived from template SED fitting at the spectroscopic redshift (see text). Black squares are for galaxies with reliability flags 3 and 4, orange filled circles are for flag 2, and blue dots are for flag 1.

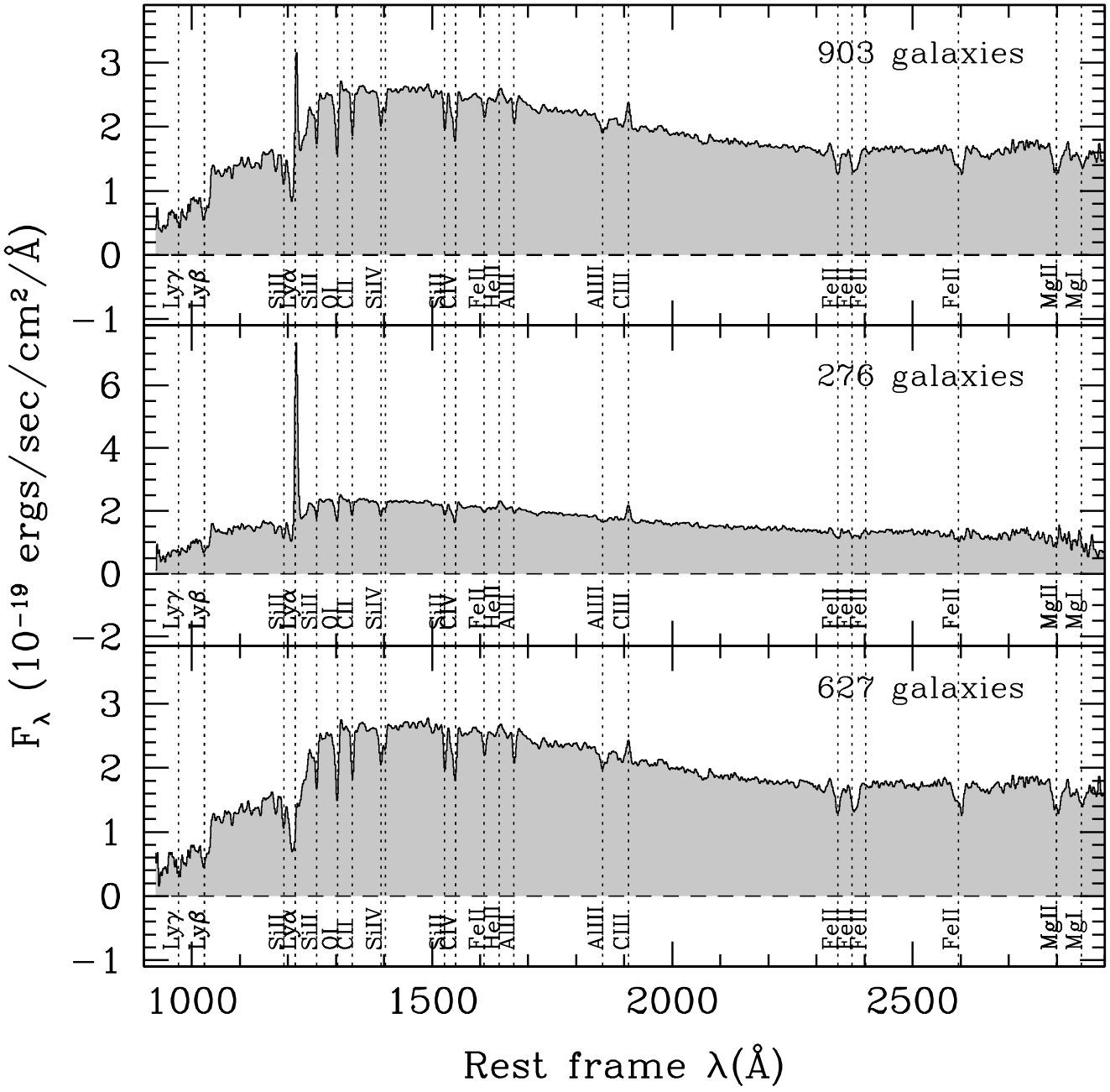


Fig. 24. Average rest-frame spectra (F_λ) of galaxies with flags 3 and 4 in VUDS with $2 \leq z \leq 3$: *(Top)*: stack of all galaxy spectra; *(Middle)*: stack of galaxies with Ly-α in emission; *(Bottom)*: stack of galaxies with Ly-α in absorption.

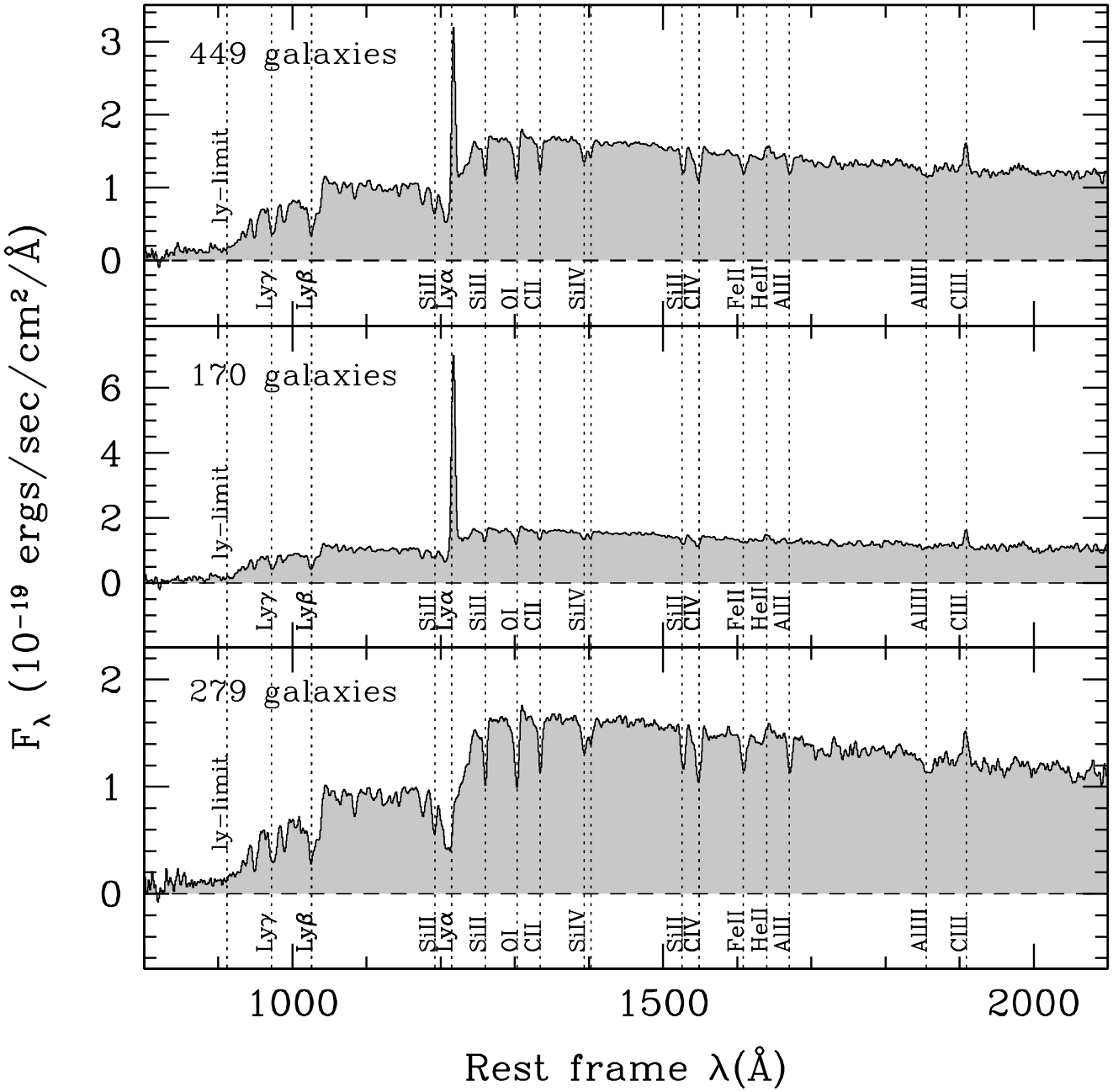


Fig. 25. Average rest-frame spectra (F_λ) of galaxies with flags 3 and 4 in VUDS with $3 \leq z \leq 4$: *(Top)*: stack of all galaxy spectra; *(Middle)*: stack of galaxies with Ly-α in emission; *(Bottom)*: stack of galaxies with Ly-α in absorption.

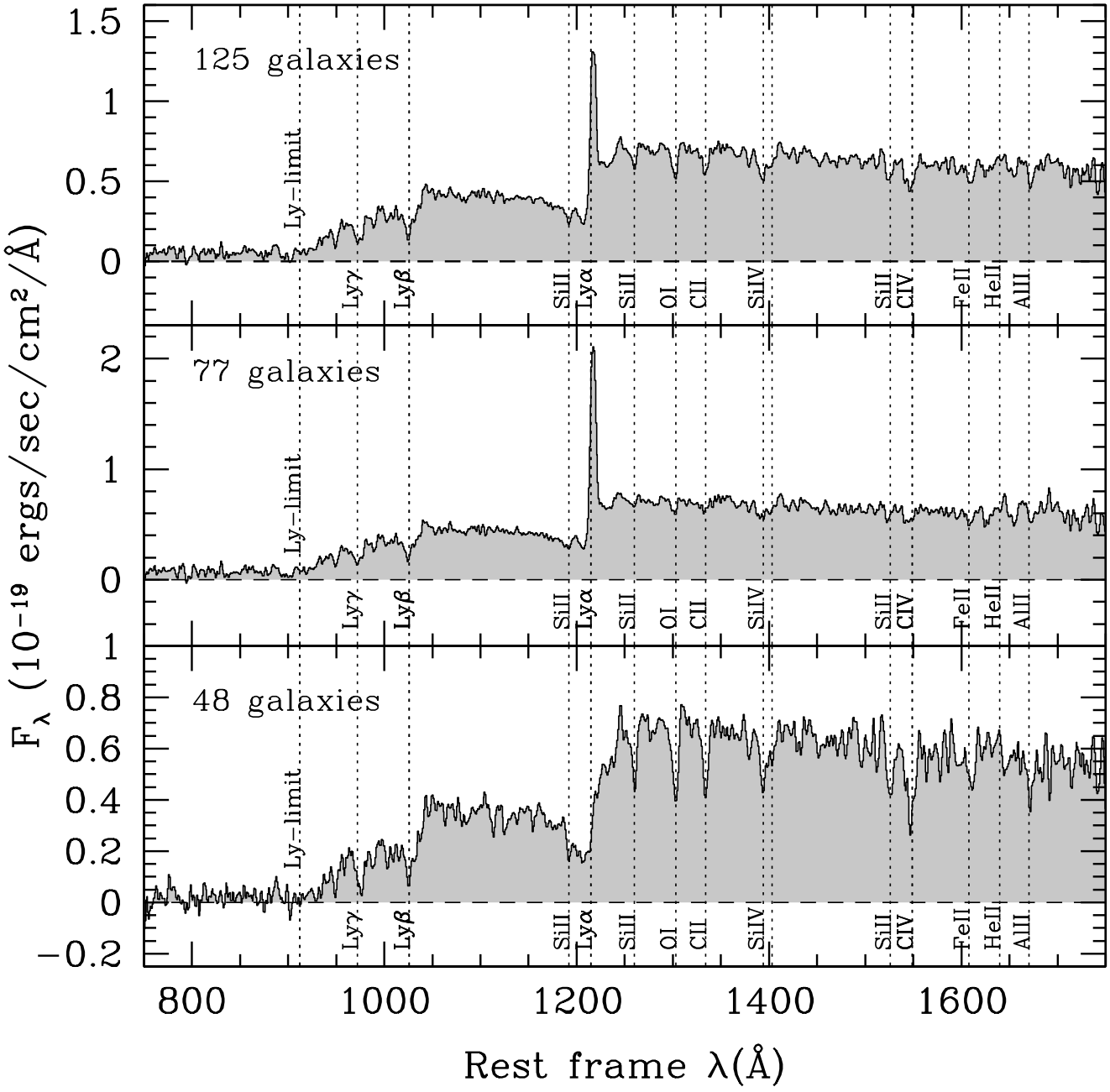


Fig. 26. Average rest-frame spectra (F_λ) of galaxies with flags 2, 3 and 4 in VUDS with $4 \leq z \leq 4.7$: *(Top)*: stack of all galaxy spectra; *(Middle)*: stack of galaxies with Ly-α in emission; *(Bottom)*: stack of galaxies with Ly-α in absorption.

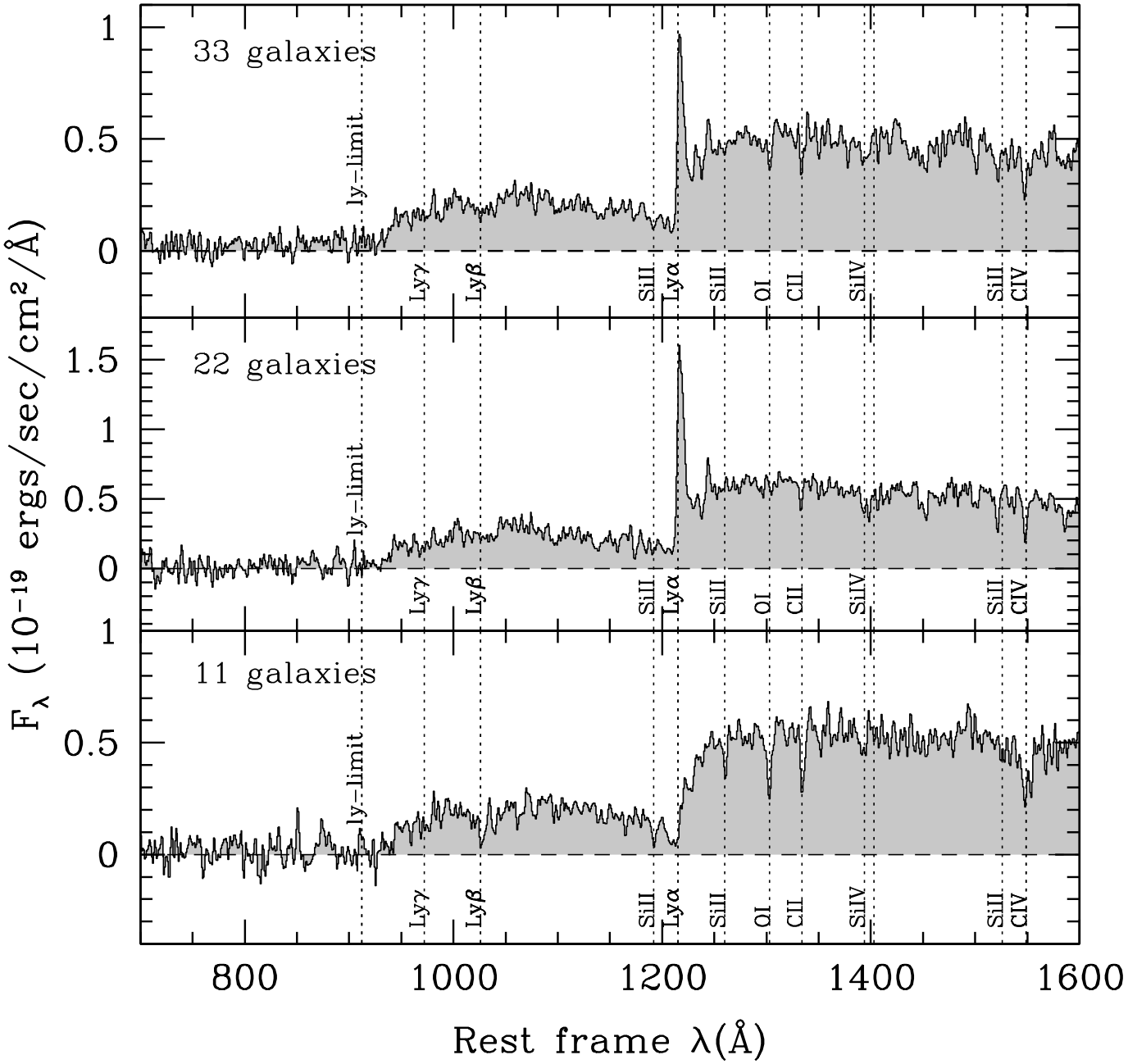


Fig. 27. Average rest-frame spectra (F_λ) of galaxies with flags 2, 3 and 4 in VUDS with $4.7 \leq z \leq 5.3$: *(Top)*: stack of all galaxy spectra; *(Middle)*: stack of galaxies with Ly-α in emission; *(Bottom)*: stack of galaxies with Ly-α in absorption.

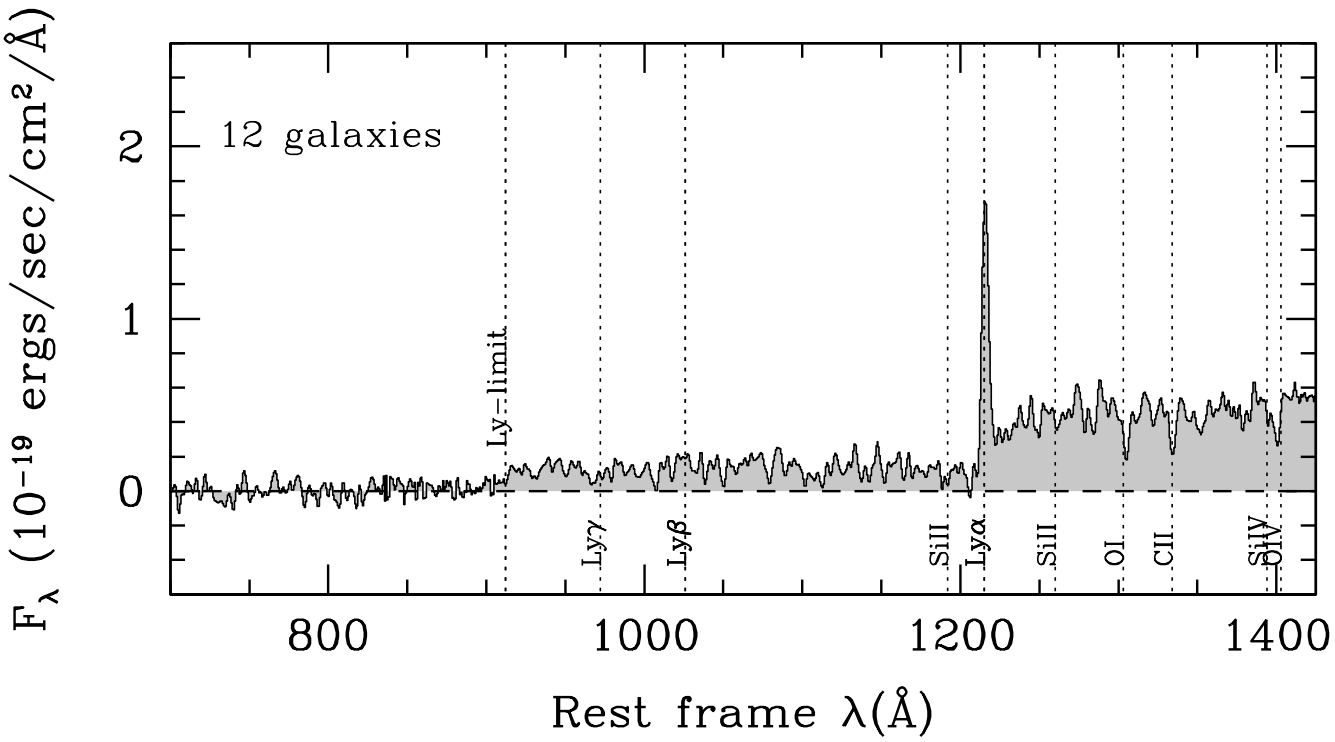


Fig. 28. Average rest-frame spectra (F_λ) of galaxies with flags 2, 3 and 4 in VUDS with $5.3 \leq z \leq 6.5$.

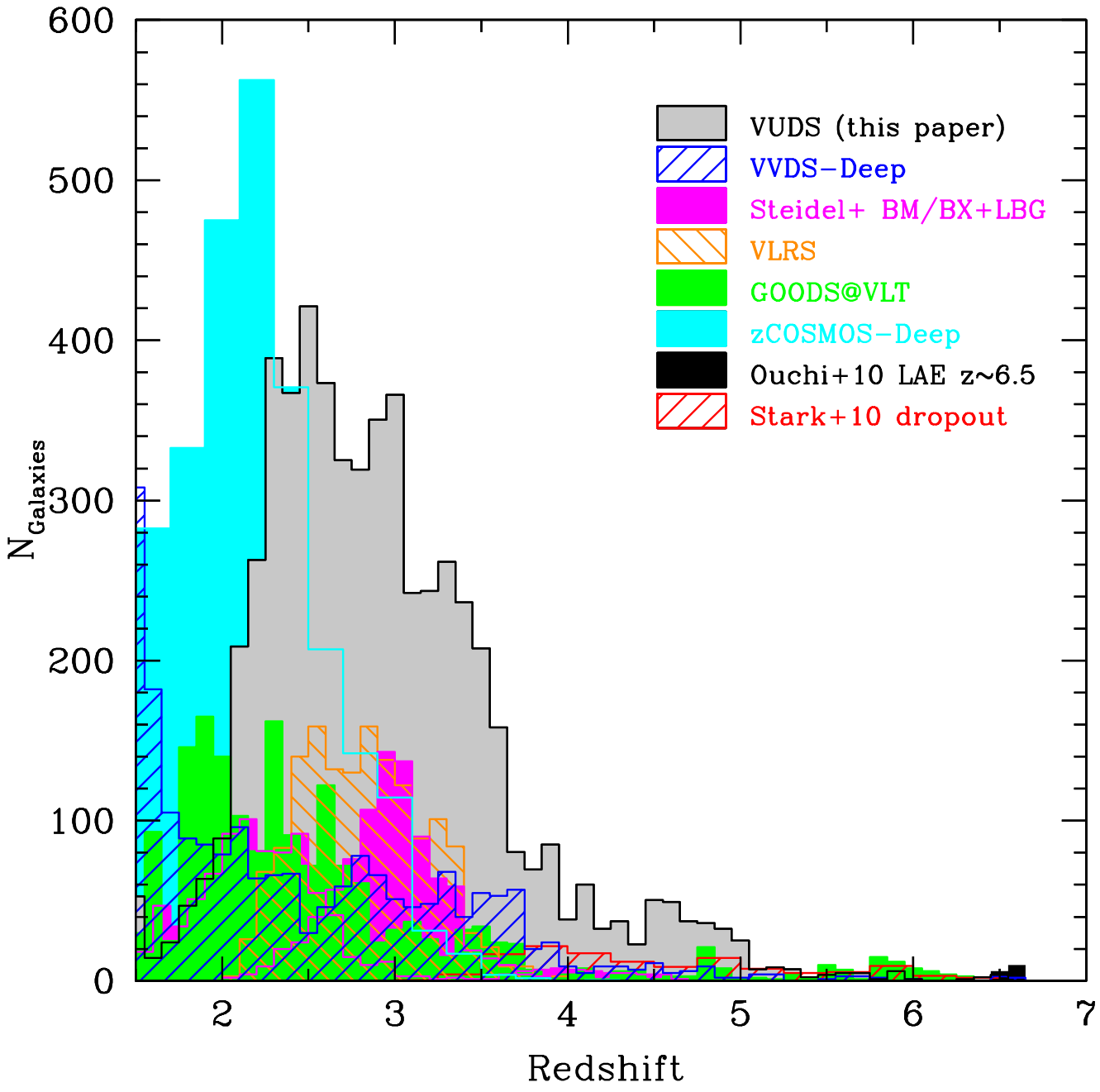


Fig. 29. Comparison of the redshift distribution of the VUDS survey with other published spectroscopic surveys at $z > 2$ (as listed in Table 5). The VUDS counts from existing measurements have been scaled by 1.2 to account for the data not yet processed at the time of this writing.

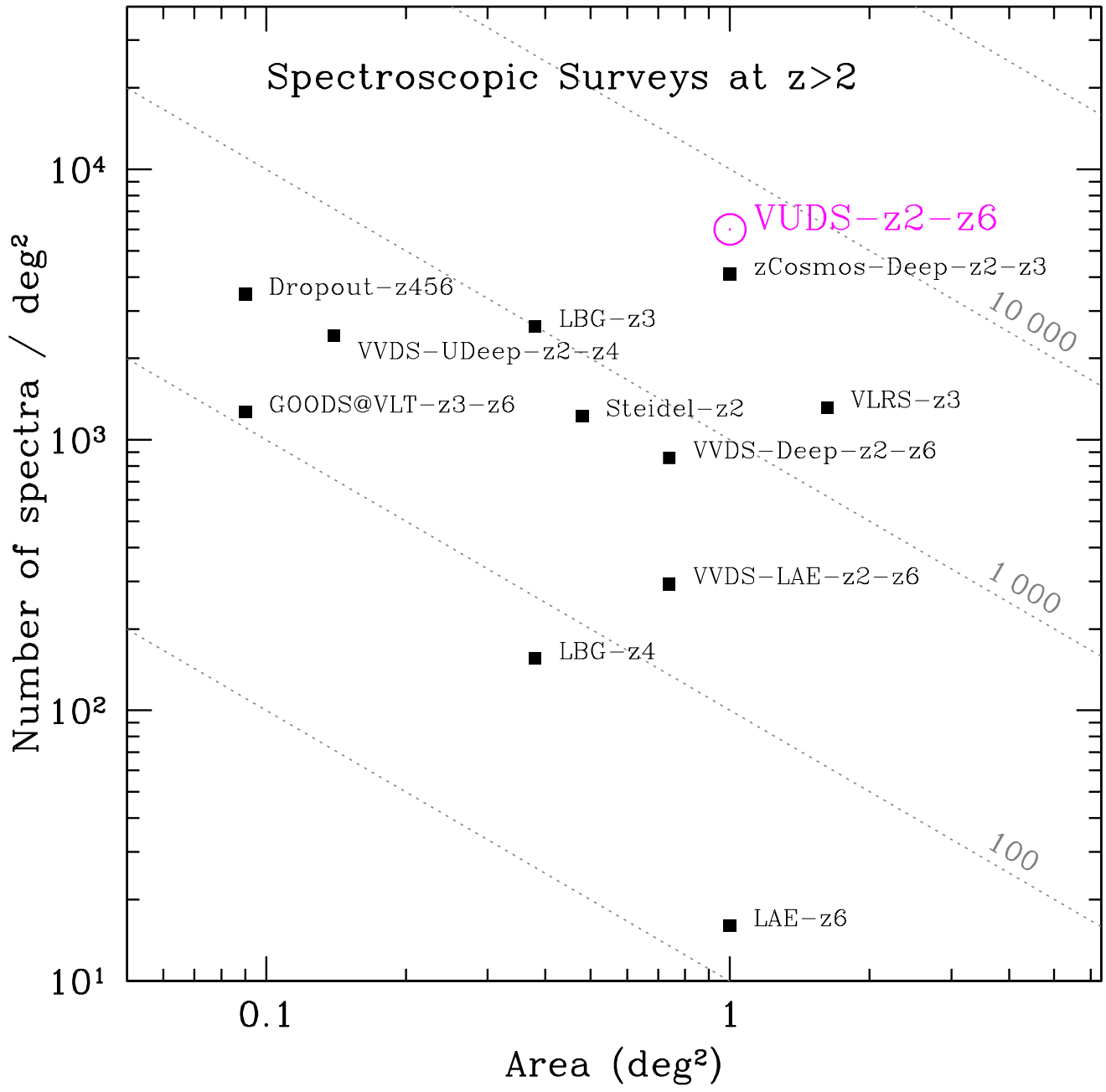


Fig. 30. Comparison of the number density of spectra in the VUDS survey with other published spectroscopic surveys at $z > 2$ (as listed in Table 5). The VUDS counts have been scaled by 1.2 to account for the data not yet processed at the time of this writing. Dashed lines are for surveys with a total number of spectra as indicated.

Astronomy & Astrophysics manuscript no. vuds˙CF˙final
November 14, 2014
© ESO 2014

The evolution of clustering length, large-scale bias and host halo mass at 2<z<5 in the VIMOS Ultra Deep Survey (VUDS)

A. Durkalec[1,⋆], O. Le Fèvre[1], A. Pollo[19,20], S. de la Torre[1], P. Cassata[1,18], B. Garilli[3], V. Le Brun[1], B.C. Lemaux[1], D. Maccagni[3], L. Pentericci[4], L.A.M. Tasca[1], R. Thomas[1], E. Vanzella[2], G. Zamorani[2], E. Zucca[2], R. Amorín[4], S. Bardelli[2], L.P. Cassarà[3], M. Castellano[4], A. Cimatti[5], O. Cucciati[5,2], A. Fontana[4], M. Giavalisco[13], A. Grazian[4], N. P. Hathi[1], O. Ilbert[1], S. Paltani[9], B. Ribeiro[1], D. Schaerer[10,8], M. Scodeggio[3], V. Sommariva[5,4], M. Talia[5], L. Tresse[1], D. Vergani[6,2], P. Capak[12], S. Charlot[7], T. Contini[8], J.G. Cuby[1], J. Dunlop[16], S. Fotopoulou[9], A. Koekemoer[17], C. López-Sanjuan[11], Y. Mellier[7], J. Pforr[1], M. Salvato[14], N. Scoville[12], Y. Taniguchi[15], and P.W. Wang[1]

[1] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France
[2] INAF–Osservatorio Astronomico di Bologna, via Ranzani,1, I-40127, Bologna, Italy
[3] INAF–IASF, via Bassini 15, I-20133, Milano, Italy
[4] INAF–Osservatorio Astronomico di Roma, via di Frascati 33, I-00040, Monte Porzio Catone, Italy
[5] University of Bologna, Department of Physics and Astronomy (DIFA), V.le Berti Pichat, 6/2 - 40127, Bologna, Italy
[6] INAF–IASF Bologna, via Gobetti 101, I–40129, Bologna, Italy
[7] Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, 75014 Paris, France
[8] Institut de Recherche en Astrophysique et Planétologie - IRAP, CNRS, Universit de Toulouse, UPS-OMP, 14, avenue E. Belin, F31400 Toulouse, France
[9] Department of Astronomy, University of Geneva, ch. d'Ă030cogia 16, CH-1290 Versoix, Switzerland
[10] Geneva Observatory, University of Geneva, ch. des Maillettes 51, CH-1290 Versoix, Switzerland
[11] Centro de Estudios de Física del Cosmos de Aragón, Teruel, Spain
[12] Department of Astronomy, California Institute of Technology, 1200 E. California Blvd., MC 249–17, Pasadena, CA 91125, USA
[13] Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA
[14] Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, D-85741, Garching bei München, Germany
[15] Research Center for Space and Cosmic Evolution, Ehime University, Bunkyo-cho 2-5, Matsuyama 790-8577, Japan
[16] SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, United Kingdom
[17] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[18] Instituto de Fisica y Astronomia, Facultad de Ciencias, Universidad de Valparaiso, Av. Gran Bretana 1111, Casilla 5030, Valparaiso, Chile
[19] Astronomical Observatory of the Jagiellonian University, Orla 171, 30-001 Cracow, Poland
[20] National Centre for Nuclear Research, ul. Hoza 69, 00-681, Warszawa, Poland

Preprint online version: November 14, 2014

ABSTRACT

We investigate the evolution of galaxy clustering for galaxies in the redshift range $2.0 < z < 5.0$ using the VIMOS Ultra Deep Survey (VUDS). We present the projected (real-space) two-point correlation function $w_p(r_p)$ measured by using 3022 galaxies with robust spectroscopic redshifts in two independent fields (COSMOS and VVDS-02h) covering in total 0.8 deg^2. We quantify how the scale dependent clustering amplitude r_0 changes with redshift making use of mock samples to evaluate and correct the survey selection function. Using a power-law model $\xi(r) = (r/r_0)^{-\gamma}$ we find that the correlation function for the general population is best fit by a model with a clustering length $r_0 = 3.95^{+0.48}_{-0.54}$ h^{-1}Mpc and slope $\gamma = 1.8^{+0.02}_{-0.06}$ at $z \sim 2.5$, $r_0 = 4.35 \pm 0.60$ h^{-1}Mpc and $\gamma = 1.6^{+0.12}_{-0.13}$ at $z \sim 3.5$. We use these clustering parameters to derive the large-scale linear galaxy bias b_L^{PL}, between galaxies and dark matter. We find $b_L^{PL} = 2.68 \pm 0.22$ at redshift $z \sim 3$ (assuming $\sigma_8 = 0.8$), significantly higher than found at intermediate and low redshifts. We fit an HOD model to the data and we obtain that the average halo mass at redshift $z \sim 3$ is $M_h = 10^{11.75 \pm 0.23}$ $h^{-1}M_\odot$. From this fit we confirm that the large-scale linear galaxy bias is relatively high at $b_L^{HOD} = 2.82 \pm 0.27$. Comparing these measurements with similar measurements at lower redshifts we infer that the star-forming population of galaxies at $z \sim 3$ should evolve into the massive and bright ($M_r < -21.5$) galaxy population which typically occupy haloes of mass $\langle M_h \rangle = 10^{13.9} h^{-1} M_\odot$ at redshift $z = 0$.

Key words. Cosmology: observations – large-scale structure of Universe – Galaxies: high-redshift – Galaxies: clustering

1. Introduction

Since the very first galaxy surveys have revealed the complex structure of the Universe, the mapping of its evolution has become an important part of Cosmology. Large surveys revealed that the Universe is composed of dense regions like clusters and filaments, but also of almost empty voids. Additionally, the underlying dark matter structure and its evolution is shown to follow the visible baryonic matter, though the luminous matter is *biased* in relation to the dark matter distribution. It is still unclear how exactly this baryonic - dark matter relation looks like, especially in the early stages of galaxy formation, and how it changed through time.

The correlation function is a commonly used tool to describe how galaxies are clustered as a function of scale, and allows to put constrains on the evolution of galaxies (Kauffmann et al. 1999, Zehavi et al. 2011). It is based on the very simple idea of measuring the probability of finding two galaxies at a given redshift (Peebles 1980). The galaxy correlation function can be interpreted by introducing two kinds of approximations. The first and most extensively used formalism is based on a simple power-law approximation of the correlation function of the form $\xi(r) = (r/r_0)^{-\gamma}$ with two free parameters: the correlation length r_0, which describes how strongly galaxies are clustered, and the slope γ (Davis & Peebles 1983). The second, more recent and detailed approximation is based on halo occupation models (Seljak 2000, Peacock & Smith 2000, Magliocchetti & Porciani 2003, Zehavi et al. 2004, Zheng et al. 2005). In this framework the correlation function is built from two components, which have their influence on different scales. The *One-halo* term dominates on small scales ($\lesssim 1.5\ h^{-1}$Mpc) and describes clustering of galaxies which reside within dark matter haloes. The *two-halo* term describes large scale ($\gtrsim 3\ h^{-1}$Mpc) galaxy clustering between different halos.

Both formulations have been extensively used in the past years. Correlation function measurements have been produced for most large galaxy surveys, such as the Sloan Digital Sky Survey (SDSS, Connolly et al. 2002, Zehavi et al. 2004), 2dF Galaxy Redshift Survey (2dFGRS, Magliocchetti & Porciani 2003), VIMOS-VLT Sky Survey (VVDS, Le Fèvre et al. 2005, Pollo et al. 2006), VIMOS Public Extragalactic Survey (VIPERS, Marulli et al. 2013), and DEEP2 (Coil et al. 2006).

Thanks to this remarkable effort it is now very well established that the strength of the galaxy clustering for the general population of galaxies is only mildly evolving from intermediate redshifts $z \sim 1$ to $z \sim 0$. It has been found that galaxy clustering depends on a variety of galaxy properties like the luminosity, morphology, colour and spectral type of galaxies: luminous galaxies tend to be more clustered than faint ones and red galaxies with old stellar population are found to be more clustered than young blue ones (Norberg et al. 2002, Zehavi et al. 2004, 2012, Pollo et al. 2006, de la Torre et al. 2007, Coil et al. 2008, Quadri et al. 2008, Meneux et al. 2006, 2008, 2009, Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012).

At redshifts higher than $z \sim 2$ the situation is less clear. Various difficulties, mainly in collecting statistically significant and representative samples, need to be overcome. While a number of attempts to measure galaxy clustering have been made at early epochs, interpreting results is not straightforward, mainly because of uncertainties connected to redshift determination, small volumes covered, and different galaxy populations selected from a range of methods. Most of the measurements at high redshifts ($z > 1$) are produced using photometric surveys targeting specific classes of galaxies or applying specific observation techniques, like BzK method (Kong et al. 2006, Lin et al. 2012) or Lyman-break galaxy (LBG) selection (Foucaud et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Savoy et al. 2011, Bielby et al. 2013). In general, within a given sample and based on angular correlation function measurements, some evidence has been presented at $z > 1$ showing that clustering also seems to depend on luminosity, stellar mass or colour as observed on lower redshift, with bright galaxies clustering more strongly than faint ones (Savoy et al. 2011) and passive galaxies clustering more strongly than star-forming galaxies at a given stellar mass (Lin et al. 2012). However the relation of galaxy samples used in these analyses to the general population of galaxies is not well established, which makes the study of evolution of the galaxy clustering at $z > 2$ difficult. Moreover these galaxies can not be easily connected with the galaxy populations at lower redshifts, which make it difficult to conduct a consistent study of galaxy clustering evolution from high redshift to $z = 0$.

From the theoretical point of view, the correlation function should evolve with time - $\xi = \xi(r, t)$, because the density field of the Universe evolves over time. In the framework of the Newtonian linear perturbation theory (and in the matter-dominated era of the history of the Universe) in the co-moving coordinates the density contrast can be decomposed into time-dependent and space-dependent factors: $\delta(\boldsymbol{r}, t) = D(t)\delta_0(\boldsymbol{r})$, where $\delta_0(\boldsymbol{r})$ is the present day value of the density contrast at a given location, and $D(t)$ is referred to as the growth factor, which depends on the parameters of the assumed cosmological models. In practice, it means that the spatial shape of the density fluctuations in co-moving coordinates does not change, and only their amplitude increases. Consequently, the amplitude of the correlation function should increase over time: $\xi(r, t) = D^2(t)\xi(r, t_0)$, with a dependence on the cosmological parameters (Peebles 1980, Schneider 2006).

The consistent measurement of the evolution of a correlation function amplitude for a given galaxy population is therefore enough - in principle - to test the paradigm of the gravitational perturbations as the origin of the large scale structure of the Universe (and a way to estimate cosmological parameters). Extending these measurements to as high redshifts as possible is necessary both for the theoretical framework of the large scale structure evolution and for the galaxy formation and evolution models.

In this work we attempt to improve our current view of the evolution of the clustering of galaxies at $z > 2$, investigating how galaxy clustering evolved from the early phases of galaxy assembly to present times. We present a clustering analysis of the VIMOS Ultra Deep Survey (VUDS), the largest spectroscopic survey covering the redshift range $2 < z < 6$ in a continuous way to date (Le Fèvre et al. 2014), using a sample of ~ 3000 galaxies with confirmed spectroscopic redshifts in the range $2 < z < 5$. The sample covers a total area of 0.81 deg^2 observed in two independent fields, COSMOS and VVDS-02h,

⋆ Based on data obtained with the European Southern Observatory Very Large Telescope, Paranal, Chile, under Large Program 185.A-0791.

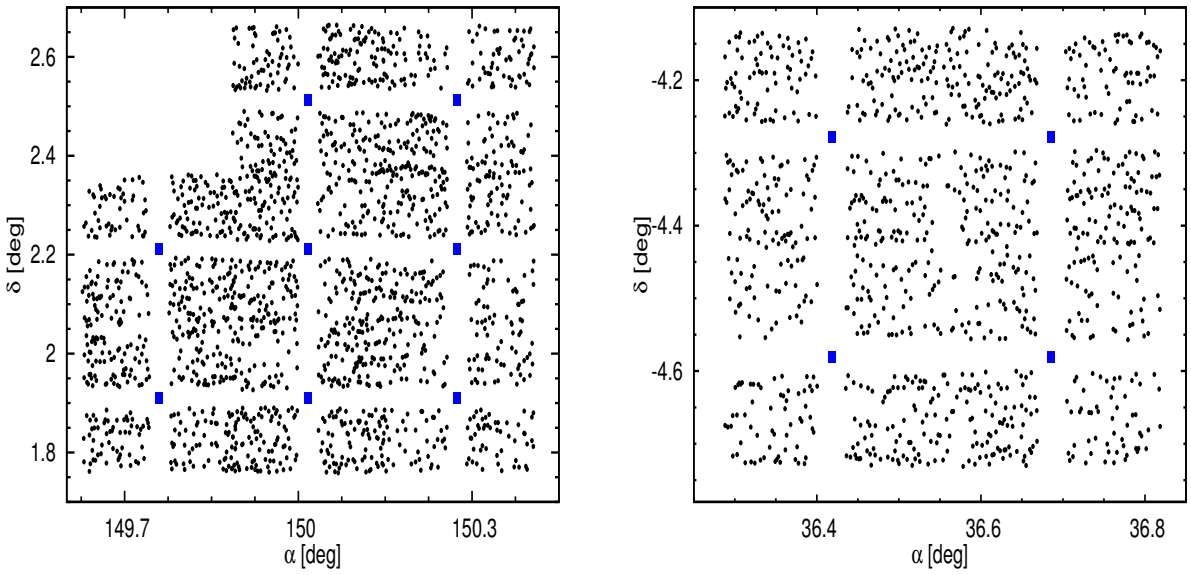


Fig. 1. The distribution of galaxies with spectroscopic redshifts $2 < z < 5$ in two independent VUDS fields: COSMOS on the *left* panel and VVDS-02h on *right* panel. The blue squares indicates VIMOS pointing centers.

with a mean redshift value of $z \sim 3$. We use a halo occupation distribution (HOD) model to put constrains on the properties of dark matter halos hosting star-forming galaxies at redshifts $z \sim 3$.

The paper is organized as follows. In section 2 we briefly describe the properties of the VUDS survey and our selected samples. The method used to measure the correlation function and derive power-law and HOD fits is presented in section 3. Results are described in section 4. We discuss our findings and compare to other work, both at low and high redshifts, in section 5, and we summarize our work in section 6.

We adopt a flat ΛCDM cosmological model with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$ (we note that using the latest cosmology parameters from e.g. Planck Collaboration et al. 2013 does not change conclusions drawn from clustering analysis, see Zehavi et al. (2011)). The Hubble constant is normally parametrized via $h = H_0/100$ to ease comparison with previous works, while a value $H_0 = 70$ km s^{-1} Mpc is used when computing absolute magnitudes and stellar masses. We report correlation length measurements in comoving coordinates and express magnitudes in the AB system.

2. The Data

2.1. The VUDS Survey summary

The VIMOS Ultra Deep Survey (VUDS, Le Fevre et al. 2014) is a spectroscopic survey targetting $\sim 10,000$ galaxies performed with the VIMOS multi-object spectograph (Le Fèvre et al. 2003) at the European Southern Observatory Very Large Telescope. The main aim of the survey is to study early phases of galaxy formation and evolution at $2 < z < 6$. The survey covers a total area of 1 deg^2 in three independent fields, reducing the effect of cosmic variance, an important property for galaxy clustering measurements. Targets are selected based on photometric redshifts ($z_{phot} > 2.4 + 1\sigma$) derived from deep multi-band photometry available in the VUDS fields, augmented by colour selection (e.g. Lyman Break Galaxies, LBG) when not selected by the photometric redshift criterion, and limit of $i_{AB} = 25$ is imposed. Details about the survey strategy, target selection, as well as data processing and redshift measurements are presented in Le Fevre et al. (2014). Here we briefly describe VUDS features which are relevant for our work.

Table 1. The number of galaxies in VUDS fields for each reliability flag in the redshift range $2 < z < 5$, as used in this study.

VUDS field	reliability flag $zflag$				Area (deg^2)
	2	3	4	9	
COSMOS	939	607	322	72	0.50
VVDS-02h	471	384	207	20	0.31
Total	1410	991	529	92	0.81

Redshift measurements were carried out in a way similar to that developed for the VVDS survey (Le Fèvre et al. 2005, Le Fèvre et al. 2013), the zCOSMOS survey (Lilly et al. 2007), and VIPERS survey (Guzzo et al. 2014). The core engine for redshift measurement is cross-correlation of the observed spectrum with reference templates using the EZ redshift measurement code (Garilli et al. 2010). At the end of the process, each redshift measurement is assigned a reliability flag which expresses the reliability of the measurement:

– Flag 0: No redshift could be assigned
– Flag 1: 50 – 75% probability to be correct
– Flag 2: 75 – 85% probability to be correct
– Flag 3: 95 – 100% probability to be correct
– Flag 4: 100% probability to be correct
– Flag 9: ∼ 80% probability to be correct; spectrum has a single emission line

The VUDS sample benefits from an extended multi-wavelength dataset (see Le Fevre et al. 2014). The multi-wavelength photometry is used to compute absolute magnitudes and stellar masses from SED fitting using the 'Le Phare' code (Arnouts et al. 1999, Ilbert et al. 2006), as described in detail by Ilbert et al. (2013) and references therein.

2.2. The sample selection for clustering analysis

In this work we are using only the objects with assigned reliability flag $zflag = 2, 3, 4, 9$ in the redshift range $2 < z < 5$. We use two independent fields, COSMOS and VVDS-02h, covering a total area 0.81 deg^2, as data for the third VUDS field in the ECDFS are still too sparse to be used for clustering analysis. Covered area corresponds to a volume $\sim 3 \times 10^7 \mathrm{Mpc}^3$ sampled in the redshift range $2 < z < 5$. Our sample consists of 3022 galaxies with reliable spectroscopic redshifts in the range $2 < z < 5$. The sample is summarized in Table 1.

The spatial distribution of the spectroscopic galaxy sample in each of these fields is presented in Figure 1, while Figure 2 presents their redshift distributions. Mean redshift value for the whole sample is $z \sim 2.96$. We use 3 different redshift bins to study our sample. In a first sample we include the whole star-forming galaxy population observed by VUDS in the range $2 < z < 5$ with a mean redshift $z \sim 2.96$. We also define a *low* redshift sample with a mean redshift $\bar{z} = 2.5$ in the range z=[2,2.9] and a *high* redshift sample with $\bar{z} = 3.45$ in z=[2.9,5], which roughly represent two equal cosmic time bins. The general properties of these samples including median luminosity M_U^{median} and stellar mass log M_*^{median} are listed in Table 2.

Because of the $i_{AB} = 25$ limit of the VUDS selection, the galaxies in the high redshift sample are ∼ 0.5 mag brighter and ∼ 0.2 $M_\odot$ more massive than those in the low redshift subsample, which is the expected effect taking into account the VUDS selection strategy. The impact of these differences on the clustering measurements of our three samples are fully discussed in Section 5.

2.3. VUDS mock catalogues

To estimate uncertainties of the correlation function (see Sec. 3.2) and to test robustness (see App. A) of our clustering measurements we make use of a large number of mock galaxy samples, which are designed to mirror the VUDS sample in the range $2 < z < 5$ in the most realistic way.

The 66 independent mock samples have been created based on a method which includes the halo occupation distribution (HOD) and the Stellar-to-Halo Mass Relation (SHMR). This method is the same as developed and introduced by de la Torre et al. (2013), and we refer the reader to the details presented in this paper. We provide below a brief summary of its most important features.

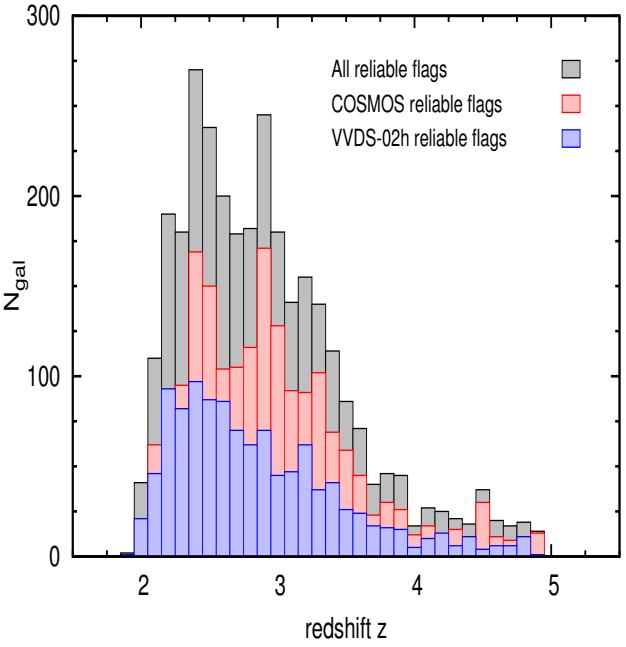


Fig. 2. The redshift distribution of the used VUDS galaxy sample for redshift range $2.0 < z < 5.0$. The filled gray histogram represents total sample of reliable galaxies (with $zflag = 2 - 4, 9$, ∼ 80% reliability), while the red and blue histograms represent the contribution from COSMOS and VVDS2h fields respectively.

In the first step the mock catalogues containing stellar masses have been created from the MultiDark *N*-body simulation (Prada et al. 2012) and Pinocchio halo lightcones (Monaco et al. 2002). We followed the stellar mass to halo mass ratio (SMHR) approach based on the assumption of a monotonic relation between halo/subhalo masses and the stellar masses of the galaxies associated with them. We first populated the haloes in lightcones with subhaloes. For this we randomly distributed subhaloes around each distinct halo following a NFW profile (Navarro et al. 1997), so that their number density satisfies the subhalo mass function proposed by Giocoli et al. (2010). Then we assigned a galaxy to each halo and subhalo, with a stellar mass given by the SHMR of Moster et al. (2013).

The procedure which was followed in the next step is similar to the one used in the VVDS, zCOSMOS and VIPERS surveys also based on VIMOS observations as the VUDS survey (Meneux et al. 2006, Iovino et al. 2010, de la Torre et al. 2011b, 2013). In order to obtain fully realistic VUDS mock catalogues we add the detailed survey selection criteria (see Sec. 2.1 and Le Fevre et al. 2014). This procedure in the end produces mock parent photometric galaxy catalogues. We then apply the slit-positioning algorithm (SSPOC) with the same settings as used for the VUDS survey (Bottini et al. 2005). This allowed us to reproduce the VUDS footprint on the sky for a proper simulation of the small-scale angular incompleteness in the mocks.

As a result of this procedure, 66 realistic mock galaxy catalogues have been produced. They contain the detailed survey

completeness function and observational biases of the VUDS survey.

3. Clustering measurements

3.1. Two-point correlation function

The two point correlation function is one of the most frequently used statistical tool used to measure and investigate galaxy clustering. It measures the probability, above Poisson, of finding two galaxies separated by a given distance r (Peebles 1980).

$$P = [1 + \xi(r)]\rho \mathrm{d}V_1 \mathrm{d}V_2, \tag{1}$$

where P is the probability of finding these two galaxies in two infinitely small volumes dV_1 and dV_2 separated by the distance r, while the average density of galaxies is given by ρ.

Although the idea is relatively simple, this definition cannot be straightforwardly applied to compute the correlation function from real data samples, because of the limitations of galaxy surveys themselves. In principle, to retrieve ideal correlation functions one would need an unlimited survey which covers the whole sky, and includes all galaxies. Naturally, in practice such surveys are unreachable. A number of estimators of $\xi(r)$, aimed at minimizing the effects related to the limited number of objects and limited areas covered by available surveys, have been proposed (see, e.g., Davis & Peebles 1983, Hamilton 1993). The estimator the most commonly used because of its well established capability to minimize the above mentioned observational limitations is the Landy & Szalay (1993) estimator, which we also apply in this work:

$$\xi(r) = \frac{N_R(N_R-1)GG(r)}{N_G(N_G-1)RR(r)} - 2\frac{(N_R-1)GR(r)}{N_G RR(r)} + 1 \tag{2}$$

Here, N_G and N_R represent, respectively, numbers of galaxies in the galaxy sample and randomly distributed objects in the same volume as observed in the survey; $GG(r)$ is the number of distinct galaxy-galaxy pairs with separations lying in the interval $(r, r+dr)$. Similarly, $RR(r)$ and $GR(r)$ are the numbers of random-random pairs and galaxy-random pairs, respectively, in the same intervals. The galaxy pairs GG, GR and RR are normalized by $N_G(N_G-1)$, $N_G N_R$ and $N_R(N_R-1)$ respectively, where N_G and N_R are the number of galaxies in the data and random catalogues.

In this work we make galaxy clustering measurements using combined data from two independent VUDS survey fields - COSMOS and VVDS2h. The final $\xi^{final}(r)$ is computed for all fields simultaneously by using Landy& Szalay estimator (Eq. 2), and the differences in size and numbers between the fields are accounted for by an appropriate weighting scheme. In particular, each pair is multiplied by the number of galaxies per unit volume for the given field:

$$\xi^{final}(r) = \frac{\sum_{i=1}^{n_{field}} w_i \cdot \Big(GG_i(r) - 2GR_i(r) + RR_i(r)\Big)}{\sum_{i=1}^{n_{field}} w_i \cdot RR_i(r)}, \tag{3}$$

where $w_i = \left(N_{gal,i}/V_i\right)^2$.

The Landy & Szalay estimator requires the creation of a random catalogue, which follows the geometrical properties of the corresponding galaxy sample and is distributed in the same volume (or on the same area on the sky, in its angular version). The number of objects generated within this catalogue is also crucial - in practice it should be significantly more numerous than the real data sample. Too few random objects introduces shot noise into $\xi(r)$ measurements, especially on small scales ($r_p < 3h^{-1}$Mpc). After a set of tests, for the measurements presented in this paper we decided to use $N_R = 100,000$, for which additional noise was found to be negligible

Another problem one has to face while measuring the real space correlation function is related to peculiar velocities of galaxies. In redshift space, these peculiar velocities distort distances computed from the redshifts z, and, as a result, they affect the shape of the correlation function $\xi(r)$ itself. The corresponding distortions are known as the coherent infall and "fingers of God" (Kaiser 1987). These differences, however, apply only to the radial separations and do not have any influence on the measured position of a galaxy on the sky. A commonly applied method to eliminate them is to split the comoving redshift space separations into two components - parallel π and perpendicular r_p to the line of sight, thus re-defining the redshift-space correlation function as $\xi(r_p, \pi)$. Integrating $\xi(r_p, \pi)$ along the line of sight gives us a projected correlation function $w_p(r_p)$, which is the two-dimensional counterpart of the real-space correlation function, free from the redshift-space distortions (Davis & Peebles 1983):

$$w_p(r_p) = 2\int_0^\infty \xi(r_p,\pi)d\pi = 2\int_0^\infty \xi\left((r_p^2+y^2)^{\frac{1}{2}}\right)dy. \tag{4}$$

Here, y is the real-space separation along the line of sight, and $\xi(r)$ stands for real space correlation function computed for $r = \sqrt{r_p^2+y^2}$. In practice, the upper integral limit π_{max} has to be finite in order to avoid adding noise to the result. After performing a number of tests, we decided to use $\pi_{max} = 20$ h^{-1}Mpc, as the value around which the correlation function is the most stable.

3.2. Error estimates

Estimating errors of the two-point correlation function is the subject of numerous discussions in the literature since the time of the very first measurements (see Hamilton 1993, Fisher et al. 1994, Bernardeau et al. 2002). The usage of a properly constrained covariance matrix is needed to account for the fact that the values of w_p for different separations r_p are not independent - pair counts in different r_p bins include partially the same galaxies. After performing a number of tests, for the VUDS data we decided to apply a combined method, which makes use of the so-called "blockwise bootstrap re-sampling" (Barrow et al. 1984) coupled to mock catalogues (see Section 2.3), similar to the method proposed by Pollo et al. (2005).

The "blockwise bootstrap re-sampling" is based on the idea of perturbing the data set by randomly creating a large number of comparable pseudo data-sets, which differ only slightly from the original sample. A number of objects is randomly selected from the data sample (objects are allowed to be drawn multiple times), and the correlation function is computed for each of these sub-samples. This procedure is repeated N_{boot} times,

giving as a result the variation around the 'real' result, which is used as an error estimate.

The second method makes use of sets of independent mock surveys - the simulated catalogues based on large N-body simulations coupled with physical definitions of galaxies, inserted e.g. by semi-analytic models. The correlation functions computed for each independent mock catalogue give the variation of the result at each scale, and are used as error estimates.

For both the bootstrap re-sampling and mocks methods the associated covariance matrix **C** between the values of w_p on ith and kth scale can be computed as:

$$\mathbf{C}_{ik} = \left\langle \left(w_p^j(r_i) - \langle w_p^j(r_i)\rangle_j\right)\left(w_p^j(r_k) - \langle w_p^j(r_k)\rangle_j\right)\right\rangle_j, \quad (5)$$

where "⟨⟩" indicates an average over all bootstrap or mock realizations, the $w_p^j(r_k)$ is the value of w_p computed at $r_p = r_i$ in the cone j, where $1 < j < N_{mock}$ for the VUDS mocks and $1 < j < N_{boot}$ for the bootstrap data.

In our case the covariance matrix reconstructed from $N_{mock} = 66$ VUDS mock catalogues could not be directly applied to the observed data, because it caused the fit to be often unstable, and not able to properly converge. This instability was caused by the fact that the diagonal elements of the matrix have realistic values but the off-diagonal non-zero elements differ significantly from those pertaining to the data sample. In other words, the detailed statistical properties of the VUDS mock catalogues are not close enough to the real VUDS data to allow for such an operation directly. For this reason we computed error bars using the scatter between the VUDS mock catalogues (since we believe it is more realistic than the scatter between the bootstrap realizations), but the off-diagonal elements of the covariance matrix were computed using the classical bootstrap method. This means that for each redshift range we measured the correlation function $w_p(r_p)$ from 1) the VUDS data; 2) $N_{mock} = 66$ VUDS mock catalogues; 3) $N_{boot} = 100$ bootstrap re-samplings of the VUDS data and in the fitting and error estimation we used the combination of these measurements.

3.3. Systematics in the CF measurement

Before introducing correlation function measurements from our data several tests have been performed on mock catalogues to investigate the influence of various VUDS survey properties on the correlation function. As the result of these tests, the correlation function correction scheme has been performed as described in the Appendix A.

In the lower panel of Figure .1 it is shown that the final measurement of the correlation function $w_p(r_p)$, obtained after introducing the full correction scheme, is still slightly underestimated, by as much as $\sim 10\%$ on small scales, with respect to the 'true' correlation function obtained for the mock parent samples (see details in App. A). This systematic underestimation may affect the value of the correlation length r_0, and all HOD parameters including an underestimation of the average halo mass $\langle M_h\rangle$ with respect to the 'true' underlying values. To estimate the size of this effect, we performed a power-law function fit on the 'true' correlation function (computed for the parent mock) as well as on the 'observed' (the mock catalogue after SSPOC selection), with the full correcting scheme, both measured as the average from 66 VUDS mock catalogues. We found that the 'true' correlation length is on average larger than the 'observed' r_0 by $\Delta r_0 = 0.42 \pm 0.34$, and that the 'observed' slope was underestimated by $\Delta\gamma = 0.06 \pm 0.04$. This, of course, would impact the galaxy bias estimation, since it is derived directly from these parameters, which we estimated to be underestimated by $\Delta b = 0.35 \pm 0.27$.

To address this problem we compute the ratio w_p^{par}/w_p^{obs} between the 'true' $w_p^{par}(r_p)$ and 'observed' $w_p^{obs}(r_p)$ correlation function (based on the VUDS mock, see appendix A) on every scale r_p for each redshift range separately. The correlation function measurements are then multiplied by this ratio for each separation r_p, resulting in the final correlation function measurement $w_p(r_p)$ on which the power-law and HOD fitting have been performed.

All results presented in this paper include the correlation function corrections described above.

3.4. Power-law model

In most cases, especially in the local Universe, the correlation function $\xi(r)$ is well described by a power law function $\xi(r) = (r/r_0)^{-\gamma}$, where r_0 and γ are correlation length and slope, respectively. With this parametrisation, the integral in Equation 4 can be computed analytically and $w_p(r_p)$ can be expressed as:

$$w_p(r_p) = r_p\left(\frac{r_0}{r_p}\right)^{\gamma} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})}, \quad (6)$$

where Γ is the Euler's Gamma Function.

The values of w_p are not independent at different separations, hence it is not possible to use simple χ^2 minimization to find the best-fit parameters. However, the covariance matrix **C** is symmetric and real, and therefore it can be inverted (if it is not singular). We can fit w_p by minimizing a generalized χ^2, which is defined as

$$\chi^2 = \sum_{i,j}^{N_D} \left(w_p^{obs}(r_i) - w_p^{mod}(r_i)\right)\mathbf{C}_{ij}^{-1}\left(w_p^{obs}(r_j) - w_p^{mod}(r_j)\right). \quad (7)$$

Here the covariance matrix **C** is computed using the method described in section 3.2 (Eq. 5). The fitting procedure to estimate the power-law parameters r_0 and γ for the projected correlation function $w_p(r_p)$ follows Fisher et al. (1994) (see also Guzzo et al. 1997 and Pollo et al. 2005).

A more detailed description of the shape of the real-space correlation function can be done e.g. in the framework of Halo Occupation Distribution (HOD) models, which we discuss in the following section 3.5. Nevertheless the use of a power-law model to describe the correlation function remains an efficient and simple approximation of galaxy clustering properties.

3.5. The halo occupation distribution (HOD) model

As we mentioned above, in practice $\xi(r)$ is usually well fitted by a power law. However, there is no strict theoretical reason to force the galaxy correlation function to assume a power law shape. Indeed, recent measurements show deviations from a power law shape. The strongest deviations are observed on small scales, especially for the most luminous galaxies (Coil et al. 2006, Pollo et al. 2006, Zehavi et al. 2011),

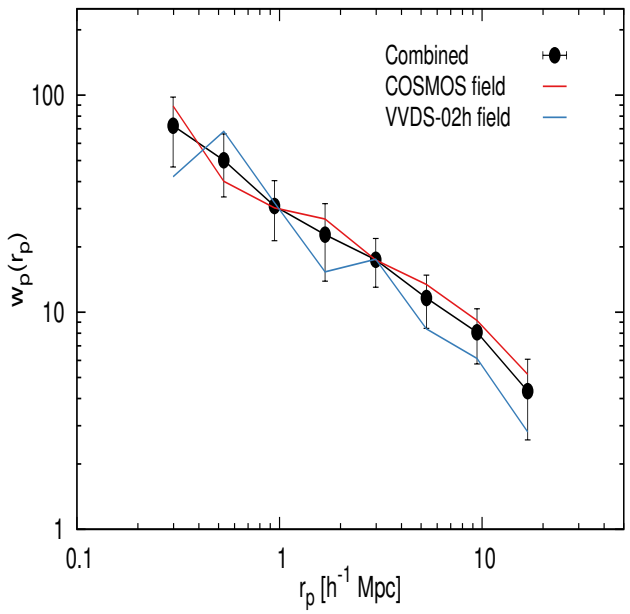


Fig. 3. The projected two-point correlation function $w_p(r_p)$ in each individual VUDS field for the redshift range $2 < z < 5$. Blue and red lines correspond to the VVDS and COSMOS field, respectively. Black points and line indicate the combined correlation function from the measurements preformed on galaxies from both fields simultaneously.

while probably the most famous ones are baryonic acoustic oscillations (BAOs) at large scales (Percival et al. 2010).

The small scale behaviour of the correlation function of galaxies can be well interpreted in the framework of the halo occupation distribution (HOD) models based on the relations between the distribution of dark matter and galaxies. The HOD models describe bias in terms of the probability $P(N|M)$ that a dark matter halo of a mass M contains N galaxies of a given type. Recently this approach has been very successfully used to model the shape of the two-point correlation function (e.g. Skibba et al. 2009, Abbas et al. 2010, Zehavi et al. 2011, Coupon et al. 2012, Kim et al. 2014).

In this work we apply the halo occupation model following the current commonly used analytical prescriptions, so that our results can be easily compared to results in the literature. We describe below the most important features and components of the HOD model we use, while for the general overview of the HOD philosophy we encourage readers to look into the review paper by Cooray & Sheth (2002). A very thorough description is also given by Coupon et al. (2012).

In the HOD framework the correlation function can be split into two components. The one-halo component $\xi^{1h}(r)$ dominates on scales smaller than the size of dark matter haloes (usually < 1.5 h^{-1}Mpc), while the two-halo component $\xi^{2h}(r)$ dominates on larger scales. The one-halo term arises from pairs of galaxies located within the same halo, while the two-halo terms is built by the pairs of galaxies hosted by different haloes. Consequently, the correlation function can be written as:

$$\xi(r) = \xi^{1h}(r) + \xi^{2h}(r). \tag{8}$$

The properties of the first term, i.e. $\xi^{1h}(r)$ can be described with the use of the halo occupation model which tries to follow the distribution of galaxies located in one dark matter halo. The second term $\xi^{2h}(r)$, depends on the statistical properties of the large scale dark matter density field and the distribution of dark matter haloes with respect to this density field.

The most important properties of the dark matter halo from this point of view are the halo mass function $n(M, z)$, for which we adopted the fitting formula proposed by Tinker et al. (2008), the halo density profile $\rho(r|M)$, for which we assumed the form described by Navarro, Frenk, & White (1997), and the halo bias $b_h(M, z, r)$, for which we use the formula proposed by Tinker et al. (2010) with the scale-dependency from Tinker et al. (2005).

We parametrized our halo occupation model in the way used by Zehavi et al. (2005b) and motivated by Kravtsov et al. (2004). In particular, we express the halo occupation function, i.e. the number of galaxies per halo, as a sum of a central galaxy and satellite galaxies. The mean occupation function for central galaxies $N_c(M)$ is represented by a step function, while the mean halo occupation function for satellite galaxies $N_s(M)$ is approximated by a Poisson distribution with the mean being a power-law:

$$\begin{aligned} \langle N_g|M\rangle &= 1 + \left(\frac{M}{M_1}\right)^{\alpha} \quad \text{for} \quad M > M_{min} \\ &= 0 \qquad\qquad\quad \text{otherwise,} \end{aligned} \tag{9}$$

where M_{min} is the minimum mass needed for a halo to host one central galaxy, and M_1 is the mass of a halo having on average one satellite galaxy, while α is the power law slope of the satellite mean occupation function.

The HOD model used in this work is simplified in comparison with the ones widely used at lower redshift ranges, e.g. by Zehavi et al. (2011) for fitting the SDSS data. However, based on the available statistics of the VUDS sample, with lower numbers of galaxies than e.g. in the SDSS, the three parameter model appears to be the best solution to retrieve robust measurements of M_{min} and M_1. A similar approach was taken by Abbas et al. (2010) in the case of the measurements from the VVDS survey at $z \sim 1$ with a similar sample size (Abbas et al. 2010). Since the details of the interplay between galaxy evolution and large scale structure growth at epochs as early as the ones probed by VUDS are still poorly understood, it is reasonable to keep a simplified description and not to dilute our analysis with models introducing too many free variables. The main results presented in this paper, however, are not expected to differ much from results where more complex HOD models are applied.

From the best-fit HOD parameters it is possible to obtain quantities describing halo and galaxy properties, like the average host halo mass M_h

$$\langle M_h|g\rangle(z) = \int dM\, M\, n(M, z) \frac{\langle N_g|M\rangle}{n_g(z)}, \tag{10}$$

the large-scale galaxy bias b_L^{HOD}

$$b_L^{HOD}(z) = \int dM\, b_h(M)\, n(M, z) \frac{\langle N_g|M\rangle}{n_g(z)} \tag{11}$$

where $n(M, z)$ is the dark matter mass function, $b_h(M, z)$ is the large-scale halo bias, and the n_g given by:

$$n_g = \int n(M) \langle N_g|M\rangle \, dM \tag{12}$$

represents the number density of galaxies.

3.6. Fitting the HOD model to the correlation function

To explore the HOD parameter space we implemented our HOD model with the publicly available CosmoPMC[1] code, which is using the "Population Monte Carlo" (PMC) technique to sample likelihood space (Wraith et al. 2009, Kilbinger et al. 2011). PMC is the adaptive importance-sampling technique (Cappé & Moulines 2007) which allows efficient sampling of the parameters space for the large number of samples. For each galaxy sample we fit the projected correlation function $w_p(r_p)$ and the number density of galaxies n_g, by summing both contributions in log-likelihood obtained by:

$$\chi^2 = \sum_{i,j}^{N_D} \left(w_p^{obs}(r_i) - w_p^{mod}(r_i)\right) \mathbf{C}_{ij}^{-1} \left(w_p^{obs}(r_j) - w_p^{mod}(r_j)\right) + \frac{\left(n_g^{obs} - n_g^{mod}\right)^2}{\sigma_{n_g}^2} \tag{13}$$

where n_g^{mod} is given by Eq. 12 at the mean redshift of the sample. The data covariance matrix is approximated taking Eq. 5. The error on the galaxy number density σ_{n_g} contains Poisson noise and cosmic variance.

3.7. Galaxy large-scale bias

According to the current cosmological paradigm of structure formation, galaxies form and evolve inside dark matter halos (White & Rees 1978). In other words, there exists a connection between the dark matter distribution and galaxies in the dense dark matter regions, where the halo clustering is important, galaxies should be more clustered. The galaxy spatial distribution, however, is biased with respect to the dark matter density field. The strength of this effect is referred to as *galaxy bias*.

As a first approximation we use a linear model for the galaxy large-scale bias. It assumes a linear relation between galaxy $\sigma_{R,g}(z)$ and mass $\sigma_{R,m}$ rms at a given redshift:

$$\sigma_{R,g}(z) = b\sigma_{R,m}(z) \tag{14}$$

where b is the galaxy bias. We also assume that b is independent of scale R, which is true especially for large scales. Usually the adopted value for the scale on which σ is measured is $R = 8h^{-1}$Mpc. Locally, at $z = 0$, the mass fluctuations reach the value $\sigma_{8,m}(z = 0) = 0.83$ (Planck Collaboration et al. 2013). In our model the redshift evolution of this quantity is described as:

$$\sigma_{8,m}(z) = \sigma_{8,m}(z = 0)D(z), \tag{15}$$

[1] http://www2.iap.fr/users/kilbinge/CosmoPMC/

where

$$D(z) = \frac{g(z)}{g(0)(1 + z)}, \tag{16}$$

and $g(z)$ is the normalized *growth factor*, which describes how fast the linear perturbations grow with the scale factor. Its value depends on the assumed cosmological model, i.e. cosmological parameters. After Carroll et al. (1992) we write down

$$g(z) = 2.5\Omega_m \left[\Omega_m^{4/7} - \Omega_\Lambda + \left(1 + \frac{\Omega_m}{2}\right)\left(1 + \frac{\Omega_\Lambda}{70}\right)\right]^{-1}, \tag{17}$$

where Ω_m and Ω_Λ are the matter and dark energy density parameters at a given redshift.

The galaxy rms at a given scale R can be retrieved from the power-law fits to the correlation function as,

$$\sigma_{8,g} = \sqrt{C_\gamma \left(\frac{r_0}{8h^{-1}\mathrm{Mpc}}\right)^\gamma}, \tag{18}$$

with

$$C_\gamma = \frac{72}{2^\gamma(3 - \gamma)(4 - \gamma)(6 - \gamma)}. \tag{19}$$

where r_0 and γ are the correlation length and the slope of power-law approximation of the correlation function.

In the following sections, we refer to the galaxy large-scale bias computed by this method as b_L^{PL}. This is also used to compute the values based on measurements of the correlation function from other surveys when the b_L estimations were not presented by the authors; in these cases we use their power-law best fit parameters r_0 and γ (e.g. from Le Fèvre et al. 2005). Another estimate of linear large-scale galaxy bias can be obtain from the HOD model. We compute the large-scale galaxy bias b_L^{HOD} using the fit of the halo occupation function (Eq. 11). The comparison of these two bias values and differences between these two procedures are discussed in section 5.

4. Results

4.1. The VUDS correlation function - general population

The projected two-point correlation function $w_p(r_p)$ is computed for the galaxy population with $i_{AB} < 25$ at $2 < z < 5$ separated into three redshift ranges. We are using a total number of 3 022 spectroscopically confirmed VUDS galaxies located in two fields. In Figure 3 we present composite correlation function $w_p(r_p)$ measurements for the full data sample in the redshift range $2 < z < 5$, along with the measurements made for each field independently. It is worth stressing that it is the first time that the correlation function has been computed with such a high accuracy from spectroscopic data at redshift $z \sim 3$. As expected, the values of the correlation function measured for galaxies located in separate fields are slightly different due to cosmic variance. However, for both fields the measurements are consistent with respect to the error bars and the signal is retrieved on all scales $0.3 < r_p < 16 \quad h^{-1}$Mpc, so that the composite function is not dominated by the signal from any of the individual fields at any separation r_p.

The correlation functions for the full sample z=[2,5] and the two z=[2,2.9] and z=[2.9,5] redshift intervals are presented in Figure 4.

Table 2. Properties of general population VUDS samples in the redshift range $2.0 < z < 5.0$.

z range	z_{mean}	M_U^{median}	$\log M_*^{median}$	N_{gal}	r_0	γ	b_L^{PL}
[2.0 – 5.0]	2.95	-21.55	9.76	3022	$3.97^{+0.36}_{-0.38}$	$1.70^{+0.09}_{-0.09}$	$2.68^{+0.22}_{-0.21}$
[2.0 – 2.9]	2.50	-21.31	9.66	1556	$3.95^{+0.48}_{-0.54}$	$1.81^{+0.02}_{-0.06}$	$2.39^{+0.32}_{-0.29}$
[2.9 – 5.0]	3.47	-21.81	9.86	1466	$4.35^{+0.60}_{-0.76}$	$1.60^{+0.12}_{-0.13}$	$3.26^{+0.47}_{-0.37}$

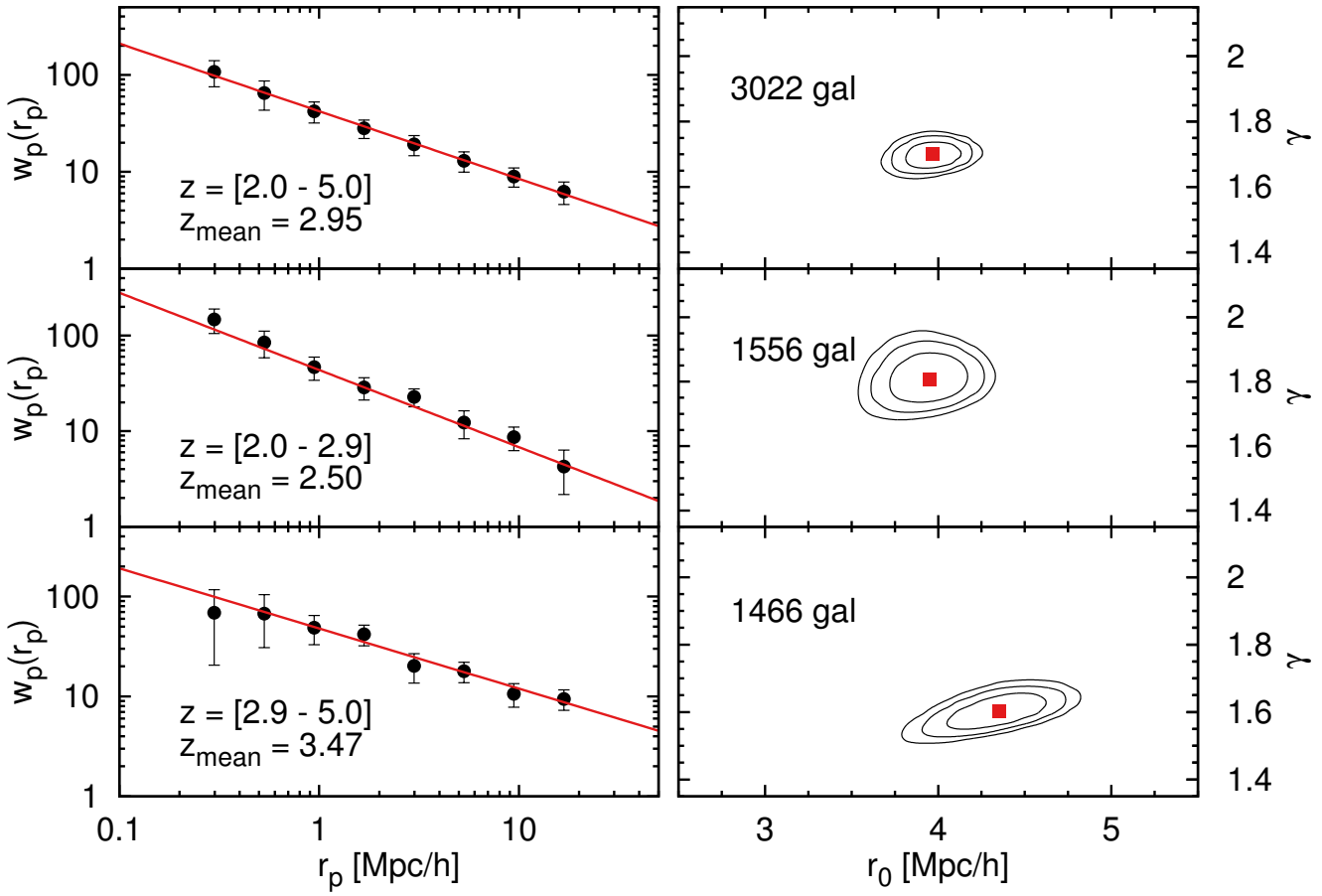


Fig. 4. *Left panel*: The projected two-point correlation function $w_p(r_p)$. The symbols and errorbars denote measurements of the composite correlation function from the VUDS survey. *Right panel*: The associated best-fit power-law parameters r_0 and γ along with error contours for the general galaxy population separated in three redshift ranges.

4.2. General population - Power law modelling

We fit a power-law to $w_p(r_p)$, with two free parameters r_0 and γ, to quantify the clustering strength. All the points in the range $0.3\ h^{-1}\mathrm{Mpc} < r_p < 16\ h^{-1}\mathrm{Mpc}$ are used (see Sec. 3.4 for the fitting method details). The measured best-fit parameters are listed in Table 2 and the power-law fit of the correlation function for each redshift sub-sample is shown in Figure 4 together with the (r_0,γ) error contours.

Within our sample we find $r_0 \sim 3.95^{+0.48}_{-0.54}$ h^{-1}Mpc in z=[2.0,2.9] and $r_0 \sim 4.35^{+0.60}_{-0.76}$ h^{-1}Mpc in z=[2.9,5]. The slight increase of r_0 with redshift is probably due to the small luminosity differences between samples (see Sec. 5 for discussion), although the difference is marginally significant given measurement errors. For these samples, the slope γ varies between $\gamma = 1.60^{+0.12}_{-0.13}$ and $\gamma = 1.81^{+0.02}_{-0.06}$, hence showing a tendency to decrease with redshift.

4.3. General population - HOD modelling

The best-fit HOD model defined in section 3.5 is fit to each of three redshift sub-samples as presented in the left panels of Figure 6. The HOD model provides a good fit to the real-space correlation function $w_p(r_p)$ in all the redshift ranges, particu-

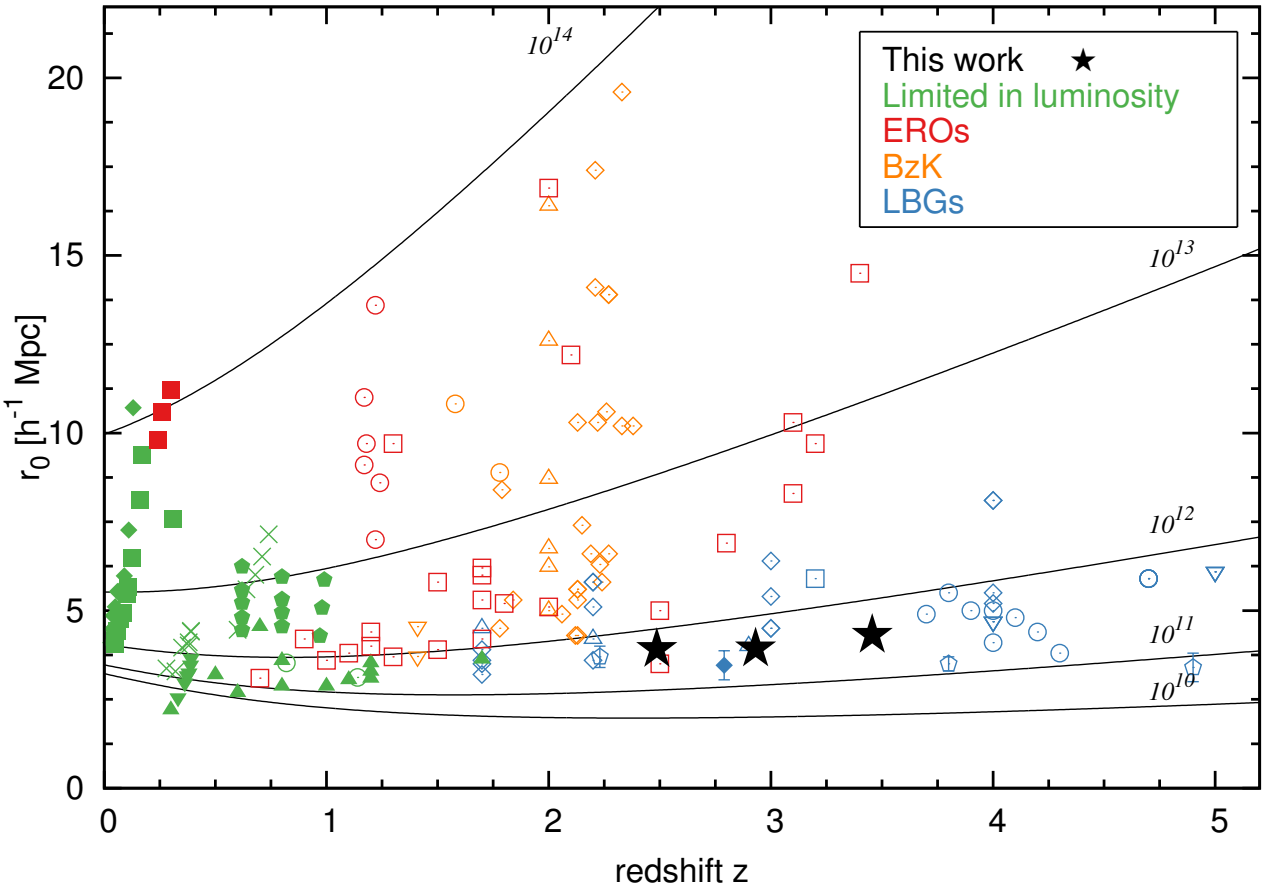


Fig. 5. Correlation length r_0 as a function of redshift from various surveys of different types of objects. Black stars indicate the measurements obtained from this work in three redshift ranges. Different colours indicate different galaxy populations targeted by using various techniques, as indicated in the upper right corner. Open symbols indicate measurements based on photometric data, while filled symbols are for measurements from spectroscopic data. *Blue:* LBG galaxies (Foucaud et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Savoy et al. 2011, Bielby et al. 2013, Barone-Nugent et al. 2014). *Orange:* BzK galaxies (Blanc et al. 2008, Hartley et al. 2010, McCracken et al. 2010, Lin et al. 2012). *Green:* Galaxy samples from surveys limited in luminosity (Norberg et al. 2002, Coil et al. 2006, Le Fèvre et al. 2005, Pollo et al. 2006, Zehavi et al. 2011, Marulli et al. 2013, Skibba et al. 2009). *Red:* EROs or massive red galaxies (Daddi et al. 2003, Zehavi et al. 2005a, Brown et al. 2008). The five solid curves show the correlation length of dark halos with different minimum masses, as labelled.

larly in z=[2,2.9] where it is an excellent representation of the data. For the highest redshift bin z=[2.9,5] the fit is not as good. This is mainly due to the more noisy measurement of the correlation function in this redshift range.

The three HOD parameters, the minimum halo mass M_{min}, the satellite occupation halo mass M_1, and the high halo mass slope α (see sec. 3.5), are inferred from the full covariance matrix (see section 3.6) and are listed in Table 3. The parameters are allowed to vary within a large range set from previous observations at high and low redshift: $10 < \log M_{min} < 14$, $10 < \log M_1 < 15$, $0.6 < \alpha < 2.0$. The right panel of Figure 6 presents the halo occupation function obtained from the HOD fit within three redshift ranges. We do not observe any significant difference in halo masses between the two redshift ranges probed for the general VUDS galaxy population. The minimum halo mass, for which a halo hosts at least one central galaxy M_{min}, has comparable value for the $z \sim 2.5$ and $z \sim 3.5$ samples.

The satellite occupation mass M_1 is noticeably higher in the higher redshift bin compared to the lower redshift sample, although the errors are quite large. These uncertainties are related to the weak one-halo term signal in the correlation function for the higher redshift measurement.

The high halo mass slope α of the satellite mean occupation function is around ~ 1.3 for z=[2,2.9], significantly steeper than $\alpha = 0.73$ found in z=[2.9,5].

The average host halo mass $< M_h|g >$, large-scale galaxy bias b_L^{HOD} and associated 1σ errors, are reported in Table 3. The redshift evolution of these parameters is presented in Figure 7 - for the average host halo mass, and in Figure 8 - for the large-scale galaxy bias. These results are fully discussed in Section 5.

Table 3. The HOD and and HOD-based parameters for the redshift sub-samples

z *range*	$\log M_{min}$	$\log M_1$	α	$\log\langle M_h\rangle$	b_L^{HOD}
[2.0 – 5.0]	$11.04^{+0.33}_{-0.36}$	$12.11^{+0.51}_{-0.45}$	$1.29^{+0.12}_{-0.11}$	$11.75^{+0.23}_{-0.28}$	$2.82^{+0.27}_{-0.16}$
[2.0 – 2.9]	$11.12^{+0.23}_{-0.35}$	$12.09^{+0.32}_{-0.35}$	$1.30^{+0.08}_{-0.10}$	$12.01^{+0.16}_{-0.18}$	$2.55^{+0.16}_{-0.21}$
[2.9 – 5.0]	$11.18^{+0.56}_{-0.70}$	$12.55^{+0.85}_{-0.88}$	$0.73^{+0.23}_{-0.30}$	$11.61^{+0.46}_{-0.40}$	$3.48^{+0.31}_{-0.26}$

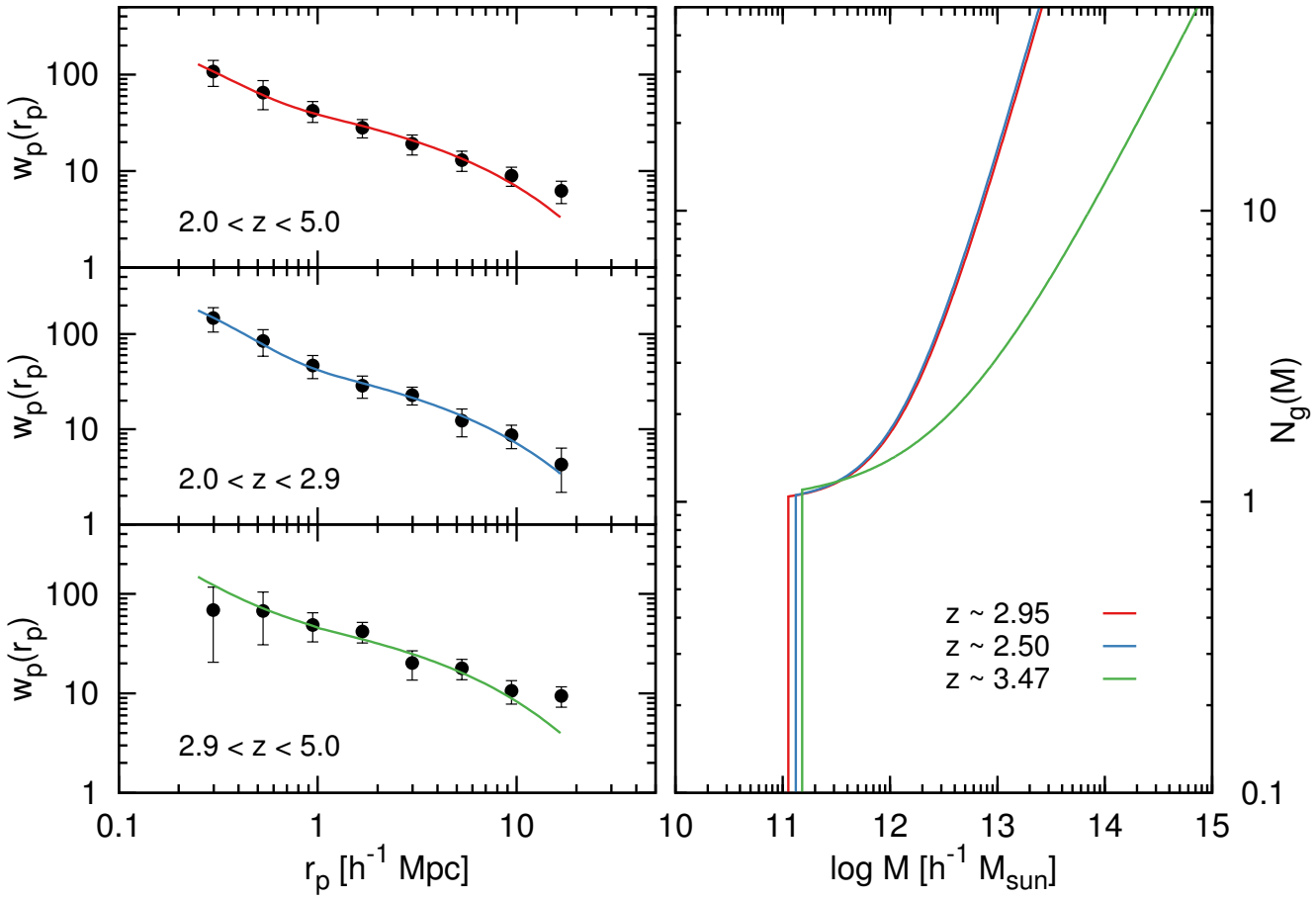


Fig. 6. *Left panel:* The correlation function for three redshift sub-samples as in Figure 4. The solid lines present the best-fit halo occupation model. *Right panel:* The evolution of the halo occupation for the three redshift ranges.

4.4. Linear large-scale bias measurements

The large-scale linear galaxy bias b_L^{PL} has been computed using the method described in Sec. 3.7, with the best-fit correlation parameters obtained for each galaxy sub-sample. The bias values obtained and the associated 1σ errors are listed in Table 2. We find that the galaxy bias is increasing from 2.39 ± 0.32 at $z \sim 2.5$ to 3.26 ± 0.47 at $z \sim 3.5$, galaxies in the higher redshift sample appearing to be more biased than those in the lower redshift sample. Additionally, we note that the bias reaches a relatively high value at $z \sim 3.5$. This is further discussed in Sec. 5.

5. Discussion

5.1. Evolution of the clustering length r_0 from $z \sim 3.5$ until the present epoch

In this paragraph we discuss the evolution of clustering properties of the general galaxy population since a redshift $z \sim 3.5$ (see Sec. 4) in the view of other measurements presented in the literature.

Within our sample, we observe the clustering strength to increase slightly with redshift over the range $z = [2.0 - 5.0]$, with r_0 being the highest at the highest redshift $z \sim 3.5$. However, this increase is marginally significant given the measurement

errors. The observed increase of r_0 with z, surprising at the first glance, may be caused by the luminosity difference between two samples at $z \sim 2.5$ and $z \sim 3.5$. Due to the the magnitude selection $i_{AB} \leq 25$ of the VUDS sample we observe systematically more luminous galaxies at higher z. As a result, our high redshift sample contains on average more luminous galaxies, with median $M_U = -21.81$, while for the lower redshift sample the median $M_U = -21.31$. The luminosity dependence of galaxy clustering is a well established fact; it has been observed both at early and late cosmic epochs (e.g. Zehavi et al. 2011, Coupon et al. 2012, Abbas et al. 2010, Pollo et al. 2006) and luminous galaxies at all z investigated so far were found to be more strongly clustered than the fainter ones which is usually related to the fact that they are located in more massive dark matter haloes and, consequently, stronger fluctuations of the dark matter density field. If this trend is also true for the redshift range investigated in this work, the observed increase of r_0 with z may be caused by the fact that our higher redshift sample is on average more luminous that the lower one. The dependency of clustering on luminosity at these high redshifts will be investigated in a future paper (Durkalec et al. in prep).

The correlation power-law parameters $r_0 = 3.97 \pm 0.48$ and $\gamma = 1.70 \pm 0.09$ at $z \sim 3$ are comparable with the clustering parameters for the star forming galaxies at intermediate ($z \sim 1$) redshift (e.g. Le Fèvre et al. 2005, Pollo et al. 2006), but the clustering length is lower than found for galaxies in the local Universe (e.g. Zehavi et al. 2011, Norberg et al. 2002). This has to be placed in the global perspective of the clustering expected to grow stronger along cosmic time.

A direct comparison of clustering strength at different epochs is a difficult task for several reasons. Different measurements are based on various galaxy samples, each with a unique selection function. Since galaxies cluster differently depending on their properties (like luminosity, colours or stellar masses), and in practice every catalogue contains a galaxy population somewhat different, the differences in clustering measurements may primarily be attributed to different galaxy sample properties as much as to evolutionary effects. The inhomogeneities between galaxy samples are especially pronounced for the high redshift samples. At $z > 2$ correlation functions have mostly been measured from surveys targeting specific classes of galaxies, e.g. extremely massive red objects or sources selected using a Lyman-break technique selecting star-forming galaxies. In addition, until recently most clustering measurements at $z > 2$ were produced from the angular correlation function computed on photometric samples with the knowledge of the redshift being only approximate and based on color selection (BzK e.g. Daddi et al. 2003, McCracken et al. 2010; LBG e.g. Adelberger et al. 2005; LAE e.g. Ouchi et al. 2004). Compared to spectroscopic samples, the lower accuracy and precision in redshift measurements in these samples and contamination in a given redshift bin from galaxies at other redshifts introduces a blurring of large scale structures which is difficult to overcome even when using large samples.

Our r_0 estimations are in excellent agreement with other available results at similar high redshifts (e.g. Daddi et al. 2003, Ouchi et al. 2004, Adelberger et al. 2005, Kashikawa et al. 2006, Hartley et al. 2010, Savoy et al. 2011), as presented in Figure 5. While r_0 in the VUDS sample is on the lower side of previous clustering measurements at these redshifts, most other samples have specific selection functions from which is it expected that they have a higher clustering bias than our sample, particularly for the redder or more luminous samples. Our measurements can be compared to the measurements by Bielby et al. (2013), based on a large sample of 2135 spectroscopically confirmed LBGs in $\sim$ 10 000 arcmin^2 at $z = 2.8$, the largest sample with a spectroscopically-based measurement from a large volume at these redshift prior to our measurements. They find a clustering length $r_0 = 3.46 \pm 0.41$ at $z = 2.79$ and a slope $\gamma = 1.5 - 1.6$, flatter than for local galaxies, in good agreement with our measurements given the associated errors. Savoy et al. (2011) measured a correlation length $r_0 = 3.5 \pm 0.5$ from $\sim$ 1400 UV-selected galaxies with photometric redshifts $z \sim 3$ and $23 < R < 27$ in the Keck Deep Fields, and Geach et al. (2012) found $r_0 = 3.7 \pm 0.3$ from the clustering of Hα emitters (HAEs) at $z = 2.23$ from HiZELS, both results being very similar to ours at comparable redshifts. We conclude that our measurements are in general agreement with the most recent and least biased measurements available in the literature. Our results are among the most reliable today given the VUDS survey area and the number of galaxies used in our study.

5.2. The host halo mass

We draw in Figure 5 the model predictions for the evolution of r_0 of dark matter halos following the prescription given by Mo & White (2002) (which is based on the Press & Schechter (1974) formalism). According to this model, dark matter halos of an average mass $\langle M_h \rangle \sim 10^{11.5} M_\odot$ have the same r_0 as observed VUDS galaxies at $z \sim 3$. If we take the rough assumption that at $z \sim 3$ haloes are occupied mostly by only one galaxy (central), this mass would be the average mass of a halo hosting the typical VUDS galaxy. However, this result is only a first approximation, since it is based on simplified assumptions of the Press & Schechter (1974) theory and needs to be refined.

The halo occupation model (HOD) is expected to give more accurate predictions, since it accounts for the actual number and distribution of galaxies occupying haloes. We obtain the number weighted average mass $\langle M_h \rangle$ of haloes hosting VUDS galaxies using the best-fit HOD parameters (Eq. 10), in three redshift ranges. We find that the halo mass decreases with redshift (see Figure 7), evolving from $\log\langle M_h \rangle = 11.61^{+0.46}_{-0.40} h^{-1} M_\odot$ at $z \sim 3.5$ to $\log\langle M_h \rangle = 12.01^{+0.16}_{-0.18} h^{-1} M_\odot$ at $z \sim 2.5$. These estimations are in reasonable agreement with the recent observations by Bielby et al. (2013) who found an average mass of $\log\langle M_h \rangle \sim 11.57 \pm 0.15 M_\odot$ for haloes hosting LBG galaxies at a redshift $z \sim 2.79$.

It can also be seen from Fig. 7 that our estimated host halo masses are much lower than the ones typically observed in the local Universe. At $z \sim 3$ the typical host halo masses are located in the range between $\langle M_h \rangle \sim 10^{11} - 10^{12} M_\odot$ while locally these masses usually reach the values between $10^{13} - 10^{14} M_\odot$. These results indicate that at some stage dark matter halos must have experienced a rapid accretion phase, as expected in the framework of the hierarchical mass assembly.

A universal formula for the mass growth of cold dark matter halos has been derived by van den Bosch (2002). The history of

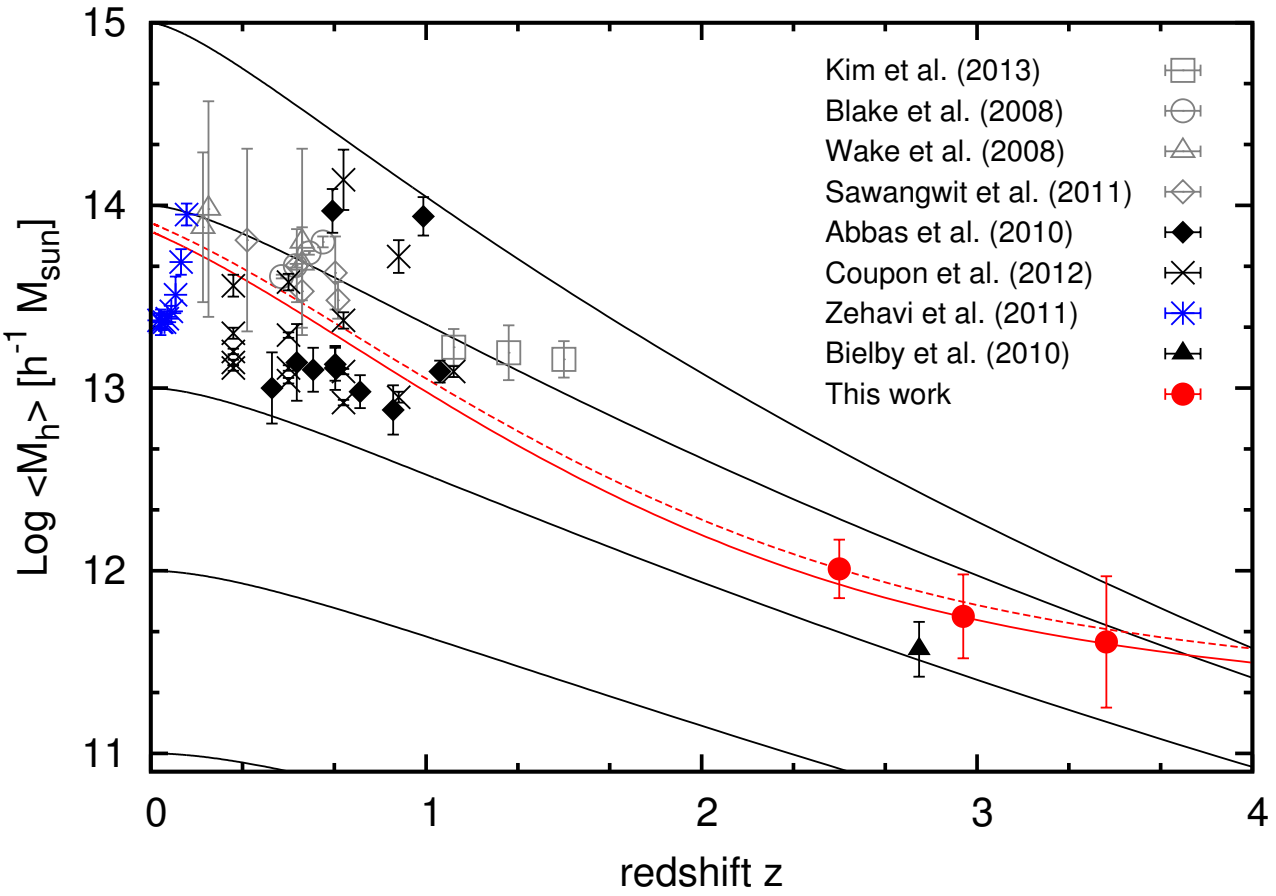


Fig. 7. The evolution of the number-weighted average host halo mass given by Eq. 10 for the three redshift ranges analysed in this study. The red filled circles indicate mass estimations from VUDS. Black and grey symbols represent the results of previous work based on spectroscopic and photometric surveys respectively. The solid black lines indicate how a host halo of a given mass M_0 at $z = 0$ evolves with redshift, according to the model given by van den Bosch (2002). The solid red line represents the halo mass evolution derived using Eq. 21, with the HOD parameters obtained from the best-fit HOD model at a redshift $z \sim 3$. The dashed red line is using the HOD best-fit parameters for $z \sim 2.5$. VUDS galaxies with a typical $L_\star$ luminosity are likely to evolve into galaxies with a luminosity $>L_\star$ today.

a halo with a given mass M_0 at $z = 0$ can be traced back in time using a simple formula:

$$\log\langle\Psi\,(M_0, z)\rangle = -0.301\left[\frac{\log\,(1+z)}{\log\left(1+z_f\right)}\right]^{\nu}, \tag{20}$$

where z_f and ν are free fitting parameters that depend on halo mass and cosmological parameters. To obtain these fitting parameters we follow the analytical formula from van den Bosch (2002). We then trace the evolution of the galaxy population sampled by VUDS from $z \sim 3$ to the present epoch to predict the mass of halos hosting the present day descendants of the VUDS galaxies. We find that in this model the typical VUDS halo with a mass $\langle M_H\rangle \sim 10^{11.75} M_\odot$ at $z \sim 3$ should evolve into a halo with a mass $\langle M_H\rangle \sim 10^{13.5} M_\odot$ at $z = 0$. In the local SDSS galaxy sample Zehavi et al. (2011) found that halos with these masses are typically occupied by star forming galaxies with luminosity $M_r < -20.5$. According to the van den Bosch (2002) model the star forming galaxies at $z \sim 3$ in VUDS with a typical characteristic luminosity $L_\star$ would likely evolve into galaxies equivalent to or brighter than $L_\star$ at the present day.

The above comparison assumes that each halo is occupied by only one galaxy. This is not expected to generally be the case, and the above picture, while broadly correct might need to be adjusted. In order to trace the evolution of dark matter haloes and the hosted galaxy population in a more realistic way we use both the halo mass growth model $\Psi(M_0, z)$ and the halo occupation function $\langle N_g|M\rangle$ at redshift $z = 3$. The average halo mass M_h as a function of redshift z is measured taking (see Eq. 10 for comparison):

$$\langle M_h\rangle(z) = \int dM\Psi^{-1}(M, z)n(M, z)\frac{\langle N_g|M\rangle}{n_g(z)}, \tag{21}$$

here the $\Psi^{-1}(M, z)$ is the inverse mass growth function proposed by van den Bosch (2002) (Eq. 20), $n(M, z)$ is the dark matter mass function, and n_g the galaxy number density. This allows us to trace the history of the typical dark matter halo hosting the average VUDS galaxy from a redshift $z = 3$ to the

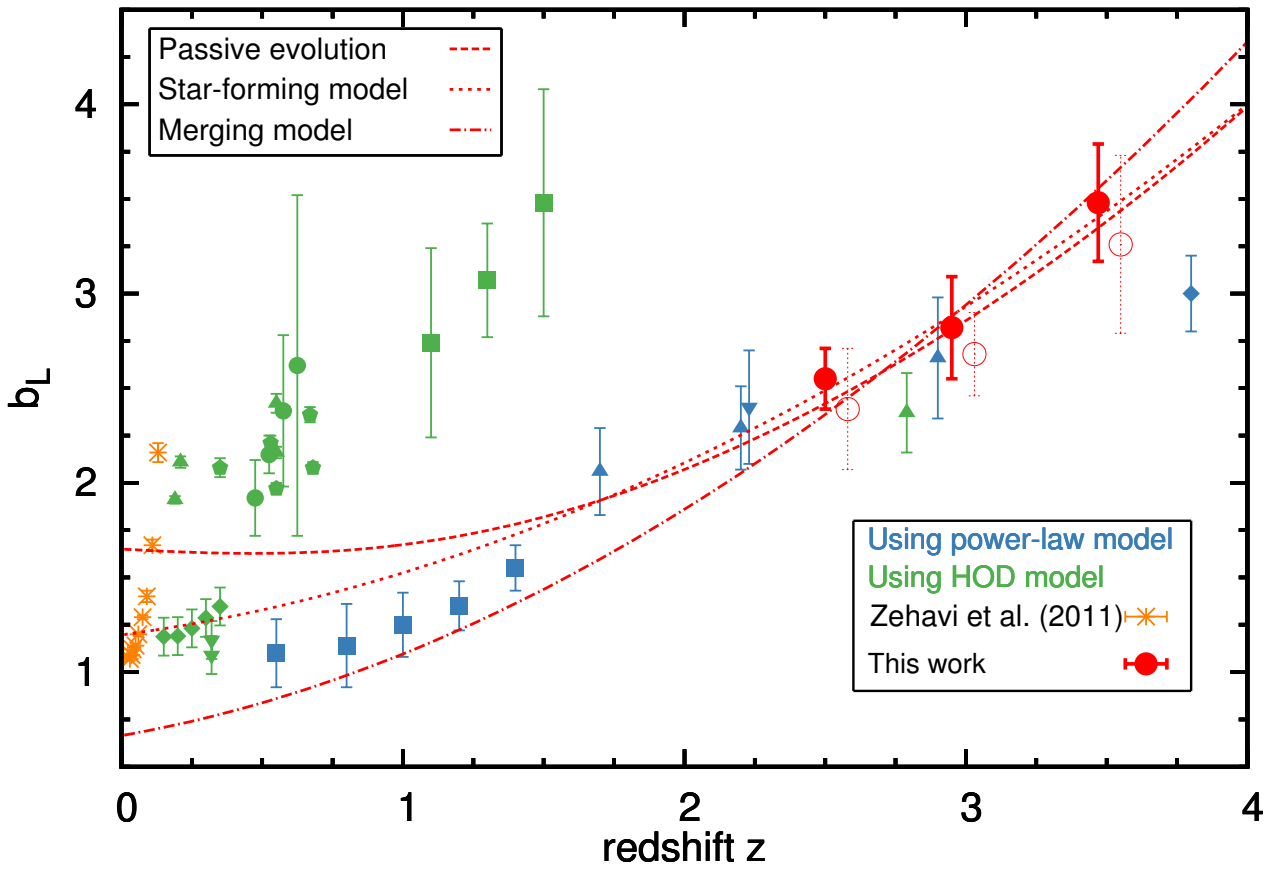


Fig. 8. Large-scale linear galaxy bias retrieved by using two methods. Red filled circles indicate measurements of b_L^{HOD}, while the open red circles represent b_L^{PL}) in three redshift ranges. Green and blue point represents the results from previous works, retrieved by using HOD (Blake et al. 2008, Wake et al. 2008, Ross & Brunner 2009, Sawangwit et al. 2011, Kim et al. 2014) and power-law frameworks (Marinoni et al. 2005, Adelberger et al. 2005, Geach et al. 2012, Barone-Nugent et al. 2014, Bielby et al. 2013) We show 3 different models for the evolution of bias: a *passive* evolution model (red dashed line) a *merging* model (red dot-dashed line), a *star-forming* model (dotted line), as described in the Sec. 5.3.

present day, as presented in Figure 7. We find that the typical halo of mass $\langle M_h \rangle = 10^{11.75}$ at $z \sim 3$ should evolve into a halo of mass $\langle M_h \rangle = 10^{13.9}$ at redshift $z = 0$. Comparing again our results from this improved model with Zehavi et al. (2011), halos with these masses are typically occupied by star forming galaxies with $M_r < -21.5$, above the characteristic luminosity $M_\star$ at $z \sim 0$. This means that the general galaxy population observed at redshift $z = 3$ evolves into the brightest and most massive galaxies, which occupy the most massive halos observed in the local Universe, as is expected in the hierarchical mass growth paradigm.

5.3. Galaxy bias

The hierarchical model of galaxy formation predicts an increase of the galaxy bias with redshift Mo & White (1996). In this section we discuss the relation between the distribution of galaxies and the underlying dark matter density field at different redshifts, by comparing the galaxy bias values measured from VUDS to the bias of galaxy populations at different redshifts from the literature.

Figure 8 shows the linear large-scale galaxy bias b_L, computed using methods based on power-law b_L^{PL} (see Sec. 3.7) and HOD b_L^{HOD} approximations (see Sec. 3.5). The results obtained from both methods are comparable within errors, and small differences are likely due to the less accurate power-law approximation. The general VUDS galaxy population shows a bias decreasing with decreasing redshift. According to our measurements, from $z \sim 3.5$ to $z \sim 2.5$ the linear bias decreased by $\Delta b_L \sim 0.9$. Both our values at $z = 2.5$ and $z = 3.5$ are significantly higher than measurements at later epochs reported in the literature. For comparison, we plot the linear large–scale bias inferred for the intermediate ($z = [0.5 - 1.5]$) redshift range by Marinoni et al. (2005), based on VVDS-Deep survey, for $z > 2$ by Geach et al. (2012) based on observations of $H\alpha$ emitters, and by Barone-Nugent et al. (2014) for $z \sim 3.8$ *LBGs*, along with the values we computed using the best-fit power-law parameters r_0 and γ from Le Fèvre et al. (2005), Adelberger et al. (2005) and Bielby et al. (2013) (the method is described in sec-

tion 3.7). The values presented in Table 3 are similar to the bias measurements recently observed for *LBGs* at redshift $z > 2$. Bielby et al. (2013) found $b_{gal} = 2.33 \pm 0.17$ from their LBG-selected spectroscopic sample, and at an even higher redshift $z \sim 3.8$ Barone-Nugent et al. (2014) estimated $b_L = 3.0 \pm 0.2$ for similarly selected galaxies.

The visible decrease of the galaxy bias with cosmic time can be explained in terms of the hierarchical scenario of structure formation. In this framework galaxies formed at early epochs in the most dense regions (highest peaks of the density field) which are the most biased with respect to the underlying average mass density field. As the mass density field evolved with time, galaxy formation is then expected to systematically move into less dense, hence less biased, regions. Additionally, it is likely that the gas in the most dense regions became too hot for collapse, and thus star formation could not take place in the strongest over-densities any more (e.g. Blanton et al. 1999). Therefore, the observed decrease of galaxy bias with cosmic time may be explained as galaxy formation is moving to less dense areas as cosmic time increases in a downsizing trend (De Lucia et al. 2006).

In Figure 8 we also draw the predicted evolution of the large-scale galaxy bias using different prescriptions. We consider three different theoretical descriptions of the biasing functions, based on different ideas of how evolution may proceed: the *passive*, the *merging*, and the *star forming* biasing models (see e.g. Marinoni et al. 2005). In the first model the number of galaxies (halo occupation) is conserved as a function of time and the bias evolves assuming the form presented in Eq. 11. The second model takes into account galaxy merging, as proposed by Mo & White (1996), who gave analytical prescriptions for computing the bias of halos using the Press & Schechter formalism. If we assume that galaxies can be identified with dark matter halos, an approximate expression for the biasing of all halos of mass $> M$ existing at redshift z (but which collapsed at redshift greater than the observation redshift, see discussion in Matarrese et al. 1997) is given by:

$$b(M,z) = 1 + \frac{1}{\delta_c}\left(\frac{\delta_c^2}{\sigma^2(M,z)} - 1\right) \quad (22)$$

where $\delta_c \simeq 1.69$ is the linear overdensity of a sphere which collapses in an Einstein-de Sitter Universe and $\sigma(M,z)$ is the linear rms fluctuations on scales corresponding to mass M at the redshift of observation. The *star-forming* model assumes that the distribution of galaxies with luminosity $> L$ is well traced by halos with mass $> M$, and predicts the biasing of objects that just collapsed at the redshift of observation (e.g., Blanton et al. 2000). In this model,

$$b(M,z) = 1 + \frac{\delta_c}{\sigma^2(M,z)} \quad (23)$$

represents the biasing of galaxies that formed in a narrow time interval around redshift z (i.e., galaxies which experienced recent star formation at redshift z).

From the halo mass analysis presented in Section 5.2 we infer that VUDS galaxies would evolve into galaxies with a luminosity $M_r < -21.5$ in the local universe (Section 5.2), for which the bias is $b_L = 1.7$. The star forming model predicts that the VUDS galaxy population would evolve into a weakly biased class of galaxies, corresponding to faint low mass galaxies. The merging model seem to have an evolution too steep compared to what is observed as it predicts an evolution to a bias significantly lower than populations of galaxies with the lowest clustering at z=0 (Zehavi et al. 2011). The model that would make the VUDS galaxies with $b_L \sim 2.8$ at z=3 evolve to $b_L = 1.7$ at z~0 is the passive model. However, it is likely that more sophisticated models would need to be developed to reproduce the joint evolution of DM halo mass and galaxy bias.

6. Summary

We examined the galaxy clustering properties of high redshift galaxies with 2<z<5, by measuring the two-point real-space correlation function $w_p(r_p)$ for 3022 galaxies from the VUDS survey. These measurements are among the first performed from spectroscopically confirmed galaxies at such high redshifts and allow to probe clustering evolution and galaxy host halos since very early times. We quantify our observations in terms of a power-law approximation and a three-parameter HOD model fitted to the correlation function computed in three redshift ranges, with median $z \sim 2.5$, $z \sim 3.0$ and $z \sim 3.5$. Our work therefore complements and extends to higher redshifts previous galaxy clustering analyses.

The results and conclusions can be summarized as follows:

- We observe a mild clustering evolution for the star-forming population in the VUDS survey. The correlation length increases from $r_0 = 3.95 \pm 0.48$ at $z \sim 2.5$ to reach $r_0 = 4.35 \pm 0.60$ at $z = 3.5$. We attribute this slight increase to the difference in absolute magnitudes between these two samples, the higher redshift sample being more luminous. The slope of the correlation function is found to be $\gamma = 1.81^{+0.02}_{-0.06}$ and $1.60^{+0.12}_{-0.13}$ at $z = 2.5$ and 3.5 respectively, similar or slightly flatter than the canonical $\gamma = 1.8$ observed at low redshifts. These measurements are in general agreement with the most recent and least biased measurements available in the literature at similar redshifts. We find that the power-law parameters we derive from the correlation function measurements are comparable to the clustering parameters for star-forming galaxies at intermediate redshifts, but the clustering length is somewhat lower than for galaxies in the local Universe.
- The average halo mass obtained from our HOD model fit reaches $\langle M_h \rangle = 10^{11.75} h^{-1} M_\odot$ at redshift $z = 3$, which is smaller than observed locally. Assuming the mass growth model from van den Bosch (2002) these halos should evolve into halos with a mass $\langle M_h \rangle (z = 0) = 10^{13.9} h^{-1} M_\odot$ at $z = 0$. As halos with such a mass are occupied by star forming galaxies with $M_r < -21.5$ in the local universe. we infer that the star-forming galaxy population observed at a redshift $z \sim 3$ has evolved into the massive and bright galaxy population observed today, as predicted in the hierarchical mass growth paradigm.
- We found that the linear large-scale galaxy bias b_L, computed using both the HOD and power-law best-fit parameters, increases with redshift from $b_L^{PL} = 2.68 \pm 0.22$ at $z = 2.5$ to $b_L^{PL} = 3.26 \pm 0.47$ at $z = 3.5$. These bias values are comparable with measurements obtained at similar redshifts but are much higher than bias values observed at low and intermediate redshift ranges where b_L reaches values in the range $1 - 2$. We discuss simple models to link the z~3 galaxies in our sample to lower redshift popula-

tions, and find that a passive halo evolution model maintaining halo occupation parameters constant will lead to a bias value close to $b_L = 1.7$ at z∼0 we inferred from the halo mass evolution. However it is likely that more sophisticated models are needed to fully interpret this evolution.

Our data provide a robust reference for the clustering of galaxies at an epoch before the peak in star formation rate. We infer that the star-forming population of galaxies at $z \sim 3$ should evolve into the massive and bright ($M_r < -21.5$) galaxy population which typically occupies haloes of mass $\langle M_h \rangle = 10^{13.9}$ at redshift $z = 0$. More complete modeling seems necessary to relate high redshift to low redshift galaxy populations, possibly taking into account non-linear effects, to better explain the galaxy clustering properties at early epochs in the life of the universe.

Acknowledgements. We thank Martin Kilbinger for assistance with implementing our HOD model into the CosmoPMC code. This work is supported by funding from the European Research Council Advanced Grant ERC-2010-AdG-268107-EARLY and by INAF Grants PRIN 2010, PRIN 2012 and PICS 2013. AC, OC, MT and VS acknowledge the grant MIUR PRIN 2010–2011. This work is based on data products made available at the CESAM data center, Laboratoire d'Astrophysique de Marseille. This work partly uses observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

References

Abbas, U., de la Torre, S., Le Fèvre, O., et al. 2010, MNRAS, 406, 1306
Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
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
Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Phys. Rep., 367, 1
Bielby, R., Hill, M. D., Shanks, T., et al. 2013, MNRAS, 430, 425
Blake, C., Collister, A., & Lahav, O. 2008, MNRAS, 385, 1257
Blanc, G. A., Lira, P., Barrientos, L. F., et al. 2008, ApJ, 681, 1099
Blanton, M., Cen, R., Ostriker, J. P., & Strauss, M. A. 1999, ApJ, 522, 590
Blanton, M., Cen, R., Ostriker, J. P., Strauss, M. A., & Tegmark, M. 2000, ApJ, 531, 1
Bottini, D., Garilli, B., Maccagni, D., et al. 2005, PASP, 117, 996
Brown, M. J. I., Zheng, Z., White, M., et al. 2008, ApJ, 682, 937
Cappé, O. & Moulines, E. 2007, ArXiv e-prints
Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499
Coil, A. L., Newman, J. A., Cooper, M. C., et al. 2006, ApJ, 644, 671
Coil, A. L., Newman, J. A., Croton, D., et al. 2008, ApJ, 672, 153
Connolly, A. J., Scranton, R., Johnston, D., et al. 2002, ApJ, 579, 42
Cooray, A. & Sheth, R. 2002, Phys. Rep., 372, 1
Coupon, J., Kilbinger, M., McCracken, H. J., et al. 2012, A&A, 542, A5
Daddi, E., Röttgering, H. J. A., Labbé, I., et al. 2003, ApJ, 588, 50
Davis, M. & Peebles, P. J. E. 1983, ApJ, 267, 465
de la Torre, S., Guzzo, L., Peacock, J. A., et al. 2013, A&A, 557, A54
de la Torre, S., Le Fèvre, O., Arnouts, S., et al. 2007, A&A, 475, 443
de la Torre, S., Le Fèvre, O., Porciani, C., et al. 2011a, MNRAS, 412, 825
de la Torre, S., Meneux, B., De Lucia, G., et al. 2011b, A&A, 525, A125
De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, MNRAS, 366, 499
Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1994, MNRAS, 267, 927
Foucaud, S., McCracken, H. J., Le Fèvre, O., et al. 2003, A&A, 409, 835
Garilli, B., Fumana, M., Franzetti, P., et al. 2010, PASP, 122, 827
Geach, J. E., Sobral, D., Hickox, R. C., et al. 2012, MNRAS, 426, 679
Giocoli, C., Tormen, G., Sheth, R. K., & van den Bosch, F. C. 2010, MNRAS, 404, 502
Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, A&A, 566, A108
Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37
Hamilton, A. J. S. 1993, ApJ, 417, 19
Hartley, W. G., Almaini, O., Cirasuolo, M., et al. 2010, MNRAS, 407, 1212
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Iovino, A., Cucciati, O., Scodeggio, M., et al. 2010, A&A, 509, A40
Kaiser, N. 1987, MNRAS, 227, 1
Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, ApJ, 637, 631
Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188
Kilbinger, M., Benabed, K., Cappe, O., et al. 2011, ArXiv e-prints
Kim, J.-W., Edge, A. C., Wake, D. A., et al. 2014, MNRAS, 438, 825
Kong, X., Daddi, E., Arimoto, N., et al. 2006, ApJ, 638, 72
Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35
Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Le Fèvre, O., Guzzo, L., Meneux, B., et al. 2005, A&A, 439, 877
Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1670–1681
Le Fevre, O., Tasca, L. A. M., Cassata, P., et al. 2014, ArXiv e-prints
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70
Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, ApJ, 756, 71
Magliocchetti, M. & Porciani, C. 2003, MNRAS, 346, 186
Marinoni, C., Le Fèvre, O., Meneux, B., et al. 2005, A&A, 442, 801
Marulli, F., Bolzonella, M., Branchini, E., et al. 2013, A&A, 557, A17
Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, MNRAS, 286, 115
McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202
Meneux, B., Guzzo, L., de la Torre, S., et al. 2009, A&A, 505, 463
Meneux, B., Guzzo, L., Garilli, B., et al. 2008, A&A, 478, 299
Meneux, B., Le Fèvre, O., Guzzo, L., et al. 2006, A&A, 452, 387
Mo, H. J. & White, S. D. M. 1996, MNRAS, 282, 347
Mo, H. J. & White, S. D. M. 2002, MNRAS, 336, 112
Monaco, P., Theuns, T., & Taffoni, G. 2002, MNRAS, 331, 587
Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Norberg, P., Baugh, C. M., Hawkins, E., et al. 2002, MNRAS, 332, 827
Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, ApJ, 611, 685
Peacock, J. A. & Smith, R. E. 2000, MNRAS, 318, 1144
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Priceton Univ. Press)
Percival, W. J., Reid, B. A., Eisenstein, D. J., et al. 2010, MNRAS, 401, 2148
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, ArXiv e-prints
Pollo, A., Guzzo, L., Le Fèvre, O., et al. 2006, A&A, 451, 409
Pollo, A., Meneux, B., Guzzo, L., et al. 2005, A&A, 439, 887
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
Quadri, R. F., Williams, R. J., Lee, K.-S., et al. 2008, ApJ, 685, L1
Ross, A. J. & Brunner, R. J. 2009, MNRAS, 399, 878
Savoy, J., Sawicki, M., Thompson, D., & Sato, T. 2011, ApJ, 737, 92
Sawangwit, U., Shanks, T., Abdalla, F. B., et al. 2011, MNRAS, 416, 3033
Schneider, P. 2006, Extragalactic Astronomy and Cosmology
Seljak, U. 2000, MNRAS, 318, 203
Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, MNRAS, 399, 966
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41
van den Bosch, F. C. 2002, MNRAS, 331, 98
Wake, D. A., Sheth, R. K., Nichol, R. C., et al. 2008, MNRAS, 387, 1045
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
Wraith, D., Kilbinger, M., Benabed, K., et al. 2009, Phys. Rev. D, 80, 023507
Zehavi, I., Eisenstein, D. J., Nichol, R. C., et al. 2005a, ApJ, 621, 22
Zehavi, I., Weinberg, D. H., Zheng, Z., et al. 2004, ApJ, 608, 16
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005b, ApJ, 630, 1
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791

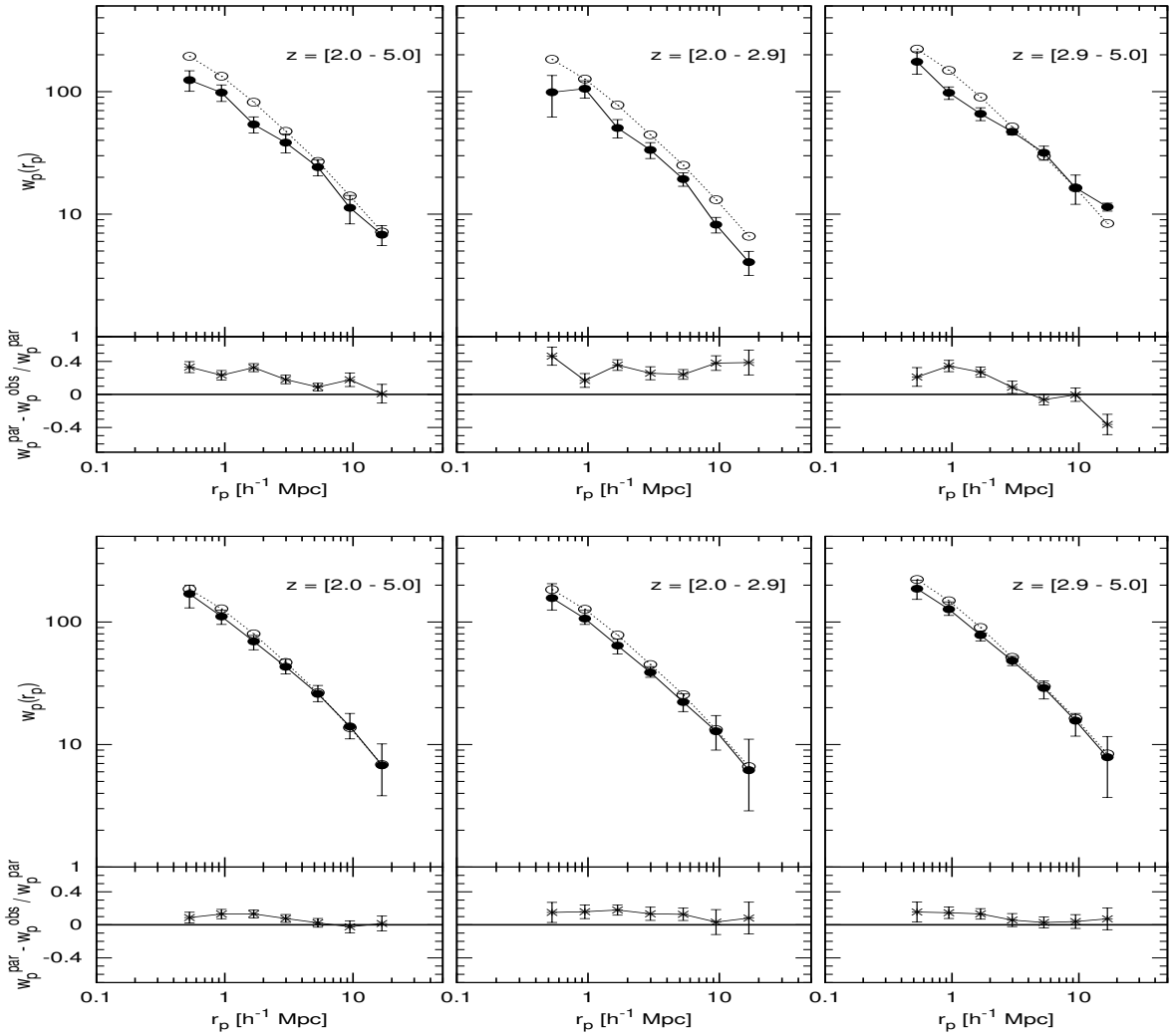


Fig. .1. Projected two point correlation function $w_p(r_p)$ measured for 66 VUDS mock catalogs in three redshift ranges. The points correspond to the mean measurement of all 66 mock catalogs, while the errors are computed as their standard deviations. The 'true' $w_p(r_p)$ computed for the whole parent sample (open circles) is compared to that measured from the 'observed' sample (filled circles). *Upper panel:* Without corrections (2) and (3). *Lower panel:* After our full correction scheme was applied.

Appendix A: Correlation function corrections scheme

In the galaxy surveys, there exists a number of observational biases which may influence the clustering measurements. These biases are related to the construction of the survey. Both the parent photometric catalogue, from which target galaxies are selected, and the final spectroscopic catalogue may be the source of these biases. Below, we list the most important survey features that cause the biases. Next, we describe the corrections applied to minimize their effects. Since the VUDS survey is performed on the same multi-slit spectrograph VIMOS as the VVDS, zCOSMOS and VIPERS surveys, the correction scheme adopted here is based on the same methodology and the ones used for the other VIMOS-based surveys in the past. In particular, the main corrections used here have been proposed and fully described by Pollo et al. (2005) and de la Torre et al. (2011a).

The main observational biases in the VUDS data can be listed as follows.

1. As shown in Figure 1, some of the areas are excised from the observations, due to the VIMOS lay-out (see A.1). The field of view of this instrument consists of four quadrants separated by 2′ gaps. No galaxies are observed in these gaps between quadrants, which influences the pair counts.
2. Obviously, not all galaxies from the photometric target candidate sample can be spectroscopically observed. Each galaxy spectrum occupies a certain area on the CCD detector, calculated from the spatial extent of the slit and the length of the spectrum due to spectral dispersion, which imposes geometrical constraints in target selection as galaxies too close to each other cannot be targeted simultaneously. It means that the galaxies from the parent photometric sample must be chosen is specific way to effectively allocate as many as possible 'spectral slits' for each observation.

In the case of the VIMOS observations, the slit allocation is performed automatically by the Super-SPOC code (Bottini et al. 2005). As a result, the spectroscopically observed sample is not a random representation of the parent photometric sample, and the introduced bias is relatively complex, especially on the small scales.

These biases, as well as a few other, less important, effects are described in detail by Pollo et al. (2005).

However, these biases can be minimized by a combination of corrections, which were presented by Pollo et al. (2005) and de la Torre et al. (2011a), and which we found the best working for the VUDS data:

1. Random sample construction. The first, most basic part of the correction scheme is the appropriate construction of the random catalog which needs to be geometrically identical (with the exception of the small scale non-linear biasses introduced by the SSPOC) with the spectroscopic sample. Generating random objects in this catalogue, we take into account the shape of the single pointings and quadrants, which are related to the shape of the VIMOS spectrograph. Additionally, we exclude the regions removed from the parent photometric sample (e.g. due to the presence of a bright star), by applying the same photometric mask to the random sample. These first-order corrections reduce most of the negative effects on the correlation function.
2. A global correction. In order to account for the missing pairs (due to the VIMOS limitations and the SSPOC strategy) we assign a "global" weight to each galaxy-galaxy pair. Assuming that the parent catalogue is free from angular incompleteness, we define a weighting function $f(\theta)$ as a ratio between the mean number of pairs in the parent photometric catalogue and the main number of pairs in the spectroscopic sample, as a function of angular separation (de la Torre et al. 2011a):

$$f(\theta) = \frac{1 + w_{par}(\theta)}{1 + w_{spec}(\theta)}, \tag{A.1}$$

   where $w_{par}\theta$ and $w_{spec}(\theta)$ are the angular correlation functions of the parent photometric and spectroscopic samples, respectively. Then, each pair from the spectroscopic sample separated by the angular distance θ is weighted by this ratio $f(\theta)$.
3. Small scale corrections. As it comes out, even this strategy does not fully account for the small-scale angular effects of the SSPOC target selection strategy. In order to account for these local small-scale biases we additionally use the local weighting scheme which is also using the parent photometric catalog as a reference. In this case, we count how many galaxies around the targeted galaxy we are missing due to the limited space for spectroscopic slits. Each targeted galaxy pair is then weighted proportionally to its representativeness of the surrounding density field. Similarly to Pollo et al. (2005) we found that the optimal size of the weighting area is the circle with a radius $\sim 40''$.

The correcting scheme including all these ingredients was confirmed to be optimal after a series of tests performed using all 66 VUDS mock catalogues. In figure .1 we present the results of the correlation function measurements for the mock

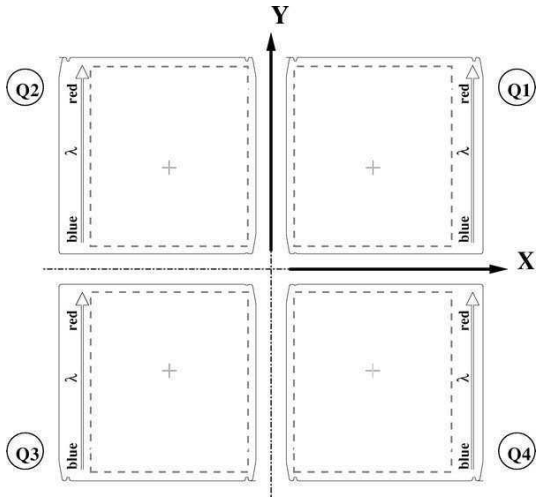


Fig. A.1. Lay-out of the VIMOS field of view. Picture from Bottini et al. (2005).

catalogues in our three redshift ranges. All the points correspond to the means of the results from 66 VUDS mock catalogues, while the errors are computed as the standard deviations between them. The three upper panels show the comparison between (1) the 'true' (computed for whole mock parent sample) correlation function and (2) the 'observed' correlation function computed only with the most obvious correction number (1), i.e. the appropriate geometrically cut random sample. The lower panels present the comparison of the same 'true' correlation function and (3) the 'observed' correlation function when the full correcting scheme described above was applied.

It is clearly seen that the introduced corrections significantly improve the retrieved signal from the 'observed' sample at all the spatial scales, and in all three redshift bins considered.

However, it can be noticed that on small scales ($r_p < 3h^{-1}$Mpc), there is still some signal missing, even after introducing the full optimal correcting scheme (see lower panels of Figure .1). For these separations, our measurement is on average underestimated by 10%±3% with respect to the 'true' value of $w_p(r_p)$ computed from the parent sample. The possible influence of this effect on the final results is discussed in section 3.3. At separations larger than $r_p \sim 3h^{-1}$Mpc the value of $w_p(r_p)$ is also slightly underestimated, but only by about 1.6% ± 0.9%.

Our correction scheme provides a significant improvement in comparison to the average 28%5% bias observed without applying it. However, the remaining systematic underestimation of the $w_p(r_p)$ has to be taken into account during the correlation function measurements (see Sec. 3.3).

Appendix B: Projected correlation function measurements

For every galaxy sub-sample used in this work we present the tables of the projected correlation function measurements $w_p(r_p)$ with associated 1σ errors at different separations r_p (in units h^{-1}Mpc).

Table B.1. General galaxy population. Projected correlation function measurements of redshift sub-samples. The first column provides the pair-weighted-projected separation of the bin r_p. Subsequent columns provide the projected correlation function values $w_p(r_p)$ along with 1σ errors.

r_p	$w_p(r_p)$		
	$2.0 < z < 2.9$	$2.0 < z < 5.0$	$2.9 < z < 5.0$
0.29	147.40 ± 42.02	107.65 ± 32.32	68.74 ± 48.22
0.53	84.84 ± 26.39	65.03 ± 21.72	67.51 ± 36.77
0.94	46.83 ± 12.79	42.22 ± 10.26	48.77 ± 15.83
1.68	28.71 ± 7.57	28.11 ± 6.07	41.94 ± 9.80
2.98	22.81 ± 4.87	19.20 ± 4.50	20.21 ± 6.58
5.31	12.32 ± 4.01	12.99 ± 3.06	17.87 ± 4.13
9.44	8.63 ± 2.39	8.97 ± 2.01	10.62 ± 2.82
16.79	4.26 ± 2.08	6.23 ± 1.62	9.44 ± 2.19

Astronomy & Astrophysics manuscript no. halo˙stellar˙mass˙v0.6
December 3, 2014
© ESO 2014

The stellar mass – halo mass relation from galaxy clustering in VUDS: a high star formation efficiency at z≃3 ⋆

A. Durkalec[1], O. Le Fèvre[1], S. de la Torre[1], A. Pollo[19,20], P. Cassata[1,18], B. Garilli[3], V. Le Brun[1], B.C. Lemaux[1], D. Maccagni[3], L. Pentericci[4], L.A.M. Tasca[1], R. Thomas[1], E. Vanzella[2], G. Zamorani[2], E. Zucca[2], R. Amorín[4], S. Bardelli[2], L.P. Cassarà[3], M. Castellano[4], A. Cimatti[5], O. Cucciati[5,2], A. Fontana[4], M. Giavalisco[13], A. Grazian[4], N. P. Hathi[1], O. Ilbert[1], S. Paltani[9], B. Ribeiro[1], D. Schaerer[10,8], M. Scodeggio[3], V. Sommariva[5,4], M. Talia[5], L. Tresse[1], D. Vergani[6,2], P. Capak[12], S. Charlot[7], T. Contini[8], J.G. Cuby[1], J. Dunlop[16], S. Fotopoulou[9], A. Koekemoer[17], C. López-Sanjuan[11], Y. Mellier[7], J. Pforr[1], M. Salvato[14], N. Scoville[12], Y. Taniguchi[15], and P.W. Wang[1]

[1] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France
[2] INAF–Osservatorio Astronomico di Bologna, via Ranzani,1, I-40127, Bologna, Italy
[3] INAF–IASF, via Bassini 15, I-20133, Milano, Italy
[4] INAF–Osservatorio Astronomico di Roma, via di Frascati 33, I-00040, Monte Porzio Catone, Italy
[5] University of Bologna, Department of Physics and Astronomy (DIFA), V.le Berti Pichat, 6/2 - 40127, Bologna, Italy
[6] INAF–IASF Bologna, via Gobetti 101, I–40129, Bologna, Italy
[7] Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie, 98b Boulevard Arago, Paris, France
[8] Institut de Recherche en Astrophysique et Planétologie - IRAP, CNRS, Universite de Toulouse, UPS-OMP, 14, avenue E. Belin, F31400 Toulouse, France
[9] Department of Astronomy, University of Geneva, ch. d'cogia 16, CH-1290 Versoix, Switzerland
[10] Geneva Observatory, University of Geneva, ch. des Maillettes 51, CH-1290 Versoix, Switzerland
[11] Centro de Estudios de Física del Cosmos de Aragón, Teruel, Spain
[12] Department of Astronomy, California Institute of Technology, 1200 E. California Blvd., MC 249–17, Pasadena, USA
[13] Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA
[14] Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, D-85741, Garching bei München, Germany
[15] Research Center for Space and Cosmic Evolution, Ehime University, Bunkyo-cho 2-5, Matsuyama 790-8577, Japan
[16] SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, United Kingdom
[17] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA Instituto de Fisica y Astronomia, Facultad de Ciencias, Universidad de Valparaiso, Av. Gran Bretana 1111, Casilla 5030, Valparaiso, Chile
[18] Astronomical Observatory of the Jagiellonian University, Orla 171, 30-001 Cracow, Poland
[19] National Centre for Nuclear Research, ul. Hoza 69, 00-681, Warszawa, Poland

Preprint online version: December 3, 2014

ABSTRACT

The relation between the galaxy stellar mass $M_\star$ and the dark matter halo mass M_h gives important information on the efficiency in forming stars and assembling stellar mass in galaxies. We present the stellar mass to halo mass ratio (SMHR) measurements at redshifts $2 < z < 5$, obtained from the VIMOS Ultra Deep Survey. We use halo occupation distribution (HOD) modelling of clustering measurements on ∼ 3000 galaxies with spectroscopic redshifts to derive the dark matter halo mass M_h, and SED fitting over a large set of multi-wavelength data to derive the stellar mass $M_\star$ and compute the SMHR=$M_\star/M_h$. We find that the SMHR ranges from 1% to 2.5% for galaxies with $M_\star$=1.3 × 10^9 $M_\odot$ to $M_\star$=7.4 × 10^9 $M_\odot$ in DM halos with M_h=1.3 × 10^{11} $M_\odot$ to M_h=3 × 10^{11} $M_\odot$. We derive the integrated star formation efficiency (ISFE), finding that the star formation efficiency is a moderate 6–9% for lower mass galaxies while it is relatively high at 16% for galaxies with the median stellar mass of the sample ∼ 7 × 10^9 $M_\odot$. The lower ISFE at lower masses may indicate that some efficient means of suppressing star formation is at work (like SNe feedback), while the high ISFE for the average galaxy at z∼3 is indicating that these galaxies are efficiently building-up their stellar mass at a key epoch in the mass assembly process. We further infer that the average mass galaxy at z∼3 will start experiencing star formation quenching within a few hundred millions years.

Key words. Cosmology: observations – large-scale structure of Universe – Galaxies: high-redshift – Galaxies: clustering

1. Introduction

Understanding processes regulating star formation and mass growth in galaxies along cosmic time remains a key issue of galaxy formation and evolution. In the ΛCDM model dark matter (DM) halos grow hierarchically, and galaxies are thought to form via dissipative collapse in the deep potential wells of these DM halos (e.g. White & Rees 1978, Fall & Efstathiou 1980). In this paradigm, cooling processes bring baryons in the high density peaks of the halos, where the conditions for gas fragmentation triggering star formation are met (Bromm et al. 2009). Current models connecting star formation and stellar mass evolution on the one hand and the formation histories of

⋆ Based on data obtained with the European Southern Observatory Very Large Telescope, Paranal, Chile, under Large Program 185.A-0791.

DM halos on the other hand are relying on approximations in the physics used and are in need to be further informed by observational data to reduce the uncertainties in the modelling process (e.g. Conroy & Wechsler 2009).

The efficiency of asssembling baryons into stars is an important ingredient to understand galaxy formation but remains poorly constrained observationally. In recent years it has been proposed to derive this efficiency comparing DM halo masses with galaxy stellar masses. With the measurement of the characteristic mass of DM halos M_h now available from observational data and of stellar mass $M_\star$ derived from the analysis of the spectral energy distribution of galaxies, coupled to the knowledge of the cosmological density of baryons and DM, one can infer the conversion rate from baryons to stellar mass.

Two methods have been used so far to make this link between $M_\star$ and M_h. Halo occupation models provide a description of how galaxies populate their host halos using galaxy clustering statistics and space densities (e.g. Zehavi et al. 2005, Leauthaud et al. 2012). Abundance matching connects galaxies to the underlying dark matter structure directly, under the assumption that the stellar masses or luminosities of the galaxies are tightly connected to the masses of dark matter halos (Conroy et al. 2009, Moster et al. 2013). The efficiency with which the galaxies converted baryons into stars is encoded in how $M_\star$ depends on M_h as a function of redshift, providing a benchmark against which galaxy evolution models can be tested. Using observed stellar mass functions, abundance matching models have led to the derivation of the Stellar Mass – Halo Mass (SMHM) relation which gives for a given halo mass the Stellar Mass to Halo Mass ratio (SMHR), SMHR=$M_\star/M_h$. Behroozi et al. (2010) find that the integrated star formation efficiency (ISFE) at a given halo mass peaks at 10-20% of available baryons for all redshifts from 0 to 4.

The shape of the SMHM is claimed not to evolve from z=0 to z=4, although it may be evolving at z>4 (Behroozi et al. 2013, Behroozi & Silk 2014). The SMHR is characterized by a peak at around $M_h = 10^{12} M_\odot$, with a lower ratio at lower and higher mass. The lower efficiency at masses below this peak may indicate that supernova feedback might be sufficient to remove gas from the galaxy as the halo gravitational potential is lower (e.g. Silk 2003, Bertone et al. 2005, Béthermin et al. 2013). At higher masses, the decrease in star formation efficiency might be produced when cold streams are replaced by isotropic cooling (e.g. Dekel & Birnboim 2006, Faucher-Giguère et al. 2011) or by some high energy feedback process like that produced by AGNs.

While this picture is attractive from a theoretical modelling standpoint, consistency with observational constraints need to be further improved. In this Letter we report on the first measurements of the SMHR from galaxies at $2 < z < 5$, using the VIMOS Ultra Deep Survey (VUDS, Le Fèvre et al. 2014). We use M_h derived from HOD modelling of the 2-point correlation function $w_p(r_p)$, and $M_\star$ obtained from SED fitting computed from ∼3000 galaxies to derive the SMHR for several galaxy samples, and compare it to SMHM models. The Letter is organized as follows: we summarize the VUDS data in Section 2, the M_h and $M_\star$ measurements are presented in Section 3, we derive the SMHR and the ISFE for several mass bins at z∼3 in Section 4, and we discuss our results in Section 5.

We use a flat ΛCDM cosmological model with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, and a Hubble constant $H_0 = 70$ km s^{-1} Mpc to compute absolute magnitudes and masses. We use $\Omega_b/\Omega_m = 0.155$ (Planck Collaboration et al. 2014).

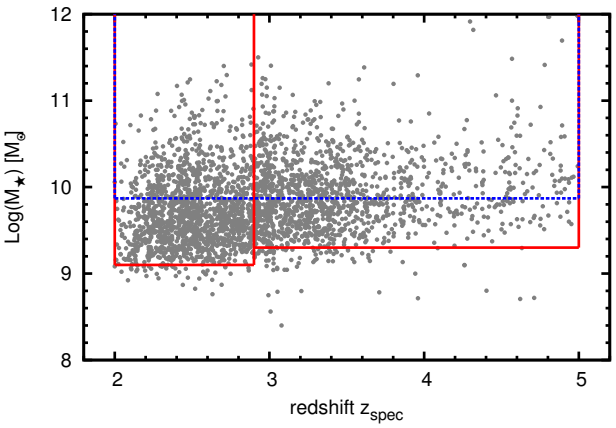


Fig. 1. The stellar mass distribution of selected VUDS sample. Vertical red solid lines indicates the boundaries of our redshift cuts, while the horizontal lines indicates the cuts in stellar mass. The dashed blue line indictate mass cut at $M_\star$=7.4×10^9 $M_\odot$ applied to define a high-mass sample (see text).

2. The VUDS Data

The VIMOS Ultra Deep Survey (VUDS) is a spectroscopic survey of ∼ 10,000 galaxies performed with the VIMOS multi-object spectrograph at the European Southern Observatory Very Large Telescope Le Fèvre et al. (2003). Its main aim is to study early phases of galaxy formation and evolution at $2 < z < 6$. Details about the survey strategy, target selection, as well as data processing and redshift measurements are presented in Le Fèvre et al. (2014).

We use data from two independent fields, COSMOS and VVDS-02h, covering a total area 0.81 deg^2. This corresponds to a volume ∼ 3×10^7Mpc^3 sampled in the redshift range $2 < z < 5$. The sample used here contains 3022 galaxies with reliable spectroscopic redshifts (spectroscopy reliability flags 2, 3, 4 and 9, see Le Fèvre et al. 2014). Galaxies have a stellar mass in the range $9 < log(M_*) < 11 M_\odot$ as presented in Figure 1. The whole sample has been divided into two redshift ranges: $2 < z < 2.9$ with $\log M_\star^{tresh} = 9.1 M_\odot$ and $2.9 < z < 5.0$ for which $\log M_\star^{tresh} = 9.3 M_\odot$, where $M_\star^{tresh}$ is the lower mass boundary of the sample resulting from the survey selection function (see below). Additionally in order to estimate the SMHR for more massive galaxies we define a subsample from galaxies in the redshift range $2 < z < 5$ and with $\log M_\star > 9.87 M_\odot$. This mass limit is the practical limit for which we can measure a galaxy correlation function signal accurately enough at each observed scale $0.1 < r_p < 10$ h^{-1}Mpc for the HOD fit to converge.

3. $M_\star$ and M_h measurements

The stellar masses in the VUDS Survey are estimated by performing SED fitting on the multi-wavelength photometry using the 'Le Phare' code (Ilbert et al. 2006), as described in details by Ilbert et al. (2013) and references therein.

Halo masses M_h are measured from a two-step process. First, the projected two-point correlation function $w_p(r_p)$ is computed for all three sub-samples in Durkalec et al. (2014). The correlation function results are then interpreted in terms of a three-parameter halo occupation model (HOD) of the form proposed by Zehavi et al. (2005) and motivated by

Table 1. The stellar mass to halo mass ratio (SMHR), and the integrated star formation efficiency (ISFE) in the VUDS survey

Redshift range	z_{mean}	Stellar mass range	$\log M_\star^{tresh}$	$\log M_h^{min}$	SMHR$\times10^{-2}$	ISFE [a]$\times10^{-2}$
[2.0 – 2.9]	2.50	[9.10 – 11.40]	$9.10^{+0.15}_{-0.16}$	$11.12^{+0.33}_{-0.36}$	$0.95^{+0.50}_{-0.35}$	$6.16^{+3.23}_{-2.26}$
[2.9 – 5.0]	3.47	[9.30 – 11.40]	$9.30^{+0.17}_{-0.19}$	$11.18^{+0.56}_{-0.70}$	$1.32^{+0.98}_{-0.57}$	$8.52^{+6.32}_{-3.68}$
[2.0 – 5.0][b]	3.00	[9.87 – 11.40]	$9.87^{+0.13}_{-0.15}$	$11.47^{+0.38}_{-0.43}$	$2.51^{+1.23}_{-0.89}$	$16.19^{+7.94}_{-5.74}$

[a] defined as SMHR$\times\Omega_m/\Omega_b$
[b] high mass sample

Kravtsov et al. (2004):

$$\langle N_g|M\rangle = 1+\left(\frac{M}{M_1}\right)^\alpha \quad \text{for} \quad M > M_{min} \qquad (1)$$
$$= 0 \qquad \text{otherwise,}$$

where M_{min} is the minimum mass needed for a halo to host one central galaxy, and M_1 is the mass of a halo having on average one satellite galaxy, while α is the power law slope of the satellite mean occupation function.

The correlation function measurements and model fitting procedures are described in Durkalec et al. (2014). By construction of the HOD model given in Eq. 2, the parameter M_{min} is the halo mass associated to galaxies with a stellar mass defined as the stellar mass threshold in the SHM relation (Zheng et al. 2005; Zehavi et al. 2005). We therefore quote the lowest mass of the sample considered as $M_\star^{tresh}$, as imposed by the survey limiting magnitude. The errors associated to this lower limit have been computed as the average of the systematic errors on $M_\star$ from the SED fitting for each redshift and mass sub-sample separately.

4. The stellar mass – halo mass relation at z∼3

Our results are presented in Table 1 and in the left panel of Figure 2. We find that for the low redshift sample z∼2.5 the stellar mass for halos of mass $\log M_h = 11.12 \pm 0.33 M_\odot$ is $\log M_\star = 9.1 M_\odot$, while at z∼3.5 the halo mass reaches $\log M_h = 11.18 \pm 0.56 M_\odot$ for a stellar mass $\log M_\star = 9.3 M_\odot$.

From these measurements we find that $\log(M_\star/M_h)$ is ranging from -2.02 ± 0.18 for the low mass sample and up to -1.6 ± 0.17 for the most massive sample, at a redshift $z \sim 3$. As shown in Figure 2 these results are compared to various measurements at low and intermediate redshift $z < 1$, obtained using different methods, including satellite kinematics (Conroy et al. 2007, More et al. 2011, weak lensing (Mandelbaum et al. 2006), galaxy clustering (Foucaud et al. 2010, Hartley et al. 2013), as well as abundance matching (Moster et al. 2013). Our measurements are in excellent agreement with the model derived from abundance matching at a redshift $z = 3$ (Moster et al. 2013).

5. Discussion and conclusion

Our SMHR measurements are one of the first to be performed at z∼3 from a clustering and HOD analysis, as made possible from the large VUDS spectroscopic redshift survey. We find that the SMHR is 1% to 2.5% for galaxies with intermediate stellar masses (at z∼3) ranging from $\sim10^9$ $M_\odot$ to $\sim7\times10^9$ $M_\odot$.

Following Conroy & Wechsler (2009) we compute the integrated star formation efficiency (ISFE) η=$M_\star/M_h/f_b$ with f_b the universal baryon fraction $f_b = \Omega_b/\Omega_m = 0.155$ (Planck collaboration, 2014). Results are reported in Table 1. We find that the ISFE range from $6.2^{+3.2}_{-2.3}$% to $16.2^{+7.9}_{-5.7}$% for galaxies with a stellar mass from $M_\star$=1.3×10^9 $M_\odot$ to $M_\star$=7.4×10^9 $M_\odot$. The IFSE therefore increases with stellar mass.

The star formation efficiency of ∼16% in a halo with M_h=3×10^{11} $M_\odot$ is quite close to a maximum of ∼20% occurring at 10^{12} $M_\odot$ in SMHM models Behroozi et al. (2010) (see also Moster et al. 2013). We use a simple mass growth model to derive the time scale for which our most massive galaxy sample would reach the maximum predicted in the SMHM relation from Behroozi et al. (2010). In this model the mass growth of DM haloes is described by a mean accretion rate $\langle \dot{M} \rangle_{mean}$ as proposed by Fakhouri et al. (2010), while galaxies grow in $M_\star$ via star formation using the median SFR for our sample from Tasca et al. (2014a), as well as through mergers with a constant accretion in stars of $\sim 1 M_\odot/yr$ (Tasca et al. 2014b). Because the halo accretion rate and SFR are dependent on redshift and halo/stellar mass, their values are computed every $\delta t = 5$ Myr to account for the changing mass and redshift.

We find that the SMHM relation would reach a maximum $\log(M_\star/M_h) \simeq -1.25$ about 350Myr after the observed epoch (i.e. at z∼2.6) and that at this time halo and stellar masses will be $M_h = 10^{11.6} M_\odot$ and $M_\star$=$2\times10^{10} M_\odot$ respectively. According to the model proposed by Moster et al. (2013) the SMHM relation is expected to reverse its trend after reaching the maximum point, but the slope of the function remains the same (see Figure 1). Therefore, assuming that the dark matter haloes continue to grow with time, the stellar mass accretion must be stopped or the dark matter accretion would need to increase with an accretion rate $\langle \dot{M} \rangle$ 10 times larger. There are no indications (e.g. in the N-body simulations), which would support the second scenario. On the other hand the stellar mass $M_\star = 2\times10^{10}$ computed at the maximum of SMHM relation is comparable to the 'quenching mass' as discussed in Peng et al. (2010). This result leads to the conclusion that, on average, the massive galaxy population with $M_\star^{min} \sim 0.5 - 1 \times 10^{10}$ $M_\odot$ observed at $z \sim 3$ will experience star formation quenching within a few hundred million years.

As discussed in this Letter, the SMHM is a simple yet efficient tool to probe star formation efficiency at the epoch of rapid stellar mass assembly provided one obtains robust measurements on both $M_\star$ and M_h, as is now possible with VUDS, complementing more indirect estimates using e.g. abundance matching. A more extensive exploration of the efficiency of star formation over a larger range of halo masses is becoming possible with new surveys coming on-line, and it would be interesting to probe higher masses than done in this paper to

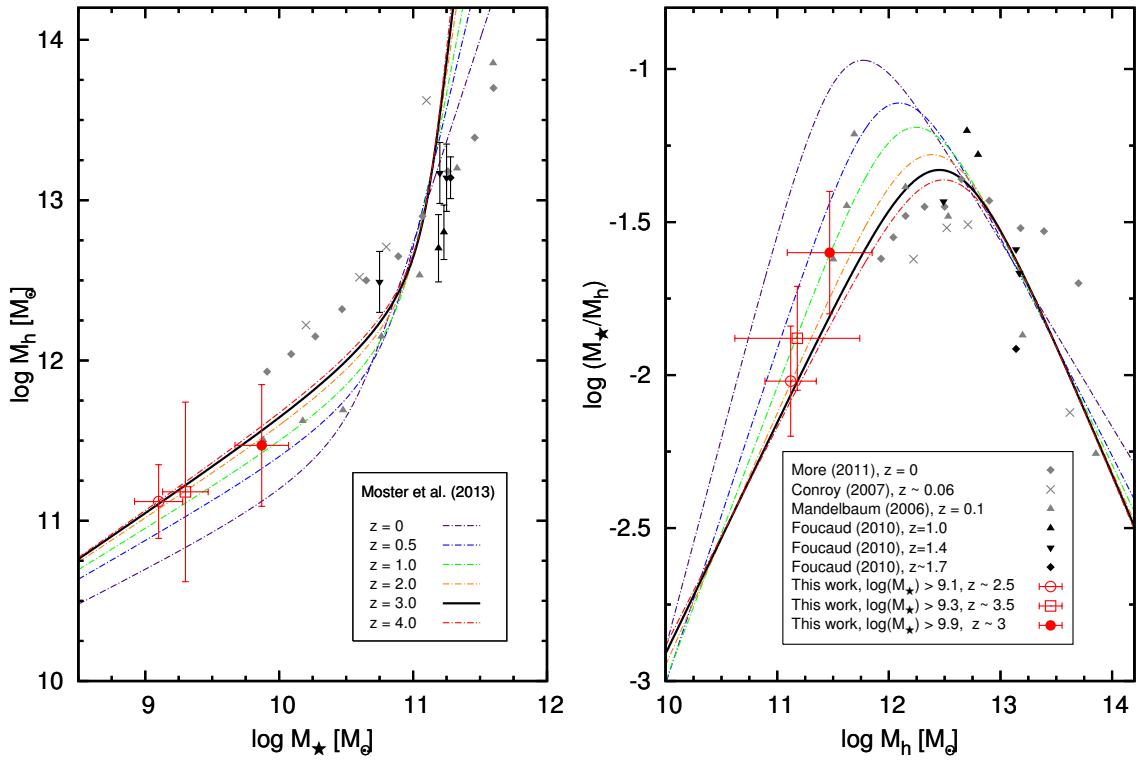


Fig. 2. *Left panel:* The relation between the stellar mass $M_\star$ and the halo mass M_h in the VUDS survey for different stellar mass and redshift cuts (red symbols). $M_\star$ is derived from SED fitting of the multi-wavelength photometric data using the known spectroscopic redshift; the error bar in $M_\star$ indicates the expected uncertainty of the SED fitting method. M_h is obtained from HOD modelling of the two-point correlation function in different redshift and mass ranges. The VUDS data is compared to measurements at low and intermediate redshift form satellite kinematics (SK, Conroy et al. 2007, More et al. 2011, weak lensing (WK, Mandelbaum et al. 2006), galaxy clustering (GC, Foucaud et al. 2010). The lines represents the model predictions derived from abundance matching at various redshift (Moster et al. 2013). *Right panel:* The halo mass M_h over stellar mass $M_\star$ ratio vs. halo mass at $z = 3$ in the VUDS survey. The colour scheme is the same as for the left panel.

evaluate the halo mass for which the star formation efficiency is the highest. Extending such measurements to higher redshifts will require the power of new facilities like Sumire-PFS, JWST or ELTs.

Acknowledgements. We thank Jean Coupon for interesting discussions. This work is supported by funding from the European Research Council Advanced Grant ERC-2010-AdG-268107-EARLY and by INAF Grants PRIN 2010, PRIN 2012 and PICS 2013. AC, OC, MT and VS acknowledge the grant MIUR PRIN 2010–2011. This work is based on data products made available at the CESAM data center, Laboratoire d'Astrophysique de Marseille, France.

References

Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379
Behroozi, P. S. & Silk, J. 2014, ArXiv e-prints
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 762, L31
Bertone, S., Stoehr, F., & White, S. D. M. 2005, MNRAS, 359, 1201
Béthermin, M., Wang, L., Doré, O., et al. 2013, A&A, 557, A66
Bromm, V., Yoshida, N., Hernquist, L., & McKee, C. F. 2009, Nature, 459, 49
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Conroy, C., Prada, F., Newman, J. A., et al. 2007, ApJ, 654, 153
Conroy, C. & Wechsler, R. H. 2009, ApJ, 696, 620
Dekel, A. & Birnboim, Y. 2006, MNRAS, 368, 2
Durkalec, A., Le Fèvre, O., Pollo, A., et al. 2014, ArXiv e-prints
Fakhouri, O., Ma, C.-P., & Boylan-Kolchin, M. 2010, MNRAS, 406, 2267
Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189
Faucher-Giguère, C.-A., Kereš, D., & Ma, C.-P. 2011, MNRAS, 417, 2982
Foucaud, S., Conselice, C. J., Hartley, W. G., et al. 2010, MNRAS, 406, 147
Hartley, W. G., Almaini, O., Mortlock, A., et al. 2013, MNRAS, 431, 3045
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35
Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1670–1681
Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2014, ArXiv e-prints
Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159
Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715
More, S., van den Bosch, F. C., Cacciato, M., et al. 2011, MNRAS, 410, 210
Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16
Silk, J. 2003, MNRAS, 343, 249
Tasca, L. A. M., Le Fèvre, O., Hathi, N. P., et al. 2014a, ArXiv e-prints
Tasca, L. A. M., Le Fèvre, O., López-Sanjuan, C., et al. 2014b, A&A, 565, A10
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791

APPENDIX B

HOD model code documentation

Software requirements

The HOD model code presented in this documentation has been developed for the PhD thesis work by Anna Durkalec. It has been created on GNU/Linux system and should run on it's architecture. However this code has not been tested on any other system than RedHat's Fedora and Debian's Ubuntu. It is recommended to use Fedora 16 (or higher) or Ubuntu 11.04 (or higher).

To proper installation required are:

- C/C++ compiler (e.g. gcc with g++ support)
- The GNU Scientific Library (GSL) (version 1.15 or higher) `http://www.gnu.org/software/gsl/`
- libconfig (version 1.14 or higher) `http://www.hyperrealm.com/libconfig/`

Optional, for automatic graphs created by code is required:

- gnuplot (version 4.6 or higher) `http://www.gnuplot.info/`

Installation

First unpack the gziped tar archive.

```
> tar xzf HODmodel_1.70.tar.gz
```

This creates the HODmodel_1.70 directory. Change to that directory and compile the code, run:

```
> make
```

If you face any problems with compilation please check if you have installed all required libraries. Also do not hesitate to contact me by email `anna.durkalec@lam.fr`.

Run

In order to run the code one need three configure files:

- `COSMO_config`.
- `HODparams_config`.
- `INPUTDATA_config`.

And one file with the input data (e.g. wp_rp.dat) results form the correlation function measurements

To run the code, change to directory with configure files and input data and type:

```
> /path/to/code/HODmodel_1.70
```

To get more familiar with the code you can run it on provided example data. Simply go to `/EXAMPLE/simple` or `/EXAMPLE/advanced` and type:

```
> ../../HODmodel_1.70
```

Configuration files

To proper run HOD model code one needs three configuration files.

- `COSMO_config` contains basic cosmology parameters.
- `HODparams_config` contains HOD parameters.
- `INPUTDATA_config` contains information about input data file and used mode.

Also necessary is to provide input data file with $w_p(r_p)$ measurements. The file format depends of chosen mode.

Table B.1: `COSMO_config`. Basic cosmology parameter file.

Z_REDSHIFT	z	Mean sample redshift
PI_MAX	π_{max}	Maximum allowed galaxy separation along the line-of-sight
OMEGA_M0	Ω_m	Matter density (cold dark matter + baryons)
OMEGA_C0	Ω_c	Cold dark matter density
OMEGA_B0	Ω_b	Baryon matter density
OMEGA_LAMBDA0	Ω_Λ	Dark energy density
OMEGA_R0	Ω_r	
H	h	Dimensionless Hubble parameter
N_S	n_s	Scalar power-spectrum index
SIG_8	σ_8	Power-spectrum normalization at small scales
DELTA_C	δ_c	Critical overdensity for spherical collapse
RHO_0	ρ_0	Main density of matter at present day in $[10^{10}\ h^2\ M_\odot\ Mpc^{-3}]$
DELTA_VIR	Δ_{vir}	Critical overdensity for virialization

Table B.2: `HODparams_config`. HOD parameters declared by user.

LOG_M_MIN	$\log_{10} M_{min}$	Minimum halo mass for central galaxies
LOG_M_1	$\log_{10} M_1$	Scale mass for satellite galaxies
LOG_M_0	$\log_{10} M_0$	Minimum halo mass for satellite galaxies
SIGMA_LOGM	$\sigma_{\log M}$	Dispersion for central galaxies
ALPHA	α	Slope of satellite occupation distribution

Table B.3: `INPUTDATA_config`. Information about input data.

MODE	1- the simple mode; 2 - the advanced mode.
WPRP_FILENAME	$w_p(r_p)$ data filename (ATTENTION! There is different file format required for different mode.
N_POINTS	Number of separation bins in the correlation function
RP_MIN	Minimum r_p value for model $w_p(r_p)$ output
RP_MAX	Maximum r_p value for model $w_p(r_p)$ output
N_BOOTS	Number of bootstrap samples (only for advanced mode)

Two modes

The code can work in two different modes depending on the input data you can provide. You declare mode in the 'INPUTDATA_config' file (MODE):

1. **MODE = 1; Simple mode.**
   In this mode the input file has three columns `'rp wprp wprp_err'`.

   The χ^2 is form as:
   $$\chi^2 = \sum_{i=1}^{N} \Big(\frac{w_p^{obs}(r_i) - w_p^{mod}(r_i)}{\sigma^i_{w_p}} \Big)^2 \tag{B.1}$$
   where N is the number of bins in the correlation function.

2. **MODE = 2; Advanced mode.**
   In this mode the input file has the format of `'*.wp'` file form correlation function code i.e. first row contains r_p values, second row associated $w_p(r_p)$ measurements, in the next rows associated bootstrap measurements.

   Based on the input data the code will compute σ_{w_p} and associated covariance matrix from bootstrap samples:
   $$\mathbf{C}_{ik} = \Big\langle \big(w_p^i(r_i) - \langle w_p^j(r_i)\rangle_j\big)\big(w_p^j(r_k) - \langle w_p^j(r_k)\rangle_j\big) \Big\rangle_j \tag{B.2}$$
   where $\langle\rangle_j$ indicates an average over all bootstrap.
   χ^2 is defined as:
   $$\chi^2 = \sum_{i,j}^{N_D} \big(w_p^{obs}(r_i) - w_p^{mod}(r_i)\big)\, \mathbf{C}_{ij}^{-1}\, \big(w_p^{obs}(r_j) - w_p^{mod}(r_j)\big) + \frac{\big(n_g^{obs} - n_g^{mod}\big)^2}{\sigma^2_{n_g}} \tag{B.3}$$

Output files

The code automatically create the directory `/HODmodel_1.70_output` in which user can find all output files. The output is slightly different for different mode.
The common files are:

- `wprp_model.out`
  It contains model results of $w_p(r_p)$ in the range declared by user (from RP_MIN to RP_MAX).

- `halo_function.out`
  It contains halo function for given HOD parameters.

- `deduced_parameters.out`
  It contains computed deduced parameters, such as mean galaxy bias, mean halo mass, satellite and central galaxies fraction along with associated errors.

- `output_graph.gp` and `graph.eps`
  The first file is the script for `gnuplot`. The second one is the graphical representation of received results. *Left panel:* Two-point projected correlation function. Black points represent data. Blue solid line represents model results. *Right panel:* Correspond halo function.

Additionaly in the **advanced mode**:

- `wprp.dat`
  It contains first two lines form input data file.

- `bootstrap.dat`
  It contains bootstrap samples from input data file.

- `wprp_and_err.out`
  It contains $w_p(r_p)$ data with computed errors in the last column.

- `covariance.out`
  It contains covariance matrix computed by using eq.B.2.

- `incov.out`
  It contains inverted covariance matrix.

List of functions

Table B.4: List of the most important functions used to model projected two-point correlation function $w_p(r_p)$. In alphabetical order.

Function name	Short description
bias_nl (σ)	Scale dependent halo bias function. Tinker et al. (2005) eq. B7 .
central_g (M)	Central galaxies fraction. Coupon et al. (2012) eq. (11).
density_profile_lambda (r,M)	Convolution of NFW profile. Sheth et al. (2001) eq. A1. ATTENTION: This function returns λ_{NFW}/M^2.
density_profile_rho (r,M)	NFW density profile. Navarro et al. (1997).
dsigma_dR_interp (R)	Derivative $\mathrm{d}\sigma/\mathrm{d}R$. Important for mass function.
find_Mstar (void)	This function finds the characteristic mass $M_\star$.
fnlps_interp (r)	Interpolated Fourier transform of nonlinear matter power spectrum
fourier_density_profile (k,M)	Fourier transform of NFW profile Sheth et al. (2001) eq. 81.
fourier_nonlinear_ps (r)	Fourier transform of evolved nonlinear power spectrum.
halo_occupation (M)	Halo occupation function.
init_fnlps ()	Initialization for fourier transform of nonlinear matter power spectrum interpolation.
init_sigma (R)	Initialization for sigma interpolation.
init_xi()	Initialization for real-space correlation function interpolation.
linear_power_spectrum_KIL (k)	Linear matter power spectrum from HALOFIT+ code (normalized).
mass_function (M)	Halo mass function. Tinker et al. (2008) Eq.3. ATTENTION: Function returns $M * dn/dM$.
nonlinear_power_spectrum (k)	Nonlinear Matter Power Spectrum. From HALOFIT+ code.
number_density ()	This function returns number density of galaxies n_g.
number_density_p_spher($M_{lim}(r)$)	Restricted number density. Spherical approach. Tinker et al. (2005) Eq. (B6) .
one_halo_xi (r)	One halo term. It returns $1 + \xi_{1h}(r) = [1 + \xi_{cs}(r)] + [1 + \xi_{ss}(r)]$.
satellite_g (M)	Satellite galaxies fraction. Coupon et al. (2012) Eq. (12)
sigma_interp (R)	Interpolated $\sigma(R)$ evolved and normalized.
two_halo_P (k)	Two halo power spectrum. Tinker et al. (2005) Eq. B5.
wp_rp_ANIA(rp)	Projected correlation function $wp(rp)$.
xi_cs (r)	This function returns central-satellite fraction of one halo term $1 + \xi_{cs}$.

Continued on next page

Table B.4 – continued from the previous page

Function name	Short description
xi_interp(r)	Interpolation of real space correlation function. It returns $\xi(r) = [1+\xi_{1h}(r)] + [1+\xi_{2h}(r)] - 1$.
xi_ss (r)	This function returns satellite-satellite fraction of one halo term $1+\xi_{ss}$.

Table B.5: List of the initial functions. In alphabetical order.

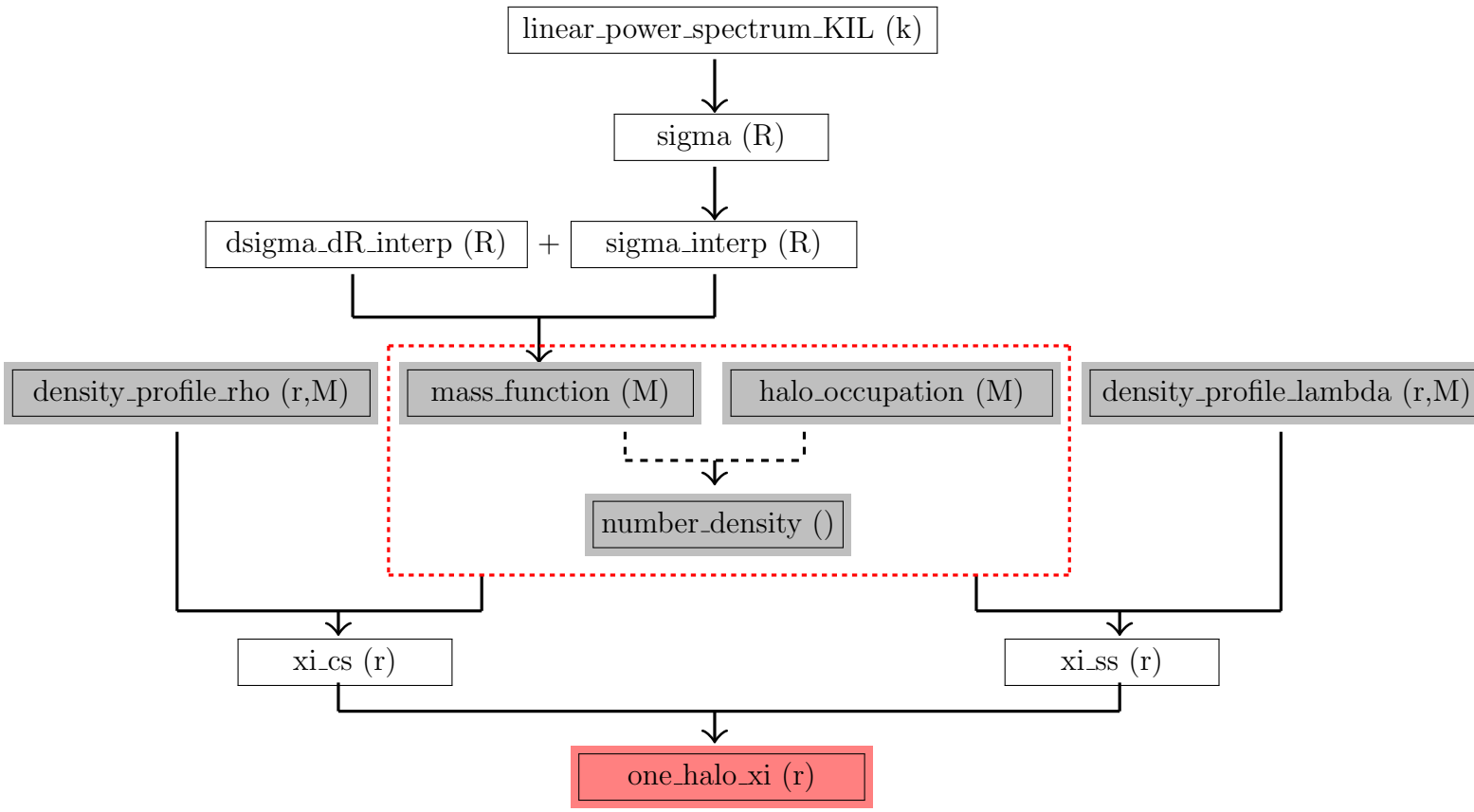
Function name	Short description
chi_square()	Compute χ^2 in the advanced mode.
chisq_inne_easy()	Compute χ^2 in the simple mode.
cov_code()	Used in the advanced mode. Compute errors for input $w_p(r_p)$ data, covariance matrix and inverted covariance matrix. Returns `covariance.out` and `invcov.out`.
graphs()	Create and returns `output_graph.gp` file.
prepare()	Reads input $w_p(r_p)$ data file in the advanced mode and returns `wprp.dat` and `bootstrap.dat`.
readingConfigFile_ANIA()	Reads the `COSMO_config` parameter file.
readingHODparams()	Reads the `HODparams_config` parameter file.
readingInfo()	Reads the `INPUTDATA_config` parameter file.
readInputData()	Reads input $w_p(r_p)$ data file in the advanced mode.
readInputData_easy()	Reads input $w_p(r_p)$ data file in the simple mode.

Table B.6: List of the functions used for compute external parameters. In alphabetical order.

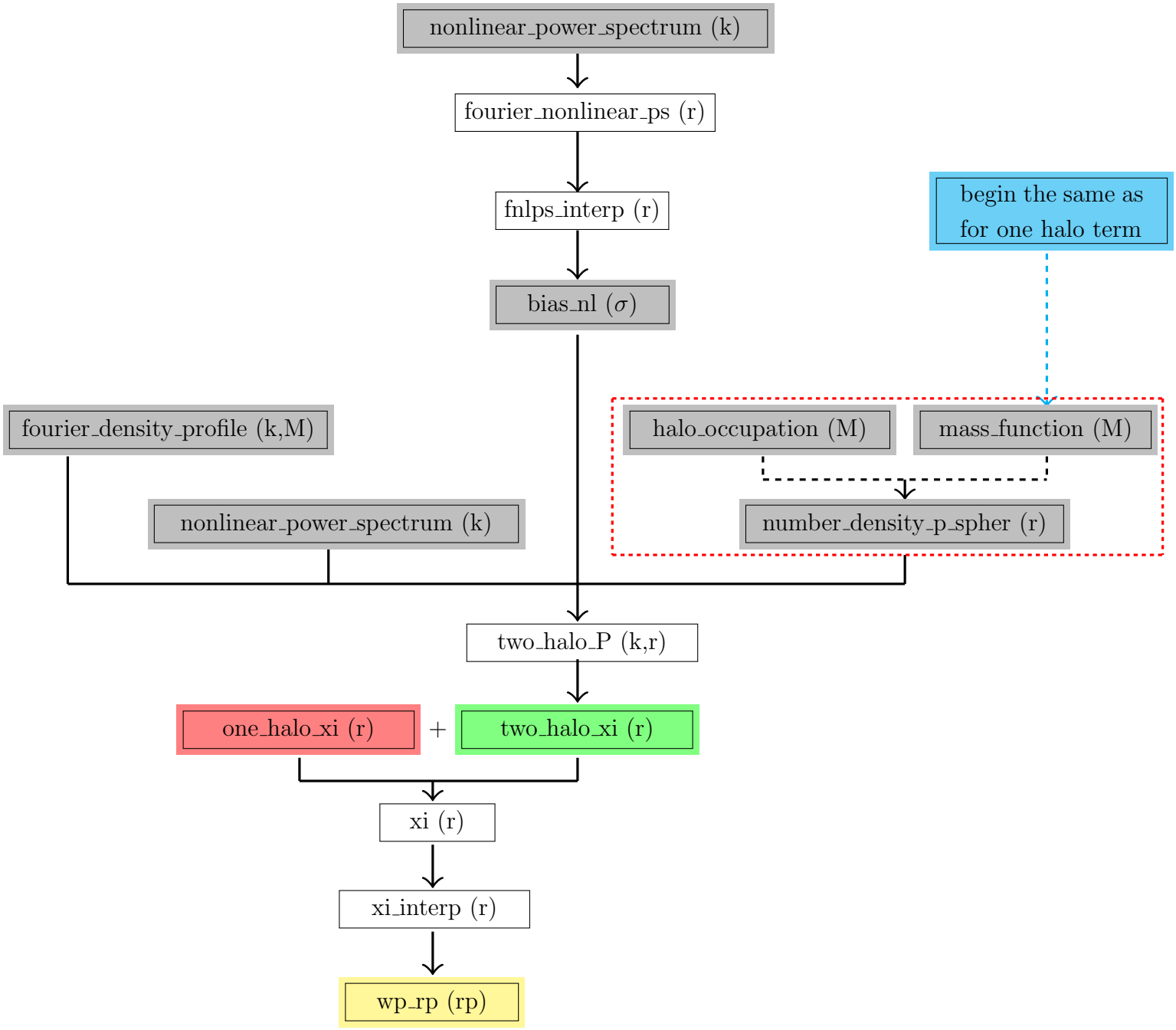
Function name	Short description
dedpar_cen_frac()	The fraction of central galaxies per halo. Coupon et al. (2012) Eq. (16).
dedpar_galaxy_bias()	Mean galaxy bias at redshift z. Coupon et al. (2012) Eq. (13).
dedpar_mean_halomass()	Mean halo mass at redshift z. Coupon et al. (2012) Eq. (15).

Flow charts

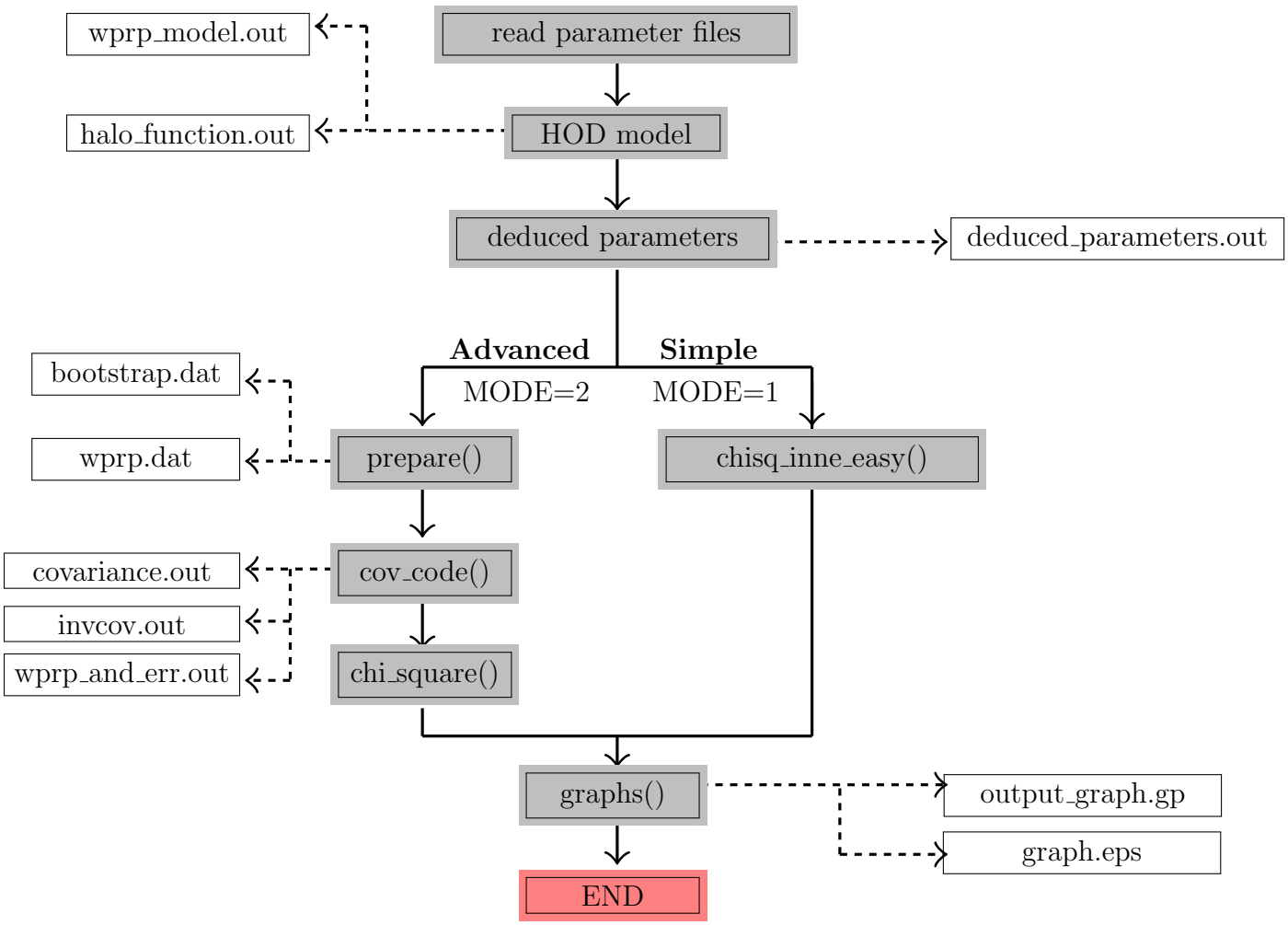
One halo term $\xi_{1h}(r)$



Two halo term $\xi_{2h}(r)$



Main function



List of Figures

1.1 Cosmic Microwave Background Radiation (CMB) as seen by ESA's Planck satellite. 6
1.2 A schematic representation of a 'merger tree' depicting the growth of a halo as the result of a series of mergers. 12
1.3 Linear growth rate of density fluctuations for different cosmologies. 15
1.4 Linear matter power spectrum $P(k)$ as a function of wavenumber k from various measurements of cosmological structure. 17
1.5 The spatial distribution of $\sim 80\,000$ galaxies in the 2dFGRS. 18

2.1 The distribution of absolute NUV magnitude (*top*) and stellar mass (*bottom*). 25
2.2 Comparison of the redshift distribution of the VUDS survey (presented in grey) with other published spectroscopic surveys at $z > 2$. 28

3.1 The real-space correlation function $w_p(r_p)$ measured for data from two independent fields. Comparison of two estimating methods. 42
3.2 The lay-out of VIMOS multiobject spectograph at the European Southen Observatory Very Large Telescope field of view. 44
3.3 Distribution of the VUDS galaxies in the rest frame angular coordinates of their pointings. 44
3.4 Example of allocation of spectra slits on one of the VIMOS quadrants . . . 45
3.5 The dispersion of projected correlation function observed for different number of objects generated in the random catalog. 48
3.6 Projected two point correlation function $w_p(r_p)$ measured for 66 VUDS mock catalogs in three redshift ranges. 49

3.7 HOD fit tests preformed on the correlation function measurements for two selected luminosity threshold subsamples $M_r < -19.5$ (*upper panel*) and $M_r < -21.5$ (*lower panel*). 54

4.1 The redshift distribution of the used VUDS galaxy sample for redshift range $2.0 < z < 5.0$. The filled gray histogram represents the total sample of reliable galaxies (with $zflag = 2 - 4, 9$, $\sim 80\%$ reliability), while the red and blue histograms represent the contribution from COSMOS and VVDS2h fields respectively. 60

4.2 The distribution of galaxies with spectroscopic redshifts $2 < z < 5$ in two independent VUDS fields: COSMOS on the *upper* panel and VVDS-02h on *lower* panel. The blue squares indicates VIMOS pointing centers. 62

4.3 Influence of the possible incorrect redshift measurements on the correlation function. 64

4.4 Cosmic variance in two VUDS fields: COSMOS and VVDS-02h. 66

4.5 The projected two-point correlation function $w_p(r_p)$ and he associated best-fit power-law parameters r_0 and γ along with error contours for the general galaxy population separated in three redshift ranges 68

4.6 The correlation function for three redshift sub-samples as in Figure 4.5 and the evolution of the halo occupation for the three redshift ranges. 70

4.7 Correlation length r_0 as a function of redshift from various surveys of different types of objects. 72

4.8 The evolution of the number-weighted average host halo mass given by Eq. 3.31 for the three redshift ranges analysed in this study. 74

4.9 Large-scale linear galaxy bias retrieved by using two methods. 78

5.1 The distribution of absolute U-band magnitudes (top) and stellar mass (bottom) in used galaxy sample. 84

5.2 The projected two-point correlation function $w_p(r_p)$ and the associated best-fit parameters r_0 and γ along with error contours for the magnitude-threshold sub-samples in two redshift ranges 85

5.3 The projected two-point correlation function $w_p(r_p)$ and the associated best-fit parameters r_0 and γ along with error contours for the stellar mass-threshold sub-samples in two redshift ranges 86

6.1 The stellar mass distribution of selected VUDS sample. Vertical red solid lines indicates the boundaries of our redshift cuts, while the horizontal lines indicates the cuts in stellar mass. The dashed blue line indictate mass cut at $M_\star$=7.4×10^9 $M_\odot$ applied to define a high-mass sample (see text). 90
6.2 The relation between the stellar mass $M_\star$ and the halo mass M_h at z=3 in the VUDS survey (red points) 92
6.3 The halo mass M_h over stellar mass $M_\star$ ratio vs. halo mass at $z = 3$ in the VUDS survey. 94

List of Tables

1.1 Current values of the cosmological parameters of the ΛCDM model and the 68% confidence limits for constrained parameters. 10

2.1 The VUDS fields. 22
2.2 VUDS spectroscopic target selection criteria. 24

3.1 Tests of HOD fitting procedure at low redshift on Luminosity-threshold Samples. 52
3.2 Tests of HOD fitting procedure on intermediate redshift range on Luminosity-threshold Samples. 55

4.1 The number of galaxies in VUDS fields for each reliability flag in the redshift range $2 < z < 5$, as used in this study. 60
4.2 Properties of general population VUDS samples in the redshift range $2.0 < z < 5.0$. 68
4.3 The HOD and and HOD-based parameters for the redshift sub-samples . . 70

5.1 Properties of luminosity limited VUDS sub-samples in the redshift range $2.0 < z < 5.0$ 85
5.2 Properties of stellar mass limited VUDS sub samples in the redshift range $2.0 < z < 5.0$ 86

6.1 The stellar mass to halo mass ratio (SMHR), and the integrated star formation efficiency (ISFE) in the VUDS survey 93

B.1 `COSMO_config`. Basic cosmology parameter file. 181

B.2 `HODparams_config`. HOD parameters declared by user. 181
B.3 `INPUTDATA_config`. Information about input data. 182
B.4 List of the most important functions used to model projected two-point correlation function $w_p(r_p)$. In alphabetical order. 184
B.5 List of the initial functions. In alphabetical order. 185
B.6 List of the functions used for compute external parameters. In alphabetical order. 185