Star-forming galaxies as tools for cosmology in new-generation spectroscopic surveys

by

Ginevra Favole

a dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Astrophysics

Instituto de Astrofísica de Andalucía (IAA) – CSIC
Granada, May 2016

Ph.D. Advisor:
Prof. Francisco Prada Martínez

Abstract

This Ph.D. thesis is a collection of clustering studies in different galaxy samples selected from the Sloan Digital Sky Survey and the SDSS-III/Baryon Oscillation Spectroscopic Survey. By measuring the two-point correlation function of galaxy populations that differ in redshift, color, luminosity, star-formation history and bias, and using high-resolution large-volume cosmological simulations, I have studied the clustering properties of these galaxies within the large scale structure of the Universe, and those of their host dark matter halos. The aim of this research is to stress the importance of star-forming galaxies as tools to perform cosmology with the new generation of wide-field spectroscopic surveys. Among the galaxies considered, I have focused my investigation on a particular class whose rest-frame optical spectra exhibit strong nebular emission lines. Such galaxies, better known as Emission-Line Galaxies (ELGs), will be the main targets of near-future missions – both ground-based, as the Dark Energy Spectroscopic Instrument, the 4-metre Multi-Object Spectroscopic Telescope, the Subaru Prime Focus Spectrograph, and space-based as EUCLID. All these surveys will use emission-line galaxies up to redshift $z \sim 2$ to trace star formation and to measure the Baryon Acoustic Oscillations as standard ruler, in the attempt to unveil the nature of dark energy. Therefore, understanding how to measure and model the ELG clustering properties, and how they populate their host dark matter halos, are fundamental issues that I have addressed in this thesis by using state-of-the-art data, currently available, to prepare the clustering prospects and theoretical basis for future experiments.

Resumen

Esta tesis doctoral presenta una colección de estudios del agrupamiento (i.e. clustering) de las galaxias en la estructura a gran escala del Universo en diferentes muestras seleccionadas de los catálogos de galaxias del Sloan Digital Sky Survey y del SDSS-III/Baryon Oscillation Spectroscopic Survey. Midiendo la función de correlación de dos puntos en las poblaciones de galaxias con diferente corrimiento al rojo, color, luminosidad, proceso de formación estelar y bias, he estudiado, utilizando simulaciones cosmológicas de alta resolución y gran volumen, las propiedades de su agrupamiento dentro de la estructura a gran escala del Universo y de los halos de materia oscura en los que residen dichas galaxias. El objetivo de esta investigación es enfatizar la importancia de las galaxias con formación estelar como instrumentos para las medidas cosmológicas en los grandes cartografiados espectroscópicos de nueva generación. Entre las galaxias seleccionadas, he enfocado mi estudio en un tipo particular cuyos espectros muestran líneas de emisión nebular. Dichas galaxias, denominadas ELGs, serán las fuentes principales que observarán los nuevos proyectos, tanto desde tierra, como son el Dark Energy Spectroscopic Instrument, el 4-metre Multi-Object Spectroscopic Telescope, the Subaru Prime Focus Spectrograph, y desde el espacio como EUCLID. Todos estos cartografiados utilizarán galaxias con líneas de emisión hasta redshift $z \sim 2$ como indicadores de formación estelar y para medir las oscilaciones acústicas bariónicas como medida de distancia, y así poder conocer la naturaleza de la energía oscura. Por lo tanto, entender cómo medir y reproducir teóricamente el agrupamiento de las ELGs, y cómo éstas galaxias pueblan sus halos, son puntos fundamentales que he estudiado en esta tesis utilizando los datos actuales para preparar las bases teóricas y el estudio de sistemáticos de cara a los experimentos futuros.

Contents

A mia madre

Acknowledgments

I wish to thank my Supervisor, Prof. Francisco Prada Martínez, for guiding me in the preparation of this thesis. I am grateful to Prof. Daniel Eisenstein, Prof. David Schlegel and Dr. Cameron McBride for the time and the efforts spent helping and hosting me during my stays in the US. It has been a pleasure working with outstanding mentors that made me realize the importance of a global vision in Science. Thank you.

Thanks to Sergio, Antonio, and Johan, who made my work much easier, and to my IFT/UAM colleagues and friends: Elisabetta, Doris, Franco, Irene, Víctor, Domenico, Susana, Paolo, Miguel P., Arianna, Jesus, Santiago, Edoardo, Sebastian, Federico S., Federico C., Miguel M., Ander, Mario, Gianluca, Ana, Josu, Xabier. Thank you for sharing coffees, lunches, beers, good moments ... and for making me smile in the bad ones.

I wish to thank Mattia, who was the first one supporting me even if I didn't know him yet. Thanks for convincing me not to throw in the towel. You were right, it has been worth. Thanks to Fabio, the italian pillar of the IAA group.

Gracias a Gabriella y Mirjana por los días de risas en La Rijana. Gracias a Chloé, Naike y Pika por ser la mejor casa de todo el Realejo.

Grazie alla mia famiglia italo-granaina, Federica, Matteo e Lorena, che mi ha adottata fin da subito senza riserve, aiutandomi a scoprire la Graná che mi piace tanto. Grazie per farmi sentire sempre a casa ... in placeta de la Parra numero 0.

Grazie a Viviana per aver condiviso la mia prima vera casa madrileña, un gruppo di amici affiatati, pensieri e viaggi memorabili. Grazie per aver contribuito a rendere Zurita un posto speciale. Grazie a Francesco e Marco, i migliori consiglieri e agenti immobiliari di Madrid. Grazie a Davide per ricordarmi sempre di *prendere la vita con leggerezza, che leggerezza non é superficialitá, ma planare sulle cose dall'alto, non avere macigni sul cuore.*

Grazie a Valentina (e a Skype), per capirmi cosí bene e per esserci sempre. Grazie a Stefania per il suo modo autentico e schietto di vedere le cose. Grazie a Camilla per gli anni passati, per quelli che verranno e per i progetti pazzi che fanno sempre bene al cuore.

Grazie a Michele per la sua gentilezza e le cene sicule a Boston. Grazie a Paolo per l'amicizia ritrovata e la foto in copertina!

Gracias a Itziar por ser una gran amiga y una gran tanguera. Gracias a Miguel por las rutas con la bici y la lista de documentales que nunca terminaré de ver. Gracias a la comunidad madrileña del tango sin la cual los ultimos años no habrían sido tan agradables. Gracias por las lindas tardes de templete, abrazos y risas. Sobretodo gracias por ayudarme a descubrir no solo un baile, sino un modo de pensar y de vivir.

Gracias a Lourdes, Javier y Pedro por abrirme su casa y acogerme como una hija y una hermana más.

Grazie alla mia famiglia italiana e a mio padre che, nonostante la distanza e le differenze, so che fa il tifo per me e mi supporta sempre. Grazie per aver contribuito a farmi arrivare qui.

Gracias a Ramon, Pessoa y Mono por hacer grandísima una casa minúscula y gracias a mi nueva familia, Pablo, por ayudarme a entender donde quiero estar y por ser dos conmigo.

Thanks to all of you!

1

A golden decade for cosmology

1.1. Introduction

There is an increasing tight connection between cosmology and particle physics that motivates understanding the fundamental properties of the Universe and justifies the development of major experiment facilities. Unveiling the nature of the dark Universe is one of the top big questions facing science over the next quarter-century. Furthermore, even those aspects for which the standard cosmological model provides a straightforward and adequate description, still pose challenging questions e.g., the biasing of the galaxies with respect to the matter distribution remains a source of uncertainty. At the same time, a more precise determination of the underlying cosmological parameters is needed to be able to asses accurately the level of agreement with those determined from the cosmic microwave background. Currently, the main goal of cosmology and astrophysics is to constrain the nature of dark matter, dark energy, and to test the predictions of the inflationary model, which could explain how the Large-Scale Structure of the Universe formed and hierarchically grows through gravitational instability. We live in a golden decade for cosmology: in the last ten years, we have experienced an unprecedented development of large spectroscopic redshift surveys facilities, together with the theoretical and computational tools for the data interpretation as N-body cosmological simulations or Semi-Analytic Models (SAMs) of galaxy formation. The first edition of the Sloan Digital Sky Survey (SDSS), the SDSS-III/Baryon Oscillation Spectroscopic Survey (BOSS), the ongoing SDSS-IV/eBOSS, and the near-future Dark Energy Spectroscopic Instrument (DESI) and EUCLID surveys are critical to achieve reliable results in all these areas. Spectroscopy is key to further astrophysical understanding. In fact, most of the fundamental physical parameters we observe (velocity, kinematics, temperature, gravity/mass, ionization state, chemical abundance, age, ...) are only feasible with spectroscopy. On the other hand, high-resolution large-volume cosmological simulations have been essential for analysing galaxy surveys to understand the properties of dark matter halos in the standard Lambda Cold Dark Matter cosmology and the growth of structure. Simulations are also an invaluable tool for studying the abundance and evolution of galaxies, their distribution and clustering properties, understanding the galaxy-halo connection and necessary for testing different cosmological models.

1.2. The cosmological framework

The standard cosmological model is based on one single assumption, confirmed by a number of observations that, on a sufficiently large scale, the Universe is isotropic and homogeneous. The Einstein field equation [e.g., 317, 227, 224, 84],

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

allows one to apply the laws of General Relativity to the matter (and energy) content of the Universe, that is to specify its dynamical state as a whole. In the expression above,

$R_{\mu\nu}$ is the *Ricci tensor*, describing the local curvature of the space-time, $g_{\mu\nu}$ the metric, R the curvature scalar, Λ the cosmological constant and $T_{\mu\nu}$ the energy-momentum tensor [see e.g., 200]. In the case of a homogeneous and isotropic Universe, the metric assumes a simple form, known also as the Friedmann-Lemaitre-Robertson-Walker metric, which can be regarded as the generalization of spherical coordinates (r, θ, ϕ) embedded in a 4 dimensional space [317, 227]:

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

This metric connects the proper distance element ds to the comoving coordinates (r, θ, ϕ), the curvature K, the time t, and the scale factor $a(t)$.

In the case of an isotropic and homogeneous Universe, the Einstein field equation with the above metric leads to the Friedmann equations [317, 227, 224, 84],

$$H^2(t) \equiv \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3P}{c^2} \right) + \frac{\Lambda c^2}{3}, \tag{1.4}$$

where $H(t)$ is the Hubble parameter, ρ is the energy density in units of c^2, and P is the pressure. By replacing $\rho \to \rho - \Lambda c^2/(8\pi G)$ in Eq. 1.3, it is immediately seen that there exists a critical density ρ_c for which the curvature is zero, i.e. $K = 0$, and this is

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G}. \tag{1.5}$$

A Universe whose density is above this critical value will have a positive curvature, that means it is spatially closed; a Universe with density below this critical threshold will have negative curvature and will be spatially open. Under the hypothesis that the Universe is an ideal adiabatic gas with pressure P and energy density ρ, we can write the continuity equation

$$\frac{d\rho}{da} + 3\left(\frac{\rho + P/c^2}{a}\right) = 0, \tag{1.6}$$

that describes how the energy density and pressure are related to one another, and how they evolve for any given component of the Universe (i.e. matter, radiation, etc ...). The relations in Eqs. 1.3 and 1.6 allow to determine the evolution with time of the fundamental parameters $a(t)$, $\rho(t)$ and $P(t)$, once a set of initial conditions is established.

According to the cold dark matter model with cosmological constant (ΛCDM; see Section 1.3), the Universe is expanding at an accelerated rate due to the presence of a "negative pressure", the dark energy, whose nature is still unknown. The evolution of this energy density is driven by the equation of state [e.g., 224, 84]

$$P = wc^2\rho \tag{1.7}$$

where, in the most general case, the parameter w is some arbitrary function of the scale factor, $w = w(a)$, with the constraint that $w \leq 0$, i.e. negative pressure. Using Eqs. 1.6 and 1.7, one can write the evolution of the energy density as [224, 84]

$$\rho(a) = \rho_0 \exp\left(-3\int_1^a [1 + w(a')]d(\ln a')\right). \tag{1.8}$$

If $w(a) = constant$, then $\rho \propto a^{-3(1+w)}$. For non-relativistic matter, including both dark matter and baryons, $w = 0$ and $\rho \propto a^{-3}$. For radiation, $w = 1/3$ and $\rho \propto a^{-4}$. In the special case of $w = -1$, $\rho \propto a^0 = constant$, and since the scale factor increases, the term Kc^2/a^2 in Eq. 1.3 will eventually become negligible with respect to the others, leading to the functional form [224, 84]

$$a(t) = a(t_0)\exp\left(\sqrt{\frac{8\pi G\rho}{3}}t\right) = a(t_0)\exp\left(\sqrt{\frac{\Lambda c^2}{3}}t\right), \tag{1.9}$$

for the scale factor, where $\Lambda = 8\pi G\rho/c^2$ is the cosmological constant or "vacuum energy".

Due to the Universe expansion, objects that are far away from us appear smaller and fainter than objects that are closer, and the Hubble parameter represents the constant of proportionality between their distance d and recession velocity v:

$$v = H_0\, d, \tag{1.10}$$

where $H_0 = 100\, h\, \mathrm{km\ s^{-1} Mpc^{-1}}$ is the Hubble parameter evaluated at present in a given cosmology. The dynamical properties of the Universe (i.e., mass density ρ and cosmological constant Λ) enter the definition of comoving distance of an object through the dimensionless density (i.e., $\Omega = \rho/\rho_{crit}$) parameters [e.g., 227, 224, 84]

$$\begin{aligned} \Omega_M &\equiv \frac{8\pi G \rho_0}{3H_0^2} \\ \Omega_\Lambda &\equiv \frac{\Lambda c^2}{3H_0^2} \\ \Omega_K &\equiv 1 - \Omega_M - \Omega_\Lambda, \end{aligned} \tag{1.11}$$

where the subscript “0” indicates that these quantities are evaluated at the present epoch, and Ω_K represents the density curvature, which in a flat Universe is zero. The critical density required for a flat Universe is $\rho_{crit} = 3H^2/(8\pi G)$ which is about 9×10^{-30} g cm^{-3} today.

Distances in cosmology are commonly expressed in terms of the redshift z. This quantity is the fractional doppler shift of its emitted light resulting from its radial motion and is defined as [e.g., 227, 142]

$$z \equiv \frac{\lambda_o}{\lambda_e} - 1 = \frac{a(t_o)}{a(t_e)} - 1, \tag{1.12}$$

where λ_e is the wavelength emitted at time t_e, when the size of the Universe was $a(t_e)$, and λ_o is the observed one at t_o, when the size of the Universe is $a(t_o)$. Redshift is also related to the radial velocity v by

$$z = \sqrt{\frac{1 + v/c}{1 - v/c}} - 1. \tag{1.13}$$

For small v/c, or small distance d, the velocity is proportional to the distance and, in linear approximation, one has

$$z \approx \frac{v}{c} = \frac{d}{D_h}, \tag{1.14}$$

where $D_h \equiv c/H_0 = 3000\, h^{-1}\mathrm{Mpc}$ is the Hubble distance.

The comoving distance between fundamental observers, i.e. observers that are comoving with the Hubble flow, does not change with time, as it accounts for the expansion of the Universe. It is obtained by integrating the proper distance elements of nearby fundamental observers along the line of sight (LOS). The comoving distance from us ($z = 0$) of an astronomical object at redshift z will be [317, 314, 227]:

$$D_C = D_H \int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda}}. \tag{1.15}$$

Two comoving objects at redshift z that are separated by an angle $\delta\theta$ on the sky are said to have the distance $\delta\theta D_M$, where the transverse comoving distance D_M is related to the line-of-sight comoving distance D_C by [317, 314, 227, 142]

$$D_M = \begin{cases} D_H \dfrac{1}{\sqrt{\Omega_K}} \sinh(\sqrt{\Omega_K} D_C/D_H) & \text{if } \Omega_\mathrm{K} > 0, \\ D_C & \text{if } \Omega_\mathrm{K} = 0, \\ D_H \dfrac{1}{\sqrt{|\Omega_K|}} \sin(\sqrt{|\Omega_K|} D_C/D_H) & \text{if } \Omega_\mathrm{K} < 0. \end{cases} \tag{1.16}$$

Using the quantities above, one can define the angular diameter distance D_A, which is the ratio of the transverse proper distance of an object to its apparent angular size, and is used to convert angular separations in telescope images into proper separations at the source. It is related to the transverse comoving distance by [317, 314, 227, 142]

$$D_A = \frac{D_M}{1+z}. \tag{1.17}$$

Analogously, we define the luminosity distance D_L as the relationship between the bolometric (i.e., integrated over all frequencies) luminosity L and the bolometric flux F measured on the Earth:

$$F = \frac{L}{4\pi D_L^2}. \tag{1.18}$$

This is linked to the transverse comoving distance and the angular diameter distance defined above by [317, 314, 227, 142]

$$D_L = (1+z)D_M = (1+z)^2 D_A. \tag{1.19}$$

If the concern is not with bolometric quantities, but rather with differential flux F_ν and luminosity L_ν, as is usually the case in astronomy, then the $K-$correction [142, 143, 35], must be applied to the flux or luminosity because the redshifted object is emitting flux in a different band than that in which we are observing. The $K-$correction depends on the spectrum of the object in question, and is unnecessary only if the object has spectrum $\nu L_\nu =$ constant. For any other spectrum, the differential flux is related to the differential luminosity by [314, 227, 142]

$$F_\nu = (1+z)\frac{L_{(1+z)\nu}}{L_\nu}\frac{L_\nu}{4\pi D_L^2}, \tag{1.20}$$

where the ratio of luminosities equalizes the difference in flux between the observed and emitted bands, and the factor of $(1+z)$ accounts for the redshifting of the bandwidth. Similarly, for differential flux per unit wavelength we have [314, 227, 142]:

$$F_\lambda = \frac{1}{(1+z)}\frac{L_{\lambda/(1+z)}}{L_\lambda}\frac{L_\lambda}{4\pi D_L^2}. \tag{1.21}$$

Another useful quantity is the distance modulus DM defined by

$$DM \equiv 5\log\left(\frac{D_L}{10\,\mathrm{pc}}\right), \tag{1.22}$$

which represents the magnitude difference between the observed bolometric flux of an object and what it would be if it were at 10 pc.

The absolute magnitude M of an astronomical object is defined to be the apparent magnitude the object in question would have if it were located at 10 pc, that is

$$M - 5\log h = m - DM(z, \Omega_m, \Omega_\Lambda, h) - K(z), \tag{1.23}$$

where K is the $K-$correction given by [142]

$$K = -2.5\log\left[(1+z)\frac{L_{(1+z)\nu}}{L_\nu}\right] = -2.5\log\left[\frac{1}{(1+z)}\frac{L_{\lambda/(1+z)}}{L_\lambda}\right]. \tag{1.24}$$

1.3. The ΛCDM model

The fundamental assumption of a homogenous Universe in the Friedmann-Lemaitre-Robertson-Walker (FLRW) model has a natural antagonist: on smaller scales the Universe is evidently highly non-homogenous, manifesting this phenomenon in a beautiful variety of structures, ranging from large clusters of galaxies many Mpc wide to stars, planets and life. This requires that small perturbations in the density of matter were already present since the very first moments after the Big-Bang, perturbations which have then grown with time, leading to the formation, throughout hierarchical clustering, of the large scale structure we see today in the Universe [e.g., 315, 189, 244]. The collisionless, cold, purely gravitational growth of these instabilities in the density field of a kind of matter still undetected by our instruments (hence dark) gave rise to large haloes which governed the assembly of ordinary baryonic matter in the formation of stars and galaxies – the so-called "Cold Dark Matter" (CDM) model [e.g., 228, 40, 78, 315, 244].

Galaxies form through the gravitational collapse and cooling of baryonic material within virialized (i.e., in equilibrium) dark matter halos [325, 270]. Under the gravitational potential, the halo contracts and heats. While compressing, the gas (i.e. the baryonic component) cools via radiative processes and eventually settles in centrifugal equilibrium at the center of the halo potential well forming a rotationally supported gas disk provided that some angular momentum is retained during the collapse [98].

This model has its most convincing support from the Cosmic Microwave Background radiation (CMB). The distribution of hot and cold spots, initially measured by COBE [276], WMAP [24] and, more recently, by Planck [236], can be related to the anisotropies in the distribution of matter when the Universe was only a few hundred thousand years old. Additional support to the CDM model has been brought by the analysis of the large scale structure (LSS) of the Universe using the widest optical surveys available to date: the 2 Degree Field Galaxy Redshift Survey [59], the Sloan Digital Sky Survey [SDSS; 329, 120, 275] and the SDSS-III/Baryon Oscillation Spectroscopic Survey [BOSS; 91, 81]. The wealth of information on the Universe from these surveys allowed the most accurate measurement of the power spectrum of galaxy clustering [299], and revealed the Baryon Acoustic Oscillation (BAO) feature [92] in the clustering of galaxies and quasars.

There is another fundamental, yet still not understood, ingredient in the current concordance cosmological model: the dark energy. Observations of distant ($z \sim 1$) supernovae, used as standard candles, have revealed that the expansion rate of the Universe is increasing with cosmic time [232, 117, 255, 233]. To take into account this effect, the cosmological constant Λ (see Eq. 1.3) was re-introduced in the FRW model, leading to the definition of the ΛCDM framework currently adopted as the standard cosmological model. The values of parameters characterizing the model are known today with a precision of $\approx 5\%$, thanks to the combination of results from a number of different projects, like the measurements of the Hubble constant [105, 168, 237] and CMB anisotropies [280, 279, 168, 238].

1.4. The observational picture

The first galaxy classification, purely based on morphological characteristics, was already proposed by Hubble in 1926 and it is still in use today. Galaxies are divided into two broad classes: ellipticals, which are systems with a rounded shape in the three axes, and spirals, that show a disk-like structure. The analysis of data in the local Universe, like the SDSS and 2dF-GRS surveys, has confirmed and in some cases shown for the first time, that this dichotomy extends to a number of fundamental characteristics of galaxies. The color-magnitude diagram shows two well separated groups of galaxies, a red cloud and a blue sequence, with elliptical galaxies populating the red region, while spiral galaxies reside in the blue part [291, 34]. This characteristic is directly linked to another important difference between the two classes. In fact, bluer spectra are the footprint of an ongoing star formation, while redder ones reflect an older stellar population, which is passively evolving [153, 328]. Moreover, the objects of each class are characterized by different masses: red/elliptical galaxies are massive systems, while blue/spiral galaxies have lower masses, with a quite clear boundary between the two classes falling at $3 \times 10^{10} \mathrm{M}_{\odot}$ [155, 33].

This bimodality in the galaxy distribution is observed also at higher redshift [144, 22, 44]. Several studies using deep surveys have shown that the stellar mass of red galaxies has grown by a factor 2 since $z \simeq 2$, while the mass distribution of blue galaxies has remained almost constant, suggesting a possible transition from the blue sequence to the red cloud with the cosmic time [22, 97]. In this scenario, red galaxies may be the result of early mass assembly and star formation, which would cause the galaxy to initially move along the blue cloud of the color-magnitude diagram, followed by quenching, that turns off star formation and moves the galaxy to the red sequence, and later by dry (i.e., gas-free) merging [53], with the result of displacing the galaxy along the red sequence towards higher masses/luminosities, with the details of these processes still not completely known.

To further complicate the framework, high-redshift galaxies can appear red not only because they are the result of old and passively evolving stars. It has been shown [288, 104], in fact, that the dust in star-forming galaxies can absorb the ultraviolet (UV) light of the young stars and re-emit it to longer wavelengths, typically in the infrared (IR) region. This class of objects, named Distant Red Galaxies, would then escape from the classical dropout selection of Lyman Break Galaxies [LGB; 287, 286], and they revealed to be more massive, older and with more dust than these latter [305, 174], providing evidence for the existence of a number of massive and evolved galaxies when the Universe was still as young as 2 - 3 Gyr.

It is a well known fact that galaxies do not reside in isolated environments, and their locations constitute what is called the large scale structure of the Universe [see e.g., 283]. When considering galaxies in their environment, there exists another important correlation between the intrinsic properties of the galaxy population, the so-called "morphology-density relation". The pioneering works by Oemler [216] and Dressler [87] showed that star-forming galaxies preferentially reside in low-density environments, while inactive elliptical galaxies are found in higher density regions.

The physical origin of this segregation is still unclear; in particular it is still unknown if the morphology-density relation generates at the time of formation of the galaxy or if it is the result of an evolution driven by the density field. There are three main processes identified for the raise of this relation [155]. First, mergers or tidal interactions can destroy galactic disks, thus converting spiral star forming galaxies into bulge-dominated quiescent elliptical galaxies. A second factor is the interaction of galaxies with the dense intra-cluster gas, which can remove the interstellar medium of the galaxy, reducing thus the star formation. Finally, gas cooling processes strongly depend on the environment [323, 31, 270].

The stellar mass function (SMF) and its proxy, the luminosity function (LF), together with the star formation rate (SFR) as a function of mass, are a primer test bench for the current knowledge on galaxy formation. The availability of wide area surveys of the local Universe and of deep surveys have allowed to draw the star formation history up to $z \sim 7$, showing that the SFR is characterized by an increase to $z = 1$, followed by a stationary period extending to $z = 3$ and a subsequent rapid decrease to $z = 7$ [186, 146, 147].

1.5. Galaxy clustering and Baryon Acoustic Oscillations

Just as Type Ia supernovae provide a standard candle[1] [232, 117, 233] for determining cosmic distances, patterns in the distribution of distant galaxies provide a "standard ruler". Imagine dropping a pebble into a pond on a windless day. A circular wave travels outward on the surface. Now imagine the pond suddenly freezing, fixing these small ripples in the surface of the ice. In an analogous fashion, approximately 370,000 years after the Big Bang, electrons and protons combined to form neutral hydrogen, "freezing" in place acoustic pressure waves that had been created when the Universe first began to form structures. These pressure waves are called Baryon Acoustic Oscillations [BAOs; 92] and the distance they have traveled is known as the sound horizon, which is defined as the speed of sound times the age of the Universe when they froze. Such acoustic oscillations in the photon-baryon fluid imprinted their signatures on both the cosmic microwave background, in form of acoustic peaks in the CMB angular power spectrum, and the matter distribution, as BAO peaks in the galaxy power spectrum. Because baryons comprise only a small fraction of matter, and the matter power spectrum has evolved significantly since last scattering of photons, BAOs are much smaller in amplitude than the CMB acoustic peaks, and are washed out on small scales by nonlinear growth of matter clustering. The BAO, or sound horizon, distance is visible as a pronounced peak in the clustering of galaxies around $150\,h^{-1}$Mpc (i.e. 450 million light years), and provides a standard ruler for cosmological distance measurements.

As there is an increased air density in a normal sound wave, there is a slight increase in the chance of finding lumps of matter, and therefore galaxies, separated by the sound horizon distance. By measuring the clustering (i.e. 3D distribution of galaxies; see e.g., [189]) on the sky of galaxies at different distances from us, we are able to precisely determine the angular scale of the sound horizon for galaxies at different redshifts. From this measurement, one can infer the cosmic expansion history defined in Eq. 1.3, $H(z) \equiv d\ln a(z)/dt$, and the growth

[1] http://www-supernova.lbl.gov/

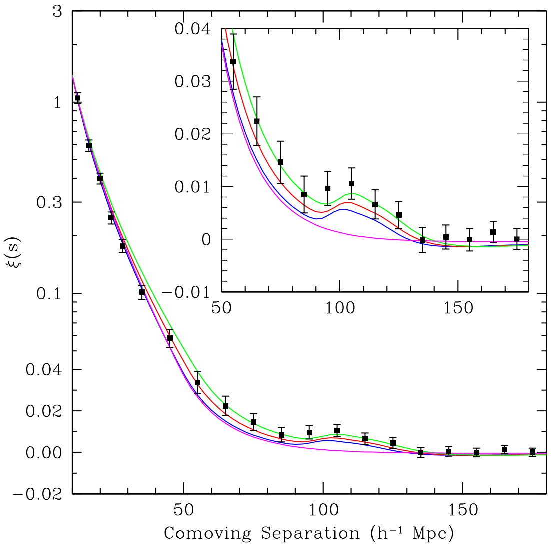

Figure 1.1: Baryon acoustic oscillation peak detected by Eisenstein et al. [92] in the clustering of SDSS Luminous Red Galaxies (LRGs).

rate of structures, $f_g(z) \equiv d\ln D(z)/d\ln a(z)$ [e.g., 130, 68, 248]. In the observed galaxy distribution, the BAO scale appears as a preferred comoving length scale, corresponding to a preferred redshift separation of galaxies in the radial direction, dz, and a preferred angular separation of galaxies in the transverse direction, $d\theta$. Comparing the observed BAO scales with the expected values (via the Hubble law in Eq. 1.10), one can derive [see e.g., 51] $H(z)$ in the radial direction, and the angular diameter distance $D_A(z)$ (defined in Eq. 1.17) in the transverse direction.

Past spectroscopic instruments, as the Sloan Digital Sky Survey [SDSS-I/II; 329, 120, 275] and the SDSS-III/Baryon Oscillation Spectroscopic Survey [BOSS; 91, 81], have mostly targeted Luminous Red Galaxies [LRGs; 90] as BAO tracers, since they are the most clustered galaxies observed in the Universe so far. Ongoing experiments as the SDSS-IV/extended Baryon Oscillation Spectroscopic Survey [eBOSS; 80], and new-generation facilities (see Section 1.7) as the Dark Energy Spectroscopic Instrument [DESI; 264], the 4-metre Multi-Object Spectroscopic Telescope [4MOST; 82], the Subaru Prime Focus Spectrograph [PFS; 294, 274] and EUCLID [176, 262], will all measure the BAO feature in the clustering of emission-line galaxies (Section 1.6) out to redshift $z \sim 2$ and Ly-α forest quasars out to $z \sim 3.5$. These new targets will allow us to better understand how structures formed in the early stages of the Universe and hierarchically evolved into the current LSS configuration, complementing the Type Ia supernova measurements as probes of cosmic expansion.

At the first order, the galaxy clustering measurement is given by the two-point correlation function (2PCF), $\xi(r)$, defined as the excess probability, over an unclustered random Poisson distribution, to find a galaxy within a volume dV at a distance r from an arbitrary chosen galaxy [e.g., 226, 128, 189],

$$dP = \bar{n}[1 + \xi(r)]dV, \tag{1.25}$$

where $\bar{n}$ is the mean number density of the galaxy sample in question. Measurements of $\xi(r)$ are generally performed in comoving space, with r having units of h^{-1}Mpc. The Fourier transform of the 2PCF is the power spectrum $P(k)$, which is used to describe the density fluctuations observed in the CMB.

The galaxy correlation function is well known to approximate a power-law across a wide range of scales,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{1.26}$$

where r_0 is the correlation length, and γ is the power-law slope or spectral index. However, improved models [see review at 69] have been shown to better match the data [332].

Several estimators for $\xi(r)$ have been proposed and tested [226, 79, 128, 159]. Throughout this work I will use the Landy & Szalay [175] one,

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}, \tag{1.27}$$

which has the advantage of minimizing the sample variance. Here DD, DR and RR are the data-data, data-random and random-random weighted and normalized pair counts computed from a data sample of N galaxies and a random catalog of N_R points. In its most general form, the two-point correlation function is given in terms of the parallel, π, and perpendicular, r_p, components of the redshift-space distance $s = \sqrt{r_p^2 + \pi^2}$ with respect to the line of sight (LOS). For further details on the 2PCF estimation see Section 3.3.2.

The clustering measurement is also a fundamental tool to understand the redshift-space distortion (RSD; see Section 3.3.1) effects as a function of the physical scale. On very large scales, galaxies fall toward overdense regions under the influence of gravity. This leads to a distortion in the redshift distribution of galaxies, with the degree of distortion proportional to the growth rate of structure, $f_g(z)$, defined above. This feature is visible in the 2PCF as a compression effect [152, 129] along the line-of-sight direction. The estimate of $f_g(z)$ can be obtained through independent measurements of the linear RSD parameter $\beta = f_g(z)/b$ [226] from the observed galaxy 2PCF, and the bias parameter $b(z)$, which describes the difference between the baryonic and the underlying dark matter distribution, and can be derived from higher order correlation functions of galaxies [308, 111], or the weak lensing shear of galaxies [306, 271]. The redshift-space distortion effects need to be modeled carefully in order to extract information on $f_g(z)$, since the random motions of galaxies on small scales also lead to RSD, which are likely coupled to the nonlinear matter clustering effects on those scales. As a consequence, random peculiar velocity differences arise between close neighbors with respect to the embedding Hubble flow resulting in structures appearing significantly stretched along the line of sight [150]. This effect is commonly referred to as the "finger-of-god"(FoG). The clustering study presented in Chapter 3 includes a straightforward model able to disentangle the different RSD effects as a function of the physical scale.

One can mitigate the impact of small-scale RSD by integrating along the line of sight to approximate the real-space clustering [79] in the projected correlation function,

$$w_p(r_p) = 2\int_0^{\infty} \xi(r_p, \pi) d\pi. \tag{1.28}$$

This integration is usually performed over a finite line-of-sight distance as a discrete sum,

$$w_p(r_p) = 2\sum_i^{\pi_{max}} \xi(r_p, \pi)\Delta\pi_i, \tag{1.29}$$

where π_i is the i^{th} bin of the LOS separation, and $\Delta\pi_i$ is the corresponding bin size.

Beside this statistic, one can measure the multipole moments of the redshift-space 2PCF, which are defined by expanding the 3D clustering estimator as

$$\xi(s,\mu) = \sum_l \xi_l(s) P_l(\mu), \tag{1.30}$$

where μ is the cosine of the angle between the redshift-space distance s and the line-of-sight direction, $\xi_l(s)$ is given by Eq. 1.27, and P_l is the l-th order Legendre polynomial. In this thesis I will focus on the monopole $\xi_0(s)$ – or, simply, $\xi(s)$ – and the quadrupole $\xi_2(s)$.

1.6. Emission-line galaxies as star formation and BAO tracers

Among the bluer, star-forming galaxies, there is a particular class of galaxies whose spectra exhibit strong nebular emission lines originating in the ionized regions surrounding short-lived but luminous massive stars [e.g., 222, 196, 157, 221]. Such Emission-line Galaxies (ELGs) are typically late-type spiral and irregular galaxies, although any galaxy that is actively forming new stars at a sufficiently high rate qualify as an ELG. Because of their vigorous ongoing star formation, the integrated rest-frame colors of ELGs are dominated by massive stars, and hence will typically be bluer than galaxies with evolved stellar populations such as luminous red galaxies [LRGs; 90]. The optical colors of ELGs at a given redshift span a larger range than LRGs due to the much greater diversity of their star formation histories and dust properties.

New-generation large-volume spectroscopic surveys (see Section 1.7), as the ongoing SDSS-IV/eBOSS [80], the near-future DESI[2] [264], 4MOST[3] [82], the Subaru PFS [294, 274] and EUCLID[4] [176, 262], will all target emission-line galaxies out to redshift $z \sim 2$, as star

[2] http://desi.lbl.gov/

[3] https://www.4most.eu/cms/

[4] http://sci.esa.int/euclid/

formation and Baryon Acoustic Oscillation (Section 1.5) tracers. Thus, observing ELGs, modeling their clustering properties and understanding how they populate their host halos are key issues explored in this thesis to prepare the basis for future missions.

By studying the strength and shape of various emission lines, we can classify these galaxies into different types and also get a handle on the composition, temperature and density of the emitting gas, as well as global properties of the galaxy such as the star formation rate, or the mass of the central black hole. Young, hot stars emit much of their energy as ultraviolet (UV) light and therefore detection of the relative brightness of UV can be used to trace their star formation rate. Many surveys have been conducted to study the SFR over time and there is strong evidence for evolution [145, 141, 298].

The source of energy that enables the gas of a galaxy to radiate is ultraviolet radiation from stars. Hot stars, with surface temperature $T_\star = 3 \times 10^4\,\mathrm{K}$, inside or in the vicinity of a gas-rich region, emit UV photons that transfer energy to the gas by photoionization [e.g., 196, 221]. Hydrogen is by far the most abundant element, and photoionization of H is thus the main energy input mechanism. A region of interstellar hydrogen that is ionized is commonly known as "H II region". This is typically a large, low-density cloud of partially ionized gas in which star formation has recently taken place. Photons with energy greater than the ionization potential of H (i.e., 13.6 eV), are absorbed in this process, and the excess energy of each absorbed photon over the ionization potential appears as kinetic energy of a new liberated photoelectron. Collisions between electrons, and between electrons and ions, distribute this energy and maintain a Maxwellian velocity distribution with temperature T in the range 5000< T< 20,000 K [221].

For historical reasons, astronomers tend to refer to the chief emission lines of gaseous nebulae (i.e., [OII] with $\lambda = 3726 - 3729\,\text{Å}$, [OIII] with $\lambda = 5007\,\text{Å}$, [OI] with $\lambda = 6300\,\text{Å}$, etc ...) as "forbidden" lines. They are forbidden since they violate one of the quantum selection rules and they are commonly denoted using brackets. Actually, it is better to think of the bulk of the lines as collisionally excited lines, which arise from levels within a few volts of the ground level and which therefore can be excited by collisions with thermal electrons. Although downward radiation transitions from these excited levels have very small transition probabilities, they are responsible for the emission lines observed. Indeed, at the low density of typical nebulae ($N_e \leq 10^4$ cm^{-3}) collisional de-excitation is even less probable. So, almost every excitation leads to emission of a photon, and the nebula thus emits a forbidden line spectrum that is quite difficult to excite under terrestrial laboratory conditions.

In addition to the collisionally excited lines, the permitted lines of H I (i.e., the 21-cm line of neutral hydrogen), He I, and He II are characteristic features of the spectra of spiral galaxies. They are emitted by atoms undergoing radiative transitions. Indeed, recaptures occur to excited levels, and the excited atoms then decay to lower and lower levels by radiative transitions, eventually ending in the ground level. The spectra of early-type spirals are characterized by an increase of the flux in the blue, to which corresponds the appearance of weak Hα λ6563 (i.e., one of the Balmer absorption lines) and [NII] λ6584 emissions, at the level of a few Å or less in equivalent width. Except for occasional weak [OII] $\lambda 3726 - 3729$ emission, no other nebular lines are detected in the integrated spectrum. Intermediate- to late-type spirals are characterized by much higher blue flux, more prominent Balmer absorption lines and nebular emission features [196, 157, 221].

Besides the galaxy UV-blue continua that help to determine the more local SFR, emission lines are commonly used as a "shortcut" method to estimate the luminosity density of the less local Universe [e.g., 145, 281]. The Hα line at $\lambda = 6563$ Å is the most solid tracer of the presence of ionized hydrogen: in H II regions, in fact, the Balmer emission line luminosities scale directly with the ionizing fluxes of the embedded stars. This line can therefore be used to derive quantitative star formation rates in galaxies [e.g., 158, 148, 206, 113].

UV radiation produced by young, massive stars provoques the photoionization of heavier elements such as neutral oxygen. The [OII] $\lambda 3726 - 3729$ emission-line doublet is the strongest feature after Hα. Its equivalent widths are well correlated with Hα, but [OII] has on average half the flux of Hα. The luminosity of the line has been calibrated [e.g., 157, 148, 113] against

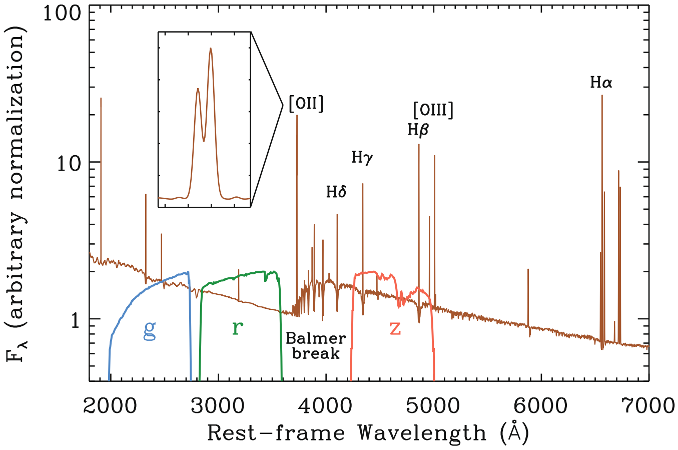

Figure 1.2: Rest-frame spectrum of an ELG showing the blue stellar continuum, the prominent Balmer break, and the numerous strong nebular emission lines. The inset shows a zoomed-in view of the [OII] doublet, which DESI (see Section 1.7) is designed to resolve over the full redshift range of interest, $0.6 < z < 1.6$. The figure also shows the portion of the rest-frame spectrum the DECam grz optical filters would sample for such an object at redshift $z = 1$. Figure from the *DESI Science final design report* at `http://desi.lbl.gov/tdr/`.

Hα and against the SFR determined from the galaxy continuum. The stochastic nature of dust extinction along the multiple sight-lines to the galaxy, and around to individual H II regions, poses problems for the calculation of the internal dust distribution. Thus the [OII] line correlation with the SFR is noisy. The SFRs derived from [OII] are less precise than those from Hα because the mean [OII]/Hα ratios in individual galaxies vary considerably, over $0.5 - 1.0$ dex [108, 157].

Figure 1.2 shows the typical rest-frame spectrum of an ELG that will be planned to be targeted by DESI (see Section 1.7) in the redshift range $0.6 < z < 1.6$, with the blue stellar continuum, the characteristic Balmer break, and the numerous nebular emission lines. The [OII] doublet is highlighted in the zoomed inset. The three closed coloured lines in the lower part of the panel represent the portion of the rest-frame spectrum the Dark Energy Camera[5] (DECam) grz optical filters would sample for such an object at redshift $z = 1$.

[5] `http://legacysurvey.org/`

1.7. Large-volume spectroscopic surveys: past, present and future

In the last decade, a huge effort has been spent in the development of wide-field spectroscopic survey facilities, both ground- and space-based, which led to amazing discoveries and made possible the construction of detailed three-dimensional maps of the Universe to probe its large scale structure.

The 2-degree-Field Galaxy Redshift Survey[6] [2dFGR; 61] (1997-2002) obtained spectra for about 220,000 objects, mainly galaxies, brighter than a nominal extinction-corrected magnitude limit of b$_J$=19.45 by scanning an area of approximately 1500 deg^2. The survey provided accurate measurements of the power spectrum of galaxies, allowing precise determinations of the total mass density of the Universe and the baryon fraction [229]. It measured the distortion of the clustering pattern in redshift space, providing independent constraints on the total mass density and the spatial distribution of dark matter [225, 133]. It also provided evidence for a non-zero cosmological constant, and constraints on the equation of state of the dark energy [89, 231].

Its successor, the Sloan Digital Sky Survey[7] [SDSS; 329, 120, 275], has created the most detailed 3D maps of the Universe ever made so far, with deep multi-color images of one third of the sky, and spectra for more than 3 million astronomical objects. Using the dedicated 2.5-m Sloan telescope [120] at the Apache Point Observatory, New Mexico, it has imaged the sky in five optical photometric bands (u, g, r, i, z) between 3000 and 10,000 Å, with a drift-scanning, mosaic CCD camera [119, 107]. During the first stages of the mission, called SDSS-I (2000-2005) and SDSS-II (2005-2008), it obtained spectra and deep, multi-color images of $\sim 930,000$ galaxies and more than 120,000 quasars. In the second phase (2009-2014), the SDSS-III Baryon Oscillation Spectroscopic Survey [BOSS; 91, 81] targeted 1.5 million galaxies up to $z = 0.7$ [8] and about 160,000 Lyman-α forest quasars in the redshift

[6] http://www.2dfgrs.net/

[7] http://www.sdss.org/

range $2.2 < z < 3$ [273]. BOSS has measured the Baryon Acoustic Oscillation (BAO) feature [92] in the clustering of galaxies and quasars with unprecedented accuracy, probing that the seeds of the large scale structure we see today in the Universe are quantum fluctuations which propagate as sound waves in the very early stages of the Universe. The SDSS high-precision maps of cosmic expansion history using baryon acoustic oscillations have been especially influential in quantifying these results, yielding exquisite constraints on the geometry and energy content of the universe. BAOs were first detected in galaxy clustering by the SDSS-I and in the contemporaneous 2dF Galaxy Redshift Survey, and have since also been detected in intergalactic hydrogen gas using Lyman-α forest techniques. These BAO measurements are complemented by the results of the SDSS-II Supernova Survey[8], which has provided the most precise measurements yet of cosmic expansion rates over the last four billion years. In addition, statistical measurements of galaxy motions and weak gravitational lensing provide some of the strongest evidence to date that Einstein's theory of General Relativity is an accurate description of gravity on cosmological scales.

Its extension, the ongoing SDSS-IV/extended Baryon Oscillation Spectroscopic Survey [eBOSS; 80], plans to target about 350,000 Luminous Red Galaxies (LRGs) in the redshift range $0.6 < z < 0.8$, 260,000 emission-line galaxies in $0.6 < z < 1$ and 740,000 Ly-α forest quasars in $0.9 < z < 3.5$. It will precisely measure the expansion history of the Universe throughout 80% of cosmic history, back to when the Universe was less than 3 billion years old, improving the current constraints on the nature of dark energy.

The near-future Dark Energy Spectroscopic Instrument[9] [DESI; 264] will use the 4-m Mayall telescope located at Kitt Peak, Arizona, to survey about 14,000 $\deg^2$ of the sky to unveil the dark ages of the Universe. It will measure the expansion of the Universe by observing the imprint of baryon acoustic oscillations set down in the first 380,000 years of its existence. This feature has the same source as the pattern seen in the cosmic microwave background,

[8] http://classic.sdss.org/supernova/aboutsupernova.html

[9] http://desi.lbl.gov/

but DESI will map it as a function of cosmic time, while the CMB can see it only at one instant. It is imprinted on all matter at large scales and can be viewed by observing galaxies of various kinds or by observing the distribution of neutral hydrogen (i.e. H II regions, see Section 1.6) across the cosmos, showing up as excess correlations at the characteristic distance of the sound horizon at decoupling. DESI will collect about 10 million spectra of LRGs up to $z = 1$, ELGs up to $z = 1.7$ and Ly-α forest quasars up to $z = 3.5$. From these will come 3D maps of the distribution of matter covering unprecedented volume. This will help to establish whether cosmic acceleration is due to a mysterious component of the Universe, the dark energy, or a cosmic-scale modification of General Relativity, and will constrain models of primordial inflation. This survey will have a dramatic impact on our understanding of dark energy through its primary measurement, that of baryon acoustic oscillations. In addition to the constraints on dark energy, the galaxy and Ly-α flux power spectra will reflect signatures of neutrino mass, scale dependence of the primordial density fluctuations from inflation, and possible indications of modified gravity. To realize the potential of these techniques requires an enormous number of redshifts over a deep, wide volume and DESI was specifically designed with such requirements.

The 4-metre Multi-Object Spectroscopic Telescope[10] [4MOST; 82] located at Cerro Paranal, Chile, will use the 4-m VISTA telescope to simultaneously measure spectra of 1 million Active Galactic Nuclei (AGN) out to $z \sim 5$, and [OII] emission-lines up to $z = 2$. It will be able to simultaneously obtain spectra of ~ 2400 objects distributed over an hexagonal field-of-view of 4 deg^2. This high multiplex of 4MOST, combined with its high spectral resolution, will enable detection of chemical and kinematic substructure in the stellar halo, bulge, thin and thick disks of the Milky Way, helping to unravel the origin of our home galaxy. The instrument will also have enough wavelength coverage to secure velocities of extra-galactic objects over a large range in redshift, thus enabling measurements of the evolution of galaxies and the structure of the cosmos. This instrument enables many science goals, but the design

[10] https://www.4most.eu/cms/

is especially intended to complement three key all-sky, space-based observatories of prime European interest: Gaia[11], EUCLID (see below), and eROSITA[12].

The Prime Focus Spectrograph [PFS; 294, 274] of the Subaru Measurement of Images and Redshifts (SuMIRe) project is a multi-fiber optical/near-infrared spectrograph that will use the Subaru 8.2-m telescope at Mauna Kea, Hawaii, to simultaneously obtain spectra of 2400 cosmological/astrophysical targets in the wavelength range from $0.38-1.3\,\mu$m, in the attempt to study galactic archaeology and galaxy/AGN evolution. Among its targets, it will collect spectra of emission-line galaxies up to $z=2$ [274].

The above ground-based surveys have been complemented by space-based missions in the near-infrared which have provided precise measurements of [OII] and Hα fluxes from emission-line galaxies over a wide range of redshifts. The advantage of observing ELGs from space is that we can get rid of the diffuse thermal emission from the atmosphere. Among these facilities, the WFC3 Infrared Spectroscopic Parallel[13] [WISP; 10] survey has collected Hα spectra [11, 86] using the two infrared grisms (G102 with $\lambda=0.80-1.17\,\mu$m, and G141 $\lambda=1.11-1.67\,\mu$m) of the Wide Field Camera 3 of the Hubble Space Telescope[14] (HST) in pure parallel mode, but for a very tiny area of the sky.

The near-future EUCLID[15] [176, 262] mission has been designed with characteristics very similar to WISP, but much larger field of view. It is a near-IR slitless spectroscopic system with two deep-field instruments, the visual imager (VIS) providing high-quality images to carry out the weak lensing galaxy shear measurement, and the near-IR spectrometer photometer (NISP) to provide photometric redshifts and slitless spectroscopy [176]. EUCLID will scan 15,000 deg^2 of the sky using a 1.2-m telescope. The forecast for the spectroscopic program is 25-50 million galaxies out to $z=2$ in one visible *riz* broad band (550-920nm)

[11] http://sci.esa.int/gaia/

[12] http://www.mpe.mpg.de/eROSITA

[13] http://wisps.ipac.caltech.edu/Home.html

[14] https://www.nasa.gov/mission_pages/hubble/main/

[15] http://sci.esa.int/euclid/

down to magnitude AB=24.5 [176, 177], and their exact number will be limited by the Hα line flux. This corresponds to a look-back time of about 10 billion years, thus covering the period over which dark energy accelerated the expansion of the Universe. This instrument is optimized for two primary cosmological probes: galaxy weak lensing and baryon acoustic oscillations. With its wide-field capability and high-precision design, EUCLID will investigate the properties of dark energy by accurately measuring both the acceleration and the variation of the acceleration at different ages of the Universe. It will test the validity of General Relativity on cosmic scales, explore the nature and properties of dark matter by mapping the 3D dark matter distribution in the Universe, and contribute to refine the initial conditions at the beginning of our Universe, which seed the formation of the cosmic structures we see today. Euclid will also deliver morphologies, masses and star-formation rates with four times better resolution and 3 NIR magnitudes deeper than possible from ground [176]. It is poised to uncover new physics by challenging all sectors of the cosmological model and can thus be thought of as the low-redshift, 3D analogue and complement to the map of the high-z Universe provided by the Planck[16] mission.

1.8. The halo-galaxy connection

The fundamental driver of progress in astronomy is through observations. The advent of large galaxy surveys has led to formidable progress in understanding galaxy formation. Nevertheless, it is difficult to link the galaxies we observe to their host dark matter halos. In fact, the dynamics of galaxy formation involves nonlinear physics and a wide variety of complex physical processes. As such, it is extremely difficult to treat the halo-galaxy connection in full detail using analytic techniques. There are three major approaches that have been developed to circumvent this problem. The first one, which makes use of hydrodynamical N-body simulations, attempts to link galaxies and halos by numerically solving the fully nonlinear equations governing the physical processes inherent to galaxy formation. The second one,

[16] http://www.cosmos.esa.int/web/planck

based on Semi-Analytic Models (SAMs), attempts to construct a coherent set of analytic approximations to describe these same physics. The third approach faces the problem in a more empirical way, by ignoring the complexity of the star formation process and providing a recipe to populate dark matter halos with the observed galaxies. In this context, two methods have been developed in this thesis: the (Sub)Halo Abundance Matching (SHAM) scheme and the Halo Occupation Distribution (HOD) model.

In what follows I give an overview of these techniques, highlighting the strengths and weaknesses of each one of them.

1.8.1 N-body/hydro simulations

The most accurate computational method for solving the physics of galaxy formation is via direct simulation, in which the fundamental equations of gravitation, hydrodynamics, and perhaps radiative cooling and transfer are solved for a large number of points arranged either on a grid or following the trajectories of the fluid flow [e.g., 29, 3, 260].

Collisionless dark matter is relatively simple to model in this way, since it responds only to the gravitational force. For the velocities and gravitational fields occurring during structure and galaxy formation, nonrelativistic Newtonian dynamics is more than adequate. Therefore, solving the evolution of some initial distribution of dark matter (usually a Gaussian random field of density perturbations consistent with the power spectrum of the CMB) reduces to summing large numbers of $1/r^2$ forces between pairs of particles. In practice, clever numerical techniques such as particle-mesh (PM), or tree algorithms are usually used to reduce this N^2 problem into something more manageable [172, 285]. Dark matter only simulations carried out primarily for the cold dark matter scenario, but see also [322, 163, 41, 77, 60, 4], have been highly successful in determining the large scale structure of the Universe, as embodied in the so-called "cosmic web". As a result, the spatial and velocity correlation properties of dark matter and dark matter halos [78, 321, 324, 88, 93, 151, 223, 13, 170, 247], together with the density profiles [208, 48, 207, 195, 242], angular momenta [16, 88, 312, 58, 182, 47, 304, 30, 109] and internal structure [204, 164, 173, 284] of dark matter halos are known to very high accuracy.

Of course, to study galaxy formation dark matter alone is insufficient, and baryonic material must be accounted for. This makes the problem much more difficult since, at the very least, pressure forces must be computed and the internal energy of the baryonic fluid tracked. Particle-based methods – most prominently smoothed particle hydrodynamics [285] – have been successful in this area, as have Eulerian grid methods [252, 106, 239, 246]. For galaxy scale simulations, the real physics of these processes is happening on scales well below the resolution of the simulation, thus the treatment of the physics is often at the "subgrid" level, which essentially means that it is introduces by hand using a semi-analytic approach [300, 156, 187, 320, 289, 73, 263, 218, 43]. Beyond this, problems such as the inclusion of radiative transfer or magnetic fields complicate the situation further by introducing new sets of equations to be solved and the requirement to follow additional fields.

Despite these complexities, progress has been made on these issues using a variety of numerical techniques [2, 52, 116, 234, 12, 85, 102, 178]. Numerous simulation codes are now able to include star formation and feedback from supernovae explosions, as Gadget-I,II,II[17] [282], Gasoline [310], HART [169] and Enzo(Zeus) [219], while some even attempt to follow the formation of supermassive black holes in galactic centers, as Gadget-III and Flash[18] [106]. More recently, the Illustris[19] [309, 209] project has achieved an unprecedented combination of resolution, total volume and physical fidelity providing simulation products with $L_{box} = 106.5\,h^{-1}$Mpc and 1820^3 particles. The future in this field points towards bigger simulations with greater dynamic range. They will provide a more detailed sub-grid physics able to characterize the chemistry of the particles involved (i.e., metals, molecules, dust).

[17] http://www.mpa-garching.mpg.de/gadget/

[18] http://flash.uchicago.edu/website/home

[19] www.illustris-project.org

1.8.2 Semi-Analytic Models

Semi-Analytic Models [SAMs; 18] address the complexity of the galaxy formation process using approximate, analytic techniques to simulate it within cosmologies in which structures grow hierarchically. They consider our best approximation for the physics that underpins galaxy formation, allowing a wide range of properties to be predicted for the galaxy population at any redshift. As with N-body hydrodynamical simulations, the degree of approximation varies considerably with the complexity of the physics being treated, ranging from precision-calibrated estimates of dark matter merger rates to empirically motivated scaling functions with large parameter uncertainty (e.g., in the case of star formation and feedback). The advantage of the semi-analytic approach is that it is computationally inexpensive compared to N-body/hydro simulations. This facilitates the construction of samples of galaxies orders of magnitude larger than possible with N-body techniques and for the rapid exploration of parameter space [139] and model space (i.e. accounting for new physics and assessing the effects). The primary disadvantage is that they involve a larger degree of approximation. The extent to which this actually matters has not yet been well assessed. Numerous studies [154, 19, 278, 57, 26, 132, 201] have extended and improved the original framework [323, 56] aiming to investigate many aspects of galaxy formation including: merger trees, halo profiles, gas infall and cooling, stellar synthesis, SN and AGN feedback mechanisms, reionization, environment, chemical evolution.

Later comparison studies of semi-analytic versus N-body/hydro calculations have shown overall quite good agreement, at least on mass scales well above the resolution limit of the simulation, but have been limited to either simplified physics as hydrodynamics and cooling only [27, 137] or to simulations of individual galaxies [293].

More recent semi-analytic models developed to evolve galaxies through a merging hierarchy of dark matter halos are e.g., GALFORM[20], Galacticus [25] and SAGE[21] [75]. These algo-

[20] www.galform.dur.ac.uk

[21] http://www.asvo.org.au/about/glossary/

rithms take the output from cosmological dark matter-only N-body simulations and build an analytic representation of the stars, gas and galaxies that are expected to live within. The resulting model galaxies can then be compared with the observed population of galaxies and used to interpret the data and test our understanding of the physics of galaxy formation.

At the Instituto de Física Teórica (IFT)/UAM we are now starting to develop our own set of semi-analytic models, called Multidark Galaxies[22], based on the available MultiDark[23] simulation products. This is a joint collaboration between the IFT/UAM, Durham, Caltech/AIP, La Plata and Swinburne Universities, whose goal is to construct and provide to the scientific community reliable and physically motivated SAMs.

1.8.3 Statistical methods

In the halo model [69], galaxies are treated as biased tracers of the underlying dark matter distribution since their clustering properties are strongly correlated. On large scales where the linear regime holds, we are able to reconstruct the dark matter clustering signal, $\xi_m(s)$, from the observed one, $\xi(s)$, by using [215]

$$\xi(s) = b^2(s)\xi_m(s), \tag{1.31}$$

where the large-scale bias $b(s)$ depends on the physical scale. Turning this concept around, knowing the physics of the simulated halos – which is straightforward because they are made of dark-matter collisionless particles interacting only gravitationally – we can reconstruct the clustering properties of the observed galaxies without dealing with the complexity of galaxy formation. This kind of modeling is purely statistical in the sense that it links galaxies to halos using only their spatial and clustering properties. There are two main schemes used to statistically populate halos with observed galaxies to create mock catalogs:

[22] www.multidarkgalaxies.pbworks.com

[23] https://www.cosmosim.org/

1. *Halo Occupation Distribution*

   The Halo Occupation Distribution [HOD; 28, 170, 336, 337] model is based on the conditional probability, $P(N|M)$, that a halo with mass M contains N galaxies of a given type. In its five-parameter formulation [337], the mean number of galaxies per halo mass is given by the sum of a central plus a satellite contribution. The central term is defined by

$$< N_{cen}(M) >= \frac{1}{2}\left[1 + erf\left(\frac{\log M - \log M_{min}}{\sigma_{\log M}}\right)\right], \tag{1.32}$$

   where the error function is defined as the integral

$$erf(x) = 2\int_0^x e^{-t^2} dt/\sqrt{\pi}. \tag{1.33}$$

   The free parameters in the central term are M_{min}, the minimum mass scale of halos that can host a central galaxy, and $\sigma_{\log M}$, the width of the cutoff profile. At a halo mass of M_{min}, 50% of halos host a central galaxy, which in terms of probability means that $P(1) = 1 - P(0)$. If the relation between galaxy luminosity and halo mass had no scatter, $< N_{cen}(M) >$ would be modeled by a hard step function. In reality, this relation must possess some scatter, resulting in a gradual transition from $N_{cen} \simeq 0$ to $N_{cen} \simeq 1$. The width of this transition is $\sigma_{\log M}$. To place the satellite galaxies, one has to assume their number in halos of a given mass follows a Poisson distribution, which is consistent with theoretical predictions [28, 170, 336]. We approximate the mean number of satellite galaxies per halo with a power law truncated at a threshold mass of M_0

$$< N_{sat} >=< N_{cen}(M) > \left(\frac{M - M_0}{M_1}\right)^{\alpha}. \tag{1.34}$$

2. *(Sub)Halo Abundance Matching*

   The (Sub)Halo Abundance Matching [SHAM; 65, 302, 165, 215] model relies on the single assumption that more luminous galaxies live in more massive halos. The assignment is performed using two proxies – usually the maximum circular velocity, V_{max}, for the halos and the luminosity, or stellar mass $M_\star$, for the galaxies. The halos are sorted according to their velocity and the fastest ones are assigned more luminous galaxies through their number densities

$$n_h(> V_{max}) = n_g(< M_r), \tag{1.35}$$

where M_r is the r-band absolute magnitude. In its basic formulation, SHAM is nothing but a one-to-one correspondence between halos and galaxy number density. In reality, to match the observations, one has to "relax" the monotonic assignment by allowing some scatter in the $V_{max} - M_r$ relation.

The advantage of using either HOD or SHAM models instead of SAMs or N-body/hydro simulations is that they are straightforward methods to connect halos to galaxies able to reproduce remarkably well [302, 334, 215, 122, 100, 99, 257] the clustering of galaxies in the Universe. The disadvantage, however, is that these models are only applicable to complete galaxy samples, i.e. samples in which all the objects have been observed. In case the sample considered is not complete, as often happens in astronomy, these prescriptions need to be modified to take into account the sample incompleteness. In Chapters 2 and 4, I will present two clustering studies on different emission-line galaxy samples, both suffering of incompleteness, for which I modify the standard SHAM procedure to correctly reproduce the observations.

In what follows, I present three clustering studies in different galaxy samples of the SDSS and the SDSS-III/BOSS surveys. I measure the 2PCF of galaxy populations that differ in redshift, color, luminosity, star-formation history, bias, and interpret the results through high-resolution cosmological simulations to better understand the galaxy halo occupation distribution and its evolution with redshift. The investigation proposed spans a redshift range going from the local Universe, with the SDSS Main and [OII] emission-line galaxy samples at $z \sim 0.1$ (Chapter 2), to $0.43 < z < 0.7$, with the BOSS CMASS DR11 galaxies (Chapter 3), up to $z \sim 0.8$, with the BOSS DR12 [OII] ELG sample (Chapter 4). The aim of this research is to stress the importance of star-forming galaxies and, among these, emission-line galaxies as tools for cosmology with new-generation wide-field spectroscopic surveys. Near-future instruments as DESI, 4MOST, Subaru PFS and EUCLID (see §1.7) will all target emission-line galaxies out to redshift $z \sim 2$ to trace star formation and to measure the baryon acoustic oscillation feature. This latter provides a standard ruler for cosmological distances, which can be used to probe the accelerated expansion of the Universe. Therefore, understanding how to measure and properly model the ELG clustering properties and how they populate their host halos are fundamental issues I address in this thesis using state-of-the-art data, currently available, to prepare the clustering prospects and theoretical basis for future experiments.

Somos un poquito más que polvo de estrellas.

El Niño de las pinturas

2

Clustering dependence on the r-band and [OII] emission-line luminosities in the local Universe

2.1. Abstract

We study galaxy clustering as a function of the [OII] emission-line luminosity in the local Universe, at redshift $z \sim 0.1$, using the SDSS DR7 Main galaxy sample extracted from the New York University -Value Added Galaxy Catalog (NYU-VAGC). We characterize the dependence of the clustering signal on the r-band absolute magnitude, M_r, and the [OII] emission-line luminosity by matching our Main galaxy selection to the available MPA-JHU DR7 release of spectrum measurements. We select several volume-limited samples, both in M_r and [OII] luminosity thresholds, and there we measure the projected, monopole and quadrupole two-point correlation functions. To model our results, we map them onto the MultiDark Planck $L_{box} = 1\,h^{-1}$Gpc cosmological simulation using a (Sub)Halo Abundance Matching approach.

We apply the SUrvey GenerAtoR (SUGAR) algorithm to build reliable light-cones including the complete redshift evolution over the range of interest, $0.02 < z < 0.22$, and accounting for those volume effects, as cosmic variance or number density fluctuations, which are observed in the data. This analysis reveals a clear dependence of galaxy clustering on both the r-band and the [OII] luminosity, generally being stronger for more luminous samples. The MultiDark mock galaxies show remarkable agreement with the data, and allow us to constrain the typical host halo masses and satellite fractions for SDSS galaxies as a function of both the $r-$band and the [OII] luminosities.

2.2. Introduction

The current standard cosmological model claims that galaxies form and evolve within the potential well of dark matter halos, which are complex structures that do not absorb, reflect or emit light and interact with ordinary, baryonic matter only gravitationally [e.g., 315]. In order to correctly predict the distribution of galaxies within their host halos and the halo distribution in the large-scale structure of our Universe, we need to build a reliable halo model accounting for all the ingredients that regulate the galaxy formation process. It is well known [e.g., 67, 188, 121, 259, 100] that galaxies can be classified by color into younger, bluer star-forming galaxies and older, redder, more clustered ones. Among the star-forming population, there is a particular class of galaxies whose rest-frame optical spectra exhibit emission lines from which detailed physical properties can be inferred. For galaxies at the peak of cosmic star formation at $z \sim 2$, these emission lines are shifted into the near-infrared which, combined with the intrinsic faintness of the source, makes them difficult to observe using ground-based facilities [192]. For this reason, relatively few near-infrared spectra of galaxies at $z \sim 2$ that cover all the important rest-frame optical emission lines have been published to date [e.g., 95, 94, 127, 256, 23, 86].

The available near-infrared spectra of star-forming galaxies at $z \sim 2$ have revealed differences in comparison with their counterparts in the local Universe [185, 212]. For example, star-forming galaxies at $z \sim 2$ tend to have higher [OIII]/Hβ ratios at a given [NII]/Hα ratio compared to local star-forming galaxies. This evidence has been attributed to more extreme interstellar medium conditions, on average, in galaxies at high redshift, possibly as a result of harder ionizing radiation field, different gas volume filling factors, higher nebular electron densities, AGN activity [268, 46, 269, 160]. The clumpy morphology and high velocity dispersions observed in many of these sources [235, 112, 179] may support the conjecture that star formation in the early Universe generally occurs in denser and higher pressure environments than those found in local star-forming galaxies. The slitless grim spectroscopy provided by the Wide Field Camera 3 (WFC3) on the Hubble Space Telescope[1] (HST) has lead to the discovery of large numbers of star-forming galaxies near the peak of cosmic star formation [10]. Grism surveys as the WFC3 Infrared Spectroscopic Parallel [WISP; 10] survey, are ideal to detect low-mass star-forming galaxies at intermediate redshift through their optical emission lines, but they lack of spectral resolution to resolve Hα from [NII] $\lambda = 6548,\, 6583\,\text{Å}$ or detect line broadening due to AGN activity. For these reasons, ground-based spectroscopy with the new generation of infrared spectrometers is required to complement these space-based facilities and help to constrain the physical properties of these galaxies.

Large-volume spectroscopic surveys – both ground-based, as the ongoing SDSS-IV extended Baryon Oscillation Spectroscopic Survey [eBOSS; 80], the near-future Dark Energy Spectroscopic Instrument [DESI; 264], the 4-metre Multi-Object Spectroscopic Telescope[2] [4MOST; 82], the Subaru Prime Focus Spectrograph [PFS; 294, 274], and space-based as the slitless, near-IR EUCLID[3] [176, 262] survey – will all target Emission-Line Galaxies (ELGs) up to $z \sim 2$, allowing us to study the evolution of their clustering properties out to very high redshifts. Measuring and modeling the ELG clustering and understanding how this particular class of star-forming galaxies populate their host halos are therefore fundamental

[1] https://www.nasa.gov/mission_pages/hubble/main/index.html

[2] https://www.4most.eu/cms/

[3] http://sci.esa.int/euclid/

issues we have to address now to set the basis for future experiments.

The goal of this work is to characterize the clustering of the well known SDSS DR7 Main galaxy sample [292, 1], both in terms of the absolute r-band magnitude and the [OII] emission-line luminosities. We derive this latter galaxy property by matching our fiducial NYU-VAGC [39] Main sample to the SDSS DR7 MPA-JHU[4] emission-line galaxy catalog. We build suitable volume-limited samples in M_r and [OII] luminosity thresholds, and there we measure the projected, monopole and quadrupole two-point correlation functions (2PCF). We then interpret our measurements building suitable light-cones by applying the SUrvey GenerAtoR algorithm [SUGAR; 257] to the high-resolution MultiDark[5] $L_{box} = 1\,h^{-1}$Gpc [167] cosmological simulation with Planck cosmology [236]. In order to populate the MultiDark halos with mock galaxies, we adopt a (Sub)Halo Abundance Matching [SHAM; 166, 302] approach.

Our model galaxies reproduce remarkably well the clustering properties of the SDSS Main galaxy sample, both in terms of the r-band absolute magnitude and the [OII] emission-line luminosity. With our analysis we are able to constrain the typical host halo masses and satellite fractions of these galaxies as a function of their M_r and [OII] luminosities. Consistently with previous results [311, 334, 122], we find that galaxies with stronger r-band luminosities show a higher clustering amplitude. The same behavior is observed in the clustering as a function of the [OII] emission-line luminosity. For both classes of measurements, we find that more luminous galaxies live in more massive halos with a lower satellite fraction, compared to their fainter counterparts. The advantage and novelty of our method is that, building a light-cone, we are able to model the evolution over the redshift range considered, $0.02 < z < 0.22$, accounting for those volume effects, as cosmic variance or galaxy number density fluctuations, that are naturally observed in the data, and a single simulation snapshot cannot capture. The cost is the volume limitation: in fact, from a simulation with

[4] https://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

[5] https://www.cosmosim.org/cms/

$V = 1\,h^{-3}\text{Gpc}^3$, the maximum light-cone aperture we can generate is about $0.02\,h^{-3}\text{Gpc}^3$, much smaller than one single MultiDark realization. This makes the model less accurate on large scales. For this reason, in the current work we focus on scales $s \lesssim 30\,h^{-1}\text{Mpc}$. The robustness of our method is demonstrated by the fact that we are able to accurately fit all our clustering statistics using a straightforward SHAM model with the satellite fraction as free parameter. From our light-cones we derive reliable clustering models that correctly fit the SDSS measurements both on small and larger scales, without introducing any velocity bias [125, 122] or additional modifications in the standard SHAM procedure. Our predictions for the typical SDSS Main satellite fraction values are overall higher than what found by Guo et al. [122] using a HOD approach, and the discrepancy is due to the different way of populating halos with galaxies in our models.

Throughout this work we adopt a flat ΛCDM cosmology with $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $n_s = 0.96$ and $\sigma_8 = 0.82$.

This chapter is organized as follows: in Section 2.3.1 we describe the SDSS Main sample selection criteria. In § 2.3.2 we explain the steps we follow to match the Main galaxy sample to the SDSS MPA-JHU emission-line galaxy catalog and how we derive the [OII] luminosity. In Section 2.3.3 we discuss the steps to define the Balmer ratio which is commonly used as a dust extinction indicator. In §2.3.4 we describe how to estimate the star formation rate using [OII] and Hα emission lines. In § 2.4 we define the clustering estimators and the tools needed to perform our measurements. Section 2.5 describes the MultiDark Planck cosmological simulation and the tools used for the analysis. We present our main results in § 3.7 and the conclusions in § 2.7.

2.3. Data

2.3.1 The SDSS DR7 Main galaxy sample

We study galaxy clustering in the local Universe as a function of r-band luminosity using the New York University Value-Added Galaxy Catalog[6] [NYU-VAGC; 39], which is based on the SDSS DR7 Main galaxy sample [1]. This sample covers an effective area of about 7300 deg^2 and contains about 520,000 galaxies satisfying the following spectroscopic target selection [292]:

$$\begin{aligned} r_{PSF} - r_{mod} &> 0.3, \\ r_p &< 17.77, \\ \mu_{50} &< 24.5, \end{aligned} \tag{2.1}$$

where r_{PSF}, r_{mod} and r_p are respectively the PSF, model and petrosian r-band apparent magnitudes and μ_{50} is the mean surface brightness within the petrosian half-light radius θ_{50},

$$\mu_{50} = r_p + 2.5 \log(2\pi\theta_{50}^2). \tag{2.2}$$

Following [334], we impose a more conservative faint magnitude limit, $r_p < 17.6$, but no bright limit. We compute the r-band absolute magnitudes of these galaxies as [36]

$$M_{^{0.1}r} - 5\log h = r_p - DM(z, \Omega_m, \Omega_\Lambda, h = 1) - K_{^{0.1}rr}(z), \tag{2.3}$$

where $K_{^{0.1}rr}(z)$ is the $K-$correction [143] from the r-band of a galaxy at redshift z to the $^{0.1}r$ band, computed using kcorrectv4.3 [38]. These magnitudes are calculated assuming $h = 1$ and are corrected including passive evolution to the median redshift of the sample, $z = 0.1$, to account that galaxy luminosities are brighter in the past [35].

In general, a magnitude-limited survey will be affected by radial-selection effects resulting

[6] http://cosmo.nyu.edu/blanton/vagc/

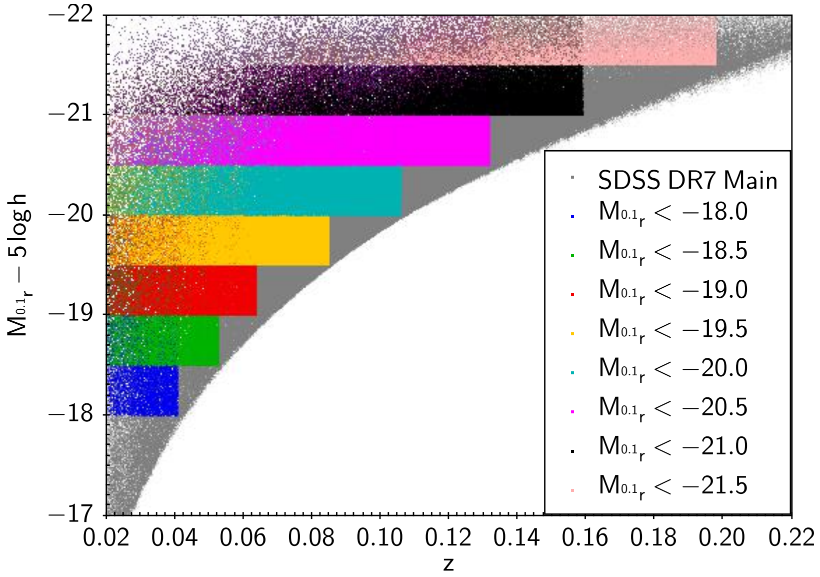

Figure 2.1: SDSS Main DR7 volume-limited samples built imposing the redshift and r-band petrosian absolute magnitude limits reported in Table2.1. All the magnitudes are computed with $h = 1$, $K-$corrected and passively evolving to $z = 0.1$. The fiber collision correction is also included.

from its inability to detect fainter galaxies at high redshifts. One way to avoid these effects is defining a volume-limited sample, in which a maximum redshift and minimum absolute magnitude are chosen, so that every galaxy in this redshift and magnitude range will be observed. We therefore build suitable volume-limited samples to measure and model the clustering dependence on galaxy luminosity. The magnitude and redshift cuts for each one of them are listed in Table 2.1 and shown in Figure 2.1. We impose a minimum redshift of $z_{min} = 0.02$ to each sample.

2.3.2 Emission-line luminosities

We assign [OII] emission-line fluxes to the SDSS DR7 Main galaxies by spectroscopically matching the NYU-VAGC catalog to the MPA-JHU[7] DR7 release of spectrum measurements. To this purpose, we consider only MPA-JHU galaxies with good spectra, i.e. those galaxies with `ZWARNING=0`. For those galaxies surviving the matching, we merge [OII] emission-line fluxes. Hereafter, the resulting galaxy catalog will be called “MPA-NYU SDSS Main”

[7] http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

$M_{0.1_r}^{min} - 5\log h$	z_{max}	N_{gal}	$\bar{n}_g$	Vol
			$[10^{-3}h^3\mathrm{Mpc}^{-3}]$	$[10^6 h^{-3}\mathrm{Mpc}^3]$
-18.0	0.041	35023	29.68	1.18
-18.5	0.053	56960	21.02	2.71
-19.0	0.064	71887	14.82	4.85
-19.5	0.085	125436	11.01	11.39
-20.0	0.106	131986	6.03	21.90
-20.5	0.132	122678	2.95	41.62
-21.0	0.159	77860	1.09	71.41
-21.5	0.198	36003	0.27	134.02

Table 2.1: Redshift and r-band absolute magnitude cuts of our SDSS Main DR7 volume-limited samples. For each sample we report the number of galaxies (N_{gal}) contained, its mean number density ($\bar{n}_g$), and its comoving volume (Vol). We impose a minimum redshift of $z = 0.02$ to each one of the samples.

catalog. Notice that all the galaxies in this samples show [OII] emission lines, and we are not including any possible elliptical, LRG, or any other type of galaxy that is central for some of the ELGs considered.

To study galaxy clustering as a function of the emission-line luminosity, we adopt the same procedure explained in Section 2.3.1 and define volume-limited samples for different [OII] luminosity thresholds. We recover the emission-line luminosities from the MPA-JHU [OII] fluxes following the procedure described below.

To this purpose, we remind that ELGs emit an intrinsic luminosity, L_{intr}, which is partially absorbed by the dust around the emitting galaxy, resulting in an observed luminosity, L_{obs}. This latter propagates to us with an observed flux, F_{obs}, which is then partially absorbed by dust around the Milky Way (this phenomenon is better known as "extinction"), and finally detected by our telescope as F_{obs}^{ext}. From the MPA-JHU DR7 measurements of F_{obs}^{ext}, we want to reconstruct the observed luminosity L_{obs} of our SDSS Main emission-line galaxies, which is linked to the observed flux through the expression [e.g., 148, 206]

$$L_{obs} = 4\pi D_L^2 10^{-0.4(m_p - m_{fib})} F_{obs}, \qquad (2.4)$$

where D_L is the luminosity distance depending on redshift and cosmology usually given in units of cm. In the equation above, the exponent is the aperture correction taking into account that only the portion of the flux “through the fiber” will be detected by the SDSS spectrograph – fibers in SDSS have an aperture of 3” [292]. The aperture correction implicitly assumes that the emission measured through the fiber is characteristic of the whole galaxy and that the star formation is uniformly distributed over the galaxy. The term m_p in Eq. 2.4 is the petrosian magnitude in the desired band-pass filter representing the total galaxy flux and m_{fib} is the fiber magnitude derived from a photometric measurement of the magnitude in an aperture the size of the fiber and corrected for seeing effects. In the SDSS *ugriz* [119, 107] optical photometric system, the [OII] doublet with wavelengths $\lambda = 3726, 3729\,\text{Å}$ lies in the u-band, Hα with $\lambda = 6563\,\text{Å}$ in the r-band Å and Hβ with $\lambda = 4861\,\text{Å}$ in the g-band.

To derive the observed luminosity through Eq. 2.4, we first need to reconstruct the observed flux F_{obs} from the flux measurements F_{obs}^{ext} available in the MPA-JHU catalog, and this is done by correcting them for extinction using the $E(B-V)$ dust maps by Schlegel et al. [265] and the extinction law by Calzetti et al. [49]

$$k(\lambda) = \begin{cases} 2.659(-2.156 + 1.509/\lambda - 0.198/\lambda^2 + 0.011/\lambda^3) + 4.05, & \text{if } \lambda < \; 6300\,\text{Å} \\ 2.659(-1.857 + 1.040/\lambda) + 4.05, & \text{if } \lambda \geq \; 6300\,\text{Å}. \end{cases} \tag{2.5}$$

The extinction-corrected flux is estimated as [148, 206]

$$F_{obs} = F_{obs}^{ext} \times 10^{-0.4A_\lambda} = F_{obs}^{ext} \times 10^{-0.4E(B-V)k(\lambda_{obs})}, \tag{2.6}$$

where F_{obs}^{ext} is usually given in units of [erg cm^{-2} s^{-1}] and can be computed from the line flux continuum (F_c) and equivalent width (EQW) as [206]

$$F_{obs}^{ext} = F_c \times |EQW|. \tag{2.7}$$

The EQW provided in the MPA-JHU catalog already account for stellar absorption. In case

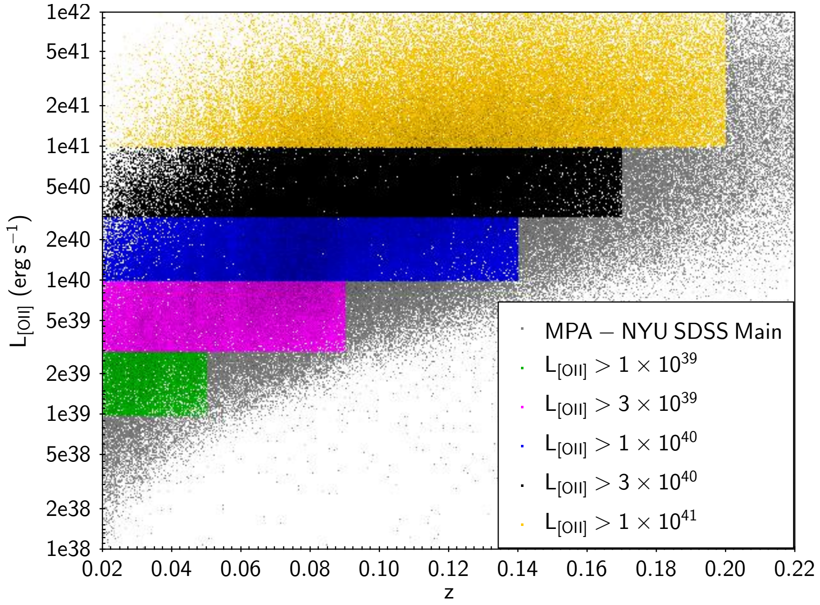

Figure 2.2: [OII] emission-line luminosity (grey dots) and volume-limited samples (coloured squares) for the NYU-MPA Main galaxies. The specific cuts used to define the samples are reported in Table 2.2. We impose a conservative minimum flux limit of 10^{-16} erg cm^{-2} s^{-1} to exclude objects with too short exposure time.

the emission-line galaxy catalog did not include stellar absorption correction, one should apply the following correction to the flux [148, 206]

$$F_{obs}^{ext,\star\, corrected} = F_{obs}^{ext}\left(\frac{EW + EW_c}{EW}\right), \tag{2.8}$$

where the factor EW$_c$ varies from $\sim 1\,\text{Å}$ for Sa galaxies, to $\sim 2\,\text{Å}$ for Sb galaxies, to $\sim 4\,\text{Å}$ for extreme late types [206, 198]. The quantity $k(\lambda_{obs})$ in Eq. 2.6 is the reddening curve defined in Eq. 2.5, $\lambda_{obs} = \lambda_{em}(1 + z)$ is the observed wavelength and λ_{em} is the emitted one. In the case of the of line doublets emitting two different wavelengths as [OII], the flux considered is the cumulative flux of both lines.

Our results for the observed [OII] emission-line luminosity of the NYU-MPA SDSS Main galaxy sample in the redshift range $0.02 < z < 0.22$ are shown in Figure 2.2. SDSS DR7 spectra[8] are combined from three or more individual exposures of 15 minutes each, corresponding to typical [OII] fluxes of $\sim 10^{-16}$ erg cm^{-2} s^{-1} [64]. We then reject those objects [OII] flux

[8] http://classic.sdss.org/dr7/products/spectra/

z_{max}	$\mathrm{L}_{[O_{II}]}^{min}$	N_{gal}	$\bar{n}_g$	Vol
	[erg s^{-1}]		$[10^{-3}h^3\mathrm{Mpc}^{-3}]$	$[10^6 h^{-3}\mathrm{Mpc}^3]$
0.05	1×10^{39}	57599	25.57	2.25
0.09	3×10^{39}	174366	12.92	13.50
0.14	1×10^{40}	244705	4.95	49.39
0.17	3×10^{40}	184626	2.13	85.59
0.20	1×10^{41}	89816	0.65	137.91

Table 2.2: Redshift and [OII] luminosity cuts that define the MPA-NYU SDSS Main volume-limited samples. For each sample we report the number of galaxies (N_{gal}) contained, its number density (n_g), and its comoving volume (Vol). We impose a minimum redshift of $z = 0.02$ and a minimum [OII] line flux of $\sim 10^{-16}$ erg cm^{-2} s^{-1} to each one of the samples.

lower than this threshold and remain with a sample of about 433,000 [OII] emission-line galaxies. The colored selections in the figure are the volume-limited samples in [OII] luminosity luminosity thresholds. The specific cuts applied to obtain them are reported in Table 2.2.

2.3.3 Balmer decrement as dust extinction indicator

The problem of quantifying extinction, i.e. the absorption and scattering of electromagnetic radiation by dust and gas between an emitting astronomical object and the observer, is directly related to the interpretation of the Balmer decrement, that is defined as the ratio of the Balmer-line intensities (i.e. luminosities) $\mathrm{L}_{H\alpha}/\mathrm{L}_{H\beta}$ [e.g., 197, 222, 220, 221]. The intensity ratios of Balmer lines in all planetary nebulae with typical gas conditions should be roughly the same, but this is not what is observed [222, 220, 221]. Interstellar extinction (or reddening) produced by dust particles selectively dims shorter bluer wavelenghts, leading to Balmer line ratios that differ systematically from the theoretical predictions. The more dust, the larger the disparity between the observed and the theoretical Balmer decrements. Turning this concept around, from the size of the discrepancy between observed and theoretical Balmer decrements, one can infer the amount of interstellar reddening and, therefore, dust between the observer and a given planetary nebula. The difference between observed and intrinsic (i.e. theoretical) nebular color in the absence of dust can be expressed as a

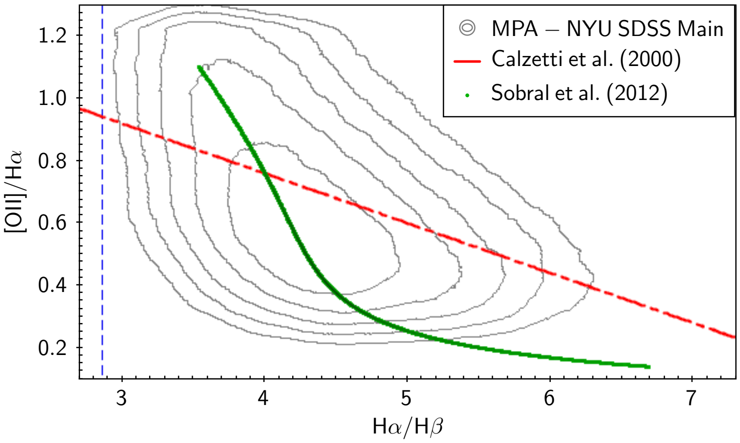

Figure 2.3: The variation of the [OII]/Hα line ratios as a function of the Balmer decrement shows they are correlated, therefore the observed [OII]/Hα ratio can be used as a robust dust extinction indicator. The vertical line is the theoretical Hα/Hβ prediction, the transverse red dot-dashed line is the prediction from the extinction law by Calzetti et al. [49], and the green solid line is the best polynomial fit by Sobral et al. [277].

$(B-V)$ color excess as [86]

$$E(B-V) = E(B-V)_{obs} - E(B-V)_{intr}. \tag{2.9}$$

This excess is then related to the Balmer decrement through the equation [86, 138]

$$\begin{aligned} E(B-V) &= \frac{E(F_{H\beta} - F_{H\alpha})}{k(\lambda_{\mathrm{H}\beta}) - k(\lambda_{\mathrm{H}\alpha})} \\ &= \frac{2.5}{k(\lambda_{\mathrm{H}\beta}) - k(\lambda_{\mathrm{H}\alpha})} \log_{10}\left[\frac{(F_{H\alpha}/F_{H\beta})_{obs}}{(F_{H\alpha}/F_{H\beta})_{int}}\right], \end{aligned} \tag{2.10}$$

where $k(\lambda_{\mathrm{H}\beta})$ and $k(\lambda_{\mathrm{H}\alpha})$ are the reddening curves defined in Eq. 2.5 evaluated at the Hβ and the Hα wavelengths. The factor $E(F_{H\beta} - F_{H\alpha})$ is analogous to the color excess, but this is defined for Hα and Hβ instead of the B- and V-bands. The quantity $(F_{H\alpha}/F_{H\beta})_{obs}$ is the observed Balmer decrement and $(F_{H\alpha}/F_{H\beta})_{int}$ is the intrinsic or unreddened one, which is computed theoretically. Under typical conditions in planetary nebulae, the intrinsic ratio remains roughly constant, $(F_{H\alpha}/F_{H\beta})_{int} = 2.86$. This value is standard for star-forming galaxies in literature and corresponds to a temperature $T = 10^4\,\mathrm{K}$ and an electron density $n_e = 10^2\,\mathrm{cm}^{-3}$ in Case B recombination [221].

Beside the [OII] emission-lines, from the MPA-NYU matching we also merge Hα and Hβ line fluxes to estimate the Balmer decrement of the SDSS Main galaxies. The variation of the [OII]/Hα ratio as a function of the Balmer decrement for the full SDSS Main sample is shown in Figure 2.3 and reveals, as expected, the presence of extinction well beyond the theoretical value (represented by the blue dashed vertical line). The transverse red dot-dashed line is the prediction from the theoretical extinction law by Calzetti et al. [49], while the green solid curve is the best polynomial fit by Sobral et al. [277]. This result shows that the observed [OII]/Hα ratio correlates well with the Hα/Hβ ratio, thus it can be calibrated as a dust extinction tracer.

2.3.4 Star formation rates

The star formation rate (SFR) of a galaxy is typically estimated by applying a scaling factor to a galaxy luminosity measurement which is star formation-sensitive [e.g., 158, 148, 206, 113]. Emission-line luminosities are therefore perfect candidates as star formation tracers. Galactic foreground obscuration (i.e. extinction) is corrected for using the dust maps by Schlegel et al. [265], but obscuration by dust intrinsic to the star-forming galaxies can cause more significant underestimates in the SFRs derived from emission lines. To include this correction, we calibrate the [OII] and the Hα luminosities defined in Eq. 2.4 using the Balmer decrement as an obscuration curve [148, 206, 113]. The resulting star formation rate for the Hα emitters is then

$$SFR_{H\alpha}(\mathrm{M_\odot\,yr^{-1}}) = \frac{10^{0.4A_{H\alpha}}}{1.5}\frac{L_{H\alpha}}{1.27\times10^{41}\mathrm{erg\,s^{-1}}}, \tag{2.11}$$

where the conversion factor corresponds to the Salpeter initial mass function (IMF) calibration of Kennicutt [158] multiplied by 1.5 to convert to the] IMF. The $A_{H\alpha}$ coefficient is the obscuration correction which, assuming an intrinsic Balmer decrement of 2.86 [221] and recalling Eqs. 2.5 and 2.10, can be written as [113]

$$A_{H\alpha} = k(H\alpha)\, E(B-V) = \frac{2.5}{k(H\beta)/k(H\alpha) - 1} \log_{10}\left(\frac{F_{H\alpha}/F_{H\beta}}{2.86}\right), \tag{2.12}$$

where $F_{H\alpha}$ and $F_{H\beta}$ are the Hα and Hβ line fluxes given in Eq. 2.6. The ratio $k(H\beta)/k(H\alpha)$ depends on the extinction law assumed and, in the case of Calzetti et al. [49], it is $k(H\beta)/k(H\alpha) = 1.895$. Analogously, for the SFR [OII] estimator we have [113]

$$SFR_{[OII]}(\mathrm{M_\odot\,yr^{-1}}) = \frac{10^{0.4A_{H\alpha}}}{1.5\, r_{lines}} \frac{L_{[OII]}}{1.27 \times 10^{41} \mathrm{erg\,s^{-1}}}, \tag{2.13}$$

where r_{lines} is the ratio of the extinguished [OII] to Hα flux. In the absence of better information, a ratio of ~ 0.5 is typically assumed [e.g., 158]. In the top panel of Figure 2.4, we show our result of the SFR$_{H\alpha}$ computed for the MPA-NYU SDSS Main sample, as a function of redshift in the range of interest $0.02 < z < 0.22$. We find good agreement with the SFRs estimates presented by Gunawardhana et al. [118] for both SDSS and GAMA galaxies at $z < 0.35$. The bottom panel shows that the SFRs computed from [OII] and Hα lines are strongly correlated. The [OII] emission lines can be then used as a SFR indicator, particularly at higher redshifts [113].

2.4. Measurements

2.4.1 Correlation functions

We measure the two-point correlation function, $\xi(r_p, \pi)$, of the volume-limited samples extracted from the SDSS DR7 Main galaxy sample using the Landy-Szalay estimator [175]

$$\xi(r_p,\pi) = \frac{DD(r_p,\pi) - 2DR(r_p,\pi) + RR(r_p,\pi)}{RR(r_p,\pi)}, \tag{2.14}$$

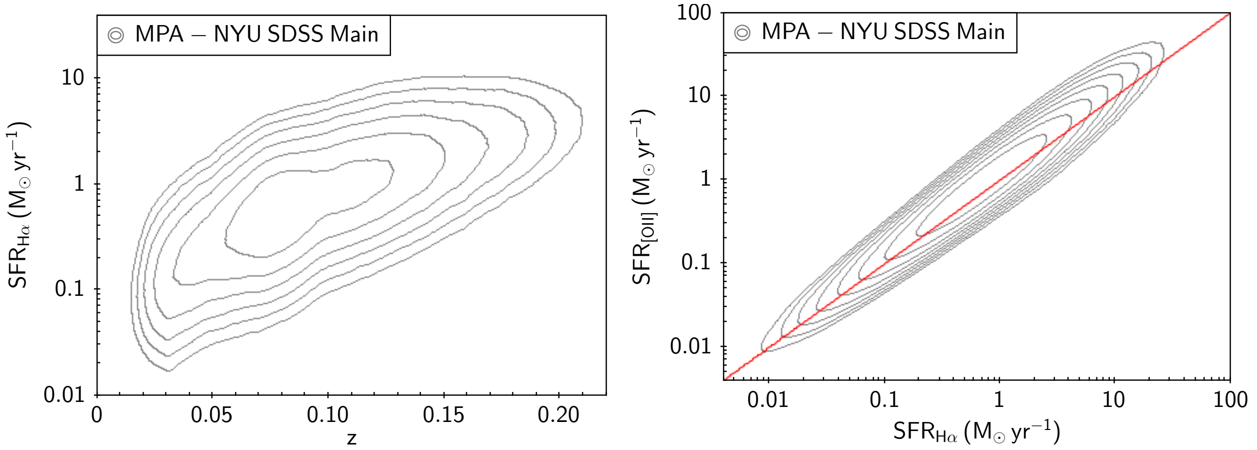

Figure 2.4: *Top:* SFR$_{H\alpha}$ as a function of redshift for the MPA-NYU SDSS Main emission-line galaxy sample. We find good agreement with previous SDSS and GAMA results [118] at $z < 0.35$. *Bottom:* SFR$_{[H\alpha]}$ versus SFR$_{[OII]}$. The two quantities are strongly correlated, then they can be both used as robust star formation indicators.

where r_p and π are, respectively, the perpendicular and parallel components of the redshift-space distance $s = \sqrt{r_p^2 + \pi^2}$, with respect to the line of sight (LOS). The quantities DD, DR and RR are the weighted and normalized data-data, data-random and random-random pair counts.

To reduce the redshift-space distortion effects visible on small scales in the clustering as an elongate feature, the so-called "finger of god", we integrate $\xi(r_p, \pi)$ along LOS and obtain the projected correlation function

$$w_p(r_p) = 2\int_0^\infty \xi(r_p, \pi)d\pi. \tag{2.15}$$

The integral above is computed by performing 15 logarithmic bins in $r_p$in the range $0.1 - 30\,h^{-1}$Mpc of width, and 20 linear π bins in the range $0 - 40\,h^{-1}$Mpc with $\Delta\pi = 2\,h^{-1}$Mpc.

Beside this statistic, we also measure the multipole moments of the redshift-space 2PCF, which is defined by expanding the 3D clustering estimator in Eq. 2.14 as [128]

$$\xi(s, \mu) = \sum_l \xi_l(s) P_l(\mu), \tag{2.16}$$

where μ is the cosine of the angle between s and the line-of-sight direction, and P_l is the

l-th order Legendre polynomial. To characterize the SDSS clustering, we focus only on the monopole $\xi_0(s)$ and the quadrupole $\xi_2(s)$ moments. For s we use the same binning scheme adopted for r_p, while for μ we do 40 linear bins in $[-1, 1]$.

2.4.2 Randoms

To correctly estimate our clustering statistics, we build suitable random catalogs including the angular and radial footprint of our data samples. We adopt the NYU-VAGC[9] `$LSS_REDUX/sample/random` catalogs that contain random points distributed with equal surface density across the area covered by the SDSS DR7 Main sample geometry and outside the bright star mask. The SDSS footprint is divided in sectors, i.e. non-overlapping regions which show a different completeness. The completeness (i.e. `FGOT` flag in the NYU-VAGC catalogs) in each sector is defined as the number of galaxies that are spectroscopic targets (i.e. have obtained redshift) over the total number of galaxies in the sector. When building our randoms, we take into account the variation of this completeness across the sky by downsampling the catalog with equal surface density in a random fashion using the completeness as a probability function. We then assign randoms the redshifts using the "shuffle" method [8].

2.4.3 Clustering weights

To compute the pair counts in Eq. 2.14, a few important corrections must be taken into account. This is done by assigning a series of weights to each object in the real and random catalogues. First, to correct for angular incompleteness, after diluting the randoms as described in Section 2.4.2, we weight both data and randoms by an angular weight given by the inverse of the sector completeness, $w_{ang} = 1/\texttt{FGOT}$.

The SDSS spectrographs are fed by optical fibers plugged on plates, which must be separated by an angular distance of 55", corresponding to $r_p \sim 0.13\,h^{-1}$Mpc at the mean redshift

[9] http://cosmo.nyu.edu/blanton/vagc/

of the sample, $z = 0.1$. We thus limit the measurements in to scales larger than that. It is then not possible to obtain spectra of all galaxies with neighbors closer than this angular distance in one single observation, and this limitation is commonly known as "fiber collision" problem [330, 191]. The problem is alleviated in sectors covered by multiple exposures but, in general, it is impossible to observe all the objects in crowded regions. Following [258], we correct the data for fiber collision by implementing a weight, w_{fc}, whose default value is 1 for all galaxies in the sample. We assign every galaxy whose redshift was not observed for fiber collision the redshift of the first neighbor closer than 55" which is good spectroscopic target, and we upweight by one the w_{fc} value of that neighbor.

Because we are using volume-limited samples, we do not need to apply any radial weight. We finally combine the two corrections above in a total weight [261]:

$$w_{tot} = w_{fc}\, w_{ang}. \tag{2.17}$$

2.4.4 Error estimation

We estimate the errors on our clustering measurements using the jackknife re-sampling technique [245, 303, 199, 213, 214, 258, 7]: we divide our data sample in $N_{res} = 200$ sub-samples containing about the same number of galaxies. We then compute the clustering of our sample excluding each time one of the re-samplings. The jackknife covariance matrix for N_{res} re-samplings is computed by

$$C_{ij} = \frac{N_{res} - 1}{N_{res}} \sum_{a=1}^{N_{res}} (\xi_i^a - \bar{\xi}_i)(\xi_j^a - \bar{\xi}_j), \tag{2.18}$$

where $\bar{\xi}_i$ is the mean jackknife correlation function estimate in the specific i^{th} bin,

$$\bar{\xi}_i = \sum_{a=1}^{N_{res}} \xi_i^a / N_{res}. \tag{2.19}$$

The overall factor in Eq. 3.9 accounts for the lack of independence between the N_{res} jackknife configurations: from one copy to the next, only two sub-volumes are different or, equivalently, $N_{res} - 2$ sub-volumes are the same [214].

2.5. Interpretation

We interpret our clustering measurements by mapping them onto the MultiDark[10] [MDPL; 167] N-body cosmological simulation with Planck cosmology [236]. The simulation box is $1\,h^{-1}$ Gpc on a side, with 3840^3 particles and a mass resolution of $1.51 \times 10^9\,h^{-1}\mathrm{M}_\odot$. It represents the best compromise between resolution and volume available to date. For the current analysis, we apply the SUrvey GenerAtoR [SUGAR; 257] algorithm to the MultiDark ROCKSTAR snapshots in the redshift range $0.02 < z < 0.22$ to produce light-cones with the same angular footprint of the SDSS data and about twice the area ($\sim 12,000$ deg^2). The advantage of using this method – instead of a single simulation snapshot at the mean redshift of the sample – is that it includes the redshift evolution, and accounts for those volume effects, as the cosmic variance or the fluctuations of the galaxy number density, that are observed in the data, and a single simulation snapshot cannot capture. In fact, because of its constant redshift and much larger volume, in a single MDPL realization the cosmic variance contribution is negligible compared to both light-cone and real data. The disadvantage of this approach is the limited volume: in fact, the maximum possible aperture for a light-cone built in a simulation volume of $1\,h^{-3}$ Gpc^3 is small compared to the original box size, i.e. $\sim 0.02\,h^{-3}$Gpc^3.

We construct light-cones using all the MultiDark halos in the redshift range $0.02 < z < 0.22$ and fix the value of the satellite fraction, f_{sat}, in our mocks to match the observed galaxy number density. We populate these halos with the galaxies of the SDSS Main volume-limited samples (2.1) by using a (Sub)Halo Abundance Matching [SHAM; 166, 302] prescription whose proxies are the galaxy luminosity and the halo maximum circular velocity over its

[10] https://www.cosmosim.org

entire history, V_{peak}. The SHAM method is based on the assumption that more luminous galaxies reside in more massive halos. We then tune the scatter parameter, σ, in the SHAM and the satellite fraction value, f_{sat}, to correctly reproduce the clustering amplitude in each one of the volume-limited samples. The satellite fraction as a free parameter in our models is necessary to correctly fit the small-scale clustering. If we do not enhance f_{sat}, our clustering predictions will be 20% lower than the data at $r \lesssim 1\,h^{-1}$Mpc, and the discrepancy will increase on smaller scales. This is the general procedure. Specifically, for the M_r and [OII] measurements there are some modifications to account for.

The advantage of modeling the SDSS Main 2PCF in M_r thresholds is that the SDSS galaxies are complete in $r-$band luminosity [203], then we only need to build one MDPL light-cone for the complete SDSS Main sample with the same number density of the data and, by tuning the scatter parameter and the f_{sat} value, we are able to precisely match the observed clustering amplitude in each one of the M_r sub-samples. We vary the satellite fraction value to optimize the agreement between data and model at the 1-halo level. From the V_{peak} values, we then derive the typical mean host halo masses for the SDSS Main galaxies (see Section 2.6.1).

Reproducing the [OII] clustering measurements is slightly more complicated, since we have to take into account that the ELG sample ($10^{38}\,\mathrm{erg\,s^{-1}} \lesssim L_{[O_{II}]} \lesssim 10^{42}\,\mathrm{erg\,s^{-1}}$) is not complete in [OII] luminosity [see 99, 64]. Thus we need to down-sample our mock galaxy catalog to the observed ELG number density in each [OII] volume-limited sample, to match the clustering measurements accounting for the ELG incompleteness. We generate a light-cone for each one of the [OII] samples, and separately compute the velocity distribution of central and satellite halos. Then, in the SHAM assignment, we force these distributions to assume a Gaussian shape depending on three parameters: the mean V_{peak}, the half-width σ_V, and the satellite fraction f_{sat}. This is done by imposing two different selections, one for centrals and one for satellites, both based on a Gaussian realization, $\mathcal{N}_{sat}(V_{peak}, \sigma_V, f_{sat})$, depending on the three parameters above, and normalized by the ELG desired number density. A similar procedure has been applied in [99], see Chapter 4. We then bin our mock catalogs in V_{peak}. In each bin, we compute the probabilities to select central and satellite mocks as

$$\begin{aligned} P_{sat}(V_{peak}, \sigma_V, f_{sat}) &= N_{sat}^{gauss}(V_{peak}, \sigma_V, f_{sat})/N_{sat}(V_{peak}) \\ P_{cen}(V_{peak}, \sigma_V, f_{sat}) &= N_{cen}^{gauss}(V_{peak}, \sigma_V, f_{sat})/N_{cen}(V_{peak}), \end{aligned} \tag{2.20}$$

where N_{sat}^{gauss} (N_{cen}^{gauss}) is the number of satellite (central) mock galaxies resulting from the Gaussian selection, and N$_{sat}$ (N$_{cen}$) is the total number of satellite (central) halos in the light-cone in the velocity bin considered. We apply the SHAM prescription drawing central and satellite halos from our MultiDark light-cone using the PDFs in Eq. 2.20. The variation of the SHAM scatter parameter, σ, is accounted for in the assignment, but its effect is highly degenerate with V_{peak} and σ_V. If the amplitude of the Gaussian realization above is higher than the amplitude of the MDPL halo velocity function, we compensate the "missing" halos by picking substitute halos in the lower tail of the V_{peak} distribution. This procedure will deform the shape of the Gaussian selection, and the effect of the distortion will be proportional to the number of missing halos one needs to replace. As a result, the mean V_{peak} of the final PDF will be displaced towards slightly lower values.

The procedure described above guarantees the reliability of our model galaxies, since it incorporates the ELG [OII] luminosity incompleteness, the scatter observed between halo velocities and galaxy luminosities (encoded in the SHAM scatter parameter, σ), and allows to correctly reproduce both the ELG number density and the clustering signal. From the V_{peak} values found from this analysis, we infer the typical mean host halo masses for the SDSS ELG sample. Our results are presented in Section 2.6.2.

2.6. Clustering results

In what follows we present our SDSS Main DR7 clustering results as a function of the M_r and [OII] luminosities. They show that galaxy clustering correlates with both M_r and [OII] luminosity – i.e., more luminous galaxies are more strongly clustered than their fainter counterparts. We find remarkable agreement with our MultiDark model galaxies. The MultiDark light-cone is built including the complete redshift evolution over the whole range of interest, $0.02 < z < 0.22$, and this allows to precisely match the 2PCF measurements as a function of the luminosity both on small and larger scales. Despite its small volume, the light-cone reproduces well the clustering measurements and gives reliable prediction of the typical host halo masses and satellite fraction for the SDSS galaxies.

2.6.1 Clustering as a function of the M_r luminosity

In the top left panel of Figure 2.5, we present the measured projected 2PCF of the SDSS Main DR7 M_r volume-limited samples defined in Section 2.3.1). In the top right panel, we compare these measurements (points) with our prediction from the MDPL light-cone (lines). Consistently with several previous works [311, 334, 122], we find that more luminous galaxies have a higher clustering amplitude. We also display the agreement between the SDSS data and our model galaxies for the monopole (bottom left plot) and quadrupole (bottom right) moments of the two-point correlation function. Just for clarity, when we compare data and models, we shift the $w_p(r_p)$ values by $0.2\,\mathrm{dex}$ and $s^2\xi_{0,2}(s)$ by $20\,h^{-2}\,\mathrm{Mpc}^2$ to avoid overlapping. From our SHAM analysis, we infer the mean host halo mass (M_h) and satellite fraction (f_{sat}) for each sample, constraining the SDSS Main galaxy halo occupation distribution as a function of the r-band luminosity. The typical mean M_h and f_{sat} values for the r-band luminosity samples are reported in Table 2.3 and indicate that more luminous galaxies reside in more massive halos where the fraction of satellites is lower. Figure 2.6 displays the satellite Halo Occupation Distribution (HOD) derived from our MDPL model galaxies. We find that our mocks are generally richer of satellites compared to the HOD analysis by Guo et al. [122]. Such a discrepancy is due to the different way of populating halos with galaxies in the SHAM and HOD models. The SHAM prescription is applied by performing a cut (see Eq. 1.35) in the halo and galaxy number densities, and excluding any object below a certain V_{peak} and corresponding luminosity. The HOD formulation does not assume such a cut, and allows one to include any kind of halo. For this reason, compared to our SHAM recipe, Guo et al. [122] assign more satellites to more massive halos or, in other words, the SHAM cut excludes satellites with small V_{peak} values in more massive halos. In order to reproduce their satellite HOD prediction (i.e. number of satellites per halo mass), we therefore need to include satellite mocks with lower V_{peak} values than the ones assigned by the SHAM. This is exactly what our model does. By increasing f_{sat}, we assign additional satellites that will distribute over the whole mass range considered. The effect of the satellite enhancement in the clustering 1-halo term is $\xi_{1h} \propto (N_{cen}\, N_{sat} + N_{sat}\, N_{sat})$ [200], which means that a difference of m satellites will result in a $\sim \mathcal{O}(m^2)$ effect in the small-scale clustering amplitude. Another important difference between our SHAM model and the HOD scheme adopted in Guo et al. [122] is that we place the satellite mocks at the sub-halo positions which are provided in the MultiDark halo catalogs, while they draw random dark matter particles for the position of their satellites (i.e. their satellite velocity distribution is consistent with that of the dark matter). To supply the peculiar velocity values to the satellites, which we take directly from the MDPL simulations, they apply the velocity bias [125] correction. In addition, our light-cones include the whole redshift evolution in the range $0.02 < z < 0.22$, and let the galaxy number density vary with redshift within the volume considered, as naturally happens in the Universe. This makes that our $n(z)$ distribution fluctuates around the mean value of the single MDPL realization, as in [334, 122].

The volume limitation of our method is visible in the high-mass tail of the distribution, where our f_{sat} curves are interrupted. Beyond $10\,h^{-1}\,$Mpc, the fluctuations due to cosmic variance are no longer negligible, and affect all the clustering results shown above. The remarkable agreement we find between the SDSS data and our MultiDark model galaxies in the quadrupole shows the robustness of our f_{sat} previsions. The satellite fraction behavior is strongly correlated with the peculiar velocities of the satellites within their parent halos, and this information is carried by $\xi_2(s)$. Our results tell us that we are correctly modeling all the clustering statistics considered by applying a straightforward SHAM prescription to

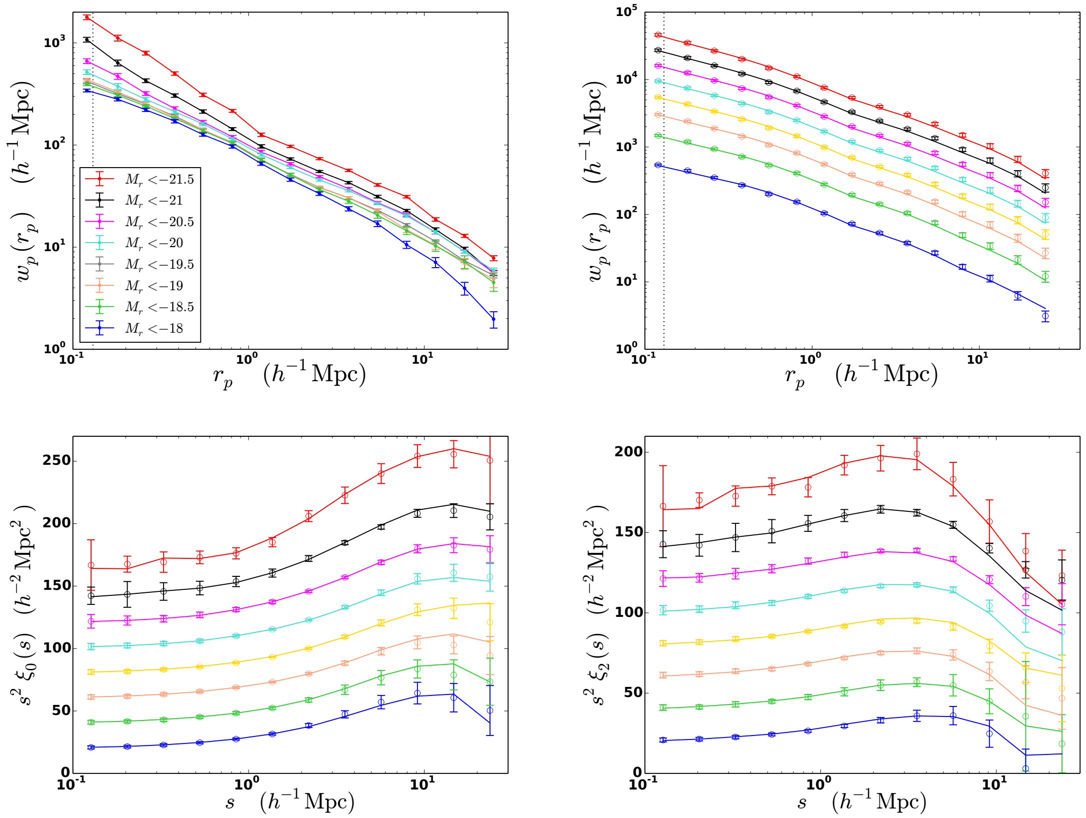

Figure 2.5: *Top left:* projected correlation function of the SDSS Main DR7 M_r samples listed in Table 2.3. The errors are estimated using 200 jackknife re-samplings. The vertical line in $w_p(r_p)$ is the fiber collision threshold computed at the mean redshift of the sample, $z \sim 0.1$. *Top right*: SDSS Main $w_p r_p$ measurements (points) versus our MultiDark model galaxies (lines). The typical host halo mass and satellite fraction values for each M_r sample are reported in Table 2.3.*Bottom left:* SDSS monopole correlation function (points) versus MultiDark mocks (lines). *Bottom right*: SDSS quadrupole (points) versus model galaxies (lines). Just for clarity, when we show both data and models, we shift the $w_p(r_p)$ values by 0.2 dex and $s^2\xi_{0,2}(s)$ by $20\,h^{-2}$ Mpc^2 to avoid overlapping.

the MultiDark light-cone, and letting vary the satellite fraction. These models naturally arise from the simulation, with no need of introducing any velocity bias modification, nor additional assumptions. Our mocks include, by construction, the redshift evolution and those volume effects, as number density fluctuations or cosmic variance, which are naturally observed in the Universe and a single simulation snapshot cannot capture. The cosmic variance contribution in our light-cones is higher than in the single MDPL realization, but still lower compared to the real effect observed in the SDSS measurements because of the volume: the light-cone covers about twice the volume of the data.

$M_{0.1_r}^{min} - 5\log h$	z_{max}	$\bar{n}_g$	mean M_h	mean f_{sat}	χ^2/dof
		$[10^{-3}h^3\mathrm{Mpc}^{-3}]$	$[h^{-1}\mathrm{M}_\odot]$		
-18.0	0.041	29.68	9.60×10^{11}	38.34	1.02
-18.5	0.053	21.02	1.49×10^{12}	34.78	2.20
-19.0	0.064	14.82	1.93×10^{12}	33.89	4.21
-19.5	0.085	11.01	2.57×10^{12}	29.72	2.54
-20.0	0.106	6.03	4.39×10^{12}	25.28	2.11
-20.5	0.132	2.95	7.84×10^{12}	18.78	3.18
-21.0	0.159	1.09	1.45×10^{13}	16.89	1.76
-21.5	0.198	0.27	3.28×10^{13}	13.33	1.64

Table 2.3: Mean host halo mass and satellite fraction (in units of percent) of the SDSS Main volume-limited samples in r-band absolute magnitude thresholds. Our mean f_{sat} values are generally higher than Guo et al. [122]. This is due to the different way of assigning galaxies halos in the SHAM and the HOD models. See the text for details. In the last column we report the χ^2 values of the $w_p(r_p)$ model fits computed with 12 dof.

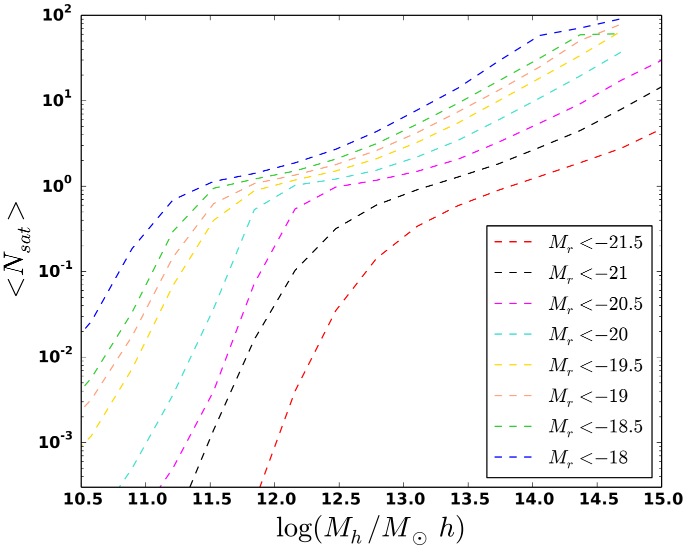

Figure 2.6: Halo occupation distribution of the MultiDark satellite mock galaxies. Our mocks are generally richer of satellites compared to the HOD analysis by Guo et al. [122], and this discrepancy is due to the different selection process in the HOD and SHAM methods. See the text for details.

z_{max}	$\mathrm{L}_{[O_{II}]}^{min}$	$\bar{n}_g$	mean M_h	mean f_{sat}	χ^2/dof
	[erg s^{-1}]	[$10^{-3}h^3\mathrm{Mpc}^{-3}$]	[$h^{-1}\mathrm{M}_\odot$]		
0.05	1×10^{39}	25.57	1.24×10^{12}	33.37	1.82
0.09	3×10^{39}	12.92	1.98×10^{12}	27.91	2.37
0.14	1×10^{40}	4.95	3.39×10^{12}	22.49	3.62
0.17	3×10^{40}	2.13	4.93×10^{12}	19.43	2.17
0.20	1×10^{41}	0.65	6.78×10^{12}	18.01	5.08

Table 2.4: Mean host halo mass and satellite fraction (in units of percent) of the SDSS [OII] ELG samples. For the $w_p(r_p)$ fits we use 11 dof.

2.6.2 Clustering as a function of the [OII] emission-line luminosity

The MPA-NYU SDSS Main clustering measurements as a function of the [OII] emission-line luminosity are presented in Figure 2.7, top left panel. We show the agreement with our MultiDark model galaxies in the projected (top right panel), monopole (bottom left) and quadrupole (bottom right) two-point correlation functions. When we compare data and models, we shift the $w_p(r_p)$ values by 0.2 dex and $s^2\xi_{0,2}(s)$ by $20\,h^{-2}\,\mathrm{Mpc}^2$ to avoid overlapping. Analogously to the M_r results, we find that more luminous [OII] galaxies are more clustered than their fainter companions. The dark matter halos hosting the SDSS [OII] ELGs, however, span a much smaller mass range than the halos hosting Our SHAM predictions for the mean [OII] host halo masses and satellite fractions are given in Table 2.4 and indicate a similar behavior to the M_r volume-limited samples: ELGs with higher [OII] luminosities tend to occupy more massive halos, with a lower satellite fraction. We find that [OII] emission-line galaxies at $z \sim 0.1$ live in halos with mass $\sim 10^{12}\,h^{-1}\mathrm{M}_\odot$, analogously to the ELG scenario found at $z \sim 0.8$ by Favole et al. [99]. The satellite fraction in the SDSS [OII] ELG samples considered in the local Universe varies between $\sim 18\%$ and $\sim 33\%$. In Figure 2.8 is displayed the clustering variation as a function of the three model parameters: V_{peak} (top left panel), σ_V (top right) and f_{sat} (bottom). In each one of the plots we allow to vary only one parameter at a time, and the other two are fixed at their best-fit values: $V_{peak} = 303\,\mathrm{km}\,s^{-1}$, $\sigma_V = 140\,\mathrm{km}\,s^{-1}$ and $f_{sat} = 18\%$.

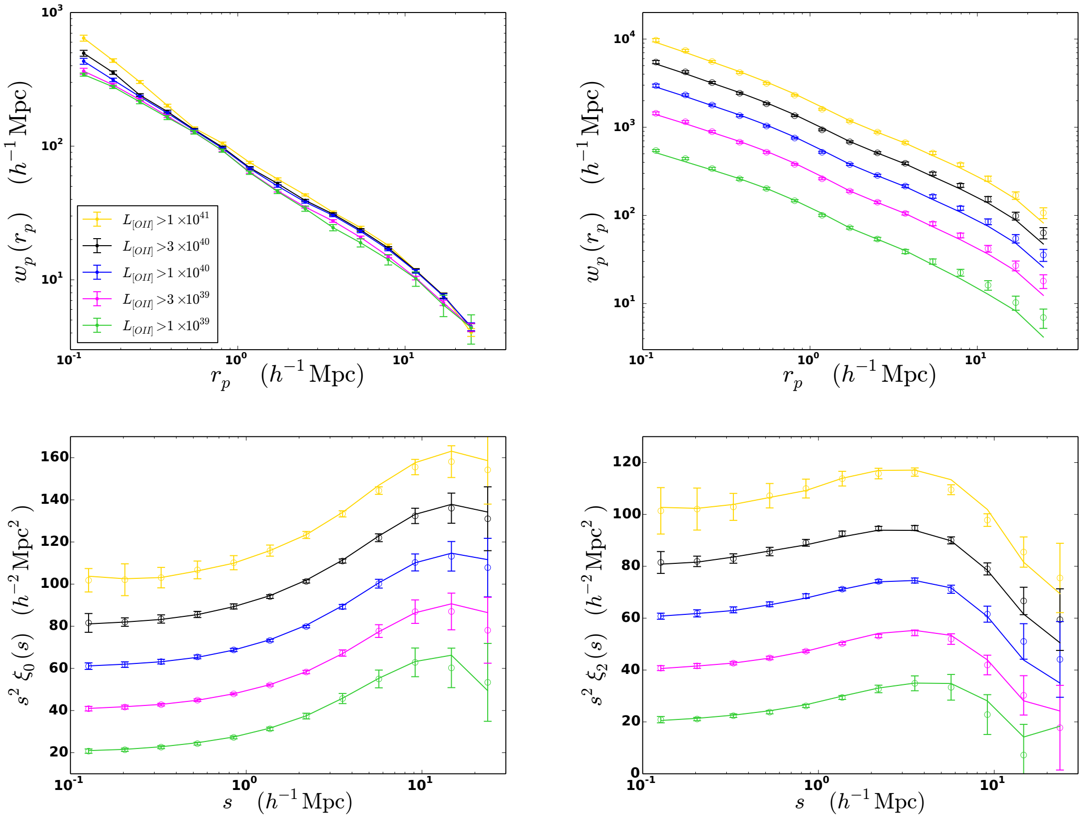

Figure 2.7: *Top line:* MPA-NYU SDSS Main projected 2PCF (points) of the volume-limited samples in [OII] luminosity thresholds defined in Table2.2 versus our MultiDark model galaxies (lines). The errors are estimated using 200 jackknife re-samplings. *Bottom line:* monopole (left) and quadrupole (right) correlation functions. The typical host halo mass and satellite fraction values for each M_r sample are reported in Table 2.4. Just for clarity, when we plot the data and the model together, we shift $w_p(r_p)$ by 0.2 dex and $s^2\xi_{0,2}(s)$ by $20\,h^{-2}$ Mpc^2 to avoid overlapping.

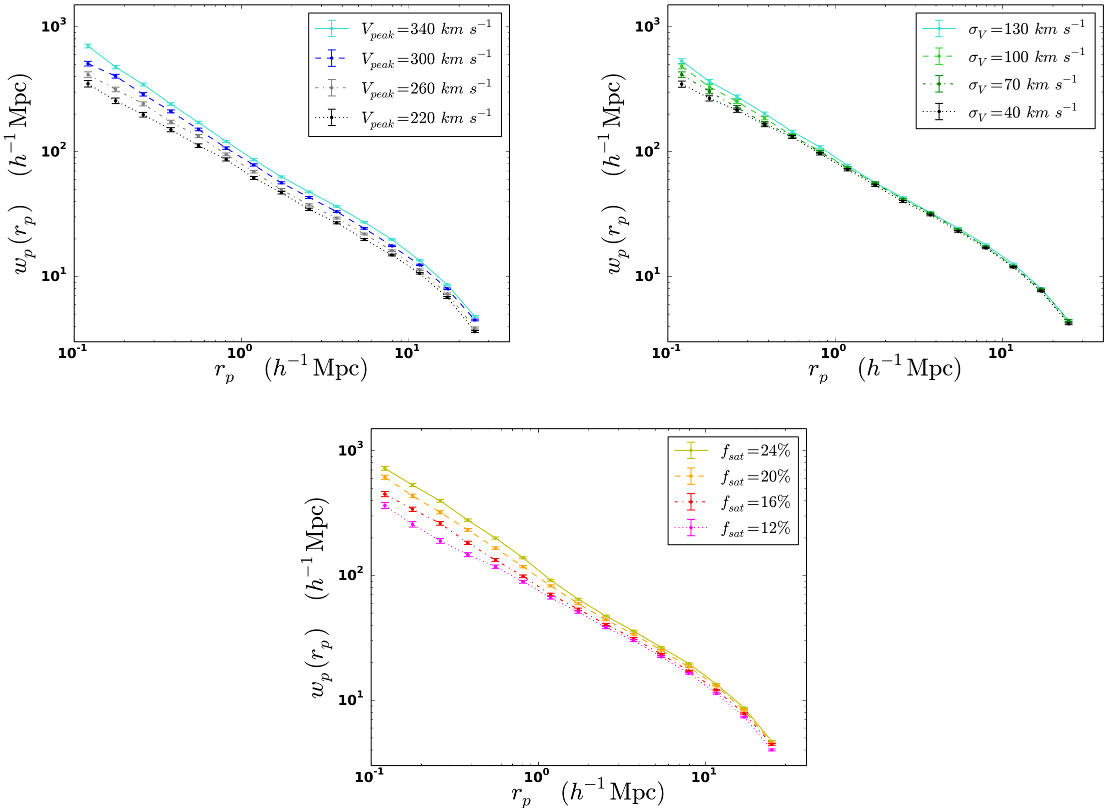

Figure 2.8: [OII] ELG clustering dependence on our model parameters: V_{peak} (top left), σ (top right) and f_{sat} (bottom). In each panel we let vary only one parameter at a time and the other two are fixed at the best-fit values: $V_{peak} = 303\,\mathrm{km}\,s^{-1}$, $\sigma_V = 140\,\mathrm{km}\,s^{-1}$ and $f_{sat} = 18\%$.

2.7. Discussion and conclusions

We have studied the dependence of galaxy clustering both on the r-band and the [OII] emission-line luminosities in the local Universe. We have selected the SDSS Main galaxy sample from the NYU Value Added Galaxy Catalog by applying the Main target selection criteria [292] defined in Section 2.3.1. The final sample contains about 520,000 galaxies, from which we have extracted eight volume-limited samples using suitable redshift cuts and r-band absolute magnitude thresholds. In those samples we have measured the projected, monopole and quadrupole two-point correlation functions. We have estimated the clustering errors using 200 jackknife re-samplings. In agreement with previous works [334, 122], we find that more luminous galaxies are more strongly clustered.

Then we have spectroscopically matched our SDSS DR7 Main galaxy selection to the MPA-JHU DR7 release of spectrum measurements and, for those galaxies surviving the matching, we have merged [OII], Hα and Hβ emission-line properties. We have computed (see Section 2.3.2) the [OII] luminosities of these galaxies, imposing a minimum [OII] flux of 10^{-16} erg cm^{-2} s^{-1} to exclude objects with too short exposure time. The final sample includes about 433,000 [OII] emission-line galaxies. We choose not to include any possible elliptical galaxy which is central for some of the [OII] ELGs considered, because our goal is to characterize the clustering properties of the emission-line galaxies only, not of both populations. From the MPA-NYU ELG merged catalog, we have selected volume-limited samples in [OII] luminosity thresholds, and there we have estimated the projected, monopole and quadrupole two-point correlation functions. We find a strong correlation between the clustering signal and the strength of the [OII] lines.

To interpret our measurements, we have built suitable light-cones (§2.5) by applying the SUGAR [257] algorithm to the snapshots of the MultiDark Planck cosmological simulation [167] with $L_{box} = 1\,h^{-1}$Gpc available in the redshift range of interest, $0.02 < z < 0.22$. In this way, we guarantee that our model galaxies account for the complete redshift evolution over the z range considered. We have adopted a (Sub)Halo Abundance Matching technique to assign the SDSS Main galaxies the MultiDark halos in the light-cones assuming, as halo proxy, its maximum circular velocity over its entire history, V_{peak}, and the luminosity for the SDSS galaxies. We build our clustering models as a function of the r-band absolute magnitude by tuning two parameters: the scatter, σ, in the SHAM assignment and the satellite fraction, f_{sat}. From this SHAM analysis, we can derive the typical mean halo masses and satellite fraction values of the SDSS Main galaxies as a function of the r-band luminosity. Our predictions are reported in Table 2.3, and indicate that more luminous galaxies reside in more massive halos where the fraction of satellites is lower. Figure 2.6 displays the satellite halo occupation distribution derived from our MDPL model galaxies. We find that our mocks are generally richer of satellites compared to the HOD analysis by Guo et al. [122]. Such a discrepancy is due to the different way of populating halos with galaxies in the SHAM and the HOD models. The SHAM prescription is applied by performing a cut (see Eq. 1.35) in the halo and galaxy number densities, which excludes any object below a certain V_{peak} and corresponding luminosity. The HOD formulation does not assume such a cut, and allows one to include any kind of halo. For this reason, compared to our SHAM recipe, Guo et al. [122] assign more satellites to more massive halos or, in other words, the SHAM cut excludes satellites with small V_{peak} values in more massive halos. In order to reproduce their satellite HOD prediction (i.e. number of satellites per halo mass), we therefore need to include satellite mocks with lower V_{peak} values than the ones assigned by the SHAM. This is exactly what our model does. By increasing f_{sat}, we assign additional satellites that will distribute over the whole mass range considered. The 1-halo term in the clustering is $\xi_{1h} \propto (N_{cen}\, N_{sat} + N_{sat}\, N_{sat})$ [200], then a difference of m satellites will result in a $\mathcal{O}(m^2)$ effect in the small-scale clustering amplitude.

Another important difference between our models is that we place the satellite mocks at the sub-halo positions provided in the MultiDark halo catalogs, while Guo et al. [122] draw random dark matter particles for the position of their satellites. To supply the peculiar velocity values to the satellites, which we take directly from the MDPL simulations, they apply the velocity bias [125] correction. The remarkable agreement we find between the SDSS data and our model galaxies in the quadrupole shows the robustness of our f_{sat} predictions, which naturally arise from the MDPL simulation, with no need of introducing any velocity bias modification, nor additional assumptions. Our mocks include, by construction, the redshift evolution and mimic those volume effects, as number density fluctuations and cosmic variance, which are observed in the data. The cosmic variance contribution in the mocks cannot achieve the real effect observed in the data because of the volume: the light-cone has twice the volume of the SDSS data. However, this is an improvement in the model reliability, compared to using a single simulation snapshot. These ingredients make our model galaxies a realistic and accurate representation of the SDSS DR7 Main galaxy sample.

Reproducing the [OII] clustering measurements is slightly more complicated than the M_r results, since emission-line galaxies are incomplete in [OII] luminosity [see 99, 64]. We therefore need to down-sample our mock galaxies to match the observed ELG number density in each one of the [OII] volume-limited samples. To do that, we calculate the satellite and central MultiDark halo velocity functions, and we separately impose them a Gaussian selection (see Section 2.5) depending on three parameters – the mean V_{peak}, the half-width σ_V, and the satellite fraction f_{sat} – and normalized to the ELG desired number density. We then bin our light-cones in V_{peak} and, in each bin, we compute the probabilities of selecting satellite and central mocks respectively as $P_{sat}(V_{peak}, \sigma_V, f_{sat}) = N_{sat}^{gauss}(V_{peak}, \sigma_V, f_{sat})/N_{sat}(V_{peak})$ and $P_{cen}(V_{peak}, \sigma_V, f_{sat}) = N_{cen}^{gauss}(V_{peak}, \sigma_V, f_{sat})/N_{cen}(V_{peak})$, where N_{sat}^{gauss} (N_{cen}^{gauss}) is the number of satellite (central) mocks resulting from the Gaussian selection, and N_{sat} (N_{cen}) is the total number of satellite (central) halos in the simulation in the velocity bin considered. We apply the SHAM prescription drawing central and satellite halos from our MultiDark light-cone using the PDFs above. The variation of the scatter parameter (σ) in the SHAM is accounted for in the assignment, but its effect is highly degenerate with V_{peak} and σ_V. This procedure guarantees the reliability of our model galaxies, since it incorporates the ELG

[OII] luminosity incompleteness, the scatter observed between halo velocities and galaxy luminosities (encoded in the SHAM parameter, σ), and allows to correctly reproduce both the ELG number density and clustering signal. Finally, from the V_{peak} values we find from this analysis, we infer the typical mean host halo masses for the SDSS ELG sample. Our results (see Section 2.6.2) demonstrate that SDSS [OII] emission-line galaxies at $z \sim 0.1$ live in halos with mass $\sim 10^{12}\,h^{-1}\mathrm{M}_{\odot}$, analogously to the ELG scenario found at $z \sim 0.8$ by Favole et al. [99] (see Chapter 4), and their mean satellite fraction varies between $\sim 18\%$ and $\sim 33\%$.

The robustness of the method presented here is demonstrated by the fact that we are able to correctly model all the three clustering statistics on small and intermediate scales, using a straightforward SHAM approach combined with light-cones. Our models naturally arise from the MultiDark simulation, with no need of additional modifications nor velocity bias corrections [125, 122]. In particular, the remarkable agreement between SDSS data and model galaxies in the quadrupole correlation function, reveals that we are modeling the satellite fraction in a reliable way. In fact, the quadrupole moment is the most sensitive statistics to the galaxy peculiar velocities on small scales, which drive the satellite fraction and the amplitude of the 1-halo term in the correlation function. Our light-cones are a reliable representation of the data because they include the full redshift evolution and those volume effects, as cosmic variance or galaxy number density fluctuations, which are visible in the real Universe and a single MultiDark realization cannot mimic. The cost, compared to using a single MultiDark snapshot, is the limitation in volume (see §2.5) that makes these models less accurate on larger scales. For this reason, we focus our analysis at $s \lesssim 30\,h^{-1}\mathrm{Mpc}$.

The models presented here provide accurate clustering prediction using a straightforward SHAM prescription applied to MDPL light-cones. The method could certainly be refined by taking into account the halo assembly bias [i.e., 183, 136] to differentiate the morphology and age of halos hosting ELGs from those of halos hosting elliptical galaxies. This is an interesting issue, already addressed by several authors [e.g., 76, 327, 9, 54, 251, 335, 136, 296, 184, 295], which we do not consider for the current analysis, and we will explore later.

This SDSS clustering study as a function of the [OII] emission-line luminosity at low redshift is particularly important in the light of new-generation wide-field spectroscopic surveys as the ongoing SDSS-IV/eBOSS survey, DESI, 4MOST, Subaru PFS and EUCLID (see Section 1.7). In fact, all these facilities will target emission-line galaxies up to redshift $z \sim 2$ to trace the baryon acoustic oscillation feature in their clustering signal. It is therefore extremely important to understand the ELG halo-galaxy connection and its evolution from the local Universe to very high redshifts. Current data lack of resolution to push the clustering analysis to very small scales, where correlations between sub-structures belonging to the same parent halo dominate. However, future space- and ground-based instruments will complement each others to provide the high-imaging quality necessary to explore those scales. Combining future data with high-resolution cosmological simulations and Semi-Analytic Models (SAMs) for galaxy formation, we will be able to better constrain and understand the galaxy halo occupation distribution on all scales and its evolution with redshift.

Beside the clustering analysis, we have used the emission-line galaxy properties to estimate the SDSS Main dust extinction and star formation rates (Sections 2.3.3 and 2.3.4). In agreement with previous studies [277, 86, 138], we find that the observed Balmer ratio Hα/Hβ exceeds the constant theoretical value $(\mathrm{H}\alpha/\mathrm{H}\beta)_{int} = 2.86$ expected for planetary nebulae in typical conditions [222, 220, 221], indicating the presence of an extinction excess in the SDSS data. Comparing the Balmer and the [OII]/Hα ratios we find they are strongly correlated, therefore this latter can be also used as a robust indicator for dust extinction. For what concerns star formation rates, our SDSS Main estimates are in good agreement with previous SDSS and GAMA results [118] at $z < 0.35$ and, consistently with [148, 206], we find that the $\mathrm{SFR}_{[OII]}$ indicator strongly correlates with the more classical estimator based on the Hα line properties. We then conclude that $\mathrm{SFR}_{[OII]}$ can be used as robust star formation tracer, especially at higher redshifts [113]. The precise determination of the star-formation history in the Universe represents one of the main goals of modern cosmology, as it is crucial to our understanding of how galactic structures form and evolve. New-generation surveys will be key to accurately determine the SFR evolution with redshift. In parallel, semi-analytic models will include SFRs as fundamental ingredient to correctly model the process of galaxy formation, allowing us to understand the complex process of structure formation and evolution.

[...] salimmo sú, el primo e io secondo, tanto ch'i' vidi de le cose belle che porta 'l ciel, per un pertugio tondo. E quindi uscimmo a riveder le stelle.

D. Alighieri, Inferno - Canto XXXIV

3

Building a better understanding of the massive high-redshift BOSS CMASS galaxies as tools for cosmology

3.1. Abstract

We explore the massive bluer star-forming population of the Sloan Digital Sky Survey (SDSS) III/BOSS CMASS DR11 galaxies at $z > 0.55$ to quantify their differences, in terms of redshift-space distortions and large-scale bias, with respect to the luminous red galaxy sample. We perform a qualitative analysis to understand the significance of these differences and whether we can model and reproduce them in mock catalogs. Specifically, we measure galaxy clustering in CMASS on small and intermediate scales ($r \lesssim 50$ h^{-1}Mpc) by computing the two-point correlation function — both projected and redshift-space — of these

galaxies, and a new statistic, $\Sigma(\pi)$, able to provide robust information about redshift-space distortions and large-scale bias. We interpret our clustering measurements by adopting a Halo Occupation Distribution (HOD) scheme that maps them onto high-resolution N-body cosmological simulations to produce suitable mock galaxy catalogs. The traditional HOD prescription can be applied to the red and the blue samples, independently, but this approach is unphysical since it allows the same mock galaxies to be either red or blue. To overcome this ambiguity, we modify the standard formulation and infer the red and the blue models by splitting the full mock catalog into two complementary and non-overlapping sub-mocks. This separation is performed by constraining the HOD with the observed CMASS red and blue galaxy fractions and produces reliable and accurate models.

3.2. Introduction

In the last decade, an enormous effort has been spent to explore the formation and evolution of the large scale structure of our Universe. The standard cold dark matter (ΛCDM) model with cosmological constant, together with the theory of cosmic inflation, has become the leading theoretical picture in which structures can form, providing a clear prediction for their initial conditions and hierarchical growth through gravitational instability [e.g., 243]. Testing this model requires one to combine large N-body simulations with measurements from last generation large-volume photometric and spectroscopic galaxy surveys, as the Sloan Digital Sky Survey [SDSS; 329, 120, 275], and the SDSS-III Baryon Oscillation Spectroscopic Survey [BOSS; 91, 81]. In particular, BOSS has been able to measure the Baryon Acoustic Oscillation (BAO) feature in the clustering of galaxies and Lyman-α forest with unprecedented accuracy, by collecting spectra of 1.5 million galaxies up to z=0.7 [8], over a 10,000 deg^2 area of sky, and about 160,000 Lyman-α forest spectra of quasars in the redshift range $2.2 < z < 3$ [273].

The ΛCDM paradigm claims that galaxies form at the center of dark matter halos, thus estimating the clustering features of such complex structures, is currently one of the main targets of modern cosmology [171]. Despite the recent dramatic improvement in the observational data, what primarily prevents us from achieving this goal immediately is the theoretical uncertainty of galaxy bias i.e., the difference between the distribution of galaxies and that of the matter. Galaxies are treated as biased tracers of the underlying matter distribution, and observations of their clustering properties are used to infer those cosmological parameters that govern the matter content of the Universe. In this context, the Halo Occupation Distribution [HOD; 28, 170, 336, 337] framework has emerged as a powerful tool to bridge the gap between galaxies and dark matter halos, providing a theoretical framework able to characterize their mutual relation in terms of the probability, $P(N|M)$, that a halo of virial mass M contains N galaxies of a given type. At the same time, it provides a robust prediction of the relative spatial and velocity distributions of galaxies and dark matter within halos. In this approach, the use of large-volume N-body cosmological simulations is crucial to produce reliable maps of the dark matter sky distribution.

In this work, we explore the red/blue color bimodality observed in the BOSS CMASS DR11 [6] galaxy sample. In order to quantify and model the differences between these two galaxy populations, we measure their clustering signal on small and intermediate scales, from $r \sim 0.1h^{-1}$Mpc up to $r \sim 50h^{-1}$Mpc. Specifically, we compute the two-point correlation function (2PCF) – both projected and in redshift-space – of these galaxies, and a new metric, $\Sigma(\pi)$, designed to extract the maximum amount of information about the small-scale nonlinear redshift-space distortions. We map our results to the MultiDark cosmological simulation [240, 253] using an HOD approach [337, 318], to generate reliable mock galaxy catalogs. In this context, we investigate whether we can find an HOD parametrization able to model both the blue and red observed clustering amplitudes, with small variations in its parameters. As an alternative to HOD models, one can interpret clustering observations with an Halo Abundance Matching (HAM) prescription [e.g., 302, 215] with the advantage of avoiding free parameters, only assuming that more luminous galaxies are associated to more massive halos, monotonically, through their number densities. HAM is a straightforward technique that provides accurate predictions for clustering measurements; nevertheless, we choose to model our CMASS clustering measurements using a five-parameter HOD scheme because it is a general method, based on a halo mass parametrization, and does not require a specific luminosity (stellar mass) function [202] to reproduce the observations.

The traditional HOD modeling reproduces well the clustering signal observed in CMASS, but it provides unphysical predictions when applied to the red and blue sub-samples, independently. In fact, in the process of populating a halo with central and satellite galaxies, it allows the same galaxy to be either red or blue i.e., to be placed in halos with different masses. To overcome this ambiguity, we propose an alternative prescription, that recovers the red and the blue models by splitting the full mock catalog into two non-overlapping sub-mocks. The separation is performed in a "natural" way by reproducing the observed CMASS red and blue galaxy fractions, as a function of the central halo mass. The resulting mocks are no longer independent – they are based on the same HOD parameter set – and the total number of degrees of freedom is reduced from 15 (three independent models, with five parameters each) to 5 (full HOD) plus 2 (galaxy fraction constraint).

We investigate the impact of redshift-space distortions on the clustering signal, both on small (1-halo term) and intermediate (2-halo level) scales. Our new metrics, $\Sigma(\pi)$, allows us to separate and quantify both the nonlinear elongation seen in the two-point correlation function below $2h^{-1}$Mpc, and the Kaiser compression at scales beyond $10h^{-1}$Mpc. We model these effects in terms of two parameters, A and G, respectively encoding the galaxy velocity dispersion with respect to the surrounding Hubble flow, and the linear large-scale bias. In agreement with several previous works [see, for instance, 311, 333, 297], we find that red galaxies are more clustered (i.e. higher peculiar velocity contribution) and biased, compared to their blue star-forming companions. Similar red/blue comparisons in terms of redshift-space distortions and linear galaxy bias have been performed in previous studies [e.g., 193, 259]. So far, however, most results for blue galaxies are for much less massive samples than CMASS. In addition, CMASS is a very large data set, and this provides a good motivation for being quantitatively exact in estimating its large-scale bias and small-scale peculiar velocities, even if the qualitative behavior is standard.

This chapter is organized as follows. In Section 3.3 we introduce the methodology used to measure and model galaxy clustering in the BOSS CMASS DR11 sample: we define the metrics used, the correlation function and the covariance estimators. We then give an overview of the MultiDark simulation, discuss the HOD formalism adopted to create mock galaxy catalogs, and introduce the analytic tools used to model both finger-of-god and Kaiser effects. In Section 3.4 we present the CMASS DR11 sample and the specific red/blue color selection used in the analysis, we illustrate how to weight the data to account for fiber collision and redshift failure effects, and outline the procedure adopted to generate randoms. Section 3.5 describes how we model our full CMASS clustering measurements building reliable mock galaxy catalogs that take into account the contribution of redshift-space distortions, and present the first results for the three metrics of interest: $\xi(s)$, $w_p(r_p)$, $\Sigma(\pi)$. In Section 3.6 we apply the traditional HOD formulation individually to the full, the red and the blue CMASS galaxy samples to create their own independent mock catalogs. Then, we present an alternative method to recover the red and blue sub-mocks from the full one, by splitting it using, as a constraint, the observed CMASS red/blue galaxy fractions. Our data versus mock $\Sigma(\pi)$ results, compared to the A, G analytic models are shown in Section 3.7. Section 3.8 reports our main conclusions.

3.3. Methods

3.3.1 Clustering measurements

We quantify the clustering of galaxies by computing the two-point correlation function i.e., the excess probability over random to find a pair of galaxies typically parameterized as a function of their co-moving separation [see, e.g., 226]. The galaxy correlation function is well known to approximate a power-law across a wide range of scales,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{3.1}$$

where r_0 is the correlation length, and γ is the power-law slope or spectral index. However, improved models [see review at 69] have been shown to better match the data [332].

The redshift-space correlation function differs from the real-space one due to the distortion effects caused by our inability to separate the peculiar velocities of galaxies from their recession velocity when we estimate distances from the redshift. These distortions introduce anisotropies in the 2PCF in two different ways. On large scales, where the linear regime holds, galaxies experience a slow infall toward an over-dense region, and the peculiar velocities make structures appear squashed in the line-of-sight direction, an effect commonly known as "Kaiser compression" [152, 129]. At smaller scales, nonlinear gravitational collapse creates virialized systems and thereby relatively large velocity differences arise between close neighbors resulting in structures appearing significantly stretched along the line-of-sight [150]. This effect is commonly referred to as the "finger-of-god"(FoG).

We are interested in using three related two-point clustering metrics: the redshift-space monopole, $\xi(s)$, the projected correlation function, $w_p(r_p)$, and a new line-of-sight focused measurement to capture small-scale redshift-space distortion effects, $\Sigma(\pi)$, which we define below. In our formalism, s represents the redshift-space pair separation, while r_p and π are the perpendicular and parallel components with respect to the line-of-sight such that $s = \sqrt{r_p^2 + \pi^2}$. We can parameterize the redshift-space correlation function as a function of redshift-space separation s or, equivalently, in terms of r_p and π. We can mitigate the impact of redshift-distortions by integrating along the line of sight to approximate the real-space clustering [79] in the projected correlation function,

$$w_p(r_p) = 2\int_0^\infty \xi(r_p, \pi) d\pi. \tag{3.2}$$

This integration is performed over a finite line-of-sight distance as a discrete sum,

$$w_p(r_p) = 2 \sum_{i}^{\pi_{max}} \xi(r_p, \pi) \Delta\pi_i, \tag{3.3}$$

where π_i is the i^{th} bin of the line-of-sight separation, and $\Delta\pi_i$ is the corresponding bin size. We use $\pi_{max} = 80h^{-1}$Mpc and $\Delta\pi = 10h^{-1}$Mpc.

Since $w_p(r_p)$ is not affected by redshift-space distortions, the best fit power-law is equivalent to a real-space measurement. One can therefore quantify the deviation of the redshift-space $\xi(r_p, \pi)$ correlation function from the real-space behavior by measuring the ratio,

$$\Sigma(\pi) = \frac{\xi(\bar{r}_p, \pi)}{\xi(\pi)}, \tag{3.4}$$

where $\xi(\pi)$ is the best-fit power law to $w_p(r_p)$, evaluated at the π scale, and $\bar{r}_p$ indicates that we perform a spherical average in the range $0.5 \leq r_p \leq 2\ h^{-1}$Mpc. This statistic illuminates the nonlinear FoG effects by normalizing out the expected real-space clustering along the line-of-sight direction. We are interested in the behavior of pairs that potentially occupy the same halo, hence our focus at small r_p values. In the attempt to interpret the small-scale nonlinear redshift-space distortions, $\Sigma(\pi)$ is preferable to measuring the quadrupole-to-monopole ratio, $\xi_2(s)/\xi_0(s)$ [128, 129, 225], because it is a compressed representation of $\xi(r_p, \pi)$ which allows to disentangle the contribution of the distortions along the line of sight from the effects across it. In Appendix 3.9.1, we provide a comparison between the $\Sigma(\pi)$ and $\xi_2(s)/\xi_0(s)$ statistics as a function of the physical scale.

3.3.2 Correlation function estimation

For our clustering statistics, we use the estimator of [175]:

$$\xi(s) = \frac{DD(s) - 2DR(s) + RR(s)}{RR(s)} \tag{3.5}$$

where DD, DR and RR are the data-data, data-random and random-random weighted pair counts computed from a data sample of N galaxies and a random catalog of N_R points. These pair counts are normalized by the number of all possible pairs, typically by dividing by $N(N-1)/2$, NN_R and $N_R(N_R-1)/2$, respectively, and weighted by [258]

$$DD(r_p,\pi) = \sum_i \sum_j w_{tot,i} w_{tot,j} \Theta_{ij}(r_p,\pi) \tag{3.6}$$

with w_{tot} given by Eq. 3.28, and $\Theta_{ij}(r_p,\pi)$ represents a step-function which is 1 if r_p belongs to the i^{th} and π to the j^{th} bin, and 0 otherwise. These weights correct the galaxy densities to provide a more isotropic selection, therefore they should not be applied to the random catalog, which is based on an isotropic distribution. For randoms $w_{tot,i} = w_{tot,j} = 1$, therefore

$$DR(r_p,\pi) = \sum_i \sum_j w_{tot,i} \Theta_{ij}(r_p,\pi) \tag{3.7}$$

$$RR(r_p,\pi) = \sum_i \sum_j \Theta_{ij}(r_p,\pi). \tag{3.8}$$

To evaluate the correlation function, we create a random catalog that has the same selection as the BOSS CMASS galaxy data matching both the redshift distribution and sky footprint [see, e.g., 8]. The method of random catalog construction is almost identical to that described in Anderson et al. [8], but constructed to be ten times as dense as the galaxy data. We down-sample random points based on sky completeness, and "shuffle" the observed galaxy redshifts assigning them to random sky positions so as to exactly reproduce the observed redshift distribution.

3.3.3 Covariance estimation

To estimate the uncertainties in our clustering measurements, we utilize the jackknife resampling technique [245, 303, 199, 213, 214]. There are known limitations to this type of error estimation [see, e.g., 213], but they have proven sufficient in analyses on scales similar to our analysis [330, 331, 334, 121, 258, 7]. The jackknife covariance matrix for N_{res} re-samplings is computed by

$$C_{ij} = \frac{N_{res}-1}{N_{res}} \sum_{a=1}^{N_{res}} (\xi_i^a - \bar{\xi}_i)(\xi_j^a - \bar{\xi}_j), \tag{3.9}$$

where $\bar{\xi}_i$ is the mean jackknife correlation function estimate in the specific i^{th} bin,

$$\bar{\xi}_i = \sum_{a=1}^{N_{res}} \xi_i^a / N_{res}. \tag{3.10}$$

The overall factor in Eq. 3.9 takes into account the lack of independence between the N_{res} jackknife configurations: from one copy to the next, only two sub-volumes are different or, equivalently, $N_{res} - 2$ sub-volumes are the same [214].

3.3.4 The MultiDark simulation

MultiDark [240] is a N-body cosmological simulation with 2048^3 dark matter particles in a periodic box of $L_{box} = 1$ Gpc h^{-1} on a side. The first run, MDR1, was performed in 2010, with an initial redshift of $z = 65$, and a mass resolution of 8.721×10^9 h^{-1}M$_\odot$. It is based on the WMAP5 cosmology [168], with parameters: $\Omega_m = 0.27$, $\Omega_b = 0.0469$, $\Omega_\Lambda = 0.73$, $n_s = 0.95$ and $\sigma_8 = 0.82$. Here Ω is the present day contribution of each component to the matter-energy density of the Universe; n_s is the spectral index of the primordial density fluctuations, and σ_8 is the linear RMS mass fluctuation in spheres of $8h^{-1}$Mpc at $z = 0$.

MultiDark includes both the Bound Density Maxima [BDM; 162, 253], and the Friends-of-Friends [FOF; 78] halo-finders. For the current analysis, we use only BMD halos that are identified as local density maxima truncated at some spherical cut-off radius, from which unbound particles (i.e., those particles whose velocity exceeds the escape velocity) are removed. According to the overdensity limit adopted, two different BDM halo catalogs are produced: (i) BDMV – halos extend up to $\Delta_{vir} \times \rho_{back}$, where $\Delta_{vir} = 360$ is the virial overdensity threshold, $\rho_{back} = \Omega_m \times \rho_c$ is the background or average matter density, and ρ_c is the critical density of the Universe. (ii) BDMW – the maximum halo density is $\Delta_{200} \times \rho_c$, where $\Delta_{200} = 200$, which implies that BDMW halos are smaller than BDMV ones. The bound density maxima algorithm treats halos and sub-halos (those sub-structures whose virial radius lies inside a larger halo) in the same way, with no distinction. In this work we use the BDMW halo catalogs, since they resolve better the distribution of sub-structures in distinct halos, leading to a clearer small-scale clustering signal.

3.3.5 Halo Occupation Distribution model using subhalos

The halo model [reviewed in 69] is a powerful tool to understand the clustering of galaxies. The Halo Occupation Distribution [HOD; 28] is a commonly used method of mapping galaxies to dark matter halos, which characterizes the bias between galaxies and the underlying dark matter distribution. The HOD is based on the conditional probability, $P(N|M)$, that a halo with mass M contains N galaxies of a given type. In our analysis, we apply the five-parameter HOD formalism presented in Zheng et al. [337] using the MDR1 simulation at $z = 0.53$. First, we populate distinct halos with central galaxies whose mean is given by the function form of:

$$\langle N_{\rm cen}(M)\rangle = \frac{1}{2}\left[1 + erf\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{3.11}$$

where the error function is defined as the integral

$$erf(x) = 2\int_0^x e^{-t^2} dt/\sqrt{\pi}. \tag{3.12}$$

The free parameters are $M_{\rm min}$, the minimum mass scale of halos that can host a central galaxy, and $\sigma_{\log M}$, the width of the cutoff profile. At a halo mass of $M_{\rm min}$, 50% of halos host

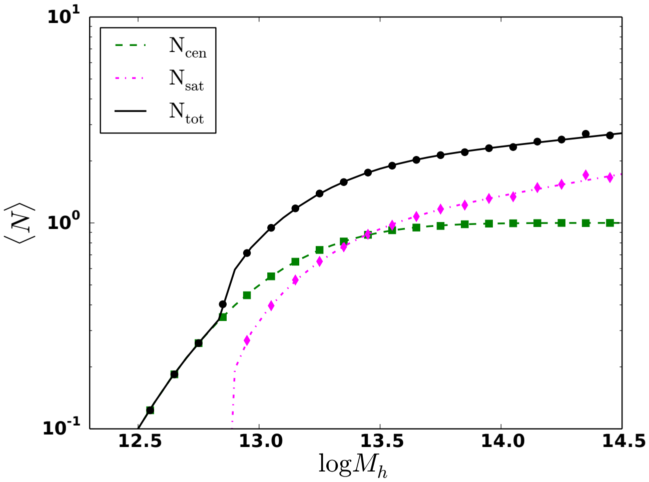

Figure 3.1: Five-parameter Halo Occupation Distribution model for MDR1, at $z = 0.53$. The parametrization is from Zheng et al. [337], and the input values from White et al. [318]. The total (solid line) population of galaxies is the sum of two contributions: central (dashed) and satellite (dot-dashed) galaxies.

a central galaxy, which in terms of probability means that $P(1) = 1 - P(0)$. If the relation between galaxy luminosity and halo mass had no scatter, $\langle N_{\rm cen}(M)\rangle$ would be modeled by a hard step function. In reality, this relation must possess some scatter, resulting in a gradual transition from $N_{\rm cen} \simeq 0$ to $N_{\rm cen} \simeq 1$. The width of this transition is $\sigma_{\log M}$. In order to place the satellite galaxies, we assume their number in halos of a given mass follows a Poisson distribution, which is consistent with theoretical predictions [28, 170, 336]. We approximate the mean number of satellite galaxies per halo with a power law truncated at a threshold mass of M_0

$$\langle N_{\rm sat}\rangle = \langle N_{\rm cen}(M)\rangle \left(\frac{M - M_0}{M_1'}\right)^{\alpha'} . \tag{3.13}$$

The parameter M_1' corresponds to the halo mass where $N_{\rm sat} \simeq 1$, when (as in our case) $M_1' > M_0$ and $M_1' > M_{\rm min}$. When $\alpha' = 1$ and $M > M_0$, the mean number of satellites per halo is proportional to the halo mass. To populate with satellite galaxies, we randomly extract from each host halo a certain number of its sub-halos, following a Poisson distribution with mean given by Eq. 3.13. The coordinates of these sub-halos become the locations for satellites. This approach, explored in previous works as [170], [318], is intrinsically different from the more commonly used procedure, in which satellites are assigned by randomly assigning the positions of dark-matter particles [see, e.g., 250]. In our case, satellite galaxies are assigned by reflecting the original halo structure made of one central halo plus none, one, or many sub-halos.

Figure 3.1 shows our HOD model built from MultiDark BDMW at $z = 0.53$, for the full CMASS sample: central galaxies are represented by the dashed curve; satellites are the dot-dashed line and the total contribution is the solid curve. As input parameters, we adopt the values consistent with the BOSS CMASS HOD modeling in White et al. [318].

3.3.6 Analytic models

Kaiser [152] demonstrated that on large scales, where the linear regime holds, the redshift-space correlation function can be factorized in terms of its real space version, $\xi(r)$, as

$$\xi(s) = \xi(r)\left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right), \tag{3.14}$$

where β is the Kaiser factor encoding the compression effect (Sec. 3.3.1) seen in the clustering signal and b is the linear bias between galaxies and the underlying matter distribution. These two quantities can be related [e.g., 226] through the following approximation:

$$\beta \approx \Omega_m^{0.6}/b. \tag{3.15}$$

In general, one can decompose the redshift-space separation s into its parallel and transverse components to the line-of-sight and approximate $\xi(r)$ with the power law in Eq. 3.1 to produce [194]:

$$\xi(r_p, \pi) = \xi(r)\left\{1 + \frac{2(1-\gamma\mu^2)}{3-\gamma}\beta + \frac{3 - 6\gamma\mu^2 + \gamma(2+\gamma)\mu^4}{(3-\gamma)(5-\gamma)}\beta^2\right\}. \tag{3.16}$$

Here γ is the power law spectral index and μ is the cosine of the angle between the separation and the line-of-sight direction. We include the small-scale nonlinear FoG by convolving with a pairwise velocity distribution [103, 129, 74], which can be modeled as an exponential,

$$f_{exp}(w) = \frac{1}{\sqrt{2}\alpha} \exp\left(-\sqrt{2}\frac{|w|}{\alpha}\right), \tag{3.17}$$

or a Gaussian form,

$$f_{norm}(w) = \frac{1}{\sqrt{2\pi}\alpha} \exp\left(-\frac{w^2}{2\alpha^2}\right), \tag{3.18}$$

where α is the pairwise velocity dispersion. The full model then becomes

$$\xi(r_p, \pi) = \int_{-\infty}^{+\infty} \xi(r_p, r_z(w)) f(w) dw, \tag{3.19}$$

with $\xi(r_p, r_z(w))$ given by Equation 3.16. The quantity $r_z(w) \equiv (\pi - w)/(aH(z))$ is the line-of-sight component of the real-space distance r, $a = (1+z)^{-1}$ is the scale factor, and $H(z)$ is the Hubble parameter evaluated at redshift z. The full $\Sigma(\pi)$ analytic model, as a function of α and β, is obtained by averaging Eq. 3.19 in the range $0.5 \leq r_p \leq 2$ h^{-1}Mpc and integrating the result in π bins, as explained in Section 3.3.1.

Combining these definitions and matching the binning in Δr_p and $\Delta\pi$, we have:

$$\Sigma(\pi) = \frac{\int \frac{dZ}{\Delta\pi} \int \frac{dR}{\Delta r_p} \int \xi\left(R, \frac{Z-w}{aH(z)}\right) f(w) dw}{\int \frac{dZ}{\Delta\pi} \int \frac{dR}{\Delta r_p} \left(\frac{r_0^2}{R^2+Z^2}\right)^{\gamma/2}} \tag{3.20}$$

Finally, we rename the parameters α and β respectively A and G to emphasize they are fitted parameters that might differ slightly from their theoretically motivated meaning. In this formalism, Eq. 3.15 simply becomes

$$G \approx \Omega_m^{0.6}/b. \tag{3.21}$$

The FoG and Kaiser effects could be overlapping and, as fit parameters in a model, they are

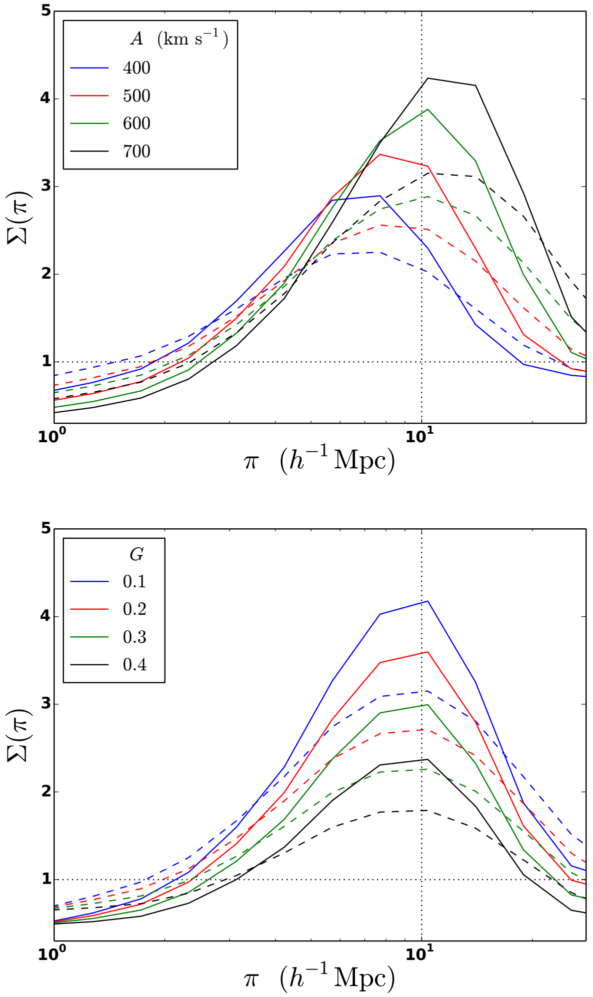

Figure 3.2: $\Sigma(\pi)$ analytic model as a function of the pairwise velocity dispersion, A, (top panel) and the parameter G, encoding the Kaiser factor (bottom panel). Solid lines represent the Gaussian model given in Eq. 3.18; dashed curves are the exponential functions in Eq. 3.17. We choose to model our $\Sigma(\pi)$ measurements using the normal functional form only, since it reproduces more accurately the small-scale feature provoqued by the FoG distortions and peak at larger scales.

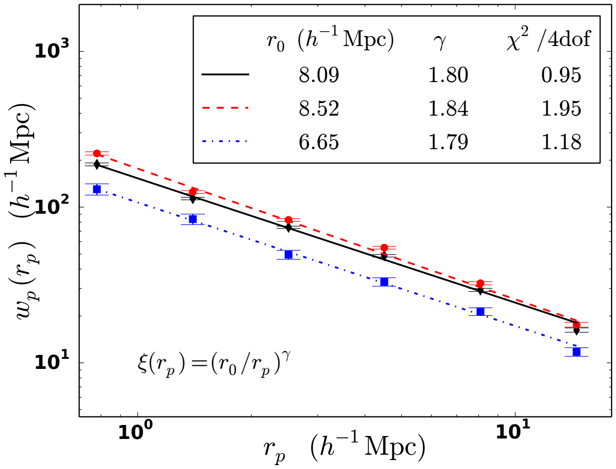

Figure 3.3: Power-law fits to the CMASS full, red and blue projected correlation functions, which define the denominator in Eq. 3.20. The r_0 and γ values we find are consistent with Zehavi et al. [331], and show that red galaxies cluster more than blue star-forming ones. The error bars correspond the 1σ uncertainties estimated using 200 jackknife resamplings (Sec. 3.3.3).

correlated. The importance of our modeling is not to isolate their value, but to differentiate between models and data with sub-populations of galaxies. Figure 3.2 shows how both effect contribute to modulate our $\Sigma(\pi)$ model. There is a degeneracy between the parameter values, in the sense that both increasing A or reducing G produces an enhancement in the $\Sigma(\pi)$ peak. This dependence prevents us from interpreting the G parameter as the only one responsible of the $\Sigma(\pi)$ amplitude.

3.3.7 Fitting $w_p(r_p)$

To implement the integral in Eq. 3.2, to estimate the projected correlation function $w_p(r_p)$, we need to truncate it at some upper value, π_{max}, above which the contribution to correlation function becomes negligible. If one includes very large scales, the measurement will be affected by noise; inversely, if we consider only very small scales, the clustering amplitude will be underestimated. In our case, CMASS results are not sensitive to $\pi \geq 80$ h^{-1}Mpc, therefore we adopt this value as our π_{max} limit. The projected auto-correlation function is related to the real-space one by [79]

$$w_p(r_p) = 2 \int_{r_p}^{\pi_{max}} \frac{r\xi(r)}{\sqrt{r^2 - r_p^2}} dr. \tag{3.22}$$

Zehavi et al. [333] demonstrates that for a generic power law, $\xi(r) = (r/r_0)^\gamma$, the equation above can be written in terms of the Euler's Gamma function as

$$w_p(r_p) = r_p \left(\frac{r_p}{r_0}\right)^\gamma \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{\gamma - 1}{2}\right) / \Gamma\left(\frac{\gamma}{2}\right). \tag{3.23}$$

allowing one to infer the best-fit power law for $\xi(r)$ from $w_p(r_p)$, corresponding to the full CMASS galaxy sample, blue and red sub-samples. Figure 3.3 presents the power-law fits to the full, red and blue CMASS projected correlation functions, and the resulting (r_0, γ) optimal values.

3.4. BOSS CMASS data

BOSS target galaxies primarily lie within two main samples: CMASS, with $0.43 < z < 0.7$ and LOWZ, with $z < 0.43$ [258, 7, 42]. These samples are selected on the basis of photometric observations done with the dedicated 2.5-m Sloan Telescope [120], located at Apache Point Observatory in New Mexico, using a drift-scanning mosaic CCD camera with five color-bands, *ugriz* [119, 107]. Spectra of the LOWZ and CMASS samples are obtained using the double-armed BOSS spectrographs, which are significantly upgraded from those used by SDSS-I/II, covering the wavelength range 3600 − 10000Å with a resolving power of 1500 to 2600 [275]. Spectroscopic redshifts are then measured using the minimum-χ^2 template-fitting procedure described in [5], with templates and methods updated for BOSS data as described in [42].

We select galaxies from CMASS DR11 [6] – North plus South Galactic caps – which is defined by a series of color cuts designed to obtain a galaxy sample with approximately constant stellar mass. Specifically, these cuts are:

$$\begin{aligned}
&17.5 < i_{cmod} < 19.9, \\
&r_{mod} - i_{mod} < 2, \\
&d_{\perp} > 0.55, \\
&i_{fib2} < 21.5, \\
&i_{cmod} < 19.86 + 1.6(d_{\perp} - 0.8),
\end{aligned} \tag{3.24}$$

where i_{cmod} is the $i-$band cmodel magnitude. The quantities i_{mod} and r_{mod} are model magnitudes, i_{fib2} is the $i-$band magnitude within a 2" aperture and $d_{\perp}$ is defined as

$$d_{\perp} = r_{mod} - i_{mod} - (g_{mod} - r_{mod})/8.0. \tag{3.25}$$

All the magnitudes are corrected for Galactic extinction using the dust maps from [265]. In addition to the above color cuts, CMASS objects must also pass two star-galaxy separation constraints:

$$\begin{aligned}
&i_{psf} - i_{mod} > 0.2 + 0.2(20.0 - i_{mod}) \\
&z_{psf} - z_{mod} > 9.125 - 0.46 z_{mod},
\end{aligned} \tag{3.26}$$

unless the objects also pass the LOWZ criteria. Therefore, to distinguish CMASS from LOWZ candidates, it is necessary to select them by redshift.

3.4.1 Color selection

The CMASS sample is mainly composed of massive, luminous, red galaxies, which are favorite subjects to study galaxy clustering. Among them, however, there is an intrinsic bluer, star-forming population of massive galaxies [258, 121], of which little is known. In the attempt to explore this bluer component to understand its contribution in the clustering properties,

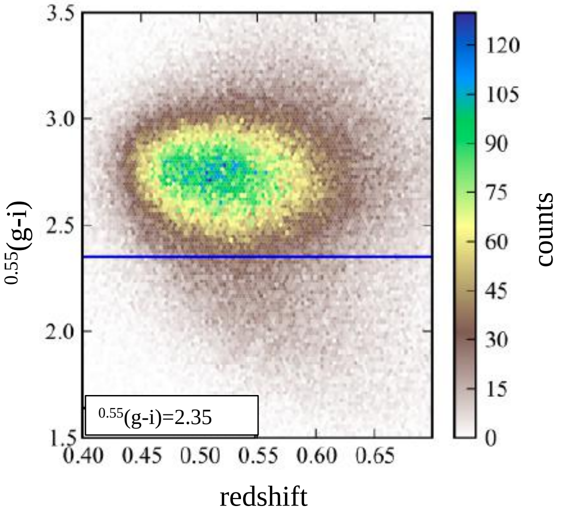

Figure 3.4: BOSS CMASS DR11 color selection: the $(g - i)$ color cut divides the full sample into a red dense population (above the blue horizontal line) and a sparse blue tail (below the line).

we split the CMASS sample into its blue and red components by applying the color cut

$$^{0.55}(g - i) = 2.35 \tag{3.27}$$

constant in redshift and K-corrected to the $z = 0.55$ rest-frame using the code by [38]. [193] applied this same color cut, with no K-corrections, to the BOSS CMASS DR8 sample to study the morphology of the LRG population; [259] used a similar selection, $^{0.55}(r-i) = 0.95$, to measure galaxy clustering at the BAO scale in CMASS DR10. Figure 3.4 presents our CMASS color selection, splitting the full sample into a red denser population (above the blue horizontal line) and a sparse blue tail (below the line), whose completeness dramatically increases when we move towards high redshift values ($z > 0.55$). For our analysis, we focus on the high-redshift tail of the CMASS sample, selecting only galaxies with redshift beyond $z > 0.55$.

3.4.2 Weights

Due to its structural features, a survey inevitably introduces some kind of spatial variation in its measurements. To avoid these distortions, we weight our pair counts by defining a linear combination of four different weights [7, 261, 258]:

$$w_{tot} = w_{FKP}\, w_{sys}(w_{fc} + w_{zf} - 1), \tag{3.28}$$

each one correcting for a different effect. In the expression above, w_{zf} accounts for targets with missing or corrupted redshift (z failure); w_{fc} corrects for fiber collision, compensating the fact that fibers cannot be placed closer than 62" on the survey plates. This limitation prevents obtaining spectra of all galaxies with neighbors closer than this angular distance in a single observation. The default value of w_{zf} and w_{fc} is set to unity for all galaxies. When a fiber collision is detected, we increment by one the value of w_{fc} for the first neighbor closer than 62". In the same way, for the nighbor we increase by one the value of w_{zf} of the nearest galaxy with a good redshift. To minimize the error in the measured clustering signal, we also require a correction based on the redshift distribution of our sample, namely the w_{FKP} factor [101], that weights galaxies according to their number density, $n(z)$. It is defined as

$$w_{FKP} = \frac{1}{1 + n(z)P_{FKP}}, \tag{3.29}$$

where P_{FKP} is a constant that roughly corresponds to the amplitude of the CMASS power spectrum $P(k)$, at $k = 0.1\,h$ Mpc^{-1}. We assume $P_{FKP} = 2 \times 10^4\,h^3$ Mpc^{-3}, in [7]. The last weight, w_{sys}, accounts for a number of further systematic effects that could cause spurious angular fluctuations in the galaxy target density. These effects are treated in detail in [258], but we do not include them in this analysis, since they are not relevant at the scales considered in this work. Therefore we set in $w_{sys} = 1$ in the following analysis.

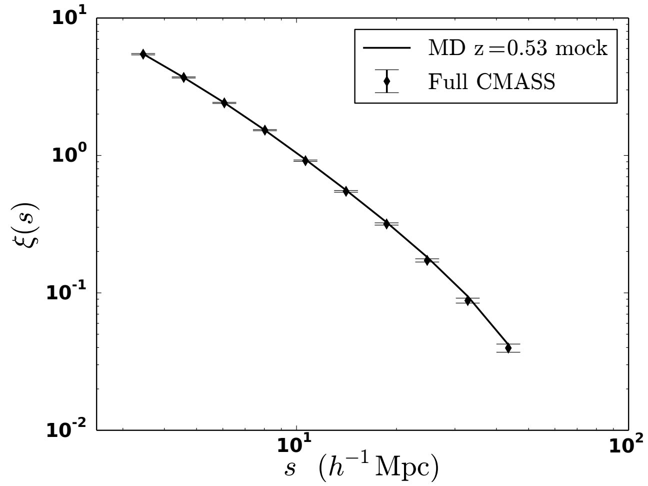

Figure 3.5: Redshift-space monopole correlation functions of our $z = 0.53$ MultiDark full mock galaxy catalog (solid line) compared to BOSS CMASS DR11 measurements (diamonds). Error bars are estimated using 200 jackknife regions.

3.5. Modeling the full CMASS sample

3.5.1 Full CMASS clustering

We construct an HOD model using MultiDark halos and sub-halos (see model description in Section 3.3.5), and produce a mock galaxy catalog which we compare to the full CMASS DR11 population. This mock is built by varying the HOD parameters to match $\xi(s)$, populating the MD simulation in each step, and using the peculiar velocities in the simulation to model redshift-space distortions. The intention is that changing the HOD will constrain the overall galaxy bias, hence we fit only one statistic. We then evaluate and further investigate these fits over the three clustering metrics: $\xi(s)$, $w_p(r_p)$ and $\Sigma(\pi)$.

However, since implementing a formal fit to determine the optimal HOD parameters is beyond the scope of this work, we improve the matching empirically, changing the input values until we find a suitable ($logM_{min}$, M_0, M_1', α', σ_{logM}) set that reproduces the observed $\xi(s)$ amplitude. We fit only M_{min} (the minimum halo mass), M_1' (the mass scale of the satellite cut-off profile) and α (the satellite slope). The remaining parameters are fixed to their default values given by White et al. [318]: $logM_0 = 12.8633$, $\sigma_{logM} = 0.5528$.

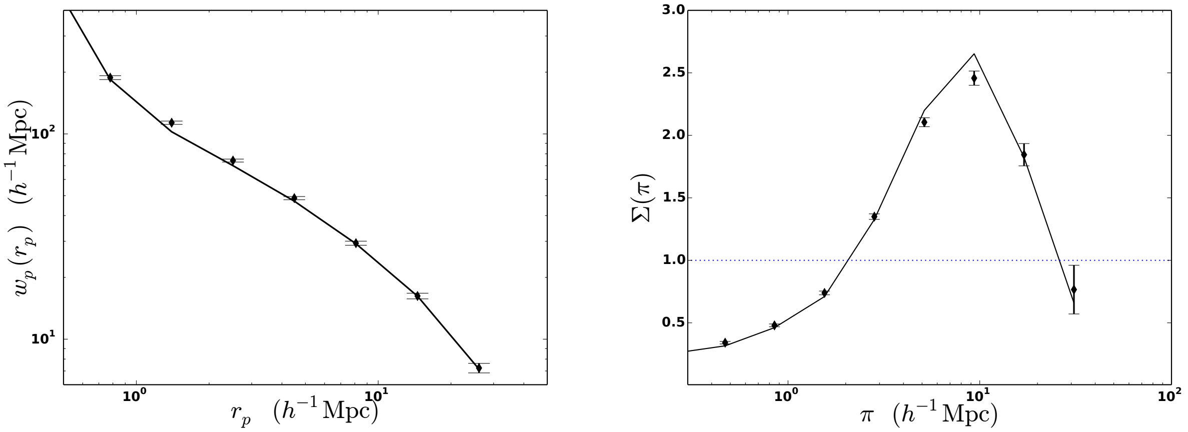

Figure 3.6: Projected correlation function (left) and $\Sigma(\pi)$ (right) for the $z = 0.53$ MultiDark full mock galaxy catalog (solid line), compared to BOSS CMASS DR11 measurements (diamonds). Error bars are estimated using 200 jackknife regions containing the same number of randoms.

The specific choice of these three parameters arises from their connection to two physical quantities we want to measure: (i) the satellite fraction, f_{sat}, that controls the slope of the 1-halo term at small scales, where sub-structures of the same halo dominate; (ii) the galaxy number density, $n(z)$, affecting the 2-halo term at larger scales, where correlations between sub-structures of different hosts become appreciable. Figure 3.20 in the Appendix illustrates how a change in M_{min}, M_1' and α affects the projected correlation function.

Figure 3.5 displays the redshift-space monopole corresponding to our empirical best fit ($\chi^2 = 11.08/7\,dof$ including the full covariance matrix computed with jackknife; the HOD parameters are given in Table 3.1) mock galaxy catalog from the MultiDark simulation. The projected correlation function, $w_p(r_p)$, and the line-of-sight statistic, $\Sigma(\pi)$, corresponding to this model are shown in Figure 3.6. In agreement with many previous works [332, 333, 121], we find that CMASS galaxies are more highly clustered at small scales (1-halo regime); then, as the spatial separation between the pairs increases, the clustering strength drops (2-halo term). Compared to White et al. [318], our best-fit mock has a much lower satellite slope, α, and M_1', resulting in a higher satellite fraction (about 27%); however, our mean satellite occupation function is compatible with results from Guo et al. [122]. Overall, the amplitude of our model galaxies is in good agreement with observations. Error bars are estimated using

200 jackknife regions gridded in right ascension and declination as follows: 10 RA×15 DEC cells for the CMASS North Galactic Cap ($N_{res} = 150$), plus 5 RA×10 DEC regions for the South Galactic Cap, ($N_{res} = 50$). This approach produces 200 equal areas of about 100 $\deg^2$ each.

To compute $\Sigma(\pi)$ through Eq. 3.4 for the full CMASS galaxy sample and MD mock, we use the best-fit power-law to their projected correlation functions, $w_p(r_p)$. The relative r_0 and γ estimates are shown in Figure 3.3. Beyond $8-10h^{-1}$Mpc, where the Kaiser squashing becomes predominant, the jackknife uncertainties on $\Sigma(\pi)$ are wider. This measurement reveals that the deviation of $\xi(\bar{r}_p, \pi)$ from the real-space behavior dramatically changes according to the scale of the problem: at very small redshift separations i.e., $\pi \leq 2h^{-1}$Mpc, where the finger-of-god dominate, the contribution of peculiar velocities pushes $\Sigma(\pi)$ below unity. Above $3h^{-1}$Mpc, $\Sigma(\pi)$ increases sharply and peaks around $8h^{-1}$Mpc. On larger scales, the correlation between pairs of galaxies is compressed along the line of sight since the Kaiser infall dominates and $\Sigma(\pi)$ drops. The $\Sigma(\pi)$ measurement shows very different and characteristic features according to the scale of interest, therefore it is a valuable tool to quantify both small and large-scale clustering effects.

3.5.2 Modeling redshift-space distortions and galaxy bias

In redshift-space, two different distortion features are observed: the finger-of-god effect which dominates below $2h^{-1}$Mpc, and the Kaiser flattening, which becomes important beyond $10 - 15h^{-1}$Mpc. These phenomena preferentially manifest themselves on different scales, but a certain degree of entanglement is unavoidable in both regimes. In order to better separate the two effects, we examine $\Sigma(\pi)$ in our MultiDark full mock catalog in three different configurations: real-space, redshift-space with only Kaiser effect and full redshift-space (FoG+Kaiser), as shown in Figure 3.7. The real-space $\Sigma(\pi)$ is defined in Eq. 3.4, omitting the peculiar velocities both in the numerator and in $w_p(r_p)$ to which we fit the power law at the denominator. Since $\Sigma(\pi)$ is the ratio between two spherically averaged

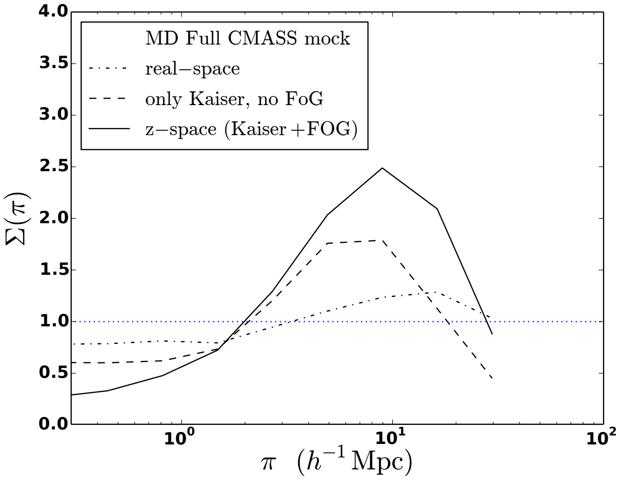

Figure 3.7: $\Sigma(\pi)$ in real-space (dot-dashed line), redshift-space with only Kaiser contribution (dashed) and Kaiser plus finger-of-god (solid). As expected, the real-space behavior is close to unity at all scales.

power laws, we expect it to be close to unity at all scales. Hence, the dot-dashed line in Figure 3.7 is compatible with expectations. The redshift-space case with only Kaiser contribution (dashed line) is computed by assigning satellite galaxies their parental v_{pec} value. In this way, each satellite shares the coherent motion of its parent, but it does not show any random motion with respect to it. The last case considered is the full redshift-space $\Sigma(\pi)$ (solid line), in which satellite galaxies have their own peculiar velocity, which is independent from their parents.

We are now able to provide a full description of our $\Sigma(\pi)$ results by modeling them through Eq. 3.20, in terms of four parameters: the power-law correlation length, r_0, its slope γ, the pairwise velocity dispersion, A and the G parameter, which is inversely proportional to the linear galaxy bias, b, through Eq. 3.21.

The linear galaxy bias is scale dependent and has been computed [e.g., 215] as the ratio between the galaxy and matter correlation functions,

$$b(s) = \sqrt{\frac{\xi(s)}{\xi_m(s)}}. \tag{3.30}$$

Our goal is to provide an estimate of both the peculiar velocity field causing the distortions

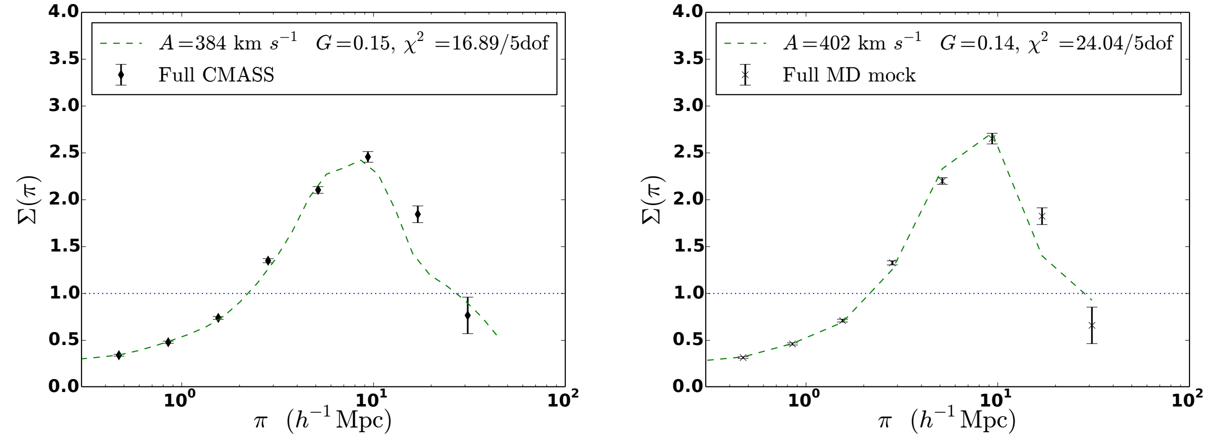

Figure 3.8: $\Sigma(\pi)$ full CMASS DR11 measurement (left panel, diamonds) and our MultiDark $z = 0.53$ mock (right panel, crosses), versus their A, G analytic model (dashed lines). For both data and mocks we assume the errors are given by our jackknife estimate, computed using 200 resamplings. The fits are performed by using the full covariance matrix. These plots reveal that the full CMASS sample and the MultiDark model galaxies share almost the same large-scale bias value, while the peculiar velocity contribution is higher in the mocks.

we observe in redshift-space in our clustering measurements and the large scale bias, using the A, G values we find from our full, red and blue CMASS and MultiDark $\Sigma(\pi)$ modeling. To this purpose, we do not compute the bias as [215], through Eq. 3.30, but we estimate it from Eq. 3.21.

Figure 3.8 displays the A, G models (dashed curves) for our full CMASS $\Sigma(\pi)$ measurement (left panel, diamonds) and full MultiDark mock catalog (right panel, crosses). All the model fits are performed including the full covariance matrix, estimated by using 200 jackknife resamplings (Sec. 3.5.1). For the MultiDark mock, we assume the same scatter of the CMASS data. Adopting a normal function (Eq. 3.18) to mimic the contribution of peculiar velocities, we find that MD model galaxies have slightly higher bias, meaning a lower G value, compared to the full CMASS population and higher peculiar velocity contribution, i.e. higher A value (see Table 3.2). This result is in agreement with the right panel in Figure 3.6: the full CMASS $\Sigma(\pi)$ observations (diamonds) experience a stronger Kaiser squashing at $\sim 10\,h^{-1}\,\mathrm{Mpc}$, i.e. they have a smaller large-scale bias compared to the MultiDark model galaxies (solid line). From these A, G values, we conclude that our full MD mock catalog can be considered a reliable representation of the full CMASS sample.

The reduced χ^2 values we derive from the full CMASS and MultiDark $\Sigma(\pi)$ model fits are less stringent compared to the estimates for $\xi(s)$ reported in the caption of Table 3.1. The main reason for this result resides in how we build the A, G model (see Eq. 3.20) to reproduce the $\Sigma(\pi)$ feature, which is the ratio of a 2PCF, spherically averaged in the range $0.5 \leq r_p < 2\,h^{-1}$Mpc, over a real-space term. To mimic this average in our model, we convolve (numerator in Eq. 3.20) a real-space correlation function with a peculiar velocity term, $f(w)$, and integrate the result to eliminate the dependence on r_p. Such an integration is performed numerically in (r_p, π) bins, by stacking r_p into a single average value per bin. The denominator in Eq. 3.20 is a real-space term, given by best-fit power law to $w_p(r_p)$, spherically averaged in the same way as the numerator. Thus, the A, G model reproduces the $\Sigma(\pi)$ measurement numerically in bins of (r_p,π) and not analytically in each point. The approximations adopted to define our $\Sigma(\pi)$ model are justified by the fact that our goal is to provide a qualitative prediction of the linear bias and redshift-space distortions in the full, red and blue CMASS samples. For this reason, we do not heavily focus on the goodness of our model fits, but instead stress the importance of a cross-comparison in terms of A, G values.

From the full CMASS model, we find a bias of $b \sim 3$, which is relatively high compared to the estimate reported in [215], $b \sim 2$. This discrepancy is due to the fact that we select only the massive bright high-redshift tail (i.e. $z > 0.55$) of the CMASS sample; for these specific galaxies the bias is expected to be higher than in [215].

3.5.3 Full CMASS covariance

We compute the full CMASS jackknife covariance matrix for the three metrics of interest using Eq. 3.9, in which ξ is either $\xi(s)$, $w_p(r_p)$, or $\Sigma(\pi)$. We estimate the goodness of our model fits to the CMASS measurements by computing the relative χ^2 values as

$$\chi^2 = A^T C_\star^{-1} A, \tag{3.31}$$

	Total	Red	Blue
$logM_{min}$	13.00	13.10	12.50
$logM_1'$	13.30	13.02	13.85
α	0.20	0.22	0.15
f_{sat}	0.27	0.33	0.11
$\langle logM_h \rangle$	12.75	13.00	12.50

Table 3.1: Our best empirical estimates of the HOD parameters for the total, red and blue independent models of the CMASS populations. We obtain these values only by fitting $\xi(s)$ with a three-dimensional grid in $logM_{min}$, $logM_1'$ and α. The resulting χ^2 values are: 11.08/7dof (full CMASS), 13.54/7dof (red) and 14.91/7dof (blue).

where $A = (\xi^i_{data} - \xi^i_{model})$ is a vector with $i = 1, ..., n_b$ components and $C_\star^{-1}$ is an unbiased estimate of the inverse covariance matrix [131, 230],

$$C_\star^{-1} = (1 - D)C^{-1}, \;\; D = \frac{n_b + 1}{N_{res} - 1}. \tag{3.32}$$

In the equation above, n_b is the number of observations and N_{res} the number of jackknife re-samplings. For the full CMASS population, the correction factor $(1 - D)$ represents a 8% effect on the final χ^2 value.

In Appendix 3.9.4, we test our jackknife error estimates using a set of 100 Quick Particle Mesh [QPM; 319] galaxy mock catalogs.

3.6. Modeling color sub-samples

We repeat the same analysis described in Section 3.5 on the red and blue color sub-samples. We first use $\xi(s)$ to fit an HOD and match the overall clustering, then use our analytic model to obtain fits for A and G. There remains a question on how to model the sub-populations in the mocks; we explore two methods.

3.6.1 Independent Red and Blue models

For simplicity, our first attempt at the color sub-samples is to individually model the red and the blue CMASS populations. That is, we assume the clustering comes from a complete sample and we generate an HOD populating halos independently of whether a galaxy is red

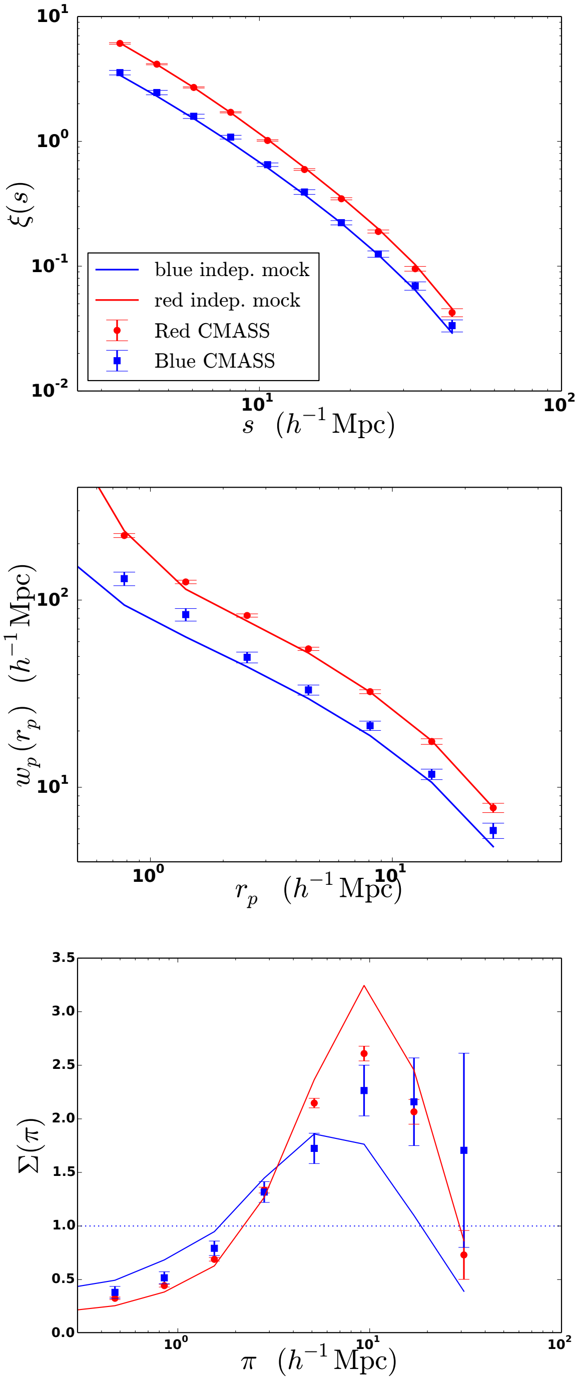

Figure 3.9: Independent mock catalogs designed to model the CMASS DR11 red and blue $\xi(s)$, $w_p(r_p)$ and $\Sigma(\pi)$ measurements (respectively indicated by points and squares). The error bars are the 1σ regions estimated using 200 jackknife re-samplings of the data. Despite we fit only $\xi(s)$, we find good agreement between data and mocks in all our three statistics. As expected, red galaxies show a higher clustering amplitude compared to the blue population.

or blue. By definition, there is no connection in the overlap and the same halo or sub-halo could host either a red and blue galaxy in the corresponding mocks. This is an over-simplified view, as clearly a galaxy can be either red or blue and not both. However, it is an assumption that is embedded within several related analyses [332, 333, 121, 123].

Figure 3.9 shows the agreement between the red and blue (points and squares) CMASS monopole, projected 2PCF and $\Sigma(\pi)$ measurements and our independent red and blue model galaxies (lines). Our empirical best-fit HOD parameter values are reported in Table 3.1, together with the satellite fraction; the fraction is higher for red than for blue galaxies, confirming that luminous red galaxies tend to live in a denser environment [311, 333, 297]. We conclude that we are able to fit correctly all our red and blue CMASS clustering results, by means of the same HOD technique, with small variations in its input parameters. However, these red and blue independent models are non-physical, because they allow the same galaxy to be either red or blue. In other words, they place both red and blue galaxies in the same hosting halos, which is not the case.

To overcome this problem, we propose an alternative halo occupation distribution approach (see next Section) in which the red and the blue models are obtained by splitting the full mock catalog into sub-populations that match the observed red/blue CMASS galaxy fractions. In this way, the red and blue model galaxies are no longer independent and, by construction, they cannot occupy the same positions in a given halo.

3.6.2 Splitting color samples using galaxy fractions

Inspired by the result in the previous section, we develop a more physically motivated model of red/blue color separation. In line with the standard halo model, we explore a splitting method based entirely on host halo mass, with each of them matching the corresponding observed CMASS galaxy fraction. By modeling these red/blue fractions, $f_{b,r}$, as a function of the central halo mass, we are able to correlate the red and the blue mock catalogs to the full one, reducing the number of free parameters from 15 (5 for each independent HOD)

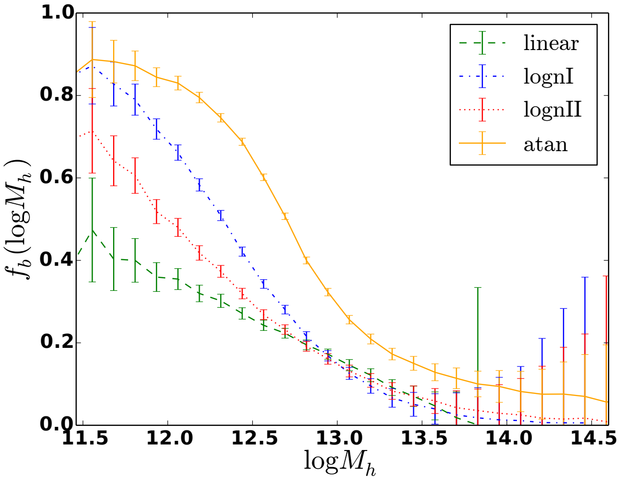

Figure 3.10: Blue galaxy fraction models, f_b, and the corresponding Poisson error, as a function of the central halo mass: linear (dashed line), log-normal I (dot-dashed), log-normal II (dotted), inverse tangent (solid). The red galaxy fractions are recovered by $f_r = 1 - f_b$.

to 5 (full HOD) plus 2 (constraint on galaxy fractions). Our galaxy fraction model must verify two conditions: (i) to obtain reliable results, the models must reproduce the overall $f_{b,r}$ values observed in our CMASS red/blue selection; this is done by requiring that

$$\begin{gathered}\Sigma_{i=1}^{N} f_b(\log M_h(i))/N = 0.25, \\ f_r(\log M_h) = 1 - f_b(\log M_h) = 0.75\end{gathered} \tag{3.33}$$

where we allow 20% of scatter, and (ii) the red (blue) fraction must approach zero at low (high) mass scales. We build our theory as a function of the central halo mass only, omitting the dependence on satellite masses. Despite this simplifying assumption, the resulting red and blue mocks match correctly the observed clustering amplitude. To mimic the red/blue split, we test different functional forms of $f_{b,r}$, starting with a basic linear one (Figure 3.10, dashed line) and two different log-normal models (dot-dashed and dotted curves) with three degrees of freedom each; they are treated in detail in Appendix 3.9.3. In order to produce a clear separation between the two populations, the best compromise is an inverse tangent-like function (solid line), with only two free parameters. The resulting functional form, as a function of the central halo mass, is

$$\begin{aligned} f_b(\log M_h) &= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left[\frac{\log M_h - D}{10^C}\right], \\ f_r(\log M_h) &= 1 - f_b(\log M_h) \end{aligned} \tag{3.34}$$

where the parameter C determines how rapidly the blue fraction drops and D establishes the half-width of the curve. Applying Eqs. 3.33 and 3.34 to the full CMASS mock catalog, we select the (C, D) combination that best fits the observed red and blue redshift-space auto-correlation functions, $\xi(s)$. The best-fit values are $C = -0.50$, $D = 12.50$, with $\chi^2_{red} = 15.43/5dof$, $\chi^2_{blue} = 6.20/5dof$ and $\chi^2_{tot} = 10.82/10dof$. We use these red and blue inverse tangent mocks to match the other two statistics, $w_p(r_p)$ and $\Sigma(\pi)$, which are shown in Figure 3.11 and the cross-correlation functions in Fig. 3.12. The $\xi(s)$ fit is performed using the full covariance matrix and the uncertainties are estimated via jackknife resampling (Sec. 3.3.3). The cross-correlations between red and blue CMASS galaxies behave similarly to the auto-correlation functions: they are stronger on small scales and weaker when the pair separation increases. These functions represent a consistency check of our red/blue fitting scheme and they provide robust information about red and blue galaxy bias: the younger and more star-forming is the galaxy, the lower are its clustering amplitude and bias.

Figure 3.13 displays the red and blue HOD models inferred by splitting the full MultiDark mock using the observed CMASS red/blue galaxy fraction. The lines are the predictions computed normalizing $\langle N_c \rangle$, $\langle N_s \rangle$, $\langle N_t \rangle$ by $f_{b,r}$. For red galaxies the HOD shape is compatible with the model shown in Figure 3.1, confirming that the red/blue separation we imposed with the galaxy fraction constraint is reliable for the red population. For blue mocks, the average number of galaxies per halo mass is ~ 10 times less compared to the red $\langle N_{cen} \rangle$, at $M_h = 10^{13.5}\, h^{-1}\mathrm{M}_\odot$ and drops almost linearily (3% factor) as the halo mass increases. Such a trend reflects the preference of blue star-forming galaxies to populate low-mass halos.

From this analysis, we estimate the conditional probability, $P(M_h|G)$, that a galaxy G

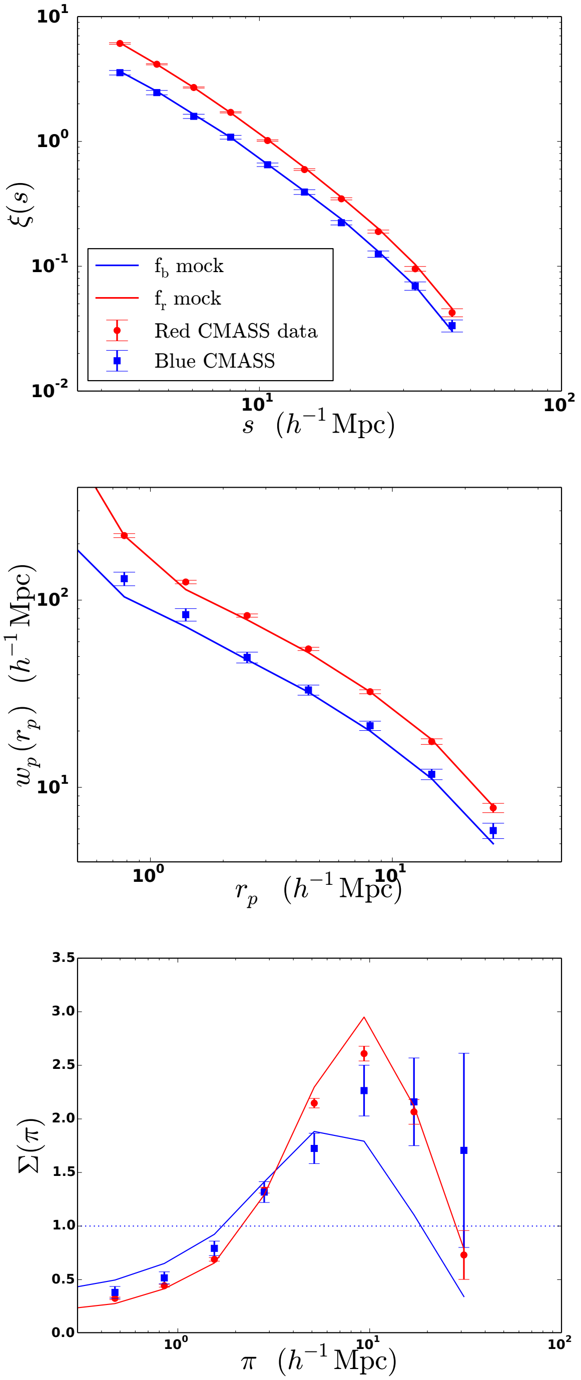

Figure 3.11: CMASS DR11 red and blue clustering measurements (points and squares) versus mocks (lines). The models are obtained by splitting the full MultiDark mock into its red and blue components, matching the observed CMASS red/blue galaxy fraction, $f_{b,r}$. In this way, we prevent the same mock galaxy to be either red or blue, and guarantee the reliability of the model. We find good agreement between the CMASS measurements and our MultiDark mocks, and confirm that red galaxies leave in more dense environments compared to the blue population.

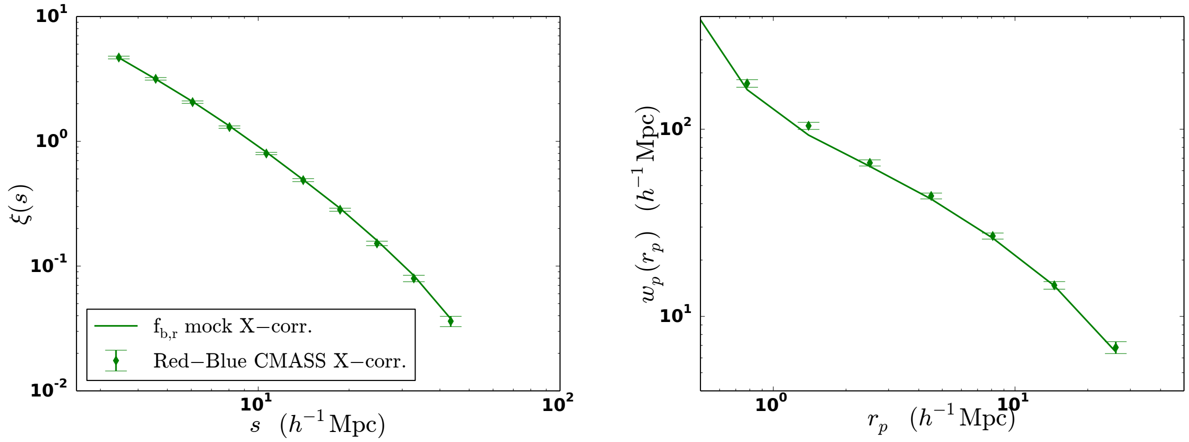

Figure 3.12: Red-blue CMASS DR11 (diamonds) versus inverse tangent mock (lines) cross-correlation functions. These plots are useful to check the mutual behavior of the the red and the blue CMASS samples. In fact, as expected, we find that the cross-correlation of these galaxies lies in between their auto-correlation functions, and the size of the errorbars (computed with 200 jackknife resamplings) is consistent with the uncertainties on their individual clustering measurements.

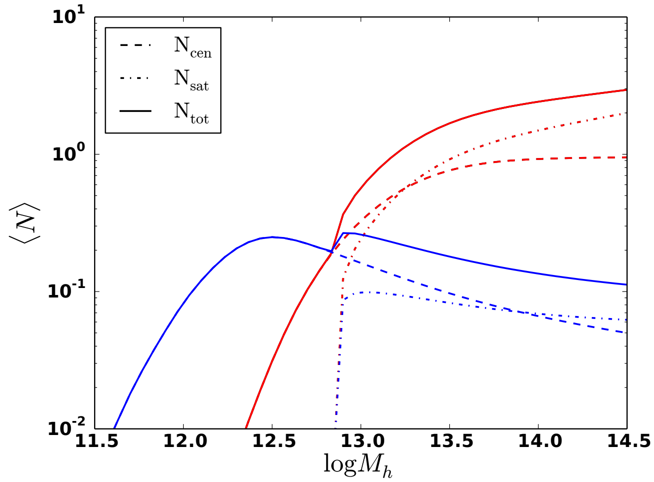

Figure 3.13: Red and blue HOD models obtained by applying the galaxy red/blue fraction condition to the MultiDark mock catalog for the full CMASS population. The lines are the predictions computed by normalizing $\langle N_c \rangle$, $\langle N_s \rangle$, $\langle N_t \rangle$ by $f_{b,r}$. For red galaxies, the HOD shape is consistent with Figure 3.1, confirming that the red/blue galaxy separation we are imposing with the satellite fraction constraint is reliable for the red population. For blue mocks, the expected average number of galaxies per halo mass is about 10 times less than for red ones at $logM_h = 13.5$, and drops almost linearly as the halo mass increases. This reveals that blue star-forming galaxies preferentially populate low-mass halos.

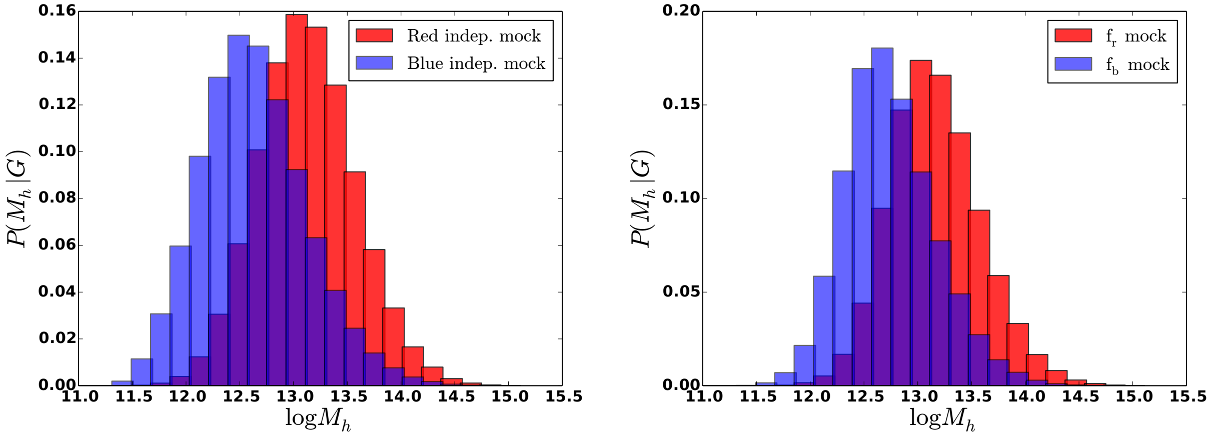

Figure 3.14: Conditional probability that a given galaxy G with a specific color is hosted by a central halo with mass M_h obtained from our red and blue independent mock catalogs (left) and applying the galaxy fraction constraint (right). In both cases, as expected, we find that red galaxies live in more massive halos compared to the blue ones.

with a specific color is hosted by a central halo having mass M_h; see Figure 3.14. As expected, the result demonstrates that CMASS early-type redder galaxies are associated to more massive halos ($M_h \sim 10^{13.1}\,h^{-1}\mathrm{M}_\odot$), compared to the late-type bluer ($M_h \sim 10^{12.7}\,h^{-1}\mathrm{M}_\odot$) companions.

3.7. Results

3.7.1 Red and Blue A, G models

We apply the same A, G modeling performed in Section 3.5.2 on the full CMASS sample and the MultiDark full mock galaxy catalog to the red and blue data samples and $f_{b,r}$ mocks, to quantify how significant their differences are at the level of large-scale bias and redshift-space distortions. Our results are presented in Figure 3.15: the top row displays the red and blue $\Sigma(\pi)$ CMASS measurements (points and squares), versus the analytic models (dashed lines); in the bottom row are the results for the red and blue MD mocks (crosses), versus their models (dashed curves). For both CMASS data and MD mocks we assume the errors are given by our jackknife estimate, done using 200 resamplings. All the model fits are fully

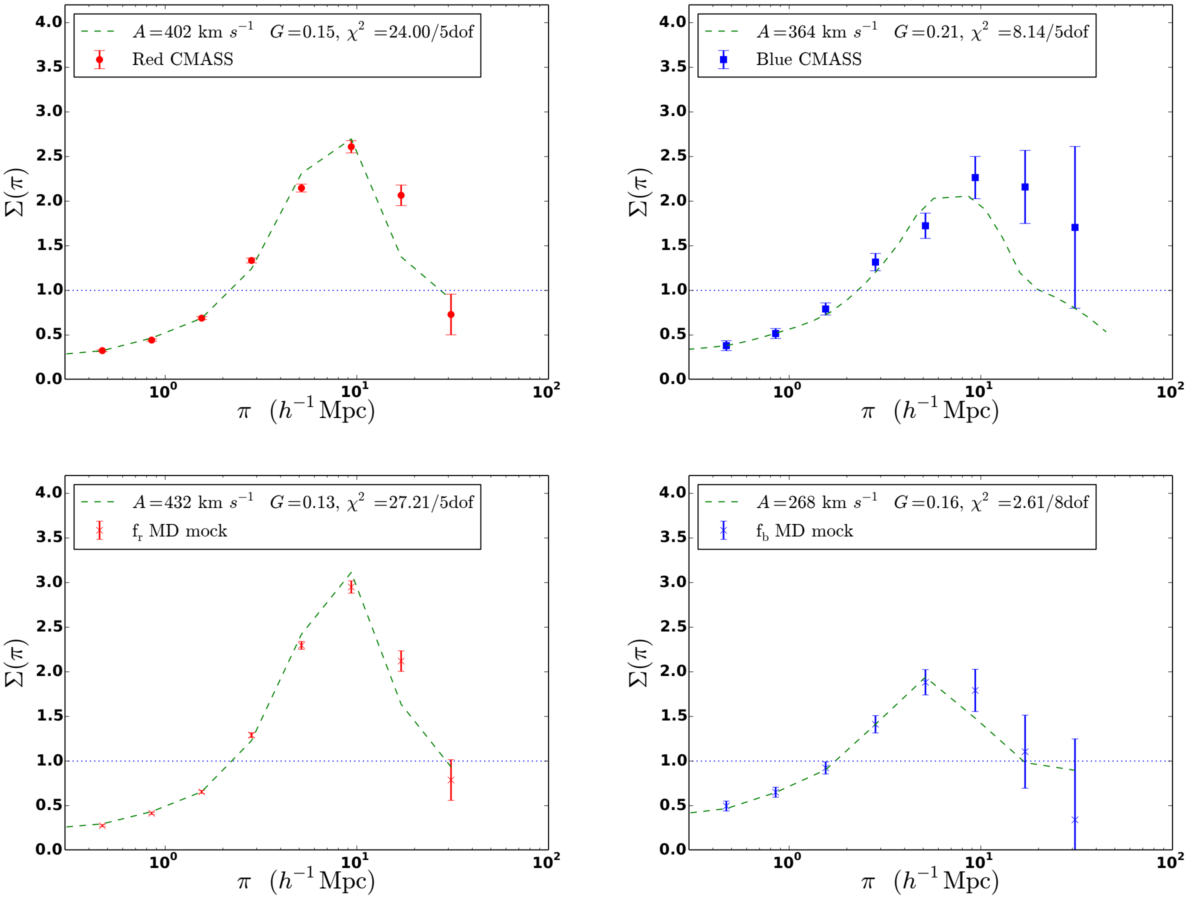

Figure 3.15: *Top row:* CMASS DR11 $\Sigma(\pi)$ red (left) and blue (right) measurements and the A, G analytic models (dashed lines). *Bottom row:* $f_{b,r}$ MultiDark mocks (crosses) and their models (dashed lines). For the mocks we adopt the jackknife errors estimated for the blue CMASS data doing jackknife. These fits are fully covariant. From these plots we conclude that blue CMASS galaxies are less biased and show a lower peculiar velocity contribution compared to the red population.

covariant and our best estimate of the A, G parameters are reported in Table 3.2.

As expected, the blue CMASS galaxies are less biased and have lower peculiar velocity contribution (i.e., smaller FoG elongation effect) compared to the red sample. A similar behavior is seen in a comparison of the red and the blue MultiDark model galaxies, suggesting that we are correctly modeling our results in terms of redshift-space distortions and large-scale bias. Our relatively high bias values are due to the specific high-redshift CMASS selection we are considering. In fact, for CMASS galaxies at $z > 0.55$, the bias is expected to be higher than the typical value reported by Nuza et al. [215], $b \sim 2$. As discussed in Section

	A $(\mathrm{km\,s^{-1}})$	G	b	χ^2
Full CMASS	384 ± 6	0.15 ± 0.01	~ 3	16.89/5dof
Full mock	402^{+9}_{-6}	$0.14^{+0.01}_{-0.02}$	~ 3	24.04/5dof
Red CMASS	402^{+8}_{-9}	$0.15^{+0.01}_{-0.02}$	~ 3	24.00/5dof
Red mock	432^{+10}_{-8}	0.13 ± 0.01	~ 3.5	27.21/5dof
Blue CMASS	364^{+47}_{-39}	$0.21^{+0.05}_{-0.04}$	~ 2	8.14/5dof
Blue mock	268 ± 35	$0.16^{+0.07}_{-0.09}$	~ 2.8	2.61/8dof

Table 3.2: Best-fit values of the A, G parameters that model $\Sigma(\pi)$ in both full, red, blue CMASS measurements and MultiDark mocks. All the fits are fully covariant. The bias is computed using the approximation given in Eq. 3.15, where β is our G parameter, see Section 3.3.6.

3.5.2, the relatively high χ^2 values we find from our model fits are due to the numerical limitations in the $\Sigma(\pi)$ definition. However, since the goal of this work is a qualitative comparison of the full, red and blue CMASS redshift-space clustering and bias features, we do not heavily focus on the goodness of the fits and give priority to the qualitative interpretation.

Figure 3.16 presents the 68% and 95% covariant confidence regions of the A, G models for the CMASS measurements. The 1σ blue region is spread out: due to their larger uncertainties, blue galaxies have less power to constrain the A, G values compared to the red and full CMASS populations. The dots indicate the position of the best-fit models for the three samples. As seen in Figure 3.15, red CMASS galaxies possess higher velocity dispersion and large-scale bias compared to the blue sample.

3.7.2 large-scale bias

The linear bias factor b, defined in Eq. 3.30, is related to the red-blue cross-correlation, $\xi_\times(s)$, by

$$b_r(s)b_b(s) = \frac{\xi_\times(s)}{\xi_m(s)}. \tag{3.35}$$

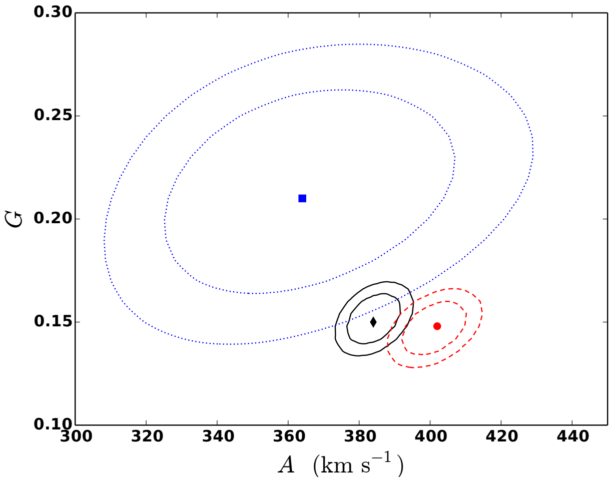

Figure 3.16: 68% and 95% confidence levels of the full (solid), red (dashed) and blue (dotted) $\Sigma(\pi)$ CMASS measurements shown in Figs. 3.8 (left panel) and 3.15 (top row). All the contours include covariances. Consistently with the size of the error bars in Figure 3.15, the blue contours are much less tight than the red and full ones. The blue CMASS galaxies are less biased and have lower velocity dispersion than the red and full populations.

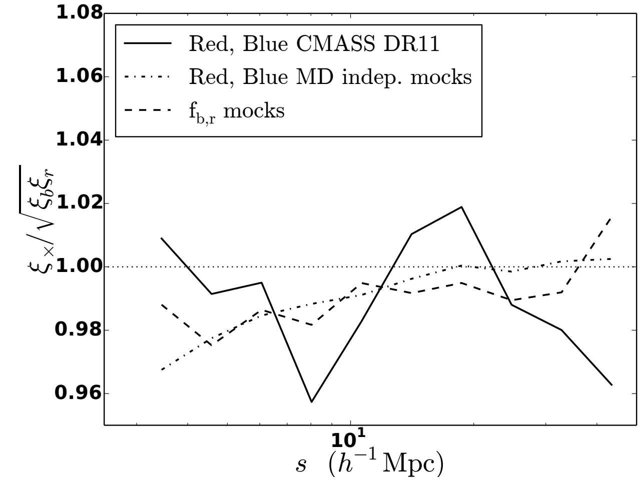

Figure 3.17: Ratio of the quantity $b_b b_r$ computed using the red-blue cross-correlation function, over the same quantity computed using the red and blue auto-correlation measurements. CMASS data (solid) versus independent (dot-dashed) and inverse tangent (dashed) mocks. Compatibly with expectations, the result is consistent with unity within 5% and the fluctuations are Poisson noise.

where the subscripts r, b indicate, respectively, red and blue galaxies, and $\xi_m(s)$ is the dark matter correlation function. We then expect that the ratio $\xi_\times(s)/\sqrt{\xi_r(s)\xi_b(s)}$ – where each term in the denominator is given by Eq. 3.30 – is close to unity. Figure 3.17 shows that our analysis produces a result that is consistent with expectations within 5%.

3.8. Discussion and conclusions

We have presented a qualitative analysis, as a function of color, of the clustering signal in the high-redshift tail (i.e., $z > 0.55$) of the BOSS CMASS DR11 massive galaxy sample. Applying the color cut defined in Eq. 3.27, we have divided the full CMASS sample into a redder and a bluer populations of galaxies, and there we have computed the redshift-space and projected correlation functions at small and intermediate scales ($0.1 \leq r \leq 50h^{-1}$Mpc). Our measurements are consistent with previous results by [311], [333], [297] and show that blue star-forming galaxies preferentially populate less dense environments, compared to the red ones. Besides the 2PCF results, we have defined and measured a new quantity, $\Sigma(\pi)$ (Eq. 3.4), which provides robust information about nonlinear small-scale redshift-space distortions and large-scale linear bias by disentangling the different effects (i.e., finger-of-god and Kaiser flattening) along and across the line of sight.

We have then mapped these results onto the MultiDark cosmological simulation using a five-parameter halo occupation distribution model to generate reliable mock galaxy catalogs that reproduce the observed clustering signal in all the CMASS sub-samples considered. First, using a traditional HOD approach, we have separately fit $\xi(s)$ of the full, red and blue CMASS populations building three independent mock catalogs (three different HOD models, with independent parameters). Instead of performing a formal fit, we have empirically tuned the HOD input parameters until we found suitable values that reproduce the observed clustering amplitude. To simplify the task, we have chosen to vary only three parameters, specifically those values related to physical quantities we want to measure: M_{min}, the minimum host halo mass, which is connected to the galaxy number density, M_1', governing the satellite fraction, and α, the slope of the satellite contribution. The remaining parameters are fixed to their default values given by White et al. [318]. Our best empirical estimates for the independent HODs are reported in Table 3.1, and confirm that red galaxies preferentially populate more massive halos, with a higher satellite fraction compared to the bluer, star-forming population. From this results we conclude that we are able to individually match the clustering of the full, the red and the blue CMASS samples with small variations in the input parameters. Using these independent mocks, we have calculated the probability, $P(M_h|G)$, that a specific galaxy G is hosted by a halo with central mass M_h (left panel of Figure 3.14), and estimated the mean central halo masses of our red and blue model galaxies. We found $M_h \sim 10^{12.5}$, $10^{13.0}\,h^{-1}\,M_{\odot}$, respectively for star-forming bluer and redder galaxies, which again confirms that red galaxies live in more massive halos.

The traditional HOD formulation reproduces both red and blue CMASS clustering results; however, it is based on a non-physical assumption: being independent, the red and blue models share a certain number of mock galaxies. This means that the same galaxy can be either red or blue, whatever its mass is. To overcome this failure, we have modified the standard HOD assignment to infer both red and blue models from the full one, in such a way they are complementary and do not overlap. We have split (see Section 3.6.2) the full mock catalog into a red and a blue sub-mocks by constraining it with an appropriate condition that mimics the observed CMASS red/blue galaxy fraction, $f_{b,r}$ (Eq. 3.33). We have tested four different functional forms of $f_{b,r}$ (see Appendix 3.9.3 for details), depending on a different number of parameters, and concluded that the best one is an inverse-tangent-like function (Eq. 3.34). It only depends on two free parameters, C and D, which respectively determine how fast the blue (red) fraction drops (grows) as the halo mass increases, and the position of the half-width point of the curve. Our results, presented in Figure 3.11, show good agreement between the MultiDark model galaxies and the CMASS observations.

We have then quantified the differences in the blue and red CMASS sub-populations from the point of view of the redshift-space distortions and large-scale bias (Section 3.7). Two regimes are interesting to this purpose: on large scales, the gravitational infall of galaxies to density inhomogeneities compresses the two-point correlation function along the line-of-sight direction; on small scales, the 2PCF experiences an elongation effect due to the nonlinear peculiar velocities of galaxies, with respect to the Hubble flow (see Sec. 3.3.1). In order to separate the two contributions and isolate the small scale elongation effect, we have built the new metric $\Sigma(\pi)$, defined in Eq. 3.4 as the ratio between a 2PCF – spherically averaged in the range $0.5 \leq r_p < 2\,h^{-1}$Mpc to maximize the FoG effect – and the best-fit power law (spherically averaged in the same way) to the projected correlation function. This quantity is preferable than the quadrupole-to-monopole ratio in the attempt to maximize the finger-of-god contribution on small scales, because it permits to separate the redshift-space features along and across the line of sight (see Section 3.9.1). We have then modeled $\Sigma(\pi)$ by convolving the real-space best-fit power law to $w_p(r_p)$, with a peculiar velocity term, assumed to be a normal function (Eq. 3.18) and the Kaiser factor (Eq. 3.16). The resulting model only depends on two parameters: G, that measures the Kaiser compression and is proportional to the inverse of the linear bias, b, and A, that is the pairwise velocity dispersion, which quantifies the FoG elongation effect. Fitting this A, G parametrization to our full, red, blue $\Sigma(\pi)$ BOSS CMASS DR11 and MultiDark mock results, we found (see Table 3.2) that blue galaxies are less biased than red ones and have a lower peculiar velocity contribution, which leads to a smaller clustering amplitude.

In conclusion, we have performed a qualitative clustering analysis as a function of color in a specific massive galaxy sample, the BOSS CMASS DR11, selecting only galaxies at $z > 0.55$. We have divided the sample in a redder and a bluer sub-populations, and here we have measured the monopole, the projected 2PCF and a new quantity, $\Sigma(\pi)$, which is a compression of $\xi(r_p, \pi)$, specifically designed to study the FoG distortions and the linear bias. We have proposed and tested a straightforward model for $\Sigma(\pi)$, depending only on two parameters, that allows to derive robust constraints on both large-scale bias and galaxy peculiar velocities, and provides a more exhaustive vision of the red/blue galaxy bimodality.

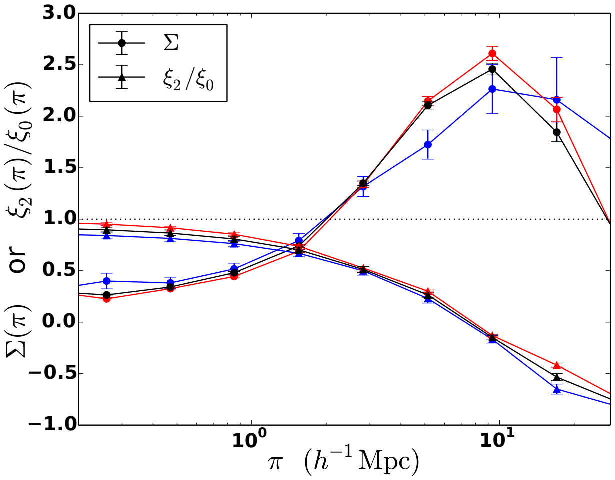

Figure 3.18: $\Sigma(\pi)$ (points) versus $\xi_2(s=\pi)/\xi_0(s=\pi)$ (triangles) measurements for the full (black), red and blue CMASS samples. The advantage of measuring $\Sigma(\pi)$ is that, by construction, it allows to disentangle the small-scale non-linear redshift-space distortion effects from the Kaiser squashing on larger scales.

3.9. Appendix

3.9.1 Quadrupole-to-monopole ratio versus $\Sigma(\pi)$ statistics

The novelty of our $\Sigma(\pi)$ statistics is that it allows one to extract the maximum contribution of small-scale redshift-space distortions separating the effects along the line of sight (i.e., finger-of-god) from the effects across it (i.e., Kaiser squashing). In fact, $\Sigma(\pi)$ is defined (see Eq. 3.4) by normalizing out the real-space contribution from the redshift-space 2PCF, spherically averaged in the range $0.5 \leq r_p \leq 2\ h^{-1}$Mpc to maximize the FoG effect. In alternative to $\Sigma(\pi)$, one could measure the quadrupole-to-monopole ratio, $\xi_2(s)/\xi_0(s)$, to extract information about the redshift-space clustering features. However, this ratio is computed as a function of the redshift-space distance $s=\sqrt{r_p^2+\pi^2}$, and does not permit to disentangle the FoG elongation from the Kaiser flattening. By modeling $\Sigma(\pi)$ in a straightforward way only as a function of two parameters A, G (see Section 3.3.6), we are able to separate the small-scale non-linear FoG regime, where the peculiar velocities (quantified by A) dominate, from the large-scale linear regime, where the Kaiser compression becomes important, and

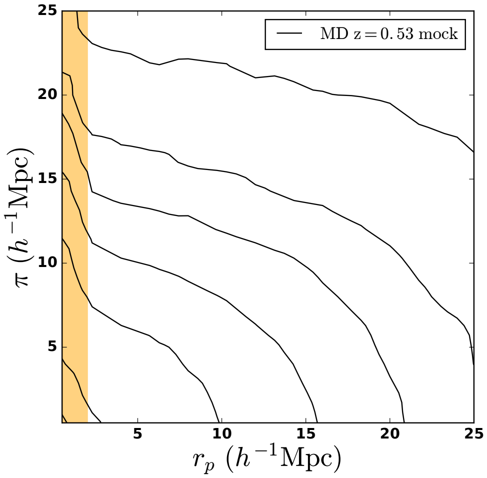

Figure 3.19: Two-point correlation function of the full CMASS MultiDark mock galaxy catalog. The orange shaded area represents the $\Sigma(\pi)$ domain where the finger-of-god effect is maximized.

the linear bias (quantified by G through Eq. 3.21) can be estimated. In Figure 3.18 we show a comparison of the full, red and blue CMASS DR11 $\Sigma(\pi)$ measurements and quadrupole-to-monopole ratios, these latter evaluated at $s = \pi$. The feature of the two metrics is not comparable, nor the information they carry. The advantage of using $\xi_2(s)/\xi_0(s)$ is the smaller size of the error bars, which would lead to tighter constraints in the analysis. On the other hand, the advantage of using $\Sigma(\pi)$ is, as explained above, that it permits to quantify both linear galaxy bias and FoG contribution in a straightforward way, disentangling the effects as a function of the physical scale. In Figure 3.19 we display the 2PCF of the MultiDark model galaxies for the full CMASS sample, given as a function of the parallel (π) and perpendicular (r_p) components to the line of sight. The orange shaded region highlights the $\Sigma(\pi)$ domain, where the finger-of-god effect is maximized.

3.9.2 Clustering sensitivity on HOD parameters

The left column in Figure 3.20 presents our HOD model (see Section 3.3.5) as a function of three parameters: M_{min} (top row), M_1' (middle), and α (bottom). We allow to vary only one parameter at a time and the remaining ones are fixed at the fiducial values given by White

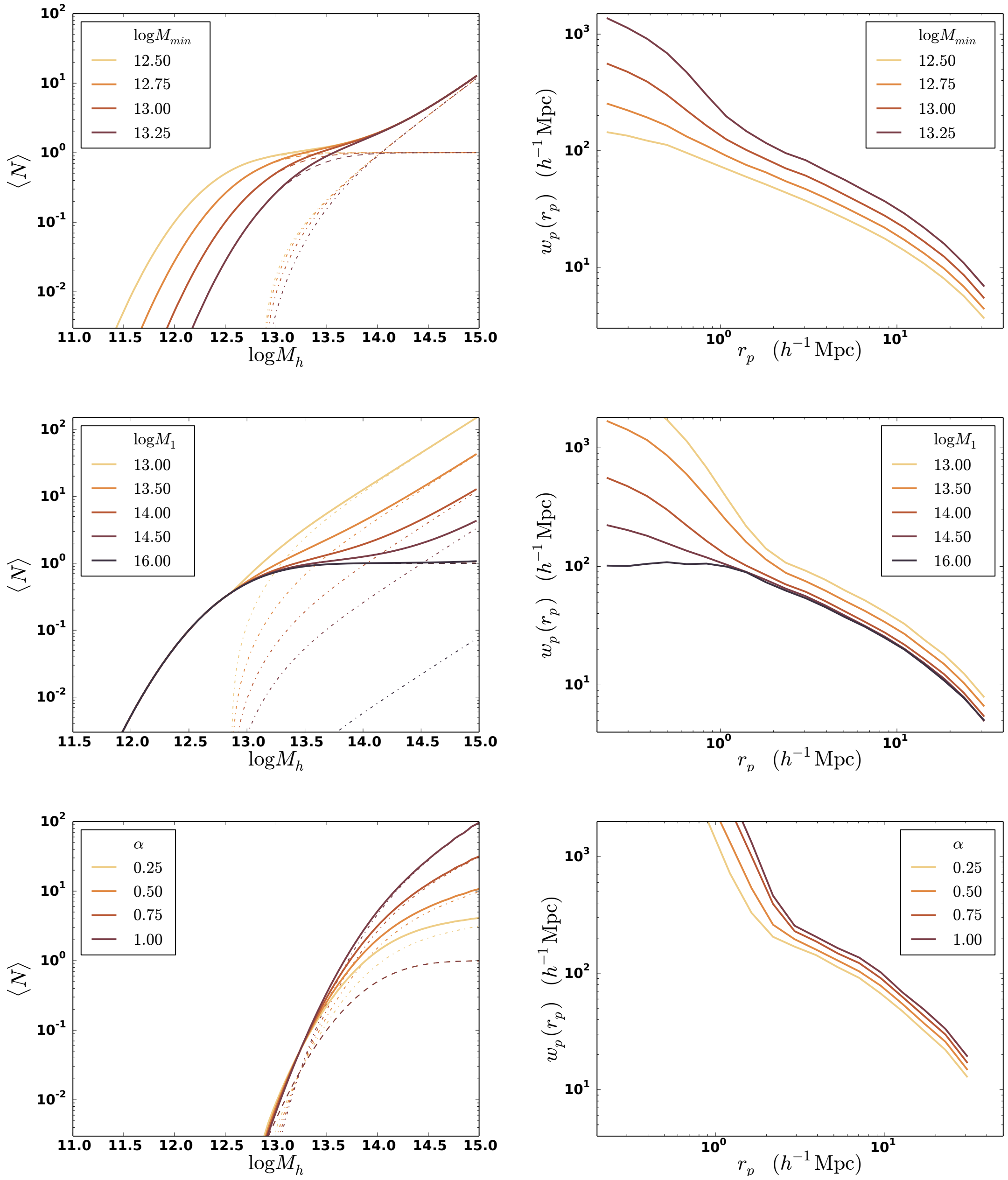

Figure 3.20: Implication of a change in the HOD input parameters (left column) on the projected correlation function (right column). We allow to vary only one parameter at a time, and fix the others to the fiducial values given by White et al. [318]. In the top row is displayed the variation of M_{min}, which especially affects the 2-halo term. A change in M_1' or α, respectively in the middle and bottom row, has a strong impact on the 1-halo term. The resulting correlation functions are degenerate with respect to these three model parameters.

et al. [318]. The projected correlation functions based on these mocks are shown in the right column. Increasing the value of M_{min} (top row, from lighter to darker solid lines) globally enhances the clustering amplitude, with a strong contribution from sub-structures belonging to different hosts (2-halo term). On the other side, the interaction between satellites belonging to the same central halo (1-halo term) is suppressed as M_1' increases (bottom row, from lighter to darker solid lines), resulting in a smoother slope at scales $r_p \leq 1h^{-1}$Mpc. The extreme case is achieved when $logM_1$ =16.00, where the satellite contribution becomes almost negligible, and $f_{sat} = 5.45 \times 10^{-4} \simeq 0$.

3.9.3 Red and Blue galaxy fraction models

In addition to the inverse tangent fraction model defined in Eq. 3.34, to mimic the red and blue galaxy fractions as a function of the central halo mass, we test also a linear model

$$f_b(\log M_h) = -M \log M_h + N, \tag{3.36}$$

and two log-normal-like functions, with three degrees of freedom each. The first one (Logn I) is given by

$$f_b(\log M_h) = \frac{P_b}{P_b + P_r}, \tag{3.37}$$

where

$$P_{b,r} = \exp\left(-\frac{(\log M_h - \mu_{b,r})^2}{2\sigma^2}\right) \tag{3.38}$$

is a density function. The parameters $\mu_{b,r}$ are the blue and red mean galaxy masses, respectively, and σ is the log-normal width. The second version (Logn II) has fixed amplitude σ, and a new parameter, k, that controls the mutual heights of the red and blue peaks. We have

$$f_b(\log M_h) = \frac{P_b}{P_b + kP_r}, \tag{3.39}$$

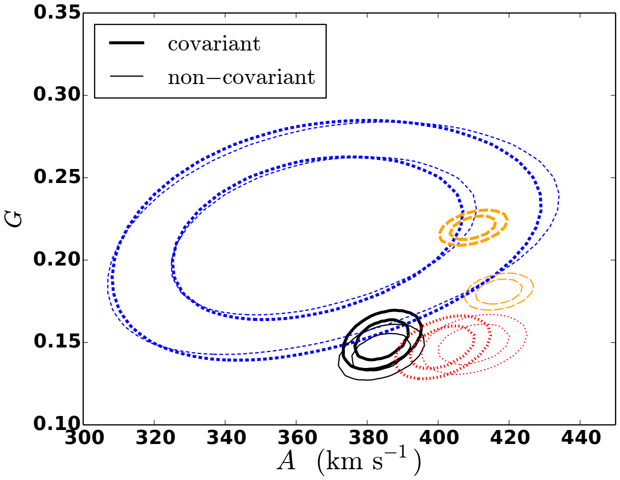

Figure 3.21: Covariant (thick contours) versus non-covariant (thin lines) 68% and 95% confidence levels of the A, G models for the $\Sigma(\pi)$ full (black solid), red (red dotted) and blue (blue dashed) CMASS measurements versus QPM mocks (orange long-dashed). QPMs have slightly different cosmology: $\Omega_m = 0.29$. The inclusion of covariances is almost negligible for the blue population, and weakly appreciable in the full case. Inversely, in the red population, covariances slightly move the fit towards higher velocity values; for QPMs, this shift is significant and drives the contours towards smaller bias values and slightly higher velocities.

where $P_{b,r}$ is given by Eq. 3.38. After applying these constraints to the full MultiDark mock catalog, we split it into its red and blue components. We then fit the clustering amplitudes of our model galaxies to the CMASS red and blue samples.

3.9.4 Testing the errors – jackknife versus QPM mocks

We test our full CMASS jackknife error estimates by computing the $\xi(s)$, $w_p(r_p)$, and $\Sigma(\pi)$ covariance matrices from a set of 100 Quick Particle Mesh [QPM; 319] mock catalogs, with slightly different cosmology: $\Omega_m = 0.29$. Since these mocks are all independent of each other, we can compute their covariance as

$$C_{kl}^{QPM} = \frac{1}{n_{QPM} - 1} \sum_{b=1}^{n_{QPM}} (\xi_k^b - \bar{\xi}_k)(\xi_l^b - \bar{\xi}_l), \tag{3.40}$$

where $n_{QPM} = 100$, and $\bar{\xi}_k$ is the mean QPM correlation function in the k^{th} bin,

$$\bar{\xi}_k = \sum_{b=1}^{n_{QPM}} \xi_k^b / n_{QPM}. \tag{3.41}$$

Figure 3.21 compares the covariant (thick lines) and the non-covariant (thin) A, G contours of the full, red and blue CMASS $\Sigma(\pi)$ models with the QPM mocks (orange, long-dashed). The inclusion of covariances is almost negligible for the blue CMASS model, while it moves the full and red models toward smaller bias values and higher velocity dispersion values, respectively. QPM contours are narrow, analogously to the full CMASS sample, and the inclusion of covariances in this case significantly moves the fit towards lower bias values and slightly higher velocities.

Figure 3.22 compares the normalized $\xi(s)$, $w_p(r_p)$, and $\Sigma(\pi)$ (from left to right) covariance matrices estimated using the QPM mocks (top row) and the jackknife re-samplings of the full, red and blue CMASS galaxy samples, to test the correlation between our observations at different scales. Overall, the QPM mocks show stronger covariances than jackknife in all three metrics. $\Sigma(\pi)$ is less correlated than the redshift-space and projected correlation functions; this is due to its definition, see Eq. 3.4. Since $\Sigma(\pi)$ is the ratio of two clustering measurements, both errors propagate in it, resulting in a smoother correlation at all scales. The red CMASS sample includes the majority of the CMASS galaxies, thus it is reasonable that its covariance matrices behave similarly to the ones of the full sample. The blue case is slightly different: errors are larger and covariances are almost negligible in all the three measurements, especially in $\Sigma(\pi)$.

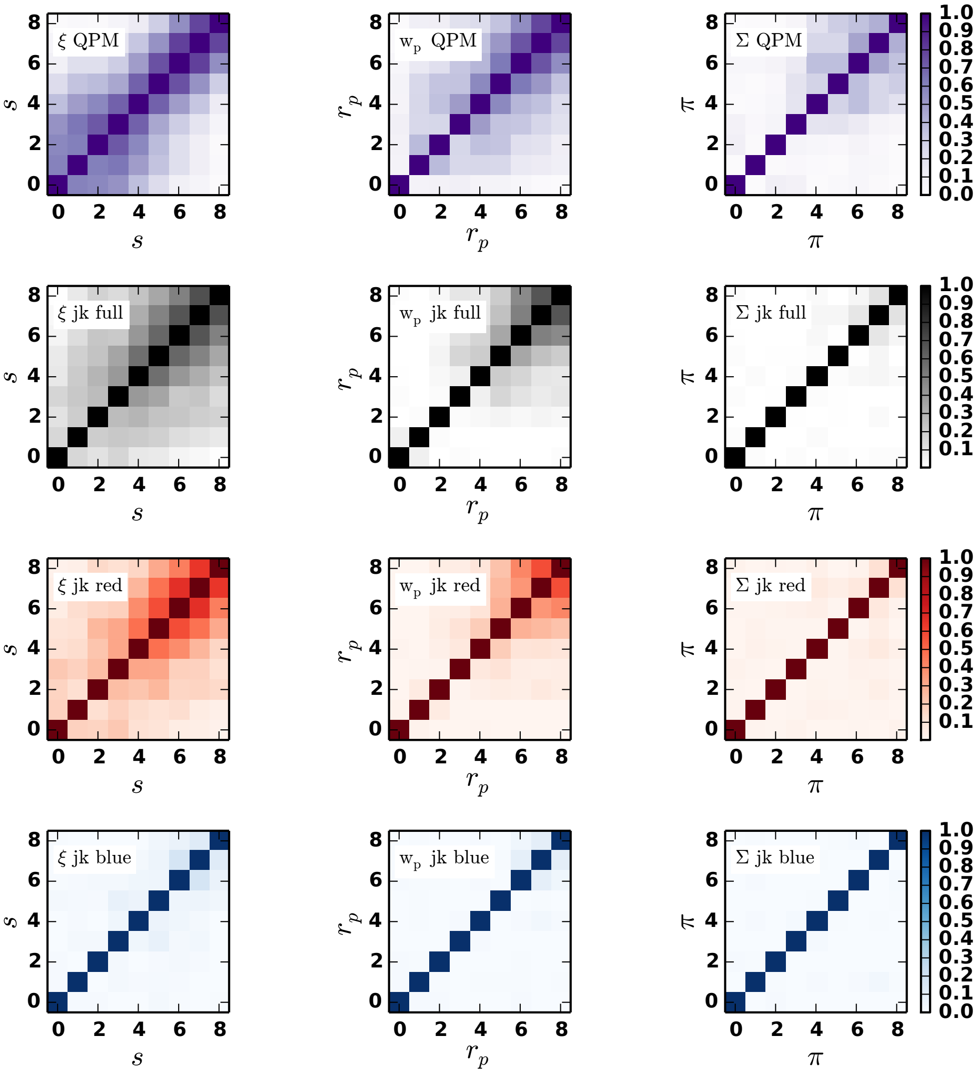

Figure 3.22: Normalized QPM (first row from the top) versus full (second row), red (third row) and blue (bottom row) CMASS jackknife covariance matrices for $\xi(s)$ (left column), $w_p(r_p)$ (central), and $\Sigma(\pi)$ (right), as a function of the s, r_p and π bins, respectively. We adopt a ten-step logarithmic binning scheme in the range $3 - 50\,h^{-1}$Mpc for s, $0.1 - 35\,h^{-1}$Mpc for r_p, and $0.1 - 40\,h^{-1}$Mpc for π. Overall, QPM mocks show higher covariances compared to the full, red, and blue CMASS samples, confirming the result shown in Figure 3.21. The left column reveals that covariances become appreciable in the red and full redshift-space 2PCFs at intermediate scales (i.e., $s \geq 8\,h^{-1}$Mpc), while they are almost negligible in the blue population. The red and full CMASS projected 2PCF are covariant at $r_p \geq 2\,h^{-1}$Mpc, while the blue case is almost covariance-free at all scales. The $\Sigma(\pi)$ measurements are significantly less covariant than the other two clustering statistics: QPM mocks show appreciable covariances only above $\pi \sim 3\,h^{-1}$Mpc, while the three CMASS samples are substantially covariance-free.

- If people sat outside and looked at the stars each night, I'd bet they'd live a lot differently.

- How so?

- Well, when you look into infinity, you realize that there are more important things than what people do all day.

Calvin & Hobbes - Stars and Infinity

4

Clustering properties of g-selected galaxies at $z \sim 0.8$

4.1. Abstract

Current and future large redshift surveys, as the Sloan Digital Sky Survey IV extended Baryon Oscillation Spectroscopic Survey (SDSS-IV/eBOSS) or the Dark Energy Spectroscopic Instrument (DESI), will use Emission-Line Galaxies (ELG) to probe cosmological models by mapping the large-scale structure of the Universe in the redshift range $0.6 < z < 1.7$. With current data, we explore the halo-galaxy connection by measuring three clustering properties of g-selected ELGs as matter tracers in the redshift range $0.6 < z < 1$: (i) the redshift-space two-point correlation function using spectroscopic redshifts from the BOSS ELG sample and VIPERS; (ii) the angular two-point correlation function on the footprint of the CFHT-LS; (iii) the galaxy-galaxy lensing signal around the ELGs using the CFHTLenS.

We interpret these observations by mapping them onto the latest high-resolution MultiDark Planck N-body simulation, using a novel (Sub)Halo-Abundance Matching technique that accounts for the ELG incompleteness. ELGs at $z \sim 0.8$ live in halos of $(1 \pm 0.5) \times 10^{12}\, h^{-1}\mathrm{M}_{\odot}$ and (22.5±2.5)% of them are satellites belonging to a larger halo. The halo occupation distribution of ELGs indicates that we are sampling the galaxies in which stars form in the most efficient way, according to their stellar-to-halo mass ratio.

4.2. Introduction

By investigating the properties of galaxy clustering within the cosmic web, it is possible to constrain cosmology and infer the growth of structure and the expansion history of the Universe [316]. In fact, galaxy clustering measurements using last-generation large-volume redshift surveys, as the Sloan Digital Sky Survey [SDSS; 329, 120, 275] and the SDSS-III Baryon Oscillation Spectroscopic Survey [BOSS; 91, 81] provide robust information about both the evolution of galaxies and the cosmological framework in which these complex structures live. In order to interpret such measurements, we need to understand the relation between the theory-predicted dark matter field and its luminous counterpart i.e., the discrete galaxy map [69].

Luminous, low-redshift galaxies have already been connected to their dark matter halos in a precise manner, through weak lensing and clustering analysis as a function of galaxy luminosity and stellar mass. Baldry et al. [14], Zehavi et al. [334] and Guo et al. [124] measured the clustering properties of the SDSS "blue cloud" and "red sequence" in the local Universe (SDSS median redshift $z \sim 0.1$; Abazajian et al. [1]), as a function of magnitude and color. Their results show that at a given luminosity, the blue sample has a lower clustering amplitude and a smaller correlation length compared to the red one.

Guo et al. [123] investigated the clustering luminosity and colour dependence of BOSS CMASS DR10 [8], and found that more luminous galaxies are more clustered and hosted by more massive halos. For luminous red galaxies (LRGs), these masses are $\sim 10^{13} - 10^{14} h^{-1}\mathrm{M}_{\odot}$, at fixed luminosity, progressively redder galaxies are more strongly clustered on small scales, which can be explained by having a larger fraction of these galaxies in the form of satellites in massive haloes. Favole et al. [100] measured galaxy clustering in the BOSS CMASS DR11 [8] sample at $z > 0.55$ as a function of color, and proposed a new statistic to extract robust information about small-scale redshift-space distortions and large-scale galaxy bias. Consistent with many previous results [e.g., 311, 333, 297], they found that, compared to the blue population, red galaxies reside in more massive halos, show a higher clustering amplitude, large-scale bias and peculiar velocities.

This type of clustering analysis has recently been extended to higher redshifts thanks to the VIMOS Public Extragalactic Survey [VIPERS; 126, 110] and DEEP2 survey [211]. Compared to DEEP2, VIPERS has a much larger volume but has a lower redshift limit however, the signal-to-noise ratio in its spectroscopic measurements is higher. Using VIPERS data, Marulli et al. [190] measured the clustering properties of galaxies at redshift $z = 0.8$ as a function of their luminosity and stellar mass, and found that the clustering amplitude and the correlation length increase with these two quantities; see also the PRIsm MUlti-object Survey (PRIMUS) results by Skibba et al. [272] and Bray et al. [45]. Mostek et al. [205] measured the clustering of the red sequence and the blue cloud at $z = 0.9$, as a function of their stellar mass and star formation history, using DEEP2 data. They argued that blue galaxies are more clustered in the local Universe than at $z = 0.9$, and red galaxies are much more clustered locally than at high redshift. They also suggested that the clustering trend observed with star formation rate (SFR) can be explained mostly by the correlation between stellar mass and clustering amplitude for blue galaxies. Coil et al. [55] studied the DEEP2 clustering dependence on color and luminosity, and found that the dependence on color is much stronger than with luminosity, and is as strong with color at $z \sim 1$ as locally. They claimed no dependence of the clustering amplitude on color for galaxies in the red sequence, but a significant dependence for galaxies within the blue cloud. Cooper et al. [66] investigated the connection between star formation (SF) and environment in DEEP2 data at $z \sim 0.1$, and $z \sim 1$. Their results indicate that, locally, galaxies in regions of higher overdensity have lower star formation rates (SFRs), and their stars form more slowly than in their counterparts in lower density regions. At $z \sim 1$, this SFR-overdensity relation is inverted; this is in part due to a population of bright, blue galaxies in dense environments, which lacks a counterpart in the local Universe, and is thought to evolve into members of the red sequence from redshift 1 to 0.

The combination of clustering with weak galaxy-galaxy lensing (see e.g., [17]) allows one to gain insight on the large-scale structure formation, and directly probe the stellar-to-halo mass relation [SHMR; 180]. The galaxy-halo connection has been measured at $z < 1$ by Leauthaud et al. [181], Shan et al. [267], and Coupon et al. [70], using three different weak lensing surveys (COSMOS, [266]; CFHT-Stripe82 and CFHTLenS[1], [140, 96]); all obtained consistent results. Leauthaud et al. [181] performed the first joint analysis of galaxy-galaxy weak lensing, galaxy clustering, and galaxy number densities using COSMOS data, and provided robust constraints on the shape and redshift evolution of the SHAM relation in the redshift range $0.2 < z < 1$. At low stellar mass, the halo mass scales proportionally to $M_\star^{0.46}$; this scaling does not evolve significantly with redshift. At $M_\star > 5 \times 10^{10}\mathrm{M}_\odot$, the SHMR rises sharply, causing the stellar mass of a central galaxy to become a poor tracer of its parent halo mass. Combining observations in the CFHT-LenS/VIPERS field from the near-UV to the near-IR, Coupon et al. [70] found that the SHMR for the central galaxies peaks at $M_{h,peak} = (1.9^{+0.2}_{-0.1} \times 10^{12}\mathrm{M}_\odot)$, and its amplitude decreases as the halo mass increases. Hearin et al. [135] presented new measurements of the galaxy two-point correlation function and the galaxy-galaxy lensing signal from SDSS, as a function of color and stellar mass, and demonstrated that the age-matching model [134], which states that older halos tend to host galaxies with older stellar populations, exhibits remarkable agreement with these and other statistics of low-redshift galaxies.

Current (Sub)Halo-Abundance Matching [SHAM; 65, 302, 165, 215] and Halo Occupation

[1] http://www.cfht.hawaii.edu/Science/CFHTLS/

Distribution [HOD; 28, 170, 336, 337] models correctly reproduce the clustering measurement mentioned above. SHAM maps observed galaxies onto dark matter halos directly from N-body cosmological simulations, according to a precise monotonic correspondence between halo and galaxy number densities. The HOD method is an analytical prescription to populate simulated halos with galaxies, in which the assignment is perfomerd by interpolating the halo occupation distribution at the values of the desired halo masses. In this sense, the SHAM approach returns a model which is built directly on the considered simulation box.

Next generation high-redshift surveys (see Section 1.7) as SDSS-IV/eBOSS [eBOSS; 80], Subaru Prime Focus Spectrograph [PFS; 294, 274] , DESI [264], 4MOST[2] and EUCLID[3] [176, 262] will use emission-line galaxies (ELGs) as BAO tracers to explore the Universe large-scale structure out to $z \sim 2$. Observing ELGs, learning how to model their clustering properties and understanding how they populate their host halos are therefore crucial points that we need to understand in order to select the targets for future experiments. From the observational point of view, the recent increment of available ELG spectroscopic data [126, 64] allows one to measure their clustering properties over about 12 deg^2 at $z = 0.8$ (corresponding to a comoving volume of $V \sim 10.6 \times 10^6 \, h^{-3}\mathrm{Mpc}^3$ in the Planck cosmology; see Section 4.4 for details), which represents a dramatic improvement.

Comparat et al. [63] demonstrated that neither a standard HOD nor a traditional SHAM technique are able to reproduce the angular clustering of ELGs on small scales. In fact, both techniques are based on the assumption that the galaxy sample to model is complete, but this is not the case of the ELGs, which are highly incomplete in stellar mass. One could instead use semi-analytic models of galaxy formation and hydrodynamic simulations, but they lack of mass resolutions to model emission line galaxies.

The aim of this work is to provide a modified version of the standard SHAM prescription, directly based on the latest MultiDark N-body simulation with Planck cosmology, that

[2] https://www.4most.eu/cms/

[3] http://sci.esa.int/EUCLID/

accounts for the ELG incompleteness and returns suitable mock galaxy catalogs able to accurately predict the ELG angular and redshift-space clustering, respectively, on small and larger scales. These mock catalogs are released to the public.

This chapter is organized as follows. Section 4.3 describes the data sets and the MultiDark simulation box used in our analysis. In Section 4.4 we present our ELG clustering and weak lensing measurements. In Section 4.5 we explain how we model the ELG clustering and we present our main results. Section 4.6 discusses the implications of our ELG clustering analysis in a galaxy evolution perspective, and Section 4.7 summarizes our main results.

Throughout the paper, we assume the Planck cosmology [236] and magnitudes in the AB system [217].

4.3. Data and simulation

4.3.1 Data sets

We build our ELG galaxy sample using the Canada-France-Hawaii Telescope Legacy Survey (CFHT-LS) Wide T0007[4] photometric redshift catalog [149, 71]. We apply a g-band magnitude cut, $20 < g < 22.8$ [107], to select galaxies with bright emission lines and low dust at $z < 1$. We also apply a color selection, $-0.5 < (u - r) < 0.7\,(g - i) + 0.1$, to remove the low-redshift galaxies. For details on the selection function, see Comparat et al. [64]. Then, to obtain the largest possible area, we convert the i-selection into the new Megacam i-band filter[5]. For the W1, W3 and W4[6] fields, we derive an average density of about 500 ELGs per deg^2, 70% of which have a photometric redshift in the range $0.6 < z < 1$. The densities of each field are reported in Table 4.1, and the errors on the photometric redshift are $\sigma_z < 0.05\,(1 + z)$ for $i < 22.5$ and $z < 1$. The $ugri$ ELG selection is brighter than $i < 22.5$.

We match the photometric targets to the available spectroscopic surveys – BOSS DR12,

[4] http://www.cfht.hawaii.edu/Science/CFHTLS/

[5] http://www4.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/megapipe/docs/filt.html

[6] http://www.cfht.hawaii.edu/Science/CFHLS/T0007/T0007-docsu10.html

Field	W1	W3	W4	all
Center α, δ	35°, -7°	215°, 54°	333°, 2°	-
area [deg^2]	63.75	44.22	23.3	131.27
N	32,808	22,195	11,025	66,028
N [deg^{-2}]	514.64	501.92	473.18	502.99
z_{phot} quartiles	0.78 / 0.88 / 1.03	0.77 / 0.88 / 1.05	0.78 / 0.88 / 1.03	0.78 / 0.88 / 1.03
z_{phot} Deciles D10, D90	0.7 / 1.24	0.68 / 1.29	0.69 / 1.25	0.69 / 1.26
N ($0.6 < z_{phot} < 1$)	23,433	15,242	7,861	46,536
N ($0.6 < z_{phot} < 1$) [deg^{-2}]	367.58	344.69	337.38	354.51
z_{phot} quartiles	0.75 / 0.83 / 0.9	0.74 / 0.81 / 0.89	0.75 / 0.83 / 0.89	0.75 / 0.82 / 0.89

Table 4.1: ELG photometric data per CFHT-LS Wide field after applying the bright star and bad field mask.

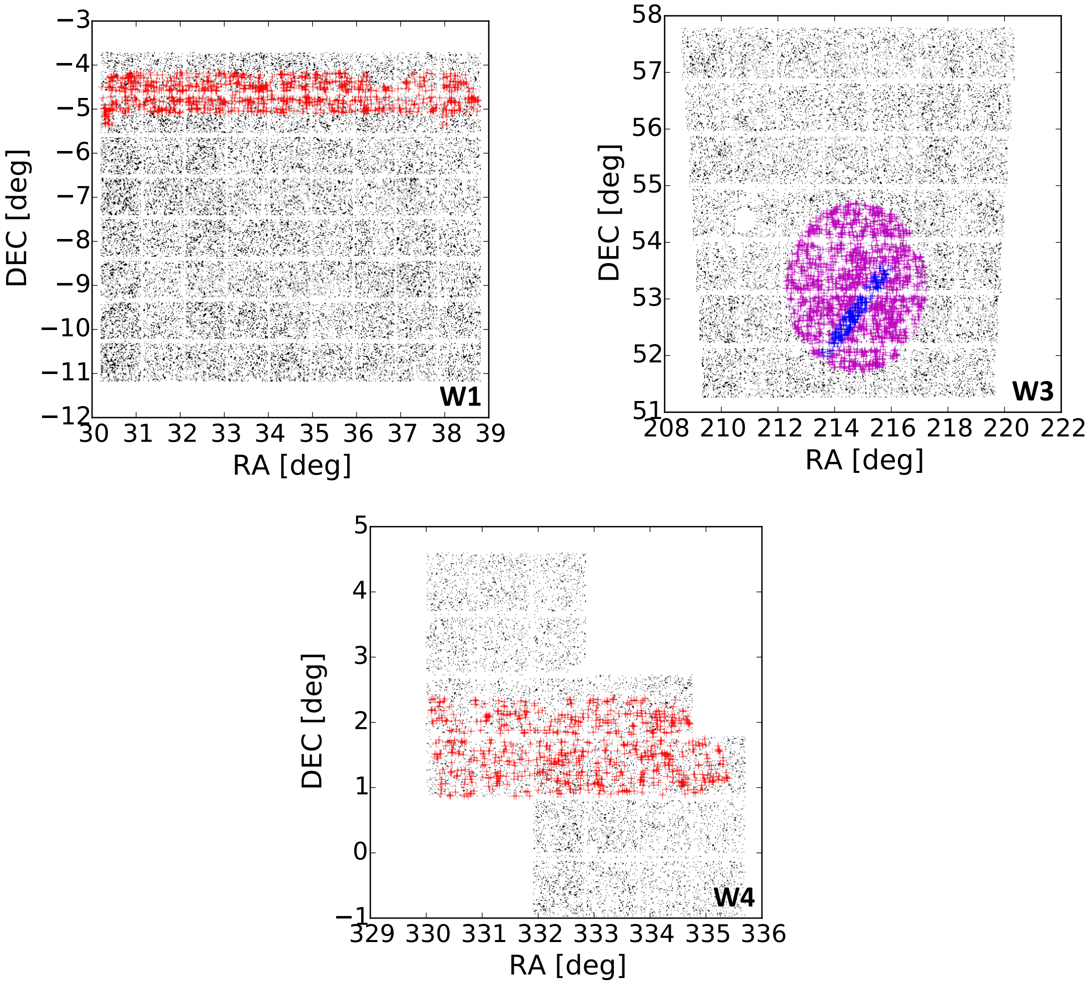

Figure 4.1: Photometric (black) and spectroscopic (VIPERS: red; BOSS: magenta; DEEP2: blue) coordinates of our ELG sample in the three CFHT-LS Wide fields.

Table 4.2: ELG spectroscopic data.

survey	match	good z	$0.6 < z < 1$	area [deg^2]	$\bar{z}$
VIPERS W1	1,223	942	760	5.478	0.803
BOSS W3	2,145	1,876	1,357	6.67	0.803
DEEP2 W3	225	222	156	0.5	0.803
VIPERS W4	1,148	846	680	5.120	0.795
All	4,741	3,886	2,953	17.668	0.803

DEEP2, VIPERS [42, 6, 211, 126] – within 1” radius; see Table 4.2. Based on KS-tests, the VIPERS, BOSS and DEEP2 spectroscopic selections constitute fair sub-samples of the complete selection: the hypothesis that they are drawn from the same distribution cannot be rejected at the 90% confidence level. For these samples, we create random catalogs with the same redshift distribution of the data and 30 times denser. Figure 4.2 displays the ELG spectroscopic redshift distribution per unit volume for the three Wide fields (solid histograms), and their mean (dashed line). Two thirds of the galaxy density is located in the redshift range $0.7 < z < 0.9$, while both the intervals $0.6 < z < 0.7$ and $0.9 < z < 1$ contain one sixth of the sample. According to the ELG selection function in [64], we select only galaxies at $z > 0.6$ since we are not interested in low-redshift objects. We have investigated further the impact of the higher redshift cut, $z < 1$, on the angular clustering by imposing to the ELG sample different redshift thresholds: $z < 1, 1.2, 1.4, 1.6$. In all these samples the lower redshift cut is fixed at $z > 0.6$ and we have imposed the $i < 22.5$ magnitude cut to eliminate bad photometric redshifts. We find that including also ELGs at $z \geq 1$, we are slightly enhancing the galaxy number density of our sample and consequently suppressing the amplitude of $w(\theta)$, but we do no see any substantial change in the angular clustering trend with respect to the $z < 1$ case. We therefore restrict the analysis to the redshift range $0.6 < z < 1$.

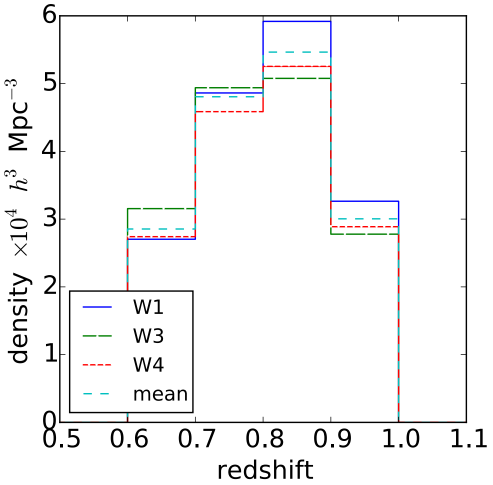

Figure 4.2: ELG weighted spectroscopic redshift distribution per unit volume for the W1, W3, and W4 Wide fields, and their mean value (dashed line).

4.3.2 MultiDark simulations

The MultiDark Planck simulation[7] (MDPL; [167]) contains 3840^3 particles in a $L = 1\,h^{-1}$Gpc box, and was created adopting a Planck ΛCDM cosmology [236]. Halos are identified based on density peaks including substructures using the Bound Density Maximum (BDM) halo finder [162, 254].

We use the MDPL halo catalogs to build a mock light-cone that matches the mean ELG redshift distribution shown in Figure 4.2. Given the high density of the ELG tracers and their expected low-mass host halos, the MDPL box is an excellent compromise between numerical resolution and volume. We apply the SUrvey GenerAtoR code [SUGAR; 257] to the 11 snapshots available from MDPL to construct a light-cone with a volume ten times the observations that covers the redshift range 0.6< z <1 ($\sim 1h^{-1}$ Gpc depth). The procedure used is analogous to the method presented by Blaizot et al. [32] and Kitzbichler & White [161], and can be summarized as follows:

1. Set the properties of the light-cone: angular mask, radial selection function (number

[7] www.MultiDark.org

density) and number of snapshots within the redshift range considered. Each slice of the light-cone is constructed by selecting all halos from every MDPL snapshot. The thickness of a slice at redshift z_i is given by $[(z_i + z_{i-1})/2, (z_i + z_{i+1})/2]$

2. Place an observer (i.e., $z = 0$) inside the box and shift the cartesian coordinates of the box in such a way that the observer occupies the central point of the box at $z = 0.8$

3. Convert from cartesian (x, y, z) to spherical (α, δ, r_c) coordinates, where r_c is the co-moving distance in real space. The redshift of each point will be:

$$r_c(z) = \int_a^b \frac{cdz'}{H_0\sqrt{\Omega_m(1+z')^3 + \Omega_\Lambda}} \tag{4.1}$$

4. From each snapshot, select the (sub)halos so that $(z_i + z_{i-1})/2 < z < (z_i + z_{i+1})/2$ and α/δ lie inside the sky window. Since the ELG observational data represent halos with typical masses $\sim 10^{12}h^{-1}\mathrm{M}_\odot$, in the light-cone we include all halos for which the simulation is complete i.e., $\log(M_h/h^{-1}\mathrm{M}_\odot) > 11.2$

5. Using the halo velocities, v_p, we compute the peculiar velocity contribution for each object along the line-of-sight and derive its distance in redshift-space as

$$s = r_c + (v_p \cdot r_c)/(aH(z)), \tag{4.2}$$

   where $a = (1+z)^{-1}$ is the scale factor and $H(z)$ is the Hubble parameter at redshift z

6. Finally, select objects from the light-cone using our selection function.

Throughout the paper we will designate our lightcone as “MDPL-LC”. Section 4.5 describes in detail the halo selection and the (Sub)Halo-Abundance Matching modeling adopted to determine the halo occupation distribution of our ELG sample.

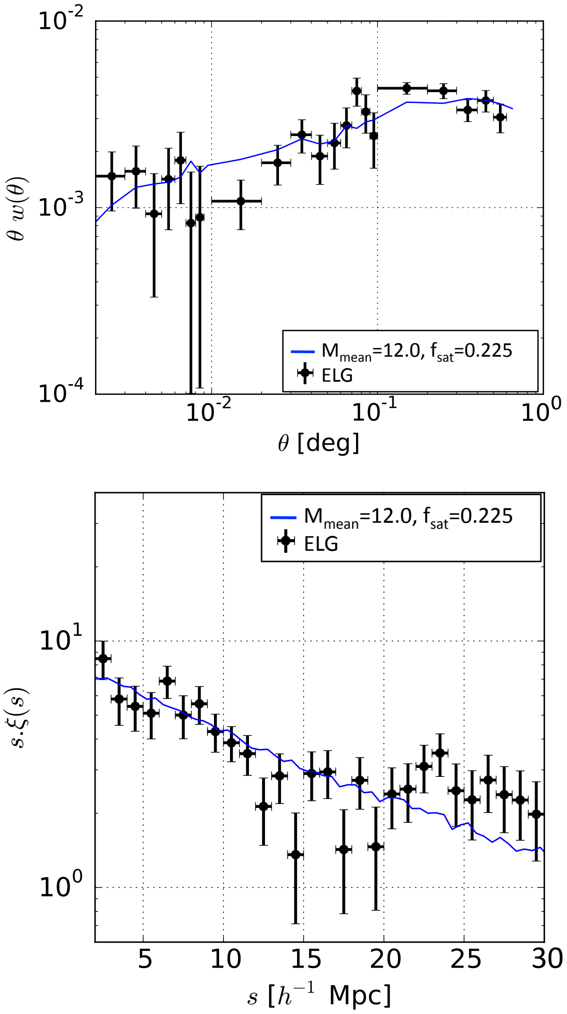

Figure 4.3: Two-point angular (top panel) and redshift-space (bottom panel) ELG correlation functions (points), together with our best-fit model (blue line), which corresponds to the point highlighted by a star in Figure 4.6.

4.4. Measurements

Using the ELG sample described in Section 4.3.1, we measure both galaxy clustering and galaxy-galaxy lensing. The following provides a detailed description of our measurements.

4.4.1 Galaxy clustering

We estimate both the angular, $w(\theta)$, and the redshift-space, $\xi(s)$ (hereafter ξ^s), two-point correlation functions following the procedures described by Landy & Szalay [175], Coupon et al. [72] and de la Torre et al. [83].

To compute ξ^s on the VIPERS and the BOSS ELG samples (see Table 4.2), independently, we create linear bins in separations of $1\,h^{-1}$Mpc at $s < 10\,h^{-1}$Mpc, and $4\,h^{-1}$Mpc for $10 < s < 40\,h^{-1}$Mpc. We then correct the impact of redshift errors and catastrophic redshifts to recover the correlation function down to $1\,h^{-1}$Mpc. The ELG we are targeting are observed using three plates overlapping the same area of the sky. This configuration guarantees that all the targets are observed at the end of the process and there is no fiber collision [37, 249]. For what concerns the finite size issue in VIPERS, we are correcting according to de la Torre et al. [83]. Finally, we combine the two measurements weighted by the projected density of each field. The resulting redshift-space correlation function is displayed in Figure 4.3 (bottom panel); fitting a power-law model, $\xi(s) = (s/s_0)^{\alpha}$, in the separation range $2 < s < 30\,h^{-1}$Mpc, we find $s_0 = (5.3 \pm 0.2)\,h^{-1}$Mpc and $\alpha = -1.6 \pm 0.1$.

Analogously, we calculate the angular 2PCF, $w(\theta)$, using photometric redshifts from the W1, W3 and W4 CFHT-LS fields. Because of the limited size of the sample, the angular correlations are biased to lower values. We correct this effect by implementing the integral constraint following Coupon et al. [72] and Tinker et al. [301]. On scales $\theta < 0.05°$, all three fields provide consistent measurements. At larger scales, the clustering signals in the W1 and W4 fields do not decrease as rapidly as expected, probably pointing to possible systematics that should need to be investigated further. We therefore use only the measurement on the W3 field, which appears the most robust.

The $w(\theta)$ of the W3 field (see top panel in Figure 4.3) is in perfect agreement with Figure 9 (panel 4) in Comparat et al. [63]. This result was computed on the Stripe 82 region [290], with three times larger area. At the mean redshift of the sample, $z = 0.8$, one degree corresponds to $18.847\,h^{-1}$Mpc; thus, $w(\theta)$ spans the range from $\sim 40\,h^{-1}$kpc up to $\sim 20\,h^{-1}$Mpc.

To estimate the errors on our galaxy clustering measurements, since the simulated light-cone area is larger than the data ($\sim$560 deg^2), we divide the best MDPL-LC model into independent (i.e., non-overlapping) realizations of our data (8 for the photometric and 24 for the spectroscopic samples), and obtain sample variance diagonal errors that we use rather than Poisson errors. We neglect a full-covariance analysis because the number of sub-samples we have is too small to produce reliable covariance estimates. Including also the off-diagonal elements of the covariance matrices would result in large fluctuations of the clustering error bars. Of course, excluding covariances we are adopting a simplified approach, but it provides a good sense of how the SDSS BOSS ELG clustering behaves. On the other hand, the ELG sample considered here is too sparse to derive tight constraints from our clustering analysis. New-generation large-volume spectroscopic surveys (see Section 1.7), as eBOSS, DESI and 4MOST, will provide new data with unprecedented statistics, sky coverage and deepness. Using those data, a fully covariant approach will return reliable and accurate error estimates.

We compare the combined ξ^s measurement from BOSS and VIPERS to previous measurements by Marulli et al. [190] to provide a first interpretation. Our result matches both the clustering signal of galaxies selected in the stellar mass range $9.5 < \log(M_*/h^{-1}\mathrm{M}_\odot) < 11$, and the clustering of galaxies selected by absolute magnitude in the interval $-22 < M_B - 5\log(h) < -20.5$. Using the stellar-to-halo mass relation from [181], [267] and [70], we can deduce a rough estimate of the halo masses populated by our ELG sample i.e., $11.6 < \log\ (M_h/h^{-1}\mathrm{M}_\odot) < 12.7$. These halo masses are typical of Milky-Way size halos, being much less massive than those hosting the LRG sample, see [215].

In the angular clustering measurement, the change of slope occurs at $\theta \sim 0.01°$, corresponding to $\sim 200\ h^{-1}$kpc. Using MDPL, we derive the relation between halo mass and virial radius at $z \sim 0.8$; halos with virial radius $\sim 200\,h^{-1}$kpc occupy the mass range $\mathrm{M}_h = (0.5-1) \times 10^{12} h^{-1}\mathrm{M}_\odot$. Since a single galaxy per halo would not induce such a change in the $w(\theta)$ slope, this result implies a satellite fraction of approximately 22.5% (see Section 4.5). Figure 4.3 displays a good agreement between our clustering measurements and predictions for ELG halos of mass $10^{12}h^{-1}\mathrm{M}_\odot$ with this satellite fraction.

4.4.2 Weak lensing

We use the latest weak lensing catalogs produced by the Canada-France-Hawaii Telescope Lensing Survey [CFHTLenS; 140, 96] on the W1 and W3 fields to measure the galaxy-galaxy lensing around 47,485 ELG lenses. This measurement allows one to constrain the halo masses. We follow Gillis et al. [114] and apply only the multiplicative correction, m_s, to the shear measurement and avoid the c2 correction. We measure the tangential shear, γ^t, around the photometric ELG sample as a function of the radial distance from the lenses using the [307] estimator:

$$\Delta\Sigma = \left[\frac{\sum_{ls} w_{ls}\gamma^t_{ls}\Sigma_c}{\sum_{ls} w_{ls}}\right] / \left[\frac{\sum_{ls} w_{ls}(1+m_s)}{\sum_{ls} w_{ls}}\right], \tag{4.3}$$

where the sum runs over the lens - source pairs (ls) and the w_{ls} values are the weight obtained by lensfit.

Since the lenses are at the higher tail of the redshift distribution and the ELGs are expected to live in low-mass halos, we recover a low signal-to-noise ratio around 2 for $R < 1$ Mpc.

We model the measurement using a truncated Navarro, Frank & White (NFW) halo profile [15] and the mass-concentration relation from Neto et al. [210] to truncate halos at half their concentration [326]. The best-fit model suggests typical halo masses of $M_{200} = 1.25 \pm 0.45 \times 10^{12} h^{-1} M_{\odot}$. The lower and upper mass limits are, respectively, $M_{200} = 5.61 \pm 7.20 \times 10^{11} h^{-1} M_{\odot}$ and $M_{200} = 1.41 \pm 0.51 \times 10^{12} h^{-1} M_{\odot}$; see Figure 4.4. This measurement is in good agreement with the first interpretations based on the clustering (see Section 4.4).

4.5. Halo occupation for emission-line galaxies

The (Sub)Halo-Abundance Matching [SHAM; e.g., 65, 302] technique is a straightforward method to link observed galaxies with dark-matter-only simulated halos. It relies in a monotonic correspondence between halo and galaxy number densities, which is based on the

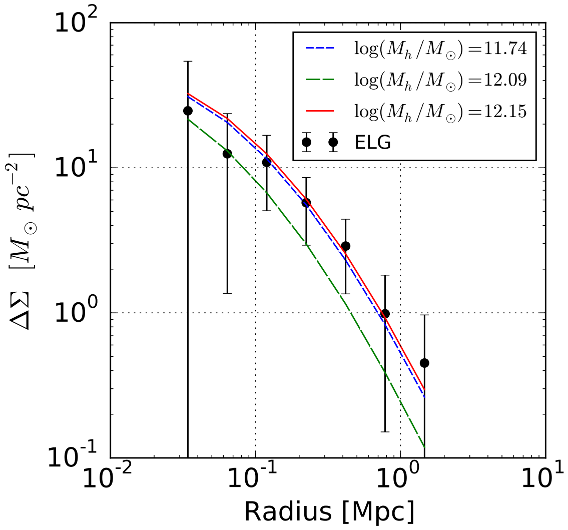

Figure 4.4: ELG surface density ($\Delta\Sigma$) as a function of the physical scale for different lens models.

assumption that more luminous galaxies reside in more massive halos. Such association is performed by choosing suitable proxies for both halos and galaxies (e.g., the halo maximum circular velocity and the galaxy luminosity or stellar mass) and includes some scatter [see e.g., 302]. The advantage of using N-body simulations, compared to analytical models, is given by the accuracy achieved in the predictions of the clustering for a given halo population. Many state-of-the art clustering measurements have been modeled using a SHAM technique that maps the observations onto suitable high-resolution N-body simulations, allowing the interpretation of the halo occupation distribution and bias [83, 215, 50]. Watson et al. [313] recently presented a method to upgrade SHAM models to account for differences between quenched and star-forming galaxies.

In the specific case of the emission-line galaxies, the traditional SHAM approach cannot be applied since it requires a complete galaxy sample, and ELGs are far from being complete in any parameter space, even in terms of their emission line luminosity, see Comparat et al. [63]. We therefore must modify the standard SHAM procedure to take into account the ELG incompleteness and match their clustering amplitude. To this purpose, we selected halos and subhalos by mass (for the subhalos we considered only the mass of the bound particles, to avoid ambiguities) to be able to compare directly with the weak lensing measurements. In the future, provided a high signal-to-noise ratio in the clustering measurement, we will properly select (sub)halos by their maximum circular velocity at accretion, [e.g., 21].

In order to model both the 1-halo and the 2-halo terms in the ELG two-point correlation functions and the weak lensing measurement, we use the MultiDark Planck $1h^{-3}$Gpc^3 box (see Section 4.3.2), which represents the best compromise between high resolution and volume, as previously described in Section 4.4.

We parametrize the probability of selecting a halo hosting an ELG as follows:

$$\begin{aligned} P(M_h, M_{mean}, \sigma_M, f_{sat}) = f_{sat}\mathcal{N}(M_h, M_{mean}, \sigma_M, \text{flag} = \text{sat})+ \\ +(1 - f_{sat})\mathcal{N}(M_h, M_{mean}, \sigma_M, \text{flag} = \text{cen}) \end{aligned} \tag{4.4}$$

where $\mathcal{N}$ is a Gaussian distribution with the variable being M_h, the halo mass. The parameters are: M_{mean}, the mean halo mass of the sample including both host and satellite halos; σ_M, the dispersion around the mean halo mass; f_{sat}, the satellite fraction. The additional parameter "flag" enables to identify among the halos the ones that are centrals (flag=cen) or the ones that are satellites (flag=sat).

To qualitatively understand the dependence of clustering on M_{mean} and f_{sat}, we impose (i) a maximum halo mass threshold to the MDPL-LC by removing all halos with $M_h > M_{max}$ and we apply the standard SHAM procedure. The higher-mass ($M_{max} > 10^{13}\,h^{-1}\mathrm{M}_\odot$) models reproduce well the observed $w(\theta)$, and that the lower-mass models ($M_{max} < 10^{13}\,h^{-1}\mathrm{M}_\odot$) match the large-scale clustering, but not the small-scale amplitude witnessed below $\theta \sim 0.01^\circ$. The top row in Figure 4.5 displays the ratio between the angular (left panel) and the monopole (right) correlation functions of the lower-mass models and the model with $M_{mean} = 10^{12}h^{-1}\mathrm{M}_\odot$. We see a mild variation in $w(\theta)$ as a function of the physical scale, and a flatter trend in the monopole.

We next (ii) fix the halo mass by selecting all the halos in the mass bin $M_h = (1 \pm 0.5 \times$

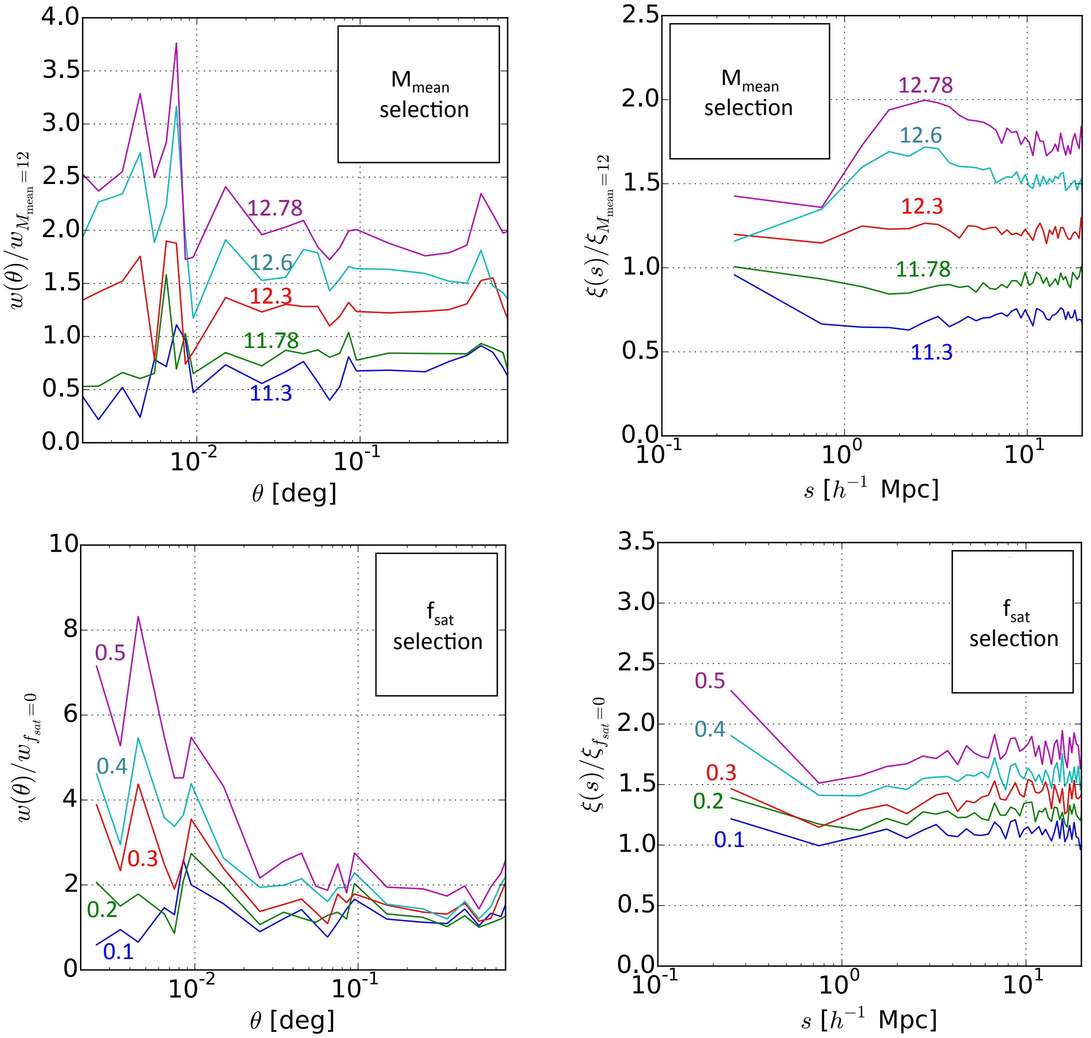

Figure 4.5: *Left column, top panel*: ratio of the angular correlation functions of the MDPL-LC halos selected by mass, to $w(\theta)$ computed at $M_{mean} = 10^{12}h^{-1}\mathrm{M}_{\odot}$. The curves in the plot go from lower mass (bottom line) to higher mass (top line). *Left column, bottom panel*: ratio of the angular correlation functions of the MDPL-LC halos with varying satellite fraction, to $w(\theta)$ computed at $f_{sat} = 0$. The lines in the plot go from lower f_{sat} (bottom line) to higher f_{sat} (top line). *Right column:* same results for the monopole. The top row presents our first experiment (see the text for details) on the lightcone: we impose different halo mass thresholds to the MDPL-LC and apply a standard SHAM. The bottom row displays SHAM in the mass bin $M_h = (1 \pm 0.5 \times 10^{12}\, h^{-1}\mathrm{M}_{\odot})$ with varying satellite fractions.

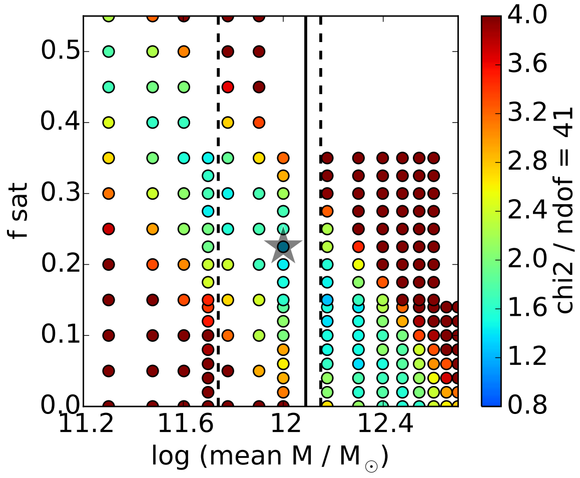

Figure 4.6: The two parameters driving the model: fraction of satellite (f_{sat}) and mean halo mass (M_{mean}). The spread around the mean halo mass is fixed at the value $\sigma_M = M_{mean}/2$. The vertical black lines represent the constraints by weak lensing (dashed: lower and upper limits; solid: mean), which rule out the majority of the low-mass and high-mass models. Our best-fit model is highlighted by the star symbol.

$10^{12}\,h^{-1}\mathrm{M}_\odot$), and vary the satellite fraction. We split this halo catalog into two catalogs, one containing only central halos ($f_{sat} = 0$) and one with satellites; then downsample both mocks to match the ELG $n(z)$. The bottom panels in Figure 4.5 present the variation of the angular and monopole clustering as a function of the scale. At small scales the amplitude of $w(\theta)$ with more than 30% satellite fraction is strongly enhanced compared to the $10-20\%$ cases. In the monopole there is almost no variation with the scale. We then combine these two products to build galaxy mock catalogs that contain a f_{sat} fraction of satellites (taken from the satellite-only mock) and (1-f_{sat}) centrals (from the central-only mock). Satellite fractions between 20% and 30% account for the clustering signal on both small and large scales; see Figure 4.6. All the selections above are done on the halo mass defined as M_{200}, which correspond to an overdensity threshold of $\Delta_{200} = 200\rho_c$ [241], where ρ_c is the critical density of the Universe.

To produce a mock catalog, we randomly select halos from the light-cone according to the probability distribution P, defined in Eq. 4.4, until the ELG redshift distribution $n(z)$ in Figure 4.2 is achieved. We then construct a grid of mocks by selecting M_{mean} in the range

$10^{11.2} - 10^{12.7}\,h^{-1}\mathrm{M}_\odot$, σ_M between the values $M_{mean}/[1., 2., 4.]\ h^{-1}\mathrm{M}_\odot$ (the sampling space is three times larger), and the satellite fraction in the interval $0 < f_{sat} < 0.5$, to obtain predictions for both ξ^s and $w(\theta)$. Finally, we compare these model predictions with our measurements by computing a combined χ^2 on scales $2 < s < 22\,h^{-1}$Mpc for the monopole, and $0.002° < \theta < 0.55°$ for the angular clustering, as follows:

$$\chi^2 = \frac{N_\xi \chi^2_\xi + N_w \chi^2_{w(\theta)}}{N_\xi + N_w}, \tag{4.5}$$

where

$$\chi^2_{w(\theta)} = \frac{1}{N_w} \sum_i^{N_w} \frac{|w_{observed}(\theta_i) - w_{halos}(\theta_i)|^2}{\sigma^2(w_{observed}(\theta_i))}, \tag{4.6}$$

and

$$\chi^2_\xi = \frac{1}{N_\xi} \sum_i^{N_\xi} \frac{|\xi_{observed}(s_i) - \xi_{halos}(s_i)|^2}{\sigma^2(\xi_{observed}(s_i))}. \tag{4.7}$$

The possible models accounting for the ELG clustering are degenerate with respect to the mean halo mass and the satellite fraction. In fact, Figure 4.6 shows that a plethora of $(\log M_{mean}, f_{sat})$ models fit the data: from $(11.3, 0.45)$ by $(12, 0.2)$ to $(12.5, 0)$. Given the 41 degrees of freedom we have, we consider acceptable those models with $\chi^2 < 1.25$. Models with a higher χ^2 value are rejected at the 90% level.

The combination with the weak lensing results breaks this degeneracy and rules out the higher- and lower-mass models. However, among these latter, there is one with $\chi^2 = 1$ and parameters: $\log M_{mean} = 12$, $\sigma_M = M_{mean}/2$, $f_{sat} = 22.5\%$ (star symbol in Figure 4.6). The angular and redshift-space correlation functions of this best-fit mock are displayed in Figure 4.3 (blue line), together with the ELG measurements. The weak lensing measurements are perfectly compatible with this best-fit model.

We provide our best-fit MDPL mock catalog to the ELG clustering measurements at http://projects.ift.uam-csic.es/skies-universes/.

4.6. Results and discussion

4.6.1 ELG clustering trends as a function of magnitude, flux, luminosity and stellar mass

We employ the complete VIPERS data sample at $z \sim 0.8$, which has about $30,000$ reliable redshifts in the range $0.6 < z < 1$, to investigate trends of the clustering amplitude (bias) with observed or rest frame broad band magnitude or emission line flux. To this purpose, we measure the emission line properties in the VIPERS spectra and find a significant [OII] flux in about two thirds of them; the rest does not show emission lines (Comparat et al., in prep.). We bin the data according to apparent and absolute magnitude, [OII] flux and luminosy, and measure the clustering in each sample (the binning scheme was set to contain between 9000 and $10,000$ data points). Figure 4.7 shows our ELG results in the observed (bottom row) and rest frame (top row). Consistently with previous analyses [e.g., 190, 205], we find that the brighter the selection in the i-band, either observed or rest-frame, the higher the bias. Analogously, the fainter the g-band limit, either observed or rest-frame, the higher the bias. The anti-correlation between [OII] flux and bias is only seen in the observed frame (the difference is ~ 1.4); in the rest frame it is not significant. It would be interesting to further investigate the correlation between [OII] luminosity and g-band magnitude in the small-scale clustering, but with the resolution of current data we are not able to push the analysis to scales $\sim 200h^{-1}$kpc, which is the typical virial radius of a halo of mass $10^{12}h^{-1}\mathrm{M}_{\odot}$. New data from eBOSS will be able to address this issue. The results above indicate that if we have a g-selected ELG sample and [OII] fluxes for a certain number of its galaxies, in order to maximize its clustering signal, we should select the ELGs with brighter i-band magnitudes.

To investigate the clustering dependence on stellar mass, we map the host halo masses for ELGs at $z \sim 0.8$, $M_h \sim 10^{12}h^{-1}\mathrm{M}_{\odot}$, onto stellar mass values using the stellar-to-halo-mass relation by [181], see their Figure 11. Our data are right before the “knee” at $M_\star \sim 3.5 \times 10^{10}h^{-1}\mathrm{M}_{\odot}$.

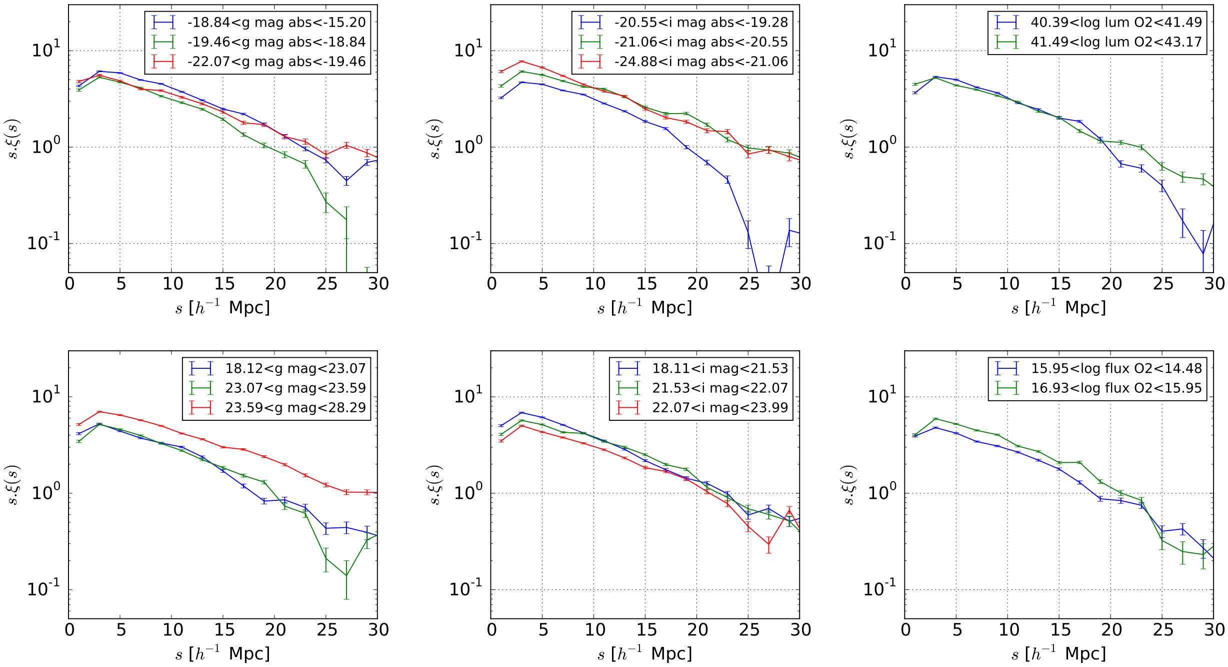

Figure 4.7: VIPERS clustering trends as a function of the $g-$band and $i-$band magnitudes (top row: rest frame; bottom row: observed frame), [OII] luminosity (top row) and [OII] flux (bottom row).

4.6.2 Star formation efficiency

From our analysis, the typical halo masses hosting ELGs at $z \sim 0.8$ are $M_h \sim (1 \pm 0.5) \times 10^{12}\,h^{-1}\mathrm{M}_{\odot}$, and (22.5±2.5)% of them are satellites belonging to a larger halo, whose central is a quiescent galaxy. Figure 4.8 provides a schematic representation of the possible ELG configurations. A total of 21.2% ELGs are single satellites belonging to a parent halo with mass $M_{hQ} \sim 2.5 \times 10^{13}\,h^{-1}\mathrm{M}_{\odot}$; only in 1.3% of the cases the parent halo hosts more than one satellite ELG. The maximum number of satellites, $n = 1.8$, is achieved in the highest-mass case, where $M_{hQ} \sim 6.8 \times 10^{13}\,h^{-1}\mathrm{M}_{\odot}$. These results imply that the mean number of ELG satellites is only slighlty larger than unity (~ 1.01). The quiescent galaxies at the center of the parent halos are not included in the sample, since the stellar masses for ELGs from the SHMR discussed above are too low for halos of $10^{13}\,h^{-1}\mathrm{M}_{\odot}$.

The typical masses for halos hosting ELGs suggest that we are sampling halos ($\sim 10^{12}\,h^{-1}\mathrm{M}_{\odot}$) that form stars in the most efficient way, according to the star formation rate discussed by

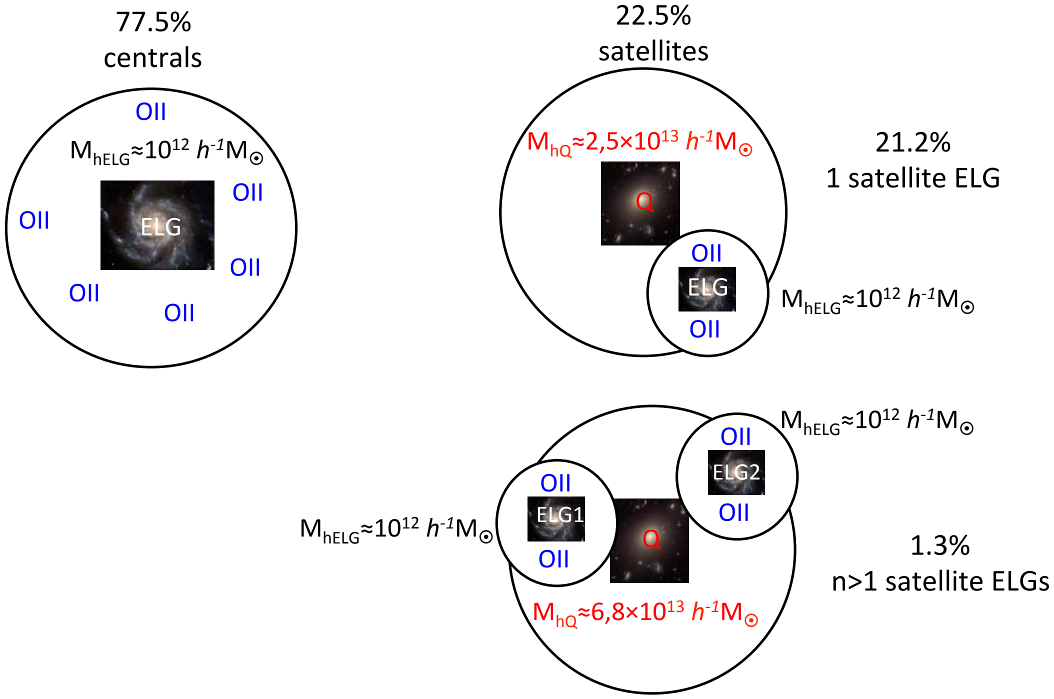
77.5%
centrals
OII
M_{hELG}≈10^{12} h^{-1}M$_\odot$
OII
OII
ELG
OII
OII
OII
22.5%
satellites
M_{hQ}≈2,5×10^{13} h^{-1}M$_\odot$
Q
21.2%
1 satellite ELG
OII
ELG
OII
M_{hELG}≈10^{12} h^{-1}M$_\odot$
M_{hELG}≈10^{12} h^{-1}M$_\odot$
OII
ELG2
OII
OII
ELG1
OII
Q
M_{hELG}≈10^{12} h^{-1}M$_\odot$
1.3%
n>1 satellite ELGs
M_{hQ}≈6,8×10^{13} h^{-1}M$_\odot$

Figure 4.8: Schematic diagram of possible ELG configurations. ELGs at $z \sim 0.8$ typically live in halos of mass $M_h \sim (1{\pm}0.5)\times 10^{12}\,h^{-1}\mathrm{M}_\odot$ and 22.5% are satellites belonging to larger halos, whose central galaxy is quiescent. Among these satellite configurations, 21.2% of parent halos with $M_{hQ} \sim 2.5\times10^{13}\,h^{-1}\mathrm{M}_\odot$ host one satellite ELG, and only 1.3% of parents host more than one satellite ELG. The maxium number of satellites, $n = 1.8$, is achieved in the highest-mass case, $M_{hQ} \sim 6.8\times10^{13}\,h^{-1}\mathrm{M}_\odot$. See the text for details.

Behroozi et al. [20] (see their Figure 1, bottom panel). This result opens a new science field and, hopefully, in the near future, integrated models combining N-body simulations with semi-analytic models (SAMs) will be able to probe star formation and shed some light on the correlations between [OII] flux and magnitude in the clustering of galaxies.

4.7. Summary

We have presented an analysis of the halo occupation distribution for emission-line galaxies, which jointly accounts for three measurements: the angular correlation function, the monopole, and the weak lensing signal around ELGs (see Section 4.4). Our procedure can be summarized in the following points:

- Apply the SUGAR [257] algorithm to the 11 snapshots available from the MDPL simulation to construct a light-cone (Section 4.3.2), with the same geometry and angular footprint of the ELG data.

- Modify the traditional SHAM technique (Section 4.5), to account for the ELG incompleteness, by selecting model galaxies by mass, until we match the observed ELG $n(z)$. In this way, our mock is constrained by the observed ELG redshift distribution, and represents a reliable model.

- Parametrize the probability of selecting a halo hosting a ELG with Eq. 4.4, in terms of the mean halo mass of the sample (M_{mean}), the dispersion around the mean (σ_M), and the satellite fraction (f_{sat}). The additional parameter "flag" enables to distinguish central and satellite halos.

- We perform two experiments (see Section 4.5) on the MDPL light-cone to derive information on which are the halo mass and satellite fraction ranges of values we need to input in our modified SHAM model to correctly fit the ELG clustering signal.

- Construct a grid of models based on these values, and jointly fit both angular and redshift-space clustering (see Section 4.5). Our best-fit models (see Figure 4.6) are degenerate with respect to M_{mean} and f_{sat}. The combination with the weak lensing analysis (see Section 3.1) breaks this degeneracy and rules out the highest and lowest mass models. Our best-fit ($\chi^2 = 1$) model is shown in Figure 4.3 together with the ELG measurements, and is given by $\log M_{mean} = 12$, $f_{sat} = 22.5\%$, $\sigma_M = M_{mean}/2$.

To conclude, we have built and released to the community a reliable galaxy mock catalog that correctly fits the clustering amplitude of the *ugri* ELG sample constructed by matching spectroscopic redshifts from BOSS DR12, VIPERS and DEEP2 (for details see Section 4.3). With these tools, we can begin building many realizations of the density field to predict errors on the BAO measurement.

The measured halo masses for halos hosting emission-line galaxies indicate that we are sampling the halos that form stars in the most efficient way, according the star formation rate discussed by Behroozi et al. [20] (see their Figure 1, bottom panel). This is an important point for the future, and opens the path to further studies to understand the correlation between clustering and the strength of emission lines. With the resolution available from current data, we are not able to push the analysis to the typical scales ($\sim 200h^{-1}$kpc) of halos of $10^{12}\,h^{-1}\mathrm{M}_{\odot}$; however, next-generation surveys (see Section 1.7), as eBOSS and DESI, will provide better resolution, and in the near future we should be able to build robust combinations of N-body simulations and SAMs that will address those questions.

5

Conclusions and future prospects

In this Ph.D. thesis, I have studied galaxy clustering in different samples of the SDSS and SDSS-III/BOSS surveys on small and intermediate scales (i.e. $r \lesssim 30\,h^{-1}\mathrm{Mpc}$). Specifically, I have measured the redshift-space two-point correlation functions – 3D $\xi(r_p, \pi)$, projected $w_p(r_p)$, angular $w(\theta)$, monopole $\xi_0(s)$ and quadrupole $\xi_2(s)$ – of these galaxies and modeled the results using the products of the MultiDark cosmological simulation to generate galaxy mock catalogs testing different approaches.

For red/blue SDSS-III/BOSS CMASS DR11 galaxies (see Chapter 3), I have applied an halo occupation distribution formalism to one single MultiDark snapshot at the mean redshift of the sample, $z \sim 0.53$, and generated galaxy mock catalogs able to reproduce the observed clustering as a function of color. The MultiDark simulation is particularly indicated to perform HOD modeling, since it includes both parent and sub-halos, thus the satellite mock galaxies can be placed randomly at the sub-halo positions. The standard

HOD prescription does not differentiate between galaxy colors, then the same mock can be either red or blue. I have circumvented this ambiguity by introducing an additional constraint that forces the MDPL model galaxies to match the observed CMASS red/blue galaxy fraction and, by consequence, it assigns each mock a specific color. I have also studied the impact of small-scale redshift-space distortions on the BOSS CMASS clustering through a straightforward two-parameter model which is able to disentangle the contribution of galaxy peculiar velocities, v_{pec}, causing the small-scale finger-of-god elongation, from the Kaiser squashing on larger scales. In agreement with several previous works [311, 333, 297], I find that bluer, star-forming galaxies have lower bias, lower v_{pec} values, lower clustering amplitude, and their host halos are less massive than their redder quenched counterparts.

For [OII] emission-line galaxies, both in SDSS at $z \sim 0.1$ and SDSS-III/BOSS at $z \sim 0.8$, I have adopted a (sub)halo abundance matching scheme and generated high-fidelity MultiDark light-cones using the SUrvey GenerAtoR algorithm developed by Rodríguez-Torres et al. [257]. The main difference of using a light-cone instead of a single MultiDark realization, is that the light-cone, by construction, includes the complete redshift evolution and is capable to mimic several volume effects – as the cosmic variance or the galaxy number density fluctuations due to the presence of voids or super clusters – that are present in the observations and a single simulation snapshot cannot emulate. A single MDPL realization, in fact, does not include evolution because it is at constant redshift and, compared to the light-cone, it is less affected by cosmic variance because its volume is larger. The small size of the light-cone volume represents the weakness of the method I have proposed. The maximum aperture achieved for the light-cone using a simulation with $V = 1\,h^{-3}\,\mathrm{Gpc}^3$ is only $\sim 0.02h^{-3}\,\mathrm{Gpc}^3$. This limitation implies that, on large scales (i.e. $r \geq 30\,h^{-1}\mathrm{Mpc}$), the clustering signal of the MDPL mock galaxies does not reproduce correctly the measurements. For this reason, I have focused the analysis on small and intermediate scales, below $30\,h^{-1}\mathrm{Mpc}$. To extend this clustering analysis to BAO scales ($150\,h^{-1}\mathrm{Mpc}$), one needs larger simulation volumes, as the 2.5Gpc BigMultiDark, but in that case the resolution will be lower.

Both HOD and SHAM models work only if the galaxy sample considered is complete, meaning that all its objects have been observed. Most of the time, however, this is not the case. SDSS and SDSS-III/BOSS emission-line galaxies, for instance, are very far from being complete both in terms of [OII] luminosity and stellar mass. To overcome this problem, I have modified the standard SHAM prescription by down-sampling the MultiDark light-cones to match the observed ELG number density, and accounting, in this way, for their incompleteness (see Chapters 2 and 4). I have characterized the galaxy halo occupation distribution model for ELGs in terms of two parameters: the satellite fraction, f_{sat}, and the mean host halo mass, M_h.

For the SDSS ELGs at $z \sim 0.1$, I have performed a clustering study (see Chapter 2) as a function of the [OII] luminosity and found a clear correlation between the amplitude of the 2PCF and the strength of the [OII] lines, with more luminous galaxies being more strongly clustered. [OII] emission-line galaxies at $z \sim 0.1$ live in halos with typical mass of $\sim 10^{12}\,h^{-1}\mathrm{M}_{\odot}$, and their satellite fraction varies between $\sim 18\%$ and $\sim 33\%$, and is lower for more luminous galaxies.

For SDSS-III/BOSS [OII] ELGs at $z \sim 0.8$ (see Chapter 4), I find a similar scenario: typically they live in halos with mean mass $M_h \sim 10^{12}\,h^{-1}$ $\mathrm{M}_{\odot}$, and 22.5% of them are satellites. In this case, I combine the clustering results with the weak-lensing measurement to reduce the degeneracy between the model parameters, $(M_{mean}, \sigma_M, f_{sat})$, and rule out most of the lower and higher mass models. I also investigate the clustering dependence on stellar mass by mapping the ELG host halo masses at $z \sim 0.8$ onto stellar mass values, using the stellar-to-halo-mass relation by Leauthaud et al. [181]. I find that typical ELG halo masses of $M_h \sim 10^{12}h^{-1}\mathrm{M}_{\odot}$ correspond to stellar masses of $M_\star \sim 3.5 \times 10^{10}h^{-1}\mathrm{M}_{\odot}$. According to the star formation rate discussed in Behroozi et al. [20] (see their Figure 1, bottom panel), I am sampling those halos that most efficiently form stars.

I have also characterized the 2PCF in the SDSS Main galaxy sample at $z \sim 0.1$ as a function of the r-band absolute magnitude. Consistently with previous works [311, 334, 122, 338] based on different model approaches, I find that more luminous galaxies are more strongly clustered than their fainter companions. Using the light-cone technique with SHAM assignment, I can correctly and accurately fit the observed SDSS clustering, reproducing its dependence on M_r. My satellite fraction predictions are overall higher than the HOD results by Guo et al. [122], and such a discrepancy is due to the different way of populating halos with galaxies in the SHAM and the HOD schemes. The SHAM prescription is applied by performing a cut (see Eq. 1.35) in the halo and galaxy number densities, which excludes any object below a certain V_{peak} and below the corresponding luminosity. The HOD formulation does not assume such a cut, and allows one to include any kind of halo. For this reason, compared to the SHAM recipe I use, Guo et al. [122] assign more satellites to more massive halos or, in other words, the SHAM cut excludes satellites with small V_{peak} values in more massive halos. In order to reproduce their satellite HOD prediction (i.e., number of satellites per halo mass), I therefore need to include satellite mocks with lower V_{peak} values than the ones originally assigned by the SHAM. This is exactly what my model does. By increasing f_{sat}, I assign additional satellites that will distribute over the whole mass range considered. The effect of the satellite enhancement in the 1-halo term of the clustering can be quantified as $\xi_{1h} \propto (N_{cen}\, N_{sat} + N_{sat}\, N_{sat})$ [200], meaning that increasing the number of satellites by m will result in a small-scale clustering enhancement of $\sim \mathcal{O}(m^2)$.

Another important difference between our methods is that I place the satellite mocks at the sub-halo positions provided in the MultiDark halo catalogs, while they draw random dark matter particles for the position of their satellites and apply the velocity bias correction [125] to mimic the peculiar velocity contribution. I take the v_{pec} values directly from the MDPL simulation.

By construction, the MDPL light-cones account for the evolution of the galaxy number density with redshift, $n(z)$, which is an effect naturally observed in the Universe. This implies that the $n(z)$ distribution of my mock galaxies fluctuates around the mean value of the single MDPL realization adopted by Zehavi et al. [334] and Guo et al. [122]. My MultiDark clustering predictions are in excellent agreement with the SDSS results, in particular with the monopole and quadrupole results. The latter carries the galaxy peculiar velocity information, which is directly linked to the f_{sat} value, and is responsible to enhance the small-scale clustering amplitude. The remarkable agreement I find in the quadrupole demonstrates that I am correctly modeling the number of satellite mocks, with no need of introducing any velocity bias correction to boost the small-scale 2PCF [125, 122]. The galaxy halo occupation distribution models I present here are robust because they naturally arise from the simulation products, by applying them a straightforward SHAM assignment, and allowing the satellite fraction to vary.

These results open the path to future studies of the correlation between galaxy clustering, bias, strength of emission lines, and star formation efficiency. Current data lack of resolution to push the analysis down to very small scales to resolve the smaller substructures. New-generation redshift surveys (see Section 1.7), as SDSS-IV/eBOSS (2014-2020), DESI (2018-2023) and EUCLID (2020-2025), will provide huge quantities of data with much better resolution and imaging quality, which will be crucial to address these questions. In the near future, it will be possible to improve the data interpretation by combining high-resolution light-cones extracted from the MultiDark cosmological simulations, as the MultiDark-Bologna Lensing factory [115], with accurate semi-analytic models for galaxy formation as Galacticus, SAGE, or the MultiDark Galaxies[1] project now under construction. In particular, the near-infrared EUCLID mission will target about 1 billion objects in one visible riz broad band (550-920nm) down to magnitude AB=24.5 [176, 177]. The forecast for the spectroscopic program is 25-50 million galaxies out to redshift $z \sim 2$, and their exact number will be limited by the Hα line flux. EUCLID will also deliver morphologies, masses and star-formation rates with four times better resolution and 3 NIR magnitudes deeper than possible from ground [176]. The high-resolution will be key to push the cluster-

[1] www.multidarkgalaxies.pbworks.com

ing study at very small scales, where correlations between sub-structures of the same halo become significant, and to do tomography of the mass distribution.

For the future, I plan to implement the Hα emission-line galaxy target selection function for EUCLID using, as pilot targets, the Hα emitters [11, 86] measured by the HST-WISP [10] mission in the redshift range $0.75 < z < 1.5$. WISP is a near-infrared slitless spectroscopic survey very similar to EUCLID, which collected data using the G102 (0.80-1.17μm) and G141 (1.11-1.67μm) grisms of the Wide Field Camera 3 of the Hubble Space Telescope. I plan to match the WISP spectroscopic sample to the photometric and ancillary data of GOODS-N[2] for the G102 and G141 grisms on the COSMOS[3] field. The latter provides photometry in 30 bands and about 70,000 accurate spectra over 25arcmin^2 in the redshift range $0 < z < 4$. I will build the Hα ELG target selection function in preparation to EUCLID on the WISP-COSMOS field and, for the first EUCLID data release, I will have all the tools in place to perform a full clustering analysis in terms of the Hα emission lines at $z \sim 2$. Using EUCLID Hα emitters, we will measure and model galaxy clustering as a function of the strength of the emission lines and the star formation rate with unprecedented accuracy. We will be able to dramatically improve the current constraints on large-scale bias and redshift-space distortions, and derive reliable and accurate estimates for the covariance matrices necessary to reduce the BAO error at the level of the systematics.

In an analogous way, we have already implemented BOSS-like and ELG selections in preparation to DESI using the Dark Energy Camera Legacy Survey[4] (DECaLS), a public high-quality imaging survey now under construction, designed to complement the SDSS, SDSS-III/BOSS and SDSS-IV/eBOSS spectroscopic database. It will image 6700deg^2 of the BOSS extragalactic footprint in the region $-20\,\mathrm{deg} < \delta < +30\,\mathrm{deg}$ using the optical grz filters, and the four WISE[5] fields. So far, about 12,000 DECaLS galaxies have been matched

[2] www.stsci.edu/science/goods/

[3] http://cosmos.astro.caltech.edu

[4] www.legacysurvey.com

[5] http://wise.ssl.berkeley.edu/

to the BOSS DR12 sample. In the near-future, DECaLS data will be complemented by the Low Redshift survey at Calar Alto [LoRCA; 62], which plans to spectroscopically observe about 200,000 galaxies at low redshift, $z < 0.2$, in the northern sky to contribute to the construction of robust galaxy samples with the best spectroscopy and photometry to date. These new data sets will be crucial to improve the baryon acoustic oscillation measurement, in the attempt to unveil the nature of dark energy, which seems to be responsible of the accelerating expansion of the Universe.

6

Conclusiones y planes futuros

En ésta tesis doctoral he estudiado el agrupamiento de galaxias en diferentes muestras de los surveys espectroscópicos SDSS y SDSS-III/BOSS, tanto a escalas pequeñas cuanto intermedias (i.e. $r \lesssim 30\,h^{-1}$ Mpc). En concreto, he medido la función de correlación de dos puntos de dichas galaxias en el espacio de redshift – 3D $\xi(r_p, \pi)$, proyectada $w_p(r_p)$, angular $w(\theta)$, monopolo $\xi_0(s)$ y cuadrupolo $\xi_2(s)$ – y modelado los resultados utilizando los productos de la simulación cosmológica MultiDark para generar catálogos mocks aplicando procedimientos diferentes.

Para las galaxias rojas y azules de la muestra SDSS-III/BOSS CMASS DR11 (Capítulo 3), he aplicado el formalismo del "halo occupation distribution" a una realización de MultiDark con redshift igual al valor medio de la muestra, $z \sim 0.53$, y he generado catálogos mocks de galaxias capaces de reproducir fielmente el clustering observado. La simulación MultiDark resulta particularmente indicada para los modelos HOD porque incluye halos distintos y sub-halos. Por lo tanto, las galaxias mocks satélites pueden ser colocadas aleatoriamente en las posiciones de los sub-halos. La prescripción HOD estándard no diferencia las galaxias por color, entonces la misma galaxia mock puede resultar tanto roja como azul. Para solventar esta ambigüedad, he introducido en mis modelos una condición adiccional que fuerce los mocks a reproducir la fracción de galaxias rojas y azules observada en CMASS, y por lo tanto, asigne a cada mock un color específico. También he estudiado el impacto en el clustering de las distorsiones presentes a pequeñas escalas en el espacio de redshift. A través de un modelo con dos parámetros, he separado la contribución de las velocidades peculiares de las galaxias, v_{pec}, responsables del efecto de elongación en el clustering conocido como "finger-of-god", de la compresión Kaiser a largas escalas.

De acuerdo con varios estudios anteriores [311, 333, 297], mis resultados demuestran que las galaxias azules con formación estelar en curso tienen menor bias, valores menores de v_{pec}, menor amplitud de clustering, y sus halos de materia oscura son menos masivos comparado con las galaxias luminosas rojas en las que la formación estelar ha cesado.

Para las galaxias con líneas de emisión [OII], tanto en SDSS con $z \sim 0.1$ cuanto en SDSS-III/BOSS con $z \sim 0.8$, he adoptado un modelo SHAM y construido light-cones MultiDark utilizando el algoritmo SUGAR desarrollado por Rodríguez-Torres et al. [257]. La diferencia principal entre usar un light-cone o una única realización MultiDark es que el light-cone, por construcción, incluye la evolución completa con el redshift y reproduce varios efectos de volumen que se observan en el Universo, como la variancia cósmica o las fluctuaciones de densidad debidas a la presencia de voids o super clusters, que una única caja MultiDark no puede emular. Una única realización MDPL no incluye evolución porque tiene un valor constante de corrimiento al rojo, y el efecto de variancia cósmica será reducido en comparación al light-cone porque el volumen es mucho mayor. El volumen pequeño del light-cone representa la desventaja del método propuesto. La máxima apertura que se puede lograr para el light-cone utilizando la simulación MultiDark con $V = 1\,h^{-3}$ Gpc^3 corresponde a un volumen de apenas $\sim 0.02\,h^{-3}$ Gpc^3. Esta limitación implica que, a largas escalas (i.e.

$r \geq 30\,h^{-1}$ Mpc), el clustering de los modelos MDPL no reproduzca tan fielmente los datos, y por este motivo, he limitado el estudio a $s \lesssim 30\,h^{-1}$Mpc. Para poder extender el análisis a la escala BAO (i.e. $150\,h^{-1}$ Mpc), es necesaria una simulación con un volumen mayor como la BigMultiDark con $L_{box} = 2.5\,h^{-1}$Gpc pero en este caso la resolución será menor.

Ambos métodos HOD y SHAM funcionan solamente cuando la muestra de galaxias considerada es completa, i.e. todas las galaxias han sido observadas. La mayoría del tiempo esto no sucede en astronomía. Por ejemplo, las muestras de galaxias con líneas de emisión de SDSS y SDSS-III/BOSS son incompletas tanto en términos de luminosidad [OII] como en masa estelar. Para solucionar este problema, he modificado la prescripción SHAM estándard reduciendo los light-cones MultiDark para que tengan la densidad de ELG observada e incluyan la incompletitud (Capítulos 2 y 4). De esta forma es posible caracterizar la distribución de ELGs en sus halos de materia oscura a través de dos parámetros: la fracción de satélites, f_{sat}, y la masa promedio de los halos, M_h.

Estudiando las galaxias SDSS con líneas de emisión [OII] en el Universo local a $z \sim 0.1$, he encontrado una clara correlación entre la amplitud de la 2PCF y la fuerza de la líneas [OII], con las galaxias más luminosas generalmente más agrupadas. Las ELGs a $z \sim 0.1$ viven en halos de masa típica $\sim 10^{12}\,h^{-1}\mathrm{M}_{\odot}$, y su fracción de satélites varía entre $\sim 18\%$ y $\sim 33\%$, y es menor para galaxias más luminosas.

Analizando las ELGs con líneas de emisión [OII] $z \sim 0.8$ en SDSS-III/BOSS (see Chapter 4), se ha observado una configuración parecida al caso del Universo local: las ELGs viven en halos de masa $M_h \sim 10^{12}\,h^{-1}$ $\mathrm{M}_{\odot}$, y 22.5% de éstas son satélites. En el caso a $z \sim 0.8$, he combinado las medidas de agrupamiento con la de weak-lensing para reducir la degeneración entre los parámetros del modelo, $(M_{mean}, \sigma_M, f_{sat})$. También he investigado la dependencia de las medidas de agrupamiento de la masa estelar poniendo en correspondencia las masas típicas de los halos para las ELGs con las masas estelares dadas por Leauthaud et al. [181]. Mis predicciones para las masas de los halos de las galaxias con líneas de emisión [OII] a $z \sim 0.8$ corresponden a masas estelares de $M_{\star} \sim 3.5 \times 10^{10} h^{-1}\mathrm{M}_{\odot}$. Según la tasa de formación estelar presentada por Behroozi et al. [20], estoy considerando aquellos halos donde la formación estelar es más eficiente.

He caracterizado también la 2PCF en la población de galaxias SDSS Main, con corrimiento al rojo $z \sim 0.1$, para estudiar el agrupamiento como función de la magnitud absolut en banda r. De acuerdo con trabajos anteriores [311, 334, 122, 338] basados en modelos distintos de los nuestros, mis resultados demuestran que las galaxias más luminosas son las más agrupadas. Generando light-cones y utilizando la técnica SHAM, he reproducido correctamente las medidas de 2PCF en SDSS y su dependencia de la luminosidad M_r.

Sin embargo, mis predicciones para la fracción de satélites son mayores comparadas con los resultados HOD de Guo et al. [122], y la discrepancia es debida al diferente método con el que poblamos los halos de materia oscura de galaxias observadas. La prescripción SHAM se aplica imponiendo un corte (ver Ecuación 1.35) en las densidades de halos y galaxias, para excluir objectos que tengan una velocidad V_{peak} y una luminosidad por debajo de cierto límite. En la formulación HOD este corte no existe, por lo tanto cualquier halo puede formar parte del catálogo virtual que se produce. Esta diferencia implica que Guo et al. [122] asigne un mayor número de satélites virtuales a los halos más masivos o, en otras palabras, el corte en el SHAM excluye satélites con valores pequeños de V_{peak} en los halos más masivos. Para poder reproducir las predicciones HOD para los satélites (i.e., el número de satélites en función de la masa del halo), necesito incluir satélites virtuales que tengan un valor de V_{peak} menor a lo que asignaría normalmente el SHAM. El modelo que he presentado hace exactamente ésto: aumentando f_{sat}, incluye satélites adiccionales que se distribuyen en todo el rango de masa considerado. El efecto de aumentar los satélites en el régimen de 1-halo del agrupamiento puede ser cuantificado como $\xi_{1h} \propto (N_{cen}\, N_{sat} + N_{sat}\, N_{sat})$ [200]. Esto significa que añadiendo m satélites, la amplitud de la función de correlación de dos puntos a pequeñas escalas aumenta de un factor $\sim \mathcal{O}(m^2)$.

Otra diferencia importante entre nuestros mtodos es que yo coloco los satélites virtuales en las posiciones de los sub-halos tomandolas, junto con las velocidades peculiares, directamente desde los catálogos MultiDark. Guo et al. [122] generan aleatoriamente partículas de materia oscura y utilizan sus coordenadas espaciales para posicionar los sub-halos. Para emular los valores de velocidades peculiares introducen una corrección conocida como “velocity bias” [125].

Por construcción, los light-cones incluyen la evolución de la distribución de galaxias $n(z)$ con el redshift, que es un efecto naturalmente observado en el Universo. Esto implica que la distribución $n(z)$ de mis galaxias virtuales fluctúa alrededor de los valores medios de la única realzación MultiDark utilizada en [334, 122]. Mis catálogos virtuales están en excelente acuerdo con las observaciones SDSS, en particular con el monopolo y el cuadrupolo de la función de correlación de dos puntos. El último incluye la información sobre las velocidades peculiares de las galaxias, que están relacionadas a los valores de f_{sat} y son responsables de aumentar la amplitud del clustering a pequeñas escalas (régimen de 1-halo). El notable acuerdo encontrado en las medidas de cuadrupolo confirma que he correctamente modelado la fracción de satélites en mis mocks, y ésto ha sido posible sin introducir ninguna corrección de “velocity bias” [125, 122] para incrementar el clustering a pequeñas escalas. Los modelos de ocupación de galaxias en halos de materia oscura que he presentado en esta tesis son sólidos porque derivan naturalmente de las simulaciones, aplicando el método SHAM a los productos MultiDark, sin necesidad de aportar modificaciones adiccionales.

Estos resultados abren el camino a estudios futuros sobre la correlación entre el clustering de galaxias, el bias, la fuerza de las líneas de emisión y la eficiencia en el proceso de formación estelar. Los datos disponibles actualmente carecen de resolución para poder extender el análisis a escalas muy pequeñas, y así poder resolver las sub-estructuras. Los surveys de nueva generación (§1.7), como SDSS-IV/eBOSS (2014-2020), DESI (2018-2023) y EUCLID (2020-2025), proporcionarán un gran número de datos con altísima resolución y calidad de imágenes que harán posible solucionar estos aspectos. En un futuro cercano será también posible mejorar la interpretación de los datos combinando light-cones de alta resolución construídos con los productos de la simulación MultiDark, como la MultiDark-Bologna Lensing factory [115], con modelos semi-analíticos de formación de galaxias, como Galacticus, SAGE o el proyecto MultiDark galaxias[1] ahora en construcción. En particular, la misión EUCLID detectará $\sim 10^9$ objetos en las bandas riz (550-920nm) hasta magnitud AB=24.5 [176, 177]. La previsión para el programa espectroscópico es de 25-50 millones de galaxias hasta redshift $z \sim 2$, y el número exacto será limitado por el flujo Hα. EUCLID proporcionará también morfologías, masas y tasas de formación estelar con resolución tres veces mejor y 3 magnitudes NIR más profundas que desde tierra [176]. La alta resolución será clave para hacer tomografía de la distribución de masa y extender el estudio del clustering a escalas muy pequeñas, donde las correlaciones entre sub-estructuras que pertenecen al mismo halo se hacen significativas.

Para el futuro planeo implementar la selección de galaxias con líneas de emisión Hα para EUCLID utilizando, como objetivos pilotos, las Hα ELGs [11, 86] detectadas por la misión espacial HST-WISP [10] en el rango de redshift $0.75 < z < 1.5$. WISP es un survey espectroscópico "slitless" en el cercano infrarrojo, con características muy similares a EUCLID, que tomó datos utilizando los "grisms" G102 (0.80-1.17μm) y G141 (1.11-1.67μm) de la Wide Field Camera 3 del Hubble Space Telescope. Pienso combinar la muestra espectroscópica de WISP con los datos fotométricos de GOODS-N[2] en los grisms G102 y G141 en el campo de COSMOS[3]. Este último survey proporciona medidas fotométricas en 30 bandas y aproximadamente 70,000 espectros con alta resolución en 25 arcmin^2 en el rango de redshift $0 < z < 4$. Mi objetivo, en preparación a EUCLID, es construir la función de selección para las galaxias con líneas de emisión Hα en el campo de WISP-COSMOS, y cuando EUCLID empezará a tomar datos, tendré todas las herramientas listas para desarrollar un estudio de agrupamiento con las Hα ELGs a alto redshift. Utilizando las Hα ELGs de EUCLID, será posible medir y modelar el clustering en términos de la fuerza de las líneas de emisión y la

[1] www.multidarkgalaxias.pbworks.com

[2] www.stsci.edu/science/goods/

[3] http://cosmos.astro.caltech.edu

tasa de formación estelar con una precisión sin precedentes. Combinando estas herramientas, lograremos mejorar radicalmente los límites actuales sobre el bias a larga escala y las distorsiones en el espacio del redshift. Podremos además derivar estimaciones precisas de las matrices de covariancia necesarias para reducir el error de las medidas BAO al nivel de los sistemáticos.

Analogamente, hemos implementado selecciones de galaxias similares a las muestras de BOSS o selecciones de ELGs en preparación a DESI, utilizando los datos del Dark Energy Camera Legacy Survey[4] (DECaLS), un survey público actualmente en construcción, que colecciona imágenes astronómicas de altísima calidad, diseñado para complementar las bases de datos espectroscópicos de SDSS, SDSS-III/BOSS y SDSS-IV/eBOSS. DECaLS escaneará 6700 deg^2 de la área de BOSS utilizando los filtros ópticos grz y los cuatro campos infrarrojos de WISE[5]. Hasta ahora aproximadamente 12,000 galaxias DECaLS tienen un correspondiente en BOSS DR12. En un futuro cercano, los datos de DECaLS serán complementados por el Low Redshift survey at Calar Alto [LoRCA; 62], que observará $\sim 200,000$ galaxias a bajo redshift, $z < 0.2$, en el emisferio norte, para contribuir a la construcción de muestras de galaxias con la mejor fotometría y espectroscopía posibles para estudios de agrupamiento. Esos datos serán determinantes para mejorar las medidas de las oscilaciones acústicas bariónicas, para entender la naturaleza de la energía oscura, responsable de la expansión accelerada de nuestro Universo.

[4] www.legacysurvey.com

[5] http://wise.ssl.berkeley.edu/

We shall not cease from exploration
and the end of all our exploring
will be to arrive where we started
and know the place for the first time.

T.S. Eliot - “Little Gidding”, Four Quartets

References

[1] Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., et al., 2009, ApJS, 182, 543

[2] Abel T., Norman M. L., Madau P., 1999, ApJ, 523, 66

[3] Agertz O., Moore B., Stadel J., et al., 2007, MNRAS, 380, 963

[4] Ahn K., Shapiro P. R., 2005, MNRAS, 363, 1092

[5] Aihara H., Allende Prieto C., An D., et al., 2011, ApJS, 193, 29

[6] Alam S., Albareti F. D., Allende Prieto C., et al., 2015, ApJS, 219, 12

[7] Anderson L., Aubourg E., Bailey S., et al., 2012, MNRAS, 427, 3435

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

[9] Angulo R. E., Baugh C. M., Lacey C. G., 2008, MNRAS, 387, 921

[10] Atek H., Malkan M., McCarthy P., et al., 2010, ApJ, 723, 104

[11] Atek H., Siana B., Scarlata C., et al., 2011, ApJ, 743, 121

[12] Aubert D., Teyssier R., 2008, MNRAS, 387, 295

[13] Bahcall N. A., Hao L., Bode P., Dong F., 2004, ApJ, 603, 1

[14] Baldry I. K., Glazebrook K., Brinkmann J., et al., 2004, ApJ, 600, 681

[15] Baltz E. A., Marshall P., Oguri M., 2009, JCAP, 1, 15

[16] Barnes J., Efstathiou G., 1987, ApJ, 319, 575

[17] Bartelmann M., 1999, in Evolution of Large Scale Structure : From Recombination to Garching, edited by A. J. Banday, R. K. Sheth, L. N. da Costa, 213

[18] Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

[19] Baugh C. M., Lacey C. G., Cole S., Frenk C. S., 1999, in The Most Distant Radio Galaxies, edited by H. J. A. Röttgering, P. N. Best, M. D. Lehnert, 265

[20] Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJL, 762, L31

[21] Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013, ApJ, 763, 18

[22] Bell E. F., Wolf C., Meisenheimer K., et al., 2004, ApJ, 608, 752

[23] Belli S., Jones T., Ellis R. S., Richard J., 2013, ApJ, 772, 141

[24] Bennett C. L., Halpern M., Hinshaw G., et al., 2003, ApJS, 148, 1

[25] Benson A. J., 2012, 17, 175

[26] Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002, MNRAS, 333, 156

[27] Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001, MNRAS, 320, 261

[28] Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

[29] Bertschinger E., 1998, ARA&A, 36, 599

[30] Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215

[31] Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

[32] Blaizot J., Wadadekar Y., Guiderdoni B., et al., 2005, MNRAS, 360, 159

[33] Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143

[34] Blanton M. R., Eisenstein D. J., Hogg D. W., et al., 2003, in American Astronomical Society Meeting Abstracts, vol. 36 of Bulletin of the American Astronomical Society, 589

[35] Blanton M. R., Hogg D. W., Bahcall N. A., et al., 2003, ApJ, 594, 186

[36] Blanton M. R., Hogg D. W., Bahcall N. A., et al., 2003, ApJ, 592, 819

[37] Blanton M. R., Lin H., Lupton R. H., et al., 2003, AJ, 125, 2276

[38] Blanton M. R., Roweis S., 2007, AJ, 133, 734

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

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

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

[42] Bolton A. S., Schlegel D. J., Aubourg É., et al., 2012, AJ, 144, 144

[43] Booth C. M., Schaye J., 2009, in American Institute of Physics Conference Series, edited by S. Heinz, E. Wilcots, vol. 1201 of American Institute of Physics Conference Series, 21–24

[44] Brammer G. B., Whitaker K. E., van Dokkum P. G., et al., 2009, ApJL, 706, L173

[45] Bray A. D., Eisenstein D. J., Skibba R. A., et al., 2015, ArXiv e-prints: 1502.01348

[46] Brinchmann J., Pettini M., Charlot S., 2008, MNRAS, 385, 769

[47] Bullock J. S., Dekel A., Kolatt T. S., et al., 2001, ApJ, 555, 240

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

[49] Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

[50] Carretero J., Castander F. J., Gaztañaga E., Crocce M., Fosalba P., 2015, MNRAS, 447, 646

[51] Chuang C.-H., Prada F., Cuesta A. J., et al., 2013, MNRAS, 433, 3559

[52] Ciardi B., Ferrara A., Marri S., Raimondo G., 2001, MNRAS, 324, 381

[53] Ciotti L., Lanzoni B., Volonteri M., 2007, ApJ, 658, 65

[54] Cohn J. D., White M., 2008, MNRAS, 385, 2025

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

[56] Cole S., 1991, ApJ, 367, 45

[57] Cole S., Benson A., Baugh C., Lacey C., Frenk C., 2000, in Clustering at High Redshift, edited by A. Mazure, O. Le Fèvre, V. Le Brun, vol. 200 of Astronomical Society of the Pacific Conference Series, 109

[58] Cole S., Lacey C., 1996, MNRAS, 281, 716

[59] Cole S., Norberg P., Baugh C. M., et al., 2001, MNRAS, 326, 255

[60] Colín P., Avila-Reese V., Valenzuela O., Firmani C., 2002, ApJ, 581, 777

[61] Colless M., Dalton G., Maddox S., et al., 2001, MNRAS, 328, 1039

[62] Comparat J., Chuang C.-H., Rodríguez-Torres S., et al., 2016, MNRAS

[63] Comparat J., Jullo E., Kneib J.-P., et al., 2013, MNRAS, 433, 1146

[64] Comparat J., Richard J., Kneib J.-P., et al., 2015, A&A, 575, A40

[65] Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

[66] Cooper M. C., Newman J. A., Weiner B. J., et al., 2008, MNRAS, 383, 1058

[67] Cooray A., 2005, MNRAS, 363, 337

[68] Cooray A., Huterer D., Baumann D., 2004, PhRvD, 69, 2, 027301

[69] Cooray A., Sheth R., 2002, PhysRep, 372, 1

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

[71] Coupon J., Ilbert O., Kilbinger M., et al., 2009, A&A, 500, 981

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

[73] Cox T. J., Jonsson P., Primack J. R., Somerville R. S., 2006, MNRAS, 373, 1013

[74] Croom S. M., Boyle B. J., Shanks T., et al., 2005, MNRAS, 356, 415

[75] Croton D. J., Stevens A. R. H., Tonini C., et al., 2016, SAGE: Semi-Analytic Galaxy Evolution, Astrophysics Source Code Library

[76] Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12

[77] Davé R., 2003, ArXiv Astrophysics e-prints

[78] Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

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

[80] Dawson K. S., Kneib J.-P., Percival W. J., et al., 2016, AJ, 151, 44

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

[82] de Jong R. S., Bellido-Tirado O., Chiappini C., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, vol. 8446, 84460T

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

[84] Dodelson S., 2003, Modern cosmology

[85] Dolag K., Borgani S., Schindler S., Diaferio A., Bykov A. M., 2008, 134, 229

[86] Domínguez A., Siana B., Henry A. L., et al., 2013, ApJ, 763, 145

[87] Dressler A., 1980, ApJ, 236, 351

[88] Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715

[89] Efstathiou G., Moody S., Peacock J. A., et al., 2002, MNRAS, 330, L29

[90] Eisenstein D. J., Annis J., Gunn J. E., et al., 2001, AJ, 122, 2267

[91] Eisenstein D. J., Weinberg D. H., Agol E., et al., 2011, AJ, 142, 72

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

[93] Eke V. R., Cole S., Frenk C. S., Navarro J. F., 1996, MNRAS, 281, 703

[94] Erb D. K., Pettini M., Shapley A. E., Steidel C. C., Law D. R., Reddy N. A., 2010, ApJ, 719, 1168

[95] Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813

[96] Erben T., Hildebrandt H., Miller L., et al., 2013, MNRAS, 433, 2545

[97] Faber S. M., Willmer C. N. A., Wolf C., et al., 2007, ApJ, 665, 265

[98] Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

[99] Favole G., Comparat J., Prada F., et al., 2015, ArXiv e-prints: 1507.04356

[100] Favole G., McBride C. K., Eisenstein D. J., et al., 2015, ArXiv e-prints: 1506.02044

[101] Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23

[102] Finlator K., Özel F., Davé R., 2009, MNRAS, 393, 1090

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

[104] Franx M., Labbé I., Rudnick G., et al., 2003, ApJL, 587, L79

[105] Freedman W. L., Madore B. F., Gibson B. K., et al., 2001, ApJ, 553, 47

[106] Fryxell B., Olson K., Ricker P., et al., 2000, ApJS, 131, 273

[107] Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748

[108] Gallagher J. S., Hunter D. A., Bushouse H., 1989, AJ, 97, 700

[109] Gao L., Navarro J. F., Cole S., et al., 2008, MNRAS, 387, 536

[110] Garilli B., Guzzo L., Scodeggio M., et al., 2014, A&A, 562, A23

[111] Gaztañaga E., Scoccimarro R., 2005, MNRAS, 361, 824

[112] Genzel R., Burkert A., Bouché N., et al., 2008, ApJ, 687, 59

[113] Gilbank D. G., Baldry I. K., Balogh M. L., Glazebrook K., Bower R. G., 2010, MNRAS, 405, 2594

[114] Gillis B. R., Hudson M. J., Erben T., et al., 2013, MNRAS, 431, 1439

[115] Giocoli C., Jullo E., Metcalf R. B., et al., 2015, ArXiv e-prints: 1511.08211

[116] Gnedin N. Y., Abel T., 2001, 6, 437

[117] Goldhaber G., Perlmutter S., 1998, PhysRep, 307, 325

[118] Gunawardhana M. L. P., Hopkins A. M., Bland-Hawthorn J., et al., 2013, MNRAS, 433, 2764

[119] Gunn J. E., Carr M., Rockosi C., et al., 1998, AJ, 116, 3040

[120] Gunn J. E., Siegmund W. A., Mannery E. J., et al., 2006, AJ, 131, 2332

[121] Guo H., Zehavi I., Zheng Z., et al., 2012, ArXiv e-prints: 1212.1211

[122] Guo H., Zheng Z., Jing Y. P., et al., 2015, MNRAS, 449, L95

[123] Guo H., Zheng Z., Zehavi I., et al., 2014, MNRAS, 441, 2398

[124] Guo H., Zheng Z., Zehavi I., et al., 2015, ArXiv e-prints: 1505.07861

[125] Guo H., Zheng Z., Zehavi I., et al., 2015, MNRAS, 446, 578

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

[127] Hainline K. N., Shapley A. E., Kornei K. A., et al., 2009, ApJ, 701, 52

[128] Hamilton A. J. S., 1992, ApJL, 385, L5

[129] Hamilton A. J. S., 1998, in The Evolving Universe, edited by D. Hamilton, vol. 231 of Astrophysics and Space Science Library, 185

[130] Hamilton A. J. S., 2001, MNRAS, 322, 419

[131] Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399

[132] Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75

[133] Hawkins E., Maddox S., Cole S., et al., 2003, MNRAS, 346, 78

[134] Hearin A. P., Watson D. F., 2013, MNRAS, 435, 1313

[135] Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, MNRAS, 444, 729

[136] Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2015, ArXiv e-prints: 1512.03050

[137] Helly J. C., Cole S., Frenk C. S., et al., 2003, MNRAS, 338, 913

[138] Hemmati S., Mobasher B., Darvish B., Nayyeri H., Sobral D., Miller S., 2015, ApJ, 814, 46

[139] Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, MNRAS, 396, 535

[140] Heymans C., Van Waerbeke L., Miller L., et al., 2012, MNRAS, 427, 146

[141] Hippelein H., Maier C., Meisenheimer K., et al., 2003, A&A, 402, 65

[142] Hogg D. W., 1999, ArXiv astro-ph/9905116

[143] Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, ArXiv astro-ph/0210394

[144] Hogg D. W., Blanton M. R., Eisenstein D. J., et al., 2003, ApJL, 585, L5

[145] Hogg D. W., Cohen J. G., Blandford R., Pahre M. A., 1998, ApJ, 504, 622

[146] Hopkins A. M., 2004, ApJ, 615, 209

[147] Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142

[148] Hopkins A. M., Miller C. J., Nichol R. C., et al., 2003, ApJ, 599, 971

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

[150] Jackson J. C., 1972, MNRAS, 156, 1P

[151] Jenkins A., Frenk C. S., Pearce F. R., et al., 1998, ApJ, 499, 20

[152] Kaiser N., 1987, MNRAS, 227, 1

[153] Kauffmann G., Heckman T. M., White S. D. M., et al., 2003, MNRAS, 341, 33

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

[155] Kauffmann G., White S. D. M., Heckman T. M., et al., 2004, MNRAS, 353, 713

[156] Kay S. T., Pearce F. R., Frenk C. S., Jenkins A., 2002, MNRAS, 330, 113

[157] Kennicutt R., 1992, in Star Formation in Stellar Systems, edited by G. Tenorio-Tagle, M. Prieto, F. Sanchez, 191

[158] Kennicutt Jr. R. C., 1998, ARA&A, 36, 189

[159] Kerscher M., 1999, A&A, 343, 333

[160] Kewley L. J., Dopita M. A., Leitherer C., et al., 2013, ApJ, 774, 100

[161] Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2

[162] Klypin A., Holtzman J., 1997, ArXiv astro-ph/9712217

[163] Klypin A., Holtzman J., Primack J., Regos E., 1993, ApJ, 416, 1

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

[165] Klypin A., Prada F., Yepes G., Hess S., Gottlober S., 2013, ArXiv e-prints: 1310.3740

[166] Klypin A., Trujillo-Gomez S., Primack J., 2010, ArXiv e-prints, arXiv:1002.3660

[167] Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

[168] Komatsu E., Dunkley J., Nolta M. R., et al., 2009, ApJS, 180, 330

[169] Kravtsov A., 2002, in APS April Meeting Abstracts

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

[171] Kravtsov A. V., Borgani S., 2012, ARA&A, 50, 353

[172] Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, ApJS, 111, 73

[173] Kuhlen M., Diemand J., Madau P., Zemp M., 2008, Journal of Physics Conference Series, 125, 1, 012008

[174] Labbé I., Huang J., Franx M., et al., 2005, ApJL, 624, L81

[175] Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

[176] Laureijs R., Amiaux J., Arduini S., et al., 2011, ArXiv e-prints: 1110.3193

[177] Laureijs R., Gondoin P., Duvet L., et al., 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, vol. 8442, 84420T

[178] Laursen P., Razoumov A. O., Sommer-Larsen J., 2009, ApJ, 696, 853

[179] Law D. R., Steidel C. C., Erb D. K., et al., 2009, ApJ, 697, 2057

[180] Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, ApJ, 738, 45

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

[182] Lemson G., Kauffmann G., 1999, MNRAS, 302, 111

[183] Li Y., Mo H. J., Gao L., 2008, MNRAS, 389, 1419

[184] Lin Y.-T., Mandelbaum R., Huang Y.-H., et al., 2016, ApJ, 819, 119

[185] Liu X., Shapley A. E., Coil A. L., Brinchmann J., Ma C.-P., 2008, ApJ, 678, 758

[186] Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388

[187] Marri S., White S. D. M., 2003, MNRAS, 345, 561

[188] Martin D. C., GALEX Science Team, 2005, in American Astronomical Society Meeting Abstracts, vol. 37 of Bulletin of the American Astronomical Society, 1235

[189] Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution

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

[191] Masjedi M., Hogg D. W., Cool R. J., et al., 2006, ApJ, 644, 54

[192] Masters D., McCarthy P., Siana B., et al., 2014, ApJ, 785, 153

[193] Masters K. L., Maraston C., Nichol R. C., et al., 2011, MNRAS, 418, 1055

[194] Matsubara T., Suto Y., 1996, ApJL, 470, L1

[195] Merritt D., Navarro J. F., Ludlow A., Jenkins A., 2005, ApJL, 624, L85

[196] Mihalas D., Binney J., 1981, Galactic astronomy: Structure and kinematics

[197] Miller J. S., 1974, Scientific American, 231, 34

[198] Miller N. A., Owen F. N., 2002, AJ, 124, 2453

[199] Miller R. G., 1974, *Biometrika*, 61, 1

[200] Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

[201] Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189

[202] Montero-Dorta A. D., Bolton A. S., Brownstein J. R., et al., 2014, ArXiv e-prints: 1410.5854

[203] Montero-Dorta A. D., Prada F., 2009, MNRAS, 399, 1106

[204] Moore B., Ghigna S., Governato F., et al., 1999, ApJL, 524, L19

[205] Mostek N., Coil A. L., Cooper M., Davis M., Newman J. A., Weiner B. J., 2013, ApJ, 767, 89

[206] Mouhcine M., Lewis I., Jones B., Lamareille F., Maddox S. J., Contini T., 2005, MNRAS, 362, 1143

[207] Navarro J. F., 2004, in Dark Matter in Galaxies, edited by S. Ryder, D. Pisano, M. Walker, K. Freeman, vol. 220 of IAU Symposium, 61

[208] Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

[209] Nelson D., Pillepich A., Genel S., et al., 2015, Astronomy and Computing, 13, 12

[210] Neto A. F., Gao L., Bett P., et al., 2007, MNRAS, 381, 1450

[211] Newman J. A., Cooper M. C., Davis M., et al., 2013, ApJS, 208, 5

[212] Newman S. F., Buschkamp P., Genzel R., et al., 2014, ApJ, 781, 21

[213] Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19

[214] Norberg P., Gaztañaga E., Baugh C. M., Croton D. J., 2011, MNRAS, 418, 2435

[215] Nuza S. E., Sánchez A. G., Prada F., et al., 2013, MNRAS, 432, 743

[216] Oemler A., 1974, The systematic properties of clusters of galaxies, Ph.D. thesis, California Institute of Technology

[217] Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

[218] Oppenheimer B. D., Davé R., 2008, MNRAS, 387, 577

[219] O'Shea B. W., Bryan G., Bordner J., et al., 2004, ArXiv astro-ph/0403044

[220] Osterbrock D. E., 1976, PASP, 88, 589

[221] Osterbrock D. E., Ferland G. J., 2006, Astrophysics of gaseous nebulae and active galactic nuclei

[222] Osterbrock D. E., Koski A. T., Phillips M. M., 1975, ApJL, 197, L41

[223] Padilla N. D., Baugh C. M., 2002, MNRAS, 329, 431

[224] Peacock J., 1999, Cosmological Physics, Cambridge Astrophysics, Cambridge University Press

[225] Peacock J. A., Cole S., Norberg P., et al., 2001, Nature, 410, 169

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

[227] Peebles P. J. E., 1993, Principles of Physical Cosmology

[228] Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815

[229] Percival W. J., Baugh C. M., Bland-Hawthorn J., et al., 2001, MNRAS, 327, 1297

[230] Percival W. J., Ross A. J., Sánchez A. G., et al., 2014, MNRAS, 439, 2531

[231] Percival W. J., Sutherland W., Peacock J. A., et al., 2002, MNRAS, 337, 1068

[232] Perlmutter S., Aldering G., della Valle M., et al., 1998, Nature, 391, 51

[233] Perlmutter S., Aldering G., Goldhaber G., et al., 1999, ApJ, 517, 565

[234] Petkova M., Springel V., 2009, MNRAS, 396, 1383

[235] Pettini M., Shapley A. E., Steidel C. C., et al., 2001, ApJ, 554, 981

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

[237] Planck Collaboration, Ade P. A. R., Aghanim N., et al., 2015, ArXiv e-prints: 1502.01589

[238] Planck Collaboration, Ade P. A. R., Aghanim N., et al., 2015, ArXiv e-prints: 1506.07135

[239] Plewa T., Müller E., 2001, Computer Physics Communications, 138, 101

[240] Prada F., Klypin A., Yepes G., Nuza S. E., Gottloeber S., 2011, ArXiv e-prints: 1111.2889

[241] Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

[242] Prada F., Klypin A. A., Simonneau E., et al., 2006, ApJ, 645, 1001

[243] Primack J. R., 1997, ArXiv astro-ph/9707285

[244] Primack J. R., 2015, ArXiv e-prints: 1505.02821

[245] Quenouille M. H., 1956, *Biometrika*, 43, 353

[246] Quilis V., 2004, MNRAS, 352, 1426

[247] Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2009, MNRAS, 394, 624

[248] Reid B. A., Samushia L., White M., et al., 2012, MNRAS, 426, 2719

[249] Reid B. A., Seo H.-J., Leauthaud A., Tinker J. L., White M., 2014, MNRAS, 444, 476

[250] Reid B. A., Spergel D. N., 2009, ApJ, 698, 143

[251] Reid B. A., Verde L., Dolag K., Matarrese S., Moscardini L., 2010, JCAP, 7, 013

[252] Ricker P. M., Dodelson S., Lamb D. Q., 2000, ApJ, 536, 122

[253] Riebe K., Partl A. M., Enke H., et al., 2011, ArXiv e-prints: 1109.000

[254] Riebe K., Partl A. M., Enke H., et al., 2013, Astronomische Nachrichten, 334, 691

[255] Riess A. G., Filippenko A. V., Challis P., et al., 1998, AJ, 116, 1009

[256] Rigby J. R., Wuyts E., Gladders M. D., Sharon K., Becker G. D., 2011, ApJ, 732, 59

[257] Rodríguez-Torres S. A., Chuang C.-H., Prada F., et al., 2015, ArXiv e-prints: 1509.06404

[258] Ross A. J., Percival W. J., Sánchez A. G., et al., 2012, MNRAS, 424, 564

[259] Ross A. J., Samushia L., Burden A., et al., 2014, MNRAS, 437, 1109

[260] Rosswog S., 2009, New A Rev., 53, 78

[261] Sánchez A. G., Scóccola C. G., Ross A. J., et al., 2012, MNRAS, 425, 415

[262] Sartoris B., Biviano A., Fedeli C., et al., 2015, ArXiv e-prints: 1505.02165

[263] Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2006, MNRAS, 371, 1125

[264] Schlegel D. J., Blum R. D., Castander F. J., et al., 2015, in American Astronomical Society Meeting Abstracts, vol. 225 of American Astronomical Society Meeting Abstracts

[265] Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

[266] Scoville N., Aussel H., Brusa M., et al., 2007, ApJS, 172, 1

[267] Shan H., Kneib J.-P., Li R., et al., 2015, ArXiv e-prints: 1502.00313

[268] Shapley A. E., Coil A. L., Ma C.-P., Bundy K., 2005, ApJ, 635, 1006

[269] Shirazi M., Brinchmann J., Rahmati A., 2014, ApJ, 787, 120

[270] Silk J., Di Cintio A., Dvorkin I., 2013, ArXiv e-prints: 1312.0107

[271] Simon P., 2013, A&A, 560, A33

[272] Skibba R. A., Coil A. L., Mendez A. J., et al., 2015, ApJ, 807, 152

[273] Slosar A., Font-Ribera A., Pieri M. M., et al., 2011, JCAP, 9, 1

[274] Smee S. A., Gunn J. E., Golebiowski M., et al., 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 9147 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2

[275] Smee S. A., Gunn J. E., Uomoto A., et al., 2013, AJ, 146, 32

[276] Smoot G. F., Bennett C. L., Kogut A., et al., 1992, ApJL, 396, L1

[277] Sobral D., Best P. N., Matsuda Y., Smail I., Geach J. E., Cirasuolo M., 2012, MNRAS, 420, 1926

[278] Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

[279] Spergel D., 2005, in American Astronomical Society Meeting Abstracts, vol. 37 of Bulletin of the American Astronomical Society, 1402

[280] Spergel D. N., Verde L., Peiris H. V., et al., 2003, ApJS, 148, 175

[281] Spinrad H., 2005, Galaxy Formation and Evolution

[282] Springel V., 2005, MNRAS, 364, 1105

[283] Springel V., Frenk C. S., White S. D. M., 2006, Nature, 440, 1137

[284] Springel V., Wang J., Vogelsberger M., et al., 2008, MNRAS, 391, 1685

[285] Springel V., White S. D. M., Jenkins A., et al., 2005, Nature, 435, 629

[286] Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., Pettini M., 1998, ArXiv astro-ph/9812167

[287] Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJL, 462, L17

[288] Stiavelli M., Scarlata C., Panagia N., Treu T., Bertin G., Bertola F., 2001, ApJL, 561, L37

[289] Stinson G. S., Kaufmann T., Quinn T., Christensen C., Wadsley J., Kazantzidis S., 2006, in American Astronomical Society Meeting Abstracts, vol. 38 of Bulletin of the American Astronomical Society, 956

[290] Stoughton C., Lupton R. H., Bernardi M., et al., 2002, AJ, 123, 485

[291] Strateva I., Ivezić Ž., Knapp G. R., et al., 2001, AJ, 122, 1861

[292] Strauss M. A., Weinberg D. H., Lupton R. H., et al., 2002, AJ, 124, 1810

[293] Stringer M. J., Brooks A. M., Benson A. J., Governato F., 2010, MNRAS, 407, 632

[294] Sugai H., Karoji H., Takato N., et al., 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8446 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0

[295] Sunayama T., Hearin A., Padmanabhan N., Leauthaud A., 2016, in American Astronomical Society Meeting Abstracts, vol. 227 of American Astronomical Society Meeting Abstracts, 407.02

[296] Sunayama T., Hearin A. P., Padmanabhan N., Leauthaud A., 2016, MNRAS, 458, 1510

[297] Swanson M. E. C., Tegmark M., Blanton M., Zehavi I., 2008, MNRAS, 385, 1635

[298] Tasca L. A. M., Le Fèvre O., Hathi N. P., et al., 2015, A&A, 581, A54

[299] Tegmark M., Blanton M. R., Strauss M. A., et al., 2004, ApJ, 606, 702

[300] Thacker R. J., Couchman H. M. P., 2000, ApJ, 545, 728

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

[302] Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16

[303] Turkey J., 1958, The Annals of Mathematical Statistics, 29, 1

[304] van den Bosch F. C., Abel T., Croft R. A. C., Hernquist L., White S. D. M., 2002, ApJ, 576, 21

[305] van Dokkum P. G., Franx M., Förster Schreiber N. M., et al., 2004, ApJ, 611, 703

[306] van Waerbeke L., 1998, in Wide Field Surveys in Cosmology, edited by S. Colombi, Y. Mellier, B. Raban, 189

[307] Velander M., Kuijken K., Schrabback T., 2011, MNRAS, 412, 2665

[308] Verde L., Heavens A. F., Percival W. J., et al., 2002, MNRAS, 335, 432

[309] Vogelsberger M., Genel S., Springel V., et al., 2014, MNRAS, 444, 1518

[310] Wadsley J. W., Stadel J., Quinn T., 2004, 9, 137

[311] Wang Y., Yang X., Mo H. J., van den Bosch F. C., 2007, ApJ, 664, 608

[312] Warren M. S., Zurek W. H., Quinn P. J., Salmon J. K., 1992, in American Astronomical Society Meeting Abstracts, vol. 24 of Bulletin of the American Astronomical Society, 1247

[313] Watson D. F., Hearin A. P., Berlind A. A., et al., 2015, MNRAS, 446, 651

[314] Weedman D. W., 1986, Quasar astronomy

[315] Weinberg D., 1997, in Dark and Visible Matter in Galaxies and Cosmological Implications, edited by M. Persic, P. Salucci, vol. 117 of Astronomical Society of the Pacific Conference Series, 578

[316] Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, PhysRep, 530, 87

[317] Weinberg S., 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity

[318] White M., Blanton M., Bolton A., et al., 2011, ApJ, 728, 126

[319] White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594

[320] White S., 2004, in KITP Conference: Galaxy-Intergalactic Medium Interactions

[321] White S. D. M., Davis M., Efstathiou G., Frenk C. S., 1987, Nature, 330, 451

[322] White S. D. M., Davis M., Frenk C. S., 1984, MNRAS, 209, 27P

[323] White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

[324] White S. D. M., Frenk C. S., Davis M., Efstathiou G., 1987, ApJ, 313, 505

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

[326] Wright C. O., Brainerd T. G., 2000, ApJ, 534, 34

[327] Wu H.-Y., Rozo E., Wechsler R. H., 2008, ApJ, 688, 729

[328] Wyder T. K., Martin D. C., Schiminovich D., et al., 2007, ApJS, 173, 293

[329] York D. G., Adelman J., Anderson Jr. J. E., et al., 2000, AJ, 120, 1579

[330] Zehavi I., Blanton M. R., Frieman J. A., et al., 2002, ApJ, 571, 172

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

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

[333] Zehavi I., Zheng Z., Weinberg D. H., et al., 2005, ApJ, 630, 1

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

[335] Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

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

[337] Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

[338] Zu Y., Mandelbaum R., 2015, MNRAS, 454, 1161