Universitat Autònoma de Barcelona

Cosmology with galaxy surveys: how galaxies trace mass at large scales

Arnau Pujol Vallribera

Cosmology with galaxy surveys: how galaxies trace mass at large scales

Arnau Pujol Vallribera

Institut de Ciències de l'Espai (ICE, IEEC/CSIC)

Advisor: Enrique Gaztañaga Balbás

Tutor: Emili Bagan Capella

Universitat Autònoma de Barcelona
Departament de Física

A thesis submitted for the degree of

Doctor of Philosophy in Physics

Barcelona, May of 2016

Acknowledgements

En primer lloc, voldria donar les gràcies a l'Enrique Gaztañaga per la supervisió de la meva tesi, per la dedicació que ha posat i per com l'ha posat. Fer la tesi doctoral amb l'Enrique ha estat un plaer, pel que he après, per com m'ha guiat, per la seva visió de la ciència i perquè es treballa molt bé amb ell. Moltes gràcies, Enrique.

I also want to give special thanks to Chihway Chang, Kai Hoffmann, Noelia Jiménez and Ramin Skibba for our collaborations and our time spent working together, discussing our projects and learning together. Also thanks to Adam Amara, Alexandre Refregier, Alexander Knebe, Frazer Pearce, Carlton Baugh, Ravi Sheth, Roman Scoccimarro and Marc Manera for multiple and useful discussions about the projects of this thesis, and for the same reason to the members of the cosmology group at ICE, specially Santi Serrano, Jorge Carretero, Martin Crocce, Pablo Fosalba and Francisco Castander. També voldria agraïr a la Guadalupe la gestió dels múltiples viatges a congressos, que m'ha facilitat molt la vida, i la seva paciència davant del fet que mai he portat tots els documents necessaris a la primera.

Hi ha moltes coses que han estat importants per acabar fent aquesta tesi. Però si alguna cosa m'ha portat a estudiar física i fer recerca en cosmologia ha estat la meva fascinació pel món, i les ganes d'entendre com funciona i per què. Per això, vull donar les gràcies als meus pares per haver-me transmès la passió per la vida i el que m'envolta, per haver-me ensenyat que el món és fantàstic, fascinant i per haver-me encomanat les ganes d'aprendre i d'entendre el perquè de les coses. I per ser uns pares collonuts.

Finalment, vull agraïr a tota aquella gent que li ha donat a aquesta etapa meva un gran dia a dia. A l'humor del Josep, les matinades del Kai, els cafès del Ferran, el pesao de Manuel, els debats polítics amb el Daniel, a los chistes de Graná tan malos de Fran, a l'Àlex, l'Antonia, la Marina Meteorita, la Gemma i molta més gent que m'estic deixant... L'Hilda! M'estic deixant l'Hilda! Con lo que me he metido contigo como para olvidarme de ti. Y a Jacobo, por ser

como eres, el puto amo. Claramente faltan Jacobos en el mundo. Al Daniele, per les nostres reflexions sobre cooperació i sobre els camins que t'obre la vida. No et desvïis del camí, que vas pel bo! Al Joni, per les nostres hores al cotxe, als partits de futbol, al cyber, per les nostres tornades a barcelona corrent, per la gran persona i companyia que ets. A Laura, por darme por el saco cada día, sin descanso, porque siempre estás aquí para quejarte de mi y por lo mucho que te echaré de menos cuando deje de tenerte en la oficina. Eres la mejor pesada que conozco. A l'Izard, el meu aprenent de runner que se m'ha fet gran. Per les nostres converses de runners, el nostre mapa de postdocs, els nostres workshops romàntics, i perquè et dec el postdoc. A la Núria, per posar-me de bon humor cada dia, per les tortus, per la hackejada que em vas fer a l'ordinador i per les múltiples pèrdues de papers. Cosines així no es descobreixen cada dia. A Carmen, por nuestras charlas de viajes, de fotos para fondos de escritorio, por tu alegría y la pasión que transmites. Y por los minions, ahí lo petaste. Y por Camboya! A Marina Stabilo por Camboya también, y por nuestras charlas de comida (y hacerme descubrir las sopas), cooperación y muchas otras. Al Meteorito, per ensenyar-me a enfocar la vida amb més gràcia, a que tot pot ser un experiment interessant, a que sempre és més avorrit dir "no" que "i per què no?". Al Padu, per les nostres converses de numpy, les tornades en cotxe per la meridiana, i les incomptables converses frikis sobre temazos als cafès de les tardes. I veure't cantant Juego de Tronos al Catan no té preu. I a l'Anna. Per cada dia, per cada cuina caiguda, pels esmorzars, per tot el que hem viscut no només durant la tesi, per tot el que aprenc de tu, pels viatges, perquè sempre hi ets. Per tot.

Abstract

Galaxy surveys are an important tool for cosmology. The distribution of galaxies allow us to study the formation of structures and their evolution, which are needed ingredients to study the evolution and content of the Universe. However, according to the standard model of cosmology, the so-called ΛCDM model, most of the matter is made of dark matter, which gravitates but does not interact with light. Hence, the galaxies that we observe from our telescopes only represent a small fraction of the total mass of the Universe. Because of this, we need to understand the connection between galaxies and dark matter in order to infer the total mass distribution of the Universe from galaxy surveys.

Simulations are an important tool to predict the structure formation and evolution of dark matter and galaxy formation. In simulations dark matter is also usually characterized by dark matter haloes, which correspond to gravitationally collapsed overdensities of dark matter. Simulations allow us to study the impact of different cosmologies and galaxy formation models on the final large scale structures that galaxies and matter form. Simulations are also useful to calibrate our tools before applying them to real surveys.

At large scales, galaxies trace the matter distribution. In particular, the galaxy density fluctuations at large scales are proportional to the underlying matter fluctuations by a factor that is called *galaxy bias*. This factor allows us to infer the total matter distribution from the distribution of galaxies, and hence knowledge of galaxy bias has a very important impact on our cosmological studies. This PhD thesis is focused on the study of galaxy and halo bias at large scales.

There are several techniques to study galaxy bias, in this thesis we focus on two of them. The first technique uses the fact that galaxy bias can be modelled from a galaxy formation model. One of the most common models is the Halo Occupation Distribution (HOD) model, that assumes that galaxies populate dark matter haloes depending only on the halo mass. With this hypothesis, and assuming a halo bias model, we can relate galaxy clustering with matter

clustering and halo occupation. However, this hypothesis is not always accurate enough. We use the Millennium Simulation to study galaxy and halo bias, the halo mass dependence of halo bias, and its effects on galaxy bias prediction. We find that he occupation of galaxies in haloes does not only depend on mass, and assuming so causes an error in the galaxy bias predictions. We also study the environmental dependence of halo bias, and we show that environment constrains much more bias than mass. When a galaxy sample is selected by properties that are correlated with environment, the assumption that halo bias only depends on mass fails. We show that in these cases using the environmental dependence of halo bias produces a much better prediction of galaxy bias. We finally suggest to use information of environment and not only halo mass to constrain bias in galaxy surveys.

Another technique to study galaxy bias is by using weak gravitational lensing to directly measure mass in observations. Weak gravitational lensing is the field that studies the weak image distortions of galaxies due to the light deflections produced by the presence of a foreground mass distribution. Theses distortions can then be used to infer the total mass (baryonic and dark) distribution at large scales. We develop and study a new method to measure galaxy bias from the combination of weak lensing and galaxy density fields. The method consists on reconstructing the weak lensing maps from the distribution of the foreground galaxies. Bias is then measured from the correlations between the reconstructed and real weak lensing fields. We test the different systematics of the method and the regimes where this method is consistent with other methods to measure linear bias. We find that we can measure galaxy bias using this technique. This method is a good complement to other methods to measure bias because it uses different assumptions. Together the different techniques will allow to constrain better bias and cosmology in future galaxy surveys.

Resum

Els cartografiats galàctics són una eina important per la cosmologia. La distribució de galàxies permet estudiar la formació d'estructures i la seva evolució, que són ingredients necessaris per a estudiar l'evolució i el contingut de l'univers. No obstant això, segons el model estàndard de cosmologia, l'anomenat model ΛCDM, la majoria de la matèria està en forma de matèria fosca, que no interacciona amb la llum. Per tant, les galàxies que observem des dels nostres telescopis són una petita fracció de la matèria total de l'univers. Per això és necessari entendre la connexió entre galàxies i matèria fosca per tal d'inferir la distribució de tota la matèria de l'univers a partir dels cartografiats galàctics.

Les simulacions són una eina important per a predir la formació i evolució de les estructures de matèria fosca i per estudiar la formació de galàxies. En les simulacions la matèria fosca sovint és caracteritzada per *halos*, que corresponen a sobredensitats de matèria fosca col·lapsades gravitacionalment. Les simulacions permeten estudiar l'impacte de diferents cosmologies i models de formació de galàxies en les estructures a gran escala finals que formen les galàxies i la matèria. Les simulacions també són útils per calibrar les nostres eines abans d'aplicar-les als cartografiats reals.

A gran escala, les galàxies tracen la distribució de matèria. En particular, les fluctuacions de densitat de galàxies a gran escala són proporcionals a les fluctuacions de matèria per un factor anomenat *bias* galàctic. Aquest factor permet inferir la distribució de matèria total a partir de la distribució de galàxies, i per tant el coneixement del bias galàctic té un impacte molt important en els nostres estudis cosmològics. Aquesta tesi doctoral està focalitzada en l'estudi del bias galàctic i el bias d'halos a grans escales.

Hi ha diferents tècniques per a estudiar el bias galàctic, en aquesta tesi ens focalitzem en dues d'elles. La primera tècnica utilitza el fet que el bias galàctic pot ser modelat a partir d'un model de formació de galàxies. Un dels models més comuns és l'anomendat *Halo Occupation Distribution* (HOD), o distribució d'ocupació en halos, que assumeix que les galàxies poblen halos de

matèria fosca només segons la massa dels halos. Amb aquesta hipòtesi, i assumint que el bias d'halos només depèn de la massa, podem relacionar l'agrupament de galàxies amb l'agrupament de matèria i l'ocupació d'halos. No obstant això, aquesta hipòtesi no sempre és suficientment precisa. Fem ús de la simulació Millennium per a estudiar el bias galàctic i d'halos, la dependència en la massa del bias d'halos i els seus efectes en les prediccions del bias galàctic. Trobem que l'ocupació de galàxies en halos no depèn només de la seva massa, i assumir això causa un error en la predicció del bias galàctic. També estudiem la dependència del bias d'halos en l'ambient, i mostrem que l'ambient restringeix molt més el bias que la massa. Quan un conjunt de galàxies és seleccionat per propietats que estan correlacionades amb l'ambient, l'assumpció de que el bias d'halos només depèn de la massa falla. Mostrem que en aquests casos utilitzant la dependència en l'ambient del bias d'halos produeix una predicció del bias galàctic molt més bona. Finalment, suggerim utilitzar informació sobre l'ambient i no només la massa d'halos per restringir el bias galàctic en cartografiats galàctics.

Una altra tècnica per estudiar el bias galàctic és utilitzant *Weak gravitational lensing* (referent a les lents gravitacionals dèbils) per mesurar directament la massa en observacions. *Weak gravitational lensing* és el camp que estudia les distorsions lleus en les imatges de les galàxies degut a la deflexió de la llum produïda per la distribució de matèria del davant de la galàxia. Aquestes distorsions permeten inferir la distribució a gran escala de la matèria total (bariònica i fosca). Desenvolupem i estudiem un nou mètode per mesurar el bias galàctic a partir de la combinació dels mapes de *weak lensing* i el camp de distribució de galàxies. El mètode consisteix en reconstruïr el mapa de *weak lensing* a partir de la distribució de les galàxies de davant del mapa. El bias és mesurat a partir de les correlacions entre el mapa de *weak lensing* reconstruït i el real. Testegem diferents sistemàtics del mètode i estudiem en quins règims el mètode és consistent amb altres mètodes per mesurar el bias lineal. Trobem que podem mesurar el bias galàctic utilitzant aquesta tècnica. Aquest mètode és un bon complement d'altres mètodes per mesurar el bias galàctic, perquè utilitza assumpcions diferents. Juntes, les diferents tècniques per mesurar el bias galàctic permetran restringir millor el bias galàctic i la cosmologia en els futurs cartografiats galàctics.

Contents

List of Figures

Chapter 1

Introduction

1.1 ΛCDM Cosmology

Our Universe offers plenty of mysteries and unanswered questions. What is Universe made of? How did the Universe begin? Which are the fundamental laws of physics? And why? For many years people have been studying cosmology in order to answer all these questions from the study of the Universe as a whole. Many discoveries have been done, such as the fact that we live in the Milky Way, that the Universe is plenty of galaxies and the accelerated expansion of the Universe.

Nowadays, the ΛCDM is the most accepted cosmological model, also known as the Standard Model of cosmology. It is based on the Einstein's theory of gravity and assumes the presence of dark matter and dark energy. According to the model, the Universe was originated from an explosion known as the Big Bang and a phase of exponential expansion called inflation. Inflation generated the primordial conditions, and the ΛCDM model describes the evolution of the Universe from these primordial conditions measured from the Cosmic Microwave Background (CMB) up to nowadays. This evolution depends on the laws of physics and the content of the Universe, which we can constrain from observations. The model is consistent with our observations, but requires the presence of the mysterious dark matter and dark energy, which nature is unknown.

In the last decades we have experienced important advances in cosmology. We discovered that the Universe is expanding and therefore is not stationary [107,136]. The discovery of the Cosmic Microwave Background (CMB) by Penzias and Wilson in 1964 and its confirmation by the Wilkinson Microwave Anisotropy Probe (WMAP) Satellite [16] allowed us to study the geometry and matter content of the Cosmos, and its density fluctuations at $380,000$ years after the Big Bang. This, together with our knowledge of General Relativity, and the cosmological principle, that states that Universe is homogeneous and isotropic at

large scales, allow us to describe the dynamics of our Universe, since the evolution of the Universe depends on its content and the laws of physics.

The ΛCDM model describes the evolution of the Universe throw the Friedmann equation, that in natural units is:

$$H^2(a) = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(a) - \frac{k}{a^2}. \tag{1.1}$$

In this equation, the scale factor a describes the size of the Universe, and H corresponds to the Hubble constant, that can be interpreted as the expansion rate of the Universe. By definition, $a = 1$ today, so at some different time t the size of the Universe is $a(t)$ times the actual size. The constant G is the gravitational constant, ρ refers to the energy density of the Universe, including all the components, and k is the curvature constant of the geometry of the Universe. Observations seem to indicate that $k \simeq 0$, meaning that Universe is flat or almost flat [16, 108, 131]. Then, equation (1.1) shows that the rate of expansion of the Universe depends on its energy density and the curvature. From the Friedmann equation we can derive the cosmic acceleration:

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}\rho(a)(1+3\omega) \tag{1.2}$$

where ω relates the pressure p with the energy density ρ as $p = \omega\rho$. We see that the cosmic acceleration depends on the energy density and pressure. The recent discovery of the present accelerated expansion of the Universe by the Nobel prizes Saul Perlmutter, Adam Riess and Brian Schmidt [169, 178] implied the addition into the energy content of the Universe of a cosmological constant Λ, also known as dark energy. If this dark energy has negative pressure, with $\omega = -1$, then it can cause an accelerated expansion. With this new element, the cosmic energy content can be decomposed as:

$$\rho(a) = \rho_m a^{-3} + \rho_r a^{-4} + \rho_\Lambda \tag{1.3}$$

where ρ_m, ρ_r and ρ_Λ are the energy density of matter (including dark and baryonic matter), radiation and dark energy at the present. Note that the energy density of dark energy is constant in time (or in the scale factor a), while matter and radiation densities are diluted as the universe expands.

Sometimes these components are shown in terms of the critical density ρ_c in the fol-

lowing way:

$$\Omega_m = \frac{\rho_m}{\rho_c} \tag{1.4}$$

$$\Omega_r = \frac{\rho_r}{\rho_c} \tag{1.5}$$

$$\Omega_k = -\frac{k}{H_0^2} \tag{1.6}$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2} \tag{1.7}$$

where

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \tag{1.8}$$

and H_0 is the Hubble expansion rate today. Using these terms, the Friedmann equation has the form:

$$H^2(a) = H_0^2[\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_k a^{-2} + \Omega_\Lambda] \tag{1.9}$$

whose redshift dependence, obtianed from the definition of the scale factor:

$$a = \frac{1}{1+z} \tag{1.10}$$

has the following expression:

$$H^2(z) = H_0^2[\Omega_m (1+z)^3 + \Omega_r (1+z)^4 + \Omega_k (1+z)^2 + \Omega_\Lambda] \tag{1.11}$$

In other words, the rate of expansion of the Universe at large scales depends on the amount of matter (dark and baryonic), the amount of radiation, the value of the cosmological constant Λ (the dark energy content), and the curvature. Then, the knowledge of these parameters (Ω_m, Ω_r, Ω_k and Ω_Λ) allows us to predict the evolution of the cosmic expansion. Or, on the other way, if we can measure the evolution of the Universe, these cosmological parameters can be constrained. But these parameters and cosmological models do not only affect to the expansion of the Universe, they also have an imprint on other properties and statistics such as the formation of structures, galaxy formation and clustering, the Cosmic Microwave Background anisotropies, the Barionic Accoustic Oscillations, or the gravitational lensing effects, to mention some examples [129]. Then, the study of all these observables is also crucial to infer the fundamental cosmological parameters, test the ΛCDM and constrain other possible cosmological models.

The actual observations indicate that our Universe is dominated by dark energy and dark matter. In Figure 1.1 we show a graphic of the fractions of dark matter, dark energy and baryonic matter in the Universe according to the last measurements from Planck [171]. We

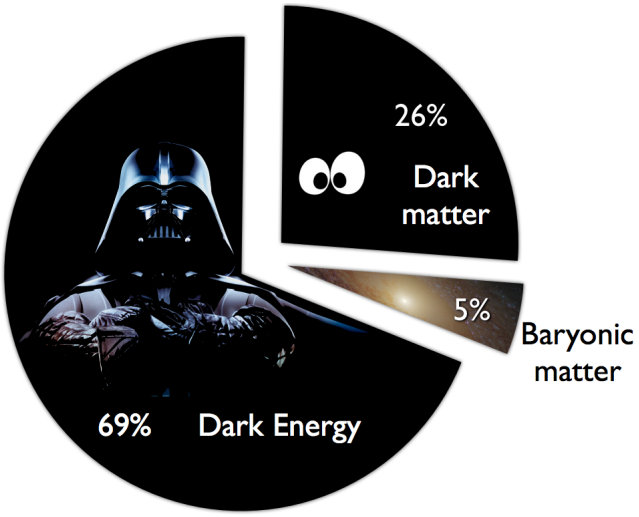

Figure 1.1: A graphic of the energy fractions of dark matter, dark energy and baryonic matter in the Universe according to Planck [171].

see that dark energy represents almost the 70% of the energy content, dark matter represents the 26% and baryonic matter, the matter of atoms, stars, planets etc., only represents the 5% of the total energy content. Note that the light of the galaxies that we observe in galaxy surveys only comes from this 5% of the content, which is only a 16% of the total matter of the Universe. Because of this, it is important to understand the relation between the light (galaxies) distribution and the matter (including dark matter) distribution in order to study our Universe from the observations of galaxies.

1.2 Large Scale Structure

1.2.1 Structure formation

According to the ΛCDM cosmology, at early times the Universe suffered an exponential expansion that is known as cosmological inflation. This expansion produced small quantum fluctuations in the matter density of the early Universe. After this, the Universe evolved and expanded according to the Friedmann equation, so equation (1.1). Due to this evolution, the small matter fluctuations started to form larger structures, producing the large scale structures that we see in the present Universe.

The matter density fluctuations are usually defined as:

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}, \tag{1.12}$$

where ρ is the matter density and $\bar{\rho}$ is the mean matter density of the Universe. When the matter density perturbations are small, $\delta \ll 1$, these perturbations grow according to linear

theory [165]:

$$\delta(\boldsymbol{x},t) = \delta(\boldsymbol{x},t_0)\frac{D(t)}{D(t_0)}, \tag{1.13}$$

where $D(t)$ represents the growth factor, that describes the growth of structures in the evolution of the Universe and depends on the scale and expansion rate of the Universe for a ΛCDM cosmological model.

Large-scale structures in the Universe arise through the gravitational clustering of matter. In the ΛCDM paradigm of hierarchical structure formation, gravitational evolution causes matter to cluster around peaks in the initial density field and collapse later into virialized objects. According to the Spherical Collapse model (e.g. [152, 167, 173]), when a perturbation exceeds the critical density $\delta_c = 1.868$, this overdensity collapses gravitationally to form a dark matter halo. This structure formation is hierarchical, in the sense that small structures form first and larger structures form later, either from a continuous mass accretion or from the merging of smaller haloes. As baryons are also affected by the gravitational potentials, they also fall and concentrate into the dark matter haloes. These systems provide the potential well and the conditions for galaxy formation [228]. It is therefore expected that the large-scale structures of galaxies are correlated with the ones of dark matter and haloes.

One of the important keys for cosmology is to understand the relation between the spatial distribution of galaxies and the underlying distribution of matter. Since galaxy formation depends on non-gravitational processes and complex baryonic physics, it is interesting to relate statistics of galaxies with the distribution of matter and haloes in order to parametrize their relation.

1.2.2 two-point correlation functions

One common way to study the distribution and clustering of galaxies is by measuring the two-point correlation function (2PCF), $\xi(r)$, which describes the amplitude of galaxy clustering as a function of scale. This quantity defines the excess of probability dP of finding a galaxy in a volume dV separated a distance r from another galaxy with respect to a random distribution:

$$dP = n(1+\xi(r))dV, \tag{1.14}$$

where n is the number density of galaxies [167]. There are several estimators of the 2PCF of a galaxy sample. An early estimator commonly used is the one from [64]:

$$\xi(r) = \frac{n_R}{n_D}\frac{DD(r)}{DR(r)} - 1, \tag{1.15}$$

where $DD(r)$ and $DR(r)$ are the number of pairs in the data (galaxies) sample and between the data and random points respectively separated a distance r, and n_D and n_R are the number density of data and random points respectively. In order to improve the statistical errors of the estimator, [98] proposed the following expression:

$$\xi(r) = \frac{DD(r)RR(r)}{(DR(r))^2} - 1, \tag{1.16}$$

where $RR(r)$ refers to the number of pairs of random points separated a distance r. Finally, the most commonly used estimator is the one from [130]:

$$\xi(r) = \left(\frac{n_R}{n_D}\right)^2 \frac{DD(r)}{RR(r)} - 2\frac{n_R}{n_D}\frac{DR(r)}{RR(r)} + 1, \tag{1.17}$$

which optimizes the statistical errors and the edge effects of the samples. If we want to focus on large scales, we can approximate the distribution of galaxies to the galaxy density field δ, obtained from the galaxy density fluctuations in space. In this case, the 2PCF becomes:

$$\xi(r_{12}) = \langle \delta(\boldsymbol{r_1})\delta(\boldsymbol{r_2})\rangle, \tag{1.18}$$

where $\delta(\boldsymbol{r}_{1,2})$ refers to the density fluctuation at the position $\boldsymbol{r}_{1,2}$, and r_{12} corresponds to the separation between $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$. This expression is consistent with the previous ones when r_{12} is large compared to the scales used to compute δ (where we would not be sensitive to the small separations between galaxies in the same δ).

In Fig. 1.2 we show the 2PCF of the dark matter field in the Millennium Simulation [206] as a function of scale. It has been calculated from equation (1.18), calculating δ in cubic grids of $500/256\,h^{-1}$ Mpc of side. We see a high amplitude at small scales, which means that dark matter tends to cluster and form structures at these scales. The smaller the distance, the stonger gravity acts to concentrate matter, and the bigger are the structures with respect to a uniform (or random) distribution. At large scales the 2PCF goes to zero, that would correspond to a uniform distribution. The fact that the Universe at large scales is homogeneous implies that the 2PCF goes to zero at large scales.

1.2.3 Galaxy bias

If galaxies and clusters are formed from the initial high-density regions of the Universe, then their clustering is an amplification of the clustering of the underlying dark matter

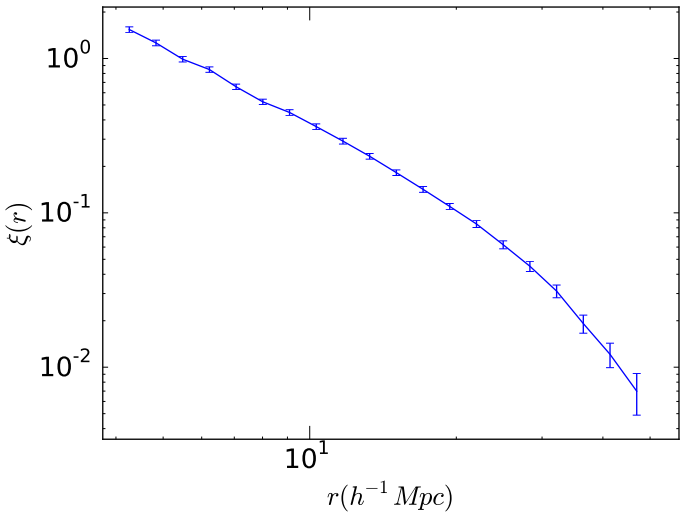

Figure 1.2: Dark matter 2PCF from the Millennium Simulation as a function of scale.

field [116]. In other words, the galaxies trace the matter distribution, and the 2PCF of galaxies is proportional to the correlation function of dark matter. This is often expressed as:

$$\xi_g(r) = b^2(r)\xi(r) \tag{1.19}$$

where $\xi_g(r)$ is the galaxy 2PCFS, $\xi(r)$ is the matter 2PCF and b is called the "bias factor". This bias factor allows us to translate the galaxy (or another tracer) distribution into the matter distribution, which is important to constrain our cosmological models from observations. Indeed, the knowledge of the bias factor produces a significant improvement on the precision of our measurements of cosmological parameters [87]. However, galaxy bias depends on many aspects, such as the scale, the galaxy properties (luminosity, colour...), the sample selection, redshift, etc.

[83] studied the relation between galaxy or halo distribution and dark matter distribution by assuming a local and non-linear relation between the density fluctuations of galaxies or haloes and the density fluctuations of dark matter. In the local bias model, the density fluctuation of galaxies (or haloes) is considered a function of the matter fluctuation at the same position:

$$\delta_h(\boldsymbol{r}) = f[\delta_m(\boldsymbol{r})] \tag{1.20}$$

where $\delta_h(\boldsymbol{r})$ and $\delta_m(\boldsymbol{r})$ are the density fluctuations at the position $\boldsymbol{r}$ of the galaxies (or haloes) and matter respectively, and $f[\delta]$ can be any function. As explained in [83, 184], this expression assumes no non-local dependencies that might come from some properties

as the local velocity field, or derivatives of the local gravitational potential. In that case, for $\delta_{h,m} \ll 1$, the equation (1.20) can be expanded in a Taylor series as:

$$\delta_h(\boldsymbol{r}) = f[\delta_m(\boldsymbol{r})] = \sum_{k=0}^{\infty} \frac{b_k}{k!}\delta_m^k(\boldsymbol{r}) = b_0 + b_1\delta_m(\boldsymbol{r}) + \frac{b_2}{2}\delta_m^2(\boldsymbol{r}) + \cdots \tag{1.21}$$

where $b_0 = 0$ since it is restricted to give $\langle\delta_h\rangle = \langle\delta_m\rangle = 0$, by definition. Then, the correlation function can be described in the following form:

$$\begin{aligned}
\xi_h(x_1,x_2) &= \langle\delta_h(x_1)\delta_h(x_2)\rangle = \sum_{j,k=0}^{\infty} \frac{b_j b_k}{j!k!}\langle\delta_m(x_1)^j\delta_m(x_2)^k\rangle & (1.22)\\
&= b_0 b_1(\langle\delta_m(x_1)\rangle + \langle\delta_m(x_2)\rangle) + b_1^2\langle\delta_m(x_1)\delta_m(x_2)\rangle + O(\delta^3) & (1.23)\\
&= b_1^2\langle\delta_m(x_1)\delta_m(x_2)\rangle + O(\delta^3). & (1.24)
\end{aligned}$$

The step form equation (1.23) to equation (1.24) is trivial because $\langle\delta\rangle = 0$ for every density fluctuation by definition. As it is shown in [88], if we are in the regime where $\delta_m \ll 1$, we can approximate equation (1.24) to:

$$\xi_h(r) \simeq b_1^2\langle\delta_m(x_1)\delta_m(x_2)\rangle = b_1^2\xi_m(r), \tag{1.25}$$

which is the definition of bias given in equation (1.19). Equation (1.25) shows that in the linear regime b_1 is the bias factor. If we define δ_h at large enough scales, where the density fluctuations are small, the local bias factor is consistent with the bias from equation (1.19). [142] showed that the local bias shows this consistency for scales larger than $R > 30 - 60Mpc\ h^{-1}$. In this linear and local bias regime, the bias factor does not depend on scale and can be considered as a constant factor:

$$\xi_g(r) = b^2\xi(r) \tag{1.26}$$

that can also be defined as:

$$\delta_g = b\delta_m. \tag{1.27}$$

This linear bias factor can be seen as a parametrization of the clustering amplitude of galaxies with respect to the dark matter. This constant factor depends on how galaxies occupy the dark matter field and depend on many properties and conditions. As galaxy clusters are generally formed in the highest density peaks of the dark matter density field, they present a high bias. In general, for a Gaussian field with initial mass fluctuations, the density peaks that first form show a higher clustering amplitude [8]. As massive galaxies and haloes are generally formed in these peaks, they also present a higher bias. Bias also depends on redshift, since at high redshift the first galaxies formed only in the highest density peaks, and because of this bias increases with redshift [82, 211].

1.2.4 Halo model and Halo Occupation Distribution

The halo model [19,55,133,166,187,188,194,195,201,209,231] is one of the most common approaches used to interpret and model the distribution and clustering of galaxies. According to this model, the dark matter field is characterized by dark matter haloes, and these haloes, although they can have different size and mass, present a universal profile [158]. With this, the dark matter 2PCF can be split into two components, the 1- and 2-halo terms:

$$\xi(r) = \xi_{1h}(r) + \xi_{2h}(r) \tag{1.28}$$

The 1-halo term of the 2PCF (ξ_{1h}) is obtained from the contribution of the matter correlation inside the same halo, which basically depends on the universal halo matter profile assumed. The 2-halo term (ξ_{2h}) contains the contribution of the correlation of matter located in different haloes, that can be predicted from linear perturbation theory. While the 1-halo term of the 2PCF is the dominant term at small scales (the characteristic scales of the haloes), the large-scale 2PCF is dominated by the 2-halo term. The large-scale clustering of dark matter haloes is biased with respect to dark matter, and halo bias is usually assumed to be a function of mass [152, 194, 218].

According to the halo model, galaxies are formed inside dark matter haloes. Hence, the 2PCF of galaxies depends on how the galaxies are populated in these haloes. The Halo Occupation Distribution (HOD) model [17, 19, 113, 187, 188, 236] is based on the halo model and describes the galaxy population in terms of central galaxies, located at the centre of haloes, and satellite galaxies, distributed in the haloes assuming some radial profile. With this, the 2PCF of galaxies can be parametrized from the occupation of galaxies in haloes. Again, the 2PCF have contributions from the 1-halo term, that comes from the galaxy pairs that are occupying the same haloes, and the 2-halo term, that comes from the pairs of galaxies that occupy different haloes. The most common approach is to describe the HOD as the probability that a halo of a given mass M hosts N galaxies, $P(N|M)$. The expected galaxy occupation can be parametrized in terms of central and satellite galaxies as:

$$\langle N(M) \rangle = \langle N_c(M) \rangle + \langle N_s(M) \rangle, \tag{1.29}$$

where N refers to the total number of galaxies per halo of mass M and N_c and N_s refer to the number of central and satellite galaxies per halo respectively. As the occupation is assumed to depend on M, $\langle N(M) \rangle$ can be parametrized in order to connect galaxy clustering with galaxy occupation using a set of free parameters. The most common parametrization is the following [233, 234, 236]:

$$\langle N(M) \rangle = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log M - \log M_{min}}{\sigma_{\log M}}\right)\right]\left[1 + \left(\frac{M - M_0}{M_1}\right)^{\alpha}\right], \tag{1.30}$$

where erf is the error function $\text{erf}(x) = \frac{2}{\pi}\int_0^x e^{-t^2}dt$ and $\sigma_{\log M}$, M_0, M_1, α and M_{min} are free parameters. M_{min} characterizes the minimum halo mass that hosts a central galaxy. M_0 and M_1 constrain the halo mass needed to be populated by satellite galaxies, and α describes how quickly the number of satellite galaxies increases with halo mass. These parameters, then, can be fitted from observations in order to learn about galaxy formation and the occupation of galaxies in haloes [57, 233].

1.3 Weak Gravitational Lensing

According to General Relativity, space and time are curved due to the presence of matter and energy. As the trajectories of the bodies (including light) follow the geodesics of space-time, light is deflected when it propagates through an inhomogeneous gravitational field. As a result of this, the images of distant galaxies that we observe are distorted due to the presence of mass in the trajectory that light follows from the source galaxies to the observer. The field that studies these distortions is called *gravitational lensing*. See [9, 157, 177] for a review of this field.

The strength of this effect depends on several aspects. First of all, it depends on the matter distribution between the source galaxies and the observer. Strong gravitational potentials cause large distortions of the image. But the strength of the light deflection and its impact on the final image distortion also depends on the relative distances between the sources and these gravitational potentials (or *lenses*). Finally, the effect is stronger if sources and lenses are better aligned with the observer.

The strongest effects can be easily recognised as distortions in the images, usually in the form of arcs or rings around a foreground galaxy or galaxy cluster (that is acting as a lens), or in the form of multiple images of the same galaxy around a given lens galaxy or cluster. *Strong gravitational lensing* is the field that studies these large distortions. These effects can be used to infer the mass of the foreground cluster. As the light deflection is caused by the gravitational potential due to the presence of any kind of mass (either dark or baryonic), this can be used to measure the presence of not only baryonic matter, but also dark matter.

However, the majority of the lensing effects are much smaller. Galaxy images are in general distorted at the percent level, causing basically a small change on its luminosity and shape. In these cases we cannot consider galaxies individually, since the main contribution to the galaxy image comes from its intrinsic shape. But instead we can study them statistically. Assuming some intrinsic properties of the galaxies (e.g. that the galaxies are randomly aligned or that they present a specific intrinsic alignment) we can study the large

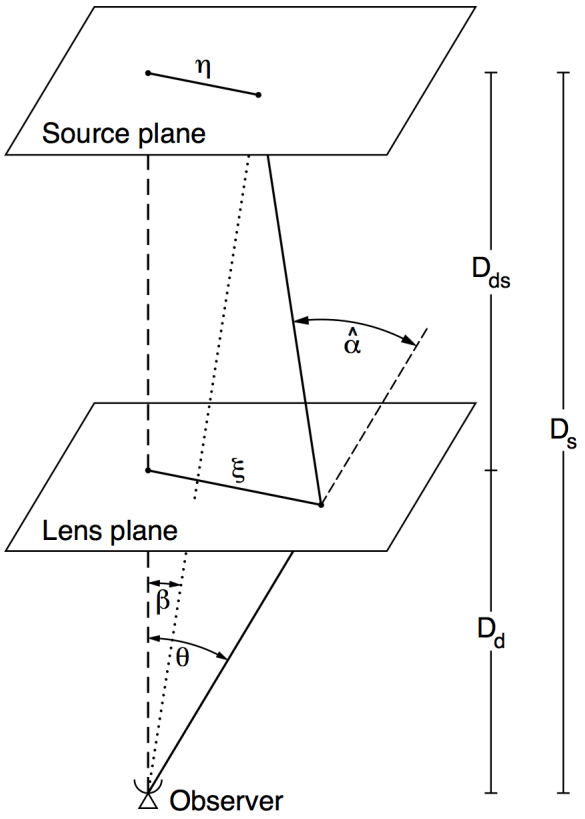

Figure 1.3: Scheme of light deflection due to gravitational lensing from [9].

scale distribution of matter. The field that studies the weak lensing distortions is called *weak gravitational lensing*, or simply *weak lensing*. The potential of this field of study depends on several statistical properties: the total area of the source galaxy samples, the number density of source galaxies, the scales that we study, intrinsic alignments, the shape noise measurements, etc.

A schematic picture of the light deflection due to the gravitational lensing effect is shown in Fig. 1.3, where we approximate the source and lens in two planes and the light trajectory as two straight lines with a deflection angle $\hat{\alpha}$. In the weak lensing regime, and in General Relativity, the angle deflection presents the following expression:

$$\hat{\alpha} = \frac{4GM}{c^2 \xi}, \tag{1.31}$$

where G is the gravitational constant, M is the mass of the lens centred in the line-of-sight and far from the light trajectory, c is the speed of light and ξ is the impact parameter specified by the distance between the position of the lens and the point where the light crosses the lens plane. When the lens is not a point in a plane but a continuous gravitational

field, $\hat{\alpha}$ is a more complicated vector function:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi}\int_{\Re^2} d^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}, \tag{1.32}$$

where $\boldsymbol{\xi} = D_d\boldsymbol{\theta}$ and κ is known as the convergence field and it is a projection of the foreground mass distribution, weighted by some geometrical factors:

$$\kappa(\boldsymbol{\theta},\chi_s) = \frac{3H_0^2\Omega_m}{2c^2}\int_0^{\chi_s} d\chi\frac{\chi(\chi_s-\chi)}{\chi_s}\frac{\delta(\boldsymbol{\theta},\chi)}{a(\chi)}, \tag{1.33}$$

where χ_s is the comoving distance to the source, $a(\chi)$ is the scale factor at the corresponding redshift in comoving distance χ, and $\delta(\boldsymbol{\theta},\chi)$ is the total (baryon and dark) mass density fluctuation. $\boldsymbol{\alpha}$ can also be expressed in terms of the lensing potential ψ:

$$\boldsymbol{\alpha} = \nabla\psi, \tag{1.34}$$

with

$$\psi = \frac{1}{\pi}\int_{\Re^2} d^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\ln|\boldsymbol{\theta}-\boldsymbol{\theta}'|. \tag{1.35}$$

With all this, the image of a source galaxy in the weak lensing regime can be described by the distortion matrix A:

$$A = (1-\kappa)\begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1-g_1 \end{pmatrix}, \tag{1.36}$$

where $g_{1,2}$ are the two components of the reduced shear $g = g_1 + ig_2$. The reduced shear can be related to the shear γ as:

$$g = \frac{\gamma}{1-\kappa}. \tag{1.37}$$

Finally, the shear can be related to the convergence κ through the lensing potential as:

$$\kappa = (\psi_{,11}+\psi_{,22})/2 \tag{1.38}$$

$$\gamma_1 = (\psi_{,11}-\psi_{,22})/2 \tag{1.39}$$

$$\gamma_2 = \psi_{,12} \tag{1.40}$$

with $\gamma = \gamma_1 + i\gamma_2$ and $\psi_{,ij} = \partial^2\psi/\partial\theta_i\partial\theta_j$.

These expressions give us a mathematical connection between the image distortions of the galaxies, the shear and the projected mass distribution in the foreground of the source galaxies. This allows to study statistics of the projected mass distribution in the Universe from the galaxy shears obtained from the images of distant galaxies [147, 223]. Weak lensing is also a good tool to constrain cosmology from observations [15, 140], and a very good complement to other statistics, since the systematics involved in weak lensing are very different compared to other comological observations.

1.4 Galaxy Surveys

Our knowledge of the large-scale structures of the Universe has increased during the last years thanks to the recent galaxy surveys. Surveys such as the 2-degree Field Galaxy Redshift Survey (2dFGRS) [50], the Sloan Digital Sky Survey (SDSS) [232] and the WiggleZ [29] supposed a step forward in the study of the distribution of galaxies in the Universe. The Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) allowed the study of weak lensing fields with unprecedented accuracy (e.g. [15, 105, 192]).

Present and up-coming surveys such as the Dark Energy Survey (DES) [75, 213], the Hyper SuprimeCam (HCS) [151], the Kilo Degree Survey (KIDS) [65], the Large Synoptic Survey Telescope (LSST) [139], the Euclid mission [132], the Dark Energy Spectroscopic Instrument (DESI) [137] and the Wide-Field Infrared Survey Telescope (WFIRST) [94] will offer the possibility to explore new fields and increase the precision of our comoslogical analyses.

Part of the work of this thesis has been implemented in DES. DES is a broad-band photometric survey designed to combine different analysis of type Ia Supernovae, Baryon Acoustic Oscillations, Weak Lensing and Galaxy cluster number counts in order to study the nature of Dark Energy and the accelerated expansion of the Universe. This survey consists on a 570 Megapixel camera (DECam) with 74 CCDs, located on the 4-m diameter Blanco Telescope, at Cerro Tololo, in Chile. The observing plan aims to cover 5000 $\deg^2$ of the sky in 525 nights. It uses the grizY filtering system and will record around 300 million galaxies with an absolute magnitude of $i < 24.5$.

1.5 Simulations

The large-scale structures that we see today in the Universe come from initial conditions that have evolved suffering complex processes involving nonlinear gravitational interactions, baryonic physics, gravitational collapses, merging of dark matter haloes, galaxy formation, etc. Although our theories can explain the fundamental laws of physics, we cannot predict analytically the evolution of such complex systems. However, the link between the early and almost uniform Universe and the final structures can be provided by numerical simulations. Simulations are an important tool to predict the structure formation of different cosmologies and theories, as well as to calibrate observations, test new tools to apply to observations and study the numerical effects of nonlinear and complex processes.

There has been a great progress in the development of simulations for the last decades. This progress consists on new codes for the numerical computation of structure formation

[36, 56, 70, 71, 120, 125, 150, 162, 168, 176, 204, 206, 207, 212], the increase in the volume of the simulations [6, 60, 78, 79], the decrease in the mass of the simulation particles [96, 205], the development of different algorithms to define dark matter haloes (see [121] and references therein) and subhaloes (see [161] and references therein), the application of different galaxy formation models (see [122] and references therein), the development of hydrodynamic simulations (e.g. [224]) and the current implementation of fast simulation codes (e.g. [143, 154, 155, 186, 210]). The advances in our simulations have an impact on the understanding of cosmological models and evolution and our precision in the analysis and measurements of observational data.

In this work we make use of two N-body simulations, the Millennium Simulation [206] and the MICE Grand Challenge [60, 78, 79] simulation.

The Millennium Simulation is a project from the Virgo Consortium[1], and at the time when it was run it was the largest N-body simulation of a ΛCDM cosmology ever done. It was run in the principal supercomputer at the Max Planck Society's Supercomputing Centre in Garching, Germany, and it consists of around 10^{10} particles in a cube of $500\,h^{-1}$ Mpc of side. Several galaxy catalogues are implemented in the simulation by following different Semi-Analytic Model (SAM) prescriptions, which are galaxy formation models based on analytic prescriptions of the baryonic physics in the process of galaxy formation.

The MICE Grand Challenge Simulation was run by the Institut de Ciències de l'Espai (ICE) at the Marenostrum Supercomuter in Barcelona. The output is a lightcone of $3\,h^{-1}$ Gpc of radius containing around 70 billion dark matter particles. Galaxies are populated in the simulations with an algorithm based on the HOD model [42] designed to reproduce clustering observations. Some galaxy catalogues of this simulations are designed in order to be comparable to DES Science Verification data. The goal of these catalogues is to use them to calibrate and study the systematics and methods that will be applied to DES data analysis.

The Millennium and MICE Grand Challenge simulations are described in more detail in §2.2 and §4.3.1 respectively.

1.6 Outline of the thesis

This thesis focuses on the study of linear bias and the relation between galaxy, halo and matter distributions. We study the large scale bias of galaxies and haloes in simulations and the connection between both, and we also develop a method to measure local bias

[1] `http://www.virgo.dur.ac.uk/`

in observations from the correlation between the galaxy distribution and the weak lensing fields.

In §2 we use the Millennium Simulation and its SAM galaxy catalogues to study galaxy and halo bias. We present a test of the halo mass dependence of bias and halo occupation that consists on a reconstruction of galaxy bias from halo bias and the HOD. We show deviations between the bias reconstruction and the galaxy bias measured in the simulation, specially for small haloes, that indicate that halo bias and occupation do not depend only on mass, and has implications in the galaxy bias predictions. In §3 we continue the analysis by studying the halo bias dependence on mass and environment. Apart from mass, we also define the environment of each halo from the surrounding dark matter density. We show that environmental density constrains much more bias than mass, meaning that environment is much more correlated with bias than mass. We see that massive haloes only occupy dense regions, while small haloes can live in any environment. The study implies that small haloes in dense environments have a high bias. If the galaxy properties are correlated with environment, galaxy bias reconstructions do not recover the real bias. We finally show that using the environmental density instead of mass for the bias reconstructions allow us to recover much better the real galaxy bias. This suggests that including information about environment when analysing galaxy clustering is important in order to avoid errors in the estimations of the HOD.

In §4 we present a new method to measure local bias from the combination of the galaxy density and weak lensing fields. We use the density field of galaxies to construct a template of the convergence field. Bias is then obtained from the zero-lag correlations between the reconstructed convergence field and the real convergence field obtained from the shape of the galaxies. We describe the theoretical contribution of the correlations as a function of redshift, that allow us to make measurements of local bias in tomographic redshift bins. We use the MICE simulations to test the method and explore the regimes where the method is valid and consistent with linear bias. We study the sensitivity of the method on different effects such as redshift binning, angular resolution, galaxy number density, nonlinearities of bias, etc. We show measurements of galaxy bias as a function of redshift using this method that are consistent with other estimators of linear bias. This method is a good complement to other measurements of bias because it does not depend strongly on σ_8, while most of the methods to measure bias in observations do. After studying this method in the MICE simulations, we discuss its application to data by presenting the main results of [44], where we have applied the method to DES Science Verification data.

We finish in §5 with a summary of the results of the thesis and the main conclusions.

Chapter 2

Halo Mass dependence of bias and its effects on galaxy clustering

Abstract

In this chapter we study how well we can reconstruct the 2-point clustering of galaxies on linear scales, as a function of mass and luminosity, using the halo occupation distribution (HOD) in several semi-analytical models (SAMs) of galaxy formation from the Millennium Simulation. We find that HOD with Friends of Friends groups can reproduce galaxy clustering better than gravitationally bound haloes. This indicates that Friends of Friends groups are more directly related to the clustering of these regions than the bound particles of the overdensities. In general we find that the reconstruction works at best to $\simeq 5\%$ accuracy: it underestimates the bias for bright galaxies. This translates to an overestimation of 50% in the halo mass when we use clustering to calibrate mass. We also find a degeneracy on the mass prediction from the clustering amplitude that affects all the masses. This effect is due to the clustering dependence on the host halo substructure, an indication of assembly bias. We show that the clustering of haloes of a given mass increases with the number of subhaloes, a result that only depends on the underlying matter distribution. As the number of galaxies increases with the number of subhaloes in SAMs, this results in a low bias for the HOD reconstruction. We expect this effect to apply to other models of galaxy formation, including the real universe, as long as the number of galaxies incresases with the number of subhaloes. We also find that the reconstructions of galaxy bias from the HOD model fails for low mass haloes with $M \lesssim 3-5\times10^{11}\,h^{-1}\,\mathrm{M}_{\odot}$. We find that this is because galaxy clustering is more strongly affected by assembly bias for these low masses.

2.1 Introduction

Understanding the link between galaxies and dark matter is one of the fundamental problems that makes precision cosmology difficult to reach. Nowadays cosmological simulations provide accurate measurements of the dark matter distribution of the Universe, but we need to relate the dark matter to galaxy distributions in order to compare to observations.

There are several empirical models of galaxy formation that allow us to populate dark matter simulations with galaxies. On one side, the Halo Occupation Distribution (HOD) (e.g. [19]) formalism uses the Halo Model (e.g. [55]) to describe the population of galaxies in haloes according to the properties of the host haloes. In many cases the models of galaxy formation assume that the properties and population of galaxies depend only on the halo mass. The population of galaxies is then described by the probability $P(N|M)$ that a halo of virial mass M contains N galaxies of a given type. One can then calculate galaxy clustering from the combination of the HOD with the clustering of halos if we assume that the clustering of haloes depends only on the halo mass. [194] used the GIF simulations [119] to model the 2-point clustering of galaxies from the clustering of haloes and the HOD. They found a good agreement with semi-analytical models.

If these assumptions are valid we can use galaxy surveys to obtain the relations between properties of galaxies and halo mass, to measure the clustering of dark matter haloes, as well as halo masses [57, 90, 233, 234, 236]. [200] used the assumptions from [234] to study the relation between halo mass and satellite and central luminosities, finding a strong relation between central galaxy luminosities and halo mass, and weak dependence for the satellite galaxies. [156] assumed the HOD to study the relation between the stellar mass of galaxies and halo mass, finding agreement with galaxy clustering in the SDSS. [123] found a relation between the halo mass dependence of populations of subhaloes and the HOD of baryonic simulations and semi-analytical models of galaxies, an indication that the distribution of galaxies can be closely related to the distribution of subhaloes.

However, some studies indicate that several properties of galaxy and halo clustering depend on properties of dark matter haloes other than mass, such as halo formation time, density concentration or subhalo occupation number [61, 84, 226]. [226] also saw that the dependence of halo clustering on halo formation time changes with mass. Some studies suggested the idea of adding a second halo property to the HOD model [1, 219, 226]. [61] studied the clustering dependence of haloes on the occupation number of subhaloes, and found that for fixed masses, the halo bias depends strongly on the number of subhaloes per halo. They found that, for fixed occupation number of subhaloes, the halo bias depends on mass. They also found a strong anticorrelation between clustering and mass for

highly occupied haloes. As galaxies possibly follow subhalo gravitational potentials, this dependence can also be found for galaxies, as we show in §2.4.1.

Moreover, the clustering of haloes have an impact on galaxy clustering. [84] showed that the clustering of haloes also depends on the halo formation time, the first indication of assembly bias. [62] studied the effects of assembly bias in galaxy clustering and showed that for fixed halo mass red galaxies are more clustered than blue galaxies. Other authors have studied the effects of assembly bias in galaxy clustering using other environmental dependencies of haloes such as density and studied these effects in observations [2, 200, 216]. [225] studied the correlation between colour and clustering of clusters of the same mass and they found the red clusters (or with a red central galaxy) to be more clustered than blue. However, [18] seem to find the opposite result.

On the other hand, semi-analytical models (SAM) populate galaxies in the dark matter haloes by modelling baryonic processes such as gas cooling, disk formation, star formation, supernova feedback, reionization, ram pressure or dust extinction [10, 11, 48] according to the potentials of dark matter. These processes contain free parameters that can be constrained by observations. Because of these processes, semi-analytical models of galaxy formation follow the evolution of the dark matter haloes, mergers, and they are more physical than HOD models in terms of environmental dependences and evolution.

In this chapter we study the consequencies of the assumption that the galaxy population and clustering only depends on the halo mass in SAMs. We use the Millennium Simulation [206] to measure the halo bias. We also compare different definitions of halo mass. We study the public SAMs of galaxies of the Millennium Simulation to see if we can reproduce the galaxy bias from this HOD assumption, thereby assuming that the clustering of galaxies only depends on the mass of the host halo. We do this by measuring both the halo bias and the HOD in the same simulation and for the same galaxies that we want to reproduce the bias. This analysis is similar to that of [194] but with some important differences. First of all, [194] model the halo bias, while we use the measurement in the simulation, and we also compare it to modelling the bias. Moreover, we include the errors of the reconstructions of galaxy clustering in order to see numerically their success. Another difference with the study of [194] is that we study several and newer SAMs in order to compare the results between them. Finally, the Millennium Simulation presents a better resolution than GIF simulations and therefore we can study smaller masses, and larger volume so we can study the 2-halo term properly. We also focus on large scales, where no assumptions are needed for the distribution of galaxies inside the haloes. [194] assumed the galaxies to be tracers of dark matter particles of the haloes, an assumption that has an impact on the 1-halo term of

the 2-point clustering. Here we only look at the 2-halo term. We also analyse the clustering dependence of the halo occupation of subhaloes, and its dependence on halo mass.

The chapter is organized as follows. In §2.2 we introduce the Millennium Simulation as well as the semi-analytical models of galaxy formation. In §2.3 we present the measurements of galaxy and halo bias and the HODs that we use to reconstruct the galaxy bias and compare it to the measurements in the simulation, which is developed in §2.4. We finish with a summary and discussion in §2.5.

2.2 Simulation

For this study we use the Millennium Simulation [206], carried out by the Virgo Consortium using the GADGET2 [207] code with the TREE-PM (Xu 1995) algorithm to compute the gravitational interaction. The simulation corresponds to a ΛCDM cosmology with the parameters: $\Omega_m = \Omega_{dm} + \Omega_b = 0.25$, $\Omega_b = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $n = 1$ and $\sigma_8 = 0.9$. It contains $2160^3 = 10,077,696,000$ particles of mass $8.6 \times 10^8\, h^{-1}\, \mathrm{M}_\odot$ in a comoving box of size $500\, h^{-1}\,$Mpc, with a spatial resolution of $5\, h^{-1}\,$Kpc. The Boltzmann code CMBFAST [190] has been used to compute the initial conditions based on WMAP [203] and 2 degree Field Galaxy Redshift Survey (2dFGRS) data [49]. The simulation output starts at $z = 127$ and it has 64 snapshots from this time to $z = 0$.

2.2.1 Haloes

In each snapshot, the haloes are identified as Friends-of-Friends (FOF) groups with a linking length of 0.2 times the mean particle separation. All the FOF with fewer than 20 particles are discarded. Then, subhaloes are identified in the FOF groups using the SUBFIND [207] algorithm, discarding all the subhaloes with fewer than 20 particles. The largest object found by SUBFIND in the FOF is located in its centre, and usually has approximately the 90% of the FOF mass. In SUBFIND all the particles gravitationally unbound to the subhalo are discarted, giving a gravitationally bound object. For this reason, the largest SUBFIND object can be seen as the gravitationally bound core of the FOF, and in this chapter we will call it halo.

We must notice that these FOF and halo catalogues are independent of the galagxy catalogues of the simulation. In Fig. 2.1 we show the mass function of FOF and haloes, in red and blue lines respectively, compared to the theoretical models of Tinker et al. (2008) [215] and Sheth, Mo & Tormen (2001) [194] (hereafter SMT 2001). The mass function has been multiplied by $M^2/\bar{\rho}$ for clarity. We define the masses of the FOF as the total number of particles belonging to the FOF, and the halo mass as the total number of particles

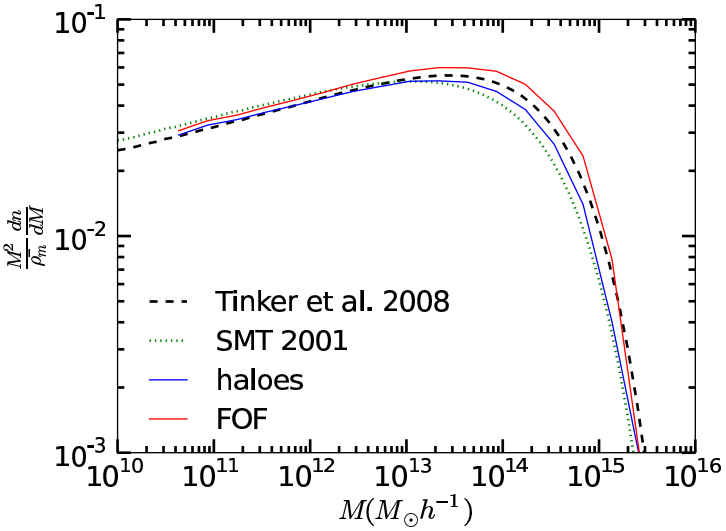

Figure 2.1: Normalized Mass Function of FOFs (red) and haloes (blue) compared to theoretical models. Black dashed line represent the theoretical mass function from Tinker et al. 2008 [215]. Green dotted line shows the model from Sheth, Mo & Tormen 2001 [194].

belonging to the halo (the largest SUBFIND of the FOF). We can see that the haloes have a mass function close to the model of SMT 2001, where the ellipsoidal collapse model has been used. On the other hand, the mass function of FOF is closer to the model of Tinker et al. (2008). We can also see that the mass function of haloes is very close to the one of FOF at low masses, meaning that most of the small haloes are contained in small FOFs, while at large masses the mass function of haloes is lower, meaning that the differences in the mass definition of haloes and FOFs become larger.

2.2.2 Galaxies

Galaxy catalogues of several semi-analytical models (SAM) are available in the public database of the simulation. In this chapter we study several SAMs ([27, 38, 66, 77, 97]. In the models of Bertone, De Lucia & Thomas 2007, De Lucia & Blaizot 2007 and Guo et al. 2011 (BDLT07, DLB07 and G11 respectively hereafter), developed at the Max-Planck-Institute for Astrophysics (MPA) in Garching[1], the galaxies are placed and evolved in the subhaloes according to their properties and their merger trees [63, 66, 206]. In the case of Bower et al. 2006 and Font et al. 2008 models (B06 and F08 hereafter), developed at the Institute for Computational Cosmology in Durham[2], the merger trees are constructed using

[1] `http://gavo.mpa-garching.mpg.de/MyMillennium/`
[2] `http://galaxy-catalogue.dur.ac.uk:8080/MyMillennium/`

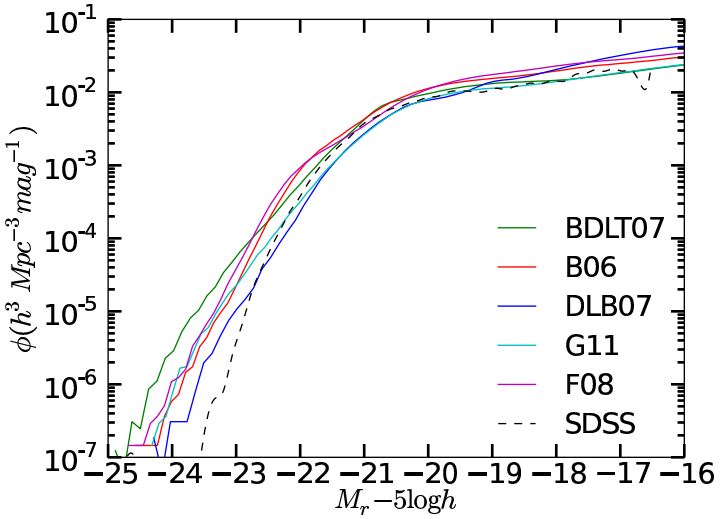

Figure 2.2: Luminosity Function of SAMs compared to the SDSS DR2 data [32]. Solid lines represent the different SAMs, and the dashed line shows the Luminosity Function from SDSS [32].

the *Dhaloes*, a different definition of halo consisting in groups of subhaloes [104, 111]. In most of the cases, the Dhaloes consist in the set of subhaloes of the same FOF, but in some cases [3] these sets are divided in different Dhaloes [99, 149]. Then the evolution of the latter models is related to halo (Dhalo) evolution, while the first models are associated to subhaloes.

In Fig. 2.2 we compare the luminosity function of all the SAMs studied with the luminosity function of SDSS DR2 [32]. To make coherent comparisons, we have used the luminosities in SDSS r filter of galaxies including dust extinction of the SAMs. We applied a factor of $5\log h$, with $h = 0.73$, in the MPA models, since this factor was not included in the database. In this Figure we can see an evident excess of bright galaxies in all the models. There is also a slight excess of galaxies in the faint end in the models DLB07, B06 and F08. However, [22] studied different algorithms to obtain galaxy magnitudes and argued that the luminosity function from [33] in r band is probably underestimated for galaxies brighter than $M_r < -22$, so these differences do not necessarily reflect problems for the SAMs. These results are in agreement with [51], where they present a deeper comparison of the different semi-analytical models of the Millennium Simulation. [51] studied how much the clustering and HOD of these semi-analytical models depend on stellar mass, cold gas mass and star formation rate.

[3] if the subhalo is outside twice the half mass radius of the parent halo or the subhalo has retained 75% of the mass it had at the last output time where it was an independent halo.

2.3 Bias and HOD

In this section we present our measurements of FOF, halo and galaxy bias that we use to analyse if galaxy bias depends only on halo mass. We estimate the 2-Point Correlation Function (2PCF) using density pixels and the expression

$$\xi(r_{12}) = \langle \delta(r_1)\delta(r_2) \rangle, \tag{2.1}$$

where $\delta(r)$ refers to the density fluctuation defined by $\delta(r) = \rho(r)/\bar{\rho} - 1$ in pixels. From that, we measure the bias using the local bias model [83]:

$$b(r) = \sqrt{\frac{\xi_g(r)}{\xi_m(r)}} \tag{2.2}$$

where $\xi_g(r)$ corresponds to the 2PCF of the studied object (haloes or galaxies), $b(r)$ is the bias factor, and $\xi_m(r)$ is the 2PCF of the dark matter field. As we assume $b(r)$ to be constant at large scales, where $\delta \ll 1$, we define the mean value by fitting $b(r)$ to a constant in the scale range $20h^{-1}\,\mathrm{Mpc} < r < 30h^{-1}\,\mathrm{Mpc}$. Although theoretically the bias may not be in the linear regime for these scales, we have checked that it behaves as a constant, so our fit can be assumed as valid. Moreover, the size of the Millennium Simulation does not allow us to go to larger distances with precision. The errors are measured with a Jack-Knife method [160] of this measurement of b, using 64 cubic subsamples. The errors are taken from the standard deviation of these subsamples. The distribution of these subsamples is close to a gaussian, and the errors obtained from the percentiles are very similar to those from the standard deviation.

2.3.1 Bias

In Fig. 2.3 we present the FOF (top) and halo (bottom) bias as a function of mass. We refer to them as $b_{FOF}(M)$, $b_h(M)$ and we compare the results with some analytical models. Using the mass function developed by [173] assuming the spherical collapse model, Mo & White (1996) [152] derived the following expression for the halo bias:

$$b(\nu) = 1 + \frac{\nu^2 - 1}{\delta_c}, \tag{2.3}$$

where $\delta_c = 1.686$ is the linear density of collapse and $\nu = \delta_c/\sigma(M)$, where $\sigma(M)$ is the linear rms mass fluctuation in spheres of radius $r = (3M/4\pi\bar{\rho})^{1/3}$. Sheth, Mo & Tormen

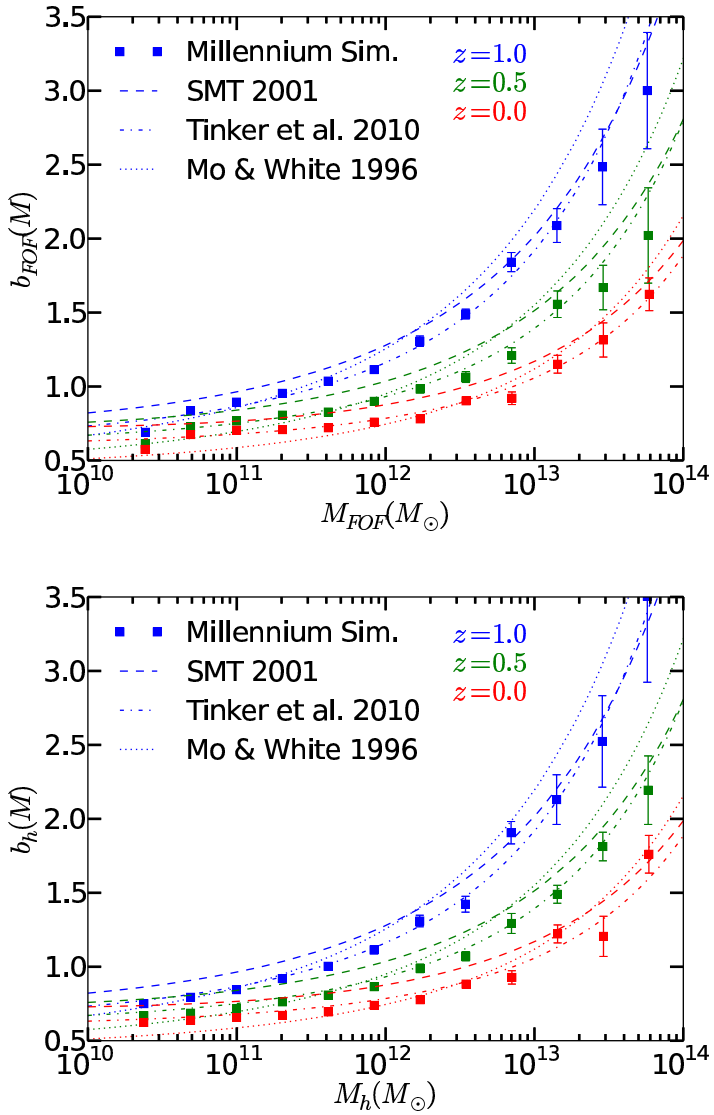

Figure 2.3: FOF (top) and halo (bottom) bias as a function of mass at 3 different redshifts compared to theoretical expressions. The squares show the measurements of bias from the Millennium Simulation. Dashed lines show the analytic model from [194], dashed-dotted lines correspond to the [218] model and dotted lines are the analytic expressions from [152]. Each colour represents a different redshift, as specified.

(2001) (hereafter SMT 2001) [194] generalized and improved the expression using an ellipsoidal collapse model and they obtained the result:

$$b(\nu) = 1 + \frac{1}{\sqrt{a}\delta_c}\sqrt{a}(a\nu^2) + \sqrt{a}b(a\nu^2)^{1-c} - \frac{(a\nu^2)^c}{(a\nu^2)^c + b(1-c)(1-c/2)}, \tag{2.4}$$

with the parameters $a = 0.707$, $b = 0.5$ and $c = 0.6$ tuned to work in N-body simulations. Finally, Tinker et al. (2010) [218] presented a more flexible expression:

$$b(\nu) = 1 - A\frac{\nu^a}{\nu^a + \delta_c^a} + B\nu^b + C\nu^c. \tag{2.5}$$

The values of the parameters of this expression used in our comparisons correspond to the values shown in Table 2 of [218] with $\Delta = 200$.

First of all, we can see that the SMT 2001 model tends to overpredict $b_{FOF}(M)$ and $b_h(M)$, especially at low masses. We can see a difference in the high mass region between FOF and haloes because for each FOF the mass of the halo is reduced (and then shifted to a lower mass) due to the SUBFIND unbinding process. On the other hand, Mo & White (1996) model tends to produce an overprediction at high masses and an underprediction at low masses for all the cases. One possible reason for this is that Mo & White (1996) assume the Press-Schechter mass function, but this mass function fails to reproduce the halo mass function in simulations [92, 95, 110, 134, 196]. Finally, the agreement of the Tinker et al. (2010) expression with $b_{FOF}(M)$ and $b_h(M)$ is remarkable.

In Fig. 2.4 we show galaxy bias as a function of luminosity, $b_g(L)$, for the SAMs. The first 5 panels show $b_g(L)$ in bins of luminosity for each of the studied SAMs, in the r band at 3 different redshifts, as specified. We also show in bottom right panel the comparison of these models with observations in the SDSS DR7 [233] using luminosity thresholds. To compute $b_g(L)$ from the SDSS DR7 data, [233] used a prediction of the dark matter correlation function from a ΛCDM cosmological model [202]. Solid lines show the different SAMs, while the dashed-dotted line shows $b_g(L)$ from [233]. As the cosmology assumed in [233] is different than the one from Millennium, the dashed line corresponds to multiply the SDSS measurement by a factor $0.8/0.9$. This factor is an approximation of the difference in the amplitude of the dark matter field of both cosmologies if we assume that $b_g(L)$ behaves as σ_8. Then, the dashed line shows an approximation of $b_g(L)$ of the SDSS galaxies normalized by the Millennium cosmology.

First of all, we can see that $b_g(L)$ increases with z, although the brightest galaxies tend to show higher clustering amplitude at low z. From the last panel, however, we observe

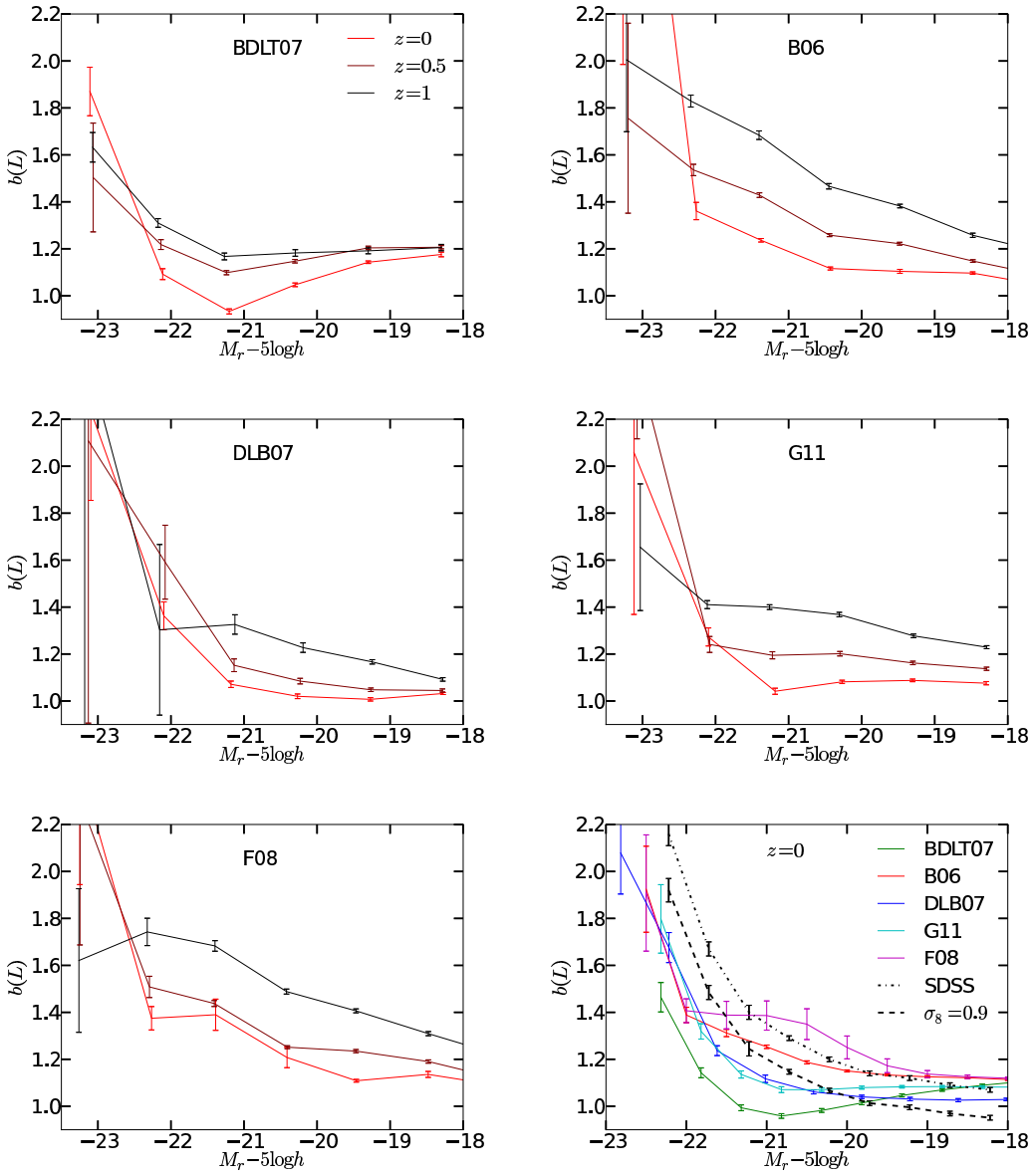

Figure 2.4: Luminosity dependence (in absolute r band magnitude) of galaxy bias for 5 SAMs and comparison with observations (bottom-right panel). The first 5 panels correspond to $b_g(L)$ at $z = 0, 0.5, 1$ for each SAM as labelled in magnitude bins. In the bottom right panel, $b_g(L)$ of all the SAMs at $z = 0$ are compared to $b_g(L)$ from SDSS DR7 [233] in magnitude thresholds instead of bins. Solid lines represent the Millennium catalogues. The black dashed-dotted line corresponds to the SDSS DR7 data. The black dashed line shows a correction of $0.8/0.9$ to approximate the amplitude of $b_g(L)$ from SDSS if the 2PCF were normalized to a cosmology of $\sigma_8 = 0.9$ instead of $\sigma_8 = 0.8$.

discrepancies for all the models with observations from SDSS DR7 presented in [233]. We notice the good agreement between F08 and B06 models in the brightest galaxies. In general, the predicted $b_g(L)$ is lower than in the observations for the brightest galaxies, and the shape of $b_g(L)$ steepens only for the brightest galaxies in the SAMs, showing a shift with respect to the SDSS DR7 data. This shift depends strongly on the different cosmologies adopted between the SDSS analysis and the Millennium Simulation. As the value of σ_8 is higher in the Millennium Simulation, their galaxy clustering is underpredicted due to the lower clustering of the dark matter of the simulation. The dashed line shows the comparison between the models and SDSS assuming the same dark matter field cosmology parameter σ_8. Here we can see that the agreement is better in the brightest galaxies, but worse in the faint end. From this panel, we can say that the agreement on $b_g(L)$ between the models and the observations is strongly dependent on the cosmologies that we assume, but anyway the shape of $b_g(L)$ of the models is different than that of SDSS data. If M_r in SDSS is shifted up as suggested by the luminosity function of Fig. 2.2, then the agreement will be better.

2.3.2 HOD

SAMs of galaxy formation are not based on the halo model and hence they do not use the HOD prescription to populate galaxies into haloes. But effectively the models produce an HOD as an output, and we can measure the occupation of galaxies in haloes and study the mass dependence of the different populations. For each galaxy catalogue, we calculate the HOD by counting the galaxies per halo as a function of the halo mass. For the reconstructions of $b_g(L)$ in §2.4, we will assume the HOD to be only dependent of halo mass (FOF or halo mass). We also analyse the luminosity dependence of these HODs. These distributions are shown in Fig. 2.5, where the HOD of some models are compared to the measurements from SDSS DR7 [233]. [233] inferred the HOD measurements from the clustering of different samples of galaxies assuming that the clustering of galaxies can be expressed in terms of the probability distribution that a halo of a given virial mass M hosts N galaxies of a given type. In this calculation they assume a ΛCDM cosmology with $\Omega_m = 0.25$, $\Omega_b = 0.045$, $\sigma_8 = 0.8$, $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $n_s = 0.95$. Dashed lines in Fig. 2.5 show the best fit of the HODs of the SDSS RD7 presented in [233] using the equation:

$$\langle N(M_h)\rangle = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\log M_h - \log M_{min}}{\sigma_{\log M}}\right)\right]\left[1+\left(\frac{M_h - M_0}{M_1'}\right)^{\alpha}\right], \tag{2.6}$$

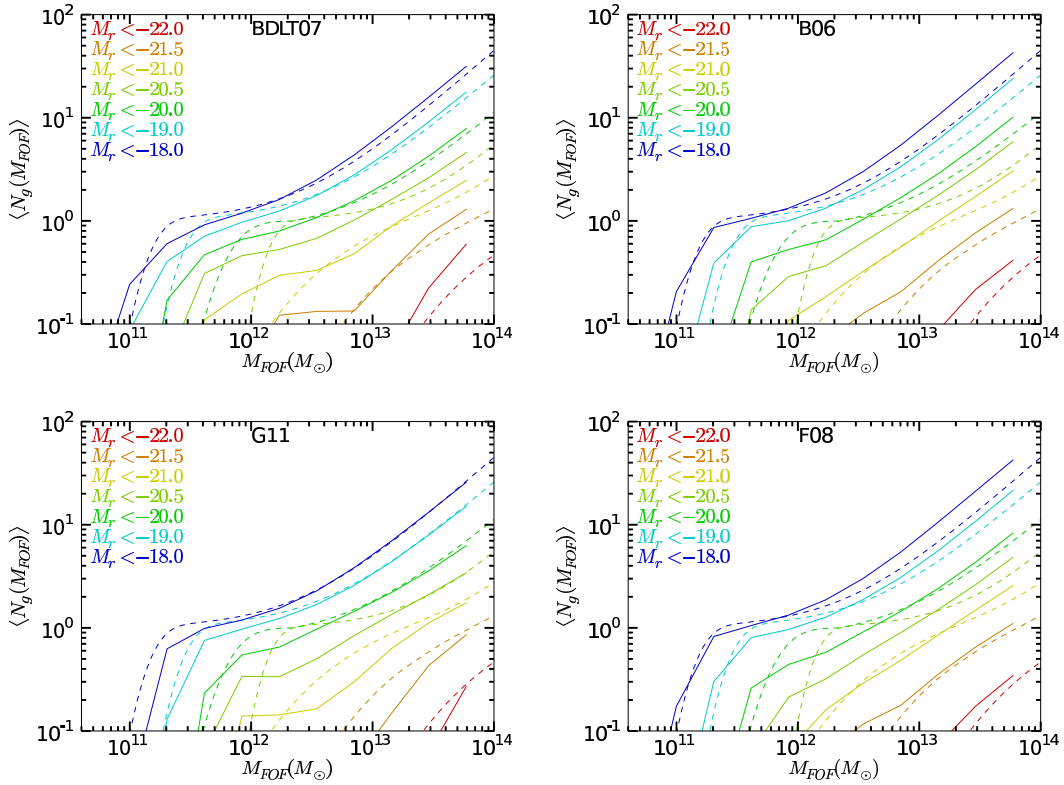

Figure 2.5: HOD of galaxies for the different SAMs (solid lines) compared to the HOD found by [233] from the SDSS DR7 data (dashed lines). Top panels correspond to the BDLT07 (left) and B06 (right) models, bottom panels show G11 (left) and F08 (right) models. Each colour corresponds to a luminosity threshold in units of $M_r - 5\log h$ as specified.

where M_{min}, $\sigma_{\log M}$, M_0, M_1' and α are parameters to be fitted. The SAM measurements are shown by solid lines and each colour corresponds to a different threshold in magnitude, M_r, using the FOF mass. The values of the HOD of the SAMs using the haloes instead of the FOF groups, although it is not shown, is pretty similar. Given the fact that the haloes have always lower mass than their respective FOF (since the difference with respect to the FOF is due to the application of the unbinding processes for dark matter particles), $N(M_h)$ always has to be higher than $N(M_{FOF})$ for a given mass if the slope of the HOD is positive, as it is. At high masses, we note that although the original SAMs tend to have a higher population, the agreement with observations is remarkable, and very good in the particular case of G11 model. At low masses, the level of agreement is model dependent, but in general the change of slope at $\langle N_g(M_{FOF})\rangle \approx 1$ tends to be softer in the SAMs than in the SDSS DR7 measurements.

We must mention, however, that the HOD measurements depend on cosmology, and different assumptions on cosmology can give different fits of the HOD in observations. Since the cosmology assumed in [233] is different than the Millennium cosmology, the HOD measurements of both cases do not need to agree.

2.4 Bias Reconstructions

In this section we want to measure if we can recover galaxy bias by assuming that the HOD and galaxy clustering depend only on halo mass. To do this we make a reconstruction of $b_g(L)$, that we call $b_{rec}(L)$, from the measurements of $b_{FOF}(M)$ (or $b_h(M)$) and the occupation of galaxies in these haloes. If we are in the linear regime (as we are) and the occupation of galaxies is only halo mass dependent, then the value of $b_g(L)$ must coincide with the reconstruction $b_{rec}(L)$ obtained from the following expression [52, 57, 187, 226]:

$$b_{rec}(L) = \int dM b_h(M) n_h(M) \frac{N_{g,h}(M,L)}{n_g(L)} \tag{2.7}$$

where n corresponds to the number density of the galaxies or haloes and $N_{g,h}(M,L)$ is the mean number of galaxies per halo (or FOF) of mass M. We test these $b_{rec}(L)$ for both FOF and haloes. As we use a range in halo mass, the galaxy sample is restricted to those objects that are inside the considered haloes, excluding those which are outside the range of halo mass. Because the FOFs and haloes represent exactly the same objects, one could think that they should give the same results. But their definitions of mass are different, and these differences are important for large and unrelaxed haloes. For this, the mass dependencies of clustering and the relations between mass and galaxy occupation can be stronger for

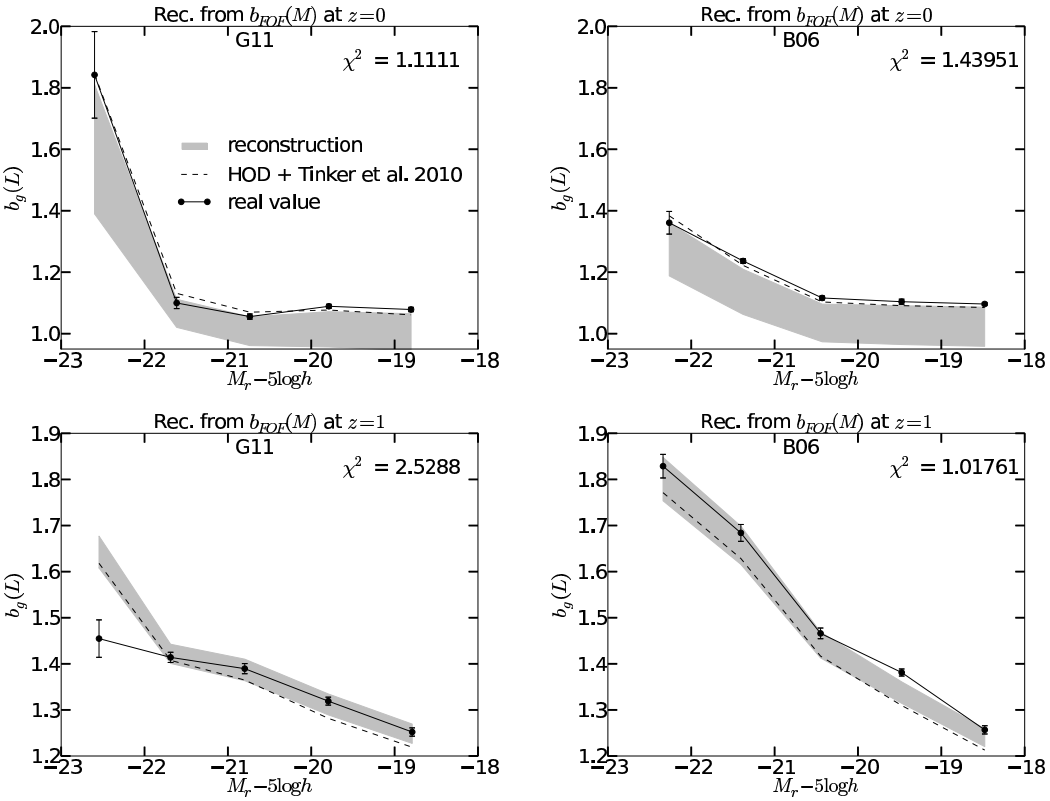

Figure 2.6: Reconstructions of $b_g(L)$ from $b_{FOF}(M)$ at $z = 0$ and $z = 1$. The grey shaded region corresponds to the predicted $b_{rec}(L) \pm 1\sigma$, while solid line represents the real measured value of $b_g(L)$, and the dashed line corresponds to the reconstructions using the values of $b_{FOF}(M)$ from the Tinker et al. (2010) model. Left panels corresponds to G11 model, while right panels show predictions for the B06. On top, $z = 0$. On bottom, $z = 1$.

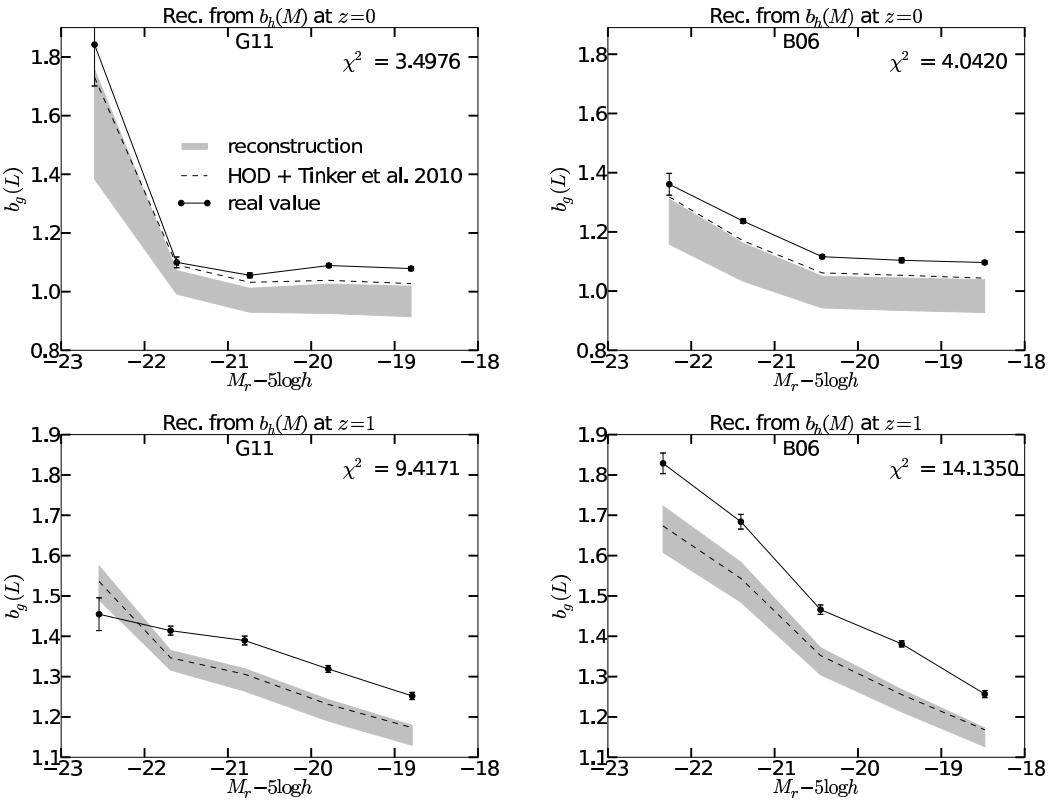

Figure 2.7: The same as in Fig. 2.6, but the reconstructions are obtained from $b_h(M)$ instead of $b_{FOF}(M)$.

one definition than for the other, and the effects of large and unrelaxed haloes will produce differences in the reconstructions.

The error is obtained by calculating the Jack-Knife error of the reconstruction using 64 cubic subsamples. Finally, in all the reconstructions χ^2/ν is calculated according to the formula:

$$\chi^2/\nu = \frac{1}{N}\sum_{i}^{N}\frac{M_i - R_i}{\sigma_{M,i}^2 + \sigma_{R,i}^2}, \tag{2.8}$$

where N is the number of data points, M and R are the measured and reconstructed points respectively, and $\sigma_{M,i}$ and $\sigma_{R,i}$ are the respective errors of M and R. We assume that the reconstructions of each subsample is independent of each other.

We focus on the G11 and B06 models in M_r as representatives of MPA and Durham models to analyse the reconstructions. Although the SAMs have clear differences in the luminosity dependence of $b_g(L)$, the results of the reconstructions of all the SAMs present similar behaviours and the same qualitative conclusions than in G11 and B06.

Fig. 2.6 shows the reconstructions of $b_g(L)$ from FOFs of these two models at $z = 0$ and 1. The reconstructions $b_{rec}(L)$ are shown as a grey shaded region representing $b_{rec}(L) \pm 1\sigma$, and they are compared to the real values, in solid lines. We also show as dashed lines $b_{rec}(L)$ using the Tinker et al. (2010) [218] model for $b(M)$ instead of the measurements of the simulation in order to compare the differences between modelling and measuring $b_{FOF}(M)$ for the reconstruction. Top panels show $z = 0$ and bottom panels are at $z = 1$. Finally, in the left panels we see the reconstructions of G11, and in the right panels we used the B06 model. In Fig. 2.7 we show the reconstructions of $b_g(L)$ from haloes instead of FOFs for the same SAMs and redshifts.

First of all, we can see that the reconstructions using the Tinker et al. (2010) [218] model differ with respect to the reconstructions from the measurements of $b_{FOF}(M)$ or $b_h(M)$ by the order of 1σ. The errors of the reconstructions reflect the fluctuations of the measurements of $b_{FOF}(M)$ and $b_h(M)$ in the simulation. Given the agreement between modelling and measuring bias from Fig. 2.3, this difference in the reconstructions can be seen as a measurement of the effects of the fluctuations of $b_{FOF}(M)$ and $b_h(M)$ in the reconstruction. But, as the galaxies are located in the haloes of the simulation, these fluctuations in the bias should be included in the reconstruction if we want to study the relation between the galaxies and their haloes, so we focus in comparing $b_g(L)$ with $b_{rec}(L)$ from the measured halo bias (i. e. shaded region instead of dashed lines).

Secondly, for the haloes we can see that the agreement between measurements and reconstructions tend to be better at low redshift, although for FOFs this is not so clear.

On the other hand, the reconstructions tend to be different from the SAM measurements by a factor of $6-7\%$ at the level of 1σ for both FOFs and haloes. This differences in clustering corresponds to a 50% difference in halo mass (see Fig. 2.3). Another important result independent of the SAMs and redshift is the fact that FOFs predict better $b_g(L)$ than haloes. This reflects that the unbinding processes, somehow, lose information about the galaxy clustering. In some sense, the mass of the FOF groups is more directly related to the clustering of these regions than the mass restricted to the bound particles in these overdensities, maybe because the FOF groups include more environment of the overdensity. Finally, we can see that the reconstructions tend to underpredict $b_g(L)$. This is a constant effect in z and appears using $b_{FOF}(M)$ or $b_h(M)$ and for all the SAMs. This effect is analysed more in detail in §2.4.1 and §2.4.2.

For equation (2.7) to be accurate we need to satisfy on of the following to conditions. The first condition is that all the haloes of the same mass have the same clustering. If this is the case, all the galaxies in these haloes have the same clustering and then we are assigning the correct clustering for the galaxies. The second condition is that galaxies populate haloes only according to their mass. If this is the case, even if the first condition is not satisfied the galaxies in the same masses must statistically have the same mean clustering. We have seen that the reconstructions differ from the measurement of $b_g(L)$, so this means that both conditions fail. So, for a fixed halo mass, different haloes must have different clustering (assembly bias). Moreover, the population of galaxies in haloes of the same mass must be correlated with the halo bias. In order to study this correlation, in §2.4.1 we study the subhalo occupation dependence of bias. We do this because we expect the number of subhaloes to be directly related with the number of galaxies but at the same time it is independent of the SAM. In §2.4.2 we study the halo mass ranges where we see the assembly bias effects in the reconstructions.

2.4.1 Subhalo population

In this section we study the subhalo occupation dependence of halo and galaxy bias. The idea is to separate the haloes and their content depending on the amount of subhaloes inside the haloes, N_{sh}, to see if clustering depends on their halo substructure. This is an indirect measurement of environment, since the amount of subhaloes in a halo depends on the merging history of the halo, which is related to the environmental abundance of haloes. This is also interesting since the number of subhaloes in the haloes is related to the number of galaxies in it.

In Fig. 2.8 we can see $b_{FOF}(M)$ (top left), $b_h(M)$ (top right) and $b_g(L)$ (bottom left) separating the samples of FOFs, haloes and galaxies according to the number of subhaloes

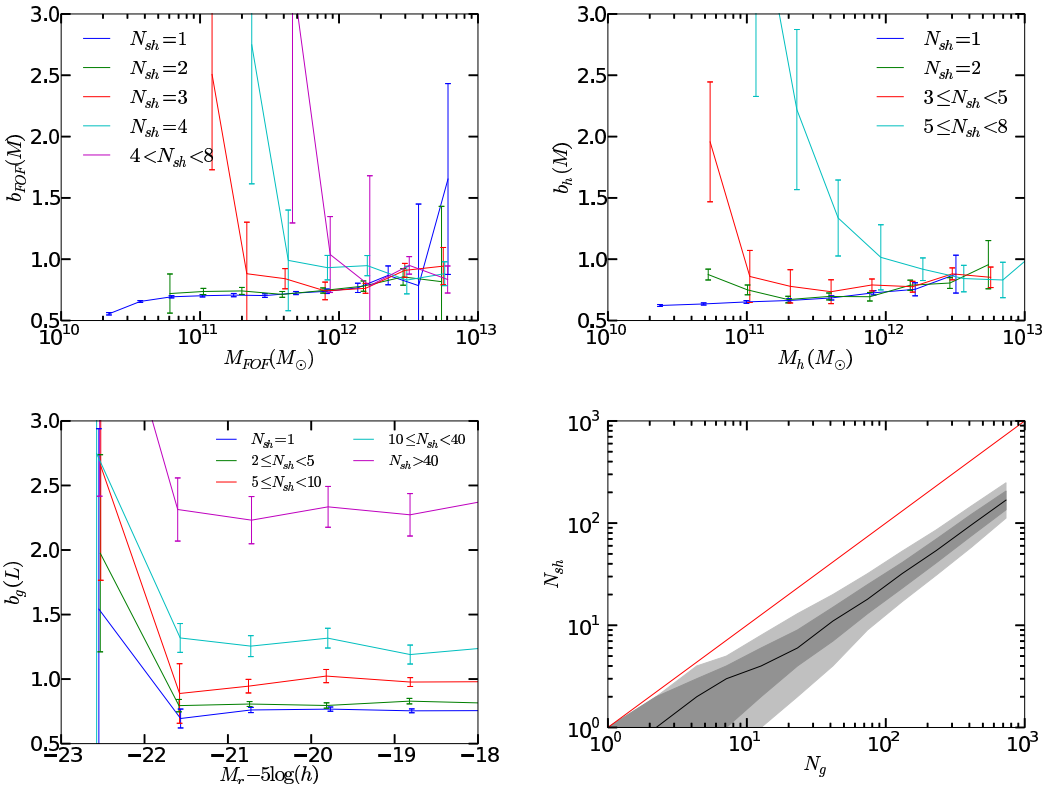

Figure 2.8: Top: $b_{FOF}(M)$ (left) and $b_h(M)$ (right) for different samples according to their number of subhaloes N_{sh} inside. Bottom left: $b(L)$ for galaxies in haloes with different N_{sh} for G11 model. Bottom right: number of galaxies (N_g) vs number of subhaloes (N_{sh}) of the FOFs. The grey shaded regions represent the 68 and 95 percentiles. The red line shows $N_g = N_{sh}$. All the panels are at $z = 0$.

in the halo. The bottom right panel shows the relation between the number of galaxies (N_g) of the G11 model and the number of subhaloes (N_{sh}) in the FOFs. The red line represents $N_{sh} = N_g$. Two conclusions can be obtained from Fig. 2.8 about galaxies and haloes: (1) for a fixed mass (at low masses at least) or luminosity, the dependence of b on subhalo occupation, N_{sh}, is very strong, and (2) for a fixed N_{sh} the dependence of $b_g(L)$ on L is weak. So, galaxy clustering has a strong dependence on N_{sh}, meaning that for a fixed luminosity we are mixing galaxies with different clustering for the same reconstruction, and this can cause a deviation between the reconstruction and the real value of $b_g(L)$. Moreover, for a fixed mass, $b_{FOF}(M)$ and $b_h(M)$ present different clustering according to the number of subhaloes, while in the reconstruction we are wrongly assuming that bias only depends on mass. This is an indication of assembly bias. From bottom right panel we can see that the number of subhaloes increases with the number of galaxies. We can see that the haloes with a high number of subhaloes tend to have even more galaxies. As the haloes with higher N_{sh} have more galaxies, the reconstructions produce an underestimation of $b_g(L)$, since we are assuming the same mean bias for these galaxies, while the haloes with more galaxies have a higher bias than the mean value of these masses. We used the number of subhaloes instead of the number of galaxies because it is independent of the SAM, and this is only dependent on the dark matter distribution, so we should see this effect in any simulated galaxy catalogue where the number of galaxies increases with the number of subhaloes. This is likely what should also happen in the real Universe. Note that this is not the case in HOD assuming $P(N|M)$, by construction.

This can be seen as a galaxy clustering consequence of assembly bias, since we see that haloes of the same mass have different clustering, and we also see that this has consequences on the galaxy clustering predictions. This means that for a fixed mass the galaxies are not randomly distributed, so they depend on other properties than mass. If their distribution were random for each mass, then the reconstruction should work by definition. If we include N_{sh} as another variable in the reconstructions, then the predictions of galaxy clustering are improved. We have checked this by reconstructing $b_g(L)$ and selecting the galaxies according to the number of subhaloes in their haloes. The results are significantly improved in this case (but they are not shown here).

From Fig. 2.8 we can see that the subhalo occupation dependence of clustering is stronger for low masses. In order to see explicitly the masses where we see this effect, in §2.4.2 we study the mass dependence of the reconstructions.

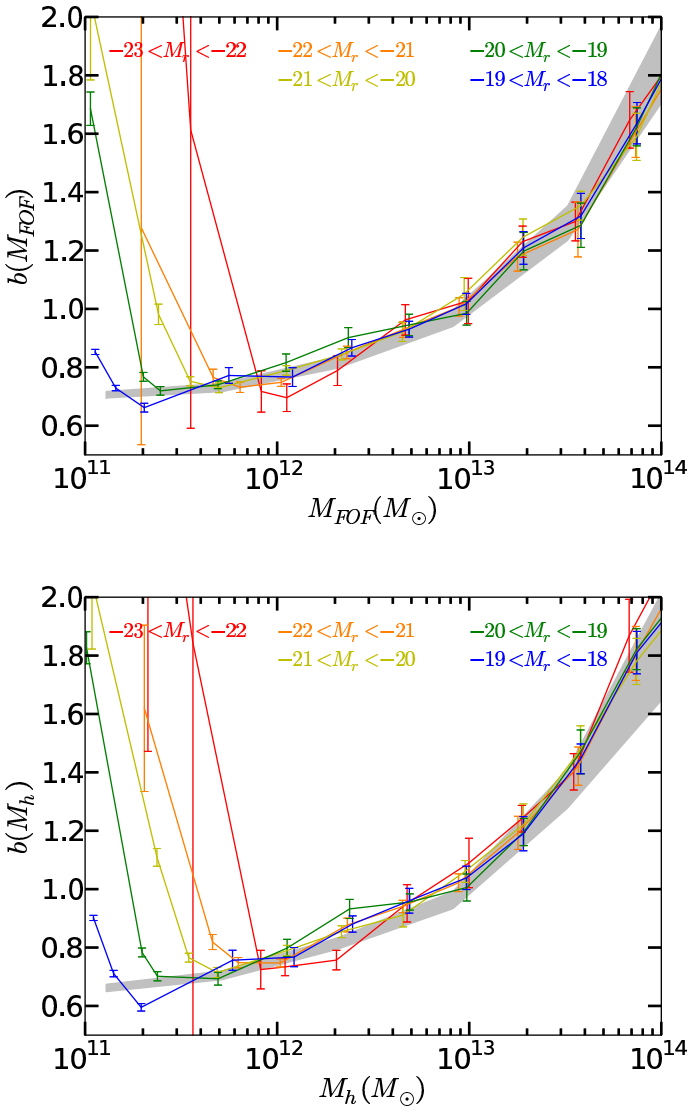

Figure 2.9: $b_g(M)$ and $b_{rec}(M)$ (from FOFs on top and haloes on bottom) for different luminosity bins for G11 galaxies. Solid lines represent galaxy bias as a function of FOF mass (top) or halo mass (bottom) of their host haloes with different luminosities represented by different colours. The grey shaded zone refer to the range of $b_{rec}(M) \pm 1\sigma$ from b_{FOF} (top) and from b_h (bottom). As the reconstructions are binned by halo mass, $b_{rec}(M) = b_{FOF}M$ (top) and $b_{rec}(M) = b_h(M)$ (bottom).

2.4.2 Halo mass dependence

To study the mass dependence of the success in the reconstructions, we measure $b_g(M_{FOF})$ and $b_g(M_h)$ in several luminosity bins. Fig. 2.9 shows $b_g(M_{FOF})$ (top) and $b_g(M_h)$ (bottom) for galaxies at different luminosities. Each colour corresponds to a luminosity. The lines show $b_g(M)$ for galaxies in FOFs or haloes with the respective luminosity, and the grey shaded regions represent the measured ranges $b_{rec}(M) \pm 1\sigma$ obtained from FOFs (top) and from haloes (bottom). This figure, then, allows to see explicitly how the reconstructions of galaxy bias work at different masses. We have only used one mass bin for each reconstruction. For a narrow mass bin, the HOD reconstruction prediction equals the halo model value, so the shaded predictions in Fig. 2.9 equal those of Fig. 2.3. Then, the reconstruction works if the values of $b_g(M)$ are close to $b_{FOF}(M)$ (and the same for $b_h(M)$). In Fig. 2.9 we can see two different behaviours. At high masses, $b_g(M)$ is close to $b_{FOF}(M)$ and $b_h(M)$, since the solid lines tend to be close or inside their shaded zones. This means that the reconstruction of $b_g(L)$ at these masses works, and the halo mass gives sufficient information to predict galaxy clustering. However, we note that there is an underestimation of b_{rec} of the order of 1σ when haloes are used instead of FOFs. In the low mass regions there is a strong disagreement between $b_g(M)$ and $b_{FOF}(M)$, and also with $b_h(M)$, especially for the brightest galaxies. The bias of the brightest galaxies is much higher than the mean one of the haloes of the corresponding mass. So, in these low masses, the galaxies are populated precisely in a way that the brightest galaxies are in the most clustered haloes of the corresponding mass. This means, again, that the clustering of these galaxies does not only depend on mass, and this is also another indication of assembly bias, since haloes of the same mass must have a different clustering. We also notice that the disagreement between $b_g(M)$ and $b_h(M)$ tends to be stronger than between $b_g(M)$ and $b_{FOF}(M)$. This is another indication that M_{FOF} is more strongly related to galaxy clustering than M_h. This means that when the haloes and FOFs present important differences in their masses, M_{FOF} tends to reflect better $b_g(L)$ than M_h.

We see that the distortions in the reconstructions appear at $M_h \lesssim 3-5\times10^{11}M_{\odot}$. Then, the reconstructions of $b_g(L)$ when we exclude haloes of these low masses and their galaxies work, and the predictions of galaxy clustering are correct. This low mass problem can be due to different aspects or a combination of them. First of all, it can reflect the consequences of assembly bias on galaxy clustering. Secondly, it can be affected by the strong stripping and mass distortions of haloes of the Millennium Simulation. When haloes interact with others or pass through high density environments, sometimes the masses are artificially distorted, and this effect is stronger for lower masses. Finally, the SAMs could be affected by assembly bias more strongly than reality. If this is the case, the clustering of SAMs at

these masses would not be correct and we would need to exclude from the analysis those galaxies that reside in low mass haloes, regardless of their properties. But if we exclude these galaxies, then $b_g(L)$ is distorted and an excess of clustering for galaxies of $M_r > -20$ is found, meaning that we need to include these galaxies for clustering studies if we want to recover observations.

2.5 Conclusions

We used the Millennium Simulation to study the clustering of galaxies as predicted by semi-analytical models (SAM) of galaxy formation and the dependence of clustering on luminosity in the SDSS r band filter. We measured the clustering of haloes and we found good agreement with theoretical models, specially for the Tinker et al. (2010) model [218]. We have found discrepancies in the galaxy clustering with respect to observations [233] that can be due in part to their excess of bright galaxies and in part due to the differences on the assumed cosmology for the analysis, in particular the difference in σ_8. Although the SAMs are not based on the HOD model, their populations agree with the observations of the HOD from the SDSS DR7 [233]. We try to reproduce galaxy bias from the bias of FOF groups and gravitationally bound haloes as a function of mass, by assuming that the population of galaxies only depends on the mass of these objects. From our study we obtain the following results:

(i) Although in some cases the reconstructions can provide a good χ^2/ν, the reconstructions tend to underpredict $b_g(L)$ by a factor of $\gtrsim 5\%$. This translates to an error in the inferred halo mass of the order of 50%.

(ii) FOF groups make better reconstructions of $b_g(L)$ than haloes, specially at high redshift, which could be due to the fact that FOF groups include more information about the environment than the main haloes.

(iii) The clustering of haloes and galaxies depends strongly on the amount of substructure in their host haloes. For a fixed halo occupation (of subhaloes), the luminosity dependence of b_g is very weak. For a fixed halo mass, there is a strong dependence of b_h on the occupation of subhaloes, an indication of assembly bias. This result is independent of the SAMs.

(iv) The reconstructions of b_g from haloes work better at high masses, but some disagreements with b_g come from the low mass haloes, where the assembly bias effect is stronger. This effect is stronger for haloes than for FOFs. This means that when the masses of haloes and FOFs are significantly different, the mass of the FOFs reflects better $b_g(L)$ than the mass of the haloes. This effect occurs for $M_h \lesssim 3-5\times 10^{11} M_\odot$. The suppression

of galaxies in the smallest haloes avoids the problems of the reconstructions of $b_g(L)$, but changes the shape of the luminosity dependence of $b_g(L)$, which makes it inconsistent with observations from SDSS DR7 [233].

Our results can also depend on the halo and subhalo finders and the SAMs. On one hand, subhaloes in the Millennium Simulation suffer very strong stripping when they interact with high density environments, and this can have consequences on the assembly bias found at low masses. On the other hand, SAMs have been modelled from these dark matter objects, and their clustering consequences could also present artificial dependences on assembly bias. A galaxy catalogue constructed from an HOD model only using halo mass would not reflect this effect. However, we expect to find this effect for all the galaxy formation models where the number of galaxies increases with the number of subhaloes, and also in the real Universe.

Recent studies [183, 220] indicate that we need to take into account baryonic effects on the dark matter haloes. The density profile of haloes change substantially when baryons are included, and this change can produce important effects on the galaxy formation models applied [220]. [183] also saw that baryonic physics reduces the mass and abundance of haloes below $M_h < 10^{12}\,h^{-1}\,\mathrm{M}_\odot$. Our study indicates that it would be premature to use the HOD interpretation for such halo masses to study these baryonic effects using the clustering of galaxies.

The agreement of Tinker et al. (2010) model, together with the convergence of Fig. 2.9 at the largest masses studied, seem to indicate that the SAMs agree with our assumptions at large masses. However, other dependences than mass are needed to predict the clustering of the SAMs on small masses, then care must be taken when assuming the HOD model at masses below $3-5\times10^{11}M_\odot$, especially when assuming that galaxy clustering only depends on mass. We have seen that the halo clustering depends strongly on N_{sh} for a fixed halo mass, and this explains the discrepancies between the measurements in SAM and the HOD modelling. For any galaxy formation model where $N_g \propto N_{sh}$ we would expect a similar assembly bias. In this case, we will underestimate the galaxy bias with the HOD modelling.

Frequently, the HOD is assumed to relate galaxy properties and halo masses [54, 57, 233]. In the case of [233], they measure the HOD parameters form the clustering of galaxies assuming that the clustering only depends on the halo mass. But these results can be affected by the halo bias dependence on the subhalo occupation for fixed masses if the number of galaxies increases with the number of subhaloes. This conclusion seems quite generic in the light of our Fig. 2.8: for a fix halo mass, clustering is stronger for halos with more substructure. The standard implementation of the HOD assumes that the clustering

of galaxies only depends on the mass of the haloes. If for a fixed halo mass there are more galaxies in the halos with more subhalos, then the mean clustering of these galaxies will be higher than the mean clustering of the haloes and we will wrongly conclude that they are in more massive haloes. This results in an overestimation of the halo mass using clustering. We have shown that this is the case in SAM, but we expect this to also be true in any other model of galaxy formation where the number of galaxies correlate with the number of subhalos.

In order to do a simple estimation of the order of magnitude of this overprediction, in Figure 2.10 we show the relation between the real mass of the haloes and the mass predicted from the clustering of their galaxies. As the mean clustering of the galaxies is higher than the mean clustering of their haloes, the galaxy bias corresponds to the bias of a higher halo mass. In this figure we see the overprediction for galaxies brighter than $M_r < -20$.

We note the degeneracy on the predicted mass over the real mass. We can see for example that a predicted mass of $10^{13} M_{\odot}$ corresponds to both a real mass of $2-3\times10^{11}\,h^{-1}\,\mathrm{M}_{\odot}$ and a real mass of $7-8\times10^{12}\,h^{-1}\,\mathrm{M}_{\odot}$. This means that we cannot strictly predict the mass from the clustering amplitude only, since this degeneracy affects all the masses. In particular we will never get the true mass for the lower mass halos.

From Table 3 of [233] we see that the best HOD parameters obtained are $\log M_{min} = 11.83$, $\log M_0 = 12.35$ and $\log M_1 = 13.08$. If the relation of Figure 2.10 were correct, the parameters should be corrected to $\log M_{min} \simeq 11.74 - 11.83$, $\log M_0 \simeq 12.18 - 12.25$ and $\log M_1 \simeq 12.90 - 12.92$. For this rough estimation we ignore the degeneracy and choose the closest M_{pred}, since for these masses most of the galaxies are populating large haloes. In detail we should weight the prediction by the number of objects. Similar considerations could be applied to other HOD analysis [54,57] where assembly could play a role, specially for the lower masses.

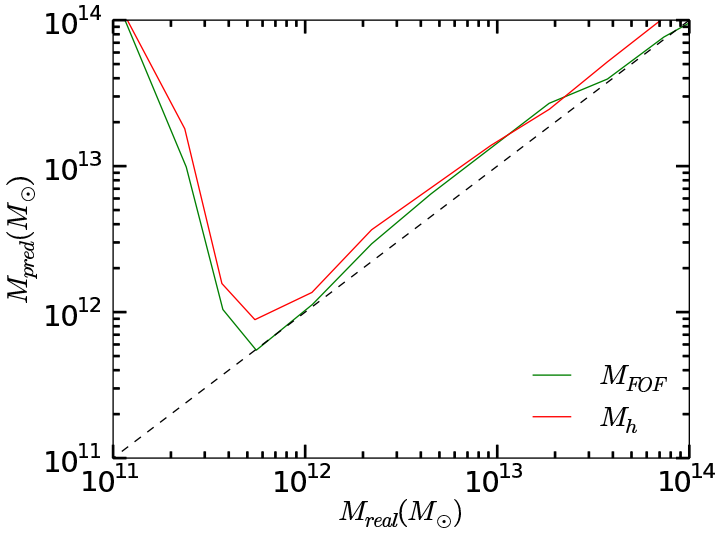

Figure 2.10: Comparison between the halo mass of the host haloes of galaxies with $M_r < -20$ with the predicted halo mass obtained from the galaxy clustering. This can be obtained from figure 2.9 by translating the bias to halo mass for galaxies with $M_r < 20$. In green we used FOF, while in red we show the predictions for haloes. The dashed line shows $M_{pred} = M_{real}$.

Chapter 3

Environmental dependence of bias

Abstract

We continue the study form §2 by studying the large scale halo bias b as a function of the environment (defined here as the background dark matter density fluctuation, $\bar{\delta}$) and we show that environment, and not halo mass m, is the main cause of large scale clustering. More massive haloes have a higher clustering because they live in denser regions, while low mass haloes can be found in a wide range of environments, and hence they have a lower clustering. Using a Halo Occupation Distribution (HOD) test, we can predict $b(m)$ from $b(\bar{\delta})$, but we cannot predict $b(\bar{\delta})$ from $b(m)$, which shows that environment is more fundamental for bias than mass. This has implications for the HOD model interpretation of the galaxy clustering, since when a galaxy selection is affected by environment, the standard HOD implementation fails. We show that the effects of environment are very important for colour selected samples in semi-analytic models of galaxy formation. In these cases, bias can be better recovered if we use $\bar{\delta}$ instead of m as the HOD variable. This can be readily applied to observations as the background density of galaxies is shown to be a very good proxy of environment.

3.1 Motivation

In the standard cosmological framework, the so-called ΛCDM paradigm, galaxies form, evolve, and reside in dark matter haloes that grow and assemble in a hierarchical way [228]. Therefore, the evolution and the properties of the host haloes will have a strong impact in determining the galaxy properties themselves. The unknown nature of the dark matter (and dark energy) can be unveiled through the study of the baryonic observables (such as galaxies) once we model the co-evolution with their host dark matter haloes. Modelling the

relation between the observed galaxy distribution and the underlying dark matter field is a fundamental problem in modern cosmology.

Given a cosmological model, the halo mass function, the halo concentration, and the halo bias can be defined, leaving the galaxy clustering as the next step to complete such a model. In order to model the galaxy clustering statistics it is necessary to specify the number and spatial distribution of galaxies within the dark matter haloes. The probability that a halo of mass M hosts N_{gal} galaxies of a given property is given by the quantity $P(N_{gal}|M)$. This statistical halo distribution is known as the 'Halo Occupation Distribution' [17, 19, 55, 113, 187, 188] (hereafter referred to as HOD). Therefore, its simplicity resides in the assumption that the derived parameters for the cosmological model and the physical properties of the galaxies are determined solely by the mass of the halo in which they reside.

The HOD has been proven to be a very powerful theoretical tool to constrain both the galaxy-halo connection and the fundamental parameters in cosmology [53, 217, 221, 231, 233, 234, 236]. By the use of the HOD mock galaxy catalogues are constructed for studies of galaxy formation as well as for the preparation and analysis of observational surveys [19, 179, 236]. However, it has become clear that the clustering of dark matter haloes depends on other properties besides mass [74, 84, 85, 114, 126, 128, 197, 226, 227].

The construction of mock galaxy catalogues through N-body simulations has shown that the amplitude of the two-point correlation function of dark matter haloes with masses lower than $10^{13}\,h^{-1}\,\mathrm{M}_{\odot}$ on large scales depends on halo formation time (e.g. [84]). Additionally, halo properties such as concentration, shape, halo spin, major merger rate, triaxiality, shape of the velocity ellipsoid and velocity anisotropy show correlations with other properties than halo mass [28, 62, 127]. Moreover, in §2 we show that haloes of equal mass can have different galaxy occupation statistics, depending on their environment (or number of substructures). Ignoring the effects of properties other than mass in the HOD modelling can distort the conclusions and interpretations of the observational results [235]. The galaxy bias reconstructions used in §2.4 can be seen as a test of the HOD standard implementation, that assumes that mass is the only halo property needed to describe galaxy clustering properly.

Several studies point in the direction of taking into account halo properties linked with the environment and their formation history to modify the standard HOD. These properties are incorporated in the formulation of more flexible schemes of HOD used to produce mock catalogues comparable with observations. Such mock catalogues have recently been presented by [62] and [146], who introduced a rank ordering of galaxy colours or SFR. They found that for fixed halo mass, red galaxies are more clustered than blue galaxies. These

and the 'age matching' models of [101, 103] have been shown to successfully reproduce a number of observed signals such as the two-point clustering and galaxy-galaxy lensing signal of SDSS galaxies [128].

In this chapter we study the dependence of large scale halo bias on mass and environment, in order to compare how well mass and environment allow to correctly determine galaxy bias. Our analysis is based on haloes from the Millennium simulation [206] and their Semi-Analytical Models (SAM) of galaxy formation of [97] as in §2. The SAM populate haloes with galaxies that are evolved and followed in time inside the complex structure of merger-trees. The baryonic processes included are laws for metal dependent gas cooling, reionization, star formation, gas accretion, merging, disk instabilities, AGN and supernovae feedback, ram pressure stripping and dust extinction, among others (e.g. [10, 86, 112]). Because of these processes, the resulting galaxy population produced by SAMs is sensitive to the environment and evolution of haloes. Depending on the galaxy selection, the information from halo mass alone could be insufficiently correlated with the clustering of galaxies. In these cases, environmental density allows to determine galaxy bias better than halo mass. Thus, we propose to use the information from environment to improve the HOD analyses on surveys and theory.

The method employed in our study is similar to the one used in §2 which performs a direct HOD reconstruction instead of a HOD fit. We compare the measured halo and galaxy bias with a HOD reconstruction based on direct measurements of the halo occupation, so there are not assumptions or fitting involved. Then, these predictions can be compared with the true bias in the simulation.

The chapter is organized as follows. In section 3.2 we describe the data used and the methodology for our measurements of clustering, bias and for our HOD reconstructions. The results are shown in section 3.3, and we summarize the conclusions of the paper in section 3.4.

3.2 Methodology

3.2.1 Simulation data

In this study we use the same data from the Millennium Simulation [1] [206] than in the previous chapter. We remind that the simulation corresponds to a Λ-CDM cosmology with the following parameters: $\Omega_m = 0.25$, $\Omega_b = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $n = 1$ and

[1] http://www.mpa-garching.mpg.de/millennium/

$\sigma_8 = 0.9$. It contains 2160^3 particles in a comoving box of $500\,h^{-1}\,\mathrm{Mpc}$ of size. We refer to §2.2 for more details about the simulation. For this chapter we focus the analysis at $z = 0$.

In this chapter we define the halo mass from the total number of particles belonging to the Friends-of-Friends (FOF) groups. Although galaxy catalogues of several Semi-Analytical Models (SAM) are available in the public database of the simulation, for this analysis we use the G11 [97] model. The conclusions of this chapter do not depend on the SAM used.

3.2.2 Clustering and bias

Spatial fluctuations of the matter or tracer density ρ are defined as normalised deviations from the mean density $\bar{\rho}$ at the position $\mathbf{r}$, i.e.

$$\delta(\mathbf{r}) \equiv \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}}, \tag{3.1}$$

We measure $\delta(\mathbf{r})$ by dividing the simulation into cubical grid cells, with side length of $500/64 \sim 8\,h^{-1}\,\mathrm{Mpc}$ and assigning a density contrast δ to each cell. We then estimate the two-point correlation function

$$\xi_{AB}(r) \equiv \langle \delta_A(\mathbf{r_1}) \delta_B(\mathbf{r_2}) \rangle, \tag{3.2}$$

which is a function of the scale $r \equiv |\mathbf{r_2} - \mathbf{r_1}|$. The average $\langle \ldots \rangle$ is taken over all pairs of δ in the analysed volume, independently of their orientation. The indices A and B refer to the different density tracer (here haloes or galaxies, which we generically call δ_G) or to the matter density contrasts, which we denote by δ_m, so that $A = B = m$ is the matter auto-correlation, ξ_{mm} and ξ_{Gm} is the cross-correlation between the tracer G and matter m. We then estimate the bias by the ratio:

$$b(r) \equiv \frac{\xi_{Gm}(r)}{\xi_{mm}r)}, \tag{3.3}$$

At large scales $r > 20\,h^{-1}\,\mathrm{Mpc}$, where $\xi < 1$, this ratio is well described by a constant, which is in good agreement with the linear bias, and tends to the value $\delta_G \simeq b_1 \delta_m$, (e.g. see [13]). We estimate this linear bias b_1, which we just call b from now on, by fitting $b(r)$ with a constant in the scale range of $20\,h^{-1}\,\mathrm{Mpc} < r < 30\,h^{-1}\,\mathrm{Mpc}$ as in the previous chapter. For larger scales, the measurements start to be noisy due to the size of the simulation. The covariance of these measurements are derived by jack-knifing [160], using 64 cubical subvolumes.

The value of b depends on how the tracers are selected. In galaxy surveys the limits of the observed galaxy luminosities roughly correspond to a selection by host halo mass. The

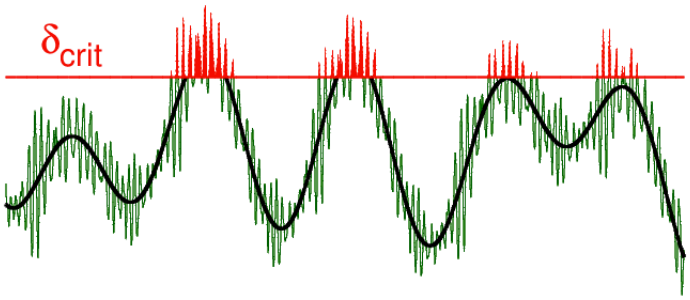

Figure 3.1: Illustration of peak-background split model, $\delta = \delta_s + \bar{\delta}$, showing in red the location of haloes that form above some critical value δ_{crit}. The clustering of those haloes is stronger than the rest (in green). But note how once the environment, given by the large-scale background density $\bar{\delta}$ (black curve) is fixed at δ_{crit}, the large scale clustering of haloes (red) does not depend on the peak-hight (halo mass).

mass dependence of the bias can be predicted from the peak-background split model (see e.g. [106] and references therein) where the matter density field is described by the superposition of small scale fluctuations (peaks) δ_s with large-scale fluctuations of the background matter density $\bar{\delta}$ around each peak (i.e. $\delta = \delta_s + \bar{\delta}$). In this model background fluctuations can lift the peak-heights above a critical density δ_{crit} at which they collapse to haloes as illustrated by Fig.3.1. The mass of these haloes then corresponds to the peak-heights. Consequently massive haloes tend to reside in environments where the background density is high, while low mass haloes can reside in a broader range of environmental densities (we demonstrate this effect with haloes from the Millennium simulation in Section 3.3.1). Massive haloes are therefore more strongly clustered than low mass haloes, which causes a mass dependence of the bias parameter.

3.2.3 Background density

The bias of a given halo sample can also depends on additional halo properties besides the halo mass, such as the concentration of the mass density profile, or on the properties of the galaxies hosted by the halo, such as the halo spin, the formation time, or the age of the stellar population (see e.g. [74, 84, 102]). These additional properties can be related to the merging history of haloes and lead to the so-called assembly bias. Furthermore, the bias of a given halo sample is affected by the tidal forces from the large-scale environment, which is known as non-local bias (see [7, 13, 43]).

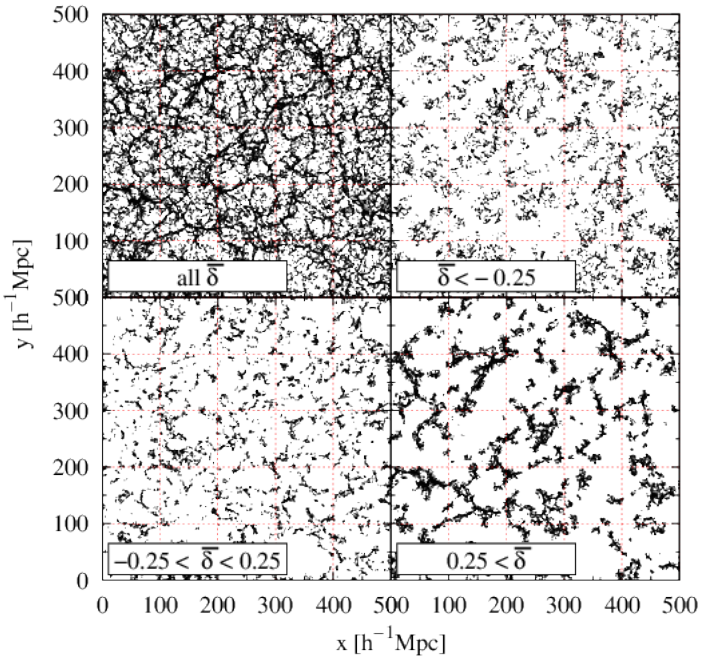

Figure 3.2: Distribution of haloes with different dark matter environmental densities $\bar{\delta}$ in a $2\,h^{-1}$ Mpc slice of the Millennium simulation. The top left panel shows all haloes, while the top right shows haloes in under-densities. The bottom left shows haloes with mean background densities while the bottom right shows haloes in over-dense regions.

These effects can lead to wrong predictions of models which rely on the assumption that the mass of a halo sample completely determines the bias, such as self-calibration techniques (e.g. [230]) or HOD models [174, 235]. A way to circumvent problems for the HOD model is to use a halo property different from the mass, which completely determines the bias of a given halo population. The peak-background split argument described in Subsection 3.2.2 suggests that the clustering of peaks (which correspond to tracers such as haloes) is determined by fluctuations of the large-scale background density $\bar{\delta}$. Hence, for fixed background densities the bias should be independent of the tracer properties (see Fig.3.1 for illustration).

For obtaining a visual impression of this argument we show in Fig. 3.2 the spatial distribution of haloes which reside in different environments, defined by the value of $\bar{\delta}$ estimated around each halo, smoothed on cubical cell of side $l = 14\,h^{-1}$ Mpc as explained below. By comparing the four panels of the figure one can see that the large-scale clustering and hence the bias strongly changes with the local dark matter background density $\bar{\delta}$ (as we will show in Section 3.3.2). On the other hand, the large-scale clustering is nearly independent of halo mass (not shown in this figure), when haloes are selected by $\bar{\delta}$. Note that a weak mass dependence of the bias for fixed $\bar{\delta}$ can be expected from non-local bias

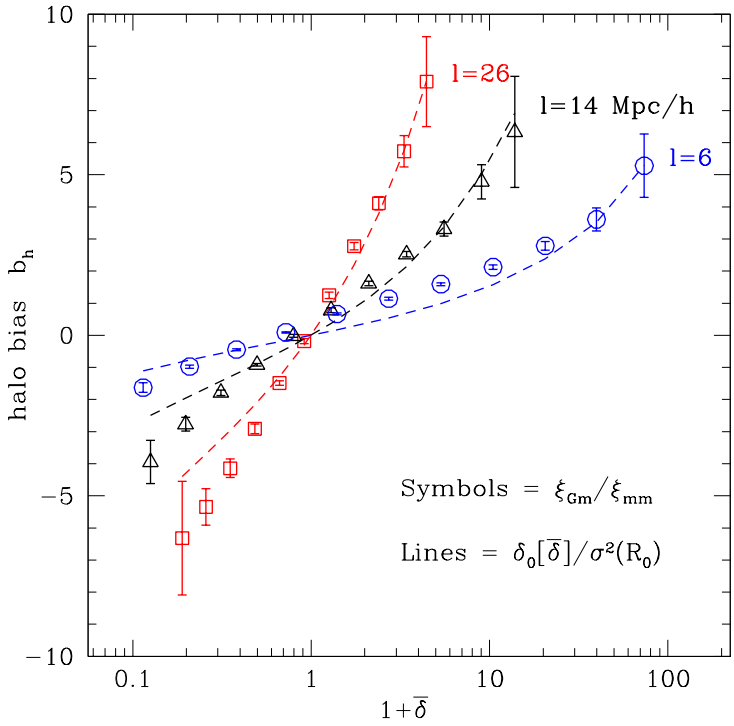

Figure 3.3: Halo bias b_h as a function of halo enviroment $\bar{\delta}$ for $l = 26h^{-1}$ Mpc (red), $l = 14h^{-1}$ Mpc (black) and $l = 6h^{-1}$ Mpc (blue). The corresponding dashed line shows a simple prediction $b \simeq \delta_0/\sigma_0^2$ where $\bar{\delta}_0$ is the linear density corresponding to $\bar{\delta}$ (according to the spherical collapse model) and σ_0^2 is the linear variance at Lagrangian scale $R_0 = R(1+\bar{\delta})^{1/3}$.

contributions.

When determining $\bar{\delta}$ around a given halo we face the problem that the dark matter density distribution of the Millennium Simulation is publicly available only as a $500/256 \simeq 2h^{-1}$ Mpc grid. We therefore assign the density in a cubical volume around each halo, with length $l \simeq 6, 14$ and $26h^{-1}$ Mpc, which have the same volume as a sphere of radius $R = 3.72, 8.68$ and $16.13h^{-1}$ Mpc respectively. The position of these volumes has a $2h^{-1}$ Mpc inaccuracy, which results from the grid cell size. We will focus on the intermediate scale $l = 14$, but we find similar results for other scales.

Inspired by [3], the linear halo bias b_h is shown as a function of $\bar{\delta}$ in Fig. 3.3 using haloes of all masses. The bias increases with the density of the environment and becomes negative in underdense environments. Note that we find negative bias because we study the halo-matter cross-correlation, while in the case of the auto-correlation the bias would remain positive. The dependence of b_h on $\bar{\delta}$ become weaker when $\bar{\delta}$ is defined at smaller scales R. For a study of the galaxy clustering for different environmental densities applied

to data see [3].

Since the large-scale clustering of the tracers corresponds to the clustering of the background density fluctuations in which they reside we can compare our measurements to predictions for the clustering of large-scale peaks, given by [116] for high fluctuations of the density field with a minimum value of $\bar{\delta}_0$,

$$b \simeq \frac{\bar{\delta}_0}{\sigma_0^2}, \tag{3.4}$$

where $\bar{\delta}_0\Delta$ is the initial linear overdensity, which is related to $\bar{\delta}$ via the spherical collapse model (see also [81]) and σ_0 is the corresponding linear variance on Lagrangian scale $R_0 = R(1+\bar{\delta})^{1/3}$. These predictions are in general in a good agreement with the measurements and do not include any free parameters. Note that some of the differences at small $\bar{\delta}$ could come from the fact that we measure the bias in bins of $\bar{\delta}$, while the prediction are for threshold values of $\bar{\delta}_0$, so we expect them to be systematically higher. This can be easily corrected, but is beyond the scope of this study.

In the next subsection we will directly use these bias measurements $b_h(\bar{\delta})$ to study the bias predictions from the HOD model, and to present a new HOD reconstruction test, based on environment $\bar{\delta}$, which is less affected by assembly bias.

3.2.4 Mass and density HOD bias reconstruction

The standard HOD model is based on the assumption that halo bias is determined solely by the halo mass m of the sample. We will therefore refer to it as the mHOD model in this chapter. In this mHOD model the bias of galaxies (b_g) which are selected by an arbitrary property P, can be reconstructed from the halo bias as a function of mass $b_h(m)$ and the mean number of galaxies with property P per halo of mass m, $\langle N_g(m|P)\rangle$ as

$$b^m_{rec}(P) = \frac{\int dm\, b_h(m)\, \langle N_g(m|P)\rangle}{\int dm\, \langle N_g(m|P)\rangle}. \tag{3.5}$$

This equation can be seen as a weighted average of $b_h(m)$, where the weight is given by $\langle N_g(m|P)\rangle$, and it represents equation 2.7 in a more general way, for any selection of galaxies. The mHOD model provides a way to infer the average mass of host haloes in which a given galaxy population is residing by varying $\langle N_g(m|P)\rangle$ in order to reproduce the observed bias, i.e. $b^m_{rec}(P) = b_g(P)$. However, it relies on the assumption that the bias of a galaxy population selected by the property P is completely determined by their host halo mass, which might not be correct as discussed in Section 3.2.2. Cosmological N-body simulations allow us to test the mHOD model, as we can measure $b_h(m)$ and $\langle N_g(m|P)\rangle$

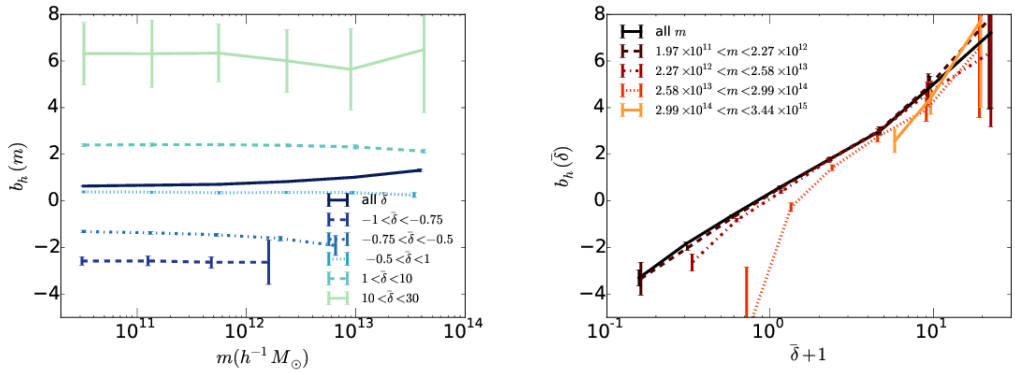

Figure 3.4: Left: halo bias as a function of mass. Each line corresponds to haloes in a fixed density environment $\bar{\delta}$. Right: halo bias as a function of the environmental density $\bar{\delta}$. Each line corresponds to a different halo mass bin. In both panels $\bar{\delta}$ is defined as the Eulerian density fluctuation around a cubic box of $14\,h^{-1}$ Mpc of side.

and compare the predicted $b^m_{rec}(P)$ to measurements of $b_g(P)$. This is what we call the bias reconstruction method and we explore in §2. The test reveals that the mHOD model fails to predict the correct bias in the low halo mass range but it works with 5-10% bias accuracy for the high mass end.

In the previous subsection we considered to redefine the HOD model using the local background density $\bar{\delta}$ around a given halo instead of the halo mass. This approach has the advantage that $\bar{\delta}$ is expected to determine the bias of a given halo or galaxy population better than the mass (we study this in §3.3). In this chapter we call the density HOD model (hereafter referred to as dHOD model) to the analogous of the mHOD model using density instead of mass. The corresponding bias reconstruction is obtained simply by replacing the halo mass in equation (3.5) by $\bar{\delta}$,

$$b^{\bar{\delta}}_{rec}(P) = \frac{\int d\bar{\delta}\; b_h(\bar{\delta})\; \langle N_g(\bar{\delta}|P)\rangle}{\int dm\; \langle N_g(\bar{\delta}|P)\rangle}, \tag{3.6}$$

In the next section we test how well our new dHOD model predicts the bias compared to the standard mHOD model.

3.3 Results

3.3.1 halo bias

In this subsection we analyse the dependence of halo bias on halo mass and dark matter density contrast $\bar{\delta}$ (smoothed with a $l = 14\,h^{-1}$ Mpc side cubical top hat filter) in the envi-

ronment of each halo. The measurement of the latter is described in §3.2. Similar results are found for other smoothing scales.

In the left panel of Fig. 3.4 we show the halo bias, measured via equation (3.3) (hereafter referred to as b_h) as a function of the halo mass m for haloes within different $\bar{\delta}$ ranges. The black line corresponds to $b_h(m)$ for all haloes, regardless of $\bar{\delta}$. This measurement is consistent with the theoretical model of [218] and with the red points of the top panel of Fig 2.3. We can see that $b_h(m)$ does not change significantly (less than a 10%) with halo mass when $\bar{\delta}$ is fixed, while it changes significantly when all the haloes are included (around a factor of 2 in the range of masses shown here). In the right panel we show the same analysis from a different point of view. Here we present $b_h(\bar{\delta})$ for different m bins. Each line corresponds to a range in m, while the black solid line shows $b_h(\bar{\delta})$ for all the haloes. We can see a strong dependence of b_h on $\bar{\delta}$, and that $b_h(\bar{\delta})$ depends weakly on m.

We conclude from these measurements that the bias of a given halo sample is well constrained when the haloes are solely selected by the density contrasts of their environment $\bar{\delta}$, almost independently of the halo mass. This important result confirms our considerations from §3.2.

If $\bar{\delta}$ is left as a free parameter a mass dependence of the bias arises from the fact that high mass haloes tend to reside in high density regions and low mass haloes in regions with lower density. This tendency can be seen in the central panel of Fig. 3.5, where we show the mass versus the environmental $\bar{\delta}$ for each halo in the simulation. The colours describe the number of haloes with the corresponding mass and density. By integrating this distribution over $\bar{\delta}$ we derive the Halo Mass Function (hereafter referred to as HMF), $\langle N_h(m,\bar{\delta})/(V\log(m))\rangle_{\bar{\delta}}$, (where V is the simulation volume and N_h the absolute number of haloes). The HMF is well described by the [215] model, as shown in the top panel of Fig. 3.5. By integrating over the mass m we obtain the Probability Distribution Function (hereafter referred to as PDF) of $\bar{\delta}$, or $\langle N_h(\bar{\delta},m)/n(m)\rangle_m$, which is roughly log-normal, as shown in the right panel of Fig. 3.5.

The aforementioned tendency that high mass haloes tend to reside in high density regions and low mass in regions with lower density can be seen more clearly in Fig. 3.6. In the top panel of this figure we show the HMF of haloes with different background densities $\bar{\delta}$. We find that the fraction of massive haloes decreases in low density environments, while fraction of low mass haloes is similar. The same effect can be seen in the bottom panel of this figure, where we show the PDF of $\bar{\delta}$ for haloes in different mass bins. We find that massive haloes (i.e $m > 10^{14}\,h^{-1}\,\mathrm{M}_{\odot}$) reside almost exclusively in very dense regions and are unlikely to be found in regions with low density, while low mass haloes can be found in a wide range of densities, with preference to average values of $\bar{\delta} = 0$.

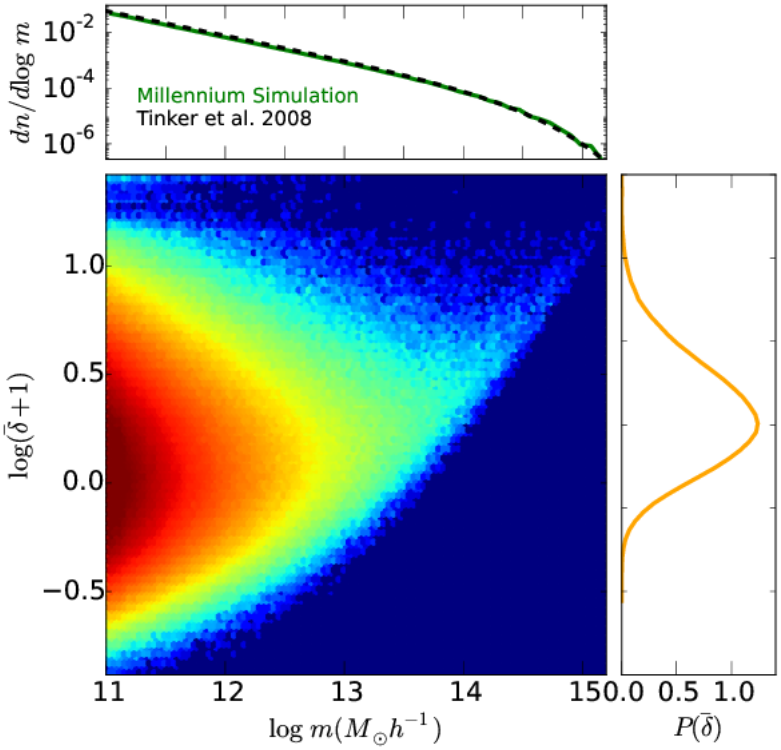

Figure 3.5: Distribution of haloes in environmental density $\bar{\delta}$ and mass m. The colours show the number of haloes that are in a density $\bar{\delta}$ and have a mass m. Right panel shows $P(\bar{\delta})$, so the $\bar{\delta}$ distribution from the contribution of all the haloes. Top panel shows the Halo Mass Function, so the contribution of all the haloes in the mass distribution.

The dependence of the $b_h(m)$ measurements on the local halo background density $\bar{\delta}$ is in consistency with [3] and the fact that the largest scales are the ones that determine better bias. The standard HOD model assumes that halo bias is determined solely by the halo mass, but the environmental density contains additional information that constraints bias. We therefore test the accuracy of mHOD reconstructions in the next subsection and compare it to reconstructions from our new dHOD model from equation (3.6), which is based on the assumption that the bias is determined solely by $\bar{\delta}$, as discussed in §3.2 and suggested by the results of this subsection.

3.3.2 HOD tests using Mass and Density

In this subsection we study how well mass m and environmental density $\bar{\delta}$ determine the linear bias of a given halo sample. We do this by testing how well the mHOD model can predict the halo bias as a function of background density $\bar{\delta}$ and how well the dHOD model can predict the halo bias as a function of halo mass m. The reconstruction for $b_h(P = \bar{\delta})$ from the mHOD model is derived by averaging the $b_h(m)$ measurement using the measured $\langle N_h(m|\bar{\delta})\rangle$ as weight, as given by equation (3.5). This reconstruction is then compared with measurements of $b_h(\bar{\delta})$, derived from the two-point correlation functions via equation (3.3). The dHOD reconstructions for $b_h(P = m)$ is tested in an analogous way using $b_h(\bar{\delta})$ and $\langle N_h(\bar{\delta}|m)\rangle$ measurements in combination with equation (3.6).

We can see in the top panel of Fig. 3.6 that the HMF depends on $\bar{\delta}$ only in the high mass end, while it is similar for different $\bar{\delta}$ in the low mass end. Since the low mass end of the HMF dominates the integral of the of the mHOD bias reconstruction $b^m_{rec}(\bar{\delta})$ from equation (3.5), we do not expect $b^m_{rec}(\bar{\delta})$ to be strongly dependent on $\bar{\delta}$. The PDF of $\bar{\delta}$ in the bottom panel of Fig. 3.6 shows that haloes of different m are differently distributed in $\bar{\delta}$. Therefore we expect a strong dependence of $b^{\bar{\delta}}_{rec}(m)$ from equation (3.6) on m.

The bias reconstructions $b^m_{rec}(\bar{\delta})$ and $b^{\bar{\delta}}_{rec}(m)$ are compared with the measured bias $b_h(\bar{\delta})$ and $b_h(m)$ in Figure 3.7. In the top panel of Figure 3.7, we show in a solid black line the measurement of $b_h(\bar{\delta})$. The hatched line corresponds to $b^m_{rec}(\bar{\delta})$. The bottom panel shows $b_h(m)$, in solid black, and $b^{\bar{\delta}}_{rec}(m)$, in hatched red lines. The region of the hatched lines corresponds to the 1σ errors of the reconstructions, derived from the jack-knife subsamples. As expected from our considerations, the $b^m_{rec}(\bar{\delta})$ reconstruction does not show a significant $\bar{\delta}$ dependence and therefore it deviates from the measured $b_h(\bar{\delta})$. This finding means that, once $\bar{\delta}$ is fixed, the halo mass (or peak height) does not contain additional information about the large-clustering, which already became apparent in the weak mass dependence of the bias of haloes with fixed $\bar{\delta}$, shown in Fig. 3.4. Hence the clustering of a given halo sample selected by $\bar{\delta}$ cannot be reconstructed via the standard mHOD model.

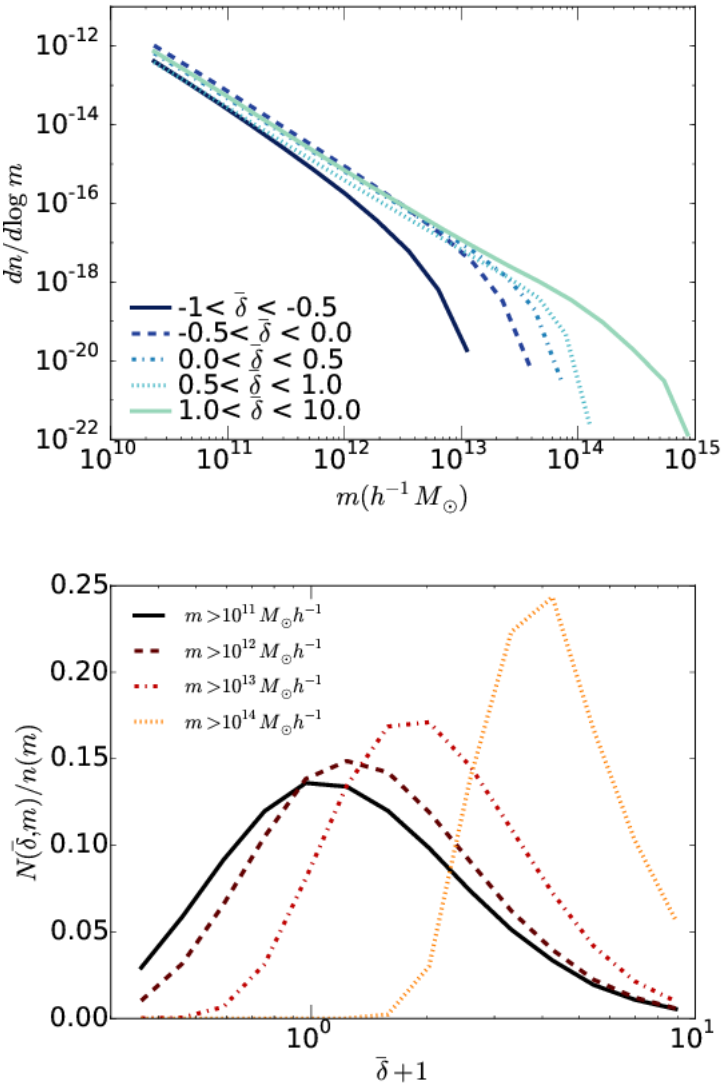

Figure 3.6: Distributions of haloes in m and $\bar{\delta}$. Top panel shows the HMF for haloes in different densities. Each colour represents haloes of a range in $\bar{\delta}$. Bottom panel show the PDF of $\bar{\delta}$ of haloes of different masses. Different colours represent haloes of different m ranges.

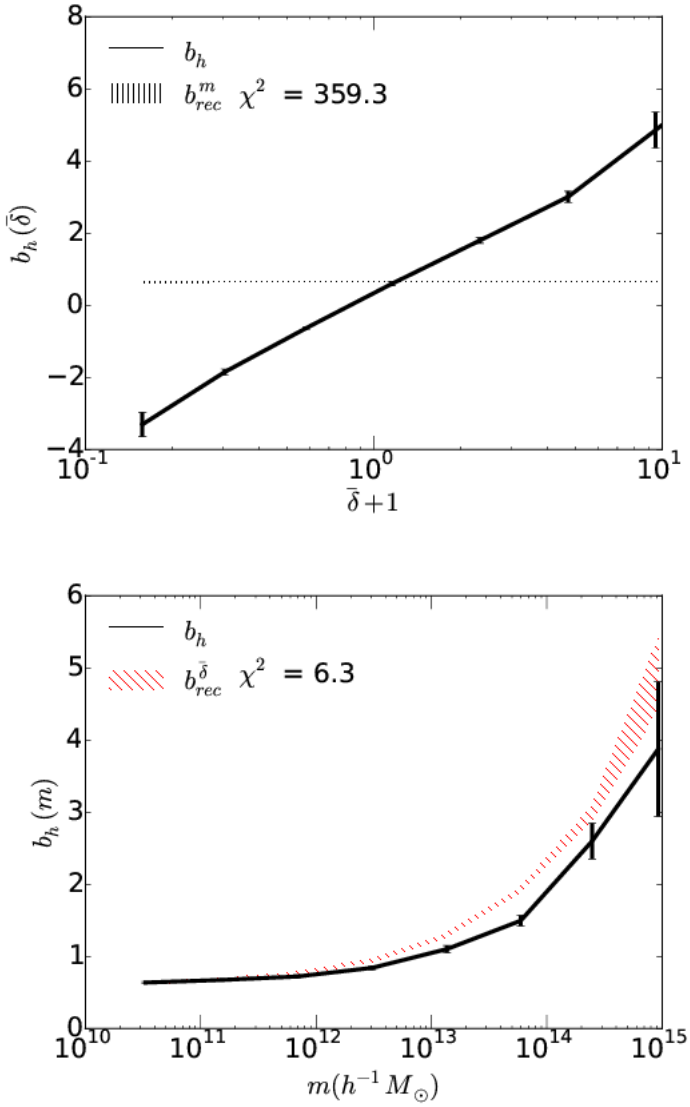

Figure 3.7: Halo bias reconstructions compared to the measured bias from the simulation. The solid black lines show the measurement of b_h from the simulation. The hatched lines show the 1σ interval of the reconstructions. Top panel shows $b^m_{rec}(\bar{\delta})$ from $b_h(m)$, while bottom panel shows $b^{\bar{\delta}}_{rec}(m)$ (in red) from $b_h(\bar{\delta})$.

The $b_h(m)$ dHOD reconstruction, $b^{\bar{\delta}}_{rec}(m)$, shown in the bottom panel of Figure 3.7, is in much better agreement with the measurement. This finding demonstrates that the bias is well determined by $\bar{\delta}$. The over prediction of b_h by the dHOD model at high halo masses results from the fact that b_h is not completely independent of the mass at fixed $\bar{\delta}$, as we see in the left panel of Fig 3.4.

3.3.3 HOD modelling of galaxy bias

The results presented in the previous subsection have important implications on HOD modelling of galaxy clustering as galaxy properties are not only determined by the mass of its host halo, but also by its interaction with the environment.

In cases where a galaxy sample is affected by the environment, assuming that bias only depends on the halo mass can lead to a misinterpretation of the HOD (e.g. the fraction of red satellites) and wrong predictions of galaxy bias as a function of galaxy properties, such as colour or luminosity [235]. In these cases, it can be worthwhile to use the environmental density for HOD bias predictions, since we have shown that it determines the bias better than halo mass.

In Fig. 3.8 we show the comparison of the different HOD reconstruction methods in two different samples of central galaxies from the G11 [97] SAM. In this analysis we focus on central galaxies, since their properties are more correlated with the halo properties than satellite or orphans galaxies, and the implications of the environmental dependence of halo bias are more directly connected to this population.

The top panel shows the bias of central galaxies as a function of the colour index $g-r = M_g - M_r$, where M_g and M_r are the absolute magnitudes in the SDSS g- and r-band, taking into account dust extinction. The bottom panel shows the bias of red central galaxies ($g-r > 0.6$) versus M_r. The cut in colour applied is arbitrary, with the goal of showing a galaxy population with fixed colour cut and different luminosities. The solid black line shows the galaxy bias b_g measured from the clustering via equation (3.3). In vertically hatched black lines we show the 1σ interval of the mHOD reconstruction for b_g from equation (3.5), that indicates how well m determines bias for these samples of galaxies. In diagonally hatched red lines we show the 1σ interval of the dHOD reconstruction of b_g from equation (3.6), that reflects how well $\bar{\delta}$ determines bias for these galaxies.

We can clearly see that the dHOD reconstruction is much closer to the measured bias than reconstruction from the mHOD model. This finding indicates that $\bar{\delta}$ determines b_g much better than m. The environmental density can be related to the formation times of the haloes and their assembly history. In terms of galaxy formation, SAMs show that galaxy colours are affected by the merging events (see [112] and references therein) which

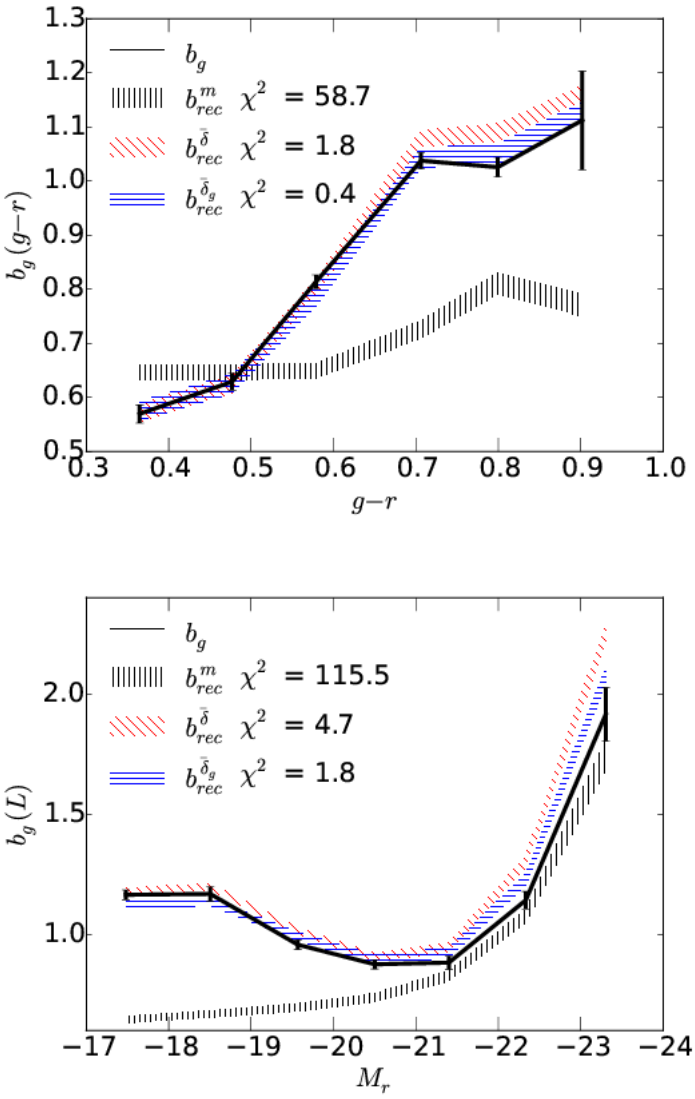

Figure 3.8: Galaxy bias compared to different reconstructions. The black solid lines show the measurements of b_g from the simulation. The hatched lines show the 1σ level of the different reconstructions of b_g from $b_h(m)$ (so b^m_{rec}, in black), from $b_h(\bar{\delta})$ (so $b^{\bar{\delta}}_{rec}$, in red) and from $b_h(\bar{\delta}_g)$ (so $b^{\bar{\delta}_g}_{rec}$, in blue). The top panel b_g is for central galaxies as a function of colour $g-r$, and in the bottom panel b_g is displayed as a function of absolute r-band magnitude for red central galaxies (red defined as $g-r>0.6$).

occur more often in high density environments. We conclude that, when galaxy properties are affected by the halo environment, the standard mHOD model can fail in predicting the bias as a function of that property. In this case HOD bias predictions based on the more fundamental halo property ($\bar{\delta}$), i.e. the dHOD model, delivers more accurate bias predictions. This effect is not always as important as in Fig. 3.8, as we can see from §2. In §2, b_g is shown as a function of absolute magnitude, and since magnitude and mass are well related for these galaxies, the mHOD reconstruction works well.

In real galaxy catalogues, it is very hard to measure $\bar{\delta}$, but instead we can easily measure $\bar{\delta}_g$, the density fluctuations of galaxies. In Fig. 3.8 we also show the bias reconstruction using $\bar{\delta}_g$ instead of $\bar{\delta}$, so $b_{rec}^{\bar{\delta}_g}$, represented as horizontally hatched blue lines. We measure $\bar{\delta}_g$ from galaxies of $M_r < -19$, and we clearly see that the result is equivalent to that of using $\bar{\delta}$, in fact it is even better. This means that $\bar{\delta}_g$ determines bias in a similar way or better than $\bar{\delta}$. This result can be expected from the fact that the large scale fluctuations $\bar{\delta}$ and $\bar{\delta}_g$ are simply related to each other by the linear bias of the background galaxy sample. Also note that $\bar{\delta}_g$ is more closely related to the galaxy distribution than $\bar{\delta}$ (the relation between $\bar{\delta}_g$ and $\bar{\delta}$ might have stochasticity, for example) so is not totally surprising that $\bar{\delta}_g$ gives a slightly better reconstruction than $\bar{\delta}$.

3.4 Conclusions

In this chapter we use the Millennium Simulation [206] and their public catalogues to study the impact of halo mass and environmental density on the prediction of linear bias.

We study the dependence of halo bias on mass (FOF mass) and environmental density, defined as the density fluctuation $\bar{\delta}$ within a given volume around each halo. Although for this study we used a cubical box of side $l = 14\,h^{-1}\,\mathrm{Mpc}$, we validated that the results are similar if we use other scales. We also find similar results when we use Lagrangian instead of Eulerian densities for the background.

We find that bias depends much more strongly on $\bar{\delta}$ than on mass, and once $\bar{\delta}$ is fixed, the halo bias depends very weakly on mass. This is important, since it reflects that the halo bias is well constrained when the haloes are selected by the environmental density, almost independently on the mass. More massive haloes have a higher clustering because they are statistically in denser regions, but not because mass is the fundamental cause of clustering. In particular, low mass haloes can be found in a wide range of densities, and hence these haloes present different clustering. This is in contradiction with the standard HOD implementation which assumes that halo bias only depends on the mass. The environmental density can be seen as a property that is sensitive to other dependencies of halo bias

apart from mass, such as assembly bias. Assembly bias usually refers to the difference in clustering from haloes of equal mass but different formation time or concentration. These differences can indeed be related to environment, as haloes in high background densities form first and have higher concentrations. This can also explain the concept of galactic conformity (see [163] and references there in) by which galaxy properties, such as luminosity and colour, are not solely determined by the mass of the halo. Thus our finding that halo environment, and not halo mass, is the key variable to understand the large scale clustering of haloes is in line with this previous results.

To study the implication of our finding for the clustering predictions in the HOD framework, we use the method of reconstructing the linear bias from the halo bias and the occupation distribution in haloes, as explained in §2.4 and in this chapter. This can be seen as a test of how well mass and environmental density constrain bias. More exactly, these bias reconstructions measure how well linear bias can be reproduced by assuming that the halo bias and occupation only depend on one variable (either mass or density around the halo). We show that we can predict $b_h(M)$ from $b_h(\bar{\delta})$, but we cannot predict $b_h(\bar{\delta})$ from $b_h(M)$. This means that $\bar{\delta}$ determines bias better than M. This is important for HOD analysis, since it is usually assumed that bias only depends on halo mass, but some galaxy populations might be affected by environment as well. According to our results, the dependencies of galaxies on the environmental density have a stronger impact on the large scale clustering than the dependencies on halo mass. With the exception of the higher mass range, which is strongly correlated with background density, as shown in Fig. 3.5, and therefore shows similar tendencies than environment.

Some of the galaxy properties can be sensitive to assembly bias and environment. In these cases, assuming that linear bias only depends on halo mass causes and error in our estimation of clustering or in the estimation of HOD parameters. Instead, we can use the environmental density as a proxy for bias, since it determines bias better than mass on large scales.

We show two examples of galaxy samples, $b(M_r)$ of red central galaxies and $b(g-r)$ of central galaxies, where the galaxy clustering does not depend only on halo mass. We see that the reconstruction using the environmental density $\bar{\delta}$ around the haloes makes a good prediction of galaxy bias, but the standard reconstruction using halo mass does not recover well the galaxy bias. This means, on one side, that the occupation of this population of galaxies in haloes is affected by the environmental density even for fixed mass, and on the other side, that even if the occupation of this galaxies in haloes depends on mass, the clustering of these galaxies is mainly due to the dependence on the environmental density. We also used $\bar{\delta}_g$ instead of $\bar{\delta}$ to measure the environmental density of haloes and the results

are equivalent to those using $\bar{\delta}$, meaning that both $\bar{\delta}$ and $\bar{\delta}_g$ are good estimators of bias. This result is expected, since at large scales $\bar{\delta}_g$ is biased with respect to $\bar{\delta}$, but the nature of both properties is the same, and hence they disclose similar information about environment. The advantage of using $\bar{\delta}_g$ is that it can be directly measured in observations, while $\bar{\delta}$ or halo mass are more difficult to estimate.

This analysis is focused on linear scales, where the 2-halo term dominates, so that we do not need to assume anything about the distribution of galaxies inside the haloes. But this does not mean that a similar analysis for the 1-halo term is not worth, since small scales can also depend on environment. For example, the satellite distribution can depend on the halo concentration for fixed mass. Also, these linear scales now become more accessible with upcoming surveys. As $\bar{\delta}$ is a better estimator of large scale clustering than mass, and it is also easier to measure it in observations (at large scales $\bar{\delta}_g$ is just a biased version of $\bar{\delta}$), this method can be applied in observations to measure the bias as a function of $\bar{\delta}$, and to study galaxy clustering from the modeling of $b_h(\bar{\delta})$ instead of $b_h(M)$ without missing the information from assembly bias and environmental dependencies of bias.

Chapter 4

Measuring local bias by combining galaxy density and weak lensing fields

Abstract

We present a new method to measure the redshift-dependent galaxy bias by combining information from the galaxy density field and the weak lensing field. This method is based on [4], where they use the galaxy density field to construct a bias-weighted convergence field κ_g. The main difference between [4] and our new implementation is that here we present another way to measure galaxy bias using tomography instead of bias parameterizations. The correlation between κ_g and the true lensing field κ allows us to measure galaxy bias using different zero-lag correlations, such as $\langle\kappa_g\kappa\rangle/\langle\kappa\kappa\rangle$ or $\langle\kappa_g\kappa_g\rangle/\langle\kappa_g\kappa\rangle$. This analysis is the first that studies and systematically tests the robustness of this method in simulations. We use the MICE simulation suite, which includes a set of self-consistent N-body simulations, lensing maps, and mock galaxy catalogues. We study the accuracy and systematic uncertainties associated with the implementation of the method, and the regime where it is consistent with the linear galaxy bias defined by projected 2-point correlation functions (2PCF). We find that our method is consistent with linear bias at the percent level for scales larger than 30 arcmin, while nonlinearities appear at smaller scales. We also find that projection along the redshift direction can cause up to a 5% deviation between the different galaxy bias estimators. This measurement is a good complement to other measurements of bias, since it does not depend strongly on σ_8 as the 2PCF measurements. We apply this method to the Dark Energy Survey Science Verification data in [44].

4.1 Introduction

The formation and evolution of the large scale structures in the Universe is an important tool for cosmology studies. But since most of the mass in the Universe is in the form of dark matter, which cannot be directly observed, we need to understand the connection between the observable universe (galaxies and stars) and dark matter. In the ΛCDM paradigm, structures form in the initial density peaks causing dark matter to gravitationally collapse and form virialized objects. Galaxies are expected to follow these gravitational potentials (e.g. [228]), and because of this they are tracers of the dark matter density peaks. The relation between the galaxy and mass distributions can be described theoretically with the galaxy bias prescription [20, 83, 116, 142, 144, 152, 196]. Galaxy bias allows us to connect the distribution of galaxies with that of dark matter, and a good knowledge of galaxy bias would be very important to improve the precision of our cosmological measurements [73].

Many papers have studied halo and galaxy bias in simulations [5, 42, 47, 74, 124, 142, 164, 174, 175, 189, 218, 235], and the different ways to measure bias [13, 21, 106, 124, 142, 172, 180]. Observationally, several measurement techniques exist for constraining galaxy bias. The most common approach is to measure galaxy bias through the 2-point correlation function (2PCF) of galaxies [30, 57, 58, 72, 145, 199, 233]. Counts-in-cells (CiC) is another method where the higher moments of the galaxy probability density function (PDF) are used to constrain galaxy bias [31, 68, 208, 229]. Alternatively, one can combine galaxy clustering with measurements from gravitational lensing, which probes the total (baryonic and dark) matter distribution. Such measurements include combining galaxy clustering with galaxy-galaxy lensing [39, 115, 141, 198] and lensing of the cosmic microwave background (CMB) [91, 185]. The method we present in this work also belongs to this class.

Gravitational lensing is the effect of light deflection due to the perturbations in the gravitational potential from mass distribution. It is a powerful tool to measure the mass distribution in the Universe, since the gravitational potential is affected by both baryonic and dark matter. Weak lensing refers to the statistical study of small distortions (around 1%) in the shapes of a large number of galaxies due to this effect. Several ongoing and future galaxy surveys aim to obtain large weak lensing data sets that will allow us to better constrain cosmology, including the Hyper Suprime-Cam (HSC) [151], the Dark Energy Survey (DES) [75, 213], the Kilo Degree Survey (KIDS) [65], the Panoramic Survey Telescope and Rapid Response System (PanSTARRS) [117], the Large Synoptic Survey Telescope (LSST) [139], Euclid [132], and Wide-Field Infrared Survey Telescope (WFIRST) [94]. From the shape of the galaxies one can statistically infer the lensing fields, which contain

information of the projected matter distribution and can be used to generate 2D and 3D mass maps [147, 222, 223].

The combination of weak lensing and galaxy density information gives us a powerful handle for measuring galaxy bias. [4] used the COSMOS field to measure galaxy bias by reconstructing a bias-weighted shear map from the galaxy density field. The galaxy bias is estimated from the zero-lag cross correlation between this bias-weighted shear map from the galaxy density field and the shear measured from galaxy shapes. Different parameterizations of bias are used to measure constant, non-linear and redshift-dependent bias. In this chapter we explore and extend the method from [4]. We analyze whether the galaxy bias measured with our method is consistent with the linear bias obtained from the projected 2-point correlation functions (2PCF). We find that our method can be affected by different parameters in the implementation such as redshift binning, the redshift range used, angular scales, survey area and shot noise. Finally, we show how to measure the redshift-dependent galaxy bias by using tomographic redshift binning. Although this method is very similar to the one presented in [4], there are few notable differences. First of all, in [4] they explore different smoothing schemes for the density field, while we explore pixelizing the maps and applying a Top Hat filter. In [4] the lensing shear is estimated for each galaxy, and the bias is measured from the predicted and measured shear of the galaxies, while we measure galaxy bias from the generated lensing maps. Finally, [4] fit different parametric biases using a wide range of redshift for the galaxy density field, while here we implement a tomographic measurement, where we measure bias in redshift bins by using the density field of galaxies in each particular bin. We apply this method to the DES Science Verification (SV) data in [44].

The chapter is organized as follows. In §4.2 we give an overview of the theory for our analysis. In §4.3 we present the method used to measure bias from the galaxy density and weak lensing fields and the numerical effects associated with the implementation of the method. In §4.4 we present the results of the different tests and the final measurement of redshift-dependent galaxy bias. We show in §4.5 the results of applying this method to DES SV data from [44]. We finally close in §4.6 with discussion and conclusions.

4.2 Theory

4.2.1 Galaxy Bias

The distribution of galaxies traces that of dark matter, and one of the common descriptions for this relation is galaxy bias, which relates the distribution of galaxies with that of dark

matter. There are several ways to quantify galaxy bias [13, 21, 106, 142, 180], and one of the most common ones is from the ratio of the 2PCFs of galaxies and dark matter:

$$\xi_g(r) = b^2(r)\xi(r), \tag{4.1}$$

where $b(r)$ is the galaxy bias, and $\xi_g(r)$ and $\xi(r)$ are the scale-dependent galaxy and matter 2PCFs respectively, which are defined as:

$$\xi_g(r_{12}) = \langle \delta_g(\boldsymbol{r_1})\delta_g(\boldsymbol{r_2})\rangle,\ \xi(r_{12}) = \langle \delta(\boldsymbol{r_1})\delta(\boldsymbol{r_2})\rangle. \tag{4.2}$$

where $\delta_g = (\rho_g - \bar{\rho}_g)/\bar{\rho}_g$ is the density fluctuation of galaxies (ρ_g is the galaxy number density), and $\delta = (\rho - \bar{\rho})/\bar{\rho}$ is the density fluctuation of dark matter (ρ is the dark matter density). As can be seen from this equation, galaxy bias generally depends on the scale r_{12} (defined as the distance between $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$). However, it has been shown that at sufficiently large scales in the linear bias regime, bias is constant (e.g. [142]).

Bias can also be defined from the projected 2PCFs:

$$\omega_g(\Theta) = b^2(\Theta)\omega(\Theta), \tag{4.3}$$

where $\omega_g(\Theta)$ and $\omega(\Theta)$ refer to the projected 2PCF of galaxies and dark matter respectively. This definition of bias will be used in the analysis of this chapter. In this case, the bias dependence is on angle Θ instead of distance r.

In the local bias model approach [83], the density field of galaxies is described as a function of its local dark matter density, so that $\delta_g = F[\delta]$. We can express this relation as a Taylor series:

$$\delta_g = b_0 + b_1\delta + \frac{b_2}{2}\delta^2 + ... = \sum_{i=0}^{\infty} b_i(z)\delta^i, \tag{4.4}$$

where b_i are the coefficients of the Taylor expansion. In the linear regime, $\delta \ll 1$, and as $b_0 = 0$ because $\langle\delta_g\rangle = \langle\delta\rangle = 0$, then the equation becomes:

$$\delta_g = b_1\delta \tag{4.5}$$

According to [142], at large scales this definition of bias is consistent with the bias obtained from the 2PCFs: for $r_{12} \gtrsim 40h^{-1}$ Mpc, b from equation (4.1) is indeed constant and equivalent to b_1 from equation (4.5). This b_1 can then be measured from the different zero-lag correlations between δ_g and δ:

$$b_1 = \frac{\langle\delta_g\delta\rangle}{\langle\delta\delta\rangle} \tag{4.6}$$

$$b_1 = \frac{\langle \delta_g \delta_g \rangle}{\langle \delta_g \delta \rangle} \tag{4.7}$$

$$b_1 = \sqrt{\frac{\langle \delta_g \delta_g \rangle}{\langle \delta \delta \rangle}} \tag{4.8}$$

Although these relations appear to measure the same parameter b_1, the results can be affected by the stochasticity in the relation between δ_g and δ, that can come from different effects, such as the intrinsic stochasticity of bias and the projection effects.

Galaxy bias from equations (4.6-4.8) depend on the smoothing angular scale θ used to measure δ and δ_g. For small angle θ nonlinearities in the relation between δ and δ_g appear, and b_1 is no longer consistent with equation (4.3). The relation between both scales of bias Θ and θ is complex, since the smoothing of δ and δ_g involves the correlations of all the scales below θ. However, in the linear and local regime bias is constant in both θ and Θ and then all the estimators can be compared.

4.2.2 Weak Lensing

Weak gravitational lensing (see e.g. [9, 177]) measures the small changes of galaxy shapes and brightnesses due to the foreground mass distribution in the line-of-sight of the (source) galaxies. By studying this effect statistically, assuming that (lensed) galaxies are randomly oriented in the absence of lensing, one can infer the mass distribution in the foreground of these source galaxies. As the light distortion is affected by gravity, weak lensing allows us to measure the total mass distribution, including baryonic and dark matter.

The gravitational potential Φ of a given density distribution δ can be defined as:

$$\nabla^2 \Phi = \frac{3H_0^2 \Omega_m}{2a} \delta, \tag{4.9}$$

where H_0 and Ω_m are the Hubble parameter and the matter density today, and a is the scale factor assuming a spatially flat Universe. Assuming General Relativity and no anisotropic stress, the lensing potential for a given source at position $(\boldsymbol{\theta}, \chi_s)$ is given by the weighted line-of-sight projection of Φ:

$$\psi(\boldsymbol{\theta}, \chi_s) = 2 \int_0^{\chi_s} \mathrm{d}\chi \frac{\chi(\chi_s - \chi)}{\chi_s} \Phi(\boldsymbol{\theta}, \chi), \tag{4.10}$$

where $\boldsymbol{\theta}$ is the angular position on the sky, χ refers to the comoving radius and χ_s is the comoving distance to the sources. The distortion of the source galaxy images can be described by the convergence κ and shear γ fields that are defined as:

$$\kappa = \frac{1}{2} \nabla^2 \psi, \tag{4.11}$$

$$\gamma = \gamma_1 + i\gamma_2 = \frac{1}{2}(\psi_{,11} - \psi_{,22}) + i\psi_{,12}, \tag{4.12}$$

where $\psi_{,ij} = \partial_i \partial_j \psi$. Focusing on the convergence field, combining equations (4.9), (4.10) and (4.11) we obtain:

$$\kappa(\boldsymbol{\theta}, \chi_s) = \frac{3H_0^2\Omega_m}{2c^2} \int_0^{\chi_s} d\chi \frac{\chi(\chi_s - \chi)}{\chi_s} \frac{\delta(\boldsymbol{\theta}, \chi)}{a(\chi)} \tag{4.13}$$

For simplicity, we define $q(\chi)$ as the lensing kernel of the integral of δ at χ:

$$q(\chi, \chi_s) = \frac{3H_0^2\Omega_m}{2c^2} \frac{\chi(\chi_s - \chi)}{\chi_s a(\chi)} \tag{4.14}$$

so that

$$\kappa(\boldsymbol{\theta}, \chi) = \int_0^{\chi_s} q(\chi, \chi_s)\delta(\boldsymbol{\theta}, \chi) d\chi. \tag{4.15}$$

Note that κ corresponds to a weighted integral of the matter density fluctuations in the line-of-sight of the source galaxies. γ and κ are related between them through equations (4.11) and (4.12). In the flat-sky approximation, conversion between γ and κ in Fourier space follows [118] KS conversion:

$$\tilde{\kappa}(\ell) - \tilde{\kappa}_0 = D^*(\ell)\tilde{\gamma}(\ell); \;\; \tilde{\gamma}(\ell) - \tilde{\gamma}_0 = D(\ell)\tilde{\kappa}(\ell), \tag{4.16}$$

where "$\tilde{X}$" indicates the Fourier transform of the field X, ℓ is the spatial frequency, $\tilde{\kappa}_0$ and $\tilde{\gamma}_0$ are small constant offsets which cannot be reconstructed and are often referred to as the "mass-sheet degeneracy". D is a combination of second moments of ℓ:

$$D(\ell) = \frac{\ell_1^2 - \ell_2^2 + i2\ell_1\ell_2}{|\ell|^2}. \tag{4.17}$$

4.3 Methodology

4.3.1 Simulation

For the analysis we use the MICE Grand Challenge simulation [60,78,79], an N-body simulation of a ΛCDM cosmology with the following cosmological parameters: $\Omega_m = 0.25$, $\sigma_8 = 0.8$, $n_s = 0.95$, $\Omega_b = 0.044$, $\Omega_\Lambda = 0.75$, $h = 0.7$. It has a volume of $(3.072\,h^{-1}\,\mathrm{Gpc})^3$ with 4096^3 particles of mass $2.927 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$. The galaxy catalogue has been run according to a Halo Occupation Distribution (HOD) and a SubHalo Abundance Matching (SHAM) prescriptions [42]. The parameters of the model have been fitted to reproduce clustering as a function of luminosity and colour from the Sloan Digital Sky Survey [233], as well as the luminosity function [33,34] and colour-magnitude diagrams [35]. We use the

MICECATv2 catalogue, an extension of the publicly available MICECATv1 catalogue[1]. The galaxy catalogue is complete for $i < 24$ from $z = 0.07$ to $z = 1.4$. The catalogue also contains the lensing quantities (γ_1, γ_2 and κ) at the position of each galaxy, calculated from the dark matter field with a resolution of Nside=8192 in healpix (corresponding to a pixel size of ~ 0.43 arcmin), so the lensing quantities of the galaxies do not have shape noise.

4.3.2 Bias estimation

In this section, we introduce the method used to estimate galaxy bias from the lensing and density maps of galaxies in the MICE simulation. It consists on the construction of a template κ_g for the lensing map κ from the density distribution of the foreground galaxies assuming equation (4.5). Substituting δ with δ_g in equation (4.13) gives:

$$\kappa_g(\boldsymbol{\theta}) = \int_0^{\chi_s} q(\chi, \chi_s)\delta_g(\boldsymbol{\theta}, \chi)d\chi \tag{4.18}$$

When computing κ_g numerically, the integral becomes a sum over all lenses in the foreground of the sources:

$$\kappa_g(\boldsymbol{\theta}) = \sum_{all\ lens\ bins} \bar{q}'\delta_g'(\boldsymbol{\theta})\Delta\chi', \tag{4.19}$$

where we have split the foreground galaxies into redshift bins. $\Delta\chi'$ refers to each redshift bin width in comoving coordinates, $\bar{q}'$ is the mean lensing weight that corresponds to each redshift bin and $\delta_g'(\boldsymbol{\theta})$ is the galaxy density fluctuation in each redshift bin at position $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ now represents a pixel in the sky plane. $\delta_g'(\boldsymbol{\theta})$ is calculated through $\delta_g'(\boldsymbol{\theta}) = (\rho_g'(\boldsymbol{\theta}) - \bar{\rho}_g')/\bar{\rho}_g'$, where $\rho_g'(\boldsymbol{\theta})$ is the density of galaxies projected in the line-of-sight in each redshift bin and position (pixel) $\boldsymbol{\theta}$. Notice that $\delta_g'(\boldsymbol{\theta})$ is calculated taking into account all the galaxies inside the volume of the cell corresponding to each pixel and redshift bin. This means that $\delta_g'(\boldsymbol{\theta})$ is constant inside the bin, and it corresponds to a projection of the galaxy density weighted by the volume of the corresponding cell.

In Figure 4.1 we show a schematic picture of the effects of equation (4.19). Dahsed black line shows $q(z, z_s)$, while red solid line shows $\bar{q}'$ in redshift bins of $\Delta z = 0.2$. We used $z_s = 1.3$ for this figure. The blue shaded region represents $\delta_g(z)$ in a random (just for the example) pixel in the sky using narrow redshift bins ($\Delta z = 0.05$). The blue solid line represents δ_g' for the redshift bins of $\Delta z = 0.2$. Equation (4.19) then is equivalent to the integral of the product of the blue and red solid lines.

Equation (4.19) is an approximation of (4.18), that assumes that the small fluctuations in redshift of δ_g inside the bins do not affect the results. It also assumes that the mean

[1] `http://cosmohub.pic.es/`

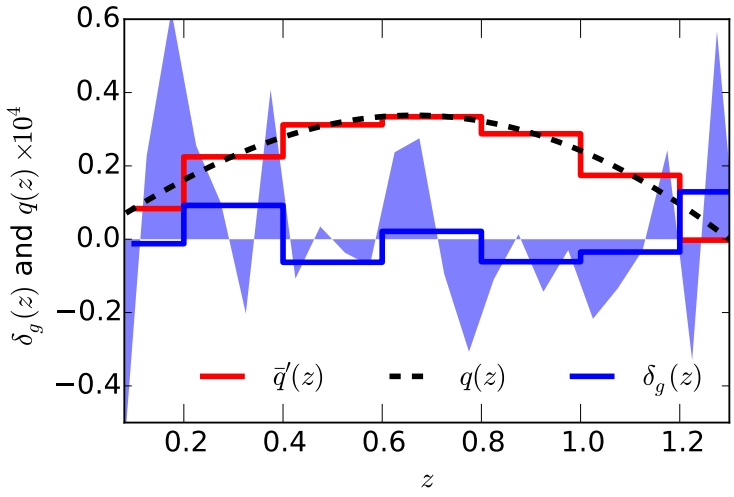

Figure 4.1: Schematic comparison of equations (4.18,4.19). Dahsed black line shows $q(z, z_s)$, while red solid line shows $\bar{q}'$ from equation (4.19) in redshift bins of $\Delta z = 0.2$. The blue shaded region represents $\delta_g(z)$ using narrow redshift bins ($\Delta z = 0.05$). The blue solid line represents δ_g' for the redshift bins of $\Delta z = 0.2$.

of $q(\chi)\delta_g(\chi)$ inside the bins is equivalent to the product of the means $\bar{q}'\delta_g'(\theta)$. These approximations are correct at large scales and when $q'(\chi)$ and $\delta_g(\chi)$ are not correlated.

We focus on the simplest case, where the galaxy bias is linear, local and redshift-independent. In this case, we can estimate b from the following zero-lag correlations of κ and κ_g:

$$b = \frac{\langle \kappa_g \kappa \rangle}{\langle \kappa \kappa \rangle - \langle \kappa^N \kappa^N \rangle} \tag{4.20}$$

$$b = \frac{\langle \kappa_g \kappa_g \rangle - \langle \kappa_g^N \kappa_g^N \rangle}{\langle \kappa_g \kappa \rangle}, \tag{4.21}$$

where κ^N and κ_g^N are the sampling and shot-noise correction factors obtained by randomizing the galaxy positions and re-calculating κ and κ_g. κ is obtained from the mean κ of the galaxies in each pixel. This is affected by the number of source galaxies in the pixel, causing a noise in $\langle \kappa\kappa \rangle$ that depends on the angular resolution used, reaching a 10% error for a pixel size of 5 arcmin. This noise is cancelled by subtracting $\langle \kappa^N \kappa^N \rangle$. On the other hand, $\langle \kappa_g \kappa_g \rangle$ is affected by shot noise, causing an error that increases with the angular resolution up to a 20% for a pixel size of 5 arcmin. This noise is cancelled by subtracting $\langle \kappa_g^N \kappa_g^N \rangle$. This correction assumes a Poisson distribution. To test how well this correction works for this method, we calculated $\langle \kappa_g \kappa_g \rangle - \langle \kappa_g^N \kappa_g^N \rangle$ using the dark matter particles instead of galaxies, and we compared the results with the true $\langle \kappa\kappa \rangle$ maps from the simulation. We did this with different dilutions (from $1/70$ to $1/700$) of the dark matter particles, and

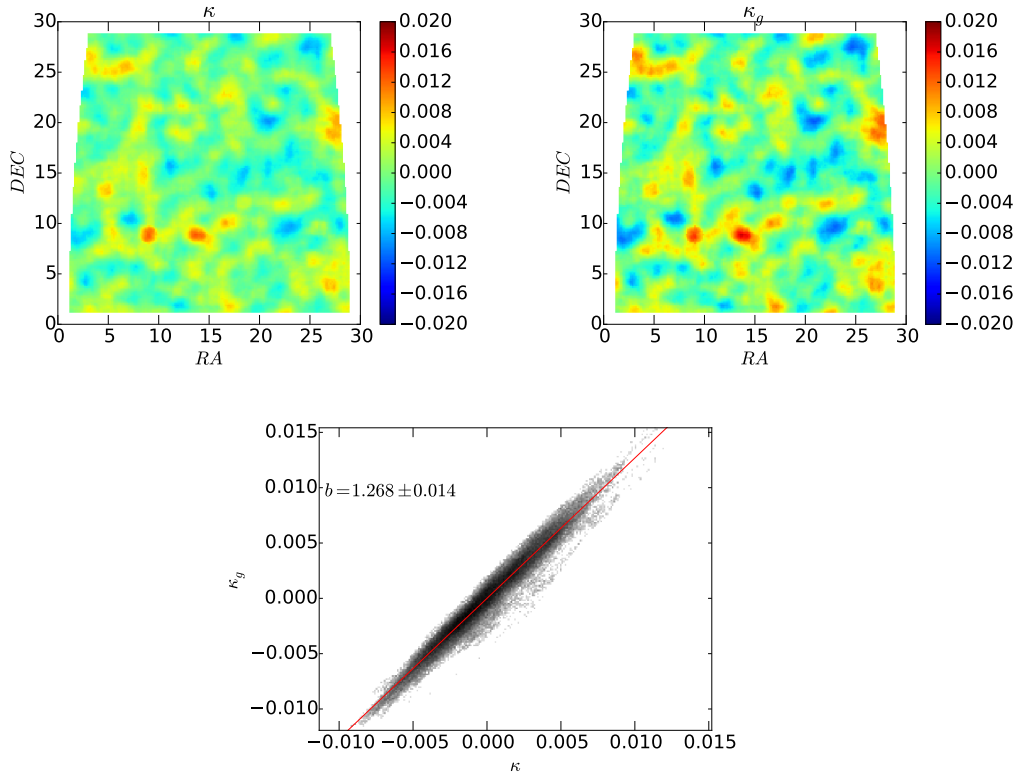

Figure 4.2: Comparison of κ vs κ_g. Top left shows the κ field from the source galaxies within $0.9 < z < 1.1$ and using a Top Hat filter of 50 arcmin of radius. Top right panel shows κ_g obtained from equation (4.19), using the same smoothing scheme. Bottom panel shows the comparison between κ_g and κ for the pixels of the maps, with the specified bias and error obtained. The red line corresponds to a line crossing the origin and its slope corresponds to b. It is consistent with the linear fit of the distribution of the points.

recover $\langle \kappa\kappa \rangle$ better than 1% independently on the dilution, indicating that the shot-noise subtraction is appropriate.

These are the estimators of bias used in this study. Since the galaxies used from the MICE simulation do not have shape noise, the estimators in this analysis are not affected by shape noise. This is not the case in observations, where shape noise is the most important source of noise of this method and needs to be corrected.

To measure the errors on b, we use the Jack-Knife (JK) method. We divide the area into 16 subsamples. We evaluate b 16 times excluding each time a different subsample. The error of b is estimated from the standard deviation of these 16 measurements as:

$$\sigma(b) \simeq \sqrt{\frac{N_{JK}-1}{N_{JK}} \sum_{i=1}^{N_{JK}} (b_i - \bar{b})^2}, \tag{4.22}$$

where N_{JK} refers to the number of Jack-Knife subsamples used and b_i is the bias measured

by excluding the *ith* subsample. We checked that the error does not change if we use a different number of subsamples (between 9 and 100) instead of 16.

Note that we can also measure bias from the following cross correlations, which was originally used in [4]:

$$b = \frac{\langle \gamma_{i,g} \gamma_i \rangle}{\langle \gamma_i \gamma_i \rangle - \langle \gamma_i^N \gamma_i^N \rangle} \tag{4.23}$$

$$b = \frac{\langle \gamma_{i,g} \gamma_{i,g} \rangle - \langle \gamma_{i,g}^N \gamma_{i,g}^N \rangle}{\langle \gamma_{i,g}^N \gamma_i^N \rangle}, i = 1,2 \tag{4.24}$$

As this is not the focus of the study, and we can obtain κ from the simulation, we measure b from equations (4.20,4.21) in this study. However, in observations we measure the shape of the galaxies, that is directly related to γ_i. Because of this, applying this method to data requires a conversion from κ_g to $\gamma_{i,g}$ or from γ_i to κ. We address this issue in [44], where we use conversions based on [118] to apply this method to DES SV data. Another aspect to take into account for data analysis is that since shape noise is the main source of noise in the measurement, we want to avoid the terms that involve variance of lensing quantities $\langle \kappa\kappa \rangle$ and $\langle \gamma_i \gamma_i \rangle$, since these terms are the most affected by shape noise.

4.3.3 Implementation

In Figure 4.2, we illustrate our procedure. We used a ~ 900 square degree area from the MICE simulation corresponding to $0^\circ < RA < 30^\circ$ and $0^\circ < DEC < 30^\circ$. The top left panel shows the convergence map κ, located at $z \simeq 1$. The top right panel shows the constructed convergence template, κ_g, derived via equation (4.19). Both maps have been smoothed using a circular top hat filter of 50 arcmin radius. We can see that κ_g is a biased version of κ at large scales. In the bottom panel we show the scatter plot of κ versus κ_g, using pixels of 7 arcmin of side in each map. The bias b shown in the plot is estimated via equation (4.20), and the error corresponds to the Jack-Knife errors from equation (4.22). In red, we show a line crossing the origin and with the slope corresponding to this estimated bias. We have checked that the b value derived from the zero-lag statistics is in agreement with a linear fit to the scatter plot at the 0.1% level. This is another indication that we are in the linear regime, where we can assume equation (4.5).

We note that the expression for the bias from equations (4.20,4.21) assumes equation (4.5). However, κ is a projection of δ in the line-of-sight weighted by the lensing kernel, as well as κ_g. Thus, the relation between κ and κ_g is a constant that comes from the redshift dependence of bias weighted by the redshift dependence of the lensing kernel. Hence, the bias obtained in this example is a weighted mean of galaxy bias as a function of redshift.

But we can take this dependence into account to measure bias at different redshifts using tomography as we explain in §4.3.5 below.

4.3.4 Numerical effects and parameters

There are different parameters that can affect our implementation presented in §4.3.2. We have studied in which regime our method is valid, or consistent with the linear bias from equation (4.3), and what are the dependences when it is not valid. With this, we can either calibrate our results or restrict to the regimes where our bias measurement is carried out. Here we describe the main numerical effects and our choice of parameters for our implementation.

Catalogue selection

We used an area of $0^o < RA, DEC < 30^o$. This is the same area we used for the fiducial bias measurements from equation (4.3), so that our comparison of both bias is not affected by differences in area or sample variance. This area is similar to DES Y1 data, so this study can be seen as an estimation of the theoretical limitations of this method on DES Y1.

We apply a magnitude cut for the foreground galaxies of $i < 22.5$, to be able to compare it with measurements in the DES SV data [59]. However, other selections can be done for this method, such us selecting galaxies by colour or luminosity, in order to measure colour and luminosity dependent bias, that would give information about galaxy formation and evolution.

Redshift bin width

We use redshift bins of $\Delta z = 0.2$ for the foreground galaxies. In this analysis we use the true redshift from the simulation, but in data this method would be also affected by photo-z errors.

For the choice of Δz we need to take into account two effects. On one side, the use of wide redshift bins would mean losing information from the small scale fluctuations of δ_g in the line-of-sight, since we project the galaxies in the same bin to measure δ_g. We have seen that this produces a deviation in the value of galaxy bias that is larger than 5% for $\Delta z > 0.2$, and it can be larger than 10% for $\Delta z > 0.3$. We explore this in Figure (4.5) and in §4.4.1. We take this effect into account when we estimate bias in tomographic bins at the end of the chapter. When we have photo-z errors, the redshift binning effect is not as important as for the ideal case. If the photo-z errors dominate, the dilution of the small scale fluctuations

come from the photo-z errors, and the redshift binning does not affect much. We address the effects of photo-z errors in [44].

On the other hand, the use of narrow redshift bins requires a smoothing of the estimation of $\bar{\rho}_g(z)$. If we calculate $\bar{\rho}_g$ for each redshift bin alone, for narrow bins $\bar{\rho}_g(z)$ is affected by the structure fluctuation in each particular redshift bin, and this causes a smoothing in the final estimation of δ_g. Some smoothing of $\bar{\rho}_g$ in redshift is needed to avoid this effect when using narrow bins. This is relevant for $\Delta z < 0.03$.

Angular scale

To generate the maps we pixelize the sky using a sinusoidal projection (which consists on redefining RA as $(RA-15)\cos(DEC)$ in order to obtain a symmetric map with pixels of equal area) with an angular resolution of 50 arcmin, so that the area of the pixels is $(50\ \text{arcmin})^2$. Then galaxies are projected in different redshift bins according to their true redshift.

The bias estimated from this method is not necessarily consistent with the bias from equation (4.3) at small scales. These two methods are only expected to agree at large scales, in the linear bias regime. Moreover, this method requires a projection in the line-of-sight, so that different scales (weighted differently according to the lensing kernel) are mixed for the same angular scale. However, we have seen that bias is constant for angular scales larger than $\Theta \gtrsim 30$ arcmin, meaning that linear scales are dominant in this regime. In Figure 4.3 we show the agreement of galaxy bias between equations (4.3) and (4.6-4.8) when we use a pixel scale of 50 arcmin, as a visual example of this.

Smoothing

An alternative way to calculate the maps in a given smoothing scale is possible by using small pixels and applying a smoothing kernel of the corresponding scale to these pixels, instead of directly using large pixels. This smoothing scheme has two advantages. First of all, the area that can be used is optimized, since the pixels affected by the edges are only the closest to the mask (see Fig. 4.9). This is important for small areas and irregular masks. The other advantage of this scheme is that the maps produced give also a good visual image of the structures and how the field changes in the sky. However, the JK estimator has to be rescaled to obtain the correct errors, since neighbour JK subsamples are correlated due to the smoothing kernel. We address this in [44].

The results for a given scale are equivalent using this smoothing scheme or just enlarging the pixels, so in this chapter we do not use any smoothing kernel in order to avoid the need of calibrating the JK estimator. For irregular masks, as in [44], smoothing allows to

optimize the area used, and because of this we apply a smoothing kernel into very small pixels.

Exceptionally, in Figure 4.2 we apply the second scheme, and we use pixels of 7 arcmin and we apply a Top Hat filter of 50 arcmin to smooth the field. We do this only in this figure in order to have a better visibility of the structures of the maps and the shape of the area used. For the rest of the analysis of the chapter, we use pixels of 50arcmin^2 and no smoothing kernel afterwards. We then obtain the values of bias, and we do not need to rescale the JK errors obtianed.

Edge effects

We use a limited area and we project the sky to obtain the maps. When we pixelize the map with a definite pixel scale, due to the projection and the shape of the area used, part of the pixels in the edges are partially affected by the edges. We exclude these pixels from the analysis.

When a smoothing kernel is applied to the pixelized map, the pixels that are close to the edges are also affected by them. We exclude the pixels whose distance to the edges is smaller than the smoothing radius.

Source redshift

We estimate the κ field at $z \simeq 1.3$ by calculating the mean κ of the source galaxies with $1.2 < z < 1.4$ in each pixel. The redshift range used ensures we have enough density of galaxies to correctly calculate κ.

Theoretically one should take into account the redshift distribution of the source galaxies so that each galaxy contributes to κ_g with its position χ_s. However, approximating these galaxies to a plane in their mean position at $z \simeq 1.3$ causes less than a 1% effect.

Foreground galaxy redshift range

Equation (4.19) is strongly dependent on the redshift range used for the foreground galaxies. If we use only a partial redshift range for the construction of κ_g, the values of κ_g obtained have a lower amplitude, since in the sum we are missing the contribution coming from the unused redshift range. Moreover, as bias depends on redshift, using a wide redshift range for the foreground galaxies involves averaging this redshift dependent bias in the final result. However, this can be corrected for and used to obtain the bias in tomographic bins, as discussed in §4.3.5.

We use single redshift bins of $\Delta z = 0.2$ for the foreground galaxies in the range of $0.2 < z < 1.2$ to estimate the bias in each of these bins. This produces a galaxy bias estimation of 5 points in the whole redshift range available (for this method) in the simulation.

4.3.5 Redshift dependence

This method involves an integral (or a sum in practice) along the redshift direction, and because of this the bias obtained is a weighted average of the redshift dependent bias. However, we can estimate galaxy bias in a given redshift bin if we restrict the calculation to the foreground galaxies in that redshift bin, assuming that bias does not change significantly in the bin. If this is the case, we can measure the redshift-dependent bias using tomographic redshift bins.

Since κ_g is obtained from the contribution of all the galaxies in front of the sources, if we restrict the redshift range for the calculation of κ_g we need to renormalize the result by taking into account the contribution from the unused redshift range. Here is a description of the correction that we apply to estimate redshift dependent bias using tomographic bins.

Taking into account the sum from equation (4.19), and using δ instead of δ_g, we have:

$$\kappa(\boldsymbol{\theta}) = \sum_{all\ lens\ bins} \bar{q}'\delta'(\boldsymbol{\theta})\Delta\chi', \tag{4.25}$$

where we remind the reader that $\bar{q}'$ is the weak-lensing efficiency kernel of each bin. If we only use the foreground galaxies (or the dark matter field) in a single redshift bin between comoving coordinates χ_{min} and χ_{max}, then we call this partial convergence field κ', where

$$\kappa'(\boldsymbol{\theta}) = \bar{q}'\Delta\chi'\bar{\delta}' = \bar{q}'\Delta\chi' \int d\chi p'(\chi)\delta(\boldsymbol{\theta},\chi), \tag{4.26}$$

and

$$\bar{q}' = \int_{\chi_{min}}^{\chi_{max}} d\chi \frac{q(\chi)}{\Delta\chi}. \tag{4.27}$$

$\Delta\chi = \chi_{max} - \chi_{min}$, and $p'(\chi)$ is the radial selection function (constant for the dark matter field in comoving coordinates, since the dark matter density is constant in these coordinates), normalized to 1 and restricted to the bin $\chi_{min} < \chi < \chi_{max}$. To simplify the notation, when the limits are not specified in the integral, the integral will go through the whole range between 0 and ∞. Note that, as $p'(\chi) = 0$ for all χ outside the bin, only the range $\chi_{min} < \chi < \chi_{max}$ contributes to the integral in equation (4.26), and $p'(\chi)$ implies a projection inside the bin. The factor $\bar{q}'$ appears to be outside the integral $\int d\chi p'(\chi)\delta(\boldsymbol{\theta},\chi)$ when working in bins. This is exact for infinitely thin bins, and is also correct if $q(\chi)$ is not correlated with $p'(\chi)\delta(\chi)$ inside the bin. So, to summarize, the expression from equation (4.26) is affected by the projection in the bin, and the correlation between $q(\chi)$ and $p'(\chi)\delta(\chi)$.

For our purpose we are interested in the factors $\langle \kappa' \kappa \rangle$, $\langle \kappa' \kappa' \rangle$ and $\langle \kappa \kappa \rangle$ to be able to measure galaxy bias in tomographic redshift bins. According to these definitions, together with equation (4.15), we have:

$$\kappa' \kappa(\Theta) = \bar{q}' \Delta\chi' \int p'(\chi_1) d\chi_1 \int_0^{\chi_s} d\chi_2 q(\chi_2) \xi(r_{12}) \tag{4.28}$$

$$\kappa' \kappa'(\Theta) = (\bar{q}' \Delta\chi')^2 \int p'(\chi_1) d\chi_1 \int d\chi_2 p'(\chi_2) \xi(r_{12}) \tag{4.29}$$

$$\kappa \kappa(\Theta) = \int_0^{\chi_s} q(\chi_1) d\chi_1 \int_0^{\chi_s} q(\chi_2) d\chi_2 \xi(r_{12}), \tag{4.30}$$

with $r_{12}^2 = \chi_1^2 + \chi_2^2 + 2\chi_1\chi_2 \cos\theta$, $\xi(r_{12})$ is the 2PCF and Θ is the angular separation.

The quantities we are interested in are the ratios:

$$f_1 = \frac{\langle \kappa' \kappa \rangle}{\langle \kappa \kappa \rangle} \tag{4.31}$$

and

$$f_2 = \frac{\langle \kappa' \kappa' \rangle}{\langle \kappa' \kappa \rangle} \tag{4.32}$$

For the general case,

$$\langle \kappa_A \kappa_B \rangle = \frac{4\pi}{\pi^2 R^4} \int_0^R dr_1 r_1 \int_0^R dr_2 r_2 \int_0^{\pi} d\eta \, \omega_{AB}(\Theta), \tag{4.33}$$

where $\Theta^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos\eta$, κ_A and κ_B can be κ, κ', κ_g or κ_g', η is the angular separation between the vectors $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ and $\omega(\Theta)$ is the two-point angular correlation function of the two fields A and B, defined as

$$\omega_{AB}(\Theta) = \int_0^{\infty} d\chi_A \int_0^{\infty} d\chi_B q(\chi_A) q(\chi_B) p'(\chi_A) p'(\chi_B) \xi_{AB}(r), \tag{4.34}$$

where $p'(\chi_{A,B})$ are the corresponding selection functions of the fields A and B, and $\xi_{AB}(r)$ is the 3D two-point cross-correlation function, that in this case corresponds to the dark matter $\xi(r)$.

Equations (4.31,4.32) can be obtained analytically and they are weakly dependent on cosmology (the only dependence comes from the ratios between the lensing kernels). Equations (4.31,4.32) describe the contribution of these zero-lag correlations of κ in a given redshift bin for the dark matter field. As the dark matter field has a bias of 1 by definition, using the galaxies instead of the dark matter field to compute these equations would give $b' f_{1,2}$ instead of $f_{1,2}$, where b' is the galaxy bias in the redshift bin used (assuming that galaxy bias is constant inside the redshift bin). Then, to estimate galaxy bias in these bins, we need to obtain the bias from equations (4.20,4.21) using only the galaxies of these bins, and then rescale the bias according to the values of f_1 (or f_2) as described here:

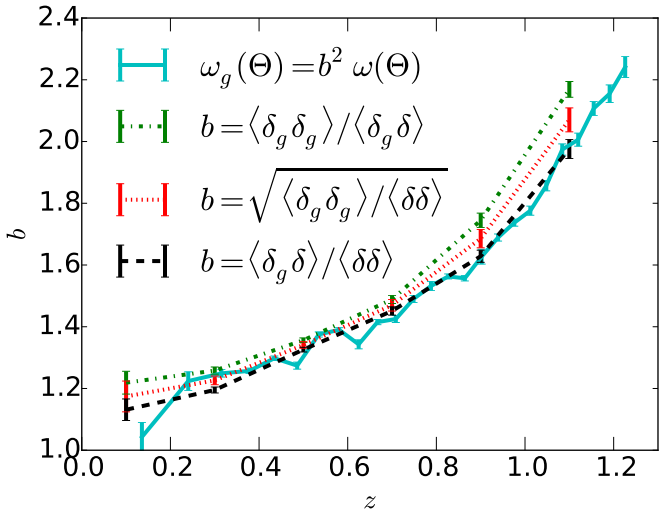

Figure 4.3: Comparison of different definitions of bias. Solid cyan line shows the bias as defined in equation (4.3). The dashed black, dash-dotted green and dotted red lines show bias according to the different definitions from equations (4.6-4.8).

$$b' = \frac{1}{f_1} \frac{\langle \kappa'_g \kappa \rangle}{\langle \kappa \kappa \rangle - \langle \kappa^N \kappa^N \rangle} \tag{4.35}$$

$$b' = \frac{1}{f_2} \frac{\langle \kappa'_g \kappa'_g \rangle - \langle \kappa'^N_g \kappa'^N_g \rangle}{\langle \kappa'_g \kappa \rangle}, \tag{4.36}$$

where κ'_g and κ'^N_g are obtained from the galaxies in a given redshift bin.

4.4 Results

4.4.1 Testing

In this study we test our method against a fiducial galaxy bias. For this, we measure $\omega(\Theta)$ and $\omega_g(\Theta)$ of dark matter and galaxies in the simulation for different redshift bins, using the same area and galaxies that we use for our method. We also estimate bias from the definitions in equations (4.6-4.8) in the same simulation to study the consistency between the different bias definitions.

In Figure 4.3 we compare different estimations of galaxy bias from the MICE Simulation, using an area of $0^o < RA, DEC < 30^o$. The solid cyan line represents the bias definition from equation (4.3). We measure $\omega(\Theta)$ and $\omega_g(\Theta)$ as a function of the angular scale, and to obtain the bias we fit the ratio as constant between 6 and 60 arcmin. The angular correlation function involves different comoving scales for different redshifts, and then

fixing the same angular scales for the galaxy bias implies a mix of physical scales. However, for large enough scales, bias is constant and is not affected by this. We have checked that bias is constant at these scales and in all redshift bins, an indication that we are in the linear regime. The galaxy bias obtained from equations (4.6-4.8) are shown in dashed black line, dotted red line and dash-dotted green line as specified in the legend. This has been calculated in each redshift bin by pixelating δ and δ_g in pixels of area $(50\,\mathrm{arcmin})^2$ using redshift bins of $\Delta z = 0.2$. The agreement between the solid cyan and the dashed black lines confirms that linear bias from $\omega(\Theta)$ converges to local bias at large scales. On the other hand, the differences in the different expressions of equations (4.6-4.8) implies a noise between δ_g and δ that affects our estimations of bias. The differences between these estimators can also be seen as an indirect measurement of this noise, that can come from stochasticity or other effects as projections and pixelization. We see that the same effect appears when using equation (4.39) to estimate galaxy bias, and this can be explained by the projection effect due to the redshift binning, as discussed below in Figures 4.4 and 4.5. We take into account this effect to estimate tomographic bias in §4.4.2.

For testing purposes, we construct here the bias-corrected κ_g map, $\hat{\kappa}_g$, defined as:

$$\hat{\kappa}_g(\boldsymbol{\theta}) = \sum_{all\ lens\ bins} q' \frac{\delta_g'(\boldsymbol{\theta})}{b'} \Delta\chi', \tag{4.37}$$

where b', or $b(z)$, corresponds to the linear bias that can be obtained from equations (4.3) or (4.6-4.8). In analogy with equations (4.20,4.21), we can calculate the corresponding normalized bias between the $\hat{\kappa}_g$ and κ fields:

$$\hat{b} = \frac{\langle \hat{\kappa}_g \kappa \rangle}{\langle \kappa\kappa \rangle - \langle \kappa^N \kappa^N \rangle} \tag{4.38}$$

$$\hat{b} = \frac{\langle \hat{\kappa}_g \hat{\kappa}_g \rangle - \langle \kappa_g^N \kappa_g^N \rangle}{\langle \hat{\kappa}_g \kappa \rangle}. \tag{4.39}$$

Under this definition, $\hat{b} = 1$ suggests that this method is measuring linear bias, since it is basically assuming equation (4.5).

Figure 4.4 shows how the estimator $\hat{b}$ changes as a function of the angular scale, defined by the pixel scale, using different estimators of $\hat{b}$ and $b(z)$. For the dashed red and blue lines we used $b(z)$ from equation (4.3) to obtain $\hat{b}$ from $\langle \hat{\kappa}_g \kappa \rangle / \langle \kappa\kappa \rangle$ and $\langle \hat{\kappa}_g \hat{\kappa}_g \rangle / \langle \hat{\kappa}_g \kappa \rangle$ respectively (we omit the $\langle \kappa^N \kappa^N \rangle$ factors for visual simplicity). We can see that the measurements are constant for $\Theta > 30$ arcmin, meaning that we are in the linear regime in these scales. However, there is a 5% difference between the two estimators at large scales (at small scales nonlinearities appear and the difference is larger). This can be interpreted

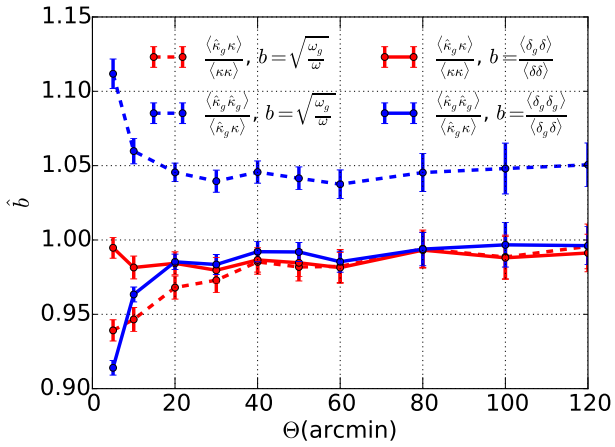

Figure 4.4: Normalized bias from the zero-lag cross correlations of κ and κ_g as a function of the angular smoothing scale, normalized by the redshift dependent bias from different estimators. Dashed red line shows $\langle\hat{\kappa}_g\kappa\rangle/\langle\kappa\kappa\rangle$, where $\hat{\kappa}_g$ has been obtained normalizing κ_g by the bias from equation (4.3). The solid red line shows the same, but normalizing κ_g from the bias obtained from $\langle\delta_g\delta\rangle/\langle\delta\delta\rangle$. For the dashed blue line, $\langle\hat{\kappa}_g\hat{\kappa}_g\rangle/\langle\hat{\kappa}_g\kappa\rangle$ has been obtained by normalizing $\hat{\kappa}_g$ by bias from equation (4.3). The solid blue line shows the same, but normalizing $\hat{\kappa}_g$ by the bias from $\langle\delta_g\delta_g\rangle/\langle\delta_g\delta\rangle$

from Figure (4.3), where we see that the estimators $\langle\delta_g\delta\rangle/\langle\delta\delta\rangle$ and $\langle\delta_g\delta_g\rangle/\langle\delta_g\delta\rangle$ are slightly different. In fact, $\langle\hat{\kappa}_g\kappa\rangle/\langle\kappa\kappa\rangle$ is indirectly measuring $\langle\delta_g\delta\rangle/\langle\delta\delta\rangle$, which is consistent with bias from equation (4.3) (at the 1% level), while $\langle\hat{\kappa}_g\hat{\kappa}_g\rangle/\langle\hat{\kappa}_g\kappa\rangle$ is indirectly measuring $\langle\delta_g\delta_g\rangle/\langle\delta_g\delta\rangle$, which is slightly higher than bias from equation (4.3). If we use $b=\langle\delta_g\delta\rangle/\langle\delta\delta\rangle$ for the bias normalization of $\langle\hat{\kappa}_g\kappa\rangle/\langle\kappa\kappa\rangle$ (shown in the solid red line) and $b=\langle\delta_g\delta_g\rangle/\langle\delta_g\delta\rangle$ for the bias normalization of $\langle\hat{\kappa}_g\hat{\kappa}_g\rangle/\langle\hat{\kappa}_g\kappa\rangle$ (shown in the solid blue line), then both estimations are consistent, as expected. As in Figure 4.3 for b, the difference between both estimators of $\hat{b}$ coming from this test can be seen as an indication (and a measurement) of the noise in the relation between δ_g and δ, giving a factor of 5%.

In order to go deeper in the analysis of these numerical effects and see whether these differences between both estimators come from the intrinsic relation between δ_g and δ or from numerical systematics, we constructed the following template κ_m:

$$\kappa_m(\boldsymbol{\theta})=\sum_{i=0}^{N} q'\delta'(\boldsymbol{\theta})\Delta\chi', \tag{4.40}$$

which corresponds to the same exact calculation than equation (4.19) for κ_g, but using dark matter particles instead of galaxies. This field κ_m is expected to reproduce κ exactly except for the numerical differences between the method and how the original κ is obtained, which basically come from the redshift binning and projection discussed below equations

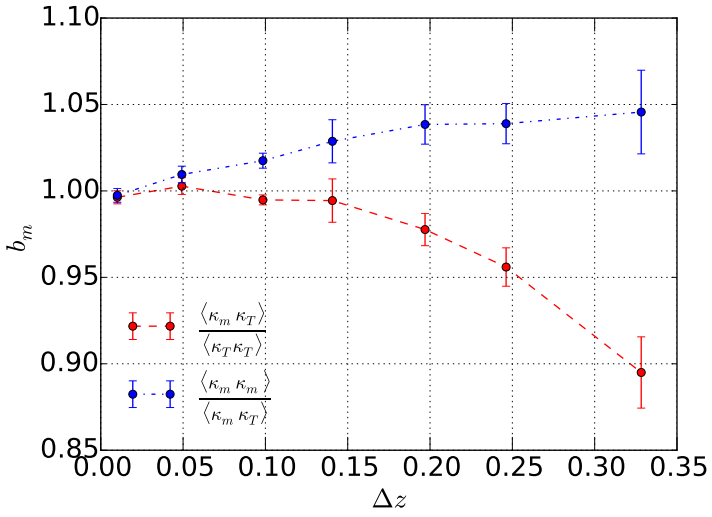

Figure 4.5: b_m, defined in equations (4.41,4.42), as a function of the redshift bin width used, Δz, for the two b_m estimators.

(4.19,4.27). In order to avoid noise in the κ map, we use κ_T, defined as the true map directly obtained from the high resolution map of the simulation [79,80,89], and calculated the bias of these two estimators of κ as:

$$b_m = \frac{\langle \kappa_m \kappa_T \rangle}{\langle \kappa_T \kappa_T \rangle} \tag{4.41}$$

$$b_m = \frac{\langle \kappa_m \kappa_m \rangle - \langle \kappa_m^N \kappa_m^N \rangle}{\langle \kappa_m \kappa_T \rangle}, \tag{4.42}$$

that should give $b_m = 1$ if there are no numerical systematics.

We have found that b_m behaves as $\hat{b}$ in our tests, meaning that the differences between the different estimators can be seen as a measurement of the numerical effects on the method. In fact, we have found that the differences mainly come from the projection effect in the redshift bins, as shown in Figure 4.5. Here we show the two estimators of b_m as a function of the redshift bin width, Δz. We use a pixel scale of 50 arcmin, a source redshift of $z_s = 1$ and we use all the dark matter particles (diluted with respect to the total number of particles, but this does not affect the result) within $z < 1$. We see that the two estimators agree when we use narrow redshift bins, but the difference between both increases with Δz. For $\Delta z = 0.2$, the difference is the 5% that we see in Figure 4.4 for the galaxies. This test measures the redshift binning and the projection impacts on this method, and can also be used to calibrate the measurements. In fact, f_1 and f_2 can be used to take into account these projections, specified by the selection function $p'(\chi)$, and the redshift binning. But in the case of Figures 4.4 and 4.5, we use all the redshift range in the foreground of the sources

and we have not corrected by f_1 and f_2. In the case of Figure 4.4, instead of using f_1 and f_2, we correct for this effect by using the bias which is estimated using the same redshift bin width as in the κ estimates. For narrow redshift bins these corrections are negligible. In the next section we will apply the f_1 and f_2 corrections to the tomographic estimations.

4.4.2 Redshift dependent bias

In Figure 4.6 we show a comparison between the theoretical predictions (in dashed black lines) of f_1 and the measurements in the MICE simulation (in green points) of $\langle \hat{\kappa}'_g \kappa \rangle / \langle \kappa \kappa \rangle$, in 6 different redshift bins of $\Delta z = 0.2$, using a redshift for the sources of $z_s = 1.3$. Here $\hat{\kappa}'_g$ is obtained from equation (4.37) but restricting the galaxies to each bin. To obtain the values for the simulation, we computed $\hat{\kappa}'_g$ in the corresponding bins, using $b = \langle \delta_g \delta \rangle / \langle \delta \delta \rangle$ for the normalization of κ_g. Then, $\hat{\kappa}'_g$ can be seen as an estimator of κ obtained from δ_g / b. We see a good agreement between theory and simulations. Note that the amplitude of f_1 is higher at the intermediate redshifts, due to the contribution of the lensing kernel, but this curve also reflects effects such as the projections due to the binning (so the fact that we ignore that $q(\chi)$ and $p'(\chi)\delta(\chi)$ might be correlated inside the bin), the correlation functions of different distances (so the fact that $\langle \kappa' \kappa \rangle$ has a contribution coming from the correlation between the dark matter distribution inside and outside the bin) or the redshift dependence of the smoothing scale of $\omega(\Theta)$ (different redshifts have different smoothing comoving scales). The final amplitude corresponds to f_1, so this reflects the contribution to $\hat{b}$ of each of these redshift bins.

Equations (4.35,4.36) give a tool that can be used for tomographic measurements of galaxy bias, since we can estimate the bias using different redshift bins of the foreground galaxies if we take this correction into account. That is, we can measure b' for a given redshift by calculating κ'_g in that bin and using equations (4.35,4.36).

Figure 4.7 shows the estimation of the tomographic bias using different redshift bins of $\Delta z = 0.2$ for both estimators from equations (4.35,4.36), represented as blue and red points as specified in the legend. We compare them with the fiducial bias from equations (4.3,4.6,4.7) shown in solid cyan, dashed black and dash-dotted green lines respectively. We see that the method we present in this chapter gives consistent results with linear bias. There are some slight differences for the estimator from equation (4.7) which, as mentioned above, is due to the effects of projection and binning. But this effect is not shown from the tomographic bias obtained from our method, because we take into account these effect in the factors f_1 and f_2. Note also that the two methods, represented by the red and blue points, give very similar results (apart from the fourth bin).

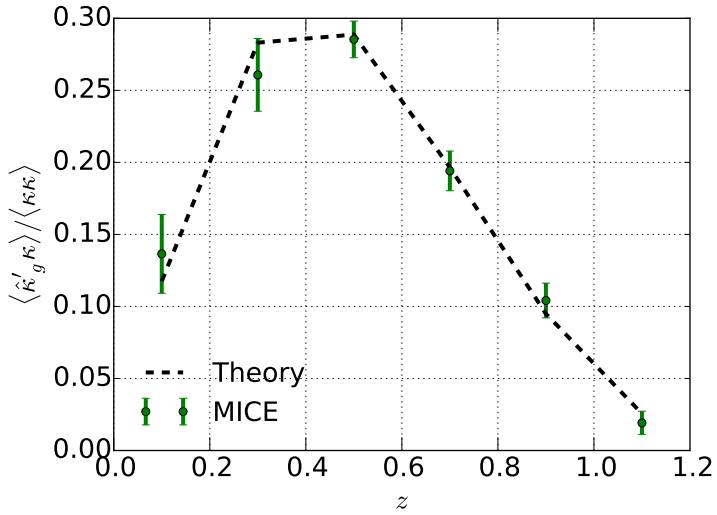

Figure 4.6: Comparison of $\langle \kappa' \kappa \rangle / \langle \kappa\kappa \rangle$ between theory (dashed black line) and simulation (green points). Each value has been obtained by using redshift bins of $\Delta z = 0.2$ to calculate κ', and using a source redshift of $z_s = 1.3$.

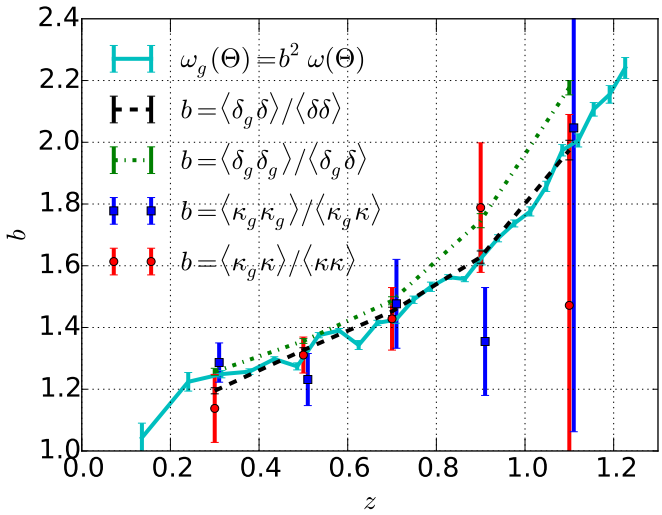

Figure 4.7: Redshift dependent bias estimated form our method, shown in red and blue points as specified in the legend. For this we used tomographic redshift bins of $\Delta z = 0.2$ and a source redshift of $z_s = 1.3$. The solid cyan line shows linear bias from equation (4.3), fitting bias as constant between 6 and 60 arcmin. The dashed black line shows bias estimated from $\langle \delta_g \delta \rangle / \langle \delta\delta \rangle$, using the same redshift bins of $\Delta z = 0.2$. The dash-dotted green line shows bias estimaed from $\langle \delta_g \delta_g \rangle / \langle \delta_g \delta \rangle$ in the same redshift bins.

We can see that the errors are very large for the highest redshift bin. This is due to the fact that, due to the lensing kernel, f_1 and f_2 are very small, and then the measurements in this bin are very sensitive to small changes. The best error bars appear where the lensing kernel is higher, so the potential of this method is optimal in the maximum of the lensing kernel. Hence, different source redshifts might be combined in order to optimize the analysis for all redshifts. In [44] we combine the results using multiple redshift bins for the source galaxies, and we fit the galaxy bias from the combination of these measurements, using both κ_g and γ_g and doing a full-covariance analysis. In this chapter we do not apply any fit, since we directly measure bias from equations (4.35,4.36) using a fixed source redshift bin.

This study proves that we can use this method to measure linear bias, since our method is consistent with the linear bias measured from equation (4.3). This method has some advantages with respect to other estimators of bias. First of all, it can be applied to observations, as discussed in §4.5. In this case, we do not need to assume a dark matter distribution, since we can obtain this information from the weak lensing maps, and hence this method is able to measure bias from the direct comparison between the galaxy and dark matter distributions. Moreover, this method does not depend strongly on σ_8, as most of the common methods to measure bias. However, it depends on Ω_m, but it is weakly dependent on the other cosmological parameters. The combination of this method with other estimators of bias that depend on σ_8 allow us to constrain at the same time bias and cosmology. Because of this, this method is a very good complement to other measurements of bias.

4.5 Application to data

When applying this method to data, we need to take into account other aspects for the measurements. First of all, we cannot measure κ directly from observations, since the lensing information comes from the ellipticity of the galaxies. Then, we need to obtain γ from the measured ellipticities, assuming some shape noise, and then convert γ to κ or κ_g to γ_g to measure bias from equations (4.20,4.21) or (4.23,4.24). [223] explored the conversion from ellipticities to κ for the DES SV data, obtaining the largest mass map from weak lensing ever observed. They also showed the consistency between the mass map and the foreground galaxy distribution. However, it is important to mention that, since shape noise is the most important source of uncertainty for this method in observations, the terms $\langle\kappa\kappa\rangle$ and $\langle\gamma\gamma\rangle$ are very noisy and are not optimal for the analysis.

Moreover, for photometric surveys we need to take into account the photo-z estimation. The uncertainty in the redshift of the galaxies causes an smoothing of κ_g that has to be

calibrated from the understanding of the photo-z errors, as well as from the distribution of the galaxies in redshift. This has to be taken into account when defining the selection functions in the calculation of f_1 and f_2.

Other aspects from observations, such as the mask and the shape measurements, can affect our measurements and must be taken into account. In this section we show the main results and a summary of [44], where we apply this method to DES SV data to measure galaxy bias.

4.5.1 Bias estimation from the galaxy density field and the weak lensing field

The information of galaxy bias can be extracted through the cross- and auto-correlation of the κ and κ'_g fields. In this case, we calculate

$$b' = \frac{\langle \kappa'_g \kappa'_g \rangle}{\langle \kappa'_g \kappa \rangle} \tag{4.43}$$

For infinitely thin redshift bins, or constant bias, b' in equation (4.43) directly measures the galaxy bias b of the lens. However, once the lens and source samples span a finite redshift range (see e.g. Figure 4.8), b' is a function of the source p_s and lens ϕ' distribution and is different from b by some factor $f(\phi', p_s)$, so that

$$b' = f(\phi', p_s) b. \tag{4.44}$$

Note that f can be determined if $b(z)$ is known. Since we have $b(z) = 1$ for the case of dark matter, we can calculate f by calculating b' and setting $b(z) = 1$, or

$$f(\phi', p_s) = \frac{\langle \kappa' \kappa' \rangle}{\langle \kappa' \kappa \rangle} = \frac{\langle \kappa'(\theta, \phi', \bar{\chi}_s) \kappa'(\theta, \phi', \bar{\chi}_s) \rangle}{\langle \kappa'(\theta, \phi', \bar{\chi}_s) \kappa(\theta, p_s) \rangle}, \tag{4.45}$$

where f here corresponds to f_2 in equation (4.32).

We use a slightly different estimator for b' compared to equation (4.43) in practice. Combined with equation (4.44), our estimator for galaxy bias is:

$$b = \frac{1}{f} \frac{\langle \gamma'_{\alpha,g} \gamma'_{\alpha,g} \rangle - \langle \gamma'^N_{\alpha,g} \gamma'^N_{\alpha,g} \rangle}{\langle \gamma'_{\alpha,g} \gamma'_{\alpha} \rangle - \langle \gamma'^N_{\alpha,g} \gamma'^N_{\alpha} \rangle}, \tag{4.46}$$

where $\alpha = 1, 2$ refers to the two components of γ.

Here we replaced κ' by γ'_α, which is possible since the two quantities are interchangeable through equation (4.16). In this work we follow the implementation of equation (4.16) as described in [223] and [45] to construct κ and γ maps as needed. The main reason to

work with γ'_α is that in our data set, γ'_α is much noisier compared to the κ'_g due to the presence of the shape noise, therefore converting γ'_α to κ'_α would be suboptimal to converting κ'_g to $\gamma'_{\alpha,g}$. This choice depends somewhat on the specific data quality at hand. In addition, the term $\langle \gamma'^N_{\alpha,g} \gamma'^N_{\alpha,g} \rangle$ is introduced to account for the shot noise arising from the finite number of galaxies in the galaxy density field. The term is calculated by randomizing the galaxy positions when calculating $\gamma'_{\alpha,g}$. Similarly, the $\langle \gamma'^N_{\alpha,g} \gamma'^N_{\alpha} \rangle$ term in the denominator is a small correction that accounts for any spurious correlation that can come from the mask.

The measurement from this method would depend on assumptions of the cosmological model in the construction of κ'_g and the calculation of f. Except for the literal linear dependence on $H_0\Omega_{\rm m}$, due to the ratio nature of the measurement, most other parameters tend to cancel out. Within the current constraints from *Planck*, the uncertainty in the cosmological parameters affect the measurements at the percent level, which is well within the measurement errors ($> 10\%$). All cosmological parameters used in the calculation of this work are consistent with the MICE GC simulations.

4.5.2 Multiple source-lens samples

We define several source and lens samples, or "bins", based on their photometric redshift (photo-z), with the lens samples labeled by i and the source samples labeled by j. We use the notation b^α_{ij} to represent the bias measured with γ_α using the source bin j and lens bin i.

Our final estimate of the redshift-dependent galaxy bias and its uncertainty is calculated by combining b^α_{ij} estimates from the two components of shear and all source redshift bins j. We estimate it by taking into account the full covariance between all the measurements of the galaxy bias in the same lens bin i. Our final estimate $\bar{b}_i$ and uncertainty $\sigma(\bar{b}_i)$ quoted are derived via:

$$\bar{b}_i = M_i^T C_i^{-1} D_i [M_i^T C_i^{-1} M_i]^{-1}, \tag{4.47}$$

$$\sigma(\bar{b}_i)^2 = (M_i^T C_i^{-1} M_i)^{-1}, \tag{4.48}$$

where D is a one dimensional array containing all the measurements $b^i_{\alpha,j}$ of galaxy bias in this lens bin i (including measurement from the two shear components and possibly multiple source bins), M is a 1D array of the same length as D with all elements being 1, and C^{-1} is the unbiased inverse covariant matrix between all $b^i_{\alpha,j}$ measurements, as estimated by Jack-Knife (JK) resampling [100]:

$$D_i = \{b^\alpha_{ij}\}, \text{ all possible } \alpha, j \tag{4.49}$$

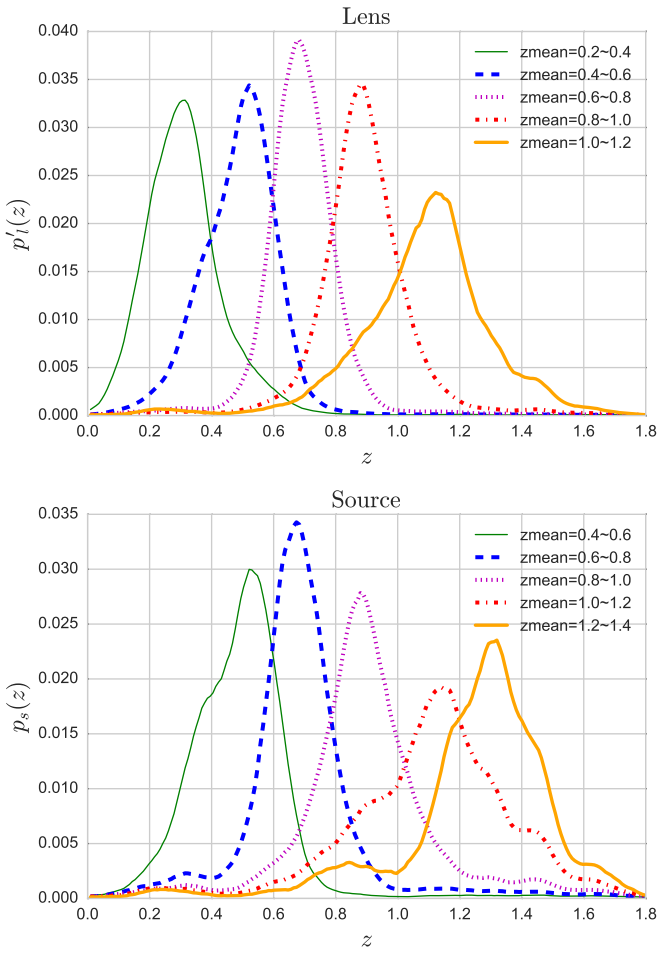

Figure 4.8: Normalized redshift distribution of the lens (top) and source (bottom) samples as estimated from the photo-z code SKYNET. Each curve represents the stacked PDF for all galaxies in the photo-z bin determined by $z_{\rm mean}$ as listed in the labels.

$$C_i^{-1} = \tau\,\mathrm{Cov}^{-1}[D_i], \tag{4.50}$$

where $\tau = (N-\nu-2)/(N-1)$. N is the number of JK samples, and ν is the dimension of C. Note that the matrix inversion of C_i becomes unstable when the measurements b_{ij}^{α} are highly correlated. This is the case in the noiseless simulations, which we will thus only show the weighted mean of the measurements for our results (Figure 4.11). For the noisy simulations and data, we quote equation (4.47) and equation (4.48) as our results. Also note that in the estimators $\bar{b}_i$ and $\sigma(\bar{b}_i)$ we have not accounted for the correlation between different lens bins i. As we discuss in §4.5.4.1, since the main contribution of the covariance comes from the coupling of the mask, noise and large-scale structures in the data, we estimate the covariance from a large number of simulations instead of using the JK method.

4.5.3 Data and simulations

We use the DES SV data collected using the Dark Energy Camera [76] from November 2012 to February 2013 and that have been processed through the Data Management pipeline described in [67,153,159,191]. Individual images are stacked, objects are detected and their photometric/morphological properties are measured using the software packages SCAMP [23], SWARP [26], PSFEx [24] and SExtractor [25]. The final product, the SVA1 Gold catalog[2] is the foundation of all catalogues described below. We use a $\sim$116.2 deg^2 subset of the data in the South Pole Telescope East (SPT-E) footprint, which is the largest contiguous region in the SV dataset. This data set is also used in other DES weak lensing and large-scale structure analyzes [12, 45, 60, 91, 214, 223].

4.5.3.1 Photo-z catalogue

The photo-z of each galaxy is estimated through the SKYNET code [93]. SKYNET is a machine learning algorithm that has been extensively tested in [182] and [37] to perform well in controlled simulation tests. To test the robustness of our results, we also carry out our main analysis using two other photo-z codes which were tested in [182] and [37]: BPZ [14], and TPZ [40,41]. We discuss in §4.5.4.3 the results from these different photo-z codes.

The photo-z codes output a PDF for each galaxy describing the probability of the galaxy being at redshift z. We first use the mean of the PDF, $z_{\rm mean}$ to separate the galaxies into redshift bins, and then use the full PDF to calculate equation (4.45). In Figure 4.8, we show the normalized redshift distribution for each lens and source bin as defined below.

4.5.3.2 Galaxy catalogue

To generate the κ_g maps, we use the same "Benchmark" sample used in [91] and [60]. This is a magnitude-limited galaxy sample at $18 < i < 22.5$ derived from the SVA1 Gold catalog with additional cleaning with color, region, and star-galaxy classification cuts (see [60] for full details of this sample). The final area is $\sim$ 116.2 square degrees with an average galaxy number density of 5.6 per arcmin^2. Five redshift bins were used from $z_{\rm mean} = 0.2$ to $z_{\rm mean} = 1.2$ with $\Delta z_{\rm mean} = 0.2$. The magnitude-limited sample is constructed by using only the sky regions with limiting magnitude deeper than $i = 22.5$, where the limiting magnitude is estimated by modelling the survey depth as a function of magnitude and

[2] https://cdcvs.fnal.gov/redmine/projects/des-sci-verification/wiki/SVA1_Gold_Catalog_v10

magnitude errors [181]. Various systematics tests on the Benchmark has been performed in [60] and [135].

4.5.3.3 Shear catalogue

Two shear catalogues are available for the DES SV data based on two independent shear measurement codes NGMIX [193] and IM3SHAPE [237]. Both catalogues have been tested rigorously in [109] and have been shown to pass the requirements on the systematic uncertainties for the SV data. Our main analysis is based on NGMIX due to its higher effective number density of galaxies (5.7 per arcmin^2 compared to 3.7 per arcmin^2 for IM3SHAPE). However both catalogues produce consistent results. We adopt the selection cuts recommended in [109] for both catalogues. This galaxy sample is therefore consistent with the other DES SV measurements in e.g., [12,214]. Similar to these DES SV papers, we perform all our measurements on a blinded catalog (for details of the blinding procedure, see [109]), and only un-blind when the analysis is finalized.

γ_1 and γ_2 maps are generated from the shear catalogues for five redshift bins between $z_{\rm mean} = 0.4$ and $z_{\rm mean} = 1.4$ with $\Delta z_{\rm mean} = 0.2$. Note part of the highest redshift bin lies outside of the recommended photo-z selection according to [37] ($z_{\rm mean} = 0.3 - 1.3$). We discard this bin in the final analysis due to low signal-to-noise (see §4.5.4.2), but for future work it would be necessary to validate the entire photo-z range used[3].

4.5.3.4 Mask

Two masks are used in this work. First, we apply a common mask to all maps used in this work, we will refer this mask as the *"map mask"*. The mask is constructed by re-pixelating the $i > 22.5$ depth map into the coarser (flat) pixel grid of 5×5 arcmin^2 we use to construct all maps. The depth mask has a much higher resolution ($nside = 4096$ `Healpix` map) than this grid, which means some pixels in the new grid will be partially masked in the original `Healpix` grid. We discard pixels in the new grid with more than half of the area masked in the `Healpix` grid. The remaining partially masked pixels causes effectively a $\sim 3\%$ increase in the total area. The partially masked pixels will be taken into account later when generating κ_g (we scale the mean number of galaxy per pixel by the appropriate pixel area). We also discard pixels without any source galaxies.

Pixels on the edges of our mask will be affected by the smoothing we apply to the maps. In addition, when performing the KS conversion, the mask can affect our results. We thus define a second *"bias mask"*, where we start from the *map mask* and further mask pixels

[3] Also note that the lowest redshift lens bin also exceeds the range that is validated. However, since the sample is brighter, the main source of error from depth variation should not be significant.

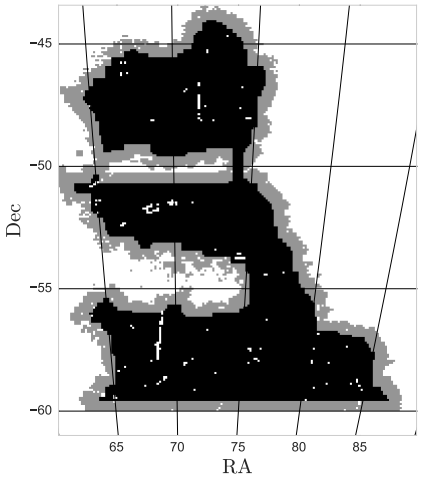

Figure 4.9: Mask used in this work. The black region shows where the galaxy bias is calculated (the *bias mask*). The black+grey map region is where all maps are made (the *map mask*).

that are closer than half a smoothing scale away from any masked pixels except for holes smaller than 1.5 pixels[4]

4.5.3.5 Simulations

In this work we use the same mock galaxy catalogue from the MICE simulations with the region 0° <RA< 30°, 0° <Dec< 30°, while we use a larger region (0° <RA< 90°, 0° <Dec< 30°) to estimate the effect from cosmic variance. In addition, we incorporate shape noise and masking effects that are matched to the data. For shape noise, we draw randomly from the ellipticity distribution in the data and add linearly to the true shear in the mock catalog to yield ellipticity measurements for all galaxies in the mock catalog. We also make sure that the source galaxy number density is matched between simulation and data in each redshift bin. For the mask, we simply apply the same mask from the data to the simulations. Note that the un-masked simulation area is ~ 8 times larger than the data, thus applying the mask increases the statistical uncertainty.

Finally, to investigate the effect of photo-z uncertainties, we add a Gaussian photo-z error to each MICE galaxy according to its true redshift. The standard deviation of the

[4]The reason for not apodizing the small masks is that it would reduce significantly the region unmasked and thus the statistical power of our measurement. We have tested in simulations that the presence of these small holes do not affect our final measurements.. We consider only pixels surviving the *bias mask* when estimating galaxy bias. Figure 4.9 shows both masks used in this work.

Gaussian uncertainty follows $\sigma(z) = 0.03(1+z)$. This model for the photo-z error is simplistic, but since we use this set of photo-z simulations mainly to test our algorithm (the calculation of f in equation (4.45)), we believe a simple model will serve its purpose.

We note that the larger patch of MICE simulation used in this work ($\sim 30 \times 30$ square degrees) is of the order of what is expected for the first year of DES data ($\sim$2,000 degree square and ~ 1 magnitude shallower). Thus, the simulation measurements shown in this work also serves as a rough forecast for our method applied on the first year of DES data.

4.5.4 Analysis and results

4.5.4.1 Simulation tests

Here we start from an ideal setup in the simulations that is very close to that used in the previous section and gradually degrade the simulations until they match our data. Below we list the series of steps we take:

1. use the full area (~ 900 deg^2) with the true γ maps
2. repeat above with photo-z errors included
3. repeat above with shape noise included
4. repeat above with SV mask applied
5. repeat above with 12 different SV-like areas on the sky, and vary the shape noise 100 times

Figure 4.10 illustrates an example of how the $\gamma_{1,g}$ and γ_1 maps degrade over these tests. The left column shows the $\gamma_{1,g}$ maps while the right column shows the γ_1 maps. Note that the color bars on the upper (lower) two maps in the right panel are 2 (4) times higher compared to the left column. This is to accommodate for the large change in scales on the right arising from shape noise in the γ_1 maps. The first row corresponds to (i) above, and we can visually see the correspondence of some structures between the two maps. Note that the $\gamma_{1,g}$ map only contributes to part of the γ_1 map, which is the reason that we do not expect even the true $\gamma_{1,g}$ and γ_1 maps to agree perfectly. The second row shows the map with photo-z errors included, corresponding to the step (ii). We find that the real structures in the maps are smoothed by the photo-z uncertainties, lowering the amplitude of the map. The smoothing from the photo-z is more visible in the $\gamma_{1,g}$ map, since the γ_1 map probes an integrated effect and is less affected by photo-z errors. The third row shows what happens when shape noise is included, which corresponds to the step (iii) above. We find the structures in the γ_1 map

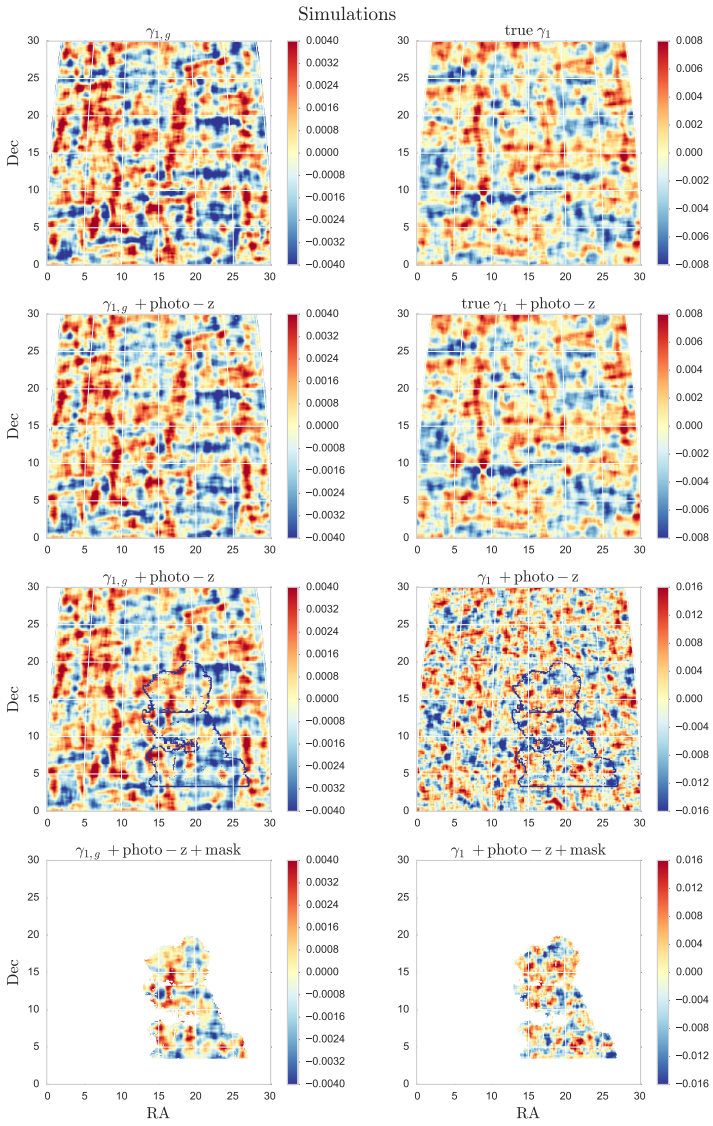

Figure 4.10: Example of simulation maps used in this work. The left column show $\gamma_{1,g}$ maps and the right column show γ_1 maps. This $\gamma_{1,g}$ maps are generated from the source redshift bin z (or $z_{\rm mean}$)$= 1.0-1.2$ and the lens redshift bin z (or $z_{\rm mean}$)$= 0.4-0.6$. The γ_1 maps are generated from the source redshift bin z (or $z_{\rm mean}$)$= 1.0-1.2$. The galaxy bias for the lens galaxies can be measured by cross-correlating the left and the right column. From top to bottom illustrates the different stages of the degradation of the simulations to match the data. The first row shows the $\gamma_{1,g}$ map against the true γ_1 map for the full 30×30 deg^2 area. The second row shows the same maps with photo-z errors included, slightly smearing out the structures in both maps. The third row shows the same $\gamma_{1,g}$ map as before against the γ_1 that contains shape noise, making the amplitude higher. Finally, the bottom row shows both maps with the SV mask applied, which is also marked in the third row for reference. Note that the color scales on the γ_1 maps is 2 (4) times higher in the upper (lower) two panels than that of the $\gamma_{1,g}$ maps.

becomes barely visible in the presence of noise, with the amplitude much higher than the noiseless case as expected. The bottom row corresponds to the step (iv) above, where the SV mask is applied to both maps. For the γ_1 map this is merely a decrease in the area. But for the $\gamma_{1,g}$ map, this also affects the conversion from κ_g to γ_g, causing edge effects in the $\gamma_{1,g}$ map which are visible in the bottom-left map in Figure 4.10. Step (v) is achieved by moving the mask around and drawing different random realizations of shape noise for the source galaxies.

With all maps generated, we then calculate the redshift-dependent galaxy bias following equation (4.47) and equation (4.48) for each of the steps from (i) to (v). In Figure 4.11 we show the result for the different stages, overlaid with the bias from the 2PCF measurement described in §4.4.1. In step (i), our measurements recover the 2PCF estimates, confirming the results from Figure 4.7, that we can indeed measure the redshift-dependent bias using this method under appropriate settings. Our error bars are smaller than that from Figure 4.7, which is due to the fact that we have combined measurements from several source bins. Since the only difference between this test and the test from Figure 4.7 is the inclusion of the KS conversion, we have also shown that the KS conversion in the noiseless case does not introduce significant problems in our measurements. Although it could be the reason for the slightly lower measurement at redshift bin $0.8-1.0$. The error bars on the highest redshift bin is large due to the small number of source and lens galaxies. In step (ii), we introduce photo-z errors. We find that the photo-z errors do not affect our measurements within the measurement uncertainties. In step (iii), the error bars increase due to the presence of shape noise. In (iv), we apply the SV mask, which causes the error bars to expand and the overall measurement to be lower by slightly more than 1σ. The larger error bars come from the smaller area and the field-to-field variation. The lowered mean value is the result of the noisiness of the denominator of our estimator (equation (4.46)) due to the presence of shape noise and the KS errors from the complicated geometry of the mask. We have additionally checked that without shape noise, the mask alone does not introduce such a strong drop. In (v), we repeat (iv) on 12 different SV-like areas in a larger (30×90 deg^2) simulation area and vary the shape noise 100 times for each. The total 1,200 simulations gives an estimate of the full covariance due to all the measurement effects considered above, and in addition, the coupling between the large-scale structure and the mask geometry. The red points in Figure 4.11 shows the mean and standard deviation of the 1,200 measurements.

The difference between the 2PCF measurements and (v) (which we refer to as Δb) as well as the error bars of (v) (which we refer to as $\sigma(\Delta b)$) will later be used to calibrate the data measurements in §4.5.4.2. We add Δb to the measured values in data, and add in quadrature $\sigma(\Delta b)$ to the uncertainty in our measurements on data. This calibration scheme

Table 4.1: Bias measurement and 1σ error bars from DES SV using the method tested in this work, with all possible lens-source combinations. We also compare here our main measurements with that using alternative shear and photo-z catalogues. Finally we compare our results with other measurement methods carried out on the same data set. The C15 estimates are from Tables 3 in that paper, while the G15 estimates are from Table 2 in that paper.

	Lens redshift ($z_{\rm mean}$)			
	$0.2-0.4$	$0.4-0.6$	$0.6-0.8$	$0.8-1.0$
This work (NGMIX+SKYNET)	1.33 ± 0.18	1.19 ± 0.23	0.99 ± 0.36	1.66 ± 0.56
This work (IM3SHAPE+SKYNET)	1.37 ± 0.23	1.39 ± 0.29	1.20 ± 0.41	1.47 ± 1.0
This work (NGMIX+TPZ)	1.34 ± 0.18	1.23 ± 0.23	1.00 ± 0.35	1.60 ± 0.55
This work (NGMIX+BPZ)	1.24 ± 0.18	1.16 ± 0.23	1.13 ± 0.37	1.41 ± 0.52
C15 [60]	1.07 ± 0.08	1.24 ± 0.04	1.34 ± 0.05	1.56 ± 0.03
G15 [91]	0.57 ± 0.25	0.91 ± 0.22	0.68 ± 0.28	1.02 ± 0.31

assumes that we have incorporated the dominant sources of statistical errors in our simulations, which we believe is a valid assumption.

With the series of simulation tests above, we have shown that although the measurement method itself is well grounded, the presence of measurement effects and noise can affect our final results from the data. In the following section we will use what is learned in this section to interpret and correct for the data measurements. Note that all our tests are based on the DES SV data set. With different data characteristics, the interplay between the different effects (photo-z, shape noise, masking, cosmic variance) could be very different.

4.5.4.2 Redshift-dependent galaxy bias of DES SV data

We now continue to measure redshift-dependent galaxy bias with the DES SV data using the same procedure as in the simulations. Figure 4.12 shows some examples of the maps. The right-most panel shows the $\gamma_{\rm l}$ map at redshift bin $z_{\rm mean} = 1.0-1.2$, while the rest of the maps are the $\gamma_{{\rm l},g}$ maps at different redshift bin evaluated for this $\gamma_{\rm l}$ map. We see the effect of the lensing kernel clearly: the left-most panel is at the peak of the lensing kernel, giving it a higher weight compared to the other lens bins. We also see correlations between $\gamma_{{\rm l},g}$ maps at different redshift bins. This is a result of the photo-z contamination.

In Figure 4.13 we show the galaxy bias measurement for our magnitude-limited galaxy sample from DES SV together with two other independent measurements with the same galaxy sample (discussed in §4.5.5). We have excluded the highest redshift bin since with only a small number of source galaxies, the constraining power from lensing in that bin is

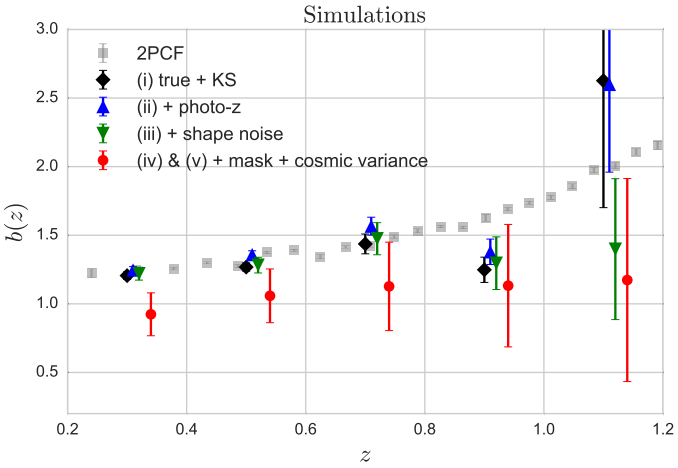

Figure 4.11: Redshift-dependent galaxy bias measured from simulations with different levels of degradation from the ideal scenario tested §4.4.2. The grey line shows the bias from the 2PCF measurement, which we take as "truth". The black, blue, green and red points corresponding to the steps (i), (ii), (iii) and (iv)+(v) in §4.5.4.1, respectively. All measurement are done using equation (4.46).

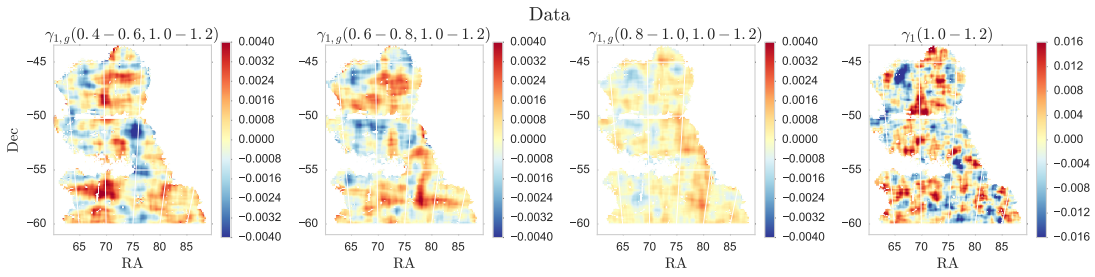

Figure 4.12: Example of maps from DES SV data. The right-most panel shows the γ_1 map generated from the source redshift bin $z_{\rm mean} = 1.0 - 1.2$, while the other panels show the $\gamma_{1,g}$ maps generated for the source redshift bin $z_{\rm mean} = 1.0 - 1.2$ and for different lens redshifts (left: $z_{\rm mean} = 0.4 - 0.6$, middle: $z_{\rm mean} = 0.6 - 0.8$, right: $z_{\rm mean} = 0.8 - 1.0$). The title in each panel for $\gamma_{1,g}$ indicate the lens and source redshift, while the title for γ_1 indicates the source redshift. Note that the color bars are in different ranges, but are matched to the simulation plot in Figure 4.10. In addition, the left-most and the right-most panels correspond to the bottom row of that figure.

very weak. Table 4.1 summarizes the results. The black data points are from this work, with a best-fit linear model of: $b(z) = 1.35^{\pm 0.28} - 0.15^{\pm 0.59} z$. As discussed earlier, our method becomes much less constraining going to higher redshift, as the source galaxies become sparse. This is manifested in the increasingly large error bars going to high redshifts. Here we only performed a simple linear fit to the data given the large uncertainties in our measurements. In the future, one could extend to explore more physically motivated galaxy bias models [46, 148].

In these measurements we include the calibration factor derived from the set of 1,200 SV-like simulations. We find that the spread in the 1,200 simulation measurements is much larger than the error bars on the JK errors on the data measurement itself. This suggests that the field-to-field variation, and how that couples with the mask and the noise is a much larger effect than what the JK errors capture (variation of structure within the field and some level of shape noise). As a result, we have taken the error bars and full covariance from the simulations as our final estimated uncertainty on the measurements.

4.5.4.3 Other systematics test

In §4.5.4.1, we have checked for various forms of systematic effects coming from the KS conversion, finite area, complicated mask geometry, and photo-z errors. Here we perform three additional tests. First, we check that the cross-correlation between the B-mode shear γ_B and γ_g is small. Lensing B-mode refers to the divergent-free piece of the lensing field, which is zero in an ideal, noiseless scenario. In Figure 4.14, we show all the b_B measurements using both shear component and all lens-source combinations. We see that all the data points as well as the weighted mean are consistent with zero at the 1–2 σ level, assuring that the B-modes in the shear measurements are mostly consistent with noise. We also show the corresponding B-modes from the simulation used in §4.5.4.1 (iv), where we see that the level and scatter in the data is compatible with that in the simulations.

Next, we check that using the second DES shear pipeline, IM3SHAPE gives consistent answers with that from NGMIX. The resulting redshift-dependent galaxy measurements are shown in Table 4.1 and are overall slightly higher than the NGMIX measurements, but they show a consistent linear fit.

Finally, we check that using two other photo-z codes also give consistent results. Table 4.1 lists the results from the different photo-z catalogues. Since SKYNET and TPZ are both machine learning codes and respond to systematic effects in a similar fashion, while BPZ is a template fitting code, we can thus view the difference between the results from

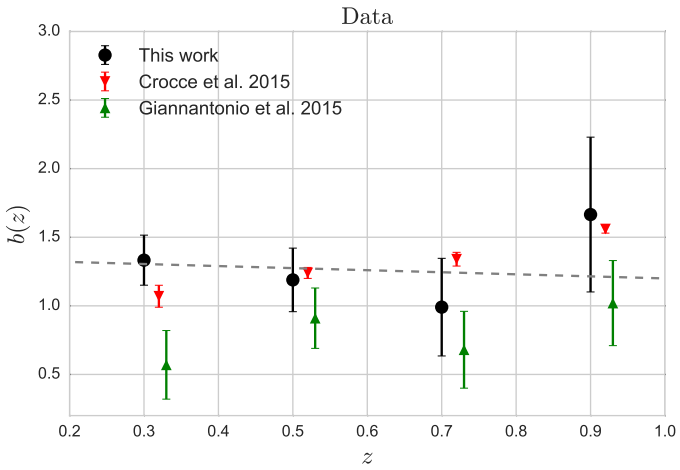

Figure 4.13: Redshift-dependent bias measured from the DES SV data. The black data points show the result from this work. The red and green points show the measurements on the same galaxy sample with different methods. The grey dashed line is the linear fit to the black data points.

BPZ and the others as a rough measure of the potential systematic uncertainty in our photo-z algorithm (see also the discussion in [37]), which is shown here to be much smaller than the other sources of uncertainties.

4.5.5 Comparison with other measurements

The redshift-dependent galaxy bias has been measured on the same data set using other approaches. Here we compare our result with two other measurements – galaxy clustering [60] (hereafter C15) and cross-correlation of galaxies and CMB lensing [91] (hereafter G15).We note that both these analyzes assumed the most recent *Planck* cosmological parameters [170], which is slightly different from our assumptions (see §4.5.3.5). But since our measurement depends very weakly on the assumption of cosmological parameters (as discussed in §4.5.1), the stronger cosmology dependencies come from the cosmological parameters assumed in C15 and G15, which are known well within our measurement uncertainties. We also note that the results we quote in Table 4.1 are based on the photo-z code TPZ, which means our redshift binning is not completely identical to theirs.

In C15, galaxy bias was estimated through the ratio between the projected galaxy angular correlation function (2PCF) in a given redshift bin and an analytical dark matter angular correlation function predicted at the same redshift. The latter includes both linear and nonlinear dark matter clustering derived from CAMB [138] assuming a flat ΛCDM+ν cosmological model based on *Planck* 2013 + *WMAP* polarization + ACT/SPT + BAO. As

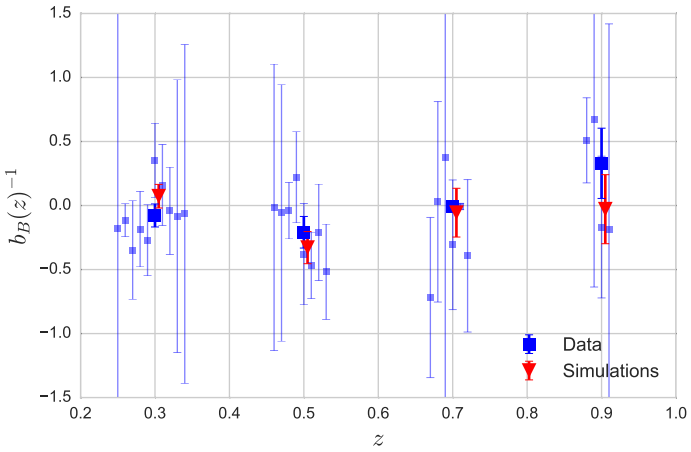

Figure 4.14: All $1/b_B(z)$ measurements from the B-mode shear and the same γ_g in our main analysis. Each small blue data point represents a measurement from a combination of lens redshift, source redshift, and shear component. Note that the low redshift bins contain more data points, as there are more source galaxies that can be used for the measurement. The large blue points are the weighted mean of all measurements at the same redshift bin from the DES SV data, while the red points are that from simulations that are well matched to data.

shown in Figure 4.13 and listed in Table 4.1, our measurements and C15 agree at the 1σ level, with the lowest redshift bin slightly above 1σ.

In G15, galaxy bias is estimated by the ratio between the galaxy-CMB convergence cross-correlation and an analytical prediction of the dark matter-CMB convergence cross-correlation, both calculated through the 2PCF (and also in harmonic space through the power spectrum). Since the lensing efficiency kernel of the CMB is very broad and the CMB lensing maps are typically noisy, this method has less constraining power than C15. However, by using an independent external data, the CMB lensing maps from the South Pole Telescope and the *Planck* satellite, this measurement serves as a good cross check for possible systematic effects in the DES data.

In calculating the theoretical dark matter-CMB convergence cross-correlation, G15 also assumed a fixed cosmology and derived all predictions using CAMB. The σ_8-b degeneracy is thus also present in their analysis. We note, however, that one could apply our method to the CMB lensing data and avoid this dependency. In our framework, the CMB lensing plane will serve as an additional source plane at redshift $\sim$1100. We defer this option to future work.

We find that G15 is systematically lower than our measurement at the 1σ level for the three highest redshift bin, and shows a much larger discrepancy for the lowest redshift bin.

G15 also has more constraining power at high redshift compared to our results, as expected. Possible reasons for the discrepancy at low redshift include systematic errors (in e.g. the photo-z estimation) that are not included in either C15, G15 or this work. In addition, the redshift bins are significantly covariant, making the overall discrepancy less significant. Finally, the scales used in the three studies are slightly different. We refer the readers to G15 for more discussion of this discrepancy.

4.6 Conclusions

In this chapter we explore a new method to measure galaxy bias from the combination of the galaxy density and weak lensing fields. This method is based on [4], where they use the galaxy density field to construct a bias-weighted convergence map κ_g in the COSMOS field. They measure different parameterizations of galaxy bias from the zero-lag correlations of the galaxy shear and a reconstruction of the shear from the galaxy density field. In this chapter we present a new way to measure tomographic bias from the zero-lag correlations between the lensing maps and a reconstruction of the lensing maps from the galaxy density field. We also study the robustness and the systematics of this method for the first time.

The implementation of this model is as follows. We construct a template of the convergence field κ_g at the source redshift by integrating the density field of the foreground galaxies in the line-of-sight weighted by the corresponding lensing kernel as specified in equation (4.18). We do this for tomographic bins in the lens distribution. We then compare to estimates of the matter convergence map κ associated to the same galaxies in the source redshift bin. We measure galaxy bias from the smoothed zero-lag cross-correlations between κ and κ_g as in equations (4.20,4.21). Instead of using the zero-lag cross-correlation we could also use the 2-point cross-correlation function.

We use the MICE simulations to study the consistency of our method by comparing our results with a fiducial galaxy bias measurement on linear scales. This is obtained from the ratio between the projected 2PCFs ($\omega(\Theta)$) of galaxies and dark matter as a function of redshift (see equation (4.3)), and fitting a constant galaxy bias between 6 arcmin and 60 arcmin. We also study local bias from equations (4.6-4.8), making use of the dark matter field of the simulation. With these comparisons we study the systematics of the method and the regimes where it is consistent with linear bias.

There are different systematic effects and numerical dependencies of the method that need to be taken into account for a correct measurement of linear bias. First of all, the method is sensitive to the redshift bin width used in the construction of κ_g, that have an impact on the galaxy bias estimators due to the projection effects of the density fields. This

causes differences in the values obtained for the different estimators, that can be larger than 5% for $\Delta z > 0.2$ and larger than 10% for $\Delta z > 0.3$. This has to be taken into account when measuring κ_g in wide redshift bins in order to obtain the correct linear bias. On the other hand, projecting the source galaxies in a plane have an insignificant impact on the results. Secondly, the angular smoothing scale of the field can be affected by nonlinearities for small enough scales. We find that the measurements are consistent with linear bias for angular scales of $\Theta > 30$ arcmin, where bias is constant. Sampling and discreteness noise is also important and needs to be taken into account (see equations (4.20,4.21)). Finally, we need to exclude from the analysis those pixels that are affected by the edges of the area used.

A correction must be applied to our estimators if we only use the foreground galaxies in a given redshift bin for the construction of κ_g. This is because the amplitude of κ_g is reduced by the fact that we do not use all the information from the complete redshift range in front of the source redshift. We predict theoretically this effect, and we find good agreement with the measurements, indicating that we can use this prediction to correct the bias obtained. The theoretical prediction describes the amplitude of the zero-lag correlations obtained using a given redshift bin for the foreground dark matter field, that by definition has a bias of 1. We can measure galaxy bias in that bin from the ratio between the zero-lag correlations of κ and κ_g (using the foregroud galaxies in that bin) and the theoretical prediction. This provides a useful tool to do tomography and measure galaxy bias in single redshift bins. We measure and show the redshift-dependent bias obtained using this method, and find good agreement with the redshift-dependent bias from equation (4.3).

Other issues associated with observational data must be addressed if we apply this method to large galaxy surveys such as the Dark Energy Survey (DES). First of all, we measure γ instead of κ from the ellipticities of the galaxies, but we obtain κ_g from the galaxy density field. Thus, we need to apply a conversion from κ_g to γ_g or from γ to κ in order to calculate the zero-lag correlations. This conversion is affected by the shape of the mask and the noise of the maps. Secondly, for photometric surveys we need to account for the photo-z estimation. The uncertainty on the redshift of the galaxies causes a smoothing in κ_g that needs to be corrected from the photo-z errors and the galaxy redshift distribution. Other effects in observations include the boundary effects of irregular masks, that affect the useful area of the analysis, and shape noise, that causes some of the estimators to be very noisy.

We applied this method to the DES Science Verification data in [44]. Following from this analysis, we carry out a series of simulation tests which incorporate step-by-step realistic effects in our data including shape noise, photo-z errors and masking. In each step, we

investigate the errors introduced in our estimation of galaxy bias. We find that shape noise and masking together affects our measurement in a non-negligible way, causing the galaxy bias to be measured low, while the photo-z affects the measurements in a predictable way if the characteristics of the photo-z uncertainties are well understood. As the measurement itself is very noisy, simulation tests where we know the "truth" provide a good anchor for building the analysis pipeline. In the main analysis of [44], we measure the galaxy bias with a $18 < i < 22.5$ magnitude-limited galaxy sample in 4 tomographic redshift bins to be 1.33 ± 0.18 ($z = 0.2 - 0.4$), 1.19 ± 0.23 ($z = 0.4 - 0.6$), 0.99 ± 0.36 ($z = 0.6 - 0.8$), and 1.66 ± 0.56 ($z = 0.8 - 1.0$). Measurements from higher redshifts are too noisy to be constraining. The best-fit linear model gives: $b(z) = 1.36^{\pm 0.31} - 0.08^{\pm 0.70} z$. The results are consistent between different shear and photo-z catalogs. This method is expected to be significantly better when applied to larger areas, such as in DES Year 1 [69] or the 5000 deg^2 from the expected total area of the DES survey.

This chapter presents the method and its first application to observations, but further studies can be done. We can explore galaxy bias for different galaxy samples, e.g. as a function of colour and luminosity. We can also explore the scale dependence of local bias by studying different angular scales and its nonlinearities, and the redshift dependence by comparing the tomographic measurements with parametric redshift-dependent bias based on [4]. In this chapter we have focused on zero-lag cross-correlations, but we could also use 2-point cross-correlations as a way to estimate the bias and include the redshift cross-correlations as a validation test.

The method studied in this chapter has several attractive features. First of all, there is no need to assume a dark matter distribution to measure bias, since this distribution is measured from the weak lensing field. This method is then a direct way to measure local bias by comparing the observed galaxy and dark matter distributions. Moreover, the method depends very weakly on σ_8 (only in non-linear corrections to f_1 and f_2), while other measurements of bias are typically strongly dependent on σ_8. On the other hand, it depends on Ω_m. Hence, a combined analysis of different measurements of galaxy bias, including this method, can be very useful to constrain better bias and cosmology. Second, the method can be applied to a situation where galaxies only cover partially the full redshift range of the lenses. Moreover, the potential of this method will rapidly increase with the data of present and upcoming surveys, such as the Hyper Suprime-Cam (HSC), the Dark Energy Survey (DES), the Kilo Degree Survey (KiDS), the Large Synoptics Survey Telescope (LSST), the Wide-Field Infrared Survey Telescope (WFIRST) and the Euclid mission.

Chapter 5

Summary and Conclusions

In this PhD thesis we focus our study on the relation between the distribution of galaxies and the halo and matter distribution at large scales. In order to do this we study galaxy and halo bias in the linear and local regimes. The bias factor is a parameter that describes the relation between the clustering amplitude of galaxies (or haloes) and the clustering amplitude of matter. The knowledge of this factor allows us to describe how galaxies trace the underlying matter distribution and improves the precision of our cosmological studies. Our study is focused on simulations, where we study linear galaxy bias, the relation between halo and galaxy bias and a new method to measure local bias that can be applied (and has been applied already) to observations.

Relation between halo and galaxy linear bias

A common assumption used to interpret galaxy clustering in surveys is the fact that galaxies populate haloes according only to the halo mass, and also that halo bias depends only on mass. These assumptions are then used to fit the Halo Occupation Distribution (HOD) from the observations of galaxy clustering [57, 233], or to populate mock galaxy catalogues in simulations [42]. However, this assumption is not always precise enough. The goal of our study is to analyse the consequences of these assumptions on the galaxy bias predictions, and also to investigate how the environmental dependence of halo bias can improve such galaxy bias predictions.

In §2 we use the Millennium Simulation [206] to study halo bias as a function of mass and galaxy bias from the SAMs available in the simulation. We present a method to test the assumptions that halo bias and galaxy occupation only depends on halo mass. The method consists on reconstructing galaxy bias from the measurements of halo bias and HOD of galaxies as a function of halo mass. These bias reconstructions recover the correct (measured) galaxy bias if halo bias and/or the HOD only depend on halo mass.

We find that the bias reconstructions tend to overpredict galaxy bias as a function of luminosity by a factor of $\gtrsim 5\%$. This implies that the assumptions that halo bias and HOD only depend on mass do not allow us to predict galaxy bias better than 5%. It also means that halo bias does not only depend on mass. In particular, we show that halo bias also depends on the amount of substructure (subhaloes) inside the haloes. For a fixed halo mass, halo bias depends strongly on the number of subhaloes, specially for low masses. This result is independent of SAMs (is dark matter only) and can be seen as a consequence of assembly bias, or the fact that haloes of a given mass present a different bias depending on other halo properties. This dependence has an impact on galaxy clustering for any sample of galaxies whose occupation depends on the amount of subhaloes, which is the case for most of the galaxy formation models. We finally show that galaxy reconstructions show important disagreements with the measured galaxy bias for halo masses below $3-5\times10^{11}\,h^{-1}\,\mathrm{M}_{\odot}$. This indicates that assembly bias affects strongly the galaxy clustering of small haloes. Then, this study suggests that the standard implementation of the HOD model would fail if the model is used for galaxies that are dominated (or are significantly affected) by these low masses.

In §3 we go further in this analysis in the Millennium Simulation and we study the impact of halo mass and environmental density on the prediction of linear bias. First of all, we study how halo bias depends on mass and environment (defined as the dark matter fluctuation around a given radius of the halo), and we find that bias depends much more strongly on environment than on mass. We see this because, when we fix the environment, bias does not depend on mass, and the environmental density dependence of halo bias is independent of mass. This result is an evidence that environmental density constrains bias much better than mass. We also see that, while massive haloes only occupy dense regions, small haloes can live in a wide range of environmental densities. This is the reason of why massive haloes have a high bias. More massive haloes have a higher clustering because they are statistically in denser regions, but not because mass is the fundamental cause of clustering. On the other hand, the fact that small haloes live in different environments is in contradiction with the standard HOD assumption that states that the clustering of haloes only depends on mass.

In order to study the implications of these results on the clustering predictions, we extend the reconstruction test to a more general method that tests if halo bias and HOD depends only on mass or only on environmental density. On one side, we reconstruct halo bias as a function of environmental density from the assumption that bias and HOD only depend on mass. We see a strong disagreement between the reconstructed and the measured bias, meaning that halo mass is not a good property to describe the halo bias of a

fixed environment. On the other hand, we reconstruct halo bias as a function of mass from the assumption that halo bias and HOD depend only on the environmental density. We find a much better agreement between the reconstruction and the measurement of halo bias, which suggests that environmental density is a good property to predict the bias of haloes of different masses. These tests suppose another indication that the environmental density is a more fundamental property to relate it to linear bias than mass. In fact, if a galaxy sample is selected from a property that follows directly halo mass, both reconstructions will be able to recover bias. However, if a galaxy is selected from a property that only follows the environmental density, the reconstruction assuming that bias and HOD only depend on mass will totally fail. We end this chapter by showing a couple of examples of galaxy selections that are sensitive to other properties than mass. When central galaxies are selected by colour, the standar HOD implementation assuming mass dependence produces a wrong reconstruction. Instead, the reconstruction from the density dependence of bias and HOD produces a much better agreement with the measured galaxy bias. This study suggests that HOD modeling could be improved by including environment as a property to describe galaxy large-scale clustering.

Measuring local galaxy bias by combining the weak lensing and density fields

In §4 we explore a new method to measure local galaxy bias from the combination of the weak lensing field and the galaxy density field. We have been testing and analysing the method in the MICE Grand Challenge Simulation [60, 78, 79] and its galaxy catalogues [42], and we finally applied the method to the DES Science Verification data [44].

The method is based on [4], where they use different parametrizations to measure galaxy bias from the zero-lag correlations between the galaxy shear and a reconstructed shear from the density distribution of the foreground galaxies. In our implementation we construct a template of the convergence map, that we call κ_g, obtained from the density field of the foreground galaxies, and we measure galaxy bias in tomographic redshift bins from the zero-lag correlations between the convergence maps κ and κ_g obtained from the galaxies in each redshift bin. The advantage of this method to measure bias is that it depends weakly on σ_8, while most of the methods to measure bias in observations depend strongly on σ_8. Hence, both σ_8 and bias can be well constrained from the combination of different techniques to measure bias, including this one, since the systematics and dependencies of this method are different than in other techniques. Another advantage with respect to [4] is that our method does not need to parametrize galaxy bias in order to measure it.

In this study we analyse the consistency of the method with other measurements of linear bias. We also study the different systematic effects and numerical dependencies that affect the measurement of bias. For this we use the MICE simulation, its galaxy catalogues and their lensing field values, the full resolution weak lensing field of the simulation and a dilution of the dark matter particles. We show that bias depends strongly on the redshift bin width used in the construction of κ_g due to the projection effects of the density field. This affects the bias estimator more than 5% for a redshift bin width $\Delta z = 0.2$ and more than 10% for $\Delta z = 0.03$. However, we correct this by modeling theoretically the effect, and we test that the theoretical calibration works in the simulation. This correction allows to measure bias in tomographic redshift bins. We also see that approximating the source galaxies in a plane produces insignificant differences in the bias estimation. We find that our estimator of bias is consistent with linear bias for scales larger than 30 arcmin, and below these scales nonlinearities appear. We also correct our measurements for shot-noise and edge effects. We finally show the measurement of galaxy bias at different redshift bins in the MICE simulation, showing a good agreement with other techniques to measure galaxy bias.

In this chapter we have studied the idealistic cases in simulations, but other effects need to be taken into account in obsevations. First of all, we never observe κ directly. Instead, we need to measure galaxy shear and obtain κ from it. An alternative is to reconstruct the shear from κ_g and measure bias form the zero-lag correlations between the observed and reconstructed shear. Moreover, in all these cases shear noise is the dominant noise that needs to be correctly taken into account. Photometric errors in the redshift estimation also affect significantly the estimation of bias. The uncertainty on the redshift of the galaxies causes a smoothing in κ_g that needs to be corrected from the photo-z errors and the galaxy redshift distribution. Finally, we need to take into account the boundary effects that can come from irregular masks. In [44] we have applied this method to measure galaxy bias as a function of redshift in the DES Science Verification data. To do this, we studied these observational effects in the MICE simulation in order to understand and calibrate each of the effects from shape noise, photo-z errors and masking. The results are consistent between different estimators of shape and redhsift. This method is expected to be significantly better when applied to larger areas, such as in DES Year 1 or the 5000 deg^2 from the expected total area of the DES survey, and its potential will rapidly increase with the data of other present and upcoming surveys, such as the Hyper Suprime-Cam (HSC), the Kilo Degree Survey (KiDS), the Large Synoptics Survey Telescope (LSST), the Wide-Field Infrared Survey Telescope (WFIRST) and the Euclid mission.

Future work

There are several possibilities to go further in these studies. On one side, we can go beyond the work from §2 and §3 on both the theoretical and observational sides. First of all, we can explore more deeply the theory concerning environmental dependences of bias. Modeling this dependence can be worth for understanding the importance of environment for bias and clustering. Moreover, as environment is very easy to measure in observations, this can be used to measure bias and cosmology directly from these measurements if we use a good model. Secondly, another step for this project could be to think and develop an alternative way to model and measure HOD that is not affected by assembly bias and other dependencies of bias. This could be done, for example, by using a model of the environmental density dependence of halo bias instead of (or including) halo mass, at least for the modelling of large-scale clustering. Finally, studies of the environmental dependence of galaxy clustering can be done in observations using weak lensing maps, where we can select galaxies according to their matter environmental density inferred from weak lensing maps.

On the other side, the work done in §4 can be continued with different implementations that would improve our analysis. First of all, the same method can be applied to DES Year 1 data and future surveys. The measurement is sensitive to the area used, in the sense that larger areas reduce the errors of the measurements. Because of this we expect to improve the measurements in future DES data. Secondly, we can use the combination of spectroscopic and photometric galaxy surveys in order to improve the sensibility of the method. Using a spectroscopic survey for the foreground galaxies would allow us to measure the galaxy density field with a much better resolution, and this would reduce the noise for κ_g significantly. Another focus could be the analysis of non-linear bias, and the study of how this local bias measurement depends on scale. Complementary to this, we could measure bias from the 2-point correlations between κ_g and κ instead of from the zero-lag correlations. This would reduce the noise in our measurements significantly (we would not be measuring variances, that are noisier), specially for irregular masks, and would allow us to study better the scale dependence of bias. Another interesting application is to measure a parametric redshift dependent bias, as in [4]. This allows us to measure the redshift evolution of bias without using tomographic bins. A comparison of the local bias measurements using tomography and parametric bias would give a test of how well these two methods measure a consistent galaxy bias, and in what conditions is better to use parametric bias or tomographic redshift bins. Finally, all these analyses can be performed for different galaxy selections, in order to study galaxy bias as a function of luminosity or

colour. This would relate galaxy properties with the mass distribution, which is useful to understand galaxy formation and evolution.

Bibliography

[1] U. Abbas and R. K. Sheth. The environmental dependence of clustering in hierarchical models. *MNRAS*, 364:1327–1336, December 2005.

[2] U. Abbas and R. K. Sheth. The environmental dependence of galaxy clustering in the Sloan Digital Sky Survey. *MNRAS*, 372:1749–1754, November 2006.

[3] U. Abbas and R. K. Sheth. Strong clustering of underdense regions and the environmental dependence of clustering from Gaussian initial conditions. *MNRAS*, 378:641–648, June 2007.

[4] A. Amara, S. Lilly, K. Kovač, J. Rhodes, R. Massey, G. Zamorani, C. M. Carollo, T. Contini, J.-P. Kneib, and et al. The COSMOS density field: a reconstruction using both weak lensing and galaxy distributions. *MNRAS*, 424:553–563, July 2012.

[5] R. E. Angulo, C. M. Baugh, and C. G. Lacey. The assembly bias of dark matter haloes to higher orders. *MNRAS*, 387:921–932, June 2008.

[6] R. E. Angulo, V. Springel, S. D. M. White, A. Jenkins, C. M. Baugh, and C. S. Frenk. Scaling relations for galaxy clusters in the Millennium-XXL simulation. *MNRAS*, 426:2046–2062, November 2012.

[7] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald. Evidence for quadratic tidal tensor bias from the halo bispectrum. *Phys.Rev.D*, 86(8):083540, October 2012.

[8] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *ApJ*, 304:15–61, May 1986.

[9] M. Bartelmann and P. Schneider. Weak gravitational lensing. *PhysRep*, 340:291–472, January 2001.

[10] C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69:3101–3156, December 2006.

[11] C. M. Baugh. Luminosity Bias: From Haloes to Galaxies. *PASA*, 30:30, April 2013.

[12] M. R. Becker, M. A. Troxel, N. MacCrann, E. Krause, T. F. Eifler, O. Friedrich, A. Nicola, A. Refregier, A. Amara, D. Bacon, G. M. Bernstein, C. Bonnett, S. L. Bridle, M. T. Busha, C. Chang, S. Dodelson, B. Erickson, A. E. Evrard, J. Frieman, E. Gaztanaga, D. Gruen, W. Hartley, B. Jain, M. Jarvis, T. Kacprzak, D. Kirk, A. Kravtsov, B. Leistedt, E. S. Rykoff, C. Sabiu, C. Sanchez, H. Seo, E. Sheldon, R. H. Wechsler, J. Zuntz, T. Abbott, F. B. Abdalla, S. Allam, R. Armstrong, M. Banerji, A. H. Bauer, A. Benoit-Levy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Fausti Neto, E. Fernandez, D. A. Finley, B. Flaugher, P. Fosalba, D. W. Gerdes, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, M. A. G. Maia, M. March, P. Martini, P. Melchior, C. J. Miller, R. Miquel, J. J. Mohr, R. C. Nichol, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas, V. Vikram, A. R. Walker, and The DES Collaboration. Cosmic Shear Measurements with DES Science Verification Data. *ArXiv e-prints*, July 2015.

[13] J. Bel, K. Hoffmann, and E. Gaztañaga. Non-local bias contribution to third-order galaxy correlations. *ArXiv e-prints*, April 2015.

[14] N. Benítez. Bayesian Photometric Redshift Estimation. *ApJ*, 536:571–583, June 2000.

[15] J. Benjamin, C. Heymans, E. Semboloni, L. van Waerbeke, H. Hoekstra, T. Erben, M. D. Gladders, M. Hetterscheidt, Y. Mellier, and H. K. C. Yee. Cosmological constraints from the 100-deg^2 weak-lensing survey. *MNRAS*, 381:702–712, October 2007.

[16] C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Greason, R. S. Hill, E. Komatsu, M. R. Nolta, N. Odegard, H. V. Peiris, L. Verde, and J. L. Weiland. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. *ApJS*, 148:1–27, September 2003.

[17] A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh, and C. G. Lacey. The nature of galaxy bias and clustering. *MNRAS*, 311:793–808, February 2000.

[18] A. A. Berlind, E. Kazin, M. R. Blanton, S. Pueblas, R. Scoccimarro, and D. W. Hogg. The Clustering of Galaxy Groups: Dependence on Mass and Other Properties. *ArXiv Astrophysics e-prints*, October 2006.

[19] A. A. Berlind and D. H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *ApJ*, 575:587–616, August 2002.

[20] F. Bernardeau. The large-scale gravitational bias from the quasi-linear regime. *A&A*, 312:11–23, August 1996.

[21] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *PhysRep*, 367:1–248, September 2002.

[22] M. Bernardi, A. Meert, R. K. Sheth, V. Vikram, M. Huertas-Company, S. Mei, and F. Shankar. The massive end of the luminosity and stellar mass functions: dependence on the fit to the light profile. *MNRAS*, 436:697–704, November 2013.

[23] E. Bertin. Automatic Astrometric and Photometric Calibration with SCAMP. In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, page 112, July 2006.

[24] E. Bertin. Automated Morphometry with SExtractor and PSFEx. In I. N. Evans, A. Accomazzi, D. J. Mink, and A. H. Rots, editors, *Astronomical Data Analysis Software and Systems XX*, volume 442 of *Astronomical Society of the Pacific Conference Series*, page 435, July 2011.

[25] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. A&AS, 117:393–404, June 1996.

[26] E. Bertin, Y. Mellier, M. Radovich, G. Missonnier, P. Didelon, and B. Morin. The TERAPIX Pipeline. In D. A. Bohlender, D. Durand, and T. H. Handley, editors, *Astronomical Data Analysis Software and Systems XI*, volume 281 of *Astronomical Society of the Pacific Conference Series*, page 228, 2002.

[27] S. Bertone, G. De Lucia, and P. A. Thomas. The recycling of gas and metals in galaxy formation: predictions of a dynamical feedback model. *MNRAS*, 379:1143–1154, August 2007.

[28] P. Bett, V. Eke, C. S. Frenk, A. Jenkins, J. Helly, and J. Navarro. The spin and shape of dark matter haloes in the Millennium simulation of a Λ cold dark matter universe. *MNRAS*, 376:215–232, March 2007.

[29] C. Blake. First Results from the WiggleZ Galaxy Redshift Survey. In *A Decade of Dark Energy*, May 2008.

[30] C. Blake, A. Collister, and O. Lahav. Halo-model signatures from 380000 Sloan Digital Sky Survey luminous red galaxies with photometric redshifts. *MNRAS*, 385:1257–1269, April 2008.

[31] M. Blanton. How Stochastic Is the Relative Bias between Galaxy Types? *ApJ*, 544:63–80, November 2000.

[32] M. R. Blanton, D. Eisenstein, D. W. Hogg, D. J. Schlegel, and J. Brinkmann. Relationship between Environment and the Broadband Optical Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 629:143–157, August 2005.

[33] M. R. Blanton, D. W. Hogg, N. A. Bahcall, J. Brinkmann, M. Britton, A. J. Connolly, I. Csabai, M. Fukugita, J. Loveday, A. Meiksin, J. A. Munn, R. C. Nichol, S. Okamura, and et al. The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1. *ApJ*, 592:819–838, August 2003.

[34] M. R. Blanton, R. H. Lupton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, M. Fukugita, and J. Loveday. The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. *ApJ*, 631:208–230, September 2005.

[35] M. R. Blanton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, D. Finkbeiner, M. Fukugita, J. E. Gunn, D. W. Hogg, Ž. Ivezić, G. R. Knapp, R. H. Lupton, J. A. Munn, D. P. Schneider, M. Tegmark, and I. Zehavi. New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *Aj*, 129:2562–2578, June 2005.

[36] P. Bode, J. P. Ostriker, and G. Xu. The Tree Particle-Mesh N-Body Gravity Solver. *ApJS*, 128:561–569, June 2000.

[37] C. Bonnett, M. A. Troxel, W. Hartley, A. Amara, B. Leistedt, M. R. Becker, G. M. Bernstein, S. Bridle, C. Bruderer, M. T. Busha, M. Carrasco Kind, M. J. Childress, F. J. Castander, C. Chang, M. Crocce, T. M. Davis, T. F. Eifler, J. Frieman, C. Gangkofner, E. Gaztanaga, K. Glazebrook, D. Gruen, T. Kacprzak, A. King, J. Kwan, O. Lahav, G. Lewis, C. Lidman, H. Lin, N. MacCrann, R. Miquel, C. R. O'Neill, A. Palmese, H. V. Peiris, A. Refregier, E. Rozo, E. S. Rykoff, I. Sadeh, C. Sánchez, E. Sheldon, S. Uddin, R. H. Wechsler, J. Zuntz, T. Abbott, F. B. Abdalla, S. Allam, R. Armstrong, M. Banerji, A. H. Bauer, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, J. Carretero, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Fausti Neto, E. Fernandez, B. Flaugher, P. Fosalba, D. W. Gerdes, R. A. Gruendl, K. Honscheid, B. Jain, D. J. James, M. Jarvis, A. G. Kim, K. Kuehn, N. Kuropatkin, T. S. Li, M. Lima, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, C. J. Miller, E. Neilsen, R. C. Nichol, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, B. Santiago, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas, V. Vikram, and A. R. Walker. Redshift distributions of galaxies in the DES Science Verification shear catalogue and implications for weak lensing. *ArXiv e-prints*, July 2015.

[38] R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, August 2006.

[39] M. Cacciato, O. Lahav, F. C. van den Bosch, H. Hoekstra, and A. Dekel. On combining galaxy clustering and weak lensing to unveil galaxy biasing via the halo model. *MNRAS*, 426:566–587, October 2012.

[40] M. Carrasco Kind and R. J. Brunner. TPZ: photometric redshift PDFs and ancillary information by using prediction trees and random forests. *MNRAS*, 432:1483–1501, June 2013.

[41] M. Carrasco Kind and R. J. Brunner. Exhausting the information: novel Bayesian combination of photometric redshift PDFs. *MNRAS*, 442:3380–3399, August 2014.

[42] J. Carretero, F. J. Castander, E. Gaztañaga, M. Crocce, and P. Fosalba. An algorithm to build mock galaxy catalogues using MICE simulations. *MNRAS*, 447:646–670, February 2015.

[43] K. C. Chan, R. Scoccimarro, and R. K. Sheth. Gravity and large-scale nonlocal bias. *Phys.Rev.D*, 85(8):083509, April 2012.

[44] C. Chang, A. Pujol, E. Gaztanaga, A. Amara, A. Refregier, D. Bacon, M. R. Becker, C. Bonnett, J. Carretero, F. J. Castander, M. Crocce, P. Fosalba, T. Giannantonio, W. Hartley, B. Jain, M. Jarvis, T. Kacprzak, A. J. Ross, E. Sheldon, M. A. Troxel, V. Vikram, J. Zuntz, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, A. Benoit-Levy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, J. Estrada, A. E. Evrard, B. Flaugher, J. Frieman, D. A. Goldstein, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, J. L. Marshall, P. Martini, P. Melchior, C. J. Miller, R. Miquel, J. J. Mohr, R. C. Nichol, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. K. Romer, A. Roodman, E. S. Rykoff, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, and A. R. Walker. Galaxy bias from the DES Science Verification Data: combining galaxy density maps and weak lensing maps. *ArXiv e-prints*, January 2016.

[45] C. Chang, V. Vikram, B. Jain, D. Bacon, A. Amara, M. R. Becker, G. Bernstein, C. Bonnett, S. Bridle, D. Brout, M. Busha, J. Frieman, E. Gaztanaga, W. Hartley, M. Jarvis, T. Kacprzak, A. Kovács, O. Lahav, H. Lin, P. Melchior, H. Peiris, E. Rozo, E. Rykoff, C. Sánchez, E. Sheldon, M. A. Troxel, R. Wechsler, J. Zuntz, T. Abbott, F. B. Abdalla, S. Allam, J. Annis, A. H. Bauer, A. Benoit-Lévy, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, F. J. Castander, M. Crocce, C. B. D'Andrea, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, A. E. Evrard, A. Fausti Neto, B. Flaugher, P. Fosalba, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. James, S. Kent, K. Kuehn, N. Kuropatkin, M. A. G. Maia, M. March, P. Martini, K. W. Merritt, C. J. Miller, R. Miquel, E. Neilsen, R. C. Nichol, R. Ogando, A. A. Plazas, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, I. Sevilla, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, J. Thaler, D. Thomas, D. Tucker, and A. R. Walker. Wide-Field Lensing Mass Maps from Dark Energy Survey Science Verification Data. *Physical Review Letters*, 115(5):051301, July 2015.

[46] L. Clerkin, D. Kirk, O. Lahav, F. B. Abdalla, and E. Gaztañaga. A prescription for galaxy biasing evolution as a nuisance parameter. *MNRAS*, 448:1389–1401, April 2015.

[47] S. Cole and N. Kaiser. Biased clustering in the cold dark matter cosmogony. *MNRAS*, 237:1127–1146, April 1989.

[48] S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk. Hierarchical galaxy formation. *MNRAS*, 319:168–204, November 2000.

[49] S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, and et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362:505–534, September 2005.

[50] M. Colless. First results from the 2dF Galaxy Redshift Survey. *Anglo-Australian Observatory Epping Newsletter*, 85:4–6, April 1998.

[51] S. Contreras, C. M. Baugh, P. Norberg, and N. Padilla. How robust are predictions of galaxy clustering? *MNRAS*, 432:2717–2730, July 2013.

[52] A. Cooray. Non-linear galaxy power spectrum and cosmological parameters. *MNRAS*, 348:250–260, February 2004.

[53] A. Cooray. Halo model at its best: constraints on conditional luminosity functions from measured galaxy statistics. *MNRAS*, 365:842–866, January 2006.

[54] A. Cooray, Y. Gong, J. Smidt, and M. G. Santos. The Near-infrared Background Intensity and Anisotropies during the Epoch of Reionization. *ApJ*, 756:92, September 2012.

[55] A. Cooray and R. Sheth. Halo models of large scale structure. *PhysRep*, 372:1–129, December 2002.

[56] H. M. P. Couchman, P. A. Thomas, and F. R. Pearce. Hydra: an Adaptive-Mesh Implementation of P 3M-SPH. *ApJ*, 452:797, October 1995.

[57] J. Coupon, M. Kilbinger, H. J. McCracken, O. Ilbert, S. Arnouts, Y. Mellier, U. Abbas, S. de la Torre, Y. Goranova, P. Hudelot, J.-P. Kneib, and O. Le Fèvre. Galaxy clustering in the CFHTLS-Wide: the changing relationship between galaxies and haloes since z ˜ 1.2*. *A&A*, 542:A5, June 2012.

[58] J. G. Cresswell and W. J. Percival. Scale-dependent galaxy bias in the Sloan Digital Sky Survey as a function of luminosity and colour. *MNRAS*, 392:682–690, January 2009.

[59] M. Crocce, J. Carretero, A. H. Bauer, A. J. Ross, I. Sevilla-Noarbe, T. Giannantonio, F. Sobreira, J. Sanchez, E. Gaztanaga, M. Carrasco Kind, and et al. Galaxy clustering, photometric redshifts and diagnosis of systematics in the DES Science Verification data. *ArXiv e-prints*, July 2015.

[60] M. Crocce, F. J. Castander, E. Gaztañaga, P. Fosalba, and J. Carretero. The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues. *MNRAS*, 453:1513–1530, October 2015.

[61] R. A. C. Croft, T. D. Matteo, N. Khandai, V. Springel, A. Jana, and J. P. Gardner. Dark matter halo occupation: environment and clustering. *MNRAS*, 425:2766–2777, October 2012.

[62] D. J. Croton, L. Gao, and S. D. M. White. Halo assembly bias and its effects on galaxy clustering. *MNRAS*, 374:1303–1309, February 2007.

[63] D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365:11–28, January 2006.

[64] M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, 267:465–482, April 1983.

[65] J. T. A. de Jong, K. Kuijken, D. Applegate, K. Begeman, A. Belikov, C. Blake, J. Bout, D. Boxhoorn, H. Buddelmeijer, A. Buddendiek, and et al. The Kilo-Degree Survey. *The Messenger*, 154:44–46, December 2013.

[66] G. De Lucia and J. Blaizot. The hierarchical formation of the brightest cluster galaxies. *MNRAS*, 375:2–14, February 2007.

[67] S. Desai, R. Armstrong, J. J. Mohr, D. R. Semler, J. Liu, E. Bertin, S. S. Allam, W. A. Barkhouse, G. Bazin, E. J. Buckley-Geer, M. C. Cooper, S. M. Hansen, F. W. High, H. Lin, Y.-T. Lin, C.-C. Ngeow, A. Rest, J. Song, D. Tucker, and A. Zenteno. The Blanco Cosmology Survey: Data Acquisition, Processing, Calibration, Quality Diagnostics, and Data Release. *ApJ*, 757:83, September 2012.

[68] C. Di Porto, E. Branchini, J. Bel, F. Marulli, M. Bolzonella, O. Cucciati, S. de la Torre, B. R. Granett, L. Guzzo, C. Marinoni, and et al. The VIMOS Public Extragalactic Redshift Survey (VIPERS). Measuring nonlinear galaxy bias at z˜0.8. *ArXiv e-prints*, June 2014.

[69] H. T. Diehl, T. M. C. Abbott, J. Annis, R. Armstrong, L. Baruah, A. Bermeo, G. Bernstein, E. Beynon, C. Bruderer, E. J. Buckley-Geer, and et al. The Dark Energy Survey and operations: Year 1. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9149 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0, August 2014.

[70] T. Doumler and A. Knebe. Investigating the influence of magnetic fields upon structure formation with AMIGA - a C code for cosmological magnetohydrodynamics. *MNRAS*, 403:453–473, March 2010.

[71] J. Dubinski, J. Kim, C. Park, and R. Humble. GOTPM: a parallel hybrid particle-mesh treecode. *New Astronomy*, 9:111–126, February 2004.

[72] A. Durkalec, O. Le Fèvre, A. Pollo, S. de la Torre, P. Cassata, B. Garilli, V. Le Brun, B. C. Lemaux, D. Maccagni, L. Pentericci, and et al. The evolution of clustering length, large-scale bias and host halo mass at 2¡z¡5 in the VIMOS Ultra Deep Survey (VUDS). *ArXiv e-prints*, November 2014.

[73] M. Eriksen and E. Gaztanaga. Combining spectroscopic and photometric surveys using angular cross-correlations II: Parameter constraints from different physical effects. *ArXiv e-prints*, February 2015.

[74] A. Faltenbacher and S. D. M. White. Assembly Bias and the Dynamical Structure of Dark Matter Halos. *ApJ*, 708:469–473, January 2010.

[75] B. Flaugher. The Dark Energy Survey. *International Journal of Modern Physics A*, 20:3121–3123, 2005.

[76] B. Flaugher, H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufore, G. M. Bernstein, R. A. Bernstein, B. Bigelow, M. Bonati, D. Boprie, D. Brooks, E. J. Buckley-Geer, J. Campa, L. Cardiel-Sas, F. J. Castander, J. Castilla, H. Cease, J. M. Cela-Ruiz, S. Chappa, E. Chi, C. Cooper, L. N. da Costa, E. Dede, G. Derylo, D. L. DePoy, J. de Vicente, P. Doel, A. Drlica-Wagner, J. Eiting, A. E. Elliott, J. Emes, J. Estrada, A. Fausti Neto, D. A. Finley, R. Flores, J. Frieman, D. Gerdes, M. D. Gladders, B. Gregory, G. R. Gutierrez, J. Hao, S. E. Holland, S. Holm, D. Huffman, C. Jackson, D. J. James, M. Jonas, A. Karcher, I. Karliner, S. Kent, R. Kessler, M. Kozlovsky, R. G. Kron, D. Kubik, K. Kuehn, S. Kuhlmann, K. Kuk, O. Lahav, A. Lathrop, J. Lee, M. E. Levi, P. Lewis, T. S. Li, I. Mandrichenko, J. L. Marshall, G. Martinez, K. W.

Merritt, R. Miquel, F. Muñoz, E. H. Neilsen, R. C. Nichol, B. Nord, R. Ogando, J. Olsen, N. Palaio, K. Patton, J. Peoples, A. A. Plazas, J. Rauch, K. Reil, J.-P. Rheault, N. A. Roe, H. Rogers, A. Roodman, E. Sanchez, V. Scarpine, R. H. Schindler, R. Schmidt, R. Schmitt, M. Schubnell, K. Schultz, P. Schurter, L. Scott, S. Serrano, T. M. Shaw, R. C. Smith, M. Soares-Santos, A. Stefanik, W. Stuermer, E. Suchyta, A. Sypniewski, G. Tarle, J. Thaler, R. Tighe, C. Tran, D. Tucker, A. R. Walker, G. Wang, M. Watson, C. Weaverdyck, W. Wester, R. Woods, B. Yanny, and The DES Collaboration. The Dark Energy Camera. *Aj*, 150:150, November 2015.

[77] A. S. Font, R. G. Bower, I. G. McCarthy, A. J. Benson, C. S. Frenk, J. C. Helly, C. G. Lacey, C. M. Baugh, and S. Cole. The colours of satellite galaxies in groups and clusters. *MNRAS*, 389:1619–1629, October 2008.

[78] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander. The MICE grand challenge lightcone simulation - I. Dark matter clustering. *MNRAS*, 448:2987–3000, April 2015.

[79] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Crocce. The MICE Grand Challenge light-cone simulation - III. Galaxy lensing mocks from all-sky lensing maps. *MNRAS*, 447:1319–1332, February 2015.

[80] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Manera. The onion universe: all sky lightcone simulations in spherical shells. *MNRAS*, 391:435–446, November 2008.

[81] P. Fosalba and E. Gaztanaga. Cosmological Perturbation Theory and the Spherical Collapse model - I. Gaussian initial conditions. *MNRAS*, 301:503–523, December 1998.

[82] J. N. Fry. The Evolution of Bias. *ApJL*, 461:L65, April 1996.

[83] J. N. Fry and E. Gaztanaga. Biasing and hierarchical statistics in large-scale structure. *ApJ*, 413:447–452, August 1993.

[84] L. Gao, V. Springel, and S. D. M. White. The age dependence of halo clustering. *MNRAS*, 363:L66–L70, October 2005.

[85] L. Gao and S. D. M. White. Assembly bias in the clustering of dark matter haloes. *MNRAS*, 377:L5–L9, April 2007.

[86] I. D. Gargiulo, S. A. Cora, N. D. Padilla, A. M. Muñoz Arancibia, A. N. Ruiz, A. A. Orsi, T. E. Tecce, C. Weidner, and G. Bruzual. Chemoarchaeological downsizing in a hierarchical universe: impact of a top-heavy IGIMF. *MNRAS*, 446:3820–3841, February 2015.

[87] E. Gaztañaga, M. Eriksen, M. Crocce, F. J. Castander, P. Fosalba, P. Marti, R. Miquel, and A. Cabré. Cross-correlation of spectroscopic and photometric galaxy surveys: cosmology from lensing and redshift distortions. *MNRAS*, 422:2904–2930, June 2012.

[88] E. Gaztanaga and C. M. Baugh. Testing deprojection algorithms on mock angular catalogues - Evidence for a break in the power spectrum. *MNRAS*, 294:229, February 1998.

[89] E. Gaztanaga and F. Bernardeau. The skewness and kurtosis of the projected density distribution function: validity of perturbation theory. *A&A*, 331:829–837, March 1998.

[90] J. E. Geach, D. Sobral, R. C. Hickox, D. A. Wake, I. Smail, P. N. Best, C. M. Baugh, and J. P. Stott. The clustering of Hα emitters at z=2.23 from HiZELS. *MNRAS*, 426:679–689, October 2012.

[91] T. Giannantonio, P. Fosalba, R. Cawthon, Y. Omori, M. Crocce, F. Elsner, B. Leistedt, S. Dodelson, A. Benoit-Levy, D. Kirk, A. H. Bauer, B. A. Benson, G. M. Bernstein, J. Carretero, T. M. Crawford, R. Crittenden, E. Gaztanaga, G. Holder, D. Huterer, B. Jain, E. Krause, H. V. Peiris, W. J. Percival, C. L. Reichardt, A. J. Ross, B. Soergel, A. Stark, K. T. Story, J. D. Vieira, J. Weller, T. Abbott, F. B. Abdalla, S. Allam, R. Armstrong, M. Banerji, R. A. Bernstein, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, J. E. Carlstrom, A. Carnero Rosell, M. Carrasco Kind, F. J. Castander, C. L. Chang, C. E. Cunha, L. N. da Costa, C. B. D'Andrea, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, A. E. Evrard, A. Fausti Neto, E. Fernandez, D. A. Finley, B. Flaugher, J. Frieman, D. Gerdes, D. Gruen, R. A. Gruendl, G. Gutierrez, W. L. Holzapfel, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, M. March, J. L. Marshall, P. Martini, P. Melchior, R. Miquel, J. J. Mohr, R. C. Nichol, B. Nord, R. Ogando, A. A. Plazas, A. K. Romer, A. Roodman, E. S. Rykoff, M. Sako, B. R. Saliwanchik, E. Sanchez, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas,

V. Vikram, A. R. Walker, R. H. Wechsler, and J. Zuntz. CMB lensing tomography with the DES Science Verification galaxies. *ArXiv e-prints*, July 2015.

[92] F. Governato, A. Babul, T. Quinn, P. Tozzi, C. M. Baugh, N. Katz, and G. Lake. Properties of galaxy clusters: mass and correlation functions. *MNRAS*, 307:949–966, August 1999.

[93] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby. SKYNET: an efficient and robust neural network training tool for machine learning in astronomy. *MNRAS*, 441:1741–1759, June 2014.

[94] J. Green, P. Schechter, C. Baltay, R. Bean, D. Bennett, R. Brown, C. Conselice, M. Donahue, X. Fan, B. S. Gaudi, and et al. Wide-Field InfraRed Survey Telescope (WFIRST) Final Report. *ArXiv e-prints*, August 2012.

[95] M. A. K. Gross, R. S. Somerville, J. R. Primack, J. Holtzman, and A. Klypin. Cold dark matter variant cosmological models - I. Simulations and preliminary comparisons. *MNRAS*, 301:81–94, November 1998.

[96] Q. Guo, S. White, R. E. Angulo, B. Henriques, G. Lemson, M. Boylan-Kolchin, P. Thomas, and C. Short. Galaxy formation in WMAP1 and WMAP7 cosmologies. *MNRAS*, 428:1351–1365, January 2013.

[97] Q. Guo, S. White, M. Boylan-Kolchin, G. De Lucia, G. Kauffmann, G. Lemson, C. Li, V. Springel, and S. Weinmann. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. *MNRAS*, 413:101–131, May 2011.

[98] A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *ApJ*, 417:19, November 1993.

[99] G. Harker, S. Cole, J. Helly, C. Frenk, and A. Jenkins. A marked correlation function analysis of halo formation times in the Millennium Simulation. *MNRAS*, 367:1039–1049, April 2006.

[100] J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *A&A*, 464:399–404, March 2007.

[101] A. P. Hearin, D. F. Watson, M. R. Becker, R. Reyes, A. A. Berlind, and A. R. Zentner. The dark side of galaxy colour: evidence from new SDSS measurements of galaxy clustering and lensing. *MNRAS*, 444:729–743, October 2014.

[102] A. P. Hearin, D. F. Watson, and F. C. van den Bosch. Beyond halo mass: galactic conformity as a smoking gun of central galaxy assembly bias. *MNRAS*, 452:1958–1969, September 2015.

[103] A. P. Hearin, A. R. Zentner, A. A. Berlind, and J. A. Newman. SHAM beyond clustering: new tests of galaxy-halo abundance matching with galaxy groups. *MNRAS*, 433:659–680, July 2013.

[104] J. C. Helly, S. Cole, C. S. Frenk, C. M. Baugh, A. Benson, and C. Lacey. Galaxy formation using halo merger histories taken from N-body simulations. *MNRAS*, 338:903–912, February 2003.

[105] H. Hoekstra, Y. Mellier, L. van Waerbeke, E. Semboloni, L. Fu, M. J. Hudson, L. C. Parker, I. Tereno, and K. Benabed. First Cosmic Shear Results from the Canada-France-Hawaii Telescope Wide Synoptic Legacy Survey. *ApJ*, 647:116–127, August 2006.

[106] K. Hoffmann, J. Bel, and E. Gaztañaga. Comparing halo bias from abundance and clustering. *MNRAS*, 450:1674–1692, June 2015.

[107] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929.

[108] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *ApJS*, 192:14, February 2011.

[109] M. Jarvis, E. Sheldon, J. Zuntz, T. Kacprzak, S. L. Bridle, A. Amara, R. Armstrong, M. R. Becker, G. M. Bernstein, C. Bonnett, C. Chang, R. Das, J. P. Dietrich, A. Drlica-Wagner, T. F. Eifler, C. Gangkofner, D. Gruen, M. Hirsch, E. M. Huff, B. Jain, S. Kent, D. Kirk, N. MacCrann, P. Melchior, A. A. Plazas, A. Refregier, B. Rowe, E. S. Rykoff, S. Samuroff, C. Sánchez, E. Suchyta, M. A. Troxel,

V. Vikram, T. Abbott, F. B. Abdalla, S. Allam, J. Annis, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, P. Doel, A. Fausti Neto, B. Flaugher, P. Fosalba, J. Frieman, E. Gaztanaga, D. W. Gerdes, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, M. March, P. Martini, R. Miquel, J. J. Mohr, E. Neilsen, B. Nord, R. Ogando, K. Reil, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas, A. R. Walker, and R. H. Wechsler. The DES Science Verification Weak Lensing Shear Catalogs. *ArXiv e-prints*, July 2015.

[110] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. *MNRAS*, 321:372–384, February 2001.

[111] L. Jiang, J. C. Helly, S. Cole, and C. S. Frenk. N-body dark matter haloes with simple hierarchical histories. *MNRAS*, 440:2115–2135, May 2014.

[112] N. Jiménez, S. A. Cora, L. P. Bassino, T. E. Tecce, and A. V. Smith Castelli. The bright end of the colour-magnitude relation of cluster galaxies. *MNRAS*, 417:785–800, October 2011.

[113] Y. P. Jing, H. J. Mo, and G. Boerner. Spatial Correlation Function and Pairwise Velocity Dispersion of Galaxies: Cold Dark Matter Models versus the Las Campanas Survey. *ApJ*, 494:1, February 1998.

[114] Y. P. Jing, Y. Suto, and H. J. Mo. The Dependence of Dark Halo Clustering on Formation Epoch and Concentration Parameter. *ApJ*, 657:664–668, March 2007.

[115] E. Jullo, J. Rhodes, A. Kiessling, J. E. Taylor, R. Massey, J. Berge, C. Schimd, J.-P. Kneib, and N. Scoville. COSMOS: Stochastic Bias from Measurements of Weak Lensing and Galaxy Clustering. *ApJ*, 750:37, May 2012.

[116] N. Kaiser. On the spatial correlations of Abell clusters. *ApJL*, 284:L9–L12, September 1984.

[117] N. Kaiser, W. Burgett, K. Chambers, L. Denneau, J. Heasley, R. Jedicke, E. Magnier, J. Morgan, P. Onaka, and J. Tonry. The Pan-STARRS wide-field optical/NIR imaging survey. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7733 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0, July 2010.

[118] N. Kaiser and G. Squires. Mapping the dark matter with weak gravitational lensing. *ApJ*, 404:441–450, February 1993.

[119] G. Kauffmann, J. M. Colberg, A. Diaferio, and S. D. M. White. Clustering of galaxies in a hierarchical universe - I. Methods and results at z=0. *MNRAS*, 303:188–206, February 1999.

[120] A. Knebe, A. Green, and J. Binney. Multi-level adaptive particle mesh (MLAPM): a c code for cosmological simulations. *MNRAS*, 325:845–864, August 2001.

[121] A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, and et al. Haloes gone MAD: The Halo-Finder Comparison Project. *MNRAS*, 415:2293–2318, August 2011.

[122] A. Knebe, F. R. Pearce, P. A. Thomas, A. Benson, J. Blaizot, R. Bower, J. Carretero, F. J. Castander, A. Cattaneo, S. A. Cora, and et al. nIFTy cosmology: comparison of galaxy formation models. *MNRAS*, 451:4029–4059, August 2015.

[123] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. Allgood, and J. R. Primack. The Dark Side of the Halo Occupation Distribution. *ApJ*, 609:35–49, July 2004.

[124] A. V. Kravtsov and A. A. Klypin. The Origin and Evolution of Halo Bias in Linear and Nonlinear Regimes. *ApJ*, 520:437–453, August 1999.

[125] A. V. Kravtsov, A. A. Klypin, and A. M. Khokhlov. Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations. *ApJS*, 111:73–94, July 1997.

[126] I. Lacerna and N. Padilla. The nature of assembly bias - I. Clues from a ΛCDM cosmology. *MNRAS*, 412:1283–1294, April 2011.

[127] I. Lacerna and N. Padilla. The nature of assembly bias - II. Halo spin. *MNRAS*, 426:L26–L30, October 2012.

[128] I. Lacerna, N. Padilla, and F. Stasyszyn. The nature of assembly bias - III. Observational properties. *MNRAS*, 443:3107–3117, October 2014.

[129] O. Lahav and A. R Liddle. The Cosmological Parameters 2010. *ArXiv e-prints*, February 2010.

[130] S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993.

[131] A. E. Lange, P. A. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, K. Coble, B. P. Crill, P. de Bernardis, P. Farese, P. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V. V. Hristov, A. Iacoangeli, A. H. Jaffe, L. Martinis, S. Masi, P. D. Mauskopf, A. Melchiorri, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J. E. Ruhl, F. Scaramuzzi, and D. Sforna. Cosmological parameters from the first results of Boomerang. *Phys.Rev.D*, 63(4):042001, February 2001.

[132] R. Laureijs, J. Amiaux, S. Arduini, J. . Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, and et al. Euclid Definition Study Report. *ArXiv e-prints*, October 2011.

[133] A. Leauthaud, J. Tinker, P. S. Behroozi, M. T. Busha, and R. H. Wechsler. A Theoretical Framework for Combining Techniques that Probe the Link Between Galaxies and Dark Matter. *ApJ*, 738:45, September 2011.

[134] J. Lee and S. F. Shandarin. Comparison of Analytical Mass Functions with Numerical Simulations. *ApJL*, 517:L5–L8, May 1999.

[135] B. Leistedt, H. V. Peiris, F. Elsner, A. Benoit-Lévy, A. Amara, A. H. Bauer, M. R. Becker, C. Bonnett, C. Bruderer, M. T. Busha, M. Carrasco Kind, C. Chang, M. Crocce, L. N. da Costa, E. Gaztanaga, E. M. Huff, O. Lahav, A. Palmese, W. J. Percival, A. Refregier, A. J. Ross, E. Rozo, E. S. Rykoff, C. Sánchez, I. Sadeh, I. Sevilla-Noarbe, F. Sobreira, E. Suchyta, M. E. C. Swanson, R. H. Wechsler, F. B. Abdalla, S. Allam, M. Banerji, G. M. Bernstein, R. A. Bernstein, E. Bertin, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, J. Carretero, C. E. Cunha, C. B. D'Andrea, D. L. DePoy, S. Desai, H. T. Diehl, P. Doel, T. F. Eifler, A. E. Evrard, A. Fausti Neto, B. Flaugher, P. Fosalba, J. Frieman, D. W. Gerdes, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, M. Jarvis, S. Kent, K. Kuehn, N. Kuropatkin, T. S. Li, M. Lima, M. A. G.

Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, C. J. Miller, R. Miquel, R. C. Nichol, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. K. Romer, A. Roodman, E. Sanchez, B. Santiago, V. Scarpine, M. Schubnell, R. C. Smith, M. Soares-Santos, G. Tarle, J. Thaler, D. Thomas, V. Vikram, A. R. Walker, W. Wester, Y. Zhang, and J. Zuntz. Mapping and simulating systematics due to spatially-varying observing conditions in DES Science Verification data. *ArXiv e-prints*, July 2015.

[136] G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59, 1927.

[137] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein, B. Flaugher, K. Honscheid, R. Kron, O. Lahav, P. McDonald, N. Roe, D. Schlegel, and representing the DESI collaboration. The DESI Experiment, a whitepaper for Snowmass 2013. *ArXiv e-prints*, August 2013.

[138] A. Lewis, A. Challinor, and A. Lasenby. Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *ApJ*, 538:473–476, August 2000.

[139] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, and et al. LSST Science Book, Version 2.0. *ArXiv e-prints*, December 2009.

[140] N. Maccrann and Dark Energy Survey Collaboration. Early Weak Lensing Results From The Dark Energy Survey. In *American Astronomical Society Meeting Abstracts*, volume 227 of *American Astronomical Society Meeting Abstracts*, page 221.02, January 2016.

[141] R. Mandelbaum, A. Slosar, T. Baldauf, U. Seljak, C. M. Hirata, R. Nakajima, R. Reyes, and R. E. Smith. Cosmological parameter constraints from galaxy-galaxy lensing and galaxy clustering with the SDSS DR7. *MNRAS*, 432:1544–1575, June 2013.

[142] M. Manera and E. Gaztañaga. The local bias model in the large-scale halo distribution. *MNRAS*, 415:383–398, July 2011.

[143] M. Manera, L. Samushia, R. Tojeiro, C. Howlett, A. J. Ross, W. J. Percival, H. Gil-Marín, J. R. Brownstein, A. Burden, and F. Montesano. The clustering of galaxies

in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the low-redshift sample. *MNRAS*, 447:437–445, February 2015.

[144] M. Manera, R. K. Sheth, and R. Scoccimarro. Large-scale bias and the inaccuracy of the peak-background split. *MNRAS*, 402:589–602, February 2010.

[145] F. A. Marín, C. Blake, G. B. Poole, C. K. McBride, S. Brough, M. Colless, C. Contreras, W. Couch, D. J. Croton, S. Croom, and et al. The WiggleZ Dark Energy Survey: constraining galaxy bias and cosmic growth with three-point correlation functions. *MNRAS*, 432:2654–2668, July 2013.

[146] S. Masaki, Y.-T. Lin, and N. Yoshida. Modelling colour-dependent galaxy clustering in cosmological simulations. *MNRAS*, 436:2286–2300, December 2013.

[147] R. Massey, J. Rhodes, R. Ellis, N. Scoville, A. Leauthaud, A. Finoguenov, P. Capak, D. Bacon, H. Aussel, J.-P. Kneib, and et al. Dark matter maps reveal cosmic scaffolding. *Nature*, 445:286–290, January 2007.

[148] S. Matarrese, P. Coles, F. Lucchin, and L. Moscardini. Redshift evolution of clustering. *MNRAS*, 286:115–132, March 1997.

[149] A. I. Merson, C. M. Baugh, J. C. Helly, V. Gonzalez-Perez, S. Cole, R. Bielby, P. Norberg, C. S. Frenk, A. J. Benson, R. G. Bower, C. G. Lacey, and C. d. P. Lagos. Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection. *MNRAS*, 429:556–578, February 2013.

[150] H. Merz, U.-L. Pen, and H. Trac. Towards optimal parallel PM N-body codes: PMFAST. *New Astronomy*, 10:393–407, April 2005.

[151] S. Miyazaki, Y. Komiyama, H. Nakaya, Y. Doi, H. Furusawa, P. Gillingham, Y. Kamata, K. Takeshi, and K. Nariai. HyperSuprime: project overview. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6269 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0, June 2006.

[152] H. J. Mo and S. D. M. White. An analytic model for the spatial clustering of dark matter haloes. *MNRAS*, 282:347–361, September 1996.

[153] J. J. Mohr, R. Armstrong, E. Bertin, G. Daues, S. Desai, M. Gower, R. Gruendl, W. Hanlon, N. Kuropatkin, H. Lin, J. Marriner, D. Petravic, I. Sevilla, M. Swanson, T. Tomashek, D. Tucker, and B. Yanny. The Dark Energy Survey data processing and calibration system. volume 8451 of *Proc. SPIE*, September 2012.

[154] P. Monaco, E. Sefusatti, S. Borgani, M. Crocce, P. Fosalba, R. K. Sheth, and T. Theuns. An accurate tool for the fast generation of dark matter halo catalogues. *MNRAS*, 433:2389–2402, August 2013.

[155] P. Monaco, T. Theuns, and G. Taffoni. The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field. *MNRAS*, 331:587–608, April 2002.

[156] B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. *ApJ*, 710:903–923, February 2010.

[157] D. Munshi, P. Valageas, L. van Waerbeke, and A. Heavens. Cosmology with weak lensing surveys. *PhysRep*, 462:67–121, June 2008.

[158] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490:493, December 1997.

[159] C. Ngeow, J. J. Mohr, T. Alam, W. A. Barkhouse, C. Beldica, D. Cai, G. Daues, R. Plante, J. Annis, H. Lin, D. Tucker, and R. C. Smith. Cyber-infrastructure to support science and data management for the Dark Energy Survey. volume 6270 of *Proc. SPIE*, page 23, June 2006.

[160] P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton. Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics. *MNRAS*, 396:19–38, June 2009.

[161] J. Onions, A. Knebe, F. R. Pearce, S. I. Muldrew, H. Lux, S. R. Knollmann, Y. Ascasibar, P. Behroozi, and et al. Subhaloes going Notts: the subhalo-finder comparison project. *MNRAS*, 423:1200–1214, June 2012.

[162] B. W. O'Shea, G. Bryan, J. Bordner, M. L. Norman, T. Abel, R. Harkness, and A. Kritsuk. Introducing Enzo, an AMR Cosmology Application. *ArXiv Astrophysics e-prints*, March 2004.

[163] A. Paranjape, K. Kovac, W. G. Hartley, and I. Pahwa. Correlating galaxy colour and halo concentration: A tunable Halo Model of galactic conformity. *ArXiv e-prints*, March 2015.

[164] A. Paranjape, E. Sefusatti, K. C. Chan, V. Desjacques, P. Monaco, and R. K. Sheth. Bias deconstructed: unravelling the scale dependence of halo bias using real-space measurements. *MNRAS*, 436:449–459, November 2013.

[165] J. A. Peacock. The evolution of clustering and bias in the galaxy distribution. *Philosophical Transactions of the Royal Society of London Series A*, 357:133, January 1999.

[166] J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *MNRAS*, 318:1144–1156, November 2000.

[167] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

[168] U.-L. Pen. A Linear Moving Adaptive Particle-Mesh N-Body Algorithm. *ApJS*, 100:269, September 1995.

[169] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999.

[170] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16, November 2014.

[171] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. Planck 2015 results. XIII. Cosmological parameters. *ArXiv e-prints*, February 2015.

[172] J. E. Pollack, R. E. Smith, and C. Porciani. A new method to measure galaxy bias. *MNRAS*, 440:555–576, May 2014.

[173] W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974.

[174] A. Pujol and E. Gaztañaga. Are the halo occupation predictions consistent with large-scale galaxy clustering? *MNRAS*, 442:1930–1941, August 2014.

[175] A. Pujol, K. Hoffmann, N. Jiménez, and E. Gaztañaga. What determines large scale clustering: halo mass or environment? *ArXiv e-prints*, October 2015.

[176] V. Quilis. A new multidimensional adaptive mesh refinement hydro + gravity cosmological code. *MNRAS*, 352:1426–1438, August 2004.

[177] A. Refregier. Weak Gravitational Lensing by Large-Scale Structure. *ARA&A* , 41:645–668, 2003.

[178] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Aj*, 116:1009–1038, September 1998.

[179] S. A. Rodríguez-Torres, F. Prada, C.-H. Chuang, H. Guo, A. Klypin, P. Behroozi, C. H. Hahn, J. Comparat, G. Yepes, and et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Modeling the clustering and halo occupation distribution of BOSS-CMASS galaxies in the Final Data Release. *ArXiv e-prints*, September 2015.

[180] N. Roth and C. Porciani. Testing standard perturbation theory and the Eulerian local biasing scheme against N-body simulations. *MNRAS*, 415:829–844, July 2011.

[181] E. S. Rykoff, E. Rozo, and R. Keisler. Assessing Galaxy Limiting Magnitudes in Large Optical Surveys. *ArXiv e-prints*, September 2015.

[182] C. Sánchez, M. Carrasco Kind, H. Lin, R. Miquel, F. B. Abdalla, A. Amara, M. Banerji, C. Bonnett, R. Brunner, D. Capozzi, A. Carnero, F. J. Castander, L. A. N. da Costa, C. Cunha, A. Fausti, D. Gerdes, N. Greisel, J. Gschwend, W. Hartley, S. Jouvel, O. Lahav, M. Lima, M. A. G. Maia, P. Martí, R. L. C. Ogando, F. Ostrovski, P. Pellegrini, M. M. Rau, I. Sadeh, S. Seitz, I. Sevilla-Noarbe, A. Sypniewski, J. de Vicente, T. Abbot, S. S. Allam, D. Atlee, G. Bernstein, J. P. Bernstein, E. Buckley-Geer, D. Burke, M. J. Childress, T. Davis, D. L. DePoy, A. Dey,

S. Desai, H. T. Diehl, P. Doel, J. Estrada, A. Evrard, E. Fernández, D. Finley, B. Flaugher, J. Frieman, E. Gaztanaga, K. Glazebrook, K. Honscheid, A. Kim, K. Kuehn, N. Kuropatkin, C. Lidman, M. Makler, J. L. Marshall, R. C. Nichol, A. Roodman, E. Sánchez, B. X. Santiago, M. Sako, R. Scalzo, R. C. Smith, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, S. A. Uddin, F. Valdés, A. Walker, F. Yuan, and J. Zuntz. Photometric redshift analysis in the Dark Energy Survey Science Verification data. *MNRAS*, 445:1482–1506, December 2014.

[183] T. Sawala, C. S. Frenk, R. A. Crain, A. Jenkins, J. Schaye, T. Theuns, and J. Zavala. The abundance of (not just) dark matter haloes. *MNRAS*, 431:1366–1382, May 2013.

[184] R. J. Scherrer and D. H. Weinberg. Constraints on the Effects of Locally Biased Galaxy Formation. *ApJ*, 504:607–611, September 1998.

[185] P. Schneider. Cosmic Shear and Biasing. *ApJ*, 498:43–47, May 1998.

[186] R. Scoccimarro and R. K. Sheth. PTHALOS: a fast method for generating mock galaxy distributions. *MNRAS*, 329:629–640, January 2002.

[187] R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain. How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *ApJ*, 546:20–34, January 2001.

[188] U. Seljak. Analytic model for galaxy and dark matter clustering. *MNRAS*, 318:203–213, October 2000.

[189] U. Seljak and M. S. Warren. Large-scale bias and stochasticity of haloes and dark matter. *MNRAS*, 355:129–136, November 2004.

[190] U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469:437, October 1996.

[191] I. Sevilla, R. Armstrong, E. Bertin, A. Carlson, G. Daues, S. Desai, M. Gower, R. Gruendl, W. Hanlon, M. Jarvis, R. Kessler, N. Kuropatkin, H. Lin, J. Marriner, J. Mohr, D. Petravick, E. Sheldon, M. E. C. Swanson, T. Tomashek, D. Tucker, Y. Yang, B. Yanny, and for the DES Collaboration. The Dark Energy Survey Data Management System. *ArXiv e-prints*, September 2011.

[192] H. Shan, J.-P. Kneib, C. Tao, Z. Fan, M. Jauzac, M. Limousin, R. Massey, J. Rhodes, K. Thanjavur, and H. J. McCracken. Weak Lensing Measurement of Galaxy Clusters in the CFHTLS-Wide Survey. *ApJ*, 748:56, March 2012.

[193] E. S. Sheldon. An implementation of Bayesian lensing shear measurement. *MNRAS*, 444:L25–L29, October 2014.

[194] R. K. Sheth, A. Diaferio, L. Hui, and R. Scoccimarro. On the streaming motions of haloes and galaxies. *MNRAS*, 326:463–472, September 2001.

[195] R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, 323:1–12, May 2001.

[196] R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. *MNRAS*, 308:119–126, September 1999.

[197] R. K. Sheth and G. Tormen. On the environmental dependence of halo formation. *MNRAS*, 350:1385–1390, June 2004.

[198] P. Simon, M. Hetterscheidt, M. Schirmer, T. Erben, P. Schneider, C. Wolf, and K. Meisenheimer. GaBoDS: The Garching-Bonn Deep Survey. VI. Probing galaxy bias using weak gravitational lensing. *A&A*, 461:861–879, January 2007.

[199] P. Simon, M. Hetterscheidt, C. Wolf, K. Meisenheimer, H. Hildebrandt, P. Schneider, M. Schirmer, and T. Erben. Relative clustering and the joint halo occupation distribution of red sequence and blue-cloud galaxies in COMBO-17. *MNRAS*, 398:807–831, September 2009.

[200] R. Skibba, R. K. Sheth, A. J. Connolly, and R. Scranton. The luminosity-weighted or 'marked' correlation function. *MNRAS*, 369:68–76, June 2006.

[201] R. A. Skibba and R. K. Sheth. A halo model of galaxy colours and clustering in the Sloan Digital Sky Survey. *MNRAS*, 392:1080–1091, January 2009.

[202] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, 341:1311–1332, June 2003.

[203] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *ApJS*, 148:175–194, September 2003.

[204] V. Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *MNRAS*, 401:791–851, January 2010.

[205] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White. The Aquarius Project: the subhaloes of galactic haloes. *MNRAS*, 391:1685–1711, December 2008.

[206] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, and et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005.

[207] V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann. Populating a cluster of galaxies - I. Results at [formmu2]z=0. *MNRAS*, 328:726–750, December 2001.

[208] M. E. C. Swanson, M. Tegmark, M. Blanton, and I. Zehavi. SDSS galaxy clustering: luminosity and colour dependence and stochasticity. *MNRAS*, 385:1635–1655, April 2008.

[209] A. Tasitsiomi, A. V. Kravtsov, R. H. Wechsler, and J. R. Primack. Modeling Galaxy-Mass Correlations in Dissipationless Simulations. *ApJ*, 614:533–546, October 2004.

[210] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein. Solving large scale structure in ten easy steps with COLA. *Journal of Cosmology and Astroparticle Physics*, 6:036, June 2013.

[211] M. Tegmark and P. J. E. Peebles. The Time Evolution of Bias. *ApJL*, 500:L79–L82, June 1998.

[212] R. Teyssier. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *A&A*, 385:337–364, April 2002.

[213] The Dark Energy Survey Collaboration. The Dark Energy Survey. *ArXiv Astrophysics e-prints*, October 2005.

[214] The Dark Energy Survey Collaboration, T. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, R. Armstrong, D. Bacon, M. Banerji, A. H. Bauer, E. Baxter, M. R. Becker, A. Benoit-Lévy, R. A. Bernstein, G. M. Bernstein, E. Bertin, J. Blazek, C. Bonnett, S. L. Bridle, D. Brooks, C. Bruderer, E. Buckley-Geer, D. L. Burke, M. T. Busha, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, C. Chang, J. Clampitt, M. Crocce, C. E. Cunha, C. B. D'Andrea,

L. N. da Costa, R. Das, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, A. Drlica-Wagner, G. Efstathiou, T. F. Eifler, B. Erickson, J. Estrada, A. E. Evrard, A. Fausti Neto, E. Fernandez, D. A. Finley, B. Flaugher, P. Fosalba, O. Friedrich, J. Frieman, C. Gangkofner, J. Garcia-Bellido, E. Gaztanaga, D. W. Gerdes, D. Gruen, R. A. Gruendl, G. Gutierrez, W. Hartley, M. Hirsch, K. Honscheid, E. M. Huff, B. Jain, D. J. James, M. Jarvis, T. Kacprzak, S. Kent, D. Kirk, E. Krause, A. Kravtsov, K. Kuehn, N. Kuropatkin, J. Kwan, O. Lahav, B. Leistedt, T. S. Li, M. Lima, H. Lin, N. MacCrann, M. March, J. L. Marshall, P. Martini, R. G. McMahon, P. Melchior, C. J. Miller, R. Miquel, J. J. Mohr, E. Neilsen, R. C. Nichol, A. Nicola, B. Nord, R. Ogando, A. Palmese, H. V. Peiris, A. A. Plazas, A. Refregier, N. Roe, A. K. Romer, A. Roodman, B. Rowe, E. S. Rykoff, C. Sabiu, I. Sadeh, M. Sako, S. Samuroff, C. Sánchez, E. Sanchez, H. Seo, I. Sevilla-Noarbe, E. Sheldon, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas, M. A. Troxel, V. Vikram, A. R. Walker, R. H. Wechsler, J. Weller, Y. Zhang, and J. Zuntz. Cosmology from Cosmic Shear with DES Science Verification Data. *ArXiv e-prints*, July 2015.

[215] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz. Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. *ApJ*, 688:709–728, December 2008.

[216] J. L. Tinker, C. Conroy, P. Norberg, S. G. Patiri, D. H. Weinberg, and M. S. Warren. Void Statistics in Large Galaxy Redshift Surveys: Does Halo Occupation of Field Galaxies Depend on Environment? *ApJ*, 686:53–71, October 2008.

[217] J. L. Tinker, A. Leauthaud, K. Bundy, M. R. George, P. Behroozi, R. Massey, J. Rhodes, and R. H. Wechsler. Evolution of the Stellar-to-dark Matter Relation: Separating Star-forming and Passive Galaxies from z = 1 to 0. *ApJ*, 778:93, December 2013.

[218] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber. The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. *ApJ*, 724:878–886, December 2010.

[219] J. L. Tinker, D. H. Weinberg, and M. S. Warren. Cosmic Voids and Galaxy Bias in the Halo Occupation Framework. *ApJ*, 647:737–752, August 2006.

[220] P. B. Tissera, S. D. M. White, S. Pedrosa, and C. Scannapieco. Dark matter response to galaxy formation. *MNRAS*, 406:922–935, August 2010.

[221] F. C. van den Bosch, X. Yang, H. J. Mo, S. M. Weinmann, A. V. Macciò, S. More, M. Cacciato, R. Skibba, and X. Kang. Towards a concordant model of halo occupation statistics. *MNRAS*, 376:841–860, April 2007.

[222] L. Van Waerbeke, J. Benjamin, T. Erben, C. Heymans, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, J. Coupon, and et al. CFHTLenS: mapping the large-scale structure with gravitational lensing. *MNRAS*, 433:3373–3388, August 2013.

[223] V. Vikram, C. Chang, B. Jain, D. Bacon, A. Amara, M. R. Becker, G. Bernstein, C. Bonnett, S. Bridle, D. Brout, and et al. Wide-Field Lensing Mass Maps from DES Science Verification Data: Methodology and Detailed Analysis. *ArXiv e-prints*, April 2015.

[224] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS*, 444:1518–1547, October 2014.

[225] Y. Wang, X. Yang, H. J. Mo, F. C. van den Bosch, S. M. Weinmann, and Y. Chu. The Clustering of SDSS Galaxy Groups: Mass and Color Dependence. *ApJ*, 687:919–935, November 2008.

[226] R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov, and B. Allgood. The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. *ApJ*, 652:71–84, November 2006.

[227] A. R. Wetzel, J. D. Cohn, M. White, D. E. Holz, and M. S. Warren. The Clustering of Massive Halos. *ApJ*, 656:139–147, February 2007.

[228] S. D. M. White and M. J. Rees. Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, May 1978.

[229] V. Wild, J. A. Peacock, O. Lahav, E. Conway, S. Maddox, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, I. Lewis, S. Lumsden, D. Madgwick, P. Norberg, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: stochastic relative biasing between galaxy populations. *MNRAS*, 356:247–269, January 2005.

[230] H.-Y. Wu, E. Rozo, and R. H. Wechsler. The Effects of Halo Assembly Bias on Self-Calibration in Galaxy Cluster Surveys. *ApJ*, 688:729–741, December 2008.

[231] X. Yang, H. J. Mo, and F. C. van den Bosch. Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *MNRAS*, 339:1057–1080, March 2003.

[232] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, N. Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. *Aj*, 120:1579–1587, September 2000.

[233] I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall, A. A. Berlind, J. Brinkmann, J. A. Frieman, and et al. Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *ApJ*, 736:59, July 2011.

[234] I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth, M. A. Strauss, I. Kayo, and et al. The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, 630:1–27, September 2005.

[235] A. R. Zentner, A. P. Hearin, and F. C. van den Bosch. Galaxy assembly bias: a significant source of systematic error in the galaxy-halo relationship. *MNRAS*, 443:3044–3067, October 2014.

[236] Z. Zheng, A. L. Coil, and I. Zehavi. Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering. *ApJ*, 667:760–779, October 2007.

[237] J. Zuntz, T. Kacprzak, L. Voigt, M. Hirsch, B. Rowe, and S. Bridle. IM3SHAPE: a maximum likelihood galaxy shear measurement code for cosmic gravitational lensing. *MNRAS*, 434:1604–1618, September 2013.