Universitat Autònoma de Barcelona
Departament de Física

Cosmology
with
Galaxy Clustering

Kai Delf Hoffmann

first Advisor: Prof. Enrique Gaztañaga
second Advisor: Prof. Francisco Javier Castander
Tutor: Prof. Enrique Fernandez

INSTITUT D'ESTUDIS
ESPACIALS
DE CATALUNYA
IEEC
CSIC
INSTITUTO DE
CIENCIAS DEL
ESPACIO
ICE

Memòria presentada al Departament de Fìsica de la
Universitat Autònoma de Barcelona per Kai Hoffmann
per a optar al grau de Doctor en Física

Kai Delf Hoffmann

first Advisor:

Prof. Enrique Gaztañaga

second Advisor:

Prof. Francisco Javier Castander

Tutor:

Prof. Enrique Fernandez

Abstract

For constraining cosmological models via the growth of large-scale matter fluctuations it is important to understand how the observed galaxies trace the full matter density field. The relation between the density fields of matter and galaxies is often approximated by a second-order expansion of a so-called bias function. The freedom of the parameters in the bias function weakens cosmological constraints from observations. In this thesis we study two methods for determining the bias parameters independently from the growth. Our analysis is based on the matter field from the large MICE Grand Challenge simulation. Haloes, identified in this simulation, are associated with galaxies. The first method is to measure the bias parameters directly from third-order statistics of the halo and matter distributions. The second method is to predict them from the abundance of haloes as a function of halo mass (hereafter referred to as mass function).

Our bias estimations from third-order statistics are based on three-point auto- and cross-correlations of halo and matter density fields in three dimensional configuration space. Using three-point auto-correlations and a local quadratic bias model we find a $\sim 20\%$ overestimation of the linear bias parameter with respect to the reference from two-point correlations. This deviation can originate from ignoring non-local and higher-order contributions to the bias function, as well as from systematics in the measurements. The effect of such inaccuracies in the bias estimations on growth measurements are comparable with errors in our measurements, coming from sampling variance and noise. We also present a new method for measuring the growth which does not require a model for the dark matter three-point correlation. Results from both approaches are in good agreement with predictions. By combining three-point auto- and cross-correlations one can either measure the linear bias without being affected by quadratic (local or non-local) terms in the bias functions or one can isolate such terms and compare them to predictions. Our linear bias measurements from such combinations are in very good agreement with the reference linear bias. The comparison of the non-local contributions with predictions reveals a strong scale dependence of the measurements with significant deviations from the predictions, even at very large scales.

Our second approach for obtaining the bias parameters are predictions derived from the mass function via the peak-background split approach. We find significant $5-10\%$ deviations between these predictions and the reference from two-point clustering. These deviations can only partly be explained with systematics affecting the bias predictions, coming from the halo mass function binning, the mass function error estimation and the mass function parameterisation from which the bias predictions are derived. Studying the mass function we find unifying relations between different mass function parameterisation. Furthermore, we find that the standard Jack-Knife method overestimates the mass function error covariance in the low mass range. We explain these deviations and present a new improved covariance estimator.

Resumen

Per à constrènyer models cosmològics mitjançant el creixement de les fluctuacions a gran escala de la matèria és cabdal entendre com les galàxies que observem tracen el camp de densitat de tot el conjunt de matèria. La relació entre el camp de densitat de matèria i el de galàxies s'acostuma a aproximar amb una expansió de segon ordre de la funció anomenada bias. La llibertat en els paràmetres d'aquesta funció redueix la informació cosmolgica que es pot extreure de les observacions. En aquesta tesi estudiem dos mètodes per determinar els paràmetres del bias independentment del creixement. L'anàlisi es basa en la distribució de matèria de la gran simulació MICE Grand Challenge. Als halos, identificats en aquesta simulació, se'ls associen galàxies. El primer mètode consisteix en mesurar diráectament els paràmetres del bias d'estadstiques de tercer ordre de les distribucions d'halos i de matèria. El segon en predir-los a partir de l'abundància d'halos en funció de la seva massa (concepte al qual ens referirem com a funció de massa).

Les nostres estimacions del bias amb estadstiques de tercer ordre es basen en les autocorrelacions i correlacions creuades de tres punts dels camps de densitat d'halos i de matèria, en l'espai de configuració tridimensional. Usant les autocorrelacion de tres punts i un model local i quadràtic del bias trobem una sobreestimació del $\sim 20\%$ en el paràmetre lineal del bias respecte a la referència provinent de correlacions de dos punts. Aquesta desviació es pot deure a ignorar contribucions no locals i d'ordre superior a la funció bias, aix com sistemàtics en les mesures.

L'efecte d'aquestes inexactituds en les estimacions del bias en les mesures del creixement són comparables amb els errors en les nostres mesures, procedents de la variància de la mostra i del soroll. També presentem un nou mètode per mesurar el creixement que no requereix un model per a la correlació de tres punts de la matèria fosca. Els resultats d'ambdós enfocaments estan en acord amb les prediccions.

Combinant les autocorrelacions i les correlacions creuades de tres punts, per una banda podem mesurar el bias lineal sense ser afectats per termes quadràtics (locals o no locals) en les funcions del bias, i de l'altra podem allar aquests termes i comparar-los amb les prediccions. Les nostres mesures de bias lineal a partir d'aquestes combinacions són molt consistents amb el bias lineal de referència. La comparació de les contribucions no lineals amb les prediccions revelen una forta dependència de les mesures amb desviacions significatives de les prediccions, incls a escales molt grans.

El nostre segon enfoc per obtenir els paràmetres de bias són prediccions derivades de la funció de massa a través de l'aproximació de "peak-background split". Trobem desviacions significatives del $5-10\%$ entre aquestes prediccions i la referència a partir de les estadstiques de dos punts. Aquestes desviacions poden ser explicades només en part a partir dels sistemàtics que afecten les prediccions de bias, provinent del "binning" de la funció de massa d'halos, l'estimació de l'error de la funció de massa i la parametrització de la funció de massa a partir de la qual se'n deriven les prediccions de bias. Estudiant la funció de massa trobem relacions entre diferents parametritzacions de la funció de massa. A més, trobem que el mètode estàndard de Jack-Knife sobreestima la covariança d'error de la funció de massa en el rang de baixa massa. Expliquem aquestes desviacions i presentem un nou i estimador de covariança millorat.

Contents

Chapter 1

Introduction

Cosmology has been a part of human culture for thousands of years. However, it remained a subject of mythologies and philosophy until around one hundred years ago the technological and theoretical advance made the study of the Universe become a subject of empirical science. An analytical approach to a description of the complete Universe became possible with the development of the theory of general relativity which relates the spacetime to the universal energy content (Einstein, 1915). Based on this theory Friedmann (1922) predicted an expanding universe as a solution of Einsteins field equations. Later Lemaître pointed out that such an expansion is consistent with the correlation between distance and redshift of galaxies, reported by Hubble (1929). The increasing precision in the measurements lead finally to the discovery that the expansion of the Universe is accelerated (Riess et al., 1998; Perlmutter et al., 1999).

Such an accelerated expansion can be explained with a constant term in Einsteins field equations which possibly results from the intrinsic energy of the vacuum with the effect of a negative pressure (e.g. Carroll, 2001). Current observations suggests that this so-called dark energy constitutes roughly 73% of the energy in the Universe today. However, it is many orders of magnitude weaker than the vacuum energy predicted by quantum field theory. Understanding the nature of the accelerated expansion therefore constitutes a major problem in physics today. Its solution might require a more complicated dark energy model, possibly with time and spatial dependency, interaction with matter or a modification of gravity at large scales. Discriminating these different theories requires an increased precision in the measurements of the cosmic acceleration, which is the purpose of several ongoing and future observational programs, such as BOSS, VIPERS, PAU, DES or Euclid. This increase in precision can be archived by observing larger volumes of the Universe, increasing the number densities of observed galaxies or improving the estimation of galaxy properties, such as shape or redshift. Furthermore one can combine independent measurements of cosmological parameters (e.g. from baryon acoustic oscillations, supernovae, cluster counts) in a joint analysis. Another important probe for cosmological models is the increase of large scale galaxy density fluctuations with time.

An inhomogeneity in the large scale distribution of galaxies (historically referred to as nebulae) has been notices already by Wright (1750), Messier (1781) or Herschel (1784). Hinks (1911) was the first who suggested to 'discuss the spiral nebula distribution more on its own merit, and less with an eye to the galactic poles'. The galaxy catalogue from the Lick survey Shane and Wirtanen (1954) revealed for the first time the filamentary distribution of galaxies in the sky. The study of these structures at different redshifts became possible with the invention of photon counting detectors, which allowed for fast redshift measurements for large numbers of objects. The first surveys using this technology where the CfA (Huchra et al., 1983) and the CfA2 surveys (Geller and Huchra, 1989). Large scale structures in the galaxy distribution can be expected from density fluctuations in the plasma, which filled the Universe at early time ($z \gtrsim 1000$). These initial fluctuations can still be observed today as temperature fluctuations in the Cosmic Microwave Background (hereafter referred to as CMB). The growth of these fluctuations is sensitive to different characteristics of the Universe, such as its mean matter density, its expansion and also how matter interacts gravitationally at large scales as outlined below.

1.1 Growth of matter density fluctuations in the expanding Universe

If the Universe is homogeneous and isotropic (which is known as the cosmological principle) its expansion can be described by the scale factor

$$a(t) = r(t)/r_0, \tag{1.1}$$

which relates a comoving scale $r(t)$ at look back time t to the value which it has at a reference time $r(t_0) = r_0$. The evolution of the scale factor therefore determines how a photon, emitted with wave length λ at look back time t is stretched as it travels through the expanding space before being observed at t_0. This stretching introduces a redshift z in the spectra of observed objects which is commonly quantified by $\lambda(t_0) = \lambda(t)(1 + z) = \lambda(t)/a(t)$.

The evolution of the scale factor can be predicted from Einsteins field equations. Assuming the cosmological principle these equations can be reduced to the Friedman equations, which relate time derivatives of the scale factor to the universal energy density content Ω_{tot}. The first Friedman equation is given by

$$\frac{H(a)^2}{H(1)^2} = \Omega_{tot}. \tag{1.2}$$

The Hubble parameter $H(t) \equiv \frac{\dot{a}}{a}$, was introduced by Hubble to describe the relation between distance and the apparent velocities of nearby galaxies, determined from their redshift (i.e. $\dot{r} \simeq \dot{a}r_0 = Hr = cz$), while c is the speed of light. The total energy density consists of contributions from the curvature k, the cosmological constant Λ (also referred to as dark energy), the radiation R and the universal cold (pressureless) matter content m:

$$\Omega_{tot} = \Omega_R a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda, \tag{1.3}$$

while it is normalised by the critical density for gravitational collapse of an Einstein-de-Sitter-Universe ($k = 0$, $\Lambda = 0$) at $z = 0$, i.e. $\Omega_{tot} = \rho_{tot}/\rho_c$, with $\rho_c = 3H(1)/8\pi G$, where G is the Gravitational constant.

From equation (1.3) it is apparent that the radiation dominates the evolution of the scale factor only in an early phase of the expansion in which $a \ll 1$ and the matter is ionised. As the scale factor increases photons are stretched and loose energy. Furthermore the total photon number decreases as matter cools down and builds neutral atoms, which is the time of the CMB emission, $z \simeq 1000$. From this time the Universe is observable as photons can travel long distances and the change of the scale factor is determined by the densities of matter and dark energy. If the dark energy is constant, while the matter density decreases with a^{-3}, the former eventually drives the accelerated expansion. However, the expansion depends on the actual values of the different energy components. Determining these values and testing if they vary with redshift, observed direction or both is therefore of high interest for investigating the nature of the accelerated expansion.

The energy densities can be determined from measurements of the Hubble parameter at different redshifts. The first Friedman equation (1.3) delivers $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$ for the present time ($a = 1$) and can therefore be rewritten as

$$H(z)^2 = H(0)^2 \sqrt{\Omega_m(z+1)^3 + (1 - \Omega_m - \Omega_\Lambda)(z+1)^2 + \Omega_\Lambda}. \tag{1.4}$$

The Hubble parameter can be obtained from observations of matter fluctuations at different redshifts in the following way. When the total energy density is dominated by Ω_m and Ω_Λ the evolution of density fluctuations with time can be well approximated in the Newtonian limit at scales much smaller than the Hubble horizon (Peebles 1980). By expressing the continuity equation, the Euler equation and the Poisson equation in comoving coordinates in terms of the normalised deviation form the mean density δ (hereafter referred to as density contrast) and combining these equations one can derive for for $|\delta| \ll 1$

$$\ddot{\delta} + 2H\dot{\delta} = 2\pi G\bar{\rho}\delta. \tag{1.5}$$

Since this equation contains no spatial derivatives the solution for the density contrast can be obtained with the ansatz $\delta(r,t) = D(t)\delta(r,t_0)$. The linear growth factor D describes how the fluctuations of the full matter distribution change with time. The solution for D, given in Section 2.3.1, depends on the energy densities. By measuring D in observations one can therefore constrain cosmological models.

1.2 Galaxies as tracers of the full matter density field

A major limitation in constraining different cosmologies via the growth factor of the full matter distribution arises from the fact that this distribution can be seen only at the discrete positions at which galaxies are located. The change of galaxy density fluctuation with time (or redshift) is not only affected by the growth factor, but also by the way galaxies form in different environments of the cosmic web. Understanding the formation of galaxies within the cosmic web is subject of intense research which typically relies on expensive simulations with many degrees of freedom. Despite being very useful for understanding galaxy formation such simulations are not practical for extracting information about the cosmological parameters from observations. This situation calls for an analytical approach for describing the galaxy-matter density relation.

Various observations, such as gravitational lensing, rotation curves of galaxies or the velocity dispersion in galaxy clusters show that galaxies are surrounded by large densities of matter which is not visible and therefore likely to consist of a new type or group of particles without or very weak electro-magnetic interaction (commonly referred to as dark matter). These observations are consistent with the standard picture of galaxy formation in which gas follows collapsing dark matter overdensities (referred to as haloes). In these high density regions the gas then cools down to form stars in the centre, which are observed as galaxies. Given that galaxies are surrounded by dark matter haloes, studies of the galaxy-matter density relation often focus on descriptions of the halo-matter density relation, which can be approached more easily theoretically and numerically.

The halo distribution can be related to the full matter distribution by simple analytic models. The freedom of the parameters in these so-called bias models reflects a degeneracy between the change of halo-density fluctuations with time and the growth of fluctuations in the underlying matter density field (see Section 2.3). For constraining cosmological models with the observed growth of galaxy densities it is therefore important to determine the parameters in the bias models independently from the growth. This can be achieved by measuring them directly in observations or predicting them from theory. In this thesis we investigate how reliable these parameters can be derived with both approaches using the cosmological dark matter N-body simulation MICE Grand Challange. The measurements of the bias parameters are studied in Chapter 2 and 3 using second- and third-order statistics of halo- and matter density distributions. In Chapter 4 we predict the bias parameters from the abundance of haloes and compare them to the measurements from the previous chapters.

Chapter 2

Growth of matter fluctuations from third-order galaxy correlations

Abstract

Measurements of the linear growth factor D at different redshifts z are key to distinguish among cosmological models. One can estimate the derivative $dD(z)/d\ln(1+z)$ from redshift space measurements of the 3D anisotropic galaxy two-point correlation $\xi(z)$, but the degeneracy of its transverse (or projected) component with galaxy bias b, i.e. $\xi_{\perp}(z) \propto D^2(z)b^2(z)$, introduces large errors in the growth measurement.

Here we study a method which breaks this degeneracy by combining second- and third-order correlations. This method uses the fact that, for Gaussian initial conditions and scales larger than 20 h^{-1}Mpc, reduced third-order matter correlations are independent of redshift (and therefore of the growth factor) while the third-order galaxy correlations depend on b. We use matter and halo catalogs from the MICE-GC simulation to test how well we can recover $b(z)$ and therefore $D(z)$ with this method in 3D real space. We also present a new approach, which enables us to measure D directly from the redshift evolution of second- and third-order galaxy correlations without the need of modelling matter correlations.

For haloes with masses lower than 10^{14} $h^{-1}\mathrm{M}_{\odot}$, we find 10% deviations between the different estimates of D and theory predictions, which are comparable to current observational errors. At higher masses we find larger differences that can probably be attributed to the breakdown of the bias model.

2.1 Introduction

Evidence that the expansion of the Universe is accelerating (Riess et al., 1998; Perlmutter et al., 1999) has revived the cosmological constant Λ, originally introduced by Einstein as an unknown fluid which may engine the observed dynamics of the Universe. Alternative explanations for the accelerated expansion could involve a modification of the gravitational laws on cosmological scales. Since these modifications of gravity can mimic well the observed accelerated expansion it is difficult to just rely on the cosmological background (i.e. the overall dynamics of the Universe) in order to verify which model is correct. However, alternative gravitational laws change the way matter fluctuations grow during the expansion history of our Universe. Measuring the growth of matter fluctuations could therefore be a powerful tool to distinguish between cosmological models (see e.g. Gaztañaga and Lobo, 2001; Lue et al., 2004; Ross et al., 2007; Guzzo et al., 2008; Song and Percival, 2009; Cabré and Gaztañaga, 2009; Samushia et al., 2012; Reid et al., 2012; Contreras et al., 2013; de la Torre et al., 2013; Steigerwald et al., 2014).

On this basis, the goal of several future and ongoing cosmological surveys, such as BOSS[1], DES[2], MS-DESI[3], PAU[4], VIPERS[5] or Euclid[6], is to measure the growth of matter fluctuations. This can be achieved by combining several observables, such as weak gravitational lensing, cluster abundance or redshift space distortions. Higher-order correlations in the galaxy distribution provide additional observables which also allow for proving the growth equation beyond linear theory from observations (e.g. see Bernardeau et al., 2002). Furthermore, higher-order correlations can be used to test the nature of the initial conditions and improve the signal-to-noise in recovering cosmological parameters (e.g. Sefusatti et al., 2006).

The relative simplicity of the fundamental predictions about amplitude and scaling of clustering statistics, must not make us overlook the fundamental difficulty that hampers large scale structure studies. The perfect, continuous (dark matter) fluid in terms of which we model the large-scale distribution of matter cannot be directly observed. Let's imagine that we are able to locate in the Universe all existing galaxies and that we know with an infinite precision their masses. Without any knowledge of how luminous galaxies trace the underlying continuous distribution of matter, even this ultimate galaxy sample would be of limited use. The problem of unveiling how the density fields of galaxies and mass map into each other is the so called galaxy biasing. Knowledge of galaxy bias, and therefore galaxy formation, can greatly improve our cosmological inferences from observations.

A common approach to model galaxy bias consists in describing the mapping between the fields of mass and galaxy density fluctuations (δ_{dm} and δ_g respectively) by a deterministic local function F. This function can be approximated by its Taylor expansion if we smooth the density field on scales that are sufficiently large to ensure that fluctuations are small,

$$\delta_g = F[\delta_{dm}] \simeq \sum_{i=0}^{N} \frac{b_i}{i!} \delta_{dm}^i, \tag{2.1}$$

where b_i are the bias parameters. It has been shown that, in this large scale limit, such a local transformation preserves the hierarchical properties of matter statistics (Fry and Gaztanaga, 1993). There is now convincing evidences about the non-linear character of the bias function (Gaztanaga, 1992; Marinoni et al., 2005; Gaztañaga et al., 2005; Marinoni et al., 2008; Kovač et al., 2011). We address non-local contributions to the bias function in Chapter 3. Since we only want to study correlations up to third order, in this thesis we shall consider bias parameters up to second order, i.e. b_1 and b_2, which is expected to be sufficient at the leading order (Fry and Gaztanaga, 1993). However, one of the goals of this Chapter is to investigate at which scale and halo mass range this expectation is fulfilled.

To study the statistical properties of the matter field we need to find the most likely value for the parameters b_i. A general approach aims at extracting them from redshift surveys using

---

[1] https://www.sdss3.org/surveys/boss.php

[2] www.darkenergysurvey.org

[3] desi.lbl.gov

[4] www.pausurvey.org

[5] http://vipers.inaf.it

[6] www.euclid-ec.org

higher-order statistics. If the initial perturbations are Gaussian and if the shape of third-order statistics are correctly described by results of the weakly non-linear perturbation theory, then one can fix the amplitude of b_i up to second order in a way which is independent from the overall amplitude of clustering (e.g. σ_8) and depends only on the shape of the linear power spectrum. This has been shown by several authors using the the skewness S_3 (Gaztanaga, 1994; Gaztanaga and Frieman, 1994), the bispectrum (Fry, 1994; Gaztanaga and Frieman, 1994; Scoccimarro, 1998; Feldman et al., 2001; Verde et al., 2002), the three-point correlation function Q, (Gaztañaga et al., 2005; Gaztañaga and Scoccimarro, 2005; Pan and Szapudi, 2005; Marín, 2011; McBride et al., 2011b; Marín et al., 2013), and the two-point cumulants C_{12} (Bernardeau, 1996; Szapudi, 1998; Gaztañaga et al., 2002; Bel and Marinoni, 2012).

The main goal of this Section is to present the bias derived from Q. We also show that, with a new approach, the growth of matter fluctuations can be measured directly from observations by getting rid of galaxy bias and without requiring any modelling of the underlying matter distribution. Despite the fact that in the present analysis we only consider real-space observables (not affected by redshift-space distortions) we argue that, as long as reduced third-order statistics are only weakly affected by redshift-space distortions at large scales and for a broad range of masses (see Fig. 2.13 Gaztañaga and Scoccimarro, 2005), the proposed method appears to be applicable on redshift galaxy surveys. Furthermore it might be applicable to an analysis of projected galaxy density maps.

This analysis is based on the new MICE-GC simulation and extends its validation presented recently by Fosalba et al. (2013a); Crocce et al. (2013); Fosalba et al. (2013b).

Detail of this simulation are presented in Section 2.2. Our estimators for both, the bias and the growth of matter fluctuations, are introduced in Section 2.3. We present our results in Section 2.4 and a summary of the work can be found in Section 2.5 together with our conclusions.

2.2 Simulation and halo samples

Our analysis is based on the Grand Challenge run of the Marenostrum Institut de Ciències de l'Espai (MICE) simulation suite to which we refer to as MICE-GC in the following. Starting from small initial density fluctuations at redshift $z = 100$ the formation of large scale cosmic structure was computed with 4096^3 gravitationally interacting collisionless particles in a 3072 h^{-1}Mpc box using the GADGET - 2 code (Springel, 2005) with a softening length of 50 h^{-1}kpc. The initial conditions were generated using the Zel'dovich approximation and a CAMB power spectrum with the power law index of $n_s = 0.95$, which was normalised to be $\sigma_8 = 0.8$ at $z = 0$. The cosmic expansion is described by the ΛCDM model for a flat universe with a mass density of $\Omega_m = \Omega_{dm} + \Omega_b = 0.25$. The density of the baryonic mass is set to $\Omega_b = 0.044$ and Ω_{dm} is the dark matter density. The dimensionless Hubble parameter is set to $h = 0.7$. More details and validation test on this simulation can be found in Fosalba et al. (2013a).

Dark matter haloes were identified as Friends-of-Friends groups (Davis et al., 1985) with a linking length of 0.2 in units of the mean particle separation. These halo catalogs and the corresponding validation checks are presented in Crocce et al. (2013).

To study the galaxy bias and estimate the growth as a function of halo mass we divide the haloes into the four redshift independent mass samples M0, M1, M2 and M3, shown in Table 2.1. They span a mass range from Milky Way like haloes (M0) up to massive galaxy

Table 2.1: Halo mass samples. N_p is the number of particles per halo, N_{halo} is the number of haloes per sample in the comoving output at $z = 0.5$. n_{halo} is the comoving number density of haloes. N_{halo} and n_{halo} are compared to the corresponding values in the light cone in Fig. 2.1.

	mass range $10^{12} M_\odot/h$	N_p	N_{halo}	n_{halo} $(10\ \mathrm{Mpc}/h)^{-3}$
M0	$0.58 - 2.32$	$20 - 80$	122300728	4.22
M1	$2.32 - 9.26$	$80 - 316$	31765907	1.10
M2	$9.26 - 100$	$316 - 3416$	8505326	0.29
M3	≥ 100	≥ 3416	280837	0.01

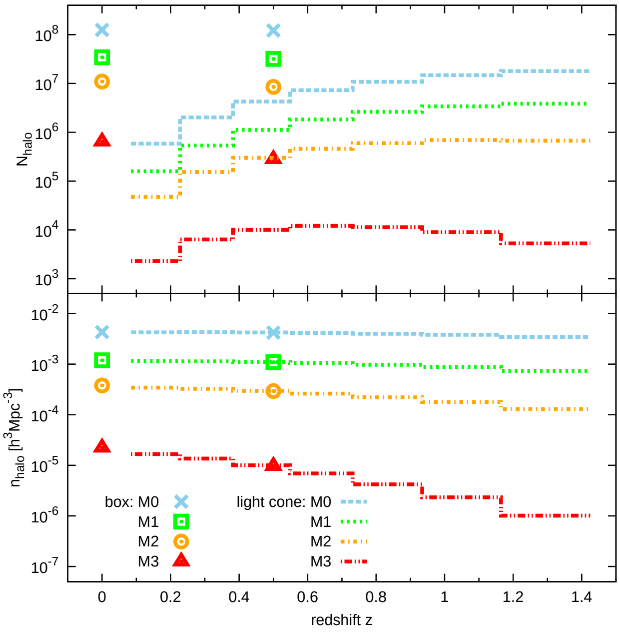

Figure 2.1: *Top:* Number of haloes in the four mass samples M0-M3 as a function of redshift in the two comoving outputs at z=0.0 and z=0.5 (symbols) and the seven redshift bins in the light cone (lines).*Bottom:* number density of the same halo mass samples as in the top panel.

clusters (M3). In the same table we show the total number and comoving number density of haloes at redshift $z = 0.5$, a characteristic redshift for current galaxy surveys, such as BOSS LRG.

We are analysing two types of simulation outputs. For a detailed study of the dark matter growth we use the full comoving output at redshift $z = 0.0, 0.5, 1.0$ and 1.5. For studying the bias estimators with minimal shot noise and sampling variance we use haloes identified in the comoving outputs at redshift $z = 0.0$ and 0.5. The investigation of the redshift evolution of the bias and growth estimators is based on seven redshift bins of the light cone output with equal width of 400 h^{-1}Mpc in comoving space over one octant of the sky. Fig. 2.1 shows the number and number density of haloes in the four mass samples for the comoving output and the light cone with respect to the redshift.

2.3 Growth and Bias Estimators

2.3.1 The growth factor

The large scale structure in the distribution of galaxies, observed today in cosmological surveys, is believed to originate from some small initially Gaussian matter density fluctuations that grew with time due to gravitational instabilities. Since the way the Universe is expanding has an impact of the growth of structures, one can use measurements of the growth to put constraints on various cosmological models.

We adopt the common definition for density fluctuations, given by $\delta(\boldsymbol{r}) = \rho(\boldsymbol{r})/\overline{\rho} - 1$, where $\rho(\boldsymbol{r})$ is the density at position $\boldsymbol{r}$ smoothed (with a spherical top-hat window) over the radius R, while $\overline{\rho}$ is the mean density of the Universe. In the linear regime (large smoothing scales) density fluctuations of matter $\delta_m(\boldsymbol{r}, z)$ evolve with the redshift z in a self similar way, thus

$$\delta_m(\boldsymbol{r}, z) = D(z)\ \delta_m(\boldsymbol{r}, z_0). \tag{2.2}$$

The reference redshift z_0 is usually arbitrarily chosen to be today, i.e. $z_0 = 0$. In the ΛCDM model the growth factor $D(z)$ depends on cosmological parameters via the Hubble expansion rate

$$H(z) = H(0)\sqrt{\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda}, \tag{2.3}$$

where Ω_m and Ω_Λ are the densities of matter and dark energy respectively, and the growth factor is then given by:

$$D(z) \propto H(z) \int_z^\infty \frac{1+z'}{H^3(z')} dz'. \tag{2.4}$$

However, in general D is also sensitive to modifications of the gravity action on cosmological scales (e.g. see Gaztañaga & Lobo 2001 and references therein). Measurements of the growth factor as a function of redshift can therefore be used to constrain cosmological models and understand the nature of cosmic expansion.

The growth factor is related to the growth rate as

$$f(z) \equiv \frac{d \ln D}{d \ln a}, \tag{2.5}$$

where $a = 1/(1+z)$.

For a spatially flat universe, we can use the growth index α defined as

$$f(z) \equiv [\Omega_m(z)]^{\alpha(z)} \,. \tag{2.6}$$

Wang and Steinhardt (1998) found that it can be approximated by

$$\alpha(z) \simeq \frac{6}{11} + \frac{30}{2662}[1 - \Omega_m(z)].$$

Recently Steigerwald et al. (2014) found an even more precise expression

$$\alpha(z) \simeq \frac{6}{11} - \frac{15}{2057}\ln[\Omega_m(z)] + \frac{205}{540421}\ln^2[\Omega_m(z)].$$

Measuring the growth factor using equation (2.2) requires knowledge of the matter density fluctuations δ_m at different redshifts, while in practice only galaxies can be observed as biased tracers of the matter field. In the following sections we describe how we quantify and measure this galaxy bias.

2.3.2 The local bias model

In this chapter our bias estimations are based on the local bias model (Fry & Gaztanaga 1993), which assumes that the galaxy (number density) fluctuation δ_g is a function of the matter density fluctuation δ_{dm} at the same location: $\delta_g = F[\delta_{dm}]$, while both fluctuations are smoothed at the same scale R. For sufficiently large smoothing scales the density fluctuations become small and we can expand this function, i.e. as in equation (2.1). For third-order statistics it is enough to stop the expansion at quadratic order Fry and Gaztanaga (see e.g. 1993)

$$\delta_g = b_1 \left\{ \delta_{dm} + \frac{c_2}{2}(\delta_{dm}^2 - \langle \delta_{dm}^2 \rangle) \right\} , \tag{2.7}$$

where b_1 and c_2 are, respectively, the linear and quadratic bias parameters which we are measuring. The term $\langle \delta_{dm}^2 \rangle$ ensures that $\langle \delta_g \rangle = 0$, where $\langle \ldots \rangle$ denotes the average over all spatial positions. Besides small density fluctuations, such a model for the bias assumes that neither the environment nor the velocity field has an impact on galaxy formation.

Recent studies have shown, that the local assumption might not be accurate for small smoothing scales when b_1 is large (Chan et al., 2012; Baldauf et al., 2012). We study such non-local contributions to the bias function in Chapter 3.

Using the information contained in the large scale distribution of galaxies at different scales we measure bias and growth with second- and third-order statistics, as described in the following sections.

2.3.3 Growth factor D from two-point matter auto-correlation

The spatial two-point auto correlation of density fluctuations can be defined as the mean product of density fluctuations δ_i at the positions $\boldsymbol{r_i}$ that are separated by the distance $r_{12} \equiv |\boldsymbol{r_1} - \boldsymbol{r_2}|$,

$$\xi(r_{12}) \equiv \langle \delta(\boldsymbol{r_1})\delta(\boldsymbol{r_2}) \rangle = \langle \delta_1 \delta_2 \rangle (r_{12}). \tag{2.8}$$

Note that the two-point correlation function depends only on the separation and thus is not sensitive to the shape of over-densities. This is in contrast with higher-order correlations: with three points we will also be able to measure deviations away from the spherically symmetric profile (Smith et al., 2006).

From equations (2.2) and (2.8) one can derive that the growth factor is related to the two-point correlation of matter as

$$\xi_{dm}(r_{12}, z) = D(z)^2 \xi_{dm}(r_{12}, z_0). \tag{2.9}$$

Our measurements of the matter correlation function in the MICE-GC simulation, presented in Fig. 2.2, indeed show a linear relation between the matter two-point correlations at different redshifts z with respect to $z_0 = 0$ on a wide range of scales. Note how at scales around $r_{12} \sim 100\ h^{-1}$Mpc the BAO peak induces some oscillations around the linear model, but the model works well for intermediate scales of $r_{12} \sim 40 - 60\ h^{-1}$Mpc.

We measured the two-point correlation by dividing the simulation volume into cubical grid cells and assigning density fluctuations to each of these cells. We then calculate the mean product of density fluctuations in grid cells that are separated by $r_{12} \pm dr$ according to definition (2.8). The measurements shown in Fig. 2.2 are based on 4 h^{-1}Mpc grid cells. Errors are derived by Jackknife resampling as described in Section 2.3.6.

As shown in Fig. 2.3, there is a good agreement between the growth factor measurements from the two-point correlation (symbols) and the theoretical prediction from equation (2.4) (dashed line) for the cosmology of the MICE-GC simulation. Deviations between predictions and measurements are at the sub-percent level and result from to non-linearities [7]. This result demonstrates that, in principle, we can obtain constrains on cosmological models by just measuring the two-point correlation function of matter.However such constraints are difficult to realise as we have to infer the correlation of the unobservable full matter field from the correlation of the observed galaxy distribution.

2.3.4 bias from halo and matter two-point correlations

A simple relation between the two-point auto correlation functions of matter and galaxies can be obtained by inserting the model for galaxy bias, given by equation (3.1), into the definition of the two-point correlation (equation 2.8). At leading order

$$\xi_g(r_{12}, z) \simeq\ b_1^2\ \xi_{dm}(r_{12}, z) + \mathcal{O}[\xi_{dm}^2]. \tag{2.10}$$

This relation only holds for sufficiently large separations r_{12}, where we can neglect terms of order ξ_{dm}^2, and small density fluctuations δ_{dm} in equation (3.1). To estimate the linear bias from ξ we define

$$b_\xi(z) \equiv \sqrt{\frac{\xi_g(r_{12}, z)}{\xi_{dm}(r_{12}, z)}} \simeq b_1 \tag{2.11}$$

which is expected to be independent of separation in the large scale limit.

The correlation functions of the halo samples M0 - M3, calculated with 4 h^{-1}Mpc grid cells, are shown in the top panel of Fig. 2.4 together with the corresponding measurement for the dark matter field. The ratios of the matter and halo correlations, shown in the bottom panel, confirm that both quantities can be related by the scale independent bias factor b_ξ between $20 \lesssim r_{12} \lesssim 60 h^{-1}$Mpc. We expect that $b_\xi(r_{12}) \simeq b_1$ at the mass and scale range of

---

[7] The increase of the deviations with redshift results from non-linearities in ξ_{dm} at $z = 0.0$. Since we use the latter as normalisation, deviations between predictions and measurements transfer to the higher redshifts. If we use ξ at high z as normalisation, as we do it later in the light cone, this effect goes into the opposite direction.

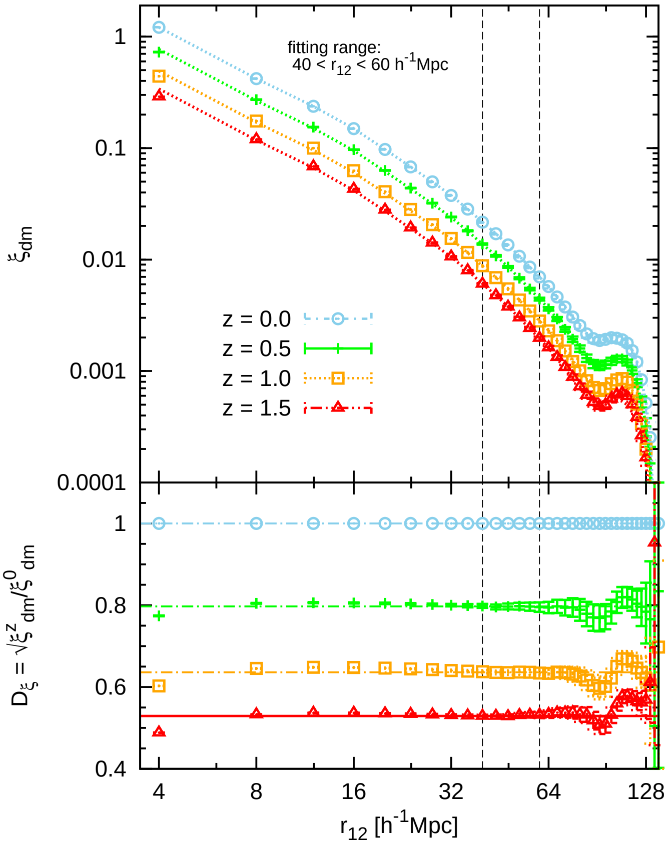

Figure 2.2: *Top:* two-point correlation ξ of the MICE-GC dark matter field measured in the comoving outputs at redshift $z = 0.0$, 0.5, 1.0 and 1.5 (blue circles, green crosses, orange squares and red triangles respectively) as a function of scale r_{12}. Dotted Lines show a fit of the amplitude of ξ at $z = 0$ to those from other redshifts between $40 - 60$ h^{-1}Mpc, via equation (2.9). *Bottom:* growth factor $D = \sqrt{\xi(r_{12}, z)/\xi(r_{12}, 0)}$ obtained from the ratio of the above correlations together with the fits displayed as dotted lines with the same colour code as the upper panel.

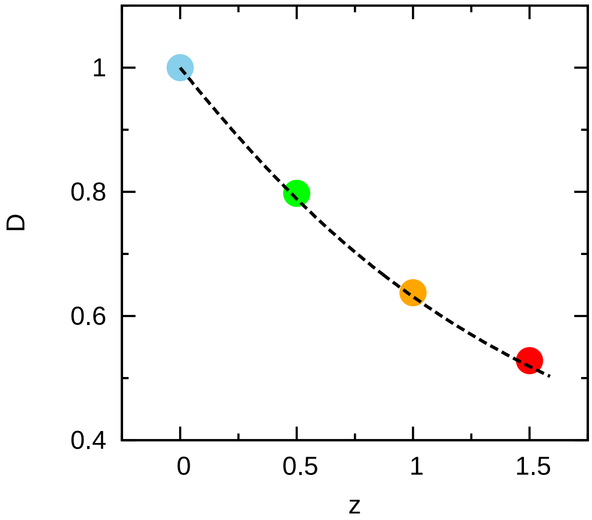

Figure 2.3: Comparison between the linear growth of matter D as a function of redshift z measured in the MICE-GC comoving outputs (symbols) and the corresponding theoretical predictions from equation (2.4) (dashed line). The MICE-GC measurements are the best fit values obtained considering the scale range 40-60h^{-1}Mpc, shown as lines with the same colour coding in Fig. 2.2.

our analysis (Crocce et al., 2013). To estimate b_ξ we perform a χ^2-fit to the ratio of the halo and matter two-point correlation in the aforementioned scale range (see Subsection 2.3.6 for details). We find the χ^2_{min} values to vary between 2.0 and 0.1. Values are smaller at $z = 0.5$ compared to $z = 0.0$. Restricting the fitting range to larger scales ($30 \lesssim r_{12} \lesssim 60h^{-1}$Mpc) also reduces the χ^2_{min}. Both findings are expected since non-linearities enter equation (2.11) at small scales and low redshift. However, restricting the fit to larger scales, as mentioned before, causes a maximum change in bias values of 1.5 percent. We therefore consider our b_ξ measurements as relatively robust, compared to the bias measurements from higher order statistics. The fitted bias factors (bottom panel of Fig. 2.4) reveal the well known increase of bias with the mass and redshift of the halo samples.

The robustness of the bias measurements from the two-point auto-correlation can be further verified by comparing them to results from the two-point cross-correlation.

The two-point cross-correlation between halo- and matter density fields, $\xi^\times$, can be measured analogously to the auto correlations as the mean product of smoothed fluctuations $\delta(\mathbf{r}) \equiv (\rho(\mathbf{r}) - \bar{\rho})/\bar{\rho}$ of each density field $\rho(\mathbf{r})$, at the positions $\mathbf{r_1}$ and $\mathbf{r_2}$ as a function of the scale $r_{12} \equiv |\mathbf{r_1} - \mathbf{r_2}|$,

$$\xi^\times(r_{12}) \equiv \langle \delta_h(\mathbf{r_1})\delta_\mathbf{m}(\mathbf{r_2})\rangle. \tag{2.12}$$

The measurements for the four halo mass samples M0-M3 at the redshifts $z = 0.0$ and $z = 0.5$ are shown in the top panels of Fig. 2.5. The amplitude increases with halo mass as expected from the peak-background split predictions from Chapter 4. The growth of matter fluctuations further contributes to an increase with redshift. At around 110 h^{-1}Mpc $\xi^\times$ shows a local maximum which results from baryonic acoustic oscillations in the initial power

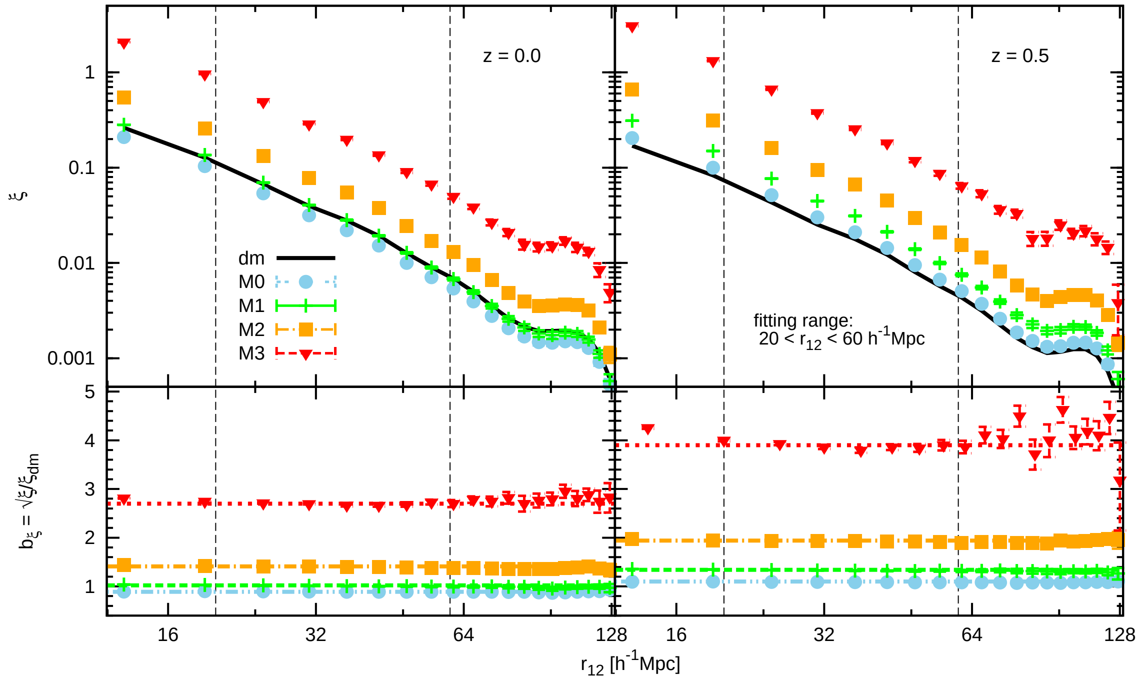

Figure 2.4: *Top*: two-point correlation ξ of the MICE-GC dark matter field (continuous lines) and the four halo mass samples M0-M3 (blue circles, green crosses, orange squares and red triangles respectively) in the comoving outputs at redshift $z = 0.0$ (left) and $z = 0.5$ (right) as a function of scale r_{12}. *Bottom*: linear bias parameter b_ξ derived from the two-point correlations via equation (2.11). Dotted lines are χ^2-fits between $20 - 60\ h^{-1}$Mpc. The minimum χ^2 values per degree of freedom are $1.05, 2.02, 0.37, 0.70$ for M0, M1, M2, M3 respectively at $z = 0.0$ and $0.42, 0.78, 0.12, 0.82$ for M0, M1, M2, M3 respectively at $z = 0.5$.

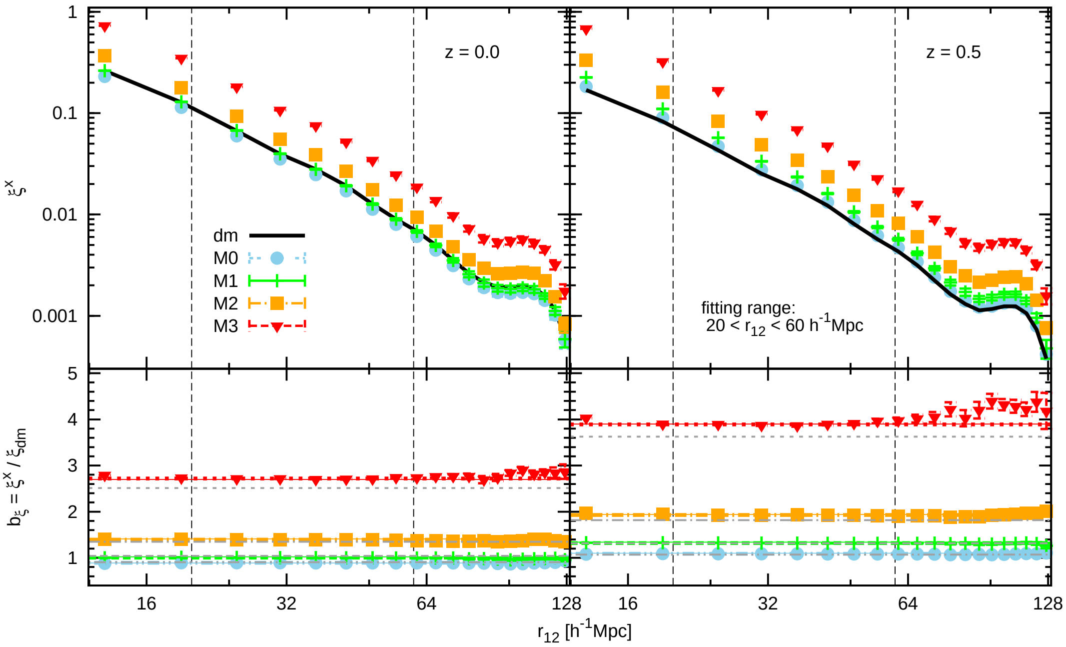

Figure 2.5: *Top*: two-point correlation ξ of the MICE-GC dark matter field (continuous lines) and the two-point halo-matter cross-correlation for the halo mass samples M0-M3 (blue circles, green crosses, orange squares and red triangles respectively) in the comoving outputs at redshift $z = 0.0$ (left) and $z = 0.5$ (right) as a function of scale r_{12}. *Bottom*: linear bias parameter b_ξ derived from the two-point correlations via equation (2.14). Colored lines are χ^2-fits between $20 - 60$ h^{-1}Mpc. The minimum χ^2 values per degree of freedom are $1.05, 1.96, 1.54, 0.23$ for M0, M1, M2, M3 respectively at $z = 0.0$ and $0.79, 1.46, 1.23, 0.77$ for M0, M1, M2, M3 respectively at $z = 0.5$. Thin solid lines are the fits to the bias from the two-point auto-correlation, shown in Fig. 2.4. Grey lines show peak-background split bias predictions from Chapter 4.

spectrum of the simulation.

A relation between the two-point halo-matter cross-correlation, the two-point matter auto-correlation, $\xi_m(r_{12}) \equiv \langle \delta_m(\mathbf{r_1})\delta_\mathbf{m}(\mathbf{r_2})\rangle$, and the halo bias can be obtained by inserting the local bias model from equation (2.1) into equation (2.12),

$$\xi^\times(r_{12}) \simeq \ b_1\ \xi_m(r_{12}) + \mathcal{O}[\xi_m^2], \tag{2.13}$$

At large scales ($r_{12} > 20h^{-1}$Mpc) we expect $\mathcal{O}[\xi_m^2]$ to be neglectable, which allows for measurements of the linear bias as

$$b_\xi^\times(r_{12}) \equiv \frac{\xi^\times(r_{12})}{\xi_m(r_{12})} \simeq b_1. \tag{2.14}$$

The measurements of b_ξ are shown in the bottom panel of Fig. 2.5. We fit b_ξ between $20-60\ h^{-1}$Mpc, where the scale-independence is a good approximation. Non-linear terms impact b_ξ at smaller scales, but also around the scale of baryonic acoustic oscillations. Comparing these bias measurements from the cross-correlation to those from the auto-correlation we find that non-linearities have a stronger effect on the auto-correlation.

The results described above allow us to estimate the growth factor of matter fluctuations from equation (2.9) in terms of galaxy (or halo) auto correlation functions as:

$$D(z) \simeq \hat{b}(z)^{-1} D_g(z), \tag{2.15}$$

where the growth factor is normalised to unity at an arbitrary redshift z_0 (i.e. $D(z_0) \equiv 1$). The bias ratio $\hat{b}(z)$ is defined as

$$\hat{b}(z) \equiv b(z)/b(z_0) \tag{2.16}$$

and the galaxy (or halo) growth factor $D_g(z)$ is:

$$D_g(z) \equiv \sqrt{\frac{\xi_g(z)}{\xi_g(z_0)}}. \tag{2.17}$$

Both definitions (2.16) and (2.17) refer to large scales, i.e r_{12} between 20-60 h^{-1}Mpc, while we find changes in the results at the percent level when we vary the fitting range. The bias at the two different redshifts z and z_0 does not need to refer to the same galaxy (or halo) populations. In Section 2.4.2 we demonstrate that taking different halo masses across the explored redshift range does not lead to unexpected growth measurements.

Equation (2.15) shows that the matter growth factor, measured from the galaxy (or halo) two-point correlation functions at different redshifts is fully degenerate with the ratio of the linear bias parameters. We therefore need an independent measurement of the bias ratio to break this degeneracy.

Note that the absolute values of the bias parameters, $b(z)$ and $b(z_0)$, do not need to be measured separately for measuring the differential growth factor between two redshift bins, as it is commonly done. Instead of the absolute bias values, we only need to measure their ratio $\hat{b}$, which can be obtained directly from third-order galaxy correlations without assumptions on the clustering of dark matter, as we will explain in Subsection 2.3.5.

By measuring the differential growth factor between two nearby redshift bins z_2 and z_1 one can also estimate the (velocity) growth rate $f(z)$ defined in equation (2.5) at the mean redshift $\bar{z} \equiv \frac{z_1+z_2}{2}$. Since the growth rate is defined as logarithmic derivative of the growth factor, it follows that

$$\begin{aligned} f(\bar{z}) &\simeq -\frac{\ln[D(z_2)/D(z_1)]}{\ln[(1+z_2)/(1+z_1)]} \\ &\simeq -\frac{\ln[b(z_1)/b(z_2)D_g(z_2)/D_g(z_1)]}{\ln[(1+z_2)/(1+z_1)]} \end{aligned} \tag{2.18}$$

Our new approach of measuring the bias ratio $\hat{b}$ with third-order galaxy correlations will enable us to measure the growth factor and the growth rate of the full matter distribution directly from the distribution of galaxies (or haloes) without assumptions on the clustering of dark matter, providing a new model independent constrain on cosmological parameters. The same approach represents an additional tool to measure $f(z)$, which is independent of redshift space distortions method (Kaiser, 1987). Note that we do not need to select the same type of objects (with respect to the halo mass) at the different redshifts. This feature allows for maximisation of the galaxy number density at each redshift.

2.3.5 Bias from the halo and matter three-point auto-correlation

In analogy to the two-point auto correlation, we can define the three-point auto correlation as

$$\zeta(r_{12}, r_{13}, r_{23}) \equiv \langle \delta(\boldsymbol{r_1})\delta(\boldsymbol{r_2})\delta(\boldsymbol{r_3})\rangle, \tag{2.19}$$

where the vectors $\mathbf{r_{12}}, \mathbf{r_{13}}, \mathbf{r_{23}}$ form triangles of different shapes and sizes. In contrast to the two-point correlation function ζ is sensitive to the shape of the matter density fluctuations. To access this additional information, we fix the length of the two triangle legs r_{12} and r_{13} while varying the angle between them, $\alpha = \mathrm{acos}(\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13})$. In the following we will therefore change the variables for characterising triangles from (r_{12}, r_{13}, r_{23}) to (r_{12}, r_{13}, α). Throughout the analysis we use triangles with $r_{13}/r_{12} = 2$ configurations, which restricts the minimum scale entering the measurements to the size of the smaller triangle leg r_{12}. Choosing configuration, such as $r_{13}/r_{12} = 1$ would introduce non-linear scales when triangles are collapsed ($\alpha = 0$).

For detecting the triples $\delta(\boldsymbol{r_1})\delta(\boldsymbol{r_2})\delta(\boldsymbol{r_3})$ we implemented the algorithm described by Barriga and Gaztañaga (2002), using the same kind of mesh as for calculating the two-point correlation with 4 and 8 h^{-1}Mpc grid cells. From the three-point correlation we then construct the reduced three-point correlation, introduced by Groth and Peebles (1977) as

$$Q \equiv \frac{\zeta(r_{12}, r_{13}, \alpha)}{\xi_{12}\xi_{13} + \xi_{12}\xi_{23} + \xi_{13}\xi_{23}}, \tag{2.20}$$

where $\xi_{ij} \equiv \xi(r_{ij})$.

Perturbation theory shows that, to leading order in the dark matter field, Q (hereafter referred to as Q_{dm}) is independent of the growth factor. This is because $\zeta \propto \langle \delta_L \delta_L \delta_L^2 \rangle \propto D^4$, so D drops in the Q ratio above (Bernardeau, 1994; Kamionkowski and Buchalter, 1999), but for galaxies Q depends on the bias parameters. These properties enable us to measure b_1 and c_2 and break the growth-bias degeneracy in equation (2.15) (Frieman and Gaztanaga, 1994; Fry, 1994; Bernardeau et al., 2002).

We test the assumption that Q_{dm} is independent of the growth factor by comparing measurements at different redshifts and scales in the MICE-GC simulation with theoretical predictions derived from second-order perturbative expansion of ξ and ζ (Bernardeau et al.,

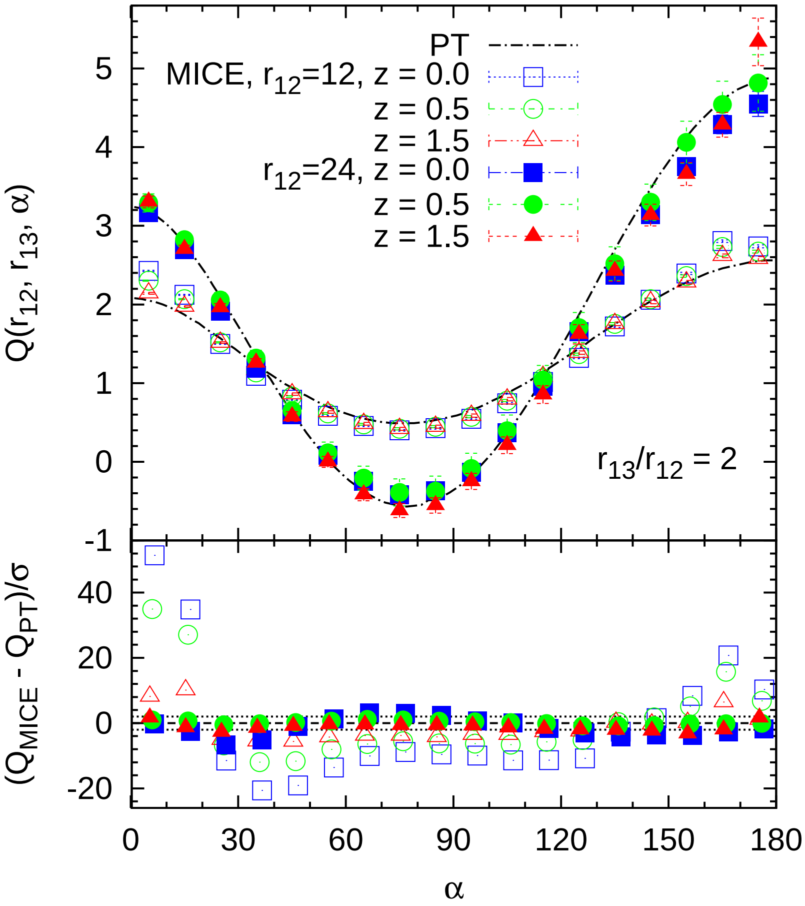

Figure 2.6: *Top*: reduced three-point correlation Q measured from the MICE-GC dark matter field in the comoving outputs at redshift $z = 0.0, 0.5, 1.5$ (blue squares, green circles, red triangles respectively) for different triangle opening angles α using $r_{12} = r_{13}/2 = 12\ h^{-1}$Mpc (open symbols) and $r_{12} = r_{13}/2 = 24\ h^{-1}$Mpc (filled symbols) compared with predictions from second-order perturbation theory (PT) using a linear power spectrum. *Bottom:* Deviations between Q from PT and measurements divided by the 1σ errors of the measurements (dashed lines correspond to $\pm 2\sigma$ discrepancies).

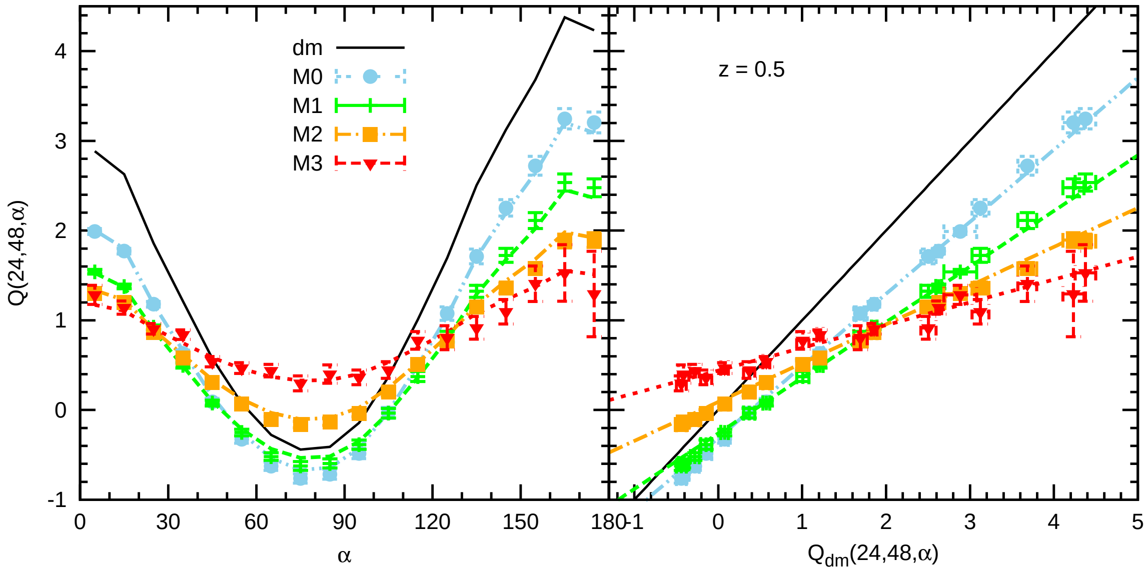

Figure 2.7: *Left*: reduced three-point correlation Q measured from the MICE-GC dark matter field and the four halo mass samples M0-M3 (black line, blue circles, green crosses, orange squares and red triangles respectively) in the comoving output at redshift $z = 0.5$ for different triangle opening angles α using triangles with $r_{12} = 24\ h^{-1}$Mpc and $r_{13} = 48\ h^{-1}$Mpc. *Right*: Q_{dm} versus Q_g at the corresponding opening angle with the same colour coding as in the left panel. Dashed lines are χ^2-fits to equation (3.6). The minimum χ^2 per degree of freedom is 6.0, 3.9, 2.0, 0.7 for M0, M1, M2 and M3 respectively. Results for redshift $z = 0.0$ are shown in Appendix 2.

2002; Barriga and Gaztañaga, 2002). The predictions are based on the MICE-GC CAMB linear power spectrum. Fig. 2.6 shows Q_{dm} at $z = 0.0$, 0.5 and 1.5 for triangles with $r_{12} = 12$ h^{-1} Mpc and $r_{13} = 24\ h^{-1}$Mpc. The measurements are based on a density mesh with 4 h^{-1}Mpc grid cells, which is the highest available resolution (see Table A.1 for details). As for the two-point correlation we derive errors for Q by Jackknife resampling (see Section 2.3.6). The values of Q show the characteristic u-shape predicted by perturbation theory, which results from the anisotropy of the shape of matter fluctuations. The amplitude of Q increases with triangle size because of the steeper slope in the two-point linear correlations at larger scales. Also Q depends only weakly on redshift while deviations between predictions and measurements become more significant at low redshift and small scales (see bottom panel of Fig. 2.6). The same effect has been reported by Fosalba et al. (2013a), who also find that the deviations decrease, when predictions are drawn from the measured instead of the CAMB power spectrum. Furthermore, these authors demonstrated that additional contributions to these deviations can result from the limited mass resolution of the simulation, especially at small scales and high redshift.

Non-linear bias

A simple relation between the bias in the local model and Q can be derived in the limit of small density fluctuations and large triangles by using equation (3.1) with the definitions (2.8), (2.19) and (2.20), and keeping second-order terms in the perturbative expansion (Frieman and Gaztanaga, 1994):

$$Q_g(\alpha) \simeq \frac{1}{b_Q}[Q_{dm}(\alpha) + c_Q]. \tag{2.21}$$

Instead of using Q_{dm}, we could also use the corresponding predictions, shown in Fig. 2.7. However, this would introduce uncertainties in the bias measurement, due to the mismatch between measurements and predictions. We interpret the parameters b_Q and c_Q as the first- and second-order bias parameters b_1 and c_2 respectively, while we expect this interpretation to be valid only in the linear regime at scales larger than roughly 20 h^{-1}Mpc. We use the notation b_Q instead of b_1 to refer to the fact that we are estimating b_1 with Q.

To measure the bias we computed Q_g for the four mass samples M0 - M3 at redshift $z = 0.0$ and $z = 0.5$ using triangles of various scales with $r_{13} = 2r_{12}$ configurations. The triangle legs consist now of 3 and 6 ±0.5 grid cells (see Table A.1 for details). We vary the size of the triangles by changing the size of the grid cells. This reduces computation time, since the number of grid cells in the simulation volume required for the measurement is minimised. Our bias measurements from Q can vary by less than 5%, when we increase the number of cells per leg instead of increasing the cell size to measure Q at larger scales. We show and discuss this effect in Appendix A.1.

Our results for $r_{12} = r_{13}/2 = 24$ h^{-1}Mpc triangles, shown in the left panel of Fig. 2.7 reveal a flattening of Q for high mass samples, as expected from equation (3.6) since b_1 increases with halo mass. In the right panel of the same figure we demonstrate that the linear relation between Q_g and Q_{dm}, given by equation (3.6) is in reasonable agreement with the measurements. We perform χ^2-fits of the dark matter results to those of the four halo samples via equation (3.6) as described in Subsection 2.3.6 and obtain the bias parameters b_Q and c_Q. These fits, shown as colored lines in Fig. 2.7, have the strongest deviations from the measurements at the smallest and highest values of Q, which might result from measurements at small angles dominating χ^2 as those have the smallest errors. The corresponding minimum values of χ^2 per degree of freedom (given in the caption of Fig. 2.7) decrease for higher mass samples as the errors of Q increase. In general we find a decrease with mass, scale and redshift. Note that these results are affected by the covariance matrix in the fit which we only know roughly from the Jackknife sampling (see Section 2.3.6 and Appendix A.1).

In order to use the bias parameter b_Q to measure the growth factor via equation (2.15) we first need to quantify deviations between b_ξ and b_Q, i.e. the linear bias b_1 inferred from the two-point function and the one from Q in the fit to equation (3.6). If the local bias model approximation works well, then we would expect $b_Q \simeq b_\xi$. A comparison is shown for different triangle scales and mass ranges in Fig. 2.8. In the top panel we show the linear bias derived with ξ and Q at redshift $z = 0.5$ as lines and symbols respectively.

In bottom panel of Fig. 2.8 we see that b_Q is up to 30% higher than b_ξ at large scales, while differences increase for smaller scales and larger values of b_ξ. Such deviations between b_ξ and b_Q have also been reported by, e.g. Manera and Gaztañaga (2011), Pollack et al. (2012), Baldauf et al. (2012), Chan et al. (2012), Moresco et al. (2014). Furthermore we find that b_Q for M3 is under predicted at small scales in contrast to results for the lower mass samples. Deviations for small triangle sizes indicate departures from the leading order perturbative expansion in which equation (3.6) is valid, while the strong deviations for the sample M3 suggest that the quadratic expansion of the bias function might not be sufficient for highly biased

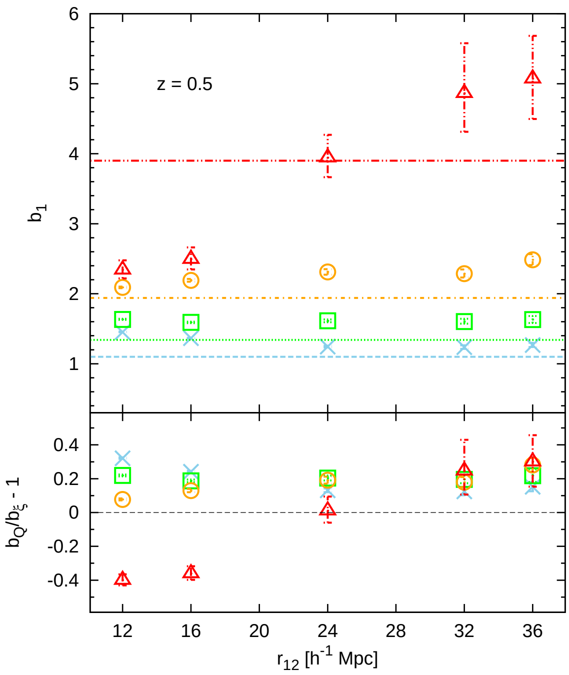

Figure 2.8: *Top*: linear bias parameter b_1 derived from the three-point correlation Q (b_Q, symbols) via equation (3.6) using triangles with $r_{13}/r_{12} = 2$ as a function of r_{12}. This is compared with b_1 derived from the ratio of dark matter and halo two-point correlations ξ (b_ξ, lines) from Fig. 2.4. Different colours denote results for the mass samples M0 to M3 (from bottom to top) with the same colour coding as in Fig. 2.4. *Bottom*: relative difference between b_Q and b_ξ. Results for redshift $z = 0.0$ are shown in the Appendix.

samples. Furthermore, differences between b_ξ and b_Q are expected due to non-local contributions to the bias function, as it has been shown in k-space by Chan et al. (2012). Performing the same analysis at redshift $z = 0.0$ gives very similar results, which are shown in Fig. A.3 of the Appendix A.1. We find in that case slightly larger deviations at small scales presumably due to a higher impact of non-linearities on the measurement. The overestimations at large scales are slightly smaller possibly as a result of smaller bias values at low redshift. We will show in Chapter 3 that deviations between b_ξ and b_Q decrease, when galaxy-matter-matter cross-correlations instead of galaxy-galaxy-galaxy auto-correlations are analysed. In the following we will focus on the results for $r_{12} = 24\ h^{-1}$ Mpc which is a compromise between having small errors and sufficiently large scales for linear bias estimation.

Despite the discrepancies between b_Q and b_ξ shown in Fig. 2.8 we will still be able to obtain a good approximation for the growth factor $D(z)$ if b_Q and b_ξ are related by the same multiplicative constant at different redshifts. This is because $D(z)$ only depends on the bias ratio, as shown in equation (2.15).

Bias ratio $\hat{b}$ from Q_g at different redshifts

A fundamental limitation for the growth factor measurement described in Section 2.3.3 is its dependence on the dark matter correlations, which cannot be directly observed. This problem is usually tackled by employing predictions for the dark matter correlations from N-body simulations or perturbation theory (see e.g. Verde et al., 2002; McBride et al., 2011b; Marín et al., 2013). Alternatively weak lensing signals can be used as a direct probe of the total matter field (Jullo et al., 2012; Simon et al., 2013). Both approaches can add uncertainties and systematic effects to the galaxy bias measurement and will therefore affect constrains of cosmological parameters derived from the growth factor.

We therefore introduce a new approach for measuring the growth factor based on the following consideration: in equation (2.15) we see that for measuring the growth factor $D(z)$ we only require knowledge about the ratio of the linear bias parameters at the redshifts z_0 and z, while the absolute bias values are irrelevant. With the three-point correlation function we can measure this ratio directly from the distribution of galaxies without knowing Q_{dm}. We can write equation (3.6) for the two redshifts z_0 and z and combine them via Q_{dm} under the assumption that Q_{dm} is independent of redshift, as shown in Fig. 2.6. We find

$$Q_g(z) = \frac{1}{\hat{b}_Q}[Q_g(z_0) + \hat{c}_Q], \tag{2.22}$$

where we have defined $\hat{b}_Q \equiv b_Q(z)/b_Q(z_0)$ and $\hat{c}_Q = [c_Q(z) - c_Q(z_0)]\,/b_Q(z_0)$. Equation (2.22) allows us to estimate the bias ratio $\hat{b}_Q$ from Q_g measurements at two different redshifts. The measurement of $\hat{b}$ can then be used in equation (2.15) to estimate $D(z)$ from the measured $D_g(z)$. The results will be shown later in Section 2.4.2.

2.3.6 Errors estimation and fitting

Since we use either one simulation at various comoving outputs ($z = 0$, $z = 0.5$, $z = 1.$ and $z = 1.5$), or one light cone, we estimate the errors of ξ, Q, S_3, C_{12}, b_ξ, b_Q, c_Q and D measurements by Jackknife resampling. The Jackknife samples of the complete comoving output are constructed from 64 cubical sub-volumes while in case of the light cone we use 100 angular regions (with equal volume at each redshift bin) in right ascension and declination on the sky. Following Norberg et al. (2009), we generate for any statistical quantity X a set of pseudo-independent measurements (X_j), from which we compute the standard deviation σ_X around the mean $\bar{X}$ (computed on the complete volume) as

$$\sigma_X = \sqrt{\frac{(n-1)}{n}\sum_{j=0}^{n}(X_j - \bar{X})^2}, \tag{2.23}$$

where n is the number of Jackknife samples.

We use the same fitting procedure for both bias estimations b_ξ, b_Q, which takes into account the covariance between ξ and Q measurements at different separations, opening angles and smoothing scales (r_{12}, α, R respectively). The covariance matrix C is computed from the deviation matrix A, which in turn is estimated by Jackknife resampling as well: a measurement in the j^{th} Jackknife sub-volume and for the i^{th} separation, angle or scale is written X_{ij}. Each element A_{ij} of the deviation matrix is calculated as $A_{ij} = X_{ij} - \bar{X}_i$. Again the mean $\bar{X}_i$ is the measurement on the complete volume. The covariance matrix can then be computed straightforwardly

$$C = \frac{n-1}{n}A^T A. \tag{2.24}$$

Deriving $\hat{b}_Q$ requires a two-parameter fit, due to the mixing of the bias coefficients (see equation (2.22)). The main problem arises from the fact that at a given redshift the errors of Q_g are correlated between the various angles. Furthermore the reduced three-point correlation can also be correlated between the two redshifts z_0 and z, where z_0 is the reference redshift. Based on equation (2.22), we define the variable

$$Z \equiv Q_g(z_0) - (\hat{b}Q_g(z) + \hat{c}), \tag{2.25}$$

and vary $\hat{b}$ and $\hat{c}$ in order to obtain $Z = 0$ for all angles α. In other words we want to measure the posterior probability distribution (hereafter referred to as likelihood $L(\hat{b}, \hat{c})$) of the two parameters $\hat{b}$ and $\hat{c}$ given that Z is expected to be $Null$. Assuming a multivariate normal distribution of Z, one can write the log-likelihood $\mathcal{L} \equiv -2\ln(L)$ as for measuring a given Z

$$\mathcal{L} = B + \ln(|C_Z|) + \chi^2, \tag{2.26}$$

where C_Z is the covariance matrix of the Z, B is a normalisation constant and $\chi^2 \equiv \sum_{i,j} Z_j C^{-1}_{Z,ij} Z_i$. Note that, if the covariance matrix does not depend on the parameters of the model, then the second term in expression (2.26) can be absorbed in the normalisation constant B. However, from definition (2.25) follows that C_Z explicitly depends on the fitting parameters $\hat{b}$ and $\hat{c}$. It can therefore be obtained from the covariance matrix of $Q_g(z_0)$, $Q_g(z_j)$ and from the cross-covariance of $Q_g(z_0)$ and $Q_g(z_j)$:

$$C_Z = C_X + \hat{b}^2 C_Y - \hat{b}\left(C_{XY} + C_{XY}^{\intercal}\right), \tag{2.27}$$

which explicitly shows the dependency of the covariance matrix C_Z on the fitting parameter $\hat{b}$. Note that C_X and C_Y are respectively the covariance matrix of $Q_g(z_0)$ and $Q_g(z_j)$ computed with equation (2.24). The cross-covariance matrix C_{XY} is defined as

$$C_{XY,ij} = \frac{n-1}{n}(X_j - \bar{X})(Y_i - \bar{Y}), \tag{2.28}$$

where n is the number of elements in both X and Y. In practice we shall neglect the correlation between redshift bins, so that $C_{XY} = C_{XY}^{\intercal} = 0$ in equation (2.27). Otherwise the inverse covariance matrix $C^{-1}_{XY,ij}$ had to be computed for each tested value of $\hat{b}$. The estimate of $\hat{b}$ and its error are obtained by marginalising over the $\hat{c}$ parameter via the posterior marginalised log-likelihood

$$\mathcal{L}(\hat{b}) = -2\ln\left\{\int L(\hat{b}, \hat{c})d\hat{c}\right\}.$$

2.4 Results

As we have pointed out in the Section 2.3 we use growth independent bias measurements from third-order statistics to break the growth-bias degeneracy that appears in growth measurements from two-point correlations. This approach is limited by the accuracy and the precision with which third-order statistics can measure galaxy bias. We study the differences between bias from second- and third-order correlations for different redshifts and halo mass ranges and present the results in Section 2.4.1. In Section 2.4.2 we show the resulting estimations for the linear growth measurements.

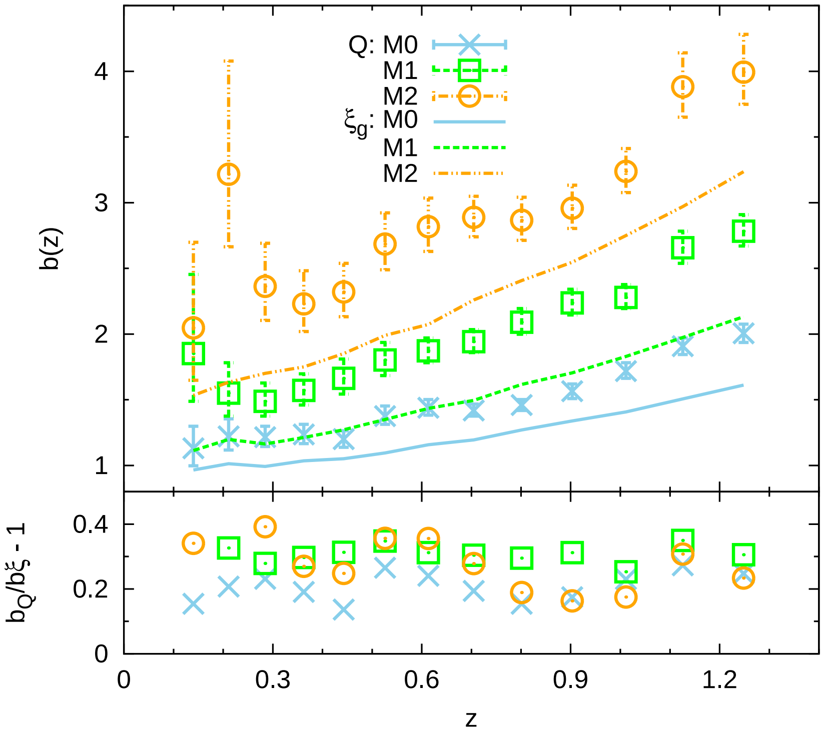

Figure 2.9: Linear bias as a function of the redshift estimated from b_Q in three mass bins of MICE-GC light cone compared to the linear bias estimated from the two-point auto correlation function b_ξ (coloured dashed lines). Blue crosses, green diamonds and orange triangles correspond respectively to M0, M1 and M2. The colour code is the same as for b_ξ in Fig. 2.4.

Alternatively to the direct approach of growth measurement described above, we have introduced a new method which does not require any modelling of third-order clustering of dark matter. It takes advantage of the fact that only the ratio of the bias parameters at two redshifts needs to be known to break the growth-bias degeneracy. This bias ratio can be directly measured from third-order statistics of the halo field (see Section 2.3.3-2.3.5). In Section 2.4.2 we compare growth factor measurements from our new method and the more common method of combining second- and third-order statistics with theoretical predictions (or simulations) for the dark matter field. In Section 2.4.2 we present growth rate measurements derived with and without third-order correlations of dark matter.

2.4.1 Bias comparison

We conduct bias measurements in a light cone, which is constructed from the MICE-GC simulation and includes redshift evolution of structures. The total volume probed by the light cone is about 15 h^{-3} Gpc^3 and we consider an octant of the sky (about 5000 deg^2). We study the deviation between the different bias estimations in five redshift bins between $0.4 < z < 1.42$ using the mass samples M0, M1 and M2. We do not present results for the highest mass sample M3 and for smaller redshifts, since the results are strongly scattered due to small numbers of haloes (see Fig. 2.1). However, this mass and redshift range will be analysed using the comoving outputs of the same simulation in Chapter 3.

For measuring the bias we use the same $(24, 48, \alpha)$ configurations for Q as in the comoving output. We estimate errors by Jackknife resampling of 100 angular regions of the light cone (see Section 2.3.6).

The results for the b_Q estimator are shown together with b_ξ in Fig. 2.9. They confirm that b_Q tends to overestimate the bias for the lower mass bins by about 30%. Moreover it shows that the ratio between b_ξ and b_Q is roughly a constant with respect to the redshift or mass bins. Since in our approach we aim at measuring the linear bias in order to extract information about the growth factor D of linear matter fluctuations, we focus on deriving a direct measurement of it in the following section.

2.4.2 Growth Measurements

In this subsection we present the growth factor D and the growth rate f, measured in the MICE-GC light cone via the equations (2.15) and (2.18) respectively. We obtained these measurements with the linear bias b, estimated with Q (equation (2.22)) at the redshift z and the reference redshift z_0. These growth measurements are compared to those from our new approach for measuring D and f via the bias ratio $\hat{b}(z) = b(z)/b(z_0)$. We can derive $\hat{b}(z)$ directly by comparing Q of galaxies (or haloes) at z and z_0 (equation (2.22)). This new approach allows us to measure the growth of dark matter using only the observable second- and third-order galaxy (or halo) correlations without the corresponding dark matter statistics (see Section 2.3.5).

Growth factor measured with Q_{dm}

Fig. 2.10 shows measurements of the growth factor D, derived from the mass samples M0, M1 and M2 in the MICE-GC light cone. Symbols denote results, which were derived by using the same mass bins at both redshifts, z_0 and z. Exploring the variation of our results for different choices of mass bins, we measure the growth factor from all combinations of mass bins. The median growth factor and the median error from all combinations are shown as grey shaded areas in the same figure. The left panel show results, derived by using the bias parameters b_Q, which were measured at each redshift separately from equation (3.6) and (2.11). This approach requires the knowledge or modelling of the dark matter Q_{dm}. Note that we normalised all measurements with respect to the highest redshift bin by setting $z_0 = 1.25$ in equation (2.15). This allows us to have a normalisation, which is performed as much as possible in the linear regime and with the lowest possible sampling variance. The measurements are compared to the theoretical prediction from equation (2.4), shown as dashed lines in the same figure. We also show that combining different halo populations (halo mass) at different redshifts we obtain results which are consistent with those derived by following the same halo population across the considered redshift range. To be independent of the normalisation we χ^2-fit the normalisation of the predictions to the median measurements from all mass sample combinations (i.e halo populations).

Our results in Fig. 2.10 show that the growth factor, measured with the bias from the third-order statistics, decreases with redshift, as expected from predictions for the linear growth factor. The good agreement between measured and predicted growth factor is remarkable since the bias estimation, on which the measurement is based on, shows a 30% over

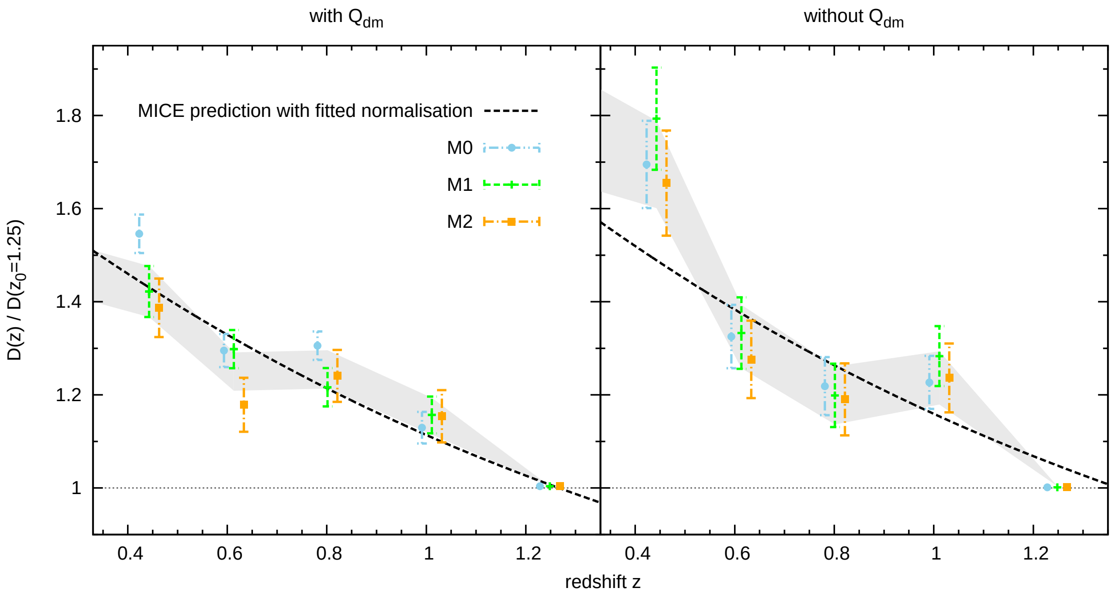

Figure 2.10: Growth factor measured from haloes in the mass samples M0, M1, M2 from the MICE-GC light cone. Measurements are normalised to be unity at the reference redshift $z_0 = 1.25$. Symbols show results derived by using the same mass bin at redshift z and z_0. Median results with median errors from combining all mass bins are shown as grey areas. Measurements are derived using the linear bias from $Q(24, 48, \alpha)$. The dashed line is the theoretical prediction from linear perturbation theory in equation (2.4) for the MICE-GC cosmology. Its normalisation was chosen to minimise deviations from the median measurements and is therefore different in each panel. Results shown in the left panel are based on separate measurements of bias at each redshift, by comparing Q in halo samples with the corresponding dark matter measurements in the same redshift bin. The results in the right panel are based on ratio measurements of bias at two redshift, by comparing Q at different redshifts (no dark matter is used).

estimation (see left panel of Fig. 2.9). We explain this finding by a cancellation in the bias ratio $\hat{b}_Q$ of the multiplicative factor by which b_Q is shifted away from b_ξ. This cancellation also happens for the median results from all mass bin combination, since this multiplicative factor is similar for all masses and redshifts. Fluctuations of the growth factor measurements at high redshifts probably result from fluctuations in the bias measurements. We expect an additional uncertainty in the growth measurement from the resolution of the density grid, used for computing b_Q (see Appendix A.1). From equation (2.15) we estimate that the 5% resolution error in b_Q propagates into D as an error of below 10%.

Growth factor measured without Q_{dm}

In the bottom panel of Fig. 2.10 we show the growth factor measurements based on the new approach, which uses the bias ratios $\hat{b}_Q$ derived from equation (2.22). This means that we compare the statistical properties only of the halo density field at different redshifts, without requiring knowledge about the dark matter quantity Q_{dm}. As in the top panel, the symbols denote measurements using the same mass bins at both redshifts, while median growth measurements and errors from all mass bin combinations are shown as grey shaded areas.

We find that the estimation a slightly larger deviation from the linear theory compared to the results from the separate bias measurement, shown in the upper panel. This discrepancy tends to be larger as the redshift is decreasing, possibly due to three effects: i) noise in the measurements of third-order galaxy (or halo) correlations Q_g enters twice, ii) non-linearities in the dark matter field become stronger at small redshift, iii) sampling variance does not cancel out since the two halo correlations, on which the measurement is based on, come from different redshifts.

In practice that last point iii) will also affect the first method, which uses Q_{dm} to get the absolute bias at each redshift. In our analysis, sampling variance cancels out between redshifts because we use Q_{dm} measured in the same simulation where we measure the corresponding halo values Q_g. In the analysis of a real survey this cancelation will not occur since one needs to use models for Q_{dm}.

We demonstrate this effect in Fig. 2.11. To study how the fitting range affects the growth estimate we now restricted to opening angles between $0 < \alpha < 60$ degree, excluding large triangles as we discuss later. The top panel shows the growth, measured with the separate bias estimates of b_Q, as shown on the top panel of Fig. 2.10. In the central panel we show the more realistic growth measurements based on the same approach, but instead of using the dark matter measurements in the same redshift as the halo measurements, we always use the dark matter results of Q from the highest redshifts bin $z_0 = 1.25$ (which is in good agreement with the results from the comoving output, as it contains more volume than the other redshift bins). Here, the sampling variance does not cancel, since the dark matter and halo correlations are measured at different redshifts. This results in a larger scatter in the central panel than in the top panel. Quantifying this scatter with respect to the predictions as $\sigma = \sqrt{\langle (D - D_{PT})^2 \rangle}$ confirms the visual impression (values are shown in Fig. 2.11). The latter approach in the central panel corresponds more closely to how the first method would be applied in a real survey: i.e. assuming a cosmology to run the dark matter model and running a simulation for that cosmology (sampling variance will not cancel as the simulation has different seeds than the real Universe). These latter growth measurements are distributed

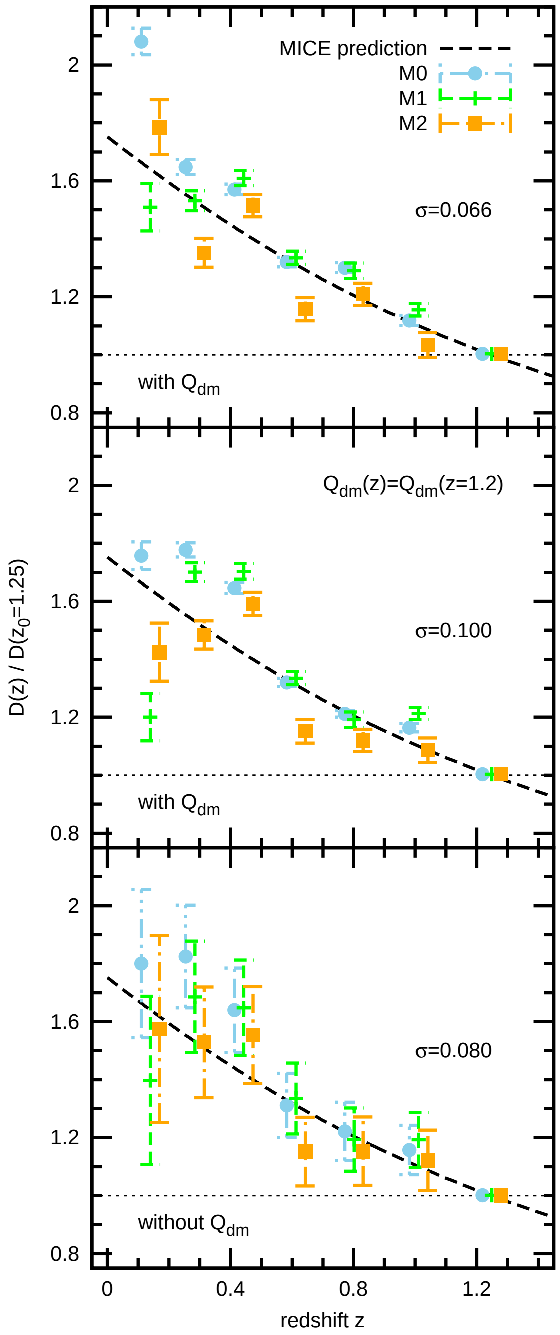

Figure 2.11: Impact of sampling variance on the linear growth factor, measured with $Q(24, 48, 0 < \alpha < 60)$ and normalised at $z_0 = 1.25$. The angular range excludes widely opened triangles which are expected to be more strongly affected by sampling variance. The top panel shows $D(z)$ estimated from separate measurements of bias at each redshift, by comparing Q in haloes with the corresponding dark matter measurements in the same redshift bin. The central panel shows the same measurements when we use the dark matter measurements from z_0.0 for all redshifts instead. In this case sampling variance between halo and dark matter fluctuations does not canceled out, as it does in the top panel. Results in the bottom panel are based on ratio measurements of bias at two redshift, by comparing Q at different redshifts (no dark matter is used). Quantifying the scatter we show the standard deviation $\sigma = \sqrt{\langle (D - D_{PT})^2 \rangle}$, where D_{PT} is the predicted growth factor (shown as black dashed line) and $\langle \ldots \rangle$ denotes the mean over all redshifts and mass samples.

in a similar way around the theoretical predictions as the results derived from the ratio bias approach using Q and ξ at the redshift z and redshift $z = 1.25$, shown in the bottom panel of Fig. 2.11. This demonstrates that most of the difference between the top and bottom panels of Fig. 2.11 comes from the artificial sampling variance cancelation in the top panel.

The restriction of the opening angles to $0 < \alpha < 60$ excludes widely opened triangles and is a possibility to decrease the impact of sampling variance on the measurements. However, comparing the growth in the bottom panel of Fig. 2.11 to the corresponding measurements for $0 < \alpha < 180$ in the bottom left panel of Fig. 2.10 we find no significant improvement of the growth measurements from restriction of the opening angles possibly due to larger errors. The latter result from the smaller number of triangles.

Growth rate measured with and without Q_{dm}

We derived the growth rate f from the measured growth factor D via equation (2.18). The product $f\sigma_8$ can also be probed by redshift space distortions, while our measurements represent an additional, independent approach to the growth rate. Especially at higher redshifts, where growth rate measurements via redshift space distortions are difficult to obtain, such additional information is valuable. Besides the relation to redshift space distortions another advantage of the growth rate with respect to the growth factor is that it is independent of the normalisation. The latter cancels out in the ratio of the growth factors, which appears in equation (2.18).

However, measuring f via D at a given redshift is not straightforward, since it depends on measurements at two different redshift. We derived f at the redshift bin z_i from growth factor measurements at z_{i+1} and z_{i-1}. This approach is motivated by the fact that the redshift bins have equal width in comoving space. Constrains of cosmological parameters with such measurements would require a more careful treatment of the assigned redshift. The employed growth factors are the median results, derived via Q from all mass combinations, which are shown as grey areas in the left panels of Fig. 2.10.

The results for f from D, measured with and without Q_{dm}, are shown in Fig. 2.12. In both cases the measurements are strongly scattered around the theoretical predictions for the MICE simulation, while the scatter is stronger for results derived without Q_{dm}. The increased scatter at lower redshifts probably results from the smaller volume of the light cone, which causes stronger fluctuations in D (see Fig. 2.10). We also expect an uncertainty from the resolution of the density grid, used for computing b_Q (see Appendix A.1). From the equations (2.15) and (2.18) we estimate that the 5% resolution error in b_Q propagates into f as an error of below 10%. Note that the errors that we find at high redshifts are comparable, or slightly better, than current errors from redshift space distortions (RSD) in the anisotropic two-point correlation function. Note that measurements from RSD directly constrain $f \times D$ and not f. Nevertheless, under some assumptions we can also infer f from RSD and the typical errors found on SDSS, BOSS and WiggleZ are around $15 - 20\%$ (Cabré and Gaztañaga, 2009; Blake et al., 2011; Tojeiro et al., 2012), which are comparable to the ones we find here.

The MICE prediction is computed from equation (2.6) with $\Omega_m = 0.25$ and $\gamma = 0.55$. To compare the scatter in our measurements with variations of the growth rate for different cosmologies we also show predictions for $\gamma = 0.35$ and 0.75. We find that our errors in the measurements are larger than the expected variations in the growth rate due to cosmology. It would be worthwhile to conduct a similar comparison using larger mass bins and combining

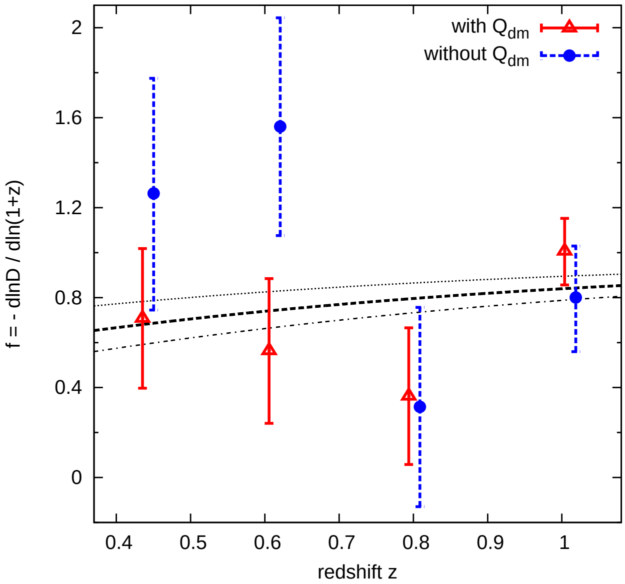

Figure 2.12: Growth rate f, estimated from the median measurements of D from all mass bin combinations (shown as grey areas in the left panels of Fig. 2.10) via equation (2.18). MICE predictions, derived from equation (2.6) with $\Omega_m = 0.25$ and $\gamma = 0.55$ are shown as thick dashed lines. By changing the values of γ to 0.35 and 0.75, we derive the predictions for different cosmologies, shown as dotted and dash-dotted lines, respectively. The large errors could be decreased by measuring the bias b_Q using a combination of different triangle configurations.

measurements from different scales and configurations of Q, to decrease the error, but the goal here is just to demonstrate the possibility of such measurements and the advantage of using it.

2.5 Summary & Discussion

The amplitude of the transverse (or projected) two-point correlation of matter density fluctuations allows us to measure the growth factor D, which can be used as a verification tool for cosmological models. Galaxies (in our study represented by haloes) are biased tracers of the full matter field as their two-point correlation at large scales is shifted by a constant bias factor b with respect to the matter two-point correlation. This bias factor is fully degenerate with D. The reduced matter and galaxy third-order statistics are independent of D, while the galaxy versions are sensitive to b. Combining second- and third-order statistics could therefore enable us to break the growth-bias degeneracy, if the difference between the effective linear bias b_1 probed by both statistics is smaller than the errors required for the growth measurements.

In this thesis we have tested these assumptions and verified how well we can recover the true growth of the new MICE-GC ΛCDM simulation (Fosalba et al., 2013a; Crocce et al., 2013; Fosalba et al., 2013b) with them. We also further validate the MICE-GC simulation by comparing the linear growth with measurements from two-point matter correlation (Fig.2.2 and 2.3) and three-point correlations of the matter field to non-linear perturbation theory predictions (Fig. 2.6).

The main goal of this Chapter is to compare bias (and the resulting growth) from the reduced three-point auto correlation Q. We estimated this quantity from density fields of matter in the MICE-GC simulation and those of haloes in different mass samples, expanding previous studies significantly to a wider range of masses (between 5.8×10^{12} and 5×10^{14} $h^{-1}\mathrm{M}_{\odot}$) and redshifts (between 0 and 1.2) with values of the linear bias b_1 between 0.9 and 4.

Our results in Fig. 2.9 show that the linear bias from Q, b_Q, systematically over estimates the linear bias from the two-point correlation, b_ξ, by roughly $20 - 30\%$ at all mass and redshift ranges. Understanding the differences between b_ξ and b_Q is crucial for constraining cosmological models with observed three-point correlations. We will therefore deepen our analysis in the following chapters by studying bias from halo-matter-matter cross-statistics (Chapter 3) and predictions from the peak-background split model (Chapter 4) to disentangle between non-linear and non-local effects on the different estimators.

For measuring the growth factor D we have introduced a new method. This new method uses the bias ratio $\hat{b}(z) = b(z)/b(z_0)$, derived directly from halo density fluctuations with reduced third-order statistics. Its main advantage with respect to the approach of measuring $b(z)$ and $b(z_0)$ separately is that it does not require the modelling of (third-order) dark matter statistics. Instead, it works with the hypothesis that

1. the reduced dark matter three-point statistics is independent of redshift z
2. the bias ratio $\hat{b}(z) = b(z)/b(z_0)$ from two- and three-point statistics is equal.

The first assumption was tested in this study numerically, while the validity of the second follows directly from our bias comparison.

In general the comparison between D from perturbation theory with measurements from our new method and the standard approach reveals a good agreement. We explain this result by a cancellation of the multiplicative factor by which b_Q is shifted away from b_ξ in the bias ratio $\hat{b}_Q$.

Our analysis shows that the new way to measure the growth factor from bias ratios is competitive with the method based on two separate bias measurements. While having larger errors the new method has the advantage of requiring much weaker assumptions on dark matter correlations than the standard method and therefore provides an almost model independent way to probe the growth factor of dark matter fluctuations in the Universe.

We demonstrated that besides the growth factor, D, the growth rate of matter, f, can also be directly measured from the galaxy (or halo) density fields with bias ratios from third-order statistics. This provides an alternative method to derive the growth rate, which is usually obtained from velocity distortions probed by the anisotropy of the two-point correlation function (RSD). The typical errors found on SDSS, BOSS and WiggleZ using RSD are around $15 - 20\%$ (Cabré and Gaztañaga, 2009; Blake et al., 2011; Tojeiro et al., 2012), which are comparable to the ones we find in Fig. 2.12 when considering the high redshift bins (20%).

Our analysis is performed in real space to have clean conditions for comparing different bias and growth estimates. This is a good approximation for the reduced higher-order correlations on the large scales considered in this study, as measurements in redshifts space always seem to be within one sigma error of the corresponding real space result (see Fig. 2.13). Note how

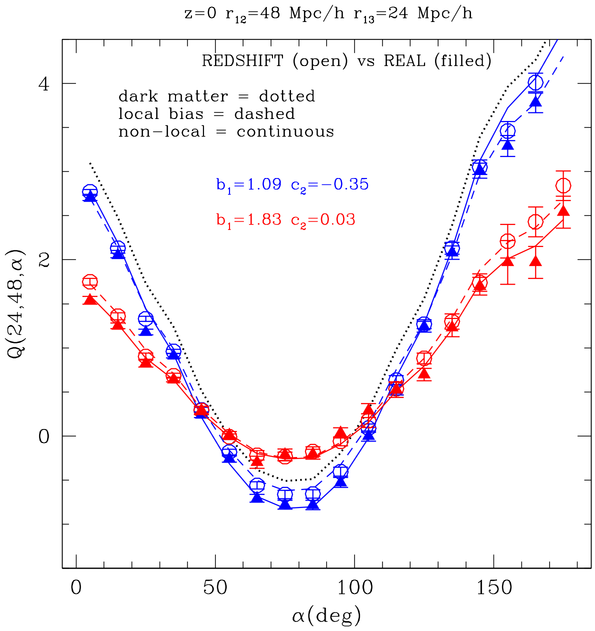

Figure 2.13: Q for dark matter (dotted) and for halo samples (symbols) with two different mass thresholds: $b_1 = b_\xi \simeq 1.09$ (blue) and $b_1 = b_\xi \simeq 1.83$ (red). We compare results in real space (filled triangles) and redshift space (open circles), which agree within the errors on these large scales ($r_{12} = r_{13}/2 = 24\ h^{-1}$Mpc at z=0). Predictions are shown for both: the local bias model (dashed lines) and non-local bias model (continuous). In both cases we have fixed $b_1 = b_\xi$ and fit for c_2.

the small, but systematic, distortions in redshifts space seem to agree even better with the local bias model than in real space on the largest scales.

Applying the method described above to obtain accurate bias and growth measurements from observations will require additional treatment of redshifts space distortions or projection effects. Two possible paths could be followed. In a three dimensional analysis redshifts space distortions need to be modelled (e.g. Gaztañaga and Scoccimarro, 2005). The projected three-point correlation can also be studied separated by in redshift bins (Frieman and Gaztañaga, 1999; Buchalter et al., 2000; Zheng, 2004). Both ways will result in larger errors, but we do not expect this to be a limitation because our error budget is totally dominated by the uncertainty in the bias. A more detailed study of this issue is beyond the scope of this thesis and will be presented elsewhere. Mock observations, like the galaxy MICE catalogues (see Crocce et al., 2013; Carretero et al., 2015) should be used to test the validity of such growth measurements under more realistic conditions.

Chapter 3

Non-local bias in three-point halo correlations

Abstract

In this Chapter we pursue our study of halo clustering bias with second- and third-order statistics of halo and matter density fields in the MICE Grand Challenge simulation. Combining three-point auto- and cross-correlations we find, for the first time in configuration space, evidence for the presence of non-local contributions to the bias function. These contributions are consistent with predicted second-order non-local effects on the bias functions originating from the dark matter tidal field. Samples of massive haloes show indications of bias (local or non-local) beyond second order. Ignoring non-local bias causes $20-30\%$ and $5-10\%$ overestimation of the linear bias from three-point auto- and cross-correlations respectively. We present a third-order bias estimator which is not affected by second-order non-local contributions. It consists of a combination of three-point auto- and cross-correlation and delivers accurate bias estimations of the linear bias. Ignoring non-local bias causes higher values of the second-order bias from three-point correlations. Our results demonstrate that third-order statistics can be employed for breaking the growth-bias degeneracy.

3.1 Introduction

With the increasing amount of data coming from current and future large-scale galaxy surveys, errors on the observed statistical properties of the spatial galaxy distribution are rapidly decreasing. This high level of precision requires at least the same level of accuracy in the modelling of the corresponding observables. An important observable is the growth of large-scale density fluctuations with time, which is sensitive to the universal matter density, the expansion of space as well as to the gravitational interaction of matter at large scales. Measurements of the this growth therefore provide constrains on cosmological parameters (e.g. Ross et al., 2007; Cabré and Gaztañaga, 2009; Song and Percival, 2009; Samushia et al., 2012; Reid et al., 2012; de la Torre et al., 2013), possible deviations from General Relativity (Gaztañaga and Lobo, 2001; Lue et al., 2004) or on alternative phenomenological description for the accelerated expansion, such as the effective field theory (Steigerwald et al., 2014; Piazza et al., 2014). Growth measurements can be undertaken by comparing the second-order correlations ξ of galaxy distributions at different redshifts. A critical aspect of this approach is the bias between the correlations of galaxies and those of the full matter density field, as we pointed out in Chapter 3. This bias can either be predicted with the peak-background split model (e.g. Bardeen et al., 1986; Cole and Kaiser, 1989; Sheth and Tormen, 1999, see also Chapter 4), or directly determined from observations using weak lensing observables, redshift space distortions or reduced third-order correlations at large scales. The latter method relies on the fact that such third-order correlations are independent of the growth at large scales, but sensitive to the bias. Third-order galaxy correlations therefore have the potential to tighten constrains on cosmological models from observations (e.g. Marín, 2011; Marín et al., 2013). However, how useful third-order correlation are for this purpose depends on the accuracy and the precision of the bias estimations they deliver (e.g. Wu et al., 2010; Eriksen and Gaztañaga, 2015).

Associating galaxies with dark matter we investigated in Chapter 2 growth measurements based on bias estimations from third-order halo auto-correlations (halo-halo-halo) using the MICE-GC simulation.

We found that linear bias estimations from the three-point auto correlations over-estimate the true linear bias (probed by two-point correlations at large scales) by 20-30%. Understanding these discrepancies between different bias estimators is crucial for constraining cosmological models with observed third-order galaxy statistics. In the light of previous studies (Manera and Gaztañaga, 2011; Pollack et al., 2012; Chan et al., 2012; Baldauf et al., 2012) we pointed out that such discrepancies might originate from non-linear and/or non-local effects on the different bias estimators. Hence, in this chapter, we focus on the analysis of three-point cross-correlations (halo-matter-matter) to decrease the impact of non-linearities on our bias estimators in order to investigate the extension of the local bias model to a possible non-local component.

3.1.1 the non-local bias model

The bias expansion, given by equation (2.1), assumes a local relation between the density contrast of matter and haloes. Inaccuracies of this deterministic relation might arise from tidal forces in the matter field, leading to a non-local contribution in the bias relation. At second order the non-local bias function can be expressed as

$$\delta_h(\mathbf{x}) = b_1 \left\{ \delta_m(\mathbf{x}) + \frac{c_2}{2}(\delta_m^2(\mathbf{x}) - \langle \delta_m^2 \rangle) + \frac{\gamma_2}{b_1}\mathcal{G}_2(\mathbf{x}) \right\}, \tag{3.1}$$

where γ_2 represents the non-local bias parameter (Chan et al., 2012; Baldauf et al., 2012). This non-local component depends on the divergence θ_v of the normalised velocity field $(\mathbf{v}/\mathcal{H}/f)$

$$\mathcal{G}_2(\mathbf{x}) = -\int \beta_{12}\theta_v(\mathbf{q}_1)\theta_v(\mathbf{q}_2)\hat{W}[q_{12}R]e^{i\mathbf{q}_{12}\cdot\mathbf{x}}d^3\mathbf{q}_1 d^3\mathbf{q}_2, \tag{3.2}$$

where $\beta_{12} \equiv 1 - \left(\frac{\mathbf{q}_1\cdot\mathbf{q}_2}{q_1 q_2}\right)^2$ represents the mode-coupling between density oscillations with wave vectors $\mathbf{q}$, which describe tidal forces. $W[q_{12}R]$ is the Fourier transform of a spherical Top-hat window with radius R. In order to be consistent with the definition of second-order bias parameter, we express to the non-local component of the biasing relation (3.1) using the quantity $g_2 \equiv 2\frac{\gamma_2}{b_1}$.

Evidence for significant contributions of such a non-local component to bias function has been reported in Fourier space for different simulations (Chan et al., 2012; Baldauf et al., 2012). However, it remains unclear how strongly these non-local contributions affect the bias and consequently third-order statistics of large-scale halo distributions in configuration space. We address this latter question in this Chapter and suggest possibilities to employ third-order statistics for accurate bias measurements, independently of non-local bias.

This chapter is organised as follows. In Section 3.2 we present the employed bias estimators. In Section 3.3 we discuss and comment our results. Finally we summarise our work and draw conclusions in Section 3.4.

3.2 Bias from three-point auto- and cross-correlation

In this section we study various bias estimators from second- and third-order clustering statistics of haloes and matter, in order to quantify and understand differences between these estimations. Such an understanding is crucial for using third-order statistics in order to break the degeneracy between the linear galaxy (or halo) bias and the linear growth of matter fluctuations, as we discussed in Chapter 2.

In this previous chapter we found that the linear bias from the reduced three-point correlation in configuration space tends to overestimate the linear bias from the two-point correlation, even when the analysis is performed at very large scales ($> 30\ h^{-1}$Mpc). Similar findings have been reported in the literature (e.g. Manera and Gaztañaga, 2011). Such deviations can be expected from non-local contributions to the bias function (Chan et al., 2012; Baldauf et al., 2012). Furthermore, non-linear terms in the perturbative expansion of correlations functions, which are usually neglected in the analysis of clustering measurements, can contribute to the deviations between the different bias estimators (Pollack et al., 2012). The goal of this section is to investigate the effect produced on various bias estimators of non-linear and non-local contributions to the biasing function.

The three-point halo-matter matter cross-correlation can be defined similar to the three-point auto correlation (equation(2.19)) as

$$\zeta^{hmm}(r_{12}, r_{13}, r_{23}) \equiv \langle \delta_h(\mathbf{r}_1)\delta_m(\mathbf{r}_2)\delta_m(\mathbf{r}_3)\rangle, \tag{3.3}$$

where the vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ form a closed triangle which can be parametrised in terms of the size of its three legs $r_{ij} \equiv |\mathbf{r_j} - \mathbf{r_i}|$ or in terms of the two legs $\mathbf{r}_{12}$, $\mathbf{r}_{13}$ and the angle $\alpha_{23} = \mathrm{acos}(\hat{\mathbf{r}}_{12} \cdot \hat{\mathbf{r}}_{13})$ between them. Analogously to equation (3.3) one can define ζ^{mhm} and ζ^{mmh}. For defining a growth independent reduced three-point cross-correlation we compared to the hierarchical three-point cross-correlation (Fry, 1984a)

$$\zeta_H^{hm} \equiv \xi_{12}^{hm}\xi_{13}^{hm} + \xi_{12}^{mh}\xi_{23}^{hm} + \xi_{13}^{mh}\xi_{23}^{mh}. \tag{3.4}$$

Note that here ξ_{ij}^{hm} referrers to the two-point cross-correlation between haloes at position $\mathbf{r_i}$ and matter at position $\mathbf{r_j}$, which is called $\xi^{\times}$ in the remainder of this thesis. Combining equation (2.19) and (3.4) one can define the symmetric reduced three-point cross-correlation function (Pollack et al., 2012)

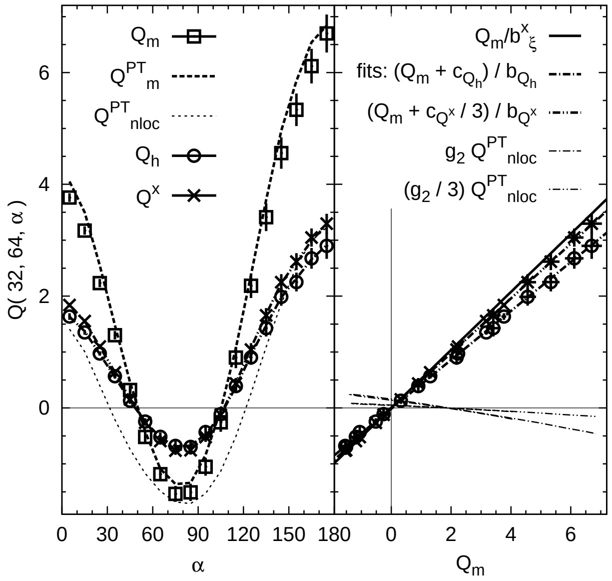

Figure 3.1: *Left*: three-point auto- and cross-correlation for matter and haloes (Q_h, $Q^\times$ and Q_m) in the mass sample M2 at redshift $z = 0.5$, measured using triangles with fixed legs of 32 and 64 h^{-1}Mpc for different opening angles α (circles, crosses and squares respectively). Fits based on the local bias model (i.e. $g_2 = 0$) to Q_h and $Q^\times$ from equation (3.6) and (3.7) are shown as thick dashed-dotted and dash-double dotted lines. Predictions for the matter three-point correlation and the non-local component from perturbation theory (Q_m^{PT} and Q_{nloc}^{PT}) are shown as thick dashed and thin dash-dotted lines respectively. *Right*: the Q_h and $Q^\times$ versus Q_m relation, used for deriving the linear and quadratic parameters b_1 and c_2 in the local bias model. Thin dashed-dotted and dash-double dotted lines show the non-local contributions to Q_h and $Q^\times$ respectively, using g_2 from ΔQ_{cg}, equation (3.12).

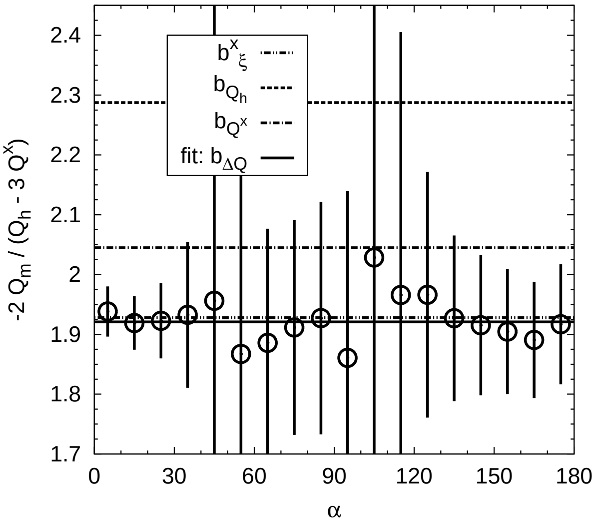

Figure 3.2: Linear bias of the halo mass sample M2 at the redshift $z = 0.5$, measured independently of second-order contributions (local and non-local) to the three-point correlation via equation (3.11) by combining Q_h, $Q^{\times}$ and Q_m (open circles). The fit to the measurements is shown as solid line. The linear bias measurements, derived from ξ, Q_h and $Q^{\times}$ within the local bias model (via equation (2.14), (3.6) and (3.7) respectively with $g_2 = 0$) are shown as dashed, dash-dotted and dash-double dotted lines respectively.

$$Q_h^{\times} \equiv \frac{1}{3}\frac{\zeta^{hmm} + \zeta^{mhm} + \zeta^{mmh}}{\zeta_H^{hm}}. \tag{3.5}$$

The reduced three-point auto-correlations for matter and halo density fields are defined analogously as $Q_m \equiv \zeta^{mmm}/\zeta_H^{mm}$ and $Q_h \equiv \zeta^{hhh}/\zeta_H^{hh}$ (see equation (2.20)). $Q_h^{\times}$, Q_h and Q_m quantify any departure from the hierarchical ansatz (Fry, 1984b). As in Chapter 2 we will refer to the reduced three-point correlation as the three-point correlation in the following.

Inserting the non-local quadratic bias model (equation (3.1)) into the definition of the three-point correlation for haloes yields, via a second-order perturbative expansion, in the limit of small density fluctuations and large triangles

$$Q_h = \frac{1}{b_1}\{Q_m + [c_2 + g_2 Q_{nloc}]\}, \tag{3.6}$$

which can be generalised to the case of three-point cross-correlation,

$$Q_h^{\times} = \frac{1}{b_1}\{Q_m + \frac{1}{3}[c_2 + g_2 Q_{nloc}]\} \tag{3.7}$$

These expressions differ significantly from the ones obtained from the local bias model, as they include the non-local contribution to the three-point halo correlation Q_{nloc}, which we present in more detail below. The expression for the local model, which we assumed in Chapter 2 corresponds to a vanishing non-local bias parameter, i.e. $g_2 = 0$. Since Q_{nloc} is a function of the opening angle $\alpha \equiv \alpha_{23}$, the three-point halo auto- and halo-matter cross-correlations are therefore no longer linearly related to the matter three-point correlation. This α dependence

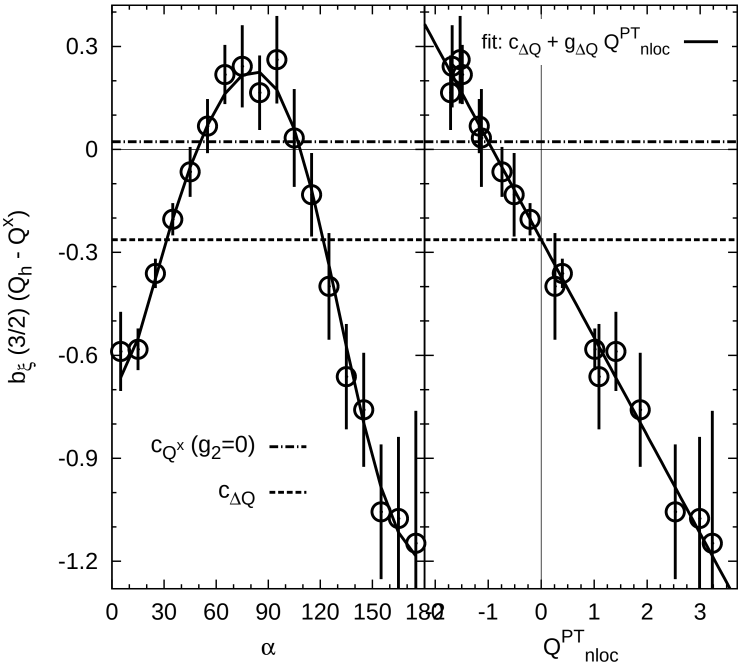

Figure 3.3: *Left:* The difference between the three-point auto- and cross-correlation Q_h and $Q^{\times}$ (ΔQ_{cg}), multiplied with the linear bias from the two-point correlation (open circles) as a function of triangle opening angle α for M2 at $z = 0.5$. According to equation (3.12) this expression is equivalent to $c_2 + g_2 Q_{nloc}$. If the bias function is quadratic and local, the measurements should correspond to the quadratic bias parameter c_2, as measured from the halo-matter cross-correlation via equation (3.7) with $g_2 = 0$ (dash-dotted line). Fits from equation (3.12), based on the PT prediction for the non-local component, Q_{nloc}, are shown as thick solid line. The quadratic bias parameter, derived from this fit (ΔQ_{cg}) is shown as dashed line. The three-point correlations were calculated from triangles with fixed leg of $r_{12} = r_{13}/2 = 32\ h^{-1}$Mpc. *Right:* same as left panel, but showing the measurements versus the Q_{nloc} prediction for each opening angle.

arises from the fact that Q_{nloc} originates from tidal forces, which modify the shape of matter fluctuations.

We can predict Q_{nloc} from the power spectrum, assuming that the perturbations of the density field δ_m are small (i.e. the divergence of the velocity field θ_v in equation (3.2) is linear and therefore equal to δ_m. It also requires to assume that the legs of the considered triangles are large compared to the smoothing radius R (large separation limit, i.e. in equation (3.2), $W(Rq_{12}) \simeq W(Rq_1)W(Rq_2)$). For such conditions the perturbation theory (hereafter also referred to as PT) offers the possibility to express the non-local three-point correlation Q_{nloc} in terms of

$$\Gamma_{123} \equiv \left[\xi(r_{12}) + 3\frac{\phi'(r_{12})}{r_{12}}\right]\left[\xi(r_{13}) + 3\frac{\phi'(r_{13})}{r_{13}}\right] L_2(\cos\alpha_{12}), \tag{3.8}$$

where

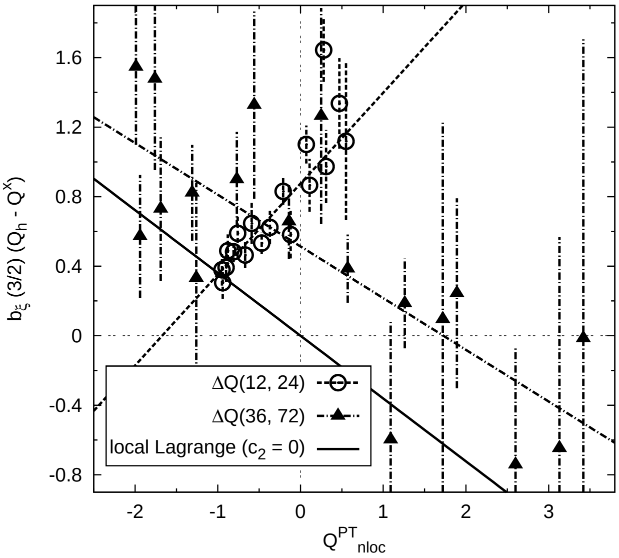

Figure 3.4: Same as right panel of Fig. 3.3, but for the halo mass sample M3 at the redshift $z = 0.0$. Results are shown for triangles with fixed legs $r_{12} = r_{13}/2$ of 12 and 36 h^{-1}Mpc (open circles and filled triangles respectively). Dashed and dashed-dotted lines are the corresponding fits to equation (3.12), used to derive the quadratic and non-local bias c_2 and g_2. The black solid line corresponds to the local Lagrangian prediction for $c_2 + g_2 Q_{nloc}$ with $c_2 = 0$ and $g_2 = 2\gamma_2/b_1$, while $\gamma_2 = -(2/7)(b_1 - 1)$. The γ_2 values from these fits are shown in Fig. 3.6.

$$\phi(r) = \int d^3\mathbf{k} \frac{P(k)}{k^2} W^2(kR) \frac{\sin(kr)}{kr} \tag{3.9}$$

and $\phi'(r) \equiv \frac{\mathrm{d}\phi}{\mathrm{d}r}(r)$. One can show that

$$Q_{nloc}(r_{12}, r_{13}, \alpha) = \frac{2}{3} \left\{ \frac{\Gamma_{123} + \Gamma_{312} + \Gamma_{231}}{\zeta_H^{mm}} - 1 \right\}. \tag{3.10}$$

The angular dependence of the non-local component of the three-point halo auto- and cross-correlation functions is encoded in equation (3.8) via the second-order Legendre polynomial $L_2(\cos\alpha_{12})$. As shown by Barriga and Gaztañaga (2002) in their equation (8), at the tree-level and for large separations, the matter three-point correlation can be expressed in the same way as expression (3.10). That is with respect to circular permutations of a function $\hat{\Gamma}_{123}$, expressed as a monopole, a dipole and a quadrupole in $\cos(\alpha_{12})$ (similar Legendre expansion have been used in Fourier space by Schmittfull et al., 2015) As a result, a non-local component, such as the tidal field $\mathcal{G}_2$ (equation (3.2)), modifies the amplitude of the quadrupole of $\hat{\Gamma}_{123}$ by an amount proportional to the non-local bias g_2.

Moreover, by comparing equations (3.6) and (3.7) one can see that quadratic and non-local contributions to the halo three-point correlation affect the cross-correlations by a factor 1/3 less than the auto-correlation. The linear bias, in contrast, affects the auto- and cross-correlation equally. We will use this property to isolate the linear from the quadratic and non-local bias, as explained below.

In the following we study non-local contributions to the halo bias in the MICE-GC simulation. Ours measurements of the three-point correlation are based on the algorithm suggested in Barriga and Gaztañaga (2002) (see also Appendix A.1 for studies of numerical effects in this algorithm and the impact of covariance between angular bins on the bias measurements). Errors are derived from 64 cubical Jack-Knife samples. We first focus on the mass bin M2 at redshift $z = 0.5$ to present our methods for extracting the parameters b_1, c_2 and g_2 from three-point correlations, which were computed using triangles with fixed legs of $r_{12} = 32$ and $r_{13} = 64\ h^{-1}$Mpc. Afterwards we will present results for all mass samples and redshifts at various scales.

3.2.1 Local bias

Our first method for measuring bias from three-point correlations is based on the local bias model ($g_2 = 0$). The linear and quadratic bias parameters are computed from the equations (3.6) and (3.7) by fitting the b_1 and c_2 parameters which allows for mapping Q_m into Q_h and $Q_h^\times$, i.e. $Q_h = (Q_m + c_2)/b_1$ and $Q_h^\times = (Q_m + \frac{c_2}{3})/b_1$. This approach is equivalent to the bias measurements from Q in Chapter 2, where we also explain the fitting procedure. The two estimations of the doublet b_1 and c_2 are respectively called b_Q, c_Q and $b_Q{}^\times$, $c_Q^\times$. In Fig. 3.1 we show how well a linear relation, expected from the local bias model, describes the mapping between the matter three-point correlation and the three-point auto- and cross-correlation functions of haloes. However, we can see that the slope of the linear relation is different when considering auto- and cross-correlations, which indicates that the two methods deliver inconsistent results (see right panel of Fig. 3.1). As explained in Chan et al. (2012) this linear relation between matter and haloes might arise from a projection effect due to the fact that we neglect the non-local component $g_2 Q_{nloc}$. In Fig. 3.1 we show that, if the contribution of Q_{nloc} is small compared to Q_m (i.e. g_2 is small, see Fig. 3.6), then they can indeed be approximately related by a linear relation. The ignored non-local contribution to halo three-point correlations can therefore be absorbed by b_1 and c_2, without substantially decreasing the goodness of the Q_h and $Q^\times$ fits. The same effect has been shown in Fourier space by Baldauf et al. (2012) in their Fig. 1. This ignorance might leads to incorrect bias measurements, unless $g_2 = 0$ as we show below.

3.2.2 Non-local bias

Our second method for measuring bias from three-point correlations is a new a approach, which combines auto- and cross-correlations. These two statistics can be combined in two different ways which allow us to isolate the linear bias from quadratic and non-local contributions to the bias model. Both combinations take advantage of the fact that the linear bias, b_1, affects Q_h and $Q_h^\times$ equally, which is not the case for the quadratic and non-local contributions, c_2 and g_2. The linear bias can be obtained by combining the equations (3.6) and (3.7),

$$b_{\Delta Q} \equiv -2\frac{Q_m}{\Delta Q}, \tag{3.11}$$

where $\Delta Q \equiv Q_h - 3Q_h^\times$. The interesting property of this linear bias estimator is that it is independent from quadratic (local and non-local) contributions to the bias function. It can therefore be used to verify if such contributions are indeed the reason for deviations between

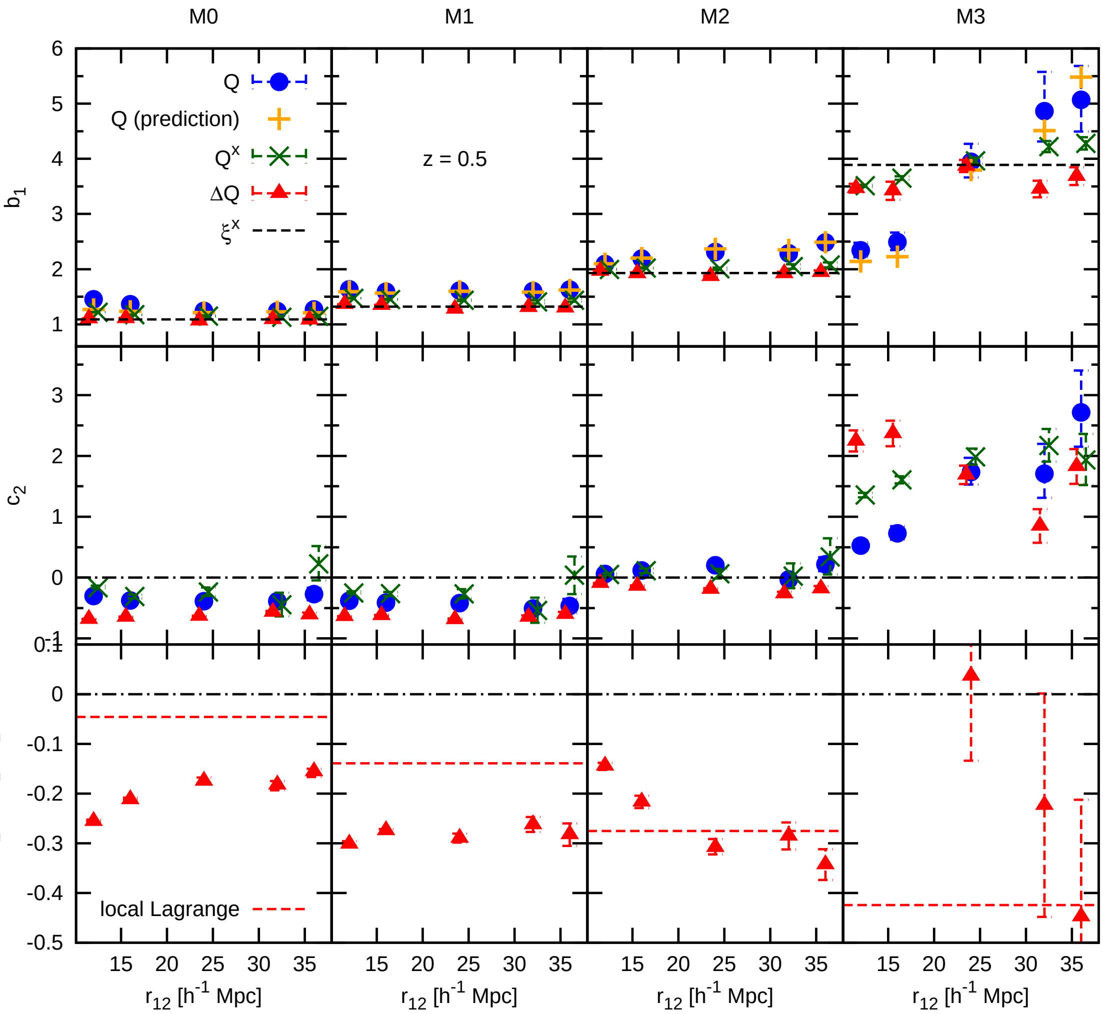

Figure 3.5: Scale dependence of bias measurements from three-point correlations. Measurements were derived using triangles with $r_{13}/r_{12} = 2$ configuration, while scale of the smaller triangle leg, r_{12}, is denoted on the x-axis (slightly shifted for clarity). Results are shown for the four mass samples M0-M3 (from left to right) at redshift $z = 0.5$. The linear and quadratic bias (top and central panel respectively), obtained from three-point auto- and cross-correlations using the local bias model ($g_2 = 0$) via the equations (3.6) and (3.7) are shown as blue circles and green crosses respectively. Predictions for effective linear bias measurements via Q while ignoring non-local bias (see Section 3.2.4) are shown as orange crosses. The linear, quadratic and non-local bias, measured using combinations of three-point auto- and cross-correlations (ΔQ, equation (3.11) and (3.12)) are shown as red triangles. The linear bias is compared with reference measurements from the two-point cross-correlation ($b_\xi^\times$), shown as black, dashed line. The non-local bias parameter $g_2 \equiv 2\gamma_2/b_1$ (bottom panel) is compared with predictions from the local Lagrangian model ($\gamma_2 = -(2/7)(b_1 - 1)$). (red dashed line). Error bars denote 1σ uncertainties.

linear bias estimations from two- and three-point correlations, as we speculated in Chapter 2. Note that the relevant quantities involved in equation (3.11) depend on the opening angle α, so does the estimator $b_{\Delta Q}$. Hence, our final $b_{\Delta Q}$ measurement is derived by fitting a constant to $b_{\Delta Q}(\alpha)$. The use of ΔQ has also the advantage that off-diagonal elements in the covariance matrix between different opening angles are smaller. This covariance is difficult to access and its Jack-Knife estimation can affect the bias estimation at the 5% level (Appendix A.1).

We show $b_{\Delta Q}$ for M2 at $z = 0.5$ in Fig. 3.2 at each angle probed by the three-point functions together with the corresponding fit. In the same figure we also display the estimations for the linear bias, derived from three-point auto- and cross-correlations (b_Q and $b_{Q^\times}$, obtained from equations (3.6) and (3.7), assuming the local bias model, i.e. $g_2 = 0$). As a reference, we also include the linear bias measurements from the two-point cross-correlation, $b_\xi^\times$, which we consider to be a reliable estimate of the true linear bias (see Section 2.3). The comparison in Fig. 3.2 reveals that the measurement and the fit of the hybrid bias estimator $b_{\Delta Q}$ are consistent with the reference b_ξ, while we see that the biases obtained from b_Q and $b_{Q^\times}$ are over estimating the linear bias. This result confirms our speculation in Chapter 2, that differences between b_ξ and b_Q are mainly due to a non-local term in the bias model.

One can notice that, as expected in case of non-local bias, the overestimation is larger in case of the auto-estimator (b_Q) compared to the cross-estimator ($b_{Q^\times}$). Note that Pollack et al. (2012) found an opposite trend, analysing a different simulation in Fourier space with a different mass resolution and cosmology. Their linear bias measurements from the bispectrum are closer to peak-background split predictions than measurements from three-point cross-correlations, while the predictions might be lower than the true linear bias (see e.g. Chapter 4 and references therein).

3.2.3 Linear and quadratic terms

In order to further verify the presence of non-local contributions to the bias function we separate Q_{nloc} from Q_m by subtracting $Q_h^\times$ from Q_h and define

$$\Delta Q_{cg} \equiv Q_h - Q_h^\times = \frac{2}{3}\frac{1}{b_1}\left[c_{\Delta Q} + g_{\Delta Q} Q_{nloc}\right], \tag{3.12}$$

where $c_{\Delta Q}$ and $g_{\Delta Q}$ are the estimators of c_2 and g_2 in equations (3.6) and (3.7). If the bias function is quadratic in δ_m and local then the non-local bias parameter g_2 is zero. Hence,

$$c_{\Delta Q} = b_1 \Delta Q_{cg}(3/2) \tag{3.13}$$

should correspond to c_2, independently of the opening angle α. In the left panel of Fig. 3.3, we show $c_{\Delta Q}$, together with c_2, estimated from Q_h and $Q^\times$ (c_Q and $c_{Q^\times}$ respectively) from the local bias model (equations (3.6) and (3.7), for $g_2 = 0$). In the same figure we also show c_2 derived from fitting ΔQ_{cg} taking into account a possible non-local bias g_2. The first important point is that the measured $c_{\Delta Q}$ shows a very clear angular dependence, with a maximum at around 80 degrees and a positive curvature. The local quadratic model therefore clearly fails in describing the impact of bias on three-point correlations. The right panel of Fig. 3.3 displays the measurements of ΔQ_{cg} with respect to the tidal, non-local component of the three-point function Q_{nloc}, as predicted by equation (3.10). It shows that measurements are compatible with a linear relation between ΔQ_{cg} and Q_{nloc}, which is expected in the second order non-local bias model. It therefore confirms the presence of non-local bias due to tidal disruption ($\mathcal{G}_2$) in the matter field. For such a linear relation, the value of ΔQ_{cg} corresponds to the second-order bias $c_{\Delta Q}$ when $Q_{nloc} = 0$, while its slope provides a direct insight to the non-local bias g_2. Finally, by comparing the various estimates of the second-order local bias c_2, one can see that, when assuming a local bias, the results differ significantly from the ones obtained when taking into account a possible non-local component into the biasing relation. However, the estimation of c_{Δ_Q} and g_2 might be affected by shortcomings of the PT predictions for Q_{nloc}. To verify how well the PT predictions describe the measurements at different scales we show the relation between ΔQ_{cg} and Q_{nloc} derived from triangles with fixed legs $r_{12} = r_{13}/2$ of 12 and 36 h^{-1} Mpc respectively in Fig. 3.4 for the mass samples M3 at $z = 0.0$, for which we find a large non-local bias amplitude. At large scales ($r_{12} = 36\ h^{-1}$ Mpc) the slope of the measured ΔQ_{cg} - Q_{nloc} relation is comparable with the local Lagrangian prediction. Interestingly at small scales the ΔQ_{cg} - Q_{nloc} relation is also linear, while the slope has the opposite sign than at large scales. The linearity at small scales indicates that higher-order terms enter the Q in a similar way as second-order non-local contributions to the bias function. This suggests that linear bias measurements can be improved by using the prediction for Q_{nloc}, while using the local Lagrangian prediction for the non-local bias g_2 is only appropriate at extremely large scales.

3.2.4 Scale dependence

Based on the methods for measuring linear, second-order and non-local bias from Q and $Q^{\times}$ (b_Q, $b_{Q^{\times}}$, $b_{\Delta Q}$, c_Q, $c_{Q^{\times}}$ and $c_{\Delta Q}$), which were presented above, we now apply our analysis to each of the mass samples M0-M3 at $z = 0.5$. We study the scale dependence of our results as before by varying the size of the triangle leg r_{12} between 12 h^{-1}Mpc and 36 h^{-1}Mpc) while fixing $r_{12}/r_{13} = 1/2$. The various bias estimations are presented in Fig. 3.5 for different triangles sizes, defined by the length of r_{12}.

From the comparison between the different bias estimations we draw similar conclusions as from the example of M2 at $z = 0.5$. On linear scales (sufficiently large triangles) the linear bias parameters obtained from each method reaches a regime in which they become scale independent. However, they do not converge to the same value. In case of the linear bias, only $b_{\Delta Q}$ is in agreement with $b_{\xi}^{\times}$, while b_Q and $b_{Q^{\times}}$ overestimate the linear bias; this overestimation is stronger in the case of Q than $Q^{\times}$. We compare this overestimation with predictions. These predictions are based on the fact that both, the local and the non-local bias model deliver relations between Q_h and Q_m, which agree with the measurements. Hence, one can obtain the effective bias derived from anlaysing Q in the local bias model with the ansatz $Q_h \simeq (Q_m + c_2 + g_2 Q_{nloc})/b_1 \simeq (Q_m + c_2^{\mathtt{eff}})/b_1^{\mathtt{eff}}$. Using Q_m and Q_{nloc} from the power spectrum and measurements of of b_1, c_2 and g_2 from the ΔQ approach one can derive the values for $b_1^{\mathtt{eff}}$ and $c_2^{\mathtt{eff}}$. The values for b_1, shown in the top panel of Fig. 3.5, agree with the b_Q measurements. Hence the overestimation of the linear bias by b_Q can indeed be attributed to the negligence of the non-local bias (which is probably contaminated with local second-order terms).

The scale dependence shown for the high mass bin M3 in Fig. 3.5 shows that, if the analysis is performed at too small scales, then one can measure a positive non-local bias, while it is in reality negative (as we see from the results derived at large scales). We indeed verified that for highly biased tracers and small triangles, the curvature of $c_{\Delta Q}$ flips from positive to negative. This scale dependence indicates a domination of the signal by non-linear

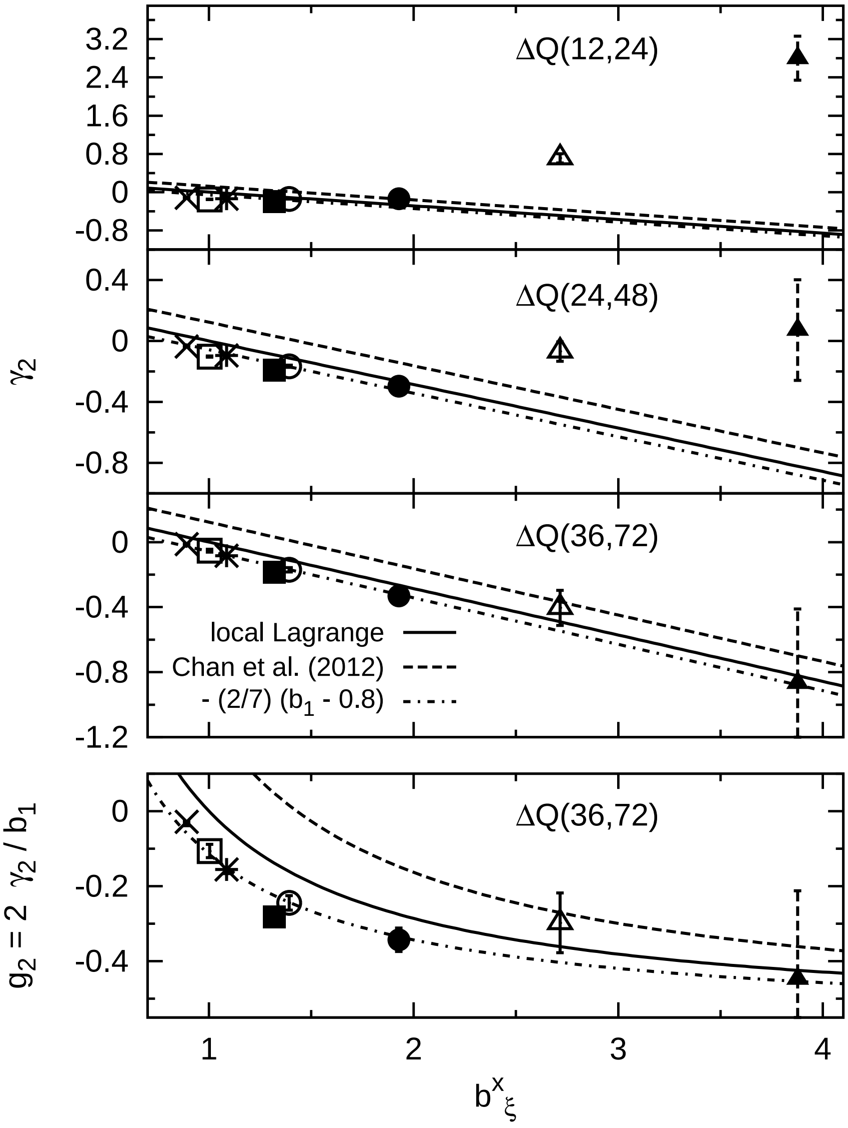

Figure 3.6: Non-local bias parameters γ_2 and $g_2 = 2\gamma_2/b_1$ versus the linear bias from the two-point cross-correlation, $b_\xi^\times$. Measurements, derived from $\Delta Q_{cg} = (Q_h - Q^\times)$ via equation (3.12) are shown as symbols with 1σ errors. Results are shown for Q measurements based on triangles with fixed legs of $r_{12} = 12$, 24 and 36 h^{-1}Mpc and $r_{12} = r_{13}/2$ configurations (from the top to the bottom). Crosses, open squares, open circles and open triangles show the non-local bias, measured for the mass samples M0, M1, M2 and M3 respectively at redshift $z = 0.0$. Stars, closed squares, closed circles and closed triangles show the corresponding M0-M3 measurements at $z = 0.5$. The measurements for $r_{12} = 36$ h^{-1}Mpc are approximated with $\gamma_2 = -(2/7)(b_1 - 0.8)$ (black dash-dotted lines) and compared with the local Lagrangian prediction ($\gamma_2 = -(2/7)(b_1 - 1)$) as well as to a fit given by Chan et al. (2012, $\gamma_2 = -(2/7)(b_1 - 1.43)$), shown as solid and dashed black lines respectively.

terms in these cases. However, for lower mass samples and large scales, for which we expect non-linear contributions to converge to zero, we still see a strong angular dependence of $c_{\Delta Q}$, which speaks for the presence of non-local bias contributions. Hence, we have shown that all the mass sample used in the present analysis exhibit a detectable non-local component.

In case of estimators of the second-order bias c_2 (middle panel of Fig. 3.5) we observe the same tendency, however we do not have a reference estimate. As a result we shall be more confident in the estimation of c_2 coming from $c_{\Delta Q}$. We compare the latter to measurements from third-order moments in Section 3.3. A comparison with c_2 derived from various methods will be presented in Bel et al. (in preparation).

The bottom panel shows that each mass sample comprises non-local bias which significantly differs from the local model $\gamma_2 = 0$. These measurements therefore constitute the first detection of non-local bias in configuration space. In the case of M0 and M1 the amplitude of the non-local bias strongly differs from the local Lagrangian biasing relation between the halo and matter field (Mo and White, 1996).

3.2.5 Non-local to linear bias relation

Following Chan et al. (2012), we compare our non-local bias measurements to the linear bias derived from the two-point cross-correlation in the top panels of Fig. 3.6. This comparison includes measurements at redshift 0.0 and 0.5 which are based on triangles with $r_{12}/r_{13} = 1/2$ configurations. For very large triangles ($r_{12} = 36\,h^{-1}$Mpc) our results indicate a linear relation between the non-local γ_2 and the linear b_1 bias, as expected for the local Lagrangian biasing. However, the amplitude of this relation lies below the local Lagrangian prediction, which is the opposite of what was reported by Chan et al. (2012). Some work is currently ongoing, aiming to explain whether these differences result from the fact that Chan et al. (2012) conduct their measurements using the Bispectrum in Fourier space, while we employ the reduced three-point correlation in configuration space. A further contribution to the discrepancies could arise from differences in the simulation, such as mass resolution effects, or differences between cosmological parameters.

Our measured $b_1 - \gamma_2$ relation shows the same tendency as those of Baldauf et al. (2012), Sheth et al. (2013) Saito et al. (2014) who also find γ_2 to be below the local Lagrangian prediction (see Table 3.1 for the cosmologies assumed in these studies). How strongly the deviations between the results from different studies are driven by differences in the cosmology, in the simulation analysis or in the halo definition remains subject of current studies.

Note also that the departures from the local Lagrangian prediction in Fig. 3.5 are strongly scale dependent for highly biased samples ($b_\xi \gtrsim 2$), which indicates the presence of non-linear contamination to Q_m and Q_{nloc} (e.g. Saito et al., 2014).

3.3 Results

In the previous section we studied measurements of the linear, quadratic and non-local bias parameters (b_1, c_2 and g_2 respectively). These measurements were derived from three-point correlations taking two different approaches.

The first approach is to compare the three-point auto- or cross-correlation with the three-point matter auto-correlation to derive the linear and quadratic bias parameters via equation (3.6) or (3.7). This method is based on the assumption of a local bias model. The linear

reference	O_m	O_b	σ_8	h	n_s
MICE-GC	0.25	0.044	0.8	0.7	0.95
Saito et al. (2014)	0.279	0.0462	0.81	0.7	0.96
Baldauf et al. (2012)	0.272	0.0455	0.81	0.704	0.967
Chan et al. (2012)	0.27	0.046	0.9	0.72	1.0
Sheth et al. (2013)	0.25	0.04	0.8	0.7	1.0

Table 3.1: Cosmological parameters of the MICE-GC simulation (analysed in this work) compared to other simulations used for studying non-local bias in the literature.

and quadratic bias parameters, derived this way from the auto-correlations are called b_Q and c_Q respectively. The corresponding quantities derived from the cross-correlation are called $b_Q^\times$ and $c_Q^\times$. The second approach is to use particular combinations of three-point auto- and cross-correlations. The linear bias parameters, derived this way via equation (3.11), $b_{\Delta Q}$, are independent of any quadratic contributions (local or non-local) to the bias function, as explained in the previous section. The quadratic and non-local bias parameters are obtained simultaneously by fitting predictions for the non-local component of the three-point correlation function, Q_{nloc}^{PT}, to ΔQ_{cg}, defined in equation (3.12). The quadratic parameters from such measurements is called $c_{\Delta Q}$.

In this section we aim at comparing the bias estimations coming from these two approaches. We therefore present the different linear and quadratic bias estimations for the four mass samples M0-M3 in the comoving simulation outputs at the redshifts 0.0 and 0.5 versus the mean halo mass in each sample in Fig. 3.7. Our bias measurements are performed using the same three-point correlations as in the previous sections with 18 opening angles between 0 and 180 degree, which are based on triangles with $r_{12} = r_{13}/2 = 36\ h^{-1}$Mpc. All error bars denote the standard deviation, derived from 64 Jackknife samples as described in Section 2.3.6.

The linear bias estimations from the different methods are presented in the upper panel of Fig. 3.7. We compare these estimations to reference measurements from the two-point cross-correlation, defined in equation (2.14) and fitted over the range $20 - 60\ h^{-1}$Mpc, which we consider as reliable (see Section 2.3). The relative deviations to this reference linear bias are shown in the central panel.

We find that the estimator b_Q, which neglects the non-local bias, overestimates the linear bias by 10-30%. This result is consistent with the expected overestimation, derived via the g_2 measurements from ΔQ_{cg} and shown as black filled symbols on Fig. 3.7. Using bias estimator $b_Q^\times$ we find weaker overestimation of 5-10%, which can be attributed to the lower impact of non-local contributions to the three-point cross-correlation compared to the corresponding auto-correlation, as discussed in the previous section. However, besides non-local terms, the discrepancies between $b_\xi \times$and b_Q can also be caused by various other effects, such as stochasticity or contributions of higher-order terms to the bias expansion (equation 3.1). The Jackknife estimation of the errors and the covariance matrix introduces an additional uncertainty in the bias measurement (see Appendix A.1).

The linear bias parameters from ΔQ is in excellent agreement with the reference for all mass ranges and at both redshifts. Deviations are in the range of the 1σ of $b_{\Delta Q}$, while the latter roughly correspond to 1% of the amplitude. For the mass sample M3 deviations become

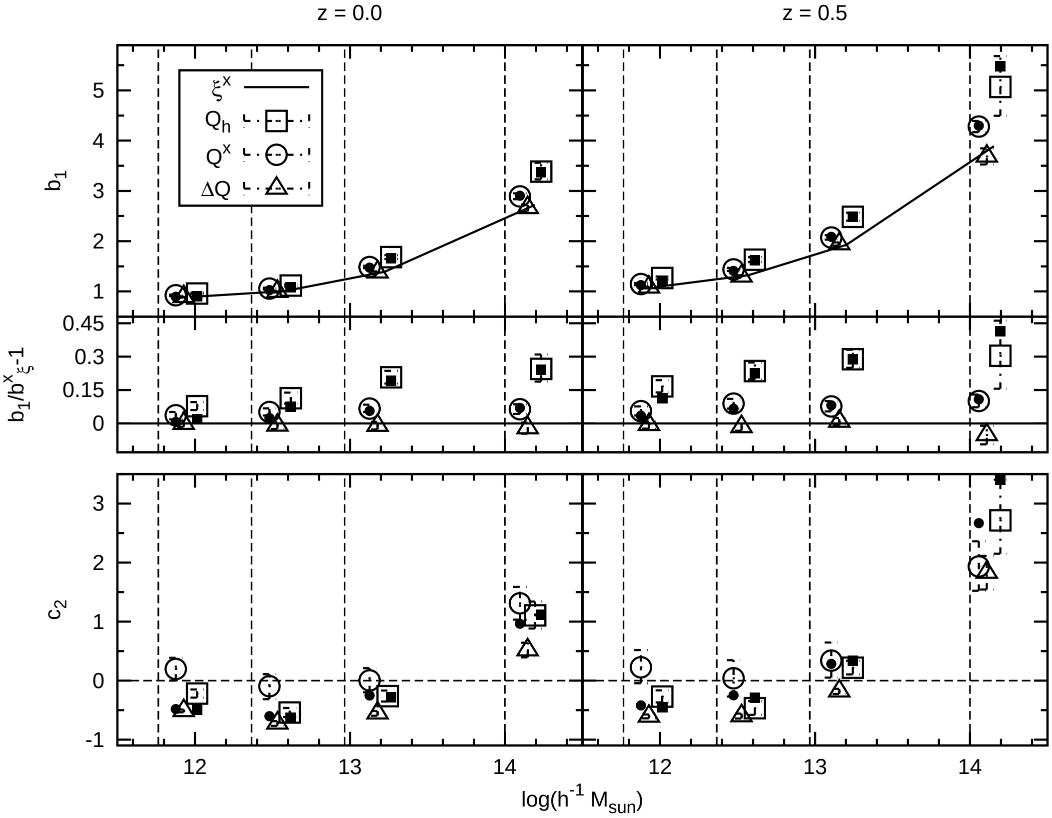

Figure 3.7: *Top:* Summary of the linear bias measurements presented in this Chapter. Results are shown for the redshifts $z = 0.0$ and $z = 0.5$ (left and right respectively) versus the mean halo mass of each mass sample M0-M3, defined in Table 2.1 (lower and upper limits of the mass samples are marked by vertical grey dashed lines. Symbols are slightly shifted along the the mass axis for clarity). The smaller panel shows the relative deviation of each estimator with respect to the linear bias from the two-point cross-correlation ($b_\xi^\times$, solid lines). The estimators b_Q, $b_Q^\times$ and $b_{\Delta Q}$ (from equations (3.6), (3.7), (??) respectively) are displayed respectively by open squares, open circles and open triangles. For measuring b_Q and $b_Q^\times$ we assumed a local bias model, i.e. $g_2 = 0$. Predictions for the effective linear and quadratic bias measurements via Q while ignoring non-local bias (see Section 3.2.4) are shown as closed symbols. *Bottom:* Summary of the second-order bias measurements c_Q, $c_Q^\times$ and $c_{\Delta Q}$ from the same bias estimation methods, as used for the linear bias in the top panel (equation (3.6)), (3.7)) and (3.12) respectively). For estimations based on Q we used triangles with $(r_{12}, r_{13}) = (36, 72)$ h^{-1}Mpc configurations. Error bars denote 1σ uncertainties.

slightly larger and more significant. We find this agreement also for smaller triangle scales, as we demonstrated for $z = 0.5$ in Fig. 3.5 and discussed in the previous section.

The lower panel of Fig. 3.7 shows how the methods compare in terms of estimating the second-order bias parameter c_2. Given the good agreement between $b_{\Delta Q}$ and $b_\xi^\times$ we assume that $c_{\Delta Q}$ is a good estimate of c_2. We find that the second-order bias estimated from the three-point correlations, neglecting non-local bias (b_Q and $b_Q^\times$), leads to significant departures from $c_{\Delta Q}$. This trend is consistent with the expected deviation, shown again black filled symbols. Interestingly the measured overestimation is stronger in case of the cross-correlation, except for M3 at redshift $z = 0.5$. This might result be caused by an incorrect estimation of the covariance between angular bins of Q, which enters the fits for the bias estimations (see Section 2.3.6). We found in Appendix A.1 that the b_Q estimation is significantly affected by the covariance estimation, especially for sample with high halo densities (low halo masses). The impact of the covariance becomes weaker for samples of massive haloes, where the error is shot-noise dominated. For this reason the c_2 estimations from three-point cross-correlations might be more strongly affected by shortcomings of the covariance estimation than estimations from the auto-correlations. Note that the covariance between angular bin of the ΔQ_{cg} estimator is lower than in the case of Q^h and $Q^\times$, as off-diagonal elements cancel out.

The bias measurements presented in this section will be compared with predictions from the peak-background split in Chapter 4.

3.4 Summary and Conclusion

We studied linear, quadratic and, for the first time in configuration space, non-local bias of halo clustering with respect to the clustering of the dark matter field. We therefore employed various second- and third-order statistics of halo and matter density fluctuations in the MICE-GC simulation. Our goal was to find if the overestimation of the linear bias parameter by the three-point auto-correlation, which we found previously in Chapter 2 (see also Manera and Gaztañaga, 2011; Pollack et al., 2012), can be attributed to shortcomings of the local quadratic bias model. Understanding this difference is crucial for breaking the degeneracy between growth and bias with three-point correlations, which would strongly amplify the statistical power of large-scale structure surveys. To achieve this goal we employ auto- and cross-statistics to disentangle the effects of linear bias on second- and third-order halo statistics from those originating from non-linear and non-local bias.

Based on our findings in Chapter 2 we consider the two-point cross-correlation as an reliable estimator for the linear bias. We therefore used it as reference in the subsequent analysis.

For studying the impact of non-linear and non-local bias on the three-point correlation we compared in Section 3.2 bias measurements from the (reduced) three-point halo-matter cross-correlation to those from the auto-correlation, using the local quadratic model. We found the linear bias from the cross-correlation to be closer to the reference than the linear bias from the auto-correlation. This is expected from second-order perturbation theory, which predicts the three-point cross-correlation to be less affected by quadratic local and non-local bias than the corresponding auto-correlation (compare equation (3.6) and (3.7)). However, the three-point cross-correlations delivers, as the corresponding auto-correlation, linear bias measurements, which lie significantly above the reference from the two-point correlation.

To further verify if this overestimation can be attributed to non-linear and non-local contributions to the bias model we take advantage of the fact that three-point auto- and cross-correlations are affected differently by non-linear and non-local bias, but equally by the linear bias (see again equation (3.6) and (3.7)). This property allows for combinations of the auto- and cross-statistics which isolate the linear from the non-linear and the non-local bias. We find the linear bias, measured by such a combination of three-point correlations independently of quadratic or non-local bias (equation 3.11) to be in excellent agreement with the reference from the two-point correlation (Fig. 3.2). This finding is a strong indication that non-local terms are indeed the reason for the overestimation of linear bias from three-point correlation, when ignoring them by assuming a local quadratic bias model. This approach could be used to measure linear bias by cross-correlating galaxy with lensing maps. The presence of non-local bias also becomes apparent in our measurements of non-linear bias contributions (local and non-local) via equation (3.12), which are in good agreement with predictions for the non-local contributions to the three-point correlation (Fig. 3.3). Our results therefore constitute the first detection of non-local bias in configuration space and demonstrate the paramount importance of taking it into account when analysing galaxy surveys. Using these non-local bias measurements we can predict how strongly the bias from thee-point correlations is overestimated by three-point auto- and cross-correlations, when non-local terms are neglected. These predictions are in good agreement with our measurements.

When the considered scales are two small ($r_{12} \lesssim 30\ h^{-1}$Mpc), the non-local bias parameter γ_2, derived from our measurements, shows a strong scale-dependence, indicating the presence of higher-order local or non-local terms in the bias function. Instead, for scales larger than 36 h^{-1}Mpc we find a linear relation between the non-local and the linear bias, over the whole mass range, as predicted by the local Lagrangian bias model (Fig. 3.6). However, the amplitude of this relation lies significantly below the local Lagrangian prediction. This is in agreement with results from Baldauf et al. (2012) and Saito et al. (2014), but in contradiction with results from Chan et al. (2012), who find the non-local bias to be above the local Lagrangian prediction. Whether this latter disagreement comes from the fact that Chan et al. (2012) analysed the bispectrum in Fourier space using different simulations is the subject of current investigations. An alternative reason for this discrepancy could be the inaccuracy of the prediction of the non-local contribution to the three-point correlation of matter, from which we derive the non-local bias parameter.

Comparing the quadratic bias from the different estimators we found the higher values from the estimators which neglect non-local bias compared to results from the ΔQ_{cg} measurements, which corresponds to our exception. However, we also find indications for other effects sources of uncertainties in the measurements, which could be shortcomings in the covariance estimation.

Our results show that the local quadratic bias model is inadequate to describe halo bias in the MICE-GC simulation. Non-local second-order terms need to be taken into account for accurate measurements of the linear bias with three-point correlation function. Two approaches are possible to do so. Non-local bias can be isolated from linear bias by combining different third-order statistics (i.e. ΔQ), or the non-local contributions need to be directly modelled. The first approach, on which we focused in this analysis, might be implemented in terms of cross-correlations between lensing and galaxy maps. We will test the second approach in a future analysis, but already provide here an expression for the non-local contribution to the three-point correlation in configuration space. At scales below $30 h^{-1}$Mpc we find indications for the presence of higher-order terms in the bias function (local or non-local). Modelling the non-local bias as a linear function of the linear bias parameter, as suggested by the local Lagrangian bias model, therefore appears to be only suitable at very large scales.

We compare the linear and second-order bias measurement, obtained in the present analysis to peak-background split predictions in Chapter 4. These various bias estimations can be used to verify the universal relation between linear and non-linear bias parameters.

Chapter 4

Comparing halo bias from abundance and clustering

Abstract

In this Chapter we model the abundance of haloes in the $\sim (3 \text{ Gpc}/h)^3$ volume of the MICE Grand Challenge simulation by fitting the universal mass function with an improved Jack-Knife error covariance estimator that matches theory predictions. We present unifying relations between different fitting models and new predictions for linear (b_1) and non-linear (c_2 and c_3) halo clustering bias. Different mass function fits show strong variations in their performance when including the low mass range ($M_h \lesssim 3\ 10^{12}\ M_{\odot}/h$) in the analysis. Together with fits from the literature we find an overall variation in the amplitudes of around 10% in the low mass and up to 50% in the high mass (galaxy cluster) range ($M_h > 10^{14}\ M_{\odot}/h$). These variations propagate into a 10% change in b_1 predictions and a 50% change in c_2 or c_3. Despite these strong variations we find universal relations between b_1 and c_2 or c_3 for which we provide simple fits. Excluding low mass haloes, different models fitted with reasonable goodness in this analysis, show percent level agreement in their b_1 predictions, but are systematically $5-10\%$ lower than the bias directly measured with two-point halo-mass clustering. This result confirms previous findings derived from smaller volumes (and smaller masses). Inaccuracies in the bias predictions lead to $5-10\%$ errors in growth measurements. They also affect any HOD fitting or (cluster) mass calibration from clustering measurements.

4.1 Introduction

In the previous chapters we studied the halo bias using second- and third-order clustering statistics. However, besides the clustering also the abundance of haloes as a function of halo mass (known as the mass function) is related to the bias function. This relation can be understood with the peak-background split model (hereafter referred to as PBS model, Bardeen et al., 1986; Cole and Kaiser, 1989; Mo and White, 1996). In this model, large-scale density fluctuations are superposed with fluctuations at small scales. These large-scale density fluctuations modulate the background cosmology (i.e. the mean density and the Hubble rate) around small-scale fluctuations (e.g. Martino and Sheth, 2009). The critical density contrast for gravitational collapse therefore depends on the environment. In regions with large-scale overdensities more small-scale fluctuations collapse to haloes than in underdense regions. This effect modifies the abundance of haloes and also their spatial distribution as they follow the pattern of the peaks of large-scale fluctuations. Haloes are therefore biased tracers of large-scale fluctuations in the full matter density field. If the matter power spectrum is known the halo bias parameters can be predicted from the mass function via the PBS model.

The PBS bias predictions can be used to determine the dark matter clustering from observed galaxy distributions if the halo masses of a given tracer sample are known (or the other way round). Such an analysis requires that the bias parameters, predicted from the mass function, are equivalent with the bias which affects the clustering. Studies of this equivalence have revealed that the PBS predictions for the linear bias b_1 are around 10% below measurements from two-point clustering statistics. Such deviations might result from assumptions of the PBS model, such as spherical collapse, or a local bias relation (e.g. Mo et al., 1997; Desjacques et al., 2010; Paranjape et al., 2013; Schmidt et al., 2013). Further numerical effects, like the definition of haloes in N-body simulation, or systematic effects such as the parametrisation and fitting procedure of the mass function might contribute to the discrepancy between the bias from PBS and clustering (e.g. Hu and Kravtsov, 2003; Manera et al., 2010). Predictions of the PBS for the relation between halo mass and bias are also employed in Halo Occupation Distribution models to predict the bias as a function of galaxy properties, such as luminosity or color (e.g. Cooray and Sheth, 2002; More et al., 2011; Coupon et al., 2012; Carretero et al., 2015). Inaccuracies of the PBS can affect such halo model predictions for galaxy bias or the average number of galaxies per halo. Moreover, halos of equal mass could have different galaxy occupation, depending on their environment (e.g. Pujol and Gaztañaga, 2014). Besides clustering analysis the PBS can be employed for estimating the lower mass threshold (or mass-observable relation) of observed galaxy samples. This so-called self calibration method (Lima and Hu, 2004, 2005) uses the fact that both, the clustering and the abundance of haloes, depend on halo mass. Inaccuracies of the PBS model can change the estimation of halo mass thresholds and therefore change the cosmological parameter estimates from such an analysis (e.g. Manera and Gaztañaga, 2011; Wu et al., 2010).

The broad application of the PBS model in large-scale structure analysis and the precision of abundance and clustering measurements from incoming observational data calls for a detailed validation of the PBS bias predictions. The purpose of this analysis is to pursue the study of deviations between halo bias measurements from clustering and PBS predictions using the wide mass range of the MICE Grand Challenge (hereafter referred to as MICE-GC) simulation. We thereby focus on the effect of mass function parametrisation and fitting on PBS bias predictions. The mass function fits are affected by the error estimations. Our analysis therefore includes a detailed study of the mass function error and covariance. This analysis leads us to an improvement of the standard Jack-Knife estimator. The study of PBS bias predictions includes non-linear bias parameters which are important for an analysis of higher-order correlations of the large-scale halo distribution and two-point correlations at small scales. We further compare the mass function fits and bias predictions with results from the literature based on different simulations, to verify a universal behavior of these quantities.

This chapter is organised as follows. In Section 4.2 we present the MICE-GC simulation and mass function fits. In Section 4.3, we present new galaxy bias predictions, compare with the literature and find a universal relation between bias parameters. In Section 4.4, we compare these predictions with the bias directly measured with two-point and three-point halo-matter cross-correlations of the MICE-GC simulation. Section 4.5 contains a summary and conclusion. In the Appendix B, we present a new method to improve the Jack-Knife covariance matrix estimation. This method can also be easily generalised to other statistics, such as the two-point correlation function (Hoffmann et al., in preparation).

4.2 Halo Mass Function

4.2.1 Mass function definition and measurement

The unconditional mass function, $dn(m)$, is defined as the comoving number density of haloes with masses between m and $m + dm$. The mass function can be written in a form which is nearly independent of redshift, cosmology and initial power spectrum (Press and Schechter, 1974; Bond et al., 1991; Sheth and Tormen, 1999) as

$$\nu f(\nu) \equiv \frac{m}{\bar{\rho}} \frac{dn(m)}{dln\ \nu}, \tag{4.1}$$

where $\bar{\rho}$ is the mean comoving mass density. The hight of density peaks is defined as

$$\nu \equiv \delta_c^2 / \sigma_m^2(m), \tag{4.2}$$

where $\delta_c = 1.686$ is the critical density for spherical collapse (which is the exact solution value for the spherical collapse in an Einstein-de Sitter universe). The variance of matter density fluctuations, $\sigma_m^2(m)$, smoothed with a spherical top-hat window with radius $R(m) = (\frac{m}{\rho}\frac{3}{4}\pi)^{1/3}$, can be calculated as

$$\sigma_m^2(m) = \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} W^2(kR(m)) \tag{4.3}$$

where $W(x) = (3/x^3)(\sin x - x \cos x)$ is the spherical top-hat window in Fourier space and $P(k)$ is the linear power spectrum. Note that m refers to the matter density field when it appears as lower index and to the mass, enclosed by $R(m)$, when used as a variable. We measure the mass function in the MICE-GC simulation at redshift z and convert it to $\nu f(\nu)$, to predict the halo bias parameters b_1, c_2 and c_3 via the PBS theory (Bardeen et al., 1986; Cole and Kaiser, 1989; Mo and White, 1996). We do not apply the halo mass correction suggested by Warren et al. (2006) for low mass resolution, since we analyse haloes down to 20 particles, while this correction was only proposed for larger numbers of particles per halo. Furthermore, it is not clear that the FoF mass, corrected in such a way is closer to the halo mass on which the PBS model is based on. More details about how we measure the mass function are given in the Appendix B.1.

Our measurements of $\nu f(\nu)$ at $z = 0.0$ and $z = 0.5$, shown as symbols in Fig. 4.1, agree visually with the expected weak redshift dependence of FoF mass functions for redshift independent linking lengths (e.g. Press and Schechter, 1974; Jenkins et al., 2001; More et al., 2011). Errors and covariances of the measurements were derived with a new estimator which combines the Jack-Knife approach with predictions for sampling variance from the power spectrum (see Appendix B.2). We also show in the Fig. 4.1 fits to the measurements, based on the mass function parametrisation of Tinker et al. (2010, equation (4.4)). The model, fitted over the mass range M123 (that is, excluding the first mass bin, see Table 4.2) is in reasonable agreement with the measurements. Including the low mass sample M1 (haloes with less than 80 particles) to the fitting range leads to poor fits of the model. The fits at both redshifts differ by less than 5% for $\nu \lesssim 3$, confirming low redshift dependence from the

model	reference	constrains
Tinker	Tinker et al. (2010)	A, a, b, c, d free
Warren	Warren et al. (2006)	$d = 0$
ST	Sheth and Tormen (1999)	$c = b$, $d = 1/2$
PS	Press and Schechter (1974)	$a = 0$, $c = 1$, $d = 1/2$
proposal	this work	$c = 1$

Table 4.1: Constrains of parameters in equation (4.4) corresponding to different mass function models. We refer to the models in the text using the abbreviations given in the left column.

measurements. When including lower masses, the redshift dependence is stronger, possibly because of redshift dependent noise in the low mass FoF detection. At larger masses ($\nu > 3$) we find up to 10% deviations, which are comparable with the mass function errors. We have verified that our conclusions also hold for fits over the higher mass range, M23, and different mass function binnings. A detailed analysis of the mass function fits, including fits of other mass function models over different mass ranges and different binnings as well as a comparison with fits compiled from the literature, can be found in Section 4.2.2.

4.2.2 Mass function fits

In order to predict the halo bias from the mass function via the PBS approach we fit different mass function models to the measurements. Several systematic effects, such as the choice of the mass function model or the mass range over which the model is fitted can limit the accuracy of the PBS bias predictions (e.g. Manera et al., 2010; Manera and Gaztañaga, 2011). The objective of the subsequent analysis is to find out how strongly these effects impact the predicted linear, quadratic and third-order bias. In particular we aim to verify if the disagreement between PBS predictions for the linear bias and the corresponding measurements from two-point correlations, presented in Section 4.4, is driven by possible shortcomings of the mass function fits. We therefore study in this subsection the fitting performance of different mass function models.

The latest model in our analysis with the highest number of free parameters is the expression given by Tinker et al. (2010) (hereafter referred to as Tinker model). It can be written as

$$\nu f(\nu) = A[1 + (b\nu)^a]\nu^d e^{-c\nu/2}, \tag{4.4}$$

where A, a, b, c, d are the free parameters. Fixing certain parameters delivers expressions which correspond to the mass function models suggested by Press and Schechter (1974), Sheth and Tormen (1999) and Warren et al. (2006) (hereafter referred to as PS, ST and Warren model respectively). The corresponding parameter constrains are summarised in Table 4.1 together with the abbreviations for the reference of each model. In Table 4.1 we also propose a new constrain, which constitute a new mass function fit. Its advantage is that it has as many free parameters as the Warren model, but matches the mass function better when we fit over the whole mass range, as we show later. In our analysis we will focus on the models of ST, Warren, Tinker and our proposal. We determine the best fitting parameters

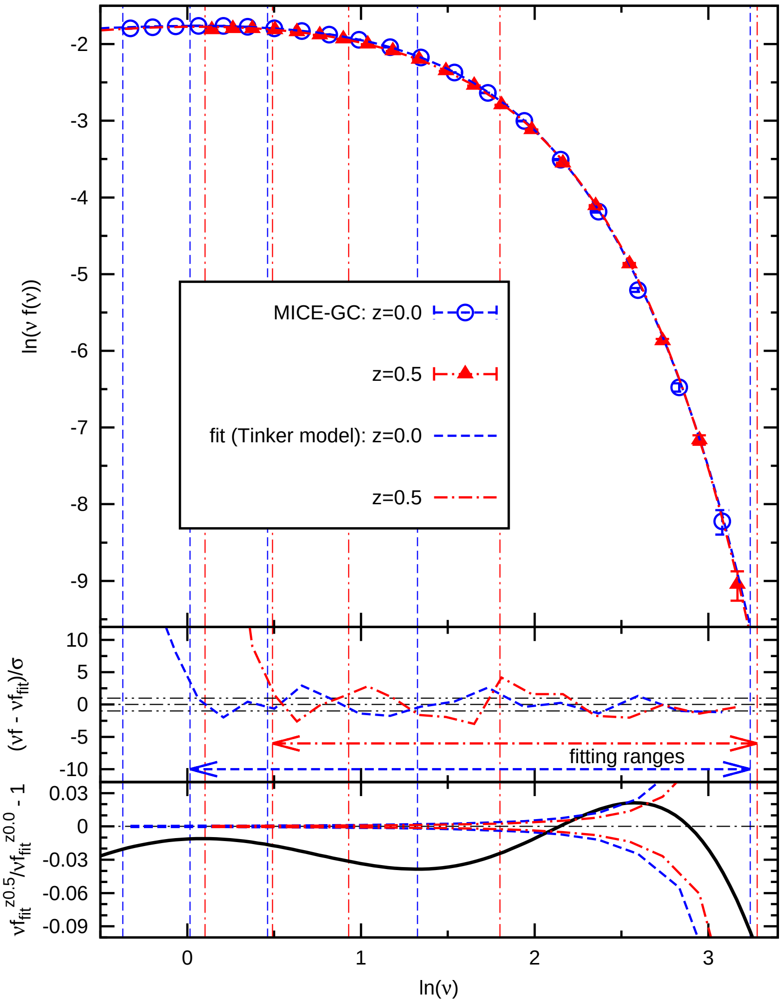

Figure 4.1: *T*op: unconditional halo mass function, defined in equation (4.1), as a function of the peak-hight $\nu \equiv \delta_c^2/\sigma^2(m)$. Symbols show MICE-GC measurements with 1σ errors based on FoF groups at the redshifts $z = 0.0$ and $z = 0.5$ (blue circles and red triangles respectively). Lines show the mass function model of Tinker et al. (2010), fitted to the measurements over the mass range M23 in the same colour coding as the symbols. *C*enter: significance of the deviation between measurements and fits. *B*ottom: relative deviation between the fits at z=0.0 and z=0.5 (black solid line). The 1σ errors of the measurements are shown as lines in the same colour coding as in the top panels. Vertical blue dashed and red dash-dotted lines denote the limits of the halo mass samples M0-M3 at z=0.0 and z=0.5 respectively, given in Table 2.1.

mass range	halo masses $[10^{12}h^{-1}M_\odot]$	N_p
M0123	≥ 0.58	≥ 20
M123	≥ 2.32	≥ 80
M23	≥ 9.26	≥ 316
M012	$0.58 - 100$	$20 - 3416$

Table 4.2: Halo mass ranges for mass function fits and clustering analysis.

for each mass function model by minimising

$$\chi^2 = \sum_{ij}^{N_{bin}} \Delta_i \hat{C}_{ij}^{-1} \Delta_j, \tag{4.5}$$

with $\Delta_i \equiv (X_i^{fit} - X_i)/\sigma_{X_i}$ and $X = \nu f(\nu)$. $\hat{C}_{ij}$ and σ_{X_i} are derived from our new JK estimator, introduced in Appendix B.2. For searching the best fitting parameters we implemented a Monte Carlo Markov Chain algorithm to explore the parameter space.

In Fig. 4.2 we show the significance of the deviations between the mass function measurements and the best fits by the different models. Results are shown at the redshifts $z = 0.0$ and $z = 0.5$, while at each redshift we fit the mass function over the different mass ranges, which are shown in Table 4.2. The first includes all halo mass samples (M0123), the second and the third exclude the low mass samples (M123, M23) and the fourth mass range excludes the highest mass sample (M012). For each fitting range we show fits based on seven different mass function binnings, dividing the mass range into $20, 25, 30, \ldots, 50$ logarithmic bins. We find that the deviations between fit and measurement can vary with the binning. However, we also see trends which are independent of this systematic effect.

All mass function models show a clear dependence of the best fit on the chosen mass range, while this dependence is weakest for the ST model. At both redshifts and all mass ranges the Tinker model fits the measurements best. This can be attributed to the fact that it contains the highest number of free parameters. The best fit parameters for the Tinker model are given in Table 4.3. For fits over the whole mass range (M0123) our proposed mass function model seems to match the measurements almost as good as the Tinker model, while having one free parameter less. It also has the advantage of producing stable values for the parameters regardless of the range used for the fit. When the fits are performed only at the highest mass range (M23) the Tinker and the Warren mass functions fit the data equally well, while the proposed model is a slightly worse fit. The ST model delivers the poorest fits in all cases. At $z = 0.5$ we find strong deviations between fits and measurements when the fitting range includes the low mass sample M0. This indicates that the FoF detection of low mass haloes can be strongly affected by shot-noise, while this effect is stronger at higher redshift.

For studying the goodness of the best fits for the different mass function models we present their best fit parameters and the corresponding χ^2 values per degree of freedom (*d.o.f.*) in Fig. 4.3, where the *d.o.f.* refer to the number of mass function bins used for the fit. Results are shown for fits over the mass ranges M0123, M123 and M23, which correspond to the different minimum peak-hights given by the x-axis. For clarity we show here only results at redshift $z = 0.0$, while we find similar results at $z = 0.5$. For each fit we show mean results with standard deviations from the seven mass binnings mentioned previously. In addition to the results derived by taking the covariance between different mass function bins into account in

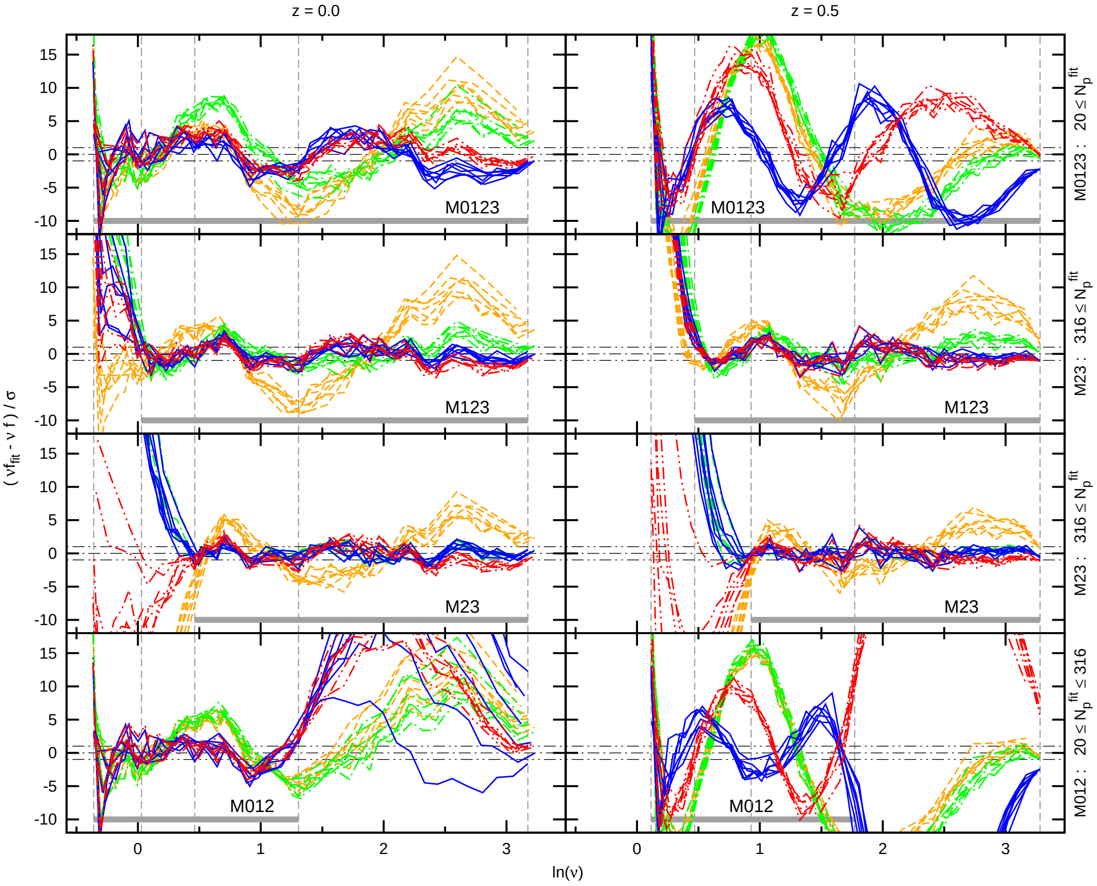

Figure 4.2: Significance of the deviations between mass function fits and measurements versus the peak-hight ν. The top, central top, central bottom and bottom panel show results from fits over the mass ranges M0123, M123, M23 and M012 respectively. These fitting ranges are marked by thick grey horizontal lines. Grey vertical lines denote the minimum and maximum peak-heights of the different halo mass samples M0-M3. Dash-dotted and dashed-double-dotted black lines denote 0 and 1σ deviations between fits and measurements respectively. Results for the redshifts $z = 0.0$ and $z = 0.5$ are shown in the left and right panels respectively. Blue, green and orange lines show fits to the models of Tinker, Warren and ST respectively, while fits to our proposed model are shown as red lines. For each fit we show seven fits, which were derived from mass function measurements based on dividing the whole mass range into $20, 25, 30, ..., 50$ bins.

z	mass range	A	a	b	c	d	$\frac{\chi^2_{min}}{d.o.f.}$
0.0	M0123	0.28	1.80	0.22	1.08	0.47	25.6
0.5	M0123	0.31	2.74	0.20	1.37	0.87	125.6
0.0	M123	0.24	1.39	0.22	0.94	0.34	3.4
0.5	M123	0.26	1.70	0.17	0.98	0.45	3.5
0.0	M23	0.17	1.10	0.55	0.85	0.01	1.5
0.5	M23	0.22	1.28	0.34	0.86	0.05	1.6

Table 4.3: Best fit parameters for the Tinker mass function model at the redshifts $z = 0.0$ and $z = 0.5$. The mass function was fitted over the mass ranges, M0123, M123 and M23, defined in Table 4.2. Results are shown as mean, derived from different mass function binnings and are also displayed in Fig. 4.3 for $z = 0.0$. The corresponding standard deviations are typically at the 2% level. The fits were performed taking covariance between different mass bins into account.

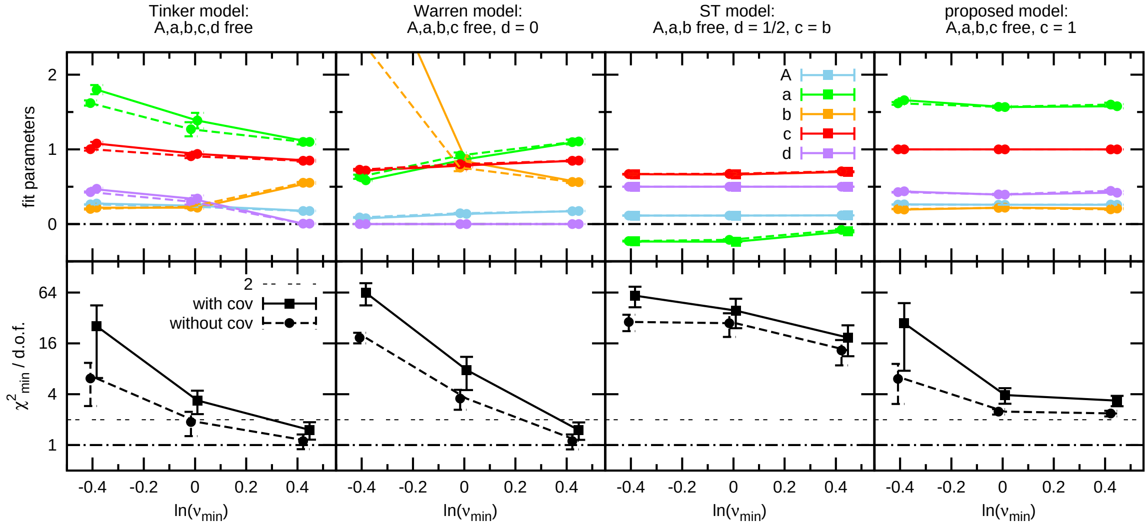

Figure 4.3: *T*op: best fit parameters for different mass function models as a function of the minimum peak-hight (corresponding to the mass ranges M0123, M123 and M23, defined in Table 4.2) used for fits at redshift $z = 0.0$. Symbols show the means with standard deviations derived from seven mass function binnings (see Fig. 4.2). Results from fits performed with and without taking the covariance between different mass function bins into account are connected with solid and dashed lines respectively. In the latter case the symbols are slightly shifted to the left for clarity. *B*ottom: minimum $\chi^2/d.o.f.$ of the fits derived using our new error estimator with 8^3 JK samples. Note that errors can be smaller than the symbol size. We find similar results at redhsift $z = 0.5$.

the fitting procedure we show results which were computed neglecting the covariance. We find that neglecting the covariance can lead to different best fit parameters, especially when the low mass range, where the off-diagonal elements of the covariance have the highest amplitudes, is included in the analysis (see Appendix B.2). However, the bias predictions are only weakly affected by the negligence of the covariance (see Fig. B.5). The final conclusions of this chapter about the comparison between bias predictions and measurements will therefore not dependent on the negligence or inclusion of the covariance in the analysis.

The $\chi^2/d.o.f.$ results, shown in the bottom panel of Fig. 4.3, are very high when the mass functions are fitted over the whole range (lowest minimum peak-hight). This poor performance, which is even apparent for the Tinker model with its five parameters, is probably related to the fact that our mass function measurements are not reliable in the low mass range. In fact the M0123 sample includes haloes with down to 20 particles. For such low numbers of particles per halo we expect strong systematic effect in the halo mass estimation and therefore on the mass function (e.g. Warren et al., 2006; More et al., 2011). Furthermore, the halo samples might be contaminated with spuriously linked FoF groups. If the analysis is performed using only the high mass sample M23 (highest minimum peak-hight), the $\chi^2/d.o.f.$ values for the best fit models drop down to values between unity and four. If we perform the fits ignoring off-diagonal elements in the covariance matrix we obtain substantially lower $\chi^2/d.o.f.$ values, especially when the fits are performed over the whole mass range. This demonstrates that the covariance cannot be neglected in the fit for the evaluation of the fitting performance of a mass function model. This statement is even true in the high mass range where the covariance is dominated by shot-noise.

We also see in Fig. 4.3 that the $\chi^2/d.o.f.$ can change for different mass function binnings, which can already be seen in Fig. 4.2. This dependence on the binning is also apparent when the off-diagonal elements of the covariance matrix are neglected in the fit. However, the best fit values of each model and the corresponding bias estimations are only weakly affected by this systematic effect.

Interestingly the best fit parameters of the Tinker model have the same values as the ones from the Warren model when the fit is performed on the higher mass M23 sample. Consequently the minimum $\chi^2/d.o.f.$ are the same in both cases. This indicates that the parameter d, which is set to zero in the Warren model is not required for fitting the high mass range, but becomes necessary, when the low mass range is included in the fit. The $\chi^2/d.o.f.$ values of our proposed model are smaller than those for the Warren model for minimum peak hights of $ln(\nu_{min}) \lesssim 0$ (M123). This agrees with the visual impression, gained from Fig. 4.2, that our proposed model delivers better mass function fits than the model of Warren, unless the analysis is restricted to the highest mass range M23. We come to the same conclusion when analysing the mass function at $z = 0.5$.

Mass function universality

In Fig. 4.1 we demonstrated that the mass function, when expressed in terms of the peak-high ν, depends only weakly on redshift. To verify if this universality also holds for other cosmologies we compare our mass function fits to the Tinker and Warren model with fits to the same models, compiled from Warren et al. (2006, Table 8), Tinker et al. (2010, Table 4, Δ_{200}), Crocce et al. (2010, Table 2) and Watson et al. (2013, Table 2, FoF Uni.). Crocce et al. (2010) and Watson et al. (2013) fit mass functions to the Warren model. Note that Crocce et al. (2010) also used simulations from the MICE simulation suite, with the same cosmology as

MICE-GC, but rely on the nested boxes approach to cover a similar mass range, while having a higher resolution in the low mass end than MICE-GC. Tinker et al. (2010) used spherical overdensities to define haloes. A universal behaviour would not only be useful for PBS bias predictions, but also for constraining σ_8 with galaxy luminosity functions, statistics of the initial density field and various other application (see e.g. White, 2002).

We compare our mass function fits with those from the literature in Fig. 4.4. We find that the different mass function fits agree at the 10% level in the low mass end, but differ by up to 60% at high masses with a significance of about 2σ in terms of error in the measurement. Departures from universality are expected for different cosmologies but also can result from systematic effects, such as the halo mass definition (e.g. Lacey and Cole, 1994; Sheth et al., 2001; Jenkins et al., 2001; White, 2002; Reed et al., 2007; Lukić et al., 2007; Tinker et al., 2008; Crocce et al., 2010; Courtin et al., 2011; More et al., 2011; Bhattacharya et al., 2011; Castorina et al., 2014). Furthermore, the fitting procedure affects the presented comparison as well.

The comparison between the Warren fit from Crocce et al. (2010) and from this work reveals the strong impact of the latter systematic effects on the fit. These two fits agree well in the high mass end, when the fit is performed over the whole mass range M0123. Interestingly we find that these fits differ more strongly from the measurements in the high mass end than fits from other simulations. Excluding lower masses from the fit (M23) leads to a better agreement between our fit and the measurement in the high mass end and therefore to a stronger difference between the results from Crocce et al. (2010) and ours. The lower amplitude of the Crocce et al. (2010) fit at low masses indicates that the low halo mass MICE-GC halo samples include more spuriously linked FoF groups, which can be expected from the low resolution as we concluded before in this section. Furthermore, a lower mass resolution leads to an overestimation of halo masses. Correcting this effect as suggested by Warren et al. (2006) and done by Crocce et al. (2010) results in a decrease of the amplitude, which is shown as grey symbols in Fig. 4.4. The fact that our Warren corrected mass function is lower than all mass function fits at in the low mass range ($ln(\nu) \lesssim 0$) indicates that the Warren correction leads to an underestimation of halo masses when it is applied on FoF groups with order of 10 particles. For intermediate masses ($0 \lesssim ln(\nu) \lesssim 2$) our Warren corrected measurements are in better agreement with the results from Crocce et al. (2010) than those without Warren correction. A comparison between the Warren corrected MICE-GC mass function at $z = 0.0$ and the Crocce et al. (2010) fit, presented by Crocce et al. (2013), also shows higher amplitude of the prediction compared to the measurement in the highest mass bin at $6\ 10^{14} M_{\odot} h^{-1}$ and an opposite trend for lower masses.

4.3 PBS bias predictions

The bias parameters b_N, introduced in equation (2.1), can be obtained from derivatives of the halo mass function via the peak-background split (hereafter referred to as PBS) approach (Bardeen et al., 1986; Cole and Kaiser, 1989; Mo and White, 1996). Following Scoccimarro et al. (2001) we derive the first-, second- and third-order bias parameters from the mass function fits as

$$b_1(\nu) = 1 + \epsilon_1 + E_1, \tag{4.6}$$

$$b_2(\nu) = 2(1 + a_2)(\epsilon_1 + E_1) + \epsilon_2 + E_2, \tag{4.7}$$

$$b_3(\nu) = 6(a_2 + a_3)(\epsilon_1 + E_1) + 3(1 + 2a_2)(\epsilon_2 + E_2) + \epsilon_3 + E_3, \tag{4.8}$$

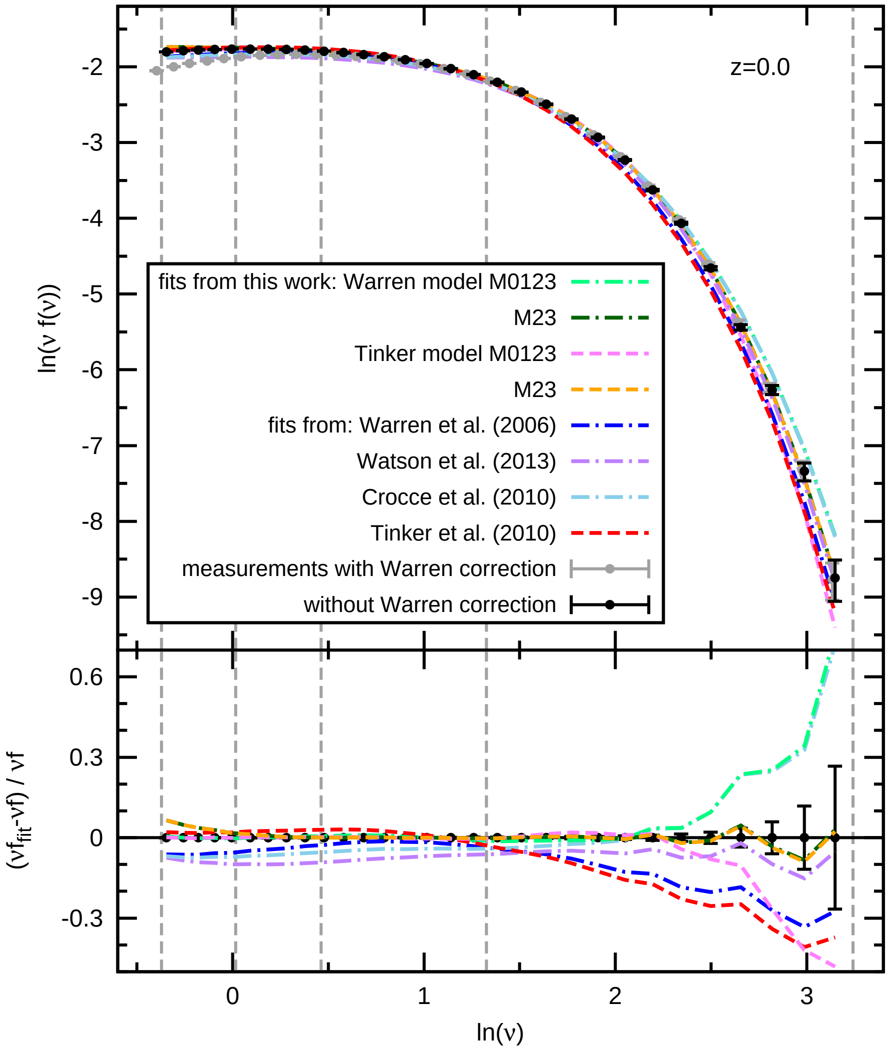

Figure 4.4: *T*op: mass function fits compiled from the literature compared with MICE-GC measurements and fits from this work over the whole mass range (M0123) and the high mass range (M23) at $z = 0.0$. Grey and black symbols show measurements computed with and without Warren correction for halo masses respectively. All fits from this work are based on the latter. *B*ottom: relative deviations between fits and measurements in the same colour coding as the top panel.

$$\epsilon_1 \equiv \frac{c\nu-2d}{\delta_c}$$

$$\epsilon_2 \equiv \frac{c\nu(c\nu-4d-1)+2d(2d-1)}{\delta_c^2}$$

$$\epsilon_3 \equiv \frac{c\nu[(c\nu)^2-6(d+1/2)c\nu+12d^2]-8d^3+12d^2-4d}{\delta_c^3}$$

$$E_1 \equiv \frac{-2a}{\delta_c[(b\nu)^{-a}+1]}$$

$$E_2/E_1 \equiv \frac{-2a+2c\nu-4d+1}{\delta_c}$$

$$E_3/E_1 \equiv \frac{4a^2+12a(d-1/2)+2(2d-1)^2+4d(d-1)-6c\nu(2d+a)+3(c\nu)^2}{\delta_c^2}$$

Table 4.4: Coefficients for computing halo bias parameters from the Tinker et al. (2010) mass function model via equations (4.6), (4.7) and (4.8). a, b, c and d are the free parameters in the Tinker model. Bias predictions for other mass function models can be obtained by using the constrains for the fitting parameters, given in Table 4.1.

where the parameters $a_2 = -17/21$ and $a_3 = 341/567$ are given by the spherical collapse model. $E_1, E_2, E_3, \epsilon_1, \epsilon_2$ and ϵ_3 are computed from the fitted parameters in the mass function models as shown in Table 4.4. Note that the expressions for the non-linear bias parameters, derived from the expressions in Table 4.4, are here presented for the first time for the Tinker model. Applying the parameter constrains from Table 4.1 delivers the equivalent expression for the PS, ST and the Warren models, as well as for our proposed model.

Predictions for b_1, $c_2 \equiv b_2/b_1$ and $c_3 \equiv b_3/b_1$, derived from the Tinker mass function fits at $z = 0.0$, are shown as a function of FoF halo mass in Fig. 4.5. The results are based on mass function fits over the whole mass range (M0123) and fits over the higher mass ranges (M123 and M23). The b_1 predictions for the different fitting ranges agree in the high mass end where the fitting ranges overlap and the mass function fits agree as well (see Fig. 4.2). In the low mass end we find a clear, but relatively weak dependence of the linear bias prediction on the fitting range. In the case of c_2 and c_3 this dependence is stronger and reaches to higher halo masses. This indicates that second- and third-order derivatives of the mass function, used to derive c_2 and c_3 cannot be measured as reliable as first-order derivatives, used to derive b_1. We see the same trends when employing the ST and Warren mass function models as well as for our proposed model, while in these cases the dependence on the fitting range is weaker (see Appendix B.3). We also find a similar behaviour of the bias predictions at $z = 0.5$.

The absolute deviations between bias prediction from the Tinker mass function, fitted over the range M123 and other predictions are shown in Fig. 4.6. These other predictions are based on Tinker and Warren fits over different mass ranges and fits for the same models compiled from the literature. We do not show relative deviations to avoid singularities at the zero crossings of c_2 and c_3. For the linear bias we find absolute deviations between the different predictions of $\Delta b_1 \simeq 0.2$, which roughly corresponds to relative deviations of around 10%. The relative deviations for c_2 and c_3 are around 50%, but can go up to more than 100%. Mass function fits over the high mass range M23 to the Tinker and Warren models deliver almost identical bias predictions, which can be expected since also the fitted

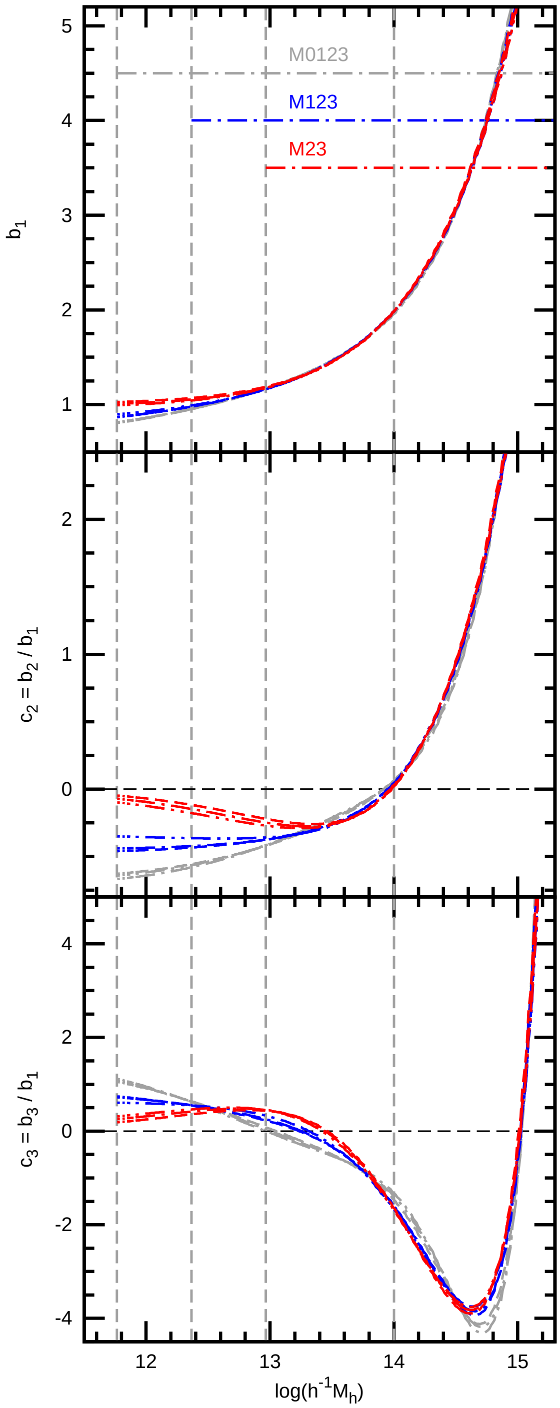

Figure 4.5: Bias parameters b_1, $c_2 \equiv b_2/b_1$ and $c_3 \equiv b_3/b_1$ (top, central and bottom panels respectively), derived from mass function fits of the Tinker model via the PBS approach at $z = 0.0$. Grey lines show results based of mass function fits over the whole mass range M0123, blue and red lines show results from mass function fits which exclude the lowest and the two lowest mass samples (M123 and M23 respectively). Results based on fits to mass function measurements with 20, 30 and 40 bins are shown as dashed, dashed-dotted and dashed-double-dotted lines respectively. Results derived from fits of other mass function models performed in this work are shown in Fig. B.5

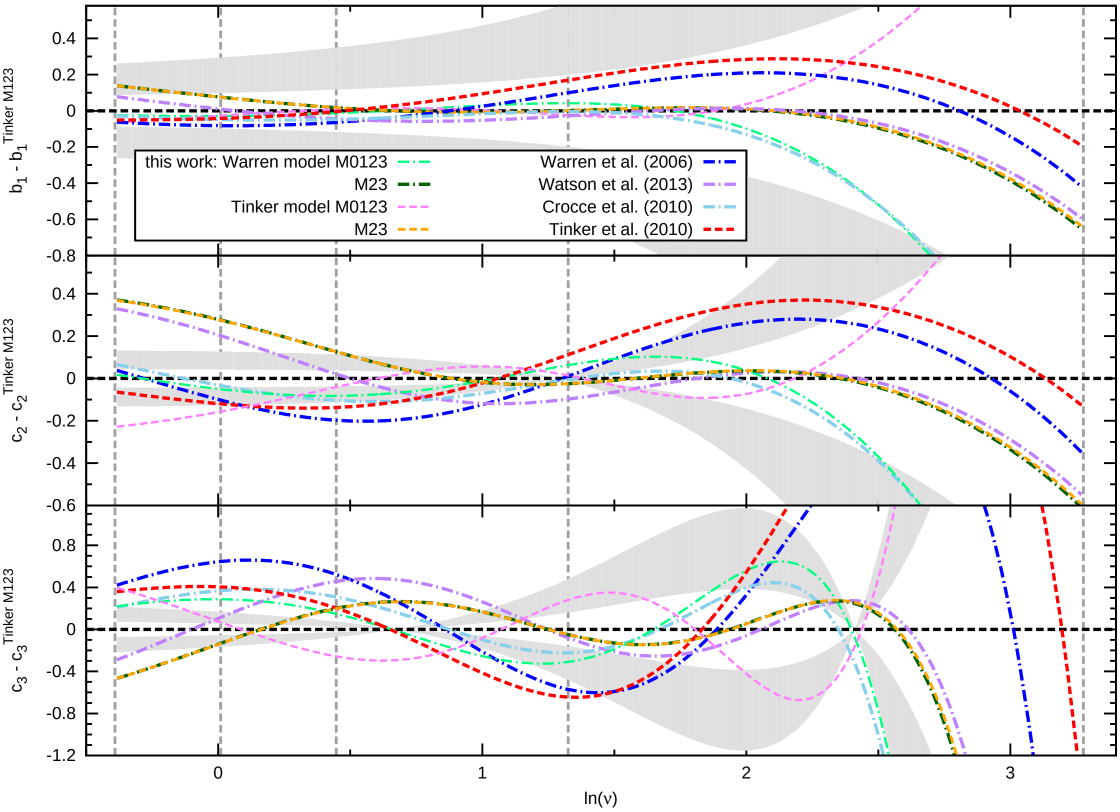

Figure 4.6: Deviations of PBS predictions for b_1, c_2 and c_3 at $z = 0.0$ (top, central and bottom panel respectively) derived from various mass functions fits with respect to the bias from our Tinker mass function fit over the range M123 (shown as blue line in Fig. 4.5) as function of the peak-hight ν. Deviations between $10 - 30\%$ are marked by grey areas. The line color coding is the same as in Fig. 4.4. Vertical dashed lines denote the ν limits of the halo mass samples M0-M3.

parameters are very similar (see Fig. 4.3). In the high mass end these two bias predictions agree with prediction from the fit to the Warren mass function given by Watson et al. (2013). Comparing our results to those of Crocce et al. (2010) we find a reasonable agreement for bias predictions based on the Warren model fitted over the whole mass range M0123.

4.3.1 Universal relation between bias parameters

A universal behaviour of the mass function, as studied in Section 4.2.2, would suggest that the bias parameters, derived from the mass function are universal as well, when they are expressed as a function of peak-hight ν. Our comparison with the literature shows that both, the mass function from different simulations and the bias parameters derived from these mass functions (especially c_2 and c_3) can differ significantly from each other. These disagreements might not only arise from different cosmologies, but also systematic effects, as discussed previously.

We now aim to verify the universality of the relation between the bias parameters. Such a universal behaviour would be useful for reducing uncertainties in linear bias measurements from third-order statistics (e.g. Manera and Gaztañaga, 2011, , Chaper 2). In Fig. 4.7 we

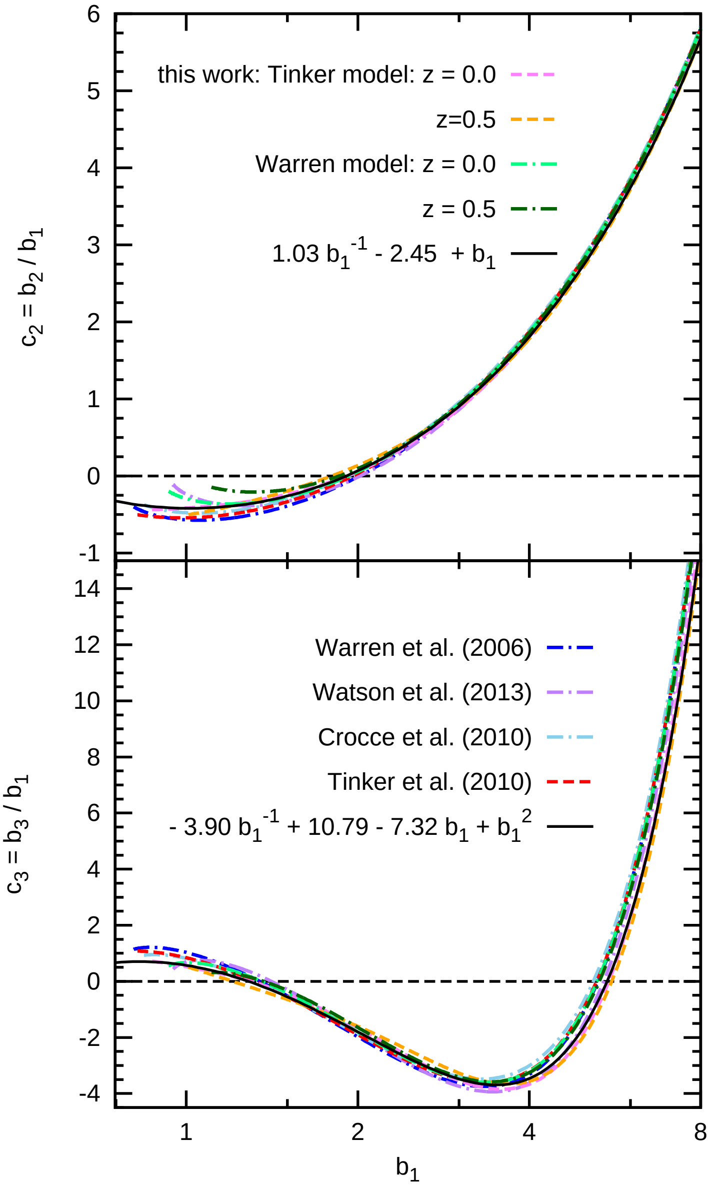

Figure 4.7: Second- and third-order bias parameters, $c_2 \equiv b_2/b_1$ and $c_3 \equiv b_3/b_1$, as a function of the linear bias parameter b_1, predicted from the PBS model (top and bottom panel respectively). Results from this work are based on mass function fits over the mass range M123. Results from the literature are shown in the same colour coding as in Fig. 4.4. Note that the mass function fits from Crocce et al. (2010) and Watson et al. (2013) are based on the Warren model. Black solid lines show polynomials (equation (4.9)), which were fitted to the PBS predictions of the Tinker model, based on MICE-GC mass function fits from this work at $z = 0.0$ (magenta dashed line) with rms per degree of freedom of 0.02 and 0.12 and for c_2 and c_3 respectively.

show the PBS prediction of the second- and third-order bias parameters, c_2 and c_3, as a function of the prediction for the linear bias b_1. We find a 10% agreement for the $b_1 - c_2, c_3$ relations for large values of the linear bias ($b_1 \gtrsim 1.5$). These relations appear to be well described by second- and third-order polynomials in the case of $c_2 \equiv b_2/b_1$ and $c_3 \equiv b_3/b_1$ respectively, with

$$b_N = \sum_{n=0}^{N} \alpha_n b_1^n, \tag{4.9}$$

as we demonstrate in the same figure. This finding can be expected from expressing b_2 and b_3 as functions of b_1 with the PS model (Table 4.1). For this model the parameters α_n can be directly predicted as $(\alpha_0, \alpha_1, \alpha_2) = (0.51, -2.21, 1)$ for $b_2 = b_1 c_2$ and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (-1.49, 8.02, -6.64, 1)$ for $b_3 = b_1 c_3$. However, we find smaller rms values with respect to the Tinker and Warren predictions for the $b_1 - c_2$ and $b_1 - c_3$ relations, when we leave $\alpha_{n<N}$ as free parameters. We show values for α_n from fits to the Tinker predictions in Fig. 4.7.

For predictions based on our fits over the whole mass range, M0123, we find deviation from this universal behaviour, while these results involve the low mass samples which we found to be unreliable previously, possibly due to low resolution and noise is the halo detection. For lower b_1 values the different predictions differ more strongly from each other. However, a weakly universal relation, especially between b_1 and c_2, might already help to improve b_1 constrains from third-oder statistics as these two parameters are usually treated as independent. We will further study this relation with direct measurements of b_1 and c_2 from simulations in a future analysis (Bel, Hoffmann & Gaztañaga in preparation).

4.4 Bias prediction versus measurements from clustering

4.4.1 Comparison with bias from ξ

In the previous section we found that the PBS bias predictions depend on the employed mass function model and the mass range over which the models are fitted. We now aim to verify how the predictions for the linear bias, b_1, in these different cases compare to linear bias measurements from the two-point halo-matter cross-correlation.

To compare the PBS predictions for the linear bias with the measurements from the two-point correlation we calculate the average bias prediction in each of the mass samples M0-M3, weighted with the halo number density $n(m)$,

$$b_N(M) = \frac{\int_{M_{low}}^{M_{up}} b_N(m) n(m) dm}{\int_{M_{low}}^{M_{up}} n(m) dm}. \tag{4.10}$$

M_{low} and M_{up} are the lower and upper limits of each mass sample M, given in Table 2.1. PBS b_1 predictions, based on fits to the Tinker model over the mass range M123, are compared with the $b_\xi(r_{12})$ measurements in the bottom panel of Fig. 2.5. For the high mass samples M2 and M3 we find clear deviations between measurements and predictions as the latter are significantly to low at all scales.

The dependence of these deviations on the mass function model and the mass range in which the models are fitted is shown in Fig. 4.8. Fitting the mass function over the whole mass range, M0123, delivers b_1 predictions which tend to be $1-15\%$, below the measurements, except for the low mass samples M0 and M1 at $z = 0.0$, for which we find up to 5% deviations in the opposite direction. We find the strongest variations between bias predictions from different models when, i) the low mass range at $z = 0.5$ is included in the mass function fitting range, or ii) when the bias is predicted for mass samples which are not within the fitting range (e.g. bias predictions for the mass sample M1, based on fits over the mass range M23). The first case i) might be explained by noise, contaminating the FoF halo detection, which results in the poor mass function fits shown in Fig. 4.2 (see discussion in Section 4.2.2). In this figure we also see that the mass function fits outside the fitting range can strongly differ for different models. This could cause the strong differences in the bias predictions, described above as case ii). We do not see that the deviations between PBS bias predictions and b_ξ measurements decrease when the analysis is restricted to the higher mass range. This is true for both redshifts $z = 0.0$ and $z = 0.5$ and consistent with results from Manera and Gaztañaga (2011).

However, restricting the fitting range to the higher mass range M23 we find a good agreement between the linear bias predictions from different mass functions models at $z = 0.0$ and $z = 0.5$. The fact that the fitting performance strongly differs for the different models (see Fig. 4.3), while all models predict a linear bias with similar deviations to the measurements, suggests that the goodness of the mass function fit is not the only reason for these deviations, as mentioned in the introduction to this chapter. These results line up with reports of Manera et al. (2010), who also find the linear PBS bias prediction to lie below measurements from the power spectrum and two-point correlations, especially at high halo masses. As in our case their result is independent of the employed mass function model and the way it is fitted to the measurements.

Furthermore, these authors investigate if differences between the predictions and measurements are related to the mass definition of haloes. They therefore perform their analysis using FoF groups with different linking lengths, as well as spherical overdensities to define halo masses. In both cases they find that the PBS model underpredicts the linear bias measurements. In fact one could expect that the halo mass should be higher than those of FoF groups in order to match the PBS predictions (since shifting the b_ξ measurements in Fig. 4.8 to higher masses would decrease the deviations between measurements and predictions). However, halo masses defined by spherical overdensities tend to be below those of FoF groups (Tinker et al., 2008). This should lead to higher measurements of the linear bias for spherical overdensities within a given mass range than corresponding measurements for FoF groups, as found by Tinker et al. (2010). The 10% underprediction of linear bias measurements by the PBS model, which we see for high mass haloes in Fig. 4.8, is therefore probably a lower bound. The consideration above also suggests that applying the Warren correction on the FoF masses could increase the differences between the PBS bias predictions and the measurements. Hence, these differences might not only be related to the halo mass definition, but also to assumptions of the PBS model, such as spherical collapse or a local bias relation (e.g. Schmidt et al., 2013; Paranjape et al., 2013). The conclusion, that bias predictions are only weakly dependent on the employed mass function model does not hold for the higher-order bias predictions c_2 and c_3 (see Fig. 4.5, 4.6 and B.5).

Bias ratios

The degeneracy between the growth of matter fluctuations and the bias of observed galaxy samples is one of the largest uncertainties for constrains of cosmological models derived from large-scale structure observations. With estimations of the typical host halo masses of such

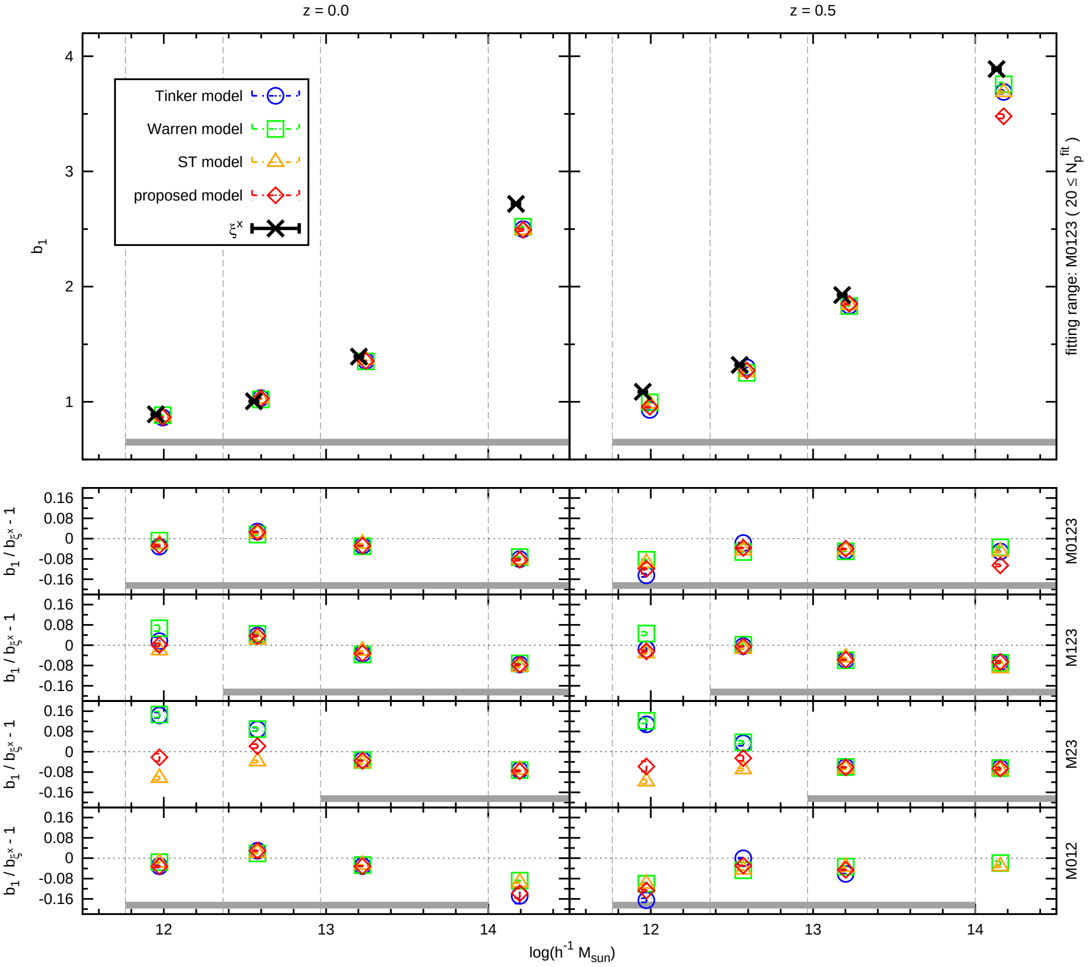

Figure 4.8: *Top*: linear bias parameters b_1 for the halo mass samples M0-M3 in the MICE-GC comoving outputs at redshift $z = 0.0$ (left) and $z = 0.5$ (right) versus the mean halo mass of each sample. Measurements from the two-point halo-matter cross-correlations, $\xi^\times$, via equation (2.14), $b_{\xi^\times}$, are shown as black crosses with 1σ errors. PBS predictions, derived from MICE-GC halo mass function fits from this work are shown as coloured symbols, which are slightly shifted to the left on the mass axis for clarity. The mass range over which the mass function is fitted is marked by a thick grey horizontal line. Error bars for the predictions are standard deviations derived from the seven different mass functions binnings shown in Fig. 4.2. *Bottom*: relative difference between b_ξ and the PBS predictions. The different panels show results for predictions based on mass function fits over the mass ranges M0123, M123, M23 and M012, from the top to the bottom respectively.

galaxy samples the PBS model can be employed to predict the bias of these samples to break the growth-bias degeneracy. Besides the galaxies host halo mass estimation, the inaccuracy of the PBS bias prediction constitutes an additional source of error in this approach. Here we aim to quantify the impact of such inaccuracies on measurement of the linear growth factor. The considered growth measurements are based on the ratio of the correlation functions of galaxy samples at two different redshifts, z_1 and z_2, multiplied with the inverse ratio of the bias of these samples (see e.g. Chapter 2). The bias ratio needs to be estimated or predicted, while its uncertainties propagate linearly into the growth measurements.

In Fig. 4.9 we show the PBS bias ratio predictions for the redshifts $z_1 = 0.0$ and $z_2 = 0.5$ and all combinations of the four halo mass samples M0-M1 at these two redshifts. The predictions are based on fits of the Tinker model to the mass function of the mass range M123, which we found to be reliable at both redshifts previously. We find an overall variation of $5 - 10\%$ for the higher mass range M123, while deviations are stronger when the low mass sample M0 at redshifts $z = 0.0$ is taken into account. This variation is stronger than uncertainties expected from the b_ξ measurements. The strong deviations for the low mass range are expected due to the poor mass function fit including M0 at $z = 0.5$ (see Section 4.2.2). The error in the bias ratio will propagate into $5 - 10\%$ error of the growth factor measurement. This uncertainty is lower than the uncertainties found for growth measurements based on bias ratio estimations from the three-point correlation (see Chapter 2). However, the estimation of the galaxies host-halo mass will introduce additional limitation in breaking the growth-bias degeneracy. Furthermore, the precision of any HOD fitting or mass interpretation from clustering measurements will be affected at similar level.

4.4.2 Comparison with bias from ΔQ

We saw in Section 4.3 that the PBS prediction for the second- and third order bias parameters is more strongly affected by the choice of the fitting range than it is the case for the linear bias. To verify in which case we can trust the prediction best we compare our measurements to those from the combination three-point auto- and cross-correlations (ΔQ), which we studied in Chapter 3. The good agreement between the linear bias from ΔQ and ξ suggests that also the quadratic bias measurements from the ΔQ_{cg} 3.12 are reliable. We thereby use measurements from $36 - 72$ h^{-1}Mpc triangle configurations, while we found in Chapter 3 that the scale dependence of the ΔQ_{cg} is weak. For comparing the measurements from ΔQ_{cg} in the four mass sample M0-M3 we derive the average predictions, weighted with the halo number density weighted with the halo number density $n(m)$ via equation (4.10). The results are shown in Fig. 4.10. The PBS predictions are based on the Tinker model, fitted over the massranges M0123, M123 and M23 (defined in Table 4.2). We see again that the prediction vary strongly in the low mass range, when it is excluded from the fitting range, while the effect is stronger at higher redshift. The variation in the low mass range is comparable with the variation which we find when employing different mass function models (see Fig. 3.7). The overall agreement between the PBS bias parameter predictions with results from clustering is best for the fitting range M123, where the low mass sample with $20 - 80$ particles haloes is excluded. Possible this fitting range constitutes a good compromise between excluding low mass haloes and covering a large mass range.

The c_2 measurements from the three-point cross-correlation enable a verification of the universal relation between b_1 and c_2, which we found in Section 4.3.1. This comparison is shown in Fig. 4.11 using the same measurements as for Q and the same PBS predictions as

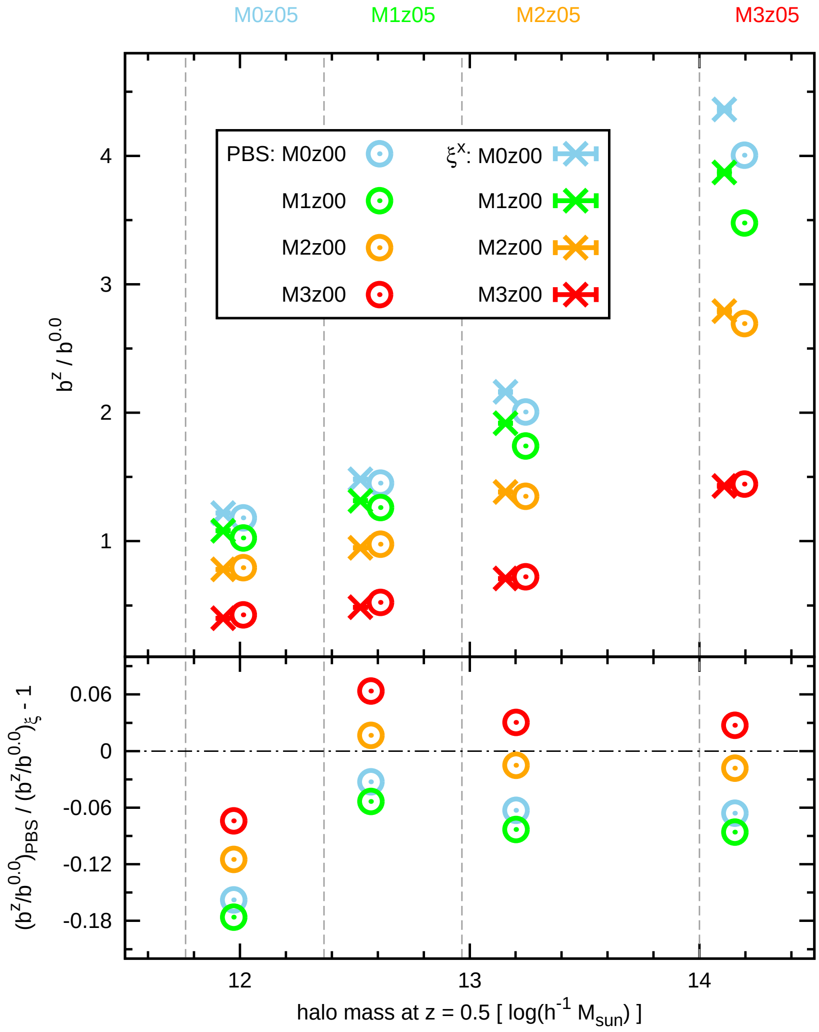

Figure 4.9: *Top*: The ratio of halo bias at the redshifts $z = 0.0$ and $z = 0.5$, which could be used to measure the linear growth factor of matter fluctuations. The PBS predictions, shown as open circles, are based on fits of the Tinker model to mass function measurements over the mass range M123. Measurements from the two-point correlation are shown as crosses with 1σ errors, derived from propagating the error of b_ξ at the two redshifts. Note that errors are smaller than the symbol size. *Bottom*: relative deviations between predictions and measurements.

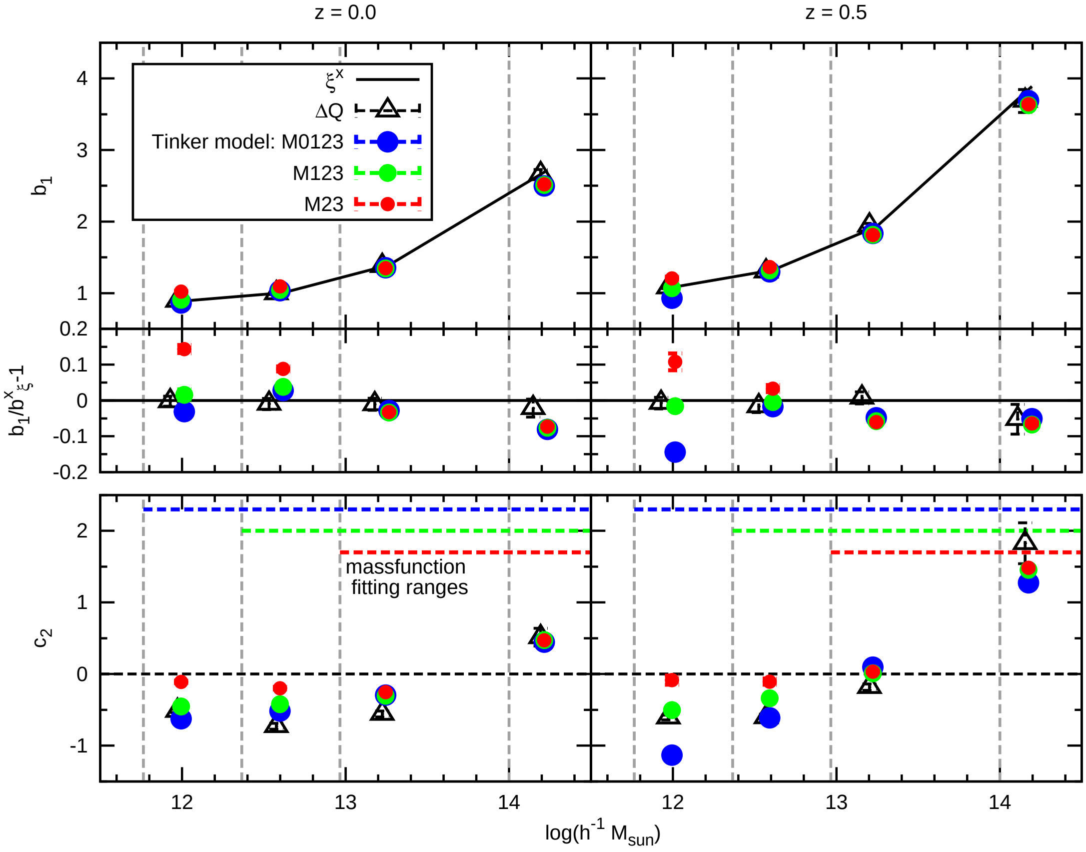

Figure 4.10: Linear and quadratic bias parameters b_1 and c_2 for the halo mass samples M0-M3 in the MICE-GC comoving outputs at redshift $z = 0.0$ (left) and $z = 0.5$ (right) versus the mean halo mass of each sample. Measurements from the two-point halo-matter cross-correlations, ξ (equation (2.14)), are shown as black solid lines. Measurements from the ΔQ method using $36 - 72\ h^{-1}$Mpc triangle configurations (equation (3.12) with 1 errors are shows as black open triangles. The PBS predictions are derived from the Tinker mass function model, fitted over the mass ranges M0123, M123 and M23 (blue green red filled circles respectively). The fitting ranges are shown in the corresponding colour coding as horizontal lines.

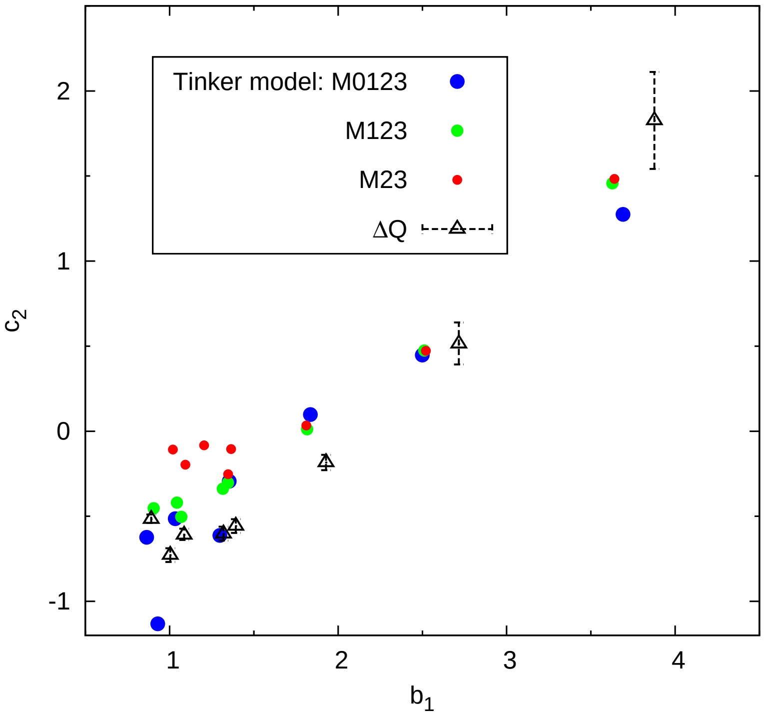

Figure 4.11: Second-order bias parameters, $c_2 \equiv b_2/b_1$, as a function of the linear bias parameter b_1. The prediction from the PBS model and the measurement from ΔQ are the same as in Fig. 4.10.

in Fig. 4.10. Regarding the PBS predictions we find the b_1-c_2 relation to be less tight for high b_1 than in our previous comparison between PBS predictions from the M123 Warren and Tinker fits with results from the literature. This might results from the inclusion of the low mass range in the M0123 fit, which we expect to be unreliable (see Section 4.3). Focusing on the comparison between the predictions from mass function fits over the ranges M123 and M23 we find them again to agree mutually in the high mass range. This this mass range also the ΔQ_{cg} measurements are consistent with a universal b_1-c_2 relation. For $b_1 \simeq 2$ the ΔQ_{cg} measurements differ significantly from the different PBS predictions, while the latter mutually agree. At $b_1 \simeq 1$ the ΔQ_{cg} measuremets are again consistent with the predictions based on M0123 and M123 mass function fits. The predictions differ at $b_1 \simeq 1$ from the measurements when the mass function fit is restricted to the high mass range (M23).

4.5 Summary and Conclusion

We investigated bias predictions from the PBS model, derived via fits to MICE-GC FoF mass functions. The accuracy of this model was tested by comparing its predictions for the linear bias to direct measurements from two-point correlations. In order to verify how the bias predictions are affected by the goodness of the mass function fit we study the performance of four mass function models, fitted over different mass ranges at the redshifts $z = 0.0$ and $z = 0.5$.

Based on the new mass function error and covariance estimator we studied the performance of different mass function models. We thereby show that the models of Press and Schechter (1974), Sheth and Tormen (1999) and Warren et al. (2006) are special cases of the mass function expression suggested by Tinker et al. (2010), as they correspond to certain values of free parameters in the Tinker model (see Table 4.1). This finding motivated us to propose a new model by fixing a different free parameter. The fitting performance of the mass function models, quantified by the minimum χ^2 values per number of analysed mass function bins (*d.o.f.*), shows strong variations among different models and fitting ranges (see Fig. 4.3). All models match the measurements better when the low mass range is excluded from the analysis. This indicates resolution effects, given that we analyse FoF groups with down to 20 particles. We find that the model of Tinker et al. (2010) shows the best overall performance, which can be expected, since it contains the highest number of free parameters. Our proposed model delivers results similar to those from the Tinker model when the whole mass range is analysed, while it has one free parameter less. These two models outperform the model of Warren et al. (2006) for fits over the whole mass range. A restriction to the high mass range ($\geq 2.32\ 10^{12} h^{-1} M_{\odot}$) leads to very similar fitting performance of the Warren et al. (2006) and Tinker et al. (2010) models with minimum $\chi^2/d.o.f.$ values close to unity, while our proposal is slightly worse. Fits to the model of Sheth and Tormen (1999) shows the most significant deviations to the measurements in all cases. These findings are independent of the mass function binning. We find that the inclusion of the covariance into the analysis substantially increases the minimum χ^2 values of the best fits and also has an impact on the best fit parameters. However, our PBS bias predictions are only very weakly affected by the mass function covariance, especially when the higher mass range is analysed, where errors are shot-noise dominated.

The the results described above can be affected by the way the mass function errors and the covariance between different mass function bins are estimated. We therefore conducted a detailed study of these quantities which is presented in Appendix B.2. Given the one MICE-GC realisation, we rely on the internal JK error estimator which we compared to predictions. The comparison reveals that the JK method is in good agreement with the predicted mass function error only in the shot-noise dominated high mass range ($\gtrsim 5\ 10^{14} M_{\odot}$), but overestimates the predictions by up to 80% in the lower mass range, where the errors are dominated by sampling variance. We explain this difference with a wrong scaling relation between sampling variance and sample volume, assumed by the standard JK estimator. By introducing an improved scaling relation, predicted from the linear matter power spectrum, we propose a new mass function error estimator. Deviations between errors of our new estimator and the predictions are less than 10% (see Fig. B.1). The advantage of the new estimator with respect to predictions is that it does not rely on a model for halo bias and the power spectrum normalisation. We plan to extend this approach to error estimations for correlation functions (Hoffmann et al. in preparation).

Verifying the dependence of the mass function on redshift, we find that our FoF mass function measurements show a significant $\lesssim 5\%$ change of the mass function amplitude between the redshifts $z = 0.0$ and $z = 0.5$ when haloes with less than 80 particles are neglected in the analysis (corresponding to a lower halo mass limit of $2.32\ 10^{12} h^{-1} M_{\odot}$, see Fig. 4.1). When including lower masses, the redshift dependence is stronger, possibly because of redshift dependent noise in the low mass FoF detection. In order to investigate a dependence on cosmology we compare our results with mass function fits from the literature. We find differences between 10% in the low mass end and up to $40 - 60\%$ in the high mass end. This is in agreement with studies on departures of the mass function universality. However, numerical and systematic effects, such as halo mass definition, resolution effects or the fitting procedure can contribute to such deviations with similar impact as the differences in cosmology (see Section 4.2.2).

After comparing fits from different mass function models to MICE-GC measurements, we study the bias prediction, derived from the best performing model of Tinker et al. (2010). Note that the non-linear bias parameter expressions for this model are presented in this work for the first time. We find that the bias prediction depends on the mass range over which the model was fitted as the amplitude of the linear bias predictions shows varies by around 10% for different fitting ranges. For the second- and third-order bias parameters the amplitude can vary by more than 50%. This dependences of the bias predictions on the fitting range are comparable with variations obtained when employing fits to other mass function models. Furthermore we find deviations with similar amplitudes in a comparison with bias prediction from mass function fits to other simulations, compiled from the literature (see Fig. 4.6).

A universal behaviour of the mass function would suggest that the bias parameters, derived from the mass function are universal as well. Despite the strong variation among different bias predictions we find a tight universal relation between b_1 and c_2 or c_3 for $b_1 \gtrsim 1.5$ across different simulations and mass function models. For smaller b_1 values, these relations are more dependent on the mass function fit, but still quite tight. The predicted $b_1 - c_2$ relation is in rough agreement with measurements from ΔQ_{cg} (see Chaper 3), especially in the low and high mass range. Using the PS mass function model we derive that the second- and third-order bias parameters b_2 and b_3 can be expressed as second- and third-order polynomials of the linear bias b_1 (see Fig. 4.7). These findings suggests that the linear bias can, at least, constrain the non-linear bias parameter. This could be used to improve the linear bias measurements from third-order statistics.

A common application of the PBS model is to predict the linear bias from clustering. We measured the latter directly from the two-point halo-matter cross-correlation at large scales in the MICE-GC and compare it to the PBS predictions. The comparison was conducted using four different mass samples at the redshifts $z = 0.0$ and $z = 0.5$. Excluding the low mass sample M1 with less than 80 particles per halo from the analysis, which we expect to be affected by noise, we find that the linear bias, predicted from the PBS, model lies $5 - 10\%$ below results from the two-point correlation (see Fig. 4.8). This effect is similar at the redshifts $z = 0.0$ and $z = 0.5$ and independent of the employed mass function model and the way it is fitted to the measurements, confirming previous findings (Manera et al., 2010; Manera and Gaztañaga, 2011). Including the low mass sample delivers similar results, but with a larger scatter among the models. From the analysis in the higher mass ranges we conclude that shortcomings in the performance of the mass function model are not the main reason for the discrepancy between PBS predictions for the linear bias and the corresponding measurements from clustering. An alternative reason for such discrepancies might be given by the overestimation of halo masses by the FoF algorithm, as those tend to be larger than halo masses of spherical overdensities (Tinker et al., 2008). However, from our results in Fig. 4.8 we conclude that shifting the linear bias measurements of FoF halo samples to lower masses would increase the deviations between measurements and PBS predictions. Hence, if FoF masses are overestimations of the halo masses described by the PBS model then the differences between linear bias predictions and measurements, found in this analysis, constitute lower bounds for the inaccuracy of the predictions. This indicates that simple assumptions of the PBS model, such as a local bias model or spherical collapse might limit the accuracy of the linear bias predictions.

We also tested how the bias predictions compare to those derived from ΔQ_{cg}. We see in Fig. 4.10 that the linear bias from ΔQ_{cg} shows in general a better agreement with $b_{\xi}^{\times}$ than the PBS predictions. Regarding the second-order bias we find measurements an prediction in a rough agreement, depending on the mass function fitting range, employed for deriving the predictions. However, its is not clear how reliable the $\Delta Q_c g$ measurements are, since they can suffer from short-coming in the Q_{nloc} predictions. A future comparison with c_2 measurements from different methods will help to verify the reliability of the measurements.

The $5 - 10\%$ deviations between linear bias predictions and measurements will affect at similar level the precision of any HOD fitting or mass interpretation from clustering measurements. We demonstrate the impact of these deviations on growth measurements from two-point correlations. Such measurements are based on the ratio of the linear bias at two different redshifts. Ignoring the unreliable low mass range we find $5 - 10\%$ deviations between PBS predictions for the bias ratios and measurements from the two-point correleation. This inaccuracy would propagate linear into measurements of the linear growth factor, based on PBS bias predictions.

Chapter 5

Discussion and conclusions

The subject of this thesis was to study the accuracy of galaxy bias measurements from three-point correlations as well as bias predictions derived from halo abundance via the peak-background split model. These two methods allow for a determination of galaxy bias independently of the growth, which breaks the growth-bias degeneracy - a major limitation in constraining cosmological models (see Section 2.3).

5.1 Bias predictions from the abundance of haloes

Bias predictions from the peak back-ground split have a broad application in cosmology besides growth measurements, such as cluster mass calibration or HOD modelling (see Section 4.5). These applications rely on the assumption that the bias parameters are predicted accurately as a function of halo mass. Our results demonstrate that these predictions are only accurate at the 10% level, which constitutes a limitation in the precise analysis of incoming large data sets from observations. We find that this inaccuracy cannot be solely attributed to systematic effects, coming from the binning or the fitting of the halo mass function. This finding indicates shortcomings of the peak-back ground split model (see Paranjape et al., 2013; Schmidt et al., 2013). In fact one might argue that the collapse of density peaks into haloes does not only depend on the local background density, as assumed in the peak-background split model, but also on the large-scale velocity field. Hence, the non-local environment might affect the formation of haloes and therefore how their density distribution is biased with respect to the full matter density field. Such effects might partly be covered by the fact that the peak-background split argument is developed in Lagrangian space. Furthermore the collapse of matter fluctuations can also be affected by the density distribution within the haloes. More complications in the comparison between bias predictions and measurements in simulations are introduced by the definition of haloes and the mass which is assigned to them (see. e.g. Knebe et al., 2011, for a comparison between haloes identified by different algorithms in N-body simulations).

5.2 Bias measurements from three-point correlations

Our bias measurements from clustering demonstrate that non-local contributions to the bias function need to be taken into account for accurate bias measurements from three-point correlations. Neglecting non-local bias causes $20-30\%$ and $5-10\%$ overestimation of the linear bias from three-point auto- and cross-correlations, respectively. This overestimation is roughly independent of halo mass and redshift at large scales (see e.g. Fig. 2.9). Therefore, it only has weak effects on the ratio of the linear bias at two different redshifts and consequently also on the measurements of the linear growth factor. Hence, we were able to measure the growth with an accuracy which is comparable to measurements from redshift-space distortions without taking non-local bias into account. However, more accurate measurements (i.e. based on larger halo samples) might require the inclusion of non-local bias.

This inclusion can be undertaken in two different ways. Non-local bias can be isolated from linear bias by combining three-point auto- and cross-correlations or the non-local contributions need to be directly modelled. In this study we focused on the first approach, obtaining accurate measurements of the linear bias. However, the application of this method to observations is not trivial since halo-matter cross-correlations cannot be directly computed. Such statistics may be accessible with cross-correlations between galaxy and lensing maps, which directly probe the full matter field. This approach introduces additional systematic effects, coming for example from galaxy luminosity or shape measurements. The second approach of modelling non-local contributions to the three-point correlation (the $g_2 Q_{nloc}$ term in equation (3.6)) is easier to implement, but also depends on assumptions on the power spectrum from which Q_{nloc} is predicted. Since the errors on the linear bias b_1 from this approach will increase, it would be useful to employ a relation between the nonlocal bias g_2 and b_1 (see e.g. Fig. 3.6). How strongly this $b_1 - g_2$ relation is affected by higher-order bias (local or non-local) or velocity bias at different scales needs to be verified in future investigations. Especially at smaller scales such non-linearities are important, which is a critical aspect when improving the statistical power of large-scale structure data sets.

5.3 Non-local galaxy bias

The non-local contribution to the halo bias function arises from the fact that the collapse of overdensities into haloes depends on the large-scale matter tidal field. However, the tidal field might not only affect the collapse of haloes, but also the properties of the galaxies which form in these haloes. This might happen due to various processes. One possibility might be that the shock heating of the gas during the collapse, which is important for its radiative cooling, depends on the matter tidal field (e.g. Ryu et al., 2003). Furthermore, the ram-pressure stripping of hot gas from haloes falling into clusters depends on the infall velocity (e.g. McCarthy et al., 2008; Font et al., 2008; Close et al., 2013). In addition streams of cold gas, falling into clusters from the surrounding filaments can have a significant impact on the properties of forming galaxies (e.g. Dekel and Birnboim, 2006; Klar and Mücket, 2012), which might also depend on the infall velocity. Also the star formation and AGN activity within galaxies, which are affected by merging events, might depend on the relative velocity of two merging galaxies, which are related to the tidal field. All these different possible relations between galaxy properties and the tidal field suggests that non-local contributions could enter the galaxy bias functions differently than the halo bias function. In order to decrease the impact of such effects it might be convenient to leave the non-local bias g_2 as a free parameter in the bias model instead of modelling it as a function of linear bias b_1.

5.4 Three-point correlations in observations

Our analysis has been performed in three-dimensional real space. However, real space is only observable perpendicular to the line of sight of the observer, since the radial distances of galaxies can not be measured directly. Distance estimates, derived from redshifts of galaxy spectra, are affected by the peculiar motions of the galaxies which distort the apparent galaxy distribution in two different ways. At small scales the thermal motion in gravitational bound galaxy groups and clusters results in an apparent elongation of the galaxy distributions along the line of sight. At large scales the coherent motion of galaxies towards overdensities causes the latter to appear squeezed along the line of sight (Kaiser, 1987). Both effects have an impact on the three-point correlation (Gaztañaga and Scoccimarro, 2005) and therefore need to be taken into account for bias measurements in redshift space. A further complication in the analysis of observational data is that in various modern surveys (e.g. DES or PAU) galaxy spectra are often measured with low precision in order to efficiently observe many objects in large volumes. Since this approach introduces strong uncertainties in galaxy distance estimates such datasets are commonly analysed with angular correlations after binning the galaxies in broad redshift ranges.

Such a two-dimensional analysis has also the advantage of being independent of assumptions on the cosmology, since redshifts are not translated into distances. The growth can therefore be measured directly from angular (or projected) correlations as a function of redshift. This approach is complementary to growth measurements derived from redshift space distortions, which rely on the line-of-sight signal in the two-point correlation. Another possible advantage of a two dimensional analysis is that non-local contributions to the three-point correlation could wash out as a consequence of the line-of-sight integration. This might lead to more precise bias measurements using the local model and could be studied empirically in simulations as well as analytically, following Zheng (2004).

However, recent studies of bias and growth measurements from three-point correlations focus on the 3D analysis (e.g. Marín et al., 2013; Moresco et al., 2014; Guo et al., 2014; Gil-Marín et al., 2014) while literature on the angular analysis is relatively sparse (Zheng, 2004; Buchalter et al., 2000; Frieman and Gaztañaga, 1999; McBride et al., 2011a,b). Estimations of the accuracy and the precision of such bias and growth measurements for modern surveys have not been presented so far, while the volume of these surveys allows for measurements at much larger scales than before, promising a better signal to noise ratios in the linear regime than in previous analysis. It would therefore be interesting to put more effort into investigation of projected three-point correlations.

5.5 Outlook

A possible source of error in our study of non-local bias with the three-point correlations Q is the assumption that the non-local contribution to Q is well described by perturbation theory, i.e. $Q_{nloc} = Q_{nloc}^{PT}$. With the methods employed in this work we are not able to verify if this assumption is true, since we can probe Q_{nloc} only in combination with b_1, c_2 and g_2 via $\Delta Q_{cg} \equiv Q_h - Q^x \simeq (c_2 + g_2 Q_{nloc})/b_1$ (see Section 3.2). In the following we present a simple method to probe g_2 and Q_{nloc} independently from each other. This method is based on the multipole expansion of the three-point correlation, suggested by Szapudi (2004), i.e. $Q = \sum Q^l \hat{P}_l$, where $\hat{P}_l$ are normalised Legendre polynomials and the coefficients

$$Q^l \equiv \langle \hat{P}_l, Q \rangle \tag{5.1}$$

are projections of the of Q on $\hat{P}_l$. Applying this projection on $\Delta Q_{cg} \equiv Q_h - Q^{\times}$ and assuming $\Delta Q_{cg} = (c_2 + g_2 Q_{nloc})/b_1$ delivers for $l > 0$

$$\Delta Q^l_{cg} = \frac{g_2}{b_1} Q^l_{nloc}. \tag{5.2}$$

The bias parameters b_1 and g_2 are independent of the order of the Legendre polynomials l, while Q^l_{nloc} is independent of the analysed halo sample, defined for example by the halo mass range with mean halo mass m. Hence, by combining different measurements of the linear bias (b_ξ) and ΔQ_{cg} from different halo mass samples m_i one can directly measure ratios of the non-local bias as

$$\frac{b_\xi(m_1)\Delta Q^l_{cg}(m_1)}{b_\xi(m_2)\Delta Q^l_{cg}(m_2)} = \frac{g_2(m_1)}{g_2(m_2)}, \tag{5.3}$$

independently of Q_{nloc}. This approach might decrease systematic effects in studies of the difference between $g_2(b_1)$ relations in simulations with different cosmologies. Furthermore one can study the relative contributions of different multipoles to Q_{nloc} among measurements in different simulations as well as theory independently of b_1 and g_2, since

$$\frac{\Delta Q^{l_1}_{cg}(m_1)}{\Delta Q^{l_2}_{cg}(m_1)} = \frac{Q^{l_1}_{nloc}}{Q^{l_2}_{nloc}}. \tag{5.4}$$

The left hand side can of this equation can be directly measured, while the right hand side can be predicted from the power spectrum.

List of Figures

List of Abbreviations

MICE	Marenostrum Institut de Ciències de l'Espai Simulations
MICE-GC	MICE-Grand Challenge
FoF	Friends-of-Friends
PBS	peak-background split
d.o.f.	degrees of freedom
JK	Jack-Knife

Acknowledgements

First of all, I want to thank Enrique Gaztañaga for supervising and supporting me during my studies at the Institut de Ciències de l'Espai (ICE). The work and discussions with him will always stay in my memory as a highly pleasant experience. It was also a great joy to work with the other members of the ICE cosmology group who were always helpful and open minded for all kinds of questions. In particular I want to thank Martin Crocce, Arnau Pujol, Jorge Carretero, Christopher Bonnett, Francisco Castander, Darren Reed and Anne Bauer. Special thanks to Arnau Pujol and Albert Izard for translating the abstract of this thesis into Catalan and for not snoring when we shared hotel rooms at conferences. I also enjoyed a lot the collaboration with researchers from other institutes, in particular Julien Bel, Roman Scoccimarro, Susana Planelles, Alexander Knebe and Noelia Jimenez. During my work at ICE I was financially supported by the FI grand from the Generalitat de Catalunya, the European Commission Marie Curie Initial Training Network CosmoComp (PITN-GA-2009-238356) and the ESP2013-48274-C3-1-P project from the Euclid mission. Having the pleasure to do music with Nikos and Hugo was a great enrichment for my life in Barcelona. The lovely support of my family has been very important for me. They are not only excellent in encouraging me to follow my interests, but also at improving the grammar in this thesis.

Thanks!

Bibliography

Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Phys.Rev.D, 86, 083540

Bardeen J.M., Bond J.R., Kaiser N., Szalay A.S., 1986, ApJ, 304, 15

Barriga J., Gaztañaga E., 2002, MNRAS, 333, 443

Bel J., Marinoni C., 2012, MNRAS, 424, 971

Bernardeau F., 1994, ApJ, 433, 1

Bernardeau F., 1996, A&A, 312, 11

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

Bhattacharya S., Heitmann K., White M., Lukić Z., Wagner C., Habib S., 2011, ApJ, 732, 122

Blake C., et al., 2011, MNRAS, 415, 2876

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

Buchalter A., Kamionkowski M., Jaffe A.H., 2000, ApJ, 530, 36

Cabré A., Gaztañaga E., 2009, MNRAS, 393, 1183

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

Carroll S.M., 2001, Living Reviews in Relativity, 4, 1

Castorina E., Sefusatti E., Sheth R.K., Villaescusa-Navarro F., Viel M., 2014, Journal of Cosmology and Astroparticle Physics, 2, 049

Chan K.C., Scoccimarro R., Sheth R.K., 2012, Phys.Rev.D, 85, 083509

Close J.L., Pittard J.M., Hartquist T.W., Falle S.A.E.G., 2013, MNRAS, 436, 3021

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

Contreras C., et al., 2013, MNRAS, 430, 924

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

Coupon J., et al., 2012, A&A, 542, A5

Courtin J., Rasera Y., Alimi J.M., Corasaniti P.S., Boucher V., Füzfa A., 2011, MNRAS, 410, 1911

Crocce M., Fosalba P., Castander F.J., Gaztañaga E., 2010, MNRAS, 403, 1353

Crocce M., Castander F.J., Gaztanaga E., Fosalba P., Carretero J., 2013, ArXiv e-prints

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

de la Torre S., et al., 2013, A&A, 557, A54

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

Desjacques V., Crocce M., Scoccimarro R., Sheth R.K., 2010, Phys.Rev.D, 82, 103529

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847., pp. 844–847

Eriksen M., Gaztañaga E., 2015, ArXiv e-prints

Feldman H.A., Frieman J.A., Fry J.N., Scoccimarro R., 2001, Physical Review Letters, 86, 1434

Font A.S., et al., 2008, MNRAS, 389, 1619

Fosalba P., Crocce M., Gaztanaga E., Castander F.J., 2013a, ArXiv e-prints

Fosalba P., Gaztanaga E., Castander F.J., Crocce M., 2013b, ArXiv e-prints

Friedmann A., 1922, Zeitschrift für Physik A: Hadrons and Nuclei, 10, 377

Frieman J.A., Gaztañaga E., 1999, ApJL, 521, L83

Frieman J.A., Gaztanaga E., 1994, ApJ, 425, 392

Fry J.N., 1984a, ApJL, 277, L5

Fry J.N., 1984b, ApJ, 279, 499

Fry J.N., 1994, Physical Review Letters, 73, 215

Fry J.N., Gaztanaga E., 1993, ApJ, 413, 447

Gaztañaga E., Lobo J.A., 2001, ApJ, 548, 47

Gaztañaga E., Scoccimarro R., 2005, MNRAS, 361, 824

Gaztañaga E., Fosalba P., Croft R.A.C., 2002, MNRAS, 331, 13

Gaztañaga E., Norberg P., Baugh C.M., Croton D.J., 2005, MNRAS, 364, 620

Gaztanaga E., 1992, ApJL, 398, L17

Gaztanaga E., 1994, MNRAS, 268, 913

Gaztanaga E., Frieman J.A., 1994, ApJL, 437, L13

Geller M.J., Huchra J.P., 1989, Science, 246, 897

Gil-Marín H., Noreña J., Verde L., Percival W.J., Wagner C., Manera M., Schneider D.P., 2014, ArXiv e-prints

Groth E.J., Peebles P.J.E., 1977, ApJ, 217, 385

Guo H., et al., 2014, ArXiv e-prints

Guzzo L., et al., 2008, Nature, 451, 541

Herschel W., 1784, Philosophical Transactions of the Royal Society of London, 74, 437, ArticleType: research-article / Full publication date: 1784 / Copyright 1784 The Royal Society

Hinks A.R., 1911, Monthly Notices of the Royal Astronomical Society, 71, 588

Hu W., Kravtsov A.V., 2003, ApJ, 584, 702

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

Huchra J., Davis M., Latham D., Tonry J., 1983, The Astrophysical Journal Supplement Series, 52, 89

Jenkins A., Frenk C.S., White S.D.M., Colberg J.M., Cole S., Evrard A.E., Couchman H.M.P., Yoshida N., 2001, MNRAS, 321, 372

Jullo E., et al., 2012, ApJ, 750, 37

Kaiser N., 1987, MNRAS, 227, 1

Kamionkowski M., Buchalter A., 1999, ApJ, 514, 7

Klar J.S., Mücket J.P., 2012, MNRAS, 423, 304

Knebe A., et al., 2011, MNRAS, 415, 2293

Kovač K., et al., 2011, ApJ, 731, 102

Lacey C., Cole S., 1994, MNRAS, 271, 676

Lima M., Hu W., 2004, Phys.Rev.D, 70, 043504

Lima M., Hu W., 2005, Phys.Rev.D, 72, 043006

Lue A., Scoccimarro R., Starkman G., 2004, Phys.Rev.D, 69, 044005

Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P.M., 2007, ApJ, 671, 1160

Manera M., Gaztañaga E., 2011, MNRAS, 415, 383

Manera M., Sheth R.K., Scoccimarro R., 2010, MNRAS, 402, 589

Marín F., 2011, ApJ, 737, 97

Marín F.A., et al., 2013, MNRAS, 432, 2654

Marinoni C., et al., 2005, A&A, 442, 801

Marinoni C., et al., 2008, A&A, 487, 7

Martino M.C., Sheth R.K., 2009, MNRAS, 394, 2109

McBride C.K., Connolly A.J., Gardner J.P., Scranton R., Newman J.A., Scoccimarro R., Zehavi I., Schneider D.P., 2011a, ApJ, 726, 13

McBride C.K., Connolly A.J., Gardner J.P., Scranton R., Scoccimarro R., Berlind A.A., Marín F., Schneider D.P., 2011b, ApJ, 739, 85

McCarthy I.G., Frenk C.S., Font A.S., Lacey C.G., Bower R.G., Mitchell N.L., Balogh M.L., Theuns T., 2008, MNRAS, 383, 593

Messier C., 1781, Connoissance des Temps Pour l'Annee Bissexile 1784, p. 227

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

Mo H.J., Jing Y.P., White S.D.M., 1997, MNRAS, 284, 189

More S., Kravtsov A.V., Dalal N., Gottlöber S., 2011, ApJS, 195, 4

Moresco M., Marulli F., Baldi M., Moscardini L., Cimatti A., 2014, MNRAS, 443, 2874

Norberg P., Baugh C.M., Gaztañaga E., Croton D.J., 2009, MNRAS, 396, 19

Pan J., Szapudi I., 2005, MNRAS, 362, 1363

Paranjape A., Sheth R.K., Desjacques V., 2013, MNRAS, 431, 1503

Perlmutter S., et al., 1999, ApJ, 517, 565

Piazza F., Steigerwald H., Marinoni C., 2014, Journal of Cosmology and Astroparticle Physics, 5, 043

Pollack J.E., Smith R.E., Porciani C., 2012, MNRAS, 420, 3469

Press W.H., Schechter P., 1974, ApJ, 187, 425

Pujol A., Gaztañaga E., 2014, MNRAS, 442, 1930

Reed D.S., Bower R., Frenk C.S., Jenkins A., Theuns T., 2007, MNRAS, 374, 2

Reid B.A., et al., 2012, MNRAS, 426, 2719

Riess A.G., et al., 1998, AJ, 116, 1009

Robertson B.E., 2010, ApJ, 713, 1266

Ross N.P., et al., 2007, MNRAS, 381, 573

Ryu D., Kang H., Hallman E., Jones T.W., 2003, ApJ, 593, 599

Saito S., Baldauf T., Vlah Z., Seljak U., Okumura T., McDonald P., 2014, Phys.Rev.D, 90, 123522

Samushia L., Percival W.J., Raccanelli A., 2012, MNRAS, 420, 2102

Schmidt F., Jeong D., Desjacques V., 2013, Phys.Rev.D, 88, 023515

Schmittfull M., Baldauf T., Seljak U., 2015, Phys.Rev.D, 91, 043530

Scoccimarro R., 1998, MNRAS, 299, 1097

Scoccimarro R., Sheth R.K., Hui L., Jain B., 2001, ApJ, 546, 20

Sefusatti E., Crocce M., Pueblas S., Scoccimarro R., 2006, Phys.Rev.D, 74, 023522

Shane C.D., Wirtanen C.A., 1954, The Astronomical Journal, 59, 285

Sheth R.K., Tormen G., 1999, MNRAS, 308, 119

Sheth R.K., Mo H.J., Tormen G., 2001, MNRAS, 323, 1

Sheth R.K., Chan K.C., Scoccimarro R., 2013, Phys.Rev.D, 87, 083002

Simon P., et al., 2013, MNRAS, 430, 2476

Smith R.E., 2012, MNRAS, 426, 531

Smith R.E., Marian L., 2011, MNRAS, 418, 729

Smith R.E., Watts P.I.R., Sheth R.K., 2006, MNRAS, 365, 214

Song Y.S., Percival W.J., 2009, Journal of Cosmology and Astroparticle Physics, 10, 004

Springel V., 2005, MNRAS, 364, 1105

Steigerwald H., Bel J., Marinoni C., 2014, Journal of Cosmology and Astroparticle Physics, 5, 042

Szapudi I., 1998, MNRAS, 300, L35

Szapudi I., 2004, ApJL, 605, L89

Tinker J., Kravtsov A.V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D.E., 2008, ApJ, 688, 709

Tinker J.L., Robertson B.E., Kravtsov A.V., Klypin A., Warren M.S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878

Tojeiro R., et al., 2012, MNRAS, 424, 2339

Valageas P., Clerc N., Pacaud F., Pierre M., 2011, A&A, 536, A95

Verde L., et al., 2002, MNRAS, 335, 432

Wang L., Steinhardt P.J., 1998, ApJ, 508, 483

Warren M.S., Abazajian K., Holz D.E., Teodoro L., 2006, ApJ, 646, 881

Watson W.A., Iliev I.T., D'Aloisio A., Knebe A., Shapiro P.R., Yepes G., 2013, MNRAS, 433, 1230

White M., 2002, ApJS, 143, 241

Wright T., 1750, *An Original Theory or New Hypothesis of the Universe*, American Elsevie, New York, 1971

Wu H.Y., Zentner A.R., Wechsler R.H., 2010, ApJ, 713, 856

Zheng Z., 2004, ApJ, 614, 527

Appendices

Appendix A

A.1 Impact of covariance and resolution in Q on linear bias estimation

The jackknife estimation of the covariance matrix for Q, C_{ij}, measured for different opening angles α_i, is a potential reason for the discrepancy between the linear bias from two- and reduced three-point correlations (b_ξ and b_Q respectively). Studying how strong our bias estimation is affected by the covariance matrix we compare b_Q derived with the jackknife covariance matrix to results measured without taking covariance into account, i.e. by setting $C_{ij} = \delta_{ij}$.

We show the covariance matrixes of Q with (24,48) configurations (see Table A.1 for details) in Fig. B.2. For the low mass sample M0 C_{ij} has a similar shape as results of Gaztañaga and Scoccimarro (2005). The off-diagonal elements are close to unity, which corresponds to Q at intermediate opening angles ($70 - 80$ deg) having covariance with values at large and small angles. For the high mass sample M3 the covariance is dominated by noise.

Examples for how well the fits to equation (3.6) match the measured relation between Q_g and Q_m are shown in Fig.A.2. The fits are shown as coloured line, while their inverse slope corresponds to b_Q and the crossing point with the y-axis marks c_Q/b_Q. Especially for the low mass sample at redshift $z = 0.0$ bias measurements performed without jackknife covariance seem to deliver better fits to the measurements. According to the covariance of Fig. B.2 the fit allows deviations in the intermediate angle that are compensated with correlated deviations at large and small scales. This produces a change in the value of the fitted bias. Whether this change is correct or not depends on whether the covariance is correct or not. For the higher mass samples and for both mass samples at redshift $z = 0.5$ results derived with and without covariance appear to be more similar. In these cases the off-diagonal regions of the covariance matrixes are less pronounced, especially for the high mass sample.

In the same figure we compare these fits to results, expected for a linear bias model with $b_Q = b_\xi$ and c_Q=0. For the low mass sample M0 at $z = 0.5$ we find that the slopes from such a model match neither the measured $Q_g - Q_{dm}$ relations nor the fits to these measurements from equation (3.6). In all other cases differences between the slopes expected from the linear bias model and the measured $Q_g - Q_{dm}$ relations are less obvious.

A comparison between b_ξ and b_Q measured with and without covariance at different scales is given in Fig. A.3. Bias measurements from Q performed without covariance tend to lie closer to the linear bias from the two-point correlation ξ, while the overall trend towards overestimation remains. The fact that b_Q measurements at large scales for low mass samples at $z = 0.0$, measured without covariance, lie very close to the corresponding b_ξ values suggests

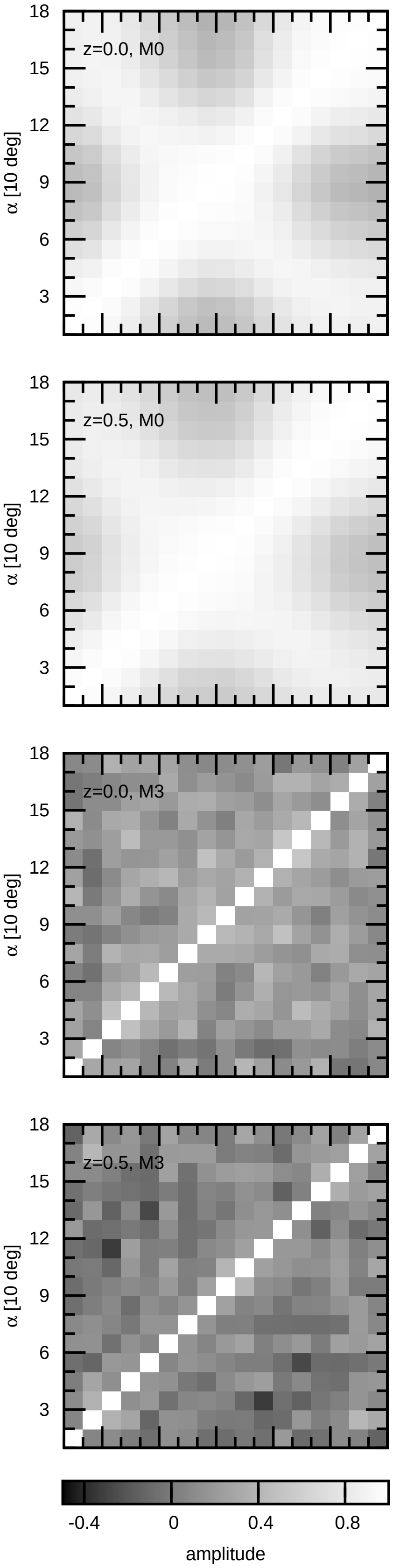

Figure A.1: Normalised covariance C_{ij} between the 18 opening angles of $Q(24, 48, \alpha)$ for the mass samples M0 and M3 at redshifts $z = 0.0$ and $z = 0.5$.

Table A.1: Characteristics of the triangles used to measure Q. r_{12} and r_{13} are the fixed sizes of two triangle legs. n_{12} and n_{13} are the numbers of cubical grid cells per triangle leg. dn is the tolerance for triangle leg sizes in units of grid cells used to define shells for constructing triangles. l_{cell} is the size of the cubical grid cells. The Figures showing results based on the different characteristics are given in the right column.

r_{12} h^{-1}Mpc	r_{13} h^{-1}Mpc	l_{cell} h^{-1}Mpc	n_{12}	n_{13}	dn_{12}	dn_{13}	Fig.
12	24	4	3	6	0.5	0.5	2.6, 2.7, 2.8, A.3
16	32	4	4	8	0.5	0.5	2.8, A.3
24	48	8	3	6	0.5	0.5	2.7, 2.8, 2.9, B.2, A.2, A.3
24	48	4	6	12	0.1	0.1	2.6, A.3
32	64	8	4	8	0.5	0.5	2.8, A.3
32	64	4	8	16	0.1	0.08	A.3
36	72	12	3	6	0.5	0.5	2.8, A.3
36	72	4	12	18	0.05	0.06	A.3

that, besides the jackknife estimation of the covariance, departures from the quadratic bias model for strongly biased halo samples with high mass at high redshift contribute in a non neglectable way to the b_ξ and b_Q discrepancy. Furthermore non-local contributions to the bias model are expected to be strongest for such highly bias samples (Chan et al., 2012). We concluded that the discrepancy between b_Q and b_ξ cannot be only due to uncertainties in the covariance matrix estimation.

In Fig. A.3 we also show examples for b_Q derived using smaller grid cell sizes and thinner shells to construct the triangles. These computations are more expensive than those based on larger grids cells, but closer to the theoretical picture. We find that the bias values change in most cases by around 5 percent. These changes can be driven by changes in the amplitude of Q, but also by changes of the covariance matrix. Especially for high mass samples and at large scales and higher redshift the amplitude of Q becomes more noisy, which can result in larger χ^2 values in the fit. The covariance becomes more diagonal since for smaller grid cells the different triangle opening angles are more independent of each other.

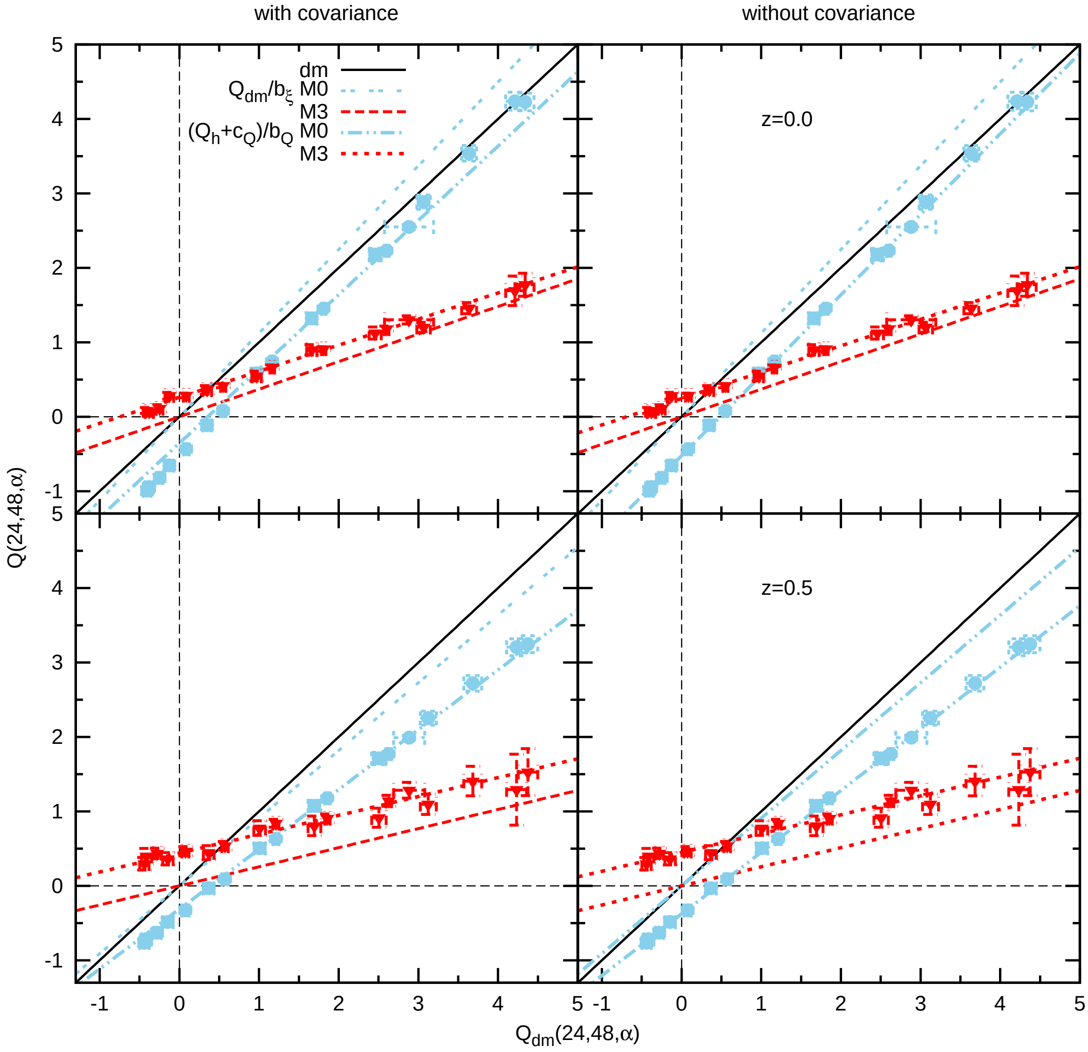

Figure A.2: Q for the high and low galaxy (or halo) mass samples M0 and M3 versus Q for dark matter at the corresponding opening angle. Dotted and dash-dotted lines are χ^2-fits to the Q_g-Q_{dm} relation expected from perturbation theory (equation (3.6)). The fits were performed with and without taking the jackknife covariance of Q_g between different opening angles into account (left and right panel respectively). Long-dashed and double dotted lines show expected results for a linear bias model, using the linear bias measurement from the two-point correlation, b_ξ. Bottom and top panels show results at redshift $z = 0.0$ and $z = 0.5$.

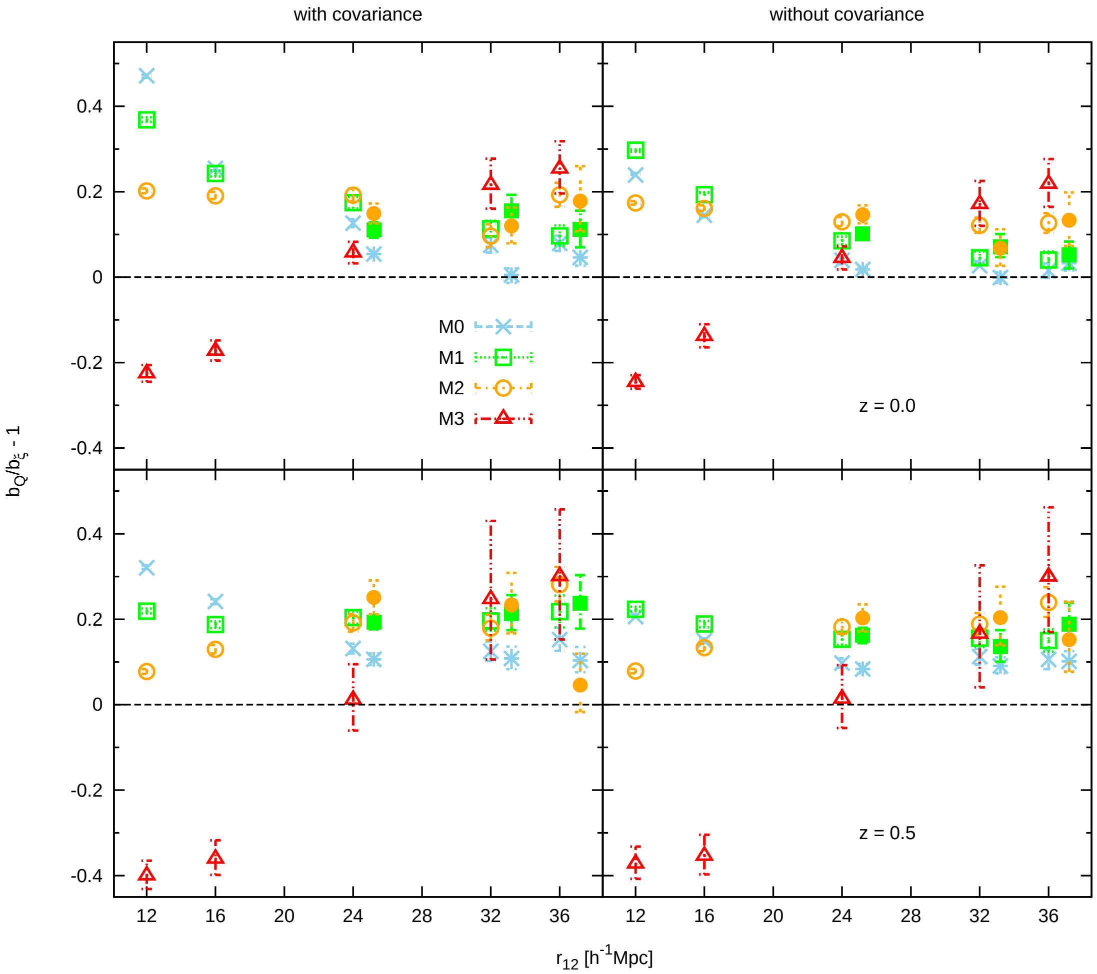

Figure A.3: Relative deviations between the linear bias parameters b_ξ and b_Q derived from two-point and reduced three-point correlations respectively. b_Q was derived using triangles with $r_{13}/r_{12} = 2$ configurations, while the r_{12} values are shown on the x-axis. Left and right panels show, respectively, results obtained with and without taking the jackknife covariance between Q at different opening angles into account. Bottom and top panels show results at redshift $z = 0.0$ and $z = 0.5$ respectively. Open symbols show results from Q using triangles consisting of 3 and 6 grid cells per leg, while the triangle scale is increased by increasing the grid cell size. Closed symbols (slightly shifted to larger scales for clarity) show results from using the smallest available grid cell size of 4 h^{-1}Mpc, while the triangle scale is increased by increasing the number of grid cells per leg. In the latter case also the shells use to contract triangles are chosen to be thinner (see Table A.1 for details). Results for M3 then become very noisy and are therefore not shown.

Appendix B

B.1 Mass function measurements

The mass function measurements are based on a rewritten form of equation 4.1:

$$\nu f(\nu) \equiv \frac{\langle m \rangle}{\bar{\rho}} \frac{dn(m,z)}{dlg\ m} \frac{dlg\ m}{dln\ \nu}, \tag{B.1}$$

where $\langle m \rangle$ is the mean halo mass in each logarithmic mass bin. If the mass bins are chosen to be exactly equal in logarithmic space the mass function amplitude slightly oscillates in the low mass range due to mass resolution effects. Since the errors are smallest in the low mass range this artifact can significantly affect the fits, causing a strong dependence of the fits on the number of mass function bins. Aiming to minimise this mass discreteness effect we determine the minimum and maximum number of particles per halo in each logarithmic mass bin, (N_p^{max} and N_p^{min} respectively. The width of the mass bin in then recalculated as $m_p(N_p^{max} - N_p^{min} + 1)$, where m_p is the particle mass. The value of ν of each bin is calculated from the mean mass of haloes in the bin. The term $\frac{dlg\ m}{dln\ \nu}$ in equation B.1 is derived directly from equation(4.3).

B.2 Covariance

In order to fit the mass function we estimate the errors and the covariance between different mass bins. A direct measurement of these quantities would require a large set of independent realisations of the simulation. Since just one realisation of the MICE-GC simulation was run we estimate the errors and the covariance using the Jack-Knife (hereafter referred to as JK) sampling technique. To validate these estimations we compare the results to theoretical predictions, which we will describe first.

B.2.1 Covariance prediction

Following Crocce et al. (2010) we derive the covariance prediction for the comoving halo number density from the linear bias relation at large scales,

$$n(m, \mathbf{r}) = \bar{n}[1 + b_1(m)\delta_m(\mathbf{r})] + \delta n^{sn}(m, \mathbf{r}), \tag{B.2}$$

where $n(m, \mathbf{r}) = N(m, \mathbf{r})/V_{tot}$ is the number density of haloes with mass m in a volume (in our case the simulation volume) around position $\mathbf{r}$, $\delta_m(\mathbf{r})$ is the matter density contrast in the same volume and $b_1 = \delta_h/\delta_m$ is the linear halo bias factor (as before m refers to the matter density field when it appears as lower index and to the halo mass when it is used as variable). The last term, $\delta n^{sn}(m, \mathbf{r})$, corresponds to noise. We will assume δn^{sn} to be Poisson shot-noise and therefore independent of $\mathbf{r}$. The predictions for the unconditional mass function can be related to those for the halo number density via equation (4.1). For the sake of simplicity the following considerations are based on the latter. The covariance matrix for number densities of haloes in different mass bins i, j is defined as,

$$C_{ij} \equiv \langle \Delta_i \Delta_j \rangle = \frac{1}{N_{samp}} \sum_{k}^{N_{samp}} \Delta_i^k \Delta_j^k, \tag{B.3}$$

where $\Delta_i^k \equiv (n_i^k - \bar{n}_i)$ and $\langle \ldots \rangle$ denotes the average over N_{samp} statistically independent volumes k. Note that the Δ introduced here is not related to the Δ used in equation (4.5) for calculating the χ^2 values of mass function fits. Inserting the expression for the number density n_i of haloes in mass bin i from equation (B.2) leads to

$$C_{ij} = \bar{n}_i \bar{n}_j b_i b_j \langle \delta_m^2 \rangle + \langle \delta_i^{sn} \delta_j^{sn} \rangle. \tag{B.4}$$

The variance of matter fluctuations $\langle \delta_m^2 \rangle = \sigma_m^2(m_{tot})$ can be derived from the power spectrum via equation (4.3), while m_{tot} is the total mass within the volume in which the mass function is measured (in our case the total mass in the simulation). Since this mass corresponds to a very large smoothing radius we can compute σ_m^2 from the linear power spectrum. The sampling variance contribution to the covariance is therefore given by

$$C_{ij}^s = \bar{n}_i \bar{n}_j b_i b_j \sigma_m^2(m_{tot}) \tag{B.5}$$

If the noise term δ^{sn} is Poissonian it averages out when taking the mean over many independent volumes. The contribution of shot-noise to the covariance is then given by

$$C_{ij}^{sn} = \delta_{ij} \frac{\sqrt{\bar{n}_i \bar{n}_j}}{V_{tot}}, \tag{B.6}$$

while here δ_{ij} is the Kronecker delta. Based on these considerations we can write the total covariance as

$$C_{ij} = C_{ij}^s + C_{ij}^{sn}. \tag{B.7}$$

A more formal derivation for this relation is given by Smith and Marian (2011), see also Robertson (2010); Valageas et al. (2011); Smith (2012). The diagonal elements of the covariance matrix correspond to the predictions for the mass function variance,

$$\sigma_i^2 = C_{ii} \tag{B.8}$$

as given by Crocce et al. (2010). For fitting the mass function we work with the normalised covariance $\hat{C}_{ij} \equiv C_{ij}/(\sigma_i \sigma_j)$ and differences normalised to σ_i. (note that here σ_i referes to the variance of the mass function in the mass bin i and not to the variance of the matter field, σ_m).

B.2.2 Jack-Knife estimation of covariance

For mass function fits in observations the covariance prediction is of limited use since it requires knowledge about the bias and the power spectrum in advance. This problem might be solved with an iterative approach for the fit, starting from an initial guess for the power spectrum and the linear bias factor. Another possibility to obtain the covariance without knowledge of the bias and the power spectrum is to estimate it with the JK sampling technique. Testing this approach we construct N_{JK} JK samples by subtracting cubical sub-volumes (hereafter referred to as JK cells) with the size V_{tot}/N_{JK} from the total simulation volume V_{tot}. The basic assumption of the JK approach is that the error scales with the size of the subtracted volume (e.g. Norberg et al., 2009). We follow the common approach by rescaling the covariance with the factor $(N_{JK}-1)$, which leads to

$$C_{ij}^{JK} \equiv (N_{JK}-1)\langle \Delta_i \Delta_j \rangle = \frac{N_{JK}-1}{N_{JK}} \sum_{k}^{N_{JK}} \Delta_i^k \Delta_j^k. \tag{B.9}$$

Again $\Delta_i^k = (n_i^k - \bar{n}_i)$, but now $\langle \ldots \rangle$ is the average over the different JK samples k, n_i^k is the comoving number density of haloes in the mass bin i in each JK sample and $\bar{n}_i$ is the corresponding halo number density in the whole simulation volume. Note that the rescaling factor, $(N_{JK}-1)/N_{JK}$, is only weakly justified and can be improved, as we show in Section B.2.4. As in the case of the predictions the diagonal elements of C_{ij}^{JK} are the JK estimation for the variance ($\sigma_i^2 = C_{ii}$) and we normalise $\hat{C}_{ij}^{JK} \equiv C_{ij}^{JK}/(\sigma_i^{JK}\sigma_j^{JK})$. Note that we can use the JK approach for studying the covariance between low and high mass bins because of the large mass range of the MICE-GC simulation. This analysis would not be possible using nested boxes, where the different mass ranges are covered by different realisations with different box sizes (e.g. Warren et al., 2006; Crocce et al., 2010; Tinker et al., 2010).

B.2.3 Covariance prediction versus Jack-Knife estimation

A comparison between the error prediction from equation (B.7) and the corresponding JK estimation from equation equation (B.9) (with $\sigma_i^2 = C_{ii}$) is shown in Fig. B.1 for the redshift $z = 0.0$. The error predictions are based on linear bias predictions from mass function fits to the Tinker model over the whole mass range, for which we expect uncertainties of around 10% (see Section 4.3). From the prediction we expect the error to be dominated by sampling variance in the low mass end and by shot-noise in the high mass end. At halo masses of $M_h \simeq 2\ 10^{13} M_\odot$ both sources are predicted to contribute equally to the total error. The JK error estimation is in good agreement with the predictions in the high mass end ($M_h \gtrsim 5\ 10^{14} M_\odot$). This indicates that the JK method is working well for different JK cell volumes when the error is dominated by shot-noise. Furthermore, the shot-noise is well described by a Poisson distribution. At halo masses lower than $5\ 10^{14} M_\odot$ we find the JK error to be up to 80% higher than the prediction.

This overestimation is consistent with results reported by Crocce et al. (2010) using the same simulation box size as the MICE-GC, while for smaller simulation boxes they find the JK error to be lower than the prediction. The fact that the overestimation of the JK error in the low mass end is larger for smaller JK cells indicates that the JK assumption of a linear relation between errors and volume is inadequate when sampling variance is the dominating source for error. However, increasing the size of the JK cells results in a smaller number of samples and therefore a stronger noise on the estimated error. In Fig. B.1 we also show a new JK error estimation, which is in good agreement with the predictions at all masses. This new JK error is based on an improved scaling between the sampling variance in a JK cell and in the whole simulation box using the linear power spectrum, as explained in Section B.2.4.

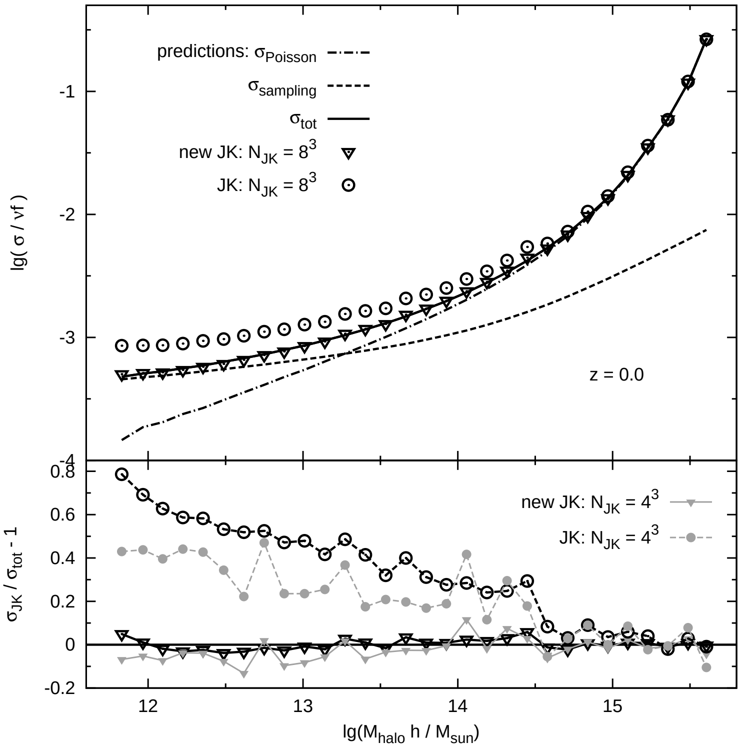

Figure B.1: *Top*: mass function error, displayed as standard deviation divided by the mass function amplitude. Black lines show predictions for the sampling variance and the Poisson shot-noise contribution together with the resulting total error, derived from the equations (B.5) - (B.7) (with $\sigma_i^2 = C_{ii}$, dashed, dashed-dotted and solid respectively). Open circles show the standard JK error estimation, open triangles show the errors from a new JK estimator, derived from the equations (B.9) and (B.14) respectively. Both estimations are based on 8^3 JK cells. *Bottom*: relative deviations between the total predicted error and the new and the standard JK estimations. The symbol types corresponds to those the top panel. Results derived from 4^3 cubical JK cells are shown as grey solid symbols.

In Fig. B.2 we compare the normalised covariance of the mass function between the mass bins i and j, predicted via equation (B.7) with the JK estimation from equation (B.9) using 8^3 JK samples at $z = 0.0$. The shape of the covariance is in good agreement with results from Smith and Marian (2011). The low mass bins are highly correlated because of sampling variance, while high mass bins are uncorrelated as their errors are dominated by shot-noise. For the comparison of the variances we find a reasonable agreement between the prediction and the JK estimations, especially in the high mass end. In the low mass end the covariance seems to be overestimated by the JK approach, while the new JK method reproduces the prediction well.

We show a more detailed comparison of the covariance amplitudes in Fig. B.3, fixing one mass bin i and varying the second mass bin j. For 8^3 JK cells we find the normalised JK covariance amplitudes to be higher than the predictions with differences of up to 0.3. Using larger JK cells this overestimation slightly decreases, while results become more noisy. Again the improved estimation is in better agreement with the prediction. We have verified that our conclusions also hold for redshift $z = 0.5$.

B.2.4 Improved Jack-Knife estimator

We now aim to understand the disagreement between the predicted mass function error and the corresponding JK estimation in order to improve the latter. The N_{JK} JK samples are constructed by subtracting haloes in JK cells of the size V_{tot}/N_{JK} from the total halo distribution. The number of haloes in the remaining JK sample is then given by $N^h_{JK} \equiv N^h_{tot} - N^h_{JKcell}$. The volume of a JK sample is given by $V_{JK} \equiv V_{tot} - V_{JKcell}$. From the definition of the number density, ($n \equiv N^h/V$), and the deviation from the mean over the total volume $\Delta \equiv n - \bar{n}$ one can derive $\Delta_{JK} = -(N_{JK} - 1)^{-1}\Delta(V_{JKcell})$. Note that this relation also holds for the number density contrast, $\delta_{JK} = -(N_{JK} - 1)^{-1}\delta(V_{JKcell})$. Hence, subtracting an overdense JK cell from the total volume generates a slightly underdense JK sample. This result leads to a relation between the variances of the number density, $\sigma^2 \equiv \langle\Delta^2\rangle$, in the JK cells, and the corresponding variance for the JK samples,

$$\sigma^2(V_{JKcell}) = (N_{JK} - 1)^2\langle\Delta^2_{JK}\rangle. \tag{B.10}$$

As for equation (B.3) $\langle\ldots\rangle$ denotes the ensemble average. The variance of the JK samples is therefore simply related to the variance at the scales of the JK cells. Note that $\langle\Delta^2_{JK}\rangle$ is not the variance at the scale of the JK sample volume, $\sigma^2(V_{JK})$, since the JK samples are not independent from each other.

From the linear bias model we assume that the variance of the halo number density results from shot-noise (σ^2_{sn}) and sampling variance (σ^2_s), as explained in Section B.2.2. The latter contribution to the total variance of the JK cells, measured via equation (B.10), is therefore given by

$$\sigma^2_s(V_{JKcell}) = (N_{JK} - 1)^2\langle\Delta^2_{JK}\rangle - \sigma^2_{sn}(V_{JKcell}). \tag{B.11}$$

Since $n/V_{JKcell} = N_{JK}n/V_{tot}$, the shot-noise for JK cells is related to the shot-noise of the whole box as $\sigma^2_{sn}(V_{JKcell}) = N_{JK}\sigma^2_{sn}(V_{tot})$. To obtain the sampling variance at the scale of the simulation box, $\sigma^2_s(V_{tot})$, we multiply $\sigma^2_s(V_{JKcell})$ with a rescaling factor

$$r_\sigma \equiv \sigma^2_s(V_{tot})/\sigma^2_s(V_{JKcell}), \tag{B.12}$$

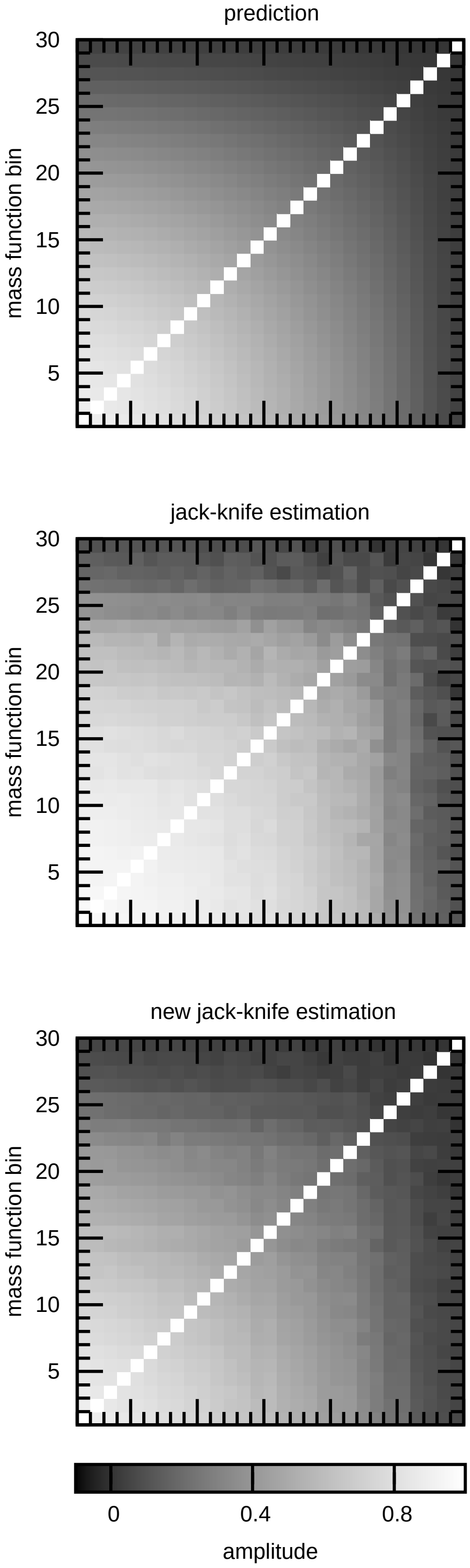

Figure B.2: Normalised covariance between mass function bins at redshift $z = 0.0$, derived from predictions (top, equations (B.5) - (B.7)), the standard JK estimator (center, equation(B.9)) and the new JK estimator (bottom, equation (B.14)). The estimations are based on 8^3 JK cells.

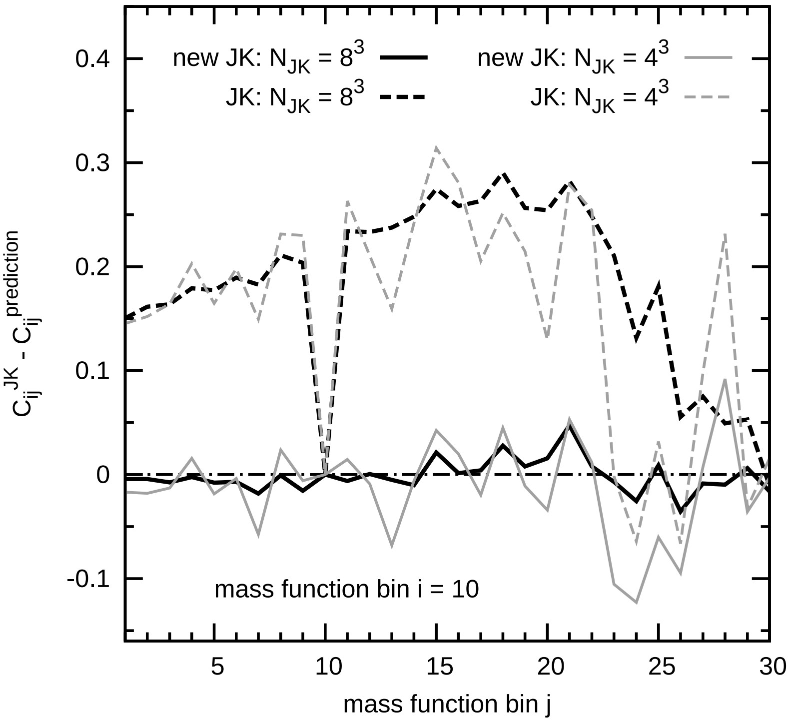

Figure B.3: Deviations between the standard and the new JK covariance estimation (solid and dashed lines respectively) and the prediction, fixing one bin to $i = 10$ and varying the other. We find similar results for different values of i. Black and grey lines show results based on 8^3 and 4^3 JK cells respectively.

which can be predicted from the linear matter power spectrum. This prediction is based on the assumption that, at large smoothing scales, the sampling variance of the halo number density is related to the dark matter variance by the linear bias factor, $\sigma_s^2 = b_1 \sigma_m^2$. Since b_1 is constant at large scales (see Fig. 2.5), it cancels out in the rescaling factor, hence $r_\sigma^h = r_\sigma^m = r_\sigma$. The prediction is then based on $\sigma_m(V)$, computed from the linear matter power spectrum via equation (4.3). We can now write the expression for the sampling variance of the simulation box, based on equation (B.11) in the general case of the covariance

$$C_{ij}^s(V_{tot}) = r_\sigma[(N_{JK}-1)^2\langle\Delta_i\Delta_j\rangle - N_{JK}C_{ij}^{sn}(V_{tot})]. \tag{B.13}$$

Note that we have now dropped the index JK in Δ for simplicity and to be consistent with equation (B.9) for the standard JK estimator. As in the latter equation the lower indices refer to the mass bins i and j. With the Poisson shot-noise term from equation (B.6), the total covariance is then given by equation (B.7) as $C_{ij} = C_{ij}^s + C_{ij}^{sn}$. The resulting expression constitutes a new error estimation for the mass function which combines direct measurements of sampling variance via the JK sampling with predictions for the rescaling factor and the shot-noise. This new estimator can be written more explicitly as

$$C_{ij}^{newJK} = r_\sigma(N_{JK}-1)^2\langle\Delta_i\Delta_j\rangle + \delta_{ij}\frac{\sqrt{n_i n_j}}{V_{tot}}(1 - r_\sigma N_{JK}). \tag{B.14}$$

As before the diagonal elements correspond to the variance, $\sigma_i^2 = C_{ii}$. Note that for Poisson shot-noise dominated errors the sampling variance can be approximated as $C_{ij}^s(V_{tot}) \simeq 0$ and the new estimator reduces to the shot-noise term, $C_{ij}^{newJK} \simeq \delta_{ij}\frac{\sqrt{n_i n_j}}{V_{tot}}$. In this case we derive from equation (B.13) that $(N_{JK}-1)^2\langle\Delta_i\Delta_j\rangle \simeq N_{JK}C_{ij}^{sn}(V_{tot})$. For large numbers of JK samples ($N_{JK} \simeq N_{JK}-1$) this expression is equivalent to $(N_{JK}-1)\langle\Delta_i\Delta_j\rangle \simeq \delta_{ij}\frac{\sqrt{n_i n_j}}{V_{tot}}$. The left hand side of this relation is the standard JK estimator. This consideration explains the good agreement between standard JK estimator with the improved JK estimator and the predictions at high masses, where the errors and the covariance are shot-noise dominated (Fig. B.1, B.2 and B.3). In the low mass range our new method is in much better agreement with the predictions than the standard JK error estimator (B.9). This can be understood with the following consideration. For large numbers of JK samples ($N_{JK} \simeq N_{JK}-1$) the new estimator corresponds to the standard JK estimator if $r_\sigma \equiv \sigma_m^2(V_{tot})/\sigma_m^2(V_{JKcell}) = 1/N_{JK}$. Since $V_{tot} = N_{JK}V_{JKcell}$, this condition is equivalent to $V_{tot}\sigma_m^2(V_{tot}) = V_{JKcell}\sigma_m^2(V_{JKcell})$. The JK approach can therefore be described as the assumption that $\sigma_m(V) \sim V^{-1/2}$ for large N_{JK}. We show $\sigma_m(V)$, computed from the linear power spectrum via equation (4.3) in Fig. B.4. The JK assumption is in a clear disagreement with the prediction which causes a too high r_σ and therefore an overestimation of sampling variances at the scale of the simulation box for the standard JK assumption.

The advantage of the new JK estimation with respect to the prediction is that it does not require knowledge of the halo bias. Furthermore, this new approach is independent of the normalisation of the power spectrum as it cancels out in the rescaling factor (equation (B.12)). However, the large scale power spectrum still needs to be known for accurate rescaling of the sampling variance via $\sigma_m(V)$. For simulations the linear power spectrum is given. In this case the new method can be used instead of running several realisations for deriving mass function errors and covariances. In observations the large scale power spectrum can only be assumed. However, with such an assumption the accuracy of the error estimation can still be

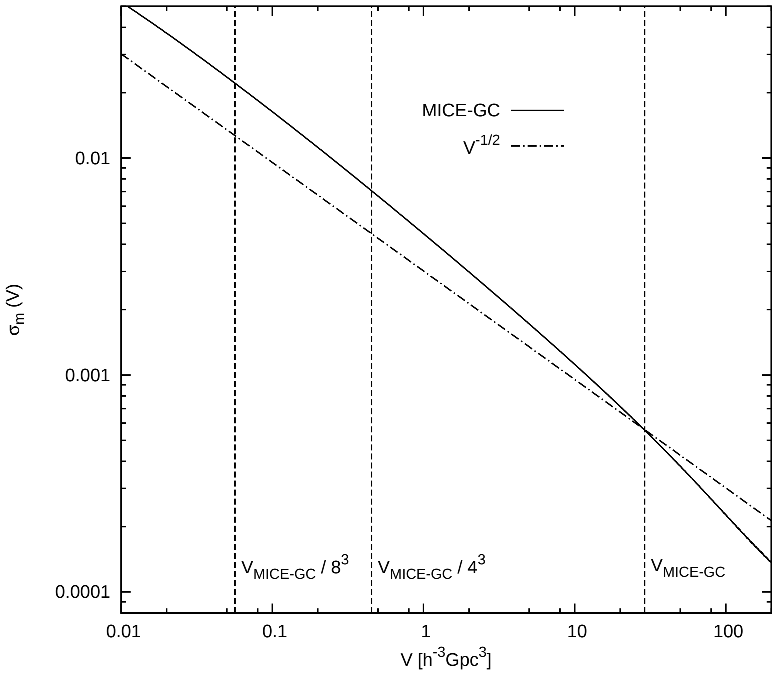

Figure B.4: The standard deviation of matter fluctuations, σ_m, predicted from the linear MICE-GC power spectrum via equation (4.3) as a function of the volume, V, of the spherical top hat smoothing window (solid line). The dash-dotted line shows the $\sigma_m(V)$ relation which corresponds to the standard JK estimator (with an arbitrary normalisation, chosen to coincidence with the predictions at the MICE-GC simulation volume V). The volumes of the simulation box and the JK cells are shown as vertical dashed lines.

improved with respect to the standard JK method, which also implies the strong assumption of $\sigma(V) \sim V^{-1/2}$. The advantage of the new JK estimation with respect to using independent subvolumes for the error estimation is that the JK samples cover larger volumes with larger average numbers of massive haloes. The covariance between the low- and high mass end of the mass function is therefore better sampled by JK samples than subvolumes. In the subsequent analysis we will employ our new method for the error and covariance estimation using $N_{JK} = 8^3$ samples.

B.3 PBS bias predictions

We show in Fig. B.5 the PBS bias predictions based on the mass function models studied in this analysis. The different predictions are based on fits over the four mass ranges M0123, M123, M23 and M012, defined in Table 4.2. The figure is analogous to Figure 4.5. We find that the linear bias parameter b_1 is less sensitive to the mass function model and the mass function fitting range than the non-linear bias parameters $c_2 \equiv b_2/b_1$ and $c_3 \equiv b_3/b_1$. I addition to Figure 4.5 we show that the bias predictions become unstable when the low mass sample, M1, is included in the analysis. Furthermore we show bias predictions, based on mass function fits over the range M123, which were derived neglecting the covariance between different mass function bins. We find that the for this example the mass function covariance has a smaller

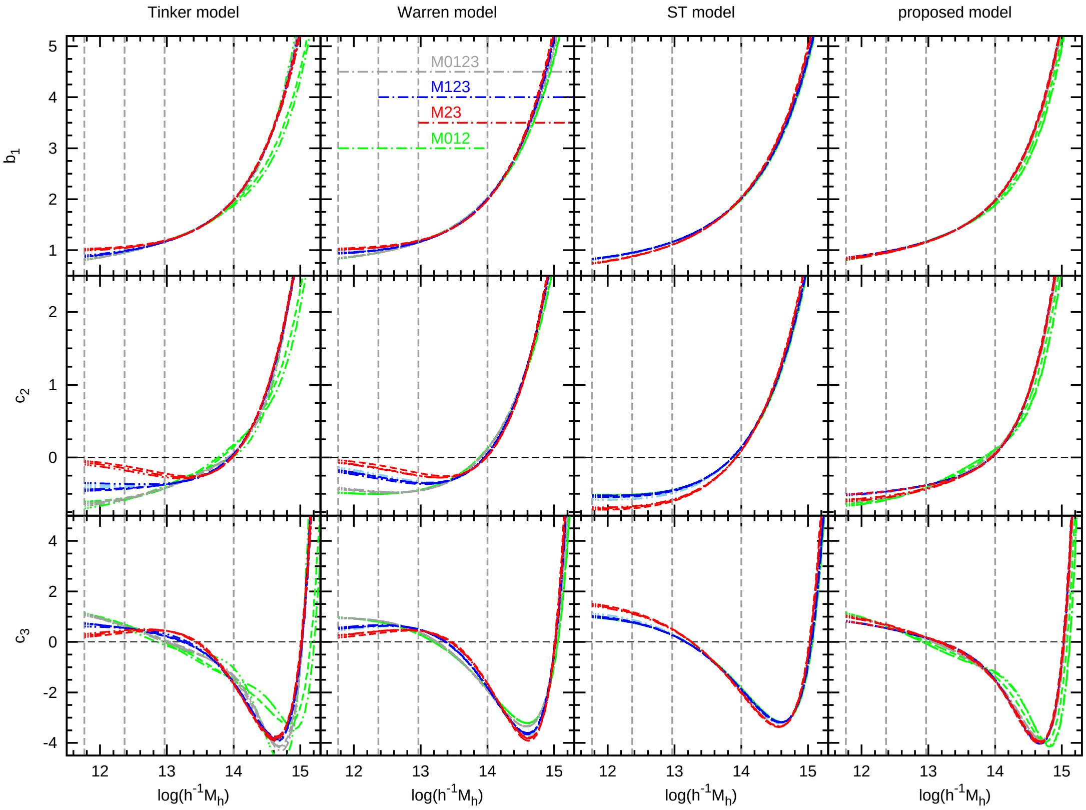

Figure B.5: Same as Figure 4.5, but for all mass function models analysed in this work. In addition we show bias predictions based on mass function fits over the whole MICE-GC mass range (M0123). For predictions from fits over the mass range M123 we show in addition results derived without taking the covariance in the measurements between different mass function bins into account as light blue lines. Note that these lines are covered by other results in most cases.

.

impact on the bias prediction than the choice of the mass function model, or the mass function fitting range. We expect the impact of the mass function covariance to increase, when low mass samples are included in the fit. However, the low mass range is hard to access for analysis of halo abundance in the MICE-GC, do to resolution effects.