New techniques for the analysis of the large scale structure of the Universe

Héctor Gil Marín

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

New techniques for the analysis of the large scale structure of the Universe

Héctor Gil Marín

Doctorat en Física
Institut de Ciències de l’Espai (IEEC-CSIC)
Facultat de Física, Universitat de Barcelona
Advisor: Raúl Jiménez
Tutor: Jaume Garriga

May 2012

Agraïments

M'agradaria agrair als meus supervisors Raúl Jiménez i Licia Verde per les incomptables recomanacions, suggeriments i ajudes en l'elaboració de tots els apartats d'aquesta tesi.

També he d'agrair a Alan Heavens per les seves contribucions en el Capítol 3, a Christian Wagner per les ajudes en l'elaboració dels treballs dels Capítols 3 i 5, a Wayne Hu i a Fabian Schmidt per l'ajuda prestada tant en el plantejament com en els resultats del treball del Capítol 4.

També vull agrair al Royal Observatory of Edinburgh, al Kavli Institute for Cosmological Physics (University of Chicago) i al Argelander-Institut für Astronomie (University of Bonn) per l'hospitalitat en els mesos que vaig passar treballant allà.

Agraeixo als membres del grup de cosmologia de l'Institut de Ciències del Cosmos de la Universitat de Barcelona: Ben, Beth, Christian, Jorge i Roland, per totes les discussions sobre cosmologia que han ajudat a fer possible que aquesta tesi fos millor.

Agraeixo també als meus companys de la Universitat de Barcelona: Andreu, Benito, Carme, Dani, Felipe, Laura, Javi, Jordi, Josep Maria, Maria, Neus, Pau, Pere, Rosa i Sinue per les incomptables hores que he passat amb ells durant l'elaboració dels treballs d'aquesta tesi.

Agraeixo a José Ramón (JR) per ajudar-me a fer tots els tràmits necessaris per dipositar aquesta tesi.

Finalment agraeixo als meus pares pel seu suport incondicional durant tot el temps (i més) que ha durat la realització d'aquesta tesi.

Resum de la Tesi

La cosmologia és la disciplina que estudia l'Univers com a conjunt. L'objectiu és indagar i entendre l'origen, l'evolució, l'estructura i el destí final del cosmos, així com les lleis que el regeixen. Actualment la cosmologia es fonamenta en la teoria del Big Bang, que engloba l'astronomia observacional i la física de partícules. A cosmologia, el terme d'estructura a gran escala es refereix a la caracterització de la distribució de matèria i radiació a escales típicament superiors a 10Mpc (desenes de milions d'anys llum). Les missions científiques de mapeig i cartografiat del cel han proporcionat informació essencial sobre el contingut i les propietats d'aquesta estructura. S'ha determinat que l'organització de l'Univers a aquestes escales segueix un model jeràrquic amb una organització superior de supercúmuls i filaments. A escales superiors no s'ha trobat evidència de cap estructura continuada. A aquest fenomen se'l coneix com el Final de la Grandesa.

Una de les troballes mes recents i sorprenents en cosmologia és l'expansió accelerada de l'Univers. Des de principis del segle XX les observacions de galàxies llunyanes havien demostrat que l'Univers s'expandia: les galàxies s'allunyen les unes de les altres a una velocitat proporcional a la distància que les separa. Com més gran és la distància, més alta és la velocitat d'allunyament. Això va suposar una revolució perquè en aquell moment la visió més estesa entre la comunitat científica era la d'un Univers estàtic i infinitament vell. El concepte d'expansió va implicar que les galàxies de l'Univers havien estat més aprop al passat i per tant tot apuntava a un origen de l'Univers que es remuntava a uns 13000 milions d'anys enrere. Tot i així, el sentit comú apuntava a que aquesta expansió havia de ser desaccelerada. A escales cosmològiques, la força gravitatòria és l'única de les 4 forces fonamentals de la física que domina. Dues d'aquestes forces, la força nuclear forta i feble, tenen un rang efectiu d'escales subatòmiques. Només la força gravitatòria i electromagnètica poden actuar a escales intergalàctiques. Tot i així, de mitjana el cosmos és elèctricament neutre i per tant només la força gravitatòria és la que juga un paper rellevant en l'evolució de l'Univers. La Relativitat General d'Einstein és per ara la teoria que proporciona una descripció més precisa de la grave-

tat. Segons aquesta teoria, la gravetat és sempre atractiva i per tant frena l'expansió de l'Univers. A finals de la dècada dels anys 90, a partir de l'observació de supernoves distants es va mesurar el ritme d'expansió de l'Univers amb l'objectiu de determinar la seva desacceleració i així conèixer el contingut de matèria i la seva geometria. El resultat d'aquestes mesures va ser el descobriment que l'Univers s'havia estat expandint de forma accelerada durant els últims 6000 milions d'anys. Des de llavors, l'explicació d'aquest fenomen encara suposa un desafiament per la comunitat científica.

Un dels models proposats per l'explicació d'aquest fenomen és el Model ΛCDM (Λ-Cold Dark Matter en anglès) que es basa en l'existència d'una forma exòtica d'energia coneguda com *energia fosca*. Segons aquest model un $\sim$ 73% del contingut de densitat d'energia de l'Univers és energia fosca. Les propietats d'aquesta forma d'energia són les d'un fluid homogeni en tot el cosmos amb una pressió negativa capaç així de produir l'expansió accelerada de l'Univers. Una de les possibles explicacions per l'energia fosca és associar-la amb l'energia del buit de teoria quàntica de camps. Tot i així, el valor predit i el valor observat difereixen en 120 ordres de magnitud i avui en dia no s'ha proposat encara cap mecanisme per explicar aquesta discrepància. A més de l'energia fosca, el model ΛCDM assumeix que un $\sim$ 22% de la densitat d'energia de l'Univers és una forma de matèria que es coneix com a *matèria fosca*. A diferència de la matèria ordinària o bariònica, que només suposa en aquest model un $\sim$ 5% de la densitat d'energia total de l'Univers, la matèria fosca només interacciona gravitatòriament i no electromagnèticament. D'aquesta manera, no és capaç de formar estrelles que emetin llum i que puguin observar-se. Així, aquest tipus de matèria només pot ser observada directament pel seu efecte sobre la matèria bariònica i els raigs de llum.

L'objectiu d'aquesta tesi és l'estudi de l'estructura a gran escala de l'Univers des d'un punt de vista teòric. En particular els diferents capítols d'aquesta tesi se centren en desenvolupar eines estadístiques per millorar l'enteniment de la natura dels constituents de l'Univers. En el Capítol 1 es presenta una introducció als conceptes bàsics de la cosmologia actual amb èmfasi en l'estructura a gran escala de l'Univers. En el Capítol 2 es presenta una extensió al models d'halos clàssic on s'inclou una dependència dels halos amb l'entorn. Amb aquest tipus de models s'espera poder explicar millor com les galàxies es distribueixen al cosmos d'acord amb les seves propietats físiques. En el Capítol 3 es presenta un mètode per determinar paràmetres cosmològics tals com el factor de creixement. L'avantatge de la tècnica presentada aquí és que permet reduir l'efecte de la variància còsmica que domina quan estudiem les escales pròximes a la grandària de l'Univers observat. En el Capítol 4 s'utilitza la tècnica del bispectre per determinar com de diferent és la Relativitat General de les teories de gravetat modificada al nivell de la funció de correlació de tres punts. En el Capítol 5 presentem una fórmula analítica pel càlcul del bispectre de matèria fosca calibrada a partir de simulacions de N-cossos. Finalment en el Capítol 6 presentem les conclusions d'aquesta tesi i les perspectives futures.

Esperem que els treballs i esforços realitzats, així com els resultats obtinguts en aquesta tesi sigui útils per futurs projectes científics. En particular, esperem que les tècniques que aquí es presenten combinades amb dades de missions científiques de cartografiat i mapeig de galàxies puguin ser útils per extreure informació rellevant sobre l'evolució i estructura del cosmos i que així puguin ajudar a desentrellar la natura i les propietats de la matèria i energia fosca.

Thesis overview

Chapter 1

A brief introduction of the basics of cosmology and large scale structure of the Universe is presented. This is the starting point for the thesis and provides a vital background material for all the following works developed in the other Chapters.

Chapter 2

This Chapter is concerned to the development of an extension of the Halo Model. We study the possibility of modifying the standard halo model dark matter haloes properties to depend not only on the halo mass but also on the halo environment. Both theoretical and observational studies indicate that properties of dark matter haloes, and specially the way they host galaxies, namely the Halo Occupation Distribution (HOD), depend not only on the mass of the halo host but also on its formation history. This formation history dependence may be related to the halo-surrounding dark matter field. In this work we present a theoretical model that allows to incorporate in a simple way this extra dependence on the environment. In this model the whole population of dark matter haloes is split in two depending on whether the haloes live in high-density environments or in low-density ones. We explore how the dark matter and the galaxy correlation function is affected by this dependence on the environment though the dark matter halo profile or the HOD respectively.

Chapter 3

In this Chapter we explore the possibility of improving the measurement of the growth factor using dark matter tracers. We compare the accuracy in the measurement of the growth factor using a single and two different biased dark matter tracers separately. We make use of realistic bias models, which include non-linear and stochastic parameters, and we calibrate them using dark matter simulations and using haloes of a certain bin-mass as tracers. We expect that using this method the sample variance could be reduced and the accuracy of the measurements improved as previous works have shown.

Chapter 4

This Chapter is concerned to exploring how possible deviations of General Relativity can be detected using the bispectrum technique. We work with a suit of cosmological simulations of modified gravitational action $f(R)$ models, where cosmic acceleration is induced by a scalar field that acts as a fifth force on all forms of matter. The goal is to see how the bispectrum of the dark matter field on mildly non-linear scales is modified by the extra scalar field. In particular we are interested in see which is the effect on the bispectrum, when different gravity models present the same power spectrum at late times.

Chapter 5

In this Chapter we propose a new simple formula to compute the dark matter bispectrum in the moderate non-linear regime ($k < 0.4\,h/\mathrm{Mpc}$) and for redshifts $z \leq 1.5$. Our method is inspired by the approach presented by Scoccimarro and Couchman (2001), but includes a modification of the original formulae and a prescription to better describe the BAO oscillations. Using ΛCDM simulations we fit the free parameters of our model. We end up with a simple analytic formula that is able to predict accurately the bispectrum for a ΛCDM Universe including the effects of Baryon Acoustic Oscillations.

Chapter 6

The major conclusions of the works presented in the thesis are summarised and discussed. Also the possible future projects are discussed.

Contents

CHAPTER 1

General Background

This Chapter is a starting point for understanding the rest of the chapters that compose this thesis. Here we set up the definitions and basics concepts to create a necessary background to understand the formation and evolution of the Cosmos.

This Chapter is organised as follows. In Section §1.1 we present a brief historical review of Cosmology for the last 100 years. In Section §1.2 we introduce the isotropy and homogeneity, two basics assumptions in the description of Universe. In Section §1.3 the Hubble law and the redshift are presented. In Section §1.4 we present Einstein Equations and in Section §1.5 a particular solution to them: the Friedmann Equations, which are the first mathematical model that describes a dynamical expanding Universe. In Section §1.6 we introduce dark matter and dark energy and the corresponding observational evidences. In Section §1.7 $f(R)$ theories are presented. Inflation is briefly presented in Section §1.8. Section §1.9 describes the basic models of structure formation, in particular, linear perturbation theory. Finally, in Section §1.10 we give a statistical description of the distribution of the density perturbations and introduce the concept of galaxy biasing.

We expect that the techniques and results presented in this thesis would be useful for future scientific surveys. In particular the statistical methods presented here may be combined with data from BOSS (Baryon Oscillation Spectroscopic Survey) or DES (Dark Energy Survey) in order to extract relevant information about the evolution, composition and structure of the Universe.

1.1 Historical review

Form a theoretical point of view, modern Cosmology starts in 1917 when Albert Einstein proposes the General Relativity theory (hereafter GR theory). Until that moment, the description of the movement of stellar bodies had been based on Newtonian laws which, now we know, provide a good description at small scales and at moderate velocities, but fail to explain: how gravity propagates, why light rays bend in space and why the speed of light propagates at the same velocity in all reference frames. Under GR, time and space are treated as a whole concept and it provides a description of how space-time is modified under the presence of matter and how the masses move according to the shape of the space-time. An exact solution to the Einstein Equations can be obtained if the Universe is assumed to be isotropic and homogeneous (we discuss this point in Section §1.2). This solution is known as the Friedmann-Lemaître-Robertson-Walker (hereafter FLRW) metric. This metric was found independently by these authors in the 20s and 30s. This solution implies that the Universe must be either expanding or contracting, but never static as Einstein initially thought. The fact that the Universe expands or contracts implies an evolution and an origin, something difficult to accept by some physicist at that time. At that epoch the idea of an expanding Universe seemed very strange because the observations only consisted of stars of our Galaxy and no expansion was observed. In order to adjust the theory with observations Einstein introduced a cosmological constant in his equations to force the Universe to be static. However some time later it was discovered that our Universe consisted of, not only our Galaxy, but many others formed also by gas and stars. This was a breakthrough because it eliminated the idea of a Universe formed only by our Galaxy. Astronomical observations made by Hubble and collaborators in 1929 estimated the distance of a set of nearby galaxies using Cepheids as standard candles. The Cepheids are a class of stars which have a luminosity that varies with a fixed periodicity. From observations in our Galaxy, it was known that the period of oscillation is proportional to the intrinsic luminosity. This feature made these stars useful as standard candles. On the other hand, Hubble had access to the radial velocity measurements of these nearby galaxies. Due to the Doppler effect, the characteristic spectral lines of the population of stars that forms a galaxy, were shifted to the blue (red) side of the spectrum if this galaxy was approaching (going away). Thus, Hubble was able to establish a relation between the recession velocity of galaxies and their distance, the so called Hubble law. The majority of the galaxies observed were moving away from us and their velocity was proportional to their distance: the more distant a galaxy were the faster was moving away from us. This constant of proportionality is called Hubble constant. Since the galaxies are now expanding from each other, it can be extrapolated that in the past all the mass was in the same starting point. Assuming that the rate of expansion is constant, the age of the Universe can be estimated. The first determination of the Hubble constant yielded to a Universe younger than the age of the Earth, which had been previously determined from geological proves. However, more accurate determinations showed that the value of the Hubble constant had been largely overestimated due to errors in the distance calibration. Once this was corrected, the value of the Hubble constant yielded to an age of the Universe about $\sim$ 10 Gyr which was in accordance with the age of the Earth and the oldest stars. As a consequence of these observations, Einstein and other scientists realised that the idea of a static Universe was wrong because GR predicted naturally this expansion. The cosmological constant that Einstein had introduced in his equation was not necessary anymore.

Until the 1990s the most common framework for cosmology was an expanding and decelerating Universe by the force of gravity. The question was whether the matter-density content and the curvature of the Universe (null, positive or negative) would yield to an eternal decelerated expansion or to a contraction and an eventual crunch. These questions motivated experiments that tried to measure with high precision the expansion of the Universe at different epochs, and thus extract the value of the deceleration rate in the recent history. The two main groups that carried these experiments were the Supernova Cosmology Project and the High-Z Supernova Search. The technique these groups used was to measure the light of a set of distant type Ia supernovae. The most accepted model states that this type of supernova results from having a white dwarf star in a binary system. Matter transfers to the white dwarf from the companion star until the white dwarf reaches a critical mass (the Chandrasekhar limit) and undergoes a thermonuclear explosion. Because all white dwarfs achieve the same mass before exploding, they all achieve the same luminosity and can be used as standard candles: observing their apparent brightness, their distance can be determined. On the other hand, the redshift of their light contains information about the expansion of the Universe from when the photon was emitted until today: the more redshifted the light is, the more the Universe has been expanding. Combining these two measurements, these groups were able to extract the expansion of the Universe at different distances (and thus at different epochs) and therefore extract the expansion rate. At the end of 90s, these two groups had been collecting enough data to compute the rate of expansion of the Universe. It was an outstanding breakthrough when these groups discovered that their measurements indicated an *accelerated* expansion rate of the Universe in the last 6000 million years (Perlmutter et al., 1999, Riess et al., 1998). Due to this discovery the Nobel prize of physics of 2011, went to Riess, Perlmutter and Schmidt who leaded these experiments. Since 1999, more data have been obtained and the accelerated rate have been confirmed (Kowalski et al., 2008, Suzuki et al., 2012). In order to explain this new phenomena, the scientific community proposed a model based on something called *dark energy*. In this model $\sim$ 73% of the density-energy content of the Universe is made of this exotic form of energy. Scientists do not know what dark energy is, but have an idea of which properties must satisfy: it should be homogeneous and does not cluster to form galaxies or clumps, such as matter does. One possibility is that dark energy might be related to the vacuum energy from quantum field theories. In that case, this dark energy would be equivalent to add a cosmological constant (again) in Einstein Equations. However, the discrepancy between the energy required to generate the observed expansion and the actual value one obtains from quantum field theories is one of the biggest in the history of physics: 120 orders of magnitude. Alternative theories to the vacuum energy propose a modification of gravity at large scales that would produce an accelerated expansion. One possibility are the $f(R)$ class of models (see Section §1.7 for details). These models produce an accelerated expansion through a modification of the Einstein Equations by an arbitrary function of the Ricci scalar R. As a consequence, an extra propagating scalar field appears that mediates a fifth force on all forms of matter. However also these models present issues, such as recovering GR limit in solar system scales.

Besides the accelerated expansion of the Universe, in the last decades the rotation and clustering of galaxies have been widely studied. It has been found that from the directly observed matter, such as gas and stars, the rotation curves of galaxies cannot be explained. These rotation curves seem to indicate that there is more matter inside galaxies than that can be actually seen. Also, independents measurements, such as strong lensing experiments and observations of galaxy collisions, seem to indicate that there is more matter than that formed by stars and gas. This extra amount of matter needed to explain these phenomena is called *dark matter*. Unlike baryonic matter (the matter that forms stars, planets, atoms), dark matter only interacts gravitationally and not electromagnetically which makes difficult to detect it. Among all the energy-density content of the Universe it seems that only$\sim$ 5% is baryonic matter, $\sim$ 22% is dark matter and the resting $\sim$ 73% is dark energy.

In the following sections of this Chapter we expand some of the concepts introduced in this historical review.

1.2 Isotropy & homogeneity

The main non-testable assumption about the Universe is the cosmological principle. In modern cosmology, the cosmological principle states that the observers on Earth do not occupy a privileged location in the Universe. Hence, viewed on sufficient large scales, the properties of the Universe should be the same for all observers regardless of their location. This concept derives from the assumptions that the humankind does not occupy a privileged point in the Universe, and the laws of physics seem to be the same everywhere around us.

From this principle two main properties of the Universe arise: isotropy and homogeneity. Isotropy means that there is no preferred direction. At sufficient large scales the same properties are observed whatever is the direction we look at. Homogeneity means that there is no preferred location. The same properties are observed no matter where you set up your observation point.

Large scale galaxy surveys and measurements of the Cosmic Microwave Background suggest that the Universe is very close to be isotropic at scales larger than 100 Mpc (Hajian and Souradeep, 2003, Zunckel et al., 2011). On the other hand, homogeneity can be also tested observing certain kind of galaxies at different times. Assuming that we know how the population of stars of these galaxies evolve, we can also test homogeneity, because looking at different times means looking at different distances (Heavens et al., 2011, Hogg et al., 2005). Also, these tests suggest that the Universe is highly homogeneous at very large scales.

1.3 Hubble law and redshift

The Hubble law states that the shift of the light of a galaxy towards the red part of the spectrum (the so called redshift) is proportional to the distance of this galaxy from the observer. This law can be considered the first observational evidence for the paradigm of the expansion of the Universe and today is still a cornerstone for the Big Bang theory. According to this law, the expansion of the Universe is characterised by the Hubble constant. Recent galaxy observations have set the value of this constant in about 73 km/s/Mpc.

The Hubble law was presented by Edwin Hubble in 1929. At that time Hubble's work had been focused on the estimation of the distance to the spiral nebulae using the Cepheids stars as standard candles. It was known that the Cepheids were stars whose period in the luminosity was related to their intrinsic luminosity. The conclusion was that the distance to those nebulae was larger than the typical size of our own Galaxy and therefore, these nebulae were in reality other galaxies. Hubble combined his own data with estimations of the velocities of these galaxies from the shift of spectral lines of the light spectrum. This shift is known as *redshift* and is defined as,

$$z \equiv \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{1.3.1}$$

where λ_{em} is the wavelength in a rest frame and λ_{obs} the observed one. The redshift can be associated to a radial velocity as if it were due to a classic Doppler effect. Thus the relation between the relative recession radial velocity v_z and redhshift is,

$$v_z = zc \tag{1.3.2}$$

where c is the speed of light.

Hubble observed that the relation between the distance and the recession velocities of those galaxies was close to be a linear relation: the more distant a galaxy is, the higher its recession velocity,

$$v_z = H_0 d \tag{1.3.3}$$

where d is the distance to that galaxy and H_0 is the value of the, so called, Hubble constant today.

1.4 A Theory of gravity for the description of the Universe

In this section we give a brief introduction to the Einstein's Equations. These are a set of equations in Einstein's general theory of relativity published in 1915. Einstein's Equations describe gravity as a result of an interaction between space-time and matter and energy. The Einstein's Equations can be written in the following tensorial form (Wald, 1984),[1]

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G\, T_{\mu\nu} \tag{1.4.1}$$

where $g_{\mu\nu}$ is the space-time metric, $T_{\mu\nu}$ is the stress-energy tensor, Λ is the cosmological constant (see Section §1.6 for details) and the terms R and $R_{\mu\nu}$ are the curvature scalar and the Ricci tensor respectively. G is the Newton gravitational constant and c the speed of light. The left-hand-side terms are related to the geometry of the space time of the Universe, whereas the right-hand-side term, $T_{\mu\nu}$ describes the energy and matter content of the Universe. Putting Λ on the left-hand-side of Eq. 1.4.1, one interprets that the origin of this cosmological constant is a vacuum energy, whereas if it is put on the

---

[1] In this section we take the speed of light c to be 1.

right-hand-side the interpretation is that its origin comes from a matter field. Assuming isotropy and homogeneity the FLRW metric is obtained. In spherical coordinates the line element reads,

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right] \tag{1.4.2}$$

where k describes the curvature of the Universe, being $k = 1$ for a closed Universe, $k = 0$ for a flat Universe and $k = -1$ for an opened Universe. $a(t)$ is called the scale factor and describes how the distances change with time. By definition it is 1 today. The time coordinate t is called the cosmic time and corresponds to the time that an observer would measure if the Universe around him were expanding uniformly. The spatial variables (r, θ, ϕ) are called the comoving coordinates of a point in the space; the angular part of this vector, $d\Omega$, is defined as $d\Omega^2 \equiv d^2\theta + \sin\theta d^2\phi$. If the Universe were perfectly homogeneous and isotropic the comoving coordinates of any point would remain constant with time. The Ricci tensor and the curvature scalar for this FLRW metric are (Carroll, 2004, Wald, 1984, Weinberg, 1972),

$$\begin{aligned} R_{tt} &= -\frac{3\ddot{a}}{a} && (1.4.3) \\ R_{ij} &= \left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + \frac{2k}{a^2} \right] g_{ij} && (1.4.4) \\ R &= 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} \right] && (1.4.5) \end{aligned}$$

where the t indices are related to the time-derivatives of the Ricci tensor, whereas the i and j indices are related to the spatial derivatives. Therefore, if the Universe is actually homogeneous and isotropic at large scales, its metric can be described only knowing its curvature k and how the scale factor changes with time.

1.5 Friedmann equations

In order to develop a model for the expansion of the Universe, we need to compute the time dependence of the scale factor $a(t)$. This requires to solve the Einstein field equations for a specific $T_{\mu\nu}$. The simplest solution is achieved by assuming that both matter and radiation in the Universe behave as a perfect fluid. The stress energy tensor for such a Universe is (Carroll, 2004),

$$T_{\mu\nu} = (\rho + p)u^\mu u^\nu - pg^{\mu\nu} \tag{1.5.1}$$

where ρ is the energy density, p is the pressure of the fluid and u^μ the comoving velocity. For an observer who moves with the expansion (comoving observer), the matter fluid is held locally at rest. This means $u^{\mu} = (1, 0, 0, 0)$. Finally, introducing this stress energy tensor in the Eq. 1.4.1 one obtains,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{1.5.2}$$

which is derived from the time-time component and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p\right) + \frac{\Lambda}{3} \tag{1.5.3}$$

which is derived from the trace. Combining these two equations the continuity equation is obtained,

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + p\right) = 0. \tag{1.5.4}$$

For normal matter, $p \geq 0$ and $\rho > 0$, and without any cosmological constant $\Lambda = 0$ we obtain $\ddot{a} < 0$. This means that the Universe is decelerating. For other models with $\Lambda > 4\pi G(\rho + 3p)$ we obtain $\ddot{a} > 0$, i.e. the Universe undergoes an acceleration process. In any case, we see that, under the assumptions of a homogeneous and isotropic Universe, governed by the laws of GR, we obtain a dynamical Universe which is either expanding or contracting but never in stable static state. Usually one works with the Hubble parameter $H(t)$ defined as,

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)}. \tag{1.5.5}$$

This $H(t)$ parameter is the same as the Hubble constant H_0 we defined in Eq. 1.3.3 when t refers to the current time. In that case, $H(t)$ relates a length scale separation between two points with their relative veolicity at some time t.

In this thesis ρ only refers to non-relativistic matter ($\rho \gg p$). The effects of radiation (relativistic matter) are only important at early Universe epochs. Since all the works of this thesis are based on late Universe, radiation can be safely ignored.

A simple analytic solution can be found for the Einstein de Sitter (EdS hereafter) Universe where $k = \Lambda = 0$. Assuming non-relativistic matter, the matter-density ρ scales as a^{-3}: the matter-density is proportional to the inverse of the volume of the sphere that contains it. Thus if ρ_0 and a_0 are the matter density and the scale factor today, Eq. 1.5.2 reads,

$$\frac{\dot{a}}{a} = \left(\frac{8\pi G\rho_0}{3}\right)^{1/2}\left(\frac{a_0}{a}\right)^{3/2}. \tag{1.5.6}$$

Integrating this equation, assuming for simplicity that at initial time $a = 0$, the scale factor depends with time as,

$$a(t) = a_0\left(6\pi G\rho_0 t^2\right)^{1/3}. \tag{1.5.7}$$

Also the matter density is,

$$\rho(t) = \rho_0 \left(\frac{a_0}{a(t)}\right)^3 = \frac{1}{6\pi G t^2} \tag{1.5.8}$$

and the Hubble parameter,

$$H(t) = \frac{2}{3t}. \tag{1.5.9}$$

The critical density

Considering a big sphere of radius R, the total matter inside this sphere is given by,

$$M = \frac{4\pi R^3}{3}\rho \tag{1.5.10}$$

where ρ is the mean matter density inside the sphere. The potential energy of a point mass m in the surface of this sphere is,

$$U = -\frac{4\pi m G \rho R^2}{3}. \tag{1.5.11}$$

On the other hand, the velocity of the point mass respect to the centre of the sphere is given by the Hubble law (Eq. 1.3.3) $v = HR$. Thus the total energy of the galaxy is the sum of kinetic energy plus gravitational energy,

$$E = mR^2 \left(\frac{1}{2}H^2 - \frac{4\pi G\rho}{3}\right). \tag{1.5.12}$$

By conservation principles, E must be constant with time. If $E < 0$ the Universe has not enough energy to expand forever and will recollapse, whereas for $E > 0$ the expansion will be eternal (but decelerated). The transition between these two scenarios is given by $E = 0$. So depending on the total amount of density-matter the Universe will recollapse or not. This characteristic density is defined as the critical density,

$$\rho_c \equiv \frac{3H^2}{8\pi G}. \tag{1.5.13}$$

The critical energy density can also be defined as the energy density necessary to make the Universe flat ($k = 0$) in the absence of cosmological constant (see Eq. 1.5.2). The density parameter Ω is defined as,

$$\Omega \equiv \frac{\rho}{\rho_c}. \tag{1.5.14}$$

However, one can also associate an effective energy density to the curvature and to the cosmological constant. Thus different Ωs can be defined,

$$\Omega_m \equiv \frac{8\pi G\rho_m}{3H^2} \tag{1.5.15}$$

$$\Omega_\Lambda \equiv \frac{\Lambda}{3H^2} \tag{1.5.16}$$

$$\Omega_k \equiv \frac{-k}{H^2} \tag{1.5.17}$$

where the subindices m, Λ and k stands for matter, cosmological constant and curvature respectively. Friedman Eq. 1.5.2 can also be written as a function of the density parameter,

$$\frac{H^2}{H_0^2} = \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda. \tag{1.5.18}$$

1.6 Dark matter and dark energy

In Section §1.1 it was pointed that about 95% of the energy density content of the Universe is dominated by forms of matter and energy that are not well understood. We label these two forms as dark matter and dark energy. In this section we present the observational proves that indicate the possible existence of these forms of matter and energy. Also, from these observational evidences we can know and put constrains on the properties of the dark sector of the Universe.

1.6.1 Dark matter

One of the goals of cosmology during the last decades has been to determine the total amount of matter density of the Universe. One of the reasons to do this is to determine the spacial curvature of the Universe. In the absence of cosmological constant, if the mass density exceeds a critical value the Universe would be positively curved, whereas if this parameter is below then it would have negative curvature. An other important point to determine this parameter is to know what the Universe is made of; which part of matter is in the form of gas and stars and which part is made of dark matter: cold stellar remnants, black holes, exotic particles. Here we enumerate some methods to determine the mass-density parameter.

1. **Mass-to-light ratio**. The amount of matter that constitutes the stellar mass can be directly observed by photometric and spectroscopic surveys. These surveys have been able to measure the luminosity density of the Universe in a certain spectral band. To convert luminosity density into mass density we need to know the mass-to-light ratio for the stars. This is, how much mass of stars takes to produce a certain amount of starlight in a certain spectral band. Knowing the star population of our and nearby galaxies this quantity can be estimated. Assuming a mixture of stars similar to the Solar neighbourhood, the mass-to-light ratio in the B-band is about $4\,M_\odot/L_{\odot,B}$. Taken the total luminosity density of stars in the B-band of our local Universe, $j_{\star,B} = 1.2 \times 10^8\,L_{\odot,B}\mathrm{Mpc}^{-3}$, the density of stars is about $5 \times 10^8\,M_\odot\mathrm{Mpc}^{-3}$ (Ryden, 2003). This value yields a very low value for

the matter density, $\Omega_\star = 0.004$: less than 1% of the matter density necessary to make the Universe flat (in the absence of cosmological constant) is made of stars. Of course not all the mass of the Universe has to form stars. We know that also a big amount forms planets, gas, black holes or brown dwarfs, which does not emit photons and there are not suitable to be detected using this method. However, other method estimation of the mass forming these objects reach at the most $\Omega_b = 0.04$[2] (Fukugita et al., 1998, Percival et al., 2010), so only 4% of the matter necessary to flatten the Universe. Independent observations indicate that the Universe is flat (or close to be flat). Therefore it seems that there is an important amount of mass hidden somewhere else. Since we know that the matter must interact gravitationally, such large amount of missing mass must have some effect on the dynamics of some objects we can observe.

2. **Rotation curves**. Stars move inside galaxies. For instance, in spiral galaxies this movement is around the central bulge. Assuming negligible collisions among stars inside galaxies (in typical galaxies the cross section is very small) the velocity of these stars is completely determined by gravity: the farther is the star from the centre the less is its velocity; and the more mass is in the centre the more is its velocity. It is then possible to estimate the amount of mass inside galaxies measuring the rotation velocity of stars in the spiral arms. Since in this kind of galaxies the majority of the baryonic mass lies in the centre of the galaxy, the mass in the outskirts can be neglected for this estimation. Thus, according to gravitational dynamics, the velocity of the stars in the outskirts should decrease as the inverse of the square root of their distance to the galactic centre. However, observations indicate that rotation curves do not follow this law but present a flat profile: the velocity seems to be constant at large distances and do not decrease (de Blok et al., 2008). This indicates that either the gravitational dynamics applied is wrong at these scales or there is more mass inside the galaxy that the one that can be observed. In the last case, the 'dark' mass distribution should increase with the radius to the centre. Usually this dark matter component of galaxies is known as galaxy dark matter halo. Taking into account this effect, the amount of mass density in the Universe reaches $\sim 20\%$ of the critical value (Burkert and Silk, 1997).

3. **Cluster dynamics**. The movement of galaxies inside clusters also is affected by the total amount of mass and can be cross-checked with the visible mass in a particular cluster. Fritz Zwicky in the 30s was the first in studying the dynamics

---

[2] The subindex b stands for baryonic mass.

of the Coma cluster and realised that the necessary mass to produce the observed dispersion was higher than the mass in visible stars. More accurate studies have recently shown that according to the cluster dynamics the total amount of mass in the Universe should be around 20% of its critical value (Rines and Diaferio, 2006, Rines et al., 2003).

4. **Gravitational lensing**. So far, the methods described for detecting dark matter are indirect methods: its presence is inferred from the movement of other objects due to the effect of gravity. However, there are techniques for detecting dark matter though the effect on light rays. According to GR, matter not only affects the trajectories of other particles but also the trajectory of photons. Dark matter (and any kind of matter) can bend and focus the light acting as a gravitational lens. However, it has not been until the last decades that this technique has been practically used to detect and measure the total amount of matter in the Universe (Benjamin et al., 2007, Falco et al., 1998).

5. **X-Ray detection.** If dark matter exists, it had to be produced by some process. Since we expect that dark matter particles do not interact electromagnetically, the more likely process is via weak force. If so, we expect that some weakly interactive process involves dark matter particles: some decay or annihilation. These kind of processes would produce an emission of X-rays or gamma-rays that could be detected by some observatories such as Chandra (Allen et al., 2002, Vikhlinin et al., 2009).

6. **CMB fluctuations**. The analysis of the photons coming from the Cosmic Microwave Background (CMB hereafter) puts constrains on the total amount of matter of the Universe. Due to the effect of the radiation pressure, photons and baryon particles oscillated together before their decoupling when the temperature of the Universe dropped down due to the expansion. At that moment, these oscillations were imprinted in the CMB photons spectrum. The shape and amplitude of these oscillations are closely related to the amount of matter in the Universe at that moment. Surveys such as COBE[3], WMAP[4] and recently Planck, have provided very accurate data of the spectrum of these photons and the amount of matter can be estimated. Recent studies suggest that the amount of total matter is around $\Omega_m = 0.23$ (Larson et al., 2011, Spergel et al., 2003).

7. **Collisions**. The presence of dark matter can also be inferred from collisions of clusters. Although these are not common phenomena, there are some observa-

---

[3] COsmic Background Explorer

[4] Wilkinson Microwave Anisotropy Probe

tions such as the Bullet cluster where a recent collision has been occurred. The gas component of the spectrum shows how the gas content of this cluster is interacting due to electromagnetic forces, which produces an effective viscosity between the two 'bullets'. On the other hand, gravitational lensing and galaxy observations show that dark matter and galaxies of the two clusters has passed without any electromagnetic interaction in such a way that now the gas-mass and dark matter and galaxies do not occupy the same place (Clowe et al., 2004, Markevitch et al., 2004). These phenomena are useful to determine whether dark matter can interact electromagnetically. It seems that the interaction via this force is very week or null. This constrains the particles that can form this kind of matter. Some theories postulate that dark matter could be composed of heavy neutrinos or other exotic particles such as axions or supersymmetric particles.

1.6.2 Dark energy

In 1929 Hubble observed that the Universe was a dynamical system that was expanding and was not static as it has been though (see Section §1.1 for details). Due to the effect of gravity, this expansion should be decelerated. However due to technical limitations of the instruments, the deceleration rate was not determined. It was in 1998 when the Universe was discovered to be speeding up its expansion. This accelerated expansion is initially counter-intuitive but the evidence has become convincing. Here we enumerate some observational proves of the accelerated expansion.

1. **SuperNovae**. In the 90s two projects were dedicated to determine the rate of the expansion of the Universe: the Supernova Cosmology Project and the High-Z Supernova Search. These groups were looking for distant Type Ia supernovae in order to measure the expansion rate of the Universe at different epochs (see Section §1.1 for details). They expected that the expansion rate would be decreasing with time. Hence, the supernovae would look brighter that their redshifts would indicate. However they found the opposite effect which indicate that the Universe is undergoing an accelerated expansion (Perlmutter et al., 1999, Riess et al., 1998).

2. **CMB anisotropies**. The second evidence of the accelerated expansion is the Cosmic Microwave Background. Measurements of the CMB indicate that (assuming a dark energy of cosmological constant kind) the Universe has a euclidian or flat geometry at large scales. Since the total amount of matter (including dark matter is $\Omega_m \simeq 0.23$) is far from the value to produce this flatness, the difference

is attributed to an exotic form of energy called dark energy (Larson et al., 2011, Spergel et al., 2003). This energy is the same that causes the acceleration of the expansion of the Universe. It is important to note that the term 'dark' refers to the ignorance about this form of energy rather than for the dark matter which refers to the absence of photons.

3. **Galaxy clusters**. The third evidence of the accelerated expansion is related to the evolution of galaxy clusters. This evolution is closely related to the expansion of the Universe. Since these clusters tend to collapse and the expansion goes against this collapse, a measurement of the amount, distribution and density of these clusters provides information of the expansion rate of the Universe. These measurements seem to indicate that the expansion is actually accelerated (Battye and Weller, 2003, Viana, 2004).

These independent proves indicate that the Universe is expanding in an accelerated way. However the physical mechanism that drives this expansion is still unknown. A possible explanation is to identify the dark energy with Einstein's cosmological constant. However, as it has been pointed out in section §1.1 there is an important discrepancy between the theoretical and observed value for this constant.

Besides the cosmological constant, other theories have been proposed to explain this acceleration. Modified gravity theories based on a modification of the Einstein Equation's (Eq. 1.4.1) changing the Ricci scalar by an arbitrary function of it, could also explain why gravity behaves repulsively at large scales. This kind of procedure is presented and widely explained in Section §1.7 and used in the work developed in Chapter 4.

1.7 $f(R)$ theories of gravity

The accelerated expansion of the Universe has still no satisfactory explanation. This phenomenon has implications not only for cosmology but also for fundamental physics: whatever is driving the accelerated expansion of the Universe must have a close connection to particle physics. Nowadays two main theories have been proposed to explain this effect. One is to assume that the $\sim$73% of the energy content of the Universe is a mysterious form of energy which is homogeneous, unclustered and with a negative pressure, which is able to drive the accelerated expansion (see Section §1.6). Among these dark energy models a special case is when the equation of state of this dark energy fluid is $P = -\rho$. In this case dark energy is equivalent to add a cosmological constant to Einstein Equations. The other way consists in dispensing with dark energy and modifying gravity at large scales. In particular, in this thesis we focus on modified gravity models where the modification is made with an arbitrary function of the Ricci scalar (see Faraoni (2008), Sotiriou and Faraoni (2010) for a review).

In this class of models, the Einstein-Hilbert action is written as[5],

$$S = \frac{1}{16\pi G}\int d^4x\,\sqrt{-g}\,[R + f(R) + L_m] \tag{1.7.1}$$

where L_m is the lagrangian of matter and $f(R)$ is an arbitrary function of the Ricci scalar R. Note that when the derivative of the function respect to R, namely f_R, tends to 0 we recover pure GR. In particular for $f_{GR}(R) = -2\Lambda$ we recover the ΛCDM gravity model. In this last case the massless spin 2 graviton is the only particle that mediates the force, whereas for a generic $f(R)$ function, in addition to this massless graviton, a massive scalar mode appears which can drive the cosmic acceleration.

A conservative point about modification of gravity should be adopted. Here we consider $f(R)$ gravity more as a toy model rather than a correct theory in the sense that these kind of theories are a proof of principle that modifying gravity is a possible alternative to dark energy. However modifying gravity is risky since unwanted consequences may appear: modifications of GR at Solar system scales, instabilities or ghosts among others fatal consequences. In any case, it has to be taken into account that none of the proposed $f(R)$ models have an exceptional support from observational data. Thus, there are many of these $f(R)$ models that are able to pass all the observational tests and fit cosmological data, but it is also true that the same happens for many dark energy models: it is not possible to discriminate among $f(R)$ models and dark energy models using current observations.

In this thesis, and more precisely in Chapter 4, we focus in particular case of modified gravity presented by Hu and Sawicki (2007), where the $f(R)$ function is given by

$$f(R) = -m^2\frac{c_1(R/m^2)^n}{c_2(R/m^2)^n + 1} \tag{1.7.2}$$

where c_1 and c_2 are free dimensionless parameters, $m^2 \equiv 8\pi G\bar{\rho}_m/3$ and for this thesis we always take $n = 1$. In particular, for high enough curvature such as $c_2(R/m^2) \gg 1$ this function can be approximated as,

$$f(R) = -m^2\frac{c_1}{c_2} - f_{R0}\frac{\bar{R}_0^2}{R} \tag{1.7.3}$$

where f_{R0} is the derivative of the function at the background curvature today $\bar{R}_0$,

$$f_{R0} \equiv \left.\frac{df}{dR}\right|_{\bar{R}0} = -\frac{c_1}{[c_2(\bar{R}_0/m^2)+1]^2} \simeq -\frac{c_1}{[c_2(\bar{R}_0/m^2)]^2}. \tag{1.7.4}$$

---

[5] In order to simplify the equations, in this section we set $c = \hbar = 1$

This form of $f(R)$ is able to reproduce the ΛCDM expansion history without requiring any cosmological constant. Also, it incorporates naturally the so-called chameleon mechanism that allows to evade solar-system tests. This mechanism allows the new force mediator to gain mass in dense environments and effectively reduce the range of the force making it negligible. However, this model presents interesting deviations from ΛCDM at intermediate scales that produce singular signatures in the large scale structure formation. These kind of deviations has been studied in (Ferraro et al., 2011, Li and Hu, 2011, Oyaizu et al., 2008, Schmidt et al., 2009) and also in the Chapter 4 of this thesis.

The free parameters c_1 and c_2 can be chosen to mimic the ΛCDM expansion history,

$$\bar{R}_{\Lambda\mathrm{CDM}} = 3m^2 \left(a^{-3} + 4\frac{\Omega_\Lambda}{\Omega_m} \right) \tag{1.7.5}$$

This requirement is equivalent to $f_{R0} \ll 1$ because the modification of gravity only plays a relevant role at late times. In that case, we can write the expansion history of the modified gravity as (Oyaizu, 2008),

$$\bar{R} = 8\pi G\bar{\rho}_m - 2f(\bar{R}) \simeq 8\pi G\bar{\rho}_m + 2\frac{c_1}{c_2}m^2. \tag{1.7.6}$$

Then, in order to match the ΛCDM expansion history one requires,

$$\frac{c_1}{c_2} = 6\frac{\Omega_\Lambda}{\Omega_m} \tag{1.7.7}$$

With this we can write $f(R)$ as,

$$f(R) = -2\Lambda - f_{R0}\frac{\bar{R}_0^2}{R}. \tag{1.7.8}$$

The variable f_{R0} is left to characterise the strength of the extra matter field. In particular for this thesis the cases $|f_{R0}| = 10^{-4}$, 10^{-5}, 10^{-6} are explored.

In Chapter 4 we study using N-body simulations, how this modification can affect the statistical moments measured in cosmology, namely the power spectrum and bispectrum (see section §1.10 for details), and we compare this results with cases where a ΛCDM is assumed in order to find measurable differences.

1.8 Inflation

In cosmology the term inflation refers to a theorised extremely fast expansion of the very early Universe (from $\sim 10^{-36}$ to $\sim 10^{-32}$ seconds after the Big Bang) driven by an effective negative-pressure fluid, similar to dark energy. After this inflationary period, the Universe continued expanding but at much slower rate. This theory was proposed by Guth (1981) in order to explain some problems of the standard model of cosmology. While several models are proposed to explain this process, nowadays the exact physical mechanism is still unknown. However, the basic picture is based on the existence of a particle that would be responsible for the inflationary process called inflaton.

As a consequence of this rapid expansion, all the observable Universe was originated in a causally connected region. Inflation was proposed to explain three main problems of the standard model of cosmology: *i)* why is the Universe flat at large scales, *ii)* why is the Universe so isotropic and homogeneous at large scales and *iii)* which is the origin of the large scale structure of the Universe. These problems are briefly discussed in this section.

1. **The flatness problem.** This is a fine-tunning problem in the Big Bang model of the Universe. The Universe is observed today very close to be flat: the total energy density of the Universe (Eq. 1.5.18) is observed to be very close to 1. This means that the total amount of matter (baryonic and dark matter) and energy is very close to the critical value. However, we know that the cosmic evolution makes the density parameter to depart rapidly from this value because is not stable. This implies that in the early Universe the density parameter has been even closer to the critical density, departing from it by one part in $\sim 10^{60}$.

2. **The horizon problem.** Measurements of the CMB show that when the Universe was about 3×10^5 years old it was extremely well termalised: very distant regions of the Universe which should have not been in contact (because the Universe was so young that any particle had no time to travel) present a very uniform temperature, with fluctuations around 10^{-5}. This effect is in direct connection with the fact that the Universe seems to be very homogeneous and isotropic with no apparent reason for that.

3. **The structure problem.** Although is assumed that the Universe is isotropic and homogeneous at large scales, at much lower scales this assumption breaks down: today we observe large amount of matter concentrations and large voids. Which is the origin of these structures? It is not satisfactory to assume that the initial conditions of the Universe were set in this way. Inflation can provide a more satisfactory answer: quantum fluctuations in early times magnified to cosmic size by the inflationary expansion, can become the seeds for the growth of these structures in the Universe. In this case, the distribution of matter in the Universe is itself statistical.

Thus, the inflationary theory explains the existence of small overdensities of matter in an almost homogeneous and isotropic Universe described by FLRW metric. In the next section we describe how these small overdensities can grow and become the clusters and galaxies we observe today.

1.9 Structure formation

Here we describe how a tiny perturbation grows to finally collapse forming a virialised structure. Throughout, we assume that at sufficiently large scales the Universe is homogeneous and isotropic, so the FLRW metric is used.

Linear perturbations in Newtonian theory

The perturbations which are useful to study the galaxy clustering are the sub-horizon fluctuations that are collapsing today. Because of that, we use Newtonian perturbation theory in an expanding frame. This theory provides an accurate description of the system when the perturbation δ is much smaller than 1.

If $p(\mathbf{r}, t)$ is the pressure, $\rho(\mathbf{r}, t)$ is the density and $\mathbf{u}(\mathbf{r}, t)$ is the velocity of the particles of this fluid at a point $\mathbf{r}$, then the equations that control the evolution of these particles are (Weinberg, 1972):

Continuity equation. This equation express the conservation of mass enclosed within a Lagrangian volume element dV as the fluid flows out through the surface,

$$\left(\frac{\partial \rho}{\partial t}\right)_{\mathbf{r}} + \nabla_{\mathbf{r}} \cdot (\rho \mathbf{u}) = 0 \tag{1.9.1}$$

where the subscript $\mathbf{r}$ denotes that the reference frame is the proper or physical coordinates.

Euler's equation. This equation describes the forces acting on a fluid element that flows with the fluid,

$$\left(\frac{\partial \mathbf{u}}{\partial t}\right)_{\mathbf{r}} + (\mathbf{u} \cdot \nabla_{\mathbf{r}})\, \mathbf{u} = -\nabla_{\mathbf{r}} \Phi - \frac{1}{\rho} \nabla_{\mathbf{r}} p \tag{1.9.2}$$

where Φ is the gravitational potential that describes the global distribution of matter in the Universe.

Poisson's equation. This expression describes how the gravitational field acts over the matter,

$$\nabla_{\mathbf{r}}^2 \Phi = 4\pi G \rho. \tag{1.9.3}$$

It is useful to express these equations in a reference frame that coexpands with the Universe: not in terms of proper coordinates r but in terms of comoving coordinates x, which are related through the cosmic scale factor a,

$$\mathbf{r} = a\mathbf{x}. \tag{1.9.4}$$

Deriving this last equation respect to the time, we obtain a similar relation for the vector $\mathbf{u}$ in terms of the Hubble constant H,

$$\mathbf{u} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} = H\mathbf{r} + \mathbf{v} \tag{1.9.5}$$

where the first term of the right-hand-side represents the Hubble flow, whereas the second one is the proper peculiar velocity $\mathbf{v}$ of a particle. Eq. 1.9.4 yields to the following relations between operators,

$$\nabla_{\mathbf{r}} \rightarrow \frac{\nabla_{\mathbf{x}}}{a} \tag{1.9.6}$$

and

$$\left(\frac{\partial}{\partial t}\right)_{\mathbf{r}} \rightarrow \left(\frac{\partial}{\partial t}\right)_{\mathbf{x}} - \frac{\dot{a}}{a}\mathbf{x} \cdot \vec{\nabla}_{\mathbf{x}}. \tag{1.9.7}$$

In order to simplify our expression we denote,

$$\nabla \equiv \nabla_{\mathbf{x}}. \tag{1.9.8}$$

Rewriting the fluid equations in terms of comoving quantities, we obtain (Weinberg, 1972)

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{a}}{a}\rho + \frac{1}{a}\nabla \cdot (\rho \mathbf{v}) = 0 \qquad \text{Continuity} \tag{1.9.9}$$

$$\ddot{a}\mathbf{x} + \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\frac{\dot{a}}{a} + \frac{1}{a}(\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{a}\frac{\nabla p}{\rho} - \frac{\nabla \Phi}{a} \qquad \text{Euler} \tag{1.9.10}$$

$$\nabla^2 \Phi = 4\pi G \rho a^2. \qquad \text{Poisson} \tag{1.9.11}$$

We focus now on finding a solution to this system of equations. Hereafter, we work with the dimensionless density perturbation variable $\delta(\mathbf{x}, t)$, defined to be

$$\rho(\mathbf{x}, t) \equiv \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)] \tag{1.9.12}$$

and in terms of the local potential associated to the perturbations $\phi(\mathbf{x}, t)$ defined as

$$\Phi(\mathbf{x}, t) \equiv \phi(\mathbf{x}, t) - \bar{\phi}(t). \tag{1.9.13}$$

In order to obtain a simple solution of our fluid equations we work in the linear limit, when the perturbations are very small. Mathematically this means that the fields δ, ϕ

and v are small, and we neglect second order terms, keeping only the leading order ones. Doing this we obtain,

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot \mathbf{v} = 0 \qquad \text{Continuity} \tag{1.9.14}$$

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}\nabla\phi = -\frac{1}{\bar{\rho}a}\nabla p \qquad \text{Euler} \tag{1.9.15}$$

$$\nabla^2\phi = 4\pi G a^2 \bar{\rho}\delta \qquad \text{Poisson} \tag{1.9.16}$$

Combining these three equations, we finally obtain the expression for the linear growth of δ,

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\bar{\rho}\delta = \frac{1}{\bar{\rho}a^2}\nabla^2 p. \tag{1.9.17}$$

This is a second order equation that describes the time-evolution of δ in the linear regime when $\delta \ll 1$. In the case of cold dark matter (with no electromagnetic forces) the fluid is effectively collisionless and pressure effects can be neglected. In the case of normal baryonic matter the fluid is assumed to be collisional and some pressure function of the density, namely $p(\delta)$, is assumed. These two cases are described below.

Collisionless matter

Usually, the collisionless matter case is used as a basic model to describes the behaviour of non-baryonic dark matter. For an EdS Universe, we have seen in Section §1.5 that $a(t) \sim t^{2/3}$, $H(t) \sim t^{-1}$ and $\bar{\rho}(t) = [3H(t)^2]/[8\pi G] \sim t^{-2}$. Furthermore for collisionless matter the pressure p of the particles is negligible and Eq. 1.9.17 becomes

$$\ddot{\delta} + \frac{4}{3t}\dot{\delta} - 4\pi G\bar{\rho}\delta = 0. \tag{1.9.18}$$

Trying a power-law solution of the form $\delta \sim t^n$, we obtain that only solutions with $n = 2/3$ and $n = -1$ are allowed. Thus, we can write the general solution as,

$$\delta(\mathbf{x}, t) = \delta_+(\mathbf{x})t^{2/3} + \delta_-(\mathbf{x})t^{-1} \tag{1.9.19}$$

where the terms δ_+ and δ_- can be fixed by the boundary conditions. For instance, if we assume that at some initial time t_i the variation of the overdensity is 0, $\dot{\delta}(t_i) = 0$, the equation above reads,

$$\delta(\mathbf{x}, t) = \delta_i(\mathbf{x})\left[\frac{3}{5}\left(\frac{t}{t_i}\right)^{2/3} + \frac{2}{5}\left(\frac{t}{t_i}\right)^{-1}\right]. \tag{1.9.20}$$

The first term is called the growing term and is proportional to the scale factor, $\delta(t) \sim a(t)$; whereas the second term is called the decaying term. Since the decaying term goes to 0 as time evolves, only the growing term has a physical relevance in the late Universe.

Collisional matter

The case of collisional matter, is usually applied to describe the behaviour of normal baryonic matter. In this case, we assume the pressure as a simple function of the density, $p = p(\rho)$. Expanding the pressure as a Taylor series, to first order about the mean density,

$$p(\rho) = p(\bar{\rho} + \delta\rho) = p(\bar{\rho}) + \left|\frac{\partial p}{\partial \rho}\right|_{\bar{\rho}} \delta\rho = p(\bar{\rho}) + c_s^2 \bar{\rho}\delta \tag{1.9.21}$$

where $c_s^2 = \partial p/\partial\rho$ is the sound speed in the fluid. Using this, Eq. 1.9.17 reads,

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - \left(\frac{c_s}{a}\right)^2 \nabla^2\delta - 4\pi G\bar{\rho}\delta = 0. \tag{1.9.22}$$

In this case is easier to work in the Fourier space rather than in real space. Is useful to express the overdensity as a Fourier sum of plane waves,

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{x}) \tag{1.9.23}$$

where $\lambda = 2\pi a/k$ is the proper wavelength of the fluctuation. Then, in Fourier space Eq. 1.9.22 reads,

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} + \omega_k^2 \delta_{\mathbf{k}} = 0 \tag{1.9.24}$$

where we have defined

$$\omega_k^2 \equiv \left(\frac{kc_s}{a}\right)^2 - 4\pi G\bar{\rho}. \tag{1.9.25}$$

Thus, for collisional matter, the time-evolution of $\delta_{\mathbf{k}}$ has the same form as a damped harmonic oscillator when $\omega_k \geq 0$. The sign of ω_k is determined by the k-mode we are considering. For small k-modes, the fluctuation grows as if the fluid were collisionless, whereas for large k-modes collisions (or the pressure) prevent the gravitational collapse. So, there exists a length scale that balances the collapse under gravity and the pressure effects. This length is called the Jeans length λ_J,

$$\lambda_J \equiv c_s \left(\frac{\pi}{G\bar{\rho}}\right)^{1/2}. \tag{1.9.26}$$

Therefore, this scale determines when the pressure effects are more important than gravity effects and vice versa. For scales $\lambda \gg \lambda_J$ the pressure effects are negligible, whereas for $\lambda \ll \lambda_J$ we have $\omega_k > 0$ and $\delta_{\mathbf{k}}$ oscillates as acoustic waves. However, as time goes on, these acoustic oscillations finally disappear due to expansion and the k-modes that satisfy $\lambda \gg \lambda_J$ finally dominate.

So far we have seen that the linear perturbation theory provides a useful, simple and analytical model to describe the process in the first stages of collapse when the perturbation is small, i.e. $\delta \ll 1$. However, once the density of the collapsing object has grown enough, this formalism is no longer applicable and a new method is needed. There are several useful approximation that allows to work when $\delta \sim 1$. One of them is the spherical collapse model.

Spherical collapse model

The main assumption of this approach is to consider a spherically symmetric perturbation that is overdense with respect to the rest of the Universe (Gunn and Gott, 1972). Therefore, the dark matter behaves as a perfect and pressureless fluid which is collapsing around a centre. In this approach the tidal forces due to nearby perturbations and the effects due to the inner substructure inside the perturbation are neglected. We also assume a collapse whose initial peculiar velocities are 0 and where the initial density profile of the perturbation is monotonically decreasing from the centre. Therefore the dynamics of the shell are determined by the mass interior to the shell. Thus, the equation of motion of a shell is given by the second Newton's law. Assuming that at some initial time t_i the perturbation has the density $\rho(\mathbf{r}, t_i) = \bar{\rho}(t_i)[1 + \delta(\mathbf{r}, t_i)]$, the time-evolution of a shell of radius r is,

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{Gm(< r)}{r^3}\mathbf{r} \tag{1.9.27}$$

where $m(< r)$ is the mass inside the shell of radius r (which is constant because no shell-crossing is allowed),

$$m(< r, t) = m(< r_i, t_i) = \frac{4}{3}\pi G\bar{\rho}(t)r_i^3[1 + \bar{\delta}_i] \tag{1.9.28}$$

where $\bar{\delta}_i$ is the main overdensity inside the shell of radius r_i. Integrating the equation of motion of the shell we obtain information about the energetics of the shell,

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 - G\frac{m(< r)}{r} = E \tag{1.9.29}$$

where E is a constant of integration that corresponds to the total energy of the perturbation. The sign of E determines whether the perturbation will end in a collapsed structure or will expand forever due to the Hubble flow. If $E > 0$, the perturbation is unbounded and it will never collapse. On the other hand, if $E < 0$ then the perturbation is bounded and as r increases $\dot{r}$ decreases until it becomes 0 and eventually negative, which means that the shell is starting a contraction. For practical interest, we only study the case when collapse is going to happen, i.e. $E < 0$. The solution for Eq. 1.9.29 when $E < 0$ is,

$$r = A(1 - \cos\theta); \qquad t = B(\theta - \sin\theta), \tag{1.9.30}$$

where θ is the evolution angle, and the constants A and B are related by $A^3 = GMB^2$. Recalling that for an EdS Universe (see Eq. 1.5.8) ,the density of the background is given by,

$$\bar{\rho}(t) = \frac{1}{6\pi G t^2} \tag{1.9.31}$$

using the Eq. 1.9.30 we can write the perturbation density δ only in terms of the angle θ,

$$\delta_{NL}(t) \equiv \frac{\rho(t)}{\bar{\rho}(t)} - 1 = \frac{9}{2}\frac{(\theta - \sin\theta)^2}{(1-\cos\theta)^3} - 1. \tag{1.9.32}$$

The above equation give us the evolution of the shell in the non-linear regime (we remark this fact with the subscript NL). On the other hand, we can recover the linear limit from these equations, simply taking the limit $\theta \to 0$. Expanding the sine and cosine about $\theta \sim 0$ until order $\mathcal{O}(\theta^7)$, Eq. 1.9.30 reads,

$$r_L \simeq A\left(\frac{\theta^2}{2!} - \frac{\theta^4}{4!} + \mathcal{O}(\theta^6)\right), \qquad t_L \simeq B\left(\frac{\theta^3}{3!} - \frac{\theta^5}{5!} + \mathcal{O}(\theta^7)\right). \tag{1.9.33}$$

Combining these two equation we obtain,

$$r_L \simeq \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right]. \tag{1.9.34}$$

Then, using

$$\rho_L = \frac{3m}{4\pi r_L^3}$$

we can say that in the linear regime, the overdensity reads

$$\delta_L = \frac{3}{5}\left(\frac{3}{4}\right)^{2/3}(\theta - \sin\theta)^{2/3}. \tag{1.9.35}$$

This equation in principle is only valid if $\delta \ll 1$. However, we can extrapolate the value of δ to non-linear epochs when $\delta \sim 1$, and keep working as if we were in linear theory. Important stages in the perturbation's evolution can now be identified:

1. **Turnaround**. At this stage the spherical shell reaches its maximum radius r_{ta} and then the recollapse starts. It is characterised by,

$$\begin{aligned} \theta_{ta} &= \pi, \quad r_{ta} = 2A, \quad t_{ta} = B\pi, \quad \delta_L^{ta} = \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \simeq 1.06, \\ \delta_{NL}^{ta} &= \frac{9\pi^2}{16} - 1 \simeq 5.54. \end{aligned} \tag{1.9.36}$$

2. **Collapse**. After the turnaround stage, the perturbation finally enters in the non-linear regime and eventually undergoes a gravitational collapse. This happens when,

$$\theta_c = 2\pi, \quad r_c = 0, \quad t_c = 2t_{ta}, \quad \delta_L^c = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \simeq 1.686, \quad \delta_{NL}^c \to \infty. \tag{1.9.37}$$

3. **Virialization**. From the mathematical point of view, if we are assuming non-interacting particles that collapse, once they arrive at the point of $\theta = 2\pi$ the particles start expanding again repeating the cycle. However, in the real physical case we do not expect that to happen. The dissipative process eventually readjusts the orbits of the particles and the system reaches to the virialised stage. This process is called *dynamical friction* or *gravitational drag*. The effect of gravity produces that light bodies accelerate and gain momentum and kinetic energy, whereas heavier bodies are slowed down to compensate. In this stage, the virial theorem is satisfied, i.e., $2K + V = 0$, where K and V are respectively the kinetic and potential energies. Applying the virial theorem one obtains,

$$\begin{aligned}\theta_v &= \frac{3\pi}{2}, \quad r_v = \frac{r_{ta}}{2}, \quad t_v = \frac{t_{ta}}{\pi}\left(\frac{3\pi}{2}+1\right), \qquad (1.9.38)\\ \delta_L^v &= \frac{3}{5}\left(\frac{9\pi}{8}+\frac{3}{4}\right)^{2/3} = 1.58 \quad , \delta_{NL}^v = \frac{9}{2}\left(\frac{3\pi}{2}+1\right)^2 - 1 \simeq 145.8.\end{aligned}$$

Here we have assumed that the virialization happens immediately after the radius reaches the virial radius but this is very unlikely. The virialization process is not an instantaneous process and while this process goes on, the mean density of the Universe decreases and therefore δ^v should be larger than the estimated value. Because of this, is better to assume that the virialization is completed not at $t_v = t_{ta}/\pi\,(3\pi/2+1)$ but at the collapse time, i.e. $t_v = t_c = 2t_{ta}$. However, we still consider that $r_v = r_{ta}/2$.

The density at the 'new' virial time is

$$\rho(t_v) = \frac{3m}{4\pi G r_v} = 8\frac{3m}{4\pi G r_{ta}} = 8\rho(t_{ta}) \qquad (1.9.39)$$

and the background density is

$$\bar{\rho}(t_v) = (6\pi G t_v^2)^{-1} = \bar{\rho}(t_{ta})/4 \qquad (1.9.40)$$

Therefore, combining these two equations and using the value of δ_{ta} of Eq. 1.9.36 we finally obtain that

$$\frac{\rho_{vir}}{\bar{\rho}} = 18\pi^2 \simeq 178. \qquad (1.9.41)$$

As long as the collapsed structure, namely halo, is defined by its virial radius r_{vir}, the mass of this structure must be given by,

$$m = \frac{4}{3}\pi r_{vir}^3 \rho_{vir} = \frac{4}{3}\pi r_{vir}^3\, 18\pi^2\bar{\rho} \qquad (1.9.42)$$

therefore,

$$r_{vir}(m) = \frac{1}{2\pi}\left(\frac{m}{3\bar{\rho}}\right)^{1/3}. \qquad (1.9.43)$$

This formalism allows then to define some collapsed and virialised structure called dark matter halo. These haloes provide an environment that allow galaxies to form evolve and interact. In Chapter 2 we widely describe and extend an analytical formalism based on haloes that is able to make useful predictions about the clustering of dark matter and galaxies in the Universe.

1.10 Statistics of galaxy clustering

1.10.1 Gaussian density perturbations

The standard picture of structure formation in cosmology assumes that the physical process that generated the initial perturbations produced fluctuations that formed a field that was Gaussian or very close to be Gaussian. This process is the responsible to produce, after some time and evolution, the galaxies and the large scale structure that can be observed today in the Universe. The current observations indicate that these initial fluctuations were very close to be Gaussian, although some signal of non-Gaussian features might be detected in the next years (Babich et al., 2004, Creminelli et al., 2006, Huterer et al., 2010, Komatsu et al., 2009). In case of this field being the overdensity of matter δ (defined in Eq. 1.9.12), the probability of obtaining a particular realisation of the field can be written in case of δ a Gaussian field as,

$$p(\delta_1,\ldots,\delta_n)d\delta_1,\ldots,d\delta_n = \frac{1}{\sqrt{(2\pi)^n\det(C)}}\exp\left[-\sum_{ij}\frac{\delta_i C_{ij}^{-1}\delta_j}{2}\right]d\delta_1,\ldots,d\delta_n \quad (1.10.1)$$

where $\delta_i \equiv \delta(\mathbf{x}_i)$ and $C_{ij} \equiv \langle\delta_i\delta_j\rangle$ is the covariance matrix between random variables. $\langle\ldots\rangle$ denotes the average over realisations. Note therefore, that the probability of finding a particular realisation of the δ field only depends on the covariance matrices of this fields in different x-points if δ is a random Gaussian field.

1.10.2 Ergodicity

Statistical quantities, such as the covariance matrix, are defined as an ensemble average of the δ-field over an infinite number of realisations. It is clear that this cannot be done in practise because we only can observe 1 single realisation of the δ-field, we only have access to our observable Universe. However, it turns out that all Gaussian fields have the property of being ergodic fields. This means that if the spatial correlation decays sufficiently rapid with separation such that, we have many statistical independent volumes in one realization, then averaging over volumes is equivalent to averaging over realizations. Since we expect that very distant regions of our observable Universe are uncorrelated, we can use ergodicity to average over different volumes as if they were different relalisations.

1.10.3 Cosmic Variance

The main issue of averaging over different disconnected patches of our Universe appears when we reach very large scales. In that case the number of patches might not be large enough to be a representative sample of the whole Universe to obtain a quantity which is statistically significant. This effect is called cosmic or sample variance. In Chapter 3 we present a technique that allows to reduce the cosmic variance effect in the measurements of cosmological parameters.

1.10.4 Correlation functions

The standard way of describing the clustering of dark matter and galaxies has been the correlation functions. The usage of these tools to study the distribution of galaxies in the Universe was first proposed in the 60s by Totsuji and Kihara and popularised by Peebles and collaborators in the 70s. The reason to do that is that the models presented in previous sections do not provide information about the presence of matter or galaxies in a specific region of the Universe because its own nature is statistical (see Section §1.8). Therefore, these models only can be tested and contrasted with observations if we use statistical tools, such as correlation functions. From a mathematical point of view, correlation functions provide a statistical description of how a set of points is distributed in the space. In cosmology, these 'points' can be galaxies, dark matter haloes, quasars, Lyman alpha emitters, among other possibilities. In this section we aim to describe the formalism and definitions of these tools.

We start by setting up the connection between correlation functions and probabilities. The simplest correlation function is the 2-point correlation. It can be defined from the joint probability $\delta^2 P_2$ of finding one object in a small volume δV_1 and another object in the volume δV_2, both separated by a vector $\mathbf{r}_{12}$ (Peebles, 1980),

$$\delta^2 P_2 \equiv n_V^2 \left[1 + \xi_2(r_{12})\right] \delta V_1 \delta V_2 \tag{1.10.2}$$

where n_V is the mean number density of points, and $\xi_2(r)$ is the two-point correlation function. Since in this thesis these tools are applied to particles in a homogeneous and isotropic Universe, the function $\xi_2(r_{12})$ does not depend on the direction of $\mathbf{r}_{12}$, only on its modulus. Note that according to Eq. 1.10.2, if the particles were distributed uniformly in the space, then $\xi_2(r_{12})$ would be null. Thus, it is clear that $\xi_2(r)$ represent the excess of probability compared with finding two particles in a uniform random distribution. In the case of $\xi_2(r) > 0$, the particles tend to be clustered at scales of r, whereas if $\xi_2(r) < 0$ particles tend to avoid each other.

In a similar way, higher order correlation functions can be defined. For instance, the 3-point correlation function arises from the probability of finding 3 particles separated by the vectors $\mathbf{r}_{12}$, $\mathbf{r}_{13}$ and $\mathbf{r}_{23}$ [6],

$$\delta^3 P_3 \equiv n_V^3 \left[1 + \xi^{(3)}(r_{12}, r_{13}, r_{23})\right] \delta V_1 \delta V_2 \delta V_3 \tag{1.10.3}$$

where $\xi^{(3)}$ is the *total* 3-point correlation function. However, this function can be written in terms of lower order correlation functions,

$$\xi^{(3)}(r_{12}, r_{13}, r_{23}) = \xi_2(r_{12}) + \xi_2(r_{13}) + \xi_2(r_{23}) + \xi_3(r_{12}, r_{13}, r_{23}) \tag{1.10.4}$$

where now ξ_3 is the so called *connected* 3-point correlation function. As before, this function does not depend on the specific orientation of the vectors due to the isotropy of the space. The connected 3-point function can be simply defined as part of the total 3-point function that does not depend on the two-point correlation functions. In other words, the terms ξ_2 in Eq. 1.10.4 represent the excess numbers of triplets one obtains compared to a random distribution. In the same way, the term ξ_3 is the excess probability respect to the expected just by a distribution with a given 2-point correlation function.

We have set up a connection between the probability of finding a set of particles distributed in a specific way and the n-point correlation function. It is useful to discuss the statistical properties of spatial fluctuations of the δ-field and establish the connection between the n-point correlation function and the density field. The two-point correlation function of particles is defined as,

$$\xi_2(r) \equiv \frac{\langle [\rho(\mathbf{x}) - \bar{\rho}][\rho(\mathbf{x}+\mathbf{r}) - \bar{\rho}] \rangle}{\bar{\rho}} \equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle \tag{1.10.5}$$

where $\bar{\rho} \equiv \langle \rho \rangle$ and the average is taken over all values and directions of $\mathbf{x}$ making use of the ergodicity of the δ-field. Now it is clear that the correlation function at r is the same quantity of the covariance matrix at $|\mathbf{x}_i - \mathbf{x}_j|$ described above in Eq. 1.10.1. In the same way we can extend this to higher order statistics. The connected 3-point correlation function can be written as,

$$\begin{aligned} \xi_3(r_{12},\, r_{13},\, |-\mathbf{r}_{12} - \mathbf{r}_{13}|) &\equiv \frac{\langle [\rho(\mathbf{x}) - \bar{\rho}][\rho(\mathbf{x}+\mathbf{r}_{12}) - \bar{\rho}][\rho(\mathbf{x}+\mathbf{r}_{13}) - \bar{\rho}] \rangle}{\bar{\rho}^3} \\ &\equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}_{12}) \delta(\mathbf{x}+\mathbf{r}_{13}) \rangle. \end{aligned} \tag{1.10.6}$$

---

[6] Note that $\mathbf{r}_{12} + \mathbf{r}_{13} + \mathbf{r}_{23} = 0$.

Again, the average is taken over all values and directions of $\mathbf{x}$.

In common cases, it is useful to work in Fourier space instead of the configuration space. The Fourier transform of the density field and its inverse are,

$$\delta(\mathbf{k}) = \int d^3\mathbf{k}\, \delta(\mathbf{x}) \exp(-i\mathbf{k}\cdot\mathbf{x}), \tag{1.10.7}$$

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{x}\, \delta(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{x}). \tag{1.10.8}$$

Thus the two- and three-point Fourier-space correlations are,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(k_1) \tag{1.10.9}$$

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_1, k_2, k_3) \tag{1.10.10}$$

$$\tag{1.10.11}$$

where δ_D is the Dirac delta function, $P(k)$ is the power spectrum, the Fourier transform of the two-point correlation function and $B(k_1, k_2, k_3)$ the bispectrum, the Fourier transform of the connected 3-point correlation function,

$$P(k) = \int d^3\mathbf{x}\, \xi_2(x) \exp(-i\mathbf{k}\cdot\mathbf{x}), \tag{1.10.12}$$

$$\begin{aligned} B(k_1, k_2, |-\mathbf{k}_1 - \mathbf{k}_2|) &= \int d^3\mathbf{r}\, d^3\mathbf{s}\, \xi_3(r, s, |-\mathbf{r} - \mathbf{s}|) \times \\ &\times \exp(-i\mathbf{k}_1\cdot\mathbf{r} - i\mathbf{k}_2\cdot\mathbf{s}). \end{aligned} \tag{1.10.13}$$

The power spectrum is a quantity that describes the clustering of points or particles in a system. On the other hand, the bispectrum also contains information about the spacial distribution and the type of the structures that are formed. In particular, Wick's theorem states that the bispectrum of a system is 0 when the fields are Gaussian. In this sense, the bispectrum is very useful to measure possible departures of Gaussianity.

In all the Chapters of this thesis the power spectrum and bispectrum are the basics statistical quantities used to describe the Universe from a statistical point of view. In particular Chapters 2 and 3 use the power spectrum (or the two-point correlation function), whereas Chapters 4 and 5 are more focused in the bispectrum.

1.10.5 Galaxy bias

The relation between the clustering properties of the dark matter and those of the galaxy field goes under the name of galaxy bias. The simplest bias model is the linear bias (Kaiser, 1984),

$$\delta_g(\mathbf{x}) = b_1 \delta_m(\mathbf{x}) \tag{1.10.14}$$

with b_1 constant and independent of position and scale. This corresponds to a deterministic and linear biasing. The main caveat of this model is that for $b_1 > 1$, the value of $\delta_g(\mathbf{x})$ can take values less than -1 which has no physical meaning by definition of δ. A more complex modelling is certainly needed to properly describe the galaxy clustering. Non-linear biasing with a bias which is no longer a constant but a function of $\delta(\mathbf{x})$ is a common way to proceed. The mean biasing function $b[\delta(\mathbf{x})]$ can be defined as the conditional mean between the galaxy and the matter field,

$$b[\delta(\mathbf{x})]\delta(\mathbf{x}) \equiv \langle \delta_g(\mathbf{x}) | \delta(\mathbf{x}) \rangle = \int d\delta_g(\mathbf{x})\, p[\delta_g(\mathbf{x})|\delta(\mathbf{x})]\delta_g(\mathbf{x}) \tag{1.10.15}$$

where $p[\delta_g(\mathbf{x})|\delta(\mathbf{x})]$ is the probability of finding $\delta_g(\mathbf{x})$ given $\delta(\mathbf{x})$. This is the natural generalisation of the deterministic linear biasing relation. Now $b[\delta(\mathbf{x})]$ characterises the non-linear bias behaviour. The function $b(\delta)$ can be characterised by the first- and second-order moment which, describe respectively the ratio among galaxy-galaxy and galaxy-matter correlation function respect to matter-matter one (Dekel and Lahav, 1999),

$$\hat{b}(r) \equiv \frac{\langle \delta(\mathbf{x}+\mathbf{r})\delta(\mathbf{x})b[\delta(\mathbf{x})] \rangle}{\langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle}, \tag{1.10.16}$$

$$\tilde{b}^2(r) \equiv \frac{\langle \delta(\mathbf{x}+\mathbf{r})\delta(\mathbf{x})b[\delta(\mathbf{x})]b[\delta(\mathbf{x}+\mathbf{r})] \rangle}{\langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})}. \tag{1.10.17}$$

These two parameters take into account the non-linearity of the system as long as one is concerned with the two-point correlation function. For the three-point correlation function 3 more parameters would be needed: galaxy-galaxy-galaxy, galaxy-galaxy-matter, galaxy-matter-matter.

Stochastic bias modelling is also needed to account for any physical or statistical process that produces a non-deterministic biasing relation between galaxies and dark matter. These processes may arise from the discrete nature of galaxies in which case it is called shot noise; if it is a Poisson process, its expression is inversely proportional to the mean density of objects. However any other process that may include all the physics in galaxy formation may produce also a stochastic relation between galaxies and dark matter. In order to account for these phenomena, a random bias field $\epsilon(\mathbf{x})$ can be defined as the differences between the galaxy field and the biased dark matter field,

$$\epsilon(\mathbf{x}) \equiv \delta_g(\mathbf{x}) - b[\delta(\mathbf{x})\delta(\mathbf{x})]. \tag{1.10.18}$$

Non-linear and stochastic bias modelling is necessary when we work with dark matter tracers (even at large scales) and accurate and realistic predictions are needed. We explore in detail the non-linear and stochastic bias and its effects on measuring cosmological parameters in Chapter 3.

CHAPTER 2

A halo model with environment dependence: theoretical considerations

In this Chapter we present the paper Gil-Marín et al. (2011a). In this work we develop of an extension of the halo model focusing on the possibility of modifying the standard halo model introducing an extra feature on haloes besides its mass: a halo environment. Thus we have more freedom to populate haloes with galaxies according to not only the halo-mass but also the halo-environment. We explore how this new property can affect the two-point correlation function at linear and non-linear scales.

This paper was done in collaboration with Raúl Jiménez and Licia Verde from the Institut de Ciències del Cosmos at the Universitat the Barcelona, Spain, and was published in Monthly Notices of the Royal Astronomical Society (MNRAS) in January 2011.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **414,** 1207–1226 (2011) doi:10.1111/j.1365-2966.2011.18456.x

A halo model with environment dependence: theoretical considerations

Héctor Gil-Marín,[1]★ Raul Jimenez[2]★ and Licia Verde[2]★

[1]*Institute of Space Sciences (IEEC-CSIC), Faculty of Science, Campus UAB, Bellaterra 08193, Spain*
[2]*ICREA & Institute of Sciences of the Cosmos (ICC), University of Barcelona, Barcelona 08024, Spain*

Accepted 2011 January 31. Received 2011 January 31; in original form 2010 August 26

ABSTRACT

We present a modification of the standard halo model with the goal of providing an improved description of galaxy clustering. Recent surveys, like the Sloan Digital Sky Survey (SDSS) and the Anglo-Australian two-degree survey (2dF), have shown that there seems to be a correlation between the clustering of galaxies and their properties such as metallicity and star formation rate, which are believed to be environment-dependent. This environmental dependence is not included in the standard halo model where the host halo mass is the only variable specifying galaxy properties. In our approach, the halo properties, i.e. the concentration, and the halo occupation distribution (HOD) prescription, will not only depend on the halo mass (like in the standard halo model) but also on the halo environment. We examine how different environmental dependence of halo concentration and HOD prescription affects the correlation function. We see that at the level of dark matter, the concentration of haloes moderately affects the dark matter correlation function only at small scales. However, the galaxy correlation function is extremely sensitive to the HOD details, even when only the HOD of a small fraction of haloes is modified.

Key words: galaxies: haloes – cosmological parameters – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

The modern language to analytically describe the clustering of galaxies is the halo model (e.g. Cooray & Sheth 2002 for a review). In its original formulation, the halo model describes non-linear clustering of dark matter and can be applied to describe also the clustering properties of galaxies.

In particular, the halo model can be calibrated to describe galaxy clustering properties in several different ways, depending on how the observed properties of galaxies are to be related to the underlying dark matter halo – the so-called halo occupation distribution (HOD): using galaxy abundance, their spatial distribution via the two-point correlation function or the luminosity dependence as a function of the halo mass. In the HOD, it is customary to classify galaxies as central or satellite. These two classes of galaxies are then assigned different occupation numbers in the dark haloes. In this description, all observational properties are completely specified by the dark matter halo mass. The original halo model has been remarkably successful at describing the first moment statistics of the clustering of galaxies.

In reality, however, galaxies are not easily divided in central or satellite, and the physical characteristics of galaxies of each type cannot be determined solely by the mass of the host halo: galaxies are more complex systems. Environment must play an important role in the process of galaxy formation, the most striking observational evidence being that clusters today have a much higher fraction of early-type galaxies than is found in the field. It has been known for more than three decades that there is a relation between galaxy morphology and density of the local environment starting from the results of Davis & Geller (1976), Dressler (1980), Postman & Geller (1984), until most recent results (see e.g. Hogg et al. 2002; Zehavi et al. 2010). It is not clear if this effect can be completely ascribed to the fact that the most massive haloes are naturally found in overdense regions, and that most massive haloes form on average earlier, or if there is some extra environmental dependence, i.e. physical mechanisms such as ram-pressure stripping, harassment, etc. (Gunn & Gott 1972; Moore et al. 1996; Moore, Lake & Katz 1998) shape the properties of galaxies and operate in dense environments. We know that, although *on average* the star formation and metallicity history are determined by the mass, their correlation properties are not (Sheth et al. 2006; Mateus, Jimenez & Gaztañaga 2008). In particular, the clustering of the properties of galaxies has been shown to depend on more parameters than mass: large-scale tidal fields

---

★E-mail: gil@ieec.uab.es (HG-M); raul.jimenez@icc.ub.edu (RJ); iciaverde@icc.ub.edu (LV)

have been shown to alter galactic spins (Jimenez et al. 2010) and ellipticity (Mandelbaum et al. 2006) but can also alter galaxy clustering properties (Hirata 2009; Krause & Hirata 2011).

The fact that clustering properties of galaxies depend on galaxy internal properties can be understood if dark matter haloes with different properties and formation histories cluster differently and host different galaxy populations. Consider, for example, the so-called halo assembly bias; e.g. Gao, Springel & White (2005) found that the amplitude of the correlation function depends on halo formation time, thus haloes assembled at high redshift are more clustered than those formed recently. Models of galaxy clustering statistics make some simplifying assumptions; mainly that (i) number and properties of galaxies populating a dark matter halo depend only on the mass of the host halo, and (ii) clustering properties of dark matter haloes are only a function of their mass and not of the larger environment (this assumption is at the basis of the excursion set formalism).

It is well known that the formation time of haloes depends on the halo mass and that objects formed at early times tend, on average, to be more concentrated than objects that formed recently. In the hierarchical galaxy formation model, there is a correlation between galaxy type and environment induced by the fact that the mass function of dark haloes in dense regions in the Universe is predicted to be top-heavy. In a simple improvement of the classic halo model (Abbas & Sheth 2005, 2006) the most massive haloes only populate the densest regions and a correlation between halo abundance and environment is introduced. This approach matches the prediction of the excursion set approach; all correlation with environment comes from mass.

But there are indications that there could be a more complex dependence on the environment which is not fully described by halo mass (see e.g. Sheth & Tormen 2004 and discussion above). It has been known for a while that the star formation history is perhaps the most affected quantity by environment, this is the so-called downsizing effect (see e.g. Heavens et al. 2004; Cowie & Barger 2008). Furthermore, the recent Large redshift Cosmological Evolution Survey (zCOSMOS) has shown how other variables are also affected by environment, in particular the shape of galaxy stellar mass function (Tasca et al. 2009; Bolzonella et al. 2010), and the formation of red galaxies at the low-mass end (Zucca et al. 2009; Cucciati et al. 2010; Galaz et al. 2011). All these new observational results seem to indicate that the physical properties of galaxies are, at least in part, determined by the environment in which they live, and that this dependence goes beyond the fact that the halo mass function is influenced by the local background density. Therefore, the analyses of future surveys with a similar sensitivity as the zCOSMOS and, especially, large redshift coverage may benefit from theoretical tools that can include environmental dependence. In practice, the definition of ‘environment’ may not be easy or univocal, and the definition may change as a function of the galaxy property under consideration. Here, we simply set up the mathematical description assuming that ‘environment’ has been already defined. We extend the standard halo model to include an extra dependence on the environment, classified as ‘cluster/node’, ‘filament’ and ‘void’ although the treatment is general enough that other choices or interpretations of environment are possible. Motivated by the excursion-set model expectations, the extra dependence comes through the halo density profile and mass function for the dark matter and also through the HOD parameters for galaxies. In the excursion-set model this is given by the formation time but in our model we introduce an extra parameter set by the environment, leaving us one more degree of freedom to tune the HOD. As a starting point, here we introduce the formalism and show the effect of each of the parameters of the model on the correlation function. However we expect that the full potential of the model and the relevance of its features will appear when considering clustering statistics beyond the two-point function of the overdensity field such as e.g. the marked correlation functions (Skibba et al. 2006; Sheth et al. 2006; Mateus et al. 2008). We defer this to the forthcoming work (Gil-Marin et al. in preparation).

This paper is organized as follow. In Section 2 we present the theoretical formulation of our extension of the model. We first introduce all the model parameters and ingredients, and then in Section 2.4 we present the expression for the correlation function of dark matter and galaxies. In Section 3, we analyse how scenario environments with different properties can affect the correlation function. In Section 4, we present the conclusions of our work. Accordingly in the Appendices we review the basics of the standard halo model and the HOD and give details of equations presented in Section 2.

2 THE MODEL

The halo model provides a physically motivated way to estimate the two-point (and higher-order) correlation function of dark matter and galaxy density field. We review the classic halo model in Appendix A. Despite its simplicity, the model is extremely successful: the model’s predictions have been compared with both simulations and observations, reproducing well the clustering properties. Nevertheless, it presents some limitations and shortcomings which we discuss next. The background material presented in Appendix A is useful to set up the stage of motivating our extension of the model and to define symbols and nomenclature used. We thus refer the reader to Appendix A for definitions of many of the symbols used.

In the halo model every halo is characterized only by its mass. This is correct at the leading order, and has been so far sufficient. Much more accurate measurements of dark matter and galaxies clustering will be available in the near future (e.g. Sloan Digital Sky Survey III [SDSSIII], EUCLID, Joint Dark Energy Mission [JDEM], Dark Energy Survey [DES], Big Barion Oscillation Spectroscopic Survey [Big Boss],[1] PannStarr, Large Synoptic Survey Telescope [LSST],[2] etc.) and a more sophisticated modelling may be needed to describe such data

---

[1] See http://bigboss.lbl.gov/index.html and http://arxiv.org/abs/0904.0468 for mote details.

[2] See http://www.lsst.org/lsst for more details.

sets. In fact, there are already indications that the properties of galaxies depend somewhat on the environment, and not only on the host halo mass (see Section 1).

It is clear that the mass has to be the primary variable in any halo model formalism. However in order to include some environmental dependence another variable needs to be introduced. At the level of the galaxy correlation function this ‘variable’ can be modulating the HOD: in some environments galaxies can populate haloes in a different way than in other. At the level of dark matter we can introduce an environmental dependence through the concentration of the density profile. It has been observed that the relation between the mass and the concentration has a very high dispersion: the distribution around the mean concentration is approximately independent of halo mass, and is well approximated by a lognormal with rms $\sigma_{\ln c} = 0.3$ (e.g. Sheth & Tormen 2004). In the halo model formulation a relation between concentration c and mass m is assumed (see equation A15), however this high dispersion may indicate that the concentration depends on extra (hidden) variables, such as halo formation history (haloes which form at high redshift have a higher concentration), tidal forces, etc., that in the end would depend on the environment. A possible approach to this problem was presented by Giocoli et al. (2010). It consists in treating the concentration as a stochastic variable and integrate not only over the mass but also over all possible values of the concentration. In this work we will instead consider that the relation between concentration and mass may depend on the environment.

In our model, we assume that the Universe contains three kinds of structures or environments: nodes, filaments and voids. Haloes lie either in a node or in filament regions. Voids contain no haloes but occupy a large fraction of the Universe’s volume, especially at $z \simeq 0$.

At the level of dark matter correlation function, we distinguish between haloes in the node regions (hereafter node-like haloes) and haloes in the filament regions (hereafter filament-like haloes) through the concentration of their profile. We assume that the node-like haloes have a profile with a concentration function $c_{\rm nod}(m)$ and filament-like haloes with $c_{\rm fil}(m)$. For simplicity, in presenting our equations and our figures, we assume that these concentrations are constants, but the formalism can straightforwardly be generalized to allow these concentrations to be functions of the mass, as it is done in the halo model. The range of these values is expected to be between $\sim$1 and $\sim$15; these values (or functions) could be calibrated by N-body simulations.

At the level of galaxy correlation function, we can decide to populate haloes in different ways, according to the type region where they lie. For instance, node-like haloes may have less satellite galaxies, or more likely to have a central galaxy than filament-like haloes.

In this section we show how we can mathematically modify the halo model in order to take into account all these effects.

2.1 Mass function

As in the standard halo model (see Appendix A for details), the halo number density is given by the mass function, which can be computed analytically in the extended Press–Schechter approach or calibrated on N-body simulations. For this application we will adopt the Sheth & Tormen (1999) mass function. More massive haloes are more likely to be found in high-density environments, thus in principle the mass function can be made to be dependent on the environment and the dependence could be calibrated on N-body simulations. At this stage, this procedure is similar to the approach presented by Abbas & Sheth (2005, 2006). For this application the environment-dependence of the mass function is accounted for through the fact that the mass function depends on the mean matter density: $\bar{\rho}_{\rm m}$, which we consider to be dependent on the local environment. However, this model is general enough to use any mass function with any environmental dependence. Let us consider a region i with a volume V_i and with a mean matter density $\bar{\rho}_{m_i}$. If we consider this region of the Universe alone, then the number of objects of mass m in this i region per unit of volume (of this region) and per unit of mass, namely $n_i(m)$, is related to the mass function of the whole Universe as (Abbas & Sheth 2005)

$$n_i(m) = [1 + b(m)(\mathcal{Y}_i - 1)]\, n(m), \tag{1}$$

where $\mathcal{Y}_i$ is the ratio between the mean matter densities of the region i and of the whole Universe: $\mathcal{Y}_i \equiv \bar{\rho}_{m_i}/\bar{\rho}_{\rm m}$. It is also useful to define the volume fraction of this region i, V_i and the volume of the whole (observable) Universe V as $\mathcal{X}_i \equiv V_i/V$. Conservation of mass imposes that the consistency relation must be satisfied:

$$\sum_i \mathcal{X}_i \mathcal{Y}_i = 1, \tag{2}$$

where the summation runs over all kinds of structures, in our case nodes, filaments and voids. Since we assume that there are no haloes in voids ($\mathcal{Y}_v = 0$), this equation reads: $\mathcal{X}_n \mathcal{Y}_n + \mathcal{X}_f \mathcal{Y}_f = 1$.[3] In other words, since the mass function is proportional to the mean matter density (see equation A1), if we consider a region with a mean matter density $\mathcal{Y}_i$ times denser than the mean density of the Universe, then the mass function of this region will be $\mathcal{Y}_i$ times the one of the whole Universe (see equation 1).

For illustration, here the adopted values for these parameters are obtained from Aragón-Calvo, van de Weygaert & Jones (2010) and are listed in Table 1.[4] In this work, we choose these numbers as a fiducial values, but other values are possible. In fact, the values of these parameters will depend mainly on the definition of environment, and on the definition of what a node and a filament is.

The filament regions are more abundant than the node regions, but less dense. The voids occupy almost 90 per cent of the volume.

---

[3] Here the sub-indices n, f and v account for the regions nodes, filaments and voids, respectively.

[4] In their work, the authors split their haloes in four types: clusters, filaments, walls and voids. Here, we adopt the node values of $\mathcal{X}$ and $\mathcal{Y}$ as their clusters haloes. For the voids we adopt $\mathcal{X}$ but we set $\mathcal{Y}$ to 0, and we combine their filaments and wall values of $\mathcal{X}$ and $\mathcal{Y}$ into what we call filaments.

Table 1. Volume fraction $\mathcal{X}$ and mean density fraction $\mathcal{Y}$. Values for node-, filament-like regions and voids used in this work. Data are from Aragón-Calvo et al. (2010).

	$\mathcal{X}$	$\mathcal{Y}$	$\mathcal{X} \cdot \mathcal{Y}$
Nodes	3.8×10^{-3}	73	0.2774
Filaments	0.1368	5.28	0.7223
Voids	$\sim$0.86	0	0

2.2 Halo density profile

In the halo model, the halo density profile can be written as a function of mass and concentration: $\rho(r|m, c)$. Usually, the NFW profile (Navarro, Frenk & White 1996, 1997) is adopted, with an empirical relation between the concentration and the mass such as equation (A15). For this application we will consider that these two variables are independent and that the concentration is set by the environment of the halo. Thus, we have two different profiles depending on whether the halo lies in a node-like region, $\rho(r|m, c_{\rm nod})$ or in a filament-like one, $\rho(r|m, c_{\rm fil})$. In principle, $c_{\rm nod}$ and $c_{\rm fil}$ can be functions of the mass, which could be calibrated from N-body simulations. Here, however, for simplicity, we will consider that these two variables are just constants.

2.3 Halo occupation distribution

In our approach, the standard HOD (described in Appendix A Section A6) must also be modified to include a possible dependence on environment.

As an example, see the discussion in Zehavi et al. (2005): blue and red galaxies populate haloes of same mass in a different way (see Appendix A Section A6 and Fig. A4); blue galaxies tend to occupy low-density regions while red galaxies the high-density ones.

Here, we are interested in modelling generic properties of galaxies (star formation, metallicity, etc.) not just colours. Whatever the property under consideration is, in our modelling galaxies can still be divided in two classes (nodes and filaments); for simplicity here we still call them ‘red’ and ‘blue’, but one should bear in mind that the argument is much more general. Thus, we make the HOD to depend not only on the mass of the host halo but also on its environment (node or filament) by adopting two different prescriptions for nodes or filaments: the parameters $M_{\rm min}$, M_1 and α of equations (A24) and (A25) must be specified for node- and filament-like haloes. Let us continue with the working example of red and blue galaxies. While in the standard halo model, the red and blue galaxies are uniformly distributed inside haloes, in the extreme case of this environmental dependence, blue galaxies only populate haloes which are in filament regions while red galaxies only populate node-like haloes. However, this is an extreme segregation and in a more realistic scenario there will be some mixing: node-like haloes host some blue galaxies and filament-like haloes can host some red ones. In order to describe this, we introduce the segregation index $\mathcal{S}$. When $\mathcal{S} = 0$, there is no segregation, i.e. red and blue galaxies are distributed equally in filament and node haloes: this corresponds to the standard approach with no environmental dependence. In the other limit, if $\mathcal{S} = 1$, blue galaxies live in filament haloes and red ones in node haloes: this is the extreme case of our extended halo model with environmental dependence.

Thus our adopted HOD equations are (see Appendix A for definition of HOD)

$$\begin{aligned} g^i_{\rm nod}(m) &= \left(\frac{1+\mathcal{S}}{2}\right) g^i_{\rm red}(m) + \frac{1-\mathcal{S}}{2} g^i_{\rm blue}(m) \\ g^i_{\rm fil}(m) &= \left(\frac{1+\mathcal{S}}{2}\right) g^i_{\rm blue}(m) + \frac{1-\mathcal{S}}{2} g^i_{\rm red}(m), \end{aligned} \tag{3}$$

where $\mathcal{S}$ is the segregation index, $g_{\rm property}(m)$ denotes the average number of galaxies of a certain *property* in a halo of mass m and $g_j(m)$ denotes the average number of galaxies in nodes ($j \to$ nod) and in filaments ($j \to$ fil). This $g(m)$ function is given by equations (A24) and (A25) for central ($i \to$ cen) and satellite galaxies ($i \to$ sat), respectively. These equations have three parameters: the minimum mass for a halo to host one (central) galaxy, $M_{\rm min}$; the average mass for a halo to host its first satellite galaxy, M_1; the power-law slope of the satellite mean occupation function, α. Taking into account that the HOD for red and blue galaxies may be different there are six HOD parameters. Moreover, if the two populations are not completely segregate but there is a degree of mixing, we have an extra parameter (see Appendix A for more details). In what follows, for some applications we will set $g_{\rm nod} = g_{\rm fil}$ (i.e. $g_{\rm red} = g_{\rm blue}$) but in general the HOD parameters for $g_{\rm red}$ will be different from those of $g_{\rm blue}$ and thus, in general, $g_{\rm nod} \neq g_{\rm fil}$.

In the following sections we will see how having two different HOD for nodes and filament regions modifies the correlation function with respect to the case of having the same HOD for all galaxies.

2.4 Two-point correlation function

Following the above assumptions, we can compute the two-point correlation function for dark matter and galaxies. Since we have two types of structures (nodes and filaments), the correlation function is split in three terms instead of two as in the halo model[5]:

$$\xi(\boldsymbol{r}) = \xi^{1h1\eta}(\boldsymbol{r}) + \xi^{2h1\eta}(\boldsymbol{r}) + \xi^{2h2\eta}(\boldsymbol{r}), \tag{4}$$

---

[5] The notation in this paper is the following: 1η means that both particles belong to the same kind of halo whereas 2η to a different one.

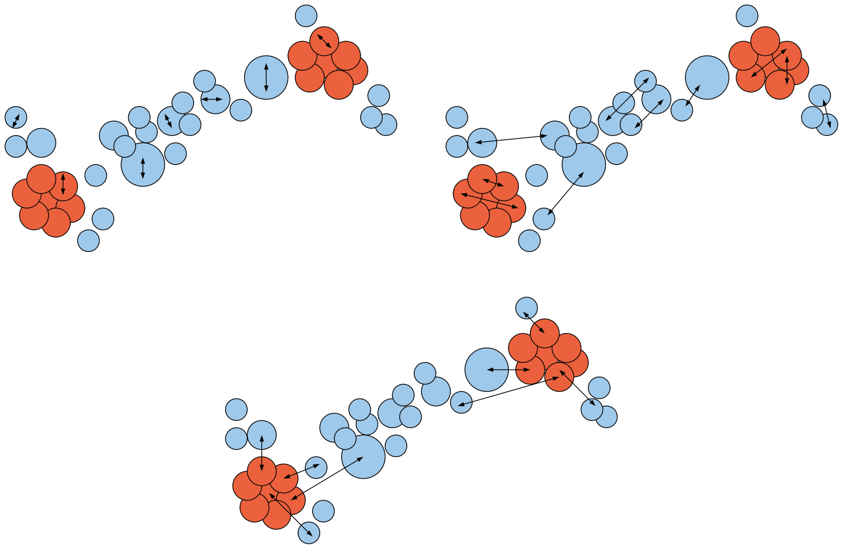

Figure 1. In our modified halo model, dark matter haloes are of two kinds: node-like (red circles) and filament-like (blue circles). Because of this, our correlation function is split in three terms: the one-halo-one-η term (top-left picture) describes the interaction of particles inside the same halo, the two-halo-one-η (top-right picture) term involves particles which are in different haloes of the same kind. Finally, the two-halo-two-η term (bottom picture) considers the contribution of particles which are in different haloes of different kinds.

where now $\xi^{1h1\eta}(\boldsymbol{r})$ accounts for particles in the same halo, $\xi^{2h1\eta}(\boldsymbol{r})$ for particles in different haloes but of the same kind and $\xi^{2h2\eta}(\boldsymbol{r})$ for particles in different haloes and of different kind separated by a distance $\boldsymbol{r}$. In Fig. 1, we show schematically the contribution of these different terms.

2.4.1 Dark matter correlation function

For dark matter particles, the correlation function reads

$$\xi_{\rm dm}(\boldsymbol{r}) = \xi_{\rm dm}^{1h1\eta}(\boldsymbol{r}) + \xi_{\rm dm}^{2h1\eta}(\boldsymbol{r}) + \xi_{\rm dm}^{2h2\eta}(\boldsymbol{r}), \tag{5}$$

where each terms is (see Appendix B for derivation)

$$\xi_{\rm dm}^{1h1\eta}(\boldsymbol{r}) = \sum_{i=1}^{2}\int {\rm d}m\, \mathcal{X}_i \frac{m^2 n_i(m)}{\bar\rho_{\rm m}^2}\int_{V_i} {\rm d}^3\boldsymbol{x}\, u(x|m, c_i)u(|\boldsymbol{x}+\boldsymbol{r}||m, c_i) \tag{6}$$

$$\begin{aligned}\xi_{\rm dm}^{2h1\eta}(\boldsymbol{r}) = \sum_{i=1}^{2}\int {\rm d}m'\, \mathcal{X}_i \frac{m' n_i(m')}{\bar\rho_{\rm m}}\int {\rm d}m''\, \mathcal{X}_i \frac{m'' n_i(m'')}{\bar\rho_{\rm m}}\int_{V_i} {\rm d}^3\boldsymbol{x}'\, u(x'|m', c_i)\int_{V_i} {\rm d}^3\boldsymbol{x}''\, u(x''|m'', c_i)\\ \times\, \xi_{hh}(|\boldsymbol{x}'-\boldsymbol{x}''+\boldsymbol{r}||m', m'')\end{aligned} \tag{7}$$

$$\begin{aligned}\xi_{\rm dm}^{2h2\eta}(\boldsymbol{r}) = 2\int {\rm d}m'\, \mathcal{X}_1 \frac{m' n_1(m')}{\bar\rho_{\rm m}}\int {\rm d}m''\, \mathcal{X}_2 \frac{m'' n_2(m'')}{\bar\rho_{\rm m}}\int_{V_1} {\rm d}^3\boldsymbol{x}'\, u(x'|m', c_1)\int_{V_2} {\rm d}^3\boldsymbol{x}''\, u(x''|m'', c_2)\\ \times\, \xi_{hh}(|\boldsymbol{x}'-\boldsymbol{x}''+\boldsymbol{r}||m', m''),\end{aligned} \tag{8}$$

where $i = 1$ refers to nodes and $i = 2$ to filaments. Here, $n(m)$ is the halo mass function, $u(x|m, c)$ is the normalized density profile of the halo of mass m and concentration parameter c at a radial distance x (see equation A13), $\bar\rho_{\rm m}$ is the mean matter density, $\xi_{hh}(d|m_1, m_2)$ is the two-point correlation function of two haloes of masses m_1 and m_2 separated by a distance d (see equation A5 for the model used here). and $\mathcal{X}_i$ and $\mathcal{Y}_i$ are the volume and density fractions defined in Section 2.1. The integration $\int_V {\rm d}^3\boldsymbol{x}$ runs over all haloes' volume and $\int {\rm d}m$ runs over all halo mass range (see Appendix A for computational details). In Fig. 2, we show the contribution of the different terms. Here we have assumed $c_{\rm nod} = 10$ and $c_{\rm fil} = 2$, just as an example to introduce the environmental dependence. According to the adopted values of $\mathcal{X}$ and $\mathcal{Y}$, the effect of the nodes is dominant at small scales ($r < 0.3$ Mpc h^{-1}) (red line in left-hand panel). At intermediate scales ($0.3 < r < 6$ Mpc h^{-1}), the filaments dominate and at large scales ($r > 6$ Mpc h^{-1}) both filaments and the cross-term play an important role in the total correlation function (blue and green lines in left-hand panel). This is expected due to the fact that the filaments are more abundant than nodes but nodes are more concentrate than filaments.

In the right-hand panel, we show the contribution of the different terms of equation (5). As we expected, the $1h1\eta$-term (red line) dominates at sub-halo scales whereas $2h1\eta$-term (blue line) dominates at large scales. The cross-correlation term between nodes and filaments, $2h2\eta$-term (green line), is less important than filaments at large scales but is not negligible. In the left-hand panel, the shape of the filament and the node contribution are set by the value of the concentration we have chosen. We will come back to this point in the next

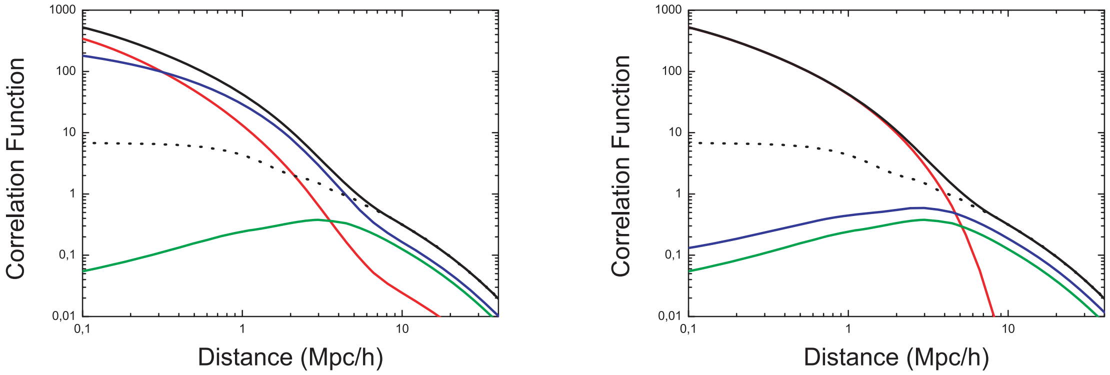

Figure 2. Different contributions to the dark matter correlation function according to our model (equations 6–8). Left-hand panel: total correlation function $\xi_{\rm dm}$ (black solid line), node contribution $\xi^{\rm nod}_{\rm dm}$ (red line), filament contribution $\xi^{\rm fil}_{\rm dm}$ (blue line), cross-contribution $\xi^{2h2\eta}_{\rm dm}$ (green line) and linear dark matter contribution $\xi_{\rm lin}$ (black-dotted line). Right-hand panel: total correlation function $\xi_{\rm dm}$ (black solid line), one-halo term $\xi^{1h1\eta}_{\rm dm}$ (red line), two-halo-same-halo term $\xi^{2h1\eta}_{\rm dm}$ (blue line), two-halo-different-halo term $\xi^{2h2\eta}_{\rm dm}$ (green line) and $\xi_{\rm lin}$ (black-dotted line).

section to see how the concentration parameter affects the shape of the correlation function. In both panels the linear correlation function is also plotted (black-dotted line). The ratio between $\xi_{\rm lin}$ and $\xi_{\rm dm}$ at large scales can be defined as an effective large-scale dark matter bias,

$$b_{\rm dm} \equiv \int {\rm d}m \frac{mn(m)}{\bar\rho_h} b(m) = 1; \tag{9}$$

by the definition of $b(m)$ (see equation A7) this dark matter bias is one if we integrate over all range of masses. Also, the effective large-scale bias can be defined for the node and filament contribution; this effective bias is given by

$$b^i_{\rm dm} \equiv \mathcal{X}_i \int {\rm d}m \frac{mn_i(m)}{\bar\rho_{\rm m}} b(m) = \mathcal{X}_i \mathcal{Y}_i\,. \tag{10}$$

Here i stands for either nodes or filaments and the relation $b_{\rm dm} = b^{\rm nod}_{\rm dm} + b^{\rm fil}_{\rm dm} = 1$ is satisfied. Note that the only difference between the dark matter bias of nodes and filaments is due to $\mathcal{X}_i$ and $\mathcal{Y}_i$. We will see that if we rescale the mean density $\bar\rho_{\rm m}$ to the mean density of nodes or filaments ($\mathcal{X}_i \bar\rho_{m_i}$), then the effective bias for nodes and filaments is the same.

Note that when we refer to the node and filament contribution to the total correlation function, we refer to the terms of the sum of equation (5). These terms are different from the correlation function we would obtain if we only took into account the node- or filament-like haloes (we call it 'pure' node and filament contribution). In the first case, each term is divided by the total dark halo matter density squared, $\bar\rho^2_{\rm m}$, while in the second ('pure') case instead it would be divided by the density of node- or filament-like haloes.[6] In Fig. 4 (left-hand panel), the contribution of these 'pure' terms is shown for the dark matter. Note that the shape is the same as in Fig. 2 and the only difference is that the normalization is shifted by a factor $(\mathcal{X}_i \mathcal{Y}_i)^{-2}$.

2.4.2 Galaxy correlation function

In this case, the environmental dependence can be introduced not only through the concentration parameter of the host halo, but also through the HOD: galaxies do not populate in the same way filament- and node-like haloes even if the host halo has the same mass. In other words, if we have two galaxy populations (red and blue) with different HOD, e.g. red galaxies are more abundant in node haloes and blue in filament haloes, then an environmental dependence arises. This is the case presented here.

As before, the galaxy correlation function reads

$$\xi_{\rm gal}(\boldsymbol{r}) = \xi^{1h1\eta}_{\rm gal}(\boldsymbol{r}) + \xi^{2h1\eta}_{\rm gal}(\boldsymbol{r}) + \xi^{2h2\eta}_{\rm gal}(\boldsymbol{r}) \tag{11}$$

where each term is

$$\xi^{1h1\eta}_{\rm gal}(\boldsymbol{r}) = \sum_{i=1}^{2} \int {\rm d}m\, \mathcal{X}_i \frac{n_i(m)}{\bar n^2_{\rm gal}} \left[2\, g^{\rm cen}_i(m) g^{\rm sat}_i(m) u(r|m, c_i) + \left(g^{\rm sat}_i(m)\right)^2 \int {\rm d}^3\boldsymbol{x}\, u(x|m, c_i)\, u(|\boldsymbol{x} + \boldsymbol{r}|m, c_i) \right] \tag{12}$$

$$\begin{aligned} \xi^{2h1\eta}_{\rm gal}(\boldsymbol{r}) = \sum_{i=1}^{2} \int {\rm d}m'\, {\rm d}m''\, \mathcal{X}^2_i \frac{n_i(m') n_i(m'')}{\bar n^2_{\rm gal}} \Bigg[& g^{\rm cen}_i(m') g^{\rm cen}_i(m'') \xi_{hh}(r|m', m'') + 2\, g^{\rm cen}_i(m'') g^{\rm sat}_i(m') \\ & \times \int {\rm d}^3\boldsymbol{x}'\, u(x'|m', c_i) \xi_{hh}(|\boldsymbol{x}' + \boldsymbol{r}||m', m'') + g^{\rm sat}_i(m') g^{\rm sat}_i(m'') \int {\rm d}^3\boldsymbol{x}' {\rm d}^3\boldsymbol{x}''\, u(x'|m', c_i) u(x''|m'', c_i) \\ & \times \xi_{hh}(|\boldsymbol{x}' - \boldsymbol{x}'' + \boldsymbol{r}||m', m'' \Bigg] \end{aligned} \tag{13}$$

---

[6] In order to obtain these 'pure' node and filament terms, we have to multiply the mean density $\bar\rho_{\rm m}$ by the factor $\mathcal{X}_i \mathcal{Y}_i$, in order to obtain the mean density of nodes ($i = 1$) or filaments ($i = 2$).

$$\xi_{\rm gal}^{2h2\eta}(\boldsymbol{r}) = 2\int {\rm d}m'\,{\rm d}m''\,\mathcal{X}_1\mathcal{X}_2\frac{n_1(m')n_2(m'')}{\bar{n}_{\rm gal}^2}\Bigg[g_1^{\rm cen}(m')g_2^{\rm cen}(m'')\xi_{hh}(r|m',m'') + g_1^{\rm cen}(m')g_2^{\rm sat}(m'')$$
$$\times\int {\rm d}^3\boldsymbol{x}''\,u(x''|m'',c_2)\xi_{hh}(|\boldsymbol{x}''+\boldsymbol{r}||m',m'') + g_2^{\rm cen}(m'')g_1^{\rm sat}(m')\int {\rm d}^3\boldsymbol{x}'\,u(x'|m',c_1)\xi_{hh}(|\boldsymbol{x}'+\boldsymbol{r}||m',m'')$$
$$+g_1^{\rm sat}(m')g_2^{\rm sat}(m'')\int {\rm d}^3\boldsymbol{x}'{\rm d}^3\boldsymbol{x}''\,u(x'|m',c_1)u(x''|m'',c_2)\xi_{hh}(|\boldsymbol{x}'-\boldsymbol{x}''+\boldsymbol{r}||m',m'')\Bigg]. \qquad (14)$$

Again $i = 1$ refers to nodes and $i = 2$ to filaments; $g^j(m)$ is the average number of galaxies that lie in a halo of mass m, the superindex j denotes the type of galaxy: central ($j \to cen$) or satellite ($j \to sat$) (see Section A6 for details); $\bar{n}_{\rm gal}$ is the mean number density of galaxies, i.e. the total number of galaxies divided by the total volume,

$$\bar{n}_{\rm gal} = \sum_{i=1}^{2}\mathcal{X}_i\,\bar{n}_{{\rm gal}\,i} \qquad (15)$$

and $n_{{\rm gal}\,i}$ is the mean number density of galaxies inside haloes of type i, i.e. the total number of galaxies inside haloes of type i divided by the volume these haloes occupy,

$$\bar{n}_{{\rm gal}\,i} = \int {\rm d}m\,n_i(m)\left[g_i^{\rm cen}(m) + g_i^{\rm sat}(m)\right]. \qquad (16)$$

As before we can define an effective large-scale galaxy bias as

$$b_{\rm gal} = \sum_{i=1}^{2}\mathcal{X}_i\int {\rm d}m\frac{n_i(m)}{\bar{n}_{\rm gal}}\left(g_i^{\rm cen}(m) + g_i^{\rm sat}(m)\right)b(m). \qquad (17)$$

In this case, the effective large-scale galaxy bias for nodes and filaments is given by

$$b_{\rm gal}^i = \mathcal{X}_i\int {\rm d}m\frac{n_i(m)}{\bar{n}_{\rm gal}}\left(g_i^{\rm cen}(m) + g_i^{\rm sat}(m)\right)b(m). \qquad (18)$$

As before $b_{\rm gal} = b_{\rm gal}^{\rm nod} + b_{\rm gal}^{\rm fil}$. Note that for the galaxy bias, the difference between the node and the filament bias not only depends on the terms $\mathcal{X}_i$ and $\mathcal{Y}_i$ as in the dark matter bias (equation 10), but also on the HOD of each population. This means that even rescaling the mean number of galaxies $\bar{n}_{\rm gal}$ to the mean number of galaxies of nodes or filaments, the biases will be different as we will see in the next section. The behaviour of these terms is illustrated in Fig. 3. For simplicity $c_1 = c_2 = c(m)$ given by equation (A15) is adopted, along with the HOD for galaxies introduced by Zehavi et al. (2005): $\log_{10}M_{\rm min} = 12.72$, $\log_{10}M_1 = 14.08$ (the masses are in ${\rm M}_\odot\,h^{-1}$), $\alpha = 1.37$, $f_0^{\rm cen} = 0.71$, $f_0^{\rm sat} = 0.88$, $\sigma_M^{\rm cen} = 0.30$ and $\sigma_M^{\rm sat} = 1.70$ (see Appendix A Section A6 for a detailed definition of these HOD parameters) with maximum segregation: $\mathcal{S} = 1$. This means that blue galaxies only populate filaments and red galaxies only populate nodes. The left-hand panel shows the correlation functions of nodes (red line), filaments (blue line), cross-term (green line) and the total (black solid line). In the right-hand panel, the different lines correspond to the terms of equation (11): $1h1\eta$ (red line), $2h1\eta$ (blue line), $2h2\eta$ (green line) and the total contribution (black solid line). For comparison, the dark matter correlation function (black-dashed line) and the linear power spectrum (black-dotted line) are also shown.

In the left-hand panel of Fig. 3 we show the effect on the total correlation function (black solid line) of red galaxies (red line), blue galaxies (blue line) and the cross-term (green line). We see that according to this HOD, red galaxies dominate the one-halo term ($r <$ 4 Mpc h^{-1}), whereas all three terms contribute to the two-halo term, the cross-term being the most important.

We can also compute the 'pure' node and filament galaxy terms, as we did in the dark matter case. These terms are shown in Fig. 4 (central panel). As before, the shape is the same in both figures but in the second one, the lines normalization is offset by $(\mathcal{X}_i\mathcal{Y}_i)^{-2}$. In the case

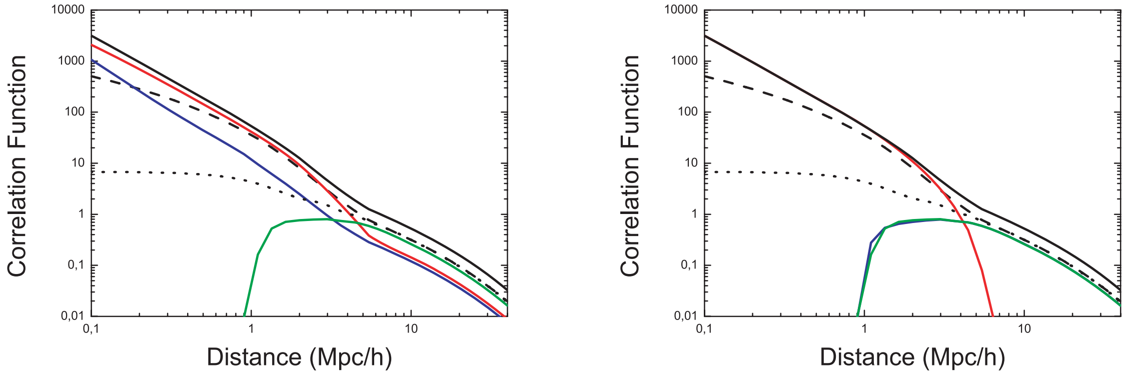

Figure 3. Same notation as in Fig. 2. Left-hand panel: $\xi_{\rm gal}$ (black solid line), $\xi_{\rm gal}^{\rm nod}$ (red line), $\xi_{\rm gal}^{\rm fil}$ (blue line) and $\xi_{\rm gal}^{2h2\eta}$ (green line). Right-hand plot: $\xi_{\rm gal}$ (black solid line), $\xi_{\rm gal}^{1h1\eta}$ (red line), $\xi_{\rm gal}^{2h1\eta}$ (blue line) and $\xi_{\rm gal}^{2h2\eta}$ (green line). In both panels $\xi_{\rm dm}$ (black dashed line) and $\xi_{\rm lin}$ (black-dotted line).

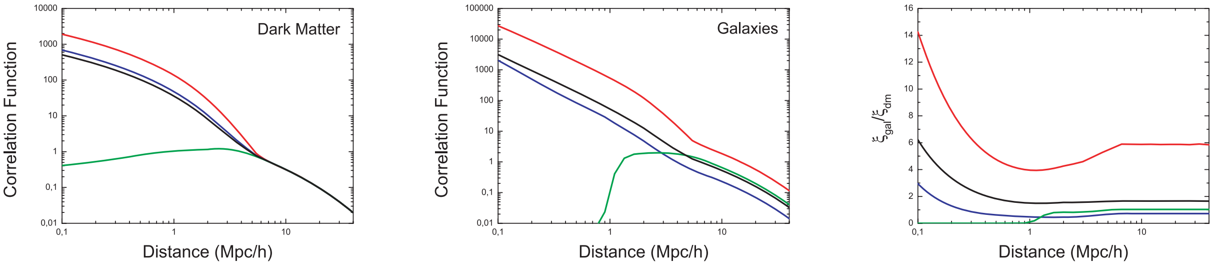

Figure 4. Separate contribution for node- (red lines), filament-like (blue lines) haloes, the cross-term (green line) for dark matter (left-hand panel) and galaxies (central panel). The ratio between galaxies and dark matter correlation function is shown in the right-hand panel. Black lines correspond to the total contribution (node+filaments+cross).

of the $2h2\eta$ term, the rescaling goes as $\xi_{\rm gal}^{2h2\eta} \rightarrow \xi_{\rm gal}^{2h2\eta}/(2\mathcal{X}_1\mathcal{Y}_1\mathcal{X}_2\mathcal{Y}_2)$ in order to obtain cross-bias at large scales defined as $b^{\rm cross} \equiv \sqrt{b^{\rm nod}b^{\rm fil}}$. This is due to the mean density definition as it is explained below.

In Fig. 4, we show the 'pure' contribution of the node and filament terms for dark matter (left-hand panel) and galaxies (central panel). The correlation function for nodes/red galaxies is considerably higher than that for filaments/blue galaxies at all scales. This is just an effect of rescaling the mean density/number of galaxies from the total mass/galaxies over the total volume to the mass/galaxies in filaments or nodes over the total volume. This rescaling goes like $\mathcal{X}_i\mathcal{Y}_i$ for the mean density/number of galaxies and like $(\mathcal{X}_i\mathcal{Y}_i)^{-2}$ for ξ. Since at small scales $\xi \sim \mathcal{X}_i\mathcal{Y}_i$, the ratio between the node and the filament 'pure' ξ terms is just $(\mathcal{X}_f\mathcal{Y}_f)/(\mathcal{X}_n\mathcal{Y}_n) \simeq 2.5$ and the pure correlation function for nodes is 2.5 times higher than that of filaments at small scales in the case of dark matter and a bit different in the case of galaxies due to the different HOD, as can be seen in Fig. 4 (left-hand and central panel). However, at large scales $\xi \sim (\mathcal{X}_i\mathcal{Y}_i)^2$, which compensates for the rescaling of the mean density/number of galaxies. Thus, in the case of dark matter both filament and node pure correlation function are exactly the same since both effective large-scale bias are the same after the rescaling (see equation 10). However, in the case of the galaxy correlation function, the effective large-scale bias is considerably different for nodes and filaments because they have different HOD (see equation 18), being higher for red galaxies than for blue ones. In the right-hand panel the ratio $\xi_{\rm gal}/\xi_{\rm dm}$ is shown for nodes (red line), for filaments (blue line), for the cross-term (green line) and for the whole sample (black line). Both at small and large scales the ratio is higher for node-like haloes.

3 ANALYSIS

In this section we perform a more exhaustive analysis to how HOD parameters, segregation and concentration affect the dark matter and galaxy correlation function. We analyse how every parameter of the HOD affects the dark matter and galaxy correlation function while the other parameters are kept fixed. We also analyse how the concentration parameter and the segregation affect both dark matter and galaxy correlation function while any other parameter is fixed. To gain insight we first start with the standard halo model (i.e. $\mathcal{S} = 0$ and only one HOD prescription for al haloes) and then we move to our extended halo model with environmental dependence.

3.1 Halo model

Here we adopt equations (A24) and (A25) as a parametrization of the HOD, and we analyse how the different parameters ($M_{\rm min}$, M_1 and α) affect $\xi_{\rm gal}$. Here $M_{\rm min}$ sets the minimum mass for a halo to have galaxies, M_1 is the mass of a halo that on average hosts one satellite galaxy and α is the power-law slope of the satellite mean occupation function (see Appendix A Section A6 for more details). We set the parameters to fiducial values which are very close to the ones proposed by Kravtsov et al. (2004): $M_{\rm min} = 10^{11}\,{\rm M}_{\odot}$, $M_1 = 22 \times 10^{11}\,{\rm M}_{\odot}$ and $\alpha = 1.00$. In the following section we analyse how changing each one of these parameters affects the galaxy correlation function and the ratio $\xi_{\rm gal}/\xi_{\rm dm}$. Since filament galaxies are the most abundant, the effects we find in this case would be qualitatively very similar to the effect of changing the HOD of filament population, keeping the nodes HOD fixed at the fiducial model.

3.1.1 Central galaxies

We start by varying $M_{\rm min}$, the minimum mass of a halo to host at least one galaxy, keeping all the other HOD parameters fixed at their fiducial values. In Fig. 5, the different solid lines correspond to different values of $M_{\rm min}$: 10^{11} (red line), 5×10^{11} (blue line), 10^{12} (green line), 5×10^{12} (pink line) and $10^{13}\,{\rm M}_{\odot}$ (orange line). The left-hand panel is the galaxy correlation function and the right-hand panel is $\xi_{\rm gal}/\xi_{\rm dm}$. For reference, in the left-hand panel, the dark matter correlation function $\xi_{\rm dm}$ is also plotted (black-dashed line).

We can see that at large scales the galaxy correlation function is just a boost of the dark matter correlation function, whereas at small scales the shape is modified. The enhancement of the amplitude of the correlation function at large scales can be explained just as a redistribution of galaxies: removing galaxies from low-mass haloes ($M < M_{\rm min}$) causes a large-scale bias increase because the function $b(m)$ (equation A7) increases with the mass. Therefore, at these large scales we expect that only the amplitude of the galaxy correlation function changes with $M_{\rm min}$ (and not the shape that is given by $\xi_{\rm lin}$). On the other hand, at small scales the one-halo term of the galaxy correlation function becomes important; this term does not depend explicitly on $\xi_{\rm dm}$, therefore the shape of the correlation function does not have to be the same. The behaviour of Fig. 5 is qualitatively very similar to the findings of Berlind & Weinberg (2002).

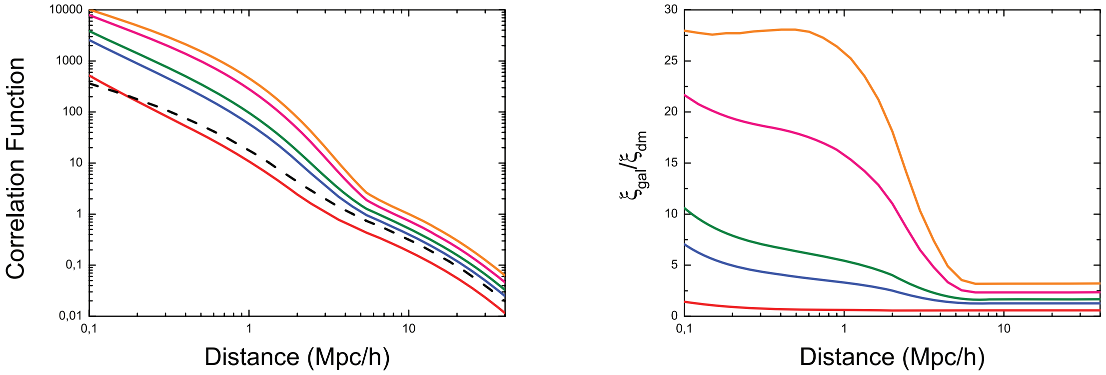

Figure 5. Effect of changing $M_{\rm min}$ in the HOD model. In the right-hand panel the galaxy correlation function $\xi_{\rm gal}$ is shown, whereas in the left-hand panel the ratio $\xi_{\rm gal}/\xi_{\rm dm}$ is shown. The different colours correspond to different $M_{\rm min}$ values: $10^{11}\,{\rm M}_\odot$ (red line), $5\times10^{11}\,{\rm M}_\odot$ (blue line), $10^{12}\,{\rm M}_\odot$ (green line), $5\times10^{12}\,{\rm M}_\odot$ (pink line) and $10^{13}\,{\rm M}_\odot$ (orange line). In the left-hand panel, black-dashed line corresponds to the dark matter correlation function $\xi_{\rm dm}$.

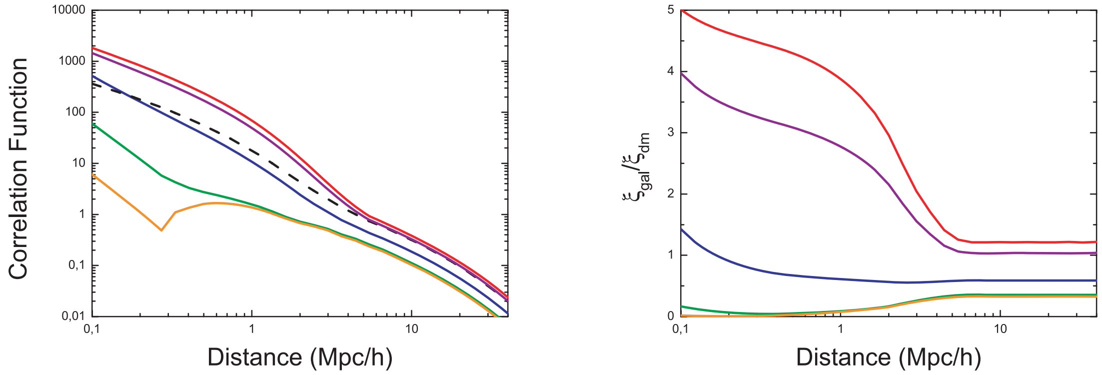

Figure 6. Effects of changing M_1 in the halo model. As in Fig. 5 the left-hand panel is $\xi_{\rm gal}$ and the right-hand panel $\xi_{\rm gal}/\xi_{\rm dm}$. Different colours are different values of M_1: $22\times10^{10}\,{\rm M}_\odot$ (red line), $44\times10^{10}\,{\rm M}_\odot$ (purple line), $22\times10^{11}\,{\rm M}_\odot$ (blue line), $22\times10^{12}\,{\rm M}_\odot$ (green line) and $22\times10^{13}\,{\rm M}_\odot$ (orange line). In left-hand plot $\xi_{\rm dm}$ is also represented (black dashed line).

3.1.2 Satellite galaxies

Here, we keep $M_{\rm min}$ and α fixed to the fiducial values $10^{11}\,{\rm M}_\odot$ and 1.00, respectively, and vary M_1 from $22\times10^{10}\,{\rm M}_\odot$ (red line) to $10^{13}\,{\rm M}_\odot$ (orange line). Recall that this is the minimum mass for a halo to host at least one satellite galaxy.

In Fig. 6, we show the effect of varying this parameter. The left-hand panel shows the galaxy correlation function and the right-hand panel $\xi_{\rm gal}/\xi_{\rm dm}$ for different values of M_1 (see the caption for details). In the left-hand panel, $\xi_{\rm dm}$ is also plotted (black-dashed line). At sub-halo scales we observe that high values of M_1 suppress the correlation function. This effect is expected: the higher M_1, the less satellites galaxies per halo and therefore the one-halo term is suppressed. We also observe that for very high values of M_1 ($>22\times10^{12}\,{\rm M}_\odot$) at very small scales $\xi_{\rm gal}$ shows an inflection point around 0.3 Mpc h^{-1}. This is due to the interaction between the central galaxies of different haloes. Since for these values of M_1 there is (almost) no satellite galaxy per halo, we can observe the interaction between central galaxies of the smallest haloes (those of a mass $\sim M_{\rm min}$), which are the only ones that at that distances can contribute to the two-halo term on those scales. Such feature is similar to the one of the two-halo term of Fig. 3, just with the difference that in that case $M_{\rm min}\sim10^{12.72}\,{\rm M}_\odot$, whereas now $M_{\rm min}=10^{11}\,{\rm M}_\odot$ and therefore such feature happens at shorter distances.

At large scales we also see that the correlation decreases as we increase M_1. This is also expected: according to equation (17), low-mass haloes are weighted more in contributing to the large-scale bias, as one decreases M_1 the large-scale correlation function also decreases. In other words, for high values of M_1 all haloes count the same (one central galaxy per halo) independently of their mass.

Now we consider the variation of α from 0.8 to 1.4, keeping $M_{\rm min}=10^{11}\,{\rm M}_\odot$ and $M_1=22\times10^{11}\,{\rm M}_\odot$. This parameter indicates how the number of satellite galaxies increases with the mass of the halo. A steeper slope α indicates more extra satellites per halo mass increment.

In Fig. 7, we show the effect of changing α in the galaxy correlation function (left-hand panel) and in $\xi_{\rm gal}/\xi_{\rm dm}$ (right-hand panel). The different colours indicate different values for α (see the caption for details) and the black-dashed line is the dark matter correlation function.

As expected, when α increases, the galaxy correlation function also increases, especially at sub-halo scales. At large scales, the effect of increasing α is the same as increasing $M_{\rm min}$: it boosts the dark matter correlation function without changing its shape. In other words, if we increase α we weight more the more massive haloes and therefore the bias increases. At small scales there is also an enhancement of the correlation function for the same reason, but in this case the shape of $\xi_{\rm gal}$ is qualitatively different from the shape of $\xi_{\rm dm}$, because as in the case of changing $M_{\rm min}$, there is not direct relation between the one-halo terms of $\xi_{\rm gal}$ and $\xi_{\rm dm}$. This behaviour was also noted by Berlind & Weinberg (2002).

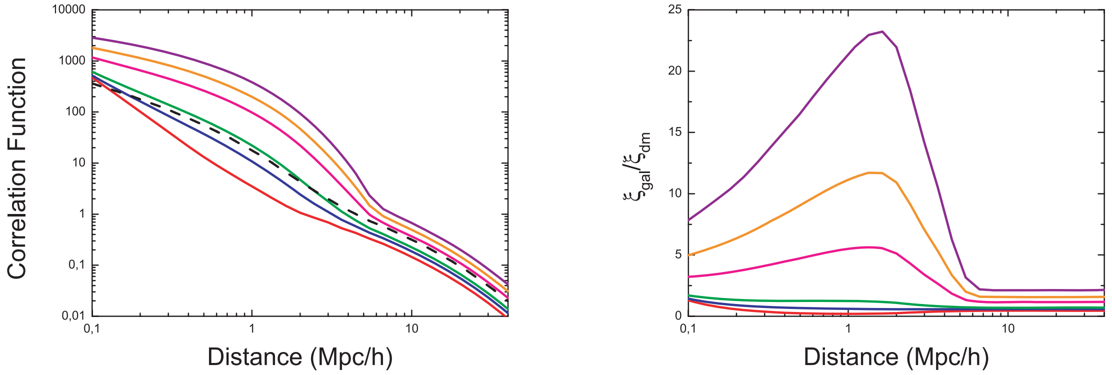

Figure 7. Effects of changing α in the halo model. As in Fig. 5 left-hand panel is $\xi_{\rm gal}$ and the right-hand panel $\xi_{\rm gal}/\xi_{\rm dm}$. Different colours are different values of α: 0.80 (red line), 1.00 (blue line), 1.10 (green line), 1.30 (pink line), 1.40 (orange line) and 1.50 (purple line). In the left-hand plot $\xi_{\rm dm}$ is the black-dashed plot.

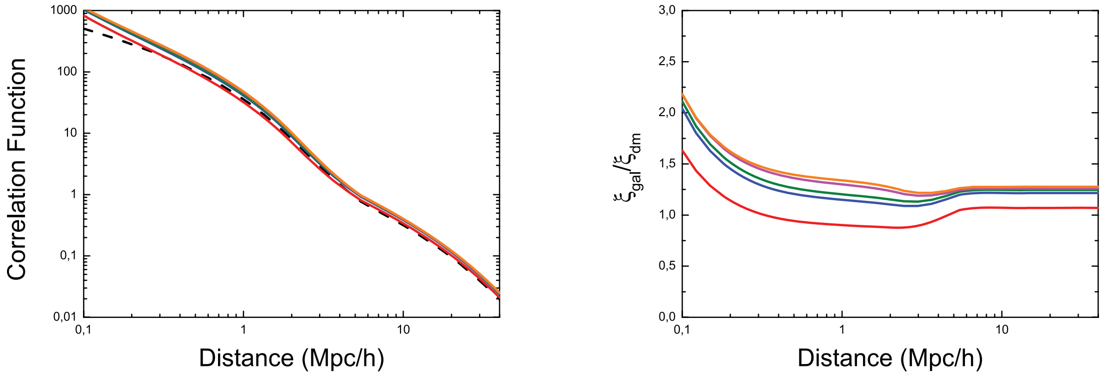

Figure 8. Effect of changing $M_{\rm min}$ for node-like haloes in the extended halo model. Left-hand panel is $\xi_{\rm gal}$; right-hand panel is $\xi_{\rm gal}/\xi_{\rm dm}$. The different colours are different values for $M_{\rm min}$: $10^{11}\,\mathrm{M}_{\odot}$ (red line), $5\times10^{11}\,\mathrm{M}_{\odot}$ (blue line), $10^{12}\,\mathrm{M}_{\odot}$ (green line), $5\times10^{12}\,\mathrm{M}_{\odot}$ (pink line) and $10^{13}\,\mathrm{M}_{\odot}$ (orange line). In the left-hand panel $\xi_{\rm dm}$ is also plotted (black-dashed line).

3.2 Extended halo model

In this section, we use again the HOD parametrized by equation (3) using as $g^{\rm cen}(m)$ and $g^{\rm sat}(m)$ for filament- and node-like haloes (equations A24 and A25) but with different parameters. In particular, we keep the HOD parameters fixed for filament haloes to the fiducial values: $M_{\rm min}=10^{11}\,\mathrm{M}_{\odot}$, $M_1=22\times10^{11}\,\mathrm{M}_{\odot}$, $\alpha=1$ and we allow to vary $M_{\rm min}$, M_1 and α only for red galaxies. We assume a concentration $c(m)$ given by equation (A15) and begin with a maximum segregation index $\mathcal{S}=1$ (this will later be relaxed). Since the filaments are the most common structure, changing their HOD would give a similar effect as changing the HOD of the whole population, which we studied in the previous section. Therefore, in this section we focus on how changing the HOD of the less abundant population (in this case the nodes) can affect the total correlation function.

3.2.1 *Central galaxies in nodes*

We set $M_1=22\times10^{11}\,\mathrm{M}_{\odot}$ and $\alpha=1.0$ for node-like satellites and vary $M_{\rm min}$ from $10^{11}\,\mathrm{M}_{\odot}$ to $10^{13}\,\mathrm{M}_{\odot}$.

In Fig. 8, we show the effect of changing $M_{\rm min}$ for node-like galaxies. The different colour lines are $\xi_{\rm gal}$ for different values of $M_{\rm min}$ for node-like galaxies (see the caption for details) and the black-dashed line is $\xi_{\rm dm}$. In the left-hand panel we show the total correlation function. We see that the total galaxy correlation function is not very sensitive to this parameter. We observe only a small to moderate change, much less than that observed when we changed $M_{\rm min}$ of the total population of haloes. This can be due to the fact that if we increase $M_{\rm min}$ less node-like haloes are populated, so there are less red galaxies and the total correlation function is dominated by the filament-like galaxies. In the right-hand plot, we show the corresponding effect on the $\xi_{\rm gal}/\xi_{\rm dm}$.

3.2.2 *Satellite galaxies in nodes*

Here, we keep $M_{\rm min}=10^{11}\,\mathrm{M}_{\odot}$ and $\alpha=1.0$ fixed and allow M_1 to vary. In Fig. 9, we show the effect of changing M_1 in the galaxy correlation function (left-hand panel) and on $\xi_{\rm gal}/\xi_{\rm dm}$ (right-hand panel). The different colour lines represent different values of M_1. As before, the black-dashed line is the dark matter correlation function. Recall that the effect of increasing M_1 is to reduce the number of satellite galaxies for low-mass haloes in the node-like regions. We see that $\xi_{\rm gal}$ is only sensitive to the change of M_1 in the range from $22\times10^{10}\,\mathrm{M}_{\odot}$ to $22\times$

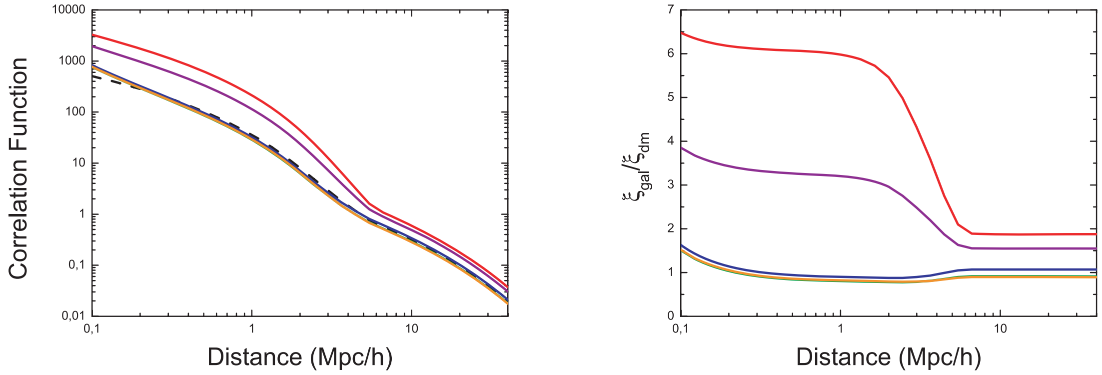

Figure 9. Effect of changing M_1 in node-like haloes. As in Fig. 8 in left-hand panel $\xi_{\rm gal}$ is shown and in the right-hand panel $\xi_{\rm gal}/\xi_{\rm dm}$ is plotted. The different colours correspond to different values for M_1: $22 \times 10^{10}\,{\rm M}_{\odot}$ (red line), $44 \times 10^{10}\,{\rm M}_{\odot}$ (purple line), $22 \times 10^{11}\,{\rm M}_{\odot}$ (blue line), $22 \times 10^{12}\,{\rm M}_{\odot}$ (green line) and $22 \times 10^{13}\,{\rm M}_{\odot}$ (orange line). In the left-hand panel, the black-dashed line represents $\xi_{\rm dm}$.

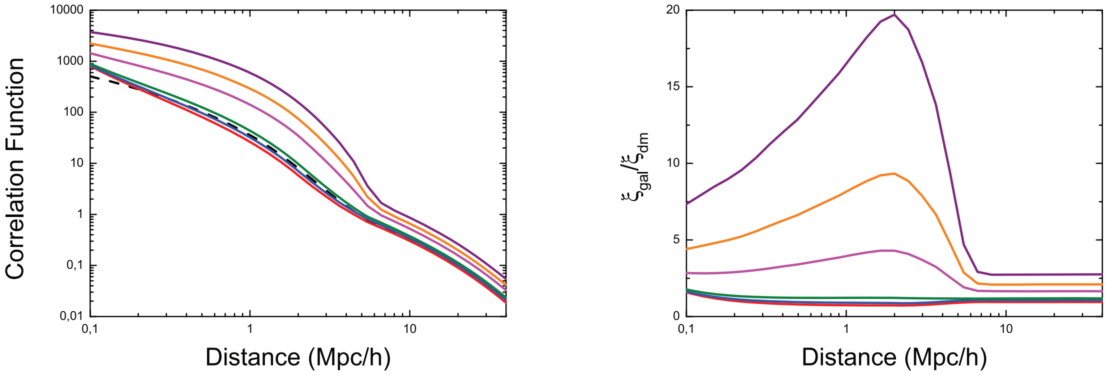

Figure 10. Effect of changing α on nodes-like haloes in our halo model. As in Fig. 8 in the left-hand panel $\xi_{\rm gal}$ is shown whereas in the right-hand panel $\xi_{\rm gal}/\xi_{\rm dm}$ is plotted. The different colours show different values of α: 0.80 (red line), 1.00 (blue line), 1.10 (green line), 1.30 (pink line), 1.40 (orange line) and 1.50 (purple line). In the left-hand plot, the black-dashed line represents $\xi_{\rm dm}$.

$10^{11}\,{\rm M}_{\odot}$. For values larger than $22 \times 10^{11}\,{\rm M}_{\odot}$ the correlation function saturates. In particular, we see that $\xi_{\rm gal}$ is especially sensitive to M_1 at small scales where the one-halo term dominates. This means that the number of satellites in the low-mass and node-like haloes plays an important role in the total correlation function.

In Fig. 10, we show the effect of changing α in node-like galaxies keeping fixed $M_{\rm min}$ and M_1 to $10^{11}\,{\rm M}_{\odot}$ and $22 \times 10^{11}\,{\rm M}_{\odot}$, respectively. We explore the regime from $\alpha = 0.80$ (red line) to $\alpha = 1.50$ (purple line) for the $\xi_{\rm gal}$ (left-hand panel) and for $\xi_{\rm gal}/\xi_{\rm dm}$ (right-hand panel). As before, $\xi_{\rm dm}$ is the black-dashed line. We observe the same effect observed in Fig. 7. As in the M_1 case, $\xi_{\rm gal}$ is especially sensitive to α at small distances where the one-halo term dominates. This may indicate that for these values of $\mathcal{X}$ and $\mathcal{Y}$ the satellite galaxies in the node-like haloes have an important role at small scales on the total galaxy correlation function in spite of not being the dominant population.

3.3 Concentration

In Fig. 11, we show the effect of changing the concentration parameter for both dark matter (left-hand panel) and galaxy correlation function (central panel). Also in the right-hand panel we show the effect on $\xi_{\rm gal}/\xi_{\rm dm}$. The different lines are for different concentration values of nodes

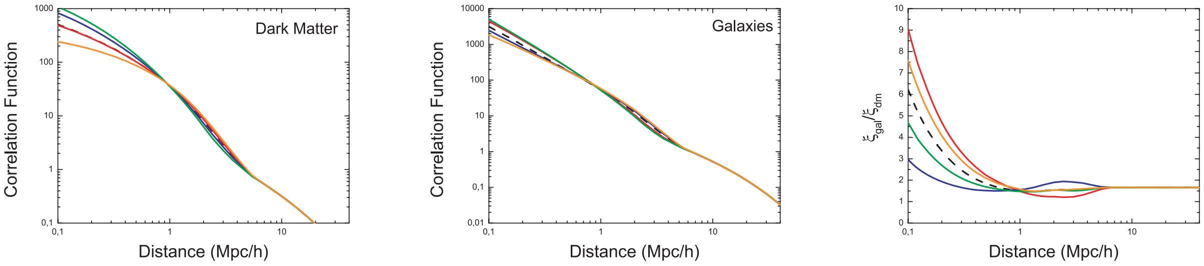

Figure 11. Correlation function for different values of the concentration parameters: $c_{\rm nod} = 10$ and $c_{\rm fil} = 2$ (red line), $c_{\rm nod} = 2$ and $c_{\rm fil} = 10$ (blue line), $c_{\rm nod} = 10$ and $c_{\rm fil} = 10$ (green line), and $c_{\rm nod} = 2$ and $c_{\rm fil} = 2$ (orange line). The black-dashed line represents the case for $c(m)$ according to equation (A15). In the left-hand panel, the dark matter correlation function is plotted, in the central panel the galaxy correlation function and in the right-hand plot $\xi_{\rm gal}/\xi_{\rm dm}$.

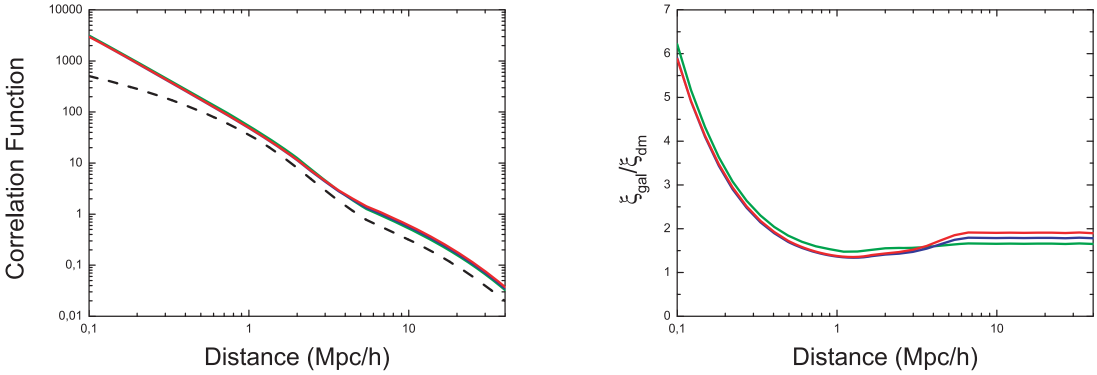

Figure 12. Effect of the segregation parameter. Left-hand panel is $\xi_{\rm gal}$ and right-hand panel $\xi_{\rm gal}/\xi_{\rm dm}$. For $\mathcal{S} = 1$ (green line) the red and blue galaxies are completely separated: red galaxies live in node-like haloes and blue galaxies in filament-like haloes. For $\mathcal{S} = 0$ (red line) the red and blue population are perfectly mixed in node- and filament-like haloes. For $\mathcal{S} = 0.5$ (blue line) an intermediate situation is presented: red galaxies are more abundant in node-like haloes and blue galaxies in filament-like haloes, but some degree of mixing between these two population is allowed. In the left-hand panel, the black-dashed line represents $\xi_{\rm dm}$.

and filaments: $c_{\rm nod} = 10$ and $c_{\rm fil} = 2$ (red line), $c_{\rm nod} = 2$ and $c_{\rm fil} = 10$ (blue line), $c_{\rm nod} = 10$ and $c_{\rm fil} = 10$ (green line) and $c_{\rm nod} = 2$ and $c_{\rm fil} = 2$ (orange line). For comparison we also have plotted $c(m)$ according to equation (A15) (black dashed line). Here, we have fixed the HOD to the one of Zehavi et al. (2005): $\log_{10}M_{\rm min} = 12.72$, $\log_{10}M_1 = 14.08$ (the masses are in $\mathrm{M}_{\odot}h^{-1}$), $\alpha = 1.37$, $f_0^{\rm cen} = 0.71$, $f_0^{\rm sat} = 0.88$, $\sigma_M^{\rm cen} = 0.30$ and $\sigma_M^{\rm sat} = 1.70$ (see Appendix A Section A6 for a detailed definition of these HOD parameters) and adopted the maximum value for the segregation index, $\mathcal{S} = 1$. Note that significant changes appear at small scales ($r < 1$ Mpc h^{-1}). Also these changes are only important for the dark matter correlation function. On the other hand, for galaxies the effect of changing c are almost negligible.

In general, the effect of changing the concentration can be summarized as a change in the slope of the small-scale correlation function (especially the dark matter one), with a more concentrated profile (green line) corresponding to a steeper correlation function.

3.4 Segregation

In Fig. 12, we explore the effect of having the two galaxy populations mixed or segregated. For two population of galaxies completely segregated the segregation index is $\mathcal{S} = 1$ (green line); when the two population are perfectly mixed this segregation index is $\mathcal{S} = 0$ (red line) (see equation 3 for definition of $\mathcal{S}$). Intermediate cases $1 > \mathcal{S} > 0$ ($\mathcal{S} = 0.5$ for blue line) describe a certain degree of mixing between the two populations, but always with red galaxies being the most abundant population in node-like haloes and blue galaxies more abundant one in filament-like haloes. The galaxy correlation function and $\xi_{\rm gal}/\xi_{\rm dm}$ are plotted in the left- and right-hand panels, respectively. In the left-hand panel, the black-dashed lines are the dark matter correlation function, $\xi_{\rm dm}$. We observe that at large scales, the more mixed the two population of galaxies are, the higher is the correlation function. This can be understood from the fact that the red galaxies are more biased (see Fig. 4). If the two population are completely mixed ($\mathcal{S} = 0$) all haloes will contain red galaxies and therefore, in total, there will be more red galaxies than if only the node-like haloes could host this red population (as it happens with $\mathcal{S} = 1$). Therefore, with increasing $\mathcal{S}$ the relative importance of the more biased galaxy population decreases, thus decreasing the total correlation function. However, this effect is very small in the total correlation function.

4 DISCUSSION AND CONCLUSIONS

In the classic halo model, the host halo mass is the only variable specifying halo and galaxy properties. This model has been remarkably successful at describing the first moment statistics of the clustering of galaxies. However, environment must play an important role in the process of galaxy formation, the most striking observational evidence being that clusters today have a much higher fraction of early-type galaxies than is found in the field. In addition, when looking at statistics beyond the two-point density correlation function, there are indications that the simplest halo model may be incomplete.

When looking at the two-point galaxy correlation function for current, low-redshift surveys, the data can be well modelled without such an additional dependence on local density (beyond that introduced by the host halo mass; see e.g. Skibba & Sheth 2009). However, the statistical power and redshift coverage of forthcoming surveys imply that an extended modelling may be needed.

In this work, we have presented a natural extension of the halo model that allows us to introduce an environmental dependence based on whether a halo lives in a high-mass density region, namely a node-like region, or in a low-mass region, namely a filament-like region. At the level of dark matter, the secondary variable (the environment-dependent variable beyond the shape of the mass function) is the concentration of the density profile of the haloes: haloes which live in node regions may have a different concentration than haloes which live in filaments regions. According to this idea we present the dark matter correlation function, $\xi_{\rm dm}$, of this new extended halo model (equations 6–8). In the classic halo model, the correlation function is usually split in two terms: the one- and two-halo terms. In our extended halo model, we have three terms: one one-halo term and two two-halo terms, depending on whether the two particles belong to haloes of the same or different type (see Fig. 1 for a graphical example). We have explored the contribution of these terms to the dark matter correlation function and have found that the contribution of particles that belong to haloes of the same and different kind is similar (see the right-hand panel of Fig. 2). We have also seen that according to the values of Table 1, the contribution of the node-like haloes starts to be dominant at small scales and is subdominant at large scales (see the left-hand panel of Fig. 2). We also have analysed how different values of the concentration affect the correlation function. At the level of dark matter, changing the concentration parameter between 2 and 10 only affects moderately the correlation function at small scales (see the left-hand panel of Fig. 11).

We have also extended our environment-dependent halo model to galaxies, computing the galaxy correlation function, ξ_{gal} (equations 12–14). This has been done by modulating the HOD recipe with the environment. In the galaxy correlation function, the HOD is our secondary variable rather than halo concentration. This means that according to our model, node-like haloes host galaxies in a different way than filament-like galaxies. In this paper we use a simple model for the HOD of only three variables: the minimum mass of a halo to host its first (central) galaxy, M_{min}; the mass of the halo to have on average its first satellite galaxy, M_1, and the way how satellite galaxies increase with the halo mass, α. Thus, we have two different HOD, one for red galaxies and another for blue galaxies. Here, we have chosen the colour as an example, although another property besides the colour, like star formation rate or metallicity, may be used instead. Of course, more sophisticated HOD could also be used, but starting with a simple prescription helps gaining physical insight. We analyse how changing the different parameters of the HOD affects the galaxy correlation function (see Figs 5–10). We see that, even when we change the parameters of the HOD of the minority population (in this case the node-like haloes), there is a considerable change in the ξ_{gal}, especially for the parameters M_1 and α, i.e. for the satellite population. Therefore, changing the HOD of only a small fraction of haloes considerably affects the total galaxy correlation function. We also have explored how the mixing or segregation between the two population of red and blue galaxies may affect ξ_{gal}. We expect that the size of this effect will depend on the HOD of each one of these populations: the more different are the HOD of red and blue galaxies, the more the effect of segregation.

It is reasonable to expect the dependence of galaxy properties and clustering on environment to be very complex in details. But, being a second-order effect, it is likely that a simplified description that captures the main trends will be all that is needed in practice. The model presented here is simplified in several ways: (1) the environment is only divided into nodes, filaments and voids, and it does not have a continuous distribution; (2) dark halo clustering only depends on mass and not on environment. The formalism introduced here allows one to implement (2) straightforwardly.

A natural extension of this model is to make a continuous dependence on the environment, instead of splitting it in nodes, filaments and voids. However, this extension presents several issues. First of all we would need a continuous dependence on the environment of the mass function and also of the HOD, which is not yet available. Secondly, nowadays N-body simulations start to provide information of haloes according a discrete number of environments: this discretization could be taken as a first approximation to the continuous distribution. While this avenue is worth pursuing, introducing a more complex model based on a continuous environment dependence is beyond the scope of this paper.

We envision that the extension of the halo model presented here will be useful for future analysis of large-scale structure of the Universe, especially the analyses that account for the physical properties of galaxies. Current photometric and spectroscopic large-scale surveys are beginning to gather not only the position of galaxies but also some of their physical properties, like colour, star formation, age or metallicity, and with future surveys the accuracy of these measurements will increase. We expect that the physical properties of galaxies depend on the environment, making these data sets the suitable ground to apply the extended halo model.

The halo model presented here uses analytic expressions for the mass function, halo density profile, and it also uses a given HOD for a certain magnitude-selected galaxies. We understand that these analytic functions may be inappropriate for comparison with data. In the present paper, we wanted to have an expression for the dark matter non-linear power spectrum, that is fast and easy to compute and a reasonable approximation to reality; we were however interested in the relative effect of including environment dependence and less concerned with absolute accuracy of the fit. In a forthcoming paper we are planning to apply this technique to fit observational data; in this case we would have to further refine our prescription to compute ξ_{dm} [see Smith et al. (2003) for a numerical approach to the halo model description]. There is still work in progress in this connection but we believe it is beyond the scope of the present paper.

The environmental halo model presented here is especially suited to model the marked (or weighed) correlation formalism (Harker et al. 2006; Skibba et al. 2006), consisting on weighting each galaxy according to a physical property and removing the dependence of clustering on local number-counts. Here, we have laid the foundations for modelling a survey's marked correlation, the treatment of which will be presented in a forthcoming work.

ACKNOWLEDGMENTS

HGM is supported by the CSIC JAE grant. LV and RJ are supported by MICCIN grant AYA2008-0353. LV is supported by FP7-IDEAS-Phys.LSS 240117, FP7-PEOPLE-2007-4-3-IRGn202182. We thank B. Reid and R. Sheth for discussions.

REFERENCES

Abbas U., Sheth R. K., 2005, MNRAS, 364, 1327
Abbas U., Sheth R. K., 2006, MNRAS, 372, 1749

Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., 2010, MNRAS, 408, 2163
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bolzonella M. et al., 2010, A&A, 524, A76
Cole S., Kaiser N., 1989, MNRAS, 237, 1127
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Cowie L. L., Barger A. J., 2008, ApJ, 686, 72
Cucciati O. et al., 2010, A&A, 524, A2
Davis M., Geller M. J., 1976, ApJ, 208, 13
Dressler A., 1980, ApJ, 236, 351
Galaz G., Herrera-Camus R., Garcia-Lambas D., Padilla N., 2011, ApJ, 728, 74
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Giocoli C., Bartelmann M., Sheth R. K., Cacciato M., 2010, MNRAS, 408, 300
Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1
Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039
Heavens A., Panter B., Jimenez R., Dunlop J., 2004, Nat, 428, 625
Hirata C. M., 2009, MNRAS, 399, 1074
Hogg D. W. et al., 2002, AJ, 124, 646
Jimenez R. et al., 2010, MNRAS, 404, 975
Krause E., Hirata C., 2011, MNRAS, 410, 2730
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Ma C.-P., Fry J. N., 2000, ApJ, 531, L87
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611
Mateus A., Jimenez R., Gaztañaga E., 2008, ApJ, 684, L61
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nat, 379, 613
Moore B., Lake G., Katz N., 1998, ApJ, 495, 139
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neto A. F. et al., 2007, MNRAS, 381, 1450
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Postman M., Geller M. J., 1984, ApJ, 281, 95
Press W. H., Schechter P., 1974, ApJ, 187, 425
Reid B. A., Spergel D. N., 2009, ApJ, 698, 143
Scherrer R. J., Bertschinger E., 1991, ApJ, 381, 349
Seljak U., 2000, MNRAS, 318, 203
Sheth R. K., 2005, MNRAS, 364, 796
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2004, MNRAS, 350, 1385
Sheth R. K., Jimenez R., Panter B., Heavens A. F., 2006, ApJ, 650, L25
Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080
Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, MNRAS, 369, 68
Smith R. E. et al., 2003, MNRAS, 341, 1311
Tasca L. A. M. et al., 2009, A&A, 503, 379
Yoo J., Tinker J. L., Weinberg D. H., Zheng Z., Katz N., Davé R., 2006, ApJ, 652, 26
Zehavi I. et al., 2005, ApJ, 630, 1
Zehavi I. et al., 2010, preprint (arXiv:1005.2413)
Zhao D. H., Jing Y. P., Mo H. J., Boumlrner G., 2009, ApJ, 707, 354
Zucca E. et al., 2009, A&A, 508, 1217

APPENDIX A: THE HALO MODEL

In this section, we review the basics of the standard halo model. This background material set up the stage for motivating and introducing our extension of the standard halo model and will define symbols and nomenclature used.

The halo model pioneered in Peacock & Smith (2000), Ma & Fry (2000) and Seljak (2000), and then thoroughly reviewed in Cooray & Sheth (2002), assumes that all the mass in the Universe is embedded into units, which are called dark matter haloes or simply haloes. These haloes are small compared to the typical distance between them (non-linear evolution makes the evolved Universe dominated by voids). For this reason, the clustering properties of the mass density field, $\delta(\boldsymbol{x})$, on small scales are determined by the spatial distribution inside the dark matter haloes, and the way they are organized in the space is not important. On the other hand, the statistics of the large-scale distribution are not affected by the matter distribution inside haloes but only by their spatial distribution. In this work, we assume all the time halo exclusion, treating haloes as hard spheres (sharp cut-off at the virial radius). In the case that two particles that belong to different haloes are separated by less distance than the sum of the virial radii of their host haloes, we avoid counting them in the computation of the correlation function.

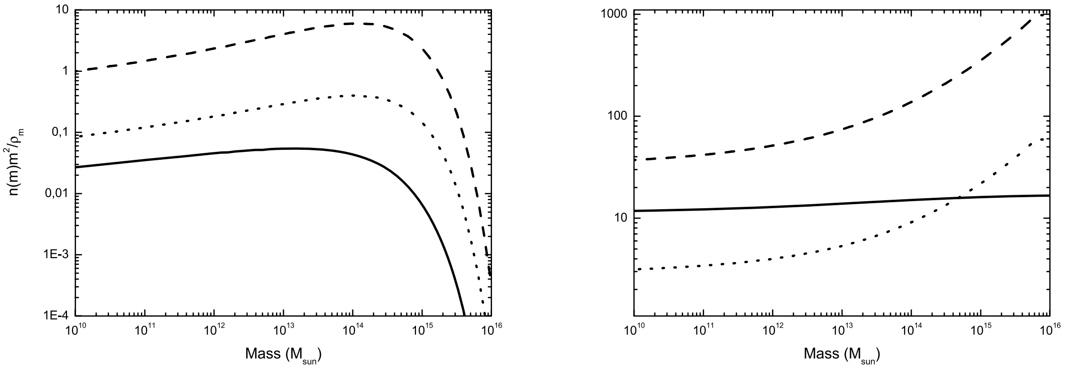

Figure A1. Left-hand panel: global mass function (solid line), mass function in node-environment (dashed line) and mass function in filament environments (dotted line). Right-hand panel: ratio of the mass functions of the left-hand panel. $n_{\rm nod}(m)/n(m)$ (dashed line), $n_{\rm fil}(m)/n(m)$ (dotted line) and $n_{\rm nod}(m)/n_{\rm fil}(m)$ (solid line).

A1 Mass function

The number density of collapsed haloes of mass m per unit of mass at a given redshift z, $n(m, z)$, can be computed using the Press–Schechter formalism (Press & Schechter 1974) identifying the present collapsed haloes with the peaks of an initially Gaussian field,

$$n(m, z) = \frac{2\bar{\rho}_{\rm m}}{m\sigma(m)} f(\nu)\nu \left| \frac{{\rm d}\sigma(m)}{{\rm d}m} \right| , \tag{A1}$$

where $\bar{\rho}_{\rm m}$ is the mean density of matter in the Universe,[7] $\sigma(m)$ is the *rms* of the power spectrum linearly extrapolated to $z = 0$ filtered with a top-hat sphere of mass m and $\nu \equiv \delta^2_{\rm sc}(z)/\sigma^2(m)$. Here, $\delta_{\rm sc}(z)$ is the linearly extrapolated critical density required for spherical collapse at z, and is given by $\delta_{\rm sc}(z) = 1.686/D(z)$, where $D(z)$ is the growth factor. Following this formalism, structures with a linearly evolved density fluctuation higher than this threshold value will collapse. The ST approach (Sheth & Tormen 1999), based on ellipsoidal collapse, predicts a mass function of the form

$$f(\nu)\nu = A(p)\left(1 + (q\nu)^{-p}\right)\left(\frac{q\nu}{2\pi}\right)^{1/2} \exp\left(-q\nu/2\right), \tag{A2}$$

with $p = 0.3$ and $q = 0.75$ and the normalization factor $A(p) = (1 + 2^{-p}\Gamma(1/2 - p)/\Gamma(1/2))^{-1}$. Note that by construction the matter density is given by

$$\bar{\rho}_{\rm m} = \int_0^{\infty} {\rm d}m\, n(m, z) m\, . \tag{A3}$$

In order to introduce an environmental dependence in the mass function, one could think of rescaling equation (A1) to account the differences of densities. However Mo & White (1996) noted that dark halo abundance in dense and underdense regions do not differ by just a factor like this. A more complex model was proposed by Abbas & Sheth (2005),

$$n(m; M_i, V_i) = [1 + b(m)\delta_i]\, n(m) \tag{A4}$$

where $n(m;\, M_i,\, V_i)$ is the mass function of a region of volume V_i which contains a mass M_i, $b(m)$ is the bias of a halo of mass m (see equation A7 for details) and δ_i is defined as $M_i/V_i \equiv \bar{\rho}_{\rm m}(1 + \delta_i)$. In Fig. A1 (left-hand panel) we have plotted the global mass function (equation A1) (solid line), a high-density environment mass function (dashed line) and a medium-density environment mass function (dotted line) using in all cases the ST mass function (equation A2). For these two last mass functions we have used, respectively, the relative density parameters of Table 1 corresponding to nodes and filaments. On the right-hand panel we show the ratios of these three different mass functions: $n_{\rm nod}(m)/n(m)$ (dashed line), $n_{\rm fil}(m)/n(m)$ (dotted line) and $n_{\rm nod}(m)/n_{\rm fil}(m)$ (solid line). We see that the ratio of the number of massive to low mass haloes is larger in dense regions (nodes) than in less dense regions (filaments) since the slope of the solid curve of the right-hand panel of Fig. A1 is positive.

A2 Bias

For large separations the bias is the relation between the correlation function of two dark matter haloes of masses m' and m'' separated by a distance r at a given redshift z, $\xi_{hh}(r, z, m', m'')$ and the underlying dark matter linear power spectrum, $\xi_{\rm lin}(r, z)$. While computing the exact form of $\xi_{hh}(r, z, m', m'')$ is a somewhat delicate matter, excellent results are obtained by using

$$\xi_{hh}(r, z, m', m'') \simeq b(m', z) b(m'', z) \xi_{\rm lin}(r, z), \tag{A5}$$

---

[7] According to a Λcold dark matter flat universe ($\Omega_{\rm m} = 0.27$, $\Omega_\Lambda = 0.73$ and $h = 0.7$) the matter density is given by $\bar{\rho}_{\rm m} = \frac{3\Omega^0_{\rm m} H_0^2}{8\pi {\rm G}}$ where $H_0 \equiv 100\, h$. Using the fiducial cosmology assumed here this value is $\bar{\rho}_{\rm m} = 7.438 \times 10^{10}\,{\rm M}_\odot\, h^{-1}\,({\rm Mpc}/h)^{-3}$.

where $b(m, z)$ denotes the large-scale linear bias factor for haloes of mass m at redshift z. While using $\xi_{\rm lin}$ as a proxy for ξ_{hh} is strictly incorrect because there may be mildly non-linear contributions and because at small separations haloes are spatially exclusive, on these scales the signal is dominated by the one-halo term. While the fit to N-body simulations can be further improved by using, e.g., higher-order perturbation-theory prediction instead of $\xi_{\rm lin}$ (e.g. Smith et al. 2008), we will not pursue this here as it is beyond the scope of the present paper.

To be consistent we should use a bias prescription derived from the extended Press–Schechter formalism using the peak background split. To the lowest order the bias is

$$b(m, z) = 1 - \left. \frac{\partial \ln[n(m, z')]}{\partial \delta_{\rm sc}(z')} \right|_{\delta_{\rm sc}(z)} . \tag{A6}$$

The bias of an object m at time z is given, at lowest order of δ, by (Cole & Kaiser 1989; Mo & White 1996)

$$b(m, z) = 1 + \frac{1}{D(z)} \left[q \frac{\delta_{\rm sc}(z)}{\sigma^2(m)} - \frac{1}{\delta_{\rm sc}(z)} \right] , \tag{A7}$$

where $D(z)$ is the linear growth factor, $\delta_{\rm sc}(z)$ is the critical threshold for collapse and is given by $\delta_{\rm sc} \simeq 1.686/D(z)$, $\sigma(m)$ is the *rms* of the power spectrum linearly extrapolated at $z = 0$ filtered with a top-hat sphere of mass m and q is the parameter introduced in the Sheth & Tormen mass function (see equation A2). This formula has been confirmed by N-body simulations giving an excellent agreement. Although the complete formula should have the extra term, $2p/(\delta_{\rm sc}(z)D(z))(1 + (\frac{q\delta_{\rm sc}^2(z)}{\sigma(m)^2})^p)^{-1}$, we have checked that the effect of this term is about 1 per cent and can be safely neglected for our application.

A3 Density profile

According to our definition, a halo is a set of particles that forms a gravitationally bound and thermodynamically stable system. Therefore, we consider that this system satisfies the virial theorem, i.e. a halo is virialized by definition. From the spherical collapse model (Gunn & Gott 1972), a halo is considered to be formed and virialized when its density reaches a certain threshold value: $\rho_{\rm vir}$. Here we will adopt $\bar{\rho}_{\rm vir} = 18\pi^2 \bar{\rho}_{\rm m}$, which is the typical value that can be found using the spherical collapse model. The corresponding virial radius is then

$$r_{\rm vir}(m) = \frac{1}{2\pi} \left(\frac{m}{3\bar{\rho}_{\rm m}} \right)^{1/3} . \tag{A8}$$

In the halo model, this will be the size of a halo of mass m.

On the other hand, cosmological simulations have shown that the density profile inside isolated haloes follows a universal profile given by (Navarro et al. 1996, 1997)

$$\rho(r|r_{\rm s}, \rho_{\rm s}) = \frac{\rho_{\rm s}}{(r/r_{\rm s})(1 + r/r_{\rm s})^2} \tag{A9}$$

where $r_{\rm s}$ is the scale radius of the halo and $\rho_{\rm s}/4$ is the halo density at scale radius. However, it is more natural to work with the concentration c, and the mass of the halo m, instead of $\rho_{\rm s}$ and $r_{\rm s}$ when we describe the halo profile. These variables are defined as

$$c \equiv r_{\rm vir}/r_{\rm s}, \tag{A10}$$

and the mass of the halo has to satisfy

$$m = \int_0^{r_{\rm vir}(m)} {\rm d}r\, 4\pi r^2 \rho(r|r_{\rm s}, \rho_{\rm s}) . \tag{A11}$$

From this last equation, we can write

$$\rho_{\rm s} = \frac{m}{4\pi r_{\rm s}^3} \left[\ln(1 + c) - \frac{c}{1 + c} \right]^{-1} . \tag{A12}$$

We prefer to use normalized profile of a halo u as a function of distance from its centre for given halo virial mass and concentration parameter, defined by

$$u(r|m, c) \equiv \frac{\rho(r|m, c)}{m}, \tag{A13}$$

which satisfies the condition

$$1 = \int_0^{r_{\rm vir}(m)} {\rm d}^3\boldsymbol{x}\, u(x|m, c) . \tag{A14}$$

A4 Concentration

In principle, the halo density profile depends on two independent parameters: either $\rho_{\rm s}$ and $r_{\rm s}$, or c and m. However N-body simulations indicate that there is a relation between the concentration and the mass. Seljak (2000) propose

$$c(m, z) = \frac{9}{1 + z} \left[\frac{m}{m_\star} \right]^{-0.2} \tag{A15}$$

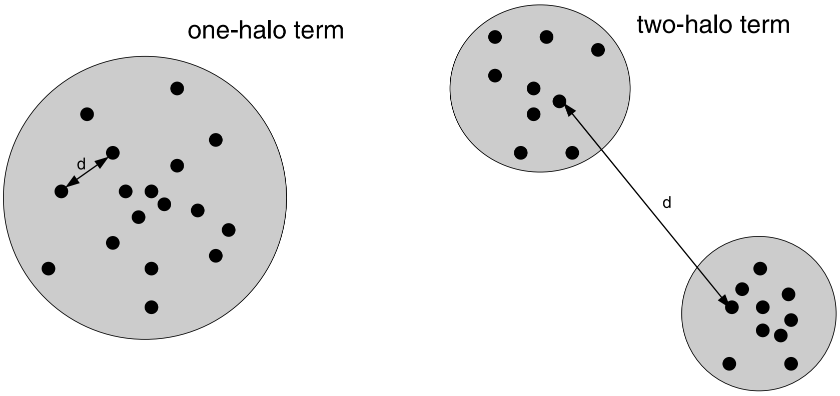

Figure A2. In the halo model, the correlation function is split in two terms. The one-halo term (left) describes the clustering of particles inside the same halo. On the other hand, the two-halo term (right) involves only particles of different haloes.

where $m_\star$ is the mass of a typical collapse halo at $z = 0$. There are many other parametrizations of the relation between the concentration and the mass (e.g. Neto et al. 2007; Zhao et al. 2009). However, we do not expect that the results of this work depend significantly on this choice. Unless otherwise stated, for the present application we thus adopt equation (A15) for the relation between concentration and mass.

This relation goes in the direction one may have expected: less massive haloes on average form earlier, when the Universe is more concentrated and therefore have a higher concentration than more massive ones. One should bear in mind, however, that there is a large dispersion around this mean relation, which may correspond to a 'hidden parameter' such as local environmental effects, tidal effects, mergers, etc.

A5 Two-point correlation function

In the halo model, the two-point correlation function of dark matter particles contained in haloes is defined as

$$\xi_{\rm dm}(\boldsymbol{r}) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle, \tag{A16}$$

where $\delta(\boldsymbol{x}) \equiv \rho(\boldsymbol{x})/\bar{\rho}_{\rm m} - 1$. This function can be split into two terms, depending on whether the two particles at distance $\boldsymbol{r}$ belong or not to the same halo:

$$\xi_{\rm dm}(\boldsymbol{r}) = \xi^{1h}_{\rm dm}(\boldsymbol{r}) + \xi^{2h}_{\rm dm}(\boldsymbol{r}). \tag{A17}$$

$\xi^{1h}_{\rm dm}(\boldsymbol{r})$ is called the one-halo term and accounts for particles in the same halo; $\xi^{2h}_{\rm dm}(\boldsymbol{r})$ is called the two-halo term and accounts for particles belonging to different haloes. Therefore, the properties of the mass density on small scales are described by $\xi^{1h}_{\rm dm}(r)$, whereas on large scales are given by $\xi^{2h}_{\rm dm}(r)$. A graphical description is shown in Fig. A2.

According to the halo model formalism, these terms are

$$\xi^{1h}_{\rm dm}(\boldsymbol{r}, z) = \int {\rm d}m \, \frac{m^2\, n(m,z)}{\bar{\rho}^2_{\rm m}(z)} \int_V {\rm d}^3\boldsymbol{x}\, u(\boldsymbol{x}|m)\, u(|\boldsymbol{x}+\boldsymbol{r}||m) \tag{A18}$$

$$\xi^{2h}_{\rm dm}(\boldsymbol{r}, z) = \int {\rm d}m' \, \frac{m'\, n(m',z)}{\bar{\rho}_{\rm m}(z)} \int {\rm d}m'' \, \frac{m''\, n(m'',z)}{\bar{\rho}_{\rm m}(z)} \int_V {\rm d}^3\boldsymbol{x}'\, u(x'|m') \int_V {\rm d}^3\boldsymbol{x}''\, u(x''|m'')\xi_{hh}(|\boldsymbol{x}'-\boldsymbol{x}''+\boldsymbol{r}|, z, m', m''). \tag{A19}$$

The integration $\int_V {\rm d}^3\boldsymbol{x}$ means over all haloes' volume, and $\int {\rm d}m$ runs over all mass range. For computational reasons, we have to adopt some limits in the mass integral. Setting $m_{\rm max} = 10^{15}\,{\rm M}_\odot$ is enough as far as the value of the integral does not change if we increase this limit. If we set the minimum mass limit to $10^9\,{\rm M}_\odot$ it is also enough for the one-halo term, because the less massive haloes contribute at very small distances (less than 0.1 Mpc h^{-1}). However the contribution of low-mass haloes to the two-halo term is not negligible. In order to compute these integrals we use the method described in Yoo et al. (2006) which consists in breaking the two-halo integral in two parts and approximate the low-mass haloes as points without inner structure,

$$\int_0^\infty {\rm d}m \frac{mn(m,z)}{\bar{\rho}_{\rm m}} b(m,z) \int_V {\rm d}^3\boldsymbol{x}\, u(x|m) \simeq \int_{m_{th}}^\infty {\rm d}m \frac{mn(m,z)}{\bar{\rho}_{\rm m}} b(m,z) \int_V {\rm d}^3\boldsymbol{x}\, u(x|m) + \left[1 - \int_{m_{th}}^\infty \frac{mn(m,z)}{\bar{\rho}_{\rm m}} b(m,z)\right], \tag{A20}$$

where we have used the approximation of equation (A5) to explicit the mass dependence of ξ_{hh}. In particular, we have set m_{th} to $10^9\,{\rm M}_\odot$ and we have checked that reducing this value does not affect the result of the integral.

In Fig. A3, we can see the contribution of these two terms. $\xi^{1h}_{\rm dm}(\boldsymbol{r})$ dominates at scales typically smaller than the virial radius, whereas $\xi^{2h}_{\rm dm}(\boldsymbol{r})$ does it for larger scales following the shape of $\xi_{\rm lin}$ slightly shifted by the effect of the bias (see equation A5).

A6 Halo occupation distribution

The halo model provides a framework to also model galaxy clustering: the complicated galaxy formation physics would determine how many galaxies form in a halo and their sampling of the dark matter halo profile. Thus, the shape of the one-halo term would be modified accordingly.

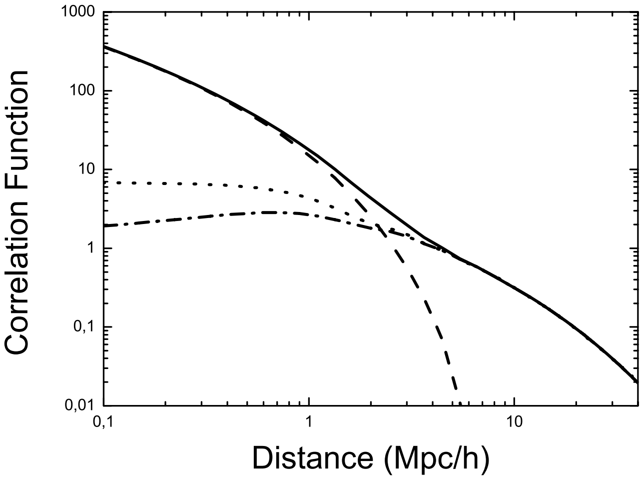

Figure A3. Total dark matter correlation function $\xi_{\rm dm}$ (solid line), one-halo term $\xi_{\rm dm}^{1h}$ (dashed line), two-halo term $\xi_{\rm dm}^{2h}$ (dot–dashed line) and linear power spectrum $\xi_{\rm lin}$ (dotted line).

Typically, the HOD assume a centre-satellite distribution of galaxies inside each halo. This means that a galaxy is placed at the centre of the halo and may be surrounded by satellite galaxies distributed according to some statistics. The average number of galaxies that lie in a halo of mass m is (e.g. Sheth 2005)

$$g_{(1)}(m) \equiv \sum_N Np(N|m) \tag{A21}$$

where $p(N|m)$ is the probability density of N galaxies formed in a halo of mass m.

We can also define the nth factorial moment of the distribution $p(N|m)$ of galaxies in haloes of mass m,

$$g_{(n)}(m) = \sum_N N(N-1)\cdots(N-n+1)p(n|m). \tag{A22}$$

Here, we assume that $p(n|m)$ follows a Poisson distribution, although other choices are also possible. Under this assumption, we can state that for satellite galaxies,

$$g_{(n)}(m) = (g_{(1)}(m))^n\,. \tag{A23}$$

In particular, we will only be interested in the first moment $g_{(1)}(m)$ (the mean number of galaxies in a halo of mass m) and in the second moment $g_{(2)}(m)$ that we will treat as the square power of the former. Therefore, hereafter we write $g_{(1)}(m)$ as $g(m)$ to simplify the notation.

The typical way to parametrize the mean number of centre galaxies is to think of the mean number of the central galaxies as a step function,

$$g^{\rm cen}(m) = \begin{cases} 1 & \text{if} \quad m > M_{\rm min} \\ 0 & \text{if} \quad m < M_{\rm min}\,. \end{cases} \tag{A24}$$

The satellite galaxy distribution can be parametrized as a power law, with the same mass cut as central galaxies,

$$g^{\rm sat}(m) = \left(\frac{m}{M_1}\right)^{\alpha} g^{\rm cen}(m), \tag{A25}$$

where $M_{\rm min}$ sets the minimum mass for a halo to have galaxies; M_1 is the mass of a halo that on average hosts one satellite galaxy and α is the power-law slope of the satellite mean occupation function. All these parameters can be tuned to fit observations and in general depend on the type of galaxy under consideration.

In the literature, there are several ways to populate haloes with galaxies depending on their type: Luminous Red Galaxies (LRGs) (Reid & Spergel 2009), field galaxies (Kravtsov et al. 2004) and many others (e.g. Zehavi et al. (2005)). On one hand, Kravtsov et al. (2004) using high-resolution dissipationless simulations find that the parameters of the HOD according to equations (A24) and A25 are $M_{\rm min} \simeq 10^{11}\,{\rm M}_{\odot}\,h^{-1}$, $M_1 \simeq 22 M_{\rm min}$ and $\alpha \simeq 1$. On the other hand, Zehavi et al. (2005) based on Sloan Digital Sky Survey (SDSS) observation state that for a volume-limited sample with magnitude in r band $(M_r) < -21$ the values of these parameters are $M_{\rm min} = 10^{12.72}\,{\rm M}_{\odot}\,h^{-1}$, $M_1 = 10^{14.08}\,{\rm M}_{\odot}\,h^{-1}$ and $\alpha = 1.37$ (see Fig. A5, left-hand panel), although these values depend strongly on the magnitude selection criteria. Therefore, it is clear that the parameters of the HOD depend strongly on the way galaxies are selected: whether one deals with a volume-limited sample of whether a magnitude selection criteria have been applied. It is interesting to note that some authors (e.g. Zehavi et al. (2005)) introduce a colour dependence through the HOD: haloes of the same mass host different kind of galaxies, red and blue. Since blue galaxies tend to live in low-density regions (filamentary regions) and red galaxies tend to live in high-density regions (node regions), this can be though as different HOD parametrization for different kind of environments. In particular, Zehavi et al. (2005) introduce the fraction of blue galaxies in a halo of mass m, $f_b(m)$, as a function of the halo mass and also as a function of galaxy type: central or satellite. Since red galaxies are more common in high-mass haloes, the authors adopt a function for f_b which is decreasing with halo mass. For central galaxies a lognormal

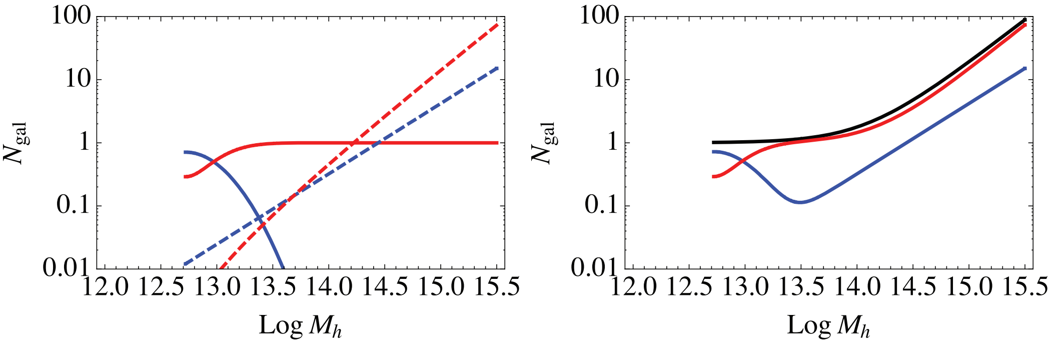

Figure A4. HOD with colour dependence according to Zehavi et al. (2005). Left plot: central and satellite galaxy contribution (solid and dashed lines, respectively) for red and blue galaxies (red and blue lines, respectively). Right-hand panel: total galaxy number for all galaxies (black line), red galaxies (red line) and blue galaxies (blue line).

function is adopted:

$$f_b^{\rm cen}(m) = f_0^{\rm cen} \exp\left[-\frac{(\log_{10} m - \log_{10} M_{\rm min})^2}{2\sigma_M^{\rm cen\,2}}\right] \tag{A26}$$

whereas for satellite a log-exponential function is used:

$$f_b^{\rm sat}(m) = f_0^{\rm sat} \exp\left(-\frac{\log_{10} m - \log_{10} M_{\rm min}}{\sigma_M^{\rm sat}}\right) , \tag{A27}$$

where m is the mass of the halo, $M_{\rm min}$ is the minimum mass of a halo to host one galaxy, f_0 is the value of the fraction for a halo with mass $M_{\rm min}$ and σ_M is a parameter that characterizes how fast the blue fraction drops. Also these parameters depend strongly on the magnitude-selection criteria. For $M_r < -21$ they are: $\log_{10}M_{\rm min} = 12.72$, $f_0^{\rm cen} = 0.71$, $f_0^{\rm sat} = 0.88$, $\sigma_M^{\rm cen} = 0.30$ and $\sigma_M^{\rm sat} = 1.70$. The average number of blue and red galaxies becomes then

$$\begin{aligned} g^i_{\rm blue}(m) &= f_b(m)^i g^i(m) \\ g^i_{\rm red}(m) &= \left[1 - f^i_b(m)\right] g^i(m), \end{aligned} \tag{A28}$$

where the average number of galaxies of type i, $g^i(m)$ is given by equations (A24) and (A25) according the values of the parameters used in Zehavi et al. (2005). In Fig. A4, we illustrate the corresponding HOD for equation (A28) for red (red lines) and blue (blue lines) galaxies, and for the total number of galaxies (black line). In the left-hand panel, we show the separate central and satellite contributions: the solid lines are central galaxies and dashed lines satellite galaxies. In the right-hand panel, we show the combined effect of the contributions. As stated by the authors, for haloes just above $M_{\rm min}$, blue central galaxies are more common. However, above $2M_{\rm min}$, central galaxies are predominantly red. Note that for blue galaxies there is a minimum in the total number of galaxies that occurs when a halo is too massive to have a blue central galaxy but not massive enough to host a significant population of blue satellite galaxies. As a general trend, we can say that according to this kind of HOD model: (i) the lowest mass haloes have blue central galaxies, (ii) higher mass haloes have red central galaxies but with a significant blue fraction in their satellite population and, finally, (iii) the highest mass haloes host red central and satellite galaxies. Zehavi et al. (2005) considered that the two populations are perfectly mixed and proceeded to compare the HOD predictions for the projected correlation function with SDSS data. This kind of difference between the HOD of blue and red galaxies motivates us to further explore the environmental dependence of the HODs of haloes of different density regions and in particular to consider that different environment can have different ratios of red and blue galaxies.

Equations (A18) and (A19) can then be modified to describe the clustering of galaxies:

$$\xi^{1h}_{\rm gal}(\boldsymbol{r}, z) = \int {\rm d}m\, \frac{n(m,z)}{\bar{n}^2_{\rm gal}(z)} \left[2\, g^{\rm cen}(m) g^{\rm sat}(m) u(r|m) + \left(g^{\rm sat}(m)\right)^2 \int {\rm d}^3\boldsymbol{x}\, u(x|m)\, u(|\boldsymbol{x}+\boldsymbol{r}|m)\right] \tag{A29}$$

$$\begin{aligned} \xi^{2h}_{\rm gal}(\boldsymbol{r}, z) = \int {\rm d}m'{\rm d}m''\frac{n(m',z)n(m'',z)}{\bar{n}^2_{\rm gal}(z)} &\left[g^{\rm cen}(m')g^{\rm cen}(m'')\xi_{hh}(r,z,m',m'') + 2\, g^{\rm cen}(m')g^{\rm sat}(m'')\int {\rm d}^3\boldsymbol{x}'\, u(x'|m'')\right. \\ &\left. \times\, \xi_{hh}(|\boldsymbol{x}'+\boldsymbol{r}|, z, m', m'') + g^{\rm sat}(m')g^{\rm sat}(m'')\int {\rm d}^3\boldsymbol{x}'{\rm d}^3\boldsymbol{x}''\, u(x'|m')u(x''|m'')\xi_{hh}(|\boldsymbol{x}'-\boldsymbol{x}''+\boldsymbol{r}|, z, m', m'')\right], \end{aligned} \tag{A30}$$

where the mean number of galaxies per halo is given by

$$\bar{n}_{\rm gal}(z) = \int {\rm d}m\, n(m,z)\left(g^{\rm cen}(m) + g^{\rm sat}(m)\right) . \tag{A31}$$

In Fig. A5 (left-hand panel) the Zehavi et al. (2005) HOD is plotted ($M < -21$ sample, $\log_{10}M_{\rm min} = 12.72$, $\log_{10}M_1 = 14.08$ and $\alpha = 1.37$). The dashed line corresponds to the contribution of central galaxies, the dotted line to satellite galaxies and the solid line to the total. In the central panel, we show the correlation function for dark matter and galaxies: $\xi_{\rm gal}$ (solid line), $\xi_{\rm dm}$ (dashed line) and $\xi_{\rm lin}$ (dotted line), according the same HOD. In the right-hand panel, the ratio between $\xi_{\rm gal}$ and $\xi_{\rm dm}$ is plotted. We see that at scales smaller than $\sim$1 Mpc h^{-1},

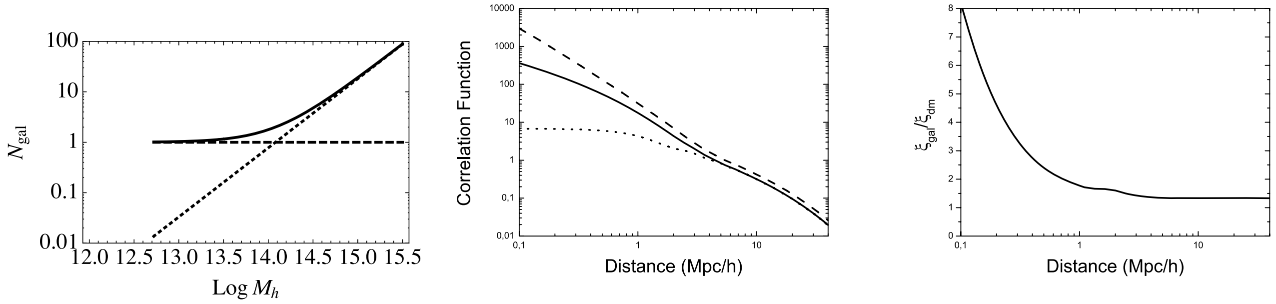

Figure A5. Left-hand panel: HOD for the $-21 > M_r$ sample. All galaxies (solid line), central galaxy (dashed line) and satellite galaxies (dotted line). Centre panel: dark matter correlation function $\xi_{\rm dm}$ (solid line), galaxy correlation function $\xi_{\rm gal}$ (dashed line) according the HOD of the left-hand panel and linear correlation function $\xi_{\rm lin}$ (dotted line). Right-hand panel: ratio between $\xi_{\rm gal}$ and $\xi_{\rm dm}$.

when typically the one-halo term dominates, the ratio $\xi_{\rm gal}/\xi_{\rm dm}$ is highly scale dependent, whereas at larger scales, the two-halo terms dominate the ratio $\xi_{\rm gal}/\xi_{\rm dm}$ is scale independent. In particular, it appears that at scales $\sim$0.1 Mpc h^{-1} there is a minimum for $\xi_{\rm gal}/\xi_{\rm dm}$ with a value of 8. At large scales ($>$10 Mpc h^{-1}) $\xi_{\rm gal}/\xi_{\rm dm}$ takes a value around 1.3. This is expected from the halo model formalism, where at small scales the one-halo term dominates whereas at large scales the two-halo term does it. From equations (A29)–(A30) it can be seen that at large scales the two-halo term can be expressed as a function of $\xi_{\rm lin}$ whereas this does not happen for the one-halo term. This yields us to the point that at large scales the ratio between galaxies and dark matter is just a constant, whereas at small scales the relation is much more complicated.

APPENDIX B

Here, we set the basics of the derivation of equations (6)–(8) of our extended halo model. This derivation is very similar to the one of classic halo model presented for the first time by Scherrer & Bertschinger (1991).

We define the overdensity $\delta(\boldsymbol{x})$ as $\delta(\boldsymbol{x}) \equiv \rho(\boldsymbol{x})/\bar{\rho} - 1$. Then the dark matter correlation function is

$$\xi_{\rm dm}(\boldsymbol{r}) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle = \frac{1}{\bar{\rho}^2}\langle \rho(\boldsymbol{x})\rho(\boldsymbol{x}+\boldsymbol{r})\rangle - 1, \tag{B1}$$

where $\langle\cdots\rangle$ denotes the ensemble average. On the other hand, the density field at $\boldsymbol{x}$ is the sum of densities of node- and filament-like regions: $\rho(\boldsymbol{x}) = \rho^{\rm nod}(\boldsymbol{x}) + \rho^{\rm fil}(\boldsymbol{x})$. We also can write the contribution of the nodes/filaments as the sum of every halo of this class:

$$\rho^j(\boldsymbol{x}) = \sum_i m_i u(\boldsymbol{x} - \boldsymbol{x}_i | m_i, c_j), \tag{B2}$$

where the summation takes place over all haloes of type j (nodes or filaments), m_i and $\boldsymbol{x}_i$ are the mass and the position (of the centre) of the ith halo, and u is the normalized profile defined in equation (A13). From equations (B1) and (B2) it is clear that we can split $\xi_{\rm dm}$ in three terms, $\xi_{\rm dm}^{1h}$, $\xi_{\rm dm}^{2h1\eta}$ and $\xi_{\rm dm}^{2h2\eta}$, depending on whether the particles belong to the same halo and depending on the particles belong to the same kind of halo (in the case of the two-halo term). The derivation of each terms to equations (6)–(8) is similar and are based on the introduction of the Dirac delta,

$$\sum_i \to \int {\rm d}m\, {\rm d}^3\boldsymbol{x} \sum_i \delta_D^3(\boldsymbol{x} - \boldsymbol{x}_i)\delta_D(m - m_i), \tag{B3}$$

and on the definition of the mass functions in the node- and filament-like regions as

$$n_j(m) \equiv \frac{1}{V_j}\int_{V_j} \sum_i {\rm d}^3\boldsymbol{x}\, \delta_D^3(\boldsymbol{x} - \boldsymbol{x}_i)\delta_D(m - m_i) \tag{B4}$$

where the index j denotes the type of environment, either node or filament, and V_j the volume of this environment.

Using these formulae and following the formalism presented by Scherrer & Bertschinger (1991), equations (6)–(8) can be obtained after some algebra.

This paper has been typeset from a TEX/LATEX file prepared by the author.

CHAPTER 3

Reducing sample variance: halo biasing, non-linearity and stochasticity

In this Chapter we present the paper Gil-Marín et al. (2010), where the possibility of improving the measurements of the growth factor computing the power spectrum of dark matter tracers in redshift space is explored. In particular, the method focuses on reducing the sample variance error by comparing different populations of dark matter tracers with different bias properties. In this formalism we account for the possibility that these tracers may present a non-linear and stochastic behaviour related to the underlying dark matter field.

This paper was done in collaboration with Christian Wagner, Raúl Jiménez and Licia Verde from the Institut de Ciències del Cosmos at the Universitat the Barcelona, Spain, with Alan Heavens from the University of Edinburgh, United Kingdom, and was published in Monthly Notices of the Royal Astronomical Society (MNRAS) in May 2010.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **407,** 772–790 (2010)

doi:10.1111/j.1365-2966.2010.16958.x

Reducing sample variance: halo biasing, non-linearity and stochasticity

Héctor Gil-Marín,[1,2,3★] Christian Wagner,[2] Licia Verde,[2,4,5] Raul Jimenez[2,4,5]
and Alan F. Heavens[3]

[1]*Institute of Space Sciences (IEEC-CSIC), Facultat de Ciències, Campus UAB, Bellaterra 08193, Spain*
[2]*Institut de Ciències del Cosmos(ICC), Universitat de Barcelona-IEEC, Martíi Franquès 1, 08028 Barcelona, Spain*
[3]*Scottish Universities Physics Alliance (SUPA), Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ*
[4]*ICREA (Institució Catalana de Recerca i Estudis Avançats)*
[5]*Institute for the Physics and Mathematics of the Universe (IPMU), the University of Tokyo, Kashiwa, Chiba 277-8568, Japan*

Accepted 2010 May 4. Received 2010 May 4; in original form 2010 March 17

ABSTRACT

Comparing clustering of differently biased tracers of the dark matter distribution offers the opportunity to reduce the sample or cosmic variance error in the measurement of certain cosmological parameters. We develop a formalism that includes bias non-linearities and stochasticity. Our formalism is general enough that it can be used to optimize survey design and tracers selection and optimally split (or combine) tracers to minimize the error on the cosmologically interesting quantities. Our approach generalizes the one presented by McDonald & Seljak of circumventing sample variance in the measurement of $f \equiv \mathrm{d}\ln D/\mathrm{d}\ln a$. We analyse how the bias, the noise, the non-linearity and stochasticity affect the measurements of Df and explore in which signal-to-noise regime it is significantly advantageous to split a galaxy sample in two differently biased tracers. We use N-body simulations to find realistic values for the parameters describing the bias properties of dark matter haloes of different masses and their number density. We find that, even if dark matter haloes could be used as tracers and selected in an idealized way, for realistic haloes, the sample variance limit can be reduced only by up to a factor $\sigma_{2\mathrm{tr}}/\sigma_{1\mathrm{tr}} \simeq 0.6$. This would still correspond to the gain from a three times larger survey volume if the two tracers were not to be split. Before any practical application one should bear in mind that these findings apply to dark matter haloes as tracers, while realistic surveys would select galaxies: the galaxy–host halo relation is likely to introduce extra stochasticity, which may reduce the gain further.

Key words: cosmological parameters – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

One of the active topics of current research is the formation and growth of large-scale structure in the Universe. Knowledge of the physical origin of the growth of structure will allow us to know about the origin of dark matter and also provides a useful way to discriminate between different theories for the origin and evolution of dark energy. In particular, comparing and combining measurements of the Universe expansion history (as given by e.g. Baryon Acoustic Oscillations, cosmic chronometers and Supernovae) with measurements of the linear growth of structure can provide a tool to test whether dark energy is an extra component with negative pressure or a manifestation of the breakdown of general relativity on large scales. To this end, usually, the goal is to measure the $f(z)$ parameter defined as $f(z) \equiv \mathrm{d}\ln D(z)/\mathrm{d}\ln a(z)$, where $D(z)$ is the linear growth factor and $a(z)$ the scalefactor.

The two main approaches to measure the growth of structure are weak gravitational lensing (e.g. Hoekstra et al. 2002; Bacon et al. 2005) and galaxy clustering, which is the technique we consider here. Galaxy clustering is a relatively simple, high signal-to-noise ratio (S/N) measurement: the angular position of galaxies can be measured using photometry and the radial position using spectroscopy. With this information the three-dimensional power spectrum of galaxies can be computed as a function of redshift. However, at the two-point level, the galaxy field can only trace the dark matter field up to a bias factor b, which may depend on scale and redshift. Thus, only once the bias

---

★E-mail: hectorgil@icc.ub.edu

is known, the galaxy power spectrum $P_{\rm gg}$ can be related to that of the dark matter $P_{\rm mm}$ and thus yield the growth of structure. The main drawback associated to this technique is that the value of this bias cannot readily be predicted from theoretical models of galaxy formation.

There are several observational techniques to measure the bias parameter (Fry 1994; Feldman et al. 2001; Verde et al. 2002; Hawkins et al. 2003; Seljak et al. 2005; Guzzo et al. 2008). In this work we use the approach that takes advantage of the redshift-space distortions. Peculiar velocities of dark matter tracers are set by the gravitational field: using the measured redshift as a distance indicator distorts clustering, enhancing it along the line of sight and the redshift-space distortion parameter is $\beta \equiv f/b$ (Davis & Peebles 1983; Kaiser 1987; Hamilton 1998). Using measurements in different directions (different Fourier modes), one can compute β. In combination with the galaxy power spectrum measurement, this approach yields the divergence power spectrum $P_{\theta\theta} = f^2 P_{\rm mm}$, which can be directly compared with theory predictions and encloses the desired dependence on the growth of structure.

There are two sources of errors in the measurement of the galaxy power spectrum: the shot noise and the sample variance (or cosmic variance). On one hand, the shot noise is due to the fact that we use a discrete set of objects to characterize the matter field. If this noise is Poisson, it is scale-independent and equal to the reciprocal of the number density of objects (Peebles 1980). On the other hand, the sample variance effect is due to the fact that the matter field has its origin in a random realization of the underlying cosmology. In a finite survey volume there are only a finite number of modes present, especially on large scales we only have a few modes to perform the averaging. Thus, the total error on the power spectrum P at a given scale k, is $\sigma_P/P = (2/N)^{1/2}\,(1 + \sigma_{\rm n}/P)$. Here, N is the number of modes measured (at the scale given by k) and $\sigma_{\rm n}$ is the shot noise contribution. We see that just reducing the shot noise (increasing the number density of objects) does not help, as there is a natural limitation on our capacity to measure P [and consequently $f(z)$], due to sample variance.

In order to reduce this limitation, a multi-tracer technique has been advocated recently (Seljak 2009). It is based on the usage of two differently biased tracers of the dark matter field. With this method the sample variance limit can be reduced. The effectiveness of this method depends on a number of factors: the ratio of these different biases; the S/N regime and on the non-linearity of the biases. With the exception of gravitational lensing, one cannot see the dark matter directly nor the dark matter haloes, so in most practical applications, tracers need to be used such as galaxies, quasars, clusters or 21-cm emission.

The goal of this paper is to study the possibility of measuring the parameter $x(z) \equiv f^2(z)D^2(z)$ using the single- and the multi-tracer formalism and see whether the reduction of the sample variance is significant. Here, we present a new formalism of how to estimate the error on x using the multi-tracer formalism, taking into account that the bias may be scale-dependent, non-linear and stochastic. This formalism may be useful for galaxy surveys, because it has been observed that the galaxy biasing is significantly non-linear and stochastic. N-body simulations and theoretical models allow us to estimate which are the bias characteristics for dark matter tracers and therefore which precision can be reached with this model.

In Section 2, we begin by introducing the formalism of our method and analyse how the different parameters affect the reduction of the sample variance effect. In Section 3 we use both analytical approximations and N-body simulations to obtain physically motivated parameters for our model and compute realistic expected errors using the single- and multi-tracer formalism. In Section 4 we conclude with a summary and a discussion of the results.

2 METHOD

We start with the basic assumptions that the tracer (galaxies or haloes) number density is given by a Poisson sampling[1] of an underlying continuous field $n_{\rm g}(\boldsymbol{x})$, with overdensity defined by $\delta_{\rm g}(\boldsymbol{x}) \equiv n_{\rm g}(\boldsymbol{x})/\bar{n}_{\rm g} - 1$, and that the galaxy overdensity field is related to the mass overdensity field δ by a conditional probability distribution $P(\delta_{\rm g}|\delta)$, including a stochastic element which we will describe later. In addition, in realistic surveys using the redshift as distance indicator, peculiar velocities distort clustering in a manner dependent on the angle with respect to the line of sight. In particular, clustering is enhanced – at least on large scales, in the linear regime – along the line of sight. It is the angular dependence of the effect that yields the signal to extract a measurement of the growth of structure. In this paper, our main goal is to estimate the error of the growth rate of perturbations, $f(z)$, generalizing the method proposed by McDonald & Seljak (2009) to measure $f(z)$, reducing the sample variance limit and using redshift-space distortions. The idea is to split the sample of objects into two subsamples with different biasing properties. In their work, they used a linear bias model and all stochasticity was due to shot noise. Here we generalize their work to the non-linear bias case and take into account the possibility of having off-diagonal noise terms, which can be introduced, for example, if there are objects in common in the two samples. In this regime, we compare the multi-tracer with the single-tracer approach, and we analyse how the noise and non-linearities affect the extent to which the sample variance limit can be improved. We also use results from simulations to set plausible values for these parameters to see how great the gains may be in practice.

2.1 Modelling of bias

The relation between clustering properties of the dark matter and those of the tracer (haloes or galaxies) goes under the name of 'bias'. The simplest, non-trivial bias model is linear bias, $\delta_{\rm g}(\boldsymbol{x}) = b_1\delta(\boldsymbol{x})$, with b_1 constant and independent of position and scale. This corresponds to a deterministic linear biasing, which has little physical motivation and is problematic if $b_1 > 1$ and the field is not very linear, since it allows an

---

[1] This model may be inadequate in details when dealing with real haloes (Smith, Scoccimarro & Sheth 2007; Seljak, Hamaus & Desjacques 2009). We will return to this point in Section 3.

unphysical $\delta_g < -1$ in voids. A more complex relation is almost certainly needed to properly describe galaxy clustering. Non-linear biasing with a bias which is no longer a constant but a function of $\delta(\boldsymbol{x})$ is a common way to improve the model.

Here we adopt the formalism proposed by Dekel & Lahav (1999). We assume that both $\delta(\boldsymbol{x})$ and $\delta_g(\boldsymbol{x})$ are random fields with one-point probability distributions functions, $P[\delta(\boldsymbol{x})]$ and $P[\delta_g(\boldsymbol{x})]$, with zero mean $\langle\delta(\boldsymbol{x})\rangle = \langle\delta_g(\boldsymbol{x})\rangle = 0$ and variances $\langle\delta^2(\boldsymbol{x})\rangle$ and $\langle\delta_g^2(\boldsymbol{x})\rangle$, respectively.

We first define the *mean* biasing function, $b[\delta(\boldsymbol{x})]$, as the conditional mean between the galaxy and the matter field,

$$b[\delta(\boldsymbol{x})]\delta(\boldsymbol{x}) \equiv \langle\delta_g(\boldsymbol{x})|\delta(\boldsymbol{x})\rangle = \int \mathrm{d}\delta_g(\boldsymbol{x}) P(\delta_g(\boldsymbol{x})|\delta(\boldsymbol{x}))\delta_g(\boldsymbol{x}). \tag{1}$$

This is the natural generalization of the deterministic linear biasing relation, where the function $b[\delta(\boldsymbol{x})]$ characterizes the non-linear bias behaviour. Note that $P(\delta_g|\delta)$ can have a width (i.e. a scatter) around the mean relation, $b(\delta)$, which is however not captured by the function $b(\delta)$. We characterize the function $b(\delta)$ by the first- and second-order moments, which are given by $\hat{b}(r)$ and $\tilde{b}(r)$ at zero lag (i.e. $r = 0$),

$$\hat{b}(r) \equiv \frac{\langle\delta(\boldsymbol{x}+\boldsymbol{r})\delta(\boldsymbol{x})b[\delta(\boldsymbol{x})]\rangle}{\langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle}, \tag{2}$$

$$\tilde{b}^2(r) \equiv \frac{\langle\delta(\boldsymbol{x}+\boldsymbol{r})\delta(\boldsymbol{x})b[\delta(\boldsymbol{x})]b[\delta(\boldsymbol{x}+\boldsymbol{r})]\rangle}{\langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle}, \tag{3}$$

where $\langle\,\rangle$ represents the averaging over the volume of the survey or over different realizations. These two parameters take into account the non-linearity of the system as long as one is concerned with the two-point correlation function (or the power spectrum) and not higher order correlations. It is useful to define their ratio as

$$R(r) \equiv \frac{\hat{b}(r)}{\tilde{b}(r)}, \tag{4}$$

which is a useful parameter to measure the non-linearity of the bias. In the linear case, $\hat{b}(r) = \tilde{b}(r)$ and $R(r) = 1$, whereas for non-linear cases, $R(r) < 1$. Note that $\hat{b}(r)$ is the bias as it would appear in the tracer–dark matter cross-correlation, while $\tilde{b}^2(r)$ would appear in the tracer auto-correlation.

We next turn to stochasticity, by which we mean any physical or statistical process that produces a non-deterministic relation between the dark matter and the galaxy (or halo) field. This may arise from the discrete nature of galaxies, in which case it is called shot noise; if it is a Poisson process, its expression is inversely proportional to the mean density of objects, $1/\bar{n}_g$, but the formalism used here allows for other stochastic processes which are encoded in the width of $P(\delta_g|\delta)$.

In order to study the stochasticity of the bias, we define the random bias field $\epsilon(\boldsymbol{x})$ as the difference between the galaxy field and the dark matter field once biased by the mean bias relation $b[\delta(\boldsymbol{x})]$,

$$\epsilon(\boldsymbol{x}) \equiv \delta_g(\boldsymbol{x}) - b[\delta(\boldsymbol{x})]\delta(\boldsymbol{x}). \tag{5}$$

If $P(\delta_g|\delta)$ is a univariate Gaussian then $b(\delta)$ and $\sigma_b^2(\delta)$, the variance of $P(\delta_g|\delta)$ at a given δ, completely specifies $P(\delta_g|\delta)$. The variance of the ϵ field is given by the average of $\sigma_b^2(\delta)$ over δ.

In general, once δ, δ_g and ϵ are defined, the corresponding correlation functions are

$$\xi_{\rm mm}(r) \equiv \langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+r)\rangle, \tag{6}$$

$$\xi_{\rm gm}(r) \equiv \langle\delta_g(\boldsymbol{x})\delta(\boldsymbol{x}+r)\rangle, \tag{7}$$

$$\xi_{\rm gg}(r) \equiv \langle\delta_g(\boldsymbol{x})\delta_g(\boldsymbol{x}+r)\rangle, \tag{8}$$

$$\xi_{\epsilon\epsilon}(r) \equiv \langle\epsilon(\boldsymbol{x})\epsilon(\boldsymbol{x}+r)\rangle, \tag{9}$$

$$\xi_{\epsilon\rm m}(r) \equiv \langle\epsilon(\boldsymbol{x})\delta(\boldsymbol{x}+r)\rangle. \tag{10}$$

In what follows, we are only interested in the two-point correlation function or the power spectrum, thus we do not need to specify further moments of $P(\delta_g|\delta)$ or higher order correlations.

In order to give an illustrative example to the bias formalism, let us consider a simple non-linear bias model given by

$$b(\delta) = b_0 + b_1\delta + b_2\delta^2. \tag{11}$$

Let us also assume that this bias model is non-stochastic. Therefore according to equation (5), we have that the galaxy overdensity must be

$$\delta_g = b_0\delta + b_1\delta^2 + b_2\delta^3. \tag{12}$$

In order to deal with the simplest scenario, we assume that δ is a Gaussian random field. This means that the n-point correlation function, $\langle\delta^n\rangle$, can be expressed as a function of the two-point correlation function, $\langle\delta^2\rangle$.

We have stated above that δ_g field has to satisfy $\langle\delta_g\rangle = 0$. Provided that $\langle\delta\rangle = \langle\delta^3\rangle = 0$, we have that b_1 must be null:

$$b(\delta) = b_0 + b_2\delta^2. \tag{13}$$

Table 1. Different sets of parameters of the non-linear model given by equation (13) and the corresponding values of $\hat{b}$, $\tilde{b}$ and R.

Set	b_0	$b_2\langle\delta^2\rangle$	$\hat{b}$	$\tilde{b}$	R
Set 1	1	0	1	1	1
Set 2	2	0	2	2	1
Set 3	0.7	0	0.7	0.7	1
Set 4	1	0.5	2.50	2.78	0.90
Set 5	0.3	0.7	2.40	2.94	0.81
Set 6	2	−0.3	1.1	1.32	0.83

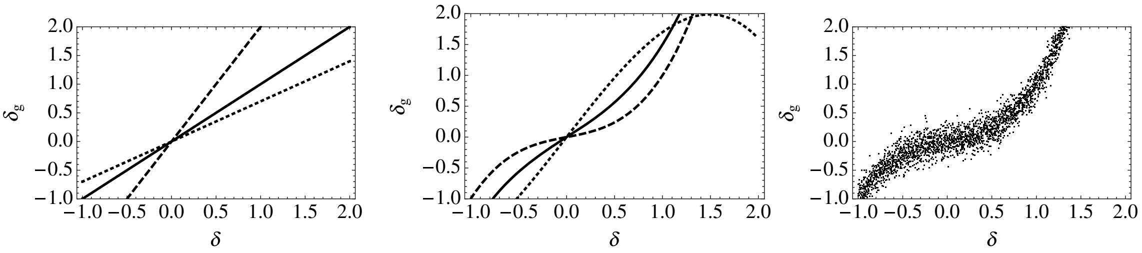

Figure 1. Different biasing models. Left-hand panel: Set 1 (solid line), Set 2 (dashed line) and Set 3 (dotted line). Middle panel: Set 4 (solid line), Set 5 (dashed line) and Set 6 (dotted line). Right-hand panel: Set 5 with stochasticity component. Sets are detailed in Table 1.

The biasing parameters given by equations (2) and (3) are

$$\hat{b} = b_0 + 3b_2\langle\delta^2\rangle, \tag{14}$$

$$\tilde{b}^2 = b_0^2 + 6b_0b_2\langle\delta^2\rangle + 15b_2^2\langle\delta^2\rangle^2, \tag{15}$$

where we have used $\langle\delta^4\rangle = 3\langle\delta^2\rangle^2$ and $\langle\delta^6\rangle = 15\langle\delta^2\rangle^3$, if δ is Gaussian.

As an illustrative example, we consider this simple biasing model with different set of parameters, b_0 and b_2, listed in Table 1. Set 1, Set 2 and Set 3 are linear biasing models ($b_2 = 0$), and therefore $\hat{b} = \tilde{b}$ and $R = 1$. These models are plotted in Fig. 1 (left-hand panel) with solid, dashed and dotted lines, respectively. On the other hand, Set 4, Set 5 and Set 6 are non-linear biasing models ($b_2 \neq 0$) plotted in Fig. 1 (central panel) with solid, dashed and dotted lines, respectively. In these cases, it is clear that $\hat{b} \neq \tilde{b}$ and therefore $R < 1$. Note also that R is a good indicator of how non-linear the model is; the more deviate R is from 1 the more non-linear $b(\delta)$ is.

Finally, the stochasticity can easily be included in this formalism just adding a random field, namely ϵ, in equation (12). The result of doing this with Set 5 is shown in Fig. 1 (right-hand panel).

2.2 Redshift-space distortions of biased tracers

So far, we have defined the bias model in configuration space. However, to deal with redshift-space distortions, it is more convenient to work with the galaxy density function δ_g in k-space. The basic relation for redshift distortions is, assuming the distant observer approximation (Kaiser 1987),

$$\delta_\mathrm{g}^\mathrm{s}(\boldsymbol{k}) = \delta_\mathrm{g}(\boldsymbol{k}) + f\mu^2(\hat{k})\delta(\boldsymbol{k}), \tag{16}$$

where $\delta_\mathrm{g}(\boldsymbol{k})$ is the Fourier transform of the overdensity of galaxies, $\delta_\mathrm{g}^\mathrm{s}(\boldsymbol{k})$ its redshift-space counterpart and $\delta(\boldsymbol{k})$ the real-space transform of the overdensity of dark matter. $\mu(\hat{k})$ is the cosine of the angle between the line of sight and $\boldsymbol{k}$. Throughout this section, we consider a single snapshot at a fixed redshift, usually we take $z = 0$. Therefore we write, for instance, f instead of $f(z)$ and so on. If we consider that we have two dark matter tracers, then

$$\delta_{\mathrm{g}i}^\mathrm{s}(\boldsymbol{k}) = \delta_{\mathrm{g}i}(\boldsymbol{k}) + f\mu^2(\hat{k})\delta(\boldsymbol{k}); \qquad i = 1, 2, \tag{17}$$

with a corresponding covariance matrix

$$C_\mathrm{2tr}(\boldsymbol{k}) = \begin{pmatrix} \left\langle\delta_\mathrm{g1}^\mathrm{s}(\boldsymbol{k})\delta_\mathrm{g1}^{\mathrm{s}\,*}(\boldsymbol{k})\right\rangle & \left\langle\delta_\mathrm{g1}^\mathrm{s}(\boldsymbol{k})\delta_\mathrm{g2}^{\mathrm{s}\,*}(\boldsymbol{k})\right\rangle \\ \left\langle\delta_\mathrm{g1}^\mathrm{s}(\boldsymbol{k})\delta_\mathrm{g2}^{\mathrm{s}\,*}(\boldsymbol{k})\right\rangle & \left\langle\delta_\mathrm{g2}^\mathrm{s}(\boldsymbol{k})\delta_\mathrm{g2}^{\mathrm{s}\,*}(\boldsymbol{k})\right\rangle \end{pmatrix}. \tag{18}$$

We define various power spectra as

$$\langle\delta_{\mathrm{g}i}(\boldsymbol{k})\delta_{\mathrm{g}j}^*(\boldsymbol{k})\rangle \equiv P_{\mathrm{g}i\mathrm{g}j}(k) \tag{19}$$

$$\langle\delta(\boldsymbol{k})\delta_{\mathrm{g}i}^*(\boldsymbol{k})\rangle \equiv P_{\mathrm{mg}i}(k) \tag{20}$$

$$\langle\delta(\boldsymbol{k})\delta^*(\boldsymbol{k})\rangle \equiv P_\mathrm{mm}(k). \tag{21}$$

Note that these quantities are related to equations (6–8) through their Fourier transforms,

$$P(k) = \int \mathrm{d}^3\boldsymbol{r}\, \xi(r)\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot r}. \tag{22}$$

The terms of the covariance matrix of equation (18) can be expressed as,[2]

$$\left\langle \delta^{\mathrm{s}}_{\mathrm{g}i}(\boldsymbol{k}) \delta^{\mathrm{s}}_{\mathrm{g}j}{}^{*}(\boldsymbol{k}) \right\rangle = P_{\mathrm{mm}}(k) \left[\frac{P_{\mathrm{g}i\mathrm{g}j}(k)}{P_{\mathrm{mm}}(k)} + f\mu^2 \frac{P_{\mathrm{mg}i}(k) + P_{\mathrm{mg}j}(k)}{P_{\mathrm{mm}}(k)} + f^2\mu^4 \right]. \tag{23}$$

In this paper, we will assume that the cross terms between the random field and the matter field are subdominant and can be assumed to be zero:

$$\langle \epsilon(\boldsymbol{x}) B[\delta(\boldsymbol{x} + r)] \rangle = 0, \tag{24}$$

where $B[\delta]$ is any function of δ. If we relate these quantities to the bias parameters defined in equations (2) and (3), we obtain

$$P_{\mathrm{mm}}(k) = \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}, \tag{25}$$

$$P_{\mathrm{mg}i}(k) = \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\mathrm{mg}i}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r} = \int \mathrm{d}^3\boldsymbol{r}\, \hat{b}_i(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}, \tag{26}$$

$$P_{\mathrm{g}i\mathrm{g}i}(k) = \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\mathrm{g}i\mathrm{g}i}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r} = \int \mathrm{d}^3\boldsymbol{r}\, \tilde{b}_i^2(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r} + \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\epsilon i \epsilon i}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}, \tag{27}$$

$$P_{\mathrm{g1g2}}(k) = \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\mathrm{g1g2}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r} = \int \mathrm{d}^3\boldsymbol{r}\, \langle \delta(\boldsymbol{x}) b_1[\delta(\boldsymbol{x})] \delta(\boldsymbol{r} + \boldsymbol{x}) b_2[\delta(\boldsymbol{r} + \boldsymbol{x})] \rangle\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r} + \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\epsilon_1\epsilon_2}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}, \tag{28}$$

$$= \int \mathrm{d}^3\boldsymbol{r}\, R_{12}(r) \tilde{b}_1(r) \tilde{b}_2(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r} + \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\epsilon_1\epsilon_2}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}, \tag{29}$$

where the parameter $R_{12}(r)$ is a new non-linearity parameter between tracers of type 1 and 2, which is defined as

$$R_{12}(r) \equiv \frac{\langle b_1[\delta(\boldsymbol{x})] \delta(\boldsymbol{x}) b_2[\delta(\boldsymbol{x} + \boldsymbol{r})] \delta(\boldsymbol{x} + \boldsymbol{r}) \rangle}{\langle b_1[\delta(\boldsymbol{x})] \delta(\boldsymbol{x}) b_1[\delta(\boldsymbol{x} + r)] \delta(\boldsymbol{x} + r) \rangle^{1/2} \langle b_2[\delta(\boldsymbol{x})] \delta(\boldsymbol{x}) b_2[\delta(\boldsymbol{x} + r)] \delta(\boldsymbol{x} + r) \rangle^{1/2}}. \tag{30}$$

For convenience we define new bias parameters,

$$\hat{b}_i(k) \equiv \frac{\int \mathrm{d}^3\boldsymbol{r}\, \hat{b}_i(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}}{\int \mathrm{d}^3\boldsymbol{r}\, \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}} = \frac{P_{\mathrm{mg}i}(k)}{P_{\mathrm{mm}}(k)} \tag{31}$$

$$\tilde{b}_i^2(k) \equiv \frac{\int \mathrm{d}^3\boldsymbol{r}\, \tilde{b}_i^2(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}}{\int \mathrm{d}^3\boldsymbol{r}\, \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}} = \frac{P_{\mathrm{g}i\mathrm{g}i}(k) - P_{\epsilon i \epsilon i}(k)}{P_{\mathrm{mm}}(k)} \tag{32}$$

$$R_i(k) \equiv \frac{\hat{b}_i(k)}{\tilde{b}_i(k)} = \frac{P_{\mathrm{mg}i}(k)}{\{P_{\mathrm{mm}}(k)[P_{\mathrm{g}i\mathrm{g}i}(k) - P_{\epsilon i \epsilon i}(k)]\}^{1/2}} \tag{33}$$

$$R_{12}(k) \equiv \frac{\int \mathrm{d}^3\boldsymbol{r}\, R_{12}(r) \tilde{b}_1(r) \tilde{b}_2(r) \xi_{\mathrm{mm}}\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}}{\left[\int \mathrm{d}^3\boldsymbol{r}\, \tilde{b}_1^2(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}\right]^{1/2} \left[\int \mathrm{d}^3\boldsymbol{r}\, \tilde{b}_2^2(r) \xi_{\mathrm{mm}}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}\right]^{1/2}} = \frac{P_{\mathrm{g1g2}}(k) - P_{\epsilon 1 \epsilon 2}(k)}{[P_{\mathrm{g1g1}}(k) - P_{\epsilon 1 \epsilon 1}(k)]^{1/2} [P_{\mathrm{g2g2}}(k) - P_{\epsilon 2 \epsilon 2}(k)]^{1/2}} \tag{34}$$

and the ϵ field power spectrum,

$$P_{\epsilon i \epsilon j}(k) \equiv \int \mathrm{d}^3\boldsymbol{r}\, \xi_{\epsilon i \epsilon j}(r)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot r}. \tag{35}$$

Thus the covariance matrix reads

$$C_{2\mathrm{tr}}(k, \mu) = \begin{pmatrix} C_{11}(k, \mu) & C_{12}(k, \mu) \\ C_{12}(k, \mu) & C_{22}(k, \mu) \end{pmatrix} \tag{36}$$

with

$$C_{11}(k, \mu) = P_{\mathrm{mm}}(k) \left[\tilde{b}_1^2(k) + 2f\mu^2 \tilde{b}_1(k) R_1(k) + f^2\mu^4\right] + P_{\epsilon_1\epsilon_1}(k), \tag{37}$$

$$C_{12}(k, \mu) = P_{\mathrm{mm}}(k) \left\{R_{12}(k) \tilde{b}_1(k) \tilde{b}_2(k) + f\mu^2 \left[\tilde{b}_1(k) R_1(k) + \tilde{b}_2(k) R_2(k)\right] + \mu^4 f^2\right\} + P_{\epsilon_1\epsilon_2}(k), \tag{38}$$

$$C_{22}(k, \mu) = P_{\mathrm{mm}}(k) \left[\tilde{b}_2^2(k) + 2f\mu^2 \tilde{b}_2(k) R_2(k) + f^2\mu^4\right] + P_{\epsilon_2\epsilon_2}(k). \tag{39}$$

Because $\tilde{b}_i(k)$ are parameters which cannot be obtained readily from observations, it is preferable to work with the redshift-space distortion parameter $\tilde{\beta}_i(k) \equiv f/\tilde{b}_i(k)$. Also $P_{\mathrm{mm}}(k, z)$ at a given z is not directly measurable. However, in the linear regime we can write it as

---

[2] In this expression the factor of 2 difference from that appearing in McDonald & Seljak (2009) is due to a different definition of $\delta_{\boldsymbol{k}}$; we consider that $\delta_{\boldsymbol{k}}$ is complex.

$P_{\rm mm}(k,z)=D^2(z)\,P^0_{\rm mm}(k)$, where $P^0_{\rm mm}(k)$ is the fiducial power spectrum and $D(z)$ is the linear growth factor. Thus, defining $x(z)\equiv D^2(z)f^2(z)$ the covariance matrix elements are

$$C_{11}(k,\mu)=x\,P^0_{\rm mm}(k)\left[\tilde\beta_1^{-2}(k)+2\mu^2\tilde\beta_1^{-1}(k)R_1(k)+\mu^4\right]+P_{\epsilon_1\epsilon_1}(k), \tag{40}$$

$$C_{12}(k,\mu)=x\,P^0_{\rm mm}(k)\left\{R_{12}(k)\tilde\beta_1^{-1}(k)\tilde\beta_2^{-1}(k)+\mu^2\left[\tilde\beta_1^{-1}(k)R_1(k)+\tilde\beta_2^{-1}(k)R_2(k)\right]+\mu^4\right\}+P_{\epsilon_1\epsilon_2}(k), \tag{41}$$

$$C_{22}(k,\mu)=x\,P^0_{\rm mm}(k)\left[\tilde\beta_2^{-2}(k)+2\mu^2\tilde\beta_2^{-1}(k)R_2(k)+\mu^4\right]+P_{\epsilon_2\epsilon_2}(k). \tag{42}$$

Note that the quantity x encompasses all the relevant cosmological information about the growth of structure. From equations (40–42) one can see that the covariance matrix can easily be separated in two parts: a signal part $\hat S$ and a noise contribution $\hat N$:

$$C_{\rm 2tr}(k,\mu)=\hat S(k,\mu)+\hat N(k), \tag{43}$$

where the noise matrix is

$$\hat N_{ij}(k)=P_{\epsilon_i\epsilon_j}(k). \tag{44}$$

If the two tracers are both a Poisson sample of the dark matter field and do not overlap then $\hat N_{ij}(k)$ is diagonal, scale-independent and its elements are $\hat N_{11}=1/\bar n_1$; $\hat N_{22}=1/\bar n_2$. This is the case considered by McDonald & Seljak (2009). Any other source of stochasticity would add to the discreteness effect and, in general, may yield non-zero off-diagonal contributions $\hat N_{12}\neq 0$.

If we can estimate the noise part and $P^0_{\rm mm}(k)$ is given, e.g. by CMB observations, then the covariance matrix depends on six parametric functions: $x,\ \tilde\beta_1(k),\ \tilde\beta_2(k),\ R_1(k),\ R_2(k)$ and $R_{12}(k)$.

Considering one dark matter tracer, the covariance matrix is simpler:

$$C_{\rm 1tr}(k,\mu)=x\,P^0_{\rm mm}(k)\left[\tilde\beta^{-2}(k)+2\mu^2\tilde\beta^{-1}(k)R(k)+\mu^4\right]+P_{\epsilon\epsilon}(k), \tag{45}$$

and depends on only x, $\beta(k)$ and $R(k)$.

An interesting point is the 'hidden' relation between the different variables. For instance, given $\beta_1(k),\beta_2(k)$ and the relative number of objects of these two tracers, $\beta(k)$ is constrained. Also, in the two-tracer case, given $R_1(k)$ and $R_2(k), R_{12}(k)$ is also constrained. We give these relations in Appendices A and B.

2.3 Forecasting errors

We use the Fisher matrix formalism (Fisher 1935) to estimate errors on x. When the means of the data are fixed (i.e. for a given fiducial model), the Fisher matrix is given by (Tegmark, Taylor & Heavens 1997)

$$F_{\lambda\lambda'}=\frac{1}{2}{\rm Tr}\left[C_{,\lambda}C^{-1}C_{,\lambda'}C^{-1}\right], \tag{46}$$

where $C_{,\lambda}\equiv {\rm d}C/{\rm d}\lambda$, C is the covariance matrix and λ the parameters of the model. The marginalized variance of parameter λ is given by

$$\sigma^2_\lambda=(F^{-1})_{\lambda\lambda}\,. \tag{47}$$

Following e.g. Feldman, Kaiser & Peacock (1994) for a survey volume V_u, in the continuum approximation,

$$F^V_{\lambda\lambda'}=\frac{V_u}{(2\pi)^3}\int_{k_{\rm min}}^{k_{\rm max}}F_{\lambda\lambda'}(\boldsymbol{k}){\rm d}^3\boldsymbol{k} \tag{48}$$

which we evaluate by a discrete sum

$$F^V_{\lambda\lambda'}\simeq\frac{V_u}{(2\pi)^2}\sum_{\mu=-1}^{+1}\Delta\mu\sum_{k=k_{\rm min}}^{k_{\rm max}}\Delta k F_{\lambda\lambda'}(k,\mu)k^2\,. \tag{49}$$

In this paper, we assume that the fiducial power spectrum, the noise matrix and the R parameters are known and we fix them to their fiducial values. We will explore the dependence of the results on the assumed fiducial values in the following sections. The lambda parameters are $\tilde b_1,\ \tilde b_2$ and x or, equivalently, $\tilde\beta_1,\ \tilde\beta_2$ and x for the two-tracers case; when reporting errors on x, we marginalize over $\tilde\beta_1$ and $\tilde\beta_2$. For the one-tracer case the parameters are $\tilde b$ and x and the reported errors on x are marginalized over $\tilde b$.

The specific values of $k_{\rm min}$ and $k_{\rm max}$ depend on the features of the survey, $k_{\rm min}$ being set by the survey volume and $k_{\rm max}$ is usually set by the onset of non-linearities. In our case, we set $k_{\rm min}=2\pi/V_u^{1/3}$ and conservatively set $k_{\rm max}=0.1\,{\rm Mpc}^{-1}\,h$ for $z=0$.

In the next section, we will compare the errors of x, obtained using the one- and the two-tracer approaches. To produce the figures, we assume we have a single snapshot at $z=0$, which corresponds to $f=0.483$ in a standard Λ cold dark matter (CDM) universe, the power spectrum is given by CAMB (Code for Anisotropies in the Microwave Background: Lewis, Challinor & Lasenby 2000), the sampling volume is set to be $V_u=1\ ({\rm Gpc}\,h^{-1})^3$ and all biases and all R coefficients are taken to be scale-independent. The relative number of tracers is $Y\equiv\bar n_1/\bar n_2$, and the S/N is ${\rm S/N}\equiv\tilde b^2P^0(k=0.1\,h\,{\rm Mpc}^{-1})\bar n$. Note that the S/N is defined relative to the *combined* sample of tracers. The relation between $\tilde b$ and $\tilde b_1,\tilde b_2$ is given in equation (A8). Since the fractional cosmic-variance error on the power spectrum (in a shell in Fourier space) is constant with redshift, the quantities reported below are valid at different z provided that the S/N and the various bias parameters are defined at the redshift of interest. Note however that the fiducial value of x (and that

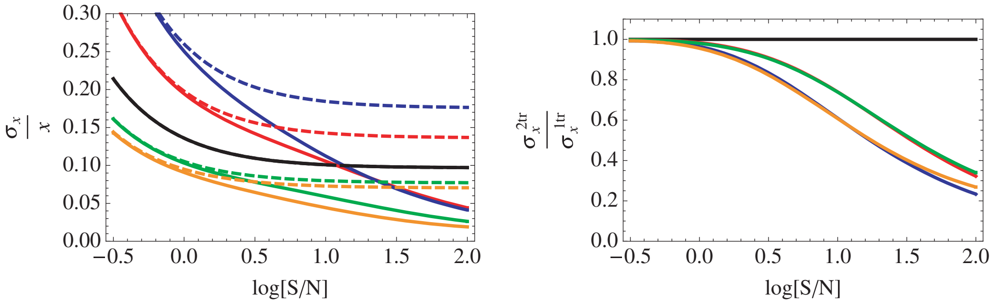

Figure 2. Left-hand panel: σ_x/x versus S/N. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In both panels $\tilde{b}_1 = 1$ and (from bottom to top) $\tilde{b}_2 = 1/3$ (orange line), 1/2 (green line), 1 (black line), 2 (red line) and 3 (blue line). For the combined sample according to equation (A8) the total bias factors are $\tilde{b} = 2/3$ (orange line), $\tilde{b} = 3/4$ (green line), $\tilde{b} = 1$ (black line), $\tilde{b} = 3/2$ (red line) and $\tilde{b} = 2$ (blue line). In the left-hand panel the solid lines represent the errors for two-tracer case, the dashed lines for one tracer.

of f) change with redshift: x increases with redshift up $z = 0.5$ and decreases for larger z, while f increases with redshift tending to 1 asymptotically. We find that the dependence of the fractional error, σ_x/x, and of the ratio of errors, $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$, on the value of x is weak. More importantly, the k_{max} at which non-linearities become important is expected to depend on redshift and to increase roughly as $(1 + z)$. The number of independent modes N in a given volume grows roughly like k_{max}^3 and the variance scales like $1/N$.

2.4 Dependence on bias

Given that the forecasted error on x depends on many variables, we will start by considering the effects of one variable at a time. The first important effect to be analysed is how the bias of the tracers (absolute and relative) affects the measurements of x (and therefore of f). For simplicity, we assume linear biasing (i.e. all R parameters equal to 1) and the same number of objects for two distinct populations ($Y = 1$) with diagonal Poisson-like noise.

In Fig. 2, we show σ_x/x (left-hand panel) and $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ (right-hand panel) versus S/N. The two-tracer case indicates that the two tracers are treated separately yielding a covariance matrix as in equations (36–42). The one-tracer case indicates that both tracers are included in the same sample and the covariance matrix is given by equation (45). In both panels, we have set $\tilde{b}_1 = 1$ (i.e. tracer 1 is unbiased) and $\tilde{b}_2$ ranges from 0.33 to 3: $\tilde{b}_2 = 1/3$ (orange line), 1/2 (green line), 1 (black line), 2 (red line), 3 (blue line). The bias factors for the combined sample according to equation (A8) are $\tilde{b} = 2/3$ (orange line), 3/4 (green line), 1 (black line), 3/2 (red line) and 2 (blue line). In the left-hand panel, the solid lines correspond to the two-tracer case, the dashed lines to the one-tracer case.

From Fig. 2 (left-hand panel), we see that for the single-tracer case (dashed lines), the lower the combined bias $\tilde{b}$, the lower the error σ_x. This can be understood if we recall that to measure f we are using the redshift-space distortion effect. Since this effect is proportional to $1/\tilde{b}$, the lower the bias, the larger the redshift-space distortions. As noted in McDonald & Seljak (2009), this means that, slightly counter-intuitively, low-bias objects may act as useful tracers, even if they are used as a single population. On the other hand, for the two-tracer case (solid lines) we can see that (i) the improvement (compared to the one-tracer case) is significant only when the signal dominates (log[S/N] $\gtrsim 0.7$) and (ii) the improvement increases as the difference in the biases of the two populations increases, also as noted by McDonald & Seljak (2009). Note that McDonald & Seljak (2009) measure the normalization of $P_{\theta\theta}$ (with the shape given by external observables) which is a different quantity from x considered here. However the fractional forecasted errors in the two quantity are the same, $\sigma_x/x = \sigma_{P_{\theta\theta}}/P_{\theta\theta}$ as well as the relative two-tracer versus one-tracer improvement $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} = \sigma_{P_{\theta\theta}}^{2\mathrm{tr}}/\sigma_{P_{\theta\theta}}^{1\mathrm{tr}}$.

Fig. 2 (right-hand panel) further quantifies the effect. The improvement between the two cases depends on the *ratio of biases* and is only significant if the S/N is large enough. In particular, for a bias ratio of 3 (blue and orange lines), the improvement is significant ($\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \leq 0.5$) for log[S/N] ≥ 1.3. However, for ratios of 2 (red and green lines), the improvement starts to be significant only when log[S/N] ≥ 1.5. Note also that for the special case $\tilde{b}_2 = \tilde{b}_1$ (black line), as expected, there is no improvement. When comparing these results – especially Fig. 2 (left-hand panel) – with those of McDonald & Seljak (2009), one should keep in mind that they define the S/N at $k = 0.4\,h\,\mathrm{Mpc}^{-1}$ while we use $k = 0.1\,h\,\mathrm{Mpc}^{-1}$, and that $P(k = 0.1\,h\,\mathrm{Mpc}^{-1}) \simeq 12\,P(k = 0.4\,h\,\mathrm{Mpc}^{-1})$. With this in mind, we reproduce their results.

We also conclude that for surveys with a low S/N (log[S/N] $\lesssim 0.7$), it is better to have a low-bias tracer; splitting the sample and using two tracers will not yield significant improvement. For example, for a bias of $\tilde{b} = 0.75$ (green dashed line in left-hand panel of Fig. 2), we obtain fractional errors of $\sigma_x/x \simeq 0.08$ for a log[S/N] = 0.7. On the other hand, for surveys with higher S/N (log[S/N] $\gtrsim 0.7$), it is better to use the two-tracer case, choosing two tracers with the highest possible bias ratio. In this case, for biases of $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 0.5$ (green solid line in left-hand panel of Fig. 2), we reach $\sigma_x/x \simeq 0.04$ for log[S/N] = 1.5. In practice, of course, the choice of which tracers to use is complicated by their number density; a high bias may be desirable, but one probably pays a penalty through low density and high shot noise.

2.5 Effect of bias non-linearities

The second interesting issue is to consider how non-linearities in the bias (i.e. the R parameters) affect σ_x. In this case, we fix $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$ with the same number of objects for each tracer ($Y = 1$) and Poisson noise. In Fig. 3 and 4 (left-hand panels), we show how $\sigma_x^{1\mathrm{tr}}/x$

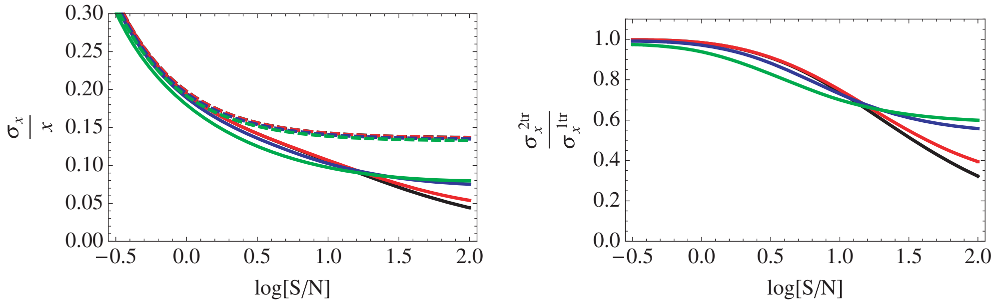

Figure 3. Left-hand panel: fractional error σ_x/x versus S/N. Dashed lines represent the single-tracer case and solid lines the two-tracer case. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In both panels the colours show different non-linear cases: $R_1 = 1.00$ (black solid line), $R_1 = 0.99$ (red solid line), $R_1 = 0.90$ (blue solid line) and $R_1 = 0.80$ (green solid line); $R_2 = 1$ and $R_{12} = R_1$. For dashed lines, R is the corresponding value for the full sample given the above values for R_2 and R_{12} as in equation (A4): $R = 1.000$ (black dashed line), $R = 0.999$ (red dashed line), $R = 0.989$ (blue dashed line) and $R = 0.978$ (green dashed line).

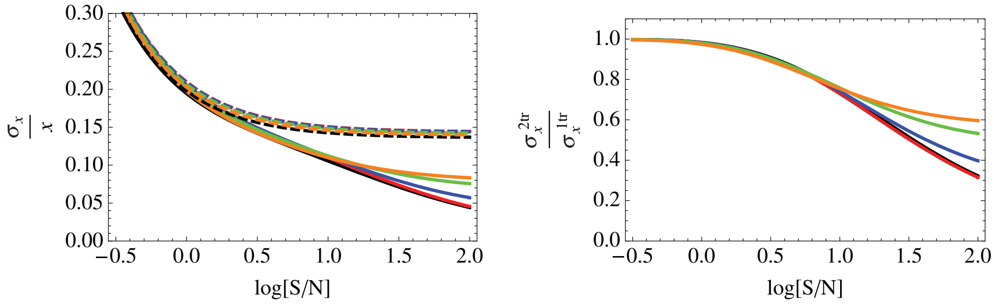

Figure 4. Left-hand panel: fractional error σ_x/x versus S/N. Dashed lines represent the single-tracer case and solid lines the two-tracer case. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In both panels, the colours show different non-linear cases: $R_{12} = 1.00$ (red solid line), $R_{12} = 0.99$ (blue solid line), $R_{12} = 0.95$ (green solid line) and $R_{12} = 0.90$ (orange solid line); $R_1 = R_2 = 0.9$. The solid black line is the prefect linear case. For dashed lines, R is the corresponding value for the full sample given the above values for R_1, R_2 and R_{12} as in equation (A4): $R = 1.000$ (black dashed line), $R = 0.900$ (red dashed line), $R = 0.902$ (blue dashed line), $R = 0.910$ (green dashed line) and $R = 0.921$ (orange dashed line).

(dashed lines) and $\sigma_x^{2\mathrm{tr}}/x$ (solid lines) vary with S/N. In both cases, the black line is for the perfect linear bias case $R_1 = R_2 = R_{12} = 1$ and colour lines show different non-linear cases (see Figs 3 and 4 captions for details). In Figs 3 and 4 (right-hand panels), we show how the ratio $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ varies with S/N.

In general, we see from Figs 3 and 4 (left-hand panels) that the two-tracer case is more sensitive to non-linear bias effects than the one-tracer case for high S/N (log[S/N] $\geq$ 1.2). In particular, in Fig. 3, for the single-tracer case a deviation from unity of R_1 produces a slight reduction of the error which is the same for all the S/N range explored. This is due to a reduction of the combined bias $\tilde{b}$: as we reduce R_{12} (because we set $R_1 = R_{12}$), $\tilde{b}$ is reduced (see equation A7) and therefore the total error is also reduced as we have seen in Section 2.4. On the other hand, for the two-tracer case, there are two opposite behaviours depending on the value of S/N: for log[S/N] $\gtrsim$ 1.2, we observe that non-linearities produce an increase in the error, whereas for log[S/N] $\lesssim$ 1.2, they produce a reduction. In the high S/N regime, reducing R_1 reduces the μ^2 coefficient in equation (40), and thus reduces the angular dependence. In the low S/N regime, this is compensated by the fact that the off-diagonal terms of the covariance matrix are reduced by non-linearities (recall that the noise off-diagonal terms are set to zero here). While this may not be clear at first sight from equation (41), we have verified it numerically.

From Fig. 3 (right-hand panel), we see that the non-linear bias increases the improvement between the two approaches for log[S/N] $\lesssim$ 1.2, and limits it for log[S/N] $\gtrsim$ 1.2. In particular, non-linear bias with $R_1 = 0.9$ and $R_2 = 1$ gives $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.6$ for log[S/N] = 1.5, whereas for the perfect linear bias case it is $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.5$. At lower S/N, namely log[S/N] = 0.7, we obtain for the same non-linear bias case, $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.85$, and for the linear bias case, $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.9$. Therefore, non-linearities affect the two-tracer approach considerably more than the single-tracer approach. Non-linearities in the mean bias relation slightly increase the precision of the x measurement for low S/N regime (log[S/N] $\lesssim$ 1.2), but they limit the effectiveness of the two-tracer approach for high S/N (log[S/N] $\gtrsim$ 1.2).

On the other hand, from Fig. 4 we observe a very similar behaviour. In this case, we have set $R_1 = R_2 = 0.9$ and we change the value of R_{12}. First of all, we observe that the one-tracer case is not very sensitive to non-linearities in this range of Rs. The small changes for one tracer are mainly due to the change of the combined bias as we have noted above for Fig. 3. The second point is that little deviations from $R_{12} = 1.0$ produce an increase on the fractional error for the two-tracer case for log[S/N] $\geq$ 1.2, as it can be seen in Fig. 4 (left-hand panel). Also in Fig. 4 (right-hand panel), we observe that the ratio of errors increases quickly for log[S/N] $\geq$ 1.2, as we leave $R_{12} = 1.00$

In summary, it is important to note (Figs 2, 3 and 4) that the two-tracer method yields a substantial improvement compared to the one-tracer approach only in the high S/N regime and if the non-linearity parameters R are close to unity. Even if R_1 is 0.9, with $R_2 = 1$ or $R_1 = R_2 = 0.9$ with $R_{12} = 0.95$, the gain saturates at about a factor of two. In Section 3.2, we will address the issue of whether dark matter

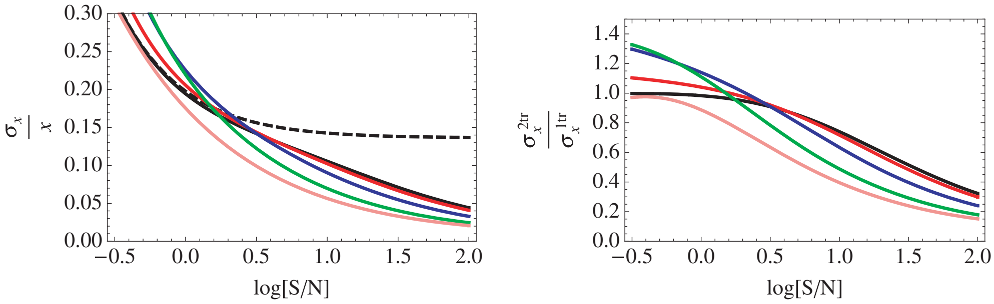

Figure 5. Left-hand panel: fractional error σ_x/x versus S/N. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. For both panels, $N_{12}/N_{11} = 0$ (black line), $N_{12}/N_{11} = 0.4$ (red line), $N_{12}/N_{11} = 0.8$ (blue line), $N_{12}/N_{11} = 0.9$ (green line) and $N_{12}/N_{11} = 1.0$ (pink line). The dashed line on the left-hand panel corresponds to the one-tracer case which is not affected by N_{12}.

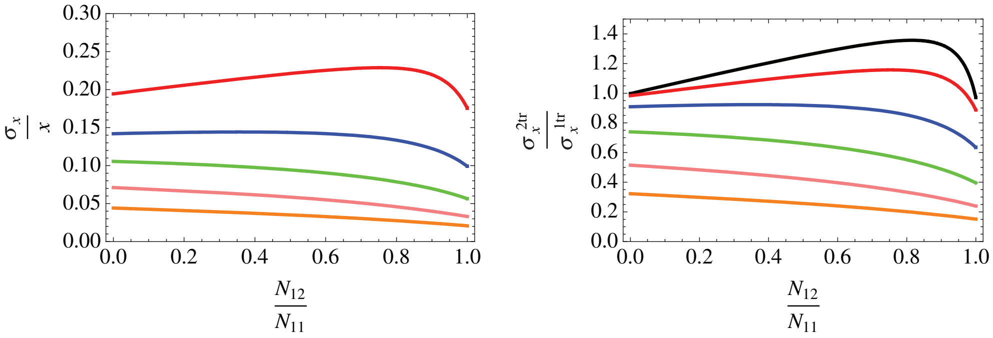

Figure 6. Left-hand panel: fractional error σ_x/x versus N_{12}/N_{11}. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus N_{12}/N_{11}. For both panels, $\log \mathrm{S/N} = -0.5$ (black line), $\log \mathrm{S/N} = 0.0$ (red line), $\log \mathrm{S/N} = 0.5$ (blue line), $\log \mathrm{S/N} = 1.0$ (green line), $\log \mathrm{S/N} = 1.5$ (pink line) and $\log \mathrm{S/N} = 2.0$ (orange line).

haloes, as seen in N-body simulations, trace the underlying dark matter with a bias that is linear enough for two tracers to be significantly advantageous compared to one.

2.6 Effect of off-diagonal noise terms

So far we have assumed that the noise matrix in equation (43) is diagonal; but any source of stochasticity in addition to Poisson sampling of two disjunct set of objects will add extra contributions to the noise matrix, which are non necessarily diagonal. Here we explore how σ_x/x and $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ change for a non-zero off-diagonal noise term. For simplicity we discuss the linear bias case with $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$ and the same number of objects for each tracer. Here, for direct comparison with previous examples, we will still set the diagonal elements of the noise matrix by the number density of the tracers (as if it was Poisson), and since $Y = 1, N_{11} = N_{22}$, but we allow $N_{12} \neq 0$. Note however that in any realistic application, a process that adds non-zero off-diagonal noise terms will also increase the diagonal matrix elements. This should be kept in mind during the following discussion.

In Fig. 5, we show how the error σ_x/x (left-hand panel) and the ratio of errors $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ (right-hand panel) change with S/N for different values of N_{12}: from the reference case $N_{12} = 0$ (black-solid line) to $N_{12} = N_{11}$ (pink-solid line).[3] The black-dashed line represents the one-tracer case, which is not affected by changes in N_{12}.

We see a different behaviour in the high S/N regime ($\log[\mathrm{S/N}] \gg 0$) and in the low S/N regime ($\log[\mathrm{S/N}] \lesssim 0$). In the high S/N regime, the higher the off-diagonal noise, the lower the error σ_x. This can be understood if we imagine the off-diagonal noise term as a correlation term between the noise terms. The more correlated the noise of the two tracers, the less the total noise; for high values of N_{12}, knowing N_{11} means knowing N_{22}. In the low S/N regime, we observe the opposite behaviour: the higher the N_{12}, the higher the error. We also observe that when the value of N_{12} is very close to N_{11}, then σ_x decreases abruptly. In the low-signal case, adding a non-perfect correlation between noise terms just means adding more noise, and only in the case this correlation between noise terms is nearly perfect ($N_{12} \simeq N_{11}$) means an improvement in the measure.

In Fig. 6, we show the fractional error of x (left-hand panel) and the ratio of errors between two- and one-tracer case (right-hand panel) versus N_{12}/N_{11} for different S/N regimes: from $\log(\mathrm{S/N}) = -0.5$ (black line) to $\log(\mathrm{S/N}) = 2.0$ (orange line). Here, the same effect is observed. For high S/N (orange, pink and green lines), increasing N_{12} decreases the error, whereas for low S/N (black, red and blue lines) the error increases. Here, the effect on the error when $N_{12} \rightarrow \mathrm{N}_{11}$ can be seen more clearly.

---

[3] Note that, the maximum value for N_{12} is $\sqrt{N_{22}N_{11}}$, as can be deduced from the Cauchy–Schwarz inequality: $|\langle\epsilon_1\epsilon_2\rangle|^2 \le \langle\epsilon_1^2\rangle\langle\epsilon_2^2\rangle$.

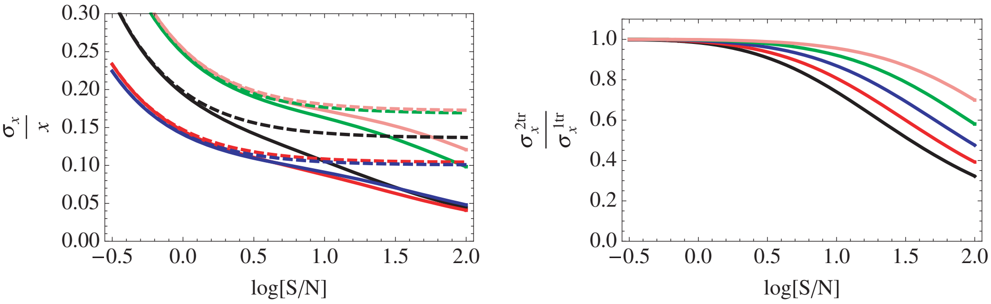

Figure 7. Left-hand panel: σ_x/x versus S/N. Right-hand panel: $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N. In the left-hand panel, the dashed-lines are the errors for the one-tracer model and the solid-lines for the two-tracer model. For both panels $Y = 1/20$ (pink line), $Y = 1/10$ (green line), $Y = 1$ (black line), $Y = 10$ (red line) and $Y = 20$ (blue line). $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$ are assumed.

On the right-hand panels of Figs 5 and 6, for low S/N and for non-zero values of N_{12}, we have that $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} > 1$. This effect is due to the fact that we are using equation (A12) to relate the noise elements of the two-tracer case with the single-tracer case. However equation (A12) only provides a correct relation among the diagonal noise matrix elements when the off-diagonal noise terms are zero, which is no longer the case. In fact, the noise for a single tracer built out of two tracers for which the noise matrix is strongly non-diagonal is not strictly Poisson. Therefore it cannot be fully described by the Poisson noise that it would have if the two tracers had a diagonal noise matrix (as equation A12 assumes). However, we have no other way to model it, and we thus stick to equation (A12). This effect is important only in the low S/N regime and where N_{12} is comparable to the diagonal terms ($N_{12} \lesssim N_{11}$)

2.7 Dependence on the relative number density

Finally it is also interesting to see how the relative number of tracers can affect the error of x. We vary the ratio between the number densities of tracers, $\bar{n}_1$ and $\bar{n}_2$, namely $Y \equiv \bar{n}_1/\bar{n}_2$, keeping the total number of tracers, $\bar{n}$, fixed. Again, we assume linear bias, Poisson noise and that the biases are $\tilde{b}_1 = 1$ and $\tilde{b}_2 = 2$.

In Fig. 7 (left-hand panel), we show the error of x versus S/N for $Y = 1/20$ (pink line), 1/10 (green line), 1 (black line), 10 (red line) and 20 (blue line), for the one-tracer model (dashed lines) and for the two-tracer one (solid lines). In the right-hand panel, we show the ratio $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}}$ versus S/N for different values of Y using the same colour scheme.

Note that $Y > 1$ means that the highly biased tracer has the lower number density; for $Y < 1$, the highly biased tracer has the higher number density. From Fig. 7 (left-hand panel), we see that the error on the one-tracer model is as expected from Fig. 2: an increase in Y causes a reduction in σ_x/x, because it is equivalent to reducing the effective bias $\tilde{b}$ (recall that $\tilde{b}_1 < \tilde{b}_2$). On the other hand, for the two-tracer model, we observe that the fractional error is lower if the low bias tracer is more abundant than the high-bias one. This is also expected because as we have seen in Fig. 2: the lower the bias, the smaller is the error of x. In Fig. 7 (right-hand panel), we see that the maximal improvement for the two-tracer approach compared to the one-tracer approach is realized when the two tracers number densities are equal, independently of the S/N. For unequal number densities, the two-tracer approach gives a better improvement over the one-tracer approach, if the number density of the highly biased tracer is lower than that of the tracer with lower bias.

3 EXPECTED VALUES FOR PARAMETERS DESCRIBING BIAS AND STOCHASTICITY

As we have seen, the improvements achievable by using two tracers, depend on various features of the tracers population, such as the S/N, the bias parameters and the amount of bias non-linearity. In the next two sections, we explore what are plausible and realistic values if dark matter haloes are taken to be the tracers. We first use analytical arguments, and in the next section, N-body simulations.

3.1 Extended Press–Schechter approach

In this section, we identify dark matter haloes with the peaks of an initially Gaussian field, and compute their number densities and biases. We assume that the tracers (haloes) are linearly biased, as the non-linear corrections to halo bias derived in this frameworks are very small. The volume effect that may arise in this formalism is due to the z-dependence of the parameters, in particular, the bias. A narrow–deep survey has a strong z-dependence in the bias and a wide–shallow has a very weak one. It is because of this, that in this section we assume a volume-limited cubic survey of comoving side 1 Gpc h^{-1} for all z, and we perform the analysis at different values of z. This way it is easier to understand how the z-dependence of the bias affects the errors of the one- and two-tracer model. We set $k_{\mathrm{min}} = 2\pi/V_u^{1/3}$ and $k_{\mathrm{max}} = 0.1D(0)/D(z)\,\mathrm{Mpc}\,h^{-1}$. $P^0(k)$ is given by CAMB for a standard Λ CDM universe. We consider a range of redshift between 0 and 4, and we parametrize the redshift dependence of f as, $f(z) = \Omega_{\mathrm{m}}(z)^{\gamma}$, with $\gamma = 0.56$. We choose the two tracers to be haloes of masses $10^{12}\,\mathrm{M}_{\odot}\,h^{-1} < M < 10^{13}\,\mathrm{M}_{\odot}\,h^{-1}$ for tracer 1 and $10^{13}\,\mathrm{M}_{\odot}\,h^{-1} < M < 10^{14}\,\mathrm{M}_{\odot}\,h^{-1}$ for tracer 2.

The number density of these haloes is related to the halo mass function, given by

$$n(M, z) = \frac{2\rho_{\rm m}}{M\,\sigma(M)} f(\nu)\nu \left|\frac{{\rm d}\sigma(M)}{{\rm d}M}\right|, \tag{50}$$

where $\sigma(M)$ is the rms of the power spectrum linearly extrapolated at $z = 0$ filtered with a top-hat sphere of mass M, $\rho_{\rm m}$ is the mean density of the Universe (we set it at $\rho_{\rm m} = 7 \times 10^{10}\,{\rm M}_\odot\,h^{-1}\,{\rm Mpc}^{-3}$) and $\nu \equiv \delta^2_{\rm sc}(z)/\sigma^2(M)$. Here, $\delta_{\rm sc}(z)$ denotes the critical threshold for collapse and is given by $\delta_{\rm sc}(z) \simeq 1.686/D(z)$. We use the Sheth & Tormen (1999) mass function:

$$\nu f(\nu) = A(p)[1 + (q\nu)^{-p}] \left(\frac{q\nu}{2\pi}\right)^{1/2} \exp(-q\nu/2), \tag{51}$$

where $p = 0.3, q = 0.75$ and the normalization factor $A(p) = [1 + 2^{-p}\Gamma(1/2 - p)/\Gamma(1/2)]^{-1}$. Thus, the number density of objects is

$$\bar{n}(M_1, M_2, z) \equiv \int_{M_1}^{M_2} n(M, z)\,{\rm d}M. \tag{52}$$

Assuming Poisson noise, we can directly relate this to the noise matrix elements,

$$N_{ii}(z) = 1/\bar{n}(M_{1i}, M_{2i}, z)\,. \tag{53}$$

As long as we are considering Poisson noise, the off-diagonal terms of the noise matrix are 0.

For the bias dependence we assume a linear bias ($\tilde{b} = \hat{b} \equiv b$). The bias of an object of mass M at redshift z is given by (Kaiser 1984; Mo, Jing & White 1997; Scoccimarro et al. 2001)

$$b(M, z) = 1 + \frac{1}{D(z)}\left[q\frac{\delta_{\rm sc}(z)}{\sigma^2(M)} - \frac{1}{\delta_{\rm sc}(z)}\right]. \tag{54}$$

The bias of a set of objects with masses between M_1 and M_2 is given by

$$\bar{b}(M_1, M_2, z) \equiv \frac{\int_{M_1}^{M_2} {\rm d}M\, n(M, z) b(M, z)}{\bar{n}(M_1, M_2, z)}\,. \tag{55}$$

In Fig. 8 (left-hand panel), we show how the bias of tracer 1 ($10^{12}\,{\rm M}_\odot\,h^{-1} < M < 10^{13}\,{\rm M}_\odot\,h^{-1}$ red-solid line), tracer 2 ($10^{13}\,{\rm M}_\odot\,h^{-1} < M < 10^{14}\,{\rm M}_\odot\,h^{-1}$ blue-solid line) and the combined sample (black-dashed line) change as a function of redshift. Bias increases with redshift roughly as $1/D^2(z)$. As there are many more lower mass haloes, the overall bias is dominated by this population. In Fig. 8 (right-hand panel), we show the S/N as a function of redshift. As before, we define ${\rm S/N}(z) \equiv \bar{b}^2(z)P^0(k = 0.1)D^2(z)[\bar{n}_1(z) + \bar{n}_2(z)]$, where $\bar{b}(z)$ corresponds to the bias of the whole sample. Note that $b^2(z)\,P^0(k = 0.1)D^2(z)$ goes roughly $\propto 1/D(z)^2$, but the number density of objects (the mass function) drops exponentially rapidly: the S/N decreases with increasing redshift; the maximum value of S/N is at $z = 0$, where ${\rm S/N} \simeq 15$.

In Fig. 9, we show the errors σ_x/x corresponding to the one- and two-tracer case (left-hand panel) and its ratio (right-hand panel) as a function of the redshift. In Fig. 9 (left-hand panel), we see that there is a minimum in the value of σ_x/x at $z \simeq 1.5$, both for the one- and for

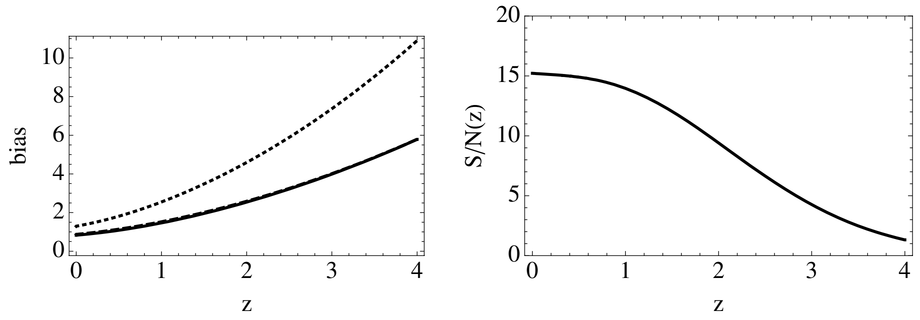

Figure 8. Left-hand panel: the bias as a function of redshift for tracer 1 ($10^{12}\,{\rm M}_\odot\,h^{-1} < M < 10^{13}\,{\rm M}_\odot\,h^{-1}$ solid line), tracer 2 ($10^{13}\,{\rm M}_\odot\,h^{-1} < M < 10^{14}\,{\rm M}_\odot\,h^{-1}$, dotted line) and for the whole sample (black-dashed line). Right-hand panel: S/N for the full sample (see text for definition) as a function of redshift.

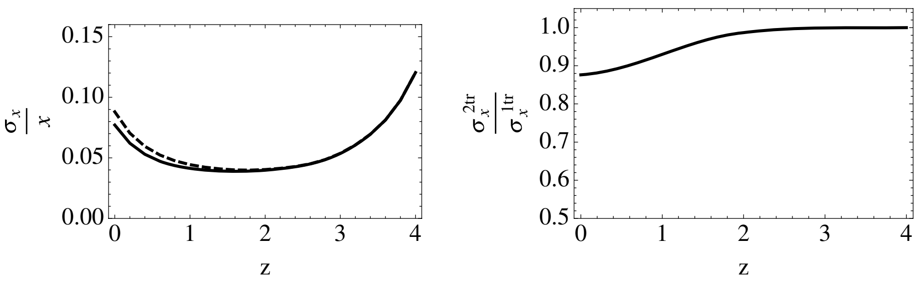

Figure 9. Left-hand panel: σ_x/x versus z for the one-tracer case (dashed line) and for the two-tracer case (solid line). Right-hand panel: $\sigma_x^{2{\rm tr}}/\sigma_x^{1{\rm tr}}$ versus z.

the two-tracer cases. This seems paradoxical (at least for the one-tracer model) because we have seen in Section 2.4 that as we increase the bias, the error on x increases. Furthermore, shot noise increases with redshift. The explanation of this seemingly paradoxical effect has to do with the fact that not only the noise and the bias change with z, but also $k_{\max}(z)$ and $x(z)$. Recall that $x(z)$ is our signal and that the fractional error per interval of k goes like the square root of the number of modes, i.e. $k^3_{\max} - k^3_{\min}$. It turns out that, as we increase the available range of k of equation (49), the error σ_x/x is reduced, scaling as these considerations indicate.

The last effect dominate at low z. At higher z, the noise and the bias effect dominate, and therefore σ_x/x increases with z as we have seen in Section 2.4. In Fig. 9 (right-hand panel), we see that only at low redshifts ($z < 1$), where the S/N is high, the improvement between the two cases is significant, reaching the minimum value of $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.88$ at $z = 0$. This result may seem worse than that we have shown in the last section for $\mathrm{S/N} \simeq 15$. However, taking into account that now the ratio of biases is not 2, but $\simeq$1.5 and also recalling that now $Y \neq 1$, the fractional error increases from $\simeq$0.7 to $\simeq$0.88.

From Fig. 9, it is clear that if the survey is dominated by low-redshift objects (wide-shallow surveys), then splitting the sample may reduce the error of x. However, the higher precision in measuring x is reached with surveys whose volume configuration results in most objects having $z \simeq 1.5$)

A mass selection of haloes as here may provide a modelling applicable to Sunyaev–Zel'dovich-selected clusters or, in an idealized way, to luminous red galaxies. But other type of surveys may select tracers in a radically different way; for example, emission-line-selected blue galaxies have a bias that evolves with redshift much more slowly than considered above. We therefore also consider a complementary, yet still highly idealized, way to select haloes and explore whether in that case the improvement in splitting the sample can be much larger. We select tracers by their peak height, i.e. their ν of equation (50), keeping the maximum and minimum ν (rather than the mass) of each tracer sample constant in redshift. We have explored different cuts and found the following. (i) The improvement in splitting the sample is maximized when the two samples have comparable number of objects. When, to maximize the bias difference between the two samples, the highly biased tracer include objects that are very rare, the shot noise for that sample become important and the gain in splitting the sample decreases. (ii) By suitably choosing the ν cuts, we have sampled the parameter space and managed to achieve $\sigma_x^{2\mathrm{tr}}/\sigma_x^{1\mathrm{tr}} \simeq 0.6$, but we have not been able to improve the gain further. For instance, choosing as tracer 1 structures with ν between 0.9 and 1.5 and as tracer 2 structures with ν between 1.5 and 20, we reach a gain of 0.6 at $z \simeq 2$.

Before any practical application of these findings, one should bear in mind that we have assumed a volume-limited sample (i.e. that all haloes in the required range are detected). In addition, we have considered a fixed survey volume seen at different redshift: for a given sky coverage, the volume per unit redshift increases with redshift for $z \lesssim 2.5$. Finally, we have selected haloes in a very idealized way. In practice, the selection will be likely applied on galaxies, which halo occupation distribution is not straightforward. We will discuss this further in Section 4.

3.2 Simulations

As we have seen in Figs 3 and 4, the gain of splitting the sample is dependent on the non-linearity parameters, specially for high S/N. In this case, Rs have to be very close to unity to exhibit substantial gain. In addition we have so far relied on the Poisson sampling assumption, which may not hold in details. For tracers that can be identified with dark matter haloes, these issues can be addressed by looking at N-body simulations.

We choose a flat ΛCDM cosmology with cosmological parameters consistent with current observational data. More specifically, the cosmological parameters of the simulation are $\Omega_{\mathrm{m}} = 0.27, \Omega_\Lambda = 0.73, h = 0.7, \Omega_{\mathrm{b}}h^2 = 0.023, n_{\mathrm{s}} = 0.95$ and $\sigma_8 = 0.8$. Our cosmological simulation consists of 1024^3 particles in a volume of $(1\,\mathrm{Gpc}\,h^{-1})^3$. This results in a particle mass of about $7 \times 10^{10}\,\mathrm{M}_\odot\,h^{-1}$. The initial conditions of the simulation were generated at redshift $z = 65.67$, by displacing the particles according to the Zel'dovich approximation from their initial grid points. The initial power spectrum of the density fluctuations was computed by CAMB (Lewis et al. 2000).

Taking only the gravitational interaction into account, the simulation was performed with GADGET-2 (Springel 2005) using a softening length of comoving $30\,\mathrm{kpc}\,h^{-1}$ and a particle mesh grid size of 2048^3. The chosen mass resolution and force resolution enable us to resolve haloes with masses above $\simeq 10^{12}\,\mathrm{M}_\odot\,h^{-1}$, i.e. each halo contains at least 15 particles. We identify haloes at redshift $z = 0$ by the friends-of-friends algorithm with a linking length of 0.2 times the mean interparticle separation. We split the haloes in two mass bins: $10^{12}\,\mathrm{M}_\odot\,h^{-1} < M < 10^{13}\,\mathrm{M}_\odot\,h^{-1}$ (M12) and $M > 10^{13}\,\mathrm{M}_\odot\,h^{-1}$ (M13). The mass bins M12 and M13 consist of about 2.1 and 0.4 million haloes, respectively.

In order to derive the mean conditional bias (see equation 1), $b(\delta)$, for the two different tracers, we first compute the halo overdensity, δ_{h}, and matter overdensity, δ, by assigning the haloes and particles, respectively, on a 512^3 grid using the cloud in cell scheme. The overdensities are then further smoothed by a Gaussian filter, $\exp(-k^2 l_{\mathrm{s}}^2/2)$, where we choose the smoothing length to be $l_{\mathrm{s}} = 2.5$, 5, 10 and 20 $\mathrm{Mpc}\,h^{-1}$. One expects that any stochasticity, non-linear or non-local effects on the bias relation should decrease as the field is smoothed with increasing smoothing lengths. The biased density as a function of the matter overdensity, $b(\delta)\delta$, is then computed by averaging the δ_{h} in the corresponding δ bin (see equation 1). Using the mean bias relation so obtained, we can compute the noise field ϵ on the grid by applying equation (5). After Fourier transforming the different fields using the same 512^3 grid, we can compute the power spectra and cross power spectra of the different quantities by spherical averaging the product of their Fourier modes, i.e. for example $P_{\epsilon\epsilon}(k) = \langle \epsilon(\boldsymbol{k})\epsilon^*(\boldsymbol{k})\rangle$ and $P_{\mathrm{hm}}(k) = \langle \delta_{\mathrm{h}}(\boldsymbol{k})\delta^*(\boldsymbol{k})\rangle$. Note that, in what follows, in evaluating the bias and R parameters from the simulations we have used explicitly equations (31–35), the noise terms are not assumed to be Poissonian but are computed directly from the ϵ field.

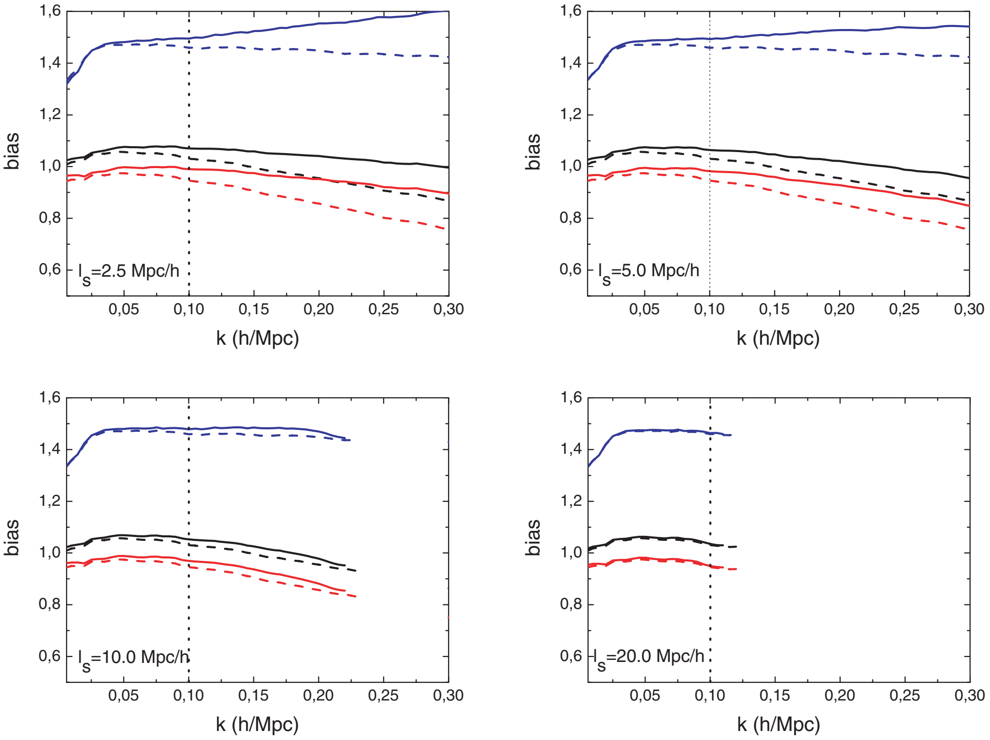

Figure 10. Bias parameters as a function of the scale k obtained from simulations. $\hat{b}$ are the dashed lines and $\tilde{b}$ the solid lines. The black lines correspond to the whole sample of haloes, whereas red lines corresponds to sample M12 ($10^{12}\,\mathrm{M}_{\odot} < M < 10^{13}\,\mathrm{M}_{\odot}$) and blue lines to sample M13 ($M > 10^{13}\,\mathrm{M}_{\odot}$). The smoothing length is 2.5 for the top-left panel, 5.0 for the top-right panel, 10.0 for the bottom-left panel and 20.0 Mpc h^{-1} for the bottom-right panel.

In Fig. 10, we show the bias parameters $\hat{b}$ (dashed lines) and $\tilde{b}$ (solid lines) versus k, for the whole sample (black lines), for M12 (red lines) and for M13 (blue lines) for different smoothing lengths.

The vertical dotted line marks $k = 0.1\,h\,\mathrm{Mpc}^{-1}$, which is the typical scale at $z = 0$ where non-linearities appear. In this paper, we always work in the linear regime, where $k < 0.1\,D(0)/D(z)\,h\,\mathrm{Mpc}^{-1}$. From these plots, we can see that $\hat{b}$ is robust to changes in l_s and also is approximately scale-independent. $\tilde{b}$ is also very close to be scale invariant and varies little with the smoothing scale. In order to have a numerical reference, $\tilde{b}$ changes by about 7 per cent and $\hat{b}$ by about 5 per cent compared with their values for $l_\mathrm{s} = 2.5$ Mpc h^{-1}, over a range from $k_\mathrm{min} = 0.01\,h\,\mathrm{Mpc}^{-1}$ to $k_\mathrm{max} = 0.1\,h\,\mathrm{Mpc}^{-1}$. For $l_\mathrm{s} = 20$ Mpc h^{-1}, both parameters change by about 7 per cent in the same k range. On the other hand, for $k \simeq 0.1\,h\,\mathrm{Mpc}^{-1}$, $\tilde{b}$ changes by about 3 per cent as l_s goes from 2.5 to 20.0 Mpc h^{-1}; $\hat{b}$ changes less than 1 per cent in the same range.

In Fig. 11, we show the values of the non-linearity parameters versus k obtained from simulations: $R(k)$ (black line), $R_1(k)$ (red line), $R_2(k)$ (blue line) and $R_{12}(k)$ (green line). Different panels correspond to different smoothing scales l_s, as in Fig. 10. All non-linearity parameters decrease as the scale decreases, but this trend disappears as we increase the smoothing length. When the smoothing length is large enough ($\simeq$20 Mpc h^{-1}), all R parameters are approximately scale-invariant and are very close to 1.

In Fig. 12 (left-hand panel), we show the different noise components obtained from simulations versus k, for different smoothing lengths: 2.5 (red lines), 5.0 (blue lines), 10.0 (green lines) and 20.0 Mpc h^{-1} (orange lines). The solid lines are $\langle \epsilon_1(\boldsymbol{k})\epsilon_1^*(\boldsymbol{k})\rangle$ (M12), the dashed lines $\langle \epsilon_2(\boldsymbol{k})\epsilon_2^*(\boldsymbol{k})\rangle$ (M13) and the dotted lines the cross terms between the two samples, $\langle \epsilon_1(\boldsymbol{k})\epsilon_2^*(\boldsymbol{k})\rangle$. The noise terms for the whole sample are not shown for clarity. In order to get a better comparison between all these noise terms, we have removed the effects of the smoothing, dividing each noise by the filter squared.

The black lines are the noise predictions assuming a Poisson-like noise ($\langle \epsilon_i(\boldsymbol{k})\epsilon_i^*(\boldsymbol{k})\rangle = 1/\bar{n}_i$). The solid line is for the M12 sample and the dashed line for M13. The cross term is relatively small (black dotted line). We observe that the M13 noise is sub-Poisson, whereas the one for M12 is super-Poisson. This is in agreement with the findings of Seljak et al. (2009). It has been noted before (Smith et al. 2007) that for massive haloes the noise could be sub-Poisson. At scales smaller than the ones of interest here, Smith et al. (2007) ascribe this to halo-exclusion effects. Noise above the Poisson level is expected if other sources of stochasticity affect halo formation. The formation and evolution of dark matter haloes is a highly complicated process: dark matter haloes grow through a mixture of smooth accretion, violent encounters and fragmentation. In the classical extended Press–Schechter/excursion set theory, haloes are identified with initial density peaks and the computation of the halo mass function (and thus as a derived quantity the halo bias) is mapped into a first passage process in the presence of a sharp barrier. This yields a deterministic halo bias, but cannot capture the full physical complications inherent to a realistic description of halo formation. In addition, numerical simulations show that there is not a good correspondence between peaks in the initial density field and collapsed haloes (see Katz, Quinn & Gelb 1993; Seljak & Warren 2004). Recently Maggiore & Riotto (2009) proposed to

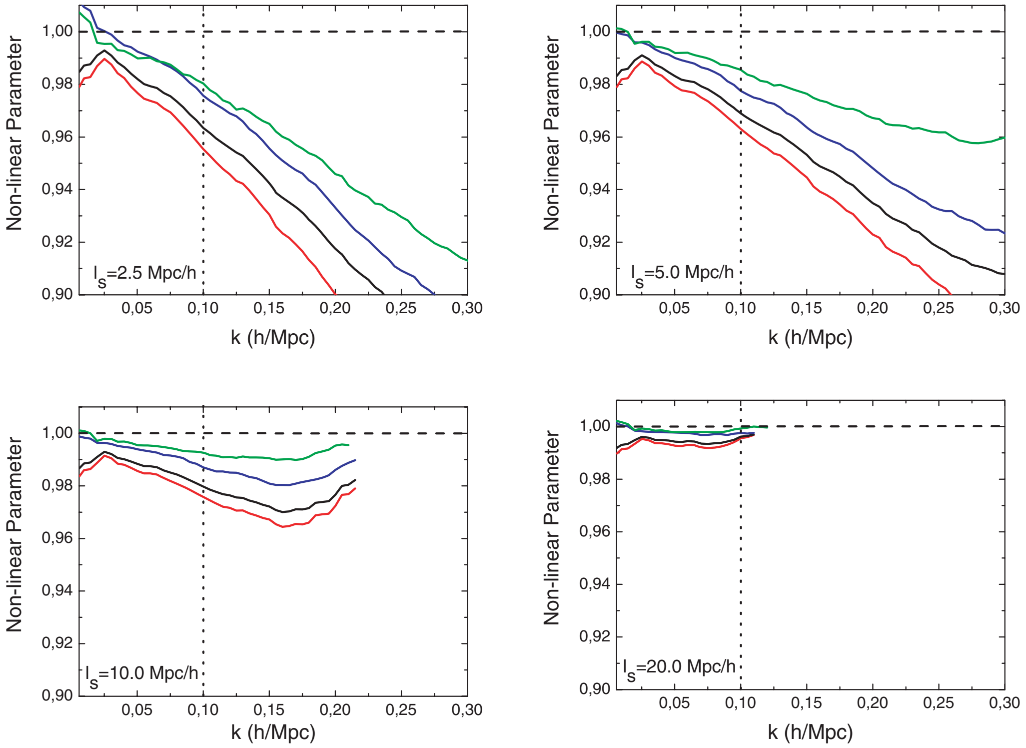

Figure 11. Non-linearity parameters, R (black line), $R_1(k)$ (red line), $R_2(k)$ (blue line) and $R_{12}(k)$ (green line), obtained from simulations for different smoothing lengths, as in Fig. 10. The subscripts 1 and 2 refer to M12 sample and M13 respectively, whereas the no subscript refers to the whole sample. The vertical dotted line, $k = 0.1\,h\,\mathrm{Mpc}^{-1}$, marks the limit of linear regime. The horizontal dashed line marks the maximum value for all Rs.

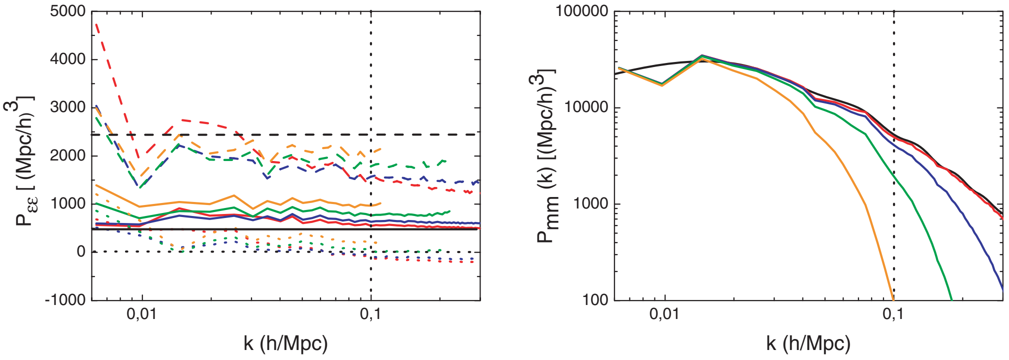

Figure 12. Noise terms (left-hand panel) and power spectrum (right-hand panel) as a function of the scale obtained from simulations. In the left-hand panel, $\langle\epsilon_1(k)\epsilon_1^*(k)\rangle$ (M12) solid lines, $\langle\epsilon_2(k)\epsilon_2^*(k)\rangle$ (M13) dashed lines and $\langle\epsilon_1(k)\epsilon_2^*(k)\rangle$ dotted lines. The black lines are the prediction for a Poisson-like noise. In the right-hand panel, $P_{\mathrm{mm}}(k)$ from CAMB (black line) and $P_{\mathrm{mm}}(k)$ from simulations (colours lines). In both panels, the colours refer to the smoothing length, 2.5 (red), 5.0 (blue), 10.0 (green) and 20.0 Mpc h^{-1} (orange).

include these effects, at least at an effective level, by taking into account that the critical value for collapse is not a fixed constant but itself a stochastic variable. This will naturally lead to an extra ‘noise’ component in the halo bias.

In the Fig. 12 (right-hand panel), we show the signal component, $P_{\mathrm{mm}}(k)$, obtained from simulations for different smoothing lengths. The colour notation is the same as that in left-hand panel. In this case, the black line is the linear theory prediction for the same cosmological parameters used in our simulations. The effect of the sampling variance can be clearly seen at large scales. The enhancement of the clustering at small scales due to the nonlinear gravitational evolution is hidden by the smoothing of the density field.

In order to apply the findings from simulations to our model, we make the following assumptions.

(i) The bias is scale-independent and does not change with the smoothing scale. We take $\tilde{b} = 1.05$ for the whole sample, $\tilde{b}_1 = 0.97$ for M12 and $\tilde{b}_2 = 1.47$ for the M13 sample. This assumption is supported by Fig. 10.

(ii) The non-linearity parameters have a linear dependence with scale k. For each smoothing length, we fitted the best linear relation up to $k = 0.1\ h\,\mathrm{Mpc}^{-1}$.

(iii) The noise is Poisson-like. Therefore the diagonal terms of the noise are $N_{ii}(k) = \alpha_i W^2(k \cdot l_{\mathrm{s}})/\bar{n}_i$, where α_i is a parameter that takes into account the deviations from the ideal Poisson noise (see Fig. 12, left-hand panel), and $W(k \cdot l_{\mathrm{s}})$ is the smoothing filter. The

Table 2. Parameters values used in plots of Fig. 13 as function of the smoothing scales used.

l_s in Mpc h^{-1}	2.5	5.0	10.0	20.0
$\tilde{b}$	1.05	1.05	1.05	1.05
$\tilde{b}_1$	0.97	0.97	0.97	0.97
$\tilde{b}_2$	1.47	1.47	1.47	1.47
$R(k)$	$1.0 - 0.39208k$	$0.99737 - 0.26854k$	$0.99666 - 0.1585k$	$0.99587 - 0.01972k$
$R_1(k)$	$0.99939 - 0.4033k$	$0.99591 - 0.31089k$	$0.99567 - 0.18733k$	$0.99479 - 0.02141k$
$R_2(k)$	$1.00752 - 0.29446k$	$1.00145 - 0.22144k$	$0.99941 - 0.11259k$	$0.99876 - 0.02276k$
$R_{12}(k)$	$0.99922 - 0.16926k$	$0.99883 - 0.12419k$	$0.99922 - 0.06535k$	$0.99905 - 0.01082k$
α	1.250	1.325	1.675	2.10
α_1	1.365	1.428	1.785	2.100
α_2	0.680	0.680	0.760	0.840

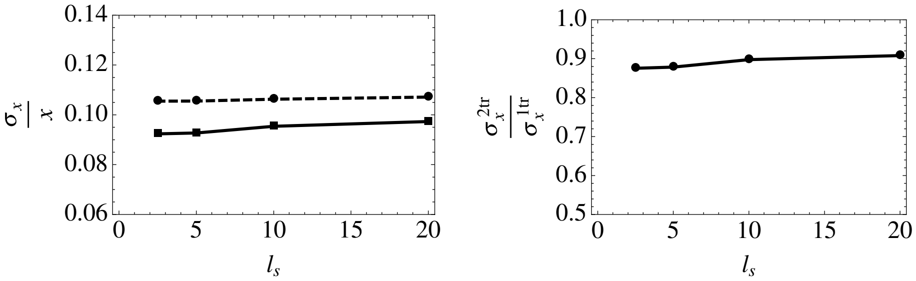

Figure 13. Left-hand panel: error of x as a function of the smoothing length l_s. The dashed line is for one-tracer model, whereas the solid line for the two-tracer one. Right-hand panel: ratio between the two models.

α values used are shown in Table 2. According to the number of haloes of the two tracers and the volume of the simulation, we have $\bar{n} = 2.5 \times 10^{-3}\, h^3/\mathrm{Mpc}^3$, $\bar{n}_1 = 2.1 \times 10^{-3}\, h^3/\mathrm{Mpc}^3$ and $\bar{n}_2 = 4.0 \times 10^{-4}\, h^3/\mathrm{Mpc}^3$.

(iv) We take the off-diagonal noise terms to be zero. This has some support from the simulations (Fig. 12, left-hand panel), where this term is $N_{12}/\sqrt{N_{11}N_{22}} < 0.2$

(v) Finally, we assume that the cross-correlation terms between the ϵ field and the matter field are 0.

We summarize all these assumptions in Table 2. Applying these conditions yield the results shown in Fig. 13. In the left-hand panel, the fractional errors for one- (dashed line) and two-tracer model (solid line) are shown as a function of the smoothing scale l_s, and the ratio of these two errors is shown in the right-hand panel.

The improvement between the two cases under these assumptions is rather modest: $\sigma_{2\mathrm{tr}}/\sigma_{1\mathrm{tr}} \simeq 0.9$. This is because it is mainly dominated by the ratio of biases – we have $\tilde{b}_2/\tilde{b}_1 \simeq 1.5$ – if we are in a region where S/N $\simeq 10$. This result is robust to small changes in the bias modelling, e.g. perfectly linear bias. We also have tried to fit the off-diagonal term to some non-zero value, but we have found that doing this does not produce any significant change in the plots of Fig. 13.

There may be some merit in splitting the sample in a different way; by choosing samples with very different biases, the gains should be larger, but in practice to do this almost certainly requires one sample to be of rare objects, which will have very high shot noise.

4 DISCUSSION & CONCLUSIONS

We have revisited the method of circumventing sample variance in the measurement of $f = \mathrm{d}\ln D/\mathrm{d}\ln a$ (D being the linear growth factor), based on comparing the clustering properties of two differently biased tracers of the dark matter distribution. This method was recently investigated by McDonald & Seljak (2009), although a similar technique in a different context was presented in Pen (2004). Along the same lines, Slosar (2009) and Seljak (2009) propose to compare clustering of differently biased tracers to circumvent sample variance in the measurement of primordial non-Gaussianity.

Most of the statistical power of these measurements comes from very large scales where cosmic variance is the dominant contribution to the statistical error-bars. By suppressing cosmic variance, this approach promises to reduce drastically error-bars on cosmologically very important quantities; for example it would allow for a high-precision determination of growth of structure as a function of redshift, as encoded in fD, and an improvement of dark energy figures of merit by large factors.

All these approaches assume that the observed objects (i.e. galaxies) trace the dark matter deterministically; the galaxy density field is assumed to be proportional to the dark matter field with the constant of proportionality given by a single parameter, the bias. This goes under the name of the linear bias model. An important underlying assumption is therefore made, that there is no stochasticity between the tracer field and the dark matter field on the scales of interest, which is expected to breakdown at some level, at least on small scales. While the linear bias model has been extremely successful in cosmology (e.g. Reid et al. 2010 and references therein), it is well known that the linear bias model might provide a good description for the galaxy power spectrum even if the relation between the galaxy and dark matter overdensities is not that of a linear bias (e.g. Heavens, Matarrese & Verde 1998).

Galaxies are believed to form inside dark matter haloes, but their formation probability as function of halo mass and their exact radial distribution is still the subject of active research. The process of halo occupation by galaxies is expected to be stochastic to some extent, but the details of the galaxy distribution within haloes are expected to become increasingly unimportant on large scales. Here, we simplify the issue by assuming (possibly with an over-simplification) that dark matter haloes can be used as tracers. The linear deterministic bias model however is known not to be a perfect description of halo clustering and that the relation between dark matter and haloes and between haloes of different masses is stochastic. For example, Seljak & Warren (2004) point out that 'the fluctuations between haloes and the initial or final matter fields are never below 10–20 per cent' and that 'the scatter between the fields in individual modes is significant and one cannot assume that the fields are simply proportional to one another'. This was further explored and quantified by Bonoli & Pen (2009). Note that the halo overdensity field is expected to have a stochastic component even if it was a perfect Poisson sampling of a linearly biased dark matter field, but the above references and N-body simulations show that there are additional sources of stochasticity beyond shot noise.

We have thus set out to generalize the approach of McDonald & Seljak (2009), by assuming that the bias of haloes may not be perfectly linear and allowing for some stochasticity. We have computed the expected error on the quantity fD achievable by comparing clustering of differently biased tracers (thus suppressing cosmic variance) and by combining the different tracers in a single sample (thus reducing shot noise and stochasticity but carrying along sample variance in full).

We have analysed how the bias, the noise, the non-linearity and stochasticity affect the measurements of fD and explored in which S/N regime it is significantly advantageous to split a galaxy sample in two differently biased tracers. We used results from simulations to set plausible values for these parameters to see how great the gains may be in practice. We find that even small amount of stochasticity (either in the form of Poisson noise or in more general form) and of non-linearity can limit significantly the performance of the two-tracers approach. In our analysis, we also have assumed a scale-independent bias. This may be enough, for the mass range studied, if we consider only dark matter haloes as tracers. This is indeed what we have seen in simulations. On the other hand, it is also true that more realistic approaches, which account for galaxies as tracers instead of haloes, should include a scale-dependent bias. However, including this in our formalism can only reduce the gain achievable by splitting the sample. We expect the ratio of errors increases from the current value of 0.9 to even closer values to 1 if the bias is strongly scale-dependent.

We have shown that only in the very high S/N regime it is significantly advantageous to split the sample and that, even though the gain is maximized by increasing the *ratio* of the biases of the two tracers, both tracers should be well sampled. We have explored different ways of selecting and splitting dark matter haloes obeying a Λ CDM mass function and found that one can achieve up to a 40 per cent reduction of the error on fD. While this would correspond to the gain from a three times larger survey volume if the two tracers were not to be split, it is much smaller that the improvement forecasted in the absence of stochasticity and bias non-linearity.

In addition, we should note that these findings apply to dark matter haloes as tracers, while realistic surveys would select galaxies: the galaxy–host halo relation is likely to introduce extra stochasticity which would reduce the gain further. The formalism we have developed, however, is general enough that can be used to optimize survey design and tracers selection and optimally split (or combine) tracers to minimize the error on the cosmologically interesting quantities.

ACKNOWLEDGMENTS

HGM is supported by a CSIC JAE grant. LV acknowledges support from FP7-PEOPLE-2007-4-3-IRG n. 202182 and FP7-IDEAS Phys.LSS 240117. LV, RJ and CW are supported by MICINN grant AYA2008-03531. The N-body simulation was performed at the Leibniz Rechenzentrum Munich using German Grid infrastructure provided by AstroGrid-D. LV & RJ acknowledge support from World Premier International Research Center Initiative (WPI initiative), MEXT, Japan.

REFERENCES

Bacon D. J. et al., 2005, MNRAS, 363, 723
Bonoli S., Pen U. L., 2009, MNRAS, 396, 1610
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Dekel A., Lahav O., 1999, ApJ, 520, 24
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Feldman H. A., Frieman J. A., Fry J. N., Scoccimarro R., 2001, Phys. Rev. Lett., 86, 1434
Fisher R. A., 1935, J. R. Statistical Soc., 98, 39
Fry J. N., 1994, Phys. Rev. Lett., 73, 215
Guzzo L. et al., 2008, Nat, 451, 541
Hamilton A. J. S., 1998, ed., Astrophys. Space Sci. Library, Vol. 231, The Evolving Universe. Kluwer, Dordrecht, p. 185
Hawkins E. et al., 2003, MNRAS, 346, 78
Heavens A. F., Matarrese S., Verde L., 1998, MNRAS, 301, 797
Hoekstra H., van Waerbeke L., Gladders M. D., Mellier Y., Yee H. K. C., 2002, ApJ, 577, 604
Kaiser N., 1984, ApJ, 284, L9
Kaiser N., 1987, MNRAS, 227, 1

Katz N., Quinn T., Gelb J. M., 1993, MNRAS, 265, 689
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Maggiore M., Riotto A., 2009, preprint (arXiv:0903.1250)
McDonald P., Seljak U., 2009, J. Cosmol. Astropart. Phys., 10, 7
Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton University Press, Princeton, NJ, p. 435
Pen U.-L., 2004, MNRAS, 350, 1445
Reid B. A. et al., 2010, MNRAS, 404, 60
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Seljak U., 2009, Phys. Rev. Lett., 102, 021302
Seljak U., Warren M. S., 2004, MNRAS, 355, 129
Seljak U. et al., 2005, Phys. Rev. D, 71, 043511
Seljak U., Hamaus N., Desjacques V., 2009, Phys. Rev. Lett., 103, 091303
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Slosar A., 2009, J. Cosmol. Astropart. Phys., 3, 4
Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
Springel V., 2005, MNRAS, 364, 1105
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Verde L. et al., 2002, MNRAS, 335, 432

APPENDIX A: RELATIONS BETWEEN ONE-TRACER CASE AND TWO-TRACER CASE

The parameters $R(r), R_1(r), R_2(r)$ and $R_{12}(r)$ are not fully independent, the same is true for the set of parameters $\tilde{\beta}(r)$, $\tilde{\beta}_2(r)$ and $\tilde{\beta}_1(r)$, and also the different noise matrices. Here we make explicit their relation.

A1 Relation between the *R* parameters

Let the tracer (galaxies, haloes...) overdensity $\delta_{\mathrm{g}}(\boldsymbol{x})$ be defined as

$$\delta_{\mathrm{g}}(\boldsymbol{x}) \equiv \frac{\rho_{\mathrm{g}}(\boldsymbol{x})}{\bar{\rho}_{\mathrm{g}}} - 1 = \frac{n_{\mathrm{g}}(\boldsymbol{x})}{\bar{n}_{\mathrm{g}}} - 1, \tag{A1}$$

where $\rho_{\mathrm{g}}(\boldsymbol{x})$ is the tracer density at $\boldsymbol{x}$. In the second equality, we have used the fact that $\rho_{\mathrm{g}}(\boldsymbol{x}) \propto n_{\mathrm{g}}(\boldsymbol{x})$ with $n_{\mathrm{g}}(\boldsymbol{x})$ the number of tracers at $\boldsymbol{x}$. The total number density of galaxies is $n_{\mathrm{g}}(\boldsymbol{x}) = n_{g1}(\boldsymbol{x}) + n_{g2}(\boldsymbol{x})$. Defining the ratio of number of galaxies as

$$Y \equiv \frac{\bar{n}_1}{\bar{n}_2}, \tag{A2}$$

we can write the overdensity of galaxies as

$$\delta_{\mathrm{g}}(\boldsymbol{x}) = \frac{\delta_{\mathrm{g}1}(\boldsymbol{x})}{1+Y^{-1}} + \frac{\delta_{\mathrm{g}2}(\boldsymbol{x})}{1+Y}. \tag{A3}$$

Recalling the definitions of $R(r), R_{12}(r), R_1(r), R_2(r)$ we find that

$$R(r) = \frac{R_1(r)Y\tilde{\beta}_2(r) + R_2(r)\tilde{\beta}_1(r)}{\sqrt{\tilde{\beta}_2^2(r)Y^2 + \tilde{\beta}_1^2(r) + 2R_{12}(r)Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)}}. \tag{A4}$$

Note that, as expected, when $Y \to 0, \mathrm{R}(r) \to R_2(r)$ and when $Y \to \infty, \mathrm{R}(r) \to R_1(r)$.

This last equation is also valid in k-space according to the definitions of $\tilde{\beta}_i(k)$, $\tilde{b}_i(k)$, $R_i(k)$ and $R_{12}(k)$. This is because the definitions of all these parameters are mathematically symmetric in Fourier and configuration space.

A2 Relation between the $\tilde{\beta}$ parameters

Similarly, we can derive the relation between $\tilde{\beta}(r)$, $\tilde{\beta}_1(r)$ and $\tilde{\beta}_2(r)$. Recalling the definition of $\tilde{\beta}$s as,

$$\tilde{\beta}_i(r) \equiv \frac{f}{\tilde{b}_i(r)} \qquad \tilde{\beta}(r) \equiv \frac{f}{\tilde{b}(r)} \tag{A5}$$

then the relation between $\tilde{\beta}$s is

$$\tilde{\beta}_i(r) = \tilde{\beta}(r)\left[\frac{\tilde{b}^2(r)}{\tilde{b}_i^2(r)}\right]^{1/2}. \tag{A6}$$

From this we obtain

$$\tilde{\beta}^{-2}(r) = \frac{\tilde{\beta}_1^{-2}(r)}{(1+1/Y)^2} + \frac{\tilde{\beta}_2^{-2}(r)}{(1+Y)^2} + \frac{2R_{12}(r)\tilde{\beta}_1^{-1}(r)\tilde{\beta}_2^{-1}(r)}{(1+Y)(1+1/Y)}. \tag{A7}$$

Again, this relation is valid for both configuration space and k-space because of reasons of symmetry in the definitions of the parameters. Clearly, we can also write down a relation between the $\tilde{b}$ parameters:

$$\tilde{b}^2 = \frac{\tilde{b}_1^2}{(1+1/Y)^2} + \frac{\tilde{b}_2^2}{(1+Y)^2} + \frac{2R_{12}\tilde{b}_1\tilde{b}_2}{(1+Y)(1+1/Y)}. \tag{A8}$$

A3 Relation between the noise terms

Let the noise matrix for one-tracer model be

$$N_{1\mathrm{tr}} = N \tag{A9}$$

and for the two-tracer model,

$$N_{2\mathrm{tr}} = \begin{pmatrix} N_{11} & N_{12} \\ N_{12} & N_{22} \end{pmatrix} \tag{A10}$$

both in k-space.

Assuming Poisson noise terms, and distinct populations, the off-diagonal terms are zero, $N_{12} = 0$. Setting $\bar{n}_1$ and $\bar{n}_2$ to be the number density of galaxies of type 1 and 2 respectively, and $\bar{n} = \bar{n}_1 + \bar{n}_2$ the total number of galaxies, we can say that

$$(N_{11})^{-1} = \bar{n}_1 \qquad (N_{22})^{-1} = \bar{n}_2 \qquad (N)^{-1} = \bar{n}. \tag{A11}$$

Finally, we obtain

$$N_{11} = N(1+Y^{-1}) \qquad N_{22} = N(1+Y). \tag{A12}$$

A4 Constraints between the non-linearity parameters

Given the definitions of the non-linearity coefficients, using the Cauchy–Schwarz inequality we find that

$$-1 \le R(r),\, R_1(r),\, R_2(r),\, R_{12}(r) \le 1. \tag{A13}$$

However, the negative values for these Rs parameters represent a negative bias for tracers relative to the dark matter, with a doubtful physical connection. For this reason, we restrict the possible values for these parameters to the range

$$0 \le R(r),\, R_1(r),\, R_2(r),\, R_{12}(r) \le 1. \tag{A14}$$

The parameters $R_1(r), R_2(r)$ and $R_{12}(r)$ are not totally independent but are related by the condition of equation (B11),

$$1 - R_1^2(r) - R_2^2(r) - R_{12}^2(r) + 2R_1(r)R_2(r)R_{12}(r) \ge 0\,. \tag{A15}$$

If we isolate $R_{12}(r)$ as a function of $R_1(r)$ and $R_2(r)$, the last equation becomes

$$R_1(r)R_2(r) - \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)} \le R_{12}(r) \le R_1(r)R_2(r) + \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)}. \tag{A16}$$

This equation is enough if we are working only with the two-tracer model. However, if we want to compare this model with the one-tracer model, we have to make sure that also the $R(r)$ parameter is between 0 and 1 (see equation A4):

$$0 \le \frac{R_1(r)Y\tilde{\beta}_2(r) + R_2(r)\tilde{\beta}_1(r)}{\sqrt{\tilde{\beta}_2^2(r)Y^2 + \tilde{\beta}_1^2(r) + 2R_{12}(r)Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)}} \le 1. \tag{A17}$$

Since $R_1(r), R_2(r), Y$ and $\tilde{\beta}_i(r)$ are always positive, the first inequality always holds, and we find that

$$R_{12}(r) \ge \frac{(R_1(r)Y\tilde{\beta}_2(r) + \tilde{\beta}_1(r)R_2(r))^2 - Y^2\tilde{\beta}_2^2(r) - \tilde{\beta}_1^2(r)}{2Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)} \tag{A18}$$

to satisfy the second.

This minimum value could be lower or higher than the one given by the equation (A16) depending on the values of the other parameters. Therefore, the limits for $R_{12}(r)$ are

$$R_{12}(r) \ge \max\left\{ R_1(r)R_2(r) - \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)},\ \frac{[R_1(r)Y\tilde{\beta}_2(r) + \tilde{\beta}_1(r)R_2(r)]^2 - Y^2\tilde{\beta}_2^2(r) - \tilde{\beta}_1^2(r)}{2Y\tilde{\beta}_1(r)\tilde{\beta}_2(r)} \right\}$$

$$R_{12}(r) \le R_1(r)R_2(r) + \sqrt{R_1^2(r)R_2^2(r) + 1 - R_1^2(r) - R_2^2(r)}. \tag{A19}$$

As we said before, also this last equation is valid for k-space parameter because of reasons of symmetry in the definitions.

APPENDIX B: CONSTRAINTS ON THE NON-LINEAR COEFFICIENTS

Suppose there are three possibly correlated fields, x, y, z (in our application these would correspond to $\delta_{\rm g1}$, $\delta_{\rm g2}$, $\delta_{\rm m}$), and the corresponding non-linear coefficients are

$$r_1^2 = \frac{\langle xz\rangle^2}{\langle x^2\rangle\langle z^2\rangle}, \tag{B1}$$

$$r_2^2 = \frac{\langle yz\rangle^2}{\langle y^2\rangle\langle z^2\rangle}, \tag{B2}$$

$$r_3^2 = \frac{\langle xy\rangle^2}{\langle x^2\rangle\langle y^2\rangle}. \tag{B3}$$

Using the Cauchy–Schwarz inequality, we can state that

$$0 \le r_i^2 \le +1. \tag{B4}$$

We want to know what are the constraints on the triplet r_1, r_2 and r_3. To solve this, consider

$$C \equiv \langle (x + \lambda y + \mu z)^2\rangle \ge 0. \tag{B5}$$

This is at least zero for all λ and μ, and in particular for the values which minimize C, namely λ' and μ':

$$\left.\frac{\partial C}{\partial \lambda}\right|_{\lambda',\mu'} = \langle (x + \lambda' y + \mu' z) y\rangle = 0, \tag{B6}$$

$$\left.\frac{\partial C}{\partial \mu}\right|_{\lambda',\mu'} = \langle (x + \lambda' y + \mu' z) z\rangle = 0. \tag{B7}$$

The system has an unique solution if and only if $r_1^2 \le 1$. In that case, the values are

$$\lambda' = \frac{1}{D}\left(\langle yz\rangle\langle xz\rangle - \langle xy\rangle\langle z^2\rangle\right) \tag{B8}$$

$$\mu' = \frac{1}{D}\left(\langle y^2\rangle\langle xz\rangle - \langle xz\rangle\langle yz\rangle\right), \tag{B9}$$

where $D = \langle y^2\rangle\langle z^2\rangle - \langle yz\rangle^2$. Substituting these values into equation (B5) for C gives

$$(1 - r_1^2)(1 - r_1^2 - r_2^2 - r_3^2 + 2r_1r_2r_3) \ge 0. \tag{B10}$$

Provided that $r_1^2 \ne 1$, we can write

$$1 - r_1^2 - r_2^2 - r_3^2 + 2r_1r_2r_3 \ge 0. \tag{B11}$$

Because of symmetry reasons we can say that this last equation holds if at least one of the rs is different from ±1. Note that if two of the rs are equal to one, so is the third. In our application, r_1, r_2, r_3 correspond to R_1, R_2, R_{12}.

This paper has been typeset from a TEX/LATEX file prepared by the author.

CHAPTER 4

The bispectrum of $f(R)$ cosmologies

In this Chapter we present the paper Gil-Marín et al. (2011b), where we explore the possibility of measuring deviations of GR using the bispectrum technique. In particular we use a suite of cosmological simulations of modified gravitational action $f(R)$ models and we compare them with ΛCDM simulations with the same power spectrum signal.

This paper was done in collaboration with Raúl Jiménez and Licia Verde from the Institut de Ciències del Cosmos at the Universitat the Barcelona, Spain; Wayne Hu from Kavli institute for cosmological physics at University of Chicago, USA; Fabian Schmidt at California Institute of Technology, USA, and was published in Journal of Cosmology and Astroparticle Physics (JCAP) in November 2011.

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

The bispectrum of $f(R)$ cosmologies

Héctor Gil-Marín,[a,b] **Fabian Schmidt,**[c] **Wayne Hu,**[d] **Raul Jimenez**[e,b] **and Licia Verde**[e,b]

[a]Institute of Space Sciences (IEEC-CSIC), Faculty of Science,
Campus UAB, Bellaterra 08193, Spain
[b]Institute of Sciences of the Cosmos (ICC-IEEC), University of Barcelona,
Barcelona 08024, Spain
[c]Theoretical Astrophysics, California Institute of Technology,
Mail Code 350-17, Pasadena, California 91125, U.S.A.
[d]Kavli Institute for Cosmological Physics, Department of Astronomy & Astrophysics,
University of Chicago,
Chicago, IL 60637, U.S.A.
[e]ICREA Institució Catalana de Recerca i Estudis Avançats,
Passeig Lluís Companys 23, 08010 Barcelona, Spain
E-mail: gil@ieec.uab.es, fabians@caltech.edu, whu@background.uchicago.edu,
raul.jimenez@icc.ub.edu, liciaverde@icc.ub.edu

Received September 9, 2011
Revised October 13, 2011
Accepted October 18, 2011
Published November 10, 2011

Abstract. In this paper we analyze a suite of cosmological simulations of modified gravitational action $f(R)$ models, where cosmic acceleration is induced by a scalar field that acts as a fifth force on all forms of matter. In particular, we focus on the bispectrum of the dark matter density field on mildly non-linear scales. For models with the same *initial* power spectrum, the dark matter bispectrum shows significant differences for cases where the final dark matter power spectrum also differs. Given the different dependence on bias of the galaxy power spectrum and bispectrum, bispectrum measurements can close the loophole of galaxy bias hiding differences in the power spectrum. Alternatively, changes in the initial power spectrum can also hide differences. By constructing ΛCDM models with very similar *final* non-linear power spectra, we show that the differences in the bispectrum are reduced ($\lesssim 4\%$) and are comparable with differences in the imperfectly matched power spectra. These results indicate that the bispectrum depends mainly on the power spectrum and less sensitively on the gravitational signatures of the $f(R)$ model. This weak dependence of the matter bispectrum on gravity makes it useful for breaking degeneracies associated with galaxy bias, even for models beyond general relativity.

Keywords: modified gravity, power spectrum, cosmological simulations

ArXiv ePrint: 1109.2115

JCAP11(2011)019

doi:10.1088/1475-7516/2011/11/019

Contents

1 Introduction

Observations of Type Ia supernovae suggest that the Universe has been accelerating since redshift $z \sim 0.5$ [11, 12]. Today the physical mechanism responsible for this process is still a mystery. The simplest model to explain the acceleration of the Universe is the ΛCDM (Lambda Cold Dark Matter model). This model assumes that the acceleration is driven by an exotic form of energy with negative pressure that might be related to the vacuum energy of quantum field theories. This theory is equivalent to adding an integration constant to the Einstein equations.

Alternative theories to the vacuum energy propose a modification of gravity in the infrared that would produce an accelerated expansion. One possibility are the $f(R)$ class of models (see [19] and references therein). These models produce accelerated expansion through a modification of the Einstein-Hilbert action by an arbitrary function of the Ricci scalar R. As a consequence, an extra propagating scalar field appears that mediates a fifth force on all forms of matter. The range of this force depends on the functional form of $f(R)$. In order to satisfy solar system tests, $f(R)$ models are often chosen to present a chameleon behavior. The chameleon mechanism makes the extra scalar field become increasingly massive in higher-curvature regions, suppressing the range of the fifth force in dense environments.

In previous works, cosmological simulations [9] have been used to study the power spectrum [10] and halo statistics [13] of these kinds of models. More recent studies with higher resolution have confirmed these previous results [21] and extended the investigation to smaller scales. In the present work we focus on how the dark matter bispectrum is modified in this class of models. While these models also predict a non-linear matter power spectrum different from the ΛCDM one, it is nevertheless interesting to look at the bispectrum for at least two reasons: *a)* except for gravitational lensing, measurements of clustering yield the galaxy or the baryon power spectrum, not the dark matter one: as baryonic physics and galaxy formation are complicated phenomena, the observed power spectrum may be biased, i.e. may differ significantly from the dark matter one; the bispectrum is well known for helping disentangle effects of gravity from effect of biasing e.g., [4, 20]. *b)* once we allow ourselves to consider non-standard models, the initial (linear) matter power spectrum does not have to be the power-law ΛCDM one to reproduce the observations. The form of the bispectrum kernel is a possible "signature" of gravity as it gets modified by any modifications from GR behavior e.g., [15].

Here we pay special attention to see whether the bispectrum can be used to break degeneracies between models with the same observed power spectrum and the same cosmology, but different gravity. We begin in section 2 with a review of non-linear gravitational dynamics in $f(R)$ models, in section 3 we briefly describe the simulations and in section 4 we introduce the density field statistics. We discuss the results in section 5 and conclude in section 6.

2 $f(R)$ gravity

The $f(R)$ class of models generalizes the Einstein-Hilbert action to include a function $f(R)$ of the Ricci scalar R,

$$S = \int d^4x \sqrt{-g} \left[\frac{R + f(R)}{16\pi G} + L_m \right] . \tag{2.1}$$

Here L_m is the Lagrangian of matter and we have assumed $c = \hbar = 1$. For standard GR with a cosmological constant, $f(R) = -16\pi G\rho_\Lambda$, whereas for modified gravity, the force modification is associated with an additional scalar degree of freedom $f_R \equiv df/dR$. In particular, in this paper we use the model for $f(R)$ proposed by [6],

$$f(R) \propto \frac{R}{AR + 1} , \tag{2.2}$$

where A is a constant with dimensions of length squared. We can write this equation as a function of its derivative evaluated at $\bar{R}_0$ (the background curvature today), namely f_{R0}. We adjust the proportionality constant to match some effective cosmological constant ρ_Λ in the limit where $f_{R0} \to 0$. For high enough curvature such that $AR \gg 1$, $f(R)$ can then be approximated as,

$$f(R) = -16\pi G\rho_\Lambda - f_{R0}\frac{\bar{R}_0^2}{R}. \tag{2.3}$$

The modified Einstein equations can be computed by varying the Einstein-Hilbert action (eq. (2.1)) with respect to the metric. We work in the quasistatic limit where the time derivatives are negligible compared to the spatial derivatives. In this regime, valid on scales much smaller than the horizon $1/H$, the trace of the modified Einstein equations yields the f_R field equation,

$$\nabla^2 \delta f_R = \frac{a^2}{3} \left[\delta R(f_R) - 8\pi G \delta\rho_m \right] , \tag{2.4}$$

where a is the scale factor, $\delta f_R = f_R(R) - f_R(\bar{R})$, $\delta R = R - \bar{R}$ and $\delta\rho_m = \rho_m - \bar{\rho}_m$. Here $\bar{R}$ is the background curvature that can be approximated by a ΛCDM universe for $|f_{R0}| \ll 1$ and ρ_m $(\bar{\rho}_m)$ is the (background) matter density.

On the other hand, the time-time component of the Einstein equations yields the modified Poisson equation,

$$\nabla^2 \Psi = \frac{16\pi G}{3} a^2 \delta\rho_m - \frac{a^2}{6} \delta R(f_R) \tag{2.5}$$

where $\Psi = \delta g_{00}/(2g_{00})$ is the Newtonian potential.

For small fluctuations of the field, we can approximate $\delta R \simeq (dR/df_R)|_{\bar{R}}\delta f_R$. We will refer to this linearization as the non-chameleon limit. Conversely if the field fluctuations are large enough such that $\delta R(f_R)$ cannot be linearized, the chameleon mechanism operates. We will refer to use of the exact, as opposed to linearized, equations as full $f(R)$ or just chameleon models.

The linearized field equations formed by eqs. (2.4)–(2.5) can be solved for the Newtonian potential as a function of the density field. In the linear approximation for δR, these two equations in Fourier space yield,

$$k^2\Psi(\mathbf{k}) = -4\pi G\left(\frac{4}{3} - \frac{1}{3}\frac{\bar{\mu}^2 a^2}{k^2+\bar{\mu}^2 a^2}\right) a^2\delta\rho_m(\mathbf{k}). \tag{2.6}$$

This equation is identical to the one in GR but with a modification of the gravitational constant,

$$G_{\text{eff}}(\mathbf{k},t) \equiv G\left(\frac{4}{3} - \frac{1}{3}\frac{\bar{\mu}(t)^2 a(t)^2}{k^2+\bar{\mu}(t)^2 a(t)^2}\right). \tag{2.7}$$

Here $\mu(R) \equiv (3df_R/dR)^{-1/2}$ is the effective mass of the scalar field f_R and $\bar{\mu}$ just stands for $\mu(\bar{R})$. The dependence on time is introduced though $\bar{R}(t)$. Note that when $f_R \to 0$, $G_{\text{eff}} \to G$ and we recover the ΛCDM limit, as expected. It is interesting to see that for a given value of f_{R0} there are two different regimes for G_{eff}, depending on whether the physical scale we are studying is larger or smaller than the inverse mass of the field. On large scales $k \ll \mu(t)a(t)$, $G_{\text{eff}} \to G$ and gravity behaves as GR, whereas on small scales $k \gg \mu(t)a(t)$ $G_{\text{eff}} \to 4G/3$ and gravity is stronger than in GR by a factor of 4/3.

In other words, in eq. (2.6), one assumes that the mass of the scalar field μ only depends on time and is the same in all regions of the Universe at a given epoch. However, for cosmologically interesting values of μ the field is then essentially massless within the Solar System. The presence of such a scalar field (fifth force) is ruled out by light deflection and time delay measurements in the Solar System, which are all consistent with GR. In the full non-linear $f(R)$ theory, $R \propto f_R^{-1/2}$ can become very large in dense environments, suppressing the field and restoring the GR relation $\delta R = 8\pi G\delta\rho_m$ (eq. (2.4)). Thus, gravity is not modified in the same way everywhere, but depends on environment. In regions with large potential wells (inside halos) the mass of the scalar field becomes large and therefore the effective range of interaction of this field shrinks recovering GR. We call this the chameleon mechanism.

3 Simulations

The simulations used in this paper are described in previous works [9, 10, 13]. Briefly, the field equation for f_R (eq. (2.4)) is solved on a regular grid using relaxation techniques and multigrid iteration. The potential Ψ is computed from the density and f_R fields following eq. (2.5) using the fast Fourier transform method. The dark matter particles are then moved according to the gradient of the computed potential, $-\nabla\Psi$, using a second order accurate leap-frog integrator.

The simulations were run using the values of $|f_{R0}| = 10^{-4}$, 10^{-5}, 10^{-6} (both for chameleon and non-chameleon cases) and 0, which is equivalent to ΛCDM.[1] The background expansion history for all cases differ from ΛCDM only at $\mathcal{O}(f_{R0})$ and are hence

---

[1] In this paper we are always using negative values for f_{R0}, so when we talk about f_{R0} we refer to its absolute value.

practically indistinguishable. The cosmology used is $\Omega_\Lambda = 0.76$, $\Omega_m = 0.24$, $\Omega_b = 0.04181$, $H_0 = 73\,\mathrm{km/s/Mpc}$ and initial power in curvature fluctuations $A_s = (4.89 \times 10^{-5})^2$ at $k = 0.05\,\mathrm{Mpc}^{-1}$ with a tilt of $n_s = 0.958$. This initial power spectrum does not include the effects of baryon acoustic oscillations. Specifically, the initial conditions for the simulations were created using ENZO [8], a publicly available cosmological N-body + hydrodynamics code. ENZO uses the Zel'dovich approximation to displace particles on a uniform grid according to the initial power spectrum. In order to propagate the initial power spectrum until late times the transfer function from [3] was used.

The simulations were started at $a = 0.02$ and are integrated in time in steps of $\Delta a = 0.002$. All simulations used here correspond to boxes of comoving size $L = 256$ and 400 Mpc/h with 512^3 grid cells and 256^3 particles. For each box size, we have 6 runs for each value of f_{R0}, with different realizations of the initial conditions.

4 Power spectrum and bispectrum

The simplest statistic of interest of the matter density field is the power spectrum $P(k)$, defined by the second moment of the Fourier amplitude of the density contrast,

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}')\rangle \equiv (2\pi)^3 \delta^D(\mathbf{k}+\mathbf{k}')P(k)\,, \tag{4.1}$$

where $\langle \dots \rangle$ denotes the ensemble average over different realizations of the Universe. By statistical isotropy, the power spectrum does not depend on the direction of the **k**-vector. In practice we only have one observable Universe, so the average $\langle \dots \rangle$ cannot be computed. However, using the isotropy of the power spectrum we can compute the average over all different directions for each **k**-vector. Note also that $P(k)$ is defined to be real. Since $\mathbf{k} = -\mathbf{k}'$, $\delta(\mathbf{k})\delta(\mathbf{k}') \sim |\delta(\mathbf{k})|^2$, which is a real number.

The second statistic of interest is the bispectrum B, defined by,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle \equiv (2\pi)^3 \delta^D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3). \tag{4.2}$$

The Dirac delta function δ^D, ensures that the bispectrum is defined only for **k**-vector configurations that form closed triangles: $\sum_i \mathbf{k}_i = 0$. Note that once the average is taken, the imaginary part of $\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)$ goes to zero.

It is convenient to define the reduced bispectrum $Q_{123} \equiv Q(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k_3})$ as,

$$Q_{123} \equiv \frac{B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{P(k_1)P(k_2)+P(k_1)P(k_3)+P(k_2)P(k_3)}, \tag{4.3}$$

which takes away most of the dependence on scale and cosmology. The reduced bispectrum is useful when comparing different models, since it has a weak dependence on cosmology and one can thus break degeneracies between cosmological parameters to isolate the effects of gravity. Hereafter, when we speak of the bispectrum we are always referring to the reduced bispectrum.

5 Results

In this paper, we present two ways of comparing the $f(R)$ and ΛCDM reduced bispectra we obtain from N-body simulations. The differences depend on whether the models are matched in their initial or final power spectra. In method A we compare the output bispectra from

N-body simulations with the same initial power spectra. Thus some of the difference in the bispectra can be attributed to the different amounts of final nonlinear power in the two sets. Method B tries to separate these contributions by generating modified initial power spectrum ΛCDM simulations whose power spectra at $z = 0$ match those of the $f(R)$ simulations.

For both methods we compute the bispectrum randomly drawing k-vectors from a specified bin, namely Δk and randomly orientating the triangle in space. We make the number of random triangles to depend on the number of fundamental triangle per bin, that scales as $k_1 k_2 k_3 \Delta k^3$ [14]. In this paper we always choose $\Delta k = 3k_{\rm min}$. For the equilateral case, at scales of $k \sim 0.65\ h/$Mpc we are generating $\sim 5 \times 10^8$ triangles. We have verified that increasing the number of triangles beyond this value does not have any effect on the measurement.

5.1 Method A (matched initial power spectrum)

Our first test of $f(R)$ vs ΛCDM bispectra utilize the same initial power spectrum. In our $f(R)$ models, modifications to gravity go to zero rapidly with redshift and the expansion history differs negligibly from ΛCDM. Thus models with the same initial power spectra as ΛCDM fit observations at high redshift, such as primary CMB anisotropy, equally well.

Since all N-body simulations start from the same initial power spectrum, the $f(R)$ modifications to gravity during the acceleration epoch lead to differences in the dark matter power spectra at low redshift that increase with $|f_{R0}|$ as was noted in figure 2 of [10]: these differences reach up to $\sim 50\%$ for $|f_{R0}| = 10^{-4}$ and $\sim 10\%$ for $|f_{R0}| = 10^{-6}$ for $k \simeq 0.5\, h/$Mpc with respect to the ΛCDM model.

Bispectra for matched initial power spectra but differing final power spectra is what is usually computed analytically [1, 2]: one predicts (using the modified Euler and continuity equations in perturbation theory) the reduced bispectra for different gravity models starting from a given initial δ_k field. One might expect that the reduced bispectra differences are independent of the power spectra, but this is only strictly true for equilateral configuration and only in the tree-level regime (for $k < 0.06\ h/$Mpc at $z = 0$). This is the main caveat of method A: the differences seen in the reduced bispectrum could be due to differences in the final matter power spectrum and not unique signatures of $f(R)$ gravity.

In figure 1 we show the dark matter reduced bispectra deviation between ΛCDM and $f(R)$ for $k_2 = 2k_1 = 0.4h/$Mpc (left panels) and for equilateral configurations (right panels) for $z = 0$ according to method A. The top panels correspond to $f(R)$ theories with $|f_{R0}| = 10^{-4}$; $|f_{R0}| = 10^{-5}$ for middle panels; and $|f_{R0}| = 10^{-6}$ for bottom panels. The blue points correspond to full $f(R)$ theories whereas the red points to non-chameleon ones. Deviations of $f(R)$ bispectra with $|f_{R0}| = 10^{-4}$ with respect to ΛCDM present a characteristic shape dependence, where the difference is maximal for $\theta_{12} \sim 0$ and π, and minimal for $\theta_{12} \sim 0.6\pi$, for both chameleon and non-chameleon and it increases as the scale is reduced. A similar trend is present for non-chameleon theories with $|f_{R0}| = 10^{-5}$. On the other hand, chameleon theories with $|f_{R0}| = 10^{-5}$ present a constant deviation from ΛCDM of $\sim 10\%$. For $|f_{R0}| = 10^{-6}$, both chameleon and non-chameleon present a constant deviation from ΛCDM of $\lesssim 5\%$ and is consistent with 0. The errors of figure 1 are suppressed compared to the individual cosmic variance errors because we are taking the ratio of N-body simulations with the same initial power spectrum and phases. Therefore, we can conclude that having the same initial conditions, the dark matter bispectra of ΛCDM and $f(R)$ theories is significantly different for $|f_{R0}| \gtrsim 10^{-5}$, especially for elongated triangles ($\theta_{12} \simeq 0, \pi$) and can reach deviations in the reduced bispectra up to $\sim 10\%$ for $|f_{R0}| = 10^{-5}$ and up to $\sim 12\%$ for $|f_{R0}| = 10^{-4}$. We

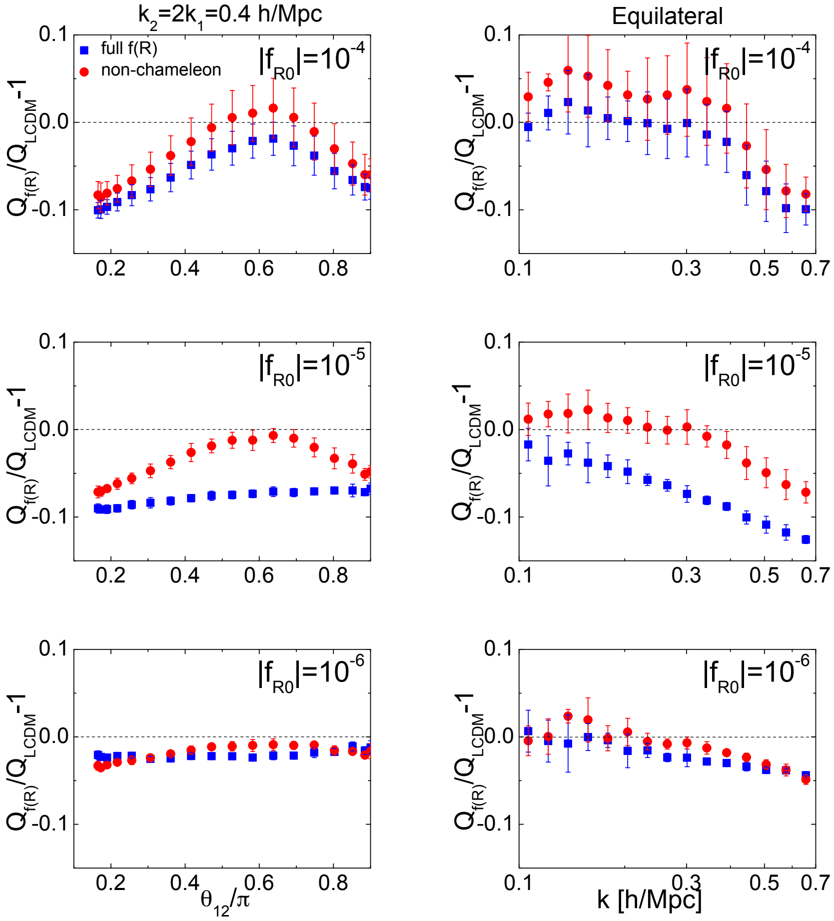

Figure 1. Relative dark matter reduced bispectrum deviations (following method A) between ΛCDM and $f(R)$ models for $k_2 = 2k_1 = 0.4\ h/\text{Mpc}$ (left panels) and equilateral configuration (right panels) at $z = 0$ as a function of the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$, namely θ_{12} (left panel) and as a function of k (right panel) for $|f_{R0}| = 10^{-4}$, 10^{-5}, 10^{-6} (top to bottom). Blue points (squares) correspond to chameleon simulations and red points (circles) to non-chameleon. Both ΛCDM and $f(R)$ bispectra have been computed from N-body simulations with the same initial conditions. As a consequence, the corresponding final ($z = 0$) power spectra of the compared models are different. Error bars are the 1-σ standard deviation of the ratio of Q values amongst the 6 independent runs. Because of that, the errors due to cosmic variance cancel out. Only $L = 400\,\text{Mpc}/h$ side-box runs are used.

see a similar dependence on triangle shape as shown in figure 5 of [1] (note that β as defined there is $1/\sqrt{6}$ for $f(R)$).

Although differences in the final dark matter power spectra between ΛCDM and $f(R)$ theories of the same initial power are large and potentially easier to test than those in the bispectra, it is possible that the galaxy power spectra for ΛCDM and $f(R)$ models could still be similar for some particular galaxy bias model [18]. Since the galaxy bias acts differently on the power spectrum and on the bispectrum, it would be very unlikely that the same galaxy bias could make $P_{\text{gal}}^{\Lambda\text{CDM}} = P_{\text{gal}}^{f(R)}$ and $Q_{\text{gal}}^{\Lambda\text{CDM}} = Q_{\text{gal}}^{f(R)}$ simultaneously.

Conversely, changes in the initial power spectra between the models might conspire to make an $f(R)$ model look like a ΛCDM model for the power spectrum at $z = 0$. These can be hidden from the CMB at high redshift if they only occur at high k. Because of that, in the next section we assess the differences between the ΛCDM and $f(R)$ dark matter reduced bispectra in the case where both models have the same final power spectrum.

5.2 Method B (matched final power spectrum)

The final power spectra of the $f(R)$ models deviate significantly from that of the ΛCDM model with the same initial conditions (see figure 2 of [10]). These deviations reach $\sim 50\%$ for $|f_{R0}| = 10^{-4}$; $\sim 10\%$ for $|f_{R0}| = 10^{-6}$; all at $k \simeq 1$ h/Mpc and at $z = 0$. That begs the question of whether bispectrum differences seen in Method A are driven by these final power spectrum differences or by uniquely gravitational modifications.

To address this question, we would like to adjust the initial conditions of the $f(R)$ simulations until the final power spectra match that of ΛCDM at $z = 0$. However, the $f(R)$ simulations are computationally very expensive (a factor of ∼20 increase over ordinary GR simulations). Instead we do the converse: we adjust the initial conditions of the ΛCDM model until its final power spectrum matches the $f(R)$ simulations. Matching ΛCDM to the $f(R)$ simulations still tests whether the remaining bispectra difference between the two models reflects gravitational modifications, independently of power spectrum differences.

5.2.1 Power spectra matching

In order to match final power spectra we need a means of quickly predicting the impact of adjusting initial conditions in ΛCDM. HaloFit [17] provides an approximate analytic mapping between the initial and final power spectra. Using HaloFit we can determine the desired initial conditions, run the matching ΛCDM simulations and compare the bispectra with those of the $f(R)$ simulations.[2]

We first test the accuracy of HaloFit in modeling the ΛCDM simulation results (see figure 2). In the HaloFit computation we use the same transfer function as employed in the ENZO code. We see that for the $L = 400$ Mpc/h runs the data-points with $k > k_N/2 \simeq 0.50$ h/Mpc underestimate the value of the predicted power spectrum by HaloFit. Here, k_N is the Nyquist mode defined as $k_N = \pi N_p^{1/3}/(2L)$. We can in principle solve this limitation by using smaller boxes. For the $L = 256$ Mpc/h runs, $k_N/2 \simeq 0.79\,h$/Mpc and we see that up to this scale the simulation agrees with the theoretical prediction. However the errors increase considerably as we reduce the box-size. Error bars in figure 2 correspond to 1-σ standard deviation amongst 6 independent runs. Likewise at low k, the simulations carry large sampling errors even for the largest boxes. To evade these problems, we use HaloFit to model only relative differences between simulations of the same $L = 400$ Mpc/h size, resolution and initial phases as we shall now describe.

In order to match the excess small scale final power in the $f(R)$ model, we add an extra running of the spectral tilt parameter to the ΛCDM initial power spectrum. Specifically, we assume a 3 free-parameter initial power spectrum model:

$$P_i(k) = P_0 \left(\frac{k}{k_p}\right)^{n_0 + \frac{1}{2}\alpha \ln(k/k_p)}, \tag{5.1}$$

---

[2] An alternative (much faster) approach would be, instead of re-running N-body simulations and computing the evolved bispectrum from there, to use fitting formulae from the literature to predict the ΛCDM bispectrum from the matched power spectrum which add a small running of the tilt. We have attempted this route; unfortunately we have found that available fitting formulae are not sufficiently accurate for our purposes [5]; as we will see we need a relative accuracy of better than 4% on the reduced bispectrum which exceeds that of available fitting formulae in the mildly non-linear regime of interest.

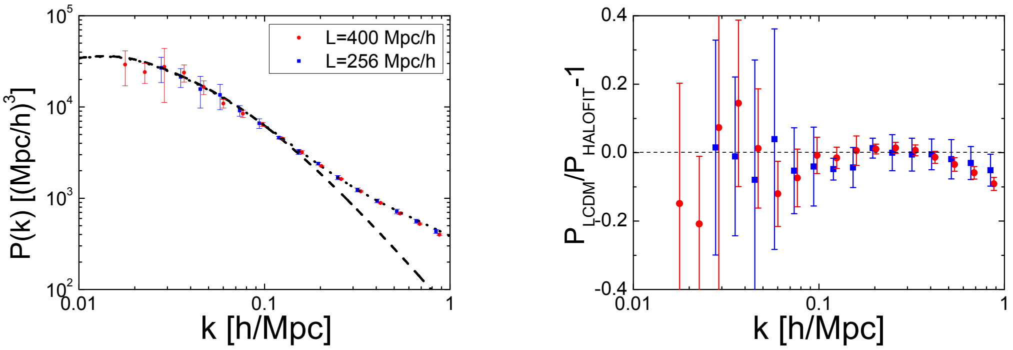

Figure 2. HaloFit accuracy for the power law ΛCDM simulations with box size $L = 256$ Mpc/h (blue squares) and $L = 400$ Mpc/h box size (red circles) *Left panel*: linear power spectrum (dashed line) and non-linear prediction from HaloFit (dotted line) plotted with the simulation results. *Right panel*: fractional difference of simulation results and the HaloFit prediction. In all cases the errors correspond to 1-σ standard deviation amongst 6 independent runs.

where P_0 is the amplitude of the power spectrum at $k_p = 0.1h\text{Mpc}^{-1}$ and $z = 0$ without the effect of the transfer function and α is the running of the tilt,

$$\frac{d\ln P_i}{d\ln k} = n_0 + \alpha \ln(k/k_p). \tag{5.2}$$

We therefore have 3 parameters $\mathbf{p} = \{P_0, n_0, \alpha\}$ which specify the initial conditions.

To find the best-fitting three parameters for a given model, we take the simulation results for the power spectrum ratio (following method A),

$$R_{\text{sim}}(k) \equiv \left\langle \frac{P^{\text{sim}}_{f(R)}(k)}{P^{\text{sim}}_{\Lambda\text{CDM}}(k)} \right\rangle. \tag{5.3}$$

Next we use the HaloFit prescription $P^{\text{HF}}(k; \mathbf{p}^{\text{match}})$ for the non-linear matter power spectrum at $z = 0$ to find the best parameter set $\mathbf{p}^{\text{match}}$, by minimizing the χ^2 given by

$$\chi^2 = \sum_j \frac{1}{\sigma^2_{R_{\text{sim}}}(k_j)} \left[\frac{P_{\text{HaloFit}}(k_j; \mathbf{p}^{\text{match}})}{P_{\text{HaloFit}}(k_j; \mathbf{p}^0)} - R_{\text{sim}}(k_j) \right]^2, \tag{5.4}$$

where the sum runs over bins in k. Here, $\mathbf{p}^0$ describes the initial power spectrum used for the $f(R)$ simulations (see first line in table 1). Finally we simulate a matched ΛCDM simulation with the same initial phases as the original but with a rescaling of the initial power

$$P_{\text{IC}}(k; \mathbf{p}^{\text{match}}) = P_{\text{IC,orig.}}(k) \frac{P_i(k; \mathbf{p}^{\text{match}})}{P_i(k; \mathbf{p}^0)}. \tag{5.5}$$

In order to avoid confusion, we designate these ΛCDM simulations as "matched"; whereas ΛCDM without this modifier denotes the standard, power law, initial conditions (the one used in section 5.1).

The advantage of this matching method is that we only model relative deviations with the HaloFit prescription. Thus the cosmic variance of the original simulations scale out as

	$\lvert f_{R0}\rvert$	$P_0/[10^3(\mathrm{Mpc}h^{-1})^3]$	n_0	α	σ_8
$\mathbf{p}^0$	0	6.31	0.958	0	0.824
cham.	10^{-4}	7.60	1.061	-0.00643	0.944
	10^{-5}	6.62	1.087	0.0427	0.878
	10^{-6}	6.34	0.985	0.00964	0.834
non cham.	10^{-4}	7.66	1.059	-0.00963	0.947
	10^{-5}	6.72	1.141	0.0620	0.898
	10^{-6}	6.37	1.012	0.0201	0.843

Table 1. Best-fit linear ΛCDM initial power spectrum parameters P_0, n_0, α (see text) that match the N-body power spectrum at $z = 0$ for each $f(R)$ simulation. These values have been computed minimizing the χ^2 of the N-body non-linear power spectrum and the non-linear HaloFit power spectrum for $k \leq k_N/2 \simeq 0.5\ h/\mathrm{Mpc}$. The same transfer function and the same cosmology is used in all cases. Also σ_8 is shown for clarity.

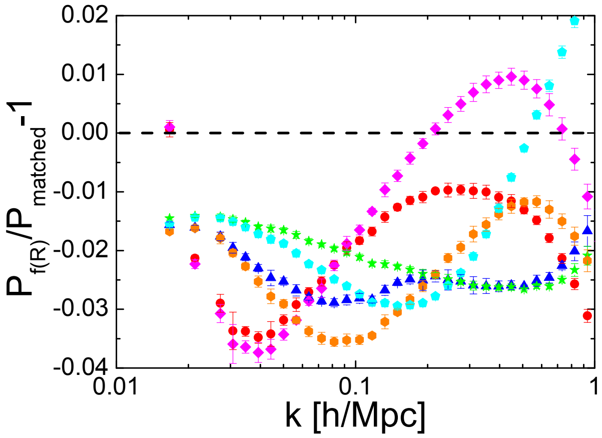

Figure 3. Relative power spectrum offset between $f(R)$ simulations and corresponding matched-ΛCDM simulations for different $\lvert f_{R0}\rvert$ values: for the chameleon simulations with $\lvert f_{R0}\rvert = 10^{-4}$ (red circles), 10^{-5} (blue triangles), 10^{-6} (green stars); and for the non-chameleon simulations with $\lvert f_{R0}\rvert = 10^{-4}$ (pink diamonds), 10^{-5} (orange hexagons), 10^{-6} (cyan pentagons). The matched simulations make use of initial power spectrum conditions shown in table 1 found by fitting relative deviations with HaloFit.

do absolute errors in HaloFit, initial condition generators, resolution, etc. In table 1 we show the best-fit values of the 3 initial power spectrum parameters. We have only used $R_{\mathrm{sim}}(k \leq 0.5\ h/\mathrm{Mpc})$ for the minimization.

In figure 3 we show $P_{f(R)}/P_{\mathrm{matched}} - 1$, where $P_{f(R)}$ is the power spectrum of the $f(R)$ simulations as before and P_{matched} is the power spectrum of the matched ΛCDM simulations. Figure 3 indicates that HaloFit is an excellent tool to predict relative differences in non-linear power spectra even for (some) non-standard ΛCDM models. We see that the differences between the matched ΛCDM and $f(R)$ power spectra are up to $\sim 4\%$ in the range $0.1h/\mathrm{Mpc} < k < 1h/\mathrm{Mpc}$, although for most of the scales and the cases are about 2-3%.

As an aside, we can also test the absolute accuracy of HaloFit's prediction for the power spectra of the matched models. Examples for different matched models are shown in figure 4.

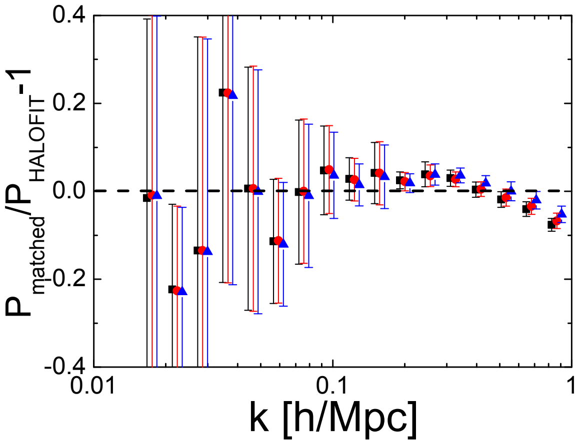

Figure 4. HaloFit accuracy for a representative set of matched ΛCDM models (matched to chameleon $|f_{R0}| = 10^{-4}$, black squares; 10^{-5} red circles; 10^{-5}, blue triangles). HaloFit is accurate within the errors for all $k < 0.5\text{Mpc}/h$ for these models which contain running of the tilt. Deviations from sample variance and resolution in the simulation seen here are largely absent in the relative matching technique shown in figure 3

.

HaloFit produces a good fit compared with the sample variance errors for all $k < 0.5\text{Mpc}/h$. As in the pure ΛCDM case, the sample variance at low k in the simulations is quite large. Deviations up to 10% for $k < 1.0$ Mpc$/h$ likewise appear due to the limited simulation resolution. Our modeling of relative effects eliminates these small differences.

In reality one does not observe at a single z but in a wide z-range. As mentioned above it is not possible to match the power spectrum at widely separated redshifts simultaneously and this feature can provide observational signatures independent from the bispectrum. We can quantify this further by estimating over what redshift interval the power spectrum matching is expected to hold. Changes in the $P_{f(R)}/P_{LCDM}$ were studied in detail by [10] and the excess evolves on the Hubble time scale. Therefore we generically expect that the matching evolves across a redshift interval of $\Delta z = 1$, i.e. no faster than any other aspect of the modeling.

5.2.2 Bispectrum

With the simulations of the matched ΛCDM models, we can now compare the bispectra for ΛCDM and $f(R)$ models whose final power spectra match to a few percent.

In figure 5 we show $Q_{f(R)}(k)/Q_{\text{matched}} - 1$ for $k_2 = 2k_1 = 0.4\,h/\text{Mpc}$ (left panel) and for equilateral triangle configuration (right panel), where $Q_{f(R)}(k)$ is the reduced bispectrum for $f(R)$ simulations, and Q_{matched} is the reduced bispectrum for the matched ΛCDM simulations. Red points show the ratio for non-chameleon simulations whereas blue points for chameleons ones. Top panels correspond to $|f_{R0}| = 10^{-4}$, middle panels to $|f_{R0}| = 10^{-5}$ and bottom panels to $|f_{R0}| = 10^{-6}$. In particular, we see that for the chameleon and non-chameleon cases with $|f_{R0}| = 10^{-4}$ and 10^{-6} the deviation is very close to 0 ($\lesssim 2\%$). For the $|f_{R0}| = 10^{-5}$ some differences appear: for the non-chameleon case there is an excess of $\sim 4\%$ and for the chameleon case there is a deficit of $\sim 4\%$ in $Q_{f(R)}$ respect to Q_{matched}, both within 5-6σ. The value $|f_{R0}| \sim 10^{-5}$ is special in that it marks the onset of the chameleon mechanism in the largest structures in the simulations. The chameleon effect may have a small but measurable impact on Q in this transition region where the chameleon effect is present for some but not all structures. Analogous transient enhancements appear in the mass function [7]. One should bear in mind though that this difference is of order the difference in the matched power spectra which varies between the full and no-chameleon cases.

Thus, for all values of f_{R0} deviations are below $\sim 4\%$. In particular we do not observe that squeezed triangles (those with $\theta_{12} \simeq 0, \pi$) present higher deviations between different gravity models as has been observed in method A (figure 1) and predicted from theoretical models that followed the same assumptions as adopted in method A [1, 2].

Finally, we found that it is better to analyze the deviation between reduced bispectra Q rather than between bispectra B. This is because the power spectrum dependence is partially canceled in the reduced bispectra. In spite of having run ΛCDM simulations to match the $f(R)$ power spectra, several percent differences between $f(R)$ and matched ΛCDM power spectra are still present (figure 3). These lead to higher deviations in B between the models (up to $\sim 8\%$ in some cases) than in Q. Thus, using Q instead of B is much more robust if we want to compare models with similar power spectra. Of course, one should keep in mind that not all the $P(k)$-dependence is cancelled when using Q, as evidenced by comparing with the results of method A. Strictly speaking, this is only true for equilateral configurations and up to tree-level in Eulerian perturbation theory.

Finally, one may want to make a connection between these results and some analytic model, namely perturbation theory (PT). Since at tree level in PT the reduced bispectrum is independent of the power spectrum (at least for equilateral configuration), the differences observed between figure 1 and 5 should be due to higher order corrections in LCDM. At 1-loop, corrections to the bispectrum can be found in e.g., [14, 16]. One can see that the leading terms depend on the linear and one-loop power spectrum and weakly on cosmology and gravity through the standard tree level bispectrum kernel (see [1] for a modification of this kernel for some $f(R)$ theories). Thus, a small modification of this formula could be expected due to $f(R)$ gravity. However, this interpretation should be considered more in a qualitative way than in a strictly quantitative way. In fact one should take into account that the precision of 1-loop PT for the bispectrum is not much better than the other (phenomenological) analytic formulae [5]. As we have already mentioned, nowadays there is no analytic model to predict the bispectrum at scales of interest here at $z = 0$ with an accuracy of few percent.

5.2.3 Discussion

If the remaining $\sim 4\%$ deviation for $|f_{R0}| = 10^{-5}$ reflects gravity and not the residual mismatch in power spectra, then it is in principle measurable with large-volume surveys. In this work, considering only the 6 runs of 400 Mpc/h box-size and provided that $h = 0.73$, the total volume is $6 \times (0.4\,\mathrm{Gpc}/h)^3 \simeq 1\mathrm{Gpc}^3$. We expect that future surveys will cover larger volumes: BOSS[3] $V \sim 5\,(\mathrm{Gpc}/h)^3$, DES[4] $V \sim 10\,(\mathrm{Gpc/h})^3$ or EUCLID $V \sim 100\,(\mathrm{Gpc}/h)^3$. As the 6 runs have different initial conditions we can use them to estimate the expected error on Q in the limit that it is dominated by cosmic variance. We have measured that the error in Q for our simulations at scales of $k \sim 0.3h/\mathrm{Mpc}$ is about 5%. We assume that the variance scales as the inverse of the number of modes, and thus the standard deviation approximately scales as $V^{1/2}$. Therefore, for a 10 Gpc^3 survey the error bars could, in principle, be as much as $\sqrt{10} \sim 3$ times smaller than our prediction. This implies that a survey with $> 10\,\mathrm{Gpc}^3$ volume (e.g., DES, EUCLID) would yield an error on $Q \sim 2\%$ at these scales. Since the

---

[3] Baryon Oscillation Spectroscopic Survey

[4] Dark Energy Survey

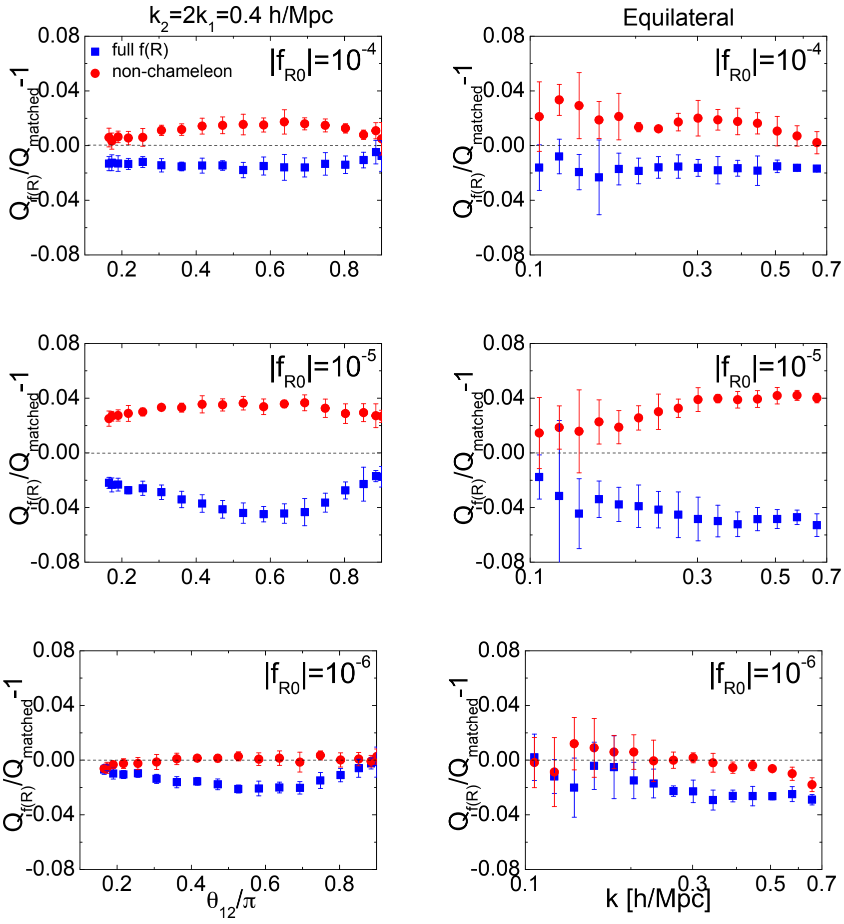

Figure 5. Relative reduced bispectrum deviation for matched final power spectra (method B) between $f(R)$ and ΛCDM simulations for $k_2 = 2k_1 = 0.4\,h/\mathrm{Mpc}$ as a function of θ_{12} (left panels), and equilateral configuration as a function of k (right panels), all at $z = 0$. Upper panels correspond to $|f_{R0}| = 10^{-4}$, middle panels to $|f_{R0}| = 10^{-5}$ and bottom panels to $|f_{R0}| = 10^{-6}$. Red points correspond to non-chameleon simulations, whereas blue points to full $f(R)$ simulations. The error-bars are the 1σ standard deviation amongst the ratio of 6 independent runs. Since we are taking the ratio between runs with the same initial phases, the cosmic variance errors are not present.

expected deviation may be of order 4%, having smaller errors would help us to confirm or discard possible deviation of the bispectrum due to modifications of gravity.

On the other hand, we have analyzed the dark matter bispectrum which is not directly observable. In practice, sources of error that we have neglected here may appear: i) galaxy-surveys provide a biased information about dark matter, ii) additional effects such as redshift distortions change the observed bispectrum (in fact we expect modified gravity to affect redshift distortions more strongly than the density field itself). Also as we go to higher z, we expect less deviations at a given scale. Conversely, the matched power spectra at $z = 0$ would become mismatched and provide other observable effects.

The results from figure 5 provide another important result. We have seen that two $f(R)$ theories of gravity with indistinguishable non-linear dark matter power spectrum, have very similar and possibly indistinguishable dark matter bispectra. This opens up the possibility of using these two statistics to break degeneracies in the galaxy bias in a way that is robust to the assumptions about the true underlying gravity model.

In fact the $f(R)$ effects on the power spectrum are at the 20-50% level. A modification of galaxy bias achieving similar effects would likely affect the reduced bispectrum at least at the 10% level (for example, a linear bias term affect the power spectrum $\propto b_1^2$ and the reduced bispectrum $\propto 1/b_1$), significantly larger than the $f(R)$ effects on the reduced bispectrum.

6 Conclusions

In this work we have analyzed the deviations in the reduced bispectrum produced by a modification of gravity, specifically the $f(R)$ class of models, both with and without the chameleon mechanism. In order to do that, we make use of a suite of $f(R)$ and ΛCDM simulations. We have proceeded in two different ways to analyze the bispectrum deviation from these simulations, methods A and B, which differ in whether the initial or final power spectra of the two cosmologies are set equal.

Method A compares the bispectrum output of $f(R)$ and ΛCDM N-body simulations with the same initial power spectrum. Figure 1 shows the bispectrum deviation obtained using this method. We observe a considerable deviation (up to $10-15\%$) in the reduced bispectrum between these $f(R)$ models and the ΛCDM one. Such differences in the bispectrum could be easily detected by surveys covering volumes $> 1\,\mathrm{Gpc}$ such as e.g., the on-going BOSS survey. Higher deviations are seen for higher values of $|f_{R0}|$ and for squeezed triangle configurations. In this method, both ΛCDM and $f(R)$ gravity runs start from the same initial δ_k values. Because of that, the different evolution of the gravity models naturally leads to different power spectra (as was observed in [10]) and also to different bispectra. This way of proceeding is equivalent to the theoretical works of [1, 2]. In order to explain discrepancies between the matter power spectrum in $f(R)$ and the observed galaxy power spectrum, one could invoke a scale-dependent galaxy bias. Since galaxy bias enters into the reduced galaxy bispectrum in a different way than in the power spectrum, bispectrum measurements can in principle close this loophole.

Alternatively, the large power spectrum differences can be eliminated by changing the shape of the initial power spectrum to instead match the final dark matter power spectrum at $z = 0$. This is at the base of method B. In this method, we compute the bispectrum deviation between a ΛCDM and a $f(R)$ model, both with the same final power spectra. Thus, we simulate a ΛCDM model with certain initial power spectrum parameters (summarized in table 1) that are adjusted to best match the $f(R)$ power spectrum at $z = 0$. From the simulations outputs we compute power spectra and bispectra. For the power spectra, residual differences are never higher than 4% in the range $0.1\,h/\mathrm{Mpc} < k < 1\,h/\mathrm{Mpc}$.

Likewise the differences in the reduced bispectrum are also smaller in the matched comparison. For the $|f_{R0}| = 10^{-4}$ and 10^{-6} cases, the Q deviation is consistent with 0 within 1σ. For $|f_{R0}| = 10^{-5}$ deviations in Q at most reach the 4% level with $5-6\sigma$ significance. These deviations are potentially a signature of the onset of the chameleon mechanism in the largest structures in the Universe. However given that this is the same order as the power spectrum difference it is unclear whether these differences indicate power-spectrum-independent modified gravity effects or that the two power spectra are not perfectly matched. In the former case, larger surveys like EUCLID will allow for a measurement of the bispectrum with enough precision to obtain a $> 6\sigma$ significance, even when exactly matching the power spectra.

On the other hand, the effect of deviations from GR gravity on the reduced bispectrum are weak compared to those on the power spectrum (at least for the cases considered here), opening up the possibility of breaking the galaxy-bias degeneracy. In fact the effect of galaxy bias is expected to be different in the power spectrum and in the bispectrum, which is why, in the context of GR gravity, the bispectrum is used to constrain galaxy bias. While the shape of the non-linear power spectrum seems to carry information about the underlying gravity model, one may always argue that a shape of the evolved power spectrum not compatible with the GR predictions could be due to biasing. For the cases we have considered here, the dependence of the reduced bispectrum on deviations from GR is weaker than the effects of bias modifications necessary to explain the deviations in the power spectrum. While we have only studied $f(R)$ models here, there is no apparent reason why this result should be specific to $f(R)$. Hence, if our findings were to remain qualitatively true for other gravity modifications, this would confirm the usefulness of employing the reduced bispectrum together with the power spectrum to constrain bias parameters.

Acknowledgments

We thank Christian Wagner for useful discussions and help with N-body simulations. HGM is supported by a CSIC JAE grant, and thanks the Kavli Institute for Cosmological Physics (KICP) at University of Chicago for hospitality. Part of this work stemmed from discussions at the Centro de Ciencias de Benasque Pedro Pascual. WH is supported by the KICP under NSF contract PHY-0114422, DOE contract DE-FG02-90ER-40560 and the Packard Foundation. LV and RJ are supported by MICINN grant AYA2008-03531. LV acknowledges support from grant FP7 ERC- IDEAS Phys.LSS 240117. FS is supported by the Gordon and Betty Moore Foundation at Caltech.

References

[1] F. Bernardeau and P. Brax, *Cosmological Large-scale Structures beyond Linear Theory in Modified Gravity*, *JCAP* **06** (2011) 019 [`arXiv:1102.1907`] [INSPIRE].

[2] A. Borisov and B. Jain, *Three-Point Correlations in f(R) Models of Gravity*, *Phys. Rev.* **D 79** (2009) 103506 [`arXiv:0812.0013`] [INSPIRE].

[3] D.J. Eisenstein and W. Hu, *Baryonic features in the matter transfer function*, *Astrophys. J.* **496** (1998) 605 [`astro-ph/9709112`] [INSPIRE].

[4] J.N. Fry, *Gravity, bias, and the galaxy three-point correlation function*, *Phys. Rev. Lett.* **73** (1994) 215.

[5] H. Gil-Marín et al., *An improved fitting formula for the dark matter bispectrum*, in preparation.

[6] W. Hu and I. Sawicki, *Models of f(R) Cosmic Acceleration that Evade Solar-System Tests*, *Phys. Rev.* **D 76** (2007) 064004 [`arXiv:0705.1158`] [INSPIRE].

[7] Y. Li and W. Hu, *Chameleon Halo Modeling in f(R) Gravity*, `arXiv:1107.5120` [INSPIRE].

[8] B.W. O'Shea, G. Bryan, J. Bordner, M.L. Norman, T. Abel, et al., *Introducing Enzo, an AMR cosmology application*, *IEEE Comput. Sci. Eng.* (2004) [`astro-ph/0403044`] [INSPIRE].

[9] H. Oyaizu, *Nonlinear evolution of f(R) cosmologies. I. Methodology*, *Phys. Rev.* **D 78** (2008) 123523.

[10] H. Oyaizu, M. Lima and W. Hu, *Nonlinear evolution of f(R) cosmologies. 2. Power spectrum*, *Phys. Rev.* **D 78** (2008) 123524 [`arXiv:0807.2462`] [INSPIRE].

[11] SUPERNOVA COSMOLOGY PROJECT collaboration, S. Perlmutter et al., *Measurements of Omega and Lambda from 42 high redshift supernovae*, *Astrophys. J.* **517** (1999) 565 [`astro-ph/9812133`] [INSPIRE].

[12] Supernova Search Team collaboration, A.G. Riess et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, Astron. J. **116** (1998) 1009 [`astro-ph/9805201`] [INSPIRE].

[13] F. Schmidt, M.V. Lima, H. Oyaizu and W. Hu, *Non-linear Evolution of f(R) Cosmologies III: Halo Statistics*, Phys. Rev. **D 79** (2009) 083518 [`arXiv:0812.0545`] [INSPIRE].

[14] R. Scoccimarro, *Cosmological perturbations: Entering the nonlinear regime*, Astrophys. J. **487** (1997) 1 [`astro-ph/9612207`] [INSPIRE].

[15] C. Sealfon, L. Verde and R. Jimenez, *Limits on deviations from the inverse-square law on megaparsec scales*, Phys. Rev. **D 71** (2005) 083004 [`astro-ph/0404111`] [INSPIRE].

[16] E. Sefusatti, *1-loop Perturbative Corrections to the Matter and Galaxy Bispectrum with non-Gaussian Initial Conditions*, Phys. Rev. **D 80** (2009) 123002 [`arXiv:0905.0717`] [INSPIRE].

[17] The Virgo Consortium collaboration, R. Smith et al., *Stable clustering, the halo model and nonlinear cosmological power spectra*, Mon. Not. Roy. Astron. Soc. **341** (2003) 1311 [`astro-ph/0207664`] [INSPIRE].

[18] Y.-S. Song, H. Peiris and W. Hu, *Cosmological Constraints on f(R) Acceleration Models*, Phys. Rev. **D 76** (2007) 063517 [`arXiv:0706.2399`] [INSPIRE].

[19] T.P. Sotiriou and V. Faraoni, *f(R) Theories Of Gravity*, Rev. Mod. Phys. **82** (2010) 451 [`arXiv:0805.1726`] [INSPIRE].

[20] L. Verde, A.F. Heavens, W.J. Percival, S. Matarrese, C.M. Baugh, et al., *The 2dF Galaxy Redshift Survey: The Bias of galaxies and the density of the Universe*, Mon. Not. Roy. Astron. Soc. **335** (2002) 432 [`astro-ph/0112161`] [INSPIRE].

[21] G.-B. Zhao, B. Li and K. Koyama, *N-body Simulations for f(R) Gravity using a Self-adaptive Particle-Mesh Code*, Phys. Rev. **D 83** (2011) 044007 [`arXiv:1011.1257`] [INSPIRE].

CHAPTER 5

An improved fitting formula for the dark matter bispectrum

This Chapter consists of the paper Gil-Marín et al. (2012), where an improved fitting formula for the dark matter bispectrum is presented. We use a set of ΛCDM simulations to calibrate the fitting parameters of a phenomenological approach similar to the one by Scoccimarro and Couchman (2001). Unlike previous works, this new proposed fit describes well the baryonic acoustic oscillation features and provides a reliable prediction for the non-linear dark matter bispectrum for ΛCDM models.

This paper was done in collaboration with Christian Wagner, Frantzeska Fragkoudi, Raúl Jiménez and Licia Verde from the Institut de Ciències del Cosmos at the Universitat the Barcelona, Spain and and was published in Journal of Cosmology and Astroparticle Physics (JCAP) in February 2012.

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

An improved fitting formula for the dark matter bispectrum

Héctor Gil-Marín,[a,b] **Christian Wagner,**[b] **Frantzeska Fragkoudi,**[b]
Raul Jimenez[c,b] **and Licia Verde**[c,b]

[a]Institut de Ciències de l'Espai (ICE), Facultat de Ciències,
Campus UAB (IEEC-CSIC), Bellaterra E-08193, Spain
[b]Institut de Ciències del Cosmos (ICC), Universitat de Barcelona (IEEC-UB),
Martí i Franqués 1, E-08028, Spain
[c]ICREA Institució Catalana de Recerca i Estudis Avançats.,
Passeig Lluís Companys 23, E-08010 Barcelona, Spain

E-mail: gil@ieec.uab.es, cwagner@icc.ub.edu, francesca.fragkoudi@gmail.com,
raul.jimenez@icc.ub.edu, liciaverde@icc.ub.edu

Received November 18, 2011
Revised February 9, 2012
Accepted February 13, 2012
Published February 29, 2012

Abstract. In this paper we present an improved fitting formula for the dark matter bispectrum motivated by the previous phenomenological approach of ref. [1]. We use a set of LCDM simulations to calibrate the fitting parameters in the k-range of $0.03\,h/\mathrm{Mpc} \leq k \leq 0.4\,h/\mathrm{Mpc}$ and in the redshift range of $0 \leq z \leq 1.5$. This new proposed fit describes well the BAO-features although it was not designed to. The deviation between the simulations output and our analytic prediction is typically less than 5% and in the worst case is never above 10%. We envision that this new analytic fitting formula will be very useful in providing reliable predictions for the non-linear dark matter bispectrum for LCDM models.

Keywords: power spectrum, cosmological simulations

ArXiv ePrint: 1111.4477

JCAP02(2012)047

doi:10.1088/1475-7516/2012/02/047

Contents

1 Introduction

The dark matter and galaxy power spectrum have been widely used to study the growth of structure, to constrain cosmological parameters and galaxy bias models. These tools have proved very successful and have contributed to crystallize the current LCDM model e.g., [2] and refs therein. With ongoing and forthcoming galaxy surveys, like BOSS[1] and EUCLID,[2] the signal-to-noise of the data will increase and the uncertainties around this model will be reduced. Higher precision data will allow the use of not only the two-point correlation function, but also of higher-order statistics, in order to constrain and improve our theories and models. The bispectrum (the three-point correlation function in Fourier space) is naturally the next statistic to consider [3, 4]. Using both the power spectrum and bispectrum we can improve our knowledge of the growth of structure and galaxy biasing [5–13], constrain possible departures from Gaussianity in the initial conditions of the matter density field [14–18] as well as constrain departures from GR e.g., [19, 20].

From a theoretical point of view, perturbation theory and subsequent improvements such as renormalized perturbation theory [21], resummed perturbation theory or time-RG flow [22] is a physically well-motivated approach to study these statistical moments. Tree-level perturbation theory has demonstrated to describe well the behavior of the power spectrum and bispectrum at large scales. However, non-trivial computations are needed to obtain predictions at non-linear scales: the one-loop correction and beyond, for the power spectrum and bispectrum. For the power spectrum, however, other phenomenological approaches have been demonstrated to work better for a wide range of redshifts and different cosmologies

---

[1] Baryon Oscillation Spectroscopic Survey.

[2] R. Laurejis et al, arXiv:1110.3193.

e.g., [23–25]. For the bispectrum there are also simple phenomenological models that predict its behavior at non-linear scales [1], but they fail to accurately reproduce the BAO-features [26] and are only precise at the 20%–30% level. Therefore better analytical models are needed to describe the bispectrum at these non-linear scales.

In this paper we improve the phenomenological description presented by [1] (hereafter SC) more than 10 years ago. Using a set of modern simulations we fit the free parameters of our proposed analytic formula. Thus, we obtain an improved description for the bispectrum in the LCDM model scenario (including baryonic acoustic oscillations) in a range of $0.03\,h/\mathrm{Mpc} \leq k \leq 0.4\,h/\mathrm{Mpc}$ and for different redshifts, $0 \leq z \leq 1.5$.

This paper is organized as follows: in §2 we begin with a description of the density field statistics and different analytic approaches to the dark matter bispectrum. In §3 we describe the simulations we use to fit the parameters. In §4 we present our results, compare them with previous fitting formulae and with 1-loop corrections and discuss the differences. We finally conclude in §5. In appendix A, we give details of how the bispectrum and its errors are computed from simulations. In appendix B we test how our formula works for other non-standard LCDM models. In appendix C we present a short description of 1-loop correction in Eulerian perturbation theory.

2 Theory

2.1 Power spectrum & bispectrum

The power spectrum $P(k)$, the Fourier transform of the two-point correlation function, is one of the simplest statistics of interest one can extract from the dark matter overdensity field $\delta(\mathbf{k})$,

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}')\rangle \equiv (2\pi)^3 \delta^D(\mathbf{k}+\mathbf{k}')P(k)\,, \tag{2.1}$$

where δ^D denotes the Dirac delta function and $\langle \ldots \rangle$ the ensemble average over different realizations of the Universe. Under the assumption of an isotropic Universe, the power spectrum does not depend on the direction of the **k**-vector. Since we only have one observable Universe, the average $\langle \ldots \rangle$ is taken over all different directions for each **k**-vector. Under the hypothesis of ergodicity both averages will yield the same result.

The second statistic of interest is the bispectrum B, defined by,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle \equiv (2\pi)^3 \delta^D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3). \tag{2.2}$$

The Dirac delta function ensures that the bispectrum is defined only for **k**-vector configurations that form closed triangles: $\sum_i \mathbf{k}_i = 0$. Note that $\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)$ is in general a complex number, however once the average is taken, the imaginary part goes to zero.

It is convenient to define the reduced bispectrum $Q_{123} \equiv Q(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k_3})$ as,

$$Q_{123} \equiv \frac{B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{P(k_1)P(k_2)+P(k_1)P(k_3)+P(k_2)P(k_3)} \tag{2.3}$$

which takes away part of the dependence on scale and cosmology.[3] The reduced bispectrum is useful when comparing different models, since it only has a weak dependence on cosmology and one can thus break degeneracies between cosmological parameters in order to isolate the effects of gravity.

---

[3] For equilateral configuration and up to tree level, Q does not depend on cosmology or scale.

The bispectrum for Gaussian initial conditions is zero and remains zero in linear theory, i.e. as long as the k-modes evolve independently.[4] However, when non-linearities start to play an important role, mode coupling is no longer negligible and the bispectrum becomes non-zero. Thus, by measuring the bispectrum one can extract information about how non-linear processes influence the evolution of dark matter clustering.

2.2 Analytic approaches in the literature

In order to understand the observational data, we need accurate theoretical predictions for $B(k_1, k_2, k_3)$. A physically well-motivated analytic theory for doing this, is perturbation theory (PT hereafter) (see [27] for a review) or subsequent improvement such as renormalized PT, resummed PT etc.

In an Einstein de-Sitter Universe (hereafter EdS Universe) and at second order (tree-level) in Eulerian perturbation theory, the bispectrum is given by [28],

$$B_{123} = 2F_2^s(\mathbf{k}_1, \mathbf{k}_2)P_1^L P_2^L + \text{cyc. perm.}, \tag{2.4}$$

where $B_{123} = B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$, $P_i^L = P^L(k_i)$ is the linear power spectrum, and the symmetrized two-point kernel F_2^s is given by

$$F_2^s(\mathbf{k}_i, \mathbf{k}_j) = \frac{5}{7} + \frac{1}{2}\cos(\theta_{ij})\left(\frac{k_i}{k_j} + \frac{k_j}{k_i}\right) + \frac{2}{7}\cos^2(\theta_{ij}), \tag{2.5}$$

where θ_{ij} is the angle between the vectors $\mathbf{k}_i$ and $\mathbf{k}_j$. This formula is the second order perturbation theory contribution to the bispectrum which is the leading order contribution. On quasi-linear scales, this expression is a very good prediction but fails in the moderate non-linear regime. The dependence on cosmology of the two-point kernel F_2^s is very weak and hence the cosmology dependence of the bispectrum is almost completely contained in P_i^L. Because of this, in this work, we use the kernel of eq. (2.5) even though we are dealing with the LCDM model.

One can improve the tree-level PT prediction by going one step further and including one-loop corrections. However, at this point the computation of the bispectrum becomes cumbersome. For an initially Gaussian δ-field this yields four additional terms to the tree-level contribution (see appendix C for details).

An alternative way of reaching these non-linear scales, without using the one-loop correction, and to even push beyond the one-loop regime of validity, is with phenomenologically motivated models. Phenomenological formulae can give simpler expressions in the non-linear regime and accurate predictions for the bispectrum. However, their physical motivation is limited and they usually have free parameters that need to be calibrated using N-body simulations.

SC proposed a fitting formula based on the structure of the formula of eq. (2.4). It consists in replacing the linear power spectrum by the non-linear one in eq. (2.4) and the EdS two-point symmetrized kernel by

$$\begin{aligned} F_2^{\text{eff}}(\mathbf{k}_i, \mathbf{k}_j) = {} & \frac{5}{7}a(n_i, k_i)a(n_j, k_j) \\ & + \frac{1}{2}\cos(\theta_{ij})\left(\frac{k_i}{k_j} + \frac{k_j}{k_i}\right)b(n_i, k_i)b(n_j, k_j) + \frac{2}{7}\cos^2(\theta_{ij})c(n_i, k_i)c(n_j, k_j), \end{aligned} \tag{2.6}$$

---

[4] Wick theorem states that the n-point correlation function of a Gaussian field is always zero when n is an odd number.

where the functions $a(n,k)$, $b(n,k)$ and $c(n,k)$ are chosen to interpolate between the tree-level results and the hyper-extended perturbation theory regime (HEPT) [29],

$$\begin{aligned}
a(n,k) &= \frac{1+\sigma_8^{a_6}(z)[0.7Q_3(n)]^{1/2}(qa_1)^{n+a_2}}{1+(qa_1)^{n+a_2}}, \\
b(n,k) &= \frac{1+0.2a_3(n+3)q^{n+3}}{1+q^{n+3.5}}, \\
c(n,k) &= \frac{1+4.5a_4/[1.5+(n+3)^4](qa_5)^{n+3}}{1+(qa_5)^{n+3.5}}.
\end{aligned} \tag{2.7}$$

Here n is the slope of the linear power spectrum at k,

$$n \equiv \frac{d\log P^L(k)}{d\log k} \tag{2.8}$$

and $q \equiv k/k_{\rm nl}$, where $k_{\rm nl}$ is the scale where non-linearities start to be important and is defined as,

$$\frac{k_{\rm nl}^3 P^L(k_{\rm nl})}{2\pi^2} \equiv 1; \tag{2.9}$$

a_i are free parameters that must be fitted using data from simulations. In particular, SC propose the values,

$$a_1 = 0.25, \quad a_2 = 3.5, \quad a_3 = 2, \quad a_4 = 1, \quad a_5 = 2, \quad a_6 = -0.2\,.$$

The function $Q_3(n)$ is given by

$$Q_3(n) = \frac{4-2^n}{1+2^{n+1}}\,. \tag{2.10}$$

With all these changes, the SC approach reads,

$$B_{123} = 2F_2^{\rm eff}(\mathbf{k}_1,\mathbf{k}_2)P_1P_2 + \text{cyc. perm.}, \tag{2.11}$$

where P_i is the non-linear power spectrum at k_i. On large scales, where the functions a, b and $c \to 1$ we recover the tree-level PT formula for the bispectrum. On the other hand, on small scales $a^2 \to (7/10)Q_3$ and b and $c \to 0$ and we obtain $Q_{123} \to Q_3(n)$, which is the prediction of HEPT.

Another approach based on phenomenological formulae, is the one presented by [26]. The main idea is to rescale the linear formula of the bispectrum, by using some scale transformation in k. This way, the tree-level formula can easily be extended up to non-linear scales using the ansatz $\tilde{k}_i = \left[1+\Delta^2_{NL}(k_i)\right]^{-1/3} k_i$; where $\Delta^2_{NL}(k) = P(k)k^3/(2\pi^2)$.

This approach has by definition the drawback that it does not preserve the BAO-features of the bispectrum. In particular, the rescaling of k produces a spurious rescaling of the peaks and troughs of the BAO wiggles that do not match with the data, producing higher deviations than the SC approach. Because of that, we do not consider this approach in this paper.

2.3 Our analytic formula

Our approach in this paper is inspired by the SC approach. It consists of not only refitting the a_i parameters from eq. (2.7) but of also modifying their expression to make it more suitable for current precision N-body data and consider the redshift range of $0 \le z \le 1.5$. In order to do

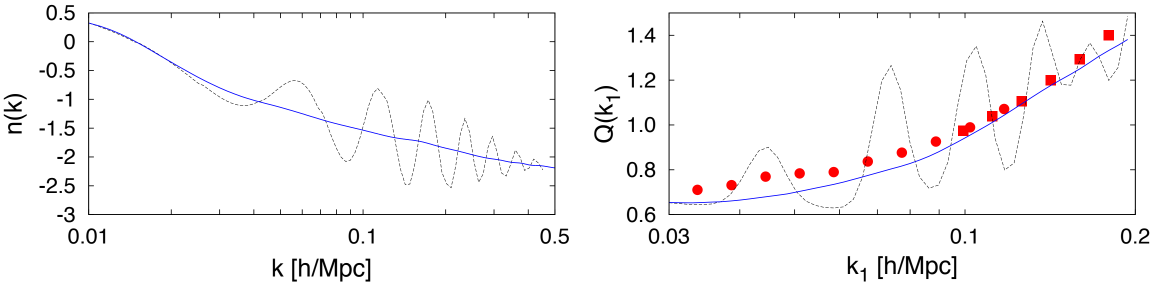

Figure 1. *Left panel*: The slope $n(k)$ (eq. (2.8)) from the linear power spectrum without smoothing (black dashed line) and with a spline smoothing (blue solid line). *Right panel*: $Q(k_1)$ for $k_2/k_1 = 2$ and $\theta_{12} = 0.6\pi$. Red circles are data from simulations A and red squares from simulations B (see table 1 for details on the simulations). Black dashed line is SC prediction without any spline in $n(k)$ and blue solid line with the spline in $n(k)$.

that, we use simulations with more particles, larger box sizes, and more realizations (and thus higher precision, better statistics and better error-control) with respect to previous works; we also consider snapshots at different redshifts. In order to improve the fitting precision, we also add 3 more parameters to the original model. The modified functions $\tilde{a}(n,k)$, $\tilde{b}(n,k)$, $\tilde{c}(n,k)$ then read,

$$\begin{aligned}
\tilde{a}(n,k) &= \frac{1+\sigma_8^{a_6}(z)[0.7Q_3(n)]^{1/2}(qa_1)^{n+a_2}}{1+(qa_1)^{n+a_2}}, \\
\tilde{b}(n,k) &= \frac{1+0.2a_3(n+3)(qa_7)^{n+3+a_8}}{1+(qa_7)^{n+3.5+a_8}}, \\
\tilde{c}(n,k) &= \frac{1+4.5a_4/[1.5+(n+3)^4](qa_5)^{n+3+a_9}}{1+(qa_5)^{n+3.5+a_9}}.
\end{aligned} \tag{2.12}$$

Note that one recovers the original SC formulae in the limit of $a_7 \to 1$ and a_8, $a_9 \to 0$.

The original SC formula was not designed to reproduce the BAO features. Applying this formula to a power spectrum with BAOs produces unphysical oscillations. These oscillations are much larger than those observed in simulations (see black dashed line in the right panel of figure 1). These oscillations are caused by the oscillatory behavior of the slope parameter n. One solution to this problem is to "dewiggle" the linear power spectrum [30]. However here we want to preserve the BAO oscillations. We propose to smooth the oscillatory behavior of the parameter n by means of splines, as is shown in the blue solid line of the left panel of figure 1. This provides an improved fit to the BAO-features, as it is shown by the blue solid line in the right panel of figure 1. In order to smooth out n we calculate its spline by taking a number of points $n(k)$, where the points are chosen to be in the middle of the amplitude of each wiggle, such that when the points are connected a smooth line would pass through them. These points are used in the spline routine, and their second order derivatives are calculated for each point k. This output is then fed into the spline routine, which returns a smoothed value of n for each value of k.

Our method consists in using this smoothed n and refit all the free a_i parameters from eq. (2.12) using the reduced bispectrum data from N-body simulations. In particular, we use the following triangle configurations at different redshifts: $\theta_{12}/\pi = 0.1, 0.2, \ldots, 0.9$, $k_2/k_1 = 1.0, 1.5, 2.0, 2.5$ and $z = 0, 0.5, 1, 1.5$.

	A	B
L_b [Mpc/h]	2400	1875
N_p	768^3	1024^3
N_r	40	3
$k_N/4$ [h/Mpc]	0.25	0.43
softening ϵ [kpc/h]	90	40
PM grid	2048^3	2048^3
ErrTolForceAcc α	0.005	0.005
initial scale factor a_i	0.05	0.02
maximum $\Delta \log a$	0.025	0.025
ErrTolIntAccuracy η	0.025	0.025
# time steps	~ 1300	~ 2500

Table 1. Simulations details for simulations A and B. L_b is the box size, N_p is the number of particles, N_r is the number of independent realizations. A quarter of the Nyquist frequency, $k_N/4$, is the upper threshold to which we trust the simulation results at the percent level. The force resolution is specified by the softening parameter ϵ and the Particle Mesh (PM) grid size. The short-range force accuracy is determined by α through the cell-opening criterion $Ml^2 > \alpha|\mathbf{a}_{\rm old}|r^4$, where M is the mass inside the cell, l its side length, $\mathbf{a}_{\rm old}$ the total acceleration of the particle in the previous time step, and r the distance between the particle and the cell. The remaining parameters set the time stepping in Gadget-2: the maximum global time step in the logarithm of the scale factor, $\max(\Delta \log a)$, and the parameter η in the individual time step criterion $\Delta a = \sqrt{2\eta\epsilon/|\mathbf{a}|}$, where $\mathbf{a}$ is the acceleration of the individual particle.

3 Simulations

The simulations in this paper consist of two different sets, namely A and B. Each simulation is characterized by the box size, L_b, the number of particles, N_p, and the number of independent runs, N_r. Details about the two simulations are given in table 1. As a rule of thumb, a maximum threshold in k for trusting the simulation data is set by a quarter of the Nyquist frequency, defined as $k_N/4 = \pi N_p^{1/3}/(4L_b)$. At this scale it has been observed that the power spectrum starts to deviate at the 1%-level with respect to higher resolution simulations [31]. We confirmed this result using our two sets of simulations. For all the plots and results shown in this paper this limit in k is never exceeded.

Both A and B simulations consist in a flat LCDM cosmology with cosmological parameters consistent with observational data. The cosmology used is $\Omega_\Lambda = 0.73$, $\Omega_m = 0.27$ $h = 0.7$, $\Omega_b h^2 = 0.023$, $n_s = 0.95$ and $\sigma_8(z = 0) = 0.7913$. The initial conditions were generated at $z = 19$ and $z = 49$ for simulations A and B respectively, by displacing the particles according to the second-order Lagrangian PT from their initial grid points. The initial power spectrum of the density fluctuations was computed by CAMB [32]. Taking only the gravitational interaction into account, the simulation was performed with GADGET-2 code [33].

As we estimate the error of the bispectrum from its dispersion among different realizations (see eq. (A.5) in appendix A for details), and given that we only have 3 simulations of

$a_1 = 0.484$	$a_2 = 3.740$	$a_3 = -0.849$
$a_4 = 0.392$	$a_5 = 1.013$	$a_6 = -0.575$
$a_7 = 0.128$	$a_8 = -0.722$	$a_9 = -0.926$

Table 2. Best-fit parameters (according to eq. (2.12)) derived by combining data from simulations A and B, using different triangle configurations $\theta_{12}/\pi = 0.1, 0.2, \ldots, 0.9$ and $k_2/k_1 = 1.0, 1.5, 2.0, 2.5$ and at different redshifts $z = 0, 0.5, 1.0, 1.5$.

type B, we divide each of these 3 boxes into 8 sub-boxes. Each of these 24 sub-boxes is then treated as if it were an independent realization with smaller box size, $L'_b = 937.5\,h/\mathrm{Mpc}$, where each of these sub-boxes contains about 512^3 particles. The measurements of the bispectrum from sub-boxes suffer from two issues: a) the measurements are not completely independent and more importantly b) the sub-boxes are affected by modes larger than sub-box size. As a consequence of this, a new source of non-Gaussian errors arises for the power spectrum and bispectrum estimation, called beat-coupling effect [34–36]. However, by using the mean density measured in each sub-box instead of the global mean density for the normalization of the density contrast, $\delta \equiv \rho/\bar{\rho} - 1$, this effect gets strongly suppressed [37]. Hence, we expect that on overlapping scales the bispectrum errors estimated from simulation B to be slightly larger than those from A. This is shown in appendix A.

In order to obtain the dark matter field from particles we discretize each box of simulation A and each sub-box of simulation B using 512^3 grid cells. Thus the size of the grid cells is 4.68 Mpc/h in A and 1.83 Mpc/h in B. We assign the particles to the cells using the count-in-cells prescription.

More details about the estimation of the bispectrum from simulations and the error bars computation are given respectively in eq. (A.4) and A.5 in appendix A.

4 Results

In order to find the best-fit parameters from eq. (2.12), namely a_i, we minimize

$$\chi^2 \propto \sum_i \left[\left(Q_i^{th} - Q_i^{sims} \right) / \sigma_{Q_i}^{sims} \right]^2$$

using a set of triangle configurations: $k_2/k_1 = 1.0, 1.5, 2.0, 2.5$ and $\theta_{12}/\pi = 0.1, 0.2, \ldots, 0.9$, at different redshifts: $z = 0, 0.5, 1.0, 1.5$.[5] The algorithm used for the minimization is *amoeba* [38]. In our analysis we neglect that the errors of the data points are correlated. However, since our errors are small (typically less than 5%) we expect that error correlations do not play an important role for this method: the dominating source of the error of our fitting formula given in eq. (2.12) comes from the imperfection of the functional form of the fitting formula and not from the uncertainties in the simulation data. We have checked that the result converges from different starting points. The resulting best-fit values are shown in table 2.

We have also checked that this is a very good fit not only for the reduced bispectrum Q but also for the bispectrum B.

In figure 2 and 3 we show the results of our fit and also the predictions of two other models for different triangles configurations for $z = 0$ (figure 2) and for $z = 1$ (figure 3).

---

[5] For simulation A the z used are $z = 0, 0.5, 1.0, 1.5$ whereas for simulation B $z = 0, 0.42, 1.0, 1.5$.

JCAP02(2012)047

These two models are 1-loop Eulerian PT (see appendix C) and the SC method (eq. (2.11)) + smoothed-n. In each plot we show the reduced bispectrum Q vs. k_1 for: N-body data (black circles for simulations A and black squares for simulations B), 1-loop correction (red solid line corresponding to data of simulations A and red dashed line to data of simulations B), SC formula (green solid line for A and green dashed line for B) and our model (blue solid and dashed line for A and B respectively). In the bottom part of each panel we show the deviation of these models with respect to the N-body data: red symbols depict the deviation of the 1-loop prediction with respect to the data, green symbols are the deviation of the SC formula and blue symbols are the deviation of our model. Circles are the deviation with respect to simulation A and squares with respect to simulation B. The error bars show the error of the bispectrum measured from the simulations (see appendix A for details). Each panel shows a different triangle configuration: from left to right $k_2/k_1 = 1.0$, 1.5, 2.0 and from top to bottom $\theta_{12}/\pi = 0.2, 0.4, 0.6, 0.8$. In order to avoid sample variance effects, we only use data points with $k_i > 0.03$ h/Mpc for simulation A and $k_i > 0.09$ h/Mpc for simulation B.

For some triangle configurations, we observe a mismatch between the data points in the overlapping region of simulations A and B . This is due to the fact that the data points do not exactly correspond to the same triangle configurations. The central value for k_i is the same for both simulations in each panel. However, since A and B have different fundamental frequencies $k_f \equiv 2\pi/L_b$ and we take the bin width to be $\Delta k = 3k_f$, the k-space over which we average the bispectrum for each bin is different (see appendix A for further explanation). Hence, when one computes the effective $\tilde{k}_i$ using eq. (A.8) one obtains that, especially for elongated triangle configurations, the simulations do not represent the same triangle configuration. Since the theoretical predictions (both 1-loop, SC and our model) are computed from $\tilde{k}_i$ there is also a mismatching between solid and dashed lines for the same reason. This effect is also noted in [30].

We have checked that this mismatch is not due to the fact that the two simulations have different resolution but to the fact that different triangle configurations are sampled. When the triangle configurations and scales coincide, we do not observe any mismatch. Note also that in the cases where a mismatch appears –due to different configurations being sampled–, it is also present (and quantitatively similar) in the the 1-loop theory prediction (see for example top left panel of figure 3).

At $z = 0$, the deviation of our model from the data is typically less than 5% and always less than 10% for $k < 0.4$ h/Mpc, whereas for SC the deviation reaches values of up to 20% and 1-loop clearly breaks down for $k \gtrsim 0.1\,h$/Mpc. We also observe that for triangles close to equilateral, both the SC approach and our work present maximum differences to the N-body data. This might have to do with the fact that for equilateral (or close to equilateral) triangles the 3 sides enter the non-linear regime at the same time, and thus, non-linearities play a stronger role than for other triangle configurations, where each side enters the non-linear regime at different redshifts. Because of this, for other configurations the differences are smaller and remain within 5% deviation.

For $z = 1$ all models work better because non-linearities do not play such an important role. However, even in this regime, our model works better than the other two. Again, for triangle configurations close to equilateral, our model reaches its maximum deviation of about 10%. At $z = 1$, all other triangle configurations typically have errors within 5%.

As a cross-check, in appendix B we compare our model with another set of simulations of non-standard LCDM model. In particular for equilateral configurations, our formula reaches

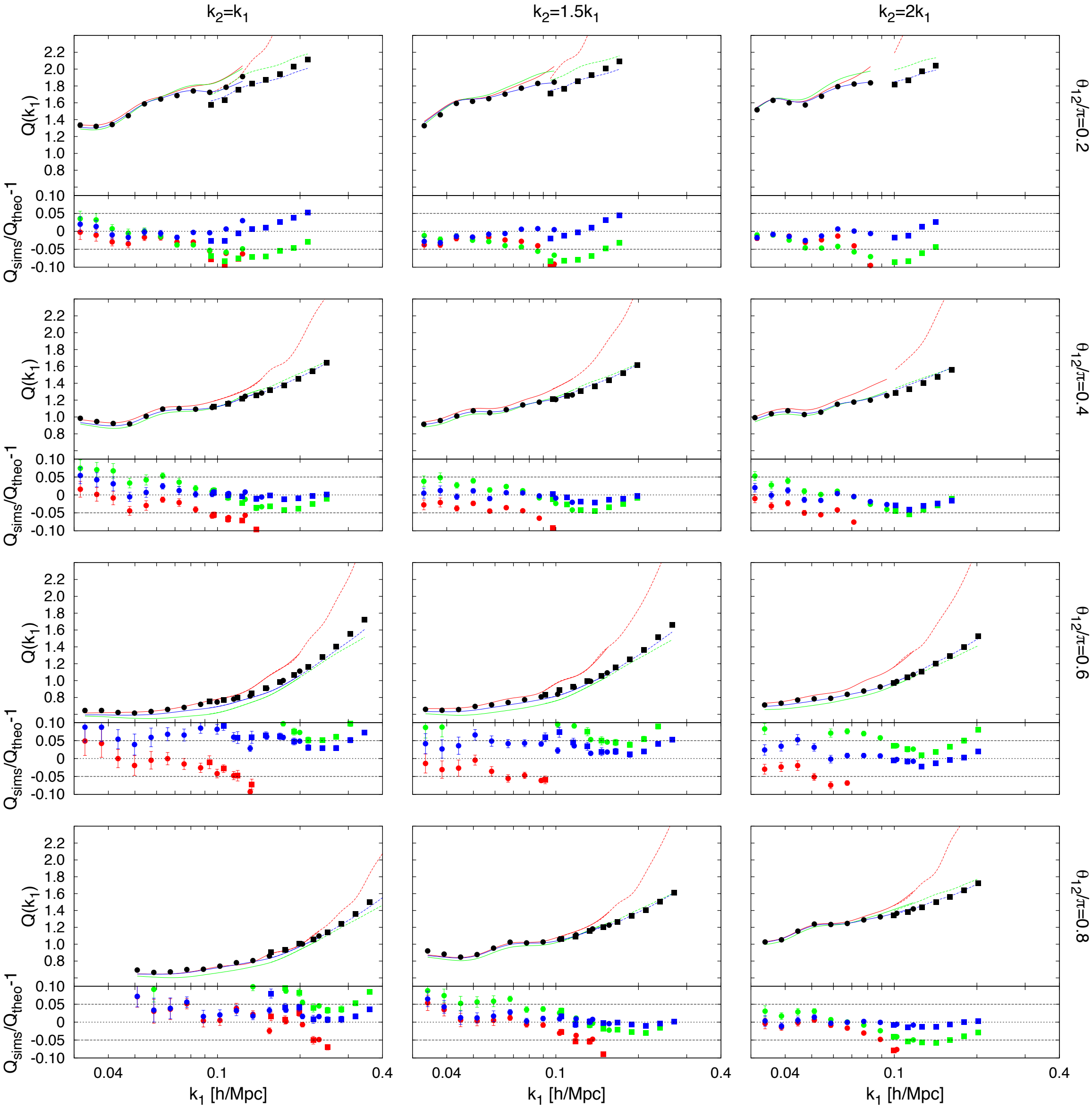

Figure 2. In the *main* panels: Q vs. k_1 for N-body simulation (black circles for simulation A and black squares for simulation B), for 1-loop correction (red solid lines to fit simulation A data and dashed line to fit simulation B), for SC formula (green solid line to fit A and green dashed line to fit B) and for our model (blue solid line to fit A and blue dashed line to fit B) at $z = 0$. In the *sub-panels*, the ratio between simulations and different theoretical models is shown: 1-loop (red points), SC (green points) and our model (blue points) for different triangles configurations. Circle symbols comes from simulation A and squares symbol from simulation B. The error bars show the measured errors from the simulations. Dashed lines mark 5% deviation from data. From left to right panels: $k_2/k_1 = 1.0, 1.5, 2.0$. From top to bottom panels: $\theta_{12}/\pi = 0.2, 0.4, 0.6, 0.8$.

deviations up to 10%. However for the scalene configurations $k_2 = 2k_1$ the deviations are only of order 3%. In the cases studied here our model improves significantly the SC fitting formula.

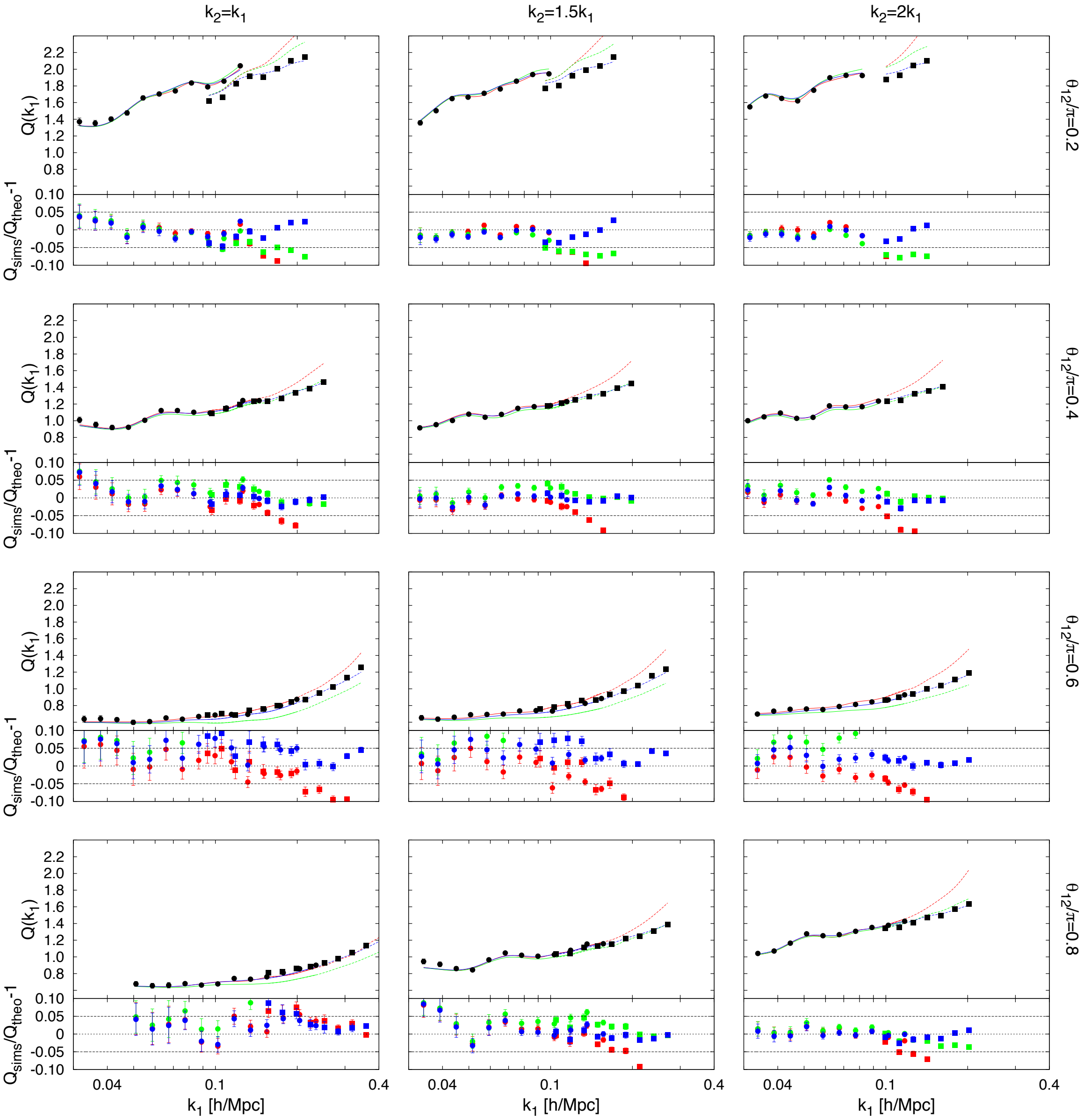

Figure 3. $Q(k_1)$ for different triangles configurations at $z = 1$. Same notation that in figure 2.

5 Conclusions

In this paper we propose a new simple formula to compute the dark matter bispectrum in the moderate non-linear regime ($k < 0.4\,h/\text{Mpc}$) and for redshifts $z \leq 1.5$. Our method is inspired by the approach presented in [1], but includes a modification of the original formulae, namely eq. (2.12), and a prescription to better describe the BAO oscillations. The cosmology dependence of the reduced bispectrum is known to be very weak, and that of the bispectrum is almost completely contained in the power spectrum. Given that the cosmological model today is well constrained by observations we have considered a single cosmology here.

Using LCDM simulations we fit the free parameters of our model. We end up with a simple analytic formula that is able to predict accurately the bispectrum for a LCDM

Universe including the effects of BAO. Our main results are summarized by eq. (2.11) where the kernel is given by eq. (2.6), the functions a, b, c are now given by $\tilde{a}, \tilde{b}, \tilde{c}$ of eq. (2.12), the fitting coefficients take the values reported in table 2 and the function $Q_3(n)$ is still given by eq. (2.10). The local slope of the linear power spectrum $n(k)$ is not any more given directly by eq. (2.8) but is a smoothed (BAO-free, but with the same broadband behavior) function of k.

The main conclusions of our work are listed below.

1. Our method is able to predict the dark matter bispectrum for a wide range of triangle configurations up to $k = 0.4\,h/\mathrm{Mpc}$ and for a redshift range $0 \leq z \leq 1.5$. In particular, for the reduced bispectrum, our fitting formula agrees within 5% with N-body data for most of the triangle configurations and always within 10% for the worst cases. This presents a considerable improvement over previous phenomenological approaches and over the prediction of Eulerian perturbation theory.

2. The equilateral and quasi-equilateral configurations are the ones for which our model deviates most strongly from N-body data. We interpret this as being due to the fact that when the 3 sides of the triangle are similar, non-linearities start to play a role at the same time, and thus, the effect on the bispectrum is stronger than when the non-linearities enter at different times, i.e. for elongated triangles. Other methods, like the one described by SC show the same behavior.

3. We have checked that our model also works well for non-standard LCDM cosmologies (see appendix B). In particular, we have checked that for $k_2 = 2k_1$ the deviation between N-body data and our model is never higher than 3% and for equilateral triangles reaches 10%. Also in these non-standard LCDM cases studied here, our model works better than the SC fitting formula.

We envision that this new analytic fitting formula will be very useful in providing a reliable prediction for the non-linear dark matter bispectrum for LCDM models. In particular, simple analytic predictions with high accuracy will be needed for the data analysis in the forthcoming era of precision data.

Acknowledgments

We thank Fabian Schmidt for providing the non-standard LCDM simulations used in appendix B. Héctor Gil-Marín thanks the Argerlander Institut für Astronomie at the University of Bonn for hospitality. Héctor Gil-Marín is supported by CSIC-JAE grant. Christian Wagner and Licia Verde acknowledge support of FP7-IDEAS-Phys.LSS 240117.

A Bispectrum estimator & error bars

Here we present details on the computation of the bispectrum and its error bars from N-body simulations. Moreover, we compare our error estimates with the Gaussian analytic predictions and discuss the differences.

We start by defining the estimator for the bispectrum as,

$$\hat{B}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \equiv \frac{V_f}{V_B} \int_{k_1} d^3q_1 \int_{k_2} d^3q_2 \int_{k_3} d^3\delta_D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)\delta_{q_1}\delta_{q_2}\delta_{q_3} \tag{A.1}$$

where $V_f = (2\pi)^3/L_b^3 \equiv k_f^3$ is the volume of the fundamental cell, k_f. The integration is defined over the bin $k_i - \Delta k_i/2 < q_i < k_i + \Delta k_i/2$. In this paper we always take $\Delta k = 3k_f$. V_B is the six-dimensional volume of triangles defined by the triangle sizes k_1, k_2 and k_3 with uncertainty Δk. Its value can be approximated by

$$V_B(k_1,k_2,k_3) = \int_{k_1} d^3q_1 \int_{k_2} d^3q_2 \int_{k_3} d^3q_3\, \delta_D(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3) \simeq 8\pi^2 k_1k_2k_3\Delta k^3 \tag{A.2}$$

which is good enough for not too small values of k_i. The variance associated to this estimator depends on higher-order correlation functions: up to the 6-point connected correlation function. However, the main contribution to the variance is given by the power spectrum. Assuming that the fields are Gaussian, the variance associated to estimator presented above is [39],

$$\Delta\hat{B}^2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = s_B\frac{V_f}{V_B}(2\pi)^3P(k_1)P(k_2)P(k_3) \tag{A.3}$$

where the symmetry factor is $s_B = 6,\ 2,\ 1$ for equilateral, isosceles or scalene configurations. The factor $(2\pi)^3$ comes from our definition of the power spectrum and bispectrum in eq. (2.1) and (2.2).

On the other hand, the discretized version of this estimator used in this paper is, $\widetilde{B}$

$$\widetilde{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \frac{L_b^6}{N_{tri}}\sum_j^{N_{tri}} \mathrm{Re}\left[\delta_j^d(k_1)\delta_j^d(k_2)\delta_j^d(k_3)\right] \tag{A.4}$$

where N_{tri} is the number of random triangle configurations used to compute the bispectrum; and j runs over these triangle configurations. For this work we use a number of random triangles that increases with k in the same way as the number of fundamental triangles: $\sim \tilde{V}_B/V_f^2$. It reaches up to $N_{tri} \sim 10^9$ for scales $k \sim 0.4\,h/\mathrm{Mpc}$. We have checked that increasing the number of random triangles beyond this value has no effect neither in the value of the bispectrum nor in its error. The index d in the δ field stands for a discrete and dimensionless quantity. Therefore the quantity $\mathrm{Re}\left[\delta_j^d(k_1)\delta_j^d(k_2)\delta_j^d(k_3)\right]$ needs to be rescaled with the factor L_b^6 to make $\widetilde{B}$ matching with the definition of the bispectrum in eq. (2.2). We compute the variance of this estimator $\widetilde{B}$ by the sample variance derived from the N_r realizations,

$$\Delta\widetilde{B}^2(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \frac{1}{N_r-1}\sum_i^{N_r}\left(\widetilde{B}_i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) - \langle\widetilde{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\rangle\right)^2 \tag{A.5}$$

where $\widetilde{B}_i$ is the bispectrum derived from the realization i and $\langle\widetilde{B}\rangle$ is the mean over all realizations,

$$\langle\widetilde{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\rangle \equiv \frac{1}{N_r}\sum_i^{N_r}\widetilde{B}_i(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\,. \tag{A.6}$$

The error on the mean $\langle\tilde{B}\rangle$ is then simply given by

$$\sigma_{\langle\tilde{B}\rangle} = \Delta\widetilde{B}/\sqrt{N_r}\,. \tag{A.7}$$

When comparing the measured N-body bispectrum with theoretical models and also when comparing eq. (A.3) with eq. (A.5), it is important to take into account the effect

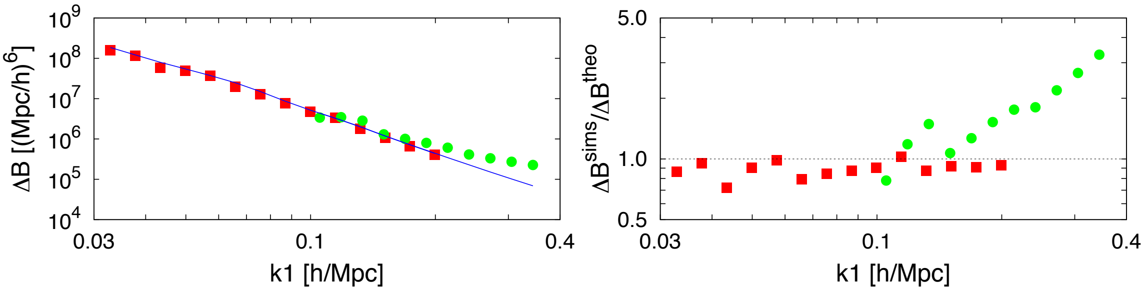

Figure 4. *Left panel*: ΔB for $k_2/k_1 = 1$ and $\theta_{12} = 0.6\pi$ triangles as a function of k_1 derived from simulations A (red squares) and from simulations B (green circles). Blue line is the theoretical prediction according to eq. (A.3). *Right panel*: $\Delta\tilde{B}/\Delta\hat{B}$ for simulations A (red squares) and B (green circles).

on the finite size of the triangle bins: each configuration is defined in terms of the sides of the triangle $k_i \pm \Delta k/2$. In this case, we are assuming $\Delta k = 3k_f$ and a large number of fundamental triangles fit into this bin. For certain configurations, it turns out that we have more triangles with k larger than the central value, k_i. Because of that, one must correct the sides of the triangles by

$$\tilde{k_i} = \frac{1}{N_{tri}} \sum_{j}^{N_{tri}} k_i^j \tag{A.8}$$

where $i = 1, 2, 3$ for each dimension and the sum is taken over all random triangle generated in the bin. This correction is extremely important at large scales and for very squeezed triangles, and less important for equilateral configuration.

In figure 4 we present a comparison of the error estimation from theoretical models (eq. (A.3)) and simulations (eq. (A.5)). In the left panel we show the error of the bispectrum associated to a volume of 1 single box for $\Delta k = 3k_f$: using the theoretical model and the simulations for the case of $k_2/k_1 = 1$ and $\theta_{12} = 0.6\pi$ triangle configurations. The blue line shows the theoretical model prediction for the error of the bispectrum of one single realization using the non-linear power spectra from simulations A and B (eq. (A.3)); whereas the red squares and green circles show the dispersion among the runs of simulations A and B respectively (eq. (A.5)). In the right panel the ratio between the errors according to the simulations and the Gaussian prediction is plotted for simulations A (red squares) and B (green circles). The error estimates of simulations A agrees well with the theoretical model. On the other hand, on small scales the error estimates of simulations B is larger than the theoretical model and further increase with decreasing scale. Similar results were found by [40]. These differences are due to the fact that eq. (A.3) neglects any higher-order contributions (because it assumes Gaussianity). However, at small scales this is no longer a good approximation as it is shown in [36]. Furthermore, the errors of simulations B have been estimated by dividing each of the 3 simulation boxes into 8 sub-boxes. This introduces extra non-Gaussian terms [36] that are not taken into account in eq. (A.3).

B Our fitting formula for non-standard LCDM models

Here we test how our model works with different LCDM simulations to those we have used to fit the a_i parameters. In particular, we test our model with LCDM simulations with a

$f(R)$-like power spectrum. For a full description of the simulations and the $f(R)$ gravity we refer the reader to [19]. These simulations were run with ENZO code and have slightly different cosmology than the ones used in the rest of this paper: $\Omega_\Lambda = 0.76$, $\Omega_m = 0.24$, $\Omega_b = 0.04181$, $h = 0.73$. They consist of 6 realizations that contain 256^3 particles in a box of 400 Mpc$/h$ per side. The one-quarter Nyquist frequency is $k_N/4 = 0.5\, h/\text{Mpc}$.

In figure 5 the reduced bispectrum Q is shown: in the right panel as a function of the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$, namely θ_{12}, for $k_2 = 2k_1$; in the left panel as a function of k_1 for equilateral configuration, both for $z = 0$. All panels correspond to a LCDM model with no BAOs whose initial conditions make their power spectrum look like a $f(R)$-like one. In order to do that a running index has been adopted in the initial conditions (see table 1 in [19] for details). Panels correspond to LCDM simulations that match with $f(R)$ models whose $|f_{R0}|$ parameter[6] is: 10^{-4} (top panels), 10^{-5} (middle panels) and 10^{-6} (bottom panels).

Black points are data from simulations, green line is the SC prediction and blue line the prediction of our model. In the right panel we only compare data points for $0.4 < \theta_{12}/\pi < 0.9$ and in the left panel $0.1\, h/\text{Mpc} < k < 0.5\, h/\text{Mpc}$ in order to ensure that all the k_i are smaller than a quarter of the Nyquist frequency.

Considering the right panels of figure 5 ($k_2 = 2k_1 = 0.4\, h/\text{Mpc}$), we see that our model describes the data within about 3%. In particular for small scales ($\theta_{12} < 0.7\pi$), both SC and our model agree with the simulations data well; however at large scales ($\theta_{12} > 0.7\pi$), our model fits the data points better. In the left panel of figure 5 (equilateral configuration), we see that our model shows deviations up to 10%. As for the standard LCDM model, the equilateral configuration is the one with the largest deviations. However even in this case, our formula behaves better than the SC model, especially at small scales.

Therefore we conclude that our formula is general enough to be applied also to some non-standard LCDM models and in particular works better than SC at small scales.

C One-loop correction terms for the power spectrum and bispectrum

Here we present a short description of the equations used to compute the one-loop correction in Eulerian perturbation theory for the bispectrum shown in figure 2 and 3. For a detailed description of PT see [27, 39].

Up to one-loop, the power spectrum can be expressed as,

$$P(k) = P^{(0)}(k) + P^{(1)}(k) + \dots \tag{C.1}$$

where $P^{(0)}(k) = P_L(k)$ is the linear term and $P^{(1)}(k) = P_{13}(k) + P_{22}(k)$ is the one loop correction. For Gaussian initial conditions, the one-loop term consist of two terms,[7]

$$P_{22} = \frac{2}{(2\pi)^3} \int d^3\mathbf{q}\, F_2^{s\,2}(\mathbf{q}, \mathbf{k}-\mathbf{q}) P_L(q) P_L(|\mathbf{k}-\mathbf{q}|) \tag{C.2}$$

$$P_{13} = \frac{6}{(2\pi)^3} P_L(k) \int d^3\mathbf{q} F_3^s(\mathbf{k}, \mathbf{q}, -\mathbf{q}) P_L(q) \tag{C.3}$$

P_{22} accounts for the mode coupling between waves with wave-vectors $\mathbf{k}-\mathbf{q}$ and $\mathbf{q}$, whereas P_{13} can be interpreted as the one-loop correction to the linear propagator.

---

[6] See eq. 2.3 in [19] for a definition of f_{R0}.

[7] The $(2\pi)^3$ in the denominator comes from the definition of the power spectrum and bispectrum in eq. (2.1) and (2.2).

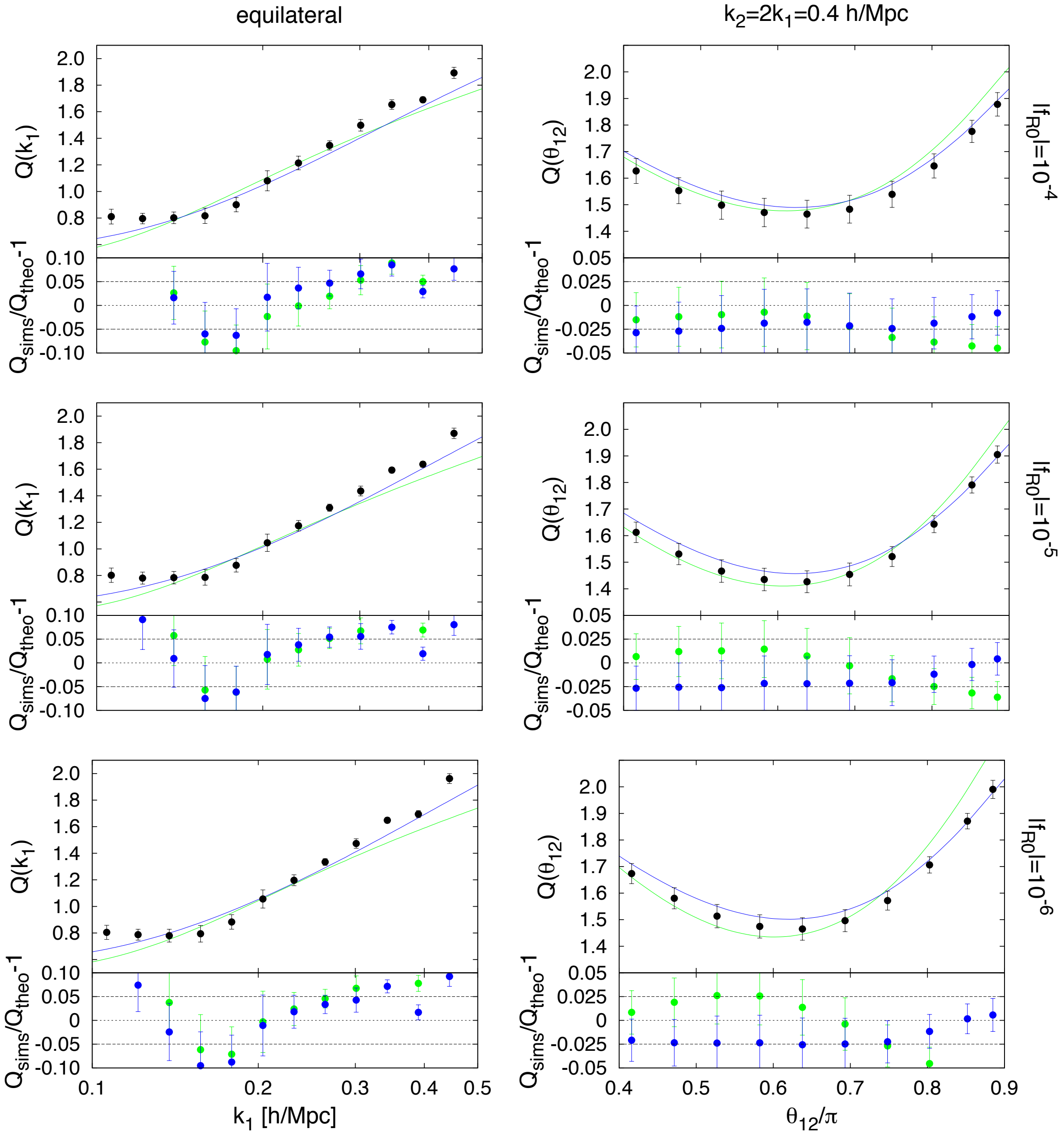

Figure 5. In the left panels: Q vs. k_1 for equilateral configuration. In the right panels Q vs θ_{12}/π for $k_2 = 2k_1 = 0.4\,h/\mathrm{Mpc}$. From top to bottom we show different non-standard LCDM models. All the panels correspond to LCDM models with $f(R)$-like power spectrum from [19]. Top panels correspond to matching to $|f_{R0}| = 10^{-4}$; middle panels to $|f_{R0}| = 10^{-5}$; bottom panels to $|f_{R0}| = 10^{-6}$. The *sub-panel* shows the corresponding ratio between simulations and different theoretical models: SC (green points) and our work (blue points). In the right sub-panels, dashed lines mark 2.5% deviation, whereas in the left panels mark the 5% deviation. The bispectrum and its error are estimated by eq. (A.6) and (A.7) (see appendix A).

Similarly to the power spectrum, the bispectrum up to one loop consists of two terms,

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = B^{(0)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + B^{(1)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + \ldots \tag{C.4}$$

For Gaussian initial conditions, the first non-zero term is the second-order contribution $B^{(0)}$

which is *tree level* correction, whereas $B^{(1)}$ is the one-loop correction. The tree level term can be expressed as,

$$B^{(0)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2F_2^s(\mathbf{k_1}, \mathbf{k_2})P_L(k_1)P_L(k_2) + 2 \text{ cyc. perm.} \tag{C.5}$$

whereas the one loop consist of four terms (only for Gaussian initial conditions),

$$\begin{aligned} B^{(1)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = {} & B^I_{222}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + B_{123}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \\ & + B^{II}_{123}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + B^I_{114}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \end{aligned} \tag{C.6}$$

Each of these terms read as,

$$\begin{aligned} B^I_{222} = {} & \frac{8}{(2\pi)^3} \int d^3\mathbf{q}\, F_2^s(-\mathbf{q}, \mathbf{q} + \mathbf{k}_1) F_2^s(-\mathbf{q} - \mathbf{k}_1, \mathbf{q} - \mathbf{k}_2) F_2^s(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) P_L(q) \times \\ & \times P_L(|\mathbf{k}_1 + \mathbf{q}|) P_L(|\mathbf{k}_2 - \mathbf{q}|) \end{aligned} \tag{C.7}$$

$$B^I_{123} = \frac{6}{(2\pi)^3} P_L(k_1) \int d^3\mathbf{q}\, F_3^s(\mathbf{k}_1, \mathbf{k}_2 - \mathbf{q}, \mathbf{q}) F_2^s(\mathbf{k}_2 - \mathbf{q}, \mathbf{q}) P_L(|\mathbf{k}_2 - \mathbf{q}) P_L(q) + 5 \text{ perm.} \tag{C.8}$$

$$B^{II}_{123} = F_2^s(\mathbf{k}_1, \mathbf{k}_2) \left[P_L(k_1) P_{13}(k_2) + P_L(k_2) P_{13}(k_1)\right] + 2 \text{ perm.} \tag{C.9}$$

$$B^I_{114} = \frac{12}{(2\pi)^3} P_L(k_1) P_L(k_2) \int d^3\mathbf{q}\, F_4^s(\mathbf{q}, -\mathbf{q}, -\mathbf{k}_1, -\mathbf{k}_2) P_L(q) + 2 \text{ perm.} \tag{C.10}$$

where F_i^s are the *symmetrized* kernels. For an EdS Universe, the non-symmetric kernels read as,

$$\begin{aligned} F_n(\mathbf{q}_1, \ldots, \mathbf{q}_n) = {} & \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \ldots, \mathbf{q}_m)}{(2n+3)(n-1)} \left[(2n+1)\alpha(\mathbf{k}, \mathbf{k}_1) F_{n-m}(\mathbf{q}_{m+1}, \ldots, \mathbf{q}_n)\right. \\ & \left. + 2\beta(\mathbf{k}, \mathbf{k}_1, \mathbf{k}_2) G_{n-m}(\mathbf{q}_{m+1}, \ldots, \mathbf{q}_n)\right] \end{aligned} \tag{C.11}$$

$$\begin{aligned} G_n(\mathbf{q}_1, \ldots, \mathbf{q}_n) = {} & \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \ldots, \mathbf{q}_m)}{(2n+3)(n-1)} \left[3\alpha(\mathbf{k}, \mathbf{k}_1) F_{n-m}(\mathbf{q}_{m+1}, \ldots, \mathbf{q}_n)\right. \\ & \left. + 2n\beta(\mathbf{k}, \mathbf{k}_1, \mathbf{k}_2) G_{n-m}(\mathbf{q}_{m+1}, \ldots, \mathbf{q}_n)\right] \end{aligned} \tag{C.12}$$

with $F_1 = G_1 = 1$. Also, $\mathbf{k}_1 \equiv \mathbf{q}_1 + \cdots + \mathbf{q}_m$, $\mathbf{k}_2 \equiv \mathbf{q}_{m+1} + \cdots + \mathbf{q}_n$, $\mathbf{k} \equiv \mathbf{k}_1 + \mathbf{k}_2$ and the functions α and β are defined as,

$$\alpha(\mathbf{k}, \mathbf{k}_1) \equiv \frac{\mathbf{k} \cdot \mathbf{k}_1}{k_1^2} \tag{C.13}$$

$$\beta(\mathbf{k}, \mathbf{k}_1, \mathbf{k}_2) \equiv \frac{k^2(\mathbf{k}_1 \cdot \mathbf{k}_2)}{2k_1^2 k_2^2} \tag{C.14}$$

In order to obtain the symmetric kernels one has to symmetrize them with respect to their arguments,

$$F_n^s(\mathbf{q}_1, \ldots, \mathbf{q}_n) = \frac{1}{n!} \sum_\pi F_n(\mathbf{q}_{\pi(1)}, \ldots, \mathbf{q}_{\pi(n)}) \tag{C.15}$$

where the sum is taken over all the permutations π of the set $\{1, \ldots, n\}$.

Finally, the reduced bispectrum up to one loop can be written as,

$$Q = \frac{B^{(0)} + B^{(1)} + \ldots}{\Sigma^{(0)} + \Sigma^{(1)} + \ldots} \simeq Q^{(0)} + Q^{(1)} + \ldots \tag{C.16}$$

where

$$\Sigma^{(0)} = P_L(k_1)P_L(k_2) + P_L(k_1)P_L(k_3) + P_L(k_2)P_L(k_3) \tag{C.17}$$

$$\begin{aligned}\Sigma^{(1)} = {} & P_L(k_1)P^{(1)}(k_2) + P^{(1)}(k_1)P_L(k_2) + P_L(k_1)P^{(1)}(k_3) \\ & + P^{(1)}(k_1)P_L(k_3) + P_L(k_2)P^{(1)}(k_3) + P^{(1)}(k_2)P_L(k_3)\end{aligned} \tag{C.18}$$

Then it is easy to show that the linear and 1-loop correction terms for Q reads as,

$$Q^{(0)} = \frac{B^{(0)}}{\Sigma^{(0)}} \tag{C.19}$$

$$Q^{(1)} = \frac{B^{(1)} - Q^{(0)}\Sigma^{(1)}}{\Sigma^{(0)}} \tag{C.20}$$

References

[1] R. Scoccimarro and H. Couchman, *A fitting formula for the nonlinear evolution of the bispectrum*, *Mon. Not. Roy. Astron. Soc.* **325** (2001) 1312 [`astro-ph/0009427`] [INSPIRE].

[2] B.A. Reid et al., *Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies*, *Mon. Not. Roy. Astron. Soc.* **404** (2010) 60 [`arXiv:0907.1659`] [INSPIRE].

[3] J.N. Fry and L.A. Melott, *Statistical comparison of galaxy formation models: the bispectrum*, *Astrophys. J.* **292** (1985) 395.

[4] I. Kayo et al., *Three-point correlation functions of SDSS galaxies in redshift space: morphology, color, and luminosity dependence*, *Proc. Astron. Soc. Jpn.* **56** (2004) 415 [`arXiv:1007.2414`] [INSPIRE].

[5] H. Guo and Y. Jing, *Determine the galaxy bias factors on large scales using bispectrum method*, *Astrophys. J.* **702** (2009) 425 [`arXiv:0907.0282`] [INSPIRE].

[6] J.N. Fry, *Gravity, bias, and the galaxy three-point correlation function*, *Phys. Rev. Lett.* **73** (1994) 215.

[7] L. Verde, A.F. Heavens, S. Matarrese and L. Moscardini, *Large scale bias in the universe. 2. Redshift space bispectrum*, *Mon. Not. Roy. Astron. Soc.* **300** (1998) 747 [`astro-ph/9806028`] [INSPIRE].

[8] S. Matarrese, L. Verde and A. Heavens, *Large scale bias in the universe: Bispectrum method*, *Mon. Not. Roy. Astron. Soc.* **290** (1997) 651 [`astro-ph/9706059`] [INSPIRE].

[9] J.N. Fry, A.L. Melott and S.F. Shandarin, *Biased power spectrum and bispectrum for an ensemble of three-dimensional scale free numerical simulations*, *Mon. Not. Roy. Astron. Soc.* **274** (1995) 745 [`astro-ph/9404037`] [INSPIRE].

[10] R. Scoccimarro, H.A. Feldman, J.N. Fry and J.A. Frieman, *The bispectrum of IRAS redshift catalogs*, *Astrophys. J.* **546** (2001) 652 [`astro-ph/0004087`] [INSPIRE].

[11] H.A. Feldman, J.A. Frieman, J.N. Fry and R. Scoccimarro, *Constraints on galaxy bias, matter density and primordial non-gausianity from the PSCz galaxy redshift survey*, *Phys. Rev. Lett.* **86** (2001) 1434 [`astro-ph/0010205`] [INSPIRE].

[12] L. Verde et al., *The 2dF Galaxy Redshift Survey: the bias of galaxies and the density of the universe*, *Mon. Not. Roy. Astron. Soc.* **335** (2002) 432 [`astro-ph/0112161`] [INSPIRE].

[13] J.E. Pollack, R.E. Smith and C. Porciani, *Modelling large-scale halo bias using the bispectrum*, to be published in *Mon. Not. Roy. Astron. Soc.* (2012).

[14] D. Jeong and E. Komatsu, *Primordial non-gaussianity, scale-dependent bias and the bispectrum of galaxies*, *Astrophys. J.* **703** (2009) 1230 [`arXiv:0904.0497`] [INSPIRE].

[15] L. Verde et al., *Large-scale structure, the cosmic microwave background, and primordial non-gaussianity*, *Mon. Not. Roy. Astron. Soc.* **313** (2000) L141 [`astro-ph/9906301`] [INSPIRE].

[16] L. Verde, R. Jimenez, M. Kamionkowski and S. Matarrese, *Tests for primordial non-gaussianity*, *Mon. Not. Roy. Astron. Soc.* **325** (2001) 412 [`astro-ph/0011180`] [INSPIRE].

[17] E. Sefusatti and E. Komatsu, *The bispectrum of galaxies from high-redshift galaxy surveys: primordial non-gaussianity and non-linear galaxy bias*, *Phys. Rev.* **D 76** (2007) 083004 [`arXiv:0705.0343`] [INSPIRE].

[18] R. Scoccimarro, E. Sefusatti and M. Zaldarriaga, *Probing primordial non-Gaussianity with large-scale structure*, *Phys. Rev.* **D 69** (2004) 103513 [`astro-ph/0312286`] [INSPIRE].

[19] H. Gil-Marín et al., *The bispectrum of $f(R)$ cosmologies*, *JCAP* **11** (2011) 019.

[20] A. Shirata, Y. Suto, C. Hikage, T. Shiromizu and N. Yoshida, *Galaxy clustering constraints on deviations from Newtonian gravity at cosmological scales II: perturbative and numerical analyses of power spectrum and bispectrum*, *Phys. Rev.* **D 76** (2007) 044026 [`arXiv:0705.1311`] [INSPIRE].

[21] M. Crocce and R. Scoccimarro, *Renormalized cosmological perturbation theory*, *Phys. Rev.* **D 73** (2006) 063519 [`astro-ph/0509418`] [INSPIRE].

[22] M. Pietroni, *Flowing with time: a new approach to non-linear cosmological perturbations*, *JCAP* **10** (2008) 036.

[23] A. Cooray and R.K. Sheth, *Halo models of large scale structure*, *Phys. Rept.* **372** (2002) 1 [`astro-ph/0206508`] [INSPIRE].

[24] C.-P. Ma and J.N. Fry, *Deriving the nonlinear cosmological power spectrum and bispectrum from analytic dark matter halo profiles and mass functions*, *Astrophys. J.* **543** (2000) 503 [`astro-ph/0003343`] [INSPIRE].

[25] VIRGO CONSORTIUM collaboration, R. Smith et al., *Stable clustering, the halo model and nonlinear cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1311 [`astro-ph/0207664`] [INSPIRE].

[26] J. Pan, P. Coles and I. Szapudi, *Scale transformations, tree-level perturbation theory and the cosmological matter bispectrum*, *Mon. Not. Roy. Astron. Soc.* **382** (2007) 1460.

[27] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [`astro-ph/0112551`] [INSPIRE].

[28] J.N. Fry, *The Galaxy correlation hierarchy in perturbation theory*, *Astrophys. J.* **279** (1984) 499 [INSPIRE].

[29] R. Scoccimarro and J.A. Frieman, *Hyperextended cosmological perturbation theory: predicting nonlinear clustering amplitudes*, *Astrophys. J.* **520** (1999) 35 [`astro-ph/9811184`] [INSPIRE].

[30] E. Sefusatti, M. Crocce and V. Desjacques, *The matter bispectrum in N-body simulations with non-gaussian initial conditions*, *Mon. Not. Roy. Astron. Soc.* **406** (2010) 1014 [`arXiv:1003.0007`] [INSPIRE].

[31] K. Heitmann, M. White, C. Wagner, S. Habib and D. Higdon, *The Coyote universe I: precision determination of the nonlinear matter power spectrum*, *Astrophys. J.* **715** (2010) 104 [`arXiv:0812.1052`] [INSPIRE].

[32] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [`astro-ph/9911177`] [INSPIRE].

[33] V. Springel, *The cosmological simulation code GADGET-2*, *Mon. Not. Roy. Astron. Soc.* **364** (2005) 1105 [`astro-ph/0505010`] [INSPIRE].

[34] A.J. Hamilton, C.D. Rimes and R. Scoccimarro, *On measuring the covariance matrix of the nonlinear power spectrum from simulations*, *Mon. Not. Roy. Astron. Soc.* **371** (2006) 1188 [`astro-ph/0511416`] [INSPIRE].

[35] C.D. Rimes and A.J. Hamilton, *Information content of the non-linear power spectrum: the effect of beat-coupling to large scales*, *Mon. Not. Roy. Astron. Soc.* **371** (2006) 1205 [`astro-ph/0511418`] [INSPIRE].

[36] E. Sefusatti, M. Crocce, S. Pueblas and R. Scoccimarro, *Cosmology and the bispectrum*, *Phys. Rev.* **D 74** (2006) 023522 [`astro-ph/0604505`] [INSPIRE].

[37] R. de Putter, C. Wagner, O. Mena, L. Verde and W. Percival, *Thinking outside the box: effects of modes larger than the survey on matter power spectrum covariance*, `arXiv:1111.6596` [INSPIRE].

[38] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical recipes in C. The art of scientific computing*, Cambridge University Press, Cambridge U.K. (1992).

[39] R. Scoccimarro et al., *Nonlinear evolution of the bispectrum of cosmological perturbations*, *Astrophys. J.* **496** (1998) 586 [`astro-ph/9704075`] [INSPIRE].

[40] H. Guo and Y. Jing, *A determination of dark matter bispectrum with a large set of n-body simulations*, *Astrophys. J.* **698** (2009) 479 [`arXiv:0904.3200`] [INSPIRE].

CHAPTER 6

Conclusions and future prospects

In this Chapter we discuss and summarise the main conclusions of this thesis. In the last subsection we also present the future prospects that would be interesting to attain given the results of this thesis.

6.1 A halo model with environment dependence: theoretical considerations

In the standard halo model the host halo mass is the only variable specifying halo and galaxy properties. This model has been remarkably successful at describing the first moment statistics of the clustering of galaxies. However, environment must play an important role in the process of galaxy formation, the most striking observational evidence being that clusters today have a much higher fraction of early type galaxies than is found in the field. In addition, when looking at statistics beyond the two-point density correlation function, there are indications that the simplest halo-model may be incomplete. Recent surveys, like the Sloan Digital Sky Survey (SDSS) and the Anglo-Australian Two-degree survey (2dF), have shown that there seems to be a correlation between the clustering of galaxies and their properties such as metallicity and star formation rate, which are believed to be environment-dependent. This environmental dependence is not included in the standard halo model where the host halo mass is the only variable specifying galaxy properties.

Following these ideas, in Chapter 2 we have presented a natural extension of the halo model that has allowed us to introduce an environmental dependence based on whether a halo lives in a high-mass density region, namely a node-like region, or in a low-mass region, namely a filament-like region. The halo properties i.e., the concentration, and the Halo Occupation Distribution (hereafter HOD) prescription, not only depend on the halo mass (like in the standard halo model) but also on the halo environment. We have examined how different environmental dependence of halo concentration and HOD prescription affect the correlation function. The main goal has been to provide an improved description of galaxy clustering. The main results from this Chapter can be summarised:

1. At the level of dark matter, the secondary variable (the environment-dependent variable beyond the shape of the mass function) is the concentration of the density profile of the haloes: haloes which live in nodes regions may have a different concentration than haloes which live in filaments regions. According to this idea we have presented the dark matter correlation function, ξ_{dm}, of this new extended halo model. In the classic halo model the correlation function is usually split in two terms: the one-halo and the two-halo term. In our extended halo model we have 3 terms: the one-halo term and 2 two-halo terms, depending whether the two particles belong to haloes of the same or different type. We have explored the contribution of these terms to the dark matter correlation function and we

have found that the contribution of particles that belong to haloes of the same and different kind is similar. We have also seen that the contribution of the node-like haloes starts to be dominant at small scales and is subdominant at large scales. We also have analysed how different values of the concentration affect the correlation function. At the level of dark matter, changing the concentration parameter only affects moderately the correlation function at small scales.

2. We also have extended our environment-dependent halo model to galaxies, computing the galaxy correlation function, ξ_{gal}. This has been done by modulating the Halo Occupation Distribution recipe with the environment. In the galaxy correlation function, the HOD is our secondary variable rather than halo concentration. This means that according to our model, node-like haloes host galaxies in a different way than filament-like galaxies. In particular, in this Chapter we have used a simple model for the HOD of only 3 variables: the minimum mass of a halo to host its first (central) galaxy, M_{min}; the mass of the halo to have on average its first satellite galaxy, M_1 and the way how satellite galaxies increase with the halo mass, α. Thus we have two different HOD, one for red-galaxies and another for blue-galaxies. We have analysed how changing the different parameters of the HOD affects to the galaxy correlation function. We have seen that, even when we change the parameters of the HOD of the minority population (in this case the node-like haloes), there is a considerable change in the ξ_{gal}, especially for the parameters M_1 and α, i.e., for the satellite population. Therefore changing the HOD of only a small fraction of haloes affects considerably to the total galaxy correlation function. We also have explored how the mixing or segregation between the two population of red and blue galaxies may affect to ξ_{gal}. We expect that the size of this effect will depend on the HOD of each one of these populations: the more different are the HOD of red and blue galaxies, the more the effect of segregation.

3. A natural extension of this model is to make a continuous dependence on the environment, instead of splitting it in nodes, filaments and voids. However this extension presents several issues. First of all we would need a continuous dependence on the environment of the mass function and also of the HOD, which is not yet available. Secondly, nowadays N-body simulations start to provide information of haloes according a discrete number of environments: this discretization could be taken as a first approximation to the continuous distribution. While this avenue is worth pursuing, introducing a more complex model based on a continuous environment dependence goes beyond the scope of this thesis.

6.2 Reducing sample variance: halo biasing, non-linearity and stochasticity

In Chapter 3 we have focused on reducing the sample variance error in the measurement of certain cosmological parameters comparing clustering of differently biased tracers of the dark matter distribution. We have developed a formalism that includes non-linear and stochastic bias. Our formalism is general enough that can be used to optimise survey design and tracers selection and optimally split (or combine) tracers to minimise the error on the cosmologically interesting quantities. Our approach generalises the one presented by McDonald and Seljak (2009) of circumventing sample variance in the measurement of $f = d\ln D/d\ln a$. Most of the statistical power of this method comes from very large scales, where cosmic variance is the dominant contribution to the statistical error-bars. By suppressing cosmic variance, this approach promises to reduce drastically error-bars on cosmologically very important quantities; for example it would allow for a high-precision determination of growth of structure as a function of redshift, as encoded in fD, and an improvement of dark energy figures of merit by large factors. We have analysed how the bias, the noise, the non-linearity and stochasticity can affect the measurements of Df and have explored in which signal-to-noise regime it is significantly advantageous to split a galaxy sample in two differently-biased tracers. We have used N-body simulations to find realistic values for the parameters describing the bias properties of dark matter haloes of different masses and their number density. The main features of these studies can be summarised as follows:

1. We have set out to generalise previous approaches of this technique by assuming that the bias of dark matter tracers may not be perfectly linear and allowing for some stochasticity. We have computed the expected error on the quantity fD achievable by comparing clustering of differently biased tracers (thus suppressing cosmic variance) and by combining the different tracers in a single sample (thus reducing shot noise and stochasticity but carrying along sample variance in full).

2. We have analysed how the bias, the noise, the non-linearity and stochasticity affect the measurements of fD and explored in which signal-to-noise regime it is significantly advantageous to split a galaxy sample in two differently-biased tracers. We have used results from N-body simulations to set plausible values for these parameters to see how great the gains may be in practice. We find that even small amount of stochasticity (either in the form of Poisson noise or in more general form) and of non-linearity can limit significantly the performance of the

two-tracers approach. In our analysis we also have assumed a scale-independent bias. This may be enough, for the mass range studied, if we consider only dark matter haloes as tracers. This is indeed what we have seen in simulations. On the other hand, it is also true that more realistic approaches, which account for galaxies as tracers instead of haloes, should include a scale-dependent bias. However, including this in our formalism can only reduce the gain achievable by splitting the sample. We expect the ratio of errors increases from the current value of 0.9 to even closer values to 1 if the bias is strongly scale dependent.

3. We have shown that only in the very high signal-to-noise regime it is significantly advantageous to split the sample and that, even though the gain is maximised by increasing the *ratio* of the biases of the two tracers, both tracers should be well sampled. We have explored different ways of selecting and splitting dark matter haloes obeying a ΛCDM mass function and found that one can achieve up to a 40% reduction of the error on fD. While this would correspond to the gain from a three times larger survey volume if the two tracers were not to be split, it is much smaller that the improvement forecasted in the absence of stochasticity and bias non-linearity.

4. In addition, we have noted that these findings apply to dark matter haloes as tracers, while realistic surveys would select galaxies: the galaxy-host halo relation is likely to introduce extra stochasticity which would reduce the gain further. The formalism we have developed, however, is general enough that can be used to optimise survey design and tracers selection and optimally split (or combine) tracers to minimise the error on the cosmologically interesting quantities.

6.3 The bispectrum of $f(R)$ cosmologies

In Chapter 4 we have analysed a suite of cosmological simulations of modified gravitational action $f(R)$ models, where cosmic acceleration is induced by a scalar field that acts as a fifth force on all forms of matter. In particular, we focus on the bispectrum of the dark matter density field on mildly non-linear scales. For models with the same *initial* power spectrum, the dark matter bispectrum has shown significant differences for cases where the final dark matter power spectrum also differs. Given the different dependence on bias of the galaxy power spectrum and bispectrum, bispectrum measurements can close the loophole of galaxy bias hiding differences in the power spectrum. Alternatively, changes in the initial power spectrum can also hide differences. By constructing ΛCDM models with very similar *final* non-linear power spectra, we have shown that the differences in the bispectrum are reduced and are comparable with differences in the imperfectly matched power spectra. These results have indicated that the bispectrum depends mainly on the power spectrum and less sensitively on the gravitational signatures of the $f(R)$ model. This weak dependence of the matter bispectrum on gravity makes it useful for breaking degeneracies associated with galaxy bias, even for models beyond general relativity.

In order to a approach to the problem, we have proceeded in two different ways to analyze the bispectrum deviation from these simulations, methods A and B, which differ in whether the initial or final power spectra of the two cosmologies are set equal. The main conclusions of the results are:

1. Method A has compared the bispectrum output of $f(R)$ and ΛCDM N-body simulations with the same initial power spectrum. We have found that the bispectrum deviation obtained using this method presents a considerable deviation (up to $10-15\%$) in the reduced bispectrum between these $f(R)$ models and the ΛCDM one. Such differences in the bispectrum could be easily detected by surveys covering volumes $> 1\,\mathrm{Gpc}$ such as e.g., the on-going BOSS survey. Higher deviations have been seen for higher values of $|f_{R0}|$ and for squeezed triangle configurations. In this method, both ΛCDM and $f(R)$ gravity runs start from the same initial δ_k values. Because of that, the different evolution of the gravity models naturally leads to different power spectra and also to different bispectra. In order to explain discrepancies between the matter power spectrum in $f(R)$ and the observed galaxy power spectrum, one could invoke a scale-dependent galaxy bias. Since galaxy bias enters into the reduced galaxy bispectrum in a different way than in the power spectrum, bispectrum measurements can in principle close this loophole.

2. Alternatively, the large power spectrum differences can be eliminated by changing the shape of the initial power spectrum to instead match the final dark matter power spectrum at $z = 0$. This is at the base of method B. In this method, we have computed the bispectrum deviation between a ΛCDM and a $f(R)$ model, both with the same final power spectra. Thus, we have simulated a ΛCDM model with certain initial power spectrum parameters that are adjusted to best match the $f(R)$ power spectrum at $z = 0$. From the simulations outputs we have computed the power spectra and bispectra. For the power spectra, residual differences have been never higher than 4% in the range $0.1\,h/\mathrm{Mpc} < k < 1\,h/\mathrm{Mpc}$. Likewise the differences in the reduced bispectrum have been also smaller in the matched comparison. For the $|f_{R0}| = 10^{-4}$ and 10^{-6} cases, the Q deviation has been consistent

with 0 within 1σ. For $|f_{R0}| = 10^{-5}$ deviations in Q at most have reached the 4% level with $5 - 6\sigma$ significance. These deviations are potentially a signature of the onset of the chameleon mechanism in the largest structures in the Universe. However given that this is the same order as the power spectrum difference it is unclear whether these differences indicate power-spectrum-independent modified gravity effects or that the two power spectra have not been perfectly matched. In the former case, larger surveys like EUCLID will allow for a measurement of the bispectrum with enough precision to obtain a $> 6\sigma$ significance, even when exactly matching the power spectra.

3. On the other hand, the effect of deviations from GR gravity on the reduced bispectrum are weak compared to those on the power spectrum (at least for the cases considered here), opening up the possibility of breaking the galaxy-bias degeneracy. In fact the effect of galaxy bias is expected to be different in the power spectrum and in the bispectrum, which is why, in the context of GR gravity, the bispectrum is used to constrain galaxy bias. While the shape of the non-linear power spectrum seems to carry information about the underlying gravity model, one may always argue that a shape of the evolved power spectrum not compatible with the GR predictions could be due to biasing. For the cases we have considered here, the dependence of the reduced bispectrum on deviations from GR is weaker than the effects of bias modifications necessary to explain the deviations in the power spectrum. While we have only studied $f(R)$ models here, there is no apparent reason why this result should be specific to $f(R)$. Hence, if our findings were to remain qualitatively true for other gravity modifications, this would confirm the usefulness of employing the reduced bispectrum together with the power spectrum to constrain bias parameters.

6.4 An improved fitting formula for the dark matter bispectrum

In Chapter 5 we have presented an improved fitting formula for the dark matter bispectrum motivated by the previous phenomenological approach of Scoccimarro and Couchman (2001). We have used a set of ΛCDM simulations to calibrate the fitting parameters in the k-range of $0.03\,h/\mathrm{Mpc} \leq k \leq 0.4\,h/\mathrm{Mpc}$ and in the redshift range of $0 \leq z \leq 1.5$. This new proposed fit has showed to describe well the BAO-features although it was not designed to. The deviation between the simulations output and our analytic prediction has been typically less than 5% and in the worst case is never above 10%. Here we list the main conclusions:

1. Our method has been able to predict the dark matter bispectrum for a wide range of triangle configurations up to $k = 0.4\,h/\mathrm{Mpc}$ and for a redshift range $0 \leq z \leq 1.5$. In particular, for the reduced bispectrum, our fitting formula has agreed within 5% with N-body data for most of the triangle configurations and always within 10% for the worst cases. This presents a considerable improvement over previous phenomenological approaches and over the prediction of Eulerian perturbation theory.
2. The equilateral and quasi-equilateral configurations have been the ones for which our model deviates most strongly from N-body data. We have interpreted this as being due to the fact that when the 3 sides of the triangle are similar, non-linearities start to play a role at the same time, and thus, the effect on the bispectrum is stronger than when the non-linearities enter at different times, i.e. for elongated triangles. Other methods, like the one described by Scoccimarro and Couchman (2001) have shown the same behavior.
3. We have checked that our model also works well for non-standard ΛCDM cosmologies. In particular, we have checked that for $k_2 = 2k_1$ the deviation between N-body data and our model is never higher than 3% and for equilateral triangles reaches 10%. Also in these non-standard ΛCDM cases studied here, our model works better than the Scoccimarro and Couchman (2001) fitting formula.

6.5 Future prospects

In this thesis we have presented several modelling techniques that can improve the current state-of-art of statistical cosmology: an improved halo modelling, a more realistic method for reducing the cosmic variance, the usage of the bispectrum for distinguishing gravity models and an improved fitting formula for the dark matter bispectrum in ΛCDM scenarios. Here we mention some possible further work along these lines.

6.5.1 The halo model

We expect that the extension of the halo model presented in this thesis can be used in the analysis of large-scale structure of the Universe, where accounting for the physical properties of galaxies, such as star formation rate, metallicity, age or colour. Current photometric and spectroscopic large-scale surveys have started to observe not only the position of galaxies, but also some of their physical properties. With future surveys the accuracy of these measurements will increase. We envision that the physical properties of galaxies depend on the environment, making these data sets the suitable ground to apply the extended halo model.

The extension of the halo model presented here uses analytic expressions for the mass function, halo density profile and HOD. We understand that these analytic functions may be inappropriate for comparison with real data, but they have to be thought more as a toy model rather than a model to fit the data. In the present thesis, we wanted to present an analytical expression for the dark matter non-linear power spectrum, that is fast and easy to compute and a reasonable approximation to reality. However, we were more interested in showing the relative effect of including environment dependence and less concerned with absolute accuracy of the fit. In a future work we may apply this technique to fit observational data; in this case we would have to further refine our prescription to compute $\xi_{\rm dm}$. There is still work in progress in this direction but we believe it is beyond the scope of the present thesis.

The environmental halo model presented in this thesis is especially suited to the marked (or weighed) correlation formalism (Harker et al., 2006, Skibba et al., 2006), consisting on weighting each galaxy according to a physical property and removing the dependence of clustering on local number-counts. Here, we have laid the foundations for modelling a survey's marked correlation, the treatment of which will be presented in a forthcoming work.

6.5.2 Cosmic variance reduction

We envision that the technique introduced by McDonald and Seljak (2009) and further developed in Chapter 3 will be very useful to reduce the effect sample variance in future galaxy surveys and missions such as BOSS or EUCLID. However, for an accurate quantification of the gain, the effects of non-linar bias of these tracers and also their possible stochasticity must be taken into account as we have shown in Chapter 3. We have demonstrated that the gain can be drastically reduced if the assumption of linear and deterministic biasing is dropped. However, a moderate improvement of the error bars is still possible.

In a future work we plan to combine the formalism developed here with real data from a large-scale galaxy survey such as BOSS. We know that dealing with galaxies instead of dark matter haloes, will produce an even more non-linear and stochastic bias, but if the signal-to-noise ratio is high enough we envision that a moderate gain in the measurements is achievable.

6.5.3 $f(R)$ theories & bispectrum

In Chapter 4 we have shown that the gravity component in the bispectrum of dark matter particles it is clearly sub-dominant and that most of the signal is dominated by the clustering of particles, namely the power spectrum. We have demonstrated that different theories of gravity with the same clustering pattern only present $\lesssim 4\%$ deviations in their bispectrum. This result must be taken into account in those studies that aim at measuring possible deviations from GR observing either galaxies or dark matter: only those surveys with extremely high signal-to-noise will be able to measure possible deviations from GR. Surveys covering volumes > 1 Gpc such as BOSS could in principle detect these differences.

An other important point for future projects is that the galaxy biasing depends very weakly on the gravity model assumed. Thus, those studies that aim to compute this bias from combined measurements of the power spectrum and bispectrum will obtain an almost gravity-independent result.

An interesting future work to do, is to predict the observed differences from a theoretical point of view. Some works (Bernardeau and Brax, 2011) have show than using cosmological perturbation theory and modifying the kernels in an suitable way could yield to a satisfactory answer.

6.5.4 The bispectrum fitting formula

We expect that the bispectrum fitting formula presented in Chapter 5 can be used in many studies related to lensing and forecasting errors of future dark matter surveys such as DES.

We also consider to improve the behaviour of the formula presented in this thesis to reach deviations below 5% respect to numerical simulations. One interesting approach is to explore the cosmological perturbation theory to higher orders and combine it with some analytic fitting part. It would be also interesting to extend the precision of this fitting formula to smaller scales $k > 0.5\,h/\mathrm{Mpc}$. Having precision tools that allows to predict analytically both power spectrum and bispectrum will be essential in the forthcoming years, in order to deal with the data from the future surveys and missions.

References

S. W. Allen, R. W. Schmidt, and A. C. Fabian. Cosmological constraints from the X-ray gas mass fraction in relaxed lensing clusters observed with Chandra. *MNRAS*, 334: L11–L15, August 2002. doi: 10.1046/j.1365-8711.2002.05601.x.

D. Babich, P. Creminelli, and M. Zaldarriaga. The shape of non-Gaussianities. *J. Cosmology Astropart. Phys.*, 8:9, August 2004. doi: 10.1088/1475-7516/2004/08/009.

R. A. Battye and J. Weller. Constraining cosmological parameters using Sunyaev-Zel'dovich cluster surveys. *Phys. Rev. D*, 68(8):083506, October 2003. doi: 10.1103/PhysRevD.68.083506.

J. Benjamin, C. Heymans, E. Semboloni, L. van Waerbeke, H. Hoekstra, T. Erben, M. D. Gladders, M. Hetterscheidt, Y. Mellier, and H. K. C. Yee. Cosmological constraints from the 100-deg^2 weak-lensing survey. *MNRAS*, 381:702–712, October 2007. doi: 10.1111/j.1365-2966.2007.12202.x.

F. Bernardeau and P. Brax. Cosmological large-scale structures beyond linear theory in modified gravity. *J. Cosmology Astropart. Phys.*, 6:19, June 2011. doi: 10.1088/1475-7516/2011/06/019.

A. Burkert and J. Silk. Dark Baryons and Rotation Curves. *ApJ*, 488:L55, October 1997. doi: 10.1086/310935.

S. M. Carroll. *Spacetime and geometry. An introduction to general relativity*. 2004.

D. Clowe, A. Gonzalez, and M. Markevitch. Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter. *ApJ*, 604:596–603, April 2004. doi: 10.1086/381970.

P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga. Limits on non-Gaussianities from WMAP data. *J. Cosmology Astropart. Phys.*, 5:4, May 2006. doi: 10.1088/1475-7516/2006/05/004.

W. J. G. de Blok, F. Walter, E. Brinks, C. Trachternach, S.-H. Oh, and R. C. Kennicutt, Jr. High-Resolution Rotation Curves and Galaxy Mass Models from THINGS. *AJ*, 136: 2648, December 2008. doi: 10.1088/0004-6256/136/6/2648.

A. Dekel and O. Lahav. Stochastic Nonlinear Galaxy Biasing. *ApJ*, 520:24–34, July 1999. doi: 10.1086/307428.

E. E. Falco, C. S. Kochanek, and J. A. Munoz. Limits on Cosmological Models from Radio-selected Gravitational Lenses. *ApJ*, 494:47, February 1998. doi: 10.1086/305207.

V. Faraoni. f(R) gravity: successes and challenges. *ArXiv e-prints*, October 2008.

S. Ferraro, F. Schmidt, and W. Hu. Cluster abundance in f(R) gravity models. *Phys. Rev. D*, 83(6):063503, March 2011. doi: 10.1103/PhysRevD.83.063503.

M. Fukugita, C. J. Hogan, and P. J. E. Peebles. The Cosmic Baryon Budget. *ApJ*, 503: 518, August 1998. doi: 10.1086/306025.

H. Gil-Marín, C. Wagner, L. Verde, R. Jimenez, and A. F. Heavens. Reducing sample variance: halo biasing, non-linearity and stochasticity. *MNRAS*, 407:772–790, September 2010. doi: 10.1111/j.1365-2966.2010.16958.x.

H. Gil-Marín, R. Jimenez, and L. Verde. A halo model with environment dependence: theoretical considerations. *MNRAS*, 414:1207–1226, June 2011a. doi: 10.1111/j.1365-2966.2011.18456.x.

H. Gil-Marín, F. Schmidt, W. Hu, R. Jimenez, and L. Verde. The bispectrum of f(R) cosmologies. *J. Cosmology Astropart. Phys.*, 11:19, November 2011b. doi: 10.1088/1475-7516/2011/11/019.

H. Gil-Marín, C. Wagner, F. Fragkoudi, R. Jimenez, and L. Verde. An improved fitting formula for the dark matter bispectrum. *J. Cosmology Astropart. Phys.*, 2:47, February 2012. doi: 10.1088/1475-7516/2012/02/047.

J. E. Gunn and J. R. Gott, III. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, 176:1, August 1972. doi: 10.1086/151605.

A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, January 1981. doi: 10.1103/PhysRevD.23.347.

A. Hajian and T. Souradeep. Measuring the Statistical Isotropy of the Cosmic Microwave Background Anisotropy. *ApJ*, 597:L5–L8, November 2003. doi: 10.1086/379757.

G. Harker, S. Cole, J. Helly, C. Frenk, and A. Jenkins. A marked correlation function analysis of halo formation times in the Millennium Simulation. *MNRAS*, 367:1039–1049, April 2006. doi: 10.1111/j.1365-2966.2006.10022.x.

A. F. Heavens, R. Jimenez, and R. Maartens. Testing homogeneity with the fossil record of galaxies. *J. Cosmology Astropart. Phys.*, 9:35, September 2011. doi: 10.1088/1475-7516/2011/09/035.

D. W. Hogg, D. J. Eisenstein, M. R. Blanton, N. A. Bahcall, J. Brinkmann, J. E. Gunn, and D. P. Schneider. Cosmic Homogeneity Demonstrated with Luminous Red Galaxies. *ApJ*, 624:54–58, May 2005. doi: 10.1086/429084.

W. Hu and I. Sawicki. Models of f(R) cosmic acceleration that evade solar system tests. *Phys. Rev. D*, 76(6):064004, September 2007. doi: 10.1103/PhysRevD.76.064004.

D. Huterer, E. Komatsu, and S. Shandera. Testing the Gaussianity and Statistical Isotropy of the Universe. *Advances in Astronomy*, 2010:697147, 2010. doi: 10.1155/2010/697147.

N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, September 1984. doi: 10.1086/184341.

E. Komatsu, N. Afshordi, N. Bartolo, D. Baumann, J. R. Bond, E. I. Buchbinder, C. T. Byrnes, X. Chen, D. J. H. Chung, A. Cooray, P. Creminelli, N. Dalal, O. Dore, R. Easther, A. V. Frolov, J. Khoury, W. H. Kinney, L. Kofman, K. Koyama, L. Leblond, J.-L. Lehners, J. E. Lidsey, M. Liguori, E. A. Lim, A. Linde, D. H. Lyth, J. Maldacena, S. Matarrese, L. McAllister, P. McDonald, S. Mukohyama, B. Ovrut, H. V. Peiris, A. Riotto, Y. Rodrigues, M. Sasaki, R. Scoccimarro, D. Seery, A. Sefusatti, K. M. Smith, A. A. Starobinsky, P. J. Steinhardt, F. Takahashi, M. Tegmark, A. J. Tolley, L. Verde, B. D. Wandelt, D. Wands, S. Weinberg, M. Wyman, A. P. S. Yadav, and M. Zaldarriaga. Non-Gaussianity as a Probe of the Physics of the Primordial Universe and the Astrophysics of the Low Redshift Universe. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, page 158, 2009.

M. Kowalski, D. Rubin, G. Aldering, R. J. Agostinho, A. Amadon, R. Amanullah, C. Balland, K. Barbary, G. Blanc, P. J. Challis, A. Conley, N. V. Connolly, R. Covarrubias, K. S. Dawson, S. E. Deustua, R. Ellis, S. Fabbro, V. Fadeyev, X. Fan, B. Farris, G. Folatelli, B. L. Frye, G. Garavini, E. L. Gates, L. Germany, G. Goldhaber, B. Goldman, A. Goobar, D. E. Groom, J. Haissinski, D. Hardin, I. Hook, S. Kent, A. G. Kim, R. A. Knop, C. Lidman, E. V. Linder, J. Mendez, J. Meyers, G. J. Miller, M. Moniez, A. M. Mourão, H. Newberg, S. Nobili, P. E. Nugent, R. Pain, O. Perdereau, S. Perlmutter,

M. M. Phillips, V. Prasad, R. Quimby, N. Regnault, J. Rich, E. P. Rubenstein, P. Ruiz-Lapuente, F. D. Santos, B. E. Schaefer, R. A. Schommer, R. C. Smith, A. M. Soderberg, A. L. Spadafora, L.-G. Strolger, M. Strovink, N. B. Suntzeff, N. Suzuki, R. C. Thomas, N. A. Walton, L. Wang, W. M. Wood-Vasey, J. L. Yun, and Supernova Cosmology Project. Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets. *ApJ*, 686:749–778, October 2008. doi: 10.1086/589937.

D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *ApJS*, 192:16, February 2011. doi: 10.1088/0067-0049/192/2/16.

Y. Li and W. Hu. Chameleon halo modeling in f(R) gravity. *Phys. Rev. D*, 84(8):084033, October 2011. doi: 10.1103/PhysRevD.84.084033.

M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56. *ApJ*, 606:819–824, May 2004. doi: 10.1086/383178.

P. McDonald and U. Seljak. How to evade the sample variance limit on measurements of redshift-space distortions. *J. Cosmology Astropart. Phys.*, 10:7, October 2009. doi: 10.1088/1475-7516/2009/10/007.

H. Oyaizu. Nonlinear evolution of f(R) cosmologies. I. Methodology. *Phys. Rev. D*, 78 (12):123523, December 2008. doi: 10.1103/PhysRevD.78.123523.

H. Oyaizu, M. Lima, and W. Hu. Nonlinear evolution of f(R) cosmologies. II. Power spectrum. *Phys. Rev. D*, 78(12):123524, December 2008. doi: 10.1103/PhysRevD.78.123524.

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

W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *MNRAS*, 401:2148–2168, February 2010. doi: 10.1111/j.1365-2966.2009.15812.x.

S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999. doi: 10.1086/307221.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038, September 1998. doi: 10.1086/300499.

K. Rines and A. Diaferio. CIRS: Cluster Infall Regions in the Sloan Digital Sky Survey. I. Infall Patterns and Mass Profiles. *AJ*, 132:1275–1297, September 2006. doi: 10.1086/506017.

K. Rines, M. J. Geller, M. J. Kurtz, and A. Diaferio. CAIRNS: The Cluster and Infall Region Nearby Survey. I. Redshifts and Mass Profiles. *AJ*, 126:2152–2170, November 2003. doi: 10.1086/378599.

B. Ryden. *Introduction to cosmology*. 2003.

F. Schmidt, M. Lima, H. Oyaizu, and W. Hu. Nonlinear evolution of f(R) cosmologies. III. Halo statistics. *Phys. Rev. D*, 79(8):083518, April 2009. doi: 10.1103/PhysRevD.79.083518.

R. Scoccimarro and H. M. P. Couchman. A fitting formula for the non-linear evolution of the bispectrum. *MNRAS*, 325:1312–1316, August 2001. doi: 10.1046/j.1365-8711.2001.04281.x.

R. Skibba, R. K. Sheth, A. J. Connolly, and R. Scranton. The luminosity-weighted or 'marked' correlation function. *MNRAS*, 369:68–76, June 2006. doi: 10.1111/j.1365-2966.2006.10196.x.

T. P. Sotiriou and V. Faraoni. f(R) theories of gravity. *Reviews of Modern Physics*, 82: 451–497, January 2010. doi: 10.1103/RevModPhys.82.451.

D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L.

Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *ApJS*, 148: 175–194, September 2003. doi: 10.1086/377226.

N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, K. S. Dawson, A. Dey, M. Doi, M. Donahue, S. Deustua, P. Eisenhardt, E. Ellingson, L. Faccioli, V. Fadeyev, H. K. Fakhouri, A. S. Fruchter, D. G. Gilbank, M. D. Gladders, G. Goldhaber, A. H. Gonzalez, A. Goobar, A. Gude, T. Hattori, H. Hoekstra, E. Hsiao, X. Huang, Y. Ihara, M. J. Jee, D. Johnston, N. Kashikawa, B. Koester, K. Konishi, M. Kowalski, E. V. Linder, L. Lubin, J. Melbourne, J. Meyers, T. Morokuma, F. Munshi, C. Mullis, T. Oda, N. Panagia, S. Perlmutter, M. Postman, T. Pritchard, J. Rhodes, P. Ripoche, P. Rosati, D. J. Schlegel, A. Spadafora, S. A. Stanford, V. Stanishev, D. Stern, M. Strovink, N. Takanashi, K. Tokita, M. Wagner, L. Wang, N. Yasuda, H. K. C. Yee, and T. Supernova Cosmology Project. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above z>1 and Building an Early-type-hosted Supernova Sample. *ApJ*, 746:85, February 2012. doi: 10.1088/0004-637X/746/1/85.

P. T. P. Viana. Cosmological Parameter Estimation with the Galaxy Cluster Abundance. *Ap&SS*, 290:149–158, February 2004. doi: 10.1023/B:ASTR.0000022170.97308.be.

A. Vikhlinin, A. V. Kravtsov, R. A. Burenin, H. Ebeling, W. R. Forman, A. Hornstrup, C. Jones, S. S. Murray, D. Nagai, H. Quintana, and A. Voevodkin. Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints. *ApJ*, 692:1060–1074, February 2009. doi: 10.1088/0004-637X/692/2/1060.

R. M. Wald. *General relativity*. 1984.

S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. July 1972.

C. Zunckel, D. Huterer, and G. D. Starkman. Testing the statistical isotropy of large scale structure with multipole vectors. *Phys. Rev. D*, 84(4):043005, August 2011. doi: 10.1103/PhysRevD.84.043005.