Durham University

Durham E-Theses

Mock Catalogues for Large Scale Structure Surveys and DESI

ALEXANDER MARK JOSEPH SMITH

How to cite:

SMITH, ALEXANDER MARK JOSEPH (2018) Mock Catalogues for Large Scale Structure Surveys and DESI. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a https://etheses.durham.ac.uk/id/eprint/12866/ is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

Mock Catalogues for Large Scale Structure Surveys and DESI

Alexander Smith

A thesis presented for the degree of

Doctor of Philosophy

Institute for Computational Cosmology

The University of Durham

United Kingdom

November 2018

Mock Catalogues for Large Scale Structure Surveys and DESI

Alexander Smith

Abstract

The upcoming Dark Energy Spectroscopic Instrument (DESI) and Euclid galaxy surveys aim to make the most precise galaxy clustering measurements yet, in order to probe the nature of the mysterious dark energy that is thought to make up the majority of the energy density of the Universe today. To reach the required precision, it is essential that the systematics that affect these measurements are understood, which requires realistic mock galaxy catalogues. This thesis focuses on building a mock catalogue for the DESI Bright Galaxy Survey (BGS), and applications of this mock. We outline the methods used to create halo merger trees from N-body and Monte Carlo simulations, which is the first step towards creating a mock catalogue. We show how these methods can be extended beyond ΛCDM to warm dark matter, and show applications. We have developed a halo occupation distribution (HOD) method for creating a BGS mock catalogue from the Millennium-XXL (MXXL) simulation, with galaxies being assigned r-band magnitudes and $g-r$ colours. The mock catalogue is able to reproduce the luminosity function and clustering of the Sloan Digital Sky Survey (SDSS) and Galaxy And Mass Assembly (GAMA) survey at different redshifts. The mock is used to quantify incompleteness in the DESI BGS due to fibre assignment, which depends on the surface density of galaxies, and to assess correlation function correction methods. An inverse pair weighting method is able to provide an unbiased correction on all scales. Finally, we show how the HOD methodology can be extended to construct mock catalogues for Euclid, and other large galaxy surveys.

Contents

List of Figures

List of Tables

Declaration

The work in this thesis is based on research carried out by the author between 2014 and 2018 while the author was a research student under the supervision of Prof. Shaun Cole, and Prof. Carlton Baugh in the Department of for any other degree or qualification.

Section 2.4.3 of Chapter 2 has been published as part of a paper:

- Lovell M. R., Bose, S., Boyarsky, A., Cole, S., Frenk, C. S., Gonzalez-Perez, V., Kennedy, R., Ruchayskiy, O., Smith, A., "Satellite galaxies in semi-analytic models of galaxy formation with sterile neutrino dark matter", 2016, MNRAS, **461**, pp. 60-72

The results of this paper are summarised in Section 2.5.

Chapter 3 has been published in the form of a paper:

- Smith, A., Cole, S., Baugh, C., Zheng, Z., Angulo, R., Norberg, P., Zehavi, I., "A lightcone catalogue from the Millennium-XXL simulation", 2017, MNRAS, **470**, pp. 4646-4661

The majority of Chapter 4 has been submitted in the form of a paper:

- Smith, A., He, J., Cole, S., Stothert, L., Norberg, P., Baugh, C., Bianchi, D., Wilson, M. J., Brooks, D., Forero-Romero, J. E., Moustakas, J., Percival, W. J., Tarle, G., Wechsler, R. H., "Correcting for Fibre Assignment Incompleteness in the DESI Bright Galaxy Survey", 2018

Figures 4.2, 4.3 and 4.6 of Chapter 4 are to be submitted as part of a paper:

- DESI Collaboration, "The DESI Bright Galaxy Survey", 2019

The GALACTICUS Hα HODs used in Chapter 5 were measured by Alex Merson.

Copyright © 2018 by Alexander Smith.

"*The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged*".

Acknowledgements

Firstly, I would like to thank my brilliant supervisors Shaun and Carlton for all their help, guidance and support over the past 4 years. Thank you for imparting your knowledge, helping me to develop as a researcher, and getting me through a stressful few months of job applications! I would also like to thank all my other collaborators, particularly Peder and Jianhua. I'm also grateful to Lydia, John and Alan for their computing expertise.

I would like to acknowledge STFC for providing the funding that made my PhD possible, and DESI UK for enabling me to travel to the DESI meeting in Ohio.

I've had the pleasure to share an office with lots of great people over the past few years. Thank you to Charles, Jacob, James T., Johannes, Peter, Stefan (when he wasn't asleep at his desk) and Steve in PH305 for making the 'shanty town' bearable over the hot summer months. I enjoyed my time in OCW131 with Andrew R., Stefan (again) and Oliver, being distracted by their random puzzles, and finally with Lee back over in PH319, on the rare occasion he was in the office. I am grateful to all the other people in the Physics department who made my time in Durham so enjoyable. There are too many people to name, but in particular I would like to thank Ruari, Stu, Mark, Jaime, Ben, Flora, James C., Rose, and of course my housemates at Dalton Crescent, Amrit and David. I will miss having our regular film and pizza nights. We watched some truly terrible films, but nothing could beat *Left Behind*, starring Nicolas Cage.

Special thanks to my housemates in first year at Wynyard Grove: Stephen, Jess, Peter, Sarah and Elodie. We had a lot of fun making house meals and doing pub quizzes.

I would also like to thank Hamish, my school Physics teacher for inspiring me to study Physics, and Chris, my master's supervisor for all her help back when I

was applying for PhDs.

Finally, I would like to thank my family for always supporting me.

Dedicated to Grandma

Chapter 1

Introduction

1.1 The ΛCDM model of cosmology

Our understanding of cosmology has changed dramatically over the past century. In the 1920s, Hubble observed that nearby galaxies are receding away from us, with a velocity that is, on average, proportional to their distance from us (Hubble, 1929). This was the first observational evidence for an expanding Universe, and is described by Hubble's law,

$$v = H_0 d, \tag{1.1}$$

where v is the recession velocity of a galaxy, d is its distance, and H_0 is the Hubble constant.

In the 1930s, Zwicky measured the velocity dispersion of galaxies within the Coma cluster in order to estimate its mass. The cluster was found to be much more massive than expected from the total luminosity. Zwicky suggested that this discrepancy was due to a mass component within the cluster that does not emit light, or 'dark matter' (Zwicky, 1933).

Another piece of evidence for dark matter came later in the 20th Century from observations of the rotation curves of galaxies (the rotational velocity of stars in the galaxy as a function of the radial distance from the centre). If all the mass in galaxies is made up of visible stars, which are mostly concentrated towards the

centre of the galaxy, then the stars should follow Keplerian dynamics, and the rotational velocity should fall with increasing distance ($v \propto r^{-1/2}$, e.g. as is seen in the planets of the Solar System). However, galaxy rotation curves are observed to be approximately flat in the outskirts of galaxies, and the rotational velocity does not depend on distance (Rubin et al., 1980). If Newtonian dynamics is correct, there must be extra mass inside the galaxy which cannot be accounted for by visible stars and gas.

In the 1990s, observations were made of Type Ia supernovae in distant galaxies, which can be used as standard candles. It was found that distant supernovae were fainter and therefore farther away than expected (Riess et al., 1998; Perlmutter et al., 1999). This was the first evidence that the expansion rate of the Universe is accelerating, driven by a mysterious additional energy density component of the Universe, or 'dark energy', which can be described as the cosmological constant, Λ.[1]

The current standard model of cosmology is the ΛCDM model, where Λ is the cosmological constant, and dark matter is in the form of cold dark matter (CDM). Each galaxy is embedded within a halo of dark matter, which is believed to be comprised of a massive particle (with mass of the order of a few GeV) that is 'cold' (i.e. at early times had a negligible thermal velocity), and only interacts with regular matter via gravity. While dark matter and dark energy are poorly understood, current observations of the Universe are consistent with the ΛCDM model (e.g. Planck Collaboration et al., 2018).

[1] There was motivation for introducing Λ before the supernova measurements (Efstathiou et al., 1990b). A low value of $\Omega_{\rm m}h \approx 0.2$ is needed to account for the measured large-scale galaxy clustering. A positive cosmological constant is therefore required for the Universe to be flat, as predicted by inflation.

1.2 The expansion history of the Universe

In the current cosmological model, the Universe is 13.8 billion years old, and was initially extremely hot and dense, expanding rapidly after the 'Big Bang'. Shortly afterwards, the Universe went through 'inflation', a period of exponential expansion. During this period, primordial fluctuations are generated, which are the seeds of the large scale structure we see today. As the Universe expands and cools, it reaches an epoch during which it is ionized, and photons are coupled with baryons. During this period, acoustic oscillations are able to propagate through the plasma until recombination, when the Universe becomes neutral and photons and baryons decouple. This happens at redshift $z \sim 1100$, or around 400,000 years after the Big Bang, when the Universe has cooled to a temperature of $\sim 3,000$ K, and the Universe becomes transparent to photons. The photons from the last scattering surface are redshifted as the Universe expands, and are observed today as the cosmic microwave background (CMB), with a temperature ~ 3 K (Mather et al., 1990). The imprint of the baryon acoustic oscillations (BAO) can be measured in the temperature anisotropies of the CMB (e.g. Hinshaw et al., 2009), and also in measurements of galaxy clustering in the low redshift Universe (e.g. Eisenstein et al., 2005b). After recombination, the Universe went through a period in which it was dominated by matter. Overdensities of matter collapse to form a cosmic web of filaments, haloes and voids. Galaxies form within dark matter haloes, and the radiation they produce reionizes the Universe. Relatively recently, as the matter density has been reduced as the Universe expands, the energy density has become dominated by dark energy, which is driving the current accelerated expansion.

Starting with the Einstein field equations of General Relativity, and assuming that the Universe is isotropic and homogeneous, the expansion of the Universe can be described using the Friedmann equations (e.g. Mo et al., 2010),

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.2}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R_0^2 a^2} + \frac{\Lambda c^2}{3}, \tag{1.3}$$

where a is the dimensionless 'scale factor', which parametrises the relative expansion of the Universe, and at the present day, t_0, is chosen to have the value $a_0 \equiv a(t_0) = 1$. The mass content of the Universe can be described as a fluid with mean density ρ and pressure p. G is the gravitational constant and c is the speed of light in a vacuum. The parameter k describes the curvature of the Universe, where $k = +1, 0, -1$, depending on whether the Universe is closed, flat, or open, and R_0 is the radius of curvature. Finally, Λ is the cosmological constant. A non-zero, positive value of Λ is required for an accelerating expansion. The Hubble parameter can be defined in terms of a as $H(a) \equiv \dot{a}/a$, where $H_0 \equiv H(a_0) \equiv 100h$ km s^{-1}Mpc^{-1} ≈ 70 km s^{-1}Mpc^{-1} is the present day value.[1]

The critical density can be defined as

$$\rho_{\rm crit} = \frac{3H_0^2}{8\pi G}, \tag{1.4}$$

which is the density required to slow the expansion of the Universe to zero as $t \to \infty$ (in a Universe with no dark energy). Each component of the Universe can be expressed relative to the critical density. For mass,

$$\Omega_{\rm m} = \frac{\rho_{\rm m}}{\rho_{\rm crit}}. \tag{1.5}$$

For radiation (i.e. photons and massless particles), $\Omega_{\rm r}$ can be defined in the same way. Similar quantities can be defined for curvature,

$$\Omega_k = -\frac{kc^2}{R_0^2 H_0^2}, \tag{1.6}$$

and for the cosmological constant,

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

[1] The most recent measurement of H_0 from the Planck satellite is $H_0 = 67.4 \pm 0.5$ km s^{-1}Mpc^{-1} (Planck Collaboration et al., 2018). This is in tension with measurements of H_0 based on the cosmological distance ladder, e.g. Riess et al. (2018) measure $H_0 = 73.5 \pm 1.6$ km s^{-1}Mpc^{-1}.

and by construction

$$\Omega_{\rm m} + \Omega_{\rm r} + \Omega_k + \Omega_\Lambda = 1. \tag{1.8}$$

The expansion history can therefore be written as

$$H^2(a) = H_0^2(\Omega_{\rm m} a^{-3} + \Omega_{\rm r} a^{-4} + \Omega_k a^{-2} + \Omega_\Lambda). \tag{1.9}$$

Measurements of the anisotropies in the CMB can be used to measure the present day contribution of each of these components to the total energy density of the Universe. The most recent results from the Planck satellite are $\Omega_{\rm m} = 0.315 \pm 0.007$ and $\Omega_\Lambda = 0.685 \pm 0.007$. The curvature of the Universe is measured to be $\Omega_k = 0.001 \pm 0.002$, which is consistent with a flat Universe (Planck Collaboration et al., 2018).

For each component of the energy density of the Universe, density and pressure are related through the equation of state $w = p/\rho c^2$. For ordinary matter which is non-relativistic, $w = 0$ since the pressure is negligible, while for photons $w = 1/3$. For the expansion of the Universe to accelerate, $\rho c^2 + 3p < 0$ (from Eq. 1.2), so the total equation of state must be $w < -1/3$. In ΛCDM, the cosmological constant is equivalent to dark energy with a constant equation of state $w = -1$. However, dark energy does not have to be constant. The equation of state could change with time, and this is commonly parametrised as

$$w(a) = w_0 + (1 - a)w_a, \tag{1.10}$$

which is independent of any particular model (Chevallier & Polarski, 2001; Linder, 2003). Alternatively, the accelerated expansion could be driven by a modification to General Relativity (GR) on large scales (Amendola et al., 2018). Large galaxy surveys aim to distinguish between these possibilities. To date, measurements from surveys have been consistent with GR (e.g. Zarrouk et al., 2018; Zhao et al., 2018), and have found w to be consistent with a constant value of -1 (e.g. Cuesta et al., 2016; DES Collaboration et al., 2017). Future surveys are driven to be even larger in order to obtain the precision measurements required to place even tighter

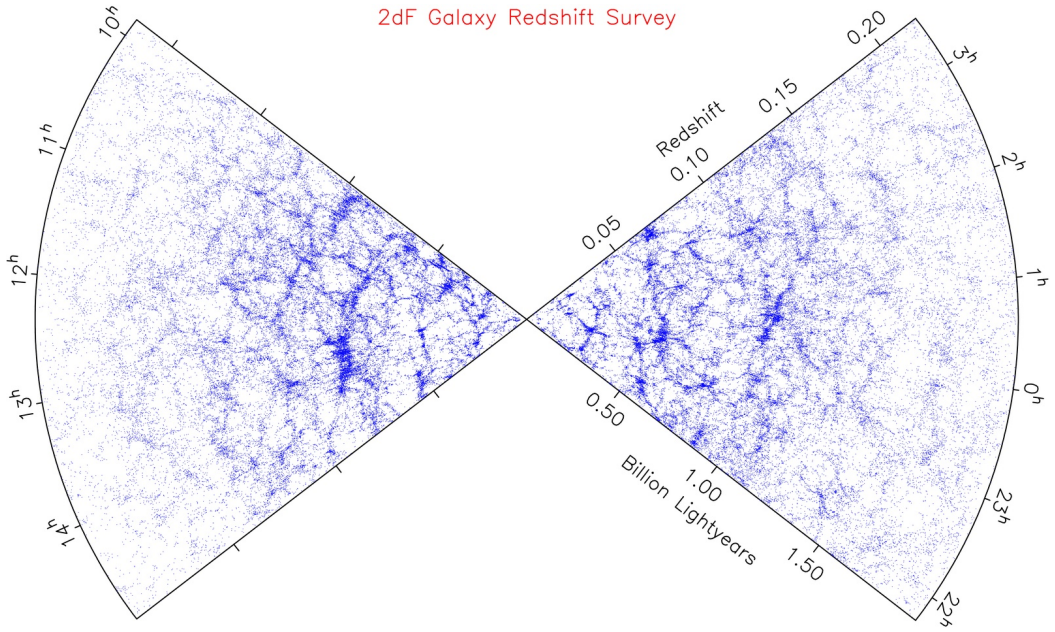

Figure 1.1: A thin slice through the 2dF survey, illustrating the large scale structure that is probed by large galaxy surveys. Figure reproduced from Colless et al. (2003).

constraints, which are needed to address the fundamental question of the nature of dark energy.

1.3 Large-scale structure surveys

Large galaxy surveys can be used to test the ΛCDM paradigm. The aim of a galaxy survey is to measure the positions, redshifts, and other properties of many hundreds of thousands or millions of galaxies, in order to create a 3D map of the large-scale structure. Predictions from ΛCDM can be compared with statistics measured from the survey, and constraints can be placed on theories beyond ΛCDM.

One of the earliest surveys was the CfA Redshift Survey (Huchra et al., 1983), which took 5 years to measure the spectra of 2,400 galaxies individually. Since then, advances in instrumentation have enabled larger and deeper surveys, with multiple objects observed simultaneously. The Two-degree-Field Galaxy Redshift

Survey (2dF) (Colless et al., 2001) obtained $\sim 250,000$ spectra, covering a total area on the sky of $1,500$ deg^2, with a median redshift ~ 0.1. A robotic multi-fibre spectrograph was used to observe 400 objects simultaneously, within a pointing of diameter 2 degrees. Fig. 1.1 shows a thin slice through the 2dF survey. The galaxies in the survey clearly trace out the filamentary large-scale structure of matter. The Sloan Digital Sky Survey (SDSS) (York et al., 2000; Abazajian et al., 2009), measured 1 million galaxy redshifts with a similar median redshift, but covering a much larger area of $\sim 10,000$ deg^2, for objects brighter than an r-band magnitude of $r = 17.7$. Each SDSS tile has a diameter of 3 degrees with 640 fibres. Before each observation, each fibre has to be, by hand, plugged into the location of each galaxy on the plate. Other surveys cover a smaller area on the sky, but are much deeper, such as the GAMA Survey (Driver et al., 2009, 2011; Liske et al., 2015), which covers 286 deg^2 with the magnitude limit $r < 19.8$, and median redshift $z_{\rm med} = 0.2$. Deeper surveys, such as GAMA, which cover a wide range of redshifts are useful for studying galaxy formation and evolution.

To map the large-scale structure at high redshift over large areas of the sky, surveys must target specific tracers, such as Luminous Red Galaxies (LRGs), Emission Line Galaxies (ELGs), and quasars, in order to reduce the total surface density of targets. The BOSS survey (Eisenstein et al., 2011; Dawson et al., 2013), which is an extension of SDSS, measured the spectra of over 1.5 million LRGs ($z < 0.7$), and the Lyman-α forest of 160,000 quasars ($2.2 < z < 3$). The BOSS LRG sample is split, using colour and magnitude cuts, into the LOWZ sample ($z \lesssim 0.4$), and the CMASS sample of massive galaxies with constant stellar mass ($0.4 < z < 0.7$). The ongoing eBOSS survey (Dawson et al., 2016) aims to fill in the intermediate redshifts by targeting 300,000 LRGs ($0.6 < z < 0.8$), 189,000 ELGs ($0.6 < z < 1.0$), and 573,000 quasars ($0.9 < z < 3.5$). The Dark Energy Survey (DES) (The Dark Energy Survey Collaboration, 2005; Dark Energy Survey Collaboration et al., 2016) is an ongoing photometric survey which aims to image 300 million objects in 5 photometric bands over $5,000$ deg^2.

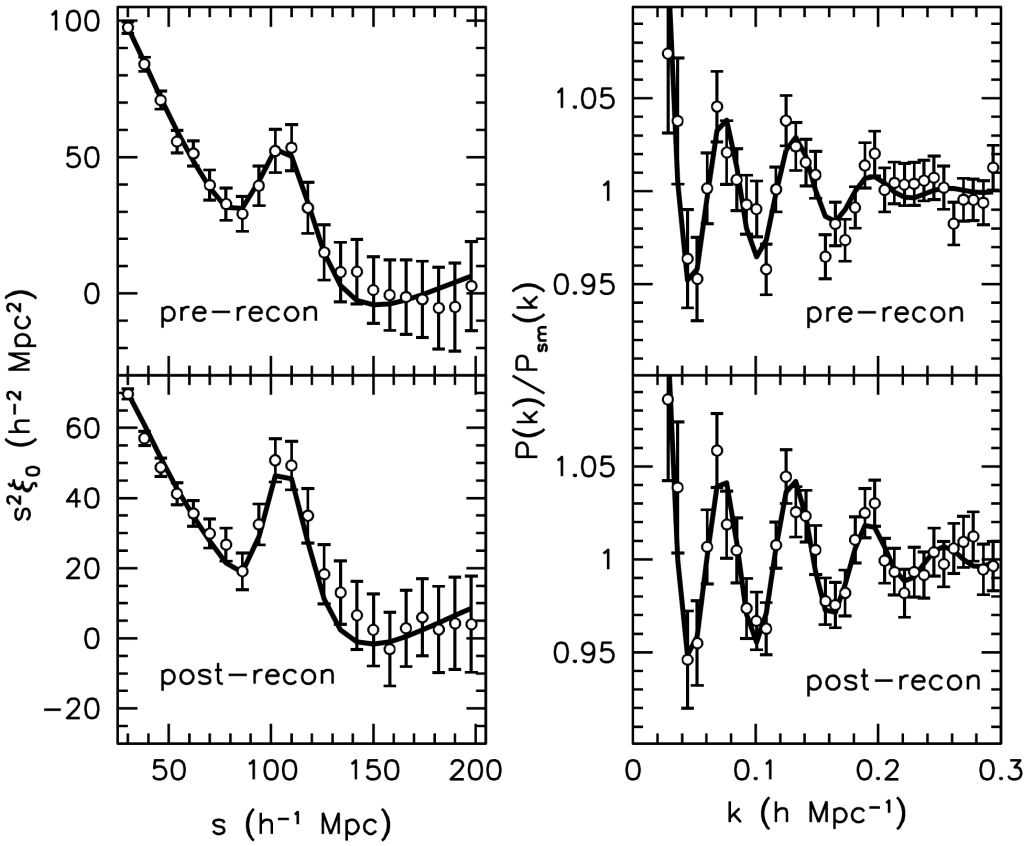

Figure 1.2: Measurement of the BAO from the CMASS sample of galaxies in the BOSS survey ($0.4 < z < 0.7$) in the two-point correlation function (left) and the power spectrum (right), before and after reconstruction. Figure reproduced from Anderson et al. (2014a).

Measurements of how galaxies are clustered in 3D space from these large galaxy surveys at different redshifts can be used to measure the expansion history of the Universe, and the growth of structure. These measurements include baryon acoustic oscillations, which can be used as a standard ruler to measure the expansion history of the Universe, and redshift space distortions, which can be used to measure the growth of structure and test general relativity.

1.3.1 Baryon acoustic oscillations as a standard ruler

During the period before recombination, photons are in thermal equilibrium with electrons and protons, and the Universe is opaque to photons, due to Thomson scattering. Overdense regions attract matter towards them, which compresses and heats up the plasma, resulting in an outward radiation pressure. This produces acoustic waves, which are able to propagate at a speed $c/\sqrt{3}$, until recombination. During this time, the waves are able to propagate a comoving distance of the order of 150 Mpc ($\sim 100\ h^{-1}$Mpc), leaving behind an excess of matter, and the imprint of the BAO can be seen as oscillations in the power spectrum of CMB anisotropies. The overdense regions collapse to form galaxies, and there is a small enhancement (of a few percent) of galaxies at the BAO separation, which can be measured in the two-point correlation function of galaxies.

The density field of matter at point $\mathbf{x}$ is given by $\delta(\mathbf{x}) = (\rho(\mathbf{x}) - \bar{\rho})/\bar{\rho}$, where $\bar{\rho}$ is the average density. The two-point correlation function is defined as the auto-correlation function of the density field at two points separated by $\mathbf{r}$,

$$\xi(\mathbf{r}) \equiv \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle \,. \tag{1.11}$$

The two-point correlation function can also be thought of as the excess probability of finding two objects with separation r, compared to a random distribution,

$$dP = n_0^2[1 + \xi(r)]dV_1 dV_2, \tag{1.12}$$

where n_0 is the average number density of objects and dV_1 and dV_2 are volume elements. Fig. 1.2 shows measurements of the BAO in the two-point correlation function from the BOSS survey. A peak can be seen, indicating that there is an enhancement of galaxy pairs separated by the BAO length scale of $\sim 100\ h^{-1}$Mpc. The displacement of galaxies from their initial positions, due to bulk flows and non-linear structure formation, dampens and broadens the BAO peak. A method called reconstruction aims to correct for this by estimating the displacement field using Lagrangian Perturbation Theory, and then moving galaxies back to their

original position (Eisenstein et al., 2007). The BAO can also be seen as oscillations in the power spectrum, which is the Fourier transform of the two-point correlation function.

Since the waves travel a fixed distance before recombination, the BAO can be used as a standard ruler to measure the expansion history of the Universe. The comoving BAO length scale is related to an angle on the sky θ,

$$s_{\rm BAO} = (1+z) d_A(z) \theta, \tag{1.13}$$

where $d_A(z)$ is the angular diameter distance to redshift z. In a flat Universe, this can be written as

$$s_{\rm BAO} = \theta \int_0^z \frac{c dz'}{H(z')}. \tag{1.14}$$

For pairs of galaxies along the line of sight, there will be a redshift separation Δz which corresponds to the BAO scale,

$$s_{\rm BAO} = \frac{c\Delta z}{H(z)}. \tag{1.15}$$

Therefore BAO measurements from large galaxy surveys can be used to measure both the expansion rate of the Universe, $H(z)$ at different redshifts, and also the angular diameter distance $d_A(z)$, which is the integrated expansion history between redshift z and the present.

1.3.2 Using redshift space distortions to measure the growth of structure

The distance to each galaxy in a survey can be inferred from its measured redshift. However, this gives a distorted view of the distribution of galaxies, as the observed redshift, $z_{\rm obs}$, is altered by peculiar motion,

$$z_{\rm obs} = z_{\rm cos} + \frac{v_r}{c}, \tag{1.16}$$

where $z_{\rm cos}$ is the cosmological redshift from Hubble's law, and v_r is the peculiar velocity along the line of sight.

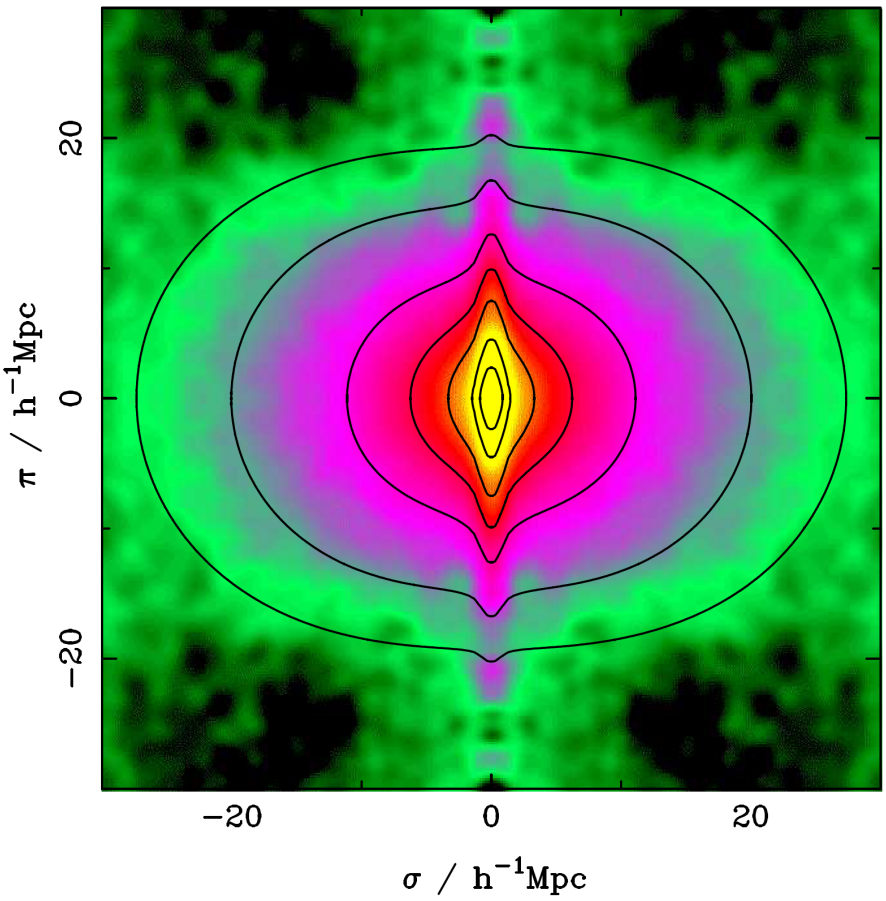

Figure 1.3: Redshift space two-point correlation function, $\xi(\sigma, \pi)$ measured from the 2dF survey, as a function of σ and π, the pair separation perpendicular to, and along the line of sight respectively. The black contours are model predictions. Figure reproduced from Peacock et al. (2001).

Consider a spherical galaxy cluster. On large scales, there is a coherent infall of galaxies towards the centre of the cluster. Galaxies on the near-side of the cluster are falling away from the observer, which increases their observed redshift. Conversely, galaxies on the far-side are falling in the direction towards the observer, and will therefore have slightly lower observed redshifts. This is the Kaiser effect (Kaiser, 1987), and results in an apparent flattening of the cluster in redshift space. On small scales, the large random velocities results in an elongation in redshift space, also known as a Fingers-of-God distortion (Jackson, 1972).

These effects can also be seen as anisotropies in the two-point correlation function $\xi(\sigma, \pi)$, where σ is the distance perpendicular to the line of sight, and π is the distance parallel to the line of sight. This is illustrated in Fig. 1.3, which shows the redshift space correlation function measured in the 2dF survey (Peacock et al., 2001). In real space, the clustering is isotropic, so contours of $\xi(\sigma, \pi)$ are circular. In redshift space, the contours are compressed on large scales, due to the Kaiser effect, while the contours on small scales are elongated along the line of sight, due to Fingers-of-God distortions.

Galaxies are biased tracers of the matter density field, $\delta_m(\mathbf{x})$, and the overdensity in galaxies can be written as $\delta_g(\mathbf{x}) = b\delta_m(\mathbf{x})$, where b is the bias factor. In Fourier space, the redshift space perturbation is related to the real space perturbation

$$\delta_g^{(s)}(\mathbf{k}) = (1 + \beta\mu^2)\delta_g(\mathbf{k}), \tag{1.17}$$

where μ is the cosine of the angle between $\mathbf{k}$ and the line of sight (Kaiser, 1987). The quantity $\beta = f/b$, where the growth rate f is the logarithmic derivative of the growth function $D(a)$,

$$f = \frac{d\ln D(a)}{d\ln a}. \tag{1.18}$$

Like Eq. 1.11, the redshift space correlation function can be defined as

$$\xi_g^{(s)}(\mathbf{s_1}, \mathbf{s_2}) = \langle \delta_g^{(s)}(\mathbf{s_1})\delta_g^{(s)}(\mathbf{s_2})\rangle, \tag{1.19}$$

so measurements of the redshift space correlation function provide a way to measure f.

The power spectrum of the density field is defined as

$$P(k) \equiv \langle|\delta(\mathbf{k})|^2\rangle. \tag{1.20}$$

It can be shown that $P(k)$ is the Fourier transform of the correlation function, and in an isotropic Universe, this can be written as

$$P(k) = 4\pi \int_0^\infty r^2\xi(r)\frac{\sin(kr)}{kr}dr. \tag{1.21}$$

The power spectrum in redshift space is

$$P_g^{(s)}(\mathbf{k}) = (b + f\mu^2)^2 P_m(k), \tag{1.22}$$

and the normalisation of the power spectrum $P_m(k)$ is proportional to $\sigma_8^2(z)$[1], so in reality, RSD measurements place constraints on the combination $f(z)\sigma_8(z)$.

In general relativity, $f \approx \Omega_{\rm m}(z)^\gamma$, where $\gamma \approx 0.55$ (e.g. Polarski & Gannouji, 2008). In modified theories of gravity, γ could differ from this value, so constraints in the $\Omega_{\rm m}$-γ plane can rule out modified gravity theories (Guzzo et al., 2008).

1.3.3 The next generation of surveys

The results of BAO and RSD analysis in galaxy surveys have, to date, been consistent with a flat ΛCDM Universe in which gravity can be described using General Relativity (e.g. Howlett et al., 2015; Ross et al., 2015; Cuesta et al., 2016; Alam et al., 2017; DES Collaboration et al., 2017; Ruggeri et al., 2018; Zarrouk et al., 2018; Zhao et al., 2018). The differences between ΛCDM and models that have not been ruled are becoming increasingly small. To distinguish between them, even more accurate measurements are needed, which requires even larger surveys.

The Dark Energy Spectroscopic Instrument (DESI) survey (DESI Collaboration et al., 2016a,b) is an upcoming survey, which aims to measure spectra of 4 million LRGs ($0.4 < z < 1.0$), 17 million ELGs ($0.6 < z < 1.6$), 1.7 million quasars ($z < 2.1$), 0.7 million higher redshift quasars ($2.1 < z < 3.5$), and ~ 10 million bright, low redshift galaxies ($r < 19.5$ at $z_{\rm med} = 0.2$). The instrument is being installed on the 4-m Mayall Telescope in Arizona, and the 5 year survey, which is being scheduled to begin at the end of 2019, will cover $\sim 14,000$ deg^2, where 5,000 objects will be able to have their spectra measured simultaneously. Unlike SDSS, where fibres are plugged by hand into each plate, DESI utilises robotic fibre positioners in the focal plane of the telescope, which can automatically place the fibres onto the position of each target galaxy in each pointing.

[1] σ_8^2 is the variance in the mass density field in spheres of radius 8 h^{-1}Mpc.

The Euclid satellite (Laureijs et al., 2011) will conduct another large survey, and is scheduled for launch in 2021. The satellite will be placed at the L2 Lagrangian point, and the 6 year mission will cover $15,000$ deg^2, measuring the spectra of 30 million ELGs ($0.7 < z < 2$), with a near-infrared slitless spectrometer. The instruments on the satellite are the visible (VIS) instrument, which will image the shapes of galaxies for studying weak lensing, and the Near Infrared Spectrometer and Photometer (NISP), which will measure near infrared photometric and spectroscopic redshifts.

1.4 Mock catalogues from cosmological simulations

In order to prepare for these upcoming large galaxy surveys, realistic synthetic galaxy catalogues are needed, and can be utilised for a variety of reasons. Firstly, mock catalogues are needed in order to develop and test the survey pipeline code. They are useful to help design the survey, and to explore different survey strategies. Mocks can also be used to test methods for measuring cosmological parameters, and to understand the systematics that will affect these measurements. Many mock catalogues are needed to estimate accurate covariance matrices of galaxy clustering and power spectrum measurements. Accurate covariance matrices are needed to obtain the uncertainty in measurements of cosmological parameters (Percival et al., 2014).

On large scales, the growth of structure is linear, and is well understood. However, in the dense regions where galaxies form, structure growth is highly non-linear, which makes it very difficult to describe analytically. The non-linear growth of structure can be simulated using numerical techniques. In an N-body simulation, such as the Millennium Simulation (Springel et al., 2005) and the Millennium-XXL Simulation (Angulo et al., 2012b), the matter distribution is represented by a set of dark matter particles. These particles are evolved from some initial conditions at high redshift within a large cosmological volume to the present day. The dark

matter particles collapse to form dark matter haloes, which merge, and through hierarchical structure formation reproduce the large-scale structure seen in the real Universe. These simulations can also be extended to include baryons, and to simulate galaxy formation physics (e.g. Genel et al., 2014; Schaye et al., 2015). However, galaxy surveys cover such large volumes that it is typically only possible to run a dark-matter only simulation. Techniques can be used to link galaxies to dark matter haloes, such as the halo occupation distribution (HOD) (e.g. Peacock & Smith, 2000; Zheng et al., 2005), subhalo abundance matching (SHAM) (e.g. Vale & Ostriker, 2004; Conroy et al., 2006), or semi-analytic models (e.g. Baugh, 2006; Benson, 2010). For estimating covariance matrices, 1000s of mocks are needed, which can be created using fast, approximate techniques (e.g. Monaco et al., 2013; White et al., 2014; Chuang et al., 2015).

Analysis methods can be tested on these idealised mock catalogues by comparing the results against the input cosmology of the mock, which is known. However, the final survey catalogue will not be ideal, and is affected by various sources of incompleteness, and observational errors. By simulating these additional effects, the analysis methods can be modified to reduce these systematic effects. To make precise BAO and RSD measurements in the era of precision cosmology, it is essential that these systematics can be corrected.

1.5 Outline of thesis

This thesis will explore mock catalogues for upcoming large surveys, with a particular focus on the DESI Bright Galaxy Survey (BGS).

The outline of this thesis is as follows. Chapter 2 will outline methods for creating halo merger trees from N-body simulations, and Monte Carlo methods. This is the first step towards the creation of mock catalogues. These methods can also be extended beyond ΛCDM to simulations with warm dark matter.

Chapter 3 describes a method for creating a halo lightcone from the snapshots of the Millennium-XXL simulation, and a HOD method to populate it with galaxies. This mock is designed to be used for the DESI BGS, and reproduces the galaxy luminosity function and galaxy clustering of the SDSS and GAMA surveys.

Chapter 4 uses this mock catalogue to explore how the DESI BGS will be affected by incompleteness due to fibre assignment. Several correlation function correction techniques are assessed by applying them to samples from the mock.

Chapter 5 extends the HOD technique to create mock catalogues of Euclid and other galaxy surveys.

The conclusions are summarised in Chapter 6.

Chapter 2

Halo merger trees in cosmological simulations

2.1 Introduction

Structure in the Universe is believed to form hierarchically, where small overdensities in the early Universe collapse through gravitational instability to form dark matter haloes. Over time, haloes grow in mass as they slowly accrete matter, and also through mergers with other haloes. At early times, structure formation can be described analytically with linear perturbation theory. However, at later times, when haloes form, structure formation is highly non-linear. Non-linear structure growth is in general very difficult to describe analytically, but can be studied using numerical techniques.

Galaxy surveys measure the positions and properties of galaxies, which form in dense regions. To build realistic mock catalogues for these surveys, e.g. with realistic galaxy clustering properties, it is therefore important that the non-linear formation of structure can be modelled accurately. The starting point of a mock catalogue is typically an N-body simulation, which represents the matter density field as a set of particles, and calculates the gravitation force on each particle over many time steps. N-body simulations are able to accurately reproduce the large-

scale structure observed in the real Universe (Springel et al., 2005). Positions and velocities of particles are output from the N-body simulation at fixed times, or snapshots. This particle information is used to identify haloes in the simulation at each snapshot, and by identifying the descendant of each halo at the next snapshot, a halo merger tree can be built. For a halo at $z = 0$, the merger tree describes the merger history of all its progenitor haloes. This information can be used to interpolate between snapshots to determine the positions of the haloes on the past lightcone of a chosen observer, as described in Chapter 3.

Merger trees can also be built using a Monte Carlo technique, the starting point of which is extended Press-Schechter theory (Bond et al., 1991), which predicts the mass function of haloes, and also the conditional mass function (the mass function of the progenitor haloes at redshift z of a halo of mass M at a later redshift). This can be used to calculate the probability that a halo will fragment into two progenitors, working backwards in time. Starting with the final halo at $z = 0$, the algorithm can be iterated over many timesteps to build up the merger tree. While Monte Carlo merger trees have the disadvantage that they do not contain spatial information for haloes, the algorithm is fast, and can be efficiently run many times. Combined with a semi-analytic model, they can be used to measure accurate statistics of a galaxy population (e.g. Cole et al., 2000).

These techniques can be extended beyond ΛCDM, for example with warm dark matter (WDM). In a Universe with warm dark matter, with an elementary particle mass of the order of a few keV, the non-negligible thermal velocities of the dark matter particles at early times would allow the particles to free stream out of, and erase, small density perturbations, while large density perturbations would be unaffected (Bode et al., 2001). This results in the suppression of the formation of small haloes on the scale of the Milky Way (MW) satellites (Lovell et al., 2012). Constraints on the mass of the MW, and of the warm dark matter properties can be made by comparing the results of WDM simulations with the number of MW satellite galaxies. In order to make predictions with WDM, the Monte Carlo

algorithm must be calibrated to reproduce the conditional mass functions of N-body simulations (e.g. Benson et al., 2013).

In this chapter, we introduce the concept of halo merger trees, which is used in Chapter 3 to build mock galaxy catalogues. As another application of merger trees, we extend the methods to models of WDM, and show that constraints can be placed on the WDM particle by comparing the number of satellite galaxies produced in the merger tree to the observed number in the Milky Way. This chapter is organised as follows: Section 2.2, gives an overview of N-body techniques in ΛCDM. We describe the spherical collapse model, which motivates the halo mass definition used in simulations, and describe methods for identifying haloes and building halo merger trees, which is the first step towards making a mock catalogue. Section 2.3 describes a fast Monte Carlo method for generating merger trees from extended Press-Schechter theory. Section 2.4 extends these methods beyond ΛCDM, for sterile neutrino warm dark matter, calibrating the Monte Carlo merger trees to N-body simulations. In Section 2.5, the sterile neutrino Monte Carlo merger trees are applied to place constraints on the properties of the sterile neutrino and Milky Way mass by comparing to the observed number of satellites around the Milky Way. The conclusions are summarised in Section 2.6.

2.2 Halo merger trees in N-body simulations

2.2.1 N-body simulations

In an N-body simulation, the density field is represented by a set of discrete particles. These particles are arranged with some initial conditions at a high redshift, and are subsequently evolved to $z = 0$ over many small time steps, where the motion of each particle depends on the gravitational field due to the other particles in the simulation. The motion of particles traces out the evolution of structure. Simulations are typically done in comoving coordinates, which factors out the ex-

pansion of the Universe, keeping the size of the simulation box fixed. Since the real Universe is filled with matter on scales larger than the box, simulating an isolated box would be unsuitable. Periodic boundary conditions are typically used, where if a particle moves beyond the boundary, it will reappear at the opposite end of the box, and it is straightforward to generate periodic initial conditions.

In a particle-particle (PP) code, the force on each particle due to the other $N-1$ particles in the simulation is calculated directly at each time step (Hockney & Eastwood, 1988). However, this calculation scales as N^2, which is highly inefficient. More efficient methods have been developed to enable simulations with as many particles as possible. The particle-mesh (PM) technique involves first calculating the density field on a grid, which is interpolated to determine the force on each particle (Efstathiou et al., 1985). While this is more efficient, as fast Fourier techniques can be used to calculate the potential and the force, it is affected by the resolution of the grid on small scales. The particle-particle-particle-mesh (P^3M) scheme overcomes this by combining the PP and PM methods. The force is calculated directly for particles in neighbouring grid cells, but uses the grid for particles in more distant cells. However, this can be slow if the PP component dominates the calculation (Hockney & Eastwood, 1988). Adaptive grid techniques alleviate this by increasing the resolution of the grid in the highest density regions. Forces can also be calculated using tree codes, which involve organising the particles in a tree structure (e.g. Barnes & Hut, 1986; Hernquist et al., 1991). GADGET (Springel et al., 2001b; Springel, 2005) is a commonly used and highly parallel code, which implements a mesh on large scales, with a tree at intermediate scales, and PP on small scales.

As the separation between two particles $\mathbf{x}_i - \mathbf{x}_j \to 0$, the force between those particles $\mathbf{F} \to \infty$. This leads to particles being spuriously scattered by large angles, when in reality the particles should be collisionless, since each of the N-body particles represents a huge number of dark matter particles. This can be mitigated by force softening, which reduces the force at sufficiently small separa-

tions. For example, the net force acting on particle i with mass m_i in a system of N particles could be softened as

$$\mathbf{F}_i = -\sum_{i \neq j}^{N} \frac{G m_i m_j (\mathbf{x}_i - \mathbf{x}_j)}{(|\mathbf{x}_i - \mathbf{x}_j|^2 + \epsilon^2)^{3/2}}, \tag{2.1}$$

where the parameter ϵ sets the separation at which the force is softened.

Simulation initial conditions are set by first generating a random density field. This is a Gaussian random field, which is generated using the initial power spectrum. The real and imaginary component of each mode is drawn from a Gaussian distribution with variance set by the power spectrum, and each mode has a random phase (Efstathiou et al., 1985). Particles are arranged in the simulation volume either on a grid, or with a glass configuration (White, 1994). First or second order Lagrangian perturbation theory (e.g. Jenkins, 2010) uses the density field to calculate a corresponding displacement for each particle from this initial position, and to also assign particle velocities.

N-body simulations are limited by computational resources, which means that there is a balance between the box size of the simulation, and the resolution. If the total number of particles is kept constant, then the box size also specifies the particle mass, in order to achieve the correct mean density of the simulated Universe. Very high resolution simulations have limited box sizes, while very large cosmological simulations have limited resolution.

The simplest N-body simulations are dark-matter-only, containing collisionless matter. The Millennium simulation (Springel et al., 2005) simulated 2160^3 particles ($m_p = 8.6 \times 10^8 h^{-1}\mathrm{M}_\odot$) in a cubic box of length 500 h^{-1}Mpc. The subsequent Millennium II simulation (Boylan-Kolchin et al., 2009) had the same number of particles, but in a smaller box of size 100 h^{-1}Mpc, with particles masses $m_p = 6.9 \times 10^6 h^{-1}\mathrm{M}_\odot$. The Millennium-XXL simulation (MXXL) (Angulo et al., 2012b) simulated 6720^3 particles a box size of 3 h^{-1}Gpc with $m_p = 6.2 \times 10^9 h^{-1}\mathrm{M}_\odot$.

Simulations can be extended to include baryoninc physics, and star formation (e.g. Genel et al., 2014; Schaye et al., 2015). These simulations are strongly affected

by the assumptions made in the subgrid physics models, which model physical processes, such as star formation and feedback, which are not resolved in the simulation (Schaye et al., 2015).

Zoom simulations, such as the Aquarius project (Springel et al., 2008), simulate a single halo at high resolution, but at large distances from the halo, use increasing lower resolution particles in order to capture the tidal field that affects the region of interest.

Mock catalogues for large galaxy surveys require simulations with a very large volume, so typically it is only feasible to do a dark matter only simulation, which is populated with galaxies later. At each simulation snapshot, positions, velocities and other information for each individual particle is output. From this particle data, dark matter haloes must first be identified, and then matched between snapshots to build a halo merger tree. Methods for identifying haloes and defining their mass are motivated by the spherical collapse model.

2.2.2 The spherical collapse model

In general, it is difficult to model the non-linear growth of structure analytically. However, the formation of dark matter haloes can be simplified by considering the process as the collapse of a spherical overdensity.

In a Universe with $\Lambda = 0$, the evolution of a spherical overdensity can be described using Newtonian physics (e.g. Mo et al., 2010),

$$\ddot{r} = -\frac{GM}{r^2}, \tag{2.2}$$

where r is the radius of the sphere, and M is the enclosed mass. This equation can be integrated to give

$$\frac{1}{2}\dot{r}^2 - \frac{GM}{r} = E. \tag{2.3}$$

For $E < 0$, the system is bound, and the solution can be written parametrically as

$$r = A(1 - \cos\theta)$$
$$t = B(\theta - \sin\theta). \tag{2.4}$$

These equations can be Taylor expanded, and $r(t)$ for small t can be written as

$$r(t) = \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3} + \cdots\right], \tag{2.5}$$

and the overdensity within the sphere is

$$\delta(t) \approx \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{2.6}$$

Initially, the sphere grows as the Universe expands, but the rate at which it grows slows, due to the overdensity enclosed within the sphere. At turnaround, the sphere reaches its maximum size, then begins to collapse. This happens when $\theta = \pi$. The radius of the sphere gets smaller as it collapses, until it has collapsed down to a point when $\theta = 2\pi$. At the collapse time, $t = 2\pi B$, and hence from Eq. 2.6, the extrapolated linear overdensity is $\delta_c = (3/20)(12\pi)^{2/3} \approx 1.69$.

In the real Universe, the sphere would never collapse to a point, but would reach virial equilibrium, where the kinetic energy, K, and potential energy, V, are related through $V = -2K$. The collapsing sphere reaches virial equilibrium when it has collapsed by a factor of 2 from its maximum size at turnaround, so the radius is $r = A$. At the collapse time ($t = 2\pi B$), the true density enhancement of the sphere with respect to the background density is therefore $[(A/2)(6t/B)^{2/3}]^3 r^{-3} = 18\pi^2 \approx 178$.

In linear theory, when a region reaches an overdensity of $\delta_c = 1.69$ at some redshift z, it can be assumed that full non-linear collapse has occurred at this redshift. In an N-body simulation, the true collapse overdensity of ~ 200 can be used to identify virialized dark matter haloes. It was shown in Cole & Lacey (1996) that this overdensity is able to accurately separate the halo from the surrounding infalling material. These values are derived assuming a flat Einstein-de Sitter

cosmological model, with no cosmological model. If a cosmological constant is introduced, the value of the density threshold does change, but only by a small amount (Eke et al., 1996).

2.2.3 Identifying haloes in N-body simulations

An N-body simulation typically outputs the particle data, which includes particle positions and velocities, at several epochs, or snapshots. This particle data can be used to build halo catalogues. There are several commonly used algorithms for building halo catalogues from the particle information.

The friends-of-friends (FOF) algorithm (Davis et al., 1985) is a commonly used halo finder that defines each dark matter halo as the set of particles separated by a linking length b, which is units of the mean interparticle separation. Typically, a value of $b = 0.2$ is chosen, since this corresponds to an average overdensity close to the value of $\delta \sim 200$ predicted from spherical collapse. The FOF algorithm has the advantages that it is simple, and makes no assumptions about the shape of haloes. However, it is unable to detect substructures within large haloes. Unbound particles that happen to be moving near a halo will be identified as being part of the halo. Nearby large structures can be linked together by a tenuous bridge of particles, even though they are clearly separate haloes.

The SUBFIND algorithm (Springel et al., 2001a) identifies gravitationally bound structures, and is able to find arbitrary levels of substructure within substructure. The starting point is with a catalogue of haloes identified using the FOF algorithm. Within each FOF group, the sets of particles that belong to each locally overdense region are identified as subhalo candidates. Unbound particles are removed, and if there are more than 20 bound particles remaining, this is identified as a subhalo. Each particle can only be assigned to a single subhalo, so to ensure arbitrary levels of substructure can be found, the process begins with the largest subhalo, working towards smaller and smaller subhaloes. A typical FOF group will be decomposed

by SUBFIND into the most massive 'main' subhalo, which is surrounded by much smaller subhaloes. A few percent of the particles form a 'fuzz' of unbound particles which are identified as being part of the FOF group, but are not part of any subhalo. The decomposition of a FOF group into SUBFIND subhaloes is illustrated in Fig. 2.1. However, SUBFIND has the issue that it can fail to detect small subhaloes if they are close to the centre of large haloes, where the background density is high.

There are several ways in which the mass of a halo can be defined. One definition is to just take the sum of the masses of all the particles that are identified as part of the halo. Alternatively, the virial mass, motivated by the spherical collapse model, can be defined as M_{200}. This is the mass enclosed by a sphere, centred on the halo, in which the average density is 200 times the critical or average density of the Universe, i.e. $M_\Delta = (4/3)\pi R_\Delta^3 \Delta\bar{\rho}$, with $\Delta = 200$. Here the density, $\bar{\rho}$, can be either the critical density, $\rho_{\rm crit}$, or the mean density $\rho_{\rm mean} = \Omega_{\rm m}\rho_{\rm crit}$.

The FOF and SUBFIND algorithms have been used to identify haloes in the Millennium and MXXL simulations. However, many other halo finders have been developed. For example, the ROCKSTAR algorithm (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement) (Behroozi et al., 2013a), finds FOF groups in 6 dimensional phase space, adaptively reducing the linking length to find FOF groups within FOF groups, in order to up a hierarchy of substructure. Particles are then assigned to haloes, starting at the deepest levels of the hierarchy, and assigning the particles of the parent group to the halo of the nearest subgroup in phase space. Finally, unbound particles are removed. Amiga's Halo Finder (AHF) (Knollmann & Knebe, 2009) identifies haloes by calculating the density on a grid. If the particle density is higher than some threshold, the grid is refined recursively, creating a hierarchy of densities. If, when moving to a finer grid, a region splits into multiple regions, the region with the most particles is the host halo, while the other regions are substructures. Particles are assigned to haloes starting at the densest grid levels. If two haloes merge at a coarser level, particles within a sphere of radius half the distance to the host halo are assigned

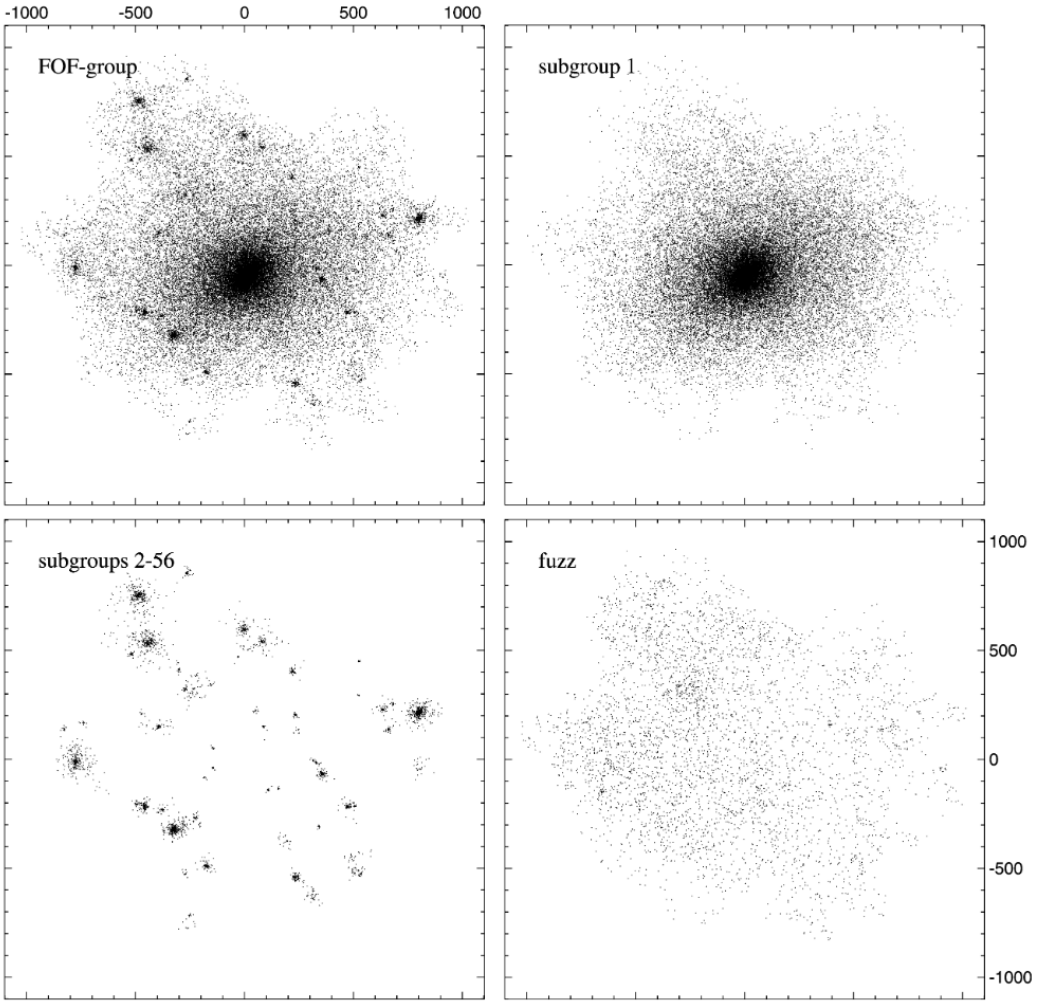

Figure 2.1: Example of the SUBFIND subhaloes identified within a FOF group. All particles belonging to the FOF group are shown in the upper left panel. Particles identified as being gravitationally bound to the main subhalo are shown in the upper right panel. The smaller subhaloes identified by SUBFIND are shown in the lower left panel. The lower right panel shows the remaining 'fuzz' of particles that are part of the FOF group, but not bound to any of the subhaloes. Coordinates are in h^{-1}kpc. Figure reproduced from Springel et al. (2001a).

as belonging to this halo, and then unbound particles are removed. See e.g. Knebe et al. (2013) for a comparison of halo finders.

2.2.4 Halo merger trees

A halo finder can be used to identify dark matter haloes at each simulation snapshot. However, haloes are also linked in time between snapshots. By finding the descendant of each halo at the next simulation snapshot, a halo merger tree can be built (e.g. Jiang et al., 2014).

A merger tree traces the evolution of each halo throughout the simulation. Each halo will first appear at the snapshot at which it is first resolved. Each halo grows in mass through the accretion of particles, and through mergers with other haloes. The merger tree tracks the descendant of each halo, so if two haloes merge, they will both have the same descendant halo at the next snapshot. In hierarchical structure formation, a halo can only increase in mass. This is not strictly true in an N-body simulation, as mass can be lost through stripping. Also, haloes should not be able to fracture into smaller haloes, but this is possible, e.g. if two FOF groups are tenuously linked together at one snapshot.

A merger tree is illustrated in Fig. 2.2, which shows the merger history of the progenitors of a final halo at snapshot s. Each circle represents a halo identified at each snapshot, where the size of the circle is proportional to its mass, and the arrows indicate the descendant of that halo at the next snapshot. The blue circles indicate the main (i.e. most massive) progenitor. If two haloes have the same descendant, this indicates that the haloes have merged between the two snapshots. Haloes can also grow in mass due to accretion. Low mass haloes will appear in the merger tree when they are massive enough to be resolved by the halo finder, and do not have progenitors.

To build a merger tree from an N-body simulation, the descendant of a halo at snapshot s needs to be identified at snapshot $s+1$. This can be done by matching

particles. The halo at snapshot $s+1$ that contains the highest number of bound particles that are identified as being in the halo at snapshot s is identified as the descendant halo. In the Millennium simulation, particles are given a weight, so that the more tightly bound a particles is, the higher the weight is. In the MXXL simulation, the descendant is the halo that contains the majority of the 15 most bound particles. However, it is not always straightforward to identify a descendant.

Sometimes, haloes can be 'lost' at one snapshot, only to reappear at a later snapshot. If a small satellite subhalo passes close to the centre of the host halo, then it can fail to be identified by the SUBFIND algorithm, since the background density is very high. Small isolated groups can also briefly drop below the resolution limit. The missing haloes can be filled in using methods such as Dhaloes (Jiang et al., 2014), which searches for descendants over several snapshots, and the Consistent Trees algorithm (Behroozi et al., 2013b), which uses the predicted evolution to add in the missing haloes.

2.3 Monte Carlo merger trees

Halo merger trees can also be generated using a Monte Carlo method. The starting point in this method is Press-Schechter theory (Press & Schechter, 1974).

The Press-Schechter formalism can be used to predict the halo mass function. This assumes that the density field evolves linearly, and when an overdensity exceeds $\delta(\mathbf{x}, t) > \delta_c \approx 1.69$,[1] it collapses to form a virialized halo. By smoothing the density field with a window function, $W(\mathbf{x}; R)$, of radius R that corresponds to a mass M, the collapsed regions can be assigned mass. Since the initial density field is a Gaussian random field, the smoothed density field is also Gaussian. By calculating the probability the smoothed density field exceeds δ_c, the Press-Schechter mass function can be derived.

[1] Since $\delta(\mathbf{x}, t) = \delta_0(\mathbf{x})D(t)$, where $D(t)$ is the linear growth rate, this condition can alternatively be written as $\delta_0(\mathbf{x}) > \delta_c(t)$, where $\delta_c(t) \equiv \delta_c/D(t)$.

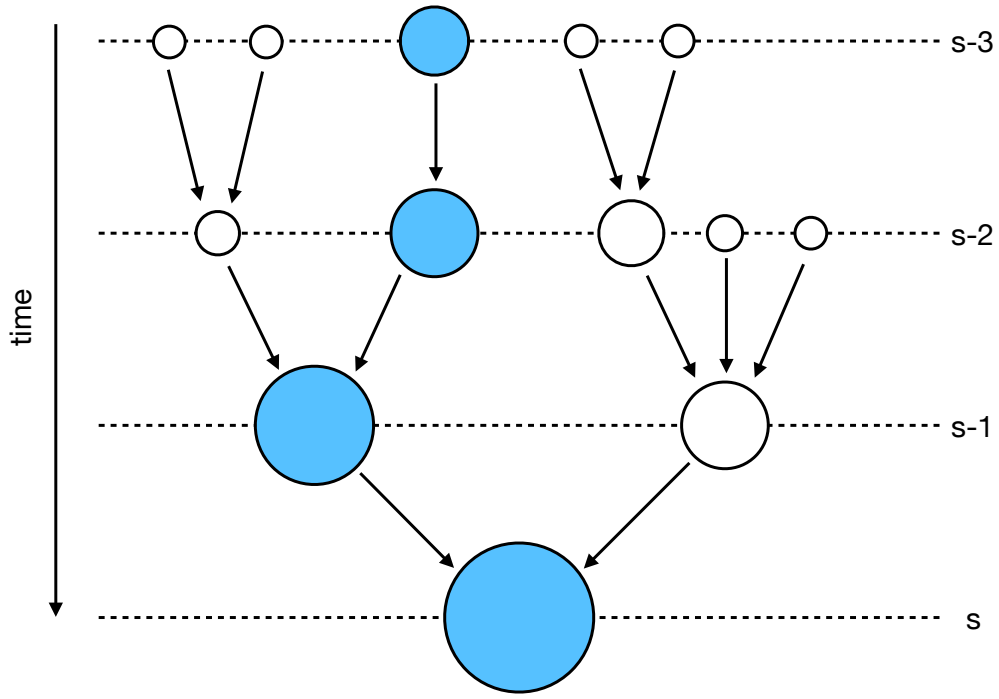

Figure 2.2: Schematic of a halo merger tree. Circles represent haloes at each snapshot, where the size of the circle is proportional to the halo mass. The arrows point to the descendant of that halo at the next snapshot. The main progenitor is coloured in blue.

The number density of bound objects with mass in the range $M \to M + dM$ is given by

$$n(M,t)dM = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M^2} \frac{\delta_c(t)}{\sigma(M)} \exp\left(-\frac{\delta_c(t)}{2\sigma^2(M)}\right) \left|\frac{d\ln\sigma(M)}{d\ln M}\right| dM, \tag{2.7}$$

where $\bar{\rho}$ is the mean density of the Universe, and

$$\sigma^2(M) = \frac{1}{2\pi^2} \int_0^\infty k^2 P(k) W^2(k;M) dk, \tag{2.8}$$

is the mass variance of the smoothed density field with power spectrum $P(k)$. $W(k;M)$ is the Fourier transform of the window function, which is a spherical top hat function in real space.

In linear theory, only initially overdense regions can collapse into virialized structures, so half of the mass of the Universe would never collapse into haloes.

However, an underdense region can be nested within a larger overdensity, so there is a non-zero probability for the matter to collapse. Press & Schechter accounted for this by introducing an arbitrary factor of 2 into the normalisation of their mass function. This also led to the development of extended Press-Schechter theory (EPS) in which the factor of 2 arose naturally (Bond et al., 1991; Bower, 1991).

The conditional mass function in EPS theory is

$$f(M_1|M_2)d\ln M_1 = \sqrt{\frac{2}{\pi}}\frac{\sigma_1^2(\delta_1-\delta_2)}{(\sigma_1^2-\sigma_2^2)^{3/2}}\exp\left(-\frac{(\delta_1-\delta_2)^2}{2(\sigma_1^2-\sigma_2^2)}\right)\left|\frac{d\ln\sigma_1}{d\ln M_1}\right|d\ln M_1. \quad (2.9)$$

This is the fraction of mass in a halo of mass M_2 at redshift z_2 that was originally in haloes of mass M_1 at the earlier redshift z_1. $\delta_1 = \delta_c(z=z_1)$ and $\delta_2 = \delta_c(z=z_2)$ are the linear theory critical density thresholds evaluated at z_1 and z_2. σ_1 and σ_2 is the variance of the smoothed density field (Eq. 2.8) evaluated at z_1 and z_2.

Following Cole et al. (2000) and taking the limit of Eq. 2.9 as $z_1 \to z_2$ gives[1]

$$\left.\frac{df}{dz_1}\right|_{z_1=z_2} d\ln M_1 dz_1 = \sqrt{\frac{2}{\pi}}\frac{\sigma_1^2}{(\sigma_1^2-\sigma_2^2)^{3/2}}\frac{d\delta_1}{dz_1}\left|\frac{d\ln\sigma_1}{d\ln M_1}\right|d\ln M_1 dz_1, \quad (2.10)$$

which is the average fraction of mass of a halo of mass M_2 that is in haloes of mass M_1 a small redshift step, dz_1, earlier. This can be used to estimate the mean number of haloes of mass M_1 that will merge to form a halo of mass M_2 in the interval dz_1, which is given by[2]

$$\frac{dN}{dM_1} = \frac{1}{M_1}\frac{df}{dz_1}\frac{M_2}{M_1}dz_1. \quad (2.11)$$

This can be integrated to find the mean number of progenitors above some mass resolution $M_{\rm res}$ (in the interval $M_{\rm res} < M_1 < M_2/2$),

$$P = \int_{M_{\rm res}}^{M_2/2}\frac{dN}{dM_1}dM_1, \quad (2.12)$$

and also the fraction of mass that lies below the resolution limit,

$$F = \int_0^{M_{\rm res}}\frac{dN}{dM_1}\frac{M_1}{M_2}dM_1. \quad (2.13)$$

[1] In the limit $z_1 \to z_2$, $\delta_1 - \delta_2 \to d\delta_1$, and $d\delta_1^2$ is negligible.

[2] The fraction of mass in haloes of mass M_2 that was previously in haloes of mass M_1 can be converted to the number of haloes of mass M_1 by multiplying by M_2/M_1. Multiplying the fraction by M_2 gives the total mass in haloes of mass M_1, and dividing by M_1 gives the number of haloes.

The Monte Carlo algorithm (Cole et al., 2000) is outlined below, which starts with the final halo, and works backwards in time over many small time steps to randomly split the halo into progenitors, building up a merger tree.

- The mass of the final halo, M_2, and the final redshift are both specified.
- The halo mass resolution limit, $M_{\rm res}$ is specified.
- A small redshift interval is chosen, so that $P \ll 1$. This makes it unlikely that a halo will have more than 2 progenitors.
- A uniform random number R in the range $0 < R < 1$ is chosen. If $R > P$, the mass of the halo is reduced by to $M_2(1-F)$, which accounts for mass accretion of unresolved haloes. If $R \leq P$, the halo is split into two progenitors. The first progenitor is given a random mass M_1 in the range $M_{\rm res} < M_1 < M_2/2$, while the other progenitor is given the mass $M_2(1-F) - M_1$.
- This process is repeated for each progenitor halo until a full merger tree has been constructed.

However, when compared to merger trees produced by N-body simulations, this method underestimates the mass of the most massive progenitors. To make the method consistent with EPS theory, Parkinson et al. (2008) modify Eq. 2.11 by a perturbing function,

$$\frac{dN}{dM_1} \rightarrow \frac{dN}{dM_1} G(\sigma_1/\sigma_2, \delta_2/\sigma_2), \tag{2.14}$$

which modifies the splitting rates, and modifies the mass distribution of the fragments. G is chosen to be of the form

$$G(\sigma_1/\sigma_2, \delta_2/\sigma_2) = G_0 \left(\frac{\sigma_1}{\sigma_2}\right)^{\gamma_1} \left(\frac{\delta_1}{\sigma_2}\right)^{\gamma_2}. \tag{2.15}$$

The parameters G_0, γ_1 and γ_2 are calibrated to reproduce the conditional mass functions measured in N-body simulations.

2.4 Merger trees with warm dark matter

2.4.1 Sterile neutrino WDM

The identity of the weakly interacting particle that makes up dark matter is currently unknown. In the ΛCDM model, particle candidates include the lightest supersymmetric particle (Ellis et al., 1984), which would have a mass of the order of GeV.

Another dark matter particle candidate well motivated by particle physics is the sterile neutrino (e.g. Dodelson & Widrow, 1994; Shi & Fuller, 1999; Asaka & Shaposhnikov, 2005). This would have a much smaller mass, of the order of keV, bringing it into the regime of warm dark matter (WDM). In a WDM universe, the DM particles would have non-negligible thermal velocities at early times, allowing them to free stream out of small density perturbations, and suppressing the formation of low mass haloes.

On large scales, CDM and WDM are indistinguishable. However, differences will become apparent on the scale of dwarf galaxies. At these smaller scales, there is some disagreement between the results of CDM simulations, and observations. For example, simulations of Milky Way (MW) sized haloes produce far more subhaloes than the number of observed satellites around the MW (Moore et al., 1999; Diemand et al., 2005; Springel et al., 2005). WDM is often motivated as a solution to this problem, due to the suppressed formation of low mass haloes (e.g. Lovell et al., 2012). However, the disagreements between simulations and observations can be resolved by introducing baryons, and the formation of galaxies in small haloes is suppressed due to reionization and feedback (Sawala et al., 2016).

The addition of a triplet of right-handed sterile neutrinos to the standard model of particle physics would explain neutrino masses and baryogenesis (Asaka & Shaposhnikov, 2005), the lightest of which would have a mass of the order of keV. The properties of this sterile neutrino are determined by its mass, the mixing angle,

and the lepton asymmetry, L_6, which is given by

$$L_6 = 10^6 \frac{(n_{\nu_e} - n_{\bar{\nu_e}})}{s}, \tag{2.16}$$

where n_{ν_e} and $n_{\bar{\nu_e}}$ are the number densities of electron neutrinos and antineutrinos, and s is the entropy density. For a given mass, there is a relationship between the mixing angle and L_6, and we will use L_6 as the free parameter. In addition to a thermal production mechanism (Dodelson & Widrow, 1994), the presence of a lepton asymmetry will boost the production of sterile neutrinos resonantly, below some momentum threshold (Shi & Fuller, 1999; Asaka & Shaposhnikov, 2005). The momentum distribution is shown in Fig. 2.3 for a 7 keV sterile neutrino with the value of L_6 varied. When $L_6 = 0$, there is no resonant production. As L_6 increases, the resonant production causes the distribution to peak at low momenta. The position of this peak moves to higher momenta with increasing L_6, until at very high values of L_6, where the production at all momenta is enhanced, and the distribution looks like that for non-resonant production.

In WDM, the power spectrum, $P(k)$ has a cutoff at large k (which corresponds to small physical scales, where the formation of structure is suppressed). The power spectrum can be written as

$$P_{\rm WDM}(k) = T^2_{\rm WDM}(k) P_{\rm CDM}(k), \tag{2.17}$$

where $P_{\rm CDM}(k)$ is the cold dark matter linear power spectrum, and $T_{\rm WDM}(k)$ is a transfer function, describing the damping introduced by the WDM particle. For a thermal relic WDM particle, the transfer function has the form of Bode et al. (2001),

$$T(k) = [1 + (\alpha k)^{2\nu}]^{-5/\nu}, \tag{2.18}$$

where $\nu = 1.12$ (Viel et al., 2005), and α is related to the WDM particle mass,

$$\alpha = 0.049 \left(\frac{\Omega_{\rm DM}}{0.25}\right)^{0.11} \left(\frac{h}{0.7}\right)^{1.22} \left(\frac{\rm keV}{M_{\rm th}}\right)^{1.11} h^{-1}{\rm Mpc}, \tag{2.19}$$

where $\Omega_{\rm DM}$ is the dark matter density parameter, and $M_{\rm th}$ is the mass of the thermal relic WDM particle, in keV. The mass of the sterile neutrino (with no

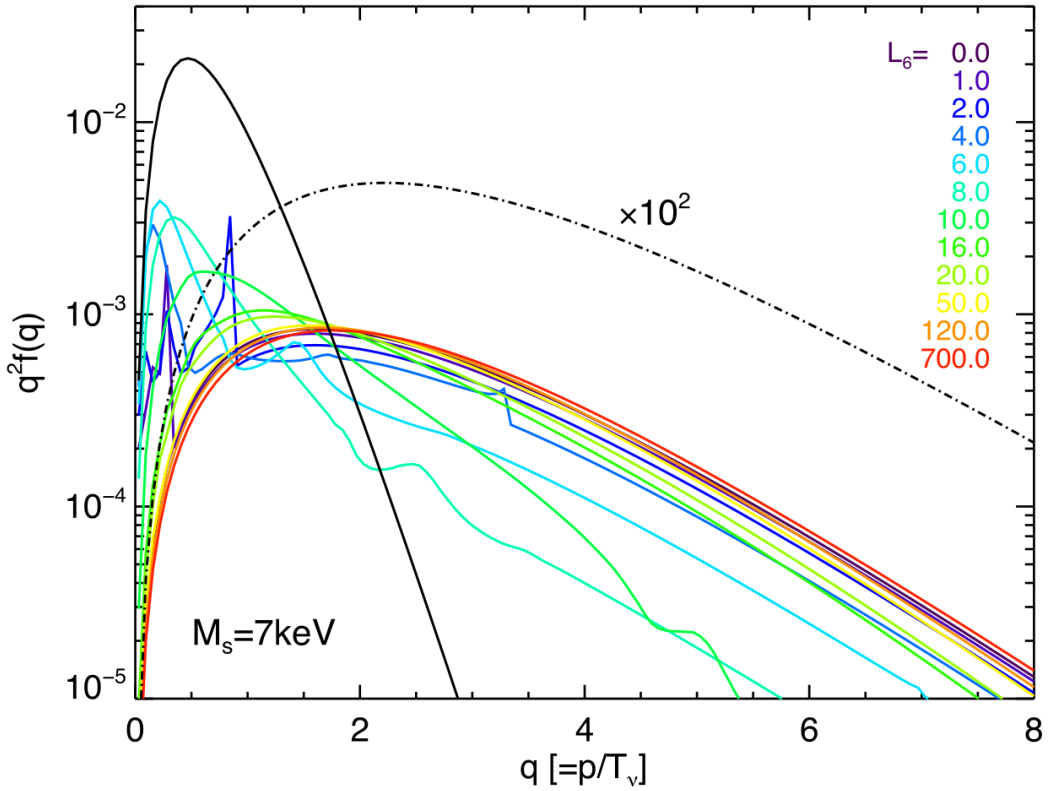

Figure 2.3: Momentum distribution for a 7 keV sterile neutrino with different values of L_6 (coloured lines). The solid black curve is for a thermal relic WDM particle with mass 1.4 keV, while the dot-dashed curve is for a thermal relic with the same temperature as the sterile neutrino, scaled by 10^2. Figure reproduced from Lovell et al. (2016)

resonant production), with the power spectrum cutoff at the same position as the thermal relic, is (from Viel et al., 2005)

$$M_{\rm s} = 4.43 \left(\frac{M_{\rm th}}{\rm keV}\right)^{4/3} \left(\frac{0.7^2 \times 0.25}{h^2 \Omega_{\rm DM}}\right)^{1/3} \rm keV. \tag{2.20}$$

The power spectrum of a 7 keV sterile neutrino is shown in Fig. 2.4. The WDM models all show a cutoff at large k, compared to the CDM power spectrum, which continues to rise to small scales. The dependence of the cutoff on L_6 shows non-monotonic behaviour, as with the momentum distribution. When $L_6 = 0$, the shape of the cutoff looks like a thermal relic. With increasing L_6, the resonant

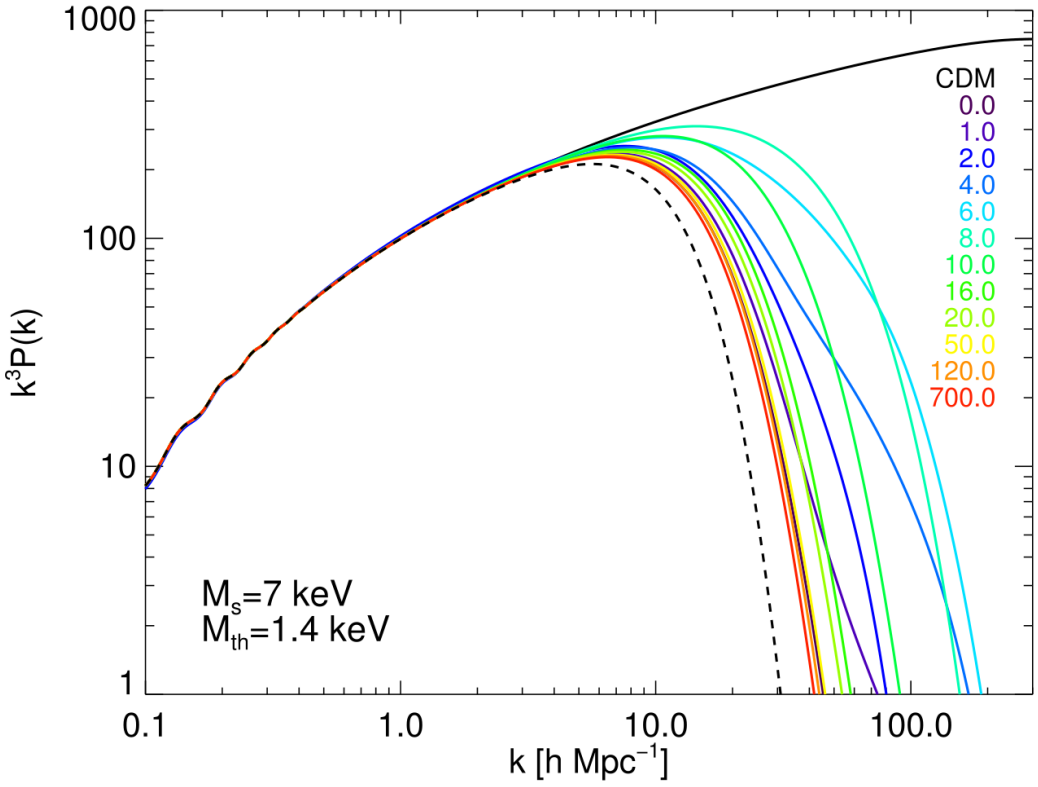

Figure 2.4: Power spectrum of a 7 keV sterile neutrino with different values of L_6 (coloured lines), which have a cutoff at large k. For comparison, the solid black curve shows the CDM power spectrum, while the dashed black curve is for a thermal relic of mass 1.4 keV, which has a cutoff at a position which is close to the sterile neutrino. Figure reproduced from Lovell et al. (2016).

production causes the WDM to become colder, and then warmer again, shifting the cutoff initially to higher k, then back to lower k.

Recently, observations have been made of an unidentified 3.5 keV feature in the X-ray spectra of galaxy clusters (Bulbul et al., 2014; Boyarsky et al., 2014). The feature has also been seen in the Milky Way (Boyarsky et al., 2015), the Andromeda galaxy (Boyarsky et al., 2014), and also in the cosmic X-ray background (Cappelluti et al., 2018). This could potentially be explained as an emission line produced by the decay of a ~ 7 keV sterile neutrino. The line has subsequently been detected in other galaxies (e.g. Neronov et al., 2016; Perez et al., 2017), and galaxy clusters

(e.g. Urban et al., 2015; Franse et al., 2016). However, other works have failed to detect the line (e.g. Anderson et al., 2015; Figueroa-Feliciano et al., 2015; Riemer-Sørensen et al., 2015).

Since the line has been observed in many objects, it is unlikely to be a statistical fluctuation. It is also unlikely to be due to instrumental systematics, since the line has been detected independently with several different instruments. The feature could also be due to atomic transitions, such as the potassium K XVIII lines (e.g. Jeltema & Profumo, 2015, 2016). Observations from Hitomi (Aharonian et al., 2017) did not find any atomic lines at 3.5 keV, but these observations were brief. More observations will be needed to understand the origin of the 3.5 keV line.

Recent results from the MiniBooNE experiment at Fermilab hint at the possible detection of a sterile neutrino, but this sterile neutrino would be too light to make up the bulk of the dark matter (MiniBooNE Collaboration et al., 2018).

2.4.2 N-body simulations with WDM

N-body simulations can be extended from CDM to WDM by changing the power spectrum that is used to set up the initial conditions. In a WDM simulation, the power spectrum of the relevant WDM model is used, which has a cutoff at high k. The particles in a WDM simulation should also initially have thermal velocities, which are negligible in the case of CDM. However, at a typical simulation resolution, these velocities are still negligible, so do not need to be included (Lovell et al., 2012; Shao et al., 2013; Leo et al., 2018).

An issue that affects WDM simulations is that filaments spuriously fragment into many small haloes. The size, and number of these spurious haloes depends on the resolution of the simulation; with increasing resolution, the mass of the spurious haloes decreases, but the total number increases. Wang & White (2007) found that the mass at which the spurious haloes appear in abundance scales as $m_{\rm p}^{1/3} k_{\rm peak}^2$, where $m_{\rm p}$ is the simulation particle mass, and $k_{\rm peak}$ is the wavenumber

at which $k^3P(k)$ reaches a maximum. Increasing the resolution to remove the spurious haloes is infeasible, as the mass of the spurious haloes scales very slowly with resolution; decreasing the particle mass by a factor of 8 only halves the mass of the spurious haloes. When particles are traced back to their positions in the initial conditions (or protohaloes), genuine haloes originate from spheroidal protohaloes, whereas the protohaloes of the spurious haloes have very flattened geometries.

A method for removing these spurious haloes is outlined in Lovell et al. (2014). This method uses the fact that the spurious haloes originate from protohaloes with very low sphericities[1], and do not have a match between low and high resolution simulations. This can be used to define cuts in sphericity and mass to remove the spurious haloes. Schneider et al. (2013) remove spurious haloes from measurements of the mass function by fitting a power law at the low mass end, where the spurious haloes dominate, and subtracting this. Hobbs et al. (2016) describe a method which uses adaptive softening, in which cells are only refined if they are undergoing collapse along all 3 axes to suppress the formation of these spurious haloes.

2.4.3 Monte Carlo merger trees with WDM

The Monte Carlo method, described in Section 2.3, can be extended to WDM by adding a cutoff to the power spectrum $P(k)$. The power spectrum is needed to calculate $\sigma(M)$, the rms density fluctuation, which is defined in Eq 2.8, where $W(k;M)$ is a window function which, in the standard EPS method, is chosen to be a top hat in real space. In Fourier space, this window function has the form

$$W(k;M) = \frac{3(\sin(kR) - kR\cos(kR))}{(kR)^3}, \tag{2.21}$$

where the mass, M, and filtering scale, R, are unambiguously related through $M = \frac{4}{3}\pi\bar{\rho}R^3$. Decreasing the mass of the window function has the effect of reweighting large k modes. This means that if the power spectrum has a sharp cutoff, $\sigma(M)$

[1]The sphericity, s, is defined as $s = c/a$, where a is the length of the largest axis, and c is the smallest axis. For a sphere, $c = a$, so $s = 1$. For a very flattened pancake-like structure, $c \ll a$, so s is close to 0.

will continue to increase with decreasing M, even though no new modes enter the filter. An alternative to the real space top hat is a sharp k-space filter. With this choice, the flattening of $\sigma(M)$ is set entirely by the sharpness of the power spectrum cutoff. However, this raises the problem that it is no longer clear how to relate the mass to the filtering scale. On dimensional grounds $M \propto k_{\rm cut}^{-3}$, and to maintain the usual relation between mass and radius we can write $M_{\rm SK} = \frac{4}{3}\pi\bar{\rho}R_{\rm SK}^3$, with $R_{\rm SK} = a/k_{\rm cut}$, where a is a constant which needs to be determined. By integrating the mean density under the window function and setting this equal to the required mass, Lacey & Cole (1993) find a value of $a = (9\pi/2)^{1/3} \approx 2.42$. Benson et al. (2013) and Schneider et al. (2013) match their results to N-body simulations, and find values of $a = 2.5$ and 2.7, respectively.

The use of a sharp k-space filter has been shown to be a suitable approach for the Viel et al. (2005) transfer function (Benson et al., 2013), which has a cutoff that is sharper than many of the sterile neutrino models. To check that the method is still valid for shallower sterile neutrino dark matter cutoffs it is necessary to check the calibration against N-body simulations.

To this end, we identified which of our set of sterile neutrino matter power spectra has the shallowest cutoff – $M_s = 3$keV and $L_6 = 14$, hereafter M3L14 – and used this as the input transfer function for re-runs of four of the Aquarius Project Milky Way dark matter haloes: Aq-A, Aq-B, Aq-C, and Aq-D (Springel et al., 2008). These were run at Aquarius resolution level 3 (softening length 120.5 pc, particle mass 5.6×10^4, 2.5×10^4, 5.4×10^4, and $5.4\times10^4 {\rm M}_\odot$, respectively) with the P-GADGET3 code; the cosmological parameters match the 7-year *Wilkinson Microwave Anisotropy Probe* constraints (*WMAP*-7; Komatsu et al., 2011). Haloes and subhaloes were identified using the gravitational potential unbinding code, SUBFIND (Springel et al., 2001a). Spurious subhaloes – those subhaloes that form by spurious fragmentation of filaments – were identified and removed from the catalogues using the Lagrangian region shape and maximum mass criteria of Lovell et al. (2014). We then compare the conditional mass functions of these simulations

with those derived from the EPS method. For a halo of mass M_2 at $z_2 = 0$, the conditional mass function gives the fraction of mass contained within progenitor haloes of mass M_1 at some earlier redshift z_1. We plot the conditional mass functions at $z_1 = 1$ in Fig. 2.5, for haloes with a final mass of $M_2 \sim 1.5 \times 10^{12} h^{-1} \mathrm{M}_\odot$. In the top panel we compare the rms density fluctuations of the M3L14 matter power spectrum with those of CDM and also three Viel et al. (2005) thermal relic power spectra with transfer function parameters $\alpha = 0.0199$, 0.0236, and 0.0340 h^{-1}Mpc, which correspond to thermal relic masses $M_{\mathrm{th}} = 2.3$, 2.0, and 1.5 keV, respectively (Lovell et al., 2014). These were calculated using a sharp k-space filter with $a = 2.7$. The M3L14 model has a different behaviour than the thermal relic models, in that the curve peels away from CDM at the same mass scale as the $\alpha = 0.0236$ model but has a slightly shallower slope for large masses. Compared to the $\alpha = 0.0236$ thermal relic, σ has a lower amplitude at intermediate mass scales but a higher amplitude for $M < 10^9 h^{-1} \mathrm{M}_\odot$. This change is reflected in the $z_1 = 1$ conditional mass functions, which are shown in the lower panels. The middle panel compares the conditional mass functions to the average of the four sterile neutrino Aquarius haloes, while the lower panel shows, for each WDM model, the conditional mass function of the Aq-A halo. For $M \gtrsim 10^9 h^{-1} \mathrm{M}_\odot$, M3L14 produces a similar number of haloes as the $\alpha = 0.0236$ thermal relic model, but below this mass the rate of decrease is much shallower such that at $M \sim 10^8 h^{-1} \mathrm{M}_\odot$, M3L14 has a greater abundance of haloes than even the $\alpha = 0.0199$ thermal relic model. In spite of this change, there is still good agreement between the number of substructures predicted by the EPS method and the number measured in the cleaned simulation halo catalogues. We choose a value of $a = 2.7$ as this produces the best agreement, but the effect of varying a is small.

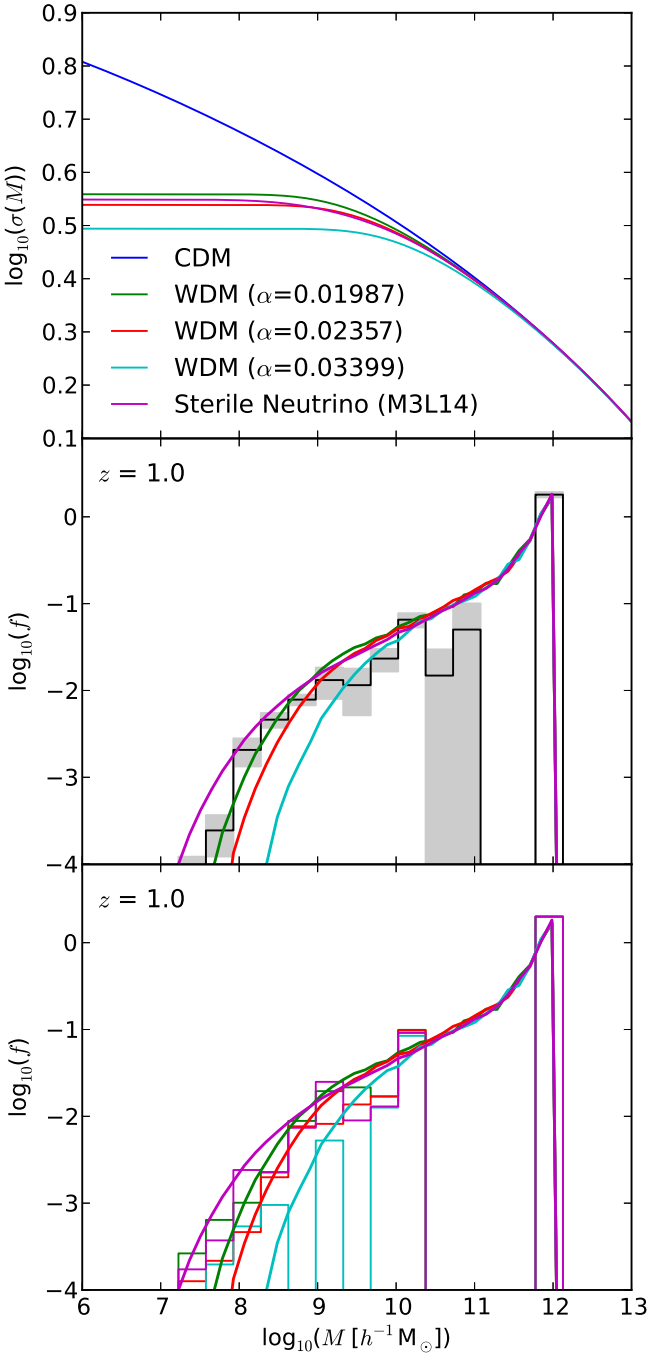

Figure 2.5: *Top panel:* $\sigma(M)$, the rms density fluctuation smoothed over mass scale, M, using a sharp k-space filter with $a = 2.7$, for CDM, WDM and sterile neutrinos where, for WDM, α determines the position of the cutoff in the power spectrum. *Middle panel:* mean conditional mass function of four N-body sterile neutrino haloes, based on the Aquarius Project with *WMAP-7* cosmological parameters (black histogram); 1σ errors are shown by the grey shaded area. Coloured lines show the mean conditional mass functions of 1000 Monte Carlo simulations, using a sharp k-space filter, for the different DM cases, with the same colours as in the top panel. *Bottom panel:* N-body conditional mass functions from halo A for the different DM cases (colour histograms). Curved lines are the same Monte Carlo conditional mass functions from the middle panel.

2.5 Milky Way satellite galaxies with sterile neutrino WDM

The modified Monte Carlo method, described in Section 2.4.3, which is calibrated to reproduce the conditional mass functions of WDM N-body simulations, is used in Lovell et al. (2016) to generate many merger trees for Milky Way (MW) sized haloes with sterile neutrino WDM. A semi-analytic model is used to predict the number of satellite galaxies around MW mass haloes. By comparing with the actual observed number of MW satellites, constraints can be placed on the properties of the sterile neutrino, and also on the mass of the MW.

The Gonzalez-Perez et al. (2014) version of the GALFORM semi-analytic model (Cole et al., 2000) is applied to the halo merger trees with different values of the MW halo mass, $M_{\rm h}$, sterile neutrino mass, $M_{\rm s}$, and lepton asymmetry, L_6. The method of Kennedy et al. (2014) is used to rule out combinations of these three parameters that are unable to produce the observed number of MW satellites.

Fig. 2.6 shows the constraints in the $M_{\rm h}$-$M_{\rm s}$ plane for different values of L_6. For a given value of L_6, the area to the lower-left of the curve is ruled out, as these values of $M_{\rm h}$ and $M_{\rm s}$ are unable to produce enough satellite galaxies in the semi-analytic model to be able to account for the observed number of satellites, while the area to the upper-right is allowed. The curves show the same non-monotonic behaviour with L_6 as described previously. Additional constraints on the sterile neutrino properties can be placed from measurements of the MW halo mass. There is still some uncertainty in the MW halo mass (e.g. as is summarised in figure 1 of Wang et al., 2015), but most measurements are in the range $5\times10^{11} < M_{\rm h} < 2\times10^{12}$ M$_\odot$. For a 7 keV sterile neutrino, the minimum halo mass of $M_{\rm h} \sim 1.5\times10^{12}$ M$_\odot$, which is well within the range of mass estimates, occurs at $L_6 \sim 8$.

Note that these results are very much dependent on the semi-analytic model, as was explored in Kennedy et al. (2014) for a thermal relic WDM particle. While

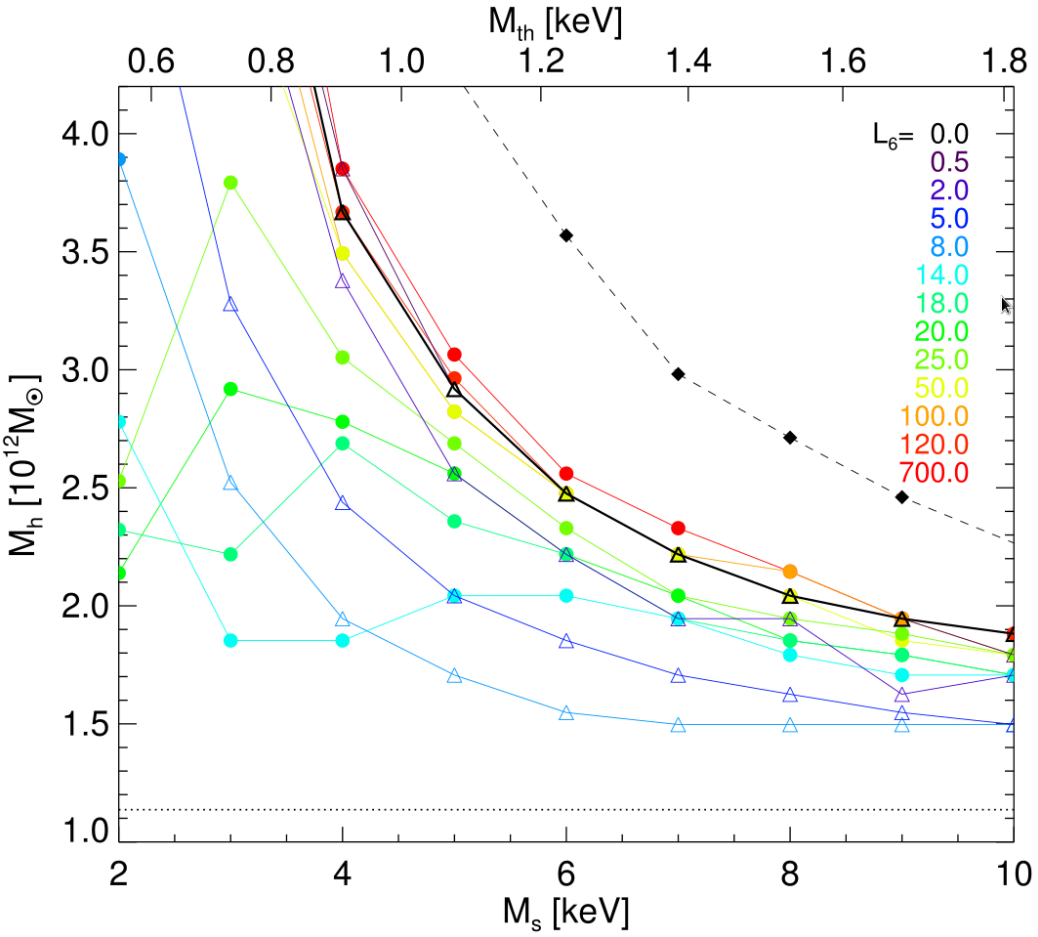

Figure 2.6: Minimum Milky Way halo mass, $M_{\rm h}$, needed to produce the number of observed MW satellites as a function of sterile neutrino mass, $M_{\rm s}$, for different values of L_6, as indicated by the legend. Empty triangles indicate combinations of $M_{\rm s}$ and L_6 that are ruled out by X-ray non-detections (Boyarsky et al., 2014), while filled circles are not ruled out. The black dashed line indicates the constraints from a thermal relic WDM particle, and the horizontal dotted line is the constraint from CDM. Figure reproduced from Lovell et al. (2016).

the general trends will remain unchanged if the parameters of the model are varied, the exact quantitative values will change.

2.6 Conclusions

Dark matter halo merger trees are the first step towards creating mock galaxy catalogues. These are typically constructed from an N-body simulation, which trace the non-linear formation of structure by evolving a set of particles over many small time steps from some initial conditions at high redshift, to $z = 0$. N-body simulations are able to reproduce the cosmic web of structure that is seen in the real Universe. In order to build up a halo merger tree, which traces the merger history of progenitor haloes at each simulation snapshot, haloes must first be identified at each snapshot, using an algorithm such as FOF or SUBFIND, and then by matching particles between snapshots, the descendant of each halo can be identified.

Halo merger trees can also be generated using a Monte Carlo algorithm. This has the disadvantage that it does not contain spatial information for haloes, but the algorithm is very fast, and can be run efficiently many times in order to build up accurate galaxy statistics when combined with a semi-analytic model, such as GALFORM. The algorithm begins with the final halo at $z = 0$, and works backwards in time, using extended Press-Schechter theory to calculate the probability that the halo will be split into two progenitors.

These methods can be extended from CDM to WDM, where the power spectrum has a cutoff at large k. The WDM power spectrum is used to set the initial conditions of the N-body simulation, and spurious haloes need to be removed. The Monte Carlo method must use a sharp k-space filter when calculating $\sigma(M)$, because of the cutoff in the power spectrum, and is calibrated to reproduce the conditional mass functions of N-body simulations.

The sterile neutrino is a WDM particle candidate that is motivated by particle physics as it would explain neutrino masses and baryogenesis. Recent observations of a 3.5 keV line in galaxies and galaxy clusters could potentially be explained as the decay of a 7 keV sterile neutrino. As an application of the WDM Monte

Carlo method for creating merger trees, Lovell et al. (2016) use the number of galaxies around the Milky Way to place constraints on the properties of the sterile neutrino, and mass of the MW. For a 7 keV sterile neutrino with lepton asymmetry $L_6 \sim 10$, the minimum halo mass required to reproduce the number of observed MW satellites is $\sim 1.5 \times 10^{12}$ $M_\odot$, which is consistent with other measurements of the mass of the MW halo. However, these results are affected by the parameters used in the semi-analytic model.

My contribution to Lovell et al. (2016) was to check the calibration of the WDM Monte Carlo merger trees against N-body simulations by computing the conditional mass functions shown in Fig. 2.5.

Chapter 3

A lightcone catalogue from the Millennium-XXL simulation

3.1 Introduction

Upcoming galaxy surveys, such as the Dark Energy Spectroscopic Instrument (DESI) survey (DESI Collaboration et al., 2016a,b) and Euclid (Laureijs et al., 2011), aim to measure the expansion history of the Universe and the growth of cosmic structures. Measurements of galaxy clustering, redshift space distortions and weak lensing will test general relativity, constrain theories of dark energy, and give us precise cosmological constraints.

In order to reach the high precision required to meet these aims, it is necessary to understand and quantify the systematic uncertainties in measurements from surveys, which requires the use of accurate mock catalogues (Baugh, 2008). Since a mock catalogue has a known cosmology and the 'true' value of a statistic can be measured directly, they can be used to develop and test the analysis tools which will be used on real observations.

Mocks are also required to test observational strategies and quantify the resultant levels of sample incompleteness. It is often not possible to assign a fibre to every galaxy due to mechanical constraints on fibre positioning (e.g. Hawkins

et al., 2003; Guo et al., 2012; Hahn et al., 2017; Burden et al., 2017; Pinol et al., 2017) and even if a fibre is assigned, a redshift measurement can fail if the galaxy has weak emission lines or low surface brightness. This incompleteness may have a significant effect on clustering measurements, and therefore in order to make precise baryon acoustic oscillation (BAO) and redshift space distortion measurements, it is important that this incompleteness is well understood, and that methods are developed and tested in order to mitigate these effects on the measured clustering. The differences in the clustering statistics expected in viable models is small, making it essential that systematics like these are understood.

Mock catalogues which have realistic galaxy clustering can be created from cosmological simulations. In order to see the BAO peak in clustering measurements, at a scale of the order of 100 h^{-1}Mpc, these simulations need to have a very large box size of the order of a few Gpc. Running a hydrodynamical simulation that has both the large volume needed to model such scales, and the resolution to produce faint galaxies down to the flux limit of the survey is infeasible, due to the large computational expense. Dark matter only simulations are much less expensive. There are several schemes which can be used to populate haloes in a dark matter only simulation with galaxies. These include the halo occupation distribution (HOD) (e.g. Peacock & Smith, 2000; Seljak, 2000; Scoccimarro et al., 2001; Berlind & Weinberg, 2002; Kravtsov et al., 2004; Zheng et al., 2005), which describes the probability a halo with mass M contains N galaxies with some property; the closely related conditional luminosity function (CLF) (e.g. Yang et al., 2003), which specifies the luminosity function of galaxies at each halo mass; subhalo abundance matching (SHAM) (e.g. Vale & Ostriker, 2004; Conroy et al., 2006), which assumes a correlation between halo or subhalo properties (e.g. mass or circular velocity), and galaxy properties (e.g. luminosity or stellar mass); and semi-analytic models (SAMs) (e.g. Baugh, 2006; Benson, 2010; Somerville & Davé, 2015), which uses analytic prescriptions to model the formation and evolution of galaxies.

In order to apply a SAM to a simulation, high resolution merger trees are

needed, and these are difficult to construct for large volume simulations. However, there are approaches which can augment the resolution of the simulation merger trees (e.g. de la Torre & Peacock, 2013; Angulo et al., 2014; Benson et al., 2016). The SHAM prescription assigns galaxies to subhaloes, requiring a complete subhalo catalogue. Since subhaloes are disrupted when they undergo mergers, this catalogue will only be complete for large subhaloes with thousands of particles, and so a very high resolution simulation is needed to resolve the low mass subhaloes that will be populated by faint galaxies. The HOD, on the other hand, can be applied to a lower resolution simulation, since satellite galaxies can be placed around the central galaxy following an analytic distribution, without knowledge of the subhaloes. The HOD method can also be applied to simulations in which the underlying cosmology has been rescaled (e.g. Angulo & White, 2010).

Ideally, these methods would be used to populate a halo lightcone that is the direct output from a simulation. However, most simulations do not output lightcones, but output snapshots at discrete times. Typically when a HOD method is used, it is applied to a single snapshot. However, this means that the halo bias is constant, and so the clustering of haloes does not evolve with redshift in the mock. Multiple snapshots can be joined together to create a lightcone, but this leads to discontinuities at the boundaries; the same halo could appear twice at either side of the boundary, or not at all (e.g. Fosalba et al., 2015).

The standard abundance matching and HOD schemes do not incorporate evolution. Attempts have been made to extend the abundance matching scheme, such as Moster et al. (2013), which reproduces the observed stellar mass function at different redshifts. There is currently no complete model for HOD evolution, as this evolution would depend on the galaxy sample under consideration. Contreras et al. (2017) use the HODs produced in SAMs to build a simple parametrisation of the evolution of the HOD parameters.

Here, we describe a HOD method which we use to populate haloes over a range of redshifts from the Millennium-XXL (MXXL) simulation with galaxies.

We first create a halo lightcone catalogue from the simulation by interpolating the positions of haloes between snapshots, which is then populated with galaxies using HODs, reproducing the observed clustering from the Sloan Digital Sky Survey (SDSS) (Abazajian et al., 2009) and the Galaxy and Mass Assembly (GAMA) survey (Driver et al., 2009, 2011; Liske et al., 2015).

This chapter is organised as follows: in Section 3.2 we describe the MXXL simulation and outline the method for generating the halo lightcone catalogue. In Section 3.3, we describe the halo occupation distribution model, and our method of evolving the HODs with redshift. In Section 3.4 we outline the method used to populate the halo lightcone with galaxies, and the method used to assign each galaxy a ${}^{0.1}(g-r)$ colour. In Section 3.5, we give examples of potential applications of the mock catalogue.

3.2 Halo lightcone catalogue

3.2.1 The MXXL simulation

The Millennium-XXL (MXXL) simulation (Angulo et al., 2012b) is a large dark-matter only N-body simulation in the same family as the Millennium simulation (Springel et al., 2005). The volume of MXXL is 216 times larger than Millennium, with a box size of 3 h^{-1}Gpc, and the particle mass is 6.17×10^9 $h^{-1}\mathrm{M}_\odot$, with a force softening of 13.7 kpc. MXXL adopts a ΛCDM cosmology with the same 1-year *Wilkinson Microwave Anisotropy Probe* (*WMAP*-1) cosmological parameters as the Millennium simulation, $\Omega_\mathrm{m} = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$, $h = 0.73$, and $n = 1$ (Spergel et al., 2003). The initial conditions were set at a starting redshift of $z = 63$, and the simulation was evolved to $z = 0$ with 63 outputs. The large volume of the simulation means that it can be used to study features such as baryon acoustic oscillations (BAOs) and redshift space distortions with good statistics.

3.2.2 Merger trees

We use the halo merger trees computed by Angulo et al. (2012b). Haloes were found using a Friends-of-Friends (FOF) algorithm (Davis et al., 1985), and bound subhaloes were identified using SUBFIND (Springel et al., 2001a). Halo merger trees were built by identifying the unique descendant of each subhalo at the subsequent snapshot. For each subhalo, the 15 most bound particles were found, and the subhalo at the next snapshot which contains the greatest number of these particles was defined as the descendant. In the case that two subhaloes contain equal numbers of these particles, the subhalo with the greatest total binding energy was chosen (Angulo et al., 2012a).

At each simulation snapshot, the subhalo merger trees are split over 3072 files. Each file contains information (e.g. position, velocity) for a subset of the haloes at that snapshot. The files also contain descendant information, i.e. in which file the descendant halo at the next snapshot is located. However, for a halo at one snapshot, its progenitors at the previous snapshot can be spread over many files, making it necessary to read in the entire merger tree at once to make the halo lightcone. To reduce the amount of memory required, we reorganise the merger tree files such that FOF groups at $z = 0$ are randomly assigned to one of 3072 files, and progenitor subhaloes are all placed in the same file as their $z = 0$ descendant. This allows us to run the lightcone code independently on each of these new files.

3.2.3 Constructing the halo lightcone catalogue

The full sky halo lightcone catalogue is created using the standard interpolation method (e.g. Merson et al., 2013, but applied to haloes rather than galaxies). An observer is firstly placed randomly inside the MXXL box. If the observer happened to be placed at the centre, haloes at the edge of the box would have a redshift $z \sim 0.5$; multiple periodic replications of the box must therefore be used in order to construct a catalogue that goes to redshifts higher than this. This

replication is done without any artificial rotation or translation in order to prevent the introduction of discontinuities. The positions and velocities of each halo at each snapshot are used to interpolate their trajectories through the simulation. From the position and redshift of a halo at two adjacent snapshots, it can be determined whether the halo crossed the observer's lightcone; if it has, a binary search algorithm is used to find the interpolated position (and velocity) at the redshift where it crosses.

The halo occupation distribution method of creating the galaxy catalogue (Section 3.3) assigns galaxies to FOF groups. Since the merger tree is defined for SUBFIND subhaloes, we need to infer the merger tree for the FOF haloes. To do this, we make the assumption that the position (and velocity) of the main subhalo (i.e. the most massive subhalo) in each FOF group is the same as that of the FOF group itself. The descendant FOF group is then found from the descendant of the main subhalo. To interpolate the position (and velocity) of each subhalo, we use cubic interpolation (i.e. use a cubic polynomial to describe the path of the halo in each dimension, using the positions and velocities at the previous and next snapshot as boundary conditions).

We use a halo mass definition of $M_{200\mathrm{m}}$ (the mass enclosed by a sphere, centred on the halo, in which the average density is 200 times the mean density of the Universe), as stored in the MXXL output for each FOF group. Since the number of galaxies in each halo depends on its mass, $M_{200\mathrm{m}}$ must be interpolated between snapshots. Below $z = 2$, the simulation snapshots are approximately spaced linearly with expansion factor. We use the descendant of the most massive subhalo to find the descendant of each halo, and then interpolate linearly in mass between snapshots, finding the mass at the redshift at which it crosses the lightcone. In the case that two or more haloes merge between snapshots, the total mass of the haloes is interpolated linearly, and each halo is assigned a constant fraction of the total mass. If the halo is not the most massive progenitor, a random time between snapshots is chosen for the merger to take place. If the halo crosses the observer's

lightcone after this time, the merger has happened, and the interpolated mass of the halo is transferred to the most massive progenitor.

The mass function of the halo lightcone at low redshifts ($z < 0.1$) is shown in Fig. 3.1 and compared to the Sheth & Tormen (1999) and Jenkins et al. (2001) analytic mass functions. At high masses, the mass function of the lightcone catalogue is in reasonable agreement with Sheth & Tormen (1999) and Jenkins et al. (2001), but there is a lower abundance of less massive haloes. This difference is because haloes in the simulation are identified using a FOF algorithm, and not a spherical overdensity (SO) finder. However, a SO mass is calculated for each FOF halo ($M_{200\mathrm{m}}$). Any small overdensities close to a large FOF group would be identified as part of the large FOF group, and therefore these would be missing from the halo catalogue.

In order to add haloes to the catalogue that are below the MXXL mass resolution (Section 3.2.5), and to evolve the HODs with redshift (Section 3.3.2), it is useful to have a smooth function which is in close agreement with the actual mass function of the catalogue. For this, we take a mass function with the same form as Sheth & Tormen (1999), but refit the parameters to the MXXL mass function. This fit is shown as the green curve in Fig. 3.1, which is in better agreement with the MXXL mass function at low masses, and is close to the fit given in equation 2 of Angulo et al. (2012b). The MXXL mass function peels away from this fit slightly at masses close to the resolution limit, but is complete for masses greater than $\sim 10^{12}\ h^{-1}\mathrm{M}_{\odot}$.

The number density of haloes as a function of redshift in the halo lightcone catalogue is shown in Fig. 3.2 for several mass thresholds. If halo masses are kept fixed between snapshots, step features can be seen, since the mass function is being kept frozen and will only change at the next snapshot. These features are most apparent for the highest mass threshold, for which the number density decreases more rapidly at high redshifts. Mass interpolation greatly reduces these features. At low redshifts, the curves become noisy, due to the small volume in each redshift

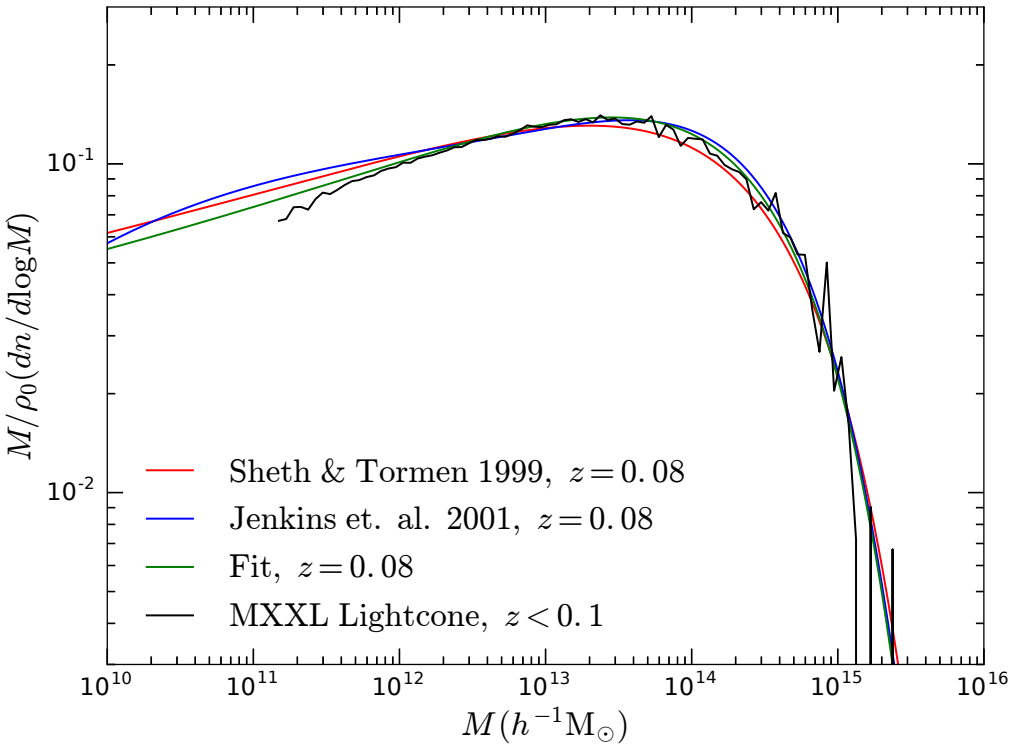

Figure 3.1: Mass function of the halo lightcone catalogue for $z < 0.1$ (black), compared to the analytic mass function of Sheth & Tormen (1999) (red), Jenkins et al. (2001) (blue), and our fit to the MXXL mass function (green), at the median redshift $z = 0.08$. Halo masses are defined as $M_{200\mathrm{m}}$, and have been interpolated linearly between simulation snapshots.

shell.

The large-scale real space correlation function of the lightcone catalogue for FOF groups with masses $M_{200\mathrm{m}} > 3\times10^{12}\ h^{-1}\mathrm{M}_\odot$ and $z < 0.5$ is shown in Fig. 3.3. This redshift limit avoids structures being repeated due to periodic replication of the box[1]. The BAO peak can be seen clearly in the clustering of haloes.

[1]Lightcones with a wide opening angle, or directed along the principle axes of the simulation, that extend beyond $z = 0.5$ will contain repeated structures. This will result in clustering errors being underestimated.

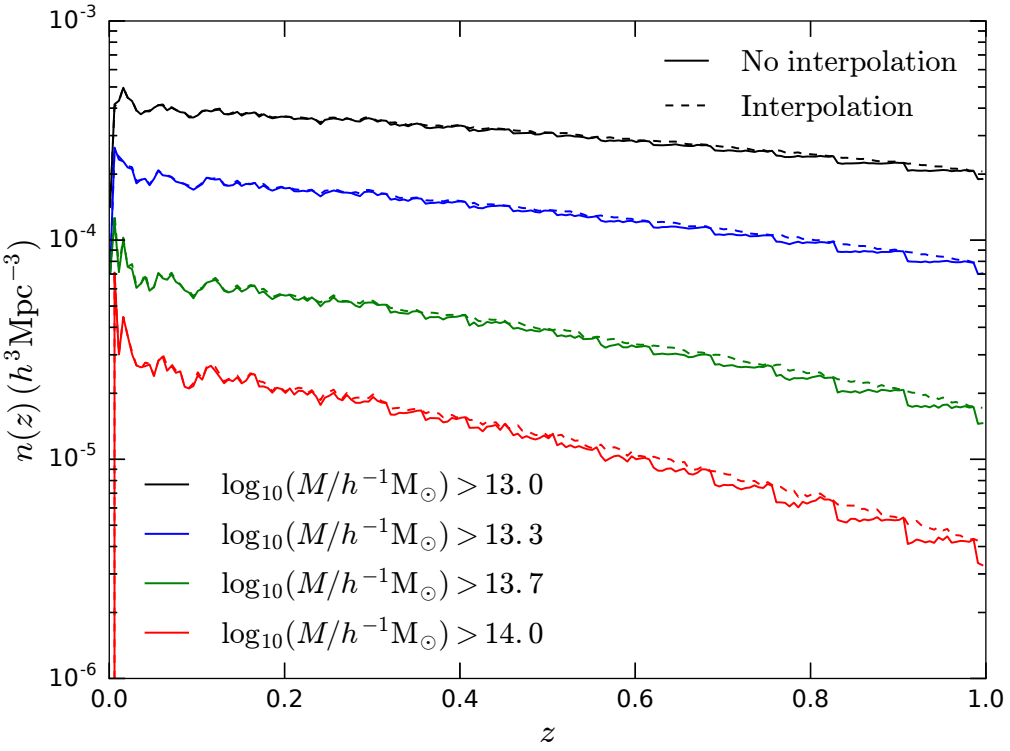

Figure 3.2: Number density of haloes in the halo lightcone catalogue as a function of redshift for haloes with mass $M_{200\mathrm{m}}$ greater than several thresholds, as indicated by the colour. Solid lines are where the halo mass has been kept frozen between snapshots, and dashed lines are where the mass has been interpolated.

3.2.4 Caveats

The interpolation scheme uses the position and velocity of a halo at two snapshots as boundary conditions in order to find the path the halo moved through in the simulation. If two haloes merge together, there is not enough information to determine when this occurs, so we assume they merge at a random time. If a new halo forms, or drops below the resolution limit, we assume this happens exactly on a snapshot.

To construct the merger trees, the descendant of a subhalo is defined as the subhalo which contains the majority of its 15 most bound particles (Angulo et al., 2012a). However, it is likely that some of the particles of the descendant subhalo

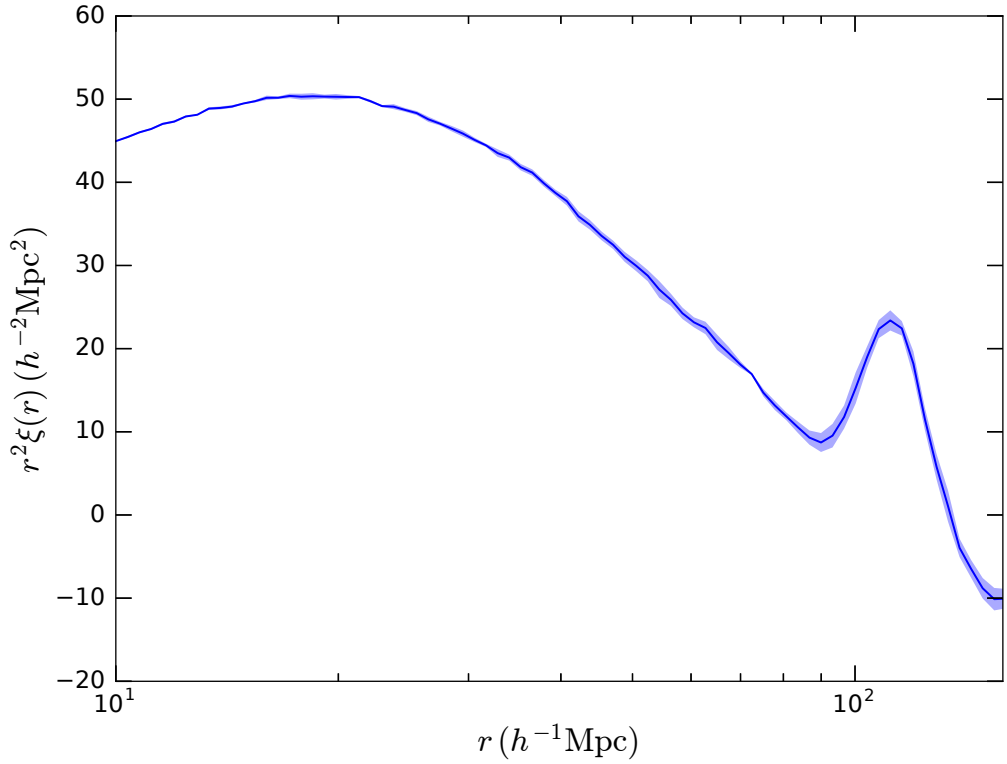

Figure 3.3: Real space correlation function, scaled by r^2, of the halo lightcone catalogue for haloes with $M_{200\mathrm{m}} > 3 \times 10^{12}\ h^{-1}\mathrm{M}_\odot$ and $z < 0.5$. The blue shaded area shows the error on the mean in the clustering, calculated from four quadrants of the sky.

were not in its progenitor, and vice versa. All of these particles are used to calculate the position and velocity of the subhalo, which can occasionally lead to jumps in the position of a subhalo that are inconsistent with its velocity.

Sometimes, a halo can be lost by the halo finder at one snapshot, but is then found again at a later snapshot. This can happen if a small halo passes very close to a more massive halo at one snapshot; the SUBFIND algorithm can fail to identify the small halo as the algorithm finds that its particles are bound to the massive halo. The MXXL merger trees we use do not make any attempt to add in these haloes lost by SUBFIND. However, since we use the most massive subhalo in a FOF group to trace the FOF merger trees, and use $M_{200\mathrm{m}}$ as the mass definition,

our results should not be affected much by small subhaloes being lost by the halo finder.

3.2.5 Haloes below the mass resolution

Populating the resolved haloes in the MXXL halo lightcone with galaxies will result in incompleteness in a magnitude limited galaxy catalogue at low redshifts. This is because intrinsically faint galaxies which are sufficiently close to the observer to be bright enough to be included in the catalogue occupy haloes which fall below the MXXL mass resolution. We use our fit to the MXXL halo mass function in order to add these haloes into the lightcone catalogue, and position them randomly in the catalogue so that they are unclustered. Other methods for augmenting the halo catalogue exist (e.g. de la Torre & Peacock, 2013; Angulo et al., 2014; Benson et al., 2016), but we find that this simple method is able to bring the dN/dz of galaxies in the catalogue into better agreement with the measured dN/dz from GAMA, while only having a very small effect on the measured clustering.

The redshift distribution of the haloes which need to be added to the lightcone catalogue can be calculated from the integral,

$$\frac{dN}{dz} = \int_{M_{\rm min}(z)}^{M_{\rm max}} n_{\rm unres}(M, z) \frac{dV}{dz} dM, \tag{3.1}$$

where $n_{\rm unres}(M, z)$ is the number density of unresolved haloes, dV/dz is the comoving volume per unit redshift, $M_{\rm min}(z)$ is the minimum halo mass that could host a galaxy brighter than the faintest observable galaxy in the survey at that redshift[1], and $M_{\rm max} = 10^{12} h^{-1} M_{\odot}$ is the mass at which the MXXL mass function is judged to be complete. If the survey is flux limited, then the faintest observable galaxy at each redshift is set by an apparent magnitude threshold; for our mock catalogue we set this threshold to $r = 20$, as this is the magnitude threshold for the

[1]The minimum halo mass that can host a galaxy brighter than $r = 20$, $M_{\rm min}(z)$, can be determined from the HODs we use to populate the lightcone. Firstly, the apparent magnitude limit can be converted to an absolute magnitude at redshift z using the k-corrections of Section 3.4.3. It can then be determined, from the HODs as a function of mass and redshift (Section 3.3), the minimum mass required to host a central galaxy brighter than this magnitude.

DESI Bright Galaxy Survey (BGS) (DESI Collaboration et al., 2016a). The number density of unresolved haloes is given by $n_{\rm unres}(M,z) = n_{\rm fit}(M,z) - n_{\rm res}(M,z)$, where $n_{\rm fit}(M,z)$ is our fit to the number density of haloes in the lightcone, extrapolated to low masses, and $n_{\rm res}(M,z)$ is the number density of haloes resolved in MXXL. We model the mass function of resolved haloes by multiplying the fit to the mass function by a cutoff at the mass resolution limit of $M_{\rm 200m} \sim 10^{11}\ h^{-1}{\rm M}_\odot$,

$$n_{\rm unres}(M,z) = [1 - {\rm cut}(M,z)] n_{\rm fit}(M,z), \tag{3.2}$$

where a good approximation to the cutoff is given by

$${\rm cut}(M,z) = 10^{(-z-2)(\log_{10}(M/h^{-1}{\rm M}_\odot)-11)^{0.6}}. \tag{3.3}$$

In order to add unresolved haloes to the catalogue, we first randomly draw a redshift for each unresolved halo from the dN/dz distribution defined in Eq. 3.1. The mass of each halo is then randomly assigned using the mass distribution at the redshift of the halo defined by $n_{\rm unres}(M,z)$. The haloes are then randomly positioned uniformly on the sky. Since the unresolved haloes are randomly positioned so that they are unclustered, redshift space distortions do not affect their clustering, and so we set the velocity of each of these haloes to zero. A random concentration is also assigned from the mass-concentration relation of MXXL (with scatter), extrapolated to lower masses.

While the introduction of unclustered haloes only has a small effect on the two-point correlation function, other statistics might also change, for example, density estimators and void statistics. We have not checked the size of this effect, but the final galaxy catalogue includes a flag which indicates whether a galaxy lives inside one of these haloes, enabling these galaxies to be removed when calculating other statistics.

3.3 Halo occupation distribution

Galaxies are biased tracers of the underlying dark matter density field. The halo occupation distribution (HOD) describes this bias between galaxies and haloes using the probability that a halo of mass M contains N galaxies with a certain property, $P(N|M)$, providing a physical interpretation of galaxy clustering measurements.

The mean number of galaxies in a halo of mass M which are brighter than some luminosity threshold, L, can be written as a sum of central and satellite galaxies (e.g. Zheng et al., 2005),

$$\langle N_{\rm gal}(>L|M)\rangle = \langle N_{\rm cen}(>L|M)\rangle + \langle N_{\rm sat}(>L|M)\rangle. \tag{3.4}$$

We use central and satellite occupation functions of the same form as Zehavi et al. (2011). The mean number of central galaxies brighter than L is described by a smoothed step function,

$$\langle N_{\rm cen}(>L|M)\rangle = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log M_{\rm min}(L)}{\sigma_{\log M}(L)}\right)\right], \tag{3.5}$$

where ${\rm erf}(x)=2\pi^{-1/2}\int_0^x e^{-x'^2}dx'$ is the error function. The parameter $M_{\rm min}$ is the halo mass for which half of haloes contain a galaxy brighter than L, and $\sigma_{\log M}$ sets the width of the step. For $M \gg M_{\rm min}$, $\langle N_{\rm cen}(>L|M)\rangle = 1$, while for $M \ll M_{\rm min}$, $\langle N_{\rm cen}(>L|M)\rangle = 0$. The mean number of satellites per halo brighter than L is given by a power law,

$$\langle N_{\rm sat}(>L|M)\rangle = \langle N_{\rm cen}(>L|M)\rangle \left(\frac{M-M_0(L)}{M_1'(L)}\right)^{\alpha(L)}, \tag{3.6}$$

where M_0 is the cutoff mass scale, M_1' the normalisation, and α the power law slope. M_1' is different to M_1, the mass of a halo which on average contains 1 satellite, although the two quantities are related[1]. The power law is also multiplied by the central occupation function, which ensures that the brightest galaxy in the halo is the central; there cannot be a satellite brighter than L without there first being a central galaxy brighter than L.

[1]Since the satellite occupation function is modified by the centrals, the relation $M_1 = M_1' + M_0$ is not exact.

3.3.1 HODs at low redshift

We use HOD parameters calculated from the SDSS using the procedure of Zehavi et al. (2011). These are calculated for different luminosity threshold galaxy samples, using an MCMC code to find the best fitting HOD parameters which reproduce the measured projected correlation functions to within the SDSS uncertainties. Since the cosmology of the MXXL simulation is different to that used by Zehavi et al. (2011), the parameter fitting was redone using the Millennium cosmology. The SDSS HODs use the absolute r-band magnitude of each galaxy, k-corrected to a reference redshift of $z_{\rm ref} = 0.1$ (see Section 3.4.3), which is the median redshift of the survey. We denote absolute magnitudes k-corrected to this redshift as ${}^{0.1}M_r$. Absolute magnitudes written as ${}^{0.1}M_r$ assume $h = 1$.

The best fitting HOD parameters, in Millennium cosmology, are shown by the points in Fig 3.4. We do not show the errors as they are misleading, due to the probability distributions being highly asymmetric. Projecting these asymmetric probability distributions to 1 dimensional errors can lead to the best fitting values of some of the parameters being outside the error bars, particularly for $\sigma_{\log M}$.

For each HOD parameter, a least squares routine is used to fit a function which describes the variation with luminosity, which are shown by the dashed lines in Fig. 3.4. The top panel shows $M_{\rm min}(L)$ and $M_1'(L)$, for which we fit curves of the same functional form as Eq. 11 from Zehavi et al. (2011),

$$L/L_* = A\left(\frac{M}{M_t}\right)^{\alpha_M} \exp\left(-\frac{M_t}{M} + 1\right), \tag{3.7}$$

where A, M_t and α_M are free parameters. We fit a power law to $M_0(L)$ (second panel). This is a poor fit for the points at ${}^{0.1}M_r = -18$, which is over 3 orders of magnitude lower than the fit, and ${}^{0.1}M_r = -19$, which is 20 orders of magnitude lower. However, increasing the value of this parameter by many orders of magnitude has a very small effect on the shape of the HODs. This is because the occupation function of central galaxies adds a second cutoff to Eq. 3.6; if M_0 is below this

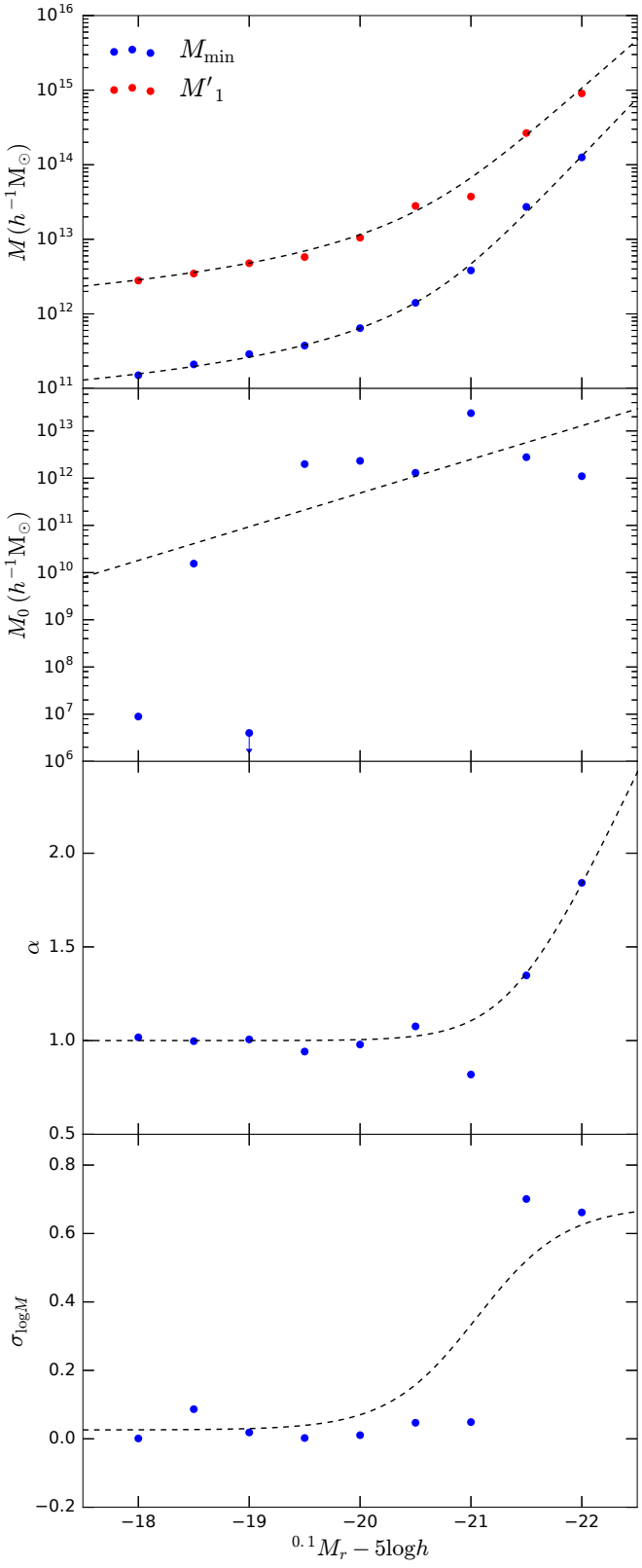

Figure 3.4: Best fitting HOD parameters to the SDSS volume limited samples in Millennium cosmology (points), and smooth functions fitted to these points (dashed lines), as a function of magnitude. Top panel: $M_{\rm min}$ (blue) and M'_1 (red). Second panel: M_0. Third panel: α. Bottom panel: $\sigma_{\log M}$. The $^{0.1}M_r - 5\log h = -19$ sample has $M_0 = 10^{-10.2}\ h^{-1}{\rm M}_\odot$, but M_0 is poorly constrained. Errors are not shown as they are misleading, due to the highly asymmetric probability distributions.

cutoff, it will not affect the shape of the HODs. The parameter $\alpha(L) \sim 1$ at low luminosities, but increases for the highest luminosity samples (third panel). We fit a linear relation, which smoothly transitions to $\alpha = 1$ at low luminosities. $\sigma_{\log M}(L)$ (bottom panel) is small at low luminosities, with a step up to ~ 0.7 for the brightest two samples. We fit a sigmoid function to $\sigma_{\log M}$, where the width of the step is set such that the HODs do not overlap.

The large step in $\sigma_{\log M}(L)$ means that, as the luminosity threshold is increased, there is a rapid jump in the amount of scatter in the luminosities of central galaxies. This results in overlapping HODs, as can be seen for the ${}^{0.1}M_r < -21$ and ${}^{0.1}M_r < -21.5$ samples in Fig. 3.5. For two luminosity thresholds L_1 and L_2, where $L_1 < L_2$, it must be true that $\langle N_{\rm gal}(> L_1|M)\rangle \geq \langle N_{\rm gal}(> L_2|M)\rangle$ since all galaxies brighter than L_2 are also brighter than L_1. However, if the two occupation functions cross, then this condition is not satisfied for haloes below the mass at which they cross. This is unphysical, as it would require a negative number of galaxies within these luminosity thresholds. We therefore must model the HODs such that there is no overlap. There exist HOD frameworks in which the occupation functions cannot overlap (see e.g. Leauthaud et al., 2011), but since we are using HOD parameters obtained using the standard HOD framework, we make a small modification to these HODs to prevent any crossing, as set out below. The HOD model we use assumes that the occupation function of galaxies depends on halo mass only, but it could also depend on some other halo property, x (e.g. formation time or halo concentration). However, this cannot solve the problem of unphysical crossing, since for the total HOD to cross, the HOD as a function of x would also have to cross for some values of x.

Eq. 3.5 assumes that the scatter set by the parameter $\sigma_{\log M}(L)$ is Gaussian. Since a Gaussian function has a long tail which extends to infinity, there will always be an overlap between HODs if $\sigma_{\log M}(L_2) > \sigma_{\log M}(L_1)$. We instead approximate

the Gaussian by using a spline kernel (Schoenberg, 1946),

$$\text{spline}(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & |x| \leq 0.5 \\ 2(1-|x|)^3 & 0.5 < |x| \leq 1 \\ 0 & |x| > 1, \end{cases} \tag{3.8}$$

which has spline(0) = 1, mean = 0, variance = 1/12, and spline(x) = 0 for $|x| > 1$. This function can be rescaled and normalised to approximate any Gaussian of mean μ and variance σ^2 as,

$$S(x) = \frac{4/3}{\sigma\sqrt{12}}\text{spline}\left(\frac{x-\mu}{\sigma\sqrt{12}}\right). \tag{3.9}$$

The HOD for central galaxies can therefore be written as

$$\langle N_{\text{cen}}(> L|M)\rangle = \frac{1}{2}\left[1 + F\left(\frac{\log M - \log M_{\text{min}}(L)}{\sigma_{\log M}(L)}\right)\right], \tag{3.10}$$

where $F(x) = 2\int_0^x S(x')dx'$. The best fitting values of $\sigma_{\log M}$ (shown by the points in the bottom panel of Fig. 3.4), suggest a sharp step between ${}^{0.1}M_r = -21$ and ${}^{0.1}M_r = -21.5$. Even using Eq. 3.10, the HODs will still overlap with this abrupt step, but they will not overlap if the step is gradual, unlike Eq. 3.5. We make the step in $\sigma_{\log M}(L)$ as narrow as we can while preventing the HODs crossing (shown by the dashed curve).

The HODs using the SDSS HOD parameters, and our fits, are shown in Fig. 3.5. Our fits produce halo occupation functions which are in reasonable agreement with the SDSS HODs, with the exception of the ${}^{0.1}M_r < -21$ and ${}^{0.1}M_r < -21.5$ samples, where the width of the step set by the parameter $\sigma_{\log M}$ is too broad and narrow respectively. This is necessary to prevent the HODs from overlapping. The SDSS HOD for the ${}^{0.1}M_r < -21$ sample appears to have a sharp transition from centrals to satellites, which is due to a large value of M_0 compared to M_1'.

3.3.2 Redshift evolution

In order to evolve the occupation functions with redshift, we first choose a target luminosity function, $\phi_{\text{target}}(L, z)$, that we would like the galaxies in the mock catalogue to reproduce as a function of redshift. This luminosity function defines a

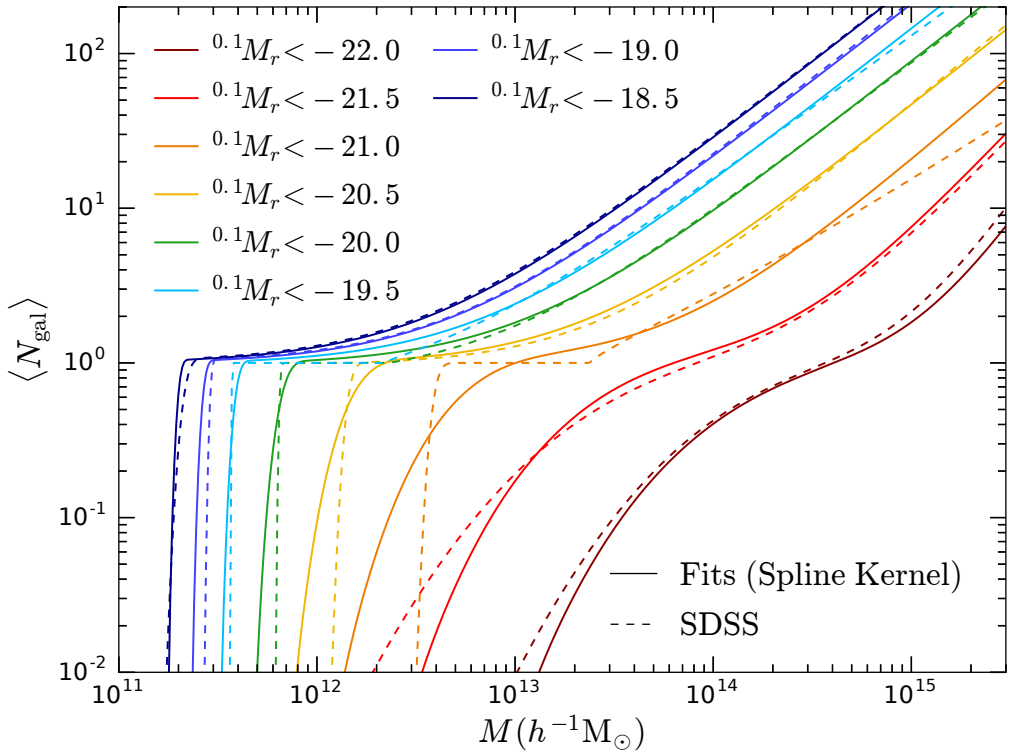

Figure 3.5: Mean halo occupation functions for luminosity threshold samples, as described by Eqs. 3.4-3.6, using SDSS HOD parameters in the Millennium cosmology (dashed lines) and our fits to the HOD parameters, using Eq. 3.10 in place of Eq. 3.5 to describe the contribution from central galaxies (solid lines). Colours indicate the luminosity threshold, as shown by the legend.

mapping between a luminosity threshold L at redshift z, and the number density of galaxies brighter than this, $n_{\rm gal}^{\rm target}(> L, z)$.

For a given HOD, the number density of galaxies brighter than L can be calculated from the integral

$$n_{\rm gal}(> L, z) = \int n_{\rm halo}(M, z) \langle N(> L|M, z) \rangle dM, \tag{3.11}$$

where $n_{\rm halo}(M, z)$ is the number density of haloes of mass M at redshift z, and $\langle N(> L|M, z) \rangle$ is the halo occupation function at redshift z. The HODs must evolve with redshift such that the condition $n_{\rm gal}(> L, z) = n_{\rm gal}^{\rm target}(> L, z)$ is satisfied.

Since the target luminosity function defines a mapping between a luminosity

threshold and the number density of galaxies, the occupation functions can be rewritten as a function of number density, $n_{\rm gal}$: $\langle N(>L|M,z)\rangle \equiv \langle N(n_{\rm gal}|M,z)\rangle$. The shape of the HOD could evolve in a complex way, but for simplicity we keep the shape of the occupation function fixed for constant galaxy number density, but slide the HODs along the halo mass axis such that the target luminosity function is achieved. That is, the HOD parameters $\sigma_{\log M}(n_{\rm gal},z)$ and $\alpha(n_{\rm gal},z)$ are kept constant, but the 3 mass parameters $M_{\rm min}$, M_0 and M_1 are all multiplied by some factor f,

$$M_{\rm HOD}(n_{\rm gal},z) = f(n_{\rm gal},z)M_{\rm HOD}(n_{\rm gal}), \tag{3.12}$$

where $M_{\rm HOD}$ is one of the HOD mass parameters. The value of f required to achieve the target luminosity function is found by finding the root of the equation

$$n_{\rm gal}(>L, f(z)) - n_{\rm gal}^{\rm target}(>L,z) = 0. \tag{3.13}$$

At high redshifts, the target luminosity function we use is the evolving Schechter function fit to the luminosity function estimated from the Galaxy and Mass Assembly (GAMA) survey. The Schechter function can be written in terms of magnitudes as

$$\phi(M) = 0.4\ln 10\,\phi^*(10^{0.4(M^*-M)})^{1+\alpha}\exp(-10^{0.4(M^*-M)}), \tag{3.14}$$

where ϕ^* is the normalisation, M^* is a characteristic magnitude and α is the faint end slope. For GAMA, Loveday et al. (2012, 2015) model the evolution of the Schechter parameters with redshift as

$$\begin{aligned} \alpha(z) &= \alpha(z_0) \\ M^*(z) &= M^*(z_0) - Q(z-z_0) \\ \phi^*(z) &= \phi^*(0)10^{0.4Pz}, \end{aligned} \tag{3.15}$$

where Q parametrises the evolution in luminosity, P parametrises the evolution in number density, and $z_0 = 0.1$ is the same reference redshift as used for the k-corrections (Section 3.4.3). The faint end slope is kept constant with redshift since

there is not enough data to constrain it at high redshifts. We use the evolving Schechter function, $\phi_{\rm GAMA}$, from Loveday et al. (2012) with $P = 1.8$ and $Q = 0.7$.

However, the shape of the GAMA Schechter luminosity function is slightly different than the SDSS luminosity function. Using it as the target at all redshifts would result in the evolution parameter $f \neq 1$ at $z = 0.1$, meaning that the HODs would change from the HODs measured from SDSS. In order to not change the HODs at $z = 0.1$, we use the luminosity function from SDSS, $\phi_{\rm SDSS}$, as the target at low redshifts. The SDSS target luminosity function we use is the result of the integral

$$\phi_{\rm SDSS}(> L) = \int n_{\rm halo}(M)\langle N(> L|M)\rangle dM, \tag{3.16}$$

where $n_{\rm halo}(M)$ is the number density of haloes at $z = 0.1$, and $\langle N(> L|M)\rangle$ is the (unevolved) occupation function. The result of this integral is close to the Blanton et al. (2003) luminosity function for absolute magnitudes brighter than ${}^{0.1}M_r = -19$, and by definition $f = 1$, so the HODs remain unchanged from SDSS at this redshift. However, at magnitudes fainter than ${}^{0.1}M_r = -19$, the result of this integral is very flat, while the Blanton et al. (2003) SDSS luminosity function is steeper; at ${}^{0.1}M_r = -17$ they differ by a factor of ~ 2. We therefore smoothly transition to the Blanton et al. (2003) luminosity function at ${}^{0.1}M_r = -19$. This is then extrapolated to fainter magnitudes with a power law.

We interpolate the target luminosity function from $\phi_{\rm SDSS}$ at low redshifts to $\phi_{\rm GAMA}$ at high redshifts,

$$\phi_{\rm target}(M, z) = (1 - w(z))\phi_{\rm SDSS}(M, z) + w(z)\phi_{\rm GAMA}(M, z), \tag{3.17}$$

where the transition between $0.1 < z < 0.2$ is set by the sigmoid function

$$w(z) = (1 + e^{-100(z-0.15)})^{-1}. \tag{3.18}$$

The evolution parameter, f, for this target luminosity function, is shown in Fig. 3.6 as a function of magnitude for different redshifts. At $z = 0.1$, f is close to 1, by definition. However, it is not exactly 1 because the function $w(z)$, which

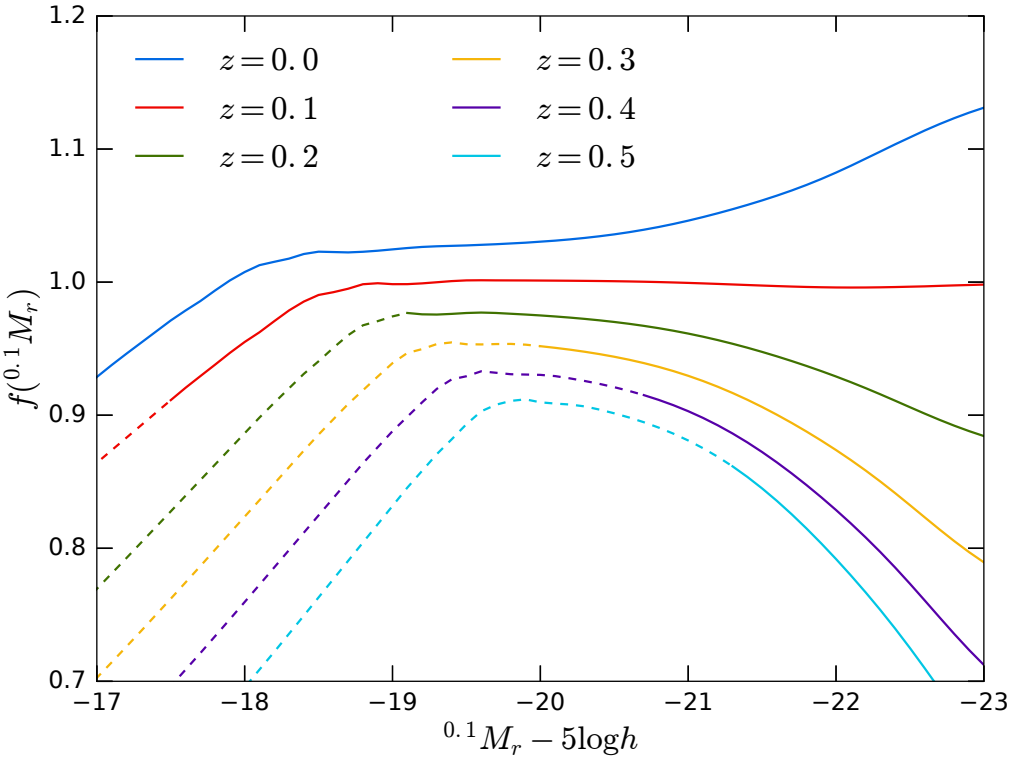

Figure 3.6: Evolution parameter, f, as a function of magnitude for different redshifts, as indicated by the colour. This is the factor by which the HOD mass parameters are multiplied in order to achieve the galaxy number density set by the target luminosity function. Dashed lines indicate absolute magnitudes which correspond to apparent magnitudes that are fainter than the $r = 20$ limit at that redshift.

sets the transition between the two target luminosity functions is close to, but not exactly 0 at $z = 0.1$. At $z = 0.1$, f is equal to 1 to within 1%. Fainter than magnitude -19, $f(z = 0.1) < 1$. At these faint magnitudes, the target luminosity function is transitioning to the Blanton et al. (2003) luminosity function. Keeping $f(z = 0.1) = 1$ at all magnitudes produces a luminosity function which, while being close to SDSS at the bright end, is too flat at the faint end, so this transition is required to bring the luminosity function of the mock into better agreement with the data.

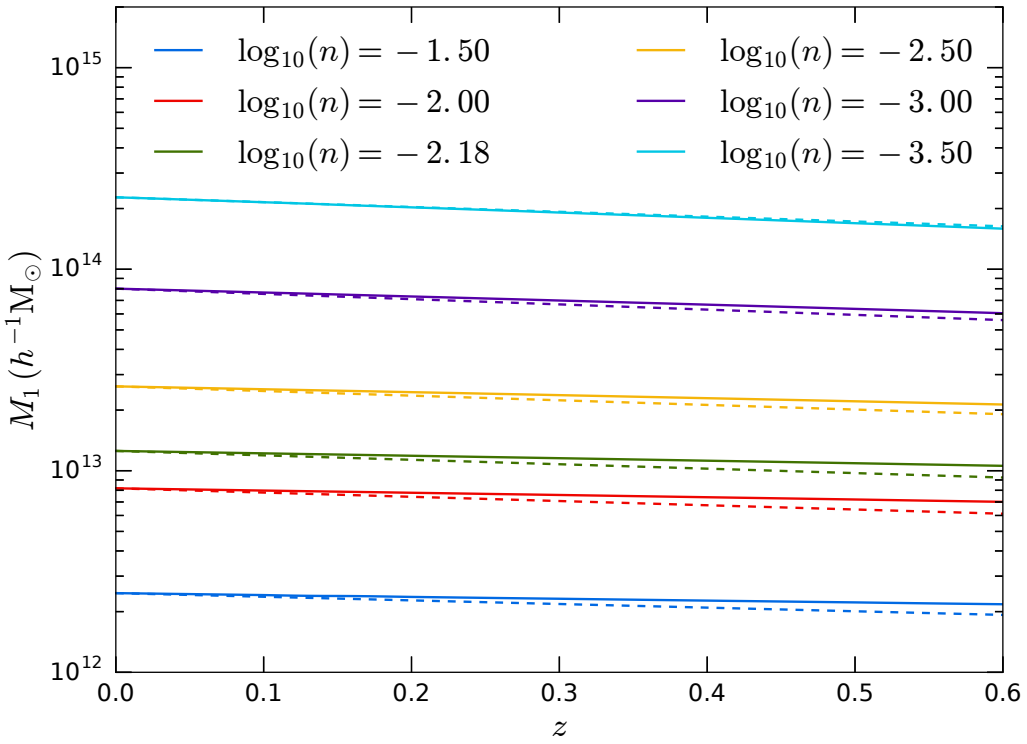

Figure 3.7: Evolution of the HOD parameter M_1 with redshift for galaxy samples of a fixed number density, where number densities, n, are in units of $h^3\mathrm{Mpc}^{-3}$. Solid lines show the evolution in the mock, as determined from the target luminosity function. Dashed lines start at the same $M_1(z = 0)$ as in the mock, but show the evolution found in Contreras et al. (2017), as predicted from the Gonzalez-Perez et al. (2014) version of the GALFORM semi-analytic galaxy formation model.

The evolution of the parameter M_1 implied by this evolution of f is shown in Fig. 3.7 for galaxy samples of a fixed number density, up to $z = 0.6$. Since the shape of the HODs are kept fixed for a fixed number density, but the HODs are evolved along the mass axis, the other mass parameters $M_{\rm min}$ and M_0 show the same evolution, while $\sigma_{\log M}$ and α are held constant. For comparison, we also show the evolution reported in Contreras et al. (2017) from their fit to the evolution found in the Gonzalez-Perez et al. (2014) version of the GALFORM semi-analytic galaxy formation model (Cole et al., 2000). We find that M_1 decreases slightly with redshift, in remarkably close agreement with what is found in Contreras et al.

(2017), although the highest number density samples show slightly less evolution. By construction, the ratio of the parameters $M_1/M_{\rm min}$ is kept constant in the mock. This is in contrast to the behaviour found in Contreras et al. (2017), where they reported that this ratio decreases over the same redshift range. The amount by which this ratio decreases depends on the semi analytic model used, and on the number density of galaxies; at most it decreases by $\sim 50\%$. The evolution of the HOD model could be extended to include this change in $M_1/M_{\rm min}$, but we find that simply keeping the mass ratio fixed produces a good match to the measured clustering (see Fig. 3.12).

3.4 Mock galaxy catalogue

Now we describe in detail the HOD method used to populate the halo catalogue with galaxies, assign each galaxy a luminosity and ${}^{0.1}(g-r)$ colour and compare the resultant clustering in the mock with measurements from SDSS and GAMA. These details can be skipped by the reader, but we give a brief summary below.

Section 3.4.1 describes the HOD method for populating the halo lightcone with galaxies with luminosities. This Monte Carlo method is based on Skibba et al. (2006), but extended to an evolving 5 parameter HOD. The number and luminosity of galaxies in each halo are randomly generated such that the input HODs are reproduced. Central galaxies are assigned the position and velocity of the halo, and satellites are randomly positioned around the central, following an NFW density profile, and assigned a random virial velocity.

The method for assigning a ${}^{0.1}(g-r)$ colour to each galaxy is described in Section 3.4.2. This is based on Skibba & Sheth (2009), and randomly assigns a colour from a parametrisation of the SDSS colour magnitude diagram. Section 3.4.2 describes our modification to the parametrisation of the colour magnitude diagram, which includes evolution, and is in agreement with measurements from GAMA. The colour assigned to each galaxy depends only on its luminosity, its redshift,

and whether it is a central or satellite galaxy; there is no explicit dependence on halo mass. Colour-dependent k-corrections derived from GAMA are described in Section 3.4.3. The colour-dependent clustering of galaxies in the mock is shown in Section 3.4.4.

3.4.1 Constructing the galaxy catalogue

We use a modified version of the method of Skibba et al. (2006) to populate the halo lightcone catalogue with galaxies, and to assign each galaxy an r-band absolute magnitude, k-corrected to $z = 0.1$. Skibba et al. (2006) use a 3 parameter HOD in which the occupation function of central galaxies is simply a step function; we have extended this method in order to reproduce the 5 parameter HOD given by Eq. 3.6 & 3.10, which adds scatter to the luminosity of central galaxies, as required by the SDSS clustering data. We also use the fits to the HOD parameters as a function of luminosity as described in Section 3.3.1 and shown in Fig. 3.4. To be consistent with the mass definition used in Zehavi et al. (2011), we take the halo mass to be $M_{200\mathrm{m}}$; i.e. the mass enclosed by a sphere in which the average density is 200 times the mean density of the Universe.

For each halo, a number, x, is randomly drawn from the spline kernel probability distribution, $S(x)$ (Eq. 3.9), with $\mu = 0$ and $\sigma = 1$. This introduces the scatter in the luminosity of the central galaxy, relative to the average luminosity in a halo of this mass. The luminosity L which is required to produce this scatter is found by solving $x\sigma_{\log M}(L)/\sqrt{2} = \log M - \log M_{\mathrm{min}}(L)$, where the factor of $\sqrt{2}$ comes from how $\sigma_{\log M}$ is defined. Finally the central galaxy is positioned at the centre of the halo, with the same velocity.

To populate a halo with satellite galaxies, a minimum luminosity, L_{min}, must first be chosen. We vary L_{min} with redshift, choosing it to be slightly fainter than the luminosity corresponding to $r = 20$. This ensures that the final mock catalogue is complete to $r = 20$ at all redshifts, while preventing galaxies that are

too faint to be observed being unnecessarily added to the catalogue. The number of satellite galaxies to be added to each halo is drawn from a Poisson distribution with mean $\langle N_{\rm sat}(> L_{\rm min}|M)\rangle$, which is given by Eq. 3.6. For each satellite, a uniform random number $0 < u < 1$ is drawn, and the luminosity is found such that $\langle N_{\rm sat}(> L|M)\rangle / \langle N_{\rm sat}(> L_{\rm min}|M)\rangle = u$. The satellite galaxies are assigned a random virial velocity, relative to the velocity of the central galaxy, which is drawn from a Maxwell-Boltzmann distribution with a line of sight velocity dispersion

$$\sigma^2(M) = \frac{GM_{\rm 200m}}{2R_{\rm 200m}}, \tag{3.19}$$

where $R_{\rm 200m}$ is the radius of the sphere, centred on the halo, in which the enclosed density is 200 times the mean density of the Universe. Finally, the satellite galaxies are positioned randomly around the centre of the halo such that they follow an NFW (Navarro et al., 1997) density profile, which is truncated at $R_{\rm 200m}$. We find that using the same concentration, c, as the halo, calculated from $c = 2.16 R_{\rm 200m}/R_{\rm Vmax}$, where $R_{\rm Vmax}$ is the radius at which the maximum circular velocity occurs, produces angular clustering which is too strong at small angular scales compared to SDSS (Wang et al., 2013). This can be improved by reducing the concentration of all haloes by a factor of 2 (see Fig. 3.8). We therefore use these reduced concentrations when positioning satellite galaxies inside each halo.

The HODs from Zehavi et al. (2011) were fit to the projected correlation functions, using the mass-concentration relation of Bullock et al. (2001), modified to be consistent with their mass definition. This mass-concentration relation is close to what is seen in MXXL. However, modifying the concentrations only has a small effect on the 1-halo term of the projected correlation functions. Down to separations of 0.1 h^{-1}Mpc, the clustering in the mock catalogue only changes by a small amount. It is the change in the clustering at physical scales smaller than this which causes the small scale angular clustering to improve, and this is below the scale at which the projected correlation functions were measured in SDSS. The angular correlation function, $\omega(\theta)$, in the mock catalogue is shown in Fig. 3.8 for galaxies in bins of apparent magnitude, compared to the angular clustering measured in

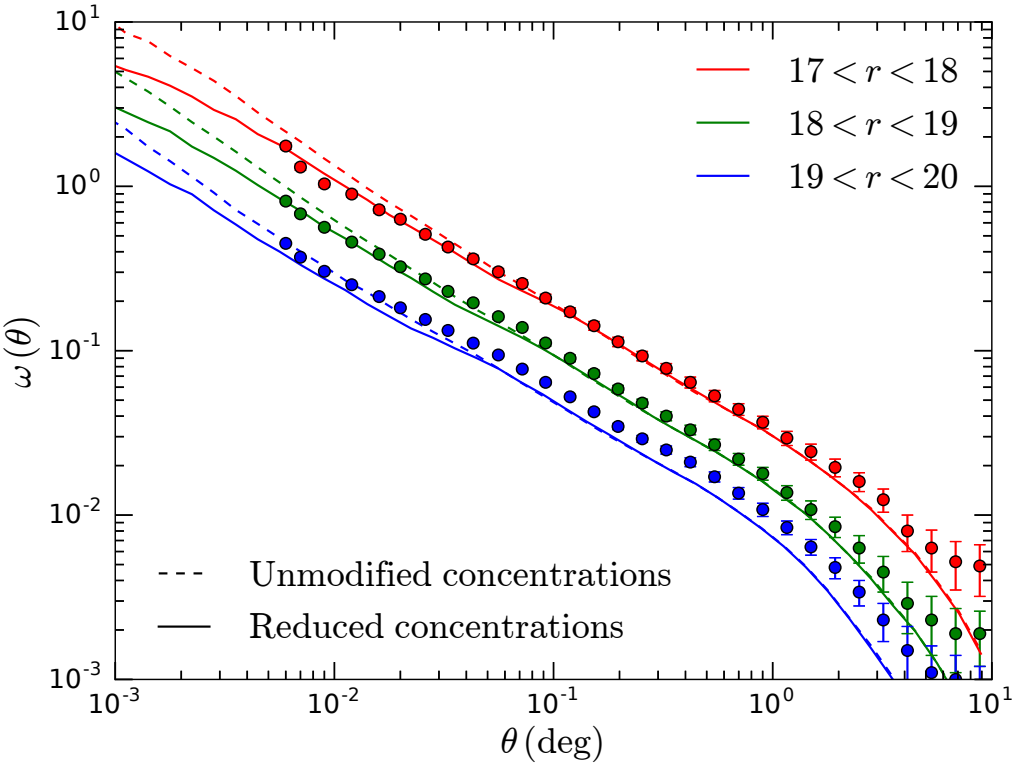

Figure 3.8: Angular clustering of galaxies in the mock catalogue in bins of apparent magnitude, as labelled (coloured lines). Points with error bars show the angular clustering of galaxies measured in the SDSS (Wang et al., 2013). Dashed curves show the angular clustering where satellite galaxies are positioned such that they follow an NFW density profile with the same, unmodified concentration as the halo. Solid curves show the resulting angular clustering when halo concentrations are reduced by a factor of 2.

SDSS (Wang et al., 2013). Solid lines show the clustering in the mock with concentrations reduced by a factor of 2, which is in good agreement with SDSS down to a small angular separation of 20 arcsec, although for the faintest sample the clustering is a little low. Using unmodified concentrations results in $\omega(\theta)$ having a slope which is steeper than the SDSS measurements, shown by the dashed curves, resulting in clustering which is too strong at small angular scales. The introduction of unclustered haloes below the mass resolution has the effect of reducing the clustering in the mock, but as we show later in Section 3.4.1.3, this effect is small.

3.4.1.1 The luminosity function of the mock

The Petrosian r-band luminosity function of the galaxy catalogue is shown in Fig. 3.9 for galaxies in three redshift bins. The dashed lines show the target luminosity at the median redshift of each bin, showing that this evolving target luminosity function is reproduced in the mock catalogue. The smaller panel in Fig. 3.9 compares the luminosity function in the mock at low redshifts with the Blanton et al. (2003) luminosity function from SDSS. Brighter than ${}^{0.1}M_r = -19$, the luminosity function in the mock is in good agreement with SDSS, which indirectly shows that the mass function of the MXXL lightcone is close to the Jenkins et al. (2001) mass function assumed by Zehavi et al. (2011), and our fits to the HOD parameters as a function of luminosity are a good approximation to the actual values. Fainter than ${}^{0.1}M_r = -19$, the luminosity functions agree by construction.

3.4.1.2 The redshift distribution of the mock

The redshift distribution of galaxies brighter than an apparent magnitude limit of $r = 19.8$ is shown in Fig. 3.10 (see Section 3.4.3 for the k-corrections used), and compared to the GAMA survey. The dN/dz of the mock catalogue is in good agreement with GAMA, within 15% of the fitted curve at most redshifts. Without adding in the low mass, unresolved haloes at low redshifts (see Section 3.2.5),

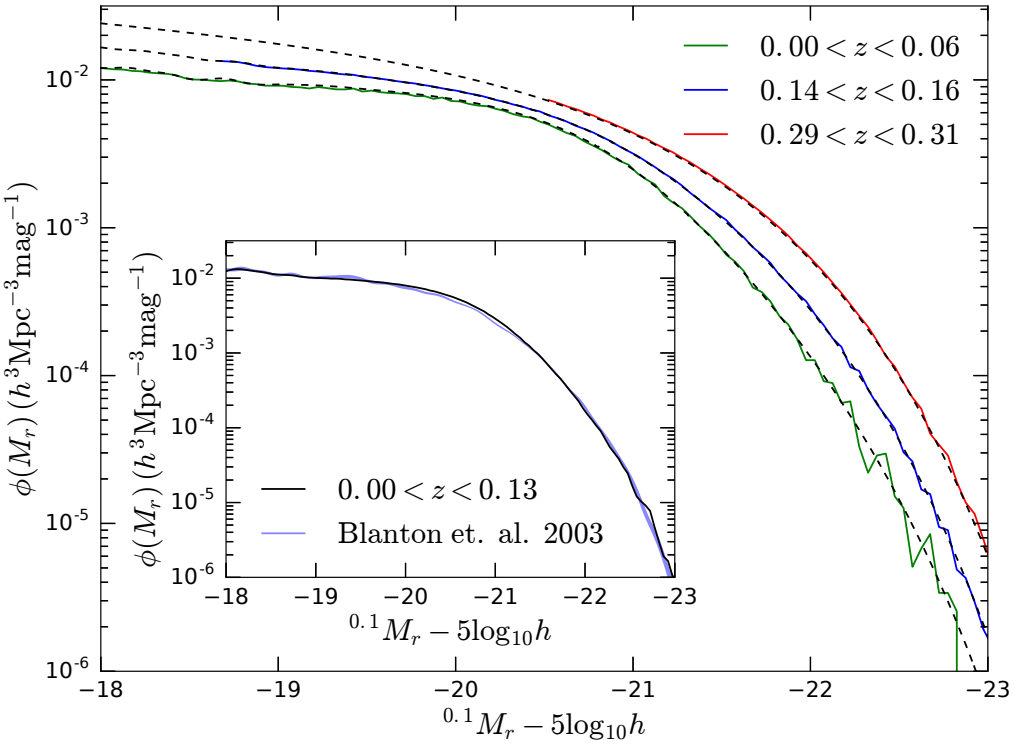

Figure 3.9: The r-band luminosity function of galaxies in the mock catalogue in different redshift bins, as indicated by the legend. Dashed lines indicate the target luminosity function at the median redshift in each bin, which transitions from the SDSS luminosity function at $z < 0.1$ to the GAMA luminosity function at $z > 0.2$. The smaller panel shows the luminosity function in the mock catalogue over the redshift range $0 < z < 0.13$, compared to the SDSS luminosity function of Blanton et al. (2003).

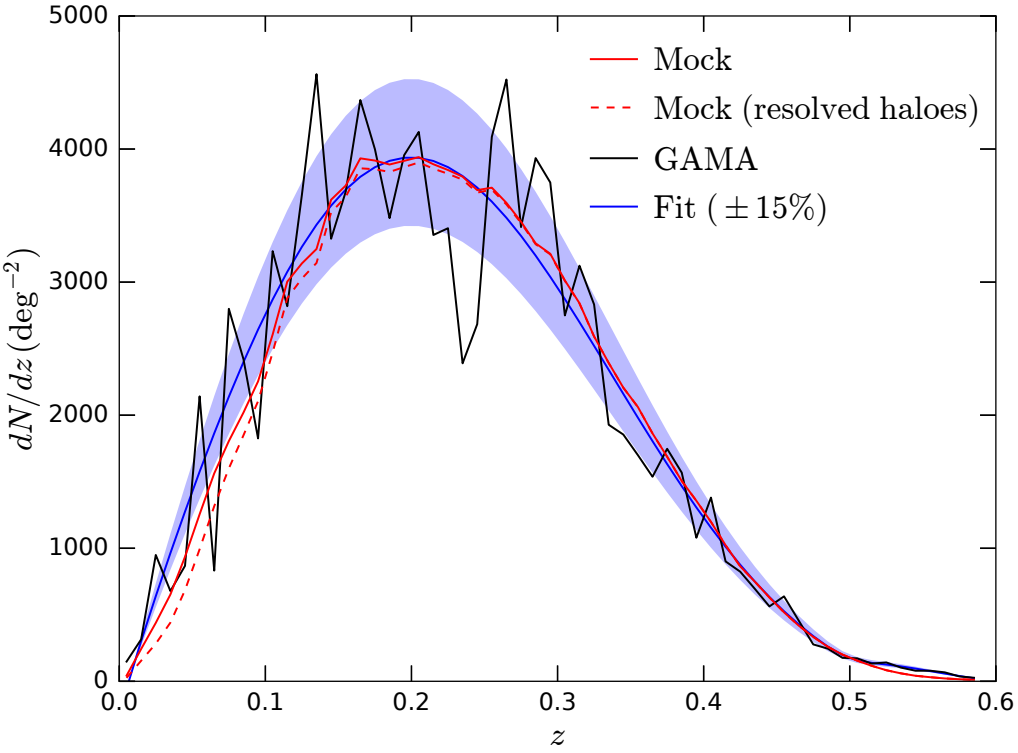

Figure 3.10: dN/dz of galaxies in the mock catalogue with $r < 19.8$ (red), compared to GAMA (black). The solid red curve shows the redshift distribution of all galaxies, including those residing in unresolved haloes below the MXXL mass resolution, while the dashed red curve only includes galaxies residing in resolved haloes. The blue curve shows a fit to the GAMA dN/dz, where the shaded region indicates ±15%.

there is a deficit in the dN/dz for $z \lesssim 0.1$ (dashed red curve); adding in these haloes increases the number of low redshift haloes, bringing the dN/dz into better agreement with GAMA (solid red curve).

3.4.1.3 Clustering of the mock

Projected correlation functions of galaxies in the mock catalogue are shown by the solid curves in Fig. 3.11 for different luminosity threshold samples at $z \sim 0.1$, where we have calculated the two point correlation functions using the publicly available

code CUTE (Alonso, 2012)[1]. These are compared to the measured clustering from SDSS (points with error bars), and the clustering predicted by the best fitting HODs (dashed lines). We use the same redshift ranges as the SDSS volume limited luminosity threshold samples (see table 2 in Zehavi et al., 2011). To be consistent with the definition of magnitude used in Zehavi et al. (2011), magnitudes are evolved to $z = 0.1$ using the evolution model $E(z) = Q_0(1 + Q_1(z - z_0))(z - z_0)$, where $Q_0 = 2$, $Q_1 = -1$ and $z_0 = 0.1$. The clustering in our galaxy catalogue is in reasonable agreement with the projected correlation functions measured from SDSS.

The small differences in the large-scale clustering between the mock catalogue and the clustering predicted by the best fitting HODs can be understood by comparing the HODs in Fig. 3.5. For example, the ${}^{0.1}M_r < -19$ sample is slightly less clustered in the mock. The fit to the HOD has a smaller $M_{\rm min}$ than the best fitting SDSS HOD, meaning that this sample contains more low mass haloes. These haloes are less biased, and therefore the clustering is reduced compared to SDSS. The ${}^{0.1}M_r < -22$ sample contains more high mass haloes, and should therefore be more clustered than SDSS, but the opposite is seen. This is because the brightest samples cover a wider redshift range, and are affected more by the evolution of the HODs. The clustering of the ${}^{0.1}M_r < -22$ sample at small scales is also affected by the evolution of the HODs over this wide redshift range.

The clustering of galaxies in the mock catalogue is also affected by the introduction of haloes below the MXXL mass resolution, which are unclustered. Adding these haloes will therefore have the effect of reducing the measured galaxy clustering. The galaxies which reside in these haloes are faint, and have low redshifts, and so the faintest galaxy samples in Fig. 3.5 are affected by this more than the bright samples. For the ${}^{0.1}M_r < -18.5$ sample, we illustrate the size of this effect: the magenta dashed curve shows the projected correlation function with galaxies residing in unresolved haloes omitted. Including these galaxies reduces the clustering,

[1] `http://members.ift.uam-csic.es/dmonge/CUTE.html`

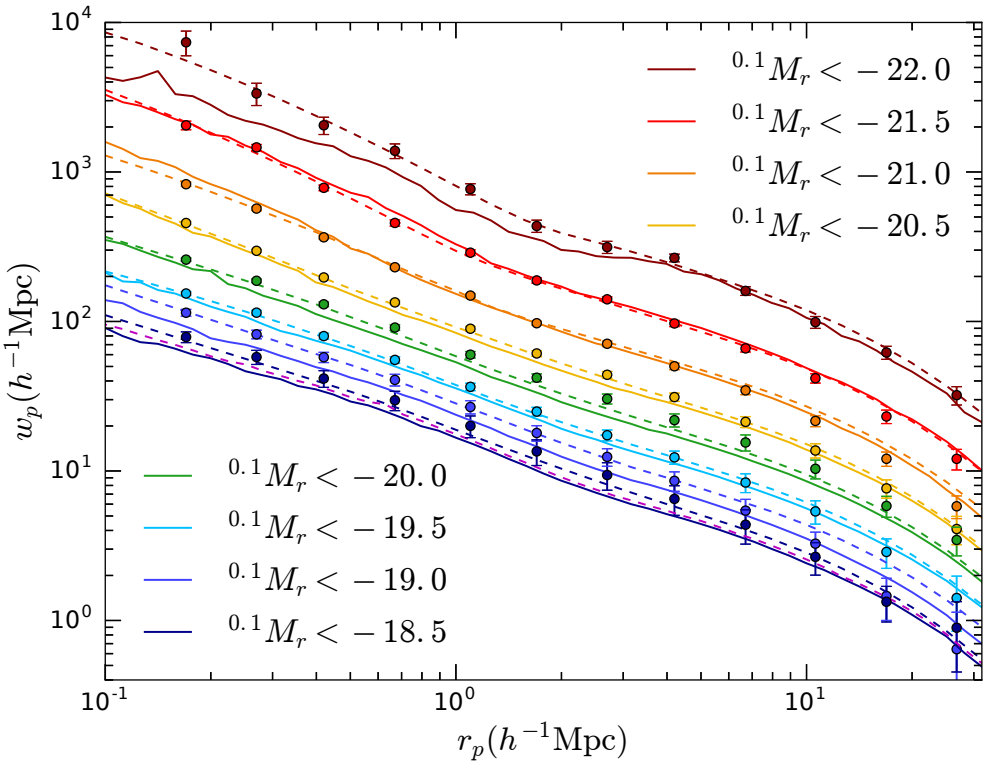

Figure 3.11: Projected correlation functions from the galaxy catalogue (solid lines), compared to the projected correlation functions from SDSS (Zehavi et al., 2011) (points with error bars) and the projected clustering predicted using the best fitting HODs in Millennium cosmology (dashed lines), for different luminosity threshold samples, as indicated by the legend. For the $^{0.1}M_r < -18.5$ sample, we also show the projected clustering in the galaxy catalogue omitting all galaxies which reside in unresolved, unclustered haloes (magenta dashed line). For clarity, the results have been offset by successive intervals of 0.15 dex, starting at the $^{0.1}M_r < -20.5$ sample.

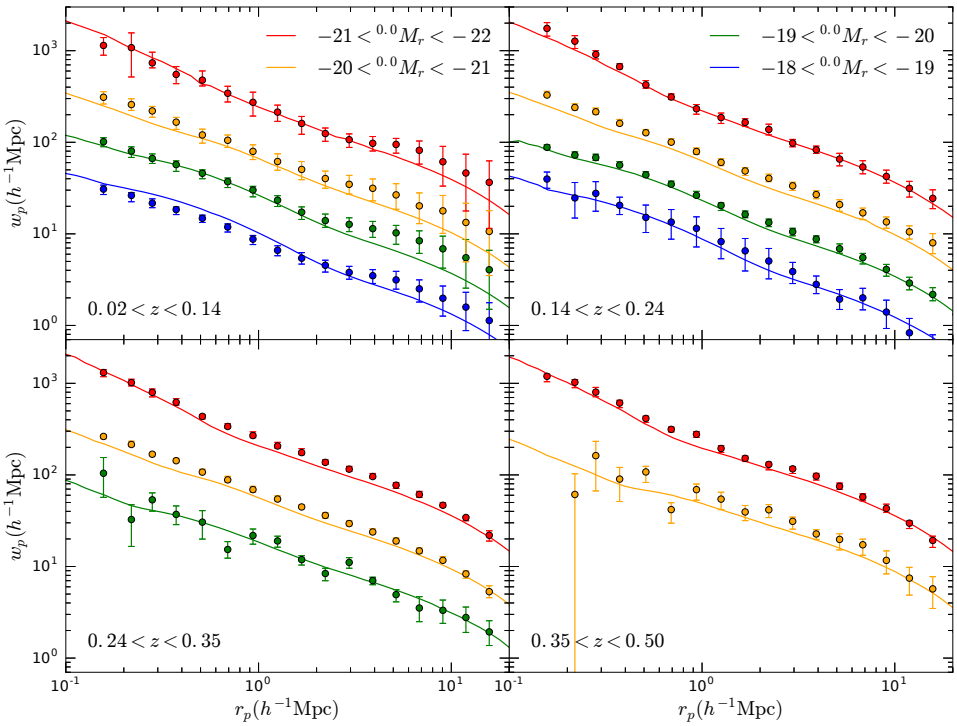

Figure 3.12: Projected correlation functions in different redshift bins for galaxies in the mock catalogue (lines). Points with error bars show the clustering of galaxies from GAMA (Farrow et al., 2015). Different colours indicate bins in ${}^{0.0}M_r$ absolute magnitude. Lines are offset by 0.4 dex relative to the $-21 <^{0.0} M_r < -20$ samples, for clarity.

but only by a very small amount.

We have checked that if we modify our fits to the HOD parameters to agree exactly with the best fitting SDSS parameters at one magnitude, and do not evolve the HODs, we reproduce the SDSS correlation functions very closely for that magnitude limit.

In Fig. 3.12 we show the projected correlation functions in the mock catalogue at high redshifts, compared to the clustering measured in GAMA by Farrow et al. (2015), which has a high completeness of galaxy pairs (Robotham et al., 2010). Here, the magnitude ranges are defined for magnitudes k-corrected to a reference

redshift of $z = 0$, denoted as ${}^{0.0}M_r$, which have also had evolutionary corrections applied. To be consistent with Farrow et al. (2015), we use the same evolutionary correction $E(z) = -Q(z - z_{\rm ref}) = -1.45z$ for this comparison. We find that the clustering at high redshifts is in good agreement with the clustering seen in GAMA. The agreement is least good in the lowest redshift bin, but at low redshifts we find good agreement with SDSS, which covers a much larger area of the sky than GAMA.

3.4.2 Assigning colours

We use the method of Skibba & Sheth (2009) to assign each galaxy a ${}^{0.1}(g - r)$ colour, where g and r are SDSS DR7 model magnitudes (Abazajian et al., 2009; Baldry et al., 2010). This method parametrises the red and blue sequence of the colour-magnitude diagram as two Gaussians with a mean and rms that are linear functions of magnitude. A galaxy is randomly chosen to be red or blue, then a colour is drawn from the appropriate Gaussian. We have modified the parametrisation of the colour-magnitude diagram given in Skibba & Sheth (2009) to bring the faint end into agreement with the colour-magnitude diagram from GAMA, and to add evolution. For those interested, this is described in detail below. For clarity, the 0.1 superscript has been omitted from the equations in the following subsections.

3.4.2.1 Low redshift

For redshifts $z < 0.1$, the parametrisation of Skibba & Sheth (2009) produces a good approximation to the SDSS colour-magnitude diagram, which we summarise below. However, we make some slight modifications to the parametrisation to also bring it into agreement with GAMA at faint magnitudes.

The mean and rms of the red and blue sequences are given by

$$\begin{aligned} \langle g-r|M_r\rangle_{\rm red}^{\rm Skibba} &= 0.932 - 0.032(M_r+20) \\ {\rm rms}(g-r|M_r)_{\rm red}^{\rm Skibba} &= 0.07 + 0.01(M_r+20), \end{aligned} \tag{3.20}$$

and

$$\begin{aligned} \langle g-r|M_r\rangle_{\rm blue}^{\rm Skibba} &= 0.62 - 0.11(M_r+20) \\ {\rm rms}(g-r|M_r)_{\rm blue}^{\rm Skibba} &= 0.12 + 0.02(M_r+20). \end{aligned} \tag{3.21}$$

The total fraction of galaxies which are blue is also parametrised as a linear function of magnitude, given by

$$f_{\rm blue}^{\rm Skibba}(M_r) = 0.46 + 0.07(M_r+20). \tag{3.22}$$

These relations from Skibba & Sheth (2009) produce a colour-magnitude diagram which is in good agreement with SDSS at the bright end. However, the faint end does not agree with what is seen in GAMA (e.g. the first panel in figure 6 of Loveday et al., 2012). Firstly, at ${}^{0.1}M_r = -16$, all the galaxies should lie on the blue sequence, while the fraction of galaxies which are blue given by Eq. 3.22 is 0.74. At faint magnitudes, we instead use a blue fraction given by

$$f_{\rm blue}^{\rm faint}(M_r) = 0.4 + 0.2(M_r+20), \tag{3.23}$$

so the total fraction of blue galaxies is

$$f_{\rm blue}(M_r) = \max\{f_{\rm blue}^{\rm faint}(M_r), f_{\rm blue}^{\rm Skibba}(M_r)\}; \tag{3.24}$$

$f_{\rm blue}(M_r)$ is capped so it is always in the range $0 \leq f_{\rm blue}(M_r) \leq 1$. Another issue with the parametrisation of Skibba & Sheth (2009) is that faint galaxies which lie on the blue sequence are too blue in comparison to the galaxies in GAMA. At ${}^{0.1}M_r = -18.7$, we transition to a flatter blue sequence, given by

$$\langle g-r|M_r\rangle_{\rm blue}^{\rm faint} = 0.4 - 0.03(M_r+16). \tag{3.25}$$

If the fraction of satellite galaxies that are blue, $f_{\rm sat}^{\rm blue}(M_r)$, is specified, then the mean colour of satellite galaxies is given by

$$\langle g-r|M_r\rangle_{\rm sat} = f_{\rm sat}^{\rm blue}(M_r)\langle g-r|M_r\rangle_{\rm blue} + (1-f_{\rm sat}^{\rm blue}(M_r))\langle g-r|M_r\rangle_{\rm red}. \quad (3.26)$$

Conversely, Eq. 3.26 can be rearranged, and the mean satellite colour can be used to specify the fraction of satellites that are blue,

$$f_{\rm sat}^{\rm blue}(M_r) = \frac{\langle g-r|M_r\rangle_{\rm sat} - \langle g-r|M_r\rangle_{\rm red}}{\langle g-r|M_r\rangle_{\rm blue} - \langle g-r|M_r\rangle_{\rm red}}, \quad (3.27)$$

(equation 8 from Skibba & Sheth, 2009, but for blue galaxies). The average colour of a satellite galaxy is parametrised by Skibba & Sheth (2009) as

$$\langle g-r|M_r\rangle_{\rm sat}^{\rm Skibba} = 0.83 - 0.08(M_r+20). \quad (3.28)$$

Modifying the mean satellite colour has the effect of changing the strength of the colour dependent clustering. We find that we get a better agreement with the clustering in SDSS by modifying the mean satellite colour to

$$\langle g-r|M_r\rangle_{\rm sat} = 0.86 - 0.065(M_r+20). \quad (3.29)$$

At, for example, ${}^{0.1}M_r = -16$, the fraction of blue satellites given by Eq. 3.27 and Eq. 3.29 is less than 1, meaning that some satellite galaxies are red. However, all galaxies at this magnitude should lie on the blue sequence, as determined from Eq 3.24. In order to achieve the correct $f_{\rm blue}$ from Eq 3.24, the fraction of satellites which are blue must be $(f_{\rm blue} - f_{\rm cen})/f_{\rm sat}$ if all central galaxies are on the blue sequence. If the value of $f_{\rm sat}^{\rm blue}$ is greater than this, it is still possible to get the correct $f_{\rm blue}$ by making central galaxies red, but $f_{\rm sat}^{\rm blue}$ cannot be less than this. To ensure that at faint magnitudes we get the total fraction of blue galaxies given by Eq 3.24, we take the fraction of blue satellites to be

$$f_{\rm sat}^{\rm blue}(M_r) = \max\left\{f_{\rm sat}^{\rm blue}(M_r), \frac{f_{\rm blue}(M_r) - f_{\rm cen}(M_r)}{f_{\rm sat}(M_r)}\right\}. \quad (3.30)$$

The fraction of central galaxies that are blue can then be determined from $f_{\rm blue}(M_r)$ and $f_{\rm sat}^{\rm blue}(M_r)$. However, Skibba & Sheth (2009) erroneously state that

the fraction of central galaxies which are blue is

$$f_{\rm cen}^{\rm blue}(M_r) = \frac{f_{\rm blue}(M_r)}{f_{\rm cen}(M_r)}, \tag{3.31}$$

where $f_{\rm cen}(M_r)$ is the fraction of galaxies which are centrals. Eq. 3.31 is only true if all satellite galaxies are red; since a significant fraction of faint satellites are blue, the fraction of blue central galaxies needs to be reduced to ensure we get the correct total fraction of blue galaxies given by Eq. 3.22. This is achieved by changing Eq. 3.31 to

$$f_{\rm cen}^{\rm blue}(M_r) = \frac{f_{\rm blue}(M_r) - f_{\rm sat}^{\rm blue}(M_r)(1 - f_{\rm cen}(M_r))}{f_{\rm cen}(M_r)}. \tag{3.32}$$

For each galaxy, a uniform random number x is drawn in the interval $0 < x < 1$. For central galaxies, if $x < f_{\rm cen}^{\rm blue}(M_r)$ (given by Eq. 3.32), the galaxy is blue, and a colour is drawn randomly from the Gaussian distribution defined by Eq. 3.21, otherwise it is red, and the colour is drawn from Eq. 3.20. Similarly, satellite galaxies are assigned to the blue sequence if $x < f_{\rm sat}^{\rm blue}(M_r)$, and the red sequence otherwise.

3.4.2.2 Evolution of colours with redshift

The colour magnitude diagram evolves with redshift, as seen for example in figure 6 of Loveday et al. (2012) from GAMA. We therefore need to evolve the expressions given in Section 3.4.2.1 in order to produce a mock which has a realistic distribution of colours at these redshifts. In the GAMA data at high redshifts, only the brightest tip of the red and blue sequences can be seen, making it difficult to constrain their slopes. We therefore keep the slope of the red and blue sequence fixed with redshift.

We keep the red and blue sequences fixed at $z < 0.1$, and evolve them with redshift as

$$\begin{aligned} \langle g-r|M_r\rangle_{\rm red}(z) &= \langle g-r|M_r\rangle_{\rm red} - 0.18(\min\{z, 0.4\} - 0.1) \\ {\rm rms}(g-r|M_r)_{\rm red}(z) &= {\rm rms}(g-r|M_r)_{\rm red} + 0.5(z-0.1) + 0.1(z-0.1)^2 \end{aligned} \tag{3.33}$$

and

$$\begin{aligned} \langle g-r|M_r\rangle_{\rm blue}(z) &= \langle g-r|M_r\rangle_{\rm blue} - 0.25(\min\{z,0.4\}-0.1) \\ {\rm rms}(g-r|M_r)_{\rm blue}(z) &= {\rm rms}(g-r|M_r)_{\rm blue} + 0.2(z-0.1), \end{aligned} \tag{3.34}$$

respectively, where we stop evolving the mean of the sequences above $z = 0.4$ in order to prevent too many high redshift galaxies being assigned as blue.

The mean satellite colour is also evolved as,

$$\langle g-r|M_r\rangle_{\rm sat}(z) = \langle g-r|M_r\rangle_{\rm sat} - 0.18(z-0.1), \tag{3.35}$$

and the fraction of blue galaxies is evolved as

$$f_{\rm blue}(M_r)(z) = 0.2M_r + 4.4 + 1.2(z-0.1) + 0.5(z-0.1)^2. \tag{3.36}$$

Fig. 3.13 shows the distribution of colours in the mock catalogue compared to GAMA for galaxies in different redshift and magnitude bins. Our parametrisation of the colour evolution is able to produce a good approximation to the GAMA colour distributions at all redshifts.

To evolve the luminosity function, we have assumed a fixed Q parameter for all galaxies. We note that Loveday et al. (2012, 2015) hint that red and blue galaxies evolve differently, with a different $Q_{\rm red}$ and $Q_{\rm blue}$. However, the assumption of fixed Q with this parametrisation is able to reproduce the observed colour-magnitude diagram.

3.4.3 Colour dependent k-corrections

In the mock catalogue, we use the HOD method to assign each galaxy an r-band absolute magnitude ${}^{0.1}M_r$, and the method outlined above to randomly generate a ${}^{0.1}(g-r)$ colour. However, the apparent magnitude, r, is the quantity which would be measured directly by the survey, and this is related to the absolute magnitude, ${}^{0.1}M_r$, through the equation

$${}^{0.1}M_r - 5\log_{10} h = r - 5\log_{10} d_L(z) - 25 - {}^{0.1}k(z), \tag{3.37}$$

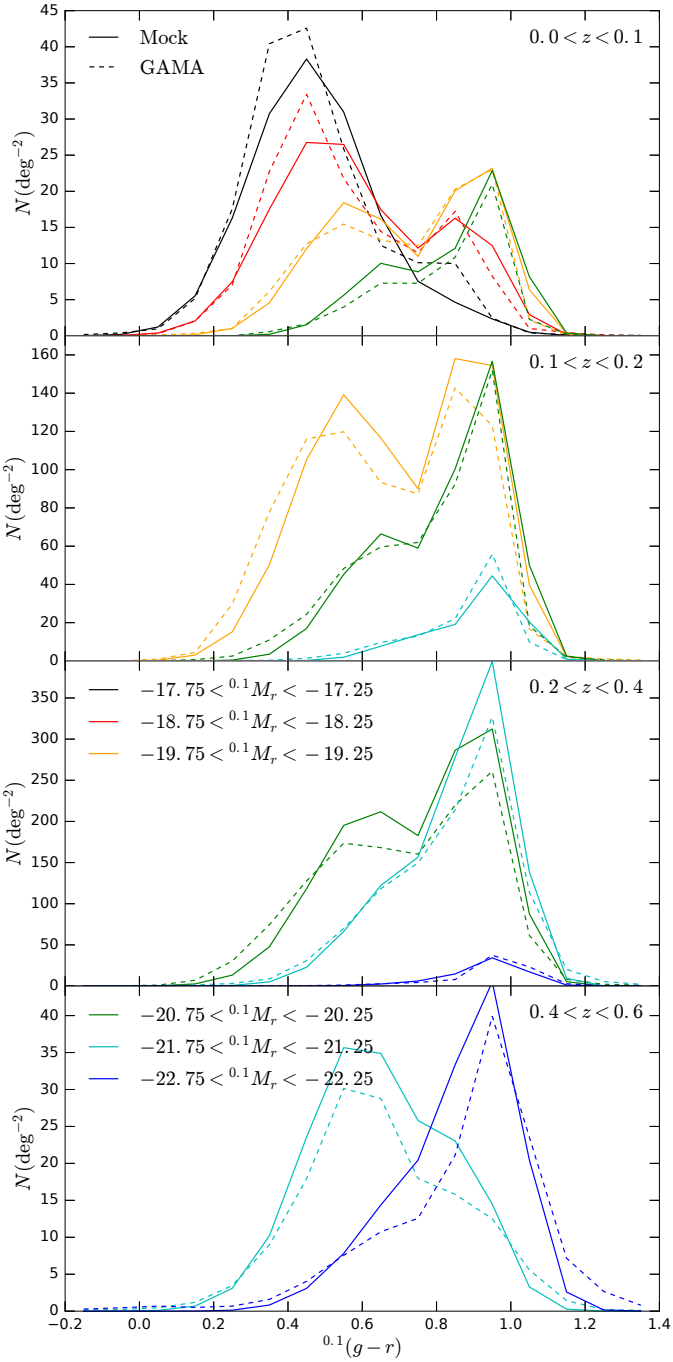

Figure 3.13: Distribution of ${}^{0.1}(g-r)$ colours in the mock catalogue (solid lines) compared to GAMA (dashed lines). Each panel shows the colour distributions of galaxies in a certain redshift range. Different ranges in absolute magnitude are indicated by the colour of the line, as shown in the legend, which is split over several panels.

where $d_L(z)$ is the luminosity distance in units of h^{-1}Mpc, and ${}^{0.1}k(z)$ is the k-correction. The superscript 0.1 denotes that the magnitude has been k-corrected to a reference redshift of $z_{\rm ref} = 0.1$. In order to calculate an apparent magnitude for each galaxy in the mock catalogue, we use colour-dependent k-corrections derived from the GAMA survey, similar to those given in table 1 of McNaught-Roberts et al. (2014), except for k-correcting to $z_{\rm ref} = 0.1$, rather than $z_{\rm ref} = 0$.

The k-correction for each individual galaxy in GAMA is fit with a 4th order polynomial of the form

$$ {}^{0.1}k(z) = \sum_{i=0}^{4} A_i (z - 0.1)^{4-i}. \tag{3.38}$$

The median k-correction is then found in 7 equally spaced bins of ${}^{0.1}(g - r)$ colour. Strictly speaking, the constant term in Eq. 3.38 should have the value $A_4 = -2.5 \log_{10}(1 + z_{\rm ref})$ (Hogg et al., 2002); McNaught-Roberts et al. (2014) do not require this, but they end up with values of A_4 close to 0 at their $z_{\rm ref} = 0$. We force our k-corrections to have the value $A_4 = -2.5 \log_{10}(1.1)$ at our $z_{\rm ref} = 0.1$, but this only has a small effect on the k-corrections.

Using 7 distinct k-corrections based on colour leads to artificial features being added to the mock catalogue; for example step features can be seen in the colour-magnitude diagram at the boundaries between colour bins. In order to remove these features, we interpolate the k-corrections between the median colour in each bin.

Fig. 3.14 shows the polynomial fits to the k-corrections as a function of redshift. By definition, all the curves cross at ${}^{0.1}k(z = 0.1) = -2.5 \log_{10}(1.1) \approx -0.103$. The polynomial coefficients are shown in Table 3.1.

3.4.4 Colour dependent clustering in the mock

The projected correlation function of galaxies in the mock at low redshifts, split by red and blue galaxies, is shown in Fig. 3.15 for different bins in absolute magnitude

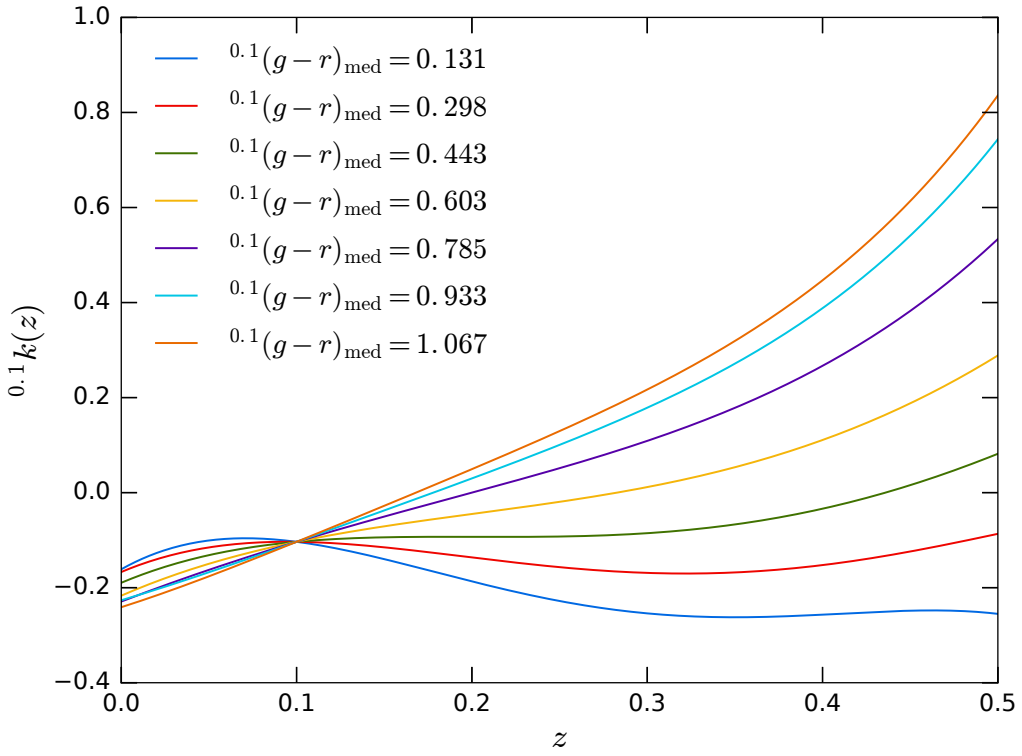

Figure 3.14: Median $^{0.1}(g-r)$ colour-dependent k-correction for galaxies in GAMA as a function of redshift, in 7 equally spaced bins of colour. The colour of each line indicates the colour bin; the median colour is indicated in the legend.

Table 3.1: Polynomial coefficients of the median k-corrections of galaxies in GAMA in equally spaced bins of $^{0.1}(g-r)$ colour, as defined in Eq. 3.38. $^{0.1}(g-r)_{\rm med}$ is the median colour in each bin, and A_i are the polynomial coefficients. The constant term $A_4 = -2.5\log_{10}(1.1) \approx -0.103$, as described in the text.

$^{0.1}(g-r)_{\rm med}$	A_0	A_1	A_2	A_3
0.131	-45.33	35.28	-6.604	-0.4805
0.298	-20.08	20.14	-4.620	-0.04824
0.443	-10.98	14.36	-3.676	0.3395
0.603	-3.428	9.478	-2.703	0.7646
0.785	6.717	3.250	-1.176	1.113
0.933	16.76	-2.514	0.3513	1.307
1.067	20.30	-4.189	0.5619	1.494

and compared to the clustering in the corresponding volume limited samples from SDSS (Zehavi et al., 2011), where the red and blue samples are defined using the same colour cut as their equation 13. In the SDSS data, red galaxies are clustered more strongly than blue galaxies, since red elliptical galaxies are more likely to reside in more massive haloes, which are more strongly biased (Eisenstein et al., 2005a). As the samples get fainter, the strength of the colour dependence becomes stronger. These trends are reproduced in the mock catalogue, using the modified satellite colour in Eq. 3.29.

Projected correlation functions for red and blue galaxies are also shown for higher redshift galaxies in Fig. 3.16, compared with the clustering seen in GAMA (Farrow et al., 2015). The galaxy samples are defined using the same ${}^{0.0}M_r$ magnitude ranges as figure 14 of Farrow et al. (2015), and using the same ${}^{0.0}(g-r)$ colour cut (their equation 4), where the superscript 0.0 denotes that these magnitudes are k-corrected to a reference redshift of $z_{\rm ref} = 0$. The clustering of the red and blue galaxies in the mock is in reasonable agreement with the GAMA data.

3.5 Applications

As shown in Section 3.4, the galaxies in the mock catalogue have realistic clustering, which is in agreement with measurements from SDSS and GAMA. The galaxies also have a realistic distribution of ${}^{0.1}(g-r)$ colours at different redshifts. Future surveys, such as DESI and Euclid, aim to make measurements of the BAO and redshift space distortions, which probe larger scales than have been considered so far. Here, we show as an example some of the measurements that can be made using this mock catalogue.

3.5.1 BAO

As described in Section 3.2.1 and shown in Fig. 3.3, the large box size of the MXXL simulation enables the BAO feature to be seen clearly in the clustering of haloes.

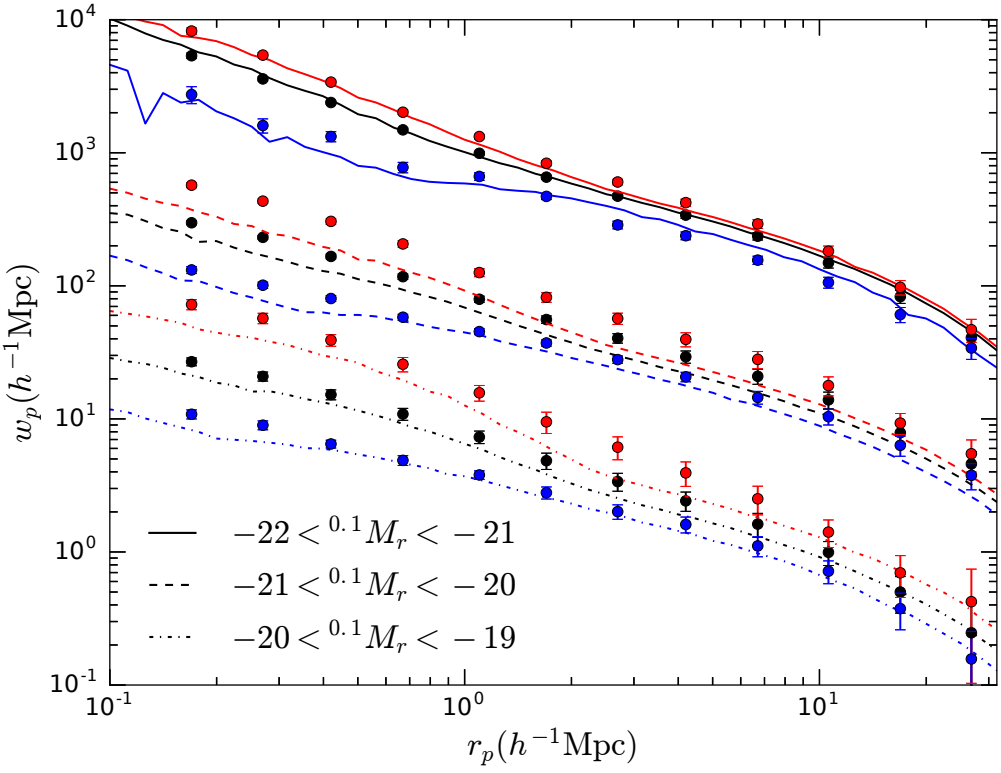

Figure 3.15: Projected correlation functions of red and blue galaxy samples in the mock catalogue at low redshifts (lines) compared to the SDSS volume limited samples (Zehavi et al., 2011) (points with error bars) for different magnitude bins. The clustering of all galaxies in a sample is shown in black, while clustering for red and blue galaxies, defined by the colour cut ${}^{0.1}(g-r)_{\mathrm{cut}} = 0.21 - 0.03\,{}^{0.1}M_r$, is shown in red and blue, respectively. Line style indicates the magnitude bin, as shown by the legend. For clarity, magnitude samples are successively offset by 1 dex from the $-21 <^{0.1} M_r < -20$ samples.

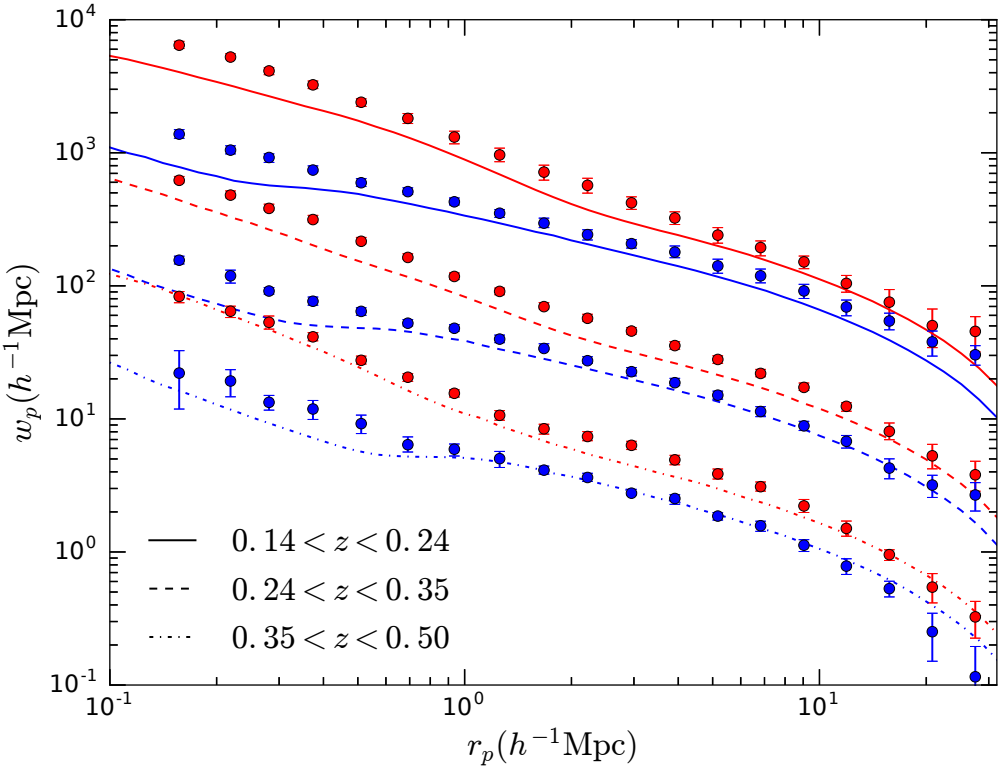

Figure 3.16: Projected correlation functions of red and blue galaxy samples at high redshift in the mock catalogue (lines), compared to GAMA (points with error bars) (Farrow et al., 2015). Red and blue lines indicate red and blue galaxy samples, where the colour cut is defined as ${}^{0.0}(g-r)_{\mathrm{cut}} = -0.03({}^{0.0}M_r + 20.6) + 0.678$. For each sample of galaxies in the mock, the same ${}^{0.0}M_r$ magnitude range is used as the GAMA galaxy sample. The style of the line indicates the redshift range, as shown in the legend. Redshift samples are successively offset from the $0.24 < z < 0.35$ samples by 1 dex for clarity.

Here, we show that the BAO can also be seen in measurements of the redshift-space galaxy clustering. Fig. 3.17 shows the large-scale redshift-space correlation function for several apparent magnitude threshold galaxy samples, using a redshift weighting $w(z) = 1/(1 + 4\pi J_3 \bar{n}(z))$ (Efstathiou et al., 1990a), where $\bar{n}(z)$ is the number density of galaxies in the sample at redshift z, and $J_3 = \int \xi r^2 dr$, where we have assumed $4\pi J_3 = 3 \times 10^4 h^3 \mathrm{Mpc}^{-3}$. The BAO peak can be seen in all samples, but the errors in the correlation function are largest for the $r < 18.0$ sample. The $r < 20.0$ sample contains fainter galaxies, and covers a larger volume, which greatly reduces the errors. For comparison, the crosses with error bars show measurements of the BAO from the Baryon Oscillation Spectroscopic Survey (BOSS) (Reid et al., 2016) for galaxies in the redshift range $0.2 < z < 0.5$ (Ross et al., 2017). The BAO scale in the mock catalogue is $\sim 7\%$ larger than is measured in BOSS. This is consistent with the difference in cosmology between that used in the MXXL simulation and the best fit to observations, including the BOSS results, and is mostly driven by the difference between $\Omega_\mathrm{m} = 0.25$ in MXXL and $\Omega_\mathrm{m} = 0.31$ in the Planck cosmology. The amplitude of the BAO peak in the mock catalogue also differs with the BOSS results, but we have not made any attempt to match the BOSS colour selection.

3.5.2 Redshift space distortions

The two-point correlation function, $\xi(s, \mu)$, in bins of s and μ, can be decomposed into multipoles (Hamilton, 1992),

$$\xi(s, \mu) = \sum_l \xi_l(s) P_l(\mu), \tag{3.39}$$

where s is the separation between a pair of galaxies in redshift space, $\mu = \cos\theta$ is the cosine of the angle between the vector $\mathbf{s}$ and the line of sight, and $P_l(\mu)$ is the l^{th} order Legendre polynomial. The multipoles can be determined from the measured $\xi(s, \mu)$ by evaluating the integral

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^{1} \xi(s, \mu) P_l(\mu) d\mu. \tag{3.40}$$

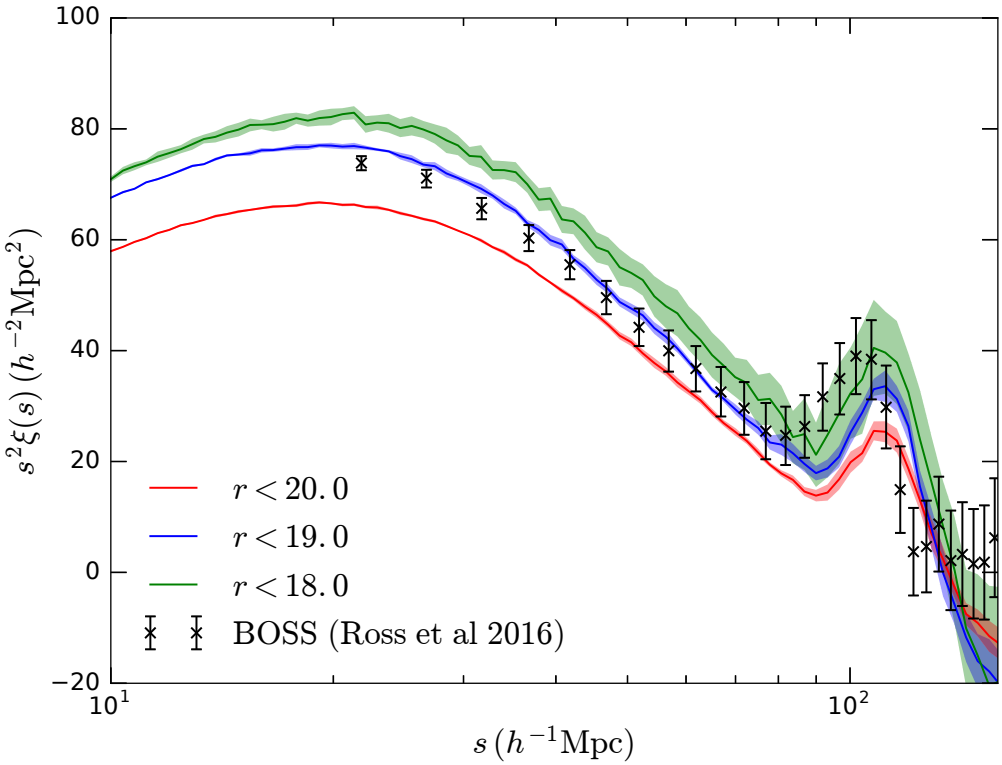

Figure 3.17: Large-scale redshift-space correlation function in the galaxy catalogue, scaled by s^2, for different apparent magnitude threshold samples, as indicated by the colour. For each sample, the solid curve is the clustering calculated over the full sky, and the shaded area is the error on the mean of four quadrants. Crosses with error bars indicate the measured clustering from BOSS in the redshift range $0.2 < z < 0.5$ (Ross et al., 2017), divided by a factor of 1.5, to make the comparison easier.

Due to the symmetry $\xi(s, \mu) = \xi(s, -\mu)$, all odd-numbered terms are zero, and in linear theory, it is only the monopole, $\xi_0(s)$, quadrupole, $\xi_2(s)$, and hexadecapole, $\xi_4(s)$, that are non-zero. The amplitude of these multipoles depends on the strength of the redshift space distortions, and can provide a way to measure $f(z)\sigma_8(z)$ (Samushia et al., 2012).

The multipoles of the redshift-space correlation function of galaxies in the mock catalogue are shown in Fig. 3.18 for different volume limited magnitude threshold samples, and compared to measurements of clustering from SDSS (Guo et al., 2015). The monopole and quadrupole show reasonable agreement with the SDSS measurements, although the amplitude of the hexadecapole is a little high for some of the samples. Overall, the redshift space distortions in the mock catalogue look reasonably realistic, showing that the catalogue will be useful for future surveys that will take redshift space distortion measurements. We have extended the predictions beyond the range of the SDSS results, where they are easier to model and can be probed by surveys like DESI and Euclid.

3.6 Conclusions

For upcoming galaxy surveys, such as DESI and Euclid, it is important to have realistic mock catalogues in order to test and verify analysis tools, assess incompleteness and determine error covariances. The mock catalogues can also be used to make predictions and set expectations in advance of the first data from the survey.

We have outlined a method for creating a mock catalogue from the Millennium-XXL (MXXL) simulation. We first created a halo lightcone catalogue from the simulation, which we then populated with galaxies using a halo occupation distribution (HOD) scheme.

The halo lightcone catalogue is created from the simulation by finding the interpolated values of the position, velocity and mass of each halo at the redshift at which it crosses the observer's lightcone. The halo lightcone catalogue covers the

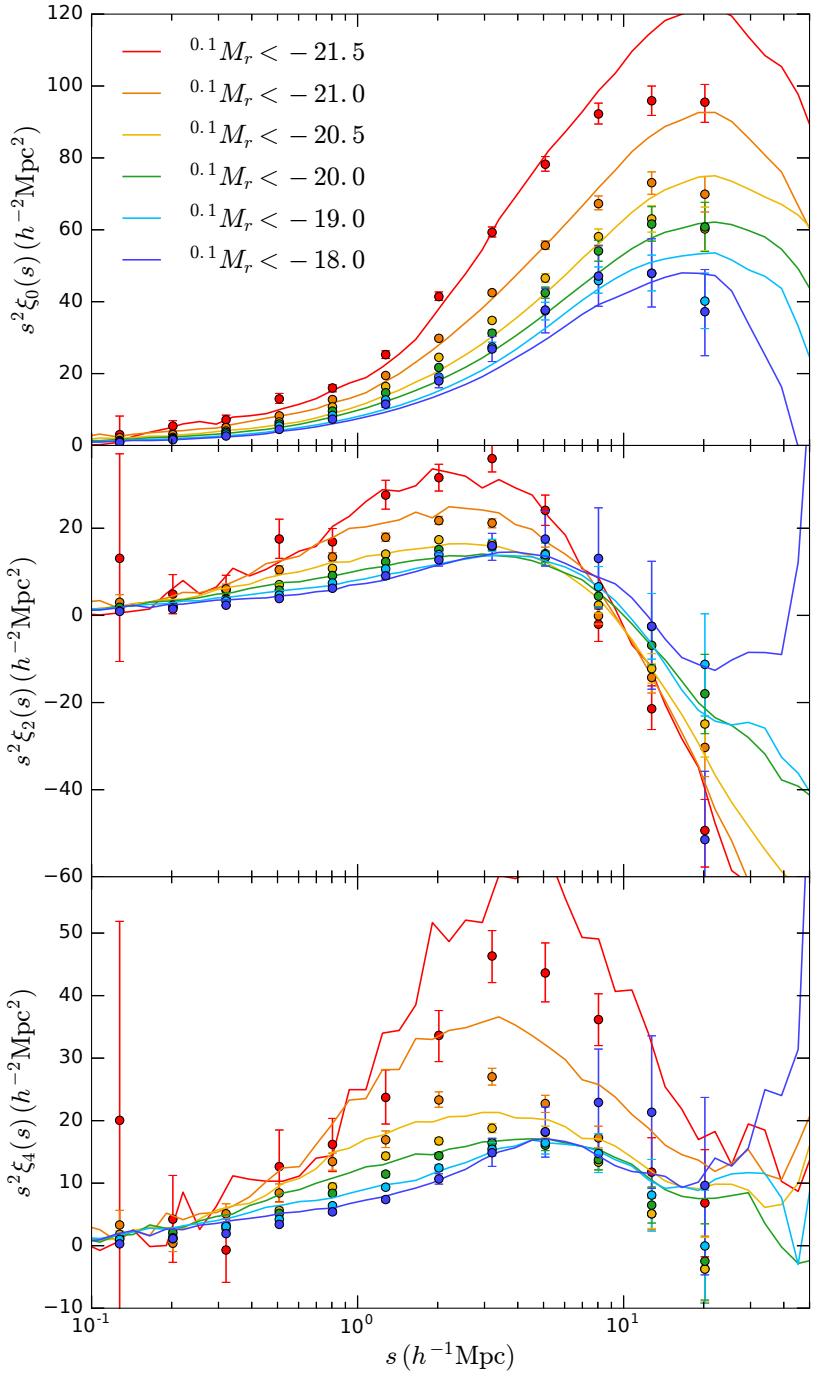

Figure 3.18: Monopole, $\xi_0(s)$, (top panel), quadrupole, $\xi_2(s)$, (middle panel), and hexadecapole, $\xi_4(s)$, (bottom panel), of the redshift space two-point correlation function for different volume limited samples (solid lines), where the colour indicates the magnitude threshold. Points with error bars show the measured clustering from SDSS (Guo et al., 2015).

full sky, and extends to redshift $z = 2.2$ with a mass resolution of $\sim 10^{11}h^{-1}\mathrm{M}_{\odot}$. Extending the catalogue to high redshifts requires multiple periodic replications of the MXXL box; these replications are only necessary to extend to redshifts greater than $z \sim 0.5$.

The halo catalogue is populated with galaxies using a Monte Carlo method, which randomly assigns galaxies with luminosities to dark matter haloes, following a HOD. This is an extension of the method outlined in Skibba et al. (2006) to a 5 parameter HOD. Galaxies in the mock catalogue are also assigned a ${}^{0.1}(g-r)$ colour, based on the Monte Carlo method of Skibba & Sheth (2009). A galaxy is assigned a colour based on its luminosity, redshift, and whether it is a central or a satellite galaxy, which is randomly drawn from a parametrisation of the SDSS colour-magnitude diagram.

The values of the HOD parameters we use are based on the best fitting HODs which reproduce the measured clustering from SDSS (Zehavi et al., 2011), but in Millennium cosmology. In the standard 5 parameter HOD model, the parameter $\sigma_{\log M}$ introduces Gaussian scatter in the luminosity of central galaxies for haloes of a fixed mass, which can lead to the unphysical crossing of HODs in different luminosity bins. We modify the HOD model so that this scatter follows a pseudo-Gaussian spline kernel, which prevents this crossing. The HODs are evolved with redshift such that they reproduce a target luminosity function, which is chosen to be the SDSS luminosity function at low redshift, and the luminosity function from GAMA at high redshifts. For a sample of galaxies with a fixed number density, the shape of the HOD is kept fixed with redshift, and the mass HOD parameters are all multiplied by the factor f which is required to produce the correct number density. By construction, the mock catalogue reproduces the SDSS and GAMA luminosity functions, and the ratio of the HOD parameters M_1/M_{min} is constant with redshift.

We modify the parametrisation of the colour-magnitude diagram outlined in Skibba & Sheth (2009), and add evolution, such that the distribution of ${}^{0.1}(g-r)$

colours in the mock catalogue agrees with measurements from GAMA at different redshifts.

The galaxy catalogue is a flux limited mock galaxy catalogue, covering the full sky with an r-band magnitude limit of $r < 20.0$ and median redshift $z \sim 0.2$. The angular and projected correlation functions of galaxies in the mock show good agreement with measurements from SDSS and GAMA, and the colour dependence of the clustering is reasonable. The BAO peak can be seen in the large-scale clustering of galaxies in the mock, and galaxies show realistic redshift space distortions, making this mock useful for upcoming surveys which will measure these.

Here we have presented one mock galaxy catalogue, but to enable model inferences and place tight constraints on cosmological parameters, error covariances need to be determined. This requires the use of many mock catalogues, of the order of 100s to 10,000s. This could be achieved by coupling the HOD component of the mock with an approximate but fast method of generating halo catalogues (e.g. Manera et al., 2013; Monaco et al., 2013; Tassev et al., 2013; White et al., 2014; Avila et al., 2015; Chuang et al., 2015; Kitaura et al., 2015).

My contribution to this work was produce the halo lightcone catalogue from the MXXL halo merger trees, which were pre-computed by Raul Angulo. I then developed the methodology and code to populate this lightcone with galaxies, using the HOD fits to the clustering measurements from SDSS, which were produced by Zheng Zheng. The methodology I developed extends the work of Skibba et al. (2006) and Skibba & Sheth (2009) for a 5 parameter HOD, with redshift evolution.

Chapter 4

Fibre Assignment Incompleteness in the DESI Bright Galaxy Survey

4.1 Introduction

The Dark Energy Spectroscopic Instrument (DESI) (DESI Collaboration et al., 2016a,b) will conduct a large spectroscopic survey with the primary science aims of making precision measurements of the baryon acoustic oscillation (BAO) scale and the large-scale redshift space distortion (RSD) of galaxy clustering. BAO will be used to measure the expansion history of the Universe and constrain dark energy (e.g. Seo & Eisenstein, 2003). RSD will be used to measure the growth rate of structure in the Universe, and place constraints on theories of modified gravity (e.g. Guzzo et al., 2008). These measurements are complementary, as they can be used to break degeneracies between models of dark energy and RSD. The instrument, which is nearing completion, will be installed on the 4-m Mayall Telescope at Kitt Peak, Arizona. DESI will consist of dark-time and bright-time programs. The dark-time survey will measure spectra of 4 million luminous red galaxies (LRGs) ($0.4 < z < 1.0$), 17 million emission line galaxies (ELGs) ($0.6 < z < 1.6$), 1.7 million quasars ($z < 2.1$) and 0.7 million high redshift quasars ($2.1 < z < 3.5$) to probe the Ly-α forest. The bright-time survey will consist of the bright galaxy

survey (BGS), a low redshift, flux limited survey of $\sim$ 10 million galaxies with a median redshift $z_{\rm med} \sim 0.2$ (BGS paper, in prep), and a survey of Milky Way stars (DESI Collaboration et al., 2016a).

The light from each target galaxy is collected by fibres located at the focal plane of the telescope, and taken to one of 10 spectrographs, where the spectrum is measured and a redshift determined. However, it is not possible to place a fibre on every single potential target, and even if it is, a redshift measurement can fail due to low surface brightness or weak spectral features. Other complications, such as observing conditions, also affect the redshift completeness in the final galaxy catalogue.[1] To make precise measurements of galaxy clustering in order to reach the primary science aims of the survey, it is essential to correct for the effects of incompleteness.

A major systematic in galaxy clustering measurements is from the effect of fibre collisions, which occur because fibres cannot be placed arbitrarily close together. Since it is not possible to place a fibre on both objects in a close pair, that pair will be missing in the final catalogue, biasing the pair counts, particularly at small scales, which can bias galaxy clustering measurements. If the same patch of sky is observed enough times, the missing galaxies will eventually be observed, removing the bias (e.g. in GAMA Robotham et al., 2010), but typically it is infeasible to do this.

In the Sloan Digital Sky Survey (SDSS) (Abazajian et al., 2009), the fibre collisions can be characterised relatively straightforwardly, since fibres can be placed anywhere on a plate, so long as they are not closer than the fibre collision scale of 55 arcsec (or 62 arcsec for BOSS). A common method to recover the redshift of missing galaxies is to simply assign them the same redshift as the nearest targeted object on the sky (e.g. Zehavi et al., 2005, 2011). However, this method produces unsatisfactory results for the redshift-space correlation function (as shown in

[1] Exposures are scaled dynamically with conditions, with the aim of achieving a consistent signal-to-noise ratio in the spectra.

Section 4.4.3.2). An alternative method that works well for SDSS involves recovering the full correlation function from the regions covered by multiple overlapping tiles (Guo et al., 2012). In dense regions, SDSS is able to target all galaxies, or an unbiased subset, but this is not true for the BGS.

Fibre collisions in DESI are more complicated, since the fibres are controlled by robotic fibre positioners, which can move each fibre anywhere in a small patrol region around a fixed set of centres, arranged in a grid. The fibre positioners can block neighbouring fibres from targeting certain objects, and objects will be missed if the number density of targets in an extended region is greater than the number density of fibres. These effects have a non-trivial impact on clustering estimates. The statistics to be measured from the survey can be modified to remove the affected scales (e.g. Burden et al., 2017; Pinol et al., 2017), but in doing so, information is lost. Bianchi & Percival (2017) have proposed a method to correct clustering measurements by estimating, from many runs of the fibre assignment algorithm, the probability that a pair of galaxies will be targeted, and have shown that this method can provide an unbiased correction to the dark-time ELG sample (Bianchi et al., 2018). The method has also been shown to be successful when applied to data from the VIPERS survey (Mohammad et al., 2018).

Galaxies in the BGS have a variety of properties, and cover a wide range of galaxy bias. Many kinds of galaxy samples can be selected from the survey, such as volume limited samples, stellar-mass selected samples and colour-selected samples. Here, we quantify the incompleteness due to fibre assignment in the DESI BGS, and assess correlation function correction techniques applied to samples from a BGS mock catalogue. This chapter is organised as follows: in Section 4.2, we describe the BGS survey strategy, DESI fibre assignment, and mock survey simulations. In Section 4.3, we quantify galaxy incompleteness in the BGS due to fibre assignment. In Section 4.4, we assess correlation function correction methods on volume limited samples from the BGS mock. Section 4.5 summarises our conclusions. Throughout, we assume the WMAP-1 cosmology of the mock catalogue

presented in Section 4.2.4, with $\Omega_{\rm m} = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$, $h = 0.73$, and $n = 1$ (Spergel et al., 2003). While this cosmology has a higher σ_8 and lower $\Omega_{\rm m}$ than measurements from Planck (Planck Collaboration et al., 2018), we use simulations tuned to produce the correct galaxy clustering, so we expect the dependence of our results on cosmology to be small.

4.2 Fibre Assignment

4.2.1 Survey Strategy

The aim of the DESI BGS is to create a highly complete flux limited catalogue of bright, low redshift galaxies, for the primary science goals of BAO and RSD analysis. The survey is expected to cover $\sim 14,000$ square degrees (Fig. 4.3) in 3 passes of the sky, measuring spectroscopic redshifts of $\sim$ 10 million galaxies, approximately 2 magnitudes deeper than the SDSS main survey (Strauss et al., 2002), with double the median redshift ($z_{\rm med} \sim 0.2$). The BGS will take place concurrently with the Milky Way Survey during bright time, when the sky is too bright for the main dark time survey due to moon phase and twilight conditions.

Fibres are currently planned to be assigned to science targets based on the following priority tiers:

1. Priority 1 galaxies ($r < r_{\rm bright}$, ~ 800 deg^{-2})
2. Priority 2 galaxies ($r_{\rm bright} < r < r_{\rm faint}$, ~ 600 deg^{-2})
3. Milky Way stars

where $r_{\rm bright} \sim 19.5$ and $r_{\rm faint} \sim 20.0$.[1]

[1] In Section 4.2.4 we use $r_{\rm bright} = 19.452$ and $r_{\rm faint} = 19.925$, which in the BGS mock catalogue gives number densities of 818 deg^{-2} and 618 deg^{-2} for the bright and faint samples respectively. We also randomly promote 10% of the faint sample to have the same priority as the bright sample (see Section 4.2).

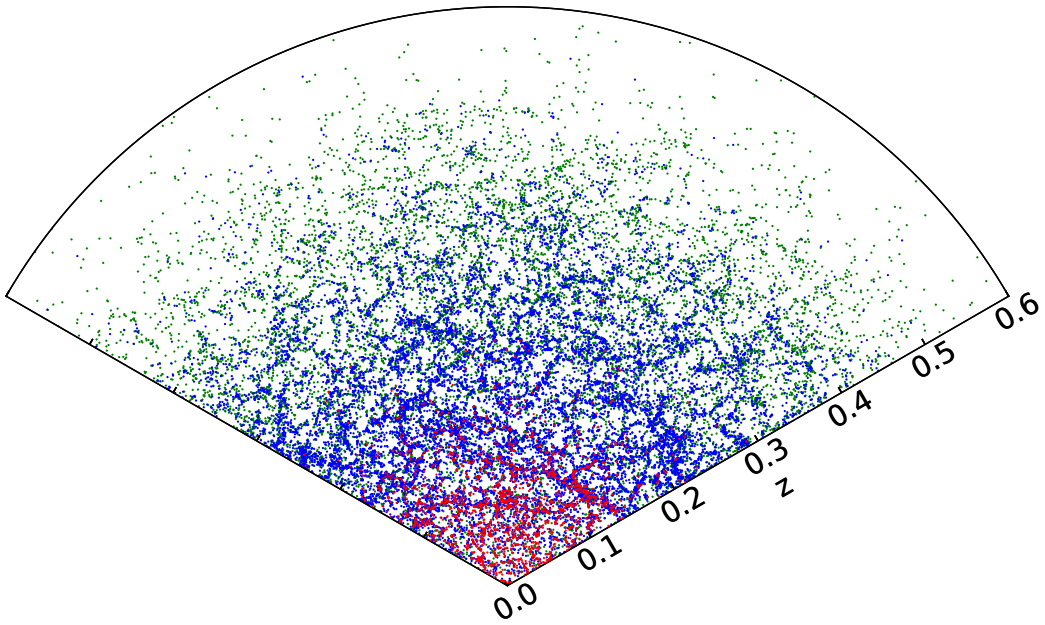

Figure 4.1: Slice through the BGS mock catalogue. Priority 1 galaxies are coloured in blue and Priority 2 galaxies are coloured in green. Galaxies with $r < 17.7$ (the magnitude limit of SDSS) are coloured in red.

The depth of the BGS, in comparison to SDSS, is illustrated in Fig. 4.3, which shows the positions of galaxies in redshift space in a thin slice of the BGS mock catalogue. The priority 1 and 2 galaxies are indicated by the blue and green points, while the red points at low redshift are at the magnitude limit of SDSS ($r < 17.7$). Most SDSS galaxies are at redshift below $z = 0.2$, while the faint BGS sample extends beyond $z = 0.5$. Fig. 4.2 shows the position on the sky of objects in a thin slice of the mock at $z = 0.3$. The left panel is cut to the number density of galaxies in the BOSS survey (Eisenstein et al., 2011; Dawson et al., 2013), while the right panel is cut to priority 1 objects, illustrates that the BGS will sample the cosmic web of structure much more densely than BOSS.

The brightest galaxies with an r-band magnitude $r < 19.5$ are preferentially

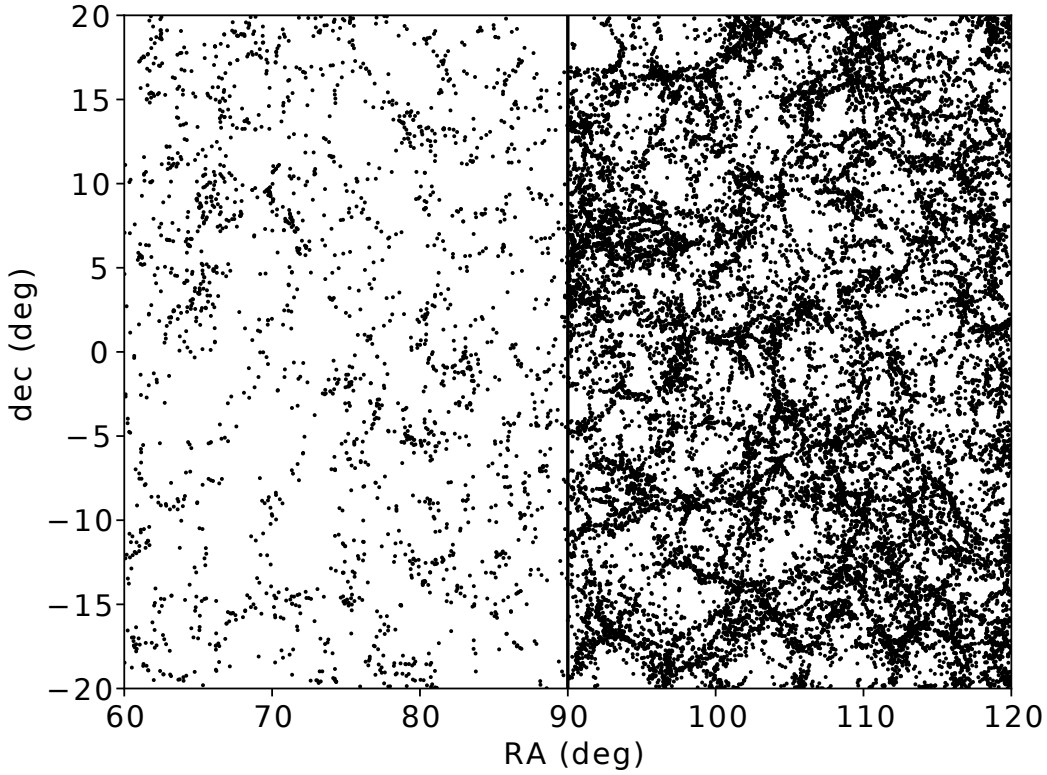

Figure 4.2: Slice through the BGS mock catalogue at $z = 0.3$. *Left panel*: galaxies with a cut in absolute number density of $3 \times 10^{-4}\ h^{-1}\mathrm{Mpc}^{-3}$, corresponding to the number density of galaxies in the BOSS survey. *Right panel*: Priority 1 galaxies.

targeted, since the redshift success rate is expected to be high, making this sample of galaxies highly complete. Fainter galaxies, with $19.5 < r < 20.0$, which have a lower redshift success rate, are given a lower priority, and will form a less complete sample. If a fibre cannot be placed on a galaxy, it will be placed on a Milky Way star.

If a galaxy fails to have its redshift measured, one possibility is for it to remain at the same priority in the next pass. If a redshift is successfully measured, it will remain a potential target in future passes to give the possibility of improving the signal-to-noise of the spectra, but its priority demoted to a fourth priority tier (below that of the Milky Way stars).

In addition to the galaxy targets, 100 fibres will be positioned on standard stars

Table 4.1: Percentage of the survey area covered by N overlapping tiles after 1 pass with 10% of tiles missing, and after the full 1, 2 and 3 passes. The total area covered by each pass is calculated by finding the fraction of objects in a random catalogue that can be potentially assigned a fibre.

N	Pass 1 (90%) (12.2k deg^2)	Pass 1 (13.5k deg^2)	Pass 2 (14.6k deg^2)	Pass 3 (14.8k deg^2)
1	89.79	88.40	13.40	3.63
2	10.20	11.59	67.32	14.91
3	0.01	0.01	18.40	55.85
4	0.0	0.0	0.87	23.14
5	0.0	0.0	0.01	2.34
6	0.0	0.0	0.0	0.12
7	0.0	0.0	0.0	0.01

and 400 on blank sky locations (sky fibres) in each exposure, with an equal number per petal, for flux calibration and sky subtraction.

The observation strategy that will be used in the BGS is still to be chosen. We assume a strategy in which the 3 complete passes are observed sequentially. Each pass consists of ~ 2000 tiles positioned over the entire survey footprint, with overlaps between neighbouring tiles. In the first pass, the tile centres are positioned on the sky with an icosahedral tiling. The tiling for subsequent passes is identical, except with a rotation on the sky, which fills in the missing area due to gaps in the focal plane (docDB-717[1]). The percentage of the survey footprint that is covered by N overlapping tiles after each full pass, and also after 90% of the first pass[2], is summarised in Table 4.1. After 1 pass, $\sim 90\%$ of the footprint is covered by a single tile. This is greatly reduced after subsequent passes, with $\sim 80\%$ covered by 3 or more tiles at the end of the survey. These numbers take into account the gaps in the focal plane.

[1] `https://desi.lbl.gov/DocDB/cgi-bin/private/ShowDocument?docid=717`

[2] 90% of 1 pass is chosen as a realistically incomplete dataset, representing what might be available one third of the way through the survey, where certain fields are missed due to observing conditions.

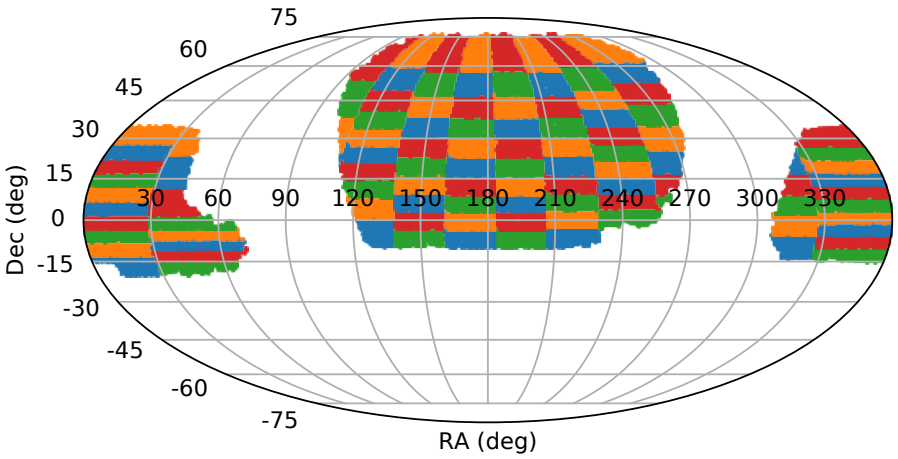

Figure 4.3: Footprint of the DESI BGS, which covers 14,800 square degrees. Colours indicate the 100 jackknife regions.

4.2.2 Robotic Fibre Positioners

Each pointing of DESI, or tile, consists of a total of 5,000 fibres, arranged on the focal plane in 10 wedge-shaped 'petals' (Schubnell et al., 2016). Each individual fibre is controlled by a robotic fibre positioner which can rotate on two arms, allowing the fibre to be placed on any object within a unique circular patrol region (see e.g. figure 3.11 of DESI Collaboration et al., 2016b), with a patrol radius corresponding to an angle on the sky of $R_{\rm patrol} = 1.48$ arcmin (0.0247 deg) (at $z = 0.2$, this is a comoving separation of 0.25 h^{-1}Mpc). The arrangement of fibres is illustrated in Fig. 4.4. There is a small overlap between the patrol regions of neighbouring fibres, and there are gaps between petals which cannot be reached by a fibre. The 'missing' squares around the edge of the tile are the location of the guide focus arrays. Each petal also contains 10 fiducials which provide light sources for the fibre view camera to calibrate fibre positioner placement (DESI Collaboration et al., 2016b).

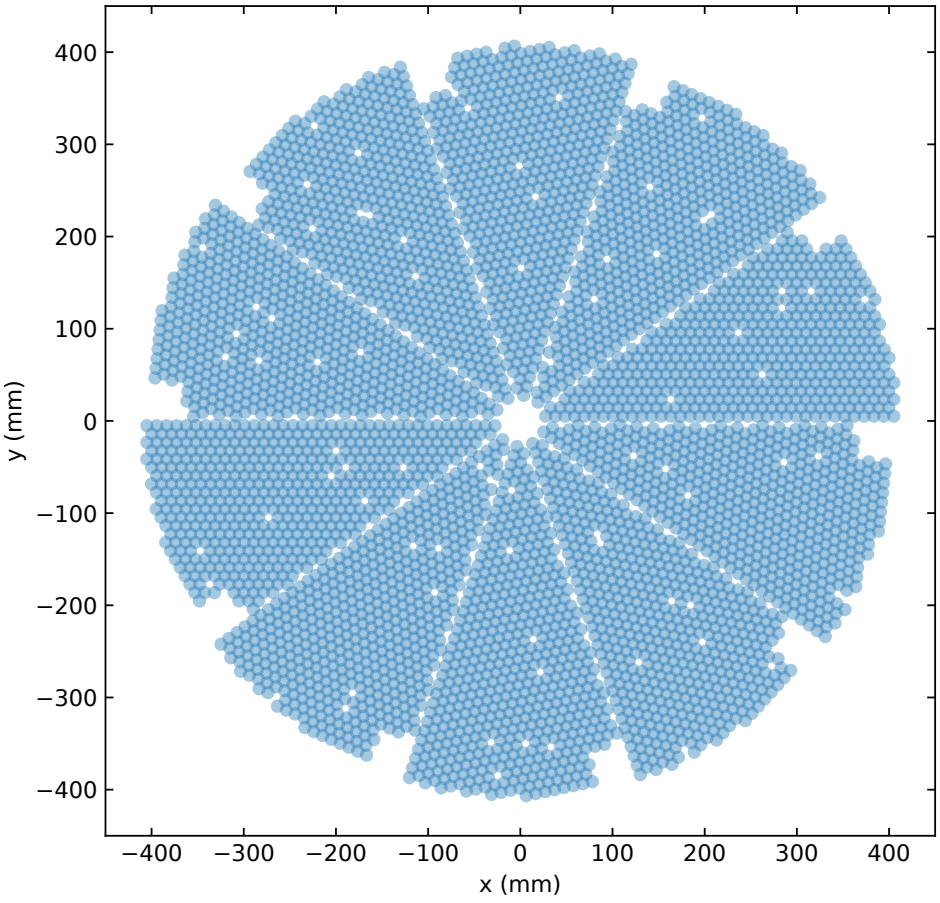

Figure 4.4: A single DESI tile, showing the arrangement of fibres in the focal plane, split into 10 petals. The blue circles indicate the patrol area of each fibre. The holes within each petal are the locations of the fiducials, which provide a light source for the fibre view camera to calibrate the placement of the fibre positioners.

4.2.3 Fibre Assignment Algorithm

To assign fibres to targets, each potential target object is first assigned a primary priority, which is an integer that is determined by the priority tier of the object, e.g. all priority 1 galaxies have the same primary priority, which is greater than the priority 2 galaxies. A uniform random sub-priority in the range $(0, 1)$ is then generated for each object, and the total priority is the sum of the primary and sub-priorities. Fibres are ordered by the highest priority object in their patrol region (from highest to lowest), and are looped through in this order. Each fibre is assigned to the object in its patrol region with the highest priority it is possible for it to target. With this scheme, the assignment of fibres to objects in the same priority tier is randomized, but if a high priority object competes for a fibre with a low priority object, the high priority object will always be assigned a fibre at the expense of the low priority object. If fibres are instead looped through in a fixed order, certain fibres would always have a high priority, and be assigned to a galaxy before its six neighbouring fibres, potentially preventing them from ever targeting certain objects due to fibre collisions.

In the current survey strategy, the entire survey is split into several epochs. In each epoch, tiles are selected by a survey planning algorithm, which determines the sequence of tiles based on date and survey conditions. The selected tiles then go through the fibre assignment algorithm. The fibre assignment algorithm loops through each tile, in a fixed order, assigning fibres to objects. At the end of this loop, there is some redistribution of fibres so that

1. the total number of targets observed is maximized
2. there are the required number of standard stars and sky fibres
3. fibres that are unused are uniformly distributed over tiles.

After fibre assignment, at the end of the epoch, galaxy priorities are updated

depending on whether the redshift measurement was successful or unsuccessful. The updated galaxy priorities is then used in the next epoch. (docDB-2742[1]).

In order to make unbiased 2-point galaxy clustering measurements using the Bianchi & Percival (2017) scheme, each pair of objects in the parent sample must have a non-zero probability of being targeted (see Section 4.4.1.3). To make sure as many pairs as possible can be targeted, we do the following:

4.2.3.1 Dithering tile positions

In regions covered by a single tile, if there are two priority 1 galaxies in the unique patrol region of a single fibre, that fibre will target the galaxy with the highest random sub-priority, but it can never target both, so the pair will always be missed.

This can be mitigated by dithering the tiling of the entire survey in each realization of the fibre assignment algorithm, i.e. randomly rotating the whole 3-pass set of survey tiles by a small angle (of the order of $R_{\rm patrol}$). This is entirely equivalent to keeping the tiling in each realization fixed, and rotating the positions of the galaxies on the sky. In some of these random dithers, the two objects in an untargetable pair will be split between two neighbouring fibres, giving the pair a non-zero probability of being targeted. Since tile centres are uncorrelated with large-scale structure, galaxy pairs of any separation in any environment are equally likely to be targeted in each realization, and therefore it is valid to average over realizations to estimate the probability. To dither the tile positions, a random rotation axis is chosen, which is uniformly distributed. The tile centres are then rotated around this axis by a small angle, which we choose to be 3 times the fibre patrol radius.

The dithering of the tile positions is only done when applying the pair weighting correction described in Section 4.4.1.3. When assigning fibres to objects in the real survey, the rotation angle is set to zero.

[1] `https://desi.lbl.gov/DocDB/cgi-bin/private/ShowDocument?docid=2742`

4.2.3.2 Priority 2 galaxies

Priority 1 galaxies always have a higher priority than priority 2 galaxies, so if it is possible for a fibre to be placed on an unobserved priority 1 galaxy, it will always target that galaxy, regardless of how many priority 2 galaxies are in the same patrol region. This means that a significant fraction of priority 2 galaxies in regions with a high density of priority 1 galaxies will always be missed.

One way of sampling these missing priority 2 galaxies is, in each fibre assignment realization, to randomly promote a certain fraction of priority 2 galaxies to the same priority as the priority 1 galaxies. This gives pairs containing at least one priority 2 galaxy in over-dense regions a small, but non-zero probability of being targeted (see Fig. 4.5).

The version of the fibre assignment algorithm we use is 0.6.0.[1]

4.2.4 Survey Simulations

To quantify incompleteness due to fibre assignment and assess correlation function correction methods, we run the fibre assignment algorithm on a BGS mock catalogue from the Millennium-XXL (MXXL) simulation, described in Chapter 3. This is a halo occupation distribution (HOD) mock, which contains galaxies to $r = 20$ over the same redshift range as the BGS, and is constructed to reproduce the luminosity function and clustering measurements from SDSS (Blanton et al., 2003; Zehavi et al., 2011) and GAMA (Loveday et al., 2012; Farrow et al., 2015).[2]

The magnitudes in this catalogue are in the SDSS r-band. These are converted to the DECam r-band (which is used in the DESI target selection) using

$$r_{\mathrm{DECam}} = r_{\mathrm{SDSS}} - 0.03587 - 0.14144(r-i)_{\mathrm{SDSS}} \tag{4.1}$$

[1] `https://github.com/desihub/desitarget`

[2] The MXXL mock is available at `http://icc.dur.ac.uk/data/` and `https://tao.asvo.org.au/tao/`

(docDB-1788[1]). Since the mock catalogue does not contain $r-i$ colours, we assume a mean colour of $(r-i) = 0.4$. To make sure the priority 1 and 2 galaxies have number densities of 818 deg^{-2} and 618 deg^{-2}, we define priority 1 and 2 galaxies using the magnitudes $r_{\rm DECam} = 19.452$ and $r_{\rm DECam} = 19.925$.[2]

The mock is first cut to the set of galaxies which are within the patrol radius of a fibre (with no dither), and therefore could potentially be assigned a fibre.[3] We run 2048 random realizations of the fibre assignment algorithm ($\sim$ 500 CPU hours), with the full 3 passes of tiles to simulate the complete survey. From the survey simulation output, it is also possible to determine which galaxies were assigned fibres in the first or second pass, allowing us to simulate a more incomplete survey without having to re-run the fibre assignment code. In addition to the full 3 passes, we also determine which galaxies are targeted in 1 pass, with a random 10% of tiles missing (which are the same tiles in each realization), to simulate a dataset that might realistically be achieved after 1/3 of the duration of the survey with a survey strategy that prioritizes area (i.e. a strategy where after 1/3 of the duration, pass 1 is completed, as opposed to a strategy where 3 passes are completed in only 1/3 of the survey area). Removing tiles reduces the overall area of the footprint and increases the fraction of the remaining area that is covered by a single tile.

In each run of the fibre assignment code, the tile positions are randomly dithered by an angle 3 times the patrol radius, and a random 10% of priority 2 galaxies are promoted to the same priority as the bright sub-sample. Unless specified, we will refer to the bright sub-sample as 'priority 1' and the faint sub-sample as 'priority 2'.

We only consider targeting incompleteness caused by the fibre assignment algorithm. Redshift incompleteness due to redshift measurement failures, and the

[1] `https://desi.lbl.gov/DocDB/cgi-bin/private/ShowDocument?docid=1788`

[2] These number densities are chosen to match assumptions made in earlier survey simulations (J. Tinker, private communication).

[3] In our clustering analysis we account for the regions of sky this process discards by applying the same criterion to the corresponding random catalogue. This differs from Bianchi et al. (2018), in which the random sample covers the full survey volume.

effects of weather, are left for future work.

4.3 Fibre Assignment Completeness

For a small region of sky, Fig. 4.5 shows the positions of targeted and untargeted galaxies in the BGS mock with the fibre patrol regions superimposed. This region is at the edge of the survey, and is mostly covered by a single tile, shown in blue, with neighbouring tiles in different colours. On the scale of the fibre patrol regions, the surface density of galaxies varies greatly. Some fibres have zero galaxies in their patrol region, leaving them free to target Milky Way stars, while fibres in dense regions can have 10 or more galaxies within their patrol region. It is clear to see that in dense regions, the fibre assignment completeness will be low, since only one galaxy can be assigned a fibre out of many potential targets. More galaxies can be targeted if there are multiple tile overlaps, which will make the completeness higher. In low density regions, the completeness will be very high, since if there is only 1 galaxy within a fibre patrol region, the fibre will always be placed on that galaxy.

Fig. 4.6 shows the position of targeted and untargeted galaxies in a larger region of the survey, after 3 passes, where the galaxies are in the redshift range $0.08 < z < 0.12$. Both panels show the same region of sky, where the overlapping grey shaded circles indicate each DESI tile (ignoring the gaps in the focal plane). Objects assigned fibres are shown as the blue points in the upper panel, while the red points in the lower panel are the objects which fail to be assigned a fibre. The untargeted galaxies in the lower panel are mostly situated within massive haloes (indicated by the black circles), and are more concentrated towards the centre, compared to the targeted galaxies in the upper panel. Again, this shows that the completeness will be low in high density regions.

The completeness due to surface density is quantified in Fig. 4.7. The upper panel shows the average completeness as a function of surface density, after 3

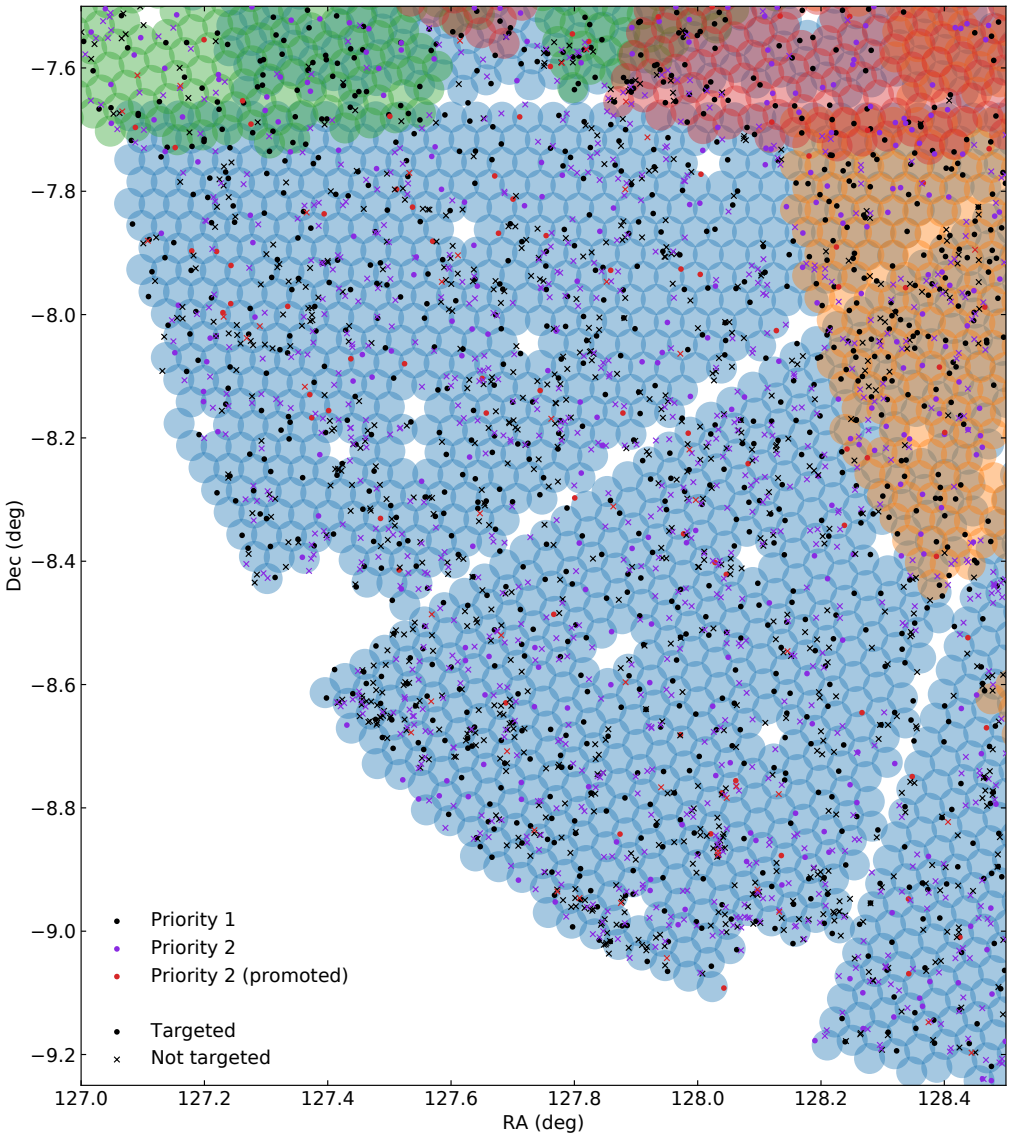

Figure 4.5: A zoom in on a small section around the edge of the survey footprint of one survey simulation, showing the positions of BGS galaxies relative to fibre patrol regions. This survey simulation has zero dither, but 10% of priority 2 galaxies are randomly promoted. Shaded circles indicate the patrol region of each fibre, with each neighbouring tile in a different colour. White regions cannot be reached by a fibre. Circles indicate galaxies which are successfully assigned a fibre, while crosses show untargeted galaxies. The bright priority 1 sample is shown in black, and the faint priority 2 sample is in purple. Promoted priority 2 galaxies are shown in red.

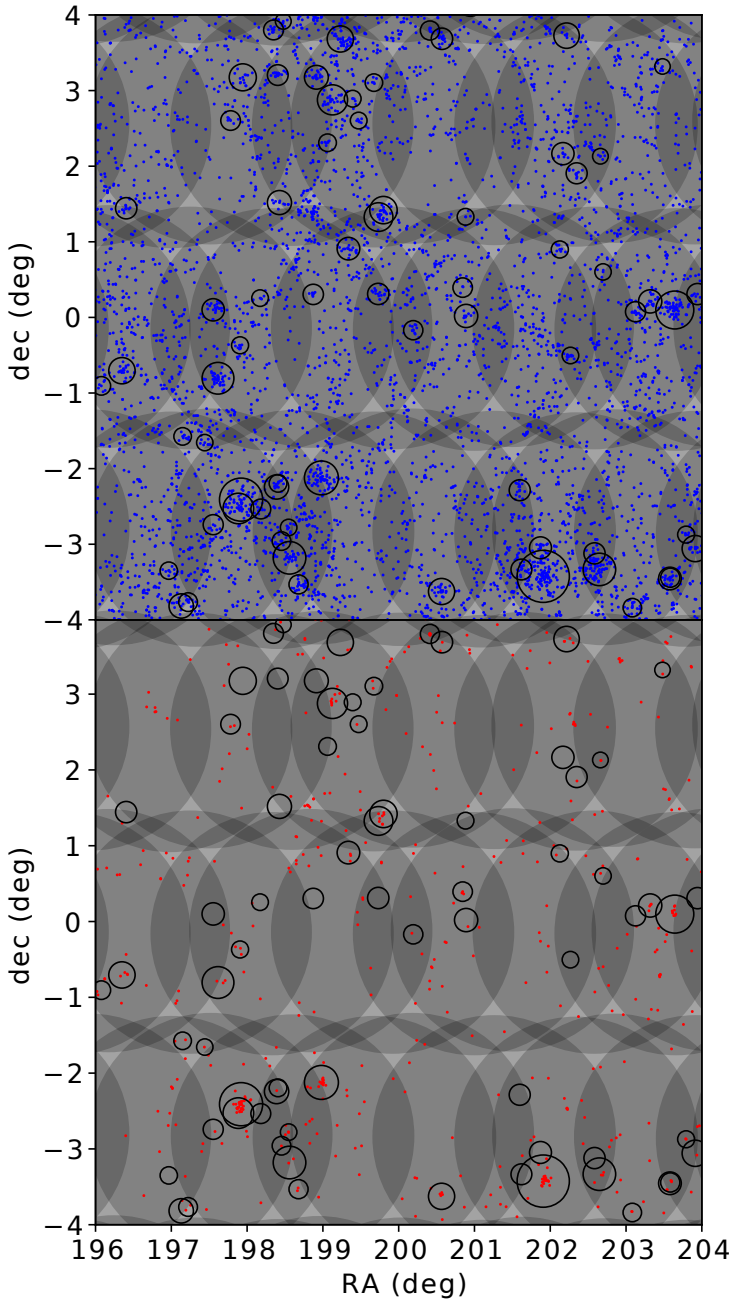

Figure 4.6: Position of DESI tiles, with radius 1.605 degrees, after 3 passes in a small area of the survey. Darker shades of grey indicate a greater number of overlapping tiles. *Top panel*: blue points show the positions of galaxies from the BGS mock catalogue in the redshift range $0.08 < z < 0.12$ which have been assigned a fibre. Black circles indicate the virial radii of halos with halo mass $M_{\rm halo} > 10^{13}\ h^{-1}{\rm M}_{\odot}$. *Bottom panel*: as above, but showing the positions of galaxies which have failed to be assigned a fibre.

passes, in HEALPIX pixels (Górski et al., 2005) with area 0.84 deg^2 ($N_{\rm side} = 64$), separately for all galaxies, and for priority 1 and 2 galaxies. The completeness decreases monotonically as the surface density of galaxies increases. Also, since priority 1 galaxies are preferentially targeted, they have a higher completeness than the priority 2 galaxies. The vertical dotted line indicates a surface density of 1436 deg^{-2}, which is the average surface density of all (priority 1 and 2) galaxies, and horizontal dotted lines show the median completeness in HEALPIX pixels, which is 88%, 94% and 80% for all, priority 1, and priority 2 galaxies respectively. The lower panel shows a histogram of the total number of galaxies, which peaks close to the average surface density. The black dotted curve shows the histogram of the densities of individual HEALPIX pixels, scaled up by a factor of 1000. The unscaled black dotted curve, multiplied by the average number of galaxies per pixel, produces the black solid curve. The variance in the surface density of pixels depends on the resolution. For pixels with area 13.4 deg^2 ($N_{\rm side} = 16$), which is larger than the area of each tile, the surface density varies from the mean by a few hundred objects per square degree.

The fibre assignment completeness of galaxies in the BGS is driven by the surface density of galaxies, since it is not possible to place a fibre on every galaxy if the density of galaxies is greater than the density of fibres.[1] With multiple passes, the same area of sky will be re-observed several times, enabling some of these previously missed galaxies to be targeted. After the full 3 passes of the BGS, most of the footprint ($\sim 80\%$) will have been covered by 3 or 4 tiles (see Table 4.1), but the targeted catalogue will still be incomplete in high density regions.

The upper panel of Fig. 4.8 shows the redshift distribution of galaxies in the BGS, before and after fibre assignment (solid and dashed curves). The lower panel shows the targeting completeness as a function of redshift, where the horizontal dotted lines indicate the average completeness. For the priority 1 and the priority

[1]Each tile of 5000 fibres has a radius of 1.605 deg, which corresponds to a fibre surface density of ~ 600 deg^{-2}.

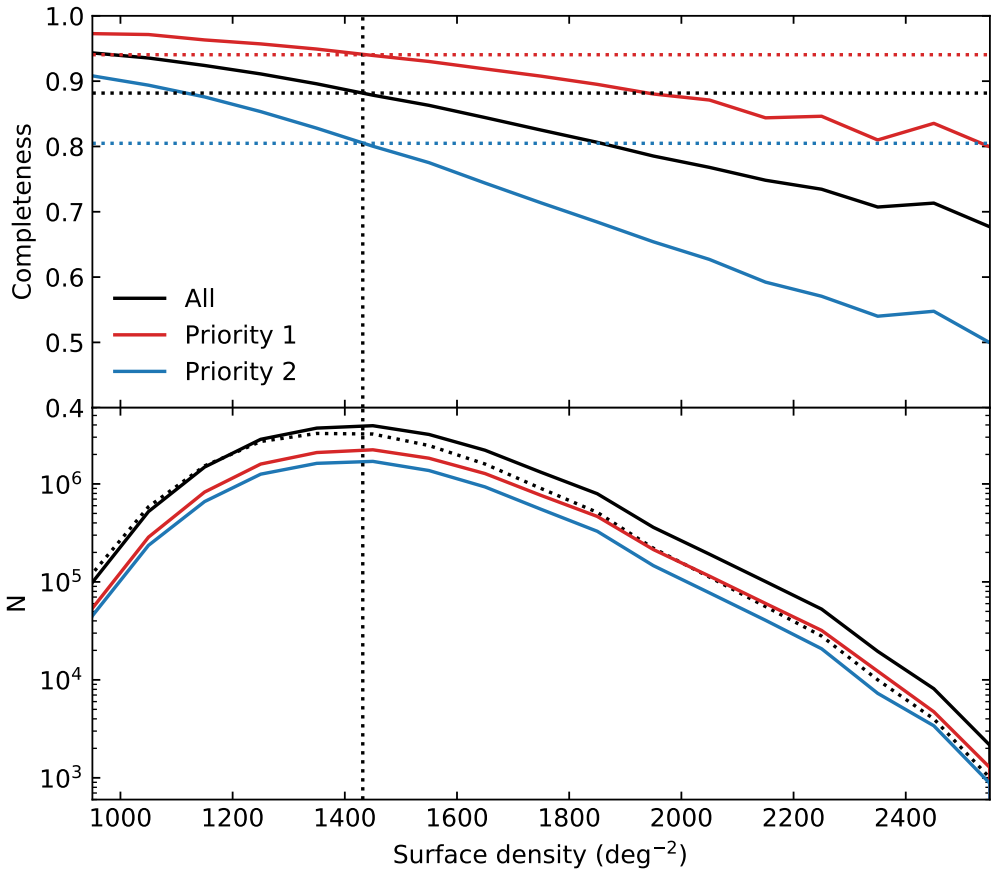

Figure 4.7: *Top panel:* average fibre assignment completeness as a function of the surface density of all BGS galaxies, in HEALPIX pixels of area 0.84 deg^2 ($N_{\rm side} = 64$) for all galaxies (black), priority 1 (red) and priority 2 (blue), after 3 passes with 10% of priority 2 galaxies promoted. The vertical dotted line indicates the average surface density of the survey (1436 deg^{-2}), and horizontal dotted lines indicate the median completeness for the three samples (88%, 94% and 80% for all, priority 1 and priority 2 galaxies respectively). *Bottom panel:* histogram of the total number of objects in bins of surface density. The dotted black curve shows the number of HEALPIX pixels, scaled up by a factor of 1000.

2 galaxies, this curve is non-monotonic. This is because haloes at high redshifts contain few galaxies brighter than the magnitude limit. These galaxies will not greatly enhance the surface density, and the completeness is high. At intermediate redshifts, many more galaxies per halo can be detected in haloes of the same mass, which will result in a much greater enhancement of the surface density, and therefore a lower completeness. At low redshifts, haloes of the same mass will contain an even greater number of galaxies brighter than the magnitude limit, but since they are nearby, they subtend a relatively large angle on the sky, and the perturbation to the surface density is low again. For the complete galaxy sample, the completeness is relatively flat at high redshifts, since the fraction of priority 2 galaxies increases with redshift.

The mean completeness (which differs slightly from the median completeness shown in Fig. 4.7) is $\sim 86\%$, while for priority 1 and 2 galaxies it is $\sim 92\%$ and $\sim 78\%$ respectively. These figures are for the case where 10% of the priority 2 galaxies are given the same priority as the priority 1 galaxies. If there was no promotion of priority 2 objects, the priority 1 galaxies would be more complete, ($\sim 93\%$) but at the expense of the low priority galaxies (see Table 4.3).

Fig. 4.9 shows the completeness of galaxies in haloes, as a function of the distance from the centre of their host halo, for haloes in different mass bins around the peak of the redshift distribution ($0.15 < z < 0.25$). The panels, from top to bottom, show the completeness for haloes with masses $M_{200\mathrm{mean}} \sim 10^{15}h^{-1}\mathrm{Mpc}$, $M_{200\mathrm{mean}} \sim 10^{14}h^{-1}\mathrm{Mpc}$, $M_{200\mathrm{mean}} \sim 10^{13}h^{-1}\mathrm{Mpc}$, and $M_{200\mathrm{mean}} \sim 10^{12}h^{-1}\mathrm{Mpc}$ respectively, plotted to the virial radius ($R_{200\mathrm{mean}}$). $M_{200\mathrm{mean}}$ is defined as the mass enclosed by a sphere of radius $R_{200\mathrm{mean}}$, in which the average density is 200 times the mean density of the Universe. Close to the centre of large haloes, the surface density of galaxies is very high, and therefore the completeness is very low. For $10^{12}h^{-1}\mathrm{Mpc}$ haloes, the average completeness near the centre is $\sim 60\%$, but for the most massive haloes, this completeness is much lower. The spike close to the centre of $M \sim 10^{15}h^{-1}\mathrm{Mpc}$ haloes is due to noise. When measuring two-point

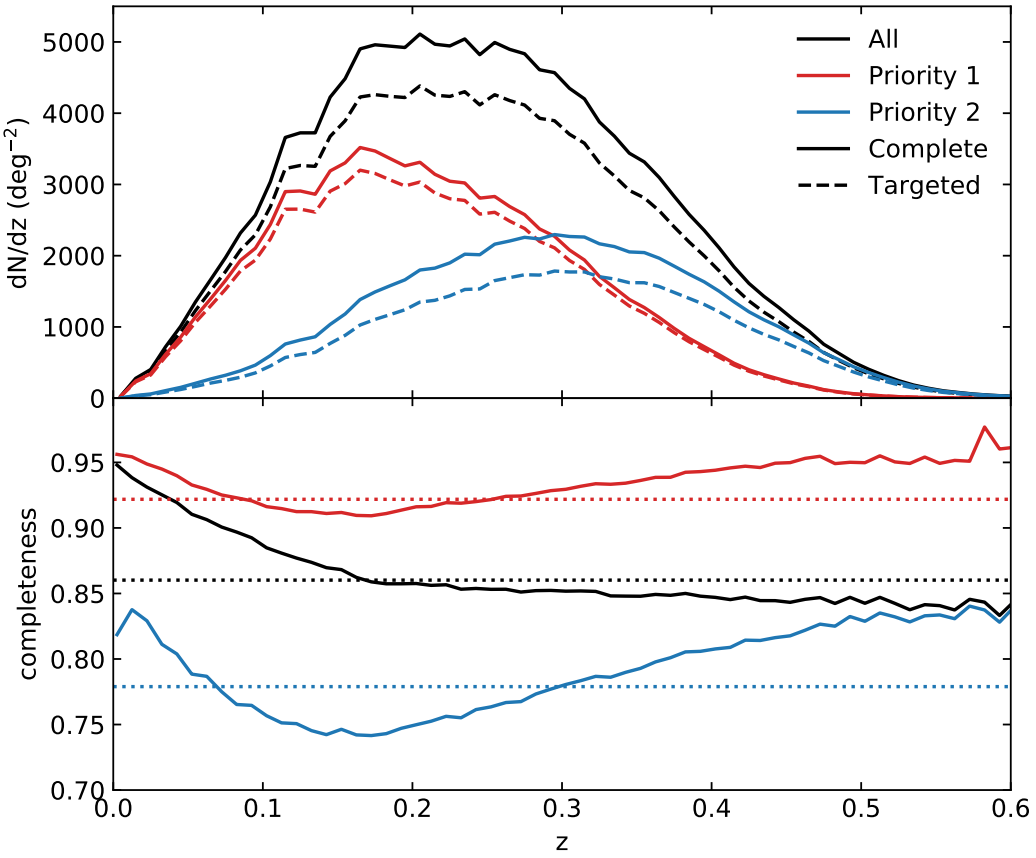

Figure 4.8: *Top panel:* Redshift distribution of galaxies before and after fibre assignment (solid and dashed curves), with the full 3 passes of tiles. The complete sample of BGS galaxies is shown in black, while priority 1 and priority 2 galaxies are in red and blue respectively. *Bottom panel:* Completeness as a function of redshift for all, priority 1 and priority 2 galaxies. Horizontal dotted lines indicate the mean completeness (86%, 92% and 78% for all, priority 1, and priority 2 galaxies respectively).

clustering statistics, as we show in Section 4.4.3, the effect of this incompleteness can be corrected, and this is unbiased so long as each galaxy pair has a non-zero probability of being targeted. Since the completeness in clusters is low, care must be taken, for example, identifying clusters and voids and estimating velocity dispersions. The incompleteness must also be taken into account when estimating higher-point statistics. Our realizations of the fibre assignment algorithm could be used to develop correction procedures for these statistics.

The total number of objects targeted, and the completeness after each pass, is shown in Table 4.2 for all galaxies, priority 1 and 2 galaxies, and the subset of priority 2 galaxies that are promoted to the same priority as priority 1. Since faint galaxies are less clustered than bright galaxies, the promoted priority 2 galaxies have a higher completeness than the priority 1 galaxies. Most of the promoted galaxies are targeted in the first pass.

Table 4.3 shows how the final completeness after 3 passes is affected by the fraction of objects in the faint sub-sample promoted to high priority. The priority 1 sample is most complete with zero promotion (92.9%), but the priority 2 sample is least complete (77%), and certain priority 2 objects will always be missed due to conflicts with high priority objects. As the fraction of priority 2 objects is increased, the percentages converge to the average completeness of $\sim 86\%$.

4.4 Correcting Two-Point Clustering Measurements

4.4.1 Mitigation Techniques

The two-point correlation function at separation $\vec{s}$ can be estimated using the estimator of Landy & Szalay (1993),

$$\xi(\vec{s}) = \frac{DD(\vec{s}) - 2DR(\vec{s}) + RR(\vec{s})}{RR(\vec{s})}, \tag{4.2}$$

where DD, DR and RR are the normalized data-data, data-random, and random-random pair counts. If galaxies in the data catalogue are missing, the resulting

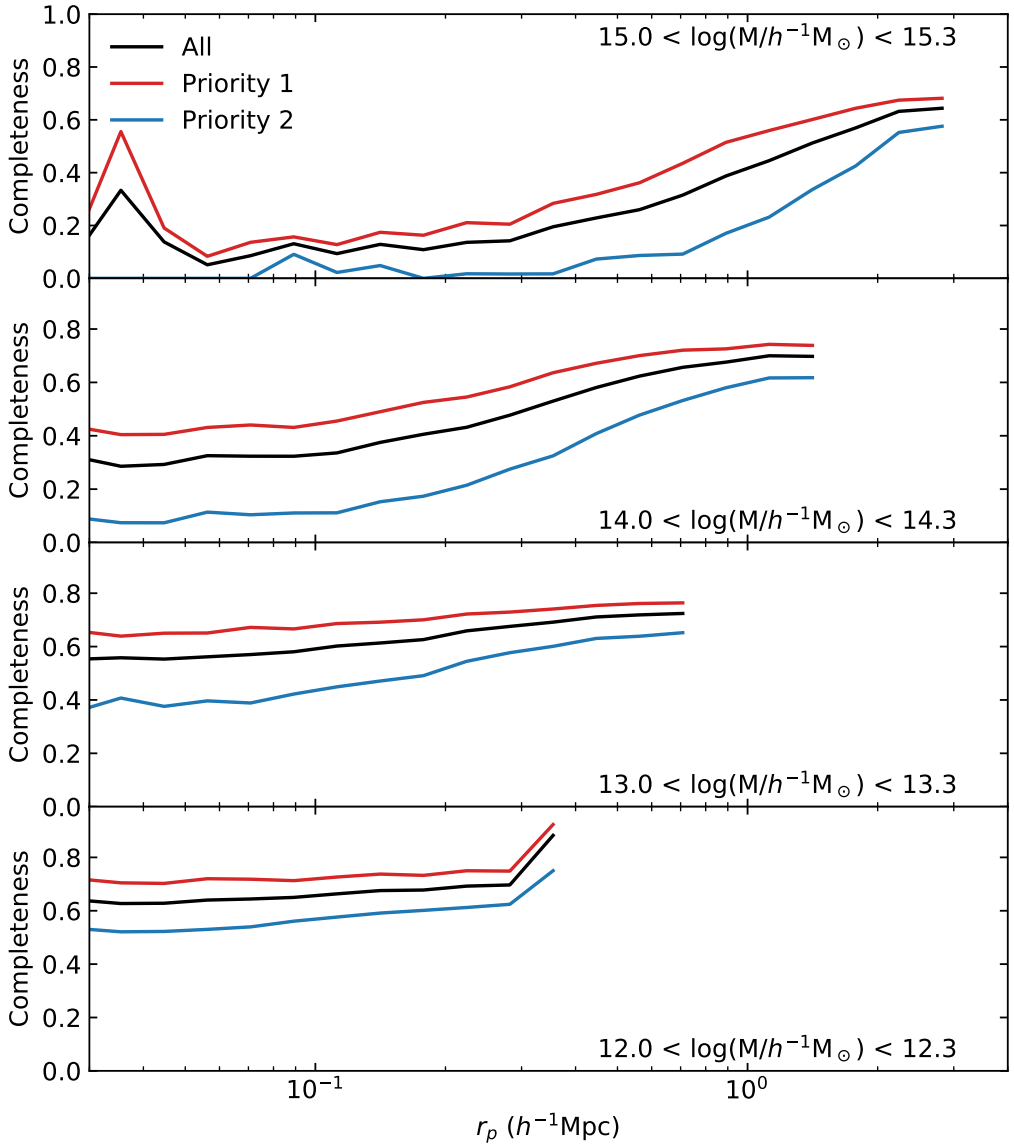

Figure 4.9: Targeting completeness of galaxies in haloes as a function of the transverse distance from the centre of their respective halo, for haloes in the redshift range $0.15 < z < 0.25$, after 3 passes. The completeness for all galaxies is shown in black, and for priority 1 and 2 galaxies in red and blue respectively.

Table 4.2: Table showing the cumulative number of objects targeted after each pass, in millions, and the completeness, as a percentage. Priority 1 and priority 2 are the intrinsic priorities based on magnitude. Priority 2 (p) is the subset of priority 2 galaxies that are promoted to have the same priority as the bright priority 1 galaxies. The final row shows the cumulative number of unused fibres which are available to target Milky Way stars (in millions) after each pass, and the percentage of fibres which are unused after each pass. A total of ~ 9 million fibres are available per pass, excluding standard stars and sky fibres (2,000 pointings, each with 4,500 available fibres).

Sample	Pass 1		Pass 2		Pass 3	
	$N_{\rm gal}$	%	$N_{\rm gal}$	%	$N_{\rm gal}$	%
All	7.54	35.6	13.78	65.0	18.24	86.0
Priority 1	5.15	42.7	8.84	73.3	11.11	92.2
Priority 2	2.39	26.1	4.95	54.1	7.12	77.8
Priority 2 (p)	0.79	86.2	0.84	92.4	0.85	93.2
Free fibres	1.49	16.5	4.30	23.8	8.89	32.8

Table 4.3: Table showing the number of objects targeted after 3 passes, in millions, and the completeness, in survey simulations where the percentage of promoted priority 2 galaxies is varied from 0% to 40%. Priority 1 and 2 galaxies are the bright and faint sub-samples, and priority 2 (p) are the promoted subset of priority 2 galaxies.

Promotion	Priority 1		Priority 2		Priority 2 (p)	
%	$N_{\rm gal}$	%	$N_{\rm gal}$	%	$N_{\rm gal}$	%
0	11.12	92.9	7.04	77.0	-	-
5	11.15	92.5	7.08	77.4	0.43	93.7
10	11.11	92.2	7.12	77.8	0.85	93.2
15	11.07	91.8	7.17	78.4	1.28	93.1
20	11.02	91.5	7.21	78.9	1.69	92.6
25	11.00	91.1	7.26	79.3	2.11	92.5
30	10.94	90.7	7.30	79.8	2.52	92.0
35	10.89	90.3	7.35	80.3	2.93	91.7
40	10.84	90.0	7.39	80.8	3.34	91.4

correlation function will be biased. Mitigation techniques attempt to recover the correlation function of the parent sample from the sample of galaxies that are targeted.

4.4.1.1 Nearest object

We use two different nearest redshift corrections. In the first correction, missing galaxies are assigned the redshift of the nearest targeted object on the sky (the approach taken in the SDSS survey analyses in e.g. Zehavi et al., 2005; Berlind et al., 2006; Zehavi et al., 2011). The catalogue of galaxies is then cut to the volume limited sample using these redshifts. Some of the untargeted objects will be assigned a redshift close to the true value, and will be correctly identified as part of the volume limited sample, but the sample will be contaminated by other galaxies which are assigned incorrect redshifts. We refer to this correction as 'nearest redshift'.

In the second correction, each galaxy is first given a weight of 1, and the weight of a missing galaxy is added to the nearest targeted object on the sky (e.g. in BAO analysis in the BOSS survey, Anderson et al., 2012, 2014b,a). For example, a targeted galaxy with no nearby untargeted galaxies would have weight 1. If there was a close galaxy that was not targeted, the weight would be transferred to the targeted galaxy, which would now have a weight of 2. We hereafter refer to this correction as 'nearest weight'. The nearest weight correction can be seen as an approximation of the pair weighting method of Section 4.4.1.3 (see Bianchi & Percival, 2017).

4.4.1.2 Angular upweighting

When estimating the correlation function, galaxy pairs are upweighted by the factor

$$W(\theta) = \frac{1 + w^{(p)}(\theta)}{1 + w(\theta)}, \tag{4.3}$$

where $w^{(p)}(\theta)$ is the angular correlation function of the complete, parent sample of galaxies, and $w(\theta)$ is the incomplete, targeted sample (e.g. the 2dFGRS analysis of Hawkins et al., 2003). This angular weighting by construction recovers the angular correlation of the parent sample. This correction makes the assumption that the targeted and untargeted galaxies are statistically equivalent in each angular bin, which is not necessarily true, and therefore it may not provide an adequate correction to the redshift space correlation function.

4.4.1.3 Pair Inverse Probability (PIP) Weights

The PIP weighting scheme (Bianchi & Percival, 2017) upweights each galaxy pair by the pair weight $w_{ij} = 1/p_{ij}$, where p_{ij} is the probability that the pair will be targeted. This probability can be estimated by running the fibre assignment code $N_{\rm real}$ times, where $N_{\rm real}$ is of the order of 100s or 1000s. For galaxy i, a vector $\vec{w}_i$ of length $N_{\rm real}$ is stored, which contains a 1 if the galaxy is assigned a fibre, and a 0 otherwise. This vector can conveniently be stored as the bits of an integer (or several integers). The pair weight for galaxies i and j can be written as the dot-product of these vectors, but can be efficiently calculated using bitwise operations,

$$w_{ij} = \frac{N_{\rm real}}{\vec{w}_i \cdot \vec{w}_j} \equiv \frac{N_{\rm real}}{\mathrm{popcount}(\vec{w}_i \& \vec{w}_j)}, \tag{4.4}$$

where & is the bitwise 'and' operator, and popcount is a bitwise operator which sums together the bits of an integer.

The corrected DD counts are calculated from summing the pair weights of galaxies in the separation bin $\vec{s}$,

$$DD_w(\vec{s}) = \sum_{\vec{s}_i - \vec{s}_j \approx \vec{s}} w_{ij} \frac{DD^{(p)}(\theta_{ij})}{DD_w(\theta_{ij})}, \tag{4.5}$$

where $DD^{(p)}(\theta_{ij})$ are the angular DD counts of the parent sample, and $DD_w(\theta_{ij})$ are the angular DD counts of the targeted sample but weighted by the pair weights

w_{ij} (from Eq. 4.4), i.e.

$$DD_w(\theta) = \sum_{\Delta\theta_{ij}\approx\theta} w_{ij}. \tag{4.6}$$

A similar correction is also applied to the DR counts, but this can be done using individual galaxy weights (see Section 4.4.1.4),

$$DR_w(\vec{s}) = \sum_{\vec{s}_i-\vec{s}_j\approx\vec{s}} w_i \frac{DR^{(p)}(\theta_{ij})}{DR_w(\theta_{ij})}. \tag{4.7}$$

In the case where there are no untargetable pairs the PIP estimator is unbiased[1] without this additional angular weighting factor. In this case the ensemble mean of the angular weighting factor is unity and its inclusion is to reduce the variance in the estimator (see Percival & Bianchi, 2017). However, in the case where there are untargetable pairs, the PIP estimator without this factor is biased.[2] Including the angular weighting corrects this bias if, at any separation, the untargeted pairs are an unbiased sample of all the pairs of that separation. The accuracy of this assumption depends on the details of the targeting algorithm. Our results provide a direct test of this for the case of the DESI BGS.

Bianchi et al. (2018) apply the PIP weighting scheme to a DESI ELG mock catalogue, and are able to recover unbiased clustering measurements. However, they do not dither the tile positions, and rely entirely on the angular weighting term to recover the small scale clustering. They also only include ELGs in their catalogue, so do not consider objects with different priorities.

4.4.1.4 Individual Inverse Probability (IIP) Weights

Each galaxy is given an individual weight, which is the inverse of the probability that the galaxy will be targeted, $w_i = 1/p_i$. This can be estimated from the same

[1] Pair weighting takes into account correlations between galaxies in a pair, and is unbiased if each pair has a non-zero probability of being targeted. E.g. if a pair is targeted n times in $N_{\rm real}$ fibre assignment realizations, its weight is $N_{\rm real}/n$, and it is targeted in $n/N_{\rm real}$ realizations, therefore the average weight is 1.

[2] Note that since the pairs with zero probability never enter the pair counts, the expectation value of the estimator is the clustering of the non-zero probability pairs.

bitwise vectors used to estimate the pair probabilities,

$$w_i = \frac{N_{\rm real}}{\rm popcount(\vec{w}_i)}. \tag{4.8}$$

If galaxies are given individual weights, the weight given to a pair of galaxies is the product of these two weights, $w_{ij} = w_i w_j$. This pair weight does not take into account any correlation between galaxy pairs, and will not produce an adequate correction on small scales where pairs are highly correlated.

4.4.2 Clustering Estimates

Correlation functions are calculated using the publicly available parallelized correlation function code TWOPCF[1], which contains an efficient implementation of the PIP weighting scheme. The code can also efficiently calculate jackknife errors in a single loop over the galaxy pairs (Stothert, 2018). To create the random catalogue, we uniformly generate random points on the sky, only keeping those that fall within the patrol region of a fibre, with no dither, so that the random catalogue covers the same footprint as the input catalogue. For illustrative purposes to compare correlation function correction techniques, we assume the parent volume limited sample is known, and assign each object in the random catalogue a redshift randomly sampled from this distribution. This ensures that the number density of objects in the random catalogue has the same evolution as the data catalogue. In the real survey, the parent sample is not known beforehand, but the redshift distribution can be determined by weighting the redshift distribution of the targeted sample by the individual galaxy weights. We have checked, and the scatter between fibre assignment realizations of the weighted $n(z)$ is within 1%. Note that in the case of a flux limited catalogue, the parent sample is known, and this is not an issue.

We also normalize the correlation function using the total number of objects in the parent sample. Again, in the real survey, this is not known, and the normaliz-

[1] `https://github.com/lstothert/two_pcf`

Table 4.4: Definition of the main and extended volume limited samples. Both samples use the magnitude range $-22 < M_r - 5\log h < -21$, where the absolute magnitudes are in the DECam r-band, and k-corrected to $z = 0.1$. $z_{\rm min}$ and $z_{\rm max}$ are the minimum and maximum redshifts, $N_{\rm gal}$ is the total number of galaxies is the sample, $f_{\rm P1}$ is the fraction of priority 1 galaxies, and $\bar{n}$ is the average number density.

sample	$z_{\rm min}$	$z_{\rm max}$	$N_{\rm gal}$	$f_{\rm P1}$	$\bar{n}$ $(h^3{\rm Mpc}^{-3})$
main	0.09	0.30	1,532,903	1.00	1.74×10^{-3}
extended	0.09	0.35	2,655,707	0.94	1.94×10^{-3}

ation factor should be obtained from the pair weights. However, we find that the difference between the normalization factor obtained from the parent sample and from the pair weights is small (a factor $\lesssim 10^{-3}$).

4.4.3 Results

We run the fibre assignment algorithm (Section 4.2) 2048 times on the BGS mock in order to generate weight vectors for each galaxy. In each realization, a random set of 10% of the priority 2 galaxies are promoted to priority 1, and the tile positions are randomly dithered by an angle 3 times the patrol radius ($3R_{\rm patrol} = 4.45$ arcmin). We apply corrections to the clustering measured from two volume limited samples, defined in Table 4.4. The maximum redshift of the main sample is chosen such that the sample only contains priority 1 galaxies, while the maximum redshift is increased for the extended sample so that it also includes priority 2 galaxies. The number densities of the two samples differ slightly, due to evolution of the number density with redshift in the mock.

4.4.3.1 Galaxy Weights

The fraction of galaxies assigned a fibre at least once after $N_{\rm real}$ realizations of the fibre assignment algorithm is shown in Fig. 4.10 for priority 1 and 2 galaxies, with

1 and 3 passes. To achieve a completeness of 99.99% for priority 1 galaxies with 3 passes, only 20 realizations are needed, while the same completeness for priority 2 galaxies requires around 180 realizations. With only a single pass of tiles, the number of realizations needed increases to 50 and 400 for priority 1 and 2 galaxies respectively. There are ~ 10 galaxies that are not assigned a fibre in any of the 2048 realizations. This number is so small that it will have a negligible effect when applying the pair weighting correction to clustering measurements. This number of realizations is sufficient to estimate accurate pair probabilities for the vast majority of galaxy pairs. However, note that the number of galaxies with zero probability, can only be used to infer a lower bound for the number of zero probability pairs.

The distribution of IIP and PIP weights for the main volume limited sample is shown in Fig. 4.11. Most of the priority 1 galaxies are targeted in every fibre assignment realization, and so the distribution of individual weights peaks at unity, with a tail extending to higher weights, due to objects in regions around the edge of the survey that are only covered by a single tile and have a low probability of being targeted. The pair weight distribution has a similar shape, but extends to higher weights. With only one pass, this distribution is very different, since $\sim 90\%$ of the survey is covered by a single tile. There are no objects targeted in every realization, and the individual weight distribution peaks at weight ~ 2, while the pair weight distribution peaks at ~ 5, with a tail extending out to very large weights.

Fig. 4.12 shows the ratio of the total DD counts in angular bins with PIP and IIP weights, for the main volume limited sample, after 1 and 3 passes, illustrating how the correlation between pairs varies as a function of angular separation. On small scales, this ratio is greater than 1, indicating that the targeting probabilities are correlated, and $w_{ij} > w_i w_j$. At intermediate scales, there is a small negative correlation, which asymptotes towards 1 on large scales, where $w_{ij} \sim w_i w_j$. However, even at 10 deg, there is a very weak correlation, and the ratio is offset from 1 by $\sim 10^{-5}$. The size of the small scale correlation depends on the galaxy sample and number of passes. After 3 passes, the DD counts differ by $\sim 4\%$. After only

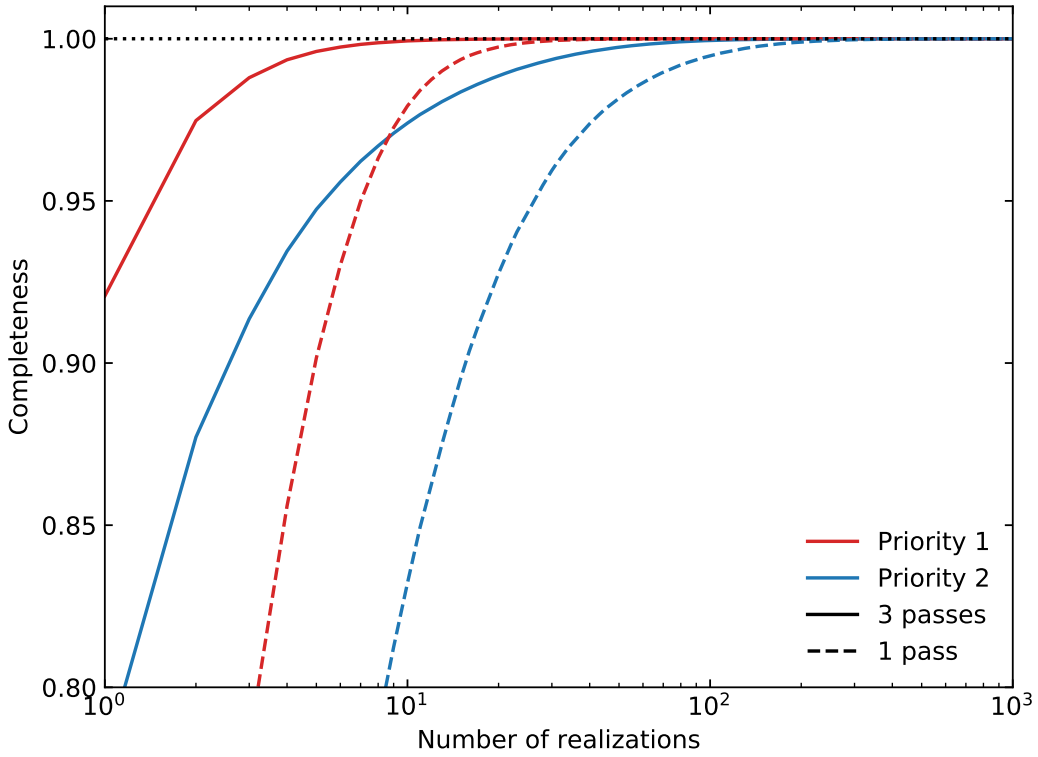

Figure 4.10: Completeness of galaxies that are assigned a fibre at least once after N random realizations of the fibre assignment algorithm. The full flux limited priority 1 and priority 2 samples are shown in red and blue respectively, where solid lines are with the full 3 passes of tiles, and dashed lines a single pass. In each realization, 10% of priority 2 galaxies are randomly promoted to priority 1, and the tile centres are randomly dithered by 3 times the patrol radius.

single pass, since most of the area has single tile coverage, correlations are much larger, and the ratio of DD counts is ~ 1.8.

4.4.3.2 Comparison of mitigation techniques

Fig. 4.13 compares the results of applying several commonly used correction methods to the monopole of the redshift space correlation function of the main volume limited sample, after 3 passes. Each correction is applied to a single realization of the fibre assignment algorithm, and errors are estimated from 100 jackknife samples

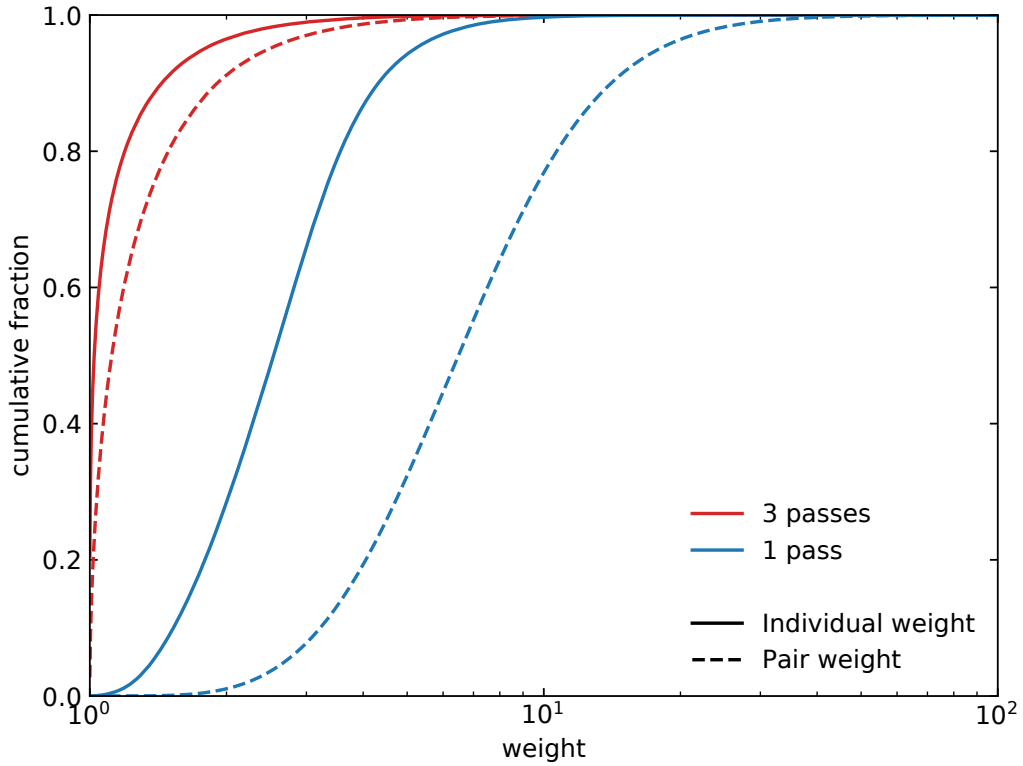

Figure 4.11: Cumulative distribution of individual galaxy weights (solid curves) and pair weights (dashed curves) of objects in the main volume limited sample with 1 (blue) and 3 (red) passes of tiles. For the individual weights, the median, 90th and 99th percentiles are 1.03, 1.44 and 3.04 respectively with 3 passes, and 2.54, 4.33 and 7.70 with a single pass. The same percentiles for the pair weights are 1.12, 1.91 and 4.39 (3 passes) and 6.50, 14.12 and 29.68 (1 pass). After 3 passes. 16% of objects are targeted in every realization, and have a weight exactly equal to 1, while 2.7% of pairs are targeted in every realization.

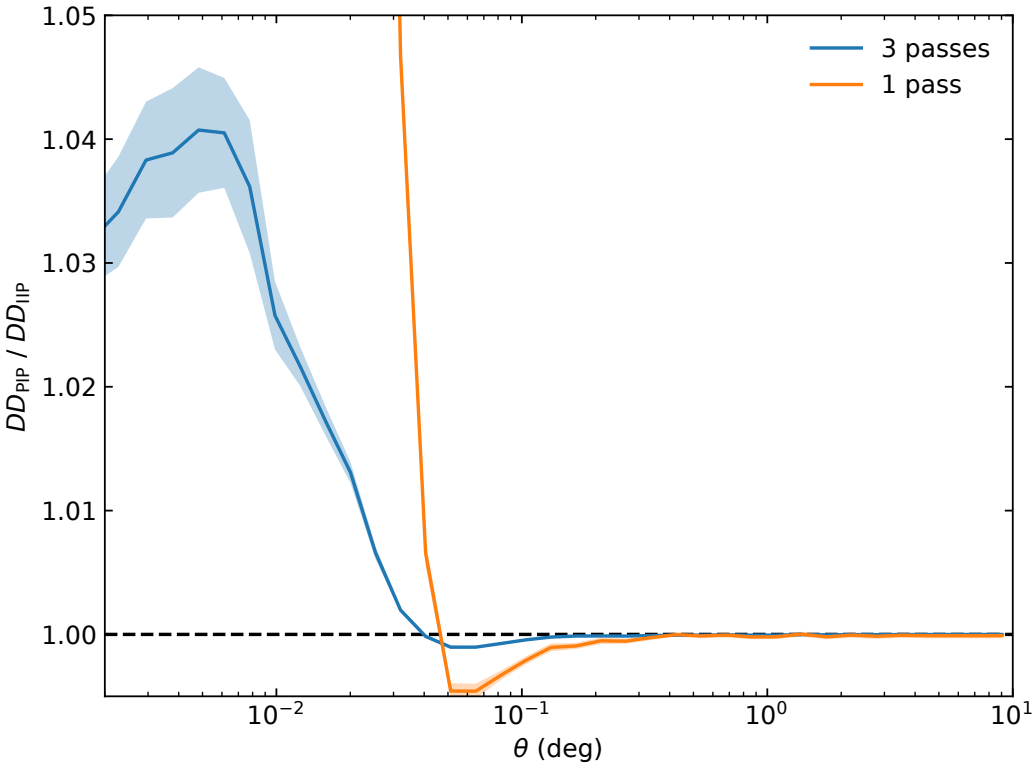

Figure 4.12: Ratio of angular DD counts calculated with pairwise, PIP, weights to that with individual IIP weights, for galaxies in the main volume limited sample, after the full 3 passes of tiles (blue), and after 90% of 1 pass (yellow). The solid curves are the average of 50 fibre assignment realizations, where the shaded regions indicate the 1σ scatter. The black horizontal dashed line indicates a ratio of unity. The ratio on small scales after 1 pass is ~ 1.8.

(see Fig. 4.3). The jackknife error is an estimate of the uncertainty in the clustering measurements due to the finite survey volume. The data is split into 100 regions of equal area, and the correlation function is calculated with each region omitted. The jackknife errors are taken from the square root of the diagonal terms of the covariance matrix. The ratio to the complete parent sample is shown in the lower panel. The purple curve shows the result of applying angular weighting, which by construction, reproduces the angular correlation function of the parent sample. However, this does not provide a satisfactory correction to the monopole. At scales of $\sim 10h^{-1}$Mpc, it differs from the parent sample by $\sim 2\%$, which is approxim-

ately twice the statistical error in the complete sample. At small scales, close to $0.1h^{-1}$Mpc, it differs by almost 10%, while the statistical error in the parent sample is $\sim 5\%$.

Assigning missing objects the redshift of the closest targeted object on the sky, shown by the green curve in Fig. 4.13, does better than angular weighting at large scales, correcting the monopole to a level of $\sim 1\%$. However, this correction produces a strong artificial boost to the clustering at small scales. Some of the untargeted galaxies will be members of clusters, and if the nearest targeted object is also a member of the same cluster, the redshift it is assigned will be close to the true redshift. However, if two galaxies at different redshifts are close together on the sky by chance, the error in the assigned redshift could be large. This chance projection of galaxies boosts the redshift space monopole at $0.1h^{-1}$Mpc by an order of magnitude.

Transferring the weight of missing galaxies to the nearest targeted galaxy on the sky, which is shown by the red curve in Fig. 4.13, produces a correction at large scales that is within 1%. The total weight of galaxy clusters is correct, and so the large-scale clustering agrees with the parent sample. However, since small separation pairs are missing, the clustering on small scales is low.

The PIP correction, shown by the brown curve in Fig. 4.13, produces a correction within $\sim 1\%$ at all scales, even on small scales below a few h^{-1}Mpc where other correction methods fail. Here, the correction is only applied to a single fibre assignment realization, but in the next section we apply the same correction to many realizations to check that is unbiased.

Note that only the monopole is shown in Fig. 4.13. We show in Section 4.4.3.4 the the PIP scheme also works well for the quadrupole and hexadecapole. The other correction methods explored in this section fare less well for the higher order multipoles, only showing agreement with the parent sample on scales larger than a few 10s of h^{-1}Mpc.

The projected correlation function,

$$w_p(r_p) = 2\int_0^{\pi_{\rm max}} \xi(r_p, \pi) d\pi, \tag{4.9}$$

is shown in Fig. 4.14, with the same corrections applied, and using $\pi_{\rm max} = 120h^{-1}$Mpc. The two nearest redshift corrections are able to correct the projected correlation function to within 1% down to a scale of $\sim 0.5h^{-1}$Mpc. Since the projected correlation function integrates along the line of sight, it reduces the impact of galaxies which are assigned the wrong redshift. Again, the PIP weighting produces a correction to within $\sim 1\%$ on all scales.

4.4.3.3 Angular clustering with PIP weights

We now apply the PIP weighting to the angular correlation function. By construction, the angular correlation function of the parent sample is recovered exactly when the pair weighting and angular correction of Eq. 4.5 are both applied. However, it is interesting to see how well the PIP weighting on its own can recover the angular correlation function for a volume limited sample, where in the real survey, the complete parent sample would not be known. To check that the correction is unbiased, we average the result of applying the correction to 50 fibre assignment realizations (which are a subset of the 2048 realizations used to estimate the pair weights). The result, after 3 passes, is shown in Fig 4.15. The left panels show the angular correlation function of the main volume limited sample, with the ratio to the complete parent sample in the bottom panel. The parent sample is shown in blue, where the shaded region is the statistical error, estimated from 100 jackknife samples. The yellow curve shows the correlation function of galaxies assigned fibres in a single realization of fibre assignment, illustrating the size of the correction that needs to be made. The green curve illustrates the result of applying only the pair weighting, without the angular upweighting term, and is the mean of 50 realizations of fibre assignment. The shaded region indicates the 1σ scatter between these realizations. This is the additional error due to measuring the clustering from a subset

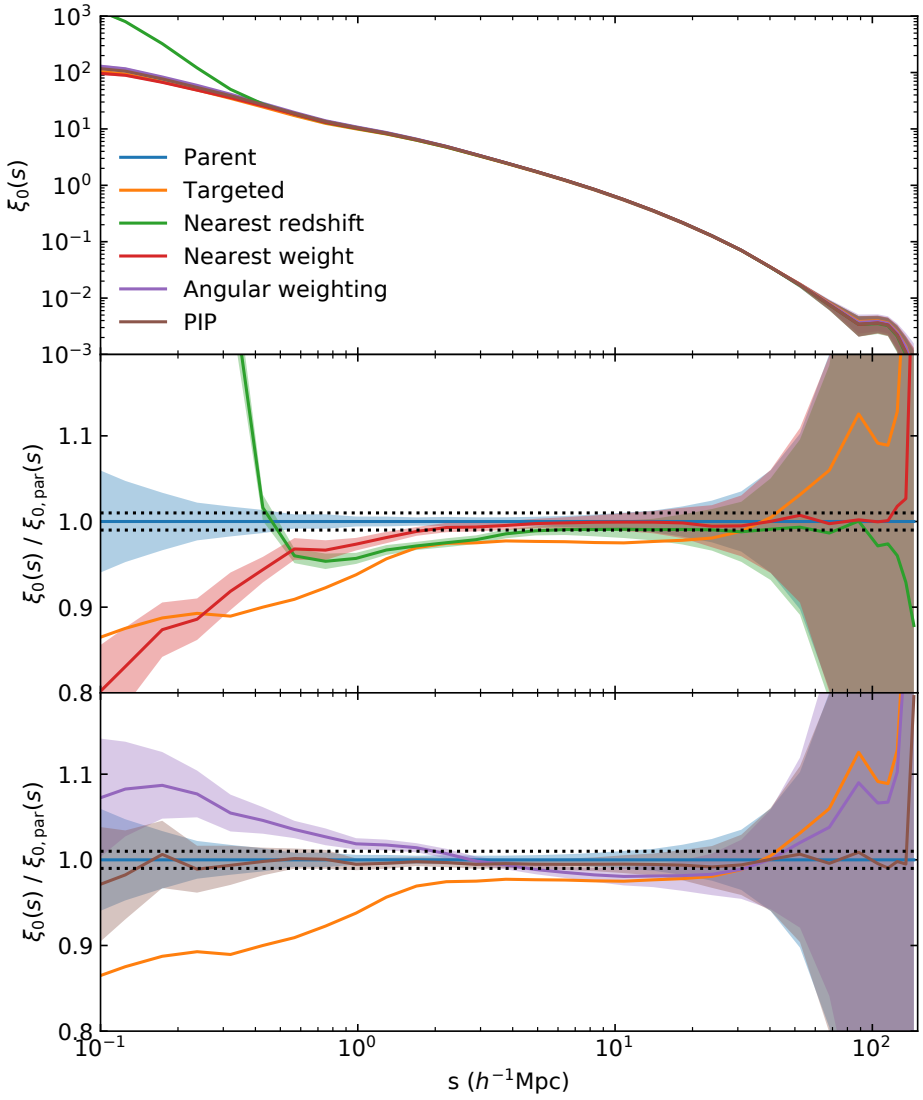

Figure 4.13: Monopole of the redshift space galaxy correlation function of the main volume limited sample, with different corrections applied. The complete parent sample is shown in blue, targeted with no correction in yellow, assigning missing galaxies the redshift of the nearest targeted galaxy on the sky in green, transferring the weight of missing galaxies to the nearest targeted galaxy in red, angular upweighting in purple, and PIP weighting in brown. The two lower panels show the ratio to the complete parent sample, for different cases. Shaded regions are errors estimated from 100 jackknife samples. Horizontal black dotted lines indicate ±1%. For $s \gtrsim 20h^{-1}$Mpc, the scatter is almost the same for all methods.

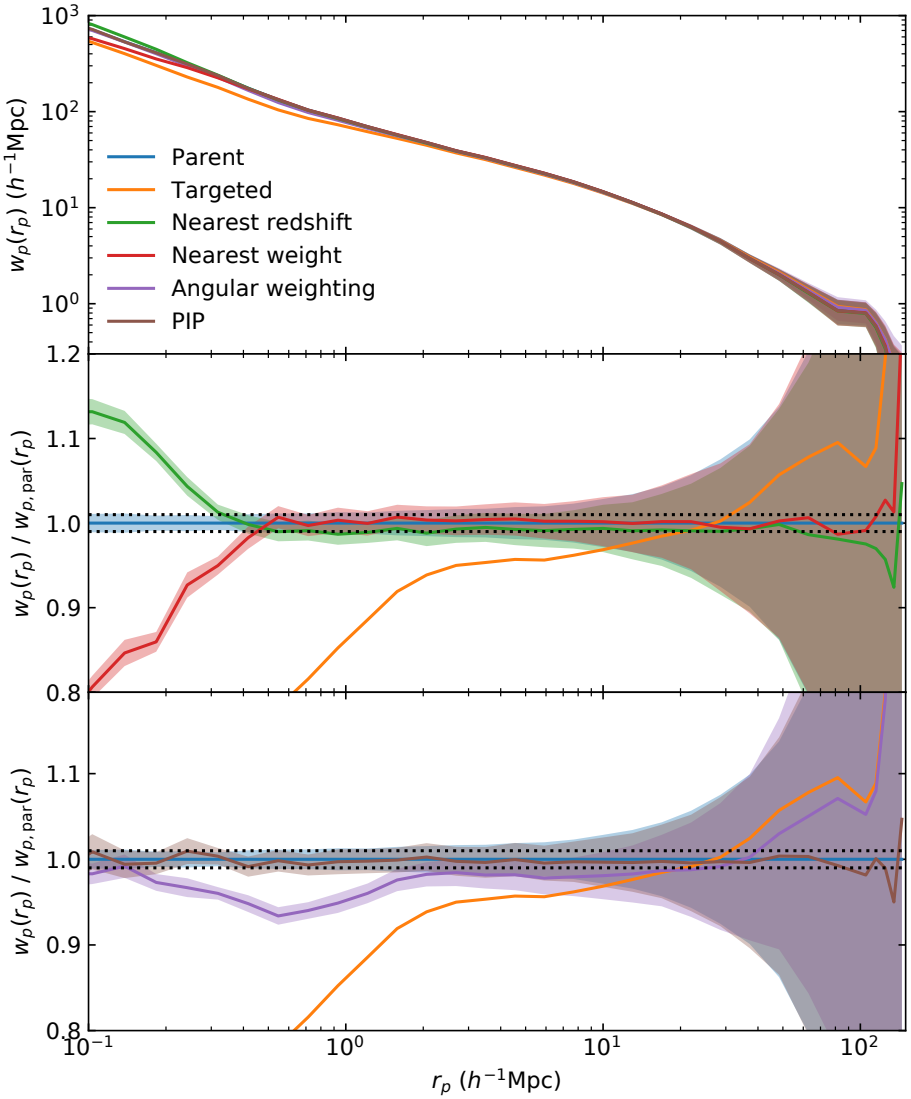

Figure 4.14: Projected correlation function of the main volume limited sample, with the same corrections applied as Fig. 4.13. Shaded regions are errors estimated from 100 jackknife samples.

of the objects in the parent sample, and we aim for this to be small compared to the statistical error in the parent sample. On large scales, the pair weighting does an excellent job of correcting the angular clustering. The mean is unbiased, and the scatter is within 1% for angular scales between ~ 0.03 deg and 1 deg. This is much smaller than the statistical error in the parent sample, which is of the order of a few percent, increasing on larger scales. However, on small scales, less than $0.5R_{\rm patrol}$, there is a small bias of a few percent. This bias is due to pairs of galaxies around the edge of the survey, in regions covered by only a single tile. Pairs of galaxies with a very small angular separation in these regions can never be targeted due to fibre collisions, even when the tiles are dithered. Since these pairs have a zero probability of being targeted, this results in a bias, which is corrected for by the angular upweighting term. It is not guaranteed that this angular correction will be accurate since, for example, missing pairs could occur preferentially in triplets, and therefore be statistically distinct from targeted pairs of the same separation. However, we find that this is not the case, and the missed pairs fall in the regions of single tile coverage. Alternatively, the edge of the survey could be trimmed, removing the regions covered by a single tile, which is only a small percentage of the footprint ($\sim 3\%$, see Table 4.1). Another alternative strategy is discussed in Section 4.4.4.

For comparison, the purple curve shows the result of applying individual galaxy weights to the same set of realizations. At small scales, applying individual weights results in a larger bias than pair weights, and this bias extends to larger angular scales. This is because individual galaxy weights do not take into account any correlation between galaxy pairs. For example, if it is difficult to target both galaxies in a pair due to fibre collisions, but relatively easy to target one or the other individually, calculating the pair probability from individual probabilities is biased since $p_i p_j > p_{ij}$. On large scales, if there are no correlations between pairs, $p_i p_j = p_{ij}$, and using individual weights should produce the same result as pair weights. However, in Fig 4.15, there is still a small difference between the green

and purple curves on large scales. Even at scales of ~ 10 deg, there is still some correlation between galaxy pairs, although this is very small, with a fractional difference in the DD counts of $\Delta DD/DD \sim 10^{-5}$. The fractional error in ξ is given by

$$\frac{\Delta\xi}{\xi} \approx \frac{\Delta DD}{DD}\frac{(1+\xi)}{\xi}. \tag{4.10}$$

On large scales, $\xi \sim 10^{-3}$, which results in a fractional difference of $\Delta\xi/\xi \sim 1\%$, which is a small, but noticeable difference in the correlation function.

The right hand panels of Fig. 4.15 shows the result of applying the same corrections to the extended volume limited sample, which also contains priority 2 galaxies. By giving the priority 2 galaxies a small probability of being promoted to priority 1, this gives every pair of priority 2 galaxies a non-zero probability of being targeted, and therefore applying the pair weighting correction produces an unbiased result on large scales. There is still a small bias on small scales for the same reason as in the main sample.

Fig. 4.16 shows the angular correlation function after only a single pass of tiles, with a random 10% of the tiles missing, for the same volume limited samples. With only 1 pass of tiles the catalogue of fibre assigned galaxies is much less complete, and a larger correction is required.

Since most of the footprint is covered by a single tile ($\sim 90\%$, see Table 4.1), the bias on scales less than $0.5R_{\rm patrol}$ is much larger than after 3 passes. Since there are overlaps between neighbouring tiles, the pair counts on these scales are low, but not zero. Pair weighting must be combined with angular upweighting in order to correct the clustering on these scales.

On larger scales, pair weights on their own are able to produce an unbiased correction, although the scatter between realizations is larger than with 3 passes, but on scales above 1 degree this scatter is approximately half of the statistical error of the parent sample.

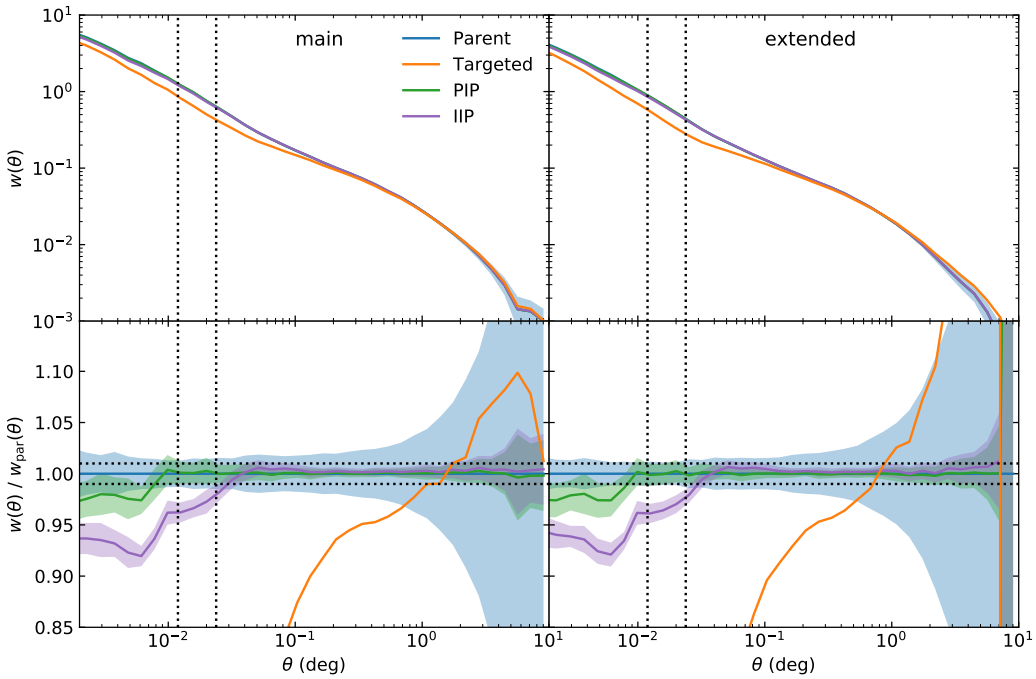

Figure 4.15: Angular correlation function for the main volume limited sample that only contains priority 1 galaxies (left), and the extended volume limited sample that also contains priority 2 galaxies (right), after the full 3 passes of tiles. The bottom panels show the ratio to the complete parent sample. The parent sample is shown in blue, where the shaded region indicates the error from 100 jackknife samples. The yellow curve illustrates the angular correlation function from one realization of fibre assignment, with no correction. Green and purple curves are the results of applying pair weighting and individual galaxy weighting, respectively, averaged over 50 realizations. The shaded regions indicate the scatter between these 50 realizations. Vertical dotted lines indicate the angular scale of $R_{\rm patrol}$ and $0.5R_{\rm patrol}$ and the horizontal lines indicate $\pm 1\%$.

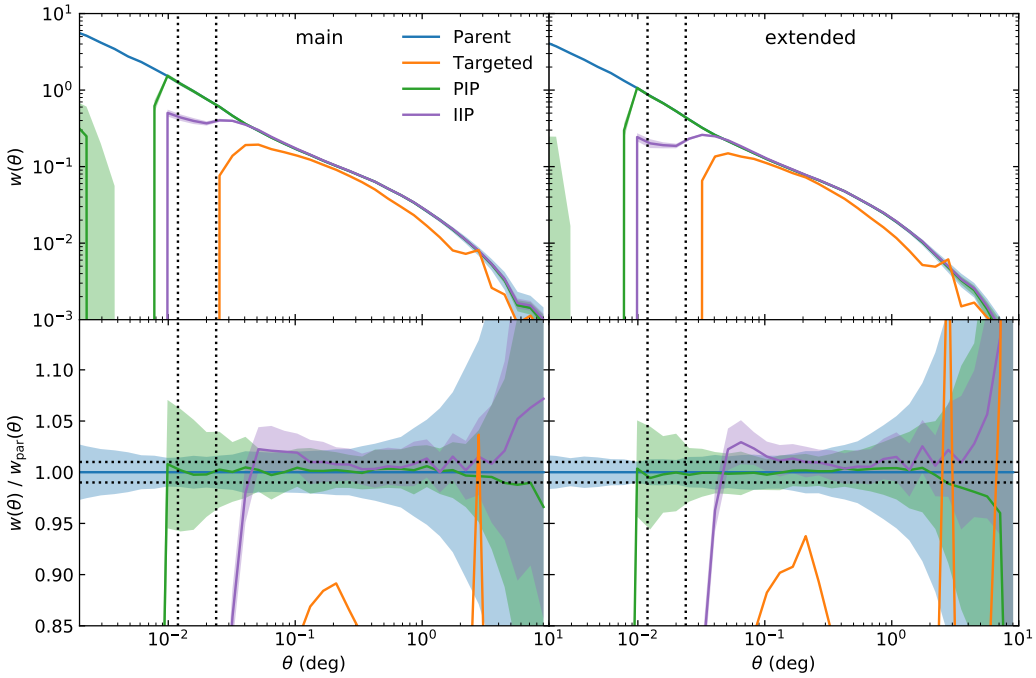

Figure 4.16: As Fig. 4.15 but after only 1 pass of tiles, and with 10% of the tiles missing. This illustrates the data that might have been obtained after one third of the complete survey, with a survey strategy that prioritized area over completeness.

4.4.3.4 Correlation function multipoles with PIP weights

The Legendre multipoles of the redshift space correlation function for the main sample after 3 passes are shown in Fig. 4.17. At large scales, the PIP weighting on its own is unbiased and does a good job of correcting the measured clustering. Between 1 and 10 h^{-1}Mpc, the scatter between realizations in the monopole is well within 1%, and even for the hexadecapole the scatter is around 1%. Note that the scatter in the quadrupole and hexadecapole appears to be large at $\sim 1h^{-1}$Mpc and $\sim 5h^{-1}$Mpc respectively, but this is just because the curves in the upper panels go through zero.

On small scales, similarly to what was seen in the angular correlation function, applying the PIP weighting on its own produces a biased result, due to pairs that cannot be targeted in regions covered by a single tile. Most of this area covered

by a single tile is located around the edge of the footprint. We again find that including the angular weighting term corrects for this small bias.

Fig. 4.18 is the same, but for the extended sample. The results look similar to that of the main sample, showing that including priority 2 galaxies does not produce any biases.

Figs. 4.19 and 4.20 show the results of applying the same corrections to the same volume limited samples, but with only 90% of 1 pass of tiles. Since the survey is much more incomplete, the correction that must be applied is larger. On large scales, applying the PIP weights on their own produces an unbiased correction, but with larger scatter between fibre assignment realizations compared to the 3 pass case. On small scales, the bias is much larger for PIP alone, but combining with angular weighting is able to correct this large small scale bias to within the errors.

After the full 3 passes of tiles, the scatter between realizations is much smaller than the statistical error in the parent sample on all scales. With only a single pass, this scatter is much larger, and on small scales becomes larger than the statistical error. The scatter is large after 1 pass because the sample is highly incomplete (e.g. for the main volume limited sample, $\sim 38\%$ of objects are assigned a fibre in each realization), and most objects have a large weight (the median weight is 2.54, see Fig. 4.11). After 3 passes, the scatter is much smaller, since the completeness of the main sample is much higher ($\sim 82\%$), and most objects have a weight close to unity. 90% of the 1 pass survey area is covered by a single tile, and the completeness of close pairs is very low, due to fibre collisions. Each pair will also have a very large weight, which results in the very large scatter on small scales. The completeness of pairs on small scales is much higher with multiple passes, and therefore the scatter is much smaller.

While the average of many fibre assignment realizations is unbiased, the real survey is only a single realization, and after 1 pass it is likely that there will be

a large scatter between the corrected clustering measurements and the true clustering at small scales. Multiple passes are therefore necessary in order to obtain precise clustering measurements on these scales. On large scales, the scatter is smaller than the statistical error after 1 pass, so it will be possible to make precise BAO and large-scale RSD measurements. However, the uncertainty in these measurements will be greatly reduced after the subsequent passes. Multiple passes will also reduce the incompleteness due to redshift measurement failures, as it will give these galaxies another chance to be targeted. To make precise small scale RSD measurements, a single pass is not sufficient.

The shot noise in these galaxy clustering measurements could potentially be reduced by capping the pair weights at some maximum value. Strictly speaking, the PIP weighting would no longer be unbiased, but this bias can be reduced by the angular weighting term, using these capped weights. We find that for the main sample after 1 pass, capping the weights at a maximum value of 100 (0.01% of pairs) has a negligible affect on the monopole, but reduces the scatter in the quadrupole and hexadecapole at scales of $\sim 1h^{-1}$Mpc by a few percent. Capping the weight at 25 ($\sim 2\%$ of pairs) introduces systematics, which are not corrected for completely by the angular weighting. On large scales, there is a negligible change in the scatter, and the small bias that is introduced is within the errors. On small scales, this bias is larger, but is still within the large errors.

4.4.4 Discussion

We have shown in the previous section that the PIP weighting scheme, in combination with angular upweighting, is able to produce an unbiased correction to clustering measurements in the BGS, even for a highly incomplete survey.

One simplifying assumption we have made is that the galaxies in the parent sample are known. The angular weighting term from Eq. 4.5 includes $DD^{(p)}$, the angular data-data pair counts of the complete parent sample (and similarly for the

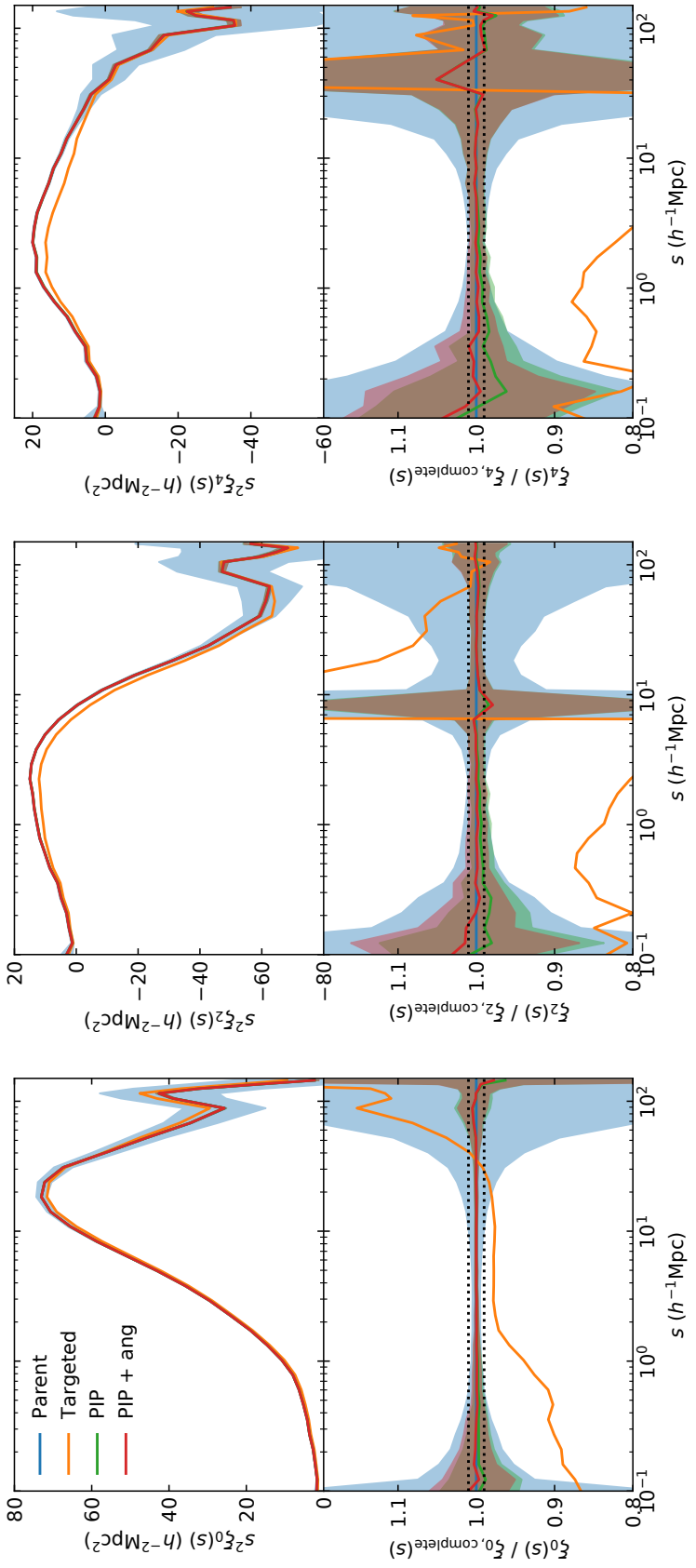

Figure 4.17: Monopole, quadrapole and hexadecapole of the redshift space galaxy correlation function for the main volume limited sample, after 3 passes. The ratio to the complete parent sample is shown in the bottom panel. The parent sample is indicated by the blue curve, where the shaded blue region is the error from 100 jackknife samples. The green curve is the average of 50 realizations, corrected with only PIP weighting. The red curve is corrected using both PIP and angular weighting. Shaded green and red regions indicate 1σ, estimated from the scatter between the 50 realizations. The horizontal dotted lines indicate $\pm 1\%$.

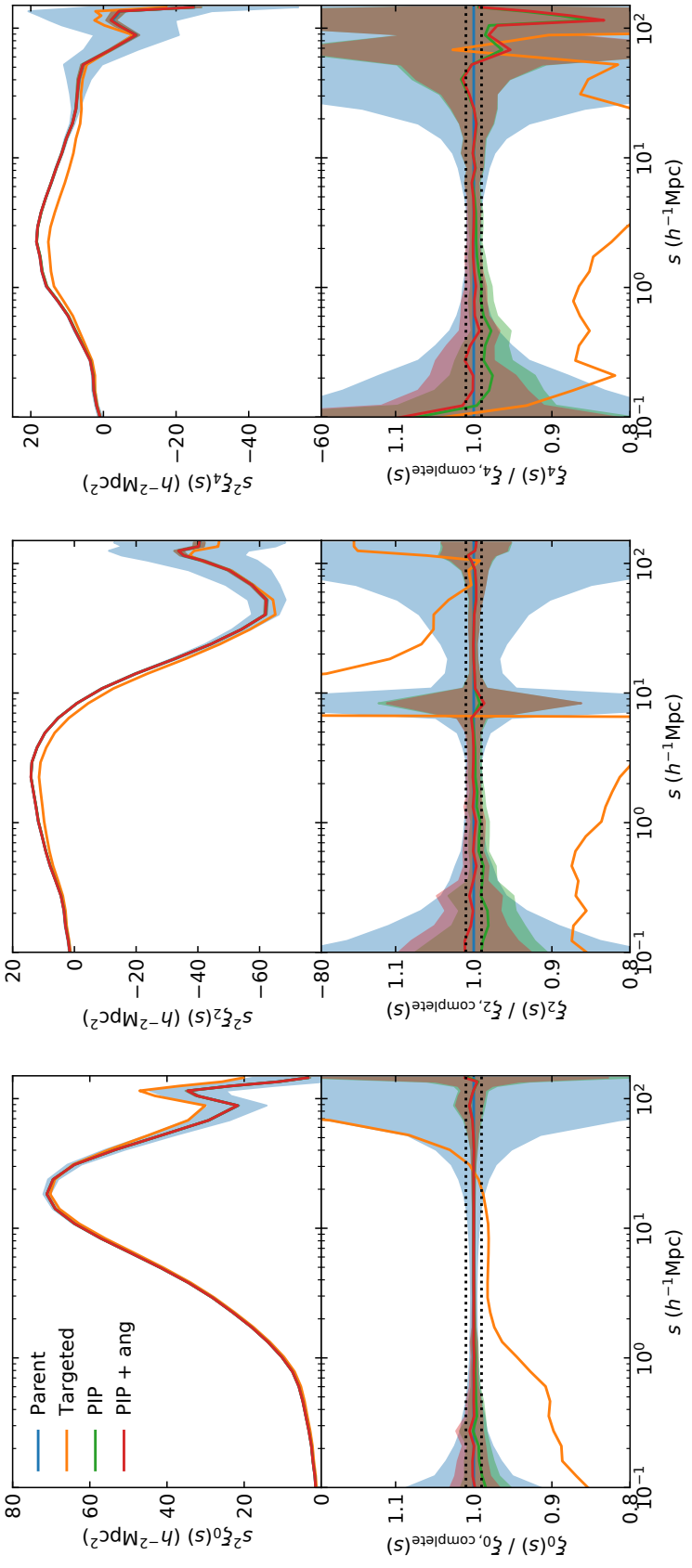

Figure 4.18: As Fig. 4.17, but for the extended volume limited sample, which contains both priority 1 and 2 galaxies

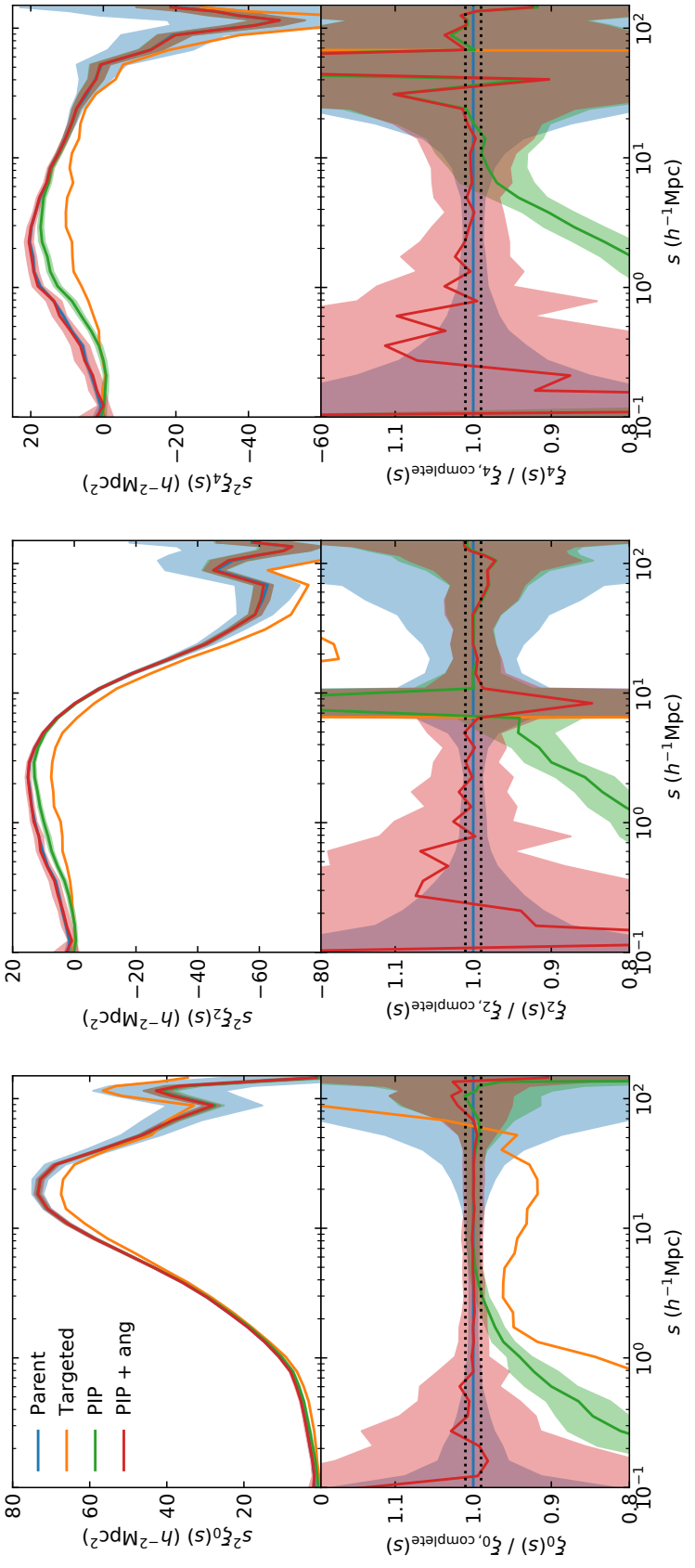

Figure 4.19: As Fig. 4.17, but for the case of only a single pass of tiles.

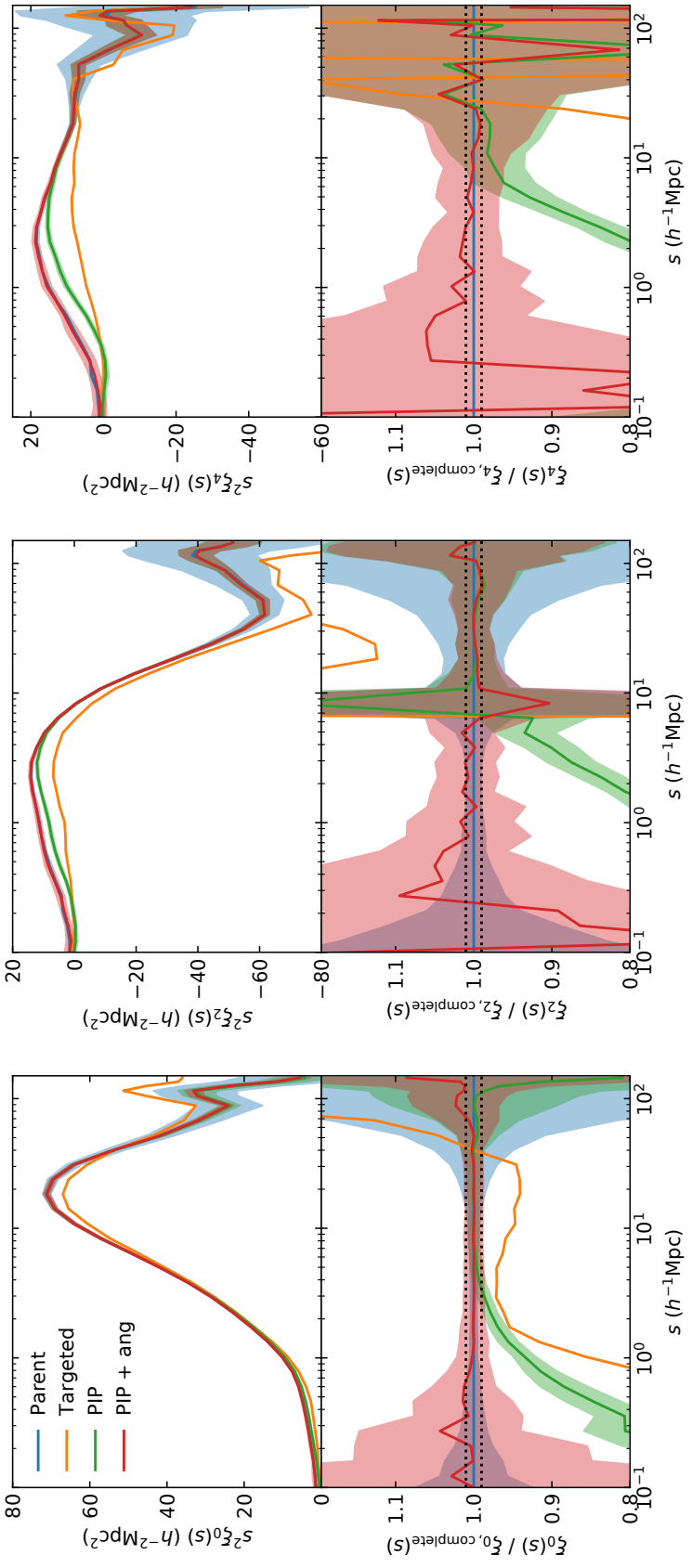

Figure 4.20: As Fig. 4.17, but for the extended volume limited sample, after only a single pass of tiles.

DR counts, $DR^{(p)}$). For a flux-limited sample, the parent sample is known, but this is not true in the case of a volume limited sample, since every redshift would need to be measured to determine an absolute magnitude, and hence which galaxies belong in the sample. When applying the angular weighting, we have used the true parent sample, which in the real survey would not be known.

In order to calculate pair weights, we dither the catalogue by a small angle in each realization of fibre assignment. For galaxies close to the edge of the survey, in half of the realizations they will fall outside the footprint, which results in these galaxies having larger weights than galaxies in the centre. In the actual survey, the dither is zero, which is a special case where no objects fall off the edge, and is not strictly represented in the ensemble of realizations. However, we find no measurable bias as only a very small fraction of objects are affected.

An issue that affects the real survey that we have not considered is stellar contamination. A small fraction of objects in the catalogue of potential targets are stars that have been misclassified as galaxies. If a fibre is placed on one of these objects, and a spectrum measured, it can be determined that it is a star and not a galaxy. Since the PIP weighting scheme can produce an unbiased correction to clustering measurements of any sub-sample of galaxies, the misclassified stars can simply be removed when estimating the correlation function. As long as the stars are included when running the fibre assignment algorithm many times to estimate the PIP weights, this will produce unbiased clustering measurements.

An alternative way to dither the catalogue would be to place the survey tiling randomly on the full sky, with a random orientation. This has the advantage that the undithered catalogue is not a special case, and could be drawn from these random tile positions. Also, every part of the sky has a non-zero probability of being in an area of the survey covered by multiple tile overlaps, giving every pair, even at very small separations, a non-zero probability of being targeted. This means that the w_{ij} pair weights without angular weighting can produce an unbiased correction, so the correction can be applied without knowledge of the complete parent sample.

However, in many of these fibre assignment realizations, the tiling would cover large areas of the sky which are outside the BGS footprint. Despite this, we expect that the total number of realizations needed to accurately estimate pair weights will be smaller, since the tail of pairs with extremely high weights are much more likely to be targeted in the realizations where they are covered by multiple tile overlaps.

A similar method to this is used in Mohammad et al. (2018), where in order to estimate pair weights for galaxies in the VIPERS survey, the parent catalogue is rotated by angles of either 0, 90, 180 and 270 degrees, and the spectroscopic mask is moved spatially. The PIP weighting scheme is shown to work well, and this is the only published example of applying the PIP weights to a real dataset.

With large dithers across the full sky, it is also necessary to modify the definition of pair weights to take into account that galaxies will fall outside the survey tiling in many of these realizations of fibre assignment. Consider a perfect survey in which if two galaxies fall within the survey tiling, it is always possible to target the pair, so all pairs should have the same weight. If the pair have a very small angular separation, then in 1/3 of realizations they will fall within the tiling and be able to be targeted, so they would have a pair weight of 3, using Eq. 4.4. However, if a pair has a very large separation, it can be unlikely that both fall within the tiling at the same time in a random realization, so the pair probability is low and therefore the weight will be much larger than 3. Eq. 4.4 incorrectly gives pairs of different separations different weights. Instead, the pair weight can be redefined as

$$w_{ij} = \frac{\vec{c}_i \cdot \vec{c}_j}{\vec{w}_i \cdot \vec{w}_j}, \tag{4.11}$$

where $\vec{c}_i$ is a bitwise coverage vector that contains a 1 if it is possible to place a fibre on galaxy i (i.e.the galaxy lies within the patrol region of a fibre though it may happen not to be targeted) in that realization, and 0 otherwise.[1] Applying this definition in the above example results in all pairs having a weight of 1, as

[1]The ability to use bitwise coverage vectors is implemented in the correlation function code TWOPCF (Stothert, 2018).

expected.

We have only shown the results of applying the correction to volume limited samples with a number density $\sim 2 \times 10^{-3} h^3 \mathrm{Mpc}^{-3}$. We have also applied the correction to volume limited samples of different number densities, and samples defined by a colour cut, and we find that applying the PIP correction with angular weighting will produce an unbiased correction.

The mock catalogue used was constructed using a set of HODs fit to clustering measurements from SDSS (Zehavi et al., 2011). These measurements of galaxy clustering are corrected for the effects of fibre collisions using the 'nearest redshift' correction, where each missing galaxy is assigned the redshift of its nearest targeted neighbour on the sky. The PIP method is not specific to any galaxy survey, and in principle could be applied to SDSS. However, in the SDSS survey, fibres can be placed anywhere on the plate, as long as they are not closer together than 55 arcsec. SDSS also covers a narrower redshift range than is expected for the BGS. In this case, the nearest redshift correction does well at correcting the projected correlation function, and it is not necessary to use the PIP weighting scheme.

4.5 Conclusions

The DESI BGS will be a highly complete, flux limited spectroscopic survey of low redshift galaxies, an order of magnitude larger than existing galaxy catalogues, with the primary science aims of BAO and RSD analysis. Fibres in the focal plane of the telescope are controlled by robotic fibre positioners, each of which can place a fibre on any galaxy within a small patrol region, leading to incompleteness in the catalogue due to fibre collisions, and the fixed density of fibres over large regions in each tile. This leaves a non-trivial impact on clustering measurements, and it is essential that these biases can be corrected.

We have quantified the targeting completeness in the BGS by applying the DESI fibre assignment algorithm to a BGS mock catalogue. To ensure each galaxy

has a non-zero probability of being targeted, and to maximize the number of pairs that can be targeted, we randomly promote 10% of faint priority galaxies to the same priority as the bright priority 1 galaxies, and dither the tile positions by a small angle of 3 times the fibre patrol radius.

The main determinant of completeness in the BGS is the surface density of galaxies. Completeness is high in low surface density regions, (e.g. over 95% for priority 1 galaxies after 3 passes), but drops significantly in the most overdense regions. Close to the centre of the very most massive haloes ($\sim 10^{15}h^{-1}\mathrm{M}_{\odot}$), the completeness can be as low as 10% or less.

We applied several correlation function correction methods to volume limited samples from the BGS mock catalogue, where the incompleteness is due to fibre assignment only. This is done for a highly complete survey with 3 passes of tiles, and a highly incomplete survey, with 1 pass and 10% of the tiles missing. Using standard angular upweighting, or assigning missing galaxies the redshift of the nearest targeted galaxy provide an unsatisfactory correction to the correlation function monopole on small scales below a few Mpc (and a few 10s of Mpc for the higher order multipoles).

After 3 passes of tiles, the method of Bianchi & Percival (2017), which combines galaxy pair weights with an angular weighting, is able to produce an unbiased correction to the angular and redshift space correlation functions, where the scatter between fibre assignment realizations is much smaller than the statistical error in the complete parent sample. The angular weighting term is required to correct a small bias on small scales caused by untargetable pairs around the edge of the survey footprint. After 1 pass, the correction is again unbiased, but the scatter between realizations is much larger, and on small scales the method relies heavily on angular weighting. More than 1 pass will be needed to make precise RSD measurements on small scales.

We propose an alternative method to dither the tiles, where the entire survey

tiling is positioned randomly on the full sky, and the pair weight definition takes into account realizations in which objects cannot be targeted. This has the advantage that pair weighting on its own can produce an unbiased correction without relying on angular weighting.

My contribution to this work was to use the MXXL mock catalogue to explore the impact of fibre assignment on the completeness of galaxies in the BGS, and to assess different correlation function correction methods. Modifications to the DESI fibre assignment code were made by Jianhua He, who also ran the code 2048 times. The correlation function code which implements the PIP correction method was developed by Lee Stothert.

Chapter 5

HOD mocks for the Euclid galaxy redshift survey

5.1 Introduction

In Chapter 3, we outlined a method for creating a halo lightcone catalogue from the MXXL simulation by interpolating halo positions, velocities and masses between simulation snapshots. This method has been used to construct a lightcone out to redshift $z = 2.2$. This catalogue was subsequently populated with galaxies using a halo occupation distribution (HOD) scheme to build a catalogue for the DESI Bright Galaxy Survey (BGS) that has realistic galaxy clustering properties. The halo lightcone can also be populated using different HOD schemes to make catalogues for other galaxy surveys. The full redshift range is not needed for the BGS, since only a small fraction of galaxies in the survey have redshifts beyond $z = 0.5$. However, the outer redshift limit of $z = 2.2$, and halo mass resolution of $\sim 10^{11}\ h^{-1}\mathrm{M}_{\odot}$, make this halo lightcone useful for making a mock catalogue of Hα sources for the European Space Agency's upcoming Euclid survey (Laureijs et al., 2011). Euclid aims to constrain the expansion history of the Universe by conducting an imaging survey, and a slitless spectroscopic survey. The imaging survey will make measurements of gravitational weak lensing, while the slitless

spectroscopic survey, which we focus on here, will be a survey of Emission Line Galaxies (ELGs).

The HOD specifies the average number of central and satellites in each halo, brighter than a luminosity threshold. As was shown in Chapter 3, a set of HODs with different luminosity thresholds can be used to randomly assign galaxies, with luminosities, to the haloes in the lightcone. In order to build the BGS mock, a 5 parameter HOD was used, where the central HOD is modelled as a smooth step function, while the satellite HOD is a power law with a cut off at low masses. This standard parametrisation is modified to use a pseudo-Gaussian spline kernel to make sure there is no unphysical crossing of HODs between different magnitude thresholds. However, any HOD parametrisation can be used, as long as there is no unphysical HOD crossing. The method can also be extended to cases where it is difficult to parametrise the HODs. If the HOD is measured in bins of mass at different redshifts and luminosity thresholds, these measured HODs can be interpolated in order to populate the haloes.

Euclid will predominantly be measuring the spectra of luminous star-forming Hα ELGs, to a flux limit of 3×10^{-16} erg s^{-1}cm^{-2}. In the DESI BGS, the r-band luminosity of central galaxies increases monotonically with halo mass (with some scatter). However, the Hα luminosity is driven by the star formation rate, which is not, making the HOD parametrisation used for the BGS unsuitable for creating a Euclid Hα mock catalogue, and motivating the need to extend the methods for a tabulated HOD.

In this chapter, we outline ongoing work in extending the HOD methodology for a tabulated HOD. The MXXL halo lightcone is populated using a set of HODs for Hα emitters that have been extracted from the GALACTICUS semi-analytic model (Benson, 2012). This chapter is organised as follows. Section 5.2 describes the HODs measured from GALACTICUS. Section 5.3 extends the HOD method for tabulated HODs. The luminosity function produced using the GALACTICUS HODs is compared to measured luminosity functions in Section 5.4. The conclusions are

summarised in Section 5.5.

5.2 Hα HODs from the GALACTICUS semi-analytic model

GALACTICUS (Benson, 2012) is a semi-analytic model of galaxy formation which creates and evolves a population of galaxies within the dark matter halo merger trees of an N-body simulation. GALACTICUS models various astrophysical processes, including gas cooling, star formation, chemical enrichment, and feedback from supernovae and active galactic nuclei, and is calibrated to produce the present day galaxy stellar mass function. GALACTICUS has been applied to the Millennium simulation (Springel et al., 2005), in order to create a catalogue of Hα emitters. A halo lightcone is first created from the Millennium simulation, which is then populated using the GALACTICUS model. By combining the semi-analytic model with a model for dust attenuation, predictions can be made of the number counts and redshift distribution of Hα sources in future surveys such as Euclid and WFIRST (Merson et al., 2018).

The HOD for Hα sources can be measured from this lightcone by simply calculating the average number of central and satellite galaxies, brighter than some luminosity threshold, in haloes, binned by halo mass and redshift. The luminosity is the blended Hα + [NII] luminosity[1], with no dust extinction, and halo masses are defined as $M_{200\mathrm{m}}$, the mass within a sphere centred on the halo in which the average density is 200 times the mean density of the Universe. The HOD is measured in 26 mass bins between $\log(M_{200\mathrm{m}}/h^{-1}\mathrm{M}_\odot) = 9.7$ and $\log(M_{200\mathrm{m}}/h^{-1}\mathrm{M}_\odot) = 14.7$, 31 redshift bins between $z = 0.7$ and $z = 2.2$, and for 30 luminosity thresholds, from $\log(L_{\mathrm{H}\alpha+[\mathrm{NII}]}/h^{-2}\mathrm{erg\ s}^{-1}) = 38$ to $\log(L_{\mathrm{H}\alpha+[\mathrm{NII}]}/h^{-2}\mathrm{erg\ s}^{-1}) = 43$. Euclid is expected to be able to target Hα sources over the approximate redshift range

[1]The low spectral resolution of Euclid ($\lambda/\Delta\lambda \approx 300$) means that the H$\alpha$ and [NII] line cannot be separated.

$0.9 \lesssim z \lesssim 1.8$, to a flux limit of 3×10^{-16} erg s^{-1}cm^{-2}, over an area of 15,000 deg^2. The Wide Field Infrared Telescope (WFIRST) (Green et al., 2012; Spergel et al., 2015) is expected to probe a similar redshift range of $1 \lesssim z \lesssim 2$, to a fainter flux limit of 1×10^{-16} erg s^{-1}cm^{-2}, but over a smaller area of $\sim 2,200$ deg^2.

The resolution of the MXXL simulation is too low to apply the semi-analytic model directly, but the HODs measured using the Millennium simulation can be applied to build a mock catalogue for the Euclid survey. However, the HODs are only measured up to a maximum mass of $\log(M/h^{-1}\mathrm{M}_\odot) = 14.7$. Since the MXXL simulation covers a much larger volume, it contains many haloes more massive than this, so the HODs need to be extrapolated to higher masses. However, we expect that the Hα emitters in such massive haloes will account for a small fraction of the total number of objects. While the most massive haloes will on average contain many Hα emitters, they are in the tail of the mass function, which is dropping rapidly, so the overall contribution is small.

5.3 Extending the HOD method for tabulated HODs

The occupation number, $\langle N_{\rm gal}(> L|M, z)\rangle$, measured for central and satellite galaxies, can be stored in a 3 dimensional array, and the values can be interpolated to find the occupation number at any luminosity, redshift or mass. In order to populate the MXXL halo lightcone, the HODs can be extrapolated to higher masses by fitting smooth functions to the high mass end of the HODs.

Fig 5.1 shows the HODs of centrals, $\langle N_{\rm cen}(> L|M, z)\rangle$. The points are the HODs measured from GALACTICUS, and each panel is at a different redshift. The shape of the Hα HODs differ from those used to create the BGS mock. For the faint samples, the occupation number is close to 1 at all masses above $10^{11}\ h^{-1}\mathrm{M}_\odot$. For the brightest samples, the HOD peaks at $\sim 10^{12}\ h^{-1}\mathrm{M}_\odot$, but then increases again at very high masses. To extrapolate these to high masses, care must be taken to ensure that there is no unphysical crossing of the HOD, especially for the fainter

luminosity bins where the curves are tightly packed together. We find that for all luminosity thresholds, $\log(-\log N_{\rm cen})$ is approximately linear with $\log M$ at high masses, and the luminosity thresholds are all approximately evenly spaced. For each luminosity threshold, a straight line is fit to $\log(-\log N_{\rm cen})$ above $10^{14}\ h^{-1}{\rm M}_{\odot}$. However, these fits could potentially cross, and also the brightest samples are poorly measured, making it difficult to fit a line. Another linear fit is made to the slope of the initial fits, as a function of $\log L$. This can be extrapolated to set the slope of $\log(-\log N_{\rm cen})$ for the poorly measured bright samples, and it also ensures that the HODs never cross. The solid curves in Fig 5.1 show these extrapolations to high masses. At high redshifts, the curves are in good agreement with the points for all luminosity thresholds. The agreement is less good at low redshifts for the brightest samples. However, only a small fraction of objects have such bright luminosities, and as we show later, the luminosity function produced by these HODs is in agreement with observations.

The occupation functions for satellites galaxies, $\langle N_{\rm sat}(> L|M,z)\rangle$, are shown in Fig. 5.2. Again, the points are the measured HODs from GALACTICUS. In the BGS HODs, the satellites are modelled as a power law, with a cutoff at low masses. The GALACTICUS HODs can be described well as a double power law. For the faint samples, a power law is fit to masses above $10^{13}\ h^{-1}{\rm M}_{\odot}$, which is used to extrapolate the HODs. However, this cannot be done for the brightest samples, which are poorly measured. A double power law is fit to the sample with $\log(L/h^{-2}{\rm erg\ s}^{-1}) > 42.14$, and this fit is offset vertically for the brighter samples. The solid curves in Fig 5.2 show theses fits, which are able to reasonably reproduce the measured HODs for the brightest samples.

The total HOD for centrals and satellites is shown in Fig. 5.3. At high masses, the smooth curves that are used to extrapolate the HODs are in good agreement with the measured points. At low masses, $\log N$ is interpolated linearly between each point. There are some cases where the measured HOD for the brightest luminosity threshold is, in one mass bin, the same as for the next brightest sample.

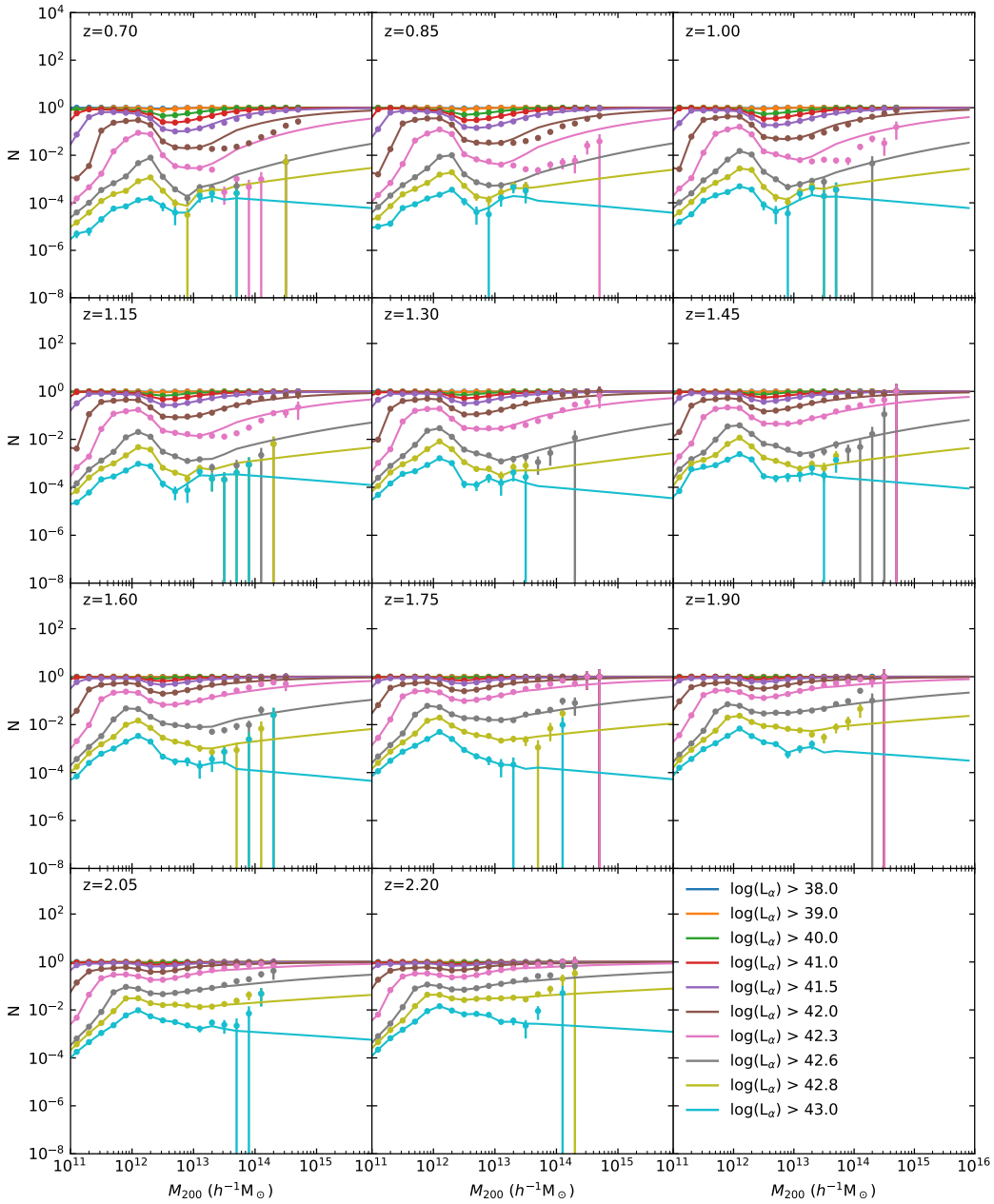

Figure 5.1: Occupation function of central Hα emitters. Points with error bars are the HODs measured from the GALACTICUS semi-analytic model, and the solid curves are the same HODs with smooth curves fit to the high mass end to enable the HODs to be extrapolated. Each panel is a different redshift, and the colours are different luminosity threshold, as indicated by the legend. Blended Hα + [NII] luminosities are in units h^{-2}erg s^{-1}.

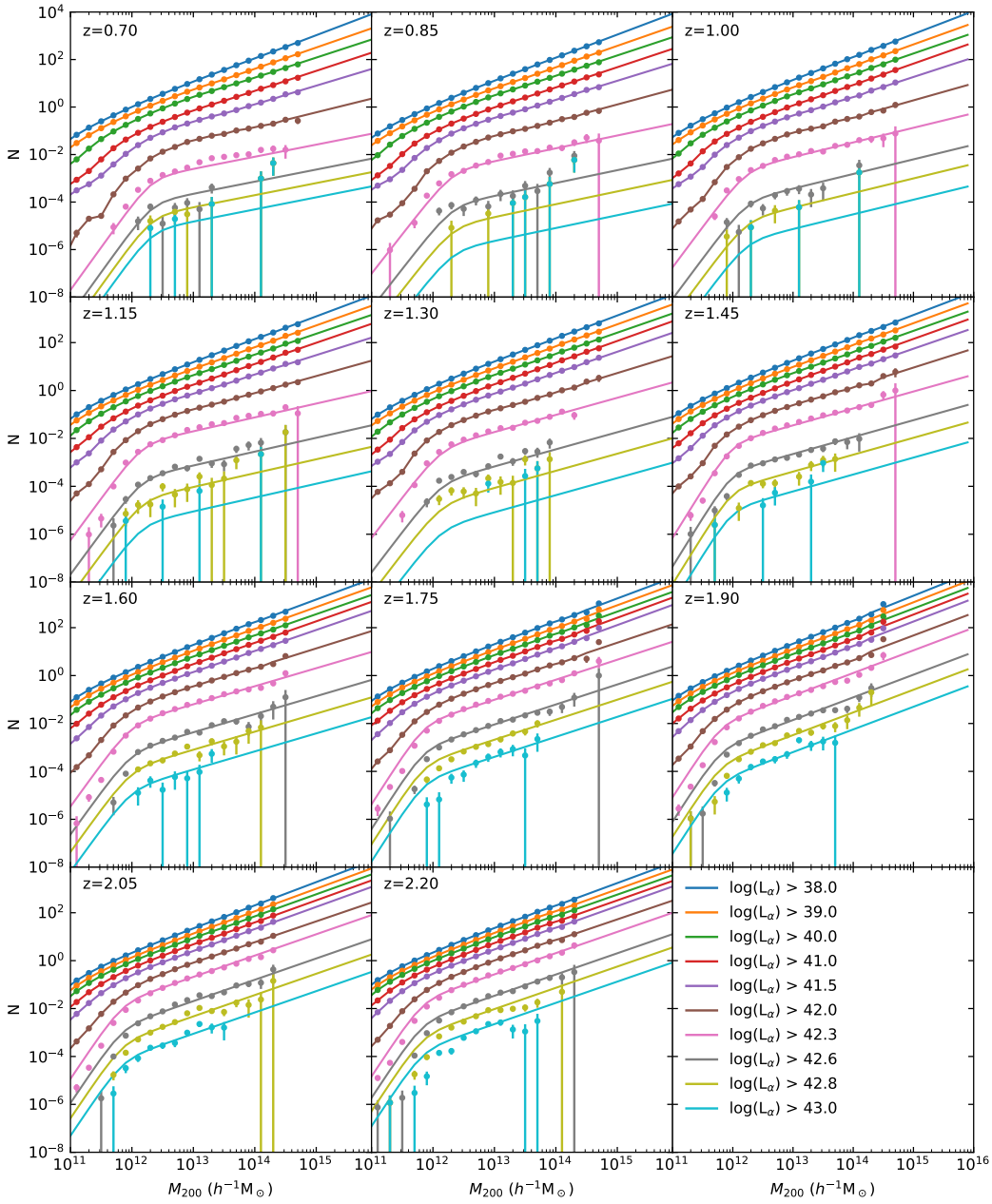

Figure 5.2: As Fig. 5.1, but for satellite galaxies.

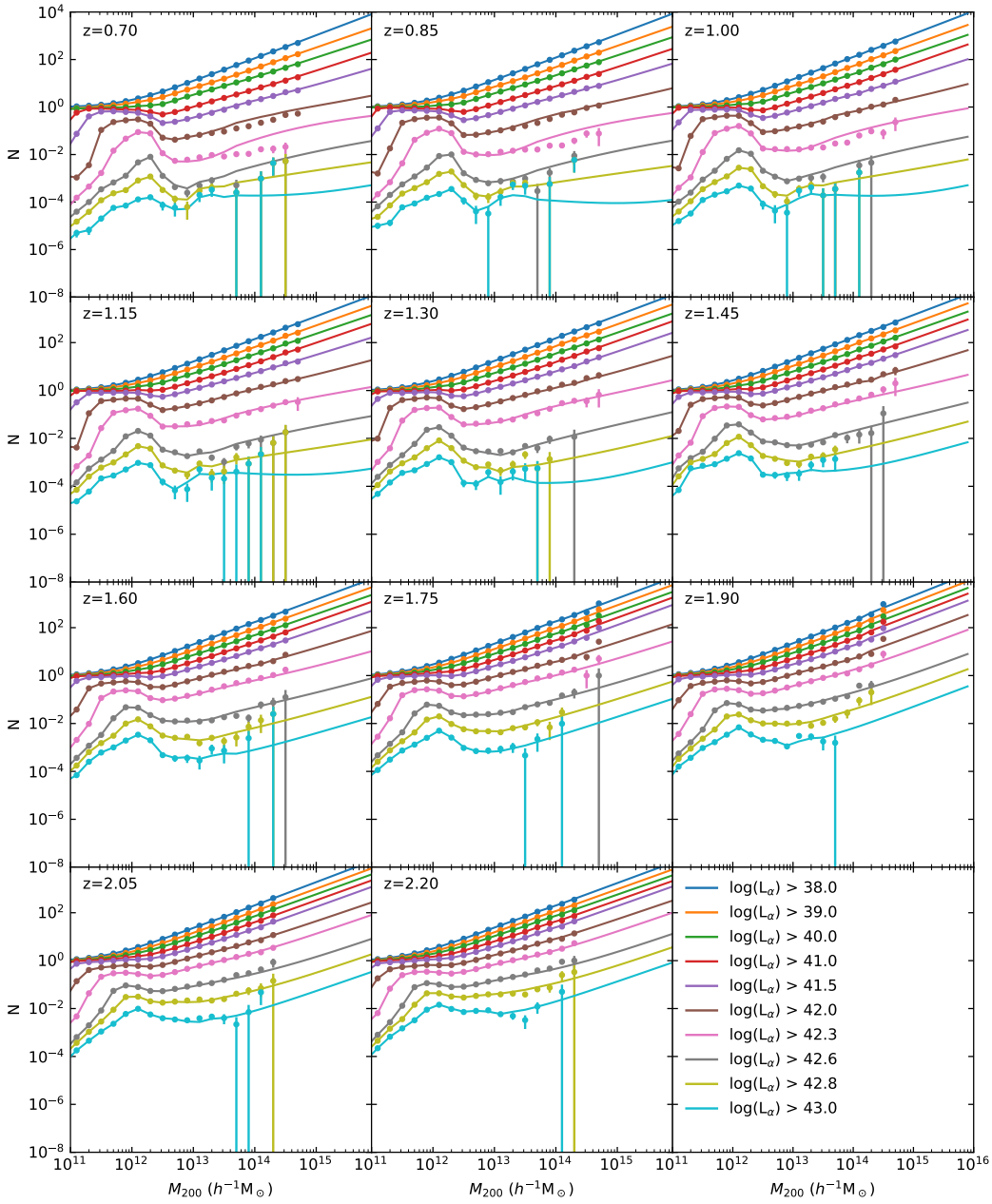

Figure 5.3: As Fig. 5.1, but for centrals and satellites.

This can cause problems when extrapolating the HOD as a function of luminosity since, if the HOD is extrapolated linearly, it would be constant with increasing luminosity. This can be resolved by, for the brightest sample, shifting the HOD down in these bins vertically, making sure the offset is at least 0.3 dex.

These HODs are then used to assign galaxies to haloes, using a method similar to that which is outlined in Chapter 3. A minimum luminosity, $L_{\rm min}$, must first be chosen. To assign central galaxies, it must first be decided which haloes contain a central galaxy brighter than $L_{\rm min}$. For each halo, a uniform random number x_1 in the range $0 < x_1 < 1$ is chosen, and compared to the average occupation number for that halo. If $x_1 < \langle N_{\rm cen}(> L_{\rm min}|M, z)\rangle$, a central galaxy will be placed in the halo, and a random luminosity must be chosen. To assign a luminosity to the central galaxy, another uniform random number, x_2, is chosen, again in the range $0 < x_2 < 1$. The luminosity is found such that

$$\frac{\langle N_{\rm cen}(> L|M, z)\rangle}{\langle N_{\rm cen}(> L_{\rm min}|M, z)\rangle} = x_2 \tag{5.1}$$

In order to speed up the process of finding the root of Eq. 5.1, a 3 dimensional array of L as a function of M, z, and random number x_2 is created, which can be searched and interpolated efficiently.

The number of satellite galaxies in each halo is drawn from a Poisson distribution with mean $\langle N_{\rm sat}(> L_{\rm min}|M, z)\rangle$. The same procedure as for the centrals is used to assign luminosities, drawing a random number for each satellite galaxy, and using Eq. 5.1 but with the satellite HODs to find the corresponding luminosity.

Central galaxies are placed in the centre of the halo, with the same velocity. Satellite galaxies are positioned randomly following a NFW profile, and assigned a random velocity, using the same methodology outlined in Chapter 3.

5.4 Luminosity function

We now use this method to populate several snapshots of the MXXL simulation, and compare the Hα luminosity function of galaxies in the mock with observational values.

The accuracy of the luminosity function depends on the number of luminosity bins in which the HOD has been measured from the semi-analytic model, and the method used to interpolate the HODs. The HOD as a function of mass, redshift, and luminosity threshold can be stored as a 3 dimensional array. Linear interpolation of a 3 dimensional array is straightforward, but other interpolation schemes are non-trivial. If $N_{\rm gal}$ is interpolated linearly, this results in step features in the luminosity function. This is because if the HOD is measured in the semi-analytic model at the luminosity thresholds L_1 and L_2 (where $L_2 > L_1$) and $N_{\rm gal}$ is interpolated between the two luminosities linearly, then the HOD in narrow luminosity bins is constant between L_1 and L_2. Alternatively, if $\log N_{\rm gal}$ is interpolated linearly, in the bin between L_1 and L_2, more galaxies will be assigned a luminosity near L_1, leading to a luminosity function with sawtooth-like features. The more luminosity thresholds that the HODs are measured at in the semi-analytic model, the smaller these features become, but they are never removed completely.

Other interpolation methods, such as cubic splines, are not trivial in 3 dimensions, but can be done easily in a single dimension. The features in the luminosity function can be smoothed, for a fixed mass and redshift, by interpolating the HOD using a monotonic cubic spline interpolation (Fritsch & Carlson, 1980). A new 3 dimensional array of the HOD can be created with much finer bins of luminosity, and the spline interpolation must be monotonic to prevent unphysical HOD crossing. Values of $\log N_{\rm gal}$ in this new array are then interpolated linearly, as before. The luminosity function is shown in Fig. 5.4 for three MXXL snapshots populated with Hα emitters. The luminosity function has been dust attenuated using the empirical law of Calzetti et al. (2000). Small bumps can be seen in the luminosity function of

the mock, particularly at the faint end, but these bumps are small compared to the errors in the measured luminosity function of Sobral et al. (2013). The luminosity function in the mock is in good agreement with the measured luminosity function. The downturn at low luminosities is below the flux limit of Euclid.

5.5 Conclusions

Chapter 3 outlined a method for creating a mock catalogue for the DESI Bright Galaxy Survey, in which a halo lightcone from the MXXL simulation is populated with galaxies using a standard 5 parameter HOD, which is modified to prevent unphysical HOD crossing. The MXXL halo lightcone catalogue extends to $z = 2.2$, making it useful for creating mock catalogues for other redshift surveys, such as Euclid and WFIRST. The HOD methodology can be extended to tabulated HODs, in which the HOD has been measured in bins of mass, luminosity, and redshift.

The GALACTICUS semi-analytic model has been applied to the Millennium simulation to create a catalogue of Hα emitters. This has been used to measure the HOD of central and satellite galaxies. These HODs can then be used to populate the MXXL lightcone. Since the MXXL simulation contains haloes much more massive than the Millennium simulation, the HODs need to be extrapolated to high masses. We fit smooth curves to the high mass end of the HODs, taking care to make sure there is no unphysical HOD crossing. Several snapshots of the MXXL simulation have been populated using these HODs, and the HODs are able to reproduce the luminosity function measured in HiZELS.

In ongoing work, the HODs will be used to populate the halo lightcone, which can be used for clustering analysis. In particular, the motivation of this work is to determine if there is any scale dependence in the form of the bias of Hα emitters on large scales.

My contribution to this work was to develop the code to extend the HOD method for a tabulated HOD. I populated the MXXL simulation using HODs that

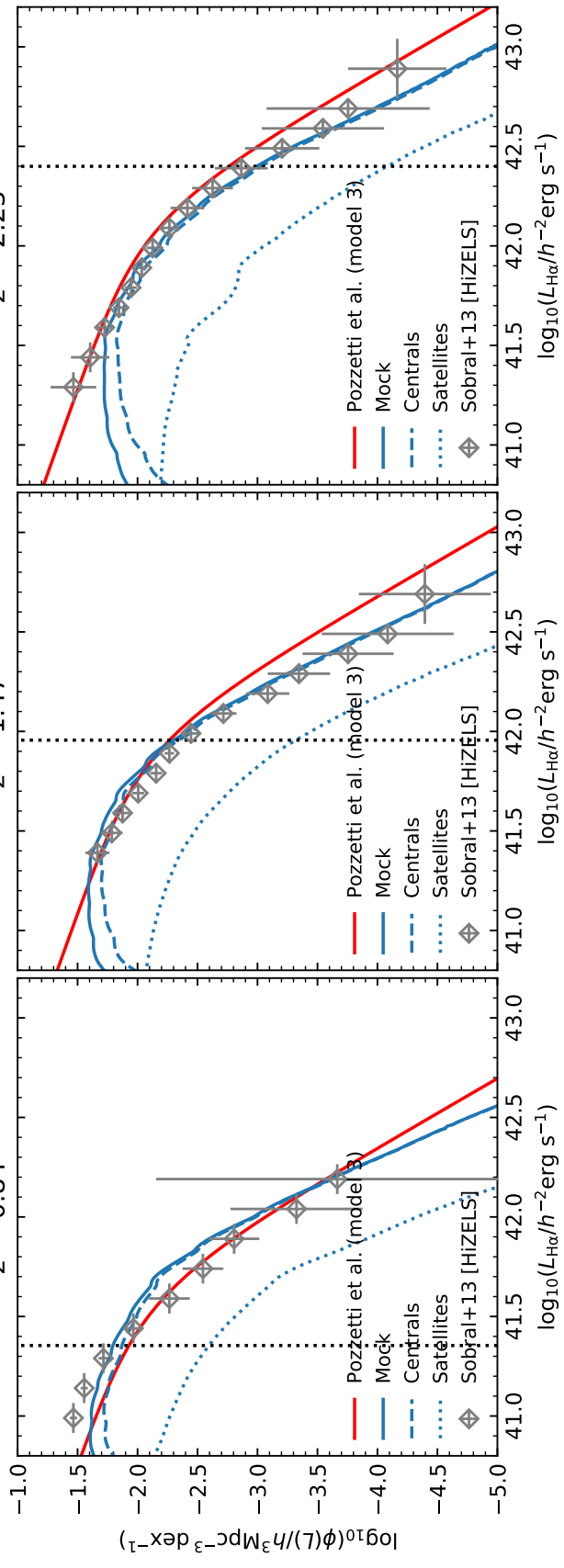

Figure 5.4: Luminosity function of Hα sources measured in three snapshots of the MXXL simulation which have been populated using the Hα HODs predicted by the GALACTICUS semi-analytic model, and dust attenuated using the model of Calzetti et al. (2000). The solid blue curve is for all galaxies, while the dashed and dotted curves are for central and satellites galaxies respectively. Red curves are the luminosity functions using the model of Pozzetti et al. (2016), and grey diamonds with error bars are the luminosity functions measured in the HiZELS survey (Sobral et al., 2013). Vertical dotted lines indicate the flux limit of 3×10^{-16} erg s^{-1}cm^{-2}.

had been measured from the GALACTICUS semi-analytic model by Alex Merson.

Chapter 6

Conclusions

To date, measurements from large galaxy surveys are consistent with the ΛCDM model, in which the dark matter is cold, with negligible thermal velocities at early times, and dark energy is described by the cosmological constant, Λ. While dark energy makes up approximately 70% of the energy density of the Universe today, and is driving the present day accelerated expansion, it is currently poorly understood (e.g. Copeland et al., 2006). Upcoming large galaxy surveys, such as the Dark Energy Spectroscopic Instrument (DESI) (DESI Collaboration et al., 2016a,b) and Euclid surveys (Laureijs et al., 2011), aim to probe the nature of dark energy by creating large 3D maps of the large-scale structure of the Universe. By analysing the clustering of galaxies, baryon acoustic oscillation measurements can be made, which can be used as a standard ruler to measure the expansion history of the Universe. Redshift space distortions can be used to measure the growth of structure, and place constraints on modified theories of gravity (e.g. Guzzo et al., 2008). To prepare for these surveys, it is necessary to utilise realistic mock galaxy catalogues, which can be used, for example, to finalise the survey strategy, develop analysis techniques, and understand systematics that will affect the statistics that are to be measured. These mock catalogues can be built using numerical techniques.

The focus of this thesis is to create mock galaxy catalogues for upcoming large galaxy surveys from large cosmological N-body simulations.

6.1 Dark matter halo merger trees

In Chapter 2, we gave an overview of how halo merger trees can be created from N-body simulations, which is the starting point for creating a mock catalogue. In an N-body simulation, the density field is represented using a set of collisionless particles, and the position and velocity of each particle is evolved depending on the gravitational force of the other particles in the simulation. The particle positions and velocities are output at several epochs, or snapshots, and dark matter haloes are identified at each snapshot using an algorithm such as FOF (Davis et al., 1985) and SUBFIND (Springel et al., 2001a). A merger tree, which describes the merger history of a halo, can then be built by identifying the descendant of each halo at the next snapshot, which can be done by matching particles.

Halo merger trees can also be built using a Monte Carlo algorithm (Cole et al., 2000), which predicts the probability that haloes will merge using extended Press-Schechter theory (Bond et al., 1991). Monte Carlo merger trees do not contain spatial information for each halo, but they can be run many times efficiently, and can be combined with a semi-analytic model of galaxy formation.

These methods can be extended beyond ΛCDM to models of warm dark matter (WDM), where the non-negligible thermal velocities of the dark matter particle at early times results in a cutoff in the power spectrum, $P(k)$, at high k, and a suppression of the formation of low mass haloes. The sterile neutrino is a warm dark matter particle candidate strongly motivated from particle physics that would act as WDM (e.g. Dodelson & Widrow, 1994; Shi & Fuller, 1999; Asaka & Shaposhnikov, 2005). Recent observations of a 3.5 keV feature in the spectra of galaxies and galaxy clusters could potentially be explained as being produced by the decay of a 7 keV sterile neutrino (Bulbul et al., 2014; Boyarsky et al., 2014). In N-body simulations with a cutoff in $P(k)$, spurious low mass haloes are formed, which need to be removed, and the Monte Carlo merger trees are calibrated to reproduce the conditional mass functions of the N-body simulations (Wang & White, 2007;

Benson et al., 2013).

As an application, Lovell et al. (2016) compare the number of satellite galaxies around the Milky Way to the number predicted by the WDM Monte Carlo merger trees, combined with the GALFORM semi-analytic model, to place constraints on the properties of the sterile neutrino and the Milky Way halo mass. For a 7 keV sterile neutrino with lepton asymmetry $L_6 \sim 10$, a minimum Milky Way halo mass of 1.5×10^{12} $\mathrm{M}_\odot$ is needed to produce the number of observed Milky Way satellites, which is consistent with measurements of the mass of the Milky Way halo. However, these results depend on the semi-analytic model used.

6.2 HOD mock catalogue for the DESI Bright Galaxy Survey

Chapter 3 outlined a method for creating a mock catalogue for the DESI Bright Galaxy Survey (BGS) from the Millennium-XXL (MXXL) simulation. The BGS will be a flux limited survey of low redshift galaxies with median redshift $z_{\mathrm{med}} \sim 0.2$, and r-band magnitude limit $r = 20$. By interpolating the positions, velocities and masses of haloes between simulation snapshots, we have constructed a full sky halo lightcone that extends to $z = 2.2$, with a halo mass resolution of $\sim 10^{11}$ $h^{-1}\mathrm{M}_\odot$.

The halo catalogue is then populated with galaxies using a halo occupation distribution (HOD) scheme. The HOD describes the average number of central and satellite galaxies in each halo as a function of mass, and we use a set of HODs measured from the Sloan Digital Sky Survey (SDSS) (York et al., 2000). The standard 5 parameter HOD is modified to prevent unphysical crossing of the HODs, and the HODs are evolved to produce a target luminosity function. By construction, the mock reproduces the luminosity function of SDSS at low redshift, and the evolving luminosity function measured in the Galaxy and Mass Assembly (GAMA) survey (Driver et al., 2009, 2011; Liske et al., 2015) at high redshifts. Galaxies are also assigned a ${}^{0.1}(g-r)$ colour, using a parametrisation of the GAMA

colour-magnitude diagram. The projected correlation functions measured in the mock for galaxies in different magnitude and redshift bins are in good agreement with measurements from SDSS and GAMA, and the mock has colour-dependent clustering. We illustrate that the BAO can be measured in the mock catalogue, and the redshift space distortions are in agreement with measurements from SDSS, making this mock catalogue useful in preparing for the DESI BGS.

6.3 Applying the BGS mock to understand fibre assignment incompleteness

In Chapter 4, we ran the DESI fibre assignment algorithm on the BGS mock catalogue to quantify incompleteness due to fibre assignment, and assess correlation function correction methods. The BGS is currently planned to cover an area of $\sim 14,000$ deg^2 in 3 passes, where each pass covers the survey area in a grid of ~ 2000 pointings of the DESI field of view, or 'tiles', each of area ~ 8 deg^2. Currently, the BGS is proposed to consist of a bright high priority sample to an r-band magnitude limit $r \sim 19.5$, with a fainter low priority sample to $r \sim 20$.

In the focal plane of the telescope, there will be a total of 5,000 fibres, arranged in 10 'petals', each of which is controlled by a robotic fibre positioner that can place the fibre anywhere within a small patrol region. Fibre positioners can block neighbouring fibre positioners from targeting certain objects, and the fixed number density of fibres on the tile results in some of the galaxies in dense regions being missed. This incompleteness has a non-trivial impact on clustering measurements. We show that completeness due to fibre assignment primarily depends on the surface density of galaxies. Completeness is high in low density regions, but in the highest density regions, close to the centre of the most massive clusters, the completeness can be 10%, or lower.

We apply the inverse pair weighting correction of Bianchi & Percival (2017) to clustering measurements from the BGS mock which has been through the fibre

assignment algorithm. By running the fibre assignment algorithm many times, we can calculate the probability that each galaxy pair is targeted. To calculate the correlation function, each pair is weighted by the inverse of this probability. This method is only unbiased if it is possible to observe every galaxy pair. To accurately estimate pair probabilities, and to ensure that as many pairs of galaxies as possible have a non-zero probability of being targeted, we randomly promote a small fraction of the fainter low priority sample to be high priority, and dither the set of tile positions by a small angle. We show that the inverse pair weighting, when combined with angular upweighting, or with regions containing untargetable pairs removed, is able to provide an unbiased correction to the galaxy clustering measurements for a complete survey with 3 passes, and also for a highly incomplete survey with a single pass. With only a single pass, the scatter between realizations on small scales is large, so multiple passes will be needed for accurate small scale clustering measurements. Other commonly used correction methods, such as a nearest neighbour correction, or angular weighting are unable to produce an unbiased correction on all scales.

6.4 Extending the HOD method to create a Euclid mock

In Chapter 5, we extended the HOD method of Chapter 3 to create a mock catalogue of Hα emission line galaxies for the Euclid redshift survey, using a set of HODs measured from the GALACTICUS semi-analytic model. Euclid will cover $15,000$ deg^2 of the sky, and will measure spectra of Hα sources to a flux limit of 3×10^{-16} erg s^{-1}cm^{-2}, covering a redshift range $0.9 \lesssim z \lesssim 1.8$. The HOD parametrisation used for the BGS mock is not applicable for star forming galaxies, and therefore the methods need to be modified for a tabulated HOD measured in bins of mass, Hα luminosity, and redshift.

The HODs are extrapolated to high masses by fitting smooth curves, taking care

to ensure there is no unphysical HOD crossing. Storing the HOD as a 3 dimensional array allows the HOD to be searched and interpolated efficiently. From populating 3 of the MXXL snapshots, we find that the luminosity functions are in agreement with the measured luminosity functions from the HiZELS survey.

6.5 Future Work

The focus of this thesis has been producing mock catalogues for upcoming large galaxy surveys. Here, we outline some of the ways this work can be extended.

- The HOD scheme of Chapter 3 has been used to populate the MXXL lightcone in order to create a BGS mock catalogue. However, this is a single mock catalogue, and, many mock catalogues are required for estimating accurate covariance matrices, of the order of 1,000 (e.g. Blot et al., 2016). Many mock catalogues can be created by combining the HOD scheme with a fast approximate method for creating a halo catalogue, such as the GLAM code (Klypin & Prada, 2018).

- The HODs used to create the BGS mock can also be used to create a mock for the Cosmology Redshift Survey, which is part of 4MOST (de Jong et al., 2016). This survey extends to $z \sim 1$, so this would require extrapolating the HODs to even higher redshifts than was done in Chapter 3.

- The Hα HODs outlined in Chapter 5 have been shown to reproduce the luminosity functions measured in HiZELS. The MXXL halo lightcone can then be populated, and this mock can be used for clustering analysis. This mock can be used to determine if there is any scale dependence in the bias of Hα emitters.

- We have shown in Chapter 4 that the inverse pair weighting is able to provide an unbiased correction to clustering measurements in the BGS, on all scales.

The method can be tested for the case of full sky dithers. More realizations of the fibre assignment algorithm will be needed to estimate accurate pair weights, but the angular weighting will not be needed.

DESI and Euclid will begin to collect data in the next few years, making this an exciting time for cosmology. Utilising mock catalogues, such as those outlined in this thesis, is essential for these surveys to reach their full potential in shedding light on the nature of dark energy.

Appendix A

Databases

The full sky MXXL halo lightcone and BGS mock catalogue outlined in Chapter 2 are made publicly available on the Theoretical Astrophysical Observatory database[2] (Bernyk et al., 2016). The catalogues are also available at `http://icc.dur.ac.uk/data/`.

A.1 MXXL halo catalogue

The halo catalogue contains a total of 5.1 billion haloes out to a redshift of $z = 2.2$, and contains the following halo properties:

- $z_{\rm obs}$, the observed redshift, which takes into account the peculiar velocity of the halo.
- $z_{\rm cos}$, the cosmological redshift, which ignores the effect of the peculiar velocity.
- Right ascension, in degrees.
- Declination, in degrees.
- $M_{\rm 200m}$, the mass enclosed by a sphere in which the average density is 200 times the mean density of the Universe, interpolated to the redshift at which the halo crosses the lightcone, in units of $10^{10}\ h^{-1}{\rm M}_{\odot}$.

[2] `https://tao.asvo.org.au/tao/`

- M_{200c}, the mass enclosed by a sphere in which the average density is 200 times the critical density of the Universe, interpolated to the redshift at which the halo crosses the lightcone, in units of $10^{10}\ h^{-1}\mathrm{M}_{\odot}$.
- V_{max}, the maximum circular velocity, in units of kms^{-1}.
- $R_{V_{\mathrm{max}}}$, the radius at which V_{max} occurs, in $h^{-1}\mathrm{Mpc}$.
- $\sigma_{R_{200\mathrm{m}}}$, velocity dispersion of particles within $R_{200\mathrm{m}}$, in units of kms^{-1}.
- Snapshot number in the MXXL simulation.
- Halo id in the MXXL simulation.

A.2 BGS galaxy catalogue

The full sky galaxy catalogue contains 58.1 million galaxies with $r < 20$, out to redshift $z = 0.8$, and contains the following properties:

- z_{obs}, the observed redshift, which takes into account the peculiar velocity of the galaxy.
- z_{cos}, the cosmological redshift, which ignores the effect of the peculiar velocity.
- Right ascension, in degrees.
- Declination, in degrees.
- $M_{200\mathrm{m}}$ of the host halo, interpolated to the redshift at which the halo crosses the lightcone, in units of $10^{10}\ h^{-1}\mathrm{M}_{\odot}$.
- Apparent r-band magnitude.
- ${}^{0.1}M_r - 5\log h$, the rest frame absolute r-band magnitude, k-corrected to a reference redshift of $z_{\mathrm{ref}} = 0.1$, with no evolutionary correction.
- ${}^{0.1}(g - r)$ colour, k-corrected to a reference redshift $z_{\mathrm{ref}} = 0.1$.

- A flag indicating whether the galaxy is a central or a satellite, and whether it is in a resolved or unresolved halo.
- Snapshot number in the MXXL simulation.
- Halo id in the MXXL simulation.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, The Seventh Data Release of the Sloan Digital Sky Survey, ApJS, 182, 543

Aharonian, F. A., Akamatsu, H., Akimoto, F., et al. 2017, Hitomi Constraints on the 3.5 keV Line in the Perseus Galaxy Cluster, ApJ, 837, L15

Alam, S., Ata, M., Bailey, S., et al. 2017, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample, MNRAS, 470, 2617

Alonso, D. 2012, CUTE solutions for two-point correlation functions from large cosmological datasets, ArXiv e-prints, arXiv:1210.1833

Amendola, L., Appleby, S., Avgoustidis, A., et al. 2018, Cosmology and fundamental physics with the Euclid satellite, Living Reviews in Relativity, 21, 2

Anderson, L., Aubourg, E., Bailey, S., et al. 2012, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample, MNRAS, 427, 3435

Anderson, L., Aubourg, É., Bailey, S., et al. 2014a, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples, MNRAS, 441, 24

Anderson, L., Aubourg, E., Bailey, S., et al. 2014b, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring D_A and H at z = 0.57 from the baryon acoustic peak in the Data Release 9 spectroscopic Galaxy sample, MNRAS, 439, 83

Anderson, M. E., Churazov, E., & Bregman, J. N. 2015, Non-detection of X-ray emission from sterile neutrinos in stacked galaxy spectra, MNRAS, 452, 3905

Angulo, R. E., Baugh, C. M., Frenk, C. S., & Lacey, C. G. 2014, Extending the halo mass resolution of N-body simulations, MNRAS, 442, 3256

Angulo, R. E., Springel, V., White, S. D. M., et al. 2012a, The journey of QSO haloes from z $\sim$ 6 to the present, MNRAS, 425, 2722

—. 2012b, Scaling relations for galaxy clusters in the Millennium-XXL simulation, MNRAS, 426, 2046

Angulo, R. E., & White, S. D. M. 2010, One simulation to fit them all - changing the background parameters of a cosmological N-body simulation, MNRAS, 405, 143

Asaka, T., & Shaposhnikov, M. 2005, The @nMSM, dark matter and baryon asymmetry of the universe [rapid communication], Physics Letters B, 620, 17

Avila, S., Murray, S. G., Knebe, A., et al. 2015, HALOGEN: a tool for fast generation of mock halo catalogues, MNRAS, 450, 1856

Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, Galaxy And Mass Assembly (GAMA): the input catalogue and star-galaxy separation, MNRAS, 404, 86

Barnes, J., & Hut, P. 1986, A hierarchical O(N log N) force-calculation algorithm, Nature, 324, 446

Baugh, C. M. 2006, A primer on hierarchical galaxy formation: the semi-analytical approach, Reports on Progress in Physics, 69, 3101

—. 2008, Creating synthetic universes in a computer, Philosophical Transactions of the Royal Society of London Series A, 366, 4381

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013a, The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores, ApJ, 762, 109

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013b, Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology, ApJ, 763, 18

Benson, A. J. 2010, Galaxy formation theory, Phys. Rep., 495, 33

—. 2012, G ALACTICUS: A semi-analytic model of galaxy formation, New A, 17, 175

Benson, A. J., Cannella, C., & Cole, S. 2016, Achieving convergence in galaxy formation models by augmenting N-body merger trees, Computational Astrophysics and Cosmology, 3, 3

Benson, A. J., Farahi, A., Cole, S., et al. 2013, Dark matter halo merger histories beyond cold dark matter - I. Methods and application to warm dark matter, MNRAS, 428, 1774

Berlind, A. A., & Weinberg, D. H. 2002, The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass, ApJ, 575, 587

Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006, Percolation Galaxy Groups and Clusters in the SDSS Redshift Survey: Identification, Catalogs, and the Multiplicity Function, ApJS, 167, 1

Bernyk, M., Croton, D. J., Tonini, C., et al. 2016, The Theoretical Astrophysical Observatory: Cloud-based Mock Galaxy Catalogs, ApJS, 223, 9

Bianchi, D., & Percival, W. J. 2017, Unbiased clustering estimation in the presence of missing observations, MNRAS, 472, 1106

Bianchi, D., Burden, A., Percival, W. J., et al. 2018, Unbiased clustering estimates with the DESI fibre assignment, MNRAS, 481, 2338

Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1, ApJ, 592, 819

Blot, L., Corasaniti, P. S., Amendola, L., & Kitching, T. D. 2016, Non-linear matter power spectrum covariance matrix errors and cosmological parameter uncertainties, MNRAS, 458, 4462

Bode, P., Ostriker, J. P., & Turok, N. 2001, Halo Formation in Warm Dark Matter Models, ApJ, 556, 93

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, Excursion set mass functions for hierarchical Gaussian fluctuations, ApJ, 379, 440

Bower, R. G. 1991, The evolution of groups of galaxies in the Press-Schechter formalism, MNRAS, 248, 332

Boyarsky, A., Franse, J., Iakubovskyi, D., & Ruchayskiy, O. 2015, Checking the Dark Matter Origin of a 3.53 keV Line with the Milky Way Center, Physical Review Letters, 115, 161301

Boyarsky, A., Ruchayskiy, O., Iakubovskyi, D., & Franse, J. 2014, Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster, Physical Review Letters, 113, 251301

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, Resolving cosmic structure formation with the Millennium-II Simulation, MNRAS, 398, 1150

Bulbul, E., Markevitch, M., Foster, A., et al. 2014, Detection of an Unidentified Emission Line in the Stacked X-Ray Spectrum of Galaxy Clusters, ApJ, 789, 13

Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, Profiles of dark haloes: evolution, scatter and environment, MNRAS, 321, 559

Burden, A., Padmanabhan, N., Cahn, R. N., White, M. J., & Samushia, L. 2017, Mitigating the impact of the DESI fiber assignment on galaxy clustering, J. Cosmology Astropart. Phys., 3, 001

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, The Dust Content and Opacity of Actively Star-forming Galaxies, ApJ, 533, 682

Cappelluti, N., Bulbul, E., Foster, A., et al. 2018, Searching for the 3.5 keV Line in the Deep Fields with Chandra: The 10 Ms Observations, ApJ, 854, 179

Chevallier, M., & Polarski, D. 2001, Accelerating Universes with Scaling Dark Matter, International Journal of Modern Physics D, 10, 213

Chuang, C.-H., Kitaura, F.-S., Prada, F., Zhao, C., & Yepes, G. 2015, EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics, MNRAS, 446, 2621

Cole, S., & Lacey, C. 1996, The structure of dark matter haloes in hierarchical clustering models, MNRAS, 281, 716

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, Hierarchical galaxy formation, MNRAS, 319, 168

Colless, M., Dalton, G., Maddox, S., et al. 2001, The 2dF Galaxy Redshift Survey: spectra and redshifts, MNRAS, 328, 1039

Colless, M., Peterson, B. A., Jackson, C., et al. 2003, The 2dF Galaxy Redshift Survey: Final Data Release, ArXiv Astrophysics e-prints, astro-ph/0306581

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time, ApJ, 647, 201

Contreras, S., Zehavi, I., Baugh, C. M., Padilla, N., & Norberg, P. 2017, The evolution of the galaxy content of dark matter haloes, MNRAS, 465, 2833

Copeland, E. J., Sami, M., & Tsujikawa, S. 2006, Dynamics of Dark Energy, International Journal of Modern Physics D, 15, 1753

Cuesta, A. J., Vargas-Magaña, M., Beutler, F., et al. 2016, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the correlation function of LOWZ and CMASS galaxies in Data Release 12, MNRAS, 457, 1770

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, The Dark Energy Survey: more than dark energy - an overview, MNRAS, 460, 1270

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, The evolution of large-scale structure in a universe dominated by cold dark matter, ApJ, 292, 371

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, The Baryon Oscillation Spectroscopic Survey of SDSS-III, AJ, 145, 10

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data, AJ, 151, 44

de Jong, R. S., Barden, S. C., Bellido-Tirado, O., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99081O

de la Torre, S., & Peacock, J. A. 2013, Reconstructing the distribution of haloes and mock galaxies below the resolution limit in cosmological simulations, MNRAS, 435, 743

DES Collaboration, Abbott, T. M. C., Abdalla, F. B., et al. 2017, Dark Energy Survey Year 1 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing, ArXiv e-prints, arXiv:1708.01530

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016a, The DESI Experiment Part I: Science,Targeting, and Survey Design, ArXiv e-prints, arXiv:1611.00036

—. 2016b, The DESI Experiment Part II: Instrument Design, ArXiv e-prints, arXiv:1611.00037

Diemand, J., Moore, B., & Stadel, J. 2005, Earth-mass dark-matter haloes as the first structures in the early Universe, Nature, 433, 389

Dodelson, S., & Widrow, L. M. 1994, Sterile neutrinos as dark matter, Physical Review Letters, 72, 17

Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, GAMA: towards a physical understanding of galaxy formation, Astronomy and Geophysics, 50, 5.12

Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release, MNRAS, 413, 971

Efstathiou, G., Davis, M., White, S. D. M., & Frenk, C. S. 1985, Numerical techniques for large cosmological N-body simulations, ApJS, 57, 241

Efstathiou, G., Kaiser, N., Saunders, W., et al. 1990a, Largescale Clustering of IRAS Galaxies, MNRAS, 247, 10P

Efstathiou, G., Sutherland, W. J., & Maddox, S. J. 1990b, The cosmological constant and cold dark matter, Nature, 348, 705

Eisenstein, D. J., Blanton, M., Zehavi, I., et al. 2005a, The Small-Scale Clustering of Luminous Red Galaxies via Cross-Correlation Techniques, ApJ, 619, 178

Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N. 2007, Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak, ApJ, 664, 675

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005b, Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, ApJ, 633, 560

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems, AJ, 142, 72

Eke, V. R., Cole, S., & Frenk, C. S. 1996, Cluster evolution as a diagnostic for Omega, MNRAS, 282, astro-ph/9601088

Ellis, J., Hagelin, J. S., Nanopoulos, D. V., Olive, K., & Srednicki, M. 1984, Supersymmetric relics from the big bang, Nuclear Physics B, 238, 453

Farrow, D. J., Cole, S., Norberg, P., et al. 2015, Galaxy and mass assembly (GAMA): projected galaxy clustering, MNRAS, 454, 2120

Figueroa-Feliciano, E., Anderson, A. J., Castro, D., et al. 2015, Searching for keV Sterile Neutrino Dark Matter with X-Ray Microcalorimeter Sounding Rockets, ApJ, 814, 82

Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015, The MICE grand challenge lightcone simulation - I. Dark matter clustering, MNRAS, 448, 2987

Franse, J., Bulbul, E., Foster, A., et al. 2016, Radial Profile of the 3.5 keV Line Out to R200 in the Perseus Cluster, ApJ, 829, 124

Fritsch, F. N., & Carlson, R. E. 1980, Monotone piecewise cubic interpolation, SIAM J. Numer. Anal., 17, 238

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, Introducing the Illustris project: the evolution of galaxy populations across cosmic time, MNRAS, 445, 175

Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., et al. 2014, How sensitive are predicted galaxy luminosities to the choice of stellar population synthesis model?, MNRAS, 439, 264

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, ApJ, 622, 759

Green, J., Schechter, P., Baltay, C., et al. 2012, Wide-Field InfraRed Survey Telescope (WFIRST) Final Report, ArXiv e-prints, arXiv:1208.4012

Guo, H., Zehavi, I., & Zheng, Z. 2012, A New Method to Correct for Fiber Collisions in Galaxy Two-point Statistics, ApJ, 756, 127

Guo, H., Zheng, Z., Zehavi, I., et al. 2015, Redshift-space clustering of SDSS galaxies - luminosity dependence, halo occupation distribution, and velocity bias, MNRAS, 453, 4368

Guzzo, L., Pierleoni, M., Meneux, B., et al. 2008, A test of the nature of cosmic acceleration using galaxy redshift distortions, Nature, 451, 541

Hahn, C., Scoccimarro, R., Blanton, M. R., Tinker, J. L., & Rodríguez-Torres, S. A. 2017, The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles, MNRAS, 467, 1940

Hamilton, A. J. S. 1992, Measuring Omega and the real correlation function from the redshift correlation function, ApJ, 385, L5

Hawkins, E., Maddox, S., Cole, S., et al. 2003, The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe, MNRAS, 346, 78

Hernquist, L., Bouchet, F. R., & Suto, Y. 1991, Application of the Ewald method to cosmological N-body simulations, ApJS, 75, 231

Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2009, Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results, ApJS, 180, 225

Hobbs, A., Read, J. I., Agertz, O., Iannuzzi, F., & Power, C. 2016, NOVel Adaptive softening for collisionless N-body simulations: eliminating spurious haloes, MNRAS, 458, 468

Hockney, R. W., & Eastwood, J. W. 1988, Computer simulation using particles

Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, The K correction, ArXiv e-prints, astro-ph/0210394

Howlett, C., Ross, A. J., Samushia, L., Percival, W. J., & Manera, M. 2015, The clustering of the SDSS main galaxy sample - II. Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at z = 0.15, MNRAS, 449, 848

Hubble, E. 1929, A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae, Proceedings of the National Academy of Science, 15, 168

Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, A survey of galaxy redshifts. IV - The data, ApJS, 52, 89

Jackson, J. C. 1972, A critique of Rees's theory of primordial gravitational radiation, MNRAS, 156, 1P

Jeltema, T., & Profumo, S. 2015, Discovery of a 3.5 keV line in the Galactic Centre and a critical look at the origin of the line across astronomical targets, MNRAS, 450, 2143

—. 2016, Deep XMM observations of Draco rule out at the 99 per cent confidence level a dark matter decay origin for the 3.5 keV line, MNRAS, 458, 3592

Jenkins, A. 2010, Second-order Lagrangian perturbation theory initial conditions for resimulations, MNRAS, 403, 1859

Jenkins, A., Frenk, C. S., White, S. D. M., et al. 2001, The mass function of dark matter haloes, MNRAS, 321, 372

Jiang, L., Helly, J. C., Cole, S., & Frenk, C. S. 2014, N-body dark matter haloes with simple hierarchical histories, MNRAS, 440, 2115

Kaiser, N. 1987, Clustering in real space and in redshift space, MNRAS, 227, 1

Kennedy, R., Frenk, C., Cole, S., & Benson, A. 2014, Constraining the warm dark matter particle mass with Milky Way satellites, MNRAS, 442, 2487

Kitaura, F.-S., Gil-Marín, H., Scóccola, C. G., et al. 2015, Constraining the halo bispectrum in real and redshift space from perturbation theory and non-linear stochastic bias, MNRAS, 450, 1836

Klypin, A., & Prada, F. 2018, Dark matter statistics for large galaxy catalogues: power spectra and covariance matrices, MNRAS, 478, 4602

Knebe, A., Pearce, F. R., Lux, H., et al. 2013, Structure finding in cosmological simulations: the state of affairs, MNRAS, 435, 1618

Knollmann, S. R., & Knebe, A. 2009, AHF: Amiga's Halo Finder, ApJS, 182, 608

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, ApJS, 192, 18

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, The Dark Side of the Halo Occupation Distribution, ApJ, 609, 35

Lacey, C., & Cole, S. 1993, Merger rates in hierarchical models of galaxy formation, MNRAS, 262, 627

Landy, S. D., & Szalay, A. S. 1993, Bias and variance of angular correlation functions, ApJ, 412, 64

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, Euclid Definition Study Report, ArXiv e-prints, arXiv:1110.3193

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

Leo, M., Baugh, C. M., Li, B., & Pascoli, S. 2018, Nonlinear growth of structure in cosmologies with damped matter fluctuations, J. Cosmology Astropart. Phys., 8, 001

Linder, E. V. 2003, Exploring the Expansion History of the Universe, Physical Review Letters, 90, 091301

Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, Galaxy And Mass Assembly (GAMA): end of survey report and data release 2, MNRAS, 452, 2087

Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, Galaxy and Mass Assembly (GAMA): ugriz galaxy luminosity functions, MNRAS, 420, 1239

—. 2015, Galaxy and Mass Assembly (GAMA): maximum-likelihood determination of the luminosity function and its evolution, MNRAS, 451, 1540

Lovell, M. R., Frenk, C. S., Eke, V. R., et al. 2014, The properties of warm dark matter haloes, MNRAS, 439, 300

Lovell, M. R., Eke, V., Frenk, C. S., et al. 2012, The haloes of bright satellite galaxies in a warm dark matter universe, MNRAS, 420, 2318

Lovell, M. R., Bose, S., Boyarsky, A., et al. 2016, Satellite galaxies in semi-analytic models of galaxy formation with sterile neutrino dark matter, MNRAS, 461, 60

Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues, MNRAS, 428, 1036

Mather, J. C., Cheng, E. S., Eplee, Jr., R. E., et al. 1990, A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite, ApJ, 354, L37

McNaught-Roberts, T., Norberg, P., Baugh, C., et al. 2014, Galaxy And Mass Assembly (GAMA): the dependence of the galaxy luminosity function on environment, redshift and colour, MNRAS, 445, 2125

Merson, A., Wang, Y., Benson, A., et al. 2018, Predicting Hα emission-line galaxy counts for future galaxy redshift surveys, MNRAS, 474, 177

Merson, A. I., Baugh, C. M., Helly, J. C., et al. 2013, Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection, MNRAS, 429, 556

MiniBooNE Collaboration, Aguilar-Arevalo, A. A., Brown, B. C., et al. 2018, Observation of a Significant Excess of Electron-Like Events in the MiniBooNE Short-Baseline Neutrino Experiment, ArXiv e-prints, arXiv:1805.12028

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Mohammad, F. G., Bianchi, D., Percival, W. J., et al. 2018, The VIMOS Public Extragalactic Redshift Survey (VIPERS). Unbiased clustering estimate with VIPERS slit assignment, A&A, 619, A17

Monaco, P., Sefusatti, E., Borgani, S., et al. 2013, An accurate tool for the fast generation of dark matter halo catalogues, MNRAS, 433, 2389

Moore, B., Ghigna, S., Governato, F., et al. 1999, Dark Matter Substructure within Galactic Halos, ApJ, 524, L19

Moster, B. P., Naab, T., & White, S. D. M. 2013, Galactic star formation and accretion histories from matching galaxies to dark matter haloes, MNRAS, 428, 3121

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, A Universal Density Profile from Hierarchical Clustering, ApJ, 490, 493

Neronov, A., Malyshev, D., & Eckert, D. 2016, Decaying dark matter search with NuSTAR deep sky observations, Phys. Rev. D, 94, 123504

Parkinson, H., Cole, S., & Helly, J. 2008, Generating dark matter halo merger trees, MNRAS, 383, 557

Peacock, J. A., & Smith, R. E. 2000, Halo occupation numbers and galaxy bias, MNRAS, 318, 1144

Peacock, J. A., Cole, S., Norberg, P., et al. 2001, A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey, Nature, 410, 169

Percival, W. J., & Bianchi, D. 2017, Using angular pair upweighting to improve 3D clustering measurements, MNRAS, 472, L40

Percival, W. J., Ross, A. J., Sánchez, A. G., et al. 2014, The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors, MNRAS, 439, 2531

Perez, K., Ng, K. C. Y., Beacom, J. F., et al. 2017, Almost closing the ν MSM sterile neutrino dark matter window with NuSTAR, Phys. Rev. D, 95, 123002

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, Measurements of Ω and Λ from 42 High-Redshift Supernovae, ApJ, 517, 565

Pinol, L., Cahn, R. N., Hand, N., Seljak, U., & White, M. 2017, Imprint of DESI fiber assignment on the anisotropic power spectrum of emission line galaxies, J. Cosmology Astropart. Phys., 4, 008

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, Planck 2018 results. VI. Cosmological parameters, ArXiv e-prints, arXiv:1807.06209

Polarski, D., & Gannouji, R. 2008, On the growth of linear perturbations, Physics Letters B, 660, 439

Pozzetti, L., Hirata, C. M., Geach, J. E., et al. 2016, Modelling the number density of Hα emitters for future spectroscopic near-IR space missions, A&A, 590, A3

Press, W. H., & Schechter, P. 1974, Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation, ApJ, 187, 425

Reid, B., Ho, S., Padmanabhan, N., et al. 2016, SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues, MNRAS, 455, 1553

Riemer-Sørensen, S., Wik, D., Madejski, G., et al. 2015, Dark Matter Line Emission Constraints from NuSTAR Observations of the Bullet Cluster, ApJ, 810, 48

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, AJ, 116, 1009

Riess, A. G., Casertano, S., Yuan, W., et al. 2018, Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant, ApJ, 861, 126

Robotham, A., Driver, S. P., Norberg, P., et al. 2010, Galaxy and Mass Assembly (GAMA): Optimal Tiling of Dense Surveys with a Multi-Object Spectrograph, Publ. Astron. Soc. Australia, 27, 76

Ross, A. J., Samushia, L., Howlett, C., et al. 2015, The clustering of the SDSS DR7 main Galaxy sample - I. A 4 per cent distance measure at z = 0.15, MNRAS, 449, 835

Ross, A. J., Beutler, F., Chuang, C.-H., et al. 2017, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: observational systematics and baryon acoustic oscillations in the correlation function, MNRAS, 464, 1168

Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1980, Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/, ApJ, 238, 471

Ruggeri, R., Percival, W. J., Gil Marin, H., et al. 2018, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measuring the evolution of the growth rate using redshift space distortions between redshift 0.8 and 2.2, ArXiv e-prints, arXiv:1801.02891

Samushia, L., Percival, W. J., & Raccanelli, A. 2012, Interpreting large-scale redshift-space distortion measurements, MNRAS, 420, 2102

Sawala, T., Frenk, C. S., Fattahi, A., et al. 2016, The APOSTLE simulations: solutions to the Local Group's cosmic puzzles, MNRAS, 457, 1931

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, The EAGLE project: simulating the evolution and assembly of galaxies and their environments, MNRAS, 446, 521

Schneider, A., Smith, R. E., & Reed, D. 2013, Halo mass function and the free streaming scale, MNRAS, 433, 1573

Schoenberg, I. J. 1946, Contributions to the problem of approximation of equidistant data by analytic functions. A. On the problem of smoothing or graduation - a 1st class of analytic approximation formula, Q. Appl. Math., 4, 45

Schubnell, M., Ameel, J., Besuner, R. W., et al. 2016, in Proc. SPIE, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 990892

Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering, ApJ, 546, 20

Seljak, U. 2000, Analytic model for galaxy and dark matter clustering, MNRAS, 318, 203

Seo, H.-J., & Eisenstein, D. J. 2003, Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys, ApJ, 598, 720

Shao, S., Gao, L., Theuns, T., & Frenk, C. S. 2013, The phase-space density of fermionic dark matter haloes, MNRAS, 430, 2346

Sheth, R. K., & Tormen, G. 1999, Large-scale bias and the peak background split, MNRAS, 308, 119

Shi, X., & Fuller, G. M. 1999, New Dark Matter Candidate: Nonthermal Sterile Neutrinos, Physical Review Letters, 82, 2832

Skibba, R., Sheth, R. K., Connolly, A. J., & Scranton, R. 2006, The luminosity-weighted or 'marked' correlation function, MNRAS, 369, 68

Skibba, R. A., & Sheth, R. K. 2009, A halo model of galaxy colours and clustering in the Sloan Digital Sky Survey, MNRAS, 392, 1080

Sobral, D., Smail, I., Best, P. N., et al. 2013, A large Hα survey at z = 2.23, 1.47, 0.84 and 0.40: the 11 Gyr evolution of star-forming galaxies from HiZELS, MNRAS, 428, 1128

Somerville, R. S., & Davé, R. 2015, Physical Models of Galaxy Formation in a Cosmological Framework, ARA&A, 53, 51

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, ArXiv e-prints, arXiv:1503.03757

Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters, ApJS, 148, 175

Springel, V. 2005, The cosmological simulation code GADGET-2, MNRAS, 364, 1105

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001a, Populating a cluster of galaxies - I. Results at [formmu2]z=0, MNRAS, 328, 726

Springel, V., Yoshida, N., & White, S. D. M. 2001b, GADGET: a code for collisionless and gasdynamical cosmological simulations, New Astron., 6, 79

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Simulations of the formation, evolution and clustering of galaxies and quasars, Nature, 435, 629

Springel, V., Wang, J., Vogelsberger, M., et al. 2008, The Aquarius Project: the subhaloes of galactic haloes, MNRAS, 391, 1685

Stothert, L. 2018, PhD thesis, University of Durham

Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample, AJ, 124, 1810

Tassev, S., Zaldarriaga, M., & Eisenstein, D. J. 2013, Solving large scale structure in ten easy steps with COLA, J. Cosmology Astropart. Phys., 6, 036

The Dark Energy Survey Collaboration. 2005, The Dark Energy Survey, ArXiv Astrophysics e-prints, astro-ph/0510346

Urban, O., Werner, N., Allen, S. W., et al. 2015, A Suzaku search for dark matter emission lines in the X-ray brightest galaxy clusters, MNRAS, 451, 2447

Vale, A., & Ostriker, J. P. 2004, Linking halo mass to galaxy luminosity, MNRAS, 353, 189

Viel, M., Lesgourgues, J., Haehnelt, M. G., Matarrese, S., & Riotto, A. 2005, Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-α forest, Phys. Rev. D, 71, 063534

Wang, J., & White, S. D. M. 2007, Discreteness effects in simulations of hot/warm dark matter, MNRAS, 380, 93

Wang, W., Han, J., Cooper, A. P., et al. 2015, Estimating the dark matter halo mass of our Milky Way using dynamical tracers, MNRAS, 453, 377

Wang, Y., Brunner, R. J., & Dolence, J. C. 2013, The SDSS galaxy angular two-point correlation function, MNRAS, 432, 1961

White, M., Tinker, J. L., & McBride, C. K. 2014, Mock galaxy catalogues using the quick particle mesh method, MNRAS, 437, 2594

White, S. D. M. 1994, Formation and Evolution of Galaxies: Les Houches Lectures, ArXiv Astrophysics e-prints, astro-ph/9410043

Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies, MNRAS, 339, 1057

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, The Sloan Digital Sky Survey: Technical Summary, AJ, 120, 1579

Zarrouk, P., Burtin, E., Gil-Marín, H., et al. 2018, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2, MNRAS, 477, 1639

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, The Luminosity and Color Dependence of the Galaxy Correlation Function, ApJ, 630, 1

—. 2011, Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity, ApJ, 736, 59

Zhao, G.-B., Wang, Y., Saito, S., et al. 2018, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: a tomographic measurement of cosmic structure growth and expansion rate based on optimal redshift weights, MNRAS, arXiv:1801.03043

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies, ApJ, 633, 791

Zwicky, F. 1933, Die Rotverschiebung von extragalaktischen Nebeln, Helvetica Physica Acta, 6, 110