Gravity to Galaxies — N-body Simulations for the DESI Survey

CAMERON GROVE

Gravity to Galaxies — N-body Simulations for the DESI Survey

Cameron Grove

A thesis presented for the degree of
Doctor of Philosophy

Department of Physics
Durham University
United Kingdom
December 2022

Gravity to Galaxies — N-body Simulations for the DESI Survey

Cameron Grove

Abstract

Simulations of the evolution of the universe under gravity are essential to our understanding of modern cosmology. These N-body simulations can contain trillions of particles and are run on some of the most powerful supercomputers in the world. N-body simulations are necessary to understand the data from large galaxy surveys that are mapping the universe in higher detail than ever before. This thesis explores the accuracy of N-body simulation methods and their applications towards the Dark Energy Spectroscopic Instrument (DESI) Survey. Firstly, the accuracy of the gravity scheme in the Smoothed Particle Hydrodynamics With Inter-dependent Fine-grained Tasking (SWIFT) simulation code is tested. These tests inform us on the limitations of running SWIFT with high values of the opening angle parameter θ. Additionally, an error in the large-scale clustering in SWIFT simulations was found and fixed. Next, a comparison between several simulation codes is presented. In this comparison, simulations were run from identical initial conditions and the level of discrepancy caused by choice of simulation code is measured. The systematic errors caused by choice of code are compared to the statistical errors in a DESI Survey volume. We find that the matter power spectra from independent codes agree to within 1% for $k < 10h\text{Mpc}^{-1}$. The halo mass functions agree to within 1% between mass limits of $10^{11.5} - 10^{14}\ h^{-1}\text{M}_{\odot}$. Halo clustering measurements are within the DESI year 5 uncertainty at scales greater than $20h^{-1}$Mpc. The results justified the choice of resolution adopted in the AbacusSummit suite of simulations that are used to create mock galaxy catalogues for DESI. To connect simulated dark matter halos to observed galaxies, we use a halo occupation distribution (HOD). We have produced a novel HOD fitting procedure that simultaneously fits HOD curves describing samples with varying absolute magnitude limits. This allows us to produce mock catalogues that reproduce a target luminosity- and colour-dependent galaxy clustering correlation function, and a target galaxy luminosity function. Finally, mock galaxy catalogues are created from the AbacusSummit suite of simulations for the DESI Bright Galaxy Survey (BGS). These mock catalogues will aid with measuring systematic effects in DESI BGS data and testing the unbiased recovery of cosmological parameters.

Contents

List of Figures

List of Tables

Declaration

The work in this thesis is based on research carried out by the author at the Institute for Computational Cosmology, Department of Physics, Durham University, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

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

- Grove, C. et al, "The DESI N-body Simulation Project I: Testing the Robustness of Simulations for the DESI Dark Time Survey", 2022, Monthly Notices of the Royal Astronomical Society, 515, pp. 1854-1870

The work presented in Chapter 5 was a collaboration between the author and Alex Smith. The author contributed all work on HOD parameter fits, produced clustering measurements of the Bright Galaxy Survey, and ran the code to create the mock galaxy catalogues.

The anomaly detection work in Section 6.1 was part of a joint project with Edward Elliott, all contributions were shared equally.

The work on pupil attainment modelling in Section 6.2 was performed in the capacity of an independent researcher and is not reflective of the Department for Education's

official position on these matters.

Acknowledgements

There are many people to whom I owe thanks for help and support over the course of my studies. Firstly, I'd like to thank my supervisors Shaun & Peder who have guided me through my research, provided assistance whenever I needed it, and pointed me in the right direction.

I'd like to thank those that I have collaborated with during my time in Durham. Alex Smith for helping me to get my head around HODs and mock catalogues, Matthieu Schaller for patience with my many SWIFT bug reports, Lehman Garrison & Chia-Hsun Chuang for their organisation of the Cosmosim group and the code comparison project, and John Helly for his software expertise.

Thanks to my college tutors at St Chad's, Paul & Sarah Chandler, who have hosted many enjoyable formals and dinner parties, keeping me involved in college life.

I am lucky to have been granted many friendships while in Durham. Thanks to Ed, Jack, Henry, and Vicky for being great housemates, pub quiz teammates, and Mario Kart opponents. Thanks to those in the department for stimulating conversations over coffee, lunch, and beer: Aidan, Jake, Ellen, Amy, Joaquin, Victor, Duncan, and many more. I'd also like to thank my friends from Cambridge, Bryn, Ben, Harry, Pier, Rob, and Shannon along with the "gerds", Abie, Adam, Charles, Fran, Kate, Miranda, and Sam for providing many memories through the years and hopefully many more in the future.

I am grateful for my parents, Simon and Rachel, who have always supported me and fostered an environment for me to learn and grow. Without them, I wouldn't be the person I am today.

Finally, Hollie, for your endless love, and willingness to spend countless hours on the East Coast Main Line to come and visit me. Thank you for your patience while I've spent

so long in the North East, you've always being there for me whenever I've needed you. I'd also like to thank you for introducing me to our fluffy pals, Vinnie, Moley, Daisy, & Nugget.

CHAPTER 1

Introduction

1.1 ΛCDM Cosmology

ΛCDM is the current standard model of cosmology. In this model the Universe started from a "Big Bang", a very hot and dense initial state, around 13.8 billion years ago. Density fluctuations were caused by quantum fluctuations in the early Universe. The Universe then expanded and cooled, the initial density fluctuations evolved to form the large-scale structure that we can observe today. In short, this model states that the universe consists of three main components: ordinary matter & radiation, dark matter, and dark energy.

Ordinary matter consists of matter that we can observe due to its interaction with light. The ease of observing objects such as stars and bright galaxies means that the study of ordinary matter in the universe extends back millennia. However, the study of the other components of the ΛCDM model have more recent origins.

1.1.1 Cold Dark Matter

Theories of the existence of dark matter began to emerge around 100 years ago. Dark matter was hypothesised to exist because of discrepancies between mass estimates from

galaxy dynamics compared to mass estimates from their brightness. There were two key observations that both indicate the presence of large quantities of non-luminous matter. Firstly, the velocity dispersion of galaxies in the Coma cluster implied that the mass of the cluster must be higher than can be explained by the observed matter for the cluster to be stable (Zwicky, 1933). Secondly, observations of flat galaxy rotation curves provided evidence for a greater amount of matter within galaxies than was observed (Rubin & W. Kent, 1970; Rogstad & Shostak, 1972). A seminal paper in the development of the theory of dark matter was Ostriker et al. (1974) which opens with "There are reasons, increasing in number and quality, to believe that the masses of ordinary galaxies may have been underestimated by a factor of 10 or more."

The theory of cold dark matter is that dark matter is made up of particles with an initial velocity that is low enough to be unimportant on cosmological scales, as opposed to hot dark matter which travels at relativistic velocities. The "warmth" of dark matter alters the scales at which overdensities initially form and therefore affects whether large-scale structure forms through a top-down approach or a bottom-up approach. Hot dark matter suppresses the growth of structure at small scales, implying that galaxies form from the collapse of larger structures. Cold dark matter allows formation of structure at small scales which means that galaxy clusters are formed from mergers between smaller clumps in a bottom-up approach (Peebles, 1982a). Hot dark matter has been ruled out (White et al., 1983). Warm dark matter, with a particle mass of order keV, remains a viable candidate (Viel et al., 2005).

1.1.2 The Cosmological Constant

Λ, the cosmological constant, or "dark energy" is related to the expansion of the universe. To understand the origin of this component, we must discuss general relativity and the Friedmann equations.

Einstein introduced the cosmological constant into the field equations of general relativity in order to preserve a static universe (Einstein, 1917), however its presence is also consistent with dynamic models of the universe.

A solution to the Einstein field equations in an isotropic and homogeneous universe produces the Friedmann equations (Friedmann, 1922)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2}, \tag{1.1}$$

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

where a is the scale factor of the universe, G is the gravitational constant, ρ is the density of the universe, p is the pressure of the universe, Λ is the aforementioned cosmological constant, c is the speed of light in a vacuum, and k/a^2 is the curvature of the universe.

The Hubble constant, H, is defined as $\dot{a}/a$, and the value at current times is defined as H_0. The Hubble constant describes the rate of expansion of the universe. The critical density of the universe, ρ_c, is defined as the density required for the universe to be flat. In the absence of dark energy, rearranging Equation 1.1 with $k = 0$ produces the definition $\rho_c = 3H^2/8\pi G$. The components of the universe can be rewritten as a ratio with respect to the critical density to produce dimensionless density parameters. For mass,

$$\Omega_m = \frac{\rho_m}{\rho_c}, \tag{1.3}$$

for radiation,

$$\Omega_r = \frac{\rho_r}{\rho_c}, \tag{1.4}$$

for Λ,

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2}. \tag{1.5}$$

and for curvature,

$$\Omega_k = -\frac{kc^2}{a_0^2 H_0^2}. \tag{1.6}$$

Equation 1.1 implies that the density parameters are constrained to add to unity. $\Omega_k = 0$ describes a flat universe, meanwhile positive and negative values represent open and closed universes respectively. The constraining of these density parameters has been a key aim for modern cosmology. Their relative sizes effectively tell us the contribution to the total energy density of the universe from each component.

For much of the 20th Century, it was assumed that $\Omega_\Lambda = 0$. However, growing evidence suggested that a non-zero value of Ω_Λ existed (Peebles, 1984; Efstathiou et al.,

1990). Evidence from Type 1A supernovae confirmed that the rate of expansion of the universe was accelerating and therefore the current scientific consensus is that Ω_Λ is non-zero (Riess et al., 1998; Perlmutter et al., 1999). Hence, the inclusion of Λ in the name of the ΛCDM model.

1.2 Evidence for the ΛCDM Model

There is ample observational evidence that is consistent with (and in some cases, initially predicted by) the ΛCDM model.

1.2.1 The Cosmic Microwave Background Radiation

The cosmic microwave background radiation (CMB) probes the early universe and contains unique evidence of inflation and the ΛCDM model.

In the early universe, shortly after the Big Bang, the universe was very hot and dense. The universe was hot enough to be fully ionised, coupling photons and baryons, and opaque to radiation. By $z \approx 1100$ the universe had cooled sufficiently for atoms to form (recombination), and photons could now travel uninterrupted through space. This radiation is a black-body spectrum, originally at a temperature of ≈3000K with a substantial component in the visible spectrum, but redshifted to a current temperature of ≈2.7K, with wavelengths corresponding to the microwave range of the electromagnetic spectrum.

The CMB was first detected in the 1960s (Penzias & Wilson, 1965). Since then, it has been an object of intense study with satellites such as COBE (Smoot et al., 1992), WMAP (Bennett et al., 2013), and Planck (Planck Collaboration, 2020) dedicated to precisely measuring small variations in the CMB.

Variations in the CMB as a function of angular separation originate from quantum fluctuations in the early universe. These fluctuations are coherent across such large scales that they could not be causally connected without the universe undergoing a period of rapid inflation. The theory of inflation provides a unique causal mechanism to explain these perturbations (Guth, 1981; Liddle, 1995). Acoustic oscillations in the photon-baryon fluid post-inflation but before recombination were theorised to leave a characteristic signature in the CMB (Peebles, 1982b, 1987; Hu & White, 1996). The characteristic lengthscale of

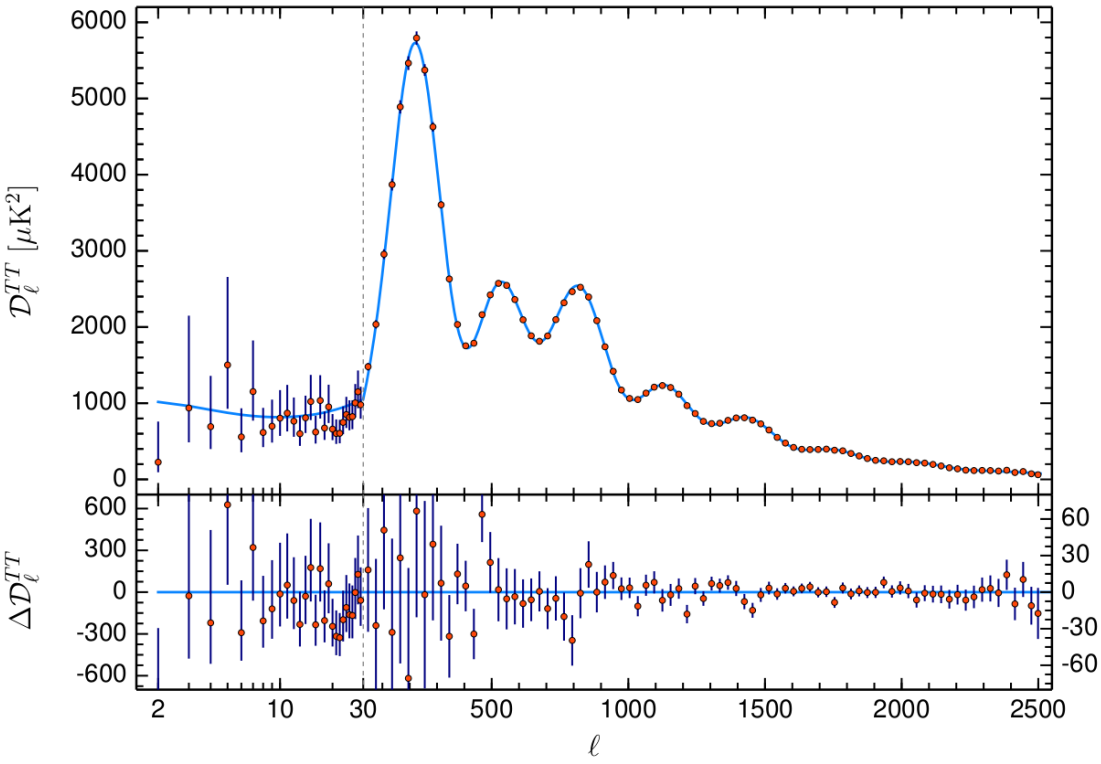

Figure 1.1: The temperature power spectrum of the cosmic microwave background radiation measured from the Planck satellite. The x-axis represents multipole number and provides an indication of angular scale. The blue line shows the computed power spectrum from ΛCDM cosmology that closely matches the observed points in red. Reproduced from Planck Collaboration (2020).

these oscillations was determined by the speed of the waves through the plasma medium, along with the time that the oscillations had to travel before recombination froze them into place. The CMB temperature power spectrum has peaks that precisely match those predicted using a ΛCDM model, as seen in Figure 1.1.

1.2.2 Baryon Acoustic Oscillations in Matter

It is not only the CMB where acoustic oscillations have left their mark. When photons and baryons decoupled at recombination the oscillation signature was left in matter too, however with a lower amplitude than in the CMB. This baryon acoustic oscillation (BAO) produced an overdensity at scales around 100 h^{-1}Mpc that influenced the growth of structure in the universe. Cold dark matter cosmology can make predictions for the strength and shape of the BAO peak (Eisenstein & Hu, 1998).

Measurements of galaxy clustering from large galaxy surveys have shown evidence for the presence of a BAO peak consistent with predictions from the ΛCDM model (Eisenstein et al., 2005; Cole et al., 2005).

Because the size of the BAO feature was fixed post-recombination, it acts as a "standard ruler" in cosmology. The scale of the BAO feature through cosmic time provides a measurement of the rate of expansion of the Universe through time and therefore precise measurements of the BAO peak at different redshifts are an aim of modern cosmology. Current constraints on BAO come from the Baryon Oscillation Spectroscopic Survey (BOSS/eBOSS) (Alam et al., 2017; Bautista et al., 2020; Raichoor et al., 2020; Hou et al., 2020).

1.2.3 Challenges to the ΛCDM Model

We have seen that the ΛCDM model is an extraordinary success, consistent with many current observations of the Universe from experiments in a wide range of physical contexts. Modern cosmology is attempting to investigate regimes where differences to ΛCDM may be detected.

A tension has emerged in measurements of the Hubble constant, H_0, between ΛCDM predictions based on CMB measurements and direct measurements of distances to galaxies

in the late-time universe (Valentino et al., 2021). The Planck value for H_0 from the CMB is $H_0 = (67.4 \pm 0.5)\,\mathrm{kms^{-1}Mpc^{-1}}$, meanwhile measurements from the local universe provide a value of $H_0 = (73.30 \pm 1.04)\,\mathrm{kms^{-1}Mpc^{-1}}$, a 5σ difference (Riess et al., 2022).

A related tension can be found in the cosmological parameter σ_8 which quantifies the strength of matter density fluctuations at a fiducial scale of 8 h^{-1}Mpc. Evidence from the CMB predicts a value of σ_8 that is higher than measurements from large-scale structure and gravitational lensing (e.g. Abdalla et al., 2022; Heymans et al., 2021).

There are also challenges to the ΛCDM model at small scales. Numerical simulations of cold dark matter predict that dark matter halos have a "cuspy" profile where density increases towards the centre of the halo. Observations of dwarf galaxies suggest that some halos contain a core of constant density (e.g. Alam et al., 2002; Oman et al., 2015). There is also the “too big to fail” problem where satellites of the Milky Way predicted by ΛCDM should be large enough to contain visible stars. Potential solutions to these problems may involve the inclusion of baryonic feedback or warm dark matter to alter small-scale structure formation (e.g. Bullock & Boylan-Kolchin, 2017).

Despite these challenges, ΛCDM remains the favoured model for structure formation in the Universe.

1.3 Clustering & the Power Spectrum

There are some important statistical quantities that can be measured from both matter tracers in the universe and in simulations that are relevant to this thesis. In this section I shall define these quantities.

The two-point correlation function is the excess probability of finding an object (usually a galaxy) at a distance r from another object, when compared to a randomly distributed sample,

$$P(r)\mathrm{d}r = P_{\mathrm{random}}(r)(1 + \xi(r))\mathrm{d}r, \tag{1.7}$$

where $P(r)$ is the probability of finding two objects separated by a distance between r and $r + \mathrm{d}r$, $P_{\mathrm{random}}(r)$ is the probability if the objects were randomly distributed, and $\xi(r)$ is the two-point correlation function.

A measured correlation function can be calculated from paircounts of galaxies and the paircounts of random particles with the same average density and volume geometry,

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1, \tag{1.8}$$

where $DD(r)$ is the number of galaxy-galaxy pairs of separation r, and $RR(r)$ is the number of random-random pairs. An improved estimator for the correlation function comes from (Landy & Szalay, 1993),

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}, \tag{1.9}$$

where galaxy-random paircounts, $DR(r)$ are also included.

The above formulae have treated r as a scalar quantity, which is appropriate when the distribution of objects is isotropic, for example the real space distribution of galaxies. However, there are also uses for non-isotropic correlation functions.

Peculiar galaxy velocities are movements relative to the Hubble flow, the expansion of the universe. Peculiar velocities parallel to the line-of-sight direction mean that redshifts cannot be exactly mapped to distances. The Doppler effect red or blueshifts light from galaxies such that their relative positions in redshift space differ from positions real space. This has the effect of decreasing the size of large density perturbations in the line-of-sight direction, because of infall of galaxies towards the overdensity. In the centre of galaxy clusters, high galaxy velocities stretch redshift space positions in the line-of-sight direction (Kaiser, 1987).

The two-dimensional redshift space correlation function splits the separation vector into two components. σ is the component perpendicular to the line-of-sight and π is the component parallel to the line-of-sight direction. Figure 1.2 shows measurements of the two-dimensional redshift space correlation function, $\xi(\sigma, \pi)$ from 2dFGRS which contains the distinctive "finger of God" (Madgwick et al., 2003).

To recover a one-dimensional correlation function from $\xi(\sigma, \pi)$ one can integrate over the π direction to produce the projected correlation function,

$$w(\sigma) = \frac{1}{H_0} \int_{-\pi_{\max}}^{\pi_{\max}} \mathrm{d}\pi \xi(\sigma, \pi), \tag{1.10}$$

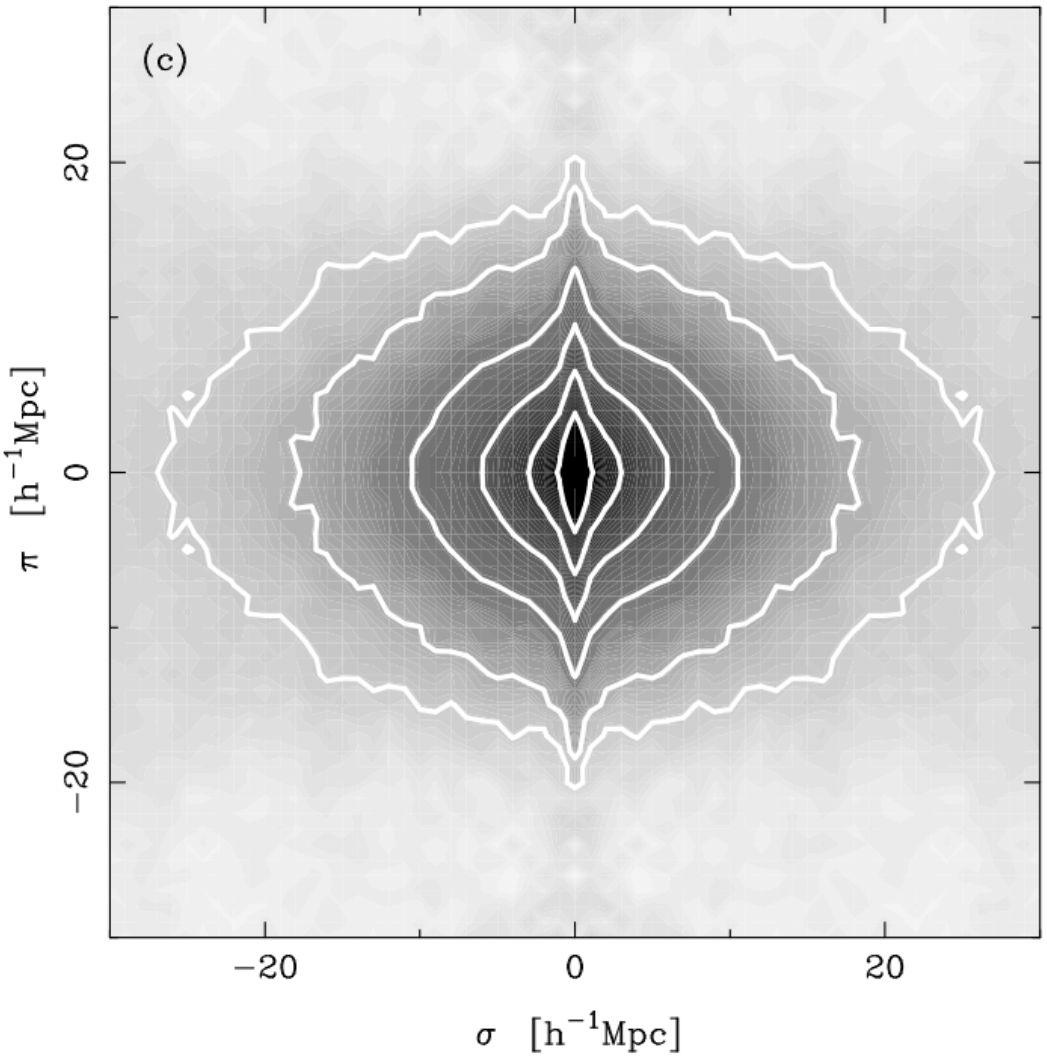

Figure 1.2: $\xi(\sigma,\pi)$ from 2dFGRS showing the presence of the "Finger of God" along the line-of-sight direction. Solid lines indicate contours of constant $\xi(\sigma,\pi)$. Reproduced from Madgwick et al. (2003).

where $\pi_{\rm max}$ defines the limits of the integration. If $\pi_{\rm max}$ is large enough to include almost all correlated pairs, and a far-observer approximation is made, then the projected correlation function can be related to the real-space correlation function (Davis & Peebles, 1983),

$$w(\sigma) = 2 \int_0^\infty \xi \left[\left(\sigma^2 + \pi^2 \right)^{1/2} \right] \mathrm{d}\pi, \tag{1.11}$$

$$w(\sigma) = 2 \int_\sigma^\infty \xi(r) \left(r^2 - \sigma^2 \right)^{-1/2} \mathrm{d}r. \tag{1.12}$$

It can also be useful to split $\xi(\sigma, \pi)$ into a multipole expansion, as a function of separation $s = \left(\sigma^2 + \pi^2 \right)^{1/2}$, and the cosine of the angle to the line of sight, μ,

$$\xi(s, \mu) = \xi_0(s) P_0(\mu) + \xi_2(s) P_2(\mu) + \xi_4(s) P_4(\mu) + ..., \tag{1.13}$$

where $\xi_0(s)$, $\xi_2(s)$, and $\xi_4(s)$ are the monopole, quadrupole, and hexadecapole respectively, and $P_i(\mu)$ is the i^{th} Legendre polynomial (Kaiser, 1987; Hamilton, 1992). There are no contributions from the odd terms because of the symmetry under the transformation $\pi \Longrightarrow -\pi$.

It is possible to extend the correlation function to three-point and higer-order statistics (Takada & Jain, 2003), however these are not explored in this thesis.

The power spectrum can be defined as the Fourier transform of the correlation function,

$$P(\mathbf{k}) = \int d^3\mathbf{r}\, e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}). \tag{1.14}$$

The measured power spectrum from a galaxy survey is the true spectrum that has been convolved with a window function that depends on the geometry of the survey. This window function must be deconvolved to produce an estimate for the true power spectrum.

The 2d anisotropic power spectrum, and its multipole expansion can be defined in a similar way to the correlation function analogues.

The correlation functions/power spectra of matter and galaxies are not the same. Therefore galaxies can be described as a *biased tracer* of matter. Understanding and parameterising the form of galaxy bias is important to relate observations of galaxy clustering to simulations (Desjacques et al., 2018). The exact form of galaxy bias is an

area of active research. At large scales, galaxy bias is close to scale-independent. The shape of the correlation function will not change between matter and galaxies for scale-independent bias. At small scales, galaxy bias is scale-dependent and depends on redshift (Weinberg et al., 2004). In addition, galaxy bias is a function of galaxy luminosity and colour (Cresswell & Percival, 2009).

1.4 Galaxy Surveys

Observations of large-scale clustering statistics have been driven by the development of galaxy surveys. Positions of galaxies are measured on the sky and the distances to galaxies are calculated using redshifts. This means that we can construct 3D maps of galaxy locations locally and back through cosmic time. Advances in engineering and technology have allowed for the positions of more galaxies to be measured more accurately and rapidly. This has ushered in the era of high-precision cosmology, where sample sizes are sufficient to suppress sample variance to low levels allowing astrophysicists to probe models in a wider range of scales. Sample variance is the noise in measurements that comes from measuring a sub-set of the total volume in the Universe. Sampling larger volumes and to fainter limits gives us more confidence that the results we measure are representative of the Universe as a whole and not a sub-region. Sampling to fainter limits in a fixed volume increases the total number of objects in the survey volume which reduces statistical noise.

The first galaxy surveys were simple catalogues of objects of interest to astronomers of the time. These included the catalogues compiled by Charles Messier and John Herschel in the 18th and 19th centuries. The small number of objects in these catalogues, and the lack of an ability to measure distance to the objects made them unsuitable for clustering analyses.

An early galaxy survey that systematically measured redshifts was the CfA Redshift Survey. This survey measured the redshifts of 2400 galaxies spectroscopically and provided evidence that galaxies were clustered in space, not evenly distributed (de Lapparent et al., 1986). A slice from the survey is shown in Figure 1.3.

Since the 1980s, galaxy surveys have progressed in scope and scale. Modern galaxy surveys can collect data on millions of galaxies, numbers that were unfeasible just a few

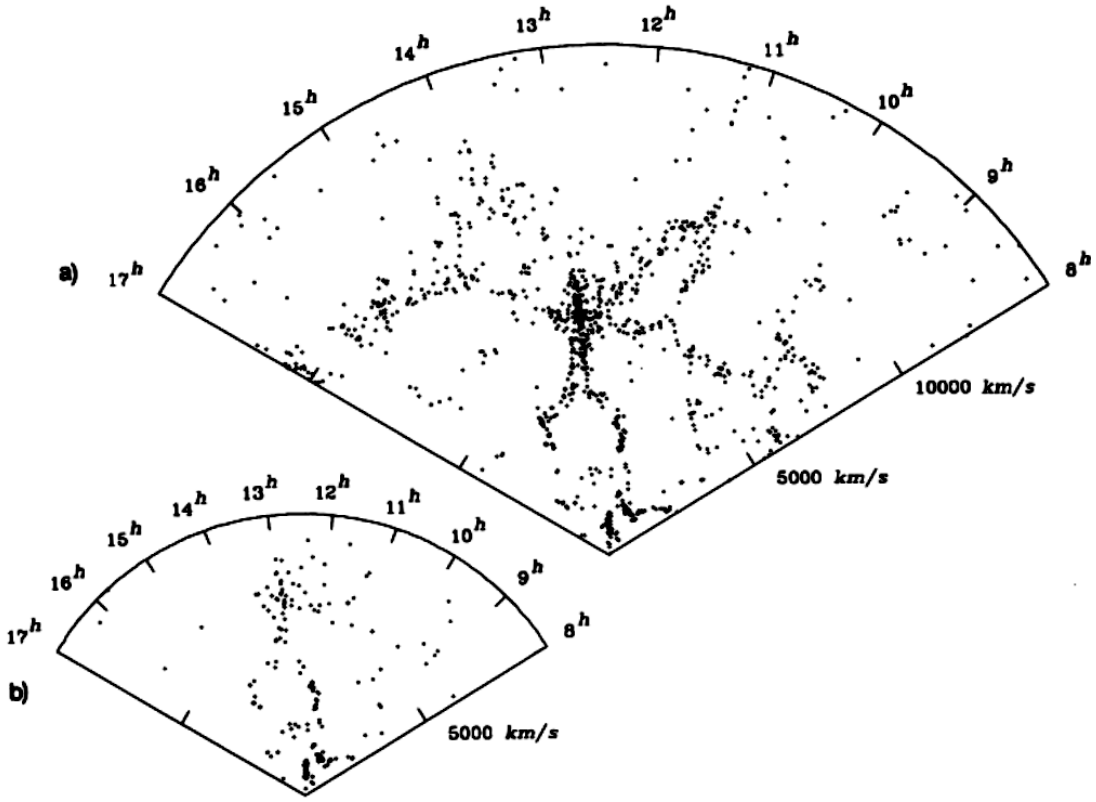

Figure 1.3: A slice through the CfA Redshift Survey showing the presence of the "CfA stick man". The image shows that galaxies are located in clusters and filaments separated by voids, not evenly distributed through space. Reproduced from de Lapparent et al. (1986).

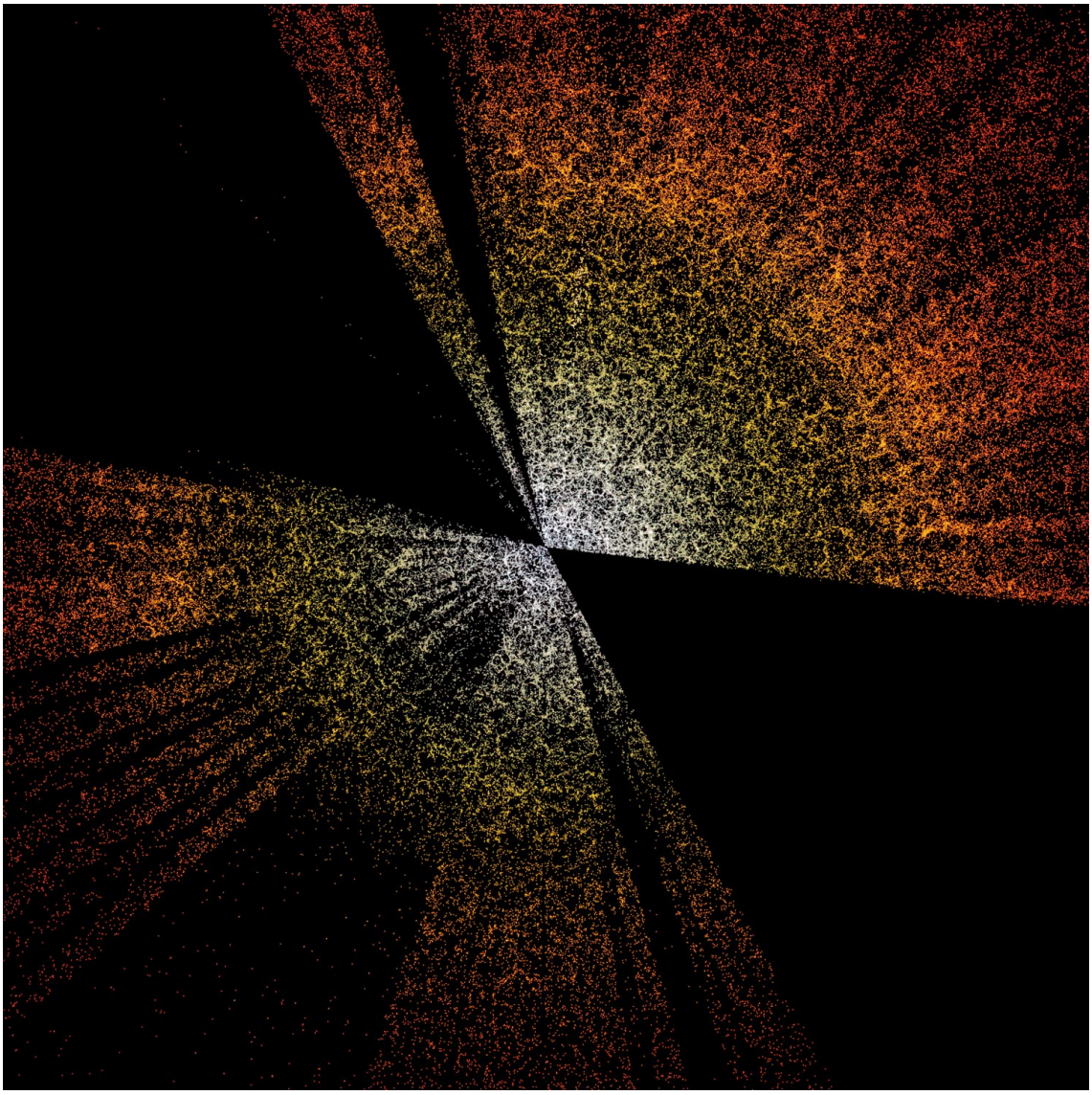

Figure 1.4: A slice through the DESI Survey containing a subset of 400,000 galaxies. The region shown is up to $z \approx 2$. Credit:D. Schlegel/Berkeley Lab using data from DESI

decades ago. Figure 1.4 shows a slice through data from the first few months of the Dark Energy Spectroscopic Instrument Survey containing 400,000 galaxies, a small subset of the total survey volume. Comparing Figure 1.3 to Figure 1.4 illustrates how the scale of galaxy surveys has expanded significantly. Examples of completed galaxy surveys include:

- The 2dF Galaxy Redshift Survey (2dFGRS) (Colless et al., 2001)
- The Sloan Digital Sky Survey (SDSS) (Abazajian et al., 2009)
- The Baryon Oscillation Spectroscopic Survey (BOSS) (Eisenstein et al., 2011)
- The Extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Dawson et al., 2016)
- The Galaxy And Mass Assembly Survey (GAMA) (Liske et al., 2015)

There are two main methods to measure redshifts in a galaxy survey. One is to use a spectrograph that splits the light from a target galaxy very finely to create a detailed spectrum of the flux density as a function wavelength. The observed spectrum of a galaxy can be matched to an existing template by identifying absorption or emission lines associated with particular atomic transitions. The wavelength difference between the observed spectrum features and the rest-frame positions of those features provides a measure of galaxy redshift. Spectroscopic methods provide the most accurate estimates for redshifts but require bright sources or long exposures to gather sufficiently many photons. Alternatively, photometric methods may be used to estimate galaxy redshifts. Light from the galaxy is passed through a series of broad filters and the total flux density in each filter is measured. These can be compared to templates to provide an estimate for the galaxy redshift. Photometric redshifts are less accurate than spectroscopic redshifts, but they allow for observation of fainter objects and require less observing time per object. The choice of observing method depends on the science goals for a particular project.

It is not only position data that is useful for constraining cosmology from galaxy surveys. The statistics of galaxy shapes and alignments can be used to probe weak gravitational lensing and the distribution of dark matter in the universe (e.g. Chang et al., 2018).

1.5 The DESI Survey

The **D**ark **E**nergy **S**pectroscopic **I**nstrument (DESI) Survey is an ongoing galaxy survey that will measure the positions of over 30 million galaxies from 2021-2026. DESI is hosted on the four-meter Mayall Telescope at the summit of Kitt Peak in Arizona. The science goals of the DESI Survey are to measure the matter power spectrum to high precision, particularly the BAO feature and the effect of redshift space distortions (RSD). The DESI instrument has set a record by measuring the redshifts of over 100,000 galaxies in a single night[1].

1.5.1 DESI Instrumentation

The Dark Energy Spectroscopic Instrument consists of 5000 fibres placed in ten petals that form a circular focal plate. Each fibre is capable of measuring a galaxy spectrum in a fifteen minute exposure[2] and the large number of fibres allows many galaxy spectra to be measured simultaneously. The arrangement of the fibre patrol regions on a tile is shown in Figure 1.5. Fibres are controlled using robotic positioners and targets are chosen using a fibre-assignment algorithm. The fibre assignment algorithm assigns as many fibres to galaxies as it can and preferentially targets those with the highest priority (DESI Collaboration, 2022). Limited patrol regions can mean that galaxies in dense environments are relatively undersampled, however it is possible to recover unbiased clustering estimates by upweighting undersampled galaxy pairs (Bianchi & Percival, 2017; Smith et al., 2019).

1.5.2 DESI Targets

There are four main extragalactic categories of targets in the DESI survey

- Bright galaxies - galaxies that are sufficiently bright to observe near twilight or when the moon is above the horizon. This target category is at low redshift (median $z = 0.2$) and contains galaxies with the broadest range of luminosities and colours.

[1] `https://perma.cc/ZM3L-VW2A`

[2] Exposure times differ between the Bright and Dark time surveys. Exposure times are scaled depending on observing conditions to achieve a specified spectral signal-to-noise ration for fiducial targets.

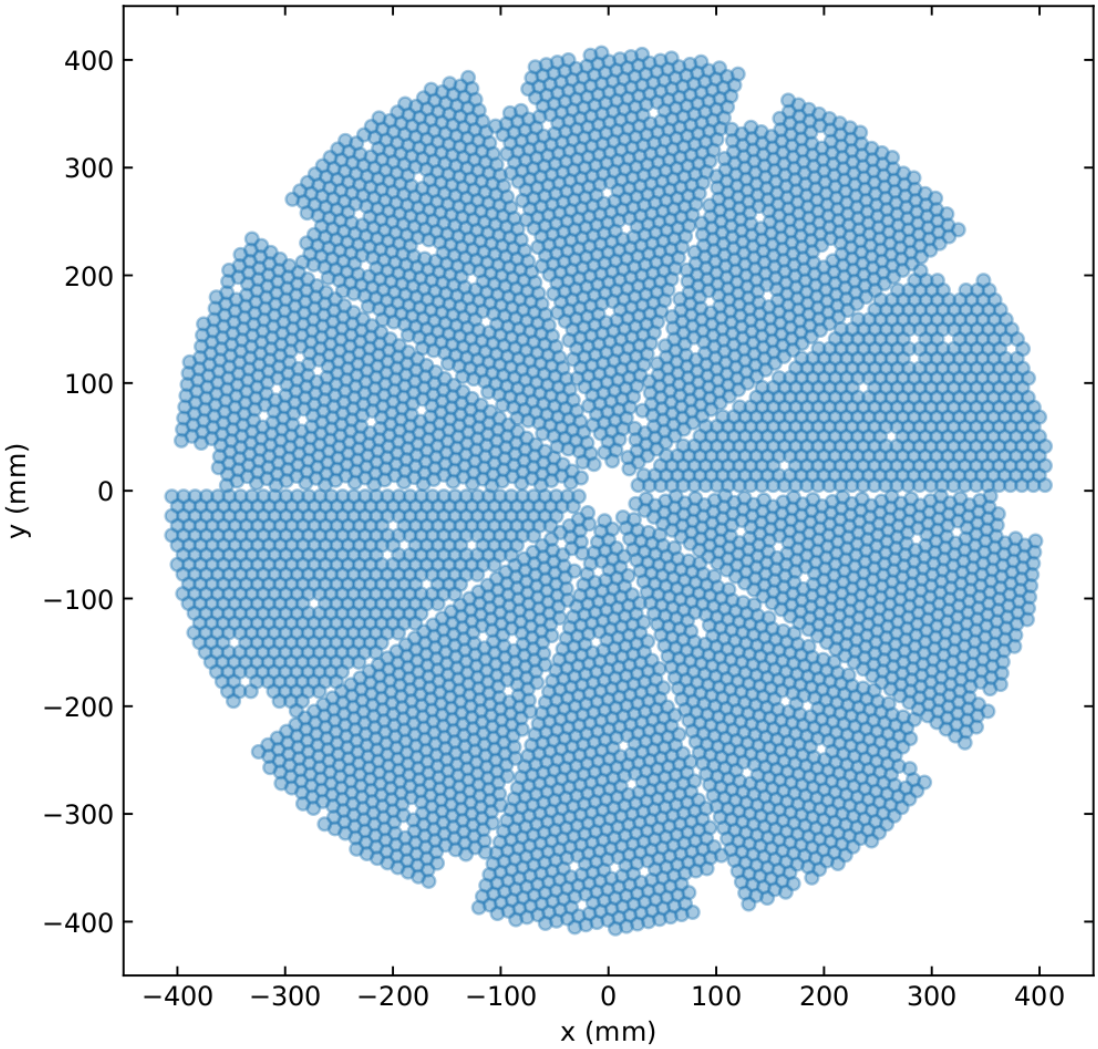

Figure 1.5: The arrangement of fibres on the DESI focal plane for a single DESI tile. The focal plane is made from ten petals. The blue regions represent patrol regions for individual fibres. Holes within petals represent the locations of fiducials that are used for calibration. Gaps around the edge represent the locations of Guide/Focus/Alignment (GFA) cameras. Reproduced from Smith et al. (2019).

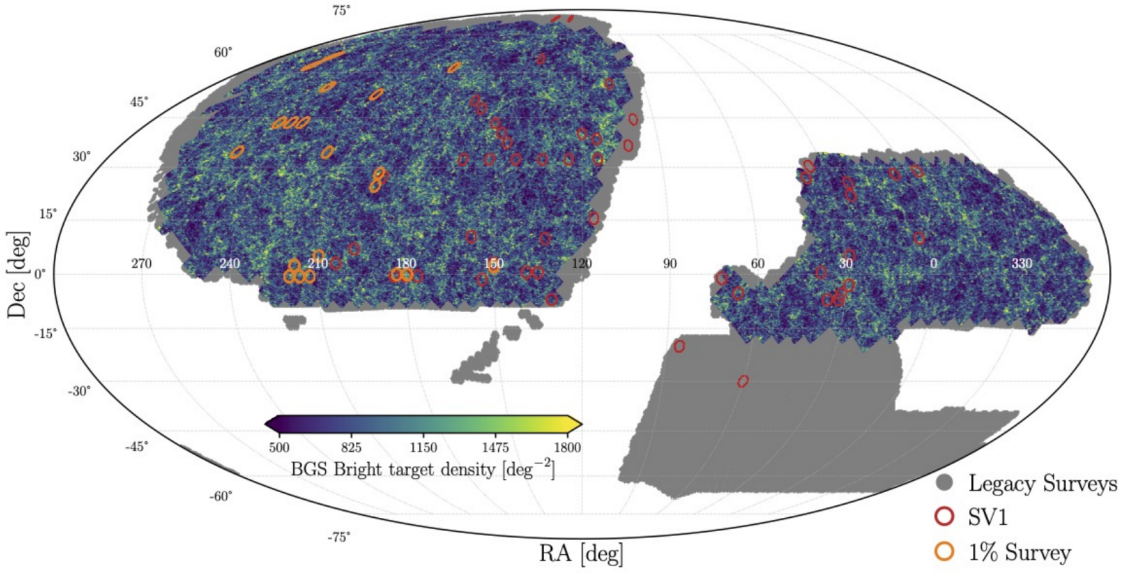

Figure 1.6: The footprint of the DESI Bright Galaxy Survey (coloured density map), along with tiles observed during initial survey validation (red) and the 1% survey (orange). Reproduced from Hahn et al. (2022).

- Luminous red galaxies (LRGs) - A homogeneous population of large and bright galaxies with a red colour
- Emission line galaxies (ELGs) - Strongly starforming galaxies at high redshift whose spectra contain features that can be matched to specific atomic transitions
- Quasars - Extremely bright sources powered by the infall of gas into a galaxy's central supermassive black hole. These sources can be observed to the largest redshifts in the survey due to their high luminosity.

In this thesis the bright galaxies are the most important target class. The survey targeting these objects is called the Bright Galaxy Survey (BGS). BGS is a nearly magnitude-limited survey of low redshift galaxies at $z < 0.6$. Two classes of target are selected, bright ($r < 19.5$ mag) and faint ($19.5 < r < 20.175$ mag with an additional fiber magnitude/colour cut to maintain high redshift efficiency). This survey will measure the redshifts of over 10 million galaxies and provide the most precise BAO and RSD measurements at low redshift (Ruiz-Macias et al., 2020, 2021; Hahn et al., 2022). Figure 1.6 shows the footprint of BGS on the sky.

In addition to the main survey, the 1% survey (also known as SV3) was completed

in May 2021. This analysed a subset of the total survey data to provide a representative sample with high completeness that can be used for validation of the survey strategy and to produce clustering measurements. The footprint of the 1% Survey, compared to the main BGS is shown in Figure 1.6. The 1% Survey covers an area of 180 $\deg^2$ and exposure times are 20% longer than the main survey in order to achieve a high spectroscopic completeness (Hahn et al., 2022).

1.5.3 Science Goals

The DESI Survey aims to measure BAO and RSD to high accuracy to constrain models of cosmology. Observed large-scale structure fits ΛCDM predictions very well but there are various theoretical models that can be tested by observations with higher precision. These include theories of modified gravity (Clifton et al., 2012) and dynamical dark energy (Zhao et al., 2017).

Deviations in the expansion history of the universe from ΛCDM could be revealed by precise measurements of the BAO scale at different redshifts. These measurements could inform whether the Universe is consistent with a constant value for Λ or not. The growth of structure in the universe is probed by redshift space distortions, this is sensitive to the nature of gravity. Modifications to gravity may help to explain the accelerated expansion of the Universe without Λ. DESI will allow us to test theories that are alternatives to general relativity (Alam et al., 2021).

1.6 N-body Simulations

This chapter so far has explored the development of the theory of ΛCDM cosmology and key observational evidence that is consistent with this theory. The creation of mock galaxy catalogues has been necessary to connect theory and observations from large galaxy surveys. Producing realistic mock galaxy catalogues is a non-trivial problem. Predictions of structure formation using perturbation theory are only accurate at early times or large scales where density perturbations are small. The formation of dark matter halos and the growth of small scale structure must be simulated numerically.

Simulations of particles interacting under gravitational forces are called N-body sim-

ulations. N-body simulations may be run at many different scales and under different conditions but this work shall focus on simulations of the evolution of large-scale structure. This means that the choice of cosmology affects the simulation as the scale factor of the universe increases through cosmic time.

1.6.1 Background & Development

Some of the first N-body simulations were run in the 1960s but these were limited to only around 100 particles (von Hoerner, 1960; Aarseth & Hoyle, 1963). Simulations of large-scale structure formation in the universe progressed rapidly in the 1980s (Davis et al., 1985; Efstathiou et al., 1985) thanks to advances in computing power and novel algorithms to simplify the necessary calculations (Eastwood et al., 1980; Appel, 1985; Barnes & Hut, 1986). Interest was also driven by observational progress in large-scale structure measurements from galaxy surveys.

Breakthroughs in N-body simulation research closely follow developments in computing. The computational complexity of N-body simulations means that there is a constant trade-off between volume and resolution. The increasing power of supercomputers over time allows for larger and higher resolution simulations to be run. Recently, individual simulations have been run with trillions of particles (Potter et al., 2017; Ishiyama et al., 2021), and suites of simulations contain particle numbers running into the tens of trillions (Maksimova et al., 2021). An example of the size of modern N-body simulation suites can be seen in Figure 1.7.

1.6.2 Initial Conditions

N-body simulations are run from initial conditions (ICs) that represent the situation in the early universe, where inflation has smoothed out the matter distribution in the universe and density perturbations are small. The initial density field is a Gaussian random field described by an initial power spectrum, which in turn is determined by the cosmology of the universe. Different realisations of the same power spectrum are possible by using different phase information in the random field (Jenkins, 2010).

The initial density field is sampled by particles that populate the N-body simulation,

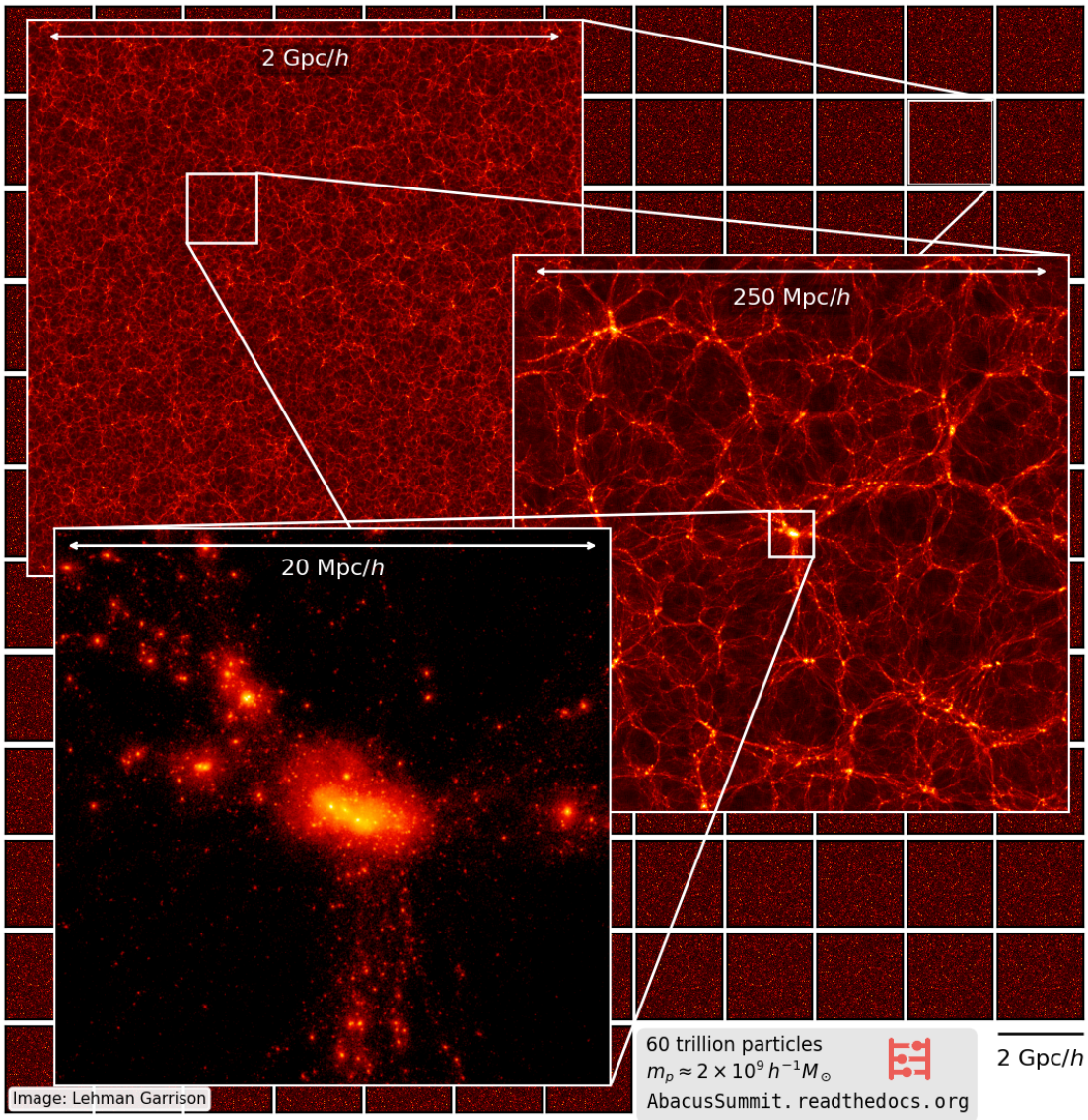

Figure 1.7: A visualisation of the AbacusSummit suite of simulations. Each box in the background represents one of the 139 base-resolution simulations. The slices shown are from snapshots at $z = 0.1$ and are $10h^{-1}$Mpc thick. Credit: The AbacusSummit Team.

a uniform grid is usually used to initialise the particles. Perturbation theory is used to translate the density field to displacements and velocities of particles. First-order perturbation theory is known as the Zel'dovich Approximation (Zel'Dovich, 1970) and second-order perturbation theory is also used (Jenkins, 2010).

1.6.3 Methods

Direct summation of inter-particle forces produces the most accurate evaluation of the gravitational force on a particle in an N-body simulation. However, this has the scaling relation $O(n^2)$ for n particles which causes runtime to increase rapidly with particle number. It would be unfeasible to run simulations at current volumes and resolutions by using direct summation alone, hence different techniques have been developed to improve the performance of simulation codes.

Particle-Mesh methods use a grid to evaluate the gravitational potential and apply forces to particles. The density field is calculated on a grid by assigning mass to grid-points. There are several methods to assign mass to the grid, such as nearest grid point or cloud-in-cell, however triangular-shaped-cloud (TSC) is the most accurate (Hockney & Eastwood, 1988). A downside to particle-mesh methods is that their accuracy is limited at small scales, by the spacing of the mesh. The runtime of the particle-mesh method scales with the runtime of the fast-Fourier transform, $O(n\log(n))$, where n is the gridsize.

Tree methods (Barnes & Hut, 1986) decompose the simulation box into sub-regions at different resolutions. The interactions between particles at large separations are then treated by approximating the particles in a sub-region as a single particle or multipole expansion which simplifies calculations. This algorithm scales as $O(n\log(n))$ for n particles. An accuracy parameter determines the level to which the sub-regions are divided before using the approximation.

The Fast Multipole Method (FMM) (Dehnen, 2014) is similar to a Barnes-Hut tree. However, instead of each particle interacting with a distant node, the node-node interactions are calculated and forces on the particles within the nodes are calculated from the expansion of the multipole field.

Particles undergoing large accelerations require more frequent force integration than those drifting undisturbed through space. N-body simulation codes make use of adaptive

timestepping to selectively calculate forces for particles in dense environments undergoing large forces without needing to repeat many similar calculations on particles in less dense environments.

Particles in N-body simulations can be treated as point masses for the majority of their interactions, however the r^{-2} scaling of the gravitational force term means that there is not a theoretical limit to the acceleration of particles if they are treated as point masses. We do not want significant two-body interactions between particles because the particles do not represent discretised units of mass in the universe, but some large number of diffuse particles of much lower individual mass. Force softening overcomes this issue by replacing the r^{-2} relation with one that does not diverge as $r \rightarrow 0$, for example $(r + \epsilon)^{-2}$.

Modern N-body simulation codes use a mixture of methods to achieve better performance at a range of scales when compared to a single algorithm in isolation. For example GADGET-2 uses direct particle-particle summation at small scales, a Barnes-Hut tree at intermediate scales, and a mesh at the largest scales (Springel, 2005).

In addition, N-body simulation codes are taking advantage of developments in computer hardware. SWIFT uses task-based parallelism to scale effectively on modern supercomputers that use thousands of connected nodes (Schaller et al., 2016). Task-based parallelism efficiently spreads computation tasks across the nodes such that dead-time is minimised, as opposed to dedicating each node to a sub-volume of the simulation box which can lead to an unbalanced load on each node. ABACUS accelerates the direct summation of forces with GPUs that are much faster than CPUs for simple arithmetic tasks (Garrison et al., 2021b).

1.6.4 Dark Matter Halos

Snapshot outputs from N-body simulations consist of particle positions and velocities, however some post-processing is necessary to extract the most value from this data. Particles can be grouped together into dark matter halos, gravitationally bound structures formed from the collapse of density perturbations or from the merging of smaller halos. Galaxies are thought to be hosted inside dark matter halos, therefore robustly being able to identify halos and characterise their properties is important to connect galactic observations and N-body simulations.

There are several halo-finding algorithms that can be used on the outputs of N-body simulations. The simplest is the friends-of-friends (FOF) algorithm (e.g. Davis et al., 1985; Springel et al., 2001; Jenkins et al., 2001). This connects particles that are separated by a smaller distance than a linking length into groups, each group being a separate halo. However, this algorithm does not distinguish separate halos that happen to be crossing one another, and two separate halos may be connected by a small bridge of particles and combined into a single group. The linking length is usually parameterised as a fraction, b, of the mean inter-particle spacing. A standard value is $b = 0.2$ (e.g. Davis et al., 1985; Jenkins et al., 2001). This value corresponds to a mean overdensity of around 180 times the mean density of the universe which is similar to the value for a virialized halo (Lacey & Cole, 1994).

An extension to the FOF algorithm uses velocity information to link particles in 6-dimensional phase-space. This can more robustly distinguish between halos on separate trajectories that are passing through one another (Behroozi et al., 2012; Elahi et al., 2019).

An alternative method for halo-finding is to define spherical regions of a particular density, usually defined as a multiple, κ, of the mean density of the simulation. $\kappa = 180$ is a standard value because it is similar to the value for virialized halos (Warren et al., 1992). Particles are assigned to halos if they lie within a region that satisfies the overdensity threshold.

The number density of halos as a function of halo mass is known as the halo mass function. Theoretical estimates for the halo mass function have been made by assuming the spherical gravitational collapse of gaussian density fluctuations. This produces a functional form that is a power law with an exponential cutoff at high masses (Press & Schechter, 1974). The Press-Schechter model predicts the halo mass function from the matter power spectrum and threshold for spherical gravitational collapse. This formalism can be generalised to a model of ellipsoidal collapse (Sheth et al., 2001) that introduces two free parameters. Analysis of the halo mass function from simulations is important because it is closely related to the observable luminosity function of galaxies, and can help to constrain models of the connection between galaxies and dark matter.

The matter distribution inside a simulated cold dark matter halo is well described by a universal profile known as the Navarro-Frenk-White (NFW) profile (Navarro et al., 1996).

The locations of specific subhalos in simulations can be found by identifying self-bound overdense groups (Springel et al., 2001). Observations of satellite & dwarf galaxies in the local universe present challenges for modelling small scales as discussed in Section 1.2.3

1.6.5 Uses of N-body simulations

N-body simulations have many uses in the study of cosmology. The clustering, mass function, and substructure of dark matter halos can be measured from the outputs of N-body simulations. These simulations allow us to see the effect of cosmology on these quantities at scales where they cannot be analytically calculated due to non-linear effects.

Realistic simulation data is necessary for modern galaxy surveys. From survey design, to final data analysis, N-body simulations provide utility. This is primarily through the creation of mock galaxy catalogues, simulated data that matches what could be observed by the survey.

Mock galaxy catalogues from N-body simulations can be used to test the unbiased recovery of cosmological parameters by comparing estimated cosmological parameters to the "true" values that were input to the simulation. Applying analysis pipelines to mock data provides estimates for the effects of systematic errors from large galaxy surveys. Various strategies to mitigate these errors can then be tested. In the era of precision cosmology, the minimisation of systematic errors is imperative as statistical errors are reduced by large sample sizes.

N-body simulations can be combined with semi-analytic models (Lacey et al., 2016) or extended with hydrodynamic effects (Schaye et al., 2014) to investigate how physical processes affect galaxy formation and evolution. Baryonic effects such as feedback from active galactic nuclei alter the matter power spectrum (van Daalen et al., 2019). However, dark-matter-only simulations are remain viable for cosmological analysis because these changes can be absorbed by models of the galaxy-halo connection (see Section 4.1.1).

Theories of modified gravity may be used in N-body simulations. If simulations of modified gravity can explain observed data better than simulations of ΛCDM then N-body simulations could help to push the boundaries of theoretical cosmology beyond the modern consensus (Hernández-Aguayo et al., 2022b).

1.7 Thesis Structure

The aim for this thesis is to explore contemporary N-body simulation codes and their applications towards making mock galaxy catalogues for the DESI survey.

Chapter 2 investigates the accuracy of the gravity scheme within the N-body simulation code SWIFT. Chapter 3[3] presents the results from comparisons between N-body simulation codes that have been run from the same initial conditions, along with a comparison of initial conditions generation codes.

Chapter 4 presents a method for fitting Halo Occupation Distribution (HOD) parameters in order to produce mock galaxy catalogues for BGS that match a target luminosity function and magnitude dependent clustering. A halo tabluation method is used to accelerate the fitting process and meta-parameters describing the evolution of HOD parameters as smooth functions of magnitude are used. In Chapter 5, the variation in HOD fits with simulation cosmology is explored, the method for populating BGS mock catalogues is presented, and comparisons between the mock catalogues and BGS data are made.

A summary of data science work undertaken in external placements is described in Chapter 6. This included building an anomaly detection model for wastewater flows, and projecting school pupil performance for the Department for Education.

Finally, conclusions are drawn and future avenues of research are suggested in Chapter 7

[3]Chapter 3 has been published in the form of a paper.

CHAPTER 2

Numerical Tests of the SWIFT Code

For N-body simulations to provide value, it is imperative that the underlying physics is being simulated with sufficient accuracy. Approximations are used to avoid the $O(N^2)$ scaling of the direct particle-particle summation of the gravitational force, and the use of discrete time-steps means that the orbits of particles undergoing extreme accelerations cannot always be accurately simulated. The accuracy of these approximations is often inversely related to the computational cost of running a simulation, therefore it is useful to understand these limitations in order to optimise code performance.

In this chapter I shall present tests of the gravity scheme in SWIFT in a variety of regimes; at early and late times, and with different force accuracy parameters.

In order to verify that the gravity schemes are correctly implemented and measure their accuracy, there are several approaches. Firstly one can compare to theoretical estimates of quantities such as the power spectrum or correlation function, however these theoretical estimates only apply accurately at large scales or early times. Secondly, one can compare the code to itself while being run in both high and low accuracy conditions (convergence testing). This does not give an objective view of the code's accuracy but does provide information on whether specific parameter choices will impact the output of the code. Finally the code can be compared to other N-body simulation codes. This also does not

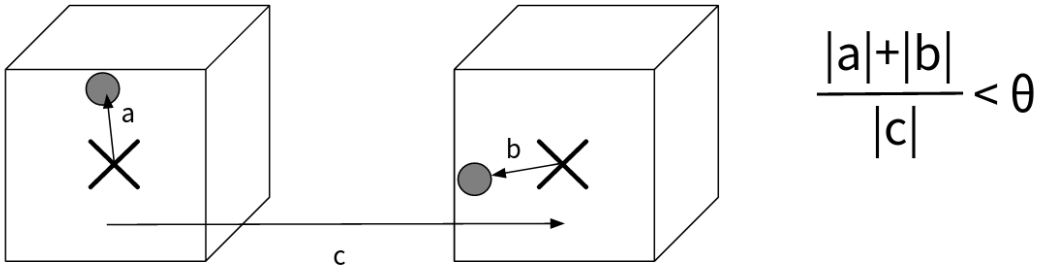

Figure 2.1: A schematic diagram of the fast multipole method's opening angle criterion. For particles at positions a & b relative to their cell centres separated by distance c, the multipole expansion of the potential is used if the inequality on the right is satisfied. Otherwise the cells are subdivided and the procedure is repeated. Direct summation is used for the closest of particle pairs.

give an objective view of accuracy, but if multiple codes with independently designed algorithms and approximations provide similar answers, then that implies control over systematic errors in the results.

The first two approaches are performed in this chapter, comparing SWIFT to linear theory and convergence testing. In the next chapter a comparison of several N-body simulation codes is presented, with a focus on applications to the DESI Survey.

2.1 SWIFT Overview

SWIFT (**S**moothed particle hydrodynamics **W**ith **I**nter-dependent **F**ine-grained **T**asking) is a simulation code that has been developed to run large simulations of galaxy formation on modern supercomputers.

SWIFT has been designed to take advantage of developments in supercomputers to run massive hydrodynamic and gravitational simulations on highly parallel architectures. SWIFT uses a task-based parallelisation strategy to maximise performance, minimise inter-node communication, and balance the load on each node used in the simulation. This allows SWIFT to achieve better performance than GADGET-2 and to have excellent strong- and weak-scaling relations when running simulations with particle numbers across many orders of magnitude on differently sized systems (Schaller et al., 2016). Strong-scaling measures the speedup when running a fixed problem size on increasing numbers of nodes,

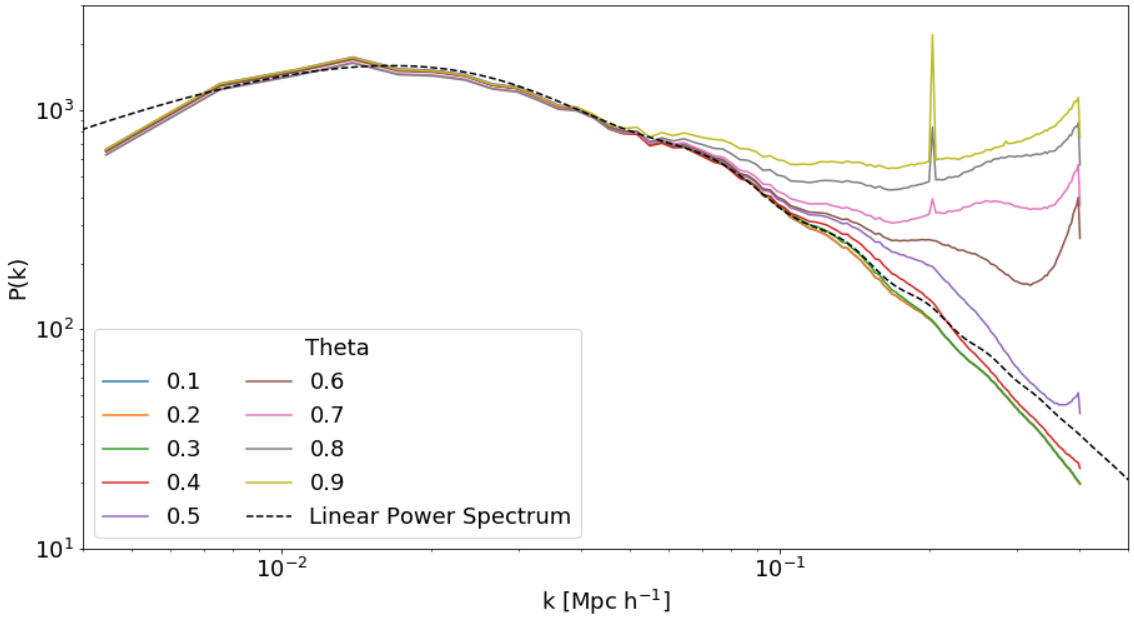

Figure 2.2: The power spectrum at z=4 for runs with different θ. The power spectrum converges for $\theta < 0.3$ and begins to diverge from the linear power spectrum at high k as θ increases. In addition spikes appear in the power spectrum at high θ.

weak-scaling measures the speedup when changing the problem size proportionally with the number of nodes.

SWIFT is not limited to only running simulations on large cosmological scales. Fluid dynamics on all scales can be simulated by SWIFT. One promising avenue is the study of planetary impacts, including the hypothetical origin of the moon (Ruiz-Bonilla et al., 2020; Kegerreis et al., 2022). There are a variety of hydrodynamics schemes which can be used by SWIFT to simulate gas and fluid properties in simulations. In these tests, however, I am purely investigating the dark-matter-only performance of SWIFT with no hydrodynamic features active.

SWIFT uses the fast multipole method (FMM) (Cheng et al., 1999; Dehnen, 2014) to quickly evaluate the gravitational forces on particles. This algorithm is $O(N)$ where N is the number of particles. This method calculates the force between particles within cells as a multipole expansion about the centre of their respective cells. An opening angle parameter θ determines whether FMM is used at a given level in the hierarchy of the cell structure, see Figure 2.1 for a diagram. Smaller θ leads to a more accurate but time consuming simulation. The short range forces are softened to avoid unwanted two body effects (Athanassoula et al., 2000). At the largest scales a particle-mesh algorithm is used.

The timestepping criterion in SWIFT cosmological runs is based on a maximal change in $\log(a)$, where a is the expansion factor of the universe. This leads to smaller timesteps at early times when accurate integration of the linear regime is important. There is also a timestep criterion based on the acceleration experienced by a particle which becomes important at late times as particles in denser environments experience larger accelerations.

2.2 Convergence Testing

Convergence testing is important for N-body simulations. Most codes could be run to extremely high accuracy however this would be prohibitively expensive in the computational cost of the simulation. In convergence testing one aims to change a parameter until the results that one is interested in stop changing to within a desired accuracy. In this section the opening angle parameter in SWIFT, θ, is convergence tested.

To do this, simulations were run from a low resolution set of initial conditions with a spectrum of θ values from 0.1 to 0.9. The simulations were run from z=63 to z=4 and the power spectra at z=4 were compared as shown in Figure 2.2. Figure 2.2 shows that the power spectrum results only converge for $\theta < 0.3$ which is relatively low and leads to long runtimes. In addition, spikes in the power spectrum are present that are related to the Nyquist frequency of the initial particle grid.

This result implied that a deeper investigation of force inaccuracies in SWIFT was necessary. Are the calculated forces close to the exact forces? Which part of the gravity calculation was causing these errors? Does the effect of the force inaccuracy change with redshift?

2.3 Force Accuracy

Firstly I investigated how the forces on particles changed with θ. I ran SWIFT for half of a timestep from the initial conditions and produced a snapshot. The particle positions of this snapshot are exactly the same as the initial conditions however the velocities have been changed by one "kick". The difference in velocities between the snapshot and the initial conditions is therefore proportional to the force felt by the particles in one timestep.

A plot of this is shown in Figure 2.3. There is a clear structure in the forces when high θ is used. Taking the difference between the high and low θ forces makes the pattern more clear and there is no grid pattern seen when low θ is used. The grid seen in Figure 2.3 is 28x28. A possible cause of this pattern is a beating between the grid used to evaluate FMM forces and the particle grid. The particle grid is 540^3 while the FMM grid is a power of two. If the FMM grid was 512^3, $540 - 512 = 28$ therefore beating between these different grids could be the cause of the observed 28x28 pattern. This grid pattern and the beating relationship between the grid sizes was also seen for other particle numbers and FMM grid sizes.

2.3.1 Comparison to Exact Forces

It is also important to test against the exact force as well as the low θ force. Just because the low theta force is converged, does not mean it is correct. I calculated the exact accelerations of 10000 particles and compared these to the calculated accelerations with varying θ. The exact accelerations were calculated by summing the Newtonian force on each of the subsampled particles from every other particle in the simulation. A histogram of the fractional force inaccuracy is shown in Figure 2.4. The forces are accurate for $\theta = 0.1, 0.3$ but become more inaccurate as θ increases beyond 0.3. This corroborates the results from the power spectra which showed that beyond $\theta = 0.3$ the power spectrum result was no longer converged. For $\theta = 0.7$ a double peaked distribution emerges as the calculated forces are systematically avoiding the correct answer. This is likely to be caused by the grid pattern inaccuracy dominating over the true force. Figure 2.5 shows that the force inaccuracy does not depend on the exact force which is consistent with the error being caused by the grid pattern which is independent of the exact force on a particle.

2.3.2 Low Redshift and Glass Initial Conditions

The above tests were performed on initial conditions of particles perturbed from regular gridpoints. It was important to investigate whether the same effect was observed at low redshift or from glass initial conditions. Glass initial conditions do not displace the particles from a regular grid but instead from points that are uniformly spaced in an

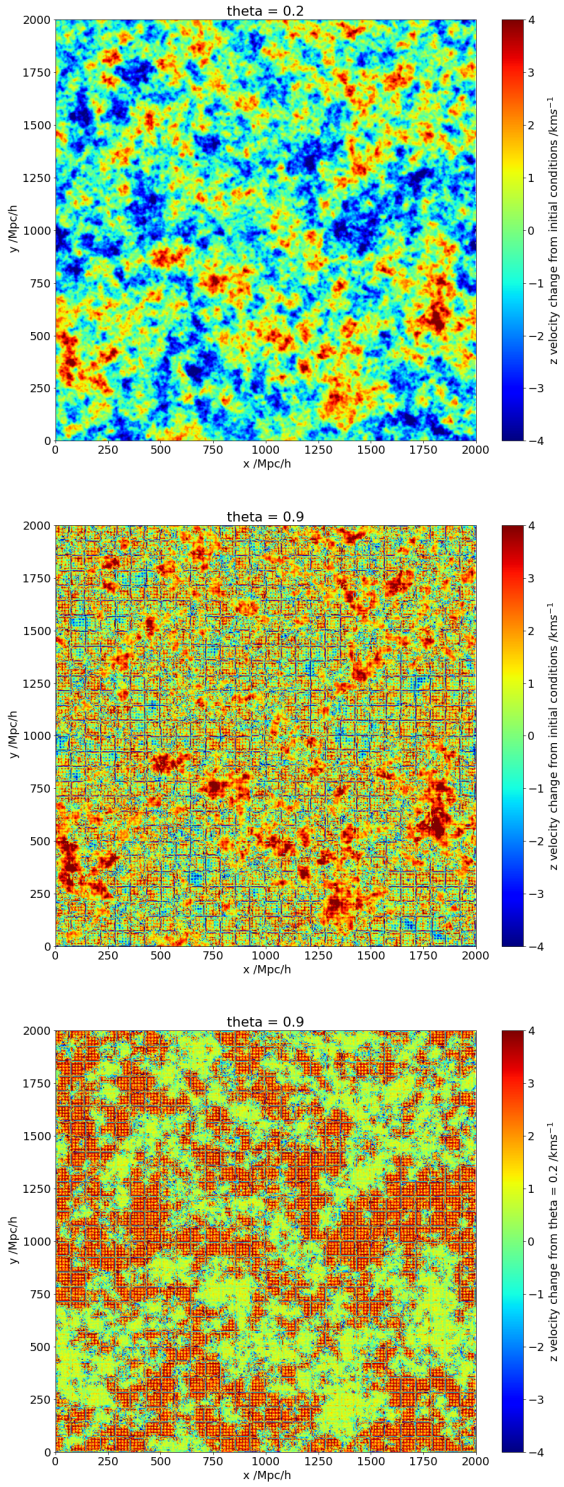

Figure 2.3: The velocity change in the z direction for particles in a single slice through the initial conditions after one timestep. Upper: $\theta = 0.2$, Middle: $\theta = 0.9$, Lower: The difference between $\theta = 0.9$ and $\theta = 0.2$ which clearly shows a 28x28 grid pattern

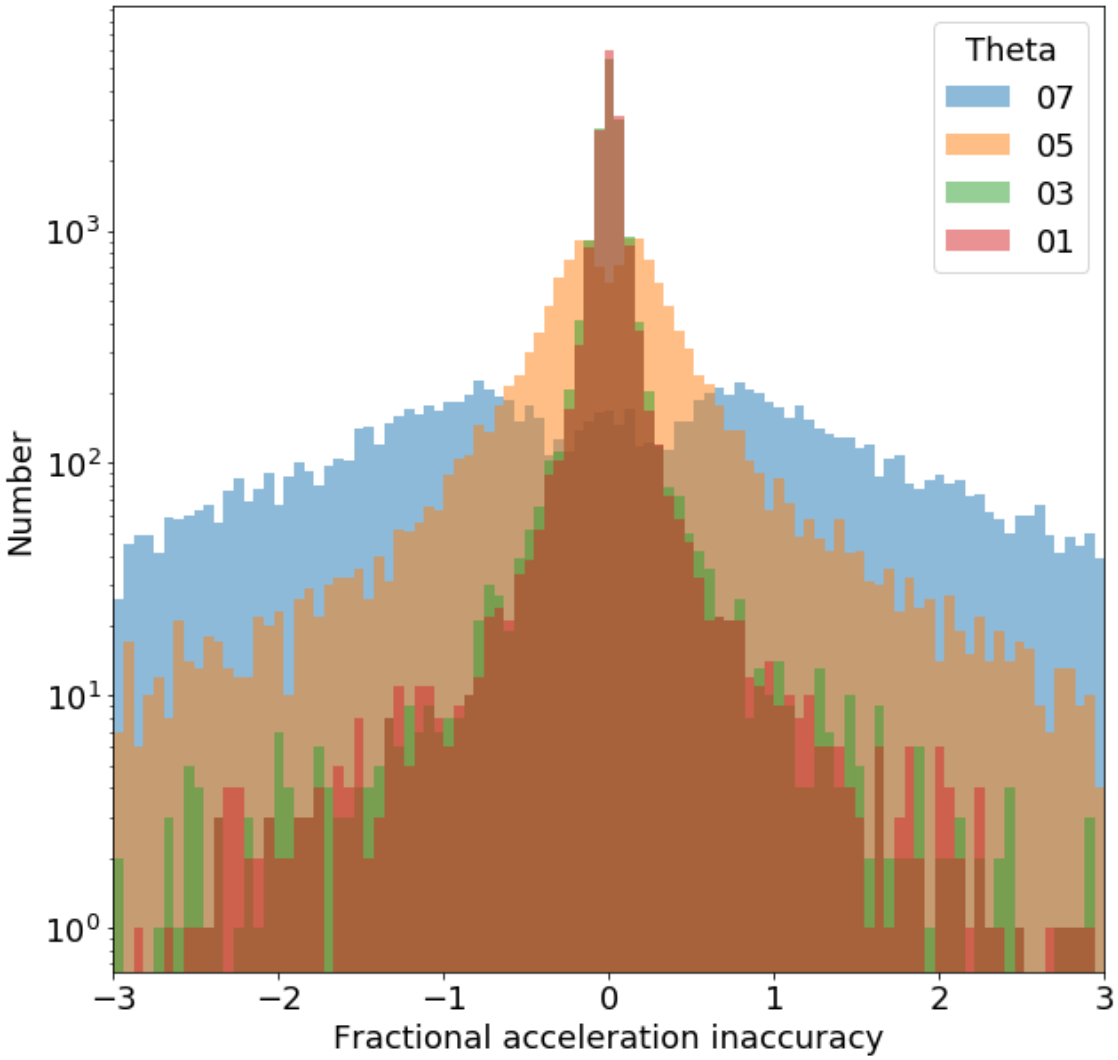

Figure 2.4: The fractional acceleration inaccuracy after one timestep for different θ values. The fractional acceleration accuracy is defined as $(a_{\text{calc}}/a_{\text{true}}) - 1$, where the z component of the acceleration vector is used in all cases. The acceleration becomes more inaccurate with θ increasing above 0.3.

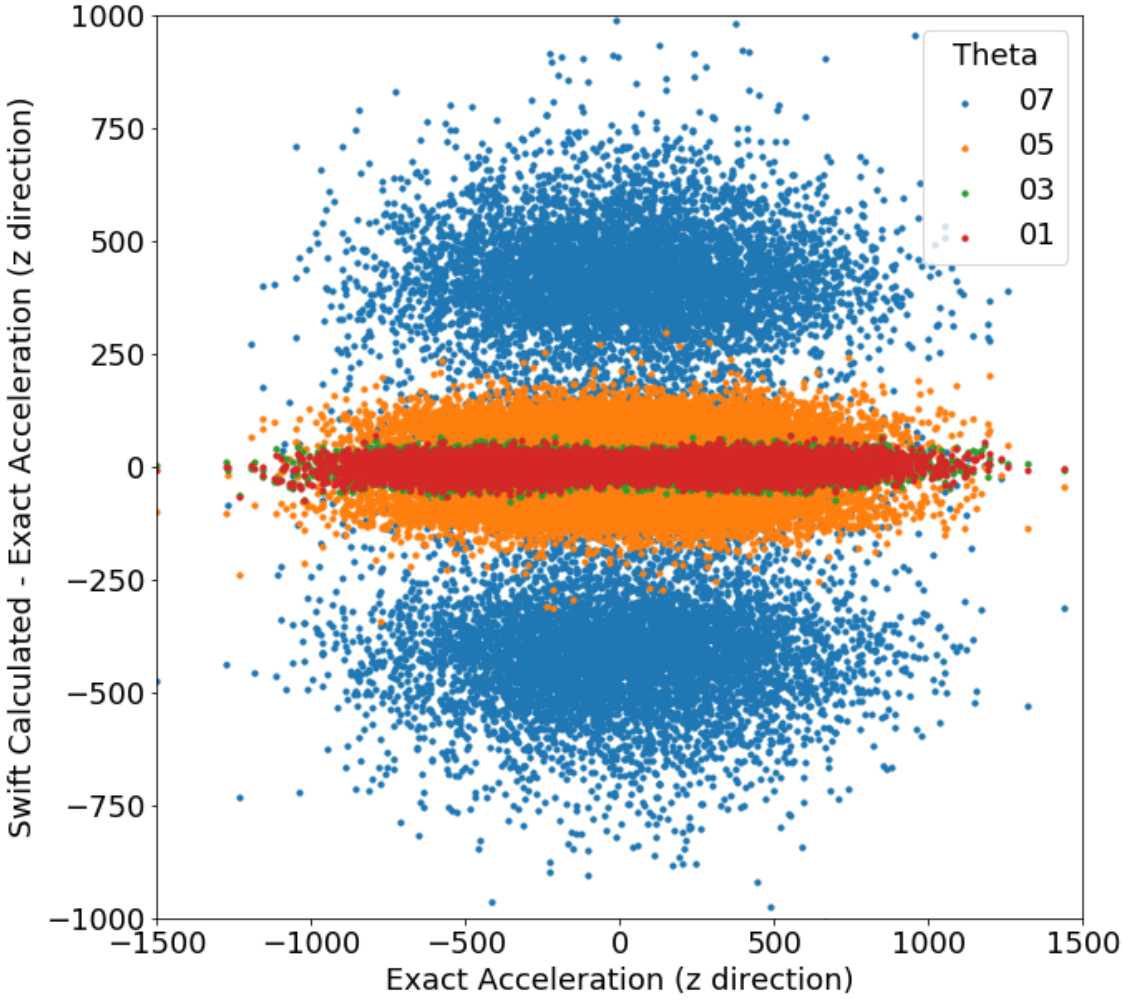

Figure 2.5: The acceleration error plotted against the exact acceleration per particle. The error in the acceleration of particles in SWIFT does not depend on the exact acceleration which the particle experiences and the error increases with θ.

isotropic manner. These initial conditions can be created by running an N-body simulation with the sign of the force of gravity inverted until all the particles are evenly spaced due to the repulsion (LHuillier et al., 2014).

Figure 2.6 shows the same as Figure 2.3 but from a higher resolution set of glass initial conditions. There is no longer a grid pattern present, this is because the absence of a particle grid now means there is no beating present and therefore the force errors are randomised.

At low redshift there is no pattern in the force inaccuracies and the fractional force inaccuracy is very low compared to the initial conditions as shown in Figure 2.7. At low redshift particles experience large forces and there are extreme variations in density across the simulation box, the lack of a regular pattern or evenly spaced particles means that the forces are now calculated more accurately by the FMM. It is possible to use much higher θ at low redshift without significantly affecting force accuracy.

In summary, I found that there were significant inaccuracies in SWIFT forces at high redshift when using θ values above 0.3. My tests show that this is likely to be caused by a beating between the particle grid and the grid used to calculate the FMM force. However, at low redshifts where the particle grid is no longer present these inaccuracies are much smaller for all values of θ.

2.4 Comparison With Perturbation Theory

Another useful test of the accuracy of an N-body simulation code is to compare to theoretical predictions of the power spectrum at large scales. Perturbation theory provides an estimate for the growth in the power spectrum as the universe evolves based on the cosmology of the universe (Bouchet et al., 1995). This is most accurate at large scales where the effect of non-linear evolution is minimised because the fluctuation amplitude is smaller.

The perturbation theory prediction in an evolved simulation was estimated by running the initial conditions code using the same phase information as was used in the simulation, but with the output set to a low redshift. Second-order perturbation theory was used to produce this power spectrum estimate (Jenkins, 2010).

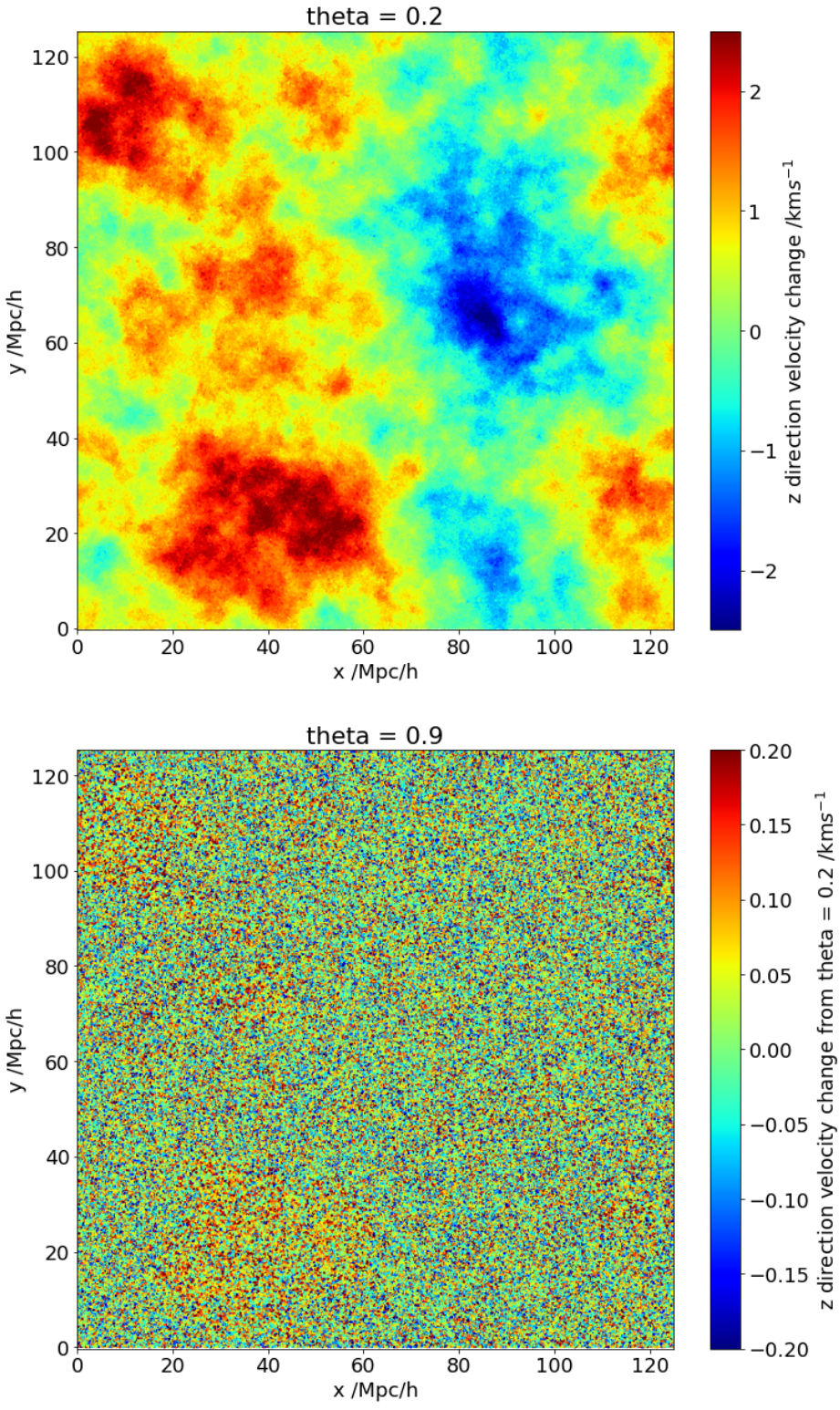

Figure 2.6: Same as Figure 2.3 but for glass initial conditions in a smaller boxsize. Upper: $\theta = 0.2$, Lower: The difference between $\theta = 0.9$ and $\theta = 0.2$. There is no grid pattern in the acceleration errors at high θ.

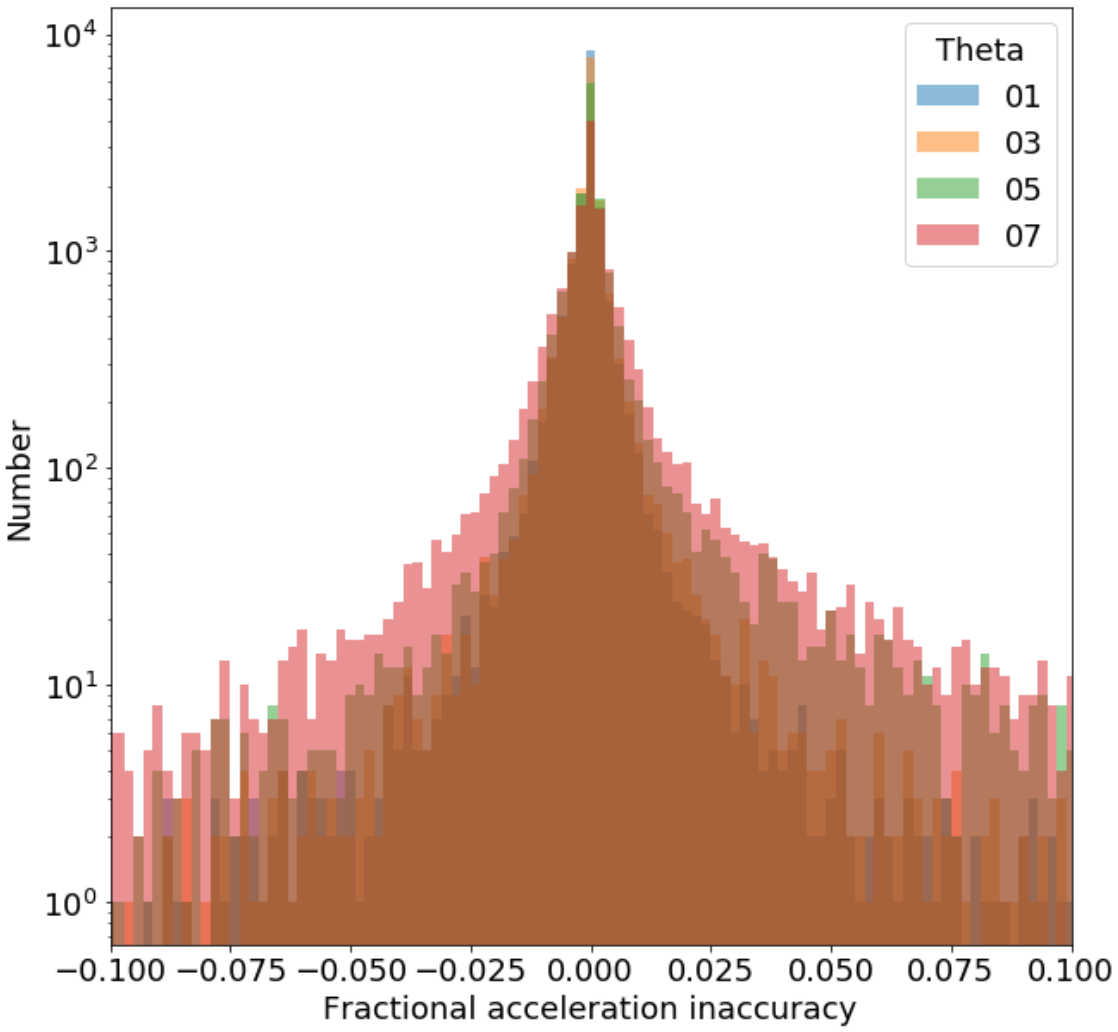

Figure 2.7: Same as Figure 2.4 but for a snapshot at redshift 0.1. The fractional acceleration errors are much smaller than in the initial conditions and only become slightly worse with increasing θ.

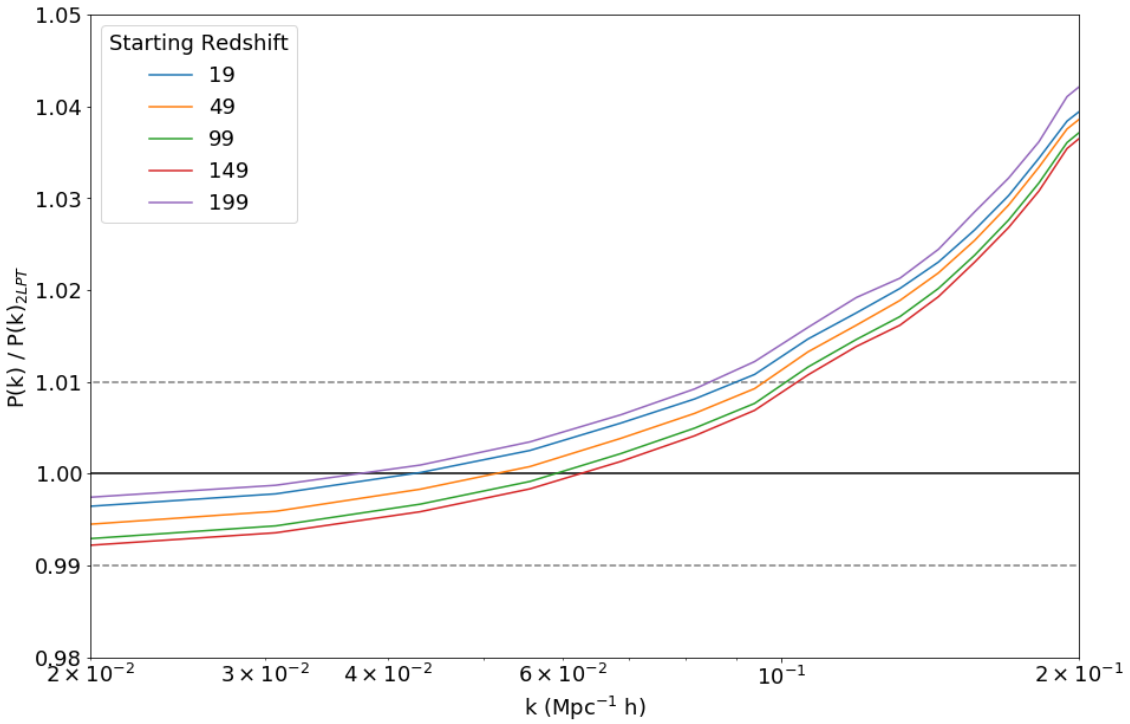

Figure 2.8: The ratio of the power spectrum from SWIFT at z=2 relative to the perturbation theory estimate at that redshift. The different line colours correspond to different starting redshifts for the simulations which affects the global timestep size. The ratio becomes asymptotically flat at large scales, as expected, but with an offset which corresponded to an error within the SWIFT code.

SWIFT showed differences to linear theory in the power spectrum at large scales which varied linearly with the size of the global timestep. The size of the global timestep was modified by starting the simulation at a different redshifts, although modifying the end time of the simulation could have been used as an alternative. This is because the global timestep size is forced to be a fraction of the global timeline (which is set by the start and end redshifts of the simulation) divisible by a power of two. The power of two used as the divisor depends on the maximum global timestep size which is defined in the input parameters to SWIFT. These effects can be seen in Figures 2.8 & 2.9.

Other codes such as GADGET-2 did not show such an effect which implied that it was due to a problem within SWIFT. This led to the developers finding a bug in the gravity scheme where the long-range gravitational potential was not updated sufficiently frequently which introduced inaccuracies in the integration scheme. This bug was fixed and this brought the large-scale power spectrum in SWIFT in line with GADGET-2 and perturbation theory with no dependence on timestep size. This updated version of the code was used in Chapter 3.

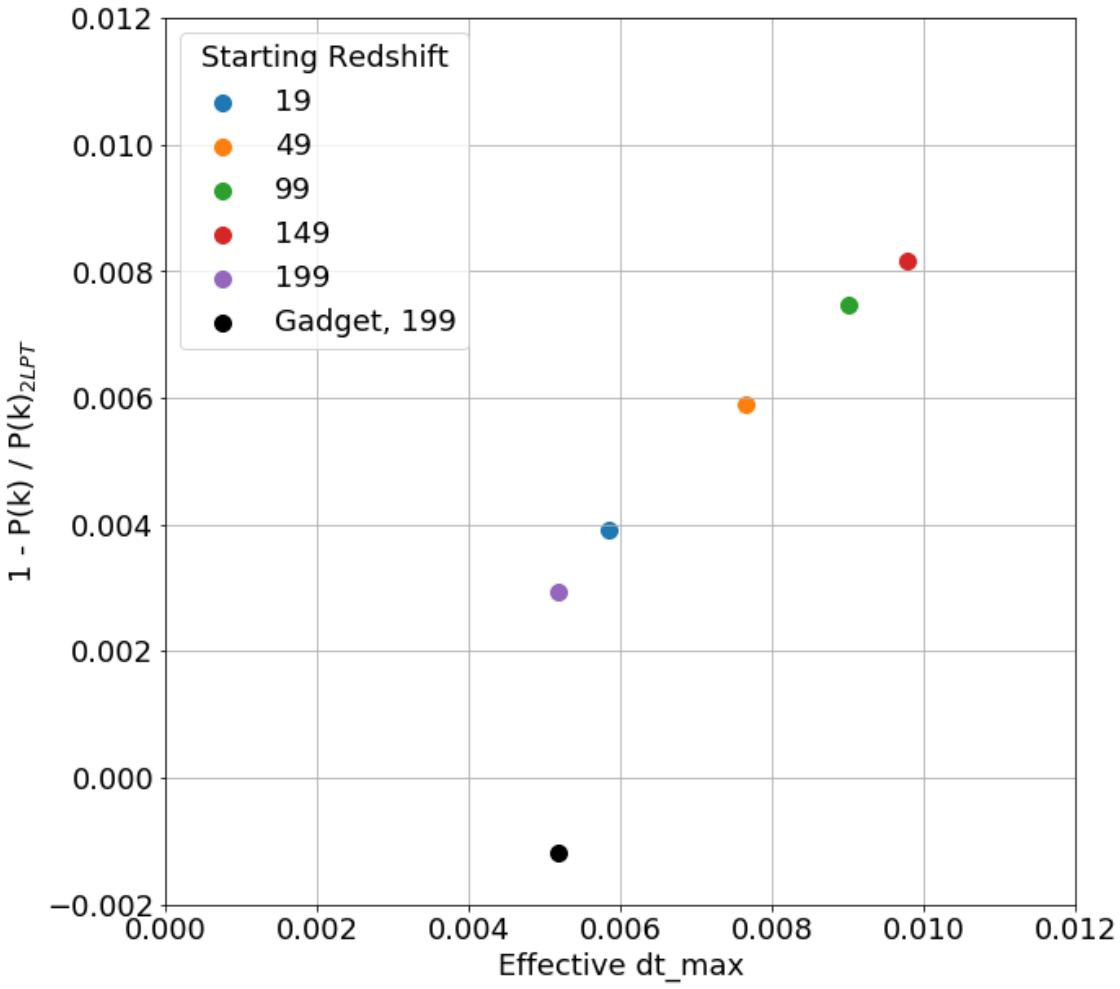

Figure 2.9: The ratio of the power spectrum in the largest scale bin ($k = 2 \times 10^{-2}$ Mpc^{-1}h) between SWIFT run from different starting redshifts and the perturbation theory estimate. The x-axis shows the global timestep size in each simulation, in units of $\log(a)$. A clear linear relationship between global timestep size and large-scale power spectrum error can be seen. The black point shows the position of GADGET-2 in this analysis, GADGET-2 does not show a sensitivity with inital simulation redshift.

2.5 Conclusions

The limitations of N-body simulations must be understood if one is to use them effectively. There is a tradeoff between accuracy and runtime, convergence testing may be used to understand how changing code parameters can affect simulation results.

In this chapter, the accuracy of the gravity scheme in the SWIFT code has been measured. A parameter that has a large effect on code accuracy is θ, the "opening angle" in the fast multipole method (FMM). The effects of varying θ on the power spectrum and on the forces experienced by particles have been measured. At high redshift, for cubic grid initial conditions, the particle accelerations are very sensitive to θ, only converging for $\theta < 0.3$. The errors in force calculations are unrelated to the exact force experienced by the particle (as calculated from direct summation). In addition, force inaccuracies form a grid pattern that is related to a beating between the particle and FMM grids. However, at low redshifts or from glass initial conditions, higher θ values do not lead to such large inaccuracies.

Comparisons between SWIFT and second order perturbation theory estimates of the power spectrum revealed a disagreement of up to 1% at large scales. Perturbation theory and N-body simulations should agree asymptotically at large scales. The power spectrum error showed a linear dependence on simulation timestep size meanwhile the corresponding case for the GADGET-2 simulation code showed no dependence on the timestep. It was found that an integration error was causing this difference in the large-scale power spectrum, and SWIFT was updated such that the calculations were corrected.

CHAPTER 3

The DESI N-body Simulation Code Comparison Project

3.1 Introduction

For many years N-body simulations have been used as a tool to explore the nonlinear evolution of the distribution of matter in the Universe (Davis et al., 1985). Their use has been invaluable in creating mock galaxy catalogues to validate the results of surveys such as 2dFGRS, SDSS, DES, KiDS, and more (Cole et al., 1998; Rodríguez-Torres et al., 2016; DeRose et al., 2021; Jong et al., 2012). The Dark Energy Spectroscopic Instrument (DESI) will be performing surveys of unprecedented size over the next five years, measuring tens of millions of galaxy spectra (DESI, 2016). The large size of the DESI survey means that the statistical errors on key measured quantities will be small and therefore keen attention must be paid to systematic errors in all stages of the data collection and analysis. Measuring the size of systematic errors introduced by N-body simulations is imperative to understand their impact on mock galaxy catalogues and analysis pipelines.

One way to measure systematic errors for an individual N-body code is using convergence testing. The code can be run with progressively more accurate parameter choices until measured statistics no longer change (Power et al., 2003; DeRose et al., 2019). However, convergence testing is limited because it only provides information on the systematic

errors which can be reduced by running a code with greater numerical accuracy. Different implementations or assumptions adopted by different codes could introduce systematic errors as well. Comparing the results from several different codes can indicate the level of control over systematic errors in converged runs of N-body simulation codes.

Code comparison projects measure the precision of N-body simulation codes when run from identical initial conditions. Measurements of the power spectrum, clustering and halo properties can be compared to indicate the precision to which N-body simulations can estimate these quantities (Winther et al., 2015; Schneider et al., 2016; Garrison et al., 2019).

This paper contains the results from multiple code comparison projects using several modern N-body simulation methods. Firstly we compare the initial conditions created using several different codes, measuring the power spectrum along with particle velocity statistics. The main N-body code comparison project compares simulations run with ABACUS, FASTPM, GADGET-2, and SWIFT from identical initial conditions. The matter power spectrum is measured, along with dark matter halo properties, abundances, clustering and power spectra. The precision of these results between different N-body codes and simulation resolutions is measured.

In Section 3.2, the initial conditions comparison is performed between three different codes. In Section 3.3, we discuss the different codes used to run the N-body simulation and we compare the results of the N-body simulations in terms of the matter power spectrum. In Section 3.4 we compare the halo catalogues produced from the different simulation codes, including measurements of the clustering, halo mass functions, and matched halo properties. Finally we summarize and conclude in Section 3.6.

3.2 Initial Conditions Codes

The starting points for N-body simulations are the initial conditions (ICs). These define the positions, velocities and masses of particles which are to be fed into the N-body simulation code. Analysis of the cosmic microwave background radiation suggests that the initial matter density field produced during inflation is a Gaussian random field (Planck Collaboration, 2020). The properties of this field are determined by its power spectrum,

Table 3.1: The number of particles and particle masses of the simulation boxes. All are cubic boxes with a side length of 500 Mpc/h.

Simulation	Particle Number	Particle Mass $/h^{-1}\mathrm{M}_{\odot}$
Low Resolution	1296^3	5.00×10^9
Medium Resolution	1728^3	2.11×10^9
High Resolution	2592^3	6.25×10^8

which in turn is determined by the cosmology of the universe. Different realisations of the same power spectrum are possible as the phases of the field are randomly generated (Abrahamsen, 1997).

In this section, we discuss three different IC generators. The considered IC codes include those used in conjunction with the ABACUS (Garrison et al., 2021b) and FASTPM (Feng et al., 2016) N-body codes, in addition to a combination of two codes, PANPHASIA[1](Jenkins, 2013) and IC_GEN which we refer to as PANPHASIA in this text, see Section 3.2.1 for details.

The cosmology is set according to the flat ΛCDM Planck 2018 results (Planck Collaboration, 2020) with the following parameters: the Hubble expansion rate at present time, $H_0 = 67.36$ km s^{-1} Mpc^{-1}, $\Omega_c h^2 = 0.1200$, $\Omega_b h^2 = 0.0223$, $\Omega_\Lambda = 0.6834$, r.m.s linear density fluctuation, $\sigma_8 = 0.81$, scalar spectral index $n_s = 0.9649$.

Considering the requirements of the DESI project, we run the simulations in boxes of side length, $L_{\mathrm{box}} = 500h^{-1}$Mpc with three different dark matter particle numbers: $N_{\mathrm{p}} = 1296^3$, 1728^3 and 2592^3. The initial conditions are generated at redshift, $z_{\mathrm{ini}} = 199$ with the input theoretical matter power spectrum generated using CAMB[2]. The details of the methods employed by each code are described below. We study the consistency between the codes in both the dark matter position and velocity distributions. The convergence test in each individual code is also performed using the three sets of resolutions. In Table 3.1, we provide the specifications of the simulations.

3.2.1 Code details

This section contains descriptions of the codes used to create initial conditions used in the comparisons in section 3.2.2.

[1]http://icc.dur.ac.uk/Panphasia.php
[2]https://camb.info/

ABACUS IC

The ABACUS IC code, called ZELDOVICH-PLT[3], generates Zel'dovich Approximation initial conditions, optionally applying Particle Linear Theory (PLT) corrections (Marcos et al., 2006; Garrison et al., 2016) for the breaking of linear theory that occurs on small scales due to the discrete representation of the density field by particles. The PLT corrections are not enabled for purposes of this code comparison. The code is written in C++ and is designed to produce initial conditions in memory-limited environments by buffering the state to disk. The code uses double precision internally, and the positions are output as single-precision displacements on a lattice. The random number generator for the initial Gaussian density field outputs 64 bits and allows synchronizing the white noise at different resolutions; this capability is employed for the three resolutions of this IC code comparison. The input power spectrum normalization is recomputed by ZELDOVICH-PLT then scaled back to the requested redshift using the growth factor ratio from ABACUS's cosmology module. Modes are filled out to the Nyquist sphere. This is the initial conditions generator used by the ABACUS code, including the AbacusSummit project[4] (Maksimova et al., 2021).

IC_GEN / PANPHASIA

There are two components to this method of creating the initial conditions which has been labelled PANPHASIA. Firstly PANPHASIA describes a method of setting the phases of cosmological initial conditions. PANPHASIA describes a pseudo random number sequence mapped to a cosmological volume from which subsections can be taken to provide the phases for initial conditions (Jenkins, 2013). This allows simple creation of resimulation initial conditions and also for the phases used in an N-body simulation to be shared so that others can recreate the same cosmological volume at any resolution without having to share the full phase information.

Once the phases have been generated using PANPHASIA the initial conditions are generated using a proprietary code called IC_GEN. This uses second order Lagrangian perturbation theory (2LPT) (Jenkins, 2010) to create the initial conditions particle distribution

[3] https://github.com/abacusorg/zeldovich-PLT
[4] https://abacussummit.readthedocs.io

according to the phases specified by the location within PANPHASIA.

FastPM IC

While not originally designed as an initial condition code, FastPM does compute an initial particle state with the traditional 2LPT-on-a-mesh approach. In addition, the initial condition code in FastPM supports producing constrained Gaussian realizations.

Particles are first placed on a uniform grid with N_c particles per side, then the first and second order LPT displacement ($\mathbf{s}_{1,2}$) and velocity terms ($\mathbf{v}_{1,2}$) are calculated from a potential induced from a Gaussian white noise field $g(\mathbf{x})$ according to an isotropic linear power spectrum $P_{\text{lin}}(k)$. The white noise field is sampled on a finite mesh of $B_{\text{LPT}}N_c$ samples per side from a pseudo-random generator compatible to the one used by Gadget's N-GenIC(Angulo et al., 2012; Springel et al., 2005).

The choice of B_{LPT}, the finite differentiation operators, and the finite Laplace operator all affect the initial particle state, even though the effect on the late time non-linear field can be quite minimal.

3.2.2 Comparison of Initial Conditions

In this section, we compare the initial conditions in terms of particle and velocity distributions.

Comparison of Particle Distribution

The power spectrum is a common measure for analysing both initial conditions and evolved N-body simulations. It measures the strength of the matter density contrast at different scales.

The power spectrum, $P(k)$ is defined as the Fourier transform of the autocorrelation function, $\xi(r)$ which in turn is derived from the matter overdensity, $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1$,

$$P(k) = \int d^3\mathbf{r}\; e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(r), \tag{3.1}$$

$$\xi(r) = \langle\delta(\mathbf{x})\delta(\mathbf{x}-\mathbf{r})\rangle = \frac{1}{V}\int d^3\mathbf{x}\;\delta(\mathbf{x})\delta(\mathbf{x}-\mathbf{r}). \tag{3.2}$$

Initial conditions sample the matter density field producing a set of particles. This limited number of particles means that there is a maximum k above which the power spectrum cannot be represented in the initial conditions. This is the Nyquist frequency. For a box with sidelength L and N^3 particles the wavenumber associated with the Nyquist frequency is

$$k_{\text{Nyquist}} = \frac{\pi N}{L}. \tag{3.3}$$

The Nyquist wavenumber, and therefore the Nyquist frequency, increases with number of particles allowing us to compare to the reference power spectrum at higher k.

The power spectra are measured using NBODYKIT [5] (Hand et al., 2018) using a triangular shaped cloud mass assignment scheme and grid sizes of 2592^3, 3456^3 5184^3 were used for the low, medium and high resolution ICs respectively. Differences near the Nyquist frequency are observed when measuring the power spectra with different grid sizes and aliasing is observed when using a grid size which does not evenly divide the particle distribution. Using large grid sizes which are multiples of the particle numbers mitigates these effects. The power spectra are truncated at the Nyquist frequency for each resolution.

Fig. 3.1 shows the power spectra measured for all ICs codes and also for different resolutions, compared to the reference power spectrum. The reference power spectrum is the theoretical linear power spectrum from CAMB at redshift, $z = 199$. The measured power spectra differ from the reference power spectrum at both large (low k) and small scales (high k). The large scale differences are mainly due to sample variance and the limited number of large scale modes in a particular realisation may lead to disagreement in the power spectrum measurement from the reference P_k at these scales. The grey envelope shows the theory sample variance. This variance decreases as k^{-1} which is consistent with the theory expectation using Poisson noise where the number of modes is proportional to k^2.

There is only one distinct line for each code at low k, this is because each code uses a different phase realization which is consistent between different resolutions. At small scales, close to the Nyquist frequency, the systematic resolution effects are different

[5] https://nbodykit.readthedocs.io/en/latest/

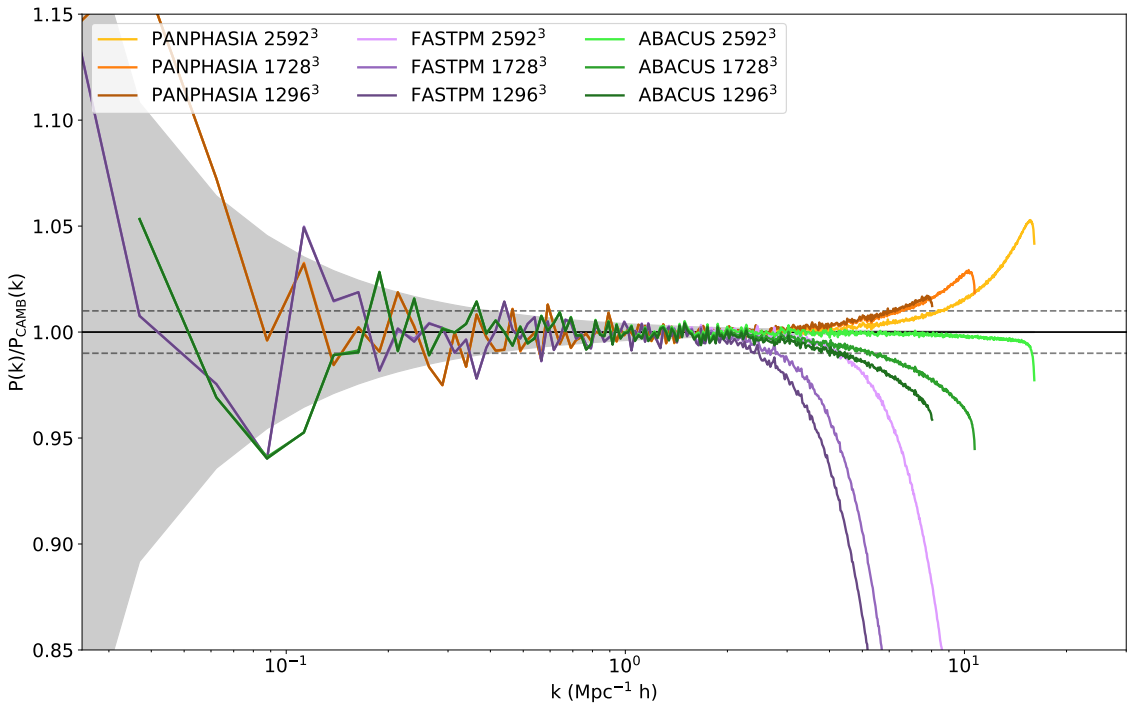

Figure 3.1: A comparison of the power spectrum of initial conditions created at redshift $z = 199$, using different codes and resolutions. The power spectra are plotted as a ratio to the input theoretical power spectrum from CAMB. At high k differences emerge between the input and the measured power spectra. At low k, the sample variance causes noise in the measured power spectrum. The grey envelope shows the theory 1-sigma sample variance.The dashed lines indicate $\pm 1\%$.

between codes. The ICs created using PANPHASIA do not disagree with the reference power spectrum by more than 5% in all resolutions up to the Nyquist frequency. The FASTPM ICs have much lower power than the reference power spectrum for $k > k_{\rm Nyquist}/2$. The ABACUS ICs have the best agreement with the reference power spectrum.

It should be noted that it is not imperative for initial conditions to exactly fit the reference power spectrum at the smallest scales in order to create a realistic N-body simulation. The majority of the power at low redshift at small scales is due to collapse of larger modes and therefore small differences in the power spectrum around $k_{\rm Nyquist}$ are mostly unimportant to the growth of structure at these scales (Neyrinck & Yang, 2013)

Comparison of Velocity statistics

The power spectrum is only a measure of the displacements of the particles produced by an initial conditions code. The particle velocities are just as important as the displacements

and affect the rate of growth of structure as the simulation progresses. In methods of generating the initial conditions, such as the Zel'dovich approximation (Zel'Dovich, 1970) and 2LPT (Crocce et al., 2006), the particle velocities are not free to vary but are fixed by the particle displacements. For example in the Zel'dovich approximation a particle's velocity is in the same direction as, and proportional to, the displacement from the particle's gridpoint. For this comparison project we thought it prudent to check the consistency between particle velocities from different IC codes. We investigated the pairwise velocity dispersion between the codes and also compared to 10 realisations of PANPHASIA ICs. There was no theoretical estimate generated therefore we aimed to verify that there were similar results between the codes with the differences being consistent with the sample variation from the 10 PANPHASIA realisations.

The pairwise velocity dispersions are calculated by taking the pairwise velocity between pairs of particles separated by a distance r along with the component of the velocity parallel and perpendicular to the vector between the particles.

For particles with positions and velocities $(\mathbf{x_1}, \mathbf{v_1})$ and $(\mathbf{x_2}, \mathbf{v_2})$ the parallel and perpendicular components are

$$v_{\parallel} = \frac{(\mathbf{x_1} - \mathbf{x_2}) \cdot (\mathbf{v_1} - \mathbf{v_2})}{|\mathbf{x_1} - \mathbf{x_2}|} \tag{3.4}$$

and

$$v_{\perp} = \sqrt{|\mathbf{v_1} - \mathbf{v_2}|^2 - v_{\parallel}^2} \tag{3.5}$$

respectively. This gives velocity dispersions in the parallel and perpendicular directions of

$$\sigma_{\parallel}^2(r) = \langle v_{\parallel}^2(r) \rangle - \langle v_{\parallel}(r) \rangle^2 \tag{3.6}$$

and

$$\sigma_{\perp}^2(r) = \langle v_{\perp}^2(r) \rangle - \langle v_{\perp}(r) \rangle^2. \tag{3.7}$$

The isotropy of the universe implies that the second term of equation 3.7 is equal to zero.

The velocity anisotropy parameter

$$\beta = 1 - \frac{\sigma_{\perp}^2}{2\sigma_{\parallel}^2} \tag{3.8}$$

measures the relative size of the parallel and perpendicular velocity dispersions.

Fig. 3.2 shows that the velocity anisotropy, as a function of pairwise inter-particle distance, is mostly consistent between the codes and at large scales the differences are mostly explained by sample variance. There is excellent agreement between the codes at small inter-particle distance. FastPM is the only outlier, having a slightly higher anisotropy than the other realisations at around $r = 150\ h^{-1}$Mpc.

It was decided to use PANPHASIA initial conditions for the N-body code comparison project in the next section. This was because PANPHASIA had reasonable agreement with the reference power spectrum close to the Nyquist frequency in all resolutions and PANPHASIA has the ability to easily share phase information to be used by other initial conditions codes in the future. These tests do not show whether the IC generators are fully accurate in absolute terms, and the impact of IC differences on evolved simulations is not measured. However these results show that the PANPHASIA ICs are a reasonable choice to use for the code comparison project when compared to the other options.

3.3 N-body Codes

This section details the N-body code comparison project. We run ABACUS, SWIFT, GADGET-2, and FASTPM from $z = 199$ to $z = 0$ at the three resolutions shown in Table 3.1. Comparisons between the codes are made in the matter power spectrum and in the dark matter halo properties, along with comparisons of the halo clustering.

The initial conditions for this code comparison project were created using PANPHASIA and IC_GEN from the previous section; all codes used identical initial particles. The cosmology used is described at the start of Section 3.2 and the resolutions of the simulation boxes are shown in Table 3.1.[6].

Codes were run by separate groups and the code parameters were decided upon by those groups.

[6]The Panphasia descriptor describing the simulation volume is [Panph1,L21,(1133107,236124,673886),S81,CH1311607586,DESI_IC_v1]

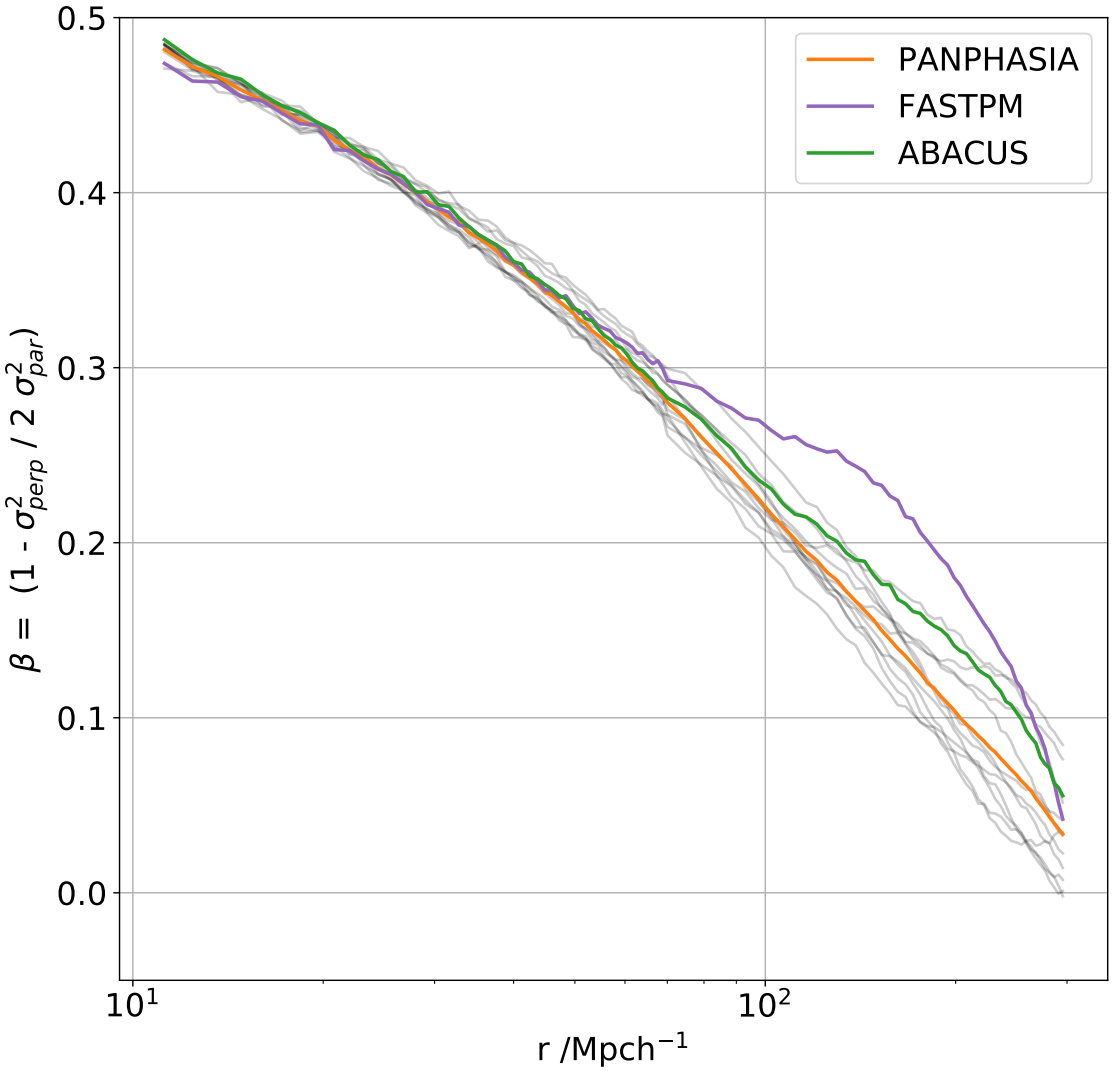

Figure 3.2: A plot showing the velocity anisotropy as a function of interparticle distance. The coloured lines show the different codes meanwhile the grey lines show the results for 10 PANPHASIA realisations to investigate sample variance. The codes agree well at small distances while at large r the differences are dominated by sample variance. The only outlier is FastPM which has a higher anisotropy than any of the other realisations around 150 h^{-1}Mpc.

3.3.1 Code details

ABACUS

ABACUS is a high force accuracy N-body code that solves the far-field force with a high-order multipole method and the near field with direct summation on GPUs (Garrison et al., 2021b). Typical RMS force accuracy is 10^{-5} to 10^{-6}. All particles in ABACUS share a single, global time step, which is chosen at the beginning of each time step. ABACUS was recently used to run the AbacusSummit simulations (Maksimova et al., 2021), and the ABACUS realization of the Euclid code comparison simulation was presented in (Garrison et al., 2019).

The numerical parameters used in this code comparison are the same as those used in AbacusSummit. Specifically, spline force softening, fixed in proper coordinates, was used such that the $z = 0$ value was 1/40th of the interparticle spacing (with the early time softening capped to 0.3 in the same units). This choice was tested using scale-free simulations in (Garrison et al., 2021a). The multipole order was $p = 8$, with 405^3 cells, and the time step parameter was set to $\eta_{\rm Accel} = 0.25$. This is mildly more conservative than ABACUS's run of the Euclid code comparison simulation, which used $\eta_{\rm Accel} = 0.3$. (Garrison et al., 2021b) show that a time step value of $\eta = 0.25$ produces sub-percent shifts in the two-point clustering for all scales larger than twice the softening length.

GADGET-2

GADGET-2(Springel, 2005) combines a tree and a particle mesh for its gravity scheme. The first order moments are used to calculate the tree forces and a fixed opening angle criterion is used to decide when to divide up the tree cells. The particle positions and velocities are updated in a leapfrog scheme and the global timesteps are evenly spaced in $\log(a)$ with dynamic shorter timesteps for particles undergoing large accelerations. A mesh size of N^3 was used for the simulation with N^3 particles. The comoving Plummer equivalent softening length used was 1/25th of the interparticle spacing. The parameters used to define the timestepping and force accuracy were MaxRMSDisplacementFac = 0.2, MaxSizeTimestep = 0.01, ErrTolTheta = 0.5, and ErrTolForceAcc = 0.0025.

SWIFT

SWIFT[7](Schaller et al., 2016) is a hydrodynamics and gravity code which uses a task based parallelisation strategy. It is primarily designed for high resolution SPH simulations of galaxy formation, however SWIFT can also be used for dark matter only simulations of large scale structure. The gravity scheme in SWIFT has three levels. At large scales a Fourier mesh is used to calculate the force on particles and account for the periodicity of the simulation box; at small scales the fast multipole method (FMM)(Dehnen, 2014) is used to calculate forces between particles in octree cells; at the smallest scales direct summation is used. A fixed opening angle criterion is used to calculate the size of the octree cells used in the FMM. A time dependent force softening is applied to the particles, the forces are softened according to a spline kernel. The global timesteps are evenly spaced in $\log(a)$ and adaptive timestepping is used for particles with large accelerations in the same way as in GADGET-2 (Theuns et al., 2015)[8]. A Fourier mesh size of 648^3 was used in the 1296^3 particle and 2592^3 particle simulations and 864^3 was used in the 1728^3 particle simulation. These values were chosen as they were the largest allowed by the code which would not introduce aliasing with the initial particle grid. The Plummer equivalent comoving softening length was 1/25th of the interparticle spacing.

FASTPM

FASTPM[9](Feng et al., 2016) is an approximated particle mesh N-body solver. By modifying the kick and drift factors in time integration, it enforces correct linear displacement evolution on large scales, and meanwhile reduces the halo stochasticity when the number of time steps is small (Feng et al., 2016). The cost is extremely low thanks to the small number of time steps required for satisfying statistics. The accuracy can be mainly controlled by two options, B, the mesh size over particle number per dimension in the force calculation, and the time steps usually specified by the number of steps T and the spacing scheme. In this comparison project, we use $B = 2$ and $T = 46$, with linear spacing in the scale factor. FastPM simulations are useful for estimating covariance matrices

[7] http://swift.dur.ac.uk/
[8] https://gitlab.cosma.dur.ac.uk/swift/swiftsim
[9] https://github.com/rainwoodman/fastpm

and other applications for which the accuracy is not a critical factor. A summary of the FastPM simulations prepared by the DESI Cosmological Simulations working group has been presented in (Ding et al., 2021).

3.3.2 Comparison of particle distribution

We first begin by examining the power spectrum dependence on the mass resolution for each code separately. As seen in Fig 3.3, this dependence is similar for all the codes tested. There is very little change in the power spectra at large scales and at small scales there are differences of 1-2%. The higher the mass resolution, the higher the power at small scales (except with FASTPM).

We then proceed to compare the power spectrum results among codes. Fig. 3.4 shows all of the power spectra relative to the ABACUS runs. There is broadly good agreement between all of the codes except FASTPM with differences below 1% at $k = 2\ h\ \mathrm{Mpc}^{-1}$. This agrees with the results found in other code comparison projects (Schneider et al., 2016; Garrison et al., 2019). We do not see the discrepancy between GADGET-2 and linear theory at large scales as seen in (Schneider et al., 2016), although this work employs GADGET-2 and (Schneider et al., 2016) employed GADGET-3. Apart from FASTPM, at $k = 10\ h\ \mathrm{Mpc}^{-1}$ the fractional differences between the power spectra are below 1%.

All codes agree on the low-k power spectrum amplitude to within 0.1%, but to ascertain that this value is that predicted by perturbation theory, the 2LPT estimates of the $z = 1$ and $z = 2$ power spectra were measured by using IC_GEN, the code used to create the initial conditions for the simulations. Using a low output redshift for the same phases gives us the estimate of the matter distribution at low redshift from which the power spectrum can be measured. The results of this test are shown in Figure 3.5. We also attempted to compare to linear theory by rescaling the power spectrum of the initial conditions by the linear growth factor, however this did not reproduce the large scale power spectrum to sub-percent accuracy, as shown in the dashed lines in Figure 3.5. The reason for this difference is likely to be mode coupling. In addition, using the Zel'dovich approximation instead of 2LPT produced a curve which was less smooth and therefore we used 2LPT for the low redshift perturbation theory power spectrum. Perturbation theory is only accurate for small density perturbations which corresponds to very large scales at low redshift. We

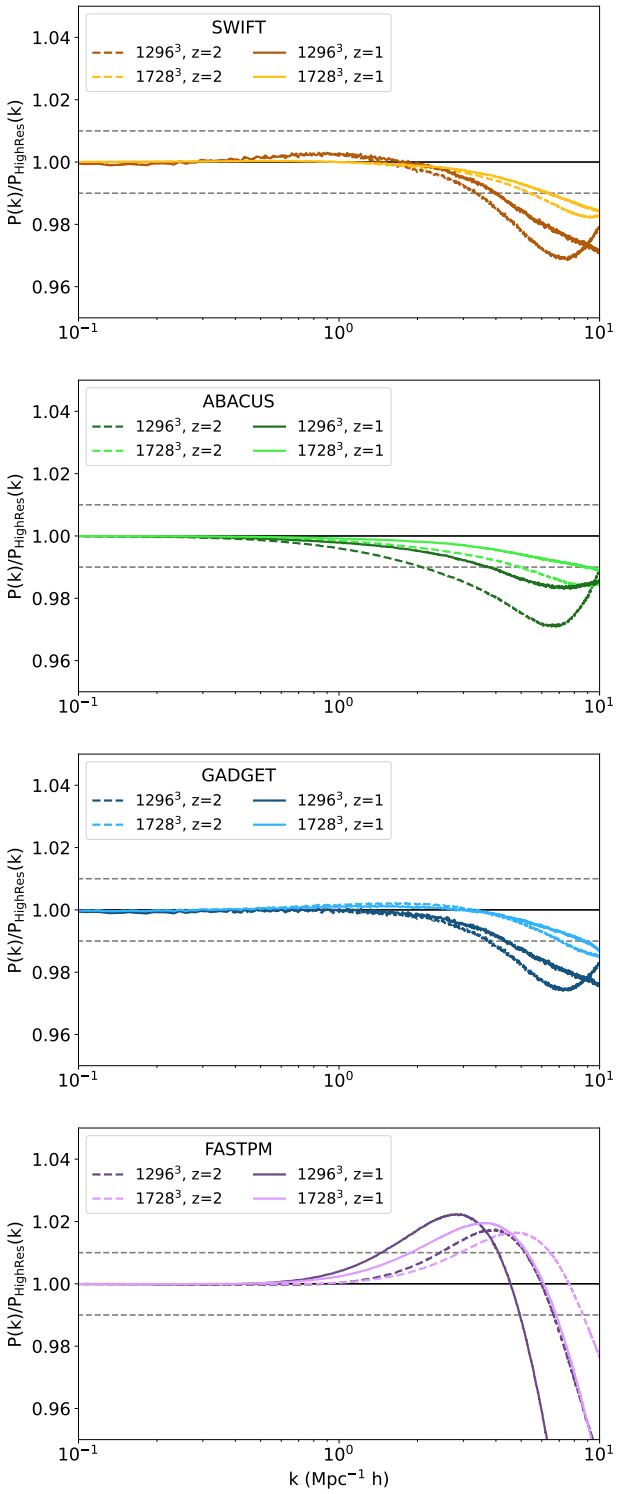

Figure 3.3: The matter power spectrum of the low and medium resolution runs relative to that of the high resolution run for each code at $z = 1$ and 2. SWIFT, ABACUS, & GADGET-2 show similar behaviour where the power spectrum decreases by up to 2% at $k = 10\ h$ Mpc^{-1} when decreasing the resolution from 2592^3 to 1728^3 particles or from 1728^3 to 1296^3 particles. The lower resolution FASTPM power spectra show an increase relative to the high resolution case at $k = 3\ h$ Mpc^{-1} before decreasing sharply above $k = 5\ h$ Mpc^{-1}.

have found that the simulation power spectra and the 2LPT power spectra asymptotically converge at low k for all the codes. Furthermore, this exercise helped identify a previously unknown error in the large-scale growth in the SWIFT code; a difference of around 0.5% in the large scale power spectrum which was caused by incorrect force integration. The error was corrected, and all results in this work use the modified code.

3.4 Dark Matter Halo Comparison

Much of the value which comes from running large N-body simulations does not come from the complete distribution of dark matter particles, but instead comes from grouping the particles into overdense regions called dark matter halos because these are the sites that host galaxies (Peacock & Smith, 2000; Benson et al., 2000; Conroy & Wechsler, 2009). Using and storing full particle data can become prohibitively expensive for large N-body simulations therefore halo catalogues are often the most useful data product to be produced from an N-body simulation. Observations and simulations suggest galaxies reside within dark matter halos and therefore they are essential to connect N-body simulations to observable results (Wechsler & Tinker, 2018).

Halo catalogues can be used to create mock galaxy catalogues which provide useful information on the expected observational results for different cosmologies, along with providing mock data which is essential for testing analysis pipelines, e.g. see (Cole et al., 1998; Rodríguez-Torres et al., 2016; DeRose et al., 2021).

In this section we perform the comparison of halo properties, mass functions and clustering. Halos are identified with the help of the phase-space temporal halo finder ROCKSTAR with force resolution parameters shown in Table 3.2 (Behroozi et al., 2013)[10], and default parameters otherwise, including `STRICT_SO_MASSES=0`. ROCKSTAR finds halos by using a friends-of-friends algorithm in six phase-space dimensions (three spatial and three velocity dimensions). Groups and subgroups are found using an adaptive hierarchical method which progressively reduces the linking length between particles. Setting `STRICT_SO_MASSES=0` affects the mass definition of halos. The halo mass in this case is defined only using particles within the friends-of-friends group, rather than also

[10]https://bitbucket.org/gfcstanford/rockstar

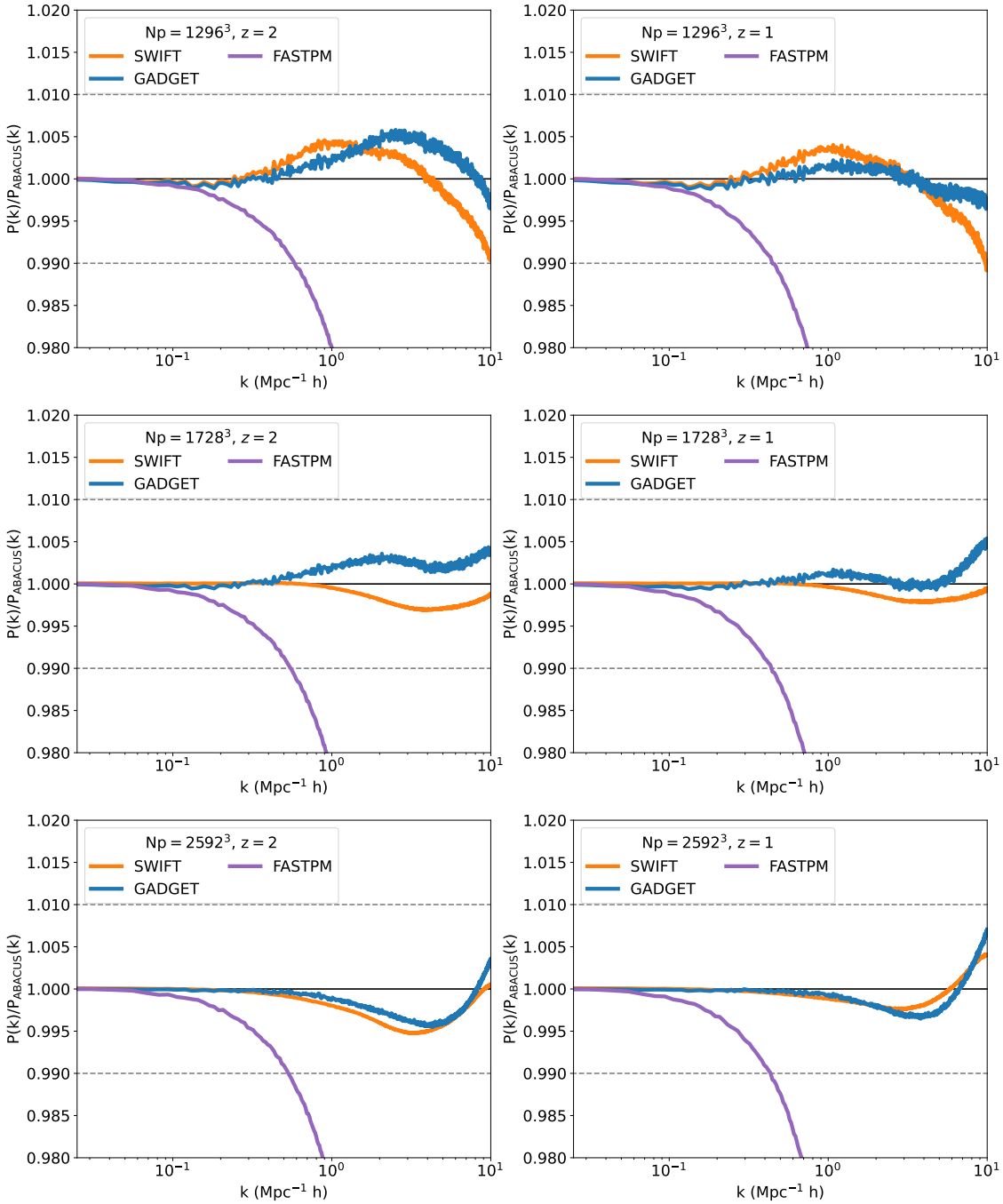

Figure 3.4: A comparison of the matter power spectrum between the different codes at $z = 1$ and $z = 2$ in all three resolutions. The power spectrum is plotted relative to ABACUS at each redshift and resolution. All the codes agree to within 0.1% at $k < 0.1\ h$ Mpc^{-1}. Differences emerge at high k but, excepting FASTPM, these are within 1% at $k = 10\ h$Mpc^{-1}.

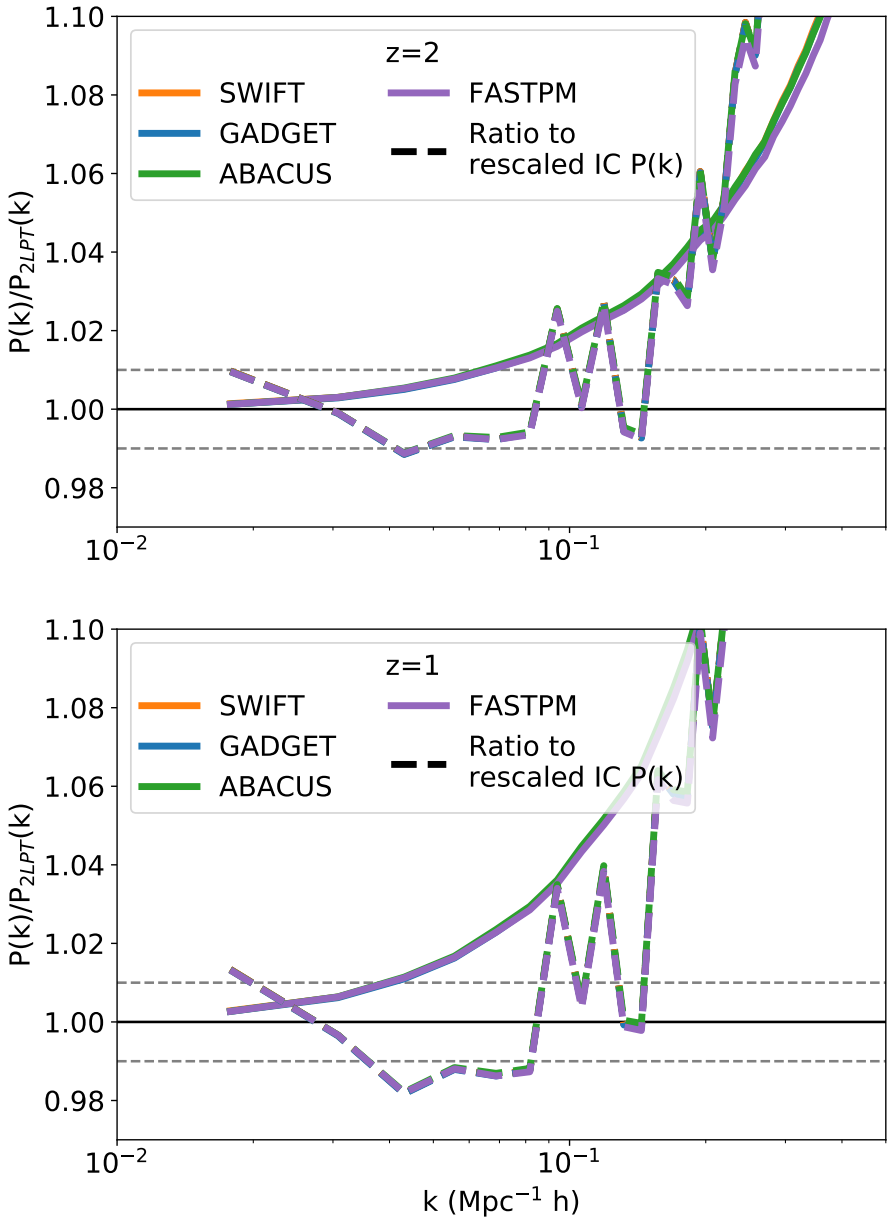

Figure 3.5: A comparison of the power spectra to perturbation theory at $z = 1$ and $z = 2$. The perturbation theory power spectrum was found using second order perturbation theory (2LPT). The power spectra should asymptotically agree with 2LPT at low k. Coloured dashed lines show the ratio of the power spectra with initial conditions rescaled by the growth factor which do not agree as closely as 2LPT.

Table 3.2: ROCKSTAR force resolution: this table shows the force resolution of the simulations in comoving h^{-1}Mpc, as input to the ROCKSTAR halo finder. Halos whose centres are closer than the force resolution are considered to be unresolved. The minimum number of particles considered to be a halo seed for 1296^3, 1728^3 and 2592^3 resolutions are 5, 10 and 10 respectively.

Simulation	ABACUS	GADGET-2	SWIFT
1296^3	0.015	0.015	0.015
1728^3	0.0072	0.011	0.011
2592^3	0.0048	0.0077	0.0077

including particles which exist outside the group but inside the spherical overdensity.

We do not include FASTPM in this comparison of halo properties and clustering. This is because performing abundance matching is outside the scope of this work, and the FASTPM results without abundance matching would be unrealistically discrepant compared with how the code is used in practice.

Our tests involving dark matter halos are performed at $z = 1$. We do not observe significant redshift evolution in our code or resolution comparisons when measuring halo mass functions and halo clustering at $z = 0$ and $z = 2$.

3.4.1 Halo properties

We execute a detailed halo-to-halo comparison by matching halos between N-body simulations and comparing their properties. There is no guarantee that halos will form in identical places, but since all simulations share the same initial particles, in many cases a robust match can be identified. Still, matching halos provides a challenge as there is no unique identifier for each halo which is well defined across several different simulation runs.

We use a nearest neighbour approach to match halos between snapshots from simulations run with different codes and resolutions. Each halo in one catalogue was matched to the halo with the nearest position in the other catalogue. All halos with a mass smaller than 5×10^{10} $M_\odot/h$ at all resolutions were removed to avoid spurious matches and an upper distance limit of $0.25\,h^{-1}$Mpc was applied for halos to be considered as matched. The matched halo properties were binned by halo mass.

Figure 3.6 shows the fraction of unmatched halos in each catalogue in a comparison

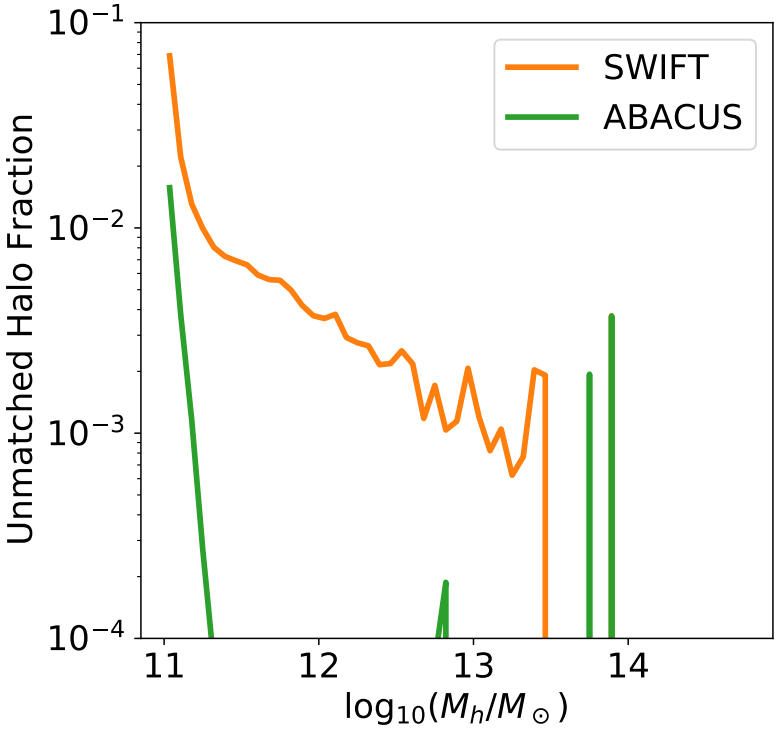

Figure 3.6: The fraction of halos which remain unmatched as a function of halo mass in the ABACUS and SWIFT catalogues when matching halos between the high resolution catalogues. The majority of halos are matched across simulations.

between the high resolution SWIFT and ABACUS simulations. This fraction increases as halo mass decreases but is only greater than 5% of halos for the lowest mass SWIFT halos at $M_h = 10^{11}\ h^{-1}\mathrm{M}_\odot$. This gives us confidence that there are not systematic effects being introduced by the nature of the halo matching algorithm.

The halo properties which are compared between matched catalogues are defined as follows. Halo separation is the distance between matched halos in units of h^{-1}Mpc, as discussed above this cannot be greater than $0.25\,h^{-1}$Mpc for any matched pair. M_{200b} is defined as the mass contained within a sphere around the halo such that the sphere has a density of 200 times the background density of the universe. r_{vir} is the radius of the halo which is similarly defined in terms of a spherical overdensity. In this case the overdensity is Δ_c times the critical density of the universe. Δ_c is the solution for a virialised cluster which is $18\pi^2$ for a critical universe but can vary with redshift otherwise. M_{vir} is the mass contained within r_{vir} (Peebles, 1980; Eke et al., 1996; Bryan & Norman, 1998). r_s is the NFW scale radius of the halo, defined in terms of the density profile (Navarro et al., 1997). σ_{v} is the halo particle velocity dispersion in physical km s^{-1}. $|v|$ is the speed of

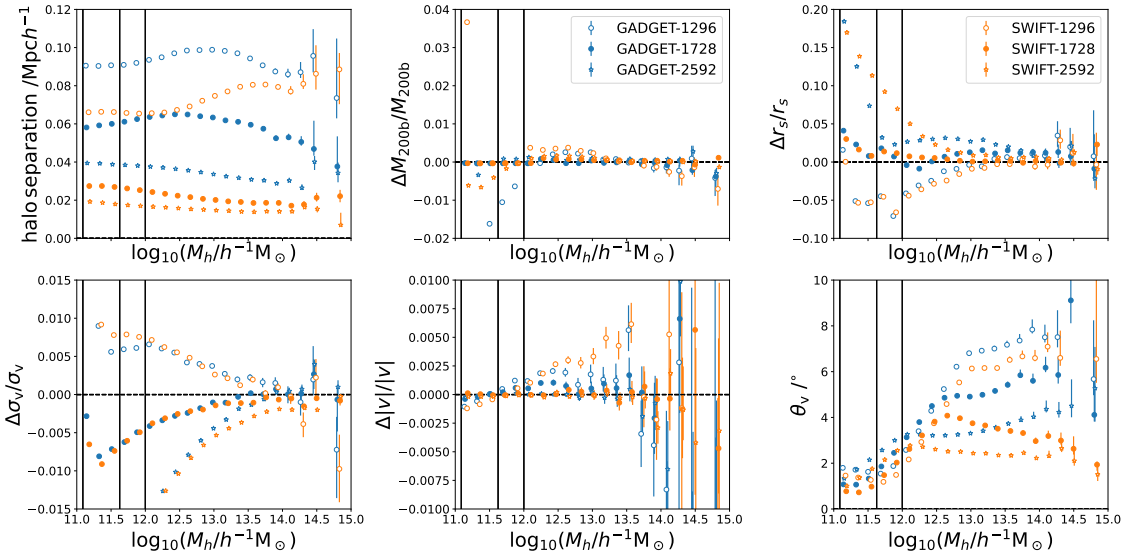

Figure 3.7: A comparison of the halo properties for halos matched between ABACUS and the other codes at fixed resolution. On the top row from left to right the panels show: 1) The distance between matched halo centres in h^{-1}Mpc. 2) The fractional difference in halo mass, as measured within a sphere of density 200 times the background density. The fractional difference in virial mass is identical to this panel, the fractional difference in virial radius is also identical to this panel but multiplied by a factor of three. 3) the fractional difference in the scale radius of matched halos. On the lower row, from left to right, the panels show: 1) The fractional difference in the velocity dispersion of matched halos. 2) The fractional difference in the magnitude of the velocities of the matched halos. 3) The difference in the direction of the velocities of matched halos in degrees. Hollow circles, filled circles, and stars represent the comparison for the 1296^3, 1728^3, and 2592^3 simulations respectively. The results are colour coded differently for each simulation code. GADGET-2: blue, SWIFT: orange, and ABACUS is used as the reference simulation. Vertical solid lines indicate the halo mass at which halos contain 200 particles in the high, medium, and low resolution simulations moving from left to right.

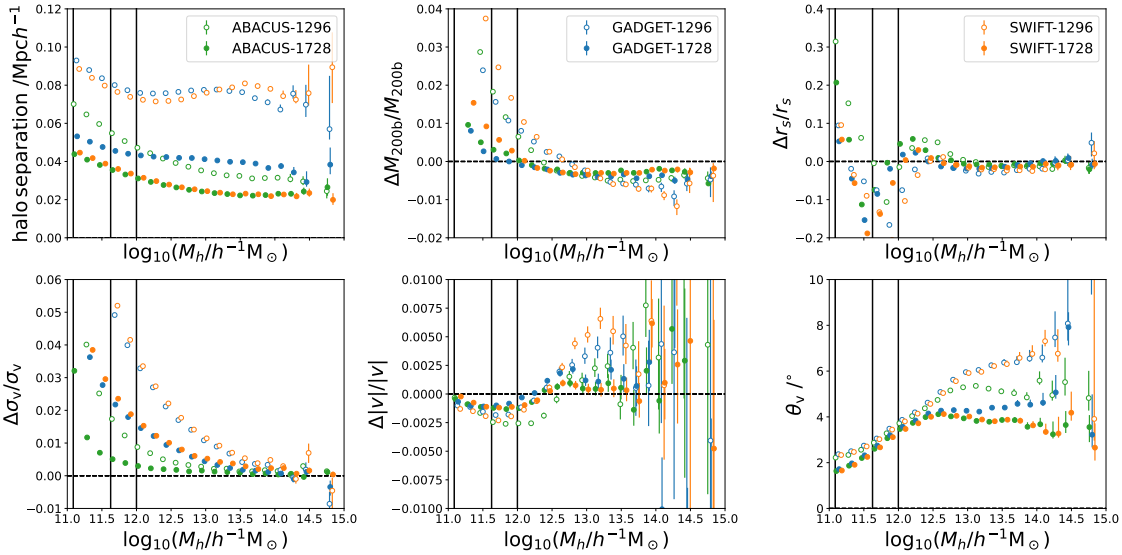

Figure 3.8: As for Fig. 3.7 but for halos matched to the high resolution simulation for each code. Note the change in range plotted in the upper-right and lower-left panels.

the halo in physical km s^{-1}. Finally θ_v is the difference in angle of the velocity vectors of the matched halos in degrees. With the exception of halo separation, $|v|$, and θ_v, which we compute ourselves, all of these properties are computed by ROCKSTAR.

Fig. 3.7 shows how the matched halo properties vary between ABACUS and the other codes at fixed resolution. The low, medium, and high resolution cases are shown as hollow circles, solid circles and stars respectively along with error bars which show the error in the mean of the matched property within each mass bin.

In general, the matched properties show greater agreement at high masses and at high resolution. This is where the halos contain a greater number of particles and therefore it is to be expected that these halos are less susceptible to variation due to differences between the codes which are primarily in the small scale forces. The halo separation is roughly constant across halo mass but varies for different codes and resolutions. The mean halo separation varies between 10 and 100 h^{-1}kpc; SWIFT, and ABACUS agree to within three softening lengths at high and medium resolution, while GADGET-2 has two to three times worse agreement.

SWIFT and GADGET show agreement with ABACUS to within 1% in the halo mass and radius at medium and high resolution for all masses. At low resolution differences emerge for halo masses below 10^{12} $h^{-1}\mathrm{M}_\odot$.

r_s shows a similar pattern of better agreement at higher halo masses. The differences at masses below 10^{12} $h^{-1}\mathrm{M}_{\odot}$ are at the 10% level and do not show clear trends with code or resolution, except that SWIFT and GADGET show progressively lower Δr_s at lower resolutions.

As resolution decreases, SWIFT and GADGET-2 have increasing $\Delta\sigma_{\mathrm{v}}$ relative to ABACUS.

The halo velocities show larger differences in both magnitude and direction between the codes for high mass halos. There are large uncertainties on these measurements due to the limited number of high mass halos within each catalogue.

Fig. 3.8 shows how the matched halo parameters vary with simulation resolution when keeping the simulation code fixed.

There is a clearer trend in the halo separation with halo mass when comparing with each code fixed, the separation is lower for higher mass halos and is also lower at higher resolution. SWIFT and GADGET-2 show larger halo separation at different resolutions compared to ABACUS.

The halo masses are systematically smaller by around 0.5% for all codes and resolutions at halo masses above 10^{13} $h^{-1}\mathrm{M}_{\odot}$. The same feature is seen in the virial radii where the difference is 0.2%. For lower mass halos the opposite effect is seen, halo masses are larger for the lower resolution simulations and this difference grows when going to lower halo masses and lower resolutions. This indicates that at lower resolutions mass is systematically moved from high mass halos into low mass halos when compared to higher resolution simulations. The origin of this effect could be that the halo finder is more likely to split off halos at low resolutions due to a lower number of particles with which to make friends-of-friends connections.

r_s is consistent between high mass halos at all resolutions for all the codes. Below 10^{12} $h^{-1}\mathrm{M}_{\odot}$, r_s becomes poorly matched as resolution is varied.

The velocity dispersion, σ_{v}, increases for lower mass halos at lower resolutions. This effect can be seen for all the codes but is most strongly present in SWIFT, and GADGET-2.

In summary, the halo properties agree well between the different codes at fixed resolution, there is better agreement for halos with mass greater than 10^{13} $h^{-1}\mathrm{M}_{\odot}$ than lower mass halos.

Lower simulation resolution systematically biases high mass halos to have slightly

lower mass and radius than those matched from higher resolution simulations. The opposite effect is observed at low masses. In general high mass halos show greater agreement in the matched halo parameters than low mass halos.

3.4.2 Halo Mass Function

The halo mass function (HMF) describes the number density of dark matter halos in different mass intervals. Here we have defined the HMF as $\mathrm{d}n/\mathrm{dlog}_{10}(M_{\rm vir})$ where n is the number density of halos in units of $\mathrm{Mpc}^{-3}h^3$ and $M_{\rm vir}$ is in units of $h^{-1}\mathrm{M}_\odot$.

The HMF has a strong effect on the number density and clustering of galaxy catalogues created using the halo catalogue, particularly at fixed mass. It is therefore important for the HMF to not change significantly between simulations run with different codes, as this could indicate underlying systematic errors.

In addition we investigate the effect of simulation resolution on the HMF. Mass resolution has the greatest effect at the low mass end of the HMF where halos contain fewer particles. It is important to find the cutoff mass below which the HMF shows significant differences due to simulation resolution.

Figure 3.9 shows the HMFs measured from the different codes at fixed resolution at $z = 1$, relative to the ABACUS HMF. The errorbars represent the estimated standard deviation in the HMF difference between phase matched simulations which are calculated using a jackknife method, see equation 3.14 in Section 3.4.3 for more details. Differences between the codes are greater at high and low masses. The HMFs agree at the 1% level between 10^{12} $h^{-1}\mathrm{M}_\odot$ and $10^{13.5}$ h^{-1} $\mathrm{M}_\odot$ for all resolutions. The HMFs do not differ by greater than 10% at masses up to $10^{14.5}$ h^{-1} $\mathrm{M}_\odot$, these differences are consistent with noise and not biased in a particular direction. At low masses there is different behaviour between the codes at each resolution. There is better agreement on the low mass end of the HMF at low resolution than high resolution.

Fig. 3.10 shows the change in HMF with resolution for the ABACUS code at $z = 1$. The effect of changing resolution is larger than changing code. The halo catalogues from different resolution simulations have similar HMFs for high mass halos. The medium resolution simulation HMF differs from the high resolution simulation HMF by more than 1% below 10^{12} h^{-1} $\mathrm{M}_\odot$. The low resolution HMF differs by a greater amount than

the medium resolution HMF, with greater than 1% disagreement below $10^{12.5}\ h^{-1}\ \mathrm{M}_\odot$. Testing the same comparison with a halo catalogue produced from a subsampled version of the high resolution simulation produces a similar relationship which indicates that discreteness effects in the halo finder are not responsible for this systematic difference.

3.4.3 Halo Clustering: Correlation Function

The clustering of dark matter halos is one of the most important statistics to be produced in an N-body simulation. Galaxy clustering is closely tied to halo clustering, especially at large scales, therefore ensuring accurate and robust halo clustering estimates is essential in order to have confidence in mock data produced from simulations.

In this section we compare the halo clustering from simulations run with different codes and at different resolutions. We compare the clustering in both real-space and in multipoles of redshift space. Estimates of the DESI statistical error are made using a jackknife method, allowing us to provide length scales above which the halo clustering from our simulations is robust for DESI analysis.

The two-point correlation function for halos, $\xi(r)$ is defined as the excess probability of finding a halo at a distance r from another halo, averaged over all halos.

$$\xi(|\mathbf{r}|) = \langle \delta(\mathbf{x})\delta(\mathbf{x}-\mathbf{r})\rangle \tag{3.9}$$

The upper panel of Fig. 3.11 shows the real-space two-point correlation functions of the halo catalogues from the different codes relative to ABACUS. The high resolution simulations were used with a mass cut of $M > 10^{11.5}\ h^{-1}\ \mathrm{M}_\odot$ and $M > 10^{11}\ h^{-1}\ \mathrm{M}_\odot$ (solid and dashed lines respectively). The correlation functions from the different simulation codes do not differ by greater than 1% for length scales above 1 h^{-1}Mpc.

In the lower panel of Fig. 3.11 we compare the real-space two-point correlation function of the halos from ABACUS simulations at different resolutions. The high and medium resolution simulations agree to within 1% for all length scales above 1 h^{-1}Mpc. A larger difference is seen in the low resolution halo clustering, the clustering increases at smaller scales to 4% greater than the high resolution below 1 h^{-1}Mpc.

The halo catalogues from the low resolution simulations have a higher clustering bias

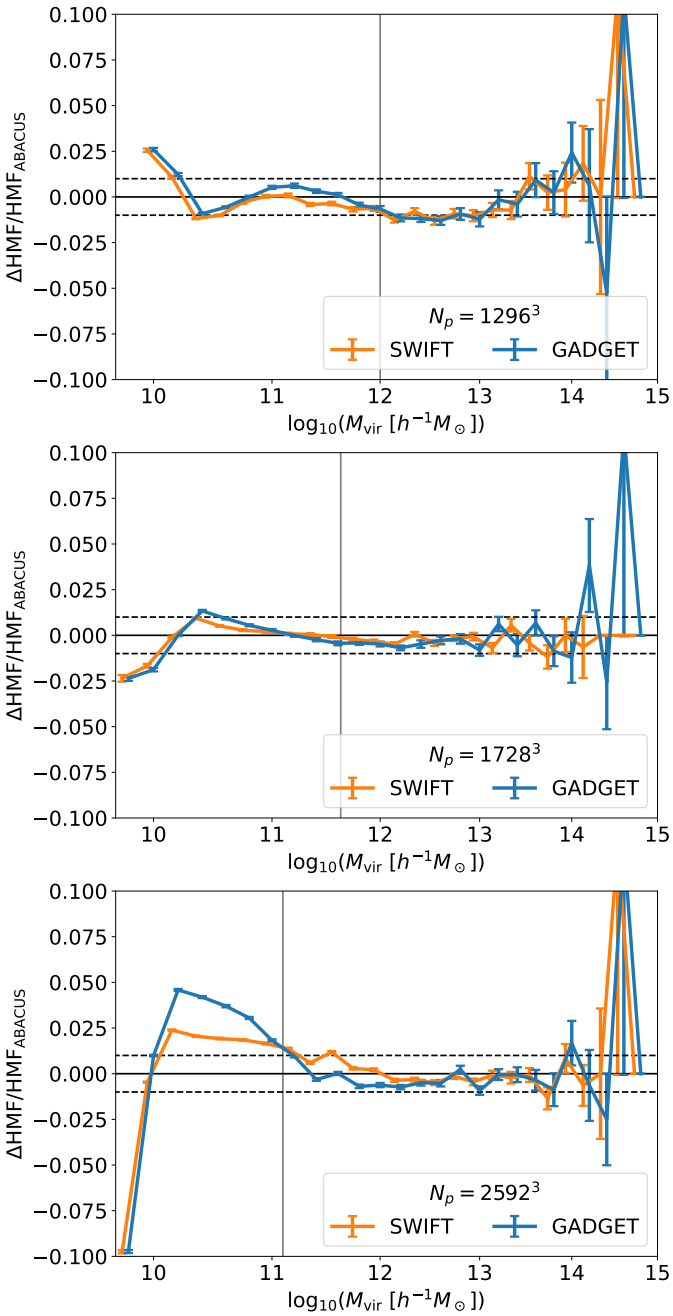

Figure 3.9: The ratio of the halo mass functions to that of the ABACUS simulation at fixed resolution at $z = 1$. The top, middle and bottom panels show the low, medium, and high resolution simulations respectively. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. Differences between the codes are below 1% between 10^{12} $h^{-1}\mathrm{M}_{\odot}$ and $10^{13.5}$ h^{-1} $\mathrm{M}_{\odot}$ at all resolutions. The differences in the HMF at large halo masses are consistent with noise. At low halo masses there is better agreement between the codes at low resolution than high resolution. Vertical lines indicate the mass at which halos contain 200 particles at each resolution.

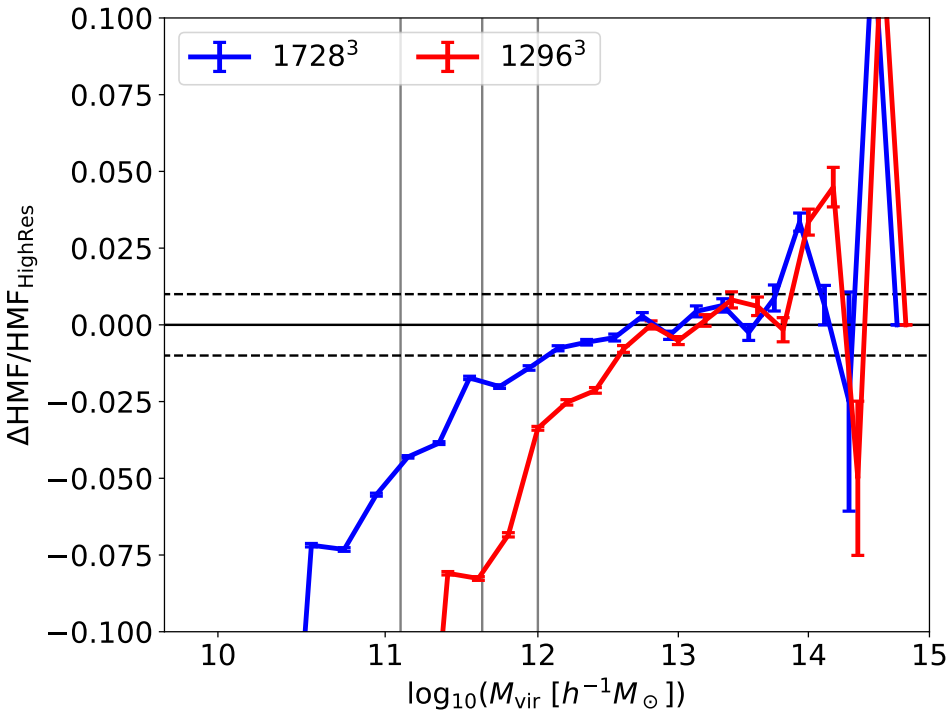

Figure 3.10: The ratio of the halo mass functions from the ABACUS simulations relative to the high resolution run at $z = 1$. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. The 1296^3 simulation shows greater differences than 1728^3 to the 2592^3 simulation. There is a deficit in low mass halos in the lower resolution simulations relative to the high resolution simulation. Vertical lines indicate the mass at which halos contain 200 particles at each resolution.

than the high resolution case. This effect can be understood with reference to the halo mass function, Fig. 3.10). There are fewer low mass halos in the low resolution case, therefore the clustering measurement is made from a sample which has proportionally more high mass halos which have a higher bias (Mo & White, 1996). Note that this clustering-resolution relation is in the opposite sense to that seen in (DeRose et al., 2019) which saw lower bias at lower mass resolution. Differences in the treatment of force softening at different resolutions may contribute to this difference.

Decreasing the halo mass cut to $M > 10^{11}\ h^{-1}\ \mathrm{M}_{\odot}$ can be seen to reduce the agreement in the clustering relative to the higher mass cut. This is seen most strongly in the low resolution simulations.

The multipoles of the redshift space correlation function, $\xi_l(s)$ are defined as

$$\xi_l(s) = \frac{2l+1}{2} \int_0^{\pi} \mathrm{d}\theta \sqrt{1-\mu^2}\xi(\sigma,\pi)\mathcal{L}_l(\mu) \tag{3.10}$$

(Peebles, 1980; Kaiser, 1987). Here l is the multipole order, θ is the angle between the line of sight and the halo separation vector, $\mu = \cos(\theta)$, $\xi(\sigma,\pi)$ is the 2d correlation function, $\mathcal{L}_l(\mu)$ is the l^{th} Legendre polynomial.

When making these comparisons of halo clustering, it is useful to understand the tolerance of observational measurements as this will provide a target accuracy. We estimate this observational tolerance by estimating the statistical error in the clustering using the jackknife method (Wu, 1986; Norberg et al., 2009).

The ABACUS high resolution box was used for jackknife estimates. This box was split into $N = 100$ distinct sub-volumes. The redshift space correlation function multipoles were measured on N sub-samples created by removing a single sub-volume each time to provide N clustering estimates. The standard error of the clustering can then be found to be:

$$\sigma(s) = \sqrt{\frac{N-1}{N}\sum_{i=0}^{N}\left(\xi^i(s) - \overline{\xi}(s)\right)^2} \tag{3.11}$$

where $\xi^i(s)$ is the i^{th} clustering estimate, produced by removing the i^{th} sub-volume and

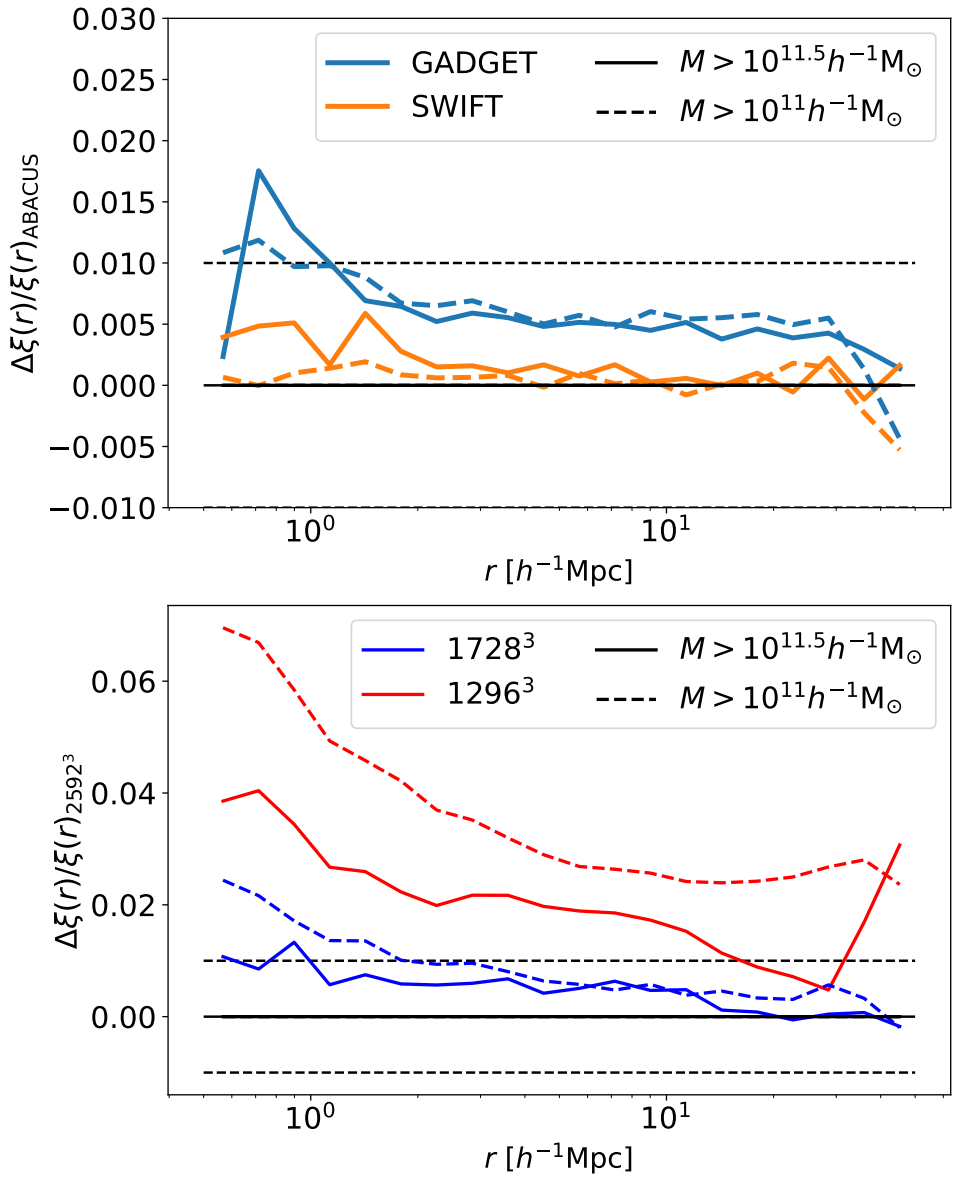

Figure 3.11: Upper Panel: Real-space 2-point correlation functions of the halos from the 2592^3 simulations at $z = 1$. The results from using two halo mass limits are shown as solid and dashed lines. Black dashed lines indicate the 1% level. SWIFT and ABACUS agree more closely than GADGET-2 and ABACUS but for $r > 1\ h^{-1}$Mpc all the codes agree to within 1%. There are not significant differences which come from changing the halo mass limit. Lower Panel: Real-space two-point correlation functions of the halos from the ABACUS simulations relative to the high resolution simulation at $z = 1$. Differences between the halo clustering measurements are larger at small scales. A higher halo mass limit and higher mass resolution both improve the agreement to the high resolution 2-point correlation function.

$\bar{\xi}(s)$ is the mean correlation function,

$$\overline{\xi}(s) = \sum_{i=0}^{N} \frac{\xi^i(s)}{N} \tag{3.12}$$

The standard error estimate, $\sigma(s)$, corresponds to the statistical error for a cubic simulation box with side length $500\,h^{-1}$Mpc. For the full DESI survey volume we use an estimated volume of $20\,h^{-3}\mathrm{Gpc}^3$ and for the year 1 DESI volume we use $4\,h^{-3}\mathrm{Gpc}^3$(Levi et al., 2013). Our estimate of the statistical error on the DESI survey is produced by multiplying by the square root of the volume ratio between the DESI survey and the simulation box

$$\sigma_{\mathrm{DESI}}(s) = \sqrt{\frac{V_{\mathrm{sim}}}{V_{\mathrm{DESI}}}}\sigma(s) \tag{3.13}$$

$\sigma_{\mathrm{DESI}}(s)$ is represented as a grey shaded region on Figs. 3.12–3.13.

This error estimate is likely to be tighter than the errors which are produced by DESI in practice. This is because over the full DESI volume the number density of observed galaxies will not be as high as the number density of halos from the simulations used here. Modelling the galaxy-halo connection will be done in a future work (Hernández-Aguayo et al., 2022a).

A source of error in the clustering comparisons is that we are comparing the results from simulation boxes of finite size. The simulations were run from identical initial conditions, which removes sample variance. However our comparisons between different simulations with the same initial conditions still contain noise in the sense that the differences between these simulations contains a noise component that would change randomly if we changed the initial conditions and compared a second matched pair of simulations. This noise has the potential to obscure the measurement of systematic differences due to code and resolution. We use jackknife method to provide an estimate for this noise by using the variance in the clustering difference between matched sub-samples

$$\sigma_{\mathrm{NOISE}}(s) = \sqrt{\frac{N-1}{N}\sum_{i=0}^{N}\left(\Delta\xi^i(s) - \overline{\Delta\xi}(s)\right)^2} \tag{3.14}$$

where $\Delta\xi^i(s)$ is the difference in the clustering measurements from two simulations with

the i^{th} sub-volume removed and $\overline{\Delta\xi}(s)$ is the mean clustering difference. A detailed investigation of the jackknife method estimating the noise between two simulations sharing the same phases will be presented in (Zhang et al., 2021).

$\sigma_{\text{NOISE}}(s)$ is represented as error bars on Figs. 3.12–3.13. The clustering difference between ABACUS and each code is used for the code comparison noise estimate, while the clustering difference to the high resolution ABACUS simulations is used for the resolution comparison noise estimate.

For the redshift space correlation function measurements, the results from three orthogonal lines of sight are averaged in order to reduce sample variance (Smith et al., 2020).

Figs. 3.12 and 3.13 show the redshift space correlation function comparison between the high resolution simulations for different codes in the monopole and quadrupole respectively. Differences between the codes in the monopole are significant relative to the DESI statistical error below 10 Mpch^{-1}, with sub 0.1% agreement required for consistency within the DESI statistical errors at these length scales. In the quadrupole the simulation noise dominates at small scales, making it difficult to establish systematic differences between the codes. The redshift space correlation function monopole and quadrupole difference between catalogues from different codes is smaller than the DESI statistical error above 20 h^{-1}Mpc. ABACUS and SWIFT have remarkable agreements down to 2 h^{-1}Mpc.

Figs. 3.12 and 3.13 show the comparison between resolutions for the ABACUS simulations. The low resolution correlation function monopole is not consistent with the high resolution case to within the DESI statistical error at any length scale. The medium resolution agrees with the high resolution correlation function to within the DESI and simulation errors for $s > 20\,h^{-1}$Mpc in the monopole and quadrupole.

One may naïvely expect that a threshold of N particles per halo is sufficient for a given clustering accuracy but that does not appear to be the case throughout these comparisons. In Fig 3.11 the medium resolution simulation with a lower halo mass cut agrees with the high resolution case more closely than the low resolution simulation with a higher mass cut. This is likely to be because the force accuracy parameters of the simulations change with resolution and and this could have effects at small length scales.

Our results suggest that our medium and high resolution simulations have sufficient

control over systematic errors in the redshift space correlation function for dark matter halos in the regime where $s > 20\,h^{-1}$Mpc for halos with mass larger than $10^{11.5}\ h^{-1}\ \mathrm{M}_{\odot}$.

3.4.4 Halo Clustering: Power Spectrum

The dark matter halo power spectrum is another useful metric to perform comparisons between N-body simulation codes. The definition of the real-space power spectrum can be found in Section 3.2.2. The multipole expansion of the redshift space power spectrum is used to understand the effects caused by peculiar velocities. The definition of the power spectrum multipole expansion is similar to the correlation function multipole expansion

$$P_l(k) = (2l + 1) \int_0^1 \mathrm{d}\mu P(k, \mu) \mathcal{L}_l(\mu), \tag{3.15}$$

where l is the multipole order, μ is the cosine of the angle between the line of sight and the halo separation vector, $P(k, \mu)$ is the 2d, anisotropic power spectrum, $\mathcal{L}_l$ is the l^{th} Legendre polynomial.

Figs. 3.14–3.16 show the results from the halo power spectra comparisons between codes and resolutions, using two different halo mass cuts, in real-space and redshift space.

Jackknife errors are used to estimate the DESI survey statistical errors along with the noise due to finite simulation volume in a similar manner as in Section 3.4.3. A detailed investigation of the jackknife method estimating the uncertainties in the power spectrum measurements will be presented in (Zhang et al., 2021).

In Fig. 3.14 the real-space power spectra are compared between codes. SWIFT, GADGET-2 and ABACUS agree at the 1% level for $0.01 \leq k \leq 1\ h\,\mathrm{Mpc}^{-1}$. There is not a large difference between the power spectra for different halo mass cuts. The same effects are seen between the codes in the redshift space power spectrum monopole in Fig. 3.15, SWIFT, GADGET-2 and ABACUS are consistent within the DESI statistical errors up to $k = 0.2\ h\,\mathrm{Mpc}^{-1}$.

Fig. 3.14 shows that the medium and low resolution real-space halo power spectra are greater than the high resolution case. The difference increases at smaller scales and is larger for the lower halo mass cut and for the lower resolution simulation. In Fig. 3.15 the same trends are seen in the redshift space monopole, with larger differences than the real-space

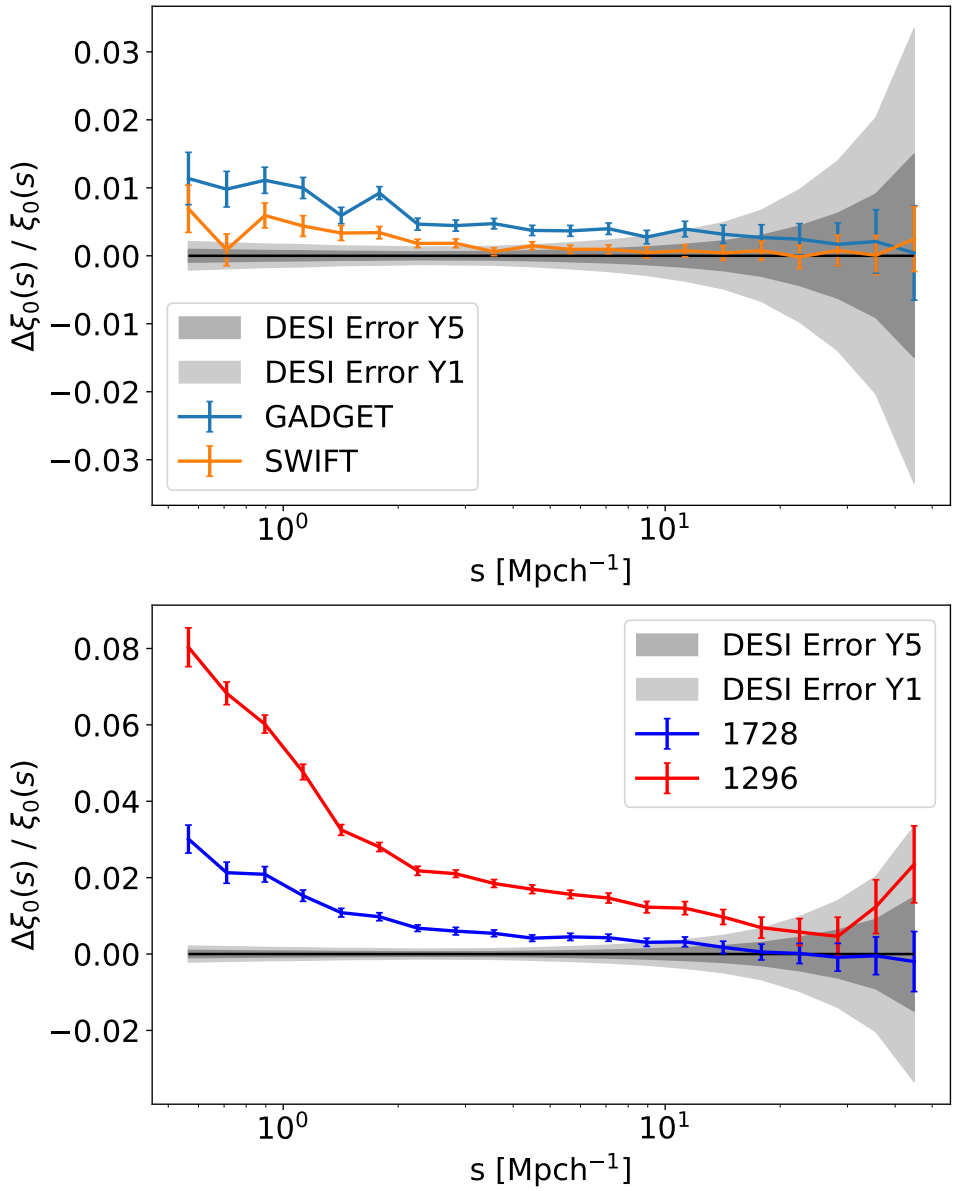

Figure 3.12: Upper panel: The redshift space clustering monopole comparison between the different codes at high resolution. ABACUS is used as the reference clustering. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. A halo mass cut of $M > 10^{11.5}\ h^{-1}\ \mathrm{M}_\odot$ was used. The systematic differences between the codes are within the expected year 5 DESI volume statistical errors for $s > 20\,h^{-1}$Mpc and within the year 1 volume statistical errors for $s > 10\,h^{-1}$Mpc. Lower panel: As above but comparing the redshift space clustering between the different resolutions for the ABACUS code, with reference to the high resolution simulation. The medium resolution simulation is consistent with the high resolution simulation to within the DESI statistical errors for $s > 20\,h^{-1}$Mpc. The low resolution clustering is not consistent with the high resolution clustering to within the expected year 5 DESI statistical errors at any of the measured scales.

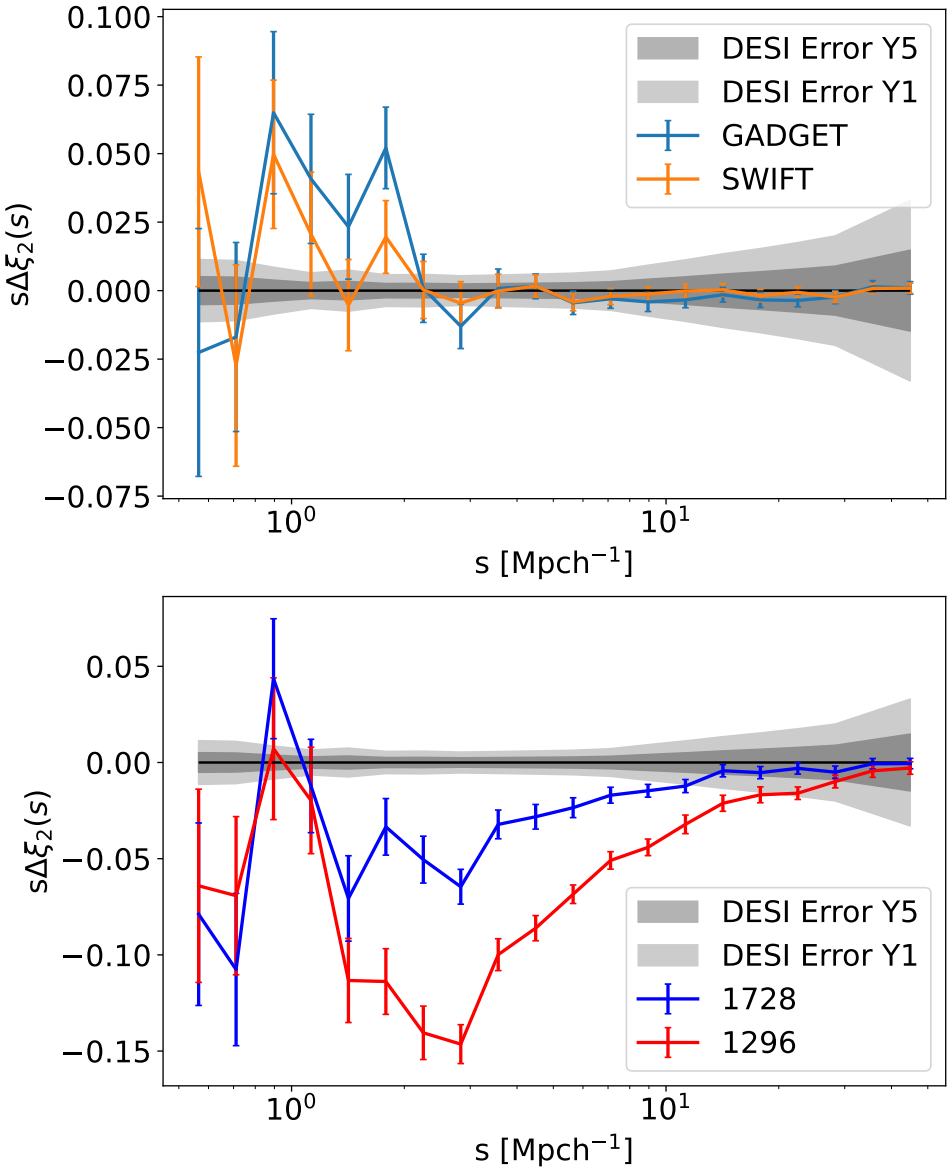

Figure 3.13: Upper panel: The redshift space clustering 2nd order multipole comparison between the different codes at high resolution. ABACUS is used as the reference clustering. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. A halo mass cut of $M > 10^{11.5}\,h^{-1}\,\mathrm{M}_{\odot}$ was used. The codes agree to within the DESI statistical errors for $s > 10\,h^{-1}$Mpc, below which noise begins to dominate the observed differences. Lower panel: As above but showing the redshift space clustering 2nd order multipole comparison between the different resolutions for the ABACUS code, with reference to the high resolution simulation. There are significant differences compared to the level of the DESI statistical errors in both the low and medium resolution simulations for $s < 10\,h^{-1}$Mpc.

power spectrum. The medium resolution simulation is consistent with the high resolution simulation for $k < 0.2\ h\,\mathrm{Mpc}^{-1}$. The low resolution simulation shows significantly greater differences than the medium resolution simulation for $k > 0.15\ h\,\mathrm{Mpc}^{-1}$.

The differences between the codes and resolutions in the redshift space quadrupole are shown in Figure 3.16. Due to the level of noise it is difficult to draw conclusions about code and resolution agreement in the quadrupole. All the codes agree within the noise level, which is comparable to the DESI statistical error for $k < 0.05\ h\,\mathrm{Mpc}^{-1}$. At smaller scales than this the simulation noise becomes larger than the expected DESI statistical error.

3.5 Comparison of Subhalo Mass Functions

The comparison of subhalo mass functions between the simulations at fixed resolution at $z = 1$ is presented in Figure 3.17. The differences between the subhalo mass functions are smaller at lower resolution, lying within 1%, 2%, and 5% in the low, medium, and high resolutions respectively between 10^{11} and $10^{12}\ h^{-1}\mathrm{M}_{\odot}$. SWIFT and GADGET-2 show good agreement within 1% for all resolutions between the mass ranges of 10^{11} and $10^{12}\ h^{-1}\mathrm{M}_{\odot}$. Above $10^{13}\ h^{-1}\mathrm{M}_{\odot}$ the subhalo mass functions become noisy. The subhalo mass function is likely to be more sensitive to differences in the softening schemes of the different simulations than the parent halo mass function.

3.6 Conclusions

We have performed two code comparison projects, comparing initial conditions generators and also N-body simulation codes.

3.6.1 IC Comparison

The initial conditions code comparison included the IC generators associated with the ABACUS, and FASTPM N-body simulation codes, along with the IC generator PANPHASIA/IC_GEN. ICs were generated at three different resolutions for each code. Figure 3.1 shows that the power spectra of the phase matched ICs from the same code agreed to within 1% up to

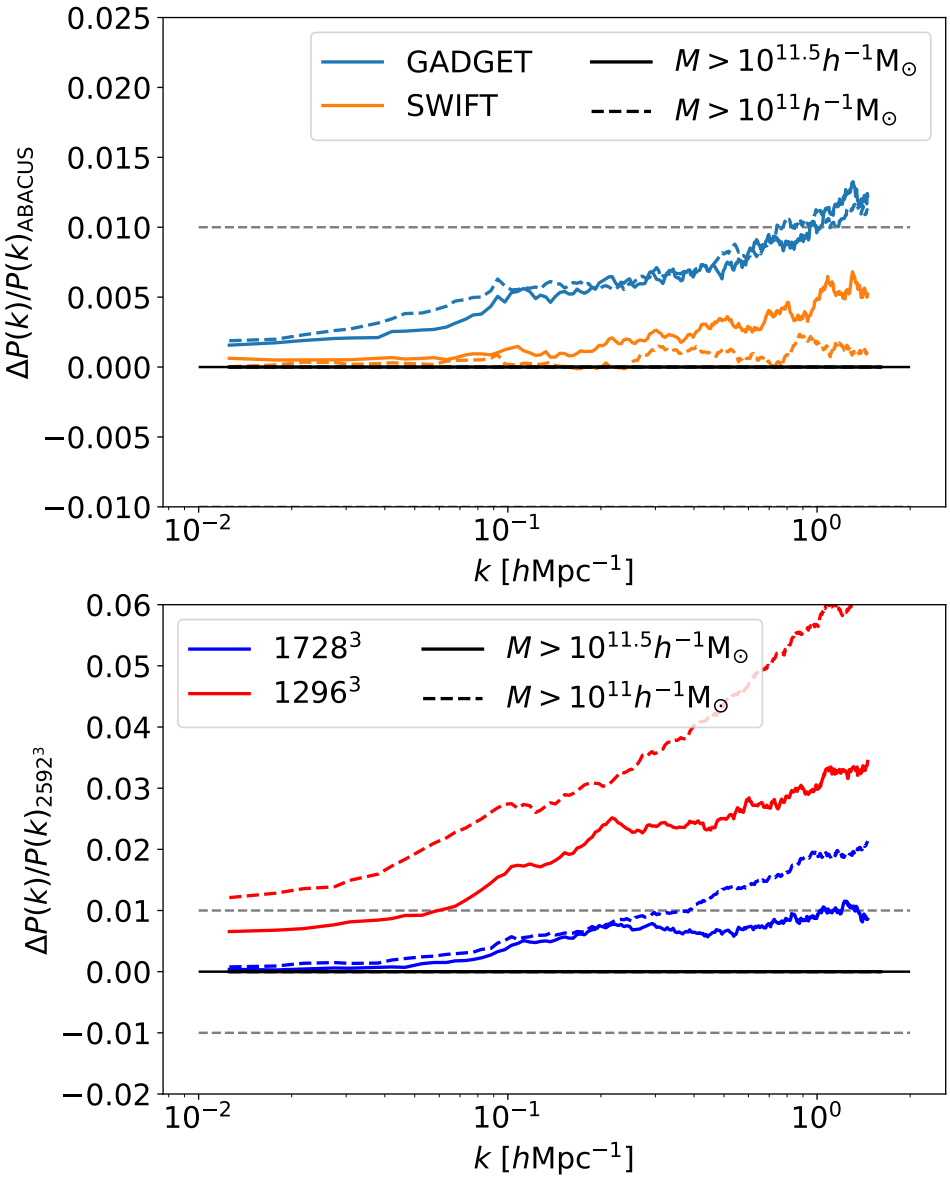

Figure 3.14: Upper panel: The ratio of the real-space power spectra of the halos from the high resolution simulations for different codes to ABACUS at $z = 1$ There is agreement between the codes to within the 1% level at both mass cuts for $k < 1\ h\,\mathrm{Mpc}^{-1}$, with SWIFT and ABACUS showing agreement to within 0.1% for $k < 0.1\ h\,\mathrm{Mpc}^{-1}$. Lower panel: The ratio of the real-space power spectra of the ABACUS halos compared to the high resolution halo catalogue. The medium resolution simulation agrees with the high resolution simulation more closely than the low resolution. The halo mass cut becomes more important for the resolution comparison than in the code comparison. The agreement becomes better with a higher halo mass cut in the resolution comparison.

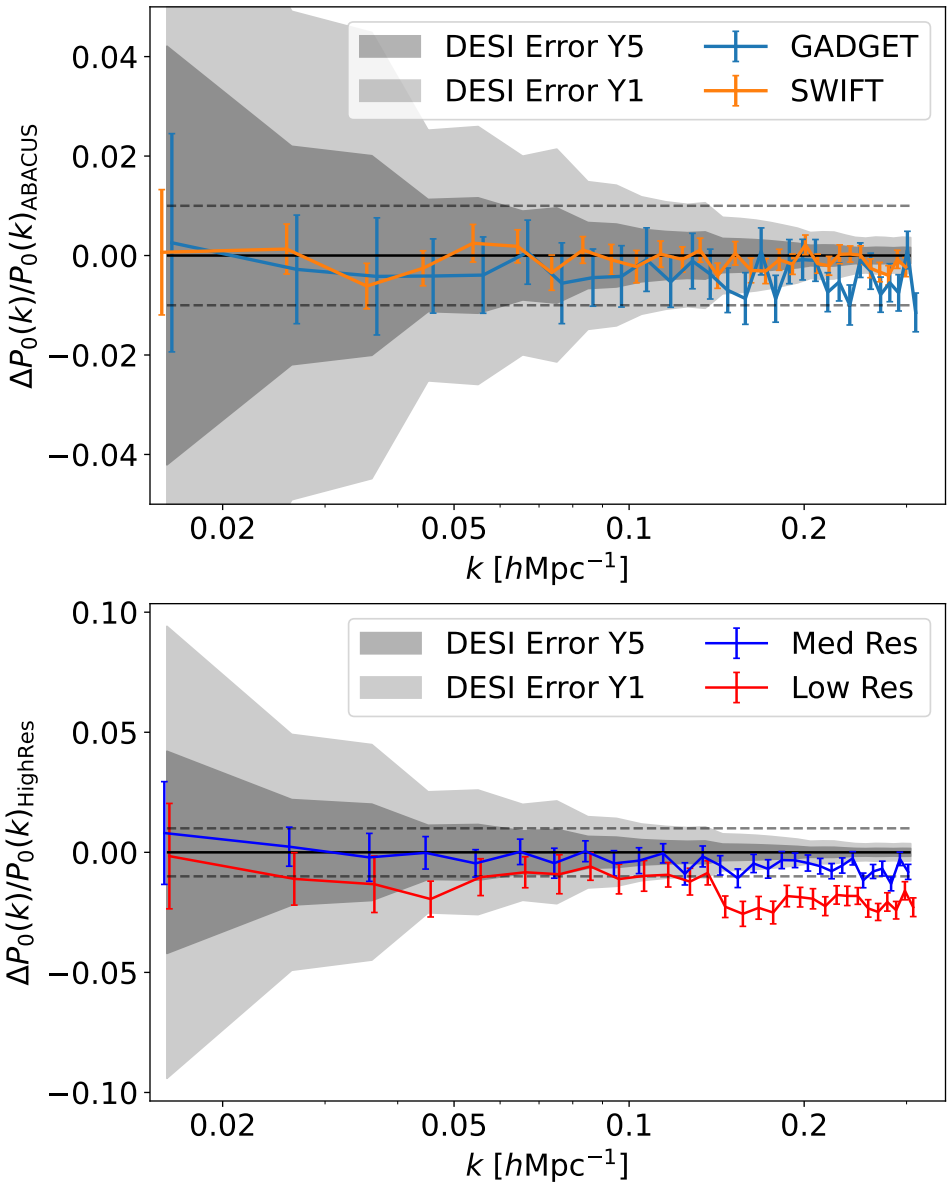

Figure 3.15: Upper panel: The ratio of the redshift space power spectrum monopoles of the halos from the high resolution simulations for different codes compared to ABACUS at $z = 1$. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. The codes show agreement to within the expected DESI statistical errors for $k < 0.1\ h\,\mathrm{Mpc}^{-1}$. Lower panel: The ratio of the redshift space power spectrum monopoles from the high resolution simulations for ABACUS at $z = 1$. The medium resolution simulation is consistent with the high resolution simulation for $k < 0.2\ h\,\mathrm{Mpc}^{-1}$. The low resolution simulation shows significant differences for $k > 0.15\ h\,\mathrm{Mpc}^{-1}$.

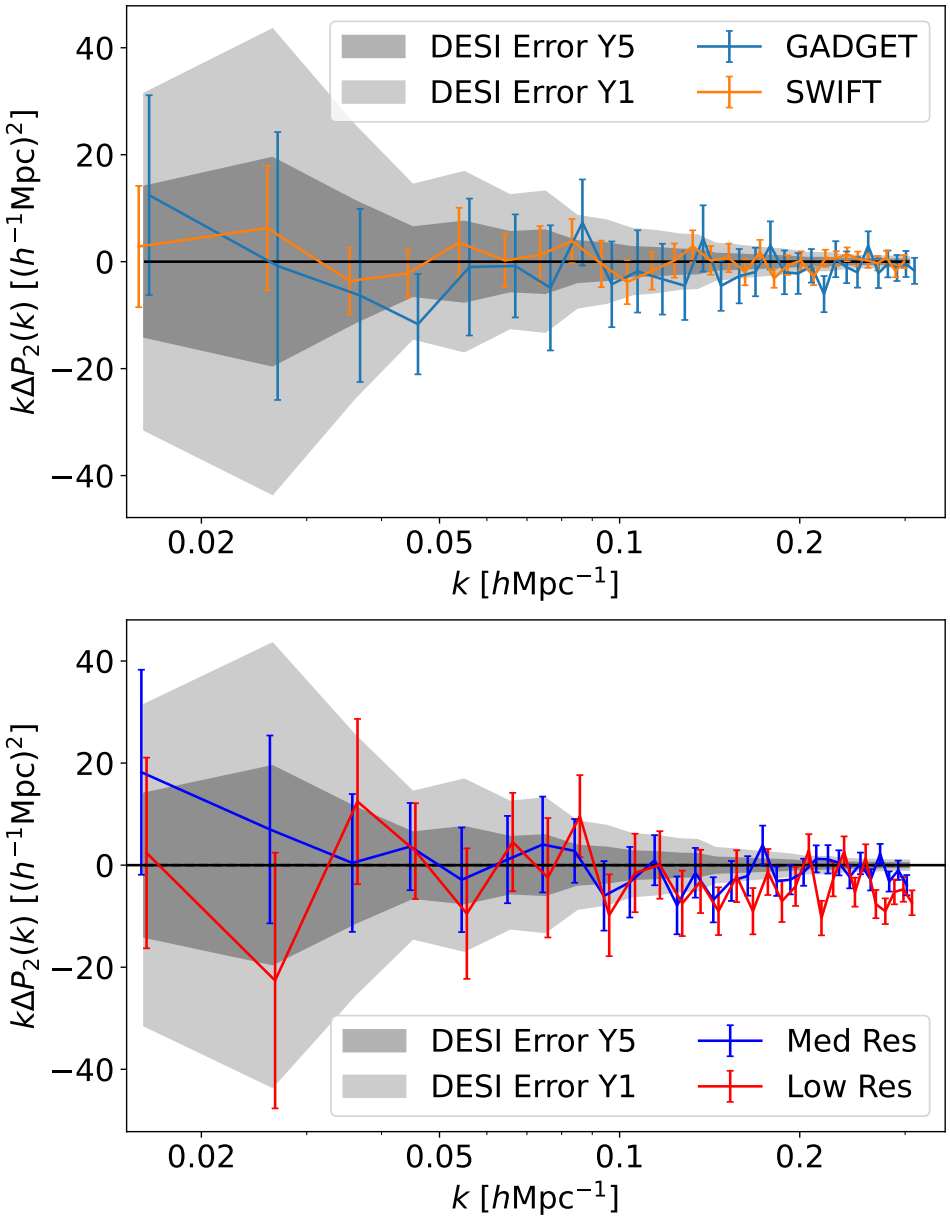

Figure 3.16: Upper panel: The difference of the redshift space power spectrum quadrupoles of the halos from the high resolution simulations for different codes compared to ABACUS at $z = 1$. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. Lower panel: The difference of the redshift space power spectrum quadrupoles from the high resolution simulations for ABACUS at $z = 1$. Due to the level of noise it is difficult to place bounds on the agreement of different codes and resolutions. For $k < 0.05\ h\,\mathrm{Mpc}^{-1}$, where the simulation noise is comparable in magnitude to the DESI errors, the code and resolution comparisons appear to be consistent with one another.

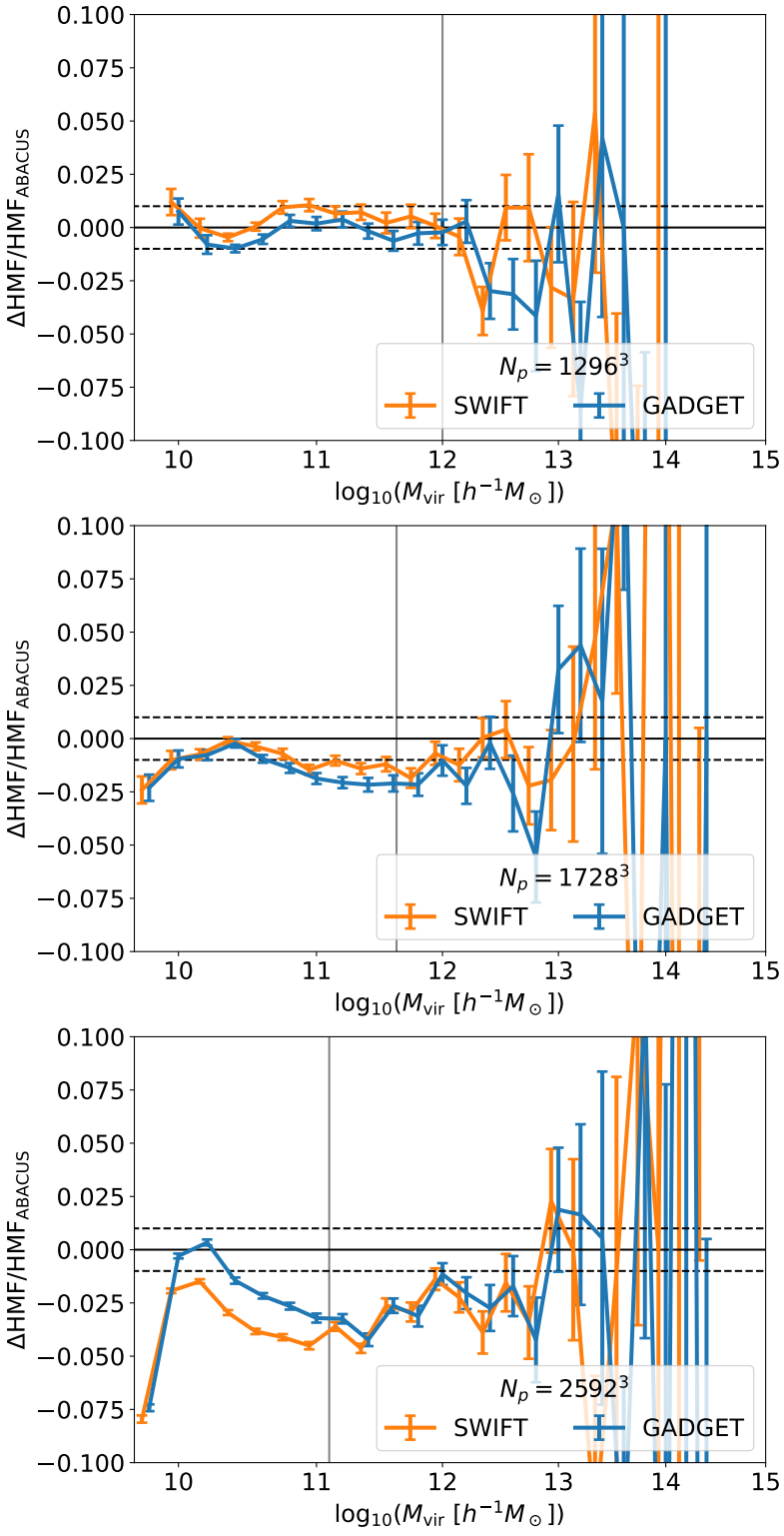

Figure 3.17: The ratio of the subhalo mass functions to that of the ABACUS simulation at fixed resolution at $z = 1$. The top, middle and bottom panels show the low, medium, and high resolution simulations respectively. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. Vertical lines indicate the mass at which halos contain 200 particles at each resolution.

20% of the Nyquist frequency. At large scales the power spectra were consistent with the reference power spectrum, with any differences being within the expected variance due to finite volume.

At small scales, ABACUS, and PANPHASIA/IC_GEN agree with the input linear power spectrum within 1% up to 50% of the Nyquist frequency. Although we chose PANPHASIA/IC_GEN to setup the N-body code comparison project, the effect of the differences among the IC codes should be small.

The pairwise velocity anisotropy was used to check the consistency of the IC codes in the particle velocities as shown in Figure 3.2. The codes were mostly consistent with one another to within the sample variance estimated by using ten realisations of the PANPHASIA/IC_GEN ICs. FASTPM showed larger pairwise anisotropy than the other codes at large scales.

3.6.2 N-body simulation code comparison

The N-body simulation codes, ABACUS, SWIFT, GADGET-2, & FASTPM were run from identical PANPHASIA/IC_GEN ICs from $z = 199$ to $z = 0$. The matter power spectra were compared at $z = 2$ and $z = 1$. Comparisons between codes at fixed resolution were made, along with comparisons between resolutions for the same code.

When comparing between resolutions for each code in Figure 3.3, the power spectra agree to within 1% at $k = 2\ h\,\mathrm{Mpc}^{-1}$ and to within 4% at $k = 10\ h\,\mathrm{Mpc}^{-1}$ with the exception of the quasi-N-body code FASTPM which shows greater differences. Comparing between the codes at fixed resolution the power spectra agree to within 0.1% at large scales as seen in Figure 3.4. Figure 3.5 shows that the large scale power spectra are consistent with perturbation theory estimates made using 2LPT for all the codes. At small scales SWIFT, GADGET-2, and ABACUS show good agreement, within 1% at $k = 10\ h\,\mathrm{Mpc}^{-1}$ for all resolutions at both redshifts. As expected, FASTPM shows the greatest difference in power spectrum relative to the other codes, disagreeing at the 1% level for $k > 0.5\ h\,\mathrm{Mpc}^{-1}$. Similar conclusions are reached at the two explored redshifts.

The dark matter halos were found using ROCKSTAR. Halo properties were compared between matched halos for fixed code and variable resolution or fixed resolution and variable code and the results were shown in Figures 3.7 & 3.8. There was the greatest

agreement between codes for high mass halos in most statistics. The difference between matched halo properties was within 1% between all the codes for halo masses greater than $10^{13}\ h^{-1}\ \mathrm{M}_{\odot}$.

Matching halo properties between different resolutions presented several systematic effects. Low mass halos in low resolution simulations had greater mass than the matched halos from the high resolution simulations, with the opposite effect observed at high masses. Matched halo properties had better agreement at high masses when comparing between simulations run with different resolutions.

The halo mass functions from SWIFT, GADGET-2, and ABACUS agreed to within 1% between mass limits of $10^{11.5} - 10^{14}\ h^{-1}\ \mathrm{M}_{\odot}$ as seen in Figure 3.9.

Halo clustering measurements were made by using both the correlation function and power spectrum. Comparisons were made between codes and resolutions for two different halo mass limits 10^{11} and $10^{11.5}\ h^{-1}\ \mathrm{M}_{\odot}$. The minimum halo mass that hosts the DESI ELG sample is expected to be between these two values (Yuan et al., 2022a).

Comparing different codes in the halo clustering, all codes agree with each other within the expected DESI year 1 uncertainty at scales larger than $10\,h^{-1}$Mpc and within the DESI year 5 uncertainty at scales larger than $20\,h^{-1}$Mpc (Figs. 3.11-3.13). ABACUS and SWIFT have remarkable agreements down to $2\,h^{-1}$Mpc. Comparing the different resolutions, the medium resolution run agrees with the high resolution one at scales larger than $20\,h^{-1}$Mpc. This indicates that one should be more careful choosing simulation resolution than choosing the N-body code.

In the halo power spectra, similar results to the halo correlation function were found as seen in Figs. 3.14-3.16. ABACUS, SWIFT, and GADGET-2 have good agreement at all the scales which we are most interested in for DESI, i.e. k <0.3 $h\,\mathrm{Mpc}^{-1}$. Medium resolution simulations agree with the high resolution simulations at a similar level to the comparisons between different codes. The low resolution simulation shows differences at least twice as large as the medium resolution simulation at all length scales. The agreement becomes closer between different resolution simulations with a higher halo mass cut.

These results indicate the expected level of systematic errors in a variety of statistics associated with N-body simulations.

We do not draw conclusions about whether certain simulations are appropriate for use

in any specific DESI analyses, this requires further propagation of the simulations through the cosmology analysis pipeline. These simulations and the results provided should act as a baseline for expected systematic errors, regardless of the application of the simulations. Future work using these simulations will explore how these errors are propagated through HOD and cosmological analysis and therefore will assess the suitability of simulations run with different codes and resolutions for use in modern galaxy surveys (Hernández-Aguayo et al., 2022a).

CHAPTER 4

Bright Galaxy Survey Mock Catalogues - HOD Fitting

4.1 Introduction

4.1.1 The Galaxy-Halo Connection

Much of the work explored so far has investigated the accuracy and reliability of N-body simulations. However, we cannot stop there if we are to apply these simulations to the analysis of galaxy surveys. In photometric and spectroscopic galaxy surveys we do not observe the distribution of dark matter in the universe, instead we observe the distribution of galaxies that satisfy the selection criteria. In order to relate dark-matter only N-body simulations to galaxy surveys we must model the galaxy-halo connection in addition to selection effects such that mocks look like survey data.

The galaxy-halo connection has been a subject of research for several decades (Cole & Kaiser, 1989; Wechsler & Tinker, 2018; Mo & White, 1996). There are many formulations of the galaxy-halo connection, from simpler Halo Occupation Distributions (HODs) to more complex treatments that aim to model the impact of various physical processes on galaxy formation in hydrodynamic and semi-analytic simulations.

Hydrodynamic galaxy formation simulations model the galaxy-halo connection directly by including both dark-matter and baryonic components interacting over time di-

rectly through gravity and other forces. For example, the heating and cooling of gas is simulated, and star formation is included. This method produces the galaxy-halo connection emergently, as a product of the physical assumptions for cosmology and for interaction of matter (Bose et al., 2019; Desmond et al., 2017). However, these simulations are much more computationally expensive than dark-matter-only simulations which significantly reduces the volumes that can be simulated in a reasonable amount of time. In addition, physics below the resolution of the simulation has to be implemented using a sub-grid model (e.g. Negri & Volonteri, 2017) which can have many degrees of freedom and an effect on the output galaxy properties. Sub-grid physics can be constrained by comparing to observations (Gibson et al., 2013). Examples of work using hydrodynamic simulations with realistic large-scale structure include: EAGLE (Crain et al., 2015), BAHAMAS (McCarthy et al., 2016), Illustris (Vogelsberger et al., 2014), and Horizon (Lee et al., 2021).

Semi-analytic galaxy formation models can use an existing dark matter-only N-body simulation and model the history of galaxy properties and halo mergers across time. Physically-informed models simulate the evolution of galaxies across timesteps through a variety of processes such as star formation, feedback, and radiative heating and cooling. This is computationally cheaper than running a full hydrodynamic simulation, allowing greater volumes to be simulated with a greater variety of parameters. There are a large number of degrees of freedom and assumptions underlying the physical models which are constrained to match observations. The choice of processes and model formulations is diverse between different semi-analytic codes, examples include GALFORM (Cole et al., 2000; Baugh et al., 2005; Bower et al., 2006; Lacey et al., 2016; Elliott et al., 2021), L-GALAXIES (Henriques et al., 2015, 2019), and others (Croton et al., 2006; Somerville et al., 2008).

HODs are a more empirical method for modelling the galaxy-halo connection. Instead of assuming that galaxies are formed by specific physical processes which affect their abundance and properties, HOD methods simply assume that the abundance of galaxies is informed by the host halo properties without reference to a physical model. This abstracts away much of the underlying physics but is sufficient if one only cares about knowing what the galaxy-halo connection is, rather than why it takes that form.

In the simplest form of HOD the number of galaxies hosted by a halo is purely a function of the halo mass. The form of the HOD is often split up into two components, one representing the probability of a central galaxy (capped at one by definition), and another representing the abundance of satellite galaxies. The specific choice of form for the HOD may depend on the type of galaxy being modelled. For example, emission line galaxies (ELGs) are modelled with a gaussian distribution for the central galaxy occupation that decreases at high masses (Avila et al., 2020), meanwhile luminous red galaxies (LRGs) are usually modelled with a smoothed step function that approaches one at high masses (Zheng et al., 2009).

More complex versions of HOD methods exist to account for the relationship between other halo properties and galaxy abundance, known as assembly bias. Decorated HODs have been explored as methods to improve HODs. These add dependence of the HOD on several more parameters such as formation time and concentration (Hearin et al., 2016). However, for the work presented in this chapter only halo mass dependence is used because the tabulation method (Section 4.3) does not scale well with greater dimensionality.

There are different methods of treating the placement of satellite galaxies in HOD methods. One approach is to place satellite galaxies randomly around the centre of the halo according to a distribution. The Navarro-Frenk-White (NFW) profile describes the mass/radius relationship for dark matter halos (Navarro et al., 1996, 1997) and satellite galaxies can be placed following this profile. In addition, random halo particles can be used to place satellite galaxies as these are tracers of the mass distribution.

Alternatively, satellite galaxies may be placed on sub-halos identified around the parent halo. Sub-halo abundance matching (SHAM) techniques can be used to link galaxy properties to sub-halo properties (Conroy et al., 2006; Reddick et al., 2013). SHAM techniques require high simulation resolution in order to resolve sub-halos sufficiently, otherwise halos may be overmerged and systematic errors can be produced in the clustering (Guo & White, 2013). An example of a simulation with sufficient resolution to use SHAM instead of HOD methods is the Uchuu simulation (Ishiyama et al., 2021).

4.1.2 AbacusSummit Overview

In this work we have used the AbacusSummit simulations (Maksimova et al., 2021) to produce HOD parameter fits and mock galaxy catalogues for BGS.

AbacusSummit[1] is a suite of N-body simulations run using the ABACUS code (Garrison et al., 2021b) on the Summit supercomputer. There are 150 simulations that have been run using different cosmologies, resolutions and initial conditions. The majority of the simulations use what is called the "Base" resolution and this was used for all of my analysis. The "Base" resolution is a cubic box with side length 2 h^{-1}Gpc and 6912^3 particles which produces a particle mass of $2.0 \times 10^9 h^{-1}\mathrm{M}_\odot$.

The AbacusSummit simulations were run from 2LPT initial conditions at z=99. Further details of the N-body simulation parameters can be found in Section 3.3.1.

AbacusSummit runs were performed in a variety of different cosmologies. The primary cosmology corresponds to Planck 2018 LCDM results (Planck Collaboration, 2020), the Hubble expansion rate at present time, $H_0 = 67.36$ km s^{-1}, $\Omega_c h^2 = 0.1200$, $\Omega_b h^2 = 0.0223$, $\Omega_\Lambda = 0.6834$, r.m.s linear density fluctuation, $\sigma_8 = 0.81$, scalar spectral index $n_s = 0.9649$. The secondary cosmologies are selected to match existing flagship N-body simulation projects and there is a wide emulator grid around the primary cosmology. Figure 4.1 shows the scatter in these cosmological parameters across the AbacusSummit cosmologies.

Halo catalogues are the primary data product from AbacusSummit, with full sets of all particles being too large to store efficiently. Halo finding in AbacusSummit was performed using CompaSO (Hadzhiyska et al., 2021), a spherical overdensity method. Subsets of field particles also exist and these were used in cases where the HOD was used below the mass resolution for resolved halos (see Section 4.3.4).

4.2 HOD Formulation

We use an HOD of the same form as Smith et al. (2017) which is closely related to Zehavi et al. (2011). This form is described by five parameters in total, two for the central galaxy

[1]https://abacussummit.readthedocs.io/

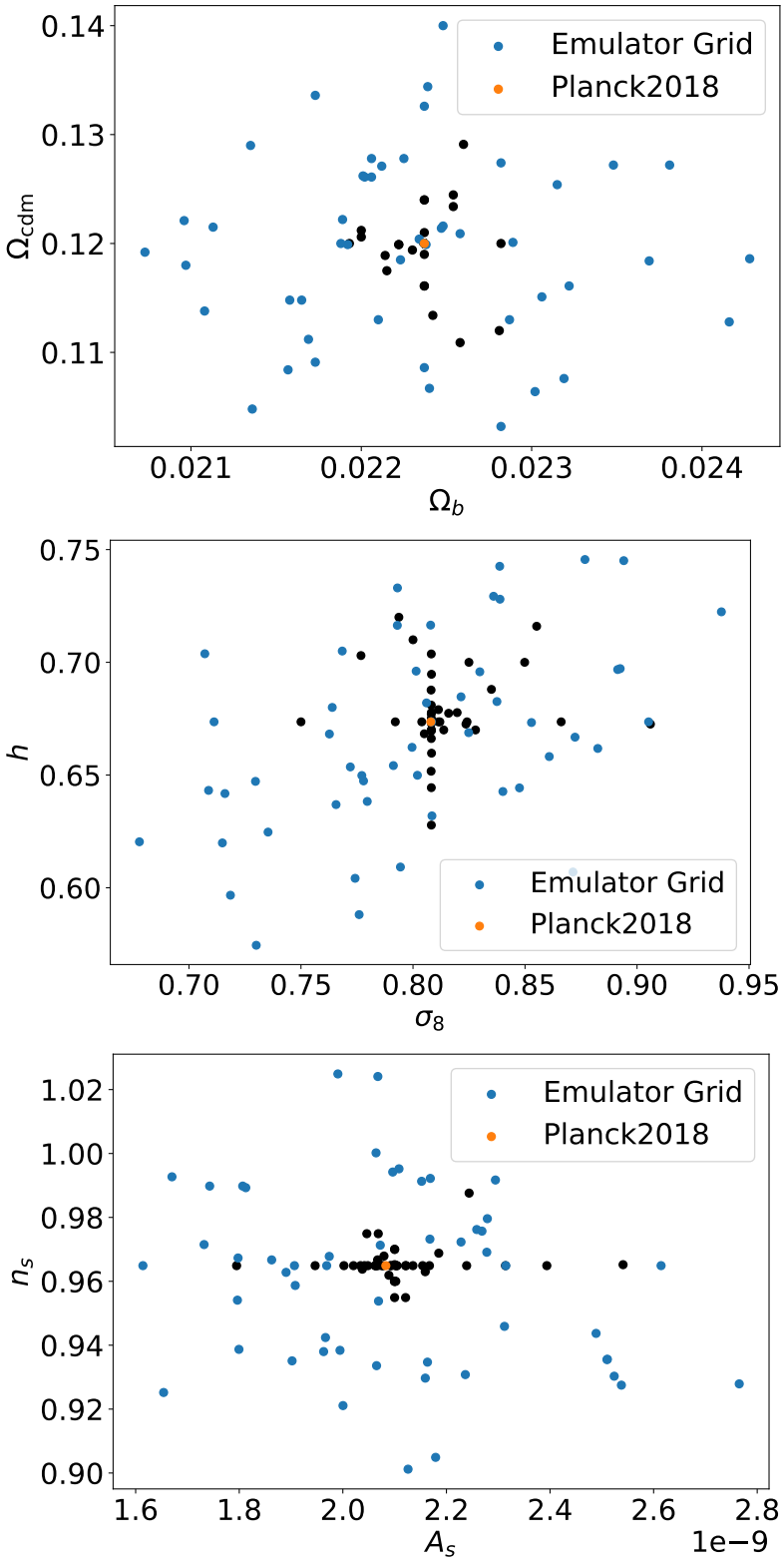

Figure 4.1: Scatter plots of the cosmologies spanned by the AbacusSummit simulations. In each panel the orange central point represents the primary Planck 2018 cosmology, the blue points are an emulator grid around the primary cosmology, and the black points represent secondary cosmologies and smaller deviations to the primary cosmology.

component and three for the satellites. The central galaxies are populated according to a smoothed step function as a function of halo mass. At low masses a pseudo-Gaussian spline kernel is used rather than a Gaussian to reduce the size of the tail of the central HOD.

The HOD for the number of central galaxies brighter than L per halo of mass M is defined as

$$\langle N_{\rm cen}(> L \mid M)\rangle = \frac{1}{2}\left[1 + F\left(\frac{\log M - \log M_{\rm min}(L)}{\sigma_{\log M}(L)}\right)\right], \tag{4.1}$$

where $M_{\rm min}$ defines the position of the step and $\sigma_{\log M}$ defines the width of the step. $F(x) = 2\int_0^x S(x')\,dx'$ and $S(x')$ is defined as the spline function to reduce the low-mass tail of the distribution:

$$\begin{aligned} S(x) &= 1 - 6|x|^2 + 6|x|^3 && |x| \leq 0.5 \\ &= 2(1-|x|)^3 && 0.5 < |x| \leq 1 \\ &= 0 && 1 < |x| \end{aligned} \tag{4.2}$$

The satellite HOD is a truncated power law weighted by the central HOD. The form of the satellite HOD is

$$\langle N_{\rm sat}(> L \mid M)\rangle = \langle N_{\rm cen}(> L \mid M)\rangle \left(\frac{M - M_0(L)}{M_1'(L)}\right)^{\alpha(L)}, \tag{4.3}$$

where M_0 is the cutoff mass scale, M_1' the normalisation, and α the power law slope.

4.2.1 Smoothly varying parameters with absolute magnitude

A central assumption in this implementation is that HOD parameters are defined to smoothly vary with an absolute magnitude limit. This is a reasonable assumption because we are not modelling a galaxy population that is sensitive to an on/off state such as star formation. This allows us to define the HOD at any magnitude and therefore to populate the mock galaxy catalogue with galaxies that have realistic luminosities, rather than populating a mock that represents a fixed magnitude cut. Fitting HOD parameters to smoothly varying curves presents some new challenges and there are a couple of options

to implement this.

One option is to fit HOD parameters for a set of different magnitudes, then to fit a smooth curve between the best-fit parameters. This method has the benefit of being relatively quick and simple, as one only has to repeat the standard HOD fitting procedure several times, once for each magnitude limit. However, there is no guarantee that the best-fit HOD parameter values can be well approximated by a smooth curve as a function of magnitude. The HOD parameters produced by fitting a smooth curve to an existing set of HOD parameter values also may not preserve the target galaxy clustering which has also been fitted. An example of this approach is Smith et al. (2017)

The fitting method chosen in this project is to instead use parameterised curves to describe how the HOD parameters vary with magnitude before fitting. These meta-parameters are then fitted, meaning that no further adjustment needs to be made to the fitted HOD parameters. This means that each HOD parameter, and therefore the shape of the total HOD, can be defined for any magnitude limit. One potential pitfall of this method is that if the parameterised curves are too constrained, then a good fit to the target clustering may not be found and therefore different forms for the curves may need to be tested in order to produce robust results. Another issue is that fitting all of the meta-parameters at once is a higher dimensional fit than fitting parameters for individual magnitudes separately. This takes up more computing time to complete the fit and the large number of parameters means that the fitting procedure is more likely to end up in a local minimum than be able to locate the global minimum. We attempted to overcome this by running the fitting procedure several times from different starting points in the parameter space. This provides more confidence that the best fit HOD parameters are stable without needing to allocate vastly greater computational resources to the problem.

The parameterised forms for the HOD parameter variation with magnitude are shown below with P_1 through P_{17} representing the fitted parameters giving 17 in total. Examples of these fitted forms can be found in Figure 4.2. Curve forms were chosen to be similar to the observed variation of best-fit parameters with magnitude change when fitted independently. $\log_{10}(M_0)$ is described by a linear function of magnitude, $\sigma_{\log M}$ is described by a smoothly varying step function, α is described by a constant value with an exponential component at bright magnitudes, finally both $\log_{10}(M_{\rm min})$ and $\log_{10}(M_1)$ are described by cubic

functions.

$$\log_{10}(M_0) = P_1(M_r + 20) + (P_2 + 11) \tag{4.4}$$

$$\sigma_{\log M} = P_3 + \frac{P_4 - P_3}{1 + e^{P_5(M_r + 20 + P_6)}} \tag{4.5}$$

$$\alpha = P_7 + P_8^{-(M_r+20)+P_9} \tag{4.6}$$

$$\log_{10}(M_{\rm min}) = (P_{10} + 12) + P_{11}(M_r + 20) + P_{12}(M_r + 20)^2 + P_{13}(M_r + 20)^3 \tag{4.7}$$

$$\log_{10}(M_1) = (P_{14} + 12) + P_{15}(M_r + 20) + P_{16}(M_r + 20)^2 + P_{17}(M_r + 20)^3 \tag{4.8}$$

4.2.2 Non-overlap condition

A benefit to using this method is that the HOD curves can be defined such that they do not produce overlaps as magnitude varies. If the curves overlap then that can imply that there are more objects brighter than a bright limit than a fainter limit which is clearly unphysical. Building this consideration into the fitting procedure ensures that it is not an issue which we have to overcome later. This is done by adding a high likelihood penalty in cases where curves overlap.

In addition, using the spline curve to truncate the low mass tails of the HOD curves helps to avoid HOD crossing at very low masses/occupations.

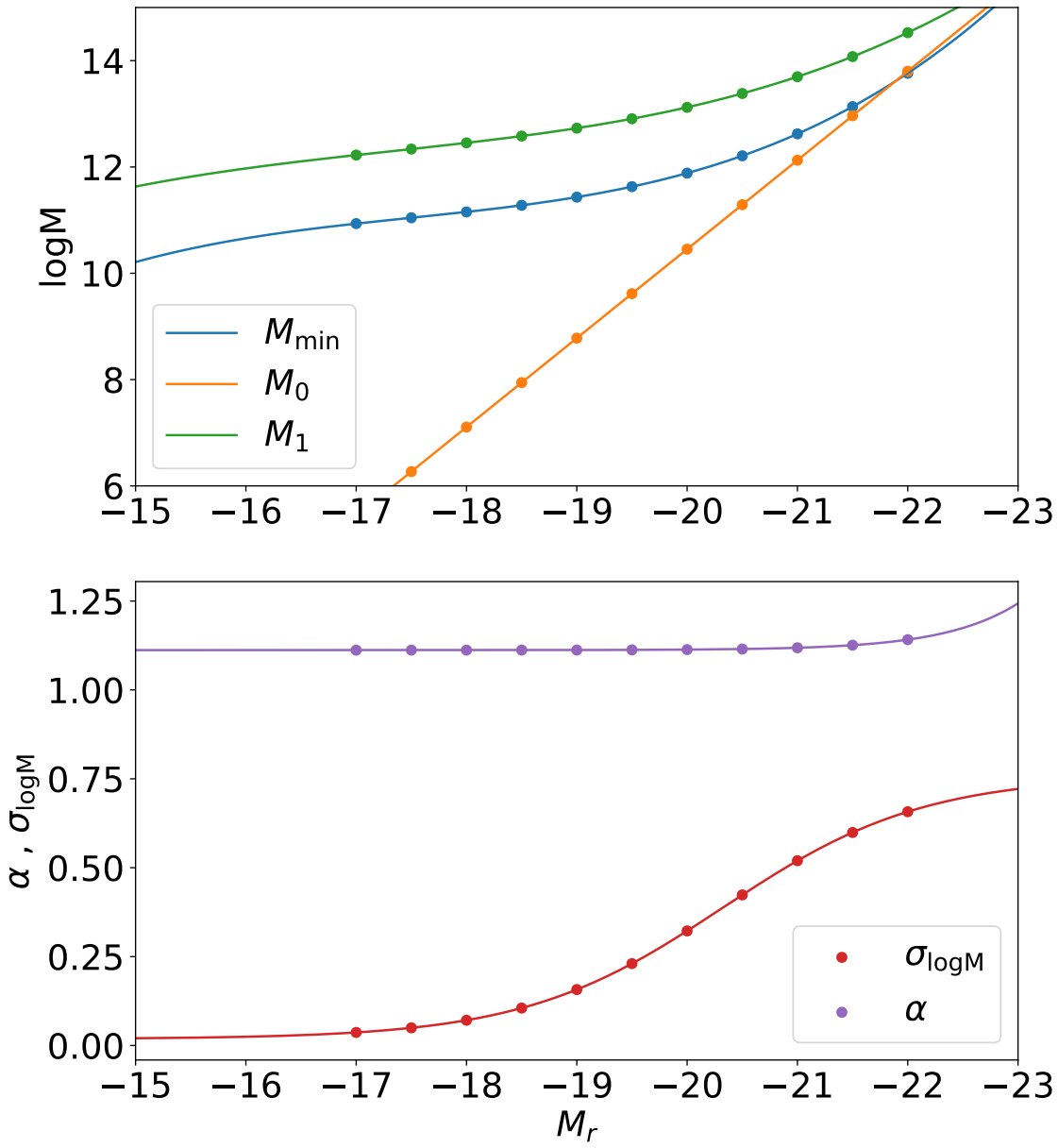

Figure 4.2: The typical shape of the HOD parameter curves as a function of absolute magnitude. Dots show the points at which fits are made. The top panel shows the mass parameters $M_{\rm min}$,M_0, and $M1$ on a logarithmic scale. The lower panel shows the shape of the curves describing the non-mass parameters, α, and $\sigma_{{\rm log}M}$.

4.3 HOD Fitting Tabulation Method

4.3.1 Introduction to the Tabulation Method

In order to efficiently explore the parameter space of possible HOD forms, a fast method to evaluate the expected galaxy clustering is needed. We have implemented a version of the halo paircounting method introduced by Zheng & Guo (2016). This method tabulates the dark matter halo paircounts by mass and separation, then the average effect of using a particular HOD can be estimated by reweighting elements of these tables and summing the contributions to the clustering.

Using this method means that the computationally expensive paircounting routine needs only to be run once in total, instead of once for each set of HOD parameters. The paircounting is run on the halos instead of the galaxies, and the larger number of halos means that there is a greater fixed cost to using this method. However when many evaluations are needed in the fitting procedure, the much quicker time for each evaluation leads to a quicker runtime in total.

Alternatives to explicitly calculating the clustering produced by a set of HOD parameters can include using emulators or analytic methods (Kwan et al., 2015; Zhai et al., 2019; Yuan et al., 2022b), but these are not guaranteed to produce accurate results for a given simulation.

This tabulation method assumes that occupancy of a halo depends on only one variable, the halo mass. We can bin halo paircounts by different combinations of halo mass and then reweight these paircounts to account for changes in the HOD. It is possible to include other halo parameters by mixing them with the halo mass to create an *effective mass* which maintains the HOD function as dependent on a single variable (Yuan et al., 2018), however this method is not tested here. HODs with two or more input variables can be used, but this increases the dimension of the tabulated paircounts and leads to an exponential increase in code runtime.

4.3.2 Tabulation Method Equations

The galaxy clustering measurement which we are aiming to estimate is the real space two point correlation function,

$$\xi(r) = \frac{N_{\rm gal}^{\rm pairs}(r)}{RR(r)} - 1. \tag{4.9}$$

Here $N_{\rm gal}^{\rm pairs}(r)$ represents the galaxy paircounts binned by pairwise distance. $RR(r)$ represents the paircounts if the same number of galaxies were distributed randomly in the box. For our periodic simulation boxes, $RR(r)$ can be calculated analytically. We can split up $N_{\rm gal}^{\rm pairs}(r)$ into a sum over pairs of galaxies from halos with varying halo masses.

We divide the total mass range of halos which we are considering into N bins which are evenly spaced in the logarithm of the halo mass. Then we can split the galaxy paircounts into a sum over all combinations of mass bin, represented by M_{ij} for the number of pairs between mass bins i and j. Pairs must not be double counted.

$$N_{\rm gal}^{\rm pairs}(r) = \sum_{M_{ij}} N_{\rm gal}^{\rm pairs}(M_{ij}, r) \tag{4.10}$$

Each term for an individual mass and distance bin can be split up into the components from central and satellite galaxy pairs in all combinations,

$$N_{\rm gal}^{\rm pairs}(M_{ij}, r) = N_{\rm CC}^{\rm pairs}(M_{ij}, r) + N_{\rm CS}^{\rm pairs}(M_{ij}, r) + N_{\rm SS}^{\rm pairs}(M_{ij}, r), \tag{4.11}$$

where the subscripts CC, CS, and SS represent central-central, central-satellite and satellite-satellite pairs respectively.

Once the total paircount is subdivided in this way it is possible to use the halo paircounts and the HOD to find the galaxy paircounts.

The HOD is defined such that for a halo of mass M, the probability of containing a central galaxy is $n_{\rm c}(M)$ and the average number of satellite galaxies is $n_{\rm s}(M)$ (there cannot be more than one central galaxy). We assume that there does not have to be a central galaxy in a halo in order for it to host satellite galaxies. This assumption simplifies the tabulation calculations significantly as otherwise one would have to consider a larger number of correlations from halos explicitly with and without central galaxies. In practice the majority of satellite galaxies are placed in halos containing a central galaxy as they

are preferentially placed in high-mass halos. Central galaxies are positioned at the centre of the halo and satellite galaxies are positioned randomly according to an NFW profile around the halo. The number of satellite galaxies per halo is chosen according to a Poisson distribution with mean $n_{\rm s}(M)$.

The central galaxies will share positions with the halos as they are placed at the centre, therefore the distribution of potential central galaxy positions is sampled by the halos themselves. For the satellite galaxies we place a fixed number (in this work, three are used) of tracer particles around each halo to represent where satellite galaxies would be placed.

Let $W_{ij}^{\rm CC}(r)$ represent the number of pairs between central halos in mass bins i and j, divided into the same radial bins which are used in the clustering measurement. $W_{ij}^{\rm CS}(r)$ represents the paircounts between central halos and satellite tracer particles. $W_{ij}^{\rm SS1}(r)$ and $W_{ij}^{\rm SS2}(r)$ represent the satellite tracer paircounts, divided into the one-halo and two-halo terms. The one-halo term is for paircounts between satellite galaxies sharing the same dark matter halo. Meanwhile the two-halo term is between satellites from separate dark matter halos. The one-halo term is important at small scales that probe the interior of dark matter halos.

First consider the case of the central-central galaxy pairs. As defined above, the probability of a halo of mass M containing a central galaxy is $n_{\rm c}(M)$. This probability is independent for all halos and therefore to find the number of galaxy pairs we can simply take the halo paircounts weighted by the product of the probability of each halo within the pair being occupied,

$$N_{\rm CC}^{\rm pairs}(r) = \sum_{M_{ij}} n_{\rm c}(M_i) n_{\rm c}(M_j) W_{ij}^{\rm CC}(r). \tag{4.12}$$

The same conditions apply for the central-satellite term and the satellite-satellite two-halo term, however we also have to account for the number of satellite tracer particles per halo,

$$N_{\rm CS}^{\rm pairs}(r) = \sum_{M_{ij}} n_{\rm c}(M_i) n_{\rm s}(M_j) W_{ij}^{\rm CS}(r) S^{-1} \tag{4.13}$$

$$N_{\mathrm{SS2}}^{\mathrm{pairs}}(r) = \sum_{M_{ij}} n_{\mathrm{s}}(M_i) n_{\mathrm{s}}(M_j) W_{ij}^{\mathrm{SS2}}(r) S^{-2} \tag{4.14}$$

where S is the number of satellite tracer particles per halo. Note the different dependence on the number of satellite tracer particles between Equation 4.13 and 4.14.

In practice, $S = 1$ is sufficient for accurate calculation of both the central-satellite term and the satellite-satellite two-halo term. For the one-halo term, we must use more than one tracer particle per halo to sample the possible galaxy pairs. The number of one-halo satellite galaxy pairs sampled per halo when using S tracer particles per halo is $S(S-1)/2$

The expected number of satellite galaxy pairs in a halo with n satellite galaxies is $\langle n(n-1)/2 \rangle$ which in the case of a Poisson distribution becomes $\bar{n}^2/2$.

Therefore in this case we relate the galaxy and halo paircounts using

$$N_{\mathrm{SS1}}^{\mathrm{pairs}}(r) = \frac{\sum_{M_{ij}} n_{\mathrm{s}}(M_i) n_{\mathrm{s}}(M_j) W_{ij}^{\mathrm{SS1}}(r)}{S(S-1)}. \tag{4.15}$$

Here all the off-diagonal terms are zero as there are no pairs of satellites between halos of different masses in the one-halo term

In each case the halo paircounts are static and only depend on the halo catalogue, meanwhile the HOD factors can vary allowing us to fit HOD parameters rapidly by evaluating the sums shown above.

The halo paircounts only have to be measured once, they can then be saved in the tabulated form shown below. This allows the galaxy paircounts to be calculated by multiplying the table elements by the relevant HOD factor and then summing over the table.

$$N_{\mathrm{CC}}^{\mathrm{pairs}}(r) = \begin{array}{c} \\ n_{\mathrm{c}}(M_1) \\ n_{\mathrm{c}}(M_2) \\ n_{\mathrm{c}}(M_3) \\ n_{\mathrm{c}}(M_4) \\ \dots \end{array} \begin{array}{c} \begin{array}{ccccc} \mathrm{HOD} & n_{\mathrm{c}}(M_1) & n_{\mathrm{c}}(M_2) & n_{\mathrm{c}}(M_3) & n_{\mathrm{c}}(M_4) & \dots \end{array} \\ \begin{bmatrix} W_{11}^{\mathrm{CC}}(r) & W_{12}^{\mathrm{CC}}(r) & W_{13}^{\mathrm{CC}}(r) & W_{14}^{\mathrm{CC}}(r) & \dots \\ W_{21}^{\mathrm{CC}}(r) & W_{22}^{\mathrm{CC}}(r) & W_{23}^{\mathrm{CC}}(r) & W_{24}^{\mathrm{CC}}(r) & \dots \\ W_{31}^{\mathrm{CC}}(r) & W_{32}^{\mathrm{CC}}(r) & W_{33}^{\mathrm{CC}}(r) & W_{34}^{\mathrm{CC}}(r) & \dots \\ W_{41}^{\mathrm{CC}}(r) & W_{42}^{\mathrm{CC}}(r) & W_{43}^{\mathrm{CC}}(r) & W_{44}^{\mathrm{CC}}(r) & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \end{array} \tag{4.16}$$

$$
SN^{\text{pairs}}_{\text{CS}}(r) = \begin{array}{c} \\ n_{\text{c}}(M_1) \\ n_{\text{c}}(M_2) \\ n_{\text{c}}(M_3) \\ n_{\text{c}}(M_4) \\ \dots \end{array} \begin{array}{c} \begin{array}{ccccc} \text{HOD} & n_{\text{s}}(M_1) & n_{\text{s}}(M_2) & n_{\text{s}}(M_3) & n_{\text{s}}(M_4) & \dots \end{array} \\ \begin{bmatrix} W^{\text{CS}}_{11}(r) & W^{\text{CS}}_{12}(r) & W^{\text{CS}}_{13}(r) & W^{\text{CS}}_{14}(r) & \dots \\ W^{\text{CS}}_{21}(r) & W^{\text{CS}}_{22}(r) & W^{\text{CS}}_{23}(r) & W^{\text{CS}}_{24}(r) & \dots \\ W^{\text{CS}}_{31}(r) & W^{\text{CS}}_{32}(r) & W^{\text{CS}}_{33}(r) & W^{\text{CS}}_{34}(r) & \dots \\ W^{\text{CS}}_{41}(r) & W^{\text{CS}}_{42}(r) & W^{\text{CS}}_{43}(r) & W^{\text{CS}}_{44}(r) & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \end{array} \tag{4.17}
$$

$$
S^2 N^{\text{pairs}}_{\text{SS2}}(r) = \begin{array}{c} \\ n_{\text{s}}(M_1) \\ n_{\text{s}}(M_2) \\ n_{\text{s}}(M_3) \\ n_{\text{s}}(M_4) \\ \dots \end{array} \begin{array}{c} \begin{array}{ccccc} \text{HOD} & n_{\text{s}}(M_1) & n_{\text{s}}(M_2) & n_{\text{s}}(M_3) & n_{\text{s}}(M_4) & \dots \end{array} \\ \begin{bmatrix} W^{\text{SS2}}_{11}(r) & W^{\text{SS2}}_{12}(r) & W^{\text{SS2}}_{13}(r) & W^{\text{SS2}}_{14}(r) & \dots \\ W^{\text{SS2}}_{21}(r) & W^{\text{SS2}}_{22}(r) & W^{\text{SS2}}_{23}(r) & W^{\text{SS2}}_{24}(r) & \dots \\ W^{\text{SS2}}_{31}(r) & W^{\text{SS2}}_{32}(r) & W^{\text{SS2}}_{33}(r) & W^{\text{SS2}}_{34}(r) & \dots \\ W^{\text{SS2}}_{41}(r) & W^{\text{SS2}}_{42}(r) & W^{\text{SS2}}_{43}(r) & W^{\text{SS2}}_{44}(r) & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \end{array} \tag{4.18}
$$

$$
S(S-1) N^{\text{pairs}}_{\text{SS1}}(r) = \begin{array}{c} \\ n_{\text{s}}(M_1) \\ n_{\text{s}}(M_2) \\ n_{\text{s}}(M_3) \\ n_{\text{s}}(M_4) \\ \dots \end{array} \begin{array}{c} \begin{array}{ccccc} \text{HOD} & n_{\text{s}}(M_1) & n_{\text{s}}(M_2) & n_{\text{s}}(M_3) & n_{\text{s}}(M_4) & \dots \end{array} \\ \begin{bmatrix} W^{\text{SS1}}_{11}(r) & 0 & 0 & 0 & 0 \\ 0 & W^{\text{SS1}}_{22}(r) & 0 & 0 & 0 \\ 0 & 0 & W^{\text{SS1}}_{33}(r) & 0 & 0 \\ 0 & 0 & 0 & W^{\text{SS1}}_{44}(r) & 0 \\ 0 & 0 & 0 & 0 & \dots \end{bmatrix} \end{array} \tag{4.19}
$$

The complete formula to calculate the expected number of galaxy pairs is shown below.

$$
\begin{aligned}
N_{\rm gal}^{\rm pairs}(r) = &\sum_{M_{ij}} n_{\rm c}(M_i) n_{\rm c}(M_j) W_{ij}^{\rm CC}(r) + \\
&\sum_{M_{ij}} n_{\rm c}(M_i) n_{\rm s}(M_j) W_{ij}^{\rm CS}(r) S^{-1} + \\
&\sum_{M_{ij}} n_{\rm s}(M_i) n_{\rm s}(M_j) W_{ij}^{\rm SS2}(r) S^{-2} + \\
&\sum_{M_{ij}} n_{\rm s}(M_i) n_{\rm s}(M_j) W_{ij}^{\rm SS1}(r) S^{-1} (S-1)^{-1}
\end{aligned}
\tag{4.20}
$$

These estimated galaxy paircounts can be combined with analytic randoms to produce the expected correlation function as shown in Equation 4.9.

The analytic expression for the number of random pairs is

$$
RR(r) = \frac{{\rm d}V(r)}{V} N(N-1), \tag{4.21}
$$

where ${\rm d}V(r)$ is the volume of a spherical shell with minimum and maximum radii corresponding to the edges of the radial bin in the clustering calculation, V is the total volume of the simulation box, and N is the expected number of galaxies in the whole simulation volume which can be estimated by applying the HOD to the halo mass function. The number of pairs of randoms vary as the HOD parameters change because the expected total number of galaxies that will be populated is altered.

4.3.3 Halo Subsampling

The halos used in the paircounting were subsampled in order to speed up computation of the tables of the binned paircounts. There are vastly more low mass halos than high mass halos in the halo catalogue and low mass halos are proportionally less occupied by galaxies. This means that HOD fits can be robust to a subsampling of low mass halos. By testing of different subsampling forms and mass cutoffs, it was found that the runtime could be improved by a factor of ten with negligble impacts to the measured number density and clustering. The form of the subsampling was a function of halo mass, with no subsampling at masses above $10^{12} h^{-1} {\rm M}_{\odot}$. The subsample fraction f as a function of

halo mass M is defined as $\log_{10}(f) = \min(1, 2(\log_{10}(Mh\mathrm{M}_\odot^{-1}) - 12))$. Changes to the number densities and clustering of HOD fits introduced by the subsampling were smaller than 0.1%.

4.3.4 Unresolved Halos

A halo mass cut at $10^{11}h^{-1}\mathrm{M}_\odot$ was applied to the AbacusSummit halo catalogues because halos are not sufficiently resolved below this limit. Three satellite tracers were generated around each halo. Randomly selected field particles were used as the locations of halos below this mass cut and these were not given satellite tracers because the satellite contribution from the HOD used in this work is negligible at low masses. Figure 4.3 shows that the clustering of low mass halos is well approximated by using field particles in this manner. The halo bias is close to one for low-mass halos at low redshift which justifies this procedure.

4.4 Emcee and Fitting Procedure

HOD fitting was performed using the EMCEE code, a Markov chain Monte Carlo sampler (Foreman-Mackey et al., 2013). All 17 HOD meta-parameters were fitted at once with target clustering and number densities taken from the BGS mock catalogue described in Smith et al. (2017) at absolute r-band magnitude limits in intervals of 0.5 from -18 to -22. These clustering and number density values were themselves tuned to fit the results from the Sloan Digital Sky Survey (SDSS) (Abazajian et al., 2009) & the Galaxy and Mass Assembly (GAMA) survey (Liske et al., 2015) during creation of the original mock catalogue (Smith et al., 2017). The halo catalogue corresponding to the $z = 0.2$ snapshot was used in all cases.

Clustering bias becomes scale-independent on large scales, this causes a problem if the simulation and target data have different shapes on large scales. This can never be fit well and therefore throws off the full HOD parameter fit. We adjust the target data to match the shape of the simulation large-scale correlation function by using the Zel'dovich clustering prediction at scales above 8 h^{-1}Mpc. In addition a volume correction was used to ensure that the target luminosity function was reproduced despite the changing size of

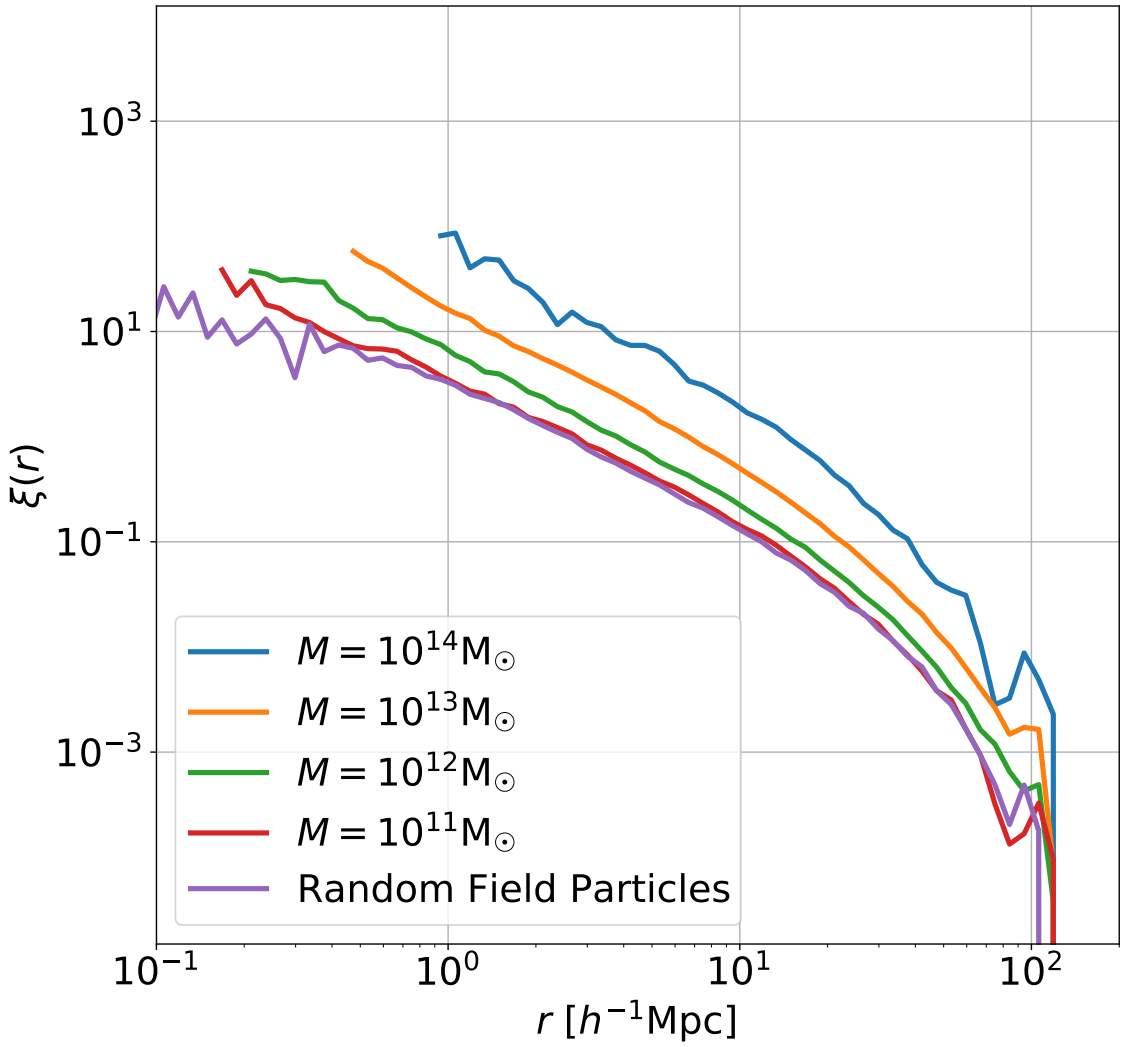

Figure 4.3: The real-space 2-point clustering of field particles along with halos of different masses from the base AbacusSummit simulation. Correlation functions are truncated at small scales where there are insufficient pairs for it to be measured. The clustering of the field particles is similar to that of halos at the low mass limit of the halos, therefore field particles can reasonably be used to extend the HOD below the mass resolution limit of the halos.

volume elements in different cosmologies.

The clustering fit was limited to a maximum scale of 50 h^{-1}Mpc. At scales larger than this Baryon Acoustic Oscillation (BAO) features become important and this shape cannot be fit to by modifying HOD parameters as it is an intrinsic feature of the cosmology.

We did not use a full error covariance matrix when finding the best-fit HOD parameters. This is because we were not trying to establish the likelihood surface for different HODs describing the data, but instead trying to find one set of HOD parameters which produces a good fit to the desired clustering measurements and luminosity function. The likelihood function used in the EMCEE fitting contained a contribution from both the clustering and the number density expected from the HOD parameters,

$$\mathcal{L}(M) = -\frac{1}{2}\left(\sum_r \left(\frac{\xi(r,M) - \xi_t(r,M)}{\sigma_\xi(r)}\right)^2 + \left(\frac{n(M) - n_t(M)}{\sigma_n}\right)^2\right) \tag{4.22}$$

where $\mathcal{L}(M)$ is the likelihood for a particular magnitude limited sample M. The subscript $_t$ is used to represent the target values for the clustering, ξ, and number density, n.

The total magnitude of the likelihood function is not important for the position of the best fit HOD parameters, but it can affect how EMCEE explores the parameter space, if the likelihood function is too steep then the walkers can become stuck in local maxima and if it is too shallow then they may not converge on the best fit solution. Through testing we were able to establish values for $\sigma_\xi(r)$ and σ_n which produced robust fits with the correct balance between fitting the clustering and the number density. $\sigma_\xi(r)$ and σ_n were the same for all nine of the magnitude fits. The form of $\sigma_\xi(r)$ was a power law which emulated the observed variation in the clustering when running the same HOD on AbacusSummit boxes run with different phases. This was to avoid overfitting to noise at very small scales. There is also a cutoff which fixes $\sigma_\xi(r)$ at scales above 0.1 h^{-1}Mpc so that systematic differences that cannot be fit to are not overweighted due to low sample variation at large scales.

4.5 Tests of the Fitting Code

It is important to test the implementation of this tabulation method to see that the galaxy clustering is accurately estimated by weighted halo paircounts. Otherwise the HOD fitting would produce incorrect parameters for the desired galaxy catalogue properties and the mocks would be miscalibrated. In addition, testing the code enables the ideal number of mass bins and subsampling rate to be found, in order to balance accuracy and performance of the code.

4.5.1 Runtime Benchmarks

It is important to balance the performance of the code between accuracy and speed. In these runtime benchmarks, the code was run on one node of the COSMA7 supercomputer that contains 512GB RAM and 28 cores (2x Intel Xeon Gold 5120 CPU).

Using a greater number of mass bins allows the HOD to be sampled more granularly and therefore the weighting factors will be more accurately calculated. However, a larger number of mass bins means that the tables containing binned paircounts grow in size and more instances of the paircounting routine need to be run which increases the overhead in the code. Figure 4.4 shows the relationship between the number of mass bins and runtime of the code. The time to paircount all of the halos is a dominant factor, but does not scale strongly with the number of mass bins. This is because the same number of pairs need to be counted in all cases, but there is only additional sampling and overhead from setting up a greater number of paircounting routines with a greater number of mass bins.

The time to evaluate the galaxy paircounts for a specific HOD does scale strongly across several orders of magnitude with the number of mass bins, due to the $O(N^2)$ size of the paircount table. However, this component of the HOD fitting procedure is rarely a dominant contributor to the overall runtime. The dominant factor in code runtime is the paircounting of the halo catalogue that occurs before the EMCEE fitting.

4.5.2 Convergence Tests of HOD Weighting Factors

The HOD weighting factors must be accurately calculated to produce a correct correlation function and number density. These inform the expected number of galaxies to be produced

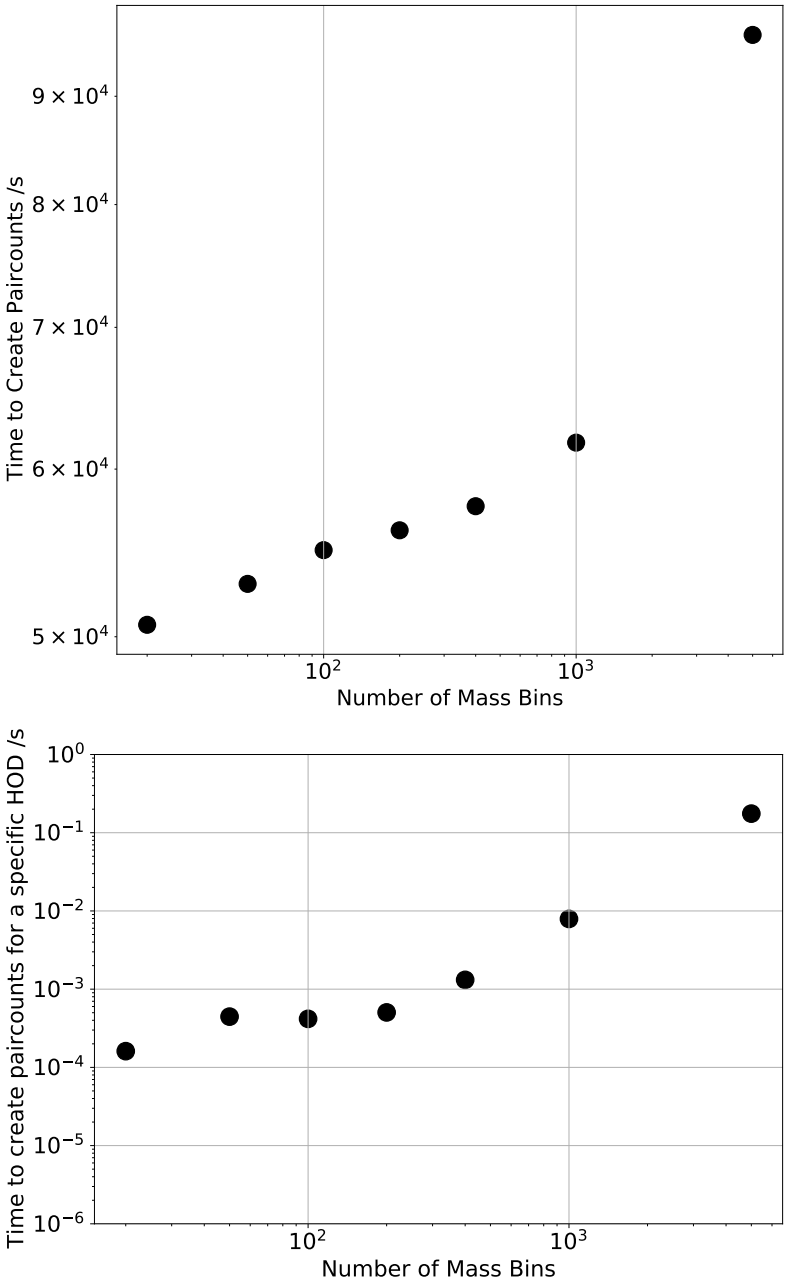

Figure 4.4: The relationship between the number of mass bins and runtime of the halo paircounting code. The upper panel shows the time to count all of the halo pairs as a function of the number of mass bins. Runtime does not change significantly up to around 1000 mass bins but beyond that a large amount of overhead increases the runtime. The overhead is due to a separate paircounting routine needing to be set up for each pair of mass bins. The lower panel shows the time to evaluate the numbers of galaxy pairs from a specific HOD for different numbers of mass bins. Increases become significant above 200 mass bins, however this procedure is often subdominant in the total time to fit HOD parameters.

per halo in a chosen mass bin which affects both the calculation of the number of galaxy pairs and analytic randoms. Runtime constraints necessitate using a small number of mass bins however this can lead to inaccuracies in calculating the number of galaxies because the HOD is not sampled frequently enough.

Equation 4.23 shows a naive method to estimate the number of galaxies within a mass range which may have insufficient accuracy for a small number of bins (large Δ),

$$N_{\rm gal}(M_0 - \Delta < M < M_0 + \Delta) \approx \mathrm{HOD}(M_0) N_{\rm halo}(M_0 - \Delta < M < M_0 + \Delta). \tag{4.23}$$

In order to overcome this issue, the halo mass function and HOD are calculated with a very high number of mass bins. The HOD weighting factor is then calculated by integrating over all of the sub-bins within each larger bin,

$$N_{\rm gal}(M_0 - \Delta < M < M_0 + \Delta) = \int_{M_0-\Delta}^{M_0+\Delta} \mathrm{HOD}(M) N_{\rm halo}(M) dM. \tag{4.24}$$

This produces a more accurate measure of the expected number of galaxies in a given mass range. The effect of integrating the HOD factors can be seen in the convergence of the paircounts in Figure 4.5. The solid lines in Figure 4.5, representing integrated HOD factors, are significantly closer to the converged case than the dashed lines, representing non-integrated HOD factors. Integrated HOD factors are also necessary to produce accurate paircounts at large scales.

4.5.3 Paircount Accuracy

This section contains results showing that the galaxy paircounts are correctly predicted as a function of distance. The paircounts are split up into several components, central-central, central-satellite, and satellite-satellite (both one- and two-halo terms). All of these expected paircounts can be compared to the paircounts arising from realisations of the mock to check their accuracy.

A set of 50 mock galaxy catalogues were created in order to reduce sample variance. The galaxies were split into central and satellite categories and the respective paircounts were calculated as a function of distance.

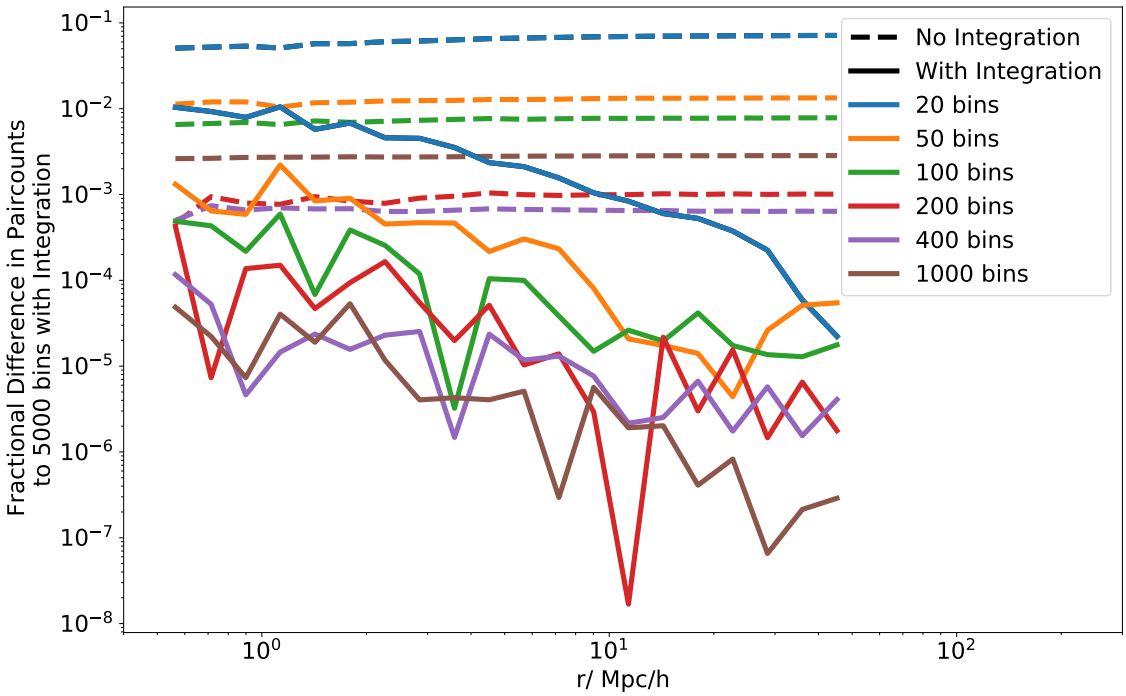

Figure 4.5: Convergence testing of central-central paircount number with the number of mass bins, and a comparison of integrated (solid lines) vs non-integrated (dashed lines) HOD factors. 5000 mass bins with integrated HOD factors is used as the reference from which differences are measured. Integration of the HOD factors is necessary for consistency with high resolution mass binning at large scales.

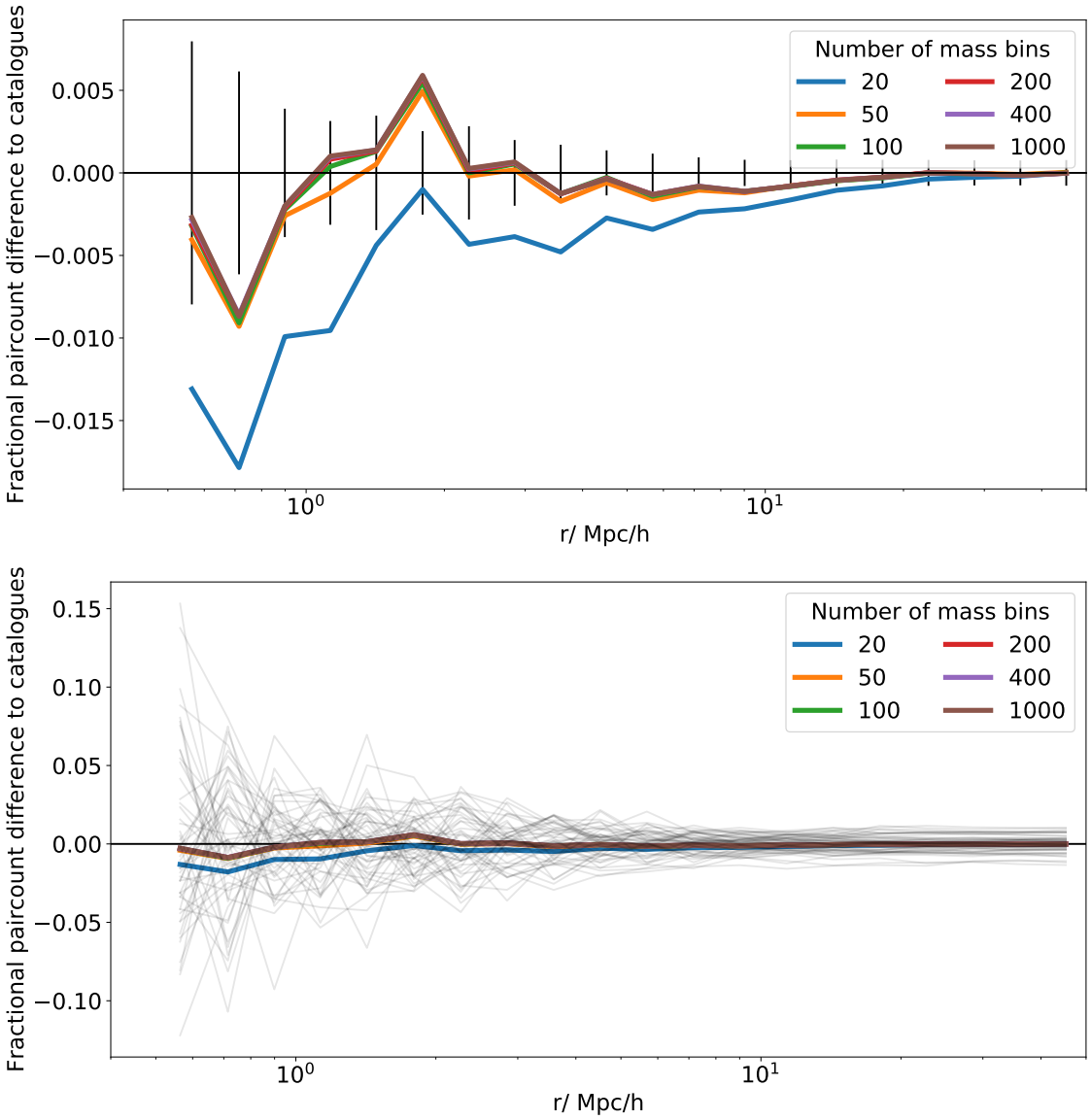

Figure 4.6: A comparison of the expected central-central paircounts to the mean from 50 catalogue realisations. The upper panel contains vertical bars indicating the standard error in the mean from the galaxy catalogues. The lower panel shows the same data but with the paircounts from each individual realisation of the galaxy catalogue shown as a grey line indicating the level of variation in the realisations.

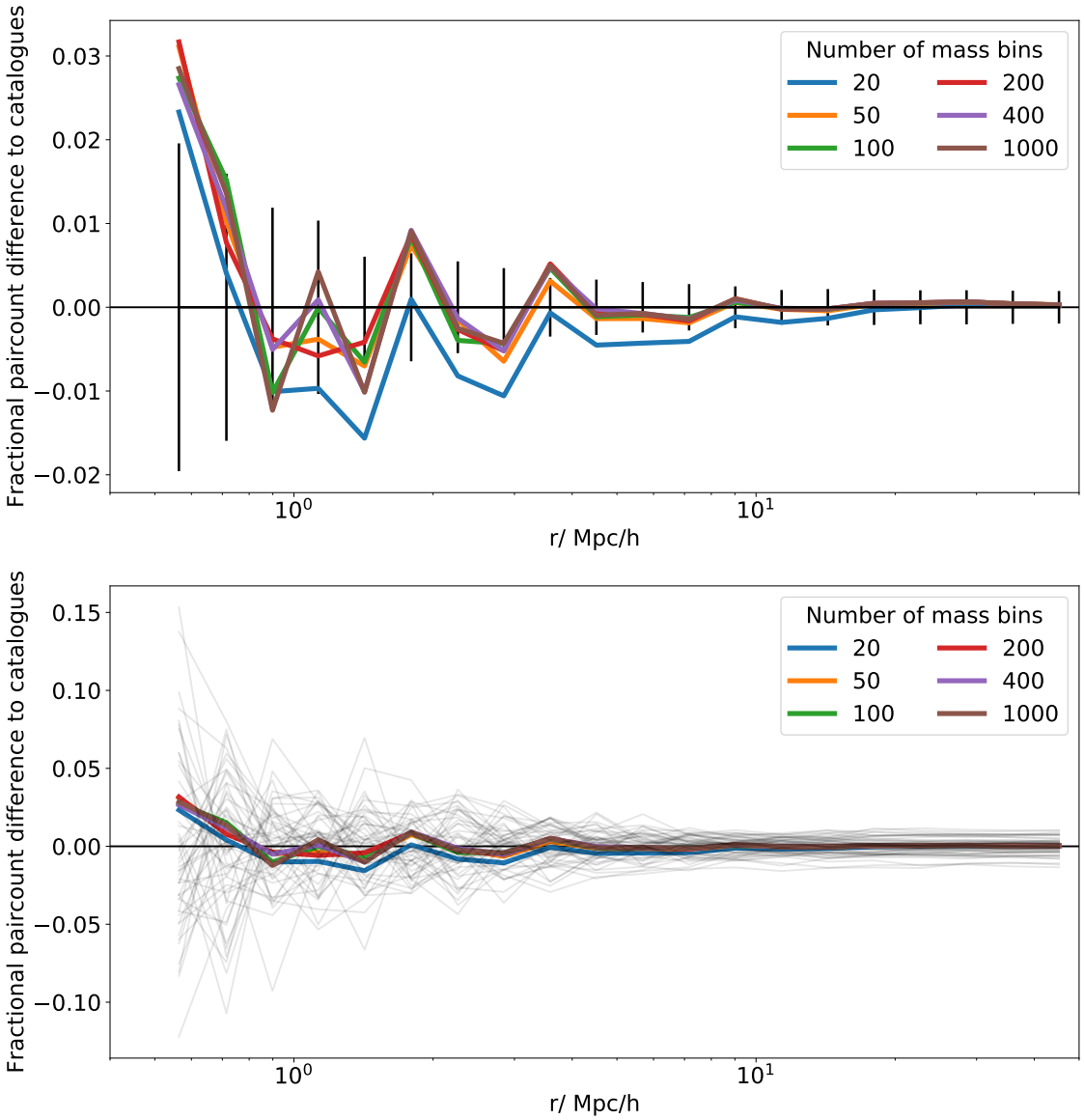

Figure 4.7: A comparison of the expected central-satellite paircounts to the mean from 50 catalogue realisations. The upper panel contains vertical bars indicating the standard error in the mean from the galaxy catalogues. The lower panel shows the same data but with the paircounts from each individual realisation of the galaxy catalogue shown as a grey line indicating the level of variation in the realisations.

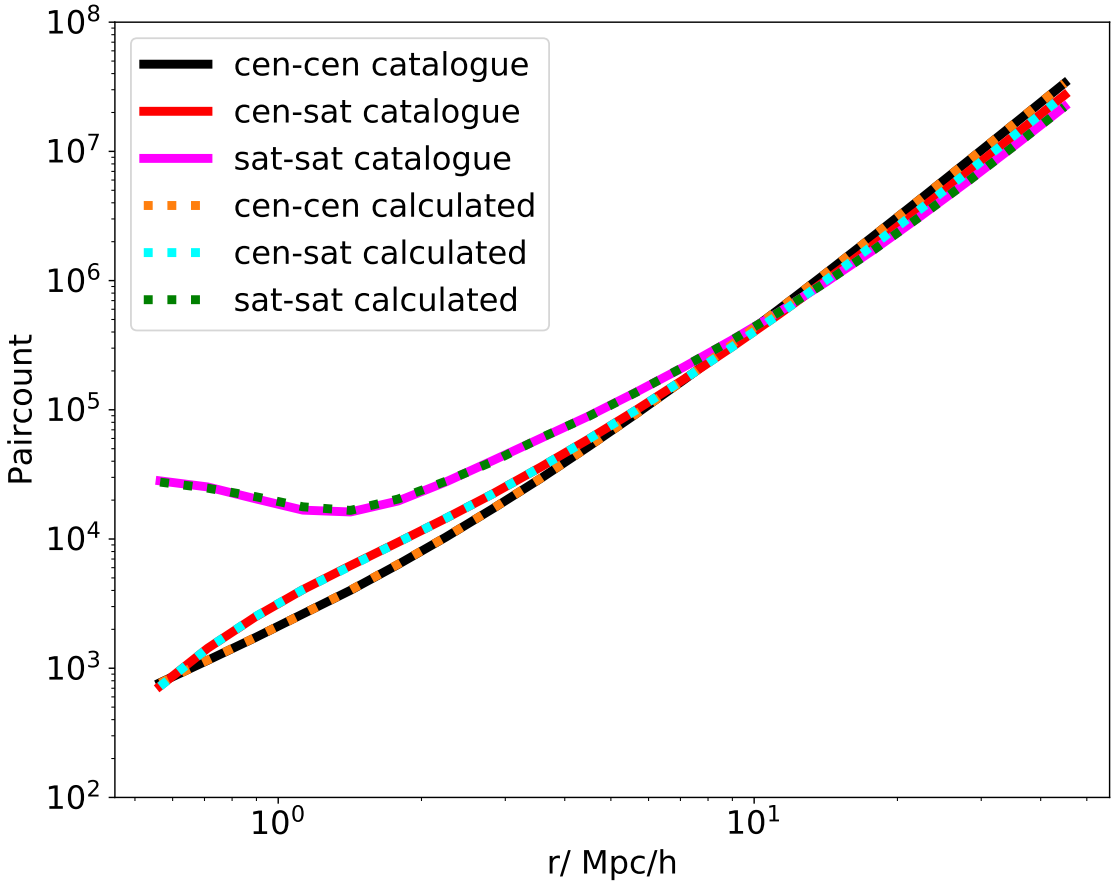

Figure 4.8: Paircount contribution from the different components in both the predicted case and measurements from galaxy catalogues. The paircounts from different central and satellite components are accurately calculated. The sat-sat component is accurate at both small- and large-scales, where the one-halo and two-halo terms dominate respectively.

Figure 4.6 shows the comparison of the predicted number of central-central pairs to the actual paircounts from catalogue realisations as a function of the number of mass bins. The expected number of pairs is generally consistent with the mean of many realisations, the larger difference at small scales is compensated by greater variance in the realised catalogues. The lower panel indicates that the expected paircount number from tabulated halo paircounting is much more consistent with the mean number of galaxy pairs than an individual galaxy catalogue realisation would be.

The same comparisons are shown for central-satellite paircounts in Figure 4.7. The expected central-satellite paircounts are consistent with the paircounts from catalogue realisations.

Figure 4.8 shows the calculated paircounts from the tabulation method and those

measured from galaxy catalogues as a function of separation. Contributions are separated by the central & satellite combinations of paircounts. The number of paircounts spans many orders of magnitude and is well approximated by the tabulation method predictions.

4.6 Best-Fit HOD parameters for AbacusSummit BGS Mock Catalogs

This section contains results for the HOD fitting procedure applied to the AbacusSummit suite of simulations. A description of the mock catalogues created using these parameters can be found in the following chapter. Separate HOD fits were created for each of the 100 cosmologies present in the simulations.

4.6.1 Best-fit Meta-parameters

The output of the fitting procedure is not the HOD parameters directly, but a set of meta-parameters describing the evolution of smooth curves describing the HOD as a function of absolute magnitude limit. Figure 4.9 shows a cornerplot of all 17 parameters described in Equations 4.4-4.8 in the case of the base AbacusSummit cosmology. The cornerplot shows the parameter space explored by the EMCEE fitting chain after an appropriate amount of burn-in is discarded. The best-fit values are indicated using red lines. Correlations between parameters indicate that degeneracies exist as the parameters vary along that axis.

The majority of the meta-parameters are well constrained, with P_3, P_8, and P_9 being relatively less constrained. Reference to Equations 4.5-4.6 explains the lack of constraint, because changes in these parameters have very little effect on the shape of the curve in certain regimes. In addition there is a long negative tail on P_1 and P_2, this is because M_0 does not affect the shape of the HOD curve if it is at sufficiently low mass that it applies when satellites have little impact.

4.6.2 Best-Fit HOD Curves

The shape of the HOD curves that are produced from the best-fit meta-parameters in the base cosmology case is shown in Figure 4.10. Curves are shown for a range of magnitude

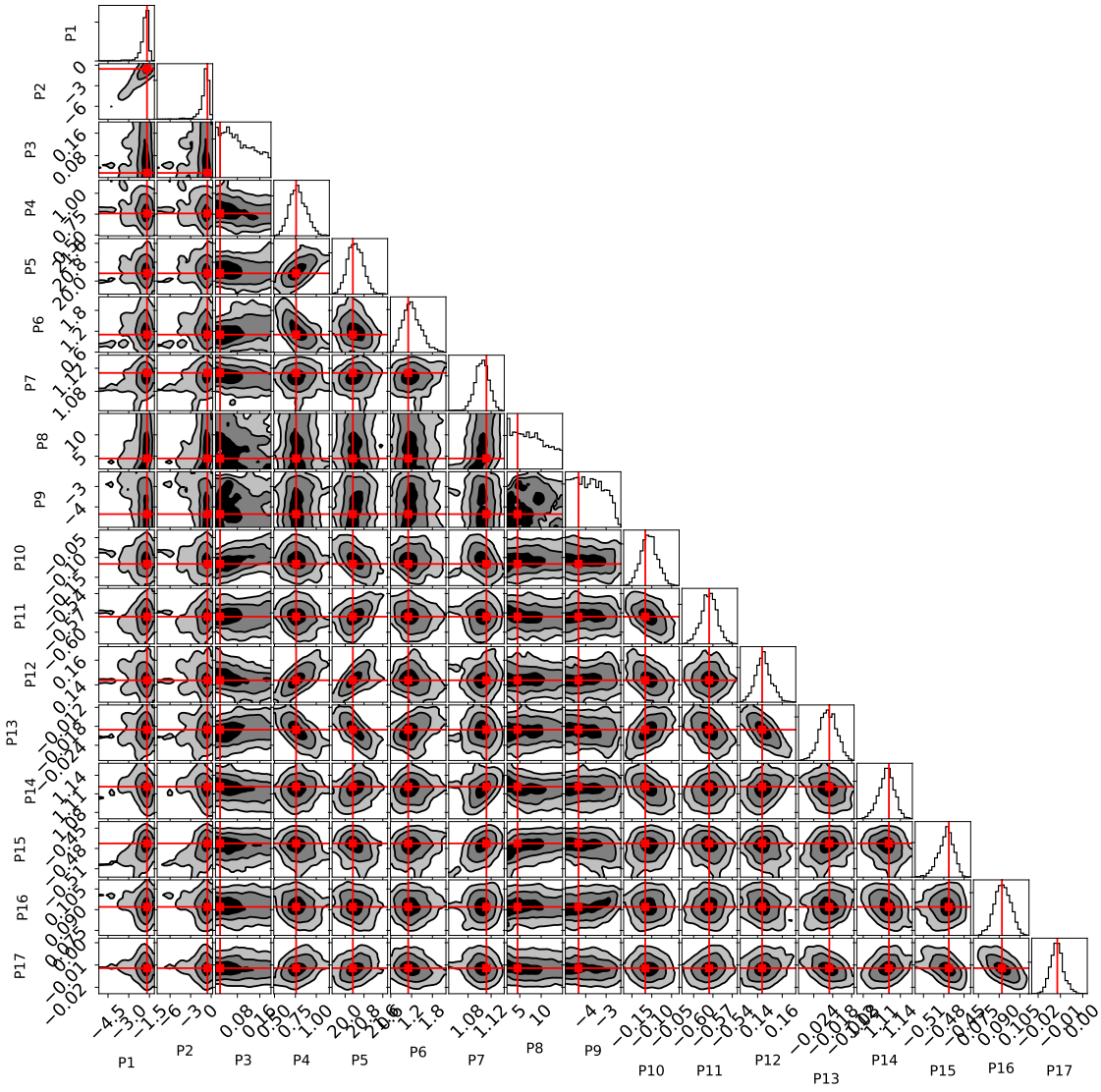

Figure 4.9: A corner plot from the EMCEE fitting chain of the 17 meta-parameters describing the smoothly varying HOD curves. Best-fit values are shown in red. This example comes from the primary AbacusSummit cosmology.

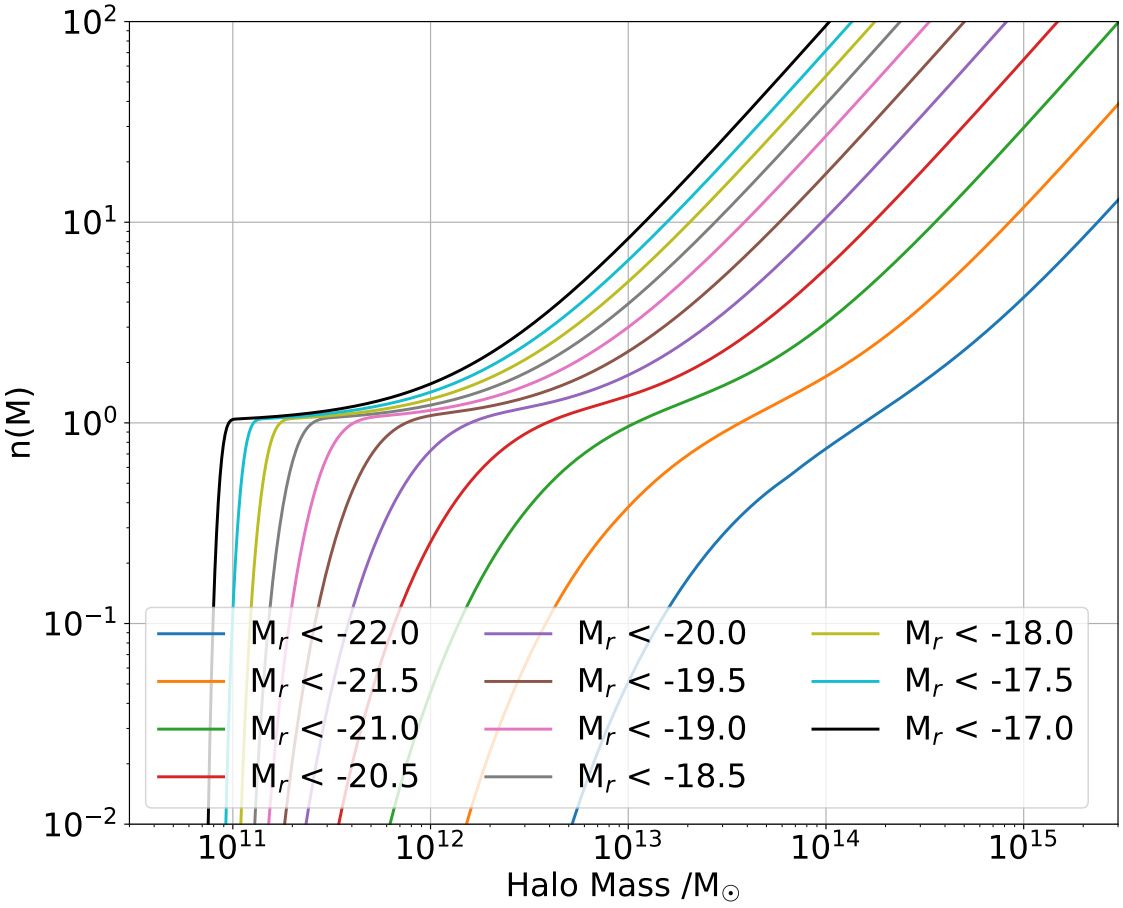

Figure 4.10: Best-fit HOD curves for the base AbacusSummit cosmology. Note that there is no overlap of the curves at any point.

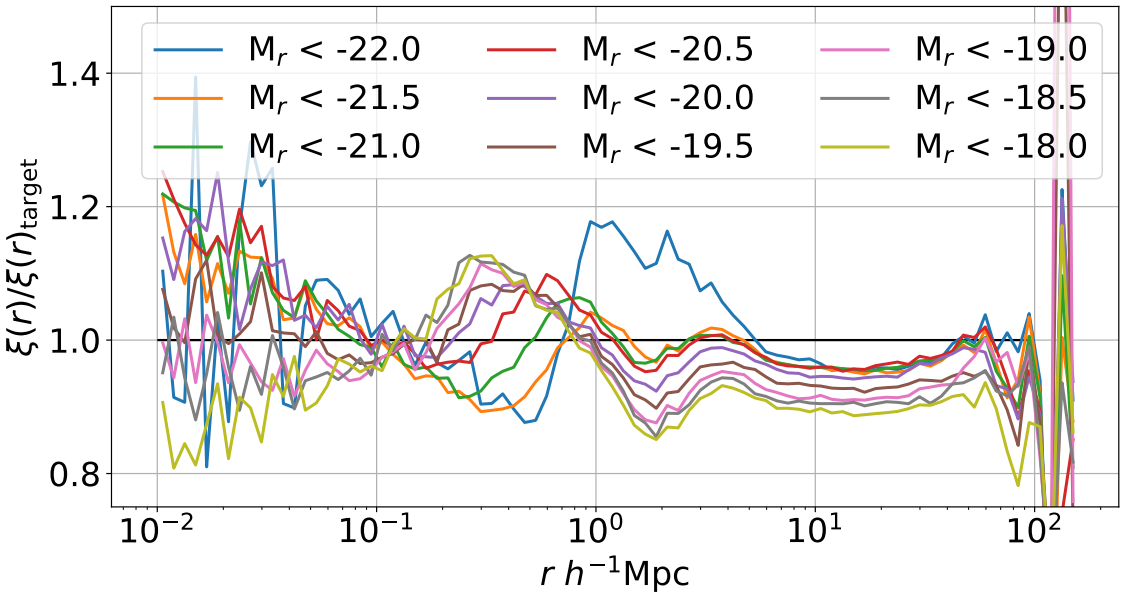

Figure 4.11: The ratio of the best-fit predicted clustering of samples with different magnitude limits to the target data for the base cosmology. There is an offset at intermediate scales where the HOD produces clustering that is around 5-10% lower than the target value. There are features in the residuals at smaller scales around the onset of the one-halo term, this is most strongly shown for the brightest sample.

limits indicated by different colours. The contributions from both the central and satellite HOD components are combined into a single curve. There are no unphysical overlaps between the HOD curves as this is disallowed by the fitting procedure.

4.6.3 Comparison to Target Data

In addition to exploring the values of the best-fit meta-parameters and shape of the HOD curves produced from them, we have also compared the quality of the best-fit with regards to the target data. Since we have fit to target data for both the clustering and the number density simultaneously, the quality of both of these aspects of the fit should be investigated.

Figure 4.11 shows the ratio of the predicted two-point clustering for the base cosmology to the target measurements for a range of magnitude limits. There is an offset at large scales where the predicted mock clustering is lower than the target clustering by between 5-10%. The features above 50Mpc are outside the range of the fitting because these are caused by BAO which cannot be mitigated by varying HOD parameters. Different behaviour is seen in the residuals to the target clustering at smaller scales depending on the magnitude limit

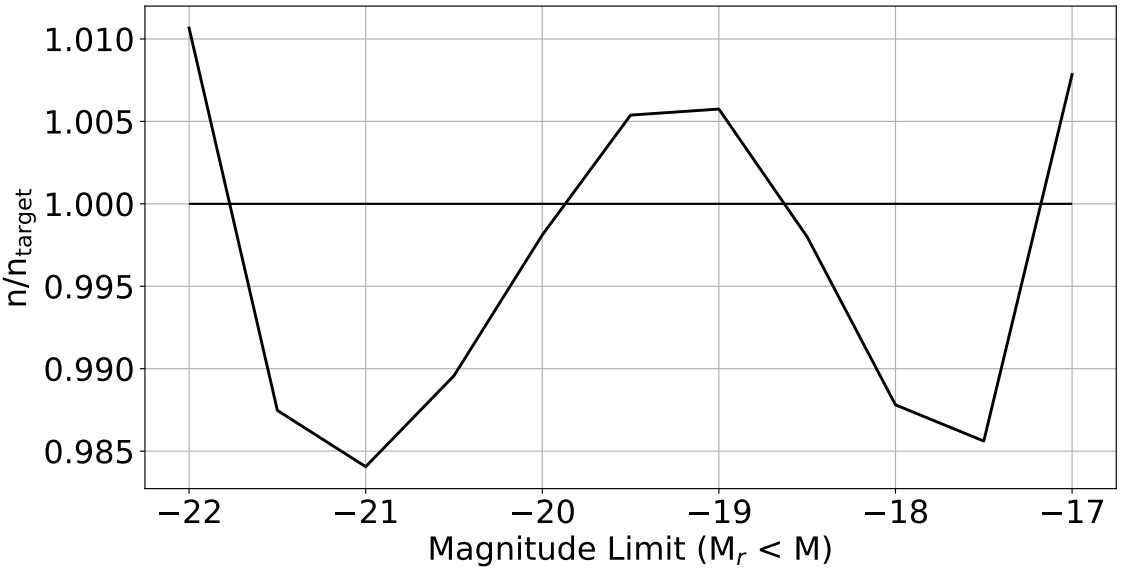

Figure 4.12: The ratio of the predicted number density of samples with different magnitude limits to the target values for the base cosmology. The number density does not differ from the target by more than 2% for any of the samples included in the fitting.

of the sample. The brightest sample displays a peak relative to the target clustering above 1 h^{-1}Mpc while the faintest samples display a peak around 0.3 h^{-1}Mpc.

In Figure 4.12, the ratio of the predicted number density to the target value is shown as a function of the magnitude limit of the sample. There is good agreement for all the samples that were included in the fitting procedure, with the number density differing by less than 2% in all cases.

4.7 Conclusions

In this chapter we have explored a method to quickly produce HOD fits that match a target luminosity function and luminosity dependent clustering across a range of different samples.

Smooth curves for HOD parameters are defined as a function of magnitude. It is the best-fit meta-parameters describing these curves that are produced instead of the HOD parameters directly. This allows the HOD to be defined for samples with any magnitude limit and a constraint such that no overlap between HOD curves representing different magnitude limits is integrated into the fitting procedure.

Halo paircounts are subsampled and tabulated as a function of halo mass. Weighted

sums of these paircounts act as an accurate estimate of the galaxy paircounts produced by applying an HOD to the halo catalogue. The weighted sums are quick to evaluate which speeds up the HOD fitting substantially. A single HOD fit takes approximately 8 hours, including the halo paircounting.

Dark matter halo catalogues from the AbacusSummit simulations are used to produce these HOD fits, however this method could be used on other halo catalogues too. Separate fits have been made for each cosmology.

The next chapter shall describe the method of producing mock galaxy catalogues from these HOD fits and comparisons to measured BGS data.

Further work could apply this method to other galaxy samples and simulation catalogues. In addition, extending the tabulation to use halo properties beyond halo mass could allow the HOD models to better account for assembly bias.

CHAPTER 5

Bright Galaxy Survey Mock Catalogues - HOD Variation, Mock Creation & Analysis

5.1 Introduction

As introduced in the previous chapter, analysis of the galaxy-halo connection is important to understand the link between simulations and observations of large-scale structure. Accurate modelling of the galaxy-halo connection allows us to turn dark-matter-only simulations into realistic simulated catalogues of galaxies. These mock catalogues are an essential part of the toolkit for analysis of cosmological surveys, and they are used in a variety of ways.

Mock galaxy catalogues can aid in the design of new galaxy surveys, providing evidence that particular survey strategies will achieve targeted scientific goals and helping to assess the impact of trade-offs in survey design (Looser et al., 2021). In addition, mocks are needed to test and optimise the key data analysis pipelines that are designed to handle the large volumes of data produced by modern surveys. Mock catalogues can be used to test the recovery of key statistics from survey data, investigating potential systematic errors and ensuring the unbiased measurement of cosmological parameters.

Mock catalogues are used to produce covariance matrices of galaxy power spectra &

clustering that are necessary to understand uncertainties in these measurements and the cosmological parameters that can be deduced from them (Lippich et al., 2018; Blot et al., 2019).

For covariance matrix applications, many thousands of mocks may be needed in order to produce sufficient accuracy (Dodelson & Schneider, 2013). High-resolution N-body simulations are too computationally expensive to simulate the volumes required therefore approximate methods or low-resolution simulations are used. Examples include Chuang et al. (2014), Avila et al. (2015), Klypin & Prada (2018), and Hernández-Aguayo et al. (2021).

These approximate methods are not sufficient for many other applications of mock catalogues where greater accuracy is required at small scales. In these cases high-fidelity mocks are created from the outputs of high-resolution N-body simulations.

There are different types of mock galaxy catalogue. In this chapter we discuss both cubic box and full-sky catalogues. Cubic box mocks place galaxies in a 3-dimensional periodic cube and do not simulate observational effects. This produces a large volume with which to test a model and one does not have to worry about additional complications related to survey geometry. Full-sky/cut-sky mocks simulate what the DESI BGS will observe. Galaxies are placed on the sky with redshifts and transformations are made into the observer's frame of reference. This realistically accounts for systematic effects on survey measurements. Full-sky/cut-sky mocks may be made from a single snapshot or a true lightcone. The true lightcone includes evolution of the simulation and dark matter halos with lookback time.

In this chapter we explore how varying the simulation cosmology affects the HOD fits. The process of producing mock galaxy catalogues for the Bright Galaxy Survey using the AbacusSummit simulations is then presented. Comparisons between the mocks and BGS data are shown.

The process of producing mock catalogues in this chapter closely follows the work in Smith et al. (2017) and Smith et al. (2022) which is referenced extensively.

5.2 HOD Fits in Varied Cosmologies

Before describing the production of the mock galaxy catalogues, we shall firstly explore the variation in HOD parameters that were used to produce the mocks. HOD parameters were fit using the method detailed in Chapter 4. A separate HOD fit was performed for simulation boxes that were created with varying cosmologies.

We want to make mocks in different cosmologies because that allows us to test the impact of cosmological parameters on mock observations. Cosmological parameter recovery is an important application of mock catalogues that requires mocks in different cosmologies. Exploring how the HOD parameters vary with cosmology informs us how our models of the galaxy-halo connection may change

An initial check that we performed was to find the level of sample variance that exists for HOD fits on multiple independent simulation boxes in the same cosmology. There are 25 AbacusSummit simulations in the base Planck ΛCDM cosmology (see Section 4.1.2) that were produced from initial conditions with different phase information. If there are large differences in best-fit HODs produced by the fitting procedure on these boxes, then it implies that there is a lack of robustness in the fitting procedure.

Figure 5.1 shows the variation in the best-fit HOD parameters from each of the base cosmology simulations with different phases. The largest variation is in M_0 at faint magnitude limits. In this region the M_0 parameter has limited or no effect on the shape of the HOD therefore this variation is to be expected. The best-fitting HODs that are produced from these parameters are shown in Figure 5.2. There is a high degree of consistency between the HODs. Differences between HODs emerge at the low mass tails where there are very few galaxies per halo. HOD differences at these scales have a small effect on the properties of galaxy catalogues because of the low occupation. Certain points have very little scatter. These locations are where the amplitude of the HOD has the greatest effect on the number density, which is constrained in the fitting process.

Figures 5.3 and 5.4 show the variation in HOD parameters and HOD shapes using simulations with different cosmologies. The level of variation in cosmological parameters is shown in Figure 4.1. The variation in HOD parameters and shapes is larger than the sample variance when the same cosmology is used. The cosmology emulator grid is sufficiently wide to lead to significant differences in the best-fit HODs. This justifies our

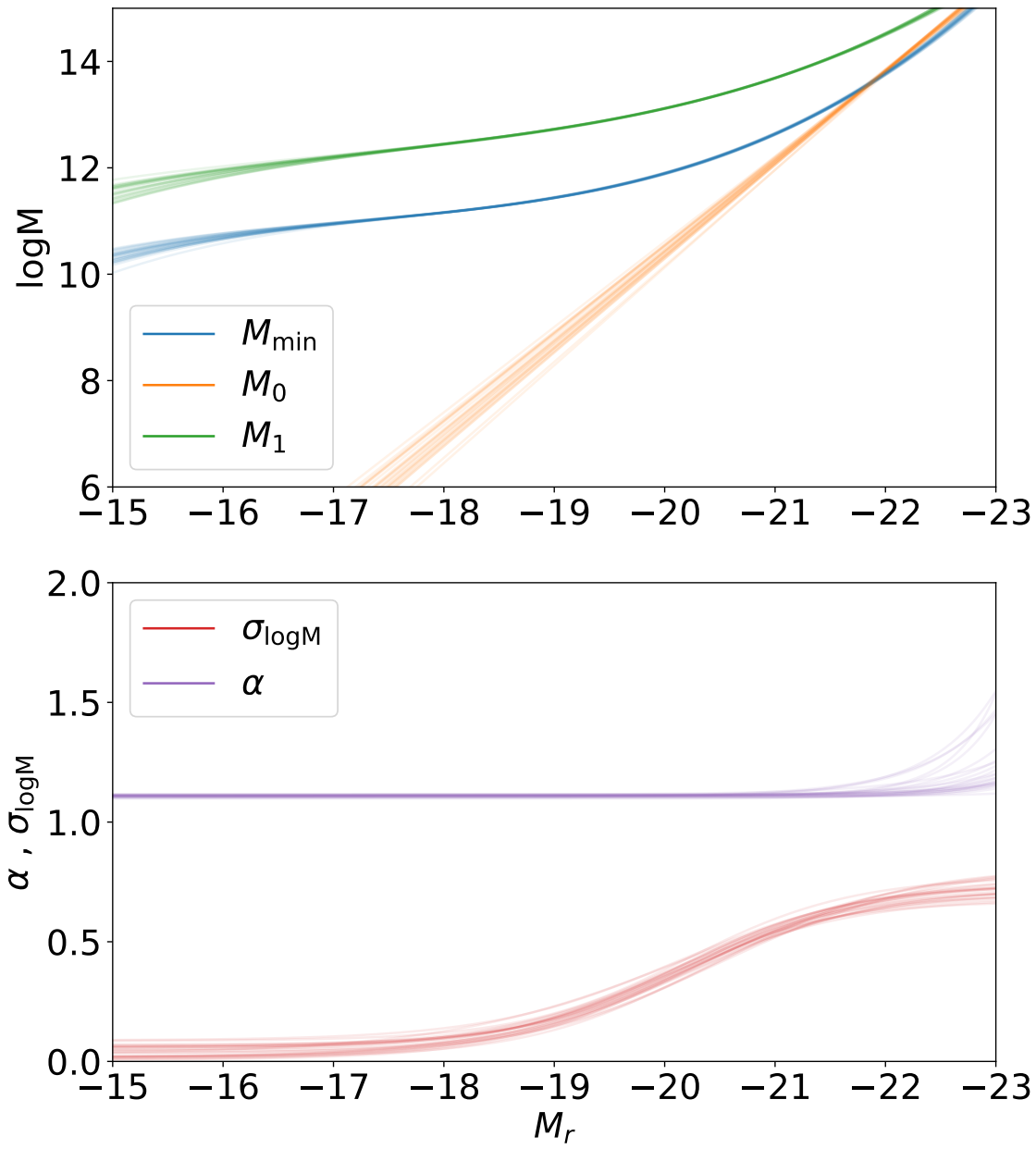

Figure 5.1: Variation of the HOD parameters as a function of absolute magnitude when using simulations with the same cosmology and different random seeds. Each line represents a single HOD fit. The top panel shows the mass parameters $M_{\rm min}$, M_0, and $M1$ on a logarithmic scale. The lower panel shows the shape of the curves describing the non-mass parameters, α, and $\sigma_{{\rm log}M}$.

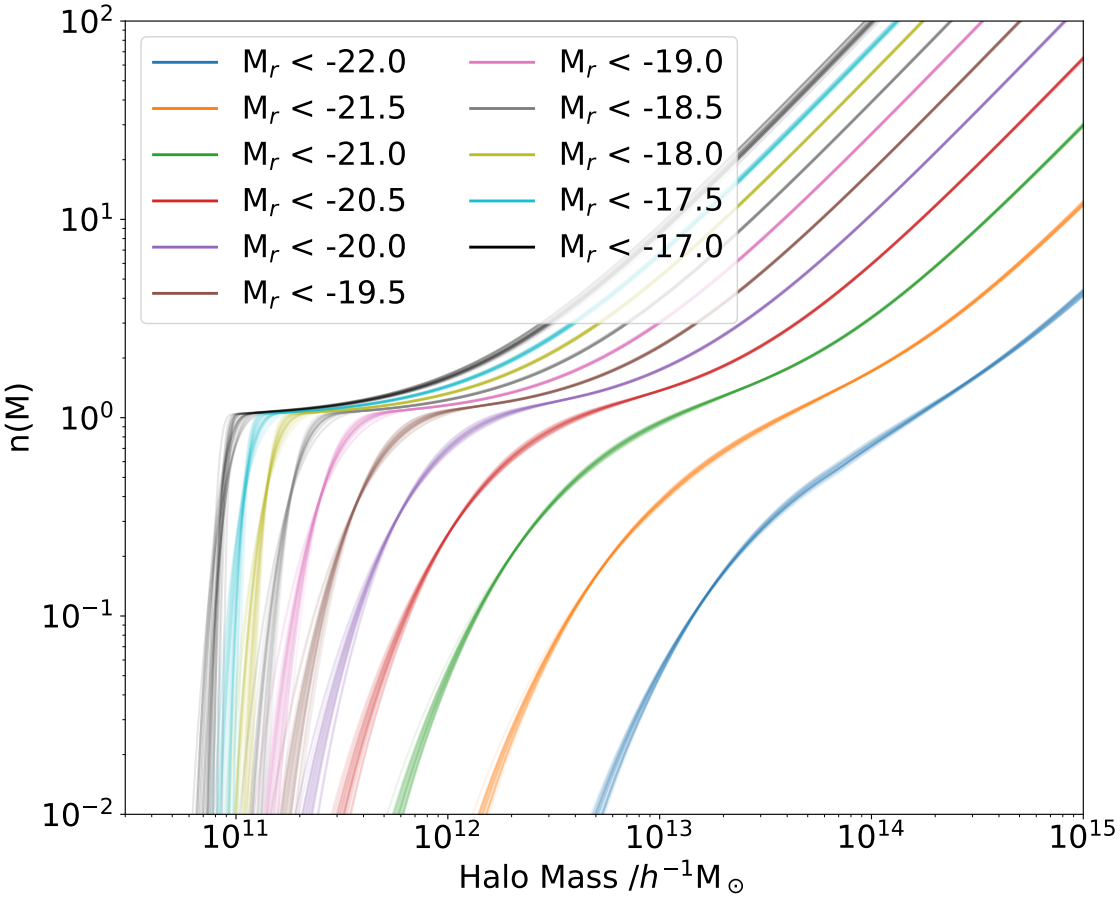

Figure 5.2: Variation of the best-fit HODs when using simulations with the same cosmology. Each line represents a single HOD fit. The HOD curves are highly consistent with one another for all the samples, with significant differences emerging only at the low mass tails.

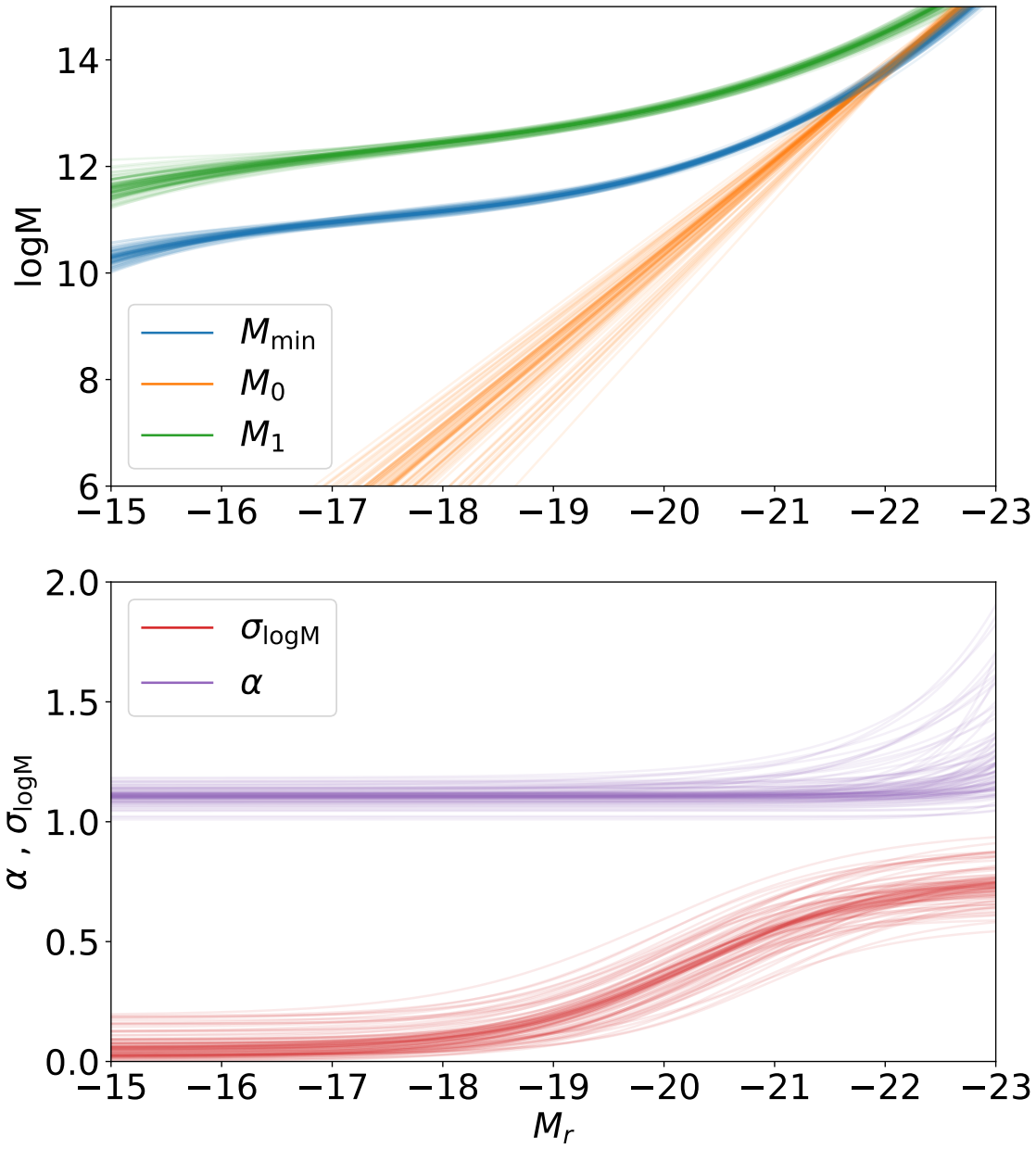

Figure 5.3: Variation of the HOD parameters as a function of absolute magnitude when using AbacusSummit simulations with different cosmologies. Each line represents a single HOD fit. The top panel shows the mass parameters $M_{\rm min}$, M_0, and M_1 on a logarithmic scale. The lower panel shows the shape of the curves describing the non-mass parameters, α, and $\sigma_{\log M}$. Comparing to Figure 5.1 there are much larger variations than can be explained by sample variance alone.

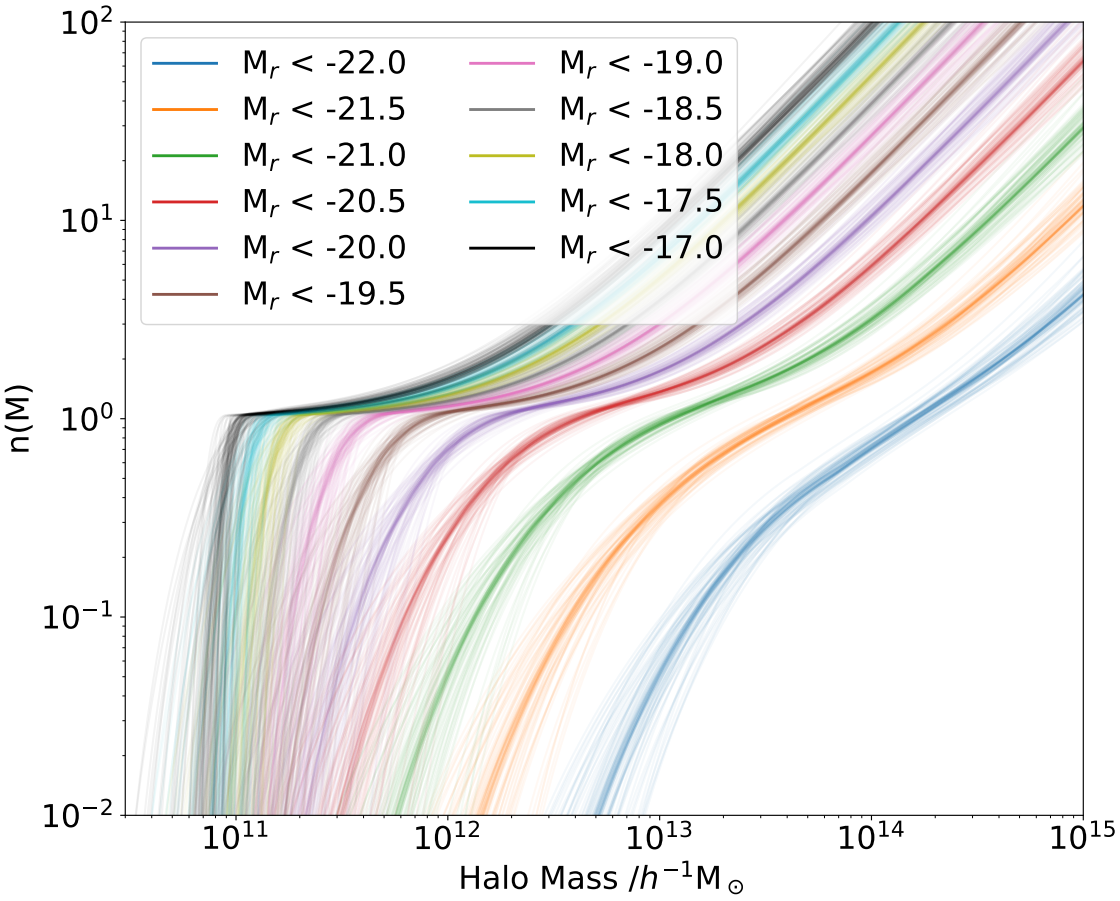

Figure 5.4: Variation of the best-fit HODs when using AbacusSummit simulations with different cosmologies. Each line represents a single HOD fit. The HOD variation is larger than the case where cosmology is held constant as in Figure 5.2.

choice to refit the HODs for simulations that used different cosmologies.

Best-fit HOD parameters are correlated to the cosmological parameters of the simulation. Figure 5.5 shows an example for σ_8 and α. A negative correlation between σ_8 and α can be seen. In addition, the increase in α from faint to bright samples is larger for small values of σ_8.

σ_8 sets the amplitude of the initial power spectrum at a fiducial scale of $8h^{-1}$Mpc. Simulations run with higher σ_8 will produce a halo catalogue with stronger clustering. Therefore, in order to reproduce the same target clustering, the galaxy clustering bias must be reduced in these cases. A method to decrease the clustering strength of galaxies is to lower the number of satellite galaxies because they are located in the highest mass (and hence most biased) halos. Therefore it makes intuitive sense that σ_8 should correlate with lower α, if M_0 and M_1 do not change significantly. In Figure 5.6, we can see the effect of simulation σ_8 on the HOD curves. Higher σ_8 is associated with fewer satellite galaxies for the reasons that have been stated.

There are too many potential combinations of cosmological and HOD parameters to investigate all the relationships in this way. These variations in HOD parameters with cosmology may be useful if one wanted to build an emulator for HOD parameters. Such an emulator could produce an estimated best-fit HOD for any cosmology that lies within the region spanned by the AbacusSummit cosmologies.

5.3 Cubic Box Mock Catalogues

In this section the creation of cubic box mock catalogues is presented. Galaxies are given 3D positions in the cubic simulation box. The nested set of HODs that we have determined are used to assign each galaxy a luminosity, as in Smith et al. (2017) and described below. In addition, the semi-empirical method of Smith et al. (2017) is used to assign each galaxy a rest frame $(g - r)$ colour. These mock catalogues are not in the same format as data observed from DESI, because DESI only sees galaxies with projected locations and redshifts on a fraction of the sky. However, they can be processed into more realistic formats. This type of mock is a precursor to creating full-sky mock catalogues that convert the 3D positions into sky coordinates and redshifts using a simulated observer.

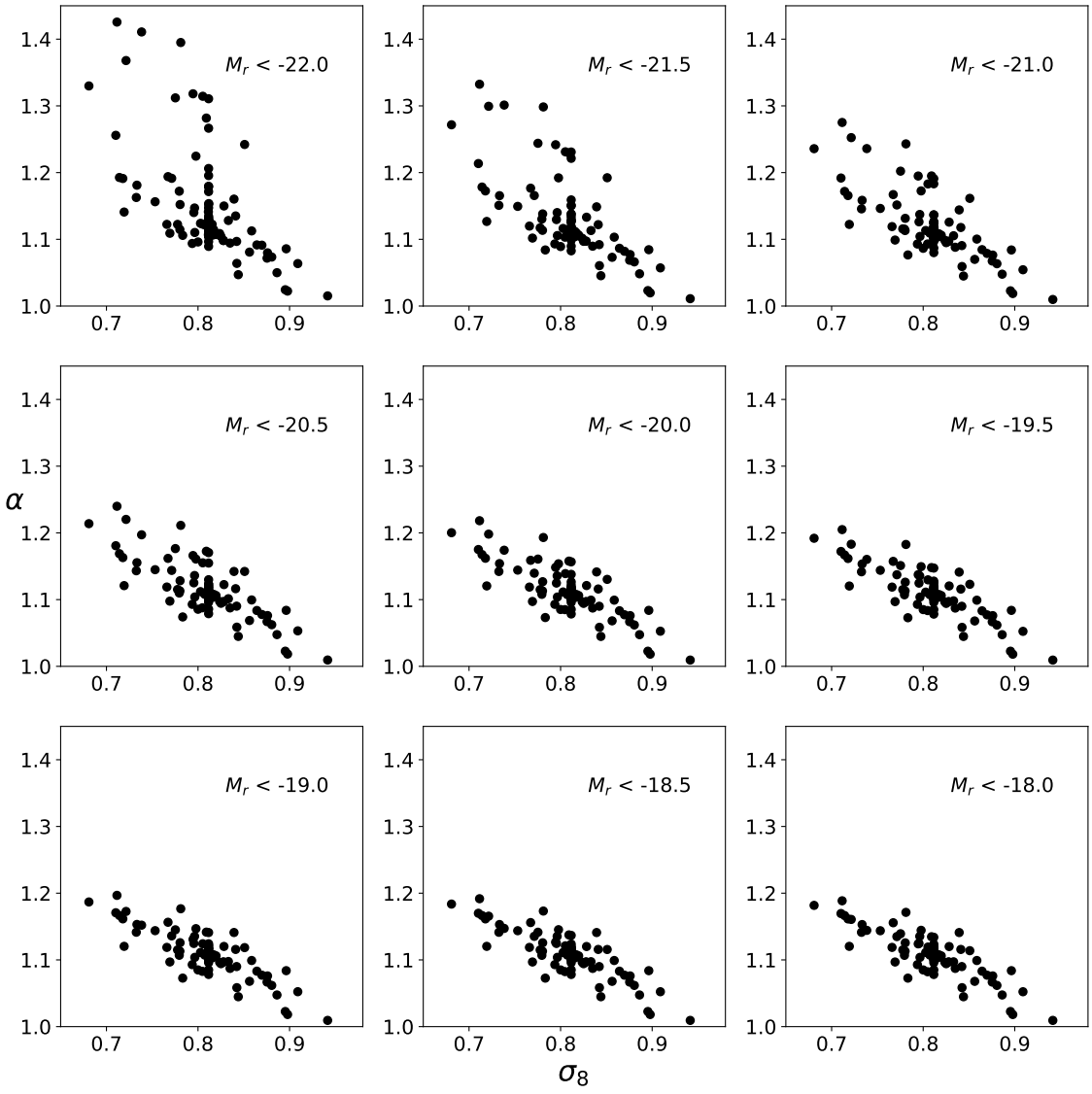

Figure 5.5: The simulation value of σ_8 vs the best-fit value of α. Individual panels show the results for each absolute magnitude sample. There is a negative correlation between α and σ_8 and low values of σ_8 are associated with a greater increase in α in bright samples.

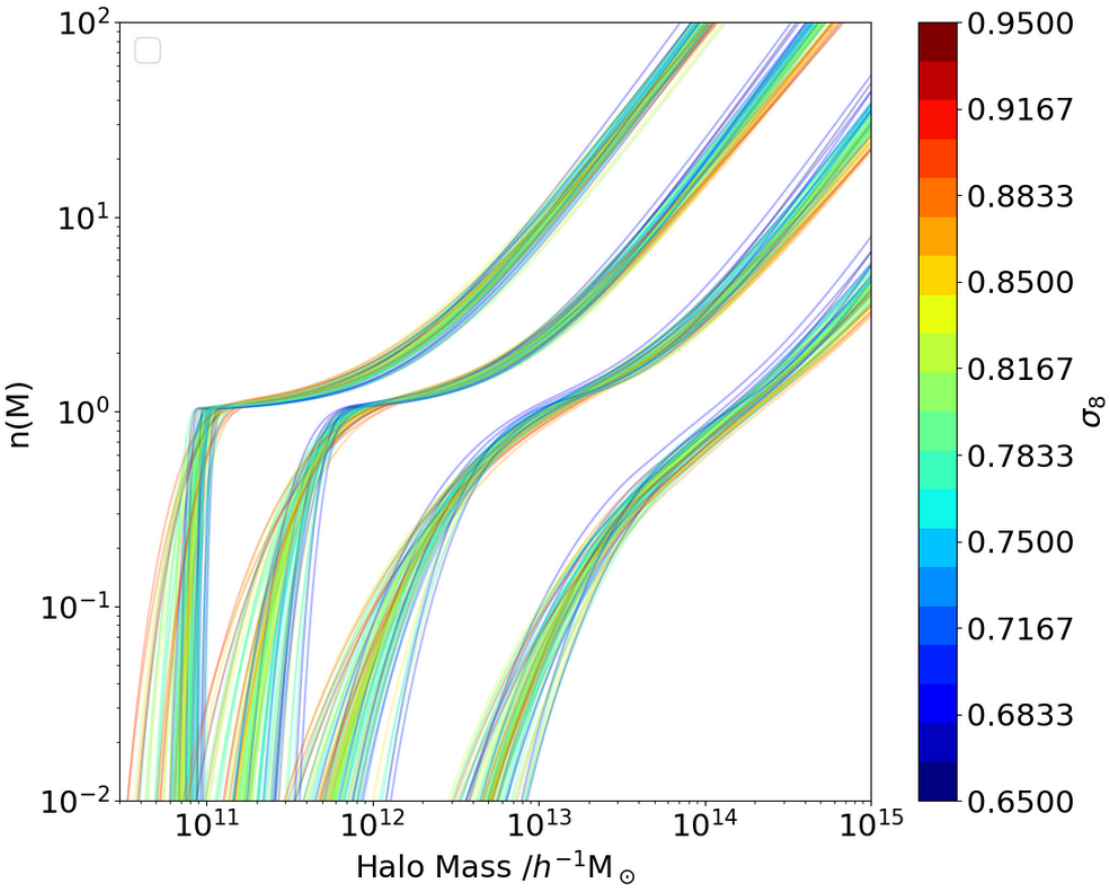

Figure 5.6: The variation in HOD curves, coloured to show the simulation value of σ_8. For clarity, only the curves representing $M_r < -22, -21, -19.5$, and -17 are shown. One can see the effect of a negative correlation between σ_8 and α in the high mass end of the HOD curve, where the satellite power law dominates. Higher σ_8 causes the best-fit HOD to produce fewer satellite galaxies.

In order to create a mock galaxy catalogue, one needs to combine the halo catalogue with the best-fit HOD that best reproduces the target luminosity function and clustering. Different mock galaxy catalogue realisations may be created from the same halo catalogue and HOD by altering the random seed used, in this work we only create one realisation of each mock galaxy catalogue.

5.3.1 Galaxy Positioning and Luminosity

The AbacusSummit halo catalogue corresponding to the cubic box snapshot at $z = 0.2$ is used to create the mock galaxy catalogues. This is because $z = 0.2$ is close to the median redshift of the DESI BGS. The probability of a halo containing a central galaxy brighter than a chosen luminosity is defined by Equation 4.1. This is a smoothed step function of halo mass, the parameters of this function depend on the luminosity limit chosen. Central galaxies are placed at the centre of the dark matter halo. The centre of a dark matter halo in AbacusSummit is defined as the location of the centre of mass of the largest sub-halo. Field particles are used to place central galaxies that would occupy halos below the mass cut at $10^{11}h^{-1}\mathrm{M}_{\odot}$ as described in Section 4.3.4. Each central galaxy is given the same velocity as its parent halo.

A modified version of the method from Skibba et al. (2006) is used to assign luminosities to galaxies. There is scatter in the luminosity of central galaxies from halos at a constant mass because the HOD describing central occupation above a magnitude is not an infinitely sharp step function as in Skibba et al. (2006). To produce this scatter, a random number, x, is sampled from a probability distribution described by the spline function in Equation 4.2. The luminosity of the central galaxy can then be found by solving the equation

$$\frac{x\sigma_{\log M}(L)}{\sqrt{2}} = \log M - \log M_{\mathrm{min}}(L), \tag{5.1}$$

where the factor of $\sqrt{2}$ comes from the definition of $\sigma_{\log M}$. This is the same as the method described in Smith et al. (2017). The condition that HOD curves cannot cross is important for this process, because it guarantees a unique solution for L, given any x and M. Interpolation of the functions $\sigma_{\log M}(L)$ and $M_{\mathrm{min}}(L)$ is used to solve for L.

Satellite galaxies are placed randomly according to an NFW profile around each halo. The number of galaxies is sampled from a Poisson distribution with the mean value calculated from the satellite HOD function shown in Equation 4.3. A minimum luminosity cut, $L_{\rm min}$ is used so that a vast number of faint satellite galaxies are not created. Satellite galaxy velocities are assigned a random virial velocity relative to the velocity of the central halo. Satellite galaxies are also populated around the central galaxies placed on field particles which represent halos below the resolved halo mass cut. However, the probability of the unresolved halos containing satellite galaxies is very low.

A slice through one of the cubic box mocks that shows galaxy positions is shown in Figure 5.7.

Satellite galaxy luminosities are assigned in a similar way to the central galaxies. This method is the same as presented in Smith et al. (2017). For each satellite galaxy, a random number, u, is generated from a uniform probability distribution between 0 and 1. u effectively corresponds to the luminosity percentile for that galaxy out of the distribution of possible values. The luminosity of the satellite galaxy is given by the solution to the equation

$$u = \frac{\langle N_{\rm sat}(> L \mid M)\rangle}{\langle N_{\rm sat}(> L_{\rm min} \mid M)\rangle}. \tag{5.2}$$

Interpolation of $\langle N_{\rm sat}(> L \mid M)\rangle$ is used to solve this equation. $\langle N_{\rm sat}(> L \mid M)\rangle$ must vary monotonically to produce a unique solution for L.

$M_{\rm min}$ and M_1 are modelled as cubic functions which means that they change rapidly at magnitudes beyond the fitting range. To avoid this, linear extrapolation of these parameters is used for faint magnitudes beyond the fitting range.

5.3.2 Colour Distributions

In addition to assigning galaxy luminosities, we also assign rest frame $(g - r)$ colours to the galaxies. This allows colour cuts to be applied to the mock and colour dependent clustering to be investigated.

The colour distributions are matched to data from GAMA. The process for assigning colours is an extension to the method of Skibba & Sheth (2009); Smith et al. (2017) and

is described in Smith et al. (2022).

The distribution of rest-frame colours is described by the sum of two Gaussian distributions, with five parameters that vary as functions of magnitude. One Gaussian describes the red population and the other describes the blue population. The parameters are the mean and standard deviation of each Gaussian, $\mu_{\rm red}$, $\sigma_{\rm red}$, $\mu_{\rm blue}$, $\sigma_{\rm blue}$, along with the fraction of blue galaxies, $f_{\rm blue}$, which defines the relative heights of the two Gaussians.

In addition, there is a luminosity-dependent bias to select satellites preferentially from the red peak and centrals from the blue peak.

Broken linear functions are used to describe how the parameters of the colour distribution vary with magnitude. In Smith et al. (2022) this was seen to improve the colour distributions relative to using a single linear function per parameter.

5.4 Conversion to Full-Sky Mocks

A mock galaxy format that is more representative of what DESI will observe is a full-sky mock. This projects the galaxies onto a simulated sky with RA and DEC coordinates, the coordinate along the line of sight becomes the redshift. Galaxy properties such as magnitude and colour are transformed into the observer's frame. The footprint of the DESI survey is overlaid as can be seen in the lower panel of Figure 5.7. This means that the effects of selection and survey geometry can be precisely modelled.

A more realistic mock creation method would be to use a lightcone which has evolution of halo and galaxy properties with redshift. Future work will be to make true lightcone mock catalogues for the BGS.

5.4.1 Galaxy Coordinates

In order to transform galaxy coordinates from the cubic-box to projections on the sky, an observer is placed at the corner of the box. The box is periodically replicated such that the sample is complete in any direction up to the maximum redshift for the mock, chosen to be $z = 0.6$ in this case.

Galaxy coordinates perpendicular to the line of sight can be projected into RA and DEC coordinates. Along the line of sight, the effects of galaxy velocity must be taken into

account. The formula for the observed redshift, $z_{\rm obs}$, is

$$z_{\rm obs} = (1 + z_{\rm cos}) \left(1 + \frac{v_{\rm los}}{c}\right) - 1, \tag{5.3}$$

where $z_{\rm cos}$ is the cosmological redshift, $v_{\rm los}$ is the line-of-sight velocity of the galaxy, and c is the speed of light. This allows the mock catalogue to accurately include the effects of redshift space distortions from peculiar velocities of galaxies.

5.4.2 Evolving Target Luminosity Function and HODs

By construction, our HODs produce a very good approximation to the target luminosity function at $z = 0.2$. However, the BGS covers $0 < z < 0.6$ and we want our mocks to have realistic evolution as a function of redshift. This means that we must model the evolution of the galaxy luminosity function when converting our cubic box to a cut-sky mock.

The target luminosity function is the luminosity function that we aim to reproduce in the mock catalogue. We aim to reproduce existing measurements from the SDSS and GAMA surveys

For $z > 0.15$, the target luminosity function is a Schechter function fit to the GAMA luminosity function (Loveday et al., 2012). Below $z = 0.15$, the target luminosity function smoothly interpolates to that from SDSS measurements (Blanton et al., 2003). For $z > 0.15$, the evolution of the Schechter parameters as a function of redshift is described by the following formulae (Lin et al., 1999)

$$M^*(z) = M^*(z_0) - Q(z - z_0), \tag{5.4}$$

$$\phi^*(z) = \phi^*(0) 10^{0.4Pz}, \tag{5.5}$$

where $z_0 = 0.1$, $P = 1.8$, and $Q = 0.7$ in this work. P sets the evolution in number density, Q sets the magnitude evolution. ϕ provides the normalisation of the luminosity function, and M^* is the cutoff magnitude.

The HOD is modified to produce the correct number densities at varying redshifts. The HOD mass parameters, M_0, M_1, and $M_{\rm min}$ are adjusted by a "slide factor", in order to reproduce the target luminosity function at different redshifts. These "slide factors" apply

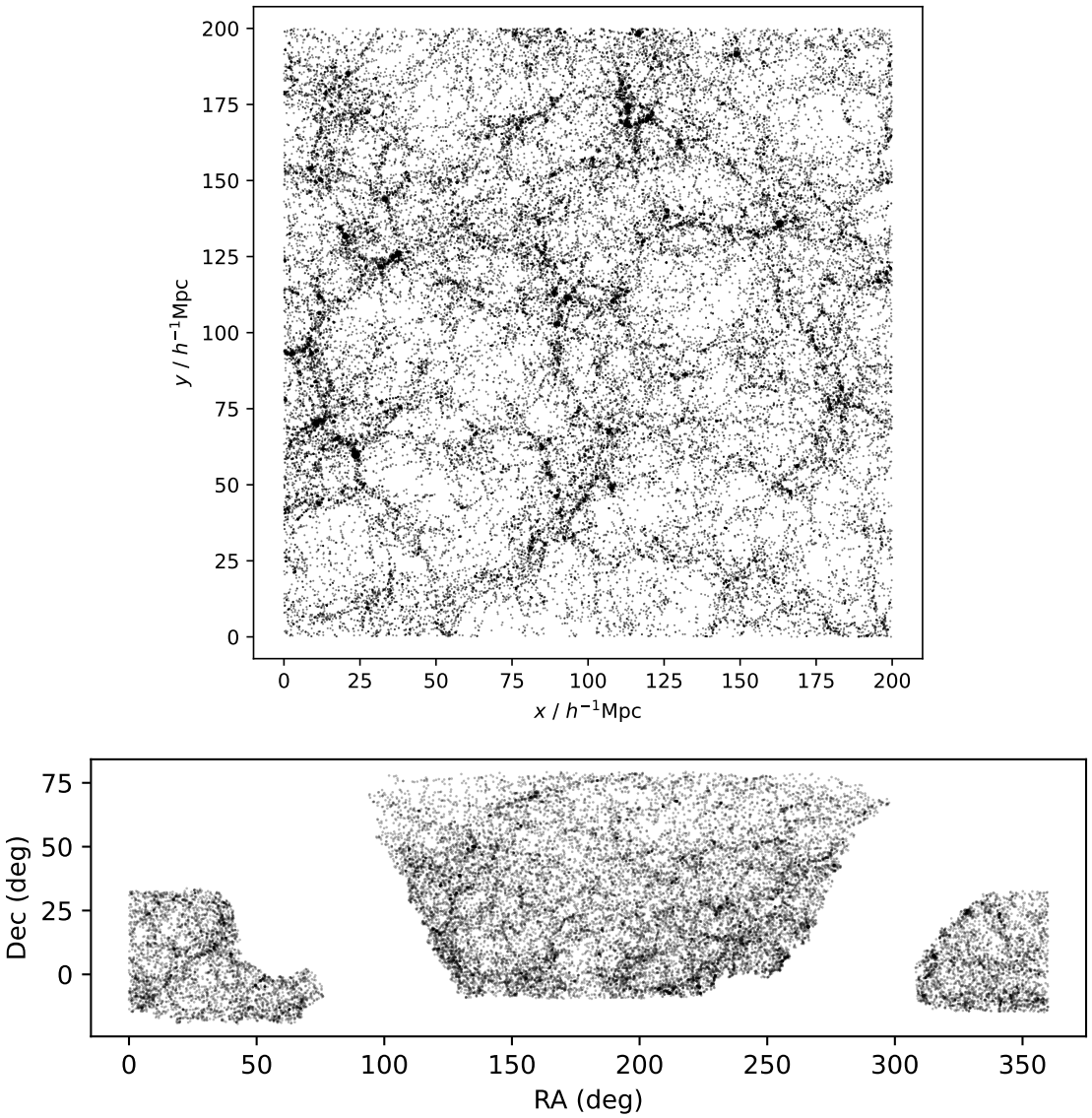

Figure 5.7: Upper panel: Galaxy positions in a small slice of the cubic box mock catalogue, of thickness $40h^{-1}$Mpc. Lower panel: Coordinates of galaxies on the sky from the full-sky mock that has been cut to only include the DESI survey footprint. The redshift slice shown is from $0.05 < z < 0.08$. Each galaxy is represented by a single point with transparency.

a fixed multiplicative factor to the HOD mass parameters such that the luminosity function is accurate. The shape parameters α, and $\sigma_{\log M}$ are unchanged.

5.4.3 Apparent Magnitudes & Colours

The absolute magnitudes of the galaxies are converted to apparent magnitudes and colours are k-corrected to the observer frame. The method presented here follows Smith et al. (2022).

A k-correction accounts for how the observed flux from an object changes at different redshifts. A single photometric band only probes a fraction of a galaxy spectrum. As a galaxy becomes redshifted, a different portion of the rest-frame spectrum is measured by a band in the observer frame. The k-correction accounts for this shift and allows us to convert between absolute and apparent magnitudes.

The exact form of the k-correction depends on the filter being used and the type of object being observed. In this work, k-corrections come from the GAMA survey and are described by fourth-order polynomials. The k-corrections are split into seven different rest frame $g - r$ colour bins with interpolation between the bins.

The formula relating apparent magnitude, r, to the absolute magnitude M_r of a galaxy is

$$r = M_r + 5 \log_{10} D_{\mathrm{L}}(z) + 25 + k_r(z), \tag{5.6}$$

where $D_L(z)$ is the luminosity distance from the observer to the galaxy at redshift, z, in units of h^{-1}Mpc. $k_r(z)$ is the r-band k-correction to redshift, z. We k-correct to a reference redshift of $z_{\mathrm{ref}} = 0.1$.

Similarly, galaxy colours can be transformed into the observer frame

$$(g - r)_{\mathrm{obs}} = (g - r)_{\mathrm{ref}} + k_g(z) - k_r(z), \tag{5.7}$$

where $k_g(z)$ is the g-band k-correction.

Figure 5.8 shows the distribution of $g-r$ galaxy colours from one of the AbacusSummit mock catalogues. Both the rest-frame colours and the observed colours are shown.

An apparent magnitude cut, $r < 20.2$ is applied to keep objects within the BGS FAINT selection criteria.

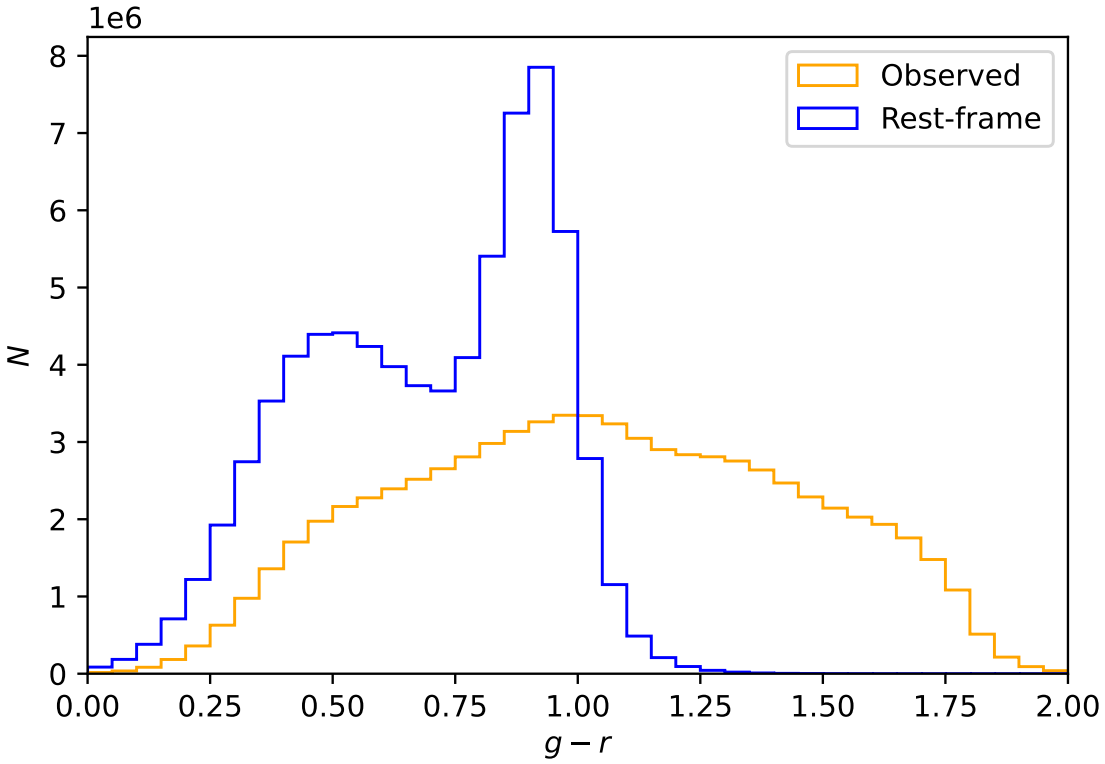

Figure 5.8: The distribution of rest-frame and observed $g - r$ colours from an AbacusSummit mock catalogue. The rest-frame distribution is described by a double Gaussian distribution.

5.5 Mock Catalogue Analysis

There are many potential uses for these AbacusSummit mock catalogues with regards to analysing BGS data. The mock catalogues can help to make predictions of galaxy clustering in a variety of cosmologies and in samples filtered by magnitude and colour.

In this section, analysis of the base AbacusSummit catalogue is presented along with comparisons to data from the BGS 1% survey that was collected during survey validation (Hahn et al., 2022). In addition, the measurements from a lightcone mock created from the Uchuu simulation (Dong-Páez et al., 2022) are included too. The Uchuu simulation was run in a cubic box of side length $2h^{-1}$Gpc and 12800^3 particles (Ishiyama et al., 2021), this leads to a mass resolution that is over six times that of AbacusSummit. The Uchuu mock catalogue uses a SHAM method to populate galaxies and their properties. The Uchuu mock catalogue used the same target data as the AbacusSummit catalogue. This allows us to compare to the data with another state-of-the-art mock catalogue. In all cases data are filtered to only include the BGS BRIGHT sample, $r < 19.5$. Comparisons are made at a variety of absolute-magnitude cuts and errors are found using a jackknife method.

5.5.1 Number Density

Firstly, the number density as a function of redshift is shown in Figure 5.9. The AbacusSummit mock catalogue and the Uchuu lightcone agree very closely with one another. The BGS data agrees well with the mock catalogues in general, especially in fainter samples. In the brightest $M_r < -22.0$ sample the BGS number density is higher than the mock catalogue value.

5.5.2 Clustering

A comparison of the projected correlation function is shown in Figure 5.10. A cut in redshift, $0.1 < z < 0.3$, is applied. There are differences in the clustering of the two mock catalogues that can be seen at both small and large scales. To a degree, this is to be expected because of the different treatment of sub-halos between the two catalogues. Uchuu uses sub-halos explicitly from a simulation with a high mass resolution, meanwhile

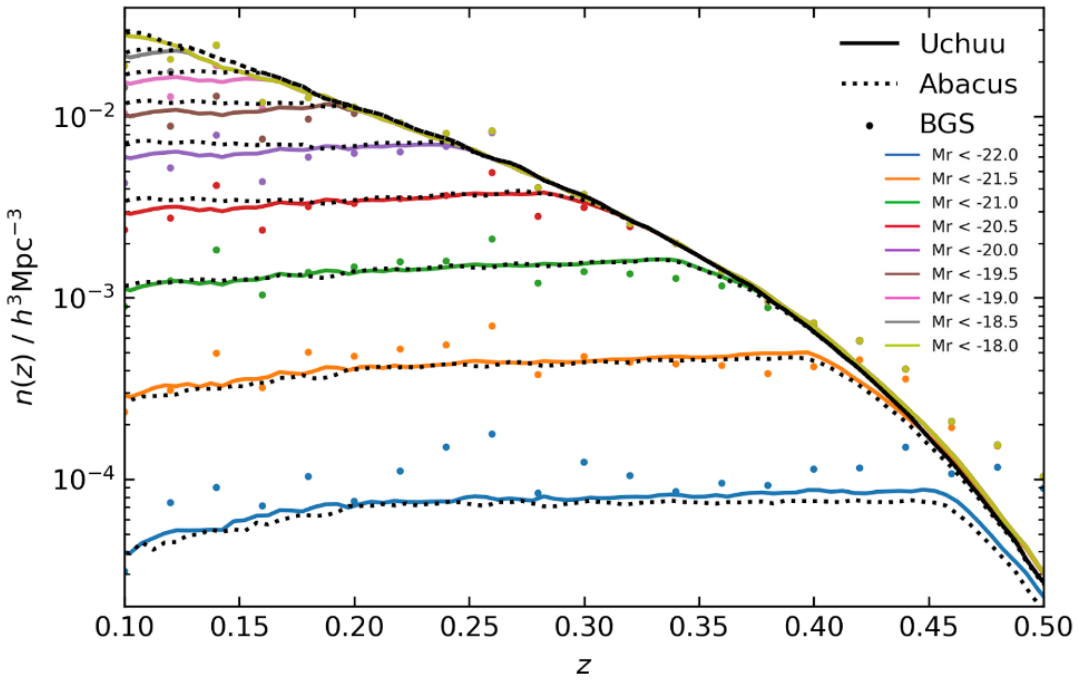

Figure 5.9: The number density of galaxies from the BGS 1% survey (points) compared to the AbacusSummit mock catalogue (dashed line) and the Uchuu lightcone (solid line). Number density is shown as a function of redshift. Each colour represents a sample with a different absolute magnitude cut. Only objects in the BGS BRIGHT sample are included. There is good agreement between the data and the mock catalogues for faint samples. In the brightest sample there is a slightly higher number density in the 1% survey than the two mock catalogues.

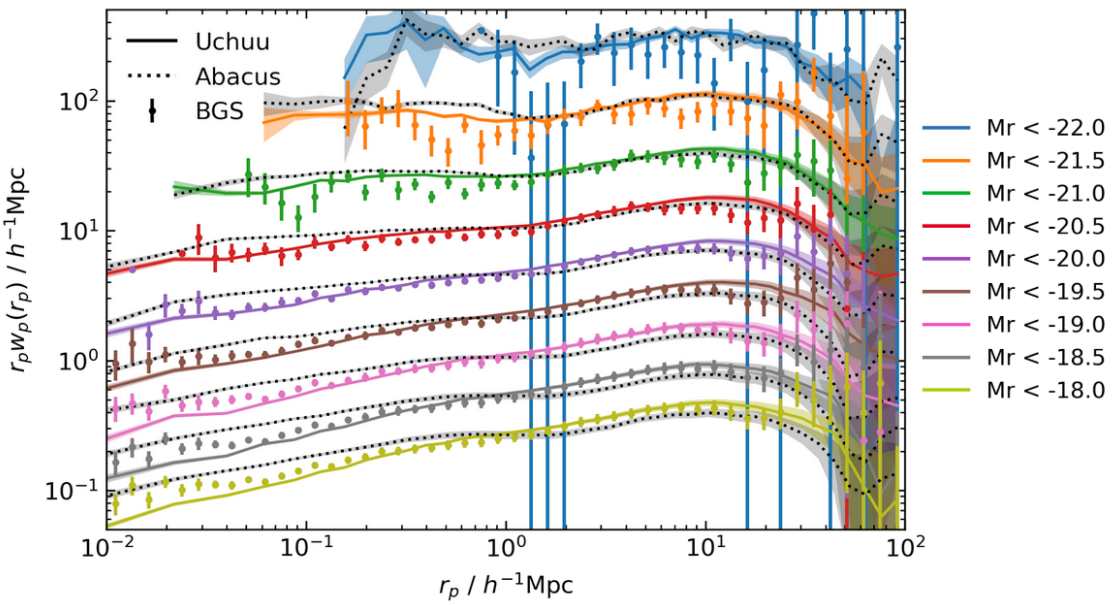

Figure 5.10: The projected correlation function of the BGS 1% survey (points), the AbacusSummit mock catalogue (dashed line), and the Uchuu mock catalogue (solid line). Each colour represents a sample with a different absolute magnitude cut. The samples have been offset from one another in the y-direction at intervals of 0.3 dex to reduce overlap and more clearly show differences between the measurements. The $M_r < -22.0$ sample is not offset. A cut in redshift, $0.1 < z < 0.3$, is applied. In faint samples, the AbacusSummit mock has lower clustering than the BGS data and the Uchuu mock at large scales, $r_p > 2h^{-1}$Mpc. The AbacusSummit mock catalogue has higher clustering than the other measurements at small scales, $r_p < 0.5h^{-1}$Mpc.

AbacusSummit has a lower mass resolution and the mocks use random samples from NFW profiles to place satellite galaxies within halos.

The BGS data is affected by incompleteness at small scales and a correction for this has been applied.

In the brightest samples there is a lot of noise in the BGS 1% survey measurements that makes it difficult to compare to the mock catalogues. In faint samples, the AbacusSummit mock has lower clustering than the BGS data and the Uchuu mock at large scales, $r_p > 2h^{-1}$Mpc. The AbacusSummit mock catalogue has higher clustering than the other measurements at small scales, $r_p < 0.5h^{-1}$Mpc. Overall, it looks like the Uchuu lightcone is more consistent with the 1% survey data than the AbacusSummit mock catalogue. The small-scale clustering difference indicates that varying the parameters of the NFW profiles that were used to position the satellites could produce a better fit. Reducing the concentration of the halos would reduce the small-scale clustering.

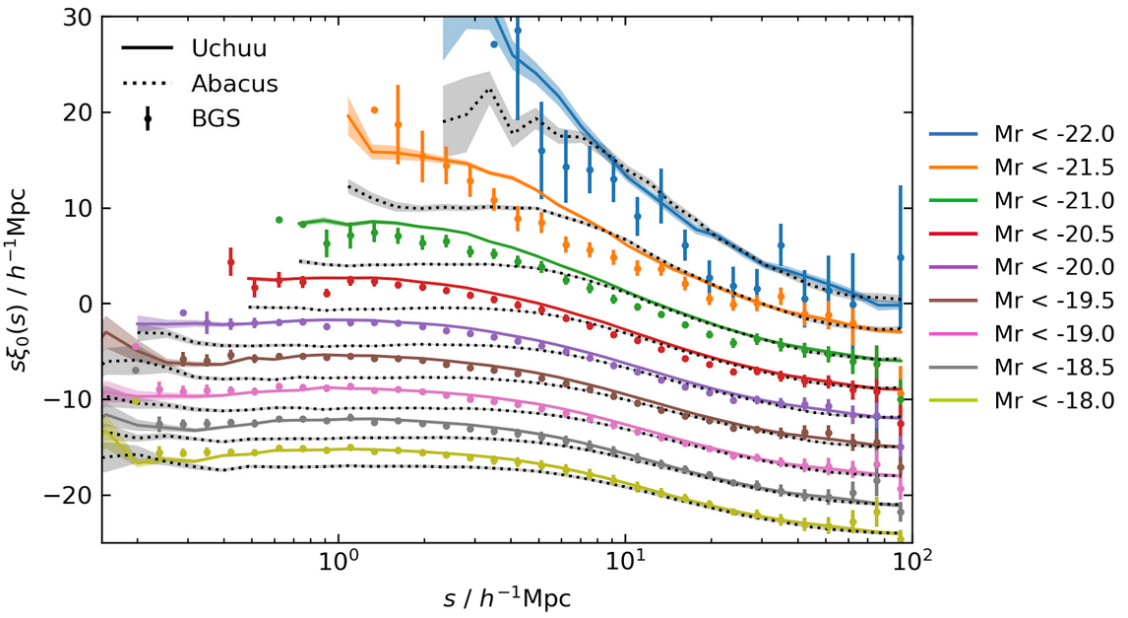

Figure 5.11: The clustering monopole of the BGS 1% survey, the AbacusSummit mock catalogue, and the Uchuu mock catalogue. The samples have been offset from one another in the y-direction at intervals of $3h^{-1}$Mpc to reduce overlap and more clearly show differences between the measurements. The $M_r < -22.0$ sample is not offset. At small scales, the AbacusSummit mock catalogue underestimates the monopole relative to the 1% survey data and the Uchuu lightcone. At large scales there is good agreement between the three measurements.

Figure 5.11 and Figure 5.12 show the clustering results for the monopole and quadrupole respectively (see Equation 1.13 for details). In the monopole, the Uchuu lightcone provides a better fit to the 1% survey data than the AbacusSummit mock catalogue for $s < 10h^{-1}$Mpc. In the quadrupole, Uchuu is again a better fit at smaller scales. In the 1% survey quadrupole, there is an increase around $s = 20h^{-1}$Mpc relative to the mock catalogues. Both the mocks and data are in good agreement at large scales.

5.6 Conclusions & Future Work

In this Chapter, we have described the process of creating mock galaxy catalogues for the DESI bright galaxy survey. Mock catalogues have been created and preliminary analysis is presented.

Firstly, HOD fits in varied cosmologies have been explored. The variation in HOD fits is greater than seen in the case of sample variance for a fixed cosmology. Best-fit HOD

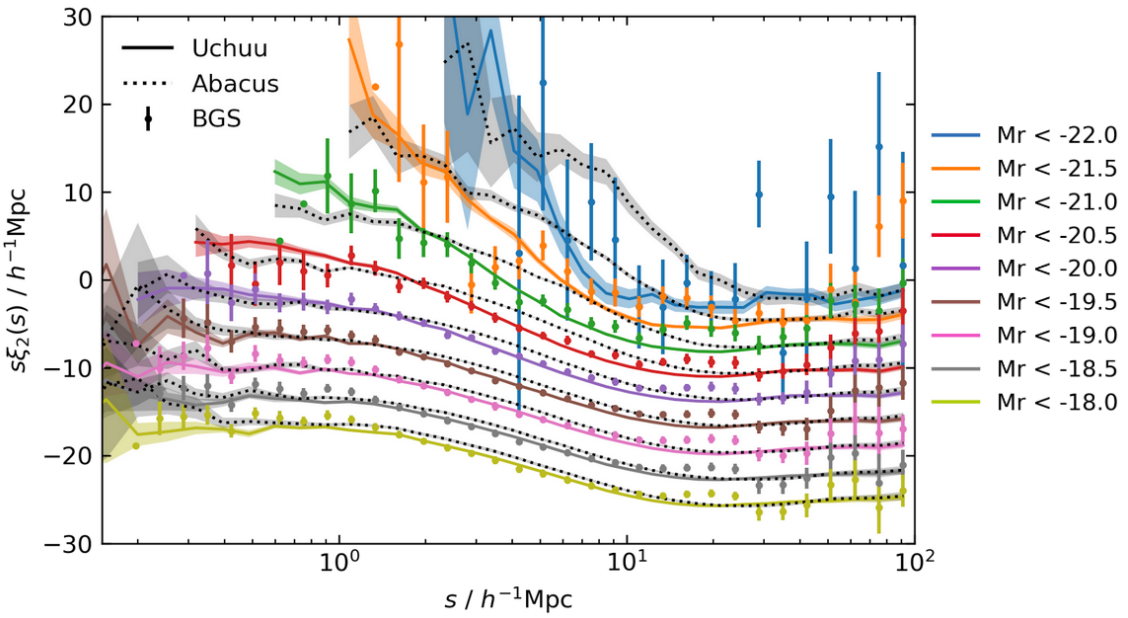

Figure 5.12: The clustering quadrupole of the BGS 1% survey, the AbacusSummit mock catalogue, and the Uchuu mock catalogue. The samples have been offset from one another in the y-direction at intervals of $3h^{-1}$Mpc to reduce overlap and more clearly show differences between the measurements. The $M_r < -22.0$ sample is not offset. Around $s = 20h^{-1}$Mpc the quadrupole from the 1% survey is larger than either of the mock catalogues.

parameters exhibit correlations with cosmological parameters of the input simulation.

Methods for the creation of both cubic-box and full-sky mock catalogues have been shown. This includes details on how to assign galaxy positions, magnitudes and colours. Galaxy properties have been transformed into the observer's frame in the full-sky mock. The mock reproduces a target luminosity function that evolves with redshift.

Number density and clustering measurements have been made from the base AbacusSummit mock catalogue. These are compared to data from the BGS 1% survey and a lightcone from the Uchuu simulation. The AbacusSummit mock catalogue generally agrees well with the BGS data, however there are some differences at small-scales.

These mock catalogues have been provided to the DESI collaboration to aid with cosmological analyses of DESI data in the future.

Future work will fit HODs to the BGS data in order to produce new mock catalogues that more closely match the BGS. In addition we aim to create mock catalogues from AbacusSummit lightcones that will include the evolution of halo properties with redshift.

CHAPTER 6

External Placements

I am part of the Centre for Doctoral Training (CDT) in Data Intensive Science at Durham. As a condition of my CDT PhD funding, it was necessary for me to undertake six months of placements with external partners. This was to facilitate closer links between the university and other institutions and to encourage the sharing of knowledge and best practices.

The two external placements which I undertook through the CDT program were:

- Northumbrian Water - flow rate predictions
- The Department for Education - pupil attainment modelling

6.1 Northumbrian Water - Flow Rate Predictions

For 8 weeks I performed a group project at Northumbrian Water (NW) with fellow student Edward Elliott. Our project was to design an anomaly detection algorithm for flow measurements from sewage treatment works (STWs).

6.1.1 Introduction

The wastewater network in the UK is a vast and complex system which requires constant monitoring. A diagram of the typical path of wastewater through the system is shown in Figure 6.1. There are many sources of wastewater: domestic, industrial, and groundwater. The water needs to be treated at an STW before it can be released back into the environment. There are many methods of treating wastewater but all involve removing organic matter and excess chemicals which would have an adverse effect on the local environment.

Each STW discharges the treated water (final effluent) into a stream or river. The Environment Agency (EA) restricts the amount of final effluent which is permitted to enter the environment based on the size of the receiving body of water. A large river can receive more final effluent than a small stream without becoming polluted. The Monitoring Certification Scheme (MCERTS) was set up by the EA in 2014 to monitor the flow of wastewater into the environment. MCERTS flow meters have been installed by NW to monitor the flow rate through all of their STWs and it is their responsibility to maintain the flow meters and produce high quality flow data. This data is provided to the EA to show that NW are fulfilling their requirements to monitor flow and protect the environment. Failing to record good quality data, or allowing a pollution incident to occur can lead to a large fine for the company.

The environment in which the flow meters are located can be hostile to the collection of good quality data. The flow meters can lose communications with the server that stores the data, they may become miscalibrated, solids can attach to the flow meter and form a blockage in the pipe. All of these problems must be swiftly dealt with to comply with MCERTS. Currently the monitoring of data quality is performed by a small team who manually look through all the data from over 100 STWs on a monthly basis. NW would like to show that they are proactive in addressing this inefficiency and therefore our project was to design a system which would automatically detect suspect data.

We were given the flow data from 121 STWs over the past three years. A flow value was recorded in 15 minute intervals at each site which gave us over 10 million data points to work with.

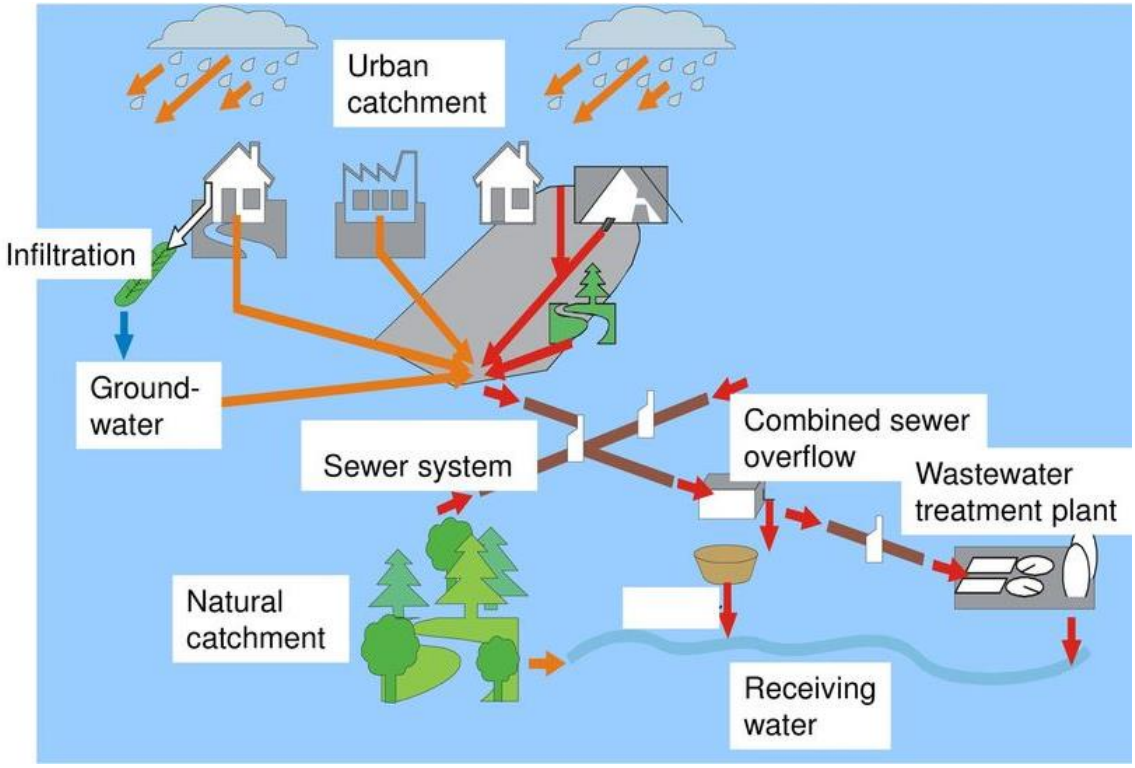

Figure 6.1: A schematic of the wastewater treatment network in the UK. Credit: Northumbrian Water

6.1.2 Methods

Initial Approach

There are many different types of suspect flow data, of varying severity and occuring across timescales from a few hours to months. Our system had to be capable of finding this suspect data reliably without flagging too many false positives. Examples of suspect data are:

- Missing data points
- Flatlining data
- Suspiciously high or low flow rate
- A moving baseline

It is possible to design very simple algorithms to detect each one of these anomalies and this is the approach which we started with. However there is an aspect of the problem which means that this is flawed. Rain is an integral component of the wastewater network

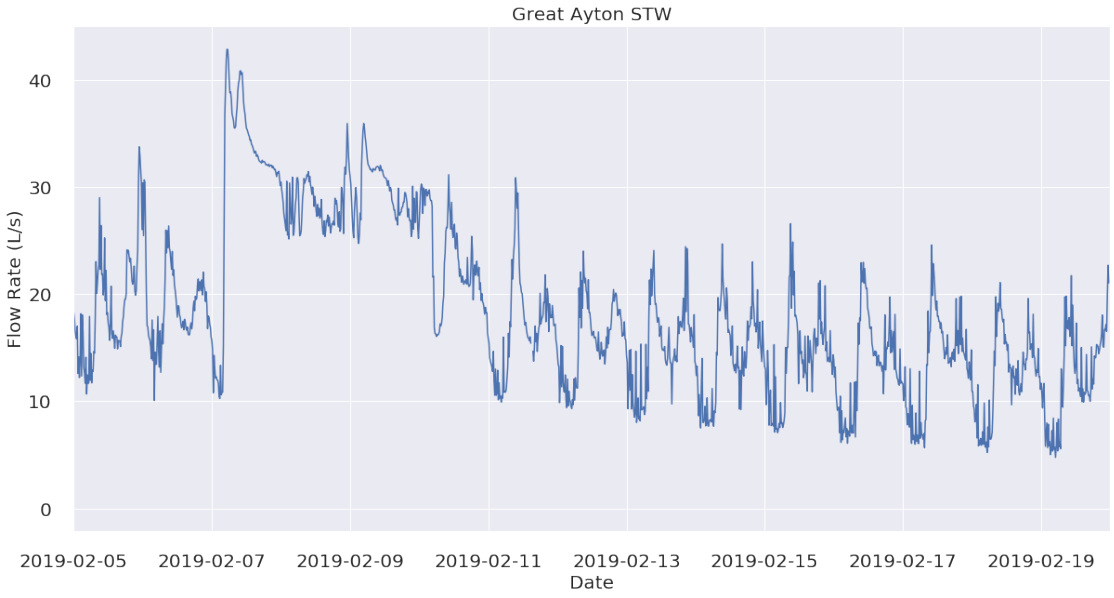

Figure 6.2: Typical, non-suspect flow data from an STW. The peak on the left is from rain entering the wastewater network. The periodicity observed is a diurnal pattern caused by domestic water usage which peaks in the morning and the evening

and this makes flow rates much less predictable when only looking at the STW flow data. Rain appears as large spikes which decay slowly and would be picked up as anomalies by this naive method. We had access to rain radar data, however rain has very different effects depending on the STW and the saturation of the ground which makes it difficult to model the effect accurately. A possible solution would be to ignore days which are rainy, however due to English weather this would not give us many days left over and it is important to be able to find suspect data on all days. Therefore we attempted a more ambitious approach to deal with this rain issue.

Machine Learning

We decided to use a machine learning based method to detect anomalies in the flow rate. This was possible as we had large amounts of data which could be used to train a neural network. This approach is very powerful because no prior model is assumed by the network, it learns how best to combine the inputs to predict the output flow rate most accurately. If the true output differs significantly from the predicted output then this suggests a possible anomaly. This approach also allowed us to easily include contextual data such as the rain radar and also data from correlated STWs.

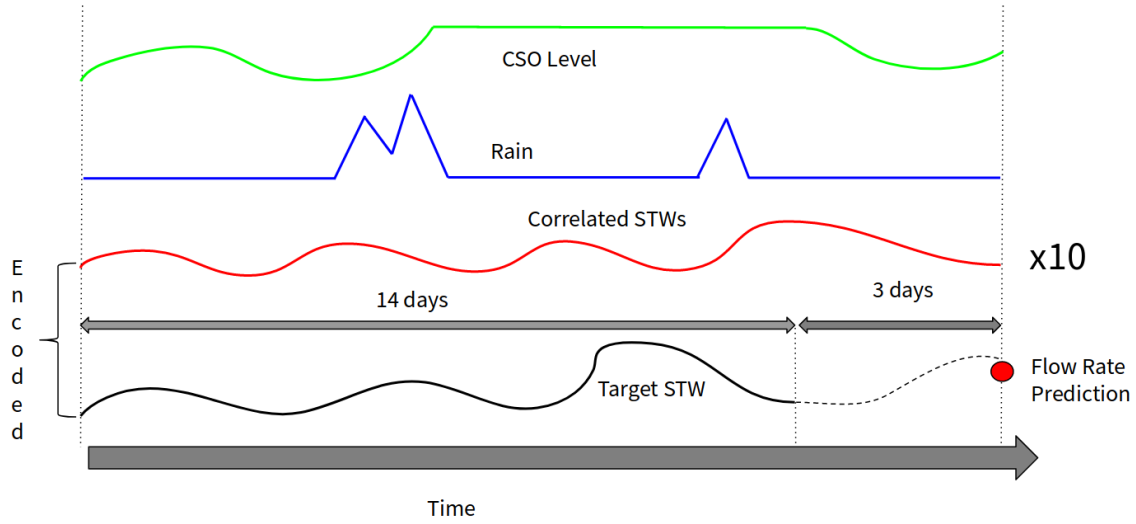

Figure 6.3: The inputs which were used by our network. Note that the target STW data is truncated three days before the prediction point while the other sources of data are not.

The inputs to our network are shown in Figure 6.3. We decided to use all the sources of data which we had access to. This included: combined sewer overflow (CSO) levels which measure the height of liquid in the sewer upstream of the STW, rain radar data which measures precipitation that can enter the STW as groundwater, and the data from 10 other STWs which are highly correlated with the target STW. The target STW data was also included as an input but this was truncated three days before the prediction point. The reason for this truncation is that we used the difference between the prediction and the actual flow rate to determine if there was suspect data or not. If there was only a short period of time between the input target STW data and the prediction then the network would learn to predict the effect of an anomaly which had appeared in its input data and therefore it would not be picked up as an anomaly.

A long-short term memory network (LSTM) was used as this has specific features which make it useful when dealing with time series data (Hochreiter & Schmidhuber, 1997). We also applied an autoencoder to the input STW data before using it in the prediction network.

An autoencoder is a type of neural network which aims to reproduce its input data. The size of the network is decreased in the middle layer relative to the input and output to create a reduced dimensionality representation of the data (Baldi, 2012). The autoencoder learns the generalised representation of the input data which retains the most information.

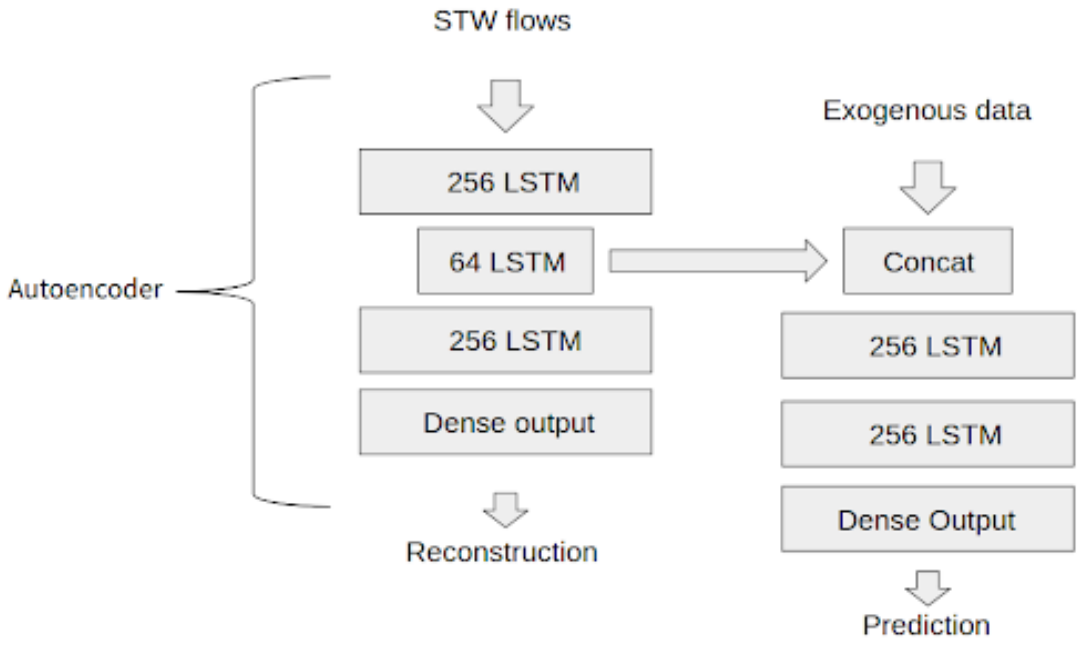

Figure 6.4: The shape of the LSTM network which we used to make predictions of the flow rate. On the left is the autoencoder, the encoded STW data is then concatenated with the other data sources and the prediction is made with two layers of 256 LSTM nodes.

This encoded data can then be passed onto the prediction network and will be more useful than the raw, unencoded data.

The network we used is shown in Figure 6.4. Smoothing and decreasing the time resolution of the data, along with only flagging anomalies which occur over multiple consecutive data points helped to reduce the false positive rate. The advantages of this method are that it is possible to perform anomaly detection on rainy days and that no assumptions are made about the type of anomaly so it is more generalisable than our initial approach.

6.1.3 Results

Figure 6.5 shows an example of an anomaly detected using this method along with a comparison to a simpler method by using a smoothed moving average. The anomaly is detected with a higher significance using the machine learning method and the flow rate prediction isn't biased by the presence of the anomaly. Due to the unsupervised nature of the problem, we did not have a list of confirmed anomalies with which to test anomaly classification accuracy. However, we were able to compare the accuracy of the predicted

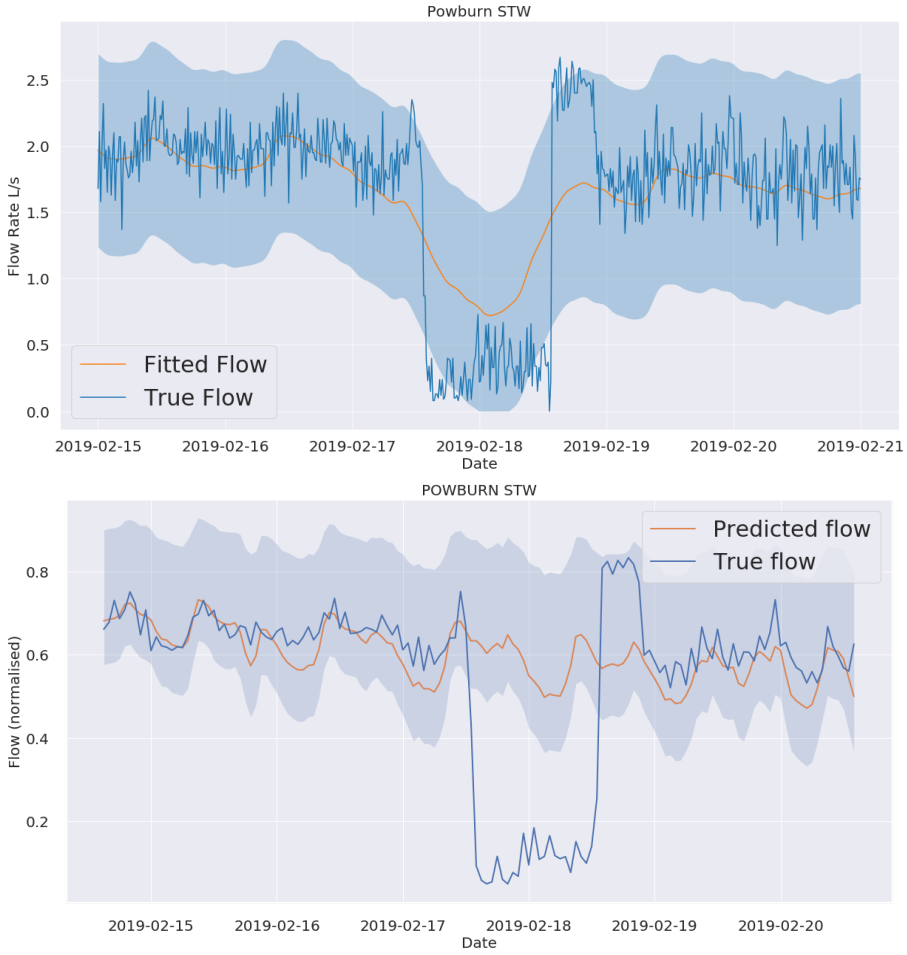

Figure 6.5: Upper panel: An anomaly detected using a smoothed moving average approach. Lower panel: The same anomaly detected using our machine learning approach. The shaded regions represent the margin of error for the prediction, outside of which an anomaly is flagged. The anomaly is detected with much higher significance using the machine learning method and the predictions are not biased around the anomaly.

flow rate in all conditions (anomalous and non-anomalous).

The inclusion of contextual data and large amount of training data allowed us to achieve a very high prediction accuracy. To measure the accuracy of our predictions we used the Nash-Sutcliffe Efficiency (NSE) (McCuen et al., 2006):

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_m^t - Q_O^t)^2}{\sum_{t=1}^{T}(Q_O^t - \overline{Q_O^t})^2}$$

where Q_m^t is the predicted flow rate, Q_O^t is the actual flow rate, $\overline{Q_0^t}$ is the mean flow rate, and the sums are taken over all timesteps from $t = 1$ to $t = T$.

NSE is often used to assess the quality of hydrological models and compares the

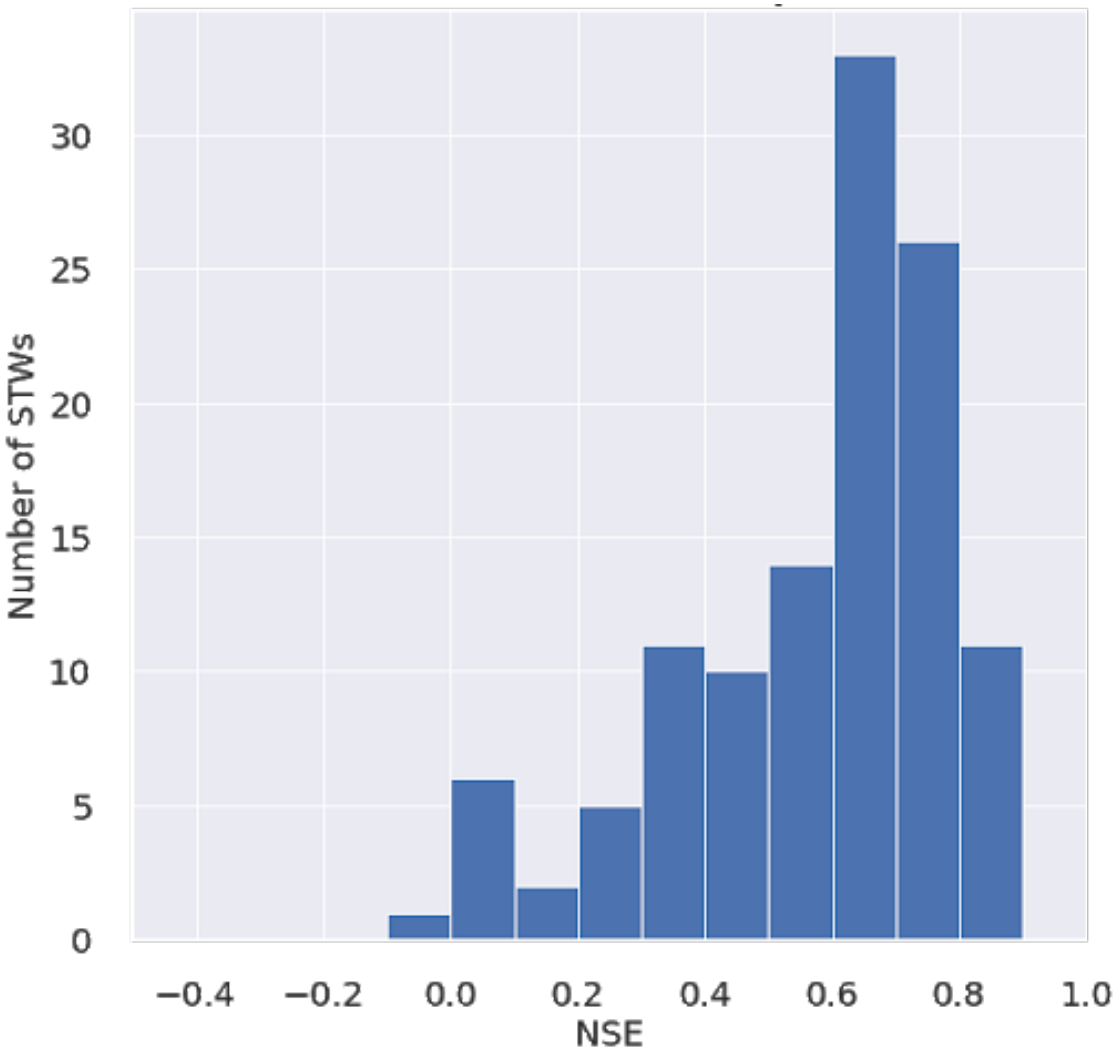

Figure 6.6: A histogram of NSE values for different STWs, over 70 have NSE>0.6 and only one has negative NSE.

performance of the model against using the mean of the data. NSE=1 corresponds to a perfect model and NSE = 0 means that the model is no better than using the mean. A commonly cited value for a model of sufficient quality is NSE > 0.6 (McCuen et al., 2006).

NSE varied between STWs as shown in Figure 6.6. In over half of the STWs our NSE was over 0.6. Figure 6.7 shows that predictions during rain and high flow were as accurate as dry weather flow predictions which allows for anomaly detection during rainy periods.

In order to assess the significance of anomalies it was necessary to produce a measure of uncertainty on our prediction. We decided to create an ensemble of networks, with the weights randomly rescaled such that the networks use the input data differently. These then create an ensemble of outputs which can be used to estimate a confidence interval. The

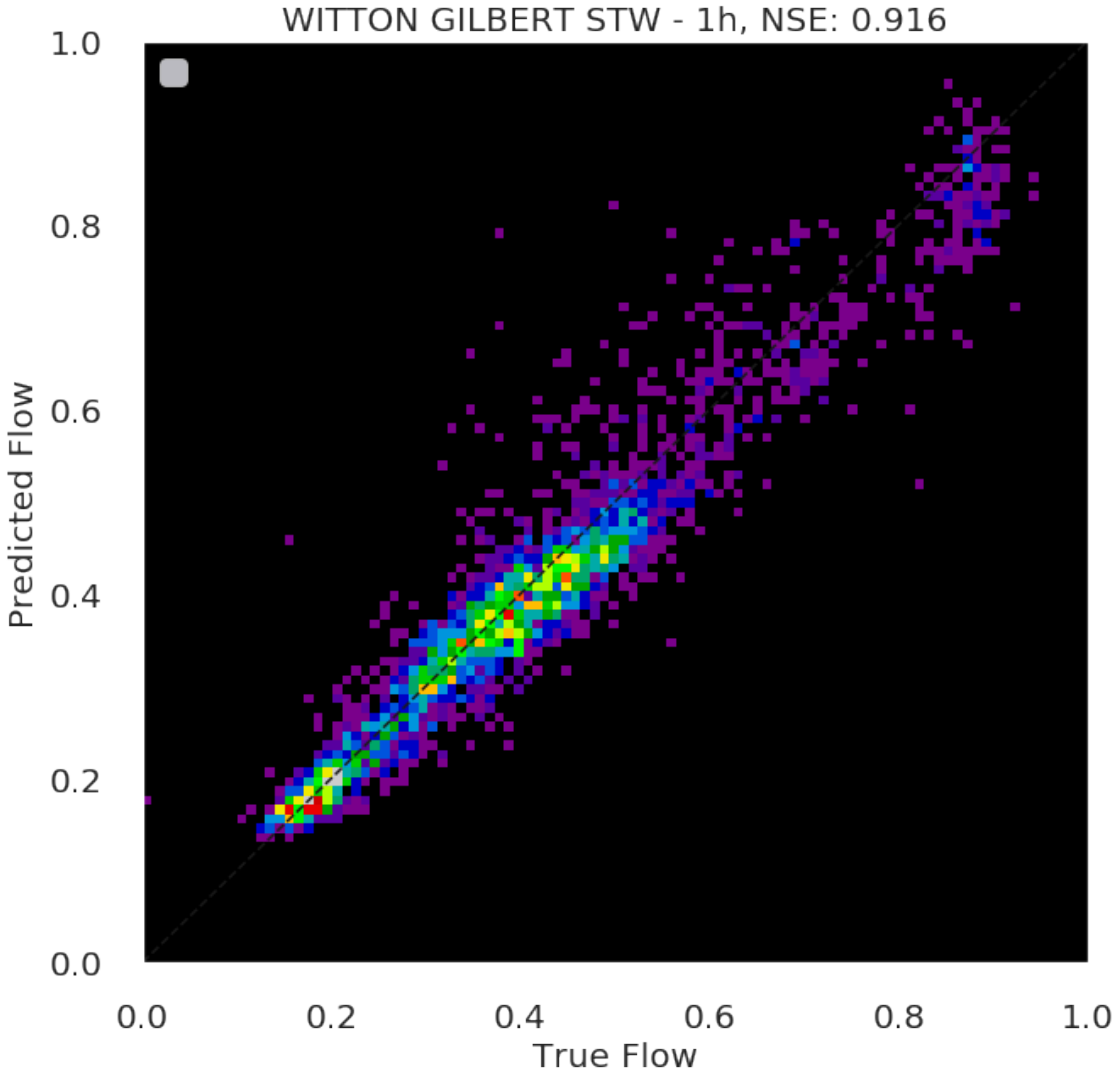

Figure 6.7: A heatmap of true flow vs predicted flow for one of the best modelled STWs. The predictions are accurate at both high and low flow rates. There was a large variation in behaviour between STWs however this plot is representative of those which had a high NSE.

benefit of this approach is that the confidence intervals become dynamic, the prediction is more certain during predictable dry spells than unpredictable rain events.

Northumbrian Water were pleased with our progress in improving the detection of suspect data in STW flows. Our project also acted as a proof of concept that machine learning models can be applied successfully to Northumbrian Water's datasets.

6.2 The Department for Education - Pupil Attainment Modelling

6.2.1 Introduction

I spent three months in early 2020 at the UK Government Department for Education, as part of the Data Science Lab (DasLab). I was given access to their databases and research literature and given the freedom to pursue whatever projects I thought would be valuable. I investigated many aspects of educational data, from the distances which students are willing to move to attend university, to the impact of educational attainment on lifetime earnings.

The main project I worked on was to project pupil outcomes from Key Stage 2 to Key Stage 4. Reducing the attainment gap between deprived and non-deprived students has been a key goal for the department and projecting attainment is a factor in measuring the impact of any interventions. In addition, this work could be used to find groups of students who outperform their expected results. These groups could be used as case studies to inform on potential methods to increase attainment in the rest of the population.

Since leaving the Department for education, my access to the documentation of this project has been limited, but a summary is contained within this chapter.

6.2.2 Modelling

Ethical Considerations

There were important ethical considerations in this research which the department took very seriously. Student data is sensitive so identifiable information was removed and I was only given access to fields which were necessary for the project in order to keep the data anonymised. The data included characteristics including race, gender, and deprivation indicies, this means that care was taken with the project such that there were no discriminatory effects as a result of the analysis. It was agreed that any potential actions taken as a result of this project would only be "positive interventions" such that any group would not be worse off as a result of actions taken.

Modelling Methods

Two main methods were considered for modelling pupil attainment, a linear model and a machine learning model using XGBoost (Chen & Guestrin, 2016; Chang et al., 2021a). The linear model should capture the effects from individual characteristics in isolation meanwhile the machine learning model may be able to capture more complex interactions between features if there is sufficient evidence for them. Finding important feature interactions would also be possible by adding terms to the linear model however the scope of possible interactions would be limited by a parametric model design which is not the case for the machine learning model.

Characteristics which were used as inputs to the models included:

- Key Stage 2 (KS2) attainment at age 11
- Income deprivation affecting children index (IDACI)
- Free school meals (FSM) eligibility and number of consecutive eligible years
- Ethnicity
- Gender
- English as a second language status (ESL)
- Year of entry into the state school system
- Urban/Rural code of school
- Special educational needs status (SEN)

The target variable was Key Stage 4 (KS4) attainment at ages 15-16.

XGBoost

An overview of the gradient boosting algorithm XGBoost is included in this section.

XGBoost is a library which provides an implementation of gradient boosted decision tree models[1]. It has become an industry standard for many data science and machine learning applications due to its speed and performance.

[1]https://xgboost.readthedocs.io/en/stable/

The outputs from a number of weak learners (individual decision trees) are combined to produce a strong learner. This is similar in principle to a random forest algorithm. The gradient boosting aspect relates to the fitting function being constructed in the form of a gradient descent algorithm. The decision trees are built in parallel and regularisation is used to minimise overfitting.

Hyperparameter tuning is an important step in training the best XGBoost model for a problem. Changing parameters such as the maximum depth of individual trees, learning rate, minimum node sample size, and number of learning rounds will alter the performance of the algorithm. Tuning these parameters with cross-validation is necessary to balance under- and over-fitting on a dataset.

6.2.3 Results

The Deparment for Education has limited my access to tables/graphs of the results from the models but I can give a summary of the findings. In general, the models mirrored existing research into pupil attainment at Key Stage 4 (Sutherland, 2015). Prior attainment at Key Stage 2 was the most important factor, and there are significant effects where lower attainment is associated with higher IDACI and rates of FSM and SEN.

There were not many cases where the XGBoost model provided significantly different predictions to the linear model, and accuracy was only slightly increased by using the more complex model architecture. The prediction accuracy was improved in the XGBoost case by finding combinations of features which described groups who defied the normal trends. As an example, pupils with English as a second language who entered the state school system late had a large attainment increase and were better predicted by the XGBoost model than the linear model. This is because the interaction of these two features selects for children of immigrant families whose performance at KS2 is artificially suppressed by unfamiliarity with the language. This effect is not fully accounted for by the linear combination of ESL and year of entry alone. However, this case does not provide useful lessons to learn about how to increase attainment across the board as other groups of pupils do not have their KS2 attainment reduced in the same manner.

6.2.4 Outcomes Dashboard & Conclusions

I created a dashboard using RShiny (Chang et al., 2021b) where one could explore the model predictions for any set of inputs and compare the predicted outcomes to the distribution of actual outcomes for pupils fitting those criteria. This would quickly and intuitively help analysts at the department to see the projected effects of different criteria on pupil attainment. Options to use either the linear or the XGBoost models were included. Using a dashboard to present information greatly increases engagement with a project and helps to allow others to explore the results and find future improvements. It also allows users to investigate the findings of the project in much greater detail and specificity than a static report and increases the breadth of questions that can be answered by analysts who are not intimately familiar with training and running the models themselves.

Although this project did not produce much actionable information with regards to methods of improving pupil attainment and shrinking the attainment gap there were several useful outcomes. Firstly, the lack of clear feature interaction effects justifies the use of linear models in much of the department's existing analysis. If a naive model has comparable performance to a complex model then Occam's Razor suggests that the simpler model should be preferred. In addition, exposure to these more complex modelling techniques should help the department to use them when suitable projects arrive, fostering bonds between the civil service and universities is key for progress to be made on this front. Finally, having an independent researcher validate many existing findings within the department provides a sanity check on research outcomes. This is of paramount importance when policy goals affecting millions of children are informed by the results of such analysis.

CHAPTER 7

Conclusions

The field of cosmology currently has an abundance of theories and a dearth of precise observations with which to test them. The ΛCDM model explains observational data exceptionally well in all but a few cases. However, it remains an unsatisfying model, with the precise nature of dark matter and dark energy being unknown. Observations of the Universe must push to new depths in order to test extensions and alternatives to ΛCDM. Large galaxy surveys such as the DESI Survey (DESI, 2016), Euclid (Laureijs et al., 2011), the 4-metre Multi-Object Spectroscopic Telescope (4MOST) (De Jong et al., 2019), and the Legacy Survey of Space and Time (LSST) (Ivezic et al., 2019), are an integral part of this drive to produce observations with greater precision. These surveys are pushing technological boundaries to map the Universe in greater detail than ever before.

N-body simulations are essential to leverage the data from galaxy surveys into information about the fundamental properties of our Universe. Realistic mock data aids our understanding of the systematic errors in measurements from galaxy surveys, which become more important as larger volume surveys decrease statistical noise to lower levels. These simulations also provide theoretical predictions for the effects of different cosmological models.

This thesis has explored N-body simulations and their application to the DESI Survey.

The analysis has spanned a variety of contexts, from abstractly testing the gravity scheme in an N-body code, to creating mock galaxies that realistically mimic DESI Survey data.

7.1 Accuracy of N-body Simulations

7.1.1 Gravity in SWIFT

In Chapter 2, tests of the gravity scheme in the SWIFT code were presented. SWIFT is a simulation code that uses a task-based parallelisation scheme to maximise performance on modern supercomputers that distribute computing power across thousands of nodes (Schaller et al., 2016).

Understanding the limitations of the gravity scheme in an N-body simulation code is an important step when considering how to run a simulation. Simulation speed must be balanced with accuracy.

The convergence of power spectrum results with the opening angle parameter, θ, was measured. θ describes the accuracy of the Fast Multipole Method (FMM) in SWIFT. Lower θ means that the force calculations are more accurate but take longer as a greater number of particle-particle interactions must be directly calculated.

The impact of θ on the forces experienced by particles was also explored. A large sensitivity to θ was seen in the forces on initial conditions created by particles perturbed from an evenly spaced grid. Acceleration inaccuracies were independent of the true force experienced by a particle. These inaccuracies were spatially represented by a grid pattern, the frequency of which suggested that a beating between the FMM grid and the particle grid. At low redshift and from glass initial conditions a lower sensitivity to θ was observed.

Differences in the measured power spectrum between SWIFT and perturbation theory were found. These differences were seen at large scales, where there is generally good agreement between perturbation theory and N-body simulation results. There was a clear linear dependence between the size of the global timestep in the SWIFT simulations and the error in the large-scale power. This error was found to be due to the long-range gravitational potential not being updated with sufficient frequency. The code was updated such that SWIFT was consistent with linear theory and other N-body simulation codes such as GADGET-2.

7.1.2 The N-body simulation code comparison project

In Chapter 3, several initial conditions generation codes and N-body simulation codes were compared against one another. This was to understand the impact of simulation code on measured results. This work built on the legacy of previous code comparison projects such as Winther et al. (2015), Schneider et al. (2016), and Garrison et al. (2019), with explicit application to the DESI Survey.

We generated initial conditions at three resolutions to test how well the codes reproduced the input power spectrum and how well the codes converged with increased resolution. We compared both the power spectrum and velocity statistics of the initial conditions that were generated.

N-body simulations were run from phase-matched initial conditions at three resolutions. We compared the codes SWIFT, ABACUS (Garrison et al., 2021b), GADGET-2 (Springel, 2005), and FASTPM (Feng et al., 2016). Comparisons were made in a range of statistics, including the matter power spectrum, matched halo properties, the halo mass function, and real- and redshift-space halo clustering statistics. The results could be compared to sample variance in the volume of the DESI Survey. This indicated the scales at which systematic errors due to choice of simulation code were below the statistical error in a DESI Survey volume. There was good agreement between the codes at large scales in the majority of the statistics. We find that the matter power spectra from the codes agree to within 1% for $k < 10h\text{Mpc}^{-1}$. The halo mass functions are consistent to within 1% between mass limits of $10^{11.5} - 10^{14}\ h^{-1}\ \text{M}_{\odot}$. Halo clustering correlation function measurements are within the DESI year 5 uncertainty at scales greater than $20h^{-1}\text{Mpc}$.

We made comparisons of the results from simulations at different resolutions in order to understand the impact of simulation resolution on measurement accuracy. There was a larger discrepancy between simulations of different resolution than simulations run with different codes. The differences due to simulation resolution were also compared to expected statistical errors in a DESI Volume. The results justified the choice of code and resolution adopted in the AbacusSummit suite of simulations (Maksimova et al., 2021) that are used to create mock galaxy catalogues for DESI.

We have shown that three, independently developed N-body simulation codes produce very similar results in matter clustering, halo properties, and halo clustering. We conclude

that N-body simulations are sufficient for the task of testing survey systematics at large scales.

7.2 HOD Fitting & Mock Galaxy Catalogues

We presented a novel method for fitting halo occupation distributions (HODs) in Chapter 4. This work implemented a tabulation method similar to that presented in Zheng & Guo (2016), where halo paircounts are precalculated and reweighted to predict the galaxy clustering when an HOD is applied. Paircounts were precalculated for central-central, central-satellite, and satellite-satellite components, with the latter being split into both the one- and two-halo term individually.

Fits were made simultaneously for multiple samples representing different absolute magnitude limits. HOD parameters were represented by smooth curves as a function of absolute magnitude. The hyperparameters describing these curves were directly fit using EMCEE (Foreman-Mackey et al., 2013). Both clustering and number density targets were included in the fitting process, these came from the GAMA Survey (Loveday et al., 2012) and SDSS (Blanton et al., 2003). Halo catalogues from the AbacusSummit simulation suite were used to create the HOD fits.

This method was successfully used to create HOD fits from simulations run with a range of cosmological parameters. In Chapter 5 the variation in these fits with simulation cosmology was explored. Correlations between best-fit HOD parameters and simulation cosmological parameters could be observed and explained. The process for creating cubic-box and full-sky mock catalogues from AbacusSummit simulations was described. Mock galaxies were given realistic luminosities and colours. A comparison was made between the clustering properties of our AbacusSummit mock catalogues and Bright Galaxy Survey data.

We have demonstrated that accurate HOD fits can be found in a range of cosmologies that span the Planck uncertainty. We have created mock galaxy catalogues from Abacus-Summit simulations that realistically simulate BGS data, including galaxy luminosities and colours. These mock catalogues will be used to measure the effects of systematics on analysis of DESI BGS data.

7.3 External Placements

Chapter 6 outlined a summary of the external placement work undertaken as part of my studies. This work applied my research skills to diverse settings outside academia.

Firstly an anomaly detection algorithm was developed for a water company. It is necessary to find and report anomalous flow readings in order to prevent pollution events and comply with environmental regulation. We created this algorithm by using a novel neural network design that accounted for a variety of input data sources such as local rainfall and sewer levels. The neural network approach was seen to have improved performance over more traditional model designs.

Secondly, a project was undertaken at the Department for Education. This project was to predict pupils' academic performance between the ages of 11 and 16. We used a machine learning approach and accounted for many variables describing pupil characteristics. This research reproduced internal departmental results. An interactive dashboard was created that allowed analysts to flexibly explore the results.

7.4 Future Work

There are clear avenues to broaden the scope of the research presented in this thesis. The N-body code comparison project from Chapter 3 can be extended to investigate the impact of the galaxy-halo connection on results. It would be useful to place bounds on the differences between mock galaxy catalogue properties that is due to the choice of simulation code. Differences between N-body codes could be marginalised over by abundance matching, or with a different galaxy-halo connection such that mock galaxy catalogue properties are not affected.

The HOD fitting method presented in Chapter 4 could be used on more simulation suites than just AbacusSummit. Different target data may be used to tune the HOD fits to the results of different surveys. For example, using data from the DESI Bright Galaxy Survey rather than GAMA & SDSS is a future application. This would then allow us to create mock galaxy catalogues that more closely match the properties of the DESI BGS. In addition, extending the HOD fitting method to use more parameters than just halo mass (e.g. environment) would allow us to more accurately model the effects of assembly bias.

The mock galaxy catalogues presented in Chapter 5 can be extended to use true lightcones from AbacusSummit as the source for dark matter halos, meaning halo properties evolve with redshift.

As galaxy surveys produce ever greater amounts of data, and as N-body simulations create universes with higher resolution in larger volumes, treatment of systematic errors will become more important. This means that researchers should remain vigilant and continue to test the accuracy and precision of N-body simulation methods.

In the coming years, DESI will begin to shed light on the evolution of dark energy through cosmic time. The consequences for the ΛCDM model are yet to be seen.

Bibliography

Aarseth S. J., Hoyle F., 1963, Monthly Notices of the Royal Astronomical Society, 126, 223 1.6.1

Abazajian K. N., et al., 2009, The Astrophysical Journal Supplement Series, 182, 543 1.4, 4.4

Abdalla E., et al., 2022, Journal of High Energy Astrophysics, 34, 49 1.2.3

Abrahamsen P., 1997, A review of Gaussian random fields and correlation functions. Norsk Regnesentral/Norwegian Computing Center, Oslo 3.2

Alam S. M. K., Bullock J. S., Weinberg D. H., 2002, The Astrophysical Journal, 572, 34 1.2.3

Alam S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617 1.2.2

Alam S., et al., 2021, Journal of Cosmology and Astroparticle Physics, 2021, 050 1.5.3

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, Monthly Notices of the Royal Astronomical Society, 426, 2046 3.2.1

Appel A. W., 1985, SIAM Journal on Scientific and Statistical Computing, 6, 85 1.6.1

Athanassoula E., Fady E., Lambert J. C., Bosma A., 2000, Monthly Notices of the Royal Astronomical Society, 314, 475 2.1

Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1856 5.1

Avila S., et al., 2020, Monthly Notices of the Royal Astronomical Society, 499, 5486 4.1.1

Baldi P., 2012, in Guyon I., Dror G., Lemaire V., Taylor G., Silver D., eds, Proceedings of Machine Learning Research Vol. 27, Proceedings of ICML Workshop on Unsupervised and Transfer Learning. PMLR, Bellevue, Washington, USA, pp 37–49, http://proceedings.mlr.press/v27/baldi12a.html 6.1.2

Barnes J., Hut P., 1986, Nature, 324, 446 1.6.1, 1.6.3

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, Monthly Notices of the Royal Astronomical Society, 356, 1191 4.1.1

Bautista J. E., et al., 2020, Monthly Notices of the Royal Astronomical Society, 500, 736 1.2.2

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2012, The Astrophysical Journal, 762, 109 1.6.4

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109 3.4

Bennett C. L., et al., 2013, The Astrophysical Journal Supplement Series, 208, 20 1.2.1

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, Monthly Notices of the Royal Astronomical Society, 311, 793 3.4

Bianchi D., Percival W. J., 2017, Monthly Notices of the Royal Astronomical Society, 472, 1106 1.5.1

Blanton M. R., et al., 2003, The Astrophysical Journal, 592, 819 5.4.2, 7.2

Blot L., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 2806 5.1

Bose S., Eisenstein D. J., Hernquist L., Pillepich A., Nelson D., Marinacci F., Springel V., Vogelsberger M., 2019, Monthly Notices of the Royal Astronomical Society, 490, 5693 4.1.1

Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, Astronomy & Astrophysics, 296, 575 2.4

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, Monthly Notices of the Royal Astronomical Society, 370, 645 4.1.1

Bryan G. L., Norman M. L., 1998, The Astrophysical Journal, 495, 80 3.4.1

Bullock J. S., Boylan-Kolchin M., 2017, Annual Review of Astronomy and Astrophysics, 55, 343 1.2.3

Chang C., et al., 2018, Monthly Notices of the Royal Astronomical Society, 482, 3696 1.4

Chang W., et al., 2021a, shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny 6.2.2

Chang W., et al., 2021b, shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny 6.2.4

Chen T., Guestrin C., 2016, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM, New York, NY, USA, pp 785–794, doi:10.1145/2939672.2939785, http://doi.acm.org/10.1145/2939672.2939785 6.2.2

Cheng H., Greengard L., Rokhlin V., 1999, Journal of Computational Physics, 155, 468 2.1

Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2014, Monthly Notices of the Royal Astronomical Society, 446, 2621 5.1

Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Physics Reports, 513, 1 1.5.3

Cole S., Kaiser N., 1989, Monthly Notices of the Royal Astronomical Society, 237, 1127 4.1.1

Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, Monthly Notices of the Royal Astronomical Society 3.1, 3.4

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, Monthly Notices of the Royal Astronomical Society, 319, 168 4.1.1

Cole S., et al., 2005, Monthly Notices of the Royal Astronomical Society, 362, 505 1.2.2

Colless M., et al., 2001, Monthly Notices of the Royal Astronomical Society, 328, 1039 1.4

Conroy C., Wechsler R. H., 2009, The Astrophysical Journal, 696, 620 3.4

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, The Astrophysical Journal, 647, 201 4.1.1

Crain R. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1937 4.1.1

Cresswell J. G., Percival W. J., 2009, Monthly Notices of the Royal Astronomical Society, 392, 682 1.3

Crocce M., Pueblas S., Scoccimarro R., 2006, Monthly Notices of the Royal Astronomical Society, 373, 369âĂŞ381 3.2.2

Croton D. J., et al., 2006, Monthly Notices of the Royal Astronomical Society, 365, 11 4.1.1

DESI 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design (arXiv:1611.00036) 3.1, 7

DESI Collaboration 2022, The Astronomical Journal, 164, 207 1.5.1

Davis M., Peebles P. J. E., 1983, The Astrophysical Journal, 267, 465 1.3

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, The Astrophysical Journal 1.6.1, 1.6.4, 3.1

Dawson K. S., et al., 2016, The Astronomical Journal, 151, 44 1.4

De Jong R. S., et al., 2019, Published in The Messenger vol. 175, pp. 3-11, March 2019. 7

DeRose J., et al., 2019, The Astrophysical Journal, 875, 69 3.1, 3.4.3

DeRose J., et al., 2021, arXiv:2105.13547 3.1, 3.4

Dehnen W., 2014, Computational Astrophysics and Cosmology, 1 1.6.3, 2.1, 3.3.1

Desjacques V., Jeong D., Schmidt F., 2018, Physics Reports, 733, 1 1.3

Desmond H., Mao Y.-Y., Wechsler R. H., Crain R. A., Schaye J., 2017, Monthly Notices of the Royal Astronomical Society: Letters, 471, L11 4.1.1

Ding Z., et al., 2021, in preparation 3.3.1

Dodelson S., Schneider M. D., 2013, Physical Review D, 88 5.1

Dong-Páez C. A., et al., 2022, The Uchuu-SDSS galaxy lightcones: a clustering, RSD and BAO study, doi:10.48550/ARXIV.2208.00540, https://arxiv.org/abs/2208.00540 5.5

Eastwood J., Hockney R., Lawrence D., 1980, Computer Physics Communications, 19, 215 1.6.1

Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, The Astrophysical Journal Supplement Series, 57, 241 1.6.1

Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705 1.1.2

Einstein A., 1917, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 142–152 1.1.2

Eisenstein D. J., Hu W., 1998, The Astrophysical Journal, 496, 605 1.2.2

Eisenstein D. J., et al., 2005, The Astrophysical Journal, 633, 560 1.2.2

Eisenstein D. J., et al., 2011, The Astronomical Journal, 142, 72 1.4

Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263 3.4.1

Elahi P. J., Cañas R., Poulton R. J. J., Tobar R. J., Willis J. S., del P. Lagos C., Power C., Robotham A. S. G., 2019, Publications of the Astronomical Society of Australia, 36 1.6.4

Elliott E. J., Baugh C. M., Lacey C. G., 2021, Monthly Notices of the Royal Astronomical Society, 506, 4011 4.1.1

Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273 3.2, 3.3.1, 7.1.2

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of the Pacific, 125, 306 4.4, 7.2

Friedmann A., 1922, Zeitschrift für Physik, 10, 377 1.1.2

Garrison L. H., Eisenstein D. J., Ferrer D., Metchnik M. V., Pinto P. A., 2016, Monthly Notices of the Royal Astronomical Society, 461, 4125 3.2.1

Garrison L. H., Eisenstein D. J., Pinto P. A., 2019, Monthly Notices of the Royal Astronomical Society, 485, 3370 3.1, 3.3.1, 3.3.2, 7.1.2

Garrison L. H., Joyce M., Eisenstein D. J., 2021a, MNRAS, 504, 3550 3.3.1

Garrison L. H., Eisenstein D. J., Ferrer D., Maksimova N. A., Pinto P. A., 2021b, MNRAS, 508, 575 1.6.3, 3.2, 3.3.1, 4.1.2, 7.1.2

Gibson B. K., Pilkington K., Brook C. B., Stinson G. S., Bailin J., 2013, Astronomy & Astrophysics, 554, A47 4.1.1

Guo Q., White S., 2013, Monthly Notices of the Royal Astronomical Society, 437, 3228 4.1.1

Guth A. H., 1981, Physical Review D, 23, 347 1.2.1

Hadzhiyska B., Eisenstein D., Bose S., Garrison L. H., Maksimova N., 2021, Monthly Notices of the Royal Astronomical Society, 509, 501 4.1.2

Hahn C., et al., 2022, DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation, doi:10.48550/ARXIV.2208.08512, https://arxiv.org/abs/2208.08512 1.6, 1.5.2, 5.5

Hamilton A. J. S., 1992, The Astrophysical Journal, 385, L5 1.3

Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, The Astronomical Journal, 156, 160 3.2.2

Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, Monthly Notices of the Royal Astronomical Society, 460, 2552 4.1.1

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2663 4.1.1

Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2019, Monthly Notices of the Royal Astronomical Society, 491, 5795 4.1.1

Hernández-Aguayo C., Prada F., Baugh C. M., Klypin A., 2021, Monthly Notices of the Royal Astronomical Society, 503, 2318 5.1

Hernández-Aguayo C., et al., 2022a, in preparation 3.4.3, 3.6.2

Hernández-Aguayo C., Ruan C.-Z., Li B., Arnold C., Baugh C. M., Klypin A., Prada F., 2022b, Journal of Cosmology and Astroparticle Physics, 2022, 048 1.6.5

Heymans C., et al., 2021, Astronomy & Astrophysics, 646, A140 1.2.3

Hochreiter S., Schmidhuber J., 1997, Neural Computation, 9, 1735 6.1.2

Hockney R. W., Eastwood J. W., 1988, Computer simulation using particles 1.6.3

Hou J., et al., 2020, Monthly Notices of the Royal Astronomical Society, 500, 1201 1.2.2

Hu W., White M., 1996, The Astrophysical Journal, 471, 30 1.2.1

Ishiyama T., et al., 2021, Monthly Notices of the Royal Astronomical Society, 506, 4210 1.6.1, 4.1.1, 5.5

Ivezic Z., et al., 2019, The Astrophysical Journal, 873, 111 7

Jenkins A., 2010, Monthly Notices of the Royal Astronomical Society, 403, 1859 1.6.2, 2.4, 3.2.1

Jenkins A., 2013, Monthly Notices of the Royal Astronomical Society, 434, 2094 3.2, 3.2.1

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, Monthly Notices of the Royal Astronomical Society, 321, 372 1.6.4

Jong J., Kleijn G. A. V., Kuijken K. H., Valentijn E. A., 2012, Experimental Astronomy, 35, 25 3.1

Kaiser N., 1987, MNRAS, 227, 1 1.3, 1.3, 3.4.3

Kegerreis J. A., Ruiz-Bonilla S., Eke V. R., Massey R. J., Sandnes T. D., Teodoro L. F. A., 2022, The Astrophysical Journal Letters, 937, L40 2.1

Klypin A., Prada F., 2018, Monthly Notices of the Royal Astronomical Society, 478, 4602 5.1

Kwan J., et al., 2015, The Astrophysical Journal, 810, 35 4.3.1

LHuillier B., Park C., Kim J., 2014, New Astronomy, 30, 79 2.3.2

Lacey C., Cole S., 1994, Monthly Notices of the Royal Astronomical Society, 271, 676 1.6.4

Lacey C. G., et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3854 1.6.5, 4.1.1

Landy S. D., Szalay A. S., 1993, The Astrophysical Journal, 412, 64 1.3

Laureijs R., et al., 2011, Euclid Definition Study Report, doi:10.48550/ARXIV.1110.3193, https://arxiv.org/abs/1110.3193 7

Lee J., et al., 2021, The Astrophysical Journal, 908, 11 4.1.1

Levi M., et al., 2013, The DESI Experiment, a whitepaper for Snowmass 2013 (arXiv:1308.0847) 3.4.3

Liddle A. R., 1995, Physical Review D, 51, R5347 1.2.1

Lin H., Yee H. K. C., Carlberg R. G., Morris S. L., Sawicki M., Patton D. R., Wirth G., Shepherd C. W., 1999, The Astrophysical Journal, 518, 533 5.4.2

Lippich M., et al., 2018, Monthly Notices of the Royal Astronomical Society, 482, 1786 5.1

Liske J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 2087 1.4, 4.4

Looser T. J., Lilly S. J., Sin L. P. T., Henriques B. M. B., Maiolino R., Cirasuolo M., 2021, Monthly Notices of the Royal Astronomical Society, 504, 3029 5.1

Loveday J., et al., 2012, Monthly Notices of the Royal Astronomical Society, 420, 1239 5.4.2, 7.2

Madgwick D. S., et al., 2003, Monthly Notices of the Royal Astronomical Society, 344, 847 1.3, 1.2

Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, MNRAS, 1.6.1, 3.2.1, 3.3.1, 4.1.2, 7.1.2

Marcos B., Baertschiger T., Joyce M., Gabrielli A., Labini F. S., 2006, Physical Review D, 73 3.2.1

McCarthy I. G., Schaye J., Bird S., Brun A. M. C. L., 2016, Monthly Notices of the Royal Astronomical Society, 465, 2936 4.1.1

McCuen R. H., Knight Z., Cutter A. G., 2006, Journal of Hydrologic Engineering, 11, 597 6.1.3

Mo H. J., White S. D. M., 1996, Monthly Notices of the Royal Astronomical Society, 282, 347 3.4.3, 4.1.1

Navarro J. F., Frenk C. S., White S. D. M., 1996, The Astrophysical Journal, 462, 563 1.6.4, 4.1.1

Navarro J. F., Frenk C. S., White S. D. M., 1997, The Astrophysical Journal, 490, 493 3.4.1, 4.1.1

Negri A., Volonteri M., 2017, Monthly Notices of the Royal Astronomical Society, 467, 3475 4.1.1

Neyrinck M. C., Yang L. F., 2013, Monthly Notices of the Royal Astronomical Society, 433, 1628 3.2.2

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, Monthly Notices of the Royal Astronomical Society, 396, 19 3.4.3

Oman K. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 3650 1.2.3

Ostriker J. P., Peebles P. J. E., Yahil A., 1974, The Astrophysical Journal, 193, L1 1.1.1

Peacock J. A., Smith R. E., 2000, Monthly Notices of the Royal Astronomical Society, 318, 1144 3.4

Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton University Press 3.4.1, 3.4.3

Peebles P. J. E., 1982a, The Astrophysical Journal, 258, 415 1.1.1

Peebles P. J. E., 1982b, The Astrophysical Journal, 263, L1 1.2.1

Peebles P. J. E., 1984, The Astrophysical Journal, 284, 439 1.1.2

Peebles P. J. E., 1987, The Astrophysical Journal, 315, L73 1.2.1

Penzias A. A., Wilson R. W., 1965, The Astrophysical Journal, 142, 419 1.2.1

Perlmutter S., et al., 1999, The Astrophysical Journal, 517, 565 1.1.2

Planck Collaboration 2020, A&A, 641, A6 1.2.1, 1.1, 3.2, 4.1.2

Potter D., Stadel J., Teyssier R., 2017, Computational Astrophysics and Cosmology, 4, 2 1.6.1

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, Monthly Notices of the Royal Astronomical Society, 338, 14âĂŞ34 3.1

Press W. H., Schechter P., 1974, The Astrophysical Journal, 187, 425 1.6.4

Raichoor A., et al., 2020, Monthly Notices of the Royal Astronomical Society, 500, 3254 1.2.2

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, The Astrophysical Journal, 771, 30 4.1.1

Riess A. G., et al., 1998, The Astrophysical Journal, 116, 1009 1.1.2

Riess A. G., et al., 2022, The Astrophysical Journal Letters, 934, L7 1.2.3

Rodríguez-Torres S. A., et al., 2016, Mon. Not. Roy. Astron. Soc., 460, 1173 3.1, 3.4

Rogstad D. H., Shostak G. S., 1972, The Astrophysical Journal, 176, 315 1.1.1

Rubin V. C., W. Kent J. F., 1970, The Astrophysical Journal, 159, 379 1.1.1

Ruiz-Bonilla S., Eke V. R., Kegerreis J. A., Massey R. J., Teodoro L. F. A., 2020, Monthly Notices of the Royal Astronomical Society, 500, 2861 2.1

Ruiz-Macias O., et al., 2020, Research Notes of the AAS, 4, 187 1.5.2

Ruiz-Macias O., et al., 2021, Monthly Notices of the Royal Astronomical Society, 502, 4328 1.5.2

Schaller M., Gonnet P., Chalk A. B. G., Draper P. W., 2016, Proceedings of the Platform for Advanced Scientific Computing Conference 1.6.3, 2.1, 3.3.1, 7.1.1

Schaye J., et al., 2014, Monthly Notices of the Royal Astronomical Society, 446, 521 1.6.5

Schneider A., et al., 2016, Journal of Cosmology and Astroparticle Physics, 2016, 047 3.1, 3.3.2, 7.1.2

Sheth R. K., Mo H. J., Tormen G., 2001, Monthly Notices of the Royal Astronomical Society, 323, 1 1.6.4

Skibba R. A., Sheth R. K., 2009, Monthly Notices of the Royal Astronomical Society, 392, 1080 5.3.2

Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, Monthly Notices of the Royal Astronomical Society, 369, 68 5.3.1

Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, Monthly Notices of the Royal Astronomical Society, 470, 4646 4.2, 4.2.1, 4.4, 5.1, 5.3, 5.3.1, 5.3.2

Smith A., et al., 2019, Monthly Notices of the Royal Astronomical Society, 484, 1285 1.5.1, 1.5

Smith A., de Mattia A., Burtin E., Chuang C.-H., Zhao C., 2020, MNRAS, 500, 259 3.4.3

Smith A., Cole S., Grove C., Norberg P., Zarrouk P., 2022, Monthly Notices of the Royal Astronomical Society, 516, 4529 5.1, 5.3.2, 5.4.3

Smoot G. F., et al., 1992, The Astrophysical Journal, 396, L1 1.2.1

Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, Monthly Notices of the Royal Astronomical Society, 391, 481 4.1.1

Springel V., 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105 1.6.3, 3.3.1, 7.1.2

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, Monthly Notices of the Royal Astronomical Society, 328, 726 1.6.4

Springel V., et al., 2005, Nature, 435, 629 3.2.1

Sutherland A., 2015, Factors Associated with Achievement 6.2.3

Takada M., Jain B., 2003, Monthly Notices of the Royal Astronomical Society, 340, 580 1.3

Theuns T., Chalk A., Schaller M., Gonnet P., 2015, SWIFT: task-based hydrodynamics and gravity for cosmological simulations (arXiv:1508.00115) 3.3.1

Valentino E. D., et al., 2021, Classical and Quantum Gravity, 38, 153001 1.2.3

Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Physical Review D, 71 1.1.1

Vogelsberger M., et al., 2014, Monthly Notices of the Royal Astronomical Society, 444, 1518 4.1.1

Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, The Astrophysical Journal, 399, 405 1.6.4

Wechsler R. H., Tinker J. L., 2018, Annual Review of Astronomy and Astrophysics, 56, 435 3.4, 4.1.1

Weinberg D. H., Dave R., Katz N., Hernquist L., 2004, The Astrophysical Journal, 601, 1 1.3

White S. D. M., Frenk C. S., Davis M., 1983, The Astrophysical Journal, 274, L1 1.1.1

Winther H. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 454, 4208âĂŞ4234 3.1, 7.1.2

Wu C. F. J., 1986, The Annals of Statistics, 14, 1261 3.4.3

Yuan S., Eisenstein D. J., Garrison L. H., 2018, Monthly Notices of the Royal Astronomical Society, 478, 2019 4.3.1

Yuan S., Hadzhiyska B., Bose S., Eisenstein D. J., 2022a, Monthly Notices of the Royal Astronomical Society, 512, 5793 3.6.2

Yuan S., Garrison L. H., Eisenstein D. J., Wechsler R. H., 2022b, Monthly Notices of the Royal Astronomical Society, 515, 871 4.3.1

Zehavi I., et al., 2011, The Astrophysical Journal, 736, 59 4.2

Zel'Dovich Y. B., 1970, A&A, 500, 13 1.6.2, 3.2.2

Zhai Z., et al., 2019, The Astrophysical Journal, 874, 95 4.3.1

Zhang H., et al., 2021, in preparation 3.4.3, 3.4.4

Zhao G.-B., et al., 2017, Nature Astronomy, 1, 627 1.5.3

Zheng Z., Guo H., 2016, Monthly Notices of the Royal Astronomical Society, 458, 4015 4.3.1, 7.2

Zheng Z., Zehavi I., Eisenstein D. J., Weinberg D. H., Jing Y. P., 2009, The Astrophysical Journal, 707, 554 4.1.1

Zwicky F., 1933, Helv. Phys. Acta, 6, 110 1.1.1

de Lapparent V., Geller M. J., Huchra J. P., 1986, The Astrophysical Journal, 302, L1 1.4, 1.3

van Daalen M. P., McCarthy I. G., Schaye J., 2019, Monthly Notices of the Royal Astronomical Society, 491, 2424 1.6.5

von Hoerner S., 1960, Zeitschrift fur Astrophysik, 50, 184 1.6.1