

Transforming cosmological N-body simulations

J. Paulavičius^{1,2}

¹*School of Physics and Astronomy, University of Edinburgh, Edinburgh, EH9 3JZ, United Kingdom*

²*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*

2 December 2016

ABSTRACT

I create realistic mock galaxy catalogues resembling the output of an N-body simulation in a given cosmology using the Zel’dovich approximation and the halo occupation distribution technique. I then implement the rescaling algorithm outlined in [Mead & Peacock \(2014a\)](#) to transform these mocks into catalogues that appear to have been the result of a simulation under a different model. This allows quick estimation of the expected degree of clustering and its error, as quantified by the covariance matrix. This expectation is generally dependent on the theory under test, which is commonly neglected by current surveys. My goal is to greatly reduce the influence of this systematic bias from future redshift-space distortions analyses, allowing tighter constraints to be placed on observables. It is found that the mock catalogues generated only possess the approximate clustering statistics expected, showing greater inaccuracies on small scales, where more accurate halo occupancy modelling could be employed. It is also found that the approach is limited to a range of redshifts in the original and target catalogue where universality of the halo mass function can be assumed and the corrections to the displacement field are small enough to avoid shell-crossing.

Declaration:

I declare that this project and report is my own work.

Signature:

Date: NOON 2 December 2016

Supervisor: Dr. M. J. Wilson

Duration: nine weeks

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

1 INTRODUCTION

The observed acceleration of the expansion of the universe is attributed to the domination of vacuum energy, but the Quantum Field Theory prediction for this vacuum energy differs greatly from that inferred from constraints on the background expansion history ([Weinberg et al. 2013](#)). The acceleration may instead be due to a breakdown of General Relativity (GR) on cosmological scales; this has motivated the creation of many modified gravity theories ([Clifton et al. 2012](#)).

A principal method for distinguishing between GR and modified gravity is to measure the rate at which density fluctuations grow, via Redshift-Space Distortions (RSD). RSD appear as an apparent statistical anisotropy of the galaxy distribution when radial positions are inferred from their

radial velocities – due to the additional Doppler effect from the peculiar velocities. On linear scales, where the overdensity of matter $\delta \ll 1$, these velocities are a consequence of the large-scale collapse of matter, due to the gravitational attraction of mass ([Kaiser 1987](#)). Testing gravity in this way has brought many successes ([Guzzo et al. 2008](#); [Peacock et al. 2001](#); [Guzzo & The Vipers Team 2013](#)) and is a fundamental goal of future surveys such as eBOSS, DESI, Euclid and WFIRST. On non-linear scales, these velocities are due to the virialised motion of galaxies in clusters ([Jackson 1972](#)).

Current RSD analyses commonly neglect the model dependence of the expected statistical errors – as quantified by the covariance matrix ([Heavens 2009](#)). If the assumption that GR is the correct theory of gravity on all scales is incorrect, and gravity is stronger than expected, this would lead

to more non-linear structure and a stronger covariance on these scales, resulting in an underestimation of the errors.

Moreover, as the statistical error on the covariance matrix estimate propagates to the errors on measured parameters (Taylor et al. 2013; Dodelson & Schneider 2013; Percival et al. 2014), scaling the variance of the measured parameters by a factor of the order $1 + N_b/N_s$, with N_b the number of bins of the measurement and N_s the number of simulations used to estimate the covariance, there simply are not enough simulations currently available to accurately estimate the covariance matrix with the range of cosmological parameters under test.

This is mainly because these simulations are prohibitively costly, especially for modified gravity models which require the calculation of the evolution of additional scalar fields (Clifton et al. 2012), which motivates the development of approximate, but fast, methods for generating mock galaxy catalogues. The aim of this project is therefore to implement and develop the rescaling method first proposed by Angulo & White (2010) and further improved upon by Mead & Peacock (2014a), in order to quantify the systematic bias resulting from neglecting the model dependence of the covariance matrix.

The first step in the process of generating mock galaxy catalogues under a given set of parameters often involves running a new simulation for each cosmology, which the rescaling algorithm aims to circumvent by directly transforming the output of simulations available to a different cosmology. This is achieved by changing the redshift and length scales of the halo catalogue to best match the halo mass function in the target cosmology while preserving the total mass enclosed by the simulation. The clustering on large scales is further amended to reflect that of the target cosmology by using the Zel'dovich approximation (ZA) (Zel'dovich 1970). Finally, the dark matter catalogue can be realistically populated by either semi-analytic or empirically motivated models. I use the latter, also called the halo occupation distribution (HOD), which is extensively employed to model RSD at non-linear scales (de la Torre & Guzzo 2012; Smith et al. 2003). No N-body simulations were used explicitly in this work, with the mock dark matter catalogues in the original cosmology obtained by applying the ZA and assigning masses to haloes according to the Sheth-Tormen halo mass function (Sheth & Tormen 1999).

The fiducial cosmology is assumed to be a Λ CDM cosmology with $\{h, \Omega_{m,0}, \Omega_{v,0}, \Omega_{r,0}, \Omega, \gamma, \sigma_8\} = \{0.678, 0.24, 0.76, 0.0, 1.0, 0.545, 0.593\}$

2 BACKGROUND THEORY

2.1 Power spectrum

I work in Fourier space and measure the anisotropic power spectrum,

$$P(\mathbf{k}) = \langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}^*(\mathbf{k}) \rangle, \quad (1)$$

where the average is taken over the modes in a spherical shell in k . The power spectrum is Fourier space counterpart of the two-point correlation function and, in the general case, carries the same information content. It is just one of the measures of the degree of clustering, but is most widely

used due to its parameter dependence making it possible to constrain both gravity and cosmological parameters, e.g. the baryon fraction, f_b , and $\Omega_m h$ (Rota et al. 2016).

Here $\delta(\mathbf{k})$ is the Fourier transform of the overdensity:

$$\delta(\mathbf{k}) = \frac{1}{V} \int_V d^3\mathbf{r} \delta(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \quad (2)$$

where the overdensity is defined as $\delta(\mathbf{r}) = \rho(\mathbf{r})/\bar{\rho} - 1$, and is therefore dimensionless. The power spectrum is related to the variance in overdensity at the inverse distance scale k , and it has no directional dependence if statistical isotropy can be assumed. A usual convention, also adopted here, is to plot the volume-independent product $VP(k)$, which is the quantity that from now the term power spectrum will refer to. A similar quantity is introduced as it is employed equally often, the dimensionless power spectrum,

$$\Delta^2(k, z) \equiv 4\pi V \left(\frac{k}{2\pi} \right)^3 P(k, z). \quad (3)$$

This has a more concrete interpretation - it gives the contribution to the overdensity variance per unit interval of $\ln |k|$.

Linear theory considers small overdensity perturbations arising from primordial fluctuations, which are perhaps of quantum origin. It holds when $\delta \ll 1$, and in this regime, according to the equations determining the large-scale linear growth of matter – the non-relativistic fluid equations and the Poisson equation, the linear evolution of the overdensity is described by

$$\partial_t^2 \tilde{\delta}(\mathbf{k}, a) + 2H \partial_t \tilde{\delta}(\mathbf{k}, a) = \tilde{\delta}(\mathbf{k}, a) \left(\frac{3}{2} H^2 \Omega(a) - \frac{c_s^2 k^2}{a^2} \right), \quad (4)$$

when matter is treated as a perfect fluid. Here $H = (\dot{a}/a)$ is the Hubble parameter and $a(t)$ denotes the cosmological scale factor (see Peacock (1999) chapter 15 for the full treatment). This equation has two solutions, the growing mode solution is the one of interest – it determines the late-time evolution:

$$\tilde{\delta}(\mathbf{k}, a) = D_+(a) \tilde{\delta}(\mathbf{k}), \quad (5)$$

which is simply an amplification with time. In an Einstein-de-Sitter (EdS) universe ($\Omega_m = 1$) this is

$$D_+ = (t/t_0)^{2/3}, \quad (6)$$

for $a \propto t^{2/3}$ and a present age of the universe given by t_0 .

A principal reason for working with the power spectrum is that it is the quantity that is predicted by inflationary theories, see Feldman et al. (1994) for further discussion. The amplitude of primordial fluctuations in overdensity, which form a Gaussian random field with the Fourier amplitudes drawn from an exponential distribution with the mean $P(k)$ (see section 16.3 of Peacock (1999)), can be directly related to the linear power spectrum that we see today through a cosmology-dependent relation similar to that given in eqn. (6). Moreover, the power of each k -mode evolves independently within linear theory, hence the covariance matrix of large-scale modes is diagonal. In addition to that, tighter constraints have been placed on cosmological parameters with the power spectrum (Alam et al. 2016; Beutler et al. 2016a) than the correlation function, and the scale at which non-linearity becomes significant is more distinct in Fourier space than in configuration space, as can be seen from mode-coupling effects on the covariance matrix for values of k above some k_{nl} (see Figure 2).

In practice, the output of a dark matter-only N -body hydrodynamics simulation are the proper comoving coordinates in units of Mpc h^{-1} , proper comoving velocities in units of $km s^{-1}$ and masses in units of $M_{\odot} h^{-1}$ for N haloes in a box of volume V employing periodic boundary conditions (PBC). The simulation box size sets the fundamental mode $k_{\text{fund}} = (2\pi/L)$, where L is the length of one side of the cubic box, with discrete modes going in units of this fundamental mode up to the Nyquist wavenumber, $k_{\text{Ny}} = \pi/\Delta$, where the cell size is denoted by Δ . I calculate the discrete Fourier transform (DFT) of the overdensity running over the lattice rather than over the individual particles, as the former allows us to use the Fast Fourier Transform (Frigo & Johnson 2005, FFT):

$$\tilde{\delta}(\mathbf{k}) = \left(\frac{1}{N_{\text{cells}}^3} \right) \sum_j \delta(\mathbf{r}_j) e^{i\mathbf{k} \cdot \mathbf{r}_j}. \quad (7)$$

In the discrete case $\delta(\mathbf{r}) = (N(\mathbf{r})/\bar{N}) - 1$, where $\bar{N} = N_{\text{particles}}/N_{\text{cells}}^3$ is the mean number density of particles (haloes or galaxies) on the lattice, and the sum is over the $N_x \times N_y \times N_z$ cells of the lattice, with \mathbf{r}_j being the position of the center of each cell. The FFT reduces the computational complexity of the resulting discrete Fourier transform from $O(N^2)$ to $O(N \ln N)$ ("An algorithm for the machine calculation of complex Fourier series", James W. Cooley and John W. Tukey).

Prior to running the FFT of the overdensity, the particles have to be assigned to cells, for which there are several choices of allocation schemes. Both the Nearest Grid Point (NGP) and Cloud in Cell (CIC) schemes (Hockney & Eastwood 1988) are employed here. The former simply allocates a particle to its nearest cell, and the latter assigns an effective weight to this cell and also its neighbours – depending on the exact particle position.

This gridding is effectively a convolution with a top-hat filter function, which must be taken into account for by deconvolving the estimated power by the appropriate function:

$$\tilde{\delta}_{\text{corr}}(\mathbf{k}) = \frac{\tilde{\delta}(\mathbf{k})}{\Pi_i \text{sinc}(\pi k_i / 2k_{\text{Ny}})^p} \quad (8)$$

The NGP case corresponds to $p = 1$, and the $p = 2$ case corresponds to CIC.

The gridding, being the sampling of a continuous function (that of the overdensity of matter), introduces the problem of aliasing, where the power from small wavelengths is superimposed on that of larger wavelengths, close to the grid scale (see Press et al. (1992) for more detail). Measurements of the power spectrum close to the Nyquist wavenumber ($\simeq k_{\text{Ny}}/4$ to k_{Ny}) are therefore unreliable. This can be reduced by using a higher-order grid allocation scheme, such as CIC, or by incorporating Jenkins folding, see Smith et al. (2003) for further description of this technique. Using Jenkins folding allows for an unbiased estimate of $P(\mathbf{k})$ to an arbitrarily high k . I utilise this technique where appropriate and use both NGP and CIC where appropriate.

The simplest example of a power spectrum that will be encountered throughout this work is the Poisson shot noise spectrum due to the random allocation of point particles. Their number density can be described as a sum of delta functions, giving the following expression for the overdensity

in real space:

$$\delta(\mathbf{r}) = \frac{\sum_{j=1}^N \delta^D(\mathbf{r} - \mathbf{r}_j)}{\bar{n}} - 1 \quad (9)$$

Taking the Fourier transform of this gives

$$\tilde{\delta}(\mathbf{k}) = \frac{1}{N} \sum_{j=1}^N e^{-i\mathbf{k} \cdot \mathbf{r}_j} - \delta^D(\mathbf{k}), \quad (10)$$

with N the total number of particles, \mathbf{r}_j their individual coordinates and with $\delta^D(\mathbf{r} - \mathbf{r}_j)$ the Dirac delta function. Finally, given $k \neq 0$ as this is unphysical, as per eqn. (1) with the convention to plot $VP(\mathbf{k})$, this becomes

$$VP(\mathbf{k}) = V \langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}^*(\mathbf{k}) \rangle = \frac{V}{N^2} \langle \sum_{jm} e^{-i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_m)} \rangle = \frac{1}{\bar{n}}, \quad (11)$$

where in the last step, the \mathbf{r}_j and \mathbf{r}_m are uncorrelated and so, for a large enough number of particles, all terms except $\mathbf{r}_j = \mathbf{r}_m$ cancel, with these simply giving N . Given that the galaxy distribution is modelled as a Poisson sampling of the underlying matter density field, the shotnoise adds to the measured clustering. This has to be taken into account when creating initial conditions for N -body simulations, for example (Springel et al. 2005). Both particles on a regular lattice, where the power spectrum is then exactly zero on all scales except that of the grid and its harmonics, and random particle positions were used when working with catalogues, with the former mainly used for testing purposes and the latter to better reflect the systematics of a real catalogue, such as that from a survey.

2.2 Redshift-space distortions

Redshift-space distortions are anisotropies in the clustering of matter when radial distances are inferred from redshifts. The assumption of statistical isotropy is not valid in redshift-space due to the effect of peculiar velocities on the line-of-sight clustering. The inferred redshift is distorted by a Doppler shift, due to the radial peculiar velocity. This effect is a small modification to the expected redshift coming from Hubble recession only,

$$(1 + z_{\text{obs}}) = (1 + z_{\text{cosm}}) \left(1 + \frac{v_{\text{pec}}}{c} \right) + \mathcal{O} \left(\frac{v_{\text{pec}}^2}{c^2} \right). \quad (12)$$

On the one hand, this greatly complicates the measurement of real-space clustering as one naturally views galaxies in redshift-space when inferring distances from their redshifts. On the other hand, these peculiar velocities carry information about the rate of the gravitational infall that the sources are experiencing, and this can be used to directly constrain theories of gravity.

One way to extract information about these peculiar velocities and therefore constrain cosmological parameters is through the relation between the power spectra in real and redshift-space – the Kaiser model of RSD (Kaiser 1987). In order to extract information from clustering in redshift-space, this apparent anisotropy must be quantified, which is done here by using the Kaiser model and expanding the real-space power spectrum in Legendre polynomials that multiply the anisotropic averages of redshift-space overdensities.

The Kaiser model provides a simple yet powerful relation between the clustering in real space and that in

redshift-space, applicable on scales where linear theory applies ($\delta \ll 1$). Starting with the observation that within linear theory dynamics (see chapter 15 of [Peacock 1999](#)), the following relation holds:

$$\nabla \cdot \mathbf{u} = -\dot{\delta}, \quad (13)$$

and combining this with eqn. (5) gives the comoving peculiar velocity in Fourier space as

$$\tilde{\mathbf{u}}(\mathbf{k}) = ifH \frac{\mathbf{k}}{k^2} \tilde{\delta}(\mathbf{k}), \quad (14)$$

where the logarithmic growth rate of density fluctuations is

$$f = \frac{d \ln |D_+|}{d \ln |a|}. \quad (15)$$

The growth rate is key for discriminating between theories of modified gravity, which commonly predict indistinguishable expansion histories ([Clifton et al. 2012](#)). In Peebles's parametrisation ([Peebles & Ratra 2003](#))

$$f(a) = \Omega_m(a)^\gamma, \quad (16)$$

where γ is predicted to be 0.545 for GR and $\simeq 0.4$ for the $f(R)$ models.

To find the mapping from real to redshift space, it is enough to think about what radial distance is measured by naively dividing the recession velocity of a distant galaxy relative to the observer by the Hubble parameter. The comoving coordinate \mathbf{r} is defined such that $\mathbf{x} = a(t)\mathbf{r}$. Differentiating this with respect to time gives $\dot{\mathbf{x}} = \dot{a}\mathbf{r} + a\dot{\mathbf{r}}$, where the first term corresponds to the Hubble recession as $H = \dot{a}/a$, so $\dot{a}\mathbf{r} = H\mathbf{x}$ and the second term $a\dot{\mathbf{r}}$ is the peculiar velocity of the galaxy. Dividing $\dot{\mathbf{x}}$ by H and taking only components of vectors along $\hat{\mathbf{r}}$ as it is the radial coordinate only that is inferred gives the proper redshift-space position of the galaxy,

$$as = \frac{\dot{\mathbf{x}}}{H} = a\mathbf{r} + \frac{a\mathbf{u} \cdot \hat{\mathbf{r}}}{H} \quad (17)$$

$$\mathbf{s} = \mathbf{r} \left(1 + \frac{\mathbf{u}(\mathbf{r}) \cdot \hat{\mathbf{r}}}{Hr} \right), \quad (18)$$

where the the peculiar comoving velocity of the observer $\mathbf{u}(\mathbf{0})$ is assumed to be equal to $\mathbf{0}$ as the redshifts have been corrected for the motion of the observer.

Given a distant object at \mathbf{r} for which the relation $kr_0 \gg 1$ holds (the distortions are plane-parallel), with r_0 being the survey depth, or, in this case, the simulation box size, the mapping of the overdensity from real to redshift space is found to be

$$\delta_s = \delta_r(1 + f\mu^2), \quad (19)$$

where $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{r}}$, and the subscripts s and r stand for redshift and real-space respectively. The relation shows that large-scale clustering is simply amplified anisotropically when viewed in redshift-space. Incorporating linear galaxy bias, $\delta_g = b\delta_m$, due to the fact that light from baryonic matter is a biased tracer of the underlying density field that mainly consists of a dark matter component, leads to the Kaiser model for redshift-space distortions (RSD):

$$P_s(\mathbf{k}) = (1 + \beta\mu^2)^2 P_g(\mathbf{k}), \quad (20)$$

with the Kaiser factor $\beta = f/b$. The Kaiser model successfully describes biased galaxy formation together with the large-scale infall of matter towards the deep gravitational wells (outflow from voids), which causes the light from galaxies on the nearside of a large cluster to be redshifted and that on the far side to be blueshifted. In addition to that, taking into consideration the virialised motion of galaxies in the centre of the well leads to the Fingers-of-God (FOG) effect ([Jackson 1972](#)), where galaxies appear to form long spindles pointing towards the observer.

In the non-linear regime ($\delta \gg 1$), on the order of the galaxy separations within a halo, these virialised motions lead to a suppression of the small-scale power, which can be modelled by adding a 1D Gaussian scatter with zero mean and a dispersion σ to the velocities of the satellites. This should depend on the mass of the host halo, and may include any measurement errors affecting redshift data in a real survey. The k -space effect is ([Peacock & Dodds 1994a](#)):

$$\tilde{\delta}(\mathbf{k}) \rightarrow \tilde{\delta}(\mathbf{k}) e^{-k^2 \mu^2 \sigma^2 / 2}, \quad (21)$$

which results in an average isotropic factor of

$$|\tilde{\delta}(\mathbf{k})|^2 \rightarrow |\tilde{\delta}(\mathbf{k})|^2 \frac{\sqrt{\pi} \operatorname{erf}(k\sigma)}{2 k\sigma} \quad (22)$$

A better alternative to model non-linear RSD is a Lorentzian damping term ([Peacock & Dodds 1994b](#); [Hawkins et al. 2003](#); [Cabr  & Gazta aga 2009](#)) with a characteristic pairwise velocity dispersion, σ_p . The resulting dispersion model,

$$P_s(\mathbf{k}) = \frac{(1 + \beta\mu^2)^2}{1 + \frac{1}{2}k^2\mu^2\sigma_p^2} P_g(\mathbf{k}), \quad (23)$$

is motivated by the distribution of pairwise velocities in N-body simulations being a good fit to a scale-independent exponential distribution ([Sheth 1996](#)).

The azimuthal symmetry and μ^2 dependence of $P_s(\mathbf{k})$ can be exploited by a compression to a limited number of multipole moments (ignoring the FOG):

$$\frac{P_s(\mathbf{k})}{P_g(\mathbf{k})} = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2 \right) + \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2 \right) L_2(\mu) + \frac{8}{35}\beta^2 L_4(\mu). \quad (24)$$

The monopole $P_0^S(\mathbf{k})$ is the spherical average of overdensity correlations, while the quadrupole $P_2^S(\mathbf{k})$ and the hexadecapole $P_4^S(\mathbf{k})$ are aspherical averages. Because of this, the shotnoise of eqn. (11) only applies to the monopole. This truncated decomposition only applies in the large-scale Kaiser regime; when damping is included there is no truncation to this series. Here $L_\ell(\mu)$ is a Legendre polynomial of order ℓ , which form a complete basis for $-1 \leq \mu \leq 1$. The quadrupole-to-monopole ratio ([Cole et al. 1995](#)) therefore contains the majority of statistical power for constraining $f\sigma_8$.

The RMS density contrast today when smoothed on a radial scale of $8h^{-1}\text{Mpc}$, σ_8 , is another key quantity that is widely used to constrain the cosmology ([Cacciato et al. 2013](#); [Beutler et al. 2016b](#)). $\sigma^2(R, z) = \langle \delta^2 \rangle$ is defined as the variance in the overdensity field smoothed with a spherical top-hat filter of radius R that would contain a mass $M = (4/3)\pi R^3 \bar{\rho}$ in a homogeneous universe:

$$\sigma^2(R, z) = \int_0^\infty \Delta^2(k, z) T^2(kR) d \ln k, \quad (25)$$

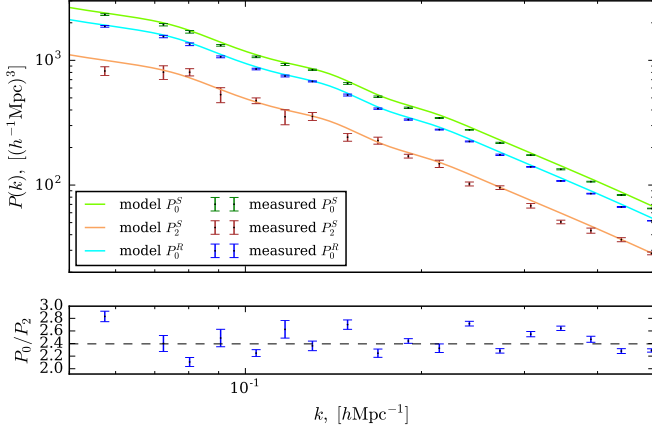


Figure 1. Averaged power spectrum measurement of 10 dark matter mock halo catalogues created assuming the ZA with mass-dependent displacements (§2.3.5), with 5×10^5 particles with random initial conditions occupying a $(256 h^{-1} \text{Mpc})^3$ box, distributed in mass according to the Sheth-Tormen halo mass function, eqn. (40). The monopole to quadrupole ratio now includes the effective bias given in eqn. (49), with $b_{\text{eff}} = 1.34$.

where $T(kR) = 3 [\sin(kR) - kR \cos(kR)] / (kR)^3$ is the Fourier transform of the filter. This provides the normalisation of the power spectrum, which is of interest in constraining the exact amplitude of the primordial density fluctuations.

I calculate the power spectrum multipoles by linear regression, minimising:

$$S = \sum_i \left[|\tilde{\delta}(\mathbf{k}, \mu_i)|^2 - P_0 - P_2 L_2(\mu_i) - P_4 L_4(\mu_i) \right]^2. \quad (26)$$

This sum-of-squares distance is minimised when

$$\frac{\partial S}{\partial P_0} = \frac{\partial S}{\partial P_2} = \frac{\partial S}{\partial P_4} = 0, \quad (27)$$

which leads to a set of simultaneous equations that are solved for P_0 , P_2 and P_4 by matrix inversion. The hexadecapole is commonly neglected due to the large statistical error on its estimate. The mean real and redshift-space monopole and the redshift-space quadrupole are shown in Figure 1, for a set of 10 mock dark matter halo catalogues, further described in §4.3.

A likelihood analysis (Heavens 2009) of the observed monopole and quadrupole requires a covariance matrix, which is generally model dependent (White & Padmanabhan 2015). The covariance matrix, when normalised to unity along the diagonal, is called the correlation matrix, and this quantity is of direct interest in Bayesian inference - it is used to find the χ^2 of the data through the relation

$$\chi^2 = \sum_{ij} (x_i - \mu_i) C_{ij}^{-1} (x_j - \mu_j), \quad (28)$$

where the covariance is given by

$$C_{ij} = \left\langle \left(P_\ell(k_i) - \langle P_\ell(k_i) \rangle \right) \left(P_{\ell'}(k_j) - \langle P_{\ell'}(k_j) \rangle \right) \right\rangle, \quad (29)$$

with the average being over the mock catalogues. The best guess for the value of a parameter of interest is then given by the maximum likelihood estimate, where the likelihood is $\mathcal{L} = \exp(-\chi^2/2)$.

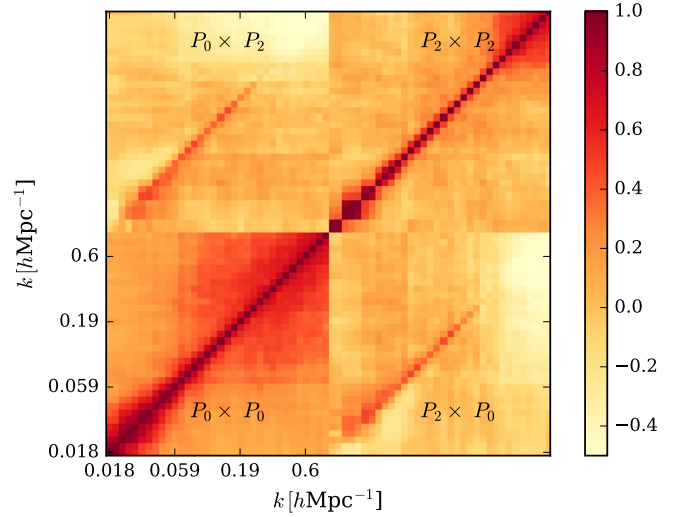


Figure 2. $P_0 - P_0$, $P_0 - P_2$ and $P_2 - P_2$ correlation matrices of 305 HOD (see section 2.3.2) mocks. This matrix demonstrates the main features that one expects to see: it is approximately diagonal for low k as these modes evolve independently according to linear theory, while high- k modes are correlated due to their interaction following the onset of non-linearity, when $\delta \gg 1$, or $k_{\text{nl}} \simeq 0.1$.

To test the machinery for calculating the covariance matrix, I estimated that of 305 Halo Occupation Distributed mocks (de la Torre, Guzzo, Peacock, Branchini & Iovino de la Torre et al.). The results of this are shown in Fig. 2, which shows the correlation matrix. This demonstrates the correlation one expects to see due to non-linearity and therefore inspires confidence in the accuracy of the calculation.

2.3 The halo model

2.3.1 Halo formation and profiles

In the halo model of large-scale structure (Seljak 2000; Peacock & Smith 2000) – see Cooray & Sheth (2002) for a review, initially overdense perturbations are assumed to form spherical virialised dark matter haloes. The initial collapse is predicted by linear theory, but following the onset of non-linearity, once the amplitude of overdensity fluctuations approaches unity, $\sigma(R, z) \simeq 1$, further evolution of the collapse can only be studied through N-body simulations or non-linear models, such as the spherical collapse model (Gunn & Gott 1972; Lahav et al. 1991). The end result is that haloes reach a virialised state with the non-linear overdensity of $\Delta_{\text{nl}} \simeq 178$, where this number is in general cosmology and redshift-dependent. Smaller mass haloes are observed to form sub-structure in large-mass haloes, which allows for the efficient cooling of baryonic matter. This is then the grounds for the formation of stars and galaxies (White & Rees 1978).

As the power spectrum is a two-point function and therefore insensitive to the underlying shape of structures, this model of dark matter being composed of spherical bodies with given masses and virial radii has proven to be a very accurate model (Takahashi et al. 2012; Klypin et al. 2016) of RSD up to scales where baryonic physics begin to have a

significant impact (Velliscig et al. 2014). The virial radius,

$$r_{\text{vir}} = \left(\frac{3M}{4\pi\bar{\rho}(z)\Delta_{\text{nl}}(z)} \right), \quad (30)$$

represents the physical boundary of the halo, beyond which there is assumed to be no mass – due to tidal stripping (Liongair 2008; Kampakoglou & Benson 2007) and the unbound nature of the trajectories at large radii. This assumption is challenged by observations of subhaloes with orbits extending to distances beyond to $3r_{\text{vir}}$ in N-body simulations (Ludlow et al. 2009). As the primary interest of this work is in the dynamics of galaxies, and not the warm baryonic gas on the cluster outskirts probed by the kinetic Sunyaev-Zel’dovich effect, the catalogues are only concerned with reproducing realistic clustering statistics, which can be achieved by populating them with galaxies with chosen positions and velocities. Therefore, r_{vir} is used as a strict truncation radius for the haloes. In this work Δ_{nl} is taken to be equal to 200 as in de la Torre & Guzzo (2012) for simplicity, noting that this can be generalised to a redshift-dependent function specific to Λ CDM, further outlined in §2 of Bullock et al. (2001).

A realistic density profile for each halo, as a function of mass, is required for the halo model to be accurate. The density profile of haloes has been shown to be universal (Navarro et al. 1996: NFW), i.e. possess a form that has a simple dependence on cosmology; this NFW profile is

$$\rho(r) \propto \frac{1}{(r/r_s)(1+r/r_s)^2}, \quad (31)$$

the cosmology dependence of which is encapsulated by the scale radius, $r_s = r_{\text{vir}}/c$; this is a characteristic length at which the density profile switches from $\rho(r) \propto r^{-1}$ to r^{-3} and therefore contains the majority of halo’s mass M (and the majority of subhaloes/satellites). The NFW profile is by no means a unique choice, and it has been suggested that the Einasto profile may be a better fit to the density of CDM haloes (Navarro et al. 2010; Dutton & Macciò 2014).

In this work, the concentration parameter is used as given in Bullock et al. (2001):

$$c(M, z) \equiv \left(\frac{r_{\text{vir}}}{r_s} \right) = \frac{c_0}{1+z} \left(\frac{M}{M^*} \right)^\beta, \quad (32)$$

where, following de la Torre & Guzzo (2012), with the values of $c_0 = 11$ and $\beta = -0.13$, where β in this case has no relation to the Kaiser model of RSD. M^* is defined such that $\sigma(M^*, z=0) = 1.686$ and it encapsulates the cosmology dependence of the dark matter concentration parameter. Although there are more advanced methods available in literature (Prada et al. 2012; Diemer & Kravtsov 2015) for parametrising the Λ CDM dark matter halo concentration that reflect discrepancies between the concentration relation used here and new data from observations and N-body simulations, I will use the relation as given due to it being a simple expression with the minimum necessary qualities – a redshift dependence and a cosmology dependence.

The decrease in concentration at higher redshifts can be related to smaller virial radii of haloes at higher redshift due to a higher background density (Bullock et al. 2001). The increase in concentration with decreasing mass is required to account for subhaloes having formed in a denser region, where the greater background density is provided by the host halo, and therefore earlier, resulting in

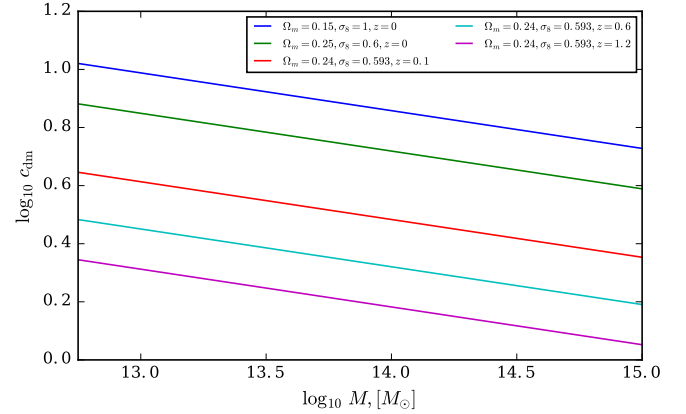


Figure 3. Shown here is the dependence of the dark matter halo concentration parameter c_{dm} , eqn. (32) on the cosmology and the redshift for haloes of masses in the range $5 \times 10^{12} - 5 \times 10^{15} M_{\odot}$. The cosmologies represented by the blue and the green lines are given in the caption, and reproduced from Ludlow et al. (2016).

a smaller core radius. In addition to that, a steepening of the outer core of subhaloes is expected due to tidal stripping (Kampakoglou & Benson 2007). Figure 3 shows the redshift and the cosmology-dependence of the dark matter concentration parameter. This form for the concentration parameter is not applicable in modified gravity models with a chameleon mechanism (screening of modifications on small scales to match GR and therefore observations that support it), and it can be amended by including an environmental dependence, redefining M^* such that $M^* = M^*(M_{\text{vir}}, \delta_{\text{env}}) = \sigma^{-1}(\delta_c(M_{\text{vir}}, \delta_{\text{env}}))$ as given in section 4.2 of Lombriser et al. (2014).

2.3.2 Galaxy occupation of haloes

The Halo Occupation Distribution (HOD) model (see e.g. Zheng et al. 2005) assumes galaxy formation is stochastic and occurs only in haloes – with a rate that is dependent only on halo mass. This neglects assembly bias, which incorporates the dependence on other factors such as the local environment (Eardley et al. 2015).

Following de la Torre & Guzzo (2012) and Zehavi et al. (2011), I divide galaxies into centrals – those that are the most massive galaxy in a given halo, and satellites – smaller mass galaxies that orbit the central. In the HOD formalism each halo is randomly assigned at most one central, and possibly a number of satellites. The expected number of each is parameterised by

$$\langle N_{\text{cen}} | M \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\ln |M| - \ln |M_{\text{min}}|}{\sigma_{\text{HOD}}} \right) \right]. \quad (33)$$

and

$$\langle N_{\text{sat}} | M \rangle = \left(\frac{M - M_0}{M_1} \right)^\alpha. \quad (34)$$

I use the parameters, e.g. M_0 and M_1 , given in Table 3 of Zehavi et al. (2011), for a maximum luminosity of $M_B = -19.5$ (see Fig. 11, 12 of the same paper for HOD fits using these parameters to the SDSS DR7 survey).

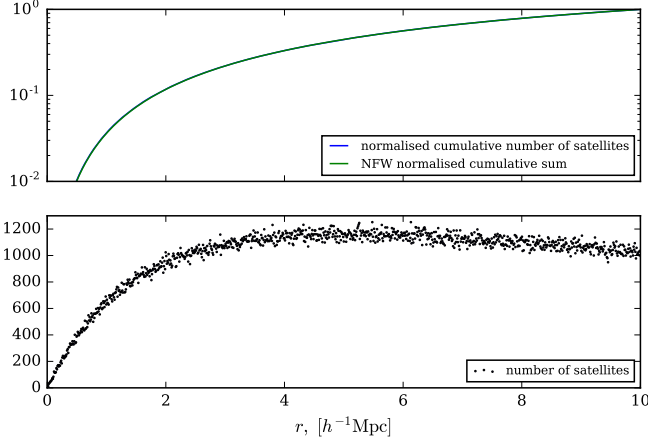


Figure 4. The upper subplot shows the cumulative number of satellite galaxies binned into 300 r bins between $r = 0 \ h^{-1}\text{Mpc}$ and $r = r_{\text{vir}} = 10 \ h^{-1}\text{Mpc}$, divided by the total number of satellites assigned. This number matches exactly the normalised cumulative sum of the NFW profile given in eqn. (35). The bottom subplot shows the actual number of satellites assigned the radial coordinate in each bin. The distribution in the radial coordinates of satellites is approximately constant at large scales, and only begins to vary noticeably at intermediate and small scales below $r = 4 \ h^{-1}\text{Mpc}$, where the corresponding effect in Figure 5 is the transition to damping above $k \simeq 0.1 \ h\text{Mpc}^{-1}$. Here $c_{\text{dm}} = 2$. The assignment accurately reproduces clustering on non-linear scales, which is why it is chosen here over the reconstitution (see §3.3 of (Mead & Peacock 2014a)) of sub-haloes.

Given a halo mass, I assign the number of centrals according to a Bernoulli distribution with the expected mean (de la Torre & Guzzo 2012); central galaxies are placed at the centre of each halo, and they have the peculiar velocity of their host halo. Similarly, the satellite number is drawn from a Poisson distribution with mean $\langle N_{\text{sat}} | M \rangle$. I assume these satellites trace the NFW density profile. The cumulative and actual number of satellites in a single mock catalogue that are assigned a given r is shown in Figure 4.

Inverse transform sampling (described in more detail in §2.3.3) is used to assign radial coordinates to the satellite galaxies, where the cumulative sum up to r for a radially symmetric profile is

$$\frac{M(r)}{M(r_{\text{vir}})} = \frac{\int_0^r r'^2 \rho(r') dr'}{\int_0^{r_{\text{vir}}} r'^2 \rho(r') dr'}, \quad (35)$$

and it is shown for the NFW profile in the upper subplot of Figure 4.

Satellites are also assigned the linear velocities of their host halo and further non-linear velocities described in §2.3.4.

The measured real-space power spectrum from such a catalogue is compared to the expectation in Fig. 5. Neglecting the clustering of centrals, the power spectrum of this assignment is a sum of two terms – the large-scale shotnoise of the random distribution of haloes (eqn. 11) convolved with the squared Fourier transform of the NFW profile describing the transition to the non-linear regime, and the small-scale shotnoise limit due to the satellites themselves randomly populating the haloes when considering scales below the virial radius.

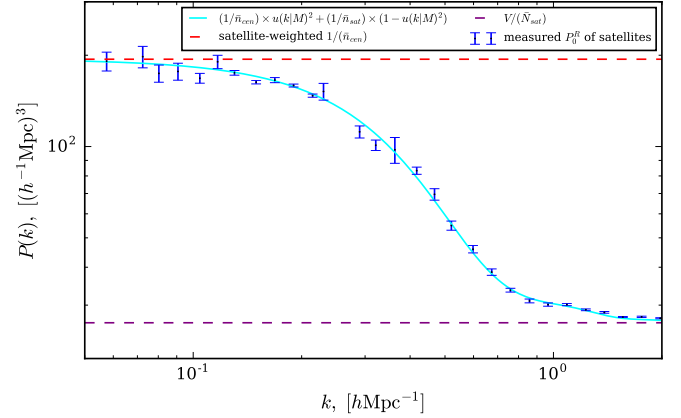


Figure 5. A plot of $P(k)^R$ of an average number of $\simeq 7 \times 10^5$, satellites per mock which populate 10^5 equal-mass ($M = 5 \times 10^{13} M_{\odot}$) dark matter haloes according to a NFW profile (eqn. 31) in a box of size $(256 \ h^{-1}\text{Mpc})^3$. Here $\Delta_{\text{nl}} = 0.004$, an artificially low value set to resolve both the 1-halo and 2-halo terms on scales before folding takes noticeable effect. This gives $\{c_{\text{dm}}, r_{\text{vir}}, r_s\} = \{3.17, 7.89, 2.49\}$. The values up to $k \simeq 0.2 \ h\text{Mpc}^{-1}$ are obtained without folding, while those for $k > 0.2 \ h\text{Mpc}^{-1}$ use a quadruple fold, hence the increase in the errorbar sizes. The mocks created can be seen to possess the expected clustering statistics.

As a number of satellites are allocated to centrals according to a Poisson distribution, the $1/\bar{n}$ shot noise power spectrum has to account for the effective weighting:

$$P_{\text{eff}}^{\text{SN}}(\mathbf{k}) = \frac{\sum_{\text{cen}} N_{\text{sat},i}^2}{(\sum_{\text{cen}} N_{\text{sat},i})^2} \quad (36)$$

which is true for any density-independent weighting. Assigning a fixed number of satellites to each halo would simply result in a $(1/\bar{n})$ spectrum on large scales. The resulting power spectrum for the correlation of satellites only is

$$P^R(k) = P_{\text{eff}}^{\text{SN}}(k) \times |\tilde{u}(k|M_0)|^2 + \frac{1}{\bar{n}_{\text{sat}}^{\text{avg}}} \times (1 - |\tilde{u}(k|M_0)|^2), \quad (37)$$

where $\tilde{u}(k|M)$ is the Fourier transform of the density profile for a halo of mass M ,

$$\tilde{u}(k|M) = \frac{\int \rho(r|M) \sin(kr) 4\pi r^2 dr}{\int \rho(r|M) 4\pi r^2 dr}, \quad (38)$$

which has the analytic form specific to the NFW profile as given on page 34 of Cooray & Sheth (2002). The last bracket in eqn. (37) ensures that the shotnoise of the satellites only appears on the relevant (small) scales.

It is clear from this equation that there are two additive contributions to the power in the halo model, that from the correlation of galaxies within the same haloes (the one-halo term), which has no affect the power on the large-scale power spectrum, and the correlation between haloes themselves (the two-halo term), which is assumed in this work to be given by linear theory. This additivity can be explained by the fact that variances can be summed for the individual components, and these are linear in the corresponding power spectra (eqn. 25). The separability into additive terms is also directly obtainable by splitting the density field into a component due to centrals and one due to satellite galaxies (Okumura et al. 2015).

2.3.3 Realistic halo masses

In producing realistic mock catalogues, it is necessary take into account the fact that haloes have a variety of masses. This is because, as described in §2.3.5, the biased clustering of haloes is mass-dependent. The halo mass function (HMF), dn/dM , gives the comoving number density of haloes as a function of mass. The mass function has been found to be well-fit by the universal form (Sheth & Tormen 1999):

$$\frac{dn}{dM}dM = \frac{\bar{\rho}}{M}f(\nu)d\nu, \quad (39)$$

where

$$f(\nu) = A \left[1 + \frac{1}{(q\nu^2)^p} \right] e^{-q\nu^2/2}, \quad (40)$$

and $A = 0.216$, $q = 0.707$ and $p = 0.3$ are free parameters calibrated to N-body simulations. The cosmology dependence is then encapsulated by the peak height only,

$$\nu = \frac{\delta_c}{\sigma(R, z)}, \quad (41)$$

with the relation $M = (4/3)\pi R^3 \bar{\rho}$ always applying (assuming a homogeneous universe). I use the Sheth-Tormen HMF to assign realistic masses to haloes in a chosen cosmology. In practice, this is done by inverse transform sampling (see pg. 287 of Press et al. (1992)). The cumulative distribution function (CDF) is calculated for all values of ν from ν_{\min} to ν_{\max} , which is achieved by integrating the mass function up to ν :

$$F(\nu) = \int_{\nu_{\min}}^{\nu} \frac{f(\nu)}{M} d\nu \quad (42)$$

with $\nu_{\min} = \nu(M_{\min})$ and $\sigma = \sigma_{\text{lin}}$ now being the linear variance in overdensity obtained from a linear power spectrum (further described in §4.2), and the cumulative probability should be calculated assuming it is zero for M_{\min} and unity for M_{\max} . The CDF is then inverted and equated with a uniform random number between 0 and 1, giving the corresponding a value of $M(\nu)$ for the halo in question.

2.3.4 Virial velocities

Following the assignment of centrals to each halo according to eqn. (33) and satellites with an NFW profile (eqns. 34, 31), both centrals and satellites are given the linear velocity of their host halo. As per §2.2, the satellites have an additional mass-dependent Gaussian random deviate added to their z -velocity to mimic the virial velocity that leads to observable effects, such as the FOG.

Eqn. (35) of Mead & Peacock (2014a) gives the variance for a given halo mass. Eqn. (22) gives the averaged isotropic effect of this damping on the overdensities, and the effect of applying this damping on the power spectrum of satellites together with the expectation is shown in Figure 6.

2.3.5 The Zel'dovich approximation

The mock galaxy catalogues as described up to this section do not have any large-scale power. Prior to HOD population of the catalogue, this large-scale power can be added with small (defined below) modifications to the particle positions. This is possible because the large-scale power is known to be

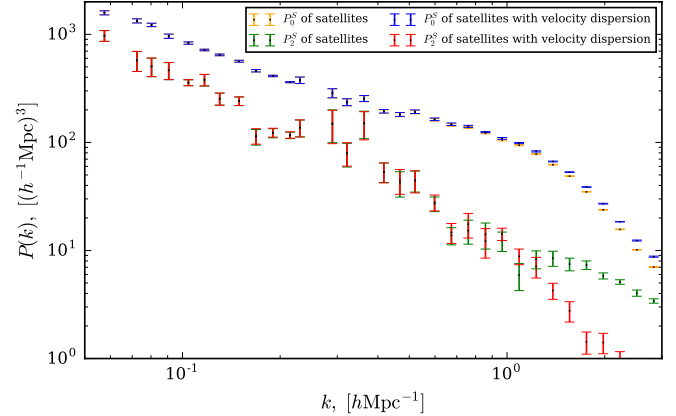


Figure 6. A Gaussian deviate with $\sigma = 5$ was added to the x , y and z -components of the satellite velocities. The plot shows the effect of this on the redshift-space monopole and quadrupole, clearly exhibiting a damping effect on small scales when compared to that measured from satellites with only linear velocities due to ZA. As the monopole is a spherical average, the effect of the damping is much better seen on the quadrupole, which is sensitive to the asphericity resulting from conversion of the peculiar motions of satellite galaxies to redshift-space positions. Here $\Delta_{\text{nl}} = 0.2$, $c_{\text{dm}} = 3.17$, $r_{\text{vir}} = 2.14$, $r_s = 0.676$.

well described by linear theory, and therefore the Zel'dovich approximation (Zel'dovich (1970), ZA) will suffice.

If a displacement field, $\Psi(\mathbf{q}, t)$, is defined that translates a set of particles (fluid elements) from their initial Lagrangian positions, \mathbf{q} , to their comoving positions $\mathbf{r}(\mathbf{q}, t)$ at any later time t i.e.

$$\mathbf{r}(\mathbf{q}, t) = \mathbf{q} + \Psi(\mathbf{q}, t), \quad (43)$$

then the linear theory prediction or Zel'dovich approximation (ZA) is a displacement field given by

$$\tilde{\Psi}(\mathbf{k}, t) = -i \frac{\tilde{\delta}(\mathbf{k}, t)}{k^2} \mathbf{k}, \quad (44)$$

where I obtain $\tilde{\delta}(\mathbf{k}, t) = D_+(t)[P_{\text{lin}}(k)]^{1/2}$ from the expectation of the linear power spectrum to accelerate the convergence and therefore reduce the number of mocks required to observe the expected clustering on average. This is expected to underestimate the variance in the displacement field and should be replaced by a value $\tilde{\delta}(\mathbf{k})$ obtained from sampling an exponential distribution with mean $P_{\text{lin}}(k)$. In the future, I will explore the added accuracy obtained by incorporating this further nuance. As with the overdensity field, I use the FFT to obtain the displacement field given its Fourier components on the grid.

The ZA is a first-order Lagrangian perturbation theory approximation to the displacement field, and it therefore extends beyond the range of applicability of first-order Eulerian perturbation theory, working even in the quasi-linear regime. I find it to work on linear and quasi-linear scales as long as the amplitude of the power spectrum used to generate the displacement field is small enough such that the resulting displacements do not, on average, allow for shell-crossing (Chan 2014) of haloes. This shell-crossing distance is estimated as

$$d_{\text{SC}} = 2 \left(\frac{V/N}{4\pi/3} \right)^{1/3} \quad (45)$$

and averaged displacements that approach or exceed this value will not in general recover the expected linear power spectrum, as the onset of shell crossing leads to the suppression of the power at high k as well as less accurate resolution of the baryon acoustic oscillation (BAO) imprint on the power spectrum.

If one was to estimate the displacement field from a measured power spectrum, that of an N -body simulation for example, one would also have to downweigh the non-linear components of the overdensity field prior to obtaining the displacement field, as the ZA only applies to the linear components of the field. This can be done by defining an R_{nl} such that $\sigma(R_{nl}, z) = 1$ and then convolving the real-space density field with a Gaussian filter of this length scale (Mead & Peacock 2014b). This becomes a multiplication in Fourier space with

$$\tilde{F}_G(k) = e^{-k^2 R_{nl}^2/2}. \quad (46)$$

The expected variance of the smoothed displacement is then given by

$$\sigma_\psi^2(R_{nl}) = \int_{k_{fund}}^{\infty} \frac{e^{-k^2 R_{nl}^2/2} \Delta_{lin}^2(k)}{k^2} d \ln k. \quad (47)$$

Before applying the displacement field to the randomly positioned haloes to give the catalogue the expected large-scale power, one needs to take into account the fact that the clustering of large-mass haloes in an N -body simulation is biased, $\delta_{haloes} = b(\nu)\delta_{matter}$, due to their preferential formation in close proximity (as a result of the effective background provided by the large-scale modes of the overdensity field (see Figure 1 of Tinker (2007))). This is a result of what is referred to as the peak-background split and has the following analytic result for the bias-mass relation as given in Sheth & Tormen (1999):

$$b(\nu) = 1 - \frac{1}{\delta_c} - \frac{\nu}{\delta_c} \left[\frac{d}{d\nu} \ln f(\nu) \right]. \quad (48)$$

The average bias of a sample of haloes of varying mass is then

$$b_{eff} = \frac{\int d\nu b(\nu) f(\nu)/M}{\int d\nu f(\nu)/M}, \quad (49)$$

where the integrals are bounded by ν_{min} and ν_{max} , determined by the least and most massive haloes incorporated. As a test, a displacement field was estimated from a model linear power spectrum. Mass-dependent bias was then applied to the displacement field before moving the haloes so as to generate the expected clustering. This is shown in figure 7. Given a model for the bias-luminosity relation, this can be used to create realistic mock catalogues at a given luminosity.

The ZA also ascribes a large-scale velocity field to the haloes to generate the expected large-scale clustering in redshift-space. Given the linear theory relations: $\dot{\delta} = -\nabla \cdot \mathbf{u}$ and $\delta(\mathbf{r}, a) = D_+(a)\delta(\mathbf{r}, 1)$, one finds that

$$\dot{\delta}(\mathbf{r}, a) = H f \delta(\mathbf{r}, a), \quad (50)$$

thus giving, as H only has a time dependence, and the combination Hf is therefore a multiplicative constant in the spatial Fourier transform,

$$\dot{\Psi} = H f \Psi, \quad (51)$$

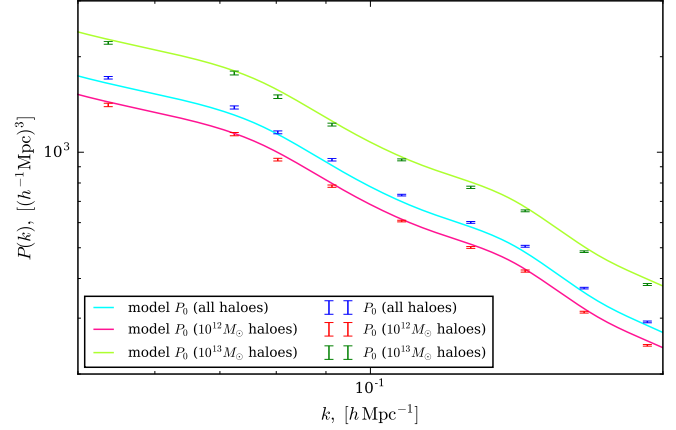


Figure 7. The plot shows the model and measured real-space $P(k)$ for a ZA displacement field with *mass-dependent* displacements given to 128^3 randomly distributed haloes in a $(256 h^{-1} \text{Mpc})^3$ box. The power spectrum is measured separately for the $5 \times 10^{10} M_\odot$ haloes, the $5 \times 10^{11} M_\odot$ haloes and all haloes, with the fraction of haloes of a given mass determined by the halo mass function of Sheth & Tormen (1999), eqn. (40). In this example $f(M_s)/f(M_l) = 1.38$, $b(M_s) = 1.13$, $b(M_l) = 1.42$ and $b_{eff} = 1.22$. The shell-crossing limitation of the ZA is compounded by the mass-dependent displacements, which increase the magnitude for the large mass haloes. Large mass haloes are more rare as they form at peaks of the overdensity field and are therefore more biased, corresponding to them having a larger $P(k)$.

and, as the peculiar velocity is related to the displacement field through $\mathbf{v} = a\dot{\Psi}$, the necessary (peculiar) velocity field is then:

$$\mathbf{v} = aHf\tilde{\Psi} \quad (52)$$

I work in the plane-parallel (distant observer) approximation, with the line of sight along the z -axis, which means that the conversion of the particle positions to redshift-space given this velocity field simply results in the following alteration of the displacements

$$z \rightarrow (1 + f)z, \quad (53)$$

with the other components of the displacement field remaining the same. Figure 1 demonstrates the application of ZA displacements and velocities to a catalogue of dark matter haloes together with the recovery of the expected monopole to quadrupole ratio as given by the Kaiser factors.

The real-space power spectrum after giving the halo catalogue mass-biased displacements becomes

$$P(k) = b_{eff}^2 P_{lin}(k) + \frac{1}{n}, \quad (54)$$

and the redshift-space power spectrum is as given by eqn. (24) with $\beta = f/b_{eff}$.

It is important to note that the mass-dependent biasing of the displacement field does not apply to the velocity field – the equivalence principle requires that locally all haloes share the same large-scale velocity field (Mead & Peacock 2014b).

3 METHOD: EXTENDING MEAD ET AL. ('14)

3.1 Rescaling the mass function

Halo mass functions have been shown to be close to universal (Sheth & Tormen 1999; Tinker et al. 2008a), with a cosmology dependence given simply by the linear variance – the evaluation of eqn. (3) when Δ^2 is that given by linear theory. As the variance is typically smooth, it may be rescaled to that in a second cosmology, but at a different redshift, by simply reinterpreting the length scale, $L'' = sL$. Here unprimed variables denote quantities in the original cosmology, primed ones give those in the target cosmology, and double primed quantities refer to those in the original cosmology after the original catalogue has been rescaled (the length scale and redshift were redefined). For example, a $(1h^{-1}\text{Gpc})^3$ simulation may be remapped to a $(1.2h^{-1}\text{Gpc})^3$ simulation, together with the corresponding movement of all particle positions: $\mathbf{r}_i \rightarrow s\mathbf{r}_i$, scaling the inter-particle displacements by s : $(\mathbf{r}_i - \mathbf{r}_j) \rightarrow s(\mathbf{r}_i - \mathbf{r}_j)$. This implies a remapping of a given halo mass according to

$$M'' = \frac{4}{3}\pi R'^3 \bar{\rho}' = s^3 \left(\frac{\Omega_m(a')}{\Omega_m(a)} \right) M, \quad (55)$$

as the remapped box must contain the appropriate mass, $M'' = \rho_{\text{cosm}} \Omega_m(a') V''$, where $\rho_{\text{cosm}} = 2.78 \times 10^{11} h^2 \text{M}_{\odot} \text{Mpc}^{-3}$. However, this neglects the differences in the halo merger histories in the two cosmologies and is therefore only an approximation.

To minimise the root-mean-square difference between the target and remapped mass functions, the function

$$\delta_{\text{rms}}^2(s, z) = \frac{1}{\ln(R'_2/R'_1)} \int_{R'_1}^{R'_2} \frac{dR}{R} \left[1 - \frac{\sigma(R/s, z)}{\sigma'(R, z')} \right]^2 \quad (56)$$

must be minimised with respect to both s and z . The integral limits, R'_1 and R'_2 , are the radial scales in the target cosmology that encompass a mass corresponding to the smallest and largest mass haloes in the target box. Figure 8 shows a comparison of the remapped linear variance with the target, together with the result of the intermediate step.

3.2 Rescaling in redshift space

The redefinition of lengths in the catalogue also has a corresponding rescaling of velocities that follows the same steps used in obtaining eqn. (52).

The comoving peculiar velocity in units of the Hubble flow across the box is

$$\mathbf{U} = \frac{\mathbf{u}}{HL}, \quad (57)$$

which will be conserved after rescaling, as a given simulation can represent any physical length scale (provided that it is sub-horizon). One has that

$$\mathbf{U} = \left(\frac{f}{L} \right) \psi. \quad (58)$$

As (ψ/L) is conserved by rescaling, \mathbf{U} simply scales with f ; i.e. $(\mathbf{U}/f) = (\mathbf{U}'/f')$. It follows that

$$\mathbf{v}'' = s \left(\frac{a' f' H'}{a f H} \right) \mathbf{v}. \quad (59)$$

Here the double prime again denotes the quantity (proper peculiar velocity) in the current catalogue after it has been rescaled.

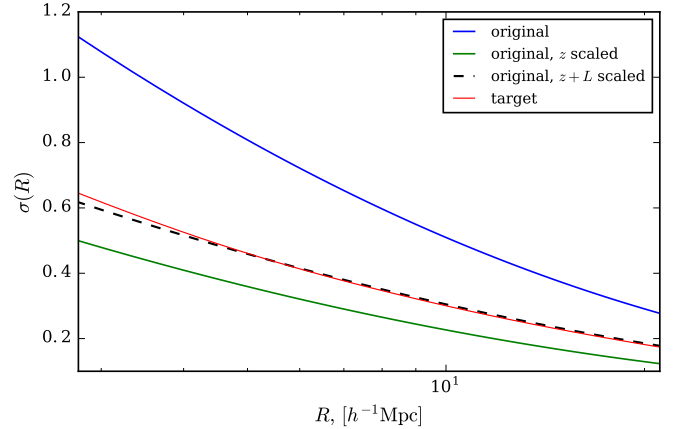


Figure 8. Linear variance for the original, remapped and target cosmologies – following Mead & Peacock (2014a), the original simulation is first scaled in redshift and then length in order to match the halo mass function as accurately as possible. The original simulation is at a redshift: $z = 0.75$ and the minimization returns best fitting parameters of $(s, z') = (1.54, 3.26)$. The algorithm clearly matches the linear variance to a high degree of accuracy. This plot was obtained during the summer project, prior to commencing the SHP.

3.3 Correcting the linear displacements and velocities

After remapping the original simulation, the large-scale power spectrum will not be exactly matched, as the mass function rescaling targets all scales and finds the optimal overall matching. To achieve the closest mass function possible, the second step is therefore to modify the particle positions to restore the correct large-scale power. This is possible using the ZA (§2.3.5) by measuring the overdensity field directly from an N-body simulation and estimating a displacement field using that information, as was done in Eisenstein et al. (2007), for example.

As I use the same linear power spectrum to apply ZA on the mocks, I do not use a filter to downweigh non-linearities that would be imprinted in the density field of an N-body simulation.

To correct the large-scale power of the rescaled simulation, the displacement field required is $\delta\tilde{\Psi}(\mathbf{k}) = \tilde{\Psi}'(\mathbf{k}) - \tilde{\Psi}''(\mathbf{k})$, with $\tilde{\Psi}''(\mathbf{k})$ equivalent to $\tilde{\Psi}(s\mathbf{k})$ according to eqn. (56), as function arguments for the current cosmology become $R \rightarrow R/s$, in Fourier space corresponding to $k \rightarrow sk$. Writing this as

$$\delta\tilde{\Psi}(\mathbf{k}) = \left[\frac{|\tilde{\Psi}'(\mathbf{k})|}{|\tilde{\Psi}''(\mathbf{k})|} - 1 \right] \tilde{\Psi}''(\mathbf{k}) \quad (60)$$

gives the required additional displacements:

$$\delta\tilde{\Psi}(\mathbf{k}) = \left[\sqrt{\frac{\Delta_{\text{lin}}'^2(k, z')}{\Delta_{\text{lin}}^2(sk, z'')}} - 1 \right] \tilde{\Psi}(s\mathbf{k}). \quad (61)$$

This assumes the expectation of the power in the original and target simulations.

As argued in §3.2 of Mead & Peacock (2014a), the mass-dependent displacement corrections must also apply to the

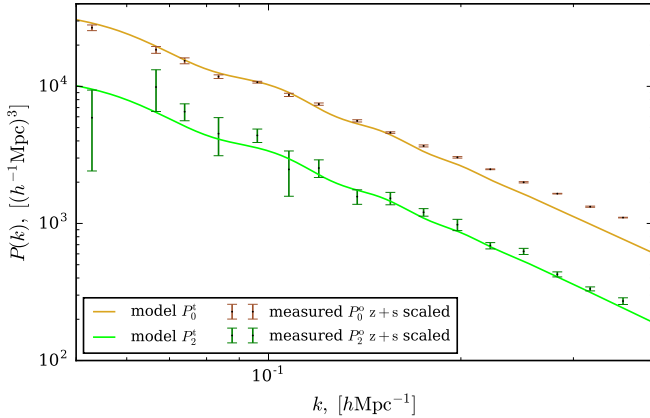


Figure 9. Shown here are the expected and measured redshift-space monopole and quadrupole for twenty ZA mocks with 5×10^5 particles in a $(256 h^{-1} \text{Mpc})^3$ volume which were rescaled from $\{\Omega_m, \Omega_v\} = \{0.24, 0.7\}$ to $\{0.15, 0.85\}$ as outlined in §3.1 and §3.2. After that, additional corrections were applied to the displacement and velocity fields as described in §3.3. The discrepancy beyond $k \simeq 0.2$ between the rescaled P_0 and the target model is due to the shotnoise of the original catalogue not being incorporated in the expectance.

fractional corrections to the displacement field,

$$\delta\Psi \rightarrow b(\nu)\delta\Psi, \quad (62)$$

as described in §2.3.5. However, given that the matching of the halo mass functions will not be perfectly accurate at the percent level (see §4.1.1 further description), this mass dependence of the additional displacements is bound to interfere with the matching of power spectra - b_{eff} will simply be different in the rescaled original and the target catalogues. It may therefore be beneficial to apply an additional correction to the biased fractional displacements that is the ratio of b_{eff}^2 for both catalogues.

The ZA also ascribes additional velocity corrections, as given in eqn. (52), which apply equally when additional displacements are given to the haloes in the original catalogue,

$$\delta\mathbf{v} = a'H'f'\delta\tilde{\Psi}, \quad (63)$$

matching the large-scale power in redshift-space. At this point in the rescaling algorithm, the original catalogue is already in the target cosmology, hence the primed quantities in the previous equation.

Figure 9 shows the rescaling of a mock dark matter-only catalogue - the redshift and length scales were first redefined with the appropriate modifications to the box size, particle positions, particle masses (eqn. 55) and velocities (eqn. 59). Then, additional mass-dependent corrections to the displacement field were applied according to eqn. (61) and eqn. (62). Finally, the velocity corrections corresponding to these additional displacements were applied, as given in eqn. (63). The power spectrum monopole is not recovered accurately beyond $k \simeq 0.2$ due to the fact that the linear power spectrum of the original simulation used for adding these displacements did not incorporate the shotnoise of the catalogue.

3.4 Population of haloes

With the linear part of the power spectrum in the rescaled original catalogue matching that of the target catalogue to a satisfactory degree of accuracy, the final step is to populate the rescaled catalogue according to the HOD (§2.3.2). This is different from the method of Mead & Peacock (2014a), where §3.3 describes a reconstitution step for the non-linear part.

In practice, having used 300 mass bins for the CDF of the ST mass function when assigning masses to haloes, I employ the same number of bins to prepare the HOD parameters ($c_{\text{dm}}, r_{\text{vir}}$) and the CDFs for the NFW radial profile of each halo in the catalogue.

Centrals and satellites are then given positions and velocities as described in full extent in §2.3.2 and §2.3.4, where no mass binning is employed for the mean numbers of centrals and satellites in each halo as this calculation is negligible in terms of computational time, noting that, when working with larger number of mocks in the future, eqns. (33, 34) will also employ binning by mass.

A small concentration dependence is also incorporated into the additional virial velocities of the satellites as per eqn. (35) of Mead & Peacock (2014a), with a further improvement to model the non-linear part more realistically being the replacement of Gaussian damping with Lorentzian damping (eqn. 23).

4 DISCUSSION

In attempting to implement the rescaling algorithm, I have found that the main challenges arise from the transmission of statistical errors through each step. To further add to the fact that I am realistically limited to a small range of k values due to the information content of either the linear power spectrum or one measured from an N-body simulation, in attempting to achieve the highest possible accuracy in that range, I am hindered by the fact the first step in the rescaling, outlined in §3.1, minimises the residual difference in the halo mass functions across the entire range of R values relating to the smallest and biggest-mass haloes in the target cosmology.

It must be noted that the cosmologies used here were similar enough that a remapping of the original power spectrum that matched smoothly the target power spectrum was possible, but it will not be the case in general, for example when transforming to cosmologies which predict a different BAO scale.

In addition to that, this first step of the algorithm is works under the assumption that the halo mass function is universal, which it is only to a certain degree of accuracy and under simplifying assumptions that will no longer hold in precision cosmology. This is discussed in the following subsection.

4.1 Limitations of the method

4.1.1 Universality of the mass function

Multiple investigations into universality of the ST (§2.3.3) halo mass function over a range of redshifts have given the clear signal that it is not, at least to the degree of accuracy

necessary for precision cosmology, as signified by a small redshift-dependence even at $z = 0$ (Lukić et al. 2007) and a much greater deviation from universality at higher redshifts (Cohn & White 2008). As shown in Tinker et al. (2008b), in particular, it is very difficult to calibrate the universal halo mass function of Sheth-Tormen to an appreciable degree of accuracy, and it is necessary to introduce a redshift dependence to do so to better than $\simeq 5\%$ accuracy. Models with collapse thresholds that depend on the environment as well as most modified gravity theories include an environmental dependence of the halo mass function which has to be taken into account to reduce the error of the rescaling and therefore of the covariance estimate at the percent level (Mead & Peacock 2014a).

4.1.2 ZA shell crossing

When creating semi-realistic mock galaxy catalogues, one has to be careful in choosing a linear power spectrum such that its amplitude is not large enough to create a displacement field that would result in shell-crossing of haloes (Chan 2014). When approaching this limit, I first notice that the BAO feature of the linear power spectrum becomes less sharply recovered, and this is a feature of large-scale structure that RSD analyses are concerned with, giving motivation for preserving it together with its effect on the relevant scale of the covariance matrix.

4.1.3 Halo occupancy statistics

In this work, I have skipped the reconstitution step described in §3.3 of (Mead & Peacock 2014a) and directly populate the haloes with galaxies (§2.3.2), aiming to model the non-linear part of the power spectrum based on the HOD prescription adapted to the original and target cosmologies. This is because the latter implementation is simpler, requires less computational resources as it is unnecessary to run a friends-of-friends (Knebe et al. (2013); FOF) or equivalent algorithm to find which particles form distinct haloes, while still ensuring the appropriate cosmology-dependent radial distribution of particles in haloes. Furthermore, the HOD prescription is usually tuned to the survey for which the mock catalogues are generated (Howlett et al. 2015), thus potentially resolving the issue of model dependence of the covariance if the survey data is used to constrain a given cosmological parameter in an range that one can safely assume to contain its value.

As mentioned before, the halo occupancy statistics used here are far from perfect and only serve to give a first estimate of the clustering on small scales. I have only considered the parametrisation for a subsample above a given luminosity threshold, when in reality the parametrisation should be a factor that is allowed to vary over the luminosity range of the catalogue, if only slightly. Not including the dependence of the occupancy of haloes on their formation time (Zhu et al. 2006) is another systematic that could be relevant to this approach due to the redefinition of redshift of the catalogue being a key part of the algorithm.

In addition to that, the stochastic law for number of satellites in a halo (eqn. 34) has been shown to have noticeable deviation from the power law at halo masses approaching

the minimum mass (Kravtsov et al. 2004). This could be corrected for by using more developed HOD prescriptions, such as those described in (Leauthaud et al. 2011; Reddick et al. 2013), or a different model altogether, for example Subhalo abundance matching, contrasted with HOD in Guo et al. (2016).

Moreover, if transforming mock catalogues to simulate the clustering under a different gravity, one would first need to better understand the additional effects, such as shielding, and what imprint they leave on halo inner profiles, especially the concentrations (Lombriser et al. 2014). To do this, it is inevitable that simulations would need to be ran under each gravity model, though one would expect a much smaller number necessary in contrast to that necessary for covariance matrix estimation given that covariance matrices propagate noise from all possible sources to the final result.

A further improvement to the modelling of the inner structure of haloes would be to drop the assumption of the density profile being spherical, allowing for distortions in the shape that one would expect realistically due to the virialised orbits of satellites having been observed to extend noticeably beyond the virial radius in N-body simulations because of tidal effects. Even in the relatively simple model of a triaxial halo (Smith & Watts 2005), a suppression of power at the level of 5% is observed on scales of $k = 1 \text{ hMpc}^{-1}$ to $k = 10 \text{ hMpc}^{-1}$, leading to a substantial underestimate of the power due to clustering on these scales if ignored. The inclusion of this, together with a minimum set of environmental effects that would have the greatest impact on small scales (Zu et al. 2008) alongside the previously outlined points, would be the key next steps to producing mock galaxy catalogues that are to be used in future surveys.

4.2 Linear power spectra

I generated the linear power spectra using the CAMB software (Challinor & Lewis 2005). CAMB integrates the Boltzmann transport equation taking into account the effects of baryons, neutrinos, as well as the decoupling of light from baryonic matter (see <http://cosmologist.info/notes/CAMB.pdf> for more information). CAMB also gives the possibility of extending this to the quasi-linear regime, or it can provide the non-linear spectrum using HALOFIT (Takahashi et al. 2012), and does so for any given (reasonable) set of cosmological parameters.

Unless stated otherwise, the original and the target cosmologies that I use have the parameters and catalogue redshifts of $\{h, \Omega_{m,0}, \Omega_{v,0}, \Omega_{r,0}, \Omega, \gamma, \sigma_8, z\} = \{0.678, 0.24, 0.76, 0.0, 1.0, 0.545, 0.593, 3.5\}$ and $\{0.678, 0.15, 0.85, 0.0, 1.0, 0.545, 0.69, 1.0\}$, respectively.

Alternatives to CAMB for generating power spectra are available (Blas et al. 2011), however, given the extensive use of CAMB in literature and its large range of parameters, it is the software of choice for generating the power spectra used in this work.

4.3 Semi-realistic mock galaxy catalogues

In order to ensure that each part of the rescaling algorithm works as expected, I create mock halo catalogues according to the ZA, outlined in §2.3.5, as is commonly done when

creating initial conditions of N-body simulations with a desired power spectrum (Springel et al. 2005), after which I populate them according to the HOD.

The mock catalogues are generated as such:

- (i) N particles are randomly placed in a given volume. This gives the shot noise power spectrum V/N , as per eqn. (11).
- (ii) The haloes are assigned masses according to the Sheth-Tormen mass function (eqn. 40).
- (iii) The ZA (see §2.3.5) displacement field for a given linear power spectrum is found by taking the inverse Fourier transform of eqn. (44) on a grid.
- (iv) The displacement field is applied, taking into account the mass bias of eqn. (48), by matching each particle to the displacement in the grid cell it is found in.
- (v) Each particle is given the correct (unbiased) linear velocity (eqn. 52).
- (vi) The halo catalogue is populated with centrals and satellites according to HOD as outlined in §2.3.2.
- (vii) Satellites have an additional Gaussian deviate added to their velocities to account for virialised motions (§2.3.4).

These steps result in a galaxy catalogue which accurately reflects clustering on both linear scales as described by linear theory and the ZA, as well as on non-linear scales, as determined by galaxy clustering statistics in N-body simulations and reflected in the HOD approach. A preliminary covariance matrix obtained following the previously outlined steps is shown in figure 10. Despite demonstrating the main features expected, it is clearly discrepant from the covariance of figure 2, and further work is needed to understand why.

I now summarise how a dark matter-only catalogue, created by carrying out the steps up to and including (v), is rescaled in the next section. Table 1 shows several example rows from an HOD-populated input catalogue.

4.4 Rescaling a galaxy catalogue

As described in §3, the rescaling algorithm as described in this work takes in a dark matter-only catalogue in the original cosmology, modifies it so as to best match the halo mass functions of the original and target cosmologies and outputs a rescaled version of the original catalogue, now in the target cosmology, which is then populated according to HOD under the parameter set of the target cosmology. The key steps of the algorithm are summarised again:

- (i) The power spectra are found in the respective cosmologies, either measuring them from N-body simulations or, as done in this work, using a linear matter power spectrum for the chosen set of cosmological parameters.
- (ii) Eqn. (56) is minimised by altering the length scales and the redshift in the original catalogue, giving values of s and z .
- (iii) The box size of the current catalogue is modified to $L'' = sL$, and all the particles are moved accordingly, $\mathbf{r}_i \rightarrow s\mathbf{r}_i$. To conserve mass, all the halo masses are scaled according to eqn. (55).
- (iv) The peculiar velocities of the haloes are rescaled according to eqn. (59) to conserve the invariant comoving velocity in units of Hubble flow across the box.

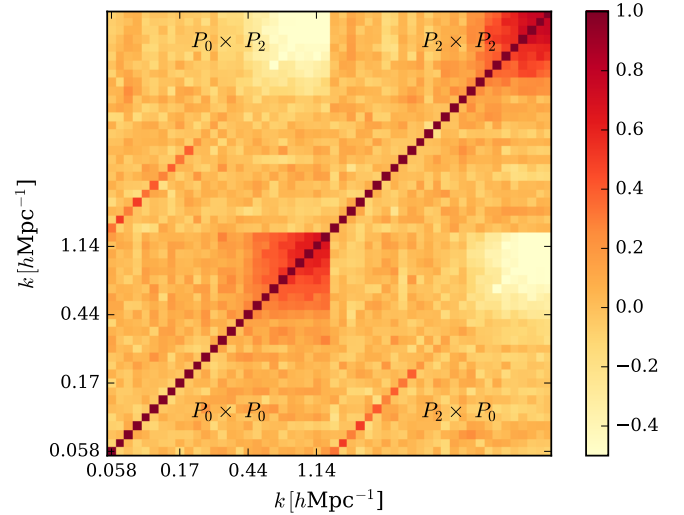


Figure 10. The preliminary covariance matrix is estimated from 200 mocks in the original cosmology at $z = 1.5$ with $\simeq 4 \times 10^5$ galaxies per catalogue generated following the steps of §4.3. No cross-talk is seen between low- k modes, as expected from linear theory which is used as the basis of generating the mocks. However, the non-linearity reflected in the mode-coupling starts only at suspiciously small scales. This is most likely due to a mistake in the allocation of satellite non-linear velocities. This is immediately obvious when contrasting with the covariance matrix of VIPERS mocks in figure 2, with the $P_0 \times P_2$ parts of the matrix being slightly less discrepant.

(v) Mass-dependent fractional displacements are applied to the haloes as per eqn. (61) and eqn. (62) to further correct the linear part of the power spectrum, not forgetting to downweigh the non-linear part of the overdensity field if it is estimated from an N-body simulation using the Gaussian filter of eqn. (46).

(vi) Additional velocities due to the ZA are applied to also correct the redshift-space power spectrum, as given in eqn. (63).

(vii) HOD is applied to the rescaled catalogue, now with the parameter set of the target cosmology: each halo is given at most one central and zero or more satellites depending on its mass, and the satellites are given virial velocities as further described in §2.3.4.

4.5 Improving covariance estimates

In continuing to develop and apply this approach, a method to increase the rate of convergence of the covariance estimate would be of use. One established way to do this is with a Shrinkage estimator (Pope & Szapudi 2008), which could be used in combination with a jackknife resampling of the mocks (Escoffier et al. 2016).

The Shrinkage estimator is applicable when there are multiple alternative sets of data available for estimating the covariance matrix, one of which may be $P(k)$ measurements from a number of N-body simulations, giving on their own an estimate that is less biased but has a high variance (possibly due to the small number of simulations available), and the other an analytic model (linear power spectra, for example) for the low- k part of the matrix, which may be biased

x	y	z	u_x	u_y	u_z	$M_{\text{halo}}/[M_{\odot}]$	$c(M)$	$b(M)$	Type	N_{sat}
100.35	247.69	80.31	8.05	-3.19	-3.77	4.2×10^{12}	4.37	1.91	central	2
97.0	144	100	6.87	-16.5	12.9	2.7×10^{13}	3.43	2.71	central	3
83.9	113	195	12.7	-6.67	2.81	1.4×10^{14}	2.77	4.00	central	19
174.56	159.49	61.16	5.01	-13.83	-8.67	6.6×10^{13}	3.05	3.31	central	12

Table 1. The table shows several example rows from the original simulation after it has been semi-realistically populated with central and satellite galaxies according to the HOD. The coordinates are comoving with the units h^{-1} Mpc and the velocities are comoving peculiar velocities with the units km s^{-1} .

but will have a substantially lower variance. This statistical technique then expresses the estimate for the covariance as a linear combination of the two data sets, with a free parameter that minimises the noise of the estimate. A clear advantage of the shrinkage estimator in the approach of this paper would be the added information available in the linear power spectra that were themselves used to generate the mocks, allowing for an implementation that would need less mocks for an estimate of equal error.

The delete-one-jackknife technique is a resampling of the data that allows an estimate of the covariance from the data itself – in the case of applying this to making an estimate of the covariance of limited number N of catalogues, a single catalogue is “covered up” and the covariance is estimated directly from the other $N - 1$. The approach outlined in this paper would not benefit from this technique as the goal is to reduce the error on the mean of the covariance through using a high number of mock datasets. However, the mock catalogues generated can be used when testing the delete-one-jackknife method in cosmology-specific applications.

In the bootstrap technique of covariance estimation (Friedrich et al. 2016), each catalogue is divided into a number N of roughly equal-sized quasi-independent sub-volumes, effectively increasing the number of simulations available for the estimate N -fold. However, this estimate of the covariance will be in the increasingly non-linear regime and will become increasingly dominated by noise due to the measurements of power spectra coming from samples with smaller populations of galaxies. Moreover, this works under the assumption that the sub-volumes are independent, which is never satisfied fully due to the presence of long-wavelength (low- k) overdensity modes. This is therefore of less use when exploring the formation of large-scale structure through linear RSD with the purpose of testing gravity on these large scales.

4.6 Adaption to modified gravity models

Many modified gravity theories add scalar, vector or tensor quantities to Einstein’s field equation in order to account for the currently observed cosmic acceleration, without the need for a cosmological constant. In future work, I intend to use the rescaling pipeline developed in order to explore a subclass of the models – the Hu-Sawicki $f(R)$ models (Hu & Sawicki 2007), for which a greater number of N-body simulations are available. In these models, the Einstein-Hilbert action: $R - 2\Lambda$ is modified to $f(R) - 2\Lambda$, resulting in equations indistinguishable from GR in the high density (shielded) regimes, which are tightly constrained by solar system tests (Clifton et al. 2012).

An important distinction between GR and $f(R)$ models is that the linear growth rate, D_+ , is scale-dependent in the latter. In this case, the linear growth equation (eqn. 4) becomes (Mead et al. 2015):

$$\frac{\partial^2}{\partial t^2} \tilde{\delta}(k, a) + 2H \frac{\partial}{\partial t} \tilde{\delta}(k, a) = \frac{3}{2} H^2 \Omega_m(a) \left[1 + \frac{1}{3} \left(\frac{\lambda^2 k^2 / a^2}{1 + \lambda^2 k^2 / a^2} \right) \right] \tilde{\delta}(k, a). \quad (64)$$

Appropriate values for the value of $\lambda(a)$ today are given in Table 1 of Mead et al. (2015), for the f_4 , f_5 and f_6 models. Our approach can be extended to transform to these models as described in the same paper.

5 CONCLUSIONS AND FUTURE WORK

In this work I have implemented and used algorithms to rapidly create a set of mock galaxy catalogues for a number of models, in order to estimate the model dependence of the expected statistical errors on the power spectrum multipole moments, as quantified by the covariance matrix. This will give a first estimate of the magnitude of any resulting systematic biases, which must be minimised in preparation for the forthcoming era of precision cosmology.

In order to achieve this, and given the scarcity of N-body simulations spanning the substantial parameter space of interest, I have also implemented and developed upon the rescaling algorithm originally proposed in Angulo & White (2010). This approach takes a N-body simulation in a given model (cosmology) and transforms it, in order to closely resemble that in a different cosmology – at least at the power spectrum level. The aim was to use mock catalogues generated as described in §4.3 so as to be able to quantify the discrepancy between the covariance when estimated from mocks in the target cosmology and that estimated from mocks transformed to best match the target cosmology. This analysis is not completed at the current stage and is to be done in future work.

The mock catalogues were shown to have the expected clustering statistics on both linear and non-linear scales, and can be generated and transformed to the desired cosmology (and, with small modifications, to a desired modified gravity of the Hu-Sawicki $f(R)$ subclass) on the timescale of $\simeq 10$ seconds per mock catalogue of 10^6 galaxies. This is a negligible amount of computational time, and I therefore expect the algorithm as described in this paper to have applications beyond providing covariance matrices for model testing in cosmology. Examples of applications may include comparing the different covariance estimation techniques outlined in §4.5 using realistic datasets, or quickly making comparisons

between how accurately current halo occupancy modelling techniques, not necessarily limited to purely empirically motivated models, simulate clustering in different cosmologies and under gravity theories.

The results of the investigations that this algorithm may facilitate are therefore bound to improve the modelling of non-linear RSD by creating the possibility of quickly exploring a large parameter space where one may have to otherwise rely on semi-analytic models only, which will be limited to models in a small subclass where the necessary results have already been obtained. This will allow future work to easily discredit models that do not show the expected clustering statistics when compared with *N*-body simulations. The method implemented in this work is adaptable to test varying cosmologies, such as WDM, for example, due to the general results for galaxy occupancy statistics, specifically the radial profile and abundance of subhaloes, being available from *N*-body simulations (Macciò et al. 2013; Lovell et al. 2014), with the universality of the NFW/Einasto profiles being preserved in cases of interest (those that are not greatly discrepant from observations).

When prepared as a black-box, the algorithm will provide the groundwork for machinery to quickly generate mock catalogues and therefore covariance matrices in an arbitrarily large parameter space. In future work, the intention is to quantify the model dependence of covariance matrices as well as adapt the rescaling algorithm to modified gravity models, as done in Mead et al. (2015), with the goal of investigating quantitatively the possible degeneracy of the redshift-space power spectrum multipole moments between varying the occupancy of haloes (HOD) and varying the mass function itself. Commonly, the error on the HOD parameters are not propagated to the resulting constraints on $f\sigma_8$ and this may be an avenue for further research – both in quantifying the resulting error, and the development of a solution to this problem.

These will be important steps for obtaining robust constraints on $f\sigma_8$, a principal discriminator between modified gravity models. Placing tight constraints on this parameter is a key goal of future galaxy redshift surveys such as Euclid and WFIRST, which will either confirm or refute the validity of GR on cosmological scales. As a result, the origin of the observed cosmic acceleration may finally be known.

6 ACKNOWLEDGEMENTS

I would like to thank my supervisor, Michael Wilson, for his patience and for pushing me during the project, as well as for providing his own relevant C code for me to use. I would also like to thank the staff at the Royal Observatory of Edinburgh for their help and willingness to answer my questions.

REFERENCES

Alam S., Ata M., Bailey S., Beutler F., 2016, ArXiv e-prints
 Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
 Beutler F., Seo H.-J., Saito S., Chuang C.-H., 2016a, ArXiv e-prints

Beutler F., Seo H.-J., Saito S., Chuang C.-H., 2016b, ArXiv e-prints
 Blas D., Lesgourgues J., Tram T., 2011, J. Cosmology Astropart. Phys., 7, 034
 Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
 Cabré A., Gaztañaga E., 2009, MNRAS, 396, 1119
 Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, MNRAS, 430, 767
 Challinor A., Lewis A., 2005, Phys. Rev. D, 71, 103010
 Chan K. C., 2014, Phys. Rev. D, 89, 083515
 Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Physics Reports, 513, 1
 Cohn J. D., White M., 2008, MNRAS, 385, 2025
 Cole S., Fisher K. B., Weinberg D. H., 1995, MNRAS, 275, 515
 Cooray A., Sheth R., 2002, Physics Reports, 372, 1
 de la Torre S., Guzzo L., 2012, MNRAS, 427, 327
 de la Torre S., Guzzo L., Peacock J. A., Branchini E., Iovino A.,
 Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108
 Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
 Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
 Eardley E., Peacock J., McNaught-Roberts T., Heymans C., Norberg P., Alpaslan M., Baldry I., Bland-Hawthorn J., Brough S., Cluver M., Driver S., Farrow D., Liske J., Loveday J., Robotham A., 2015, Monthly Notices of the Royal Astronomical Society, 448, 3665
 Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, ApJ, 664, 675
 Escoffier S., Cousinou M.-C., Tilquin A., Pisani A., Agüichine A., de la Torre S., Ealet A., Gillard W., Jullo E., 2016, ArXiv e-prints
 Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
 Friedrich O., Seitz S., Eifler T. F., Gruen D., 2016, MNRAS, 456, 2662
 Frigo M., Johnson S. G., 2005, Proceedings of the IEEE, 93, 216
 Gunn J. E., Gott III J. R., 1972, ApJ, 176, 1
 Guo H., Zheng Z., Behroozi P. S., Zehavi I., Chuang C.-H., Comparat J., Favole G., Gottloeber S., Klypin A., Prada F., Rodríguez-Torres S. A., Weinberg D. H., Yepes G., 2016, MNRAS, 459, 3040
 Guzzo L., Pierleoni M., Meneux B., Branchini E., 2008, Nature, 451, 541
 Guzzo L., The Vipers Team 2013, The Messenger, 151, 41
 Hawkins E., Maddox S., Cole S., Lahav O., Madgwick D. S., Norberg P., Peacock J. A., 2003, MNRAS, 346, 78
 Heavens A., 2009, ArXiv e-prints
 Hockney R. W., Eastwood J. W., 1988, Computer simulation using particles
 Howlett C., Ross A. J., Samushia L., Percival W. J., Manera M., 2015, MNRAS, 449, 848
 Hu W., Sawicki I., 2007, Phys. Rev. D, 76, 064004
 Jackson J. C., 1972, MNRAS, 156, 1P
 Kaiser N., 1987, MNRAS, 227, 1
 Kampakoglou M., Benson A. J., 2007, MNRAS, 374, 775
 Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

- Knebe A., Pearce F. R., Lux H., Ascasibar Y., Behroozi P., 2013, *MNRAS*, 435, 1618
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, *MNRAS*, 251, 128
- Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R. H., 2011, *ApJ*, 738, 45
- Lombriser L., Koyama K., Li B., 2014, *??jnlJ. Cosmology Astropart. Phys.*, 3, 021
- Longair M. S., 2008, *Galaxy Formation*
- Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, *MNRAS*, 439, 300
- Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, *MNRAS*, 460, 1214
- Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, *ApJ*, 692, 931
- Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, *ApJ*, 671, 1160
- Macciò A. V., Ruchayskiy O., Boyarsky A., Muñoz-Cuertas J. C., 2013, *MNRAS*, 428, 882
- Mead A. J., Peacock J. A., 2014a, *MNRAS*, 440, 1233
- Mead A. J., Peacock J. A., 2014b, *MNRAS*, 445, 3453
- Mead A. J., Peacock J. A., Lombriser L., Li B., 2015, *MNRAS*, 452, 4203
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Navarro J. F., Ludlow A., Springel V., Wang J., Vogelsberger M., White S. D. M., Jenkins A., Frenk C. S., Helmi A., 2010, *MNRAS*, 402, 21
- Okumura T., Hand N., Seljak U., Vlah Z., Desjacques V., 2015, *Phys. Rev. D*, 92, 103516
- Peacock J. A., 1999, *Cosmological Physics*
- Peacock J. A., Cole S., Norberg P., Baugh C. M., 2001, *Nature*, 410, 169
- Peacock J. A., Dodds S. J., 1994a, *MNRAS*, 267, 1020
- Peacock J. A., Dodds S. J., 1994b, *MNRAS*, 267, 1020
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P. J., Ratra B., 2003, *Reviews of Modern Physics*, 75, 559
- Percival W. J., Ross A. J., Sánchez A. G., 2014, *MNRAS*, 439, 2531
- Pope A. C., Szapudi I., 2008, *MNRAS*, 389, 766
- Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, *MNRAS*, 423, 3018
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, *Numerical recipes in C. The art of scientific computing*
- Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, *ApJ*, 771, 30
- Rota S., Granett B. R., Bel J., Guzzo L., Peacock J. A., 2016, *ArXiv e-prints*
- Seljak U., 2000, *MNRAS*, 318, 203
- Sheth R. K., 1996, *MNRAS*, 279, 1310
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- Smith R. E., Peacock J. A., Jenkins A., White S. D. M., 2003, *MNRAS*, 341, 1311
- Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efsthathiou G., Couchman H. M. P., 2003, *MNRAS*, 341, 1311
- Smith R. E., Watts P. I. R., 2005, *MNRAS*, 360, 203
- Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, *Nature*, 435, 629
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 152
- Taylor A., Joachimi B., Kitching T., 2013, *MNRAS*, 432, 1928
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008a, *ApJ*, 688, 709
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008b, *ApJ*, 688, 709
- Tinker J. L., 2007, *MNRAS*, 374, 477
- Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, *MNRAS*, 442, 2641
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Physics Reports*, 530, 87
- White M., Padmanabhan N., 2015, *??jnlJ. Cosmology Astropart. Phys.*, 12, 058
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Zehavi I., Zheng Z., Weinberg D. H., Blanton M. R., Bahcall N. A., Berlind A. A., Brinkmann J., Frieman J. A., Gunn J. E., Lupton R. H., Nichol R. C., Percival W. J., Schneider D. P., Skibba R. A., Strauss M. A., Tegmark M., York D. G., 2011, *ApJ*, 736, 59
- Zel'dovich Y. B., 1970, *??jnlA&A*, 5, 84
- Zheng Z., Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Davé R., Frenk C. S., Katz N., Lacey C. G., 2005, *ApJ*, 633, 791
- Zhu G., Zheng Z., Lin W. P., Jing Y. P., Kang X., Gao L., 2006, *ApJ*, 639, L5
- Zu Y., Zheng Z., Zhu G., Jing Y. P., 2008, *ApJ*, 686, 41

EXAMPLE C CODE

Approximate lines of code written during SHP: 1500.

```

Catalogue *HOD_satellites(Catalogue *current_catalogue, Spline *variance, Parameters *params) {
    double mean_N_sats, sum_weights, sum_weights_sq, cos_theta_sat, phi_sat, theta_sat;
    double NFW_cumsum, r_sat, x_sat, y_sat, z_sat, vx_sat, vy_sat, vz_sat, sigma_sq;
    double sigma, c, velDisp, cdm_vel_par;
    int satellite_no, satellite_index, cur_massBin;
    sum_weights = sum_weights_sq = mean_r_sat = mean_r_sq_sat = 0.0;
    satellite_no = satellite_index = 0;
    double *mass_lims = mass_min_max(current_catalogue);
    int massBins = 300;
    BinInfo *mass_binInfo = prep_bins(mass_lims[0], mass_lims[1], massBins, LOG_BIN);
    Spline **NFW_cumsums = nfw_cumsums_catalogue(mass_binInfo, 200, variance, params);
    HOD_Parameters **HOD_parameters_catalogue = HOD_params_catalogue(mass_binInfo, variance, params);

    int *no_satellites = malloc((size_t)current_catalogue->particle_no * sizeof(*no_satellites));
    for (int i = 0; i < current_catalogue->particle_no; i++) {
        mean_N_sats = pow((current_catalogue->particles[i]->mass - M0)/M1, alpha);
        no_satellites[i] = gsl_ran_poisson(gsl_ran_r, mean_N_sats);
        sum_weights += (double) no_satellites[i];
        sum_weights_sq += (double) (no_satellites[i] * no_satellites[i]);
        satellite_no += no_satellites[i];
    }
    double weighted_shotnoise_sats = volume*sum_weights_sq / pow(sum_weights, 2.0);
    Particle **satellites = malloc((size_t)satellite_no * sizeof(*satellites));

    for (int i = 0; i < current_catalogue->particle_no; i++) {
        cur_massBin = x_to_bin(mass_binInfo, current_catalogue->particles[i]->mass);
        c = HOD_parameters_catalogue[cur_massBin]->cdm;
        cdm_vel_par = (c*(1.0 - pow(1.0+c, -2.0) - 2.0*log(1.0+c)/(1.0+c))/
            (2.0*pow(log(1.0+c) - c/(1.0+c), 2.0)));
        sigma_sq = ((G*current_catalogue->particles[i]->mass)/
            (3.0*HOD_parameters_catalogue[cur_massBin]->rvir))*cdm_vel_par;
        sigma = sqrt(sigma_sq);
        for (int j = 0; j < no_satellites[i]; j++) {
            cos_theta_sat = randomDouble(-1.0, 1.0);
            theta_sat = acos(cos_theta_sat);
            phi_sat = randomDouble(0.0, 2.0*pi);
            r_sat = splint_generic(NFW_cumsums[cur_massBin], randomDouble(0.0, 1.0));
            x_sat = r_sat * sin(theta_sat) * cos(phi_sat) + current_catalogue->particles[i]->x;
            y_sat = r_sat * sin(theta_sat) * sin(phi_sat) + current_catalogue->particles[i]->y;
            z_sat = r_sat * cos_theta_sat + current_catalogue->particles[i]->z;

            velDisp = gsl_ran_gaussian(gsl_ran_r, sigma);
            vx_sat = current_catalogue->particles[i]->vx + velDisp;
            vy_sat = current_catalogue->particles[i]->vy + velDisp;
            vz_sat = current_catalogue->particles[i]->vz + velDisp;

            char *label = malloc(20 * sizeof(*label));
            sprintf(label, "satellite_%d", satellite_index);
            Particle *satellite = new_particle(x_sat, y_sat, z_sat, vx_sat, vy_sat, vz_sat,
                0.0, label, SATELLITE, current_catalogue->particles[i]->parent_id);
            PBC(satellite);
            satellites[satellite_index] = satellite;
            satellite_index += 1;
        }
    }
    free_all();
    return toReturn;
}

```