An Exploration of Clustering Statistics and Tidal Alignments in the Context of Group Catalogues

by

Nick DeMarchi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The use of three-dimensional galaxy surveys and large cosmological simulations of the galaxy distribution has advanced our understanding of large-scale structure over the last few decades. At the forefront of this are clustering measurements that seek to determine cosmological parameters including the linear growth function, f. When making clustering measurements using group catalogues, whether they be dark matter halos or galaxy clusters, various selections in our catalogues are generally performed that can alter the outputs for our measurements. An example of this is when objects are selected based on a certain mass threshold, a primary bias is introduced that alters our perceived two-point clustering statistics. Additionally, when using spectroscopic data to measure the positions of groups in the Universe, an observational effect caused by peculiar velocities, known as redshift space distortions, can also alter our perceived clustering and create an anisotropic effect in the power spectrum. Furthermore, many recent works have focused on secondary properties, independent of mass, that can alter clustering measurements for groups. In this work, we will explore selection-based contamination effects when finding groups in redshift space with a focus on intrinsic alignments for our selections. The selections used in this work consider halo size as a proxy for its mass and could lead to intrinsic alignments altering the anisotropic signal. We will utilize a current model that includes selection-based effects to the standard redshift space distortion model where this contamination incorporates a dependence on the configuration of the large-scale tidal fields and thus intrinsic alignments. If the selection of our groups is further biased, this can result in systematic errors to f, altering growth of structure measurements. To investigate, we primarily use ΛCDM simulations for dark matter only particles and explore how the selection of groups with preferential halo alignments can produce an anisotropic signal, even in real space. We will also create a variety of mock galaxy catalogues using a halo occupation distribution for our dark matter groups. This population technique is optimized and modified to include similar intrinsic alignment statistics as our dark matter halo catalogue. We will then perform various tests of intrinsic alignments on these galaxy catalogues and attempt to differentiate between anisotropic effects induced by such alignments in our selection with respect to a catalogue that lacks their presence. We find that utilizing various selection-based algorithms produced a weak differential anisotropic signal between these catalogues, most likely as a consequence of significant noise sources. These noise sources will be discussed and potential advancements to the limitations faced are considered for future work.

Acknowledgements

I would like to thank my supervisor Will Percival for over-seeing this project along with Andrej Obuljen for his help and collaboration. I also thank Francisco Villaescusa-Navarro for providing the Quijote simulation snapshots used in this work. I also thank James Taylor and Niayesh Afshordi for agreeing to be on my committee and for their useful discussion. This research was supported by the Waterloo Centre for Astrophysics and the University of Waterloo. I also acknowledge the support provided by Compute Ontario (www.computeontario.ca) and Compute Canada (www.computecanada.ca). Finally, I acknowledge the use of `nbodykit` (Hand et al. 2018), `astropy` (Robitaille et al. 2013), `matplotlib` (Hunter 2007), `numpy` (Harris et al. 2020) and `scipy` (Virtanen et al. 2020).

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction To Large Scale Structure and Galaxy Surveys

1.1 Introduction

The physics of the early Universe can be described in terms of its chemical composition, geometry and density through a combination of theories including Big Bang nucleosynthesis and general relativity, along with observations of the Cosmic Microwave Background. The present day Universe, however, is more complex and the use of galaxy surveys help create a link between the physics of the early Universe and the current observed large-scale structure. By measuring the properties of the early and late Universe, we are able to test models of how it may have formed and the components that make it up. The ΛCDM model is our current benchmark for describing these components, tested and measured in several

experiments such as *Planck* (for recent results see: Aghanim et al. 2020) and the *Baryon Oscillation Spectroscopic Survey*, Schlegel et al. (2009).

Over the last few decades our understanding of large-scale structure in the Universe has seen rapid progress, utilizing precise measurements of the galaxy distribution in surveys in combination with large cosmological simulations. A main focus has been the use of certain statistics to measure the clustering properties of collapsed objects including dark matter halos and/or galaxy clusters. On large-scales, the clustering of collapsed objects is said to be linearly biased with respect to the clustering of the full matter density field. This is based on a simple model stating that the galaxies trace the dark matter up to this overall linear bias constant. This linear relation tells us that larger over-densities should be more strongly clustered, well-described by the peak-background-split model, Bardeen et al. (1986) (see appendix C).

Of late, a considerable amount of attention has been put on understanding secondary clustering properties of dark matter halos and/or galaxy clusters that can produce an additional bias. It has been shown that the clustering properties of collapsed objects can additionally depend on various internal properties as reviewed in Wechsler and Tinker (2018). This secondary bias, often referred to as assembly bias, is a more complicated picture since internal halo properties in turn can be correlated with themselves. Several works have shown that the large-scale tidal fields are a key physical ingredient in the assembly history of collapsed objects. In Obuljen et al. (2019) we see that of these secondary properties, certain non-scalar quantities such as the velocity dispersion tensor, inertia tensor and angular momentum tensor strongly correlate with the large-scale tidal fields. In particular, the strongest correlation was seen between the dimensionless tidal field tensor

and the moment of inertia tensor for a dark matter halo. This is due to the fact that the halo or cluster shapes are dictated by the stretching axis of the tidal field tensor which allows objects to align and cluster in specific ways. We will utilize this correlation with the inertia tensor to make several measurements of alignments in chapters 2 and 3.

In this work, my initial analysis for selection-based contamination as a function of orientation will be performed on dark matter group catalogues utilizing the Quijote simulations, Villaescusa-Navarro et al. (2020). This type of analysis was previously accomplished in Obuljen et al. (2019) where selection contamination from a Friends-of-Friends group finder produced a systematic difference between the redshift space quadrupole and linear Kaiser prediction (see Appendix G). It was also shown that this effect can be very large, even in real space, if we have preferential selection for objects orientated along the line-of-sight. I will present such selection dependencies in both real and redshift space in chapter 2. I will include further insight to this selection effect by examining the stacked halo orientations along the line-of-sight when using a similar group finder along with an example of what a combination of anisotropies from Kaiser theory and selection contamination based on orientation may look like in redshift space. A simple model for this selection contamination will also be briefly derived in this chapter that follows the one found in Hirata (2009). I will then take this analysis one step further and populate our dark matter halos with galaxies using a halo occupation distribution, similar to the one found in Reid et al. (2014). However, I will modify the population technique by creating an algorithm that includes the three-dimensional shape of our halo upon population (code can be found in Appendix H). Thus, our newly populated galaxy clusters should contain similar internal information as the halos regarding their orientation and alignment.

In chapter 3 I will utilize similar statistics as presented in Mandelbaum et al. (2010) but consider halo and cluster shapes as a function of the moment of inertia tensor rather than galaxy imaging data based on pixels. The inertia tensor can be used instead in the case that we have the three-dimensional particle positions. Furthermore, following the analysis by Hopkins et al. (2005), I will use the alignment functions to test for two types of intrinsic alignments for both my dark matter catalogue in chapter 3 and my galaxy cluster catalogues in chapter 4. Pair-counting and alignment function codes were created to accomplish this task and are also attached in Appendix H. From Hirata (2009) it is argued that selection effects due to intrinsic alignments will lead to anisotropic clustering, altering the amplitude of the multipoles. The amplitude of this effect will depend on the selection method used and I will investigate such effects for multiple selection algorithms used on dark matter and galaxy catalogues. Publicly available group finders were used from the python package *nbodykit* (Hand et al. 2018) to make these selections. This package was also used to make measurements the power spectrum. In chapter 4 I will conclude with a comparative analysis between various galaxy group catalogues I create from our dark matter halos to test for potential deviations from linear redshift distortions that intrinsic alignments may have on our groups found in redshift space.

1.2 Motivation for group catalogues

In this work, in order to identify a dark matter halo or galaxy cluster, and to study these internal properties better, various group finding algorithms will be used. At the forefront of such algorithms are the Friends-of-Friends based algorithms which will be discussed

in chapter 4. A few fundamental questions drive the development of these algorithms, including: i) How does one identify a group ? ii) Where is the central object of a group? iii) How can we account for redshift distortions when identifying these groups ? This third question has significant impact to this work and also for measurements of the growth of cosmic structure. When selecting objects using our group finders, redshift space distortions add an additional complication for galaxy survey data in that we do not observe the real space positions of objects. Instead, we observe a combination of their recessional velocity and peculiar velocity from spectroscopic data. To compare between galaxy surveys and simulation data we rely heavily upon the capability in identifying these groups properly in redshift space (for a review of various group finding algorithms and their success in accomplishing this task see Knebe et al. (2011)). Furthermore, in order to precisely measure certain cosmological parameters such as the growth function and primary bias (discussed in more detail in sections 1.6 and 1.9 respectively) we depend on the ability of obtaining real groups that lack any contamination upon selection.

The use of group catalogues is a very interesting topic to study since almost all galaxies exist in groups, forming the building blocks of our Universe. It is essential to extract groups of galaxies from surveys to understand criteria including the evolution of galaxies, halo and cluster properties, the growth of large-scale structure and thus the underlying cosmological model. For instance, when using group catalogues, since we can obtain halo masses, we may split our sample based on this quantity to explore clustering as a function of mass (see figure 1.1). We may also split our group catalogues on other internal properties besides mass, such as their shapes. Thus, we are able to study the clustering of halos split on different bias populations (briefly discussed in chapter 2). We generally refer to this bias

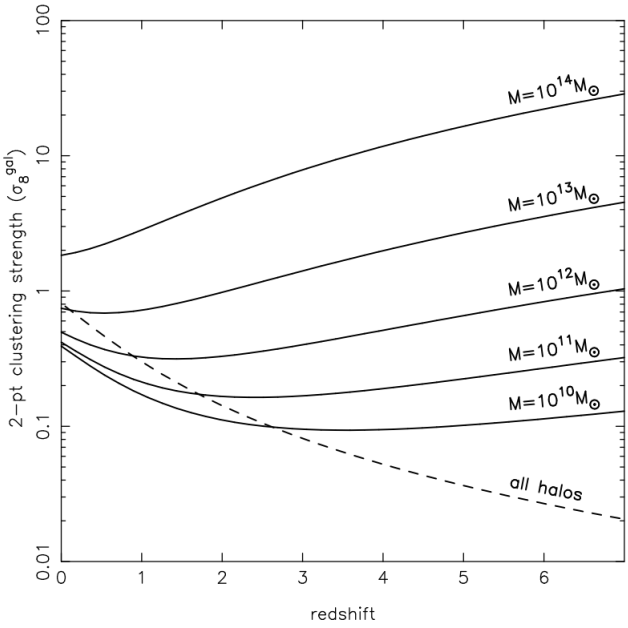

Figure 1.1: Expected large-scale clustering amplitude for halos of different masses as a function of redshift. We notice that halos of different masses will cluster differently and in particular that more massive halos exhibit a stronger clustering amplitude. At low redshifts, halos are expected to grow in mass and will not evolve along these solid lines which are each for halos of a specified mass over time. The dashed line shows the expected growth in clustering for all the mass described by the growth rate, Percival (2013)

as deviations from the linear matter density field for our tracer used, and on large-scales, reduces to a local, linear function (briefly discussed in chapter 1 and in Appendix C by the peak-background split model). For example, splitting our catalogue based on mass is important if one wishes to use a multi-tracer analysis to measure redshift-space distortions, McDonald and Seljak (2009). This approach requires that the split must be made in a clean way that is not affected by systematic errors that would change the clustering, such as selection-based contamination and intrinsic alignment effects. If one could determine this bias with sufficiently small error then halo clustering would become a leading method for measuring this parameter. Therefore, I will utilize groups selected based on some mass threshold and seek to extract information for other internal properties such as preferential orientations and alignments that may be present upon selection. As mentioned, intrinsic alignments are a potential cause of systematic error by contaminating our selection and thus the inferred value of the redshift space distortion parameter when making clustering measurements. Thus, we need to explore the presence of selection-based contamination for our redshift space groups and the size of the effect intrinsic alignments may have in this contamination. This will be the *main focus* of this work and we will aim at extracting this potential effect from our group catalogues utilizing various selections in redshift space.

1.3 The density field and distribution of matter

On large scales, the Universe can be described by the density field, δ, which gives information about the massive clusters and filamentary structures that make up our Cosmic Web (see Bond et al. 1996). Given the quantities $\rho_b(\boldsymbol{r})$ and $\rho(\boldsymbol{r})$ representing the expected

mean density and observed density respectively, we can define $\delta(\boldsymbol{r})$ as:

$$\delta(\boldsymbol{r}) = \frac{\rho(\boldsymbol{r})}{\rho_b(\boldsymbol{r})} - 1 \tag{1.1}$$

In this definition, regions with $\delta > 0$ are considered to be over-dense, while $\delta < 0$ are under-dense. An important property of this density field is the overall spatial distribution of galaxies. However, one important piece of information while studying the distribution of galaxies is that there is no reason to assume that their distribution exactly traces the distribution of mass in the Universe. One can make the following two assumptions about the galaxy distribution:

- Since galaxy formation involves processes like radiative cooling of hot gas and gravitational collapse, one can expect that the efficiency of galaxy formation is related to the density of matter, δ_m.
- Galaxies may only form in special regions and thus the relation between the galaxy distribution and large-scale matter distribution is not a complete description.

Various works have shown, that on large-scales, galaxies are biased tracers of the mass in the Universe, such as in Kaiser (1984). We will discuss this bias in further detail in section 1.9. In general, detailed studies of the distribution of galaxies in surveys along with studies of the overall matter distribution in simulations can help us understand the overall picture of structure formation in the Universe. From *Planck*, at early times, it was shown that δ is well described by a Gaussian distribution and structure formation via gravitational collapse of these over-dense regions gives rise to anisotropic regions at smaller scales where $\delta >> 1$

and isotropic regions at larger scales where $\delta << 1$. In terms of the collapsed regions, we refer to these as dark matter halos which contain several galaxies that have formed within them. A simple model for the spherical collapse of mass into these concentrated regions we call halos is described by the Press-Schechter theory (see Press and Schechter 1974).

In terms of the distribution of matter, figure 1.2 gives an example of observational and simulation data of matter in the Universe. At the top of this figure we have slices in the distribution of galaxies out to a redshift of $z = 0.15$ compiled from the Sloan Digital Sky Survey (SDSS), Blanton et al. (2017). Galaxies are colored according to the ages of their stellar population ($g - r$ magnitude) and the distances of each galaxy are estimated from their observational redshift. Here the redder, more strongly clustered points are galaxies that are made of older stars. The empty regions were not mapped by the SDSS because of dust from our own galaxy that obscures the view of the distant universe. On the bottom of figure 1.2 is a comparison of this galaxy survey with one of the most accurate cosmological simulations, the Bolshoi simulation, which adopts a ΛCDM framework (see Klypin et al. 2011). A crucial observation from these distributions is the fact that the distribution of material is not random but instead shows a variety of structures. Also, a large portion of the galaxies are distributed in low-density filamentary structures that surround larger voids, regions where the galaxy population is low. Embedded in the filamentary structure are the high-density regions containing clusters of galaxies which we described as halos.

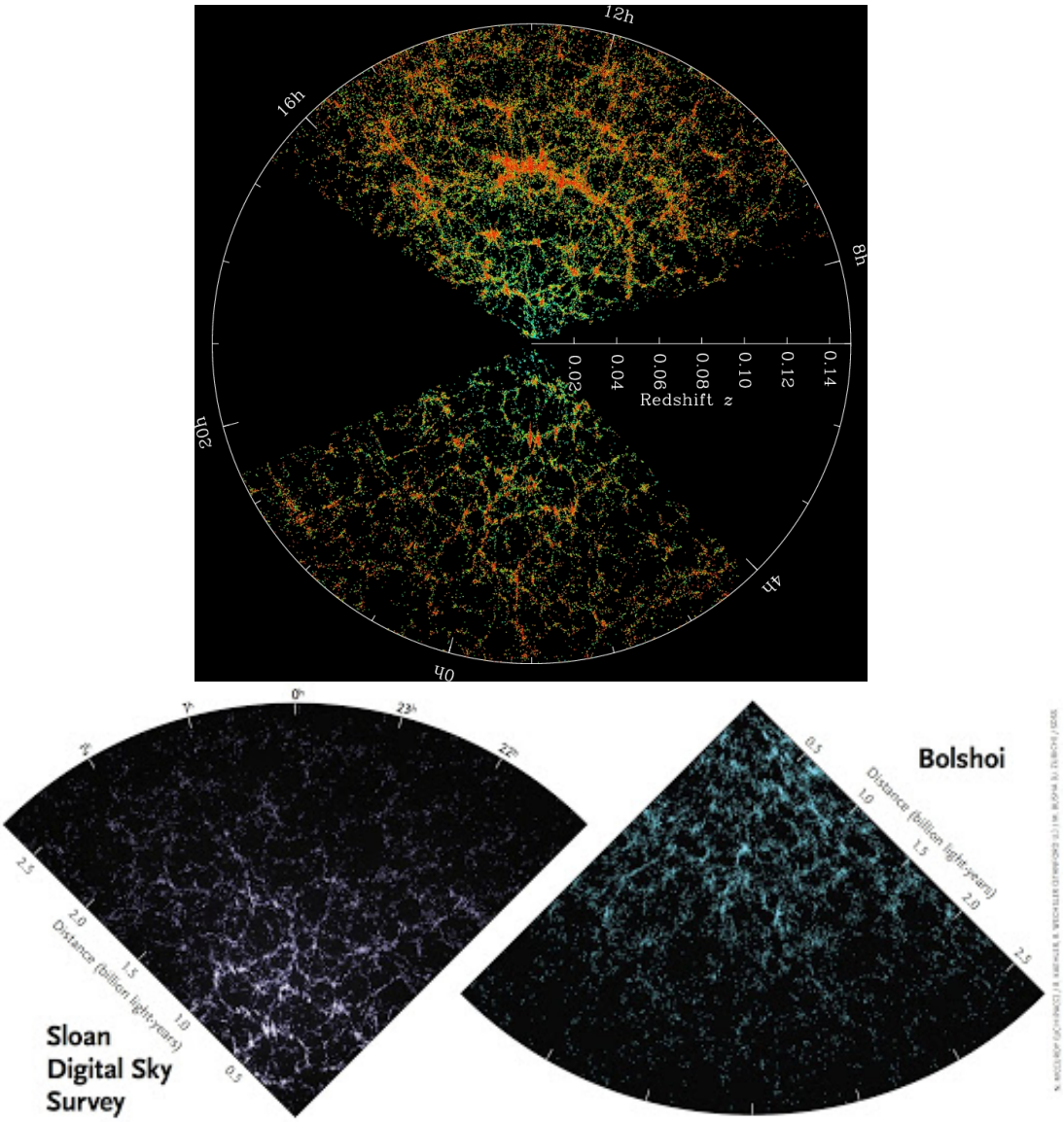

Figure 1.2: *Top:* Slices through the three dimensional map of the distribution of galaxies from SDSS-3. Earth is at the center, and each point represents a galaxy, Blanton et al. (2017). *Bottom:* Slices through the same SDSS survey (left) and the ΛCDM Bolshoi simulation (right), Riebe et al. (2011).

1.4 The correlation function and power spectrum

In order to quantify the distribution of material in the Universe one can adopt the statistics of the two-point correlation function, $\xi(\boldsymbol{r})$, as described in Peebles (1980). In general, $\xi(\boldsymbol{r})$ is defined as the excess number of galaxy pairs found at a given separation, r, with respect to an expected random distribution of pairs. This two-point statistic can be described from the over-density field by:

$$\xi(\boldsymbol{r}) = \langle \delta(\boldsymbol{r}) \delta(\boldsymbol{r'}) \rangle \tag{1.2}$$

In an actual galaxy survey, we do not directly measure the over-densities of each region. However, since we know the positions of objects, we can use this as an estimate for the over-density field. There exists various estimators such as the Landy-Szalay estimator (for a comparison of various estimators, see Kerscher et al. 2000) which is described by:

$$\hat{\xi}_{LS}(\boldsymbol{r}) = \frac{DD - 2DR + RR}{RR} \tag{1.3}$$

Here, DD, DR, RR are the amount of data-data, data-random and random-random pairs respectively, normalized by the total possible pairs for each calculation.

In addition to the correlation function, another important observable that I will mainly be focused on is the power spectrum, $P(\boldsymbol{k})$. In the case of the formation of large-scale structure (LSS) where fluctuations in the density field are drawn from a Gaussian distribution, the power spectrum gives a complete statistical description of these fluctuations. We can define the power spectrum from the expected two-point function of the over-density in

Fourier space. Therefore, we will first consider the Fourier transform of the dimensionless over-density:

$$\delta(\boldsymbol{k}) = \int_{-\infty}^{+\infty} \delta(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} d^3r \tag{1.4}$$

The power spectrum is then defined as:

$$P(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \langle \delta(\boldsymbol{k}) \delta(\boldsymbol{k'}) \rangle \tag{1.5}$$

From equation 1.4 and 1.5, we can deduce that the correlation function and power spectrum form a Fourier pair where the power spectrum is the Fourier transform of the correlation function. In general, we can reduce the power spectrum into a one-dimensional function due to the cosmological principle, such that, $P(\boldsymbol{k}) = P(k)$. However, when calculations of the power spectrum or correlation function are made in a galaxy survey, the redshift distances used are a combination of both the Hubble flow and *peculiar velocity* of the galaxies. Therefore, it should be noted that this relation is not entirely true. For the observed power spectrum, peculiar velocities break down the rotational invariance, changing the expected isotropic clustering. The peculiar velocities of galaxies are deviations from the uniform Hubble expansion generated by the gravitational field of surrounding matter such that the Hubble relation becomes: $v_{rad} = zc = H_o r + v_{pec}$ where r is the radial distance in Mpc. We will discuss in the next section the effects of peculiar velocities on the observed quantities $P(k)$ and $\xi(r)$.

1.5 Peculiar velocities and redshift space distortions

When the 3D distribution of galaxies is mapped out in a galaxy survey, the radial, comoving distance to an object can be determined from the measured observational redshift, z_{ob}. In general, the 3D position of an object is inferred from its right ascension (RA), declination (DEC) and redshift (z_{ob}). There exists two contributions for the observed redshift that affect the measured radial distance of an object. They are:

- The expansion of the Universe (or Hubble flow):

$$r(z) = \int_0^z \frac{cdz'}{H(z')} \tag{1.6}$$

- Change in position (or deviation of Hubble flow) caused by peculiar motions of an object:

$$\boldsymbol{s} = \boldsymbol{r} + (1+z)\frac{\boldsymbol{v}\cdot\hat{r}}{H(z)} \tag{1.7}$$

Here, $\boldsymbol{v}\cdot\hat{r} = v_{los}$ is the peculiar velocity of the object along the observers radial direction, or line-of-sight (LOS). The definition of $\boldsymbol{s}$ tells us that we do not observe galaxies in their *real space* positions, $\boldsymbol{r}$, but instead their *redshift space* positions as a result of v_{los}. The contribution of the peculiar velocity term is generally very small, however, it has a non-negligible impact on the clustering statistics of the matter density field, δ_m, and thus the power spectrum or correlation function. As previously mentioned, the peculiar velocities break down the rotational invariance as defined by the cosmological principle and thus make redshift space clustering anisotropic. This anisotropy is referred to as redshift space distortions (RSD).

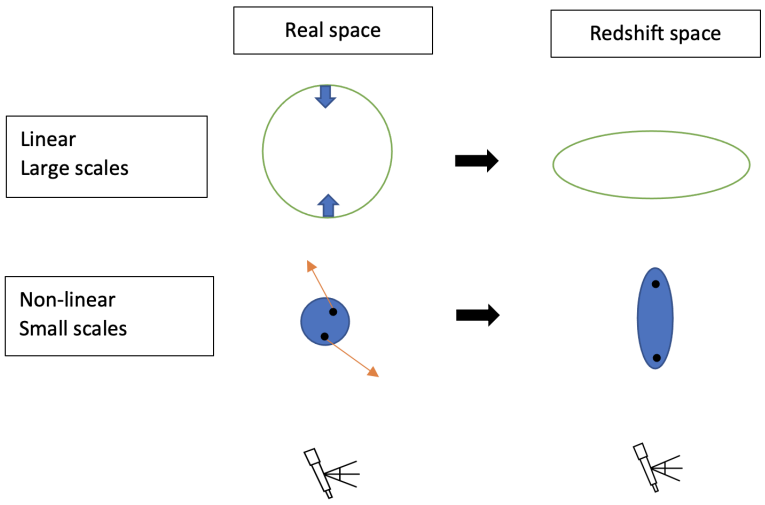

Figure 1.3: Schematic of redshift space distortions for an over-density and the Kaiser effect (*top*) and a virialized object and the FoG effect (*bottom*).

The effects of RSD on clustering can be described on both large scales and small scales in the Universe. At large scales, objects that reside within an over-dense region of matter tend to coherently fall into the over-density, squashing the appearance of the density field along an observers LOS. This squashing along the LOS changes the clustering amplitude such that clustering appears stronger in this direction. This effect was first derived in 1987 by Kaiser (see Kaiser 1987) and is often referred to as the *Kaiser effect*. At small scales, virialized objects have random motions and in this case, the peculiar velocities can stretch out the density field along an observers LOS. This stretching decreases the clustering amplitude in this direction. This effect is often referred to as the Fingers-of-God (FoG) effect (see Jackson 1972). Figure 1.3 gives a visual of these two effects for redshift space observations. These effects are important because in a galaxy survey one can only measure

the clustering quantities as a function of their redshift space position. Therefore, we must consider this effect when calculating the observed power spectrum and correlation function as mentioned. It should be noted that although RSD affects the true, real space positions of objects, we can gain useful insight on theories of gravity and the growth of structure using the linear redshift space velocity field (see Sargent and Turner 1977 and Lahav et al. 1991). This is because, in linear theory, velocities are related to nearby over-densities via the continuity equation:

$$\dot{\delta} + ikv = 0 \tag{1.8}$$

Making a change of variable for our derivative and solving for the velocity, one can obtain (in 1D):

$$v(k) = ifaH\frac{\delta(k)}{k} \tag{1.9}$$

Here, $f = \frac{d\ln D}{d\ln a}$ is the linear growth rate (or growth function) and D is the growth factor for a growing mode. Equation 1.9 presents us with a relation between the linear velocity field and the linear growth rate. From this, we can extract information about the growth function and the cosmological density parameter, which are of importance in developing a standard set of cosmological parameters describing our Universe.

1.6 Modeling RSD in the linear regime

To model RSD in the linear regime, we can first consider the divergence of the three dimensional velocity field, $\nabla \cdot \boldsymbol{v}(\boldsymbol{r}) = -aHf\theta(\boldsymbol{r})$, and its Fourier transform, $\boldsymbol{v}(\boldsymbol{k}) = -iaHf\frac{\boldsymbol{k}}{k^2}\theta(\boldsymbol{k})$. From equation 1.8 and 1.9 it is clear that $\delta(\boldsymbol{k}) = \theta(\boldsymbol{k})$ in the linear regime

where perturbations in δ are small. Next, density conservation from real to redshift space implies that:

$$\{1 + \delta_m^s(\boldsymbol{s})\}d^3\boldsymbol{s} = \{1 + \delta_m(\boldsymbol{r})\}d^3\boldsymbol{r} \tag{1.10}$$

which comes from linear conservation laws for ODE's (in general, we define the Jacobian as, $J = |\frac{\partial \boldsymbol{s_i}}{\partial \boldsymbol{r_j}}|$). Using equation 1.7 and that $\frac{1}{a} = (1 + z)$, and considering a global plane-parallel approximation (i.e. that RSD only depends on one direction, fixed as the LOS of the observer) we can solve this Jacobian matrix to be:

$$J \approx \left(1 - \frac{1}{aH}\frac{\partial v_z}{\partial z}\right)^{-1} \tag{1.11}$$

Here we have chosen the observers LOS to be in the $\hat{\boldsymbol{z}}$ direction. Therefore, we have the relation, $\delta_m^s(\boldsymbol{s}) = J^{-1}\{1 + \delta_m(\boldsymbol{r})\} - 1$. The Fourier component, at linear order, can be derived to give [1]:

$$\begin{aligned}\delta_m^s(\boldsymbol{k}) &= \int \left(\delta_m(\boldsymbol{r}) - \frac{1}{aH}\frac{\partial v_z(\boldsymbol{r})}{\partial z}\right) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} d^3r \\ &= \delta_m(\boldsymbol{k}) - \frac{1}{aH}\int e^{i\boldsymbol{k}\cdot\boldsymbol{r}} d^3r \frac{\partial}{\partial z}\int \frac{v_z(\boldsymbol{k'})}{(2\pi)^3} e^{-i\boldsymbol{k'}\cdot\boldsymbol{r}} d^3k'\end{aligned} \tag{1.12}$$

Substituting $v_z(\boldsymbol{k}) = -iaHf\frac{k_z}{k^2}\theta(\boldsymbol{k})$ and taking the partial derivative in the integrand gives:

$$\delta_m^s(\boldsymbol{k}) = \delta_m(\boldsymbol{k}) + f\int \frac{d^3k'}{(2\pi)^3}\int \frac{k_z k'_z}{k^2} e^{i(\boldsymbol{k}-\boldsymbol{k'})\cdot\boldsymbol{r}}\theta(\boldsymbol{k})d^3r \tag{1.13}$$

[1] See Appendix A for a full derivation for the redshift space density field

Since we have fixed our LOS direction to be $\hat{\boldsymbol{z}}$, we can define the directional cosine as $\mu = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{z}} = \frac{k_z}{k}$. Using the relation above that, $\delta(\boldsymbol{k}) = \theta(\boldsymbol{k})$, we arrive at the following description:

$$\delta_m^s(\boldsymbol{k}) = (1 + f\mu^2)\delta_m(\boldsymbol{k}) \tag{1.14}$$

Equation 1.14 tells us about the transformation of our matter density field in redshift space in terms of the real space density field, the growth function and the directional cosine for our observation.

1.7 Current result of RSD measurements

In general, a key test for dark energy against modified gravity models is through the growth of structure using RSDs. Our current models assume general relativity (GR) to be the correct theory of gravity where predictions about RSDs are made. Our derivation of equation 1.14 which includes the growth function parameter f, assumes GR to be correct and predicts the growth rate of cosmological structure. There are a number of galaxy surveys that have made such cosmological measurements for the growth of structure and having precise measurements of f are of great importance. Such surveys include: WiggleZ (Drinkwater et al. 2010), BOSS (Schlegel et al. 2009), VIMOS public extra-galactic redshift survey (Scodeggio et al. 2018) and DES (Collaboration et al. 2021). As such, constraining the parameter f (or $f\sigma_8$ in a survey where σ_8 is a normalization for our power spectrum calculated as the root-mean-square of mass fluctuation in spheres with radius 8 h^{-1}Mpc) is of importance to test our models against observation. For instance, an analysis to measure

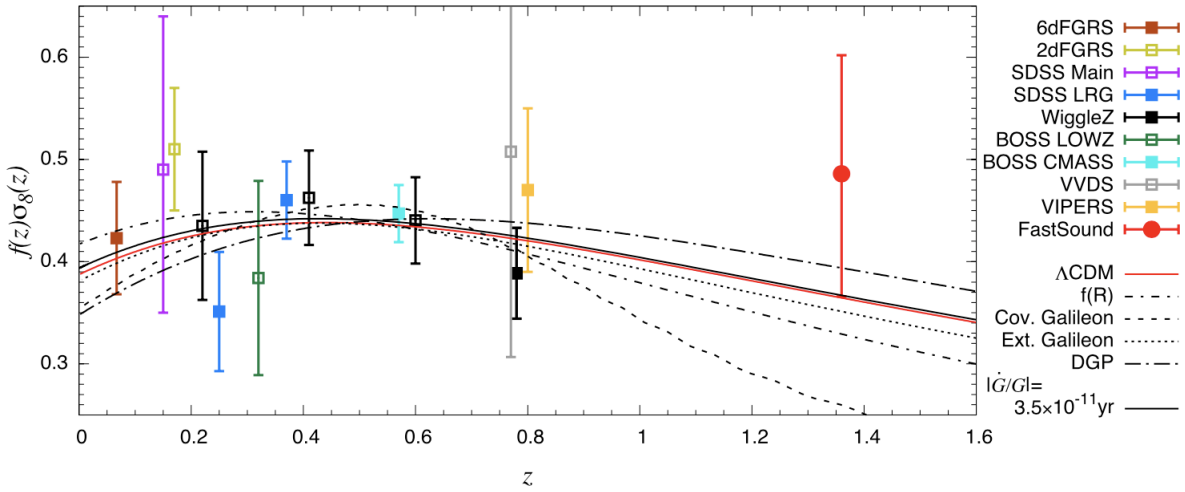

Figure 1.4: Various surveys and their measurements of the logarithmic growth function $f\sigma_8$ for their specific survey depths, Okumura et al. (2016)

$f(z)\sigma_8(z)$ from the SDSS luminous red galaxy data, gives $f(z = 0.25)\sigma_8(z = 0.25) = 0.3930 \pm 0.0457$ and $f(z = 0.37)\sigma_8(z = 0.37) = 0.4328 \pm 0.0370$ where no prior is imposed on the growth rate, Samushia et al. (2012). In comparison, our ΛCDM model with general relativity predicts $f(z = 0.25)\sigma_8(z = 0.25) = 0.4260 \pm 0.0141$ and $f(z = 0.37)\sigma_8(z = 0.37) = 0.4367 \pm 0.0136$, which is consistent with these prior survey measurements. Figure 1.4 presents a variety of surveys and their measurements of f.

1.8 Aside: Quijote simulations and dark matter catalogues

To investigate the properties of dark matter halos, we will use 100 dark matter only Quijote N-body simulations to create the skeleton of a ΛCDM universe Villaescusa-Navarro

et al. (2020). The subset of simulations we use are output snapshots at $z = 0$ of dark matter particles in a periodic box of length 1 h^{-1}Gpc. Each simulation contains 512^3 particles of mass $M_p = 6.57 \times 10^{11}\ h^{-1}M_\odot$. All simulations were run using the following set of cosmological parameters: $\{\Omega_m, \Omega_b, \Omega_\Lambda, n_s, h = 0.3175, 0.049, 0.6825, 0.9624, 0.6711\}$. These parameters are in agreement with the constraints from the Planck satellite (Aghanim et al., 2020).

In order to create dark matter group catalogues we will use an isotropic Friends-of-Friends group finder. This group finder utilizes the three-dimensional particle positions in our simulations to link pairs of particles within some sphere of radius $b_{ll} = 0.2 \times mps$, where mps is the mean particle separation for our data. If we link together a pair of particles then this pair will exist in the same group. Any particles that subsequently also link to either particle in that pair are then also added to the same group. Thus, this group finder will look for regions in our data where objects are clumped together. In terms of defining a group size, this is achieved by setting a minimum 'mass cutoff'. Since all particles in our simulation are of the same mass, we may set a minimum number of particles limit in order to be classified a group. For majority of our catalogues, this minimum is set to be groups that contain more than 20 particles.

1.9 Primary linear bias

In the section 1.6 we went through a *brief* derivation of the redshift space Kaiser factor that describes the anisotropies we see on large scales due to peculiar velocities. However, one final variable must be included in this derivation for equation 1.14 and that is the

galaxy bias term. When we observe galaxies to probe the density field, we must consider that all the mass in the Universe is not within the galaxies and thus they are only *biased* tracers of the full matter density field. It is worth noting that this bias encodes the physics of galaxy formation and so it can be considered in general a non-linear function. However, on large scales, we can reduce this bias term into a linear function as shown in Bardeen et al. (1986) and Cole and Kaiser (1989) by means of the peak-background split model (see Appendix C for further discussion). Thus, on large scales, where density fluctuations are small, it is common to assume a linear bias, denoted b_g, between fluctuations in the mass, δ_m, and fluctuations in the galaxy distribution, δ_g, given by:

$$\delta_g(\boldsymbol{k}) = b_g \delta_m(\boldsymbol{k}) \tag{1.15}$$

In this definition our subscript g will denote that the tracer of the matter density field will be our groups rather than galaxies[2]. In general, the value of b_g depends on the mass of collapsed objects and the redshift at which they are observed such that the clustering of objects is stronger for more massive objects and for objects that formed earlier. Figure 1.5 gives an example of the calculated linear bias used for a dark matter simulation only for two separate mass cuts. We see that for more massive objects we obtain a larger value for our linear bias on large scales (i.e $k > 0.1\ h\mathrm{Mpc}^{-1}$). Therefore, if we start from our linear conservation law described in section 1.6 and use the relation between the density fluctuation of our tracer and the mass given by equation 1.15, we can derive the density

[2]Objects such as galaxies, halos/clusters or voids can be used as biased tracers of the density field

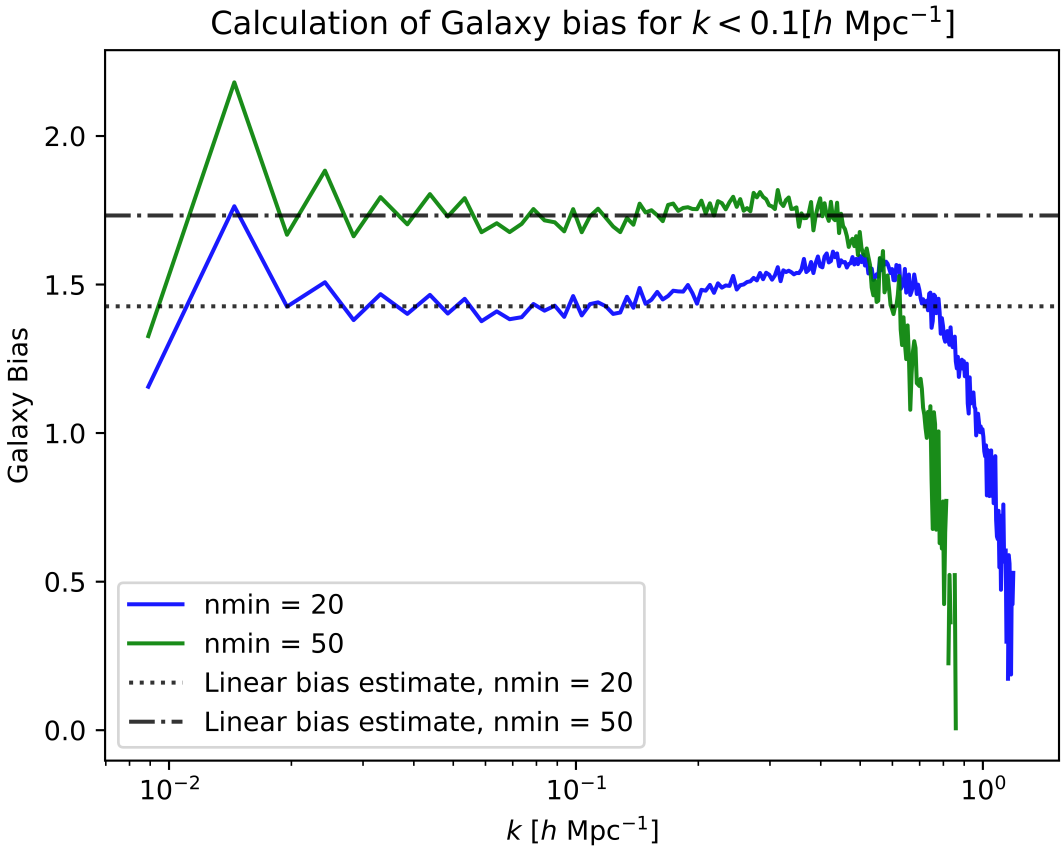

Figure 1.5: An example for calculations of b_g using an N-body dark matter only simulation. The green solid line is for groups with more than 50 particles ($M_h > 3.28 \times 10^{13} h^{-1} M_\odot$) while the blue solid line is for groups with more than 20 particles ($M_h > 1.31 \times 10^{13} h^{-1} M_\odot$). The dotted-dashed (dotted) black line is the approximated linear galaxy bias for scales $k < 0.1\ h\mathrm{Mpc}^{-1}$ for the green (blue) lines.

field to be:

$$\delta_g^s(\boldsymbol{k}) = (b_g + f\mu^2)\delta_m(\boldsymbol{k}) \tag{1.16}$$

Equation 1.16 is the simplest model to described the large scale density fluctuations in redshift space, derived in Kaiser (1987), and thus, is often referred to as the Kaiser formula to describe the Kaiser effect. Comparing equation 1.16 above with equation 1.14, we notice that the anisotropic term in the transformation between real and redshift space comes from the velocity field and does not depend on this primary bias. Our bias will only affect the observed δ_m term and not the anisotropy caused by the peculiar velocity. Using this equation, one can then easily derive the redshift space power spectrum at linear order:

$$P_g^s(\boldsymbol{k}) = P_g^s(k,\mu) = (b_g + f\mu^2)^2 P_m(k) \tag{1.17}$$

Finally, in order to calculate the anisotropic effect that RSD causes, we can decompose our power spectrum, equation 1.17, using the Legendre polynomials [3]:

$$P_\ell^s(k) = \frac{2\ell+1}{2}\int_{-1}^{+1} P^s(k,\mu)\mathcal{L}_\ell(\mu)d\mu \tag{1.18}$$

Only even ℓ values up to $\ell = 4$ survive in this multipole expansion for the Kaiser formula. Thus, we have terms for the monopole, quadrupole and hexadecapole. These are:

$$P_0^s(k) = \left(b_g^2 + \frac{2}{3}b_g f + \frac{1}{5}f^2\right) P_m(k) \tag{1.19}$$

[3] See Appendix B for discussion on Legendre polynomials

$$P_2^s(k) = \left(\frac{4}{3}b_g f + \frac{4}{7}f^2\right) P_m(k) \tag{1.20}$$

$$P_4^s(k) = \frac{8}{35}f^2 P_m(k) \tag{1.21}$$

Since f determines the amplitude of peculiar velocity flows and RSD, in real space we would set $f = 0$ in 1.19, 1.20 and 1.21. Below in figure 1.6 we show the monopole and quadrupole computed for dark matter halos only in a ΛCDM simulation. Although anisotropies are not so obvious from the monopole, we notice that RSD introduces an overall correction of, $P_0^s(k)/P_0^r(k) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)$ where $\beta = f/b_g$. This correction is outlined by the separation in the black solid line and the black dashed line. When observing the quadrupole however, we know that for an isotropic Universe, $P_2^r(k) = 0$ (as shown by the black solid line). RSD that squashes over-densities on large scales produce anisotropies in the radial direction (or along the $\hat{z}$ direction in the plane-parallel approximation). This anisotropy results in a non-zero quadrupole on large scales, which can be seen in the red curve (our measured $P_2^s(k)$) and the corresponding black dashed line given by equation 1.20. Because figure 1.6 shows the measured power spectrum for dark matter halos only by using their centre of mass positions, we do not see the small scale suppression in the redshift space monopole and quadrupole due the FoG effect. We also notice that $P_0^r(k)$ and $P_0^s(k)$ begin to disagree with the measured monopoles on small scales since we start to shift from linear to non-linear theory at larger k values. For our theory curves over-plotted in black described by equations 1.19 and 1.20 we will compute the linear bias as outline in this section in figure 1.5 for groups with more than 20 particles. We will also use a current

theoretical model for f from Lahav et al. (1991) described by:

$$f(z) = \Omega_m(z)^{4/7} + \left(1 + \frac{\Omega_m(z)}{2}\right)\frac{\Omega_\Lambda(z)}{70} \approx \Omega_m(z)^{0.57} \tag{1.22}$$

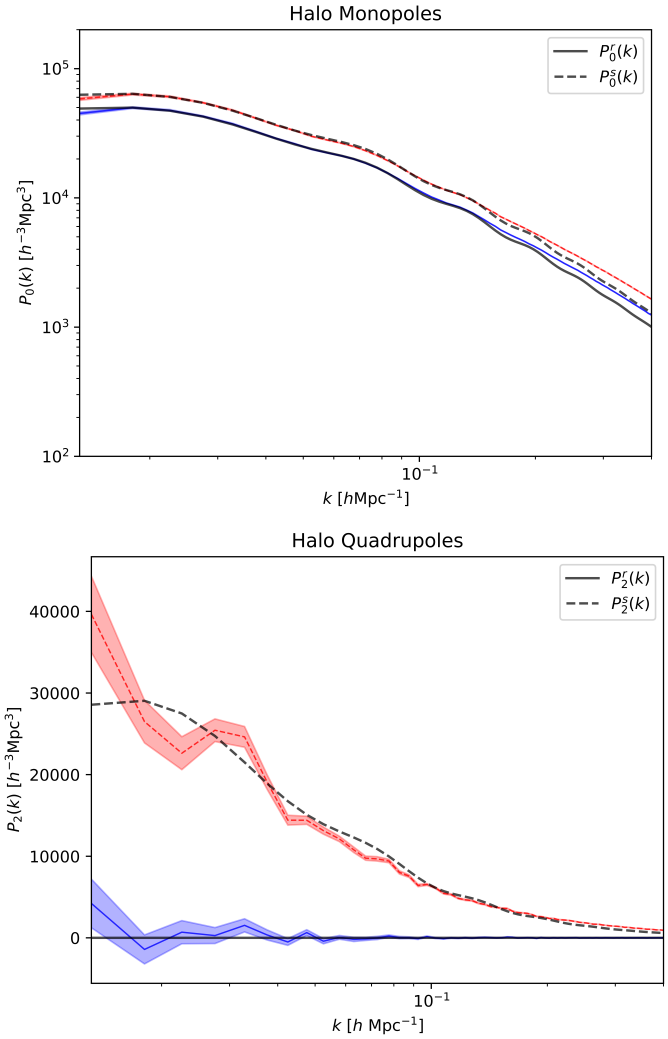

Figure 1.6: *Top:* The measured dark matter halo monopole in real space (blue solid) and redshift space (red dashed) for a ΛCDM simulation. Over plotted are the linear theory curves for real space (black solid) and redshift space (black dashed). *Bottom:* The measured quadrupole in real and redshift space. Line colours are the same as for the monopole.

Chapter 2

Introduction to Assembly Bias

In modern cosmological models, the vast majority of the mass in the Universe (~ 85 %) consists of dark matter (see measured matter densities from Aghanim et al. 2020). In these models, dark matter forms the skeleton of the Universe that can be traced in observations by baryonic material such as galaxies that will form, merge and evolve within dark matter halos. The relationship between galaxies and their host dark matter halo is known as the galaxy-halo connection. These halos represent high density regions in the Universe, seeds from the primordial fluctuations in the matter distribution, which have grown through gravitational instability over time, as discussed in section 1. The galaxy-halo connection began with an understanding of modern galaxy formation theories and large-scale structure formation theories, which is well-described by our benchmark ΛCDM model. There now exist several, large numerical, semi-analytic and hydrodynamic simulations that are used to piece together this connection. In terms of clustering, early works recognize a strong dependence on halo mass and redshift (Mo and White 1996 and Jenkins et al. 1998) which

are assembly history properties for structures in the Universe. In chapter 1 we discussed that a primary property of clustering is related to the mass and this gives rise to a primary bias. However, there also exists a range of secondary halo properties that can influence the assembly histories and clustering. Some of these secondary properties have been shown to depend on the large-scale environment (Hahn et al. 2009) and the large-scale tidal fields (Hirata 2009, Martens et al. 2018, Obuljen et al. 2019). In these works, the alignment of halos or galaxies are studied and shown to correlate with the large-scale tidal fields and can alter the measured large-scale clustering statistics. In this chapter, we will discuss these secondary properties of halos along with the tidal fields and how they can influence our measurements.

2.1 Basics of assembly bias

In section 1 we introduced the primary bias (or galaxy bias), b_g, which tells us that the clustering strength of halos (or galaxies in a survey) is approximately linearly biased compared to the clustering of matter. We have also introduced that more massive halos or early type halos are more strongly clustered than small mass halos or late type halos Kaiser (1984). However, the spatial distribution of galaxies can also correlate with various secondary properties of halos, such as their concentration, spin, formation epoch, angular momentum, shape, velocity dispersion etc. (for example, Faltenbacher and White 2009 and Mao et al. 2017). Furthermore, at a fixed halo mass, M_h, we find that the clustering of halos can exhibit a strong dependence on these secondary properties. We often refer to

this dependence as halo assembly bias, or secondary biasing[1]. In general, we can define assembly bias as the dependence of halo clustering on a halo property other than its mass, M_h. Halo assembly bias is a key effect in understanding the clustering of virialized objects in the Universe and was initially detected while sorting through various halo properties at a given mass in Wechsler et al. (2002) and Gao et al. (2005). Figure 2.1 gives an example of the dependence of clustering based on the selection of halos at a fixed mass as a function of certain halo assembly properties. The figures to the left and middle are slices through a ΛCDM simulation where the red and green dots are halos with high concentration and low concentration respectively. We notice that for high concentrated

[1]Secondary bias is a preferred term since halo assembly histories are not solely responsible for the clustering of halos/clusters. However, most works have continued to use the term assembly bias in describing this effect.

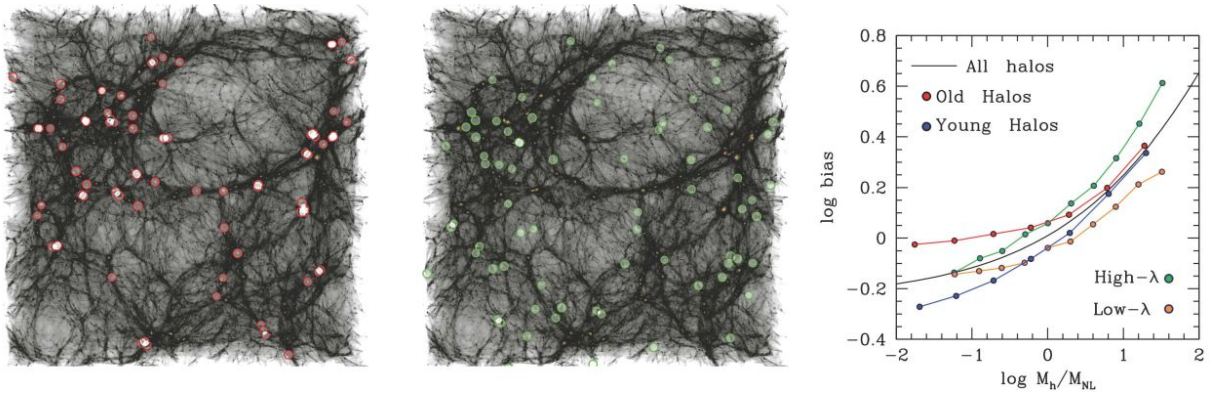

Figure 2.1: Illustration of halo assembly bias based on concentration, formation time, and angular momentum. *Left:* Slice through cosmological simulation at $z = 0$ where red dots are the top 5% of high concentration halos of mass, $m_h = 10^{10.8}$. *Middle:* Slice through same simulation with green dots the top 5% of low concentration halos of the same mass. *Right:* Dependence of halo bias when we split on formation time (red and blue) and angular momentum (green and orange). The solid black line is the overall bias for all halos as a function of their halo mass, Wechsler and Tinker (2018).

halos at a fixed mass, objects tend to be more clustered within the high density regions or filaments, while for the low concentration halos their exists a more random distribution of their locations. On the right of figure 2.1 is a plot generated by Wechsler and Tinker (2018) with data taken from Bett et al. (2007) and Li et al. (2008). This plot shows the dependence of halo bias on secondary halo properties such as formation time (red and blue dots) and angular momentum (orange and green points). It should be noted that this figure considers relatively low mass halos, $M_h = 6.3 \times 10^{10} M_\odot$. The general assembly bias trend seen in these numerical studies is that for smaller mass halos, highly concentrated regions, or older halos, tend to exhibit stronger clustering, whereas for high mass halos, or younger halos, the opposite is true and they are weakly clustered. As mentioned, there exist other properties in the literature such as spin, velocity dispersion and shape that can produce secondary bias trends. Some of these other properties are highlighted in figure 2.2.

In terms of the clustering of massive halos based on concentration and formation time, simple models for structure formation predict this trend such as peaks theory (Dalal et al. 2008) and ellipsoidal dynamics (Desjacques 2008). In these theories, assembly bias arises from the strong correlation between the density of proto-halo structures that will eventually become virialized and collapse and its large scale density environment. The correlation between these produce halos with a large inner density (i.e. high concentration) which will form early and live in under-dense regions compared to halos of the same mass which have low inner density and form late, living inside denser regions of the Universe. In the peaks theory, it is expected that this trend should stay consistent at any fixed halo mass and goes against the inversion we see of the assembly bias trend at fixed high and low masses.

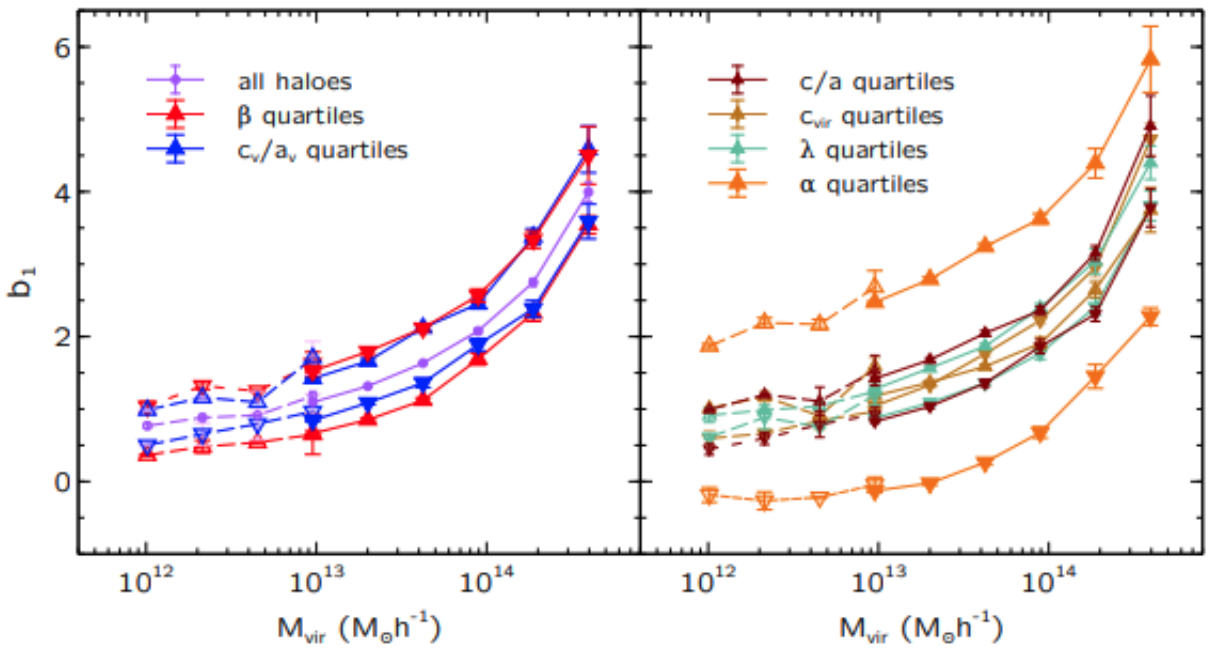

Figure 2.2: Secondary halo bias, b_1 as a function of different populations. Upwards triangles are for the top 25 percent of objects exhibiting a certain halo characteristic and the downwards triangles of for the lowest 25 percent. *Left:* The purple circles show the full halo catalog in each mass bin. Blue triangles are halos split on the internal anisotropic velocity dispersion of dark matter particles and red triangles are split on the halo velocity anisotropy. *Right:* Orange triangles are halos split on tidal anisotropy, dark brown is halos split on shape, light brown is halos split on concentration and light green is halos slit on spin, Ramakrishnan et al. (2019).

A possible explanation for this inversion was proposed in Hahn et al. (2009) where mass accretion is suppressed for low mass halos in filaments due to mass flows being redirected along the filament towards the massive halo as a result of the large-scale tidal environment. This tidal effect slows down the assembly of halos along the large-scale filaments that 'feed' the massive halo along with mass loss for halos in filaments that pass through or close to the massive halo. This suggests that the formation epoch and environment density are

secondary effects induced by the enhanced density of halos in filaments near a massive halo where tidal effects are strongest. This allows for small mass halos to end up being clustered together near over-dense regions where the massive halo exists as seen in the left hand image of figure 2.1. This interesting result, studied using N-body simulations, was further established in a sequence of works using ZOMG simulations of low mass halos in different large scale tidal environments; Borzyszkowski et al. (2014), Romano-Díaz et al. (2017) Garaldi et al. (2017). In this picture, it is argued that structure formation extends beyond the density peaks model and the assembly behaviour of halos can be modulated by their tidal environment[2]. In the next section we will look further into the role of the tidal fields and the large-scale clustering of halos.

2.2 Halo alignments and the tidal field

We have now introduced the systematic effects of galaxy bias, b_g, velocity flows and redshift space distortions, described by f, and selection effects based on internal halo properties referred to as assembly bias. When studying N-body simulations of dark matter only, we introduced in section 2.1 that the assembly behaviour of halos can be modulated by the tidal fields with effects such as the suppression of halo growth and tidal mass loss due to encounters with massive halos. Further, it is shown in Hirata (2009) and Obuljen et al. (2019) that the large-scale tidal fields are responsible for additional systematic errors

[2]Recall that in section 1 we introduced the hierarchical structure formation model based on the Press-Schechter formalism where the mass of a halo is determined by the density peaks on large-scales that exceed some threshold for collapse. In this model, formation time of halos depends only on the halo mass without influence of the large scale environment and our primary bias arises form the peak background split model.

in power spectrum measurements due to the preferential selection of objects aligned along the stretching axis of the tidal field for instance. This secondary effect is understood by orientation-dependant selection effects and can be studied when one makes a selection based on certain non-scalar internal halo properties. For example, objects that exist in under-dense regions will align towards the denser regions (i.e. the stretching axis of the tidal field will be pointing towards the over-densities), while objects in the denser regions will align along the overdensity. Now imagine a case where the stretching axis of the tidal field aligns objects in these dense regions along our LOS. If we have a selection method where objects that are aligned with the LOS are more likely to be picked up, an anisotropic signal will be produced. This is because, for k-modes perpendicular to our LOS, regions of low-density will have more objects that were not aligned with our LOS resulting in deeper troughs and an over-estimate of the amplitude of our Fourier density modes. Alternatively, k-modes parallel to our LOS, where more dense regions contain objects perpendicular to our LOS will dampen the peaks giving an under-estimate for the amplitude of our Fourier density modes. The resulting average of Fourier modes over various angles can then give us this anisotropic signal. Thus, the resulting anisotropy can be defined by the following two criteria:

- Objects must be intrinsically aligned along the stretching axis of the tidal field (known as 'linear alignment' in weak lensing literature)
- Our sample must have a selection effect that depends on this orientation relative to the line of sight

The resulting product (i.e. an inequality in clustering and an inequality in selection) will

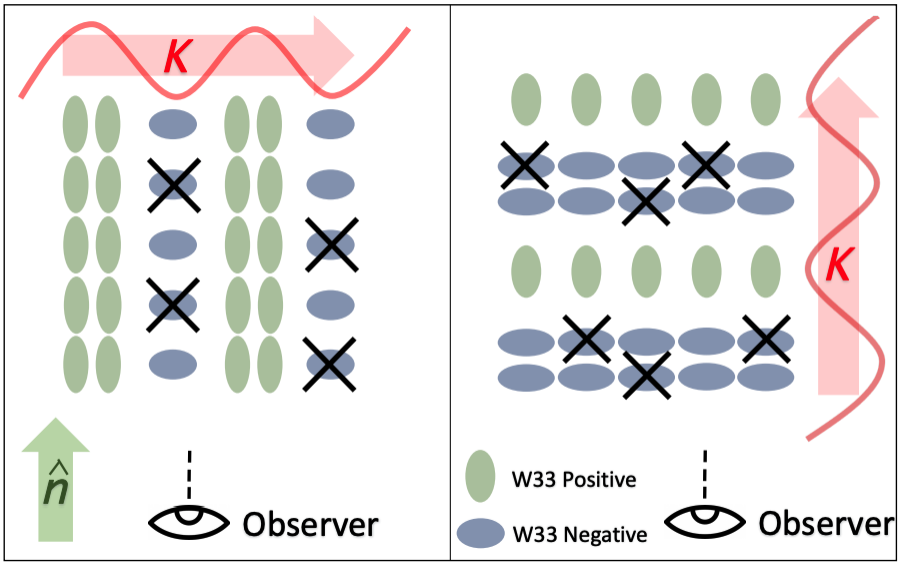

Figure 2.3: Two situations of selection bias that result in anisotropic effects on the resulting k-modes. Assuming objects are less likely to be observed if aligned perpendicular to our line of sight, then some objects will not appear to the observer (crossed out). Fourier modes perpendicular to the line of sight will have deeper troughs (*left*) while Fourier modes parallel to the line of sight will have shallower peaks (*right*), Martens et al. (2018).

create an additional systematic error in the measurement of f which gives information about redshift space distortions. A simple and exaggerated example of this is shown in figure 2.3. In this figure, objects that are aligned perpendicular to the LOS, $\hat{n}$, are less likely to be selected. For k-modes across the LOS, under dense regions will appear even less dense, giving us a higher inferred amplitude, while a k-mode parallel to the LOS can have shallower peaks and thus a lower inferred amplitude. As mentioned, the end result can be an anisotropic signal based on the orientation-dependant selection of these objects and tidal field alignments. Let us now further inspect how we can quantify this effect in the observed power spectrum.

2.3 Intrinsic alignment effects on RSD measurements

The dynamics of dark matter can be described by the gravitational potential, ϕ and its gradient, Laplacian and tidal tensor, the latter given by:

$$T_{ij}(\boldsymbol{x}) = \left[\nabla_i \nabla_j - \frac{1}{3}\delta^K_{ij}\nabla^2\right]\phi(\boldsymbol{x}) \tag{2.1}$$

where δ^K_{ij} denotes the Kronecker delta function. The presence of massive dark matter halos that dominate the local gravitational potential can generate a strong tidal field to its smaller neighbouring halos. Over time, this strong tidal field can induce a shear in the velocity field of the surrounding objects. This field will affect the flow of dark matter such that neighbouring objects will become stretched and aligned in the direction of the massive halo and compressed in the plane perpendicular to it. From Hahn et al. (2007), it was shown that the majority of halos tend to reside in filaments and sheets and so the tidal forces will accumulate coherently. This creates a well-known intrinsic alignment effect of small mass objects being stretched and pointing towards more massive objects.

Let us now take a closer look at how RSDs can be altered by objects with tidal alignments and direction dependant selection effects. From Hirata (2009), it is shown that the best model for intrinsic alignments is to suppose that the large-scale tidal fields can induce a preferred orientation for objects. The selection of a certain orientation would then depend on the observers LOS and configuration of the tidal field. A plausible model for tidal alignments can be derived where the large scale tidal field is taken as a perturbation and then Taylor expanded to its lowest non-vanishing term. The observed galaxy density

field in the linear regime will thus include a direction dependant selection function, $\epsilon(\hat{\boldsymbol{n}}|\boldsymbol{x})$, which was derived to be:

$$\epsilon(\hat{\boldsymbol{n}}|\boldsymbol{x}) = \frac{A}{4\pi G a^2 \rho_m}\left(\hat{n_i}\hat{n_j}\nabla_i\nabla_j - \frac{1}{3}\delta^K_{ij}\nabla^2\right)\phi(\boldsymbol{x}) = \frac{A}{4\pi G a^2 \rho_m}T_{ij}(\boldsymbol{x})\hat{n_i}\hat{n_j} \tag{2.2}$$

where A is the Taylor expansion coefficient. Using the Poisson equation we can simplify equation 2.2 to :

$$\epsilon(\hat{\boldsymbol{n}}|\boldsymbol{x}) = A s_{ij}(\boldsymbol{x})\hat{n_i}\hat{n_j} \tag{2.3}$$

where

$$s_{ij}(\boldsymbol{x}) = \left(\nabla_i\nabla_j\nabla^{-2} - \frac{1}{3}\delta^K_{ij}\right)\delta_m(\boldsymbol{x}) \tag{2.4}$$

is the dimensionless tidal field tensor. Finally, considering the Fourier transform of $\epsilon(\hat{\boldsymbol{n}}|\boldsymbol{x})$ we can now estimate the effect of tidal alignments on the observed galaxy power spectrum in redshift space. In Fourier space, equation 2.3 becomes:

$$\epsilon(\hat{\boldsymbol{n}}|\boldsymbol{k}) = A s_{ij}(\boldsymbol{k})\delta_m(\boldsymbol{k}) = A\left((\hat{n}\cdot\hat{k})^2 - \frac{1}{3}\right)\delta_m(\boldsymbol{k}) \tag{2.5}$$

In Hirata (2009), it is argued that the final observed redshift space galaxy density in the linear regime will be:

$$\delta^s_g(\boldsymbol{k}) = (b_g + f\mu^2)\delta_m(\boldsymbol{k}) + \epsilon(\hat{\boldsymbol{n}}|\boldsymbol{k}) \tag{2.6}$$

Since $s_{ij}(\boldsymbol{k})$ is a rank-2, 3×3, traceless matrix (i.e. $s_{xx}+s_{yy}+s_{zz}=0$) it can be described by a quadrupole moment tensor and thus is a projection of the quadrupole. Therefore, when expanding $\epsilon(\hat{\boldsymbol{n}}|\boldsymbol{k})$ in terms of spherical harmonics as described in Hirata (2009),

the expansion gives a quadrupole term only with all others being zero. This leads to the following:

$$P_g^s(\boldsymbol{k}) = (b_g + f\mu^2)^2 P_m(\boldsymbol{k}) + \left[\frac{4}{3}A(b_g + f\mu^2)\mathcal{L}_2(\mu)P_m(\boldsymbol{k}) + \frac{4}{9}A^2\mathcal{L}_2(\mu)^2 P_m(\boldsymbol{k})\right] \tag{2.7}$$

where the terms in square brackets are from the surviving $l = 2$ components of ϵ which transform as a quadrupole. Since our two tidal alignment terms present are completely determined by δ_m in this model, we can simplify the above expression to:

$$P_g^s(\boldsymbol{k}) = P_g^s(k,\mu) = \left(b_g - \frac{A}{3} + (f + A)\mu^2\right)^2 P_m(k) \tag{2.8}$$

We notice that the redshift space power spectrum retains the same functional form as equation 1.17 in Chapter 1, however, the addition of the intrinsic alignments alter the parameters b_g and f. Equation 2.8 is known as the *tidal alignment model*. In this model, the constant A will affect the apparent rate of structure growth, f, depending on the selection of objects. If objects are selected in a region where the tidal fields compress objects along the LOS, then $A > 0$, enhancing RSD, while if in regions where the tidal field stretches objects along the LOS, then $A < 0$, suppressing RSD. In general, the parameter A describes how tidal fields feed into the orientation of objects and ultimately affect their selection. Thus, the constant A encodes the physical correlation of halo properties with the tidal fields adding to the Kaiser distortion. In the case where tracer objects are not correlated with the tides, $A = 0$ and we recover the general Kaiser model for RSDs.

In figure 2.4 we present an example where we consider selected objects in real space

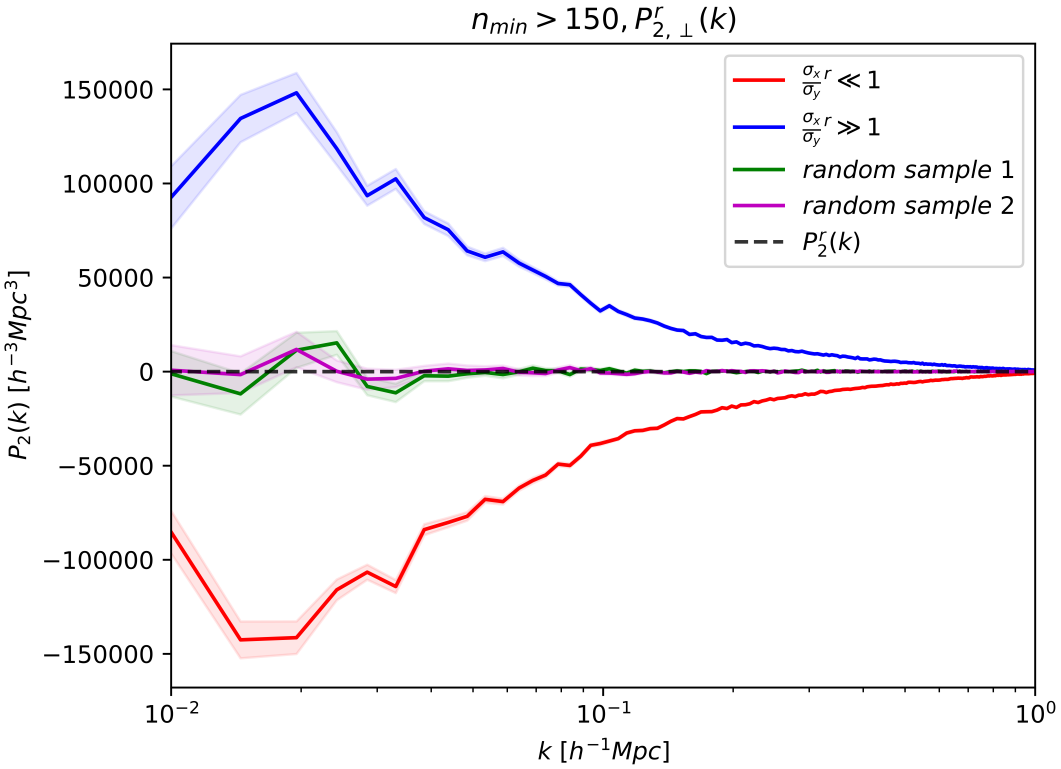

Figure 2.4: Anisotropic signal for real space selection of halos based on their orientation with respect to the LOS (for this example we choose the x-axis to be our LOS). Halos orientated parallel to the LOS (red solid, $\parallel$ halos), perpendicular to the LOS (blue solid, $\perp$ halos), and randomly orientated (green and magenta solid), are composed of halos of mass $M_h > 9.86 \times 10^{13} h^{-1} M_\odot$. The random distributions are of the same size as each aligned halo catalogue. The solution to the Kaiser RSD term when your LOS is perpendicular to the direction RSD is applied is shown by the black dashed line. Our two kinds of selection clearly deviate from this line.

that are either orientated perpendicular or parallel to the LOS (similar to what is shown in figure 2.3). To split our halo catalogue based on orientation we consider the ratio of the dispersion of particles within a halo, $R = \sigma_x/\sigma_y$. Wherever this ratio is greater than one we will have selected a halo that is more aligned with the x-axis and wherever this

ratio is less than one we will have selected a halo that is more aligned with the y-axis. We will also consider the case of the most extreme orientations and select only the top 25% of halos aligned parallel or perpendicular to our LOS (in order to remove the large number of spherical halos in real space at $z = 0$). In this figure, we have also decided to make a mass cut by using only objects with more than 150 particles ($M_h > 9.86 \times 10^{13} h^{-1} M_{\odot}$). The reason for this mass cut is to highlight this effect for some of the more massive halos that will not exist close by in filaments. From this plot, it is obvious that even in real space, we will produce a large scale anisotropic signal based on the selection of objects with a preferred orientation. As described in Obuljen et al. (2019), the correlations seen between non-scalar properties and the tidal fields implies that halos will exhibit an anisotropic assembly bias signal if selected based on these properties. Ultimately, since the tidal field tensor along the LOS has a similar weighting to the quadrupole (i.e. $s_{los}/\delta_m = (\mu^2 - \frac{1}{3})$ and $\mathcal{L}_2(\mu) = \frac{3}{2}(\mu^2 - \frac{1}{3})$), if using an object selection algorithm that is correlated with the tidal field, we will complicate the interpretation of measurements for the growth of cosmic structure using RSDs.

2.4 Tidal alignment in redshift space for our simulation data

In figure 2.4 we illustrated how our selection can introduce an anisotropy to the real space power spectrum, contaminating redshift space clustering measurements of the growth of structure for group catalogues. Next, we would like to make the same selection but on

objects in redshift space space as would be the case in a survey. First, we can show that on average, the orientation of all halos in redshift space will not be random but instead are inherently aligned along the LOS (or where RSD is applied).

In figures 2.5-2.8 we show contours for the shapes of our low mass halos in a 2D plane that includes the direction RSD was applied to (in our case the z-axis in a cube). If we group find in real space, and move objects to redshift space, we notice that on average most objects will be aligned and orientated along the LOS. This is expected for individual halo shapes while using an isotropic group finder. However, if we instead were to make a selection of halos in redshift space (more realistic case) and then take those halos and move them back into real space, we notice that our selection is no longer isotropic and we have added in some preferred orientation and alignment along our LOS. This means that when selecting in redshift space, our group finder preferentially selects object aligned along the LOS in real space, even in our case of low mass halos. For more massive halos, this selection effect becomes even stronger and we have a greater deviation from isotropic selection. In order to avoid this selection dependence we can attempt to select objects based on their orientations perpendicular to our LOS and RSD in redshift space.

In figure 2.9 we show an example of the same selection as figure 2.4 but now for the more realistic case in redshift space. In this example, we can see that our anisotropy is now a combination of the RSD Kaiser effect and the anisotropic assembly bias effect. However, for this example we have identified objects not on their LOS orientations as in a survey but based on their orientation perpendicular to the LOS. This is because if we apply RSD along the z-axis (our LOS) then the xy-plane will not be as affected by RSDs. We can re-derive the quadrupole for the Kaiser effect when our LOS is set to be perpendicular to

RSD to be:

$$
\begin{aligned}
P^{s}_{2,\perp}(k) &= \frac{1}{2\pi}\frac{5}{2}\int_{-1}^{+1}\int_{0}^{2\pi}\left[b_g + f(1-\mu^2)\cos\phi^2\right]^2 P_m(k)\mathcal{L}_2(\mu)d\mu d\phi \\
&= -2P_m(k)\big[\frac{1}{3}b_g f + \frac{1}{7}f^2\big] \\
&= -\frac{1}{2}P^{s}_{2,\parallel}(k)
\end{aligned}
\tag{2.9}
$$

Unfortunately, this is a main advantage that we can easily use in a simulation box but not for a galaxy survey. More care would be needed into setting up a Fast-Fourier-Transform (FFT) algorithm where we can compute the power spectrum in the sky-plane (i.e. perpendicular to our LOS) and then take into consideration the orientation of objects in this plane. Ultimately, we have shown here that anisotropic assembly bias is present where you have a specific selection and RSD can provide us with that selection where objects are preferentially aligned along the LOS.

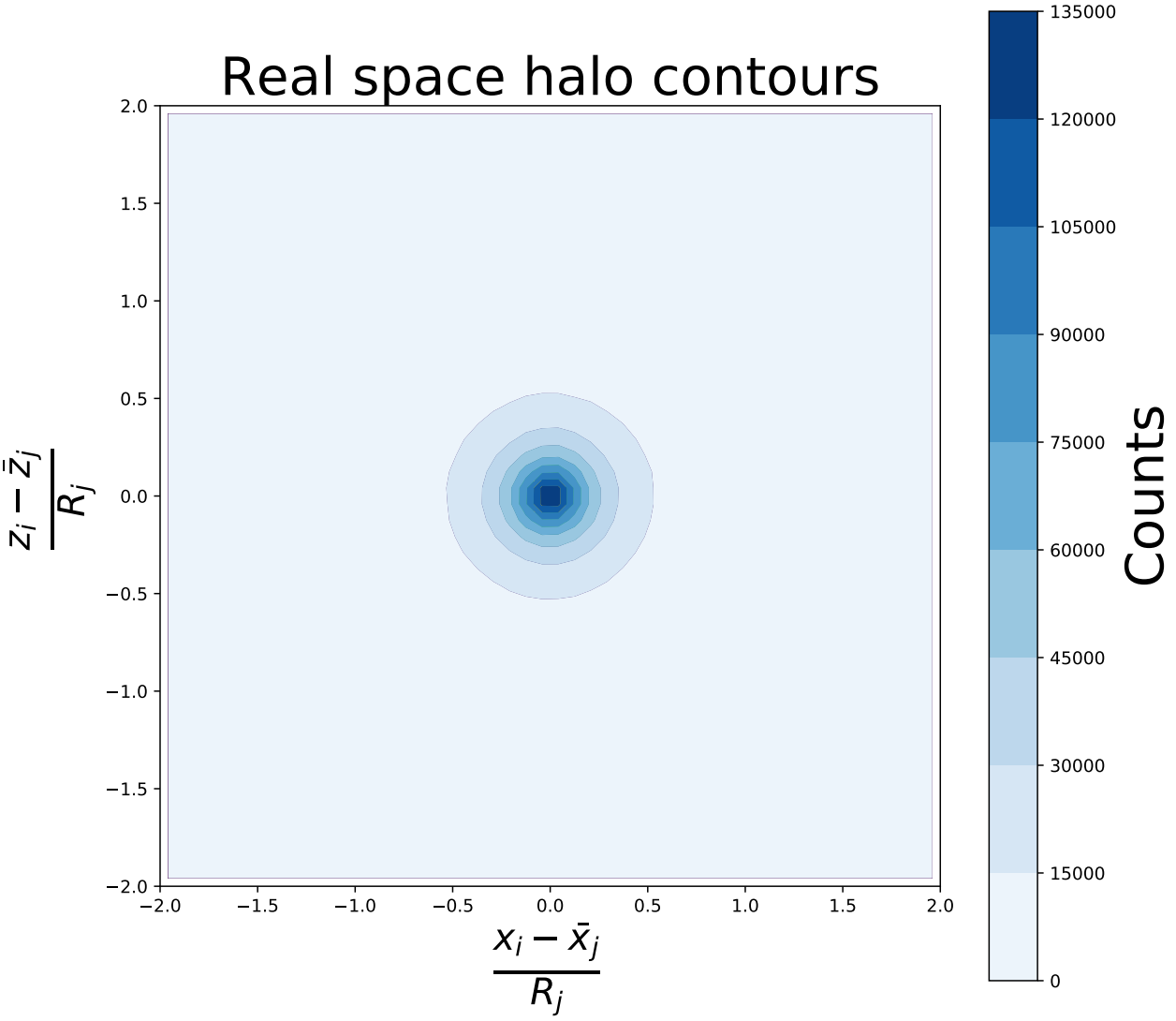

Figure 2.5: Real space contours (10 levels) from Friends-of-Friends selection in the xz-plane for halos with mass $M_h > 9.86 \times 10^{13} h^{-1} M_\odot$. On average, we notice halo shapes give a spherical distribution signifying their orientations are isotropically selected.

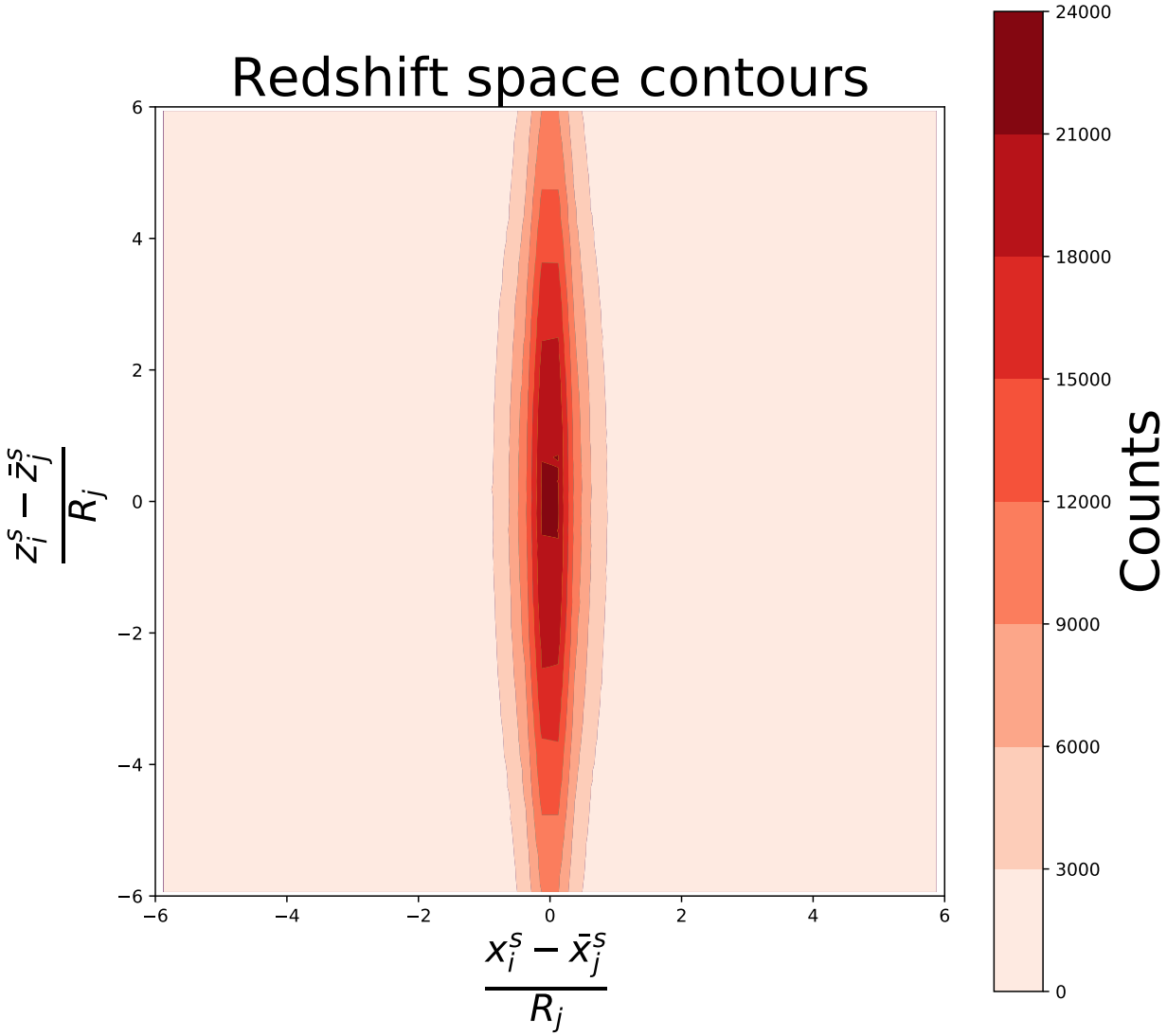

Figure 2.6: Orientation of halos in the same plane as figure 2.5 when we move the particles of each halo into redshift space. Since we use the plane-parallel approximation particles will only be dispersed in one direction, elongating structures along the LOS.

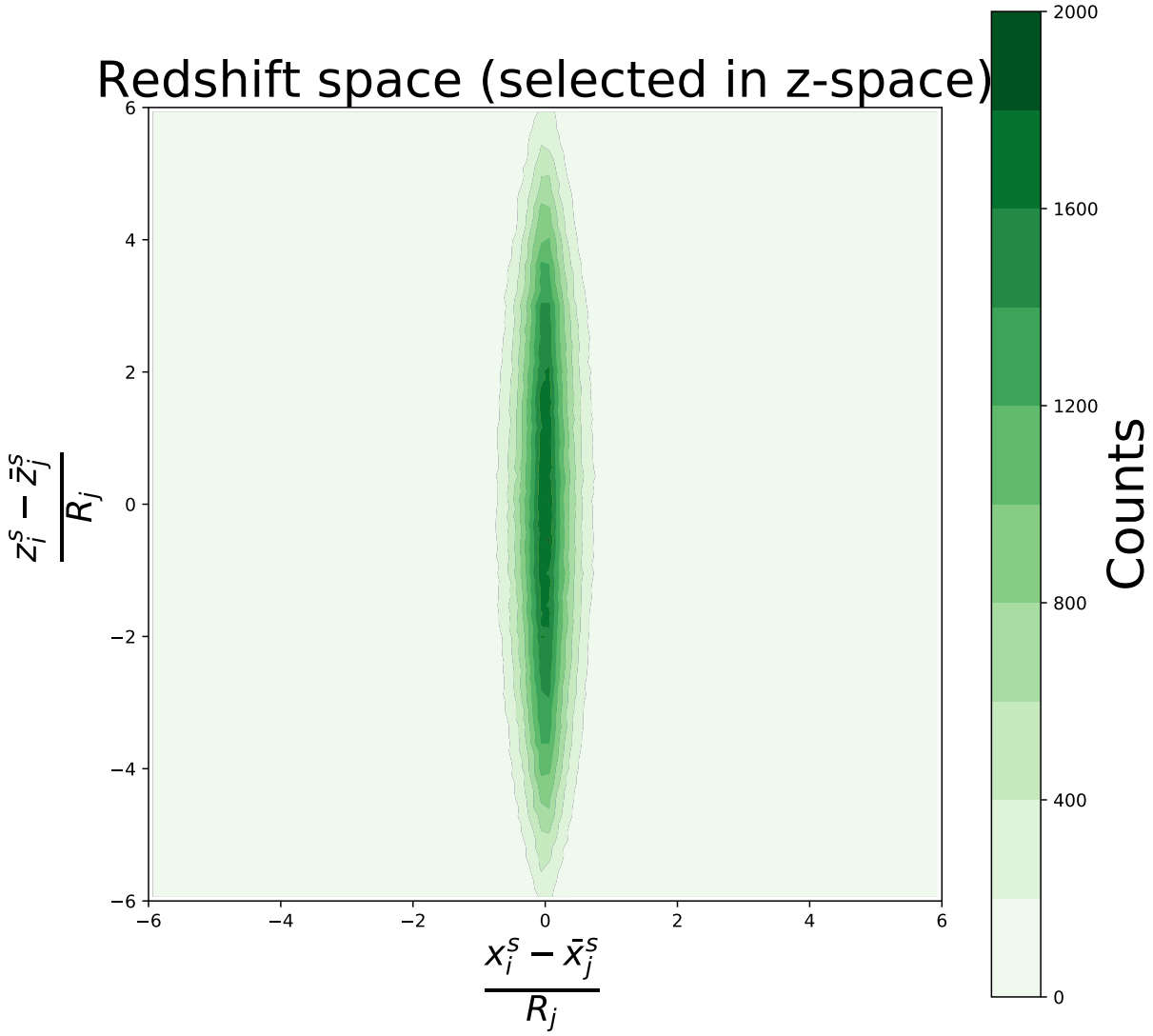

Figure 2.7: Contours for halos found in redshift space using the same isotropic Friends-of-Friends group finder. Overall distribution is equivalent to the case where halos are found in real space and moved to redshift space.

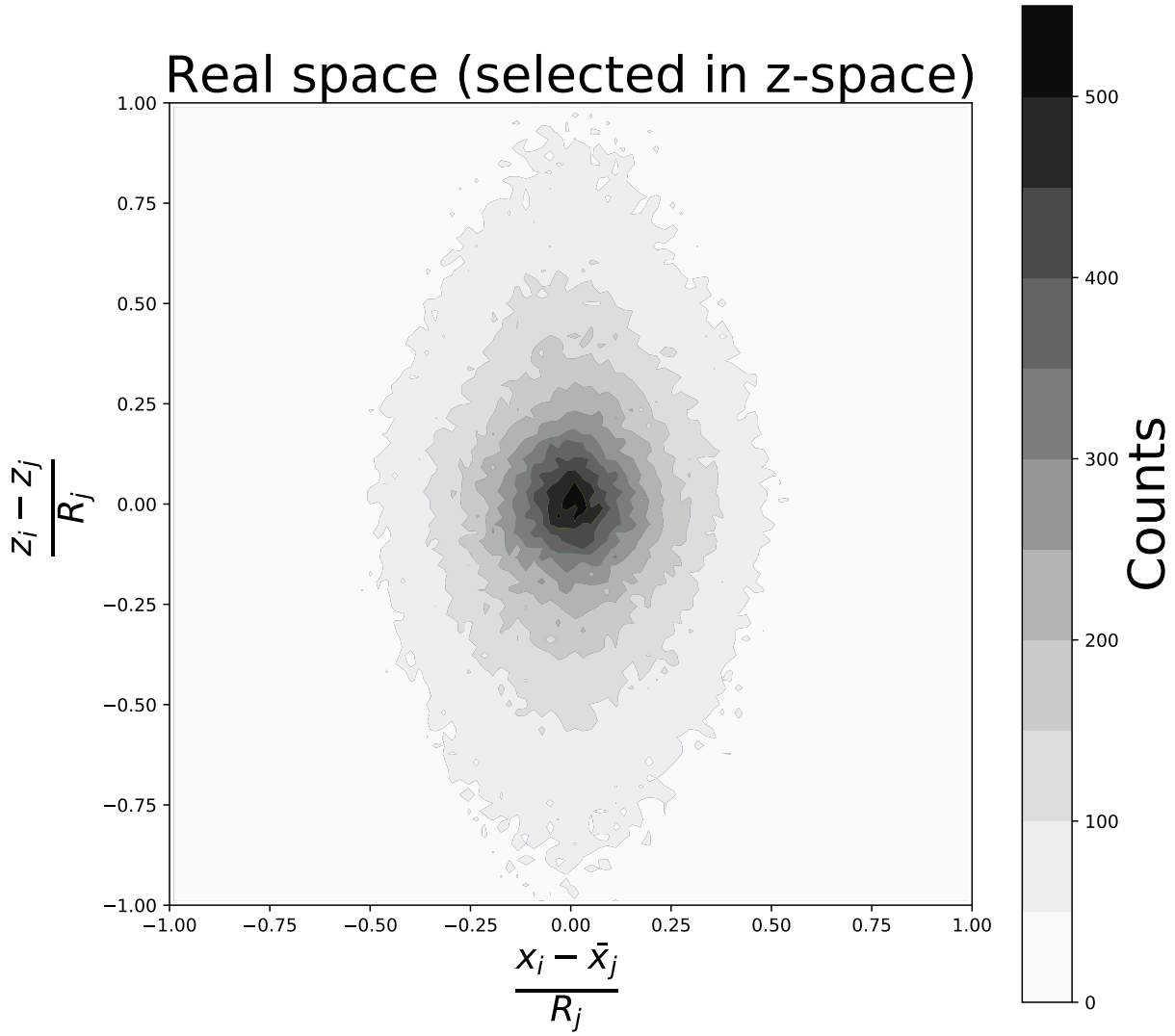

Figure 2.8: Halos found in redshift space moved back into real space. Preferred orientation of halos no longer random and redshift space halos show a preference of being aligned along the LOS. This is the result of redshift space selection effects from a isotropic group finder.

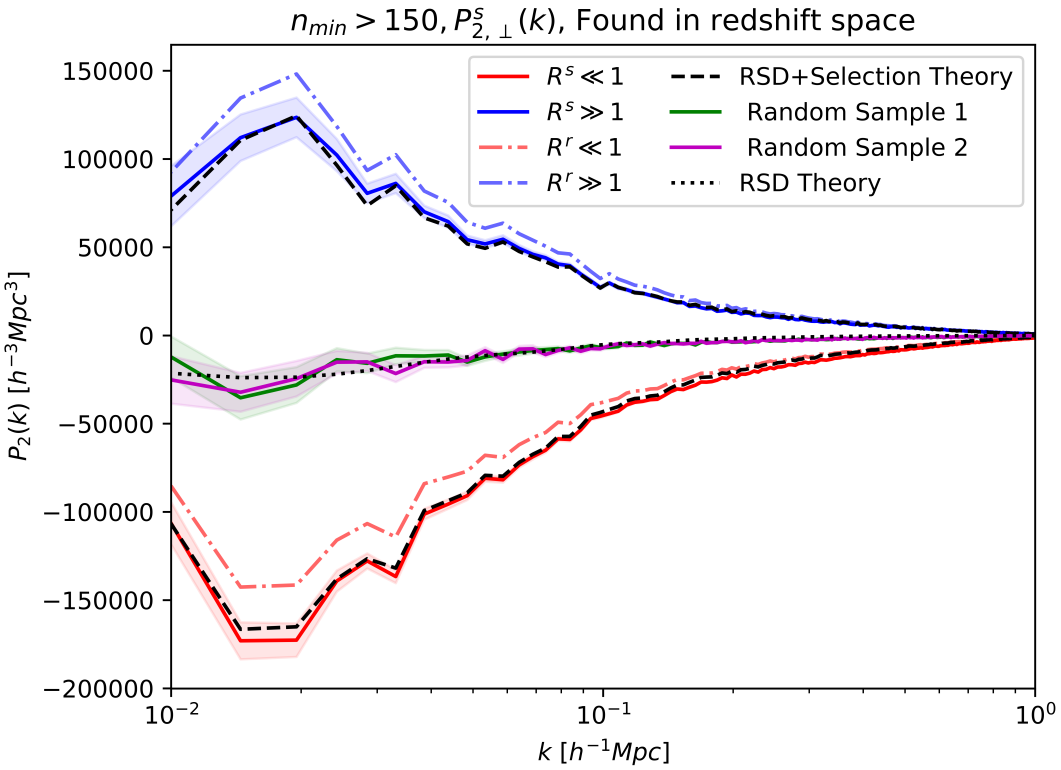

Figure 2.9: Combination of Kaiser RSD and anisotropic selection effects. Catalogues are split based on their orientations in the 'sky-plane' (i.e. the xy-plane in our cube). Here, P_2 is measured perpendicular to the LOS in order to better distinguish object orientation and RSD effects. The solid red (blue) curved are for selection of objects in redshift space that are orientated parallel (perpendicular) to the LOS. The dot-dashed red (blue) lines are the same lines from figure 2.4 for real space selection. The green and magenta lines are two random samples for objects selected in redshift space. As we can see the random samples to not exhibit an additional anisotropic selection effect and follow the linear RSD theory curve (dotted black line). The dashed black line is the expected anisotropic signal if we combine real space selection effects and Kaiser RSD.

Chapter 3

Halo Alignments From a Weak Lensing Perspective

In this Chapter we will explore a different way of detecting the alignments responsible for our additional anisotropic signal. This will be achieved using the statistics from weak gravitational lensing studies, and in particular, intrinsic alignment (IA) measurements such as the the shear-shear and galaxy-shear correlation functions. We will also use the alignment functions from IA studies which is a statistic that can detect the presence of tidal alignments for a range of scales. In Chapter 2 we introduced the tidal field alignments and how an anisotropic signal can be measured if splitting or selecting your catalogue based on specific internal properties that correlate with the large-scale tidal field. In Obuljen et al. (2019), it was shown that the strongest correlation is seen between the tidal field tensor, s_{ij}, and the inertia tensor, I_{ij}. For this chapter, without having to explicitly split our catalogue based on orientation, we are able to show the presence of these alignments

in the statistics mentioned above. Utilizing the inertia tensor we are able to find the tidal fields stretching axis of our halos and compute our various statistics.

3.1 Intrinsic alignment measurements

The variety of galaxy morphologies we observe is related to their interactions with the local environment influencing their evolution along with the initial conditions of their formation. The connection of galaxy morphology and the local environment was made early on by Hubble (1926) and further established in work by Dressler (1980) with the morphology-density relation. Furthermore, the orientations of galaxies, their alignments with local structure, and the impact of their shapes on measurements of gravitational weak lensing has been a prominent area of research in the 21st century as reviewed in Joachimi et al. (2015). We will adopt the statistics from such studies in the context of orientations and shapes of dark matter halos and galaxy clusters.

A crucial ingredient in weak lensing studies is the measurement of the alignments of galaxy shapes. This can be quantified by using a galaxies ellipticity, ϵ_i, and position angle, $\theta_{p,i}$, with respect to some local coordinate system. We should note that although most works of intrinsic alignments derived from weak lensing focus on galaxy shapes and orientations as contamination, we may apply similar principles and statistics to galaxy clusters or dark matter halos to see if orientations of clusters/halos are correlated.

The ellipticity of a galaxy or cluster can be calculated using the quadrupole moments of the surface brightness distribution of an image. In weak lensing studies this can be

expressed by,

$$Q_{ij} = \frac{1}{F_0} \int x_i x_j W(\boldsymbol{x}) \Sigma(\boldsymbol{x}) d\boldsymbol{x} \tag{3.1}$$

where $\boldsymbol{x}$ is the two-dimensional position on the sky, $\Sigma(\boldsymbol{x})$ is the surface brightness of the galaxy, F_0 is the weighted monopole moment (corresponding to the flux in the case of unweighted moments) and $W(\boldsymbol{x})$ is a weight function. In other words, Q_{ij} will give information on the shapes of a galaxy or cluster such that a circular image will have $Q_{xx} = Q_{yy}$ and $Q_{xy} = 0$. However, in the case of our N-body simulation data, where we do not have the surface brightness distribution of objects, we can easily quantify the shapes of our clusters using the traceless moment of inertia tensor instead,

$$I_{ij} = \frac{1}{N} \sum_{k=1}^{N} \Delta x_{k,i} \Delta x_{k,j} \tag{3.2}$$

where a cluster has N particles and $\Delta x_{k,i}$ is the i^{th} component of the unit vector $\Delta \hat{x}_k$ pointing from the cluster centre of mass to the position of the k^{th} particle. From here, if we consider the sky-plane to be the xy-plane in our simulation, a good measure for the ellipticities ϵ_1, ϵ_2 (sometimes referred to as third flattening) can be calculated as:

$$\epsilon_1 = \frac{I_{xx} - I_{yy}}{I_{xx} + I_{yy} + 2\sqrt{I_{xx} I_{yy} - I_{xy}^2}} \tag{3.3}$$

$$\epsilon_2 = \frac{2 I_{xy}}{I_{xx} + I_{yy} + 2\sqrt{I_{xx} I_{yy} - I_{xy}^2}} \tag{3.4}$$

Figure 3.1 is a sample of the orientations of objects in this plane for different values of ϵ_1

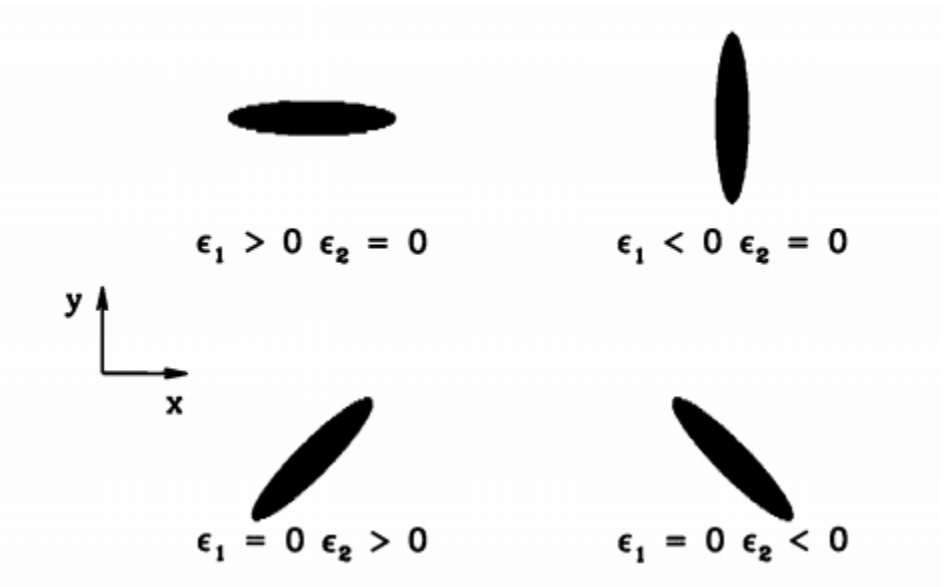

Figure 3.1: Ellipticities ϵ_1 and ϵ_2 in the xy-plane. Circular objects will have $\epsilon_1 = \epsilon_2 = 0$, Dodelson (2003).

and ϵ_2. As mentioned at the beginning of this chapter, using both the ellipticities and the position angle we may derive our statistic for intrinsic alignments in a local environment. The position angle (or pointing angle[1]), θ_p, can be defined as the radial vector that joins a pair of objects together. In the sky-plane we will define the position angle relative to the x-axis. Thus, using our ellipticities and position angle for a pair of objects we can now compute the tangential and cross ellipticities to quantify our shape correlation measurements[2]. These are:

$$\epsilon_+ = -\big(\epsilon_1 \cos 2\theta_p + \epsilon_2 \sin 2\theta_p\big) \tag{3.5}$$

[1]Depends on the context of use, but both the position angle defined in section 3.1 and pointing angle that will be defined in section 3.4 are equivalent angles.

[2]For GI correlations this is usually defined as the shear $\gamma_{+,\times}$. For the rest of this work we will refer to the use of the ellipticities for intrinsic alignment measurements as a pure shape statistic rather than a shear statistic as in weak lensing.

$$\epsilon_\times = -\big(\epsilon_1 \sin 2\theta_p - \epsilon_2 \cos 2\theta_p\big) \tag{3.6}$$

In the next section we will use the tangential and cross ellipticities to calculate the most common statistic in the intrinsic alignment and weak lensing literature: the large-scale two-point correlation functions know as the galaxy-shear and shear-shear correlations. As mentioned, we will refer to these as the galaxy-shape and shape-shape correlations. Combined with the standard galaxy-galaxy correlation function, these make up what is known as the three-two-points (3-2PCF).

3.2 Ellipticity correlations

If our density field, δ and our ellipticities, ϵ, are Gaussian, then all cosmological information will be contained in the auto and cross correlations between positions and shapes of our halos/clusters. Thus, we can derive estimators of the two-point correlation functions between the positions of our tracers and their corresponding shapes. Using the ellipticities with respect to axes orientated tangentially(+) or at 45 degrees ($\times$) to a line joining a pair of objects we can define the ellipticity correlation function as:

$$\xi_\pm(r_p, \Pi) = \langle \epsilon_+ \epsilon_+ \rangle (r_p, \Pi) \pm \langle \epsilon_\times \epsilon_\times \rangle (r_p, \Pi) \tag{3.7}$$

It should be noted that in clustering studies, ξ is commonly used to describes the 3D correlation function while w_p describes the projected correlation function. Here we use the symbol ξ to describe projected quantities such as the ellipticities and will make it clear by writing ξ as a function of r_p and Π which are the projected sky-plane and LOS separations

respectively. Thus, the 3D redshift space separation between a pair of objects would be $s = \sqrt{r_p^2 + \Pi^2}$. In our analysis, r_p is chosen to be the projected separation of a pair of objects that do not include the axis RSD is applied.

The ellipticity correlations are pretty straight-forward to estimate from a survey or simulation box. In terms of the galaxy-galaxy correlation which considers objects linked together based on their separation in r_p and Π we can use the Landy-Szalay estimator discussed in chapter 1:

$$\hat{\xi}_{gg}(r_p, \Pi) = \frac{DD(r_p, \Pi) - 2DR(r_p, \Pi) + RR(r_p, \Pi)}{RR(r_p, \Pi)} \tag{3.8}$$

Where DD, DR, RR are the pair counts normalized by the total possible pairs of each data set. For example, given a catalogue D that contains n_D objects with a given position, if we make P_{DD} possible pairs in various bins of r_p and Π, then our calculation for $DD(r_p, \Pi)$ would be:

$$DD(r_p, \Pi) = \frac{P_{DD}(r_p, \Pi)}{n_D(n_D - 1)/2} \tag{3.9}$$

Next, we can adopt a version of this estimator for galaxy-shape correlations that modifies the Landy-Szalay estimator as described in Mandelbaum et al. (2010). This estimator makes use of equation 3.5 to see on average how many objects are either parallel or perpendicular with the vector joining it to another object in your data set. To do this we will define the following calculation for a pair of objects denoted by indices i and j:

$$S_+D = \sum_{i \neq j | r_p, \Pi} = \epsilon_+(i|j) \tag{3.10}$$

Simply put, S_+D is the sum of the tangential ellipticity over all pairs of objects with separations r_p and Π, where one object is in the shape sample and one object is in the position sample. The modified galaxy-shape correlation then reads:

$$\hat{\xi}_{g+}(r_p, \Pi) = \frac{S_+D - S_+R}{RR} \tag{3.11}$$

Similarly, we can define an estimator for the shape-shape correlation for both the tangential and cross ellipticities. These are:

$$\hat{\xi}_{++}(r_p, \Pi) = \frac{S_+S_+}{RR} \tag{3.12}$$

$$\hat{\xi}_{\times\times}(r_p, \Pi) = \frac{S_\times S_\times}{RR} \tag{3.13}$$

where,

$$S_+S_+ = \sum_{i \neq j | r_p, \Pi} = \epsilon_+(i|j)\epsilon_+(j|i) \tag{3.14}$$

and likewise for $S_\times S_\times$, where the cross ellipticities are used instead. Alternatively, the above correlation functions can use the position angle or spin parameters, however, a direct relation for the impact of intrinsic alignments requires knowledge on the shapes of our objects, and thus the ellipticities are most useful in our study.

3.3 Projected correlation functions

As with our previous analysis in chapter 2, we wish to find a way to measure our statistic with minimal influence of RSDs. To do this in the previous chapter we considered the sky-plane (recall this is the xy-plane in our simulation) to select objects based on their orientations. This had allowed us to minimize the effects of RSDs from our selection. In terms of our ellipticity correlations, we can collapse this statistic along our axis where RSD is applied (i.e the Π axis) to obtain the projected correlation function, $w(r_p)$. The projected correlation function is unaffected by RSD since it will only depend on r_p which is in-fact the xy-plane where we will measure ϵ_1 and ϵ_2. To do this, an integrator is created to collapse our matrix of values stored in Π bins for various r_p bins[3]. For some pair $a, b \in \{\delta, g, +, \times\}$, the corresponding projected correlation function, $w_{ab}(r_p)$ is given by integrating the correlation function as follows,

$$w_{ab}(r_p) = \int_{-\Pi_{\max}}^{+\Pi_{\max}} \xi_{ab}(r_p, \Pi) d\Pi \tag{3.15}$$

where we have established that Π is the distance between objects along the LOS. A detailed physical understanding of our projected correlation function for our shape and position catalogues requires a comparison with theoretical alignment models, such as the ones detailed in Kiessling et al. (2015) and Samuroff et al. (2020). These are,

$$w_{gg}(r_p) = b_g^2 \int J_0(kr_p) P_\delta(k) \frac{kdk}{2\pi} \tag{3.16}$$

[3]see Appendix H for source code of `correlations` function.

$$w_{g+}(r_p) = -b_g \int J_2(kr_p) P_{GI}(k) \frac{kdk}{2\pi} \tag{3.17}$$

$$w_{++}(r_p) = \int \left[J_0(kr_p) + J_4(kr_p) \right] P_{II}(k) \frac{kdk}{2\pi} \tag{3.18}$$

where $J_n(kr_p)$ is the n^{th} order Bessel function of the first kind and b_g is the linear galaxy bias. In these integrals, we define the large-scale galaxy-shape and shape-shape power spectrum's from Hirata and Seljak (2004) which have the form,

$$P_{GI}(k) = -\frac{C_1 \bar{\rho}(z) a^2(z)}{D(z)} P_\delta(k) \tag{3.19}$$

$$P_{II}(k) = \left(\frac{C_1 \bar{\rho}(z) a^2(z)}{D(z)} \right)^2 P_\delta(k) \tag{3.20}$$

where C_1, is a normalization constant, $\bar{\rho}(z), a(z), D(z)$, are the mean density in the Universe, scale factor and growth function at some redshift z respectively, and $P_\delta(k)$ is the linear matter power spectrum. We follow many previous analyses in fixing C_1 to the value from Brown et al. (2002) and parameterizing deviations in strength of alignment from this baseline, such that we have a free amplitude A_1 when numerically computing the integrals where $P_{GI}(k) \to A_1 P_{GI}$ and $P_{II} \to A_1^2 P_{II}$.

Figure 3.2 is our measured projected correlation functions using our estimators described by equation 3.8, 3.11 and 3.12 with our theory curves over-plotted. The presence of non-zero galaxy-shape and shape-shape correlation functions over all scales indicates the presence of both small and large scale alignments as a result of the tidal fields. It should be noted that the large scale tidal fields will orient and stretch massive objects certain ways which will then dictate the orientation and alignment of objects on smaller scales, such as

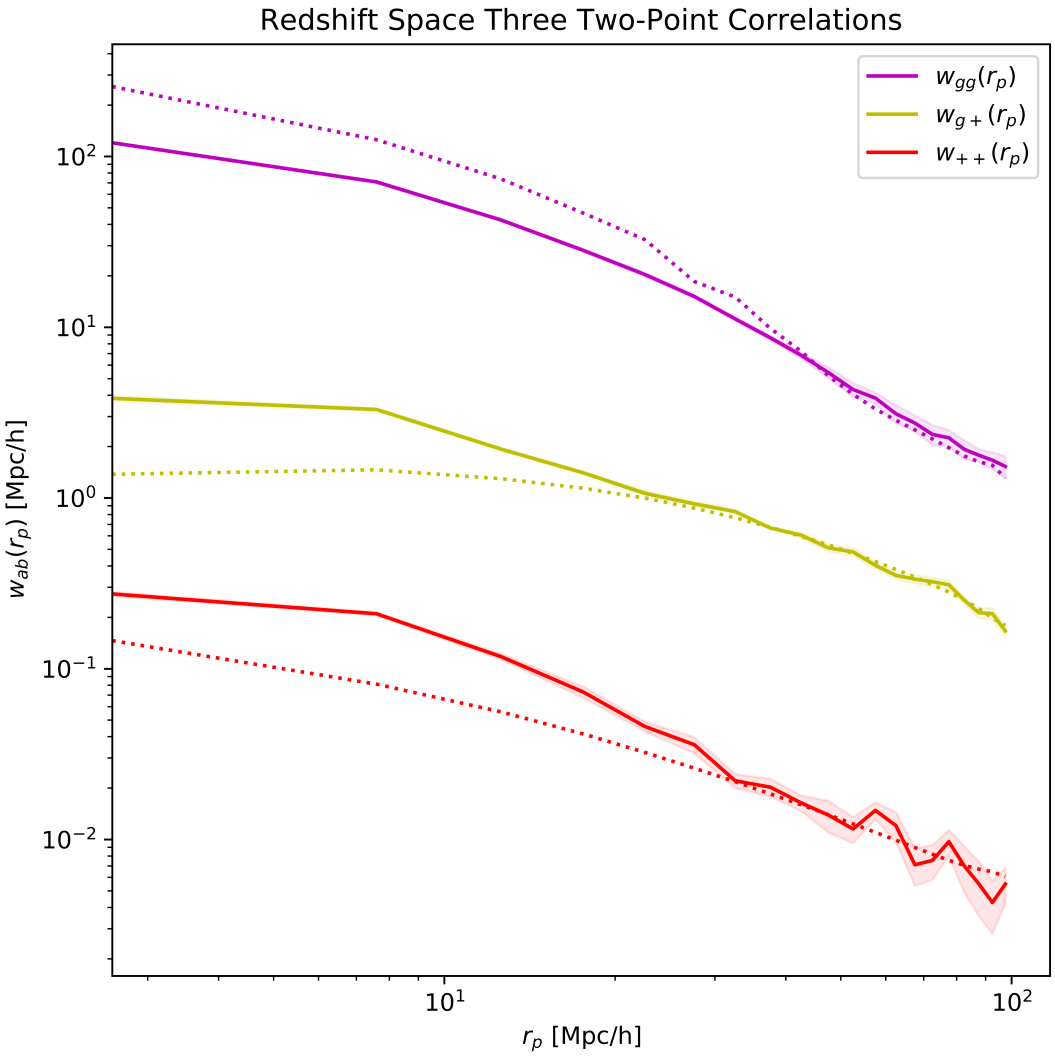

Figure 3.2: Our three 2-point projected correlation functions (solid lines). Shaded areas represent 1σ uncertainties, computed as the error on the mean of our simulations. Dashed lines of similar colour to the solid lines are the associated linear alignment theory curves (Note that our fits are only seen for scales $> 20-30\ h^{-1}$Mpc). Non-linear alignment theory can predict the shape of our projected correlations for small scales.

the halos/clusters residing in filaments. Thus, we can say that our large scale alignment amplitude A presented in equation 2.8 in chapter 2 is related to the constant A_1 used in equations 3.17 and 3.18 and holds information about our alignment signal (this is presented in more detail in Samuroff et al. 2020). In the next section we will look at an alternative statistic that is easier to use when measuring the alignments of massive/compact objects based on tidal interactions within a local environment.

3.4 The alignment functions

As discussed, structures in the Universe are predicted to have shapes that are aligned in specific ways with each other due to tidal forces and physical processes like the coherent infall of matter along filaments for instance. These alignments can be manifested in various ways such as the relative orientation and alignment of pairs of galaxies or cluster shapes with one another or the alignment of their shapes with the largest cluster in its local environment (usually chosen to be the Brightest Cluster Galaxy (BCG) in a survey). As mentioned, most of the intrinsic alignment studies have put a focus on galaxies mainly because of its effects on weak lensing measurements, Hirata and Seljak (2004). In this section we once again consider a statistic that is discussed in terms of intrinsic alignments in weak lensing literature. We will measure the large-scale intrinsic alignments of halos in our simulations and search for two types of alignments discussed in Hopkins et al. (2005) using pairs of halos. The first statistics we consider is the alignment between the projected major axes of a pair of halos, commonly referred to as the correlation alignment. The second is the alignment between one halos major axis and the radial vector connecting it

to the other halo in its pair, referred to as the pointing alignment.

To start, for each identified halo in our simulation we can find pairs of separation $r_p < 100\ h^{-1}$Mpc. To group pairs separated by their projected distance r_p we must also consider 'redshift' bins along the LOS where RSD is applied. This will be a maximum separation in what we have called the Π direction for this analysis. From Smargon et al. (2012) we can consider redshift bins of $\Delta z \leq 0.015$ that are used. In the adopted ΛCDM cosmology for our simulations, this works out to be about $\Pi \leq 100\ h^{-1}$Mpc. We then compute the two statistics mentioned where the pointing angle, θ_p, indicates whether halos tend to point towards one another and the correlation angle, θ_c, which tells us about the relative alignment of cluster shapes of the pair. Ultimately, we will compute the quantity $\langle \cos^2 \theta_n \rangle$ for $n \in \{c, p\}$ in each r_p bin[4]. This is the dot product of the vector that points along the largest axis of our halo with either another halos largest axis vector, $\langle \cos^2 \theta_c \rangle$, or the vector joining it to another halo, $\langle \cos^2 \theta_p \rangle$. These two statistics are referred to as the *alignment functions* in the literature.

In order to identify halo shapes and the stretching axis vector as a result of the tidal fields we once again use the inertia tensor in our simulation given by equation 3.2. We then solve our characteristic equation: $\det(I_{ij} - \lambda \mathcal{I})$ where $\mathcal{I}$ is the identity matrix to find the eigenvalues λ_i and corresponding eigenvectors, $\vec{v_i}$. These eigenvalues are sorted such that $\lambda_1 > \lambda_2 > \lambda_3$ where our largest eigenvalue λ_1 corresponds to the largest length of our halos axes. Thus, the normalized eigenvector $\vec{v_1}$ associated with eigenvalue λ_1 is the stretching or long axis of our inertia tensor.

[4]In terms of projected shapes, $\pm cos(\theta_n)$ are physically identical, since there is no preferred 'positive' direction for the unit vectors used. Thus, $\langle \cos^2 \theta_n \rangle$ is a good statistic to include the full range of values without boosting our amplitude.

In figure 3.3 we show our measurements of both alignment functions $\langle \cos^2 \theta_n \rangle$. For our projected case, this statistic would be equal to $\langle \cos^2 \theta_n \rangle = 0.5$ for a purely random distribution of cluster shapes. Thus, the error bars on each point is estimated by assigning all halos a random orientation angle and redoing the measurement with the random angles (i.e. lack of any physical alignments). As mentioned, the random signal is a known value for both the pointing and correlation angles and so deviations from 0.5 in our random orientation angle measurements can be used to quantify the noise in our measurement of each function. We notice that for separations $r_p < 30\ h^{-1}$Mpc we have a strong indication that alignments are not random for both statistics. As we move to larger and larger separations we tend to the mean signal for randomly orientated objects where information about the environment for a pair of halos becomes less correlated. Also, equivalent to Splinter et al. (1997), Hopkins et al. (2005) and Smargon et al. (2012) we find a stronger alignment signal for the pointing angle and that the pointing angle signal extends to larger scales. The features of these measurements indicate that tidal alignments are present, orientating our halos in specific ways, such as halos in filaments pointing towards over-densities and the tidal fields stretching axis influencing the alignments of our pairs.

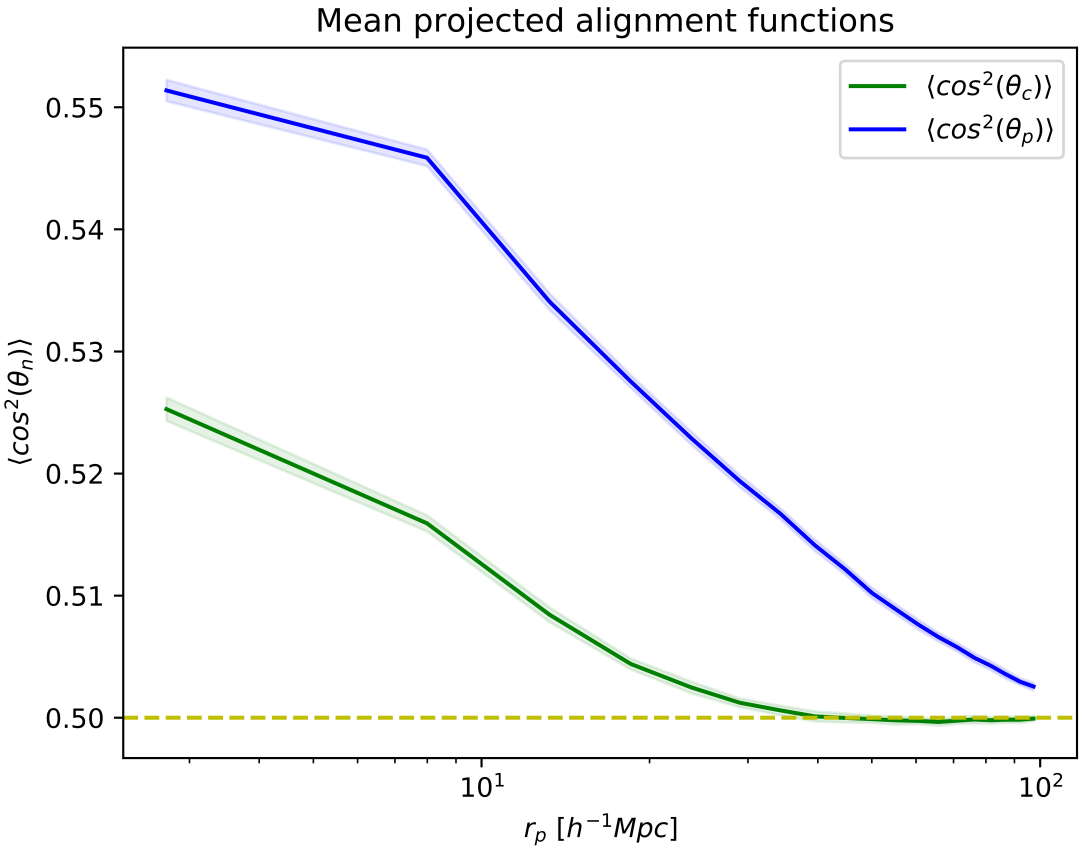

Figure 3.3: Correlation alignment(green) and pointing alignment (blue) averaged over 100 Quijote simulations for halos of mass $M_h > 6.57 \times 10^{13}\ h^{-1}M_\odot$. The shaded regions are the uncertainties for deviations from $\langle cos^2(\theta_n)\rangle$. We have plotted the 5σ uncertainty here since the width of our solid line is larger than the 1σ uncertainty. We have used $z = 0$ snapshots in this scenario and at late times we expect halos to be more spherical on average and to obtain the weakest signal for both alignment functions. As we move to early times in the Universe the halo ellipticities become more defined and our alignment function signal becomes stronger. The yellow dashed line represents the mean of $\langle cos^2(\theta_n)\rangle$ which would be the expected value of our calculation if we do not have the presence of alignments.

Chapter 4

On the Difficulty of Using Group Catalogues to Make Clustering Measurements

In this chapter, we will create various galaxy catalogues that mimic a BOSS galaxy survey (such as the one used in Reid et al. (2014)) in attempt to differentiate between pure RSD anisotropy and tidal alignment anisotropy based on selection, introduced in the previous chapters. To create these catalogues, we will use our Quijote ΛCDM simulations and then populate our dark matter groups with galaxies using a Halo Occupation Distribution (HOD). We will include information about the shapes of the parent dark matter halo, since, as mentioned in chapters 2 and 3, this internal property strongly correlates with the tidal field and can be used to infer information about how clusters are aligned.

These alignments can add an additional contamination to our selection and thus redshift space clustering measurements. In this chapter we wish to present a measurement of how alignments can produce an additional anisotropic effect in the multipoles and its dependence on selection method. In section 4.1 we will *briefly* present the simulation data used to set up the framework of our dark matter catalogues. In section 4.2 and 4.3 we discuss our population technique for three separate galaxy catalogues that will be used. We will describe a standard HOD analysis, a modified HOD that was used in order to recover information about the parent dark matter halo shapes, lost with a standard HOD, along with a control catalogue that serves to differentiate between pure RSD and RSD+IA anisotropies. In section 4.4 we present the algorithms that participated in both our real and redshift space group finding. In section 4.5 we use our statistics from chapters 2 and 3 to confirm the presence of alignments and shape information in our galaxy catalogue. Finally, in section 4.6 we combine our selection effects and alignments analysis in attempt to recover an anisotropic assembly bias signal for a survey-like galaxy catalogue.

4.1 Quijote simulations and dark matter halos

To investigate the properties of galaxy clusters, we use 100 Quijote N-body simulations to create the skeleton of a ΛCDM universe Villaescusa-Navarro et al. (2020) in a periodic box of length 1 h^{-1}Gpc. . These snapshots are similar to those used in previous chapters, however, the subset of simulations we use here are output snapshots at $z = 0.5$. The choice of this redshift is based on the subsequent choice of HOD parameters that follow Reid et al. (2014) (described in section 4.3). Each simulation contains 512^3 particles of mass

$M_p = 6.57 \times 10^{11}\ h^{-1} M_\odot$. All simulations were run using the following set of cosmological parameters: $\{\Omega_m, \Omega_b, \Omega_\Lambda, n_s, h = 0.3175, 0.049, 0.6825, 0.9624, 0.6711\}$. These parameters are in agreement with the constraints from the Planck satellite (Aghanim et al., 2020). We will then group dark matter particles in real space using an isotropic Friends-of-Friends (FoF) group finder with linking length parameter $b_{ll} = 0.2$. We consider a dark matter halo to be a group with more than 20 particles and thus set $nmin = 20$. These catalogues are saved and then used when populating galaxies into our halos. More details on further use of the grouping algorithms is discussed in 4.4.

4.2 Galaxy catalogues and the halo occupation distribution

Using our dark matter catalogues, we wish to create mock galaxy catalogues that resemble a galaxy survey. For this work we have decided to consider the analysis conducted in Reid et al. (2014) in order to populate our dark matter halos. Here the authors create a mock catalogue to mimic the CMASS sample from SDSS data release 10 (DR10) which covers galaxies in the redshift range $0.43 < z < 0.7$. For details on DR10 see Ahn et al. (2014) and for details of the sub-sample used see Anderson et al. (2014). Using our dark matter snapshots, we will populate our halos with galaxies based on the *Zheng07Model* from Zheng et al. (2007). This HOD will be our baseline used in this work where the amount of centrals and satellites populated within a dark matter halo is given by the

Parameter	Choice
$\log M_{min}$	13.031
$\sigma_{\log M}$	0.38
α	0.76
$\log M_o$	13.27
$\log M_1$	14.08

Table 4.1: List of the HOD parameters used in this work.

following numerical calculations:

$$N_{cen}(M) = \frac{1}{2}\left(1 + \text{erf}\left[\frac{\log M - \log M_{min}}{\sigma_{\log M}}\right]\right) \tag{4.1}$$

$$N_{sat}(M) = \left(\frac{M - M_o}{M_1}\right)^{\alpha} \tag{4.2}$$

Here, $\log M_{min}$ is the minimum mass required for a dark matter halo to host a central galaxy, $\sigma_{\log M}$ is the rate of transition from $N_{cen} = 0$ to $N_{cen} = 1$, α is the power law slope of the relation between the dark matter halo mass and the number of satellites, M_o is the low-mass cutoff for a dark matter halo of mass M to host a satellite and M_1 is the characteristic halo mass where the function N_{sat} begins to assume a power law. To mimic the SDSS DR10 data and the LowRes simulation used in Reid et al. (2014), we adopt similar parameters used for their HOD as outlined in table 4.1. For our Quijote simulations, this HOD population technique and parameter choice gives us a number density of approximately, $n = 3.3 \times 10^{-4}$ for each simulation. This is about a factor of 1.3 times smaller than what this HOD gives on the simulations we are comparing to which has a number density of $n_{\text{LowRes}} = 4.2 \times 10^{-4}$. From here on out we will refer to this populated galaxy catalogue as our *Zheng catalogue*. In general, this population technique assumes a spherical Navarro-

Frenk-White (NFW) profile (Navarro et al. 1997) in which a central galaxy is placed in the halo centre of mass position and satellites are randomly placed within some sphere of maximum radius, R_{vir}. We should note that from our lower mass limit of dark matter halos, some halos will only receive a central galaxy and thus it is not possible to retrieve information about the shape for such objects. Also, the clustering description of this HOD assumes the galaxy-halo connection depends only on the halo mass removing the 3D information of our halo and does not allow our tracer galaxies to give us information about the orientation of the host halo of a particular mass. We will discuss a modification to work around this in the next section, 4.3, along with a control catalogue that serves as a better comparison to the modifications we make. In general, the modifications will change the outputs of the *Zheng07Model* to include information of the shape of our parent dark matter halo that galaxies are populated in. Simply put, this standard Zheng07 HOD does not include information about intrinsic alignments and we wish to include this in our catalogue to see if we will introduce anisotropic assembly bias, similar to what was shown for dark matter halos in Obuljen et al. (2019).

4.3 Modified HOD and control catalogue

As mentioned, in our standard HOD analysis, information about the orientation and shape of our parent dark matter halo, that should be reflected in our galaxy clusters after population, disappears. In order to create a galaxy catalogue that includes this information we will modify the HOD output used by placing satellites in locations that trace out the 3D distribution of dark matter particles in the halo. Following the same

formalism as described in section 4.2, we populate centrals the same way but alter the location of our satellites. Instead of having our satellites randomly placed within a sphere of radius R_{vir} for each halo, we will move these satellites into randomly selected dark matter particle positions of that associated halo. Thus, when we populate our satellites, we essentially are sub-sampling the dark matter halos. We also make sure to include the dark matter particle velocities that are selected to become our satellites. Since the standard HOD randomly places satellites, the standard HODs velocities are thus also randomized, removing redshift space shape information. To verify that we have maintained majority of the information about the stretching axis of a halo with our new galaxy clusters, we will use the calculation of our stretching axis (i.e. largest eigenvector associated with largest eigenvalue of I_{ij}) of the parent dark matter halo with its newly populated galaxies. To compare, we then compute the dot product between these vectors. If our galaxy clusters maintain similar shape information then this calculation should be approximately 1 since the angle ϕ between these eigenvectors should be zero. Figure 4.1 shows the distribution of $\cos\phi$ for the case where we use the standard HOD cluster shapes and the new modified HOD cluster shapes. As expected, since the standard HOD removes information about the dark matter halo shapes, we have a flat, random distribution of galaxy cluster shapes compared to the parent dark matter halo shapes. For our modified HOD we see a skewed distribution towards $\cos\phi = 1$ were we have successfully recovered shape information for majority of our halos. Values of $\cos\phi \to 0$ are dominated by our 2-particle galaxy clusters since these will add a great deal of noise to the shape information as these only allow 1 satellite to be placed relative to the central with a very low probability of tracing out the correct parent dark matter halo shape. We find that 98% of clusters that give back

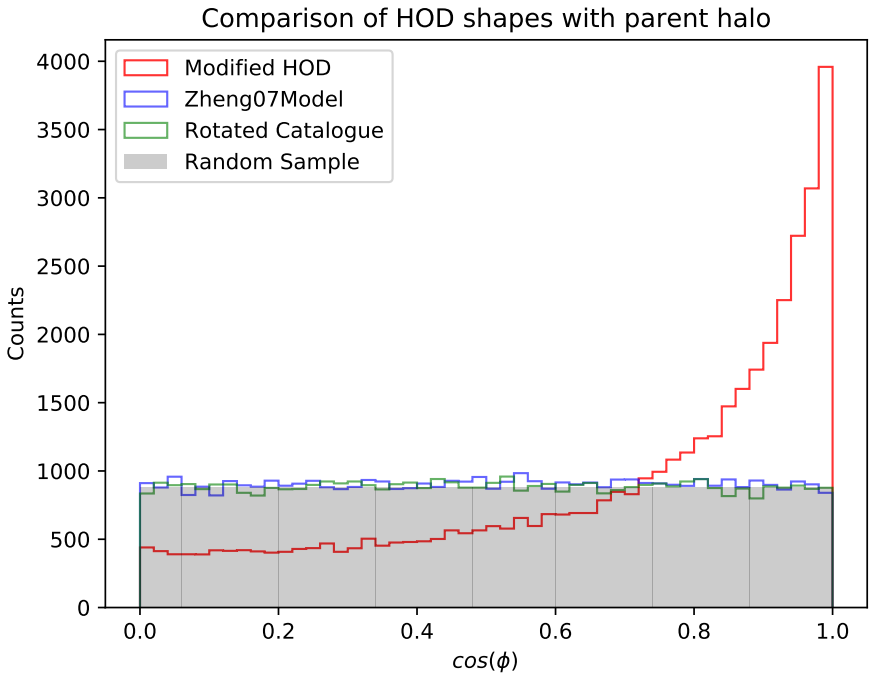

Figure 4.1: Histogram of the misalignment angle which is the dot product of parent dark matter halo shape with corresponding populated cluster shape (in terms of the stretching axis or longest axis eigenvector used to calculate shape correlations and alignment functions). The solid red line is for our modified HOD which is designed to give back similar eigenvectors used to compute alignments of dark matter halos. The shaded grey region is if we had a random uniform distribution of cluster shapes (i.e. the newly populated clusters do not represent the proper shape of the parent halo). Over-plotted is the Zheng07Model HOD output and my rotated catalogue galaxy cluster shapes that are random relative to their parent dark matter halo (purple and green lines respectively). Notice that the Zheng07Model or rotated catalogue do not give back information of shapes equivalent to their parent halo shapes or an alignment signal and thus follows the random distribution given by the shaded grey region.

values of $\cos\phi \leq 0.5$ are from 2-particle clusters. This number drops of exponentially for $\cos\phi \to 1$. It is true that this adds significant noise to the alignment signal, however, for many of the low mass halos, we still recovered some 2-particle clusters that ended up in our final bin of $0.98 < \cos\phi \leq 1.0$. In order to utilize the entire galaxy catalogue, we decide to keep these clusters and make our measurements. If we attempt to remove these 2-particle clusters and make similar measurements, this will remove 85% of all clusters in the sample. We will show in 4.5 that even if we include these noisy cluster shapes, we still recover a clear alignment signal in our galaxy catalogue. We will refer to this galaxy catalogue as our *Modified catalogue*.

Our third and final catalogue to be used will be a control catalogue that utilizes our modified HOD technique but removes any alignment signal. In order to do this, after our modified population technique, we rotate the satellites in each halo in order to randomize our alignments. The rotation is achieved using a randomized orthogonal rotation matrix that follows a Haar distribution (Stewart 1980). This is the only known uniform distribution for a n-dimensional special orthogonal group[1]. We will also make sure to rotate the satellite velocities by the same random matrix used to rotate its positions. Thus, once this 3×3 rotation matrix is applied to each cluster, this catalogue will only contain RSD effects and be equivalent to the standard HOD technique. However, this catalogue makes for a better comparison to our modified HOD in the analysis since it adopts the same population technique and then removes the presence of intrinsic alignments as described. It also maintains the same radial separations between satellites and the central galaxy

[1]The **scipy.stats** python package contains a function **special_ortho_group()** that can calculate this rotation matrix for us.

Group Finder	Linking length choice
$\mathrm{FoF}_{\text{zheng, mod, rot}}$	$b_{ll} = 0.215$
$\mathrm{CGM}_{\text{zheng, mod, rot}}$	$r_\perp = 0.1, r_\parallel = 0.513$

Table 4.2: The different galaxy group catalogues used and their corresponding linking lengths. The subscripts 'zheng', 'mod' and 'rot' represent the standard Zheng HOD output, the modified HOD output and the rotated shape catalogues respectively.

after rotation in real space. We will refer to this catalogue in our analysis as our *Rotation catalogue*. Appendix E gives a brief visualization of how this rotation is applied in creating our control catalogue.

4.4 Group finding algorithms

In order to analyze large cosmological simulations, we adopt the hierarchical Friends-of-Friends (FoF) algorithm that is publicly available with the Python package *nbodykit* adapting techniques from Efstathiou et al. (1985). We have created our group catalogue in real space utilizing the snapshots described in section 4.1. We consider groups with at least 20 particles ($M_h > 1.314 \times 10^{13}\ h^{-1}M_\odot$) selected with an isotropic linking length, $b_{ll} = 0.2$. This results in grouping $> 300,000$ dark matter halos per simulation. Once these dark matter halos have been found in real space we then proceed to populate our halos with galaxies using our standard HOD. Using our populated galaxy catalogues, we then move our galaxies to redshift space by the transformation: $\boldsymbol{s} = \boldsymbol{r} + \frac{\boldsymbol{v}\cdot\hat{z}}{aH}$. We then proceed to group our galaxies in each catalogue considering both an isotropic FoF and a Cylindrical Grouping Method (CGM) algorithm. The names of the different group catalogues and the linking lengths used on the galaxy catalogues is outlined in table 4.2.

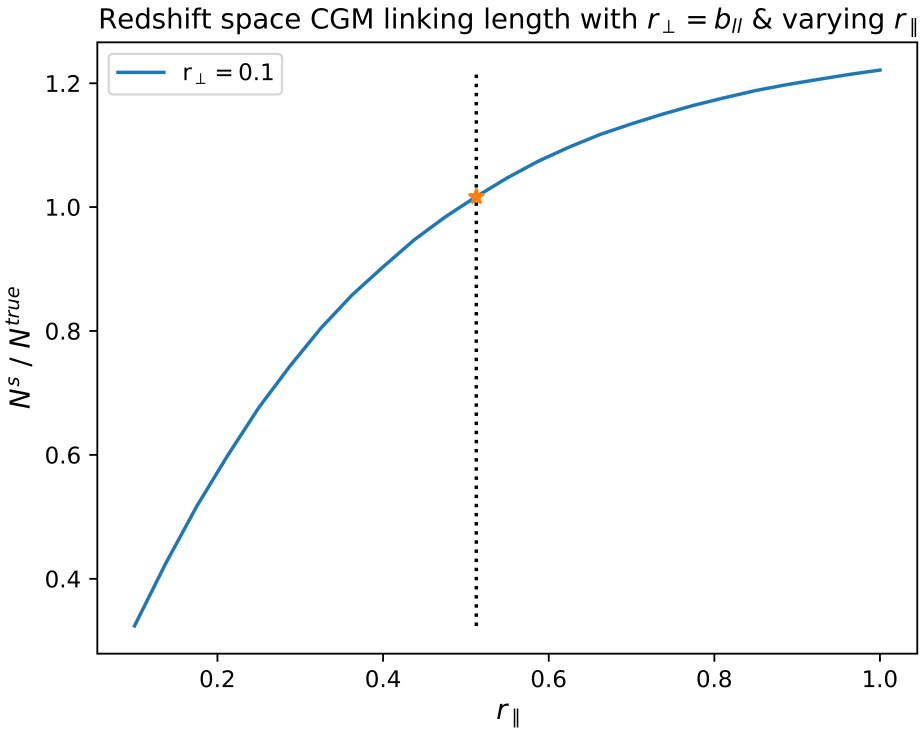

Figure 4.2: Keeping $r_\perp$ fix at a value if 0.1 we adjust $r_\parallel$ in order to recover an equivalent amount of redshift space clusters as the ones that exist when we initially populated galaxies into our groups.

We arrive at the choice of and isotropic FoF linking length of 0.215. This choice is based on a few criteria. First, in real space, b_{ll} will no longer be the most ideal linking length since we have greatly sub-sampled our dark matter catalogue and the mean particle separation between our populated galaxies and the original dark matter particles has changed substantially. The second is that we have then moved particles into redshift space where the apparent field is non-physical and so we must consider a wider range of values. Thus, by altering our value of b_{ll} we can attempt to group together galaxies that were

populated into a specific dark matter halo by extending the linking length a bit[2]. For our CGM linking lengths we consider an optimal redshift space grouping algorithm that keeps the transverse linking length the same as our standard FoF in real space (i.e. found to be $r_{\perp} = b_{ll} = 0.1$ as discussed in appendix D for our galaxies) and then change the LOS linking length ($r_{\parallel}$) to recover an identical number of real space groups in redshift space. Figure 4.2 shows the ratio of real space groups recovered in redshift space and the corresponding choice of $r_{\parallel}$. We found the most optimal choice based on this criteria to be $r_{parallel} = 0.513$.

4.5 Identifying alignments in galaxy catalogues

In order to confirm that we have indeed included the proper cluster shapes and halo alignments in our galaxy catalogue, we can use the alignment functions described in chapter 3. A good comparison for our case is Smargon et al. (2012) where the alignment functions were used on two pre-existing galaxy survey catalogues, a maxBCG catalogue from SDSS photometric imaging data (Koester et al. 2007) and an adaptive matched filter catalogue that identifies galaxy over-densities (Dong et al. 2008). Figures 1 and 2 in Smargon et al. (2012) show a comparison of their correlation and pointing alignment signal relative to a ΛCDM simulation (see Appendix F for plots).

In figures 4.3 and 4.4 we try to depict the same sort of plot where we have calculated our galaxy alignment functions and include the original dark matter halo alignment signal recovered in chapter 3 using our ΛCDM simulations. As we see, the alignment function

[2]See appendix D for further discussion on FoF linking length on our galaxy catalogue

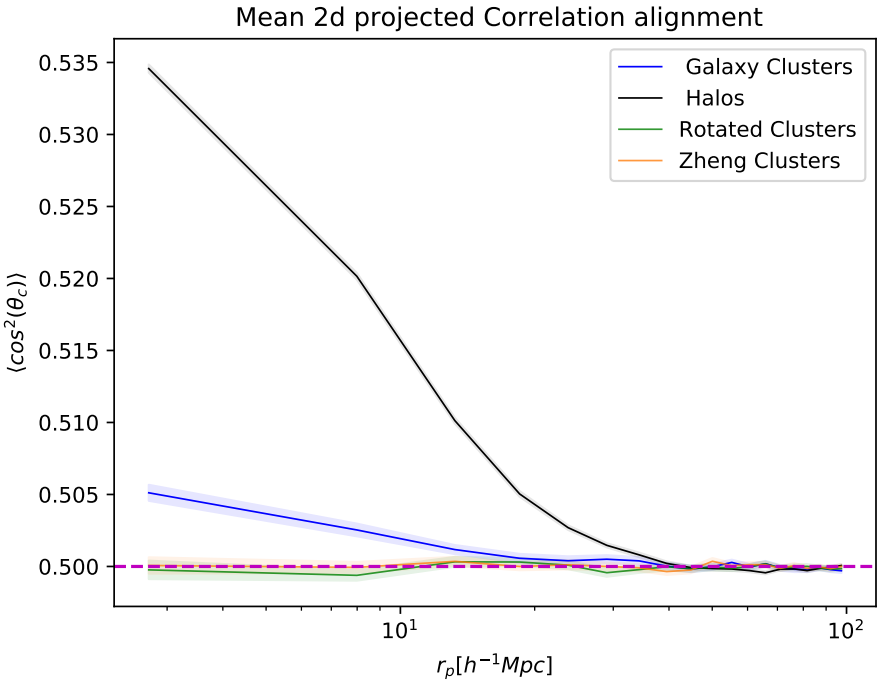

Figure 4.3: Similar to figure 3.3 but for the correlation angle alignment of galaxy clusters (blue) compared to dark matter halos (black). Also included are the Zheng catalogue and control catalogue in orange and green respectively. These catalogues lack any IA signal. After making use of our modified population technique we see that we have a clear signal for galaxy cluster alignments that follows a similar trend as to the dark matter halos. Here, our galaxy cluster correlation alignment signal is suppressed, most likely due to the noise our 2-particle clusters adds to our sample. Shaded errors are measured using the uncertainty on 100 simulations. For a comparison to Smargon et al. (2012) see figure F.1 in Appendix F.

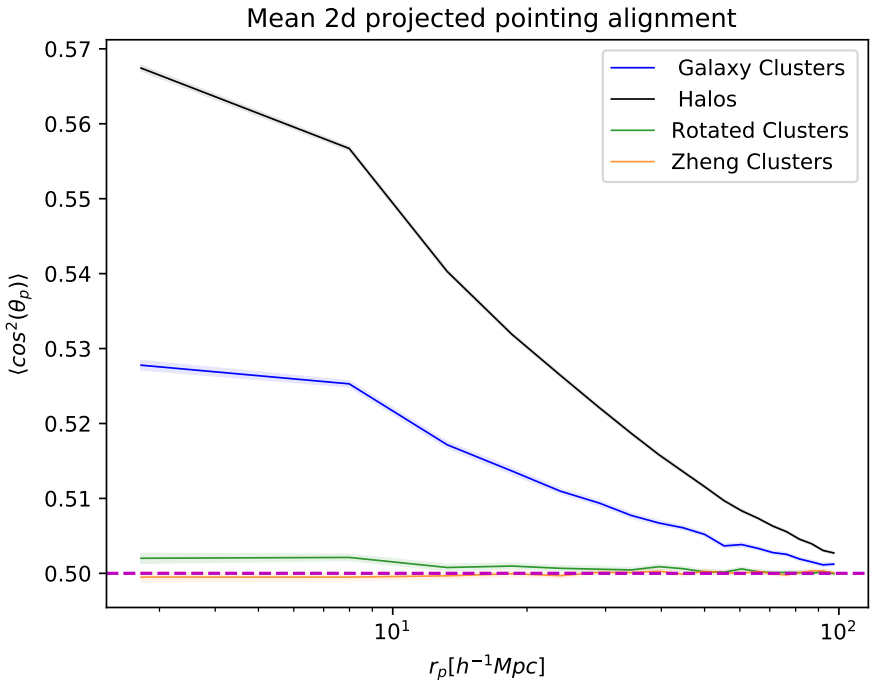

Figure 4.4: Similar to figure 3.3 but for the pointing angle alignment of galaxy clusters (blue) compared to dark matter halos (black). The Zheng and control catalogue is also included in orange and green respectively. Once again we have a clear signal for galaxy cluster alignments that is suppressed compared to the halos but present for all scales. Can also compare with figure F.2 in Appendix F.

signals become suppressed since we have sub-sampled our dark matter halos and added noise to our shape calculations. This is dominated by our 2-particle clusters as mentioned in section 4.3, and is particularly weak for the correlation alignment. From Smargon et al. (2012) this same trend is noticeable as a result of the smaller amount of cluster pairs identified in the survey compared to a simulation.

From this, we can confirm we are clearly recovering a signal for the pointing alignment on all scales along with a correlation alignment as a consequence of galaxy cluster alignments not being completely random. As mentioned, this also includes the shot-noise from the 2-particle clusters that do not perfectly trace the proper parent halo shapes.

4.6 Large-scale anisotropy: RSD + alignments

In chapter 2, we have that shown that if a particular selection of halos in real and redshift space couple to certain internal halo properties, such as the inertia tensor, one can recover an additional anisotropy in the power spectrum (seen in the quadrupoles of figure 2.4 and figure 2.9). These internal properties themselves correlate with the large-scale tidal fields (Hirata 2009, Martens et al. 2018, Obuljen et al. 2019) and selections independent of these properties, will not result in this additional anisotropy and be dominated by RSD selection effects. Thus, if our selection method is more likely to pick up halos orientated a certain way then we will introduce an anisotropic assembly bias in our measurements Obuljen et al. (2019). We have also shown in figure 2.8 that once we move into redshift space and make a selection, we inherently are making an anisotropic selection since objects will preferentially align along our LOS where RSD effects are present (for reference, figure

3 in Obuljen et al. (2019) gives back similar information where if we find groups in redshift space and move back to real space, the resulting quadrupole will be different than the real space power and Kaiser prediction due to this anisotropic selection[3]). This is because once the halos (or galaxies) are systematically moved from real to redshift space, performing our group finding using an isotropic FoF for instance in redshift space will produce biased groups. Thus, we create a catalogue of objects found in redshift space that are preferentially orientated a particular way. In this section we plan to utilize all three of our galaxy catalogues created (as discussed in section 4.3) in an attempt to differentiate between the anisotropies caused by RSD selection effects with various group finders compared to RSD and intrinsic alignment selection effects with the same group finders. Therefore, regardless of the selection effects that may be caused by RSDs, we plan on extracting the contamination of our groups based solely on the inclusion of intrinsic alignments.

To make a comparison between each catalogue, we will use our *Rotation Catalogue* which serves as our control group for our galaxy catalogues. This control group should only include RSD effects and removes any intrinsic alignments present and thus should not include any additional selection effects based on these alignments. First, to make sure that our control group is not changing our power spectrum measurements prior to any group finding and selection, we will measure the multipoles and compare to our modified catalogue. This will allow us to confirm that both catalogues contain the same RSD effects and that we do not inherently add in some discrepancy that is selection-independent between them. Figure 4.5 shows that we have identical monopole and quadrupole measurements for each catalogue and thus can now proceed with our selections on each.

[3]See Appendix G

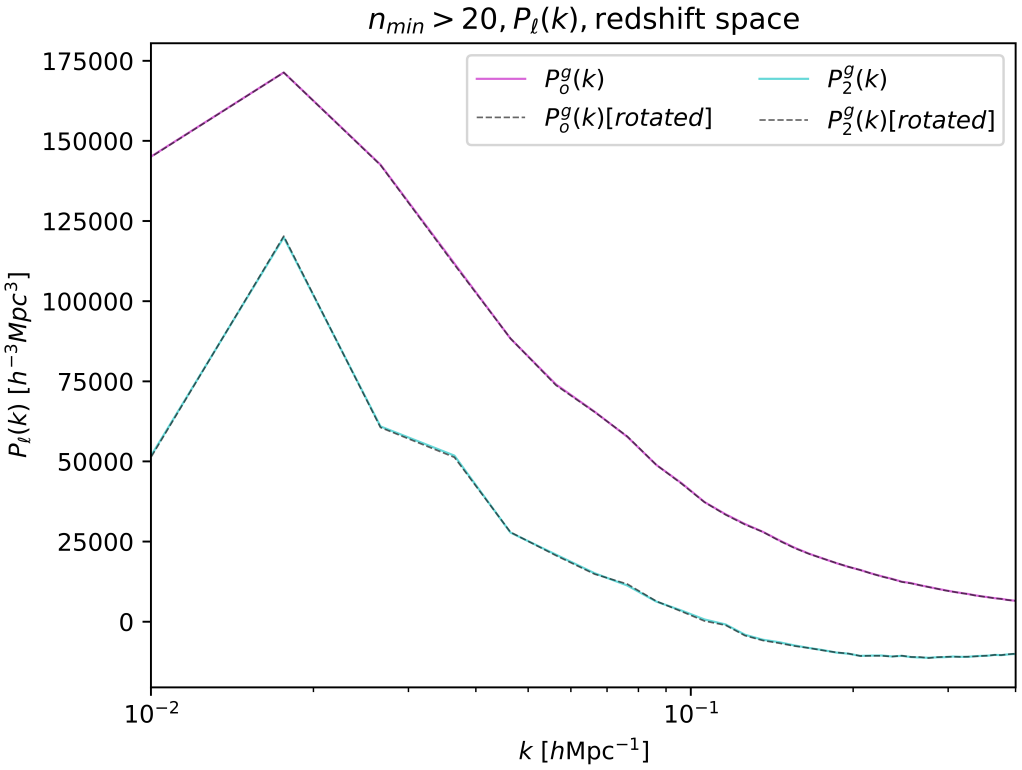

Figure 4.5: The monopole and quadrupole for our baseline model before and after rotation on a log-linear scale. The monopole (quadrupole) before rotation is given by the magenta (cyan) lines. The monopole and quadrupole after rotation are over-plotted given by the black dashed lines. Clusters are from populated halos that had more than 20 dark matter particles.

To analyze the difference between our catalogues and selection method adopted, we will plot the ratio,

$$\mathcal{R}_{l,j} = \frac{\frac{1}{n}\sum_i^n \left(P^s_{l,i,j}(k) - P^s_{l,i,\text{control}}(k)\right)}{\langle P^s_{l,\text{control}}(k)\rangle} \quad \text{for } l \ \in \{0,2\} \tag{4.3}$$

where n is the amount of simulations used, $P^s_{l,i,\text{control}}(k)$ is the redshift space multipole for the i^{th} simulation of the rotation catalogue and $\langle P^s_{l,\text{control}}(k)\rangle$ is the average, smoothed multipole for the rotation catalogue. Here, $P^s_{l,i,j}(k)$ is the i^{th} simulation measurement on the j^{th} catalogue where $j \in \{\text{FoF}_{\text{mod}}, \text{FoF}_{\text{Zheng}}, \text{CGM}_{\text{mod}}, \text{CGM}_{\text{Zheng}}\}$. Equation 4.3 is a way of determining the size of the effect our parameter A can have on the selection of groups from equation 2.8. For all three catalogues we will use both our isotropic FoF and CGM group finder. Therefore, we will have a total 6 group catalogues and 4 lines per plot (2 galaxy catalogues will be compared relative to our control $\times$ 2 different selection methods for each). In figure 4.6 we show these ratios for each catalogue for both the monopole and quadrupole (for reference to the actual individual multipole outputs for all catalogues, see appendix J.). Since we do not expect much difference in the monopoles, information regarding an additional anisotropic signal will be seen in the quadrupole plot, at the bottom of figure 4.6. However, from our quadrupole measurements, we can see that we lack a strong and noticeable difference between the control catalogue and my modified galaxy catalogue that includes intrinsic alignments (red dashed and solid blue lines). We believe this can be a result of the already suppressed intrinsic alignment signal being washed out on large-scales by RSDs for such small group sizes. As mentioned, we are dominated by 2-particle clusters that add noise to our shape statistics and may not know

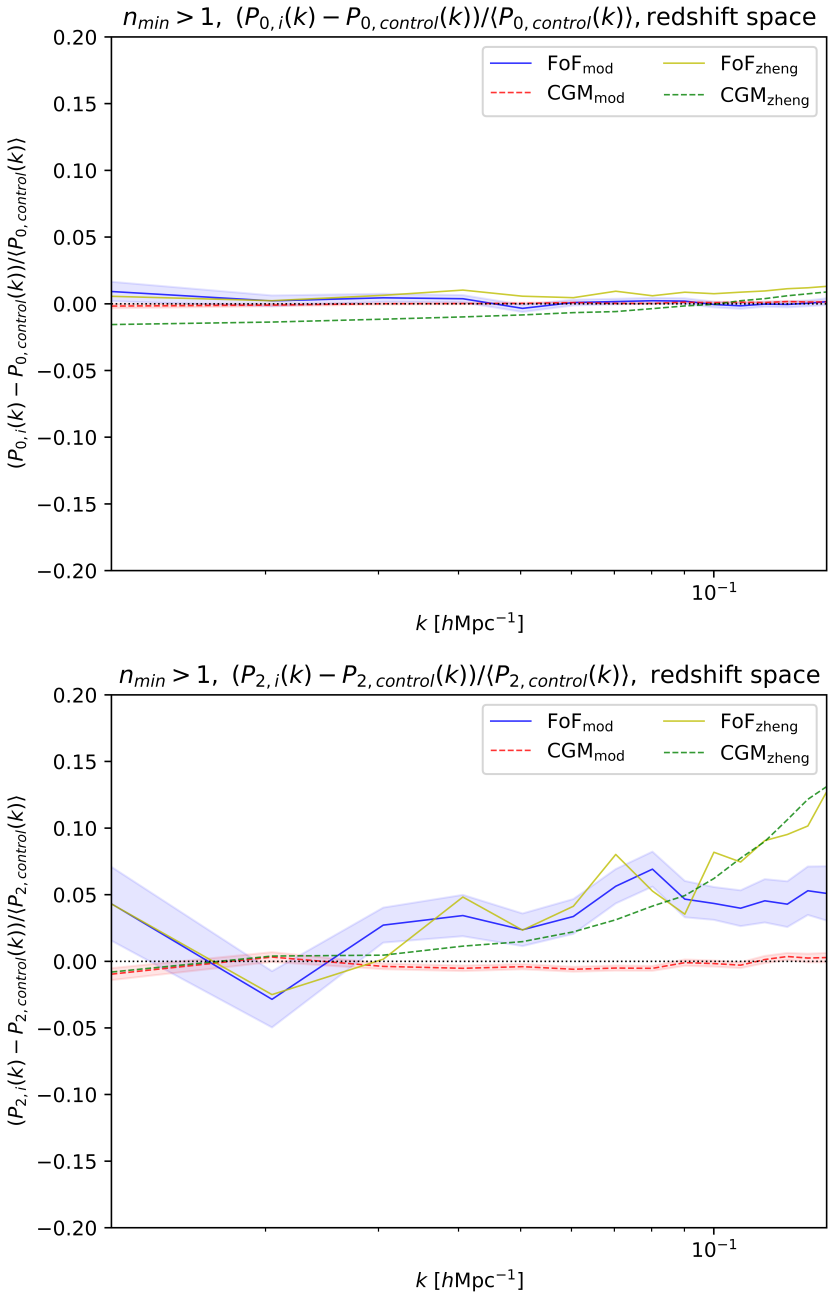

Figure 4.6: The monopole fractions (*top*) and quadrupole fractions (*bottom*) for our modified and Zheng07 catalogues relative to the control catalogue. Uncertainty for groups found with CGM and FoF in the case of our modified catalogue comparison is calculated as the standard deviation of the mean for equation 4.3 over 100 simulations.

about the proper halo shapes in real space, let alone in redshift space. Since we modify the shapes of our clusters in real space and then move these clusters into redshift space, a cluster with 2-particles will distort the alignment and become preferentially aligned along the LOS in our plane-parallel approximation.

For our FoF selection, we see very minor differences between the control and modified catalogue (solid blue line) in our quadrupole which can be a result of a combination of a weak intrinsic alignment signal and sub-optimal selection method to extract this potential systematic effect in redshift space. We also notice that the Zheng07 catalogue comparison using this selection (solid yellow line) tells us that this difference is also present for a catalogue lacking intrinsic alignments. As discussed in chapter 2, we find that an isotropic FoF gives us an anisotropic selection along our LOS in redshift space. This group finder will mainly group the most compact 'massive' objects in redshift space that exist in large over-densities relative to CGM. If we do not change the linking length, this would result in only approximately 25% of the more massive clusters being grouped. For our case, the small change to the linking length allows us to recover more clusters, however, we must be obtaining similar objects, whether it be the modified catalogue, Zheng07 catalogue or control catalogue, dominated by RSD selection effects. For the CGM finder, which does a better job at recovering clusters in redshift space, and is more commonly used on galaxy catalogues, there also is a lack of a strong signal for differences between the quadrupoles on large-scales as seen by the dashed red line. We note that at small scales for the Zheng07 catalogue with CGM, given by the dashed green line, we begin to diverge from the control and this is believed to be a result of changing the sizes of our halos with our new population technique. Ultimately, this signal is so faint for each group finder that it currently has very

little consequence on our measurements.

At some point the intrinsic alignment signal has to be present as we are essentially sub-sampling the halos. It is clear there is a strong signal in the halos as seen in Obuljen et al. (2019) and for preferential selection as seen from our plots in chapters 2. For our sub-sampled galaxy catalogue measurements, the weak signal may be a result of an insufficient selection method along with our catalogue being dominated by noise. For our modification technique, we are dominated by two noise sources. The first is the fact that we drastically sub-sampled the halos and have many clusters with particles that do not know about the halo orientation. Also, we have the probability of 'chance pairs' because of such a small minimum clusters size (i.e $nmin > 1$ in our selection for clusters). Therefore, we may have non-real groups from halos that only received 1 galaxy for instance and were then grouped together in redshift space. If we compare to our halos where we use much larger values for $nmin$, we have a greater probability of obtaining real groups that are composed of 100 or more particles rather than 2 or more. Secondly, we are utilizing an HOD in which the choice of parameters are set to match the CMASS SDSS DR10 sample. Future galaxy samples that are deeper than CMASS will pick up fainter satellite galaxies, altering the number density and thus parameter choices for the HOD, better reflecting the number of objects typically in galaxy cluster. Also, for our low-resolution simulations we generally only obtain massive halos and so for deeper surveys a simulation with better resolution can create a mock galaxy catalogue utilizing a smaller M_1 in terms of the HOD population and thus we can have more objects in our halos to better trace the dark matter. In doing so these selection effects may become more prominent and of interest in contaminating our groups. Therefore, it is clear that in order to obtain a clear signal for anisotropic

assembly bias in a mock galaxy catalogue, more care in our population and group sizes could be important, along with our selection and cuts to our sample in order to extract such a signal.

Chapter 5

Summary and Discussion

Throughout this work we have studied the significance of group finders and selection effects when making clustering measurements for LSS. In chapter 1 we introduced the theory behind RSD anisotropy known as the Kaiser effect along with primary bias, b_g. In chapter 2 we explored some background theory on secondary bias and clustering along with the relation of tidal field alignments and selection dependant contamination effects. This selection effect can cause additional anisotropies in the power spectrum as seen in the quadrupole and is referred to as anisotropic assembly bias, Obuljen et al. (2019). In chapter 3 we explored the relation between large-scale tidal fields and the alignments of halos utilizing some weak lensing statistics. We observe that the large-scale alignments and orientations that are present within an over or under-dense region can dictate the alignments of individual halos or clusters. Utilizing our simulations, we detect the presence of various intrinsic alignments using an adopted form for the projected shape correlations along with the alignment functions. In chapter 4 we created a mock galaxy catalogue

that includes information about halo shapes and orientations in order to include intrinsic alignments in our sample (not typically present in standard HOD analyses). We compare this catalogue with a control catalogue that only contains RSD effects along with a standard and commonly used Zheng07 HOD Model that lacks intrinsic alignments. We believe that intrinsic alignments play a role in the selection of clusters when using group finders and results in additional anisotropic effects as described by the tidal alignment model in Hirata (2009). We found that extracting this systematic effect was not so trivial. Our attempt was to separate selection effects between a catalogue that only contains RSD effects along our LOS and a catalogue that includes both RSD and intrinsic alignment selection effects, the latter being a more realistic and physical description of a survey. Ultimately, each catalogue contained similar selection effects dominated by RSDs that inherently align objects with our LOS. However, as initially shown in Obuljen et al. (2019), RSD selection effects can also provide us with an additional anisotropic signal and thus each catalogue has the presence of contamination in our selection of groups, albeit similar, and lacking a strong indication of intrinsic alignment contamination. This makes the picture of using these groups to make clustering measurements rather difficult as one would need to separate the presence of such a selection effect. Thus, care must be taken when using group catalogues where our selection is coupled to specific internal halo or cluster properties, as these may add additional anisotropies, especially for more massive objects. There exists room for improvement and advancement in this work in terms of addressing some of the noise sources and limitations encountered to extract the intrinsic alignment effect. In general, selection on halo size could lead to intrinsic alignments altering the anisotropic signal and should be of interest in future, deeper surveys, like Euclid (Scaramella et al. 2021) and DESI (Levi

et al. 2019) when using group catalogues where this systematic error could become more significant.

References

N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, and et al. Planck 2018 results. *Astronomy Astrophysics*, 641:A6, Sep 2020. ISSN 1432-0746. doi: 10.1051/0004-6361/201833910. URL http://dx.doi.org/10.1051/0004-6361/201833910.

C. P. Ahn, R. Alexandroff, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, Aubourg, S. Bailey, F. A. Bastien, and et al. The tenth data release of the sloan digital sky survey: First spectroscopic data from the sdss-iii apache point observatory galactic evolution experiment, Mar 2014. ISSN 1538-4365. URL http://dx.doi.org/10.1088/0067-0049/211/2/17.

L. Anderson, Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, and et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data releases 10 and 11 galaxy samples, Apr 2014. ISSN 0035-8711. URL http://dx.doi.org/10.1093/mnras/stu523.

J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields, May 1986.

P. Bett, V. Eke, C. S. Frenk, A. Jenkins, J. Helly, and J. Navarro. The spin and shape of dark matter haloes in the millennium simulation of a cold dark matter universe, Mar 2007. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2007.11432.x.

M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, A. Almeida, J. Alonso-García, F. Anders, S. F. Anderson, B. Andrews, and et al. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. , 154:28, July 2017. doi: 10.3847/1538-3881/aa7567.

J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web, Apr 1996. ISSN 1476-4687. URL http://dx.doi.org/10.1038/380603a0.

M. Borzyszkowski, A. D. Ludlow, and C. Porciani. The formation of cold dark matter haloes – ii. collapse time and tides, Nov 2014. ISSN 0035-8711. URL http://dx.doi.org/10.1093/mnras/stu2033.

M. L. Brown, A. N. Taylor, N. C. Hambly, and S. Dye. Measurement of intrinsic alignments in galaxy ellipticities, Jul 2002. ISSN 1365-2966. URL http://dx.doi.org/10.1046/j.1365-8711.2002.05354.x.

S. Cole and N. Kaiser. Biased clustering in the cold dark matter cosmogony., Apr. 1989.

D. Collaboration, T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila, D. Bacon, E. Baxter, K. Bechtol, M. R. Becker,

G. M. Bernstein, S. Bhargava, S. Birrer, J. Blazek, A. Brandao-Souza, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, H. Camacho, A. Campos, A. C. Rosell, M. C. Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, A. Chen, R. Chen, A. Choi, C. Conselice, J. Cordero, M. Costanzi, M. Crocce, L. N. da Costa, M. E. da Silva Pereira, C. Davis, T. M. Davis, J. D. Vicente, J. DeRose, S. Desai, E. D. Valentino, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, C. Doux, A. Drlica-Wagner, K. Eckert, T. F. Eifler, F. Elsner, J. Elvin-Poole, S. Everett, A. E. Evrard, X. Fang, A. Farahi, E. Fernandez, I. Ferrero, A. Ferté, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, G. Giannini, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, I. Harrison, W. G. Hartley, K. Herner, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, E. M. Huff, D. Huterer, B. Jain, D. J. James, M. Jarvis, N. Jeffrey, T. Jeltema, A. Kovacs, E. Krause, R. Kron, K. Kuehn, N. Kuropatkin, O. Lahav, P. F. Leget, P. Lemos, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, J. L. Marshall, P. Martini, J. McCullough, P. Melchior, J. Mena-Fernández, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, J. Muir, J. Myles, S. Nadathur, A. Navarro-Alsina, R. C. Nichol, R. L. C. Ogando, Y. Omori, A. Palmese, S. Pandey, Y. Park, F. Paz-Chinchón, D. Petravick, A. Pieres, A. A. P. Malagón, A. Porredon, J. Prat, M. Raveri, M. Rodriguez-Monroy, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. S. Rykoff, S. Samuroff, C. Sánchez, E. Sanchez, J. Sanchez, D. S. Cid, V. Scarpine, M. Schubnell, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, T. Shin, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, M. Tabbutt, G. Tarle, D. Thomas, C. To, A. Troja, M. A. Troxel, D. L. Tucker, I. Tutusaus, T. N. Varga, A. R. Walker,

N. Weaverdyck, J. Weller, B. Yanny, B. Yin, Y. Zhang, and J. Zuntz. Dark energy survey year 3 results: Cosmological constraints from galaxy clustering and weak lensing, 2021.

N. Dalal, O. Doré, D. Huterer, and A. Shirokov. Imprints of primordial non-gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects, Jun 2008. ISSN 1550-2368. URL http://dx.doi.org/10.1103/PhysRevD.77.123514.

V. Desjacques. Environmental dependence in the ellipsoidal collapse model, Aug 2008. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2008.13420.x.

V. Desjacques, D. Jeong, and F. Schmidt. Large-scale galaxy bias, Feb 2018. ISSN 0370-1573. URL http://dx.doi.org/10.1016/j.physrep.2017.12.002.

S. Dodelson. Modern cosmology, 2003.

F. Dong, E. Pierpaoli, J. E. Gunn, and R. H. Wechsler. Optical Cluster Finding with an Adaptive Matched-Filter Technique: Algorithm and Comparison with Simulations, Apr. 2008.

A. Dressler. Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies., Mar. 1980.

M. J. Drinkwater, R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, K. Glazebrook, R. Sharp, M. B. Pracy, S. Brough, M. Colless, and et al. The wigglez dark energy survey: survey design and first data release, Jan 2010. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2009.15754.x.

G. Efstathiou, M. Davis, S. D. M. White, and C. S. Frenk. Numerical techniques for large cosmological N-body simulations, Feb. 1985.

A. Faltenbacher and S. D. M. White. Assembly bias and the dynamical structure of dark matter halos, Dec 2009. ISSN 1538-4357. URL http://dx.doi.org/10.1088/0004-637X/708/1/469.

L. Gao, V. Springel, and S. D. M. White. The age dependence of halo clustering, Oct 2005. ISSN 1745-3933. URL http://dx.doi.org/10.1111/j.1745-3933.2005.00084.x.

E. Garaldi, E. Romano-Díaz, M. Borzyszkowski, and C. Porciani. Zomg - iii. the effect of halo assembly on the satellite population, Sep 2017. ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stx2489.

O. Hahn, C. Porciani, C. M. Carollo, and A. Dekel. Properties of dark matter haloes in clusters, filaments, sheets and voids, Feb 2007. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2006.11318.x.

O. Hahn, C. Porciani, A. Dekel, and C. M. Carollo. Tidal effects and the environment dependence of halo assembly, Oct 2009. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2009.15271.x.

N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian. nbodykit: An open-source, massively parallel toolkit for large-scale structure, Sep 2018. ISSN 1538-3881. URL http://dx.doi.org/10.3847/1538-3881/aadae0.

C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van

Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy, Sept. 2020. URL https://doi.org/10.1038/s41586-020-2649-2.

C. M. Hirata. Tidal alignments as a contaminant of redshift space distortions, Oct 2009. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2009.15353.x.

C. M. Hirata and Seljak. Intrinsic alignment-lensing interference as a contaminant of cosmic shear, Sep 2004. URL https://link.aps.org/doi/10.1103/PhysRevD.70.063526.

P. F. Hopkins, N. A. Bahcall, and P. Bode. Cluster alignments and ellipticities in cdm cosmology. *The Astrophysical Journal*, 618(1):1–15, Jan 2005. ISSN 1538-4357. doi: 10.1086/425993. URL http://dx.doi.org/10.1086/425993.

E. P. Hubble. Extragalactic nebulae., Dec. 1926.

J. D. Hunter. Matplotlib: A 2d graphics environment, 2007.

J. C. Jackson. A critique of Rees's theory of primordial gravitational radiation, Jan. 1972.

A. Jenkins, C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson. Evolution of Structure in Cold Dark Matter Universes. , 499(1):20–40, May 1998. doi: 10.1086/305615.

B. Joachimi, M. Cacciato, T. D. Kitching, A. Leonard, R. Mandelbaum, B. M. Schäfer,

C. Sifón, H. Hoekstra, A. Kiessling, D. Kirk, and et al. Galaxy alignments: An overview, Jul 2015. ISSN 1572-9672. URL http://dx.doi.org/10.1007/s11214-015-0177-4.

N. Kaiser. On the spatial correlations of Abell clusters. , 284:L9–L12, Sept. 1984. doi: 10.1086/184341.

N. Kaiser. Clustering in real space and in redshift space, July 1987.

M. Kerscher, I. Szapudi, and A. S. Szalay. A comparison of estimators for the two-point correlation function. *The Astrophysical Journal*, 535(1):L13–L16, May 2000. ISSN 0004-637X. doi: 10.1086/312702. URL http://dx.doi.org/10.1086/312702.

A. Kiessling, M. Cacciato, B. Joachimi, D. Kirk, T. D. Kitching, A. Leonard, R. Mandelbaum, B. M. Schäfer, C. Sifón, M. L. Brown, and et al. Galaxy alignments: Theory, modelling simulations, Sep 2015. ISSN 1572-9672. URL http://dx.doi.org/10.1007/s11214-015-0203-6.

A. A. Klypin, S. Trujillo-Gomez, and J. Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation, Oct. 2011.

A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, S. Colombi, J. Diemand, and et al. Haloes gone mad: The halo-finder comparison project, Jun 2011. ISSN 0035-8711. URL http://dx.doi.org/10.1111/j.1365-2966.2011.18858.x.

B. P. Koester, T. A. McKay, J. Annis, R. H. Wechsler, A. Evrard, L. Bleem, M. Becker, D. Johnston, E. Sheldon, R. Nichol, and et al. A maxbcg catalog of 13,823 galaxy

clusters from the sloan digital sky survey, May 2007. ISSN 1538-4357. URL http://dx.doi.org/10.1086/509599.

O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees. Dynamical effects of the cosmological constant., July 1991.

M. E. Levi, L. E. Allen, A. Raichoor, C. Baltay, S. BenZvi, F. Beutler, A. Bolton, F. J. Castander, C.-H. Chuang, A. Cooper, J.-G. Cuby, A. Dey, D. Eisenstein, X. Fan, B. Flaugher, C. Frenk, A. X. Gonzalez-Morales, O. Graur, J. Guy, S. Habib, K. Honscheid, S. Juneau, J.-P. Kneib, O. Lahav, D. Lang, A. Leauthaud, B. Lusso, A. de la Macorra, M. Manera, P. Martini, S. Mao, J. A. Newman, N. Palanque-Delabrouille, W. J. Percival, C. A. Prieto, C. M. Rockosi, V. Ruhlmann-Kleider, D. Schlegel, H.-J. Seo, Y.-S. Song, G. Tarle, R. Wechsler, D. Weinberg, C. Yeche, and Y. Zu. The dark energy spectroscopic instrument (desi), 2019.

Y. Li, H. J. Mo, and L. Gao. On halo formation times and assembly bias, Sep 2008. ISSN 1365-2966. URL http://dx.doi.org/10.1111/j.1365-2966.2008.13667.x.

R. Mandelbaum, C. Blake, S. Bridle, F. B. Abdalla, S. Brough, M. Colless, W. Couch, S. Croom, T. Davis, M. J. Drinkwater, and et al. The wigglez dark energy survey: direct constraints on blue galaxy intrinsic alignments at intermediate redshifts, Oct 2010. ISSN 0035-8711. URL http://dx.doi.org/10.1111/j.1365-2966.2010.17485.x.

Y.-Y. Mao, A. R. Zentner, and R. H. Wechsler. Beyond assembly bias: exploring secondary halo biases for cluster-size haloes, Dec 2017. ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stx3111.

D. Martens, C. M. Hirata, A. J. Ross, and X. Fang. A radial measurement of the galaxy tidal alignment magnitude with boss data, May 2018. ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/sty1100.

P. McDonald and U. Seljak. How to evade the sample variance limit on measurements of redshift-space distortions, Oct 2009. ISSN 1475-7516. URL http://dx.doi.org/10.1088/1475-7516/2009/10/007.

H. J. Mo and S. D. M. White. An analytic model for the spatial clustering of dark matter haloes. , 282(2):347–361, Sept. 1996. doi: 10.1093/mnras/282.2.347.

J. F. Navarro, C. S. Frenk, and S. D. M. White. A universal density profile from hierarchical clustering, Dec 1997. ISSN 1538-4357. URL http://dx.doi.org/10.1086/304888.

A. Obuljen, N. Dalal, and W. J. Percival. Anisotropic halo assembly bias and redshift-space distortions. *Journal of Cosmology and Astroparticle Physics*, 2019(10):020–020, Oct 2019. ISSN 1475-7516. doi: 10.1088/1475-7516/2019/10/020. URL http://dx.doi.org/10.1088/1475-7516/2019/10/020.

T. Okumura, C. Hikage, T. Totani, M. Tonegawa, H. Okada, K. Glazebrook, C. Blake, P. G. Ferreira, S. More, A. Taruya, and et al. The subaru fmos galaxy redshift survey (fastsound). iv. new constraint on gravity theory from redshift space distortions atz 1.4. *Publications of the Astronomical Society of Japan*, 68(3):38, Apr 2016. ISSN 2053-051X. doi: 10.1093/pasj/psw029. URL http://dx.doi.org/10.1093/pasj/psw029.

J. A. Peacock. Large-scale surveys and cosmic structure, 2003.

P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

W. J. Percival. Large scale structure observations, 2013.

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation, Feb. 1974.

S. Ramakrishnan, A. Paranjape, O. Hahn, and R. K. Sheth. Cosmic web anisotropy is the primary indicator of halo assembly bias, Aug 2019. ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stz2344.

B. A. Reid, H.-J. Seo, A. Leauthaud, J. L. Tinker, and M. White. A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of sdss-iii cmass galaxies, Aug 2014. ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stu1391.

K. Riebe et al. The MultiDark Database: Release of the Bolshoi and MultiDark Cosmological Simulations. 9 2011. doi: 10.1002/asna.201211900.

T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, and et al. Astropy: A community python package for astronomy, Sep 2013. ISSN 1432-0746. URL http://dx.doi.org/10.1051/0004-6361/201322068.

E. Romano-Díaz, E. Garaldi, M. Borzyszkowski, and C. Porciani. Zomg – ii. does the halo assembly history influence central galaxies and gas accretion?, Apr 2017. ISSN 1365-2966. URL http://dx.doi.org/10.1093/mnras/stx878.

S. Samuroff, R. Mandelbaum, and J. Blazek. Advances in constraining intrinsic alignment models with hydrodynamic simulations, 2020.

L. Samushia, W. J. Percival, and A. Raccanelli. Interpreting large-scale redshift-space distortion measurements. *Monthly Notices of the Royal Astronomical Society*, 420(3): 2102–2119, Jan 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.20169.x. URL http://dx.doi.org/10.1111/j.1365-2966.2011.20169.x.

W. L. W. Sargent and E. L. Turner. A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies., Feb. 1977.

R. Scaramella, J. Amiaux, Y. Mellier, C. Burigana, C. S. Carvalho, J. C. Cuillandre, A. D. Silva, A. Derosa, J. Dinis, E. Maiorano, M. Maris, I. Tereno, R. Laureijs, T. Boenke, G. Buenadicha, X. Dupac, L. M. G. Venancio, P. Gómez-Álvarez, J. Hoar, J. L. Alvarez, G. D. Racca, G. Saavedra-Criado, J. Schwartz, R. Vavrek, M. Schirmer, H. Aussel, R. Azzollini, V. F. Cardone, M. Cropper, A. Ealet, B. Garilli, W. Gillard, B. R. Granett, L. Guzzo, H. Hoekstra, K. Jahnke, T. Kitching, M. Meneghetti, L. Miller, R. Nakajima, S. M. Niemi, F. Pasian, W. J. Percival, M. Sauvage, M. Scodeggio, S. Wachter, A. Zacchei, N. Aghanim, A. Amara, T. Auphan, N. Auricchio, S. Awan, A. Balestra, R. Bender, C. Bodendorf, D. Bonino, E. Branchini, S. Brau-Nogue, M. Brescia, G. P. Candini, V. Capobianco, C. Carbone, R. G. Carlberg, J. Carretero, R. Casas, F. J. Castander, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, C. J. Conselice, L. Conversi, Y. Copin, L. Corcione, A. Costille, F. Courbin, H. Degaudenzi, M. Douspis, F. Dubath, C. A. J. Duncan, S. Dusini, S. Farrens, S. Ferriol, P. Fosalba, N. Fourmanoit, M. Frailis, E. Franceschi, P. Franzetti, M. Fumana, B. Gillis, C. Gio-

coli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, F. Hormuth, P. Hudelot, S. Kermiche, A. Kiessling, M. Kilbinger, R. Kohley, B. Kubik, M. Kümmel, M. Kunz, H. Kurki-Suonio, S. Ligori, P. B. Lilje, I. Lloro, O. Mansutti, O. Marggraf, K. Markovic, F. Marulli, R. Massey, S. Maurogordato, M. Melchior, E. Merlin, G. Meylan, J. J. Mohr, M. Moresco, B. Morin, L. Moscardini, E. Munari, R. C. Nichol, C. Padilla, S. Paltani, J. Peacock, K. Pedersen, V. Pettorino, S. Pires, M. Poncet, L. Popa, L. Pozzetti, F. Raison, R. Rebolo, J. Rhodes, H. W. Rix, M. Roncarelli, E. Rossetti, R. Saglia, P. Schneider, T. Schrabback, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, J. Skottfelt, L. Stanco, J. L. Starck, P. Tallada-Crespí, D. Tavagnacco, A. N. Taylor, H. I. Teplitz, R. Toledo-Moreo, F. Torradeflot, M. Trifoglio, E. A. Valentijn, L. Valenziano, G. A. V. Kleijn, Y. Wang, N. Welikala, J. Weller, M. Wetzstein, G. Zamorani, J. Zoubian, S. Andreon, M. Baldi, S. Bardelli, A. Boucaud, S. Camera, G. Fabbian, R. Farinelli, J. Graciá-Carpio, D. Maino, E. Medinaceli, S. Mei, C. Neissner, G. Polenta, A. Renzi, E. Romelli, C. Rosset, F. Sureau, M. Tenti, T. Vassallo, E. Zucca, C. Baccigalupi, A. Balaguera-Antolínez, P. Battaglia, A. Biviano, S. Borgani, E. Bozzo, R. Cabanac, A. Cappi, S. Casas, G. Castignani, C. Colodro-Conde, J. Coupon, H. M. Courtois, J. Cuby, S. de la Torre, S. Desai, D. D. Ferdinando, H. Dole, M. Fabricius, M. Farina, P. G. Ferreira, F. Finelli, P. Flose-Reimberg, S. Fotopoulou, S. Galeotta, K. Ganga, G. Gozaliasl, I. M. Hook, E. Keihanen, C. C. Kirkpatrick, P. Liebing, V. Lindholm, G. Mainetti, M. Martinelli, N. Martinet, M. Maturi, H. J. McCracken, R. B. Metcalf, G. Morgante, J. Nightingale, A. Nucita, L. Patrizii, D. Potter, G. Riccio, A. G. Sánchez, D. Sapone, J. A. Schewtschenko, M. Schultheis, V. Scottez, R. Teyssier, I. Tutusaus, J. Valiviita, M. Viel, W. Vriend, and L. Whittaker. Euclid preparation: I. the euclid

wide survey, 2021.

D. Schlegel, M. White, and D. Eisenstein. The baryon oscillation spectroscopic survey: Precision measurements of the absolute cosmic distance scale, 2009.

F. Schmidt, D. Jeong, and V. Desjacques. Peak-background split, renormalization, and galaxy clustering, Jul 2013. ISSN 1550-2368. URL http://dx.doi.org/10.1103/PhysRevD.88.023515.

M. Scodeggio, L. Guzzo, B. Garilli, B. R. Granett, M. Bolzonella, S. de la Torre, U. Abbas, C. Adami, S. Arnouts, D. Bottini, and et al. The vimos public extragalactic redshift survey (vipers), Jan 2018. ISSN 1432-0746. URL http://dx.doi.org/10.1051/0004-6361/201630114.

R. K. Sheth and G. Tormen. Large-scale bias and the peak background split, Sep 1999. ISSN 1365-2966. URL http://dx.doi.org/10.1046/j.1365-8711.1999.02692.x.

A. Smargon, R. Mandelbaum, N. Bahcall, and M. Niederste-Ostholt. Detection of intrinsic cluster alignments to 100 h-1 mpc in the sloan digital sky survey, 06 2012.

R. J. Splinter, A. L. Melott, A. M. Linn, C. Buck, and J. Tinker. The ellipticity and orientation of clusters of galaxies inn-body experiments, Apr 1997. ISSN 1538-4357. URL http://dx.doi.org/10.1086/303896.

G. W. Stewart. The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators, June 1980.

A. Taruya, T. Nishimichi, and S. Saito. Baryon acoustic oscillations in 2d: Modeling redshift-space power spectrum from perturbation theory, Sep 2010. ISSN 1550-2368. URL http://dx.doi.org/10.1103/PhysRevD.82.063522.

F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys, and et al. The quijote simulations, Aug 2020. ISSN 1538-4365. URL http://dx.doi.org/10.3847/1538-4365/ab9d82.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, 2020.

X. Wang, H. Wang, H. J. Mo, J. Shi, and Y. Jing. Secondary bias of dark matter halos: Perspective in the linear density field, 2021.

R. H. Wechsler and J. L. Tinker. The connection between galaxies and their dark matter halos, Sep 2018. ISSN 1545-4282. URL http://dx.doi.org/10.1146/annurev-astro-081817-051756.

R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel. Concentrations of dark halos from their assembly histories, Mar 2002. ISSN 1538-4357. URL http://dx.doi.org/10.1086/338765.

Z. Zheng, A. L. Coil, and I. Zehavi. Galaxy evolution from halo occupation distribution modeling of deep2 and sdss galaxy clustering, Oct 2007. ISSN 1538-4357. URL http://dx.doi.org/10.1086/521074.

APPENDICES

Appendix A

Full Derivation of the Redshift Space Density Field

Starting from the conservation relation, $\delta_m^s(\boldsymbol{s}) = J^{-1}\{1+\delta_m(\boldsymbol{r})\} - 1$, and taking the Fourier transform, we get:

$$\begin{aligned}
\delta_m^s(\boldsymbol{k}) &= \int \delta_m^s(\boldsymbol{s}) e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3s \\
&= \int \left(J^{-1}\{1+\delta_m(\boldsymbol{r})\} - 1\right) e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3s \\
&= \int \frac{d^3r}{d^3s}\{1+\delta_m(\boldsymbol{r})\} e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3s - \int e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3s \\
&= \int \{1+\delta_m(\boldsymbol{r})\} e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3r - \int e^{i\boldsymbol{k}\cdot\boldsymbol{s}} \frac{d^3r}{d^3r} d^3s \\
&= \int \{1+\delta_m(\boldsymbol{r})\} e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3r - \int \left\{1 + \frac{1}{aH}\frac{\partial v_z}{\partial z}\right\} e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3r \\
&= \int \delta_m(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3r - \int \frac{1}{aH}\frac{\partial v_z}{\partial z} e^{i\boldsymbol{k}\cdot\boldsymbol{s}} d^3r
\end{aligned} \tag{A.1}$$

From here, we will make the substitution, $\boldsymbol{s} = \boldsymbol{r} + \frac{\boldsymbol{v}\cdot\hat{z}}{aH}$, such that, $i\boldsymbol{k}\cdot\boldsymbol{s} = i\boldsymbol{k}\cdot\boldsymbol{r} + \frac{ikv_z\mu}{aH}$, where $\mu = \hat{k}\cdot\hat{z} = \cos\theta$. Thus,

$$\delta_m^s(\boldsymbol{k}) = \int \left\{ \delta_m(\boldsymbol{r}) - \frac{1}{aH}\frac{\partial v_z}{\partial z} \right\} e^{i\boldsymbol{k}\cdot\boldsymbol{r} + ik\mu v_z/(aH)} d^3r \tag{A.2}$$

This expression is exact under the distant observer and plane-parallel approximations. In the derivation of equation 1.16, the extra term in the exponent is dropped out at linear order to obtain our Kaiser formula. This exact expression is very difficult to evaluate and is sensitive to non-linear physics. There are works that have sought to evaluate the full form of this integral, such as the Taruya-Nishimichi-Saito (TNS) formula, Taruya et al. (2010).

Appendix B

Legendre Polynomials

Recall the compact expression for the Legendre polynomials is:

$$\mathcal{L}_n(\mu) = \frac{1}{2^n n!} \frac{d^n}{d\mu^n} \left(\mu^2 - 1\right)^n \tag{B.1}$$

Equation B.1 is known as the Rodrigues' formula. The first few Legendre polynomials are:

n	$\mathcal{L}_n(\mu)$
0	1
1	μ
2	$\frac{1}{2}\left(3\mu^2 - 1\right)$
3	$\frac{1}{2}\left(5\mu^3 - 3\mu\right)$
4	$\frac{1}{8}\left(35\mu^4 - 30\mu^2 + 3\right)$
5	$\frac{1}{8}\left(63\mu^5 - 70\mu^3 + 15\mu\right)$

Table B.1: The first 5 Legendre Polynomials

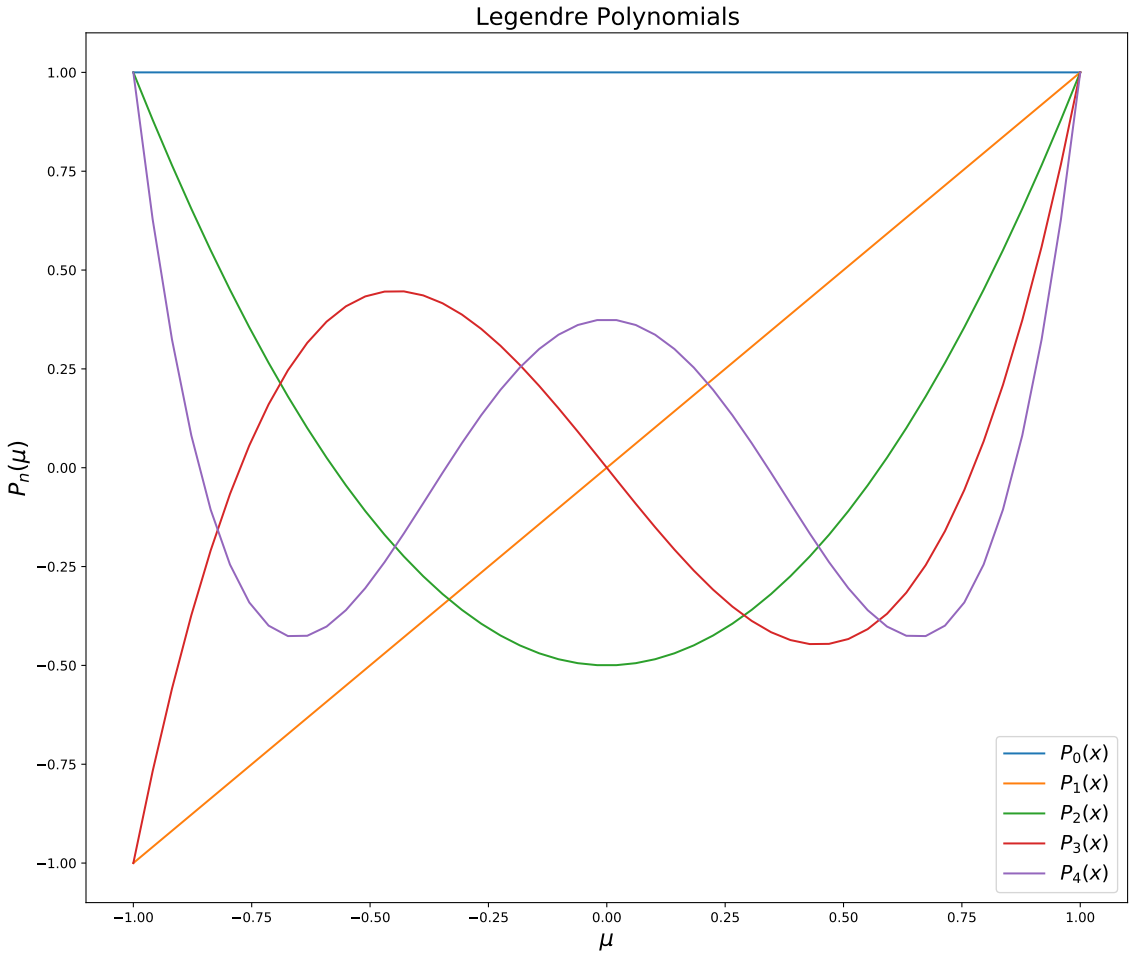

Figure B.1: Illustration of the first 5 Legendre polynomials. Curved produced using the python package **scipy.special.eval_legendre**

Appendix C

Peak-Background Split Model

In cosmological theories, we generally describe the dark matter distribution and not the galaxy distribution. However, what we observe from surveys is galaxies which we say are biased tracers of the underlining dark matter density field. The general picture of galaxy bias is that dark matter halos form at the peaks of the density field and then galaxies can form inside these dark matter halos (although not every halo necessarily will host a galaxy). However, not all peaks will form halos since the peak has to have a sufficiently high enough over-density. If the peak of the smoothed density field is greater than this threshold, a halo of mass M_h will form in this region. Figure C.1 gives an example where we consider the 2D smoothed matter density field and identify the peaks that exist above some threshold where dark matter halos form and thus where galaxies will form. Likewise, figure C.2 gives a 1D example of how regions that lie above a value of ν times the root-mean-squared of density fluctuations, $\sigma(M)$, will be clustered in specific (shaded) regions. Objects that form inside these peaks will give rise to a non-uniform spatial distribution that

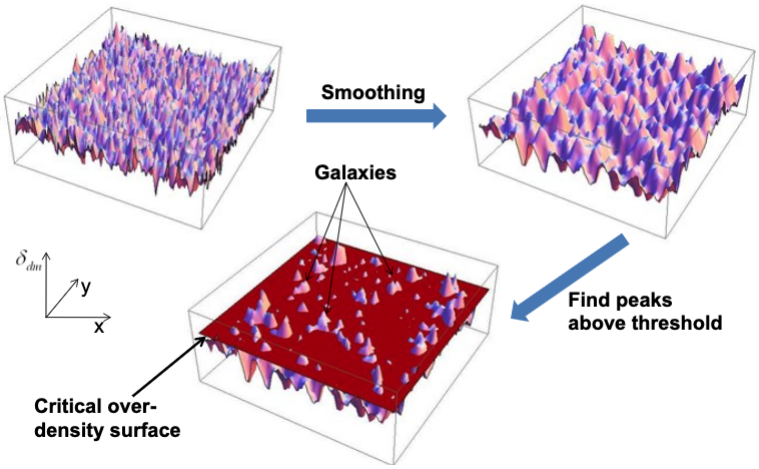

Figure C.1: A 2D example of the perturbations in the matter density field where the peak densities after smoothing will form halos that will then host galaxies. Here a halo of mass M_h is defined as a region in the space around the peak of the smoothed density field where its over-density δ is greater than some threshold δ_c, Desjacques et al. (2018).

traces specific regions of the matter density field (i.e. a bias). Although this description gives us a simple picture of where halos will form and thus a description of the observed biased galaxy distribution, this galaxy bias can be both a non-linear and non-local function. However, on large enough scales we are able to reduce this bias to a local, linear function via the *peak background split model*, Bardeen et al. (1986). The general idea of this model is to decompose the density field into peaks (fluctuations on small scales, δ_s) and background (large scale matter fluctuations, δ_l) as shown in figure C.3. According to this model, a central mechanism of biased clustering is related to rare high density fluctuations that will have collapsed *sooner* if it lies in a large scale over-dense region. This means that long wavelength modes can help enhance the abundance of massive objects that have collapsed in an over-density with respect to a mean mass of collapsed objects. As described above in the simple peaks model, let us first consider a density threshold to be some critical value δ_c

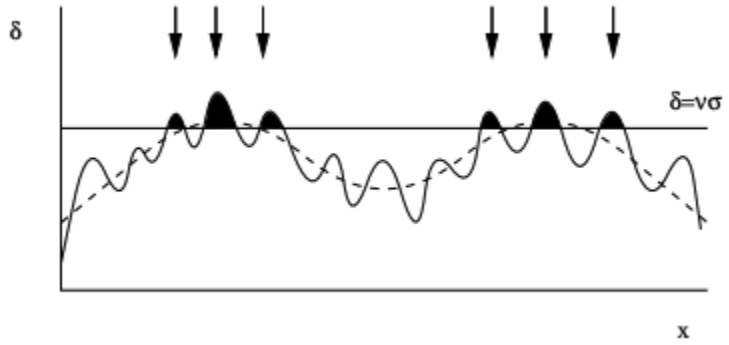

Figure C.2: 1D picture but with peaks shown as shaded regions that exist above some threshold. Notice in these shaded areas where halos are expected to form, the peaks are clustered together, Peacock (2003).

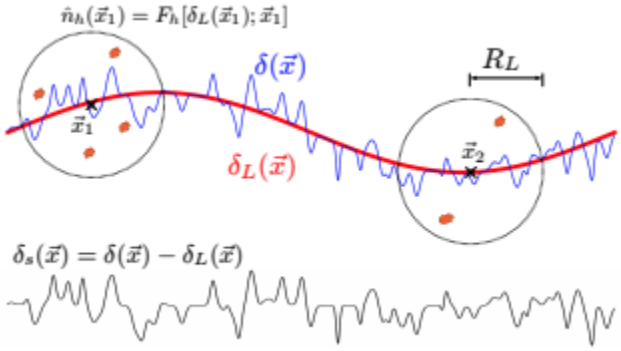

Figure C.3: An example of 1D perturbations in the density field (blue line) that are separated into large scale, δ_l (red line), and small scale, δ_s (black line), modes. The number of objects that form in the highlighted region depends on the function $F_{h,l}$ which encodes a departure from the deterministic number of 'tracer' objects that will form. Schmidt et al. (2013)

(generally, $\delta_c = 1.686$ in spherical collapse models). We can then assume a perturbed case where we add some large scale shift to the density perturbations, δ_l. This shift is meant to perturb the density threshold altering the local number density of halos $n(M)$, described by:

$$n \to n - \frac{dn}{d\delta_c}\delta_l \tag{C.1}$$

In this sense, a positive background, δ_l, will effectively reduce the threshold overdensity, δ_c, such that fluctuations only needed to achieve a value of $\delta_c - \delta_l$, allowing for more peaks to be pushed above this threshold, ultimately enhancing the number of objects that collapsed in that region. A key ingredient for this analysis is the mass-function of dark matter halos. Using the Press-Schechter formalism (see Press and Schechter 1974) we can describe the number density of halos in the mass range dM by:

$$n(M)dM = \sqrt{\frac{2}{\pi}}\frac{\rho_o}{M^2}\frac{\delta_c}{\sigma}\left|\frac{d\ln\sigma}{d\ln M}\right| e^{-\frac{\delta_c^2}{2\sigma^2}} dM \tag{C.2}$$

To first order, using the peak background split method, the long wavelength mode leads to a bias described by:

$$b = \frac{\delta_h}{\delta_l} = 1 - \frac{1}{n}\frac{dn}{d\delta_c} \tag{C.3}$$

Here we have used a boosted over-density for halos denoted as δ_h on large scales. Using the mass function described in equation C.2 we can arrive at a bias term:

$$b = \frac{\delta_c}{\sigma^2} - \frac{1}{\delta_c} = \frac{\nu - 1}{\delta_c} \tag{C.4}$$

Here we have used the substitution of $\nu = \frac{\delta_c}{\sigma}$(the same ν detailed in figure C.2). This is the general theory and formalism for calculating the linear galaxy bias. More accurate mass functions such as that of Sheth and Tormen (1999) have also been used in the literature to predict the value of b.

Appendix D

FoF for Galaxy Catalogue

In chapter 4 we discussed the use of an isotropic FoF to group together dark matter particles for our Quijote simulations. The standard choice of linking length parameter b_{ll} is usually set to be 0.2 times the mean particle separation. However, after populating our dark matter halos (that were found with this linking length) with galaxies, we have now changed the mean particle separation of our new galaxy catalogue that we will be using to once again apply our grouping algorithms to. Thus, taking into account the new positions of objects we arrive at a new linking length parameter to use of $b_{ll} = 0.1$ in real space. This choice is based on trying to match the amount of galaxy groups we find to the amount that are already present in the catalogue. Below in figure D.1 we plot the ratio of the groups identified by FoF for various choices of b_{ll} and the true groups that we know already existed after populating halos (i.e by use of the existing group ID's).

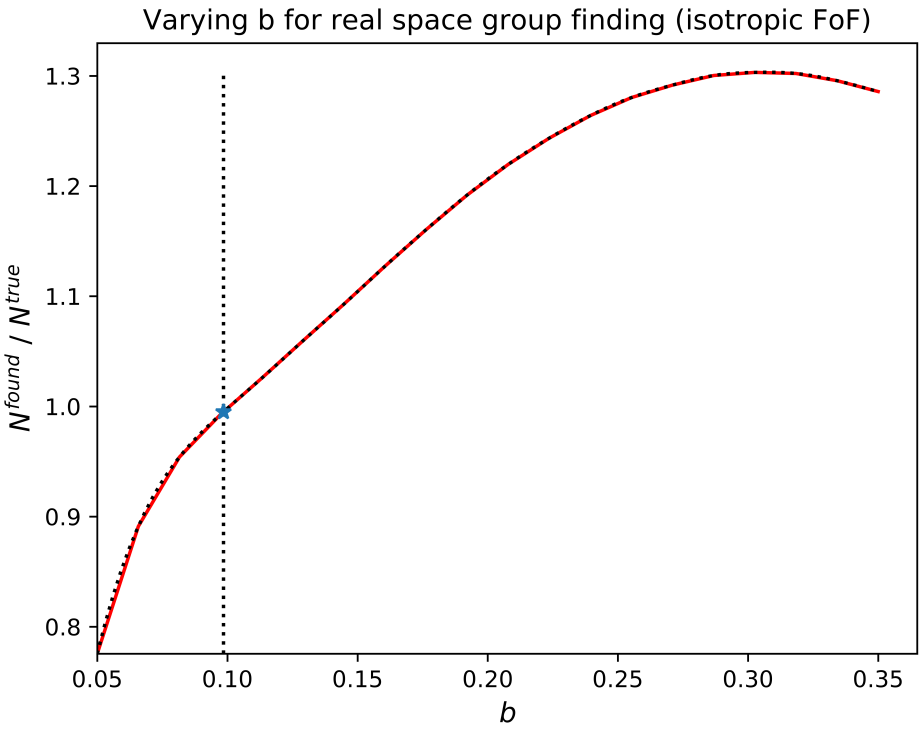

Figure D.1: The ratio of groups found to the true number of groups in our galaxy catalogue for various choices of b_{ll}. This takes into account the new mean particle separation in real space. The over-plotted dashed line that follows the solid red line is an attempted interpolation technique to better estimate this ratio. The star is our best choice of $b_{ll} = 0.099 \approx 0.1$.

Appendix E

Uniform Rotation of Galaxy Clusters

Consider a dark matter halo that has been populated with galaxies using our HOD population technique described in chapter 4. Our central galaxy position of a cluster will be given by the centre of mass position $\vec{r}_{com}$. If we consider a rotation matrix $\boldsymbol{M}$, then our rotation of the satellites about the centre of mass position (i.e. the central galaxy) will be:

$$\vec{r}_{rot} = \vec{r}_{com} + \boldsymbol{M}(\vec{r}_{sat,i} - \vec{r}_{com}) \tag{E.1}$$

where $\vec{r}_{sat,i}$ is the position of the i^{th} satellite of our cluster. Let us consider a randomly populated galaxy cluster where we will plot all satellite positions relative to the central position in units of h^{-1}Mpc as shown in the left of figure E.1. We will then apply our rotation matrix and compute equation E.1 to recover a randomly rotated cluster. This is shown to the right of figure E.1 where the orange points is our new rotated satellites and the blue points are the original satellite positions from the left plot. We can visualize this

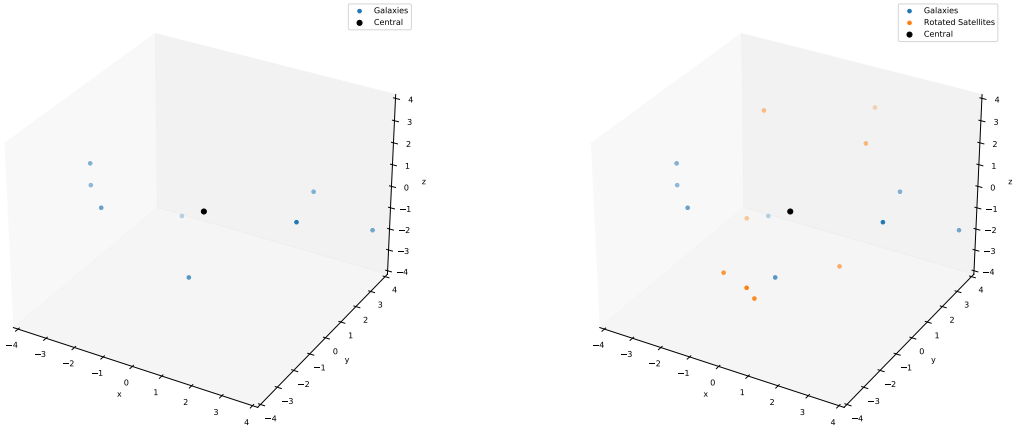

Figure E.1: Visual of central and satellite galaxies of a random cluster inside a cube. Positions of satellites are plotted relative to the centre of mass position (i.e. the central position) in h^{-1}Mpc.(*left*) Satellite positions before rotation in blue.(*right*) Satellite positions after rotation in orange (relative to original positions in blue).

uniform rotation by considering a central galaxy and one satellite galaxy. If we rotate the satellite N times by our matrix, then in 2D we should trace out a uniform circle about the central galaxy. Likewise, in 3D, this process should create a sphere of N points about our central. This is shown in figures E.2 and E.3.

Lastly, we must also apply this rotation matrix to the velocities since we have rotated the positions by this matrix. The matrix must be the same matrix we used to randomize the positions. In doing so, we must make sure the the x and y positions are not changing and only the z-component is affected by our velocity rotation. This is due to the fact that we consider a plane-parallel approximation for RSD in our simulations that only affects the z-component of our positions. To prove that we have not altered the xy-plane, we can

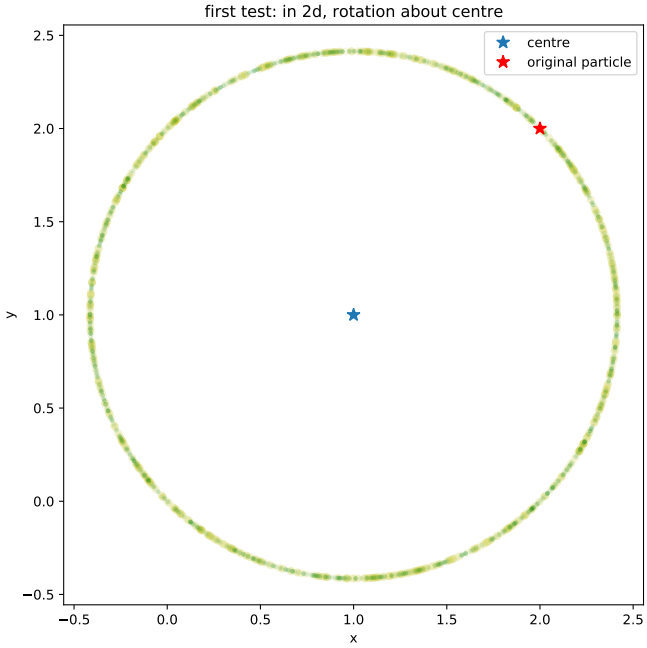

Figure E.2: Visual representation of satellite rotated 1000 times by our matrix about the central.

first apply our rotation matrix to our satellite velocities only and then plot the positions of our satellites before and after rotation in this plane. The positions should match up which is shown by figure E.4. [t!] Finally, if we combine rotation to our positions and rotation to the velocities we will get two independent clusters with different redshift space positions and orientations (albeit still preferentially aligned along the z-axis since RSD provides us with this alignment). This is shown in figure E.5.

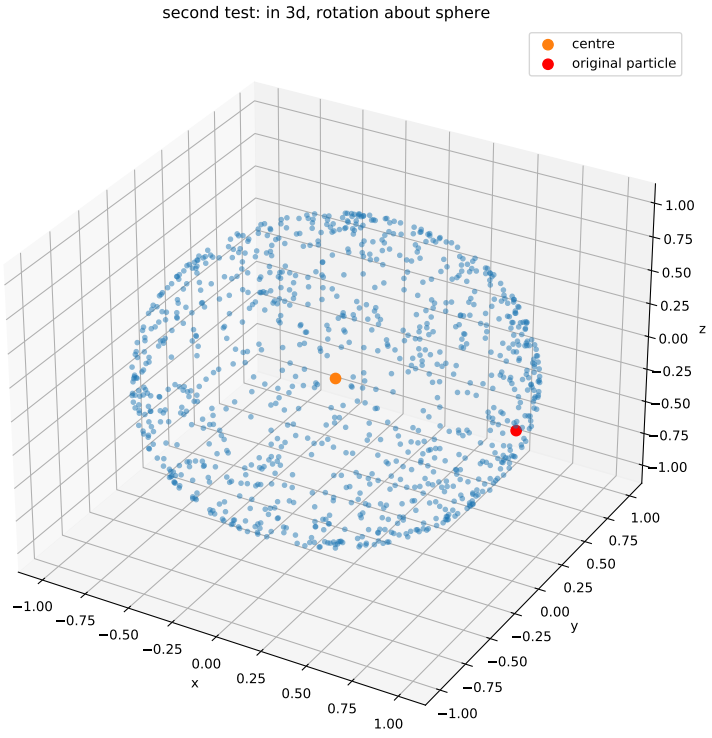

Figure E.3: Same visual as figure E.2 but now in 3 dimensions.

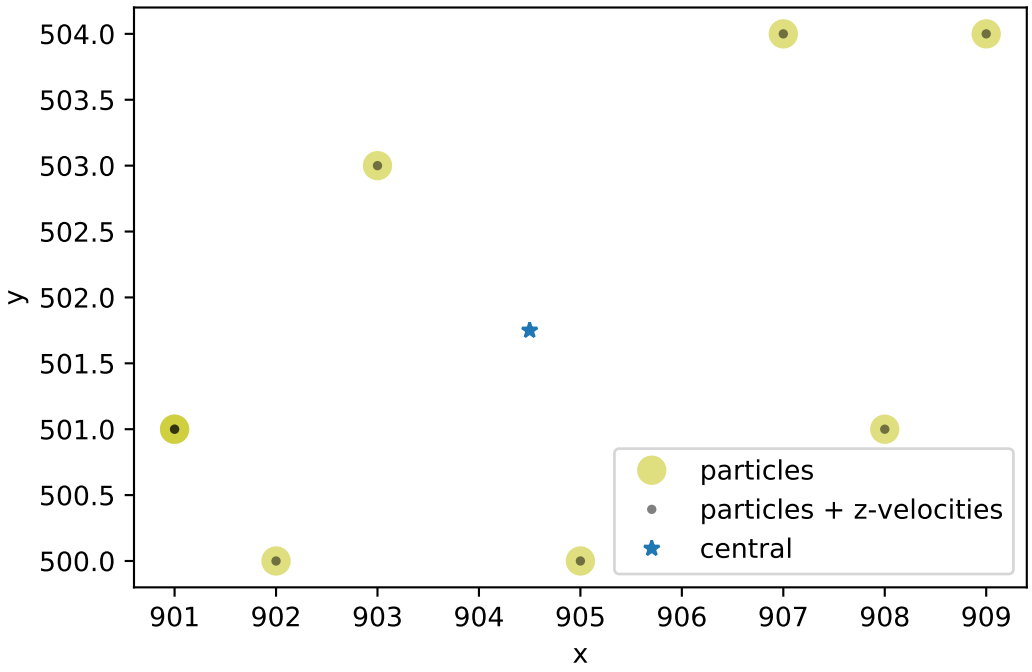

Figure E.4: Case where we only rotate the velocities of our satellites. Shown here is the xy-plane where rotation will only affect the z-components of our velocities. Satellite and central positions are plotted based on their positions in a cube of 1 h^{-1}Mpc.

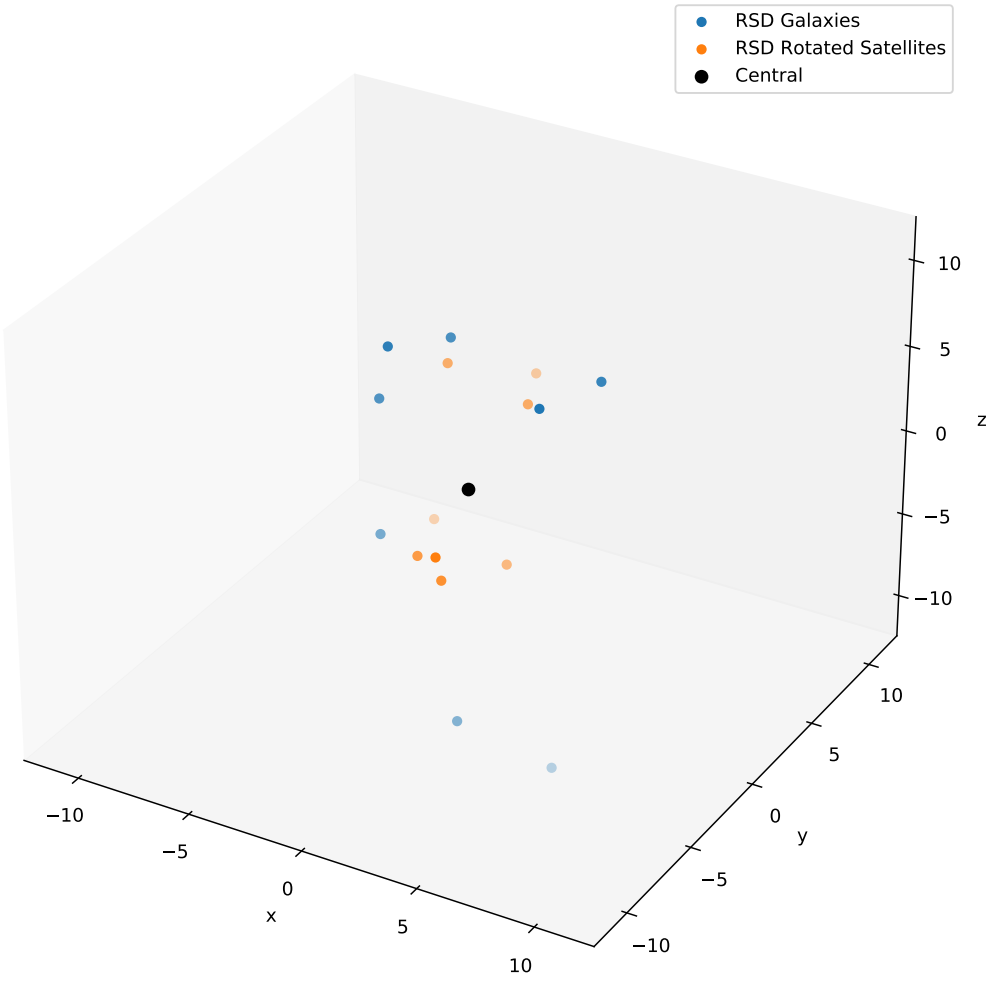

Figure E.5: Combination of rotation applied to both the satellite positions and velocities. Satellite positions are plotted relative to the centre of mass positions in h^{-1}Mpc.

Appendix F

Alignment Function Measurements in Galaxy Surveys

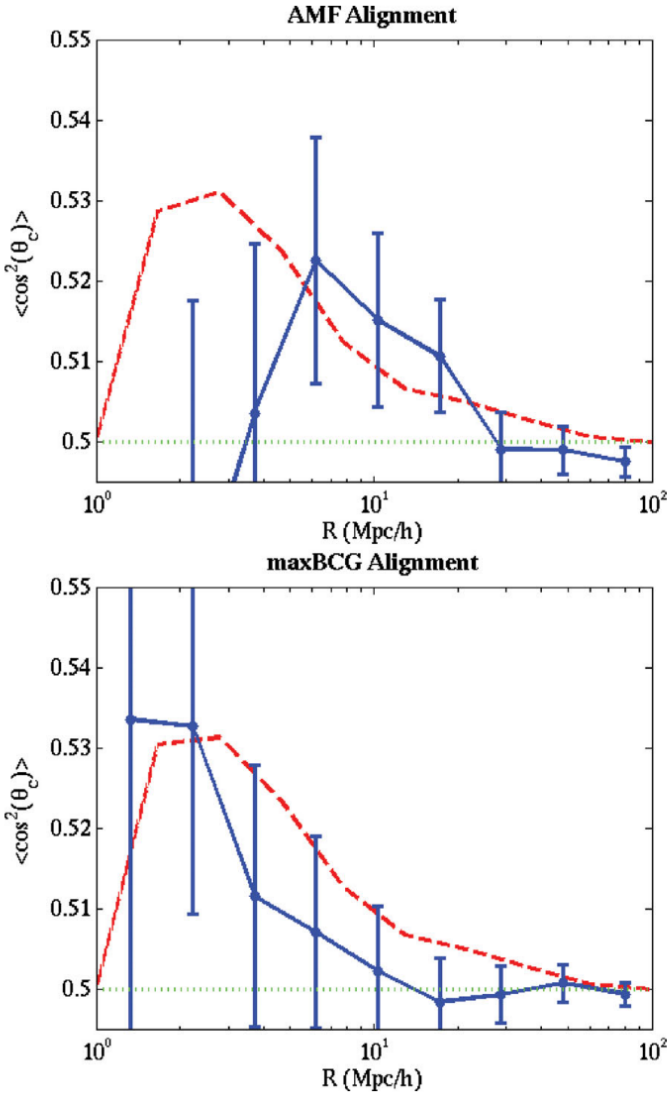

Figure F.1: Cluster correlation angle alignments as a function of comoving separation for adaptive matched filter (AMF) catalogue (*top*) and maxBCG catalogue (*bottom*). Blue points with error bars are the observational results with 2.5 (AMF) and 2σ (maxBCG) uncertainties. The red dashed lines are theoretical predictions from comparative simulation data.

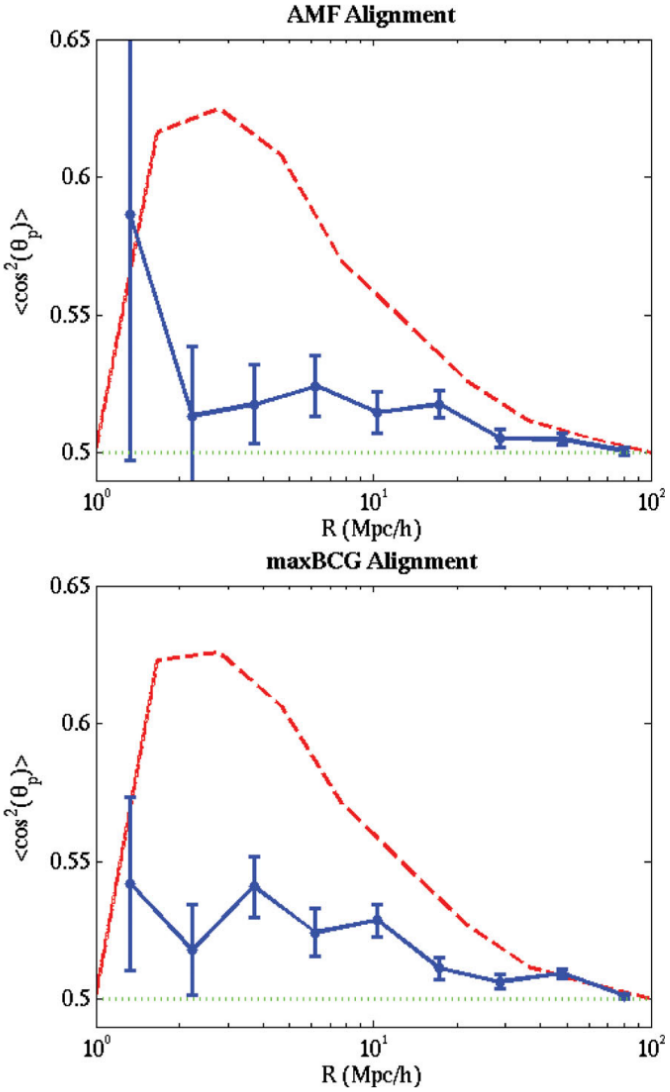

Figure F.2: Cluster pointing angle alignment as a function of comoving separation For the same two catalogues with 6σ (AMF) and 10σ (maxBCG) uncertainties. Line format is the same as in figure F.1.

Appendix G

Redshift Space to Real Space Quadrupole

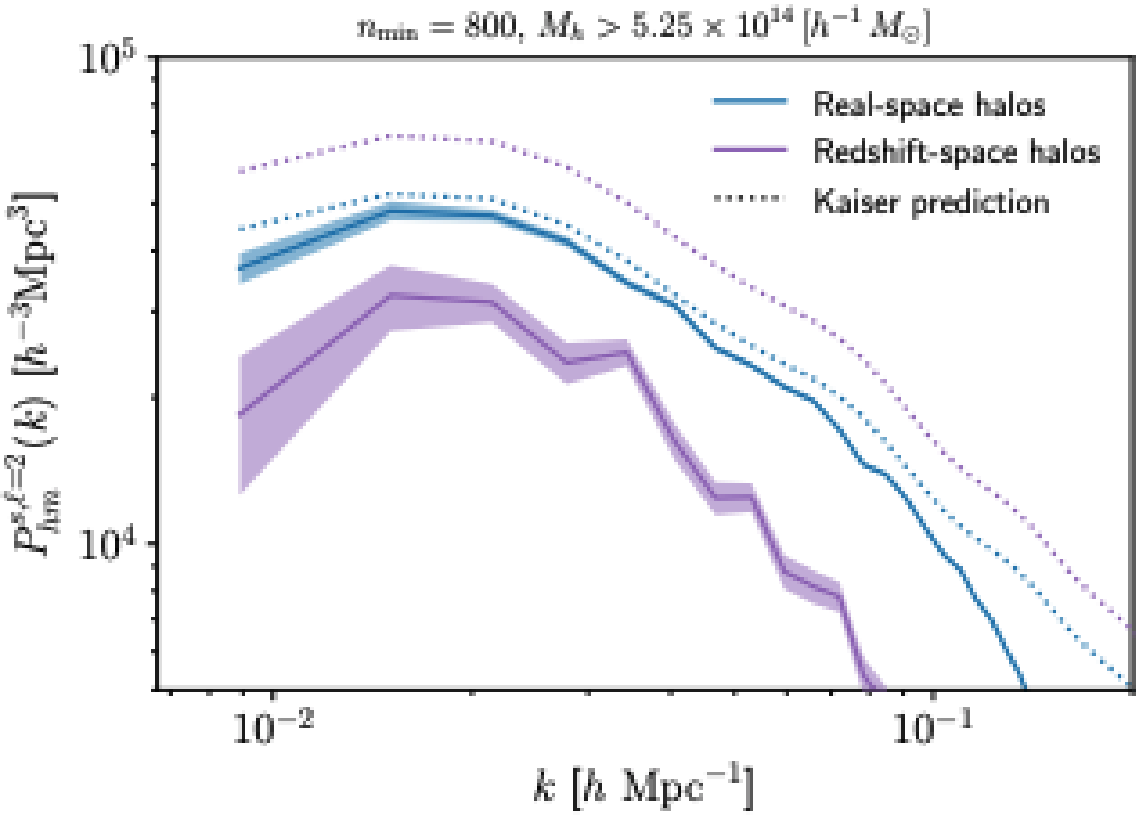

Figure G.1: Quadrupole of the halo-matter cross power spectrum from Obuljen et al. (2019). Shown are the cases of halos found in redshift space and moved back to real space (purple) versus halos found in real space (blue). As we can see, these quadrupoles are different and is based upon anisotropic selection of halos in redshift space. Over-plotted are the Kaiser theory predictions for each.

Appendix H

Python Module: Algorithms

Below is majority of the code that was used to modify our HOD by including halo shape statistics along with our own pair counting algorithm to compute the auto and cross correlations with information of shapes.

```
1
2  from nbodykit.lab import *
3  import numpy as np
4  from nbodykit.algorithms.fof import equiv_class
5  from numba import jit
6
7  # Algorithm to replace satellite positions with new positions from halo
       catalog.
8  # Requires particle catalog, hod catalog and halo labels starting from 0
9
10 class new_galaxies(object):
11
12     def __init__(self, cat, hod, labels):
13
14         self.cat = cat
15
16         self.hod = hod
17
18         self.labels = labels
19
20         self.run()
21
22     def run(self):
23
24         new_data = new_satellites(self)
25
26         self.hod_pos = new_data[0]
```

```
27
28          self.hod_vel = new_data[1]
29
30          self.hod_id = new_data[2]
31
32          self.cen_pos = new_data[3]
33
34          self.cen_vel = new_data[4]
35
36          self.cen_id = new_data[5]
37
38          self.sat_pos = new_data[6]
39
40          self.sat_vel = new_data[7]
41
42          self.sat_id = new_data[8]
43
44
45
46  # The functions used
47
48  def new_satellites(self):
49
50      print("starting new satellite positions function")
51
52      u_sat = np.unique(self.hod['halo_id'][self.hod['gal_type']==1].
    compute(), return_counts = True)[0]
```

```
53
54      u_cen = np.unique(self.hod['halo_id'][self.hod['gal_type']==0].
        compute(), return_counts = True)[0]
55
56      if np.isin(u_sat, u_cen).all() != True:
57
58          print("Not all satellite labels have a corresponding central
        label")# this shows a few satellites do not have centrals
59
60      else:
61
62          print(" All satellites have a central")
63
64      new_u_sat = np.array(list(set(u_sat).intersection(u_cen))) # change
        sats to only be ones where we have a cen
65
66      if np.isin(new_u_sat, u_cen).all() == True:
67
68          print("Fixed: all satellite labels are now associated with
        central labels") # now this shows all satellites have a central
69
70      if np.isin(u_cen, new_u_sat).all() == True:
71
72          print("All centrals have a satellite")
73
74      else:
75
```

```
76          print("Some centrals do not have a satellite")
77
78      sat_ids = self.hod['halo_id'][self.hod['gal_type']==1].compute()
79
80      if np.isin(sat_ids, u_cen).all() != True:
81
82          print("Some satellite ID's are not associated with central ID") #
        shows the satellite by number not paired with cen
83
84      wrong_ids = np.array(list(set(u_sat).difference(u_cen)))
85
86      new_sat_ids = np.array([i for i in sat_ids if i not in wrong_ids])
87
88      new_sat_counts = np.unique(new_sat_ids, return_counts = True)[1]
89
90      alternate_counts = np.bincount(new_sat_ids)[np.bincount(new_sat_ids)
        != 0]
91
92      if np.any(alternate_counts == new_sat_counts):
93
94          print("Satellite counts are complete")
95
96      if np.isin(new_sat_ids, u_cen).all() == True:
97
98          print("Fixed: all satellite ID's are associated with Central ID")
        # check that new sat_ids exist with centrals
99
```

```
100     new_pos = []
101
102     new_vel = []
103
104     new_id = []
105
106     for i,j in enumerate(new_u_sat):
107
108         Ns = new_sat_counts[i]
109
110         S = self.cat['Position'][np.where(self.labels==j)].compute()
111
112         V = self.cat['Velocity'][np.where(self.labels==j)].compute()
113
114         ind = np.random.choice(S.shape[0], Ns, replace = False)
115
116         new_pos.append(S[ind,:])
117
118         new_vel.append(V[ind,:])
119
120         l = [j]*Ns
121
122         new_id.append(l)
123
124     sat_pos = np.concatenate(new_pos)
125
126     sat_id = np.concatenate(new_id)
```

```
127
128     sat_vel = np.concatenate(new_vel)
129
130     u_cen = np.unique(self.hod['halo_id'][self.hod['gal_type']==0].
    compute(), return_counts = True)[0]
131
132     cen_pos = self.hod['Position'][np.where(self.hod['gal_type'] == 0)].
    compute()
133
134     cen_vel = self.hod['Velocity'][np.where(self.hod['gal_type'] == 0)].
    compute()
135
136     cen_id = np.array([j for i,j in enumerate(u_cen)])
137
138     hod_pos = np.concatenate((cen_pos,sat_pos))
139
140     hod_vel = np.concatenate((cen_vel, sat_vel))
141
142     hod_id = np.concatenate((cen_id,sat_id))
143
144     return hod_pos, hod_vel, hod_id, cen_pos, cen_vel, cen_id, sat_pos,
    sat_vel, sat_id
145
146 # Fast 3d Inertia tensor
147
148 def three_d_inertia_values(halos, particles, label):
149
```

```
150 Nhalo0 = label.max()+1
151
152 N = np.bincount(label, minlength=Nhalo0)
153
154 lenN = len(N)
155
156 dx = (particles[:,0]-halos[label][:,0])
157
158 dy = (particles[:,1]-halos[label][:,1])
159
160 dz = (particles[:,2]-halos[label][:,2])
161
162 dr2 = np.sum((particles[:,:3]-halos[label][:,:3])**2, axis=1)
163
164 Ixy = equiv_class(label, dx*dy/dr2, op=np.add, dense_labels=True,
     minlength=lenN)/N
165
166 Ixz = equiv_class(label, dx*dz/dr2, op=np.add, dense_labels=True,
     minlength=lenN)/N
167
168 Iyz = equiv_class(label, dy*dz/dr2, op=np.add, dense_labels=True,
     minlength=lenN)/N
169
170 Ixx = equiv_class(label, dx**2/dr2, op=np.add, dense_labels=True,
     minlength=lenN)/N
171
```

```
172   Iyy = equiv_class(label, dy**2/dr2, op=np.add, dense_labels=True,
       minlength=lenN)/N
173
174   Izz = equiv_class(label, dz**2/dr2, op=np.add, dense_labels=True,
       minlength=lenN)/N
175
176   return Ixx,Iyy,Izz,Ixy,Ixz,Iyz
177
178 # compute max 3d eigenvector. Points along major axis
179
180 def max_3d_eigenvector(Ixx,Iyy,Izz,Ixy,Ixz,Iyz):
181
182     eigenvectors = np.zeros((Ixx.size,3))
183
184     for i in range(0,Ixx.size):
185
186         tensor = np.array(((Ixx[i],Ixy[i],Ixz[i]),(Ixy[i],Iyy[i],Iyz[i])
      ,(Ixz[i],Iyz[i],Izz[i])))
187
188         output = np.linalg.eig(tensor)
189
190         array_max = np.amax(abs(output[0]))
191
192         ind = np.where(abs(output[0])==array_max)[0]
193
194         eigenvectors[i] = output[1][:,ind].T[0]
195
```

```
196     return eigenvectors
197 
198 @jit(nopython=True)
199 def two_d_cos_thetas(data, eigenvectors, rp_edges, pimax):
200 
201     # Bins in rp
202 
203     Nrpbins = rp_edges.size-1
204 
205     # Number of halos for loop
206 
207     n1 = len(data)
208 
209     bin_centres = (rp_edges[1:]+rp_edges[:-1])/2
210 
211     #Set all arrays
212 
213     cos_thetac, cos_thetap, counts = np.zeros((3,Nrpbins))
214 
215     # Double loop for pair counts with data
216 
217     for i in range(0,n1):
218 
219         for j in range(i+1,n1):
220 
221             dd_rp = np.sqrt((data[i,0]-data[j,0])**2 + (data[i,1]-data[j
    ,1])**2)
```

```
222
223             dd_pi = np.sqrt((data[i,2]-data[j,2])**2)
224
225             bin_i = int(Nrpbins*(dd_rp-rp_edges[0])/(rp_edges[-1]-
    rp_edges[0]))
226
227             if (bin_i <= Nrpbins-1) & (dd_rp >= rp_edges[0]) & (dd_pi <=
    pimax):
228
229                 rx, ry = (data[i,0] - data[j,0]), (data[i,1] - data[j,1])
230
231                 rps = np.array((rx,ry)) / dd_rp
232
233                 counts[bin_i] += 1
234
235                 cos_thetac[bin_i] += np.dot(eigenvectors[i],eigenvectors[
    j])*np.dot(eigenvectors[j],eigenvectors[i])
236
237                 cos_thetap[bin_i] += np.dot(eigenvectors[i],rps)*np.dot(
    rps,eigenvectors[i])
238
239     return cos_thetac, cos_thetap, bin_centres, counts
240
241 @jit(nopython=True)
242 def three_d_cos_thetas(data,eigenvectors,edges):
243
244     # Bins in r
```

```
245
246     Nrbins = edges.size-1
247
248     # Number of halos for loop
249
250     n1 = len(data)
251
252     bin_centres = (edges[1:]+edges[:-1])/2
253
254     #Set all arrays
255
256     cos_thetac, cos_thetap, counts = np.zeros((3,Nrbins))
257
258     # Double loop for pair counts with data
259
260     for i in range(0,n1):
261
262         for j in range(i+1,n1):
263
264             dd_r = (np.sqrt((data[i,0]-data[j,0])**2 + (data[i,1]-data[j
    ,1])**2 + (data[i,2] - data[j,2])**2))
265
266             bin_i = int(Nrbins*(dd_r-edges[0])/(edges[-1]-edges[0]))
267
268             if (bin_i <= Nrbins-1) & (dd_r >= edges[0]):
269
270                 counts[bin_i] += 1
```

```
271
272                 rx,ry,rz = (data[i,0] - data[j,0]), (data[i,1] - data[j
    ,1]), (data[i,2] - data[j,2])
273
274                 R = np.array((rx,ry,rz))
275
276                 rs = R / dd_r
277
278                 cos_thetac[bin_i] += (np.dot(eigenvectors[i],eigenvectors
    [j]))**2
279
280                 cos_thetap[bin_i] += (np.dot(eigenvectors[i],rs))**2
281
282     return cos_thetac, cos_thetap, bin_centres, counts
283
284
285
286 @jit(nopython=True)
287 def PairCounts(data,randoms,edges,pimax,e1,e2):
288
289     # Set my bins
290
291     Nrpbins = edges.size-1
292
293     Npibins = pimax
294
295     # Use unit depth spacing for pi bins ( 1 Mpc/h)
```

```
296
297     pi_values = np.linspace(0,pimax,pimax+1)
298
299     # Number of halos for each loop
300
301     n1 = len(data)
302
303     n2 = len(randoms)
304
305     #Set all arrays
306
307     dd_counts,dr_counts,rr_counts,ss,sd,sr  = np.zeros((6,Nrpbins,Npibins
    ))
308
309     #Bin centres for rp
310
311     bin_centres = (edges[1:]+edges[:-1])/2.
312
313     # Double loop for pair counts with data
314
315     for i in range(0,n1):
316
317         for j in range(i+1,n1):
318
319             dd_rp = (np.sqrt((data[i,0]-data[j,0])**2+(data[i,1]-data[j
    ,1])**2))
320
```

```
321             dd_pi = (np.sqrt((data[i,2]-data[j,2])**2))
322
323             bin_i = int(Nrpbins*(dd_rp-edges[0])/(edges[-1]-edges[0]))
324
325             bin_j = int(Npibins*(dd_pi-pi_values[0])/(pi_values[-1]-
    pi_values[0]))
326
327             if (bin_i <= Nrpbins-1) & (dd_rp >= edges[0]) & (bin_j <=
    Npibins-1):
328
329                 dd_counts[bin_i,bin_j] += 1
330
331                 rx,ry = (data[i,0] - data[j,0]), (data[i,1] - data[j,1])
332
333                 r_squared = rx**2 + ry**2
334
335                 sin2phi = (2*rx*ry)/r_squared
336
337                 cos2phi = 1 - 2*(ry**2)/r_squared
338
339                 SplusSplus = (-1*(e1[i]*cos2phi + e2[i]*sin2phi) * (-1*(
    e1[j]*cos2phi + e2[j]*sin2phi)))
340
341                 SplusD = -1*(e1[i]*cos2phi + e2[i]*sin2phi)
342
343                 ss[bin_i,bin_j] += SplusSplus
344
```

```
345                 sd[bin_i,bin_j] += SplusD
346
347     print(" Completed dd ")
348
349     #Double loop for randoms
350
351     for i in range(0,n2):
352
353         for j in range(i+1,n2):
354
355             rr_rp = (np.sqrt((randoms[i,0]-randoms[j,0])**2+(randoms[i
    ,1]-randoms[j,1])**2))
356
357             rr_pi = (np.sqrt((randoms[i,2]-randoms[j,2])**2))
358
359             bin_i = int(Nrpbins*(rr_rp-edges[0])/(edges[-1]-edges[0]))
360
361             bin_j = int(Npibins*(rr_pi-pi_values[0])/(pi_values[-1]-
    pi_values[0]))
362
363             if (bin_i <= Nrpbins-1) & (rr_rp >= edges[0]) & (bin_j <=
    Npibins-1):
364
365                 rr_counts[bin_i,bin_j] += 1
366
367     print(" Completed rr ")
368
```

```
369     # Double loop for data-randoms
370 
371     for i in range(0,n1):
372 
373         for j in range(0,n2):
374 
375             dr_rp = ((np.sqrt((data[i,0]-randoms[j,0])**2 + (data[i,1]-
    randoms[j,1])**2)))
376 
377             dr_pi = ((np.sqrt((data[i,2]-randoms[j,2])**2)))
378 
379             bin_i = int(Nrpbins*(dr_rp-edges[0])/(edges[-1]-edges[0]))
380 
381             bin_j = int(Npibins*(dr_pi-pi_values[0])/(pi_values[-1]-
    pi_values[0]))
382 
383             if (bin_i <= Nrpbins-1) & (dr_rp >= edges[0]) &(bin_j <=
    Npibins-1):
384 
385                 dr_counts[bin_i,bin_j] += 1
386 
387                 dr_tan_phi = (data[i,1]-randoms[j,1])/(data[i,0]-randoms[
    j,0])
388 
389                 dr_cos2phi = (1 - dr_tan_phi**2)/(1 + dr_tan_phi**2)
390 
391                 dr_sin2phi = (2*dr_tan_phi)/(1 + dr_tan_phi**2)
```

```
392
393                 SplusR = -1*(e1[i]*dr_cos2phi + e2[i]*dr_sin2phi)
394
395                 sr[bin_i,bin_j] += SplusR
396
397     print(" Completed dr ")
398
399     return dd_counts, dr_counts, rr_counts, ss, sd, sr, bin_centres
400
401 # defined separate correlation function outside of my own pair counter
        function that takes a long time ( O(N^2)).
402
403 def correlations(DataSize,RandomsSize,dd,dr,rr,ss,sd,sr,dpi,pimax,edges):
404
405     DD = dd/(0.5*DataSize*(DataSize-1))
406
407     RR = rr/(0.5*RandomsSize*(RandomsSize-1))
408
409     DR = dr/(DataSize*RandomsSize)
410
411     SS = ss/(0.5*DataSize*(DataSize-1))
412
413     SD = sd/(0.5*DataSize*(DataSize-1))
414
415     SR = sr/(DataSize*RandomsSize)
416
417     LS = (DD-2.*DR+RR)/RR
```

```
418
419     LSc = np.concatenate((LS))
420
421     XiPlusPlus1 = np.concatenate((SS/RR))
422
423     XiPlusG1 = np.concatenate(((SD - SR)/ RR))
424
425     nrpbins = edges.size - 1
426
427     npibins = dd.size // nrpbins
428
429     if npibins != pimax:
430
431         msg = 'pibins shouls equal pimax value, check pimax, dd_counts
    and rpbins'
432
433         raise ValueError(msg)
434
435     wp,wp_plusplus,wp_plusg = np.zeros((3,nrpbins))
436
437     for i in range(nrpbins):
438
439         wp[i] = 2.0 * dpi * np.sum(LSc[i * npibins: (i+1) * npibins])
440
441         wp_plusplus[i] = 2.0*dpi*sum(XiPlusPlus1[j] for j in range(i*
    npibins,npibins*(i+1)))
442
```

```
443         wp_plusg[i] = 2.0*dpi*sum(XiPlusG1[j] for j in range(i*npibins,
    npibins*(i+1)))
444
445     return wp, wp_plusg, wp_plusplus
```

Listing H.1: Python example

Appendix I

Systematic Error in f

Although the use of our various catalogues did not provide us with a distinct difference in our multipole measurements, the change to a larger positive ratio than say a theoretical calculation for $P_2^2(k)$ using Kaiser can indicate that there is some additional large-scale bias that is contaminating f when using FoF for our RSD+IA catalogue. This contamination is related to the additional anisotropic assembly bias term, A, in equation 2.8. If we consider equation 1.19 and 1.20, our Kaiser predictions include a positive non-zero galaxy bias, b_g, and a positive, non-zero f that will boost both our multipole measurements. If we only contain b_g and f then these ratio values between each catalogue should be similar for P_0 and P_2. However, if we include the presence of tidal alignments and direction dependant selection, we must re-derive our multipoles using equation 2.8. In doing so we obtain the following:

$$P_0^s(k) = \left[(b_g - \frac{A}{3})^2 + \frac{2}{3}(b_g - \frac{A}{3})(f + A) + \frac{1}{5}(f + A)^2\right] P_m(k) \tag{I.1}$$

$$P_2^s(k) = \left[\frac{4}{3}(b_g - \frac{A}{3})(f + A) + \frac{4}{7}(f + A)^2\right] P_m(k) \tag{I.2}$$

From these predictions, by inspection we notice that if our constant A is non-zero and negative, this can result in the suppression of our quadrupole measurements and perceived f. However, if $A > 0$ we can boost RSD allowing for the quadrupole mainly to become larger while the monopole would remain unaffected (i.e. for different values of A in equation I.1 the terms out-front of our linear power would be equivalent). For example, if we were to use the theoretical prediction of f, estimate b_g from real space and then estimate A for our measured P_2 from our RSD+IA catalogue grouped with FoF, we find for our simulations we would obtain a value of $A = 0.4$ for instance which greatly changes the perceived growth function. In figure I.1 we show this estimation with standard kaiser over-plotted. Here the black solid line represents our measurements and the orange dashed line would be equation I.2 with a positive anisotropic assembly bias term. If we were to have pure RSD (the blue dashed line), since our monopole ratio is zero with respect to Kaiser then this quadrupole ratio would have to follow the same pattern. It is clear this would not be the case here. This difference may be of interest for future work if we have a stronger intrinsic alignment signal that contaminated our redshift space measurements. In doing so, one may then use these models to extract a value for A based on your selection of objects in redshift space.

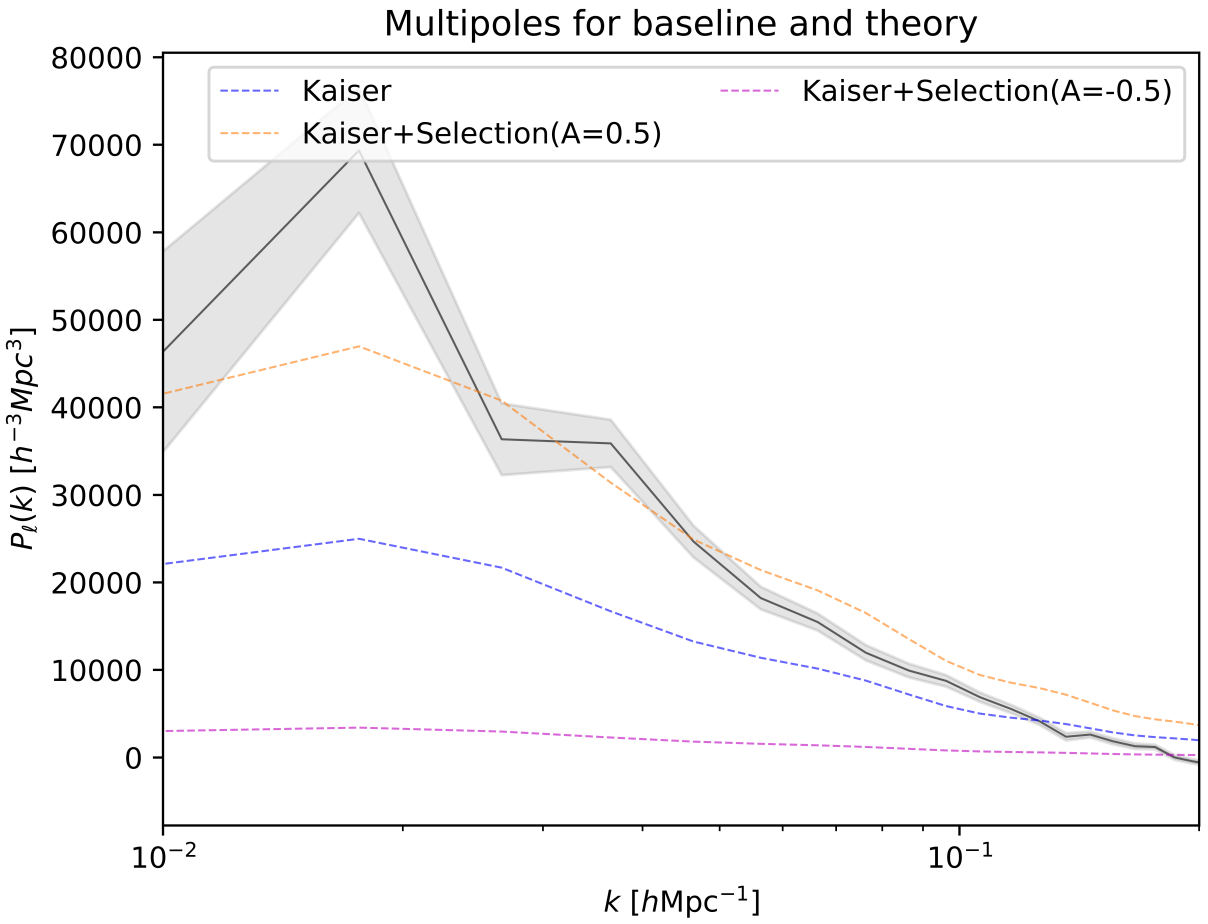

Figure I.1: Example of our measured quadrupole (solid black) against various theory curves. In blue is the standard Kaiser prediction in redshift space with a calculated bias, b_g, from the monopole and theoretical prediction of f for a redshift of $z = 0.5$. In orange (magenta) we have included our anisotropic assembly bias term, A with an example for a positive value (orange) and negative value (magenta) that would either enhance or suppress RSD respectively.

Appendix J

Multipoles for Galaxy Catalogues

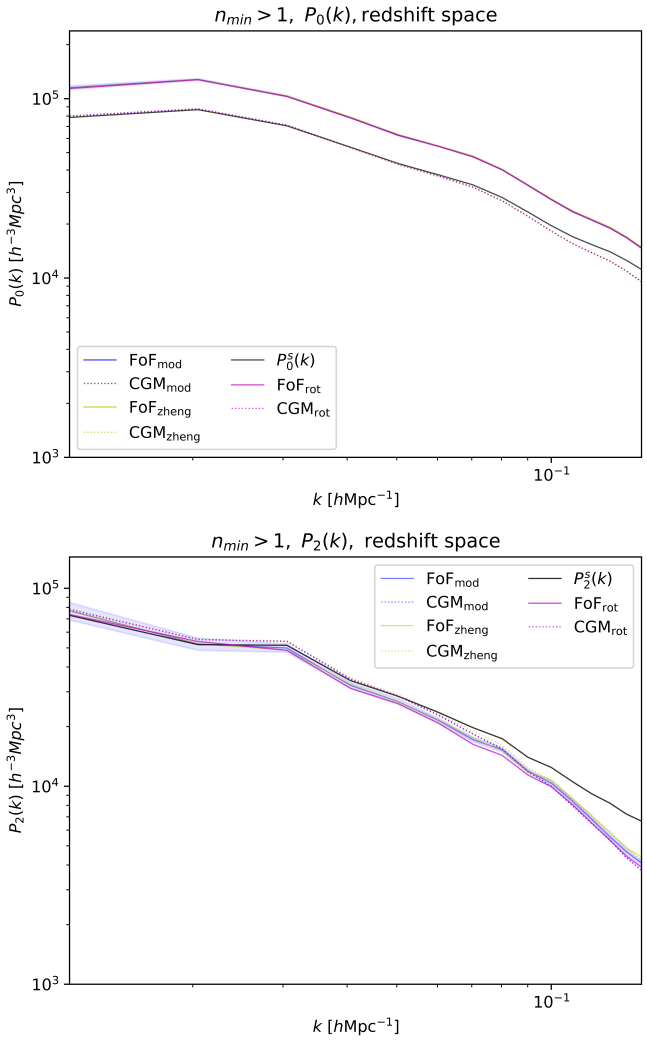

Figure J.1: Monopole and quadrupole for all galaxy catalogues. Included are our modified catalogue (blue), rotated catalogue (green), Zheng07 catalogue (yellow) along with a baseline model (black). The baseline model if the power spectrum computed on our galaxy catalogue before any grouping using FoF or CGM. The different group finders are separated by line style (Solid lines are for groups found using FoF and dashed lines are for groups found using CGM).