

## ABSTRACT

In order to restore the galaxy clustering of eBOSS galaxies, we apply the SubHalo Abundance Matching (SHAM) method to the UNIT N-body simulation. The eBOSS galaxy sample contains 174,816 LRGs at  $0.6 < z < 1.0$  and 173,736 ELGs at  $0.6 < z < 1.1$ . To mitigate the fibre collision effects, their correlation functions are calculated using the pairwise-inverse-probability and angular up-weighted (PIP+ANG) pair counts. The covariance matrices of the correlation function multipoles are obtained from 1000 realisations of the multi-tracer Ezmocks. We use the peak maximum circular velocity over the mass accretion history (i.e.,  $V_{\text{peak}}$ ) in the SHAM study. The SHAM models are the same for LRGs and ELGs, with a scattering parameter  $\sigma$  and the massive-end truncate parameter  $V_{\text{cut}}$ . The SHAM ELGs agree well with the observations in both NGC and SGC, with a typical  $V_{\text{peak}} \approx 230 \text{ km/s}$  for their host halos. But there are discrepancies for LRG quadrupoles on small scales. This problem is solved by taking into account the redshift uncertainty in LRG SHAM models. This uncertainty is modelled by Gaussian smearing in SHAM LRGs' peculiar velocity. The host halos of LRG has a typical  $V_{\text{peak}} \approx 700 \text{ km/s}$ , three times larger than ELG host halos. It implies typical halos that host an LRG are 9 times more massive than the typical host halos of ELGs.

*Keywords:* Cosmology:large-scale structure of Universe - galaxies: abundances - galaxies: halos - methods: numerical

## 1. INTRODUCTION

$\Lambda$ -Cold Dark Matter ( $\Lambda$ CDM) is the standard model of the cosmology today. It is a well-established framework that describes the gravitational evolution of the Universe.  $\Lambda$  refers to the cosmological constant, which has been interpreted as a kind of dark energy that has a constant energy density. Cold dark matter is the dark matter particles that spread slowly when the primordial perturbation started to grow. Dark energy results in the accelerated expansion of the Universe while dark matter governs the gravitational evolution of the cosmic structures. Given the primordial field,  $\Lambda$ CDM gives an accurate prediction of the dark matter clustering on large scales (i.e., the large-scale structure, LSS). The LSS has an excess signal at a characteristic comoving scale of about 150 Mpc today, which can be used as the ‘Standard Ruler’ (Eisenstein et al. 2005). The signal is the imprint of the Baryon Acoustic Oscillation (BAO). BAO is the sound waves that propagate in the early universe and is fixed by the gravitational instability after the decoupling of photons (Eisenstein et al. 2007). It provides a direct measurement of the Hubble parameter and the angular diameter distance. Meanwhile, the Redshift-Space Distortion (RSD, Kaiser 1987) is another tool to study the LSS. RSD is caused by the peculiar velocity of the matter, resulting in the observed clustering anisotropy. The anisotropy is a probe of the structure growth rate (Dawson et al. 2016).

The LSS traced by galaxies and quasars can be revealed by redshift surveys. By measuring the redshift of millions of tracers, the surveys give a 3D map of the Universe and enable the statistical studies of the LSS. The redshift survey is an important part of the Sloan Digital Spectroscopic Survey<sup>1</sup> (SDSS). SDSS has four phases to date and the first two phases (i.e., SDSS-I and SDSS-II) have provided over 1.6 million spectra in total until 2008 (Abazajian et al. 2009). The SDSS-III project (2008-2014) consists of four surveys and the Baryon Oscillation Spectroscopic Survey (BOSS) is the largest one (Dawson et al. 2013). BOSS studies BAO and RSD and gives constraints on cosmological parameters at 2% level (Alam et al. 2017). This project has taken the spectra of more than 1.5 million Luminous Red Galaxies (LRG), the most luminous galaxies in the Universe, with luminosity as faint as  $i = 19.9$  mag at redshift  $0.15 < z < 0.75$ , and over 150,000 quasars as faint as  $g = 22$  mag at  $2.15 < z < 3.5$ .

<sup>1</sup> <http://www.sdss.org/>

As the extension of BOSS, the extended Baryon Oscillation Spectroscopic Survey (eBOSS, 2014-2019, Dawson et al. 2016) probes another 2.5 million LRGs ( $0.6 < z < 1.0$ ) and over 5 million quasars (Ross et al. 2020). Additionally, eBOSS has observed around 2 million Emission Line Galaxies (ELGs) at  $0.6 < z < 1.1$  (Raichoor et al. 2020). ELGs are blue, star-forming galaxies that have a characteristic OII doublet emission. They also have a much higher number density than LRG samples (Dawson et al. 2016). The convenient redshift measurement and a large amount of ELGs make them a practical BAO tracer for next-decade redshift surveys like the Dark Energy Spectroscopic Instrument (DESI, 2020-2025, DESI Collaboration (2016)). With new tracers at a higher redshift range, eBOSS consequently expects a more precise distance scale measurement in an epoch when the dark energy started to dominate in our universe (Dawson et al. 2016).

Compared to large scales, constructing analytical theories on small scales to explain the non-linear phenomena is difficult. N-body simulations are a kind of useful tool to study the structure formation there. It typically simulates the motions of dark matter particles in a box and thus models the matter density evolution under a given cosmology framework.

However, N-body simulations and observations cannot be directly compared. N-body simulations trace particles that only interact through gravity, i.e., the dark matter particles. While observations (e.g., galaxy surveys) detect the luminous objects, i.e. the baryonic matter. The clustering of baryons does not necessarily follow that of the dark matter on all scales. For instance, JING et al. (1998) find that the correlation function of galaxies is lower than that of the N-body simulation. In order to bridge the gap, they introduce a galaxy bias model that describes the galaxy distribution in dark matter halos, i.e., the Halo Occupation Distribution (HOD) method. This bias model enables the description of baryonic distribution under the  $\Lambda$ CDM framework and the model parameters can be constrained by observations. This becomes increasingly important as the massive spectroscopic surveys like eBOSS and DESI map the Universe with higher and higher precision.

HOD originates from a phenomenological assumption that for a dark matter halo with a given mass  $M_h$ , there is a certain probability to host  $N$  galaxies, i.e.,  $P(N|M_h)$  (see Peacock & Smith 2000; Berlind & Weinberg 2002, etc. for the pioneering studies). HOD has been successfully applied to BOSS researches like Nuza et al. (2013); Guo et al. (2014). SubHalo Abundance Matching (SHAM) is another way to assign galaxies to dark matter halos (see Kravtsov et al. 2004; Conroy et al. 2006; Behroozi et al. 2010, etc. for the pioneering studies). The assumption is that massive galaxies reside in massive halos and one halo (or subhalo) can only have one galaxy in its centre. Rodríguez-Torres et al. (2016); Campbell et al. (2018); Tinker et al. (2019) have used SHAM to study LRGs and quasars for BOSS.

This master project modifies a SHAM method proposed by Rodríguez-Torres et al. (2016) to calibrate the galaxy-halo relation between the eBOSS galaxy tracers (Mohammad et al. 2020) and the UNIT N-body simulation (Chuang et al. 2019). Section 2 introduces the basis of  $\Lambda$ CDM and Section 3 presents the relation between the galaxy bias and the 2-point correlation function. Section 4 introduces the observational data and simulations included in this project and Section 5 describes the SHAM method and the galaxy-halo relation calibration with `pymultinest`. Section 6 shows the results of the SHAM model and the LRG SHAM improvement achieved by considering the redshift uncertainty. Section 7 discusses another two possible solutions to reduce the difference between the LRG SHAM model and observation. Finally, we conclude in Section 8.

## 2. $\Lambda$ CDM COSMOLOGY

$\Lambda$ CDM is the standard model that describes the gravitational evolution of the universe with General Relativity. Its basic assumption is that the universe is isotropic and homogeneous on large scales. It is a realistic assumption that has been confirmed by various observations like the Cosmic Microwave Background (CMB, e.g., Planck, Planck Collaboration 2019).

### 2.1. Friedmann-Lemaître-Robertson-Walker Metric

The Friedmann-Lemaître-Robertson-Walker metric (i.e., FLRW metric) is a general metric that describe an isotropic and homogeneous spacetime as follows:

$$\begin{aligned} ds^2 &= c^2 dt^2 - dl^2 \\ &= c^2 dt^2 - a^2(t) \left( \frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right) \end{aligned} \quad (1)$$

where  $dt$  is the time interval and  $dr$ ,  $d\Omega$  are the radial interval and transverse spatial interval. The spatial coordinate  $dr$  and  $d\Omega$  is the comoving coordinate which does not evolve. In comoving coordinates, the distances between objects

evolve together with the universe expansion, which is quantified by the scale factor  $a(t) \equiv \frac{R(t)}{R(t_0)}$ .  $R(t_0)$  is the scale of the universe today and  $R(t)$  is the scale of the universe at time  $t$ .  $K = -1, 0, 1$ , represent an open, flat and closed universe respectively.

## 2.2. Friedmann equations<sup>2</sup>

The Einstein-Hilbert action gives the equation of motion for matter in the gravitational field, i.e., the Einstein equation

$$(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R) - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}. \quad (2)$$

The left-hand side represents the deformation of the spacetime due to the presence of matter. The deformation creates the gravitational field and leads to the motion of matter inside.  $g_{\mu\nu}$  is the metric for an arbitrary coordinate;  $R_{\mu\nu}$  and  $R$  are the Ricci tensor and scalar, and  $R_{\mu\nu}$  represents the second derivative of  $g_{\mu\nu}$ .  $\Lambda$  is the cosmological constant. The right-hand side is the matter property embodied by the energy-momentum tensor  $T_{\mu\nu}$ . In the first order approximation, the universe can be regarded as the perfect fluid. In this case,  $T_{\mu\nu} = (\rho + p/c^2)u_\mu u_\nu - pg_{\mu\nu}$  and the Einstein equation given the FLRW metric becomes:

$$\frac{\dot{a}^2}{a^2} + \frac{Kc^2}{a^2} - \frac{\Lambda c^2}{3} = \frac{8\pi G}{3}\rho, \quad (3)$$

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{Kc^2}{a^2} - \Lambda c^2 = -\frac{8\pi G}{c^2}p. \quad (4)$$

where  $\rho$  and  $p$  are the energy density and pressure of the fluid respectively. Eq. 3 and Eq. 4 are known as the Friedmann equations. The energy density  $\rho$  includes the matter ( $\rho_m$ ) and the radiation ( $\rho_\gamma$ ) that obey the law of energy conservation, i.e.,  $\nabla T_{\mu\nu} = 0$ . With FLRW metric, the law of energy conservation can be expressed as

$$\frac{\partial}{\partial t}(\rho c^2 a^3) + p \frac{\partial a^3}{\partial t} = 0. \quad (5)$$

The equation of state (EOS) for the matter is  $p = 0$ , for the radiation is  $p = \rho c^2/3$ . Feeding their EOSs to Eq. 5, we obtain the matter energy density  $\rho_m a^3 = \text{const.}$  and the radiation energy density  $\rho_\gamma a^4 = \text{const.}$ . In other words, the relations between the energy density at time  $t$  and the present  $t = 0$  are  $\rho_m(t) = \rho_{m,0}/a^3(t)$  and  $\rho_\gamma(t) = \rho_{\gamma,0}/a^4(t)$ . The cosmological constant can also be interpreted as a component of the universe. So we define its energy density as  $\rho_\Lambda(t) = \frac{\Lambda c^2}{8\pi G}$ . Similarly, the curvature energy density can be defined as  $\rho_K(t) = -\frac{3Kc^2}{8\pi G a^2(t)}$ . By doing so, we can move those terms in Eq. 3 to its right-hand side and obtain a relation between the expansion rate of the universe and its contents.

The Hubble parameter is defined as

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \quad (6)$$

and the Hubble constant  $H_0$  is the Hubble parameter at present.  $H_0$  is often expressed as  $100h$  (km/s)/Mpc. With the Hubble constant, we define the critical energy density as  $\rho_c = \frac{3H_0^2}{8\pi G}$  and introduce a dimensionless quantity to quantifies the amount of those contents - the density parameter  $\Omega_X = \frac{\rho_X}{\rho_c}$ . So Eq. 3 turns into  $H^2(t) = H_0^2(\Omega_m + \Omega_\gamma + \Omega_\Lambda + \Omega_K)$ .

The photon's wavelength evolve with the universe expansion, inducing the cosmological redshift  $z$ . So the redshift  $z$  is related to the scale factor  $a(t)$  as

$$a(t) = \frac{1}{1+z}. \quad (7)$$

Then Eq. 3 can be specified by the measurement at  $t=0$  as

$$\frac{H^2(z)}{H_0^2} = \Omega_{m,0}(1+z)^3 + \Omega_{\gamma,0}(1+z)^4 + \Omega_{\Lambda,0} + \Omega_{K,0}(1+z)^2. \quad (8)$$

<sup>2</sup> Referred to the lecture notes of *General Relativity and Cosmology II* delivered by Professor Mikhail Shaposhnikov

### 2.2.1. Distances

Distance is an important quantity to match the theoretical description and the observation. Assuming photons emitted by an object in  $r_1$  at  $t_1(z = z_1)$  travel along a radial trajectory and arrive at the observer in  $r_0$  at  $t_0(z = 0)$ , then the comoving distance  $\chi$  they have travelled is

$$\chi = \int_{r_1}^{r_0} \frac{dr}{\sqrt{1 - Kr^2}} = \int_{t_1}^{t_0} \frac{cdt}{a(t)} = \int_0^{z_1} \frac{cdz}{H(z)}. \quad (9)$$

With Eq. 8, the comoving distance  $\chi$  in Eq. 9 can be further written as

$$\chi = \int_0^{z_1} \frac{cdz}{H(z)} = \frac{c}{100h} \int \frac{dz}{\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\gamma,0}(1+z)^4 + \Omega_{\Lambda,0} + \Omega_{K,0}(1+z)^2}}. \quad (10)$$

The physical distance between two objects observed at a fixed time (i.e., the proper distance) is

$$d_{pro}(t) = \int ds(dt = 0) = a(t) \int \frac{dr}{\sqrt{1 - Kr^2}} \quad (11)$$

The proper distance increases as time flows.

### 2.2.2. The Evolution of the Universe

Since the Friedmann equations are non-linear, it is not easy to solve it analytically unless we consider only one dominate component at a time. In the early universe, the radiation was the main driver of the universe expansion. In the radiation-dominated epoch, the scale factor  $a(t)$  evolves as

$$a(t) = a(t_i) \left( \frac{t}{t_i} \right)^{1/2}, \quad (12)$$

where  $t_i$  is the initial time for the expansion. The expression indicates a decelerating expansion because  $\ddot{a} < 0$ . Since the energy density of the matter is diluted slower than that of the radiation, the matter takes over at redshift around 3000. In the matter-dominated era, the pressure term also vanishes. So the solution to the Friedmann equation is

$$a(t) = a(t_i) \left( \frac{t}{t_i} \right)^{2/3}. \quad (13)$$

It implies the expansion is also decelerating ( $\ddot{a} < 0$ ). According to the measurements of redshift surveys (e.g., SDSS-III/BOSS, [Alam et al. 2017](#)) and CMB (e.g., Planck, [Planck Collaboration \(2019\)](#)),  $\Omega_{m,0} = 0.31$  (Baryon  $\Omega_{b,0} = 0.05$ ),  $\Omega_{K,0}, \Omega_{\gamma,0} \approx 0$ ,  $\Omega_{\Lambda,0} = 0.69$ , i.e., the Universe today is  $\Lambda$ -dominated. The transition from a matter-dominated epoch and to such epoch happend at redshift around 0.5. Considering only the cosmological constant energy density in Friedmann equations, the scale factor evolves exponentially as

$$a(t) = a(t_i) \exp\left(\sqrt{\frac{\Lambda}{3}} c(t - t_i)\right). \quad (14)$$

It means that our universe is expanding with acceleration ( $\ddot{a} > 0$ ). Given the constant energy of  $\Lambda$  and the decreasing energy density of other contents, the exponential expansion will continue.

To conclude, in the  $\Lambda$ CDM theory, the universe has been expanding since the early universe. The driver of the expansion has been changed from the radiation, the matter to the cosmological constant. The first two epochs were decelerating expansions epochs with different rates. The present  $\Lambda$ -dominated universe is experiencing an accelerating expansion.

## 3. THE GALAXY CLUSTERING

The isotropy and homogeneity of the universe holds on large scales. For smaller scales, the galaxy distribution forms a cosmic web that has galaxy clusters, filaments and voids. Such large-scale structure originates from the primordial quantum field fluctuations which are frozen by the inflation. At the end of the inflation, the field decays to particles. They become the seeds of the gravitational growth and eventually evolve to the present cosmic web. So the primordial field can only determines the statistical properties instead of the accurate shape the LSS. The primordial field is assumed to be a Gaussian field. Since the properties of a Gaussian field is completely determined by the two-point statistics, the two-point correlation functions (2PCF) and the power spectra are often used in the LSS analysis.

### 3.1. Two-Point Correlation Function<sup>3</sup>

The 2PCF  $\xi(\mathbf{s}, t)$  is a function of separation and time:

$$\xi(\mathbf{s}, t) = \langle \delta(\mathbf{x}, t) \delta(\mathbf{x}-\mathbf{s}, t) \rangle, \quad (15)$$

where  $\delta(\mathbf{x}, t)$  is the density contrast that quantifies the density field fluctuations, which is defined as

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \quad (16)$$

$\bar{\rho}(t)$  is the average density over all the scales which is equivalent to the density of a uniform density field. The density contrast under linear growth can be separated to  $\delta(\mathbf{x}, t) = D(t)\tilde{\delta}(\mathbf{x})$  in which  $D(t)$  is the overdensity growth factor and  $\tilde{\delta}(\mathbf{x})$  is the contrast in position  $\mathbf{x}$ . So the galaxy 2PCF can also be written as

$$\xi(\mathbf{s}, t) = D^2(t) \langle \tilde{\delta}(\mathbf{x}) \tilde{\delta}(\mathbf{x}-\mathbf{s}) \rangle. \quad (17)$$

Observations have shown that dark matter is the main component of matter (e.g., [Planck Collaboration 2019](#)) and they only interact through the gravitational force. So the structure evolution process is dominated by the dark matter field. For galaxy redshift surveys, the dark matter field cannot be directly detected due to its ‘darkness’. But the observed galaxy distribution roughly follows this field. So in practice, it is the galaxy number density, instead of the matter density that is used. Since the 2PCF measures the excess probability of finding a galaxy pair compared to a uniform galaxy sample, random catalogues should also be included in the calculation. For simplicity, we drop the time component from  $\xi$  and only keep the spatial component hereafter. There have been several estimators that use the galaxy pairs to compute the galaxy 2PCF. For instance, the Peebles-Hauser estimator ([Peebles & Hauser 1974](#))

$$\xi_{PH}(s, \mu) = \frac{DD_n(s, \mu)}{RR_n(s, \mu)} - 1 \quad (18)$$

is often used for the simulation.  $DD_n(s, \mu)$  and  $RR_n(s, \mu)$  are normalised galaxy–galaxy pairs (DD) and random–random pairs (RR). They are both functions of the pair separation  $s$  and pair angles  $\beta = \cos^{-1}(\mu)$ . And  $\beta = \pi/2 - \langle \mathbf{s}_1 + \mathbf{s}_2, \mathbf{s}_1 - \mathbf{s}_2 \rangle$  where  $\langle \rangle$  represents the intersection angle of vectors in it,  $\mathbf{s}_1$  is the position vector of galaxy 1 and  $\mathbf{s}_2$  is the position vector of galaxy 2. The SHAM DD pairs are counted by `Corrfunc` Python module ([Sinha & Garrison \(2019\)](#); [Sinha & Garrison \(2020\)](#)) and normalised by  $N_{gal}^2$ . The random samples for simulations are uniformly distributed in a cubic box. Hence the RR term can be evaluated analytically as

$$RR_n(s, \mu) = \frac{4\pi}{3} \frac{s_{\max}^3 - s_{\min}^3}{V_{\text{box}}} \frac{1}{N_\mu}, \quad (19)$$

where  $s_{\max}$  and  $s_{\min}$  are the boundaries of the pair separation bins,  $V_{\text{box}}$  is the volume of the simulation box and  $N_\mu$  is the number of  $\mu$  bins. The correlation function of our SHAM galaxies are calculated with Eq. 18 and Eq. 19.

For galaxy surveys with irregular observation footprints, the Landy–Szalay estimator ([Landy & Alexander 1993](#)) is a preferable choice ([Kerscher et al. 2000](#)):

$$\xi_{LS}(s, \mu) = \frac{DD_n(s, \mu) - 2DR_n(s, \mu) + RR_n(s, \mu)}{RR_n(s, \mu)}. \quad (20)$$

The extra term  $DR_n$  is the normalised galaxy–random galaxy pairs. The random–random pairs  $RR_n$  here is the pair counts from the random catalogue instead of the analytical expression Eq. 19. We use the PIP+ANG corrected galaxy pair counts ([Mohammad et al. \(2020\)](#)) to calculate the correlation functions for eBOSS LRGs and ELGs.

The correlation function can be decomposed by the Legendre polynomials  $P_l(\mu)$  as

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^1 \xi(s, \mu) P_l(\mu) d\mu. \quad (21)$$

Due to the pair-wise symmetric, the odd terms vanish. We consider in this work only the monopole ( $l = 0$ ) and quadrupole ( $l = 2$ ).

<sup>3</sup> Referred to the lecture notes of *AstroIV: Observational Cosmology* delivered by Professor Jean-Paul Kneib in EPFL

### 3.2. Galaxy Bias

Galaxies are biased tracers of the dark matter field. A galaxy bias model explicitly reveals how galaxies distribute in the dark matter field in different scales at different time. Its general expression is

$$\tilde{\delta}_{gal} = f(\tilde{\delta}_{halo}) \quad (22)$$

where  $\tilde{\delta}_{gal}$  is the density contrast of galaxies and  $\tilde{\delta}_{halo}$  is the density contrast of dark matter halos. The function of  $\tilde{\delta}_{halo}$  is the galaxy bias model. The linear galaxy bias model is

$$\tilde{\delta}_{gal} = b \times \tilde{\delta}_{halo}. \quad (23)$$

As a result, the galaxy correlation functions are expected to be  $b^2$  times of the dark matter correlation function in the linear regime.

### 3.3. Redshift Space Distortion<sup>4</sup>

The Redshift Space Distortion (RSD) is the deformation of the galaxy spatial distribution in a space where the positions of galaxies are altered apparently by their peculiar velocity, i.e., in the redshift space. The distortion happens because the peculiar velocity along the line of sight will contaminate the cosmological redshift measurement. In contrast, the real space is where the objects reside in their true positions. In a flat universe, the time derivative of the proper distance  $\dot{\mathbf{d}}_{pro} = \dot{a}(t)\mathbf{r} + a(t)\dot{\mathbf{r}}$  where  $\mathbf{r}$  is the comoving distance expressed in real space and  $a(t)$  is the scale factor. The first term  $\dot{a}(t)$  is the Hubble flow and the second one  $a(t)\dot{\mathbf{r}}$  is the peculiar velocity.  $\dot{\mathbf{d}}_{pro} = \dot{a}(t)\mathbf{d}_{redshift}$  is the object velocity expressed in the redshift space. So the relation between the position in the redshift space and the real space is

$$\mathbf{d}_{redshift} = \mathbf{r} + \frac{\mathbf{v}_{pec}}{a(t)H(z)}. \quad (24)$$

Objects in over-density regions tend to have peculiar velocity pointing to their mass centre. Fig. 1 illustrates the effect of peculiar velocities on the correlation functions on different scales. On large scales, the apparent shift brought by the peculiar velocity is small compared to the separation scale. So it results in a squeezing effect along the line of sight and a negative galaxy correlation function quadrupole (Kaiser 1984). But on small scales, the relatively large apparent displacement of the peculiar velocity extends the separation scales. So the over-density regions appear elongated along the line of sight. This is the Finger-of-God effect, leading to the positive galaxy quadrupoles (Hamilton 1998). In our study, we assume that an observer measures the clustering from the direction that is parallel to the Z-axis (the third Cartesian coordinate). So Eq. 24 are all projected to the Z direction, i.e.,

$$Z_{redshift} = Z_{real} + \frac{V_z}{a(t)H(z)}, \quad (25)$$

where  $Z$  is the position along the Z-axis and  $V_z$  is the peculiar velocity in the Z-axis.

Due to the RSD effect, the density contrast of galaxies in redshift space  $\delta_{gal,obs}$  becomes (Kaiser 1987)

$$\delta_{gal,obs} = \delta_{gal} - \frac{\partial_d(\mathbf{v} \cdot \mathbf{n})}{H}, \quad (26)$$

where  $\delta_{gal}$  is the real-space density contrast,  $H$  is the Hubble parameter,  $\mathbf{v}$  is the peculiar velocity of objects, and  $\mathbf{n}$  is the line of sight and the partial derivative is w.r.t. the proper distance  $d_{pro} = a(t)r$  by assuming a flat  $\Lambda$ CDM. So the observed correlation function

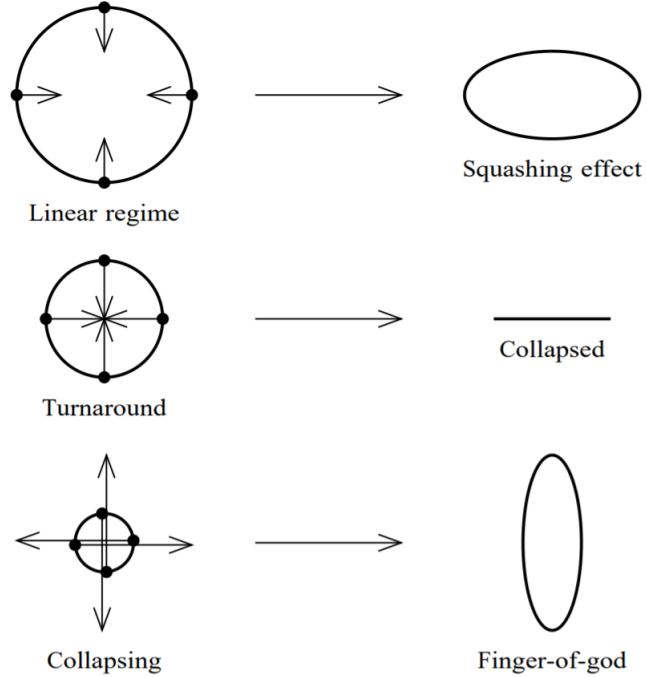
$$\begin{aligned} \xi_{gal,obs} &= D^2 \langle \tilde{\delta}_{gal,obs} \times \tilde{\delta}'_{gal,obs} \rangle \\ &= D^2 \langle (\tilde{\delta}_{gal} - \frac{\partial_d(\mathbf{v} \cdot \mathbf{n})}{H}) \times (\tilde{\delta}'_{gal} - \frac{\partial_d(\mathbf{v}' \cdot \mathbf{n})}{H}) \rangle. \end{aligned} \quad (27)$$

Decomposing it up to the hexadecapole ( $l = 6$ ) (Hamilton 1992), the observed correlation function is

$$\xi_{gal,obs} = D^2 \left\{ \left(1 + \frac{2f}{3} + \frac{f^2}{5}\right) \mu_0(s) - \left(\frac{4f}{3} + \frac{4f^2}{7}\right) \mu_2(s) P_2(\cos\beta) + \left(\frac{8f^2}{35}\right) \mu_4(s) P_4(\cos\beta) + \dots \right\}. \quad (28)$$

<sup>4</sup> Referred to the CUSO lecture: *large-scale structure* delivered by Professor Camille Bonvin in 2016: <http://www.blue-shift.ch/teaching/>





**Figure 1.** The Redshift-Space Distortion effects on the over-density region on different scales. The left panel is the correlation function in the real space while the right panel is that in the redshift space. The dots are galaxies with peculiar velocities denoted by arrows. The top panel is on large scales where the peculiar velocities are small compared to the radius, leading to a squashing effect along the line of sight. The middle panel is the turnaround scale where the peculiar velocity is equal to the radius. The lower panel is on small scales where the peculiar velocities are large enough to produce ‘outflows’. So the correlation function is elongated along the line of sight (i.e., the finger-of-god effect). It is the Figure 2 from [Hamilton \(1998\)](#).

where  $f = \frac{a}{D} \frac{dD}{da}$  is the growth rate of the cosmological perturbation,  $\mu_l$  is a function of galaxy pair separation  $s$ . Furthermore, if we use the linear bias model in Eq. 23 to transfer  $\tilde{\delta}_{gal}$  in Eq. 27 to  $b\tilde{\delta}_{DM}$ , the observed correlation function becomes

$$\xi_{gal,obs} = D^2 \left\{ (b^2 + \frac{2bf}{3} + \frac{f^2}{5})\mu'_0(s) - (\frac{4bf}{3} + \frac{4f^2}{7})\mu'_2(s)P_2(\cos\beta) + (\frac{8f^2}{35})\mu'_4(s)P_4(\cos\beta) + \dots \right\}. \quad (29)$$

We are going to effectively model the galaxy bias using the SubHalo Abundance Matching (SHAM) method. The SHAM bias models are calibrated by the correlation function multipoles of eBOSS galaxies in the redshift space. In our study, we assume a flat universe with  $\Omega_m = 0.31$  and  $\Omega_\Lambda = 1 - \Omega_m = 0.69$ .

## 4. DATA DESCRIPTION

### 4.1. eBOSS Observations

#### 4.1.1. Target Selection

Our observational data are the galaxy samples from the Sloan Digital Sky Survey IV - Extended Baryon Oscillation Spectroscopic Survey (SDSSIV-eBOSS). The SDSS telescope ([Gunn et al. 2006](#)) has a  $5^\circ$  field of view, and it has observed  $7500 \text{ deg}^2$  of the sky in both the North Galactic Cap (NGC) and the South Galactic Cap (SGC) for eBOSS ([Dawson et al. 2016](#)). The telescope has two spectrographs that collect spectra from 1000 fibres at a time. All the fibres are plugged in a plate that covers the focal plane, and each fibre points to a given target, i.e., an eBOSS tracer that has been pre-selected by its photometric properties. The aim of the selection process is to have the desired type of tracer at the expected redshift range. They should also be able to use in high-precision BAO studies. So the selected samples should be statistically uniform and have specific physical properties embodied by their photometric information.

The eBOSS LRG selection criteria include the signal-to-noise ratio selection, the flux limits cut, the colour selection and the star exclusion ([Prakash et al. 2016](#)) using the SDSS imaging data ([Gunn et al. \(2006\)](#)). In particular, the flux

cut

$$19.9 \leq i \leq 21.8 \quad (30)$$

has a lower limit in  $i$  band, aiming at avoiding the BOSS CMASS galaxies. This lower limit of magnitude corresponds roughly to an upper limit of the stellar mass, meaning that the most massive LRGs may have been excluded from the eBOSS LRG samples. Its effect on our LRG SHAM model is discussed in Section 5.1. The colour selection

$$\begin{aligned} r - i &> 0.98, \\ i - z &> 0.625, \end{aligned} \quad (31)$$

is expected to pick out the samples at  $0.6 < z < 1.0$  and the star exclusion

$$r - W1 > 2(r - i) \quad (32)$$

is applied to remove the stars.  $W1$  is the 3.4-microns infrared magnitude from the Wide-Field Infrared Survey Explorer (WISE, Wright et al. (2010); Lang et al. (2014)).

The eBOSS ELGs target selection (Raichoor et al. 2017) is based on the imaging data from the DECam Legacy survey (DECaLS, Dey et al. (2019)). The selection criteria are defined so that samples have clean photometry, OII doublets emissions, and redshift at  $0.6 < z < 1.1$ . The standards in NGC and SGC are slightly different. In NGC, the ideal OII emitter samples have

$$21.825 < g < 22.9. \quad (33)$$

Their redshift range is roughly constrained by

$$\begin{aligned} -0.068 \times (r - z) + 0.457 &< g - r < 0.112 \times (r - z) + 0.773, \\ 0.637 \times (g - r) + 0.399 &< r - z < -0.555 \times (g - r) + 1.901. \end{aligned} \quad (34)$$

In SGC, the selection criteria are:

$$\begin{aligned} 21.825 &< g < 22.825, \\ -0.068 \times (r - z) + 0.457 &< g - r < 0.112 \times (r - z) + 0.773, \\ 0.218 \times (g - r) + 0.571 &< r - z < -0.555 \times (g - r) + 1.901. \end{aligned} \quad (35)$$

#### 4.2. Galaxy Clustering

The eBOSS clustering catalogues consist of the 3D position of each tracer, including its Right Ascension, Declination and redshift (RA, DEC,  $z$ ). The redshift can be converted to the distance given the cosmological parameters following Eq. 9. In addition, four different weights are provided:  $w_{\text{sys}}$ ,  $w_{\text{noz}}$ ,  $w_{\text{CP}}$  and  $w_{\text{FKP}}$ . They account for the correction of the photometric systematics, the redshift failure, fibre collisions and the non-uniform galaxy number density (Ross et al. 2020; Raichoor et al. 2020). The total weight used for clustering measurements is

$$w_{\text{gal}} = w_{\text{sys}} \times w_{\text{noz}} \times w_{\text{CP}} \times w_{\text{FKP}} \quad (36)$$

After counting the weighted galaxy pairs, the correlation functions can be obtained by using Eq 20. The fibre collision effect is particularly important in our SHAM calibration study because it may affect the correlation function in our interested scales. The close-pair weight  $w_{\text{CP}}$  is introduced because the distance between fibres should be larger than  $62''$  due to the physical sizes of the fibres, i.e., the fibre collision. As a result, the number of galaxy pairs with angular separation below  $62''$  is highly underestimated. There are also galaxy pairs that have one galaxy observed but the other locates in the collision region. So the correlation function on small scales (e.g., smaller than 5Mpc/h) is also biased. The traditional solution is to apply the close-pair weight  $w_{\text{CP}}$ :

$$w_{\text{CP}} = \frac{N_{\text{CP}} + N_{\text{good}}}{N_{\text{good}}} \quad (37)$$

where  $N_{\text{CP}}$  is the number of the half-missing tracer pairs and  $N_{\text{good}}$  the number of tracers that have good redshifts (Ross et al. 2020).



The pairwise-inverse-probability (PIP) correction is a better way to account for the missing galaxy pairs. The probability of a galaxy to be observed by the survey is obtained by running multiple fibre assignment simulations. By weighting the observed galaxy pairs with the inverse selection probability, Bianchi & Percival (2017) successfully recovered the two-point correlation functions of simulations on scales where the selection probabilities are not zero. However, galaxy pairs with a separation smaller than  $62''$  might have null selection probability, and PIP weighting cannot be used in this case. Percival & Bianchi (2017) proposed the angular up-weighting scheme to recover the correlation functions on small scales. By combining those two schemes, Mohammad et al. (2020) is able to provide unbiased correlation functions for eBOSS galaxies and quasars on scales down to  $0.1 \text{ Mpc/h}$ . So we apply the PIP-ANG-corrected galaxy pair-counts from Mohammad et al. (2020) to Eq 20, and obtain the correlation function multipoles of eBOSS data for our SHAM models.

#### 4.3. EZmocks

The correlation function measures the universe, i.e., only one realisation under  $\Lambda\text{CDM}$  framework. So alternative methods have been developed to accurately estimate statistical errors. One of them is to use multiple realisations of simulations with the same cosmological parameters. However, N-body simulations are too expensive for covariance matrix estimations, in which they require thousands of realisations. Chuang et al. (2015) proposed a fast methodology to generate mock halo or galaxy catalogues: the Effective Zel'dovich approximation mocks (EZmocks). They rely on the Zel'dovich approximation (Zel'dovich 1970) to generate dark matter density fields at a given redshift. Tracers are populated in the density field using both the deterministic and the stochastic bias models. EZmocks are able to reproduce the two-point and three-point statistics of N-body simulations down to non-linear scales.

In our study, we use the EZmocks for eBOSS tracers Zhao et al. (2020). They are constructed with a flat  $\Lambda\text{CDM}$  fiducial cosmology with  $\Omega_m = 0.307115$ ,  $\Omega_b = 0.048206$ ,  $h = 0.6777$ ,  $\sigma_8 = 0.8225$ ,  $n_s = 0.9611$ . The mocks come from simulations with  $1024^3$  particles in  $5^3 \text{ (Gpc/h)}^3$ . There are 1000 realisations for each tracer and in each galactic cap.

#### 4.4. UNIT Simulation

UNIT is the abbreviation of Universe N-body simulations for the Investigation of Theoretical models from galaxy surveys<sup>5</sup>. It is a large-volume and high-resolution N-body simulation that aims at predict the non-linear effects of galaxy statistics (Chuang et al. 2019). They use the N-body code Gadget (Springel 2005) with  $1^3 \text{ (Gpc/h)}^3$  cubic box and  $4096^3$  particles. The particle mass resolution is  $1.2 \times 10^9 \text{ M}_\odot / \text{h}$ . UNIT uses the fix amplitude and inverse phase technique to significantly reduce the cosmic variance. Its effective volume is 5 times larger than that of some massive surveys like DESI and Euclid. The cosmological parameters are  $\Omega_m = 0.3089$ ,  $h \equiv H_0/100 = 0.6774$ ,  $n_s = 0.9667$ ,  $\sigma_8 = 0.8147$ . The simulation starts at  $a(t) = 0.01$  ( $z = 99$ ) and is evolved to  $a(t) = 1$  ( $z = 0$ ). There are 128 snapshots at different redshift, and we use the ones at  $a = 0.53780$  ( $z = 0.859$ ) for ELGs SHAM and  $a = 0.58760$  ( $z = 0.702$ ) for LRGs SHAM.

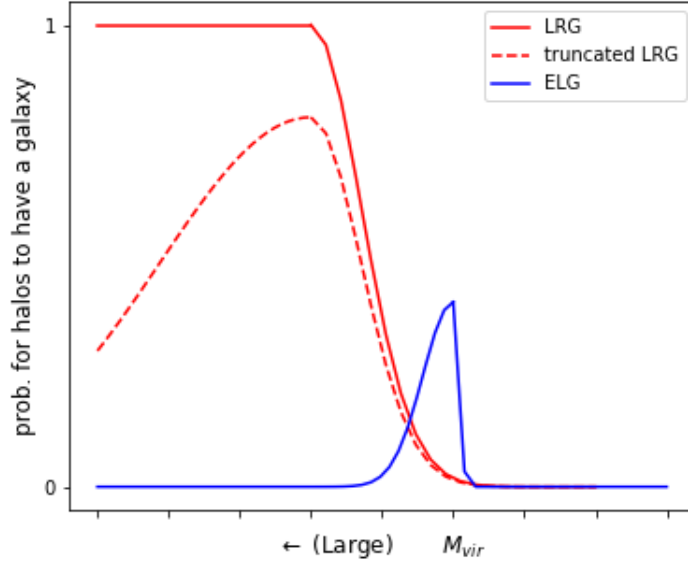
### 5. METHODOLOGY

Our study is to reproduce the eBOSS galaxy clustering with N-body simulation. SubHalo Abundance Matching (SHAM) is employed to realise our galaxy bias hypothesis from the dark matter distribution to the galaxy distribution. The bias hypothesis is a combination of the linear model and the stochastic model. The linear bias implies that luminous (massive) galaxies reside in massive halos. While the stochastic bias is introduced due to effects like the halo merger history and feedback processes (Behroozi et al. 2010). Different bias models will change the clustering of SHAM galaxies as Eq. 29 suggests. By calibrating the bias hypothesis with the observations, we can determine the physical stochastic bias model.

#### 5.1. SubHalo Abundance Matching

SubHalo Abundance Matching is a method to select dark matter halos and assign a galaxy in its centre. The galaxy is expected to have the same peculiar velocity as its host halo. The deterministic bias model means only the most massive halos can host galaxies, but the stochastic process also allows light halos to have one. A simple but physical description of such combined bias model is

<sup>5</sup> <http://www.unitsims.org/>



**Figure 2.** The  $V_{peak}$  probability distribution function (PDF) hypotheses of LRG (the red solid line), the truncated LRG (the red dashed line) and ELG (the blue line).

1. Massive halos have a higher probability of hosting a massive LRG inside, but no star-forming galaxy (ELG) resides there due to some quenching mechanisms(e.g., [Kauffmann et al. 2004](#); [Dekel & Birnboim 2006](#));
2. Low-mass halos cannot hold any galaxy due to their weak gravitational potential;
3. halos with intermediate mass have a certain probability to have a galaxy inside.

We expect that the final LRG and ELG halo probability distributions to follow the red solid line and the blue solid line in Fig. 2 respectively. However, the eBOSS LRG target selection might have removed the most massive LRGs as Section. 4.1.1 shows. Therefore, the LRG  $V_{peak}$  probability distribution used in this work is also truncated in the massive end as the red dashed line of Fig. 2 shows.

Since the halo mass  $M_{vir}$  is not explicitly defined for subhalos, we choose the peak maximum circular velocity over the mass accretion history (i.e.,  $V_{peak}$ ) to complete the SHAM ([Rodríguez-Torres et al. 2016](#)). The galaxy probability distribution can be achieved by the generalised SHAM steps (the original version is from [Rodríguez-Torres et al. \(2016\)](#)):

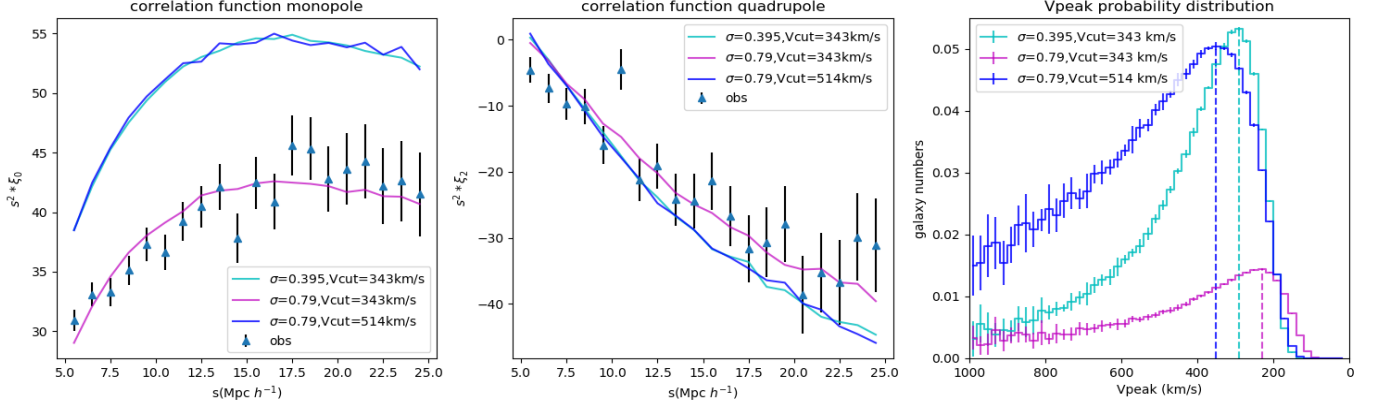
1. Scatter the  $V_{peak}$  with a normal distribution, i.e.,

$$V_{peak}^{scat} = V_{peak} \times (1 + N(0, \sigma)). \quad (38)$$

2. Keep halos with  $V_{peak}^{scat}$  that are below a threshold  $V_{cut}$  to cut the massive end
3. Assign  $N_{gal}$ -th galaxies to the remaining halos that have the largest  $V_{peak}^{cut}$
4. Construct a SHAM galaxy catalogue, the position and peculiar velocity of galaxies are the same as their host halos.

Here,  $N_{gal}$  is the number of galaxies obtained from the UNIT simulation. It is fixed by the effective number density  $n_{eff}$  of galaxies of the observational data as follows:

$$N_{gal} = n_{eff} \times V_{sim} \quad (39)$$



**Figure 3.** The impacts of SHAM parameters  $\sigma$  and  $V_{\text{cut}}$  on the correlation function monopole (*the left panel*), the quadrupole (*the middle panel*) and the  $V_{\text{peak}}$  probability distribution (*the right panel*). The data points are the observation of ELG in SGC, along with its best-fit line in magenta. The blue lines represent the SHAM model with the same  $\sigma$  as the best-fit and a 50% larger  $V_{\text{cut}}$ . The cyan lines are the SHAM model having the same  $V_{\text{cut}}$  as the best-fit but 50% smaller  $\sigma$ .

where  $V_{\text{sim}}$  is the comoving volume of the UNIT simulation, which is 1 (Gpc/h) $^3$ . The effective number density of the eBOSS tracers are evaluated as

$$n_{\text{eff}} = \sqrt{\frac{\int n^2(z) dV}{\int dV}} = \sqrt{\frac{\int n^2(\chi) \chi^2 d\chi}{\int \chi^2 d\chi}} \quad (40)$$

where  $dV = A_{\text{eff}}((\chi + d\chi)^3 - \chi^3)/3 \approx A_{\text{eff}}\chi^2 d\chi$ .  $\chi$  is the comoving distance for an object at redshift  $z$ , and  $d\chi$  is the comoving distance between  $z$  and  $z + dz$ . For ELGs,  $n_{\text{eff}} = 2.93 \times 10^{-4}$  (Mpc/h) $^{-3}$ , while for LRG,  $n_{\text{eff}} = 6.26 \times 10^{-5}$  (Mpc/h) $^{-3}$ .

### 5.1.1. Parameter Impacts

In our SHAM models, the scattering parameter  $\sigma$  represents the stochasticity of the stellar mass–host halo mass relation; the  $V_{\text{cut}}$  for ELGs comes from the quenching effect for the star-forming process, while for LRGs,  $V_{\text{cut}}$  indicates the eBOSS LRG truncation in the massive end. For the same  $\sigma$ , a larger  $V_{\text{cut}}$  means more massive halos are selected, i.e., those that have larger galaxy bias remain. Thus the selected halos (i.e., galaxies) will have a larger monopole and a smaller quadrupole. While for the same  $V_{\text{cut}}$ , a smaller  $\sigma$  leads to the similar effects. This similarity also indicates that  $\sigma$  and  $V_{\text{cut}}$  are degenerate. Fig. 3 shows the influences of parameters on the SHAM correlation functions, and their corresponding probability distributions. The maximum-probability  $V_{\text{peak}}$  (i.e., the typical  $V_{\text{peak}}$ ) of the the larger- $V_{\text{cut}}$  model and smaller- $\sigma$  model are larger than the best-fit, meaning that more massive halos are included as we expected. Comparing to the monopoles ( $\propto b^2$ ), the relatively small differences between the quadrupoles ( $\propto b$ ) has been implied in Eq. 29. Correlation functions of the larger- $V_{\text{cut}}$  model is close to those of the smaller- $\sigma$  model. But their  $V_{\text{peak}}$  probability distributions and the typical  $V_{\text{peak}}$  are significantly different.

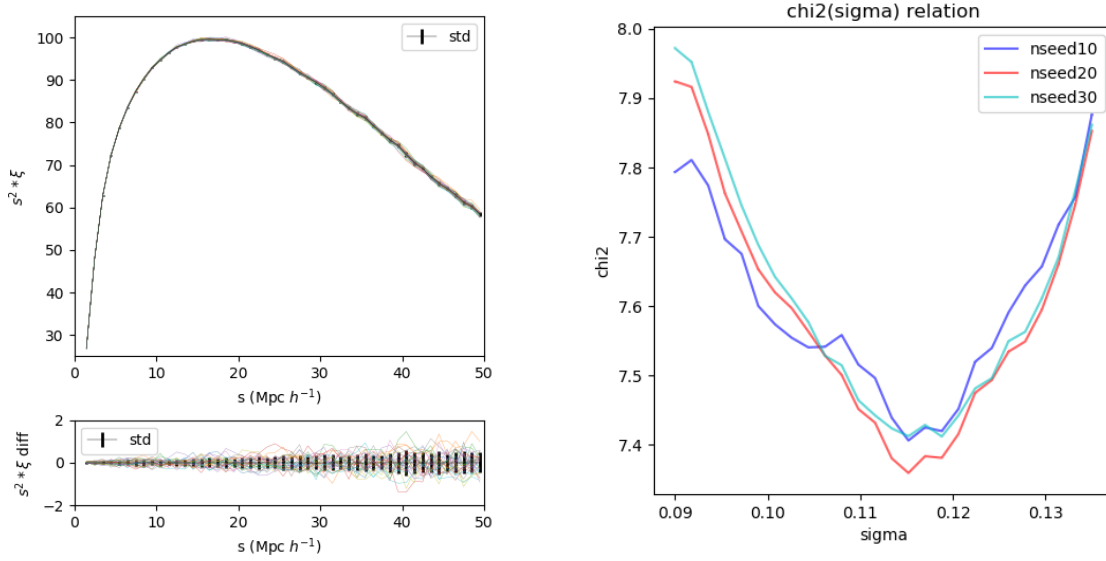
### 5.2. SHAM Model Calibration

In this section, we will explain how to use eBOSS galaxy multipoles in 5-25Mpc/h to calibrate our galaxy bias models. The best SHAM parameter set is obtained by minimising the  $\chi^2$  value (i.e., the maximum log-likelihood). Our  $\chi^2$  for a given parameter set  $\Theta$  is defined as

$$\chi^2(\Theta) = (\xi_{\text{data}} - \xi_{\text{model}}(\Theta))^T \mathbf{C}^{-1} (\xi_{\text{data}} - \xi_{\text{model}}(\Theta)) \quad (41)$$

where  $\xi = [\xi_0, \xi_2, \dots]$  denotes the vector that are composed of the two-point correlation function multipoles.  $\xi_{\text{data}}$  is multipoles measured from the PIP+ANG corrected pair counts and  $\xi_{\text{model}}$  is the average two-point correlation multipoles over several realisations of SHAM galaxy catalogues.  $\mathbf{C}$  is the unbiased covariance matrix (Hartlap et al. 2007):

$$\mathbf{C}^{-1} = \mathbf{C}_s^{-1} \times \frac{N_m - N_s - 2}{N_m - 1} \quad (42)$$



**Figure 4.** *Upper left panel:* Correlation function monopoles measured from 15 realisations of the same SHAM parameters but different random seeds. The black line and error bars indicates their mean and the standard deviation. *Lower left panel:* The difference between the monopole of 15 realisations and their mean value, the error bars are still from the 15 realisations. *Right panel:* the  $\chi^2 - \sigma$  relation variation w.r.t. the number of averaged realisations. The blue line is for 10, the red line is for 20 and the cyan line is for 30 realisations.

where  $N_s$  is total number of bins for  $\xi_{data}$  and  $N_m$  is the number of EZmocks.  $\mathbf{C}_s$  is the covariance matrix of EZmocks, calculated as

$$\mathbf{C}_{s,ij} = \frac{1}{N_m - 1} \sum_{k=1}^{N_m} [\xi_{k_{th}}(s_i) - \bar{\xi}(s_i)][\xi_{k_{th}}(s_j) - \bar{\xi}(s_j)] \quad (43)$$

where  $\xi_{k_{th}}$  is the correlation function measured from the  $k_{th}$  mock, and  $\bar{\xi}(s_i)$  is the average of the mock correlation function in a given distance bin  $s_i$ .

We assume a Gaussian likelihood  $\mathcal{L}(\Theta)$  for our parameters, which is

$$\mathcal{L}(\Theta) \propto e^{-\frac{\chi^2(\Theta)}{2}}. \quad (44)$$

#### 5.2.1. SHAM Code Optimisation

We use `iminuit`, a fast and robust  $\chi^2$  minimiser (`iminuit team` (Accessed: 2018-03-05), James & Roos (1975)) to do a preliminary test of our SHAM code. `iminuit` can estimate the errors of parameters even with non-parabolic  $\chi^2 - \Theta$  relations as long as the relations converge. A simplified SHAM model without the second step (i.e.,  $\sigma$  is the only variable) is employed and we only fit the monopole. We use the Box-Muller transformation in the  $\chi^2$  calculation function to calculate our Gaussian distributed random arrays, with only one uniform random array generated per calibration. We also note the statistical fluctuation brought by the random scattering might affect the precision of the best parameter. The left panel of Fig. 4 shows the monopole dispersion of SHAM galaxies with the same  $\sigma$  but different random seeds. So we average over multiple realisations for one parameter set  $\Theta$  to stabilise the statistical fluctuation. The right panel of Fig. 4 shows how the number of realisations (i.e., nseed) affects the shape of  $\chi^2(\sigma)$  relation. None of them has a parabolic shape near the global minimum. This is possibly caused by the random array fluctuation and this noisy  $\chi^2$  relation can easily trap the minimiser in a local minimum. But the  $\chi^2$  curve of 30 realisations is smoother than 10 and 20 realisations. Considering the coding efficiency and stability, we fix the number of realisations as 30 for `iminuit`  $\chi^2$  minimisation.

#### 5.2.2. Bayesian Analysis

Galaxy	$\sigma$	$V_{\text{cut}}$ (km/s)	$\chi^2$	Reduced $\chi^2$
ELG NGC	$0.513^{+0.443}_{-0.081}$	$268^{+124}_{-30}$	52.296	1.376
ELG SGC	$0.790^{+0.200}_{-0.285}$	$342^{+58}_{-61}$	51.526	1.356
LRG NGC	$0.800^{+0.035}_{-0.056}$	$1167^{+29}_{-63}$	72.785	1.915
LRG SGC	$0.710^{+0.144}_{-0.029}$	$994^{+167}_{-12}$	54.593	1.437

**Table 1.** The maximum-likelihood parameters and their  $\chi^2$  obtained from `pymultinest`. The reduced  $\chi^2$  value is the  $\chi^2$  value divided by 38 with  $N_s = 40$  and  $N_{\text{param}} = 2$ .

`pymultinest` (Buchner et al. 2014) is the python implementation of Multinest (Feroz & Hobson (2008), Feroz et al. (2009), Feroz et al. (2019)), an efficient Nested sampling technique especially for multi-modal posteriors (Skilling 2004). The Bayesian inference defines the posterior probability  $P(\Theta|D)$  as follows (Feroz & Hobson 2008):

$$P(\Theta|D) = \frac{P(D|\Theta)P(\Theta)}{P(D)} \quad (45)$$

where  $D$  is the data,  $P(D|\Theta) \equiv \mathcal{L}(\Theta)$  is the likelihood for a given parameter set  $\Theta$ ,  $P(\Theta)$  is the parameter prior (i.e., the range and distribution of the parameters).  $P(D) \equiv \mathcal{Z} = \int_{\Omega_\Theta} P(D|\Theta)P(\Theta)d\Theta$  is the Bayesian evidence, the normalisation factor of the posterior. During the Monte Carlo sampling, particles walk randomly in the parameter space. Finally, particles will be converged to the vicinity of the maximum-likelihood parameter set. The sampling will stop if the product of the maximum likelihood and the remaining prior volume is smaller than a given value (0.5 by default). The results of `pymultinest` are analysed by its built-in analyser and the posterior distributions will be plotted by `Getdist` (Lewis 2019). We use a uniform  $\sigma$  prior at  $[0,1]$  for both LRG and ELG. As for the  $V_{\text{cut}}$ , LRGs have a uniform prior in  $[800,1200]$  km/s and ELGs have that in  $[100,500]$  km/s.

## 6. RESULTS

Table. 1 concludes the maximum-likelihood parameters with errors, their  $\chi^2$  and reduced  $\chi^2$  provided by results from `pymultinest`. The results from `iminuit` are presented in the following discussion only for reference.

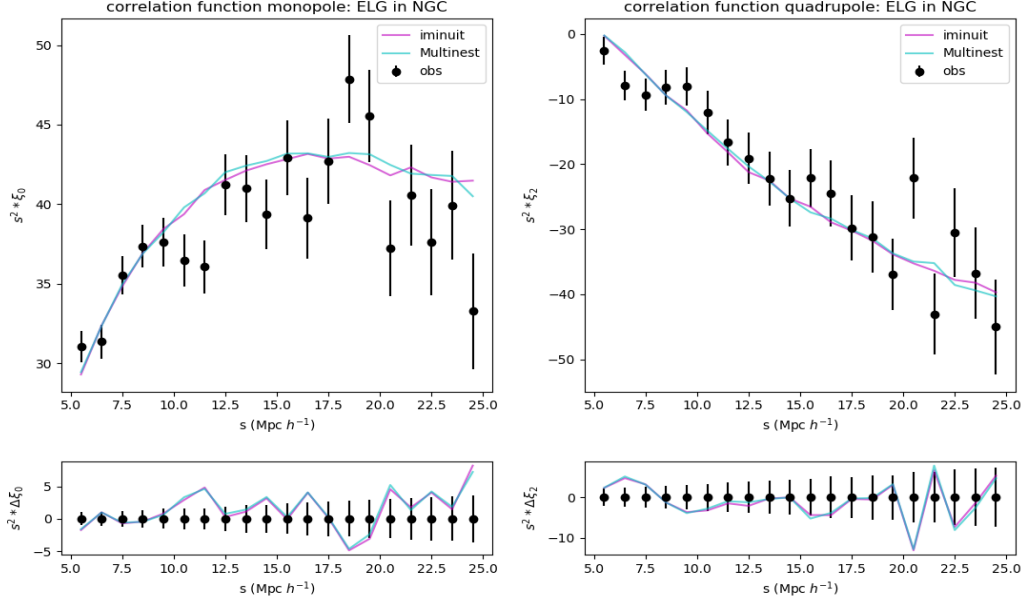
### 6.1. ELG SHAM Model

The effective redshift of eBOSS ELG is  $z_{\text{eff}} = 0.845$ . Given the impact of the growth factor  $D(z)$  on the correlation functions, we choose the UNIT simulation snapshot with the closest redshift  $z = 0.859$  to conduct the SHAM. We average over 20 realisations for `pymultinest` as a trade-off between the efficiency and statistical stability. SHAM ELGs have  $N_{\text{gal}} = 2.93 \times 10^5$  in both NGC and SGC. The best-fit correlation functions in NGC and SGC are presented in Fig. 5 and Fig. 6. `pymultinest` and `iminuit` produce almost the same correlation function multipoles here, which are mostly within  $1\sigma$  error estimated by EZmocks. Fig. 7 and Fig. 8 are the  $V_{\text{peak}}$  PDFs and distribution functions for NGC and SGC. Their shapes are asymmetric peak, consistent with our expectation in Fig. 2. Fig. 9 are the `pymultinest` posterior distributions. The shape and ascending trend of the posteriors show that  $\sigma$  and  $V_{\text{cut}}$  are highly-degenerate as we expect.

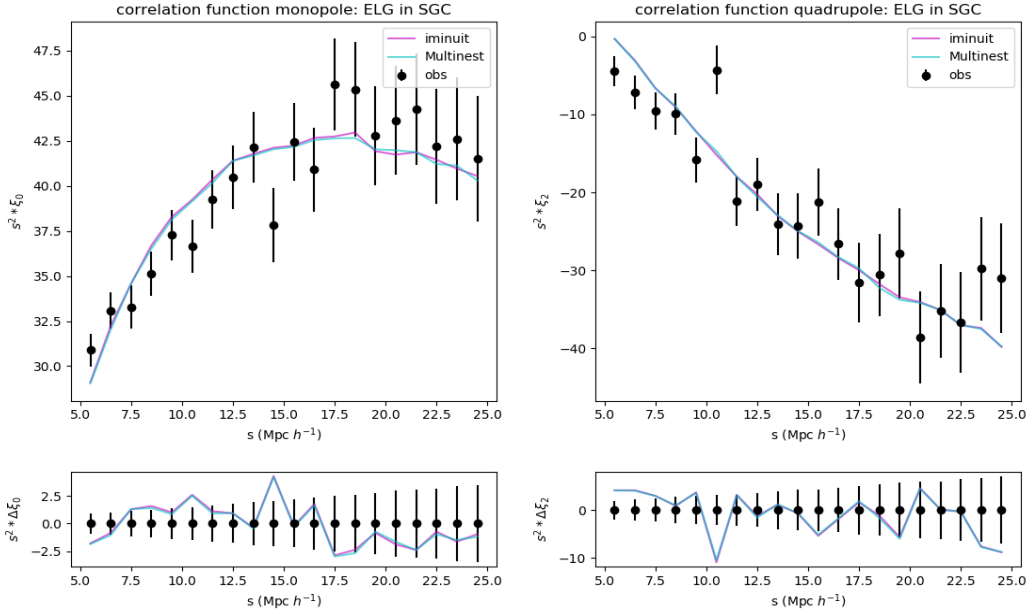
For ELGs in NGC, `iminuit` best-fit parameters are  $\sigma = 0.694$  and  $V_{\text{cut}} = 319$  km/s, within the 68% confidence interval of the `pymultinest` results. `iminuit`  $\chi^2$  value is 55.294, meaning that `pymultinest` fitting is slightly better than it. Despite the large difference on  $\sigma$ , `pymultinest` and `iminuit` have similar correlation function monopoles. This is because the effect of a larger  $\sigma$  to include more low-mass halos is canceled by a larger  $V_{\text{cut}}$  to maintain more massive halos. It is embodied in the  $V_{\text{peak}}$  PDF and distribution (Fig. 7) where `iminuit` has a bit more massive halos but `pymultinest` has a lot more halos with intermediate mass. As a result, their bias are quite similar. The `iminuit` results for ELG in SGC is also within the 68% reliable region of `pymultinest` results. But the  $\chi^2$  is still slightly larger than `pymultinest`. We regard the probability distribution function peak at  $V_{\text{peak}} \approx 230$  km/s as the typical  $V_{\text{peak}}$  value for the host halos of ELGs.

### 6.2. LRG SHAM Model

The effective redshift of eBOSS LRG is  $z_{\text{eff}} = 0.698$ , so UNIT simulation snapshot with the closest redshift  $z = 0.702$  is employed. The number of realisations we use for the LRG SHAM model calibration is the same as that of ELGs.



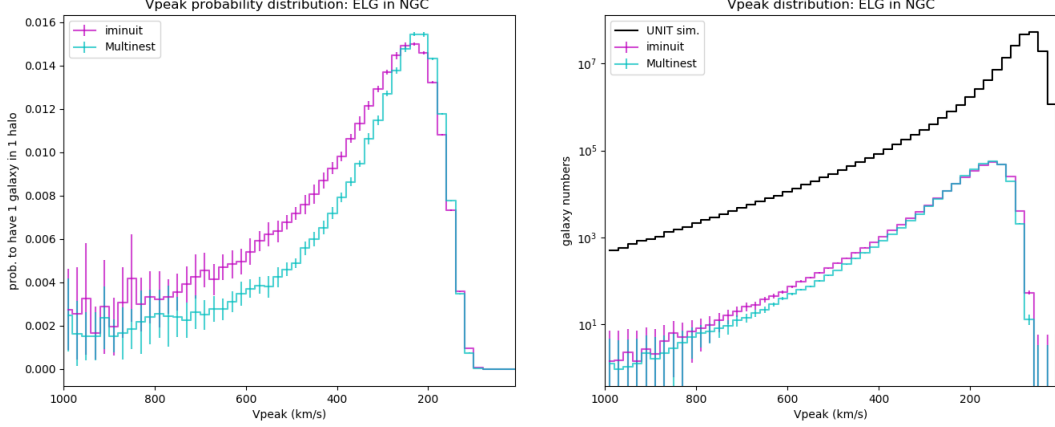
**Figure 5.** The optimal SHAM correlation function monopoles (*upper left panel*) and quadrupoles (*upper right panel*) for ELG in NGC. The black data points are observations and the error bars are the standard deviation of 1000 EZmocks. The cyan lines are the averages of 20 SHAM realisations with the `pymultinest` maximum-likelihood parameters. The magenta lines are those with the `iminuit` best-fit parameters. *lower panels*: The difference between the best-fits and the observational data points for the monopole (*lower left*) and the quadrupole (*lower right*).



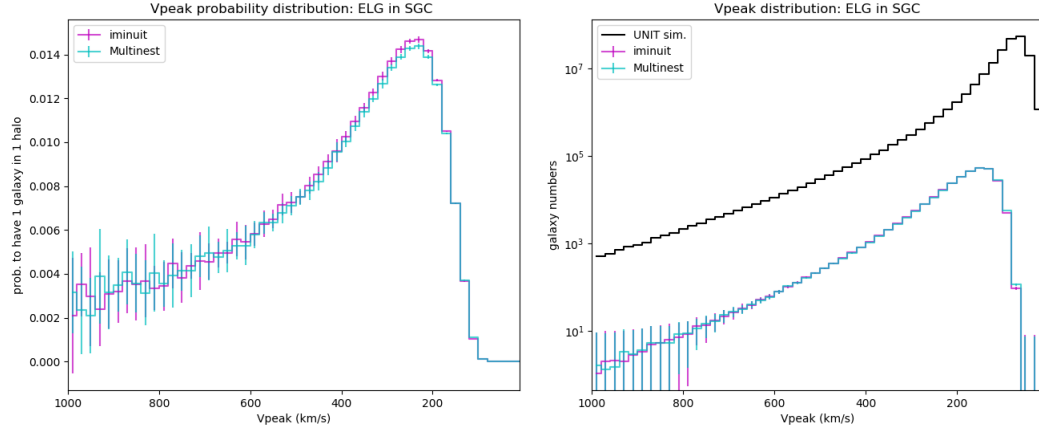
**Figure 6.** The same figure as Fig. 5 for ELG in SGC

$N_{gal} = 6.26 \times 10^4$  for LRGs in both NGC and SGC. Fig. 10 and Fig. 11 are the best-fit correlation functions, where the quadrupole shows a significant deviation from the observation especially in 5-12 Mpc/h. We will discuss the solution to it in Section 13. Fig. 12 and Fig. 13 are the  $V_{peak}$  PDF and distribution functions. Fig. 14 shows the LRG `pymultinest` posteriors in NGC and SGC separately.





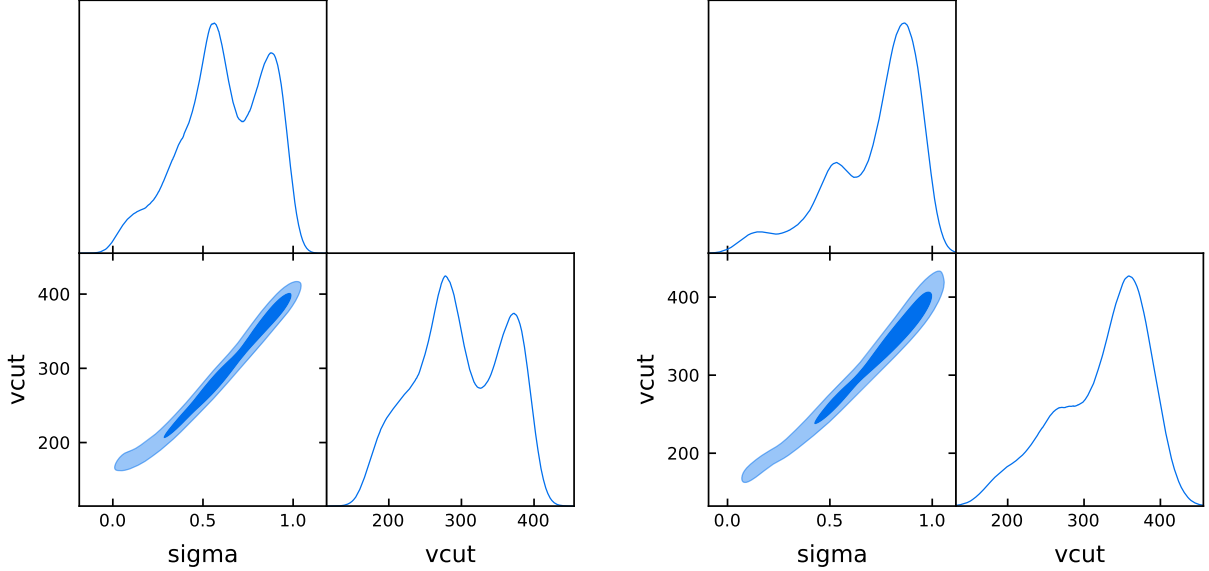
**Figure 7.** *left panel:* The halo  $V_{\text{peak}}$  PDF of the best-fit SHAM for ELG in NGC. The probability is the number of selected halos divided by that of the UNIT halos. The cyan lines and error bars are the averages and standard deviations of 20 SHAM realisations with the `pymultinest` maximum-likelihood parameters. The magenta lines are those with the `iminuit` best-fit parameters. *right panel:* The best-fit  $V_{\text{peak}}$  distribution histogram for ELG in NGC. Besides the cyan and magenta lines with the same meanings as the *left panel*, the extra black histogram is the  $V_{\text{peak}}$  distribution of UNIT halo catalogue. It is the denominator of PDF, and its numerators are the  $V_{\text{peak}}$  distribution of galaxies (i.e., the selected halos) presented in cyan or magenta.



**Figure 8.** The same figure as Fig. 7 for ELG in SGC

The best parameter set of `iminuit` for LRG in NGC is  $\sigma = 0.550$  and  $V_{\text{cut}} = 870$  km/s with  $\chi^2(\Theta) = 85.982$ ; For LRG in SGC, the best parameter set is  $\sigma = 0.566$ ,  $V_{\text{cut}} = 861$  km/s with  $\chi^2(\Theta) = 59.068$ . They are all out of the 68% confidence interval of `pymultinest` parameters. This is probably because `iminuit` fall into local minimum instead of searching for the global minimum. But the correlation functions of `iminuit` are not obviously worse than the `pymultinest`. The correlation function monopoles of `pymultinest` and `iminuit` both agree with the LRG observations. Similar to the explanation for ELG in NGC, a smaller  $\sigma$  means higher bias, but the smaller  $V_{\text{cut}}$  of `iminuit` pull back the bias because it removes more massive halos. The  $V_{\text{peak}}$  PDF in NGC (the left panel of Fig. 12) shows a significant difference between `iminuit` and `pymultinest` in the high-mass end implied by the different  $V_{\text{cut}}$ . But their  $V_{\text{peak}}$  distributions and correlation function monopoles do not illustrate a large difference. So the relatively big  $V_{\text{peak}}$  PDF difference is mainly due to the small number of UNIT halos in that range.

The LRG  $V_{\text{peak}}$  PDF starts to increase after 230 km/s and reach the turning point at around 700 km/s. This characteristic value is around 3 times larger than that of ELG. It means the LRG host halos are around 9 times more massive than the ELG halos. The multi-tracer HOD results produced by Alam et al. (2019) shows that LRG halos mass have a mean value 17 times larger than ELG halos. Another point to notice is the PDF values of LRG is around



**Figure 9.** The posterior contour for ELG in NGC (*left panel*) and for ELG in SGC (*right panel*). The darker blue areas are the 68% ( $1\sigma$ ) confidence region and the light blue areas are the 95% ( $2\sigma$ ) confidence region.

10 times larger than ELG, but the number density of LRG is 3 times smaller than ELG. The reason can be found in the  $V_{\text{peak}}$  distribution: LRG SHAM mainly select heavy halos, the UNIT halos numbers of which are from  $10^3$  to  $10^5$ ; while ELG is assigned in light halos, the UNIT halos numbers of which are one order of magnitude larger. As a result, the ELG fractions are about 10 times smaller.

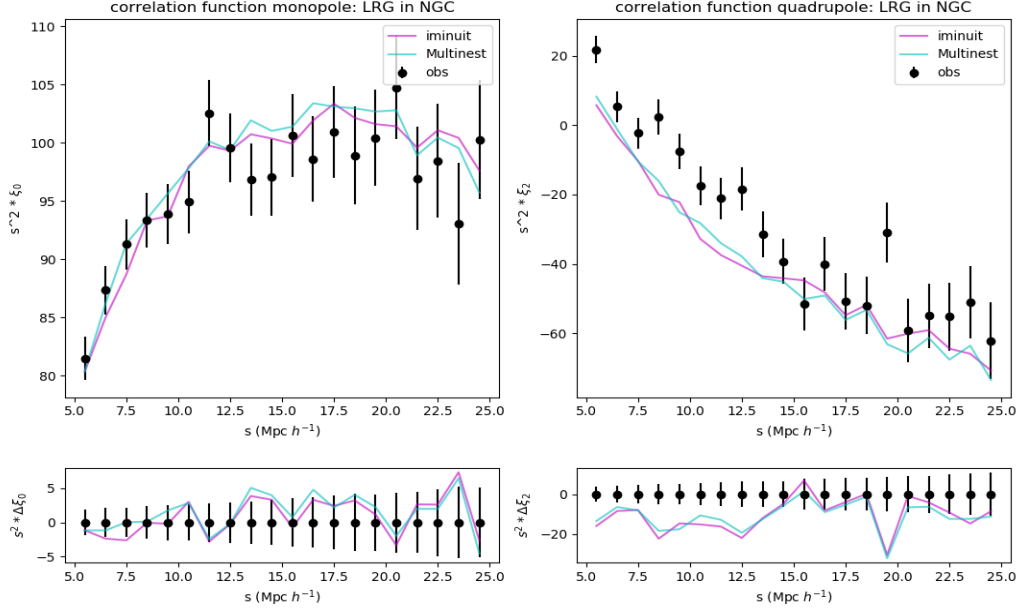
### 6.3. LRG SHAM Model Improvement

Smith et al. (2020) investigates the effect of redshift uncertainty on the correlation function in the redshift space. They specify that impact of redshift uncertainty on the quadrupole is significant but the monopole is small on small scales. This result shows a possible solution to our LRG SHAM problem. Ross et al. (2020) finds the redshift uncertainty of 11556 pairs of eBOSS LRGs at  $0.6 < z < 1.0$  can be well-fitted by a Gaussian distribution. The redshift uncertainty  $\Delta z$  is converted to  $\Delta v = c\Delta z/(1+z)$ , and the distribution has its mean value  $\mu = 1.3$  km/s and its standard deviation  $\sigma = 91.8$  km/s. To model this effect in our SHAM galaxies, we scatter their peculiar velocity  $v_z$  with a Gaussian distribution with  $\mu = 0$  and  $\sigma_v = 91.8$  km/s. By simply applying this correction to the previous best-fit SHAM model of LRG in NGC,  $\chi^2$  decreases from 72.785 to 40.609. The improvement is mainly from the quadrupole, especially in range 5-12 Mpc/h. The monopole is not sensitive to the peculiar velocity, so it does not change significantly. The smeared SHAM model has also reduced the SHAM  $\chi^2$  for LRG in SGC from 54.593 to 37.626. This finding is consistent with the results in Smith et al. (2020). Therefore, we introduce the smearing parameter  $\sigma_v$  in our SHAM, but we keep  $\sigma_v = 91.8$  km/s in our present work. With this parameter, the maximum-likelihood parameters becomes  $\sigma = 0.806$  and  $V_{\text{cut}} = 1170$  km/s, within the  $1\sigma$  confidence interval of the previous LRG NGC results. The  $\chi^2$  decreases from 72.785 to 33.910. The smeared LRG SHAM correlation functions are presented in Fig. 15. In contrast, the redshift uncertainty of 11967 pairs of ELGs is represented by a Gaussian distribution with its dispersion  $\sigma_v = 21.3$  km/s. It is much smaller than the dispersion of LRGs. Because ELGs use the emission lines that are narrower than the absorption lines of LRGs for the redshift determination, making the ELG redshift measurements more robust. Our follow-up study will investigate the ELG SHAM with redshift uncertainty as well.

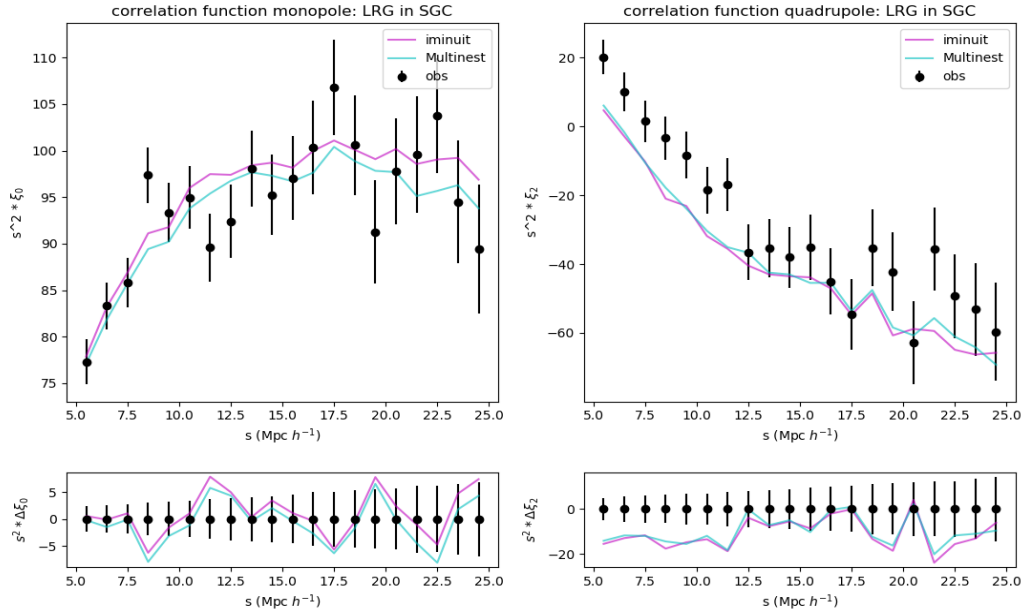
## 7. DISCUSSION

In addition to the redshift uncertainty, the amplitude of quadrupole can also be affected by the redshift (i.e.,  $D(a)$  and  $f(a)$ ) and the galaxy bias ( $b$ ) as Eq. 29 shows. So we investigate their effects on the SHAM model in this section.

The present LRG samples have a large redshift span ranging from 0.6 to 1.0. The correlation functions in different redshift range are not the same due to the evolution of large scale structure and target selection effects. A reliable SHAM model should also fit well in separated redshift bins. Additionally, the discrepancy in quadrupole might be solved



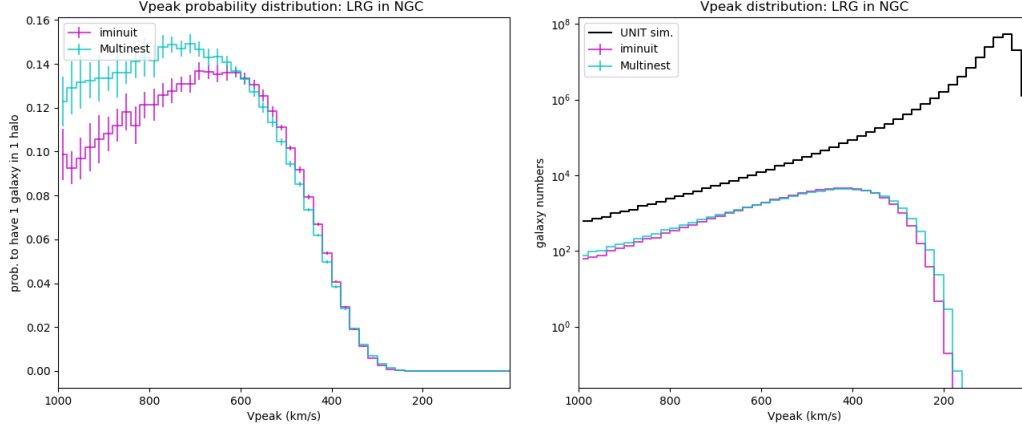
**Figure 10.** The same figure as Fig. 5 for LRG in NGC



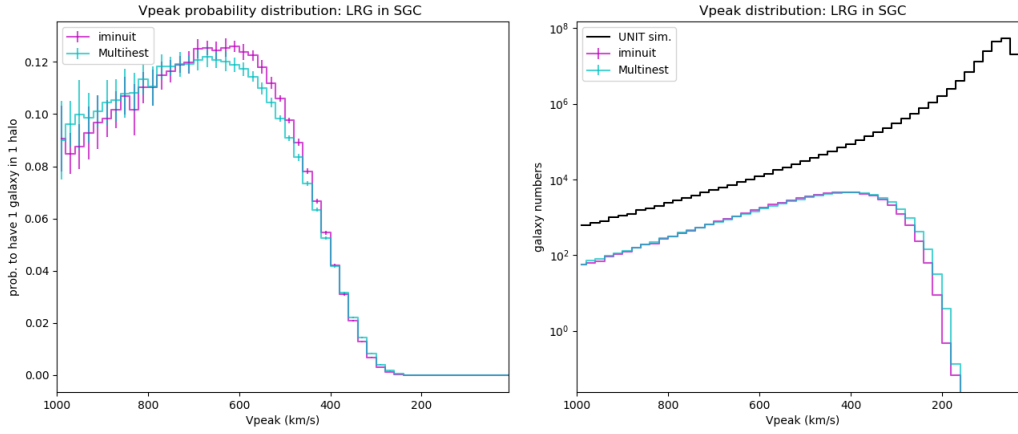
**Figure 11.** The same figure as Fig. 5 for LRG in SGC

in this case. So we plot the 2PCF in different redshift bins as well as the best-fit 2PCF of the LRG SHAM. We intend to find a 2PCF that is near the best-fit quadrupole. However, Fig. 16 shows that the correlation function evolution is significant in monopoles but not in quadrupoles, i.e., all of the 2PCF observation is away from the LRG SHAM model. Although it does not help to reconcile the quadrupole discrepancy, the SHAM models in different redshift bins are still worth trying. Because we can use the smeared LRG SHAM model to study the redshift uncertainty in different redshift range.

Finally, we can also change the LRG  $V_{\text{peak}}$  probability distribution hypothesis (the bias model) to fit the quadrupole. Favole et al. (2016) introduces two sub-types of LRG in their Halo Occupation Distribution study for BOSS. They have



**Figure 12.** The same figure as Fig. 7 for LRG in NGC

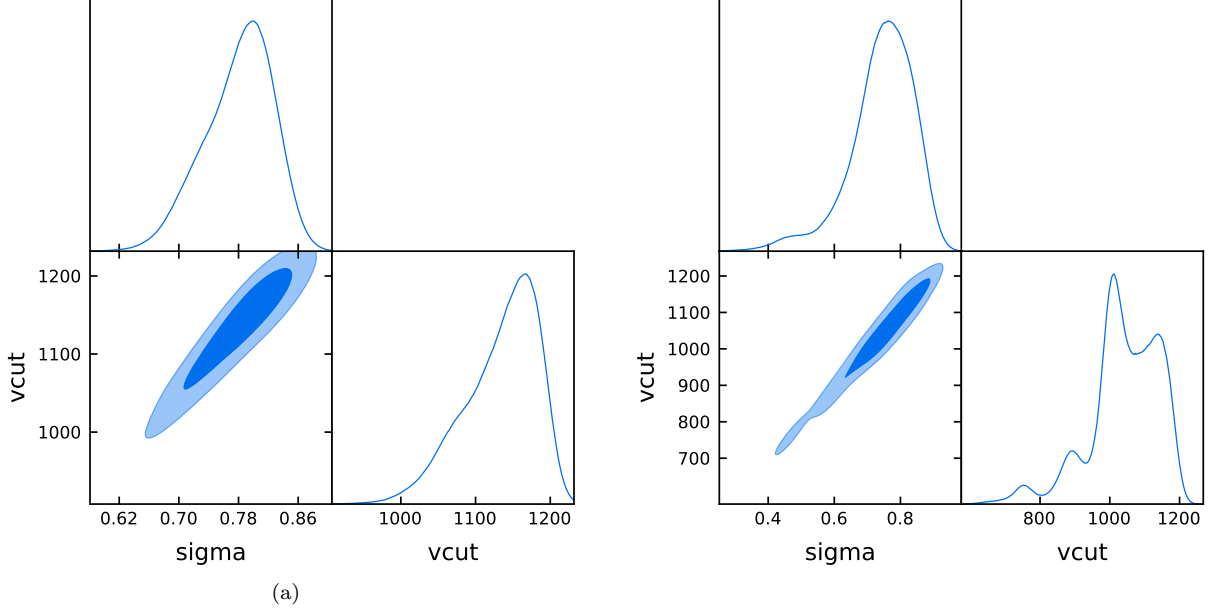


**Figure 13.** The same figure as Fig. 7 for LRG in SGC

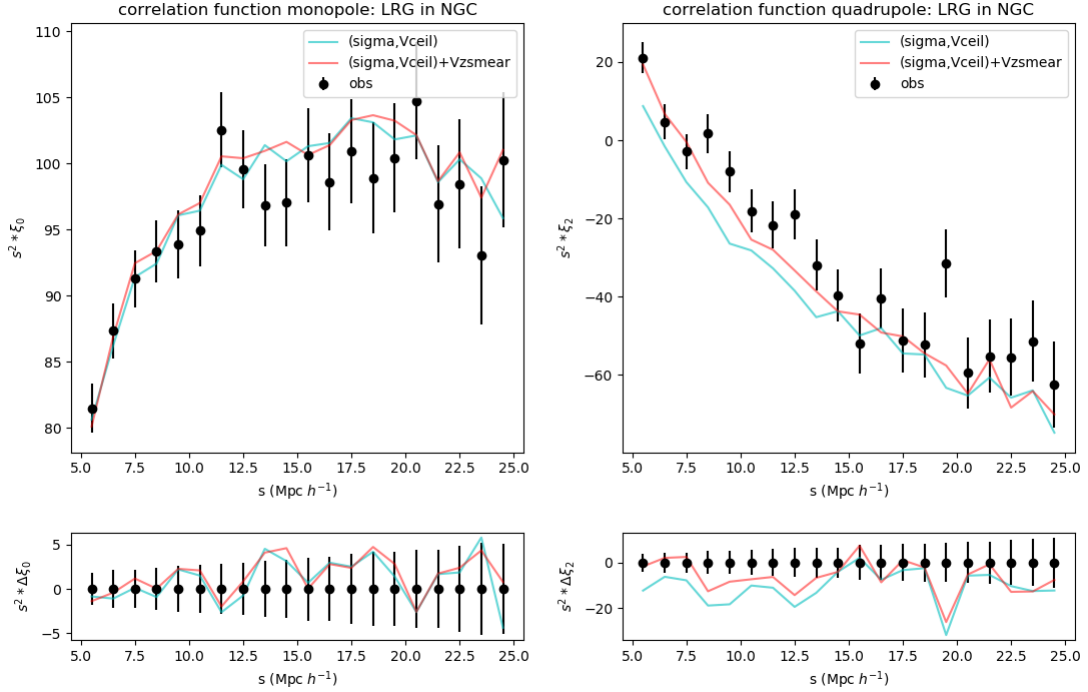
shown that the monopoles of those two subtypes are significantly different. eBOSS LRGs might also have different species. So we shall implement a bi-model SHAM for LRGs, in which we expect the  $N_1$  LRGs reside in the most massive halos and  $(N_{\text{gal}} - N_1)$  LRGs are hosted by relatively low-mass halos. As Fig. 17 shows, the model with  $N_1 = 10000$  has a larger monopole in all the fitting range and quadrupole in 5-8 Mpc/h than  $N_1 = 1000$  and  $N_1 = 0$  (i.e., the previous SHAM). However, the discrepancy for quadrupole extends to 8-25 Mpc/h. This method is another way of playing with the galaxy bias. So a larger  $N_1$  (i.e., a larger proportion of large bias galaxies) has the a similar effect as increasing the  $V_{\text{cut}}$  or decreasing  $\sigma$  as Fig. 3 illustrates. But like  $\sigma$  and  $V_{\text{cut}}$ , the additional  $N_1$  parameter cannot achieve the expected improvement for the LRGs in quadrupoles.

## 8. CONCLUSION

The modified Halo Abundance Matching method is applied to the UNIT N-body simulation (the original SHAM is proposed by Rodríguez-Torres et al. (2016)). The target is to resume the clustering of eBOSS LRGs and ELGs in 5-25 Mpc/h. The galaxy probability distributions in dark matter halos (see Fig. 2) are constructed considering both the physical process and the target selection effect (for LRG only). Such distribution functions are realised by the Halo Abundance Matching method, i.e., scattering ( $\sigma$ ) and selecting ( $V_{\text{cut}}$ ) the  $V_{\text{peak}}$  of halos (including subhalos) from the UNIT halo catalogue and assigning galaxies in their centre. By fitting the two-point correlation multipoles (monopole and quadrupole) of SHAM ‘galaxies’ to the observation, we obtain the best-fit PDF parameters and their errors. Those PDFs will be used in the future multi-tracer SHAM studies.

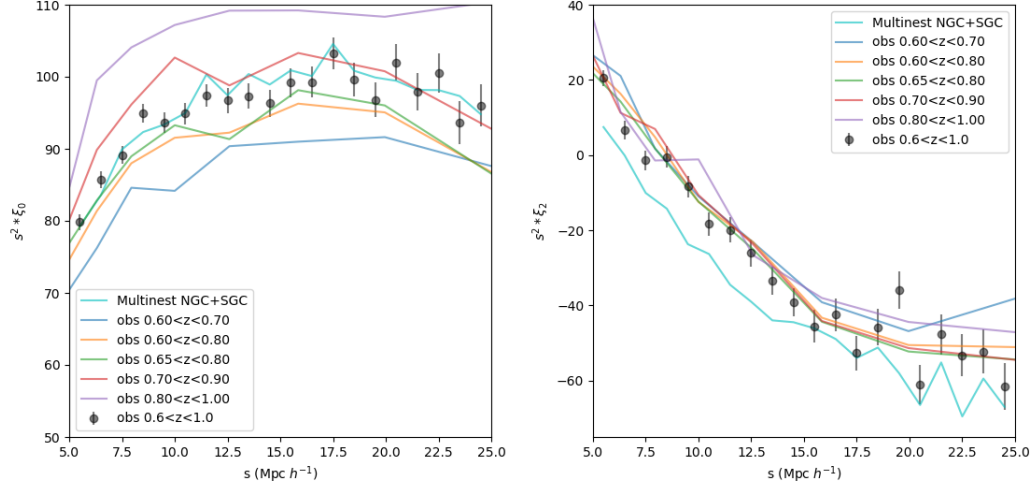


**Figure 14.** The same figure as Fig. 9 for LRG in NGC (*left panel*) and LRG in SG (*right panel*).

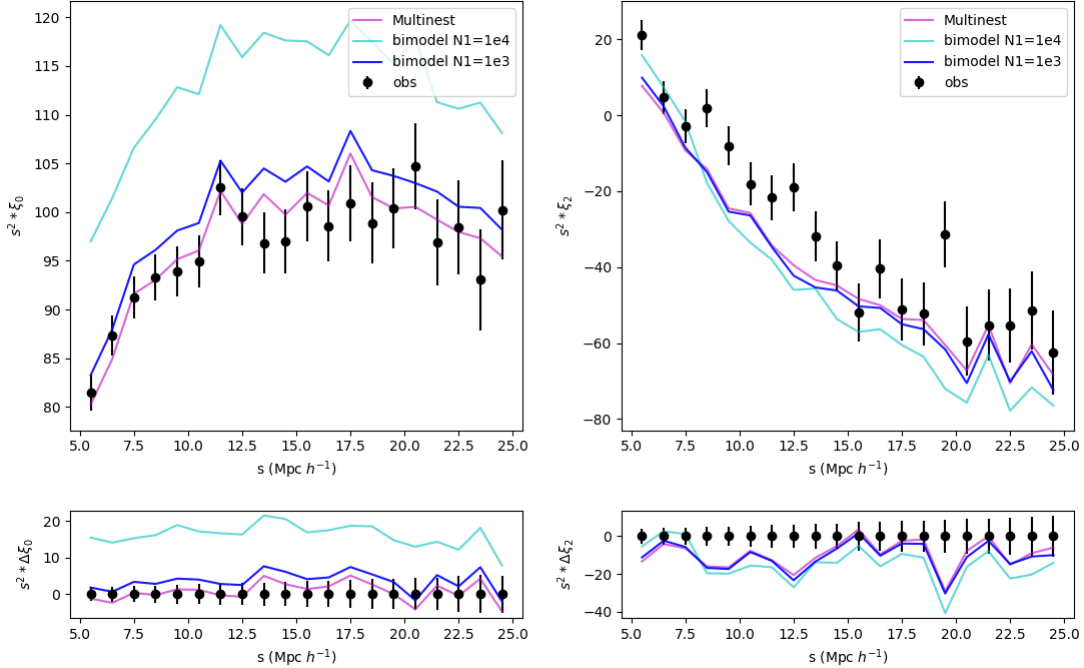


**Figure 15.** The correlation function monopole (*left panel*) and quadrupole (*right panel*) of LRG in NGC. The black dots with error bars are PIP corrected observation, the cyan curves are the best-fit SHAM multipoles and the blue curves are the smeared best-fit SHAM multipoles.

The effective number density and the effective redshift of the eBOSS ELG are  $n_{\text{eff}} = 2.93 \times 10^{-4} / \text{Mpc}^3$ ,  $z_{\text{eff}} = 0.845$ , the corresponding SHAM ELG number and UNIT simulation redshift are  $N_{\text{gal}} = 2.93 \times 10^5$  and  $z_{\text{sim}} = 0.859$ . The maximum likelihood parameters for ELG in NGC are  $\sigma = 0.513^{+0.443}_{-0.081}$  and  $V_{\text{cut}} = 268^{+124}_{-30}$  km/s, while in SG the results are  $\sigma = 0.790^{+0.200}_{-0.285}$  and  $V_{\text{cut}} = 342^{+58}_{-61}$  km/s. Their fittings are almost within  $2\sigma$  errors provided by the mocks.



**Figure 16.** The correlation function monopole (*left panel*) and quadrupole (*right panel*) of LRG NGC+SGC in different redshift bins. The multipoles of HAM models are also the combination of the best-fit in NGC and SGC.



**Figure 17.** The correlation function monopoles (*left panel*) and quadrupoles (*right panel*) of LRG in NGC. The magenta lines are the best-fit HAM model, the blue lines are the HAM bimodel with the parameters of the best-fit and  $N_1 = 1000$  and the cyan lines are the best-fit with  $N_1 = 10000$ .

The ELG  $V_{\text{peak}}$  PDFs both have single-peak shapes with their peaks near 230 km/s, the typical  $V_{\text{peak}}$  for halos that host a galaxy inside.

For eBOSS LRG,  $n_{\text{eff}} = 6.26 \times 10^{-5} / \text{Mpc}^3$  and  $z_{\text{eff}} = 0.698$ , so  $N_{\text{gal}} = 6.26 \times 10^4$  and  $z_{\text{sim}} = 0.702$ . In NGC, the maximum likelihood results are  $\sigma = 0.800^{+0.035}_{-0.056}$  and  $V_{\text{cut}} = 1167^{+29}_{-63}$  km/s and in SGC the results are  $\sigma = 0.710^{+0.144}_{-0.029}$  and  $V_{\text{cut}} = 994^{+167}_{-12}$  km/s. The large  $V_{\text{cut}}$  values lead to the upside-down tick shape of the LRG PDF and a larger typical  $V_{\text{peak}}$  around 700 km/s. To solve the LRG SHAM problem for quadrupole in 5-12 Mpc/h, we take into account the redshift uncertainty of LRG. By smearing the peculiar velocity of SHAM LRGs, we successfully improve the performance of LRG SHAM model in quadrupoles. We also discuss other possible solutions by considering the



correlation function evolution and the LRG bi-model. Although they do not show a significant improvement in the quadrupole, SHAM should still be tested in different redshift bins.

We also compare the behaviors of the  $\chi^2$  minimisation python package `iminuit` to the Multinest python implementation `pymultinest`. The best-fit values of `iminuit` can go out of the 68% confidence intervals of `pymultinest`. `pymultinest` always produces parameter sets that have smaller  $\chi^2$  than `iminuit`.

## REFERENCES

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, 182, 543
- Alam, S., Peacock, J. A., Kraljic, K., Ross, A. J., & Comparat, J. 2019, 16, 1
- Alam, S., Ata, M., Bailey, S., et al. 2017, *MNRAS*, 470, 2617
- Alam, S., Ata, M., Bailey, S., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 2617
- Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, *Astrophys. J.*, 717, 379
- Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, *ApJ*, 717, 379
- Berlind, A. A., & Weinberg, D. H. 2002, *ApJ*, 575, 587
- Bianchi, D., & Percival, W. J. 2017, *MNRAS*, 472, 1106
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, 564, A125
- Campbell, D., van den Bosch, F. C., Padmanabhan, N., et al. 2018, *Mon. Not. R. Astron. Soc.*, 477, 359
- Chuang, C. H., Kitaura, F. S., Prada, F., Zhao, C., & Yepes, G. 2015, *Mon. Not. R. Astron. Soc.*, 446, 2621
- Chuang, C. H., Yepes, G., Kitaura, F. S., et al. 2019, *Mon. Not. R. Astron. Soc.*, 487, 48
- Chuang, C.-H., Yepes, G., Kitaura, F.-S., et al. 2019, *MNRAS*, 487, 48
- Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, *Astrophys. J.*, 647, 201
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, 145, 10
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, 151, 44
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, 368, 2
- DESI Collaboration. 2016, DESI Final Design Report Part I: Science, Targeting, and Survey Design, Tech. rep. <https://desi.lbl.gov/tdr/>
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168
- Eisenstein, D. J., Seo, H.-J., & White, M. 2007, *ApJ*, 664, 660
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *ApJ*, 633, 560
- Favole, G., McBride, C. K., Eisenstein, D. J., et al. 2016, *MNRAS*, 462, 2218
- Feroz, F., & Hobson, M. P. 2008, *Mon. Not. R. Astron. Soc.*, 384, 449
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, 398, 1601
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, *The Open Journal of Astrophysics*, 2, 10
- Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, *AJ*, 131, 2332
- Guo, H., Zheng, Z., Zehavi, I., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 2398
- Hamilton, A. J. S. 1992, *The Astrophysical Journal*, 385, L5
- Hamilton, A. J. S. 1998, *Astrophysics and Space Science Library*, Vol. 231, *Linear Redshift Distortions: a Review*, ed. D. Hamilton, 185
- Hartlap, J., Simon, P., & Schneider, P. 2007, *Astron. Astrophys.*, 464, 399
- iminuit team. Accessed: 2018-03-05, iminuit - A Python interface to Minuit, <https://github.com/scikit-hep/iminuit>, ,
- James, F., & Roos, M. 1975, *Computer Physics Communications*, 10, 343
- JING, Y. P., MO, H. J., & G.BORNER. 1998, *Astrophys. J.*, 10, 1
- Kaiser, N. 1984, *ApJL*, 284, L9
- Kaiser, N. 1987, *Monthly Notices of the Royal Astronomical Society*, 227, 1
- Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, *MNRAS*, 353, 713
- Kerscher, M., Szapudi, I., & Szalay, A. S. 2000, *The Astrophysical Journal*, 535, L13
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, *Astrophys. J.*, 609, 35
- Landy, S., & Alexander, S. 1993, *the astropysical journal*, 412, 64
- Lang, D., Hogg, D. W., & Schlegel, D. J. 2014, *arXiv e-prints*, arXiv:1410.7397
- Lewis, A. 2019, *arXiv e-prints*, arXiv:1910.13970
- Mohammad, F. G., Percival, W. J., Seo, H.-J., et al. 2020, *MNRAS*
- Nuza, S. E., Sánchez, A. G., Prada, F., et al. 2013, *Mon. Not. R. Astron. Soc.*, 432, 743

- Peacock, J. A., & Smith, R. E. 2000, *Mon. Not. R. Astron. Soc.*, 318, 1144
- Peebles, P. J. E., & Hauser, M. G. 1974, *The Astrophysical Journal Supplement Series*, 28, 19
- Percival, W. J., & Bianchi, D. 2017, *MNRAS*, 472, L40
- Planck Collaboration. 2019, [arXiv:arXiv:1807.06209v2](https://arxiv.org/abs/1807.06209v2)
- Prakash, A., Licquia, T. C., Newman, J. A., et al. 2016, *Astrophys. J. Suppl. Ser.*, 224, 34
- Raichoor, A., de Mattia, A., Ross, A. J., et al. 2020, submitted
- Raichoor, A., Comparat, J., Delubac, T., et al. 2017, *Mon. Not. R. Astron. Soc.*, 471, 3955
- Rodríguez-Torres, S. A., Chuang, C. H., Prada, F., et al. 2016, *Mon. Not. R. Astron. Soc.*, 460, 1173
- Ross, A., Bautista, J., Tojeiro, R., et al. 2020, submitted
- Sinha, M., & Garrison, L. 2019, in *Software Challenges to Exascale Computing*, ed. A. Majumdar & R. Arora (Singapore: Springer Singapore), 3–20
- Sinha, M., & Garrison, L. H. 2020, *MNRAS*, 491, 3022
- Skilling, J. 2004, in *American Institute of Physics Conference Series*, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405
- Smith, A., Burtin, E., Hou, J., et al. 2020, submitted
- Springel, V. 2005, *MNRAS*, 364, 1105
- Tinker, J. L., Cao, J., Alpaslan, M., et al. 2019, 21, 1
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
- Zel’dovich, Y. B. 1970, *A&A*, 5, 84
- Zhao, C., et al. 2020, submitted