

# Modelling Galaxy Velocity Distributions in FLAMINGO

MSc thesis

Sowmya Parthasarathy



Universiteit  
Leiden  
Observatory

# Modelling Galaxy Velocity Distributions in FLAMINGO

A thesis submitted in partial fulfillment  
of the requirements for the degree of

Master of Science  
in  
Astronomy

Author:  
Student ID:  
Supervisor:  
Second corrector:  
Project duration:

Sowmya Parthasarathy  
s3908739  
Dr. M.P. van Daalen  
Prof. Dr. M. T. Kriek  
September 2024 – June 2025

Cover image: Andrew Pontzen and Fabio Governato / Wikimedia Commons (CC BY 2.0)

Template style: Thesis style by Richelle F. van Capelleveen  
Template licence: Licenced under CC BY-NC-SA 4.0



Universiteit  
Leiden  
Observatory



# Abstract

Redshift-space distortions (RSD) such as the Fingers-of-God effect arise due to peculiar velocities of galaxies. In particular, the Fingers-of-God effect arises due to the random motions of sub-halos within larger halo groups. Accurate redshift-space distortion modelling requires well-describing models for the velocity difference distributions, which are hard to derive from observations. Traditional approaches often rely on simplified, scale-independent analytical functions that fail to accurately model the observed velocity difference distribution and redshift-space clustering statistics. In this study, we attempt to improve upon existing models for the velocity difference distributions by using the hydrodynamical, cosmological simulation FLAMINGO. We model the velocity difference distributions by decomposing the contributions to clustering statistics like the power spectrum into one-halo and two-halo contributions. We find that a closed-form, two-component Gaussian mixture model captures the observed behaviour of the velocity difference distribution at fixed scales dominated by the Fingers-of-God effect. We also observe that the model effectively captures the observed behaviour of the underlying distribution across multiple halo mass bins, thereby providing a unified description of the velocity difference. We extend the model further to also include scale dependence such that it encodes clustering information, making the calculation of redshift-space clustering statistics straightforward. Compared to standard streaming-model PDFs, which mix random motions of sub-halos with the coherent infall of halo groups while modelling, our model is both computationally simple and physically motivated.

# Acknowledgements

This thesis would not have been possible without the help and support of my supervisor, Dr. Marcel van Daalen. I would like to express my gratitude for the many individual and group meetings that helped me during my project, and I have learned a lot from everyone. I would also like to thank my family and friends for always being there for me.

I dedicate this thesis to my late paternal grandfather and late maternal grandmother, M. Govindan and L.Mohanam, both of whom I lost in this past year and a half. It has been really hard, and I hope you both know that I did my best. I would not be the person I am today without either of you, and I hope I make you both proud.

# Contents

<b>Abstract .....</b>	<b>iv</b>
<b>Acknowledgements .....</b>	<b>v</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 Large Scale Structure	1
1.2 Redshift-Space Distortions	3
1.3 FLAMINGO simulations	6
1.4 Goals of this work	8
1.5 Structure	8
<b>2 Methods .....</b>	<b>9</b>
<b>3 One-Halo Distributions .....</b>	<b>13</b>
3.1 Test of Common Analytical Models - Scale Independent	13
3.2 Analytical Scale-dependent Modelling	19
3.2.1 Conditional modelling .....	19
3.2.2 Classification of sub-halos .....	28
3.2.3 Joint modelling .....	31

<b>4</b>	<b>Two-halo distributions .....</b>	<b>36</b>
<b>5</b>	<b>Conclusion .....</b>	<b>44</b>
<b>A</b>	<b>Appendix .....</b>	<b>47</b>
A.1	Conditional Model Fits	47
	<b>Bibliography .....</b>	<b>60</b>

# 1 . Introduction

## 1.1 Large Scale Structure

The lofty goal of understanding the formation and evolution of the universe is the cornerstone of modern cosmology. The examination of the large-scale structure of the universe and its characteristics can perhaps be traced back to the first efforts to cataloguing fuzzy ‘nebulae’ by Charles Messier in the late 18th century, Herschel in 1864, and Dreyer in 1888 (Mo et al., 2010). This led to a debate of the nature of the ‘nebulae’, dubbed the *Great Debate*, as people were unsure if the ‘nebulae’ were objects inside the Milky Way or ‘island universes’ just like our galaxy (Cimatti et al., 2020). This problem was resolved in 1925 by Edwin Hubble’s discovery that the ‘nebulae’ were extragalactic using Cepheids to estimate the distances of the ‘nebulae’, which in turn started the study of extragalactic objects.

In addition to identifying external galaxies, Hubble also contributed to understanding the nature of the Universe itself (Hubble, 1926). He showed that the velocity of external galaxies were receding from us. The recession velocities of galaxies, also called Hubble velocities  $v_H$ , were found to be proportional to their distances  $d$ . This is now known as the Hubble-Lemaître law:  $v_H = H_0 d_p$ , where  $d_p$  is the proper distance and  $H_0$  is the Hubble constant. This led to the revelation of an expanding Universe. Moreover, Hubble also took the first steps in identifying the extent of homogeneity on larger scales. Using a catalogue of 400 ‘nebulae’, he found that it was mostly uniform on large scales (Hubble, 1926). A larger sample of bright galaxies in the Shapley-Ames catalogue was used by Hubble in 1934 to show that on smaller scales, the universe appeared to be inhomogeneous (Coil, 2013). Simultaneously, advances in the theoretical work, especially that of Albert Einstein on the theory of general relativity in 1916, helped construct mathematical models of the Universe itself. Subsequent breakthroughs, such as identifying the accelerated expansion of the universe, helped in developing modern cosmology.

Now, one of the fundamental assumptions to describe the nature of homogeneity observed is the *statistical cosmological principle*, that states that the universe is statistically homogeneous and isotropic (Mo et al., 2010). Typically, we start by considering the cosmic density field  $\rho(\vec{x}, t)$ , generated by random

processes. The density contrast field can then be used describe the amplitude of density fluctuations, defined as

$$\delta(\vec{x}, t) \equiv \frac{\rho(\vec{x}, t) - \bar{\rho}}{\bar{\rho}}$$

Here, the mean density  $\bar{\rho}$  only depends on time due to the statistical cosmological principle. If the universe were perfectly homogeneous and isotropic, structures like galaxies would not have formed. So, the *statistical* cosmological principle can account for structure formation caused by perturbations of the initial cosmic density field. Currently, the formation of structures is generally understood as inflation, the initial rapid expansion phase on the universe, producing macroscopic density perturbations. In an expanding universe dominated by matter, perturbations grow with time. Regions with a higher initial density than the mean density will pull objects from its surroundings towards itself, increasing its density further. Conversely, regions with a lower initial density will become more underdense as matter is pulled away from it. This mechanism is gravitational instability, playing a vital role in the formation of large scale structure. The evolution of the perturbations depends on the cosmological epoch dictated by the properties of different components, such as matter and dark matter in the cosmological model.

Standard inflation theory predicts Gaussian initial perturbations of the density field (Mo et al., 2010). The probability for  $\delta$  at  $n$  different points in space is then an  $n$ -variate Gaussian, and the covariance matrix is  $C_{ij} = \langle \delta_i \delta_j \rangle = \xi(r_{ij})$ . The correlation function  $\xi(r_{ij})$ , therefore completely specifies the Gaussian random field, especially in the linear regime  $|\delta| \ll 1$ . In the non-linear regime  $|\delta| > 1$ , density contrast field becomes non-Gaussian, so it is no longer completely specified by  $\xi(r_{ij})$ . The two-point correlation function of the density contrast

$$\xi(|\vec{r}_1 - \vec{r}_2|) = \langle \delta(\vec{r}_1) \delta(\vec{r}_2) \rangle$$

only depends on the distance between two objects or the amplitude of  $\vec{r}_2 - \vec{r}_1$  due to the statistical cosmological principle. Typically, the two-point correlation function is measured by using observations of distribution of tracers like galaxies (Coil, 2013).

Focusing on tracers like galaxies that have a discrete distribution, and using similar methods as that of the continuous density contrast field, the number density field of galaxies in the universe can be taken as  $n(|\vec{r}|)$  (Mo et al., 2010). In an infinitesimally small volume  $dV$  such that it can only contain zero or one galaxy and never more, i.e., the occupation number is either 0 or 1. The probability of finding a galaxy in that volume is  $\bar{n}dV$ , where  $\bar{n}$  is the mean of the number density field that only depends on time due to the statistical cosmological principle. Now, if one thinks about the probability  $P_{12}$  of finding an object in a volume  $dV_1$  with a center at  $\vec{r}_1$  and another in  $dV_2$  that has a center  $\vec{r}_2$ ,

$$P_{12} = \langle n(\vec{r}_1) n(\vec{r}_2) \rangle dV_1 dV_2 = \bar{n}^2 [1 + \xi(|\vec{r}_1 - \vec{r}_2|)]$$

If the process of 'assigning' galaxies in different infinitesimal volumes across space is a random Poisson process, the probability would have been  $P_{12,\text{Poisson}} = \bar{n}^2 dV_1 dV_2$ , so the two-point correlation function  $\xi(|\vec{r}_1 - \vec{r}_2|)$  is a measure of the excess probability over what we would expect for an uncorrelated random Poisson distribution. So, the two-point correlation function describes the extent of clustering as a function of scale, and has been used as one of the most important tools to analyze large-scale structure of the universe (see e.g. (Coil, 2013)). Alternatively, one can also utilize its Fourier space transformation, the power spectrum  $P(\mathbf{k})$ :

$$\langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P(\mathbf{k}). \quad (1.1)$$

Here,  $\delta_D$  refers to the Dirac delta function,  $\delta^*$  is the complex conjugate, and  $P(\mathbf{k}) = P(k)$  due to the statistical cosmological principle, as the power spectrum must be invariant under the rotation of the wavevector  $\mathbf{k}$ .

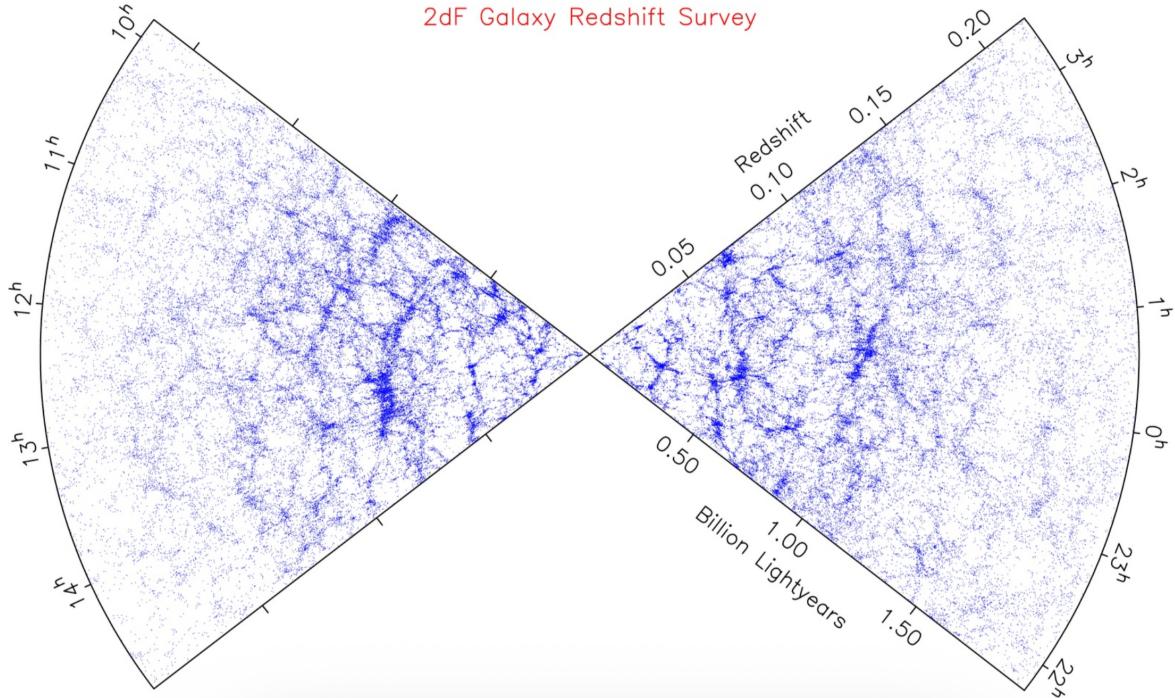
## 1.2 Redshift-Space Distortions

To be able to use the two-point correlation function to study clustering, one typically uses galaxy redshift surveys to obtain observations of spatial distributions of galaxies such as the galaxy distribution measured by the 2DF survey (Percival et al., 2001) as seen in fig 1.1. The distance measurement for objects in the surveys are obtained using the observed redshift using  $r(z) = \int_0^z c dz / H(z)$ , where  $H(z)$  is the Hubble parameter and  $c$ =speed of light (Hang et al., 2022). This would provide the true distance measurement if the proper velocities of a galaxy is equal to the Hubble velocity. However, the inhomogeneity that can be observed in the distribution of galaxies causes galaxies to have an excess over the Hubble velocity, called peculiar velocities (Mo et al., 2010). These peculiar velocities Doppler-shift objects along our line-of-sight, modifying the measured redshift as seen below:

$$1 + z_{\text{obs}} = (1 + z_{\text{cos}}) \left(1 - \frac{v_{\parallel}(\mathbf{r})}{c}\right)^{-1}$$

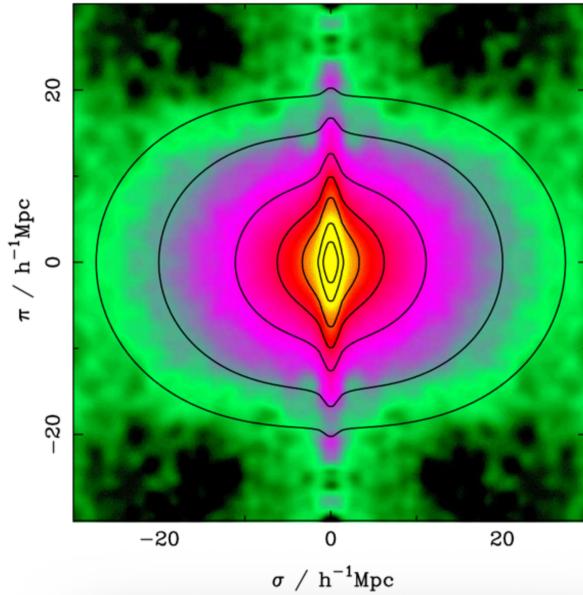
Here,  $z_{\text{obs}}$  is the observed redshift,  $z_{\text{cos}}$  is the redshift we would observe if objects had no peculiar velocity,  $\mathbf{r}$  is the real-space position, and  $v_{\parallel}(\mathbf{r})$  is the peculiar velocity component along our line-of-sight (Hang et al., 2022). Thus the distance measurement that we obtain from it, which is called *redshift-space distance*, can be transformed to obtain the true distance measurement or *real-space distance* as seen in equation 1.2. As we can see, the map from redshift-space distance  $s$  to real-space distance  $r$  involves the peculiar velocity  $v$  (along our line-of-sight).

$$\mathbf{s} = \mathbf{r} + \frac{(1 + z_{\text{cos}}) v_{\parallel}(\mathbf{r})}{H(z_{\text{cos}})} \hat{\mathbf{r}} \quad (1.2)$$



**Figure 1.1:** The galaxy distribution in redshift-space measured by the 2dF survey. Each blue dot represents a galaxy. Source: Percival et al. (2001)

This transformation of variables gives rise to a non-linear mapping of clustering (or any information for that matter) in terms of the peculiar velocity, and causes clustering in redshift-space distorted in an anisotropic way. These distortions are called *Redshift space distortions* (RSD) (Mo et al., 2010). The distortions have been examined in broadly two regimes in terms of scale. On larger scales, objects tend to fall towards regions of higher density. For a distant observer, this will make the clustering appear stronger in redshift space than it actually is along the line of sight. This effect is called the *Kaiser effect* (Kaiser, 1987). On the other hand, on smaller scales, random motions dominate and cause the clustering to appear lesser than it is along our line of sight. Since the density field appears stretched along our line of sight, this effect is called the *Fingers-of-God effect* (Jackson, 1972). The distortions can be more clearly understood by looking at redshift-space  $\xi$  as seen in figure 1.2. In the absence of RSD,  $\xi$  in real space would appear as roughly circular iso-contours in the plane of radial and transverse separations. However, the Fingers-of-God effect elongate the contours along the line-of-sight on small scales, and the Kaiser effect squash the contours on larger radii.



**Figure 1.2:** The redshift-space correlation function for the 2DF redshift survey as a function of radial separations on the  $y$ -axis and transverse separations on the  $x$ -axis. The contours are model predictions for the redshift distortions. Source: Peacock et al. (2001)

As mentioned before, redshift-space distortions affect all relevant information that we obtain in redshift space. So any attempt to study relevant cosmological information from observations requires a good understanding of the mapping from redshift to real space. In particular, since redshift space distortions are induced by peculiar velocities of objects, the use or development of an accurate model of the velocity differences of pairs of objects (sometimes also called pairwise velocities) is at the core of accurate redshift-space distortion modelling using two-point statistics like  $\xi$ . However, it is incredibly hard to find the functional form of the velocity difference distribution from first principles alone (see e.g. Bianchi et al. 2015). Additionally, trying to obtain a functional form for the velocity difference from measurements of redshift space is also ill-advised, especially if the velocity difference distribution is scale-dependent (see e.g. Scoccimarro 2004). In addition to observational limits and sample bias, if velocity difference distributions are scale-dependent, one usually recovers a distribution that has moments that are a combination of the moments of the true physical distribution from observations (see Scoccimarro 2004 for a full derivation of the proof). As a consequence, the most commonly used redshift space distortion models make strong assumptions about the characteristics of the velocity difference distributions.

The simplest and perhaps the most popular RSD model is the dispersion model (Fisher, 1995; Scoccimarro, 2004), which usually treats the redshift space power spectrum as a linear expansion of the redshift space power spectrum  $P_s^L(k, \mu)$  multiplied by a damping factor  $D_{\text{FOG}}$  to account for the Fingers-of-God

effect under the distant-observer approximation as seen in equation 1.3:

$$P_s(k, \mu) = P_g(k) \underbrace{\left(1 + \beta\mu^2\right)^2}_{P_s^L(k, \mu)} \underbrace{\frac{1}{1 + k^2\mu^2\sigma_p^2/2}}_{D_{FOG}} \quad (1.3)$$

Here,  $P_g(k)$  is the real-space galaxy power spectrum, the Kaiser distortion parameter  $\beta = f/b$ , with  $f$  being the linear growth rate and  $b$  is the linear bias factor due to the fact that galaxies are biased tracers of the underlying matter distribution. Additionally,  $\mu = k_i/k$ , where  $i$  denotes the chosen line of sight direction, and  $\sigma_p$  is the *scale-independent* velocity difference dispersion.

The dispersion model assumes that a scale-independent velocity difference distribution gives rise to the damping factor, and that the Kaiser effect, which is captured by  $P_s^L(k, \mu)$  and the Fingers-of-God effect are independent of each other. In equation 1.3, the  $D_{FOG}$  is a Lorentzian function. Variations of this model include using a Gaussian  $D_{FOG}$  and non-linear expansions of the redshift space power spectrum just to name a few.

Another popular redshift-space distortion model among many others is the *streaming* model, which treats the redshift-space two-point correlation function as the result of the convolution of  $\xi(r)$  with a *scale-dependent* peculiar velocity difference or pairwise velocity distribution as seen in equation 1.4 (Kaiser, 1987; Scoccimarro, 2004).:

$$1 + \xi_s(s_\perp, s_\parallel) = \int_{-\infty}^{\infty} \frac{H(z) dr_\parallel}{(1+z)} \times \mathcal{P}\left(v_\parallel = \frac{H(z)(s_\parallel - r_\parallel)}{(1+z)} \mid s_\perp, r_\parallel\right) \left[1 + \xi\left(\sqrt{s_\perp^2 + r_\parallel^2}\right)\right] \quad (1.4)$$

Here,  $\xi_s(s_\parallel, s_\perp)$  represents the redshift-space two-point correlation function,  $\xi(r)$  refers to the real-space correlation function, and  $\mathcal{P}(v_\parallel \mid s_\perp, r_\parallel)$  refers to the pairwise velocity probability distribution function.

Again, the most-used version of this streaming model is the Gaussian streaming model (GSM) (Reid & White, 2011) that uses a scale-dependent Gaussian model for  $\mathcal{P}(v_\parallel \mid s_\perp, r_\parallel)$ . However, previous works on analytical considerations using N-body simulations (typically dark matter-only simulations) show that the assumption of a scale-dependent and a scale-independent Gaussian pairwise velocity distributions do not work well, both on small and large scales (see e.g. Scoccimarro 2004; Bianchi et al. 2015; Kuruvilla & Porciani 2018). Thus, in this work, we aim to improve redshift-space distortion modelling by attempting to find well-describing analytical velocity-difference distributions that work on small and large-scales using hydrodynamical, cosmological simulations.

### 1.3 FLAMINGO simulations

As mentioned in the previous section, it is hard to derive what the velocity difference distribution from observations are a priori, so simulations can be used to find

a model for the distribution. An well-defined analytical model in particular would make the calculations of the mapping of redshift-space to real-space power spectrum or two-point correlation function more straight-forward. With this goal in mind, this work utilized data from the hydrodynamical cosmological simulations project FLAMINGO (Schaye et al., 2023). FLAMINGO accounts for baryonic effects that are not taken into account in dark matter only simulations when trying to study the evolution of large scale structure. Baryonic effects do affect the center-of-mass peculiar velocities of tracers of the matter distribution, so it is important that they are handled properly. The project offers three levels of baryonic particle masses:  $8.6 \times 10^9 M_{\odot}$ ,  $1.1 \times 10^9 M_{\odot}$  and  $1.3 \times 10^8 M_{\odot}$ , 2 different simulation box sizes, 1 Gpc and 2.8 Gpc, and several different cosmologies. For the purpose of this project, we used data from the simulation with intermediate mass resolution  $1.1 \times 10^9 M_{\odot}$  (identifier L1\_m9) with  $\Lambda$ CDM cosmology (D3A cosmology), which corresponds to a universe that is flat with  $\sim 75\%$  dark energy,  $\sim 20\%$  cold dark matter and  $\sim 5\%$  of baryonic matter at  $z=0$ . Overall, the simulation contains  $1800^3$  baryonic particles,  $1000^3$  neutrino particles and  $1800^3$  cold dark matter (CDM) particles.

In order to model peculiar velocity differences, we use halos and subhalos as the tracers of interest. Halos are roughly stable, isolated objects that are supported by random motions of its particles against gravitational collapse, and consist mostly of dark matter (Benson, 2010). Halos can host substructures that are gravitationally bound called *sub-halos*, which are typically assumed to host a galaxy. In this work, we assume that each sub-halo hosts one galaxy, and will only refer to (sub-) halos from now. The simulation data that was used in this work utilized halos that were identified using the Friends-of-Friends (FOF) halo finder HBT+ (Hierarchical Bound-Tracing) code (Han et al., 2017; Kugel et al., 2025).

Briefly, HBT+ identifies halos in *history space* rather than phase space. FOF group particles are identified by assuming that they belongs to a group if the distance to the nearest group particle is smaller than a linking length  $b=0.2$  times the mean inter-particle separation. Afterwards, particles are tracked across time based on the idea that every sub-halo was once the only one hosted by a halo (Han et al., 2017; Chandro-Gómez et al., 2025). The main advantage of this halo finder over the use of the previous halo finder VELOCIRAPTOR used in Schaye et al. (2023) is that it is better at separating overlapping subhalos, which helps while finding halo properties such as center-of-mass positions (see e.g. Chandro-Gómez et al. 2025). Finally, for each (sub-)halo in the halo catalogue produced by HBT+, relevant properties of (sub-)halos were obtained using the Spherical Overdensity and Aperture Processor (SOAP), specifically designed for FLAMINGO. This includes exclusive sphere quantities that calculate properties such as the center-of-mass peculiar velocities of (sub-)halos for bound particles after applying a radial cut from a halo's center.

## 1.4 Goals of this work

The goal of this project is to take a step towards improving redshift-space distortion modelling by finding analytical models for the distribution of velocity differences. To accomplish this, we use FLAMINGO simulation data to first examine the performance of simple analytical functions that are popular in redshift-space distortion modelling. We examine if the assumption of scale-independence is valid in RSD models. Additionally, we also attempt to find analytical models for the (scale-dependent) velocity difference distribution in order to improve redshift-to-real space mapping of information.

## 1.5 Structure

This thesis is structured as follows. In chapter , we explain the methods used to obtain the relevant peculiar velocity difference quantities and the fitting methods used. Chapter 3 examines the assumption of scale-independent models, and presents the results and analysis of fitting different scale-dependent analytical models to the one-halo velocity difference distributions. After that, chapter 4 presents the analysis of two-halo distributions. A summary of the findings, along with a discussion on the future outlook of the work done in this thesis is presented in chapter 5

# 2 . Methods

To find velocity difference distributions, we start by assuming that contributions to the power spectrum can be split into two as seen in equation 2.1 (see e.g. Tinker 2007; Hang et al. 2022) and this approach is typical in redshift-space distortion modelling using halo models (Seljak 2000; Peacock & Smith 2000). This also applies to the two-point correlation function. Here,  $P_{1h}(k)$  or the 1-halo power spectrum term, corresponds to contributions on smaller scales where the peculiar velocity difference inside halo groups matter (Fingers-of-God effect is most relevant). Additionally,  $P_{2h}(k)$  or the 2-halo power spectrum term corresponds to contributions where the peculiar velocity differences of pairs of halo groups matter the most.

$$P(k) = P_{1h}(k) + P_{2h}(k) \quad (2.1)$$

Thus, the fitting and examination of the characteristics of the peculiar velocity difference distributions is also split into two based on relevant scales. The peculiar velocity difference distribution inside halo groups will be referred to as the *one-halo* distribution, and the peculiar velocity differences of pairs of halo groups will be referred to as the *two-halo* distribution from now on.

First focusing on modelling of the one-halo distribution, to find a well-describing one-halo velocity difference distribution, for each sub-halo, we find the peculiar velocity difference  $\vec{v}$  with  $\vec{v} = \vec{v}_{\text{com}} - \vec{V}_{\text{com}}$ , where  $\vec{V}_{\text{com}}$  is the center-of-mass peculiar velocity of the halo group that contains the sub-halo, and  $\vec{v}_{\text{com}}$  is the center-of-mass peculiar velocity of the sub-halo. Assuming the distant observer approximation on scales where the Fingers-of-God effect dominates, and assuming that each axis  $\vec{v}$  are independent samples from the same peculiar velocity difference distribution, we can then model  $v$  as:

$$v_i = v_{\text{com},i} - V_{\text{com},i} \quad (2.2)$$

We do not assume a preferred axis. This increases the number of available samples by a factor of 3. An important thing to note here is that we do expect parameters such as the velocity dispersion to be affected by halo mass (see e.g. Hang et al. 2022), so we first bin the available peculiar velocity difference information based on the halo group mass (FOFMass values found in the original halo catalogue). Since the resolution of baryonic particles in the simulation used

is  $\sim 10^9 M_\odot$ , we choose to model FOF groups in the mass range  $10^{13-15} M_\odot$ , and mass bins are chosen to have a bin width of 0.5 dex.

In each halo group mass bin the sub-halos in each halo group are identified using the `HostHaloIndex` and `HaloCatalogueIndex` properties available in the halo catalogue. A halo group can be uniquely identified using its `HostHaloIndex`, and the sub-halos contained in one halo group will all share the same `HostHaloIndex` that corresponds to the halo group's `HaloCatalogueIndex` (`HaloCatalogueIndex` = -1 for the halo groups). As mentioned before, we use the exclusive sphere properties of subhalos. The peculiar velocities of the sub-halo and the halo group are stored in `CentreofMassVelocity`.

For the purpose of examining the scale dependence  $r$  of the parameters of any peculiar velocity difference distribution, we use  $r = |\vec{x}_{\text{COM}} - \vec{X}_{\text{COM}}|$ . Both the COM position of the sub-halo ( $\vec{x}_{\text{COM}}$ ) and the COM position of the halo group ( $\vec{X}_{\text{COM}}$ ) are obtained from the halo catalogue property `CentreofMass`.

Then, to fit analytical models, we use the binned maximum likelihood method (Baker & Cousins, 1984). In the case of functions with one variable, we start by assuming that we have  $N$  bins with a constant bin width  $h$ . We aim to fit any function  $f(v; \theta)$ , where  $v$  refers to the velocity difference data and  $\theta$  refers to the fit parameters. In each bin  $i$ , the measured content is  $n_i$ . On the other hand, the fitted bin content is:

$$f_i(\theta) = \int_{v_0+h(i-1)}^{v_0+hi} f(v; \theta) dv \quad (2.3)$$

If we assume a Poisson distribution for the number of events in each bin,  $n_i$ , the histogram likelihood is:

$$\mathcal{L}_i(\theta) = \prod_{i=1}^N \frac{f_i^{n_i} e^{-f_i}}{n_i!} \quad (2.4)$$

and the negative log-likelihood is:

$$\mathcal{F}_i(\theta) \equiv -\ln \mathcal{L}_i(\theta) = -\sum_{i=1}^N (n_i \ln f_i - f_i - \ln n_i!) \quad (2.5)$$

Here,  $\sum_{i=1}^N f_i = A(\theta)$  is just the fitted contents of all bins, and  $\sum_{i=1}^N \ln n_i!$  is independent of  $\theta$ , so it can be ignored. Ignoring this constant, the fitting is done by minimizing the following function:

$$\mathcal{F}_i(\theta) = -\sum_{i=1}^N n_i \ln f_i + A(\theta). \quad (2.6)$$

The minimization in this work was done with the L-BFGS-B algorithm (Zhu et al., 1997)), using the `minimize` function in SciPy(Virtanen et al., 2020), and applying bounds on parameters when appropriate. Any function is normalized such that the total number of objects  $M$  is equal to the total area under the function:

$$M = \int_{-\infty}^{\infty} f(v; \theta) dv. \quad (2.7)$$

Extending the above maximum-likelihood method to a function  $f(r, v; \theta)$  of two variables  $r$  and  $v$ , we can bin the velocity difference data into  $N_r$  radial bins  $\{[r_{i-1}, r_i]\}_{i=1}^{N_r}$  and, within each radial bin  $i$ , we can have  $N_v^{(i)}$   $v$ -bins  $\{[v_{j-1}^{(i)}, v_j^{(i)}]\}_{j=1}^{N_v^{(i)}}$ . Assuming that  $n_{ij}$  is the observed count in bin  $(i,j)$ , the expected counts per bin is:

$$P_{ij}(\theta) = \int_{r_{i-1}}^{r_i} dr \int_{x_{j-1}^{(i)}}^{x_j^{(i)}} dx f(r, x; \theta) \quad (2.8)$$

Then, the total log-likelihood is:

$$\ln \mathcal{L}(\theta) = \sum_{i=1}^{N_r} \sum_{j=1}^{N_v^{(i)}} \left[ n_{ij} \ln P_{ij} - P_{ij} - \ln(n_{ij}!) \right].$$

Ignoring the constant  $\sum \ln n_{ij}!$  in a similar way as before, the negative log-likelihood to minimize is then

$$\mathcal{F}(\theta) = -\ln \mathcal{L} = -\sum_{i,j} n_{ij} \ln P_{ij} + \sum_{i,j} P_{ij}. \quad (D.8)$$

Equation 2.8 is evaluated using `dblquad` in SciPy , and  $\mathcal{F}$  is minimized with the L-BFGS-B algorithm option in SciPy's `minimize` as mentioned before (Virtanen et al., 2020).

Finally, we use a slightly different velocity difference quantity for 2-halo distributions. Without assuming a distant observer approximation, for a pair of halo groups, the velocity difference vector is given by  $\vec{V} = \vec{V}_{1,\text{COM}} - \vec{V}_{2,\text{COM}}$ . Similarly, the separation vector between halo group 1 and halo group 2 is  $\vec{X} = \vec{X}_{1,\text{COM}} - \vec{X}_{2,\text{COM}}$ . Now, we aim to model the projection of the velocity difference vector along the (radial) separation of halo groups  $\tilde{v} = \vec{V} \cdot \frac{\vec{X}}{\|\vec{X}\|}$  using the aforementioned binned maximum likelihood method. However, 2-halo distributions mass binning choices are more complex, as we now need to consider mass  $M_1$  of halo 1 and mass  $M_2$  of halo 2 in each pair of halo groups. We choose to focus on a few mass bins for  $M_1$  and  $M_2$  in this work, which will be discussed further in chapter 4. It is

also important to note that since we are working with a simulation box of size 1 Gpc, periodic boundary conditions apply to the positions of objects in the box. The simulation box is treated as repeating at the edges, so distances are always measured along the shortest path across this repeating space. Consequently, the largest possible separation between any two objects in the simulation box is half the box length or 500 Mpc.

# 3 . One-Halo Distributions

This chapter presents the different velocity dispersion distributions of sub-halos with respect to their larger FOF group center-of-mass velocities. First, the fitting trials of common analytic velocity dispersion models to the simulation data are shown, illustrating how they do not capture the observed dispersion pattern. Then, the results of proposed *1-halo* dispersion model of this project is used to show why we recommend this model for the scales where the *Fingers-of-God* effect dominates. We also explore the scale-dependence of the model per mass bin. For every distribution where a model was fit, the normalization was done such that the area under the model is equal to the total number of samples in the distribution.

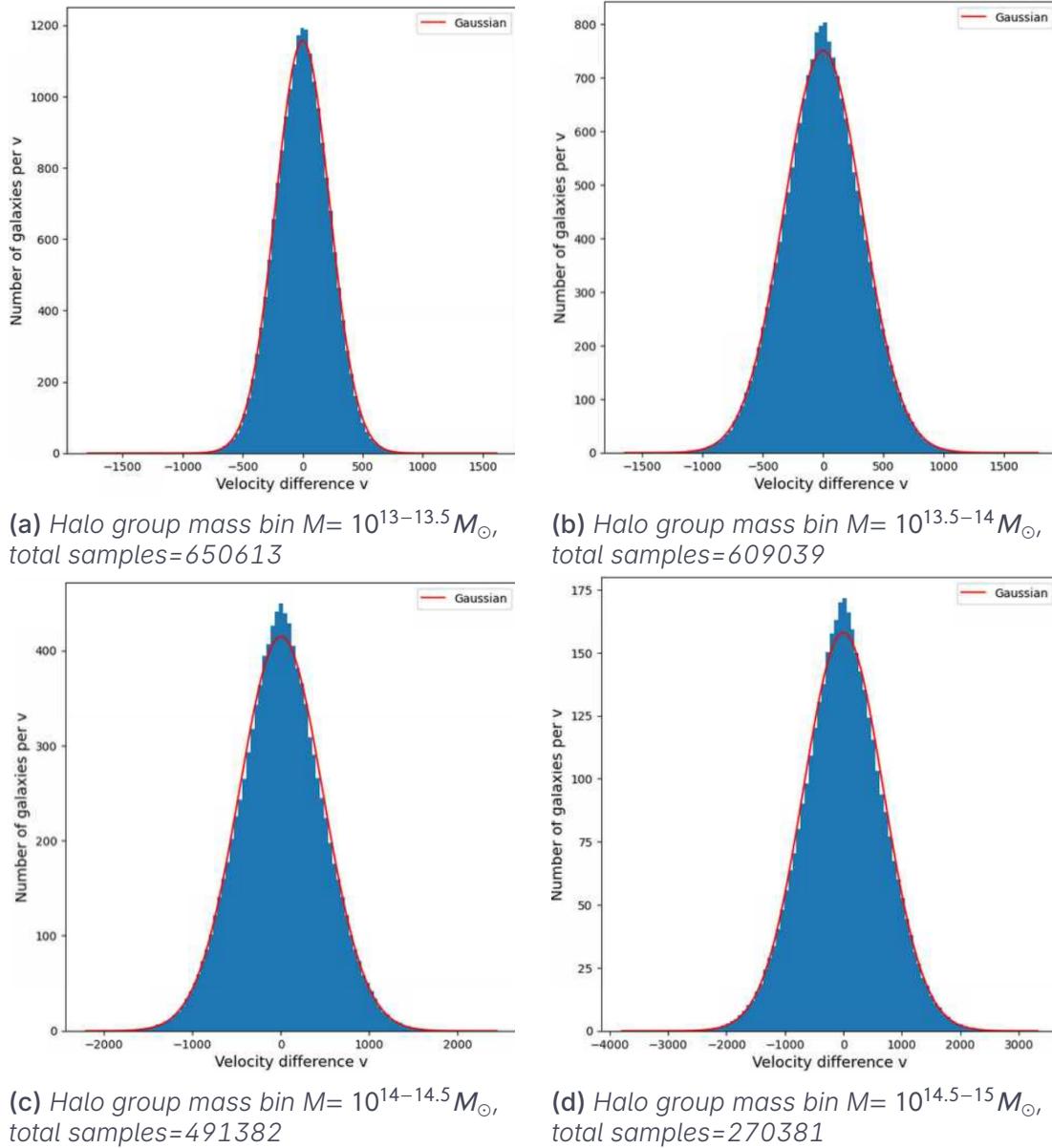
## 3.1 Test of Common Analytical Models - Scale Independent

Starting with examining whether a scale-independent simple analytical models commonly used in the context of the dispersion model as explained before in section 1.2 are enough, we examine the results of fitting a simple scale-independent Gaussian, Lorentzian or a Laplacian distribution. Firstly, looking at the Gaussian fits to 1-halo peculiar velocity differences across different mass bins using the function, where  $\mu$  = mean and  $\sigma$ :

$$f_G(v; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(v - \mu)^2}{2\sigma^2}\right] \quad (3.1)$$

It must be noted that typically, for the FOG convolution factor in the dispersion model, most applications assume that the odd moments of the distribution are zero, but we include the mean for the sake of completeness. Figure 3.1 shows the best-fit Gaussian models for halo groups with masses between  $10^{13-15} M_\odot$ , and the best-fit parameters are shown in table 3.1. We first note that an assumption of no mean infall can be a good approximation, as the mean values across different halo mass bins are extremely close to zero. However, as seen in figure 3.1, the distribution does not quite capture the excess kurtosis seen in the data. The underlying velocity difference data here appears to have Gaussian-like tails, but consistently shows an excess kurtosis that increases with increase in halo group

mass. It is interesting to note that the underlying data does not appear show a strong skewness. This means that the use of a Gaussian is not fully sufficient for accurate FOG redshift-space modelling under the scale-independent assumption, which is in line with the existing non-Gaussian expectation of the distribution from previous N-body simulations (see e.g. Scoccimarro 2004; Bianchi et al. 2016).



**Figure 3.1:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation. The red line indicates the best-fit, scale-independent Gaussian to the velocity distribution.

Mass bin [ $M_{\odot}$ ]	$\mu$	$\sigma$
$10^{13-13.5}$	-0.14	224.7
$10^{13.5-14}$	0.059	324.3
$10^{14-14.5}$	1.41	473.5
$10^{14.5-15}$	0.079	684.4

*Table 3.1:* Best-fit parameters for a scale-independent Gaussian distribution for different halo group mass bins

Moving on to a Laplacian scale-independent fit that is often used to an improved model to account for the non-Gaussian characteristics of the velocity difference distribution on smaller scales (see e.g. Scoccimarro 2004; Hang et al. 2022), we discuss the fitting done using the following function, where  $\mu$  = mean and  $b$  is the scale:

$$f_{\text{Laplace}}(v; \mu, b) = \frac{1}{2b} \exp\left[-\frac{|v - \mu|}{b}\right] \quad (3.2)$$

The best-fit parameters are provided in table 3.2, and similar to the Gaussian fits, the means are again extremely close to zero. Figure 3.2 presents the best-fit Laplacians to the underlying data, and as one can see, it provides a much poorer fit than a Gaussian does across all halo group mass bins. A Laplacian has an excess positive kurtosis compared to a Gaussian, and this leads to an overestimation of the contribution close to  $v = 0$ , and heavier tails that do not match the behaviour of the data. Thus, the use of a scale-independent Laplace distribution in the context of a dispersion model does not match the observed behaviour of the underlying velocity difference on FOG scales.

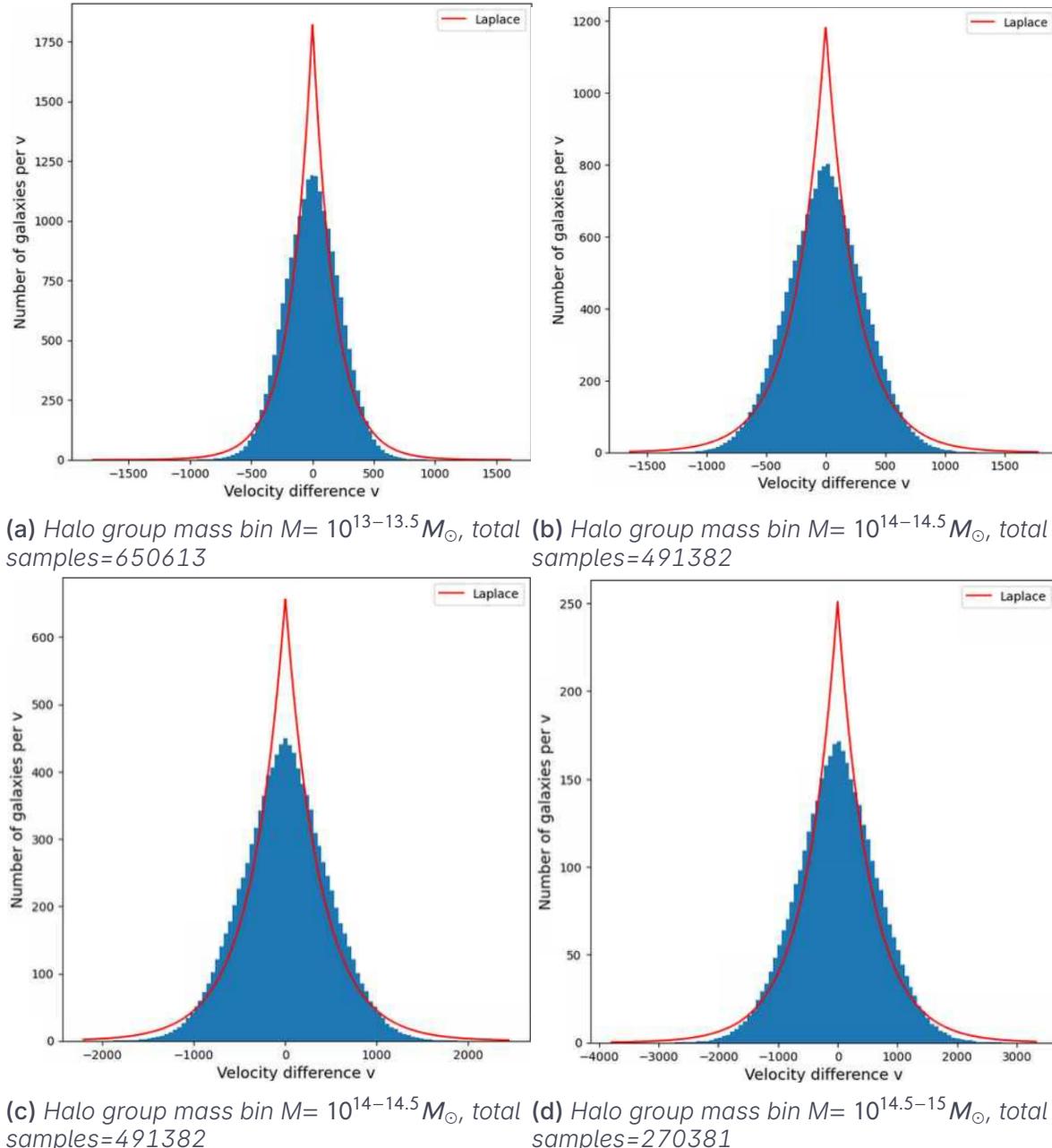
One could also examine the fits of a Lorentzian distribution as shown below, where  $v_0$  = location and  $\gamma$  = scale:

$$f(v; v_0, \gamma) = \frac{1}{\pi} \frac{\gamma}{(v - v_0)^2 + \gamma^2} \quad (3.3)$$

As we can see in figure 3.3, heavy-tailed distributions like a Lorentzian or a Laplace do not capture the underlying behaviour of the velocity difference on FOG scales well. The data as mentioned before, match that of a Gaussian and drops off much faster than the best-fit Lorentzian or Laplacian distribution. Both a Lorentzian and a Laplace predict more high-velocity differences than what actually exists in the simulation data. Additionally, the peaks of the underlying distribution are much flatter than the best-fit Lorentzian or a Laplace predict. While the Lorentzian distribution lacks of a sharp peak compared to a Laplace distribution across all mass bins, it still places too much weight near  $v = 0$ .

So, a scale-independent Gaussian, Lorentzian or a Laplacian distribution are not sufficient to provide an accurate description of the underlying velocity

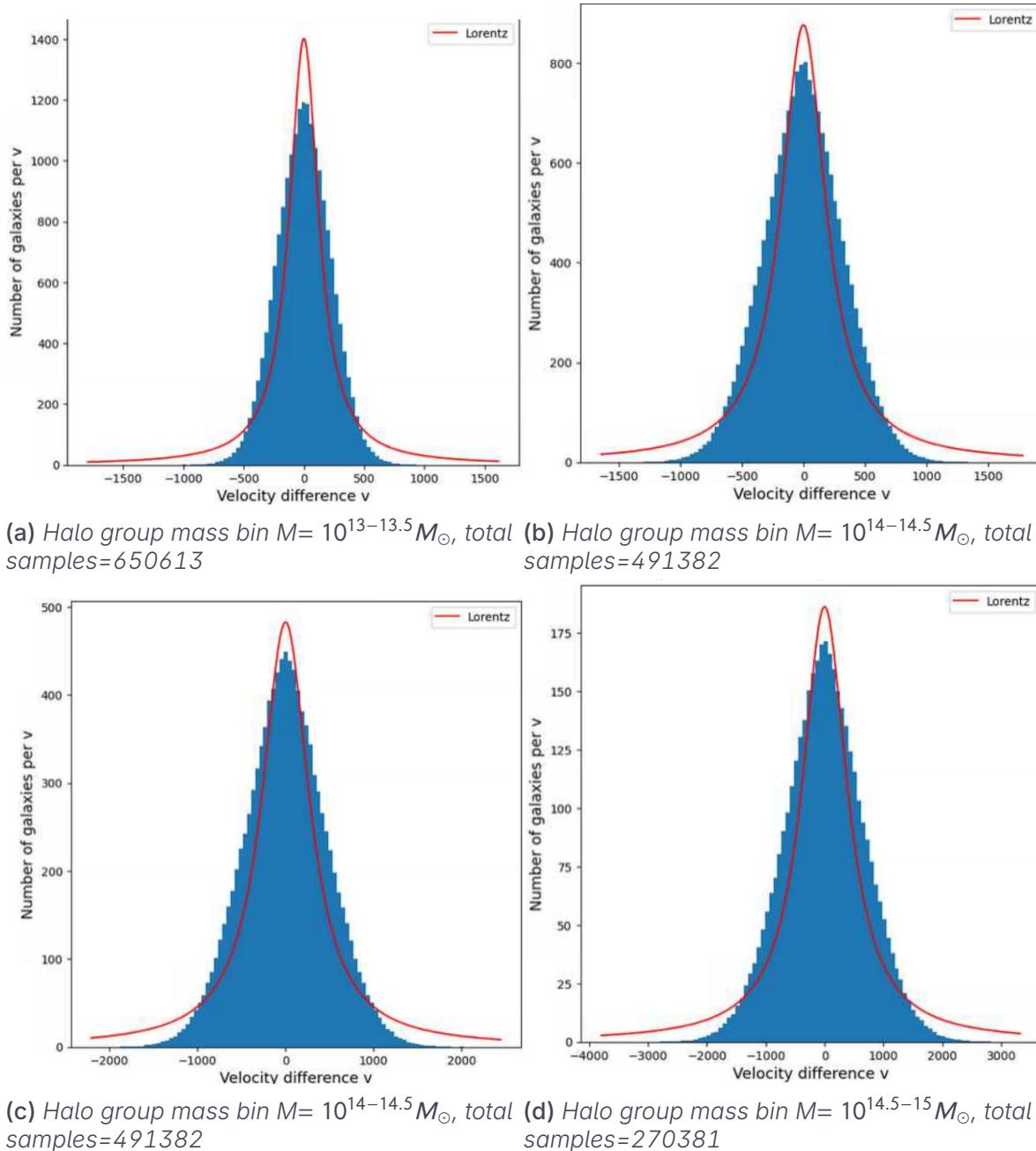
difference distribution across different mass bins. However, this is perhaps incongruent with the findings that a Lorentzian damping factor in Fourier space matches the ratio of redshift space and real space power spectra in galaxy redshift surveys if one uses the dispersion model in 1.3 (see e.g. Landy 2002). This can be explained if the underlying peculiar velocity difference distribution was indeed scale-dependent. This is because if it were scale-dependent, the velocity difference distribution can no longer be obtained by a simple ratio of redshift space and real space power due to the more complicated relationship between the two as seen in the streaming model (equation 1.4). Thus, we examine the scale dependence of the underlying distribution, and propose a working analytical model for the scale-dependent distribution in the next section.



**Figure 3.2:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation. The red line indicates the best-fit, scale-independent Laplacian to the velocity distribution.

Halo group mass bin [ $M_{\odot}$ ]	$\mu$	$b$
$10^{13-13.5}$	0.5	177.9
$10^{13.5-14}$	-0.2635	256.6
$10^{14-14.5}$	0.725	373.9
$10^{14.5-15}$	-1.798	539.6

*Table 3.2: Best-fit parameters for a scale-independent Laplacian distribution for different halo group mass bins*



**Figure 3.3:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation. The red line indicates the best-fit scale-independent Lorentzian to the velocity distribution.

## 3.2 Analytical Scale-dependent Modelling

### 3.2.1 Conditional modelling

The examination of the scale ( $r$ ) dependent velocity distribution can be done by binning the available velocity difference information based on the separation of a

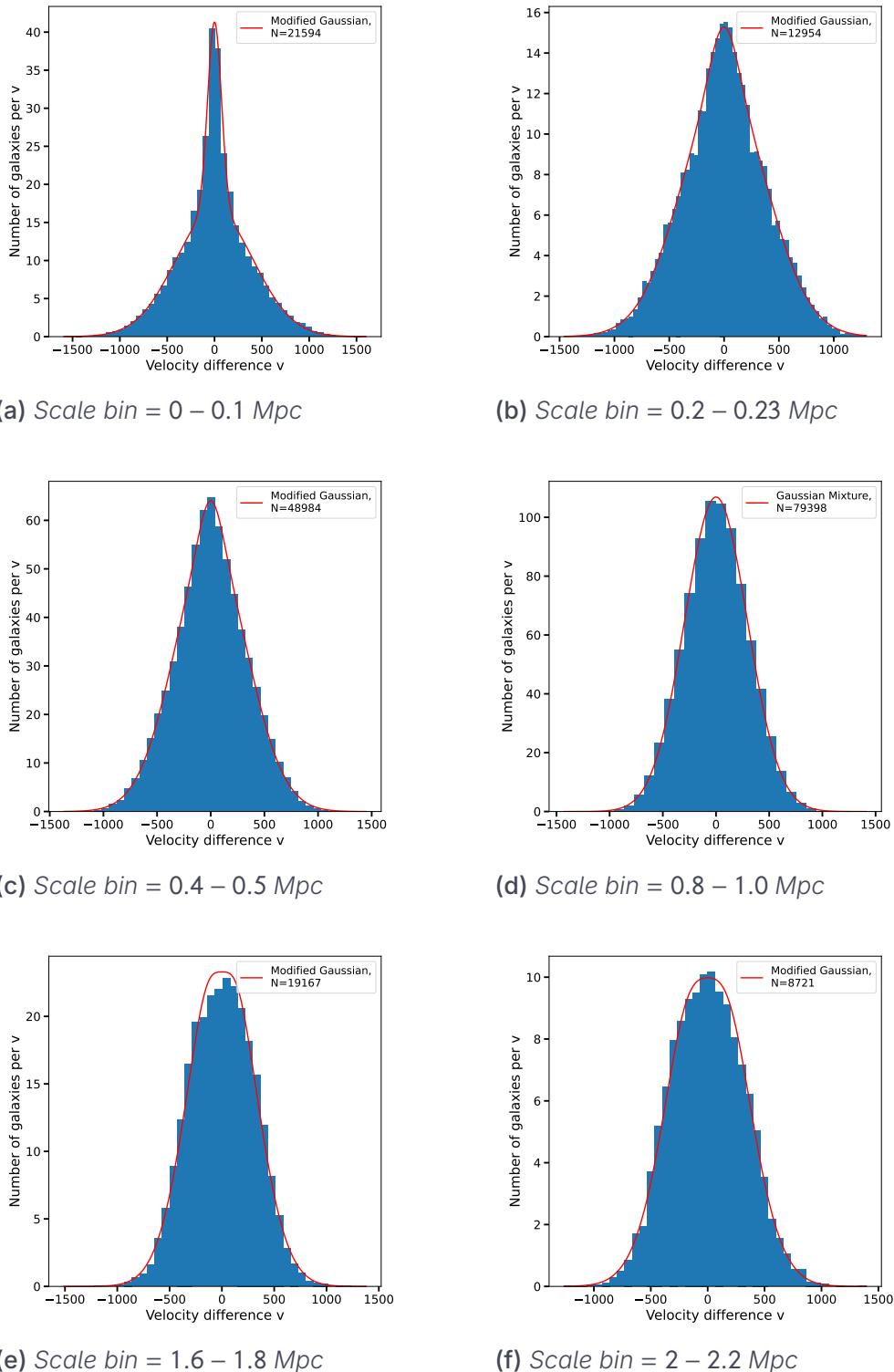
Halo group mass bin [ $M_{\odot}$ ]	$x_0$	$\gamma$
$10^{13-13.5}$	0.537	147.9
$10^{13.5-14}$	-0.664	221.4
$10^{14-14.5}$	-0.340	324.6
$10^{14.5-15}$	0.0237	463.5

*Table 3.3:* Best-fit parameters for a scale-independent Lorentzian distribution for different halo group mass bins

sub-halo with respect to the center of mass position of its halo group as described in chapter 2. Considering the properties of the distribution as observed in the previous section, we find that a weighted Gaussian sum as seen in 3.4 provides the best fits out of many other analytical forms (mostly with an exponential base) per  $r$  bin. This distribution will be referred to as the one-halo conditional distribution. In 3.4,  $\lambda$  controls the relative contribution of the two different Gaussians, each with different widths signified by  $\sigma_1$  and  $\sigma_2$ .

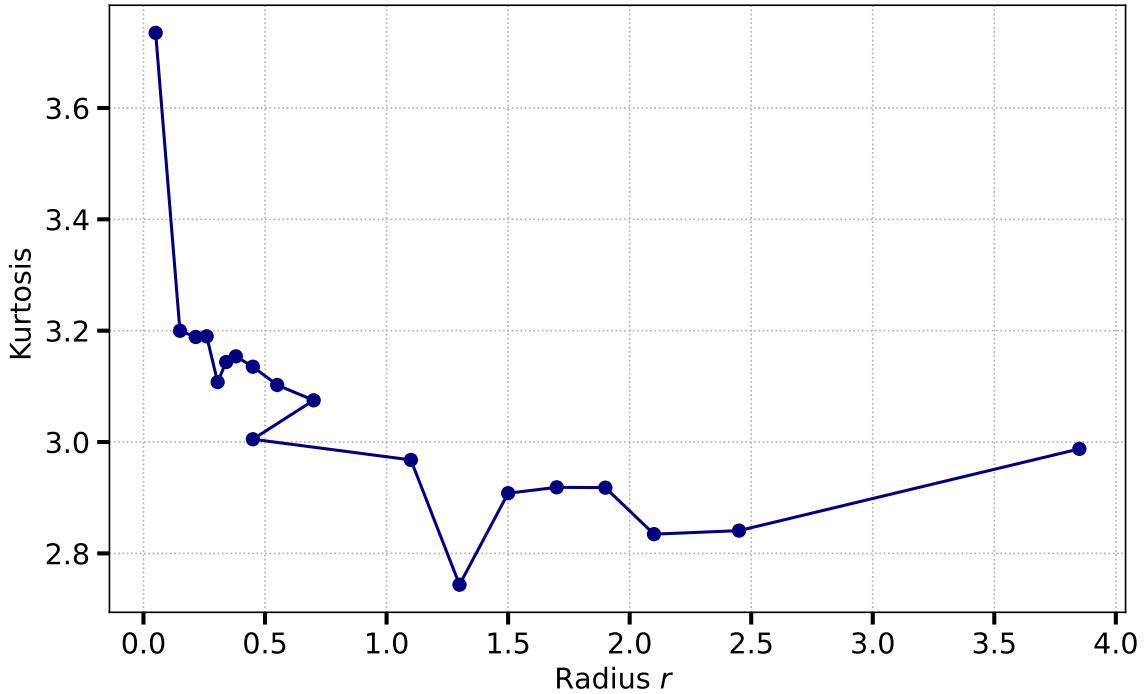
$$f(v | r) = \frac{1}{\sqrt{2\pi}} \left[ \frac{(1 - \lambda)}{\sigma_1} \cdot e^{-\frac{v^2}{2\sigma_1^2}} + \frac{\lambda}{\sigma_2} \cdot e^{-\frac{v^2}{2\sigma_2^2}} \right] \quad (3.4)$$

Focusing on the data and model fits from the halo group mass bin of  $10^{13.5-14} M_{\odot}$  first, as seen in figure 3.4, the underlying data shows clear variations across different  $r$  bins. This means that the assumption of a scale-independent velocity difference function is not consistent with the nature of the peculiar velocity difference of sub-halos, and the dispersion model does not provide an accurate picture of the redshift space distortions. So, the underlying scale-dependence helps explain why scale-independent models such as the Gaussian or Laplacian distributions considered before do not provide a good fit to the velocity distribution data. The scale-dependent distributions are consistent with the theoretical expectation and findings of other RSD models such as the streaming model (Scoccimarro, 2004) and the halo redshift-space distortion models (Seljak, 2000; Peacock & Smith, 2000; Tinker, 2007).



**Figure 3.4:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation for halo group mass bin  $M=10^{13.5-14} M_{\odot}$ . The red line in each sub-figure indicates the best-fit scale-dependent weighted sum of Gaussians shown in equation 3.4 to the velocity distribution, and  $N$  specifies the total number of samples

The scale dependence of the distribution is captured by the variation of the three fit parameters in equation 3.4 as seen in tables 3.4, 3.5, 3.6 and 3.7. The distribution has a zero-mean Gaussian with a broader width  $\sigma_1$  that contributes most of the weight, while the other zero-mean Gaussian typically has a smaller width  $\sigma_2$  across all  $r$  bins. This shape captures both a Gaussian-like tail and a concentrated central peak observed in the data at any  $r$ . As shown in the plots, the peak becomes less prominent as  $r$  increases, which corresponds to a decreasing weight  $\lambda$  of the narrower Gaussian component in the distribution. As each Gaussian component is symmetric and centered at zero, the overall skewness of the distribution is zero across all the bins. On the other hand, as seen in figure 3.5 kurtosis across all  $r$  bins is generally non-Gaussian (A Gaussian has a kurtosis of 3), which allows the distribution to reproduce the sharp central peaks observed in certain scales. The kurtosis values roughly match the expectations of non-Gaussian kurtosis at all scales from other N-body simulation work, such as Scoccimarro (2004), where the kurtosis values lie between  $\sim 2.6 - 4$ .

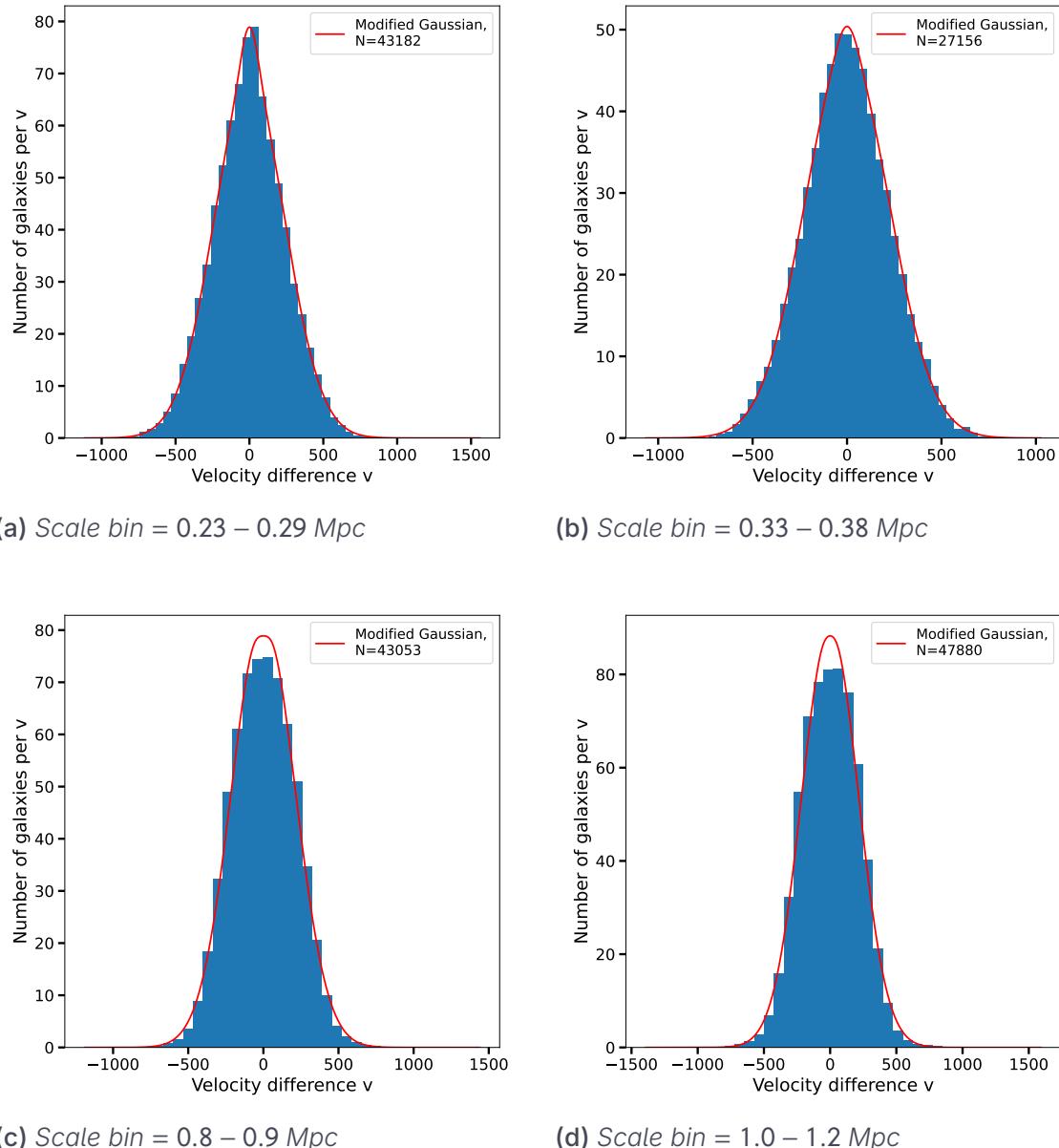


**Figure 3.5:** Kurtosis of the conditional distribution in equation 3.4 as a function of radius or scale  $r$  for the halo group mass bin  $M = 10^{13.5-14} M_{\odot}$  bin

Now, this analytical distribution was also fit to the velocity difference data in other halo mass bins. Figure 3.6 shows some representative best fits of the Gaussian mixture model for halo group mass bin  $M=10^{13-13.5} M_{\odot}$  and  $r$ , and the remaining fits are provided in the appendix A.1. The figures in fig 3.6 and the best-fit parameters at different scales as seen in tables 3.4, 3.5, 3.6 and 3.7 together reveal that the underlying distribution also has the same characteristics of decreasing relative weight near the peak  $v = 0$  with scale, i.e. non-Gaussian kurtosis at all scales, zero mean and skewness as seen in the data and model fits from the halo group mass bin of  $10^{13.5-14} M_{\odot}$ . So, the Gaussian mixture is a unified model for the 1-halo peculiar velocity difference distribution that works well across different mass bins.

Generally, the average  $\sigma_1$  increases with halo group mass as seen in tables 3.4, 3.5, 3.6 and 3.7, which is what we would expect from first principles, as the characteristic dispersion for a halo group is expected to increase with mass (Hang et al., 2022). Additionally, it must be noted that the largest radial bin is typically wider than the smaller radial bins. This is because there is a sharp decrease in the number of sub-halos beyond the typical virial radius of a halo group with mass  $M$ . The virial radius of a halo of mass  $M$  defines the characteristic boundary of the halo, and in this work, we adopt the definition of the virial radius being the radius at which the density is equal to the critical density multiplied by an overdensity factor of  $18\pi^2$  (Mo et al., 2010). To improve fitting by having a sufficient sample size to reduce sensitivity to noise, the larger radial bins were used. While this might introduce smoothing of the scale dependence of the 3 parameters, it does not affect the result that the functional form of the distribution still captures the features of the underlying distribution effectively.

So, we have now found a fully analytical expression that can effectively describe the observed distribution of the peculiar velocity difference of subhalos at fixed separations. Unlike purely numerical approximations for the distribution, the closed-form weighted sum of two Gaussians has well-known mathematical properties which include analytical expressions for moments, and allows easy parameter estimation and evaluation. As a result, any application of the model to evaluate the one-halo power spectrum or the correlation function, after a careful inclusion of mass dependence in the model, is thus more straightforward and quicker to evaluate.



**Figure 3.6:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation for halo group mass bin  $M=10^{13-13.5} M_{\odot}$ . The red line in each sub-figure indicates the best-fit scale-dependent weighted sum of Gaussians shown in equation 3.4 to the velocity distribution, and  $N$  specifies the total number of samples

Additionally, the non-zero  $\lambda$  and different  $\sigma_1$  and  $\sigma_2$  also clearly demonstrate that the assumption of a single Gaussian for the peculiar velocity different distributions at fixed scales is insufficient to fully capture the observed characteristics of  $v$  across different scales where the Fingers-of-God effect dominates. This means that RSD models such as the most commonly used Gaussian streaming model, especially for galaxy surveys, [see e.g. Samushia et al. 2014; Tamone

et al. 2020] that use a simple Gaussian would not perform well on small scales. Reid & White (2011) identified that the scale-dependent Gaussian streaming model (GSM) was insufficient to reproduce the redshift-space correlation function at scales smaller than  $s \sim 30h^{-1}\text{Mpc}$ . However, Reid & White (2011) suggested that the issue was with the use of perturbative theory predictions to compute mean  $v_{12}$  and the velocity dispersion  $\sigma_{12}$  that were used as inputs for the GSM, as perturbative theory is expected to break down for small scales. Other works on the streaming model (see e.g. Scoccimarro 2004; Bianchi et al. 2016; Kuruvilla & Porciani 2018) have identified that the assumption of a Gaussian itself is an oversimplification on different scales, and that is also what we observe in this work.

Now, the subsequent research on the streaming model such as Bianchi et al. (2016); Kuruvilla & Porciani (2018); Cuesta-Lazaro et al. (2020) sought to improve the performance of a Gaussian distribution for peculiar velocity differences, especially on smaller scales by introducing extensions of a Gaussian distribution can capture non-zero skewness and non-Gaussian kurtosis. For example, Kuruvilla & Porciani (2018) introduce a variation of the streaming model that uses generalized hyperbolic distributions (GHD), as it offers more flexibility than a simple Gaussian at fixed separations to model variable skewness, kurtosis and heavy tails as required, but increases the complexity in utilizing the distribution in the streaming model. Another example of a variation that introduced skewness was done by Cuesta-Lazaro et al. (2020) by assuming that the velocity difference distribution is a skew-t distribution, which is a generalization of the Student's t-distribution that includes an additional parameter to describe skewness.

Both Kuruvilla & Porciani (2018) and Cuesta-Lazaro et al. (2020) adopt this as they do find negatively skewed line-of-sight velocity distributions on smaller scales, and this is because they focus on the contribution of *pairs of dark matter halo groups*. Other formulations of variations of the GSM like Bianchi et al. (2015); Bianchi et al. (2016) also restrict themselves to halo group pairs. This can explain why there is a difference between the properties of the symmetric model found in this work and the skewed distribution functions of the more recent streaming models. It is worth noting that even if one were to consider the random motions inside halo groups directly, the streaming model still 'mixes' the 1-halo and 2-halo contributions by trying to capture the *full* observed profile at fixed scales. This mixing of contributions from random motions inside halo groups and the coherent motion of pairs of halo groups at smaller scales might again lead to a version of a skewed distribution like a GHD if one tries to work within the streaming model, as the velocity distribution would have to capture the skewness that arises from halo groups, while also trying to capture the characteristics, such as peakedness of the distribution. Ultimately, the approach of the streaming model as presented in Scoccimarro (2004) might require more complicated and less physically motivated velocity difference distributions to model redshift space distortions, unlike the distribution presented in this work.

On the other hand, the separation of the two contributions adopted in this work is the foundation idea of the halo model, which acknowledges the physical distinction between the dominant processes inside and outside halo groups. However, popular analytical models for the one-halo contribution typically again assume a Gaussian at fixed separations and masses (see e.g. Tinker 2007). As explained before, the assumption of a Gaussian distribution is not justified to model the underlying peculiar velocity differences of sub-halos inside halo groups accurately. Throughout the work on the one-halo velocity difference distribution, we have adopted the plane-parallel or distant observer i.e., that all lines of sights to different objects across the sky are effectively parallel. This allows us to fix a line of sight direction and project each velocity difference vector along the fixed direction to use in redshift space distortion modelling. Additionally, we also made the assumption that the axis components of the velocity difference vector are independent of each other. The advantage of this assumption is that each velocity difference vector contributes three independent samples to the fitting routine. Both of these assumptions are not expected to affect the results obtained here. Firstly, the assumption of a distant observer only breaks down larger scales, as the angular separation between any two tracers is not negligible for larger scales, which gives rise to the *wide-angle* effect where lines of sights to different objects are no longer approximately parallel (Castorina & White, 2020). Additionally, cross-axis correlations between each Cartesian component of velocity difference vectors are expected to be negligible inside halo groups, as there are no preferred directions because random motions dominates on the scales considered here.

Radial bin	$\sigma_1$	$\sigma_2$	$\lambda$
0-0.1	278.3	62.4	0.26
0.1-0.17	265.3	60	0.04
0.17-0.23	253.7	59.7	0.034
0.23-0.29	243.4	58	0.036
0.29-0.33	232.2	56.2	0.03
0.33-0.36	227.1	55.1	0.02
0.36-0.4	222.8	54	0.016
0.4-0.45	218.6	53.1	0.012
0.45-0.5	217.6	52	0.005
0.5-0.55	212.2	51	-0.0056
0.55-0.6	212.6	50	-0.014
0.6-0.65	210.6	48.8	-0.014
0.65-0.7	211.2	47	-0.01
0.7-0.8	212.7	46	-0.01
0.8-0.9	211.3	44.9	-0.008
0.9-1.0	210	43.4	-0.01
1.0-1.2	213.6	39.9	-0.003
1.2-5	221.8	36.8	-0.002

**Table 3.4:** Best-fit parameter values of equation 3.4 for halo group mass bin  $10^{13-13.5} M_\odot$

Radial bin	$\sigma_1$	$\sigma_2$	$\lambda$
0-0.1	412.6	72.8	0.21
0.1-0.2	395.9	84	0.068
0.2-0.23	392.2	122.3	0.073
0.23-0.29	378.5	122.3	0.074
0.29-0.32	365.5	77	0.038
0.32-0.36	362.6	102	0.054
0.36-0.4	353	122.3	0.063
0.4-0.5	341.5	105.1	0.053
0.5-0.6	331.2	104.7	0.041
0.6-0.8	316	79.9	0.028
0.8-1.0	297.6	99.1	0.0021
1.0-1.2	300	80	-0.013
1.2-1.4	280.6	178.6	-0.26
1.4-1.6	299.8	100	-0.04
1.6-1.8	304.8	100.1	-0.035
1.8-2.0	309.5	100.1	-0.035
2.0-2.2	303.2	134.2	-0.09
2.2-2.7	311.9	143.8	-0.09
2.7-5.0	335.4	101.8	-0.005

**Table 3.5:** Best-fit parameter values of equation 3.4 for halo group mass bin  $10^{13.5-14} M_\odot$

Radial bin	$\sigma_1$	$\sigma_2$	$\lambda$
0-0.3	590.4	103.5	0.072
0.3-0.5	545.7	140.7	0.05
0.5-0.6	518.8	138.7	0.048
0.6-0.8	517.1	121.9	0.0396
0.8-1.0	466.7	102.9	0.029
1.0-1.2	441.3	111.6	0.019
1.2-1.4	428.4	100.5	0.0241
1.4-1.6	420.88	111.2	0.006
1.6-1.8	424.7	120	0.007
1.8-2.0	430.1	100.3	-0.002
2.0-2.2	433.4	126.3	-0.002
2.2-2.7	449.3	99.9	-0.005
2.7-9	465.4	99.8	-0.05

**Table 3.6:** Best-fit parameter values of equation 3.4 for halo group mass bin  $10^{14-14.5} M_\odot$

Radial bin [Mpc]	$\sigma_1$	$\sigma_2$	$\lambda$
0-0.4	864.7	113.6	0.05
0.4-0.6	802.7	175.9	0.04
0.6-0.8	801.5	104.3	0.03
0.8-1.0	730	127.7	0.026
1.0-1.2	696.2	101.1	0.015
1.2-1.4	670.5	92.1	0.013
1.4-1.6	648.5	164.6	0.01309
1.6-1.8	629	170	0.024
1.8-2.0	613.5	100	0.02
2.0-2.2	604.7	98.79	0.017
2.2-2.7	604.3	96.8	0.017
2.7-4.0	644.1	90	-0.003
4.0-12.0	679.7	88.96	-0.003

**Table 3.7:** Best-fit parameter values of equation 3.4 for halo group mass bin  $10^{14.5-15} M_\odot$

### 3.2.2 Classification of sub-halos

We now have a well-describing analytical form for the conditional distribution, but it is also important to understand the possible physical reason behind the distribution having the form of a weighted sum of two Gaussian. For this purpose, since we focus on the random motions of sub-halos inside their halo groups, we can classify sub-halos based on their dynamical history and observe the differences in the populations with scale. Broadly, we can classify sub-halos into infalling, orbiting, and background sub-halos.

Firstly, infalling sub-halos are sub-halos that are on their initial approach towards the halo group and have not yet crossed the pericenter (Aung et al., 2023). Typically, infalling sub-halos experience minor removal of mass that occurs through dynamical processes such as tidal stripping, which is the process where tidal interactions within the halo group remove mass (Mo et al., 2010). So sub-halos that are infalling at a certain moment retain most of the mass that they had before they started falling towards the halo groups. Secondly, orbiting sub-halos are sub-halos that have crossed the pericenter at least once, and have experienced significant mass stripping by dynamical processes (Aung et al., 2023). Finally, background sub-halos are those that currently lie beyond the turnaround radius  $r_{ta}$ , which is defined as the radius at which the average physical radial velocity of objects is zero (Aung et al., 2023). By definition, one can only find background sub-halos beyond the turnaround radius.

Now, accurate classification of sub-halos requires the construction of halo merger trees, which track the history of halo assembly and evolution of the sub-halos across time by carefully considering the different physical effects such as dynamical friction and tidal stripping (Jiang & van den Bosch, 2016a). However, constructing accurate merger trees is a challenging and intensive process that is beyond the scope of this project. So, we adopt a simpler classification scheme based on the dynamical information of sub-halos available in the snapshot of redshift  $z = 0$  that allows us to gain basic insight into the possible physical motivation behind the conditional distribution.

#### Method

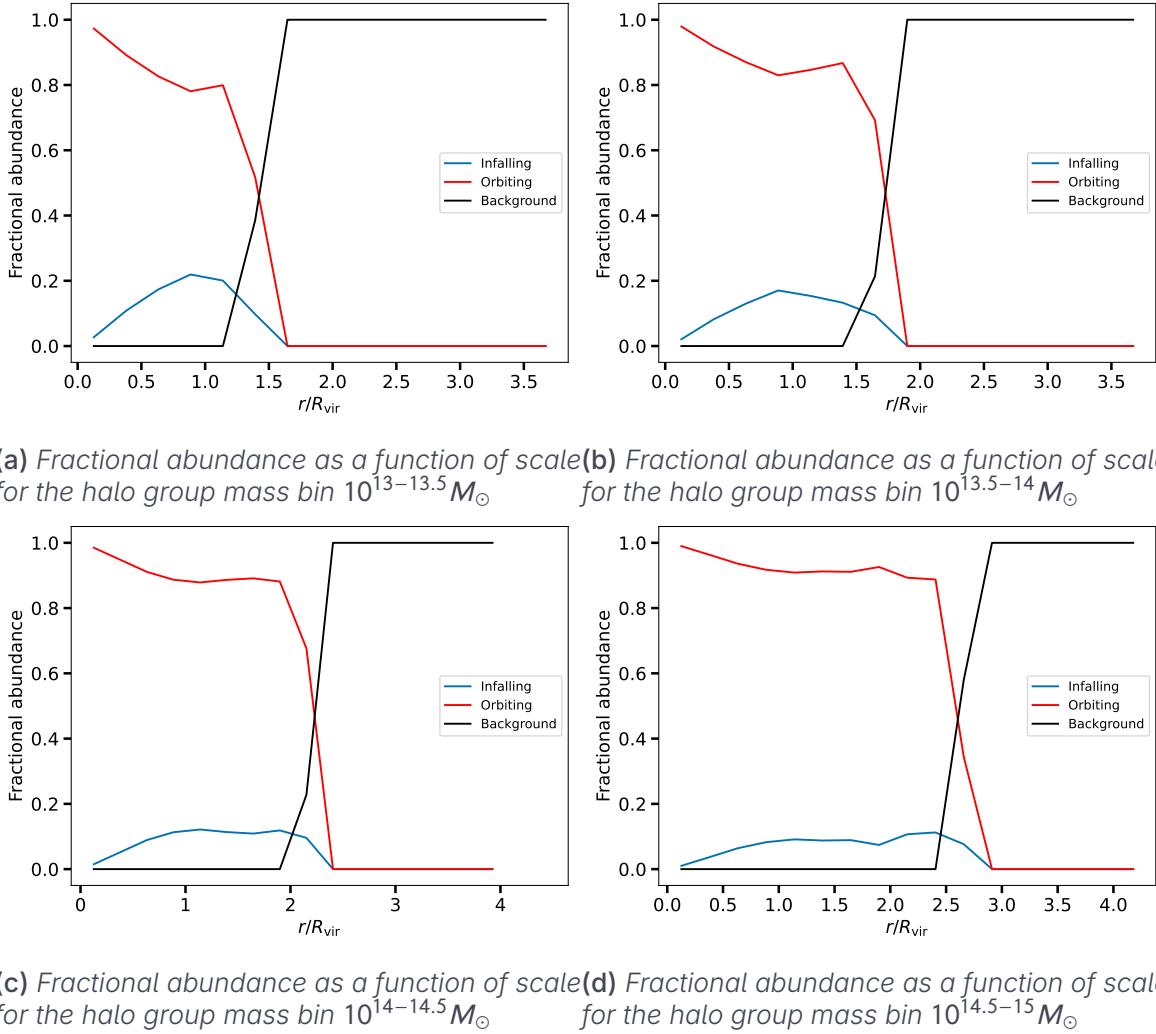
In order to separate background sub-halos from orbiting and infalling in each halo group, we utilize an average turnaround radius for a certain halo group mass bin  $M$ . To estimate the turnaround radius, the radial positions and corresponding radial velocities of all of the sub-halos from different halo groups were grouped into radial bins of fixed width to obtain stable average velocity profiles. For each bin, the mean radial velocity was calculated, which produced a radial velocity profile as a function of radius. The turnaround radius was then identified as the smallest radius at which the average radial velocity changes from negative to positive values. If no such transition was detected, the radius corresponding to the bin with the least negative average radial velocity was adopted instead. This derived average turnaround radius was then used to identify background sub-halos as any sub-halos with a radial separation  $r > r_{ta}$  inside its halo group.

Additionally, orbiting and infalling sub-halos by definition lie inside  $r < r_{ta}$ , and further distinction of the two was made primarily using the last maximum mass of the sub-halo throughout its lifetime,  $m_{max}$  and its current mass  $m$ . As mentioned before, one would expect orbiting sub-halos to have experienced significant mass loss due to dynamical processes such as tidal stripping, resulting in a current mass that is much smaller than its last maximum mass. On the other hand, infalling sub-halos, are expected to retain most of their (maximum) mass. Hence, we adopt a ratio of  $m/m_{max} = 0.6$  as the threshold value that separates infalling and orbiting sub-halos, i.e., sub-halos with  $m/m_{max} > 0.6$  are categorized as infalling and sub-halos with  $m/m_{max} \leq 0.6$  are categorized as orbiting (Jiang & van den Bosch, 2016b).

### Results and discussion

Figure 3.7 shows the fractional abundance of each category as a function of scale for the four different halo group mass bins considered in this work. Within each halo group mass bin, the scale in fig. 3.7 is normalized by the virial radius of the halo group mass  $M$ , and the turnaround radii for each mass bin of 0.5 dex from  $10^{13-15} M_\odot$  is  $1.4r_{vir}$ ,  $1.7r_{vir}$ ,  $2.2r_{vir}$  and  $2.6r_{vir}$  respectively. For each halo group mass bin, we can see that infalling and orbiting populations both contribute on smaller scales as one would expect, with the fractional abundance of orbiting sub-halos dominating. As  $r$  increases, the fractional abundance of both infalling and orbiting sub-halos decreases, and we only find background sub-halos beyond the turnaround radius  $r_{ta}$  by construction.

Since the peculiar velocity difference distribution is well-described by a weighted sum of two Gaussians as shown in section 3.2.1, up to the turnaround radius, it appears that the contribution of the broad component that has a relatively large weight can be associated with orbiting sub-halos. Processes like tidal stripping and dynamical friction that remove mass generally randomize the motions of orbiting sub-halos, which results in a wider spread in velocities. On the other hand, the narrower Gaussian component can be associated with infalling sub-halos. The infalling sub-halos generally have more coherent radial motions as they have not yet been strongly affected by the halo group, which could result in a much narrower velocity dispersion. This means that the combined contribution of two different dynamical sub-halo types at any  $r < r_{ta}$  is why we require a weighted sum of two Gaussians with different dispersions, where the infalling and orbiting sub-halos affect the peak and tail of the distribution respectively. At scales beyond the turnaround around, since we only have one dynamical type, this explains why the weight  $\lambda$  of the narrower Gaussian approaches zero as discussed before, suggesting that the velocity distribution of background sub-halos can be described mostly by the broader Gaussian component in the mixture.



**Figure 3.7:** Fractional abundances of orbiting sub-halos (red), infalling sub-halos (blue) and background sub-halos (black) as a function of scale for different halo group mass bins

It is important to acknowledge that the relatively simple classification scheme adopted here rather than using halo merger trees could have led to incorrect classification of some sub-halos. For example, we could have sub-halos that are currently infalling but were stripped of the mass in another group. The classification based on  $m/m_{max}$  could incorrectly identify these sub-halos as orbiting sub-halos. Additionally, estimating  $r_{ta}$  by averaging radial velocities over many halo groups within a mass bin neglects real differences between individual halo groups and is affected by the finite number of sub-halos and choice of radial bin width. This means that the classification of some sub-halos with radial positions very close to the average  $r_{ta}$  could be incorrect. Nevertheless, this simplified scheme provides a first-order idea of the physical motivation behind the weighted sum of two Gaussians in the one-halo conditional distribution.

### 3.2.3 Joint modelling

We can now extend the working velocity difference distribution model at fixed scales to also include scale dependence. This can be done by fitting smooth functions to the three parameters  $\sigma_1$ ,  $\sigma_2$ , and  $\lambda$  as seen in equation 3.5, and e.g. 3.5 will be referred to as the joint distribution. After trying different versions for the smooth functions across different halo group mass bins, we find that equations 3.6, 3.7, and 3.8 work well for  $\sigma_1$ ,  $\sigma_2$  and  $\lambda$  respectively. Tables 3.8 and 3.9 provides the best-fit parameters for equations 3.6, 3.7, and 3.8. Figure 3.8 shows the distributions marginalized over the full range of scales ( $r = 0$  to  $r_{max}$  based on the maximum separation of sub-halos in each halo group mass bin) for the four different mass bins considered, so the underlying data in 3.8 is identical to the histograms presented in section 3.1. The marginalized distributions show that the best-fit models perform well on average across the different halo group mass bins.

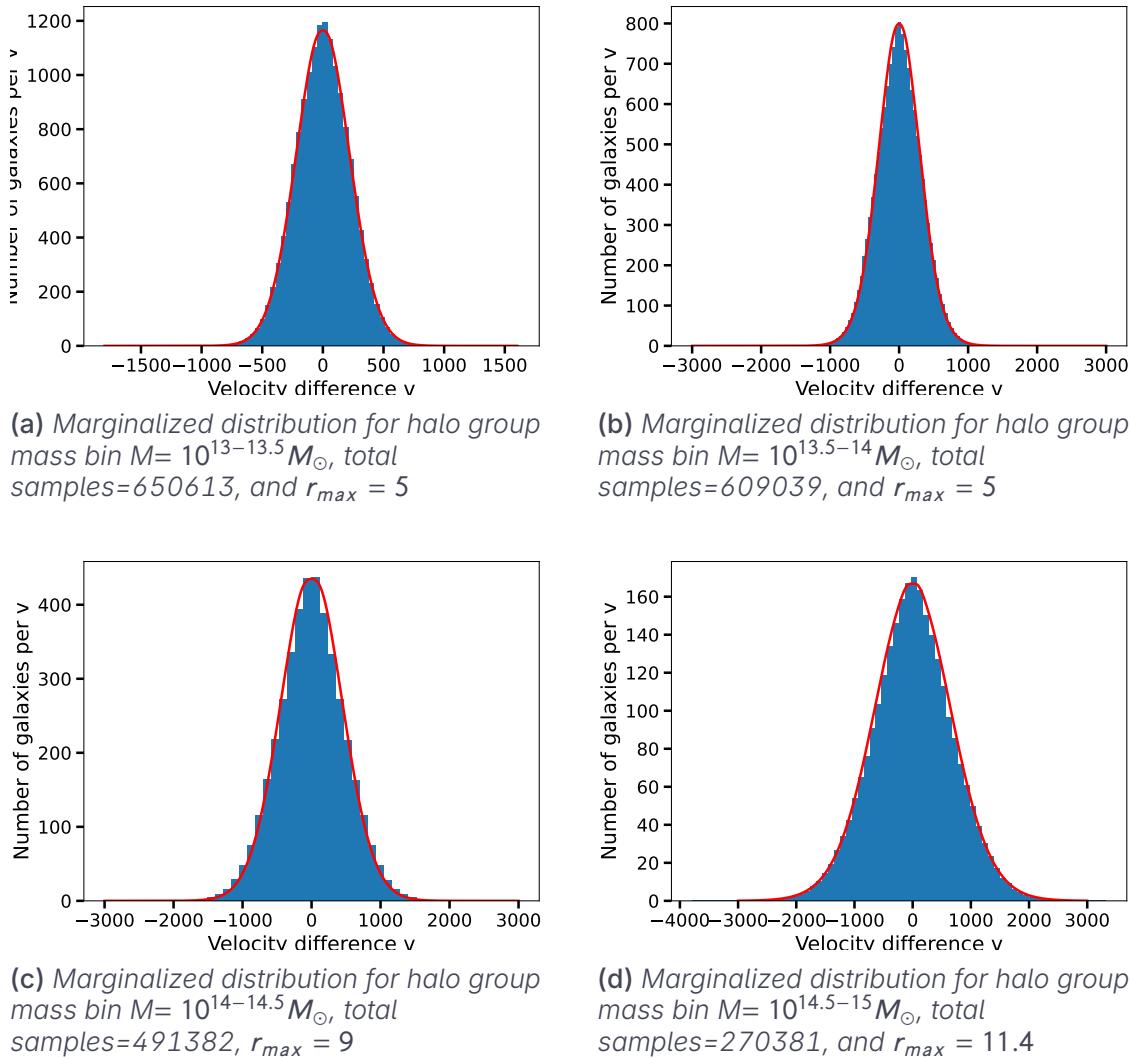
$$f(v, r) = \frac{1}{\sqrt{2\pi}} \left[ \frac{1 - \lambda(r)}{\sigma_1(r)} \exp\left(-\frac{v^2}{2\sigma_1(r)^2}\right) + \frac{\lambda(r)}{\sigma_2(r)} \exp\left(-\frac{v^2}{2\sigma_2(r)^2}\right) \right] \quad (3.5)$$

$$\sigma(r) = p r^n + q r + b \quad (3.6)$$

$$\sigma_1(r) = m r + c \quad (3.7)$$

$$\lambda(r) = A e^{-B r} + C \quad (3.8)$$

Additionally, we also show representative bin-width marginalized distributions in figures 3.9 and 3.10, which are one-dimensional distributions produced by marginalizing the joint distribution over a small range of  $r$ . The bin-width marginalized distributions are identical to the distributions produced in section 3.2.1, allowing us to validate the joint distribution by checking if they can reproduce the conditional distributions at different fixed scales. Again, we can see that the joint distribution model can reproduce the underlying characteristics of the data at different fixed scales for all four halo group mass bins. It must be noted that the performance of marginalization of the joint distribution was observed to slightly become worse at larger separations due lower number of sub-halos, but the overall joint distribution model remains an adequate description for most scales of interest for the one-halo regime.



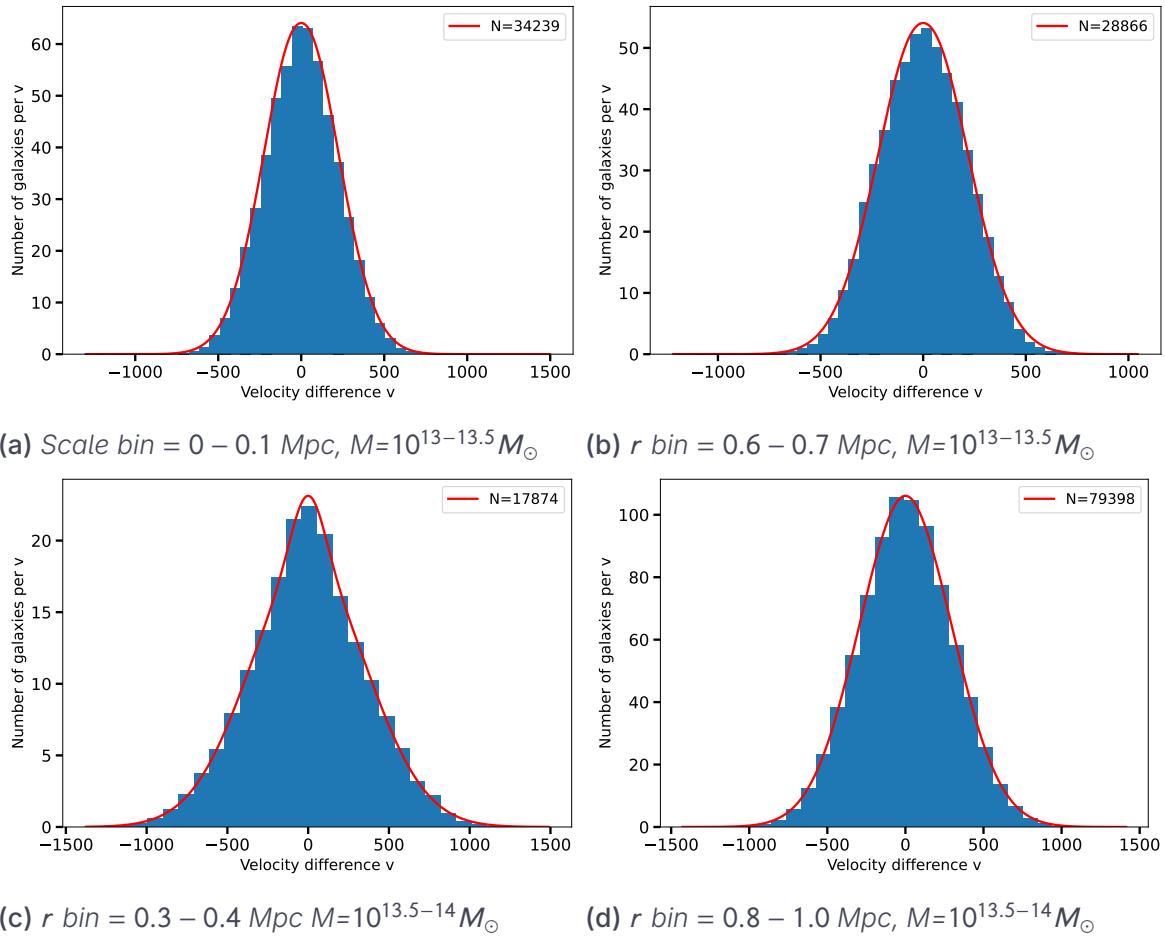
**Figure 3.8:** Marginalized distributions for all halo group mass bins from 0 to maximum separation of sub-halos  $r_{max}$

It is very important to note that the defined joint distribution of velocity difference and scale  $f(v, r)$  encodes clustering information. This is because the amplitude of  $f(v, r)$  at fixed scales in this work depends on the total number of sub-halos, which is proportional to  $1 + \xi(r)$  at fixed scales. So,  $f(v, r)$  simultaneously contains information on both real-space clustering statistics and the velocity difference distribution, making it incredibly useful to calculate the redshift-space correlation function or the power spectrum. For example, under the assumption of statistical isotropy, the 1-halo redshift-space two-point correlation function can be calculated using the halo and streaming model presented in Hang et al. (2022) with equation 3.9, where  $r_{||}$  is the line-of-sight component of the real separation  $r$ ,  $s_{||}$  and  $s_{\perp}$  are the line-of-sight and transverse components of the redshift-space separation  $s$  respectively,  $\xi_{1h}^s$  is the one-halo redshift-space

correlation function, and  $r$  is the magnitude of the 3D separation vector.

$$1 + \xi_{1h}^s(s_\perp, s_\parallel) = \int_{-\infty}^{\infty} dr_\parallel f(v_\parallel, r) \quad (3.9)$$

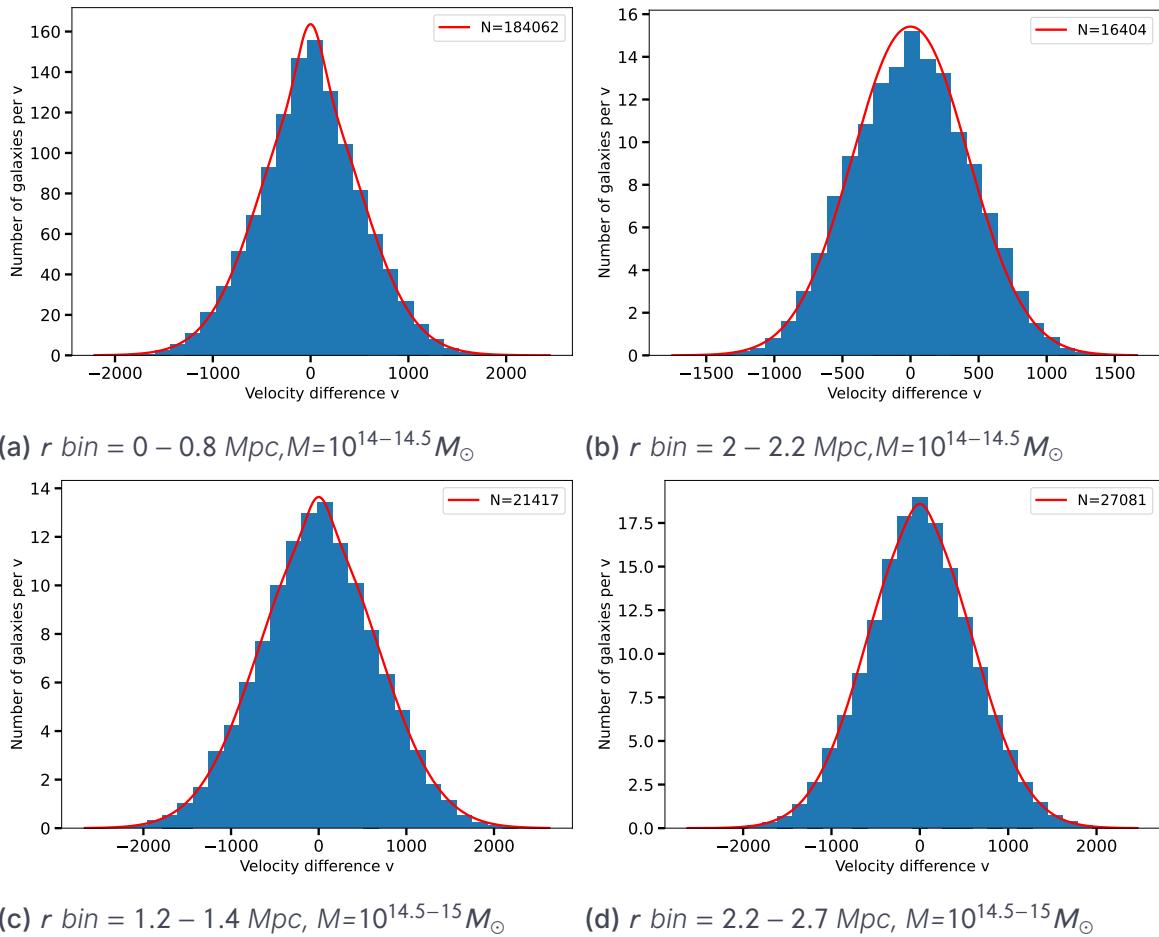
Thus, we now have the one-halo conditional and joint peculiar velocity difference distributions, both of which are unified models that work across different mass bins. The analytical forms of the distributions in this work were obtained after careful examination of the full velocity distributions. The obtained conditional and joint distribution models capture the underlying distribution while being as simple as possible. With the two functions, we can construct a full and exact mapping of the one-halo contribution of the power spectrum or the two-point correlation function of galaxies from real space to redshift-space after careful consideration of the mass dependence, improving on current methods of redshift-space distortion modelling.



**Figure 3.9:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation for halo group mass bin  $M=10^{13-13.5} M_\odot$  and  $M=10^{13.5-14} M_\odot$ . The red line in each sub-figure indicates the radial bin-width marginalized distribution of 3.5 in a specific radial bin, and  $N$  specifies the total number of samples

Mass bin $M_{\odot}$	$p$	$n$	$q$	$b$	$m$	$c$
$10^{13-13.5}$	-227.7	0.22	37.6	396.4	7.8	70.9
$10^{13.5-14}$	511.3	1.15	-653.4	437	-9.42	104.7
$10^{14-14.5}$	-505.3	0.19	40	921.9	-4.6	125.7
$10^{14.5-15}$	-554.6	0.7	262.4	1007.8	-7.9	140

**Table 3.8:** Best-fit parameter values for equations 3.6 and 3.7 for different halo group mass bins.



**Figure 3.10:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation for halo group mass bin  $M=10^{14-14.5} M_{\odot}$  and  $M=10^{14.5-15} M_{\odot}$ . The red line in each sub-figure indicates the radial bin-width marginalized distribution of 3.5 in a specific radial bin, and  $N$  specifies the total number of samples

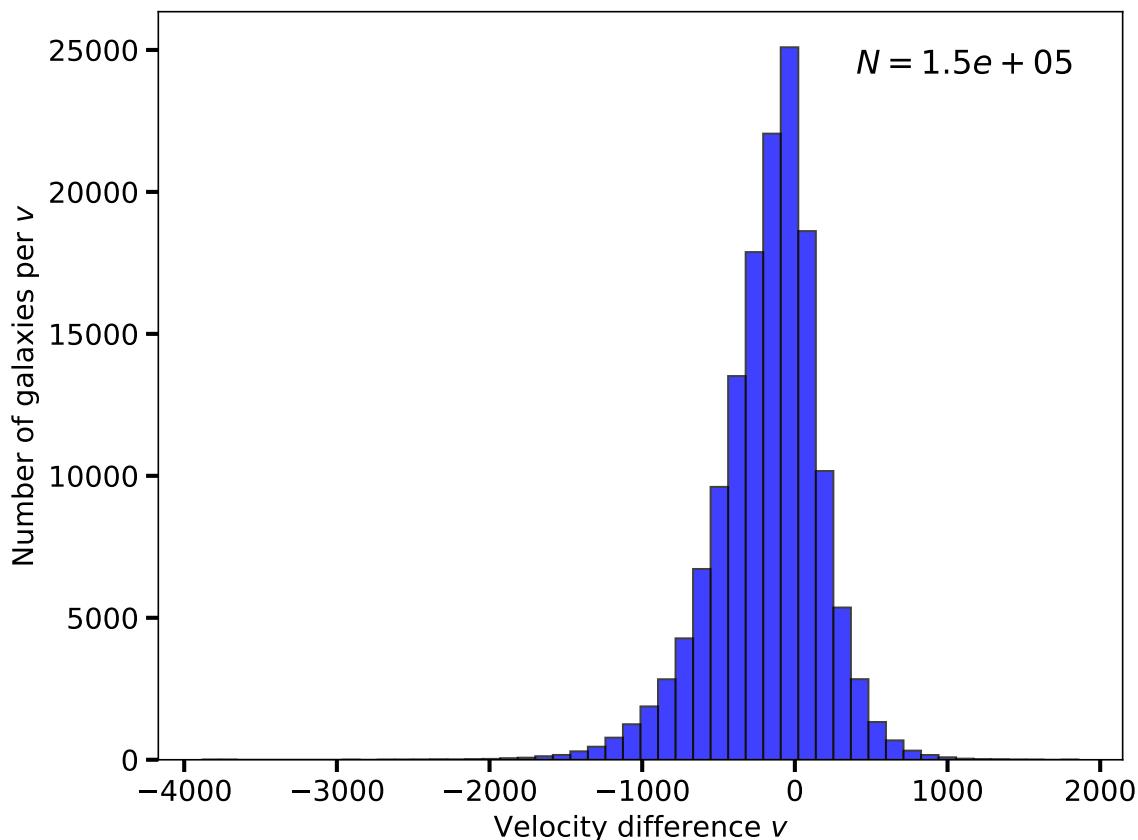
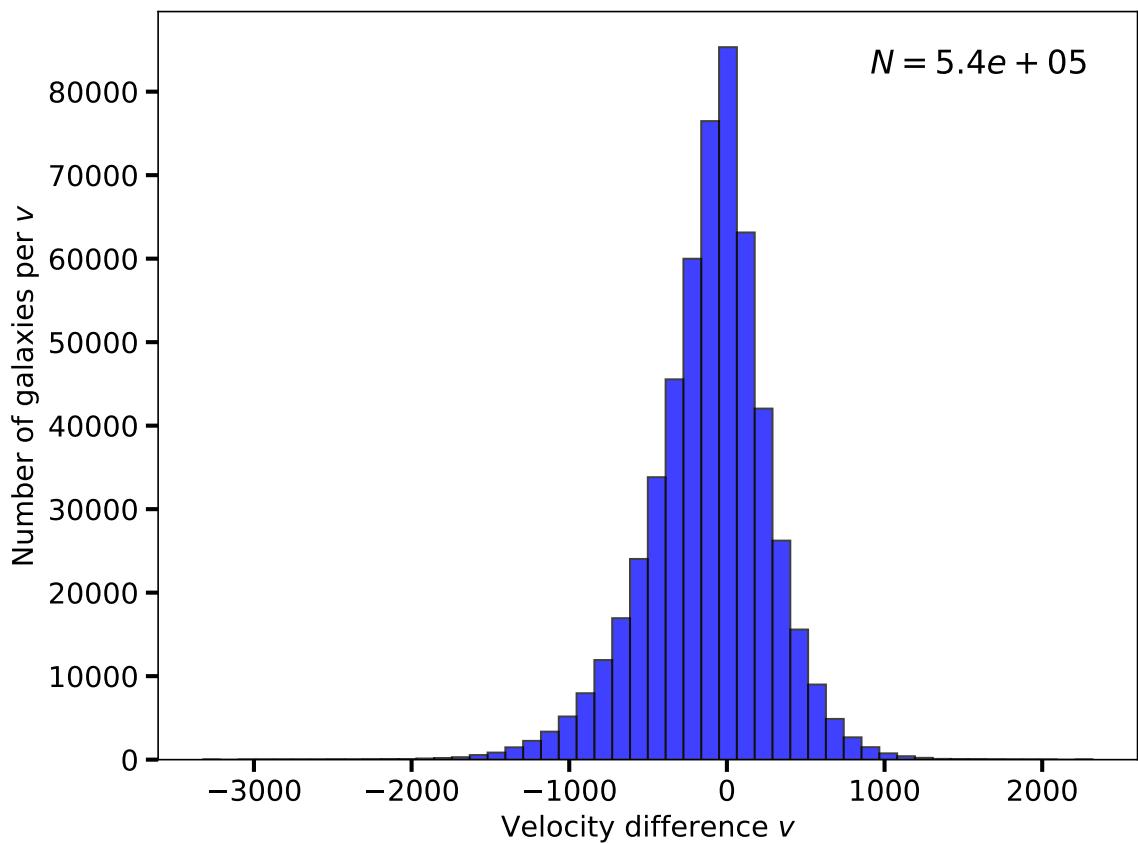
Mass bin $M_{\odot}$	<i>A</i>	<i>B</i>	<i>C</i>
$10^{13-13.5}$	0.69	20	-0.005
$10^{13.5-14}$	0.22	3.03	-0.009
$10^{14-14.5}$	0.1	1	-0.015
$10^{14.5-15}$	0.046	0.45	-0.005

*Table 3.9:* Best-fit parameter values for *A*, *B*, and *C* in equation 3.8 for different halo group mass bins.

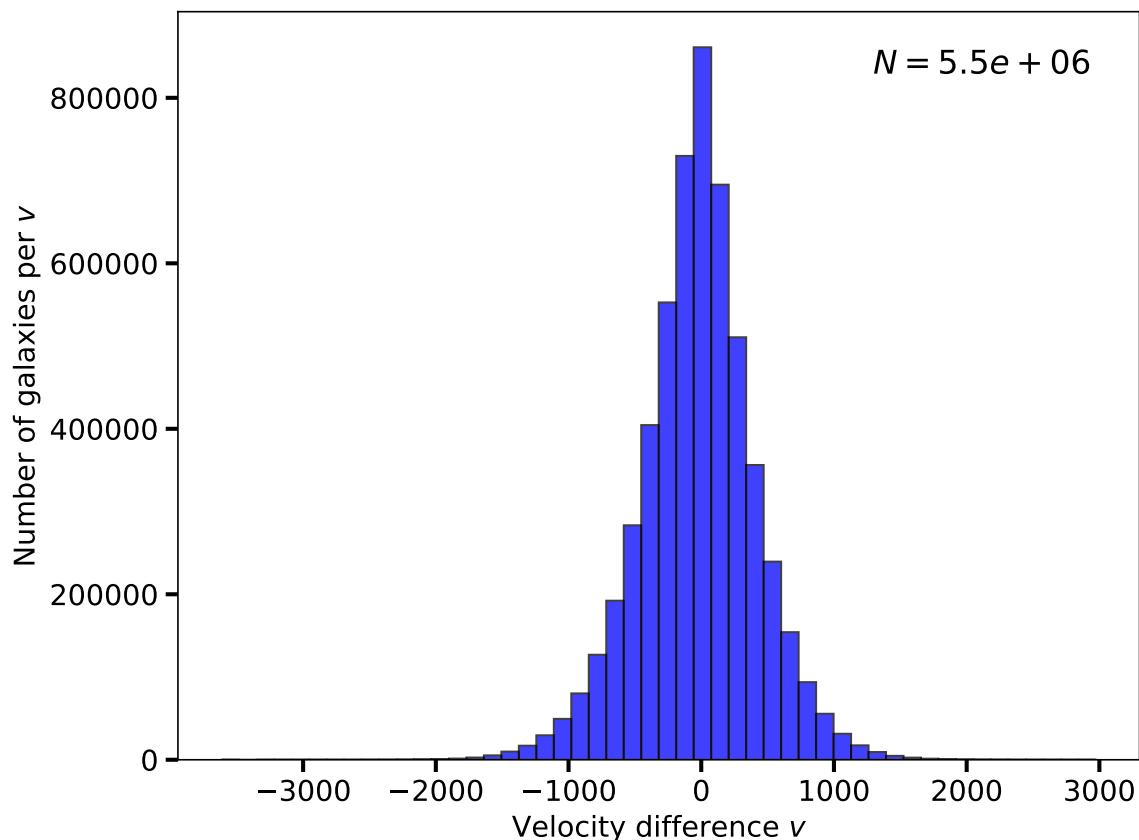
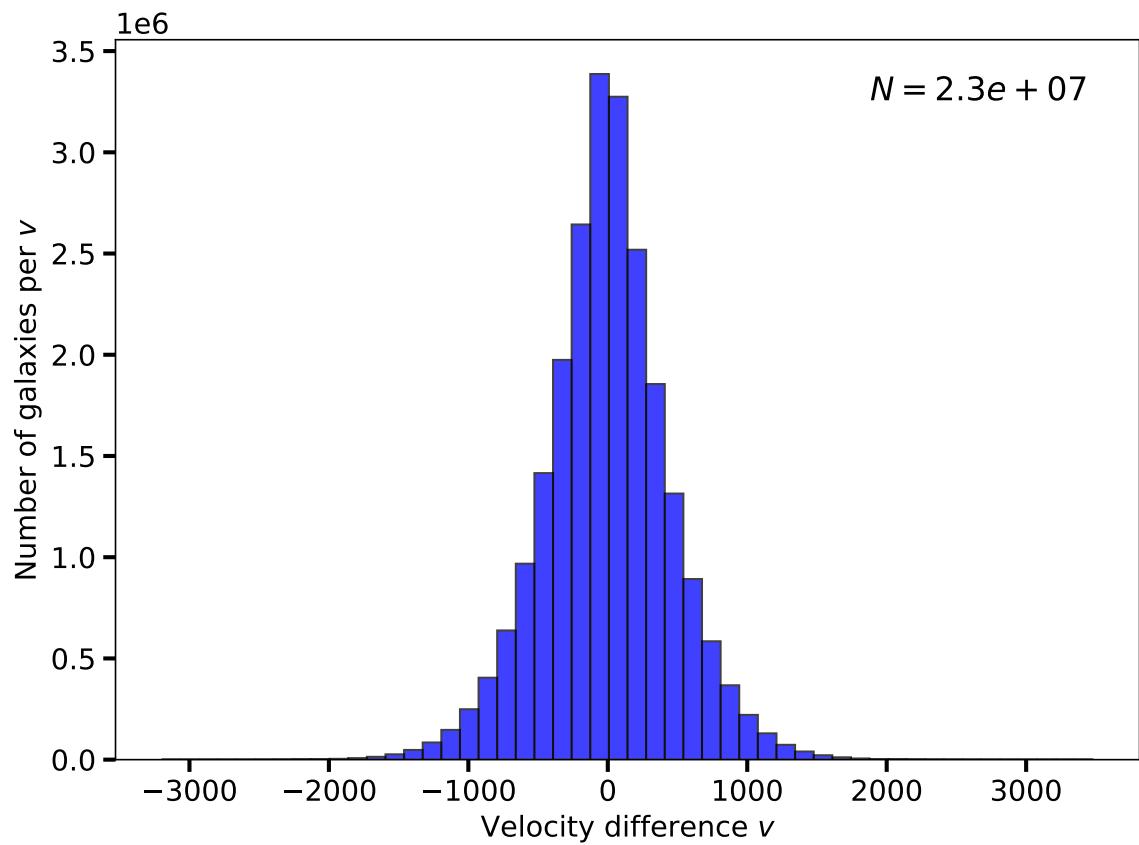
# 4 . Two-halo distributions

This chapter focuses on the velocity difference distribution of pairs of halo groups that is relevant for the two-halo contribution to the power spectrum or the correlation function. Here, the velocity quantity being utilized is slightly different from that of the one-halo distribution presented in chapter 3. As explained in the methods chapter 2, we utilize the component of the velocity difference vector projected along the 3D separation vector. This is different from the conventional quantity  $v_r$  used in typical RSD models such as the streaming model (see e.g. Reid & White 2011), where  $r$  actually represents the line-of-sight direction. As explained before, the distant observer approximation breaks down on very large scales, so we choose to utilize the velocity component projected along the separation of each pair. According to the adopted definition of  $v$  for the 2-halo distribution in this work,  $v < 0$  corresponds to pairs that are moving towards each other, and  $v > 0$  corresponds to pairs that are away from each other along the separation vector.

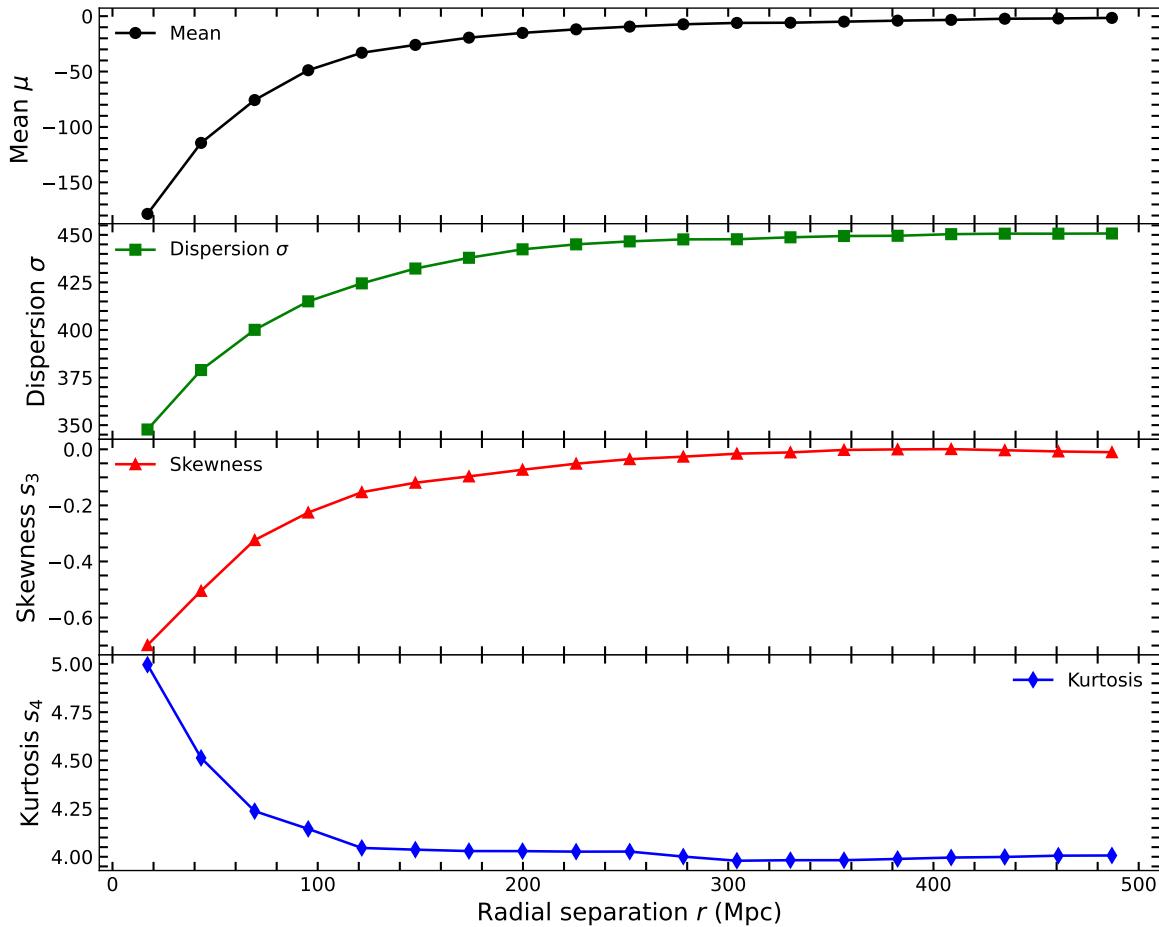
Although we successfully identified a model that performs well on the one-halo distribution, modelling the two-halo distribution has proven significantly more challenging. To better understand the underlying causes of this difficulty, this chapter focuses on examining the fundamental characteristics of the data distribution itself. Firstly, considering the 2-halo distributions at different scales for  $M_1 = 10^{14-14.5} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$ , we present a few of the histograms in figures 4.1 and 4.2 and the first four moments, the mean, dispersion, skewness, and kurtosis of the velocity difference data in fig. 4.3. Similar to the one-halo distributions, the characteristics of the distributions vary with scale, demonstrating that any two-halo model has to account for scale dependence. Interestingly, on scales up to  $r \sim 300\text{Mpc}$ , the distributions have a negative mean, indicating mean coherent infall towards each other. The mean approaches the limit of zero with an increase in separation. Additionally, the skewness is also negative up to  $r \sim 200\text{Mpc}$  and approaches zero as separation increases. This means that the distribution becomes symmetric at large separations, which is similar to a normal distribution. However, the kurtosis  $s_4 > 4$  for all scales, and the kurtosis of a Gaussian is  $s_4 = 3$ , so the distribution never approaches a Gaussian limit.

(a)  $r$  bin =  $4.0 - 30.1 \text{ Mpc}$ (b)  $r$  bin =  $30.1 - 56.2 \text{ Mpc}$ 

**Figure 4.1:** Two-halo velocity histograms for halo group mass bins of  $M_1 = 10^{14-14.5} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$ ,  $N$  represents the total number of samples in each histogram

(a)  $r$  bin =  $134.5 - 160.6 \text{ Mpc}$ (b)  $r$  bin =  $291.2 - 317.3 \text{ Mpc}$ 

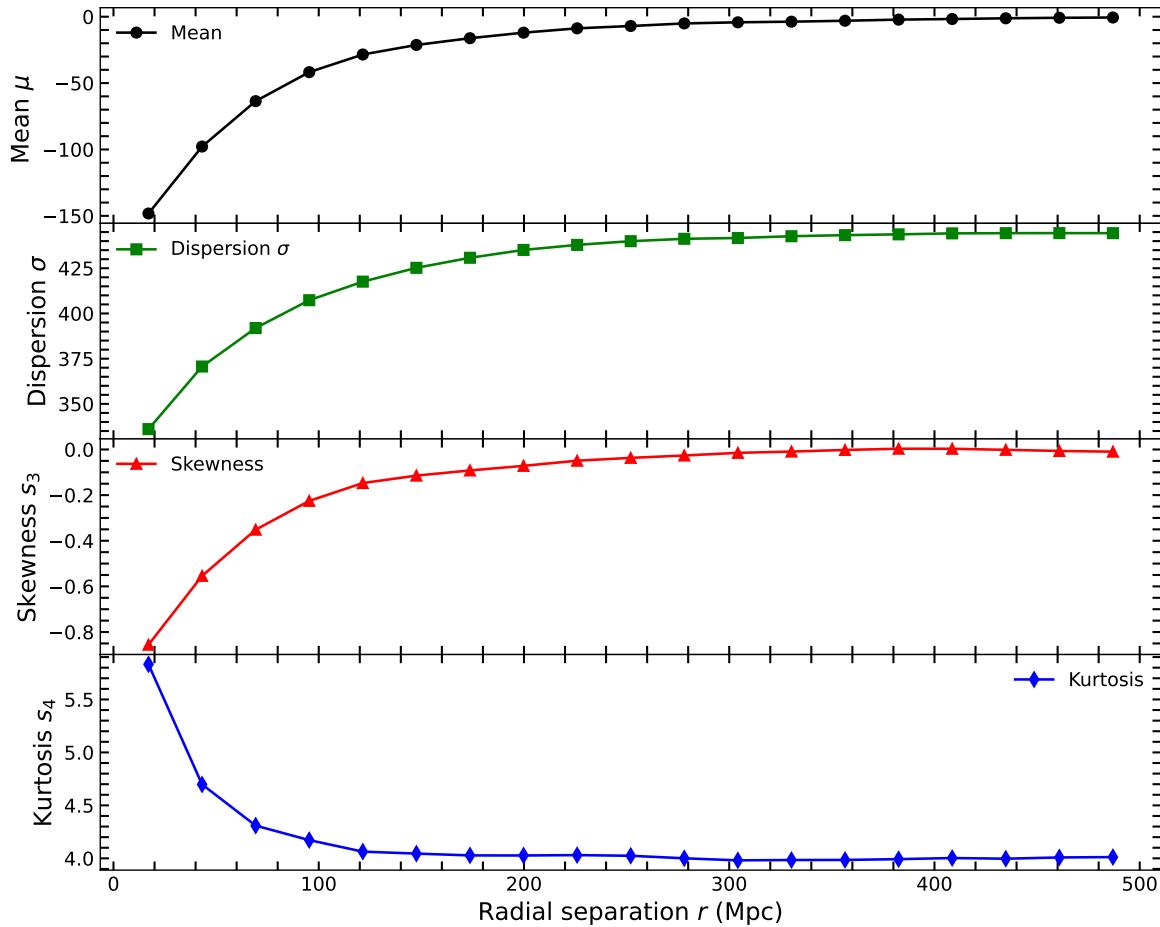
**Figure 4.2:** Two-halo velocity histograms for halo group mass bins of  $M_1 = 10^{14-14.5} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$ ,  $N$  represents the total number of samples in each histogram



**Figure 4.3:** The first four moments, the mean, dispersion, skewness and kurtosis as a function of separation  $r$  for halo group mass bins of  $M_1 = 10^{14-14.5} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$

Similarly, the two-halo distributions for  $M_1 = 10^{13.5-14} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$  also exhibit similar characteristics of negative skewness and non-zero mean up to intermediate scales, both approaching the limit of zero for large separations and a non-Gaussian kurtosis  $s_4 > 4$  on all scales as seen in figures 4.5, 4.7, and 4.6. Together, the moments as seen in fig. 4.4 highlight how a Gaussian is insufficient on all scales, and a Gaussian limit is never achieved even at the largest scales. This means that the assumption of a Gaussian velocity distribution in the Gaussian streaming model is also not accurate enough when we consider halo pairs. The observed patterns of non-zero higher order moments, negative skewness that decreases with scale and  $s_4 > 4$  in this work match the findings of e.g. Scoccimarro (2004) and Kuruvilla & Porciani (2018). This means that an ideal 2-halo distribution function must allow for variable mean, skewness, and be leptokurtic, i.e., allow for a higher peak than a Gaussian and heavier tails. These moment characteristics could be well-described by the previously discussed complicated distributions such as generalized hyperbolic distributions and skew-T distributions that have been used with the streaming model (Kuruvilla

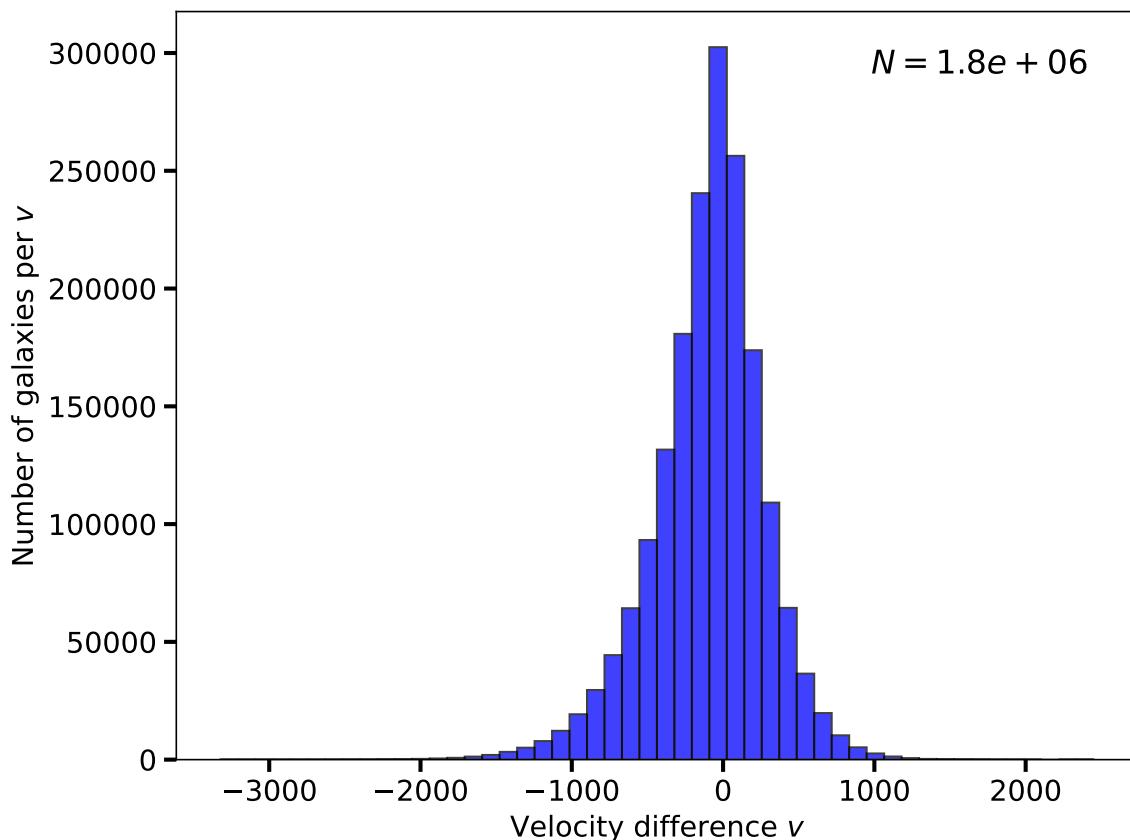
& Porciani 2018; Cuesta-Lazaro et al. 2020). However, since the models are used as conditional distributions (distributions for the velocity difference quantity at fixed scales) within the framework of the streaming model, the use of the model is further complicated by the dependence of each parameter on scale, mass  $M_1$ , mass  $M_2$ , and potentially even redshift  $z$ . So, it is worth exploring other models that might reduce the complexity of existing models while faithfully reproducing the observed characteristics.



**Figure 4.4:** The first four moments, the mean, dispersion, skewness and kurtosis as a function of separation  $r$  for halo group mass bins of  $M_1 = 10^{13.5-14} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$

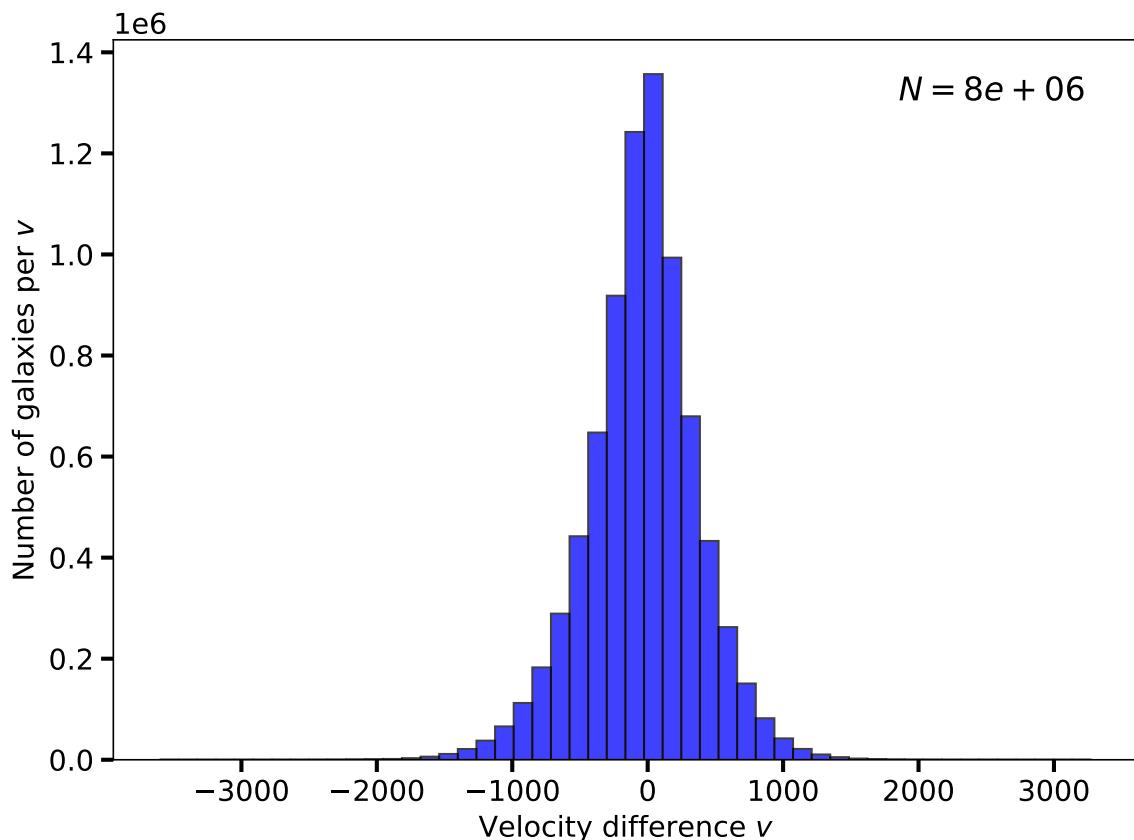
Physically, on small and intermediate scales, the observed negative skewness reflects that pairs of halos are more likely to move toward each other rather than away from each other. This can be understood by examining the scale dependence of clustering. The clustering strength (described by  $\xi(r)$ ) increases with decreasing scale for tracers like halos (Mo et al., 2010). This means that on smaller scales, pairs of halos are more likely to be found in high-density regions rather than low-density regions, so they are more likely to experience coherent infall toward each other due to gravitational attraction. This contributes

to the negative mean pairwise velocity and the skewness observed in the velocity distribution (Tinker, 2007). However, since the correlation between halos decreases with an increase in scale, pairs do not have a preferential direction of motion anymore, so the velocity difference distribution becomes more symmetric. Additionally, a non-Gaussian kurtosis *on all scales* could be because the final velocity difference distribution is a possible superposition of distributions of pairs in different density regions, and non-linear effects inside each halo (Tinker 2007; Scoccimarro 2004)



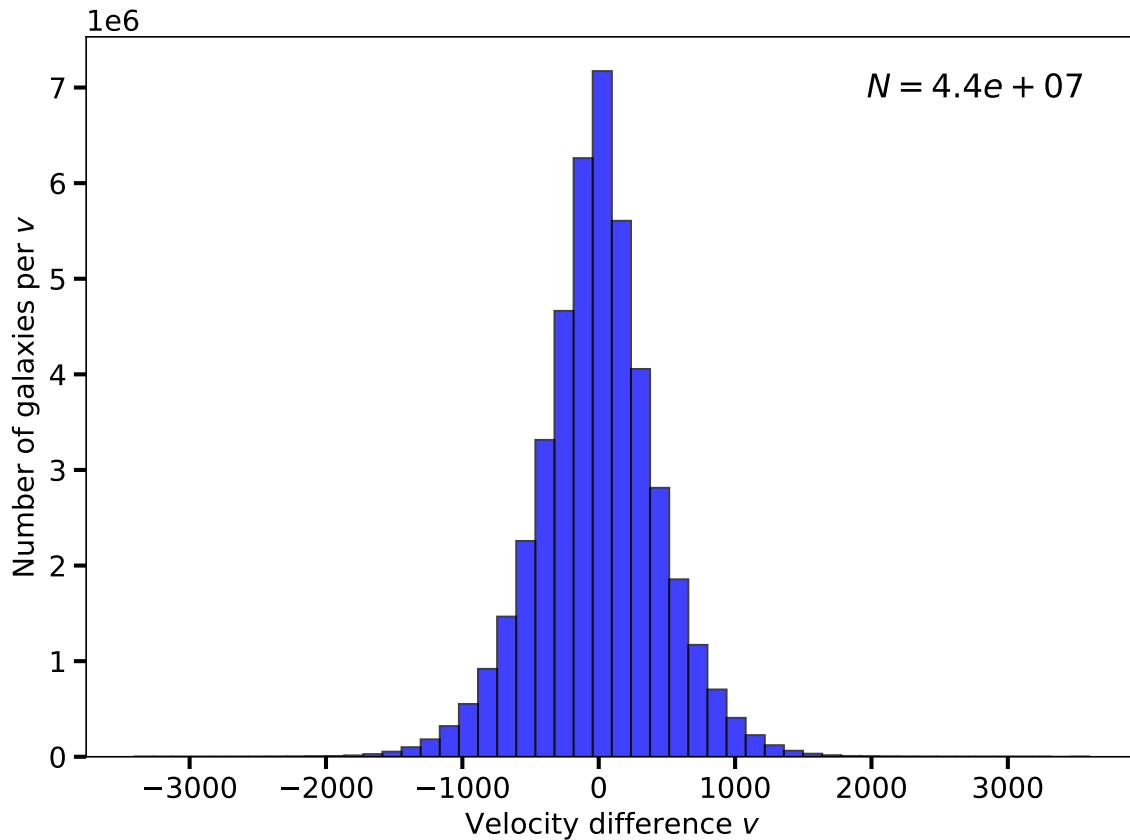
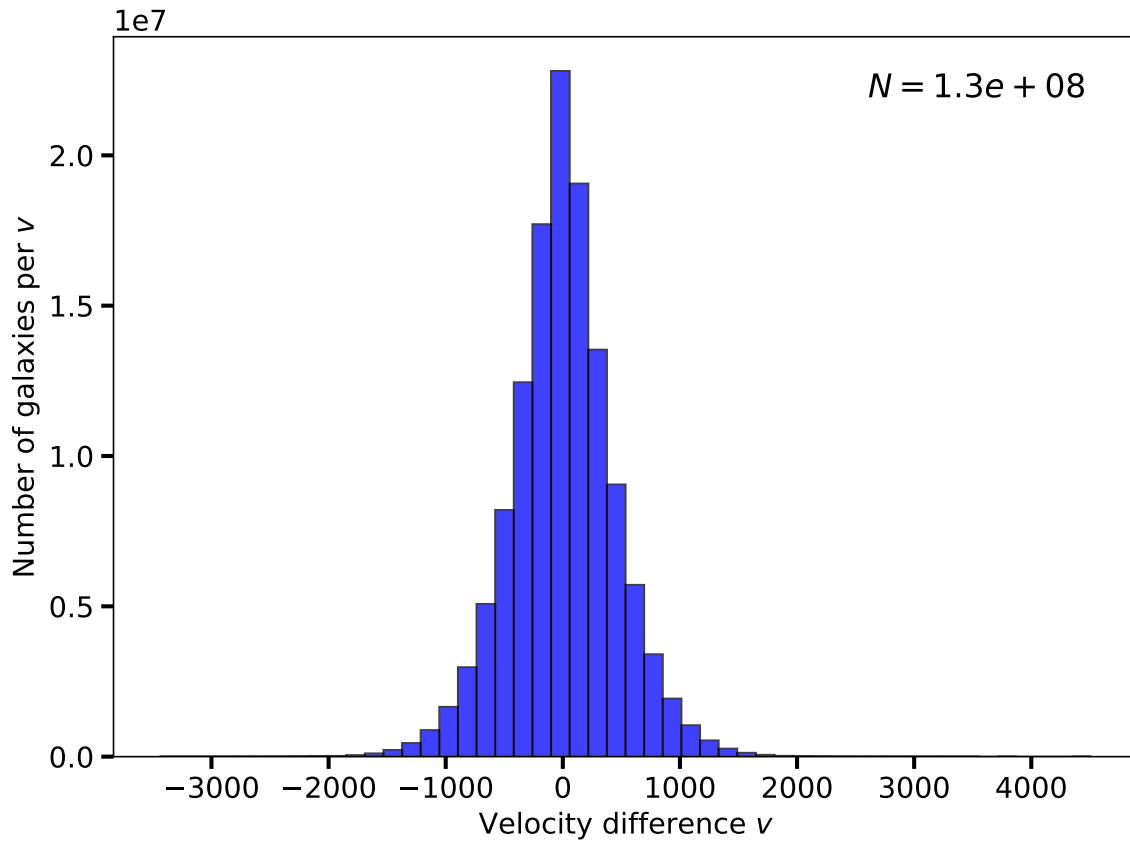
(a)  $r$  bin =  $30.1 - 56.2 \text{ Mpc}$

**Figure 4.5:** Two-halo velocity histograms for halo group mass bins of  $M_1 = 10^{13.5-14} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$ ,  $N$  represents the total number of samples in each histogram



(a)  $r$  bin = 82.3 – 108.4 Mpc

**Figure 4.6:** Two-halo velocity histograms for halo group mass bins of  $M_1 = 10^{13.5-14} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$ ,  $N$  represents the total number of samples in each histogram

(a)  $r$  bin =  $212.8 - 212.8 \text{ Mpc}$ (b)  $r$  bin =  $369.5 - 395.6 \text{ Mpc}$ 

**Figure 4.7:** Two-halo velocity histograms for halo group mass bins of  $M_1 = 10^{13.5-14} M_\odot$  and  $M_2 = 10^{13.5-15.5} M_\odot$ ,  $N$  represents the total number of samples in each histogram

# 5 . Conclusion

This thesis aimed at improving redshift-space distortion modelling by attempting to find well-describing analytical velocity-difference models for the one-halo and two-halo distributions using the hydrodynamical, cosmological simulation FLAMINGO. We summarize the main findings of this thesis below.

Starting with the one-halo distributions, we find that commonly used scale-independent analytical models such as a Gaussian or a Laplacian distribution fail to capture the characteristics of a Gaussian-like tail with a more pronounced peak simultaneously across 4 different halo group mass bins  $M = 10^{13-15} M_\odot$ . We identify that the assumption of scale independence is invalid, as the characteristics of the velocity difference distribution vary with scale as seen in section 3.2.1. Using the full velocity difference information available in each halo group mass bin, we find that a weighted sum of two zero-mean Gaussians with different widths,  $\sigma_1$  and  $\sigma_2$ , provides an excellent fit at fixed scales. The most commonly used simple scale-dependent Gaussian is therefore insufficient for accurate modelling of the one-halo distribution. The weighted sum of two Gaussians also works across different mass bins at fixed scales, thereby providing a unified one-halo conditional distribution. The weighted sum captures the observed non-Gaussian kurtosis of the underlying distribution across different scales very well.

We also find that our one-halo model is physically motivated. By classifying the sub-halos in each halo group mass bin into infalling, orbiting, and background sub-halos based on a simple classification scheme, we find that in the smaller radial bins, both orbiting and infalling sub-halos contribute to the observed velocity difference distribution. As we cross the turnaround radius, only background sub-halos contribute to the velocity difference distribution. This helps us understand why the weight  $\lambda$  of the narrow Gaussian component declines with increasing separation. In smaller radial bins, the narrower Gaussian component reflects the coherent, inward radial motions of infalling sub-halos, while the broader Gaussian component captures the randomized velocities of orbiting, mass-stripped sub-halos. Beyond the turnaround radius, the broader Gaussian again captures the randomized velocities of background sub-halos. This distinguishes our model from more complex and less physically motivated extensions of the Gaussian streaming model, such as a generalized hyperbolic

distribution (Kuruvilla & Porciani, 2018).

We then found the joint probability distribution  $f(v, r)$  by fitting smooth functions to  $\sigma_1(r)$ ,  $\sigma_2(r)$ , and  $\lambda(r)$ . This joint form not only reproduces both the full range marginalized velocity distribution and the narrow bin conditional histograms, but it also explicitly encodes clustering information. Thus, calculations of the one halo redshift space correlation function or the power spectrum are straightforward and analytic, enabling more efficient redshift - space distortion modelling.

The obtained one-halo conditional distribution contrasts with distributions presented in extensions of the Gaussian streaming model extensions such as those by Bianchi et al. (2016), Kuruvilla & Porciani (2018) and Cuesta-Lazaro et al. (2020), and this is because the velocity distributions were examined for pairs of halo groups, which is the two-halo distribution. Even if the one-halo contribution were to be considered, streaming models aim to simultaneously model the one-halo random motions and two-halo coherent infall, even on small scales where the random motions are dominant. This mixing inevitably produces non-zero skewness on small scales and requires very complex functional forms such as a generalized hyperbolic distribution to accommodate for the non-Gaussian kurtosis and skewness. By separating the two different physical processes, we obtain a simpler, more interpretable, analytical velocity distribution to utilize in redshift - space distortion modelling.

In the two-halo distributions, we see clear scale-dependent non-Gaussian features, including negative skewness and leptokurtic kurtosis across all scales. The negative skewness at smaller scales reflects halo pairs preferentially moving toward each other due to gravitational infall in clustered regions, which weakens with increasing separation. The persistent excess kurtosis suggests the distribution cannot be approximated by a Gaussian at any scale, likely due to a mix of pairs from different density environments and non-linear effects (Tinker, 2007; Scoccimarro, 2004). These results also show that a Gaussian is insufficient for the 2-halo distribution, motivating more complex models like skew-t distributions (Kuruvilla & Porciani, 2018; Cuesta-Lazaro et al., 2020). However, the dependence of model parameters on scale and the mass of each halo group in pairs highlights the difficulty of using such complex modelling and the need for simpler yet accurate alternatives.

There are many ways to build on the work done in this thesis. The most immediate first step in improving on the current work is modelling the mass dependence in the one halo model. This would allow us to directly compute the one-halo power spectrum or correlation function contribution and is a vital step in building a redshift-space distortion model. Our sub-halo classification, based on a single snapshot and mass ratio, could be improved by using halo merger trees. The next step would be finding a (preferably) analytical model for the 2-halo distribution that works across different mass bins and scales. This is a very

---

complex task, but it is necessary in order to construct a better and self-consistent redshift-space distortion model. Additionally, both the one-halo and two-halo distribution analyses were done for  $z = 0$  and for the fiducial cosmology. It would be interesting to see if the distributions vary with redshift and cosmology. We could also examine the effects of varying AGN feedback using FLAMINGO to understand how baryonic effects modify the underlying velocity difference distribution. Together, examining the possible variations with the aforementioned properties would help us obtain a more complete understanding of redshift space distortions.

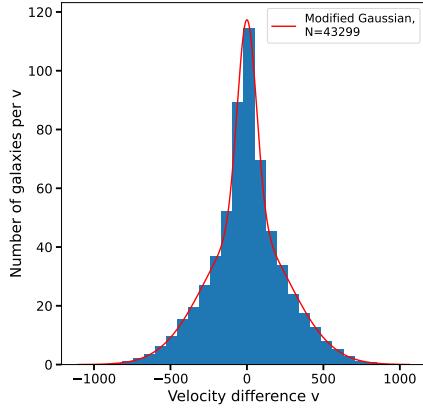
In summary, the Gaussian mixture model we have presented offers a closed-form, physically grounded description of the one-halo velocity difference distribution, that performs better than scale-independent dispersion models, scale-dependent Gaussian, and more elaborate skewed velocity difference distributions. By tying each component to distinct sub-halo populations and deriving an analytic joint distribution  $f(v, r)$ , we provide the necessary tools for precision RSD analysis.

# A . Appendix

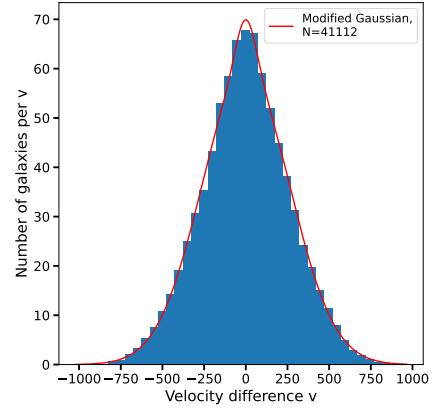
## A.1 Conditional Model Fits

The following graphs were made as a part of fitting equation 3.4 in chapter 3 for different radial bins inside each halo group mass bin. In each graph, N represents the total number of samples, and the red line in each sub-figure indicates the best-fit scale-dependent weighted sum of Gaussians shown in equation 3.4 to the velocity distribution.

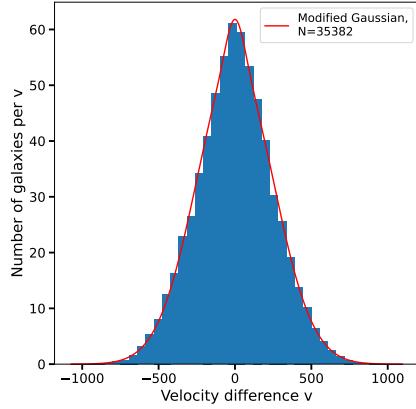
Halo group mass bin  $M=10^{13-13.5} M_{\odot}$ :



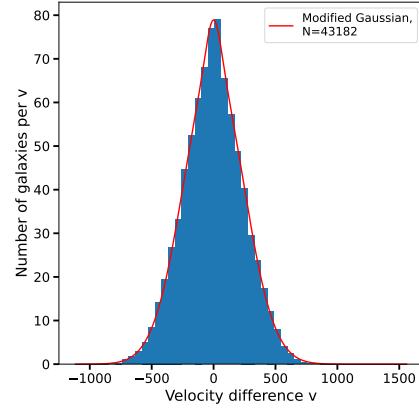
(a) Scale bin = 0 – 0.1 Mpc



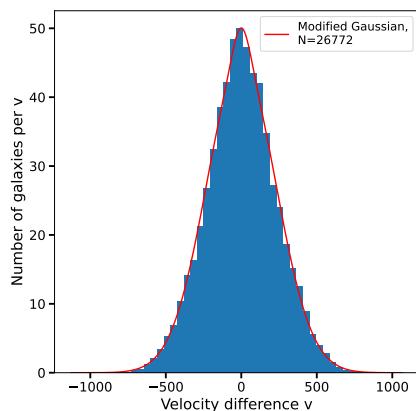
(b) Scale bin = 0.1 – 0.17 Mpc



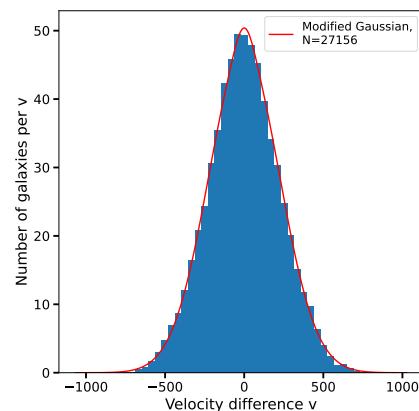
(c) Scale bin = 0.17 – 0.23 Mpc



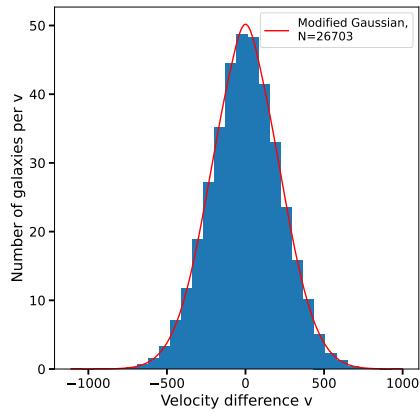
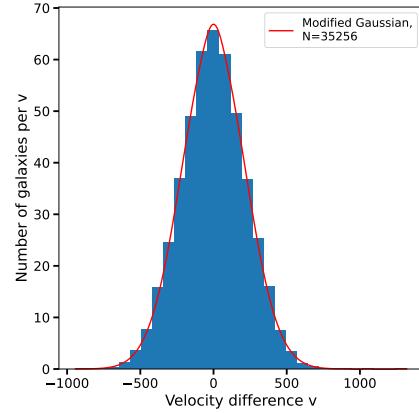
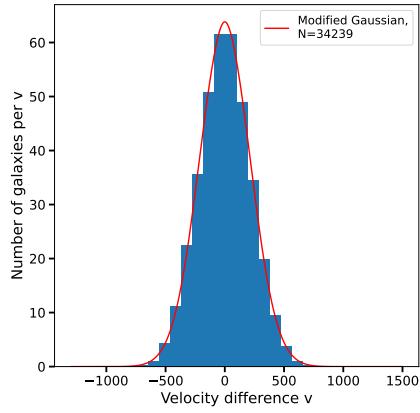
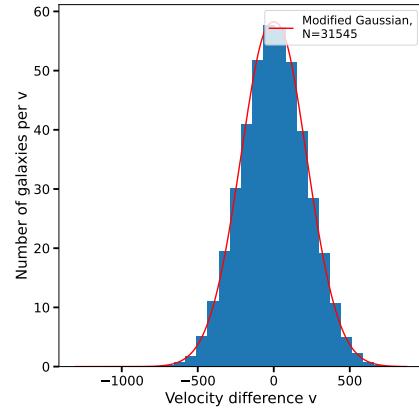
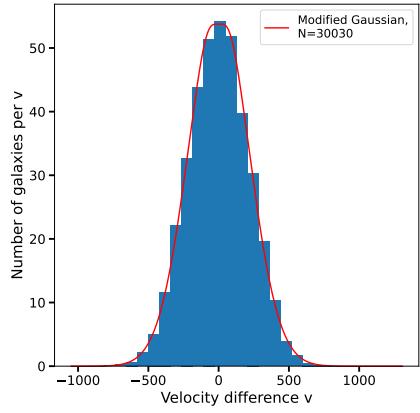
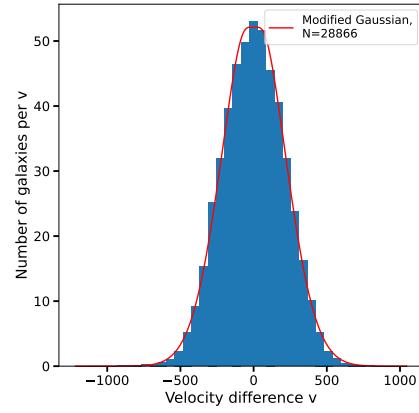
(d) Scale bin = 0.23 – 0.29 Mpc

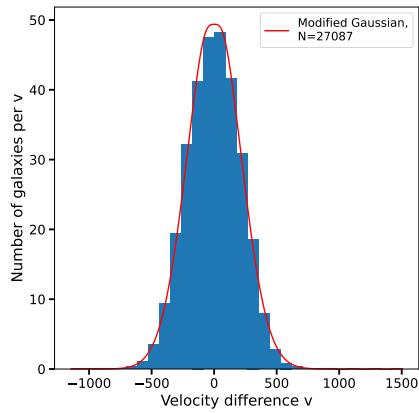
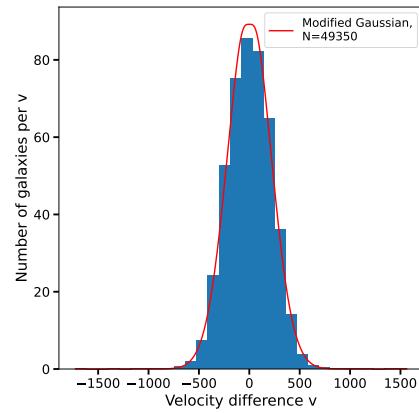
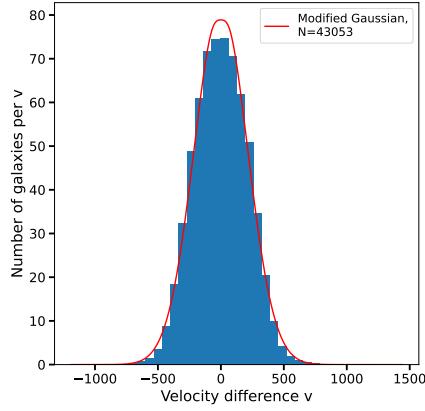
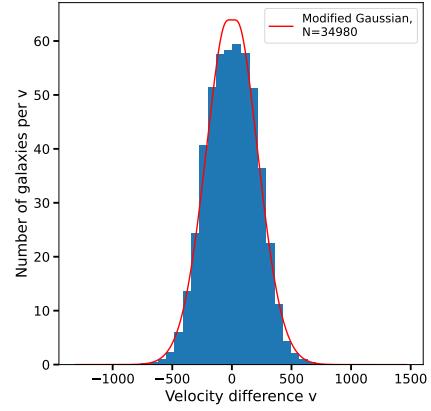
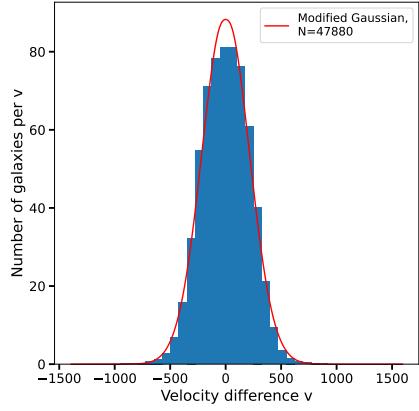
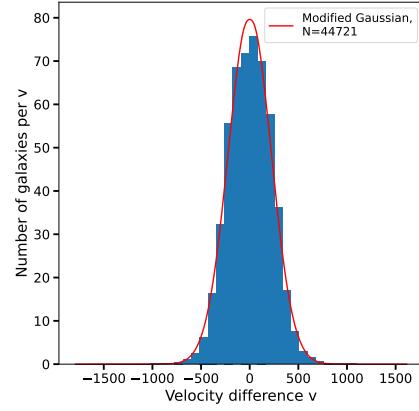


(e) Scale bin = 0.29 – 0.33 Mpc

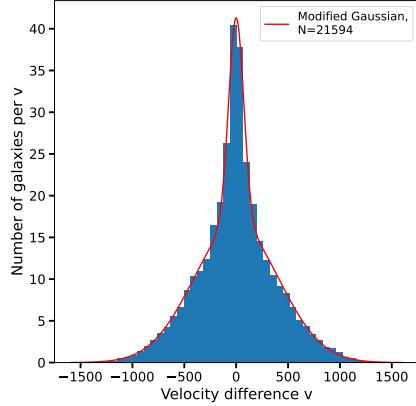


(f) Scale bin = 0.33 – 0.4 Mpc

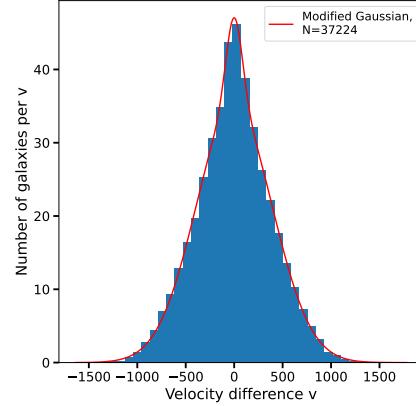
(a) Scale bin =  $0.4 - 0.45$  Mpc(b) Scale bin =  $0.45 - 0.5$  Mpc(c) Scale bin =  $0.45 - 0.5$  Mpc(d) Scale bin =  $0.5 - 0.55$  Mpc(e) Scale bin =  $0.55 - 0.6$  Mpc(f) Scale bin =  $0.6 - 0.65$  Mpc

(a) Scale bin =  $0.65 - 0.7 \text{ Mpc}$ (b) Scale bin =  $0.7 - 0.8 \text{ Mpc}$ (c) Scale bin =  $0.8 - 0.9 \text{ Mpc}$ (d) Scale bin =  $0.9 - 1.0 \text{ Mpc}$ (e) Scale bin =  $1.0 - 1.2 \text{ Mpc}$ (f) Scale bin =  $1.2 - 5 \text{ Mpc}$

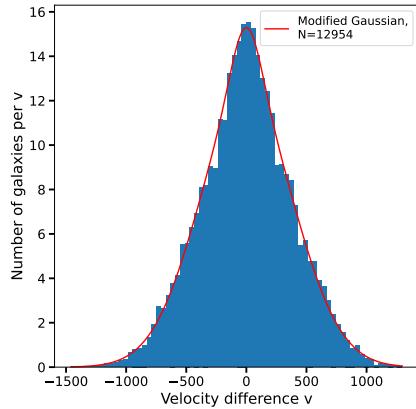
Halo group mass bin  $M=10^{13.5-14} M_{\odot}$ :



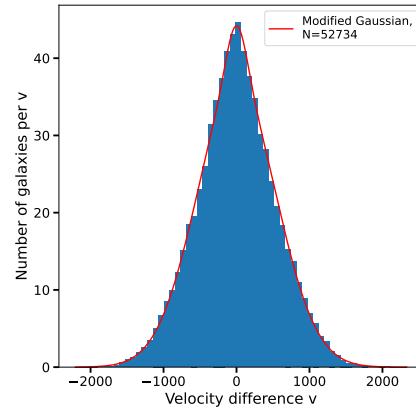
(a) Scale bin = 0 – 0.1 Mpc



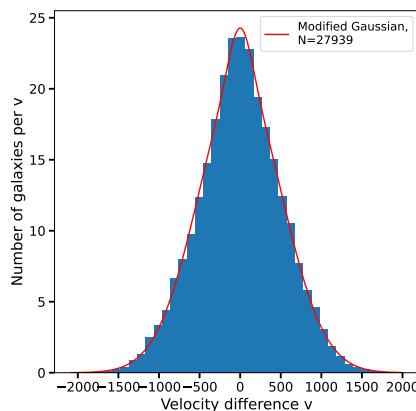
(b) Scale bin = 0.1 – 0.2 Mpc



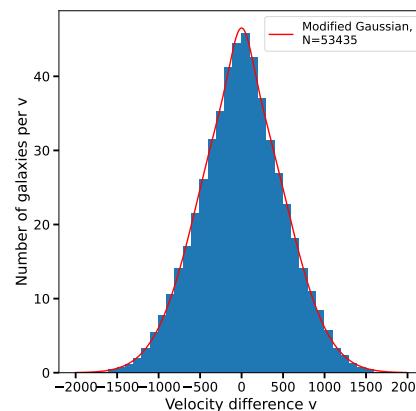
(c) Scale bin = 0.2 – 0.23 Mpc



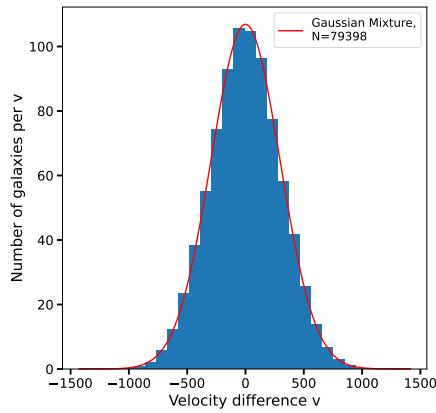
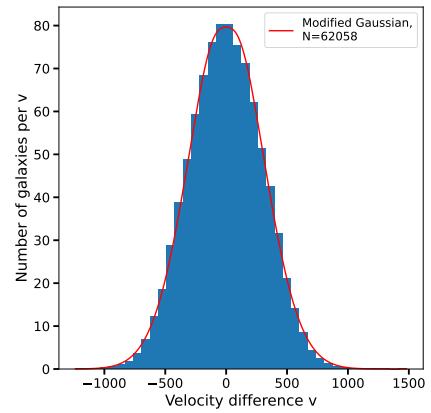
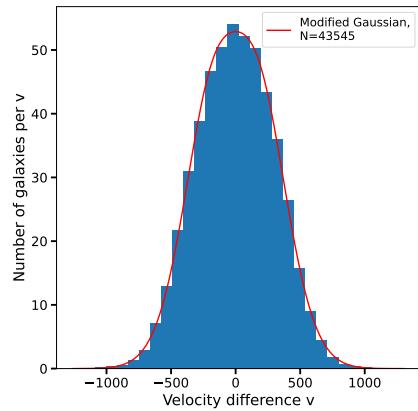
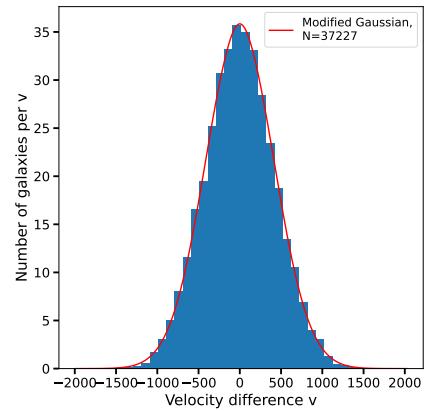
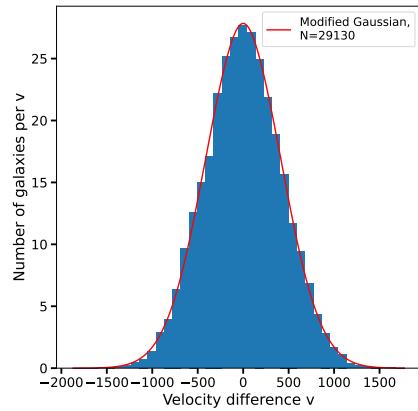
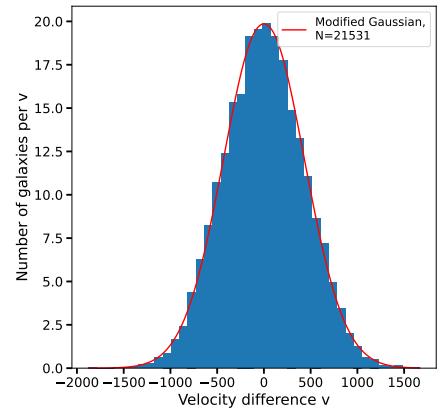
(d) Scale bin = 0.23 – 0.29 Mpc

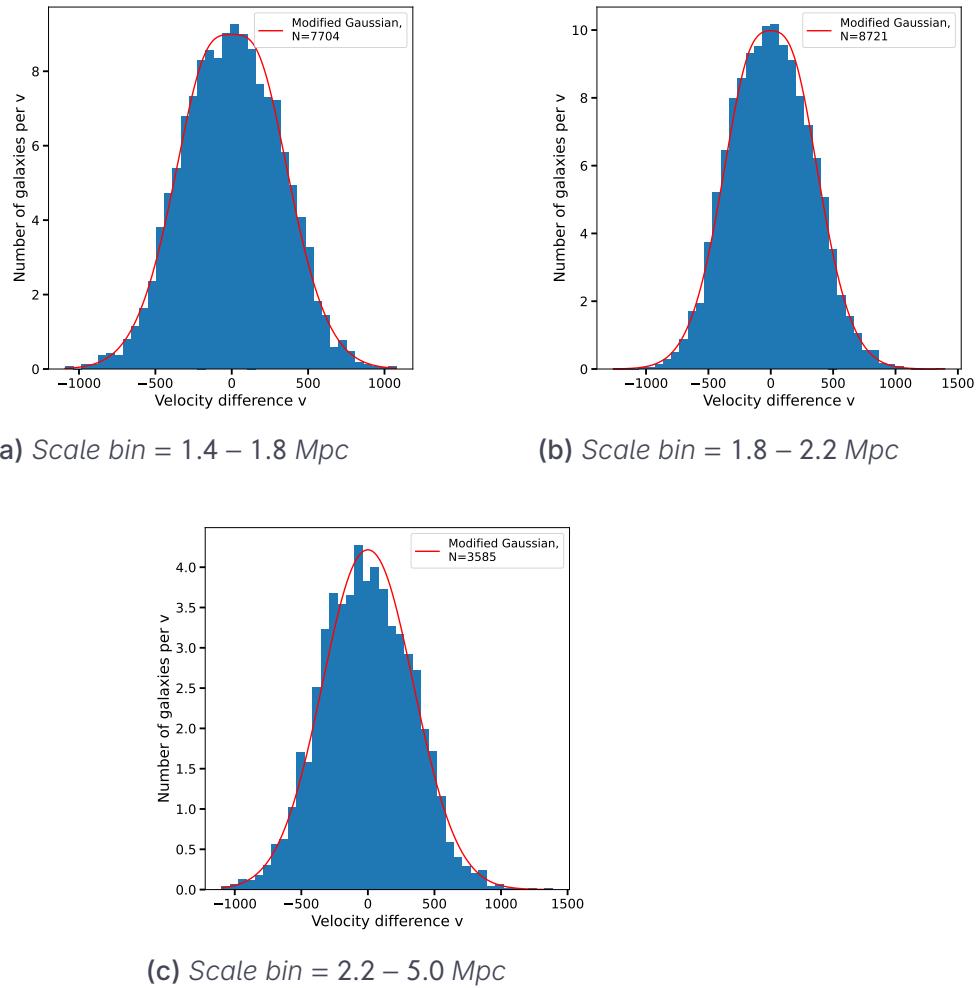


(e) Scale bin = 0.29 – 0.32 Mpc



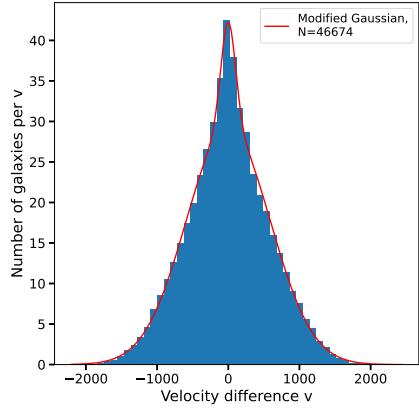
(f) Scale bin = 0.32 – 0.4 Mpc

(a) Scale bin =  $0.4 - 0.5 \text{ Mpc}$ (b) Scale bin =  $0.5 - 0.6 \text{ Mpc}$ (c) Scale bin =  $0.6 - 0.8 \text{ Mpc}$ (d) Scale bin =  $0.8 - 1.0 \text{ Mpc}$ (e) Scale bin =  $1 - 1.2 \text{ Mpc}$ (f) Scale bin =  $1.2 - 1.4 \text{ Mpc}$

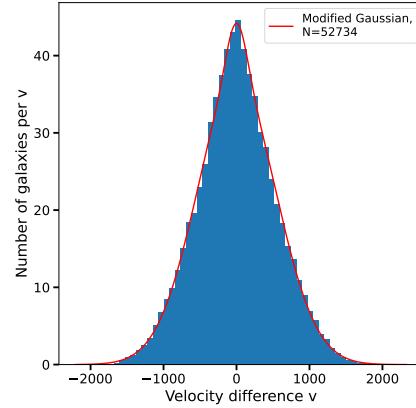


**Figure A.6:** Histograms of the one-halo peculiar velocity differences for halo pairs at fixed separation for halo group mass bin  $M=10^{13.5-14} M_{\odot}$ . The red line in each sub-figure indicates the best-fit scale-dependent weighted sum of Gaussians shown in equation 3.4 to the velocity distribution, and  $N$  specifies the total number of samples

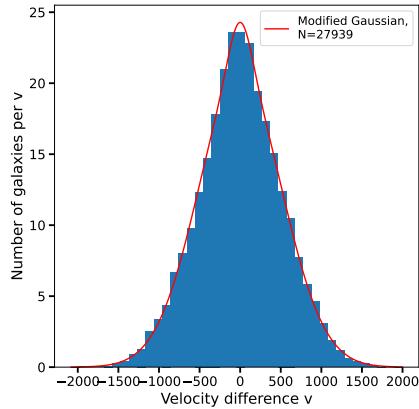
Halo group mass bin  $M=10^{14-14.5} M_{\odot}$ :



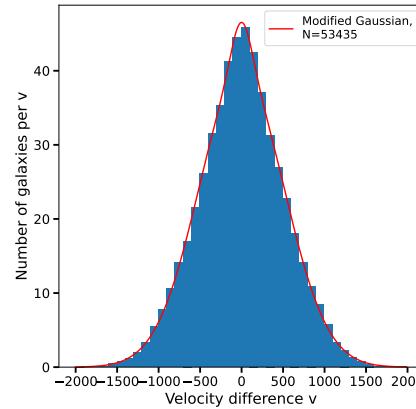
(a) Scale bin = 0 – 0.3 Mpc



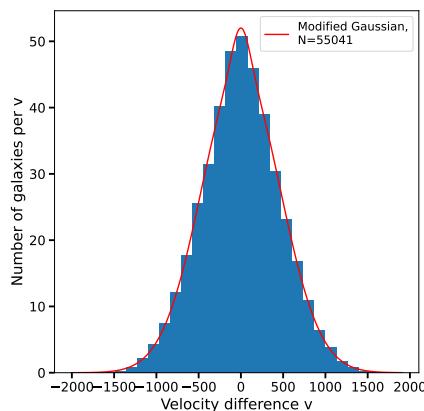
(b) Scale bin = 0.3 – 0.5 Mpc



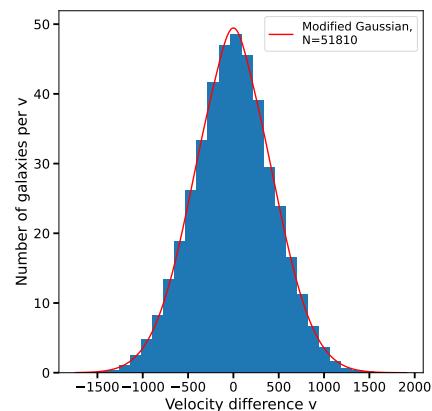
(c) Scale bin = 0.5 – 0.6 Mpc



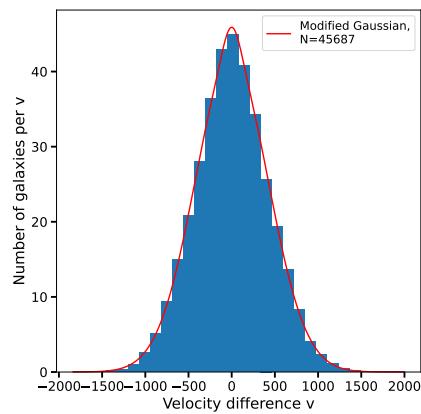
(d) Scale bin = 0.6 – 0.8 Mpc



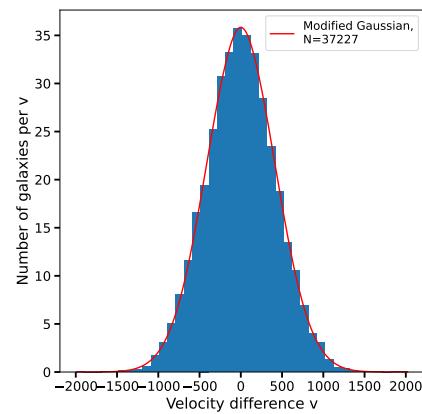
(e) Scale bin = 0.8 – 1.0 Mpc



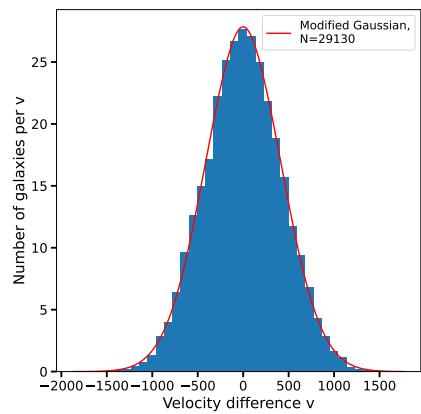
(f) Scale bin = 1.0 – 1.2 Mpc



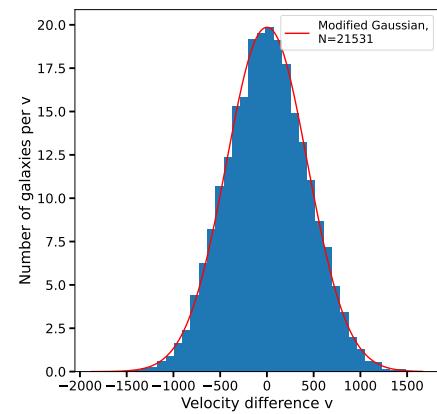
(a) Scale bin = 1.2 – 1.4 Mpc



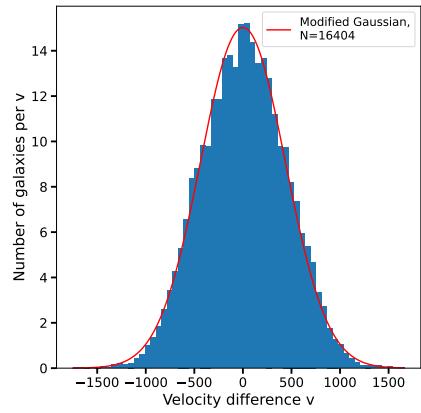
(b) Scale bin = 1.4 – 1.6 Mpc



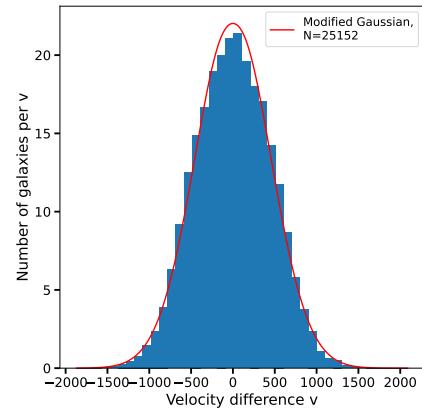
(c) Scale bin = 1.6 – 1.8 Mpc



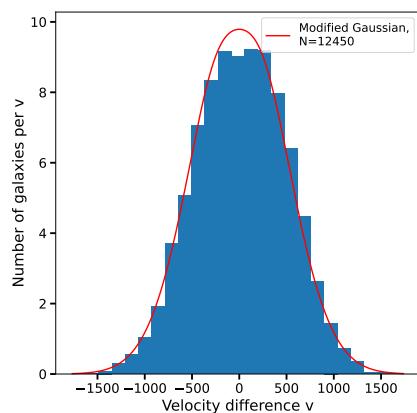
(d) Scale bin = 1.8 – 2.0 Mpc



(e) Scale bin = 2.0 – 2.2 Mpc

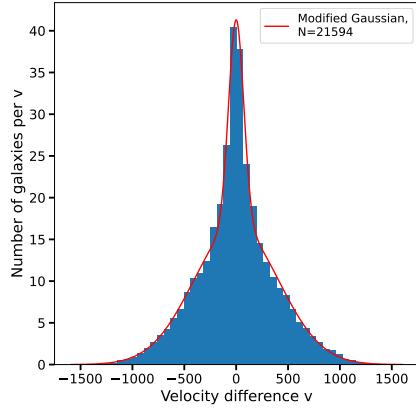


(f) Scale bin = 2.2 – 2.7 Mpc

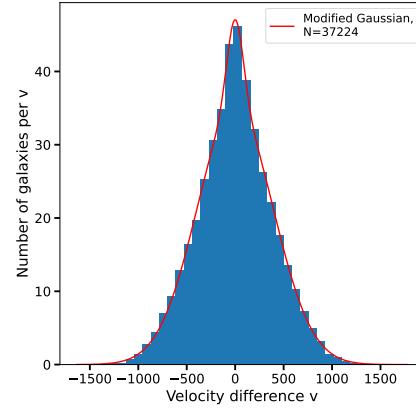


(a) Scale bin = 2.7 – 9 Mpc

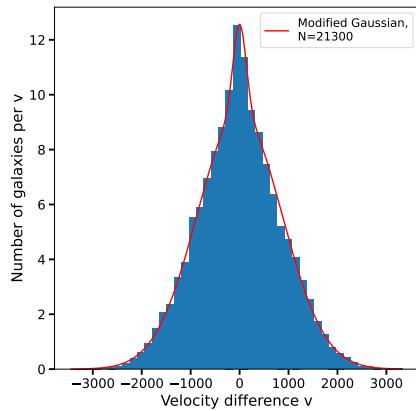
Halo group mass bin  $M=10^{14.5-15} M_{\odot}$ :



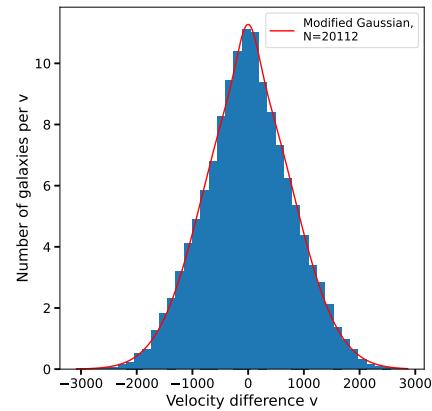
(a) Scale bin = 0 – 0.3 Mpc



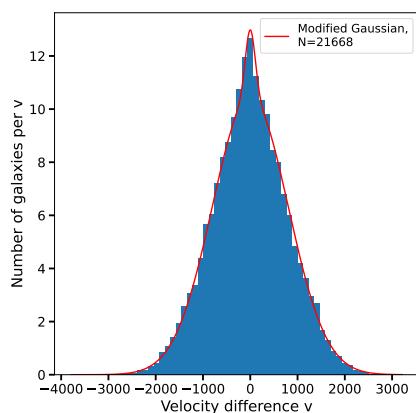
(b) Scale bin = 0.3 – 0.5 Mpc



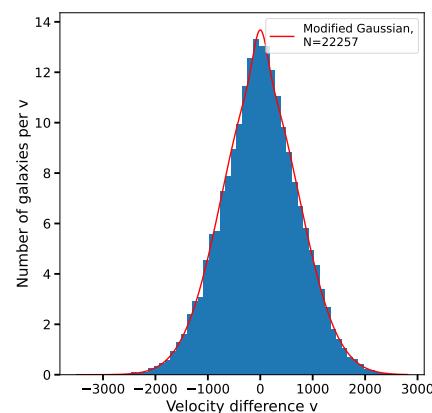
(c) Scale bin = 0.5 – 0.6 Mpc



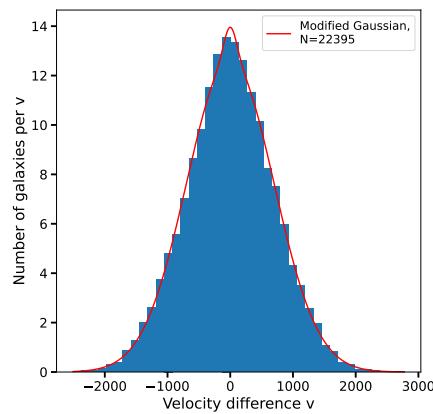
(d) Scale bin = 0.6 – 0.8 Mpc



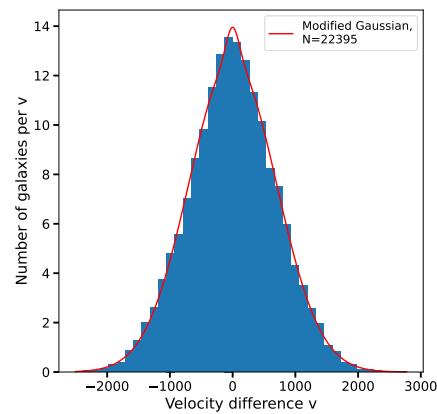
(e) Scale bin = 0.8 – 1.0 Mpc



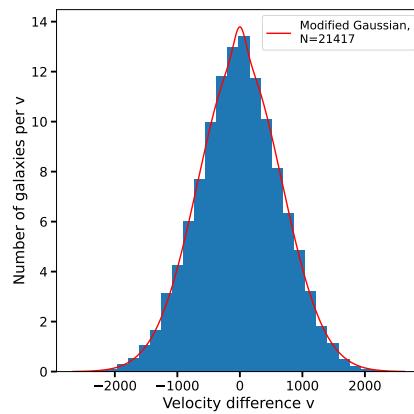
(f) Scale bin = 1.0 – 1.2 Mpc



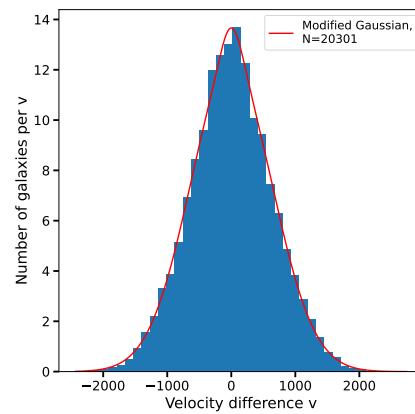
(a) Scale bin = 1.2 – 1.4 Mpc



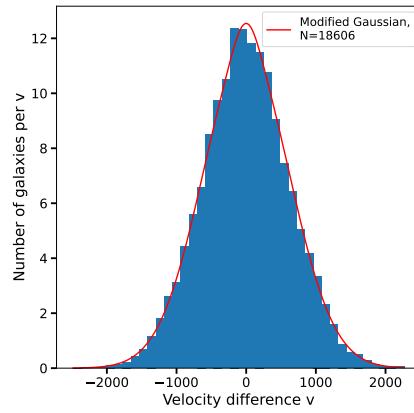
(b) Scale bin = 1.4 – 1.6 Mpc



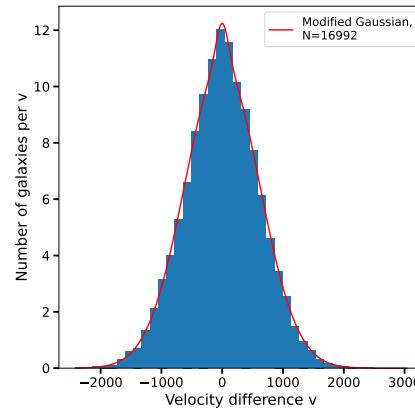
(c) Scale bin = 1.6 – 1.8 Mpc



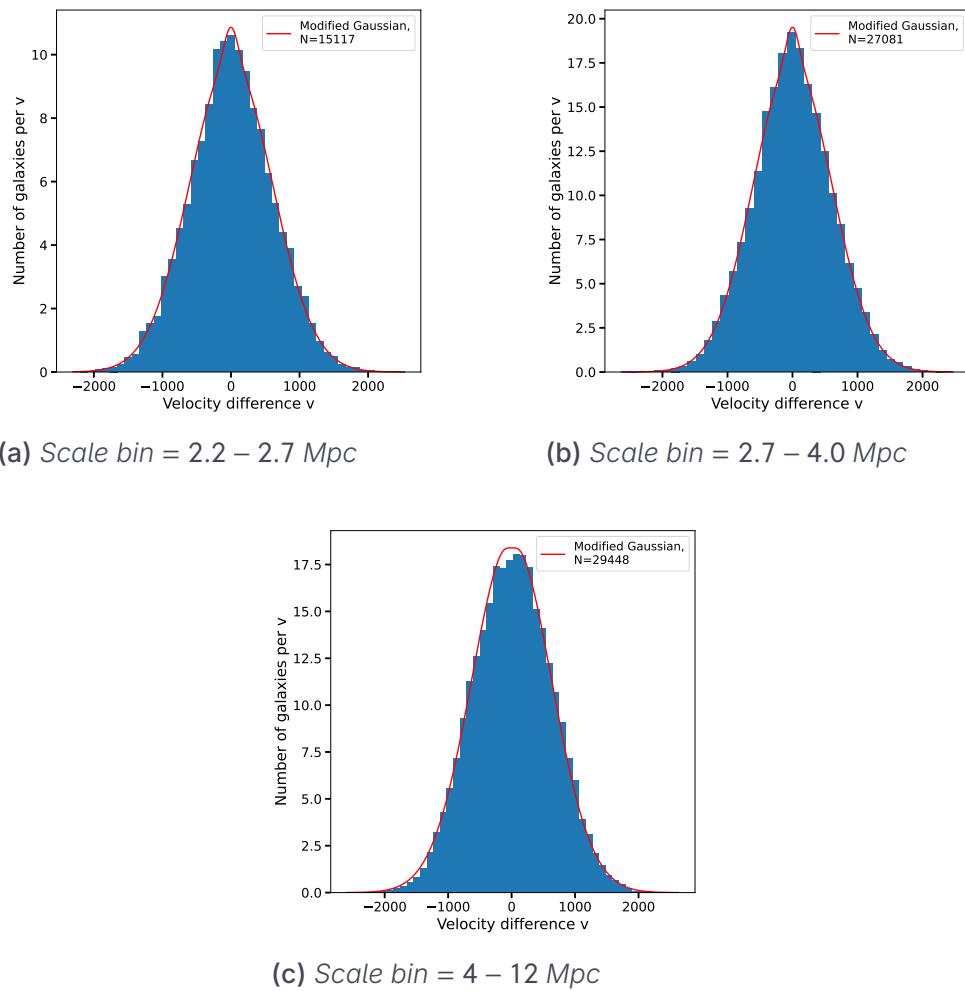
(d) Scale bin = 1.8 – 2.0 Mpc



(e) Scale bin = 2.0 – 2.2 Mpc



(f) Scale bin = 2.2 – 2.7 Mpc



# Bibliography

- Aung, H., Nagai, D., Rozo, E., Wolfe, B., & Adhikari, S. 2023, Monthly Notices of the Royal Astronomical Society, 521, 3981, doi: [10.1093/mnras/stad601](https://doi.org/10.1093/mnras/stad601)
- Baker, S., & Cousins, R. D. 1984, Nuclear Instruments and Methods in Physics Research, 221, 437
- Benson, A. J. 2010, PhR , 495, 33, doi: [10.1016/j.physrep.2010.06.001](https://doi.org/10.1016/j.physrep.2010.06.001)
- Bianchi, D., Chiesa, M., & Guzzo, L. 2015, MNRAS , 446, 75, doi: [10.1093/mnras/stu2080](https://doi.org/10.1093/mnras/stu2080)
- Bianchi, D., Percival, W. J., & Bel, J. 2016, Monthly Notices of the Royal Astronomical Society, 463, 3783–3798, doi: [10.1093/mnras/stw2243](https://doi.org/10.1093/mnras/stw2243)
- Castorina, E., & White, M. 2020, MNRAS , 499, 893, doi: [10.1093/mnras/staa2129](https://doi.org/10.1093/mnras/staa2129)
- Chandro-Gómez, , Lagos, C. d. P., Power, C., et al. 2025, Monthly Notices of the Royal Astronomical Society, 539, 776, doi: [10.1093/mnras/staf519](https://doi.org/10.1093/mnras/staf519)
- Cimatti, A., Fraternali, F., & Nipoti, C. 2020, Introduction to galaxy formation and evolution: from primordial gas to present-day galaxies
- Coil, A. L. 2013, in Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology, ed. T. D. Oswalt & W. C. Keel, Vol. 6, 387, doi: [10.1007/978-94-007-5609-0\\_8](https://doi.org/10.1007/978-94-007-5609-0_8)
- Cuesta-Lazaro, C., Li, B., Eggemeier, A., et al. 2020, MNRAS , 498, 1175, doi: [10.1093/mnras/staa2249](https://doi.org/10.1093/mnras/staa2249)
- Fisher, K. B. 1995, ApJ , 448, 494, doi: [10.1086/175980](https://doi.org/10.1086/175980)
- Han, J., Cole, S., Frenk, C. S., Benitez-Llambay, A., & Helly, J. 2017, Monthly Notices of the Royal Astronomical Society, 474, 604–617, doi: [10.1093/mnras/stx2792](https://doi.org/10.1093/mnras/stx2792)
- Hang, Q., Peacock, J. A., Alam, S., et al. 2022, MNRAS , 517, 374, doi: [10.1093/mnras/stac2569](https://doi.org/10.1093/mnras/stac2569)

- Hubble, E. P. 1926, ApJ , 64, 321, doi: [10.1086/143018](https://doi.org/10.1086/143018)
- Jackson, J. 1972, Monthly Notices of the Royal Astronomical Society, 156, 1P
- Jiang, F., & van den Bosch, F. C. 2016a, MNRAS , 458, 2848, doi: [10.1093/mnras/stw439](https://doi.org/10.1093/mnras/stw439)
- . 2016b, MNRAS , 458, 2848, doi: [10.1093/mnras/stw439](https://doi.org/10.1093/mnras/stw439)
- Kaiser, N. 1987, MNRAS , 227, 1, doi: [10.1093/mnras/227.1.1](https://doi.org/10.1093/mnras/227.1.1)
- Kugel, R., Schaye, J., Schaller, M., Forouhar Moreno, V. J., & McGibbon, R. J. 2025, MNRAS , 537, 2179, doi: [10.1093/mnras/staf111](https://doi.org/10.1093/mnras/staf111)
- Kuruvilla, J., & Porciani, C. 2018, MNRAS , 479, 2256, doi: [10.1093/mnras/sty1654](https://doi.org/10.1093/mnras/sty1654)
- Landy, S. D. 2002, ApJL , 567, L1, doi: [10.1086/339900](https://doi.org/10.1086/339900)
- Mo, H., Van den Bosch, F., & White, S. 2010, Galaxy formation and evolution (Cambridge University Press)
- Peacock, J. A., & Smith, R. E. 2000, MNRAS , 318, 1144, doi: [10.1046/j.1365-8711.2000.03779.x](https://doi.org/10.1046/j.1365-8711.2000.03779.x)
- Peacock, J. A., Cole, S., Norberg, P., et al. 2001, Nature, 410, 169–173, doi: [10.1038/35065528](https://doi.org/10.1038/35065528)
- Percival, W. J., Baugh, C. M., Bland-Hawthorn, J., et al. 2001, Monthly Notices of the Royal Astronomical Society, 327, 1297
- Reid, B. A., & White, M. 2011, MNRAS , 417, 1913, doi: [10.1111/j.1365-2966.2011.19379.x](https://doi.org/10.1111/j.1365-2966.2011.19379.x)
- Samushia, L., Reid, B. A., White, M., et al. 2014, Monthly Notices of the Royal Astronomical Society, 439, 3504
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, Monthly Notices of the Royal Astronomical Society, 526, 4978, doi: [10.1093/mnras/stad2419](https://doi.org/10.1093/mnras/stad2419)
- Scoccimarro, R. 2004, PhRvD , 70, 083007, doi: [10.1103/PhysRevD.70.083007](https://doi.org/10.1103/PhysRevD.70.083007)
- Seljak, U. 2000, MNRAS , 318, 203, doi: [10.1046/j.1365-8711.2000.03715.x](https://doi.org/10.1046/j.1365-8711.2000.03715.x)
- Tamone, A., Raichoor, A., Zhao, C., et al. 2020, Monthly Notices of the Royal Astronomical Society, 499, 5527
- Tinker, J. L. 2007, MNRAS , 374, 477, doi: [10.1111/j.1365-2966.2006.11157.x](https://doi.org/10.1111/j.1365-2966.2006.11157.x)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. 1997, ACM Transactions on mathematical software (TOMS), 23, 550