UNIVERSITY OF CALIFORNIA
SANTA CRUZ

GALAXY SPATIAL DISTRIBUTIONS: IMPROVEMENTS AND HOW THEY CAN BE USED TO INFORM THE GALAXY–HALO CONNECTION

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

James Kakos

March 2024

The Dissertation of James Kakos
is approved:

_____
Professor Joel R. Primack, Chair

_____
Professor Steve Ritz

_____
Professor Brant Robertson

_____
Professor Aldo Rodriguez-Puebla

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Galaxy Spatial Distributions: Improvements and How They Can Be Used to Inform the Galaxy–Halo Connection

by

James Kakos

In the modern ΛCDM model of cosmology, galaxies form in the centers of overdense regions in the cosmic web, known as dark matter halos. The formation and evolution of galaxies are believed to be connected to the formation and evolution of the halos they occupy. This concept is referred to as the galaxy–halo connection, and it provides us with an avenue for understanding the complex physics involved in galaxy formation. Because we assume every galaxy is located in the center of a halo, drawing parallels between the spatial distributions of galaxies and halos is an effective way of illuminating how halo properties may be connected to galaxy properties. However, three-dimensional spatial information is difficult to obtain accurately in the real Universe, as all information must be extracted from the emitted light of distant galaxies. In this paper, we apply the stochastic order redshift technique (SORT) to mock redshift surveys to test how well it recovers the true distribution of galaxies. SORT relies on a small (10%) reference sample of high-quality redshifts that outline the underlying structure of galaxies to determine new estimates of low-quality redshifts. We find that SORT overall improves redshifts, recovers the redshift-space clustering on scales $\gtrsim 2.5\ h^{-1}$Mpc, and provides improved estimates of local densities. Then, we study the clustering properties of central SDSS galaxies as a function of specific star formation rate (sSFR). We find that central

galaxy auto-correlations show little dependence on sSFR, with the established result of quiescent galaxies clustering more strongly than star-forming galaxies attributable to satellites. Because halo assembly history is known to affect distinct halo clustering, this result implies there is little net correlation between halo assembly history and central galaxy sSFR. We also find that cross-correlations of centrals with satellites increase with lower sSFR, suggesting that quiescent centrals have more satellites than star-forming centrals of the same mass. We compare our findings to the predictions of empirical models of sSFR using the Bolshoi–Planck N-body simulation and find that models dependent on halo assembly history disagree with observations while a model independent of halo assembly history reproduces well the observed clustering properties of centrals.

Acknowledgments

The text of this dissertation includes a reprint of one first-author paper I led and the text of another first-author paper I led that has not yet been published but submitted for peer review. I would like to acknowledge and thank the co-authors of these two papers. For Kakos et al. (2022), the co-authors were Joel Primack, Aldo Rodriguez-Puebla, Nicolas Tejos, L. Y. Aaron Yung, and Rachel Somerville. For Kakos et al. (2024), the co-authors were Aldo Rodriguez-Puebla, Joel Primack, Sandra Faber, David Koo, Peter Behroozi, and Vladimir Avila-Reese. All of these co-authors have kindly given their permission to include our papers in this dissertation.

I would like to especially highlight and thank Joel Primack and Aldo Rodriguez-Puebla. Joel served as my research adviser during my time as a graduate student at the University of California, Santa Cruz. His vast wealth of knowledge regarding the literature and general ideas surrounding the main concepts we researched were invaluable to my progress. Along the way, I spent countless hours trying to think and work through the technical aspects of these research projects. This effort was aided tremendously by Aldo and his understanding of the interplay of the finer details. This extended not only to the physics we worked on but to the more laborious tasks of interpreting intermediate results and debugging the many, many lines of code that were written for these analyses.

Beyond my collaborators, I would also like to thank all of my family and friends, both present and past. We are molded by the company we keep, and I consider myself fortunate to have been surrounded by so many good people. I will not attempt to make an exhaustive list of names here, but I offer my thanks and gratitude to all those who have helped shape me into who I am today. This accomplishment is a reflection of the

collective support and guidance I have received from them throughout my life.

Chapter 1.

Introduction

§ 1.1. General Overview

In this document, I present the research I did as a graduate student at the University of California, Santa Cruz. My research focused broadly on the galaxy–halo connection, which is a widely studied aspect of cosmology aiming to better understand the physics involved in galaxy formation. The idea behind the galaxy–halo connection is as follows. In modern ΛCDM cosmology, the matter in the early Universe is comprised of a mixture of $\sim\frac{5}{6}$ dark matter and $\sim\frac{1}{6}$ baryonic matter. Over time, gravity pulls this matter together into a large structure that we refer to as the cosmic web. In overdense regions of the cosmic web, matter becomes gravitationally bound and forms dark matter halos. Within sufficiently massive halos, the baryonic gas cools, collapses to the halos' centers, and begins to form galaxies. Galaxy masses continue to grow as they form stars with available gas and through mergers as halos grow hierarchically, accreting smaller, less massive halos over time. In this picture, the dark matter and baryonic matter of a halo have a shared formation and evolutionary history. As such, we expect that properties

of galaxies will be connected in some way to the properties of the halos in which they form. The way in which these properties are connected, however, is still uncertain.

Attempts to constrain the galaxy–halo connection range from simple empirical models to complex hydrodynamical simulations. With empirical models, connections are made between observations of galaxies and properties of halos as determined by dark matter cosmological simulations. A simple example of this is abundance matching. With abundance matching, one connects a galaxy property to a halo property based solely on the relative abundance of those properties within samples of galaxies and halos. A standard use case of this is abundance matching galaxy stellar masses to halo masses, from which we can derive a relation that describes the expected stellar mass of a galaxy that resides in a halo of some given mass. With hydrodynamical simulations, attempts are made to model the physics involved in galaxy formation. This can involve processes like the cooling of gas, formation of stars within a galaxy, and feedback mechanisms that can eject or recycle gas and affect the rate and duration of star formation (e.g., feedback effects of black holes which can become effective at quenching galaxies at high masses).

In this paper, we focus on another aspect of the galaxy–halo connection: spatial distributions. In our current understanding of galaxy formation, galaxies reside at the centers of dark matter halos. This means that the spatial distribution of galaxies should mirror that of the underlying structure of halos. In this way, we can connect galaxies and halos using their environments. This could be done using various methods for measuring local densities or, commonly, through two-point statistics which measure the clustering on different length scales. It has been established in various studies that halo properties (e.g., mass or concentration) can be linked to their environments and clustering behavior. Searching for similar links between galaxies and their environments

could help draw connections between which halo properties may be influencing which galaxy properties. For reviews of galaxy formation and the galaxy–halo connection, see, e.g., Mo et al. (2010), Somerville and Davé (2015b), and Wechsler and Tinker (2018b).

Each of the two first-author papers I was involved in was motivated by this idea of the galaxy–halo connection. The first paper was indirectly related, focusing solely on improvements of galaxy spatial distributions without any analysis surrounding galaxy or halo properties. The second paper was more directly related, with an emphasis on the clustering behavior of galaxies as a function of their specific star formation rate (sSFR) and how it might be influenced by halo properties. As such, while they both fall under the same umbrella of the galaxy–halo connection, they have fairly distinct introductions that set up their respective analyses. I will give a brief overview here of each of the papers but otherwise reserve their introduction for their respective chapters.

§ 1.2. Description of Research

Chapter 2 will present my first first-author paper which focuses on galaxy redshifts. Galaxy environments are thought to be one of the key aspects in determining galaxy properties (e.g., Balogh et al., 2004; Kauffmann et al., 2004; Blanton et al., 2005a), and we can only know environments as well as we can know galaxy locations. Positions on the sky are relatively easy to measure, but accurately determining the distance of a galaxy along the line of sight is much trickier. For this, we depend on redshift estimates of galaxies. Redshifts are inferred from the intensity of light received from distant galaxies at various wavelengths. As light travels through our expanding Universe, its wavelength increases in a known way that can be directly related to the distance it

travels. These redshift estimates are sensitive to the methods used to obtain them, and naturally, the easiest and cheapest methods will yield the greatest number of redshifts, whereas the more expensive but more accurate methods will be comparatively limited (e.g., Spergel et al., 2015; Ivezić et al., 2019). The goal of this paper was to develop a method that could leverage the more limited sample of high-quality redshifts that will be available from future surveys and use them to improve estimates of the lower-quality redshifts we will obtain in a much greater abundance. This ideally would lead to a clearer picture of the full three-dimensional spatial distribution of galaxies and allow for improved studies of galaxy environments.

Chapter 3 will present my second first-author paper which focuses more directly on the galaxy–halo connection. Here, we used clustering statistics to investigate the dependence of central galaxy specific star formation rate (sSFR) on halo properties. It has been well-established in the literature that red/quiescent galaxies tend to be more clustered than blue/star-forming galaxies (e.g., Li et al., 2006; Heinis et al., 2009; Zehavi et al., 2011; Coil et al., 2017; Berti et al., 2021b), but the extent to which halo properties are responsible for this is still uncertain (e.g., Blanton and Berlind, 2007; Tinker et al., 2011; O'Donnell et al., 2021, 2022). The simplest models of the galaxy–halo connection rely solely on halo mass to be the main determining factor for galaxy properties, but there are more complicated models that search for dependencies beyond halo mass (see Wechsler and Tinker, 2018b for a review). It has been shown that the clustering amplitude of dark matter halos can scale with secondary halo properties (e.g., halo concentration) at a fixed halo mass; this is referred to as assembly bias (e.g., Wechsler et al., 2006). Given the established result that quiescent galaxies cluster more strongly than star-forming galaxies at a fixed stellar mass, studies have been done attempting

to tie this relation to different halo properties (e.g., Hearin and Watson, 2013; Masaki et al., 2013; Becker, 2015). The conventional wisdom surrounding this argues that early-forming halos will host early-forming galaxies. These galaxies will, consequently, reach the ends of their life cycles (i.e., quench) earlier as they will run out of gas and cease star formation. A known assembly bias effect is that early-forming halos at a fixed halo mass are more clustered than late-forming halos (e.g., Gao et al., 2005; Wechsler et al., 2006). This creates a very natural connection to be drawn to explain why quiescent galaxies are more clustered than star-forming galaxies. The problem with this argument is that it attempts to explain a relation (quiescent vs star-forming clustering) that is based on galaxy samples consisting of both centrals and satellites. The theory surrounding assembly bias, however, has been developed focusing on distinct halos, excluding sub-halos. This means that we should be studying central galaxies, excluding the satellites. Satellite galaxies will have formed within their own sub-halos that are later accreted by larger halos when they become satellites. This means the satellites will generally have distinct formation and evolutionary histories from the halo in which they eventually reside. This makes the connection between satellites and their host halos more complex. In this paper, we separated centrals from satellites and investigated the clustering properties of only the centrals as a function of sSFR. Additionally, we tested simple models of central galaxy sSFR based on the conventional arguments of assembly bias, as well as a model that assumes no assembly bias, to see how well these models could reproduce the clustering of observed central galaxies.

Chapter 2.

Galaxy Correlation Function and Local Density from Photometric Redshifts Using the Stochastic Order Redshift Technique (SORT)

§ 2.1. Introduction

In modern cosmology, the large-scale distribution of galaxies arises from the gravitational evolution and hierarchical clustering of primordial fluctuations. Large ΛCDM N-body simulations of cold dark matter and dark energy predict how these structures evolve. Such simulations show that many properties of dark matter halos are correlated with the local density of the regions in which they form on scales of a few megaparsecs (e.g., Lee et al., 2017a). But baryonic physics is complex, and we are still seeking to understand how galaxies form and evolve and how that is connected with the properties of their host dark matter halos and the environments in which they reside (e.g., Somerville and Davé, 2015a; Wechsler and Tinker, 2018a and references therein). This can perhaps be clarified by comparing how halo properties and galaxy properties, such

as stellar radius, depend on local density and small-scale clustering (e.g., Behroozi et al., 2022).

We anticipate that this effort will be tremendously aided by the immense quantity of data that will flow from the giant LSST imaging survey at the Vera Rubin Observatory (Ivezić et al., 2019) and the higher-resolution imaging surveys by the Euclid Space Telescope* and the Roman Space Telescope (Spergel et al., 2015). These surveys will provide multi-waveband photometric redshifts for billions of galaxies, of accuracy $\sigma_z/(1+z) \approx 0.02$ or better. Euclid will also measure grism redshifts of accuracy $\sigma_z/(1+z) \approx 10^{-3}$ for ~30 million galaxies (Scaramella et al., 2021). Meanwhile, the Dark Energy Spectroscopic Instrument (DESI Collaboration et al., 2016) will measure redshifts of accuracy $\sigma_z/(1+z) \approx 10^{-4}$ for ~10 million QSOs and ~20 million galaxies, including ~17 million emission-line galaxies in the redshift interval $0.6 < z < 1.6$. In the same regions of the sky where these accurate spectroscopic redshifts are being measured, the imaging surveys will produce more than an order of magnitude more photometric redshifts. It is therefore very important to develop methods that can make efficient use of the combination of a small fraction of spectroscopic redshifts and a much larger fraction of photometric or grism redshifts in order to measure the local environments and correlations of distant galaxies. The present paper discusses one such method.

The basic idea behind these methods is that galaxies cluster, especially on scales of a few megaparsecs. The idea of estimating redshifts using clustering was first developed by Seldner and Peebles (1979); Phillipps and Shanks (1987), and Landy et al. (1996). More recently, Ménard et al. (2013) proposed a method using a small set of reference

*https://sci.esa.int/web/euclid

galaxies with spectroscopic redshifts to estimate redshifts for a larger set of galaxies that are nearby on the sky to the reference galaxies. This was tested with simulations by Schmidt et al. (2013), compared with spectroscopic redshifts by Rahman et al. (2015), used to reconstruct redshift distributions from measurement of the angular clustering of galaxies using a subset of spectroscopic redshifts by Scottez et al. (2016), and tested with simulations by Scottez et al. (2018). A related method was proposed by Morrison et al. (2017) See also Gatti et al. (2022), Cawthon et al. (2020), and Hildebrandt et al. (2021) for more recent applications of clustering redshifts.

A method to estimate redshifts of galaxies with photometric redshifts using proximity to the cosmic web defined by a subset of galaxies with spectroscopic redshifts was proposed by Aragon-Calvo et al. (2015), who applied this PhotoWeb method to the SDSS out to redshift $z \approx 0.12$. Shuntov et al. (2020) applied this method to a larger sample of galaxies with spectroscopic redshifts from the SDSS and BOSS surveys out to redshift $z = 0.4$ to reconstruct the cosmic web using the DisPerSE algorithm (Sousbie, 2011), and they used a convolutional neural network trained with the SDSS and GAMA surveys to obtain photometric redshifts with mean absolute deviation $\sigma_{\rm MAD} \approx 0.01$ out to redshift $z \approx 0.3$ for bright galaxies with $r < 17.8$. They claimed that their version of the PhotoWeb method improved the accuracy of the redshifts by about a factor of two, to $\sigma/(1+z) \approx 0.004$.

The stochastic order redshift technique (SORT; Tejos et al., 2018) is complementary to these approaches. It considers a patch on the sky where initially two kinds of galaxy redshift measurements exist: less accurate (e.g., photometric) and precise (spectroscopic). The galaxies with precise redshifts are used as a "reference sample," and it is of course expected that these correspond to a small fraction of the total number of galaxies.

New "recovered" redshifts are sampled from the distribution of precise redshifts and matched one-to-one with the uncertain redshifts such that the rank order of the uncertain redshifts is preserved. This step is motivated by the fact that random variables drawn from Gaussian probability density functions (PDFs) with equal, arbitrarily-large standard deviations satisfy stochastic ordering. By construction, SORT is non-parametric as it does not need to assume any functional form for either the distribution of redshifts or the relationship between spectroscopic and photometric redshifts. Thus, the power of SORT relies on its simplicity and versatility.

In this paper, we test how well the SORT method can use photometric redshifts plus a smaller set of reference galaxies with spectroscopic redshifts to estimate the correlations of galaxies and the local densities of their environments out to high redshifts (here we focus on a redshift interval from 0.5 to 2.5). We test the SORT method using mock galaxy surveys from backward light cones extracted from the Small MultiDark–Planck and Bolshoi–Planck cosmological ΛCDM simulations (Klypin et al., 2016a; Rodríguez-Puebla et al., 2016a). The dark matter halos were populated with central and satellite galaxies using a current version of the Santa Cruz semi-analytic model (SAM), which has been shown to reproduce well the properties of observed galaxies out to high redshifts (Somerville et al., 2021 and references therein). We show that SORT is indeed robust and that it can provide unbiased measurement of the redshift-space two-point correlation function on scales $\gtrsim 2.5\ h^{-1}$Mpc while also recovering the local galaxy and mass density, especially in regions of higher than average density where most galaxies reside.

This paper is organized as follows. In Section 2.2, we briefly describe the method, while in Section 2.3 we describe the mock galaxy surveys used to study its performance.

In Section 2.4, we present the results of applying SORT to a 2 square degree, mock wide-field light cone including galaxy redshifts, two-point correlation functions, and inferred three-dimensional densities of galaxy neighborhoods. In Section 2.5, we provide a discussion regarding preservation of the redshift rank order, the effects of the SORT parameters, limitations of the method, and potential future improvements. Section 2.6 presents a summary and main conclusions. Appendix A provides the results of applying SORT to a mock CANDELS light cone of area 0.2 square degrees. Appendix B describes our method for assigning 3D coordinates to satellite galaxies in the Santa Cruz SAM. Appendix C describes SORT performance with larger photometric redshift uncertainties. Appendix D provides several supplementary figures. All reported distances hereafter are comoving unless specified otherwise.

§ 2.2. The SORT Method

Here we present a brief overview of the SORT method. For a more complete discussion with illustrative figures, we refer the reader to Tejos et al. (2018).

2.2.1. General Idea

Consider a set of N galaxies comprised of a mixture of low-quality (referred to as photometric) and high-quality (referred to as spectroscopic) redshifts within a volume V. Assume that there are $N_{\rm ph}$ galaxies with photometric redshifts and $N_{\rm sp}$ galaxies with spectroscopic redshifts. When observing galaxies along some pencil-beam-like sub-volume, each of these subsets of galaxies will have a redshift probability distribution, $P_{\rm ph}$ or $P_{\rm sp}$, dependent upon their respective redshift uncertainties. In principle, both $P_{\rm ph}$ and

$P_{\rm sp}$ can be considered representations of the same underlying true probability distribution with different levels of noise. Due to the greater expense of obtaining spectroscopic redshifts than photometric redshifts, the statistics for $P_{\rm sp}$ are comparatively limited. However, if $N_{\rm sp}$ is large enough to be statistically relevant to the total set of N galaxies – i.e. accurately traces the cosmic structure within the volume – the higher quality of the spectroscopic redshifts will provide us with a higher resolution look at the true galaxy distribution. In this way, $P_{\rm ph}$ can be considered a noisier version of $P_{\rm sp}$.

We can leverage the relationship between $P_{\rm ph}$ and $P_{\rm sp}$ to try to improve the estimates of the photometric redshifts. To do this, we rely on stochastic ordering, which is defined as follows. Given two PDFs $P_A(x)$ and $P_B(x)$, the variable X_A is stochastically less than X_B if

$$P_A(X_A > x) \leqslant P_B(X_B > x) \quad \forall x. \tag{2.2.1}$$

To relate this to redshift estimates, consider two observed photometric redshifts $z_i^{\rm ph}$ and $z_j^{\rm ph}$ where $z_i^{\rm ph} < z_j^{\rm ph}$. We can think of each of these as being random variables sampled from Gaussian* PDFs centered on $z_i^{\rm true}$ and $z_j^{\rm true}$, respectively, with equal standard deviations determined by the measurement uncertainties.** Even with potentially overlapping PDFs, $z_i^{\rm ph}$ and $z_j^{\rm ph}$ will satisfy stochastic ordering. Therefore, we can say the most likely scenario is that the underlying true redshifts satisfy $z_i^{\rm true} \leqslant z_j^{\rm true}$. By extension, if we have $N_{\rm ph}$ redshift estimates ordered such that $z_1^{\rm ph} \leqslant z_2^{\rm ph} \leqslant \ldots \leqslant z_{N_{\rm ph}}^{\rm ph}$, we would also expect the true redshifts to most likely have the same rank ordering such that $z_1^{\rm true} \leqslant z_2^{\rm true} \leqslant \ldots \leqslant z_{N_{\rm ph}}^{\rm true}$.

*The redshift PDFs are not required to be Gaussian, but this is used for simplicity.

**Alternatively, one can think of the Gaussian PDFs as centered on $z^{\rm ph}$ where the PDF corresponds to the probability of finding $z^{\rm true}$ at a given z.

Of course, we do not have the "true" redshifts for galaxies, so we rely on high-quality spectroscopic redshifts instead. To apply the idea, we search in pencil-beam-like sub-volumes to determine $P_{\rm sp}$ in that sub-volume. We then randomly sample $N_{\rm ph}$ "recovered" redshifts, $z_i^{\rm rec}$, from $P_{\rm sp}$. Both the photometric redshifts and the recovered redshifts are rank ordered and matched one-to-one such that $z_i^{\rm rec} \leftrightarrow z_i^{\rm ph}$ for all $N_{\rm ph}$ redshifts. In doing this, we take advantage of the higher resolution provided by $P_{\rm sp}$ and simultaneously preserve the rank ordering.

We note that there may be cases where Equation (2.2.1) does not hold true, e.g., in situations where there is high variance among the individual photometric PDF widths and/or shapes. However, we can expect that for state-of-the-art photometric redshift uncertainties, the PDFs will be well-behaved and obey stochastic ordering for the majority of cases. We also note that SORT is a statistical model that should only be applied to sets of galaxies rather than individual measurements. Overall, SORT can improve redshift estimates of a set, but it can also make individual measurements worse than the original photometric estimates. Indeed, in some cases, SORT may return individual measurements that are inconsistent with a galaxy's original PDF (i.e. redshifts with errors larger than three times the photometric uncertainty).

2.2.2. The SORT Algorithm

For each galaxy i in the sample with photometric redshifts, the following steps are taken (see also Figure 1 from Tejos et al., 2018):

1. A circle with radius R is drawn on the sky around the ith galaxy.

2. Galaxies that fall within a cylinder defined by the radius R and a redshift range

$z_i \pm \Delta_z$ are selected and used for the remaining steps.

3. From the selected galaxies, a check is made to ensure there are at least $N_{\rm ref}^{\rm min}$ galaxies with spectroscopic redshifts. If there are not at least $N_{\rm ref}^{\rm min}$ spectroscopic redshifts, the circle radius is incremented by δR until the criterion is met or R exceeds some $R_{\rm max}$. If R exceeds $R_{\rm max}$, SORT is considered to have failed and does not return any redshifts. The algorithm then moves to the next galaxy.

4. A redshift histogram of the spectroscopic galaxies is made using a binning of $dz/3$. The histogram is then convolved with a Gaussian with $\sigma = dz$ to have a smooth version of the discrete $\mathrm{d}N/\mathrm{d}z$ distribution associated with the reference sample.

5. For each of the photometric galaxies within the cylinder, a new recovered redshift is sampled from the histogram of spectroscopic redshifts.

6. The selected galaxies' redshifts and the recovered redshifts are each rank ordered and matched one-to-one so each photometric galaxy is assigned a recovered redshift. This is the key "sorting" step of SORT.

As this procedure is carried out for the remaining galaxies, every time a given galaxy is within the cylinder of one of its neighbors, it will gain another recovered redshift based on that selection. In the data presented, there was a median of 43 recovered redshifts for any given galaxy. After the algorithm completes, each galaxy is assigned the median of all its recovered redshifts as its sorted redshift, $z_{\rm sort}$. The values used for the algorithm parameters are discussed in Section 2.4.

Table 2.1.: Comparison of the two mock light cones used. Each light cone was restricted to the redshift range $0.75 < z < 2.25$. Galaxies were selected from three complete redshift bins (as shown in Figure 2.1). The light cones were extracted from different simulations, though the cosmological parameters are the same for both with the exception of σ_8, as described in Section 2.3.

Light Cone	Size (deg^2)	Galaxies	Completeness	Simulation	σ_8
Wide Field	2	1,058,366	$H < 27$	SMDPL	0.829
CANDELS	0.2	47,404	$H < 25.5$	BolshoiP	0.823

§ 2.3. Mock Galaxy Surveys

2.3.1. Simulations and Backward Light Cones

We use mock galaxy surveys constructed by extracting dark matter halos along backwards light cones from the Small MultiDark–Planck (SMDPL) and Bolshoi–Planck (BolshoiP) N-body simulations (Klypin et al., 2016a; Rodríguez-Puebla et al., 2016a). The cosmological parameters of SMDPL and BolshoiP are $\Omega_\Lambda = 0.693$, $\Omega_\mathrm{m} = 1 - \Omega_\Lambda$, $\Omega_\mathrm{b} = 0.048$, $h = 0.678$, $n_s = 0.96$, and $\sigma_8 = 0.829$ for SMDPL and $\sigma_8 = 0.823$ for BolshoiP. A brief summary of the light cones are shown in Table 2.1. The dark matter halos in the simulations were identified using ROCKSTAR (Behroozi et al., 2013a). The backward light cones were constructed using the LIGHTCONE package* released by Behroozi et al. (2019a), and further details are described in Somerville et al. (2021) and Yang et al. (2021).

The merger histories of the dark matter halos were constructed using an algorithm based on the extended Press–Schechter formalism (Somerville and Kolatt, 1999; Somerville et al., 2008). The formation and evolution of galaxies within these halos was then modeled using the Santa Cruz SAM (Somerville and Primack, 1999; Somerville

*https://bitbucket.org/pbehroozi/universemachine/src/master/

et al., 2008, 2015). Somerville et al. (2021) presented a suite of light cones that was designed to represent the geometry and approximate areas of the five fields from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS)*. They compared the mock survey predictions with the CANDELS observed counts, stellar mass functions, rest-frame luminosity functions from $0.1 \lesssim z \lesssim 2$, and found generally good agreement. Yung et al. (in preparation) present a suite of 2 square degree mock light cones that have been populated with galaxies using the same approach (Yung et al., 2019a,b). In this work, we make use of one of the 2 square degree mock light cones and one of the mock CANDELS catalogs with field geometry similar to the COSMOS field, covering an area on the sky of 17×41 square arcmin in right ascension and declination.

The Santa Cruz SAM does not make use of the N-body positions and velocities for dark matter halos once they become "sub-halos" (or satellites) within a larger halo. Instead, it estimates the galactocentric radius of each satellite from the center of the halo and its decay due to dynamical friction using an analytic model (see Somerville et al., 2008 and Somerville et al., 2021). As a result, in order to compute separate redshifts for the satellites, the 3D positions and velocities must be assigned in post-processing. For details on our method for assigning these properties to the satellite galaxies, see Appendix B.

The full catalogs span the range $0 < z < 10$, though this work uses only galaxies in the range $0.75 < z < 2.25$ based on their mock observed redshifts. The lower redshift limit was imposed to ensure the light cones had large enough cross-sectional areas to measure the two-point correlation function on scales of $\gtrsim 3\ h^{-1}$Mpc. Mock observed

*http://arcoiris.ucolick.org/candels/

redshifts were calculated using

$$z_{\rm obs} = z_{\rm los} + \delta_z(1 + z_{\rm los}) \tag{2.3.1}$$

where $z_{\rm los}$ is the redshift that includes distortions from peculiar velocities along the line of sight and δ_z is a random sample from a Gaussian centered at zero with standard deviation σ_z (either photometric or spectroscopic). Note that we do not model catastrophic failures in the photometric sample as we do not expect them to have a significant effect on the net result of the SORT method.

Apparent magnitudes are provided in the mock light cones. For this work, we use H-band magnitudes given by the "wfc3f160w_dust" output of the Santa Cruz SAM. The observed-frame IR luminosities are calculated based on the star formation histories predicted by the Santa Cruz SAM and stellar population synthesis models of Bruzual and Charlot (2003). Dust attenuation is modeled using a standard "slab" model as described in Somerville et al. (2012). For more details, we refer the reader to Somerville et al. (2021).

We adopt a completeness of $H < 25.5$ for the mock CANDELS light cone, which is a rough limit to which we expect CANDELS photometric redshifts to be accurate. We expect future surveys to improve this and thus adopt $H < 27$ for the wide-field light cone. Galaxies were selected from three volume-complete regions within the light cone, as shown in Figure 2.1. In each of the three regions, 10% of galaxies were randomly chosen to have mock spectroscopic redshifts while the remaining 90% were given mock photometric redshifts. After preparing the mock catalogs, the 2 square degree wide-field light cone had 1,058,366 galaxies (~147 galaxies per square arcmin) and the mock CANDELS light cone had 47,404 galaxies (~68 galaxies per square arcmin). All results

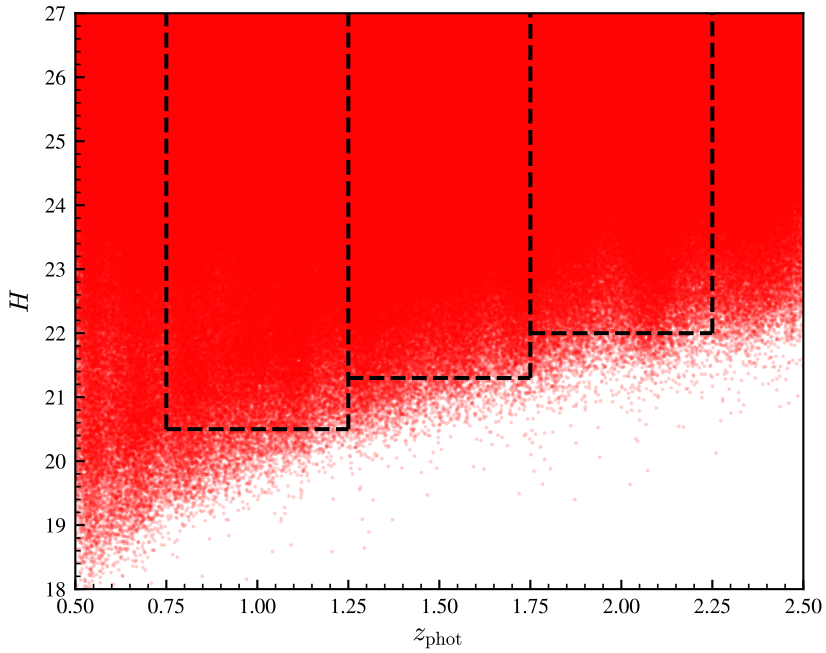

Figure 2.1. Galaxies were selected from the three volume-complete regions defined by the dashed lines. Spectroscopic redshifts were assigned within each region randomly to 10% of the galaxies in that region.

in the main text of this paper are drawn from the wide-field light cone as this provides better overall statistics. Parallel results for the mock CANDELS light cone are shown in Appendix A and are more representative of present galaxy surveys.

2.3.2. Redshift Types

Here we define several different redshift types that will be discussed in our results:

$z_{\rm cos}$: These are redshifts that are purely cosmological and include neither redshift-space distortions from line-of-sight peculiar velocities nor measurement uncertainty.

$z_{\rm spec}$: These are simulated spectroscopic redshifts that include a small measurement uncertainty according to Equation (2.3.1).

$z_{\rm ref}$: These are the reference sample redshifts. They comprise a relatively small fraction of the total number of redshifts and have spectroscopic quality.

$z_{\rm phot}$: These are simulated photometric redshifts. They are generated the same way as $z_{\rm spec}$ but with larger uncertainties.

$z_{\rm sort}$: These are the results of running the SORT method.

$z_{\rm ctrl}$: These are the results of the controlled SORT algorithm that excludes rank ordering (see Section 2.5.1 for details).

§ 2.4. Results

Most of the results in this paper were obtained assuming a spectroscopic redshift fraction of 10%, although we also explored larger and smaller spectroscopic fractions (see Figure D.2). The spectroscopic and photometric uncertainties used were $\sigma_z^{\rm sp}/(1+z) = 0.0001$ and $\sigma_z^{\rm ph}/(1+z) = 0.01$, respectively, but we also provide results for $\sigma_z^{\rm ph}/(1+z) = 0.02$ in Figure 2.7 and Appendix C. The minimum required number of reference galaxies for each sub-volume was set to $N_{\rm ref}^{\rm min} = 4$. We found SORT to be effective with this value as low as $N_{\rm ref}^{\rm min} = 2$, but increasing to 4 provided a better overall estimate of the two-point correlation function while other results remained similar. The initial search radius was set to $R = 0.01^\circ$ and the redshift bin width was set to $dz = 0.0003$. These correspond to length scales of around 0.3–0.7 h^{-1}Mpc for R and 0.3–0.6 h^{-1}Mpc for dz in the range $0.75 < z < 2.25$. These values were chosen to be able to capture relevant scales of the cosmic web. The search radius increment was set to $\delta R = 0.1R$ with a maximum possible radius of $R_{\rm max} = 0.1^\circ$. If the $N_{\rm ref}^{\rm min}$ criterion was not met within $R \leqslant R_{\rm max}$ for a

given galaxy, that galaxy was removed from the results.* The search depth was limited to $z_i \pm \Delta_z$ with $\Delta_z = 2.5\sigma_z^{\rm ph}$. This depth was chosen to be large enough to capture nearly all photometric redshifts and their true environments within the same sub-volume. See Section 2.5.2 for details on these parameters.

Our primary comparison for the results of SORT is to $z_{\rm spec}$, as spectroscopic redshifts represent our best estimates of galaxy redshifts and SORT uses these to trace the cosmic web. However, in some cases, we also show results of $z_{\rm cos}$ despite these redshifts not being directly observable due to redshift-space distortions. These results are shown for comparison as they represent the true underlying distribution of galaxies. Figures 2.3 and D.4 show $z_{\rm cos}$ to illustrate the effects of redshift-space distortions and the alignment of reference galaxies with the true cosmic web. Density estimates of $z_{\rm sort}$ are compared to those of $z_{\rm cos}$ in Figure 2.8 as Lee et al. (2017a) showed that many halo properties correlate with local densities using the true N-body positions of halos (which are replicated by using $z_{\rm cos}$, not $z_{\rm spec}$).

2.4.1. Improving Redshift Estimates

A general look at how well SORT is able to improve redshift estimates can be seen in Figure 2.2. Each panel shows a different redshift type plotted against the full declination of the light cone. The middle panel shows the reference sample, $z_{\rm ref}$, which is comprised of 10% of the spectroscopic sample and is assumed to be known when SORT is applied. This is the structural outline that SORT uses to reassign redshifts. In the $z_{\rm phot}$ panel, the cosmic structure is almost entirely smoothed out. Even with an optimistic photometric uncertainty of $\sigma_z^{\rm ph}/(1+z) = 0.01$, one can only get a very rough sense of high- or

*In the data presented, no such galaxies were removed.

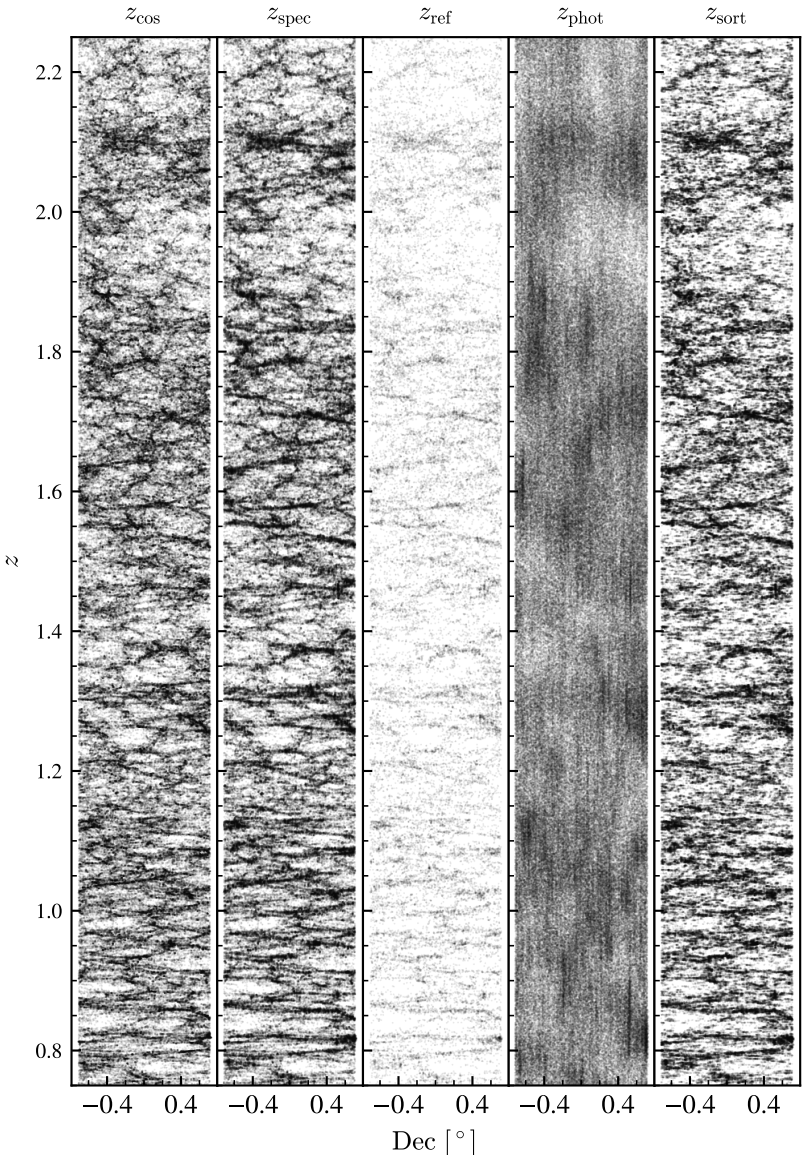

Figure 2.2. Scatter plot of the projected two-dimensional distribution of galaxies for z_{cos}, z_{spec}, z_{ref}, z_{phot}, and z_{sort}. Each panel shows a 0.5° slice in right ascension and the full declination of the light cone. The middle panel, z_{ref}, corresponds to 10% of the total galaxies, and the remaining panels show the 90% non-reference galaxies. The large-scale features of the cosmic web are much more identifiable with z_{sort} than z_{phot}. However, SORT's tendency to group galaxies closely together means that it struggles to recover low-density regions. Note that the horizontal cuts slightly visible in the z_{phot} panel are a result of the completeness condition shown in Figure 2.1.

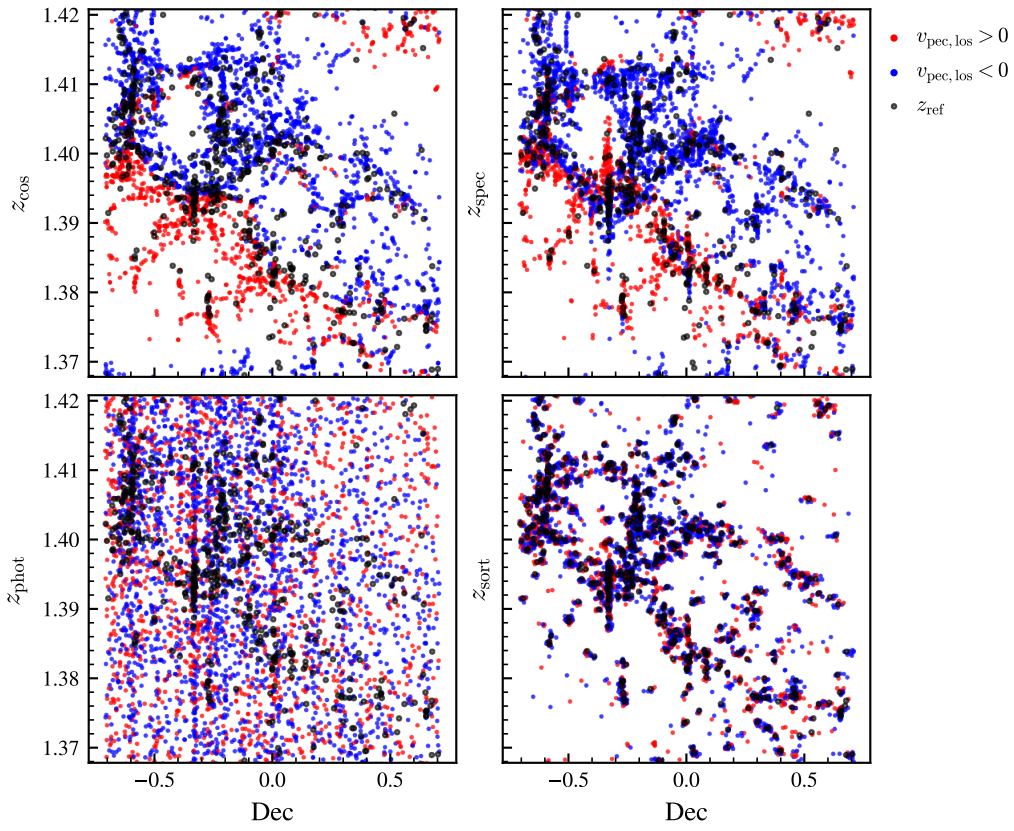

Figure 2.3. Right ascension slices (thickness 0.1°) of galaxy distributions using different redshifts in a roughly $75\times75\ h^{-1}$Mpc region of space. The red and blue coloring denotes the direction of the peculiar velocity along the line of sight (red is positive and blue is negative). The black rings with empty centers are reference galaxies. This region is dominated by a high-density ring of galaxies that surrounds a void in the upper left quadrant. We can see that the accurate tracing of this ring by the reference sample allows SORT to recreate it while also preserving the void in the center. We expect that such voids surrounded by a sufficiently high density of galaxies should largely be preserved in $z_{\rm sort}$. Reference galaxies are rarely found in voids, but may be shifted into them by redshift-space distortions in cases where dense clusters are positioned along the line of sight to the void. As such, SORT will primarily place galaxies around the voids where the reference galaxies reside. On the other hand, with photometric redshifts, the large uncertainties in the measurements of galaxies surrounding the void smooth out the region, obscuring the underlying structure as shown in the bottom left panel.

low-density regions. The $z_{\rm sort}$ panel shows a significant improvement on $z_{\rm phot}$. We see more accurate clustering of galaxies, as well as signs of filamentary structure and voids. We note that SORT's reconstruction of low-density regions is not particularly good. This is primarily due to SORT's tendency to place galaxies near other galaxies. Lower-density regions will be populated with few galaxies, and only a small fraction of those will be reference galaxies.

A more zoomed-in view of the different redshift types can be seen in Figure 2.3. Each panel shows a square region of space, roughly $75 \times 75\ h^{-1}$Mpc. The red and blue coloring represents the direction of the peculiar velocities along the line of sight; red points have positive velocities and blue points have negative velocities. When comparing $z_{\rm cos}$ to $z_{\rm spec}$, we see that galaxies in denser regions become spread out vertically. The severity of these distortions will directly impact SORT's ability to reconstruct the cosmic web. Redshift-space distortions in the reference sample will inherently affect how SORT assigns redshifts. For example, there is a dense cluster of galaxies in Figure 2.3 in front of a void. The redshift-space distortions cause a number of galaxies, including some reference galaxies, to be shifted into the void. This results in SORT placing galaxies in the void where they otherwise should not be placed.

We notice also how SORT clusters galaxies tightly to the reference sample. In the lowest density environments, there are cases where galaxies build up around one or two reference galaxies – e.g., around (0.5, 1.46) in the $z_{\rm sort}$ panel. Galaxies are pulled along the line of sight to a nearby reference galaxy, leading to horizontal structures in a plane perpendicular to the line of sight. This is most prominent in low density environments because SORT has to increase its search radius to find reference galaxies. This allows galaxies at a wider range of angular separations to be placed at roughly the same

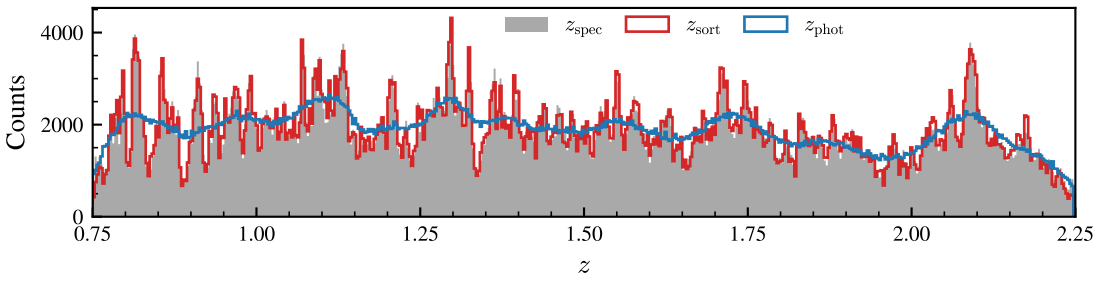

Figure 2.4. Redshift distributions for $z_{\rm spec}$ (grey), $z_{\rm sort}$ (red), and $z_{\rm phot}$ (blue) with arbitrary binning of 0.003. The large uncertainty of the photometric redshifts blurs out the structure of the distribution, which becomes more or less flat over the entire range. The distribution produced by SORT much more closely follows the distribution obtained with spectroscopic measurements. This is by design, as SORT samples new redshifts based on the distribution of the spectroscopic-quality reference sample within each sub-volume.

redshift. Overall, though, we see that SORT does a fairly good job at reconstructing the main features in this region of space, especially compared to the photometric redshifts. In the $z_{\rm phot}$ panel on the lower left, any sign of the main features in this region is almost completely lost.

This is further shown when looking at the recovery of the spectroscopic $\mathrm{d}N/\mathrm{d}z$ distribution. The one-dimensional redshift distributions are shown in Figure 2.4 for $z_{\rm spec}$ (grey region), $z_{\rm sort}$ (red), and $z_{\rm phot}$ (blue). The peaks and valleys are smoothed out in the photometric distribution while $z_{\rm sort}$ shows significant improvement in ability to outline large-scale structure along the line of sight. By construction, $z_{\rm sort}$ is meant to follow the $\mathrm{d}N/\mathrm{d}z$ distribution of the spectroscopic reference sample, and that is what we observe here. Some of the discrepancy is a result of SORT pulling galaxies from low-density regions, where reference galaxies are scarce, to high-density regions.

Figure 2.5 shows the error $\Delta z/(1+z)$ with respect to $z_{\rm spec}$. The grey shaded region shows the error distribution for $z_{\rm phot}$. Since the photometric redshifts were generated

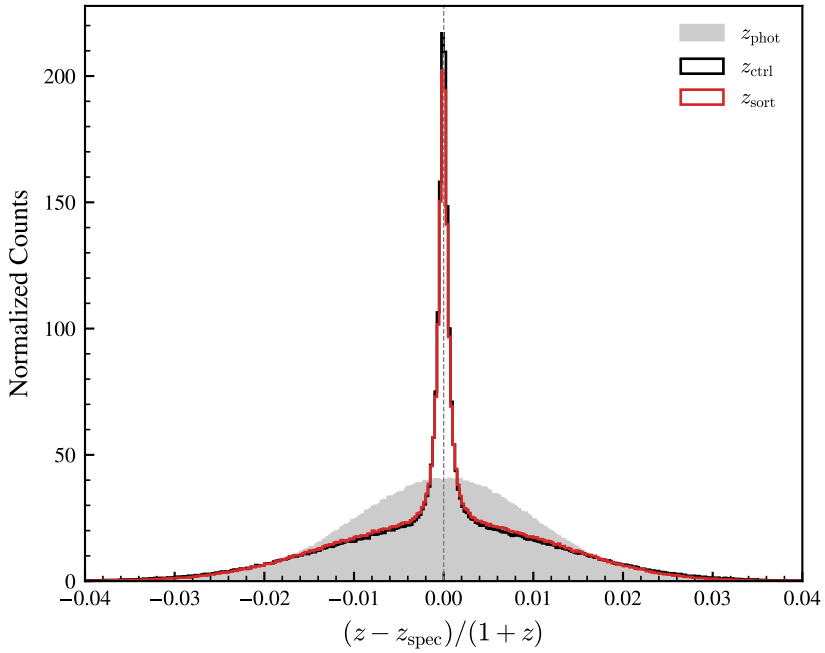

Figure 2.5. Normalized distribution of Δz (excluding the spectroscopic sample) for $z_{\rm sort}$, $z_{\rm phot}$, and $z_{\rm ctrl}$ (see Section 2.5.1 for details on $z_{\rm ctrl}$.). The photometric distribution essentially recovers the Gaussian used to create the photometric sample. SORT is able to produce a tall peak surrounding $\Delta z = 0$ where a significant fraction of redshifts have been improved. The overall standard deviation of $\Delta z_{\rm sort}$ is comparable to $\Delta z_{\rm phot}$ as shown by the broader base of the distribution.

using a Gaussian distribution, the recovered distribution is Gaussian with a standard deviation of $\sim 0.01(1 + z)$. In red, the results of SORT show a significant fraction of redshifts that have been improved. Overall, $\Delta z_{\rm sort}$ and $\Delta z_{\rm phot}$ share a similar standard deviation; however, the large peak shows that $z_{\rm sort}$ provides much more information than $z_{\rm phot}$. This is shown clearly in both Figures 2.2 and 2.4 as $z_{\rm sort}$ is able to more accurately outline large-scale structure that is washed out by $z_{\rm phot}$.

A direct comparison of redshifts can be seen in Figure 2.6. The left and right panels show the two-dimensional histograms of $z_{\rm phot}$ and $z_{\rm sort}$ compared to $z_{\rm spec}$. We continue to see improvement in redshift estimates after applying SORT. The large peak shown in

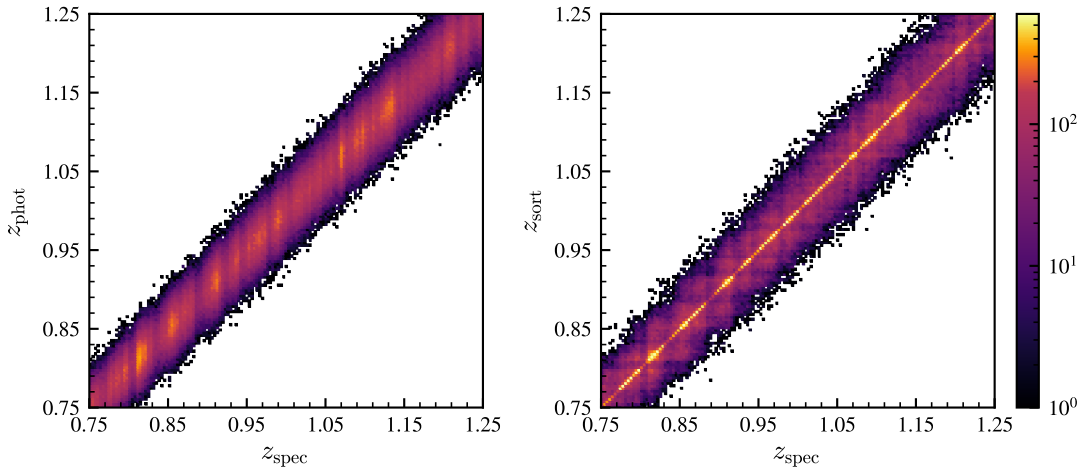

Figure 2.6. Two-dimensional redshift histograms for $z_{\rm phot}$ and $z_{\rm sort}$ relative to $z_{\rm spec}$ with binning of 0.003. The color bar represents the total number of counts in each bin. We observe significant improvement in redshift estimates by $z_{\rm sort}$ compared to $z_{\rm phot}$. There are much higher counts along the line of equality for $z_{\rm sort}$, and this effect is consistent across the entire redshift range of the light cone. All redshift bins can be seen in Figure D.5.

Figure 2.5 is now represented by a bright, narrow band of redshifts along the line of equality where errors are small. This improvement is seen in all redshift bins, which are shown in Figure D.5.

2.4.2. Recovering The Two-Point Correlation Function

The two-point correlation function (2PCF) is a relatively simple metric that provides information about the three-dimensional spatial clustering of galaxies. The large uncertainties associated with photometric redshifts lead to smoothing of spatial clustering and a highly biased estimate of the 3D 2PCF on relevant scales. As a result, using only photometric redshifts, one typically calculates the 2D angular 2PCF. Here we test SORT's ability to recover the 3D 2PCF. We note, however, that this test is somewhat

conservative because redshift distortions and SORT only affect positioning along the line of sight. Angular correlations do not deviate from their true values.

Estimates of the 2PCF were calculated using various redshift types as a function of redshift-space distance s. Both $\xi_{\rm cos}(s)$ and $\xi_{\rm spec}(s)$ assume 100% of the galaxies have a known cosmological or spectroscopic redshift.* $\xi_{\rm ref}(s)$ uses only the reference sample – i.e. only 10% of galaxies with spectroscopic redshifts. For $\xi_{\rm phot}(s)$, $\xi_{\rm sort}(s)$, and $\xi_{\rm ctrl}(s)$ (see Section 2.5.1 for details on $\xi_{\rm ctrl}(s)$), 2PCFs were calculated using their respective 90% non-reference sample redshifts plus the 10% spectroscopic-quality reference sample. The 2PCFs were calculated using CORRFUNC (Sinha and Garrison, 2020a) from scales of $\sim$1 h^{-1}Mpc to $\sim$18–30 h^{-1}Mpc (larger scales are calculated in higher redshift bins). Figure 2.7 shows 2PCF results for $0.75 < z < 1.25$ (see Figure D.3 for 2PCFs in all redshift bins). We note that because SORT is dependent upon $z_{\rm ref}$, which comprises a small fraction of the total redshifts, SORT's ability to reconstruct the three-dimensional distribution of galaxies is susceptible to sample variance in $z_{\rm ref}$. As such, SORT was run on the same light cone with 10 different random seeds (which determine the selection of $z_{\rm ref}$) to find an average result for the 2PCF. These averages are shown in Figure 2.7 along with error bars calculated as the standard deviation across the 10 random seeds. We also note that averaging over the 10 random seeds was done for all redshift types, though results will not vary much when using $z_{\rm cos}$, $z_{\rm spec}$, and $z_{\rm phot}$.

We see that using only photometric redshifts leads to a very poor estimate of the 2PCF. This is not surprising due to the large uncertainty associated with $z_{\rm phot}$. On the other hand, we observe that SORT accurately recovers the spectroscopic 2PCF on scales of $s \gtrsim 2.5\ h^{-1}$Mpc. At the smallest scales, however, SORT overestimates the 2PCF. This

*It is not expected that SORT should recover $\xi_{\rm cos}(s)$ since $z_{\rm ref}$ traces $z_{\rm spec}$, not $z_{\rm cos}$. These results are shown for the sake of comparison.

result stems from the design of the SORT algorithm – namely, galaxies will be placed near other galaxies, resulting in high clustering on smaller scales.* The lower limit to which SORT is accurate will depend on the choice of binning the method uses. As our chosen bin width corresponds roughly to 1 h^{-1}Mpc, we can only expect to be reasonably accurate at scales larger than this.

2.4.3. Estimating Local Densities

Densities were calculated by searching for neighbors within cylindrical sub-volumes surrounding each galaxy. The total length of each cylinder was set to 4 h^{-1}Mpc. Photometric redshift uncertainties correspond to scales of $\sigma_z^{\rm ph} = 0.01(1+z) \approx 30\,h^{-1}$Mpc for $0.75 < z < 2.25$. This is much larger than the scale of the cylinder and thus photometric density estimates will be particularly poor. Nevertheless, we use this cylinder length to test the limits of SORT. We also provide a sample of density estimates using a much longer cylinder defined by $l = 2\frac{\Delta v}{c}(1+z)$ with $\Delta v = 1000$ km s^{-1} in Figure D.8.

The radius of the cylinder is initially set to $r = 0.02°$. If there are not at least n galaxies within the cylinder, the radius is incremented by $\delta r = 0.001°$ until that condition is met or r reaches $r_{\rm max} = 0.04°$. This method was chosen to make the calculation adaptive. The range of densities across the entire light cone is large, and having an adaptive aperture allows for probing different scales. The radius can start small to inspect high-density regions and expand in low-density regions to estimate an average density where there may otherwise be only one or two galaxies in the cylinder. The values for r, δr, and $r_{\rm max}$ were arbitrarily chosen to be similar to the parameters used by SORT. Likewise,

*In the previous SORT paper (Tejos et al., 2018), the 2PCF was *underestimated* on small scales. This difference stems from the fact that the previous paper did not include satellite galaxies while this one does.

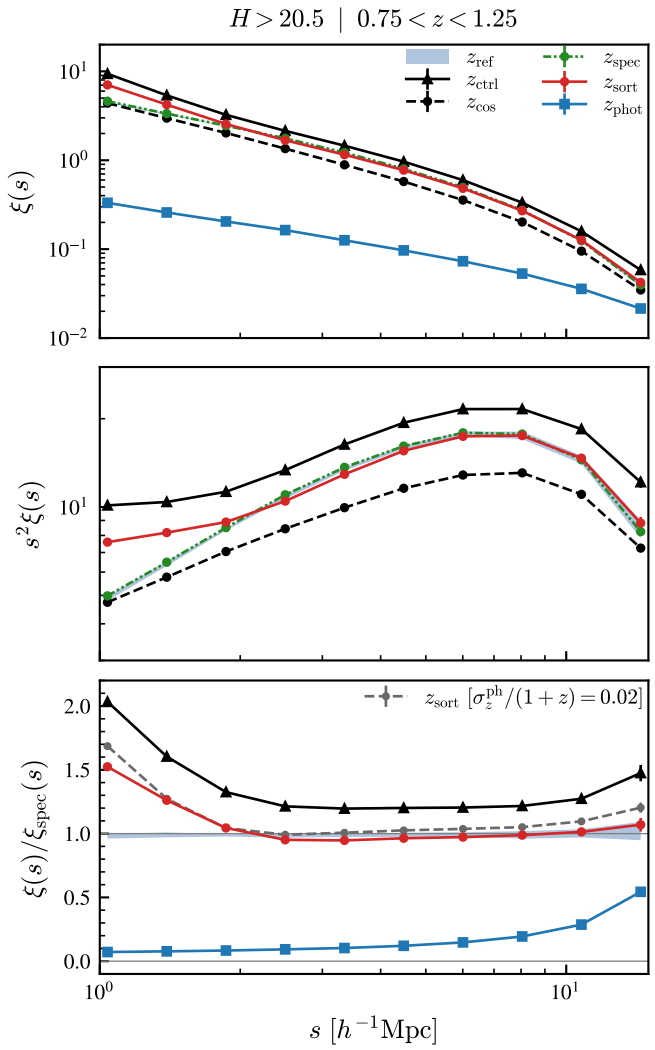

Figure 2.7. Two-point correlation functions (2PCFs) of various redshift types as a function of redshift-space distance s shown in three different ways. The results show the mean value of the 2PCFs along with 1σ error bars after running SORT with 10 different random seeds to determine the reference sample selection. Note that the error bars are too small to be seen. We observe that $\xi_{\rm phot}(s)$ is a poor estimate of the 2PCF at all shown length scales, and $z_{\rm ctrl}$ consistently overestimates the 2PCF while $\xi_{\rm sort}(s)$ is accurate (relative to $\xi_{\rm spec}(s)$) for $s \gtrsim 2.5\ h^{-1}$Mpc. See Section 2.5.1 for details on $z_{\rm ctrl}$ (shown as black triangles) and Appendix C for details on $\xi_{\rm sort}(s)$ using $\sigma_z^{\rm ph}/(1+z) = 0.02$ (shown as the grey dashed line in the bottom panel).

the minimum threshold of neighbors was arbitrarily chosen to be $n = 5$. In principle, these parameters are all adjustable depending on how much ones wishes to constrain the densities. The results of SORT's estimations of local densities compared to photometric estimates are generally not sensitive to the choice in these parameters, though.

Figure 2.8 shows three different density estimates in one redshift bin with $\rho_{\rm phot}$ in the top panels and $\rho_{\rm sort}$ in the bottom panels. The left panels show number densities, the middle panels show stellar mass densities, and the right panels show halo mass densities only considering central galaxies. The color and contours are proportional to the maximum bin value within each of the individual subplots. The dashed contour (in red) is set to a limit equal to the minimum contour level in the corresponding $\rho_{\rm phot}$ subplot.

As expected, the photometric densities are underestimated in high-density regions. The high uncertainty of the photometric redshifts has the effect of smoothing out high- and low-density regions causing them to take on a more average density. After applying SORT, low bias in the high-density regime is greatly improved and the distributions become more aligned with the line of equality. As shown by the dashed contour, SORT struggles in the low-density regime, and the scatter is comparable to the photometric estimates. SORT tends to overestimate its lowest local densities, which is a side effect of the clustering nature of the method. This is likely not something that can easily be remedied due to the simplicity of the SORT method. By design, SORT places galaxies near where it finds spectroscopic redshifts. Most spectroscopic redshifts will be in areas of higher density because this is where most galaxies are located. This tends to develop a cosmic structure that is highly clustered. SORT's ability to reconstruct low-density regions is dependent upon the quantity of high-quality redshifts found there, which will

tend to be fairly limited.

SORT's estimates for central halo mass densities are not quite as good as its number and stellar mass densities, particularly for the mock CANDELS light cone where statistics are more limited (see Figure A.4). The likely cause of this is the removal of satellite galaxies. Halo masses for satellites are not tracked once they become sub-halos and therefore were not considered in these calculations. The problem with this is that SORT is effective on average for the *full ensemble* of galaxies and does not discriminate different demographics (e.g., centrals versus satellites). We would not expect results to be as effective for a given subset of data since there is no mechanism within the method to treat different subsets differently. By removing satellites, we are decreasing the reliability of SORT, particularly in high-density regions where most satellites reside. However, we note that despite this, SORT still shows improvement over photometric density estimates.

§ 2.5. Discussion

2.5.1. The Effects of Preserving the Rank Order

One of the key aspects of SORT is the sorting itself. While it is clear that the reference sample provides a significant amount of information to SORT, one might wonder what the effect of sorting is (in step 6 of Section 2.2.2). To test this, we ran a control algorithm that excluded step 6 where the rank ordering is done. The control results were computed simultaneously with the standard SORT results and are identical in every way with the exclusion of the sorting; hence, the only difference in these two sets of results lies solely in the rank ordering. We call the results of this control algorithm $z_{\rm ctrl}$.

The first result to consider is the Δz histogram shown in Figure 2.5. Performing a

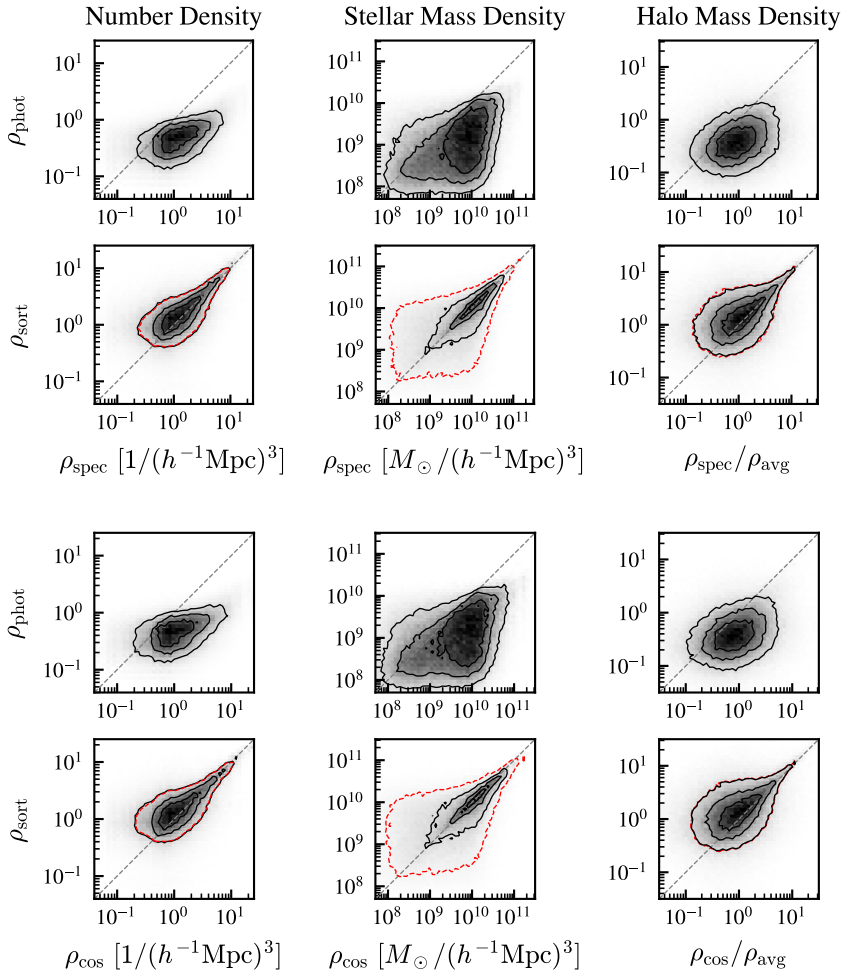

Figure 2.8. Two-dimensional density histograms for $\rho_{\rm phot}$ and $\rho_{\rm sort}$ in the range $0.75 < z < 1.25$. The left panels show number densities, the middle panels show stellar mass densities, and the right panels show halo mass densities using only central galaxies. The top six panels compare densities to estimates using $z_{\rm spec}$, and the bottom six panels compare densities to estimates using $z_{\rm cos}$. The solid contours represent limits of 25, 50, and 75 percent of the maximum bin value in each subplot. The dashed contour (red) is set at a limit equal to the minimum contour level in the corresponding $\rho_{\rm phot}$ subplot. As expected, the photometric densities estimates are all poor as the cylinder length scales are much smaller than the typical photometric redshift error. SORT densities show significant improvement in regions with average or higher density. Overall scatter is similar when comparing $\rho_{\rm sort}$ to $\rho_{\rm phot}$, but $\rho_{\rm sort}$ displays much better alignment with the line of equality and a more peaked distribution surrounding it. See Section 2.4.3 for details on the density estimates.

two-sided Kolmogorov–Smirnov test on Δz_{sort} and Δz_{ctrl} yields a p-value of $p < 0.001$. This indicates with a high level of certainty that sorting indeed changes the distribution of Δz. To understand the differences, we consider the two features of Δz: the narrow, central peak and the broader base.

When looking at the peak around $\Delta z = \pm 0.001$, we see the control sample outperforming SORT. To interpret this result, we plotted the same diagram with the data broken into a set of central galaxies and a set of satellite galaxies, shown in Figure 2.9. The left panel shows only satellites and the right panel shows only centrals. Looking at the peaks shows that the difference between Δz_{sort} and Δz_{ctrl} arises in the satellites. Satellite galaxies will tend to be situated more closely to their neighbors than a central galaxy. As such, the peak of Δz^{sat} will tend to favor environments that are more densely packed. This is precisely what the control sample provides.

To illustrate this, consider some region of space containing a dense cluster of galaxies. If we assume there are N galaxies along a pencil-beam-like sub-volume encompassing this dense cluster, we would expect each of those galaxies to have $\sim N$ recovered redshifts after SORT is complete. In other words, since the cluster is dense, we expect most of the galaxies to fall within the sub-volumes of their neighbors. The sorting aspect of SORT will always assign the lowest-redshift galaxies in this region the lowest recovered redshifts, and likewise assign the highest-redshift galaxies the highest recovered redshifts. This is simply following the condition laid out by stochastic ordering. Recalling that z_{sort} is taken to be the median of a galaxy's assigned recovered redshifts, galaxies on the lower-redshift end of the cluster are biased to have a lower z_{sort} and vice-versa at the higher-redshift end. In contrast to this, the control sample has no such bias. Each of the galaxies in the cluster will receive a random recovered redshift. After the algorithm

completes, each galaxy will have a mixture of $\sim N$ high and low recovered redshifts which will tend to have a median towards the center of the cluster. This centralization makes the cluster more dense than SORT would make it, thus favoring $\Delta z^{\rm sat}$.

Let us now consider the right panel of Figure 2.9, $\Delta z^{\rm cent}$. In this case, we have observed that there is no appreciable difference between $\Delta z_{\rm sort}$ and $\Delta z_{\rm ctrl}$ when it comes to the peak. However, if we look beyond the peak, we can see that SORT performs better than the control sample. $\Delta z_{\rm sort}$ tends to have higher counts than $\Delta z_{\rm ctrl}$ up until the point where the tails of their distributions become broader than that of $\Delta z_{\rm phot}$, around $\Delta z^{\rm cent} = \pm 0.015$. Beyond this point, $\Delta z_{\rm sort}$ has a steeper distribution, signifying its better overall recovery of redshift estimates. This relative shape is also present for the Δz histogram of the entire set of galaxies, though difficult to see in Figure 2.5.

To further investigate the effects of sorting, we can consider the 2PCF. This metric provides a better characterization of the full three-dimensional distribution of galaxies than Δz. Figure 2.7 shows a clear distinction between $z_{\rm sort}$ and $z_{\rm ctrl}$. SORT is able to accurately reproduce $\xi_{\rm spec}(s)$ on scales of $s \gtrsim 2.5\ h^{-1}$Mpc. Due to the centralization and higher density produced by the control sample, $\xi_{\rm ctrl}(s)$ ends up on average around 25% higher than $\xi_{\rm sort}(s)$. In other words, $z_{\rm ctrl}$ is overestimating the clustering while $z_{\rm sort}$ is not.

We conclude that while much of the information is provided by the reference sample, the sorting aspect of SORT certainly provides useful information as well. This information is most evident when considering the 2PCF where the lack of sorting leads to over-clustering by around 25%. The only drawback to sorting comes with the Δz histogram of satellite galaxies. This is a difficult issue to resolve because SORT uses one prescription to treat two distinct demographics, and information about which galaxies are centrals

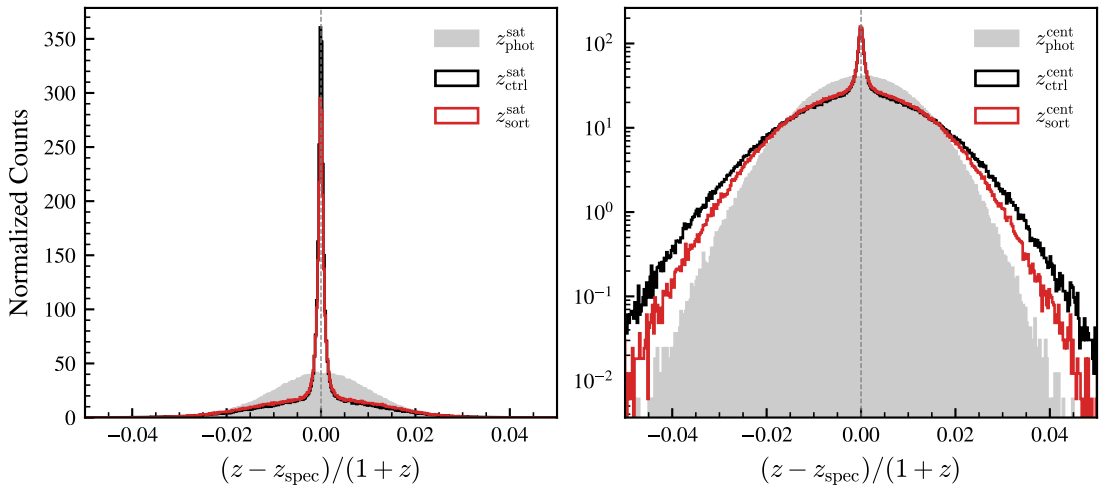

Figure 2.9. Normalized distribution of Δz for $z_{\rm sort}$, $z_{\rm ctrl}$, and $z_{\rm phot}$. The left panel shows the results using only satellite galaxies and the right panel shows the results using only central galaxies. See Section 2.5.1 for details. The left panel shows that the more highly-clustered results of $z_{\rm ctrl}$ favor satellite galaxies. The right panel shows similar results between $z_{\rm sort}$ and $z_{\rm ctrl}$ at smaller errors; however, SORT displays better treatment at larger errors ($\Delta z/(1+z) \gtrsim 0.02$) with more rapidly declining tails.

or satellites is not readily available for real observations.

2.5.2. Sub-volume Parameters

The SORT parameters $N^{\rm min}_{\rm ref}$, R, and Δ_z determine the sizes of the cylindrical sub-volumes that surround each galaxy during the SORT procedure. A balance must be struck for these parameters in order for SORT to produce reasonable results.

Having a larger cylinder radius allows for more of the environment to be taken into consideration when looking for reference galaxies. This can be useful in cases where galaxies are near the outer edge of a large cluster of galaxies. If the radius is too small, the inner region of the cluster may not be detected by the pencil-beam-like sub-volume. This leaves the galaxy more susceptible to being pulled toward denser regions that may

be close on the sky but not in redshift. On the other hand, making the radius too large can also be problematic. With a large radius, galaxies that are not particularly nearby on the sky, but still within the sub-volume, will be pulled toward the redshifts of denser regions. Because SORT only moves galaxies along the line of sight, these galaxies will be placed around the same redshift as another group of galaxies but with a seemingly "incorrect" position on the sky. The result is a distribution of galaxies that becomes elongated in a plane perpendicular to the line of sight.

This effect can be seen in Figures 2.3 and D.4 and is a signature of the SORT method. In low-density regions (e.g., the upper right corner of the $z_{\rm sort}$ panel), we see horizontal formations of galaxies. The magnitude of this effect can be limited by adjusting $R_{\rm max}$ or $N_{\rm ref}^{\rm min}$. As $N_{\rm ref}^{\rm min}$ becomes smaller, the radius of the average sub-volume will also be smaller, leading to narrower horizontal formations. This may come at the expense of SORT's overall performance, though. We found that increasing $N_{\rm ref}^{\rm min}$ from two to four provided a better estimate of the 2PCF, for example. If, instead, $R_{\rm max}$ is adjusted, one must take care to not make it too small relative to $N_{\rm ref}^{\rm min}$. If $R_{\rm max}$ is too small, the fail rate of the SORT algorithm will increase as the maximum sub-volume size is too constrained to find enough reference galaxies. Likewise, the fail rate will also increase if $N_{\rm ref}^{\rm min}$ is too large for a given sub-volume size.

The redshift cut imposed by Δ_z is a new addition to SORT. In the original SORT paper (Tejos et al., 2018) which looked at nearby galaxies in a wider, shallower field, an apparent magnitude cutoff was imposed such that only galaxies within $\pm\delta m$ of the ith galaxy were considered in the sub-volume. When SORT is applied to a deeper field, the magnitude cut is not sufficient to allow SORT to perform well. The same range of magnitudes can be found at opposite ends of the light cone, which leaves too large of a

range of possible recovered redshifts for a given galaxy.

To limit the range of redshifts that are considered neighbors of the ith galaxy, a redshift cut based on a galaxy's photometric redshift was implemented to replace the magnitude cut. The length chosen for the cylinder should be based on the photometric uncertainty. In this case, we have assumed the photometric uncertainties are Gaussian. As such, we have chosen $\Delta_z = 2.5\sigma_z^{\rm ph}$ to allow the majority of photometric galaxies the potential of recovering their true redshift. The value of this parameter was not thoroughly tested, however, and may not be optimal. We emphasize that this parameter, as well as other SORT parameters, should be tested to find optimal values when applied to different surveys. The values used in this paper correspond to sensible length scales, but optimal values will likely vary depending on the metric one wishes to optimize.

2.5.3. Limitations

One main limitation of the SORT method is its dependence on a reference sample. The limitations of this dependence are twofold. First, there is a limitation to the length scale that SORT will be able to properly recover. Dispersion velocities of $v \approx 200$ km s^{-1} correspond to lengths of ~4–6 h^{-1}Mpc for $z = 1$–2. This is a rough limit of SORT's accuracy relative to the *true* distribution of galaxies (i.e. not the spectroscopic distribution to which results were compared in this paper) due to redshift-space distortions. Second, SORT requires a structural outline by the reference sample to recover an accurate distribution of galaxies. If the reference sample does not outline a feature of the cosmic web, then SORT will not be able to reconstruct it. This effect is most significant in low-density regions. The fraction of galaxies found in low-density regions will naturally be low. An even lower fraction of those galaxies will be reference

galaxies. Without reference galaxies, SORT will not be able to reproduce an accurate distribution of galaxies in these regions.

A second limitation of the SORT method is the fact that it can only improve redshift estimates collectively for ensembles of galaxies. Figure 2.5 shows a tall peak surrounding $\Delta z = 0$ where a significant fraction of redshifts have been improved, but there is no way to tell which galaxies are in this peak or which galaxies are in the tails. Despite SORT doing a fairly good job of recovering the large-scale structure of galaxies, there will undoubtedly be some galaxies placed in the wrong environments. Higher accuracy redshift estimates are still required to properly place galaxies on an individual basis. However, as shown in the previous section, SORT can still be used to infer local densities (particularly, average or higher densities) reasonably well.

2.5.4. Future Considerations

One possible next step for SORT is updating the assignment of spectroscopic redshifts to create the reference sample. In this paper, spectroscopic redshifts were assigned randomly to 10% of galaxies within three complete redshift bins. To make tests of the SORT method more realistic, one could model the reference galaxy selection using the methods chosen by large imaging surveys (e.g., selecting a mixture of brighter galaxies and galaxies with high star formations rates that produce strong emission lines).

Another possible step is to improve redshift assignments within sub-volumes as the SORT method is carried out. There are currently no considerations given to the angular correlations within each sub-volume. As shown in the left panel of Figure 2.9, the current treatment of satellite galaxies by SORT is not optimal.* This could potentially

*That is not to say that $z_{\rm ctrl}$ *is* optimal, but it is enough to demonstrate that SORT is not.

be improved by assigning similar recovered redshifts to galaxies that appear highly clustered on the sky. This could also be implemented when determining the final $z_{\rm sort}$ redshift of a galaxy. The final selection from a galaxy's pool of recovered redshifts at the end of the method could be biased to redshifts where the galaxy appears more clustered on the sky, as opposed to taking a simple median. Such considerations could also prove useful in reducing the horizontal structures produced by SORT discussed in Section 2.5.2.

Yet another improvement would be to treat satellite galaxies more realistically than we have done, as described in Appendix B. For example, an improved semi-analytic treatment of satellite galaxies could be based on the recent `SatGen` papers (Jiang et al., 2021; Green et al., 2021b,a).

Recently, it has been shown that correlations between galaxy and halo properties create observable signatures in local environments (Behroozi et al., 2022). In particular, halo spin, concentration, growth rate, and interaction history have all been shown to leave scale-dependent signatures in both 2PCFs and the distributions of distances to galaxies' kth nearest neighbors out to $z \sim 2.5$. These determinations were based on projected two-dimensional environments so as to make them observationally accessible with low-resolution spectroscopy ($\sigma_z/(1+z) \lesssim 0.005$). We have shown that SORT is able to recover the full three-dimensional 2PCF as estimated with high-resolution spectroscopy. We also provide in Appendix D results using SORT to estimate three-dimensional distances to kth nearest neighbors (see Figure D.9). We expect SORT's ability to reasonably-well recover local environments may allow for further observationally-accessible detections of environmental signatures that result from galaxy–halo property correlations.

In this paper, we have not taken advantage of the fact that galaxy properties could

depend on environment and/or location within the cosmic web. Indeed, it is well known that, for example, star forming / blue galaxies are less clustered than more quiescent/red galaxies (see, e.g., Li et al., 2006; Zehavi et al., 2011; Coil et al., 2017; Berti et al., 2019, 2021a) and that more spheroid-like morphologies are more frequently in denser environments (e.g., Dressler, 1980; Pearson et al., 2021). Thus, a natural next step within the framework of SORT would be to divide the reference sample by galaxy properties. By doing so, it is expected that SORT would be able to determine even more accurate redshifts than when not considering galaxy properties.

Finally, we expect to test the performance of SORT using real data sets from highly complete spectroscopic galaxy surveys, e.g., GAMA (Baldry et al., 2018) and DESI (DESI Collaboration et al., 2016), in order to account for systematics that are present in real surveys but not properly modeled by our mock experiment. For instance, we expect that the so-called "catastrophic redshift" failures in photometric redshift methods will have a minor effect in the performance of SORT as these are typically only a small fraction of the total sample. Other systematic differences include having non-Gaussian PDFs for the photometric redshifts and having a set of photometric galaxies with variable $\sigma_z^{\rm ph}$ in the sample. For example, COSMOS2020 (Weaver et al., 2022) obtained photometric redshift precision of ∼4% for the faintest galaxies and better than 1% for the brightest galaxies.

§ 2.6. Summary and Conclusions

In this paper, we have tested the performance of the SORT method (Tejos et al., 2018) in mock high-redshift surveys. SORT is a simple, efficient, and robust method that

can be used to improve redshift estimates. It relies upon a reference sample of high-quality spectroscopic redshifts for which a precise distribution $\mathrm{d}N/\mathrm{d}z$ is known within pencil-beam-like sub-volumes of the survey. Within each sub-volume we:

1. sample new "recovered" redshifts from the $\mathrm{d}N/\mathrm{d}z$ distribution of high-quality redshifts

2. match the recovered redshifts one-to-one with the low-quality (photometric) redshifts such that the rank order is preserved.

The second step is motivated by the fact that random variables drawn from Gaussian PDFs with equal, arbitrarily-large standard deviations satisfy stochastic ordering. In other words, if two redshift estimates z_i and z_j satisfy $z_i < z_j$, then their *true* redshift values most likely satisfy $z_i^{\mathrm{true}} \leqslant z_j^{\mathrm{true}}$. Thus, preserving the rank order makes the assigned recovered redshifts more likely to be close to their underlying true value. This process is repeated for sub-volumes surrounding each galaxy in the survey. The result is every galaxy with a low-quality redshift is assigned multiple recovered redshifts from which a new redshift estimate can be determined.

We ran the SORT method on a wide-field 2 square degree mock light cone and a mock CANDELS light cone extracted from the Small MultiDark–Planck and Bolshoi–Planck N-body simulations, respectively, to test its performance in a pencil-beam-like survey spanning the redshift range $0.75 < z < 2.25$. After applying SORT, we observe similar improvement from both mock catalogs and make the following determinations:

1. We observed overall improvement in redshift estimates, allowing for better reconstruction of the three-dimensional distribution of galaxies than photometric

redshifts alone provide. This can be seen broadly in Figure 2.2 or more close up in Figure 2.3.

2. We also observed that SORT produces much better agreement (by design) in the one-dimensional $\mathrm{d}N/\mathrm{d}z$ distribution, allowing it to better identify large-scale structure along the line of sight as shown in Figure 2.4.

3. Distributions of redshift errors with respect to spectroscopic estimates were overall improved throughout the light cone, while the standard deviations remained about the same. One- and two-dimensional histograms of this are shown in Figures 2.5 and 2.6, respectively.

4. SORT accurately recovers the spectroscopic redshift-space 2PCF down to scales of $\gtrsim 2.5\ h^{-1}$Mpc while photometric redshifts (with errors of $\sigma_z^{\mathrm{ph}}/(1+z) = 0.01$ corresponding to scales of ~20–30 h^{-1}Mpc) drastically underestimate galaxy clustering. This is shown clearly in the top and bottom panels of Figure 2.7.

5. SORT is able to recover three-dimensional local densities in regions of average or higher density at scales of $\gtrsim 4\ h^{-1}$Mpc. Three different density histograms are shown in Figure 2.8, and additional histograms are shown in Figure D.8 at a larger length scale of $l = 2\frac{1000\ \mathrm{km\ s^{-1}}}{c}(1+z)$.

We expect that such improved determinations of local galaxy environments will help to distinguish the effects of environmental properties (e.g., local density) on galaxy evolution from other effects, such as galaxy stellar or halo mass (e.g., Peng et al., 2010; Woo et al., 2013; Chartab et al., 2020; Behroozi et al., 2022).

Chapter 3.

Star-forming and Quiescent Central Galaxies Cluster Similarly: Implications for the Galaxy–Halo Connection

§ 3.1. Introduction

According to the ΛCDM paradigm, galaxy formation and evolution take place within massive dark matter halos. A key component of our understanding of galaxy formation and evolution within this context is the stellar-to-halo mass relation (SHMR). The SHMR relates the mass of stars within a galaxy to the mass of its host halo. This serves as a vital tool in connecting the observed properties of galaxies with the underlying dark matter halos – known as the galaxy–halo connection – allowing us to explore the processes governing galaxy formation, the impact of feedback mechanisms, galaxy bias, and the cosmological context in which galaxies form and evolve (for a review, see Somerville and Davé, 2015b).

The SHMR is derived by measuring the stellar masses of a large sample of galaxies

and associating them with their corresponding halo masses, which are inferred through various methods such as gravitational lensing (Mandelbaum et al., 2006, 2016), galaxy clustering (Berlind and Weinberg, 2002a), galaxy kinematics (More et al., 2011; Wojtak and Mamon, 2013; Lange et al., 2019), empirical modeling (Behroozi et al., 2019b), abundance matching techniques (Conroy et al., 2006), or any combination of these methods (for a review, see Wechsler and Tinker, 2018b). These probes of the galaxy–halo connection provide a statistical description of the SHMR, which is assumed to be an increasing one-to-one monotonic relationship. In addition, the SHMR exhibits scatter, indicating that there is a range of stellar masses for a given halo mass. This scatter may arise from a combination of factors such as the stochasticity of the star formation process, varying merger histories, and the influence of environmental effects on galaxy formation. Previous attempts to constrain the scatter around the SHMR have found it to be of the order ∼0.15 dex (see, e.g., Rodríguez-Puebla et al., 2015; Behroozi et al., 2019b; Porras-Valverde et al., 2023).

Whatever is influencing a galaxy's position in the SHMR should be related to the assembly history of the galaxy and, ultimately, its star formation activity (or color), i.e., linked to its position within the specific star formation rate (sSFR)–stellar mass plane. Observationally, several previous studies have identified a robust segregation in halo mass at fixed stellar mass, where quiescent/red central galaxies inhabit more massive halos compared to star-forming/blue central galaxies (e.g., More et al., 2011; Tinker et al., 2013; Rodríguez-Puebla et al., 2015; Mandelbaum et al., 2016; Lange et al., 2019). At least part of the explanation for this segregation is the fact that while a galaxy's star formation may cease, its host halo can continue to grow hierarchically, especially in the case of more massive halos. Consequently, for a given stellar mass, galaxies that

ceased star formation earlier tend to have larger halo masses, even if they had high star formation efficiency during their active star-forming phase. Notably, this segregation in the SHMR becomes more pronounced in massive galaxies, with quiescent central galaxies residing in halos that are a factor of ∼2 more massive than halos of star-forming central galaxies (More et al., 2011; Rodríguez-Puebla et al., 2015; Mandelbaum et al., 2016).

Another way to describe the SHMR segregation is that at fixed halo mass, $M_{\rm h}$, star-forming galaxies have a higher stellar mass, M_*, than quiescent galaxies. Previous studies by Moster et al. (2018, 2020) have found an opposite relation, that quiescent galaxies have a *higher* stellar mass than star-forming galaxies at a fixed halo mass. This would seem to imply that quiescent galaxies reside in halos of *lower* mass than star-forming galaxies at a fixed stellar mass. However, these authors point out that due to the scatter in the SHMR, $\langle M_*(M_{\rm h})\rangle$ cannot simply be inverted to obtain $\langle M_{\rm h}(M_*)\rangle$ (see also Rodríguez-Puebla et al., 2013, 2015). The combined effect of Eddington bias and a higher fraction of star-forming galaxies at lower masses results in the average halo mass being higher for quenched galaxies than star-forming galaxies at a fixed stellar mass, which is consistent with observational constraints. This phenomenon has been described in the literature as the inversion problem (Cui et al., 2021). In addition to an opposite relation, we note that some papers have found little evidence of such a segregation in the SHMR (see, e.g., Figure 38 of Behroozi et al., 2019b).

Do specific halo properties determine a galaxy's position in the SHMR, or is it instead influenced by other stochastic factors? Is the strong segregation observed in the SHMR a genuine phenomenon? If so, does it align with the notion of halo assembly bias – the concept that the clustering behavior of haloes depends not only on halo mass but also

on formation history – as a main driver of the assembly histories of galaxies? Two-point correlation functions are well-explored tools for investigating these questions. By assuming that the centers of halos and sub-halos serve as the locations of central and satellite galaxies, we can focus on modeling how galaxies inhabit halos of varying masses to understand galaxy clustering. In particular, if the segregation is indeed present, it could be interpreted that the SHMR is primarily a reflection of halo mass as the determining factor for stellar mass and sSFR. This would suggest that, at a given stellar mass, the enhanced clustering of quiescent galaxies could be attributed to their occupancy of more massive halos (Rodríguez-Puebla et al., 2015). On the other hand, if differences in clustering are influenced by other halo properties, two possibilities emerge:

1. The scatter around the SHMR may be merely random variation unrelated to galaxy assembly history, with halo assembly history being the key factor determining how galaxies cluster (e.g., Hearin and Watson, 2013).

2. There could exist a segregation in the SHMR, and clustering information is shared between this segregation and halo assembly bias.

These considerations present scenarios that can be tested using various approaches for measuring two-point correlation functions, as discussed in the main body of this paper.

It is observationally established that red/quiescent galaxies tend to be more clustered than blue/star-forming galaxies (see, e.g., Li et al., 2006; Heinis et al., 2009; Zehavi et al., 2011; Coil et al., 2017; Berti et al., 2021b). Recent works by Coil et al. (2017) and Berti et al. (2021b) studied the auto-correlation functions (ACFs) of PRIMUS, DEEP2, and SDSS galaxies as a function of their sSFR and stellar mass. They found that in bins

of stellar mass, the ACF decreases monotonically as sSFR increases, including within the star-forming and the quiescent populations, which Berti et al. (2021b) refer to as intrasequence relative bias. They interpret these differences in galaxy clustering as a function of sSFR as evidence that the scatter in galaxy sSFR is physically connected to the large-scale cosmic density field.

The degree to which halo properties are responsible for the differing clustering of star-forming and quiescent galaxies is still uncertain (Blanton and Berlind, 2007; Tinker et al., 2011; O'Donnell et al., 2021, 2022). The simplest models of the galaxy–halo connection depend solely on halo mass to determine the properties of galaxies (for a discussion, see Wechsler and Tinker, 2018b). However, there has been an effort to expand models of the galaxy–halo connection to include halo properties beyond mass (e.g., Hearin and Watson, 2013; Masaki et al., 2013; Becker, 2015). Some such studies aiming to identify halo properties that influence the regulation of SFRs in galaxies have highlighted two potentially key factors: halo accretion rate and halo concentration.

Halo accretion rate controls the influx of gas into a galaxy's interstellar medium (see, e.g., Avila-Reese and Firmani, 2000; Bouché et al., 2010; Dekel and Mandelker, 2014; Wetzel and Nagai, 2015; Rodríguez-Puebla et al., 2016). One might expect halo accretion rate to be linked to SFR, as having more gas available may allow a galaxy to form more stars. It has also been shown that, at a fixed halo mass, halos with lower accretion rates tend to be in higher density environments at low redshifts (Lee et al., 2017b), so it may also be expected that quiescent galaxies are more clustered than star-forming galaxies when modeling sSFR based on halo accretion rate. This, indeed, has been shown by Becker (2015) where the relative clustering amplitudes of quiescent and star-forming galaxies agree with observations. Additionally, the dispersion in the halo

accretion rate has been shown to reproduce the width of the SFR distribution in star-forming galaxies at various redshifts (Dekel and Mandelker, 2014; Rodríguez-Puebla et al., 2016).

Halo concentration, on the other hand, is linked to the timing of gas infall into the halo, with more concentrated halos experiencing earlier infall (see, e.g., Avila-Reese et al., 1998; Wechsler et al., 2002; Gao et al., 2004; Dutton et al., 2010; Matthee and Schaye, 2019). A prevailing explanation for the difference in clustering for quiescent and star-forming galaxies is that galaxies that form earlier reach the ends of their life cycles (i.e., become quiescent) earlier, and earlier-formed halos tend to be more clustered. At a fixed halo mass, halos with high concentration tend to be the ones that formed earlier and are more clustered (Wechsler et al., 2006; Gao et al., 2008; Montero-Dorta et al., 2021), so it is natural to connect halo concentration to galaxy SFR. Indeed, models based on halo concentration have successfully reproduced the observed clustering of red and blue galaxies (Hearin and Watson, 2013; Masaki et al., 2013).

In both cases of halo accretion rate and halo concentration, halo assembly bias is assumed to be important for understanding the distribution and behavior of galaxies within dark matter halos (Sheth and Tormen, 2004; Gao et al., 2005; Wechsler et al., 2006; Montero-Dorta et al., 2021). It is important to note that existing descriptions of assembly bias and density fields have been primarily developed for distinct halos (i.e., excluding sub-halos). This would correspond to a sample of only central galaxies, excluding satellites. Despite sharing the same host halo, centrals and satellites are expected to have distinct formation and evolutionary histories, which makes the connection between satellites and the halos in which they reside more complex. We highlight the fact that past studies of correlation function trends with sSFR or galaxy

color have generally not made distinctions between centrals and satellites. Notably, the findings of Coil et al. (2017) and Berti et al. (2021b) are obtained using samples of central and satellite galaxies together. However, Berti et al. (2021b) showed that a modified version of the UniverseMachine model (Behroozi et al., 2019b) agrees with their ACF results and that in this model, central galaxies contribute substantially to the dependence of clustering on sSFR at a given stellar mass. In this paper, we will test the sSFR-dependence on galaxy clustering, focusing on central galaxies only. This will allow for a more direct comparison with the theoretical framework of the galaxy–halo connection. As we will show, the presence of satellite galaxies introduces additional variability in the correlations. Furthermore, we find that ACFs alone may not be sufficient to discriminate models of the galaxy–halo connection when considering only central galaxies, while cross-correlation functions (CCFs) between central and satellite galaxies prove to be a powerful tool in this regard.

This paper is organized as follows. In Section 3.2, we describe the main data set we use, including the cuts that were made to define our samples and the group catalogs used to identify central and satellite galaxies. In Section 3.3, we describe our methods for calculating correlation functions and binning the data into sub-samples as a function of stellar mass and sSFR. Section 3.4 presents our results of ACFs and CCFs within these various sub-samples. Using these results, we test different models of the galaxy–halo connection in Section 3.5. In Section 3.6, we summarize our results and provide a discussion of their robustness and comparisons with the literature. Finally, in Section 3.7, we present our conclusions.

§ 3.2. Data

3.2.1. Sloan Digital Sky Survey

In this study, we use observations from the Sloan Digital Sky Survey (SDSS) with galaxy redshifts taken from SDSS Data Release 7 (Abazajian et al., 2009). Stellar masses and star formation rates (SFRs) are taken from the MPA-JHU catalog (Kauffmann et al., 2003; Brinchmann et al., 2004) and have a Kroupa (2001) initial mass function. Brinchmann et al. (2004) used observed emission lines, including Hα, within the central fiber of the SDSS and modeled them based on the Charlot and Longhetti (2001) stellar population synthesis model. They showed that using the standard Kennicutt (1998) conversion factor from Hα to SFR is a good average correction for most of the star-forming galaxies. The same is not true for quiescent galaxies, however, for which their SFRs are mostly given by the D4000 break.

3.2.2. Group Catalogs

One of the main goals of this paper is to quantify, separately, the contributions of central and satellite galaxies to the observed two-point auto-correlation function (ACF) as well as their cross-correlation function (CCF). Therefore, an important designation we use for our sample is whether a galaxy is a central or satellite. In this paper, we utilize the halo-based group catalog by Yang et al. (2012, hereafter Y12) as our primary database to identify central and satellite galaxies within the SDSS. It is important to note that we exclusively employ their spectroscopic group sample, comprising a total of 593,227 galaxies at $z \leq 0.2$. Since galaxy group finders operate in redshift space, they are inherently susceptible to errors from redshift-space distortions, making

it difficult to achieve a perfect galaxy-to-group assignment and introducing systematic uncertainties in the identifications. To ensure the robustness of our results, we also incorporate two additional group catalogs: Tempel et al. (2017, hereafter T17) and Rodriguez and Merchán (2020, hereafter R&M20), containing 571,291 and 648,480 galaxies, respectively, at $z \leqslant 0.2$. It is essential to clarify that our primary aim in using these alternative catalogs is not to critique or evaluate which group finder performs best in identifying groups but rather to examine whether all three catalogs yield consistent results. In each of these catalogs, we define central galaxies as the most massive galaxy within a group in terms of stellar mass, with any remaining galaxies in the group taken to be satellites. Note that by this definition, isolated galaxies that have no satellites will be considered centrals.

3.2.3. Data selection

To calculate two-point correlation functions, we use a random catalog of the SDSS provided by Blanton et al. (2005b). The random catalog has 1 million points distributed with constant surface density over the area of SDSS. The SDSS projected sky distribution has irregular edges and some holes where no objects are catalogued. In order to ensure consistency between the observations and the random catalog when calculating two-point correlation functions, we trim the edges of both distributions following Varela et al. (2012) and Cebrián and Trujillo (2014). The following four cuts are applied:

1. Southern limit: $\delta > 0°$
2. Western limit: $\delta > -2.555556(\alpha - 131°)$
3. Eastern limit: $\delta > 1.70909(\alpha - 235°)$

4. Northern limit: $\delta < \arcsin\left[\frac{0.93232\sin(\alpha - 95.^{\circ}9)}{\sqrt{1-[0.93232\cos(\alpha - 95.^{\circ}9)]^2}}\right]$.

These cuts can be seen in Figure 2 of Varela et al. (2012) and Figure 1 of Cebrián and Trujillo (2014). Additionally, we mask out the holes in both the observations and the random catalog with small rectangular cuts over the affected regions.

After applying these cuts to the data, we then narrow our selection down to the five mass bins. Each bin will correspond to a volume-limited sample that is complete in stellar mass over the range of the bin. The mass bins cover the range $10.0 < \log(M_*/M_\odot) < 11.5$, where the two lower mass bins have widths of 0.375 dex and the three higher mass bins have widths of 0.25 dex (as listed in Table 3.1). We choose the lower mass bins to be wider to improve statistics since the lower masses require smaller volumes to maintain completeness.

To define the sub-volumes, we follow the "Volume1" procedure in Appendix A of Y12.* The redshift limits of the sub-volumes are defined based on a minimum threshold in the galaxy number density, $n_{\rm vol}(z)$, for a given sub-volume. In this way, sub-volumes will have relatively flat $n_{\rm vol}(z)$ distributions, i.e., number densities that are near-constant. To determine the redshift limits, in each mass bin, we locate the redshift corresponding to the peak of $n_{\rm vol}(z)$ and then travel to lower and higher redshifts until the distribution falls to 50% of the maximum. For this process, we calculate $n_{\rm vol}(z)$ in redshift bins of width $dz = 0.005$ and apply a Gaussian smoothing with $\sigma = dz$ to the overall distribution. This smoothing helps to avoid spikes within any single redshift bin of $n_{\rm vol}(z)$ biasing the results of the process (e.g., regions of low density due to cosmic

*Additionally, Y12 defines a "Volume2" procedure following van den Bosch et al. (2008), which accounts for the differing selection effects of red and blue galaxies in flux-limited surveys, owing to their different mass-to-light ratios. We find our resulting volumes to be generally consistent with this alternative completeness limit.

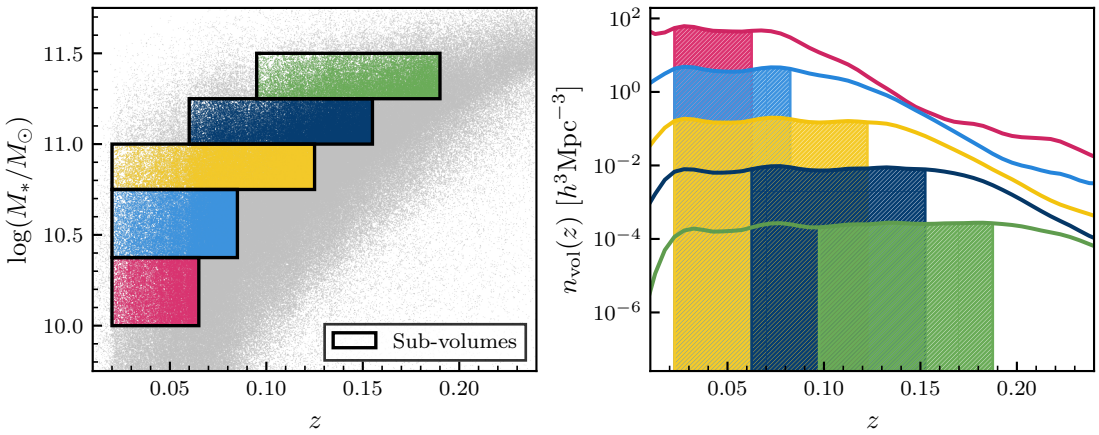

Figure 3.1. *Left panel*: stellar mass as a function of redshift for the SDSS. The rectangles with colored points show the five volume-limited samples used for measuring correlation functions. *Right panel*: number densities as a function of redshift for the five sub-volumes. The solid lines show the histograms over the entire redshift range, while the colored regions show the selected redshift ranges spanned by the sub-volumes. The redshift limits of the sub-volumes are determined by traveling in each direction from the maximum of the distribution to points where the counts drop to 50% of the maximum. This process was applied to various sub-samples within each sub-volume to ensure completeness in all of them (see Section 3.2.3 and Figure E.1 for details). Note that each successive distribution from front to back is scaled up by an additional order of magnitude for clarity.

variance, rather than a lack of observations, can cross the 50% threshold resulting in an artificially smaller sub-volume). Using a threshold of 50% was chosen as a balance between completeness and statistics. Choosing a higher threshold would yield sub-volumes with a higher degree of completeness; however, the two-point statistics would become less reliable as the sub-volumes would be smaller with fewer galaxies. Within each sub-volume, this process is repeated for the individual sub-samples of galaxies that we work with: all, central, satellite, star-forming, green valley, and quiescent galaxies. We choose the redshift limits that allow us to satisfy our completeness condition in all sub-samples simultaneously within a mass bin.

Table 3.1.: Comparison of the five sub-volumes used. Stellar mass bins cover the range $10.0 < \log(M_*/\mathrm{M}_\odot) < 11.5$, with the lower two mass bins having widths of 0.375 dex and the upper three mass bins having widths of 0.25 dex. The redshift ranges correspond to the ranges where the number densities of galaxies are at least 50% of the maximum in a given mass bin (see Section 3.2.3 and Figure E.1 for details).

M_* Limits $[\log(M_*/M_\odot)]$	z Limits	N_{gal}
10.0 – 10.375	0.02 – 0.065	23,969
10.375 – 10.75	0.02 – 0.085	48,221
10.75 – 11.0	0.02 – 0.125	57,533
11.0 – 11.25	0.06 – 0.155	51,738
11.25 – 11.5	0.095 – 0.19	27,269

The five sub-volumes are shown in the left panel of Figure 3.1 along with the number densities in the right panel. The colored lines show the number densities across the full redshift range in each mass bin, and the shaded regions show the limits that define the sub-volumes based on our completeness condition. Note that the number densities are scaled up by increasing orders of magnitude from front to back for clarity. A more detailed look at the number densities for the individual sub-samples can be seen in Figure E.1. The result of our selection process is a catalog of 208,730 galaxies across the five sub-volumes. Specifics for each sub-volume are described in Table 3.1.

As a check of the completeness in our resulting sub-volumes, we calculate the galaxy stellar mass function (GSMF) across our full mass range. Figure 3.2 shows the GSMF in each of our five sub-volumes (colored points) compared to a double Schechter function using the Dragomir et al. (2018) best-fit parameters of SDSS (black line). Overall, we find generally good agreement between our sub-volumes and the fit, though there is some discrepancy at low mass. This is likely the result of the 50% limit being too low and creating a volume that is too large for fainter low-mass galaxies. While this leads to

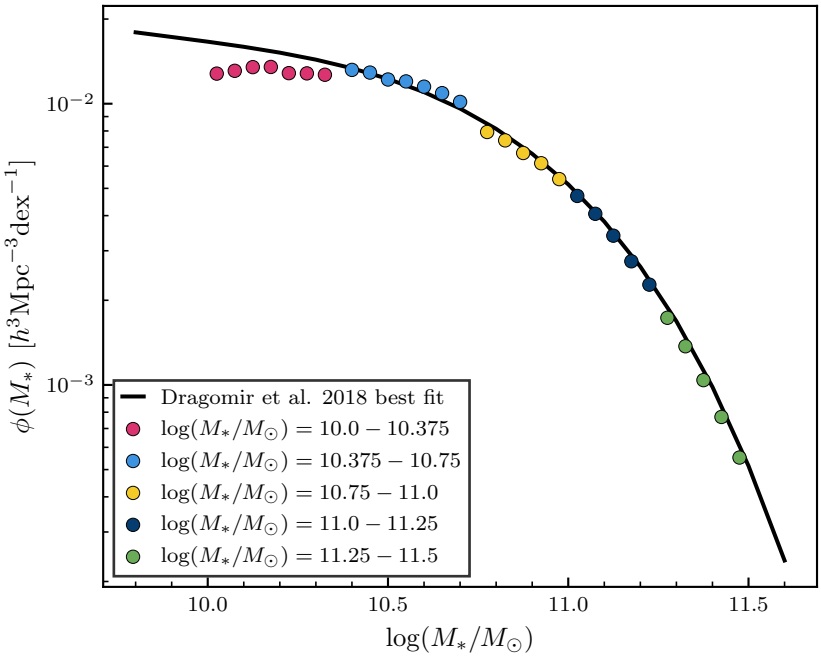

Figure 3.2. Galaxy stellar mass function (GSMF) of our five sub-volumes compared to a double Schechter function with the best-fitting parameters from Dragomir et al. (2018). Our sub-volumes match well with the SDSS GSMF, with a slight discrepancy at the lowest masses. This is the result of a trade-off between completeness and sufficient statistics to calculate correlation functions.

a lower level of completeness than our other sub-volumes, the larger volume is required to obtain reliable two-point statistics.

§ 3.3. Methods

3.3.1. Clustering Measures

Galaxy clustering was measured using two-point correlation functions, which measure the excess probability over random of finding pairs of galaxies separated by a distance r. To mitigate the effects of redshift-space distortions, the correlation function is calculated

in a projected area broken into components perpendicular to (r_p) and parallel to (r_π) the line of sight. Assuming two galaxies are positioned at $\mathbf{s_1}$ and $\mathbf{s_2}$, a line-of-sight vector can be defined as $\mathbf{l} = \frac{1}{2}(\mathbf{s_1} + \mathbf{s_2})$ and a separation vector can be defined as $\mathbf{s} = \mathbf{s_1} - \mathbf{s_2}$. With these definitions, r_p and r_π can be calculated as

$$r_\mathrm{p} = \sqrt{\mathbf{s}\cdot\mathbf{s} - r_\pi^2} \tag{3.3.1}$$

$$r_\pi = \frac{\mathbf{s}\cdot\mathbf{l}}{|\mathbf{l}|}. \tag{3.3.2}$$

Correlation functions are calculated using the Landy and Szalay (1993) estimator

$$\xi(r_\mathrm{p}, r_\pi) = \frac{\mathrm{D_1D_2} - \mathrm{D_1R_2} - \mathrm{D_2R_1} + \mathrm{R_1R_2}}{\mathrm{R_1R_2}} \tag{3.3.3}$$

where $\mathrm{D_1D_2}$ is the data–data pair counts, $\mathrm{D_1R_2}$ and $\mathrm{D_2R_1}$ are the data–random pair counts, and $\mathrm{R_1R_2}$ is the random-random pair counts. For CCFs, the subscripts denote the two different data sets that are being cross-correlated. For ACFs, the subscripts are all the same and Equation (3.3.3) simplifies to

$$\xi(r_\mathrm{p}, r_\pi) = \frac{\mathrm{DD} - 2\mathrm{DR} + \mathrm{RR}}{\mathrm{RR}}. \tag{3.3.4}$$

These are then integrated over r_π to find the two-dimensional projected correlation functions

$$w_\mathrm{p}(r_\mathrm{p}) = 2\int_0^{r_\pi^\mathrm{max}} \xi(r_\mathrm{p}, r_\pi)\mathrm{d}r_\pi. \tag{3.3.5}$$

For our calculations, we integrate to $r_\pi^\mathrm{max} = 20\,h^{-1}$Mpc. The calculations were done using CORRFUNC (Sinha and Garrison, 2020b) in 10 r_p bins from $0.1\,h^{-1}$Mpc to $20\,h^{-1}$Mpc. Error bars for correlation functions show the 1σ error after running 200 bootstrap samples. Since we will show results of both projected ACFs and CCFs, we will use w_a to

denote projected ACFs and w_c to denote projected CCFs.

Correlation functions can be affected on small scales by fiber collisions. In the SDSS, galaxies within 55 arcsec of each other cannot receive fibers on the same plate, but some regions (roughly a third of the sky) were tiled with overlapping plates. We tested for the impact of fiber collisions by upweighting pair counts for galaxies that had pairs on small scales by a factor of 3. The largest impact was, as expected, on scales below $r_p \approx 0.1$ Mpc for galaxies around $\log(M_*/M_\odot) = 10$ and scales below $r_p \approx 0.2$ Mpc for galaxies around $\log(M_*/M_\odot) = 11$, with changes at the few-percent level on larger scales. Hence, we do not expect fiber collisions to affect conclusions for any of the correlation function analyses in this paper.

3.3.2. Fitting the Star-forming Main Sequence

In this section, we describe how we define the star-forming main sequence (SFMS). We follow Stephenson et al. (in preparation, see also Rodríguez-Puebla et al., 2020a and Fang et al., 2018) and briefly describe the method below. This definition will be the basis for how we break down the sSFR–M_* plane into a grid of sub-samples (see Section 3.3.3). The idea is to create bins in sSFR based on distance (in dex) from the mean of the SFMS at a fixed M_*. To do this, we use an iterative process of fitting the sSFR–M_* relation of star-forming galaxies until the fit becomes stable. The general process is as follows:

1. Fit a straight line to the data sample of $\log(\mathrm{sSFR/yr^{-1}})$ vs $\log(M_*/M_\odot)$.

2. Shift the line down 0.45 dex in $\log(\mathrm{sSFR/yr^{-1}})$ and select all galaxies above the line to be the new sample.

3. Repeat until the fit parameters are stable to a maximum tolerance of 10^{-3}.

Once stable parameters are obtained for the straight line fit, we calculate the median sSFR of all galaxies above the line (i.e., star-forming galaxies) as a function of M_* in 10 mass bins. These medians are used to fit a function of the form

$$\langle \log[\mathrm{sSFR}_{\mathrm{MS}}(M_*)] \rangle = \log\left(\frac{\psi_0}{M_*}\right) - \log\left[1 + \left(\frac{M_*}{M_0}\right)^{\gamma}\right] \tag{3.3.6}$$

where ψ_0, M_0, and γ are fitting parameters (Lee et al., 2015). For our data set, we find the best-fit parameters $\log(\psi_0/M_\odot \mathrm{yr}^{-1}) = 0.829$, $\log(M_0/M_\odot) = 10.914$, and $\gamma = -0.897$. Equation (3.3.6) defines the mean of the SFMS, which we use to create all other sSFR bins. We take all galaxies 0.45 dex below this curve or higher to be the SFMS.

3.3.3. The sSFR–M_* Grid

In order to systematically study ACFs and CCFs in the sSFR–M_* plane, we break our galaxy sample into a grid binning by stellar mass and distance from the mean of the SFMS. The five mass bins defined in Section 3.2.3 are used to mitigate the effects of mass dependence on the correlation functions and isolate the dependence on sSFR. Figure 3.3 shows the mean and median masses within each cell of the grid. In every grid cell, the mean and median mass are similar in value to each other, and in any given mass bin, they each remain relatively constant as a function of sSFR (the largest variation within any mass bin is $\lesssim 0.05$ dex). This ensures that, within a given mass bin, any disparities in the two-point correlation functions across the sSFR ranges will be *solely attributable* to their distinct clustering properties rather than being influenced by internal mass effects within the bin.

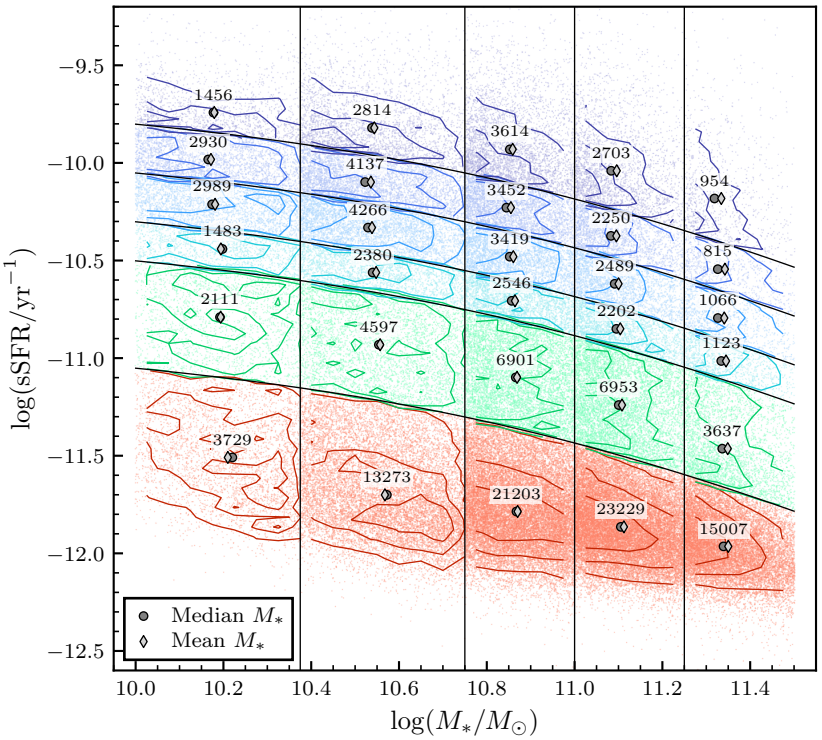

Figure 3.3. Scatter plot of the sSFR–M_* plane for Y12 central galaxies showing the grid binning scheme. Correlation functions are calculated within individual cells or combinations of cells in the grid. The data are broken horizontally into the five sub-volumes shown in Figure 3.1 and vertically by distance from the SFMS (specifics shown in Table 3.2). The text boxes show the number of galaxies within each grid cell. Gray circles and diamonds show the median and mean stellar masses and median sSFR within each cell. Overall, we find the median and mean stellar masses to be similar in value and close to constant as a function of sSFR, which helps remove mass effects in the trends of the correlation functions.

Table 3.2.: List of sub-samples in the sSFR–M_* plane and the corresponding ΔMS bin definitions.

Sub-sample	ΔMS bin [dex]
Highly star-forming (HSF)	ΔMS > 0.25
Upper main sequence (UMS)	0 < ΔMS < 0.25
Lower main sequence (LMS)	-0.25 < ΔMS < 0
Bottom of the main sequence (BMS)	-0.45 < ΔMS < -0.25
Green valley (GV)	-1 < ΔMS < -0.45
Quiescent (Q)	ΔMS < -1

Using Equation (3.3.6) as an initial bin edge, we choose four additional parallel curves to break sSFR into six bins. In this way, we are binning galaxies by their distance from the mean of the SFMS, which we define as ΔMS, where $\Delta\mathrm{MS} \equiv \log[\mathrm{sSFR}(M_*)] - \langle\log[\mathrm{sSFR}_{\mathrm{MS}}(M_*)]\rangle$. We follow the bin definitions used in Stephenson et al. (in preparation) and describe them here. The divisions for star-forming galaxies are made based on the assumption that the SFMS has a width of $\sigma \sim 0.25$ dex (see, e.g., Speagle et al., 2014). Highly star-forming (HSF) galaxies lie at least 1σ above the mean of the SFMS and potentially include starburst galaxies. Upper main sequence (UMS) and lower main sequence (LMS) galaxies lie above and below the mean of the SFMS, respectively, but are constrained to be within $\pm 1\sigma$. Bottom of the main sequence (BMS) galaxies lie at least 1σ below the mean of the SFMS but above -0.45 dex. Green valley (GV) galaxies are those between 0.45 dex and 1 dex below the mean of the SFMS, and quiescent (Q) galaxies are all galaxies more than 1 dex below the mean of the SFMS. These bin definitions are summarized in Table 3.2. Figure 3.3 shows the sSFR–M_* plane with our grid plotted as black lines on top of it. The second horizontal curve from the top shows our fit of the mean of the SFMS, from which all other horizontal curvess are derived. See Table E.1 for details on galaxy counts and mean and median values of

stellar mass and ΔMS for both centrals and satellites in the grid.

§ 3.4. Results

The primary objective of this paper is to investigate the clustering properties of galaxies across the sSFR–M_* plane, as shown in Figure 3.3. Previous studies have primarily focused on the clustering properties of galaxies by categorizing them into star-forming/blue or quiescent/red galaxies (see, e.g., Li et al., 2006; Zehavi et al., 2011; Guo et al., 2011; Hearin and Watson, 2013; Tinker et al., 2013), with only a few exceptions examining correlations across the sSFR–M_* plane (Coil et al., 2017; Berti et al., 2021b). To the best of our knowledge, there have been no reports on two-point correlations across the sSFR–M_* plane when distinguishing between central and satellite galaxies or on the cross-correlations between centrals and satellites. As we will discuss in this paper, a detailed examination of cross-correlations is essential for understanding how galaxies are connected to their dark matter halos. In the following sections, we present and discuss our results on galaxy clustering.

As mentioned in Section 3.2.2, we use three different group catalogs for this work. All results in Section 3.4 were calculated using each of the group catalogs, and while we do not show all of these results, we find that they all produce clustering trends that are consistent with each other and lead to the same conclusions we draw in this paper. A discussion comparing the group catalogs can be found in Section 3.6.1.

Finally, note that for clarity in all correlation function figures, we do not plot points where the error exceeds the value of the point itself (unless specified otherwise). Additionally, the label at the top of each column in each correlation function figure shows

the value of the center of that mass bin in log space as a general guide for the stellar masses to which each column corresponds.

3.4.1. Auto-correlation Functions

Figure 3.4 shows the ACFs of all galaxies (centrals and satellites together) as a function of ΔMS. The dark and light blue lines in the upper two rows correspond to galaxies in the upper (HSF+UMS) and lower (LMS+BMS) halves of the SFMS, respectively. The green lines in the third row correspond to green valley galaxies, and the red lines in the bottom row correspond to quiescent galaxies. In each column, there is a black dashed line that remains the same in every row. These lines represent the ACFs of galaxies in the center of the SFMS (LMS+UMS) and serve as a reference to see how the ACFs evolve from high to low ΔMS at a fixed mass. We observe a strong and consistent trend of increasing clustering going below the SFMS. This effect is most pronounced at lower masses where the clustering amplitude increases up to an order of magnitude above the SFMS amplitude. In every mass bin, the increase in the clustering amplitude occurs most strongly at smaller separations, i.e., in the one-halo term where the highly clustered nature of satellites has the greatest impact on the ACFs. This signal, however, becomes weaker in the higher mass bins as there are fewer satellites in the sub-samples at higher masses.

We now investigate the relative contributions of central and satellite galaxies to this observed trend in the ACFs. Figure 3.5 shows the ACFs of central galaxies only. With this sample, we notice a drastic reduction in the trends as a function of ΔMS. The remaining weak trend is only somewhat noticeable in the most quiescent centrals, especially at lower masses. Note that since these ACFs are of centrals only, we focus

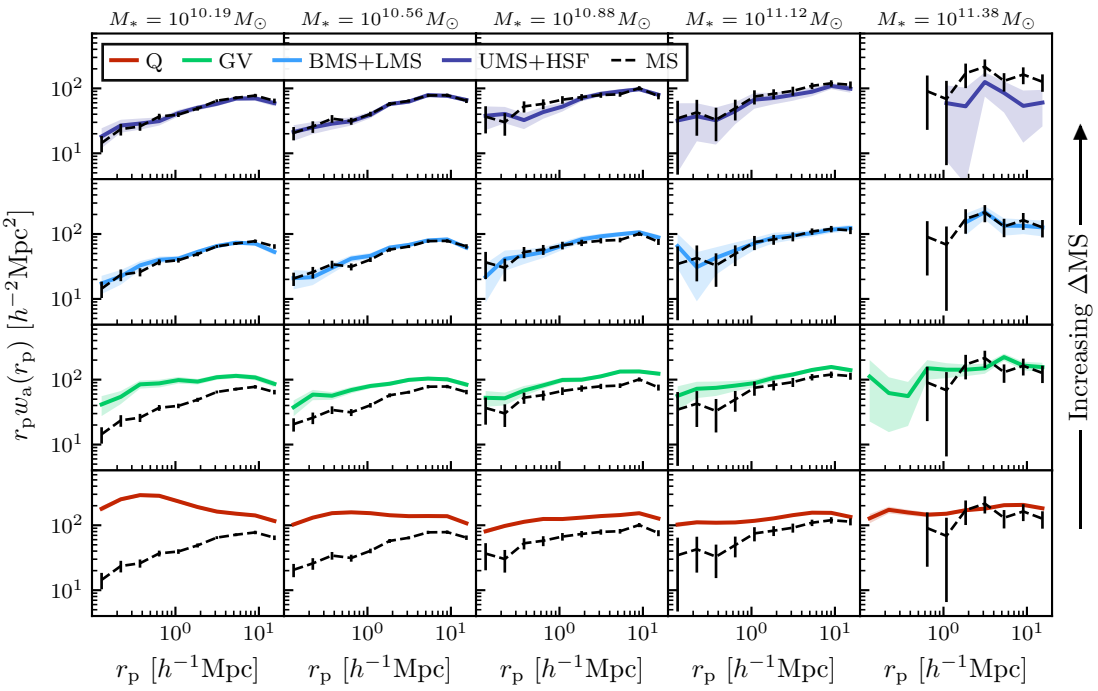

Figure 3.4. Projected auto-correlation functions (ACFs) of all galaxies as a function of ΔMS and M_*. The rows correspond from top to bottom to UMS+HSF galaxies (dark blue), BMS+LMS galaxies (light blue), GV galaxies (green), and quiescent galaxies (red). The colored lines show the results within a grid cell, and the black dashed lines show the ACFs of all galaxies in the center of the SFMS (LMS+UMS). While showing no evolution across the SFMS, there is a clear trend of increasing clustering amplitude as ΔMS decreases below the SFMS. The magnitude of this effect tends to decrease with mass due to the differing galaxy selections in the different sub-volumes. Greater clustering at small separations is primarily the result of the presence of satellite galaxies, and the relative fraction of satellite galaxies within each mass bin decreases as the mass increases (see Figure 3.6).

primarily on scales larger than $r_p \sim 1\ h^{-1}$Mpc as we do not expect multiple centrals to occupy the same halo. Signals below this scale are likely the result of the projection used to calculate the ACFs and potentially satellites that are misidentified as centrals. We discuss this further in Section 3.6.1. We conclude that the strong trend and main contributor of increased clustering amplitude below the SFMS is therefore the product of satellite galaxies in the sample. Further, the ratio of satellite galaxies to central galaxies increases with distance below the SFMS. This effect is strongest at lower masses, and these galaxy samples are precisely where we see the strongest clustering in Figure 3.4.

Figure 3.6 shows the weighted ACFs of all (black), central (dash-dot blue), and satellite (magenta) galaxies to illustrate the relative contributions of centrals and satellites to the overall ACF of all galaxies. Formally, this is given by

$$w_{\mathrm{a,\,all}} = f_{\mathrm{cen}}^2 w_{\mathrm{a,\,cen}} + f_{\mathrm{sat}}^2 w_{\mathrm{a,\,sat}} + 2 f_{\mathrm{cen}} f_{\mathrm{sat}} w_{\mathrm{c,\,cen–sat}} \tag{3.4.1}$$

where f_{cen} and f_{sat} are the fractions of central and satellite galaxies, respectively.* The last term in Equation (3.4.1) is a weighted CCF between centrals and satellites. We have plotted separately the first two terms but omitted this last term in Figure 3.6 to ensure clarity regarding the relative contributions of the ACFs. We will revisit central–satellite CCFs in Section 3.4.2.

In Figure 3.6, the ACFs of centrals (dash-dot blue lines) are seen to remain relatively constant as a function of ΔMS while the ACFs of satellites (magenta lines) grow substantially with decreasing ΔMS, particularly at small separations and most pronounced

*To derive Equation (3.4.1), we use the fact that for two samples i and j, the number of galaxy pairs between them, N_{ij}, can be related to the number density of objects in each sample, n_i and n_j, and the correlation function between them, w_{ij}, as $N_{ij} = n_i n_j (1 + w_{ij})$. Note that for auto-correlations, $i = j$, which yields pair counts $N_i = \frac{1}{2} n_i^2 (1 + w_i)$ where the factor of $\frac{1}{2}$ is to avoid double counting. Equation (3.4.1) can then be derived using the relations $N_{\mathrm{all}} = N_{\mathrm{cen}} + N_{\mathrm{sat}} + N_{\mathrm{cen\text{-}sat}}$ and $n_{\mathrm{all}} = n_{\mathrm{cent}} + n_{\mathrm{sat}}$ (see also Zehavi et al., 2011).

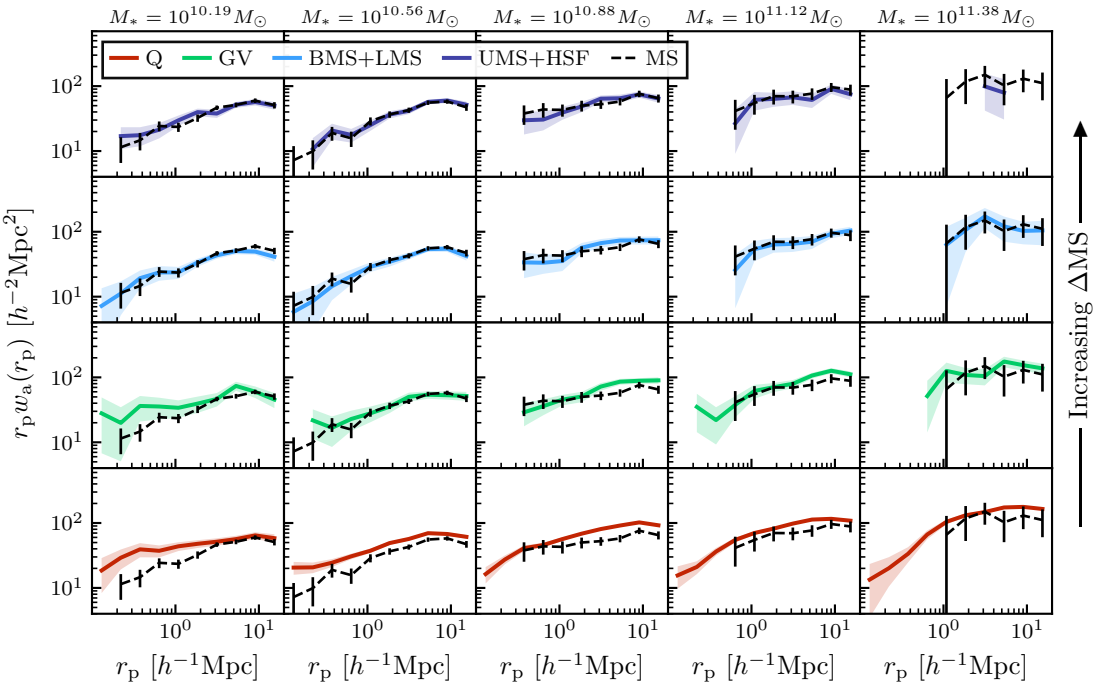

Figure 3.5. Projected auto-correlation functions (ACFs) of central galaxies as a function of ΔMS and M_*. The rows correspond from top to bottom to UMS+HSF galaxies (dark blue), BMS+LMS galaxies (light blue), GV galaxies (green), and quiescent galaxies (red). The colored lines show the results within a grid cell, and the black dashed lines show the ACFs of central galaxies in the center of the SFMS (LMS+UMS). We observe a drastic reduction in the trends as a function of ΔMS compared to the ACFs of all galaxies (see Figure 3.4). The remaining trend is only clear in the most quiescent galaxies. However, because these are ACFs of central galaxies, only scales larger than $r_\mathrm{p} \sim 1\ h^{-1}\mathrm{Mpc}$ should be considered as we do not expect multiple centrals to exist within the same halo. Signals below this scale can be caused by the projection used to calculate w_a and by satellites misidentified as centrals. We discuss this further in Section 3.6.1.

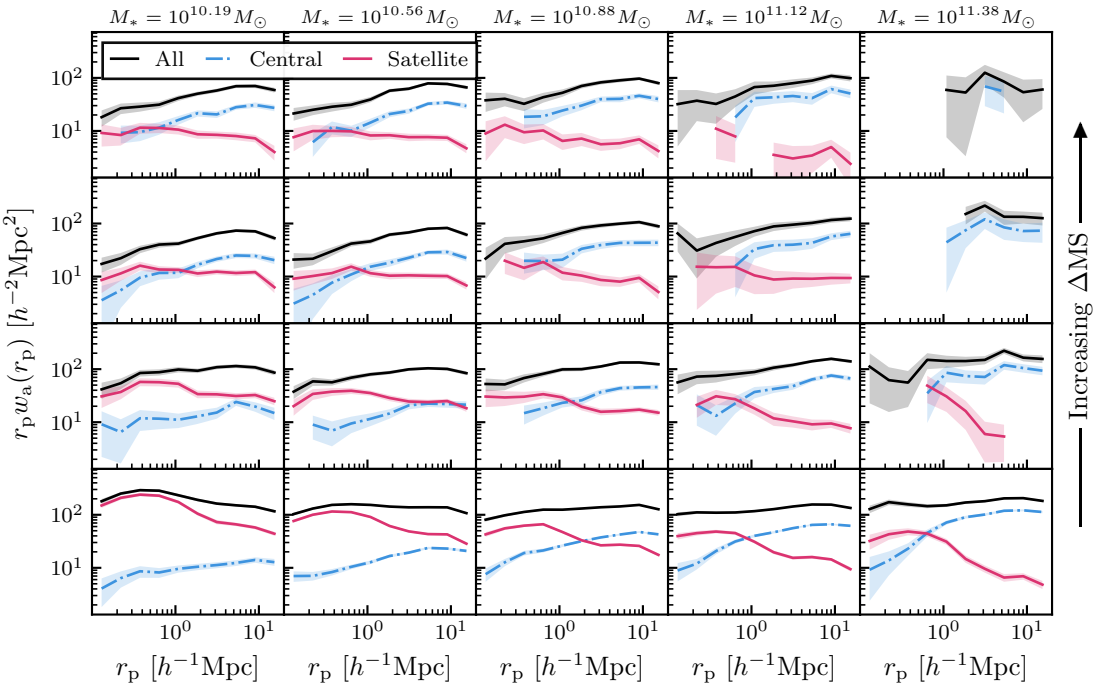

Figure 3.6. Projected auto-correlation functions (ACFs) of all (black), central (dash-dot blue), and satellite (magenta) galaxies as a function of ΔMS and M_*. ACFs of centrals and satellites are scaled by their fractional numbers according to Equation (3.4.1) to show their relative contributions to the overall ACF. The rows correspond from top to bottom to UMS+HSF galaxies, BMS+LMS galaxies, GV galaxies, and quiescent galaxies. The increased clustering with decreasing ΔMS in the ACFs of all galaxies is driven primarily by satellite galaxies, which are highly clustered and tend to be quiescent. On the other hand, the ACFs of central galaxies show little-to-no trend with ΔMS at a fixed stellar mass.

at lower masses where most satellites reside. This reaffirms our conclusion that satellites are the main driver of the trends shown in Figure 3.4 and explains why the trend weakens with increasing mass. As we consider bins of higher mass, the satellite fraction drops, causing the ACFs to become more weighted by contributions from centrals, which generally do not change with ΔMS.

In the existing literature, there is substantial empirical evidence indicating that, on average, a galaxy's stellar mass increases monotonically with halo mass. Because halo clustering also increases with halo mass – often referred to as halo bias – one would expect that the clustering of central galaxies should similarly increase with stellar mass. Indeed, this is what we observe in Figures 3.5 and 3.6. Apart from halo bias, the clustering amplitude of halos can vary based on other factors, such as formation redshift and accretion history, a phenomenon known as assembly bias (see, e.g., Wechsler et al., 2006). The empirical evidence presented in Figures 3.5 and 3.6 – that is, the almost nonexistent dependence of the central galaxy ACFs with ΔMS at a fixed stellar mass – suggests that assembly bias is not producing any significant net effect in central galaxy clustering in the sSFR–M_* plane.

3.4.2. Cross-correlation Functions

Since the ACFs of central galaxies show little trend with ΔMS, we further investigate the clustering properties of centrals by considering cross-correlations of centrals with satellites. However, we want to do this in a way that will not be influenced by the clustering trends of the satellites. If, for example, we calculate CCFs within each cell of the Figure 3.3 grid, we know the clustering amplitudes will be highest in the quiescent cells – similar to Figure 3.4 – because satellites are highly clustered and tend to be

quiescent. To avoid this bias, we instead cross-correlate centrals in a grid cell with *all* satellites in the same mass bin, regardless of satellite sSFR. By doing this, the CCFs will be independent of satellite sSFR and, consequently, unbiased by the clustering trends shown in Figures 3.4 and 3.6.

The central–satellite CCFs are shown in Figure 3.7 using all six ΔMS bins from Figure 3.3. The upper four rows (blue) show results for the SFMS broken into four sections (HSF, UMS, LMS, and BMS), followed by green valley (green) and quiescent (red) centrals. The black dashed lines show the CCFs of centrals in the center of the SFMS (LMS+UMS) within that mass bin with all satellites in the same mass bin. Across the SFMS, we see similar behavior in the CCFs. However, looking at green valley and especially quiescent centrals, there is a general trend of increasing clustering amplitude at smaller separations, i.e., in the one-halo term. For quiescent centrals, the CCF clustering amplitude is greater than that of SFMS centrals by up to a factor of ∼3.

Interpreting this increase in clustering is challenging as there are multiple factors at play in the CCFs. First is the projection that is used when calculating the CCFs. Since they are calculated in a projected environment, pair counts of centrals with satellites of another group along the line of sight can lead to higher clustering amplitudes on small scales. If quiescent centrals tend to live in higher density environments, it may be more likely to find cases of quiescent centrals along the line of sight of a larger group than star-forming centrals. Another possibility is pair counts of centrals at the outskirts of larger halos. Since we calculate the CCFs in mass bins, we are cross-correlating centrals with satellites of similar masses. It is, therefore, likely that satellites of a similar mass to a given central will belong to a different, more massive central. There may then be pair counts from backsplash galaxies (see, e.g., Borrow et al., 2023). These galaxies are

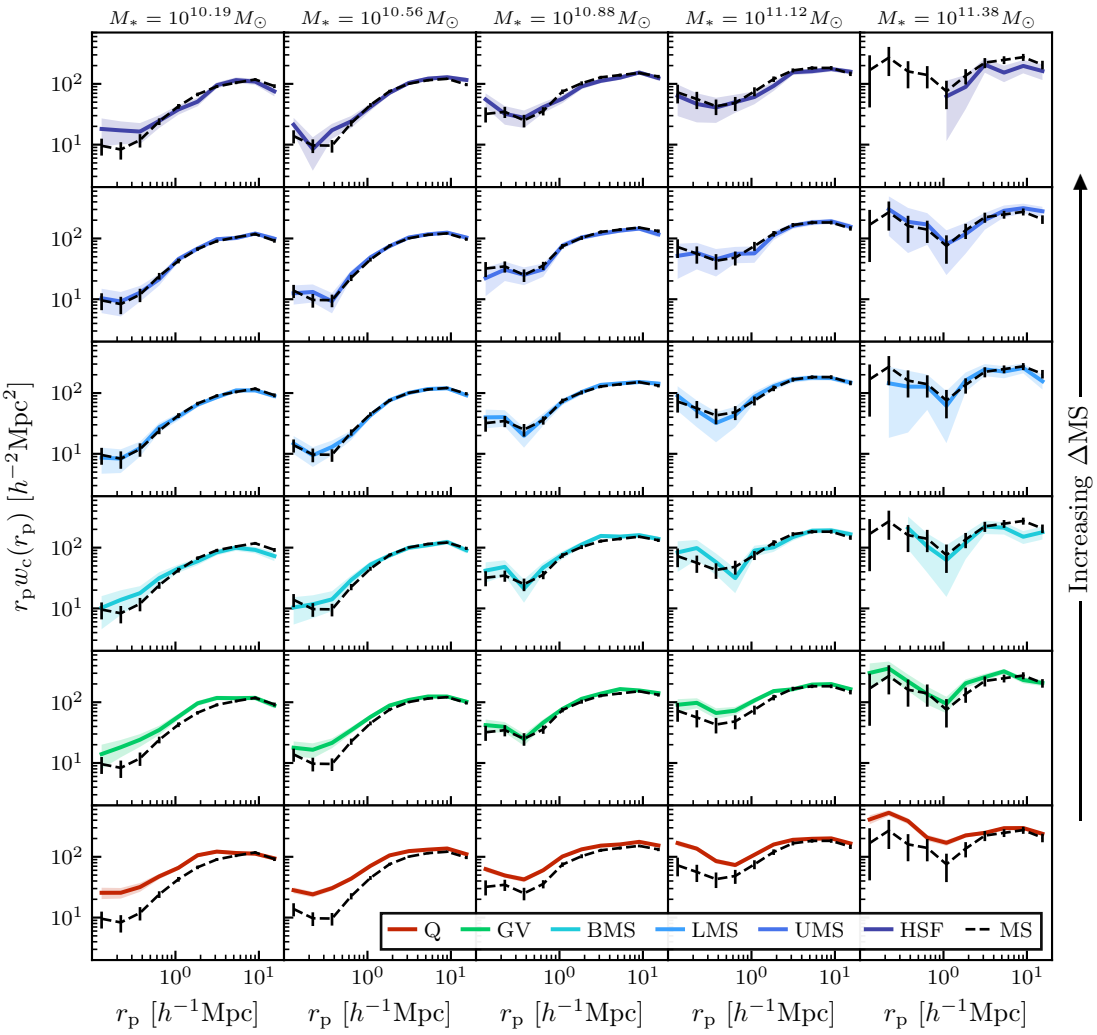

Figure 3.7. Projected cross-correlation functions (CCFs) of central galaxies as a function of ΔMS and M_*. The centrals of a given cell in the grid are cross-correlated with all other satellites in the same mass bin. For comparison, a black dashed line is included that represents the CCF of all centrals in the center of the SFMS (LMS+UMS) with all satellites in the same mass bin. While there is no strong trend in the CCFs across the SFMS (top four rows in blue), we do see a trend of increasing clustering at small separations for GV centrals (green) and a further increase for quiescent centrals (red). Because this increase occurs at small separations, it suggests that more quiescent centrals will tend to have more satellite galaxies.

ones that once resided in a larger cluster but have moved outward on their orbits. In the process of this migration, their gas becomes stripped, resulting in these galaxies tending to be quiescent. A final contributor to the small-scale CCFs is centrals that indeed have satellites with masses that are similar enough to be in the same mass bin. While these cases may be rare, it is possible that they could contribute in a significant way to the small-scale CCFs depending on the frequency of occurrences of the other previously mentioned factors. If this is the case, it would suggest that quiescent centrals tend to have more satellites than star-forming centrals.

To check this conclusion explicitly, we calculate the mean number of satellites per group above a stellar mass of $\log(M_*/M_\odot) = 10$ for all, SFMS, green valley, and quiescent centrals. For this, we create a new volume-limited sample that spans a mass range of $10 < \log(M_*/M_\odot) < 12$ using the same methodology described in Section 3.2.3. The left panel of Figure 3.8 shows the results, plotting only the mass bins that contain at least 50 centrals. As suggested by the CCFs, we see a clear separation in the number of satellites for SFMS and quiescent centrals, with green valley centrals lying generally in the middle at higher central masses. The separation is ~0.15 dex at lower masses, growing up to ~0.5 dex at higher masses. Note that in the halo occupation distribution (HOD) framework, $\langle N_{\rm sat} \rangle$ will be proportional to halo mass as $\langle N_{\rm sat} \rangle \propto M_{\rm vir}^{\alpha}$ with $\alpha \sim 1$ (e.g., Moster et al., 2010; Zehavi et al., 2011). Under the assumptions of HOD, where halo mass determines all galaxy properties, the segregation we see in $\langle N_{\rm sat} \rangle$ with ΔMS reflects the existence of the SHMR segregation, where quiescent galaxies live in higher mass halos than star-forming galaxies with the same stellar mass.

The right panel of Figure 3.8 further investigates $\langle N_{\rm sat} \rangle$ by dividing the SFMS into the HSF, UMS, LMS, and BMS sub-samples. In the stellar mass range of $10.3 \lesssim$

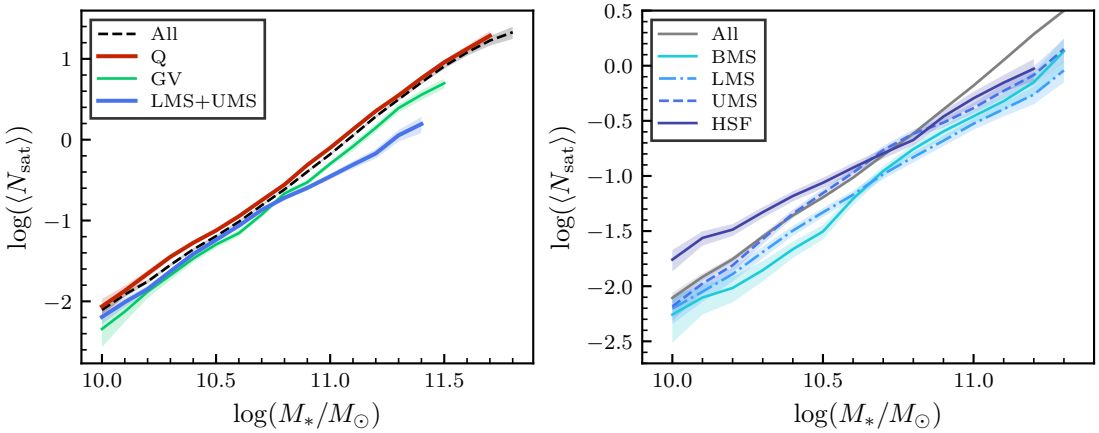

Figure 3.8. Average number of satellites per central as a function of central galaxy stellar mass. Only mass bins that contain at least 50 centrals are plotted. *Left panel*: The sample is broken into sub-samples of all, LMS+UMS, GV, and quiescent centrals. We see that quiescent centrals on average have more satellites than SFMS centrals, which agrees with our conclusion from the CCFs in Figure 3.7. At lower masses ($\log(M_*/M_\odot) \lesssim 10.5$), the average difference is ~0.15 dex, while at higher masses, it grows up to ~0.5 dex. *Right panel*: The SFMS is broken into four sub-samples as a function of ΔMS. Note that the axes have changed to better show the results. For $10.3 \lesssim \log(M_*/M_\odot) \lesssim 10.6$, a clear monotonic trend is observed with SFR activity, where higher ΔMS sub-samples have higher $\langle N_{\rm sat}\rangle$. For lower masses, the trend is less evident for UMS, LMS, and BMS centrals though HSF centrals clearly have the most satellites, exceeding the average of the entire sample (gray solid line, equivalent to the black dashed line in the left panel). At higher masses, the trend is less clear for all four sub-samples.

$\log(M_*/M_\odot) \lesssim 10.6$, a clear monotonic trend is observed with SFR activity, where higher ΔMS sub-samples have higher $\langle N_{\rm sat} \rangle$. However, for masses below this range, the trend is less evident for UMS, LMS, and BMS. Similarly, for masses above this range, the trend is less distinct for all four sub-samples. For central galaxies with masses $\log(M_*/M_\odot) \lesssim 10.6$, HSF centrals host more satellites than UMS, LMS, and BMS centrals. Moreover, on average, HSF centrals host ∼1.6 times more satellite galaxies compared to the entire sample (gray solid line), while UMS, LMS, and BMS centrals generally host fewer satellites than the overall sample. This trend for HSF centrals aligns with our findings in Figure 3.7, where the one-halo term of HSF centrals in the lowest mass bin is marginally above that of all SFMS galaxies. The excess of satellite galaxies for HSF centrals may contribute to their elevated SFR values due to close interactions with their satellites (see, e.g., Lin et al., 2008; Yesuf et al., 2021; Bottrell et al., 2024).

§ 3.5. Implications for the Galaxy–Halo Connection

With the signal we have seen in the clustering properties of central galaxies – namely, that more quiescent centrals tend to have more satellites than star-forming centrals (shown in Figures 3.7 and 3.8) – we can now use this to test models of the galaxy–halo connection. To do this, we use the Bolshoi–Planck (BP) N-body simulation (Klypin et al., 2016b; Rodríguez-Puebla et al., 2016b). The BP simulation box is 250 h^{-1}Mpc on each side, containing 2048^3 particles with a mass resolution of $1.55 \times 10^8\ h^{-1}\mathrm{M}_\odot$ and force resolution of 1.0 h^{-1}kpc. The cosmological parameters of BP are $\Omega_\Lambda = 0.693$, $\Omega_{\rm m} = 1 - \Omega_\Lambda$, $\Omega_{\rm b} = 0.048$, $h = 0.678$, $n_s = 0.96$, and $\sigma_8 = 0.823$. Dark matter halos and sub-halos in the simulation were identified using ROCKSTAR (Behroozi et al., 2013b).

Since our correlation functions thus far have been calculated within five different sub-volumes, we use five different sub-volumes created from BP snapshots for these tests. The goal is to have each snapshot cover the same stellar mass and redshift ranges as our five SDSS sub-volumes.

3.5.1. Mock Catalogs: Redshift Space and Stellar Masses

To create our five samples using BP, we first project snapshots into redshift space. We start by defining an origin that is centered on one side of the simulation box. Assuming the snapshot corresponds to a redshift $z_{\rm BP}$, we shift all halos away from the origin a distance $d_{\rm c}(z_{\rm BP})$, where $d_{\rm c}(z_{\rm BP})$ is the comoving distance to $z_{\rm BP}$. This sets the redshift of the near side of the box equal to $z_{\rm BP}$. Then, we infer cosmological redshifts for all halos moving across to the far side of the box based on their comoving distances from the origin. Additionally, we add redshift-space distortions according to the equation

$$z_{\rm obs} = z_{\rm cos} + \frac{v_{\rm los}}{c}\,(1 + z_{\rm cos}) \tag{3.5.1}$$

where $z_{\rm obs}$ is the observed redshift with redshift-space distortions, $z_{\rm cos}$ is the cosmological redshift inferred from a halo's comoving distance from the origin, $v_{\rm los}$ is the line-of-sight peculiar velocity of a halo relative to the origin, and c is the speed of light. Once every halo has an observed redshift assigned to it, we then cut the snapshots so each one has the same redshift range as the corresponding sub-volume in our SDSS samples. In some cases, the snapshot is not large enough to encompass the entire redshift range of a sub-volume. For these instances, we choose BP snapshots that give

us a redshift range that is as close as possible to our SDSS sub-volume redshift ranges.*

Once all the snapshots and redshift ranges have been determined, we assign stellar masses to all of the halos. To do this, we follow Dragomir et al. (2018, see also Reddick et al., 2013; Calette et al., 2021) and assume that the halo property that best correlates with stellar mass and reproduces the observed ACF is

$$V_{\rm DM} = \begin{cases} V_{\rm max} & \text{for distinct halos} \\ V_{\rm peak} & \text{for sub-halos} \end{cases} \tag{3.5.2}$$

where $V_{\rm max}$ is the maximum circular velocity of dark matter in a distinct halo at the observed time and $V_{\rm peak}$ is the maximum circular velocity throughout the entire history of a sub-halo. Moreover, we follow Rodríguez-Puebla et al. (2012, see also Rodríguez-Puebla et al., 2013; Guo et al., 2016; Devi et al., 2019) to separately derive $V_{\rm DM}$–M_* relations for centrals and satellites. As shown in Rodríguez-Puebla et al. (2012), using separate relations for centrals and satellites results in ACFs that are more consistent with observations than using the same relation for both.

We follow the same procedure described in Section 2.2 of Calette et al. (2021) to derive the $V_{\rm DM}$–M_* relations based on the numerical deconvolution algorithm discussed in Appendix D of Rodríguez-Puebla et al. (2020b). One difference with respect to Calette et al. (2021) is that we assume that their Equation (3) can be written separately for centrals and satellites as

$$\phi_i(M_*, z) = \iint \frac{\mathcal{H}_i(M_*|V_{\rm DM})\phi_{{\rm DM},i}(V_{\rm DM})}{\mathcal{V}_{\rm c}(z_f) - \mathcal{V}_{\rm c}(z_i)} {\rm d}\log(V_{\rm DM}){\rm d}\mathcal{V}_{\rm c}(z), \tag{3.5.3}$$

where the subscript i refers to either centrals/distinct halos or satellites/sub-halos,

*For the BP snapshots we use, all redshift boundaries are within 0.01 of the corresponding SDSS sub-volume redshift boundaries.

$\mathcal{V}_c$ is the comoving volume, and $z_i = 0.001$ and $z_f = 0.2$ correspond to the redshift ranges over which the GSMF that we use has been constrained. The function $\phi_{\mathrm{DM},i}(V_{\mathrm{DM}})$ is the halo/sub-halo velocity function and $\mathcal{H}_i(M_*|V_{\mathrm{DM}})$ refers to the log-normal conditional probability distribution function that a halo/sub-halo with velocity $\log(V_{\mathrm{DM}}) \pm \frac{1}{2}\mathrm{d}\log(V_{\mathrm{DM}})$ hosts a galaxy with a mass $\log(M_*) \pm \frac{1}{2}\mathrm{d}\log(M_*)$. We assume that the dispersion around the V_{DM}–M_* relation is $\sigma = 0.15$ dex and the same for both central and satellite galaxies. Note that Equation (3.5.3) returns the volume-weighted V_{DM}–M_* relation of centrals and satellites.

To derive the V_{DM}–M_* relationship, we use the GSMF of all galaxies, $\phi(M_*)$, from Dragomir et al. (2018). We calculate the GSMF of central and satellite galaxies by computing the fraction of satellites, f_{sat}, as a function of stellar mass using the volume-limited samples described in Section 3.2. The GSMF of central and satellite galaxies will then, by definition, be given by $\phi_i(M_*) = f_i\phi(M_*)$ with $f_{\mathrm{cen}} = 1 - f_{\mathrm{sat}}$. Using $\phi_i(M_*)$ on the left side of Equation (3.5.3) allows us to deconvolve the equation to isolate $\mathcal{H}_i(M_*|V_{\mathrm{DM}})$ and calculate stellar masses for all the dark matter halos. Once the halos have stellar masses assigned to them, we select our final BP samples by choosing the corresponding stellar mass and redshift ranges to match our five SDSS sub-volumes.

3.5.2. Assigning sSFRs to Dark Matter Halos

We now test different models of sSFR for central galaxies. For this step, we consider only centrals/distinct halos as we expect satellites/sub-halos to have different formation and evolutionary histories from the host halos in which they reside. Note that we model sSFR empirically using our snapshots and SDSS observations at $z \leq 0.2$; we do not attempt to model the histories of the haloes. For a given model, we choose

a halo property, H, with which sSFR will correlate as either a monotonic increasing or decreasing relationship.* Then, in bins of stellar mass, we rank order all halos by $H(M_*)$. Assuming there are N halos in a given mass bin, we sample N sSFRs from the corresponding mass bin of SDSS observations. These sampled sSFRs are also rank ordered and then matched to the halos based on rank (i.e., if H and sSFR are positively correlated, the highest sampled sSFR will be assigned to the halo with the highest value of $H(M_*)$).

Using this procedure, we test three different models of sSFR: halo accretion rate** (averaged over a dynamical time), $\dot{M}_{\rm h}$, halo concentration, $C_{\rm vir}$, and peak circular velocity, $V_{\rm peak}$, which are summarized in Table 3.3. As mentioned in Section 3.1, halo accretion rate controls the influx of gas into a galaxy's interstellar medium (see, e.g., Avila-Reese and Firmani, 2000; Bouché et al., 2010; Dekel and Mandelker, 2014; Rodríguez-Puebla et al., 2016) and has been shown to reproduce the width of the SFR distribution in star-forming galaxies at various redshifts (Rodríguez-Puebla et al., 2016) and, to a lesser extent, the clustering of galaxies (Becker, 2015). Previous studies of $C_{\rm vir}$ have shown that it is linked to the timing of gas infall into halos (see, e.g., Avila-Reese et al., 1998; Wechsler et al., 2002; Gao et al., 2004; Dutton et al., 2010; Matthee and Schaye, 2019), and models of galaxy color based on $C_{\rm vir}$ have successfully reproduced the observed ACFs of red and blue galaxies (Hearin and Watson, 2013).

The $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models will be used as tests of whether halo assembly history is the key factor determining how galaxies cluster in the sSFR–M_* plane. At a fixed halo mass, halos with low $\dot{M}_{\rm h}$ will tend to be the ones in denser environments comprised of

*These are the most straightforward trends that have a physical basis for sSFR to be correlated with halo property H.

**Since we are modeling *specific* SFR, it is also sensible to correlate this with specific halo accretion rate ($\dot{M}_{\rm h}/M_{\rm h}$). We tested this model as well and found similar results to the $\dot{M}_{\rm h}$ model.

Table 3.3.: Description of the three sSFR models of the galaxy–halo connection. The correlation direction tells whether a given halo property correlates with sSFR positively (higher halo property corresponds to higher sSFR) or negatively (higher halo property corresponds to lower sSFR).

Halo Property	sSFR Correlation	Motivation
Halo accretion rate ($\dot{M}_{\rm h}$)	Positive	Halo accretion rate controls the influx of gas into a galaxy's interstellar medium. We might expect galaxies in halos with higher accretion rates may have higher sSFR as they have more gas available to form stars.
Halo concentration ($C_{\rm vir}$)	Negative	Halo concentration is linked to the timing of gas infall into halos where more concentrated halos experience earlier infall. We might expect galaxies in highly concentrated halos have low sSFR as they have run out of gas and ceased star formation.
Halo peak circular velocity ($V_{\rm peak}$)	Negative	$V_{\rm peak}$ is strongly correlated with halo mass. If quiescent galaxies reside in more massive halos, we can assign lower values of sSFR to halos with higher $V_{\rm peak}$.

already-formed halos (see, e.g., Maulbetsch et al., 2007; Lee et al., 2017b) and sub-halos with less dark matter available for accretion. We might expect that the lower accretion rate leads to less gas available to a halo's hosted galaxy and therefore a lower sSFR. In the $\dot{M}_{\rm h}$ model, we then expect quiescent galaxies to be more clustered than star-forming galaxies at a fixed stellar mass. Similarly, at a fixed halo mass, halos with high $C_{\rm vir}$ will tend to be the ones that formed at earlier times and are more clustered. We then might expect that the galaxies hosted by such halos also formed earlier and reached the ends of their life cycles (i.e., became quiescent) earlier. In the $C_{\rm vir}$ model, we can then expect that quiescent galaxies will again be more clustered than star-forming galaxies at a fixed stellar mass.

Contrary to these models, the $V_{\rm peak}$ model will be a test that assumes only a segregation in the SHMR. $V_{\rm peak}$ is strongly correlated with halo mass; therefore, assigning lower sSFR to halos with higher $V_{\rm peak}$ will lead to quiescent galaxies occupying more massive halos than star-forming galaxies at the same stellar mass. It is important to note that the motivation for this model is empirical according to the results from gravitational weak lensing (Mandelbaum et al., 2006), galaxy groups and clustering (Tinker et al., 2013; Rodríguez-Puebla et al., 2015), and galaxy kinematics (More et al., 2011), as discussed in Section 3.1. In this way, the difference in clustering for quiescent and star-forming centrals will be driven by halo bias. This could be tested more directly with a model based on halo mass rather than $V_{\rm peak}$. However, we tested such a model and found that $V_{\rm peak}$ overall provided results more consistent with observations.

3.5.3. Clustering Predictions

For each of the models, we calculate ACFs and CCFs using the same methodology described in Sections 3.3 and 3.4. Note that because we cross-correlate centrals with all satellites in the same mass bin regardless of satellite sSFR, these results do not depend on a model of sSFR for satellites. We show the results of the ACFs and CCFs for the $\dot{M}_{\rm h}$, $C_{\rm vir}$, and $V_{\rm peak}$ models in Figures 3.9 and 3.10, respectively. The rows correspond to HSF+UMS, LMS+BMS, GV, and quiescent centrals from top to bottom. We plot the ACFs and CCFs of SDSS observations as circles and the models as lines.

We first consider the ACFs of the three models, shown in Figure 3.9. The predicted ACFs agree fairly well with the SDSS observations, showing little to no trend with ΔMS at a fixed stellar mass. This result is initially counterintuitive for all models. In the case of the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models, it may be expected that the assembly bias effects built into them would result in quiescent galaxies clustering more strongly than star-forming galaxies of the same stellar mass. For the $V_{\rm peak}$ model, we effectively assign lower sSFRs to higher mass haloes at the same stellar mass, leading to the expectation that halo bias would cause quiescent galaxies to cluster more strongly than star-forming galaxies at a given stellar mass. However, neither of these effects is observed. To understand this result, one must consider the behavior of halo bias as a function of halo mass. The magnitude of halo bias is a function of halo mass and ramps up significantly at higher masses around $\log(M_{\rm h}/M_{\odot}) \gtrsim 13.5$ (see, e.g., Wechsler and Tinker, 2018b). In the $V_{\rm peak}$ model, the highest halo masses for star-forming and quiescent galaxies are, on average, around $\log(M_{\rm h}/M_{\odot}) \sim 13$ and $\log(M_{\rm h}/M_{\odot}) \sim 13.5$, respectively (see Section 3.5.4). At these halo masses, the difference in halo bias is not significant enough to produce a signal in the auto-correlations. For the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models, the highest

halo masses for star-forming and quiescent galaxies are around $\log(M_{\rm h}/M_{\odot}) \sim 13.9$ and $\log(M_{\rm h}/M_{\odot}) \sim 13.4$, respectively. Note that these models predict that quiescent galaxies reside in *lower* mass haloes than star-forming galaxies at fixed stellar mass. Because these halo masses are higher, the difference in halo bias is stronger. It may then be the case that halo bias increases the clustering of star-forming galaxies while assembly bias increases the clustering of quiescent galaxies, resulting in no net difference in the ACFs as a function of ΔMS. It is also worth noting that assembly bias is an effect at fixed *halo* mass, and it is not straightforward how it translates when working at fixed *stellar* mass as we are here.

Looking now at the CCFs of the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models in Figure 3.10, we see that these models produce very similar results. They agree well with the observations at the lowest masses, aside from the quiescent centrals where they predict higher clustering in the one-halo term. In the second-lowest mass bin, there is an upturn in clustering amplitude in the one-halo term for the higher star-forming centrals. This upturn grows with stellar mass and disagrees strongly with what we observed in the SDSS. As mass increases, the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models produce a trend with ΔMS *opposite* of what we see in observations; clustering is highest for the star-forming centrals and lowest for quiescent centrals.

In contrast to the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models, the $V_{\rm peak}$ model CCFs produce reasonable agreement with the SDSS observations at all masses but especially in the three higher mass bins. At lower masses, the SFMS clustering has a tendency to be over-predicted in the one-halo term. However, the main success of this model is that it generally shows increased clustering amplitude for quiescent galaxies, which agrees with the trend in the observations, unlike what the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models predicts.

An important thing to note from these correlation function results is the fact that the

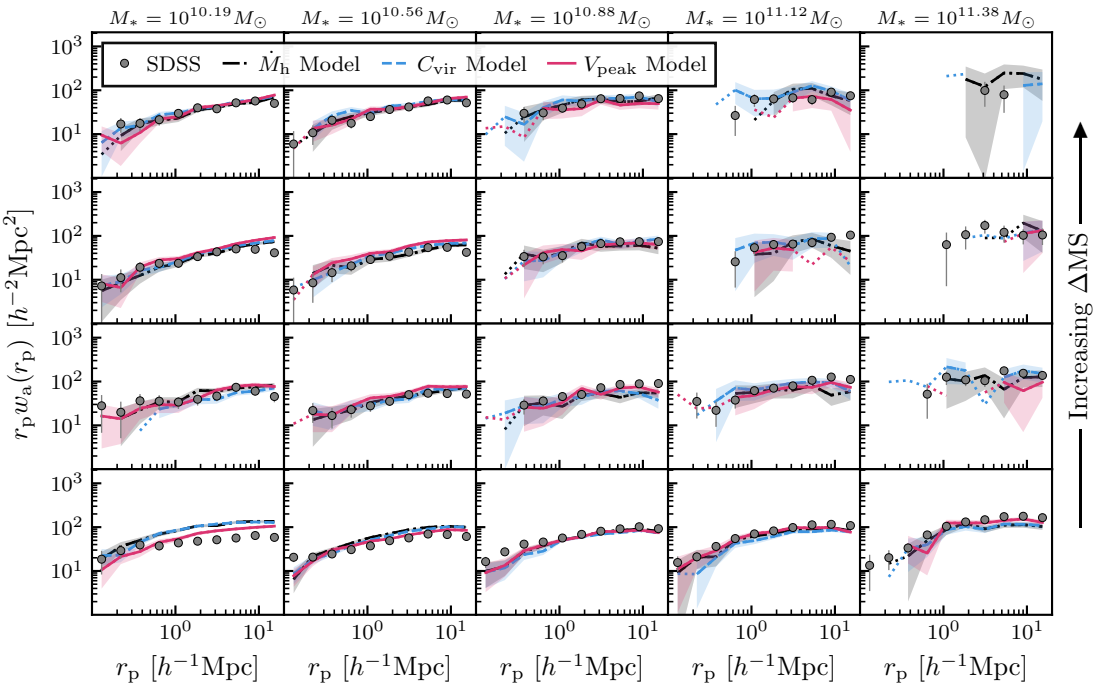

Figure 3.9. Projected auto-correlation functions (ACFs) of central galaxies as a function of ΔMS and M_*. The rows correspond from top to bottom to UMS+HSF, BMS+LMS, GV, and quiescent centrals. The gray circles show results for SDSS observations and the lines show results of three different models of sSFR based on different halo properties using the Bolshoi–Planck simulation. The dash-dot black lines show the $\dot{M}_\mathrm{h}$ model, the dashed blue lines show the C_vir model, and the solid magenta lines show the V_peak model. Dotted lines show results (without error bars) where the error reaches up to $2\omega_\mathrm{a}$. We see that all models are consistent with SDSS and produce ACFs that are very similar.

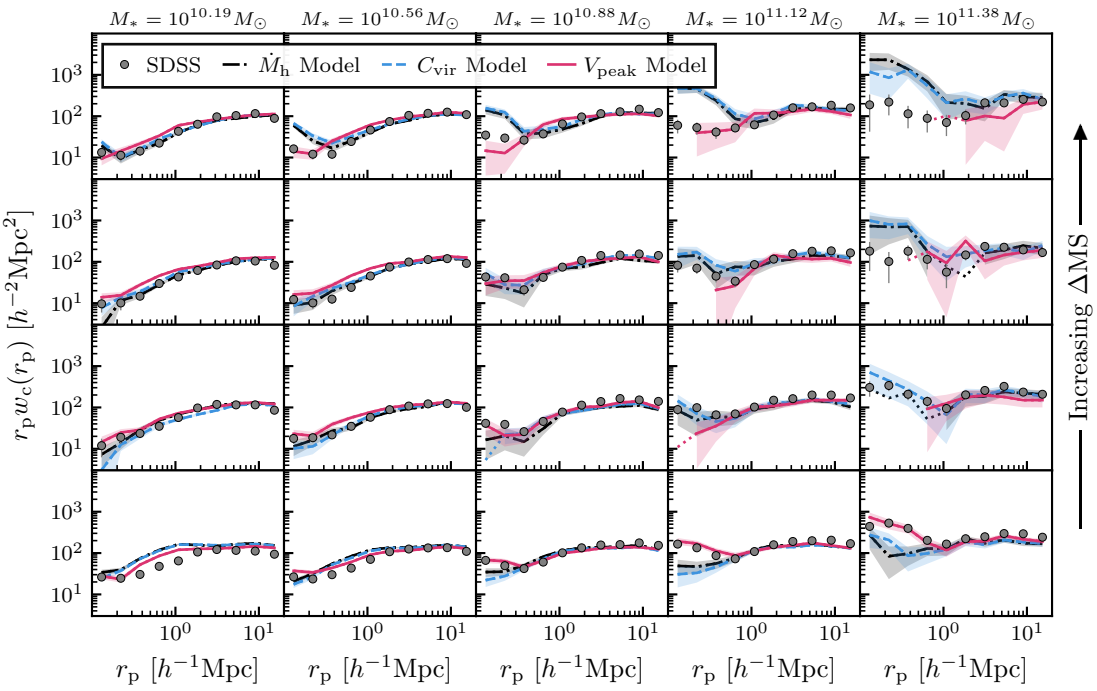

Figure 3.10. Projected cross-correlation functions (CCFs) of centrals in a given grid cell with all satellites in the same mass bin. The rows correspond from top to bottom to UMS+HSF, BMS+LMS, GV, and quiescent centrals. The gray circles show results for SDSS observations and the lines show results of three different models of sSFR based on different halo properties using the Bolshoi–Planck simulation. The dash-dot black lines show the $\dot{M}_{\rm h}$ model, the dashed blue lines show the $C_{\rm vir}$ model, and the solid magenta lines show the $V_{\rm peak}$ model. Dotted lines show results (without error bars) where the error reaches up to $2w_{\rm c}$. We see that the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models for sSFR fail to reproduce the CCF trend in ΔMS that we find in observed data. At higher masses, clustering amplitude is highest for star-forming centrals, disagreeing strongly with SDSS. However, the $V_{\rm peak}$ model is consistent with SDSS at all masses.

CCFs showed differing results among the models, whereas the ACFs did not. In almost every region of our sSFR–M_* plane, all three models predict ACFs that are nearly the same while the $V_{\rm peak}$ model was able to clearly distinguish itself via CCFs. What this tells us is that when considering only central galaxies, ACFs alone may not be sufficient to discriminate different models of galaxy sSFR. We do add, however, that since the presence of satellites has a strong impact on clustering trends with sSFR, it is possible that ACFs of *all* galaxies (i.e., centrals and satellites together) could differentiate models of galaxy sSFR, though we are unable to test this given that we do not attempt to model sSFRs for sub-halos.

Finally, we consider the predictions of $\langle N_{\rm sat} \rangle$ by the three models, shown in Figure 3.11. In the SDSS observations, there is a consistent trend at all masses of quiescent galaxies having more satellites than star-forming galaxies, and this trend is weakest at lower masses and strongest at higher masses. In these regards, we again see clear failures in the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models. While the $\dot{M}_{\rm h}$ model does predict lower $\langle N_{\rm sat} \rangle$ for star-forming galaxies at lower masses, there is a reversal of this trend above $\log(M_*/M_\odot) \approx 10.4$, contrary to what the observations show. Additionally, the $\langle N_{\rm sat} \rangle$ values for quiescent and star-forming galaxies are diverging at lower masses rather than converging as the observations do. The $C_{\rm vir}$ model incorrectly predicts that star-forming galaxies have greater $\langle N_{\rm sat} \rangle$ at all masses, opposite of what the observations show, and has converging values of $\langle N_{\rm sat} \rangle$ for quiescent and star-forming galaxies at both high and low masses. On the other hand, the $V_{\rm peak}$ model agrees with the observations in its prediction that quiescent galaxies have greater $\langle N_{\rm sat} \rangle$ than star-forming galaxies at all masses, with the difference being the greatest at higher masses. The failures of the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models show that conventional assembly bias models of sSFR do not reproduce the observed

clustering of central galaxies. The successes of the $V_{\rm peak}$ model, however, show that the clustering of central galaxies in the sSFR–M_* plane can be largely explained by a mass-based model, independent of assembly bias.

3.5.4. Stellar-to-Halo Mass Relations

In this section, we analyze the SHMR predictions of the three models. Figure 3.12 shows the resulting SHMRs in the left panels and the inverted SHMRs in the right panels for each of the three models described in the preceding sections. We note that none of the three models is affected by the inversion problem (Cui et al., 2021). In each panel, the blue and red lines represent the SFMS centrals and a sample consisting of quiescent and green valley centrals (Q+GV), respectively. When using $V_{\rm peak}$ (top panels), SFMS galaxies exhibit a larger stellar mass than Q+GV galaxies at a fixed halo mass (left panel). This outcome is a consequence of assigning increasing values of sSFR to decreasing values of $V_{\rm peak}$, as illustrated in the right panel of Figure 3.12. Since $V_{\rm peak}$ strongly correlates with halo mass, this effectively results in assigning quiescent galaxies to more massive halos than star-forming galaxies at the same stellar mass (consequently, quiescent galaxies will then have lower stellar masses than star-forming galaxies at the same halo mass). The $C_{\rm vir}$ and $\dot{M}_{\rm h}$ models (middle and bottom panels) result in opposite trends to the $V_{\rm peak}$ model, both in the SHMR and the inverted SHMR. These trends can be understood as follows. In the $C_{\rm vir}$ model, we assign increasing values of sSFR to decreasing values of concentration at a fixed stellar mass. Since at $z \sim 0$ halo concentration correlates with halo mass as $C_{\rm vir} \propto M_{\rm vir}^{-0.1}$ (Macciò et al., 2008; Klypin et al., 2016b), we are effectively assigning star-forming galaxies to high-mass halos. For the $\dot{M}_{\rm h}$ model, increasing values of sSFR were assigned to increasing values

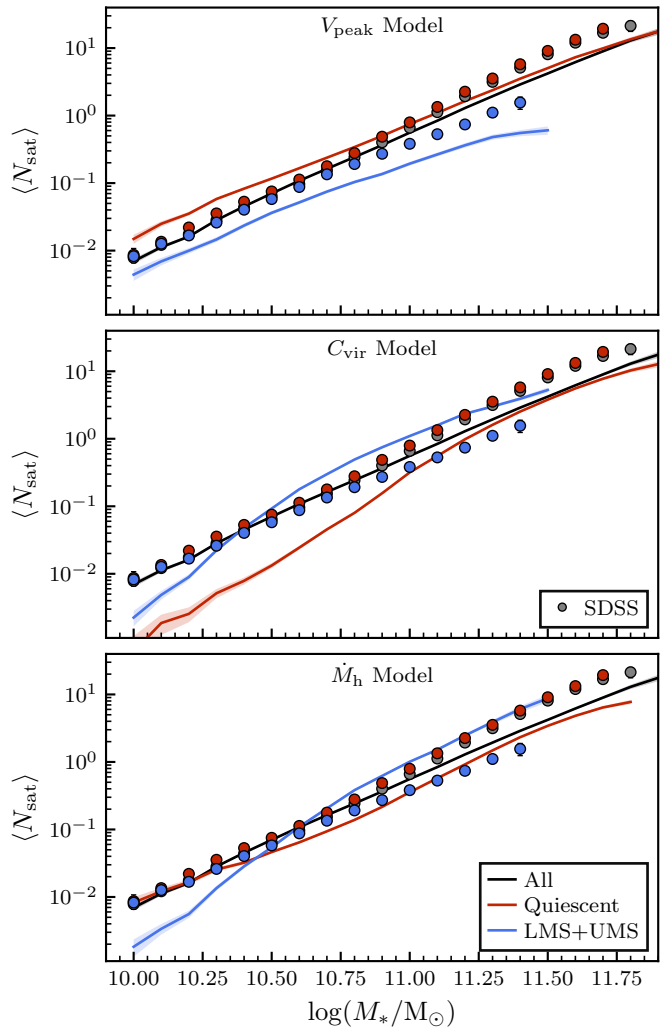

Figure 3.11. Average number of satellites per central as a function of central galaxy stellar mass. Only mass bins that contain at least 50 centrals are plotted. The sample is broken into sub-samples of all, LMS+UMS, and quiescent centrals. The circles show results for SDSS observations and the lines show results of three different models of sSFR based on different halo properties using the Bolshoi–Planck simulation. The $C_{\rm vir}$ model predicts that star-forming centrals have a greater $\langle N_{\rm sat} \rangle$ than quiescent centrals, contrary to the observations. This also occurs in the $\dot{M}_{\rm h}$ model above $\log(M_*/M_\odot) \approx 10.4$ where the $\langle N_{\rm sat} \rangle$ relation inverts. The $V_{\rm peak}$ model, on the other hand, agrees with the observations that quiescent centrals have more satellites than star-forming centrals at all masses.

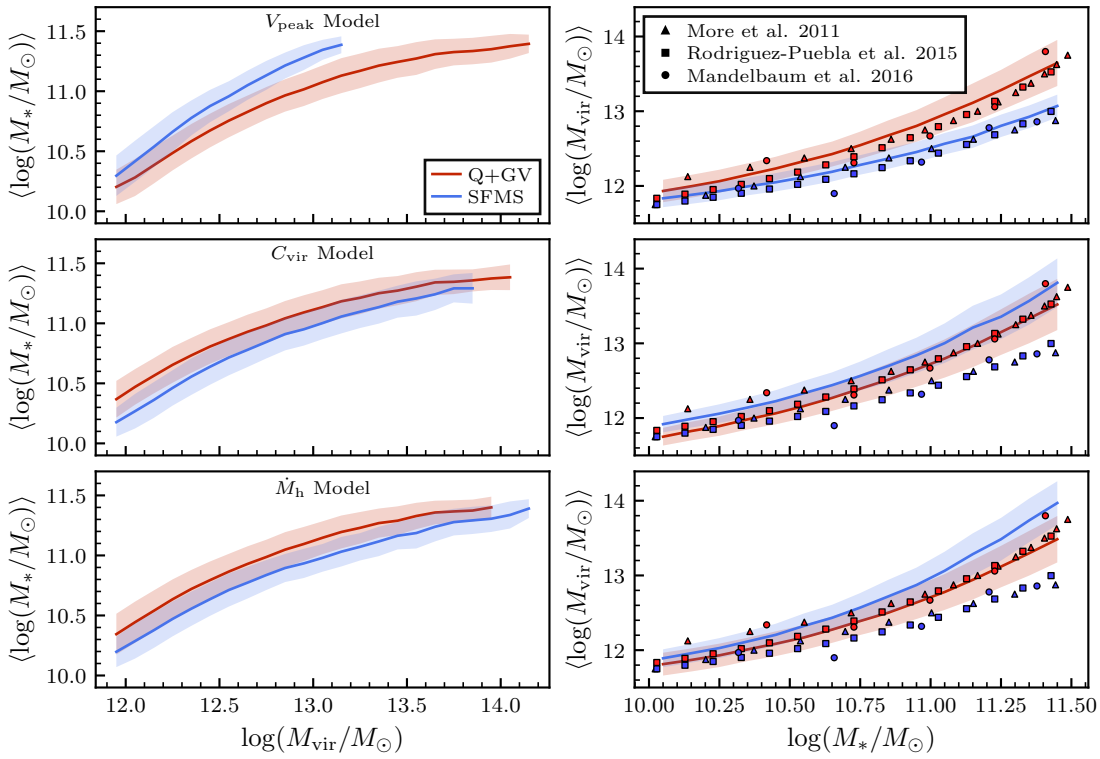

Figure 3.12. The stellar-to-halo mass relation (SHMR), $\langle M_*(M_{\rm vir}) \rangle$ (left panels), and the inverted SHMR $\langle M_{\rm vir}(M_*) \rangle$ (right panels), for quiescent and green valley (Q+GV) and SFMS centrals using the $V_{\rm peak}$, $C_{\rm vir}$, and $\dot{M}_{\rm h}$ models of sSFR. Only mass bins with at least 50 galaxies are plotted for the different models. Also shown in the right panels are determinations using satellite kinematics (More et al., 2011), galaxy groups combined with galaxy clustering (Rodríguez-Puebla et al., 2015), and weak lensing (Mandelbaum et al., 2016). We see that while all models produce inverted SHMRs that agree with observations for Q+GV galaxies, only the $V_{\rm peak}$ model reproduces the observed relation for SFMS galaxies. We also note that our model results are not affected by the inversion problem.

of halo accretion rate at a given stellar mass. Since $\dot{M}_{\rm h} \propto M_{\rm vir}^{1.1}$ (Fakhouri et al., 2010; Rodríguez-Puebla et al., 2016b), we are again effectively assigning star-forming galaxies to high-mass halos.

Our findings reveal a distinct segregation in the SHMR when considering SFMS and Q+GV centrals across the three models of sSFR. When compared to earlier determinations using satellite kinematics (More et al., 2011), galaxy groups combined with galaxy clustering (Rodríguez-Puebla et al., 2015), and weak lensing (Mandelbaum et al., 2016), we observe that all models agree for Q+GV galaxies, but only the $V_{\rm peak}$ model aligns with the trend where quiescent galaxies inhabit more massive halos compared to their star-forming counterparts.

3.5.5. Caveats and interpretations

In this section, we briefly address some caveats that may affect some of the results of our models. Regarding the $C_{\rm vir}$ model, there may be some fraction of halos that exhibit low values of $C_{\rm vir}$ because they have not yet relaxed after mergers. These merging halos are typically found in higher density environments and are expected to host quiescent galaxies. However, due to their temporarily low $C_{\rm vir}$, they could be erroneously assigned high sSFR values, which might contribute to the unexpected trends seen in Figures 3.10 and 3.11. As for the $\dot{M}_{\rm h}$ model, in some instances, high mass accretion rates may involve major mergers, particularly at high masses where the halo should actually host a quiescent central galaxy. In the model, however, such halos would be assigned high sSFR. This scenario may also contribute to the unexpected trends in Figures 3.10 and 3.11. While these caveats may impact the results quantitatively, we expect the qualitative aspects would remain unchanged.

Turning now to the $V_{\rm peak}$ model, one might initially assume that the sSFR segregation imposed in the $V_{\rm peak}$–M_* relation could be attributed, in part, to assembly bias related to $V_{\rm peak}$. This assumption arises from the close relationship between $V_{\rm peak}$ and $C_{\rm vir}$ (see, e.g., Klypin et al., 2016b), where higher values of $C_{\rm vir}$ will generally correspond to higher values of $V_{\rm peak}$ for a given halo mass. However, as shown in Figure 3.12, this may not necessarily hold true, as the segregation of the SHMR differs between the two models.

To investigate this further, we consider the formation redshift, $z_{\rm form}$, for halos hosting SFMS and Q+GV galaxies in our models, where $z_{\rm form}$ is defined as the redshift at which a dark matter halo reached 50% of its current mass. Figure 3.13 presents $z_{\rm form}$ as a function of halo mass (left panels) and stellar mass (right panels). The top row shows the results of the $V_{\rm peak}$ model, where halos hosting SFMS and Q+GV galaxies exhibit minimal difference in $z_{\rm form}$ as a function of halo or stellar mass. In contrast, both the $C_{\rm vir}$ and $\dot{M}_{\rm h}$ models demonstrate a strong correlation of sSFR with $z_{\rm form}$. This is consistent with the expected behavior of halo assembly bias, where early-forming halos tend to host quiescent galaxies. These results support our initial hypothesis that the $V_{\rm peak}$ model is significantly associated with halo mass.

We conclude that the empirical evidence obtained in this section *strongly* suggests that the $V_{\rm peak}$ model, in which the enhanced clustering of quiescent galaxies is attributed to the fact that they occupy more massive halos, offers a realistic model of how normal to massive central galaxies (i.e., $\log(M_*/M_\odot) > 10$) inhabit dark matter halos.

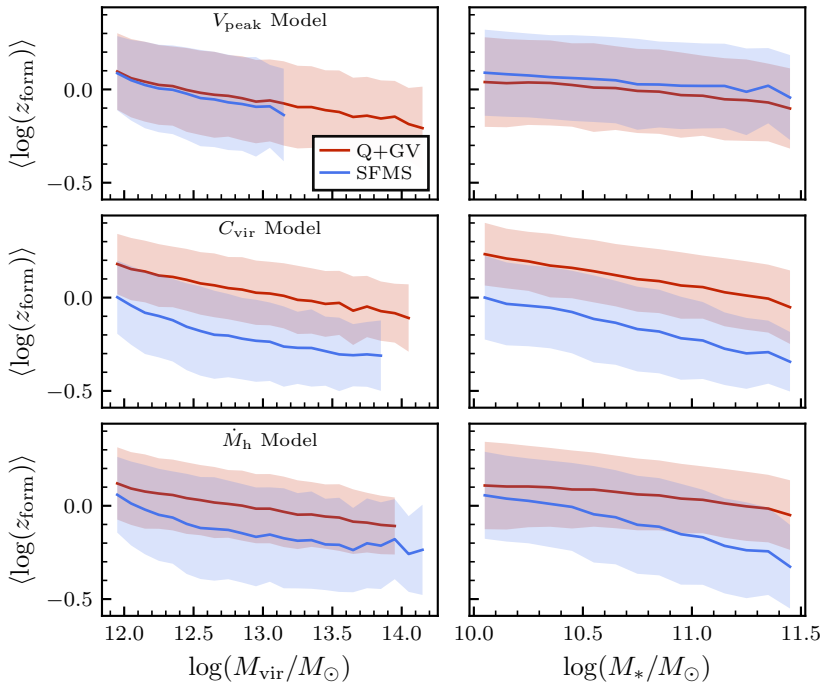

Figure 3.13. Formation redshift, $z_{\rm form}$, as a function of halo mass (left panels) and stellar mass (right panels) for three models of sSFR broken into quiescent and green valley (Q+GV) and SFMS centrals. Formation redshift is defined as the redshift at which a halo attained 50 per cent of its current mass. The $C_{\rm vir}$ and $\dot{M}_{\rm h}$ models each create a strong segregation in formation redshift between Q+GV and SFMS galaxies. This is consistent with the expected behavior of the assembly bias effect that is built into these models, where early-forming haloes host quiescent galaxies. In contrast, the $V_{\rm peak}$ model has no clear bias in formation redshift with sSFR, which supports our hypothesis that the $V_{\rm peak}$ model is strongly associated with halo mass.

§ 3.6. Discussion

In this paper, we have investigated how the clustering properties of galaxies vary across the sSFR–M_* plane. To do this, we defined five volume-limited sub-samples that are complete in stellar mass within their respective mass bins, as outlined in Section 3.2.3. Additionally, we confirmed that our sub-samples reproduce the SDSS GSMF (Dragomir et al., 2018), as shown in Figure 3.2.

For each of these five sub-samples, we analyzed the ACFs of all galaxies together (Figure 3.4) as well as individually for centrals (Figure 3.5) and satellites, all of which are summarized in Figure 3.6. Our analysis revealed that the ACFs of all galaxies strongly depend on sSFR at a given stellar mass, while the ACFs of central galaxies are largely independent of sSFR. We concluded that the presence of satellite galaxies and the fraction of satellites as a function of sSFR play a crucial role in these differences.

Next, we explored the cross-correlation of central galaxies with satellites across the sSFR–M_* plane (Figure 3.7). We demonstrated that there is a consistent trend in all mass bins of increased clustering amplitude in the one-halo term for lower ΔMS at a fixed stellar mass. The clustering amplitude of quiescent centrals was as much as $\sim$3 times greater than that of SFMS centrals. This observation aligns with the expectation that quiescent central galaxies tend to have more satellites than those on the SFMS. We supported this observation by showing a difference in the average number of satellites for SFMS and quiescent central galaxies in Figure 3.8. We found that the average number of satellites for quiescent centrals is about 2–3 times larger than that of SFMS centrals at higher masses.

In the following sections, we discuss the robustness of our results and comparisons

Table 3.4.: Percent agreement of central and satellite designations across all three group catalogs in our five sub-volumes.

M_* Limits [$\log(M_*/M_\odot)$]	Agreement
10.0 – 10.375	68.3%
10.375 – 10.75	72.8%
10.75 – 11.0	78.3%
11.0 – 11.25	84.4%
11.25 – 11.5	90.1%

with the literature.

3.6.1. Robustness of Results: Group Catalog Comparison

In this section, we will discuss comparisons of the group catalogs used and the results obtained from them. Since the results of this paper depend on the identification of central and satellite galaxies, it is important to consider the same analysis using other group catalogs. While it is not possible to directly check the accuracy of any given group catalog, checking for systematic trends in our results among different catalogs will test the robustness of our findings.

The simplest metric for comparing different group catalogs is to compare the central or satellite designation of a given galaxy across the catalogs. Note that not every galaxy in our sample has a designation in every group catalog. About 3% of galaxies have designations in only two catalogs, 3% have no designations, and fewer than 1% have designations in only one catalog. For comparison, we consider only galaxies that have a designation in all three catalogs. We find generally good agreement across the three group catalogs, ranging from ~70% to 90% with the agreement improving with mass. The values for each mass bin are shown in Table 3.4.

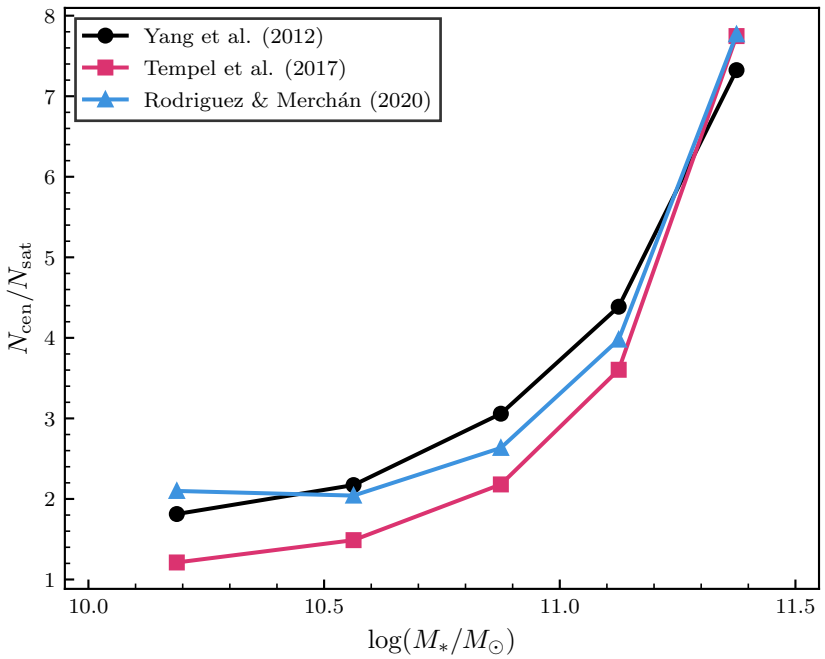

Figure 3.14. Ratios of central to satellite galaxies in each mass bin for each of the three group catalogs. T17 generally has the lowest ratio, i.e., the most galaxies designated as satellites. This is useful for testing central-only samples as one would expect T17 to have a more conservative estimate of centrals, leaving a sample with less contamination from satellites misidentified as centrals.

Another relevant statistic in the group catalogs is the relative numbers of central and satellite galaxies. Group catalogs that identify a higher ratio of centrals to satellites may have a higher rate of contamination where satellites are misidentified as centrals. On the other hand, group catalogs that identify a lower ratio of centrals to satellites may have a more pure sample of centrals. The ratios of centrals to satellites for the three group catalogs are shown in Figure 3.14. All three catalogs are fairly similar, though T17 has a clear lower ratio.

As a comparison to Figure 3.5, we show the same results of ACFs of centrals using T17 in Figure 3.15. We see the same lack of a distinct trend in the central ACFs that we

saw using the Y12 catalog. One notable difference, however, is the lack of a signal at small separations. If, indeed, T17 provides a more pure sample of only central galaxies, this would suggest the apparent signal at small separations for low-mass centrals in Figure 3.5 is most likely the result of satellites misidentified as centrals in the Y12 group catalog. We note also that this apparent signal occurs primarily in the quiescent sub-sample, and we know the satellite fraction is highest when considering quiescent galaxies. Thus, any systematic bias in the misidentification of satellites as centrals would manifest most strongly in the quiescent galaxy sub-sample. Additionally, we can compare to the ACFs of centrals using R&M20. From Figure 3.14, we know that R&M20 has a ratio of centrals to satellites that is comparable to Y12, with a slightly higher ratio in the lowest mass bin and a slightly lower ratio in the next three mass bins. In the results of the R&M20 central ACFs, we saw the strongest small separation signal of the three catalogs for low-mass quiescent (and green valley) centrals, as well as a signal in the next three mass bins similar to Y12. This result is consistent with our expectations of potential satellite contamination based on the ratios of centrals to satellites.

In Figure 3.16, we show the results of the CCFs for the three different group catalogs. Each row shows a simplified version of Figure 3.7 for each group catalog. Since we observe no trends in the CCFs across the SFMS in any of the group catalogs, we show only a single blue line per each grid cell corresponding to the CCF of all SFMS centrals with all satellites. As before, the green lines show the CCFs of GV centrals and the red lines show the CCFs of quiescent centrals. The values of the CCFs are generally consistent across the group catalogs, and, similar to Figure 3.7, we see the same trend of increasing clustering amplitude in the one-halo term with decreasing ΔMS present

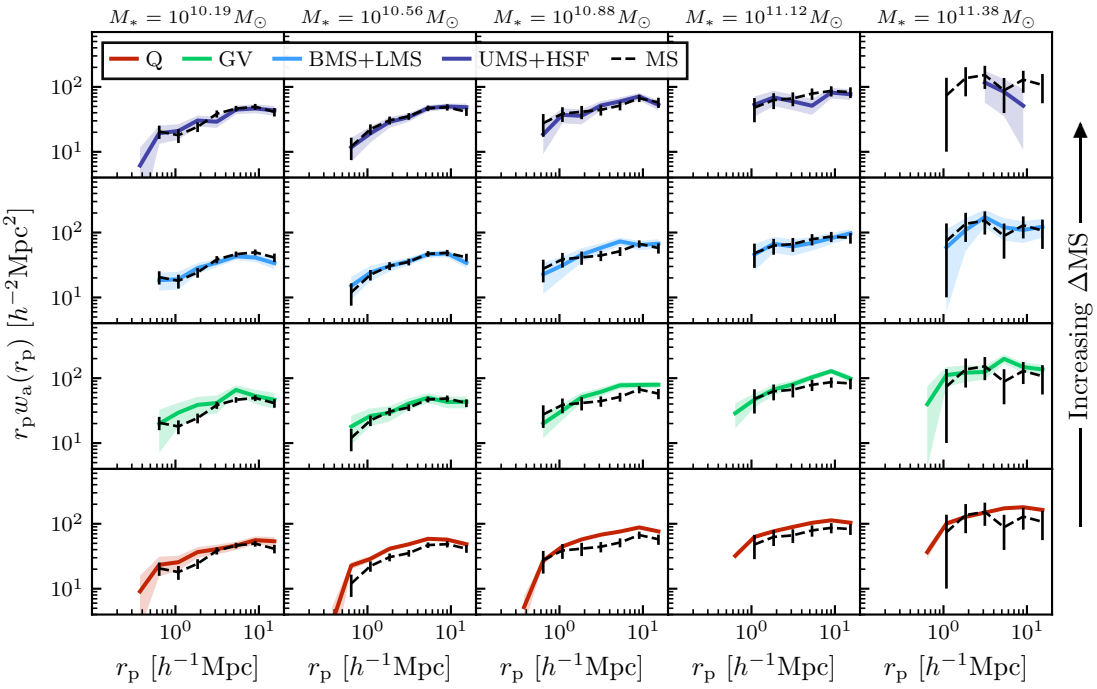

Figure 3.15. Projected auto-correlation functions (ACFs) of central galaxies identified by T17 as a function of ΔMS and M_*. The rows correspond from top to bottom to UMS+HSF galaxies (dark blue), BMS+LMS galaxies (light blue), GV galaxies (green), and quiescent galaxies (red). The colored lines show the results within a grid cell, and the black dashed lines show the ACFs of central galaxies in the center of the SFMS (LMS+UMS). Similar to Y12, there is no strong trend in the ACFs as a function of ΔMS, though there is a slight bias to higher clustering for the quiescent centrals. Unlike Y12, there is no signal in the quiescent centrals above the SFMS ACFs at small separations. See Section 3.6.1 for details on this.

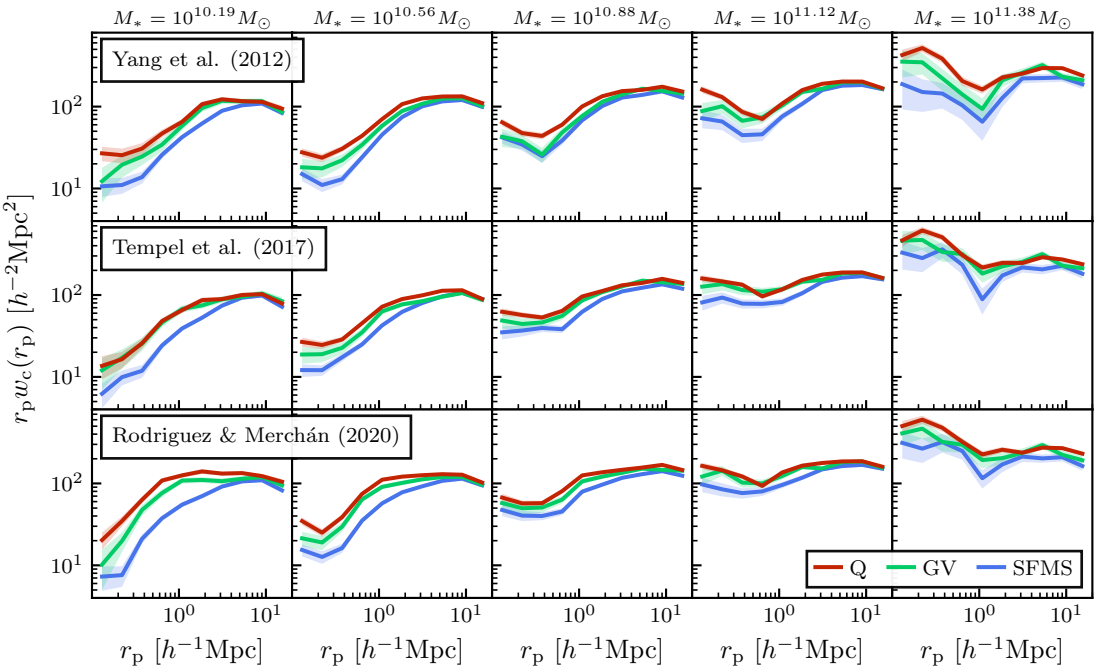

Figure 3.16. This shows a simplified version of Figure 3.7 where the grid has been collapsed vertically and each row now presents the results of a different group catalog. Shown are the cross-correlation functions of SFMS (blue), GV (green), and quiescent (red) centrals with all satellites in the same mass bin. This highlights the consistent trend across all group catalogs of increasing clustering at small separations as ΔMS decreases.

in both T17 and R&M20. These results reinforce our conclusion that quiescent central galaxies are more likely to have a greater number of satellites than central galaxies located on the SFMS.

3.6.2. Comparison with the Literature

Projected ACFs of galaxies as a function of sSFR and stellar mass have been measured by Coil et al. (2017, hereafter C17) and Berti et al. (2021b, hereafter B21). C17 study higher redshift galaxies from the PRIMUS and DEEP2 galaxy redshift surveys

spanning the redshift range $0.2 < z < 1.2$. B21 study lower redshift SDSS galaxies in the redshift range $0.02 < z < 0.04$ using the same stellar mass and SFR measurements we use in this paper. While different galaxy samples are studied in these two papers, they ultimately draw similar conclusions about the clustering trends with sSFR and stellar mass: namely, that clustering amplitude is as strong (or stronger) a function of sSFR as it is of stellar mass. However, this conclusion is based on ACFs of *all* galaxies without distinguishing between centrals and satellites. As we have shown in Figure 3.4, we confirm the presence of a strong sSFR-dependent clustering amplitude when considering centrals and satellites together. However, as we have shown in Figure 3.6, this trend is almost entirely produced by satellites which are predominantly quiescent galaxies. This possible explanation for the observed trend is mentioned in Section 6.3 of C17, and we confirm it in this paper for the stellar mass and sSFR ranges we study.

Another conclusion of C17 and B21 is the existence of intrasequence relative bias (ISRB), i.e., that within the star-forming (or quiescent) population, galaxies with higher sSFR have lower clustering amplitude than galaxies with lower sSFR. While we do not attempt to make any determination about this for quiescent galaxies, we can compare results of similar sub-samples for star-forming galaxies. In general, we do not observe ISRB within the SFMS, even when measuring clustering using centrals and satellites together. However, we do not believe our results are necessarily at odds with their conclusion. This effect may be more prominent at lower masses as C17 and B21 use samples that go well below our lower limit of $\log(M_*/M_\odot) = 10$. We attempt here to illustrate this but note that the wider mass bins used by C17 and B21 make it difficult to give this as the definitive explanation.

We can investigate this by comparing our results with Figure 1 of B21. In the rightmost panels (their "M_*/sSFR grid"), we can compare the results of the black and blue regions, which are in a fixed mass bin from $9.75 < \log(M_*/M_\odot) < 10.4$ with mean sSFRs of $\log(\mathrm{sSFR/yr}^{-1}) = -9.82$ and -10.33, respectively. These regions roughly correspond to star-forming galaxies above and below the center of the SFMS in our lowest mass bin: $10.0 < \log(M_*/M_\odot) < 10.375$ with mean sSFRs of $\log(\mathrm{sSFR/yr}^{-1}) \approx -9.9$ and -10.3, respectively. While these sub-samples do not directly correspond to each other, they are sufficiently similar for the sake of this comparison. We can see their clustering amplitudes are fairly similar, as we also observed in our results. There is a difference in clustering amplitude at larger separations for the higher and lower sSFR sub-samples, but the difference looks to be within the error bars. However, if we compare their purple and magenta regions, which are at a lower mass of $9.25 < \log(M_*/M_\odot) < 9.75$ but in the *same* sSFR bins as the black and blue regions, we see a much larger increase in clustering going from the purple (higher sSFR) to the magenta (lower sSFR) sub-sample. This suggests that the ISRB within the SFMS may indeed be much more prominent at lower stellar masses, which we do not probe in this paper.

A final consideration we give to this effect is the potential mass trend within a mass bin. In this paper, we have chosen narrower mass bins to attempt to limit the extent to which the mean stellar mass within a given mass bin can vary. Clustering amplitude will generally increase as stellar mass increases, so it is important to bear this in mind when attempting to isolate clustering trends with sSFR. We do not attempt to quantify the magnitude of the mass effect on clustering amplitude, but we simply note that the wider mass bins used by C17 and B21 are more susceptible to intra-bin mass trends. The aforementioned black and blue sub-samples used in B21 have a difference in mean

stellar mass of 0.08 dex while our comparable bins have a difference of only 0.016 dex. This difference may also play a role in explaining the differences between our results and those of B21. We do note, however, that the mean mass difference for the purple and magenta sub-samples is only 0.04 dex, which suggests that intra-bin mass trends in this case likely are not the cause of the ISRB observed in C17 and B21.

We note that Figure 12 of B21 plots the 3D ACFs of central and satellite galaxies separately predicted by the revised UniverseMachine, showing that the ACFs of both centrals and satellites are higher for quiescent and lower for star-forming galaxies. In initial tests, we found that the SFR-dependence of the clustering signal for central galaxies in the UniverseMachine is largely driven by backsplash halos that have lost at least 10% of their peak mass and that non-backsplash halos have nearly no SFR-dependence in their clustering. As backsplash galaxies are expected to be drawn from the same population as satellite galaxies (and hence have much higher quenched fractions than field galaxies), it would be difficult to remedy the clustering of the UniverseMachine by adjusting backsplash galaxies alone. This would suggest, along the same lines as in Cui et al. (2021), that non-backsplash centrals must actually have an inverse SFR-dependence in their clustering so that the clustering of the entire central sample has no SFR dependence (see also O'Donnell et al., 2021, 2022, who reached similar conclusions). We will follow up on this possibility in future work.

§ 3.7. Conclusions

In this paper, we analyzed the clustering properties of galaxies in the sSFR–M_* plane by calculating two-point auto-correlation functions (ACFs) of all, central, and satellite

galaxies and cross-correlation functions (CCFs) of centrals with satellites as a function of distance from the mean of the star-forming main sequence (SFMS) in various mass bins. These calculations were done using a spectroscopic sample of galaxies from the SDSS in the redshift range $0.02 < z < 0.2$ with stellar masses and star formation rates from the MPA-JHU catalog (Kauffmann et al., 2003; Brinchmann et al., 2004). To define central and satellite galaxies, we used three different group catalogs from Yang et al. (2012), Tempel et al. (2017), and Rodriguez and Merchán (2020). In each of these group catalogs, central galaxies were defined as the galaxy with the largest stellar mass in each group. For details on the catalogs and data selection, see Section 3.2. We draw the following conclusions from our analysis based on the Yang et al. (2012) group catalog for galaxies with stellar masses in the range $10.0 < \log(M_*/M_\odot) < 11.5$:

1. At a fixed stellar mass, ACFs of all (central and satellite) galaxies together depend strongly on sSFR (Figure 3.4). Green valley and quiescent galaxies have greater clustering amplitudes than SFMS galaxies, with quiescent clustering amplitudes increasing up to as much as an order of magnitude larger than that of SFMS galaxies.

2. When considering only central galaxies, ACFs show little-to-no dependence on sSFR at a fixed stellar mass (Figures 3.5 and 3.15). The ACFs do, however, increase with stellar mass. Since galaxy stellar mass increases monotonically with halo mass, and halo clustering also increases with halo mass (known as halo bias), we conclude that the increased galaxy clustering with stellar mass is driven by halo bias. However, halo clustering can also vary depending on secondary factors such as formation redshift and accretion history (known as assembly bias).

Because the observed central galaxy ACFs do not vary with sSFR at a fixed stellar mass, this suggests that assembly bias is not producing any significant net effect in central galaxy clustering in the sSFR–M_* plane.

3. By calculating the ACFs of centrals and satellites separately, we found that satellites are the main contributor to the strong sSFR dependence of all-galaxy ACFs (Figure 3.6). Since satellites are innately highly clustered and tend to be quiescent, they drive up the clustering amplitudes as lower sSFR samples are considered at a fixed stellar mass.

4. Cross-correlations of central galaxies as a function of sSFR with satellite galaxies of any sSFR show that more quiescent centrals have higher clustering amplitudes in the one-halo term than SFMS centrals in all stellar mass bins studied in this paper (Figure 3.7). CCFs of quiescent centrals are larger than those of SFMS centrals by up to a factor of ∼3 in the one-halo term. This suggests that quiescent centrals tend to host more satellites in their dark matter halos on average than SFMS centrals. We confirmed this conclusion by directly calculating $\langle N_{\rm sat} \rangle$ as a function of sSFR (Figure 3.8). Within the context of the halo occupation distribution, this implies that quiescent centrals reside in higher mass halos than SFMS centrals of the same stellar mass, which is consistent with observations.

All results were additionally checked using the group catalogs from Tempel et al. (2017) and Rodriguez and Merchán (2020) to test for robustness. We find that our conclusions are valid regardless of the choice in group catalog employed for clustering analysis (see Section 3.6.1 and Figure 3.16).

We used these conclusions to test whether different models of the galaxy–halo con-

nection could reproduce, at the same time, the sSFR–M_* plane and the ACFs, CCFs, and $\langle N_{\rm sat} \rangle$ as a function of distance from the mean SFMS. Additionally, we directly calculated the resultant stellar-to-halo mass relations (SHMRs) from these models for star-forming and quiescent galaxies. We assigned sSFR values to dark matter halos in the Bolshoi–Planck N-body simulation using three different models based on (1) halo accretion rate, $\dot{M}_{\rm h}$, (2) halo concentration, $C_{\rm vir}$, and (3) peak circular velocity, $V_{\rm peak}$. From comparing the models to the observations, we draw the following conclusions:

1. ACFs of central galaxies from the three different models of sSFR all show *similar* results that agree well with observations (Figure 3.9). It may be expected that the assembly bias built into the $\dot{M}_{\rm h}$ and $C_{\rm vir}$ models would lead to trends in the ACFs with sSFR at a fixed stellar mass, but no trends were observed. This may be the result of competing biases within the models or the assembly bias effect being lessened or removed when binning by *stellar* mass rather than *halo* mass.

2. Modeling sSFR based on $\dot{M}_{\rm h}$ or $C_{\rm vir}$ did not reproduce trends in the CCFs of observed SDSS galaxies (Figure 3.10). Instead, these models predicted a trend *opposite* to what is observed. We also saw an opposite trend when considering $\langle N_{\rm sat} \rangle$ for star-forming and quiescent galaxies (Figure 3.11). However, modeling sSFR based on $V_{\rm peak}$ produced CCFs that were consistent with observations and showed quiescent centrals with greater $\langle N_{\rm sat} \rangle$ than star-forming centrals at all stellar masses, also agreeing with SDSS observations. These results suggest that the cross-correlation clustering of centrals with satellites in the sSFR–M_* plane may be primarily driven by halo mass, as $V_{\rm peak}$ is strongly correlated with halo mass.

3. While ACFs showed similar results for all models, the capability of CCFs and $\langle N_{\rm sat} \rangle$ to differentiate models of sSFR based on halo properties has proven to be a powerful tool for constraining different models of the galaxy–halo connection.

4. Comparing the SHMRs of the different models showed that all models agree with observations of the SHMR of quiescent galaxies, but only the $V_{\rm peak}$ model agreed that star-forming galaxies reside in halos of lower mass than quiescent galaxies at a fixed stellar mass. The accretion rate and concentration models both predicted that star-forming galaxies reside in halos of *higher* mass at a fixed stellar mass, again, opposite of what the observations show.

5. In the $V_{\rm peak}$ model, halos hosting SFMS and quiescent galaxies show minimal difference in formation redshift. This suggests that the $V_{\rm peak}$ model is primarily associated with halo mass and rules out being influenced by assembly bias related to the relationship between $V_{\rm peak}$ and $C_{\rm vir}$.

In our $V_{\rm peak}$ model, the difference in clustering for quiescent and star-forming centrals is driven primarily by a segregation in the SHMR, wherein quiescent centrals reside in halos of higher mass than star-forming centrals at the same stellar mass. The evidence obtained from both observations and our models of sSFR suggests that the $V_{\rm peak}$ model offers a realistic model of how massive ($\log(M_*/M_\odot) \geq 10$) centrals inhabit dark matter halos.

Early-forming halos form in denser regions of the Universe, and they are therefore more clustered. It is tempting to imagine that this assembly bias of early-forming halos helps explain why galaxies with lower sSFR are more clustered. In more detail, the idea is that early-forming halos host early-forming galaxies, which also quench

early. But if this were true of central galaxies, the quenched centrals would be more correlated than SFMS centrals, which we have shown is not true. This implies that central galaxies in early-forming halos do not, in general, quench earlier than central galaxies in later-forming halos of the same mass.

Our finding that the central galaxy ACFs are unchanged as a function of sSFR at a fixed stellar mass leaves the door open for multiple possible explanations that require further testing. While the $V_{\rm peak}$ model does not include assembly bias, one could imagine a model that uses both a segregation in the SHMR and assembly bias to reproduce the observed ACFs. With a segregation in the SHMR such that quiescent galaxies reside in more massive halos than star-forming galaxies at a fixed stellar mass, it may be expected that quiescent centrals should be more clustered than star-forming centrals of the same stellar mass, as a result of halo bias. Since we do not observe this, it suggests, perhaps, that there could be an additional effect increasing the clustering amplitude of star-forming galaxies, leading to an overall lack of clustering trend with sSFR. Interestingly, this would be an example of assembly bias favoring higher clustering for *star-forming* galaxies rather than quiescent galaxies, as is often considered.

An astrophysical example of this is provided by the Simba large-volume cosmological hydrodynamic simulation (Davé et al., 2019; Cui et al., 2021), in which late-quenching galaxies live in early-formed halos and early-quenching galaxies live in later-formed halos. In Simba, this occurs because of a combination of astrophysical effects: central galaxies in early-forming halos accrete more cold gas and AGNs in such galaxies are less effective in quenching them, while late-forming halos have mainly hot-mode accretion and the associated jet-mode AGN feedback quenches them early. For example, Section 3 of the Supplementary Information for Cui et al. (2021) describes three types of halos of

the same mass at $z = 0$, in which the central galaxies in the early-forming halos quench last, and the central galaxies in the late-forming halos quench earliest. Consistent with this, Wang et al. (2023) note that both observations and the EAGLE simulation indicate that early-formed halos tend to have a higher ratio of stellar mass to halo mass, and their analysis of SDSS galaxies implies that late-formed halos tend to host quiescent galaxies. We leave to a future paper further study of such large-volume simulations, including the ACFs and CCFs of their central and satellite galaxies. Our preliminary finding is that the Simba central galaxy ACFs and the dependence of $N_{\rm sat}$ on the sSFR of the central are consistent with our analysis of the SDSS.

Appendix

A. SORT Performance in a Mock CANDELS Light Cone

In addition to the 2 square degree wide-field light cone, SORT was tested on a narrower 17×41 square arcmin light cone with 47,404 galaxies (~68 galaxies per square arcmin). All model parameters were kept the same, and we observed that SORT continues to perform well at improving redshift estimates and determining local galaxy environments. Here we provide parallel results of the main text for SORT applied to this light cone with minor adjustments detailed hereafter.

Figures A.1 and A.2 show the one- and two-dimensional redshift histograms comparing $z_{\rm sort}$ and $z_{\rm phot}$ to $z_{\rm spec}$. Figure A.1 is a direct parallel to Figure 2.5 and shows similar overall improvement in redshift errors. For Figure A.2, however, because statistics are more limited in the mock CANDELS light cone, all three redshift ranges of width $\Delta z = 0.5$ were stacked on top of each other to create composite histograms. This allows recovery of similar distributions to those shown in Figure 2.6. In particular, the photometric redshifts maintain their Gaussian error distributions while SORT redshifts build up along the line of equality.

As before, SORT was run on the same light cone with 10 different random seeds (which determine the selection of $z_{\rm ref}$) to find an average result for the 2PCF. The

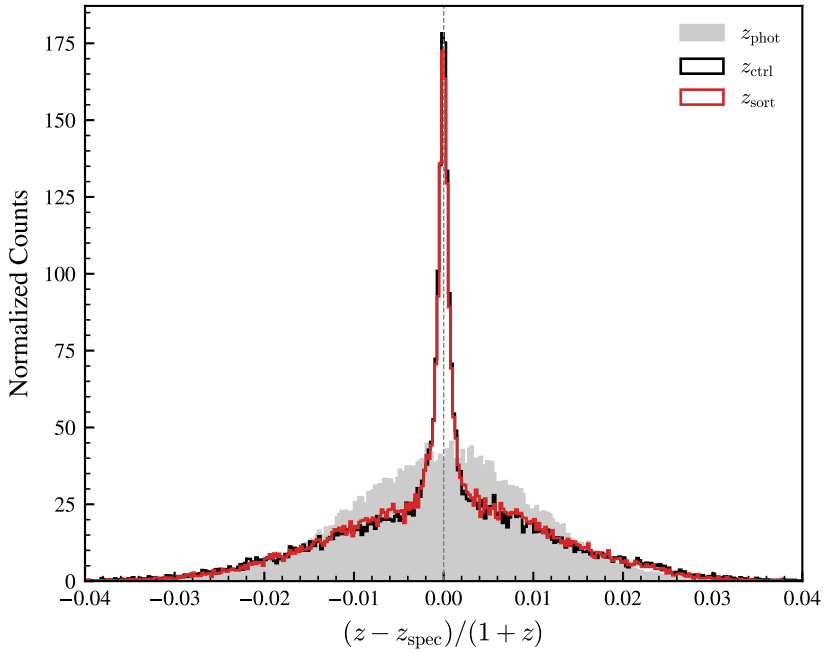

Figure A.1. Normalized distribution of Δz (excluding the spectroscopic sample) for $z_{\rm sort}$, $z_{\rm phot}$, and $z_{\rm ctrl}$ using the mock CANDELS light cone (see Section 2.5.1 for details on $z_{\rm ctrl}$). We recover a distribution of redshift errors similar to Figure 2.5 using the wide-field light cone. In particular, the $\Delta z_{\rm sort}$ distribution is dominated by a tall, central peak of improved redshifts.

average 2PCF estimates are shown in Figure A.3 for each of the redshift types in the range $0.75 < z < 1.25$. The error bars represent the standard deviations of $\xi(s)$ for each of the 10 seeds within each bin. Due to the narrower geometry of this light cone compared to the wide-field light cone, the 2PCFs were only calculated out to 8 h^{-1}Mpc. Similar to Figure 2.7, we see that SORT provides a much better estimate of the 2PCF than using photometric redshifts.

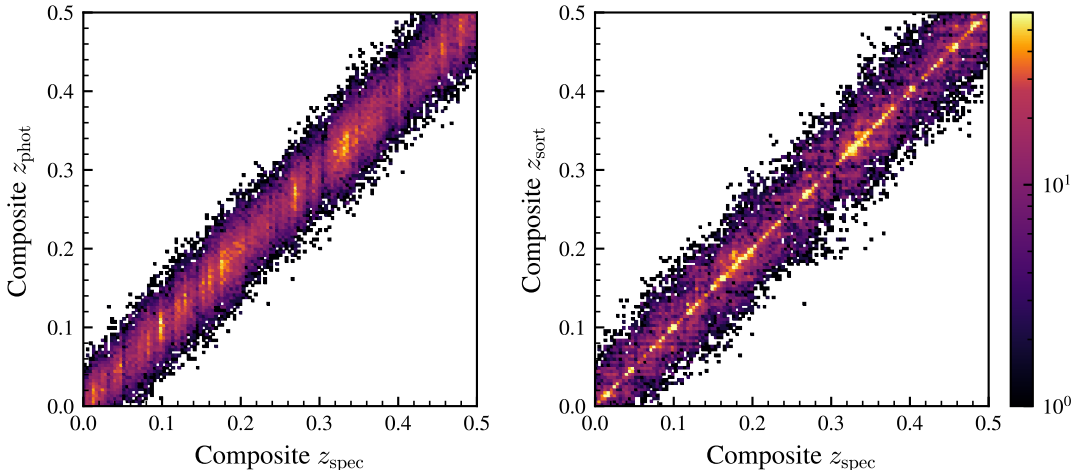

Figure A.2. Two-dimensional redshift histograms for $z_{\rm phot}$ and $z_{\rm sort}$ relative to $z_{\rm spec}$ with binning of 0.004 using the mock CANDELS light cone. The color bar shows the total number of counts in each bin. The data represent the full catalog of redshifts broken into the three complete redshift bins of size $\Delta z = 0.5$ that have been stacked on top of each other. In doing so, we are able to observe similar improvement in redshift estimates to the wide-field light cone after applying SORT.

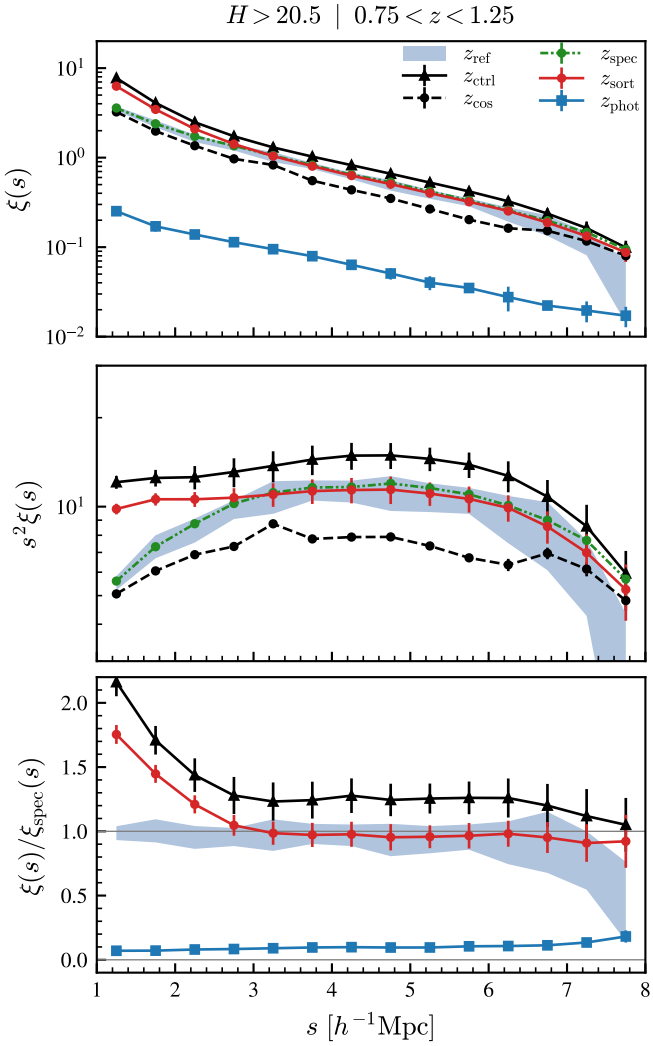

Figure A.3. Two-point correlation functions (2PCFs) of various redshift types as a function of redshift-space distance s shown in three different ways using the mock CANDELS light cone. In each panel, the values plotted represent the mean result of running SORT with 10 different random seeds to average out sample variance when selecting the reference galaxies. We continue to see that $\xi_{\rm phot}(s)$ is a poor estimate of the 2PCF and $z_{\rm ctrl}$ overestimates the 2PCF while $\xi_{\rm sort}(s)$ is accurate (relative to $\xi_{\rm spec}(s)$) for $s \gtrsim 2.5\ h^{-1}$Mpc. See Section 2.5.1 for details on $z_{\rm ctrl}$ (shown as black triangles.

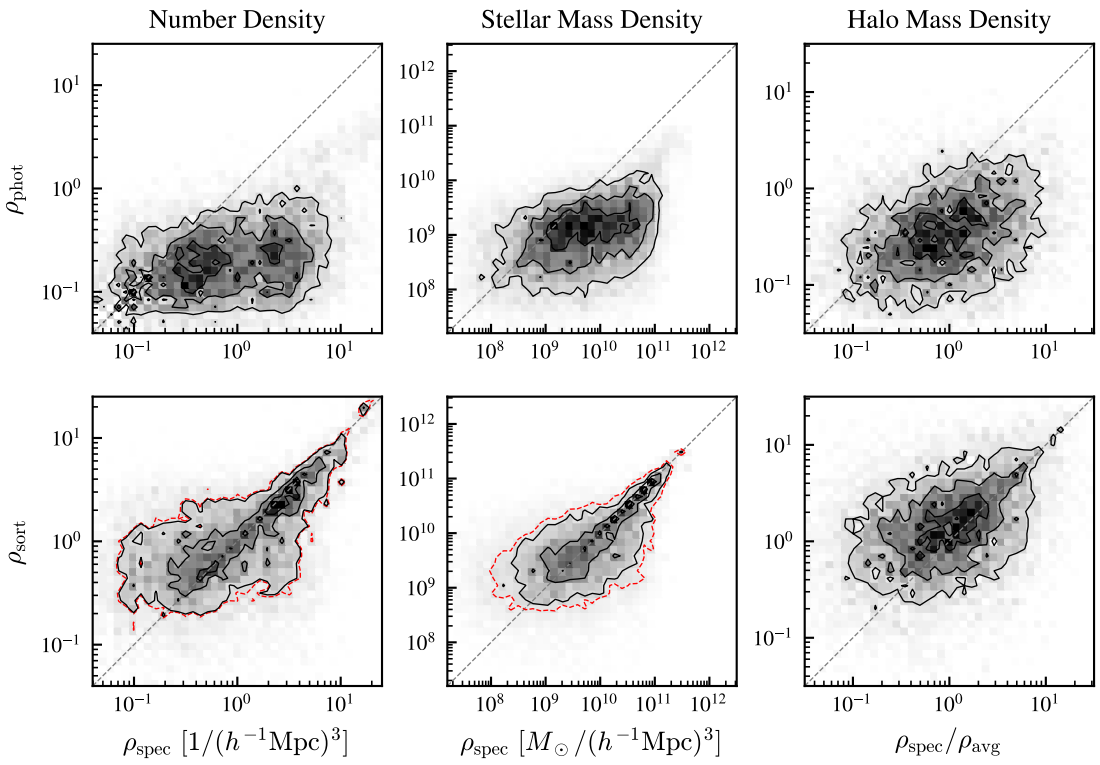

Figure A.4. Two-dimensional density histograms for $\rho_{\rm phot}$ and $\rho_{\rm sort}$ in the range $0.75 < z < 1.25$ using the mock CANDELS light cone. The left panels show number densities, the middle panels show stellar mass densities, and the right panels show halo mass densities using only central galaxies. The solid contours represent limits of 25, 50, and 75 percent of the maximum bin value in each subplot. The dashed contour (red) is set at a limit equal to the minimum contour level in the corresponding $\rho_{\rm phot}$ subplot. Densities were calculated with a fixed cylinder length of 4 h^{-1}Mpc and a radius check this starting at ~0.5 h^{-1}Mpc and expanding up to ~2 h^{-1}Mpc as needed to encompass at least five galaxies. As with the wide-field light cone, we continue to see improvement in density estimates by SORT compared to photometric estimates. See Section 2.4.3 for details on the central halo mass densities.

B. Assigning Three-Dimensional Coordinates to Satellite Galaxies

The mock galaxy surveys used for this paper do not provide full three-dimensional coordinates for satellite galaxies. Instead, all satellite galaxies are assigned the same redshift as the dark matter halo they occupy. We used the following procedure to assign new three-dimensional coordinates and velocities to each of the satellites.

§ B.1. Calculating the Position

New positions were calculated for each satellite assuming that the satellites have the same radial distribution as the dark matter (see, e.g., Berlind and Weinberg, 2002b; Cooray and Sheth, 2002).* The radial density profile was approximated using the NFW formula (Navarro et al., 1996, 1997)

$$\rho_{\rm NFW}(r) = \frac{4\rho_s}{(r/R_s)(1+r/R_s)^2}. \tag{B.1.1}$$

*Improved treatments based on observations are discussed in Watson et al. (2012); Wechsler and Tinker (2018a); Lange et al. (2019); van den Bosch et al. (2019).

It is determined by two parameters, in this case ρ_s and R_s. Alternatively, it can be determined by the halo mass, $M_{\rm vir}$, and the halo concentration, $c_{\rm vir}$, which is defined as

$$c_{\rm vir} = \frac{R_{\rm vir}}{R_s}. \tag{B.1.2}$$

The scale radius, R_s, is the radius at which the log-space derivative of $\rho_{\rm NFW}(r)$ is -2. This could be found by fitting the NFW profile to each halo in the simulation. However, a more robust method is to find the Klypin scale radius using the $M_{\rm vir}$–$V_{\rm max}$ relation under the assumption of an NFW profile (Klypin et al., 2011). The parameter $V_{\rm max}$ is the maximum circular velocity of the halo (i.e. the maximum of $\sqrt{GM(r)/r}$, where $M(r)$ is the mass enclosed within a radial distance r). For an NFW profile, the maximum circular velocity occurs at $R_{\rm max} = 2.1626 R_s$ (Klypin et al., 2011; Behroozi et al., 2013a). With this, we calculated the Klypin concentration, $c_{\rm vir,K}$, by numerically solving

$$\frac{c_{\rm vir,K}}{f(c_{\rm vir,K})} = V_{\rm max}^2 \frac{R_{\rm vir}}{GM_{\rm vir}} \frac{2.1626}{f(2.1626)} \tag{B.1.3}$$

where

$$f(x) = \ln(1+x) - x/(1+x). \tag{B.1.4}$$

The radial distribution for an NFW profile can also be written in terms of the halo's mass as

$$M_h(r) = M_{\rm vir} \times u_{\rm vir}(r) \tag{B.1.5}$$

where $u_{\rm vir}(r)$ is

$$u_{\rm vir}(r) = \frac{\ln(1 + c_{\rm vir,K} x) - c_{\rm vir,K} x/(1 + c_{\rm vir,K} x)}{\ln(1 + c_{\rm vir,K}) - c_{\rm vir,K}/(1 + c_{\rm vir,K})} \tag{B.1.6}$$

with $x = r/R_{\rm vir}$. We can use this to sample new radial positions for the satellites within a halo. For each satellite in a given halo, the following procedure was followed.

1. Generate three random numbers U_r, U_θ, and U_ϕ, each uniformly distributed between 0 and 1.

2. Sample a radius from the distribution $u_{\rm vir}(r)$. This can be done by finding the value r such that $U_r - u_{\rm vir}(r) = 0$.

3. Assign the new spherical coordinates (r, θ, ϕ) to the satellite relative to the halo's center where $\theta = \pi U_\theta$ and $\phi = 2\pi U_\phi$.

4. Assign new Cartesian coordinates $\mathbf{r} = (x, y, z)$ relative to the halo using

$$x = r \sin\theta \cos\phi \tag{B.1.7}$$

$$y = r \sin\theta \sin\phi \tag{B.1.8}$$

$$z = r \cos\theta. \tag{B.1.9}$$

5. Get the position of the satellite relative to the box of the simulation using $\mathbf{R} = \mathbf{R}_{\rm h} + \mathbf{r}$, where $\mathbf{R}_{\rm h}$ is the position of the halo relative to the simulation box. A cosmological redshift, $z_{\rm cos}$, can be inferred from the new satellite position.

§ B.2. Calculating the Velocity

To find the line-of-sight redshift, $z_{\rm los}$, of each satellite, we must account for the effects of the peculiar velocity along the line of sight. The peculiar velocity of each satellite will depend on its position within a halo. Using the new radial distribution of satellites,

as well as properties of the halos within which the satellites reside, we can estimate new satellite velocities.

1. By assuming that the satellite velocities trace the dark matter particle velocities within an NFW halo, we can calculate the velocity dispersion of the satellites at a distance r from the halo's center using

$$\sigma^2(r) = \frac{c_{\rm vir,K} V_{\rm vir}^2}{\mu(c_{\rm vir,K})} \frac{r}{R_s} \left(1 + \frac{r}{R_s}\right)^2 \int_{r/R_s}^{\infty} \frac{\mu(x)\mathrm{d}x}{x^3(1+x)^2} \tag{B.2.1}$$

where $\mu(x)$ is defined as

$$\mu(x) = \ln(1 + c_{\rm vir,K} x) - c_{\rm vir,K} x/(1 + c_{\rm vir,K} x). \tag{B.2.2}$$

2. Sample a velocity v from the Gaussian distribution

$$P(v) = \frac{1}{\sqrt{2\pi\sigma^2(r)}} \exp\left(-\frac{v^2}{2\sigma^2(r)}\right). \tag{B.2.3}$$

3. Generate two random numbers U_θ and U_ϕ, each uniformly distributed between 0 and 1.

4. Using $\theta = \pi U_\theta$ and $\phi = 2\pi U_\phi$, the components of the satellite's velocity vector $\mathbf{v} = (v_x, v_y, v_z)$ relative to the halo's center are

$$v_x = v \sin\theta \cos\phi \tag{B.2.4}$$

$$v_y = v \sin\theta \sin\phi \tag{B.2.5}$$

$$v_z = v \cos\theta. \tag{B.2.6}$$

5. With respect to the box of the simulation, the satellite's velocity is $\mathbf{V} = \mathbf{V}_{\rm h} + \mathbf{v}$, where $\mathbf{V}_h$ is the halo's velocity with respect to the simulation box.

The component of the velocity along the line of sight can be found by the new position and velocity vectors:

$$v_{\rm los} = \mathbf{V} \cdot \hat{\mathbf{R}} \tag{B.2.7}$$

where $\hat{\mathbf{R}}$ is the unit vector pointing to the satellite's position. The final redshift can be calculated using

$$z_{\rm los} = z_{\rm cos} + \frac{v_{\rm los}}{c}(1 + z_{\rm cos}), \tag{B.2.8}$$

where c is the speed of light.

C. SORT Performance with Larger Photometric Uncertainties

Our fiducial photometric uncertainty is somewhat optimistic at $\sigma_z^{\rm ph}/(1+z) = 0.01$, though not entirely unrealistic as future redshift estimates are expected to have photometric uncertainties of $\sigma_z^{\rm ph}/(1+z) \approx 0.02$ or better. Nevertheless, here we present brief results of SORT for larger photometric uncertainties.

We reiterate that stochastic ordering holds true for Gaussian PDFs with arbitrarily-large standard deviations. We should therefore expect to see similar redshift improvement when increasing the photometric uncertainty. The results for Δz are shown in Figure C.1 with the fiducial results in black and the results with higher photometric uncertainties in red and blue. The histograms have been normalized by the photometric uncertainty to show the *relative* performance of SORT as $\sigma_z^{\rm ph}$ is increased. To deal with biases from sample variance in $z_{\rm ref}$, the histograms show the collection of SORT results using 10 different random seeds. We observe that SORT's improvement of redshifts with respect to a given photometric uncertainty remains largely unchanged as $\sigma_z^{\rm ph}$ increases. In all three cases we see the same general features: (i) a similar overall standard deviation in $\Delta z_{\rm sort}$ and $\Delta z_{\rm phot}$, (ii) a modest increase in scatter at the tail ends of $\Delta z_{\rm sort}$ compared to $\Delta z_{\rm phot}$, and (iii) a tall central peak of improved redshifts.

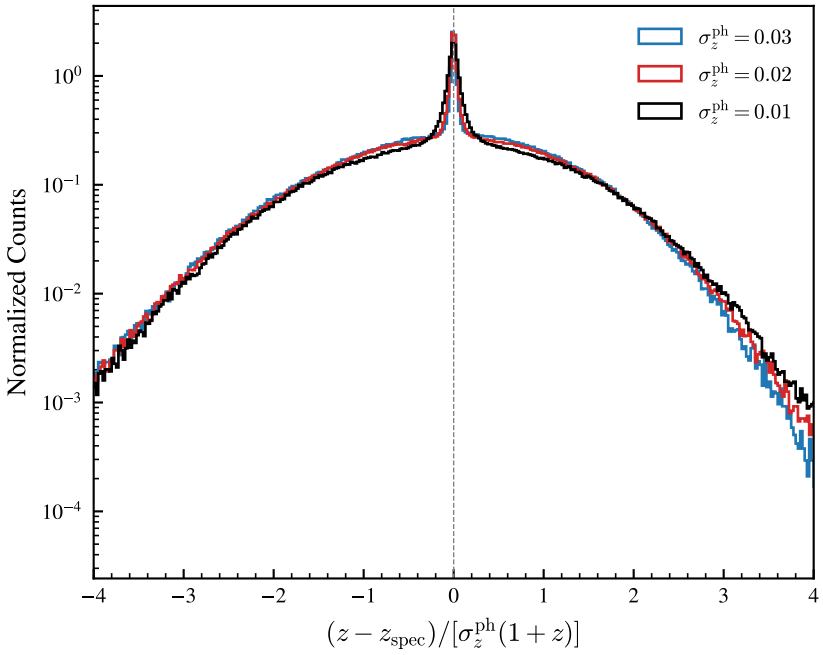

Figure C.1. Normalized distributions of Δz (excluding the spectroscopic sample) for $z_{\rm sort}$ and $z_{\rm phot}$ using three different photometric uncertainties. The distributions are normalized by their respective photometric uncertainties. We observe that the *relative* improvement of redshifts by SORT with respect to a given $\sigma_z^{\rm ph}$ is generally independent of $\sigma_z^{\rm ph}$. There are some dissimilarities, however. In particular, asymmetry in the tails of the distributions grows with $\sigma_z^{\rm ph}$, and the width of the peak decreases with increasing $\sigma_z^{\rm ph}$.

The two most notable differences are an increase in asymmetry in the tails of the histograms and a decrease in peak width as $\sigma_z^{\rm ph}$ increases. The asymmetry of the tails will be mostly irrelevant to the net result of SORT as the counts are around two orders of magnitude lower than the peak which dominates the distribution. Though we do not test this here, the width of the peak is likely more relevant to the final results of SORT. However, even with a photometric uncertainty of $\sigma_z^{\rm ph}/(1+z) = 0.02$, SORT is still able to fairly well recover the 2PCF at similar scales of $s \gtrsim 2.5\ h^{-1}$Mpc (shown in the bottom panel of Figure 2.7).

D. Additional Figures

Here we provide additional supplementary figures that support the main text. Figure D.1 shows two-dimensional histograms of the redshift errors as a function of the chosen redshift – either $z_{\rm phot}$ or $z_{\rm sort}$. Figure D.2 shows redshift errors for SORT with varying spectroscopic fractions. Figure D.3 shows results of the 2PCF in each of the three complete redshift bins. Figure D.4 shows another square region of space (similar to Figure 2.3) using different redshift types. Figure D.5 shows two-dimensional redshift histograms for each of the three complete redshift bins. Figure D.6 shows two-dimensional histograms that correspond to the errors of densities shown in Figure 2.8. Figure D.7 shows stellar mass densities in each of the three complete redshift bins using cylinders of length 4 h^{-1}Mpc. Figure D.8 shows densities similar to Figure 2.8 but calculated at a larger length scale of $l = 2\frac{1000\text{km s}^{-1}}{c}(1+z)$. Figure D.9 show two-dimensional histograms of 3D distances to kth nearest neighbors for k = 3, 5, and 7.

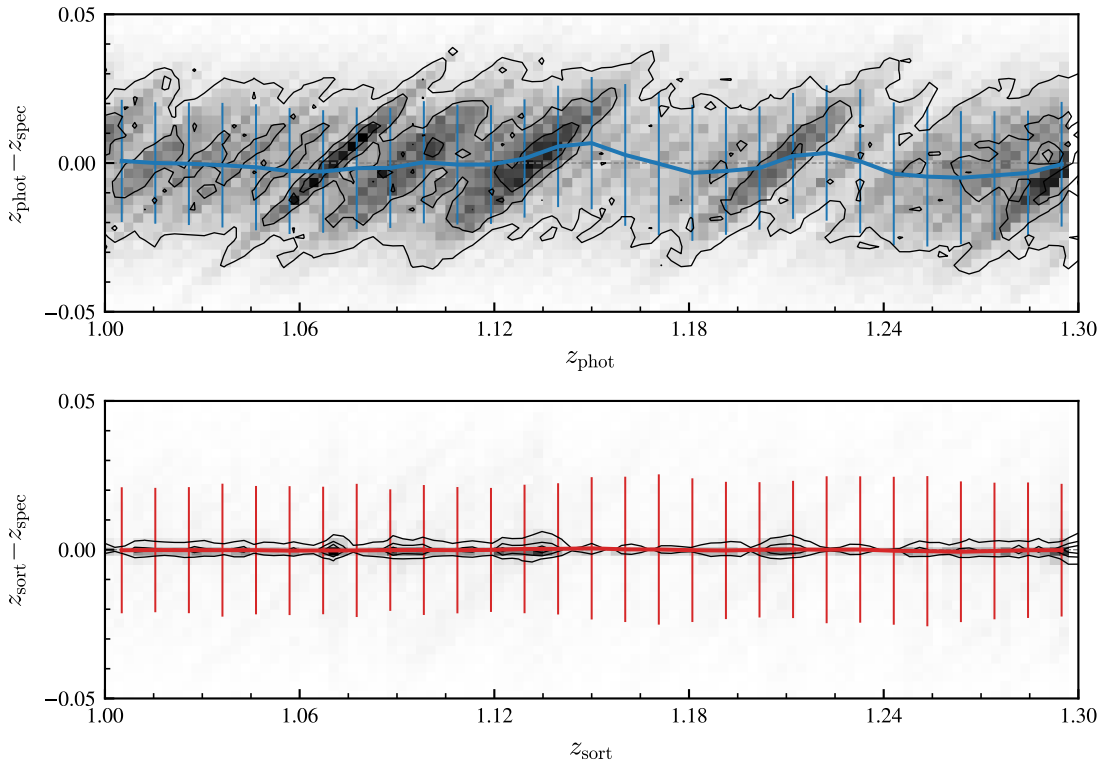

Figure D.1. Normalized two-dimensional histograms for errors in z_{phot} and z_{sort} relative to z_{spec}. The contours show the limits where counts are at least 25, 50, or 75 percent of the maximum value in each of the two subplots. While only the range $z = 1$–1.3 is shown, the results are representative of the entire light cone. There is a clear bias in the error of z_{phot} in regions of higher density. This bias is shown as a blue line which designates the median value of all redshifts within a series of bins along with 1σ error bars. As with Figure 2.5, both Δz_{phot} and Δz_{sort} have similar standard deviations. After applying SORT, though, the error bias is almost completely removed for the entire redshift range of the light cone.

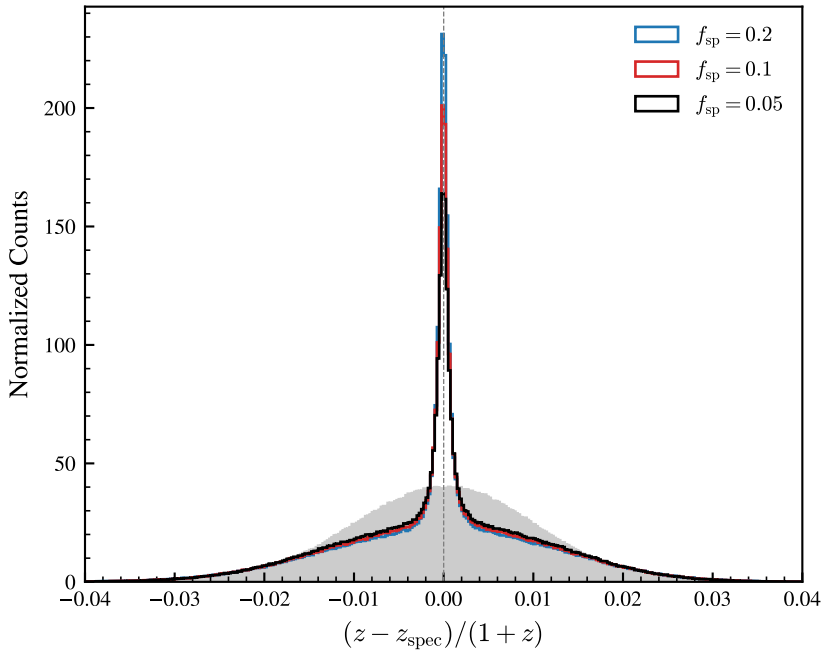

Figure D.2. Normalized distribution of Δz (excluding the spectroscopic sample) for $z_{\rm phot}$ and $z_{\rm sort}$ using three different spectroscopic fractions. As the spectroscopic fraction increases, SORT produces a taller peak surrounding $\Delta z = 0$. Even with a spectroscopic fraction as low as 5%, SORT still improves redshift estimates for a significant fraction of galaxies. The efficiency of SORT is rooted in the fact that most galaxies will tend to occupy a relatively small volume. Therefore it only takes a relatively small fraction of galaxies to reasonably trace the underlying distribution.

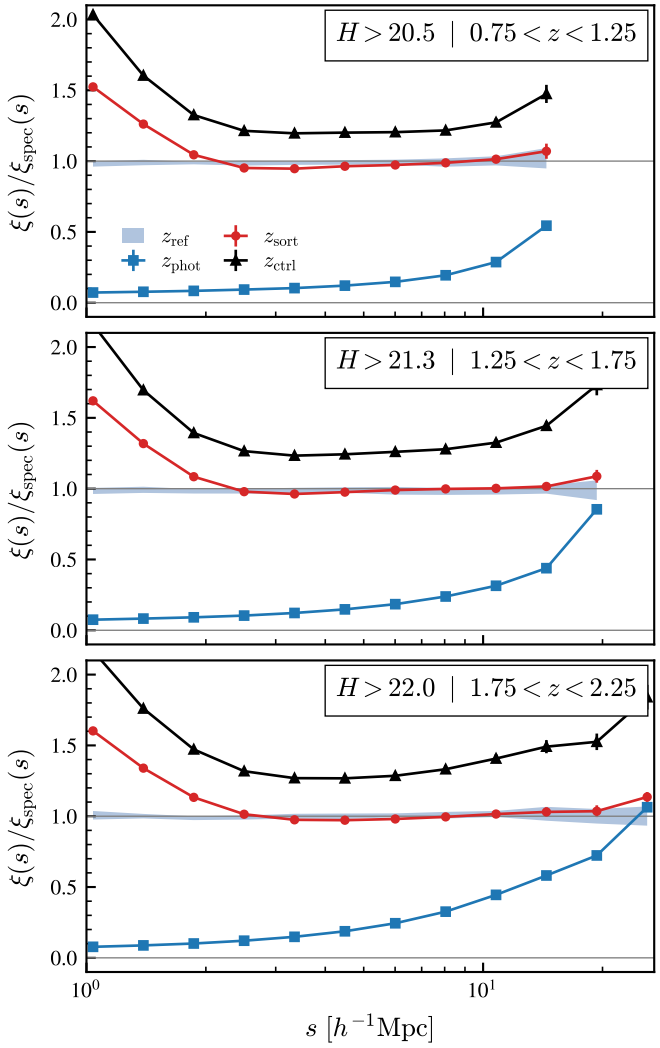

Figure D.3. Two-point correlation function (2PCF) ratios using $z_{\rm ref}$, $z_{\rm phot}$, $z_{\rm sort}$, and $z_{\rm ctrl}$ with respect to $z_{\rm spec}$ as a function of redshift-space distance s in three complete redshift bins. The 2PCFs were calculated out to distances of ~18–30 h^{-1}Mpc, limited by the sizes of each redshift bin. The results show the mean value of the 2PCFs along with 1σ error bars after running SORT with 10 different random seeds to determine the reference sample selection. Note that the error bars are too small to be seen. The 2PCF estimates provided by $\xi_{\rm sort}(s)$ show significant improvement over $\xi_{\rm phot}(s)$ and accurately recover $\xi_{\rm spec}(s)$ at scales of $s \gtrsim 2.5$ h^{-1}Mpc. We also observe the continued trend of $\xi_{\rm ctrl}(s)$ overestimating the 2PCF at all scales relative to $\xi_{\rm sort}(s)$ (see Section 2.5.1 for details).

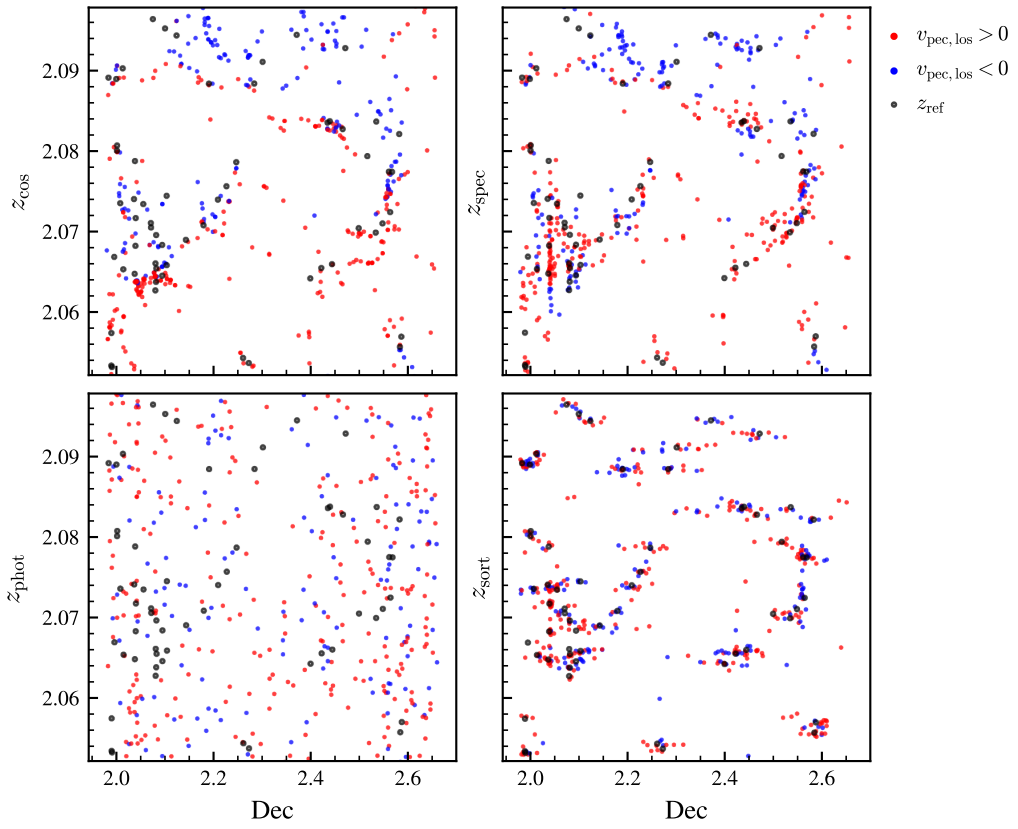

Figure D.4. Right ascension slices (thickness 0.1°) of galaxy distributions using different redshifts in a roughly $44 \times 44\ h^{-1}$Mpc region of space. The red and blue colouring denotes the direction of the peculiar velocity along the line of sight (red is positive and blue is negative). The black rings with empty centers are reference galaxies. Using the outline of the reference galaxies, SORT is able to recover the distinctive features in this region – in particular, the large filamentary structure across the top and right side of the panels, as well as the more dense group of galaxies in the lower left. We also note the presence of a characteristic feature found in SORT galaxy distributions – namely, horizontal rows of galaxies where there are few reference galaxies. In these areas, the radii of the sub-volumes within which SORT searches must expand to find reference galaxies. Galaxies are then pulled along the line of sight to an incorrect redshift, creating elongated features in a plane perpendicular to the line of sight. See Section 2.5.2 for details.

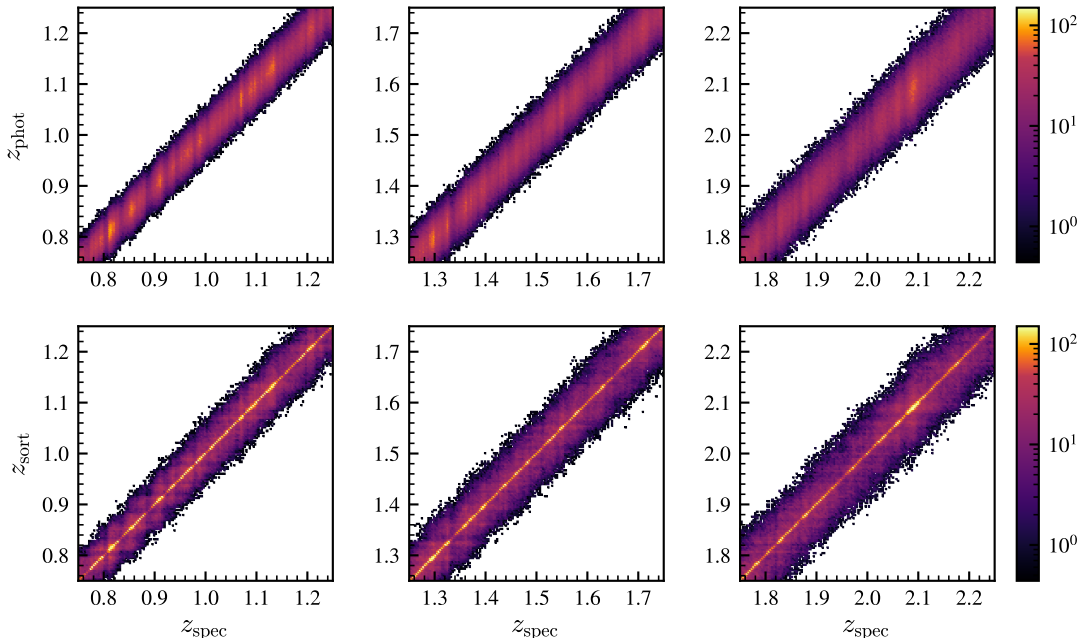

Figure D.5. Normalized two-dimensional redshift histograms for z_{phot} and z_{sort} compared to z_{spec} in all redshift bins. The z_{sort} distributions show significant improvement as counts build up along the line of equality while overall scatter for larger redshift errors is only modestly increased. This effect is consistent across all redshift ranges.

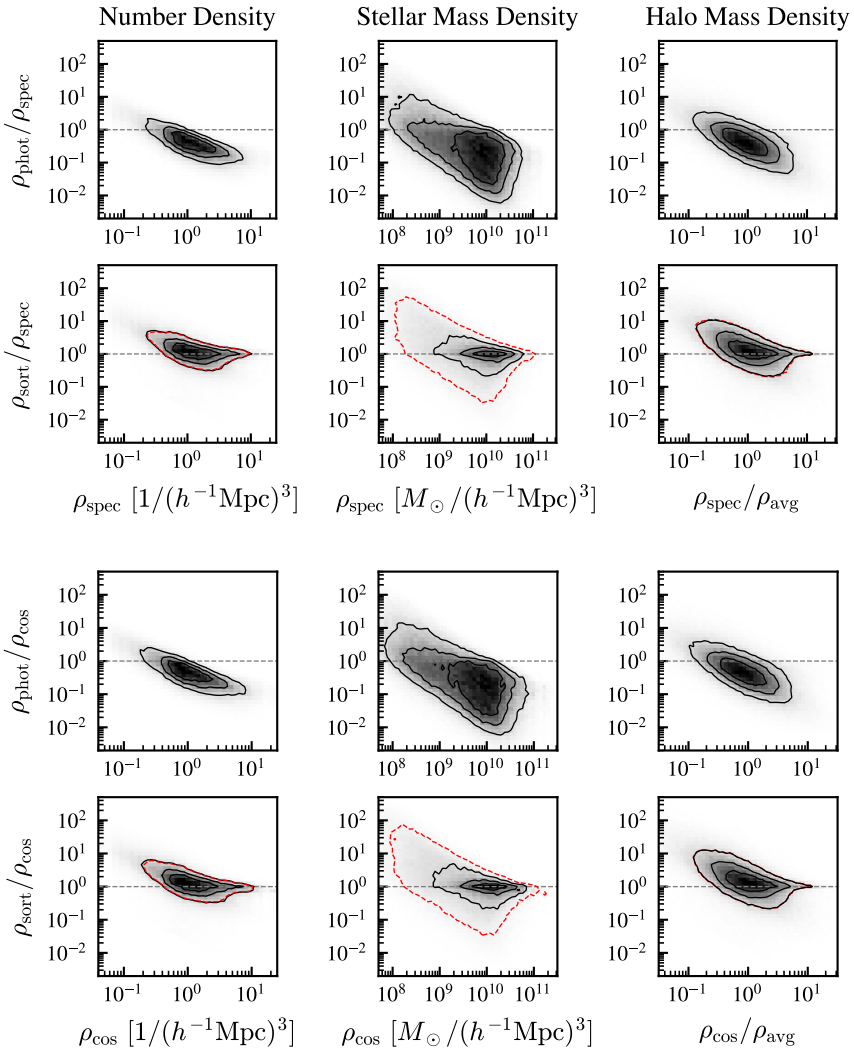

Figure D.6. Two-dimensional density error histograms for $\rho_{\rm phot}$ and $\rho_{\rm sort}$. Densities were calculated in cylinders of length 4 h^{-1}Mpc. The left panels show number density, the middle panels show stellar mass densities, and the right panels show halo mass densities using only central galaxies. The solid contours represent limits of 25, 50, and 75 percent of the maximum bin value in each subplot. The dashed contour (red) is set at a limit equal to the minimum contour level in the corresponding $\rho_{\rm phot}$ subplot. The horizontal dashed line represents zero error. While SORT struggles with lower densities, we observe much improvement from the highest densities down to average densities. SORT distributions show better alignment with the zero error line while photometric densities all tend to be underestimated except in the lowest-density environments.

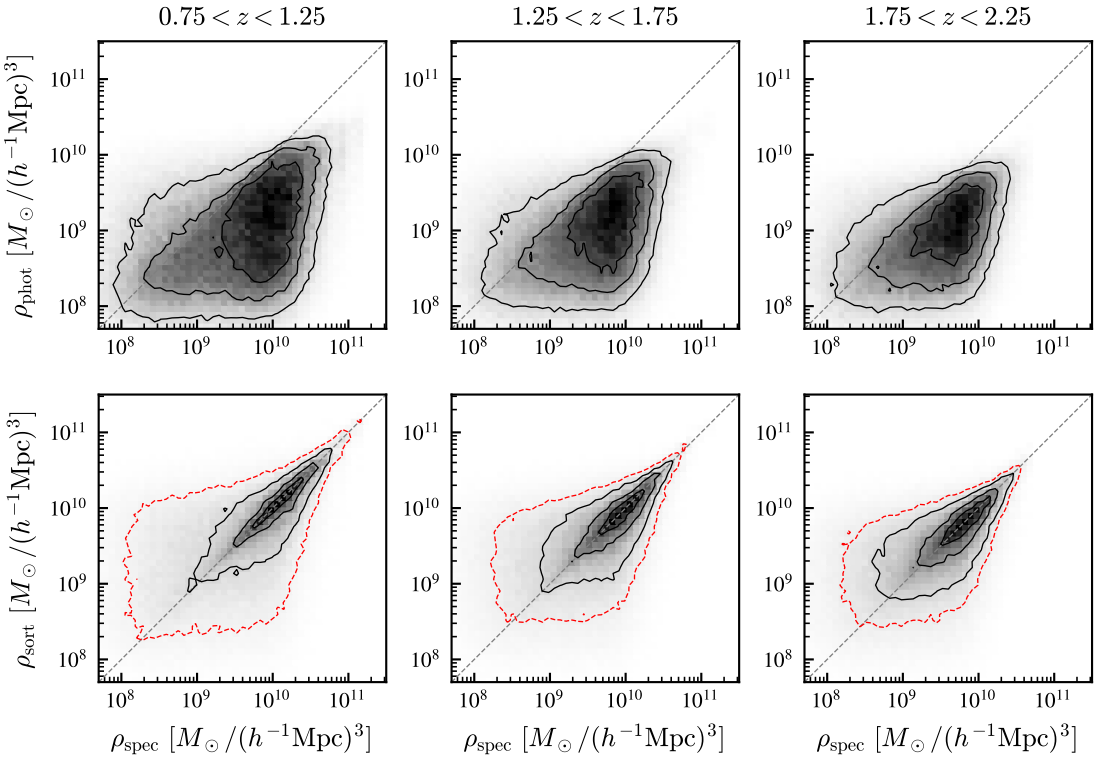

Figure D.7. Two-dimensional stellar mass density histograms for $\rho_{\rm phot}$ (top panels) and $\rho_{\rm sort}$ (bottom panels) for all redshift ranges of the light cone using a cylinder length of 4 h^{-1}Mpc. The solid contours represent limits of 25, 50, and 75 percent of the maximum bin value in each subplot. The dashed contour (red) is set at a limit equal to the minimum contour level in the corresponding $\rho_{\rm phot}$ subplot. We observe consistent improvement in density estimates with SORT at all redshifts. Biases in regions of average or higher density are greatly reduced. SORT distributions are more symmetric across the line of equality and overall scatter is lower.

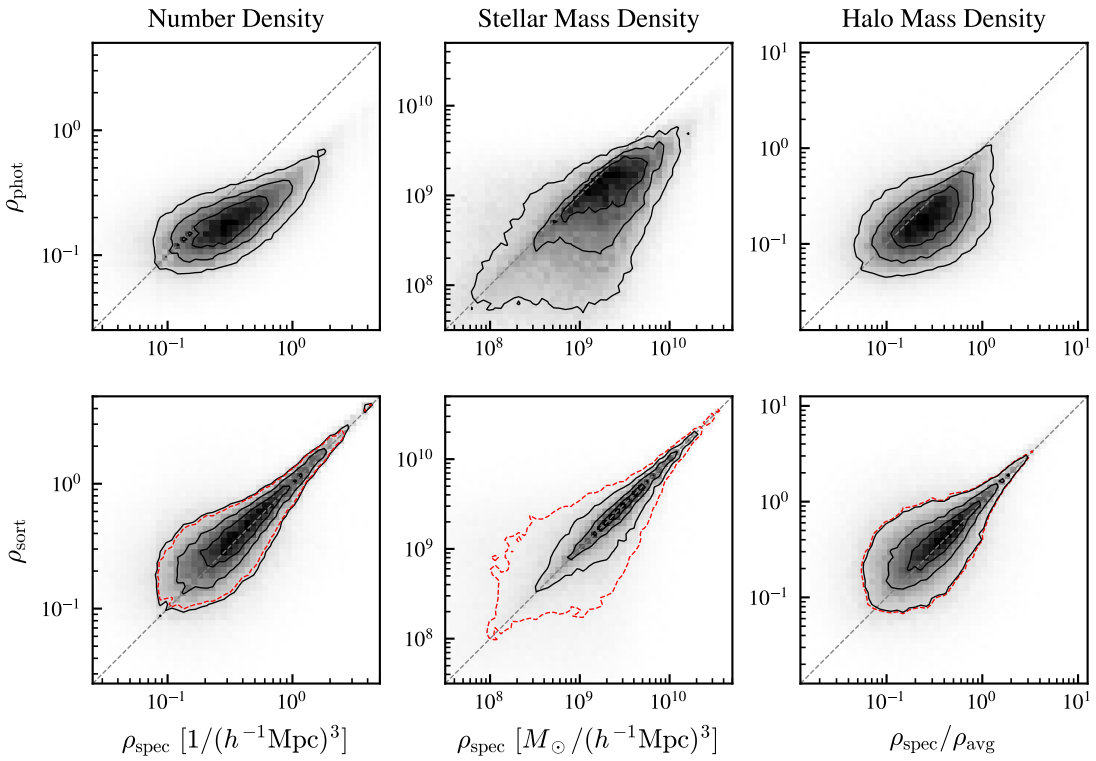

Figure D.8. Two-dimensional density histograms for $\rho_{\rm phot}$ (top panels) and $\rho_{\rm sort}$ (bottom panels) in the range $0.75 < z < 1.25$. Densities were calculated within cylinders of length $l = 2\frac{1000\ {\rm km\ s^{-1}}}{c}(1+z)$. The left panels show number density, the middle panels show stellar mass densities, and the right panels show halo mass densities using only central galaxies. The solid contours represent limits of 25, 50, and 75 percent of the maximum bin value in each subplot. The dashed contour (red) is set at a limit equal to the minimum contour level in the corresponding $\rho_{\rm phot}$ subplot. The longer length of the cylinder significantly improves SORT density estimates, most notably for stellar mass densities.

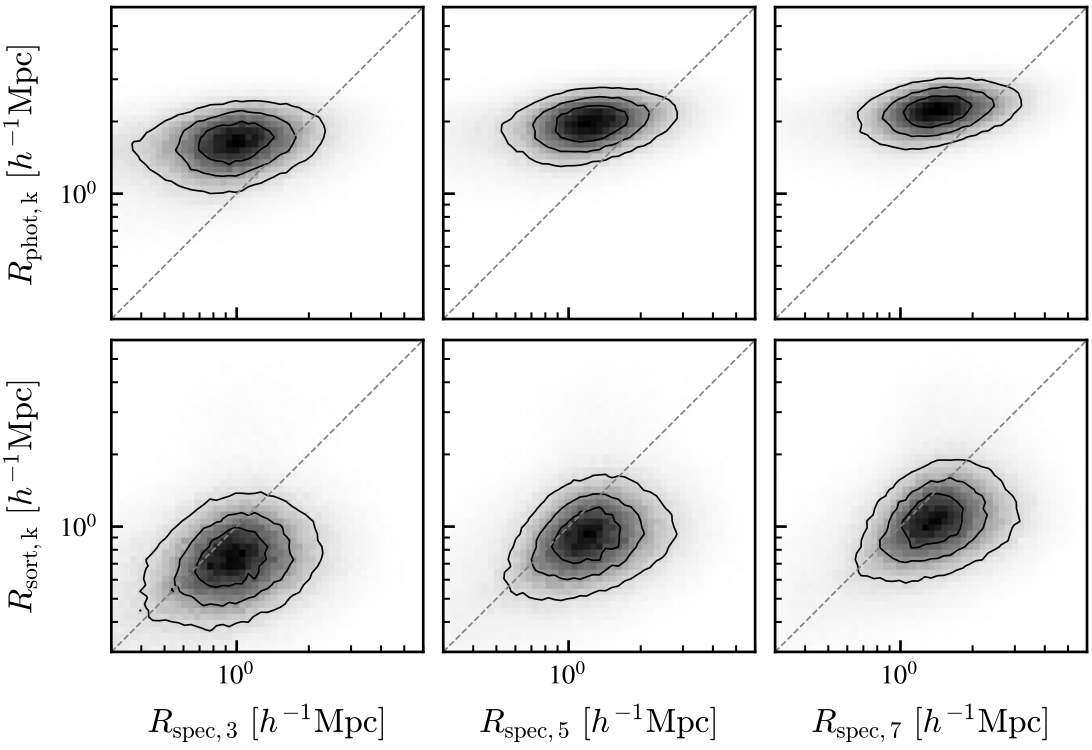

Figure D.9. Two-dimensional histograms of 3D distances to kth nearest neighbors using z_{phot} (top panels) and z_{sort} (bottom panels) compared to z_{spec} for $k = 3$, 5, and 7. Overall scatter is slightly increased using z_{sort}, but alignment with the line of equality is improved, particularly at smaller scales. At larger scales (corresponding to lower densities), SORT underestimates R_k as it packs galaxies too closely together in low-density environments. See Section 2.5.4 for details.

E. Sub-volume Completeness

The main goal of this paper was to study the clustering properties of centrals in the sSFR–M_* plane. Doing this involved breaking down our full galaxy sample into different sub-samples: primarily central, satellite, star-forming, green valley, and quiescent galaxies. Our objective was to isolate clustering trends in sSFR and therefore worked within five bins of stellar mass. Each bin would then represent a volume-limited sample complete in stellar mass over the mass range of the bin. To create such sub-volumes, we used the number density distributions as a function of redshift, $n_{\rm vol}(z)$, for each of the various sub-samples to find redshift bounds within which every sub-sample was complete. The specifics for our definition of completeness are described in Section 3.2.3. The $n_{\rm vol}(z)$ distributions used to define the redshift limits of our sub-volumes are shown in Figure E.1 (note that they have been normalized and scaled for clarity). The colored lines show $n_{\rm vol}(z)$ for the full redshift range, and the shaded regions show the limits where each sub-sample is complete. The final limits of each sub-volume, shown as vertical black lines, are taken as the redshift limits within which every sub-sample is complete.

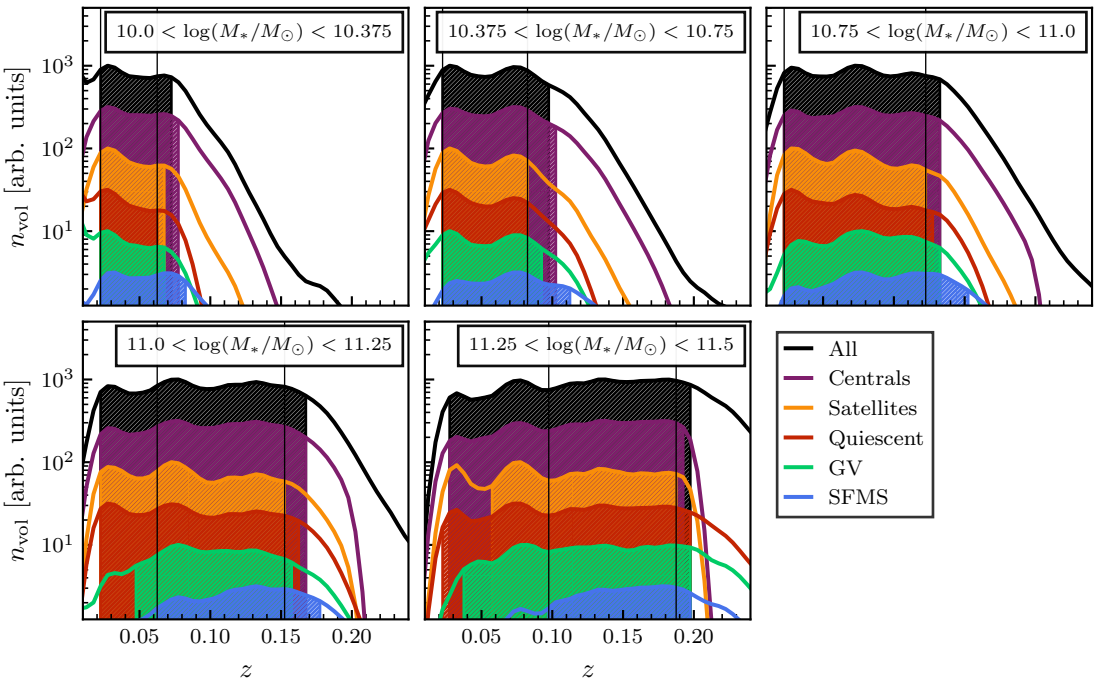

Figure E.1. Joint completeness of different sub-samples within each of our five sub-volumes. The full ranges of number densities are shown by the colored lines, and the shaded regions show the limits of completeness for each sub-sample. The vertical black lines denote the chosen redshift limits for each sub-volume, which are selected as the highest low-redshift limits and lowest high-redshift limits that allow for completeness in every sub-sample within each sub-volume. Note that the distributions have been normalized and scaled by increments of 0.5 dex for clarity.

Table E.1.: List of sub-samples in the sSFR–M_* plane with the corresponding mean and median values of $\log(M_*/M_\odot)$ and ΔMS and counts of central and satellite galaxies from Yang et al. (2012).

M_* Limits [$\log(M_*/M_\odot)$]	Sample	$\log(M_*/M_\odot)$		ΔMS [dex]		$N_{\rm gal}$
		Mean	Median	Mean	Median	
		Centrals				
10.0 – 10.375	HSF	10.179	10.177	0.389	0.353	1,456
	UMS	10.172	10.165	0.115	0.111	2,930
	LMS	10.181	10.174	-0.120	-0.117	2,989
	BMS	10.194	10.198	-0.344	-0.342	1,483
	GV	10.194	10.192	-0.705	-0.691	2,111
	Q	10.210	10.220	-1.404	-1.406	3,729
10.375 – 10.75	HSF	10.544	10.539	0.428	0.392	2,814
	UMS	10.536	10.523	0.115	0.111	4,137
	LMS	10.538	10.529	-0.121	-0.121	4,266
	BMS	10.548	10.540	-0.348	-0.347	2,380
	GV	10.557	10.554	-0.720	-0.714	4,597
	Q	10.568	10.572	-1.455	-1.478	13,273
10.75 – 11.0	HSF	10.859	10.852	0.474	0.425	3,614
	UMS	10.855	10.845	0.124	0.125	3,452
	LMS	10.860	10.852	-0.125	-0.124	3,419
	BMS	10.864	10.857	-0.349	-0.348	2,546
	GV	10.869	10.866	-0.735	-0.737	6,901
	Q	10.870	10.868	-1.412	-1.424	21,203
11.0 – 11.25	HSF	11.096	11.084	0.503	0.454	2,703
	UMS	11.095	11.083	0.120	0.119	2,250
	LMS	11.100	11.091	-0.125	-0.123	2,489
	BMS	11.104	11.095	-0.353	-0.351	2,202
	GV	11.108	11.101	-0.736	-0.739	6,953
	Q	11.112	11.105	-1.353	-1.360	23,229
11.25 – 11.5	HSF	11.334	11.319	0.534	0.475	954
	UMS	11.342	11.326	0.117	0.119	815
	LMS	11.341	11.326	-0.131	-0.132	1,066
	BMS	11.345	11.334	-0.351	-0.350	1,123
	GV	11.349	11.336	-0.768	-0.796	3,637
	Q	11.350	11.339	-1.294	-1.296	15,007

		Satellites				
10.0 – 10.375	HSF	10.172	10.168	0.404	0.358	437
	UMS	10.167	10.165	0.112	0.109	871
	LMS	10.170	10.163	-0.126	-0.130	971
	BMS	10.191	10.195	-0.349	-0.345	607
	GV	10.179	10.176	-0.723	-0.722	1,367
	Q	10.191	10.191	-1.452	-1.467	3,855
10.375 – 10.75	HSF	10.539	10.524	0.433	0.403	761
	UMS	10.537	10.526	0.115	0.113	1,080
	LMS	10.533	10.516	-0.123	-0.119	1,206
	BMS	10.543	10.535	-0.352	-0.354	813
	GV	10.547	10.539	-0.732	-0.731	2,192
	Q	10.555	10.551	-1.499	-1.527	8,433
10.75 – 11.0	HSF	10.859	10.849	0.477	0.428	715
	UMS	10.855	10.849	0.119	0.119	724
	LMS	10.856	10.851	-0.125	-0.125	776
	BMS	10.863	10.857	-0.356	-0.357	626
	GV	10.866	10.860	-0.737	-0.744	2,200
	Q	10.863	10.858	-1.437	-1.453	8,409
11.0 – 11.25	HSF	11.088	11.073	0.499	0.458	384
	UMS	11.091	11.078	0.109	0.102	357
	LMS	11.098	11.086	-0.130	-0.126	443
	BMS	11.101	11.090	-0.354	-0.354	471
	GV	11.100	11.089	-0.728	-0.722	1,521
	Q	11.101	11.091	-1.379	-1.385	5,903
11.25 – 11.5	HSF	11.319	11.307	0.516	0.456	115
	UMS	11.327	11.304	0.118	0.124	96
	LMS	11.338	11.321	-0.129	-0.127	114
	BMS	11.339	11.325	-0.354	-0.358	144
	GV	11.333	11.320	-0.765	-0.785	482
	Q	11.338	11.325	-1.308	-1.313	2,135

Bibliography

Kevork N. Abazajian, Jennifer K. Adelman-McCarthy, Marcel A. Agüeros, Sahar S. Allam, Carlos Allende Prieto, Deokkeun An, Kurt S. J. Anderson, Scott F. Anderson, James Annis, Neta A. Bahcall, C. A. L. Bailer-Jones, J. C. Barentine, Bruce A. Bassett, Andrew C. Becker, Timothy C. Beers, Eric F. Bell, Vasily Belokurov, Andreas A. Berlind, Eileen F. Berman, Mariangela Bernardi, Steven J. Bickerton, Dmitry Bizyaev, John P. Blakeslee, Michael R. Blanton, John J. Bochanski, William N. Boroski, Howard J. Brewington, Jarle Brinchmann, J. Brinkmann, Robert J. Brunner, Tamás Budavári, Larry N. Carey, Samuel Carliles, Michael A. Carr, Francisco J. Castander, David Cinabro, A. J. Connolly, István Csabai, Carlos E. Cunha, Paul C. Czarapata, James R. A. Davenport, Ernst de Haas, Ben Dilday, Mamoru Doi, Daniel J. Eisenstein, Michael L. Evans, N. W. Evans, Xiaohui Fan, Scott D. Friedman, Joshua A. Frieman, Masataka Fukugita, Boris T. Gänsicke, Evalyn Gates, Bruce Gillespie, G. Gilmore, Belinda Gonzalez, Carlos F. Gonzalez, Eva K. Grebel, James E. Gunn, Zsuzsanna Györy, Patrick B. Hall, Paul Harding, Frederick H. Harris, Michael Harvanek, Suzanne L. Hawley, Jeffrey J. E. Hayes, Timothy M. Heckman, John S. Hendry, Gregory S. Hennessy, Robert B. Hindsley, J. Hoblitt, Craig J. Hogan, David W. Hogg, Jon A. Holtzman, Joseph B. Hyde, Shin-ichi Ichikawa, Takashi Ichikawa, Myungshin Im, Željko Ivezić, Sebastian Jester, Linhua Jiang, Jennifer A. Johnson, Anders M. Jorgensen, Mario Jurić, Stephen M. Kent, R. Kessler, S. J. Kleinman, G. R. Knapp, Kohki Konishi, Richard G. Kron, Jurek Krzesinski, Nikolay Kuropatkin, Hubert Lampeitl, Svetlana Lebedeva, Myung Gyoon Lee, Young Sun Lee, R. French Leger, Sébastien Lépine, Nolan Li, Marcos Lima, Huan Lin, Daniel C. Long, Craig P. Loomis, Jon Loveday, Robert H. Lupton, Eugene Magnier, Olena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Bruce Margon, John P. Marriner, David Martínez-Delgado, Takahiko Matsubara, Peregrine M. McGehee, Timothy A. McKay, Avery Meiksin, Heather L. Morrison, Fergal Mullally, Jeffrey A. Munn, Tara Murphy, Thomas Nash, Ada Nebot, Jr. Neilsen, Eric H., Heidi Jo Newberg, Peter R. Newman, Robert C. Nichol, Tom Nicinski, Maria Nieto-Santisteban, Atsuko Nitta, Sadanori Okamura, Daniel J. Oravetz, Jeremiah P. Ostriker, Russell Owen, Nikhil Padmanabhan, Kaike Pan, Changbom Park, George Pauls, Jr. Peoples, John, Will J. Percival, Jeffrey R. Pier, Adrian C. Pope, Dimitri Pourbaix, Paul A. Price, Norbert Purger, Thomas Quinn, M. Jordan Raddick, Paola Re Fiorentin, Gordon T. Richards, Michael W. Richmond, Adam G. Riess, Hans-Walter Rix, Constance M.

Rockosi, Masao Sako, David J. Schlegel, Donald P. Schneider, Ralf-Dieter Scholz, Matthias R. Schreiber, Axel D. Schwope, Uroš Seljak, Branimir Sesar, Erin Sheldon, Kazu Shimasaku, Valena C. Sibley, A. E. Simmons, Thirupathi Sivarani, J. Allyn Smith, Martin C. Smith, Vernesa Smolčić, Stephanie A. Snedden, Albert Stebbins, Matthias Steinmetz, Chris Stoughton, Michael A. Strauss, Mark SubbaRao, Yasushi Suto, Alexander S. Szalay, István Szapudi, Paula Szkody, Masayuki Tanaka, Max Tegmark, Luis F. A. Teodoro, Aniruddha R. Thakar, Christy A. Tremonti, Douglas L. Tucker, Alan Uomoto, Daniel E. Vanden Berk, Jan Vandenberg, S. Vidrih, Michael S. Vogeley, Wolfgang Voges, Nicole P. Vogt, Yogesh Wadadekar, Shannon Watters, David H. Weinberg, Andrew A. West, Simon D. M. White, Brian C. Wilhite, Alainna C. Wonders, Brian Yanny, D. R. Yocum, Donald G. York, Idit Zehavi, Stefano Zibetti, and Daniel B. Zucker. The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182(2):543–558, June 2009. doi: 10.1088/0067-0049/182/2/543.

M. A. Aragon-Calvo, Rien van de Weygaert, Bernard J. T. Jones, and Bahram Mobasher. Submegaparsec individual photometric redshift estimation from cosmic web constraints. *MNRAS*, 454(1):463–477, November 2015. doi: 10.1093/mnras/stv1903.

V. Avila-Reese and C. Firmani. Properties of Disk Galaxies in a Hierarchical Formation Scenario. *Rev. Mex. Astron. Astrofis.*, 36:23, April 2000.

V. Avila-Reese, C. Firmani, and X. Hernández. On the Formation and Evolution of Disk Galaxies: Cosmological Initial Conditions and the Gravitational Collapse. *ApJ*, 505:37–49, September 1998. doi: 10.1086/306136.

I. K. Baldry, J. Liske, M. J. I. Brown, A. S. G. Robotham, S. P. Driver, L. Dunne, M. Alpaslan, S. Brough, M. E. Cluver, E. Eardley, D. J. Farrow, C. Heymans, H. Hildebrandt, A. M. Hopkins, L. S. Kelvin, J. Loveday, A. J. Moffett, P. Norberg, M. S. Owers, E. N. Taylor, A. H. Wright, S. P. Bamford, J. Bland-Hawthorn, N. Bourne, M. N. Bremer, M. Colless, C. J. Conselice, S. M. Croom, L. J. M. Davies, C. Foster, M. W. Grootes, B. W. Holwerda, D. H. Jones, P. R. Kafle, K. Kuijken, M. A. Lara-Lopez, Á. R. López-Sánchez, M. J. Meyer, S. Phillipps, W. J. Sutherland, E. van Kampen, and S. M. Wilkins. Galaxy And Mass Assembly: the G02 field, Herschel-ATLAS target selection and data release 3. *MNRAS*, 474(3):3875–3888, March 2018. doi: 10.1093/mnras/stx3042.

Michael Balogh, Vince Eke, Chris Miller, Ian Lewis, Richard Bower, Warrick Couch, Robert Nichol, Joss Bland-Hawthorn, Ivan K. Baldry, Carlton Baugh, Terry Bridges, Russell Cannon, Shaun Cole, Matthew Colless, Chris Collins, Nicholas Cross, Gavin Dalton, Roberto de Propris, Simon P. Driver, George Efstathiou, Richard S. Ellis, Carlos S. Frenk, Karl Glazebrook, Percy Gomez, Alex Gray, Edward Hawkins,

Carole Jackson, Ofer Lahav, Stuart Lumsden, Steve Maddox, Darren Madgwick, Peder Norberg, John A. Peacock, Will Percival, Bruce A. Peterson, Will Sutherland, and Keith Taylor. Galaxy ecology: groups and low-density environments in the SDSS and 2dFGRS. *MNRAS*, 348(4):1355–1372, March 2004. doi: 10.1111/j.1365-2966.2004.07453.x.

Matthew R. Becker. Connecting Galaxies with Halos Across Cosmic Time: Stellar mass assembly distribution modeling of galaxy statistics. *arXiv e-prints*, art. arXiv:1507.03605, July 2015. doi: 10.48550/arXiv.1507.03605.

Peter Behroozi, Risa H. Wechsler, Andrew P. Hearin, and Charlie Conroy. UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10. *MNRAS*, 488(3):3143–3194, September 2019a. doi: 10.1093/mnras/stz1182.

Peter Behroozi, Risa H. Wechsler, Andrew P. Hearin, and Charlie Conroy. UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10. *MNRAS*, 488(3):3143–3194, September 2019b. doi: 10.1093/mnras/stz1182.

Peter Behroozi, Andrew Hearin, and Benjamin P. Moster. Observational measures of halo properties beyond mass. *MNRAS*, 509(2):2800–2824, January 2022. doi: 10.1093/mnras/stab3193.

Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *ApJ*, 762(2):109, January 2013a. doi: 10.1088/0004-637X/762/2/109.

Peter S. Behroozi, Risa H. Wechsler, Hao-Yi Wu, Michael T. Busha, Anatoly A. Klypin, and Joel R. Primack. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. *ApJ*, 763(1):18, January 2013b. doi: 10.1088/0004-637X/763/1/18.

Andreas A. Berlind and David H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *ApJ*, 575 (2):587–616, August 2002a. doi: 10.1086/341469.

Andreas A. Berlind and David H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *ApJ*, 575 (2):587–616, August 2002b. doi: 10.1086/341469.

Angela M. Berti, Alison L. Coil, Andrew P. Hearin, and John Moustakas. PRIMUS: Clustering of Star-forming and Quiescent Central Galaxies at 0.2 < z < 0.9. *ApJ*, 884 (1):76, October 2019. doi: 10.3847/1538-4357/ab3b5d.

Angela M. Berti, Alison L. Coil, Andrew P. Hearin, and Peter S. Behroozi. Main-sequence Scatter is Real: The Joint Dependence of Galaxy Clustering on Star Formation and Stellar Mass. *AJ*, 161(1):49, January 2021a. doi: 10.3847/1538-3881/abcc6a.

Angela M. Berti, Alison L. Coil, Andrew P. Hearin, and Peter S. Behroozi. Main-sequence Scatter is Real: The Joint Dependence of Galaxy Clustering on Star Formation and Stellar Mass. *AJ*, 161(1):49, January 2021b. doi: 10.3847/1538-3881/abcc6a.

Michael R. Blanton and Andreas A. Berlind. What Aspects of Galaxy Environment Matter? *ApJ*, 664(2):791–803, August 2007. doi: 10.1086/512478.

Michael R. Blanton, Robert H. Lupton, David J. Schlegel, Michael A. Strauss, J. Brinkmann, Masataka Fukugita, and Jon Loveday. The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. *ApJ*, 631(1):208–230, September 2005a. doi: 10.1086/431416.

Michael R. Blanton, David J. Schlegel, Michael A. Strauss, J. Brinkmann, Douglas Finkbeiner, Masataka Fukugita, James E. Gunn, David W. Hogg, Željko Ivezić, G. R. Knapp, Robert H. Lupton, Jeffrey A. Munn, Donald P. Schneider, Max Tegmark, and Idit Zehavi. New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *AJ*, 129(6):2562–2578, June 2005b. doi: 10.1086/429803.

Josh Borrow, Mark Vogelsberger, Stephanie O'Neil, Michael A. McDonald, and Aaron Smith. There and back again: Understanding the critical properties of backsplash galaxies. *MNRAS*, 520(1):649–667, March 2023. doi: 10.1093/mnras/stad045.

Connor Bottrell, Hassen M. Yesuf, Gergö Popping, Kiyoaki Christopher Omori, Shenli Tang, Xuheng Ding, Annalisa Pillepich, Dylan Nelson, Lukas Eisert, Hua Gao, Andy D. Goulding, Boris S. Kalita, Wentao Luo, Jenny E. Greene, Jingjing Shi, and John D. Silverman. IllustrisTNG in the HSC-SSP: image data release and the major role of mini mergers as drivers of asymmetry and star formation. *MNRAS*, 527(3): 6506–6539, January 2024. doi: 10.1093/mnras/stad2971.

N. Bouché, A. Dekel, R. Genzel, S. Genel, G. Cresci, N. M. Förster Schreiber, K. L. Shapiro, R. I. Davies, and L. Tacconi. The Impact of Cold Gas Accretion Above a

Mass Floor on Galaxy Scaling Relations. *ApJ*, 718(2):1001–1018, August 2010. doi: 10.1088/0004-637X/718/2/1001.

J. Brinchmann, S. Charlot, S. D. M. White, C. Tremonti, G. Kauffmann, T. Heckman, and J. Brinkmann. The physical properties of star-forming galaxies in the low-redshift Universe. *MNRAS*, 351(4):1151–1179, July 2004. doi: 10.1111/j.1365-2966.2004.07881.x.

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344(4):1000–1028, October 2003. doi: 10.1046/j.1365-8711.2003.06897.x.

A. R. Calette, Aldo Rodríguez-Puebla, Vladimir Avila-Reese, and Claudia del P. Lagos. The galaxy H I-(sub)halo connection and the H I spatial clustering of local galaxies. *MNRAS*, 506(1):1507–1525, September 2021. doi: 10.1093/mnras/stab1788.

R. Cawthon, J. Elvin-Poole, A. Porredon, M. Crocce, G. Giannini, M. Gatti, A. J. Ross, E. S. Rykoff, A. Carnero Rosell, J. DeRose, S. Lee, M. Rodriguez-Monroy, A. Amon, K. Bechtol, J. De Vicente, D. Gruen, R. Morgan, E. Sanchez, J. Sanchez, I. Sevilla-Noarbe, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, S. Avila, D. Bacon, E. Bertin, D. Brooks, D. L. Burke, M. Carrasco Kind, J. Carretero, F. J. Castander, A. Choi, M. Costanzi, L. N. da Costa, M. E. S. Pereira, K. Dawson, S. Desai, H. T. Diehl, K. Eckert, S. Everett, I. Ferrero, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. Huterer, D. J. James, A. G. Kim, J. P. Kneib, K. Kuehn, N. Kuropatkin, O. Lahav, M. Lima, H. Lin, M. A. G. Maia, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, J. Muir, J. Myles, A. Palmese, S. Pandey, F. Paz-Chinchón, W. J. Percival, A. A. Plazas, A. Roodman, G. Rossi, V. Scarpine, S. Serrano, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, C. To, M. A. Troxel, R. D. Wilkinson, and the DES Collaboration. Dark Energy Survey Year 3 Results: Calibration of Lens Sample Redshift Distributions using Clustering Redshifts with BOSS/eBOSS. *arXiv e-prints*, art. arXiv:2012.12826, December 2020.

María Cebrián and Ignacio Trujillo. The effect of the environment on the stellar mass-size relationship for present-day galaxies. *MNRAS*, 444(1):682–699, October 2014. doi: 10.1093/mnras/stu1375.

Stéphane Charlot and Marcella Longhetti. Nebular emission from star-forming galaxies. *MNRAS*, 323(4):887–903, May 2001. doi: 10.1046/j.1365-8711.2001.04260.x.

Nima Chartab, Bahram Mobasher, Behnam Darvish, Steve Finkelstein, Yicheng Guo, Dritan Kodra, Kyoung-Soo Lee, Jeffrey A. Newman, Camilla Pacifici, Casey Papovich, Zahra Sattari, Abtin Shahidi, Mark E. Dickinson, Sandra M. Faber, Henry C.

Ferguson, Mauro Giavalisco, and Marziye Jafariyazani. Large-scale Structures in the CANDELS Fields: The Role of the Environment in Star Formation Activity. *ApJ*, 890(1):7, February 2020. doi: 10.3847/1538-4357/ab61fd.

Alison L. Coil, Alexander J. Mendez, Daniel J. Eisenstein, and John Moustakas. PRIMUS+DEEP2: The Dependence of Galaxy Clustering on Stellar Mass and Specific Star Formation Rate at 0.2 < z < 1.2. *ApJ*, 838(2):87, April 2017. doi: 10.3847/1538-4357/aa63ec.

Charlie Conroy, Risa H. Wechsler, and Andrey V. Kravtsov. Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. *ApJ*, 647(1):201–214, August 2006. doi: 10.1086/503602.

Asantha Cooray and Ravi Sheth. Halo models of large scale structure. *Phys. Rep.*, 372 (1):1–129, December 2002. doi: 10.1016/S0370-1573(02)00276-4.

Weiguang Cui, Romeel Davé, John A. Peacock, Daniel Anglés-Alcázar, and Xiaohu Yang. The origin of galaxy colour bimodality in the scatter of the stellar-to-halo mass relation. *Nature Astronomy*, 5:1069–1076, October 2021. doi: 10.1038/s41550-021-01404-1.

Romeel Davé, Daniel Anglés-Alcázar, Desika Narayanan, Qi Li, Mika H. Rafieferantsoa, and Sarah Appleby. SIMBA: Cosmological simulations with black hole growth and feedback. *MNRAS*, 486(2):2827–2849, June 2019. doi: 10.1093/mnras/stz937.

A. Dekel and N. Mandelker. An analytic solution for the minimal bathtub toy model: challenges in the star formation history of high-z galaxies. *MNRAS*, 444:2071–2084, November 2014. doi: 10.1093/mnras/stu1427.

DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey,

Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier

Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, art. arXiv:1611.00036, October 2016.

N. Chandrachani Devi, Aldo Rodríguez-Puebla, O. Valenzuela, Vladimir Avila-Reese, César Hernández-Aguayo, and Baojiu Li. The galaxy-halo connection in modified gravity cosmologies: environment dependence of galaxy luminosity function. *MNRAS*, 488(1):782–802, September 2019. doi: 10.1093/mnras/stz1664.

Radu Dragomir, Aldo Rodríguez-Puebla, Joel R. Primack, and Christoph T. Lee. Does the galaxy-halo connection vary with environment? *MNRAS*, 476(1):741–758, May 2018. doi: 10.1093/mnras/sty283.

A. Dressler. Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies. *ApJ*, 236:351–365, March 1980. doi: 10.1086/157753.

Aaron A. Dutton, Frank C. van den Bosch, and Avishai Dekel. On the origin of the galaxy star-formation-rate sequence: evolution and scatter. *MNRAS*, 405(3):1690–1710, July 2010. doi: 10.1111/j.1365-2966.2010.16620.x.

Onsi Fakhouri, Chung-Pei Ma, and Michael Boylan-Kolchin. The merger rates and mass assembly histories of dark matter haloes in the two Millennium simulations. *MNRAS*, 406(4):2267–2278, August 2010. doi: 10.1111/j.1365-2966.2010.16859.x.

Jerome J. Fang, S. M. Faber, David C. Koo, Aldo Rodríguez-Puebla, Yicheng Guo, Guillermo Barro, Peter Behroozi, Gabriel Brammer, Zhu Chen, Avishai Dekel, Henry C. Ferguson, Eric Gawiser, Mauro Giavalisco, Jeyhan Kartaltepe, Dale D. Kocevski, Anton M. Koekemoer, Elizabeth J. McGrath, Daniel McIntosh, Jeffrey A. Newman, Camilla Pacifici, Viraj Pandya, Pablo G. Pérez-González, Joel R. Primack, Brett Salmon, Jonathan R. Trump, Benjamin Weiner, S. P. Willner, Viviana Acquaviva, Tomas Dahlen, Steven L. Finkelstein, Kristian Finlator, Adriano Fontana, Audrey Galametz, Norman A. Grogin, Ruth Gruetzbauch, Seth Johnson, Bahram

Mobasher, Casey J. Papovich, Janine Pforr, Mara Salvato, P. Santini, Arjen van der Wel, Tommy Wiklind, and Stijn Wuyts. Demographics of Star-forming Galaxies since z ∼ 2.5. I. The UVJ Diagram in CANDELS. *ApJ*, 858(2):100, May 2018. doi: 10.3847/1538-4357/aabcba.

L. Gao, S. D. M. White, A. Jenkins, F. Stoehr, and V. Springel. The subhalo populations of ΛCDM dark haloes. *MNRAS*, 355:819–834, December 2004. doi: 10.1111/j.1365-2966.2004.08360.x.

Liang Gao, Volker Springel, and Simon D. M. White. The age dependence of halo clustering. *MNRAS*, 363(1):L66–L70, October 2005. doi: 10.1111/j.1745-3933.2005.00084.x.

Liang Gao, Julio F. Navarro, Shaun Cole, Carlos S. Frenk, Simon D. M. White, Volker Springel, Adrian Jenkins, and Angelo F. Neto. The redshift dependence of the structure of massive Λ cold dark matter haloes. *MNRAS*, 387(2):536–544, June 2008. doi: 10.1111/j.1365-2966.2008.13277.x.

M. Gatti, G. Giannini, G. M. Bernstein, A. Alarcon, J. Myles, A. Amon, R. Cawthon, M. Troxel, J. DeRose, S. Everett, A. J. Ross, E. S. Rykoff, J. Elvin-Poole, J. Cordero, I. Harrison, C. Sanchez, J. Prat, D. Gruen, H. Lin, M. Crocce, E. Rozo, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, S. Avila, D. Bacon, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, A. Choi, C. Conselice, M. Costanzi, M. Crocce, L. N. da Costa, M. E. S. Pereira, K. Dawson, S. Desai, H. T. Diehl, K. Eckert, T. F. Eifler, A. E. Evrard, I. Ferrero, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, T. Giannantonio, R. A. Gruendl, J. Gschwend, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. Huterer, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, M. Lima, N. MacCrann, M. A. G. Maia, M. March, J. L. Marshall, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, R. L. C. Ogando, A. Palmese, F. Paz-Chinchón, W. J. Percival, A. A. Plazas, M. Rodriguez-Monroy, A. Roodman, G. Rossi, S. Samuroff, E. Sanchez, V. Scarpine, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, T. N. Varga, J. Weller, R. D. Wilkinson, R. D. Wilkinson, and DES Collaboration. Dark Energy Survey Year 3 Results: clustering redshifts - calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS. *MNRAS*, 510(1):1223–1247, February 2022. doi: 10.1093/mnras/stab3311.

Sheridan B. Green, Frank C. van den Bosch, and Fangzhou Jiang. Erratum: The tidal evolution of dark matter substructure - II. The impact of artificial disruption on subhalo mass functions and radial profiles. *MNRAS*, 508(2):2944–2945, December 2021a. doi: 10.1093/mnras/stab2786.

Sheridan B. Green, Frank C. van den Bosch, and Fangzhou Jiang. SatGen - II. Assessing the impact of a disc potential on subhalo populations. *MNRAS*, 509(2):2624–2636, October 2021b. doi: 10.1093/mnras/stab3130.

Hong Guo, Zheng Zheng, Peter S. Behroozi, Idit Zehavi, Chia-Hsun Chuang, Johan Comparat, Ginevra Favole, Stefan Gottloeber, Anatoly Klypin, Francisco Prada, Sergio A. Rodríguez-Torres, David H. Weinberg, and Gustavo Yepes. Modelling galaxy clustering: halo occupation distribution versus subhalo matching. *MNRAS*, 459(3):3040–3058, July 2016. doi: 10.1093/mnras/stw845.

Qi Guo, Simon White, Michael Boylan-Kolchin, Gabriella De Lucia, Guinevere Kauffmann, Gerard Lemson, Cheng Li, Volker Springel, and Simone Weinmann. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. *MNRAS*, 413(1):101–131, May 2011. doi: 10.1111/j.1365-2966.2010.18114.x.

Andrew P. Hearin and Douglas F. Watson. The dark side of galaxy colour. *MNRAS*, 435(2):1313–1324, October 2013. doi: 10.1093/mnras/stt1374.

Sébastien Heinis, Tamás Budavári, Alex S. Szalay, Stéphane Arnouts, Miguel A. Aragón-Calvo, Ted K. Wyder, Tom A. Barlow, Karl Foster, Friedman G. Peter, D. Christopher Martin, Patrick Morrissey, Susan G. Neff, David Schiminovich, Mark Seibert, Luciana Bianchi, José Donas, Timothy M. Heckman, Young-Wook Lee, Barry F. Madore, Bruno Milliard, R. Michael Rich, and Sukyoung K. Yi. Spatial Clustering from GALEX-SDSS Samples: Star Formation History and Large-Scale Clustering. *ApJ*, 698(2):1838–1851, June 2009. doi: 10.1088/0004-637X/698/2/1838.

H. Hildebrandt, J. L. van den Busch, A. H. Wright, C. Blake, B. Joachimi, K. Kuijken, T. Tröster, M. Asgari, M. Bilicki, J. T. A. de Jong, A. Dvornik, T. Erben, F. Getman, B. Giblin, C. Heymans, A. Kannawadi, C. A. Lin, and H. Y. Shan. KiDS-1000 catalogue: Redshift distributions and their calibration. *A&A*, 647:A124, March 2021. doi: 10.1051/0004-6361/202039018.

Željko Ivezić, Steven M. Kahn, J. Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F. Anderson, John Andrew, James Roger P. Angel, George Z. Angeli, Reza Ansari, Pierre Antilogus, Constanza Araujo, Robert Armstrong, Kirk T. Arndt, Pierre Astier, Éric Aubourg, Nicole Auza, Tim S. Axelrod, Deborah J. Bard, Jeff D. Barr, Aurelian Barrau, James G. Bartlett, Amanda E. Bauer, Brian J. Bauman, Sylvain Baumont, Ellen Bechtol, Keith Bechtol, Andrew C. Becker, Jacek Becla, Cristina Beldica, Steve Bellavia, Federica B. Bianco, Rahul Biswas, Guillaume Blanc, Jonathan Blazek, Roger D. Blandford, Josh S. Bloom, Joanne Bogart, Tim W. Bond, Michael T. Booth, Anders W. Borgland, Kirk Borne, James F.

Bosch, Dominique Boutigny, Craig A. Brackett, Andrew Bradshaw, William Nielsen Brandt, Michael E. Brown, James S. Bullock, Patricia Burchat, David L. Burke, Gianpietro Cagnoli, Daniel Calabrese, Shawn Callahan, Alice L. Callen, Jeffrey L. Carlin, Erin L. Carlson, Srinivasan Chandrasekharan, Glenaver Charles-Emerson, Steve Chesley, Elliott C. Cheu, Hsin-Fang Chiang, James Chiang, Carol Chirino, Derek Chow, David R. Ciardi, Charles F. Claver, Johann Cohen-Tanugi, Joseph J. Cockrum, Rebecca Coles, Andrew J. Connolly, Kem H. Cook, Asantha Cooray, Kevin R. Covey, Chris Cribbs, Wei Cui, Roc Cutri, Philip N. Daly, Scott F. Daniel, Felipe Daruich, Guillaume Daubard, Greg Daues, William Dawson, Francisco Delgado, Alfred Dellapenna, Robert de Peyster, Miguel de Val-Borro, Seth W. Digel, Peter Doherty, Richard Dubois, Gregory P. Dubois-Felsmann, Josef Durech, Frossie Economou, Tim Eifler, Michael Eracleous, Benjamin L. Emmons, Angelo Fausti Neto, Henry Ferguson, Enrique Figueroa, Merlin Fisher-Levine, Warren Focke, Michael D. Foss, James Frank, Michael D. Freemon, Emmanuel Gangler, Eric Gawiser, John C. Geary, Perry Gee, Marla Geha, Charles J. B. Gessner, Robert R. Gibson, D. Kirk Gilmore, Thomas Glanzman, William Glick, Tatiana Goldina, Daniel A. Goldstein, Iain Goodenow, Melissa L. Graham, William J. Gressler, Philippe Gris, Leanne P. Guy, Augustin Guyonnet, Gunther Haller, Ron Harris, Patrick A. Hascall, Justine Haupt, Fabio Hernandez, Sven Herrmann, Edward Hileman, Joshua Hoblitt, John A. Hodgson, Craig Hogan, James D. Howard, Dajun Huang, Michael E. Huffer, Patrick Ingraham, Walter R. Innes, Suzanne H. Jacoby, Bhuvnesh Jain, Fabrice Jammes, M. James Jee, Tim Jenness, Garrett Jernigan, Darko Jevremović, Kenneth Johns, Anthony S. Johnson, Margaret W. G. Johnson, R. Lynne Jones, Claire Juramy-Gilles, Mario Jurić, Jason S. Kalirai, Nitya J. Kallivayalil, Bryce Kalmbach, Jeffrey P. Kantor, Pierre Karst, Mansi M. Kasliwal, Heather Kelly, Richard Kessler, Veronica Kinnison, David Kirkby, Lloyd Knox, Ivan V. Kotov, Victor L. Krabbendam, K. Simon Krughoff, Petr Kubánek, John Kuczewski, Shri Kulkarni, John Ku, Nadine R. Kurita, Craig S. Lage, Ron Lambert, Travis Lange, J. Brian Langton, Laurent Le Guillou, Deborah Levine, Ming Liang, Kian-Tat Lim, Chris J. Lintott, Kevin E. Long, Margaux Lopez, Paul J. Lotz, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ashish Mahabal, Rachel Mandelbaum, Thomas W. Markiewicz, Darren S. Marsh, Philip J. Marshall, Stuart Marshall, Morgan May, Robert McKercher, Michelle McQueen, Joshua Meyers, Myriam Migliore, Michelle Miller, David J. Mills, Connor Miraval, Joachim Moeyens, Fred E. Moolekamp, David G. Monet, Marc Moniez, Serge Monkewitz, Christopher Montgomery, Christopher B. Morrison, Fritz Mueller, Gary P. Muller, Freddy Muñoz Arancibia, Douglas R. Neill, Scott P. Newbry, Jean-Yves Nief, Andrei Nomerotski, Martin Nordby, Paul O'Connor, John Oliver, Scot S. Olivier, Knut Olsen, William O'Mullane, Sandra Ortiz, Shawn Osier, Russell E. Owen, Reynald Pain, Paul E. Palecek, John K. Parejko, James B. Parsons, Nathan M. Pease, J. Matt

Peterson, John R. Peterson, Donald L. Petravick, M. E. Libby Petrick, Cathy E. Petry, Francesco Pierfederici, Stephen Pietrowicz, Rob Pike, Philip A. Pinto, Raymond Plante, Stephen Plate, Joel P. Plutchak, Paul A. Price, Michael Prouza, Veljko Radeka, Jayadev Rajagopal, Andrew P. Rasmussen, Nicolas Regnault, Kevin A. Reil, David J. Reiss, Michael A. Reuter, Stephen T. Ridgway, Vincent J. Riot, Steve Ritz, Sean Robinson, William Roby, Aaron Roodman, Wayne Rosing, Cecille Roucelle, Matthew R. Rumore, Stefano Russo, Abhijit Saha, Benoit Sassolas, Terry L. Schalk, Pim Schellart, Rafe H. Schindler, Samuel Schmidt, Donald P. Schneider, Michael D. Schneider, William Schoening, German Schumacher, Megan E. Schwamb, Jacques Sebag, Brian Selvy, Glenn H. Sembroski, Lynn G. Seppala, Andrew Serio, Eduardo Serrano, Richard A. Shaw, Ian Shipsey, Jonathan Sick, Nicole Silvestri, Colin T. Slater, J. Allyn Smith, R. Chris Smith, Shahram Sobhani, Christine Soldahl, Lisa Storrie-Lombardi, Edward Stover, Michael A. Strauss, Rachel A. Street, Christopher W. Stubbs, Ian S. Sullivan, Donald Sweeney, John D. Swinbank, Alexander Szalay, Peter Takacs, Stephen A. Tether, Jon J. Thaler, John Gregg Thayer, Sandrine Thomas, Adam J. Thornton, Vaikunth Thukral, Jeffrey Tice, David E. Trilling, Max Turri, Richard Van Berg, Daniel Vanden Berk, Kurt Vetter, Francoise Virieux, Tomislav Vucina, William Wahl, Lucianne Walkowicz, Brian Walsh, Christopher W. Walter, Daniel L. Wang, Shin-Yawn Wang, Michael Warner, Oliver Wiecha, Beth Willman, Scott E. Winters, David Wittman, Sidney C. Wolff, W. Michael Wood-Vasey, Xiuqin Wu, Bo Xin, Peter Yoachim, and Hu Zhan. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

Fangzhou Jiang, Avishai Dekel, Jonathan Freundlich, Frank C. van den Bosch, Sheridan B. Green, Philip F. Hopkins, Andrew Benson, and Xiaolong Du. SatGen: a semi-analytical satellite galaxy generator - I. The model and its application to Local-Group satellite statistics. *MNRAS*, 502(1):621–641, March 2021. doi: 10.1093/mnras/staa4034.

James Kakos, Joel R. Primack, Aldo Rodríguez-Puebla, Nicolas Tejos, L. Y. Aaron Yung, and Rachel S. Somerville. Galaxy correlation function and local density from photometric redshifts using the stochastic order redshift technique (SORT). *MNRAS*, 514(2):1857–1878, August 2022. doi: 10.1093/mnras/stac1307.

James Kakos, Aldo Rodriguez-Puebla, Joel R. Primack, Sandra M. Faber, David C. Koo, Peter Behroozi, and Vladimir Avila-Reese. Star-forming and quiescent central galaxies cluster similarly: Implications for the galaxy-halo connection, 2024.

Guinevere Kauffmann, Timothy M. Heckman, Simon D. M. White, Stéphane Charlot, Christy Tremonti, Jarle Brinchmann, Gustavo Bruzual, Eric W. Peng, Mark Seibert,

Mariangela Bernardi, Michael Blanton, Jon Brinkmann, Francisco Castander, Istvan Csábai, Masataka Fukugita, Zeljko Ivezic, Jeffrey A. Munn, Robert C. Nichol, Nikhil Padmanabhan, Aniruddha R. Thakar, David H. Weinberg, and Donald York. Stellar masses and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey. *MNRAS*, 341(1):33–53, May 2003. doi: 10.1046/j.1365-8711.2003.06291.x.

Guinevere Kauffmann, Simon D. M. White, Timothy M. Heckman, Brice Ménard, Jarle Brinchmann, Stéphane Charlot, Christy Tremonti, and Jon Brinkmann. The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *MNRAS*, 353(3):713–731, September 2004. doi: 10.1111/j.1365-2966.2004.08117.x.

Jr. Kennicutt, Robert C. Star Formation in Galaxies Along the Hubble Sequence. *ARA&A*, 36:189–232, January 1998. doi: 10.1146/annurev.astro.36.1.189.

Anatoly Klypin, Gustavo Yepes, Stefan Gottlöber, Francisco Prada, and Steffen Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *MNRAS*, 457(4):4340–4359, April 2016a. doi: 10.1093/mnras/stw248.

Anatoly Klypin, Gustavo Yepes, Stefan Gottlöber, Francisco Prada, and Steffen Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *MNRAS*, 457(4):4340–4359, April 2016b. doi: 10.1093/mnras/stw248.

Anatoly A. Klypin, Sebastian Trujillo-Gomez, and Joel Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. *ApJ*, 740 (2):102, October 2011. doi: 10.1088/0004-637X/740/2/102.

Pavel Kroupa. On the variation of the initial mass function. *MNRAS*, 322(2):231–246, April 2001. doi: 10.1046/j.1365-8711.2001.04022.x.

Stephen D. Landy and Alexander S. Szalay. Bias and Variance of Angular Correlation Functions. *ApJ*, 412:64, July 1993. doi: 10.1086/172900.

Stephen D. Landy, Alexander S. Szalay, and David C. Koo. Strong Angular Clustering of Very Blue Galaxies: Evidence of a Low-Redshift Population. *ApJ*, 460:94, March 1996. doi: 10.1086/176954.

Johannes U. Lange, Frank C. van den Bosch, Andrew R. Zentner, Kuan Wang, and Antonio S. Villarreal. Updated results on the galaxy-halo connection from satellite kinematics in SDSS. *MNRAS*, 487(3):3112–3129, August 2019. doi: 10.1093/mnras/stz1466.

Christoph T. Lee, Joel R. Primack, Peter Behroozi, Aldo Rodríguez-Puebla, Doug Hellinger, and Avishai Dekel. Properties of dark matter haloes as a function of local environment density. *MNRAS*, 466(4):3834–3858, April 2017a. doi: 10.1093/mnras/stw3348.

Christoph T. Lee, Joel R. Primack, Peter Behroozi, Aldo Rodríguez-Puebla, Doug Hellinger, and Avishai Dekel. Properties of dark matter haloes as a function of local environment density. *MNRAS*, 466(4):3834–3858, April 2017b. doi: 10.1093/mnras/stw3348.

Nicholas Lee, D. B. Sanders, Caitlin M. Casey, Sune Toft, N. Z. Scoville, Chao-Ling Hung, Emeric Le Floc'h, Olivier Ilbert, H. Jabran Zahid, Hervé Aussel, Peter Capak, Jeyhan S. Kartaltepe, Lisa J. Kewley, Yanxia Li, Kevin Schawinski, Kartik Sheth, and Quanbao Xiao. A Turnover in the Galaxy Main Sequence of Star Formation at $M_* \sim 10^{10}$ $M_\odot$ for Redshifts z < 1.3. *ApJ*, 801(2):80, March 2015. doi: 10.1088/0004-637X/801/2/80.

Cheng Li, Guinevere Kauffmann, Y. P. Jing, Simon D. M. White, Gerhard Börner, and F. Z. Cheng. The dependence of clustering on galaxy properties. *MNRAS*, 368(1):21–36, May 2006. doi: 10.1111/j.1365-2966.2006.10066.x.

Lihwai Lin, David R. Patton, David C. Koo, Kevin Casteels, Christopher J. Conselice, S. M. Faber, Jennifer Lotz, Christopher N. A. Willmer, B. C. Hsieh, Tzihong Chiueh, Jeffrey A. Newman, Gregory S. Novak, Benjamin J. Weiner, and Michael C. Cooper. The Redshift Evolution of Wet, Dry, and Mixed Galaxy Mergers from Close Galaxy Pairs in the DEEP2 Galaxy Redshift Survey. *ApJ*, 681(1):232–243, July 2008. doi: 10.1086/587928.

Andrea V. Macciò, Aaron A. Dutton, and Frank C. van den Bosch. Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results. *MNRAS*, 391(4):1940–1954, December 2008. doi: 10.1111/j.1365-2966.2008.14029.x.

Rachel Mandelbaum, Uroš Seljak, Guinevere Kauffmann, Christopher M. Hirata, and Jonathan Brinkmann. Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies. *MNRAS*, 368(2):715–731, May 2006. doi: 10.1111/j.1365-2966.2006.10156.x.

Rachel Mandelbaum, Wenting Wang, Ying Zu, Simon White, Bruno Henriques, and Surhud More. Strong bimodality in the host halo mass of central galaxies from galaxy-galaxy lensing. *MNRAS*, 457(3):3200–3218, April 2016. doi: 10.1093/mnras/stw188.

S. Masaki, Y.-T. Lin, and N. Yoshida. Modelling colour-dependent galaxy clustering in cosmological simulations. *MNRAS*, 436:2286–2300, December 2013. doi: 10.1093/mnras/stt1729.

Jorryt Matthee and Joop Schaye. The origin of scatter in the star formation rate-stellar mass relation. *MNRAS*, 484(1):915–932, March 2019. doi: 10.1093/mnras/stz030.

Christian Maulbetsch, Vladimir Avila-Reese, Pedro Colín, Stefan Gottlöber, Arman Khalatyan, and Matthias Steinmetz. The Dependence of the Mass Assembly History of Cold Dark Matter Halos on Environment. *ApJ*, 654(1):53–65, January 2007. doi: 10.1086/509706.

Brice Ménard, Ryan Scranton, Samuel Schmidt, Chris Morrison, Donghui Jeong, Tamas Budavari, and Mubdi Rahman. Clustering-based redshift estimation: method and application to data. *arXiv e-prints*, art. arXiv:1303.4722, March 2013.

Houjun Mo, Frank van den Bosch, and Simon White. *Galaxy Formation and Evolution*. Cambridge University Press, 2010.

Antonio D. Montero-Dorta, Jonás Chaves-Montero, M. Celeste Artale, and Ginevra Favole. On the influence of halo mass accretion history on galaxy properties and assembly bias. *MNRAS*, 508(1):940–949, November 2021. doi: 10.1093/mnras/stab2556.

Surhud More, Frank C. van den Bosch, Marcello Cacciato, Ramin Skibba, H. J. Mo, and Xiaohu Yang. Satellite kinematics - III. Halo masses of central galaxies in SDSS. *MNRAS*, 410(1):210–226, January 2011. doi: 10.1111/j.1365-2966.2010.17436.x.

C. B. Morrison, H. Hildebrandt, S. J. Schmidt, I. K. Baldry, M. Bilicki, A. Choi, T. Erben, and P. Schneider. the-wizz: clustering redshift estimation for everyone. *MNRAS*, 467(3):3576–3589, May 2017. doi: 10.1093/mnras/stx342.

Benjamin P. Moster, Rachel S. Somerville, Christian Maulbetsch, Frank C. van den Bosch, Andrea V. Macciò, Thorsten Naab, and Ludwig Oser. Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. *ApJ*, 710(2):903–923, February 2010. doi: 10.1088/0004-637X/710/2/903.

Benjamin P. Moster, Thorsten Naab, and Simon D. M. White. EMERGE - an empirical model for the formation of galaxies since z $\sim$ 10. *MNRAS*, 477(2):1822–1852, June 2018. doi: 10.1093/mnras/sty655.

Benjamin P. Moster, Thorsten Naab, and Simon D. M. White. EMERGE - empirical constraints on the formation of passive galaxies. *MNRAS*, 499(4):4748–4767, December 2020. doi: 10.1093/mnras/staa3019.

Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, May 1996. doi: 10.1086/177173.

Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490(2):493–508, December 1997. doi: 10.1086/304888.

Christine O'Donnell, Peter Behroozi, and Surhud More. Observing correlations between dark matter accretion and galaxy growth - I. Recent star formation activity in isolated Milky Way-mass galaxies. *MNRAS*, 501(1):1253–1272, February 2021. doi: 10.1093/mnras/staa3654.

Christine O'Donnell, Peter Behroozi, and Surhud More. Observing correlations between dark matter accretion and galaxy growth: II. testing the impact of galaxy mass, star formation indicator, and neighbour colours. *MNRAS*, 509(3):3285–3300, January 2022. doi: 10.1093/mnras/stab3170.

W. J. Pearson, L. Wang, S. Brough, B. W. Holwerda, A. M. Hopkins, and J. Loveday. Galaxy and Mass Assembly: Group and field galaxy morphologies in the star-formation rate - stellar mass plane. *A&A*, 646:A151, February 2021. doi: 10.1051/0004-6361/202039177.

Ying-jie Peng, Simon J. Lilly, Katarina Kovač, Micol Bolzonella, Lucia Pozzetti, Alvio Renzini, Gianni Zamorani, Olivier Ilbert, Christian Knobel, Angela Iovino, Christian Maier, Olga Cucciati, Lidia Tasca, C. Marcella Carollo, John Silverman, Pawel Kampczyk, Loic de Ravel, David Sanders, Nicholas Scoville, Thierry Contini, Vincenzo Mainieri, Marco Scodeggio, Jean-Paul Kneib, Olivier Le Fèvre, Sandro Bardelli, Angela Bongiorno, Karina Caputi, Graziano Coppa, Sylvain de la Torre, Paolo Franzetti, Bianca Garilli, Fabrice Lamareille, Jean-Francois Le Borgne, Vincent Le Brun, Marco Mignoli, Enrique Perez Montero, Roser Pello, Elena Ricciardelli, Masayuki Tanaka, Laurence Tresse, Daniela Vergani, Niraj Welikala, Elena Zucca, Pascal Oesch, Ummi Abbas, Luke Barnes, Rongmon Bordoloi, Dario Bottini, Alberto Cappi, Paolo Cassata, Andrea Cimatti, Marco Fumana, Gunther Hasinger, Anton Koekemoer, Alexei Leauthaud, Dario Maccagni, Christian Marinoni, Henry McCracken, Pierdomenico Memeo, Baptiste Meneux, Preethi Nair, Cristiano Porciani, Valentina Presotto, and Roberto Scaramella. Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function. *ApJ*, 721(1):193–221, September 2010. doi: 10.1088/0004-637X/721/1/193.

S. Phillipps and T. Shanks. On the variation of galaxy correlations with luminosity. *MNRAS*, 229:621–626, December 1987. doi: 10.1093/mnras/229.4.621.

Antonio J. Porras-Valverde, John C. Forbes, Rachel S. Somerville, Adam R. H. Stevens, Kelly Holley-Bockelmann, Andreas A. Berlind, and Shy Genel. Why do semi-analytic models predict higher scatter in the stellar mass-halo mass relation than cosmological hydrodynamic simulations? *arXiv e-prints*, art. arXiv:2310.11507, October 2023. doi: 10.48550/arXiv.2310.11507.

Mubdi Rahman, Brice Ménard, Ryan Scranton, Samuel J. Schmidt, and Christopher B. Morrison. Clustering-based redshift estimation: comparison to spectroscopic redshifts. *MNRAS*, 447(4):3500–3511, March 2015. doi: 10.1093/mnras/stu2636.

R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi. The Connection between Galaxies and Dark Matter Structures in the Local Universe. *ApJ*, 771:30, July 2013. doi: 10.1088/0004-637X/771/1/30.

Facundo Rodriguez and Manuel Merchán. Combining friend-of-friend and halo-based algorithms for the identification of galaxy groups. *A&A*, 636:A61, April 2020. doi: 10.1051/0004-6361/201937423.

A. Rodríguez-Puebla, N. Drory, and V. Avila-Reese. The Stellar-Subhalo Mass Relation of Satellite Galaxies. *ApJ*, 756(1):2, September 2012. doi: 10.1088/0004-637X/756/1/2.

A. Rodríguez-Puebla, V. Avila-Reese, and N. Drory. The Galaxy-Halo/Subhalo Connection: Mass Relations and Implications for Some Satellite Occupational Distributions. *ApJ*, 767(1):92, April 2013. doi: 10.1088/0004-637X/767/1/92.

A. Rodríguez-Puebla, J. R. Primack, P. Behroozi, and S. M. Faber. Is main-sequence galaxy star formation controlled by halo mass accretion? *MNRAS*, 455:2592–2606, January 2016. doi: 10.1093/mnras/stv2513.

Aldo Rodríguez-Puebla, Vladimir Avila-Reese, Xiaohu Yang, Sebastien Foucaud, Niv Drory, and Y. P. Jing. The Stellar-to-Halo Mass Relation of Local Galaxies Segregates by Color. *ApJ*, 799(2):130, February 2015. doi: 10.1088/0004-637X/799/2/130.

Aldo Rodríguez-Puebla, Peter Behroozi, Joel Primack, Anatoly Klypin, Christoph Lee, and Doug Hellinger. Halo and subhalo demographics with Planck cosmological parameters: Bolshoi-Planck and MultiDark-Planck simulations. *MNRAS*, 462(1):893–916, October 2016a. doi: 10.1093/mnras/stw1705.

Aldo Rodríguez-Puebla, Peter Behroozi, Joel Primack, Anatoly Klypin, Christoph Lee, and Doug Hellinger. Halo and subhalo demographics with Planck cosmological parameters: Bolshoi-Planck and MultiDark-Planck simulations. *MNRAS*, 462(1):893–916, October 2016b. doi: 10.1093/mnras/stw1705.

Aldo Rodríguez-Puebla, Vladimir Avila-Reese, Mariana Cano-Díaz, S. M. Faber, Joel R. Primack, José Franco, I. Aretxaga, and Eder Santiago-Mayoral. The Star-forming Main Sequence and the Contribution of Dust-obscured Star Formation since z $\sim$ 4 from the Far-UV+IR Luminosity Functions. *ApJ*, 905(2):171, December 2020a. doi: 10.3847/1538-4357/abc7c2.

Aldo Rodríguez-Puebla, A. R. Calette, Vladimir Avila-Reese, Vicente Rodriguez-Gomez, and Marc Huertas-Company. The bivariate gas-stellar mass distributions and the mass functions of early- and late-type galaxies at z $\sim$ 0. *Publ. Astron. Soc. Australia*, 37:e024, June 2020b. doi: 10.1017/pasa.2020.15.

R. Scaramella, J. Amiaux, Y. Mellier, C. Burigana, C. S. Carvalho, J. C. Cuillandre, A. Da Silva, A. Derosa, J. Dinis, E. Maiorano, M. Maris, I. Tereno, R. Laureijs, T. Boenke, G. Buenadicha, X. Dupac, L. M. Gaspar Venancio, P. Gómez-Álvarez, J. Hoar, J. Lorenzo Alvarez, G. D. Racca, G. Saavedra-Criado, J. Schwartz, R. Vavrek, M. Schirmer, H. Aussel, R. Azzollini, V. F. Cardone, M. Cropper, A. Ealet, B. Garilli, W. Gillard, B. R. Granett, L. Guzzo, H. Hoekstra, K. Jahnke, T. Kitching, M. Meneghetti, L. Miller, R. Nakajima, S. M. Niemi, F. Pasian, W. J. Percival, M. Sauvage, M. Scodeggio, S. Wachter, A. Zacchei, N. Aghanim, A. Amara, T. Auphan, N. Auricchio, S. Awan, A. Balestra, R. Bender, C. Bodendorf, D. Bonino, E. Branchini, S. Brau-Nogue, M. Brescia, G. P. Candini, V. Capobianco, C. Carbone, R. G. Carlberg, J. Carretero, R. Casas, F. J. Castander, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, C. J. Conselice, L. Conversi, Y. Copin, L. Corcione, A. Costille, F. Courbin, H. Degaudenzi, M. Douspis, F. Dubath, C. A. J. Duncan, S. Dusini, S. Farrens, S. Ferriol, P. Fosalba, N. Fourmanoit, M. Frailis, E. Franceschi, P. Franzetti, M. Fumana, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, F. Hormuth, P. Hudelot, S. Kermiche, A. Kiessling, M. Kilbinger, R. Kohley, B. Kubik, M. Kümmel, M. Kunz, H. Kurki-Suonio, S. Ligori, P. B. Lilje, I. Lloro, O. Mansutti, O. Marggraf, K. Markovic, F. Marulli, R. Massey, S. Maurogordato, M. Melchior, E. Merlin, G. Meylan, J. J. Mohr, M. Moresco, B. Morin, L. Moscardini, E. Munari, R. C. Nichol, C. Padilla, S. Paltani, J. Peacock, K. Pedersen, V. Pettorino, S. Pires, M. Poncet, L. Popa, L. Pozzetti, F. Raison, R. Rebolo, J. Rhodes, H. W. Rix, M. Roncarelli, E. Rossetti, R. Saglia, P. Schneider, T. Schrabback, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, J. Skottfelt, L. Stanco, J. L. Starck, P. Tallada-Crespí, D. Tavagnacco, A. N. Taylor, H. I. Teplitz, R. Toledo-Moreo, F. Torradeflot, M. Trifoglio, E. A. Valentijn, L. Valenziano, G. A. Verdoes Kleijn, Y. Wang, N. Welikala, J. Weller, M. Wetzstein, G. Zamorani, J. Zoubian, S. Andreon, M. Baldi, S. Bardelli, A. Boucaud, S. Camera, G. Fabbian, R. Farinelli, J. Graciá-Carpio, D. Maino, E. Medinaceli, S. Mei, C. Neissner, G. Polenta, A. Renzi, E. Romelli, C. Rosset, F. Sureau, M. Tenti, T. Vassallo,

E. Zucca, C. Baccigalupi, A. Balaguera-Antolínez, P. Battaglia, A. Biviano, S. Borgani, E. Bozzo, R. Cabanac, A. Cappi, S. Casas, G. Castignani, C. Colodro-Conde, J. Coupon, H. M. Courtois, J. Cuby, S. de la Torre, S. Desai, D. Di Ferdinando, H. Dole, M. Fabricius, M. Farina, P. G. Ferreira, F. Finelli, P. Flose-Reimberg, S. Fotopoulou, S. Galeotta, K. Ganga, G. Gozaliasl, I. M. Hook, E. Keihanen, C. C. Kirkpatrick, P. Liebing, V. Lindholm, G. Mainetti, M. Martinelli, N. Martinet, M. Maturi, H. J. McCracken, R. B. Metcalf, G. Morgante, J. Nightingale, A. Nucita, L. Patrizii, D. Potter, G. Riccio, A. G. Sánchez, D. Sapone, J. A. Schewtschenko, M. Schultheis, V. Scottez, R. Teyssier, I. Tutusaus, J. Valiviita, M. Viel, W. Vriend, and L. Whittaker. Euclid preparation: I. The Euclid Wide Survey. *arXiv e-prints*, art. arXiv:2108.01201, August 2021.

Samuel J. Schmidt, Brice Ménard, Ryan Scranton, Christopher Morrison, and Cameron K. McBride. Recovering redshift distributions with cross-correlations: pushing the boundaries. *MNRAS*, 431(4):3307–3318, June 2013. doi: 10.1093/mnras/stt410.

V. Scottez, Y. Mellier, B. R. Granett, T. Moutard, M. Kilbinger, M. Scodeggio, B. Garilli, M. Bolzonella, S. de la Torre, L. Guzzo, U. Abbas, C. Adami, S. Arnouts, D. Bottini, E. Branchini, A. Cappi, O. Cucciati, I. Davidzon, A. Fritz, P. Franzetti, A. Iovino, J. Krywult, V. Le Brun, O. Le Fèvre, D. Maccagni, K. Małek, F. Marulli, M. Polletta, A. Pollo, L. A. M. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, J. Bel, J. Coupon, G. De Lucia, O. Ilbert, H. J. McCracken, and L. Moscardini. Clustering-based redshift estimation: application to VIPERS/CFHTLS. *MNRAS*, 462(2):1683–1696, October 2016. doi: 10.1093/mnras/stw1500.

V. Scottez, A. Benoit-Lévy, J. Coupon, O. Ilbert, and Y. Mellier. Testing the accuracy of clustering redshifts with simulations. *MNRAS*, 474(3):3921–3930, March 2018. doi: 10.1093/mnras/stx3056.

M. Seldner and P. J. E. Peebles. Statistical analysis of catalogs of extragalactic objects. XI. Evidence of correlation of QSOs and Lick galaxy counts. *ApJ*, 227:30–36, January 1979. doi: 10.1086/156699.

Ravi K. Sheth and Giuseppe Tormen. On the environmental dependence of halo formation. *MNRAS*, 350(4):1385–1390, June 2004. doi: 10.1111/j.1365-2966.2004.07733.x.

M. Shuntov, J. Pasquet, S. Arnouts, O. Ilbert, M. Treyer, E. Bertin, S. de la Torre, Y. Dubois, D. Fouchez, K. Kraljic, C. Laigle, C. Pichon, and D. Vibert. PhotoWeb redshift: boosting photometric redshift accuracy with large spectroscopic surveys. *A&A*, 636:A90, April 2020. doi: 10.1051/0004-6361/201937382.

Manodeep Sinha and Lehman H. Garrison. CORRFUNC - a suite of blazing fast correlation functions on the CPU. *MNRAS*, 491(2):3022–3041, Jan 2020a. doi: 10.1093/mnras/stz3157.

Manodeep Sinha and Lehman H. Garrison. CORRFUNC - a suite of blazing fast correlation functions on the CPU. *MNRAS*, 491(2):3022–3041, Jan 2020b. doi: 10.1093/mnras/stz3157.

Rachel S. Somerville and Romeel Davé. Physical Models of Galaxy Formation in a Cosmological Framework. *ARA&A*, 53:51–113, August 2015a. doi: 10.1146/annurev-astro-082812-140951.

Rachel S. Somerville and Romeel Davé. Physical Models of Galaxy Formation in a Cosmological Framework. *ARA&A*, 53:51–113, August 2015b. doi: 10.1146/annurev-astro-082812-140951.

Rachel S. Somerville and Tsafrir S. Kolatt. How to plant a merger tree. *MNRAS*, 305 (1):1–14, May 1999. doi: 10.1046/j.1365-8711.1999.02154.x.

Rachel S. Somerville and Joel R. Primack. Semi-analytic modelling of galaxy formation: the local Universe. *MNRAS*, 310(4):1087–1110, December 1999. doi: 10.1046/j.1365-8711.1999.03032.x.

Rachel S. Somerville, Philip F. Hopkins, Thomas J. Cox, Brant E. Robertson, and Lars Hernquist. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *MNRAS*, 391(2):481–506, December 2008. doi: 10.1111/j.1365-2966.2008.13805.x.

Rachel S. Somerville, Rudy C. Gilmore, Joel R. Primack, and Alberto Domínguez. Galaxy properties from the ultraviolet to the far-infrared: Λ cold dark matter models confront observations. *MNRAS*, 423(3):1992–2015, July 2012. doi: 10.1111/j.1365-2966.2012.20490.x.

Rachel S. Somerville, Gergö Popping, and Scott C. Trager. Star formation in semi-analytic galaxy formation models with multiphase gas. *MNRAS*, 453(4):4337–4367, November 2015. doi: 10.1093/mnras/stv1877.

Rachel S. Somerville, Charlotte Olsen, L. Y. Aaron Yung, Camilla Pacifici, Henry C. Ferguson, Peter Behroozi, Shannon Osborne, Risa H. Wechsler, Viraj Pandya, Sandra M. Faber, Joel R. Primack, and Avishai Dekel. Mock light-cones and theory friendly catalogues for the CANDELS survey. *MNRAS*, 502(4):4858–4876, April 2021. doi: 10.1093/mnras/stab231.

T. Sousbie. The persistent cosmic web and its filamentary structure - I. Theory and implementation. *MNRAS*, 414(1):350–383, June 2011. doi: 10.1111/j.1365-2966.2011.18394.x.

J. S. Speagle, C. L. Steinhardt, P. L. Capak, and J. D. Silverman. A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ˜ 0-6. *ApJS*, 214:15, October 2014. doi: 10.1088/0067-0049/214/2/15.

D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, B. Macintosh, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, D. Benford, M. Hudson, W. S. Jeong, Y. Mellier, W. Traub, T. Yamada, P. Capak, J. Colbert, D. Masters, M. Penny, D. Savransky, D. Stern, N. Zimmerman, R. Barry, L. Bartusek, K. Carpenter, E. Cheng, D. Content, F. Dekens, R. Demers, K. Grady, C. Jackson, G. Kuan, J. Kruk, M. Melton, B. Nemati, B. Parvin, I. Poberezhskiy, C. Peddie, J. Ruffa, J. K. Wallace, A. Whipple, E. Wollack, and F. Zhao. Wide-field infrarred survey telescope-astrophysics focused telescope assets wfirst-afta 2015 report, 2015.

Nicolas Tejos, Aldo Rodríguez-Puebla, and Joel R. Primack. Stochastic Order Redshift Technique (SORT): a simple, efficient and robust method to improve cosmological redshift measurements. *MNRAS*, 473(1):366–379, January 2018. doi: 10.1093/mnras/stx2317.

E. Tempel, T. Tuvikene, R. Kipper, and N. I. Libeskind. Merging groups and clusters of galaxies from the SDSS data. The catalogue of groups and potentially merging systems. *A&A*, 602:A100, June 2017. doi: 10.1051/0004-6361/201730499.

Jeremy Tinker, Andrew Wetzel, and Charlie Conroy. Are Halo and Galaxy Formation Histories Correlated? *arXiv e-prints*, art. arXiv:1107.5046, July 2011. doi: 10.48550/arXiv.1107.5046.

Jeremy L. Tinker, Alexie Leauthaud, Kevin Bundy, Matthew R. George, Peter Behroozi, Richard Massey, Jason Rhodes, and Risa H. Wechsler. Evolution of the Stellar-to-dark Matter Relation: Separating Star-forming and Passive Galaxies from z = 1 to 0. *ApJ*, 778(2):93, December 2013. doi: 10.1088/0004-637X/778/2/93.

Frank C. van den Bosch, Daniel Aquino, Xiaohu Yang, H. J. Mo, Anna Pasquali, Daniel H. McIntosh, Simone M. Weinmann, and Xi Kang. The importance of satellite quenching for the build-up of the red sequence of present-day galaxies. *MNRAS*, 387(1):79–91, June 2008. doi: 10.1111/j.1365-2966.2008.13230.x.

Frank C. van den Bosch, Johannes U. Lange, and Andrew R. Zentner. Basilisk: Bayesian hierarchical inference of the galaxy-halo connection using satellite kinematics - I. Method and validation. *MNRAS*, 488(4):4984–5013, October 2019. doi: 10.1093/mnras/stz2017.

Jesús Varela, Juan Betancort-Rijo, Ignacio Trujillo, and Elena Ricciardelli. The Orientation of Disk Galaxies around Large Cosmic Voids. *ApJ*, 744(2):82, January 2012. doi: 10.1088/0004-637X/744/2/82.

Kai Wang, Yangyao Chen, Qingyang Li, and Xiaohu Yang. Late-formed haloes prefer to host quiescent central galaxies - I. Observational results. *MNRAS*, 522(2):3188–3200, June 2023. doi: 10.1093/mnras/stad1175.

Douglas F. Watson, Andreas A. Berlind, Cameron K. McBride, David W. Hogg, and Tao Jiang. The Extreme Small Scales: Do Satellite Galaxies Trace Dark Matter? *ApJ*, 749(1):83, April 2012. doi: 10.1088/0004-637X/749/1/83.

J. R. Weaver, O. B. Kauffmann, O. Ilbert, H. J. McCracken, A. Moneti, S. Toft, G. Brammer, M. Shuntov, I. Davidzon, B. C. Hsieh, C. Laigle, A. Anastasiou, C. K. Jespersen, J. Vinther, P. Capak, C. M. Casey, C. J. R. McPartland, B. Milvang-Jensen, B. Mobasher, D. B. Sanders, L. Zalesky, S. Arnouts, H. Aussel, J. S. Dunlop, A. Faisst, M. Franx, L. J. Furtak, J. P. U. Fynbo, K. M. L. Gould, T. R. Greve, S. Gwyn, J. S. Kartaltepe, D. Kashino, A. M. Koekemoer, V. Kokorev, O. Le Fèvre, S. Lilly, D. Masters, G. Magdis, V. Mehta, Y. Peng, D. A. Riechers, M. Salvato, M. Sawicki, C. Scarlata, N. Scoville, R. Shirley, J. D. Silverman, A. Sneppen, V. Smolčić, C. Steinhardt, D. Stern, M. Tanaka, Y. Taniguchi, H. I. Teplitz, M. Vaccari, W. H. Wang, and G. Zamorani. COSMOS2020: A Panchromatic View of the Universe to z 10 from Two Complementary Catalogs. *ApJS*, 258(1):11, January 2022. doi: 10.3847/1538-4365/ac3078.

R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel. Concentrations of Dark Halos from Their Assembly Histories. *ApJ*, 568:52–70, March 2002. doi: 10.1086/338765.

Risa H. Wechsler and Jeremy L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. *ARA&A*, 56:435–487, September 2018a. doi: 10.1146/annurev-astro-081817-051756.

Risa H. Wechsler and Jeremy L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. *ARA&A*, 56:435–487, September 2018b. doi: 10.1146/annurev-astro-081817-051756.

Risa H. Wechsler, Andrew R. Zentner, James S. Bullock, Andrey V. Kravtsov, and Brandon Allgood. The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. *ApJ*, 652(1):71–84, November 2006. doi: 10.1086/507120.

Andrew R. Wetzel and Daisuke Nagai. The Physical Nature of the Cosmic Accretion of Baryons and Dark Matter into Halos and Their Galaxies. *ApJ*, 808(1):40, July 2015. doi: 10.1088/0004-637X/808/1/40.

Radosław Wojtak and Gary A. Mamon. Physical properties underlying observed kinematics of satellite galaxies. *MNRAS*, 428(3):2407–2417, January 2013. doi: 10.1093/mnras/sts203.

Joanna Woo, Avishai Dekel, S. M. Faber, Kai Noeske, David C. Koo, Brian F. Gerke, Michael C. Cooper, Samir Salim, Aaron A. Dutton, Jeffrey Newman, Benjamin J. Weiner, Kevin Bundy, Christopher N. A. Willmer, Marc Davis, and Renbin Yan. Dependence of galaxy quenching on halo mass and distance from its centre. *MNRAS*, 428(4):3306–3326, February 2013. doi: 10.1093/mnras/sts274.

Shengqi Yang, Rachel S. Somerville, Anthony R. Pullen, Gergö Popping, Patrick C. Breysse, and Abhishek S. Maniyar. Multitracer Cosmological Line Intensity Mapping Mock Light-cone Simulation. *ApJ*, 911(2):132, April 2021. doi: 10.3847/1538-4357/abec75.

Xiaohu Yang, H. J. Mo, Frank C. van den Bosch, Youcai Zhang, and Jiaxin Han. Evolution of the Galaxy-Dark Matter Connection and the Assembly of Galaxies in Dark Matter Halos. *ApJ*, 752(1):41, June 2012. doi: 10.1088/0004-637X/752/1/41.

Hassen M. Yesuf, Luis C. Ho, and S. M. Faber. What is Important? Morphological Asymmetries are Useful Predictors of Star Formation Rates of Star-forming Galaxies in SDSS Stripe 82. *ApJ*, 923(2):205, December 2021. doi: 10.3847/1538-4357/ac27a7.

L. Y. Aaron Yung, Rachel S. Somerville, Steven L. Finkelstein, Gergö Popping, and Romeel Davé. Semi-analytic forecasts for JWST - I. UV luminosity functions at z = 4-10. *MNRAS*, 483(3):2983–3006, March 2019a. doi: 10.1093/mnras/sty3241.

L. Y. Aaron Yung, Rachel S. Somerville, Gergö Popping, Steven L. Finkelstein, Harry C. Ferguson, and Romeel Davé. Semi-analytic forecasts for JWST - II. Physical properties and scaling relations for galaxies at z = 4-10. *MNRAS*, 490(2):2855–2879, December 2019b. doi: 10.1093/mnras/stz2755.

Idit Zehavi, Zheng Zheng, David H. Weinberg, Michael R. Blanton, Neta A. Bahcall, Andreas A. Berlind, Jon Brinkmann, Joshua A. Frieman, James E. Gunn, Robert H. Lupton, Robert C. Nichol, Will J. Percival, Donald P. Schneider, Ramin A. Skibba, Michael A. Strauss, Max Tegmark, and Donald G. York. Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *ApJ*, 736(1):59, July 2011. doi: 10.1088/0004-637X/736/1/59.