1-20-2024

Buzzard to Cardinal: Improved Mock Catalogs for Large Galaxy Surveys

Chun-Hao To
Ohio State University

Joseph DeRose
Lawrence Berkeley National Laboratory

Risa H. Wechsler
Kavli Institute for Particle Astrophysics and Cosmology

Eli Rykoff
Kavli Institute for Particle Astrophysics and Cosmology

Hao-Yi Wu
Boise State University

See next page for additional authors

Authors

Chun-Hao To, Joseph DeRose, Risa H. Wechsler, Eli Rykoff, Hao-Yi Wu, Susmita Adhikari, Elisabeth Krause, Eduardo Rozo, and David H. Weinberg

CrossMark

Buzzard to Cardinal: Improved Mock Catalogs for Large Galaxy Surveys

Chun-Hao To[1,2,3] , Joseph DeRose[4] , Risa H. Wechsler[5,6,7] , Eli Rykoff[5,7] , Hao-Yi Wu[8] , Susmita Adhikari[9] ,
Elisabeth Krause[10] , Eduardo Rozo[11] , and David H. Weinberg[1,3]

[1] Center for Cosmology and AstroParticle Physics (CCAPP), Ohio State University, Columbus, OH 43210, USA; to.87@osu.edu
[2] Department of Physics, Ohio State University, Columbus, OH 43210, USA
[3] Department of Astronomy, Ohio State University, Columbus, OH 43210, USA
[4] Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[5] Kavli Institute for Particle Astrophysics and Cosmology, Stanford, CA, USA
[6] Physics Department, Stanford University, Stanford, CA 94305, USA
[7] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[8] Department of Physics, Boise State University, Boise, ID 83725, USA
[9] Indian Institute of Science Education and Research Pune, Pune, Maharashtra, India
[10] Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
[11] Department of Physics, University of Arizona, Tucson, AZ 85721, USA

Abstract

We present the Cardinal mock galaxy catalogs, a new version of the Buzzard simulation that has been updated to support ongoing and future cosmological surveys, including the Dark Energy Survey (DES), DESI, and LSST. These catalogs are based on a one-quarter sky simulation populated with galaxies out to a redshift of $z = 2.35$ to a depth of $m_r = 27$. Compared to the Buzzard mocks, the Cardinal mocks include an updated subhalo abundance matching model that considers orphan galaxies and includes mass-dependent scatter between galaxy luminosity and halo properties. This model can simultaneously fit galaxy clustering and group–galaxy cross-correlations measured in three different luminosity threshold samples. The Cardinal mocks also feature a new color assignment model that can simultaneously fit color-dependent galaxy clustering in three different luminosity bins. We have developed an algorithm that uses photometric data to further improve the color assignment model and have also developed a novel method to improve small-scale lensing below the ray-tracing resolution. These improvements enable the Cardinal mocks to accurately reproduce the abundance of galaxy clusters and the properties of lens galaxies in the DES data. As such, these simulations will be a valuable tool for future cosmological analyses based on large sky surveys.

Unified Astronomy Thesaurus concepts: Cosmology (343); Dark matter (353); N-body simulations (1083); Large-scale structure of the universe (902); Galaxy dark matter halos (1880); Galaxy evolution (594)

1. Introduction

Over the past two decades, large galaxy surveys have systematically mapped hundreds of millions of galaxies with unprecedented precision, allowing us to establish the standard cosmological model that describes the universe's evolution over billions of years. However, analyzing these data to their full potential requires advanced theoretical models and excellent control of systematics. Achieving these requirements is challenging because most of the information lies in the scales where the theory is highly nonperturbative. Furthermore, one galaxy survey can be analyzed with multiple cosmological probes, which could share the same sources of systematics. Consistently modeling these systematics with different cosmological probes is essential to yielding unbiased cosmological constraints. Finally, developing accurate theoretical models is more challenging when considering blind analyses, in which the data is transformed to obscure actual cosmological signals during the development of the models.

Synthetic sky catalogs, also known as mock catalogs or mocks, provide a valuable tool for quantifying systematics and developing analysis techniques. They consist of plausible universes that can serve as a sandbox for researchers to test and develop methods for analyzing survey data. Accomplishing this task places several requirements on the synthetic catalogs. First, one would want to use these catalogs to control systematics so that they are much smaller than the statistical uncertainties of the data. Therefore, the volume of the mocks has to be larger (ideally, much larger) than the volumes probed by the targeted surveys. Second, the galaxies in the mocks should be realistic, although the level of realism required depends on the specific survey and analysis techniques used. Third, a fast generation of new mocks is desirable. When analyzing survey data, new techniques might be developed, and new systematics might be found. These developments might require new mocks that meet newly defined requirements. Further, fast mock generation allows the creation of various plausible realizations, allowing one to marginalize over uncertain physical processes.

Many techniques have been developed over the past two decades to generate synthetic catalogs (see, e.g., Wechsler & Tinker 2018, for a review). Ideally, one would want to simulate galaxies directly from numerical solutions of coupled dark matter and baryon evolution to generate a realistic galaxy catalog. Unfortunately, while significant progress has been made over the past two decades, this method is still too computationally demanding to produce synthetic galaxy catalogs larger than the volume observed by galaxy surveys (see, e.g., Vogelsberger et al. 2020, for a review). On the other

hand, several practical alternative methods have been developed to simulate galaxy formation processes using phenomenological models. Ordered from the least computationally demanding to the most computationally demanding, these alternative methods include:

1. The halo occupation model (HOD; Berlind & Weinberg 2002; Zheng et al. 2005), where one adopts phenomenological models to describe the statistical relations of galaxy properties and properties of the largest dark matter halos hosting these galaxies;
2. The subhalo abundance matching (SHAM) model (Kravtsov et al. 2004; Conroy et al. 2006), where one relates galaxy properties to subhalo properties via simple rankings;
3. Semi-analytic models (SAMs; see, e.g., Baugh 2006; Somerville & Davé 2015, for reviews), where one simulates galaxy formation physics using analytical prescriptions and integrates galaxy properties through halo merger histories.

Combinations of these alternative methods have led to a proliferation of synthetic catalogs that meet the aforementioned requirements for galaxy survey cosmology analyses (e.g., Fosalba et al. 2015; Potter et al. 2017; DeRose et al. 2019; Korytov et al. 2019). The most direct predecessor of this work is the Buzzard simulations (DeRose et al. 2019), which combine SHAM (Lehmann et al. 2017) and low-resolution large-volume N-body simulations using a machine-learning-based technique (ADDGALS; Wechsler et al. 2022). The Buzzard simulations produce realistic galaxy properties, allowing one to run target selections (such as redMaGiC galaxy selections, Rozo et al. 2016) on the simulations in the same way as survey data. Further, the simulations are relatively computationally inexpensive, making it possible to generate large numbers of realizations of survey data. Because of these features, the Buzzard simulations have facilitated end-to-end validations of cosmological analysis pipelines from main galaxy catalogs to cosmological constraints (e.g., MacCrann et al. 2018; To et al. 2021b; Chen et al. 2022; DeRose et al. 2022b; White et al. 2022).

While the Buzzard simulations and other catalogs built with the ADDGALS[12] algorithm have facilitated analyses in multiple cosmological surveys (see Wechsler et al. 2022, and references therein), the amplitude of galaxy clustering on scales less than $1\,h^{-1}$ Mpc in this simulation is smaller than the Sloan Digital Sky Survey (SDSS) measurements up to 50%. This suppressed galaxy clustering significantly impacts the properties of optically selected clusters, where one relies on the overdensity of red-sequence galaxies at $<1\,h^{-1}$ Mpc scales to identify galaxy clusters. Specifically, the number of redMaPPer galaxy clusters at a given richness (λ) in the Buzzard simulation is a factor of 3–4 smaller than the observed value in the DES Y1 data (DeRose et al. 2019; Abbott et al. 2020). The lack of a redMaPPer cluster problem is not unique in Buzzard simulations. CosmoDC2 (Korytov et al. 2019), the only alternative simulation that currently has been tested with

its redMaPPer cluster properties, also underpredicts the richness of optically selected clusters unless one artificially boosts red-sequence galaxies in cluster environments. While the additional boosting solves the lack of clusters problem (Korytov et al. 2019), it creates discontinuities between galaxy properties in cluster environments and in the field.

The unrealistic galaxy population in cluster environments is a critical limitation for studies that use these simulations to quantify the performance of optically selected clusters (Shin et al. 2019; Abbott et al. 2020; Myles et al. 2021b; To et al. 2021b; Wu et al. 2022; Zhang et al. 2023). This shortcoming can be partly mitigated by abundance matching: one compares the Nth richest clusters in the data to those in the simulations. The abundance matching technique makes the comparison of simulations and data sensitive only to the rank of richness instead of its absolute value, thereby reducing the problem of mismatched galaxy abundances in clusters. However, these effects cannot be calibrated reliably from simulations in which cluster richnesses are a factor of 2 lower than observed values.

The exact reason for this lack of cluster galaxies in the Buzzard simulation was previously unknown. DeRose et al. (2019) and Wechsler et al. (2022) hypothesize that it arises from artificial disruptions of subhalos that experience close pericentric passages (van den Bosch & Ogiya 2018). Following this line, DeRose et al. (2021) developed a new SHAM model that includes orphan galaxies to address this problem. However, while the SHAM model with orphan galaxies can fit the galaxy clustering measured in SDSS in several stellar mass bins, no model was found to fit all three stellar mass bins considered in that work simultaneously. Further, DeRose et al. (2021) also found that the color assignment model used in Wechsler et al. (2022) and DeRose et al. (2022b) can lead to an underestimation of red galaxy clustering in the lowest stellar mass bins ($\log M^* \in$ [9.8, 10.2] $h^{-1}\,M_\odot$). Because red galaxies more likely reside in cluster environments, this reduced red galaxy clustering can also lead to a lack of cluster galaxies.

In this paper, we solve the lack of cluster galaxies problem in the Buzzard simulation by quantifying and addressing both contributing factors: (1) artificial subhalo disruption in the SHAM model and (2) the color assignment model. Our new model can simultaneously fit galaxy clustering and group–galaxy cross-correlations measured at three luminosity thresholds and also fit color-dependent galaxy clustering. We propagate this model through the ADDGALS algorithm (Wechsler et al. 2022) and generate the Cardinal simulations. Figure 1 shows a flowchart that summarizes the key steps in generating Cardinal. A list of improvements from Buzzard v2.0 to Cardinal is also presented in Appendix K. Finally, we compare the properties of redMaGiC galaxies and redMaPPer clusters in Cardinal and Dark Energy Survey (DES) Y3 data (Sevilla-Noarbe et al. 2021) and find excellent agreement.

This paper is organized as follows. In Section 2, we detail the construction of the new SHAM models that Cardinal is based on. In Section 3, we detail the steps in generating Cardinal using the new SHAM model. Specifically, the improved color assignment method is presented in Section 3.4. In Section 3.6, we address the remaining problems in our color assignment models, including the lack of redshift evolution in training spectra and the inadequacy of summarizing colors using current spectral energy distribution (SED) templates. In Section 4, we compare the properties of redMaGiC galaxies and redMaPPer clusters in Cardinal

---

[12] We note that there are two existing versions of Buzzard: the Buzzard Flock containing 18 DES Y1 realizations, and Buzzard v2.0, containing 18 DES Y3 realizations. These catalogs are based on 18 realizations of the Chinchilla N-body simulations, each of which is generated from a different random seed. In the remainder of this paper, we use "the Buzzard simulation" to refer to one of the Buzzard v2.0 mocks that shares the same N-body simulation as the Cardinal.
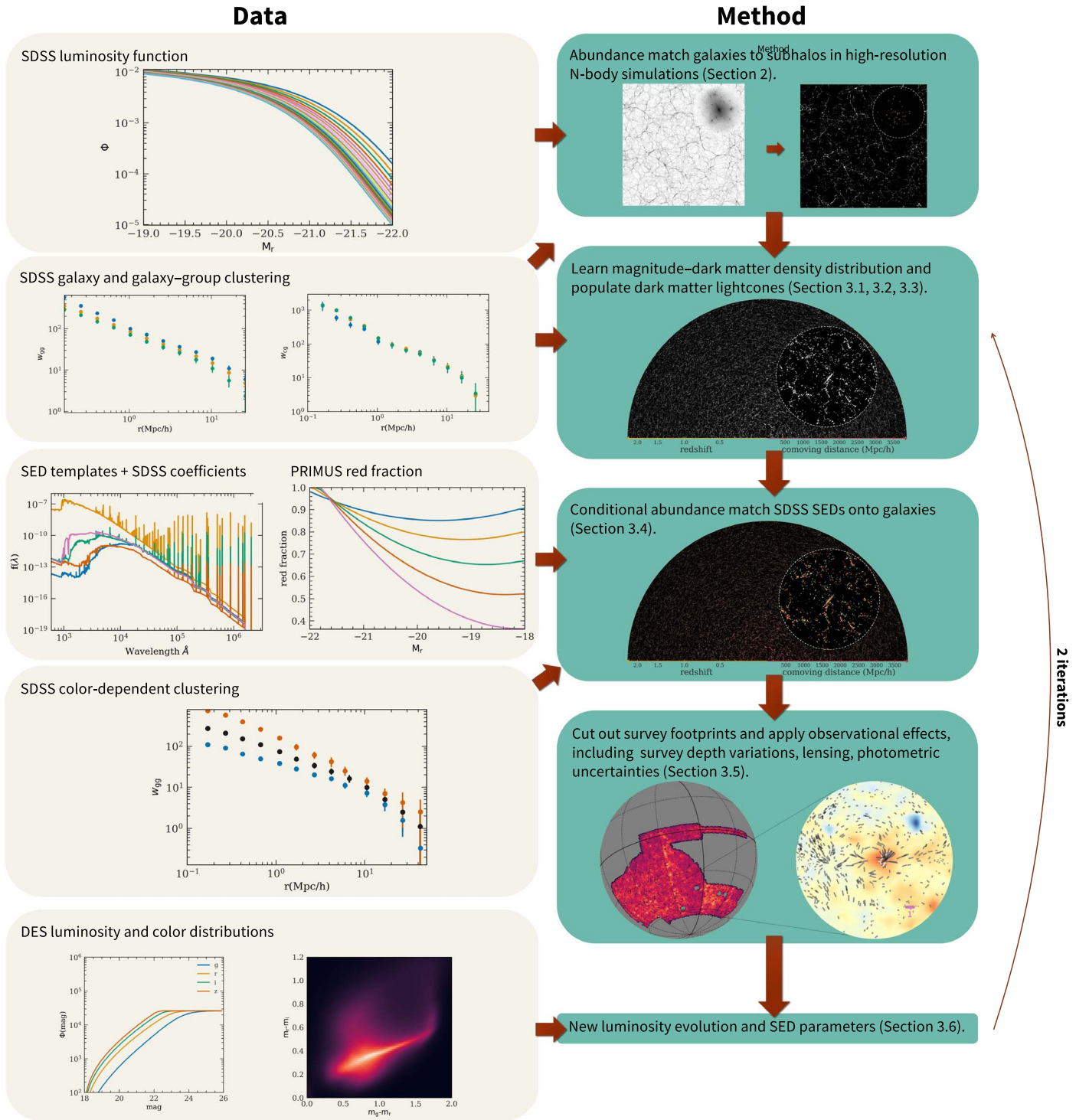
Figure 1. Flowchart of the algorithm. The left column shows the observational inputs, and the right column shows the algorithm. The algorithm can be broadly categorized into five steps. First, we develop an extended SHAM model to populate galaxies on *N*-body simulations with resolved structure (Section 2). Second, we measure the statistical relations of galaxies' luminosities and local density tracer R_δ in the SHAM galaxy catalogs. We then paint luminosity onto particles of a light-cone simulation using the measured relation (Section 3.3). Third, we use conditional abundance matching techniques to assign colors to galaxies in the mock using observed galaxy colors in SDSS (Section 3.4). Fourth, we apply observational effects in the mocks, including lensing, photometric uncertainties, and varying survey depths (Section 3.5). Finally, we use DES's photometric galaxy catalog to fine-tune the galaxy colors (Section 3.6).

and DES Y3 data. Finally, we conclude in Section 5 with a discussion of future improvements.

2. A SHAM-Based Galaxy–Halo Connection Model

We use a modified `SHAM` algorithm to construct the training data for the galaxy–halo connection model. We describe the

data and simulations used to construct the SHAM model below.

2.1. Calibrating Data

We use the NYU Value-Added Galaxy Catalog (VAGC; Blanton et al. 2005) constructed from the SDSS DR7

<div align="center">Table 1</div>

Descriptions of the Simulations Used for This Analysis, Including the Size of the Box ($L_{\rm box}$), the Mass Resolution of the Particles ($M_{\rm part}$), and the Force Softening Length ($\epsilon_{\rm Plummer}$)

Name	$L_{\rm box}$	$M_{\rm part}$	$\epsilon_{\rm Plummer}$
SHAM galaxy catalogs generation and tests			
Chinchilla-T1	$400\,h^{-1}$ Mpc	$5.9 \times 10^{8}\,h^{-1}\,M_{\odot}$	$5.5\,h^{-1}$ kpc
SMDPL	$400\,h^{-1}$ Mpc	$9.6 \times 10^{7}\,h^{-1}\,M_{\odot}$	$1.5\,h^{-1}$ kpc
Light-cone mock generation			
L1 ($z < 0.315$)	$1.05\,h^{-1}$Gpc	$3.3 \times 10^{10}\,h^{-1}\,M_{\odot}$	$20\,h^{-1}$ kpc
L2 ($0.315 < z < 0.955$)	$2.6\,h^{-1}$Gpc	$1.6 \times 10^{11}h^{-1}\,M_{\odot}$	$35\,h^{-1}$ kpc
L3 ($0.955 < z < 2.35$)	$4\,h^{-1}$Gpc	$5.9 \times 10^{11}\,h^{-1}\,M_{\odot}$	$53\,h^{-1}$ kpc
Cosmological parameters			
$\Omega_m = 0.286,\ \Omega_b = 0.047,\ \sigma_8 = 0.82,\ n_{\rm s} = 0.96,\ h = 0.7,\ A_s = 2.145 \times 10^{-9}$			

(Abazajian et al. 2009) main galaxy catalog to constrain the parameters in the SHAM model. We consider three volume-limited galaxy samples: $M_r < -19$, $M_r < -20$, and $M_r < -21$ with $0.026 < z < 0.067$, $0.026 < z < 0.106$, and $0.026 < z < 0.106$, respectively. We limit our analysis to the north galactic cap to avoid modeling differences in target selection between the north and south galactic caps. From these three volume-limited samples, we measure the projected correlation function given by

$$w_{\rm gg}(r) = 2 \int_0^{\pi_{\rm max}} \xi(r, \pi)\, d\pi, \qquad (1)$$

where π is the line-of-sight distance between pairs, r is the distance between pairs perpendicular to the line of sight, and $\pi_{\rm max}$ is $40\,h^{-1}$ Mpc. We measure $\xi(r, \pi)$ using the Landy–Szalay estimator (Landy & Szalay 1993), with 12 logarithmically spaced bins between $r = 0.13$ and $r = 32.6\,h^{-1}$ Mpc and 40 linearly spaced bins in π. The smallest scale is chosen to avoid systematics caused by the size of the SDSS fibers.

In addition to galaxy clustering ($w_{\rm gg}$), we also use the cross-correlation between galaxy groups and galaxies ($w_{\rm cg}$) to constrain the SHAM model. Berlind & Weinberg (2002) suggest that the group multiplicity function provides complementary information on galaxy–halo connections relative to galaxy clustering (see also Sinha et al. 2018). Given the number densities of galaxies, group multiplicity functions are simply integrations of galaxy–galaxy group cross-correlations across spatial separations. Therefore, we include the galaxy–galaxy group cross-correlations to better constrain SHAM parameters important to galaxy occupations in cluster environments. We first measure galaxy groups in the VAGC catalogs using the Self-Calibrated Galaxy Group Finder (Tinker 2020, 2021). We restrict the measurement to galaxies with $0.026 < z < 0.067$ and $M_r < -19$ to ensure the galaxy groups are complete. We remove all color information used in the Self-Calibrated Galaxy Group Finder because this information does not exist in the mock galaxy catalogs constructed from the SHAM model. This might degrade the performance of the group finders, but it allows an apples-to-apples comparison between the measurements in data and mocks. With the group catalogs, we select groups with mass greater than $5 \times 10^{13}\,h^{-1}{\rm M}_{\odot}$, the mass range of $\lambda > 20$ redMaPPer clusters. We then cross-correlate the centers of these groups with galaxies at $0.026 < z < 0.067$ and $M_r > -19$, $M_r > -20$, and $M_r > -21$ to obtain $\xi_{\rm cg}(r, \pi)$, which we integrate using Equation (1) to get $w_{\rm cg}(r)$.

We measure the covariance matrix of the data using the jackknife resampling technique. We first employ a K-means

algorithm implemented in TreeCorr (Jarvis 2015) on random points to separate the sky into 128 uniform patches. We then perform the measurements by excluding one patch at a time. The covariance matrix is then estimated as

$$\mathrm{cov}\,(d_i,\ d_j) = \frac{N-1}{N} \sum_{k=1}^{k=N} (d_{i,k} - \langle x_i \rangle)(d_{j,k} - \langle x_j \rangle), \qquad (2)$$

where d_i is the ith element of the data vector, N is 128, $d_{i,k}$ is the ith element of the data from measurements that exclude the kth patch, and $\langle x_i \rangle$ is the mean of the element over $N = 128$ patches. Jackknife estimations introduce noise into the covariance matrix, which leads to biases in the inversion of the covariance. Thus, one has to regularize the covariance matrix before inverting it. Here, we adopt an approach similar to that in Behroozi et al. (2019). In short, we perform an eigenvalue decomposition of the jackknife-estimated covariance matrix to obtain eigenvalues D_n and the associated eigenvectors v_n. Due to various possible sources of noise (such as variations in sky backgrounds, variations in fiber assignment efficiency, and systematics in galaxy photometry), we do not expect the error estimated to be better than 10%. We therefore rank order the eigenvalues and find the eigenvalues whose square rooted values are below 0.1 of the data projected onto the eigenspace. We then replace those eigenvalues with 0.1 of the data projected onto the eigenspace and multiply the new eigenvalues with the eigenvectors to form a regularized covariance matrix.

2.2. Simulations and Models

2.2.1. Simulations

To generate the training model, we use the Chinchilla-T1 dark matter simulation (see Table 1 for a summary), which has a volume of $(400\,h^{-1}\,{\rm Mpc})^3$ with particle resolution $5.9 \times 10^8 h^{-1}\,M_{\odot}$. The simulation is generated using L-GADGET2 (Springel 2005) with a ΛCDM cosmology that has $\Omega_m = 0.286$, $\Omega_b = 0.047$, $\sigma_8 = 0.82$, $n_s = 0.96$, $h = 0.7$, and three massless neutrino species with $N_{\rm eff} = 3.046$. Halo finding is performed using ROCKSTAR (Behroozi et al. 2013), and merger trees were generated using Consistent Tree (Behroozi et al. 2012). We refer to Wechsler et al. (2022) for the details of this simulation. Given that our galaxy samples are selected at low redshift, we use the snapshot corresponding to $z = 0$ for the work described in this section.

Evidence has shown that subhalos in dark matter-only simulations are susceptible to physical and unphysical

disruptions (Klypin et al. 1999; Weinberg et al. 2008; van den Bosch & Ogiya 2018). We account for this effect using a prescription similar to that in Behroozi et al. (2019), which adds disrupted subhalos back into the simulation. We first identify subhalos that are no longer detected by ROCKSTAR in each snapshot. Then, we locate the host halos that contain these subhalos within their virial radius and use the SAM from Behroozi et al. (2019) to simulate the evolution of the subhalos' position, mass, and maximal circular velocity. These procedures produce a catalog containing standard halos that can be found and tracked using ROCKSTAR and disrupted subhalos (also known as orphans) that have properties calculated semi-analytically.

2.2.2. Models

Using the SHAM technique, we populate subhalos with galaxies (tracked subhalos and orphans). Here, we first describe the general concept of this technique and then describe the extensions we develop in this paper. In the most basic form, SHAM assigns each subhalo with a luminosity by enforcing the relation

$$n(L > x) = n(X_h > y), \quad (3)$$

where $n(L > x)$ represents the number density of galaxies with luminosity greater than x and $n(X_h > y)$ represents the number density of subhalos with properties X_h greater than y. Following Lehmann et al. (2017), we adopt X_h as v_α defined as

$$v_\alpha = v_{\rm vir} \left(\frac{v_{\rm max}}{v_{\rm vir}} \right)^\alpha, \quad (4)$$

where $v_{\rm vir}$ is the virial velocity of the halos, $v_{\rm max}$ is the maximum circular velocity, and α is a free parameter. These quantities are evaluated at the epoch when the halo's mass is at the maximum to avoid complicated physics subhalos experienced when falling into a big halo. The free parameter α allows additional flexibility for the SHAM model. In particular, $\left(\frac{v_{\rm max}}{v_{\rm vir}} \right)$ can be viewed as a proxy of halo concentration. The free parameter α controls the dependence of galaxy luminosity on halo concentrations.

The $n(L > x)$ in Equation (3) is estimated by fitting a modified double-Schechter function with a Gaussian tail to the luminosity function measured in the SDSS DR7 galaxy catalogs. For details on constructing $n(L > x)$, we refer the readers to Appendix E.1 of DeRose et al. (2019). We cannot directly apply the estimated $n(L > x)$ in Equation (3) because the relations between galaxy luminosity and halo properties are stochastic. This stochasticity comes from observational uncertainties, complicated astrophysical processes that affect galaxy evolution within dark matter halos, and additional halo properties that are correlated with galaxy luminosities. One can model these complicated processes as

$$n(L > x) = \int P(L|L') n(L' > x) dL', \quad (5)$$

where $n(L > x)$ is the measured luminosity function. The intrinsic galaxy luminosity L' in the above expression is determined solely by the selected halo properties X_h by comparing the number density of galaxies with luminosity above L', $n(x > L')$, to the number density of halos with the selected property above X_h, $n(x > X_h)$. In most of the SHAM

works (e.g., Reddick et al. 2013; Lehmann et al. 2017; Contreras et al. 2021; DeRose et al. 2021), one parametrizes $P(L|L')$ as a log-normal distribution with mean $L = L'$ and scatter σ. With this assumption, Equation (5) can be viewed as a convolution problem and can be solved using standard deconvolution algorithms. However, numerous lines of evidence based on observations and simulations have shown that the scatter in $P(L|L')$ might depend on halo mass (see Wechsler & Tinker 2018, for a review). Wechsler & Tinker (2018) show that the value of this scatter is a constant at the high-mass end and increases at the low-mass end. We therefore parameterize $P(L|L')$ as a log-normal distribution with scatter

$$\sigma(L') = \max(\sigma_v + \sigma_{\rm vs}(-2.5 \log(L') + \sigma_{\rm vp}), \sigma_v), \quad (6)$$

where σ_v, $\sigma_{\rm vs}$, and $\sigma_{\rm vp}$ are free parameters, and $\max(A, B)$ represents the maximum of A and B. Further, observational constraints based on galaxy groups (Tinker 2021) and satellite kinematics (Lange et al. 2019) constrain this scatter to be less than 0.4 with 95% confidence over the range of luminosities considered here. We, therefore, set an additional prior on $\sigma(L')$ to be below 0.4. With a functional form of $P(L|L')$, we then estimate $n(L' > x)$ based on the measured $n(L > x)$ using the algorithm presented in Appendix A.

When subhalos (tracked subhalos and orphans) fall onto big halos, they might be tidally disrupted, the gas within them could be stripped, and the galaxy living inside them might be destroyed. Thus, we must allow additional flexibility in the model to capture these physical processes. Previous work has been done on mitigating these issues by applying cuts in halo properties on tracked subhalos (Reddick et al. 2013), orphans (Behroozi et al. 2019; DeRose et al. 2021), or both (Contreras et al. 2021). In this paper, we treat tracked subhalos and orphans equally. This approach has the benefit that the result depends less on the resolution of the simulations. We mitigate the issues of physical disruptions using a procedure similar to those in Behroozi et al. (2019) and DeRose et al. (2021). For each subhalo, we compare the maximum circular velocity at the current time ($v_{\rm max,now}$) to the maximum circular velocity at the time when the halo mass is at the maximum ($v_{{\rm max},M_{\rm peak}}$). When $v_{\rm max,now}$ of a subhalo is much smaller than $v_{{\rm max},M_{\rm peak}}$, the subhalo is likely tidally stripped and is less likely to host a galaxy. We therefore set a threshold of the ratio between $v_{\rm max,now}$ and $v_{{\rm max},M_{\rm peak}}$, below which the subhalos do not host galaxies. Further, because galaxies with different luminosities will have different timescales of dynamical frictions and resilience to tidal disruptions, we allow this threshold to depend on v_α, the combination of halo properties used in SHAM. With these insights, our parameterization of the probability that a subhalo is physically disrupted reads as

$$P({\rm disrupt}) = \Theta \left(T_{\rm dis}(v_\alpha) - \frac{v_{\rm max,now}}{v_{{\rm max},M_{\rm peak}}} \right), \quad (7)$$

$$T_{\rm dis}(v_\alpha) = T_l + (T_h - T_l) \left(0.5 + 0.5 \, {\rm erf} \left(\frac{\log_{10} v_\alpha - v_m}{\sqrt{2}\, \sigma_d} \right) \right), \quad (8)$$

where Θ is the Heaviside step function. In the above expression, T_l and T_h are free parameters, with $T_{\rm dis}(v_\alpha)$ interpolating between asymptotic behaviors at the high and low

Table 2
Parameters and Priors of the Extended SHAM Model Described in Section 2.2.2

Parameter	Prior	Best-fit Value	Description	Relevant Equations
			Halo properties for SHAM	
α	flat(0.0, 1.0)	$0.16^{+0.12}_{-0.26}$	Parameters in SHAM that control the dependence of galaxy luminosity on halo concentration	Equation (4)
			Scatter in luminosity–halo mass relation	
σ_v	flat(0.1, 0.25)	0.20 ± 0.017	Scatter in luminosity–halo mass relation at $M_r = \sigma_{\rm vp}$	
$\sigma_{\rm cs}$	flat(-0.01, 0.3)	$0.17^{+0.05}_{-0.06}$	Luminosity dependence of the scatter	Equation (6)
$\sigma_{\rm vp}$	flat(-20.5, -20.0)	$-20.39^{+0.07}_{-0.16}$	Pivot point of the scatter	
			Subhalo disruption	
T_l	flat(0.01, 1.0)	0.01 ± 0.082	Asymptotic value of $T_{\rm dis}$ at $v_\alpha = 0$	Equations (7) and (8)
T_h	flat(0.2, 1.6)	$0.78^{+0.17}_{-0.27}$	Asymptotic value of $T_{\rm dis}$ at $v_\alpha = \infty$	
v_m	flat(1.9, 3.3)	$2.00^{+0.23}_{-0.16}$	Value of $\log v_\alpha (km/s)$ when $T_{\rm dis} = 0.5(T_l + T_h)$	
σ_d	flat(0.1, 1.0)	$0.47^{+0.17}_{-0.33}$	Steepness of the transition	

v_α ends, v_m governs where the transition from T_l to T_h occurs, and σ_d controls how steep the transition is. Although Equations (7) and (8) may appear complex, the underlying physical behavior is simple. For halos with large v_α, T_h sets the threshold of $v_{\rm max,now}/v_{{\rm max},M_{\rm peak}}$ to determine whether the halos are tidally disrupted. Conversely, for halos with small v_α, T_l determines the threshold. Equation (8) ensures a smooth transition between small and large v_α values, and two additional parameters control the location and slope of the transition.

To summarize, our extended SHAM model has eight free parameters, whose values are given in Table 2. Given these parameters, we can generate a galaxy catalog using the following procedure. We first employ Equation (4) to calculate the v_α of each subhalo in a halo catalog, including tracked subhalos and halos identified by ROCKSTAR and orphan subhalos generated using a SAM from Behroozi et al. (2019). We then remove subhalos according to Equations (7) and (8). For each of the remaining subhalos, we populate a galaxy at the position of the subhalo with luminosity determined by Equation (5).

2.2.3. Measurements in Simulations

We select galaxies in simulations using the same luminosity cut as the data. To account for the effect of redshift-space distortions, we first transform the coordinates of halos along the line of sight using the velocity of the halos. We then apply the periodic boundary condition on the transformed coordinates and calculate $w_{\rm gg}(r)$ using the same radial binning as the data. Finally, the $w_{\rm gg}(r)$ is estimated using the natural estimator, with an analytic calculation of the random–random pairs. To minimize cosmic variance, we repeat the above process by choosing the line-of-sight direction along the x, y, and z axes of the box and taking the average of the measurements. Further, we minimize the stochasticities due to the Monte Carlo nature of the SHAM model by repopulating the simulations 10 times with the same parameters.

Because the simulation box size is small, cosmic variance cannot be ignored in the total error budget. We estimate this cosmic variance using the jackknife technique similar to the procedure described in Section 2.1. The main difference is that there is no survey mask in the simulations, so we generate 125

jackknife subsamples by equally partitioning the box. Further, the jackknife subsampling breaks the periodic boundary conditions. One has to consider this when analytically estimating the random–random terms in the natural estimator. Here, we adopt the approach described in He (2021) to estimate the random–random terms for each jackknife subsample. We estimate the covariance matrix from this cosmic variance using the best-fit parameters to the data presented in Section 2.3. This covariance matrix is then added to the observed covariance matrix while performing the likelihood inferences.

Regarding the galaxy group samples, we run the Self-Calibrated Galaxy Group Finder (Tinker 2020, 2021) on the simulated galaxies with the same setting as run on SDSS data. As pointed out before, in both runs on simulations and data, we remove steps that use color information to ensure an apples-to-apples comparison.

2.3. SHAM Results and Discussion

We fit the model described in Section 2.2.2 to the data described in Section 2.1 assuming a Gaussian likelihood with the priors described in Table 2. The challenge is that the fitting is hugely time-consuming because each step of the likelihood inference requires populating halos 10 times and computing the correlation function 30 times. We tackle this challenge by first building an emulator of the SHAM model and using it for likelihood inferences. We detail the procedure for constructing this emulator in Appendix B. Then, the likelihood inference is performed using an ensemble slicing sampling method implemented in ZEUS (Karamanis & Beutler 2020; Karamanis et al. 2021).

Figure 2 compares the best-fit model and the data. The minimum χ^2 is 79.4 with a degree of freedom of 64, yielding a probability to exceed (PTE) of 0.09. Table 2 shows the best-fit parameters. Overall, our model describes the data well, but a small difference can be seen at $r < 0.7\,h^{-1}$ Mpc of $w_{\rm cg}$ for the faintest galaxy sample. We want to determine whether the difference we have observed is due to limitations of the N-body simulations, such as force softening or finite resolution. To do this, we repeat the entire SHAM analysis using the Small MultiDark Planck (SMDPL) N-body simulation (Klypin et al. 2016), which has a higher resolution, a different softening scale, and a different fiducial cosmology than the previous
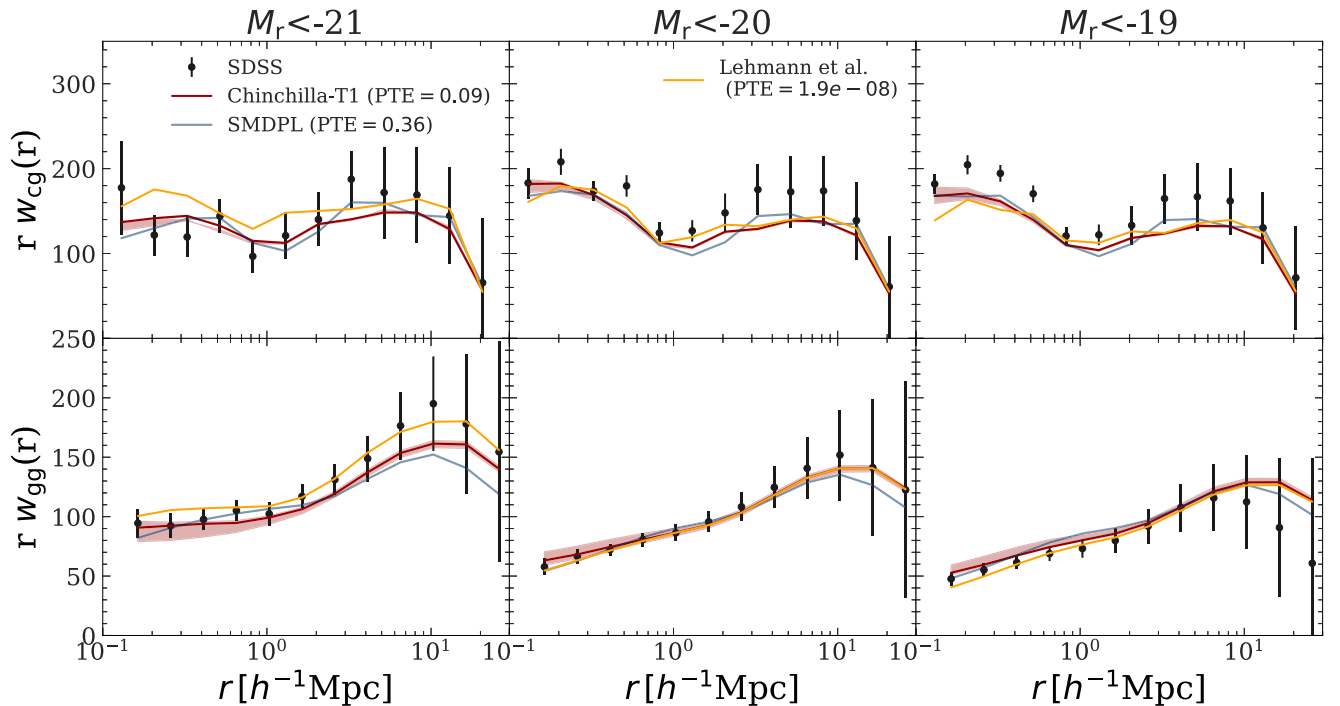
Figure 2. Comparison of the data and the best-fit SHAM model. Different columns correspond to different magnitude cuts in galaxies. The top row shows galaxy–galaxy group cross-correlations and the bottom shows galaxy autocorrelations. Black dots are measurements using SDSS data, while error bars show the 1σ uncertainties. Red lines show the best-fit SHAM models based on the Chinchilla-T1 simulation. Shaded regions show 1σ uncertainties. The blue line shows the best-fit SHAM models based on the SMDPL simulations. For comparison, orange lines show the predictions of Lehmann et al. (2017)'s best-fit model, which does not consider mass-dependent scatter in the luminosity–halo mass relations and orphan subhalos. The PTE value in the legend corresponds to the PTE.

simulation. This difference persists, as shown in the blue line in Figure 2. One possibility is that this difference comes from tensions between w_{cg} and w_{gg}. Small-scale w_{gg} is dominated by the one-halo term, which has significant contributions from galaxies in high-mass halos. In Figure 2, one can see that the small-scale w_{gg} for the faintest galaxy samples prefers a slightly smaller one-halo term than the best-fit SHAM model (shown as the red line), while the small-scale w_{cg} prefers the opposite. This mild tension might be related to the findings in Hearin et al. (2013), where they find tensions between group multiplicity functions, which are an integrated version of w_{cg}, and galaxy clustering, under the assumption of the basic SHAM model. This tension is much smaller in our extended SHAM model, and the constraining power of the data cannot distinguish it from statistical fluctuations. We therefore leave further investigations to future work.

Figure 3 shows the posteriors of SHAM parameter constraints based on Chinchilla-T1 (red) and SMDPL (blue). Most of the SHAM parameters based on Chinchilla-T1 and SMDPL are consistent, indicating the robustness of the result to details of N-body simulations, including cosmology, resolution, and force softening. The only parameter that is slightly inconsistent is σ_{vp}. Chinchilla-T1 (the lower resolution simulations) prefers a brighter (more negative) value than SMDPL (the higher resolution simulations). One possible explanation is that SMDPL has more low-mass halos than Chinchilla-T1. For a given magnitude bin, the missing low-mass halos in Chinchilla-T1 can be compensated by a larger scatter. Thus, Chinchilla-T1 has a brighter pivot point for the scatter. In both Chinchilla-T1 and SMDPL, the inclusion of orphan galaxies is strongly preferred. For the galaxy sample considered in this work, ~20% are orphan galaxies for Chinchilla-T1 and ~15% for SMDPL. Another interesting

result is that both Chinchilla-T1 and SMDPL prefer a varying scatter in the luminosity–v_α relation at the 1σ level. The positive value of σ_{cs} indicates that the scatter increases for lower-mass halos. This is consistent with results based on group finders (Tinker 2021), satellite kinematics (Lange et al. 2019), and galaxy clustering (Xu et al. 2018).

3. Populating Galaxies in Low-resolution Simulations

The SHAM model presented in Section 2 allows us to create high-fidelity galaxy catalogs based on high-resolution simulations. However, the high-resolution simulations typically have a volume much smaller than the volume accessible with current and upcoming galaxy surveys, making them insufficient to validate models with the required accuracy. In this section, we describe the formalism to transfer the knowledge learned in the SHAM catalogs to populate galaxies in large simulations with low resolution. This way, one can generate multiple realizations of mock galaxy catalogs with a modest computational expense.

3.1. Simulations

We use light cones with an area of 10,313 deg² constructed from the L1, L2, and L3 simulations detailed in Table 1. These simulations are generated with the same cosmological parameters as Chinchilla-T1. Details of the light-cone construction are presented in Appendix B1 of DeRose et al. (2019).

3.2. Targets

We aim to generate a mock that supports sciences in large galaxy surveys, such as DES and Vera Rubin Observatories' Legacy Survey of Space and Time (LSST). The various science
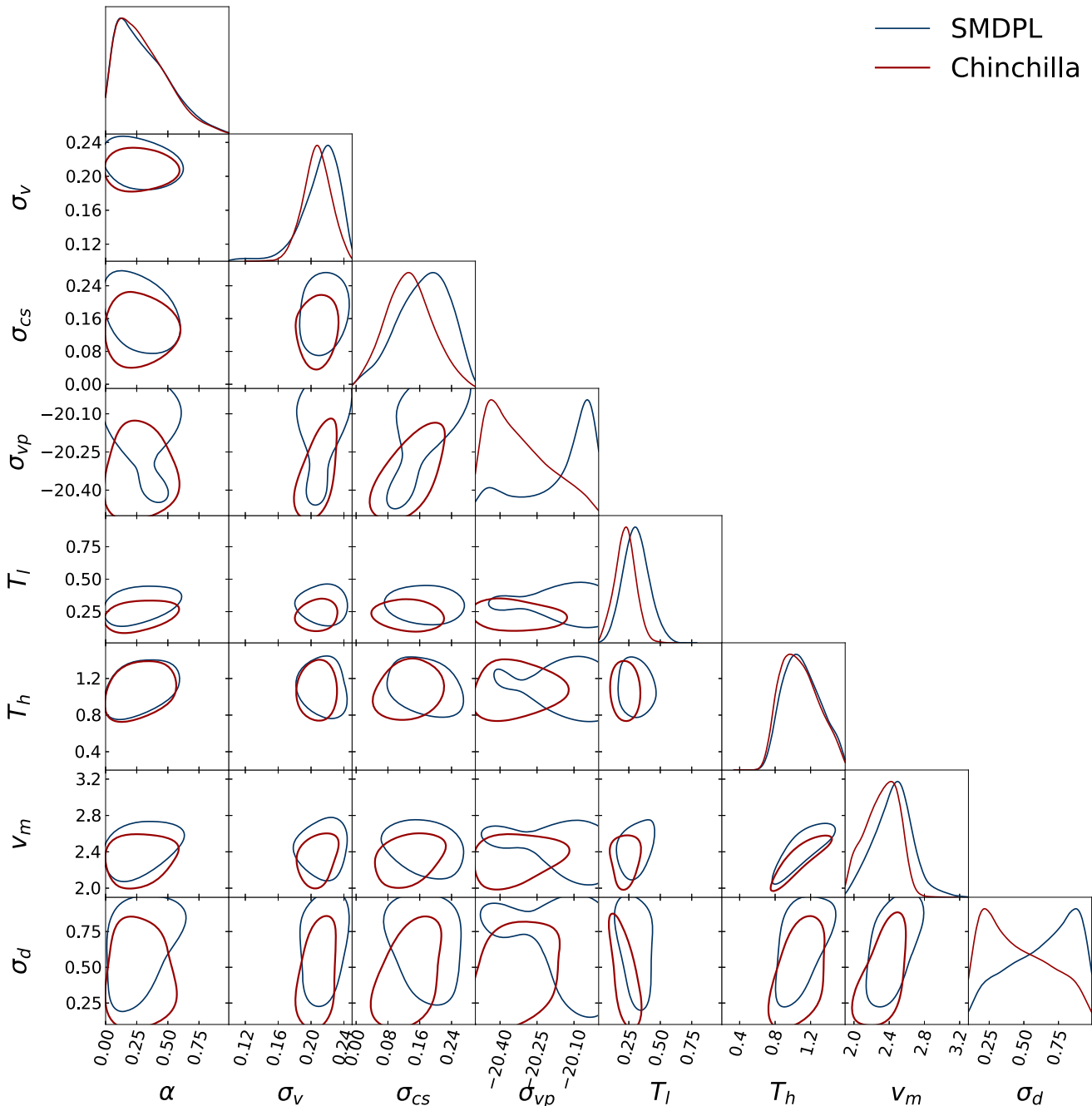
Figure 3. Posteriors of the extended SHAM model given by SDSS galaxy clustering and group–galaxy cross-correlations. Contours show the 68% confidence region of the posteriors given by w_{gg} and w_{cg}. Red lines show constraints based on Chinchilla-T1, and blue lines show constraints based on SMDPL. Details of these simulations can be found in Table 1. Descriptions of these parameters are given in Table 2.

cases in these surveys place stringent constraints on the galaxy properties in the simulations. In this paper, we mainly focus on the two main samples in DES: redMaPPer clusters (Rykoff et al. 2014, 2016) and redMaGiC galaxies (Rozo et al. 2016). As shown in DeRose et al. (2019), these two samples place the most stringent constraints on the galaxy models using DES data. We provide a brief description of these two samples below.

3.2.1. redMaPPer Clusters

In optical surveys, galaxy clusters appear to be spatial and redshift concentrations of red-sequence galaxies. The relatively tight color–absolute magnitude relations of red-sequence galaxies enable one to identify galaxy clusters using galaxies without redshift information. The primary algorithm used by the DES Collaboration (Rykoff et al. 2016) and the LSST Dark Energy Science Collaboration (Kovacs et al. 2022) to select clusters is redMaPPer (Rykoff et al. 2014), which employs a matched filter algorithm to select overdensities of red-sequence galaxies. Here, we briefly summarize the algorithm. First, the redMaPPer algorithm uses spectroscopic data to construct a red-sequence template empirically. It then computes the redshift of each galaxy by matching its color to the template. Second, redMaPPer identifies bright and red galaxies as cluster centers and determines the probability of each galaxy

being a member ($p_{\rm mem}$) by comparing its spatial distribution, color, and luminosity to a model. Third, the algorithm removes clusters that have $p_{\rm mem} > 0.5$ of another cluster and repeats the above processes. Finally, redMaPPer assigns a richness value (λ) to each cluster, calculated by summing $p_{\rm mem}$ of each member galaxy. This richness value is used as a primary cluster mass proxy (McClintock et al. 2019; Abbott et al. 2020; To et al. 2021a) in cosmological analyses given their expected tight relation to halo masses (Rozo & Rykoff 2014; Rozo et al. 2015a). redMaPPer also provides the most probable redshift of each galaxy cluster (z_λ) based on the colors of its member galaxies.

3.2.2. redMaGiC Galaxies

The red-sequence model derived from the redMaPPer clusters can be used to select galaxy samples with excellent photometric redshift uncertainties ($\sigma_z \approx 2\%$). To achieve this, one first constructs a color model using redMaPPer member galaxies with high membership probabilities ($p_{\rm mem}$). The redshift of each galaxy can be estimated by maximizing the consistency of galaxy colors and color models. One can then select bright galaxies with colors consistent with the color model. The galaxy samples selected in this way are called redMaGiC (Rozo et al. 2016) and are one of the primary lens samples in the DES cosmology analyses (Abbott et al. 2018, 2022). Based on the consistency with the color model, each redMaGiC galaxy is associated with a redshift probability distribution $p(z_{\rm redmagic})$.

3.3. Painting Galaxy Luminosities onto Dark Matter Particles

We use the ADDGALS algorithm to populate the dark matter light cones presented in DeRose et al. (2019) and Wechsler et al. (2022) with galaxies. The algorithm is detailed in Wechsler et al. (2022). Here, we briefly summarize the general formalism and highlight modifications.

ADDGALS populates galaxies in low-resolution N-body simulations based on three distributions:

1. The distribution of dark matter overdensities around galaxies given their absolute r-band magnitude M_r and redshift z, $P(R_\delta | M_r < x, z)$, where x is the specified magnitude threshold. The local dark matter overdensity is estimated using R_δ, distances to the k nearest dark matter particles such that the enclosed dark matter mass is $1.3 \times 10^{13}\, h^{-1} M_\odot$,
2. A galaxy luminosity function $\phi(M_r, z)$, and
3. The distribution of central galaxy absolute magnitudes given their host halos' virial mass ($M_{\rm vir}$) and redshift, $P(M_r | M_{\rm vir}, z)$.

In this work, the galaxy luminosity function is the same as the one used to create SHAM models at $z = 0$. At $z > 0$, the Schechter function's characteristic luminosity (L^*) is shifted based on a third-order polynomial. The parameters of this polynomial are constrained based on DES Y1 data (Abbott et al. 2018). Details of constructing this luminosity function are given in DeRose et al. (2019).

Given the redshift-dependent luminosity function, we populate galaxies on each snapshot of Chinchilla-T1 simulations using the best-fit SHAM model presented in Section 2. We then use these galaxies to determine $P(M_r | M_{\rm vir}, z)$ for central galaxies in resolved halos, defined as halos containing

more than 200 particles, and $P(R_\delta | M_r < x, z)$ for satellite and central galaxies in unresolved halos.

To determine the distribution of central galaxy absolute magnitudes at fixed host halo virial mass $M_{\rm vir}$ and redshift, we assume that $P(M_r | M_{\rm vir}, z)$ is a Gaussian distribution with a mass-dependent scatter. The mass dependency of the scatter is determined by fitting a straight line to the measured M_r scatter of $M_{\rm vir} > 10^{13}\, h^{-1} M_\odot$ halos in the SHAM galaxy catalogs. The mean of the Gaussian distribution is given by

$$\langle M_r(M_{\rm vir}) \rangle = A - 2.5\left(a \log_{10}\left(\frac{M_{\rm vir}}{b} \right) \right) \tag{9}$$

$$- \frac{1}{c} \log_{10}\left(1 + \left(\frac{M_{\rm vir}}{b} \right)^{(c \times d)} \right), \tag{10}$$

where A, a, b, c, and d are free parameters determined at each snapshot of the SHAM galaxy catalogs. Using the best-fit function, we populate central galaxies on resolved halos in the low-resolution light-cone simulations.

Similar to $P(M_r | M_{\rm vir}, z)$, we determine $P(R_\delta | M_r < x, z)$ in the SHAM galaxy catalogs at each snapshot. We model $P(R_\delta | M_r < x, z)$ as a log-normal and normal distribution sums, given by

$$P(R_\delta | M_r < x, z)$$
$$= (1 - p) \frac{e^{-(\ln(R_\delta) - \mu_c)^2 / 2\sigma_c^2}}{R_\delta \sqrt{2\pi}\, \sigma_c} + p \frac{e^{-(R_\delta - \mu_f)^2 / 2\sigma_f^2}}{\sqrt{2\pi}\, \sigma_f}, \tag{11}$$

where p, μ_c, μ_f, σ_c, and σ_f are free parameters and are measured in each snapshot of the SHAM galaxy catalogs at grids of magnitude thresholds from $M_r = -18$ to -24. Note that σ_f has length units, and σ_c is dimensionless. The additional R_δ in the denominator of the first term ensures the consistency of the units. We then build a Gaussian process emulator of these parameters to enable accurate interpolations. We detail the emulator construction in Appendix C.

With $P(M_r | M_{\rm vir}, z)$ and $P(R_\delta | M_r < x, z)$, we paint r-band luminosities onto dark matter particles in the light-cone simulations. For the resolved halos, we paint luminosities at the center of each halo using the learned $P(M_r | M_{\rm vir}, z)$. The resolved halos are defined as halos with $M_{\rm vir} > 6 \times 10^{12}\, h^{-1} M_\odot$ for the L1 and L2 boxes and $M_{\rm vir} > 1 \times 10^{13}\, h^{-1} M_\odot$ for the L3 box. These choices of halo masses were justified in Wechsler et al. (2022). For unresolved galaxies, we paint luminosities onto dark matter particles. We first generate random realizations of galaxies' redshifts by inverse transform sampling the measured redshift distributions of dark matter particles in the light-cone simulations. Next, we generate random realizations of galaxies' r-band luminosities (M_r) from the luminosity function $\phi(x, z)$ after subtracting the number densities of resolved halos. By doing so, we avoid double-painting central galaxies. We then convert the cumulative conditional probability $P(R_\delta | M_r < x, z)$ to $P(R_\delta | M_r, z)$ using finite difference estimation

$$P(R_\delta | M_r = x, z) = \frac{N(x + \delta x) P(R_\delta | M_r < x + \delta x, z)}{M}$$
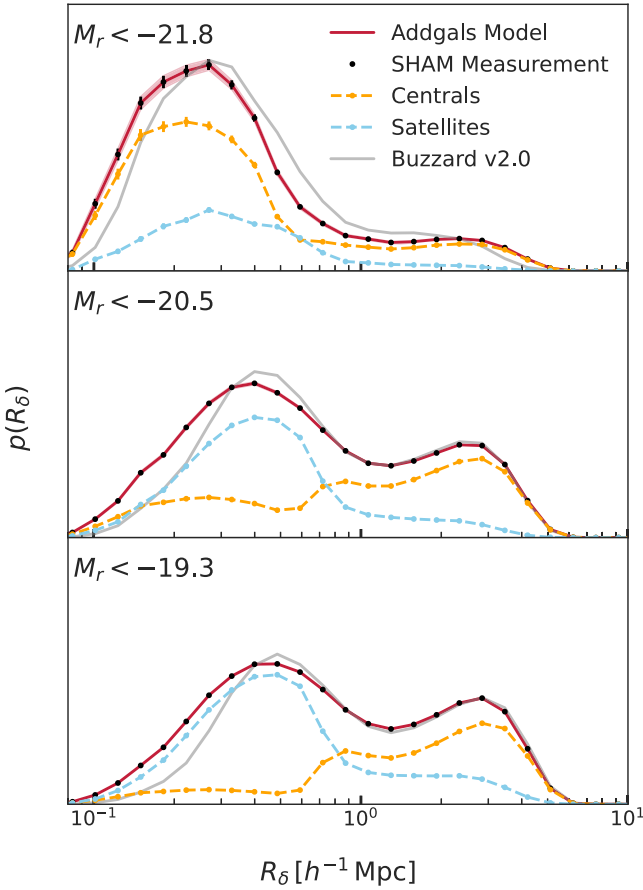$$- \frac{N(x) P(R_\delta | M_r < x, z)}{M}, \tag{12}$$

Figure 4. Distribution of R_δ, the radius that encloses a dark matter mass of $1.3 \times 10^{13} \, h^{-1} M_\odot$, at $z = 0$. Each panel shows a magnitude cut, with the brightest samples on the top and the faintest samples on the bottom. Black dots show measurements in SHAM catalogs with error bars showing a 1σ Poisson error. Red lines show the measurements of R_δ using mock galaxies in light-cone simulations. We further break the R_δ distributions in SHAM mocks into centrals (blue) and satellites (orange). As a comparison, the previous version Buzzard v2.0 (Wechsler et al. 2022) is shown as gray lines.

where M is a normalization constant and $N(x) = \int_{-\infty}^{x} \phi(x, z)$, the cumulative luminosity function. Each galaxy is randomly assigned an R_δ based on its M_r and z by inverse transform sampling of $P(R_\delta | M_r, z)$. We arrive at a list of galaxy (R_δ, M_r, z) values. The remaining task is to put galaxies in the correct positions. In the N-body light cones, we measure R_δ for each dark matter particle. We then divide the dark matter light cone and galaxy samples into $\Delta z = 0.01$ bins. For each redshift bin, we assign galaxies to dark matter particles using the closest match of R_δ in orders of galaxies' brightness.

Figure 4 shows the comparison of R_δ distributions of the painted galaxies (red line) onto those measured in the SHAM model (black dots). The agreement between the red line and the black dots indicates that Equation (11) provides a reasonable description of the measurement and the galaxy assignment algorithm works reasonably well. Similar to Wechsler et al. (2022), the R_δ distributions show double bump features for the two faintest galaxy thresholds. Central galaxies dominate one bump, and satellite galaxies dominate the other. This double bump feature of R_δ distributions indicates the effectiveness of R_δ in separating centrals and satellites at a given luminosity. We further compare the R_δ distribution in this work with the previous version (Buzzard v2.0; Wechsler et al. 2022). We find

that R_δ is shifted to smaller values at a given luminosity than Buzzard v2.0, indicating a more significant satellite fraction at a given luminosity. This is likely because we include orphan satellites in the SHAM model while Buzzard v2.0 did not.

3.4. Painting Colors Onto Galaxies

So far, we have generated a galaxy mock with a realistic spatial distribution and rest-frame r-band luminosity distribution. This section describes an algorithm for assigning colors to these galaxies.

3.4.1. Overall Color Distribution

Following Wechsler et al. (2022), we assume that galaxy SEDs can be described by the product of five coefficients and five kcorrect spectral templates (Blanton et al. 2003; Blanton & Roweis 2007). One can then directly assign the measured kcorrect coefficients in real data to galaxies in the simulations. This approach has a couple of benefits. First, the small number of coefficients for each galaxy reduces the requirements of memory and computational power to generate and store mock catalogs. Second, the product of coefficients and kcorrect templates provides full SED information for each galaxy, from which one can compute the observed magnitude by applying band shifts and the observed band-passes without regenerating galaxy SEDs. Third, the direct assignment from data guarantees reasonable matches between mocks and data. However, this approach requires a pool of measured kcorrect coefficients representative of the targeted survey data. Generating this pool of kcorrect coefficients requires representative spectroscopic redshifts down to the photometric survey depths. This is usually unachievable.

Following DeRose et al. (2019), we adopt a hierarchical approach to associate kcorrect coefficients of an SDSS galaxy to a simulated galaxy. We first generate a representative sample at $z < 0.2$ down to DES survey depth using galaxies with $z = (0.005, 0.2)$ in the SDSS DR7 VAGC catalog (Blanton et al. 2005). Each galaxy has five kcorrect coefficients according to Blanton et al. (2003) and Blanton & Roweis (2007). We then use PRIMUS (Coil et al. 2011) galaxies to quantify the redshift evolution of this coefficient pool. Specifically, we employ PRIMUS galaxies to calculate the ratio $w_r(M_r, z)$ of the probability of being red at z to the probability at $z < 0.2$. We define galaxies as red when $^{0.1}g - r > 0.15 - 0.03 M_r$. With $w_r(M_r, z)$ at hand, we can generate kcorrect coefficients for each simulated galaxy using a rejection sampling algorithm. We first bin SDSS galaxy samples and simulated galaxies using their rest-frame r-band luminosities (M_r) with widths $\Delta M_r = 0.1(22.5 + M_r)$. For each simulated galaxy in the ΔM_r bin, we first remove the SDSS galaxy samples in the same bin if

$$\text{rand} > w_r(M_r, z) p_{\text{red}},$$
$$p_{\text{red}} = \frac{N_{\text{red}}(M_r \in \Delta M_r)}{N(M_r \in \Delta M_r)}, \qquad (13)$$

where rand is a random number drawn from a uniform distribution from 0–1, p is an empirically measured red fraction in SDSS galaxy samples, z is the redshift of the simulated galaxy, and N_{red} is the number of red galaxies in a given magnitude bin. We associate the kcorrect coefficients of a

galaxy in the remaining SDSS galaxy sample to the simulated galaxy by selecting the one with the closest value of M_r. If there is no SDSS galaxy in the ΔM_r bin, we expand the width of the bin to $\Delta M_r = 0.1(22.5 + M_r)^2$ and repeat the rejection sampling step. By doing so, we paint `kcorrect` coefficients to each simulated galaxy. With `kcorrect` coefficients for each simulated galaxy, we can compute the galaxy's rest-frame $g - r$ color by convolving the SED generated from these `kcorrect` coefficients and the SDSS's bandpass. The simulated galaxies will have the same color–M_r relation in the SDSS galaxy samples and the same red fraction–redshift relation as constrained by the PRIMUS galaxy samples. These steps have been validated in Wechsler et al. (2022) and DeRose et al. (2019).

3.4.2. Environmentally Dependent Galaxy Color Distribution

Once we create mock galaxies with realistic overall color distribution, we shuffle galaxy colors to create environmental dependencies of the color. We use the conditional abundance matching technique (Hearin & Watson 2013; Masaki et al. 2013) and the rest-frame $g - r$ color to perform the shuffling. Specifically, we assign the SED that corresponds to the rest-frame $g - r$ color to a galaxy such that the following equation is satisfied:

$$p(<\text{rank}(g - r)|M_r) = p(>\text{rank}(e_p)|M_r), \quad (14)$$

where e_p is a proxy to quantify galaxy environments. DeRose et al. (2022a) found that R_h, the distance to the nearest massive halos, provides a good proxy of galaxy environments. Using R_h and the conditional abundance matching technique, they find comparable color-dependent galaxy clustering to SDSS measurements (DeRose et al. 2022a; Wechsler et al. 2022). However, this proxy is inadequate for galaxies in galaxy clusters. First, more massive halos are bigger. On average, galaxies that live in a more massive halo would have larger R_h than those living in small mass halos. Therefore, using R_h as a color proxy would make galaxies *bluer* in *more massive* halos. Second, for galaxies living in massive clusters, R_h is the distance to the central galaxy. Using R_h as a color indicator would create a strong radial color profile in a massive halo that is independent of cluster mass. This strong correlation breaks the self-similarity of galaxy clusters of different masses and lacks observational support. One hint of these problems is shown in Wechsler et al. (2022)'s comparison of red galaxy clustering at $r < 1 \, h^{-1}$ Mpc in the faintest magnitude bin (see also the gray line in Figure 5). The ratio of red galaxy clustering amplitude to all galaxies is low compared to the data. Since most of the pairs that contribute to small-scale clustering are from galaxies residing in large-mass halos, the small clustering ratio indicates that the galaxies in clusters are too blue.

Given the aforementioned shortcomings of R_h, we define a new galaxy environment proxy e_{hc}. e_{hc} is constructed based on two insights. First, assuming galaxy clusters with different masses are self-similar, the distance to a cluster should be measured in units of the cluster's radius. In practice, we use the ratio of the distance to a cluster and the cluster's virial radius (R_{vir}) to some power (c_0) as the color proxy. The additional power of the virial radius allows the possibility that more

massive halos are stronger at quenching galaxies. The second insight is that the galaxy color gradient is shallower at the inner part of the clusters compared to the outskirts (Adhikari et al. 2021). Further, given the success of producing reasonable large-scale red galaxy clustering using R_h (Wechsler et al. 2022), we would like the new environment proxy e_{hc} to be similar to R_h on large scales. Therefore, we design a mapping from R_h to e_{hc} such that e_{hc} approaches $R_h + c$ when R_h is infinity and becomes independent of R_h when R_h is zero. c is an arbitrary constant irrelevant to color assignments because only the ranks of R_h are related to galaxy colors Equation (14). Given these two insights, the environment proxy has the following functional form:

$$x = d_h/R_{vir}^{c_0}$$

$$e_{hc} = 0.5x + c_1 \left(\frac{x - 1}{2c_1} \, \text{erf} \left(\frac{x - 1}{2c_1} + \frac{\exp\left(-\left(\frac{x-1}{2c_1}\right)^2\right)}{\sqrt{\pi}} \right) \right),$$

$$(15)$$

where d_h is the distance to the nearest massive halo with mass greater than M_{ch}, and M_{ch}, c_0, and c_1 are three free parameters. c_1 controls the sharpness of transition from $de_{hc}/dx = 0$ to $de_{hc}/dx = 1$.

One remaining problem of Equation (15) is that all central galaxies with $M > M_{ch}$ have $e_{hc} = 0$, making them the reddest galaxies at a given magnitude. Given that M_{ch} is usually $10^{12.8} \, h^{-1} M_\odot$, assigning all central galaxies with $M > M_{ch}$ the reddest SEDs will contradict observations (e.g., Wetzel et al. 2012). We must make some central galaxies living in low-mass halos blue. The challenge is that our light cone has different resolutions at different redshifts, and a simple mass cut might transfer this reshift-dependent resolution onto galaxy colors. Fortunately, R_δ gives a nice separation of centrals and satellites that is less sensitive to the resolution of the simulations (see Figure 4). Specifically, using Chinchilla-T1, we find that none of the satellite galaxies in $M > 10^{12.8} \, h^{-1} M_\odot$ halos and none of the central galaxies with $M > 2 \times 10^{13} \, h^{-1} M_\odot$ have R_δ greater than 2. That is, galaxies with $R_\delta > 2$ are likely to be low-mass isolated centrals whose color should be preferentially blue. Built on these insights, we can increase the value of the environmental proxy e_{hc} of a galaxy living in a low-density environment ($R_\delta > 2$) and likely to be red ($e_{hc} < 1$) in the original algorithm. In this way, these galaxies will be bluer because of a larger environmental proxy value. We, therefore, increase galaxies' e_{hc} by c_2 if $R_\delta > 2$ and $e_{hc} < 1$.

Finally, we allow the possibility that color ranks (rank $(g - r)$) and galaxy environment proxy ranks (rank(e_p)) are not perfectly correlated. Instead of implementing Equation (14), we implement

$$p(<\text{rank}(g - r)|M_r) = p(<\widetilde{\text{rank}}(e_p)|M_r). \quad (16)$$

In the above equation, $\widetilde{\text{rank}}(e_p)$ is constructed such that it is an unbiased estimator of rank$(g - r)$ and is correlated with rank $(g - r)$ with correlation coefficient r_c. We construct $\widetilde{\text{rank}}(e_p)$ using `Halotools` (Hearin et al. 2017). DeRose et al. (2022a) find that galaxies with different stellar masses prefer a different r_c using SDSS color-dependent clustering measurements (Zehavi et al. 2011). Motivated by their findings, we
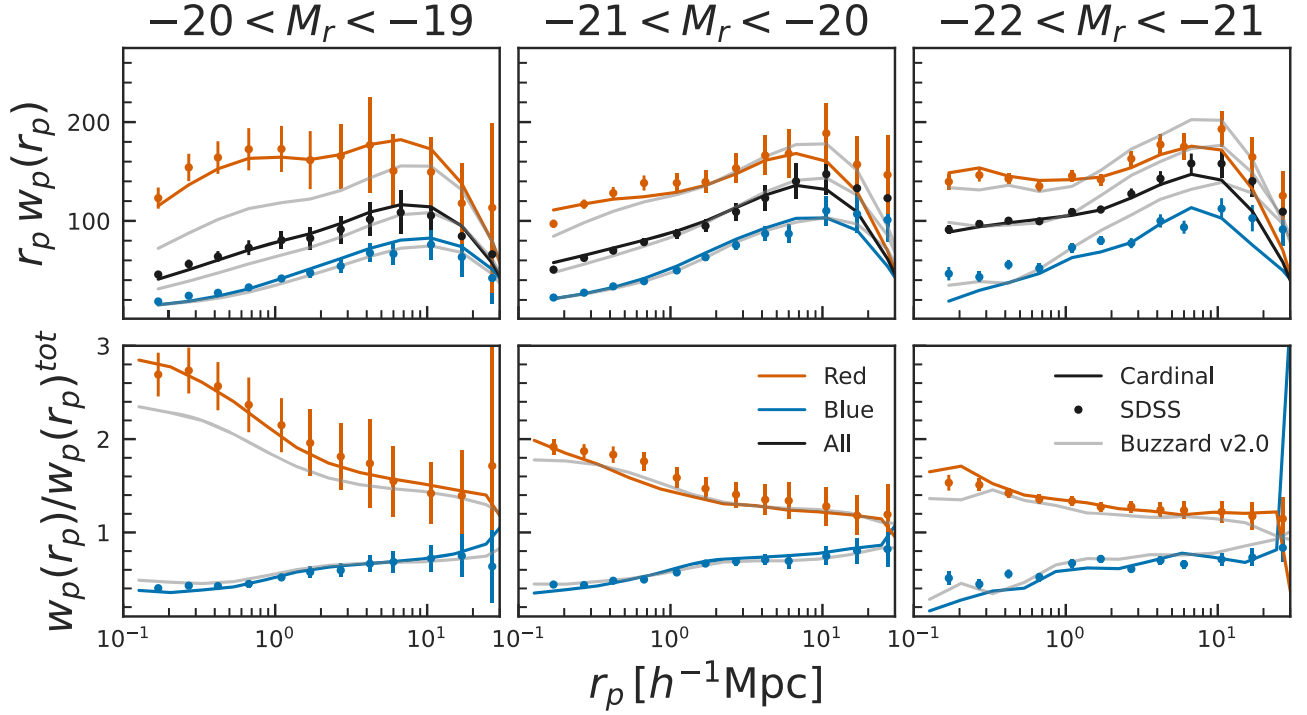
Figure 5. Comparison of color-dependent clustering measured in Cardinal (line) and data (dots with error bars, Zehavi et al. 2011). The error bar shows 1σ uncertainties. Red lines and dots correspond to red galaxies, while blue lines and dots correspond to blue galaxies. Different columns correspond to different magnitude bins in galaxies. For comparison, the results from Buzzard v2.0 are shown in gray.

Table 3
Parameters of the Color Model Described in Section 3.4.2

Parameter	Best-fit Value	Description	Relevant Equations
M_{ch}	12.62	Mass threshold of halos	Equation (15)
c_0	2.00	Power index of halo radius	Equation (15)
c_1	0.50	Sharpness of transition from $de_{hc}/dx = 0$ to $de_{hc}/dx = 1$	Equation (15)
c_2	10.3	Offset of e_{hc} for low-mass centrals	
c_l	0.47	Bright-end asymptotic value of color–e_{hc} correlations	Equation (17)
c_h	0.90	Faint-end asymptotic value of color–e_{hc} correlations	Equation (17)
c_m	19.73	Transition magnitude of color–e_{hc} correlations	Equation (17)
c_σ	1.12	Sharpness of the transition of color–e_{hc} correlations as a function of magnitude	Equation (17)

parameterize r_c to depend on M_r via

$$r_c = c_l + (c_h - c_l)\left(0.5 + 0.5\,\mathrm{erf}\left(\frac{M_r - c_m}{\sqrt{2}\,c_\sigma}\right)\right), \qquad (17)$$

where c_l and c_h are free parameters that govern the asymptotic behaviors at low and high M_r, and c_m and c_σ govern the transition M_r and the steepness of this transition. Equation (17) has the same functional form as Equation (8). Again the underlying physical behavior is pretty simple. For bright galaxies (small M_r), r_c is determined by c_l. Conversely, for faint galaxies, r_c is determined by c_h. Equation (17) simply ensures a smooth transition from bright to faint objects, and two additional parameters control the location and slope of the transition.

In summary, our color assignment model has eight parameters (summarized in Table 3):

1. M_{ch}: a parameter that defines the mass threshold of halos to which we calculate distances (d_h) of each galaxy.
2. c_0 and c_1: two free parameters control the mapping of d_h to the environment proxy e_{hc}.
3. c_2: a parameter is added to e_{hc} of low-mass centrals to make it bluer.
4. c_l, c_h, c_m, and c_σ: parameters that control the luminosity dependence of the correlation between colors and environment proxies.

We use color-dependent clustering measured in Zehavi et al. (2011) to constrain these parameters. Specifically, to separate the uncertainties of the color model and the galaxy clustering model, we use the ratio of galaxy clustering of red and blue galaxies to all galaxies to constrain the color model parameters. A simple downhill simplex algorithm is used to find the best-fit parameters. At each step of the downhill simplex algorithm, we shuffle the galaxy SEDs in the mocks using Equation (16), regenerate galaxy colors using SEDs and SDSS filters, select

galaxy samples according to Zehavi et al. (2011), calculate galaxy clustering using Corrfunc (Sinha & Garrison 2020), and compare the measurements with the data assuming a Gaussian likelihood. Finally, our best-fit model has $\chi^2 = 84.9$ with a degree of freedom of 58, corresponding to a PTE value of 0.01. In Figure 5, we compare the best-fit model to the measurement. We find that our best-fit model can accurately describe the data where the predicted clustering differs from that of Buzzard v2.0. The differences between our model and Buzzard v2.0 are especially pronounced for the lowest luminosity sample.

3.5. Observational Effect

3.5.1. Photometric Noise

So far, we have generated a quarter sky galaxy light cone with realistic color, luminosity, and spatial distributions. We add galaxy shapes and sizes following methods described in DeRose et al. (2019). Finally, we apply lensing effects using the ray-tracing code CALCLENS (Becker 2013). Specifically, we apply deflection, rotation, shear, and magnification for all galaxies to alter their positions, shapes, and photometry.

We cut out two DES Y3 regions from this light cone and rotate them into the DES Y3 footprint (Sevilla-Noarbe et al. 2021). We then apply several observational effects on the mock catalogs. First, we apply the survey masks on the mock using the DES Y3 survey mask that is in the form of a HEALPix map with $N_{side} = 4096$, corresponding to a resolution of 0.73 arcmin². For each HEALPix pixel, we randomly select galaxy samples according to the fracgood value, which describes the amount of masking within the HEALPix pixel. Second, we add photometry noise to each galaxy, a process that will be detailed later. This step is essential because magnitude noise can drastically affect the number of galaxies near survey depth limits due to Eddington biases. For analyses that use galaxies near survey depth limits, which is the case for most weak lensing analyses, properly modeling noise is essential. Finally, we cut out galaxies with observed magnitude fainter than the survey depth.

The remaining task is to add a realistic magnitude error on each galaxy. We follow the prescription described in DeRose et al. (2019) and Wechsler et al. (2022) that uses the effective exposure time (t_{eff}) and 10σ limiting magnitudes (m_{lim}) from survey data. Here, we briefly summarize the prescriptions. First, we assume that the observed total number of photons of a galaxy comes from two contributions: photons from galaxies I_{gal} and photons from the noise, such as sky, readout noise, etc. I_{sky}. The I_{gal} can simply be related to galaxy magnitude (m_{gal}) via

$$I_{gal} = 10^{-0.4(m_{gal}-ZP)} \times t_{eff}, \tag{18}$$

where $ZP = 22.5$. Following Rykoff et al. (2015), I_{sky} can be empirically determined from data. Sevilla-Noarbe et al. (2021) provide the 10σ limiting magnitudes (m_{lim}) and the effective exposure time (t_{eff}) in the form of HEALPix maps with a resolution of 0.73 arcmin². Using this information, we calculate I_{sky} for each HEALPix pixel by

$$I_{sky} = \frac{10^{-0.8(m_{lim}-ZP)} \times t_{eff}}{100} - 10^{-0.4(m_{lim}-ZP)}. \tag{19}$$

We assume the observed number of photons follows a Poisson distribution. The noisy observed flux F_{obs} is then given

by

$$F_{obs} = \frac{Poisson(I_{gal} + I_{sky}) - I_{sky}}{t_{eff}}, \tag{20}$$

where Poisson denotes a random draw from the Poisson distribution. The second term in the above equation is to mimic the process of sky background subtractions in observations. We note that this implementation is different from DeRose et al. (2019) and Wechsler et al. (2022), where the authors performed a random draw from a Gaussian distribution with a scatter as the square root of the mean. The Gaussian approximation is valid for high signal-to-noise galaxy samples but could lead to a bias for low signal-to-noise galaxies. The noise of the observed flux σ_{flux} is given by $\sqrt{I_{gal} + I_{sky}}/t_{eff}$. The associated magnitude and error are

$$m_{obs} = ZP - 2.5\log_{10} F_{obs}$$
$$m_{err,obs} = \frac{2.5}{\log(10)} \frac{\sigma_{flux}}{F_{obs}}. \tag{21}$$

3.5.2. Gravitational Lensing

The matter along the line of sight will distort the light from galaxies. This distortion will modify the position of galaxies, distort their shapes, and magnify their brightness. We include these effects by performing full-sky multi-plane ray tracings using CALCLENS (Becker 2013). This step has been detailed in Appendix C of DeRose et al. (2019). Here, we provide a brief summary. First, we decompose the full-sky particle light cones into HEALPix maps with $n_{side} = 8192$, corresponding to a resolution of 0.46 arcmin. Next, for each pixel, we divide the particle light cones into equally spaced lens planes from $z = 0$–2.4 with separations of $25\,h^{-1}$ Mpc. Then, for each of the simulated galaxies, we calculate the deflection angle and distortion matrix of the light at each lens plane. To achieve this, we first calculate the lensing potential from integrated dark matter density fields by solving a two-dimensional Poisson equation (Jain et al. 2000). Next, we calculate the deflection angle using the lens equation (Teyssier et al. 2009), which relates the deflection angles to the first derivative of the lensing potential. Then, the distortion matrix is calculated using the second derivative of the lensing potential (Jain et al. 2000; Hilbert et al. 2009; Becker 2013). Finally, these deflection angles and distortion matrices at each plane are combined to produce the final total deflections, shears, and magnifications (Equations (10) and (12) of Becker 2013).

We validate the ray-tracing shears by computing the tangential shear γ_t around halos with $M_{vir} > 2 \times 10^{14}\,h^{-1}\,M_\odot$ and $z = [0.6, 0.62]$. This tangential shear can be related to the excess surface density ($\Delta\Sigma$) of the matter around halos, which can be measured directly using particles in the simulation. We relate the tangential shear to $\Delta\Sigma$ using the estimator presented in Sheldon et al. (2004), which reads as

$$\Delta\Sigma(r) = \frac{\sum_i^{N_{halos}}\sum_j^{N_{Sources}}\gamma_t^{i,j}\Sigma_{crit}^{i,j}}{\sum_i^{N_{halos}}\sum_j^{N_{Sources}}\left(\Sigma_{crit}^{i,j}\right)^2}, \tag{22}$$

where N_{halos} is the number of halos in the bin, $N_{Sources}$ is the number of galaxies around halos at distance r, and $\gamma_t^{i,j}$ is the tangential shear of galaxy j around halos i. To avoid contamination, we only use galaxies with a redshift of 0.1
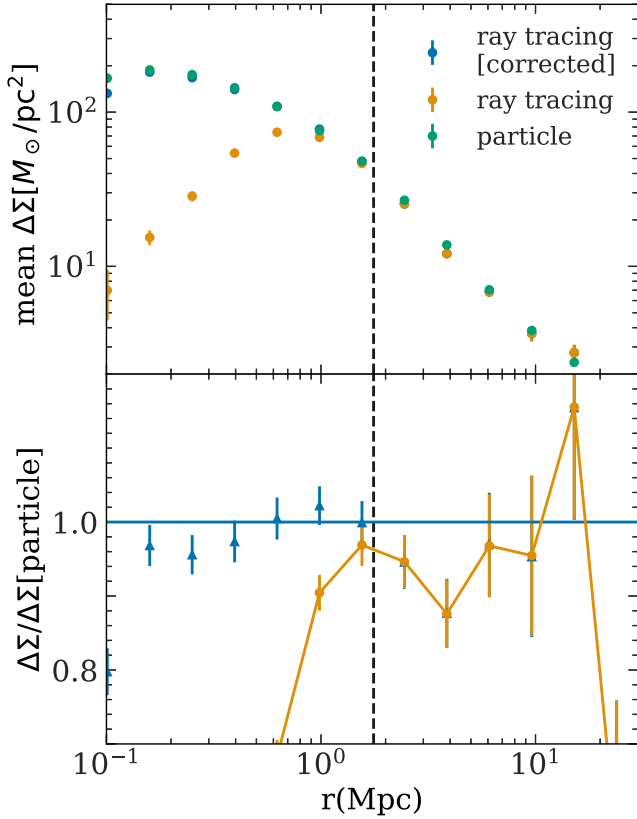
Figure 6. Mean $\Delta\sigma$ profiles measured around halos with $M_{vir} > 2 \times 10^{14} \, h^{-1} M_\odot$ and $z = [0.6, \, 0.62]$. The x-axis is measured in physical coordinates. The green dots are derived from dark matter particle–halo cross-correlations. The orange dots are calculated from ray-tracing shapes (and are similar to previous catalog versions). The blue dots show our improved model, calculated from corrected ray-tracing shears. Error bars are estimated with 10,000 bootstrap resampling. The black vertical dashed lines show 10 times the pixel size for ray tracings.

greater than the redshift of the halos. In the above equation, $\Sigma_{crit}^{i,j}$ is given by

$$\Sigma_{crit}^{i,j} = \frac{c^2}{4\pi G} \frac{D_s}{D_h D_{hs}}, \tag{23}$$

where D_h, D_S, and D_{hs} are angular diameter distances to halos, galaxies, and between halos and galaxies. The $\Delta\sigma$ can also be calculated from the matter-halo cross-correlation functions ξ_{hm} using

$$\Delta\Sigma(r) = \bar{\rho}\frac{4}{r^2}\int_0^r x \, dx \int_0^\infty d\chi \, \xi_{hm}(\sqrt{x^2 + \chi^2}) \\ - 2\bar{\rho}\int_0^\infty d\chi \, \xi_{hm}(\sqrt{r^2 + \chi^2}). \tag{24}$$

Figure 6 shows the comparison of $\Delta\Sigma$ measured via ray tracing (Equation (22), orange dots) and via direct measurements of particles (Equation (24), green dots). We find that the ray-tracing and particle calculations agree well on large scales but deviate on small scales. This deviation has been shown in the literature (Takahashi et al. 2017; Kovacs et al. 2022) and likely comes from the finite angular resolution of 0.46 arcmin when we calculate the lensing potentials for ray tracing. This resolution problem in lensing can be problematic for cluster lensing analyses when most of the signal lies below $1 \, h^{-1}$ Mpc. Because of the resolution problem, cluster lensing studies have

relied on the dark matter particle–halo cross-correlation (Abbott et al. 2020; Wu et al. 2022). However, Wu et al. (2019) point out that while the particle–halo cross-correlation produces the right mean $\Delta\Sigma$, it significantly underestimates the halo-to-halo variance of $\Delta\Sigma$ at large scales. This is because particle–halo cross-correlations ignore the line-of-sight structure's contributions to the lensing profile.[13]

We empirically correct the shears of each galaxy using the measured particle–halo cross-correlations. We first bin the halos in simulations with redshift from $z = 0.18$–0.67 and the virial mass M_{halo} above $10^{13} \, h^{-1} M_\odot$ into $\Delta z = 0.02$ and $\Delta \log_{10}(M) = 0.15 \, h^{-1} M_\odot$ bins. Next, we calculate $\Delta\Sigma$ for halos in each bin using particle–halo cross-correlations ($\Delta\Sigma_p$) using Equation (24) and ray tracing derived shears using Equation (22) ($\Delta\sigma_\gamma$). We calculate the differences between the two and apply a correction on the two shear components of each galaxy γ_1 and γ_2. Specifically, our algorithm reads as

Algorithm 1.

---

for Galaxies with redshift z
    for halos h with mass $M_h > 10^{13} \, h^{-1} M_\odot$, redshift $z_h < z$, and distance to galaxies $r_h < 4'.6$ (10 times the resolution of the ray tracings) **do**
        $\Sigma_{crit}^{h,g} \leftarrow \frac{c^2}{4\pi G}\frac{D_g}{D_h D_{hg}}$
        $\Delta\gamma_t \leftarrow (\Delta\Sigma_p - \Delta\Sigma_\gamma)/\Sigma_{crit}^{h,g}$
        $\Delta\gamma_{1,h} = \Delta\gamma_t/(\cos(2\phi) + \sin(2\phi)\tan(2\phi))$
        $\Delta\gamma_{2,h} = \Delta\gamma_{1,h}\tan(2\phi)$
    end for
    $\gamma_1 \leftarrow \gamma_1 + \sum_h \Delta\gamma_{1,h}$
    $\gamma_2 \leftarrow \gamma_2 + \sum_h \Delta\gamma_{2,h}$
end for

In the above algorithm, D_g, D_h, and $D_{h,g}$ are angular diameter distances to galaxies, halos, and between halos and galaxies. Index h goes through halos with mass greater than $10^{13} \, h^{-1} M_\odot$ along the line of sight with angular separations less than 4.6 arcmin of the galaxies. The two equations relating $\Delta\gamma_t$ to $\Delta\gamma_1$ and $\Delta\gamma_2$ are derived assuming $\Delta\gamma_\times = 0$. This is motivated by the fact that gravitational lensing due to the localized mass distribution does not generate $\gamma_\times$ modes. The blue dots in Figure 6 show the $\Delta\Sigma$ measurements using the corrected galaxy shears. We find that it is consistent with the derivation from particle–halo cross-correlations, indicating the effectiveness of our algorithm. In Appendix E, we further show that this correction of galaxy shears has well below 0.1% impact on cosmic shear ($\xi_{+/-}$) at large scales. At small scales, $\xi_{+/-}$ based on corrected galaxy shears is more consistent with theoretical predictions.

3.6. Conditional Abundance Matching Color

In this section, we first diagnose mismatches between modeled and observed color distributions. We then describe our schemes for correcting these mismatches.

Figure 7 compares the total magnitude distributions (the left column) and color–color distributions (the right column) of

---

[13] We note that one could include line-of-sight contribution into particle–halo cross-correlations by calculating two-dimensional projected correlations as implemented in `Halotools` (Hearin et al. 2017). However, as shown in Wu et al. (2019), one must consider particles with distances to clusters much greater than $100 \, h^{-1}$ Mpc to avoid underestimating variances, which might be computationally challenging.
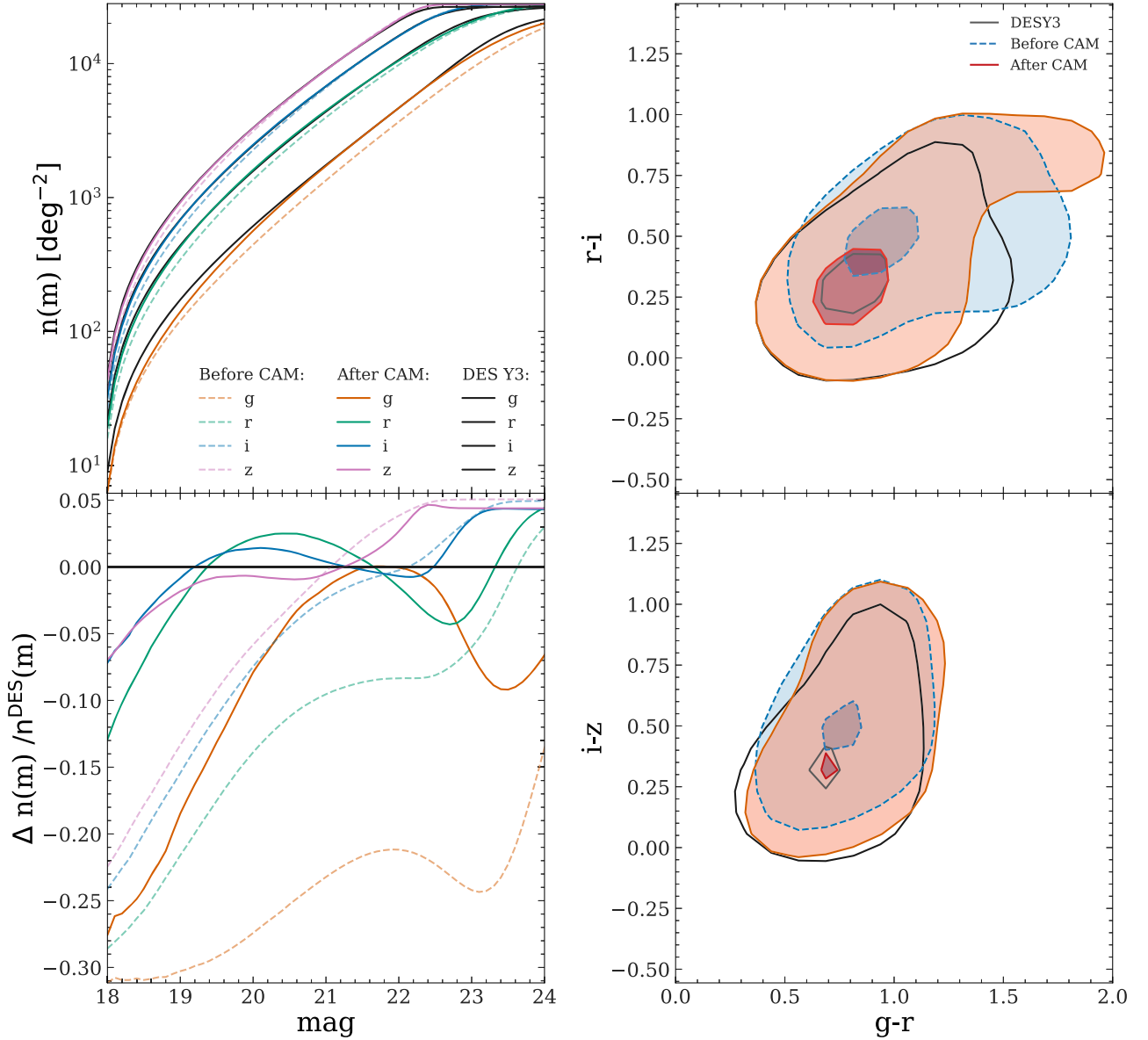
Figure 7. Comparison of apparent magnitude between Cardinal and DES Y3 data. The top left panel compares overall magnitude distributions, and the bottom left panel shows the fractional differences. The top right panel compares $g - r$ and $r - i$ distributions, and the bottom right compares $g - r$ and $i - z$ distributions. The black lines show measurements of DES Y3 data, the dashed lines show the Cardinal data before applying the conditional abundance matching scheme (detailed in Section 3.6), and the solid lines show the Cardinal data after applying the conditional abundance matching scheme. Different colors correspond to different color bands. Contours show 1σ and 2σ boundaries of the density distributions. The new conditional abundance matching scheme greatly improves the consistency between mocks and data.

galaxies in the mock (dashed lines) and the data (solid black lines). In general, the galaxies in the mock have redder colors. This trend has also been shown in Figure 5 of DeRose et al. (2019), where the authors compared Buzzard to COSMOS data. As described in Section 3.4.1, the overall color distribution is determined by two different factors: (a) kcorrect spectral templates in Blanton & Roweis (2007), and (b) SDSS kcorrect coefficients with redshift evolution controlled by PRIMUS' red fraction measurements. To identify the exact cause of this mismatch of colors, we first test the effectiveness of kcorrect spectral templates in describing colors of high-redshift galaxies. We calculate kcorrect coefficients of redMaPPer member galaxies in clusters with a richness greater than 20 measured in the DES Y3 data. In this calculation, we use the redshift of the host clusters to minimize

the error of photometric redshifts. We then reconstruct the magnitude of galaxies using these kcorrect coefficients. Figure 8 compares the $g - r$ color measured using the reconstructed magnitudes to those measured with the input magnitudes. We choose to show $g - r$ for simplicity but find that the trend is consistent between different colors. Figure 8 shows that the kcorrect reconstructed color is unbiased for the low redshift bin. However, for the high-redshift bin, while the reconstructed color of blue galaxies seems to be mostly unbiased, the reconstructed color is increasingly biased for redder galaxies. This indicates that summarizing colors with kcorrect spectral templates is valid for low-redshift galaxies but biases the colors of high-redshift red galaxies. We further test this finding using 31 COSMOS SED templates (Ilbert et al. 2009), and find a similar result. Interestingly, this bias makes
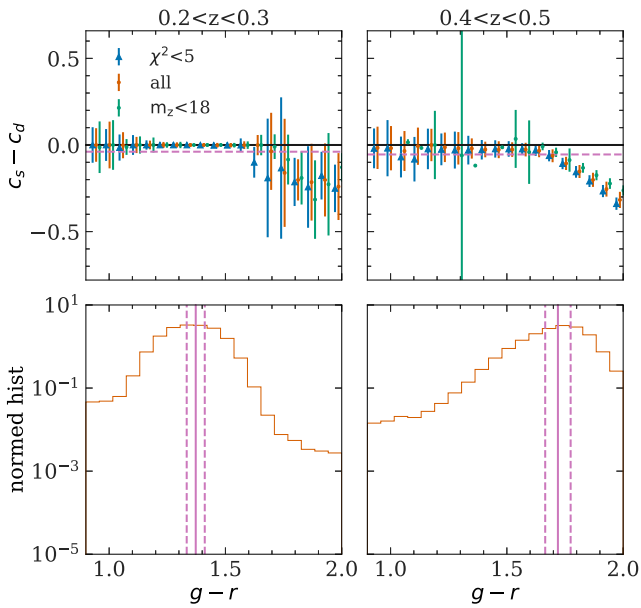
Figure 8. Color ($g - r$) residuals as a function of $g - r$ color (c_d). c_s is the $g - r$ color constructed using k-correction templates and best-fit coefficients. This calculation uses `redMaPPer` member galaxies with `redMaPPer` run on DES Y3 data. Each column corresponds to a redshift slice using redshifts of host galaxy clusters. Top panels: dots show the median c_s–c_d in each color bin, while error bars show 68% scatter. Magenta lines show 1σ widths of red sequence reported by `redMaPPer`. We split the samples into two subsets: galaxies that are bright with $m_z < 18$ (green), and galaxies that best match `redMaPPer`'s red-sequence template (blue). Both subsets are artificially shifted horizontally to improve clarity. Bottom panels: histograms show the measured color distributions of all galaxies in the calculation. Magenta lines show the mean (solid) and width (dashed) of the red sequence reported by `redMaPPer`. For all samples, the reconstructed color is mostly unbiased at $z = 0.2$–0.3. At $z = 0.4$–0.5, the reconstructed color is biased for red galaxies and mostly unbiased for blue galaxies.

red galaxies bluer, which contradicts the trend shown in Figure 7. Therefore, we conclude that the trend shown in Figure 7 is likely due to the insufficiency of the combination of SDSS `kcorrect` coefficients and PRIMUS' red fraction measurements for describing the colors of high-redshift galaxies.

In summary, for red galaxies, two competing effects affect their colors: the insufficiency of SDSS `kcorrect` coefficients and PRIMUS' red fraction make them too red, and summarizing the color with `kcorrect` spectral templates makes them too blue. The cancellation of these two competing effects on the colors of red galaxies can potentially explain why the mean color of red-sequence galaxies in Buzzard (DeRose et al. 2019) is consistent with the data. In contrast, the overall colors of galaxies are biased red. Further, while the cancellation of these two competing effects can make the colors of red galaxies unbiased, it could boost the total number of red galaxies because blue galaxies are more abundant than red-sequence galaxies. Figure 5 of DeRose et al. (2019) shows that the number of red galaxies in Buzzard is larger than in the COSMOS data. To further understand this hypothesis, we present a toy model in Appendix J, showing that our hypothesis can qualitatively reproduce Buzzard's $g - r$ color distribution shown in Figure 5 of DeRose et al. (2019).

A more comprehensive and physical solution to these two causes of color mismatches between mocks and data is important but requires further research. Here we provide an empirical solution using DES Y3 photometric data. The DES

Y3 data constrain galaxies' overall color and apparent magnitude distributions. Because DES Y3 data do not have redshift information for each galaxy, we must make some assumptions to calibrate our simulations with photometric data. We make the following assumptions:

1. The ranks of galaxy colors given the observed magnitude in mocks are valid.
2. The relative colors of galaxies in different redshifts are correct.

With these two assumptions, we can then employ the DES Y3 data to tune the multidimensional color and magnitude distributions of galaxies in Cardinal using the conditional abundance matching technique. We first match the observed z-band magnitude distributions in Cardinal and data because the z band is the reference magnitude used to select `redMaGiC` galaxies and `redMaPPer` clusters. To achieve this, we retune the third-order polynomial parameters that govern the redshift evolution of L^* by matching the luminosity functions of the mock galaxy catalogs to the data. We use a downhill simplex algorithm to minimize a Gaussian likelihood with Poisson errors. Second, we match the color distributions by enforcing the equality of the following probabilities between mocks and data:

$$P(<c_i | m_z, c_{j<i}), \qquad (25)$$

where $c_i = [g - r, r - i, i - z]$ denotes the three colors used for lens galaxies and cluster selections. We use the method implemented in `Halotools` (Hearin et al. 2017) to perform this matching. With this second step, we can ensure the matches of overall color distributions between mocks and data. The above algorithm has one important caveat. The red galaxies live in a tight color–magnitude space known as the red sequence. Because we match overall galaxy colors and magnitude, this process will widen the color–magnitude relations of red galaxies. In Appendix G, we present an algorithm that uses `redMaPPer` to reduce this broadening.

So far, we have obtained galaxy catalogs with realistic magnitude and color distributions. Unfortunately, this catalog likely has incorrect environmentally dependent galaxy colors because of additional noise introduced in the various conditional abundance matching processes. Despite this problem, we can still use this catalog to generate a new color template that captures the redshift dependence of color distributions more accurately than the original SDSS catalogs. Naively, we could rebuild the SED template coefficients using the catalog. But, as shown in Figure 8, we find that summarizing galaxy colors using SED templates could lead to a bias of colors for red galaxies up to 0.3 dex. We, therefore, adopt a different approach. Instead of building new SED template coefficients using the abundance-matched catalog, we build an observed color table as a function of the galaxy's observed z-band magnitude m_z and redshifts. We then use this table and repeat steps starting from Section 3.4.2 to build a new mock catalog.

While we use a special treatment of red-sequence galaxies in the conditional abundance matching method (detailed in Appendix G), the additional noise caused by this process makes the width of the red sequence too wide compared to the data. This too-wide red sequence could lead to overly pessimistic estimations of photometric errors of red galaxies and could cause massive background contamination of optical
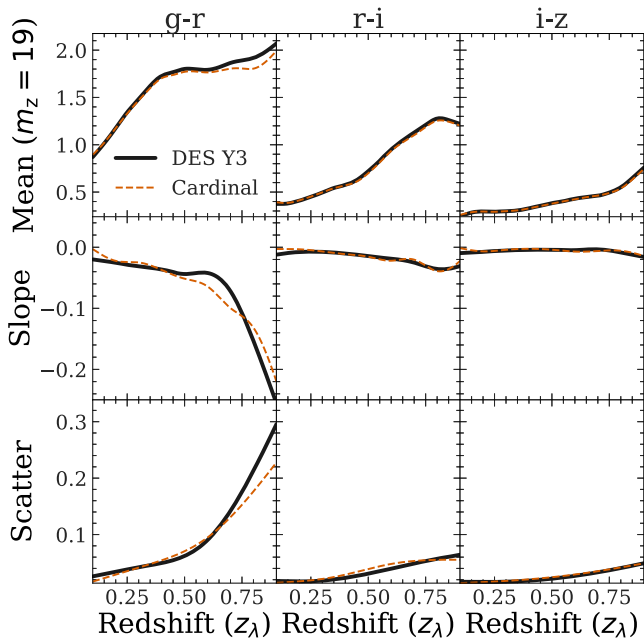
Figure 9. Comparison of red-sequence galaxies in Cardinal (red) and DES Y3 data (black) estimated by `redMaPPer`. Three columns correspond to g–r, r–i, i–z colors, respectively. The top row shows the mean colors at $m_z = 19$ as a function of redshift. The middle row shows the slope of colors–m_z relation as a function of redshifts estimated by `redMaPPer`. The bottom row shows the scatter of red-sequence colors. The largest discrepancy occurs at high-redshift $g - r$ colors, where the data is particularly noisy.

cluster identifications. To solve this problem, we further improve the consistency of the red sequence in mocks and the data using the algorithm presented in Appendix H. The resulting red sequence in Cardinal is compared with DES Y3 data in Figure 9. We find that the red sequences in simulations and data are very consistent.

After the color adjustment, we add observational noise to generate the new mock catalog using the method described in Section 3.5. The new mock catalog's overall color and magnitude distributions are shown as solid lines in Figure 7. The agreement between mocks and data is greatly improved. In most bands, we achieve a fractional error in galaxy luminosity functions of ~5%.

4. Comparison to DES Y3 Observations

Now that we have a realistic galaxy catalog with observed magnitudes down to survey depth limits, we proceed to select various cosmic structure tracers and compare their properties with the DES Y3 data.

4.1. redMaPPer Clusters

In this section, we compare clusters in mocks and DES Y3 data. As described in Section 3.2.1, `redMaPPer` is the main cluster sample in DES cluster cosmology analyses. Observationally, `redMaPPer` clusters are selected based on their richness (λ). For example, DES Y1 cosmology analyses (Abbott et al. 2020; To et al. 2021a) consider $\lambda > 20$ `redMaPPer` clusters as cosmological samples.

The simplest way to generate `redMaPPer` clusters in mocks is to select dark matter halos containing more than 20 bright red-sequence galaxies or some other value to match the observed abundance. However, the richness value of

`redMaPPer` clusters in the data contains significant components from correlated large-scale structure along the line of sight due to redshift uncertainties (Costanzi et al. 2019; Sunayama 2023). Further, Abbott et al. (2020), Sunayama et al. (2020), To et al. (2021b), and Wu et al. (2022) show that these projected components in richness can cause biases in the two-point correlation functions of `redMaPPer` clusters, including cluster lensing, cluster–galaxy cross-correlations, and cluster clustering. Therefore, counting the numbers of red-sequence galaxies within three-dimensional distance to halos in simulations is not the same as `redMaPPer` richness, making comparisons with observed `redMaPPer` clusters hard to interpret.

Costanzi et al. (2019), Sunayama et al. (2020), and Wu et al. (2022) present an improved way to simulate `redMaPPer` richness by counting the numbers of red-sequence galaxies within cylinders along the line of sight. This approach provides a numerically efficient way to simulate `redMaPPer` richness that does not require light-cone simulations. The problem with this approach is that `redMaPPer` does not weigh galaxies uniformly along the line of sight; instead, galaxies that are further away from clusters along the line of sight usually have smaller p_{mem}. Thus, counting galaxies within a cylinder is likely to overestimate the richness of clusters. One way to remedy this problem is to measure the p_{mem} distributions of galaxies as a function of spectroscopic redshifts. However, obtaining spectroscopic redshifts for galaxies down to $0.2L^*$ is challenging.

Finally, one could run `redMaPPer` algorithm on realistic mock galaxy catalogs to ensure an apples-to-apples comparison of simulated and observed clusters. However, the richness value of `redMaPPer` clusters depends on member galaxies' colors, magnitudes, and positions, making it hard to reproduce in simulations. Therefore, `redMaPPer` has only been applied to a limited number of mock catalogs (DeRose et al. 2019; Korytov et al. 2019).

Given the realistic red-sequence galaxies in Cardinal, we adopt this last approach to generate `redMaPPer` clusters. We select galaxy clusters using `redMaPPer` v0.8.4,[14] which includes several improvements relative to its predecessors (Rykoff et al. 2014, 2016; McClintock et al. 2019). This includes a complete adaptation of the code from IDL to Python and improved models of red-sequence galaxies. This additional improvement will affect the richness of `redMaPPer` clusters; therefore, we run the same `redMaPPer` on both `Buzzard` v2.0 (DeRose et al. 2022b) and publicly available DES Y3 data (Sevilla-Noarbe et al. 2021) to ensure apples-to-apples comparisons. Figure 10 compares the cluster abundances as a function of richness in four different redshift bins measured in Cardinal and DES Y3 data. We find that the data and Cardinal generally agree except for the lowest redshift bins. This might indicate differences between the redshift evolution of the richness–mass relations in Cardinal and the data. In comparison, the orange lines in Figure 10 show that the cluster abundances measured in Cardinal demonstrate a much better agreement with the data in all redshift bins. There are three main differences between the Cardinal model and the Buzzard model, which are

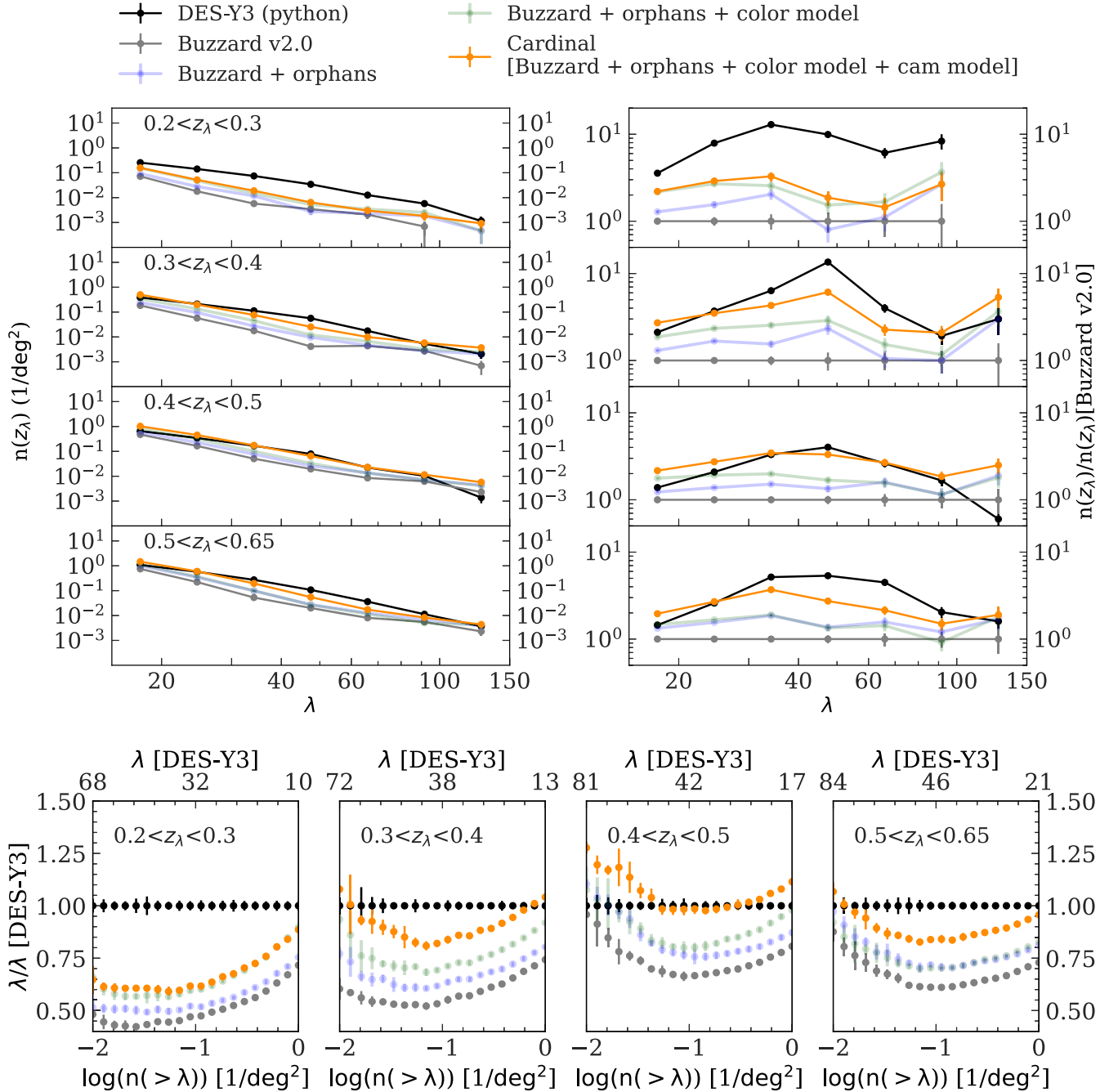1. The training SHAM model in Cardinal includes orphan subhalos as detailed in Section 2.

Figure 10. Comparison of cluster number density in Cardinal and DES Y3 data. The top left plot shows redMaPPer cluster abundances per square degree as a function of richness (λ) and redshift (z_λ). The top right plot shows the relative cluster number density of each realization relative to Buzzard v2.0. Four different rows in the top two panels show four redshift bins based on z_λ, the most probable redshifts of clusters estimated by redMaPPer. The black lines correspond to the DES Y3 data, and the gray lines correspond to Buzzard v2.0 (DeRose et al. 2022b). We note that in both cases, we re-run redMaPPer using the same version of redMaPPer as used in Cardinal to ensure apples-to-apples comparisons. The orange lines show the number of redMaPPer identified clusters in Cardinal. Evidently, one can see that the cluster abundances in Cardinal are much more consistent with the data. The blue and green lines show two different realizations that are intermediate steps from Buzzard v2.0 to Cardinal. Specifically, the blue line corresponds to a realization where the only difference in Buzzard v2.0 is including orphans in the SHAM model (detailed in Section 2). The green line represents one step forward by changing the galaxy color assignment models detailed in Section 3.4. Error bars show the Poisson noise. For further comparison, the bottom panel shows the richness ratio as a function of cumulative cluster abundance between simulations and DES Y3 data. The top x-axis shows the corresponding richness of DES Y3 data at a given cumulative number density. Error bars are estimated with 50 jackknife resampling.

2. The color assignment model that produces environment-dependent galaxy color distributions is improved. We detailed this step in Section 3.4.

3. We include an additional abundance matching step to address the remaining color inconsistencies due to the insufficient training spectra and SED templates. We detailed this step in Section 3.6.

To better understand how these steps affect the cluster abundances, we run redMaPPer on two additional realizations. First, we only update the training SHAM model in Cardinal but fix other color assignment models the same as Buzzard v2.0. The redMaPPer clusters identified in this realization are shown as the blue line in Figure 10. We can then compare this with clusters in Buzzard v2.0 to see the impact of
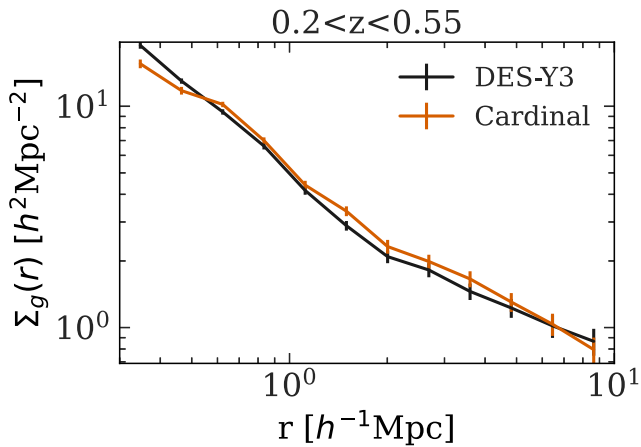
Figure 11. Stacked surface densities of galaxies around `redMaPPer` clusters with $\lambda > 20$ and redshift 0.2–0.55. Error bars show 1σ uncertainties estimated using 64 jackknife resampling. Black dots show measurements for DES Y3 galaxies and red dots show measurements for Cardinal.

SHAM models on the `redMaPPer` cluster abundances. We note that this comparison depends on redshift and richness. For the simplicity of the argument, let us focus on the $z_\lambda = [0.4, 0.5]$ bin. We find that including orphan models can boost cluster abundance with richness $\lambda > 20$ by roughly 50%, which is insufficient to account for the factor of three differences in cluster abundances between Buzzard v2.0 and the data. Second, we update the training SHAM and color assignment models for environment-dependent galaxy color distributions. The `red-MaPPer` clusters in this realization are shown as the green lines in Figure 10. Combining the color assignment models and the SHAM model can boost the richness by ∼100%. Finally, comparisons of the full Cardinal model and the green lines show that the additional color abundance matching step can boost the richness by another ∼100%, making the cluster abundances in Cardinal agree with the DES Y3 data. We, therefore, conclude that all three modifications of the Buzzard v2.0 model are needed to reproduce the observed cluster abundances. Cluster abundances are sensitive to richness value. To better illustrate the change of richness values that correspond to the change of abundances, we show the differences in richness value of Cardinal and Buzzard v2.0 at a fixed cluster abundance in Figure 10. In Appendix F, we further compare the conditional luminosity functions (CLFs) of satellites and centrals of `redMaPPer` clusters in Cardinal, Buzzard v2.0, and DES Y3 data.

Figure 11 further compares the stacked galaxy surface density around `redMaPPer` clusters in DES Y3 data and Cardinal, calculated using the method detailed in Appendix I. We find decent agreements between Cardinal and DES Y3 data. Compared to Buzzard v2.0 (e.g., Figure 8 in Wechsler et al. 2022), the one-halo regime agrees with the data much better. This is likely due to the better match of richness values for a given cluster number density in Cardinal. On the other hand, the consistency on large scales is very interesting. As shown in To et al. (2021a, 2021b), `redMaPPer`–`redMaGiC` cross-correlations in Buzzard have a much larger large-scale bias compared to DES Y1 data. If this excess of large-scale bias is related to galaxy cluster samples, Figure 11 suggests that Cardinal will not have this problem. We leave further investigations on this aspect in future work.

4.2. *redMaGiC Galaxies*

As described in Section 3.2.2, `redMaGiC` galaxies are one of the main galaxy samples in the DES cosmological analyses (Abbott et al. 2018, 2022). Given the realistic galaxy cluster properties in Cardinal, we apply the `redMaGiC` algorithm on Cardinal galaxies in the same way as `redMaGiC` was run on the DES Y3 data. Specifically, `redMaGiC` galaxies are selected using redshifts of `redMaPPer` clusters instead of the true redshifts in the simulations. This allows us to have realistic galaxy selection-related systematics in Cardinal. Following the DES Y3 cosmology analysis, we further bin `redMaGiC` galaxies into five tomographic bins with edges, [0.15, 0.35, 0.5, 0.65, 0.8, 0.9], using $z_{\rm redmagic}$, the redshift that maximizes $p(z_{\rm redmagic})$. We then Monte Carlo sample $p(z_{\rm redmagic})$ for each redshift bin to estimate their redshift distributions.

Figure 12 compares the redshift distributions of `redMaGiC` galaxies in Cardinal and the data. We find that the `redMaGiC` estimated redshift distributions are very similar to those in the DES Y3 data. This agreement is nontrivial. Although the same algorithm produces both samples, this algorithm is applied to two different galaxy catalogs. This similarity in redshift distributions demonstrates the realism of red galaxy properties in Cardinal. In the bottom panel of Figure 12, we compare the redshift distributions estimated by `redMaGiC` and the true redshift distributions in the simulation. While the redshift distributions estimated by `redMaGiC` are slightly narrower compared to the true redshift distributions, the overall shapes of redshift distributions are very consistent between `redMaGiC`'s estimations and the true distributions.

We proceed to access the clustering properties of `redMaGiC` in Cardinal. We first measure the galaxy correlation functions $w(\theta)$ using the weighted version of the Landy–Szalay estimator (Landy & Szalay 1993),

$$
\begin{aligned}
w(\theta_{\min} < \theta < \theta_{\max}) = &\sum_{i,j} \frac{w_i w_j}{\sum_i w_i (\sum_j w_j - 1)} \\
&- 2\sum_{i,R} \frac{w_i}{\sum_i w_i N_R} \\
&+ \sum_{RR} \frac{1}{N_R(N_R - 1)},
\end{aligned} \tag{26}
$$

where i, j goes through each galaxy pair that have angular separation $\theta_{\min} < \theta < \theta_{\max}$. The w_i and w_j are systematic weights associated with galaxies to remove spurious correlations of galaxies and survey depths. We detail the construction of these weights in Appendix D. The index R corresponds to randoms that describe the survey footprint. In this analysis, we make the number of randoms (N_R) 10 times greater than the number of galaxies in each tomographic bin to remove additional shot noise caused by the randoms. The i, R, and RR in Equation (26) go through each galaxy–random and random–random pairs that are separated by θ. We compare the measured $w(\theta)$ of Cardinal, Buzzard v2.0, and DES Y3 data in Figure 13. To avoid cosmic variance, we compare Cardinal and Buzzard v2.0 generated with the same dark matter simulation. In the lower two redshift bins, the `redMaGiC` clustering in Cardinal has a stronger one-halo term than Buzzard v2.0. This is likely due to the combination of changes in the color assignment and the inclusion of orphan galaxies. This is
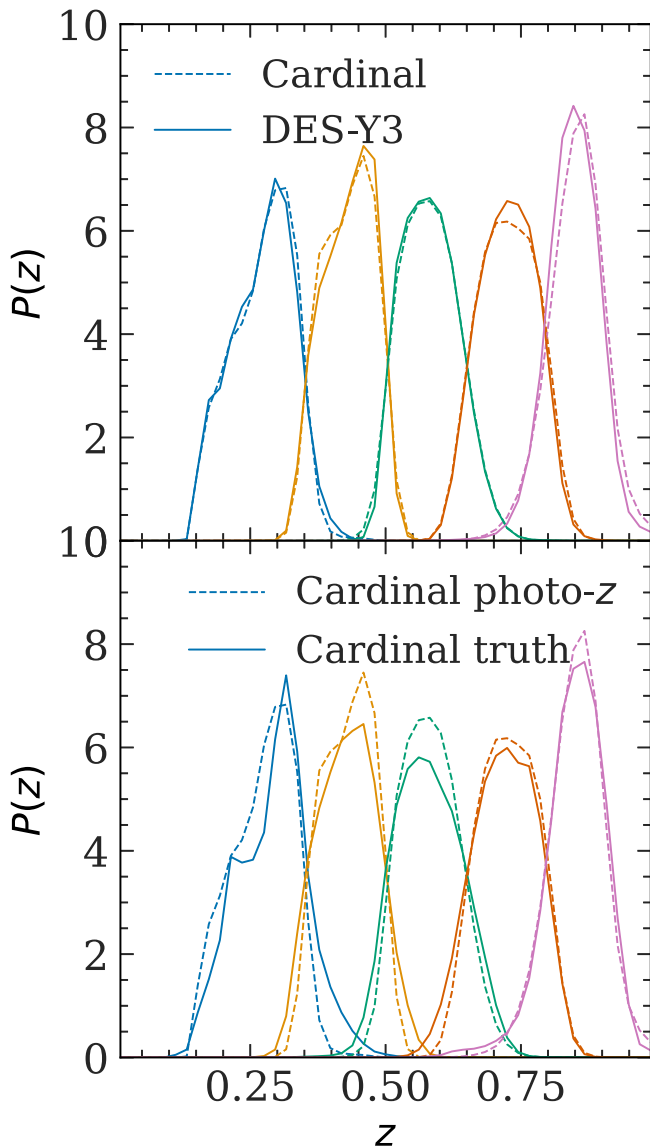
Figure 12. Comparison of redshift distributions. The top panel shows redshift distributions estimated by `redMaGiC` in Cardinal (dashed) and data (solid). The bottom panel compares true (solid) and photometric redshift distributions estimated by `redMaGiC` (dashed) in Cardinal. Different colors show different tomographic bins.

consistent with the finding compared to SDSS galaxies, which are galaxies at $z \simeq 0.1$ (Figure 5). For all redshift bins, the clustering in Cardinal is somewhat smaller than that of the DES Y3 data. To better understand the origin of this inconsistency, we compare the clustering of `redMaGiC` in Cardinal, and Buzzard v2.0 generated with the same dark matter simulation. We find that for the first and second redshift bins, the clustering of `redMaGiC` in Cardinal is consistent with Buzzard v2.0. For the third redshift bin, the deficit of clustering is likely due to the slightly worse photometric redshift performances in Cardinal compared to Buzzard v2.0. For the fourth and fifth redshift bins, we find that the `redMaGiC` samples in these redshift ranges are slightly fainter than Buzzard v2.0 and data, which could potentially explain the differences in clustering. Finally, to illustrate the impact of cosmic variance, we show the mean w (θ) measured in 18 Buzzard v2.0 realizations (magenta lines) in

Figure 13. Evidently, some of the differences between Cardinal and DES Y3 data are likely due to cosmic variance.

5. Conclusions and Outlook

Large simulations with realistic galaxies are increasingly essential in precision cosmological analyses based on large surveys. One of the widely used multipurpose simulations is the Buzzard simulation, which produces DES, Dark Energy Spectroscopic Instrument (DESI), and LSST-like mock catalogs out to $z = 2.35$ to a depth of $m_r = 27$ (DeRose et al. 2019, 2022b; Wechsler et al. 2022). This simulation played an essential role in the core cosmology analyses for DES and in planning and methodology development for several other surveys. In this paper, we introduce several model improvements to the Buzzard catalogs and generate a new set of mock catalogs, the Cardinal mocks. The main improvements are listed below, with a more detailed list in Appendix K.

1. We update the SHAM model used to generate the Buzzard simulation. The new SHAM model considers orphan galaxies and a flexible disruption model and incorporates mass-dependent scatter between galaxy luminosity and halo properties. For the first time, the SHAM model can simultaneously fit galaxy clustering and group–galaxy cross-correlations in three different luminosity thresholds measured in SDSS (Section 2).

2. A new color assignment model is developed to produce the environmentally dependent galaxy colors accurately. For the first time, the color assignment model can simultaneously fit color-dependent galaxy clustering in three different luminosity bins measured in SDSS (Section 3.4).

3. We identify two causes of the discrepancy between the color distribution in mocks and data. These include the need for redshift evolution in the spectroscopic training data and the insufficiency of summarizing galaxy colors using current SED templates. We provide a solution that uses photometric data and conditional abundance matching techniques. Applying this conditional abundance matching scheme, the apparent magnitude and color–color distributions are much more consistent with the DES Y3 data (Section 3.6).

4. We address the lack of lensing shear due to limited ray-tracing resolution. We develop a novel method that uses dark matter particle–halo cross-correlations to fix this problem. We find that the $\Delta\sigma$ around massive halos is more consistent with expectations after this correction is applied (Section 3.5.2).

We incorporate these improvements into the `ADDGALS` algorithm to generate the Cardinal mock, a one-quarter sky simulation out to $z = 2.35$ to a depth of $m_r = 27$. We further cut out one DES Y3 footprint and apply realistic DES Y3 photometric errors and sky backgrounds. The latest `redMaPPer` cluster finding algorithm and `redMaGiC` lens galaxy algorithm are also run on the catalogs to produce realistic DES Y3-like cluster and lens samples (summarized in Figure 1).

We compare the Cardinal mocks with DES Y3 and SDSS data. These comparisons include the abundance of `redMaPPer` clusters, the projected galaxy density profiles around `redMaPPer` clusters, the redshift distribution of `redMaGiC` galaxies, and galaxy clustering for various samples. In the cluster abundance comparison, we find that the Cardinal
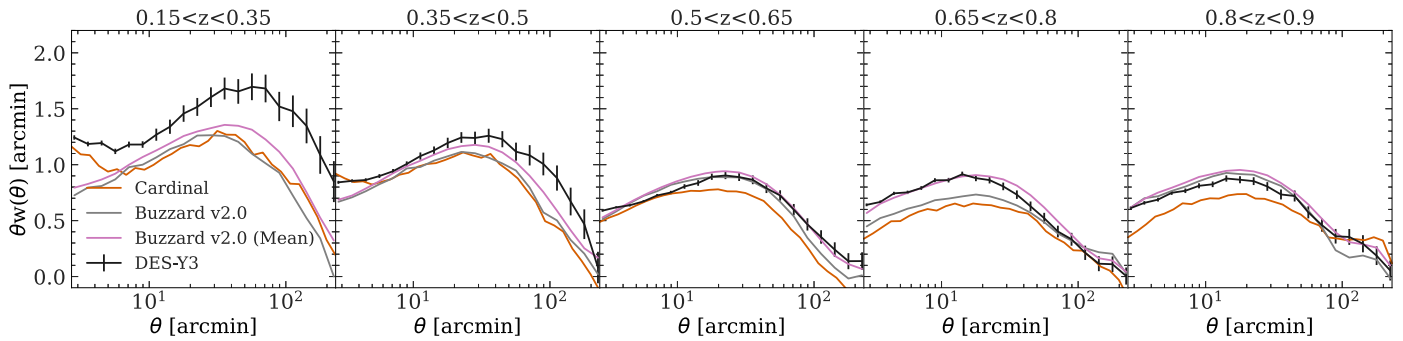
Figure 13. Comparison of `redMaGiC` clustering in Cardinal (red lines) and DES Y3 data (black lines). Five different panels show the five tomographic bins based on $z_{\rm redmagic}$, with edges at [0.15, 0.35, 0.5, 0.65, 0.8, 0.9]. Error bars show a 1σ confidence region. For comparison, gray lines show the galaxy clustering of Buzzard v2.0 using the same dark matter simulations as Cardinal. The magenta lines show galaxy clustering of Buzzard v2.0 averaged over 18 DES Y3 realizations (DeRose et al. 2022b).

clusters have a much more consistent number density to the data as a function of richness and redshift than those identified in the Buzzard simulation (Figure 10). We further make the comparison by changing one model component at a time. We find that two main factors contribute to the long-standing issue of lower cluster number densities in the Buzzard simulation, where the cluster abundance in simulations is 10%–25% of the cluster abundance in the data at the same richness. One of these is the previously postulated lack of orphans in the SHAM model (DeRose et al. 2019; Wechsler et al. 2022), and another is the galaxy color assignment model. We further find galaxy profiles around `redMaPPer` clusters in Cardinal are in excellent agreement with the data (Figure 11). As for the `redMaGiC` galaxy comparison, we find that the Cardinal algorithm produces reasonable `redMaGiC` galaxy properties compared to the DES Y3 data. These include the redshift distributions of the lens samples and galaxy clustering in each tomographic bin. Finally, we compare color-dependent galaxy clustering between Cardinal and SDSS data and find excellent consistency.

With the updated model, which includes more realistic small-scale galaxy clustering and lensing properties, we expect many applications using the Cardinal mock. We list a few potential applications below.

1. Optically selected cluster cosmology: Cardinal is a valuable tool to quantify the selection bias in optically selected clusters, which affects observables such as weak lensing profiles because cluster selection is based on richness. This bias has been identified as the dominant systematic in optical cluster cosmology (Sunayama et al. 2020; To et al. 2021a, 2021b; Wu et al. 2022). Cardinal is one of the very few mocks that the `redMaPPer` cluster finder can be run on, and it is so far the *only* mock that yields realistic `redMaPPer` clusters as a function of richness and redshift without artificially boosting cluster member galaxies. Future work will use Cardinal simulations to understand the selection functions of `red-MaPPer` clusters and the level of selection biases and to motivate a flexible model to mitigate these biases.

2. Photometric reshifts: Given the more consistent galaxy color distributions to the data (Figure 7), Cardinal is a valuable tool to assess the uncertainties of photometric redshifts (Myles et al. 2021a). For example, one could use a Cardinal mock to perform controlled experiments to quantify and develop programs to mitigate various photometric redshift calibration systematics (see, e.g.,

Newman & Gruen 2022, for a review). This includes (a) how heterogeneous selections of spectroscopic redshift samples and redshift errors in deep fields of cosmological surveys can affect photometric redshift calibrations of wide-field galaxies, (b) how astrophysical systematics can affect the effectiveness of cross-correlation redshifts, such as lensing magnifications, nonlinear clustering, and bias evolutions, and (c) how the sample variances due to limited area of deep fields propagate to redshift uncertainties of wide-field galaxies.

3. Quantify small-scale lensing systematics: The new empirical correction of the ray-tracing algorithm (Section 3.5.2) makes Cardinal accurately produce one-halo term lensing profiles. Cardinal can then serve as a tool to study systematic effects of the small-scale galaxy–galaxy and cluster lensings. This includes (a) developing methods to quantify boost factor measurements of cluster lensing (Varga et al. 2019) using different photometric redshift estimations and (b) validating small-scale galaxy–galaxy lensing and clustering models based on HOD or CLF.

4. End-to-end tests of multi-probe analyses: with realistic clusters, galaxies, and lensing properties in the same simulation, Cardinal is the ideal mock to develop a combine-probe analysis pipeline that aims to perform joint analyses of auto- and cross-correlations of these observables (MacCrann et al. 2018; To et al. 2021a, 2021b; DeRose et al. 2022b).

5. Defining optimal galaxy samples for clusters, galaxies, and lensing joint analyses: one of the limitations of forward modeling multi-probe analyses is that different cosmological probes rely on different galaxy samples. Under the framework of HOD models, different galaxy samples require a different set of galaxy–halo connection parameters, making parameter inferences harder and diluting cosmological signals. One interesting question is whether it is possible to define a common galaxy sample that serves as lens galaxies, source galaxies and parent samples for cluster finding. Cardinal can serve as a sandbox to develop this project.

Cardinal can be further improved in the following directions. First, the current model for generating the colors and magnitudes of galaxies that are dependent on their environment only considers information measured in the same snapshot, even though the formation history of a galaxy plays an important role in determining its colors and brightness. The

reason for this choice is purely computational: accurately determining the formation history of halos requires simulations with a high level of resolution, which in turn requires significant computing resources. However, recent developments in the hybrid Lagrangian perturbation theory model can help bypass this problem. For example, Modi et al. (2020) and Kokron et al. (2021) show that using the local density of dark matter with proper weights based on initial conditions can model galaxy clustering of mock galaxies generated using the HOD down to $k = 0.7\,h\,\mathrm{Mpc}^{-1}$. The success of these models in producing small-scale galaxy clustering indicates a significant amount of information about galaxies encoded in the initial conditions beyond the dark matter density measured in the same snapshot. Furthermore, Lucie-Smith et al. (2022) find that the initial conditions play an essential role in determining the inner part of the dark matter profiles. Thus, we expect that combining initial conditions and local density will significantly improve the accuracy of the color and magnitude assignment models. Moreover, incorporating initial conditions does not place additional requirements on simulation resolution, and thereby does not require significant additional computational power. Second, we find that the SED templates cannot summarize galaxy colors with uniform accuracy across colors. In Cardinal, we use DES photometric catalogs to solve this problem empirically. However, by doing this, we lose some ability to generate mock catalogs with arbitrary photometric bands. Future work will expand the SED templates using advanced stellar population synthesis models and DESI data. Third, we empirically address the lack of redshift evolution in the training spectra using the photometric data. Although this method works well in the end, it requires significant tuning to remove additional noise that broadens the red sequences. The main problem is that cluster finders require accurate red galaxy colors down to $0.2L^*$, which is $m_z = 21.8$ at $z = 0.6$. The spectroscopic training samples must go down to the same magnitude with uniform selections. This data is unavailable even with Stage 4 spectroscopic surveys (e.g., DESI). However, the Cardinal algorithm only requires a small area coverage for these spectroscopic samples. These requirements align well with the need for training spectra for photometric redshift calibrations. We anticipate that future campaigns to obtain photometric redshift training spectra will be helpful for generating more accurate simulations.

Finally, the current approach to generating simulations is bottom-up. We first obtain a decent luminosity-dependent galaxy clustering and group–galaxy cross-correlations. We employ conditional abundance matching techniques to paint colors to ensure reasonable color-dependent galaxy clusterings. We then perform several conditional abundance matching steps to fix problems in the color model when matching to photometric data sets. This approach has been successful but may not be scalable with the increasing amount of data. A top-down approach might scale better with a plethora of data coming in. With all the data sets and targeted summary statistics in hand, one can optimize a galaxy–dark matter connection model. Here, we specifically chose galaxy–dark matter connections instead of galaxy–halo connections because halo finding and halo definitions could suffer from limited resolution. For this top-down approach to perform well, work has to be done to avoid overfitting the data with an overly flexible model and to increase the interpretability of the constrained model. For the latter, one can derive conditional

probabilities from the constrained model and combine that with the bottom-up approach to gain physical insights. We leave this interesting direction for future work.

Although Cardinal was primarily developed for validating cosmological analyses, the method has the potential for broader applications. The ultimate goal of the program is to learn the physics of the universe from the vast amounts of data collected by cosmological surveys and use this knowledge to create a mock universe with higher fidelity. However, to achieve this goal, it is necessary to have high resolution to resolve important physical processes and sufficient volumes to leverage the full constraining power of the data, making the program computationally demanding. One of the critical features of Cardinal is connecting galaxy properties between high-resolution, small-volume simulations and low-resolution, large-volume simulations. While the current connection in Cardinal is limited to galaxy luminosity, one can extend this to other properties, such as rest-frame galaxy colors, sizes, ellipticities, gas densities, and temperatures. Using improved connections, replacing the SHAM training model with first principle-driven models (e.g., SAMs or hydrodynamical simulations), and employing emulator techniques, one can constrain physics of the baryonic component of our universe, the expansion histories of the universe, the growth of cosmic structures, and the interplays between these models using the full power of the survey data. At the same time, one can use this learned physics to create mocks with rich information to facilitate multiwavelength and multi-probe cosmology analyses. However, significant challenges remain in making the first principle-driven models of baryons computationally efficient, identifying relevant physics, and judiciously selecting informative observational data or summary statistics derived from them.

Finally, although we have been focusing on optical surveys, the Cardinal simulations can also inform many multiwavelength studies. The dark matter halos in Cardinal are resolved down to $M_{200\mathrm{m}} = 10^{13}\,h^{-1}\,M_\odot$, making the simulation sufficient to paint Sunyaev–Zel'dovich (SZ) and X-ray signals. Future work will generate realistic SZ and X-ray groups and clusters, making Cardinal useful for combined probe analyses of SZ, X-ray, and optical observations.

Appendix A
Varying Scatter in the SHAM Model

This section describes an algorithm to incorporate varying scatter in the SHAM model. Given an observed luminosity function $\hat{\phi}(L)$ and a halo mass function $N(M)$, one can relate the two functions via

$$\hat{\phi}(L) = \text{Poisson}\left(\int P(L|L')N(L')dL'\right),$$
$$N(L') := \int D(L', M)N(M)\, dM, \tag{A1}$$

where Poisson(x) represents a realization from the Poisson distribution with mean x, and $D(L, M)$ represents the mapping such that $N(>L) = N(>M)$. This mapping ($D(L, M)$) can be understood as assigning the brightest galaxy to the most massive halos. In Equation (A1), L' is the luminosity without scatter, and $P(L|L')$ represents the relations between L' and the observed luminosity of galaxies L. In most of the SHAM models, $P(L|L')$ is assumed as a Gaussian distribution with a constant scatter. Under this assumption, Equation (A1) can be considered a convolution problem. One can then obtain $N(L')$ from the observed $\hat{\phi}(L)$ using the Richardson–Lucy algorithm (Lucy 1974).

However, much evidence has shown that the scatter in $P(L|L')$ varies according to halo masses (Stiskalek et al. 2021; see Wechsler & Tinker 2018 for a review). Thus, we want to incorporate this modeling flexibility into our model. We adopt $P(L|L')$ as a Gaussian distribution with a scatter $\sigma(L')$ defined as

$$\sigma(L') = A + B(-2.5\log_{10}(L') + C), \tag{A2}$$

where A, B, and C are free parameters. With this parameterization, Equation (A1) remains linear. The linearity allows one to solve it using the same Richardson–Lucy algorithm, even though it cannot be phrased as a convolution problem.
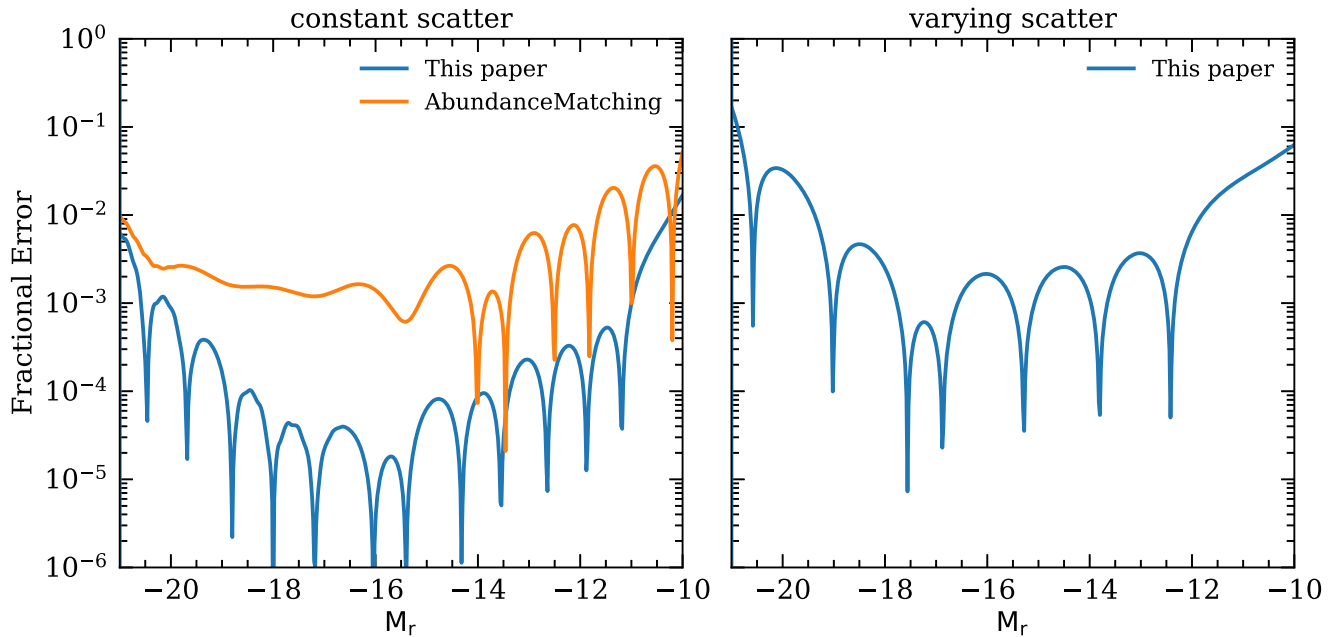
Figure 14. Performance of the algorithm. The y-axis shows the relative error caused by the Richardson–Lucy process, evaluated as $\mathrm{conv}(\mathrm{deconv}(\phi(L)))/\phi(L) - 1$, where conv shows the process in Equation (A1), `deconv` shows the Richardson–Lucy process, and $\phi(L)$ shows the original luminosity function. If $P(L|L')$ is a Gaussian distribution with a constant scatter, conv is a convolution operation and `deconv` is a deconvolution operation. The x-axis shows $M_r = 22.5 - 2.5 \log L$. In the first panel, we perform the process on a constant scatter (0.2 dex), where we can compare the performance of our implementation (blue line) to other existing codes (orange). In the second panel, we perform the process on varying scatter using the best-fit parameters in this paper.

Figure 14 compares the performance of our implementation of the Richardson–Lucy algorithm to a widely used abundance matching code `AbundanceMatching`. Evidently, our implementation reaches a better accuracy than the existing code. Further, we evaluate the performance of our algorithm in the case of nonconstant scatter. We find that our algorithm can achieve an accuracy of 10% for the best-fit model of this paper.

Appendix B
SHAM Emulator

We build an emulator for the SHAM model to facilitate the likelihood inference. We first generate the training data by Latin hypercube sampling of the priors described in Table 2. We then generate the data vector at each sample following the procedure described in Section 2.2.2. We build an emulator for each radial bin of the data vector. Specifically, given the training data and the associated data vector at a given radial bin, we construct the model using the polynomial chaos expansion implemented in `Chaospy` (Feinberg & Langtangen 2015). In short, we first model the objective function that maps

training data to the associated data vector as

$$f = \sum_\alpha c_\alpha m_\alpha, \tag{B1}$$

where c_α are free parameters and m_α is a set of orthogonal polynomials up to a certain degree constructed using the discretized Stieltjes procedure (Stieltjes 1884). We determine c_α by minimizing the mean squared error of the input data vector. Because m_α are polynomials, this minimization can be done analytically. In practice, we find that using 1000 training data and including polynomials up to the third degree is sufficient to achieve accuracy so that the uncertainties associated with the emulator are subdominant of the error budget.

To assess the accuracy of the emulator, we perform the leave-one-out test: iteratively removing one training point off the training sets, training an emulator, and calculating the differences between the removed training point and the emulator's prediction. Figure 15 shows the result of the leave-one-out test. The mean of the error is shown as white
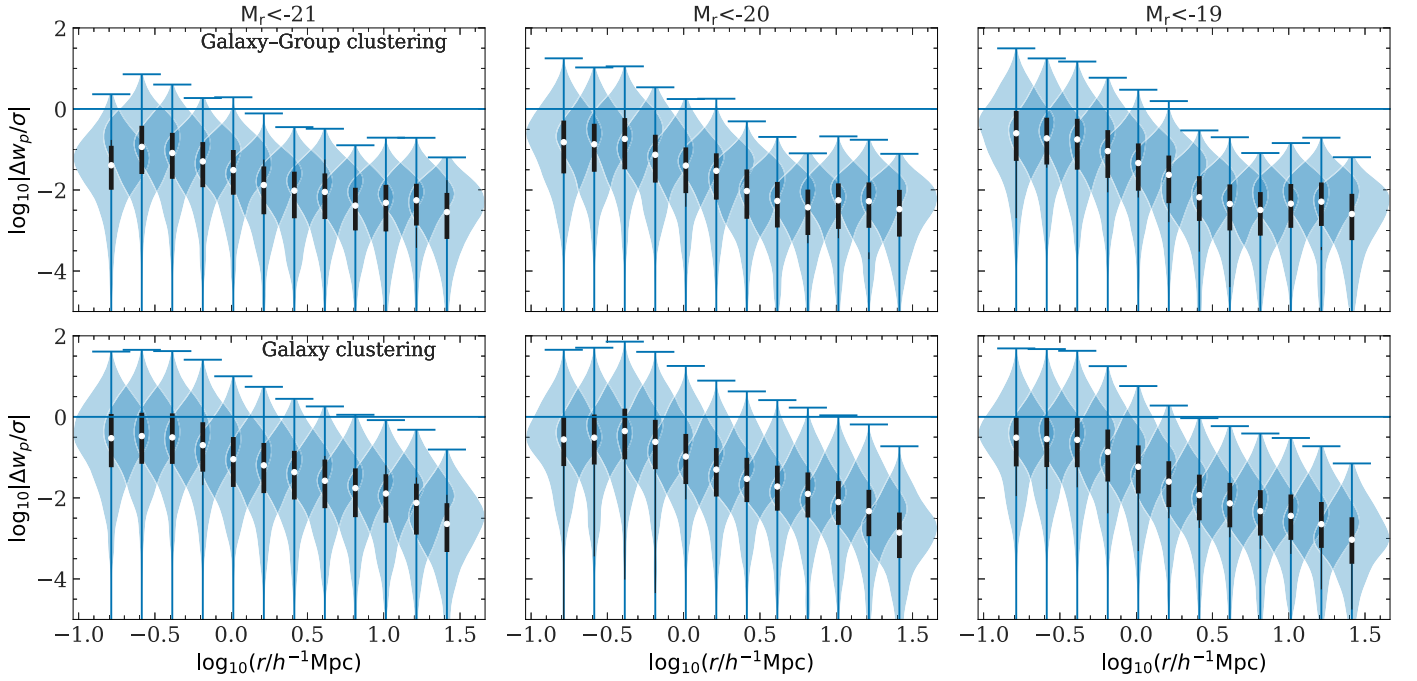
Figure 15. Accuracy tests of the SHAM emulators. The top row corresponds to galaxy–galaxy group correlations in each magnitude bin and the bottom row shows galaxy autocorrelations. Each panel shows emulator errors relative to the total error budget. Shaded blue regions are constructed by performing the leave-one-out test: removing one training data, training emulators, and calculating the error of the emulator at the removed point. The mean of the distribution is indicated by white dots and the black line shows the 25%–75% quantiles. Evidently, the emulator errors are much smaller than the total error budget in all radial bins and can therefore be ignored in the likelihood inferences.

dots. Evidently, the error is much smaller than the total error budget. We, therefore, ignore the emulator errors in the likelihood inferences.

Appendix C
Gaussian Process of the $P(R_\delta | M_r < x, z)$ Model

Following Wechsler et al. (2022), we employ a Gaussian process model to interpolate parameters $\theta(M_r, z)$ of the $P(R_\delta | M_r < x, z)$ model determined from the SHAM simulation. We first generate the training sample by measuring $\theta(M_r, z)$ on 18 evenly spaced magnitude bins from $M_r = -22.5$ to -18 at each snapshot of Chinchilla-T1. We then interpolate the training data using the Gaussian process algorithm implemented using George (Ambikasaran et al. 2015). That is, we model the joint probability of the targeted θ_p at parameter space $x = (M_r, z)$, and the training set θ_t as

$$\begin{bmatrix} \theta_p \\ \theta_t \end{bmatrix} \sim \text{Gaussian}\left(0, \begin{bmatrix} K(x, x) + \sigma_n^2 I & K(x, x_t) \\ K(x_t, x) & K(x_t, x_t), \end{bmatrix}\right), \quad (C1)$$

where $\theta_t = (\theta_1, \theta_2, ..., \theta_n)$ is the measured θ at $x_t = ((M_r, z)_1, (M_r, z)_2, ..., (M_r, z)_n)$, I is an identity matrix, σ_n

are free parameters to prevent overfitting. One can then find the θ_p that maximizes the probability and use it as the interpolated value of θ at $x = (M_r, z)$. The remaining task is then to determine K and σ_n. We model $K(x_1, x_2)$ a combination of exponential kernel and Matérn kernel that has the following forms:

$$K = \exp\left(\frac{-\|x_1 - x_2\|^2}{l_1^2}\right)$$
$$+ \left(1 + \frac{\sqrt{3}\|x_1 - x_2\|}{l_2}\right)\exp\left(-\frac{\sqrt{3}\|x_1 - x_2\|}{l_2}\right), \quad (C2)$$

where l_1 and l_2 are free parameters. We can then determine l_1, l_2, and σ_n by maximizing Equation (C1) evaluated on the training set θ_t. Once l_1, l_2, and σ_n are determined, we can predict θ_t at other parameters by maximizing Equation (C1) analytically.

Figure 16 shows the Gaussian process interpolations of the training data on grids of test points as well as the training data. We find that the Gaussian process model agrees with the training data and does not show signs of overfitting.
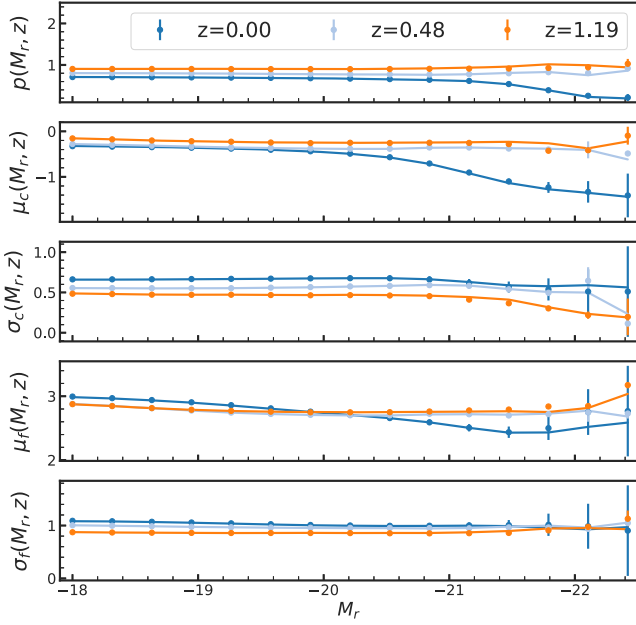
Figure 16. Parameters of the $P(R_\delta|M_r, z)$ model. The line shows the Gaussian process models for the redshift and magnitude dependence. Points with 1σ error bars show the measurement in the SHAM mocks.

Appendix D
Systematic Weight

The observed galaxy densities can have spurious correlations due to varying observing conditions across the sky. While we do not have most of the varying observing conditions in the simulations, such as dust reddening, seeing, and stars, we have spatially varying survey depths in the simulation. These spatially varying survey depths will produce spurious
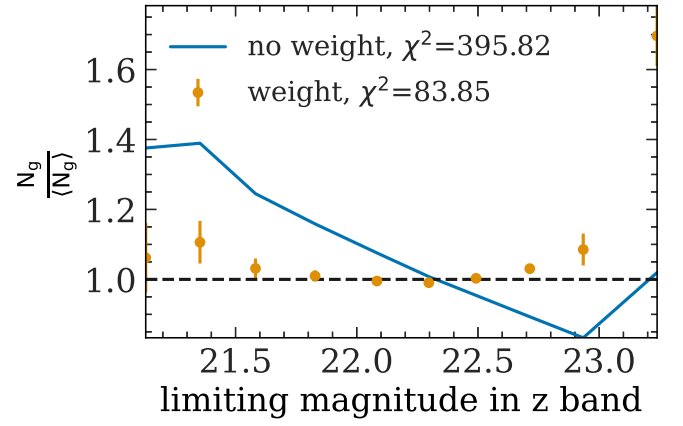

Figure 17. Performance of systematic weights on removing spurious correlations. The plot shows the observed pixel number densities evaluated on nside = 512 maps as a function of the limiting magnitudes of each pixel. The original `redMaGiC` number density is shown as the blue line, and the weighted number density is shown as the orange dots. Error bars show the 1σ error estimated with 150 jackknife resamples.

correlations of galaxies. We remove this signal following the methods described in Rodríguez-Monroy et al. (2022). First, we bin the galaxy density into `HEALpix` maps with nside = 512 corresponding to a resolution of $0°.11$. Next, we measure the relation of galaxy number density and the input survey depth s used to generate Cardinal. Finally, we fit a linear function to this relation. The fitted function reads as the form $F = ms + c$, where m and c are free parameters. The weight of each galaxy to remove the spurious correlations is then given by $1/F$. Figure 17 shows the performance of this algorithm. One can see that the weights significantly suppress the correlations between galaxy density and survey depths.

Appendix E
Validation of Cosmic Shear

In Section 3.5.2, we presented an algorithm to fix the small-scale lensing around halos (cluster lensing and galaxy–galaxy lensing) using particle–halo cross-correlations. In this appendix, we verify that this correction has a negligible impact on large-scale cosmic shears and makes small-scale cosmic shears less affected by ray-tracing resolutions. Figure 18 compares $\xi_{+/-}$ measured in Cardinal using corrected and original shears from ray tracing. The corrected and uncorrected $\xi_{+/-}$ are consistent on large scales. However, on small scales, the corrected $\xi_{+/-}$ are much more consistent with the theories based on CAMB, and Cosmic Emu calculated using the Core Cosmology Library (Chisari et al. 2019). This indicates that the $\xi_{+/-}$ calculated using corrected shapes is less affected by the limited resolution of ray tracing. This finding is consistent with the results in Takada & Bridle (2007), where the authors show that the lensing effect from massive clusters ($M > 10^{13} M_{\odot}$) significantly contributes to the one-halo term of cosmic shear power spectra.
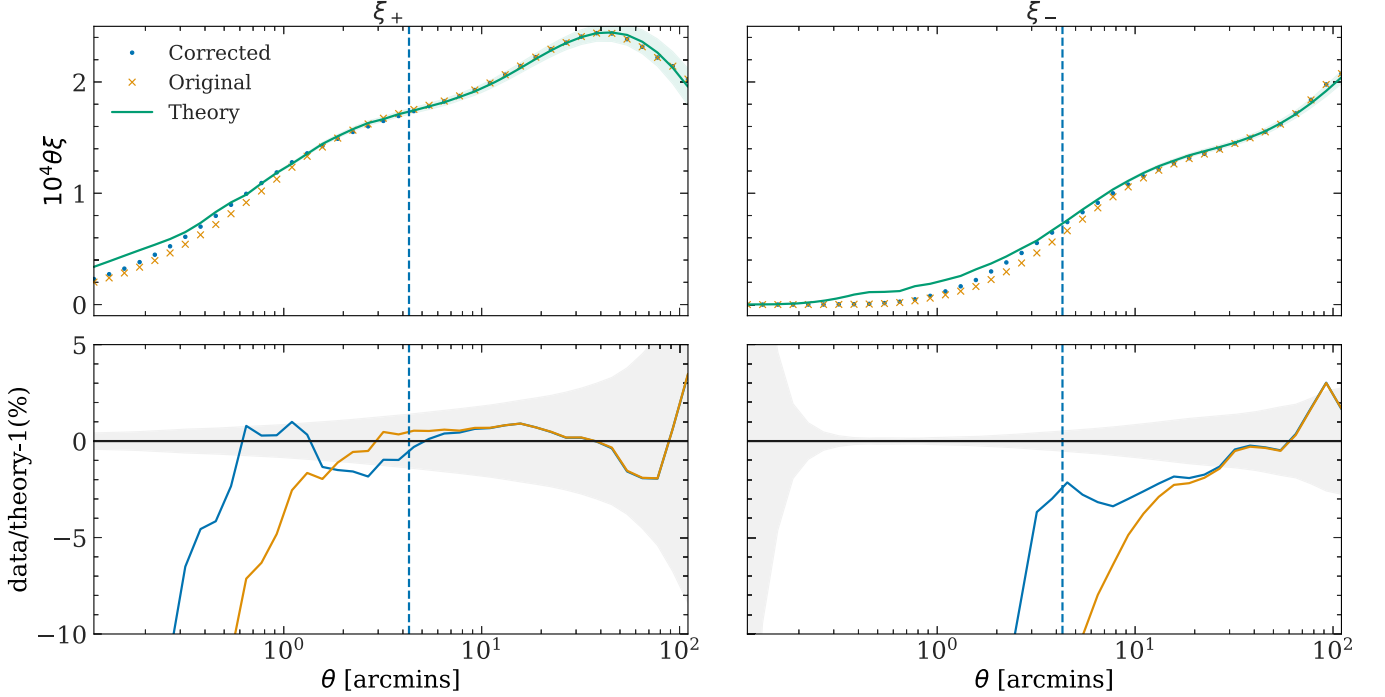
Figure 18. Comparison of $\xi_{+/-}$ calculated using original shear from ray-tracing (orange) and corrected shears (blue) of galaxies in Cardinal with $z = [0.6, 0.9]$. Green lines show theory prediction using CAMB and Cosmic Emulator. Shaded regions show 1σ error estimated with 204 jackknife resampling. Bottom panels show fractional differences between $\xi_{+/-}$ measurements and theory. The shaded region shows a 1σ error. The vertical dashed line shows 10 times the ray-tracing resolution.

Appendix F
Conditional Luminosity Function

We compare the conditional luminosity function in Cardinal, Buzzard v2.0, and DES Y3 data. The measurements follow the prescriptions detailed in To et al. (2020). Figure 19 shows the result. We find that for the lower two redshift bins, the central galaxy luminosity distributions are consistent between Buzzard v2.0 and Cardinal. For the highest redshift bin, the central galaxy luminosity in Cardinal is somewhat fainter than in Buzzard v2.0. However,

the relative brightness of centrals and satellites in Cardinal is more consistent with the DES Y3 data than with Buzzard v2.0. This is an important feature for the performance of cluster finders, as indicated in Kovacs et al. (2022) and To et al. (2021b). In terms of the satellites, we find that the satellite luminosity distributions in Cardinal are more consistent with the DES Y3 data than Buzzard v2.0 on the faint end. On the bright end, Cardinal and Buzzard's satellite luminosity distributions are consistent but are fainter than the DES Y3 data.
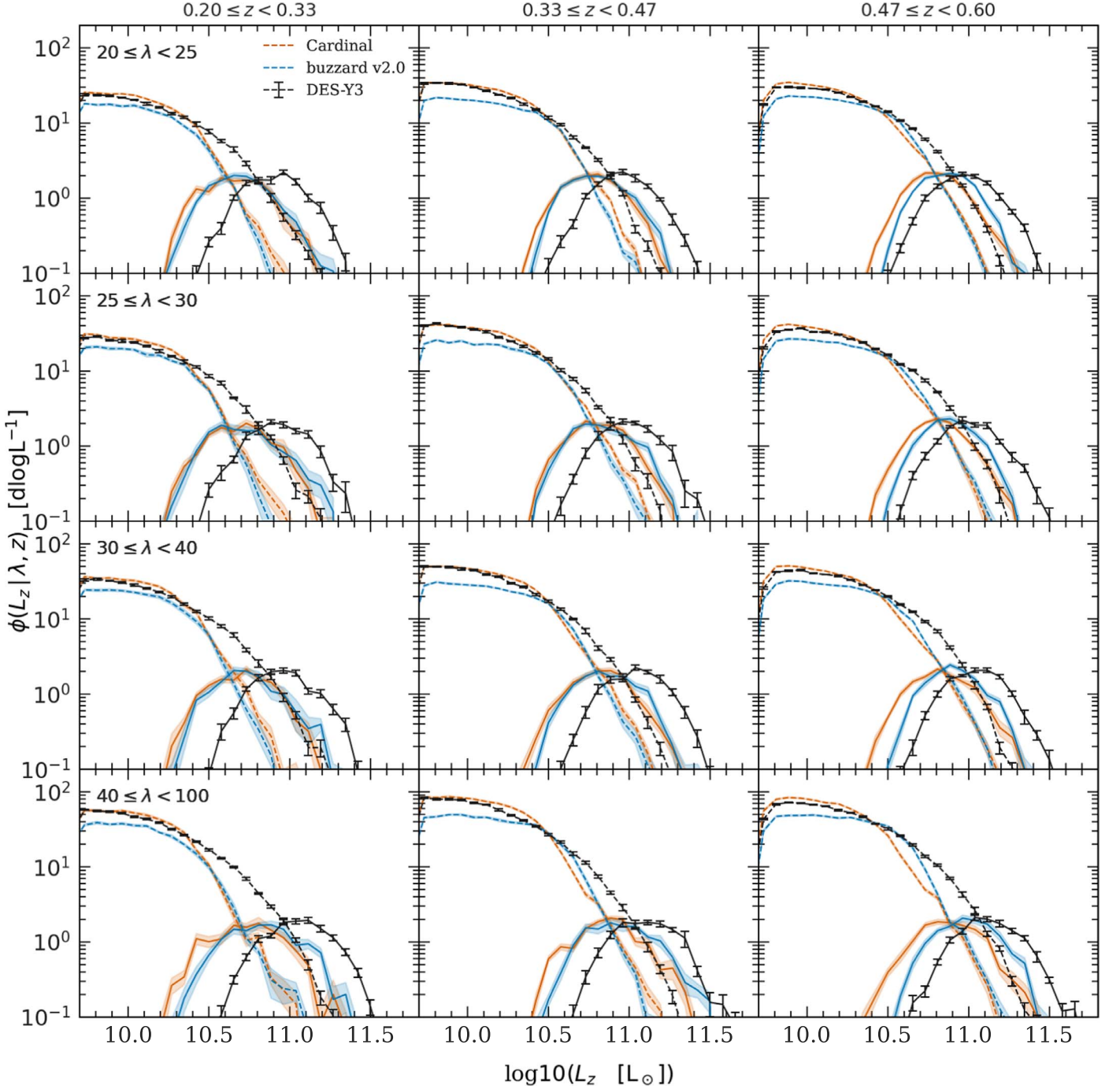
Figure 19. The conditional luminosity function for central galaxies (solid lines) and satellite galaxies (dashed lines) for `redMaPPer` clusters in Buzzard v2.0 and Cardinal (blue and orange lines, respectively) and the DES Y3 data (black lines). Different panels show the different richness and redshift bins, as indicated in the legend in each panel. Richness increases from top to bottom, and redshift increases from left to right. Error bars and shaded regions show the 1σ errors estimated from 50 jackknife resampling.

Appendix G
Algorithm of Avoiding Widening Red Sequence with Conditional Abundance Matching

Here, we present an algorithm that uses `redMaPPer` to avoid widening the width of the red sequence in the conditional abundance matching step. First, we run the `redMaPPer` cluster finder in the mocks generated before applying the conditional abundance matching. For each galaxy in the mock, we obtain a χ^2, quantifying the consistency of the multi-dimensional colors with the empirically constructed red-sequence model, and $z_{\rm red}$, the redshift that minimizes the χ^2 (see the details in Rykoff et al. 2014). For the galaxies consistent with the red sequence (i.e., small χ^2), $z_{\rm red}$ provides a good estimator of galaxies' redshift, with a typical uncertainty of ~ 0.02. Second, we abundance match this χ^2 measured in mocks to χ^2 measured in the data by enforcing the equality of $p(<\chi^2|m_z)$. In this way, we can avoid the possibility that χ^2 distributions in the mocks and data might differ. Third, we select red galaxies in the data and mocks using the abundance-matched χ^2. According to Rozo et al. (2015b), we select red galaxies as $\chi^2 < 20$. Third, we perform abundance matching separately for blue and red galaxies. For blue galaxies with $\chi^2 > 20$, we enforce the equality of Equation (25) between the mocks and data. For red galaxies with $\chi^2 < 20$, we use the $z_{\rm red}$ information. To avoid the possibility that $z_{\rm red}$ distributions in mocks and data differ, we first compute $z_{\rm a,red}$, the abundance-matched $z_{\rm red}$ in Cardinal by matching $p(<z_{\rm red}|m_z)$ between Cardinal and the data. We then match the overall color distributions of mocks and data by enforcing the equality of the following equation:

$$P(<c_i|m_z, c_{j<i}, z_{\rm a,red}). \tag{G1}$$

Appendix H
Algorithm of Matching Red Sequences in Mocks and Data

In this appendix, we describe an algorithm that matches red sequences in mocks and the data. We first run the `redMaPPer` algorithm on mocks and data to obtain a red-sequence model. This model has the following functional form:

$$
\begin{aligned}
P(\boldsymbol{c}|z, m_z) &\propto \exp(-0.5\chi^2), \\
\chi^2 &= (\boldsymbol{c} - \langle \boldsymbol{c}|z, m_z\rangle)(\boldsymbol{C}_{\rm int}(z) + \boldsymbol{C}_{\rm err}(z))^{-1} \\
&\times (\boldsymbol{c} - \langle \boldsymbol{c}|z, m_z\rangle)^T,
\end{aligned}
\tag{H1}
$$

where $\boldsymbol{c} = [g-r, r-i, i-z]$, $\boldsymbol{C}_{\rm int}(z)$ is the intrinsic scatter of the red sequence, and $\boldsymbol{C}_{\rm err}(z)$ is the observational noise. The mean color $\langle \boldsymbol{c}|z, m_z\rangle$ is modeled as a simple power law $c(z) + s(z)(m_z - m_p(z))$, where $c(z)$, $s(z)$, and $m_p(z)$ are free parameters. With $P(\boldsymbol{c}|z, m_z)$ models for mocks and data, we can enforce the consistency of red sequence using the following algorithm:

Algorithm 2.

---
for galaxies with redshift z and observed magnitudes $m_{g,r,i,z}$ **do**
  $m^o_{g,r,i,z} \leftarrow$ noisy realizations of $m_{g,r,i,z}$ using Equation (20).
  Calculate χ^2 using $m^o_{g,r,i,z}$ and $P_m(\boldsymbol{c}|z, m_z)$
  if $\chi^2 < 100$ and $R_n < R(z)$ and $M_n > M(z)$ **then**
    for $c \in \{g-r, r-i, i-z\}$ **do**
$$c += \frac{F(1-\frac{\chi^2}{100})}{G(z)}\left((c - \langle \boldsymbol{c}|z, m_z\rangle_m)\left(\frac{s_d}{s_m}\right)^2 + \langle \boldsymbol{c}|z, m_z\rangle_d - c\right)$$

(Continued)

---
    end for
    $m_g \leftarrow c_{g-r} + c_{r-i} + c_{i-z} + m_z$
    $m_r \leftarrow c_{r-i} + c_{i-z} + m_z$
    $m_i \leftarrow c_{i-z} + m_z$
  end if
end For

In the above algorithm, x_d represents the model evaluated using DES Y3 data, x_m represents the model evaluated using mocks, R_n is the distance to the nearest resolved halo whose mass is M_n, and s is the square root of the diagonal terms of $\boldsymbol{C}_{\rm int}(z)$. The $(1 - \frac{\chi^2}{100})$ term is to ensure a smooth transition between red galaxies and blue galaxies. F controls the nonlinear dependence of the color shift and $(1 - \frac{\chi^2}{100})$, which is given by $F(x) = 0.5 + 0.5\,{\rm erf}((x - 0.1)/0.05)$, where erf is the error function. The $(s_d/s_m)^2$ term ensures consistency of the scatter in the observed colors, which are proportional to the square root of the scatter of c due to the Poisson draws. The $G(z)$ removes additional noise in the red sequence introduced by the conditional abundance matching (CAM) method. This additional noise becomes large when the galaxies' magnitude approaches the survey depth limits. Therefore, $G(z)$ must be larger at higher redshifts when galaxies are fainter than their low-redshift counterparts. We empirically find that $G(z) = 2(0.5 + 0.5\,{\rm erf}((z - 0.8)/0.3) + 1.2$ produces reasonable galaxy properties. Further, shifting all galaxies with $\chi^2 < 100$ will produce too many red galaxies. We, therefore, control the number of red galaxies and their clustering with two additional cuts $M_n > M(z)$ and $R_n < R(z)$. Again, because the additional noise introduced by the CAM model becomes larger at higher redshifts, $M(z)$ should decrease, and $R(z)$ should increase with redshifts. If $M(z)$ becomes too small, the clustering of red-sequence galaxies will become too small. We increase $R(z)$ to ensure enough red galaxies and strong enough clustering. Empirically, we find

$$
\begin{aligned}
R(z) &= 6(0.5 + 0.5\,{\rm erf}((z - 0.54)/0.1)) + 1.2 \; h^{-1}\,{\rm Mpc} \\
M(z) &= 10^{-0.25(0.5+0.5\,{\rm erf}((z-0.65)/0.05))+12.90} \; h^{-1}\,M_\odot
\end{aligned}
\tag{H2}
$$

give reasonable red galaxy properties.

Appendix I
Galaxy Radial Profiles Around Clusters

This appendix details our analyses on constructing galaxy radial profiles around `redMaPPer` clusters. We construct radial profiles for `redMaPPer` clusters in simulation and data using exactly the same algorithm to ensure an apples-to-apples comparison. Our calculation pipeline is similar to the one presented in Chang et al. (2018) but has important differences in the details. We emphasize that it is necessary to recalculate the galaxy profile around `redMaPPer` clusters for data because `redMaPPer` has been sufficiently changed.

We wish to measure galaxy density profiles around `redMaPPer` clusters using photometric data. This density profile (Σ_g) can be related to correlation functions via the following equation:

$$\Sigma_g(R) = \langle \Sigma_g \rangle w(R), \tag{I1}$$

where R is the comoving distance to galaxy clusters, $\langle \Sigma_g \rangle$ is the mean density of the galaxy samples, and $w(R)$ are the correlation functions between galaxies and clusters. Without knowing galaxy redshifts, we cannot evaluate Equation (I1) directly. For a photometric data set like DES, we must construct an estimator that gives Equation (I1) but does not require knowledge of galaxy redshifts. For the simplicity of the argument, let us first assume galaxy cluster samples all have the same redshifts z_c. We can then define our estimator of $\Sigma_g(R)$ as

$$\hat{\Sigma}_g(R) = \left(\frac{\mathrm{DD_{CT}}(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_T} - 1 \right) \frac{N_T}{A}, \tag{I2}$$

where $\mathrm{DD_{CT}}$ denotes the number of pairs between clusters and galaxy samples in photometric data, $\mathrm{DD_{CR}}$ denotes the number of pairs between clusters and random points describing survey footprints, N_T is the total number of considered galaxies, N_R is the total number of randoms, and A is the survey area in the comoving unit at redshift z_c. In the above equation, θ is the angular separation of the pairs corresponding to a comoving distance R at redshift z_c.

We now show that $\hat{\Sigma}_g(R)$ is an unbiased estimator of $\Sigma_g(R)$. The full galaxy sample (T) can be divided into two groups: (1) galaxies physically associated with galaxy clusters (g) and (2) other galaxies (f). Because pair counting is an additive process, $\mathrm{DD_{CT}}$ is then a sum of $\mathrm{DD_{Cg}}$ (pairs between clusters and associated galaxies) and $\mathrm{DD_{Cf}}$ (pairs between clusters and other galaxies). Further, clusters and unassociated galaxies are not clustered, so $\mathrm{DD_{Cf}}$ is then $N_c N_f \zeta(\theta)$, where N_f is the number of unassociated galaxies and ζ is some geometric factor that depends on angular separation. By the same argument, $\mathrm{DD_{CR}}$ is $N_c N_R \zeta(\theta)$. With these, we have

$$\hat{\Sigma}_g(R) = \left(\frac{\mathrm{DD_{CT}}(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_T} - 1 \right) \frac{N_T}{A}$$

$$= \left(\frac{\mathrm{DD_{Cg}}(\theta) + \mathrm{DD_{Cf}}(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_g + N_f} - 1 \right) \frac{N_g + N_f}{A}$$

$$= \left(\frac{\mathrm{DD_{Cg}}(\theta) + N_c N_f \zeta(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_g + N_f} - 1 \right) \frac{N_g + N_f}{A}$$

$$= \left(\frac{\mathrm{DD_{Cg}}(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_g + N_f} + \frac{N_c N_f \zeta(\theta)}{N_c N_R \zeta(\theta)} \frac{N_R}{N_g + N_f} - 1 \right)$$

$$\times \frac{N_g + N_f}{A}$$

$$= \left(\frac{\mathrm{DD_{Cg}}(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_g} - 1 \right) \frac{N_g}{A}$$

$$= \left(\frac{\mathrm{DD_{Cg}}(\theta)}{\mathrm{DR_{CR}}(\theta)} \frac{N_R}{N_g} - 1 \right) \langle \Sigma_g \rangle, \tag{I3}$$

In the above equation, we can see that the terms in the parenthesis are the Davis–Peebles estimator (Davis & Peebles 1983) of $w(R)$. We note that the above derivation is only exact when considering the Davis–Peebles estimator. Using the Landy–Szalay estimator (Landy & Szalay 1993) to estimate $\Sigma_g(R)$ like the one described in Chang et al. (2018) would include additional terms and potentially lead to biases.

In survey data, not all clusters live at the same redshifts. We, therefore, follow the prescription described in Chang et al. (2018) for the full survey data. We first divide the redMaPPer clusters with $\lambda > 20$ from $z = 0.2$–0.55 into $\Delta z = 0.025$ bins. We estimate galaxy density profiles around clusters for each redshift bin ($\hat{\Sigma}_g(R)_i$) using Equation (I3). We then estimate the $\Sigma_g(R)$ using the weighted sum of $\hat{\Sigma}_g(R)_i$, with the number of clusters in each redshift bin as the weight. To avoid mixing galaxies with different luminosities across the entire redshift range, we further make a cut on galaxy samples at each redshift bin $M_r - 5\log(h) < -20$. The M_r is calculated assuming galaxies are at the same redshift as galaxy clusters.

In Cardinal, we repeat all the above calculations.

Appendix J
Color Template Bias Simulation

This section employs a simple simulation to offer valuable insights into the potential impact of the color biases presented in Figure 8 on galaxy color distribution. We accomplish this by generating mock galaxy samples through a combination of two Gaussian distributions. Specifically, we draw 100k blue galaxies with a mean $g - r$ color of 1 and a standard deviation of 0.3, followed by drawing 5k red galaxies with a mean $g - r$ of 1.65 and a standard deviation of 0.05. The resulting galaxy color distribution is presented as the blue histogram in Figure 20, designed to mirror the cosmos samples' color distribution at $z = 0.43$–0.63, as depicted in Figure 5 of DeRose et al. (2019). Next, we shift the color up by 0.1 magnitude, a typical value when comparing the $g - r$ color distribution between Buzzard and DES Y3 data. The resulting color distribution is illustrated as the orange histogram in Figure 20. Subsequently, we incorporate the color biases displayed in the right panel of Figure 8, resulting in a galaxy sample with the color distribution depicted as the black histogram in Figure 20. Strikingly, the black histogram reproduces two key features in Figure 5 of DeRose et al. (2019). First, the red-sequence galaxies corresponding to the black histogram have a similar mean color to those in the unaltered galaxy samples. Second, the color distribution of red-sequence galaxies is significantly more peaked than those of the unaltered samples. This simple yet compelling simulation further supports our hypothesis
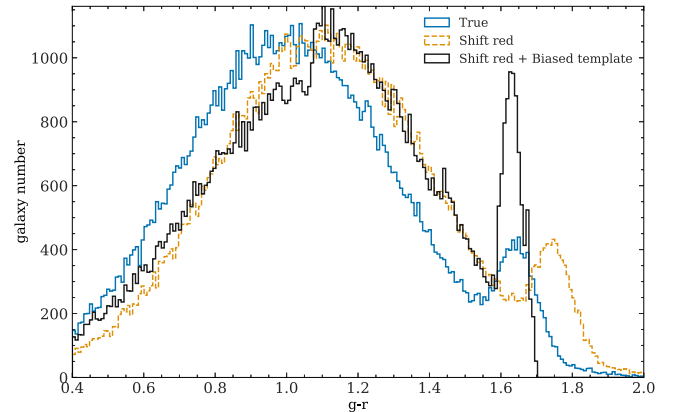

Figure 20. Color distributions of mock samples with the blue histogram representing the true color distribution. The orange histogram corresponds to the samples when colors are shifted by 0.1 magnitude. The black histogram shows the case when we summarize the color using SED templates. To generate the black histogram, we apply the biases shown in the right panel of Figure 8 to the galaxies in the orange histogram.

regarding the color discrepancies' origin between Buzzard and the data.

Appendix K
Differences between Cardinal and Buzzard v2.0

This section summarizes the main improvements from Buzzard v2.0 to Cardinal. We divide these improvements into four categories, summarized below.

1. SHAM:
   (a) We consider a luminosity-dependent scatter.
   (b) Orphan subhalos are included, whose abundances are controlled by a specially designed orphan model.
   (c) The group–galaxy cross-correlation function is used in addition to the galaxy–galaxy correlation function to constrain the model.
2. Environmentally dependent galaxy color model:
   (a) We consider a luminosity-dependent correlation between galaxy color and distances to nearby massive halos at fixed galaxy luminosity.
   (b) When using distances to nearby massive halos as a local environment proxy, we normalize them by the halo size.
   (c) We apply a nonlinear transform on distances to nearby massive halos to break the degeneracy of large-scale color-dependent clustering and galaxy color gradients around massive halos.
   (d) We introduce an additional parameter that controls the color of low-mass centrals.
3. Observational effects:
   (a) We increase the fidelity of faint galaxy photometric uncertainties by swapping the Gaussian distribution used to generate noise of galaxy fluxes to a Poisson distribution.
   (b) We empirically correct artifacts in the ray-tracing shears due to limited resolution.
4. Afterburner:
   (a) We demonstrate the limitations of `kcorrect` SED templates in describing red galaxy colors.
   (b) To mitigate this problem, we develop a new algorithm that employs the conditional abundance matching technique and a photometric data set to correct galaxy colors empirically.
   (c) We develop a new algorithm for matching the red sequence between mocks and data that takes into account the slope of the color–magnitude relations.

Appendix L
Environment Proxy Visualization

In Figure 21, we visualize Equation (15) with the parameters used to generate the Cardinal mock.
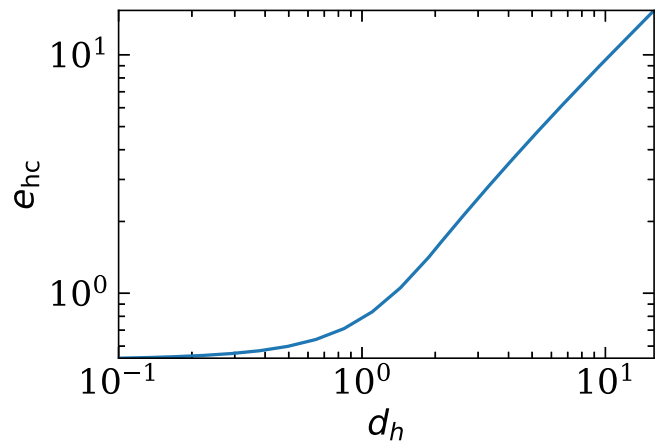

Figure 21. Visualization of Equation (15), with the parameters used to generate the Cardinal mock. The x-axis is the distance to the nearest massive halo, and the y-axis is the environment proxy used to generate galaxy color.

ORCID iDs

Chun-Hao To ⓘ https://orcid.org/0000-0001-7836-2261
Joseph DeRose ⓘ https://orcid.org/0000-0002-0728-0960
Risa H. Wechsler ⓘ https://orcid.org/0000-0003-2229-011X
Eli Rykoff ⓘ https://orcid.org/0000-0001-9376-3135
Hao-Yi Wu ⓘ https://orcid.org/0000-0002-7904-1707
Susmita Adhikari ⓘ https://orcid.org/0000-0002-0298-4432
Elisabeth Krause ⓘ https://orcid.org/0000-0001-8356-2014
Eduardo Rozo ⓘ https://orcid.org/0000-0002-1666-6275
David H. Weinberg ⓘ https://orcid.org/0000-0001-7775-7261

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, PhRvD, 98, 043526
Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2020, PhRvD, 102, 023509
Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, PhRvD, 105, 023520
Adhikari, S., Shin, T.-h., Jain, B., et al. 2021, ApJ, 923, 37
Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, ITPAM, 38, 252
Astropy Collaboration 2013, A&A, 558, A33
Astropy Collaboration 2018, AJ, 156, 123
Baugh, C. M. 2006, RPPh, 3101, 69
Becker, M. R. 2013, PhD thesis, Univ. Chicago
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al., 2012, Consistent Trees: Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology, Astrophysics Source Code Library, ascl:1210.011
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Blanton, M. R., Brinkmann, J., Csabai, I., et al. 2003, AJ, 125, 2348
Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734
Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, AJ, 129, 2562
Chang, C., Baxter, E., Jain, B., et al. 2018, ApJ, 864, 83
Chen, S.-F., White, M., DeRose, J., & Kokron, N. 2022, JCAP, 2022, 041
Chisari, N. E., Alonso, D., Krause, E., et al. 2019, ApJS, 242, 2

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Contreras, S., Angulo, R. E., & Zennaro, M. 2021, MNRAS, 508, 175
Costanzi, M., Rozo, E., Rykoff, E. S., et al. 2019, MNRAS, 482, 490
Davis, M., & Peebles, P. J. E. 1983, ApJ, 267, 465
DeRose, J., Becker, M. R., & Wechsler, R. H. 2022a, ApJ, 940, 13
DeRose, J., Wechsler, R. H., Becker, M. R., et al. 2019, arXiv:1901.02401
DeRose, J., Wechsler, R. H., Becker, M. R., et al. 2022b, PhRvD, 105, 123520
Feinberg, J., & Langtangen, H. P. 2015, JComS, 11, 46
Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015, MNRAS, 447, 1319
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
He, C.-C. 2021, ApJ, 921, 59
Hearin, A. P., Campbell, D., Tollerud, E., et al. 2017, AJ, 154, 190
Hearin, A. P., & Watson, D. F. 2013, MNRAS, 435, 1313
Hearin, A. P., Zentner, A. R., Berlind, A. A., & Newman, J. A. 2013, MNRAS, 433, 659
Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31
Hunter, J. D. 2007, CSE, 9, 90
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Jain, B., Seljak, U., & White, S. 2000, ApJ, 530, 547
Jarvis, M., 2015 TreeCorr: Two-point Correlation Functions, Astrophysics Source Code Library, ascl:1508.007
Karamanis, M., & Beutler, F. 2020, arXiv:2002.06212
Karamanis, M., Beutler, F., & Peacock, J. A. 2021, MNRAS, 508, 3589
Klypin, A., Gottlöber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 516, 530
Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
Kokron, N., DeRose, J., Chen, S.-F., White, M., & Wechsler, R. H. 2021, MNRAS, 505, 1422
Korytov, D., Hearin, A., Kovacs, E., et al. 2019, ApJS, 245, 26
Kovacs, E., Mao, Y.-Y., Aguena, M., et al. 2022, OJAp, 5, 1
Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lange, J. U., van den Bosch, F. C., Zentner, A. R., Wang, K., & Villarreal, A. S. 2019, MNRAS, 487, 3112
Lehmann, B. V., Mao, Y.-Y., Becker, M. R., Skillman, S. W., & Wechsler, R. H. 2017, ApJ, 834, 37
Lucie-Smith, L., Adhikari, S., & Wechsler, R. H. 2022, MNRAS, 515, 2164
Lucy, L. B. 1974, AJ, 79, 745
MacCrann, N., DeRose, J., Wechsler, R. H., et al. 2018, MNRAS, 480, 4614
Masaki, S., Lin, Y.-T., & Yoshida, N. 2013, MNRAS, 436, 2286
McClintock, T., Varga, T. N., Gruen, D., et al. 2019, MNRAS, 482, 1352
Modi, C., Chen, S.-F., & White, M. 2020, MNRAS, 492, 5754
Myles, J., Alarcon, A., Amon, A., et al. 2021a, MNRAS, 505, 4249
Myles, J., Gruen, D., Mantz, A. B., et al. 2021b, MNRAS, 505, 33
Nelson, D., Springel, V., Pillepich, A., et al. 2019, ComAC, 6, 2
Newman, J. A., & Gruen, D. 2022, ARA&A, 60, 363
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Pérez, F., & Granger, B. E. 2007, CSE, 9, 21
Potter, D., Stadel, J., & Teyssier, R. 2017, ComAC, 4, 2

Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30
Rodríguez-Monroy, M., Weaverdyck, N., Elvin-Poole, J., et al. 2022, MNRAS, 511, 2665
Rozo, E., & Rykoff, E. S. 2014, ApJ, 783, 80
Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, MNRAS, 461, 1431
Rozo, E., Rykoff, E. S., Bartlett, J. G., & Melin, J.-B. 2015a, MNRAS, 450, 592
Rozo, E., Rykoff, E. S., Becker, M., Reddick, R. M., & Wechsler, R. H. 2015b, MNRAS, 453, 38
Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104
Rykoff, E. S., Rozo, E., Hollowood, D., et al. 2016, ApJS, 224, 1
Rykoff, E. S., Rozo, E., & Keisler, R. 2015, arXiv:1509.00870
Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, ApJS, 254, 24
Sheldon, E. S., Johnston, D. E., Frieman, J. A., et al. 2004, AJ, 127, 2544
Shin, T., Adhikari, S., Baxter, E. J., et al. 2019, MNRAS, 487, 2900
Sinha, M., Berlind, A. A., & McBride, C. K. 2018, MNRAS, 478, 1042
Sinha, M., & Garrison, L. H. 2020, MNRAS, 491, 3022
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
Springel, V. 2005, MNRAS, 364, 1105
Stieltjes, T. J. 1884, Ann. Sci. de l'École Norm. Supérieure, 1, 409
Stiskalek, R., Desmond, H., Holvey, T., & Jones, M. G. 2021, MNRAS, 506, 3205
Sunayama, T. 2023, MNRAS, 521, 5064
Sunayama, T., Park, Y., Takada, M., et al. 2020, MNRAS, 496, 4468
Takada, M., & Bridle, S. 2007, NJPh, 9, 446
Takahashi, R., Hamana, T., Shirasaki, M., et al. 2017, ApJ, 850, 24
Teyssier, R., Pires, S., Prunet, S., et al. 2009, A&A, 497, 335
Tinker, J. L. 2020, arXiv:2007.12200
Tinker, J. L. 2021, ApJ, 923, 154
To, C., Krause, E., Rozo, E., et al. 2021a, PhRvL, 126, 141301
To, C.-H., Krause, E., Rozo, E., et al. 2021b, MNRAS, 502, 4093
To, C.-H., Reddick, R. M., Rozo, E., Rykoff, E., & Wechsler, R. H. 2020, ApJ, 897, 15
van den Bosch, F. C., & Ogiya, G. 2018, MNRAS, 475, 4066
Varga, T. N., DeRose, J., Gruen, D., et al. 2019, MNRAS, 489, 2511
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, NatRP, 2, 42
Wechsler, R. H., DeRose, J., & Busha, M. T. 2022, ApJ, 931, 145
Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435
Weinberg, D. H., Colombi, S., Davé, R., & Katz, N. 2008, ApJ, 678, 6
Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, MNRAS, 424, 232
White, M., Zhou, R., DeRose, J., et al. 2022, JCAP, 2022, 007
Wu, H.-Y., Costanzi, M., To, C.-H., et al. 2022, MNRAS, 515, 4471
Wu, H.-Y., Weinberg, D. H., Salcedo, A. N., Wibking, B. D., & Zu, Y. 2019, MNRAS, 490, 2606
Xu, H., Zheng, Z., & Guo, H. 2018, MNRAS, 481, 5470
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zhang, Z., Wu, H.-Y., Zhang, Y., et al. 2022, MNRAS, 523, 1994
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791