Combination of Cluster Number Counts and Two-Point Correlations: Validation on Mock Dark Energy Survey

Chun-Hao To  
Stanford University

Elisabeth Krause  
University of Arizona

Eduardo Rozo  
University of Arizona

Heidi Wu  
Boise State University

Daniel Gruen  
Stanford University

See next page for additional authors
Combination of cluster number counts and two-point correlations: validation on mock Dark Energy Survey

Chun-Hao To 1,2,3, Elisabeth Krause 4,5, Eduardo Rozo 5, Hao-Yi Wu 6,7, Daniel Gruen 2,3, Joseph DeRose 8,9, Eli Rykoff 2,3, Risa H. Wechsler 1,2,3, Matthew Becker 10, Matteo Costanzi 11,12, Tim Eifler 4, Maria Elidaiana da Silva Pereira 13, and Nickolas Kokron 1,2,3

(DES Collaboration)

1 Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
2 Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA
3 SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
4 Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
5 Department of Physics, University of Arizona, Tucson, AZ 85721, USA
6 Center for Cosmology and Astroparticle Physics, The Ohio State University, Columbus, OH 43210, USA
7 Department of Physics, Boise State University, Boise, ID 83725, USA
8 Center for Cosmology and Astroparticle Physics, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA
9 Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
10 Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA
11 INAF – Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy
12 Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy
13 Physics Department, Brandeis University, 415 South Street, Waltham, MA 02453, USA

Accepted 2021 January 21. Received 2021 January 4; in original form 2020 August 25

ABSTRACT

We present a method of combining cluster abundances and large-scale two-point correlations, namely galaxy clustering, galaxy–cluster cross-correlations, cluster autocorrelations, and cluster lensing. This data vector yields comparable cosmological constraints to traditional analyses that rely on small-scale cluster lensing for mass calibration. We use cosmological survey simulations designed to resemble the Dark Energy Survey Year 1 (DES-Y1) data to validate the analytical covariance matrix and the parameter inferences. The posterior distribution from the analysis of simulations is statistically consistent with the absence of systematic biases detectable at the precision of the DES-Y1 experiment. We compare the \( \chi^2 \) values in simulations to their expectation and find no significant difference. The robustness of our results against a variety of systematic effects is verified using a simulated likelihood analysis of DES-Y1-like data vectors. This work presents the first-ever end-to-end validation of a cluster abundance cosmological analysis on galaxy catalogue level simulations.

Key words: cosmological parameters – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

The simple cosmological model of a vacuum dark energy and Lambda cold dark matter (ΛCDM) is able to describe a variety of observations from the high- to low-redshift Universe. Despite its success, the two pillars of this model, dark energy and cold dark matter, lack a fundamental theory to connect to the standard model of particle physics. Without a compelling candidate for such a theory, one way to test the ΛCDM paradigm is by comparing its predictions to precise measurements of both the growth of structure and the expansion history of the Universe over the past several Gyr, when dark energy dominates the total energy budget of the Universe.

Commonly used probes of growth and/or cosmic expansion include Type-Ia supernovae, galaxy clustering, weak gravitational lensing, redshift-space distortions, and the abundance of galaxy clusters (see e.g. Weinberg et al. 2013 for a review). A large body of work has shown that the combination of these different probes is particularly powerful. For example, Abbott et al. (2018, hereafter DESY1KP) combine three two-point correlation functions – galaxy clustering, galaxy–galaxy lensing, and cosmic shear – resulting in a precise constraint on the growth of the structure. Similar analyses have also been carried out for the Kilo-Degree Survey (KiDS; Joudaki et al. 2018; van Uitert et al. 2018). In this work, we extend this type of analysis by incorporating cluster abundances and cluster-based large-scale structure statistics into the data vector of the combined probe analysis.

Galaxy clusters form at peaks of the primordial matter density field, and their space density over time reflects the gravitational growth of the coupled fluctuations of dark matter and baryons. As such, the abundance and spatial distribution of galaxy clusters are sensitive to the growth of structure and the expansion history of the

* E-mail: chto@stanford.edu (CHT); krausee@arizona.edu (EK)
Universe (see e.g. Allen, Evrard & Mantz 2011 for a review). Due to their independent information, different systematic uncertainties, and different degeneracies, it is expected that the combination of cluster statistics with other cosmological probes will yield cosmological constraints that are both more precise and more robust (e.g. Takada & Bridle 2007; Oguri & Takada 2011; Schaan, Takada & Spergel 2014; Lacasa & Rosenfeld 2016; Krause & Eifler 2017; Nicola, Dunkley & Spergel 2020; Salcedo et al. 2020). Critically, however, despite the extensive theoretical work on this front, no implementation of these techniques has been neither validated on realistic cosmological survey simulations, nor applied to data.

This paper presents an essential step towards accurate cosmological parameter constraints from combined cluster statistics. We develop a model and covariance matrix to combine cluster information with two-point correlation functions, including galaxy clustering, galaxy–cluster cross-correlation, cluster clustering, and cluster lensing. We validate the applicability of this model for a Dark Energy Survey Year 1 (DES-Y1)-like experiment, which comprises 1321–deg$^2$ area of the sky in five broad-band filters, $g$, $r$, $i$, $z$, and $Y$. In this study, we consider cluster samples built using the red-sequence Matched-filter Probabilistic Percolation cluster finder algorithm (redMaPPer; Rykoff et al. 2014) and galaxy samples built using the automated algorithm for selecting luminous red galaxies (redMaGiC; Rozo et al. 2016).

Much of the information on structure growth available in current surveys lies beyond the regime where the theoretical modelling is perturbative, making a theoretical prediction of the observations challenging. The theory is challenged further when using galaxies and galaxy clusters as dark matter tracers, since such analyses require a sufficient understanding of their statistical connection to the dark matter. Moreover, the overdensities of galaxies and galaxy clusters can be subject to significant systematic biases due to observational constraints. For galaxy clusters in particular, it is known that photometrically selected samples suffer from projection effects: massive dark matter haloes are more easily identified as galaxy clusters when their own galaxy overdensities are enhanced in the plane of the sky by the projections of unassociated galaxies along the line of sight (Costanzi et al. 2019a; Sunayama et al. 2020). This selection effect biases both the observed galaxy and matter overdensities about the selected galaxy clusters relative to randomly selected haloes of the same mass. A similar argument leads one to conclude that orientation biases, the major axes of detected clusters aligned with the line of sight, must also be present in photometrically selected cluster samples (Wu et al., in preparation). Without a proper model of these systematics, the cosmological constraints from cluster abundances would be biased (Abbott et al. 2020, DES2020 hereafter).

The existence of important observational systematic biases implies that robust cosmological analyses that rely on galaxies and galaxy clusters should be tested on simulated data sets that explicitly incorporate as many of these systematics as possible (see e.g. MacCrann et al. 2018, for an example applied to the DES 3 × 2 point measurement). Such simulations are intended to provide plausible realizations of a given cosmology, allowing one to test the robustness of the analysis against both theoretical and observational systematics. This approach is particularly powerful when one considers the essentials of a blind analysis of survey data: simulations allow us to finalize analysis choices on simulated data sets prior to applying the method to the real data.

However, there is an important caveat to this approach. No simulation is perfect. When analysing synthetic data, it can be difficult to disentangle biases in cosmological parameters coming from flaws in the analysis from those due to differences between simulations and data. The latter is often due to uncertainties in the underlying galaxy population or in the models for how galaxies trace the underlying dark matter density field (Wechsler & Tinker 2018). In this paper, we take a conservative approach and use three sets of cosmological simulations, each populated with galaxies using different assignment schemes, to develop and test our theoretical model. Specifically, we will demonstrate that the theoretical model developed in this work is capable of correctly recovering the underlying cosmological parameters of several simulated data sets irrespective of the details of the galaxy population model.

This work considers four tracers that can be measured by an imaging survey: the abundance of galaxy clusters, the spatial distributions of galaxies and galaxy clusters, and the lensing shear field ($\gamma$). These tracers are related to fluctuations of the matter density field, making them sensitive to the growth of structure in the Universe. Because haloes that host galaxy clusters form from rare peaks in the matter density field, their abundance (the halo mass function) is highly sensitive to the amplitude of matter fluctuations in the Universe. By binning galaxy clusters based on an observable proxy for halo mass (such as the richness of redMaPPer clusters), one may simultaneously calibrate the relation between this observational proxy and halo mass, as well as the underlying cosmological parameters.

Turning to the spatial distribution of galaxies and clusters, the main challenge to our ability to extract cosmological information from the corresponding correlation function is that both galaxies and clusters are biased tracers of the matter density field. Fortunately, on sufficiently large scales, their overdensities are simply proportional to the matter fluctuations. Even then, however, the amplitude of matter fluctuations is degenerate with the galaxy and cluster bias.

The above degeneracy can be broken using weak gravitational lensing. Weak gravitational lensing shear is the coherent distortion of the shapes of distant galaxies (often called source galaxies) due to fluctuations of the matter density along their line of sight. The cross-correlation of shear and galaxy cluster, often called cluster lensing, is related to cluster–matter cross-correlation. On large scales, the cluster–matter cross-correlation is linearly related to the matter two-point correlation function via the cluster bias. Since the cluster lensing and cluster two-point correlations depend on the matter two-point correlation function with different powers of bias, they are complementary to each other: the combination of galaxy and cluster clustering with cluster weak lensing enables us to self-calibrate the clustering bias and measure the amplitude of matter fluctuations simultaneously.

This paper is organized as follows: In Section 2, we detail the construction of cosmological survey simulations, including a brief summary of the creation of the mock catalogues, a description of the sample selection, and a comparison of galaxy cluster properties in different versions of the mock catalogues. In Section 3, we present a model describing the redMaPPer selection bias, an additional large-scale bias that can be present if the redMaPPer cluster finder preferentially selects clusters with properties that are correlated with the mass observables. We detail the model and covariance matrix calculation in Section 4. In Section 5.1, we describe the construction of the data vector in this analysis from galaxy and galaxy cluster catalogues. We summarize the analysis choice, including minimum scale cuts, estimation of samples’ redshift distributions, the procedures to generate covariance matrices, and final model parameters in Section 5.2. In Section 5.3, we test the robustness of the constraints on $\Omega_m$ and $\sigma_8$ against various potential systematics. In Sections 5.4
and 5.5, we summarize the main results from analysing simulated data sets, including an estimation of the theoretical systematic of our analysis pipeline, and a check of the validity of the theoretically derived covariance matrix employed in our analysis. Section 6 summarizes our findings.

2 SIMULATIONS AND SAMPLE SELECTION

This analysis uses the BUZZARD mock catalogues, which are described in detail in DeRose et al. (2019) and Wechsler et al. (in preparation). Here, we briefly summarize the key characteristics of the simulations and focus on the properties that are related to the performance of the cluster finder.

The creation of the BUZZARD mock catalogues involves six steps. First, the $N$-body simulation is generated assuming a flat $\Lambda$CDM cosmology with $\Omega_m = 0.286$, $\sigma_8 = 0.82$, $\Omega_b = 0.046$, $h = 0.7$, and $n_s = 0.96$. Secondly, the galaxies are populated into high-resolution $N$-body simulations by the subhalo abundance matching model presented in Lehmann et al. (2017), which matches brighter galaxies to haloes with higher peak circular velocities while allowing for some scatter between the two. Thirdly, the model connecting galaxies’ $r$-band absolute magnitudes ($M_r$) and local matter density is generated based on galaxies populated in the second step. In addition, a $M_r$–halo mass relation is fitted to central galaxies residing in resolved haloes. Fourthly, the low-resolution $N$-body light-cone simulation is populated with galaxies in the following ways: central galaxies are populated on resolved haloes according to the $r$-band magnitude–halo mass relation; satellites and field galaxies are populated on dark matter particles according to the $r$-band magnitude–local matter density relation obtained from the previous step. Fifthly, a spectroscopic sample of galaxies is used to populate a spectral energy distribution (SED) to each galaxy. This procedure is done by ranking galaxies in the spectroscopic data and galaxies in the simulation by their distances to the fifth nearest galaxy ($\Sigma_5$). In each magnitude bin, the SEDs of the spectroscopic galaxies are put on the simulated galaxies that have the same ranking. Sixth, we apply the DES survey depth and photometry uncertainties to each galaxy. We then run ray-tracing code CALCLENS\(^1\) (Becker 2013) to obtain the lensed magnitudes and galaxy shapes. The Bayesian photometric redshift (BPZ; Benitez 2000) method is then run to obtain the photometric redshift of each galaxy (BPZ is the fiducial method used to estimate photometric redshifts of the source galaxies in DESY1KP, DES2020).

In this paper, we adopt the set of simulations created with this procedure as the baseline simulation (BUZZA, version 1.9.2), which contains 11 realizations of the DES-Y1 survey created from two sets of $N$-body simulations. In each realization, we run redMaGiC and redMaPPer on the galaxies in the same way as is done on the data. The end products are redMaGiC galaxies, redMaPPer clusters, and the shapes and photometric redshifts of all galaxies.

The galaxies in BUZZA are found to have red-sequence colours with less scatter at fixed redshift than what is observed in the DES-Y1 data (see fig. 11 in DeRose et al. 2019). We expect that a less-scattered red sequence will lead to a more mild projection effect in redMaPPer clusters. This is because a red sequence with less scatter helps redMaPPer distinguish cluster galaxies from foreground and background contamination. To verify this expectation, we calculate the fraction of galaxies along the line of sight of a redMaPPer cluster that would be counted as member galaxies of the given cluster as a function of their redshift separations. The width of this distribution $S_z$, called $\sigma_z$; in DES2020, is expected to directly relate to the line-of-sight length-scale within which redMaPPer counts galaxies as cluster members. This is also the quantity used to construct the projection effect model in DES2020. In this paper, we follow the same procedure as in DES2020 and Costanzi et al. (2019a) to measure the $S_z$ in the BUZZARD mock catalogues. As shown in Fig. 1, the value of $S_z$ for redMaPPer clusters in BUZZA is smaller than in the data, consistent with our expectation from the width of the red sequence in BUZZA.

One of the goals of this paper is to test the robustness of the model developed in this paper against systematics introduced by the redMaPPer cluster finder. Therefore, we create a new simulation, BuzzB (version 1.9.2 + 2), which increases the impact of projection

\(^1\)https://github.com/beckermr/calc lens

Figure 1. Projection effects and orientation biases in different versions of the simulations. Left-hand panel: Mean $S_z$ as a function of redshift. $S_z$, defined in Section 2, is expected to relate to the projection effects of redMaPPer clusters. The error bar is the error in the mean. The black line represents the measurement in DES-Y1 data (DES2020). The blue, green, and orange lines correspond to the measurements in each of the three different versions of BUZZARD. The differences between these simulations are summarized in Table 1. Right-hand panel: Mean $\cos \theta$ for redMaPPer clusters, where $\cos \theta$ is the cosine of angle between the halo’s major axis and the line of sight. The blue, green, and orange lines are the measurement in three different versions of BUZZARD. Again, error bars represent the error on the mean. In all three versions of BUZZARD, $\cos \theta$ is greater than 0.5, indicating that the redMaPPer cluster finder selects clusters that are preferentially oriented along the line of sight. The amount of orientation is consistent across simulations despite very different underlying galaxy–halo connections.
Table 1. Summary of the simulations adopted in this analysis. First, we employ three different galaxy–halo connection models: in BuzzA, we employ our baseline model; in BuzzB, we adjust the width of red sequence by adding scatter to red galaxies’ luminosity; in BuzzC, we adjust the colour-dependent galaxy clustering and the width of red sequence. Secondly, in this analysis, we run redMaPPer in two different modes. In ‘Fullrun,’ redMaPPer is run treating BUZZARD galaxies as real data, and the run includes both cluster finding, centre identification, and richness calculation. In ‘Halonur,’ redMaPPer is run fixing the cluster centres at the halo centres to avoid the ambiguity introduced when associating galaxy clusters to dark matter haloes; richness is calculated at the halo centres. A comparison of these two different run modes of redMaPPer is presented in Appendix B. Halonur is used to develop the selection bias model. For the rest of the analysis in this paper, we use redMaPPer catalogues in the Fullrun mode.

| Simulation name | BuzzA | BuzzB | BuzzC |
|----------------|------|------|------|
| BUZZARD version number | v1.9.2 | v1.9.2 + 2 | v1.9.8 |
| RedMaPPer mode | Fullrun/Halonur | Fullrun/Halonur | Fullrun/Halonur |
| Footprint | DES-Y1 | DES-Y1 | DES-Y3 |
| Survey depth | DES-Y1 | DES-Y1 | DES-Y3 |
| Number of realizations | 10 | 10 | 1 |

2.1 Sample selection

We select two galaxy samples and one cluster sample from the simulations. The first galaxy sample is comprised of redMaGiC galaxies, obtained by running the redMaGiC algorithm (Rozo et al. 2016) on the simulations with the same settings as the DES-Y1 run. We then cut galaxies with redshift $z > 0.6$, the highest redshift of the redMaPPer clusters. We further split the galaxies into three bins using the redMaGiC photometric redshift ($z_{RMG}$) estimate: $0.15 < z_{RMG} < 0.3$, $0.3 < z_{RMG} < 0.45$, and $0.45 < z_{RMG} < 0.6$. These redshift bins are consistent with the first three redshift bins of lens galaxies in DESY1KP. Since we focus exclusively on galaxies with redshift less than 0.6, we use the redMaGiC high-density sample (luminosity, $L > 0.5L_*$; number density, $n = 1 \times 10^{-3} \ h^3 \text{Mpc}^{-3}$) for this analysis.

The second galaxy sample consists of source galaxy samples. Here, we do not run through the source galaxy selection procedure as described in Zuntz et al. (2018), which requires performing image simulations of the BUZZARD mock catalogues. Instead, we use a procedure similar to that described in DeRose et al. (2019), applying size and magnitude cuts to yield a similar source density as the DES-Y1 data. The cuts we apply are

(i) Mask all regions where the limiting magnitude and point spread function (PSF) size cannot be estimated,
(ii) $\sigma(m_{r,i,z}) < 0.25$,
(iii) $\sqrt{r_{gal}^2 + r_{psf}^2} > 1.05 r_{psf}$, and
(iv) $m_{i} < 21.25 + 2.13 z$, 

where $\sigma(m_{r,i,z})$ are magnitude errors in the $r, i, z$ bands, $r_{psf}$ is the $i$-band PSF full width at half-maximum (FWHM) estimated from the data at the position of each galaxy, $r_{gal}$ is the half-light radius of the galaxy, and $z$ is the BPZ photo-$z$ of each galaxy. Note that these cuts are slightly different from the cuts in MacCrann et al. (2018) and DeRose et al. (2019). We find that these cuts reproduce better the galaxy number densities in the data. We then use the BPZ photo-$z$ to split the samples into four redshift bins, defined as $0.2 < z < 0.43$, $0.43 < z < 0.63$, $0.63 < z < 0.9$, and $0.9 < z < 1.3$.

The cluster samples are selected using the redMaPPer (Rykoff et al. 2014) algorithm with the same settings as those described in McClintock et al. (2019c). We then split the redMaPPer clusters into three redshift bins using the redMaPPer photometric redshift: $0.2 < z < 0.6$, $0.6 < z < 0.9$, and $0.9 < z < 1.3$.
we see all of the simulations predict simulations span the range of values of their mean. The amount of projection in the redMaPPer catalog is related to \( \lambda \) and \( z \). Fig. 1 shows the comparison of mean \( \cos \lambda \) for the three versions of BUZZARD. We assume that \( \cos \lambda \) is redshift independent. This choice was made based on our analysis of BuzzA, where no redshift evolution of \( b_{sel} \) is observed. In subsequent analysis of the BuzzB and BuzzC simulations we found that 3 out of 11 simulations exhibited redshift evolution at 2σ to 3σ significance, as determined from a direct fit to the galaxy and particle data. While these realizations exhibit redshift evolution, the noise in the DES-Y1 data set is sufficiently large that the bias on \( \Omega_m \) and \( \sigma_8 \) incurred from assuming no redshift evolution is small. In particular, in Fig. 5, we find that our posteriors are consistent with the input cosmology. We have also explicitly tested the impact of adding the redshift evolution in our posteriors through a reanalysis of the realization 4a of BuzzB, the realization that exhibits the largest amount of redshift evolution among all realizations. Relative to the model that assumes redshift independent \( b_{sel} \), allowing for redshift evolution in \( b_{sel} \) shifts the posteriors towards the input cosmology. In particular, the smallest confidence contours containing the true \( \sigma_8 \) and \( \Omega_m \) parameters are the 81 and 93 per cent confidence contours for the model with and without redshift evolution, respectively. The small difference in the contours demonstrates that these shifts are small relative to the statistical errors. In the Appendix F, we detail the investigation of how the redshift-dependent \( b_{sel} \) affects the cosmological constraint.

We further investigate the connection between \( b_{sel} \) and the two known systematics in redMaPPer clusters: projection effects and ori-
expect that future work on quantifying these two systematics can put a tighter prior on the selection bias and hence tighten the cosmological constraints derived from the data.

Here, although we measure the $b_{\text{rel}}$ based on redMaPPer–redMaGiC cross-correlations, this selection bias is not limited to this part of the data vector. Given that galaxies are biased tracers of the dark matter density field, we expect the selection bias $b_{\text{rel}}$ to apply for cluster–cluster and cluster–shear correlations as follows:

$$y_{\lambda \text{ selected}} = b_{\text{rel}}(M)s_{\lambda \text{ selected}}$$

$$w_{cc\lambda \text{ selected}} = b_{\text{rel}}^2(M)w_{cc\lambda \text{ selected}}.$$ (2)

While we expect the above argument is valid on sufficiently large scales, we do not expect this simple model to hold at small scales. For example, the redMaGiC galaxies clustering signal may be correlated with the richness of redMaPPer clusters at a fixed redMaPPer mass. This correlation would introduce an additional redMaPPer selection effect on redMaGiC–redMaPPer clustering, but not on cluster lensing and cluster clustering. We find that in the DES-Y1 data (DES1Y1KP, DES2020), the fraction of redMaGiC galaxies in redMaPPer clusters with richness above $\lambda = 20$ is $\approx 2–3$ per cent. Thus, we expect that any such redMaPPer selection effect has negligible effects on the clustering of redMaGiC galaxies scales greater than $8h^{-1}$ Mpc, the minimum scale cut in this analysis. We also note that the above argument is only valid in the linear regime. Further analysis of the impact of selection bias on cluster lensing and cluster clustering beyond the linear regime needs to be done to extend this framework to small angular scales.

4 MODEL AND COVARIANCE MATRIX

We assume the probability distribution $P(D|p)$ of the observed data vector $D$ given the model parameters $p$ is Gaussian. Therefore, the likelihood function takes the form

$$L(p|D) \propto \exp \left( -\frac{1}{2} [D - M(p)]^T C^{-1} [D - M(p)] \right).$$ (4)

where $M(p)$ is the model prediction and $C$ is the covariance matrix. In this section, we describe the construction of the model and the covariance matrix.
4.1 Model

The data vector of this analysis consists of the abundance of redMaPPer clusters (N), as well as four distinct two-point correlations. These are: (1) the autocorrelation of redMaGiC galaxies \( w_{gg}(\theta) \); (2) the redMaPPer–redMaGiC cross-correlation \( w_{cg}(\theta) \); (3) the autocorrelation of redMaPPer clusters \( w_{cc}(\theta) \); and (4) the redMaPPer cluster–shear cross-correlation \( \gamma_{\perp,\ell}(\theta) \).

4.1.1 Cluster abundance

The redMaPPer cluster abundance in a given richness (\( \delta \lambda \)) and redshift (\( \delta z \)) bin is given by

\[
N^i = \int_0^\infty d\ztrue \frac{dV}{d\ztrue} \langle \phi_i | \ztrue \rangle \int_{z_h}^\infty dz \int_0^\infty dM P(\lambda | M, \ztrue) \frac{dn}{dM}(M, \ztrue),
\]

(5)

where

\[
\langle \phi_i | \ztrue \rangle = \int_{z_h}^\infty dzobs \int \frac{dV}{d\ztrue} p(zobs | \ztrue) \phi_i(zobs)
\]

(6)

and \( \phi_i \) is the redshift binning function for bin \( i \), i.e. \( \phi_i(z_{obs}) = 1 \) if \( z_{obs} \) is in bin \( i \), and zero otherwise. In the above expression, \( \ztrue \) denotes true redshifts of galaxy clusters, \( M \) represents the cluster mass, \( d\Vd\ztrue \) is the survey volume per unit redshift, and \( dn/dM \) is the Tinker halo mass function (Tinker et al. 2010). In Appendix D, we verify that replacing the Tinker mass function by an emulator (McClintock et al. 2019a) has a negligible impact on our results.

In equation (5), the redshift distribution \( p(z_{obs} | \ztrue) \) is averaged over all clusters

\[
p(z_{obs} | \ztrue) = A^{-1} \sum_a \mathcal{N}_a z_{obs} | \ztrue \sigma_{z,a}^2,
\]

(7)

where \( A \) is a normalization constant; \( \mathcal{N}(x|A, B) \) represents the Gaussian distribution with mean \( A \) and variance \( B \); \( z_{RMCG} \) and \( \sigma_z \) are reported photometric redshift estimation and uncertainty in the redMaPPer; the sum is over all clusters \( a \) in a thin redshift shell of width \( \pm 0.005 \) centred on \( \ztrue \). The survey volume per unit redshift is estimated by

\[
dV/d\ztrue = \Omega_{\text{mask}}(z) f(z) x^2(z),
\]

(8)

where \( x \) represents the comoving distance, and \( \Omega_{\text{mask}}(z) \) is the effective survey area obtained from the redMaPPer algorithm (Costanzi et al. 2019b). We model the richness–mass relation \( \langle F(\lambda | M, \ztrue) \rangle \) as a lognormal model with scatter

\[
\sigma_{\text{log}} = \sigma_{\text{intrinsic}} + (e^{\text{log}} - 1)/e^{\text{log}},
\]

(9)

and mean

\[
\langle \ln(\lambda | M) \rangle = \ln\lambda_0 + A_1 \ln(M/M_{\text{vir}}) + B_1 \ln((1 + z)/1.45).
\]

(10)

In Section 5.3, we demonstrate that this model is sufficient to obtain unbiased cosmological constraints.

4.1.2 Two-point clustering correlation functions

A two-point correlation function can be related to the corresponding angular power spectrum via

\[
w_{gg}^i(\theta) = \sum_i P_i(\cos(\theta)) C_{\alpha, \beta}^i(\ell),
\]

(11)

where \( \alpha \) and \( \beta \) denote the two tracers being correlated (galaxy overdensity \( g \) or cluster overdensity \( c \)), \( i \), \( j \) represents the tomographic bins of the two tracers, \( \theta \) is the angular separation, and \( P_\ell \) is the Legendre polynomial of order \( \ell \).

In the linear regime, the angular power spectrum \( C_{\alpha, \beta}^i(\ell) \) of two density tracers at redshift bins \( \delta z_i \) and \( \delta z_j \) can be written as

\[
C_{\alpha, \beta}^i(\ell) = \frac{2}{\pi} \int dz_1 \int dz_2 \int \frac{dk}{k} \Delta_{\alpha}^i(\ell, \delta z_i, z_1) \Delta_{\beta}^j(\ell, \delta z_j, z_2) P_{\delta}(k, z_1, z_2),
\]

(12)

where

\[
P_{\delta}(k, z_1, z_2) = P_{\phi}(k) T_{\text{sel}}(z_1, k) T_{\text{sel}}(z_2, k),
\]

(13)

\[
\Delta_{\alpha}^i(k, \delta z_i, z) = \Delta_{\alpha}^D(k, \delta z_i, z) + \Delta_{\alpha}^{\text{RSD}}(k, \delta z_i, z),
\]

\[
\Delta_{\alpha}^{D/s}(k, \delta z_i, z) = q_\alpha'(z) b_{\alpha}'(z) j_i(k x(z)),
\]

\[
\Delta_{\alpha}^{\text{RSD}}(k, \delta z_i, z) = -q_\alpha'(z) f(z) j_i(k x(z)).
\]

(14)

In the above expression, \( b_{\alpha}'(z) \) is the linear bias of the tracer \( \alpha \) in redshift bin \( i \), \( q_{\alpha}'(z) \) is the unit-normalized redshift distribution of the tracer \( \alpha \), \( k \) is the 3D wavenumber, \( P_{\phi}(k) \) is the power spectrum of the primordial curvature perturbations, \( T_{\text{sel}}(z,k) \) is the matter overdensity transfer function, \( f(z) = d\ln D/d\ln a \) is the scale independent growth rate, and \( j_i(k x(z)) \) is the second-order derivative of the spherical Bessel function. The density tracer’s number density \( \Delta_{\alpha}^N \) includes two contributions: a tracer’s density contribution \( \Delta_{\alpha}^D \) and a linear contribution from redshift space distortions \( \Delta_{\alpha}^{\text{RSD}} \). We refer the reader to section 2.4.1 of Chisari et al. (2019) for a more comprehensive description. In practice, we include the contribution of equal-time non-linear matter power spectra while evaluating the tracer’s density contribution in equation (12). That is, we ignore the contribution of unequal-time non-linear matter power spectra, which have been shown to be subdominant (Chisari & Pontzen 2019; Fang et al. 2020).

The unit-normalized redshift distribution of the redMaPPer clusters \( q_{\alpha}'(z) \) is calculated by

\[
q_{\alpha}'(z) = A^{-1} \langle \phi_i | \ztrue \rangle \frac{dV}{dz},
\]

(15)

where \( A \) is a normalization constant. Following DESY1KP, we calculate the unit-normalized redshift distribution of redMaGiC galaxy \( q_{\alpha}'(z) \) by stacking \( p(z_{\text{obs}}) \), which is approximated by a Gaussian distribution with mean \( \alpha \) and scale \( \sigma_z \) given by the redshift \( z_{\text{RMCG}} \) and photometric uncertainty reported by redMaGiC algorithm.

We treat the redMaGiC galaxy bias \( b_{\alpha}'(z) \) in each tomographic bin as a nuisance parameter, which is a constant in each redshift bin. Unlike the galaxy bias, the bias of galaxy clusters is a predicted quantity in our model. We relate the bias to the mass of the galaxy clusters via measurement in \( N \)-body simulations (Tinker et al. 2010). Here again replacing the Tinker bias by an emulator (McClintock et al. 2019b) has a negligible impact on our conclusions (see Appendix D). As pointed out in Section 3, redMaPPer clusters are subject to selection effects that manifest as an additional mass-dependent clustering bias. Thus, the net clustering bias of clusters in a given richness bin \( \delta \lambda \) at redshift \( z \) is given by

\[
b\delta(\delta \lambda, z) = \langle b_{\delta}(z) \rangle = \frac{1}{n(\lambda, z)} \int dM b_\delta(M, z) b_{\text{sel}}(M, z) \frac{dn}{dM} \int_{z_h}^\infty dz P(\lambda | M, z),
\]

(16)

where the normalization is given by

\[
n(\lambda, z) = \int dM \frac{dn}{dM} \int_{z_h}^\infty dz P(\lambda | M, z),
\]

(17)

\( b_\delta(M, z) \) is the Tinker bias function, and \( b_{\text{sel}}(M, z) \) is the selection bias model defined in equation (1).
We evaluate equation (12) using the fast generalized FFTLog² algorithm presented in Fang et al. (2020).

4.1.3 Cluster lensing
Cluster lensing is the measurement of the tangential shear of source galaxies around galaxy clusters. Here, we utilize the Limber approximation (Limber 1953) to convert the 3D power spectrum to the angular power spectrum. This analysis choice is justified in Fang et al. (2020), which shows that the galaxy–galaxy lensing model with Limber approximation is sufficiently precise to derive unbiased cosmological parameters from a Rubin Observatory LSST Y1-like survey. Given the large number density of galaxies relative to the number of galaxy clusters in a survey, as well as the steepness of the halo mass function relative to the bias–mass relation, the galaxy–galaxy lensing signal has a higher signal-to-noise ratio than the cluster lensing signal at the same scale. Thus, we expect the Limber approximation to be sufficient for modelling the cluster lensing signal in this analysis. Under the Limber approximation, the tangential shear of the background galaxies in redshift bin $j$ around the galaxy clusters in redshift bin $\delta z_i$ and richness bin $\delta \lambda$ at an angular separation $\theta$ can be written as

$$\gamma_{\ell\ell}^j(\theta) = \frac{3}{2} \Omega_m \left( \frac{H_0}{c} \right)^2 \int \frac{d\ell}{2\pi} J_2(\ell \theta) \int dz \frac{g(z)q(z)}{a(z)\chi(z)} P_{\text{hm}}(k = \frac{\ell + 1/2}{\chi(z)}\zeta),$$

(18)

where $J_2$ is the second-order Bessel function of the first kind, $a$ is the scale factor, $\chi$ is the comoving distance, $(P_{\text{hm}})$ is the averaged cluster–matter power spectrum. In the above expression $g(z)$ is the lensing efficiency for source galaxies in redshift bin $j$, computed as

$$g^j(z) = \int_z^{\infty} dz' q(z') \frac{\chi(z') - \chi(z)}{\chi(z')},$$

(19)

where $q^j(z)$ is the unit-normalized redshift distribution of source galaxies in redshift bin $j$, which is estimated using the BPZ photo-z PDF estimates.

Similar to equation (16), the averaged cluster–matter power spectrum $(P_{\text{hm}})$ in redshift bin $i$ and richness bin $\delta \lambda$ can be written as

$$(P_{\text{hm}})(k, z) = \frac{1}{\rho_m} \int dM P_{\text{hm}}(k, M, z) \frac{dn}{dM} \int_z^{\infty} dz P(m|M, z),$$

(20)

where $P_{\text{hm}}(k, M, z)$ is the halo–matter power spectrum of haloes with mass $M$ at redshift $z$.

Following Krause & Eifler (2017), the halo–matter power spectrum is modelled in the halo model framework (Cooray & Sheth 2002). In this model, $P_{\text{hm}}(k, M, z)$ can be written as

$$P_{\text{hm}}(k, M, z) = \frac{M}{\rho_m} u(k, c, z) + b_f(M, z)b_m(M)P_{\text{NL}}(k),$$

(21)

where $\rho_m$ is the mean matter density of the Universe, $b_m$ is defined in Section 3. and $u(k, c, z)$ is the Fourier transform of the NFW profile with halo concentration $c$, for which we use the concentration–mass relation of Bhattacharyya et al. (2013).

4.2 Covariance matrix
The Gaussian likelihood (equation 4) indicates that the covariance matrix is a key quantity that determines the error on the inferred cosmological parameters. As summarized in Krause et al. (2017), the covariance matrix can be generated by three different methods: estimation from simulations, estimation from data, and analytical calculations. While the first two approaches require less theory assumptions, the covariance estimators are inherently noisy. The noise in covariance estimations leads to additional uncertainties to the inferred cosmological parameters estimated from the Gaussian likelihood (Hartlap, Simon & Schneider 2007; Dodelson & Schneider 2013). In this paper, we analytically compute the covariance matrix. This approach is motivated by the following arguments. First, unlike the estimation from simulations or data, there is no estimator noise in the theoretically derived covariance matrix, allowing the use of the Gaussian likelihood, instead of using a multivariate t-distribution (Sellentin & Heavens 2016). Second, as pointed out in Wu et al. (2019), the non-Gaussian terms in our covariance matrix are subdominant, and thus the corresponding uncertainties are not important.

In this section, we summarize the analytical covariance matrix computation. The analytical covariance matrix can be separated into three components: angular two-point statistics $\times$ angular two-point statistics, angular two-point statistics $\times$ cluster abundance, and cluster abundance $\times$ cluster abundance. The covariance of two angular two-point functions $w_1, w_2 \in \{w_{gg}, w_{cg}, w_{cc}, \gamma_{\ell\ell} \}$ is related to the covariance of the angular power spectra by

$$\text{Cov}(w_{1}^j(\theta), w_{2}^j(\theta)) = \int \frac{d\ell d\ell'}{2\pi} J_{\ell_1}(\theta) J_{\ell_2}(\theta) \left[ \text{Cov}(C_{1}^{\ell_1}(\ell), C_{2}^{\ell_2}(\ell)) \right],$$

(22)

where $n = 0$ for $w_{gg}$, $w_{cg}$, $w_{cc}$, and $n = 2$ for $\gamma_{\ell\ell}$. The term $\text{Cov}(C_{1}^{\ell}(\ell), C_{2}^{\ell_2}(\ell))$ is the covariance of angular power spectra given by the sum of a Gaussian and a non-Gaussian covariance, including supersample variance (Krause & Eifler 2017). The covariance of angular two-point functions and cluster abundance $(N)$ can be related to the covariance of the angular power spectrum and cluster abundance via

$$\text{Cov}(w_{1}^j(\theta), N^{i}) = \int \frac{d\ell d\ell'}{2\pi} J_{\ell_1}(\theta) J_{\ell_2}(\theta) \left[ \text{Cov}(C_{1}^{\ell_1}(\ell), N^{i}) \right],$$

(23)

where $\text{Cov}(C_{1}^{\ell}(\ell), N^{i})$ is the covariance of the angular power spectrum and cluster abundance. The cluster abundance cross cluster abundance terms are the sum of Poisson shot noise terms and supersample variance terms. We refer the reader to Krause & Eifler (2017) for more details.

5 RESULTS
5.1 Measurement
We measure the two-point correlation functions – galaxy clustering, galaxy–cluster cross-correlations, cluster clustering – using the Landy–Szalay estimator (Landy & Szalay 1993)

$$\hat{w}(\theta) = \frac{DD - 2DR + RR}{RR},$$

(24)

where $DD$ is the number of pairs of tracers (galaxies or galaxy clusters) with angular separation $\theta$, $RR$ is similarly defined for a catalogue of points whose positions are randomly distributed within

²https://github.com/xfangcosmo/FFTLog-and-beyond
the survey volume (random points), and $DR$ is the number of cross pairs between tracers and random points. The correlation functions are calculated in 20 logarithmic angular bins between 2.5 and 250 arcmin to match the analysis in DESY1KP. The pair counting is done by CORRFUNC$^3$ (Sinha & Garrison 2020).

The cluster lensing tangential shear signal $\gamma^{1,1}_c(\theta)$ is measured by averaging the tangential shear ($\gamma_t$) of source galaxies over all cluster–source galaxy pairs with an angular separation $\theta$. The $\gamma^{1,1}_c(\theta)$ estimator is written as

$$\hat{\gamma}^{1,1}_c(\theta) = \frac{\sum_{i \in DS^{1,1}(\theta)} e^{\alpha}_C c_i}{DS^{1,1}(\theta)} - \frac{\sum_{t \in RS^{1,1}(\theta)} e^{\alpha}_T}{RS^{1,1}(\theta)},$$

where $DS^{1,1}(\theta)$ is the number of cluster–source galaxy pairs at redshift bins $i$ and source galaxies at redshift bins $j$ that are separated by an angular separation $\theta$, and $RS^{1,1}(\theta)$ is similarly defined as $DS^{1,1}(\theta)$ but on random–source galaxy pairs. $c_i$ is the tangential shear of the source galaxies in cluster–source galaxy pair $\alpha$.

This estimator is biased due to photometric redshift uncertainties. Due to uncertainties in the redshift estimations, some of the source galaxies are members of galaxy clusters. These galaxies are not lensed by galaxy clusters, thus diluting the lensing signal. Sheldon et al. (2004) point out that this dilution effect can be measured by the following estimator

$$B^{1,1}(\theta) = \frac{N_i}{N_c} \frac{DS^{1,1}(\theta)}{RS^{1,1}(\theta)},$$

where $N_i$ is the number of random points, $N_c$ is the number of galaxy clusters, and $RS^{1,1}(\theta)$ is the number of random points–source galaxy pairs with angular separation $\theta$. The $B^{1,1}(\theta)$ is usually called boost factor in the literature, and $1/B^{1,1}(\theta)$ is the amount of dilution due to photometric uncertainties. Since the boost factor is measured in the data, we apply this correction directly on the estimator. Using this correction, our cluster lensing estimator is written as

$$\hat{\gamma}^{1,1}_c(\theta) = \frac{\sum_{i \in DS^{1,1}(\theta)} e^{\alpha}_C c_i}{DS^{1,1}(\theta)} - \frac{\sum_{t \in RS^{1,1}(\theta)} e^{\alpha}_T}{RS^{1,1}(\theta)},$$

where $DS^{1,1}(\theta)$ is the number of cluster–source galaxy pairs calculated in 20 logarithmic angular bins between 2.5 and 250 arcmin. The calculation is done by TREECORR$^4$ (Jarvis, Bernstein & Jain 2004).

5.2 Analysis choices

We summarize our analysis choices below. We expect these analysis choices to be carried through to the analysis of real data.

(i) Minimum angular scale cuts. For $w_{gg}(\theta)$, $w_{cg}(\theta)$, and $\gamma_{1,1,c}(\theta)$, we adopt a minimum scale cut corresponding to $8 h^{-1}$ Mpc at the mean redshift of each lens redshift bin. In Section 5.3, we justify this scale cut by verifying that the cosmological posteriors derived from our analysis are robust to a variety of systematics when adopting this cut. For $w_{cc}(\theta)$, we adopt a minimum scale cut corresponding to $16 h^{-1}$ Mpc at the mean redshift of each cluster redshift bin. The scale cut is chosen such that the $\chi^2$ of the best-fitting model is consistent with the $\chi^2$ between data vectors measured in different realizations of the same simulation scheme. In this way, we obtain the minimum scale where the best-fitting model provides a good description of the $w_{cc}(\theta)$ without relying on the exact value of the theory covariance matrix. We further confirm that the distribution of $\chi^2$ values is consistent with the expected $\chi^2$ distribution based on the number of degrees of freedom. Finally, we apply an additional scale cut on $w_{cc}(\theta)$ to avoid biases and large fluctuations to the correlation function measurement from to sparseness issues when there are only few clusters pairs in the angular bins. Thus, we cut out the angular bins when the expected number of pairs are less than one hundred. We find that this additional scale cut largely improves the $\chi^2$ of the best-fitting model.

(ii) Redshift distributions. The redshift distributions of the lens samples ($\tilde{P}_{gg}(\ell)$) are calculated based on photometric redshifts estimated by redMaPPer and redMaGiC. The redshift distributions of the source galaxies are estimated from photometric redshifts estimated by BPZ (Benitez 2000). Following DESY1KP, we introduce two sets of nuisance parameters to account for systematics in photometric redshift estimations. The systematics are modelled through shift parameters $\Delta^{g,s}_i$, so that

$$q^{g,s}_i(z) = \hat{q}^{g,s}_i(z - \Delta^{g,s}_i), \quad \alpha \in \{g, s\},$$

where $g$ denotes redMaGiC galaxies, $s$ denotes source galaxies, and $\hat{q}^{g,s}_i(z - \Delta^{g,s}_i)$ denotes the estimated redshift distributions based on photometric redshifts. Note that we do not account for the redshift systematic of redMaPPer clusters, since DESY2020 demonstrate that this systematic is subdominant. These shift parameters $\Delta^{g,s}_i$ are marginalized over using Gaussian priors of width [0.008, 0.007, 0.007] for redMaGiC galaxies and [0.016, 0.013, 0.011, 0.022] for source galaxies. The mean of the Gaussian prior is estimated by comparing the true redshift distribution in simulations and the photometric redshift estimations. This is clearly not possible in a real data analysis. In the analysis of real data, the mean of the Gaussian prior is estimated using cross-correlations of galaxy samples with spectroscopic samples (Cawthon et al. 2018; Hoyle et al. 2018). Because we focus on cluster-related systematics in this paper, we do not repeat this process in the simulation.

(iii) Matter power spectrum. We evaluate the non-linear matter power spectrum using the Eisenstein & Hu (1998) approximation for the transfer function and the revised HALOFIT fitting formula of Takahashi et al. (2012) for the non-linear evolution. To validate this model, we compare the theory data vector generated at the true cosmology to that generated from CLASS (Blas, Lesgourgues & Tram 2011) and HALOFIT. We find that the $\Delta \chi^2$ between the two data vectors is 1.45. Thus, we conclude that this theory approximation does not affect the conclusion of this paper.

(iv) Theory covariance matrix. The covariance matrix is calculated assuming a fixed set of cosmological and nuisance parameters. Carron (2013) shows that when approximating the true data likelihood with a Gaussian likelihood, the parameter posteriors better match the true uncertainty in the measurement when the cosmological independence of the covariance matrix is ignored. In particular, allowing the covariance matrix to vary with cosmology results in overoptimistic constraints. In this analysis, we fix the cosmological parameters for the covariance matrix at the true cosmology in the BUZZARD mock catalogues. This is clearly not possible in an analysis of real data. However, DESY1KP shows that there is negligible change in the parameter constraints in the $3 \times 2pt$ analysis while

$^3$https://github.com/manodeep/Corrfunc

$^4$https://github.com/rmjarvis/TreeCorr
using two different cosmologies to calculate the covariance matrix. Because our data vectors are more shot-noise dominated than the $3 \times 2pt$ data vectors, we expect our conclusions to be insensitive to this analysis choice.

Unlike a $3 \times 2pt$ analysis, however, our observable also depends on the richness–mass relation and the selection bias, for which we do not have good a priori estimates. We use an iterative approach to obtain the richness–mass relation and selection bias parameters used to compute the covariance matrix. We start by setting $b_{\text{ini}} = 1$ and $b_{\text{ini}} = 0$. The fiducial richness–mass relation is obtained from fitting the cluster abundance data only assuming only shot noise, and adopting the true cosmology of the simulations. This selection bias and richness–mass relation are then used to generate a covariance matrix, which is adopted while fitting cluster abundance and cluster clustering simultaneously at the true cosmology of the simulation. We update the richness–mass relation and selection bias parameters with the best-fitting parameters, and re-fit the cluster abundance and cluster clustering data. Thus, we obtain a new estimate of the parameters and priors considered in this analysis. Flat represents the flat prior in the given range and Gauss($\sigma$) denotes the Gaussian prior with width $\sigma$. The means of the Gaussian priors are determined by comparing the redshift bias of redMaGiC galaxies, and four shift parameters for the redshift bias of source galaxies. Finally, our model has four richness–mass relation parameters, and two selection bias parameters. Here, we do not consider intrinsic alignment effect for the redMaGiC galaxies, and four shift parameters for the redshift bias of source galaxies. Parameter Prior

(iii) Measure the bias in $\Omega_m$ and $\sigma_8$ due to the unaccounted systematics.

All analyses are done using the analytical covariance matrix generated using the best-fitting cosmology of DESY1KP and the richness–mass relation parameters in Bleem et al. (2020).

Fig. 4 quantifies the cosmological parameter bias due to these three systematic effects that are not fully captured by the BUZZARD mock catalogues. The first systematic is the sensitivity to the anomalously low lensing signal at small scales for low richness clusters described in DES2020. A similar low lensing signal has been detected in analyses of SDSS spectroscopic galaxies (Leauthaud et al. 2017). Importantly, DES2020 points out that the BUZZARD mock catalogues do not exhibit a similar feature. We test the sensitivity of our analysis to this systematic by reducing the amplitude of the one-halo term of our lowest richness bins by 50 per cent.

The second systematic is the functional form of the richness–mass relation. The richness–mass relation depends on both the galaxy–halo connection as well as the performance of the cluster finder (Costanzi et al. 2019a). If the functional form of the richness–mass relation in the simulations does not reproduce that of the data, then our simulation tests may leave us blind to this possible source of systematic uncertainty. For example, DeRose et al. (2019) find a deficit in richness at fixed halo mass when comparing BUZZARD to DES-Y1 data (McClintock et al. 2019c), leading to a factor of 2 fewer $\lambda > 20$ galaxy clusters in BUZZARD than in the data. Thus, while the lognormal richness–mass relation is sufficient to describe the richness–mass relation in the BUZZARD mock catalogues, it does not guarantee that it will adequately describe the data. To check the

| Parameter | Prior |
|-----------|-------|
| $\Omega_m$ | Flat (0.1, 0.9) |
| $\sigma_8$ | Flat (0.5, 1.2) |
| $n_s$ | Flat (0.87, 1.07) |
| $\Omega_b$ | Flat (0.03, 0.07) |
| $h$ | Flat (0.55, 0.91) |
| $b_1^1$ | Flat (0.8, 3.0) |
| $b_1^2$ | Flat (0.8, 3.0) |
| $b_1^3$ | Flat (0.8, 3.0) |
| $b_2^1$ | Flat (0.8, 3.0) |
| $\Delta_1^1$ | Gauss($\sigma = 0.008$) |
| $\Delta_1^2$ | Gauss($\sigma = 0.007$) |
| $\Delta_2^1$ | Gauss($\sigma = 0.007$) |
| $\Delta_2^2$ | Gauss($\sigma = 0.007$) |
| $\Delta_3$ | Gauss($\sigma = 0.018$) |
| $\Delta_4$ | Gauss($\sigma = 0.013$) |
| $\sigma_{\text{intrinsic}}$ | Flat (0.1, 1.0) |

### 5.3 Robustness

We quantify the impact of potential systematics that cannot be fully be tested by the simulations. Specifically, we consider the following systematics: small-scale lensing systematics, the functional form of the richness–mass relation, and beyond linear bias expansions. In all cases, our robustness tests follows the same procedure:

(i) Generate a noiseless data vector that includes the systematic effect being tested.

(ii) Analyse this data vector with the baseline model.

(iii) Measure the bias in $\Omega_m$ and $\sigma_8$ due to the unaccounted systematics.
robustness of our cosmological constraints against this systematic, we generate the input data vector using the richness–mass relation in DES2020, which we then analyse with our baseline lognormal richness–mass relation model.

The third systematic is the possible presence of non-linear galaxy and galaxy cluster biases. This systematic is partially tested by the simulations; however, the size of the effect depends on the mean mass of the galaxy cluster, thus depending on the exact value of the richness–mass relation. Because we do not expect the simulation to perfectly reproduce the richness–mass relation of the real data, we check the robustness of this model using a theoretical calculation. The input data vector is generated including the next-to-leading order contribution from quadratic bias $b_{2}$, tidal bias $b_{s}$, and the third-order non-local bias $b_{3d}$ (McDonald & Roy 2009; Baldauf et al. 2012). The non-linear contributions are evaluated using the FAST-PT code (McEwen et al. 2016) with $b_{2}$ determined by $b_{2} = b_{1}$ relation measured in $N$-body simulations (Lazeyras et al. 2016). The $b_{s}$ and $b_{3d}$ are determined by their relation to $b_{1}$ derived from the equivalence of Lagrangian and Eulerian perturbation theory (Saito et al. 2014). This data vector is then analysed by the baseline linear bias model.

Fig. 4 shows that none of the above systematics bias our posteriors by more than 0.5σ. We therefore conclude that our model is sufficiently flexible to enable us to derive robust cosmological constraints at the precision achievable by DES-Y1-like surveys. More detailed modelling may be required for future, more constraining, analyses.

In this analysis, we do not consider redMaPPer mis-centring as a potential systematic for two reasons. First, the scale of mis-centring of redMaPPer clusters is $\approx 0.2 \, h^{-1} \text{Mpc}$ (Zhang et al. 2019), much smaller than the smallest scales included in our analysis. Secondly, any additional scatter in richness estimates due to miscentring effects can be absorbed by the richness–mass scaling relation parameters (Section 4.1).

5.4 Fiducial cosmological parameter constraints

We test whether our pipeline can correctly recover the cosmological parameters in simulated data following the method developed in MacCrann et al. (2018). We assume that the potential systematically biased posterior on parameters $\theta$, $P^{sys}(\theta \mid s_{i})$ inferred from analysing a simulated data vector $s_{i}$ can be related to the true posterior $P(\theta \mid s_{i})$ by a shift $\Delta \theta$ in the parameter space

$$P^{sys}(\theta \mid s_{i}) = P(\theta - \Delta \theta \mid s_{i}), \quad (29)$$

To quantify the significance and the size of potential systematics we estimate the posterior of $\Delta \theta$ by analysing a set of simulated data vectors $\{s_{i}\}$, all generated from the same true cosmological parameters $\theta_{true}$. In the following, we use Bayes’ theorem to relate the posterior of $\Delta \theta$ ($P(\Delta \theta \mid s_{i}, \theta_{true})$) to the potential systematically biased posterior ($P^{sys}(\theta \mid s_{i})$) inferred from analysing individual simulated data vectors

$$P(\Delta \theta \mid s_{i}, \theta_{true}) = \frac{P(\{s_{i}\} \mid \Delta \theta, \theta_{true}) P(\Delta \theta)}{P(\{s_{i}\})}$$

$$= \frac{P(\Delta \theta) \prod_{i=1}^{N} P(s_{i} \mid \Delta \theta, \theta_{true})}{\prod_{i=1}^{N} P(s_{i})}$$

$$\propto \prod_{i=1}^{N} P(\theta_{true} \mid s_{i})$$

$$\propto \prod_{i=1}^{N} P^{sys}(\theta_{true} + \Delta \theta \mid s_{i}), \quad (30)$$

where we assume that we analyse $N$ simulated universes. In the second equality, we assume that realizations of the simulated universe are mutually independent; in the third equality, we assume that the parameter shift ($\Delta \theta$) is drawn from a flat prior and the generation of simulated universe only depends on the true cosmological parameters $\theta_{true}$; in the fifth equality, we substitute equation (29). From equation (30), we can estimate the probability of systematic biases on a parameter by computing the product of the parameter posteriors from analysing each simulation realization.

In this paper, we focus on estimating systematic biases in $\Omega_{m}$ and $\sigma_{8}$, the two cosmological parameters we expect to be well constrained by analyzing the DES-Y1 data. Fig. 5 shows 68 and 95 percent constraints on $\Omega_{m}$ and $\sigma_{8}$ from the three versions of the BUZZARD mock catalogues, marginalizing over 16 nuisance parameters and 3 cosmological parameters as described in Section 5.2. The grey contours show the constraints from individual realizations and the orange contours represent the combination of constraints from all realizations. The black dashed lines indicate the true cosmology of the BUZZARD mock catalogues. For comparison, in all panels in Fig. 5, the blue contours denote the expected 68 and 95 percent constraints from the DES-Y1 data, which are estimated from analysing noiseless theory data vectors with the covariance matrix generated based on DES-Y1 data (Abbott et al. 2018; Bleem et al. 2020).

From equation (30), the orange contours can be related to the posterior of systematics ($P(\Delta \theta \mid s_{i}, \theta_{true})$) on $\Omega_{m}$ and $\sigma_{8}$. Given the posterior of systematics, we can calculate what is the minimal confidence interval containing the null hypothesis ($\Delta \theta = 0$). We find that the 96.9, 93.0, and 99.2 percent confidence intervals contain $\Delta \theta = 0$ for BuzzA, BuzzB, and BuzzC, respectively. These results indicate the presence of biases in our simulation results at an $\approx 95$ percent confidence. We note, however, that the statistical
Figure 5. Constraints on $\Omega_m$ and $\sigma_8$ from cluster number counts and four two-point correlation functions (redMaGiC autocorrelations, redMaPPer-redMaGiC cross-correlations, redMaPPer autocorrelations, and cluster lensing) measured in three versions of the BUZZARD mock catalogues: BuzzA (top), BuzzB (lower left), and BuzzC (lower right). Contours show the 68 and 95 per cent confidence levels. The differences between these simulations are summarized in Table 1. In each panel, grey contours show the constraints from individual realizations and the orange contours show the combination of these posteriors (equation 30). The blue contours show the expected DES-Y1 constraints shifted to centre on the input cosmology of the simulation. The black dashed lines indicate the true cosmology, i.e. the input cosmology used to generate the simulation. This indicates that our method is unbiased at the accuracy level needed for DES-Y1 data.

power of the simulations is 10 times that of DES-Y1 for BuzzA and BuzzB, and 3 times that of DES-Y1 for BuzzC. To determine whether this level of systematic is important for a DES-Y1 experiment, we calculate the probability that the systematic shift in $\sigma_8$ and $\Omega_m$ is contained within the 68 per cent likelihood contour of the DES-Y1 experiment. Roughly speaking, this is the probability that systematic shifts in the parameters are smaller than the statistical uncertainties. We find $P(\text{sys} < \sigma_{Y1}) = 0.65$ for BuzzA, $P(\text{sys} < \sigma_{Y1}) = 0.57$ for BuzzB, and $P(\text{sys} < \sigma_{Y1}) = 0.19$ for BuzzC. Our analysis are consistent with negligible to modest parameter biases.

To decide whether enlarging the error of the analysis on real data is needed to accommodate this systematic, we need to understand whether this systematic is due to flaws in analysis pipeline or other sources, such as statistical fluctuations or flaws in the simulations. We therefore perform an analysis combining galaxy–galaxy lensing and galaxy clustering ($2 \times 2\text{pt}$ analysis) on the same set of simulations. We find a similar level of bias for $\Omega_m$ and $\sigma_8$ (see Appendix G1 for details). From this analysis, we believe that the modest parameter biases found in Fig. 5 are not due to flaws in the analysis pipeline, but rather reflect an unlucky draw and/or possible flaws in the simulations that impact multiple large-scale structure analyses in similar ways. For this reason, we decide not to enlarge the error bar on the analysis of real data. More simulations are required to understand the sources of this parameter bias, and we plan to increase the number of simulations in future work. We also caution that we are not able to combine the three versions of the BUZZARD simulations to arrive at a stronger statement because they share the same underlying matter distribution, and are therefore not mutually independent realizations.

We summarize the estimation of systematic biases from simulations in Table 3.
allow the extraction of more cosmological information than that accessible to individual probes, but these analyses also provide an opportunity to internally calibrate possible systematics. Despite the promise, combined probe analyses face many challenges, particularly with regard to the necessity of theoretical assumptions and the need for reliable covariance matrix estimates. Thus, it is important to systematically validate methods of any new combined probe analysis.

In this work, we develop and validate a method of combining the abundance of galaxy clusters with four two-point statistics: galaxy clustering (\(w_{gg}\)), galaxy–galaxy cluster cross-correlations (\(w_{gc}\)), galaxy cluster autocorrelations (\(w_{cc}\)), and cluster lensing (\(y_{lc}\)). Our methodology is validated using three versions of the BUZZARD mock catalogues (DeRose et al. 2019; summarized in Table 1) with realistic galaxy and galaxy cluster selections. Based on this simulation analysis, we identify a boost in the clustering amplitude of redMaPPer galaxy clusters due to selection effects in redMaPPer. Specifically, we find that redMaPPer clusters in simulations are impacted by contamination with structures along the line of sight (projection effects), and that they are preferentially aligned along the line of sight. Sunayama et al. (2020) find similar biases in cluster lensing and cluster autocorrelations based on their mock redMaPPer catalogues. While the performance of mock redMaPPer catalogues is arguably less sensitive to the complicated galaxy–halo connection model, their mock redMaPPer algorithm is not the same and will not include all of the same systematics as the full redMaPPer algorithm being run on the data and on the simulations herein.

The three versions of the BUZZARD mock catalogues that are used to validate our pipeline include many of the complexities in the real data: varying magnitude errors due to survey depth variations affect galaxy selection; correlations between mass tracers and the cluster selection efficiency impact the distribution of galaxy cluster samples; photometric redshift estimations of source galaxies are implemented; and realistic galaxy–halo connection models that are more sophisticated than the developed theoretical model are considered. Our analyses of these synthetic data sets are the first end-to-end tests of a cluster cosmology pipeline on realistic simulated galaxy data sets.

Our main results and conclusions can be summarized as follows:

(i) We identify an additional large-scale cluster bias due to redMaPPer selection. We find that this selection bias can be well explained by projection effects and halo orientation biases, the two known systematics affecting the weak lensing signal of optically selected galaxy clusters (DES2020; Wu et al., in preparation).

(ii) We develop a model and a theory covariance matrix to combine the galaxy cluster abundance and two-point statistics (galaxy clustering, galaxy–cluster cross-correlations, cluster autocorrelations, and cluster lensing).

(iii) We validate the model by analysing three versions of the BUZZARD mock catalogues. Our simulation analysis is statistically

### Table 3. A summary of constraints on the size of parameter biases in \(\Omega_m - \sigma_8\) parameter spaces inferred from different versions of the BUZZARD mock catalogues. The first two rows show the bias in \(\Omega_m\) and \(\sigma_8\) with 68 per cent uncertainties. The third row shows the minimal confidence interval containing the null hypothesis (\(\Delta \theta = 0\)). The forth and fifth rows show the probability that the systematic is smaller than the expected DES-Y1 68 and 95 per cent uncertainties.

| Simulations | BuzzA | BuzzB | BuzzC |
|-------------|-------|-------|-------|
| \(\Delta \Omega_m\) | 0.037 ± 0.016 | 0.043 ± 0.019 | 0.080 ± 0.037 |
| \(\Delta \sigma_8\) | −0.056 ± 0.022 | −0.061 ± 0.028 | −0.055 ± 0.048 |
| Minimal confidence interval encompasses \(\Delta \theta = 0\) | 0.969 | 0.930 | 0.992 |
| \(P(\text{sys} < \sigma Y1)\) | 0.65 | 0.57 | 0.19 |
| \(P(\text{sys} < 2\sigma Y1)\) | 0.97 | 0.93 | 0.55 |

### Figure 6. Comparison of the distribution of \(\chi^2\)s at the best-fitting values in each realizations of BuzzA (the blue vertical lines) and the expected distribution based on the number of degrees of freedom (the orange line). The agreement between the orange line and the distribution of blue lines indicates the validity of the theory covariance matrix.

### 5.5 Performance of the theory covariance matrix

One of the key components of this paper is the computation of the theory covariance matrix, which is used to combine the cluster abundance measurements and two-point statistics. In this section, we validate the performance of the covariance matrix using the BUZZARD mock catalogues. Because we do not expect the performance of the theoretical covariance matrix to depend on the details of galaxy–halo connection, we focus on BuzzA in this section. In each realization of BuzzA, we use Nelder–Mead algorithm (Gao & Han 2012) to find the best-fitting parameters. We then compute the distribution of \(\chi^2\) at the best-fitting parameters recovered from the 10 independent realizations, and compared it to the \(\chi^2\) distribution with the appropriate number of degrees of freedom. The number of degrees of freedom is set to the number of entries in the data vector minus the number of parameters that are constrained by the data. Fig. 6 shows this comparison. To quantify the difference between these two distributions, we perform the Kolmogorov–Smirnov (K–S) test. The resulting \(p\) value is 1.1 per cent, indicating that the two probability distributions are consistent with one another. The low probability may suggest the presence of unmodelled uncertainties, though confirming the existence of such errors will require increasing the number of simulations analysed.

### 6 CONCLUSIONS

Combined probe analyses have been demonstrated to be a powerful cosmological inference tools (Abbott et al. 2018). Not only do they
consistent with no systematics biases, though there is ≈2σ evidence for a bias comparable to the statistical uncertainties in DES-Y1. We argue in Appendix G that this bias is unlikely due to flaws in the analysis pipeline but rather due to unfortunate draws and/or flaws in the simulations that impact other large-scale structure probes similarly. A definitive conclusion must await simulations with more constraining power.

(iv) We validate the theory covariance matrix by comparing the distribution of $\chi^2$'s at the best-fitting parameters from analyses of each realization of the BUZZARD mock catalogues and the expected $\chi^2$ distributions. We perform the Kolmogorov–Smirnov (K–S) test and find the $P$-value is 1.1 per cent, indicating that the two probability distributions are consistent with one another.

(v) We stress test the analysis pipeline by analysing theory data vectors contaminated by systematics that are not fully captured by the BUZZARD mock catalogues: lowering the cluster lensing one-halo term in the lowest richness bin by 50 per cent, adding a non-linear clustering term, and using a more complicated functional form of the richness–mass relation of redMaPPer clusters. We find that the inference pipeline is robust against these possible systematics.

In the near future, we plan to apply the pipeline developed in this paper to the DES-Y1 data set. The covariance matrix developed here enables us to combine our results with the constraints from the $3 \times 2pt$ DES-Y1 key project (DESY1KP). Although we have been focusing on the application on cosmological constraints from optically identified galaxy clusters, we note that there are many potential applications. First, the comparison of different two-point correlations can shed additional light on systematics of galaxy clusters as a cosmological probe. For example, for cluster cosmology analyses using large-scale information, we can test the robustness of bias models by comparing different two-point correlations. Secondly, the same pipeline can be applied to galaxy clusters selected in other wavelengths, such as X-ray and microwave. We forecast these possibilities in a companion paper (Krause et al., in preparation).

Our results suggest that despite the surprising results of the DES-Y1 cluster abundance analysis (DESB2020), a multiprobe cluster cosmology approach based on photometrically selected samples may recover unbiased cosmological parameter information when restricting the analysis to large scales only. Moreover, because our analysis is especially well suited for being combined with the now popular $3 \times 2pt$ analysis (combining cosmic shear, galaxy–galaxy lensing, and galaxy clustering), we expect the approach highlighted here may become the standard for the cosmological analysis of near-future photometric cluster samples.

ACKNOWLEDGEMENTS

We thank Sebastian Bocquet, August Evrard, Oliver Friedrich, and Xiao Fang for helpful discussions and comments on the manuscript. We thank Vivian Miranda for early implementations of parts of the code in this study. The original idea was discussed during the Lighthouse workshop, organized by DG, EK, and Adam Mantz, held during 2017 March at Point Montara. We thank the workshop participants for early discussions, and the KIPAC workshop programme for support. This paper has gone through internal review by the DES collaboration. This work was supported in part by the U.S. Department of Energy contract to SLAC no. DE-AC02-76SF00515 (CH, DG, and RW). CH and EK were supported in part by NASA ROSES ATP 16-ATP16-0084. EK was supported in part by the Department of Energy grant DE-SC0020247. ER was supported by DOE grants DE-SC0015975 and DE-SC0009913, and by NSF grant AST-2009401. ER also acknowledges funding from the Cottrell Scholar programme of the Research Corporation for Science Advancement. HW is supported by NSF Grant AST-1516997. This work was supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515 and as part of the Panofsky Fellowship awarded to DG. Some of the computing for this project was performed on the Sherlock cluster; we thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these results.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NFS’s NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

This work is based in part on observations at Cerro Tololo Inter-American Observatory at NSF’s NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1133766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grant numbers ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA programme of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).
DATA AVAILABILITY

The simulation data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Abbott T. M. C. et al., 2020, Phys. Rev. D, 102, 023509
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Phys. Rev. D, 86, 083540
Becker M. R., 2013, MNRAS, 435, 115
Benitez N., 2000, ApJ, 536, 571
Bhattacharya S., Habib S., Heitmann K., Vikhlinin A., 2013, ApJ, 766, 32
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 2011, 034
Bleem L. E. et al., 2020, ApJS, 247, 25
Carron J., 2013, A&A, 551, A88
Cawthon R. et al., 2018, MNRAS, 481, 2427
Chisari N. E., Pontzen A., 2019, Phys. Rev. D, 100, 023543
Chisari N. E. et al., 2019, ApJS, 242, 2
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Costanzi M. et al., 2019a, MNRAS, 482, 490
Costanzi M. et al., 2019b, MNRAS, 488, 4779
DeRose J. et al., 2019, preprint (arXiv:1901.02401)
Dietrich J. P. et al., 2014, MNRAS, 443, 1713
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Fang X., Krause E., Eifler T., MacCrann N., 2020, J. Cosmol. Astropart. Phys., 2020, 010
Gao F., Han L., 2012, Comp. Opt. Appl., 51, 259
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Hoyle B. et al., 2018, MNRAS, 478, 592
Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
Joudaki S. et al., 2018, MNRAS, 474, 4894
Krause E., Eifler T., 2017, MNRAS, 470, 2100
Krause E. et al., 2017, preprint (arXiv:1706.09359)
Lacasa F., Hoekstra H., Cacciato M., Viola M., Köhlinger F., van der Burg R. F. J., Sand D. J., Graham M. L., 2015, A&A, 575, A48
Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022
Sunayama T. et al., 2020, MNRAS, 496, 4468
Takada M., Bridle S., 2007, New J. Phys., 9, 446
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
To C.-H., Reddick R. M., Rozo E., Rykoff E., Wechsler R. H., 2020, ApJ, 897, 15
van Uitert E. et al., 2018, MNRAS, 476, 4662
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., 2013, Phys. Rev., 530, 87
Wu H.-Y., Weinberg D. H., Salcedo A. N., Walbing B. D., Za Y., 2019, MNRAS, 490, 2606
Zehavi I. et al., 2011, ApJ, 736, 59
Zhang Y. et al., 2019, MNRAS, 487, 2578
Zuntz J. et al., 2018, MNRAS, 481, 1149

APPENDIX A: INVESTIGATION OF BUZZARD MOCK CATALOGUES

During the analysis, we find one realization being different than other realizations. In the specific realization (realization 3b of BUZZARD and BUZZB), we find that the combination of galaxy clustering and galaxy–galaxy lensing recover cosmological parameters that are 3σ away from the true cosmology. The top panel of Fig. A1 shows the 68 and 95 per cent constraints from the combination of galaxy clustering and galaxy–galaxy lensing. We can see that realization 3b is clearly an outlier compares to other realizations. Since the combination of galaxy clustering and galaxy–galaxy lensing pipeline has been thoroughly validated (Krause et al. 2017; MacCrann et al. 2018), we interpret this as an indication that galaxies behave differently in realization 3b. We are then interested in whether it is the galaxy clustering or galaxy–galaxy lensing data vector that causes this bias, because only galaxy clustering goes into the data vector in the analysis of this paper. We analyse the galaxy clustering data vector alone by fixing the galaxy biases at galaxy biases measured from

| Simulations | BuzzA | BuzzB |
|-------------|-------|-------|
| ∆Ω_m | 0.051 ± 0.016 | 0.051 ± 0.016 |
| ∆σ_8 | -0.077 ± 0.020 | -0.067 ± 0.023 |
| Minimal confidence interval encompasses Δθ = 0 | 1.000 | 0.997 |
| P(sys < σ_Y1) | 0.28 | 0.50 |
| P(sys < 2σ_Y1) | 0.88 | 0.94 |

Table A1. This table is the same as Table 3 but includes realization 3b, a realization known to be problematic as shown in Appendix A.
cross-correlations of redMaGC galaxies and dark matter particles. We find that the galaxy clustering in realization 3b is problematic. The recovered cosmology is more than 3σ away from the truth. As a comparison, we show the galaxy clustering constraints in realization 4c, the second-most biased realization. The bias of cosmological parameters in realization 3b does not present in realization 4c. This finding indicates that the galaxy clustering behaves differently in realization 3b than other realizations. Further, as shown in the bottom panel of Fig. A1, we find a 3σ tension between galaxy clustering and galaxy–galaxy lensing in realization 3b, consolidating our conclusions that galaxy clustering in realization 3b is problematic. Although it is not included in our main analysis, we show the result of including this problematic realization in Table A1.

APPENDIX B: COMPARISON OF THE HALORUN AND THE FULL RUN

In Section 3, we develop the selection bias model based on redMaPPer Halorun, where redMaPPer is run fixing the cluster centres at the halo centres to avoid the ambiguity of associating galaxy clusters to dark matter haloes. The redMaPPer Halorun allows us to quantify the mass distribution of richness-selected galaxy clusters perfectly, which is important to quantify the selection bias. However, we will never be able to apply the same procedure on the data. Thus, it is crucial to understand the difference between redMaPPer Halorun, and the actual redMaPPer run (Fullrun), where redMaPPer is run with the same setting as the run on real data. Fig. B1 shows the comparison of the parameter constraints from the cluster abundance and two-point statistics measured from the Halorun and the Fullrun in one BUZZARD realization. We note that we do not expect the parameter

Figure A1. Top panel: Constraints on \( \Omega_m \) and \( \sigma_8 \) from galaxy clustering and galaxy–galaxy lensing (2 × 2pt) measured in each realizations in BUZZA; contours show the 68 and 95 per cent confidence levels. The blue contours denote the most biased realization and the orange contours denote the second most biased realization. Middle panel: Constraints on \( \Omega_m \) and \( \sigma_8 \) from galaxy clustering; contours show the 68 and 95 per cent confidence levels. In this analysis, we fix the galaxy biases at galaxy biases measured from cross-correlations of redMaGC galaxies and dark matter particles in each realization. Clearly, the figure shows that the galaxy clustering in realization 3b is biased. The same bias is not shown in the second most biased realization. This finding makes us believe that the galaxy clustering behaves differently in realization 3b from other realizations. Bottom panel: Constraints on \( \Omega_m \) and \( \sigma_8 \) from galaxy clustering and galaxy–galaxy lensing in realization 3b; contours show the 68 and 95 per cent confidence levels. In this analysis, we fix the galaxy biases at galaxy biases measured from cross-correlations of redMaGC galaxies and dark matter particles. Clearly, the figure shows that the galaxy clustering is in tension with the galaxy–galaxy lensing in realization 3b.

Figure B1. Constraints on cosmological parameters (\( \Omega_m \) and \( \sigma_8 \)), richness–mass relation parameters (\( \ln \alpha_0, A_\lambda, \sigma_{\text{intrinsic}} \), and \( B_{\lambda} \)), and selection bias parameters (\( b_{00}, b_{11} \)); contours show the 68 and 95 per cent confidence levels. Both colours are analysis on data vectors generated from the same BUZZARD realization. The orange lines are analysis on the redMaPPer Halorun, in which redMaPPer is run fixing the cluster centres at the halo centres. The blue lines indicate an analysis on the redMaPPer Fullrun, where redMaPPer is run with the same setting as the run on real data. The agreement between the orange and blue contours indicates that the Halorun is sufficient for constructing a model that describes the selection bias well in the Fullrun.
constraints from the Halorun and the Fullrun agree perfectly, since there is noise on the richness estimation in the Fullrun. Therefore, clusters whose richness is greater than 20, the lower richness cut in this analysis, in the Halorun have a different mass distribution from the clusters in the Fullrun. From Fig. B1, despite small differences, we can see both cosmological parameters ($\Omega_m$ and $\sigma_8$), richness–mass relation parameters ($\ln\lambda_0$, $A_\lambda$, $\sigma_{\text{intrinsic}}$, $B_\lambda$), and selection bias parameters ($b_{s0}$, $b_{s1}$), agree well. This indicates that the Halorun is sufficient for constructing a model that describes the selection bias well in the Fullrun.

APPENDIX C: COMPARISON OF SIMULATIONS TO THE DATA

In this section, we compare the conditional luminosity function in the simulations to the DES-Y1 data. The conditional luminosity function is a good tool to understand the relative brightness of satellites and central galaxies. Clearly, this quantity is closely related to the performance of the cluster finder. The measurement is done according to the method described in To et al. (2020). To account for the different richness–mass relations in simulations and data, we abundance match the simulations to the data. That is, for each

![Figure C1. Conditional luminosity function as a function of redshift $z$ and richness $\lambda$. Richness increases from top to bottom and redshift increases from left to right. The dashed line and solid lines represent the luminosity function of satellite galaxies and central galaxies, respectively. The shaded region and error bar correspond to $1 - \sigma$ uncertainties. The different colours correspond to different versions of BUZZARD as summarized in Table 1. For comparison, we overplot the measurement of the DES-Y1 data. To account for the differences of richness–mass relations in simulations and data, we abundance match the simulations to the data. That is, for each richness bin, we consider the top $X$ per cent most massive haloes in the simulation, where there are $X$ per cent of the clusters in the data that have richness greater than the given richness. In this plot, we focus on the relative brightness between centrals and satellites, an important quantity related to redMaPPer performance. The range between the three simulations spans the data.](image)
richness bin, we consider the top $X$ per cent most massive haloes in the simulation, where there are $X$ per cent of the clusters in the data that have richness greater than the given richness. Fig. C1 shows that while the conditional luminosity functions in the simulations do not match the data exactly, the range spanned by different simulations well covers the data.

**APPENDIX D: TINKER VERSUS EMULATOR**

One potential systematic in this analysis is the use of the Tinker halo mass function and the Tinker bias (Tinker et al. 2010), which are known to have 5–10 per cent systematic uncertainties (McClintock et al. 2019a, b). We test whether these uncertainties are ignorable in our analysis by comparing the parameter constraints from one BUZZARD realization estimated using the Tinker halo mass function and the Tinker bias with the constraints estimated using the halo mass function emulator (McClintock et al. 2019a) and the halo bias emulator (McClintock et al. 2019b). Fig. D1 shows the comparison of constraints on both cosmological parameters and nuisance parameters. The constraints from the Tinker halo mass function and the Tinker bias are consistent with the constraints from the emulator, indicating that the theory systematics from the Tinker halo mass function and the Tinker bias are subdominant in this analysis. We note that the above conclusion is only valid for a DES-Y1 like survey with cosmological parameters considered in this work. Additional tests are needed for applications on more sophisticated cosmological models and on future surveys with higher precision.

**APPENDIX E: COVARIANCE MATRIX**

The cluster shot noise in the theory covariance matrix depends on the expected cluster abundance, which is sensitive to the richness–mass relation parameters. To determine the richness–mass relation parameters of a given cosmology, we adopt an iterative approach. Given a cosmology, we first decide a fiducial richness–mass relation parameters to generate a covariance matrix. This covariance matrix is then used to analyse the cluster abundance and the clustering of galaxy clusters to determine the richness–mass relation parameters. We repeat this process until convergence. Since there is noise in the shot-noise estimation, it will be worrisome if the parameter constraints are sensitive to noise in the shot-noise estimation. To test this, we run our analysis with the covariance matrix generated in each iterations of the above iterative process. Fig. E1 shows that the parameter constraints are not sensitive to noise in the shot-noise estimation.

**APPENDIX F: REDSHIFT DEPENDENT SELECTION BIAS**

We consider the redshift-dependent $b_{sel}$, defined as

$$b_{sel}(M) = b_{sel}(M/M_{piv})^{b_{sel}(1+z)^{b_{sel}(1.45)}}$$

where $b_{sel}$ describes the redshift dependency of $b_{sel}$. We fit this model to the measured $b_{sel}$ in the Halorun of each realization and find 5 out of 21 realizations show $b_{sel} \neq 0$ at 2–3σ significance. To
Figure F1. Constraints on cosmological parameters ($\Omega_m$ and $\sigma_8$) and selection bias parameters ($b_{s0}$, $b_{s1}$, $b_{s2}$); contours show the 68 and 95 per cent confidence levels. $b_{s2}$ is the redshift dependence of $b_{sel}$ defined in equation (F1). The dashed lines denote the input cosmology of generating the simulation. The blue contours show the constraint assuming redshift independent $b_{sel}$, and the orange contours show the constraint assuming redshift dependent $b_{sel}$.

Figure G1. Constraints on $\Omega_m$ and $\sigma_8$ from galaxy clustering and galaxy–galaxy lensing (2 × 2pt) measured in BuzzA (top panel), BuzzC (middle panel), and BUZZARD v1.6 (bottom panel); contours show the 68 and 95 per cent confidence levels. In each panel, grey contours show the constraints from individual realizations and orange contours show the combination of these posteriors (equation 30). As a comparison, green contours show the constraints from cluster number counts and four two-point correlation functions (redMaGiC autocorrelations, redMaPPer–redMaGiC cross-correlations, redMaPPer autocorrelations, and cluster lensing), same as the orange contours in Fig. 5. The blue contours show the expected DES-Y1 constraints shifted to centre on the input cosmology of the simulation (same as the blue contours in Fig. 5). The black dashed lines indicate the true cosmology, i.e. the input cosmology used to generate the simulation. The bottom panel shows the result presented in MacCrann et al. (2018), which validates the 2 × 2pt pipeline on an older version of the BUZZARD mocks (BUZZARD v1.6).

APPENDIX G: COMPARISON WITH GALAXY–GALAXY LENSING AND GALAXY CLUSTERING

To understand whether the 2–3σ systematics shown in Fig. 5 are due to flaws in the analysis pipeline or other sources, such as statistical fluctuations or flaws in simulations, we run an analysis combining galaxy–galaxy lensing and galaxy clustering (2 × 2pt analysis) on each realization of BuzzA and BuzzC. We combine the constraints on $\sigma_8$ and $\Omega_m$ from analysis of each realization in the same way as we did in section 5. The result is shown in Fig. G1. We find that even in the most extreme case, including $b_{s2}$ into the posterior only shifts the contours of $\sigma_8$ and $\Omega_m$ by $\approx 0.5 \sigma$.