

Fermilab

Galaxy Clustering with LSST: Effects of Number Count Bias from Blending

FERMILAB-PUB-24-0864-V

arXiv:2411.14564

This manuscript has been authored by Fermi Research Alliance, LLC
under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy,
Office of Science, Office of High Energy Physics.

GALAXY CLUSTERING WITH LSST: EFFECTS OF NUMBER COUNT BIAS FROM BLENDING

BENJAMIN LEVINE,^{*1,2} JAVIER SÁNCHEZ,^{†3,4,5} CHIHWAY CHANG,^{2,4} ANJA VON DER LINDEN,¹
EBONI COLLINS,⁶ ERIC GAWISER,⁷ KATARZYNA KRZYŻAŃSKA,⁸ BORIS LEISTEDT,⁹
AND THE LSST DARK ENERGY SCIENCE COLLABORATION

¹Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA

²Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA

³Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

⁴Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

⁵Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA

⁶Department of Physics, Dillard University, 2601 Gentilly Blvd, New Orleans, LA 70122, USA

⁷Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA

⁸Department of Physics, Cornell University, 109 Clark Hall, Ithaca, NY 14853, USA and

⁹Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK

Version November 25, 2024

Abstract

The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) will survey the southern sky to create the largest galaxy catalog to date, and its statistical power demands an improved understanding of systematic effects such as source overlaps, also known as blending. In this work we study how blending introduces a bias in the *number counts* of galaxies (instead of the flux and colors), and how it propagates into galaxy clustering statistics. We use the 300 deg² DC2 image simulation and its resulting galaxy catalog (LSST Dark Energy Science Collaboration et al. 2021) to carry out this study. We find that, for a LSST Year 1 (Y1)-like cosmological analyses, the number count bias due to blending leads to small but statistically significant differences in mean redshift measurements when comparing an observed sample to an unblended calibration sample. In the two-point correlation function, blending causes differences greater than 3 σ on scales below approximately 10', but large scales are unaffected. We fit Ω_m and linear galaxy bias in a Bayesian cosmological analysis and find that the recovered parameters from this limited area sample, with the LSST Y1 scale cuts, are largely unaffected by blending. Our main results hold when considering photometric redshift and a LSST Year 5 (Y5)-like sample.

Subject headings: cosmological parameters – large-scale structure of the universe

1. INTRODUCTION

The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al. 2019) is projected to begin its 10-year survey of the southern hemisphere sky in late 2025. Both the depth and sky coverage of LSST are significantly enhanced with respect to Stage-III dark energy experiments (Albrecht 2006) such as the Dark Energy Survey (DES, Flaugher 2005), the Kilo-Degree Survey (KiDS, de Jong et al. 2015) and the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP, Aihara et al. 2018). These improvements are expected to lead to the detection of approximately 20 billion galaxies, improving the statistical power from a variety of dark energy probes (see the LSST Dark Energy Science Collaboration Science Requirement Document, The LSST Dark Energy Science Collaboration et al. 2018, hereafter “DESC SRD”). However, the improved statistical power will also lead to an increased sensitivity to systematics as previously-subdominant effects become relevant.

One systematic effect of particular interest to static probes of cosmology is blending systematics, an effect in which the detection and/or measurement algorithms mischaracterize—or fail to detect entirely—overlapping luminous sources. This effect is especially prominent

in faint galaxy populations, as the number density of objects increases steeply with fainter magnitude, meaning that the likelihood of overlapping another source with similar brightness also increases steeply. Blending can bias the centroid position, color, and shape of sources—all quantities that are used in cosmological probes of the large-scale structure (e.g., Abbott et al. 2018; Heymans et al. 2021). In the HSC-SSP, more than 50% of detected objects were “deblended” (Bosch et al. 2018); in image simulations of a joint Roman Space Telescope and Rubin Observatory survey, 20%–30% of Rubin single detections were identified as multiple blended objects in Roman data (Troxel et al. 2023). For a more detailed review of blending in cosmic surveys, see Melchior et al. (2021).

The effects of blending in the context of weak lensing analyses have seen interest from the community in recent years (e.g., Dawson et al. 2016; Sheldon & Huff 2017; Gruen et al. 2019; Sanchez et al. 2021; Nourbakhsh et al. 2022). Various techniques have been developed to minimize the impact to cosmic shear estimation (e.g., Bernstein et al. 2016; Sheldon et al. 2020) and photometric redshifts (e.g., Jones & Heavens 2019; MacCrann et al. 2021). New deblending techniques are also being developed in order to mitigate the effects of blending in different probes (e.g., Merlin et al. 2015; Joseph et al. 2016; Melchior et al. 2018; Arcelin et al. 2021). However, the

*benjamin.c.levine@stonybrook.edu

†jsanchez@stsci.edu

impact of blending in the study of two-point statistics of galaxy positions (i.e., galaxy clustering) has not received such attention. This lack of interest is partially due to the fact that blending is mostly thought to be important on very small scales where the separation between galaxies is comparable to the size of the galaxies themselves, and these spatial scales are usually not used in cosmological analyses. In addition, the effect is not seen in most current datasets due to the lower galaxy number density. However, as we approach the era of LSST, it is important to revisit the question of whether and how blending could affect galaxy clustering measurements in Stage-IV surveys.

In principle, blending will have three main effects on galaxy clustering measurements: 1) Unrecognized blends cause us to “lose” galaxies from the catalog; we thus count fewer pairs on the scales at which blending occurs, as well as the characteristic clustering scales of galaxies that cause blending and galaxies that are susceptible to blending. 2) Blending can cause errors in the brightness measurements of a galaxy, causing it to scatter into or out of a magnitude-selected sample. 3) Blending can cause errors in the color measurements of a galaxy, resulting in incorrect photometric redshift estimation.

The third effect is both complex to measure and highly algorithm-dependent (MacCrann et al. 2021); in addition, characterizing it would require very high accuracy and precision in simulating the galaxy population, which is hard to achieve. There are, however, ongoing efforts to produce reliable photometric redshifts for LSST (e.g., Schmidt et al. 2020). In Appendix A we explore a photometric redshift scenario using Bayesian Photometric Redshift (BPZ, Benítez 2000) redshift estimates from simulated photometry. However, due to the aforementioned complications, we leave a thorough examination of photometric redshift impacts for future study.

The first two effects are easier to assess and are the focus of this work. The LSST Data Challenge 2 (DC2, LSST Dark Energy Science Collaboration et al. 2021) image simulations allow us to implicitly capture both effects by comparing the “observed”¹ DC2 Year 1 (Y1) correlation function to the corresponding correlation function in the truth catalog, without needing to quantify some of the details (e.g., measuring how much the magnitudes are typically biased, or deciding how to identify a blend in detail). Then, to track down the specific effects of blending on galaxy clustering (as opposed to any other systematics introduced by the image processing pipeline), we choose a relatively simple definition to approximately distinguish between blended and unblended objects.

We emphasize here that, although we will spend considerable time describing our selected definition of blended and unblended objects, many of our results—in particular those concerning any shifts in the cosmology inference—are independent of this definition because they rely only on a comparison between the truth and observed catalogs. The definition of blends serves mainly to distinguish between biases that arise from blending and biases that arise from some other source.

We focus particularly on the effect on the redshift dis-

tribution and inferred values of matter density Ω_m and linear galaxy bias $b(\bar{z})$. We will not investigate any additional related issues (including color shifts, shape measurements, and centroid offsets), but future studies could in principle adopt the same tools and methodology to analyze them.

This work is structured as follows: in Section 2 we present the DC2 samples used for our analysis, as well as the methodology followed. In Section 3 we show the results regarding the bias to the redshift distributions of our samples. In Section 4 we present the clustering results obtained for the different samples considered in this work. In Section 5 we discuss the limitations of our analysis. Finally, in Section 6 we present concluding remarks.

2. ANALYSIS FRAMEWORK

2.1. Data

It is challenging to infer the impact of blends in LSST using existing survey data: some existing ground-based datasets are deep enough to achieve the required number densities to evaluate the impact of blending (such as the HSC-SSP deep and ultra-deep fields), but their small area is inadequate to test the full cosmological parameter space to the precision required by LSST. Using larger but shallower data (such as HSC-SSP wide) does not guarantee mapping the full parameter space relevant to blending (e.g., morphology, size, surface brightness, and clustering amplitude). One solution is to supplement shallower surveys with overlapping data from deeper, better-resolved images, e.g., with existing space-based imaging from HST; or very large telescopes with adaptive-optics systems. However, this approach quickly becomes limited by the expense of telescope time given the large number of galaxies we wish to study. In addition, there is no guarantee that there are no sources below the noise level that contribute to the measured flux of our detected sources. Future space-based surveys such as Roman (Spergel et al. 2015; Akeson et al. 2019) and Euclid (Laureijs et al. 2011) will provide the required data, but their data products will not be available in time to prepare for LSST. To this end, realistic end-to-end image simulations offer a unique opportunity to study the blending problem because we have access to both the catalog produced by our detection and deblending algorithms and, most importantly, the truth. In particular, the DESC DC2 simulations are developed specifically for LSST. Despite their limited area (approximately 300 deg^2) compared to the full planned area for LSST (approximately 18000 deg^2), the access to ground truth and their underlying number density makes them extremely useful tools with which to study the effects of blending.

DC2 is to an end-to-end photometric catalog based on the Outer Rim simulation (Heitmann et al. 2019). It includes a 440 deg^2 cosmological mock galaxy catalog (CosmoDC2) and a synthetic stellar catalog, simulated single-epoch and coadd images and their corresponding catalogs in six bands (u, g, r, i, z, y), and a combined forced-photometry catalog. A detailed description of the different data products from DC2 can be found in LSST Dark Energy Science Collaboration et al. (2021). We focus on the Wide-Fast-Deep (WFD) area shown in Fig-

¹ Throughout this work we will refer to the simulated DC2 photometric catalog as “observed” data, since it corresponds directly to what we expect to observe with LSST.

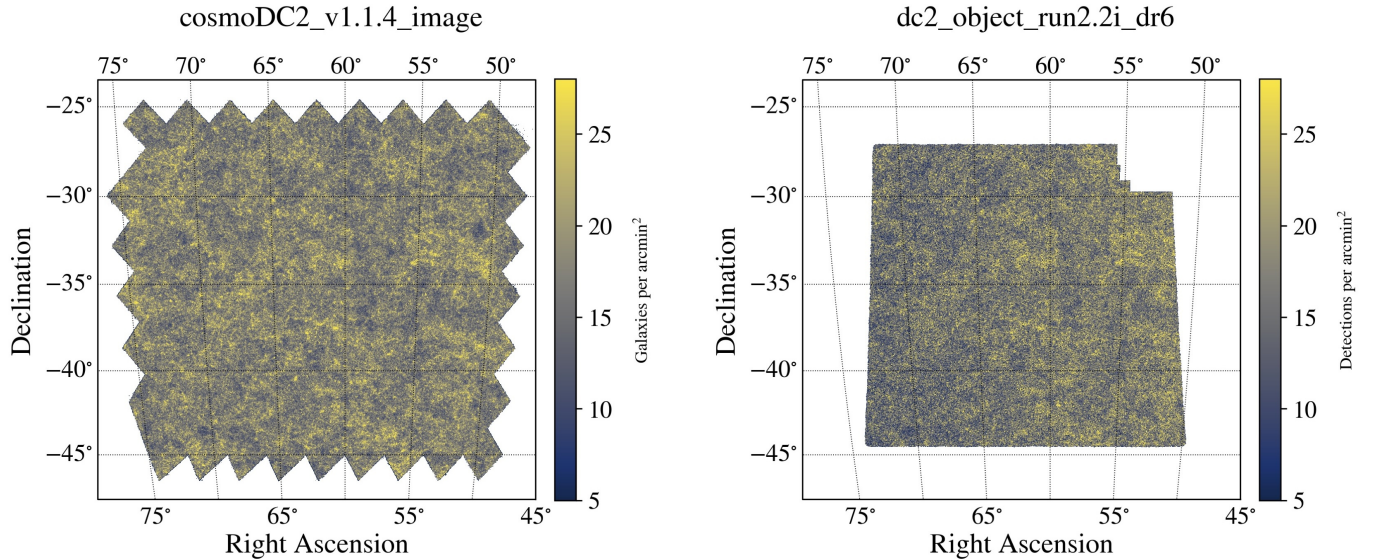


FIG. 1.— *Left*: map of the input (truth) galaxies from the CosmoDC2 simulation. *Right*: map of detected sources in the DC2 DR6 object catalog. The maps are shown using the McBryde projection.

ure 1. We use the `dc2_run2.2i_object_dr6` 5-year LSST depth “object catalog,” produced by detecting sources in the coadded images, for our observed data. Using this 5-year depth catalog guarantees a high completeness for the Y1-like subsample that we analyze here, minimizing potential impacts of incompleteness that may dominate over blending effects. This catalog is processed with version 19.0.0 of the Rubin Science Pipelines² (Jurić et al. 2017), and is designed to resemble the catalogs that will be produced during the operations of LSST. Detailed validation tests were carried out on the DC2 products as described in Kovacs et al. (2022), and additional validation is shown in Appendix B. To avoid contamination and systematic impacts on the sample by bright stars, we apply the star mask developed by Du (2023). Excluding this mask does not significantly affect our results.

The end-to-end simulation framework of DC2 offers us the unique opportunity to study blending in a realistic yet controlled way: we can match the galaxy catalog as detected by the Rubin software to the input “true” galaxy sample, in principle allowing us to determine the effects of blending on any given object. It is important to note that some of the effects of blending are tied to the choice of detection and deblending algorithm used for the analysis.³ When we measure properties based on the truth catalog, such as the intrinsic redshift distribution of isolated vs. blended objects (Section 3), any observed biases will be present independently of the detection and deblending algorithms, since our definition of isolated and blended objects is independent of these algorithms. However, for properties based on derived quantities in the observed catalog, such as flux and position (both of which propagate into the two-point correlation function; Section 4), the bias will be dependent on the specifics of the detection and deblending algorithms.

2.2. Matching truth and observed catalogs

The information of which photon was emitted by which source is not propagated through to the simulated images, and therefore there is no straightforward way to determine which truth objects were blended together into a single detection. As a consequence, we match objects in the observed catalog to objects in the truth catalog by selecting the truth object closest in r -band magnitude within a given radius R_{\max} , making the assumption that only objects within R_{\max} will likely be blended. We match in the r band because it is the deepest photometric band. The radius size $R_{\max} = 1''$ is chosen to be approximately the size of the typical point-spread function (PSF) of LSST images (estimated to be approximately $0.7''$ for the r band, Ivezić et al. 2019). We select galaxies with an observed magnitude of $i < 24.1$, which follows the LSST Y1 Gold sample magnitude cut as specified in the DESC SRD. We also require the selected objects to be measured as extended in each one of the g , r , i , and z bands (`<band>_base.ClassificationExtendedness_value > 0.5`), which removes the majority of bright stars that are incorrectly identified as galaxies by a simple `extendedness` cut.

Figure 2 shows that nearly 100% of our selected observed objects in DC2 are within $0.4''$ from their best truth match, suggesting that the pipeline adequately distinguishes between centroid separations greater than $1''$ —if not, we would expect to see more matches at larger radii. $R_{\max} = 1''$ is therefore a conservative choice of matching radius. Figure 2 also shows that over 95% of our observed objects are within 0.25 magnitudes of their best match in the truth catalog, providing further evidence for the reliability of our matching process.

We use a KDTree algorithm from SCIKIT-LEARN (Pedregosa et al. 2011) to count the number of neighbors within R_{\max} in the truth and observed catalogs for each observed detection. Our definition of neighbor includes the centered object; i.e., an isolated observed object with no other detections within R_{\max} will have one observed

² For additional details on the Rubin Science Pipelines see pipelines.lsst.io.

³ Additional details about the detection and deblending algorithms that we use can be found in Bosch et al. (2018).

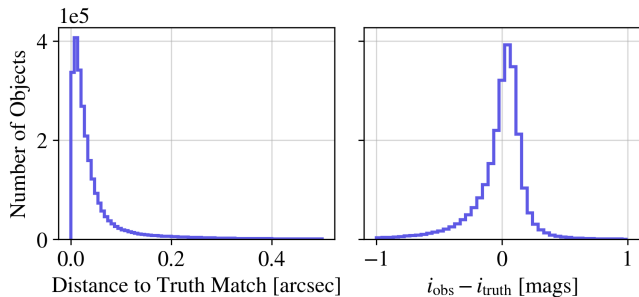


FIG. 2.— *Left*: histogram of the distance between an observed object and its best truth match. *Right*: histogram of the difference between the i -band magnitude of an observed object and its best truth match. For the right plot we use objects from `cosmoDC2_v1.1.4_small`, a representative subset of CosmoDC2. For both, the best match is defined as the truth object nearest in r -band magnitude to the detection within $1''$.

neighbor (itself). Then, we sort the matches into three broad categories (see Figure 3 for an illustration).

- **One-to-one matches:** all sources in the observed catalog that have only one neighboring source in the truth catalog within a radius $R_{\max} = 1''$, and one neighboring source within $R_{\max} = 1''$ in the observed catalog. These sources will most likely be unaffected by flux from any other objects (i.e., we can consider them isolated). The tails of very bright sources could in principle extend beyond R_{\max} and influence the determination of other sources of interest (Gawiser et al. 2006). However, we expect that the majority of the sources impacted by these effects will be masked by the bright star mask and thus excluded from the sample. Approximately 42% of observed galaxies in our sample are one-to-one matches.
- **Multiple-to-one matches:** all sources in the observed catalog that have more than one neighboring source within $R_{\max} = 1''$ in the truth catalog and one neighboring source within $R_{\max} = 1''$ in the observed catalog. The photometry and astrometry of these sources could be biased and the uncertainties may be correlated with neighboring sources. These detections are considered to be blended. Approximately 57% of observed galaxies in our sample are multiple-to-one matches.
- **Ambiguous matches:** all sources in the observed catalog which have more than one neighbor in the observed catalog. This category may also include shreds, in which a single true object may lead to multiple detections in the observed catalog. We leave the analysis of this category to future work.⁴ Given their rarity (approximately 1% of our sample) we do not expect ambiguous blends to significantly affect the results of this study. In addition, an analysis of this category would require a more sophisticated matching algorithm that takes into account the physical overlap of the measured shapes of objects; such algorithms are currently in

⁴ Note that the nomenclature for categories of blends is not standardized. The so-called “ambiguous blends” described in Dawson et al. (2016) correspond to our multiple-to-one matches; Mandelbaum (2018) uses “unrecognized blends” to refer to the same.

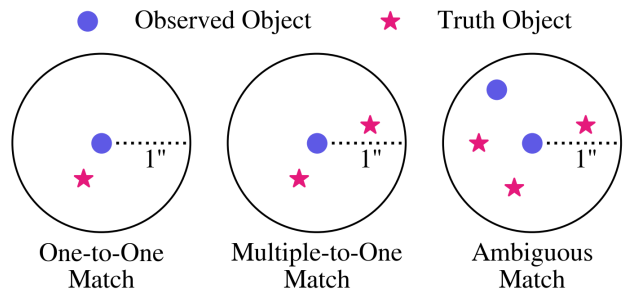


FIG. 3.— The selection criteria for matching objects between the observed and truth catalogs. We perform a search within $1''$ of the observed object. One-to-one matches (left) are defined as $N_{\text{truth}} = N_{\text{observed}} = 1$ and multiple-to-one matches (center) are defined as $N_{\text{truth}} > N_{\text{observed}} = 1$. Ambiguous matches (right) are defined as $N_{\text{observed}} > 1$.

development (e.g., Ramel et al. 2023; Liang, Adari et al. in prep.).

We count all neighbors in the truth and observed catalogs, up to a magnitude $i < 30$. In practice, there will not be significant bias to a galaxy blended with an object many magnitudes fainter than itself. However, determining what magnitude cutoff should be used to identify a “significant” or “strong” blend is nontrivial. Rather, the work here should be interpreted as the worst-case scenario—i.e., that almost every blended galaxy has unreliable photometry. We also reiterate that the exact definition of our one-to-one and multiple-to-one samples serves mainly to identify which biases are related to blending and which are not. The majority of our results are independent of this definition.

A benefit of using these distance-based categories compared to other criteria based, for example, on the degree of blending (such as blendedness; Bosch et al. 2018) is that they do not depend as strongly on the choice of detection and deblending algorithms.

It should be noted that it is extremely challenging to identify these three categories in real (unsimulated) data (though, as we will discuss in Section 3, a subsample of bright, isolated galaxies, such as those used for spectroscopic redshift calibration, should be close to a one-to-one sample). To emulate the expected LSST Y1 observations, we also present results for the “all observed” category, which consists of all galaxies detected in the DC2 observed catalog (the union of the three categories of blending described above).

We introduce a final category, defining “lost” objects as truth objects that are not a nearest match for a detection in the observed catalog. Note that the objects in this category do depend on the particular choice of detection and deblending algorithms.

In Figure 4 we show the distribution of observed sources as a function of the number of neighbors in the truth and observed catalogs. The histogram cells above the diagonal are spurious detections due to source shredding or poor pixel masking. The cells on the diagonal are sources that have been “correctly” detected, i.e., there are no biases between the input and output number of sources. The cells below the diagonal correspond to sources that have undetected neighbors. Approximately 42% of the sources are one-to-one matches, i.e., are most likely not blended, and approximately 57% of the de-

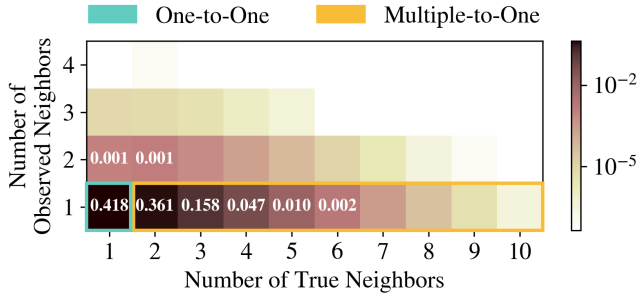


FIG. 4.— Distribution of neighbors in the truth and observed catalogs for each detected object, with a magnitude cut of $i < 24.1$. Cells corresponding to our one-to-one and multiple-to-one samples are highlighted.

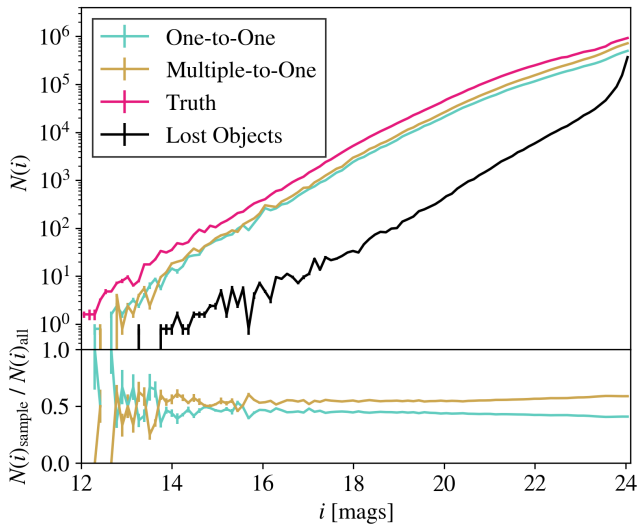


FIG. 5.— *Top*: number of objects in our samples at a given i -band magnitude. For the observed samples, we plot the observed magnitudes derived from the Rubin Science Pipelines. For the truth and lost samples, we plot the true magnitudes from the CosmoDC2 simulation. *Bottom*: deviation of the one-to-one and multiple-to-one samples from the all observed sample.

tected sources have more than one truth neighbor within $1''$, meaning that these sources likely present some degree of blending. This means that more than half of the detected objects in DC2 Data Release 6 (DR6) present some degree of blending (Melchior et al. 2021).

Note that this fraction is substantially higher than that found by Troxel et al. (2023), which is likely due the fact that we match to fainter objects and have a different definition of blends. Indeed, when we cut our truth catalog at $i < 27$ before matching, we find that only 17.5% of objects exhibit blending. Nonetheless, note that Figure 10 in Troxel et al. (2023) is qualitatively similar to our Figure 4.

In Figure 5 we show the i -band magnitude distribution of the one-to-one matches, multiple-to-one matches, and lost samples. For reference, we also plot the magnitude distribution of the truth catalog. Both detected samples closely follow the true distribution, while the lost objects remain comparatively-rare up to $i = 23$. Fainter objects are more difficult to detect, even if isolated, and are more sensitive to the presence of nearby luminous sources.

3. BIAS IN REDSHIFT DISTRIBUTION

Cosmological analyses that rely on properties of ensembles of galaxies, such as galaxy clustering and weak lensing shear, require an accurate description of the overall redshift distribution of the sample of galaxies involved. In this section, we test whether blending affects the calibration of these distributions using reference spectroscopic samples, such as in the DIR method (Lima et al. 2008), in a measurable way. We limit our analysis to testing for changes in the ensemble redshift distributions knowing the underlying truth redshifts and colors. However, blending can affect photometric redshift calibration in other ways (e.g., using unrecognized blends in the calibration sample, which can lead to biased redshift estimates). These effects, which are dependent on the calibration method used, are left for future work.

Figure 6 shows the probability density of each of our samples as a function of redshift. For the observed catalog, we use the true redshift of the best match in the truth catalog (as described in Section 2.2). The redshift of the multiple-to-one and lost samples are skewed to higher redshifts (to an extreme degree for the lost sample) compared to the one-to-one, all observed, and truth samples; i.e., blending tends to affect objects at higher redshifts. Given our definition of blends, one would naively expect blending to occur independently of both redshift and magnitude, since neither the redshift nor magnitude of objects ever enter into our $1''$ matching scheme. There are two factors which contribute to the observed bias. First, at higher redshifts, the angular diameter distance becomes smaller, meaning that physical clustering within a halo occurs at smaller angular scales. When the separation between halo members becomes smaller than $1''$, these high redshift galaxies will be identified as blends by our algorithm. Second, blending increases the measured flux of dim objects close to our magnitude cutoff, causing objects that would otherwise be excluded from our sample to be falsely included. Since these faint objects tend to lie at higher redshifts, we expect the high redshift end of our sample to be preferentially blended. These effects combined qualitatively explain the effect observed, though we will not attempt any explicit forward-modeling in this paper.

The mean and median redshifts of each observed distribution are shown in Table 1. The uncertainties are computed via jackknife with 10 bins, and are order $\mathcal{O}(10^{-6})$ —much smaller than the errors on the photometric redshifts that will be used in the LSST cosmological analysis. The measurements show small but significant differences between each sample, with the multiple-to-one sample having the highest mean redshift.

One consequence of this difference is that redshift calibration methods that rely on the characterization of isolated (one-to-one) galaxies may produce biased estimates of $N(z)$ in the presence of blends. Indeed, Figure 6 demonstrates that a one-to-one calibration sample overestimates the normalized $N(z)$ of the all observed sample by 2.27% for $z < 0.4$, and underestimates $N(z)$ by 5.92% for $z > 1.0$. These biases are statistically significant in DC2, and therefore will also be significant at the statistical level of LSST when trying to characterize the whole ensemble of galaxies at once. This means that the inclusion of nuisance parameters to allow certain perturbations in shape or mean of the $N(z)$, and/or alternative methods to cross-check the overall $N(z)$ will become even

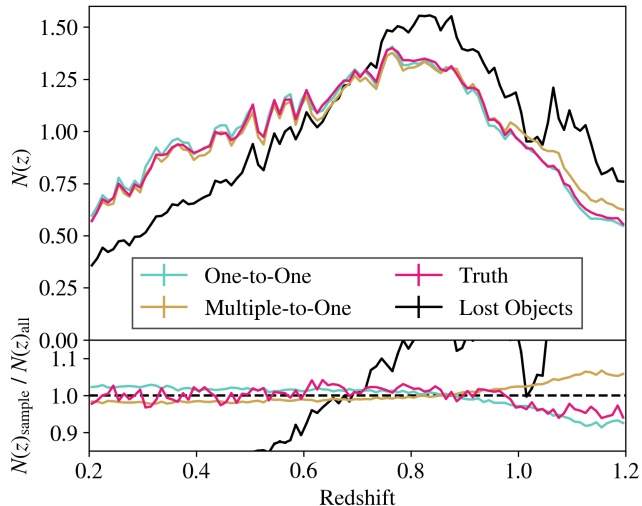


FIG. 6.— *Top*: number of objects in our samples at a given redshift, normalized such that the integral over all bins is equal to unity. For the observed samples, the redshift is given by the true redshift of the best truth match. *Bottom*: deviation from the all observed sample.

TABLE 1

MEAN AND MEDIAN REDSHIFT VALUES OF OUR SAMPLES. FOR THE OBSERVED CATALOG, WE USE THE TRUE REDSHIFT OF THE BEST MATCH IN THE TRUTH CATALOG (AS DESCRIBED IN SECTION 2.2). UNCERTAINTIES ARE NEGLIGIBLE.

Sample	Mean z	Median z
One-to-One	0.6982	0.7092
Multiple-to-One	0.7092	0.7220
All Observed	0.7042	0.7160
Truth	0.7021	0.7141

more important (e.g., Cawthon et al. 2022; García-García et al. 2023; Ruiz-Zapatero et al. 2023, among others).

We subdivide our samples into 5 different redshift tophat bins with 0.2 bin width, between $z = 0.2$ and $z = 1.2$, to estimate the effects on a tomographic analysis. This follows the Y1 analysis choices described in the DESC SRD. The mean redshifts for each bin, as well as the differences between the one-to-one and all observed samples, are shown in Table 2. The deviations in redshift between the tomographically-binned samples are smaller than in the unbinned case, but still statistically significant (except for the case of the $0.4 < z < 0.6$ bin, in which the one-to-one sample is consistent with the all observed sample).

The DESC SRD specifies the maximum uncertainty on the mean tomographic redshift to be $\sigma_{\bar{z}} < 0.005(1 + \bar{z})$. Our observed deviations between the one-to-one and all observed samples are about 10% of the upper bound on $\sigma_{\bar{z}}$. Biased estimates of $N(z)$ due to an isolated calibration sample may therefore contribute several percent of the DESC SRD redshift error budget. Numerous other systematic effects are also expected to bias $N(z)$, such as photometric redshift uncertainties and survey non-uniformity. If these effects correlate with blending, the impact could increase even further, but such analysis is beyond the scope of this work. Our results demonstrate the importance of minimizing and/or mitigating the impact of blending in cosmological analyses of LSST.

4. BIAS IN GALAXY CLUSTERING

4.1. Measuring the galaxy clustering signal

In LSST, two-point measurements of galaxy clustering are planned to be a pillar of the so-called “ 3×2 -point” cosmology analyses (Heymans et al. 2021; Abbott et al. 2022; Sugiyama et al. 2023), along with galaxy-galaxy lensing (two-point correlation of galaxy position and galaxy weak lensing) and cosmic shear (two-point correlation of galaxy weak lensing).

In this work we use the Landy & Szalay estimator (Landy & Szalay 1993) to estimate the two-point correlation function of the galaxy density field:

$$w(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (1)$$

where D stands for “data”, or the galaxy sample that traces large-scale structure, and R stands for “random”, a sample of random points that occupy the same footprint as D but are spatially randomly distributed. $DD(\theta)$ are the number of data-data pairs separated by an angle θ , $DR(\theta)$ is the number of data-random pairs, and $RR(\theta)$ is the number of random-random pairs. The correlation functions are estimated using TREECORR (Jarvis et al. 2004) and are computed in 12 log-spaced angular bins between $0.1'$ and $250'$. For our nominal analysis we will focus on scales where $k < 0.3h \text{ Mpc}^{-1}$, following the DESC SRD.

The covariance matrices are estimated via jackknife resampling on the larger SkySim5000 simulation⁵, which expands the footprint of the CosmoDC2 catalog to 5000 deg^2 . We use 150 jackknife patches. SkySim5000 is not an end-to-end image simulation, but it is a direct extension of the truth catalog; therefore, it allows us to obtain a covariance estimate that is as close as possible to the truth covariance. This estimate also has lower statistical noise thanks to SkySim5000’s larger footprint. We rescale the SkySim5000 covariances by the area ratio between the simulations in order to obtain the final DC2 covariance, using the so-called f_{sky} approximation (Knox 1995). Due to the conservative nature of the dataset in this analysis and the simplistic geometry of our footprint, the contributions to the covariance matrix in our analysis are dominated by the cosmic variance component, and therefore SkySim5000 provides a good approximation to the covariance matrix for the analysis.

4.2. Bias in the correlation function

We shift our focus to analyze and compare the two-point correlation functions obtained for the different galaxy samples (one-to-one, multiple-to-one, all observed, and the underlying truth galaxy catalog) that we introduce in Section 2.1, across the five tomographic bins described in Section 3. The results are shown in Figure 7.

The multiple-to-one sample has a larger amplitude than the one-to-one sample. The correlation function of the all observed sample lies in between, as expected given that it is a superset of the two. These differences are expected given our definition of the multiple-to-one

⁵ For additional information on SkySim5000 we refer the reader to github.com/LSSTDESC/gcr-catalogs.

TABLE 2

MEAN REDSHIFT VALUES FOR OUR TOMOGRAPHICALLY-BINNED SAMPLES. FOR THE OBSERVED CATALOG, WE USE THE TRUE REDSHIFT OF THE BEST MATCH IN THE TRUTH CATALOG (AS DESCRIBED IN SECTION 2.2). UNCERTAINTIES ARE NEGLIGIBLE. WE ALSO SHOW THE DEVIATION OF THE ONE-TO-ONE SAMPLE FROM THE ALL OBSERVED SAMPLE.

Mean Redshift					
Redshift	One-to-One	Multiple-to-One	All Observed	Truth	$z_{\text{all}} - z_{1\text{-to-1}}$
0.2-0.4	0.3078	0.3079	0.3080	0.3080	0.0002
0.4-0.6	0.5036	0.5037	0.5036	0.5036	0.0000
0.6-0.8	0.7037	0.7039	0.7038	0.7037	0.0001
0.8-1.0	0.8936	0.8947	0.8943	0.8939	0.0007
1.0-1.2	1.0890	1.0908	1.0901	1.0897	0.0011

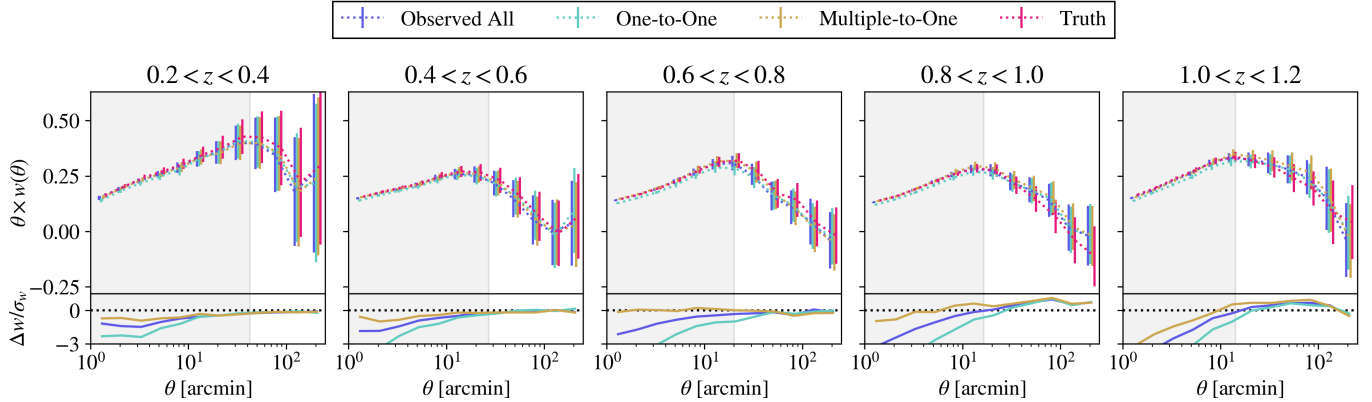


FIG. 7.— *Top*: angular autocorrelation function measurements for each of our samples. *Bottom*: deviation from the true correlation function. Error bars are computed as the square root of the covariance matrix diagonal, with the covariance matrix estimated from SkySim5000 as described in Section 4.1. Gray shaded regions denote small scales ignored in cosmological analyses, as specified by the DESC SRD ($k < 0.3h \text{ Mpc}^{-1}$). Note that the apparent redshift dependence on the bias in the bottom panel is largely dominated by the improved statistics in the higher redshift bins, which lead to reduced uncertainties.

sample: projected alignments between objects will occur more often in environments with a higher clustering bias, resulting in a higher correlation function. The one-to-one sample, which selects isolated objects, should have a comparatively lower correlation function.

At small scales ($< 10'$), all three observed samples exhibit less clustering than the truth; for the all observed sample (what we will measure in LSST) the deviation can be beyond 3σ as measured in DC2.

At larger scales, the observed samples are mostly consistent with the truth but show deviations on the order of 0.5σ in the higher two redshift bins. However, these deviations at large scales are consistent between the one-to-one and multiple-to-one samples, suggesting that the origin is likely not blending.

We extrapolate these DC2-based results to the full LSST area by rescaling the uncertainties by the area ratio, $A_{\text{ratio}} = \sqrt{A_{\text{LSST}}/A_{\text{DC2}}} \approx 7.8$, where $A_{\text{LSST}} = 18000 \text{ deg}^2$ and $A_{\text{DC2}} = 294 \text{ deg}^2$. On small scales, this may translate into deviations of up to 21σ or larger. These deviations may complicate efforts to measure and model nonlinear galaxy bias (Nicola et al. 2024) and other small scale effects (e.g., Chisari et al. 2019a).

4.3. Bias in inferred cosmology

We perform a cosmological analysis on our measured correlation functions following the analysis choices from the DESC SRD. We model the correlation function using the CORE COSMOLOGY LIBRARY (CCL, Chisari et al. 2019b) with a linear galaxy bias for each redshift bin. First, we obtain the angular power spectrum:

$$C_\ell = b^2(\bar{z}) \frac{2}{2\ell + 1} \int dz \left(\frac{dn}{dz} \right)^2 H^2(z) P \left(k = \frac{\ell + 1}{r(z)}, z \right), \quad (2)$$

where $b(\bar{z})$ is the galaxy bias at the mean redshift of the sample, dn/dz is the redshift distribution, $H(z)$ is the Hubble parameter, and $P(k)$ is the matter-power spectrum as computed by CAMB (Lewis et al. 2000). An inverse Fourier transform yields the correlation function:

$$w(\theta) = \sum_{\ell \geq 0} \left(\frac{2\ell + 1}{4\pi} \right) \mathcal{P}_\ell(\cos \theta) C_\ell, \quad (3)$$

where \mathcal{P}_ℓ is the Legendre polynomial of order ℓ .

We obtain the posterior probability for the cosmological analysis by evaluating the Bayesian likelihood:

$$\log \mathcal{L} = -0.5 \sum_i \sum_{j,k} (w_i(\theta_j) - w_i^{\text{th}}(\theta_j)) C_{i,jk}^{-1} (w_i(\theta_k) - w_i^{\text{th}}(\theta_k)), \quad (4)$$

where i runs over the redshift bins and j, k each run over the angular bins. $C_{i,jk}^{-1}$ is the inverse covariance matrix for the bin i . We ignore the correlations between redshift bins given that we are using top-hat bins in true redshift. However, we do not expect the conclusions from this work to be affected by these approximations. This is due to the fact that we are ignoring the effects in photometric redshifts, and that in a typical cosmological analyses the information comes from both the auto- and cross-power

TABLE 3

TOP: PRIORS USED FOR OUR LIKELIHOOD ESTIMATION. $\mathcal{U}(a, b)$ REPRESENTS A UNIFORM (TOP-HAT) PRIOR BETWEEN a AND b . $\mathcal{N}(\mu, \sigma)$ REPRESENTS A GAUSSIAN PRIOR WITH MEAN μ AND STANDARD DEVIATION σ . BOTTOM: FIXED VALUES FOR COSMOLOGICAL PARAMETERS USED IN OUR MODEL. THE PARAMETERS ARE SET TO THEIR DC2 INPUT VALUES (KORYTOV ET AL. 2019).

Parameter	Prior/Value
Ω_m	$\mathcal{U}(0.1, 0.5) \times \mathcal{N}(0.2648, 0.2)$
b_i	$\mathcal{U}(0.3, 4.0) \times \mathcal{N}(1.9, 0.9)$
Ω_b	0.0448
σ_8	0.8
h	0.71
n_s	0.963

spectra; however, for the case of top-hat bins the same information is encoded in the auto spectra.

Using the package EMCEE (Foreman-Mackey et al. 2013), we run a Markov Chain Monte Carlo process (MCMC) to estimate the posterior distributions for Ω_m and the galaxy bias parameters $b_i(\bar{z})$. We set truncated Gaussian priors for Ω_m and b_i ; the parameters of the Gaussian are as specified by the DESC SRD, while the truncation boundaries are set so as to prevent walkers from wandering beyond physically meaningful values. These priors are shown in Table 3. The free parameters in our model are those to which galaxy clustering is the most sensitive (Percival et al. 2001). All other cosmological parameters are set to their DC2 simulation input values for the purposes of modeling the correlation function.

Our likelihood function is given in Equation 4. For each chain we use gradient minimization to set starting values, run 32000 samples, and discard 100 samples for burn-in. We confirm that the burn-in period is adequate and that the chains converge by visually inspecting the sample distributions to check that there is no evolution in mean and variance over time. To visualize the posteriors we use the Kernel Density Estimator function of CHAINCONSUMER (Hinton 2016).

The MCMC posteriors are shown in Figure 8. The cosmological constraints obtained for all of our observed samples are consistent with those obtained for the truth sample within 1σ . Additionally, all of the constraints obtained for the galaxy samples considered here are consistent with the DC2 input ($\Omega_m = 0.265$) within $< 2\sigma$, although we find a small preference towards higher than expected Ω_m values. In Appendix B we explore this deviation, finding that this preference is due to sample variance. The most significant differences in the inferred parameters between the samples are found in the galaxy biases, particularly in the one-to-one sample at high redshifts. This is expected, as the one-to-one sample is selected in a way that is naturally less clustered than the other two observed samples. We show the fractional uncertainty for the biases for each sample in Table 4. From the posterior distributions in Figure 8 we conclude that blending has no measurable impact on the inferred cosmology using the DC2 footprint with LSST Y1 fiducial analysis choices.

As discussed in Section 4.2, the deviation between the truth and observed samples is amplified at smaller scales which are not used in the cosmological analysis. To test

TABLE 4

FRACTIONAL DEVIATION FROM THE TRUTH SAMPLE FOR EACH OF THE COSMOLOGICAL PARAMETERS ($\Delta X = X_{\text{Truth}} - X_{\text{Sample}}$).

	One-to-One	Multiple-to-One	All Observed
$\Delta\Omega_m/\sigma$	0.355	0.253	0.224
$\Delta b_1/\sigma$	0.162	0.212	0.135
$\Delta b_2/\sigma$	0.330	0.161	0.139
$\Delta b_3/\sigma$	0.571	0.064	0.160
$\Delta b_4/\sigma$	0.709	0.278	0.439
$\Delta b_5/\sigma$	0.425	0.100	0.201

how the choice of small-scale cut could impact the inferred cosmology, we compute the best-fit Ω_m for various scale cutoffs using gradient descent minimization. The results are shown in Figure 9; we find that the best-fit Ω_m value at smaller scale cuts is compatible with our fiducial analysis within 1σ for all but the most conservative scale cut (20 Mpc). Note, however, that the increased uncertainty on the inferred cosmology for the 20 Mpc scale cut would likely bring it into 1σ concordance with our fiducial analysis as well.

5. LIMITATIONS

Our fiducial analysis includes two major simplifications to the data set. First, we use the true redshifts of each observed galaxy, generated by matching the observed catalog to the true catalog. Photometric redshifts will add a further complication to the measurements. Second, we perform a magnitude cut at the Y1 Gold limit, despite the fact that DC2 is measured at a Y5 depth.

To test whether these simplifications affect our conclusions, we re-run the analysis using the LSST Y5 magnitude cut ($i < 24.92$) and a BPZ photometric redshift catalog for DC2. Our main results hold in these secondary analyses. The procedure and results of these tests are described in detail in Appendix A.

These tests provide evidence that blending will not significantly impact Ω_m measurements from galaxy clustering analyses in the context of a Y5 analysis with photometric redshift effects. However, for the reasons discussed in Section 1, it is difficult to isolate the exact impact of blending in these measurements as opposed to the effects of incompleteness and biases introduced by the photometric redshift algorithm itself.

In addition to these simplifications, we also highlight three additional areas for future study. First, most cosmological analyses performed to date limit the usage of galaxy-clustering results to linear scales (e.g., Heymans et al. 2021; Abbott et al. 2022) on their baseline analyses. However, there are significant efforts in the community to improve our understanding of small scales (Nicola et al. 2024) and unlock additional cosmological information. In our baseline analysis we focus on linear scales and the impact of blending on the linear bias. However, Figure 7 clearly shows that the impact of blending is more significant at small (non-linear) scales. Thus, beyond improvements in modeling non-linearities, additional work understanding and modeling the impact of blending at these scales will be required in order to push further into the non-linear regime using LSST data.

Second, we consider galaxy clustering in isolation and do not address the effect of blending in the full $3 \times 2pt$ analyses. Nourbakhsh et al. (2022) find significant im-

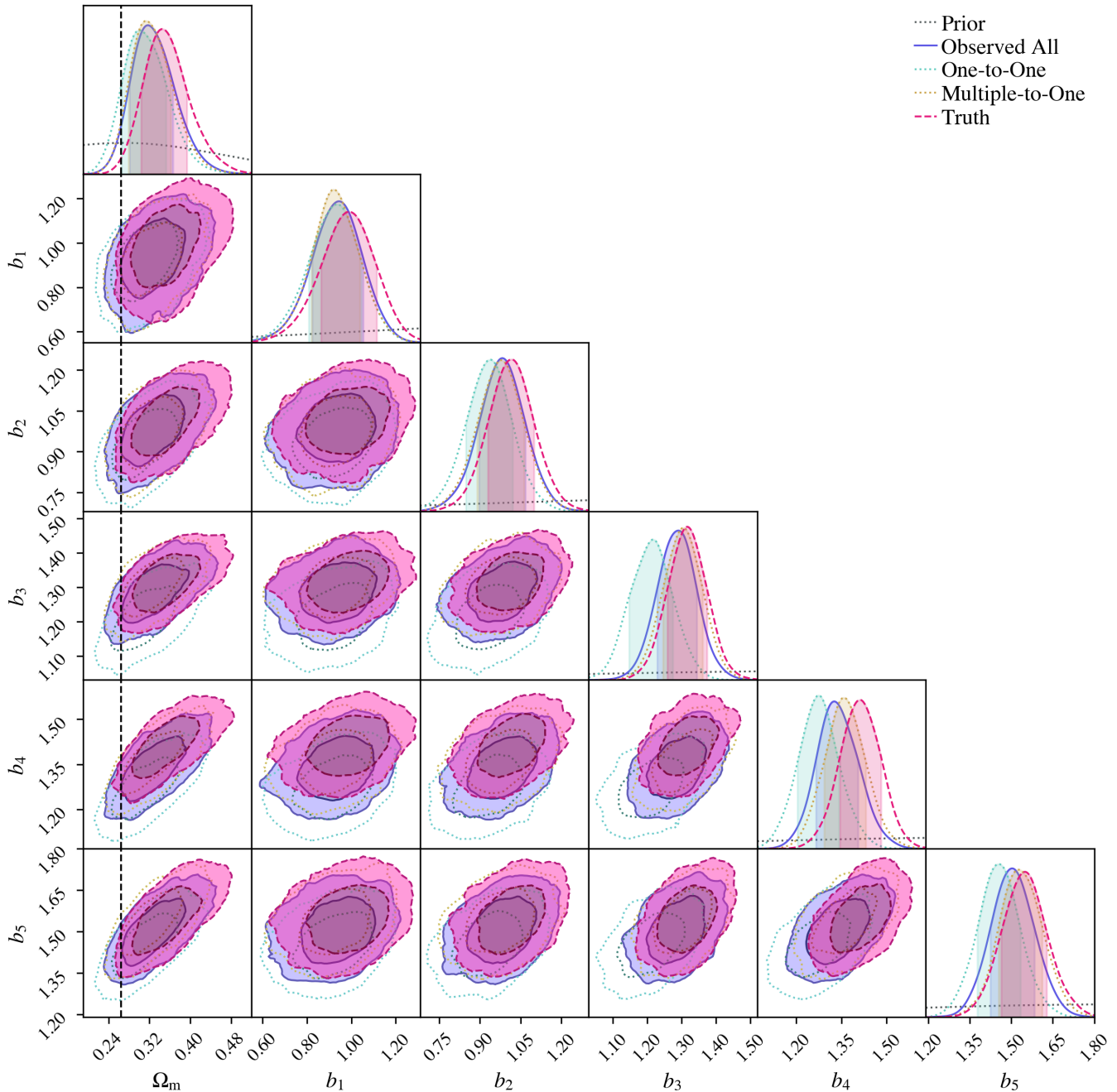


FIG. 8.— Contours for cosmological parameters assuming Λ CDM model with Ω_m and galaxy bias b_i freed, computed for each of our samples. We use DESC SRD priors and show the simulation input value of $\Omega_m = 0.265$ as a vertical dashed line.

pect in cosmic-shear-only analyses, and our results point to significant impact in clustering-only analyses for the full LSST area (though not in the smaller footprint of DC2), which motivates a deeper study of the impact in 3×2 pt analyses.

Finally, we note that the galaxies used for the clustering analysis are detected and deblended using a particular detection and deblending algorithm (those included in the Rubin Science Pipelines v19.0.0), and the specifics of the effect of blending found in the two-point correlation function do depend ultimately on these. However, potential improvements in these algorithms and/or combinations with space-based higher-resolution imaging can

mitigate the impact of blending in galaxy clustering measurements found in our work.

6. SUMMARY & CONCLUSIONS

In this work we study the impact of the number count bias from blending in two-point galaxy clustering analyses for the Vera C. Rubin Observatory LSST.⁶

Using data from the DC2 simulated sky survey gives us the opportunity to analyze LSST-like data while still having access to the ground truth, which provides an

⁶ The code used for our analysis of catalogs, correlation functions, and cosmology, as well as plotting routines, is available at github.com/bclevine/BLxClustering.

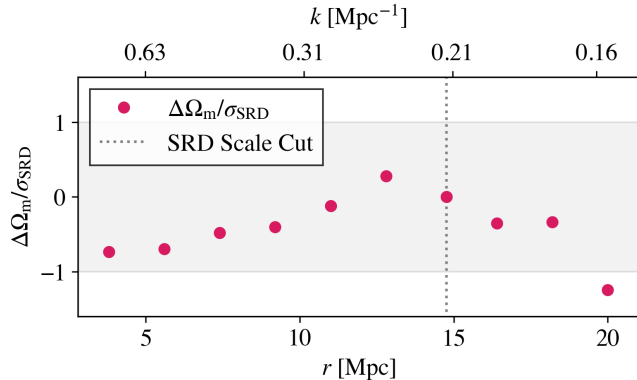


FIG. 9.— Relative deviation of best-fit Ω_m from our fiducial cosmological analysis ($\Delta\Omega_m = \Omega_{m,\text{best-fit}} - \Omega_{m,\text{SRD}}$) compared to the uncertainty at our fiducial scale cut (σ_{SRD}), as a function of scale cut. The gray shaded region shows the 1σ range. The DESC SRD scale cut used in our cosmological analysis is shown with a gray dotted line.

ideal scenario to test the effects of blending. After comparing and matching with the truth catalog we subdivide the observed galaxies into two main samples: the one-to-one matches, i.e., observed galaxies that have a single neighbor in the truth catalog within a $R_{\text{max}} = 1''$ radius; and the multiple-to-one matches, i.e., those observed galaxies that have more than one neighbor in the truth catalog within a $R_{\text{max}} = 1''$ radius. These two categories roughly correspond to likely-not-blended and likely-blended galaxies, respectively. By using a distance-based classification criterion we attempt to make our conclusions as deblender-independent as possible.

We follow the fiducial Y1 analysis presented in the DESC SRD and analyze potential differences between the one-to-one sample, the multiple-to-one sample, and the combination of both (the all observed sample), as well as the underlying truth. We first look for differences in the ensemble redshift distributions, finding that the one-to-one sample has a different mean redshift than the multiple-to-one and the all observed samples. These biases, while statistically significant, will likely be subdominant compared to other sources of mean redshift bias such as photometric redshift errors. Further study of how different redshift calibration algorithms interact with blending will be required in order to quantify the full impact to $N(z)$ measurements.

Second, we compare the two-point correlation functions obtained for the one-to-one, multiple-to-one, all observed, and truth galaxies. We find that at small, nonlinear scales the differences between the observed and truth correlation functions are highly significant ($> 3\sigma$), but at linear scales we do not find any measurable systematic biases due to blending. When extrapolating these results to the full LSST area, these differences at small scales become extremely significant ($> 21\sigma$). These differences advise for caution when pushing to the small-scale regime using galaxy clustering two-point measurements in future cosmological analyses, as systematic biases may arise if the effect of blending is not accounted for properly. We find some mild deviation in the measured correlation functions between the observed galaxies and the truth at linear scales. However, these are most likely not due to blending, but rather to other observational effects, as the three observed samples (one-to-one,

multiple-to-one, and all observed) all exhibit the same effect.

We use the measured two-point data vectors to obtain constraints on the combination of Ω_m and galaxy bias parameters. We find that the inferred posterior distributions for Ω_m for all samples (one-to-one, multiple-to-one, all observed, and truth) are statistically compatible. Furthermore, we obtain best-fit Ω_m values statistically compatible with the input $\Omega_m = 0.265$ used for DC2. The inferred best-fit Ω_m values are slightly higher than the simulation input, but we confirm that this is due to sample variance present in the DC2 area. We obtain the same results from cosmology analyses using a Y5 magnitude cut and photometric redshifts.

Our results provide evidence that, for the LSST Y1 Gold sample with the fiducial scale cuts, the number count bias from blending will not significantly impact the results of galaxy clustering analyses. However, for cosmology beyond the fiducial Y1 analysis, it will be important to 1) measure the correlation of blending with other sources of redshift bias to confirm that blending remains subdominant; and 2) study and model the impact of blending on angular scales smaller than our cutoff of $k = 0.3h \text{ Mpc}^{-1}$, allowing us to take advantage of the statistical power of the correlation function on nonlinear scales.

ACKNOWLEDGMENTS

This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The internal reviewers were Cyrille Doux, Eric Gawiser, and Boris Leistedt. The authors acknowledge the usage of ASTROPY, CCL, EMCEE, HEALPY, MATPLOTLIB, NUMPY, PANDAS, SCIKIT-LEARN, SCIPY, SKYPROJ, and TREECORR. B. Levine and A. von der Linden are supported by DOE grant DE-SC0023387. C. Chang is supported by DOE grant DE-SC0021949. E. Gawiser is supported by DOE grant DE-SC0010008. B. Leistedt is supported by the Royal Society through a University Research Fellowship.

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne - France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BEIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

The contributions from the authors are as follows. B. Levine performed the main analysis of redshift, angular correlation function, and cosmological measurements. J.

Sánchez supervised the project, performed the nearest-neighbors matching, and produced correlation function measurements for SkySim5000 and some instances of DC2. B. Levine and J. Sánchez led the paper writing. J. Sánchez and C. Chang proposed the project and organized collaborators. C. Chang and A. von der Linden

provided guidance on both the analysis and text. E. Gawiser and B. Leistedt offered extensive feedback as internal reviewers, leading to paper updates. C. Chang, E. Gawiser, and J. Sánchez are DESC builders. K. Krzyżńska and E. Collins contributed to initial stages of the project and initial analyses.

REFERENCES

- Abbott T. M. C., et al., 2018, *Phys. Rev. D*, **98**, 043526
 Abbott T. M. C., et al., 2022, *Phys. Rev. D*, **105**, 023520
 Aihara H., et al., 2018, *PASJ*, **70**, S8
 Akeson R., et al., 2019, arXiv e-prints, p. arXiv:1902.05569
 Albrecht A., 2006, in APS April Meeting Abstracts. APS Meeting Abstracts. p. G1.002
 Arcelin B., Doux C., Aubourg E., Roucelle C., LSST Dark Energy Science Collaboration 2021, *MNRAS*, **500**, 531
 Benítez N., 2000, *ApJ*, **536**, 571
 Bernstein G. M., Armstrong R., Krawiec C., March M. C., 2016, *MNRAS*, **459**, 4467
 Bosch J., et al., 2018, *PASJ*, **70**, S5
 Cawthon R., et al., 2022, *MNRAS*, **513**, 5517
 Chisari N. E., et al., 2019a, *The Open Journal of Astrophysics*, **2**, 4
 Chisari N. E., et al., 2019b, *ApJS*, **242**, 2
 Dawson W. A., Schneider M. D., Tyson J. A., Jee M. J., 2016, *ApJ*, **816**, 11
 Du Z., 2023, PhD thesis, UC Riverside
 Flaugh B., 2005, *International Journal of Modern Physics A*, **20**, 3121
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
 García-García C., Alonso D., Ferreira P. G., Hadzhiyska B., Nicola A., Sánchez C., Slosar A., 2023, *J. Cosmology Astropart. Phys.*, **2023**, 025
 Gawiser E., et al., 2006, *ApJS*, **162**, 1
 Gruen D., et al., 2019, *MNRAS*, **488**, 4389
 Heitmann K., et al., 2019, *ApJS*, **245**, 16
 Heymans C., et al., 2021, *A&A*, **646**, A140
 Hinton S. R., 2016, *The Journal of Open Source Software*, **1**, 00045
 Ivezić Ž., et al., 2019, *ApJ*, **873**, 111
 Jarvis M., Bernstein G., Jain B., 2004, *MNRAS*, **352**, 338
 Jones D. M., Heavens A. F., 2019, *MNRAS*, **483**, 2487
 Joseph R., Courbin F., Starck J. L., 2016, *A&A*, **589**, A2
 Jurić M., et al., 2017, in *Astronomical Data Analysis Software and Systems XXV*. p. 279 (arXiv:1512.07914), doi:10.48550/arXiv.1512.07914
 Knox L., 1995, *Phys. Rev. D*, **52**, 4307
 Korytov D., et al., 2019, *ApJS*, **245**, 26
 Kovacs E., et al., 2022, *The Open Journal of Astrophysics*, **5**, 1
 LSST Dark Energy Science Collaboration et al., 2021, *ApJS*, **253**, 31
 Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64
 Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
 Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, **538**, 473
 Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, *MNRAS*, **390**, 118
 MacCrann N., et al., 2021, *Monthly Notices of the Royal Astronomical Society*, **509**, 3371–3394
 Mandelbaum R., 2018, *Annual Review of Astronomy and Astrophysics*, **56**, 393
 Melchior P., Moolekamp F., Jerdee M., Armstrong R., Sun A. L., Bosch J., Lupton R., 2018, *Astronomy and Computing*, **24**, 129
 Melchior P., Joseph R., Sanchez J., MacCrann N., Gruen D., 2021, *Nature Reviews Physics*, **3**, 712
 Merlin E., et al., 2015, *A&A*, **582**, A15
 Nicola A., et al., 2024, *J. Cosmology Astropart. Phys.*, **2024**, 015
 Nourbakhsh E., Tyson J. A., Schmidt S. J., LSST Dark Energy Science Collaboration 2022, *MNRAS*, **514**, 5905
 Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, **12**, 2825
 Percival W. J., et al., 2001, *MNRAS*, **327**, 1297
 Ramel M., Doux C., Kuna M., 2023. p. arXiv:2310.02079 (arXiv:2310.02079), doi:10.48550/arXiv.2310.02079
 Ruiz-Zapatero J., Hadzhiyska B., Alonso D., Ferreira P. G., García-García C., Mootoivaloo A., 2023, *MNRAS*, **522**, 5037
 Sanchez J., Mendoza I., Kirkby D. P., Burchat P. R., LSST Dark Energy Science Collaboration 2021, *J. Cosmology Astropart. Phys.*, **2021**, 043
 Schmidt S. J., et al., 2020, *MNRAS*, **499**, 1587
 Sheldon E. S., Huff E. M., 2017, *ApJ*, **841**, 24
 Sheldon E. S., Becker M. R., MacCrann N., Jarvis M., 2020, *ApJ*, **902**, 138
 Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
 Sugiyama S., et al., 2023, *Phys. Rev. D*, **108**, 123521
 The LSST Dark Energy Science Collaboration et al., 2018, arXiv e-prints, p. arXiv:1809.01669
 Troxel M. A., et al., 2023, *Monthly Notices of the Royal Astronomical Society*, **522**, 2801–2820
 de Jong J. T. A., et al., 2015, *A&A*, **582**, A62

redshifts (determined as the mean of the BPZ probability density function), and then feed the true $N(z)$ distribution for each of these photometric-redshift-selected bins into CCL. This corresponds to a scenario where we assume that we still have access to the true redshift distribution (i.e., we ignore any blending-related biases in the $N(z)$ distribution), since trying to estimate the impact of blending on $N(z)$ depends on the choice of $N(z)$ estimation method and is therefore beyond the scope of this paper. It is possible that blending-related errors in the calibrated $N(z)$ due to an imperfect estimator may lead to larger overall biases than we measure in this section. The 1σ confidence intervals we obtain are shown in Figure 10.

Our main conclusions from the fiducial analysis are also apparent in these Y5 and photometric redshift tests. In all cases, the true value of Ω_m is recovered within $< 2\sigma$, and the fits to the observed data are consistent with the fits to the truth data within 1σ . The inferred values of galaxy bias are slightly lower than the fiducial anal-

APPENDIX

A. RESULTS USING YEAR 5 MAGNITUDE CUTS AND PHOTOMETRIC REDSHIFTS

In the main text we assert that the use of Y1 magnitude cuts and true redshifts does not significantly affect our final conclusions (see Section 5). Here, we describe in detail our results when using Y5 depth and photometric redshifts. Note that the DESC SRD only contains Y1 and Y10 cuts, so to obtain the Y5 magnitude cut we interpolate between the Y1 and Y10 cuts, using $\sqrt{\text{time}}$ to scale the flux density, to obtain $i < 24.92$.

We exactly repeat our fiducial analysis using the LSST Y5 magnitude cut and a BPZ photometric redshift catalog for DC2. The latter can enter into the pipeline in two ways: first through the binning of galaxies when computing the correlation functions, and second through the $N(z)$ provided to CCL. For this test, we bin galaxies in the correlation functions according to their photometric

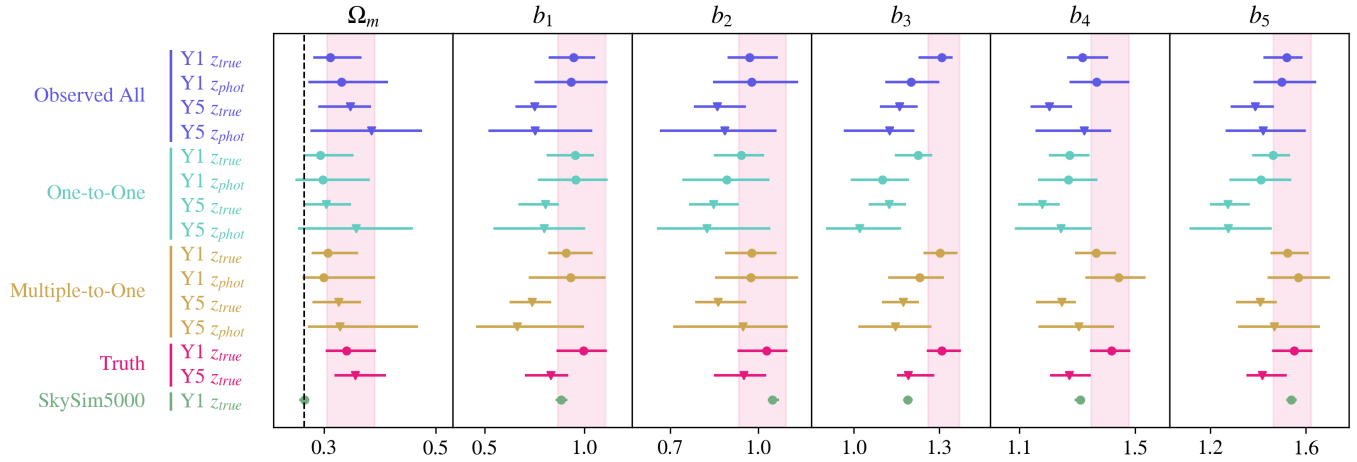


FIG. 10.— 1σ ranges for the inferred cosmological parameters of each sample, using different combinations of magnitude cutoff (Y1/Y5) and binning redshift (true redshift/photometric redshift). The first three samples (“observed all,” “one-to-one,” and “multiple-to-one”) are drawn from the DC2 catalog as described in Section 2.2. The magenta “truth” sample is drawn from the CosmoDC2 truth catalog. The green sample in the bottom row corresponds to the SkySim5000 catalog, which we detail in Appendix B. The simulation input value of $\Omega_m = 0.265$ is shown as a vertical dashed line. The magenta shaded regions show the 1σ range for the truth data in our fiducial (Y1) analysis.

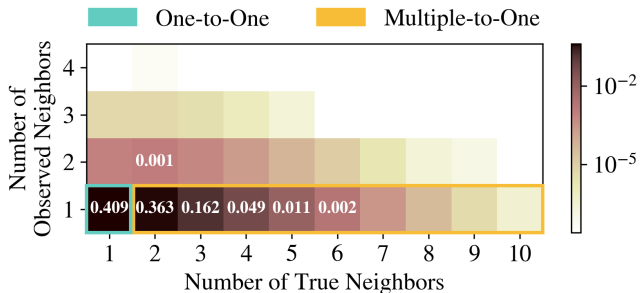


FIG. 11.— Distribution of neighbors in the truth and observed catalogs for each detected object, with a Y5 magnitude cut of $i < 24.92$. Cells corresponding to our one-to-one and multiple-to-one samples are highlighted.

ysis for the Y5 tests by up to 1σ . However, since these shifts are consistent between all three of our observed samples, they are unlikely to be caused by blending. We do not attempt further analysis to determine the source of the shifts in galaxy bias, but they are expected: using the Y5 magnitude cut introduces to the sample fainter galaxies that are less massive and less clustered, lowering the overall bias.

We now describe the characteristics of the blended samples in the Y5 and photometric redshift analyses and compare them to our main analysis. When using Y5 depth, the number of galaxies in our analysis between $0.2 < z_{\text{true}} < 1.2$ nearly doubles from 13.1 million to 22.1 million. The fraction of blended objects remains approximately the same (Figure 11), due to the fact that our definition of blends is independent of the sample magnitude cut and the blending rate is nearly uniform across magnitudes (Figure 5). The $N(z)$ distribution within the defined redshift bins is not significantly affected by the use of the Y5 cut. The qualities of the distribution are consistent with those we find in Y1: the multiple-to-one sample remains skewed to higher redshifts as compared to the one-to-one sample. In addition, the proportion of lost objects significantly increases (as expected, given that incompleteness has a larger effect for Y5).

For the case of photometric redshifts, we use a cata-

log of redshifts estimated via BPZ using the measured photometry from the DC2 coadded images. Using these data, the overabundance at high redshifts of multiple-to-one objects as compared to one-to-one objects is reduced (Figure 12) as compared to the fiducial analysis (Figure 6). The $N(z)$ bias due to photometric redshift calibration with an isolated one-to-one sample is therefore below the $N(z)$ bias due to true (spectroscopic) calibration found in our fiducial analysis, so our main results provide an upper bound for the redshift bias in Y1. It is noteworthy that the one-to-one and multiple-to-one redshift distributions are similar when using photometric redshifts, given that every object in the multiple-to-one distribution has contaminated flux. Our results suggest that either the particular photometric redshift algorithm used here is robust to such contaminants, or that the majority of contaminants are too faint to significantly bias the measured redshift.

One feature of interest in Figure 12 is the periodic offset between the true redshift distribution and the photometric distribution. This offset is a consequence of the photometric redshift algorithm and results in incorrect bin assignment when computing the correlation functions. This offset is consistent between the one-to-one and all observed samples. Figure 13 shows the true redshift distribution of galaxies in each of our five tomographic bins, with bins assigned using photometric redshift. The true redshift distributions of the all observed and one-to-one samples are extremely similar. As shown in Figure 10, incorrect bin assignment does not significantly impact the inferred value of Ω_m for any of our samples.

All the results stated above are qualitatively the same when applying photometric redshifts to the Y5 catalog.

B. VALIDATION USING THE SKYSIM5000 COSMOLOGICAL SIMULATION

The truth catalog used in this analysis (CosmoDC2) is a subset of the SkySim5000 simulated sky catalog, which has a footprint of one eighth of the sky. An illustration of CosmoDC2 and SkySim5000 (as well as the DC2

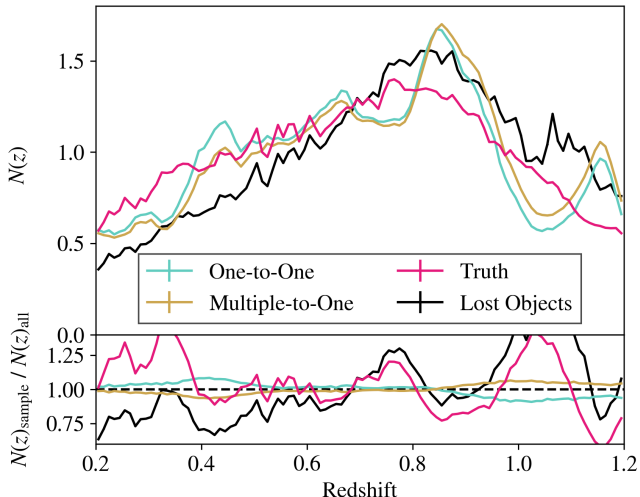


FIG. 12.— *Top*: number of objects in our Y1 samples at a given redshift, normalized such that the integral over all bins is equal to unity. For the observed samples, the redshift is given by the photometric redshift. For the truth and lost samples, the redshift is given by the true redshift. *Bottom*: deviation from the all observed sample.

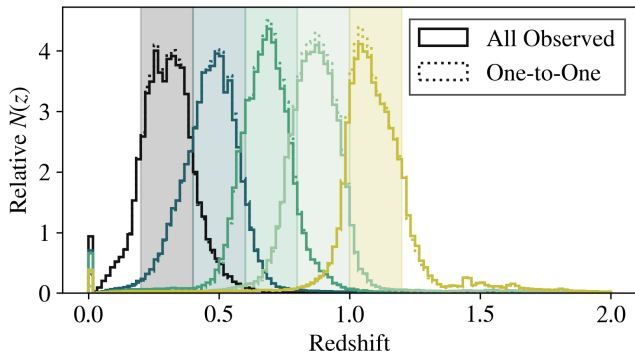


FIG. 13.— True redshift distribution of Y1-selected galaxies in each of our five tomographic bins, with bins assigned using photometric redshift. Each histogram is normalized such that the integral over all its bins is equal to unity. The shaded regions correspond to the selections for each tomographic bin. The overdensities at $z = 0$ correspond to stars that are misidentified in the DC2 catalog and are not removed by our cuts.

observed catalogs) is shown in Figure 14. The two simulations use the same physics and differ only by their sizes, meaning that the measured correlation functions and inferred cosmological parameters should agree up to the level of statistical variance. We validate our measurements by comparing the aforementioned measurements in CosmoDC2 to those of `skysim5000_v1.1.2`.

The angular correlation function for galaxy clustering is shown in Figure 15. The CosmoDC2 measurements are statistically compatible within 1σ of SkySim5000 for all but the second redshift bin, which differs significantly. In order to confirm that the discrepancy is likely due to cosmic variance, we randomly select 12 patches of sky from SkySim5000 with areas matching that of CosmoDC2. The 1σ distribution of correlation functions from these patches is plotted in Figure 15. CosmoDC2 lies within $2 - 3\sigma$ of the patches for the second redshift bin, and well within 1σ for the other bins. It is likely that the deviations in CosmoDC2 are simply an effect of cosmic

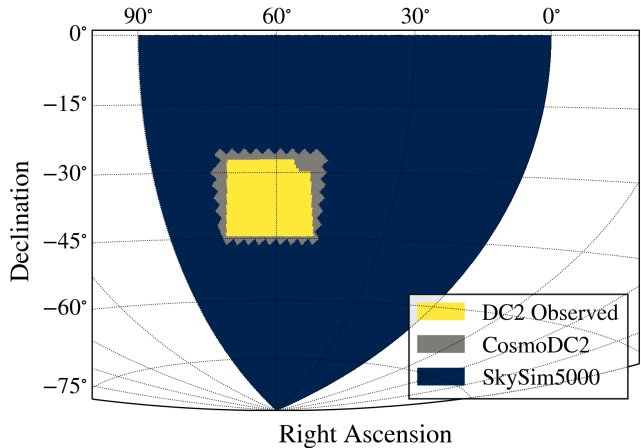


FIG. 14.— Footprints of the observed galaxies from the DC2 catalog (yellow), truth galaxies from the CosmoDC2 simulation (gray), and truth galaxies from SkySim5000 (blue). The footprints are shown using the Hammer-Aitoff projection.

variance, rather than a systematic issue with the catalog.

We repeat the MCMC analysis described in Section 4.3 for SkySim5000 and show the results in Figure 16. The contours for SkySim5000 are more precise than for CosmoDC2 thanks to the greater statistical power, and we recover the simulation input Ω_m within 1σ . The contours agree with those of CosmoDC2 within $2 - 3\sigma$, resembling the level of statistical deviation in the correlation function’s second redshift bin.

This paper was built using the Open Journal of Astrophysics \LaTeX template. The OJA is a journal which provides fast and easy peer review for new papers in the `astro-ph` section of the arXiv, making the reviewing process simpler for authors and referees alike. Learn more at <http://astro.theoj.org>.

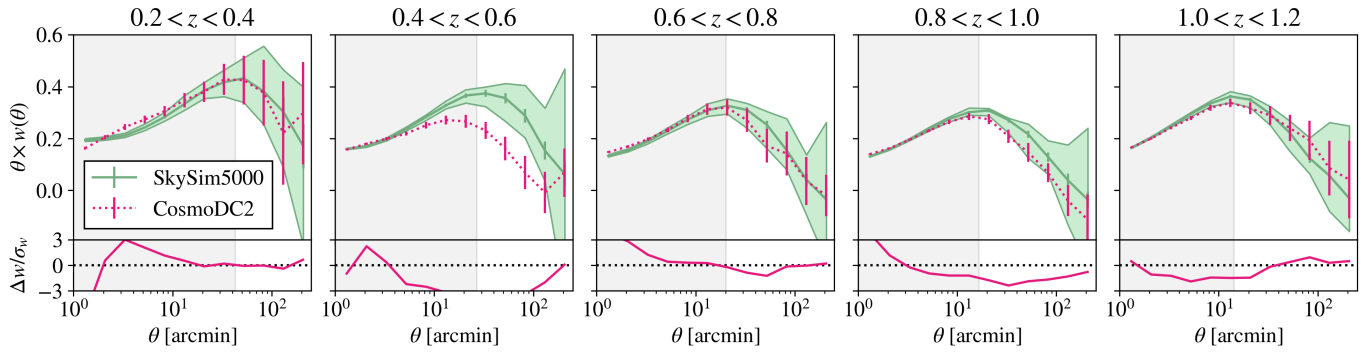


FIG. 15.— *Top*: angular correlation function measurements for CosmoDC2 (magenta) and SkySim5000 (green). The 1σ distribution for 12 randomly-selected DC2-sized patches of SkySim5000 is shaded in light green. *Bottom*: deviation from the full SkySim5000 correlation function. Gray shaded regions denote small scales ignored in cosmological analyses, as specified by the DESC SRD ($k < 0.3h \text{ Mpc}^{-1}$).

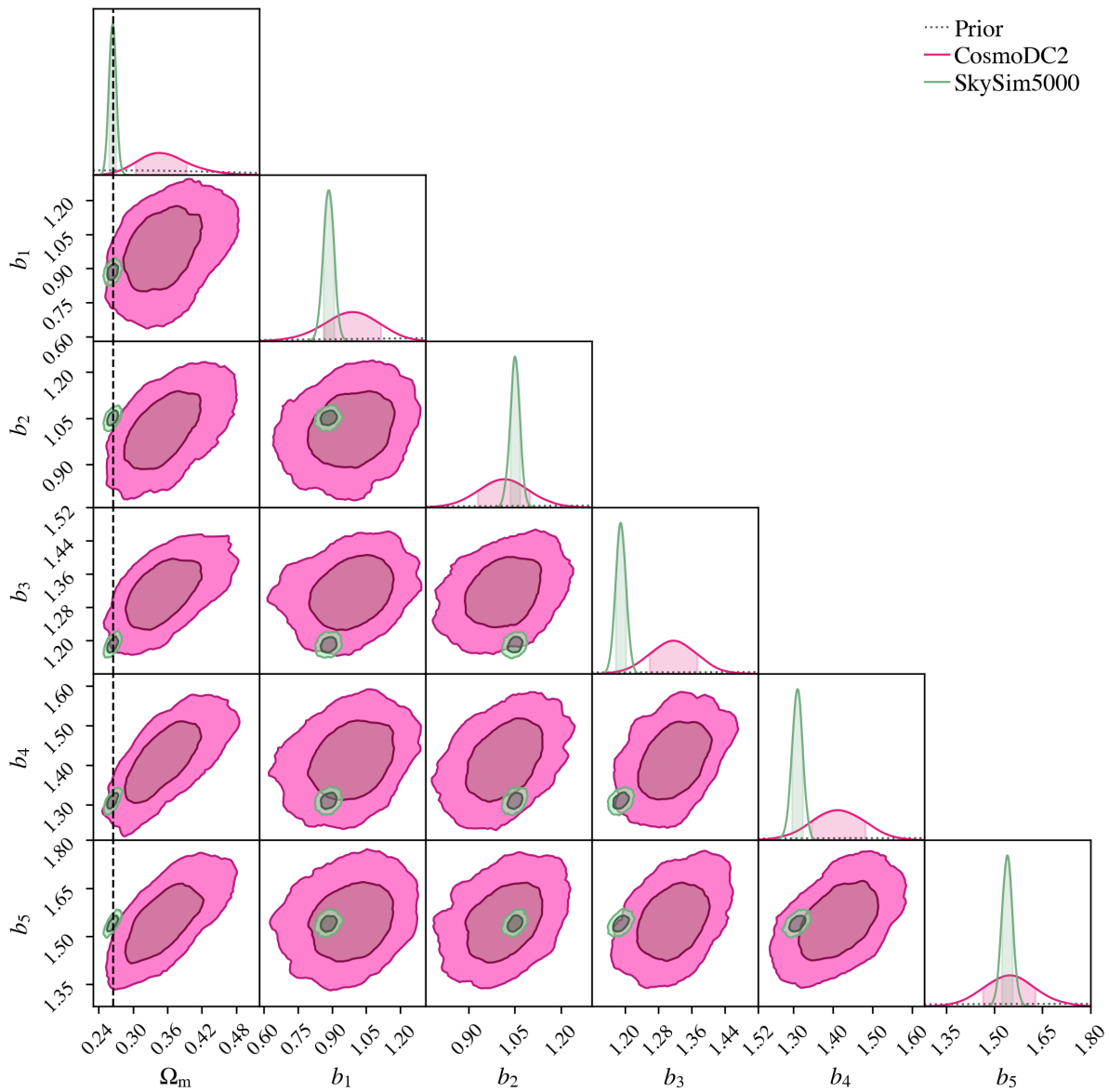


FIG. 16.— Contours for cosmological parameters assuming Λ CDM model with Ω_m and galaxy bias b_i freed, computed for CosmoDC2 (magenta) and SkySim5000 (green). We use DESC SRD priors and show the simulation input value of $\Omega_m = 0.265$ as a vertical dashed line.