Dark Energy Survey Year 3 results: Cosmology from combined galaxy clustering and lensing validation on cosmological simulations

J. DeRose[1,*] R. H. Wechsler,[2,3,4] M. R. Becker,[5] E. S. Rykoff,[3,4] S. Pandey,[6] N. MacCrann,[7] A. Amon,[3] J. Myles,[2,3,4] E. Krause,[8] D. Gruen,[2,3,4] B. Jain,[6] M. A. Troxel,[9] J. Prat,[10,11] A. Alarcon,[5] C. Sánchez,[6] J. Blazek,[12,13] M. Crocce,[14,15] G. Giannini,[16] M. Gatti,[6] G. M. Bernstein,[6] J. Zuntz,[17] S. Dodelson,[18,19] X. Fang,[8] O. Friedrich,[20] L. F. Secco,[6,11] J. Elvin-Poole,[21,22] A. Porredon,[21,22] S. Everett,[23] A. Choi,[21] I. Harrison,[24,25] J. Cordero,[25] M. Rodriguez-Monroy,[26] J. McCullough,[3] R. Cawthon,[27] A. Chen,[28] O. Alves,[28,29,30] F. Andrade-Oliveira,[29,30] K. Bechtol,[27] H. Camacho,[29,30] A. Campos,[18] A. Carnero Rosell,[31,30,32] M. Carrasco Kind,[33,34] H. T. Diehl,[35] A. Drlica-Wagner,[10,35,11] K. Eckert,[6] T. F. Eifler,[8,36] R. A. Gruendl,[33,34] W. G. Hartley,[37] H. Huang,[38] E. M. Huff,[36] N. Kuropatkin,[35] M. Raveri,[6] R. Rosenfeld,[39,30] A. J. Ross,[21] J. Sanchez,[35] I. Sevilla-Noarbe,[26] E. Sheldon,[40] B. Yanny,[35] B. Yin,[18] Y. Zhang,[35] P. Fosalba,[14,15] M. Aguena,[30] S. Allam,[35] J. Annis,[35] S. Avila,[41] D. Bacon,[42] S. Bhargava,[43] D. Brooks,[44] E. Buckley-Geer,[10,35] D. L. Burke,[3,4] J. Carretero,[16] F. J. Castander,[14,15] C. Chang,[10,11] M. Costanzi,[45,46,47] L. N. da Costa,[30,48] M. E. S. Pereira,[28] J. De Vicente,[26] S. Desai,[49] J. P. Dietrich,[50] P. Doel,[44] A. E. Evrard,[51,28] I. Ferrero,[52] A. Ferté,[36] B. Flaugher,[35] J. Frieman,[35,11] J. García-Bellido,[41] E. Gaztanaga,[14,15] T. Giannantonio,[53,20] J. Gschwend,[30,48] G. Gutierrez,[35] S. R. Hinton,[54] D. L. Hollowood,[23] K. Honscheid,[21,22] D. Huterer,[28] D. J. James,[55] K. Kuehn,[56,57] O. Lahav,[44] M. Lima,[58,30] M. A. G. Maia,[30,48] J. L. Marshall,[59] P. Melchior,[60] F. Menanteau,[33,34] R. Miquel,[61,16] J. J. Mohr,[50,62] R. Morgan,[27] A. Palmese,[35,11] F. Paz-Chinchón,[33,53] D. Petravick,[33] A. Pieres,[30,48] A. A. Plazas Malagón,[60] E. Sanchez,[26] V. Scarpine,[35] S. Serrano,[14,15] M. Smith,[63] M. Soares-Santos,[28] E. Suchyta,[64] G. Tarle,[28] D. Thomas,[42] C. To,[2,3,4] and T. N. Varga[62,65]

(DES Collaboration)

[1]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*
[2]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, California 94305, USA*
[3]*Kavli Institute for Particle Astrophysics & Cosmology, P.O. Box 2450, Stanford University, Stanford, California 94305, USA*
[4]*SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*
[5]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, Illinois 60439, USA*
[6]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[7]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom*
[8]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721-0065, USA*
[9]*Department of Physics, Duke University Durham, North Carolina 27708, USA*
[10]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA*
[11]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*
[12]*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*
[13]*Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland*
[14]*Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain*
[15]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain*
[16]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain*
[17]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, United Kingdom*
[18]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*
[19]*NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[20]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*
[21]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[22]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[23]*Santa Cruz Institute for Particle Physics, Santa Cruz, California 95064, USA*

[24]*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom*

[25]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom*

[26]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid 28040, Spain*

[27]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, Wisconsin 53706-1390, USA*

[28]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*

[29]*Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*

[30]*Laboratório Interinstitucional de e-Astronomia—LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, Rio de Janeiro 20921-400, Brazil*

[31]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*

[32]*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain*

[33]*Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark Street, Urbana, Illinois 61801, USA*

[34]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, Illinois 61801, USA*

[35]*Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, Illinois 60510, USA*

[36]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA*

[37]*Department of Astronomy, University of Geneva, Chemin d'Écogia 16, CH-1290 Versoix, Switzerland*

[38]*Department of Physics, University of Arizona, Tucson, Arizona 85721, USA*

[39]*ICTP South American Institute for Fundamental Research Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*

[40]*Brookhaven National Laboratory, Building 510, Upton, New York 11973, USA*

[41]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain*

[42]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, United Kingdom*

[43]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, United Kingdom*

[44]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*

[45]*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*

[46]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*

[47]*Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy*

[48]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, Rio de Janeiro 20921-400, Brazil*

[49]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*

[50]*Faculty of Physics, Ludwig-Maximilians-Universität, Schneiderstrasse 1, 81679 Munich, Germany*

[51]*Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA*

[52]*Institute of Theoretical Astrophysics, University of Oslo. P.O. Box 1029 Blindern, NO-0315 Oslo, Norway*

[53]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

[54]*School of Mathematics and Physics, University of Queensland, Brisbane, Queensland 4072, Australia*

[55]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, Massachusetts 02138, USA*

[56]*Australian Astronomical Optics, Macquarie University, North Ryde, New South Wales 2113, Australia*

[57]*Lowell Observatory, 1400 Mars Hill Road, Flagstaff, Arizona 86001, USA*

[58]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, São Paulo 05314-970, Brazil*

[59]*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA*

[60]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, New Jersey 08544, USA*

[61]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

[62]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany*

[63]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, United Kingdom*

[64]*Computer Science and Mathematics Division, Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37831, USA*
[65]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München,
Schneiderstrasse 1, 81679 München, Germany*

We present a validation of the Dark Energy Survey Year 3 (DES Y3) 3×2-point analysis choices by testing them on BUZZARD2.0, a new suite of cosmological simulations that is tailored for the testing and validation of combined galaxy clustering and weak-lensing analyses. We show that the BUZZARD2.0 simulations accurately reproduce many important aspects of the DES Y3 data, including photometric redshift and magnitude distributions, and the relevant set of two-point clustering and weak-lensing statistics. We then show that our model for the 3×2-point data vector is accurate enough to recover the true cosmology in simulated surveys assuming the true redshift distributions for our source and lens samples, demonstrating robustness to uncertainties in the modeling of the nonlinear matter power spectrum, nonlinear galaxy bias, and higher-order lensing corrections. Additionally, we demonstrate for the first time that our photometric redshift calibration methodology, including information from photometry, spectroscopy, clustering cross-correlations, and galaxy–galaxy lensing ratios, is accurate enough to recover the true cosmology in simulated surveys in the presence of realistic photometric redshift uncertainties.

I. INTRODUCTION

Having effectively exhausted the information encoded in the linear modes probed by the cosmic microwave background, cosmologists have lately turned to wide-field galaxy surveys as their preferred tool for studying dark matter and dark energy. These galaxy surveys, which probe the distribution of matter in the low-redshift universe, encode a great deal of cosmological information, but their complex observational and modeling systematics pose significant challenges to unbiased inference.

Recently, combinations of galaxy clustering and weak lensing have been realized as powerful mechanisms for extracting cosmological information from photometric galaxy surveys such as the Dark Energy Survey[1] (DES), Kilo Degree Survey,[2] and Hyper Suprime Cam.[3] In particular, the combination of shear-shear [cosmic shear, $\xi_{\pm}(\theta)$], galaxy position–galaxy position [galaxy clustering, $w(\theta)$], and tangential shear–galaxy position [galaxy–galaxy lensing, $\gamma_t(\theta)$] two-point functions into a 3×2-point analysis has proven powerful, in part because of its ability to break degeneracies between nuisance parameters and cosmological parameters [1,2]. The resultant increase in constraining power comes at a cost, though: great care must be taken to ensure that all of the components that feed into the 3×2-point analysis are robustly determined, lest the inferred cosmological constraints be biased.

While some analysis validation can be undertaken analytically [3], or using the data itself [4], validation against realistic cosmological simulations is an essential component of modern galaxy survey cosmology analyses [5–8]. In this work, we validate three main components of the 3×2-point analysis being performed on the first three years of DES data (DES Y3): the two-point function measurement pipeline, photometric redshift calibration methodology, and the likelihood and modeling framework used to obtain cosmological constraints from two-point functions and redshift distributions.

The first challenge that must be overcome in 3×2-point analyses is the accurate measurement of the cosmic shear, galaxy–galaxy lensing, and galaxy angular clustering statistics that make up the 3×2-point data vector. The measurement of galaxy ellipticities is especially important for weak-lensing analyses, and dedicated image simulations that test for and constrain biases that appear in this process are an essential ingredient for modern weak-lensing analyses [5,9–13]. In this work, we focus on ensuring that our two-point measurement pipelines deliver unbiased correlation functions in the absence of these shear systematics. Mitigating angular systematics imprinted on galaxy clustering and galaxy–galaxy lensing statistics by survey and foreground inhomogeneities is also important, and while we do imprint DES depth variations on our simulations that lead to systematics of this kind, the DES Y3 methodology for removing such systematics is thoroughly investigated independently in Refs. [14–16].

The next key component required for accurate cosmological inference is a characterization of the source and lens galaxy sample redshift distributions. The robust calibration of source redshift distributions is a challenge shared by all

---

*Corresponding author.
jderose@lbl.gov
[1]https://www.darkenergysurvey.org/.
[2]http://kids.strw.leidenuniv.nl/.
[3]https://hsc.mtk.nao.ac.jp/ssp/.*

weak-lensing surveys [17–19]. A critical component of such a calibration is correctly accounting for incompleteness of the spectroscopic galaxy catalogs that form the foundation of all photometric redshift estimation. Previous versions of the simulations presented in this work have been used to study these effects [20], but here we assume that we have complete, albeit realistically sized, spectroscopic redshift catalogs, upon which we build our photometric redshift estimations. We proceed to test the three separate components of the DES Y3 photometric redshift calibration: photometric redshift estimation using self-organizing maps [self-organizing map photometric redshift (SOMPZ) and three-step dirichlet sampling (3sDir)], clustering redshifts [Clustering redshift (WZ)], and the ratios of galaxy–galaxy lensing signals (shear ratios). These three components are thoroughly validated by Amon *et al.* [21], Gatti *et al.* [22], Myles *et al.* [18], Sánchez *et al.* [23], and Sánchez *et al.* [24]. In this work, we present further tests showing that the combination of these three methods is self-consistent and unbiased in our simulations and for the first time show that the source redshift distributions calibrated using these methods deliver unbiased cosmological constraints in simulations.

The challenge of lens galaxy photometric redshift estimation is also formidable. The lens sample that we use in this work, REDMAGIC, is a sample of luminous red galaxies selected to be constant in comoving number density as a function of redshift. As they are a bright subset of galaxies, photometric redshift estimation is relatively straightforward compared to that for source galaxy, but nonetheless must be thoroughly validated. In the DES Y3 data, clustering redshifts are again used for this task, and this methodology is tested using the simulations presented here by Cawthon *et al.* [25]. The REDMAGIC redshift estimation is further validated in this work, where we show that use of REDMAGIC photometric redshift distributions does not bias our cosmological inference. We note that after unblinding the DES Y3 REDMAGIC analysis, the fiducial DES Y3 analysis was shifted to use a different lens galaxy sample, called the MAGLIM sample [26,27]. Validation of that analysis, including the use of one of the simulations, described in this work is presented by Porredon *et al.* [26].

With robust two-point function and redshift distribution measurements in hand, one remaining challenge is to accurately predict the dependence of the 3×2-point data vector on the cosmological parameters of interest. Weak lensing is particularly sensitive to nonlinearities induced by gravitational collapse during the process of large-scale structure formation, and so validation of the assumed matter power spectrum model is a particularly important aspect of 3×2-point analyses. The N-body simulations used in this work were run with different settings and resolutions than those used in the Halofit model for the matter power spectrum that the DES Y3 3×2-point

analysis employs and thus provide an independent validation of this modeling alongside that presented by Krause *et al.* [28]. The ray-tracing algorithm that we employ [29] allows us to incorporate higher-order lensing effects such as magnification, angular deflection, and reduced shear in the two-point functions that we measure in our simulations, validating the approximations of these effects that are made in our modeling framework that are also validated by Elvin-Poole *et al.* [30], Krause *et al.* [28], and Prat *et al.* [15].

Robust marginalization over astrophysical nuisance parameters such as galaxy bias, redshift-space distortions (RSDs), galaxy intrinsic alignments, and baryonic effects on the matter distribution is another important challenge when interpreting the measurements made in 3×2-point analyses. As the models used for these effects in the DES Y3 3×2-point analysis are largely perturbative in nature, it is essential to validate them against fully nonlinear solutions for these physical processes as implemented in cosmological simulations. In this work, we focus on testing our models for galaxy bias and RSD, while Krause *et al.* [28] and Seccof *et al.* [31] test our models for intrinsic alignments and baryonic effects. These tests complement those performed by Pandey *et al.* [6], further validating the bias modeling choices that are used in the analysis of the DES Y3 data by Pandey *et al.* [32] and the accompanying 2×2-point analyses. In particular, we show here that the scale cuts used for our analyses are robust to nonlinearities beyond those assumed in our model, by showing that the cosmological constraints obtained from the measurements on our simulations are unbiased at high significance.

Our presentation will be organized as follows. In Sec. II, we describe the BUZZARD2.0 simulations, which represent a significant upgrade over the simulations used in DES Year 1 analyses. In Sec. III, we describe how 3×2-point measurements are made and how redshift distributions are estimated in our simulated analyses. In Sec. IV A, we describe the model used to obtain cosmological constraints from the DES Y3 3×2-point measurements that we wish to test on these simulations. In Sec. V, we present the results of our simulated analyses, including varying levels of realism. Finally, in Sec. VI, we summarize our work and conclude by discussing future directions for improvement.

II. BUZZARD2.0 SIMULATIONS

In this work, we make use of a suite of 18 N-body simulations that are designed to reproduce the lens and source samples used in the DES Y3 3×2-point analysis. These are a new version of the BUZZARD simulations [33], implementing a number of improvements over those used in analyses of the first year of DES data (DES Y1). In this section, we briefly summarize the pertinent details from DeRose *et al.* [33] and outline the main improvements, relegating a more detailed description of these changes to Appedixes A–C. The main improvements of the

simulations presented in this work over those used in DES Y1 analyses are as follows:

(1) improved color-dependent clustering, based on a conditional abundance matching model;
(2) explicit matching of red-sequence color distributions to the DES data;
(3) more realistic photometric errors enabled by the BALROG [9] image simulation framework.

Each BUZZARD simulation is constructed from three independent N-body simulations, with sizes of 1.05^3, 2.6^3, and 4.0^3 h^{-3} Gpc^3 and particle loads of 1400^3, 2048^3, and 2048^3, respectively. All simulations were run using the L-GADGET2 code [34] and initialized with independent seeds at $z = 50$ using second-order Lagrangian perturbation theory as implemented in 2LPTIC [35] and linear power spectra computed with CAMB [36], assuming a flat Lambda Cold Dark Matter (LCDM) cosmology with $\Omega_m = 0.286$, $\Omega_b = 0.046$, $h = 0.7$, $n_s = 0.96$, and $\sigma_8 = 0.82$. A single set of these simulations is sufficient to generate a unique light cone with an area of 10,413 square degrees out to $z = 2.35$. For further details regarding simulation specifications and validation of relevant observables, see Secs. 3.1 and 3.2 in DeRose et al. [33].

Galaxies are included in these light cones using the ADDGALS algorithm [37], which imbues each galaxy with a position, velocity, size, ellipticity, and spectral energy distribution. Functionally, the algorithm for producing our source and lens galaxy samples proceeds as follows:

(1) Assign galaxy positions, velocities, and absolute magnitudes.
(2) Assign galaxy Spectral Energy Distributions (SEDs).
(3) Assign intrinsic galaxy sizes and ellipticities.
(4) Perform ray tracing and lensing of the galaxy catalog.
(5) Generate survey-specific photometry and ellipticities.
(6) Select source and lens galaxy samples.

Step 1 is unchanged from previous versions of the BUZZARD simulations and is fully explained by Wechsler et al. [37]. We generate a list of galaxy absolute magnitudes by integrating a luminosity function that has been fit to a variety of spectroscopic datasets. Additionally, we have optimized the luminosity function in order to match observed galaxy counts as a function of magnitude in the DES Y3 data, using an algorithm analogous to that used in DES Y1, described in Appendix E.1 of the work by DeRose et al. [33]. Galaxies from this list are assigned phase-space coordinates using a model for $P(\delta|M_r, z)$ that is tuned to a subhalo abundance matching model, where $P(\delta|M_r, z)$ is the probability that a galaxy in the simulation with an absolute magnitude M_r is found in an overdensity δ. In practice, δ is defined as the inverse of the radius enclosing a 3×10^{13} $h^{-1} M_\odot$. In halo masses around and below this mass, ADDGALS tends to be slightly less accurate as described by Wechsler et al. [37].

One of the main galaxy samples used for DES cosmological constraints is the REDMAGIC sample, which is

designed to have robust photometric redshifts made possible by preferentially selecting bright red galaxies whose redshift distributions have been shown to be accurately and precisely characterized [25,38,39]. To better model this sample, we have made significant improvements to the color-dependent clustering model with respect to previous versions of these simulations. In particular, we adopt a conditional abundance matching algorithm that assigns SEDs from the Sloan Digital Sky Survey (SDSS) Main Galaxy Sample with redder rest frame $g - r$ colors to galaxies that are closer to dark matter halos with masses greater than $M_{200b} \sim 10^{13} h^{-1} M_\odot$. A more detailed description of this algorithm can be found in Appendix A as well as in the work by Wechsler et al. [37] and DeRose et al. [40].

Following this procedure, we compute broadband magnitudes by integrating each SED over DES $ugrizY$ and VISTA JHK bandpasses. This procedure leads to color distributions that well approximate those observed in the DES Y3 data. In particular, color distributions for BUZZARD and the DES Y3 deep field data are compared in Fig. 1, in which we have binned the deep field data in redshift using the most precise redshift estimate available for each galaxy. Agreement is excellent except for in $u - g$, where there are known deficiencies in the SED templates used in our simulations [33].

In DeRose et al. [33], we showed that the galaxy red sequence, e.g., $P(r - i|z)$ for red galaxies, in the DES Y1 BUZZARD simulations was significantly narrower than that found in the DES Y1 data. This led to overoptimistic photometric redshift uncertainties for red-sequence galaxies such as the REDMAGIC sample, as the uncertainty in
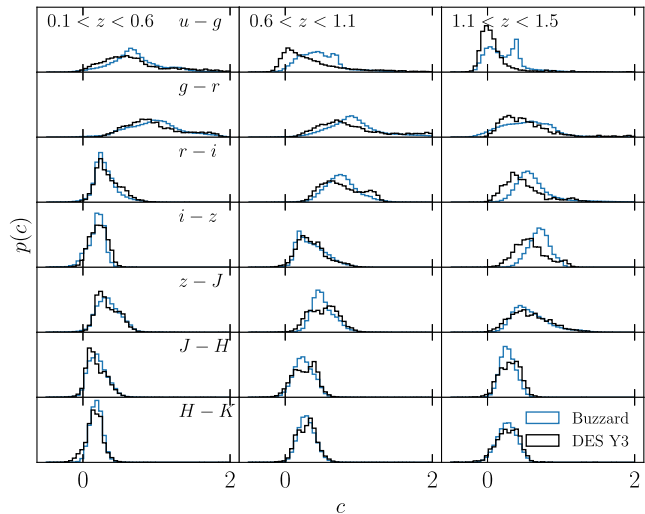

FIG. 1. Comparison of $ugrizJHK$ color distributions as a function of redshift between BUZZARD (blue) and the DES Y3 redshift sample (black). Different rows depict color distributions for different band combinations (listed in the left column), while different columns show different redshift bins. Agreement is good, except for in the u-band and at redshifts $z > 1$, where the SED templates used in BUZZARD are poorly constrained by data [41].
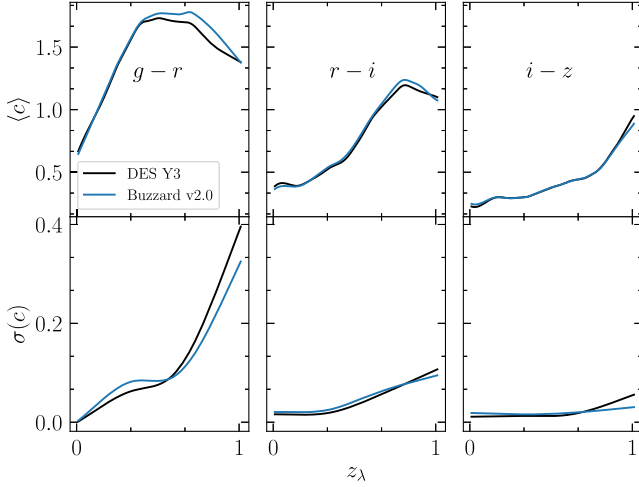
FIG. 2. Comparison of red-sequence colors between BUZZARD and DES Y3 as measured by REDMAPPER. Top: mean red-sequence color as a function of redshift for the DES Y3 data (black) compared to the BUZZARD simulations (blue). The largest differences occur at high redshift in $g - r$, where the mean color is poorly constrained in the data. Bottom: scatter in red-sequence colors as a function of redshift for the DES Y3 data and BUZZARD.

these photometric redshift estimates is directly proportional to the width of the red sequence. Additionally, at high redshifts, galaxies in the DES Y1 BUZZARD simulations were significantly redder than those found in the data, leading to a deficit of bright, red galaxies. To remedy these issues, we explicitly force the mean and width of the red sequence to match that found in the data. The agreement between the BUZZARD2.0 simulations and the DES Y3 data in this respect is shown in Fig. 2. The algorithm for performing this matching is described in Appendix B.

Once SEDs have been assigned and magnitudes generated in each DES bandpass, we assign galaxy half-light radii and intrinsic ellipticities as a function of observed i-band magnitude as described in Appendix E3 in DeRose *et al.* [33]. After this step, we have a catalog of galaxies with true positions, velocities, SEDs, DES magnitudes sizes, and ellipticities.

Before applying any survey-specific masking or photometric errors to our simulations, we compute weak-lensing quantities by ray tracing through our simulations with the CALCLENS code [29] using the same configuration as described in DeRose *et al.* [33]. We then compute deflection, rotation, shear, and magnification for each galaxy from the lensing distortion tensor at the position of that galaxy. These quantities are then used to deflect the angular positions, rotate and shear the intrinsic ellipticities, and magnify the sizes and magnitudes of each galaxy.

After lensing has been performed, the simulations are rotated into the DES Y3 footprint and masked using the DES Y3 REDMAGIC mask. We are able to cut two DES Y3 footprints per 10,000 square degree light cone, each with an area of 4143.17 square degrees.

In addition to these improvements, we have implemented a more realistic model for photometric errors in order to provide a better testing ground for the photometric redshift methodologies employed in the DES Y3 3×2-point analysis. Along with the Gaussian photometric errors that we produce using the model described in Appendix E.4 of DeRose *et al.* [33], we produce independent realizations of photometric errors by making use of the photometric error distributions measured by the BALROG image simulation framework [4,9]. BALROG injects low-noise images of galaxies observed in the DES deep fields [42] into wide-field images and remeasures their photometry. This allows for estimates of detection efficiency and photometric errors as a function of the nearly noiseless deep-field galaxy properties. These relationships are precisely what are required in order to apply photometric errors to our simulations. For each simulated galaxy, we find a BALROG injected galaxy that is closest in riz bands and apply the magnitude offsets between the true BALROG injected photometry and the measured wide-field detection. For a detailed description of this algorithm, see Appendix C. At this point, we have galaxy catalogs with lensed positions, velocities, SEDs, sizes, ellipticities, and $ugrizYJHK$ magnitudes with realistic noise from which we can select DES-like galaxy samples.

III. MEASUREMENT OF THE 3 × 2-POINT DATA VECTOR

The 3×2-point data vector consists of the combination of galaxy ellipticity and galaxy density autocorrelations, the cross-correlation between galaxy density and galaxy ellipticities, and the associated source and lens galaxy redshift distributions. In our simulations, we have kept as close as possible to the measurement pipelines used in the DES Y3 data, with the main exception being that our simulated source catalog does not have inverse variance weights or METACALIBRATION shear responses [4,43]. We have also opted to measure all shear correlation functions without shape noise in order to maximize the constraining power afforded us by this suite of simulations. Appendix D describes how we select source and lens galaxy catalogs from the BUZZARD simulations, and Appendix E describes the estimators used to construct the two-point functions that make up the 3×2-point data vector.

In Figs. 3 and 4, we compare the mean 3×2-point data vector from the BUZZARD simulations to the 3×2-point data vector measured from the DES Y3 data, finding agreement at the 10%–20% level for nearly all scales and redshift bins. The first redshift bin in $w(\theta)$ is the major exception; here, the BUZZARD simulations disagree with the DES Y3 measurements by a factor of approximately 1.5 on large scales. This bin contains the intrinsically faintest lens galaxies in our sample, so it is possible that this is a sign of a breakdown in our color-dependent galaxy clustering model at fainter magnitudes than the samples used to
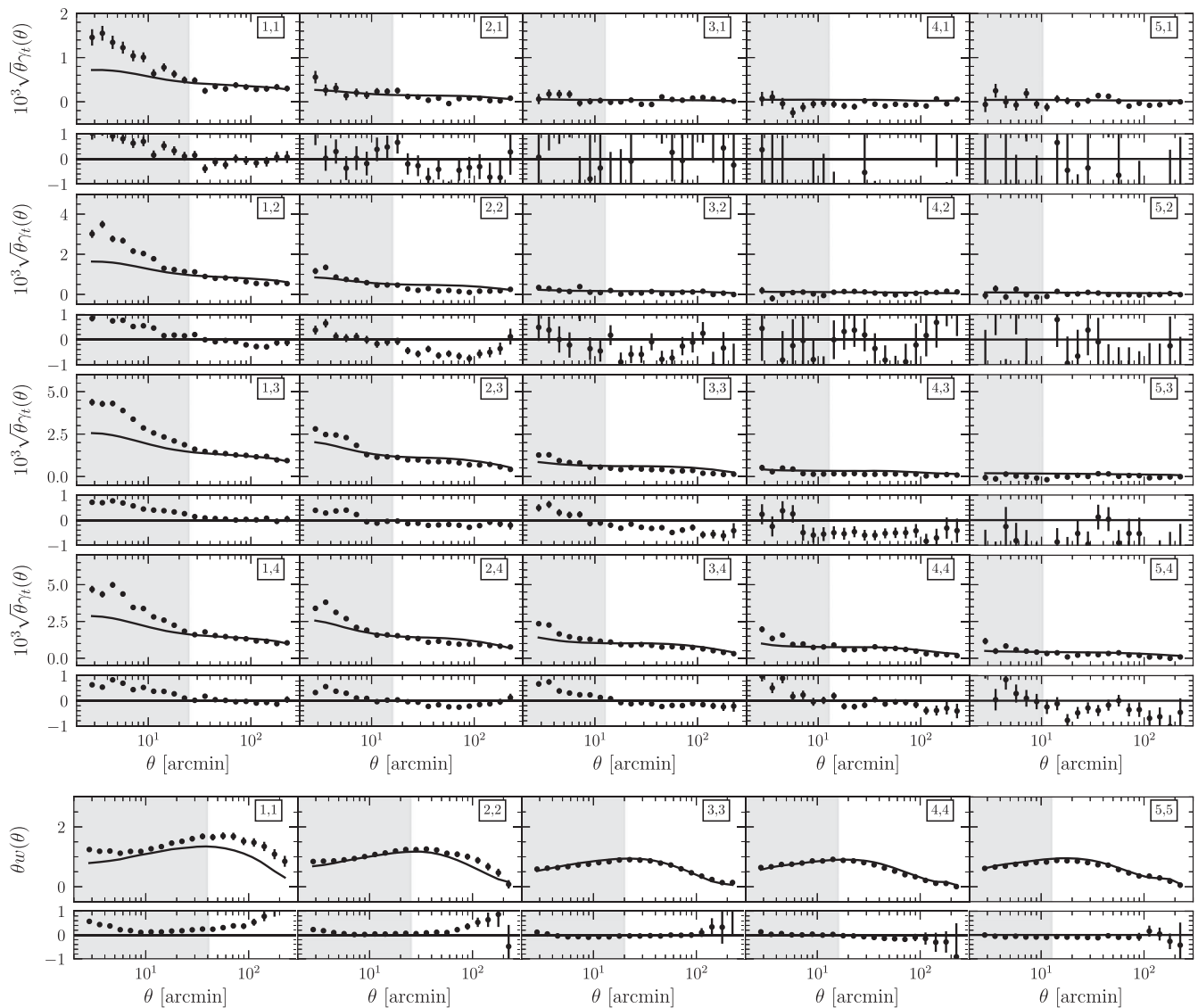
FIG. 3. Comparison of $\gamma_t(\theta)$ and $w(\theta)$ between BUZZARD (lines) and the measurements from the Y3 data using the REDMAGIC lens sample (points), with error bars given by the fiducial Y3 covariance matrix. Shaded regions indicate the angular scale cuts applied in the fiducial 3×2-point analysis. Rows alternate between showing the measured signals and the fractional difference between data and simulations. Numbers in each panel label the bin combinations shown, with the numbers in the $\gamma_t(\theta)$ panels representing the lens-source bin pair and numbers in $w(\theta)$ panels denoting the lens bins alone. The intent of this comparison is to gauge how well the galaxy clustering and lensing properties in BUZZARD match those observed in the data, with the caveat that disagreement may arise due to differences in cosmology, and source and lens galaxy redshift distributions. Agreement in $w(\theta)$ is generally better than for $\gamma_t(\theta)$ especially for the second through fifth lens bins, which may be a result of slightly different source redshift distributions between the simulations and data, especially where there is significant overlap between source and lens redshift distributions. We also observe a marked deficit of power on small scales in $\gamma_t(\theta)$ for the first lens bin at smaller scales than those used in the analysis, part of which is likely due to resolution effects in BUZZARD. The excellent match between $\gamma_t(\theta)$ on larger scales in the first lens bin is interesting in light of the large discrepancy in the amplitude of $w(\theta)$ between the DES Y3 data and BUZZARD for this lens bin. The DES Collaboration [44] demonstrates that under the assumption of LCDM the REDMAGIC $w(\theta)$ and $\gamma_t(\theta)$ imply bias values that differ by 12%, so this may play some role in the discrepancies seen between the BUZZARD and DES Y3 $w(\theta)$ measurements in the first lens bin.

constrain the model parameters in the work by DeRose et al. [40]. It is also interesting to note that the BUZZARD predictions for $\gamma_t(\theta)$ in this lens bin match the data quite well. In light of the observed 12% discrepancy between the bias inferred from $\gamma_t(\theta)$ and $w(\theta)$ in the DES Y3 REDMAGIC

sample [44], the fact that BUZZARD agrees with the DES Y3 REDMAGIC $\gamma_t(\theta)$ measurement in the first lens bin implies that we should see a 24% discrepancy in $w(\theta)$ on large scales under the assumption of LCDM and linear bias, which is approximately what is observed.
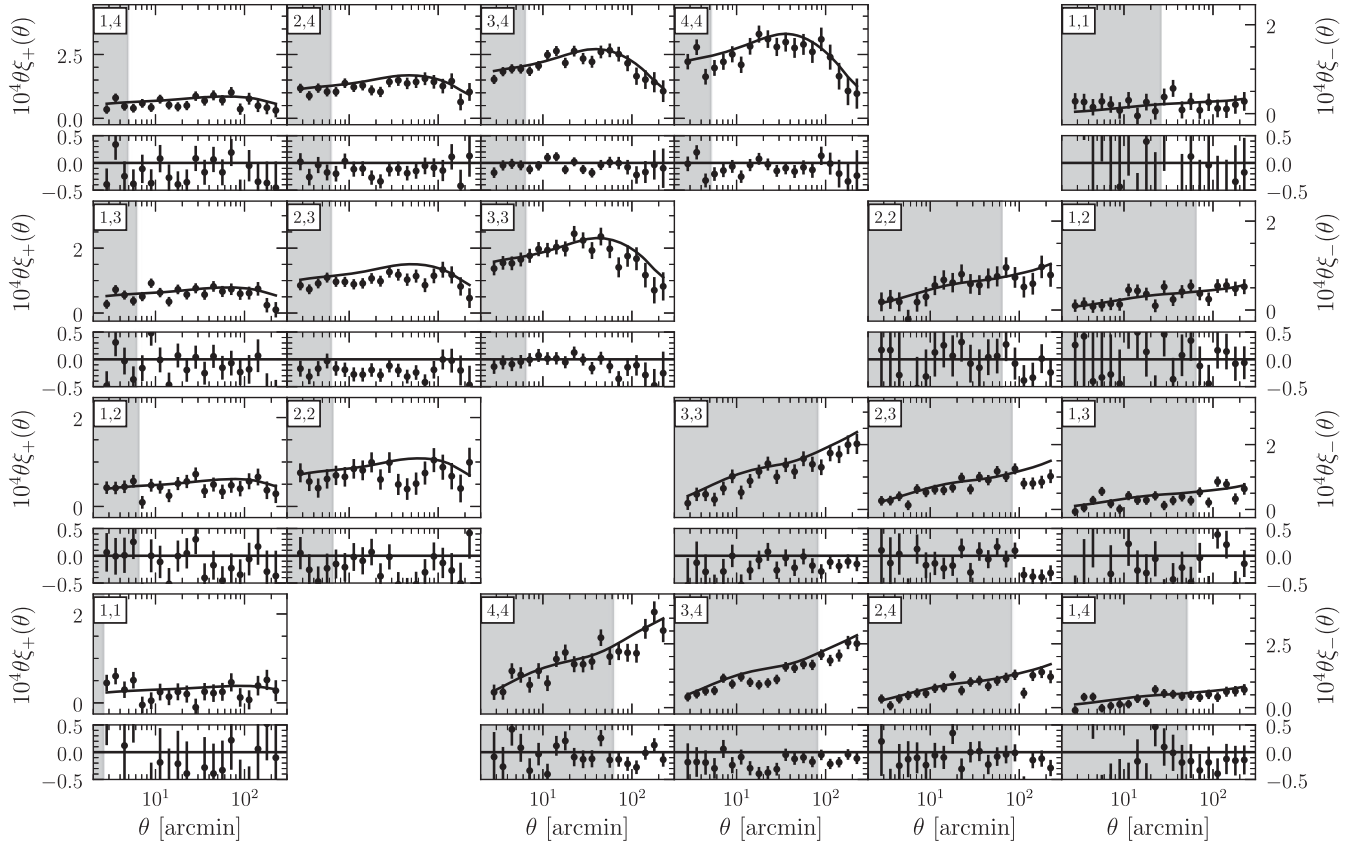
FIG. 4.   Comparison of $\xi_\pm(\theta)$ between BUZZARD and the DES Y3 data. As in Fig. 3, rows alternate between $\xi_\pm(\theta)$ and the fractional difference between simulations and data, while the numbers in each panel denote which source bin pairs are plotted. The intent of this comparison is to show that the cosmic shear measurements from our simulations largely agree with those in the DES Y3 data, so analyses performed on the simulations can be trusted to have similar constraining power to those performed on the DES Y3 data. The residuals shown here are largely scale independent, which is likely a representation of the imperfect match in source redshift distributions between our simulations and data, along with a small difference in the best-fit cosmology from the $\xi_\pm(\theta)$ analysis in the DES Y3 data and that used in the BUZZARD simulations.

We also see that there is a deficit of power in the BUZZARD measurements on small scales in $\gamma_t(\theta)$ for bin combinations that include the first lens bin. This deficit is likely a result of resolution effects in our N-body simulations on one-halo scales, as these measurements probe the smallest physical scales shown in these figures due to the nonlocal nature of $\gamma_t(\theta)$. We note that for $\xi_\pm(\theta)$ there are no large deviations from our simulation measurements on scales smaller than the DES Y3 scale cuts, as one might expect from baryonic effects on the matter power spectrum, which have been shown to affect the matter power spectrum at the 10%–30% level [45].

To confront our two-point function measurements with theoretical predictions, we require estimates of the redshift distributions of our source and lens galaxies for each tomographic bin. We have made validation of these algorithms on the simulations presented here a focus for the DES Y3 3×2-point analysis. Each component of our redshift estimation framework including photometric redshift estimation and tomographic binning with SOMPZ

[18,46], redshift distribution uncertainty propagation with 3SDIR [18,24], and complementary redshift distribution information from clustering cross-correlations [22,25] and galaxy–galaxy lensing ratios [23] has been tested extensively on the BUZZARD simulations. Here, we briefly summarize how redshift distributions are obtained from our simulations, relegating a more extensive discussion to Appendix F and the papers where each individual redshift estimation component are validated and applied to the DES Y3 dataset [10,12,27,47,48].

Our lens galaxy photometric redshift estimation is relatively unchanged from that used in DES Y1. We briefly describe it here and refer the reader to the work by Rozo *et al.* [39], Cawthon *et al.* [25], and Rodríguez-Monroy *et al.* [16] for more details. As the REDMAGIC sample is a set of bright galaxies for which we have abundant spectroscopy, we place significant confidence in REDMAGIC in the photometric redshift estimates provided by the algorithm itself, $p(z_{\mathrm{redMaGiC}})$. These are obtained by constructing a red-sequence spectral template from a combination of

spectroscopy and galaxy cluster members. In our simulations, we assume that we have a sparse, but unbiased, spectroscopic training set of similar size to that used in the data. We bin lens galaxies into five tomographic bins with edges, $\{0.15, 0.35, 0.5, 0.65, 0.8, 0.9\}$, using the mean of $p(z_{\mathrm{redMaGiC}})$. To estimate the $n(z)$ for each tomographic bin, we stack four Monte Carlo samples drawn from the $p(z_{\mathrm{redMaGiC}})$ for each galaxy.

Source galaxy tomographic binning and mean redshift distribution estimation is performed by a self-organizing-map-based algorithm called SOMPZ. SOMPZ leverages information from many-band photometry taken in small deep fields in the DES Y3 footprint in combination with secure spectroscopic and multiband photometric redshifts from PAU + COSMOS [49] to estimate the redshift distributions of wide field weak-lensing source galaxies, accounting for selection and noise biases using the BALROG image simulation framework. All of these pieces of information are included in our BUZZARD simulations, with the simplifying assumption that the redshift catalog

that is used by sompz is sparse but unbiased. We use 3SDIR to generate samples of $n(z)$ for each source tomographic bin, propagating sample variance and shot noise errors. Clustering cross-correlations between lens and source galaxies are then used to select $1000 n(z)$ samples, and the distribution of the means of these samples is used as a prior for the analyses presented in Sec. V D.

The top row of Fig. 5 compares the source and lens redshift distributions measured from a single realization of the BUZZARD simulations to those measured in the DES Y3 data. The agreement between the source redshift distributions in BUZZARD and the DES Y3 data is quite good; the mean redshifts of each tomographic bin in BUZZARD are $[0.326, 0.511, 0.744, 0.871]$ compared with $[0.382, 0.563, 0.759, 0.913]$ in the Y3 data. The high- and low-redshift tails are also captured well in the simulations, except in the last source bin where there is a tail to high redshift in the DES Y3 data that is not as extended in the BUZZARD simulations. This discrepancy is consistent with the observation that colors in BUZZARD approximate DES data more
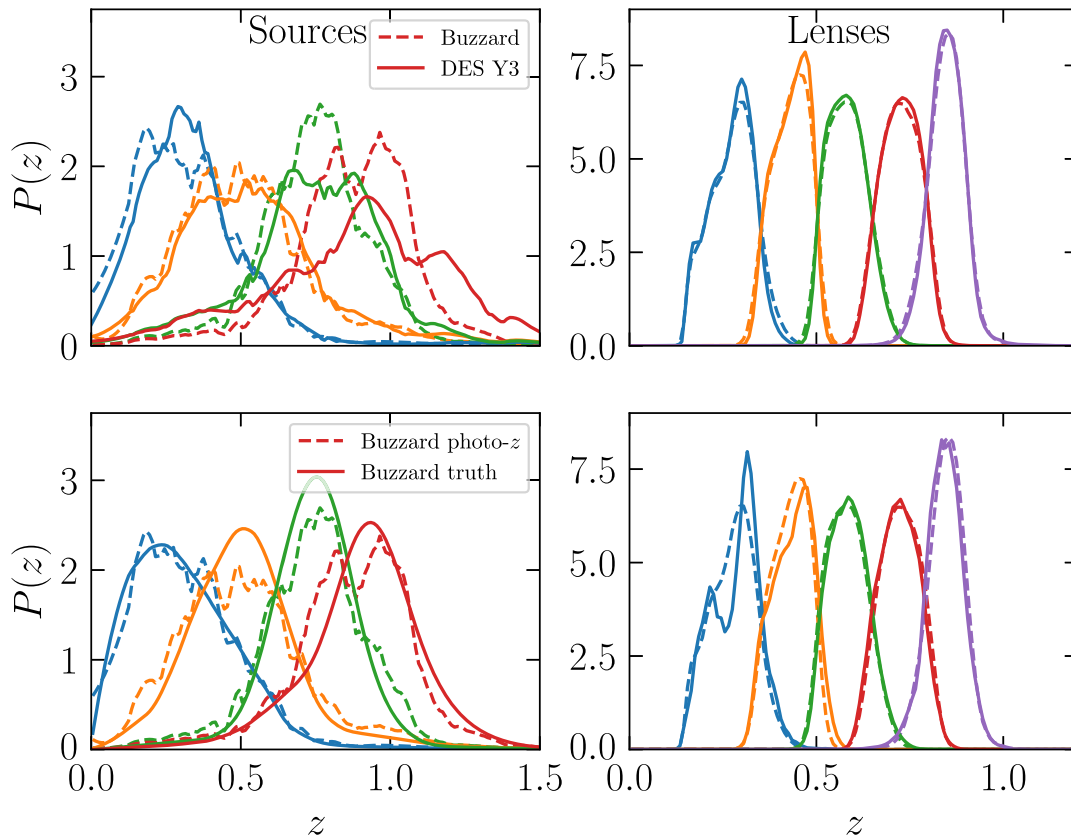

FIG. 5.   Top: comparison of source (left) and lens (right) redshift distributions in BUZZARD (dashed) to DES Y3 data (solid). The mean redshifts of the source tomographic bins are $[0.326, 0.511, 0.744, 0.871]$ in BUZZARD compared with $[0.382, 0.563, 0.759, 0.913]$ in the Y3 data. We have matched the effective source number density and shape noise in BUZZARD to that found in the DES Y3 METACALIBRATION sample [10]. In combination with the close match in mean redshifts of each tomographic bin, this means that the total signal to noise of our lensing measurements in BUZZARD should be approximately the same as that found in the DES Y3 data. Bottom: comparison between true (solid) and photometric (dashed) redshift distributions in BUZZARD for sources (left) and lenses (right). The differences between true and photometric redshift distributions illustrated here are shown to be negligible for the simulated analyses presented in this work in Sec. V D.

poorly at $z > 1$, as seen in Fig. 1. A key component necessary for obtaining this level of agreement in redshift distributions was the implementation of a more realistic photometric error model, described in Appendix C. The agreement between the lens redshift distributions in our simulations and those found in the data is near perfect. Tuning of the red-sequence color model using the algorithm described in Sec. II and Appendix B was necessary in order to obtain this level of agreement.

IV. 3 × 2-POINT LIKELIHOOD

In this section, we describe the theoretical model that that we employ in the simulated analysis of the BUZZARD2.0 3 × 2-point measurements.

A. Theoretical model

Here, we provide an overview of the model that we will use to describe the 3 × 2-point measurements and refer the reader to the work of Amon *et al.* [21], Elvin-Poole *et al.* [30], Krause *et al.* [28], Pandey *et al.* [32], Prat *et al.* [15], and Secco *et al.* [31] for complete technical specifications of the model.

1. Field-level description

There are two main fields that we must accurately describe in order to model 3 × 2-point measurements: the scalar projected galaxy overdensity field measured in a tomographic bin i at position $\hat{\mathbf{n}}$, $\delta_{\mathrm{obs}}^i(\hat{\mathbf{n}})$, and the spin-2 galaxy shape field in tomographic bin j at position $\hat{\mathbf{n}}$, $\gamma_\alpha^j(\hat{\mathbf{n}})$.

We account for three contributions to the galaxy overdensity field,

$$\delta_{\mathrm{obs}}^i(\hat{\mathbf{n}}) = \underbrace{\int d\chi\, W_\delta^i(\chi)\delta_{\mathrm{g}}^{(3D)}(\hat{\mathbf{n}}\chi,\chi)}_{\delta_{\mathrm{D}}^i(\hat{\mathbf{n}})} + \delta_{\mathrm{RSD}}^i(\hat{\mathbf{n}}) + \delta_\mu^i(\hat{\mathbf{n}}), \quad (1)$$

where χ is the comoving distance and $W_\delta^i = n^i(z)dz/d\chi$ is the normalized radial window function for galaxies in tomographic bin i. The first term in this sum, $\delta_{\mathrm{D}}^i(\hat{\mathbf{n}})$, is the projection of the three-dimensional (3D) density contrast, $\delta_{\mathrm{g}}^{(3D)}$, and the following two terms are contributions from RSD and magnification, respectively.

The configuration-space three-dimensional density contrast can be perturbatively expanded in terms of operators that satisfy the symmetries of general relativity. We account for the following dependencies in this work,

$$\delta_{\mathrm{g}}^{3D} \sim f(\delta_{\mathrm{m}}, \nabla_i\nabla_j\Phi, \nabla_i v_j)$$
$$\sim f^{(1)}(\delta_{\mathrm{m}}) + f^{(2)}(\delta_{\mathrm{m}}^2, s^2) + f^{(3)}(\delta_{\mathrm{m}}^3, \delta_{\mathrm{m}}s^2, \psi, st), \quad (2)$$

where δ_m, Φ. and v are the three-dimensional matter density contrast, potential, and velocity fields and ψ, s^2

and t are the scalar quantities constructed from contractions of the shear and velocity divergences $\nabla_i\nabla_j\Phi$ and $\nabla_i v_j$ [50]. On the right-hand side of Eq. (2), we have organized these fields by the order at which they contribute to the galaxy density contrast, where all arguments of the function f^i contribute at ith order. If we remain agnostic about the absolute magnitude of the contribution from these fields to the total galaxy density contrast and treat them to one-loop order in perturbation theory, then we must introduce four bias coefficients per lens bin: b_1, b_2, b_{3nl}, and b_{s^2} [50,51]. The parameters b_1 and b_2 describe the dependence of the galaxy density contrast on the matter density contrast and its square, b_{3nl} governs the dependence of all third-order fields that contribute at one loop, and b_{s^2} dictates the dependence on s^2. Pandey *et al.* [6] showed that for the accuracy of DES Y3 it is sufficient to leave only b_1 and b_2 entirely free and to relate b_{3nl} and b_{s^2} to b_1 through the so-called Lagrangian coevolution relations [52]:

$$b_{s^2} = (-4/7) \times (b_1 - 1) \quad (3)$$

$$b_{3nl} = b_1 - 1. \quad (4)$$

Additionally, we multiply b_1 by the fully nonlinear matter density contrast, whose power spectrum is modeled using Halofit [53,54]. In Sec. V, we explore a fiducial model that assumes only linear bias, as well as a higher-order bias model that leaves b_1^i and b_2^i free per lens bin, where the superscript enumerates the lens redshift bin that each coefficient contributes to.

RSD contributes to this projected field through the apparent bulk motion of galaxies across redshift bins due to the large-scale coherent infall of galaxies towards each other [55]. As our redshift bins are broad, this effect can be treated at linear order and thus depends only on the linear matter density field and the Hubble parameter, $H(z)$. As such, inclusion of RSD contributes no additional free parameters to our model. During the preparation of this work, a bug in the implementation of RSD in the BUZZARD simulations was identified. As our analysis was quite advanced, we opted to correct for this by adjusting our model to account for it, rather than correcting our simulations. Additional details are contained in Appendix H.

The magnification contribution to δ_{obs}^i is sourced by gravitational lensing of galaxies by matter along the line of sight and contributes to the density both in a purely geometric manner by increasing/decreasing the apparent surface area around dense/underdense lines of sight and by modulating the galaxy selection function through the (de)magnification of galaxy magnitudes and sizes around (under)dense lines of sight. The first of these effects depends only on the underlying cosmological model and matter power spectrum, but the second depends on the number density of the galaxy sample in question as a function of flux and size. To account for this dependence,

we introduce the proportionality constant C^i and write the magnification term as

$$\delta_\mu^i(\hat{\mathbf{n}}) = C^i \kappa^i(\hat{\mathbf{n}}), \qquad (5)$$

where $C^i = 2[\alpha^i(m) - 1]$ and α is the slope of the intrinsic magnitude–size distribution, $n^{\text{intr}}(m)$:

$$\alpha = 2.5 \frac{d}{dm}[\log n^{\text{intr}}(m)]. \qquad (6)$$

We have also introduced the tomographic convergence field

$$\kappa^i(\hat{\mathbf{n}}) = \int d\chi \, W_\kappa^i(\chi) \delta_{\mathrm{m}}(\hat{\mathbf{n}}\chi, \chi), \qquad (7)$$

where δ_{m} is the 3D matter density contrast, and the tomographic lensing efficiency

$$W_\kappa^i(\chi) = \frac{3\Omega_m H_0^2}{2} \int_\chi^{\chi_H} d\chi' n^i(\chi') \frac{\chi}{a(\chi)} \frac{\chi' - \chi}{\chi'}. \qquad (8)$$

It is important to note that we assume constant galaxy bias, magnification, and enclosed mass (see Appendix G) parameters in each lens bin. The validity of this assumption is tested in our simulated analyses.

We can write the galaxy shape field as

$$\gamma_\alpha^j(\hat{\mathbf{n}}) = \gamma_{\mathrm{G},\alpha}^j(\hat{\mathbf{n}}) + \epsilon_{\mathrm{I},\alpha}^j(\hat{\mathbf{n}}) + \epsilon_{0,\alpha}^j(\hat{\mathbf{n}}), \qquad (9)$$

where $\gamma_{\mathrm{G},\alpha}^j$ is the contribution sourced by the gravitational field, i.e., the gravitational lensing signal, and the latter two terms are sourced by the intrinsic shapes of galaxies, where we have decomposed this contribution into a spatially coherent alignment with the large-scale matter distribution $\epsilon_{\mathrm{I},\alpha}^j(\hat{\mathbf{n}})$ and a stochastic contribution $\epsilon_{0,\alpha}^j(\hat{\mathbf{n}})$ which will only contribute to the noise in our correlation function measurements.

To model the spatially coherent "intrinsic alignment" (IA) term, $\epsilon_{\mathrm{I},\alpha}^j(\hat{\mathbf{n}})$, we first construct a three-dimensional shape field $\bar{\gamma}_{ij}^{\mathrm{IA}}$ using the tidal alignment and tidal torquing (TATT) model as [31]

$$\bar{\gamma}_{ij}^{\mathrm{IA}} = A_1 s_{ij} + A_{1\delta} \delta s_{ij} + A_2 s_{ik} s_{kj} + \cdots, \qquad (10)$$

where s_{ij} is the traceless tidal field tensor. This three-dimensional field must then be projected to obtain the two-dimensional field described in Eq. (9), which is then used to compute angular correlation functions.

We let the coefficients in Eq. (10) evolve as a power law in redshift:

$$A_1(z) = -a_1 \bar{C}_1 \frac{\rho_{\text{crit}}\Omega_m}{D(z)} \left(\frac{1+z}{1+z_0}\right)^{\eta_1}, \qquad (11)$$

$$A_2(z) = 5a_2 \bar{C}_1 \frac{\rho_{\text{crit}}\Omega_m}{D^2(z)} \left(\frac{1+z}{1+z_0}\right)^{\eta_2}; \qquad (12)$$

$\bar{C}_1$ is a normalization constant, by convention fixed at a value $\bar{C}_1 = 5 \times 10^{-14} \, M_\odot \, h^{-2} \, \text{Mpc}^2$, obtained from SuperCOSMOS (see the work by Brown et al. [56]). The parameter z_0 is a pivot redshift, which we fix to the value 0.62; ρ_{crit} is the critical density; and $D(z)$ is the linear growth function.

We then relate $A_{1\delta}$ to A_1 via

$$A_{1\delta} = b_{\text{TA}} A_1, \qquad (13)$$

where b_{TA} can be interpreted as the source galaxy bias, although we have allowed it to assume a broader prior than would be physically possible if it were a galaxy bias parameter in order to give this term more flexibility in our intrinsic alignment (IA) model. In total, we have five IA parameters: three amplitude parameters in a_1, a_2, b_{TA} and two parameters governing their redshift evolution, α_1 and α_2. This three-dimensional field can be projected into the two-dimensional shape field that is observed.

Note that we include modeling of IAs in our simulated analyses solely for sake of comparison with the expected constraining power of the analyses being performed on the data. No IA signal is included in the BUZZARD simulations. Marginalizing over the full IA model significantly impacts the constraining power of cosmic shear analyses, but its impact on 3×2-point is much less severe [31,44]. Nevertheless, we have verified that none of the results presented here changes in a qualitative way when we do not marginalize over IAs.

2. Angular two-point statistics

We are not interested in these fields themselves but rather in their angular two-point auto- and cross-correlations. In general, an angular two-point function $\xi_{AB}^{ij}(\theta)$, where i and j are tomographic bin indices and A and B specify the fields being correlated, can be related to its corresponding angular power spectrum $C_{AB}^{ij}(\ell)$ via the relation

$$\xi_{AB}^{ij}(\theta) = \sum_\ell \frac{2\ell + 1}{4\pi} C_{AB}^{ij}(\ell) d_{mn}(\theta), \qquad (14)$$

where d_{mn} is the Wigner D matrix and $m = n = 0$ for $w(\theta)$; $m = 0$, $n = 2$ for γ_t; and $m = 2$, $n = \pm 2$ for $\xi_\pm$.

We compute bin-averaged predictions for these two-point functions by computing $d_{mn}(\theta)$ averaged over the width of the angular bin as described by Krause et al. [28]. The angular power spectra are in turn computed by projecting three-dimensional power spectra P_{AB} along the line of sight weighted by the relevant projection kernels. For $\xi_\pm$ and γ_t, we use the Limber approximation [57]

$$C_{AB}^{ij}(\ell) = \int d\chi \frac{W_A^i(\chi)W_B^j(\chi)}{\chi^2} P_{AB}\left(k = \frac{\ell + 0.5}{\chi}, z(\chi)\right),$$
(15)

but for $w(\theta)$, the accuracy of the DES Y3 data demands that we compute the full non-Limber projection integral on large scales as described by Fang *et al.* [47] and Krause *et al.* [28]. Our simulations do not assume the Limber approximation for $w(\theta)$, but our ray-tracing algorithm does implicitly assume it, and thus the $\gamma_t(\theta)$ and $\xi_\pm(\theta)$ measurements in our simulations also use the Limber approximation. Additionally, wherever the nonlinear matter power spectrum appears in our analysis, we use the recalibrated Halofit model [53], as described by Takahashi *et al.* [54].

We also wish to mitigate the nonlocal nature of the galaxy–galaxy lensing signal, as $\gamma_t(\theta)$ at a fixed value of θ is sensitive to the total mass enclosed at all angles less than θ. There are a number of similar methods that aim to remove the sensitivity of $\gamma_t(\theta)$ to this enclosed mass [5,58,59], but we opt to analytically marginalize over a single parameter in each lens bin in order to account for this,

$$\gamma_t^{ij}(\theta) = \gamma_{t,\text{model}}^{ij}(\theta) + \frac{C_{ij}}{\theta^2},$$
(16)

where

$$C_{ij} = B_i \int dz_l dz_s n^i(z_l) n^j(z_s) \Sigma_{\text{crit}}^{-1}(z_l, z_s) D_A^{-2}(z_l)$$
(17)

$$\equiv B_i \beta_{ij},$$
(18)

where B_i is a free parameter describing the mass enclosed within the minimum scale used in the $\gamma_t(\theta)$ part of our data vector per lens bin, Σ_{crit} is the critical surface mass density, and $D_A(z_l)$ is the angular diameter distance to z_l. Notably, we have made the assumption that this enclosed mass does not evolve significantly over the width of our lens redshift bin, an assumption that is implicitly tested in this work.

Instead of sampling over these additional enclosed mass parameters, analytically marginalize over them by adding an extra component to our covariance matrix as described in Appendix G.

B. Scale cuts

We have determined a set of scale cuts such that our fiducial linear bias modeling choices detailed above are sufficient to deliver unbiased cosmological constraints in a series of simulated analyses on noiseless data vectors. When making choices about whether a particular analysis assumption is acceptable for the constraining power afforded us by the DES Y3 data, we set a criterion that the assumption under consideration must result in no greater than a 0.3σ bias in the two-dimensional spaces

of $S_8 - \Omega_m$ and $w - \Omega_m$ for LCDM and wCDM analyses, respectively. Formally, our criterion can be expressed as

$$P(\hat{S}_8, \hat{\Omega}_m | \mathbf{d}_{\text{fid}}) > 0.235,$$
(19)

where $(\hat{S}_8, \hat{\Omega}_m) = \text{E}[P(S_8, \Omega_m | \mathbf{d}_{\text{cont}})]$ for the LCDM case and likewise for $w - \Omega_m$ for the wCDM case. Procedurally, this test is performed as follows. We generate a data vector, $\mathbf{d}_{\text{cont}}$, by breaking a set of assumptions that are made in our fiducial analysis. We then analyze $\mathbf{d}_{\text{cont}}$ using the fiducial model where those assumptions hold, producing the posterior $P(S_8, \Omega_m | \mathbf{d}_{\text{cont}})$. We then generate a data vector using COSMOSIS, $\mathbf{d}_{\text{fid}}$, using the same cosmology as that assumed when generating $\mathbf{d}_{\text{cont}}$, but now making the assumptions that were previously broken. We analyze this with the same model, giving us the posterior $P(S_8, \Omega_m | \mathbf{d}_{\text{fid}})$. Our scale cut criterion, Eq. (19), then requires that $\text{E}[P(S_8, \Omega_m | \mathbf{d}_{\text{cont}})]$ must fall within the 0.3σ ($P > 0.235$) confidence region of $P(S_8, \Omega_m | \mathbf{d}_{\text{fid}})$.

The contaminated data vector that we use, $\mathbf{d}_{\text{cont}}$, breaks our matter power spectrum and linear bias assumptions. Instead of our fiducial assumptions, we use a model for the nonlinear matter power spectrum that takes into account baryonic effects as measured in the Overwhelmingly Large Simulations (OWLS) Active Galactic Nuclei (AGN) simulation [60]. OWLs AGN possesses feedback effects that are more significant than many more recent hydrodynamic simulations but are still within the realm of possibility. Additionally, we contaminate with the nonlinear bias model described in Sec IV A and bias coefficients as described by Pandey *et al.* [32].

We have determined that scale cuts on $\xi_\pm(\theta)$ that yield a χ^2 difference between our contaminated and fiducial models for $\xi_\pm(\theta)$ of 0.5 are sufficient to pass Eq. (19). For $\gamma_t(\theta)$ and $w(\theta)$, we make angular scale cuts that correspond to 6 h^{-1} Mpc and 8 h^{-1} Mpc [or 4 h^{-1} Mpc for both $\gamma_t(\theta)$ and $w(\theta)$ when testing nonlinear bias modeling], respectively, at the low edge of the redshift range for each lens bin. This procedure, along with a number of additional stress tests of our fiducial model, including tests of different matter power spectra, higher-order lensing effects, and a more complex IA redshift scaling, are discussed further by Krause *et al.* [28].

V. VALIDATION OF THE DES Y3 3 × 2-POINT ANALYSIS

We now proceed to investigate whether the models described in Sec. IV are sufficient to recover unbiased constraints on cosmological parameters in our simulations. We investigate four different analysis configurations:

(A) fixed cosmology;
(B) linear bias, true redshift distributions;
(C) nonlinear bias, true redshift distributions;
(D) linear bias, calibrated photo-z distributions.

The results of these tests are described in the following subsections. The parameters that we leave free for each of these configurations are listed with their priors in Table I. We also list the part of the data vector that each parameter contributes to, i.e., $\xi_{\pm}(\theta)$, $w(\theta)$, or $\gamma_t(\theta)$.

The simulations presented in this work contain complexities beyond the assumptions made in our fiducial model, including

(1) beyond one-loop galaxy bias, including stochasticity;
(2) redshift evolution of galaxy bias within each lens bin;
(3) redshift evolution of enclosed mass parameter, B_i, within each lens bin;

(4) source galaxy clustering;
(5) reduced shear;
(6) source galaxy magnification;
(7) multiple lens plane deflection ("beyond Born" approximation);
(8) anisotropic source and lens $n(z)$;
(9) non-Gaussian distributed data vectors.

For all posterior parameter distributions presented in this paper, we make use of the POLYCHORD nested sampler [61] with the same settings used by the DES Collaboration [44], which have been shown to yield converged posteriors and evidences. All fits are done on data vectors with the fiducial scale cuts described in Sec. IV B unless explicitly stated otherwise. We make

TABLE I. Parameters and priors.

Parameter	Prior	Data vector	Analysis configuration
		Cosmology	
Ω_m	Flat $(0.1, 0.9)$	$\xi_{\pm}(\theta), w(\theta), \gamma_t(\theta)$	B, C, D
A_s	Flat $(5 \times 10^{-10}, 5 \times 10^{-9})$	$\xi_{\pm}(\theta), w(\theta), \gamma_t(\theta)$	B, C, D
n_s	Flat $(0.87, 1.07)$	$\xi_{\pm}(\theta), w(\theta), \gamma_t(\theta)$	B, C, D
Ω_b	Flat $(0.03, 0.07)$	$\xi_{\pm}(\theta), w(\theta), \gamma_t(\theta)$	B, C, D
h	Flat $(0.55, 0.91)$	$\xi_{\pm}(\theta), w(\theta), \gamma_t(\theta)$	B, C, D
		Lens galaxy bias	
b_1^i $(i = 1, 5)$[a]	Flat $(0, 3.0)$	$w(\theta), \gamma_t(\theta)$	A1/2, B, C, D
b_2^i $(i = 1, 5)$[b]	Flat $(-5, 5)$	$w(\theta), \gamma_t(\theta)$	A2, C
		Intrinsic alignment	
a_1	Flat $(-5, 5)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	B, C, D
a_2	Flat $(-5, 5)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	B, C, D
α_1	Flat $(-5, 5)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	B, C, D
α_2	Flat $(-5, 5)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	B, C, D
b_{ta}	Flat $(0, 2)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	B, C, D
		Magnification	
α_{mag}^i	Flat $(-4, 4)$	$\gamma_t(\theta), w(\theta)$	A1/2
		Point mass	
B^i	Flat $(-50, 50)$	$\gamma_t(\theta)$	A1/2[c]
		Lens photo-z	
Δz_l^1	Gauss $(0.000, 0.004)$	$w(\theta), \gamma_t(\theta)$	D
Δz_l^2	Gauss $(0.000, 0.003)$	$w(\theta), \gamma_t(\theta)$	D
Δz_l^3	Gauss $(0.000, 0.003)$	$w(\theta), \gamma_t(\theta)$	D
Δz_l^4	Gauss $(0.000, 0.005)$	$w(\theta), \gamma_t(\theta)$	D
Δz_l^5	Gauss $(0.000, 0.01)$	$w(\theta), \gamma_t(\theta)$	D
σz_l^5	Gauss $(1.000, 0.054)$	$w(\theta), \gamma_t(\theta)$	D
		Source photo-z	
Δz_s^1	Gauss $(0.000, 0.018)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	D
Δz_s^2	Gauss $(0.000, 0.013)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	D
Δz_s^3	Gauss $(0.000, 0.006)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	D
Δz_s^4	Gauss $(0.000, 0.013)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	D
		Shear calibration	
m^i $(i = 1, 4)$	Gauss $(0.0, 0.015)$	$\xi_{\pm}(\theta), \gamma_t(\theta)$	D

[a]Analysis setup C samples over $b_1^i \sigma_8$.
[b]Analysis setup C samples over $b_2^i \sigma_8^2$.
[c]Marginalized over analytically in B/C/D.

use of the COSMOSIS[4] likelihood and sampling framework [62], which is one of two essentially interchangeable implementations of the model described in Scc. IV A, along with COSMOLIKE [63]. COSMOSIS makes use of the CAMB Boltzmann solver [36,64].

We make use of the covariance matrix appropriate for a single survey realization even though we are fitting to the mean of 18 simulations. This is because we wish to keep as close to the actual analysis that will be performed on the DES Y3 data as possible. Analyzing each of the 18 simulations independently with this covariance matrix and taking the product of their posteriors would have yielded the closest approximation to the analyses performed on the DES Y3 data, but doing so would have proven extremely computationally expensive. Instead, we have opted to run analyses on the mean data vector from all 18 simulations but use a covariance matrix appropriate for a single realization. This is in the spirit of the rest of the simulated analyses used to validate the DES Y3 cosmological constraints.

Because of this choice, we are susceptible to the same parameter projection effects in marginalized posteriors as those discussed by Krause *et al.* [28] and Appendix I. These parameter projection effects are a consequence of correlations between cosmological parameters and poorly constrained nuisance parameters, such that when the nuisance parameters are marginalized over they impart a bias in the marginalized posterior mean of a parameter of interest. We emphasize that these effects are not a systematic error, in the sense that they decrease in size as the constraining power of an analysis increases. For a concrete example of parameter projection effects, see the discussion in Sec. V B and Appendix I.

As such, we always compare posteriors obtained from our BUZZARD simulated data vectors with posteriors obtained by running the same analysis on a synthetic data vector generated by COSMOSIS assuming the BUZZARD cosmology, true redshift distributions, and best-fit nuisance parameters from analysis configuration A1 or A2, as defined in Sec. V A, depending on whether the analysis under consideration uses a linear or nonlinear bias model. We will often refer to the COSMOSIS simulated data vector as an *uncontaminated* data vector, as it is produced from the same model that is being used to perform the analysis. When comparing the posteriors obtained from the BUZZARD simulated data vector to those obtained from the uncontaminated data vector, we can disentangle the parameter biases resulting from parameter projection effects that will be present in *both* marginalized posteriors from parameter biases resulting from unmodeled systematics in BUZZARD which will only be present in the BUZZARD posteriors.

To illustrate the increased constraining power available to us through the use of the mean measurement from 18 simulations, we show $1/\sqrt{18}\sigma$ and $2/\sqrt{18}\sigma$ confidence intervals when plotting constraints run on the BUZZARD data vectors. These roughly correspond to the 1σ and 2σ confidence intervals that we would have obtained had we performed analyses using the single Y3 covariance scaled by the inverse of the number of simulations. In fact, these sets of confidence intervals are identical in the limit of flat priors, which we use for the majority of our parameters. If our inference framework worked perfectly, we would expect the mean of the posteriors derived from 18 simulations to match the mean of the posteriors from the uncontaminated analysis at a level defined by these tighter confidence regions, even though the uncertainties from a Y3-sized data set are a factor of $\sqrt{18}$ larger.

Because we have a finite number of simulations, we need to apply a slightly different criterion than Eq. (19) in order to validate our analysis assumptions. In particular, considering the data vector generated by taking the mean over all of our simulations as $\mathbf{d}_{\mathrm{cont}}$ in Eq. (19), we note that the quantity $E[P(S_8, \Omega_m | \mathbf{d}_{\mathrm{cont}})]$ now has an uncertainty associated with it due to the noise that remains in $\mathbf{d}_{\mathrm{cont}}$ even after averaging over 18 simulations. We wish to take this uncertainty into account when determining whether an analysis passes or fails our criterion. To do this, we can generalize Eq. (19) to

$$
\begin{aligned}
\mathrm{PTE} &\equiv 1 - \int_{P(S_8,\Omega_m|\mathbf{d}_{\mathrm{fid}})>0.235} P(S_8, \Omega_m | \mathbf{d}_{\mathrm{cont}}; \Sigma/N) d\theta \\
&\approx 1 - \int_{P(S_8,\Omega_m|\mathbf{d}_{\mathrm{fid}})>0.235} P(S_8, \Omega_m | \mathbf{d}_{\mathrm{cont}}; \Sigma)^N d\theta, \quad (20)
\end{aligned}
$$

where Σ is the covariance for a single Y3 BUZZARD realization and N is the number of simulation realizations (18). The proportionality in the second line holds in the limit that the prior in the two-dimensional space of $S_8 - \Omega_m$ is flat over the region where $P(S_8, \Omega_m | \mathbf{d}_{\mathrm{fid}}) > 0.235$, and $\mathbf{d}_{\mathrm{fid}}$ is the uncontaminated data vector generated by COSMOSIS. This quantity is the probability that the analysis performed on the mean buzzard data vector, $\mathbf{d}_{\mathrm{cont}}$, results in a parameter bias that is more than 0.3σ, i.e., a probability to exceed (PTE) 0.3σ cosmological parameter bias.

Values for these probabilities are quoted in Table II, and these constitute the main result of this paper. In particular, we find that in all cases the probability for any analysis to exceed a $0.3\sigma(1\sigma)$ bias is less than 62% (2%) for LCDM and 58% (1%) in wCDM. In the following subsections, we break these results down by analysis configuration.

We supplement these values by calculating the mean two-dimensional offsets in $\Omega_m - S_8$ and $S_8 - w$ in Table III for LCDM and wCDM, respectively. These offsets are given by

---

[4]https://bitbucket.org/joezuntz/cosmosis/.

TABLE II.   PTE $0.3/1\sigma$ cosmological parameter bias. PTE values less than 1% are indicated as <0.01.

	LCDM B	wCDM B	LCDM C	wCDM C	LCDM D	wCDM D
$\xi_\pm(\theta)$	$0.43/<0.01$	$0.19/<0.01$	...	...	$0.14/<0.01$	$0.24/<0.01$
$w(\theta), \gamma_t(\theta)$	$0.25/<0.01$	$0.35/<0.01$	$0.20/<0.01$	$0.26/<0.01$	$0.08/<0.01$	$0.05/<0.01$
$\xi_\pm(\theta), w(\theta), \gamma_t(\theta)$	$0.61/0.02$	$0.49/<0.01$	$0.35/<0.01$	$0.58/0.01$	$0.15/<0.01$	$0.12/<0.01$

TABLE III.   Mean $\Omega_m - S_8 / w - S_8$ parameter bias.

	LCDM B	wCDM B	LCDM C	wCDM C	LCDM D	wCDM D
$\xi_\pm(\theta)$	0.19σ	0.04σ	...	...	0.07σ	0.15σ
$w(\theta), \gamma_t(\theta)$	0.13σ	0.14σ	0.05σ	0.11σ	0.01σ	0.01σ
$\xi_\pm(\theta), w(\theta), \gamma_t(\theta)$	0.23σ	0.18σ	0.09σ	0.21σ	0.07σ	0.05σ

$$\frac{|\hat{\mathbf{X}}_{\rm fid} - \hat{\mathbf{X}}_{\rm cont}|}{\sigma_{X,\rm fid}},$$

where

$$\hat{\mathbf{X}}_{\rm fid} = \mathrm{argmax}[P(\mathbf{X}|\mathbf{d}_{\rm fid})]$$
$$\hat{\mathbf{X}}_{\rm cont} = \mathrm{argmax}[P(\mathbf{X}|\mathbf{d}_{\rm cont})];$$

$\mathbf{X}$ is either (Ω_m, w) or (S_8, Ω_m), and $\sigma_{X,\rm fid}$ is the 1σ confidence interval in the direction of $\hat{\mathbf{X}}_{\rm fid} - \hat{\mathbf{X}}_{\rm cont}$.

These mean biases are more easily interpretable, but we caution that these are less robust than the PTE values, as the mean posterior values for the BUZZARD chains have non-negligible uncertainty of approximately $1/\sqrt{18}\sigma$ due to residual noise in the mean BUZZARD data vector. That uncertainty is an upper bound on the actual uncertainty on these numbers, since our measurements do not include shape noise, while our covariance matrix does.

A. Fixed cosmology

Before we obtain cosmological constraints from our simulated measurements, we first wish to demonstrate that our models for galaxy bias and lens magnification provide good fits to our measurements when the cosmology in our model is fixed to the true cosmology of our simulations. Additionally, we fix all photometric redshift and shear calibration parameters and use the measured true source and lens redshift distributions to make our model predictions. We set all IA parameters to zero as our simulations contain no IA contribution. We refer to this as analysis configuration A.

Within this configuration, we consider two variants: one using a linear bias model and one using our full nonlinear bias model described in Sec. IV A, which we refer to as configurations A1 and A2, respectively. A1 assumes one linear bias coefficient b_1^i and one magnification coefficient C^i for each lens bin. Analysis configuration A2 also fits for one second-order bias parameter per lens bin, b_2^i. In this

configuration, we also sample over the enclosed-mass parameters, B^i, described in Sec. IV A, rather than analytically marginalizing over them in order to aid in presentation of our results. Analytic marginalization results in the same best-fit likelihood values as sampling over the enclosed-mass parameters, but interpreting model residuals in a visual format is made much easier if we can obtain best-fit values for these parameters. The priors that we assume on our bias, magnification, and enclosed-mass parameters are listed in Table I.

The best-fit models for configurations A1 and A2 are shown in Figs. 6 and 7, and the error bars are obtained from the covariance of a single Y3 simulation, which is what is used to fit these models. The prediction for $\xi_\pm(\theta)$ has no free parameters in this configuration, and we find a chi-squared of 1.4 for 207 data points. The differences on large scales in $\xi_\pm(\theta)$ are likely caused by a combination of source galaxy clustering and source galaxy magnification [28,31]. On scales below the scale cuts used in this analysis, the observed differences are likely sourced by ray-tracing resolution effects [33]. For analysis configurations A1 and A2, the predictions for $w(\theta)$ and $\gamma_t(\theta)$ have 15 and 20 free parameters, respectively, with 53 and 232 data points for $w(\theta)$ and $\gamma_t(\theta)$ for A1 and 68 and 272 data points for $w(\theta)$ and $\gamma_t(\theta)$ for the $r_{\rm min} = 4~h^{-1}$ Mpc scale cuts used for the A2 analysis. A1 results in a chi-squared of 4.5 and 9.1 for $w(\theta)$ and $\gamma_t(\theta)$, while A2 results in chi-squared values for $w(\theta)$ and $\gamma_t(\theta)$ of 7.2 and 8.4. The fact that the reduced chi-squared values for $w(\theta)$ and $\gamma_t(\theta)$ are larger than for $\xi_\pm(\theta)$ is expected, as $w(\theta)$ and $\gamma_t(\theta)$ have an additional contribution from shot noise in the lens galaxy sample, which is not present in $\xi_\pm(\theta)$. In all cases, the residuals are significantly smaller than the expected errors on our Y3 measurements. In the following sections, we investigate the effect of these residuals on cosmological parameter constraints, where we have formal requirements on acceptable parameter biases. We have also checked that the values for the magnification coefficients obtained in these analyses are consistent with the expected values, as
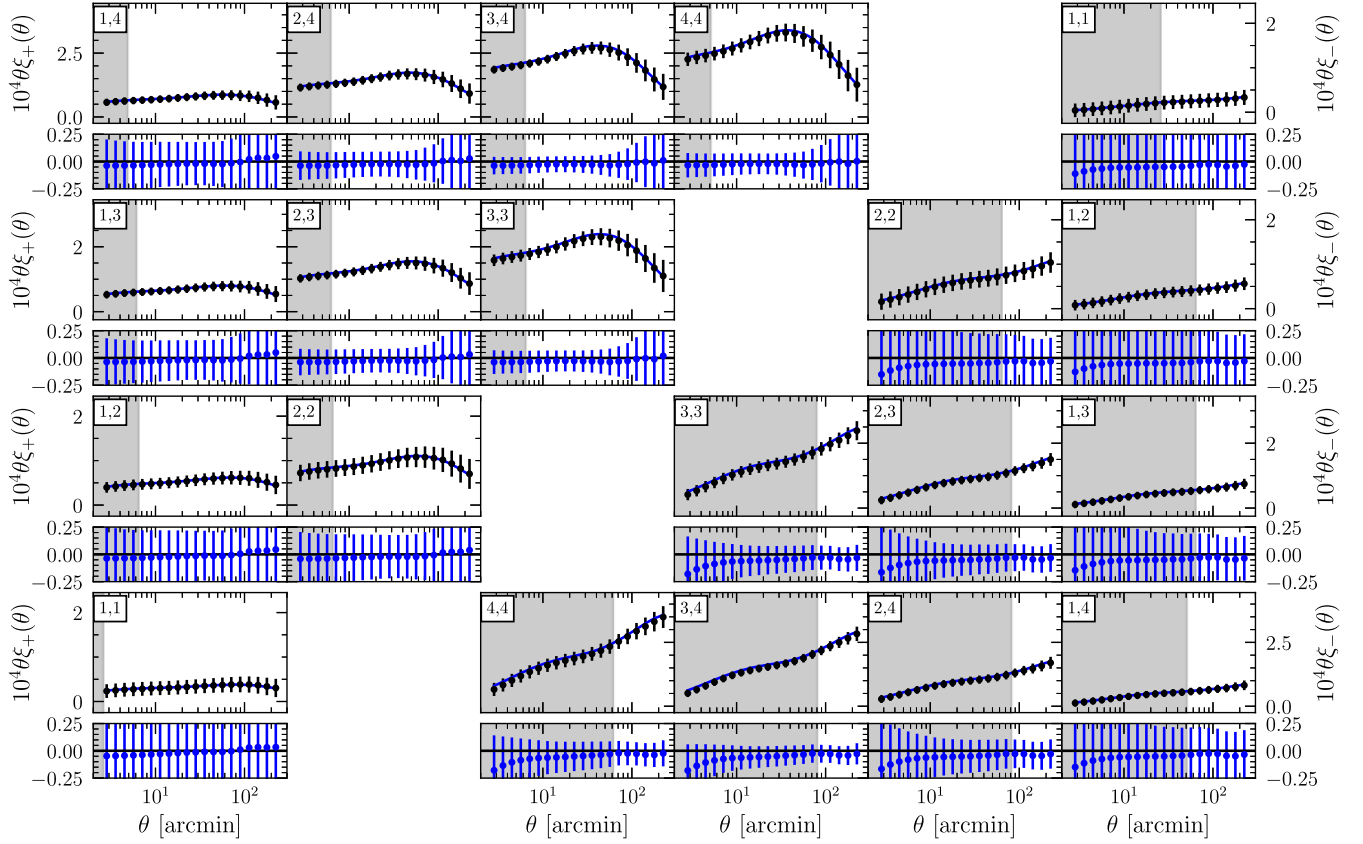
FIG. 6.   Comparison of our fiducial $\xi_{\pm}(\theta)$ model prediction (lines) at the true BUZZARD cosmology to the mean measurement from our simulations without shape noise (points). Rows alternate between the signals themselves and fractional deviations between models and simulations, and the gray regions are our fiducial scale cuts. We find $\chi^2 = 0.78$ for ξ_+ and $\chi^2 = 0.59$ for ξ_-. The differences on large scales are likely caused by a combination of source galaxy clustering and source galaxy magnification [28,31]. On scales below the scale cuts used in this analysis, the observed differences are likely sourced by ray-tracing resolution effects [33].

further explored by Elvin-Poole *et al.* [30]. These results are summarized in Table IV.

B. Linear bias

We proceed to analyze our simulations assuming the true measured redshift distributions with analysis configuration B. Configuration B samples over LCDM or *w*CDM cosmological parameters, linear galaxy bias and the TATT IA model while fixing nuisance parameters related to photometric redshift uncertainty and weak-lensing shear calibration. We also fix the parameters describing lens magnification to the values obtained for the BUZZARD simulations by Elvin-Poole *et al.* [30].

The results of this analysis are summarized in the top row of panels in Figs. 8 and 9, in which we display posteriors obtained from analyses of $\xi_{\pm}(\theta)$, $w(\theta) + \gamma_t(\theta)$, and $\xi_{\pm}(\theta) + w(\theta) + \gamma_t(\theta)$ in the three different subfigures for a LCDM and *w*CDM analysis, respectively. The solid contours in these figures are obtained by running analyses on $\mathbf{d}_{\text{cont}}$, i.e., the mean BUZZARD data vector, and the inner and outer contours represent the $1/\sqrt{18}\sigma$ and $2/\sqrt{18}\sigma$

confidence regions using a covariance matrix appropriate for a single simulation realization. We emphasize that these confidence regions are equivalent to the 1σ and 2σ confidence regions we would have obtained using a covariance appropriate for the mean of 18 simulations in the limit of flat priors.

The dashed contours in each panel are constraints obtained by running on an uncontaminated data vector assuming the covariance of a single Y3 simulation and so approximately represent the uncertainties on S_8, Ω_m, and w that we would expect from the Y3 data in the absence of photometric redshift and shear calibration systematics. The PTE values obtained by plugging these posteriors into Eq. (20) are summarized in the first two columns of Table II. All three analyses pass the criterion in Eq. (20). We also quote biases in the mean $\Omega_m - S_8$ and $w - \Omega_m$ posteriors for the LCDM and *w*CDM analyses with respect to the black uncontaminated posteriors in Table III. We find that all data combinations result in mean posterior biases that are less than 0.3σ.

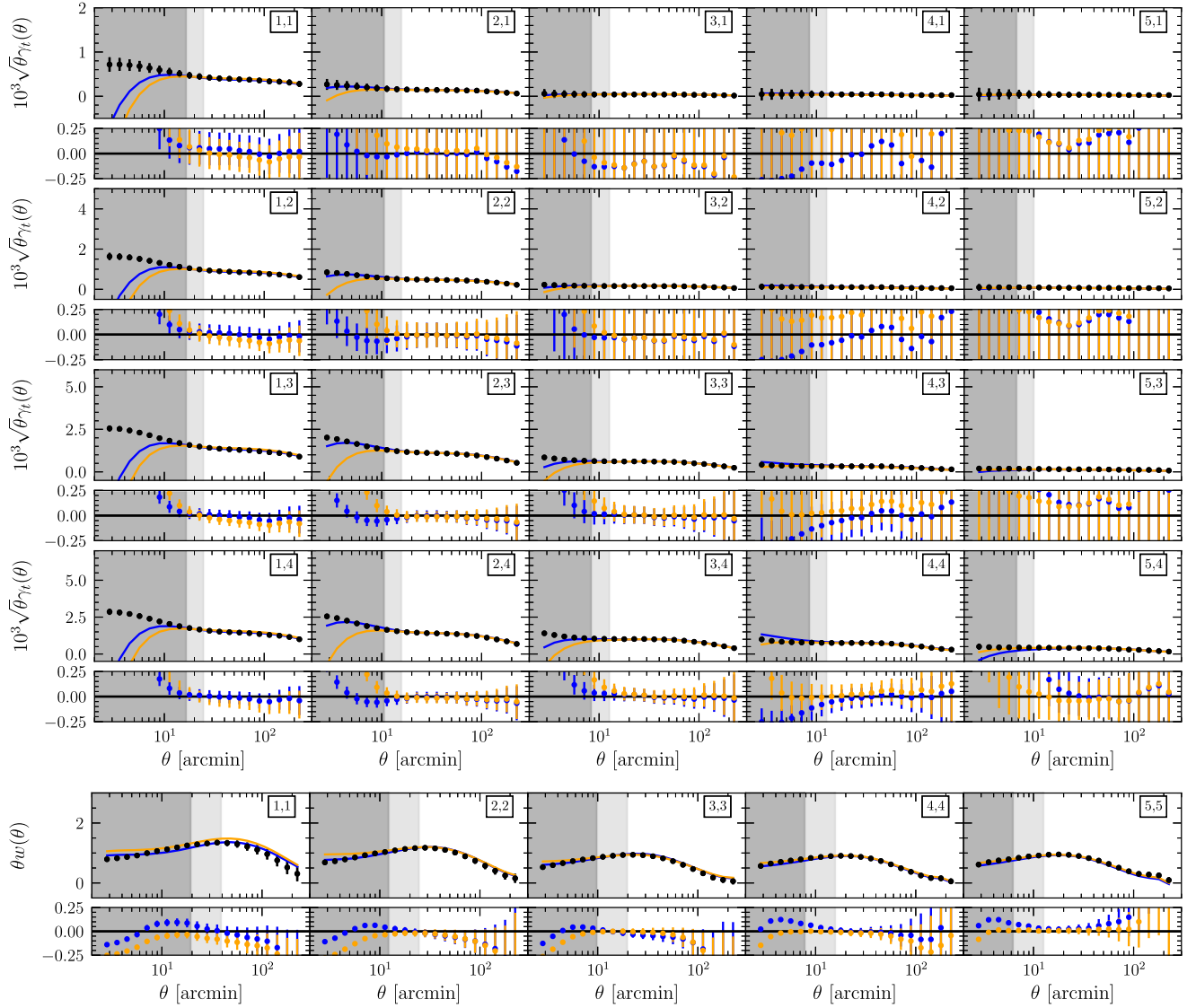In *w*CDM, we see that the uncontaminated black-dashed posteriors are biased with respect to the truth for

FIG. 7. Same as Fig. 6, but for $\gamma_t(\theta)$ and $w(\theta)$. Best-fit configuration A1 (linear bias, blue) and A2 (nonlinear bias, yellow) models fixed to the true BUZZARD cosmology are compared to the mean BUZZARD data vector without shape noise. Rows alternate between the signals themselves and fractional deviations between models and simulations. For configuration A1, we find $\chi^2 = 4.5$ for $\gamma_t(\theta)$ and $\chi^2 = 9.1$ for $w(\theta)$ for our fiducial scale cuts, shown as light gray shaded regions. For A2, we find $\chi^2 = 7.2$ for $\gamma_t(\theta)$ and $\chi^2 = 8.4$ for $w(\theta)$ using $r_{\min} = 4\ h^{-1}$ Mpc scale cuts, depicted by the dark gray shaded regions.

2×2-point and 3×2-point analyses. This is a so-called projection effect that results from degeneracies between our nuisance parameters and cosmological parameters, particularly w and S_8. In configuration B, poorly constrained

TABLE IV. Best-fit configuration $A1/2$ χ^2/data vector dimensionality ($N_{\rm DV}$)/number of free parameters ($N_{\rm param}$).

	A1 χ^2	A1 $N_{\rm DV}$	A1 $N_{\rm param}$	A2 χ^2	A2 $N_{\rm DV}$	A2 $N_{\rm param}$
$\xi_\pm(\theta)$	1.4	207	0	...	...	...
$w(\theta)$	4.5	53	15	7.2	68	20
$\gamma_t(\theta)$	9.1	232	15	8.4	272	20

nonlinear IA parameters in our TATT model couple with S_8 and w, such that when we marginalize over these IA parameters differences in posterior volume sourced by correlations between IA parameters and S_8 and w lead to apparent biases. It is for this reason that it is important to quote offsets with respect to the uncontaminated posteriors, rather than the true parameter values in our simulations, lest we neglect the impact of these projection effects. The direction and size of these projection effects depends sensitively on the mean nuisance and cosmological parameter values assumed when generating our data vectors, as they are entirely a function of the parameter volume that is projected over to obtain marginalized posteriors. As such,

it is also important that our data vector that is used to obtain uncontaminated posteriors is generated with nuisance parameters that match those found in the BUZZARD simulations. For a concrete demonstration of these effects in the case where they are most significant in our analyses, see Appendix I.

If our forward model exactly matched the BUZZARD measurements at the true BUZZARD cosmology, then the projection effects in the contaminated and uncontaminated posteriors would match, as the posteriors would then agree perfectly and thus projection effects would be identical. Because we do not have an analytic forward model that exactly describes the BUZZARD measurements, this sets a limit on the precision with which we can perform the tests presented in this work because differences in projection effects in combination with biases in our model can then source marginalized posterior offsets. This is particularly true in the limit where parameter projection effects are large, as the PTE values and parameter offsets that we quote depend on the nuisance parameter values assumed in our uncontaminated data vectors.

While these effects are counterintuitive and limit our ability to interpret posteriors at the precision of fractions of 1σ, we can take solace in the fact that they are not a systematic bias, as the size of these effects will shrink proportionally to the constraining power of the data.

C. Nonlinear bias

In configuration C, we add nonlinear bias to the model used in configuration B. In particular, we use the nonlinear bias model described in Sec. IV A and sample over $\sigma_8 b_1^i$ and $\sigma_8^2 b_2^i$ rather than the bare bias parameters in order to mitigate parameter projection effects as discussed in Ref. [32]. We again assume the true measured redshift distributions and fix nuisance parameters related to photometric redshift uncertainty and weak-lensing shear calibration. In this case, we relax our angular scale cuts so that they correspond to $r_{\min} = 4h^{-1}$ Mpc at the low edge of each lens redshift bin for $w(\theta)$ and $\gamma_t(\theta)$, as motivated by Pandey *et al.* [6].

The results of this analysis are shown in the middle rows of Figs. 8 and 9 for LCDM and wCDM, respectively. The posteriors here are analogous to those produced using analysis configuration B. The PTE values obtained by plugging these posteriors into Eq. (20) are summarized in the third and fourth columns of Table II. Biases in the mean $\Omega_m - S_8$ and $w - \Omega_m$ posteriors for the LCDM and wCDM analyses with respect to the black uncontaminated posteriors are quoted in Table III. There is a moderate decrease in $\Omega_m - S_8$ parameter bias for the full 3×2-point analysis in this configuration compared with the constraints obtained using configuration B, but the shift that we observe is approximately the same size as that expected from the tests performed by Pandey *et al.* [32], $\sim 0.2\sigma$, that were

originally used to determine the scale cuts that we are further testing here.

In wCDM, we see that the PTE values and $w - \Omega_m$ parameter biases increase with respect to those found in configuration B. This is likely a result of a mismatch in projection effects occurring in our BUZZARD analysis and our uncontaminated analysis. The first piece of evidence for this is that we see the opposite effect, i.e., a decrease in PTE and parameter biases in our configuration C LCDM analyses where projection effects are less important. Furthermore, we see no evidence that our nonlinear bias model is a worse fit for the relaxed scale cuts used in analysis C than our linear bias model is when using our fiducial scale cuts. This is evidenced by the very small change in reduced chi-squared values between configurations A1 and A2, as described in Sec. V A. Finally, the maximum likelihood w value for the configuration C BUZZARD 3×2-point analysis is -1, indicating that there are indeed large projection effects in this analysis, as this maximum likelihood value is significantly offset from the mean of the w posterior in the middle right panel of Fig. 9. In light of this, we caution against overinterpretation of the larger PTE and parameter bias values for these wCDM constraints. This issue further emphasizes the importance of quoting parameter constraints where the effects of parameter projection effects are minimized, as has been a focus for the DES Y3 analyses.

D. Calibrated photometric redshifts

Finally, we test whether our photometric redshift marginalization methodology is sufficient to recover unbiased cosmological constraints on our simulations in the presence of realistic photometric redshift uncertainties. The methodology for obtaining the calibrated photometric redshift distributions that we employ in this analysis is described in detail in Appendix F as well as in the work by Buchs *et al.* [46], Cawthon *et al.* [25], Gatti *et al.* [22], Myles *et al.* [18], Sánchez *et al.* [23], and Sánchez *et al.* [24]. In summary, we use lens redshift distributions as estimated by the REDMAGIC algorithm and samples of our source redshift distributions generated by the 3SDIR algorithm [18,24] using the SOMPZ redshift distribution estimates, weighted by the likelihood of those samples given the cross-correlation of our source galaxies with REDMAGIC and spectroscopic galaxies.

We have made a significant simplifying assumption in our simulations: that we have sparse but unbiased spectroscopic redshift calibration samples for both REDMAGIC and our source photo-z estimation methodology. This is not assumed in the Y3 analysis on data, where we have added additional uncertainty in order to encapsulate possible biases in these calibration samples.

The lens photo-z uncertainties are incorporated by shifting the means of the fiducial lens $n(z)$ estimates, an approximation that has been shown to be sufficient for the
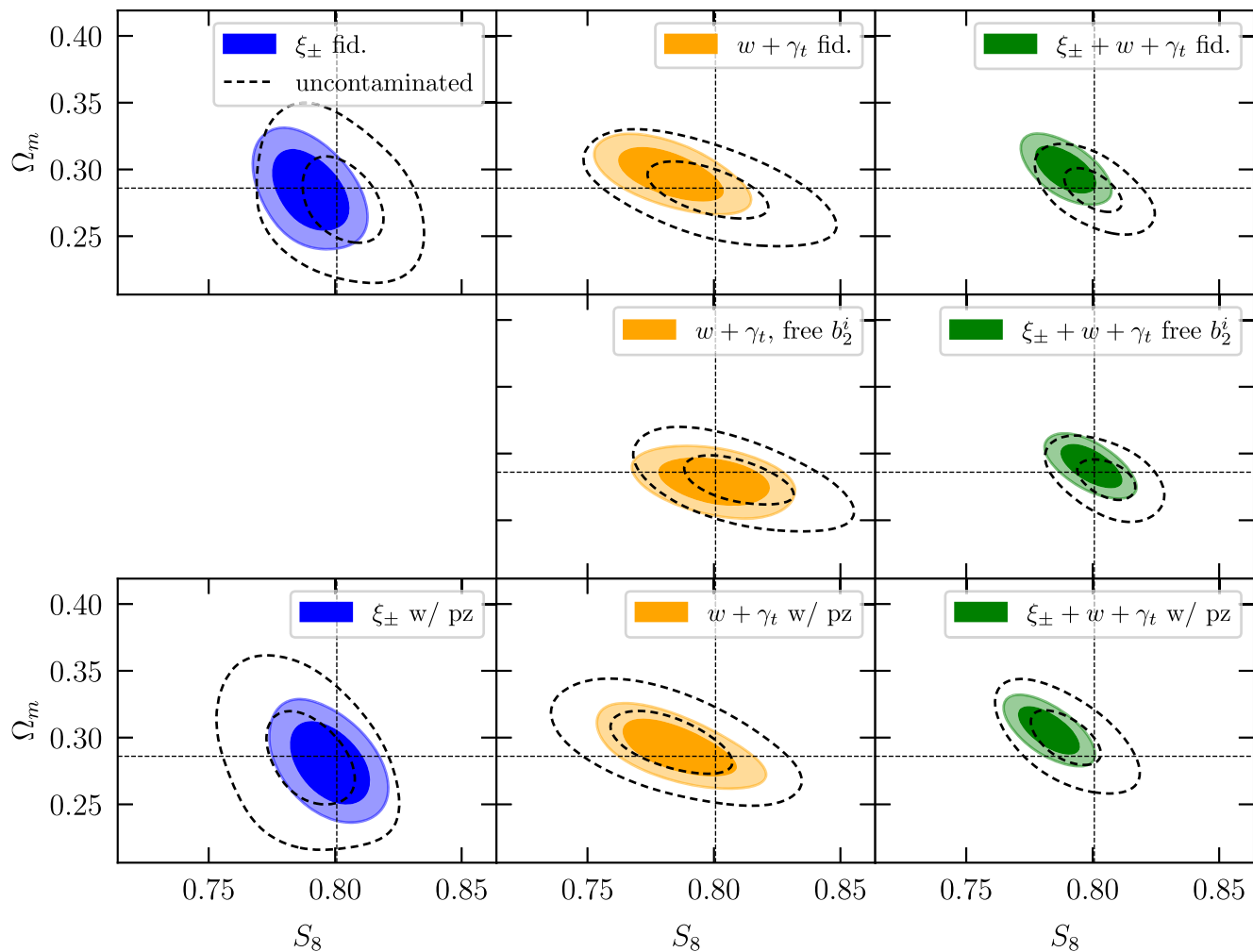
FIG. 8.   Constraints on S_8 and Ω_m from 1×2- (left), 2×2- (middle), and 3×2-point (right) analyses on the mean data vector from the full suite of simulations. Top: constraints marginalizing over only cosmology, linear bias, and IAs, assuming the true source and lens redshift distributions. Posteriors obtained using the mean BUZZARD data vector are shown as solid lines, while dashed contours use a data vector generated at the true cosmology of the simulations with the best-fit linear bias model from analysis configuration A1, i.e., the blue line in Fig. 7. The shaded BUZZARD contours are the $1/\sqrt{18}$ and $2/\sqrt{18}$ confidence regions, while the dashed contours represent the 0.3 and 1σ confidence regions for a single simulation realization. The crosshairs represent the true BUZZARD cosmology, and the difference between the dashed contours and these is a product of parameter projection effects. Middle: same as the top row, but posteriors are obtained using analysis configuration C (nonlinear bias), where the uncontaminated data vector is the best-fit nonlinear bias model from analysis configuration A2, i.e., the yellow line in Fig. 7. Bottom: same as the top and middle rows, but using analysis configuration D, i.e., using calibrated photometric redshift distributions to make our model predictions, and marginalizing over source and lens photometric redshift uncertainties. This isolates the effect of photometric redshift biases on our analysis. Dashed contours are the same as the solid contours in the top row, but scaled to represent the constraining power of a single Y3 simulation. In all cases, the probability to exceed a parameter bias of more than 0.3σ is less than 60%, as summarized in Tables II and III.

REDMAGIC lens sample in the work by Rodríguez-Monroy *et al.* [16]. We also marginalize over a rescaling of the width of $n(z)$ for the highest redshift lens bin, as was determined necessary in Cawthon *et al.* [25]. The source redshift distributions have significantly more uncertainty than our lens redshift distributions, and so we have developed methodology that enables us to explicitly marginalize over samples of this redshift distribution that encapsulate complexity that is more significant than shifts in the means of our redshift distributions. Cordero *et al.* [65]

have shown that this additional uncertainty is negligible for our DES Y3 analysis and that marginalizing over shifts in the means of our source redshift distributions is sufficient both in the BUZZARD simulations and in the DES Y3 data. As such, for this analysis, we marginalize over a Gaussian prior on these mean shift parameters. Priors on all photometric redshift related parameters are listed in Table I. In particular, we assume values for these priors based on preliminary characterizations of these effects in the DES Y3 data, although we note that the size of these priors
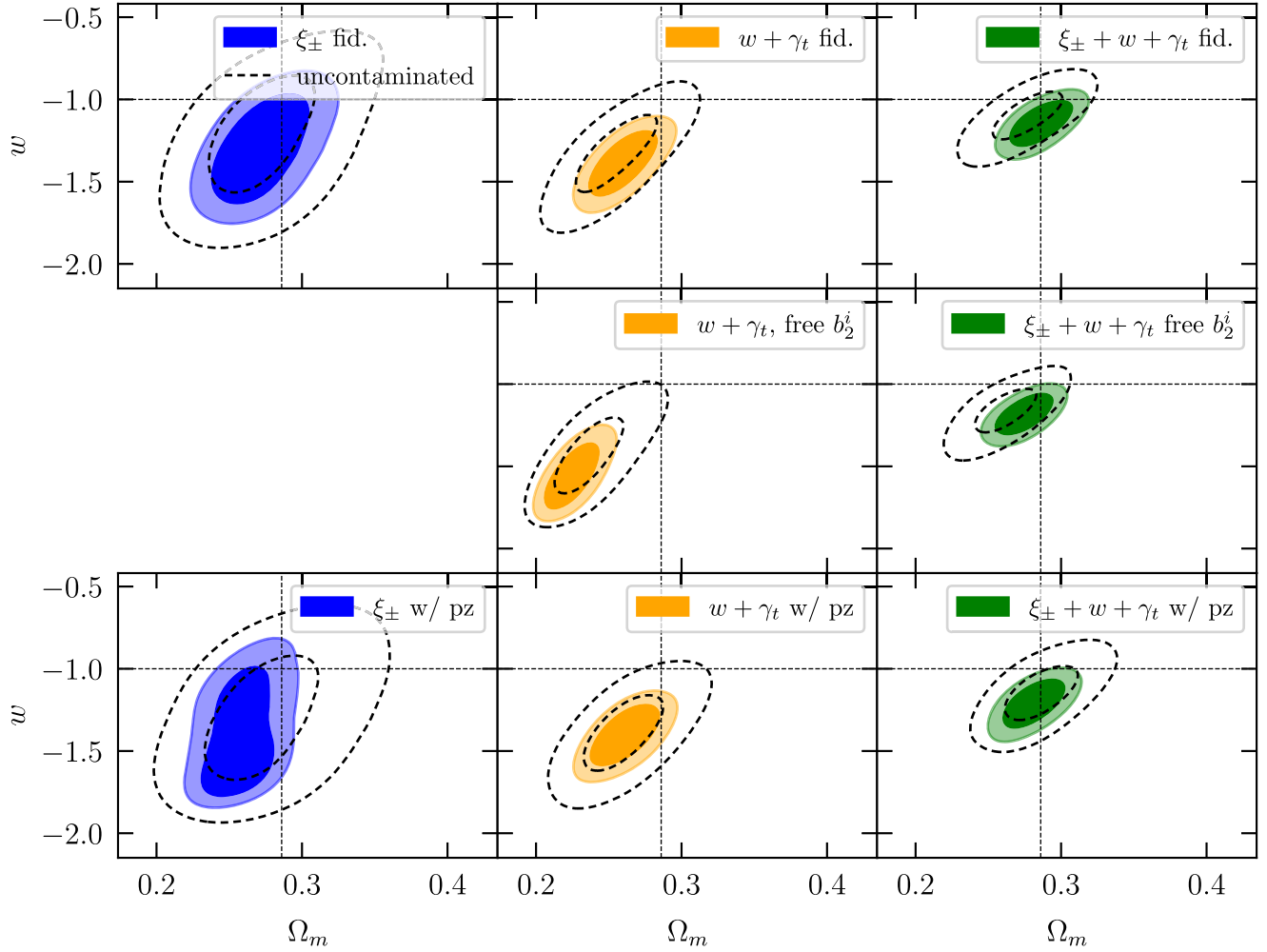
FIG. 9. Same as Fig. 8, but for a wCDM parameter space, showing constraints on Ω_m and w. In all cases, the probability to exceed a parameter bias of more than 0.3σ is less than 59%, as summarized in Table II. The wCDM constraints in the presence of nonlinear bias as shown for configuration C (middle row) are impacted by parameter projection effects, as discussed further in Sec. V C. The inclusion of shear ratios improves the constraining power of $\xi_\pm(\theta)$ alone in configuration D (bottom left) with respect to configuration B, mostly by partially breaking degeneracies with intrinsic alignment parameters, even though configuration D also marginalizes over additional nuisance parameters.

contributes negligibly to the size of our posteriors, as IA- and bias-related nuisance parameters are our dominant systematics.

Additional information on our redshift distributions as well as IA parameters is gained through the inclusion of small-scale shear-ratio measurements at the likelihood level, as described in Appendix F 4 and by Sánchez *et al.* [23]. We also marginalize over multiplicative shear biases in this analysis configuration, with uncertainties determined by a preliminary estimate of these on our Y3 data [66]. These multiplicative biases have no discernible impact on our posteriors due to their very small uncertainties.

Results of this analysis are summarized in the bottom rows of Figs. 8 and 9, where now we compare the posteriors obtained using analysis configuration D (solid) with those obtained with configuration B (dashed). The dashed contours are now approximately the 0.3σ and 1σ

constraining power of a single Y3 analysis. We choose to compare configuration D to configuration B rather than an uncontaminated analysis because we wish to isolate parameter biases that are sourced by photometric redshift (mis)estimation. We see that in all cases the shifts in posteriors are negligible, indicating that our photometric redshift calibration methodology has been successful in encapsulating the photo-z biases contained in our simulations. PTE values and parameter biases for analysis configuration D are listed in Tables II and III.

VI. SUMMARY AND CONCLUSIONS

We have presented the BUZZARD2.0 simulations, a suite of 18 synthetic DES Y3 galaxy catalogs tailored for the validation of combined clustering and lensing analyses. We have used these simulations to test the robustness of a

number of choices made in the DES Y3 3×2-point analysis, showing in particular the following:

(1) Our model can fit the measurements from our simulations at the simulations' true cosmology.

(2) Using the true redshift distributions in our simulations and sampling over cosmology, linear bias, and intrinsic alignments, we can recover the true cosmology of our simulations using cosmic shear alone, 2×2-point, and 3×2-point analyses.

(3) Using the true redshift distributions in our simulations and sampling over cosmology, *nonlinear* bias, and intrinsic alignments, we can recover the true cosmology of our simulations, using 2×2-point and 3×2-point analyses.

(4) We are able to recover unbiased cosmological constraints when assuming calibrated photometric redshift distributions, making use of the full calibration methodology applied to the Y3 data.

In Sec. II, we describe the new suite of BUZZARD simulations used in this work and elsewhere in our DES Y3 analyses, highlighting the improvements that have been made over the simulations used in DES Y1. In particular, we demonstrate improved agreement between color and magnitude distributions for our source and lens galaxy samples in Figs. 1, 2, and 10.

In Sec. III, we describe how we measure our 3×2-point data vectors, including the photometric redshift calibration methodology that was applied to the simulations. We present comparisons between redshift distributions for source and lens galaxies (Fig. 5) and the 3×2-point measurements (Figs. 3 and 4), showing significant improvements in the level of agreement from similar comparisons in DES Y1. We also highlight the four-step calibration that is applied to the source photometric redshift distributions in our simulations: Self-organizing map (SOM)-based photometric redshift estimation via SOMPZ and 3SDIR, clustering redshifts, and shear ratios. We have kept as close to the procedure employed in the DES Y3 data as possible, and we note the important aspects of these algorithms that were validated using the BUZZARD simulations that are presented in greater detail by Alarcon *et al.* [49], Cawthon *et al.* [25], Gatti *et al.* [22], Myles *et al.* [18], Sánchez *et al.* [23], and Sánchez *et al.* [24].

In Sec. IV A, we describe the models applied to the simulations, and in Sec. V, we show that these models produce constraints on S_8, Ω_m, and w that are biased at $<0.3\sigma$ (1σ) with a probability of at least $38\%(98\%)$ in LCDM and $42\%(99\%)$ in wCDM, while accounting for residual noise in the measurements from our simulations. These results are summarized in Table II and Figs. 8 and 9. Mean two-dimensional parameter biases are less than 0.3σ for all analysis configurations and are summarized in Table III. These results demonstrate that our 3×2-point analysis is robust to the assumptions made regarding bias modeling, nonlinearities in the matter distribution, higher-order lensing effects, non-Gaussianity of the likelihood function, and approximations made in photometric redshift estimation in a realistic simulated analysis setting.

We note that the probabilities quoted above would asymptote to either 0 or 1 in the limit of infinite simulations, which suggests the question of why we have not made an effort to generate more simulations in order to more precisely determine our systematic biases. The reason for this choice is partially pragmatic, as generating these simulations requires significant computational and human time. An equally important reason is that the analyses presented in this work suffer from systematic errors that contribute non-negligible uncertainty to these probabilities.

One such systematic, namely, interpretation of posterior probability distributions in the presence of projection effects, has already been discussed and argued to be important for many of the analyses presented here. Significant effort has been made to minimize projection effects in these analyses. For example, wherever possible, we have reduced the complexity of our models by removing additional parameters that are not required at the precision of our data. Although these projection effects limit the interpretability of our posteriors at the fraction of 1σ level, we reiterate that they are not a systematic bias in the traditional sense, as the size of these effects will decrease proportionally to the constraining power of our analyses.

In addition to systematics intrinsic to our posterior distributions, systematics in our simulations are also important. It is necessary to make simplifying assumptions when generating large suites of simulations as we have done here, and ruling out the contribution of these assumptions to the parameter offsets in this analysis is a time-intensive task. Efforts to systematize these tasks [67] are extremely important to the continued ability to perform analysis validation tests such as those presented in this work.

A number of important effects have been left out of these simulations, including shear calibration biases, spectroscopic incompleteness in photo-z calibration, more realistic survey inhomogeneity, galaxy intrinsic alignments, and baryonic effects on the matter distribution. The treatment of these effects is thoroughly validated elsewhere for our DES Y3 3×2-point analysis, but work is ongoing to incorporate many of these into future versions of these simulations. With significant investment, we anticipate that improvements in the methodologies used to generate these simulations will continue to meet the validation needs of upcoming joint clustering and lensing analyses.

ACKNOWLEDGMENTS

APPENDIX A: COLOR-DEPENDENT CLUSTERING

We impart a color-dependent clustering signal to our simulations in a two-step manner. First, SEDs are assigned to each simulated galaxy by finding a galaxy in the SDSS Main Galaxy Sample (MGS) with a close match in M_r and assigning its SED to the simulated galaxy, preferentially choosing blue SEDs over red ones proportional to $\frac{f_{\rm red}(z)}{f_{\rm red}(z=0)}$, where $f_{\rm red}(z)$ is the fraction of red galaxies found at redshift z as described in Appendix E.2 in DeRose *et al.* [33]. Each SED is represented in our simulations by a set of five KCORRECT templates [41].

Once each galaxy has a SED, we perform a conditional abundance matching procedure. In particular, for every galaxy, we compute R_h, the distance to the nearest halo above a mass cut of $M_{h,{\rm cut}}$. We then shuffle SEDs between galaxies in order to enforce the relation

$$P(< g - r|M_r) = P(< R_h|M_r), \qquad (A1)$$

where $g - r$ is the rest frame $g - r$ color of each galaxy. In practice, we introduce an extra parameter in this model to allow for a nonunity correlation, r between R_h and $g - r$, as described in Eqs. (12) and (13) in the work by DeRose *et al.* [40]. We use the best-fit values of r and $M_{h,{\rm cut}}$ from DeRose *et al.* [40], where this model is fit to SDSS MGS redshift space clustering measurements. This procedure makes use of conditional abundance matching algorithms implemented by Hearin *et al.* [68].

APPENDIX B: RED-SEQUENCE COLOR MODEL

We match the mean and scatter of the red-sequence in our simulations to that observed in DES Y3 data by applying the following algorithm:

---

for galaxy g with redshift z and absolute magnitude M_r **do**
    if $g - r > 0.095 - 0.035 M_r$ **then**
        for band $b \in \{r, i, z\}$ **do**
            Add mean offset, $\langle \Delta c_b(z) \rangle$, to magnitude m_b
            Add noise with variance $\Delta \sigma_b(z)^2$ to m_b

---

Here, $\langle \Delta c_b(z) \rangle$ and $\Delta \sigma_b^2$ are the mean offset in in red-sequence color and difference between scatter in the red-sequence as a function of z between the DES Y3 data and the unaltered version of these simulations. Since there are four photometric bands in DES, but only three unique colors, we can assign the mean differences in $g - r$, $r - i$, and $i - z$ to r, i, and z without loss of generality. When adding additional noise to the magnitudes in our simulations, we must account for the fact that noise in a band b contributes to two colors $X - b$ and $b - Y$, where X and Y are adjacent bands to b. Thus, to match the width of $P(X - b|z)$ and $P(b - Y|z)$, we must take $\Delta \sigma_b(z)^2 = \sigma(X - b|z)^2 - \sigma(b - Y|z)^2$. This procedure leads to very good matches to the mean and scatter of the red sequence observed in DES Y3 as shown in Fig. 2.

APPENDIX C: PHOTOMETRIC ERROR MODEL

We make use of the relationship between true photometry and noisy wide-field photometry as measured by BALROG to add photometric noise to our simulations. The algorithm for doing so is as follows:

1: **for** each simulated galaxy g **do**
2:     Find galaxy g' in DES Y3 deep fields by matching $\mathbf{m_g}$
3:     Randomly choose a wide-field injection $\hat{g}'$ of g'
4:     **If** $\hat{g}'$ was not detected **then**
5:         Set all of g's observed magnitudes to 99
6:     **else**
7:         Compute error in the wide-field magnitudes $\Delta \mathbf{m_{g'}}$
8:         Add $\Delta \mathbf{m_{g'}}$ to true simulated magnitudes $\mathbf{m_g}$

---

where $\mathbf{m_g}$ are riz magnitudes of the simulated galaxy g and $\Delta \mathbf{m_{g'}}$ is the difference between the true injected riz magnitudes from g' and the magnitudes measured from the wide-field injection $\hat{g}'$. In this way, we reproduce the relation $p(\Delta \mathbf{m_g}|\mathbf{m_g})$ measured in the DES Y3 data exactly.

APPENDIX D: DES Y3 SOURCE AND LENS GALAXY SAMPLE SELECTION

In this work, we focus on reproducing the properties of the METACALIBRATION source galaxy sample [10] and REDMAGIC lens galaxy sample [16] in order to mimic the fiducial DES Y3 3×2-point analysis. The REDMAGIC galaxy sample is selected using wide-field $griz$ photometry, and so we are able to apply the same selection algorithm described by Rozo *et al.* [39] and Rodríguez-Monroy *et al.* [16] to our simulated galaxy catalogs. In particular, we select two different REDMAGIC catalogs, that we refer to as the HIGHDENS and HIGHLUM samples.

The first of these has the highest number density and is used for the first three lens bins of our 3×2-point data vector. The second is lower number density but can be selected out to $z < 0.9$ and so is used for our two higher redshift lens bins. Applying this procedure results in magnitude and photometric redshift distributions that are
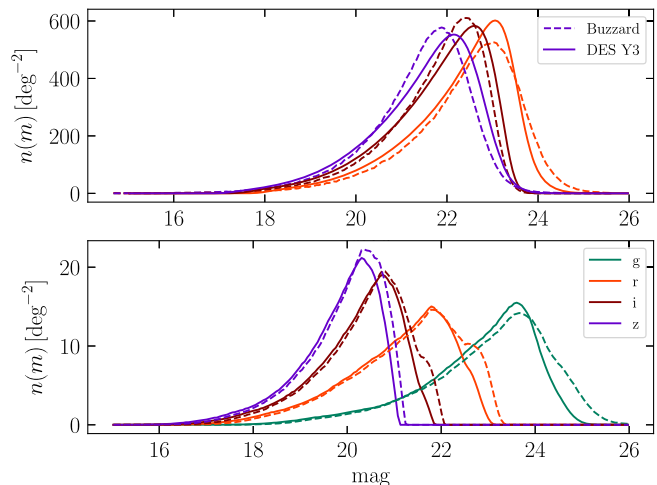

FIG. 10. Comparison of magnitude distributions for sources (top) and REDMAGIC lenses (bottom) between BUZZARD (dashed) and the DES Y3 data (solid). Different colors represent different photometric bands, where we compare riz for sources and $griz$ for lenses, because these are the bands that are used for each respective selection. Overall agreement is good, with fractional differences between BUZZARD and the DES Y3 data not exceeding 20%.

nearly identical to those found in the DES Y3 data, as shown in the right panels of Figs. 5 and 10.

Selecting a source galaxy sample that matches the properties of the DES Y3 METACALIBRATION sample is more challenging. This is partially due to the fact that this sample pushes up against the detection limit of the wide-field survey and so includes implicit selection on many properties such as surface brightness, morphology, proximity to bright objects, and observing conditions that are not modeled in our simulations. This is also a fainter sample than REDMAGIC in general and so involves galaxies in our simulations that are more sparsely observed in the spectroscopic samples that we use to inform our simulated galaxy samples. Given these limitations, we focus on constructing a simulated source sample that matches the effective number density, shape noise, and redshift distributions of the DES Y3 METACALIBRATION sample, as these are the properties that largely govern the cosmological constraining power of the cosmic shear and galaxy–galaxy lensing measurements that we wish to analyze.

To do this, we apply two cuts to our simulated galaxy sample motivated by cuts that are performed to construct the DES Y3 METACALIBRATION sample:

(1) $10 < f_i/\sigma(f_i) < 1000$,

(2) $r_{\mathrm{eff}}/r_{\mathrm{psf}} > x_1/(1 + x_2 z) + x_3$.

One of the main cuts that influences the number density of the Y3 METACALIBRATION sample is a cut on signal-to-noise ratio (SNR), $10 < \mathrm{SNR} < 1000$. The first cut above approximates this SNR cut, where we compute SNR as $\mathrm{SNR} = f_i/\sigma(f_i)$ and f_i is the noisy simulated i-band flux and $\sigma(f_i)$ is the error in that flux, computed as described in

Appendix C. Applying only this cut results in a sample with an overabundance of galaxies, so we apply an additional cut on galaxy size, motivated by the cut in the ratio of galaxy size to Point spread function (PSF) size applied to the Y3 data. Here, we approximate the Gaussian size used in the Y3 data with the simulated half-light radius r_{eff} where $r_{\text{eff}} = \sqrt{r_{50}^2 + r_{\text{psf}}^2}$ is the PSF convolved galaxy half-light radius, r_{50}, and the PSF size, r_{psf}, is taken from the DES Y3 maps of PSF size at the position of the simulated galaxy. We empirically found this functional form to provide a good fit to the tomographic source number densities in the DES Y3 data. The x_i are free parameters in this cut that we fit to the total response and inverse variance weighted number density of a preliminary version of the DES $Y3$ METACALIBRATION catalog, $n_{\text{eff}} = 5.9$. This fit yields a number density in our simulations of $n_{\text{eff}} = 5.84$. Before measuring correlation functions with this source catalog, we also explicitly match the effective shape noise, $\sigma_e = [0.247, 0.266, 0.263, 0.314]$, in the four tomographic bins used. This procedure is performed in two steps. First, we assign ellipticities using the method described in Appendix E.3 of DeRose *et al.* [33]. As the ellipticity distribution sampled from in this process is not binned in redshift, the resulting tomographic shape noise values do not exactly match those observed in the DES Y3 METACALIBRATION catalog and coincidentally are always an underestimate of those found in the data. Thus, to match the tomographic shape noise values measured in the data perfectly, we compute

$$\sigma_{e,\text{off,i}} = \sqrt{\sigma_{e,\text{data,i}}^2 - \sigma_{e,\text{fid,i}}^2},$$

where $\sigma_{e,\text{data,i}}$ and $\sigma_{e,\text{fid,i}}$ are the shape noise in the *i*th tomographic bin in the DES Y3 METACALIBRATION catalog and the fiducial shape noise in the BUZZARD simulations, respectively. We then add an additional random component to the unlensed ellipticities of galaxies binned tomographically in the simulations with mean zero and variance equal to $\sigma_{e,\text{off,i}}^2$, thus matching correcting the tomographic shape noise values in the simulations to match those in the data by definition.

This full procedure leads to good matches to the magnitude and redshift distributions observed in the DES Y3 Metacalibration sample as shown in Figs. 5 and 10. The photometric redshift estimation methodology for the simulated source sample is discussed at more length in Sec. F and in the work by Myles *et al.* [18].

APPENDIX E: TWO-POINT FUNCTION ESTIMATION

To compute the galaxy angular autocorrelation function for a single tomographic bin, $w(\theta)$, we use the Landy-Szalay estimator [69]

$$
\begin{aligned}
\hat{w}(\theta) &= \frac{DD - 2DR + RR}{RR} \\
&= \sum_{ab} \frac{w_a w_b}{N_L(N_L - 1)} - 2\sum_{aR} \frac{w_a}{N_L N_R} \\
&\quad + \sum_{RR} \frac{1}{N_R(N_R - 1)}.
\end{aligned}
$$

The first sum runs over pairs of lens galaxies separated by an angle $\theta_{\min} < \theta < \theta_{\max}$, where $\theta_{\min}$ and $\theta_{\max}$ are the edges of the angular bin, $w_{a/b}$ are the systematic weights associated with galaxies a and b, and N_L is the total number of lens galaxies in the tomographic bin under consideration. The lens weights are determined using the same algorithm described in Ref. [16]. The second and third sums run over lens-random and random-random pairs in the same angular bin, and N_R is the total number of randoms. We always take this to be $N_R = 20N_L$, as opposed to the $N_R = 40N_L$ used in the measurements on the data. We work with the mean measurement from many simulations, so extra noise resulting from the use of fewer randoms per lens galaxy is offset by this fact and allows us to save a significant amount of computing time when measuring angular clustering in the BUZZARD simulations.

Following the choices made in the DES Y3 data, we bin the REDMAGIC sample into five distinct redshift bins, with edges $z = \{0.15, 0.35, 0.5, 0.65, 0.8, 0.9\}$. Galaxies are divided between these bins using the mean of the REDMAGIC photometric redshift Probability distribution function (PDF), $z_{\text{mean,redMaGiC}}$. For the first three bins, we use the REDMAGIC HIGHDens sample; the two highest redshift bins use the REDMAGIC HIGHLUM sample.

Our estimator for tangential shear around lens galaxies includes boost factors and random point subtraction as

$$\hat{\gamma}_t(\theta) = \hat{\gamma}_{t,\text{no boosts}}(\theta)B(\theta) - \hat{\gamma}_{t,\text{rand}} \tag{E1}$$

$$= \frac{N_r}{N_l} \frac{\sum_{LS} w_L \epsilon_{t,LS}(\theta)}{N_{\text{rs}}} - \frac{\sum_{RS} \epsilon_{t,RS}(\theta)}{N_{\text{rs}}}, \tag{E2}$$

where in Eq. (E1) we have written the estimator in terms of the uncorrected tangential shear estimator $\hat{\gamma}_{t,\text{no boosts}}$, the boost factor $B(\theta)$, and the estimator for tangential shear around random points $\hat{\gamma}_{t,\text{rand}}$. We further expand this in Eq. (E2), where N_r is the number of randoms, $N_l = \sum w_L$ is the effective number of lenses given by the sum over the lens weights for all lens galaxies w_l, $\epsilon_{t,LS}$ is the tangential shear of a source-lens pair, $\epsilon_{t,RS}$ is the tangential shear of a source-random pair, and N_{rs} is the number of random-source pairs. This expression is a simplified version of that shown by Prat *et al.* [15], where we have set all Metacalibration responses as well as source and random weights to 1. We measure γ_t for each of four source galaxy tomographic bins, as described in Sec. F, around the five

tomographic lens galaxy bins, resulting in a total of 20 unique cross-correlation measurements.

The estimator of the shear-shear correlation functions can be written in terms of the measured radial, $\epsilon_\times$, and tangential, ϵ_t, components of ellipticities defined per galaxy pair along the line of separation between the galaxies:

$$\hat{\xi}_\pm(\theta) = \langle \epsilon_t \epsilon_t \pm \epsilon_\times \epsilon_\times \rangle(\theta). \tag{E3}$$

This is determined by averaging over all galaxy pairs (a, b) separated by an angle θ as

$$\hat{\xi}_\pm(\theta) = \frac{\sum_{ab}[\epsilon_{t,a}\epsilon_{t,b} \pm \epsilon_{\times,a}\epsilon_{\times,b}]}{N_{\text{pair}}}, \tag{E4}$$

where N_{pair} is the number of galaxy pairs separated by an angle θ. This is again a simplified version of the expression used in the measurements on the DES Y3 data, as we do not use METACALIBRATION responses or inverse variance weights in our simulations. We measure all ten unique auto- and cross-correlations of the four tomographic source galaxy bins.

In all cases, correlation functions are measured in 20 logarithmically spaced bins between 2.5 and 250 arc min, and the mean angle of the bin is reported as the pair-count averaged separation within that bin. We describe our redshift binning algorithms for our source and lens galaxy samples in the following Appendix.

APPENDIX F: PHOTOMETRIC REDSHIFT CALIBRATION

Compared with our source galaxy photometric redshift estimation algorithm, our lens galaxy photometric redshift estimation is relatively unchanged from that used in DES Y1. We briefly describe it here and refer the reader to the work by Rozo *et al.* [39], Cawthon *et al.* [25], and Rodríguez- Monroy *et al.* [16] for more details. As the REDMAGIC sample is a set of bright galaxies with abundant spectroscopy, we place significant confidence in the REDMAGIC photometric redshift estimates provided by the algorithm itself, $p(z_{\text{redMaGic}})$. These are obtained by constructing a red-sequence spectral template from a combination of spectroscopy and galaxy cluster members. In our simulations, we assume that we have a sparse but unbiased spectroscopic training set of similar size to that used in the data. We bin lens galaxies into five tomographic bins with edges, $\{0.15, 0.35, 0.5, 0.65, 0.8, 0.9\}$, using the mean of $p(z_{\text{redMaGic}})$. To estimate the $n(z)$ for each tomographic bin, we stack four Monte Carlo samples drawn from the $p(z_{\text{redMaGic}})$ for each galaxy.

We now describe how the source galaxy photometric redshift calibration is performed in our simulations, highlighting important similarities and differences to what is done in the data. As this methodology is new to DES Y3, much of it was developed and tested against the simulations

presented here. The calibration methodology, excluding calibration internal to the 3×2 point, is broken into four distinct steps described in the following four sections. These individual steps are described in significantly greater detail by Buchs *et al.* [46], Myles *et al.* [18], Alarcon *et al.* [49], Sánchez *et al.* [24], Gatti *et al.* [22], and Sánchez *et al.* [23].

1. SOMPZ

We aim to determine the redshift distribution $n(z)$ of the source galaxy sample, proportional to the probability, $p(z)$, that a galaxy in that sample is at redshift z. We estimate $n(z)$ by reweighting the redshift distribution of a sample with known redshift information in a suitable way. In the DES Y3 data, this reweighting is performed in two steps using three different galaxy samples: a sample of wide-field galaxies that form our weak-lensing source sample (the *weak-lensing source* sample); a set of galaxies with deep, many-band photometry (the *deep-field* sample); and a sample with deep, many-band photometry as well as securely determined redshifts (the *redshift* sample).

First, we characterize the redshift distributions of low noise galaxy detections from our deep-field photometry where we have additional photometric information in the form of near-infrared photometric measurements in $yJHK$ from the Ultra Vista survey. We discretize the $ugrizYJHK$ color space spanned by this deep-field sample using a self-organizing map (SOM) [70]. We then estimate the redshift distribution, $p(z|c)$ in each cell, c, of the so-called deep SOM by stacking redshift estimates from our redshift sample. This sample is a combination of spectroscopic surveys and COSMOS + PAU galaxies with photometric redshifts [49,71]. In our simulations, we construct analogs of these deep-field photometric catalogs by selecting patches of a single BUZZARD simulation with the same area as the deep-field catalogs in the DES Y3 data. We apply a constant level of photometric noise to these catalogs, derived from the median depth in each deep field and construct the deep SOM from this photometry. To estimate $p(z|c)$, we assume that we have a redshift sample that is free of selection biases and is the same size as the redshift sample used to estimate $p(z|c)$ in the DES Y3 data.

Analogously, we construct a wide-field SOM using our entire wide-field galaxy catalog, labeling cells in this SOM as $\hat{c}$. We wish to estimate $p(z|\vec{m})$ where $\vec{m}$ is a vector of wide-field magnitudes, or in the same SOM language, $p(z|\hat{c})$, where $\hat{c}$ is the SOM cell, or phenotype, that $\vec{m}$ is placed in. This process is complicated by the fact that the wide-field sample has significantly larger photometric uncertainties than the deep-field sample and is not supplemented by the near-IR photometry. As such, we must connect the magnitudes that we measure for our wide-field sample to those measured in our deep fields. In the data, this is performed by injecting deep-field galaxies into wide-field observations with BALROG allowing for the estimation

of $p(c|\hat{c})$, i.e., the probability that a wide-field galaxy that is placed in the wide-field SOM cell $\hat{c}$ would be placed in the deep-field SOM cell c, had it been observed in the deep fields rather than the wide field. In our simulations, we can estimate $p(c|\hat{c})$, by determining which deep SOM cell c each wide-field galaxy in cell $\hat{c}$ falls into. This relation is determined using the same number of galaxies in our simulations as we have BALROG injections in DES Y3 data.

Once galaxies have been assigned to cells $\hat{c}$ based on their photometric information, we construct tomographic bins and assign each cell to a bin. For our fiducial redshift distributions, we construct these bins according to the following procedure:

(1) To construct a set of n tomographic bins $\hat{b}$, begin with an arbitrary set of $n+1$ bin edge values e_j.

(2) Assign each galaxy in the redshift sample to the tomographic bin $\hat{b}$ that contains the median of its $p(z)$ (or its spectroscopic redshift z, if it has one). This yields a set of $N_{\text{spec},(\hat{c},\hat{b})}$ galaxies satisfying the dual condition of membership in a wide SOM cell $\hat{c}$ and a tomographic bin $\hat{b}$. This can be written as a sum over BALROG realizations i of redshift galaxies:

$$N_{\text{spec},(\hat{c},\hat{b})} = \sum_i \delta_{\hat{c},\hat{c}_i} \delta_{\hat{b},\hat{b}_i}. \tag{F1}$$

(3) Assign each wide cell $\hat{c}$ to the bin $\hat{b}$ to which a majority of its constituent redshift sample galaxies are assigned:

$$\hat{b} = \{\hat{c}|\text{argmax}_{\hat{b}} N_{\text{spec},(\hat{c},\hat{b})}\}. \tag{F2}$$

(4) Adjust the edge values e_j such that the numbers of galaxies in each tomographic bin $\hat{b}$ are approximately equal and repeat the procedure from step (ii) with the final edges e_j.

After completing this procedure, our final bin edges are $z = [0.0, 0.358, 0.631, 0.872, 2.0]$ for the Y3 weak-lensing source catalog. Because of slight differences in the Y3 source galaxy catalog and the simulated BUZZARD equivalent, the bin edges in the equivalent BUZZARD catalog are $z = [0.0, 0.346, 0.628, 0.832, 2.0]$.

Finally, we combine all these pieces of information together into the redshift distribution for each bin $\hat{b}$:

$$p(z|\hat{b},\hat{s}) \approx \sum_{\hat{c}\in\hat{b}} \sum_c p(z|c,\hat{b},\hat{s}) p(c|\hat{c},\hat{s}) p(\hat{c}|\hat{s}) \tag{F3}$$

2. 3SDIR

Once we have the SOMPZ formalism in place, we can estimate the redshift distribution of a tomographic bin using Eq. (F3). An alternative way of writing that equation, highlighting which sample is used to inform each term, is

$$p(z|\hat{b},\hat{s}) \approx \sum_{\hat{c}\in\hat{b}} \sum_c \underbrace{p(z|c)}_{\text{Redshift}} \underbrace{p(c)}_{\text{Deep}} \underbrace{\frac{p(c,\hat{c})}{p(c)p(\hat{c})}}_{\text{Balrog}} \underbrace{p(\hat{c})}_{\text{Wide}}, \tag{F4}$$

where the $\hat{b}, \hat{s}$ selections are implicit in the rhs terms. One of the main uncertainties on our estimate of $n(z)$ comes from the sample variance and shot noise present in the deep and redshift samples, which inform the probabilities $p(z|c)$ and $p(c)$. We denote the joint probability of redshift and color informed by the deep and redshift galaxy samples as a set of coefficients $\{f_{zc}\}$, with $0 \leq f_{zc} \leq 1$ and $\sum_{zc} f_{zc} = 1$, where z represents a redshift bin, and c a deep SOM color cell.

We implement the 3SDIR method, an approximate model that produces samples of $\{f_{zc}\}$ given the observed number counts of redshift and color from the deep and redshift samples including the uncertainty from shot noise and sample variance [18]. This method was developed and validated first in simulations [24], but for a nontomographic sample with a different selection than the DES Y3 source sample, where all galaxies in the deep fields had redshift information, and without a transfer function to reweight the colors in the deep field. Here, we test this method with simulations tailored to the DES Y3 samples, using an extended version of 3SDIR described by Myles *et al.* [18]. For each sample of the coefficients $\{f_{zc}\}$, we can compute a sample of the redshift distribution in each tomographic bin, propagating the uncertainty to the full shape of the $n(z)$.

To test the performance of the 3SDIR method, we perform a similar test to what was used to validate the SOMPZ algorithm, using the 300 BUZZARD realizations of the deep fields. In each realization, we draw 10^4 samples of the coefficients $\{f_{zc}^i; i = 1, \ldots, 10^4\}$ using 3SDIR and the measured number counts of redshift and color in this realization. From it, we estimate the mean redshift of each f_z^i sample, $\bar{z}^i = \sum_z z f_z^i$, and its average value $\bar{z}^{\text{3SDir}} \equiv \langle\bar{z}^i\rangle$ in each BUZZARD realization. We also compute the $\bar{z}^{\text{SOMPZ}}$ value of the single $n(z)$ from SOMPZ in each realization, which we obtain by fixing the probabilities to the number counts. We are able to verify that the expected value of the mean redshift across the 300 realizations agrees between 3SDIR and SOMPZ. We also find the pull distribution between individual $\bar{z}$ samples from 3SDIR and the fiducial SOMPZ $\bar{z}$ to be very close to a Gaussian with zero mean and unit variance.

Note that we are changing how we parametrize the uncertainty; instead of fixing the $n(z)$ and modeling the uncertainty with a shift to the distribution, we are modeling the uncertainty in $p(z,c)$ observed in the deep fields and fully propagating it to an uncertainty on the shape of the $n(z)$. Therefore, it is particularly important to show that no significant biases are introduced to the mean redshift, which is the $n(z)$'s leading-order statistic affecting the cosmological constraints of cosmic shear analysis. We have

verified this point using the BUZZARD simulations; here, we can marginalize over the effects of sample variance by producing multiple versions of the DES deep fields in different lines of sight. For a more detailed presentation of these results, we refer the reader to the work by Myles *et al.* [18].

3. Clustering redshifts

We use clustering redshift methods to further constrain the $n(z)$ samples produced by 3SDIR. Clustering redshift methods exploit the 3D overlap between a target sample (the weak-lensing sample) and a reference sample with accurate redshift estimates to infer the $n(z)$ of the former. Clustering redshift methods have been used in the past both to provide independent estimates of the $n(z)$ and to calibrate the mean of the redshift distributions obtained from photometric estimates [17,38,48,72–76].

We fully describe the DES Y3 clustering methodology for the source sample in Ref. [22]. We make use of two reference samples: red, luminous REDMAGIC galaxies with high quality photo-z estimates [16,39] and spectroscopic galaxies from the BOSS and eBOSS surveys [77–80]. The two samples complement each other. The former has a higher number density and covers the full DES Y3 footprint, but it has a limited redshift coverage, and it comes with a small uncertainty related to the REDMAGIC photo-z estimates. The latter is a spectroscopic sample and spans a larger redshift interval, allowing us to calibrate redshift distributions at higher redshift, but has a lower number density and only ∼700 squrea degrees of overlap with the DES Y3 footprint. Both samples have been reproduced in the simulations, and the BOSS and eBOSS selections are described by Gatti *et al.* [22].

For DES Y3, we explored two different approaches to using clustering redshift information. The first approach computes the mean redshift each tomographic bin in order to compare these estimates with the mean redshifts of the 3SDIR $n(z)$ samples. The mean redshift is only estimated in a limited redshift interval covered by the reference samples, excluding the tail of the distributions to mitigate the effect of magnification. This first approach is only used to cross-check the $n(z)$ from the 3SDIR.

The second method includes the clustering redshift information in a likelihood analysis, joint with sample variance and shot noise from the 3SDIR method, that returns samples of probable redshift distributions, while marginalizing over a flexible model of the redshift evolution of source galaxy clustering bias, the dominant systematic in such clustering redshift studies. The second method generates an ensemble of redshift distributions used in the DES Y3 cosmological analysis [18], and it is shown to vastly improve the accuracy of the shape of $n(z)$ derived from photometric data alone. Both methods have been tested in simulations and proved unbiased within uncertainties prior to application to data [22].

4. Shear ratio

In the DES Y3 cosmological analysis, we use the ratios of small-scale galaxy–galaxy lensing measurements sharing the same lens redshift bin and two different source tomographic bins to constrain redshift uncertainties and other systematics or nuisance parameters of our model. We briefly summarize this probe and direct the reader to detailed description by Sánchez *et al.* [23] for more information and robustness tests.

Lensing ratios or *shear ratios* have the advantage that if the lens galaxies are tightly binned in redshift they are mostly geometrical and can be modeled in the small, nonlinear scale regime where we are not able to accurately model the original tangential shear quantity. Lensing ratios at small scales are therefore able to provide very valuable independent information that otherwise would be discarded. These shear-ratio measurements have been used before, especially as a test of the source redshift distributions, but this is the first time they are fully integrated as an additional probe within the 3×2-point project. This permits us to use the entire constraining power of the lensing ratios, not only exploiting the dependency on the redshift distributions.

The shear-ratio data vector consists of nine numbers, each one corresponding to the scale-averaged lensing ratio for a given lens and two source bins. In this work, we use three lens redshift bins and four source redshift bins. Note that the DES Y3 3×2-point project uses five lens bins, but we choose to discard the two highest redshift bins to construct lensing ratios both because they do not increase the S/N substantially and because the impact of lens magnification is much stronger for the highest redshift lens bins, and we prefer not to be dominated by lens magnification even though it is included in the modeling. From these redshift bins, given that we construct ratios of tangential shear measurements with a given fixed lens bin and two different source bins for each lens bin we can construct three independent ratios to make a total set of nine independent ratios. Regarding the angular scales, we measure the lensing ratios in the same angular binning used for the 3×2-point analysis, and then we apply the scale cuts as detailed in Table 2 in the work by Sánchez *et al.* [23]. Summarizing, we discard the scales already used in the 3×2-point analysis for the galaxy–galaxy lensing probe and scales where the IA model is not applicable for the ratio combinations that have significant overlap between source and lens galaxy redshift distributions, as these are most affected by IA.

This data vector is used in a separate lensing ratio likelihood, assumed to be Gaussian. The covariance for this data vector comes from the propagation of the theoretical galaxy–galaxy lensing covariance and is independent of the 3×2-point covariance, as detailed and validated by Sánchez *et al.* [23]. The constraints on the mean redshifts of each source tomographic bin from this likelihood are

used as an independent validation of the SOMPZ and WZ estimates of these values. Furthermore, we combine this shear-ratio likelihood with the 3×2-point likelihood described below to self-consistently constrain source redshift distributions, as well as other parameters, such as shear multiplicative biases and intrinsic alignments. The BUZZARD simulations were used to validate a number of assumptions made when using shear-ratio data, as described by Sánchez *et al.* [23], and we further validate the use of this information in this work showing that its combination with all other redshift information leads to unbiased cosmological constraints in our BUZZARD simulated analyses.

APPENDIX G: COVARIANCE MATRIX

We model the statistical uncertainties of the two-point function measurements considered in this paper as a multivariate Gaussian distribution. Our model of the disconnected four-point function part of the covariance matrix of that distribution (also known as the Gaussian part of the covariance) is described by Friedrich *et al.* [81] and includes analytic treatment of bin averaging and sky curvature. The connected four-point function part of the covariance matrix and the contribution from supersample covariance are modeled as described by Krause and Eifler [63].

In Friedrich *et al.* [81], we demonstrate the robustness of our analysis with respect to the details of our covariance model and show that deviations from the Gaussian likelihood assumption are negligible in our analysis setup. In that paper, we identify approximations in our treatment of survey geometry to be the main source of inaccuracy in the covariance model. When considering the full 3×2-point data vector, this leads to an underestimation of our uncertainties in key cosmological parameters by about 3%–4%, and on average it also increases the χ^2 between measurements and the corresponding maximum likelihood models by about 4%.

We analytically marginalize over all enclosed mass parameters, which is possible under our assumptions of a Gaussian likelihood and a Gaussian prior for these parameters, characterized by a width of σ_{B^i}. We use very broad priors on these parameters, as motivated by Pandey *et al.* [32]. This would make our covariance numerically unstable to invert. To circumvent this, we use the Shermann-Morrison formula to analytically marginalize over B^i directly in the inverse covariance matrix, N^{-1}, via

$$N^{-1} = C^{-1}U(I + U^T C^{-1}U)^{-1}U^T C^{-1}, \qquad \text{(G1)}$$

where C^{-1} is the inverse covariance matrix without the enclosed mass contribution, I is the identity matrix, and U is an $N_d \times N_{\text{lens}}$ with ith column given by $\sigma_{B^i}\vec{t}^i$ and

$$t_a^i = \begin{cases} 0 & \text{if lens redshift bin for element a is not } i \\ \beta_{ij}\theta_a & \text{otherwise,} \end{cases},$$
$$\text{(G2)}$$

where a is the data vector index, j is the source galaxy bin, and θ_a is the angular separation associated with the ath element of the data vector. We take $\sigma_{B^i} = 10000$ for all i.

We generate the covariance used for all of the analyses in this work using true BUZZARD cosmology, source and lens number densities, shape noise values, and redshift distributions.

APPENDIX H: REDSHIFT-SPACE DISTORTIONS IN BUZZARD

In the process of preparing this paper for publication, an error in our implementation of redshift space distortions was discovered. Specifically, the proper motions of galaxies were included in their observed redshifts as

$$z_{\text{obs}} = z_{\text{cos}} + \frac{v_{\text{LOS}}}{c} \qquad \text{(H1)}$$
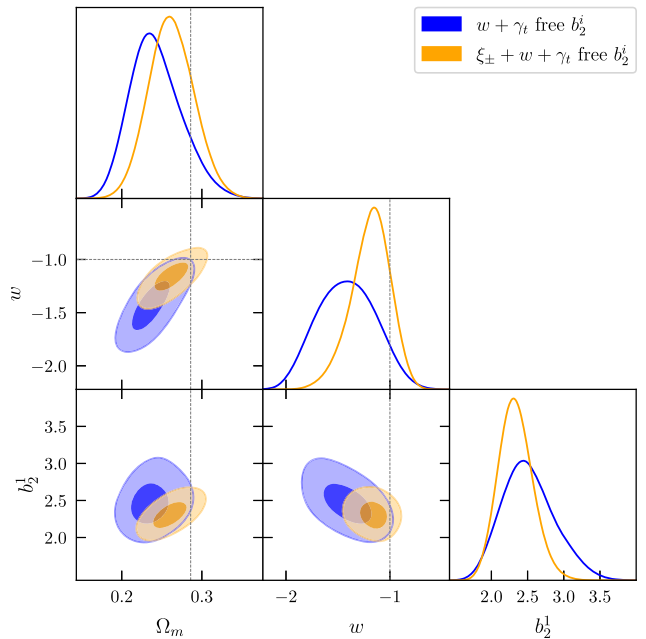
instead of using the correct expression

FIG. 11. 1 and 0.3σ posteriors of our analysis configuration C, wCDM model fit to uncontaminated 2×2-point and 3×2-point data vectors. These are the same posteriors as the dashed lines in the second and third columns of the middle row of fig. 9. There is a degeneracy between b_2^1 (the b_2 nonlinear bias parameter in the first lens bin) and w in both cases. In the less-constraining 2×2-point case, there is a tail to high values of b_2^1, which correlate with low values of w. When including the cosmic shear data, this tail is truncated, symmetrizing the b_2^1 posterior and thus reducing the projection effect on the w posterior.

$$z_{\rm obs} = z_{\rm cos} + (1 + z) \frac{v_{\rm LOS}}{c}, \qquad (\text{H2})$$

where $z_{\rm cos}$ is the pure cosmological redshift and $v_{\rm LOS}$ is the line-of-sight velocity. This error impacts the effect of redshift-space distortions on $w(\theta)$ and $\gamma_t(\theta)$ measurements. As our analysis was nearly completed when we discovered this error, we opted to correct for this effect in our model for RSD rather than change our simulations. This is accomplished by rescaling the linear growth rate that is used in the Kaiser model for RSD by a factor of $(1 + z)$, i.e., $f \to \frac{f}{(1+z)}$.

The impact of this error on the analyses presented in this work is negligible; not accounting for this error increases the chi-squared for the analyses presented in Sec. VA by 2.1 for $w(\theta)$, and the impact on all posteriors presented in this work is virtually unnoticeable.

APPENDIX I: PROJECTION EFFECTS

We have argued that projection effects cause offsets between the maxima of our one- and two-dimensional posteriors and the input cosmology of our simulations. Such effects arise because of correlations between parameters of interest, such as cosmological parameters, and asymmetrically constrained nuisance parameters. When the nuisance parameters are marginalized over, the asymmetry of the constraint on these parameters leads to more posterior volume in one direction in the parameter of interest than the other, thus shifting the mean of the posterior of the parameter of interest away from its true value.

Figure 11 demonstrates this effect using the uncontaminated posteriors from the analysis configuration C, wCDM model. The less constraining 2×2-point analysis shows very significant projection effects, with the true cosmology of the simulated data vectors lying on the edge of the 1σ $w - \Omega_m$ posterior, while the more constraining 3×2-point analysis is affected less significantly by projection effects. There is a large degeneracy between w and the b_2 parameter in the first lens bin for both analyses, although it is less apparent for the 3×2-point analysis where b_2^1 is much better constrained. In the 2×2-point case, the correlation between b_2^1 and w in conjunction with the long tail to high b_2 prevents very negative values of w from being ruled out. When marginalizing over b_2, this shifts the mean of the one-dimensional posterior of w to more negative values. In the 3×2-point case, high values of b_2^1 are ruled out, and so the tail to small w is removed, thus reducing the projection effect on this analysis compared with the 2×2-point analysis.

[1] T. M. C. Abbott, F. B. Abdalla, A. Alarcon *et al.*, Phys. Rev. D **98**, 043526 (2018).
[2] C. Heymans, T. Tröster, M. Asgari *et al.*, Astron. Astrophys. **646**, A140 (2021).
[3] E. Krause, T. F. Eifler, J. Zuntz *et al.* (to be published).
[4] E. Suchyta, E. M. Huff, J. Aleksić *et al.*, Mon. Not. R. Astron. Soc. **457**, 786 (2016).
[5] N. MacCrann, J. Blazek, B. Jain, and E. Krausea, Mon. Not. R. Astron. Soc. **491**, 5498 (2020).
[6] S. Pandey, E. Krause, B. Jain *et al.*, Phys. Rev. D **102**, 123522 (2020).
[7] G. Rossi, P. D. Choi, J. Moon *et al.*, Mon. Not. R. Astron. Soc. **505**, 377 (2021).
[8] S. Sugiyama, M. Takada, Y. Kobayashi, H. Miyatake, M. Shirasaki, T. Nishimichi, and Y. Park, Phys. Rev. D **102**, 083520 (2020).
[9] S. Everett *et al.*, Astrophys. J. Suppl. Ser. **258**, 15 (2022).
[10] M. Gatti, E. Sheldon *et al.*, Mon. Not. R. Astron. Soc. **504**, 4312 (2021).
[11] A. Kannawadi, H. Hoekstra, L. Miller *et al.*, Astron. Astrophys. **624**, A92 (2019).
[12] R. Mandelbaum, F. Lanusse, A. Leauthaud, R. Armstrong, M. Simet, H. Miyatake, J. E. Meyers, J. Bosch, R. Murata, S. Miyazaki, and M. Tanaka, Mon. Not. R. Astron. Soc. **481**, 3170 (2018).
[13] S. Samuroff, S. L. Bridle, J. Zuntz *et al.*, Mon. Not. R. Astron. Soc. **475**, 4524 (2018).
[14] J. Elvin-Poole, M. Crocce, A. J. Ross *et al.*, Phys. Rev. D **98**, 042006 (2018).
[15] J. Prat *et al.*, Phys. Rev. D **105**, 083528 (2022).
[16] M. Rodríguez-Monroy *et al.*, Mon. Not. R. Astron. Soc. **511**, 2665 (2022).
[17] H. Hildebrandt, J. L. van den Busch, A. H. Wright *et al.*, Astron. Astrophys. **647**, A124 (2021).
[18] J. Myles, A. Alarcon *et al.*, Mon. Not. R. Astron. Soc. **505**, 4249 (2021).
[19] M. Tanaka, J. Coupon, B.-C. Hsieh, S. Mineo, A. J. Nishizawa, J. Speagle, H. Furusawa, S. Miyazaki, and H. Murayama, Publ. Astron. Soc. Jpn. **70**, S9 (2018).
[20] W. G. Hartley, C. Chang, S. Samani *et al.*, Mon. Not. R. Astron. Soc. **496**, 4769 (2020).
[21] A. Amon *et al.*, Phys. Rev. D **105**, 023514 (2022).
[22] M. Gatti, G. Giannini *et al.*, Mon. Not. R. Astron. Soc. **510**, 1223 (2021).
[23] C. Sánchez, J. Prat *et al.*, Phys. Rev. D **105**, 083529 (2022).
[24] C. Sánchez, M. Raveri, A. Alarcon, and G. M. Bernstein, Mon. Not. R. Astron. Soc. **498**, 2984 (2020).
[25] R. Cawthon *et al.* (to be published).
[26] A. Porredon *et al.* (to be published).
[27] A. Porredon *et al.*, Phys. Rev. D **103**, 043503 (2021).
[28] E. Krause *et al.* (to be published).
[29] M. R. Becker, Mon. Not. R. Astron. Soc. **435**, 115 (2013).
[30] J. Elvin-Poole, N. MacCrann *et al.* (to be published).

[31] L. F. Secco, S. Samuroff *et al.*, Phys. Rev. D **105,** 023515 (2022).

[32] S. Pandey *et al.* (to be published).

[33] J. DeRose, R. H. Wechsler, M. R. Becker *et al.*, arXiv: 1901.02401.

[34] V. Springel, Mon. Not. R. Astron. Soc. **364,** 1105 (2005).

[35] M. Crocce, S. Pueblas, and R. Scoccimarro, Mon. Not. R. Astron. Soc. **373,** 369 (2006).

[36] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538,** 473 (2000).

[37] R. H. Wechsler, J. DeRose, M. T. Busha *et al.*, arXiv: 2105.12105.

[38] R. Cawthon, C. Davis, M. Gatti *et al.*, Mon. Not. R. Astron. Soc. **481,** 2427 (2018).

[39] E. Rozo, E. S. Rykoff, A. Abate *et al.*, Mon. Not. R. Astron. Soc. **461,** 1431 (2016).

[40] J. DeRose, M. R. Becker, and R. H. Wechsler, arXiv: 2105.12104.

[41] M. R. Blanton and S. Roweis, Astrophys. J. **133,** 734 (2007).

[42] W. G. Hartley, A. Choi *et al.*, Mon. Not. R. Astron. Soc. **509,** 3547 (2021).

[43] E. S. Sheldon and E. M. Huff, Astrophys. J. **841,** 24 (2017).

[44] T. M. C. Abbott *et al.* (DES Collaboration), Phys. Rev. D **105,** 023520 (2022).

[45] M. P. van Daalen, I. G. McCarthy, and J. Schaye, Mon. Not. R. Astron. Soc. **491,** 2424 (2020).

[46] R. Buchs, C. Davis *et al.*, Mon. Not. R. Astron. Soc. **489,** 820 (2019).

[47] X. Fang, E. Krause, T. Eifler, and N. MacCrann, J. Cosmol. Astropart. Phys. 05 (2020) 010.

[48] D. J. Bates, R. Tojeiro, J. A. Newman, V. Gonzalez-Perez, J. Comparat, D. P. Schneider, M. Lima, and A. Streblyanska, Mon. Not. R. Astron. Soc. **486,** 3059 (2019).

[49] A. Alarcon, C. Sánchez, G. M. Bernstein, and E. Gaztañaga, Mon. Not. R. Astron. Soc. **498,** 2614 (2020).

[50] P. McDonald and A. Roy, J. Cosmol. Astropart. Phys. 08 (2009) 020.

[51] S. Saito, T. Baldauf, Z. Vlah, U. Seljak, T. Okumura, and P. McDonald, Phys. Rev. D **90,** 123522 (2014).

[52] T. Matsubara, Phys. Rev. D **90,** 043537 (2014).

[53] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, Mon. Not. R. Astron. Soc. **341,** 1311 (2003).

[54] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Astrophys. J. **761,** 152 (2012).

[55] N. Kaiser, Mon. Not. R. Astron. Soc. **227,** 1 (1987).

[56] M. L. Brown, A. N. Taylor, N. C. Hambly, and S. Dye, Mon. Not. R. Astron. Soc. **333,** 501 (2002).

[57] D. N. Limber, Astrophys. J. **117,** 134 (1953).

[58] T. Baldauf, R. E. Smith, U. Seljak, and R. Mandelbaum, Phys. Rev. D **81,** 063531 (2010).

[59] Y. Park, E. Rozo, and E. Krause, Phys. Rev. Lett. **126,** 021301 (2021).

[60] M. P. van Daalen, J. Schaye, C. M. Booth, and C. Dalla Vecchia, Mon. Not. R. Astron. Soc. **415,** 3649 (2011).

[61] W. J. Handley, M. P. Hobson, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **453,** 4385 (2015).

[62] J. Zuntz, M. Paterno, E. Jennings *et al.*, Astron. Comput. **12,** 45 (2015).

[63] E. Krause and T. Eifler, Mon. Not. R. Astron. Soc. **470,** 2100 (2017).

[64] C. Howlett, A. Lewis, A. Hall, and A. Challinor, J. Cosmol. Astropart. Phys. 04 (2012) 027.

[65] J. P. Cordero, I. Harrison *et al.*, Mon. Not. R. Astron. Soc. **511,** 2170 (2022).

[66] N. MacCrann *et al.*, Mon. Not. R. Astron. Soc. **509,** 3371 (2021).

[67] Y.-Y. Mao, Tomuram, R. Zhou *et al.*, Lsstdesc/Descqa: Descqa V2 Alpha Release (V2.0.0-0.4.5), 2017.

[68] A. P. Hearin, D. Campbell, E. Tollerud *et al.*, Astrophys. J. **154,** 190 (2017).

[69] S. D. Landy and A. S. Szalay, Astrophys. J. **412,** 64 (1993).

[70] T. Kohonen, Biol. Cybern. **43,** 59 (1982).

[71] C. Laigle, H. J. McCracken, O. Ilbert *et al.*, Astrophys. J. Suppl. Ser. **224,** 24 (2016).

[72] C. Davis, M. Gatti, P. Vielzeuf *et al.*, arXiv:1710.02517.

[73] C. Davis, E. Rozo, A. Roodman *et al.*, Mon. Not. R. Astron. Soc. **477,** 2196 (2018).

[74] H. Hildebrandt, M. Viola, C. Heymans *et al.*, Mon. Not. R. Astron. Soc. **465,** 1454 (2017).

[75] A. Johnson, C. Blake, A. Amon *et al.*, Mon. Not. R. Astron. Soc. **465,** 4118 (2017).

[76] J. L. van den Busch, H. Hildebrandt, A. H. Wright, C. B. Morrison, C. Blake, B. Joachimi, T. Erben, C. Heymans, K. Kuijken, and E. N. Taylor, Astron. Astrophys. **642,** A200 (2020).

[77] R. Ahumada, C. Allende Prieto, A. Almeida *et al.*, Astrophys. J. Suppl. Ser. **249,** 3 (2020).

[78] K. S. Dawson, D. J. Schlegel, C. P. Ahn *et al.*, Astrophys. J. **145,** 10 (2013).

[79] K. S. Dawson, J.-P. Kneib, W. J. Percival *et al.*, Astrophys. J. **151,** 44 (2016).

[80] S. A. Smee, J. E. Gunn, A. Uomoto *et al.*, Astrophys. J. **146,** 32 (2013).

[81] O. Friedrich *et al.*, Mon. Not. R. Astron. Soc. **508,** 3125 (2021).