Dark Energy Survey Y3 results: Blending shear and redshift biases in image simulations

N. MacCrann,[1,†] M. R. Becker,[2] J. McCullough,[3,4,5] A. Amon,[3,4,5] D. Gruen,[4,3,5] M. Jarvis,[6] A. Choi,[7]
M. A. Troxel,[8] E. Sheldon,[9] B. Yanny,[10] K. Herner,[10] S. Dodelson,[11] J. Zuntz,[12] K. Eckert,[6] R. P. Rollins,[13]
T. N. Varga,[14,15] G. M. Bernstein,[6] R. A. Gruendl,[16,17] I. Harrison,[13,18] W. G. Hartley,[19] I. Sevilla-Noarbe,[20]
A. Pieres,[21,22] S. L. Bridle,[13] J. Myles,[4] A. Alarcon,[2] S. Everett,[23] C. Sánchez,[6] E. M. Huff,[24] F. Tarsitano,[25]
M. Gatti,[26] L. F. Secco,[6] T. M. C. Abbott,[27] M. Aguena,[28,21] S. Allam,[10] J. Annis,[10] D. Bacon,[29]
E. Bertin,[30,31] D. Brooks,[32] D. L. Burke,[3,5] A. Carnero Rosell,[33,34] M. Carrasco Kind,[16,17] J. Carretero,[26]
M. Costanzi,[35,36] M. Crocce,[37,38] M. E. S. Pereira,[39] J. De Vicente,[20] S. Desai,[40] H. T. Diehl,[10]
J. P. Dietrich,[41] P. Doel,[32] T. F. Eifler,[42,24] I. Ferrero,[43] A. Ferté,[24] B. Flaugher,[10] P. Fosalba,[37,38]
J. Frieman,[10,44] J. García-Bellido,[45] E. Gaztanaga,[37,38] D. W. Gerdes,[46,39] T. Giannantonio,[47,48]
J. Gschwend,[21,22] G. Gutierrez,[10] S. R. Hinton,[49] D. L. Hollowood,[23] K. Honscheid,[7,50] D. J. James,[51]
O. Lahav,[32] M. Lima,[28,21] M. A. G. Maia,[21,22] M. March,[6] J. L. Marshall,[52] P. Martini,[7,53,54] P. Melchior,[55]
F. Menanteau,[16,17] R. Miquel,[56,26] J. J. Mohr,[41,14] R. Morgan,[57] J. Muir,[3] R. L. C. Ogando,[21,22]
A. Palmese,[10,44] F. Paz-Chinchón,[47,17] A. A. Plazas,[55] M. Rodriguez-Monroy,[20] A. Roodman,[3,5]
S. Samuroff,[11] E. Sanchez,[20] V. Scarpine,[10] S. Serrano,[37,38] M. Smith,[58] M. Soares-Santos,[39] E. Suchyta,[59]
M. E. C. Swanson,[17] G. Tarle,[39] D. Thomas,[29] C. To,[4,3,5] and R.D. Wilkinson[60]

(DES Collaboration)

ABSTRACT

As the statistical power of galaxy weak lensing reaches percent level precision, large, realistic and robust simulations are required to calibrate observational systematics, especially given the increased importance of object blending as survey depths increase. To capture the coupled effects of blending in both shear and photometric redshift calibration, we define the effective redshift distribution for lensing, $n_\gamma(z)$, and describe how to estimate it using image simulations. We use an extensive suite of tailored image simulations to characterize the performance of the shear estimation pipeline applied to the Dark Energy Survey (DES) Year 3 dataset. We describe the multi-band, multi-epoch simulations, and demonstrate their high level of realism through comparisons to the real DES data. We isolate the effects that generate shear calibration biases by running variations on our fiducial simulation, and find that blending-related effects are the dominant contribution to the mean multiplicative bias of approximately −2%. By generating simulations with input shear signals that vary with redshift, we calibrate biases in our estimation of the effective redshift distribution, and demonstrate the importance of this approach when blending is present. We provide corrected effective redshift distributions that incorporate statistical and systematic uncertainties, ready for use in DES Year 3 weak lensing analyses.

Key words: gravitational lensing: weak – cosmology: large-scale structure of Universe

1 INTRODUCTION

While weak gravitational lensing of galaxies has enormous potential as a cosmological probe (e.g. Weinberg et al. 2013; Albrecht et al. 2006), measurements of the weak lensing shear have proven

arXiv:2012.08567v2 [astro-ph.CO] 26 Jan 2022

to be extremely difficult in practice (e.g. Mandelbaum et al. 2014). The shear manifests as a small distortion in the observed shape of a galaxy. Its measurement is subject to numerous biases, and requires accurate calibrations of many properties of the input images. Typically, these biases have been quantified by assuming a linear relation between component x of the measured shear $\bar{g}_x^{\text{obs}}$ (averaged over an ensemble of galaxies), and component y of the true shear g_y^{true} (e.g. Heymans et al. 2006)

$$\bar{g}_x^{\text{obs}} = (1 + m_{xy})g_y^{\text{true}} + c_x, \tag{1}$$

where m_{xy} is known as the multiplicative bias, and c is known as the additive bias. This linear relation is expected to hold in the weak lensing regime, where g_y^{true} is small (and so contributions of order $(g_y^{\text{true}})^2$ can be neglected). One can also consider the multiplicative term as quantifying the linear response, $R_{xy} \equiv 1 + m_{xy}$ of the shear estimate to a change in the input shear i.e.

$$R_{xy} \equiv \frac{\partial \bar{g}_x^{\text{obs}}}{\partial g_y^{\text{true}}}. \tag{2}$$

The off-diagonal elements of m_{xy} (and R_{xy}) are often assumed (and empirically found) to be zero, allowing for the more common notation where $m_x \equiv m_{xx}$. In the following we will usually drop the shear component subscripts for brevity, with expressions involving $\bar{g}^{\text{obs}}$ and g^{true} generally holding for either component of the shear.

The most thoroughly studied biases in weak lensing measurements have mainly been performed with simulations of isolated objects. These include noise bias (e.g. Refregier et al. 2012; Kacprzak et al. 2012, model bias (e.g. Voigt & Bridle 2010), selection biases (e.g. Kaiser 2000; Bernstein & Jarvis 2002; Hirata & Seljak 2003), and biases from miscorrecting for the image point-spread function (PSF) (e.g. Paulin-Henriksson et al. 2008). These problems were tackled by community-driven efforts like the STEP (Heymans et al. 2006; Massey et al. 2007) and GREAT (Bridle et al. 2010; Kitching et al. 2013; Mandelbaum et al. 2015) challenges, and aided by the development of the widely used GALSIM[1] software for simulation of astronomical images (Rowe et al. 2015). For isolated objects, the aforementioned biases have largely been solved by methods like METACALIBRATION (Huff & Mandelbaum 2017; Sheldon & Huff 2017) and "Bayesian Fourier Domain" (BFD) (Bernstein et al. 2016), at least for sufficiently well understood data (i.e. with accurately-characterized noise and background levels and PSF). The METACALIBRATION method is particularly powerful because it does not rely on calibration simulations, which inevitably rely on assumptions about the properties of the faint, often poorly resolved galaxies used in weak lensing analyses.

More recently, some studies have begun to study shear calibration biases in the context of multiple objects and blending. It has generally been assumed that in this case the use of image simulations will be essential, and these have been used for the calibration of recent weak lensing cosmology analyses, for example by Fenech Conti et al. (2017); Kannawadi et al. (2019) (for the Kilo-Degree Survey[2]), Mandelbaum et al. (2018) (for the Hyper Suprime-Cam Subaru Strategic Program[3]), and Samuroff et al. (2018); Kacprzak et al. (2020) (for DES Year 1 analyses). Works such as Hoekstra et al. (2017) and Euclid Collaboration (2019), with an eye to deeper upcoming datasets, have used image simulations to study effects such as the impact of undetected galaxies on the shear calibration.

In parallel to these simulation-based calibration efforts, Sheldon et al. 2020 developed new measurement methodology, METADETECTION, which corrects for much of the impact of blending, in particular the significant shear biases imparted by detection and deblending algorithms, as well as the impact of blending at the shape measurement stage. The METADETECTION method does not require simulation-based calibration, and exhibits extremely low levels of shear calibration bias even on (constant shear) simulations designed to match the depth of Rubin Observatory Legacy Survey of Space and Time[4] (LSST) data.

Current and future surveys will have large amounts of blending of objects at different redshifts (e.g. Dawson et al. 2016). The component galaxies in blended systems will therefore often experience different shears. As we will discuss in Section 2, the impact of this on weak lensing statistics cannot be fully accounted for by the use of simplified constant shear simulations used thus far in the field to calibrate shear measurements, or corrections from shear estimation methods like METADETECTION. In order to gain intuition into the possible effects, consider the following simplified situation, shown in Figure 1. In both panels we input a pair of galaxies at different redshifts; for the sake of illustration we arbitrarily set one at low redshift ($z = 0.25$) and one at high redshift ($z = 0.75$). The high redshift galaxy is placed at the center of the stamp in both panels. For simplicity, we have fixed both galaxies to be round before lensing. In the top panels, the galaxies are not blended together, and in the bottom panels they are. In both cases we assume we can unambiguously detect two separate objects and precisely know their centroids.

Let us consider the response to shear of the central ($z = 0.75$) object in each stamp. We can apply shear separately at the two redshifts from which the light in the image is sourced (which in this case just means applying shear separately to the two galaxies present in the stamp). The right-hand panels of Figure 1 show the response to shear of the measured shape of the central galaxy, as a function of the redshift of the applied shear. The response is defined here as above, as $R = \left(\partial \bar{g}^{\text{obs}} / \partial g^{\text{true}}\right)$. We estimated these responses numerically from our simple simulations using the METACALIBRATION method. In the top right panel, we see that the high redshift object does not respond to the shear of the low redshift one (it is zero at $z = 0.25$) and has a unit response to a shear applied at its own redshift (the peak at $z = 0.75$). This result makes sense since METACALIBRATION is known to be unbiased at high precision for idealized cases such as this, and the objects do not overlap.

The more interesting case is when the galaxies are blended, shown in the lower panels of Figure 1. In this case we see that the high redshift object responds to the shear of the low redshift object (the small peak at $z = 0.25$). It also has an apparent greater than unity response to the shear applied at its own redshift (the peak at $z = 0.75$ that is greater than one). Both of these effects are due to the galaxy being blended with the low redshift neighbor. The latter effect is likely due to the positioning of the neighboring galaxy in the positive g_1 direction, which is the shear component for which we compute the response.

However, the response of the high redshift object to a low redshift shear is a qualitatively different effect, distinct from standard multiplicative biases due to blending or detection; it is a bias that depends not only on the presence of the neighbor, but also on the shear applied to the neighbor. This indicates that the shape

---

measurement of the high redshift object is carrying information about the low redshift shear.

In fact, we assert that it is the response to shear that defines how we should weight the redshifts to which we assign the shear information for a given object in a weak lensing analysis. This insight is a key subject of this work, where we will definitively measure these effects in simulations of the DES Year 3 (Y3) analysis (note that "Year 3" includes the first three years of DES observations). There are also important implications for analysis of future surveys. As the amount of blending increases with increased depth, inferring the redshift distribution relevant for lensing and the shear calibration biases will be a joint analysis task. In the example above, we have described how a single detected object can have a (non-unity) response to shear at multiple redshifts. This effect cannot be fully described by the traditional multiplicative bias, m. The shear calibration and the effective redshift distribution cannot be fully decoupled.

In this work, we expand upon the ideas from this simple example and apply them to the DES Year 3 shear analysis. We introduce our formalism for accounting for blending in Section 2. In Section 3 we describe realistic simulations of the DES Year 3 survey, validating them against the data. Then in Section 4 we describe and investigate the source of the traditional shear calibration biases estimated from constant shear simulations. In Section 5 we show that the biases described above, due to blended sources with different applied shears, are present in these DES Y3 simulations. We then present a method to combine mean shear measurements from the simulations with estimated redshift distributions in order to jointly infer corrections to both the shear calibration and the redshift distributions. We apply this method to the DES Year 3 simulations, producing a parameterized model of these effects that can be used to interpret the DES Year 3 shear catalogs, which we describe in Section 6. We summarize and discuss directions for future work in Section 7.

2 QUANTIFYING SHEAR CALIBRATION BIASES FOR WEAK LENSING SHEAR STATISTICS

In the following, we describe our formalism and methodology of using image simulations to calibrate gravitational lensing measurements. To this end it is useful to distinguish *galaxies* from *detections*. We take a galaxy to be emitting light *of fixed redshift z*, with a particular surface brightness profile local to a position θ on the sky. A detection on the other hand is, simply put, a thing identified by an algorithm designed to detect and deblend astronomical sources, such as that employed by SEXTRACTOR (Bertin & Arnouts 1996). Due to blending, measurements made on a detection may be affected by light from multiple galaxies or stars and therefore multiple redshifts. We only have access to detections in an imaging survey, and we measure statistics averaged over ensembles of detections. We thus must determine how effects such as blending affect weak lensing shear statistics of ensembles of detections, rather than individual galaxies.

In the following, we will work with the simplest such statistic, $\bar{g}^{\rm obs}(\theta)$, the mean measured shear over all detections within some pixel at some angular position θ. This is sufficient for our purposes since more cosmologically interesting 2-point statistics can be written as functions of such pixel-averaged mean shears, allowing the straightforward propagation of shear calibration biases, at least in the case that spatial correlations of the biases can be neglected (see e.g. Kitching et al. 2019 for an investigation of higher-order effects from relaxing this assumption). We assume there is some

true, redshift-dependent shear field $g^{\rm true}(\theta, z)$. We can assume that in the weak lensing regime, where contributions of order $(g^{\rm true})^2$ or higher can be neglected, the mean measured shear is some linear function of the true shear field,

$$\bar{g}^{\rm obs}(\theta) = \int_0^\infty dz \, n_\gamma(z) g^{\rm true}(\theta, z) + c + \eta, \tag{3}$$

where c is an additive bias, and η is some measurement noise (due to e.g. the intrinsic shapes of galaxies, or pixel noise, that averages to zero over many such measurements). In the following we will often drop η for compactness, such that $\bar{g}^{\rm obs}(\theta)$ is really the expectation value of the mean measured shear. The response of the mean measured shear to the true shear field, which has traditionally been described as the multiplicative bias, m, is now described by the function $n_\gamma(z)$, which we call the *effective redshift distribution for lensing*, or *effective redshift distribution* for short. In the next sections, we consider what this function is in idealized conditions, and how to estimate it from image simulations under less idealized conditions.

2.1 Isolated galaxies

It is useful to begin by considering the simplified case where each detection corresponds to a single galaxy. The image of each galaxy, and thus each detection, is subject to a gravitational shear field $g^{\rm true}(\theta, z)$, which depends only on its position θ in the sky and on its redshift z. Consider an ensemble of detections, e.g. a single photometric redshift bin in a gravitational lensing analysis. If each detection's shear is measured without bias, and each detection contributes to the mean with equal weight, then we have

$$\bar{g}^{\rm obs}(\theta) = \int_0^\infty dz \, n(z) g^{\rm true}(\theta, z). \tag{4}$$

Here $n(z)$ is the derivative of the number of detections per unit area with respect to redshift

$$n(z') \propto \left. \frac{dN}{dz} \right|_{z=z'} \tag{5}$$

and is normalized to unity, such that no denominator is required in equation 4. Hence in this case, as expected the effective redshift distribution is simply what is usually called the *redshift distribution* i.e. $n_\gamma(z) = n(z)$. One can estimate $n(z)$ as a finely sampled histogram of redshifts of individual galaxies associated with the detections,

$$n(z)\Delta z \approx \frac{\text{\# of galaxies with } z < z_j < z + \Delta z}{N}, \tag{6}$$

where N is the total number of galaxies, or

$$n(z) \propto \sum_j \delta(z - z_j). \tag{7}$$

2.2 Isolated galaxies with shear measurement biases

Now suppose that for detections in the shear catalog (which we assume still have a one-to-one correspondence to galaxies) at redshift z, there is a mean multiplicative bias ($\bar{m}(z) \neq 0$), or equivalently non-unity mean response to shear ($\bar{R}(z) \neq 1$). This could be due to any imperfect correction of the biases discussed in Section 1 in the measurement of observed shear. The mean measured shear is now given by

$$\bar{g}^{\rm obs}(\theta) = \int_0^\infty dz \, \bar{R}(z) n(z) g^{\rm true}(\theta, z). \tag{8}$$
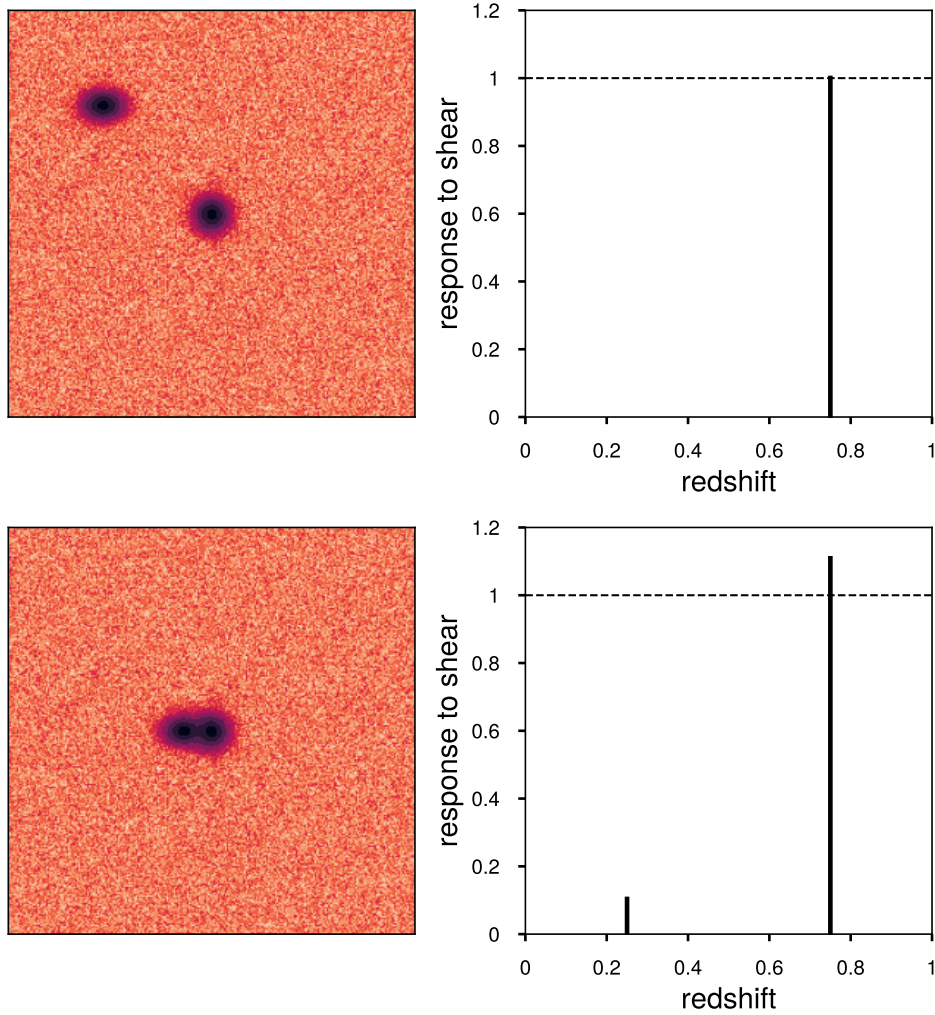
Figure 1. Simple example of the interplay between blending, shear calibration, and photometric redshift distributions. In both rows, the left panel shows an image of a pair of simulated galaxies, with the central object at higher redshift, $z = 0.75$, and a neighboring object at low redshift, $z = 0.25$. The right-hand panels show the response of the shape measurement of the central ($z = 0.75$) object, to applied shear, as a function of the redshift at which that shear is applied. In the top-row, where the objects are not blended, the the high-redshift object only responds to a shear at its own redshift, and since we use METACALIBRATION shear estimation, the response is unity (i.e. the shear estimation is unbiased) to very high precision. In the bottom right panel, we show the response when the objects are blended. We see two effects here. First, due to blending, the METACALIBRATION estimator is now biased (the right peak at z=0.75 has non-unit height). Second, the high-redshift object now responds to input shears at other redshifts (the small peak at z = 0.25). We show in this work that these responses define the effective redshift distribution for lensing predictions, in addition to quantifying the multiplicative bias of the measurements.

The effective redshift distribution is now given by

$$n_\gamma(z) = \bar{R}(z)n(z) \propto \bar{R}\frac{\mathrm{d}N}{\mathrm{d}z} \tag{9}$$

and accounts not only for the fraction of galaxies in an ensemble at a given redshift, but also for how sensitive to shear our measurements of the shapes of those galaxies are (this will depend on the shape measurement method). For example, all galaxies at redshift z' could be unresolved. Our catalog could contain some number of detections corresponding to them, whose shapes (and thus the shear they are subject to) we cannot measure. The mean observed shear of our catalog (or any other weak lensing statistic) would not respond to any true shear applied to light from z'. Thus, while these galaxies contribute to $dN/dz|_{z'}$, they should not contribute to $n_\gamma(z)$.

The METACALIBRATION method provides estimates of the shear response for each detection, so it is straightforward to generalize

equation 7 to the case of non-unity response:

$$n_\gamma(z) = \frac{\sum_j R_j \delta(z - z_j)}{\sum_j R_j}, \tag{10}$$

where R_j is the METACALIBRATION shear response for detection j, and the use of $\sum_j R_j$ in the denominator ensures $n_\gamma(z)$ is normalized to unity.

2.3   Blended galaxies and the general case

The definition of $n_\gamma(z)$ of equation 9 seems natural and has been used not only in DES Year 1 (Hoyle & Gruen et al., 2018), but in some form by most other weak gravitational lensing analyses to date. It is accurate under the assumptions made – that each detection corresponds to light from a single redshift, and that while response to shear may be imperfect, it can be described by its mean as a

function of the redshift, or estimated based on some properties of the images of actual detections, each associated with a single redshift.

Let us now consider the less idealized, ubiquitous case of blending. Here, a detection is not always associated with a single galaxy. Rather, measurements on a single detection may be influenced by the light of multiple galaxies at different redshifts along the line of sight. The observed shape may respond to shear applied to any of these multiple galaxies. It is not clear which galaxy's redshift to associate with the detection in order to estimate its contribution to dN/dz and $\bar{R}(z)$. While methods exist for inferring distinct redshift components in blended systems (Jones & Heavens 2019; Padmanabhan et al. 2019), in theory allowing the assignment of a detection to multiple redshifts, we still would not know the relative weight to assign to each redshift component in the ensemble $n(z)$. For a detection j, with components k at redshift z_k, the correct weight R_j^k is the linear response of the measured shape of the detection, $\bar{g}_j^{\text{obs}}$, to a shear applied to component k only, or equivalently at redshift z_k only i.e.,

$$R_j^k = \frac{\partial \bar{g}_j^{\text{obs}}}{\partial g(z_k)}. \tag{11}$$

The effective redshift distribution $n_\gamma(z)$ for the ensemble would be formed by summing over components k for each detection j i.e.

$$n_\gamma(z) = \frac{\sum_j \sum_k R_j^k \delta(z - z_k)}{\sum_j \sum_k R_j^k}. \tag{12}$$

Here R_j^k is a generalization of the response used by Metacalibration, which is the response of the measured shape of a detection to a constant (i.e. redshift independent) shear. Unfortunately, unlike the Metacalibration response which can be computed numerically by applying an artificial shear to galaxy images, we cannot measure this generalization on real galaxy images, since it would require deblending the redshift components in order to shear them individually, which is impossible to do perfectly.

In the presence of blending, we can thus no longer assume a separable effective redshift distribution $n_\gamma(z) = \bar{R}(z)n(z)$. Blending however does not invalidate equation 3, which assumes only that in the weak lensing regime, the observed mean shear $\bar{g}^{\text{obs}}$ can always be approximated as some linear function of the true shear field $g^{\text{true}}(z)$. Equation 3 in fact provides a definition of $n_\gamma(z)$ via a functional derivative

$$n_\gamma(z) = \frac{\mathrm{d}\bar{g}^{\text{obs}}}{\mathrm{d}g^{\text{true}}(z)}. \tag{13}$$

One can also define $n_\gamma(z)$ in the limit $\Delta z \to 0$,

$$n_\gamma(z)\Delta z = \frac{\Delta\bar{\gamma}^{\text{obs}}}{\Delta\gamma^{\text{true}}(z, z + \Delta z)}. \tag{14}$$

We will later also use $N_\gamma(z_1, z_2)$, the integral of $n_\gamma(z)$ over some redshift interval (z_1, z_2),

$$N_\gamma(z_1, z_2) = \int_{z_1}^{z_2} \mathrm{d}z\, n_\gamma(z). \tag{15}$$

This is the response of the mean measured shear of the ensemble $\bar{\gamma}^{\text{obs}}$ to an applied shear $\Delta\gamma^{\text{true}}$ in redshift interval (z_1, z_2).

Equation 13 and equation 14 constitute *definitions* of the effective redshift distribution that should be used to describe a sample of detections in a weak gravitational lensing analysis e.g. for predicting shear correlation functions or tangential shear signals. Note

that the normalization of $n_\gamma(z)$, i.e. $\int_0^\infty dz\, n_\gamma(z)$, is now meaningful. The normalization is the response of the mean measured shear to a true shear that is constant in redshift. This corresponds to the traditional $1+m$, where m is the mean multiplicative bias for the ensemble. Note that numerical codes for making theoretical predictions of weak lensing statistics often internally normalize the provided redshift distribution. In this case one would need to apply the normalization of $n_\gamma(z)$ to the predicted statistic as an additional correction.

The use of $n_\gamma(z)$ unifies two effects: biases in photometric redshifts and shear calibration, that have traditionally been treated separately. As we enter the era of deeper galaxy surveys, where blending becomes more and more important, we no longer have the luxury of treating these two systematic effects separately.

2.4 Calibrating n_γ with image simulations

The definition of $n_\gamma(z)$ in equation 9 is superseded by that of equation 13. In the presence of blending, the two are expected to disagree, not just in their normalization (i.e. as an overall multiplicative bias), but also in their shape.

Photometric and/or clustering-based redshift calibration, can at best (when combined with a Metacalibration response estimate for each detection) aim to measure $\bar{R}(z)n(z)$. Image simulations must be used to check whether the difference between $\bar{R}(z)n(z)$ and the true $n_\gamma(z)$ is small, as one might hope as long as blending is rare or mild. In case that difference is not negligible, image simulation methods can be used to infer corrections to $\bar{R}(z)n(z)$ that improve its agreement with the true $n_\gamma(z)$; this is the approach taken in Section 5 of this work.

Equation 14 allows us to operationalize the updated definition of n_γ. From suitable image simulations we can measure $N_\gamma(z_1, z_2)$, that is, the integral of $n_\gamma(z)$ over some interval α between (z_1^α, z_2^α), in which we vary the applied shear. To this end, we need to run a separate image simulation for each interval we would like to study, with a differential shear $\Delta\gamma_\alpha^{\text{true}}$ applied to galaxies within that interval. We will use more compact notation in the following, with N_γ^α denoting the integral of $n_\gamma(z)$ over redshift interval α,

$$N_\gamma^\alpha \equiv N_\gamma(z_1^\alpha, z_2^\alpha) \tag{16}$$

$$\equiv \int_{z_1^\alpha}^{z_2^\alpha} \mathrm{d}z\, n_\gamma(z). \tag{17}$$

This is estimated from our simulations via

$$N_\gamma^\alpha = \frac{\Delta\bar{\gamma}^{\text{obs}}}{\Delta\gamma_\alpha^{\text{true}}}, \tag{18}$$

where $\Delta\bar{\gamma}^{\text{obs}}$ is the change in mean measured shear of the ensemble.

Thus far, we have only considered a single ensemble of detections, and its effective redshift distribution for lensing, $n_\gamma(z)$. In Section 5, we will estimate these quantities for multiple subsets of our simulated detections, for example true or photometric redshift bins. In this case, we assign these subsets a label i, and denote as $\bar{g}_i^{\text{obs}}$, $n_{\gamma,i}(z)$ and $N_{\gamma,i}^\alpha$, the mean measured shear, effective redshift distribution and integrated effective redshift distribution for the subset i of our detections.

In the absence of blending, and if there were no overlap between the redshift interval α, and the redshift range of galaxies contributing to detections in subset i, $\Delta\bar{\gamma}_i^{\text{obs}}$ and thus $N_{\gamma,i}^\alpha(z)$ would vanish i.e. there would be no response of the mean shear for subset i to shear in redshift interval α. Blending, however, causes a non-zero response for any realistic selection of an ensemble i of detections. This is

because there will always be some amount of blending between galaxies in redshift interval α, and the detections in i, and shear applied to the former will impact the measurement of the shapes of the latter.

3   DES Y3 IMAGE SIMULATIONS

Having introduced and motivated our formalism for quantifying observational biases, we now turn to describing and validating our suite of DES Year 3-like image simulations. We note again here that DES Year 3 refers to the first three years of DES data processed together. In our simulation design we follow closely the real DES Year 3 data, by simulating complete sets of single-epoch images required to form DES Year 3 *tiles* in all four photometric bands *griz*, and then applying the same software for coadding, object detection and object measurement as is applied on the real data in Sevilla-Noarbe et al. (2021) and Gatti & Sheldon et al., (2021). This consistent simulation of weak lensing data across multiple photometric bands is key a step forward in the Dark Energy Survey's joint shear and photo-z bias characterization. We describe the main steps in our fiducial simulation pipeline below.

3.1   Exposures, single-epoch images, and tiles

DES wide-field images are processed in *tiles*, square sky regions of side length $10,000$ pixel (≈ 0.72 degree, see Morganson et al. 2018). There are 10,338 such tiles included in the Y3 dataset. For each tile region, all images overlapping that region are included in a coadded image for each band. These input images to the coadd come from many *exposures*, each of which is a collections of images, one for each CCD in the Dark Energy Camera (DECam, Flaugher et al. 2015) focal plane. We refer to these individual CCD images as *single-epoch images*. We also organize our image simulation using this tiling system.

We select at random 400 of the tiles entering the Y3 dataset, and for each tile, generate simulated versions of all the single-epoch images with any overlap of the tile region. For a given version of our simulation, this constitutes 62.8 million simulated objects, of which 15.4 million are detected in our shear pipeline, and 4.1 million pass shear catalog quality cuts (see Section 3.4). We simulate a total of 20 versions of this 400 tile set simulation, with versions differing only in their applied shear field (see Section 3.2), or whether objects are placed on a regular grid and whether detection is performed (see Section 3.5). We were limited in producing more simulation volume by time and computing resources, but find that this 400 tile set is sufficient volume such that statistical uncertainties in the simulation measurements are not the dominant uncertainty on our inferred bias corrections.

3.2   Single-epoch Image generation

Our simulation pipeline starts by generating a simulated version of each single-epoch image using GALSIM[5] (Rowe et al. 2015), with the addition of various custom modules. Briefly, we create simulated images with noise, point-spread-functions and world-coordinate system (WCS) estimated from the DES Y3 data, and insert parametric models for stars and galaxies. The images are simulated with the following properties:

- **Pixel geometry:** Each simulated single-epoch image is generated with the pixel geometry of the corresponding image in the real DES data. This is simply set by the DECam CCD properties, all of which have 4096×2048 pixels.

- **Noise:** Noise is assumed to be Gaussian and is drawn from the weight maps estimated for the corresponding image in the real DES data. For pixels that are masked in the corresponding image, the median of the weight map is used as the inverse noise variance. This noise field constitutes the only background onto which simulated objects are drawn — we do not simulate e.g. non-zero sky background. This choice implicitly assumes that the background subtraction performed on the Y3 data is sufficiently accurate.

- **WCS:** Our input objects to the simulation, for which we generate a model in sky coordinates, are consistently drawn into all the single-epoch images they overlap, using the WCS (world-coordinate system) solution for the the corresponding image in the real data.

- **PSF:** The drawn objects are convolved with a smoothed version of the PSF model estimated from the real data by Piff[6] (Jarvis et al. 2021). See Appendix C for the details and justification of this procedure. A significant simplification of our analysis here is that we do not attempt to measure the PSF from our simulations for use in shape measurement, rather we use the input PSF models. While we believe this is well justified by the stringent PSF validation performed in Jarvis et al. (2021), jointly simulating the inference of PSF and shear would be a natural extension of the work presented here.

- **Masking:** We package with the simulated images the bad pixel masks taken from the corresponding real data image files. These indicate pixels to exclude or interpolate for downstream processing and measurement codes.

- **Input galaxies:** We randomly draw galaxy models from a catalog of "bulge+disk", two-component parametric fits to galaxies in the COSMOS field[7]. This catalog is described in Hartley & Choi et al., (2020); we provide a brief description here. The morphological parameters (half-light-radius, bulge-to-disc-ratio, ellipticity) of the parameteric galaxy models were fit to Hubble Space Telescope Advanced Camera for Surveys (HST-ACS) imaging (Scoville et al. 2007; Koekemoer et al. 2007), specifically we use the re-processed HST-ACS data of Leauthaud et al. (2007). These model fits were then used to estimate fluxes for the DECam *griz* filter bandpasses, using forced photometry at the same sky positions in deep stacks of DECam imaging (described in Hartley & Choi et al., 2020). The use of both HST imaging and DECam imaging gives us a catalog of parametric galaxy models with realistic and well-constrained morphology (from HST) and realistic fluxes and colors in the DES filters (from the deep DES imaging). When drawing a parametric model galaxy into the simulations, we apply a random rotation to each simulated object.

We additionally match this catalog (using a 0.75 arcsecond matching radius) to the Laigle et al. (2016) redshift catalog, and apply area masks for the problematic regions identified by Hartley & Choi et al., (2020), after which 222116 unique input objects remain. We remove 60 (0.03%) of these for which the parametric model fits failed. We additionally apply a selection $r_{50} > -0.25(\mathrm{mag}_i - 22) - 1.35$, which we find effectively removes the stars (we separately simulate stars using a different catalog described below).

In our fiducial simulation we include only galaxies with i-band magnitude < 25.5, which is two magnitudes fainter than our thresh-

---

old for inclusion in the eventual shape catalog. While we do not perform an analysis of sensitivity to this choice, the results of e.g. Hoekstra et al. (2017) suggest the absence of fainter objects than this in our simulations will probably bias our multiplicative bias estimates at the $\sim 0.1\%$ level, well below our current uncertainties.

In each tile, the number of galaxies simulated is drawn from a Poisson distribution with mean 170000, which corresponds to a number density of 88 galaxies per square arcminute. The density of input objects was tuned such that the number of detected objects in the simulations matched the number of detected objects for the same set of tiles in the real DES Y3 data. The majority of these galaxies will be either undetected or removed via cuts, resulting in a number density of roughly 6 galaxies per square arcminute used for shear estimation.

Galaxies are placed randomly on the sky and so are not clustered. Compared to the real Universe, we expect this to result in less blending of objects at similar redshifts. We discuss the implications of this approximation in Sections 6.3 and 7.

• **Input stars:** We use a catalog of stars simulated using the Trilegal[8] code, with the best-fitting models from the *MWFitting* method presented in Pieres et al. (2020). This catalog contains both sky positions and fluxes for a simulated population of stars complete in the magnitude range 14-26 in the *g*-band. In our fiducial simulation we include only stars with *i*-band magnitude < 25.5.

• **Input shear:** We run otherwise identical realizations of each simulation with constant input shears of either +0.02 or −0.02 in a single shear component (the other component being 0). This allows us to follow the approach of Pujol et al. (2019), who propose computing multiplicative shear biases via measuring the difference in recovered shear between two simulations which are identical apart from a small change in input shear. Using this procedure greatly reduces the noise (both shape noise and measurement noise) on the multiplicative bias estimate, since much of it cancels when taking the difference. Note this is closely related to the idea of the 'ring-test' introduced by Nakajima & Bernstein (2007). We additionally generate simulations where the applied shear depends on the redshift of the input galaxy. More specifically, we generate simulations where galaxies with redshift within some redshift interval α have a difference in applied shear of $\Delta g_\alpha^{\mathrm{true}} = 0.04$ with respect to the rest of the simulated galaxies. This allows us to measure the N_γ^α as described in Section 5.1.

3.3 Image reduction, coaddition and detection

The processing and measurements applied to the simulated images closely follows that performed on the real DES Y3 images (described in Morganson et al. 2018; Sevilla-Noarbe et al. 2021; Gatti & Sheldon et al., 2021). We summarize here:

• A 10000×10000 pixel weighted-mean coadd image is generated for each tile, in each band, using SWarp (Bertin et al. 2002). The weight maps used are the same as those used to generate the noise on the simulated images. The *r*, *i* and *z* coadd images are then themselves combined in a CHI-MEAN coadd image, again using SWarp.

• The *riz* coadd is used for object detection and segmentation, which is performed by SExtractor (Bertin & Arnouts 1996). SExtractor outputs a catalog of detected objects, various measured quantities for these objects, as well as segmentation maps (which

---

[8] http://stev.oapd.inaf.it/cgi-bin/trilegal

indicate which pixels in the images are assigned to which catalog objects).

• Multi-epoch data structure (MEDS, Jarvis et al. 2016) files are generated for each band for each tile. For each detected object in the SExtractor catalog, the MEDS file contains a postage-stamp cutout from each of the single-epoch images in which that object appears. This data format makes convenient the fitting of models simultaneously to multiple observations of a given object.

3.4 Shear estimation

To generate a shear catalog for each tile, we run Metacalibration on the *r*, *i* and *z* MEDS files, which fits an elliptical Gaussian profile, convolved with the PSF model, to the observed light profile of each detection. The parameters of the profile are fit jointly to square regions (*stamps*) extracted from each single-epoch image for all bands, apart from a free amplitude allowing an independent flux in each band. Stamps with a masked fraction of more than 0.1 are not used in the fits. Estimates of the shear response, R_{ij} are also generated by Metacalibration which is the response of the measured shear in component *i* to an applied shear in component *j*. From the shear catalogs generated by Metacalibration we select a sample suitable for weak lensing measurements by applying the identical catalog cuts as those applied to the DES Y3 data in Gatti & Sheldon et al., (2021). The most significant (in terms of number of objects removed) of these are cuts on the object signal-to-noise ratio, $S/N > 10$, and the ratio of PSF-deconvolved galaxy size, T, to the PSF size, T_{psf}, $T/T_{\mathrm{psf}} > 0.5$. T is an area measure (equal to the trace of the covariance) for the Gaussian profile. The signal-to-noise ratio cut is required to minimize biases associated with shear-dependence of the SExtractor selection, and the size cut reduces the impact of any PSF modeling errors. See Gatti & Sheldon et al., (2021) for more discussion of the motivation and details for the shear catalog cuts, as well as detailed descriptions of the quantities such as T. This sample can then be used to estimate the shear recovered from the simulation, which can then be compared to the true shear input to the simulation to estimate any biases in the shear recovery.

3.5 Simulation variants

In order to better understand the source of shear calibration biases, we generate and analyse two sets of simulations additional to the `fiducial` simulation described thus far. In the `grid` simulations, objects are placed on a regular grid with spacing ≈ 35 pixels (≈ 9 arcseconds). We would expect any biases related to blending to be absent in this variant. The second variant is the `grid-truedet` simulations, which again places objects on this regular grid, and in addition the SExtractor detection catalog used as input to the shape measurement is replaced by a catalog containing the true positions of input objects. We would expect this to remove any biases related to possible shear dependence of the SExtractor detection probability (e.g. if rounder objects were more likely to be detected).

3.6 Simulation Validation

The final part of our simulation pipeline is validating our simulation's realism by comparing it to the real data. We infer below that the shear biases present are related to the blending of sources and the possible shear-dependence in how sources are selected, segmented and modelled. This blending must therefore be accurately

characterized in the simulations. We ensure here that the number density of sources, the noise levels in the image, and the distribution of measured source properties like flux and size are well-matched between the fiducial simulation and real DES data. We expect the effects of blending to be sensitive to the properties of neighbouring objects (such as the number of them within a given distance, and their brightness) encountered by our target source galaxies, so we additionally study statistics sensitive to these in Section 6.3.

3.6.1   *Aesthetics*

We generate *gri* color images of our simulated coadds, using the `desimage`[9] package. These are useful for visual inspection of the simulations, but also for visual comparison to images made via the same process on the real data. Figure 2 shows color images for the same 1000x1000 pixel region of the coadd tile `DES0003-3832` in the real data (left panel) and the fiducial simulation (right panel). A lack of bright stars in the simulation is apparent. These are masked out of the real data before analysis, so we do not believe that their absence from the simulations can impact our results.

3.6.2   *SEXTRACTOR comparisons*

SExtractor forms a crucial part of our pipeline in detecting and segmenting objects in the coadd images, so it is important that the simulated images analysed using SExtractor resemble the real data. Here, we perform comparisons of the measured properties of objects detected in the simulations.

The top-left and top-middle panels of Figure 3 shows the joint distribution of magnitude and size estimated by SExtractor, for the *g* and *i*-bands respectively. Specifically, we use `MAG_AUTO` (an elliptical aperture magnitude), and `FLUX_RADIUS` (with `PHOT_FLUXFRAC = 0.5`) an estimate of the PSF-convolved half-light radius of the object[10]. Some clear features are apparent in both simulations and data, such as the shift of the distribution to larger size in the *g*-band compared to the *i*-band, which is due primarily to the larger PSF. We note that the distributions plotted share the same normalization, so any differences in absolute number density would be apparent.

3.6.3   *METACALIBRATION quantities*

For shear estimation and photometry, we use elliptical Gaussian models fit to single-epoch images (as opposed to the coadds used for the SExtractor quantities discussed above). The distributions of size and signal-to-noise of a weak lensing galaxy sample have often been identified as key to the expected level of bias in the shear recovery (e.g. Refregier et al. 2012; Kacprzak et al. 2012). The top-right panel of Figure 3 compares the joint distribution of signal-to-noise and size between simulations and data. The distributions are smoother in the log of these quantities, so we actually compare distributions of $\log_{10}(S/N)$ and (since T can take slightly negative values) $\log_{10}(1 + T)$.

As well as size and signal-to-noise, it is important to verify that other quantities used to select sub-samples of the source galaxies are

well matched between simulations and data. A key example of this is selection of objects into bins based on their photometric redshift (photo-*z* henceforth). Photo-*z* algorithms usually use some discrete set of broadband fluxes, with ratios of those fluxes (or differences in magnitudes), known as *colors* considered particularly informative for redshift estimation. The bottom three panels of Figure 3 demonstrate that the simulations reproduce well the joint distributions of (from left to right): *i*-band magnitude and $r - i$ color, *i*-band magnitude and $i - z$ color, and $r - i$ color and $i - z$ color.

3.7   Photo-z inference

The photometric redshift distributions for the simulation sample are inferred using the *SOMPZ* method outlined in Myles & Alarcon et al., (2021) (see also Buchs & Davis et al., 2019). Following the methodology applied to DES Y3 data there, detected objects in the image simulations are assigned to two self organizing maps (SOMs), corresponding to "wide" (32 x 32 cells, using *riz* flux information) and "deep" (64 × 64 cells, using *ugrizJHK$_s$* colour information from the DES Deep Fields measurements presented in Hartley & Choi et al., 2020). In the Y3 data anaysis, galaxies with spectroscopic redshifts or flux measurements in a large number of photometric bands ("deep" galaxies) are assigned to the higher resolution SOM. When those same galaxies are assigned to the lower resolution SOM based on their wide field flux information, they can be used to infer information about other galaxies with only the limited wide-field photometry available in order to calibrate the color-redshift relation.

We use the SOMs constructed from the Y3 data catalogs and the same software to assign the simulated galaxies (Myles & Alarcon et al., 2021). We assume that all simulation input galaxies have precise redshifts (equal to the photo-*z* point estimates by Laigle et al. 2016) that, together with their measured deep *ugrizJHK$_s$* fluxes, describe the deep color-redshift relation. Matching simulation detections to input galaxies allows us to generate a transfer function that connects redshift, deep SOM cell, and wide SOM cell. The wide cells and their contained galaxies are then grouped together by their mean redshift in a variety of ways (see below) to form tomographic bins.

The procedure for matching injection galaxies to measured detections in these simulations begins with a nearest neighbors search to identify the three closest objects in the truth catalog to a given detection. For a detection to be matched to a true object, it must have a close match within two pixels. If it does not have one, it is ignored in the rest of the photo-*z* inference (roughly 0.5% of detections). Detections with an exclusive close match (i.e. one that is not a close match to any other detection) make up approximately 30% of the simulated sample. To discriminate between detections with multiple close truth matches, we loop over injections from brightest to faintest in *i*-band magnitude and assign the brightest close truth match that has not yet been assigned to another detection. Roughly 70% of detections fall into this category. If all close matches have already been assigned to other detections via this loop, no truth match is assigned, but this happens rarely enough to be negligible. For example, this only occurs for 4 cases in the entire fiducial $(g_1, g_2) = (-0.02, 0.00)$ simulation.

The wide SOM occupancy for both the simulations and the data can be seen in Figure 4, showing very good general agreement. This is a consequence of the close alignment of color and magnitude distributions between the data and the simulations discussed in Section 3.6.2. The largest point of discrepancy is due to cells composed of very large, very blue galaxies at low redshifts, which likely
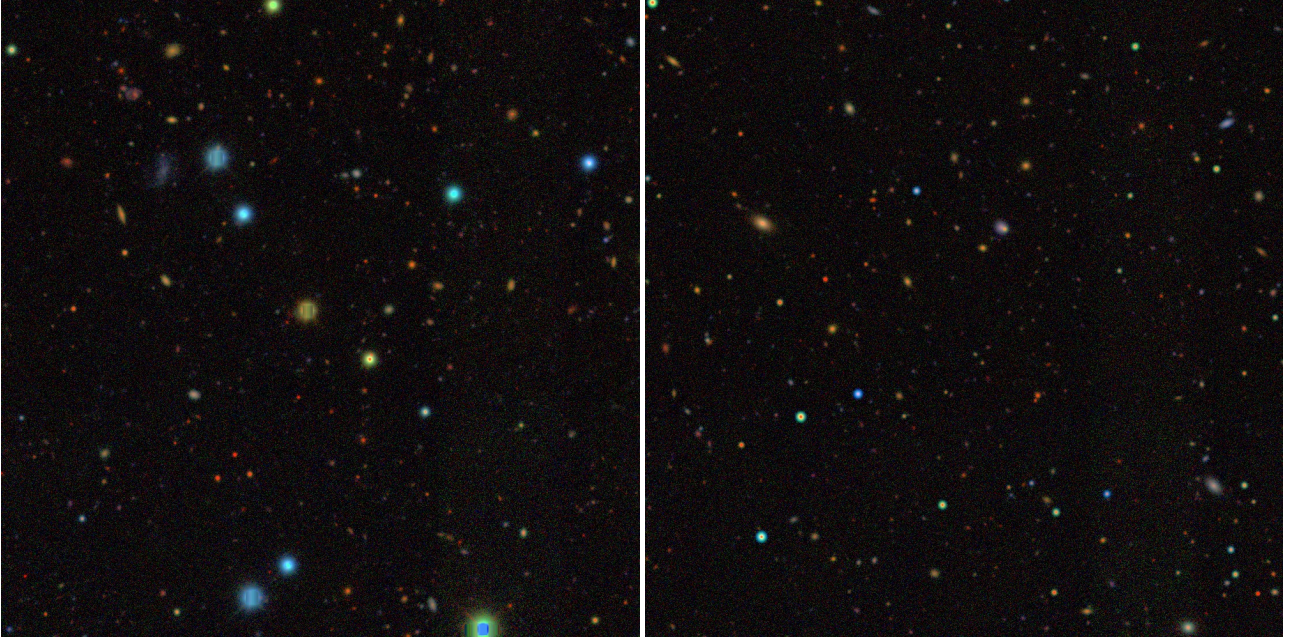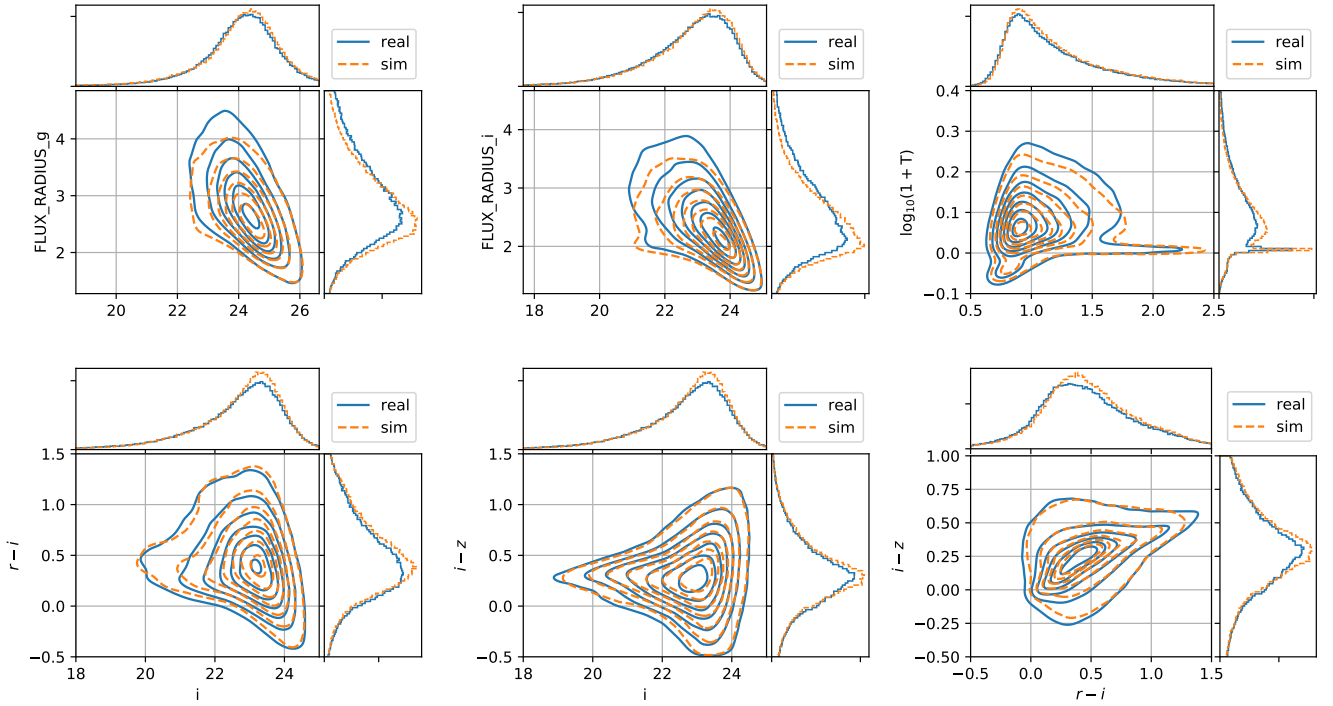

Figure 2. gri color image of 1000×1000 pixel region of tile DES0003-3832 for the real Y3 data (left panel) and the fiducial simulation (right panel). .

Figure 3. Comparisons of joint distribution of measured quantities between simulations (orange dashed lines) and DES Y3 data (blue lines). The top-left and top-middle panels show joint distributions of SExtractor measured quantities `MAG_AUTO` and `FLUX_RADIUS` g-band and i-band respectively. The stellar locus is at larger `FLUX_RADIUS` in the g-band because of the larger PSF. The remaining panels show comparison of quantities estimated by the shape measurement code (Metacalibration). The top-right panel shows the joint distribution of signal-to-noise and size (in fact, $\log_{10}(S/N)$ and $\log_{10}(1+T)$). The bottom-left panel shows the joint-distribution of i-band magnitude and $r-i$ color. The bottom-middle panel shows the joint distribution of i-band magnitude and $i-z$ color. The bottom right panel shows the joint distribution of $r-i$ and $i-z$ colors.

Sample	Binning Style	$\bar{z}_0$	$\bar{z}_1$	$\bar{z}_2$	$\bar{z}_3$
Data	Fiducial	0.334	0.517	0.749	0.936
Sim	Fiducial	0.311	0.460	0.723	0.894
Sim	Equal	0.322	0.510	0.745	0.920
Sim	Z-Match	0.325	0.511	0.753	0.930
Sim	W-Match	0.294	0.463	0.714	0.895

Table 1. Table describing the mean redshift per photo-z bin for each binning algorithm as compared to that found in the data.

do not significantly contribute to the redshift-dependent effects of blending that are key to this analysis.

A principal aim of this work is to quantify the shear calibration separately for each of the photometric redshift bins used in the DES Y3 cosmology analyses. This requires that we perform an equivalent photometric redshift binning on the image simulations. One can motivate several different choices of photo-z binning algorithms in order to estimate the shear calibration biases that may differ due to the small differences in color-redshift relation between the sims and the data. Presented here is a description of the four binning algorithms we use and their resulting summary statistics like the mean redshift per bin.

• **Fiducial**: We use the same mapping between wide SOM cell and photo-z bin as in the Y3 data. We make this our fiducial choice, since we think the philosophy of treating the real data and the simulations as similarly as possible is a sensible one. However, small differences in wide SOM occupancy between simulations and real data are apparent with this procedure. Whereas in the data, each photo-z bin has an equal number of objects (25% of the total, by construction), there are deviations from 25% occupancy in the simulations, with the four photo-z bins receiving 19%, 24%, 25% and 30% of the objects respectively.

• **Equal Count** (`equal`): Instead of taking the mapping from the data, the wide SOM cells are ordered by mean redshift and then grouped such that an approximately equal number of galaxies end up in each photo-z bin, much like how the mapping is chosen in the data.

• **Mean redshift matching** (`z-match`): This is a binning that chooses the wide SOM cells such that they closely match the mean redshift found in the data photo-z bins, with approximately equal counts in each bin. By necessity, in this binning scheme a galaxy may end up assigned to multiple bins, or not assigned to any bin.

• **SOM Occupancy Matching** (`w-match`): This preserves the mapping used in the fiducial case, but also reweights galaxies to reproduce the relative wide SOM cell occupancy found in the Y3 data, using the ratio between the left and center panels of Figure 4.

The mean redshift for each of these photo-z binning choices on the simulations are reported in Table 1, and the full distributions can be seen in Figure 5. We see broad similarity between the binning recipes, but revisit the alternative options in Section 6.2 where we test the sensitivity of our calibration corrections to the choice of binning recipe.

4   RESULTS I: SHEAR CALIBRATION BIASES FROM CONSTANT SHEAR SIMULATIONS

In this section, we begin by examining the shear calibration biases apparent in constant shear simulations, and we report average shear calibration bias estimates for the full DES Y3-like sample, as well as for individual photo-z bins. As described in Section 2, these bias estimates are not sufficient in general to correct theoretical

predictions for weak lensing shear statistics, hence in Section 5, we present estimates of biases in $n_\gamma(z)$, using simulations where the input shear varies with redshift.

For these calculations, we use four 400 tile constant shear simulations with input shear

$$(g_1, g_2) \in \{(0.02, 0), (-0.02, 0), (0, 0.02), (0, -0.02)\} \tag{19}$$

(see Section 3.2). Each simulation is identical (i.e. random elements of the simulation use the same random seeds) apart from the applied shear, reducing the noise on the differences between the mean shear measured between pairs of the simulations (see Pujol et al. 2019). See Table 2 for more details.

For each shear component μ, the multiplicative and additive biases, m_μ and c_μ are calculated by fitting the model

$$\bar{g}_x^{\mathrm{obs}} = (1 + m_x)g_x^{\mathrm{true}} + c_x \tag{20}$$

to the pair of simulations with $g_x^{\mathrm{true}} = \pm 0.02$, where $\bar{g}_x^{\mathrm{obs}}$ is the measured mean shear for component x. The uncertainty on $\bar{g}_x^{\mathrm{obs}}$, inferred from jackknifing over the simulated tiles, is included in the fit.

Table 3 contains mean multiplicative biases for our `fiducial`, `grid`, and `grid-truedet` simulations. In the `grid` simulation, blending is removed by placing down objects on a grid. In the `grid-truedet` simulation, objects are again placed on a grid, and in addition we use their true positions for the detection catalog, rather than the detection catalog estimated using SEXTRACTOR. We expect this to additionally remove any selection biases due to shear-dependence of the SEXTRACTOR selection.

While the fiducial simulations exhibit a mean multiplicative bias of $\approx -2\%$, the multiplicative bias for both the `grid` and the `grid-truedet` simulations are greatly reduced, with a remaining bias of around -0.4%. We can draw a few conclusions from this. Firstly, the fact that we see consistent biases for the `grid` and `grid-truedet` cases implies that we do not have significant SEXTRACTOR biases for isolated objects. This is not unexpected, since we apply a $S/N > 10$ cut to our catalogs that likely dominates over the threshold for detection used by SEXTRACTOR, and METACALIBRATION is able to accurately correct for selection biases due to this cut.

Beyond that, we can attribute most of the multiplicative bias we see in the fiducial simulation to the presence of blending. As explored in Sheldon et al. (2020), the presence of blending can likely generate biases through various (related) mechanisms. Firstly there is 'detection' bias due to shear-dependence of the detection algorithm, including the decision of how many different objects to assign to a blend. Secondly one may also expect an additional bias due to the presence of a neighbor in the shape measurement process, either due to its contaminating flux, or due to some unaccounted-for shear dependence of the neighbor masking algorithm. With these simulations, we cannot fully decouple these effects. A simulation with randomly-placed objects as in the fiducial simulation, but using true detection to remove detection biases, would shed light on the issue, but we did not have the resources to run this variation.

Figure 6 shows the multiplicative bias (averaged over shear component) as a function of signal-to-noise ratio. For the `grid` and `grid-truedet` cases there is no clear trend with S/N, while for the fiducial case, there do appear to be variations with S/N, although only at the percent level, and we do not attempt to explain this behaviour. We also do not have a conclusive explanation for the $\approx -0.4\%$ biases remaining in the `grid` and `grid-truedet` simulations. Appendix B explores various potential sources of small (sub-percent) multiplicative biases using idealized simulations, and we do see the potential for biases due to masking at the $\sim 0.1\%$, so
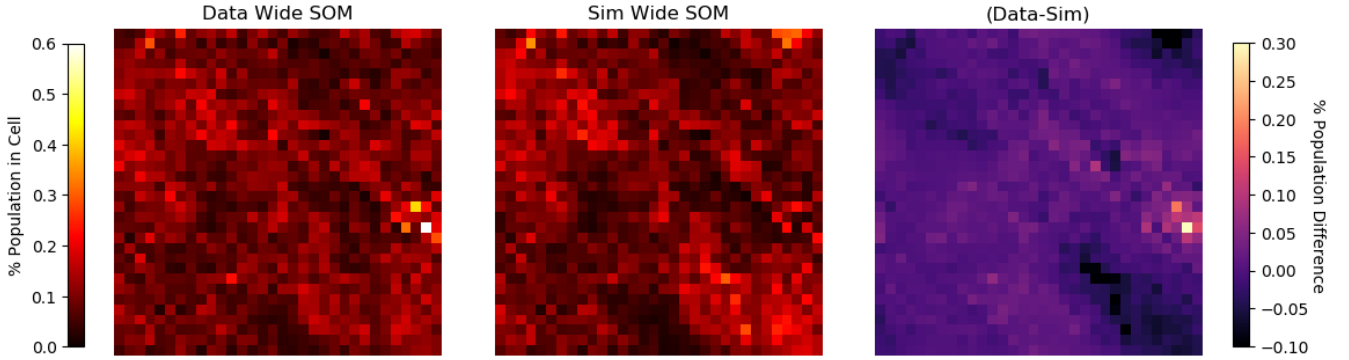
Figure 4. The wide SOM population distribution in the data (left) compared to the simulations (center), with a residual (right) that shows the majority of cells agree to within 0.05% in total population.
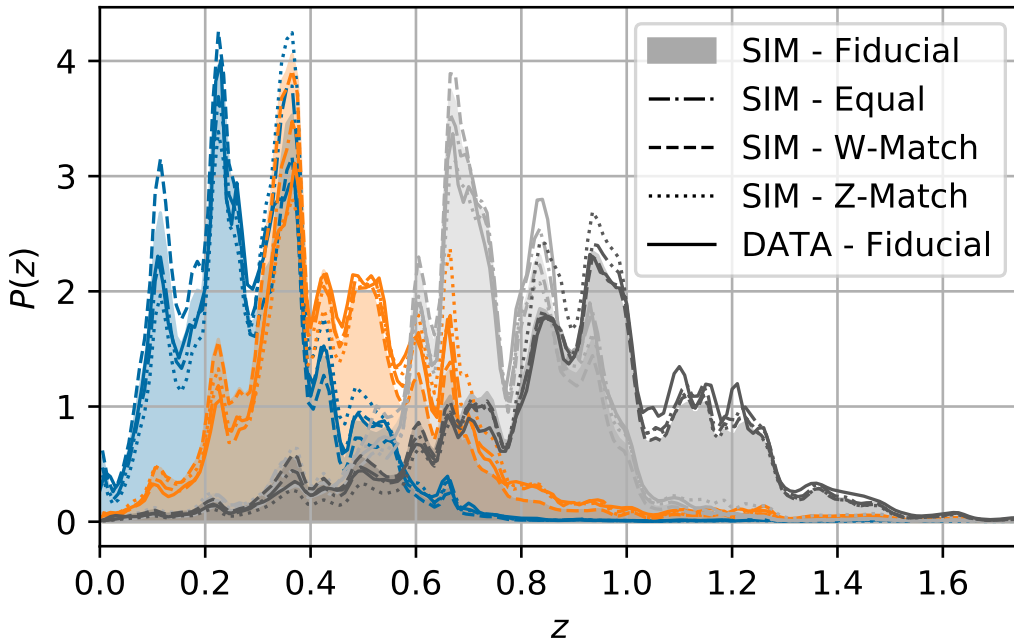

Figure 5. The photometric redshift distributions for the four photo-z binning schemes applied in the simulations (described in Section 3.7, as well as the estimated Y3 data redshift distributions (labelled "DATA - Fiducial"). For plotting purposes all curves have been smoothed using a Gaussian kernel of width 0.01 in z.

this may be contributing some of the remaining bias. Given that we directly apply the masks from the DES Y3 data to our simulations, we are confident that the biases in the simulations due to masking will be at a similar level to that present in the real data.

In order to provide some more detail on the impact of blending on the multiplicative bias, we attempt to compute the bias as a function of object separation as follows. Using the galaxies in the input truth catalogs for our simulations, we find pairs of input galaxies that are separated by some distance and are no closer to any other galaxy in the simulation than the other galaxy in its pair. We then define a small region in the simulated image encompassing each pair, such that detections in this region include light from just these two galaxies. Specifically, we used a circle of radius 0.75 times the pair separation, centered on the midpoint between the two galaxies.

Detections in these regions were used to compute the multiplicative bias as a function of separation. The results of this exercise are shown in Figure 7. For comparison, we show the predicted effects of detection bias from Sheldon et al. (2020) for a simplified simulation setup involving only pairs of galaxies with equal size and flux. Note also that Sheldon et al. (2020) used a different procedure to remove the light of neighboring objects. At small separations we see the qualitative effects of detection bias, that is a large negative bias. At large separations, the bias is consistent with zero within the statistical noise. At intermediate scales, we also see a positive bias that was not apparent in the Sheldon et al. (2020) simulation results. This can occur when a neighbor is not detected, so may only be significant when pairs with widely disparate fluxes are included.

To summarize, we see multiplicative biases in our simulations
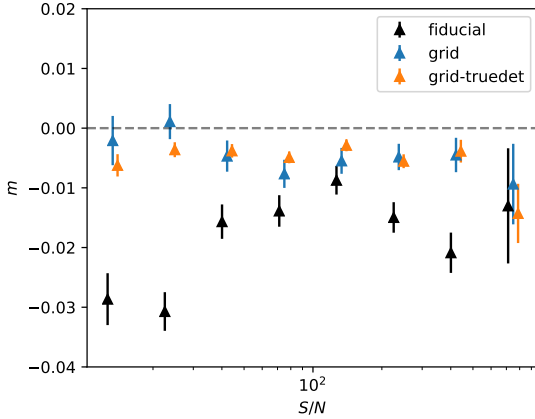
Figure 6. Multiplicative bias as a function of S/N for the `fiducial`, `grid`, and `grid-truedet` simulations. The `fiducial` simulation places the objects in the images at random positions and detects them with SEXTRACTOR. The `grid` simulation places them on a grid with ≈ 9 arcsecond spacing and still employs SEXTRACTOR. The `grid-truedet` simulation uses a grid for the objects, but uses their true positions instead of detecting them with SEXTRACTOR.

from at least three sources. First, as shown in Appendix B, the effects of masking and their corrections used in the DES Y3 analysis cause a small, few tenths of a percent bias. Second, for very close pairs of truth objects, we see a strong negative multiplicative bias in Figure 7 that is qualitatively consistent with the detection biases studied in Sheldon et al. (2020). At intermediate separations we see some positive multiplicative bias which is presumably blending-related, involving pairs with widely different fluxes.

We note here that the mean multiplicative bias of ≈ -0.02 we estimate here is somewhat different from the multiplicative bias prior of 0.012 ± 0.012 inferred by Zuntz & Sheldon et al., (2018) for the DES Year 1 shear catalog, which used very similar shape measurement methodology. We believe the significant improvements in simulation realism presented here explain this discrepancy. We note also that Sheldon et al. (2020), which used an independent set of image simulations to those presented here, that we believe also contained significant improvements with respect to those used by Zuntz & Sheldon et al., (2018), reported multiplicative biases much close to those presented here for DES-like simulations.

In Section 6.3, we use a re-weighting procedure to estimate systematic uncertainty (due to potential simulation inaccuracy) in the impact of blending, which is then propagated to our final priors on the shear calibration.

5 RESULTS II: : ESTIMATES OF $n_\gamma(z)$ BIASES FROM REDSHIFT-DEPENDENT SHEAR SIMULATIONS

We explained in Section 2 that for theory predictions of shear statistics involving ensembles of detections over a range of redshifts, we require an estimate of $n_\gamma(z)$, the effective redshift distribution for lensing. In this section we describe the extra simulations and methodology used to infer biases in the DES Y3 methodology for estimating $n_\gamma(z)$. To summarize, we take the following approach (with much more detail given in the following section). For each photo-z bin i:

• We measure $N_{\gamma,i}^\alpha$ for 4 redshift intervals α (with ranges
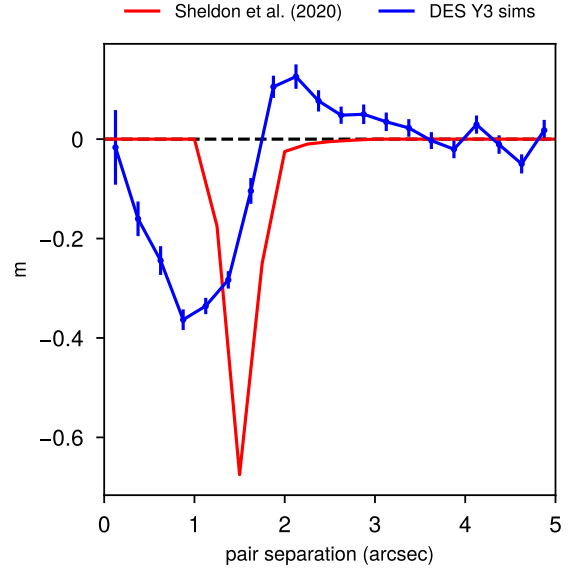
Figure 7. Multiplicative bias for detections corresponding to pairs of truth objects separated by a given angle. To make this plot, we find pairs of true objects separated by some distance in the truth catalog in our fiducial simulations and not near any other objects. We then use detected objects near this location to measure the multiplicative bias as a function of separation. At small scales, we find signals in our simulations that appear to be from detection bias. For comparison, we show the detection bias result from Sheldon et al. (2020) as the red line. At intermediate scales, we see a positive bias, which is probably from pairs where one source is much fainter than the other. The bias converges to be consistent with zero at large separations.

$(z_1^\alpha, z_2^\alpha))$, from simulations with a change in applied shear within that interval. We use redshift intervals $(z_1^\alpha, z_2^\alpha) \in \{(0.0, 0.4), (0.4, 0.7), (0.7, 1.0), (1.0, 3.0)\}$. Note these redshift intervals α have no specific correspondence to the four photo-z bins i we use.

• We compare this measurement to the prediction based on integrating the METACALIBRATION response-weighted redshift distribution, $n_\gamma^{\mathrm{mcal}}(z)$ (which is how $n_\gamma(z)$ is estimated in the real data).

• Due to computing limitations, we can only measure $N_{\gamma,i}^\alpha$ in the four coarse aforementioned redshift intervals. However, we need a finely sampled $n_\gamma(z)$ to make theory predictions. Therefore we use a model that parameterizes deviations from $n_\gamma^{\mathrm{mcal}}(z)$ of the form $n_\gamma^{\mathrm{model}}(z) = f(z)n_\gamma^{\mathrm{mcal}}(z) + g(z)$, where $f(z)$ and $g(z)$ are smooth functions of z. We fit this model to our $N_{\gamma,i}^\alpha$ measurement.

We start in Section 5.1 by describing the simulation inputs and procedure for estimating N_γ^α for a given ensemble of detections. In Section 5.2 we present the $N_{\gamma,i}$ estimates for the case without redshift binning, and compare these to the prediction from the METACALIBRATION response-weighted redshift distribution, $n_\gamma^{\mathrm{mcal}}(z)$. In Section 5.3 we present measurements of $N_{\gamma,i}^\alpha$ for ensembles of galaxies restricted to true redshift intervals, a case which most clearly demonstrates the response of galaxies assinged one redshift to a shear applied at another. In Section 5.4 we present measurements of $N_{\gamma,i}^\alpha$, for each photo-z bin i, again comparing these to the $n_\gamma^{\mathrm{mcal}}(z)$. In Section 5.5 we describe our modeling approach using fitting functions to infer $n_\gamma(z)$ corrections from these measurements, and finally in Section 5.6 present the resulting $n_\gamma(z)$ bias model constraints.

variant	sheared redshift interval	(g_1, g_2) in redshift interval	(g_1, g_2) outside redshift interval	object placement	SExtractor detection
`grid-truedet`	[0.0, 3.0]	(+0.02, 0.00)	–	grid	no
`grid-truedet`	[0.0, 3.0]	(−0.02, 0.00)	–	grid	no
`grid-truedet`	[0.0, 3.0]	(0.00, +0.02)	–	grid	no
`grid-truedet`	[0.0, 3.0]	(0.00, −0.02)	–	grid	no
`grid`	[0.0, 3.0]	(+0.02, 0.00)	–	grid	yes
`grid`	[0.0, 3.0]	(−0.02, 0.00)	–	grid	yes
`grid`	[0.0, 3.0]	(0.00, +0.02)	–	grid	yes
`grid`	[0.0, 3.0]	(0.00, −0.02)	–	grid	yes
`fiducial`	[0.0, 3.0]	(+0.02, 0.00)	–	random	yes
`fiducial`	[0.0, 3.0]	(−0.02, 0.00)	–	random	yes
`fiducial`	[0.0, 3.0]	(0.00, +0.02)	–	random	yes
`fiducial`	[0.0, 3.0]	(0.00, −0.02)	–	random	yes
`fiducial`	[0.0, 0.4]	(+0.02, 0.00)	(−0.02, 0.00)	random	yes
`fiducial`	[0.4, 0.7]	(+0.02, 0.00)	(−0.02, 0.00)	random	yes
`fiducial`	[0.7, 1.0]	(+0.02, 0.00)	(−0.02, 0.00)	random	yes
`fiducial`	[1.0, 3.0]	(+0.02, 0.00)	(−0.02, 0.00)	random	yes
`fiducial`	[0.0, 0.4]	(0.00, +0.02)	(0.00, −0.02)	random	yes
`fiducial`	[0.4, 0.7]	(0.00, +0.02)	(0.00, −0.02)	random	yes
`fiducial`	[0.7, 1.0]	(0.00, +0.02)	(0.00, −0.02)	random	yes
`fiducial`	[1.0, 3.0]	(0.00, +0.02)	(0.00, −0.02)	random	yes

Table 2. Image simulation properties. Each line lists a specific image simulation produced for this work. The `fiducial` simulation places objects at random into the image and uses SEXTRACTOR for object detection. We also ran variations for some of the simulations where objects were placed on a grid (`grid`) or we used their true locations for shear measurements (`grid-truedet`).

variant	$m_1 \times 100$	$m_2 \times 100$	$m \times 100$	$c_1 \times 10^4$	$c_2 \times 10^4$
`grid-truedet`	−0.48 ± 0.07	−0.40 ± 0.07	−0.44 ± 0.05	−2.06 ± 1.48	−2.77 ± 1.46
`grid`	−0.33 ± 0.15	−0.35 ± 0.15	−0.34 ± 0.11	−2.14 ± 1.34	−3.59 ± 1.40
`fiducial`	−2.23 ± 0.16	−1.93 ± 0.17	−2.08 ± 0.12	−0.86 ± 1.35	−1.34 ± 1.45
`fiducial` bin 0	−1.55 ± 0.44	−0.96 ± 0.45	−1.25 ± 0.31	1.57 ± 3.05	−4.38 ± 2.75
`fiducial` bin 1	−1.77 ± 0.56	−1.87 ± 0.54	−1.82 ± 0.39	−1.86 ± 2.66	−0.69 ± 2.79
`fiducial` bin 2	−2.51 ± 0.64	−2.03 ± 0.67	−2.27 ± 0.44	−3.05 ± 2.39	1.11 ± 2.43
`fiducial` bin 3	−0.038 ± 0.79	−0.034 ± 0.88	−3.60 ± 0.59	0.31 ± 2.78	−1.24 ± 2.89

Table 3. Average multiplicative (m) and additive (c) biases for the `fiducial`, `grid` and `grid-truedet` simulations. While we expect no bias in the shear measurements for simulations of objects on a grid with no detection employed, we do find a small bias due to masking corrections. See Appendix B for details. The `fiducial` simulations show non-trivial multiplicative biases due to a combination of blending, object detection effects, masking, and potentially other unknown causes.

5.1 Estimating N_γ^α

As discussed in Section 2, we can estimate N_γ^α, which is $n_\gamma(z)$ integrated over the interval $z_1^\alpha < z < z_2^\alpha$, by generating a simulation which we can label α, that has constant true shear g^{const} apart from in redshift interval α which has an applied true shear $g^{\mathrm{const}} + \Delta g_\alpha^{\mathrm{true}}$. Our estimate for N_γ^α is then given by

$$N_\gamma^\alpha = \frac{\Delta \bar{g}_\alpha^{\mathrm{obs}}}{\Delta g_\alpha^{\mathrm{true}}}, \qquad (21)$$

where $\Delta \bar{g}_\alpha^{\mathrm{obs}}$ is the change in measured mean shear w.r.t a simulation with constant shear g^{const} at all redshifts.

In practice, we always use $\Delta g_\alpha^{\mathrm{true}} = 0.04$ and $g^{\mathrm{const}} = -0.02$. The latter is chosen because we already generated simulations with constant shear -0.02 (separately for both shear components) for the constant shear results in Section 4. That is, for each redshift interval α, we generate an extra simulation with

$$(g_1, g_2) = \begin{cases} (0.02, 0.00) & \text{if } z_1^\alpha < z < z_2^\alpha \\ (-0.02, 0.00) & \text{otherwise.} \end{cases} \qquad (22)$$

This allows us to estimate N_γ^α for the first shear component; we also generate an analogous simulation to estimate it for the second shear

component, with

$$(g_1, g_2) = \begin{cases} (0.00, 0.02) & \text{if } z_1^\alpha < z < z_2^\alpha \\ (0.00, -0.02) & \text{otherwise,} \end{cases} \qquad (23)$$

and then average over the two sets of per-shear-component N_γ^α measurements.

We use four redshift intervals α with lower and upper redshift limits $(z_1^\alpha, z_2^\alpha) \in \{(0.0, 0.4), (0.4, 0.7), (0.7, 1.0), (1.0, 3.0)\}$. The bottom section of Table 2 summarizes the simulation inputs for these redshift-dependent shear simulations.

Since we have only a finite volume of simulation and thus noisy estimates of $\Delta \bar{g}_\alpha^{\mathrm{obs}}$, and the estimates for different α are correlated (for two reasons - they are differences with a common constant shear simulation, and their random seeds are matched with that constant shear simulation to reduce noise on the differences), we construct the covariance matrix $\mathrm{Cov}(\Delta \bar{g}_\alpha^{\mathrm{obs}}, \Delta g_\beta^{\mathrm{obs}})$ by jackknifing over simulated tiles. We then have

$$\mathrm{Cov}(N_\gamma^\alpha, N_\gamma^\beta) = \mathrm{Cov}(\Delta \bar{g}_\alpha^{\mathrm{obs}}, \Delta g_\beta^{\mathrm{obs}})/(0.04^2). \qquad (24)$$

5.2 Effective redshift distribution for lensing: Simulation measurements vs. predictions

We start with the simplest case of N_γ^α for the full ensemble of detected objects in our simulations which pass the standard selection cuts - we call this the 'non-tomographic' case, since no tomographic redshift binning is applied. As discussed in Section 2, we are interested foremost in measuring biases in the estimation of $n_\gamma(z)$ via the method used on the real data - that is, a METACALIBRATION response-weighted histogram of redshift estimates, which we denote $n_\gamma^{\mathrm{mcal}}(z)$. From this, one can make a prediction for N_γ^α

$$N_\gamma^{\alpha,\mathrm{mcal}} = \int_{z_1^\alpha}^{z_2^\alpha} \mathrm{d}z \, n_\gamma^{\mathrm{mcal}}(z). \tag{25}$$

Since in the simulations we work with discrete detections with assigned redshifts, we can estimate $N_\gamma^{\alpha,\mathrm{mcal}}$ via summing the METACALIBRATION responses of detections assigned to redshift interval α:

$$N_\gamma^{\alpha,\mathrm{mcal}} = \frac{\sum_{z_1^\alpha < z_j < z_2^\alpha} R_j}{\sum_j R_j}, \tag{26}$$

where the sum in the numerator is over all detected objects with an assigned redshift in the interval $z_1^\alpha < z_j < z_2^\alpha$, and the sum in the denominator is over all detected objects. $N_\gamma^{\alpha,\mathrm{mcal}}$ then is simply the fraction of the response-weighted detections in redshift interval α. Any differences between this fraction, and the N_γ^α measured directly from the simulations via equation 21 would imply a bias in $n_\gamma^{\mathrm{mcal}}(z)$ as an estimate of $n_\gamma(z)$.

In the top panel of Figure 8, we show the simulation measurements of N_γ^α, as well the $N_\gamma^{\alpha,\mathrm{mcal}}$ and for reference the redshift distribution $n_\gamma^{\mathrm{mcal}}(z)$. On this scale, the simulation measurements, N_γ^α, are indistinguishable from the prediction, $N_\gamma^{\alpha,\mathrm{mcal}}$, so in the bottom panel we show the fractional difference between the two quantities. The blue rectangles are the fractional difference between N_γ^α measured from the simulations (equation 21) and, $N_\gamma^{\alpha,\mathrm{mcal}}$ (estimated via equation 26) i.e. $N_\gamma^\alpha/N_\gamma^{\alpha,\mathrm{mcal}} - 1$ for the four redshift intervals α. The height of the rectangles denotes the $1-\sigma$ error bounds on this fractional difference, propagated from the covariance on the measured N_γ^α. We see that $N_\gamma^{\alpha,\mathrm{mcal}}$ is a percent-level overestimate of N_γ^α, with the disparity increasing for the higher redshift intervals.

What is going on here? Let us consider the highest redshift interval. Figure 8 implies that our ensemble of detections responds to a shear in this redshift interval around 6% more weakly than one would expect from the METACALIBRATION response-weighted redshift distribution. This could be an indication of either of two subtly different effects (or both):

(i) The measured shear for detections assigned redshift z depends only on the applied shear at redshift z, but has some (redshift-dependent) mean multiplicative bias $\bar{m}(z) \neq 0$ or equivalently non-unity mean response $\bar{R}(z) \neq 1$ (that is not captured by METACALIBRATION), such that $\bar{\gamma}^{\mathrm{obs}}(z) = \bar{R}(z)\gamma^{\mathrm{true}}(z)$, and therefore $n_\gamma(z) = \bar{R}(z)n_\gamma^{\mathrm{mcal}}(z)$.

(ii) Due to blending, the measured shear for detections assigned redshift z has some additional non-zero response to the applied shear at other redshifts z' (due to contamination by light from galaxies at z'), such that $\bar{\gamma}^{\mathrm{obs}}(z) = \int R(z,z')\gamma^{\mathrm{true}}(z')$. Here $R(z,z')$ is some function that determines the strength of linear response to shear applied at redshift z', for detections assigned redshift z. We might expect this sort of effect from blending, since the measured
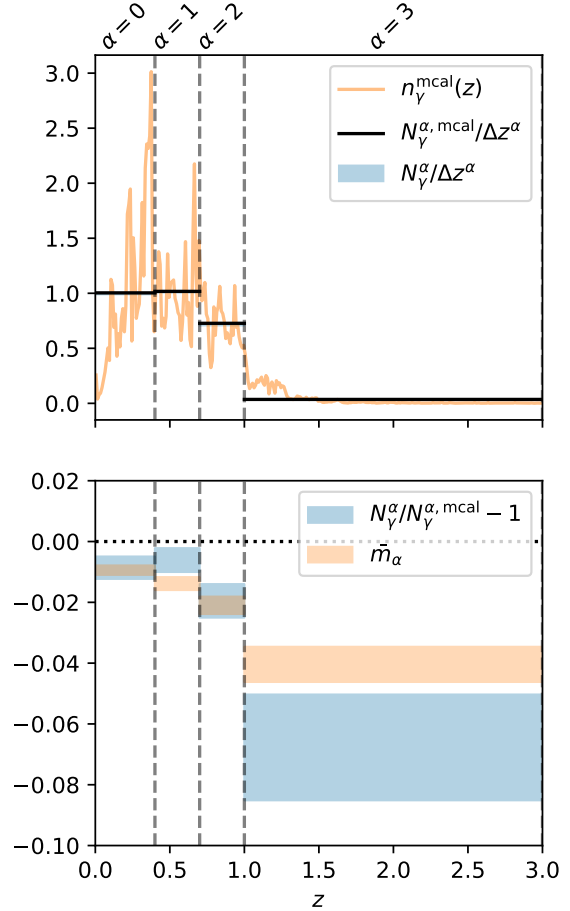

Figure 8. Measurements of N_γ^α without photo-z binning. Top panel: The orange line is $n_\gamma^{\mathrm{mcal}}(z)$ (the METACALIBRATION response-weighted $n(z)$), and the black horizontal lines show $N_\gamma^{\alpha,\mathrm{mcal}}$ i.e. the prediction for N_γ^α based on integrating $n_\gamma^{\mathrm{mcal}}(z)$ over each interval α. We also show as blue rectangles the direct measurements of N_γ^α from the simulations, with the height of the rectangle indicating the 1σ uncertainty region. In both cases we divide by the width of the redshift interval α for consistency with the plotted $n_\gamma^{\mathrm{mcal}}(z)$. On this scale the percent level differences between N_γ^α (black lines) and $N_\gamma^{\alpha,\mathrm{mcal}}$ (blue rectangles) are difficult to perceive, hence we show fractional difference in the bottom panel.

Bottom panel: Blue rectangles indicate the fractional difference between the measured N_γ^α, and $N_\gamma^{\alpha,\mathrm{mcal}}$. $N_\gamma^{\alpha,\mathrm{mcal}}$ is biased high, especially for $z > 1$, where the bias is around 6%. This means that the response of the shear catalog to shear at $z > 1$ is around 6% less than predicted. Orange rectangles show, $\bar{m}_\alpha$ the mean multiplicative bias for objects assigned to each redshift interval, as inferred from constant shear simulations. The fact that the two sets of measurements differ implies the presence of cross-redshift blending - where galaxies assigned to one redshift interval have some non-zero response to shear applied to a different interval.

shear of objects assigned redshift z may be influenced by the shear applied to light from other redshifts z'. This is precisely the effect we demonstrated with our simple simulation in Section 1, where the shape measurement of the high z galaxy had non-zero response to shear applied to the low z galaxy.

We know to some extent the first effect is present - we do see a multiplicative bias in the constant shear simulations in Section 4. In the next section we show measurements of $N_{\gamma,\alpha}^i$, for true redshift bins i, which gives us a clear insight into whether the second effect is
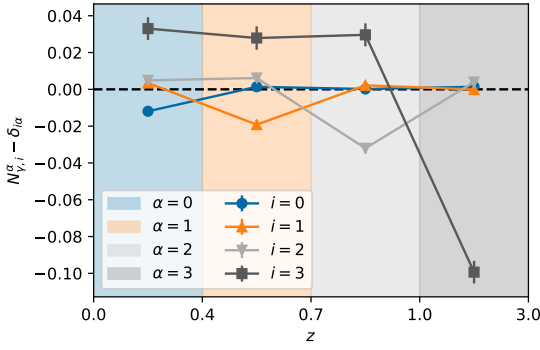
Figure 9. Measurements of $N_{\gamma,i}^{\alpha}$ for *assigned true* redshift interval i. $N_{\gamma,i}^{\alpha}$ is the response of the mean shear of galaxy ensemble i, to a shear applied in redshift interval α. In this case our ensembles i correspond to intervals in true redshift. The METACALIBRATION prediction for $N_{\gamma,i}^{\alpha}$ in this case is simply the identity matrix: $N_{\gamma,i}^{\alpha} = \delta_{i\alpha}$, since one would expect unity response of galaxies to shear applied in their own redshift interval, and zero response otherwise. Hence we have subtracted $\delta_{i\alpha}$ in this plot to reduce the dynamic range. The y-axis then demonstrates the biases from assuming the METACALIBRATION-response weighted $n(z)$ as the effective redshift distribution.

present. But before that, is useful to consider $\bar{m}_{\alpha}$, the multiplicative bias for detections assigned to redshift interval α as measured from constant shear simulations. If only the first mechanism above was present, we would expect this measurement to be equivalent to the fractional difference $N_{\gamma}^{\alpha}/N_{\gamma}^{\alpha,\mathrm{mcal}} - 1$ shown in Figure 8. The orange rectangles in Figure 8 show $\bar{m}_{\alpha}$, and the difference with respect to $N_{\gamma}^{\alpha}/N_{\gamma}^{\alpha,\mathrm{mcal}} - 1$ implies the presence of the second mechanism above, as we will explore further in the next section.

5.3 N_{γ}^{α} measurements for assigned true redshift bins

We noted in Section 2.4 that it will be useful to measure N_{γ}^{α} for subsets i of our detections. While using photometric redshift bins as those subsets, as we do in the next section, is more directly applicable to an analysis of real data, it is instructive to study $N_{\gamma,i}^{\alpha}$ for bins i in "true" redshift, or really *assigned* true redshift. The qualification here is important because we have already noted that *detections* do not necessarily correspond to only one galaxy, so may not have a unique true redshift, but they are assigned a unique true redshift in the matching procedure described in Section 3.7.

A measurement of $N_{\gamma,i}^{\alpha}$ for $i \neq \alpha$ directly probes the presence of effect (ii) discussed above – the linear response of the shear measured for detections assigned a redshift z, to the applied shear at other redshifts $z' \neq z$. It is calculated via

$$N_{\gamma,i}^{\alpha} = \frac{\Delta\bar{\gamma}_{\alpha,i}^{\mathrm{obs}}}{\Delta\gamma_{\alpha}^{\mathrm{true}}}, \tag{27}$$

where $\Delta\bar{\gamma}_{\alpha,i}^{\mathrm{obs}}$ is the change in mean measured shear for galaxies assigned to redshift interval i only, for the simulation in which redshift interval α is sheared. $N_{\gamma,i}^{\alpha}$ can be related to the $R(z, z')$ postulated in Section 5.2 via

$$N_{\gamma,i}^{\alpha} = \int_{z_1^i}^{z_2^i} dz \int_{z_1^{\alpha}}^{z_2^{\alpha}} dz' R(z, z'). \tag{28}$$

The prediction for $N_{\gamma,i}^{\alpha}$ based on the METACALIBRATION

response-weighted $n(z)$ for detections assigned to redshift interval i, $n_{\gamma,i}^{\mathrm{mcal}}(z)$, is simple. Since in this case the full sample is within redshift interval i, we should see unity response to shear applied to the assigned interval i i.e. when $\alpha = i$, and zero response otherwise i.e. when $\alpha \neq i$. That is, $N_{\gamma,i}^{\mathrm{mcal}} = \delta_{i,\alpha}$.

In Figure 9, we plot as points with errorbars the $N_{\gamma,i}^{\alpha}$ for the four redshift intervals used. Since we expect $N_{\gamma,i}^{\alpha}$ to be close to unity for $i = \alpha$, we subtract $\delta_{i\alpha}$ from the measurements so that the results are all near zero. We see that the diagonal elements of $N_{\gamma,i}^{\alpha}$ are all less than unity, implying that detections assigned to a given redshift interval do not have unity response to shear applied in that same redshift interval, with the size of the effect increasing from around 1% at low redshift to around 10% at high redshift. This implies the presence of a multiplicative shear bias e.g. due to a dilution of the applied shear due to contamination of the shape measurement by light from other (unsheared) redshift intervals.

The off-diagonal terms in $N_{\gamma,i}^{\alpha}$ are predominantly positive, especially for $i = 3$, that is for detections assigned to the highest redshift interval. This means that the detections exhibit a positive response to shear applied in the other redshift intervals $\alpha \neq i$, a clear detection of the mechanism (ii) described in the previous section, and demonstrated in our simple simulation in Section 1. This effect is potentially important since it implies there will be extra correlation in the shears measured at different redshifts w.r.t what one would expect from the estimated redshift distribution $n_{\gamma}^{\mathrm{mcal}}(z)$. Or put differently, there will be additional tails or broadening in the effective redshift distribution for weak lensing.

5.4 N_{γ}^{α} measurements for photo-z bins

We now repeat the above measurement, but applied to photometric redshift bins, rather than bins of assigned true redshifts. These are the most directly applicable measurements for the DES Y3 weak lensing analyses. We use the same simulations and shear measurements to calculate the integrated effective density $N_{\gamma,i}^{\alpha}$ for photometric redshift bin i. $N_{\gamma,i}^{\alpha}$ is estimated from the simulations in the same way as above but now simply restricting the mean shear calculation to detections assigned to photometric redshift bin i, according to the procedure described in Section 3.7. Note the change in meaning of index i, which now labels the photo-z bin rather than the bin in assigned true redshift.

Figure 10 shows the difference between the $N_{\gamma,i}^{\alpha}$ measurements and the $N_{\gamma,i}^{\alpha,\mathrm{mcal}}$ predictions for each photo-z bin i. In this case, $N_{\gamma,i}^{\alpha,\mathrm{mcal}}$ is no longer $\delta_{i,\alpha}$ as before, since for any (i, α) pair, the photo-z bin i includes galaxies outside of the range $(z_1^{\alpha}, z_2^{\alpha})$. Instead, we now calculate $N_{\gamma,i}^{\alpha,\mathrm{mcal}}$ using equation 26, summing over the METACALIBRATION responses for the galaxies in each photo-z bin i. The blue rectangles in Figure 10 show the resulting $N_{\gamma,i}^{\alpha} - N_{\gamma,i}^{\alpha,\mathrm{mcal}}$ values with their estimated uncertainties.

The measurements are noisier for this case, especially where there is little weight in the underlying $n(z)$, e.g. at high redshift ($\alpha = 2$ and $\alpha = 3$) for the first photo-z bin, $i = 0$. Due to the fact that our photometric redshift pipeline cannot group galaxies perfectly in redshift, the effects from source blending in Figure 9 are diluted. The model fits that also appear in Figure 10 are described in the next section.
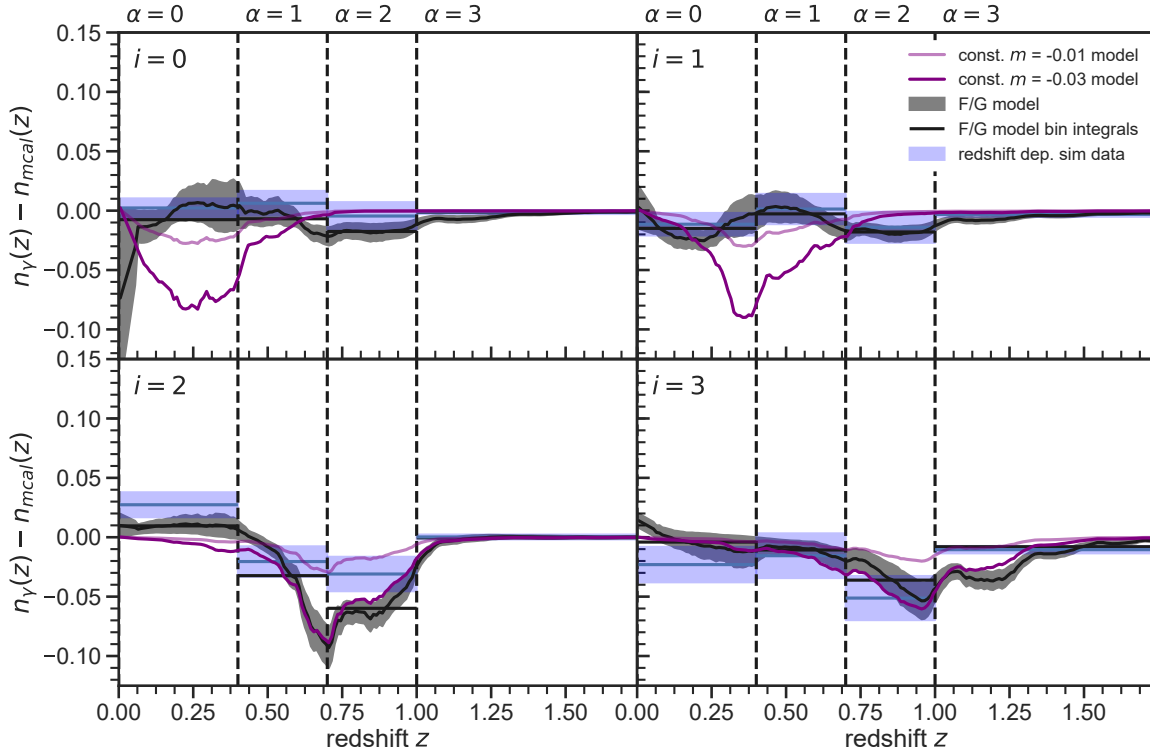
Figure 10. Modeling $n_{\gamma,i}^{\mathrm{model}}(z)$ using $N_{\gamma,i}^{\alpha}$ from our image simulations. In each panel, the blue rectangles show $N_{\gamma,i}^{\alpha} - N_{\gamma}^{\alpha,\mathrm{mcal}}$ for the photometric redshift bin i. The width represents the redshift extent of the interval α and height indicates the 1σ uncertainty on the measurement. The grey band in each panel shows the posteriors of our $n_{\gamma,i}^{\mathrm{model}}(z)$ and the horizontal black lines are the integrals over this model which form the predictions for the $N_{\gamma,i}^{\alpha}$ measurements from the simulations. The light and dark purple lines show the models that would result if one used a constant multiplicative bias of -0.01 and -0.03 respectively. It is clear from the simulation data, especially in bin 2, that no constant multiplicative bias model would be sufficient. In particular, at lower redshift in bin 2, they cannot capture the increased response to low redshift shear due to blending.

5.5 $n_{\gamma}(z)$ bias model constraints for photo-z bins

Above, we have made measurements of $N_{\gamma,i}^{\alpha}$ in a set of coarse redshift bins. While these measurements constitute direct constraints on the shear systematics in these bins, we expect the underlying effects to be smooth functions of redshift, and we require a continuous corrected $n_{\gamma}(z)$ to make theoretical predictions. Thus we now fit a smooth model to the measurements. We posit that this smooth model has two terms as follows

$$n_{\gamma,i}^{\mathrm{model}}(z) = (1 + F_i(z))\, n_{\gamma,i}^{\mathrm{mcal}}(z) + G_i(z) \ . \tag{29}$$

The term $F_i(z)$ captures changes in the effective weighting of the discrete sources used to construct $n_{\gamma,i}^{\mathrm{mcal}}(z)$. These effects could include things like constant multiplicative biases or redshift-dependent multiplicative bias effects. The term $G_i(z)$ captures responses to shear at redshifts not present in the naive $n_{\gamma,i}^{\mathrm{mcal}}(z)$ distribution computed from the detected sources. This term is designed to capture blending effects where high-redshift sources respond to input shear at lower redshifts. Note though that a sufficiently flexible $F_i(z)$ could also capture such effects.

5.5.1   The model

Due to the limited volume of image simulation data, we can only employ relatively constrained models. For $G_i(z)$, we fit a template with a single free amplitude. The template is built from the flux

density of the COSMOS sample in the *riz*-bands as a function of redshift. This model is motivated by the idea that for random projections, blending effects will be proportional to the flux density of objects.

For $F_i(z)$, we employ a model with two parts. First, we have a constant term that is meant to capture overall multiplicative biases. Second, we employ a perturbative method detailed in Berberan-Santos (2007) to build a set of correction terms that capture mean shifts in the redshift distribution and possibly changes in its width. The overall idea of this method is that we can expand $n_{\gamma,i}^{\mathrm{model}}(z)$ perturbatively about the input $n_{\gamma,i}^{\mathrm{mcal}}(z)$ with the coefficients of the expansion corresponding to changes in the moments of $n_{\gamma,i}^{\mathrm{mcal}}(z)$. The form of this series for two general PDFs $h(x)$ and $p(x)$ is

$$h(x) \approx p(x) \left(1 + a_1 \frac{1}{p(x)} \frac{dp(x)}{dx} + a_2 \frac{1}{2p(x)} \frac{d^2 p(x)}{dx^2} + \dots \right). \tag{30}$$

In our application, we use either the first or the first two terms of this series to generate template functions to put into $F_i(z)$. We then fit for the coefficients a_1 and/or a_2. Given that directly differentiating the simulation $n_{\gamma,i}^{\mathrm{mcal}}(z)$ will be quite noisy, we fit a model to this quantity and then differentiate that model. If we call this model $\phi(z;\theta_i)$, then the final form for $F_i(z)$ is

$$F_i(z) = a_0 + a_1 \frac{1}{\phi(z;\theta_i)} \frac{d\phi(z;\theta_i)}{dz} + a_2 \frac{1}{2\phi(z;\theta_i)} \frac{d^2\phi(z;\theta_i)}{dz^2} \ . \tag{31}$$

In this expression, a_0 is approximately an overall multiplicative

bias term, a_1 approximately controls changes in the mean redshift of $n_{\gamma,i}^{\mathrm{mcal}}(z)$, and a_2 approximately controls a combination of the mean and the width of $n_{\gamma,i}^{\mathrm{mcal}}(z)$. We use the following distribution (Baugh & Efstathiou 1993; Brainerd et al. 1996),

$$\phi(z; a, b, c) \propto z^a \exp\left(-\left(\frac{z}{b}\right)^c\right) \tag{32}$$

for our fiducial model for ϕ, keeping only the terms up to a_1 in equation 31. As an alternative, we also considered using a Student's-t distribution with $\nu = 1$ degree of freedom, keeping all terms up to a_2. We found that for the Student's-t distribution we needed to increase the width of the distribution by a factor of 10 when generating the model terms in order to avoid extremely strong model effects in the tails of the distribution.

From these smooth models, we can estimate $N_{\gamma,i}^{\alpha}$ via

$$N_{\gamma,i}^{\alpha,\mathrm{model}} = \int_{z_1^{\alpha}}^{z_2^{\alpha}} \mathrm{d}z\, n_{\gamma,i}^{\mathrm{model}}(z). \tag{33}$$

We can also make a prediction for the mean multiplicative bias for photo-z bin i, $\bar{m}_i$, or equivalently mean response $\bar{R}_i = 1 + \bar{m}_i$, which we measured from the constant shear sims, which is simply

$$\bar{R}_i^{\mathrm{model}} = \int_0^{\infty} n_{\gamma,i}^{\mathrm{model}}(z), \tag{34}$$

i.e. the normalization of $n_{\gamma}(z)^{\mathrm{model}}$. Therefore we can use both $N_{\gamma,i}^{\alpha}$ and $\bar{R}_i$ measurements to constrain the $F_i(z)$ and $G_i(z)$ components of our $n_{\gamma}(z)$ bias model.

5.5.2 Fitting procedure

Equipped with simulation measurements of $N_{\gamma,i}^{\alpha}$ and $\bar{R}_i$, an estimate of their joint covariance matrix C_i (again estimated by jackknifing over simulated tiles) and a model for $n_{\gamma}(z)$ with which we can predict them, we can now form a likelihood to constrain that model.

We use Markov chain Monte Carlo (MCMC) to produce samples of $F(z)$ and $G(z)$ and thus $n_{\gamma}^{\mathrm{model}}(z)$. Our likelihood is

$$\log L = \mathcal{N}\left(\left[\bar{R}_i, N_{\gamma,i}^{\alpha}\right]^{\mathrm{sim}} - \left[\bar{R}_i, N_{\gamma,i}^{\alpha}\right]^{\mathrm{model}}, C_i\right), \tag{35}$$

where $\mathcal{N}(\mu, \Sigma)$ indicates a Gaussian distribution with mean μ and covariance Σ. We use wide, uninformative priors on the free parameters of $F(z)$ and $G(z)$, and use EMCEE (Foreman-Mackey et al. 2013) to sample the posterior. We use a configuration of 12 walkers taking 10^4 steps each. We then discard the first 5×10^3 samples and thin the rest of the samples by a factor of 10. Our final chains have autocorrelation lengths of approximately $40 - 50$ and are thus long enough to generate a sufficient number of independent samples. Having sampled the posteriors of the model parameters, we can then produce samples of $F(z)$, $G(z)$, and $n_{\gamma,i}^{\mathrm{model}}(z)$ that are conditioned on our simulation measurements.

5.6 Model constraints

The posterior constraints on our models are show in Figure 10 as the grey bands. We also show the predictions for $N_{\gamma,i}^{\alpha}$ as the horizontal grey lines. We find that our model generally provides a good fit to the data. In this figure we also show the predictions of a model which employs a constant m per tomographic bin, i.e. $n_{\gamma,i}^{\mathrm{model}}(z) = (1 + m_i)n_{\gamma,i}^{\mathrm{mcal}}(z)$. While this model may be sufficient for the lower redshift tomographic bins, in bin $i = 2$ we find that

quantity	tomo. bin	fiducial model	Student's-t model
$m \times 100$	0	-1.36 ± 0.29	-1.37 ± 0.29
$m \times 100$	1	-1.78 ± 0.36	-1.77 ± 0.36
$m \times 100$	2	-2.48 ± 0.41	-2.52 ± 0.42
$m \times 100$	3	-3.34 ± 0.52	-3.31 ± 0.52
$\Delta \bar{z}_0 \times 100$	0	-0.65 ± 0.27	-0.96 ± 0.30
$\Delta \bar{z}_0 \times 100$	1	-0.80 ± 0.31	-0.76 ± 0.29
$\Delta \bar{z}_0 \times 100$	2	-0.36 ± 0.26	-0.40 ± 0.19
$\Delta \bar{z}_0 \times 100$	3	-1.13 ± 0.60	-0.45 ± 0.37

Table 4. Inferred constraints on m and $\delta\bar{z}$ from the $n_{\gamma}(z)$ model samples fit to the image simulations measurements. These $n_{\gamma}(z)$ models are generated by perturbing the image simulation $n_{\gamma}^{\mathrm{mcal}}(z)$.

it cannot capture the low redshift tails introduced into $n_{\gamma,i}^{\mathrm{model}}(z)$, perhaps due to blending.

From each $n_{\gamma,i}^{\mathrm{model}}(z)$ sample we can calculate an effective multiplicative bias m, and a change in mean redshift $\Delta\bar{z}$. m is equal to the normalization of $n_{\gamma,i}^{\mathrm{model}}(z)$ minus 1, i.e.

$$m = \int \mathrm{d}z\, n_{\gamma,i}^{\mathrm{model}}(z) - 1 \tag{36}$$

and

$$\Delta\bar{z} = \frac{\int \mathrm{d}z\, z n_{\gamma,i}^{\mathrm{model}}(z)}{\int \mathrm{d}z\, n_{\gamma,i}^{\mathrm{model}}(z)} - \frac{\int \mathrm{d}z\, z n_{\gamma,i}^{\mathrm{mcal}}(z)}{\int \mathrm{d}z\, n_{\gamma,i}^{\mathrm{mcal}}(z)}. \tag{37}$$

Note we have chosen the sign convention of our definition of $\Delta\bar{z}$ to match that used in the cosmology analyses (eg. DES Collaboration (2021)), which means it is the correction one would apply to $n_{\gamma,i}^{\mathrm{mcal}}(z)$, rather than the bias in $n_{\gamma,i}^{\mathrm{mcal}}(z)$ (which would have the opposite sign).

These two summary statistics of $n_{\gamma}(z)$ are probably the two most important for weak lensing analysis, where higher-order biases in the redshift distribution are likely to have a subdominant effect to changes in normalization (i.e. multiplicative bias) or mean redshift. The two models give largely similar constraints in m and $\delta\bar{z}$, with the fiducial model in general allowing slightly larger shifts in redshift. The first two lines of Table 4 contains these inferred m and $\delta\bar{z}$ constraints for the two $n_{\gamma}(z)$ bias models (fiducial and Student's-t), when applied to the fiducial $n(z)$s from the image simulations.

The $n_{\gamma}(z)$ bias model constraints encapsulate our characterization of shear calibration biases for each DES Y3 photometric redshift bin, including redshift dependent effects such as the impact of blending on the effective redshift distribution for lensing. They also include the statistical uncertainty on the calibration due to finite simulation volume. When using the $n_{\gamma}(z)$ bias model for the real DES Y3 data, we apply it instead to the $n(z)$s estimated from the DES Y3 data. In the next section, we describe this procedure, and explore potential systematic uncertainties in our corrections, in order to provide statistical and systematic uncertainty priors on the corrections measured from the simulations here.

6 $n_{\gamma}(z)$ CORRECTIONS FOR THE DES Y3 SHEAR CATALOG

In Section 5, we generated estimates of the bias in our methodology for constructing the effective redshift distribution for lensing, $n_{\gamma}(z)$. We did this by comparing direct estimates of $n_{\gamma}(z)$ from the image simulations, possible only because we know the input shear to the simulations, to the method we use on the Y3 data: the METACAL-IBRATION response-weighted redshift distribution, $n_{\gamma}^{\mathrm{mcal}}(z)$. These

bias estimates are intended to be used to make corrections to the redshift distributions used in theory predictions in weak lensing cosmology analyses. In our case, those redshift distributions are those estimated from the DES Y3 data, summarized in Myles & Alarcon et al., (2021), which provide ensembles of redshift distributions constrained by both galaxy flux information, and clustering (Gatti & Giannini et al., 2020). In Section 6.1 we describe our procedure for applying our corrections to these redshift distributions. This procedure results in a set of $n_\gamma(z)$ samples, which can be sampled over when performing cosmological parameter inference (e.g. using the HYPERRANK method described in Cordero et al. 2021).

The accuracy of our corrections depends on how realistically we have simulated the DES Y3 data. In Sections 6.2-6.4 we explore this issue by inspecting the sensitivity of our $n_\gamma(z)$ bias corrections to variations in the simulation measurements or simulation analysis choices. Results from these variations feed into our systematic uncertainties.

Finally, in Section 6.5 we summarize our final corrections which include both statistical and systematic uncertainties.

6.1 Applying corrections to the DES Y3 $n(z)$s

Myles & Alarcon et al., (2021) provide an ensemble of 1000 samples (for each of the four photo-z bins) drawn from the posterior of the set of photo-z bin redshift distributions, constrained by both galaxy flux (including color) and clustering information. These distributions are already weighted by the METACALIBRATION response estimated from the DES data, so in fact correspond to samples of $n_{\gamma,i}^{mcal}(z)$. For each of these 1000 samples, we draw 100 randomly chosen samples of our bias model inferred from our simulation measurements, quantified by the functions $F(z)$ and $G(z)$. Then we apply these model biases to $n_{\gamma,i}^{mcal}(z)$ using equation 29 to produce estimates of $n_\gamma(z)$. This process leaves us with $1000 \times 100 = 100,000$ $n_\gamma(z)$ samples for each photo-z bin, but note that in practice one could downsample this set of $n_\gamma(z)$s to a more manageable number depending on the application.

In Figure 11 we show the distributions of mean multiplicative bias, m (top panels), and mean redshift, $\bar{z}$ (bottom panels) inferred from these $n_\gamma(z)$ samples. m is calculated as in equation 36 i.e. is simply the normalization of $n_\gamma(z)$ minus 1. In all cases the height of the violin shape is proportional to the fractional weight in the distribution at a given value of the x-axis. The lines labelled "fiducial" represent the inferred m and $\bar{z}$ distributions for this case. We also list the mean and standard deviation of these m and $\bar{z}$ distributions in Table 6.1. We note here that we are in some sense extrapolating the corrections we have inferred from our simulations onto the DES Y3 data, and one must always be careful with extrapolation. However, the fact that the mean m values inferred from the image simulation $n_\gamma(z)$ models (first row of Table 4) are within the 1σ bounds of the "fiducial" Y3 data m priors in Table 6.1, gives us confidence that this extrapolation is well-behaved.

In Table 6.1 we also list the mean and standard deviation of two other quantities, labelled "$\delta\bar{z}$" and "$\delta\bar{z}_{FG}$". These are defined as follows. For a given input $n_\gamma^{mcal}(z)$ sample, which we will call $n_s^{mcal}(z)$, we generate an output $n_\gamma(z)$ sample, $n_{\gamma,s}(z)$, by applying samples of the $F(z)$ and $G(z)$ perturbation functions. For a given $n_{\gamma,s}(z)$, $\delta\bar{z}_s$ is defined as the change in mean redshift with respect to the mean redshift of the *ensemble* of input $n_\gamma^{mcal}(z)$ samples i.e.

$$\delta\bar{z}_s = \bar{z}\left[n_{\gamma,s}(z)\right] - \left\langle \bar{z}\left[n_s^{mcal}(z))\right]\right\rangle_s, \tag{38}$$

where $\bar{z}[\psi(z)]$ represents the mean redshift of the function $\psi(z)$,

and $\langle\rangle_s$ represents an average over all samples s. Note the distribution of this $\delta\bar{z}$ will include the scatter in the input $n_\gamma^{mcal}(z)$ samples.

The quantity $\delta\bar{z}_{FG}$ on the other hand does not; it is defined as

$$\delta\bar{z}_{FG,s} = \bar{z}\left[n_{\gamma,s}(z)\right] - \bar{z}\left[n_s^{mcal}(z)\right] \tag{39}$$

for a given sample s, and thus isolates the impact of applying the $F(z)$ and $G(z)$ correction functions. Hence it is this quantity that is more directly comparable to the $\Delta\bar{z}$ we defined in Section 5.6 for the image simulation $n_\gamma(z)$, and listed values for in Table 4. The image simulation $\Delta\bar{z}$ from Table 4 are all within the 1σ bounds of the $\delta\bar{z}_{FG}$ distributions, except for photo-z bin 1, where it is within 2σ.

In the next three sub-sections we test for sensitivity of the inferred m and $\delta\bar{z}$ priors to various potential systematics in our simulation corrections.

6.2 Alternative photo-z-binning schemes

Given the range of viable different photo-z binning schemes presented in Section 3.7, we test for the sensitivity of our corrections to this choice, and include a corresponding contribution to our systematic uncertainties related to this choice. For each of the photo-z binning schemes, we repeat the inference of $n_\gamma(z)$ corrections performed in Section 5.5, but now using the slightly different $N_{\gamma,i}^\alpha$ and $\bar{R}_i$ measurements resulting from changing the photo-z binning scheme. This results in sets of $F(z)$ and $G(z)$ samples for each photo-z binning scheme, from which we can generate $n_\gamma(z)$ samples for the DES Y3 data, following the procedure in Section 6.1. The inferred ms and $\bar{z}$s for these cases are also shown in Figure 11, as the lines labelled "zmatch pz-binning", "wmatch pz-binning " and "equal pz-binning" (see Section 3.7 for the details of the binning methods).

6.3 Simulation re-weighting

The overall agreement between object properties and density in the simulation and real data is very good, both globally and within each redshift bin. To validate that any remaining differences would not significantly modify our inferred shear calibration, we focus on several properties most likely to correlate with the amount of blending, which is our main source of shear calibration bias.

We implement a re-weighting scheme designed to match the small-scale counts of close pairs in the DES Y3 data, including the magnitude dependence of this behaviour. We believe this should provide a first order estimate of the potential impact of ignoring galaxy clustering in our simulations, which although weak for our wide-in-redshift photometric redshift bins, may have non-negligible impact on the amount of blending.

We choose three quantities on which to re-weight: the magnitude (calculated from the mean of the flux in the r, i and z bands), the distance to the nearest neighbor, and the magnitude of the nearest neighbor (again based on the r, i and z bands). The neighbor can be any detected object with non-negative flux (i.e. we do not restrict the neighbor candidates to objects passing shape catalog cuts). We then aim to produce a set of weights to apply to the simulation results that will improve the match of the joint distributions of these quantities between simulations and data.

To accomplish this, we use k-means clustering to define clusters of these three quantities based on 200,000 randomly selected objects in each photo-z bin. We then assign all objects in both the
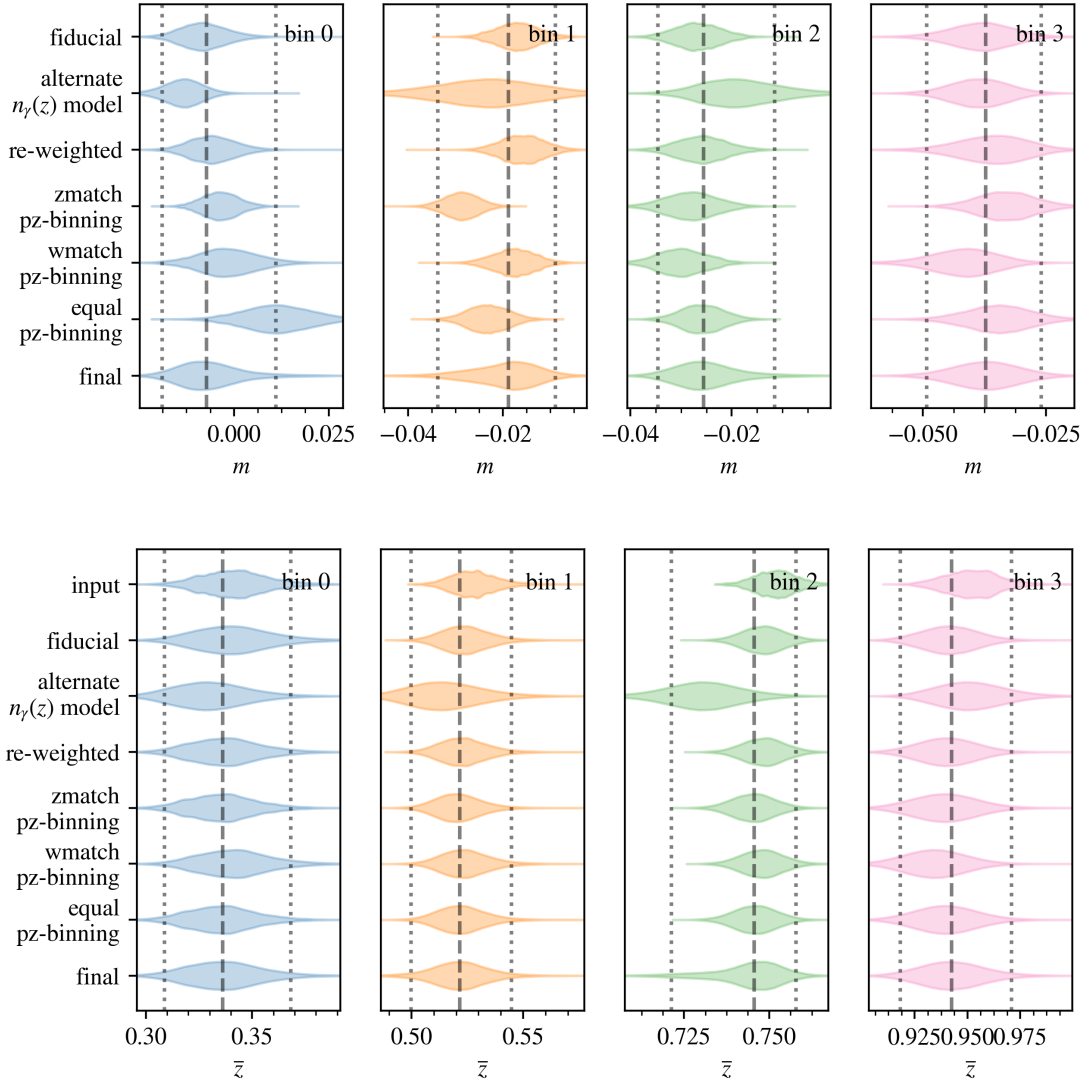
Figure 11. Distributions of the inferred m (top panels) and mean redshift $\bar{z}$ (bottom panels) of our DES Y3 redshift $n_\gamma(z)$ estimates, generated by applying our simulation corrections to an input ensemble of redshift distributions. For each line $n_\gamma(z)$ corrections are applied from a variation on the fiducial simulation analysis, which have been chosen to expose potential systematic errors in the fiducial approach. The lines marked 'final' represent the mixture approach described in Section 6.5, which incorporates systematic uncertainty by mixing together samples from the alternative approaches with the fiducial analysis. In the lower panel, we show also in the first line the distribution of mean redshifts for the input (i.e. before correction) redshift distributions.

simulations and DES Y3 data to these clusters. Weights are produced for each photo-z bin by taking the ratio of the number of the number of objects assigned to each data cluster to the number assigned to each simulation cluster. We show in Figure 12 that this re-weighting of objects improves agreement with the data.

We update the weights of our simulated shear catalogs, and re-derive the $n_\gamma(z)$ correction model (by re-performing the $N_{\gamma,i}^\alpha$ and $\bar{R}_i$ measurements and the $n_{\gamma,i}^{\mathrm{model}}(z)$ fits). We again apply the resulting $F(z)$ and $G(z)$ corrections to the Y3 data $n(z)$ ensemble to get an ensemble of $n_\gamma(z)$ samples. The inferred ms and $\bar{z}$s for these cases are also shown in Figure 11, as the lines labelled "re-weighted". The impact on $\bar{z}$ is small, with at most 0.2% shifts in the mean. There is also a small (between 0.1 and 0.2%), coherent shift in the mean of the m distributions to less negative values for all photo-z bins. While small, we find the sign surprising, since we

expect this re-weighting procedure to increase the number of close pairs in the simulation, and thus increase the impact of blending on the multiplicative bias.

We note that this re-weighting procedure will not include all the potentially relevant effects present in a galaxy sample with realistic clustering. We discuss in detail in Section 7 the possibility that we have underestimated blending-related biases as a result of this.

6.4 Alternative $n_\gamma(z)$ bias parameterization

We chose a fiducial functional form for the multiplicative correction $F(z)$ in our $n_\gamma(z)$ bias model. This was based on the perturbative expansion in equation 31, applied to the $n(z)$ model in equation 32. Note that we only use this simple parametric model to generate a functional form for the perturbations from $n_\gamma^{\mathrm{mcal}}(z)$; we do not use

Model	photo-z bin	$m \times 100$	$\bar{z}$	$\delta\bar{z} \times 100$	$\delta\bar{z}_{FG} \times 100$
input	0	N/A	0.341 ± 0.016	0.000 ± 1.608	0.000 ± 0.000
fiducial	0	-0.789 ± 0.625	0.341 ± 0.019	-0.021 ± 1.938	-0.021 ± 0.711
alternate $n_\gamma(z)$	0	-1.346 ± 0.541	0.330 ± 0.018	-1.054 ± 1.834	-1.054 ± 0.991
re-weighted	0	-0.574 ± 0.615	0.337 ± 0.017	-0.406 ± 1.750	-0.406 ± 0.353
zmatch pz	0	-0.368 ± 0.481	0.335 ± 0.016	-0.565 ± 1.644	-0.565 ± 0.143
wmatch pz	0	-0.172 ± 0.790	0.341 ± 0.018	-0.009 ± 1.766	-0.009 ± 0.361
equal pz	0	1.117 ± 0.908	0.335 ± 0.017	-0.580 ± 1.651	-0.580 ± 0.212
final	0	-0.627 ± 0.908	0.336 ± 0.018	-0.467 ± 1.846	-0.467 ± 0.758
input	1	N/A	0.528 ± 0.013	-0.000 ± 1.280	0.000 ± 0.000
fiducial	1	-1.703 ± 0.446	0.524 ± 0.013	-0.367 ± 1.344	-0.367 ± 0.215
alternate $n_\gamma(z)$	1	-2.315 ± 1.080	0.515 ± 0.018	-1.267 ± 1.812	-1.267 ± 1.441
re-weighted	1	-1.589 ± 0.461	0.523 ± 0.013	-0.467 ± 1.341	-0.467 ± 0.193
zmatch pz	1	-2.869 ± 0.375	0.521 ± 0.012	-0.725 ± 1.212	-0.725 ± 0.358
wmatch pz	1	-1.763 ± 0.452	0.524 ± 0.013	-0.385 ± 1.335	-0.385 ± 0.218
equal pz	1	-2.308 ± 0.435	0.522 ± 0.013	-0.561 ± 1.311	-0.561 ± 0.282
final	1	-1.984 ± 0.779	0.521 ± 0.015	-0.671 ± 1.506	-0.671 ± 0.833
input	2	N/A	0.750 ± 0.006	-0.000 ± 0.629	0.000 ± 0.000
fiducial	2	-2.667 ± 0.458	0.746 ± 0.006	-0.365 ± 0.632	-0.365 ± 0.209
alternate $n_\gamma(z)$	2	-1.649 ± 0.892	0.729 ± 0.012	-2.048 ± 1.173	-2.048 ± 1.251
re-weighted	2	-2.570 ± 0.494	0.746 ± 0.006	-0.355 ± 0.641	-0.355 ± 0.152
zmatch pz	2	-2.759 ± 0.511	0.743 ± 0.006	-0.676 ± 0.637	-0.676 ± 0.204
wmatch pz	2	-2.967 ± 0.483	0.745 ± 0.006	-0.411 ± 0.627	-0.411 ± 0.202
equal pz	2	-2.558 ± 0.430	0.743 ± 0.006	-0.613 ± 0.626	-0.613 ± 0.256
final	2	-2.412 ± 0.760	0.741 ± 0.011	-0.830 ± 1.071	-0.830 ± 0.959
input	3	N/A	0.946 ± 0.015	0.000 ± 1.524	0.000 ± 0.000
fiducial	3	-3.794 ± 0.718	0.934 ± 0.013	-1.170 ± 1.317	-1.170 ± 1.046
alternate $n_\gamma(z)$	3	-3.780 ± 0.606	0.945 ± 0.019	-0.060 ± 1.913	-0.060 ± 0.977
re-weighted	3	-3.574 ± 0.841	0.933 ± 0.014	-1.304 ± 1.378	-1.304 ± 1.333
zmatch pz	3	-3.253 ± 0.589	0.933 ± 0.017	-1.313 ± 1.698	-1.313 ± 0.902
wmatch pz	3	-4.179 ± 0.852	0.927 ± 0.015	-1.875 ± 1.490	-1.875 ± 1.575
equal pz	3	-3.442 ± 0.690	0.933 ± 0.015	-1.277 ± 1.548	-1.277 ± 1.037
final	3	-3.692 ± 0.761	0.936 ± 0.017	-1.008 ± 1.672	-1.008 ± 1.290

Table 5. Summary statistics of the DES Y3 data $n_\gamma(z)$ distributions, or in the case of the lines labelled "input", input ensemble of redshift distributions i.e. before correction. The "input" lines hence have no associated multiplicative bias m. The column labelled $\bar{z}$ is the mean and standard deviation of the distribution of mean redshift of $n_\gamma(z)$. The column labelled $\delta\bar{z}$ is the mean and standard deviation of the distribution of $\delta\bar{z}$, which for a given $n_\gamma(z)$, is the difference in mean redshift w.r.t to the mean redshift of the input ensemble of $n(z)$. The column labelled $\delta\bar{z}_{FG}$ is the mean and standard deviation of the distribution of $\delta\bar{z}_{FG}$, which for a given $n_\gamma(z)$, is the difference in mean redshift w.r.t the mean redshift of the specific $n(z)$ sample used to generate the $n_\gamma(z)$.

the parameteric fit directly. We have also investigated the use of a Student's-t model, which we use to generate an alternative set of $F(z)$ and $G(z)$ samples with which we correct the DES Y3 $n(z)$ ensemble. The resulting inferred m and $\delta\bar{z}$ distributions are show in Figure 11, as the lines labelled "alternate $n_\gamma(z)$ model". We observe some non-negligible shifts with respect to the fiducial m and $\bar{z}$. The largest are for photo-z bin 2, with a shift in the mean of the m distribution of 1.0%, and a shift in the mean $\bar{z}$ of -1.7%. We also see broadening of the $\bar{z}$ distributions for all photo-z bins.

6.5   Final $n_\gamma(z)$ priors

As discussed in Section 6.1, the $n_\gamma(z)$ samples inferred from applying the results of our fiducial simulation to the DES Y3 data $n(z)$ include our fiducial correction and statistical uncertainties. We incorporate systematic uncertainties as follows. We generate a final prior on $n_\gamma(z)$ which is a mixture of the priors inferred from our fiducial approach, and the variations described in Sections 6.2-6.4. More precisely, we formulate the prior $P(n_\gamma(z))$ as

$$P(n_\gamma(z)) \propto P(n_\gamma(z)|\text{Fiducial assumptions})$$
$$+P(n_\gamma(z)|\text{Simulation re-weighting})$$
$$+P(n_\gamma(z)|\text{Alternative } n_\gamma(z) \text{ bias parameterization}) \quad (40)$$
$$+P(n_\gamma(z)|\text{Alternative photo-}z \text{ binning scheme}).$$

Each term on the right-hand-side represents the prior on $n_\gamma(z)$ given some set of analysis assumptions. We normalize each term on the right-hand-side such that they contribute equal weight to the total prior on $n_\gamma(z)$. Note that this means the priors based on the three alternative photo-z binning schemes described in Section 6.2 each contribute only one third of the weight to the mixed prior, compared to the other variations. This choice, while somewhat arbitrary, ensures we give equal weight to the three categories of systematic uncertainty defined in Sections 6.2-6.4.

We can generate the final prior on $n_\gamma(z)$ simply by combining the samples from the individual cases in the correct proportions. The inferred m and $\bar{z}$ distributions for this final prior is shown in the bottom rows of Figure 11. The grey dotted vertical lines show the 5th and 95th percentiles of this final prior, while the dashed vertical line shows its median.

One can see qualitatively by eye that this mixture prior accounts for systematic uncertainties with a longer tail where there is an outlier in one of the variations. For example, the m distribution for bin 0, while still centered close to the fiducial case, gains extra weight in the tail to high m due to the outlying m distribution for the "equal" photo-z binning variation. Similarly, the significant shift for the alternate $n_\gamma(z)$ variation in the $\bar{z}$ distribution for bin 2 is accounted for by a large tail to low $\bar{z}$ values.

We note that the m and $\delta\bar{z}$ statistics presented here are sum-
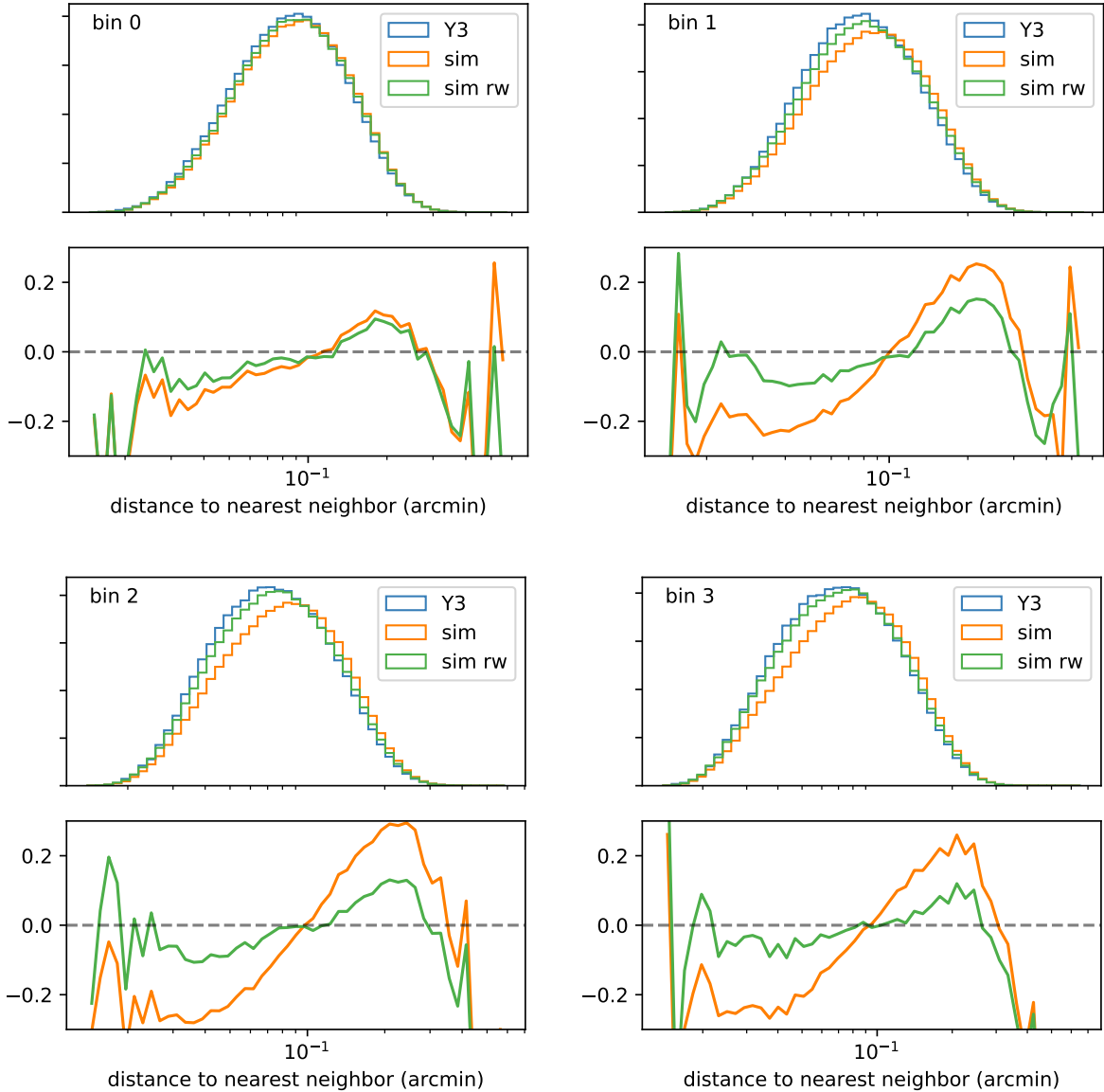
Figure 12. Distribution of distance to nearest neighbor for each photo-z bin. The first and third rows show these distributions (as normalized histograms) for the Y3 data ("Y3", blue), the image simulations ("sim", orange) and the re-weighed simulations ("sim rw", green - using the re-weighting procedure described in Section 6.3). The second and fourth rows show the fractional difference of the "sim" and "sim rw" distributions with respect to the Y3 data distributions. In all cases the agreement between simulations and data is improved by the re-weighting procedure.

mary statistics of the full corrections, which are incorporated into perturbed $n_\gamma(z)$ samples. It is these $n_\gamma(z)$ samples that should be used in theoretical predictions of weak lensing statistics (e.g. shear correlation functions or tangential shear) for the DES Year 3 shear catalog. An algorithm for efficiently sampling over discrete $n_\gamma(z)$ samples, such as that presented in Cordero et al. (2021) can be used (or alternatively one could directly sample the likelihood in equation 35, simultaneously with the likelihood for the Year 3 weak lensing statistics). However, for some applications, it may be desirable and sufficiently accurate to directly use the inferred m and $\delta\bar{z}$ statistics as approximate corrections to the effective redshift distribution, and sample over them as nuisance parameters. The accuracy of such an approach should be validated against the use of the full corrections.

7 DISCUSSION

Image simulations have long been recognized as an essential tool for calibrating shear estimation pipelines, and have played a role in providing the shear calibration for all recent weak lensing cosmology analyses. While current methodologies like METACALIBRATION have made major strides in dealing with well-studied shear estimation biases like noise bias and model bias, we are now entering an era of weak lensing data that is both deep, resulting in significant blending of galaxy surface brightness profiles, and voluminous, thus statistically powerful enough to have very strict accuracy requirements.

In this work we present realistic image simulations designed to accurately calibrate many of the complexities in analysing real imaging data. By using a combination of morphological information

from HST and flux information from DES Deep Fields, we make use of an input galaxy sample with realistic joint distributions of flux (in multiple bands) and morphology. This, along with careful simulation of the observational characteristics of real DES Y3 data, gives us confidence that measurements of biases can be reliably applied as calibration corrections. This multi-band weak lensing calibration simulation allows us to perform the same photometric redshift inference as is performed on the real data, allowing reliable and consistent corrections for each photometric redshift bin.

While we believe these simulations are sufficiently realistic given the requirements of the DES Year 3 cosmology analyses, they do of course involve approximations and shortcuts that may not be sufficiently accurate for future analyses. We do not simulate the estimation of the WCS solution, PSF, background levels or noise levels in our simulations, under the assumption that any errors in our estimation of these in the real data are negligible. These assumptions should be carefully re-examined for more precise upcoming weak lensing datasets.

In addition, we simulated galaxies with random sky positions, rather than a realistic level of clustering. While our re-weighting procedure in Section 6.3 can increase the number of pairs of close detected objects, it does not explicitly match the number of groups of more than two close detected objects, however, at current number densities, we expect those to be much less frequent. The re-weighting also cannot impact the statistics of undetected objects, and so will not change the frequency of detected galaxies having close, undetected neighbours. This specific case is investigated in detail by Euclid Collaboration (2019), who demonstrate that for Euclid-like data, the clustering of faint galaxies around the bright ones used for shape measurement can induce a multiplicative bias of $\sim -0.5\%$ (with respect to the case where the faint galaxies are unclustered). Since this is of the same order as our statistical uncertainties, it is worth careful consideration.

We firstly note that the only conclusive way to judge the impact of the clustering of undetected galaxies is to simulate it. However, generating a weak lensing calibration image simulation with realistic clustering statistics, as well as the realistic distributions of measured properties demonstrated here (Section 3.6) is an ongoing challenge. Current cosmological simulations which include the former have not been demonstrated to produce the high level of agreement with real data at the image level that is required for weak lensing calibration. In the absence of such a simulation, we make some arguments that the size of the biases estimated by Euclid Collaboration (2019) are likely an upper limit on the size of biases likely to be generated by clustering of undetected objects in DES Year 3.

Firstly, the DES Year 3 data is shallower than the Euclid-like simulations in Euclid Collaboration (2019), hence in general blending-related biases are lessened. Secondly, the "faint" galaxies in Euclid Collaboration (2019) are defined as having $S/N < 10$. We include all objects detected by SExtractor (and with positive flux) in our re-weighting procedure, which therefore only misses out on the $S/N << 10$ objects that are undetected by SExtractor. Thirdly, we are using Metacalibration, which has some robustness to blending - for very close pairs which are always detected as a single object, Metacalibration performs extremely well (as demonstrated in Sheldon et al. (2020)), because it calibrates the response of the measurement to shear of all the light in the system. This is not the case for the shape measurement methods used in Euclid Collaboration (2019), which we would expect to be more sensitive to "model bias" arising from the unusual apparent morphology of blended systems. Similarly, Metacalibration's insensitivity to noise bias means faint objects that are always undetected

are unlikely to cause significant biases; their contribution behaves largely as extra noise.

While these arguments are somewhat based on intuition rather than extensive simulation results, given constraints on time and computing resources, we think they justify our approach and priorities for the calibration of this intermediate dataset. Nonetheless, we recommended that Amon et al. (2021) include a robustness test where an additional multiplicative bias with prior width 1% was marginalized over in the cosmological inference from cosmic shear. This extra bias was assumed to be coherent across redshift bins, and represents a pessimistic interpretation of the potential bias arising from the effect explored by Euclid Collaboration (2019). They find negligible change in the constraint on $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$, likely due to the dominance of other sources of systematic uncertainty.

Another potential limitation of our simulations is the limited size of our input galaxy catalog (see e.g. Kannawadi et al. 2015 for an investigation of the impact of limited input catalog size in weak lensing image simulations), which came from the COSMOS region, motivated by the availability of HST imaging which provides precise morphological information for our input galaxies. While our detailed comparisons of the simulation outputs and the real DES Year 3 data suggest that cosmic variance in the input catalog is not a huge effect (and this conclusion agrees with that of Kannawadi et al. 2019), we would recommend investigations of the possibility of larger reliable input catalogs from deep ground-based imaging only (as measured by Hartley & Choi et al., 2020 and used in injection simulations by Everett et al. 2020). The planned LSST deep-drilling fields would be a valuable source of such data.

Conceptually, we have stressed the importance of characterizing biases due to blending via their impact on the effective redshift distribution. When blending becomes an important contributor to shear calibration biases, photo-z and shear inference can no longer be cleanly decoupled. Folding both photo-z and shear calibration biases into an effective redshift distribution for lensing, $n_\gamma(z)$, is both a compact approach, and very general in that $n_\gamma(z)$ is the key observational quantity required to make theoretical predictions for most weak lensing signals of interest. The normalization of $n_\gamma(z)$ corresponds to the traditional mean multiplicative bias, $1 + m$.

Via simulations with redshift-dependent shear signals, we have provided a methodology for directly estimating $n_\gamma(z)$ integrated over finite redshift intervals. The key point in this approach is that the value of the effective redshift distribution at a given redshift is defined by the response of the ensemble to shear at that redshift. Therefore only by simulating such a shear signal can the effective redshift distribution be estimated robustly. We believe that with the increased depth (and therefore increased blending) and stricter accuracy requirements of future surveys such as the Rubin Observatory and Euclid, such an approach will become essential i.e. $n_\gamma(z)$ is the quantity that must be accurately characterized for reliable cosmological inference, which is a requirement beyond the traditional approach of characterizing the multiplicative bias m.

In Section 6 we propagated our simulation measurements through to corrections for the DES Y3 redshift distributions, producing from an input ensemble of redshift distributions, an output ensemble of effective redshift distributions, $n_\gamma(z)$ ready for use in the Year 3 weak lensing cosmology analyses. This output ensemble includes both statistical (due to finite simulation volume) and systematic (due to potential simulation inaccuracies) uncertainties. These $n_\gamma(z)$ should be used directly in the theoretical predictions for weak lensing statistics like the shear correlation functions and galaxy-galaxy lensing tangential shear, as the effective redshift distribution entering the lensing kernel. Note that the (in general non-

unity) normalization of the effective redshift distributions should be retained in the calculations, or applied to the resulting theory prediction if using a numerical code that normalizes the redshift distribution internally. This data will be made available along with the DES Year 3 shear catalogs, on publication of the key Year 3 cosmology analyses, as part of the DES Y3 coordinated release[11].

8 DATA AVAILABILITY

All DES Year 3 cosmology catalogs, including calibrations, will be made available at https://www.darkenergysurvey.org/the-des-project/data-access/ on publication of the key cosmological results.

REFERENCES

Albrecht A., et al. 2006, arXiv e-prints, pp astro–ph/0609591
Amon A., et al. 2021, arXiv e-prints, p. arXiv:2105.13543
Baugh C. M., Efstathiou G., 1993, MNRAS, 265, 145
Berberan-Santos M., 2007, Journal of Mathematical Chemistry, 42, 585
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bernstein G. M., et al. 2016, MNRAS, 459, 4467
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bertin E., et al. 2002, in Bohlender D. A., Durand D., Handley T. H., eds, Astronomical Society of the Pacific Conference Series Vol. 281, Astronomical Data Analysis Software and Systems XI. p. 228
Brainerd T. G., Blandford R. D., Smail I., 1996, ApJ, 466, 623
Bridle S., et al. 2010, MNRAS, 405, 2044
Buchs R., et al. 2019, MNRAS, 489, 820
Cordero J. P., et al. 2021, arXiv e-prints, p. arXiv:2109.09636
DES Collaboration 2021, arXiv e-prints, p. arXiv:2105.13549
Dawson W. A., et al. 2016, ApJ, 816, 11
Euclid Collaboration 2019, A&A, 627, A59
Everett S., et al. 2020, arXiv e-prints, p. arXiv:2012.12825
Fenech Conti I., et al. 2017, MNRAS, 467, 1627
Flaugher B., et al. 2015, AJ, 150, 150
Foreman-Mackey D., et al. 2013, PASP, 125, 306

Gatti M., et al. 2020, arXiv e-prints, p. arXiv:2012.08569
Gatti M., et al. 2021, MNRAS, 504, 4312
Hartley W. G., et al. 2020, arXiv e-prints, p. arXiv:2012.12824
Heymans C., et al. 2006, MNRAS, 368, 1323
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hoekstra H., Viola M., Herbonnet R., 2017, MNRAS, 468, 3295
Hoyle B., et al. 2018, MNRAS, 478, 592
Huff E., Mandelbaum R., 2017, arXiv e-prints, p. arXiv:1702.02600
Jarvis M., et al. 2016, MNRAS, 460, 2245
Jarvis M., et al. 2021, MNRAS, 501, 1282
Jones D. M., Heavens A. F., 2019, MNRAS, 483, 2487
Kacprzak T., et al. 2012, MNRAS, 427, 2711
Kacprzak T., et al. 2020, Phys. Rev. D, 101, 082003
Kaiser N., 2000, The Astrophysical Journal, 537, 555
Kannawadi A., Mandelbaum R., Lackner C., 2015, MNRAS, 449, 3597
Kannawadi A., et al. 2019, A&A, 624, A92
Kitching T. D., et al. 2013, ApJS, 205, 12
Kitching T. D., et al. 2019, The Open Journal of Astrophysics, 2, 5
Koekemoer A. M., et al. 2007, ApJS, 172, 196
Laigle C., et al. 2016, ApJS, 224, 24
Leauthaud A., et al. 2007, ApJS, 172, 219
Mandelbaum R., et al. 2014, ApJS, 212, 5
Mandelbaum R., et al. 2015, MNRAS, 450, 2963
Mandelbaum R., et al. 2018, MNRAS, 481, 3170
Massey R., et al. 2007, MNRAS, 376, 13
Morganson E., et al. 2018, PASP, 130, 074501
Myles J., et al. 2021, MNRAS, 505, 4249
Nakajima R., Bernstein G., 2007, AJ, 133, 1763
Padmanabhan N., et al. 2019, arXiv e-prints, p. arXiv:1903.01571
Paulin-Henriksson S., et al. 2008, A&A, 484, 67
Pieres A., et al. 2020, MNRAS, 497, 1547
Pujol A., et al. 2019, A&A, 621, A2
Refregier A., et al. 2012, MNRAS, 425, 1951
Rowe B. T. P., et al. 2015, Astronomy and Computing, 10, 121
Samuroff S., et al. 2018, MNRAS, 475, 4524
Scoville N., et al. 2007, ApJS, 172, 38
Sevilla-Noarbe I., et al. 2021, ApJS, 254, 24
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Sheldon E. S., et al. 2020, ApJ, 902, 138
Voigt L. M., Bridle S. L., 2010, MNRAS, 404, 458
Weinberg D. H., et al. 2013, Phys. Rep., 530, 87
Zuntz J., et al. 2018, MNRAS, 481, 1149

config. #	gal. mag.	gal. size	PSF model	masking effects	outlier rej.	mcal PSF	m ($\times 10^{-3}$)	c ($\times 10^{-5}$)
1	17	0.5″	Gaussian	no	no	gauss-fit	g1: 0.225 ± 0.005	0.1 ± 0.1
							g2: 0.440 ± 0.004	-0.1 ± 0.1
2	17	variable	Gaussian	no	no	gauss-fit	g1: 0.314 ± 0.007	-0.2 ± 0.2
							g2: 0.415 ± 0.007	-0.1 ± 0.02
3	17	variable	Gaussian	no	no	symmetrize	g1: 0.381 ± 0.009	0.1 ± 0.2
							g2: 0.383 ± 0.009	-0.4 ± 0.5
4a	17	variable	Piff	no	no	gauss-fit	g1: 0.219 ± 0.025	-0.1 ± 0.3
							g2: 0.294 ± 0.127	0.1 ± 0.3
4	17	variable	Piff	no	no	symmetrize	g1: 0.405 ± 0.075	0.2 ± 0.3
							g2: 0.414 ± 0.014	1.4 ± 0.4
5	17	variable	Piff	yes	no	symmetrize	g1: -1.95 ± 0.30	0.6 ± 0.4
							g2: -1.94 ± 0.29	6.8 ± 1.0
6	18	variable	Piff	no	yes	symmetrize	g1: 8.56 ± 0.16	0.0 ± 0.9
							g2: 8.70 ± 0.13	-9.5 ± 1.0
7	18.75	variable	Piff	no	yes	symmetrize	g1: 3.79 ± 0.33	-2.0 ± 2.3
							g2: 3.78 ± 0.13	-2.3 ± 1.9
8	19.5	variable	Piff	no	yes	symmetrize	g1: 2.44 ± 0.16	-2.8 ± 2.4
							g2: 2.50 ± 0.11	-1.0 ± 2.6
9	20.5	variable	Piff	no	yes	symmetrize	g1: 0.89 ± 0.29	-5.9 ± 6.1
							g2: 0.56 ± 0.26	0.1 ± 6.6

Table B1. Validation simulation configurations and results. All simulations place the galaxies on a grid and use "true detection" where the shear measurement code is fed the true position of the galaxy in the image. We then run variations with and without bad pixel masking and corrections ("masking effects"), pixel-level outlier rejection ("outlier rej."), the galaxy and PSF models ("gal. mag.", "gal. size", and "PSF model"), and the Metacalibration treatment of the PSF model ("mcal PSF"). The last two columns list the multiplicative and additive biases we find in these simulation setups. See the text in Appendix B for details.

APPENDIX A: SHEAR AND REDSHIFT BIAS PARAMETERS USED PRIOR TO UNBLINDING THE ANALYSIS

A final technical improvement to the priors on m and $\delta\bar{z}$ was completed in parallel with the unblinding of the DES Y3 3x2pt cosmology analysis. These updates were not expected to impact the analysis in a significant way. The original chains at unblinding used a different (but very similar) set of m and $\delta\bar{z}$ priors. Gaussian priors on the per-bin m were marginalized over with mean, μ and width, σ for the four photo-z bins given by $(\mu, \sigma) = [(-0.01044, 0.00576), (-0.01579, 0.00414), (-0.02489, 0.00532), (-0.03802, 0.00801)]$. The mean of the $n_\gamma(z)$ samples generated using our fiducial correction was used, with a Gaussian $\delta\bar{z}$ prior estimated from their statistical scatter only, with width $[0.018, 0.013, 0.006, 0.013]$.

DES Collaboration (2021) describe the unblinding criteria and procedure in detail, and conclude that updating the priors had negligible impact on the final cosmological constraints, and would not have affected the unblinding criteria (and therefore the decision to unblind).

APPENDIX B: SIMULATION VALIDATION TESTS

One of the main concerns when working with the complex simulation suite presented in this work is that the simulations themselves could contain bugs which would cause non-zero multiplicative or additive biases. To guard against this possibility, we tested the performance of our pipelines and the simulations in configurations where previous studies have demonstrated sub-0.1% performance.

In these simulations, all objects have exponential light profiles. We fix the objects to have the same magnitude in each band, but different simulations vary the magnitude overall. The galaxies have a half-light radius that is either fixed to 0.5″ or allowed to vary according to the input catalogs for the more realistic simulations, although in the latter case, we limit the half-light radius to at most 0.8″ in order to prevent adjacent objects from overlapping. We test both a Gaussian PSF with a FWHM of 0.9″ and the smoothed Piff PSF models. We also vary whether or not we apply the bad pixel masks from the data to the simulations (and then correct for them). See Sevilla-Noarbe et al. (2021) and Gatti & Sheldon et al., (2021) for more details on the bad pixel masks and how our measurement codes correct for them. Further, we vary whether or not we perform pixel-level outlier rejection on the individual postage stamps for each object. Finally, we also vary a detail of how Metacalibration handles the computation of the PSF it uses in its reconvolution step. These variants are denoted internally as symmetrize or gauss-fit. The simulation configurations we tested are listed in Table B1.

The measurements for m and c from these simulations are in the last two columns of Table B1. We find that in the simplest case (configuration 1), we recover the applied shear to better than a part in a thousand with no evidence of any additive bias. The same conclusion holds for configurations 2, 3, and 4a which vary the galaxy sizes, the PSF model, and the Metacalibration reconvolution PSF treatment. Comparing configurations 4a and 4, we find that galaxy size variation coupled with the symmetrize Metacalibration PSF treatment introduces a small additive error in $g2$, but one that is too small to impact cosmology. Configuration 5, which introduces pixel masking effects relative to configuration 4, demonstrates that our masking corrections cause a small $\sim -0.2\%$ multiplicative bias. This effect is well below our requirements and is subdominant relative to detection and/or blending, as described above. Configurations 6, 7, 8, and 9 use the pixel-level outlier rejection and employ a series of fainter objects from magnitude 17 to magnitude 20.5. We find that extremely bright objects exhibit some non-trivial multiplicative bias in this case. This effect was traced back to Poisson noise from the object's flux falsely triggering the pixel-level outlier rejection (which only considered the sky noise in its threshold). In the data, these bright objects are extremely rare. Further, this effect is correctly modeled in our simulations for any objects that do eventually pass the cuts we employ on the shear catalogs.
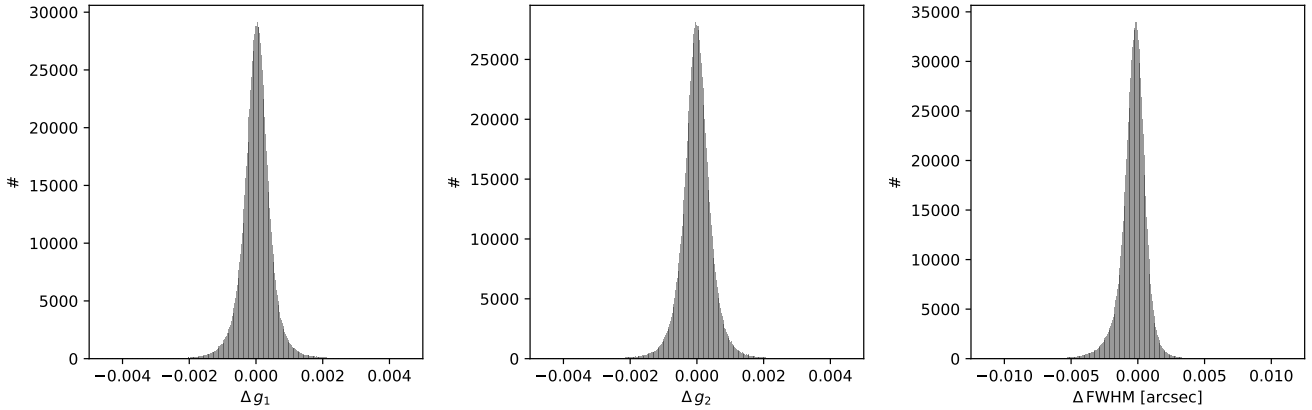
Figure C1. Performance of the smoothed `Piff` PSF models. Each panel from left to right shows a histogram of the differences in the shape (g_1 or g_2) or FWHM of the smoothed PSF model versus the underlying `Piff` PSF model.

APPENDIX C: `PIFF` PSF MODEL SMOOTHING FOR SIMULATIONS

The simulations in this work use the observed `Piff` PSF models. We found a bug in the process of doing the DES Y3 analysis, detailed in Jarvis et al. (2021, §3.2), which causes some images of the `Piff` PSF model to have what is effectively extra background noise. In tests of the simulations, we found that when using the model images as the true PSF for the simulation, sometimes this extra noise caused shear recovery biases with METACALIBRATION even in cases where that method is known to work at extremely high precision. In order to avoid any potential problems in the simulations presented here, we applied the following smoothing procedure to the `Piff` PSF models. For each single-epoch image, we fit the rendered image of the `Piff` model with an `ngmix turb` model (which is an approximation to a Kolmogorov-like PSF for a long exposure taken through atmospheric turbulence). This fit is done over a grid of locations on the CCD and the parameters of it are interpolated. We then use these interpolated fit parameters to draw the image of the PSF model at any location on the CCD. We include a small correction for the size of pixel in the fit parameters in order to more closely match the output PSF sizes when the PSF is drawn.

Figure C1 shows the differences in the shapes and FWHM for the raw `Piff` models and the rendered smooth models. To make these histograms, we selected 10^4 CCDs at random, fit our smoothed, interpolated models, and then drew images of both the smooth model and the interpolated one at 10 random points per CCD. We find an extremely small negative bias in the FWHM, but otherwise the models provide a good approximation to the true PSF size and shape variations.

APPENDIX D: THE EFFECT OF SMOOTH RESPONSE WEIGHTING ON REDSHIFT DISTRIBUTIONS

As previously noted in Section 2.3, the redshift distributions are weighted by the METACALIBRATION response. It was found in general that the noise in response measurement could result in unrealistic redshift distributions, where negatively measured responses could lead to an unphysical negative $N(z)$ at several points. Noise in response measurement then, tracks directly into N(z) scatter.

The DES Year 3 analysis resolves this problem by smoothing the same two sets of weights (shear and response) over a log grid in signal to noise `mcal_s2n_r` and the size ratio with respect to the point spread function, T_{mcal}/T_{psf}. One such grid, generated from the mesh average in the data is plotted in Fig. D1. For each detection in the image simulations, we assign it to a cell on this grid of SNR and size ratio, and take the mean response of all detections in that cell. This mean response (or *smooth response*) is then used when weighting the redshift distribution.

For this weighting scheme to be valid, we tested the shift in mean redshift for each tomographic bin in the image simulations from the reweighting, as this data set has a very well understood input redshift distribution. The shift applied as a result of smoothing in each bin was on average 0.0025 in mean redshift, with the first bin being the most affected with a shift of 0.004, and the third being the least affected at 0.0002. Response weighting as a whole shifts the mean redshift in each bin by an order of magnitude larger, so the majority of the effect is preserved. In conclusion, given the reduction in $N(z)$ scatter achieved with a smooth response weight and the small amplitude of bias ($\sim 10^{-3}$), this approximation can be made when evaluating the Year 3 data.

APPENDIX E: AFFILIATIONS

[1] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK

[2] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

[3] Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

[4] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA

[5] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

[6] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
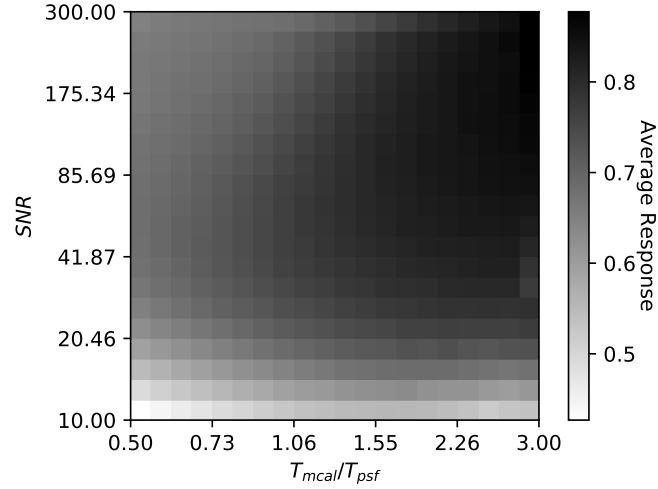
Figure D1. The smooth response grid generated from the DES Y3 catalog shows correlation between mean response and size ratio, and mean response and signal to noise. This 20 x 20 log-spaced grid supplied the smoothed weighting scheme evaluated in App. D.

[7] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

[8] Department of Physics, Duke University Durham, NC 27708, USA

[9] Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA

[10] Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA

[11] Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA

[12] Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK

[13] Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

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

[15] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany

[16] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

[17] National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA

[18] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

[19] Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Geneva, Switzerland

[20] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

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

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

[23] Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA

[24] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA

[25] Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland

[26] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain

[27] Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile

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

[29] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK

[30] CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France

[31] Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France

[32] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

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

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

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

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

[37] Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

[38] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain

[39] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

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

[41] Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany

[42] Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA

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

[44] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

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

[46] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

[47] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[48] Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[49] School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

[50] Department of Physics, The Ohio State University, Columbus, OH 43210, USA

[51] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

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

[53] Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA

[54] Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138

[55] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

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

[57] Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390

[58] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK

[59] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

[60] Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK

This paper has been typeset from a TEX/LATEX file prepared by the author.