Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527,** L115–L121 (2024) https://doi.org/10.1093/mnrasl/slad143
Advance Access publication 2023 October 10

Detection of the significant impact of source clustering on higher order statistics with DES Year 3 weak gravitational lensing data

M. Gatti,[1★] N. Jeffrey,[2] L. Whiteway,[2] V. Ajani,[3] T. Kacprzak,[3] D. Zürcher,[3] C. Chang,[4,5] B. Jain,[1] J. Blazek,[6] E. Krause,[7] A. Alarcon,[8] A. Amon,[9,10] K. Bechtol,[11] M. Becker,[8] G. Bernstein,[1] A. Campos,[12] R. Chen,[13] A. Choi,[14] C. Davis,[15] J. Derose,[16] H. T. Diehl,[17] S. Dodelson,[12,18] C. Doux,[19] K. Eckert,[1] J. Elvin-Poole,[20] S. Everett,[21] A. Ferte,[22] D. Gruen,[23] R. Gruendl,[24,25] I. Harrison,[26] W. G. Hartley,[27] K. Herner,[17] E. M. Huff,[21] M. Jarvis,[1] N. Kuropatkin,[17] P. F. Leget,[15] N. MacCrann,[28] J. McCullough,[15] J. Myles,[15,22,29] A. Navarro-Alsina,[30] S. Pandey,[1] J. Prat,[4,5] M. Raveri,[31] R. P. Rollins,[32] A. Roodman,[15,22] C. Sanchez,[1] L. F. Secco,[5] I. Sevilla-Noarbe,[33] E. Sheldon,[34] T. Shin,[35] M. Troxel,[36] I. Tutusaus,[37,38,39] T. N. Varga,[40,41,42] B. Yanny,[17] B. Yin,[12] Y. Zhang,[43,44] J. Zuntz,[45] S. S. Allam,[17] O. Alves,[46] M. Aguena,[47] D. Bacon,[48] E. Bertin,[49,50] D. Brooks,[2] D. L. Burke,[15,22] A. Carnero Rosell,[47,51,52] J. Carretero,[53] R. Cawthon,[54] L. N. da Costa,[47] T. M. Davis,[55] J. De Vicente,[33] S. Desai,[56] P. Doel,[2] J. García-Bellido,[57] G. Giannini,[4] G. Gutierrez,[17] I. Ferrero,[58] J. Frieman,[5,17] S. R. Hinton,[55] D. L. Hollowood,[59] K. Honscheid,[60,61] D. J. James,[62] K. Kuehn,[63,64] O. Lahav,[2] J. L. Marshall,[65] J. Mena-Fernández,[33] R. Miquel,[53,66] R. L. C. Ogando,[67] A. Palmese,[12] M. E. S. Pereira,[68] A. A. Plazas Malagón,[15,22] M. Rodriguez-Monroy,[33] S. Samuroff,[6] E. Sanchez,[33] M. Schubnell,[46] M. Smith,[69] F. Sobreira,[30,47] E. Suchyta,[70] M. E. C. Swanson,[2] G. Tarle,[46] N. Weaverdyck,[16,46] and P. Wiseman[69] (DES Collaboration)

Affiliations are listed at the end of the paper

Accepted 2023 October 2. Received 2023 September 23; in original form 2023 August 17

ABSTRACT

We measure the impact of source galaxy clustering on higher order summary statistics of weak gravitational lensing data. By comparing simulated data with galaxies that either trace or do not trace the underlying density field, we show that this effect can exceed measurement uncertainties for common higher order statistics for certain analysis choices. We evaluate the impact on different weak lensing observables, finding that third moments and wavelet phase harmonics are more affected than peak count statistics. Using Dark Energy Survey (DES) Year 3 (Y3) data, we construct null tests for the source-clustering-free case, finding a p-value of $p = 4 \times 10^{-3}$ (2.6σ) using third-order map moments and $p = 3 \times 10^{-11}$ (6.5σ) using wavelet phase harmonics. The impact of source clustering on cosmological inference can be either included in the model or minimized through ad hoc procedures (e.g. scale cuts). We verify that the procedures adopted in existing DES Y3 cosmological analyses were sufficient to render this effect negligible. Failing to account for source clustering can significantly impact cosmological inference from higher order gravitational lensing statistics, e.g. higher order N-point functions, wavelet-moment observables, and deep learning or field-level summary statistics of weak lensing maps.

Key words: cosmology: observations.

1 INTRODUCTION

Weak gravitational lensing from large-scale structures in the Universe induces small distortions in the observed shape of background source galaxies. The weak lensing signal can be measured using large samples of galaxies to observe correlated distortions in observed galaxy ellipticities (see Bartelmann & Schneider 2001). The angular distribution of source galaxies is not uniform; it is modulated by observational and selection effects (such as varying observing depth) and by clustering due to galaxies tracing the underlying density field. The latter effect, called 'source clustering' (Schneider, van Waerbeke & Mellier 2002; Schmidt et al. 2009; Valageas 2014; Krause et al. 2021), causes the galaxy number density to be correlated with the target lensing signal: since we expect a larger lensing signal along overdense lines of sight, we preferentially sample the shear field where its value is larger. For pixelized shear maps, this results in two distinct effects: (1) the average noise-free lensing signal

★ E-mail: marcogatti29@gmail.com

is modulated by a different effective redshift distribution, and (2) the 'shape noise' (due to the intrinsic ellipticities of galaxies) is correlated with the lensing signal.

Higher order statistics have recently been growing in popularity as powerful tools for efficiently extracting cosmological information from current weak lensing data (e.g. Liu et al. 2015; Vicinanza et al. 2016; Martinet et al. 2018; Fluri et al. 2019; Cheng et al. 2020; Jeffrey, Alsing & Lanusse 2021a; Gatti et al. 2022; Lu, Haiman & Li 2023; Zürcher et al. 2023). Their use can improve constraints on cosmological parameters (relative to standard two-point statistics), can help discriminate between general relativity and modified gravity theories (Cardone et al. 2013; Peel et al. 2018), and can help self-calibrate astrophysical and observational nuisance parameters (Pyne & Joachimi 2021). Given the increasing precision of these measurements, the impact of systematic errors on higher order statistics is a subject of careful consideration.

The impact of source clustering has generally been neglected in the forward model, as it has often been considered a small, higher order contribution to weak lensing observables. The efficiency of lensing peaks roughly halfway between the source and the observer, and vanishes at the source location; any correlation between the shear field 'seen' by the source galaxy and the density field it lives on is suppressed. Source clustering has been studied in the context of two-point correlation functions, and theoretical calculations by Krause et al. (2021) have shown it to be negligible for Stage III surveys for catalogue-based Gaussian statistics. Whether its impact on weak lensing higher order statistics is also negligible is less clear, although some early estimates suggested a stronger impact on three-point correlation functions (Valageas 2014). The effect of source clustering has not to date been explicitly included in the suites of simulations used for simulation-based cosmological analyses (e.g. Martinet et al. 2018; Zürcher et al. 2021), although a few peak statistics analyses (Kacprzak et al. 2016; Harnois-Déraps et al. 2021; Zürcher et al. 2021) performed initial tests of this effect (under some simplifying assumptions), showing no significant effect on their cosmological constraints.

This work develops a forward-modelling procedure to introduce source clustering effects into the simulated maps. We consider the impact of source clustering on several non-Gaussian observables, looking primarily at map-based estimators. We show that source clustering generates a clear signature on higher order summary statistics for specific analysis choices, we demonstrate this effect in the Dark Energy Survey (DES) Year 3 (Y3) data, and we discuss the impact of this effect on previously published DES measurements.

2 DATA AND SIMULATIONS

2.1 DES Y3 weak lensing catalogue

We use the DES Y3 weak lensing catalogue (Gatti et al. 2021); it consists of 100 204 026 galaxies, with a weighted $n_{\rm eff} = 5.59$ galaxies arcmin^{-2}, over an effective area of 4139 deg^2. It was created using the METACALIBRATION algorithm (Huff & Mandelbaum 2017; Sheldon & Huff 2017), which provides self-calibrated shear estimates starting from (multiband) noisy images of the detected objects. A residual small calibration (via a multiplicative shear bias) is provided; based on sophisticated image simulations (MacCrann et al. 2022), it accounts for blending-related detection effects. An inverse variance weight is further assigned to each galaxy in the catalogue to enhance the overall signal-to-noise ratio. The sample is divided into four tomographic bins of roughly equal number density (Myles et al. 2021). Redshift distributions are provided by the SOMPZ method, in combination with clustering redshift constraints (Myles et al. 2021).

2.2 Simulations

We rely on simulations produced using the PKDGRAV3 code (Potter, Stadel & Teyssier 2017).We use 50 independent realizations at the fixed cosmology $\Omega_{\rm m} = 0.26$, $\sigma_8 = 0.84$, $\Omega_{\rm b} = 0.0493$, $n_{\rm s} = 0.9649$, and $h = 0.673$ from the DARKGRIDV1 suite, described in detail in Zürcher et al. (2021, 2022). All simulations include three massive neutrino species with a mass of $m_\nu = 0.02$ eV per species. The simulations were obtained using 14 replicated boxes in each direction (14^3 replicas in total) so as to span the redshift interval from $z = 0$ to $z = 3$. Each individual box contains 768^3 particles and has a side-length of 900 h^{-1} Mpc. For each simulation, lens planes $\delta_{\rm shell}(\hat{\boldsymbol{n}}, \chi)$ are provided at ∼87 redshifts from $z = 3$ to $z = 0$. The lens planes are provided as HEALPIX (Górski et al. 2005) maps and are obtained as the overdensity of raw number particle counts; for this work, we downsample the original resolution of NSIDE = 2048 to NSIDE = 1024 (with pixel size ≈3.4 arcmin). The lens planes are converted into convergence planes $\kappa_{\rm shell}(\hat{\boldsymbol{n}}, \chi)$ under the Born approximation (e.g. equation 2 from Fosalba et al. 2015). Lastly, shear planes $\gamma_{\rm shell}(\hat{\boldsymbol{n}}, \chi)$ are obtained from the convergence maps using a full-sky generalization of the Kaiser & Squires (1993) algorithm (Jeffrey et al. 2021b).

3 SOURCE CLUSTERING IMPLEMENTATION

In the limit of high source galaxy density, the observed projected shear in direction $\boldsymbol{\theta}$ will be

$$\gamma(\boldsymbol{\theta}) = \frac{\int n(\boldsymbol{\theta}, z)\, \gamma(\boldsymbol{\theta}, z)\, \mathrm{d}z}{\int n(\boldsymbol{\theta}, z)\, \mathrm{d}z}, \tag{1}$$

where $n(\boldsymbol{\theta}, z)$ is the unnormalized galaxy density (i.e. $\int_V n(\boldsymbol{\theta}, z)\, \mathrm{d}\boldsymbol{\theta}\mathrm{d}z$ is the number of source galaxies in the volume V). The observed shear γ is the sum of signal γ_s and noise ϵ_n:

$$\gamma(\boldsymbol{\theta}) = \frac{\int n(\boldsymbol{\theta}, z)\Big(\gamma_s(\boldsymbol{\theta}, z) + \epsilon_n(\boldsymbol{\theta}, z)\Big)\, \mathrm{d}z}{\int n(\boldsymbol{\theta}, z)\, \mathrm{d}z} = \gamma_s(\boldsymbol{\theta}) + \gamma_n(\boldsymbol{\theta}). \tag{2}$$

It has been standard in many previous analyses to use the spatial average

$$\bar{n}(z) = \frac{\int n(\boldsymbol{\theta}, z)\, \mathrm{d}\boldsymbol{\theta}}{\int \mathrm{d}\boldsymbol{\theta}} \tag{3}$$

as an approximation to $n(\boldsymbol{\theta}, z)$; however, this approximation cannot include the effect of source clustering. We instead model the directional variation of the source galaxy distribution arising from its dependence on the overdensity field $\delta(\boldsymbol{\theta}, z)$, i.e. $n(\boldsymbol{\theta}, z) = \bar{n}(z)\,[1 + f(\delta(\boldsymbol{\theta}, z))]$ for some function f. This leads to a relation between $n(\boldsymbol{\theta}, z)$ and the observed shear $\gamma(\boldsymbol{\theta}, z)$, as they both depend on δ. This relation has a direct impact on the expected value γ_s (i.e. the signal is modulated). Additionally, as the variance of the noise term γ_n depends on n (more source galaxies lead to reduced noise), this relation will have an impact on the expected value of terms such as $\gamma_s\gamma_n^2$. A simulation that does not include source clustering effects is in danger of incorrectly modelling these expected values.

Below, we describe how to create pixelized shear maps both without and with source clustering effects. We consider one fixed tomographic bin. We assume as inputs a noiseless pixelized simulated shear map and a separate galaxy shape catalogue. The latter is needed to supply shape noise information (as the simulated shear map is not

assumed to have an associated simulated galaxy catalogue); in our case, the DES Y3 shape catalogue serves this purpose. We then add a source clustering effect by amending both signal and noise terms using factors related to the matter overdensity in the shear simulation.

An alternative method for creating shear simulations with source clustering would be to use the results of the N-body simulation (i.e. the simulation used to create the simulated shear field) to directly create a galaxy catalogue (using some Haloc Occupation Distribution prescription, for example), to assign shape noise to these galaxies, and to use this information to add noise to the shear simulation.However, this task is complex, and therefore we opt for the simpler approach implemented in this work.

In what follows, let p be a pixel, s a thin redshift shell, $\gamma(p, s)$ the noiseless shear from the shear simulation, and $\bar{n}(s)$ the galaxy count across the whole footprint (Myles et al. 2021). From the galaxy catalogue, let g denote a galaxy, w_g its weight, and e_g its ellipticity after the application of a random rotation to erase the shear signal.

3.1 Mock shear maps with no source clustering

The output simulated shear for a given pixel p is the sum of signal and shape noise contributions:

$$\gamma(p) = \frac{\sum_s \bar{n}(s)\gamma(p,s)}{\sum_s \bar{n}(s)} + \frac{\sum_g w_g e_g}{\sum_g w_g}. \tag{4}$$

In the signal term the sum is over all shells s, and in the noise term the sum is over all the shape catalogue (i.e. DES Y3) galaxies g in p.

3.2 Mock shear maps with source clustering

Let $\delta(p, s)$ be the matter overdensity in the shear simulation. Let b_g be the galaxy-matter bias; for simplicity we assume linear biasing to hold and moreover for our main tests we assume $b_g = 1$ (a reasonable choice for the blue field galaxies that constitute most of the galaxies in the shear catalogue). The factor $\bar{n}(s)\left[1 + b_g\delta(p, s)\right]$ is then the relative galaxy count in pixel p and shell s; it is generated from the shear simulation and is therefore consistent with the shear signal. In the output simulated shear, both the signal and the shape noise contributions have been amended to account for source clustering as follows:

$$\gamma_{\rm SC}(p) = \frac{\sum_s \bar{n}(s)\left[1 + b_g\delta(p,s)\right]\gamma(p,s)}{\sum_s \bar{n}(s)\left[1 + b_g\delta(p,s)\right]} + \left(\frac{\sum_s \bar{n}(s)}{\sum_s \bar{n}(s)\left[1 + b_g\delta(p,s)\right]}\right)^{1/2} F(p)\frac{\sum_g w_g e_g}{\sum_g w_g}. \tag{5}$$

The signal term is a weighted average over shells; here, the average has been amended to include a shear-correlated source galaxy count. In the shape noise term, there are two additional factors. The first, a source clustering factor, results in the shape noise variance scaling as the inverse of the relative galaxy count, as desired; this gives a correlation between the shear signal in a pixel and the square of the shape noise that was not present before. The second, $F(p)$, is a near-unity scale factor introduced to avoid double-counting source clustering effects. The DES Y3 catalogue used to model the shape noise of the pixels is already affected by source clustering. In practice, this means that the noise of the catalogue is already modulated by $1/\sqrt{\sum_s \bar{n}(s)\left[1 + b_g\delta_{\rm data}(p, s)\right]}$. This modulation is not correlated with the large-scale structure of the simulations. However, since equation (5) introduces a similar modulation, the net effect is that the even moments of the pixel's simulated noise (variance, kurtosis, etc.) are slightly enhanced with respect to data, mostly at small scales and low redshifts. The function $F(p)$ corrects this enhancement. We opted for the following expression:

$$F(p) = A\sqrt{1 - B\sigma_e^2(p)}, \tag{6}$$

where the coefficients A and B are per-bin constants, and $\sigma_e^2(p)$ is the variance of the pixel noise. This correction is (only mildly) cosmology dependent; we used our simulations at fixed cosmology to estimate the two sets of constants for the four bins: $A = [0.97, 0.985, 0.990, 0.995]$ and $B = [0.1, 0.05, 0.035, 0.035]$.

3.3 Remarks concerning our implementation

We generate shear maps for each tomographic bin. The 50 independent simulations at fixed cosmology for our main tests yield 200 independent simulated DES Y3 shear catalogues (as we can cut four independent DES Y3 footprints from each full-sky map). The simulations have not been run at the best-fitting cosmology for the data. However, based on the results presented in Gatti et al. (2022), the cosmology chosen for the simulations should still provide a reasonable fit to the data. Moreover, for simplicity, we did not include any intrinsic alignments and we assumed zero shear and redshift biases; we do not expect this to affect any of our conclusions.

4 RESULTS

In this work, we consider the following summary statistics:

(i) **Second and third map moments**: second moments are a Gaussian statistic (i.e. a function only of the power spectrum), whereas third moments probe additional non-Gaussian features of the field (Van Waerbeke et al. 2013; Petri et al. 2015; Chang et al. 2018; Peel et al. 2018; Vicinanza et al. 2018; Gatti et al. 2020, 2022). Second and third moments of the DES Y3 weak lensing mass maps were used in Gatti et al. (2022) to infer cosmology; we use the same implementation of the moment estimator.

(ii) **Peaks**: the peak statistic counts the number of peaks of the smoothed map above a certain threshold. We follow the implementation of peak counts in Zürcher et al. (2022).

(iii) **Wavelet phase harmonics (WPH)**: WPH are second moments of smoothed weak lensing mass maps that have undergone a non-linear transformation. WPH are often associated with machine learning methods as they were designed to emulate information capture in the manner of a convolutional neural network, without the need for training data (Mallat 2016).
WPH statistics characterize the coherent structures in non-Gaussian random fields, by quantifying the phase alignment at different spatial scales (Mallat, Zhang & Rochette 2020; Zhang & Mallat 2021), and they can provide useful insights as a direct analogy with deep learning. We follow the implementation of WPH in Allys et al. (2020), which has already found success with astrophysical applications (Regaldo-Saint Blancard et al. 2021; Jeffrey et al. 2022).

These map-based statistics are applied to reconstructed weak lensing mass maps, using a full-sky generalization of the Kaiser & Squires (1993) algorithm that recovers a noisy estimate of the lensing convergence field κ from pixelized shear maps (see Jeffrey et al. 2021b). The statistics are applied to 'smoothed' versions of the maps. More details about the specific implementation of each statistic are provided in the supplementary material online.

For each statistic, we assess in Fig. 1 the impact of source clustering by comparing the measurements from the simulations

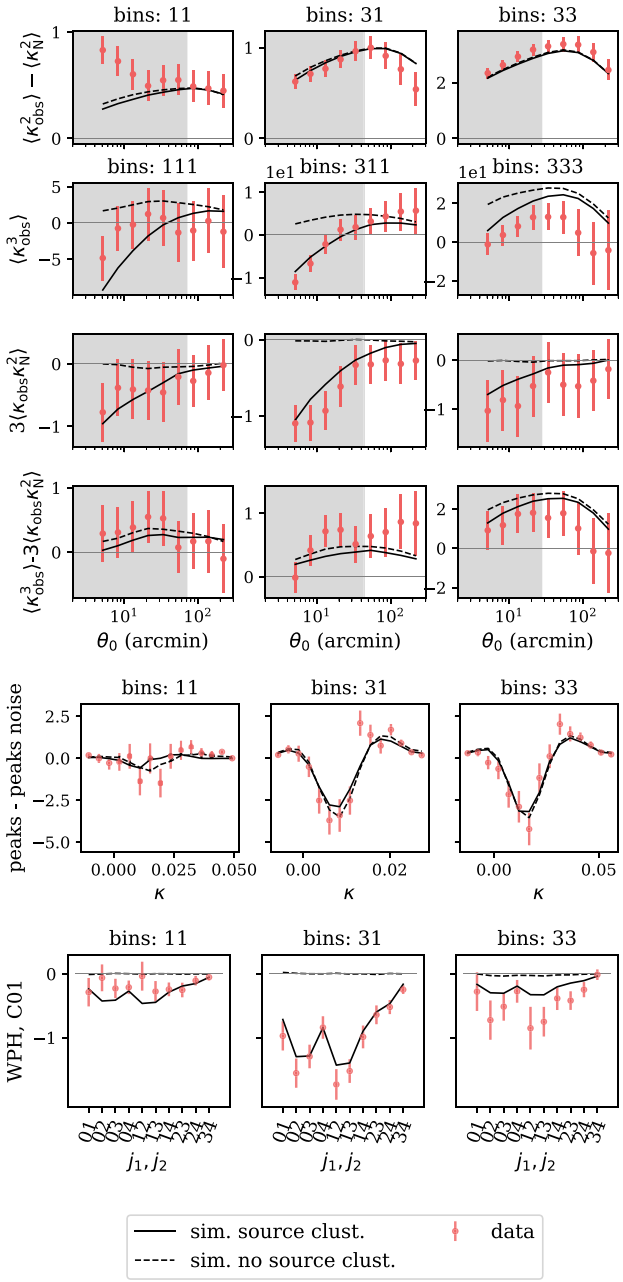

Figure 1. From top to bottom: second and third moments, peak functions, and WPH for different combinations of maps involving different bins (e.g. 'bins: 311' means that two maps for the first bin and one for the third bin have been used). The points represent the measurement in DES Y3 data and solid (dashed) lines represent the average in simulations with (without) source clustering. We multiplied the amplitude of each statistic by a constant to rescale the dynamical range on the y-axis for plotting purposes. Where present, the shaded regions represent the scales that have been not considered in the cosmological analyses using moments (Gatti et al. 2022). Third moments involving noise are labelled $3\langle\kappa\kappa^2_{\rm N}\rangle \equiv \langle\kappa\kappa_{\rm N}\kappa_{\rm N}\rangle^{i,j,k}$ + cycl., with 'cycl.' referring to the cyclic permutation of the indexes of the bins.

with and without source clustering (solid and dashed lines); these measurements are then compared to data (red points). When possible, we highlight the part of the measurements not included in the DES Y3 cosmological analyses (grey regions in Fig. 1).

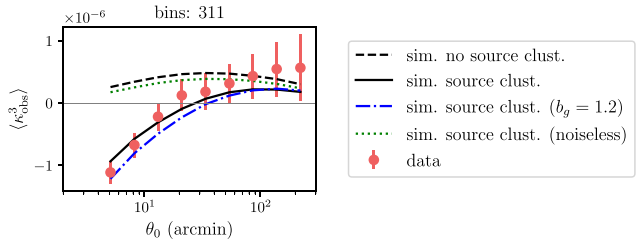

Figure 2. Third moments as measured in simulations using different mocks: with no source clustering, with source clustering, and with source clustering but assuming a stronger linear clustering for the source galaxies (i.e. assuming a galaxy-matter bias of $b_g = 1.2$).

4.1 Second and third map moments

Given current measurement uncertainties, the impact of source clustering on second moments is negligible (first row of Fig. 1), in line with the findings of Krause et al. (2021). It only slightly dampens the signal at small scales and in moments that include a low-redshift bin, for both 'auto' and 'cross' moments. For third moments, the impact is more dramatic (second row of Fig. 1), particularly for moments that include low-redshift bins. The data clearly follow better the simulations with source clustering, and the difference between the two sets of simulations is often significantly larger than measurement uncertainties.

Most of the effect induced by source clustering is due to a non-zero correlation between the convergence field and the noise. The effect of source clustering for a mock sample with no shape noise is significantly smaller (but does not vanish completely, see Fig. 2). The non-zero noise–signal correlation follows from the noise modulation introduced in equation (5), and it is a consequence of the map-making procedure. This can also be tested in data by looking at third moments that combine the noisy convergence maps and 'noise-only' maps created by randomly rotating the galaxy ellipticities of the shape catalogue. The rotation erases the shear signal but preserves the source clustering modulation of the noise. In simulations, we find that while moments of the form $\langle\kappa^2\kappa_{\rm N}\rangle$ or $\langle\kappa^3_{\rm N}\rangle$ are consistent with zero within uncertainties, $\langle\kappa\kappa^2_{\rm N}\rangle$ are not (in the presence of source clustering). This is shown in the third row of Fig. 1, where simulations with source clustering provide a good match to the data. That $\langle\kappa\kappa^2_{\rm N}\rangle$ is non-zero was already noted in Gatti et al. (2020, 2022), although the nature of the effect was not then understood. To compare the measurements to theory predictions, the authors of those papers subtracted $\langle\kappa\kappa^2_{\rm N}\rangle$ from the estimated third moments $\langle{\kappa_{\rm obs}}^3\rangle$. The result of this procedure is shown in the fourth row of Fig. 1; the impact of source clustering is greatly minimized, although the measurement errors are now larger. This procedure completely removes the contribution due to the non-zero correlation between the convergence field and the noise, and leaves the part of the effect associated with the modification of the average shear signal in the pixels, which is subdominant. Using the simulations produced in this work, we verified that the scale cut adopted in the Gatti et al. (2020, 2022) analysis, in combination with the subtraction of $\langle\kappa\kappa^2_{\rm N}\rangle$ terms, makes the analysis robust against source clustering effects (neglecting source clustering effects produces only a 0.08σ shift in the marginalized two-dimensional posterior of $\Omega_{\rm m}$ and S_8).

4.2 Peaks

The fifth row of Fig. 1 shows the impact of source clustering on the peak count function. We show the measurements only for the

smoothing scale $\theta_0 = 13.2$ arcmin, intermediate among the several smoothing scales included in the DES Y3 peak analysis in Zürcher et al. (2022); the trend with scales (not shown here) and redshift is similar to the moments' case, i.e. the difference between the two sets of simulations increases with smaller smoothing scales and when low-redshift bins are considered. Noise and signal are non-trivially mixed together due to the strong non-linearity of the peak function, and so, unlike the moments' case, we did not try to create a procedure to minimize the impact of source clustering, nor did we try to single out the effects due to the extra noise–signal correlations. We found that for peak statistic the effect is less striking than the moments' case. We verified that for the scales considered in the analysis by Zürcher et al. (2021), i.e. [7.9, 31.6] arcmin, the difference between two simulated data vectors with and without source clustering is small enough to not bias the cosmological inference (neglecting source clustering effects produces only a 0.18σ shift in the marginalized two-dimensional posterior of Ω_m and S_8).

4.3 WPH

The last row of Fig. 1 shows the WPH statistics obtained computing the second moments between a noisy convergence map and a noise-only map. The WPH statistics can involve a non-linear transformation of the input fields; in this case, before computing the second moments, we applied the modulus operation to the noise-only map (Allys et al. 2020). These statistics are consistent with zero in the absence of source clustering; however, we detect a clear signal in data due to noise–signal correlations, and this is well reproduced by the simulations with source clustering.

4.4 Significance

Using the moments and the WPH coefficients, we can construct two null tests for source clustering. The $C01$ coefficients of the WPH statistics of noisy convergence maps and noise-only maps are expected to be zero in the absence of source clustering (consistent with the simulation without source clustering). Using this null test for the bin combination (3, 1), we find a p-value for our observed χ^2 of $p = 3 \times 10^{-11}$, which corresponds to 6.5σ significance. This result assumes a mean-zero Gaussian likelihood with covariance matrix Σ estimated from simulations with no source clustering, where $\chi^2 = d^\mathrm{T}\Sigma^{-1}d$ with measured observable vector d. The same null test for the third moment $\langle\kappa\kappa_\mathrm{N}^2\rangle$ for the bin combination (3, 1, 1) yields $p = 4 \times 10^{-3}$ (2.6σ). No trivial null test can be constructed with the peak statistics.

Finally, we note that the magnitude of the source clustering effect also depends on the clustering properties of the source sample, which should be marginalized over when analysing map-based weak lensing higher order statistics (Fig. 2).

5 DISCUSSION AND CONCLUSION

We have demonstrated the impact of source galaxy clustering on map-based higher order summary statistics of weak gravitational lensing observables. Source clustering affects the mean shear field estimated from galaxy catalogues, as the noise-free lensing signal is modulated by a different effective redshift distribution; moreover, it induces a strong correlation between a pixel's shear signal and its noise properties. The latter effect is the dominant one in map-based higher order statistics. Using simulations with galaxies that either trace or do not trace the underlying density field, we show that the effect induced in the signals of common higher order statistics can exceed the current measurement uncertainties, depending on the choice of scale cut and of summary statistic redshift range. We find that third moments and WPH coefficients are the most affected ones, whereas peak counts are less affected. Source clustering effects are larger at small scales and for statistics applied to combinations of low- and high-redshift samples, and diminish at high redshift.

Further, we have shown a clear source clustering feature using DES Y3 data. Due to the induced correlation between the shear signal and the noise properties of the maps, third moments combining the noisy convergence maps and 'noise-only' maps no longer vanish. We detected a similar feature at high statistical significance for WPH. Mocks with source clustering were well able to reproduce these features; mocks without source clustering provided a poor fit to the data (p-values of 4e−3 for third moments and 3e−11 for WPH).

Cosmological analyses using map-based higher order statistics have two strategies for dealing with source clustering: either minimize its effect by introducing ad hoc scale cuts and/or de-noising procedures, or fully forward model it, incorporating it into simulations. This work presents a recipe for efficiently incorporating source clustering effects into simulations, and also shows how to minimize the impact of source clustering for third moments using a de-noising procedure. If left unaccounted for, or if not tested, this effect could impact cosmological inference made with statistics using weak gravitational lensing observables, especially map-based higher order statistics (including ones not considered here, e.g. scattering transforms, deep learning summary statistics, Minkowski functionals, etc.). In the case of the DES Y3 higher order statistics analyses – moments (Gatti et al. 2022) and peaks (Zürcher et al. 2023) – we verified that the scale cuts and de-noising procedures adopted were sufficient to render this effect negligible.

Other effects could cause noise–signal correlations in map-based estimators, e.g. any selection effect depending on the local value of the matter and shear fields modulating the source number density. Source magnification induces an extra modulation proportional to $1 + \kappa(p, s)$; however, our tests show this to be negligible (owing to a lower signal amplitude compared to the density field). Blending effects are also likely negligible, as they are expected to affect only a small fraction of the sample. In general, any deviation from the simple $1 + b_g\delta(p, s)$ modulation considered here would lead to a specific redshift evolution and/or amplitude signature in the measurements, and we do not see this. Other astrophysical effects such as intrinsic alignment and baryonic feedback can impact $\gamma(s, p)$ and $\delta(s, p)$, but they do not directly modulate the number of galaxies. They could, however, enhance the source clustering effects: intrinsic alignment, in particular, is a local effect modulated by the same density fluctuations that modulate the source clustering (Blazek et al. 2019), and hence it could boost the amplitude of the noise–signal correlations.

This work focused on map-based statistics. Source clustering is expected to affect catalogue-based statistics differently: there should be no noise–signal contributions (as these are due to averaging the shear in pixels before estimating the summary statistics), but sources would still be preferentially sampled in regions with high shear/convergence. The impact is thus expected to be smaller; we leave this investigation to future works.

DATA AVAILABILITY

All DES Y3 data used in this work are publicly available at https://des.ncsa.illinois.edu/releases/y3a2/. The code and mocks used in this work can be made available upon request.

REFERENCES

Allys E., Marchand T., Cardoso J.-F., Villaescusa-Navarro F., Ho S., Mallat S., 2020, Phys. Rev. D, 102, 103506
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, Phys. Rev. D, 100, 103506
Cardone V. F., Camera S., Mainini R., Romano A., Diaferio A., Maoli R., Scaramella R., 2013, MNRAS, 430, 2896
Chang C. et al., 2018, MNRAS, 475, 3165
Cheng S., Ting Y.-S., Ménard B., Bruna J., 2020, MNRAS, 499, 5902
Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofmann T., Schneider A., 2019, Phys. Rev. D, 100, 063514
Fosalba P., Gaztañaga E., Castander F. J., Crocce M., 2015, MNRAS, 447, 1319
Gatti M. et al., 2020, MNRAS, 498, 4060
Gatti M. et al., 2021, MNRAS, 504, 4312
Gatti M. et al., 2022, Phys. Rev. D, 106, 083509
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Harnois-Déraps J., Martinet N., Castro T., Dolag K., Giblin B., Heymans C., Hildebrandt H., Xia Q., 2021, MNRAS, 506, 1623
Huff E., Mandelbaum R., 2017, preprint (arXiv:1702.02600)
Jeffrey N., Alsing J., Lanusse F., 2021a, MNRAS, 501, 954
Jeffrey N. et al., 2021b, MNRAS, 505, 4626
Jeffrey N., Boulanger F., Wandelt B. D., Regaldo-Saint Blancard B., Allys E., Levrier F., 2022, MNRAS, 510, L1
Kacprzak T. et al., 2016, MNRAS, 463, 3653
Kaiser N., Squires G., 1993, ApJ, 404, 441
Krause E. et al., 2021, preprint (arXiv:2105.13548)
Liu J., Petri A., Haiman Z., Hui L., Kratochvil J. M., May M., 2015, Phys. Rev. D, 91, 063507
Lu T., Haiman Z., Li X., 2023, MNRAS, 521, 2050
MacCrann N. et al., 2022, MNRAS, 509, 3371
Mallat S., 2016, Phil. Trans. R. Soc. A, 374, 20150203
Mallat S., Zhang S., Rochette G., 2020, Inf. Inference: J. IMA, 9, 721
Martinet N. et al., 2018, MNRAS, 474, 712
Myles J. et al., 2021, MNRAS, 505, 4249
Peel A., Pettorino V., Giocoli C., Starck J.-L., Baldi M., 2018, A&A, 619, A38
Petri A., Liu J., Haiman Z., May M., Hui L., Kratochvil J. M., 2015, Phys. Rev. D, 91, 103511
Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophys. Cosmol., 4, 2
Pyne S., Joachimi B., 2021, MNRAS, 503, 2300
Regaldo-Saint Blancard B., Allys E., Boulanger F., Levrier F., Jeffrey N., 2021, A&A, 649, L18
Schmidt F., Rozo E., Dodelson S., Hui L., Sheldon E., 2009, ApJ, 702, 593
Schneider P., van Waerbeke L., Mellier Y., 2002, A&A, 389, 729
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Valageas P., 2014, A&A, 561, A53
Van Waerbeke L. et al., 2013, MNRAS, 433, 3373
Vicinanza M. et al., 2016, preprint (arXiv:1606.03892)
Vicinanza M., Cardone V. F., Maoli R., Scaramella R., Er X., 2018, Phys. Rev. D, 97, 023519
Zhang S., Mallat S., 2021, Appl. Comput. Harmon. Anal., 53, 199
Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, J. Cosmol. Astropart. Phys., 2021, 028
Zürcher D. et al., 2022, MNRAS, 511, 2075
Zürcher D., Fluri J., Ajani V., Fischbacher S., Refregier A., Kacprzak T., 2023, MNRAS, 525, 761

SUPPORTING INFORMATION

Supplementary data are available at *MNRASL* online.

SC_supplementary_.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

[1]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[2]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[3]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[4]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[5]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[6]*Department of Physics, Northeastern University, Boston, MA 02115, USA*
[7]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[8]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*
[9]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[10]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[11]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390, USA*
[12]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA*
[13]*Department of Physics, Duke University, Durham, NC 27708, USA*
[14]*NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA*
[15]*Kavli Institute for Particle Astrophysics & Cosmology, Stanford University, PO Box 2450, Stanford, CA 94305, USA*
[16]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[17]*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*
[18]*NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[19]*Department of Physics, Université Grenoble Alpes, CNRS, LPSC-IN2P3, F-38000 Grenoble, France*
[20]*Department of Physics and Astronomy, University of Waterloo, 200 University Avenue W, Waterloo, ON N2L 3G1, Canada*
[21]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[22]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[23]*University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstraße 1, D-81679 Munich, Germany*
[24]*Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St, Urbana, IL 61801, USA*
[25]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[26]*School of Physics and Astronomy, Cardiff University, Cardiff, CF24 3AA, UK*
[27]*Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland*
[28]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
[29]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[30]*Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Campinas, SP 13083-859, Brazil*
[31]*Department of Physics, University of Genova and INFN, Via Dodecaneso 33, I-16146 Genova, Italy*
[32]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[33]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), 28040, Madrid, Spain*
[34]*Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA*

[35]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[36]*Department of Physics, Duke University, Durham, NC 27708, USA*
[37]*Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, F-31400 Toulouse, France*
[38]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[39]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[40]*Excellence Cluster Origins, Boltzmannstraße 2, D-85748 Garching, Germany*
[41]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[42]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstraße 1, D-81679 München, Germany*
[43]*Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile*
[44]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[45]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[46]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[47]*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[48]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[49]*UMR 7095, Institut d'Astrophysique de Paris, CNRS, F-75014 Paris, France*
[50]*UMR 7095, Institut d'Astrophysique de Paris, Sorbonne Universités, UPMC Univ Paris 06, F-75014 Paris, France*
[51]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[52]*Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain*
[53]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[54]*Physics Department, William Jewell College, Liberty, MO 64068, USA*
[55]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[56]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[57]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[58]*Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, NO-0315 Oslo, Norway*
[59]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[60]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[61]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[62]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[63]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[64]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[65]*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*
[66]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[67]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[68]*Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029 Hamburg, Germany*
[69]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[70]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

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