Lensing without borders – I. A blind comparison of the amplitude of galaxy–galaxy lensing between independent imaging surveys

A. Leauthaud, A. Amon, S. Singh, D. Gruen, J.U. Lange, S. Huang, N.C. Robertson, T.N. Varga, Y. Luo, C. Heymans, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **510,** 6150–6189 (2022) https://doi.org/10.1093/mnras/stab3586
Advance Access publication 2021 December 14

Lensing without borders – I. A blind comparison of the amplitude of galaxy–galaxy lensing between independent imaging surveys

A. Leauthaud [1★] A. Amon,[2★] S. Singh,[3,4] D. Gruen [2,5] J. U. Lange [1,2] S. Huang [6]
N. C. Robertson,[7] T. N. Varga [5,8] Y. Luo [1] C. Heymans,[9,10] H. Hildebrandt,[10] C. Blake,[11]
M. Aguena [12,13] S. Allam,[14] F. Andrade-Oliveira,[13,15] J. Annis,[14] E. Bertin [16,17] S. Bhargava,[18]
J. Blazek,[19,20] S. L. Bridle,[21] D. Brooks,[22] D. L. Burke,[2,23] A. Carnero Rosell [13,24,25]
M. Carrasco Kind [26,27] J. Carretero [28] F. J. Castander,[29,30] R. Cawthon [31] A. Choi,[32]
M. Costanzi [33,34,35] L. N. da Costa,[13,36] M. E. S. Pereira,[37] C. Davis [2] J. De Vicente [38]
J. DeRose,[4,39,40] H. T. Diehl,[14] J. P. Dietrich,[41] P. Doel,[22] K. Eckert,[42] S. Everett,[40] A. E. Evrard [43,44]
I. Ferrero [45] B. Flaugher,[14] P. Fosalba [29,30] J. García-Bellido [46] M. Gatti,[42] E. Gaztanaga [29,30]
R. A. Gruendl,[26,27] J. Gschwend,[13,36] W. G. Hartley,[47] D. L. Hollowood [40] K. Honscheid,[32,48] B. Jain,[42]
D. J. James,[49] M. Jarvis,[42] B. Joachimi,[50] A. Kannawadi [51] A. G. Kim [39] E. Krause,[52] K. Kuehn,[53,54]
K. Kuijken,[55] N. Kuropatkin,[14] M. Lima,[13,56] N. MacCrann [57] M. A. G. Maia,[13,36] M. Makler [58,59]
M. March [42] J. L. Marshall,[60] P. Melchior [6] F. Menanteau,[26,27] R. Miquel,[28,61] H. Miyatake [62,63]
J. J. Mohr,[8,41] B. Moraes,[64] S. More,[65] M. Surhud,[63] R. Morgan,[31] J. Myles [2,23,66] R. L. C. Ogando [36]
A. Palmese [14,67] F. Paz-Chinchón,[26,68] A. A. Plazas Malagón [6] J. Prat,[69] M. M. Rau [70] J. Rhodes,[71]
M. Rodriguez-Monroy [38] A. Roodman,[2,23] A. J. Ross [32] S. Samuroff [70] C. Sánchez [42]
E. Sanchez [38] V. Scarpine,[14] D.J. Schlegel,[72] M. Schubnell,[44] S. Serrano,[29,30] I. Sevilla-Noarbe,[38]
C. Sifón [73] M. Smith [74] J. S. Speagle,[75,76,77] E. Suchyta [78] G. Tarle,[44] D. Thomas [79] J. Tinker [80]
C. To [2,23,66] M. A. Troxel [81] L. Van Waerbeke,[82] P. Vielzeuf[28] and A. H. Wright [10]

Affiliations are listed at the end of the paper

Accepted 2021 November 30. Received 2021 November 19; in original form 2021 May 13

ABSTRACT

Lensing without borders is a cross-survey collaboration created to assess the consistency of galaxy–galaxy lensing signals ($\Delta\Sigma$) across different data sets and to carry out end-to-end tests of systematic errors. We perform a blind comparison of the amplitude of $\Delta\Sigma$ using lens samples from BOSS and six independent lensing surveys. We find good agreement between empirically estimated and reported systematic errors which agree to better than 2.3σ in four lens bins and three radial ranges. For lenses with $z_{\rm L} > 0.43$ and considering statistical errors, we detect a $3–4\sigma$ correlation between lensing amplitude and survey depth. This correlation could arise from the increasing impact at higher redshift of unrecognized galaxy blends on shear calibration and imperfections in photometric redshift calibration. At $z_{\rm L} > 0.54$, amplitudes may additionally correlate with foreground stellar density. The amplitude of these trends is within survey-defined systematic error budgets that are designed to include known shear and redshift calibration uncertainty. Using a fully empirical and conservative method, we do not find evidence for large unknown systematics. Systematic errors greater than 15 per cent (25 per cent) ruled out in three lens bins at 68 per cent (95 per cent) confidence at $z < 0.54$. Differences with respect to predictions based on clustering are observed to be at the 20–30 per cent level. Our results therefore suggest that lensing systematics alone are unlikely to fully explain the 'lensing is low' effect at $z < 0.54$. This analysis demonstrates the power of cross-survey comparisons and provides a promising path for identifying and reducing systematics in future lensing analyses.

Key words: cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

The pursuit to constrain the equation of state of dark energy has motivated a number of imaging weak-lensing surveys. A number of the surveys are now complete such as the Sloan Digital Sky Survey

★ E-mail: alexie@ucsc.edu (AL); alexandra.amon@ast.cam.ac.uk (AA)

(SDSS; Gunn et al. 1998), the Canada France Hawaii Telescope (CFHT) Lensing Survey (CFHTLenS;[1] Heymans et al. 2012), the Deep Lens Survey (Jee et al. 2013), the Red-Sequence Cluster Lensing Survey (Hildebrandt et al. 2016), and the CFHT Survey of Stripe 82 (CS82; Leauthaud et al. 2017). Analysis of a number of weak-lensing surveys are ongoing including the Dark Energy Survey (DES;[2] The Dark Energy Survey Collaboration 2015), the Kilo Degree Survey (KiDS;[3] Kuijken et al. 2015), and the Hyper Suprime Cam survey (HSC;[4] Aihara et al. 2018b). Looking forward, a number of Stage 4 surveys will also be carried out within the next decade including the Legacy Survey of Space and Time (LSST;[5] LSST Science Collaboration 2009), the *Euclid*[6] mission (Laureijs et al. 2011), and the *Roman*[7] mission (Spergel et al. 2015).

As the statistical precision of these surveys has grown, intriguing differences with respect to the best-fitting model from the *Planck* experiment have begun to emerge. Assuming a standard six-parameter ΛCDM model, recent cosmic shear measurements (Hikage et al. 2019; Asgari et al. 2021; DES Collaboration 2021) appear to prefer slightly lower values for the S_8 cosmological parameter compared to the best-fitting *Planck* cosmology (Planck Collaboration 2020). Another such difference is the 'lensing is low' effect. This is the observation that the lensing amplitude around luminous red galaxies is lower than predicted by their clustering in a *Planck* cosmology (Cacciato et al. 2013; Leauthaud et al. 2017; Lange et al. 2019; Singh et al. 2020). Measurements of the E_G statistic (e.g. Blake et al. 2016; Amon et al. 2018b; Blake et al. 2020; Singh et al. 2020) and joint cosmic shear and Baryon Oscillation Spectroscopic survey (BOSS) clustering analyses (see for example Heymans et al. 2021 and references therein) draw similar conclusions.

Using data from the SDSS main survey, Cacciato et al. (2013) studied clustering and galaxy–galaxy lensing measurements (hereafter 'gg-lensing'). While their constraints on S_8 were consistent with Wilkinson Microwave Anisotropy Probe (WMAP; Dunkley et al. 2009) at the time of publication, their results correspond to a lower value of S_8 compared to *Planck*. Also using the SDSS main sample, Mandelbaum et al. (2013) obtained cosmological constraints using large-scale measurements of lensing and clustering. However, due to the limited volume of the main sample and the radial scale cuts employed, their constraints have relatively large errors and are consistent with both WMAP and *Planck*. Using the larger CMASS sample from the BOSS (Alam et al. 2017) and lensing data from a combination of CFHTLenS and CS82, Leauthaud et al. (2017) showed that the observed lensing signal around CMASS LRGs is lower than predicted from the clustering. Specifically, they found $\Delta\Sigma_{\rm predicted}/\Delta\Sigma_{\rm obs} \sim 1.2-1.3$ (20–30 per cent level differences in the lensing amplitude) where $\Delta\Sigma_{\rm predicted}$ is the signal predicted from a variety of galaxy halo models applied to the clustering. Lange et al. (2019) confirmed and extended these results to a wider range in redshift using CFHTLenS data. Lange et al. (2019) also showed the effect to be relatively independent of galaxy stellar mass (albeit with lower signal-to-noise due to smaller sample sizes when dividing by galaxy mass). Singh et al. (2020) extended the results of Mandelbaum et al. (2013) and studied the lensing and clustering of the BOSS LOWZ sample using lensing from SDSS, as well as *Planck* CMB lensing. Using only the large-scale signal, they constrain the S_8 parameter to be ∼15 per cent lower than predicted by *Planck* at the 2–4σ level. Their CMB lensing analysis prefers a 10 per cent (1σ) lower value of S_8 but with lower significance due to the larger errors from CMB lensing. Finally, Lange et al. (2021) showed the 'lensing is low' effect to be independent of both halo mass ($M_{\rm halo} > 10^{13.3}$ h^{-1} M$\odot$) and radial scale ($r < 60$ h^{-1} Mpc).

Taken together, these results could offer tantalizing hints of physics beyond ΛCDM. However, lensing measurements are notoriously difficult and understanding (and controlling for) systematic errors is one of the most challenging aspects for any lensing analysis. The weak-lensing community is acutely aware of the need to quantify and mitigate systematic errors and has been actively engaged in reducing systematic errors. There have been a number of community efforts to combat systematics, such as the Shear TEsting Programme (Heymans et al. 2006; Massey et al. 2007), the GRavitational lEnsing Accuracy Testing (GREAT) challenges (Bridle et al. 2009; Mandelbaum et al. 2014), and the PHoto-z Accuracy Testing program (Hildebrandt et al. 2010). As underscored by existing efforts on this front, two key challenges are: (1) the accurate measurement of the lensing shear from galaxy shapes (in the presence of noise, the point spread function, and galaxy blends), and (2) the determination of photometric redshifts (or redshift distributions), for source galaxies. While systematic errors from shape measurements and redshifts have in the past been sub-dominant compared to statistical errors, the increase in statistical precision afforded by larger survey areas means that even greater attention must be paid to systematic errors. Of particular concern is the possibility that the data may be affected by an unknown systematic that has yet to be quantified.

Systematic errors may be categorized into three types: the 'known knowns', the 'known unknowns', and the 'unknown unknowns'. The 'known knowns' are effects already accounted for in systematic error budgets. The 'known unknowns' are effects that are currently being studied and will be incorporated into future systematic error budgets. The 'unknown unknowns' are effects that have not been thought about and may not accounted for. If the differences with respect to *Planck* (after including the known systematic errors) continue to increase in significance, then the question of 'unknown unknowns' will become of considerable interest.

Lensing without borders (hereafter LWB) is an inter-survey collaboration, exploiting the overlap on the sky of existing lensing surveys with the BOSS spectroscopic survey to perform direct and empirically motivated tests for systematic effects in measurements of gg-lensing, following the methodology in Amon et al. (2018b). LWB has two goals: (1) to empirically search for systematic trends that could be used to reduce systematic floors, and (2) to empirically test if large 'unknown unknowns' systematic effects are present in the data.

The premise underlying LWB is that the gg-lensing signal around BOSS galaxies measures $\Delta\Sigma$, the excess differential surface mass density, a physical quantity. As such the measured $\Delta\Sigma$ values for BOSS galaxies should agree, independently of the lensing data employed (modulo sample variance and inhomogeneity in the lens sample). BOSS provides spectroscopic redshifts for lenses which enables a more accurate measurement of $\Delta\Sigma$. We perform a blind comparison of the amplitude of the $\Delta\Sigma$ signal using four lens samples from BOSS and using the sources catalogues and methodologies from six distinct lensing surveys (SDSS, CS82, CFHTLenS, DES, HSC, and KiDS). As shown in Luis Bernal & Peacock (2018), constraints on systematic errors improve when considering a large

[1] https://www.cfhtlens.org
[2] https://www.darkenergysurvey.org
[3] http://kids.strw.leidenuniv.nl
[4] https://hsc.mtk.nao.ac.jp/ssp
[5] https://www.lsst.org
[6] https://sci.esa.int/web/euclid
[7] https://roman.gsfc.nasa.gov

number of independent measurements, even if some measurements are more uncertain than others.

LWB provides an empirical end-to-end test of systematics in gg-lensing that is sensitive to both the shear calibration of the data, the redshift estimation, as well as the methodology for computing $\Delta\Sigma$. The framework developed here also provides a first handle on determining the origin of amplitude offsets (shear calibration, redshift estimation, and methodology), however, future work will focus more specifically on developing methodologies for disentangling such effects.

In the radial range of consideration in this paper ($r < 10$ Mpc), statistical constraints on the amplitude of the gg-lensing signal vary from $\sigma_{\rm amp} \sim 0.04$–$0.1$ depending on the survey at hand (these numbers will depend on which lensing data set is being used, the cuts made on the lens sample, and the radial range under consideration). Here, $\sigma_{\rm amp}$ is the statistical error on the ratio $\Delta\Sigma_{\rm predicted}/\Delta\Sigma_{\rm obs}$, where $\Delta\Sigma_{\rm predicted}$ is the predicted signal based on galaxy clustering (which should be the same for all surveys). With reported tensions between lensing and clustering in a *Planck* cosmology being at the 10–30 per cent level (Leauthaud et al. 2017; Lange et al. 2019; Singh et al. 2020; Lange et al. 2021), the tests proposed here will be able to check for large unknown systematics that could lead to such differences. However, our tests rely on the assumption that all of the lensing surveys are independent, have been analysed independently, and are not subject to confirmation bias.

The goal of this paper is to provide the first direct and empirically motivated test of the consistency of the galaxy–galaxy lensing amplitude across lensing surveys and to develop a framework for such comparisons. While the precision of the tests in this paper is limited by the existing overlap between various lensing surveys and BOSS, the LWB methodologies developed in this paper will become more powerful both as the overlap between lensing surveys increases, as well as the overlap between lensing surveys and spectroscopic surveys, such as Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016).

Our methodology is outlined in Section 3. Section 4 describes the foreground lens sample and Section 5 gives a brief description of the weak-lensing data used in this paper. The various methodologies used to compute $\Delta\Sigma$ are described in Section 6. Section 7 presents tests on the homogeneity of the BOSS samples. Our results are presented in Section 8 and discussed in Section 9. Section 10 presents a summary and our conclusions. We use a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $H_0 = 70$ km s^{-1} Mpc^{-1}. We assume physical coordinates to compute $\Delta\Sigma$.[8]

2 GENERAL METHODOLOGY FOR GALAXY–GALAXY LENSING

Here, we describe in general terms how to convert tangential shear into $\Delta\Sigma$. The full details, including team specific approaches, are presented in Section 6.

2.1 From γ to $\Delta\Sigma$

The shear signal induced by a given foreground mass distribution on a background source galaxy will depend on the transverse proper distance between the lens and the source and on the redshift configuration of the lens-source system. A lens with a projected surface mass density, $\Sigma(r)$, will create a shear that is proportional to the *surface mass density contrast*, $\Delta\Sigma(r)$:

$$\Delta\Sigma(r) \equiv \overline{\Sigma}(< r) - \overline{\Sigma}(r) = \Sigma_{\rm c} \times \gamma_{\rm t}(r). \quad (1)$$

Here, $\overline{\Sigma}(< r)$ is the mean surface density within proper radius r, $\overline{\Sigma}(r)$ is the azimuthally averaged surface density at radius r (e.g. Miralda-Escude 1991; Wilson et al. 2001), and $\gamma_{\rm t}$ is the tangentially projected shear. The geometry of the lens-source system intervenes through the critical surface mass density $\Sigma_{\rm c}$:

$$\Sigma_{\rm c} = \frac{c^2}{4\pi G} \frac{D_{\rm A}(z_{\rm s})}{D_{\rm A}(z_{\rm L}) D_{\rm A}(z_{\rm L}, z_{\rm s})}, \quad (2)$$

where $D_{\rm A}(z_{\rm L})$ and $D_{\rm A}(z_{\rm s})$ are angular diameter distances to the lens and source, and $D_{\rm A}(z_{\rm L}, z_{\rm s})$ is the angular diameter distance between the lens and source. When the redshifts (or redshift distribution) of source galaxies are known, each estimate of $\gamma_{\rm t}$ can be directly converted to an estimate of $\Delta\Sigma(r)$.

To measure $\Delta\Sigma(r)$ with high signal to noise, the lensing signal must be stacked over many foreground lenses and background sources. In order to optimize the signal to noise of this stacking process, an inverse variance weighting scheme is commonly employed when $\Delta\Sigma_{ij}$ is summed over many lens–source pairs. Each lens–source pair is attributed a weight w_{ij} that is often (but not always) the estimated variance of the shear measurement. The excess projected surface mass density is the weighted sum over all lens–source pairs:

$$\Delta\Sigma = \frac{\sum_{j=1}^{N_{\rm Lens}} \sum_{i=1}^{N_{\rm Source}} w_{ij} \times \gamma_{,ij} \times \Sigma_{{\rm c},ij}}{\sum_{j=1}^{N_{\rm Lens}} \sum_{i=1}^{N_{\rm Source}} w_{ij}}. \quad (3)$$

To remove systematic bias and obtain the optimal covariance (Mandelbaum et al. 2005; Singh et al. 2017), it has become common to subtract the measurement around random points (see Section 6.). In addition, sometimes additional weights may be applied to the lens sample (e.g. see Section 4.1).

2.2 Correction terms due to imperfect knowledge of source redshifts

To compute $\Delta\Sigma$ we must select background source galaxies. However, source galaxies typically only have photometric redshifts. These redshifts may be biased and the source selection may be imperfect. A number of correction terms are applied to $\Delta\Sigma$ estimates to account for such effects. These are:

(i) The boost factor. A 'boost correction factor' is sometimes applied in order to account for the dilution of the signal by physically associated sources (e.g. Kneib et al. 2003; Hirata et al. 2004; Sheldon et al. 2004; Mandelbaum et al. 2006). This correction factor is usually computed by comparing the weighted number density of source galaxies for the lens sample to the weighted number density of source galaxies around random points. However, the validity of boost correction factors is debated (e.g. Applegate et al. 2014; Melchior et al. 2015; Simet & Mandelbaum 2015; Leauthaud et al. 2017).

(ii) The dilution factor. The 'background' sample may contain a number of galaxies that are actually in the foreground ($z_{\rm s} < z_{\rm L}$). Because foreground galaxies are unlensed, the inclusion of these galaxies will cause $\Delta\Sigma$ to be underestimated.

(iii) The $f_{\rm bias}$ correction factor. $\Delta\Sigma$ estimates can be biased due to imperfect calibration of photo-z's. Furthermore, even with perfectly calibrated point source photo-z's, $\Delta\Sigma$ can be biased because of the non linear response of $\Delta\Sigma$ to source redshifts (via the $\Sigma_{\rm c}$ factor). Instead of using a point source estimate, some teams prefer to integrate over a source redshift probability distribution function

[8] See appendix C in Dvornik et al. (2018)

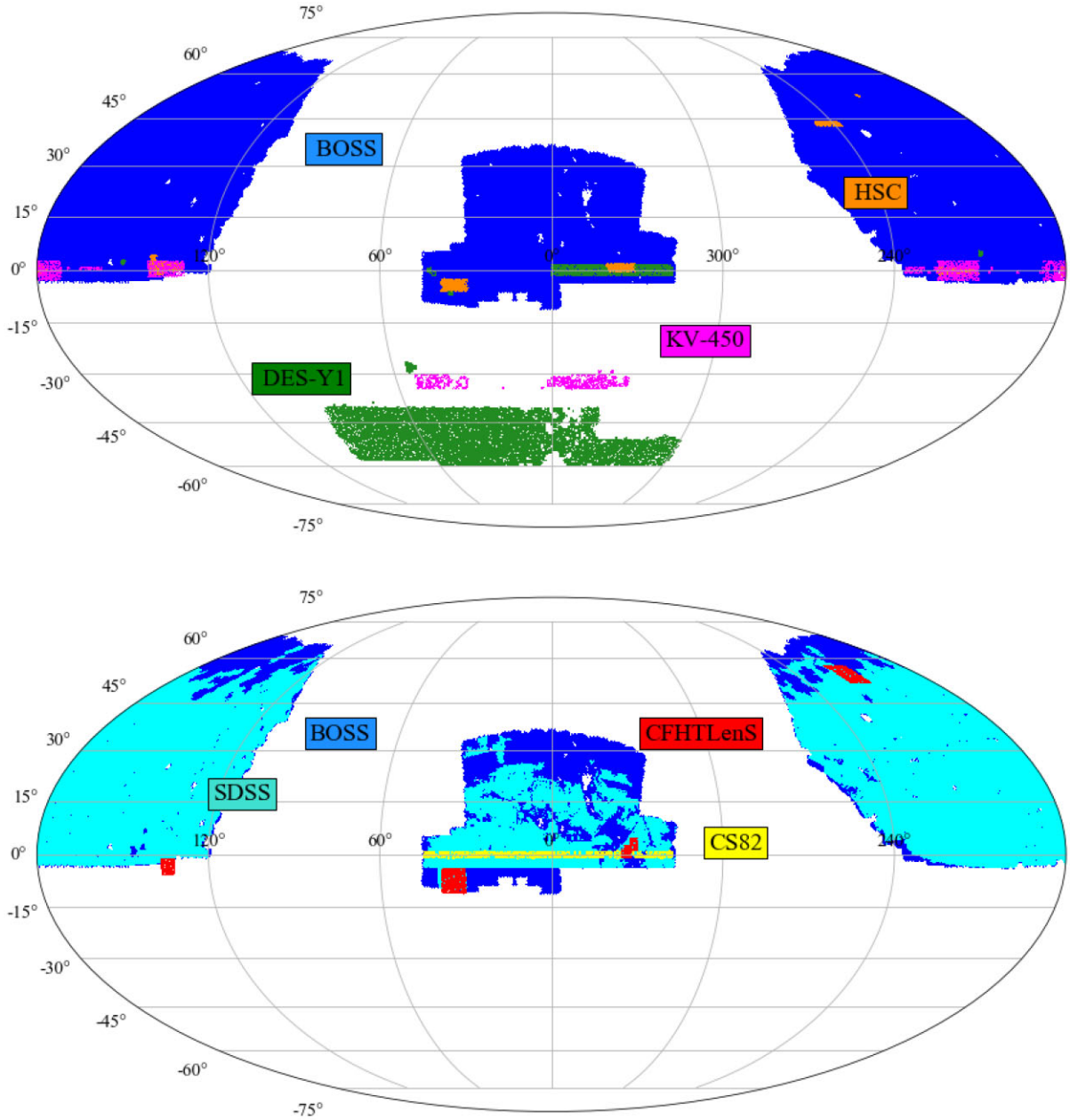

Figure 1. Footprints of a subset of the weak-lensing surveys and their overlap with the BOSS survey. BOSS is shown in dark blue, SDSS weak-lensing in cyan, HSC in orange, CFHTLenS in red, CS82 in yellow, DES-Y1 in green, and KV-450 (KiDS) in magenta.

(PDF). However, this integration will only be accurate if the full shape of the PDF is well calibrated. In other terms, *an unbiased mean P(z) does not guarantee an unbiased* $\Delta\Sigma$. For these reasons, a correction factor called $f_{\rm bias}$ is sometimes applied (see Section 6.1.3). This correction factor is computed using a representative sample of galaxies with spectroscopic redshifts. Often, $f_{\rm bias}$ is written in a way that also corrects for the dilution factor.

3 METHODOLOGY

3.1 General approach and limitations

In this paper, we use weak-lensing data from CS82, CFHTLenS, HSC, KiDS, SDSS, and DES. A description of these data are given in Section 5.

For the lens sample, we select a common set of lenses from BOSS (see Section 4). Fig. 1 displays the footprints of different surveys considered in this paper. Table 1 gives the overlap between BOSS and various lensing surveys.[9] Currently, apart from the SDSS lensing catalogue, the overlap between BOSS and existing lensing surveys is

[9]Binary masks with nside = 2048 were used.

Table 1. Overlap between current weak-lensing surveys and the spectroscopic BOSS, reported in deg^2. These area values are just rough estimates and should not be used for computations.

	SDSS lensing	HSC Y1	DES Y1	KV-450	CFHTLenS	CS82
BOSS	8359	166	160	204	118	144
SDSS lensing	-	160	134	196	108	130
HSC Y1	160	-	26	68	32	11
DES Y1	134	26	-	0	20	67
KV-450	196	68	0	-	3	0
CFHTLenS	108	32	20	3	-	7
CS82	130	11	67	0	7	-

typically of order 100–200 deg^2, however, this overlap will rapidly expand over the next few years to reach of order $\sim$1000 deg^2.

One of the main assumptions behind our methodology is that the lens sample selects a homogeneous sample of dark matter haloes across the BOSS footprint. However, there may be inhomogeneity in the BOSS lens sample. This is tested in Sections 7 and 8.3. There are two other caveats to our analysis. First, we do not account for cross-covariance between surveys (overlap areas are modest and are quoted in Table 1). Section 9.3.2 outlines a methodology for accounting for cross-covariance when the overlap between survey footprints increases. Ignoring this cross-covariance means that our systematic errors may be overestimated (and that our main conclusions are conservative). Secondly, our tests rely on the assumption that all of the lensing surveys are independent, and have been analysed independently. However, there may be systematic errors that are common between different lensing surveys (e.g. a common redshift calibration sample, such as COSMOS-30 and/or similar shear measurement methods) which cannot be tested here.

3.2 Computation of $\Delta\Sigma$

Prior to computing $\Delta\Sigma$, we agreed that all teams would compute the signal under the following set of assumptions:

(i) A fixed fiducial cosmology (as given in Section 1).
(ii) A fixed radial binning scheme. We use 10 logarithmically spaced bins from 0.05 to 15 Mpc.
(iii) Physical transverse distances are used for the computation of $\Delta\Sigma$.
(iv) Data points are compared at the mean r value of the bin (see justification below), where r is a physical transverse radius. This value is the same for all surveys.
(v) The lens and random files provided to each team correspond to the intersection between the BOSS footprint and the footprint of each shear catalogue.
(vi) Our fiducial test uses systematic weights that are applied to lenses to ensure that the spatial variations of the lenses follow those of the randoms (see Section 4.1).
(vii) We also perform an additional test for the CMASS sample in which we measure the lensing signal without systematic weights.

The effective value of r within bins depends on the scaling of the underlying signal ($\Delta\Sigma$) which is same for all the surveys. It also depends on the weighting imposed by the survey window (or the distribution of source galaxies), which can be different for different surveys. On small scales, the effects of survey masks are expected to be small, in which case the mean value of r within the bins, $\bar{r} = (r_{\rm high} - r_{\rm low})/2$, is close to the effective value for the measured $\Delta\Sigma$ (see equation D3 and fig. D2 in Singh et al. 2020).[10] In the mock tests performed by Singh et al. (2020) for SDSS, binning effects with $\bar{r}$ were < 1 per cent at $r < 60$ Mph h^{-1}, smaller than the 10–30 per cent differences of concern for this paper.

There are also a number of other choices required for a $\Delta\Sigma$ calculation. The following aspects were intentionally not discussed and were not homogenized among teams:

(i) How to write the estimator for $\Delta\Sigma$.
(ii) How to use the redshift information for each source.
(iii) How (and if) to compute and apply boost factors.
(iv) How (and if) to compute and apply dilution factors.
(v) How (and if) to apply any further correction factors for photo-z biases.
(vi) Computation of the covariance matrix.

Each team was responsible for the computation of $\Delta\Sigma$. Teams were asked to perform all tests deemed necessary before unblinding. Section 6 provides the specific details on how each team computed $\Delta\Sigma$.

3.3 Blinding strategy

We agreed that each team would compute $\Delta\Sigma$ independently. In the blinded phase, each team applied a multiplicative scale-dependent offset to their $\Delta\Sigma$ values. We opted for a scale-dependent offset so that no guesses could be made as to which scales were in better agreement. Each team randomly drew two numbers α and β with values between [0.80, 1.2] and then multiplied their $\Delta\Sigma$ values by a radially dependent factor $f(r)$:

$$f(r) = \frac{1}{9}\left[(\beta - \alpha)r + 10\alpha - \beta\right], \quad (4)$$

where r is expressed in physical Mpc. This blinding strategy results in a radial-dependent offset between signals at the 20 per cent level. All figures were made with this blinding strategy during the blinded phase.

There are already $\Delta\Sigma$ values published for CMASS and LOWZ. Hence, our tests could not be made 100 per cent blind. But to make tests as blind as possible, we imposed redshift cuts on the BOSS samples so that it was not possible to compare directly with other published values. These are described in Section 4.2.

[10]The impact of binning in the context of cosmic shear measured in angular bins has been discussed in other work (e.g. Krause et al. 2017; Troxel et al. 2018b; Asgari et al. 2019) but conclusions from these papers are not directly relevant to the case of gg-lensing which measures the signal in physical bins and over a much narrower redshift range.

3.4 Aspects of tests agreed to before analysis

This section describes aspects of the tests that were decided upon before the analysis was conducted.

(i) Small scales were distinguished from large scales when comparing signals. This is because smaller scales are subject to boost factor correction uncertainties, whereas large scales will be more affected by error estimates (correlated shape noise and sample variance). The scales $R_1 = [0.05,1]$ Mpc and $R_2 = [1,15]$ Mpc were analysed both separately and jointly. The motivation for these scales is based on the idea that boost correction factors should mainly only affect $\Delta\Sigma$ below 1 Mpc.

(ii) Data from each survey were fit with a model in which only the overall amplitude was allowed to vary (see next section). This is because the current errors on gg-lensing do not provide good constraints on slope variations.[11] Hence, we only tested for amplitude shifts. The radial ranges R_1 and R_2 were fit both separately and jointly. The resulting amplitudes are noted A_1 (for the R_1 range), A_2 (for the R_2 range), and A (for the full range).

(iii) Amplitudes were compared across different surveys.

(iv) A set of post-unblinding tests was also defined and is described further in Section 8. It was agreed to use 3σ as a threshold for determining trends to be significant.

(v) It was agreed to not comment on any survey being deemed either 'high' or 'low'. Doing so would amount to sigma-clipping and would introduce confirmation bias into the results by lowering the estimated value of $\sigma_{\rm sys}$.

(vi) Monte Carlo tests were used to show that given the number of bins, the errors, and the number of surveys used, there is ∼6 per cent probability of having one survey appear either 'high' or 'low' across all lens bins. It was therefore agreed to not comment on this aspect and we also strongly encourage readers not to do so.

3.5 Amplitude fitting

Our goal is to detect differences between the amplitudes of the $\Delta\Sigma$ signal, as measured by different surveys. One common, yet fairly insensitive way is a direct χ^2 test between the data points. Given knowledge about the shape of $\Delta\Sigma(r)$, and its covariance, a more stringent test can be done based on a *matched filter*.[12] Here, we opt to use the latter because we are primarily interested in comparing the amplitudes of measurements from different surveys.

For a data vector d, with covariance matrix C, a linear combination A can be written as $A = w^{\mathsf{T}}d = \sum_i w_i d_i$ with a weight vector w. The variance of this linear combination is $\sigma_{\rm A}^2 = w^{\mathsf{T}}\mathsf{C}w = \sum_{ij} w_i w_j C_{ij}$. When the true shape of the noiseless signal (i.e. the expectation value of d) is known as t, one can show that the linear combination of d with the highest possible signal-to-noise ratio is given by the matched filter amplitude A with weights $w \propto \mathsf{C}^{-1}t$.

In our case, d is the difference between the $\Delta\Sigma$ data vectors measured by two surveys. To define a matched filter, we need to know both the true shape and the covariance matrix of d. For the first ingredient, we expect that a potential non-zero d is primarily due to multiplicative errors, e.g. arising from shear or redshift calibration errors. That is, d has a radial shape close to that of $\Delta\Sigma$ itself. For the true profile assumed for our matched filter, we adopt $\Delta\Sigma$ as predicted by a Halo Occupation Distribution (HOD) analysis of the CMASS clustering signal from Leauthaud et al. (2017), hereafter noted $\Delta\Sigma_{\rm HOD}$. This model was obtained by fitting a standard HOD model to the two-point clustering of the CMASS sample and then by population a dark matter simulation with this HOD and predicting $\Delta\Sigma(r)$. The redshift range over which the clustering was measured (full CMASS sample) is different from the redshift ranges of the lens samples used in this paper. We should thus not expect the lensing amplitude here to match the prediction from clustering, but the general shape of $\Delta\Sigma_{\rm HOD}$ for BOSS samples does not vary strongly with redshift (e.g. Leauthaud et al. 2017), and so this model is good enough for our purpose.

The second ingredient for the matched filter is the covariance matrix of d. We assume that any pair of surveys 1 and 2, who have measured $\Delta\Sigma$ with covariances C_1 and C_2, are uncorrelated, such that the covariance of d is simply $(\mathsf{C}_1 + \mathsf{C}_2)$. We have verified empirically that the optimal filter defined this way for any pair of surveys is not too different from the filter assuming $\mathsf{C} \propto \sum_j \mathsf{C}_j$, where the sum runs over all surveys j. We will use the latter in order to be able to compare the amplitudes of all surveys on the same footing.

In summary, for each survey j, and one of three radial ranges (small, large, and all radii), we will determine a matched amplitude

$$A_j = \frac{w^{\mathsf{T}}\Delta\Sigma_j}{w^{\mathsf{T}}\Delta\Sigma_{\rm HOD}}, \tag{5}$$

and its uncertainty,

$$\sigma_j^2 = w^{\mathsf{T}}\mathsf{C}_j w, \tag{6}$$

where,

$$w = \left(\sum_j \mathsf{C}_j\right)^{-1} \Delta\Sigma_{\rm HOD}. \tag{7}$$

Note that because the operations are linear, the difference between two amplitudes is the same as the amplitude of the difference between the two corresponding data vectors (for which this matched filter was derived). In line with our focus on inter-comparing lensing surveys, our figures will report $A - \overline{A}$ only, which is not sensitive to an amplitude difference between the lensing signal and the clustering-based prediction. Here, $\overline{A}$ is the mean amplitude averaged over the lensing surveys.

The validity of our tests do not rely on the model having the correct shape – it remains a test on a linear combination of the data that should be zero in the absence of biases. The sensitivity of the test, however, does depend on $\Delta\Sigma_{\rm HOD}$. Had we used the matched filter amplitude for each individual survey, i.e. with $\boldsymbol{w}_j = \mathsf{C}_j^{-1}\Delta\Sigma_{\rm HOD}$, then this would not be the case: each survey would weight the signal differently as a function of radius, and an offset between $\Delta\Sigma_{\rm HOD}$ and the correct model could manifest as a non-zero difference in amplitudes for mutually consistently calibrated surveys with differently structured covariance matrices.

3.6 Searching for trends caused by correlated systematic errors

One of the key goals of this paper is to investigate if correlations exist between measured-lensing amplitudes and survey properties (e.g. $n_{\rm star}$, survey depth) that should, in principle, have no impact on $\Delta\Sigma$ (Sections 8.3 and 8.4). If found, such correlations could provide important clues as to the origin of systematic errors. These could be both known or unknown systematic errors. We seek to pin-point trends caused systematic errors. For this, we both use the reported

[11] A slope variation would result when a measured $\Delta\Sigma$ does not have the same shape as $\Delta\Sigma_{\rm HOD}$.

[12] Fitting the amplitude of a model to the data can be thought of as an optimal linear combination the data vector, yielding one number of interest. We then perform tests, such as χ^2, using this one number.

statistical errors, as well as the sum in quadrature of statistical and systematic errors to conduct these tests.

The left-hand side of Fig. 2 illustrates an example in which a systematic error correlates with a given parameter X (e.g. redshift, $n_{\rm star}$, etc.) and causes a trend in the lensing amplitude versus X (in this example, the measured trend is detected with a positive slope $\beta > 0$). The errors in the left hand figure are the statistical errors on the measurements. The green line indicates the true level of systematic error in these data (the rms deviation between the horizontal line and the blue data points).

The right-hand side of Fig. 2 now considers the addition of the estimated systematic errors. Systematic errors are educated guesses and may underestimate or overestimate the true value. For example, current lensing surveys rarely report estimates of the *error on the systematic error*. If the estimated systematic error underestimates the true value, then the trend with $\beta > 0$ may still be detected. If the estimated systematic error is equal or larger than the true value, then the trend may no longer be detected. Whether or not a trend would be detected will depend on how close the estimated systematic error is to the true value and how many data points are available. Thus, using the sum in quadrature of the statistical and the estimated systematic errors may not provide any insight into sources of systematic error.

In the case of a single dominant systematic error that correlates with parameter X, the statistical errors will increase the probability of detecting the trend, as illustrated on the left-hand side of Fig. 2. However, the picture will be more complicated if multiple kinds of systematic error with distinct physical origins are present in the data. In this case, the correct errors to use would be the sum in quadrature of the statistical error and the true values of those systematic errors that do not correlate with the parameter under investigation (e.g. parameter X in Fig. 2). However, systematic errors are not known at this level of detail (and the true values are not usually known). Because we are working in the regime of systematic uncertainties, where the true errors are not exactly known, there is no perfect way of carrying out these tests. The use of statistical errors will enhance the probability of the detecting trends if they are present in the data, but the significance of these trends could be overestimated, especially if multiple different kinds of systematic error are present in the data. In this paper, we will carry out tests both using statistical errors, as well as the sum in quadrature of statistical and systematic errors, keeping in mind the advantages and disadvantages of both choices.

3.7 Estimate of global systematic error

A second key goal in the paper is to use the measured spread between the amplitudes of $\Delta\Sigma$ as an empirical and end-to-end estimate of systematic errors. This global estimate will be noted $\sigma_{\rm sys}$ and is computed as follows. We first compute the reduced χ^2 between amplitudes (measured following the methodology described in Section 3.5). When the reduced χ^2 of the data, χ^2_ν, is greater than 1, we report the value of $\sigma_{\rm sys}$ that yields $\chi^2_\nu = 1$. We assume that each amplitude data point is drawn from a normal distribution with $\sigma^2 = \sigma^2_{\rm stat} + \sigma^2_{\rm sys}$ where $\sigma_{\rm stat}$ is the error on the amplitude for each survey. We also derive 68 per cent and 95 per cent confidence intervals on $\sigma_{\rm sys}$. For this, we consider the expected probability distribution for χ^2 with degrees of freedom $\nu = n - 1$ where n is the sample size ($n = 6$ for LOWZ and $n = 5$ for CMASS). We find the range of $\sigma_{\rm sys}$ values that produces a χ^2 that is within the central 68 per cent and 95 per cent of the distribution. Monte Carlo tests were used to validate this methodology.

Because we use the spread between the data points as a means to estimate the overall systematic error, the number we quote should be thought of as an ensemble estimate over all of the surveys under consideration. Monte Carlo tests were used to show that the $\sigma_{\rm sys}$ value that we estimate is roughly equal to the mean systematic error among surveys.

Our empirical estimate is a multiplicative bias on the amplitude of $\Delta\Sigma$. More specifically, if we consider a variable $S_{\rm sys}$ that is drawn from a Gaussian of width $\sigma_{\rm sys}$ and unknown mean, the relation between the true value $\Delta\Sigma_T$ and the measured value $\Delta\Sigma_M$ is:

$$\Delta\Sigma_M(r) = (1 + S_{\rm sys}) \times \Delta\Sigma_T(r), \tag{8}$$

where $S_{\rm sys}$ is independent of r and where $S_{\rm sys}$ can take on a different value for each survey. The mean of $S_{\rm sys}$ is unknown because we cannot use the methods here to determine the absolute value of $\Delta\Sigma$. For example, we would not be able to detect a systematic bias if this bias were common to all of the lensing surveys and had a similar impact on $\Delta\Sigma$.

3.8 Effective redshift weighting of lens samples

Different surveys apply a different effective weight to the lens sample (e.g. Nakajima et al. 2012; Mandelbaum et al. 2013; Simet et al. 2016; Leauthaud et al. 2017). However, amplitude variations in $\Delta\Sigma$ across the CMASS redshift range have been found to be small (Leauthaud et al. 2017; Blake et al. 2020). We also use narrow redshift bins for our lens samples in order to mitigate this effect. This topic is discussed further in Section 8.2.

4 FOREGROUND LENS DATA

4.1 BOSS survey

BOSS is a spectroscopic survey of 1.5 million galaxies over 10 000 deg^2 that was conducted as part of the SDSS-III program (Eisenstein et al. 2011) on the 2.5 m aperture Sloan Foundation Telescope at Apache Point Observatory (Gunn et al. 1998, 2006). A general overview of the BOSS survey can be found in Dawson et al. (2013), the BOSS spectrographs are described in Smee et al. (2013), and the BOSS pipeline is described in Bolton et al. (2012). BOSS galaxies were selected from Data Release 8 (DR8, Aihara et al. 2011) *ugriz* imaging (Fukugita et al. 1996) using a series of colour-magnitude cuts.

The BOSS selection uses the following set of colours:

$$c_\parallel = 0.7(g_{\rm mod} - r_{\rm mod}) + 1.2[(r_{\rm mod} - i_{\rm mod}) - 0.18] \tag{9}$$

$$c_\perp = (r_{\rm mod} - i_{\rm mod}) - (g_{\rm mod} - r_{\rm mod})/4 - 0.18 \tag{10}$$

$$d_\perp = (r_{\rm mod} - i_{\rm mod}) - (g_{\rm mod} - r_{\rm mod})/8.0 \tag{11}$$

The subscript 'mod' denotes model magnitudes, which are derived by adopting the better fitting luminosity profile between a de Vaucouleurs and an exponential luminosity profile in the r-band (Stoughton et al. 2002). All magnitudes are corrected for Galactic extinction using the dust maps of Schlegel, Finkbeiner & Davis (1998).

BOSS targeted two primary galaxy samples: the LOWZ sample at $0.15 < z < 0.43$ and the CMASS sample at $0.43 < z < 0.7$. The LOWZ sample is an extension of the SDSS I/II Luminous Red Galaxy (LRG) sample (Eisenstein et al. 2001) to fainter magnitudes and is defined according to the following selection criteria:

$$|c_\perp| < 0.2 \tag{12}$$

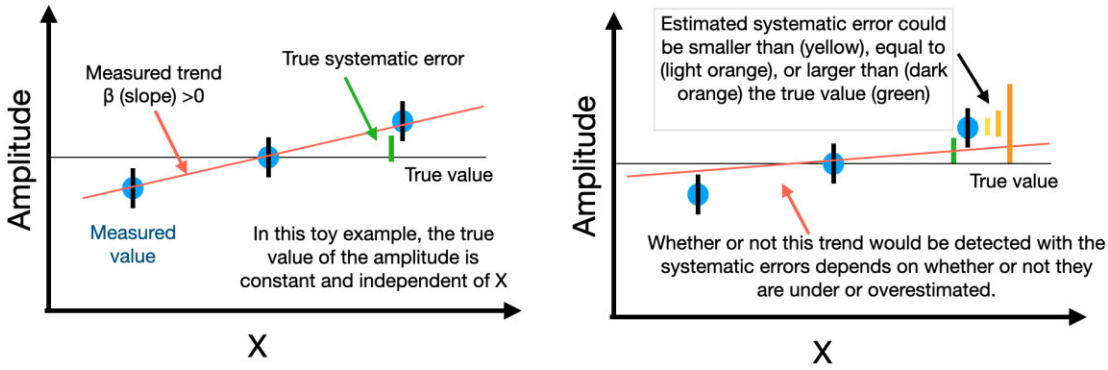

Figure 2. Illustrative example of the detection of a trend originating from systematic effects. Left: amplitude of the lensing signal versus parameter X. In this toy example, the true value of the amplitude is constant with X. However, a systematic error that correlates with X causes a trend (here with slope $\beta > 0$) in the relationship between amplitude and slope. The true systematic error (green horizontal line) is the rms spread between the measured data points (blue) and the horizontal line. Right: the estimated systematic error ($\sigma_{\rm sys,est}$) could be smaller, equal to, or larger than the true value ($\sigma_{\rm sys,true}$). Indeed, the true level of systematic error is rarely known. If $\sigma_{\rm sys,est} < \sigma_{\rm sys,true}$, the trend may be detected. If $\sigma_{\rm sys,est} \geq \sigma_{\rm sys,true}$, the trend may no longer be detected. Trends may or may not be detected if the estimated systematic error is summed in quadrature with the statistical error. The use of the statistical error bar (left) will yield a detection of trends if they are present in the data, but the significance of the trend could be overestimated.

$$r_{\rm cmod} < 13.6 + c_{\parallel}/0.3 \quad (13)$$

$$16 < r_{\rm cmod} < 19.6 \quad (14)$$

$$r_{\rm psf} - r_{\rm cmod} > 0.3 \quad (15)$$

Here, PSF magnitudes are denoted with the subscript 'psf'. The subscript 'cmod' denotes composite model magnitudes, which are calculated from the best-fitting linear combination of a de Vaucouleurs and an exponential luminosity profile (Abazajian et al. 2004). Equation (12) sets the colour boundaries of the sample; equation (13) is a sliding magnitude cut which selects the brightest galaxies at each redshift; equation (14) corresponds to the bright and faint limits; and equation (15) is to separate galaxies from stars. In a similar fashion to the SDSS I/II LRG sample, the LOWZ selection primarily selects red galaxies (Reid et al. 2016).

The CMASS sample targets galaxies at higher redshifts with a surface density of roughly 120 deg^{-2}. CMASS targets are selected from SDSS DR8 imaging according to the following cuts:

$$|d_{\perp}| > 0.55 \quad (16)$$

$$i_{\rm cmod} < 19.86 + 1.6(d_{\perp} - 0.8) \quad (17)$$

$$17.5 < i_{\rm cmod} < 19.9 \quad (18)$$

$$r_{\rm mod} - i_{\rm mod} < 2 \quad (19)$$

$$i_{\rm fib2} < 21.5 \quad (20)$$

where $i_{\rm fib2}$ is the estimated i-band magnitude in a 2 arcsec aperture diameter assuming 2 arcsec seeing. Star-galaxy separation on CMASS targets is performed via:

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20.0 - i_{\rm mod}) \quad (21)$$

$$z_{\rm psf} - z_{\rm mod} > 9.125 - 0.46 z_{\rm mod}. \quad (22)$$

In this paper, we use catalogues from Data Release 12 (DR12; Alam et al. 2015). We use the large-scale structure catalogues described in Reid et al. (2016) that were generated via the MKSAMPLE code and that can be found at https://data.sdss.org/sas/dr12/boss/lss/.[13]

These large-scale structure catalogues include information about the BOSS selection function, survey masks, and imaging quality masks, as well as weights to correct for various selection effects. In this paper, we will be concerned with understanding if inhomogeneities in the BOSS samples may lead to variations in the mean halo mass of the sample across different regions. We will return to this topic in Section 7.

Veto masks are applied to the LSS catalogues (Reid et al. 2016). These masks reject regions where BOSS galaxies cannot be observed. Among other things, these masks impose a cut that rejects areas of the survey that are too close to bright stars (the bright star mask), that have non photometric imaging conditions, where the seeing is poor, and with high extinction.

In the early phase of the survey, an incorrect star-galaxy separation scheme was used for LOWZ. We do not use any LOWZ galaxies in regions where this happened (chunks 2–6 corresponding to the LOWZE2 and LOWZE3 samples). As a result, the areas covered by CMASS and LOWZ are different. See appendix A in Reid et al. (2016).

In DR12, a 'combined' sample was also created. We do not use the combined sample here. The reason for this is because the CMASS sample is more subject to observational effects (seeing, stellar density). We wish to study the impact of these effects on the lensing signal in isolation from the LOWZ sample. Also, we do not wish to use the LOWZE2 and E3 samples which are in the combined sample.

The BOSS LSS catalogues include various weights designed to minimize the impact of artificial observational effects that can impact estimates of the true galaxy over-density field. A full description of these weights is given in Reid et al. (2016). We briefly summarize the weights here:

[13]Exact file names are galaxy_DR12v5_CMASS_North.fits.gz and so on and so forth.

(i) $w_{\rm cp}$: accounts for galaxies that did not obtain redshifts due to fibre collisions by up-weighting the nearest galaxy from the same target class.

(ii) $w_{\rm noz}$: weighting scheme designed to deal with galaxies for which the spectroscopic pipeline failed to obtain a redshift.

Taking these two weights together, the overall redshift weight is $w_{\rm z} = w_{\rm cp} + w_{\rm noz} - 1$. In addition, there is also a set of weights that are designed to correct for variations in the CMASS samples with stellar density and seeing. Because the LOWZ sample is brighter than CMASS, it does not require these extra weights. The angular systematic weights for CMASS are:

(i) $w_{\rm star}$: a weight to account for variations in the CMASS number density with stellar density. $w_{\rm star}(n_{\rm s}, i_{\rm fib2}) = (A_{i\rm fib2} + B_{i\rm fib2} n_{\rm s})^{-1}$. Variations in the number density with stellar density were found to correlate with galaxy surface brightness, in particular, the $i_{\rm fib2}$ magnitude. As the stellar density increases, on average, galaxies with lower magnitudes in a 2 arcsec fibre are lost from the sample.

(ii) $w_{\rm see}$: the seeing based weight. There is a correlation between the number density and local seeing, due to star galaxy separation. For CMASS, the effect is such that in poor seeing conditions, the number density decreases because compact galaxies are classified as stars and are removed from the sample.

The total angular systematic weight for each galaxy is $w_{\rm systot} = w_{\rm star} w_{\rm see}$. Finally, the total weight for CMASS is constructed as $w_{\rm tot} = w_{\rm systot}(w_{\rm cp} + w_{\rm noz} - 1)$.[14]

The BOSS systematic weights were designed to up-weight galaxies to create a sample with constant number density. We apply the systematic weights to our lens samples so that the spatial distribution of the randoms follow that of the lens sample. However, applying the BOSS weights will not guarantee a sample with fixed halo mass across the survey – indeed selection effects could lead to spatial inhomogeneity in the mean halo mass across the survey. In Section 7, we explore the impact of the inhomogeneity of the BOSS samples on $\Delta\Sigma$. We also design a set of post-unblinding tests that can be found in Section 8.3.

4.2 Lens samples

We use four distinct lens samples. Two are based on LOWZ and two are based on CMASS. Specifically, the samples we use are:

(i) L1: LOWZ sample with $0.15 < z < 0.31$
(ii) L2: LOWZ sample with $0.31 < z < 0.43$
(iii) C1: CMASS sample with $0.43 < z < 0.54$
(iv) C2: CMASS sample with $0.54 < z < 0.7$

These redshift cuts are designed to ensure that the signals cannot be compared with any other published values. Fine redshift bins were also desirable in order to minimize differences in the mean effective redshift across surveys (see Section 8.2).

We apply $w_{\rm tot}$ to the lens samples to ensure that the distribution of the randoms follows the variations in the lens samples. We also further test how our results vary if $w_{\rm tot}$ is not applied.

In BOSS, redshift dependent effects are taken into account with the systematic weights. For example, the $w_{\rm star}$ weight includes a magnitude dependence via $i_{\rm fib2}$ which accounts for redshift dependent variations in the number density. Previous analyses of BOSS data have binned by redshift, most notable is the final DR12 cosmological analysis which had arbitrary redshift binning across the combined LOWZ and CMASS samples (Alam et al. 2017).

Each lensing team has provided a HEALPIX mask (Górski et al. 2005) corresponding to the footprint of their shear catalogue. The BOSS lens and random catalogues are masked by each of the survey HEALPIX masks before computing $\Delta\Sigma$.

5 WEAK-LENSING DATA

This section provides brief descriptions on the various lensing data sets used in this paper. Readers are referred to the original survey and shear catalogue papers for the full details. The footprints of each of the lensing surveys involved in this collaboration are shown in Fig. 1 together with the footprint of BOSS. These lensing surveys differ in terms of their location on the sky, coverage area, data quality, depth, and number of source galaxies. Beyond that, their analyses differ in shear and redshift calibration techniques. These differences are summarized in Table 2.

5.1 SDSS

The SDSS survey (York et al. 2000) imaged ∼9000 deg^2 of the sky. We use the shape catalogue provided by Reyes et al. (2012) which is based on the re-gaussianization technique developed by Hirata & Seljak (2003). Briefly, the algorithm uses adaptive moments to measure the PSF-convolved galaxy shapes and then corrects for the PSF using the adaptive moments of the measured PSF, while also accounting for the non-gaussianity of both PSF and the galaxy light profiles. The shear calibration factor ($1 + m \sim 1.04 \pm 0.02$) is derived using simulations performed by Mandelbaum et al. (2012) and Mandelbaum et al. (2018b).

Photometric redshift estimates for source galaxies were obtained by Nakajima et al. (2012), using the template fitting method ZEBRA (Feldmann et al. 2006) on *ugriz* SDSS DR8 photometry. Following Nakajima et al. (2012), a representative spectroscopic sample is used to estimate and correct for the bias in $\Delta\Sigma$ caused by imperfect photometric redshifts.

5.2 HSC

The Wide layer of the Hyper Suprime-Cam Subaru Strategic Program aims to cover 1400 deg^2 of the sky in *grizy* using the Hyper Suprime-Cam (Komiyama et al. 2018; Miyazaki et al. 2018) Subaru 8.2 m telescope. The survey design is described in Aihara et al. (2018a), the HSC analysis pipeline is described in Bosch et al. (2018), and validation tests of the pipeline photometry are described in Huang et al. (2018a).

In this paper, we use the shear catalogue associated with the first data release (DR1; Aihara et al. 2018b). This catalogue covers an area of 136.9 deg^2 split into six fields (see Fig. 1) and has a mean *i*-band seeing of 0.58 arcsec and a 5σ point-source depth of $i \sim 26$. We refer the reader to Mandelbaum et al. (2018a) for details regarding the first-year shear catalogue. Only a brief description is given here. For HSC Y1, galaxy shapes are estimated on the co-added *i*-band images using a moments-based shape measurement method and the re-Gaussianization PSF correction method (Hirata & Seljak 2003). The shear calibration is described in Mandelbaum et al. (2018c). The HSC Y1 shear catalogue uses a conservative source galaxy selection including a magnitude cut of $i < 24.5$. The unweighted and weighted source number densities are 24.6 and 21.8 arcmin^{-2}, respectively.

[14]There are also the so-called 'FKP' weights ($w_{\rm FKP}$) based on Feldman, Kaiser, and Peacock 1994. These are weights that are designed maximize the signal to noise of 3D clustering statistics, not to correct for systematic effects, and are not relevant for the present study.

Table 2. Overview of lensing surveys used in this paper and methodologies used to compute $\Delta\Sigma$. First section: general properties of weak-lensing surveys. We quote the survey area in deg^2 (after masking out bright stars and other artefacts), the characteristic seeing [full-width at half-maximum (FWHM)], the photometric bands available for photometric redshift estimation, the median unweighted redshifts of the source distribution, and the effective weighted galaxy number density (see equation 1 in Heymans et al. 2012) after photo-z quality cuts measured in galaxies per square arc-minutes. Second section: method used to compute photometric redshifts, calibration samples used to ensure unbiased redshifts (z-reference sample), and choices regarding whether the mean redshift or the full $p(z)$ distribution was calibrated to be unbiased (z-calibration type). Third section: choices for the selection of background galaxies (also see Section 6). All surveys use a galaxy-by-galaxy $z_{\rm phot}$ point estimate to select background galaxies. Fourth section: choices regarding the computation of $\Delta\Sigma$. This includes the redshift adopted for the computation of $\Sigma_{\rm c}$ (equation 2). Here, choices differ with respect to the use of a point source estimate or the $p(z)$. As detailed in Section 2.2, integrating over the $p(z)$ does not guarantee an unbiased estimate of $\Delta\Sigma$ unless z-calibration type is also of type $p(z)$. Finally, we also specify choices regarding the boost factor correction, $f_{\rm bias}$, and the dilution factor. For $f_{\rm bias}$, KiDS is marked with a star symbol to indicate that the method employed should be equivalent to an $f_{\rm bias}$ and dilution correction, but the methodology used is different (see Section 6.6).

	SDSS	HSC-Y1	CS82	CFHTLenS	KiDS-VIKING-450	DES-Y1
Area (deg^2)	9243	137	129.2	126	341	1321
FWHM (arcsec)	1.2	0.58	0.6	0.6	0.66	0.96
filters	*ugriz*	*grizY*	*ugriz*	*ugriz*	*ugriZYJHK$_s$*	*griz*
$z_{\rm med}$	0.39	0.80	0.57	0.7	0.67	0.59
$n_{\rm eff}$	1.18	21.8	4.5	15.1	6.93	6.3
z-name	ZEBRA	FRANKENZ	BPZ	BPZ	BPZ+DIR	BPZ
z-method	SED	Machine learning	SED	SED	kNN	SED
z-reference sample	SPECZ	SPECZ + COSMOS30	SPECZ	none	SPECZ	COSMOS30
z-calibration type	none	full $p(z)$ shape	mean $z_{\rm phot}$	none	full $n_{\rm s}(z)$ shape	mean of $p(z)$
z-usage in source selection	Point estimate	Point estimate	Point estimate	Point estimate	Point estimate	Point estimate
Source selection cut 1	$z_{\rm s} > z_{\rm L}$	$z_{\rm s} > z_{\rm L} + 0.1$	$z_{\rm s} > z_{\rm L} + 0.1$	$z_{\rm s} > z_{\rm L} + 0.1$	$z_{\rm s} > z_{\rm L} + 0.1$	$z_{\rm s} > z_{\rm L} + 0.1$
Source selection cut 2	none	$z_{\rm s} > z_{\rm L} + \sigma_{68}$	$z_{\rm s} > z_{\rm L} + \sigma_{95}/2.0$	none	$0.1 < z_{\rm s} \leq 1.2$	none
$\Sigma_{\rm c}$ computation	Point estimate	Point estimate	Point estimate	$p(z_{\rm s})$	$n(z_{\rm s})$	$p(z_{\rm s})$
Boost factor correction	yes	no	no	no	no	yes
dilution correction	yes	yes	yes	no	yes*	yes
$f_{\rm bias}$	yes	yes	yes	no	yes*	yes

A variety of photometric redshifts have been computed for the HSC Y1 catalogue (Tanaka et al. 2018). Here, we use the FRANKENZ photo-z's described in Speagle et al. (2019), which uses a hybrid method that combines Bayesian inference with machine learning. In brief, FRANKENZ derives photo-z's for each object by computing a posterior-weighted average of the redshift distributions of its nearest photometric neighbours in the training set, taking into account observational uncertainties. The S16A HSC photo-z's were trained on a catalogue of ∼300k sources including a combination of spectroscopic, grism, prism, and many-band photometric redshifts covering a wide redshift, colour, and magnitude range. Using the `best` photo-z value from Speagle et al. (2019), the source distribution in this paper has a mean redshift of $z_{\rm s} = 0.95$ and a median of $z_{\rm s} = 0.8$. A series of tests validating our galaxy–galaxy lensing measurements using FRANKENZ photo-z's can be found in Speagle et al. (2019).

5.3 CS82

The CS82 survey is 160 deg^2 (before masking cuts are applied) of imaging data along the SDSS Stripe 82 region. We briefly summarize the key features of the CS82 weak-lensing catalogue and refer the reader to Leauthaud et al. (2017) for further details. CS82 is is built from 173 MegaCam (Boulade et al. 2003) i'-band images taken under excellent seeing conditions (median seeing is 0.6 arcsec). The limiting magnitude of the survey is i'∼24.1. The images were processed based on the procedures presented in Erben et al. (2009) and shear catalogues were constructed using the same weak-lensing pipeline developed by the CFHTLenS collaboration using the *lens*fit Bayesian shape measurement method (Miller et al. 2013). A series of quality cuts is applied to construct the CS82 source catalogue (see Leauthaud et al. 2017 for details). Shear calibration was performed using the same methodology as CFHTLenS.

Photo-z's were computed from SDSS *ugriz* imaging by Bundy et al. (2015) using the Bayesian photometric redshift software BPZ (Benítez 2000; Coe et al. 2006). The peak of the posterior distribution given by BPZ, $z_{\rm B}$, is used for sources redshifts, and a fiducial photo-z quality cut of ODDS > 0.5 is applied to reduce the catastrophic outlier rate. The CS82 survey overlaps with a number of spectroscopic surveys. Among these, the DEEP2 (Newman et al. 2013) catalogue spans the magnitude range of the CS82 and was the most useful in terms of assessing the photometric redshifts. A representative spectroscopic sample was used to estimate and correct for the bias caused by photometric redshifts (Leauthaud et al. 2017). After applying photo-z quality cuts, the CS82 source catalogues corresponds to an effective weighted galaxy number density[15] of $n_{\rm eff} = 4.5$ galaxies arcmin^{-2}.

5.4 CFHTLenS

CFHTLenS analysed 172 deg^2 of imaging data from the wide component of the CFHT Legacy survey (*ugriz* imaging to a 5σ point source limiting magnitude of $i_{\rm AB} = 25.5$). The observing strategy reserved the best seeing (seeing <0.8 arcsec) conditions for the lensing i-band filter, the primary object detection filter, and follow-up with the other bands in the poorer seeing conditions.

The data reduction for CFHTLenS was conducted with the THELI pipeline (Schirmer et al. 2004; Erben et al. 2005) following the procedures outlined in Erben et al. (2013). The data set shares a similar data processing pipeline to KiDS, where the shape measurement of

[15]Here, we use $n_{\rm eff}$ as defined by equation (1) in Heymans et al. (2012)

galaxies was conducted using the *lens*fit model fitting code (Miller et al. 2013). Shear multiplicative bias terms were characterized as a function of the signal-to-noise ratio and galaxy size using image simulations, thereby allowing for the calculation of the multiplicative bias term for an arbitrary selection of galaxies.

Photometric redshifts, z_B, were estimated using the Bayesian photometric redshift algorithm (BPZ; Benítez 2000) and *ugriz*-band data. A probability distribution of true redshifts was estimated from the sum of the uncalibrated BPZ redshift probability distributions. As such the CFHTLenS analysis represents a snapshot of our best understanding of photometric redshift accuracy in 2012 (Hildebrandt et al. 2012). This approach has since been demonstrated to carry systematic error (Choi et al. 2016). Current weak-lensing surveys focus on optimal methods to calibrate their photometric redshift distributions (e.g. Hildebrandt et al. 2016; Hoyle et al. 2018; Tanaka et al. 2018; Buchs et al. 2019; Speagle et al. 2019; Wright et al. 2020).

For cosmic shear, Choi et al. (2016) found the largest bias in the mean redshift of the source sample to be 0.04. This corresponds to a shift of 0.6σ in the cosmological constraints for cosmic shear. However, the response of galaxy–galaxy lensing to redshift errors is different and the Choi et al. (2016) results cannot be directly translated into errors on $\Delta\Sigma$. Instead, here we evaluate the impact of this photo-z bias on $\Delta\Sigma$ and include this in the reported CFHTLenS systematic error budget (Section 6.5).

5.5 KiDS

The KiDS survey (Kuijken et al. 2015) will span 1350 deg^2 on completion, in two patches of the sky with the *ugri* optical filters, as well as forced-aperture photometry on five infrared bands from the overlapping VISTA Kilo-degree Infrared Galaxy (VIKING) survey (Edge et al. 2013), yielding the first well-matched wide and deep optical and infrared survey for cosmology and more accurate photometric redshifts. It uses the wide-field camera, OmegaCAM, at the VLT Survey Telescope at ESO Paranal Observatory, optimally designed for lensing with high-quality optics and seeing conditions in the detection r-band filter with a median of <0.7 arcsec.

This paper uses 450 deg^2 of KiDS-VIKING nine-band imaging data (KV-450; Wright et al. 2019). With an effective, unmasked area of 360 deg^2, this data set has an effective number density of $n_{\rm eff} = 6.93$ galaxies arcmin^{-2}. Galaxy shapes were measured from the r-band data using a self-calibrating version of *lens*fit (Miller et al. 2013; Fenech Conti et al. 2017). A weight, w_s, is also assigned based on the quality of the shape measurement. Utilizing a large suite of image simulations, the multiplicative shear bias was deemed to be at the per cent level for the entire KiDS ensemble (Kannawadi et al. 2019).

The redshift distribution for KiDS galaxies was determined via four different approaches, which were shown to produce consistent results in a cosmic shear analysis (Hildebrandt et al. 2020). The preferred method of that analysis, used here, is the 'weighted direct calibration' [direct calibration method (DIR)] method, which exploits an overlap with deep spectroscopic fields. Following the work of Lima et al. (2008), the spectroscopic galaxies are re-weighted in nine-band colour space to obtain a true redshift distribution. A sample of KiDS galaxies is selected using their associated z_B value, estimated from the nine-band photometry as the peak of the redshift posterior output by BPZ (Benítez 2000). The resulting redshift distribution is well-calibrated in the range $0.1 < z_B \leq 1.2$ (see Wright et al. 2020 for a detailed mock catalogue analysis that quantified the accuracy of the DIR method for a KV-450 like survey).

5.6 DES

The DES survey conducted its first year of survey operation (Y1) between 2013 August 31 and 2014 February 9 (Drlica-Wagner et al. 2018) from the 4-meter Blanco Telescope and the Dark Energy Camera (Flaugher et al. 2015). DES Y1 covers two non-contiguous areas near the southern galactic cap: The 'SPT' area (1321 deg^2), which overlaps the footprint of the South Pole Telescope Sunyaev-Zel'dovich Survey (Carlstrom et al. 2011), and the 'S82' area (116 deg^2), which overlaps the Stripe-82 deep field of the SDSS (Annis et al. 2014). Each area within these footprints was revisited three to four times to reach sufficient photometric depth in the four *griz* DES bands. In this paper, we only use the S82 area which overlaps with BOSS.

For the DES Y1 data, two independent shape catalogues were created: METACALIBRATION (Sheldon & Huff 2017; Huff & Mandelbaum 2017) and IM3SHAPE (Zuntz et al. 2013) both of which were found suitable for cosmological analyses. In the present study, we only consider the METACALIBRATION shape catalogue as it provides the larger surface source density of 6.28 arcmin^{-2} over the full Y1 footprint. The METACALIBRATION approach, instead of relying on calibrating shear bias from image simulations, makes use of the actual observed galaxy images to de-bias shear estimates, estimating a response R of measured ellipticity to shear. METACALIBRATION also provides a photometric catalogue derived from its internal galaxy model fits.

Photometric redshifts for the Y1 source catalogue were initially estimated using the BPZ algorithm, and the mean redshift of the resulting sample of galaxies calibrated by matching to galaxies with high-quality photometric redshifts in COSMOS (Laigle et al. 2016) by magnitude, colour and size (Hoyle et al. 2018), and by cross-correlation with a photometric LRG sample (Davis et al. 2017; Gatti et al. 2018). To properly account for selection effects, the photometric redshifts were calculated with two different input photometries, one using the fiducial DES Y1 GOLD photometry catalogue (Drlica-Wagner et al. 2018, for $n(z)$ estimation), and one using the METACALIBRATION derived photometry catalogues (for selection and weighting of galaxies). The performance of the redshift estimates have been validated and McClintock et al. (2019) quantified the COSMOS-derived bias correction for $\Delta\Sigma$.

6 COMPUTATION OF $\Delta\Sigma$

This section describes how each team computed $\Delta\Sigma$. This section provides a snap-shot picture of each different team's approach to the computation of $\Delta\Sigma$ (also see Table 2). For the full details on the methodology, and tests regarding the validity of each computation, the reader is referred to survey specific papers. See Section 2 for an introduction to terminology and for the definition of $\Delta\Sigma$.

6.1 Computation of $\Delta\Sigma$ and notation

Here we define common notation used in the computation of $\Delta\Sigma$. We then give the details of each team's specific computation.

6.1.1 Redshifts and critical surface mass density

Lenses have spectroscopic redshifts and their redshifts are noted z_L. For source galaxies, redshift probability distributions are denoted

as $p(z_s)$, point source estimates of redshifts are denoted z_s, and an ensemble redshift distribution is denoted $n(z_s)$. Photometric redshifts are a noisy and, in some cases, a biased estimate of the true source redshift. For this reason boost, dilution, and $f_{\rm bias}$ corrections are sometimes required when computing the critical surface density.

Teams employ three different approaches for the computation of the critical surface mass density. First, the critical surface mass density can be computed for each lens–source pair and with a source point source estimate following equation (2). This is the methodology employed by SDSS, HSC, CS82, and DES.

Secondly, the critical surface density may be computed for a lens-source pair but using a $p(z)$. Here the inverse critical surface density is estimated:

$$\Sigma^{\rm inv}_{\rm c,pz}(z_{\rm L}, z_{\rm s}) = \frac{4\pi G D_{\rm A}(z_{\rm L})}{c^2} \int_{z_{\rm L}}^{\infty} dz_{\rm s}\, p(z_{\rm s}) \frac{D_{\rm A}(z_{\rm L}, z_{\rm s})}{D_{\rm A}(z_{\rm s})}, \tag{23}$$

and the critical surface density is then:

$$\Sigma_{\rm c,pz} = 1/\Sigma^{\rm inv}_{\rm c,pz}. \tag{24}$$

If the per-source photometric redshift probability distributions are an accurate representation of the statistical and systematic redshift error, then this approach removes the necessity for a dilution or $f_{\rm bias}$ correction, when $p(z_s)$ is normalized as $\int_0^\infty p(z_s) dz_{\rm s} = 1$. As shown, for example in Hildebrandt et al. (2020), however, the posterior redshift PDFs estimated by BPZ, are inherently biased. As such, this approach is not recommended, but we include it nevertheless as this was the methodology originally employed by CFHTLenS in Ford et al. (2015), where additionally the $p(z_s)$ was normalized as $\int_{z_{\rm L}}^\infty p(z_s) dz_{\rm s} = 1$ such that the dilution factor was unaccounted for.

Thirdly, the critical surface mass density may also be computed for each lens galaxy with redshift $z_{\rm L}$ but for the ensemble source population (after lens-source separation cuts). In this case, the effective inverse critical surface mass density is noted $\overline{\Sigma}^{\rm inv}_{\rm c,nz}$ and is computed following:

$$\overline{\Sigma}^{\rm inv}_{\rm c,nz}(z_{\rm L}, n(z_{\rm s})) = \frac{4\pi G D_{\rm A}(z_{\rm L})}{c^2} \int_{z_{\rm L}}^{\infty} dz_{\rm s}\, n(z_{\rm s}) \frac{D_{\rm A}(z_{\rm L}, z_{\rm s})}{D_{\rm A}(z_{\rm s})}, \tag{25}$$

where $\int_0^\infty n(z_{\rm z}) = 1$. The effective surface density is then:

$$\overline{\Sigma}_{\rm c,nz} = 1/\overline{\Sigma}^{\rm inv}_{\rm c,nz}. \tag{26}$$

If the ensemble redshift distribution estimate is an accurate and unbiased measurement of the true ensemble distribution (for example through calibration with an external spectroscopic sample) then both the dilution and $f_{\rm bias}$ correction are automatically accounted for with this approach. This is the methodology employed by KiDS.

Testing of the equivalence between these different approaches is warranted and will be carried out using mock simulations in the DESI lensing mock challenge (Lange et al., in preparation).

6.1.2 *Weighting schemes*

An inverse variance weight is applied to lens–source pairs and is noted:

$$w_{\rm Ls} = \frac{\Sigma_{\rm c}^{-2}}{\sigma_{\rm e}^2 + \sigma_{\rm rms}^2} \equiv \frac{\Sigma_{\rm c}^{-2}}{\sigma_{\rm s}^2}, \tag{27}$$

where $\sigma_{\rm s}$ is the total shape noise, $\sigma_{\rm rms}$ is the intrinsic shape dispersion per component, and σ_e is the per-component shape measurement error. For shape catalogues that use *lens*fit, the *lens*fit weight is $w_{\rm lf}^{-1} \sim \sigma_{\rm e}^2 + \sigma_{\rm rms}^2$ and $w_{\rm lf}$ is used for weighting (note that in the notation used here, $w_{\rm Ls}$ includes the $\Sigma_{\rm c}^{-2}$ term, whereas $w_{\rm lf}$ is the *lens*fit approximation to the total shape noise).

DES uses a different weight, first because they choose to normalize the individual source's contribution to shear rather than in units of Σ, and secondly because they do not weight by the inverse shape noise variance of the individual source. The equation for the DES weight applied to each source's shape is

$$w^{\gamma}_{\rm Ls} = \Sigma^{-1}_{\rm c,MCAL}\left(z_{\rm L}, z^{\rm MCAL}_{\rm s,mean}\right) \tag{28}$$

where MCAL indicates a metacalibration redshift (see Section 6.7). This weight can be thought of as a weight on γ, hence the $\Sigma_{\rm c}^{-1}$ term instead of the $\Sigma_{\rm c}^{-2}$ term used in equation (27).

6.1.3 *f_{bias} correction factor*

The $f_{\rm bias}$ correction factor accounts for biases that arise when converting γ to $\Delta\Sigma$ using sources with photometric redshifts (see Section 2.2 and a more detailed derivation in Appendix B). This term is computed using a representative sample of galaxies (hereafter called the 'calibration catalogue') following:

$$f^{-1}_{\rm bias} = \frac{\sum_{\rm Ls} w_{\rm calib,s}\, \sigma_s^2\, \Sigma^{-1}_{\rm c,Ls,P}\, \Sigma^{-1}_{\rm c,Ls,T}}{\sum_{\rm Ls} w_{\rm calib,s}\, \sigma_s^2\, \Sigma^{-2}_{\rm c,Ls,P}}, \tag{29}$$

where $\sigma_{\rm s}$ is shape noise of calibration sources, $\Delta\Sigma_{\rm crit,Ls,P}$ represents the (possibly biased) value of $\Delta\Sigma_{\rm crit}$ measured with photo-zs, $\Delta\Sigma_{\rm crit,Ls,T}$ represents the true value of $\Delta\Sigma_{\rm crit}$, and the sum is performed over all possible pairs of lenses and sources from the calibration catalogue. The calibration weight, $w_{\rm calib}$ may account for: (a) the sample variance of the calibration sample or (b) colour differences between the overall source sample and the calibration sample. The form of $f_{\rm bias}$ written here includes the dilution effect by sources that scatter above $z_{\rm L}$ but which are actually located at lower redshifts than $z_{\rm L}$. equation (29) is written in terms of $f^{-1}_{\rm bias}$ because of the dilution factor and to avoid issues in the computation of $\Delta\Sigma_{\rm T}$ when $z_s < z_{\rm L}$ (resulting in an ill defined $\Sigma_{\rm c}$ term). The relation between $\Delta\Sigma_{\rm P}$ and $\Delta\Sigma_{\rm T}$ is:

$$\Delta\Sigma_{\rm T} = f_{\rm bias}\Delta\Sigma_{\rm P}. \tag{30}$$

DES employs a similar equation but without the shape noise weight.[16] Specifically, DES uses $\tilde{f}$ defined as:

$$\tilde{f}^{-1}_{\rm bias} = \frac{\sum_{\rm Ls} w_{\rm calib,s} \Sigma^{-1}_{\rm c,Ls,P} \Sigma^{-1}_{\rm c,Ls,T}}{\sum_{\rm Ls} w_{\rm calib,s} \Sigma^{-2}_{\rm c,Ls,P}}, \tag{31}$$

6.1.4 *Effective lens redshift*

Finally, each survey also computes the effective lens redshift for each of the samples. The effective redshift of each lens sample is

$$z_{\rm eff} = \frac{\sum_{\rm Ls} w_{\rm sys} w_{\rm Ls} z_{\rm L}}{\sum_{\rm Ls} w_{\rm sys} w_{\rm Ls}}, \tag{32}$$

where the sum is taken over all lens source pairs and $w_{\rm sys}$ is the BOSS systematic weight applied to each lens.

6.2 SDSS

The methodology of Singh et al. (2018) is used to compute $\Delta\Sigma$. A photo-z point estimate, $z_{\rm s}$, is used to select source galaxies behind lenses ($z_{\rm s} > z_{\rm L}$), as well as to compute the $\Sigma_{\rm c}$ factors and to

[16]The DES Y1 catalogue does not have shape noise estimates for source galaxies but it is expected that later versions of DES source catalogues will include such estimates.

weight each lens–source pair with the weighting scheme given in equation (27). The maximum likelihood redshift is taken as the point source estimate. The representative spectroscopic sample from Nakajima et al. (2012) is used to correct for biases arising from photometric redshifts. These corrections are of order 10 per cent (estimated at ~ 2 per cent accuracy, see also tests in Singh et al. 2018) and increase with the effective redshift of the lens sample.

Following Mandelbaum et al. (2005), the measurement around random points is subtracted to remove the additive systematic bias and also to obtain the optimal covariance (Singh et al. 2017). $\Delta\Sigma$ is computed as a function of physical radius r as:

$$\Delta\Sigma(r) = \frac{f_{\rm bias}}{2\mathcal{R}(1+m)}(\Delta\Sigma_{\rm L}(r) - \Delta\Sigma_{\rm R}(r)), \tag{33}$$

where $\Delta\Sigma_{\rm L}$ is the stacked signal around lens galaxies, $\Delta\Sigma_{\rm R}$ is the stacked profile around a much larger number of random positions that share the same redshift distribution as lenses, and $f_{\rm bias} \approx 1.09(1.2)$ is the correction for photo-z calibration errors for the L1 and the L2 samples, respectively. This factor corrects both for photo-z bias and the dilution of the signal caused by sources that are below the lens redshift but get scattered above it due to photo-z error. The $1 + m$ term is the correction for the shear multiplicative bias with $1 + m \sim 0.96$. $\mathcal{R}$ is the shear responsivity factor. The SDSS lensing catalogue employs a single m and $\mathcal{R}$ value for all galaxies, defined at the full shape catalogue level.

The signal around lens galaxies is computed as:

$$\Delta\Sigma_{\rm L}(r) = \frac{\Sigma^{r}_{\rm Ls} w_{\rm sys} w_{\rm Ls}\, \epsilon_{\rm t,Ls} \Sigma_{\rm c,Ls}}{\Sigma^{r}_{\rm Rs} w_{\rm sys} w_{\rm Ls}}, \tag{34}$$

where $\sum^{r}_{\rm Ls}$ indicates a sum over all lens–source pairs with separation r. The sum in the denominator is taken over random source pairs ($\Sigma^{r}_{\rm Rs}$) which applies a boost correction which is important at small scales [$r \leq 1$ (Mpc)]. The signal around random points, $\Delta\Sigma_{\rm R}(r)$, is computed in a similar fashion to equation (34) but the sums are taken over random–source pairs instead of lens–source pairs.

Shear calibration and photo-z's are both estimated to be around the 2 per cent level (Reyes et al. 2012; Nakajima et al. 2012). From tests using cross correlations, photo-z calibration uncertainty is around 5 per cent. We therefore quote 5 per cent as upper limit on the photo-z calibration systematics. Adding these in quadrature yields an estimated ~ 6 per cent systematic error.

The covariance of the measurements is estimated using jackknife method with 100 approximately equal area regions. The weighted mean redshifts of the L1 and L2 lens samples are $z_L = 0.223$ and $z_{\rm L} = 0.357$.

6.3 HSC

The methodology described in Speagle et al. (2019) is used to compute $\Delta\Sigma$. The HSC calculation closely follow the SDSS approach with a few differences that are highlighted below. The full details of the calculation, as well as a number of tests validating the robustness of the signals, can be found in Speagle et al. 2019. The `best` photo-z value from FRANKENZ is used as a point estimate for the photometric redshift for each source galaxy, $z_{\rm s}$. The `medium` photo-z quality cut from Speagle et al. (2019) is applied. This cut requires $\chi^2_5 \leq 6$ and $z_{\rm risk} \leq 0.25$, where χ^2_5 describes the goodness of fit using a five-degree χ^2 distribution and $z_{\rm risk}$ is the 'risk' that the point estimate is incorrect as defined in Tanaka et al. (2018). These photo-z cuts keep about 75 per cent of all source galaxies. Source-lens separation is performed by requiring $z_{\rm s} > z_{\rm L} + 0.1$ and $z_{\rm s} > z_{\rm L} + \sigma_{68}$ where σ_{68} is the 1σ confidence limit of the photo-z. In a similar fashion to equation (33), $\Delta\Sigma$ is computed as a function of physical radius r following

$$\Delta\Sigma(r) = f_{\rm bias}(\Delta\Sigma_{\rm L}(r) - \Delta\Sigma_{\rm R}(r)). \tag{35}$$

The signal around lens galaxies is computed as:

$$\Delta\Sigma_{\rm L}(r) = \frac{1}{2\mathcal{R}(r)[1+K(r)]} \frac{\Sigma^{r}_{\rm Ls} w_{\rm sys} w_{\rm Ls}\, \epsilon_t^{\rm (Ls)} \Sigma_{\rm c}^{\rm (Ls)}}{\Sigma^{r}_{\rm Ls} w_{\rm sys} w_{\rm Ls}}. \tag{36}$$

This equation is similar to equation (34) with three differences. First, the normalization in the denominator is $\Sigma^{r}_{\rm Ls}$ instead of $\Sigma^{r}_{\rm Rs}$ (summed weights over lens–source pairs instead of random–source pairs) because boost factor corrections are not applied. Secondly, whereas SDSS uses a single value for $\mathcal{R}$, here we compute:

$$\mathcal{R}(r) = 1 - \frac{\Sigma^{r}_{\rm Ls} w_{\rm sys} w_{\rm Ls}\, \sigma^2_{\rm rms,s}}{\Sigma^{r}_{\rm Ls} w_{\rm sys} w_{\rm Ls}}, \tag{37}$$

This is because in the HSC shape catalogue, $\sigma_{\rm rms}$ depends on galaxy properties like SNR and resolution (also see equation 23 in Speagle et al. 2019). Thirdly, in HSC, the correction for multiplicative bias is $1/[1 + K(r)]$ instead of $1 + m$. This is because in HSC, each galaxy has an m value (see Mandelbaum et al. 2018a for details about the calibration of HSC weak-lensing catalogue). As described in Speagle et al. (2019), $K(r)$ is computed following:

$$K(r) = \frac{\Sigma^{R}_{\rm Ls} w_{\rm sys} w_{\rm Ls}\, m_{\rm s}}{\Sigma^{R}_{\rm Ls} w_{\rm sys} w_{\rm Ls}}. \tag{38}$$

The signal around random points, $\Delta\Sigma_{\rm R}(r)$, is computed in a similar fashion to equation (34) but the sums are taken over random–source pairs instead of lens–source pairs.

The COSMOS many-band catalogue (Laigle et al. 2016) is used to compute corrections due to photo-z's biases and dilution effects (the $f_{\rm bias}$ term). For these signals, the values for $f_{\rm bias}$ range between $f_{\rm bias} = 1.00$ and $f_{\rm bias} = 1.02$. In Speagle et al. (2019), a number of tests were performed on the robustness of the gg-lensing signal with regards to the photo-z calibration. Each source galaxy has quantities denoted $P_{\rm phot}$ and $F_{\rm phot}$ which indicate what kind of redshift it was primarily trained on (e.g. photo-z, spec-z, grism-z). By computing the gg-lensing signal with various values of $P_{\rm phot}$ and $F_{\rm phot}$, Speagle et al. (2019) showed that the gg-lensing signals are stable with respect to the origin of the training redshifts.

To compute the uncertainty of the $\Delta\Sigma$ signal, lens and random samples are grouped into 41 roughly equal-area sub-regions. A $N = 10\,000$ bootstrap re-sampling is used to estimate errors for $\Delta\Sigma$. The weighted mean redshifts of the four lens samples are $z_{\rm L} = 0.23$, $z_{\rm L} = 0.36$, $z_{\rm L} = 0.49$, $z_{\rm L} = 0.59$. The code used to compute $\Delta\Sigma$ (DSIGMA) is publicly available at https://github.com/johannesulf/dsigma. The systematic error is estimated to be of order 5 per cent (roughly Gaussian and 1σ).

6.4 CS82

The CS82 lensing signals are computed using the same code as HSC (DSIGMA). The main difference with Leauthaud et al. (2017) is that here the signal around random points in subtracted. But this does not have a large effect on the results and the derived signals are consistent with those derived in Leauthaud et al. (2017). Photometric redshifts are derived using *ugriz* photometry and the BPZ algorithm. Each source galaxy is assigned a point source redshift corresponding to the $z_{\rm B}$ value from BPZ. A cut of ODDS > 0.5 was applied to the source catalogue in order to reduce the number of source galaxies with catastrophic redshift failures. Source background selection is

performed by requiring that $z_s > z_L + 0.1$ and $z_s > z_L + \sigma_{95}/2.0$ where σ_{95} is the 95 per cent confidence limit on the source redshift. Leauthaud et al. (2017) showed that the CMASS lensing signal did not vary when a more stringent lens-source separation scheme was employed. Boost factors were not applied. The $f_{\rm bias}$ term was applied with values ranging from $f_{\rm bias} = 0.98$ to $f_{\rm bias} = 1.03$ using a representative sample of spectroscopic redshifts (reweighed to match the colour and magnitude distribution of the source sample) described in Leauthaud et al. (2017).

Errors on $\Delta\Sigma$ are computed via jack-knife. Because the same code is used as for HSC (`dsigma`), all other aspects of the calculation are as given in Section 6.3. The weighted mean redshifts of the four lens samples are $z = 0.227$, $z = 0.362$, $z = 0.488$, and $z = 0.586$. The systematic error is roughly estimated to be ∼6 per cent (roughly Gaussian and 1σ).

6.5 CFHTLenS

The photometric redshift probability distribution for each galaxy, $p(z_s)$, is computed from *ugriz*-band photometry using the BPZ algorithm, as well as a point estimate redshift per galaxy, $z_s = z_B$ (Hildebrandt et al. 2012). Galaxies where the peak of their $p(z_s)$ are in the range $0.15 < z_B < 1.3$ are used. The full redshift probability distribution is used to measure $\Delta\Sigma$ and $z_B > z_L + 0.1$ is required. This lens-source separation has been shown to significantly reduce the amplitude of the boost correction (see for example, Amon et al. 2018a). The $p(z_s)$ is used to estimate $\Sigma^{\rm inv}_{\rm c,pz}$ (equation 23) for each source pair. The weighted stacked $\Delta\Sigma$ is then calculated via

$$\Delta\Sigma_{\rm L}(r) = \frac{1}{1+\overline{m}_{\rm s}} \frac{\sum_{\rm Ls} w_{\rm sys} w_{\rm lf}\, \epsilon_{\rm t}\, \Sigma^{\rm inv}_{\rm c,pz}}{\sum_{\rm Ls} w_{\rm sys} w_{\rm lf} (\Sigma^{\rm inv}_{\rm c,pz})^2}, \tag{39}$$

The multiplicative bias correction, $\overline{m}_{\rm s}$, is calculated for a given lens sample as

$$\overline{m}_{\rm s} = \frac{\sum_{\rm s} w_{\rm sys} w_{\rm lf}\, m_{\rm s}}{\sum_{\rm s} w_{\rm sys} w_{\rm lf}}, \tag{40}$$

where $m_{\rm s}$ is the per galaxy multiplicative bias and $w_{\rm lf}$ is the *lens*fit weight. The difference with regards to equation (38) used by HSC is that this equation uses $w_{\rm lf}$ instead of $w_{\rm Ls}$. The difference between these two quantities is that $w_{\rm Ls}$ includes a $\Sigma_{\rm c}^{-2}$ term.

The signal around random lenses is subtracted from the signal around the lenses,

$$\Delta\Sigma(r) = \Delta\Sigma_{\rm L}(r) - \Delta\Sigma_{\rm R}(r). \tag{41}$$

Following Ford et al. (2015), boost, dilution, and $f_{\rm bias}$ correction factors are not calculated or applied. The error that is then incurred is accounted for in this analysis with a significant systematic error budget. The methodology of Xia et al. (2020) is used to compute a systematic error due to the error in the uncalibrated photometric redshifts, $p(z_s)$. A photo-z shift of $\delta z_s = 0.04$ is used to capture the photo-z bias found by Choi et al. (2016). The $p(z_s)$ is shifted by $\pm\delta z_s$ and two new functions $\Sigma^{\rm inv}_{\rm c}(z_{\rm l})^{\pm}$ are computed. The full measurement and error analysis is repeated using both the $\Sigma^{\rm inv}_{\rm c}(z_{\rm l})^{+}$ and $\Sigma^{\rm inv}_{\rm c}(z_{\rm l})^{-}$. The difference, $\Delta\Sigma_{\rm bias}$, is averaged over all scales. This photo-z uncertainty is the main systematic uncertainty for CFHTLenS. This systematic error is estimated to be up to 6 per cent for the LOWZ lens sample and up to 10 per cent for CMASS (roughly Gaussian and 1σ). After unblinding, a 5 per cent systematic error on m (Kuijken et al. 2015; Kilbinger et al. 2017) was also included. This increased the systematic errors but did not change any of the main conclusions. The final numbers are reported in Table 5.

Statistical errors are computed via bootstrapping over measurements using 1000 patches. The weighted mean redshifts of the four lens samples are $z_L = 0.23$, $z_L = 0.36$, $z_L = 0.49$, and $z_L = 0.60$.

6.6 KiDS

The KiDS lensing signal is computed similarly to the methodology outlined in Dvornik et al. (2018) and Amon et al. (2018b). A point estimate of the photometric redshift per galaxy, z_B, is derived using *ugriZYJHK*$_s$ photometry and the BPZ algorithm. This redshift is used to define the source samples and for source-lens separation. The source galaxy sample is first limited to $0.1 < z_B \leq 1.2$. Then, further source-lens separation cuts are applied to significantly reduce the amplitude of the boost correction. These are defined as $z_B > z_L + 0.1$, following tests in Amon et al. (2018a).

The ensemble redshift distribution of the source sample behind each lens, $n(z_s|z_L)$, is estimated using a DIR that employs a diverse and representative set of spectroscopic samples (Hildebrandt et al. 2020). Specifically DIR calibrated redshift distributions $n(z)$ are determined for a series of photometric redshift z_B slices of width 0.1. A critical surface density is then computed (equation 26) for a series of discrete lens values ($z_{\rm i,L} = 0.0, 0.1, 0.2$, etc.) and a composite DIR-calibrated source redshift distribution of 'background' galaxies with $z_B > z_{\rm i,L} + 0.1$. Linear interpolation is then used to compute the critical surface density for each lens in the full KiDS-BOSS sample. If the DIR-calibration results in an unbiased and accurate representation of the true source redshift distribution then both the dilution and $f_{\rm bias}$ correction are already included with this approach.

The multiplicative shear calibration correction (Kannawadi et al. 2019) is estimated for the ensemble source and lens galaxy population. Extending the method described in Dvornik et al. (2018) to the higher KV-450 redshifts, the shear calibration is estimated for 11 linear source photometric redshift bins between $0.1 < z_B \leq 1.2$. These corrections are then optimally weighted and stacked following:

$$\overline{m} = \frac{\sum_{\rm i} w'_{\rm i} m_{\rm s}}{\sum_{\rm i} w'_{\rm i}}, \tag{42}$$

where $w' = w_{\rm s} D(z_{\rm L}, z_{\rm s})/D(z_{\rm s})$. The resulting correction $\overline{m} \approx -0.014$ is independent of the distance r from the lens, and reduces the effects of multiplicative bias to within ±2 per cent (Kannawadi et al. 2019).

The signal around lens galaxies is computed:

$$\Delta\Sigma_{\rm L}(r) = \frac{1}{1+\overline{m}} \frac{\sum_{\rm Ls} w_{\rm sys} w_{\rm Ls}\, \epsilon_{\rm t}\, \overline{\Sigma}_{\rm c,nz}}{\sum_{\rm Ls} w_{\rm sys} w_{\rm Ls}}, \tag{43}$$

where $\overline{\Sigma}_{\rm c,nz}$ is given in equation (26).

The signal around random lenses is subtracted as follows:

$$\Delta\Sigma(r) = \Delta\Sigma_{\rm L}(r) - \Delta\Sigma_{\rm R}(r). \tag{44}$$

Errors are computed using a bootstrap method using regions of 4 deg^2. The weighted mean redshifts of the four lens samples are $z_L = 0.23$, $z_L = 0.36$, $z_L = 0.49$, $z_L = 0.58$.

Similar to the method employed by CFHTLenS, KiDS computes a contribution to the systematic uncertainty due to the error in the sample's calibrated redshift distribution, $n(z_s)$, by reporting an additive systematic error. This is determined by propagating $-0.06 < \delta z_s < 0.014$, as advised by Wright et al. (2020). The difference between the two measurements, $\Delta\Sigma_{\rm sys}$, is averaged over all scales and taken as the systematic error. This systematic error is estimated to be up to 2 per cent for LOWZ and up to 3 per cent for CMASS. This is the dominant systematic uncertainty for the KiDS measurements. After unblinding, the systematic error on m

as estimated in Kannawadi et al. (2019) was also included. This increased the systematic errors by 1 per cent but did not change any of the main conclusions. The final numbers are reported in Table 5.

6.7 DES

The DES lensing signal is computed following the methodology outlined in McClintock et al. (2019) and using the METACALIBRATION weak-lensing source galaxy catalogue for DES Y1 (Zuntz et al. 2018).

The METACALIBRATION algorithm (Huff & Mandelbaum 2017; Sheldon & Huff 2017) provides estimates on the ellipticity ϵ of galaxies, the response of the ellipticity estimate on shear R_γ, and of the ensemble mean ellipticity on shear-dependent selection $\mathsf{R}_{\rm sel}$. These are applied in the shear estimator to correct for the bias of the mean ellipticity estimates.

The DES shear response is broken into two terms: $\mathsf{R}^{\rm T}_{\gamma,s}$ is the shear response measured for individual galaxies, averaged over both ellipticity components, and $\langle \mathsf{R}^{\rm T}_{\rm sel} \rangle$ is the shear response of the source selection. The latter is a single mean number computed for each source galaxy ensemble. The DES catalogue also contains a multiplicative bias correction term (one number per source catalogue, similar to SDSS).

Two different photometric redshift estimates are used. The first is based on fluxes measured in the METACALIBRATION process. In this case, the redshift used is the mean of the $p(z)$ estimated from the METACALIBRATION photometry and is denoted $z^{\rm MCAL}_{\rm mean}$. The second is a random draw from the $p(z)$ estimated from the Y1 GOLD MOF photometry (Drlica-Wagner et al. 2018; hereafter denoted $z^{\rm MOF}_{\rm MC}$). Both are estimated using the BPZ algorithm (Hoyle et al. 2018). In order to properly account for the selection response term, the METACALIBRATION redshifts are used for source selection ($z^{\rm MCAL}_{\rm mean} > z_{\rm L} + 0.1$) and for the weight, $w^{\gamma}_{\rm Ls}$. The $z^{\rm MOF}_{\rm MC}$ redshifts, preferable due to the higher quality of the photometric information, are used to convert shear to $\Delta\Sigma$. This sample can be at $z^{\rm MOF}_{\rm MC} < z_{\rm L}$ despite the $z^{\rm MCAL}_{\rm mean}$-based source selection selection. That fact that different redshifts are used to weight the signal and to compute $\Delta\Sigma$ requires a modified $\Delta\Sigma$ estimator is described below.

Equation (2), and point source redshifts, are used to compute the critical surface density. However, $z^{\rm MCAL}_{\rm mean}$ is used for $w^{\gamma}_{\rm Ls}$ and $z^{\rm MOF}_{\rm MC}$ is used for $\Sigma_{\rm c,MOF}$.

Photometric redshift estimates and their associated uncertainties are calibrated using the Laigle et al. (2016) COSMOS photometric redshifts and using the algorithms described in Hoyle et al. (2018) and McClintock et al. (2019). Unlike the DES shear two-point functions (Abbott et al. 2018; Prat et al. 2018; Troxel et al. 2018a), the calibration of redshifts for $\Delta\Sigma$ are not refined by the result of the cross-correlation techniques (Davis et al. 2017; Gatti et al. 2018).[17] The consistency of the two (Hoyle et al. 2018), however, is evidence for the validity of the former.

The lensing estimator is given by

$$\Delta\Sigma(r) = \frac{1}{1+f_{\rm cl}} \frac{1}{\hat{f}^{-1}_{\rm bias} + m} \left(\Delta\Sigma_L(r) - \Delta\Sigma_R(r)\right), \tag{45}$$

with the signal around lens galaxies estimated as

$$\Delta\Sigma_L(r) = \frac{\sum\limits_{\rm Ls} w_{\rm sys}\, w^{\gamma}_{\rm Ls}\, \Sigma^{-1}_{\rm c,Ls,MOF}}{\sum\limits_{\rm Ls} w_{\rm sys}\, w^{\gamma}_{\rm Ls}\, \Sigma^{-1}_{\rm c,Ls,MOF}\, \left(\mathsf{R}^{\rm T}_{\gamma,s} + \langle \mathsf{R}^{\rm T}_{\rm sel} \rangle\right)} \times \frac{\sum\limits_{\rm Ls} w_{\rm sys}\, w^{\gamma}_{\rm Ls}\, \epsilon^{\rm Ls}_{\rm t}}{\sum\limits_{\rm Ls} w_{\rm sys}\, w^{\gamma}_{\rm Ls}\, \Sigma^{-1}_{\rm c,Ls,MOF}}. \tag{46}$$

Equation (46) is equivalent to equation (12) in McClintock et al. (2019). Here, we have ordered the terms for comparison with the estimators used by other surveys. For instance, the correction by the mean response in the first term here is similar to the $1/(1+K)$ term in equation (36) and the $1/(1+m)$ term in equation (43). The second term can be interpreted as a weighted mean tangential ellipticity in the nominator, normalized by a weighted mean $\Sigma^{-1}_{\rm c}$ estimated from MOF photometry. These use the METACALIBRATION-derived weights of equation (28).

Equation (45) subtracts the signal around random points and corrects it for systematic errors in photometric redshifts through $\hat{f}^{-1}_{\rm bias}$ and shear through a multiplicative bias correction m. Terms proportional to $\hat{f}^{-1}_{\rm bias} \times m$ are neglected. The signal is divided by $(1 + f_{\rm cl})$ to apply a boost factor. Here, $f_{\rm cl}$ is the fractional contribution from galaxies falsely identified as sources to the weighted mean shear, estimated using $p(z)$ decomposition (Gruen et al. 2014; Varga et al. 2019). All correction terms are defined and estimated as in McClintock et al. (2019).[18] The analysis setup used in these calculations is made publicly available in the XPIPE package.[19]

The measurement used here differs from various other DES analyses where systematic uncertainties were incorporated at the model/likelihood level, and their amplitudes varied according to their respective prior. In the present study, we apply the correction directly to the data vector, while estimating the corresponding systematic uncertainties for each lens redshift bin and for the inner and outer radial ranges, respectively. Shear calibration and photometric redshift systematic errors are estimated using the methodology of McClintock et al. (2019). The combined systematic uncertainty is estimated to be 2 per cent for the three lower redshift bins, and 3 per cent for the highest redshift bin. When incorporating the covariance of boost factor estimates to the net systematic error budget of the different radial ranges, we find a combined upper limit for the different radial ranges across all lens redshift bins respectively at the level of 2–3 per cent.

The weighted mean redshifts of the four lens samples are $z_{\rm L} = 0.23$, $z_{\rm L} = 0.36$, $z_{\rm L} = 0.49$, $z_{\rm L} = 0.59$.

7 HOMOGENEITY OF BOSS SAMPLES

The validity of the tests we seek to perform rely on the assumption that BOSS selects a homogeneous sample of foreground galaxies living in similar dark matter haloes. This assumption may be invalid if the properties of the CMASS and LOWZ samples (e.g. luminosity, colour, and stellar mass) vary spatially (each survey's submitted measurement is performed on a different patch of the sky and therefore with different samples drawn from BOSS). The BOSS clustering team identified several factors leading to inhomogeneity

[17] This analysis uses source galaxies at $z > 0.9$ where there is a dearth of spectroscopic galaxies for calibration purposes. See fig. 3 in Gatti et al. (2018)

[18] The impact of systematic weights is expected to be minor on the recovered boost factors, and as a computational simplification were assumed to be unity with respect to the boost factor calculation.

[19] https://github.com/vargatn/xpipe

in the BOSS samples (Ross et al. 2011, 2012; Reid et al. 2016; Ross et al. 2017). The goal of this section is to investigate inhomogeneity in the sub-regions probed by each survey footprint.

7.1 Overall homogeneity

We first study the overall homogeneity of the CMASS and LOWZ samples. We apply the masks of each of the lensing surveys to the BOSS catalogues to extract distributions of colour, $i_{\rm cmod}$, $i_{\rm fib2}$, z, and $\log_{10}(M_*)$ within each of the sub-regions. For M*, we use the Granada masses[20] described in Ahn et al. (2014). Fig. 3 and Table 3 demonstrate that the basic properties of the two samples are spatially homogeneous across the regions of interest. A further visual confirmation of this is provided by Fig. C2.

7.2 Galactic hemisphere

SDSS imaging is carried out into two large contiguous areas in the North Galactic Cap (NGC) and the South Galactic Cap (SGC). Fig. 1 displays the overlap between the NGC, the SGC, and the lensing surveys. The SDSS lensing catalogues have the most overlap with BOSS and cover most of the NGC and about half of the SGC. CFHTLenS and HSC have fields in both the NGC and the SGC. KiDS only overlaps with BOSS in the NGC. CS82 and DES only overlap with BOSS in the SGC. It is thus important to understand whether or not BOSS samples give rise to the same lensing signals in the NGC and the SGC.

Schlafly et al. (2010) and Schlafly & Finkbeiner (2011) found photometric offsets between the NGC and the SCG. The photometric calibration of the DR13 catalogue (which came after BOSS targeting was complete) was tied to Pan-STARRS 1 (PS1), as described in Finkbeiner et al. (2016). This procedure led to new flat fields and zero points in the g, r, i, and z bands, and new flat fields (but not new zero points) in the u band. The updated photometry results in a 0.015 magnitude difference in $c_\parallel$ compared to when BOSS targeting was performed.

Ross et al. (2017) find a 1 per cent difference in the number density of CMASS between the NGC and the SGC. Differences for LOWZ are larger: the projected density of LOWZ is 7.6 per cent higher in the SGC compared to the NGC. Ross et al. (2012) find these differences to be consistent with the level of colour offsets determined by Schlafly & Finkbeiner (2011) and that the sliding cut $c_\parallel$ imparts the largest differences. Because of the small colour offsets, the North and the South may correspond to slightly different galaxy populations. As a result, the BOSS team treats galaxies in the North and in the South as two separate samples.

Fig. A1 in Alam et al. (2017) investigates the differences in the clustering scales between the NGC and the SGC (also see Lee et al. 2019 for angular clustering $w(\theta)$). The power spectrum of CMASS is consistent for both hemispheres over the scales $0.05 < k\,[h^{-1}{\rm Mpc}] < 0.3$. There is, however, an amplitude shift in the power spectrum $P(k)$ (a 4 per cent shift in the amplitude of the power spectrum monopole) for LOWZ but it can be explained by taking into account the colour shifts between SDSS photometry in the north and south described previously (see appendix A in Alam et al. 2017 and Lee et al. 2019).

Because galaxy number counts are steep,[21] a small colour offset can easily result in variations in number density, but the variations in the galaxy (dark matter halo) selection may still be comparatively small. Using the calculation outlined in Appendix A, and assuming the Leauthaud et al. (2012) stellar-to-halo mass relation, a 0.015 shift in flux corresponds to a halo mass shift of 0.006 dex. This imparts less than a 1 per cent shift on $\Delta\Sigma$. However, these are only rough estimates based on the impact of magnitude shifts on M^*. Without a detailed understanding of the impact of the colour offsets on the galaxy (and underlying halo) selection, it is not trivial to translate differences in clustering amplitudes or number densities into differences in $\Delta\Sigma$. We therefore implement additional tests of the impact of North versus South via two alternative methods. First, we estimate the mean shift in M^* in different regions directly. Secondly, we use lensing from SDSS to perform a direct and empirical test on potential shifts in $\Delta\Sigma$.

We first consider the DR14 catalogue with updated and better calibrated photometry from Pan-STARRS 1 (PS1), as described in Finkbeiner et al. (2016). We take the existing LOWZ and CMASS catalogues and cross-correlate them with the DR14 photometry.[22] Fig. 3 shows that the updated photometry does not have a large impact on distributions in colour, $i_{\rm cmod}$, $i_{\rm fib2}$, z, and $\log_{10}(M_*)$. Table 1 lists differences between the mean values of these quantities and the overall BOSS sample. Differences are small and photometric calibration should therefore not impact $\Delta\Sigma$ (also see Appendix A).

We now carry out a more direct test of the impact of differences between the North and the South on $\Delta\Sigma$. The SDSS survey is the only lensing survey with enough coverage in both hemispheres to perform a direct test (see Fig. 1). SDSS is too shallow to accurately measure lensing for CMASS, so this test is limited to the LOWZ sample. However, as described above, we expect such effects to be more important for LOWZ than for CMASS. Fig. 4 shows $\Delta\Sigma$ measurements obtained for the full LOWZ sample, as well as from the North and South regions separately. Errors shown in the figure are obtained using jackknife (68 regions for the North, 32 regions for the South, and 100 regions for the full sample). When computing the ratio, we add the covariances from the North and South regions assuming that they are independent following: $\mathrm{cov}(R)/R^2 = \mathrm{cov}(\Delta\Sigma_N)/\Delta\Sigma_N^2 + \mathrm{cov}(\Delta\Sigma_S)/\Delta\Sigma_S^2$. The results from both regions are consistent with the difference being 5 ± 8 per cent. Lensing from SDSS yields the highest signal-to-noise measurements for LOWZ – hence if such differences are not detectable with SDSS, they are also not detectable with our other lensing data sets. Based on Figs 3 and 4, we conclude that differences between North and South are not a concern for the present study. None the less, for completeness, we also perform a post-unblinding test on the impact of North versus South in Section 8.3.

7.3 Stellar density and seeing

Ross et al. (2012) investigated how the number density ($n_{\rm gal}$) of DR9 BOSS galaxies varied with stellar density, seeing, Galactic extinction, and sky background (in the imaging that was used for targeting). No effects were found for the LOWZ sample.[23] Larger effects were detected for CMASS. Differences between CMASS and LOWZ are explained by the fact that LOWZ galaxies are on average considerably brighter than CMASS galaxies (Tojeiro et al. 2014).

[20] https://www.sdss.org/dr16/spectro/galaxy_granada/

[21] The number of galaxies in a sample rises steeply as a function of the limiting magnitude of the sample.

[22] Lee et al. (2019) studies how the sample changes with cuts applied to DES photometry, and a similar behaviour might be seen in the case of cuts applied to Pan-STARRS photometry.

[23] With the exception of the LOWZE2 and LOWZE3 samples which are excluded from the present study.

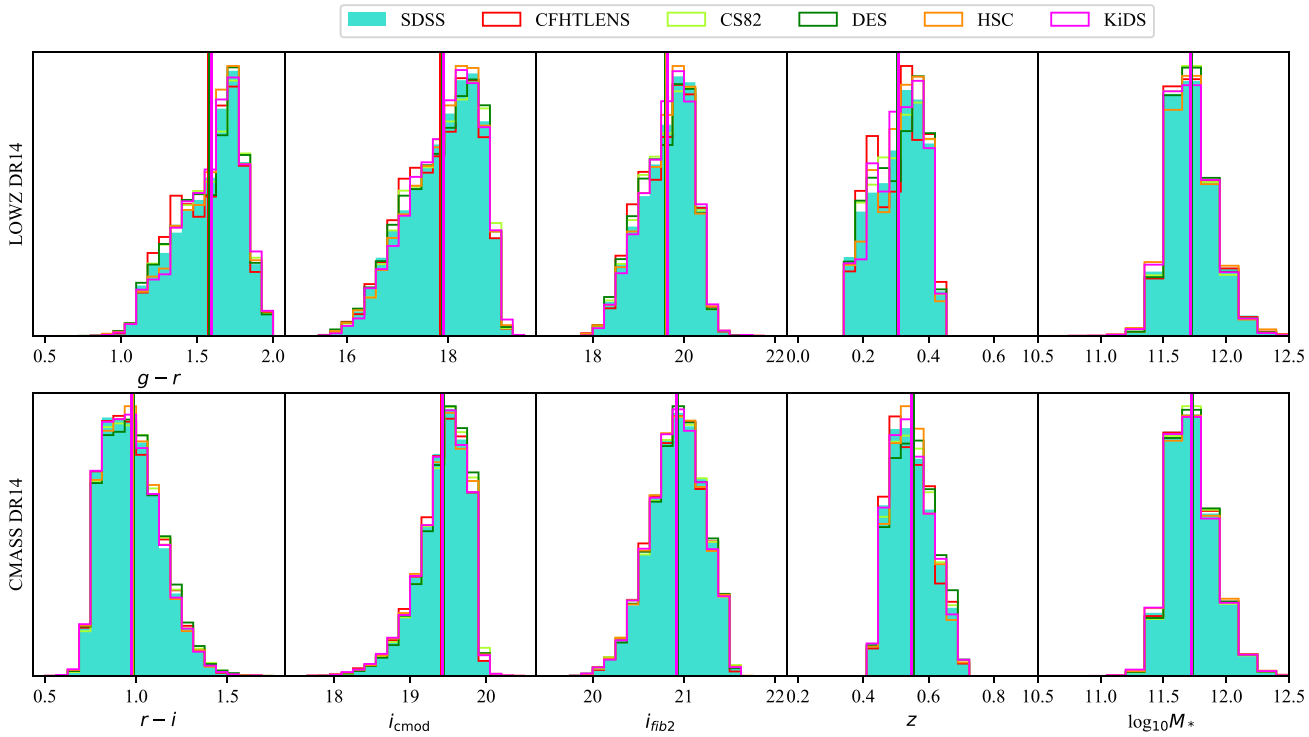

Figure 3. Distributions of colour, $i_{\rm cmod}$, $i_{\rm fib2}$, z, and $\log_{10}(M_*)$ for LOWZ (upper panels) and CMASS (lower panels) in the regions of overlap with each of the lensing surveys. The DR14 catalogue with photometry tied to Pan-STARRS 1 (PS1) as described in Finkbeiner et al. (2016) was used for magnitudes. The CMASS selection includes a cut at $i_{\rm fib2} < 21.5$ (equation 16). The reason this sharp cut is not apparent in this figure is because of scatter between DR14 and the original CMASS targeting catalogue.

Table 3. Mean values of photometric quantities for the overall CMASS and LOWZ samples. This table also provides differences between the mean quantities in each lensing survey footprint and the overall BOSS samples. We consider the colour differences used in the LOWZ (CMASS) selections. We use $(r - i)$ for CMASS and $(g - r)$ for LOWZ.

Mean values	$i_{\rm cmod}$	$g - r$	$r - i$	$\log_{10}M_*$
SDSS CMASS	19.4186 ± 0.0004	–	0.9780 ± 0.0002	11.7264 ± 0.0002
SDSS LOWZ	17.694 ± 0.001	1.5126 ± 0.0004	–	11.7227 ± 0.0003
Differences	$\Delta(i_{\rm cmod})$	$\Delta(g - r)$	$\Delta(r - i)$	$\Delta(\log_{10}M_*)$
		CMASS		
HSC Y1	0.0054 ± 0.003	–	0.0028 ± 0.001	0.001 ± 0.001
DES Y1	0.014 ± 0.003	–	0.0044 ± 0.001	0.005 ± 0.001
KiDS	0.0104 ± 0.002	–	−0.0011 ± 0.001	−0.004 ± 0.001
CFHTLenS	−0.00036 ± 0.003	–	−0.0017 ± 0.001	0.003 ± 0.002
CS82	0.020 ± 0.002	–	0.0064 ± 0.001	0.003 ± 0.001
		LOWZ		
HSC Y1	−0.0033 ± 0.01	0.009 ± 0.003	–	−0.006 ± 0.002
DES Y1	0.0202 ± 0.008	−0.007 ± 0.002	–	−0.004 ± 0.002
KiDS	0.0205 ± 0.009	0.011 ± 0.003	–	−0.013 ± 0.002
CFHTLenS	−0.01∗ ± 0.01	−0.013 ± 0.004	–	−0.002 ± 0.002
CS82	0.024 ± 0.007	−0.002 ± 0.002	–	−0.008 ± 0.002

For CMASS, stellar density ($n_{\rm star}$) was found to have the largest impact on $n_{\rm gal}$. The relationship between $n_{\rm gal}$ and $n_{\rm star}$ was found to depend on galaxy surface brightness where $i_{\rm fib2}$ was used as a proxy for surface brightness. As the stellar density increases, on average, galaxies with lower magnitudes in a 2 arcsec fiber are lost from the sample. In Reid et al. (2016), the functional form for $w_{\rm star}$ is given by:

$$w_{\rm star}(n_{\rm star}, i_{\rm fib2}) = (A_{\rm ifib2} + B_{\rm ifib2} n_{\rm s})^{-1} \quad (47)$$

where $A_{\rm ifib2}$ and $B_{\rm ifib2}$ depend on $i_{\rm fib2}$. Ross et al. (2012) found that applying $w_{\rm star}$ accounts for observed variations in $n_{\rm gal}$ with other quantities.

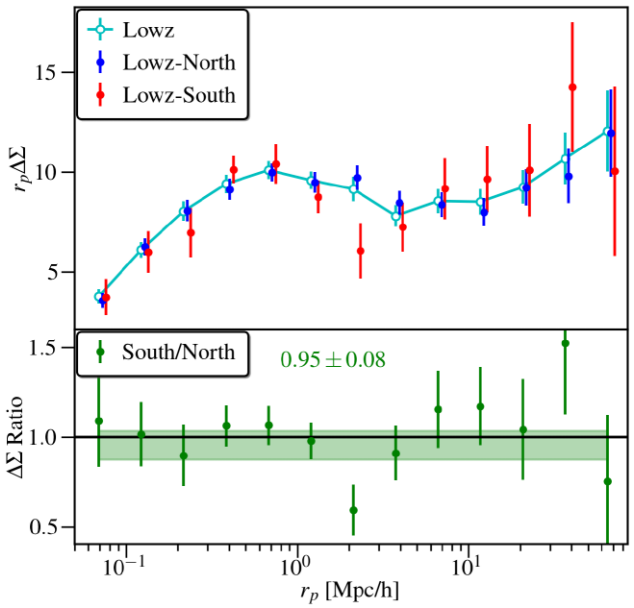

Figure 4. Impact of North versus South photometry differences on the $\Delta\Sigma$ signal for LOWZ. Upper panel: $\Delta\Sigma$ measured for the full LOWZ sample, as well as for the North and South regions separately. Signals are computed using the SDSS lensing catalogue. Lower Panel: ratio of $\Delta\Sigma$ from each hemisphere. The bands and the number quoted in the text are the mean values of the ratio obtained by fitting a constant to values between $0.5 < r_p < 15$ Mpc. The amplitude of $\Delta\Sigma$ is consistent between both hemispheres.

Ross et al. (2012) further show that the stellar density weights have a large impact on the clustering of CMASS which suggests that spatial variations in $n_{\rm star}$ might also be an important systematic effect for the present study. Clustering studies in BOSS adopt systematic weights that are designed to re-weight galaxies to a fixed overall number density. However, this scheme may not be appropriate for lensing. Indeed, the dark matter halo mass function displays a non linear relationship between number density and halo mass. Hence, weights designed to maintain a constant number density will not guarantee distributions of equal halo mass.

Fig. 5 displays distributions of $n_{\rm star}$ for regions that overlap with lensing surveys (we study variations in $n_{\rm star}$ directly rather than $w_{\rm star}$). There are clear differences in $n_{\rm star}$ distributions across lensing surveys.

As originally detailed in Ross et al. (2011), the observed density of the CMASS sample also correlates with local seeing because of the star galaxy separation cuts. There is no detected corresponding effect for LOWZ.[24] For CMASS, the effect is such that in poor seeing conditions, the number density decreases because compact galaxies are classified as stars and are removed from the sample. BOSS clustering studies employ a systematic weight, $w_{\rm see}$, to account for the seeing dependence of n_g for CMASS.

We tried a variety of tests to reweigh surveys to different effective distributions in $n_{\rm star}$ and seeing. However, we found such tests to be limited by (a) the small sizes of the lensing surveys at hand, and (b) strong spatial variations in the PSF associated with BOSS targeting. Fig. C1 in the appendix shows an example of the spatial variation of the PSF in one of the HSC fields. Instead, we opted for a more straightforward post-unblinding test. This is described in Section 8.3 and Figs 8 and 10.

[24]Except for the LOWZE3 sample which we removed from our analysis.

Ross et al. (2017) do not find clear evidence for correlations between $n_{\rm gal}$ and sky background, air mass, or extinction. In the future, it would be instructive to investigate whether or not these quantities have an impact on the lensing, but we do not explore these aspects in this paper.

8 RESULTS

The first goal of this paper is to search for trends in the data that could be due to systematic effects following the methodology outlined in Section 3.6. If found, such correlations could provide important clues as to the origins and level of systematic effects (including those effects that are 'known knowns'). Because we are specifically seeking to pin-point trends caused by systematic errors, in these tests, we use the reported statistical errors (Figs 8 through 13 display statistical errors). Section 8.5 discusses trends including both statistical and estimated systematic errors. Our second goal is to use the measured spread between the amplitudes of $\Delta\Sigma$ as an empirical and end-to-end estimate of systematic errors following the methodology outlined in Section 3.7. This estimate is then compared with the systematic errors as reported by each survey (Section 8.6). This test will help to determine if unknown systematic effects are present in the data. Unless mentioned otherwise, all of the results in this section were blinded according to the scheme presented in Section 3.3. It was agreed before unblinding that any homogeneity trends in the lens samples greater than 3σ would be discussed. Our lensing signals and code used to make the main figures are available at https://github.com/alexieleauthaud/lensingwithoutborders.

8.1 Comparison of $\Delta\Sigma$ and computation of amplitudes

Fig. 6 displays $\Delta\Sigma$ for the four lens samples. Results from SDSS are only shown for LOWZ because it was agreed before unblinding that SDSS might not be able to measure an accurate lensing signal for CMASS. Table 4 gives the overall signal to noise of the various measurements in each of the radial ranges. SDSS and HSC have the highest signal to noise for LOWZ and HSC has the highest signal to noise for CMASS.

We fit the overall amplitudes of each of the lensing signals using $\Delta\Sigma_{\rm HOD}$ as reference and for each of the three radial ranges (see Section 3.5). The results are shown in Fig. 7.

As an additional test, we also compute the amplitudes using only pairs of surveys (when using pairs, the weighting scheme changes, see Section 3.5). We use SDSS for LOWZ and HSC for CMASS as the reference survey. Results are unchanged when using a pair-wise weighting scheme, demonstrating that our amplitude fits are robust to the specifics of individual survey covariances.

8.2 Effective lens redshift

Each of the lensing catalogues has a different mean source redshift, with SDSS being the most shallow, and HSC the deepest. As a result, each survey imparts a different lens weight on each of the different samples. The effective lens redshift (see Section 6.1.4) for each of the bins is computed for each survey in Section 6. However, due to the relatively narrow lens redshift bins, we find only minor differences in the effective lens redshifts among surveys. Differences in $z_{\rm eff}$ are always less than $\Delta z = 0.02$. Using the Stripe 82 Massive Galaxy Catalogue (Bundy et al. 2015), we estimate how much the mean stellar mass of galaxies varies across this Δz. We find that ΔM^* variations are less than 0.015 dex over this the maximum Δz. As

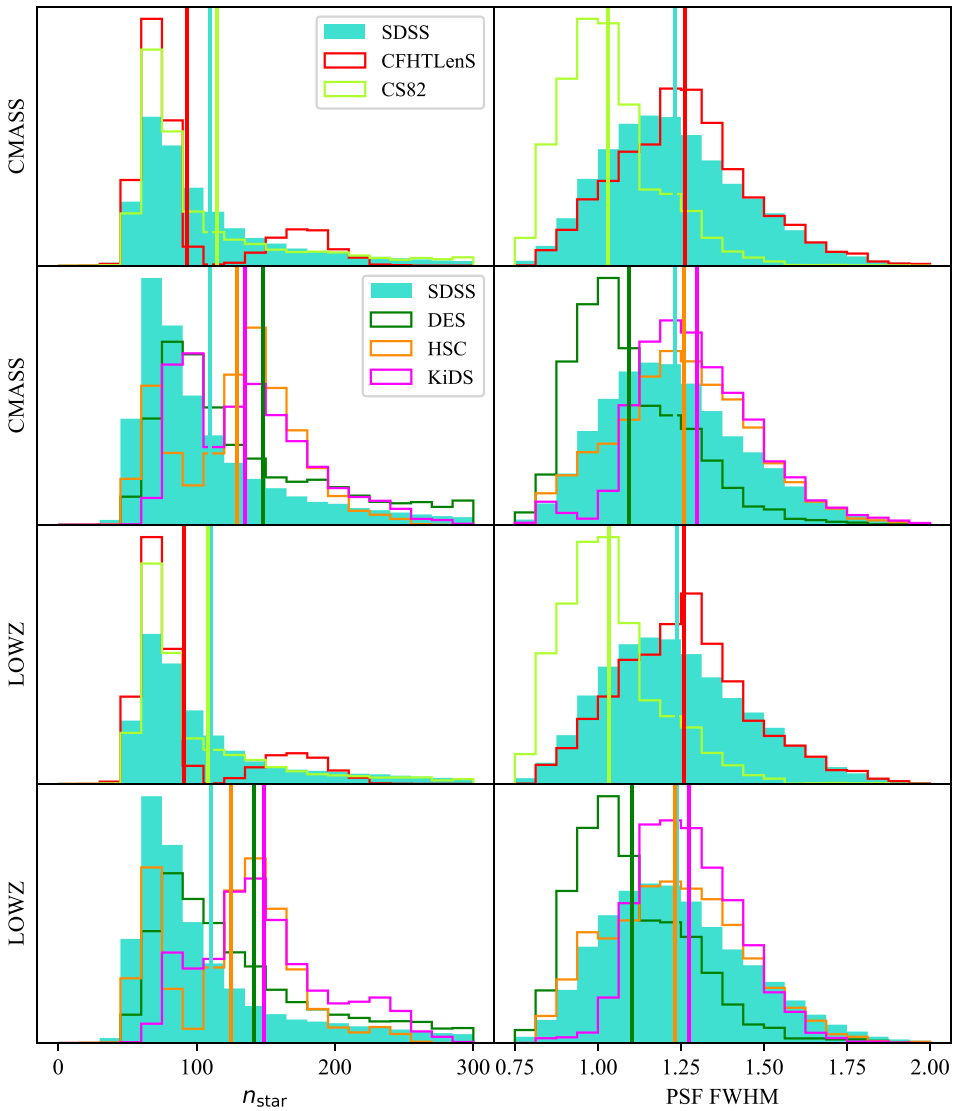

Figure 5. The distributions of stellar density (left) and *i*-band seeing FWHM (right) from the BOSS targeting catalogues are not homogeneous in the regions of overlap with lensing surveys. Vertical lines indicate the mean value for each survey.

discussed in Section 7.2 and Appendix A such differences are not a concern for the present study.

8.3 Tests related to homogeneity of lens samples

We investigate whether or not the amplitude of the lensing signals vary according to spatially varying properties of the lens samples. We show post-unblinding results for variations in the amplitudes as a function of $n_{\rm star}$, the PSF FWHM of the imaging used in BOSS targeting, as well as tests related to position on the sky (North versus South). For these tests, we include measurements both with, and without, the BOSS weights, $w_{\rm tot}$. Although BOSS did not find any trends in number density variations for LOWZ with $n_{\rm star}$ or PSF FWHM, we include test for LOWZ here for completeness. We fit the trends with a linear relation with slope β and comment on whether or not the slope, β, is consistent with zero. It was agreed before unblinding that trends greater than 3σ would be considered significant. In all figures, trends greater or equal to 3σ are highlighted in red.

8.3.1 SDSS PSF

Fig. 8 shows the amplitudes of each survey as a function of the PSF FWHM of the imaging used in BOSS targeting. No significant trends are found, suggesting that number density variations of BOSS with the SDSS PSF do not correlate with the halo mass properties of CMASS galaxies.

8.3.2 Galactic hemisphere

Fig. 9 shows the amplitudes from each survey as a function of the fraction of area in the Northern galactic cap. No significant trends are found for L1 and L2 which is consistent with the pre-blinding tests we carried out with SDSS (see Appendix D).

A trend is found for C2 (4.0σ for low-R, 1.6σ for high-R, and 3.5σ for all-R). However, there is a mild correlation between $n_{\rm star}$ and $A_{\rm North}/(A_{\rm South}+A_{\rm North})$ where $A_{\rm North}$ is the area in the North and $A_{\rm South}$ is the area in the South. A trend is also found in C2 with $n_{\rm star}$ and the $n_{\rm star}$ trend is more significant than the one in Fig. 9. For this

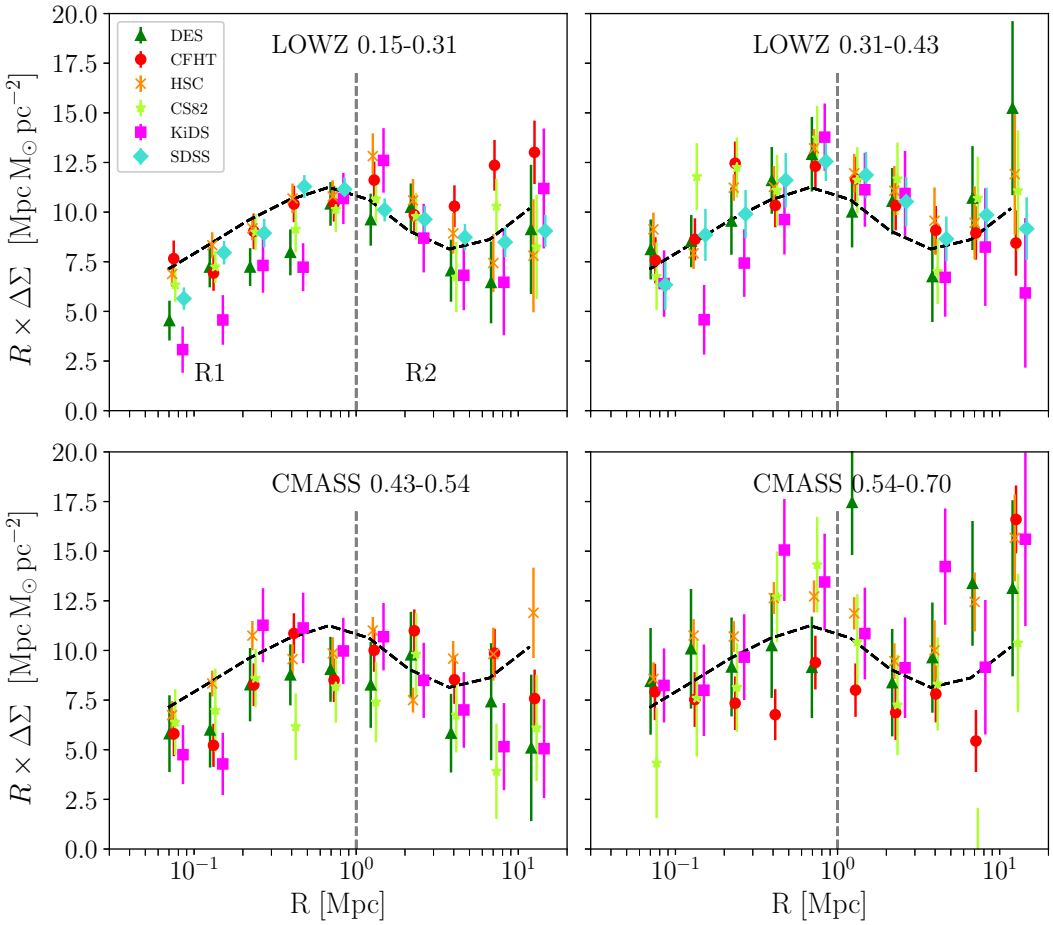

Figure 6. Galaxy–galaxy lensing signal around four LRG lens samples from six different lensing surveys with statistical uncertainties. Dashed vertical grey lines delineate our two scale cuts ($R_1 = [0.05,1]$ Mpc and $R_2 = [1,15]$ Mpc). Dashed grey lines show $\Delta\Sigma_{HOD}$ (the predicted shape of $\Delta\Sigma$ based on an HOD to CMASS clustering as described in Section 3.5). This paper does not test for amplitude shifts between lensing and clustering and the amplitude of $\Delta\Sigma_{HOD}$ has been arbitrarily normalised to match the lensing signals. Data points have been shifted slightly along the *x*-axis for visual clarity.

Table 4. Signal-to-noise ratio of our $\Delta\Sigma$ measurements.

	LOWZ 0.15–0.31			LOWZ 0.31–0.43			CMASS 0.43–0.54			CMASS 0.54–0.7		
Survey	R1	R2	All R	R1	R2	All R	R1	R2	All R	R1	R2	All R
CFHTLenS	9.98	9.78	9.88	9.34	7.64	8.49	7.17	7.60	7.38	5.96	5.95	5.95
CS82	8.56	6.38	7.47	6.70	5.30	6.00	4.26	3.15	3.71	3.83	2.74	3.29
DES	7.12	5.15	6.13	6.18	4.25	5.22	4.39	2.97	3.68	3.68	4.07	3.87
HSC	13.53	7.55	10.54	12.39	7.39	9.89	12.20	9.60	10.90	13.61	9.36	11.48
KiDS	5.25	4.54	4.90	4.80	3.64	4.21	4.72	3.64	4.18	4.55	3.63	4.10
SDSS	14.80	13.27	14.03	8.28	7.66	7.97	–	–	–	–	–	–

reason, we will consider n_{star} to be the driving trend in this redshift range.

8.3.3 Lensing amplitudes versus stellar density

Fig. 5 shows that the n_{star} distribution varies considerably between, and within, each of the lensing surveys (due to the location of each of survey on the sky in relation to the Galactic Plane). The density of stars varies between $50 < n_{star} < 300$ overall, and the mean value for each survey varies between $90 < n_{star} < 150$. The mean value of n_{star} is different for LOWZ and CMASS because they cover slightly different areas.

Fig. 10 shows the lensing amplitudes from each survey as a function of n_{star}. No trends are found for L2 and C1 but trends are found in L1 and C2. For L1, the lensing signal is found to decrease with n_{star} with a slope of $\beta \sim 0.0025$. The trend is 3.1σ for low-R, 1.2σ for high-R, and 3.4σ for all-R. For C2, the amplitude of the lensing signal is found to increase with n_{star} with a slope of $\beta \sim 0.007$. The trend is 5σ for low-R, 3.9σ for high-R, and 6σ for all-R. Applying the BOSS w_{tot} weights does not impact the lensing amplitudes and thus does not correct for this effect.

After unblinding, we decided to test the sensitivity of the trend to data points at the extremities (at low n_{star} and high n_{star}). For L1, when removing the low n_{star} data point (CFHTLenS) the trend is

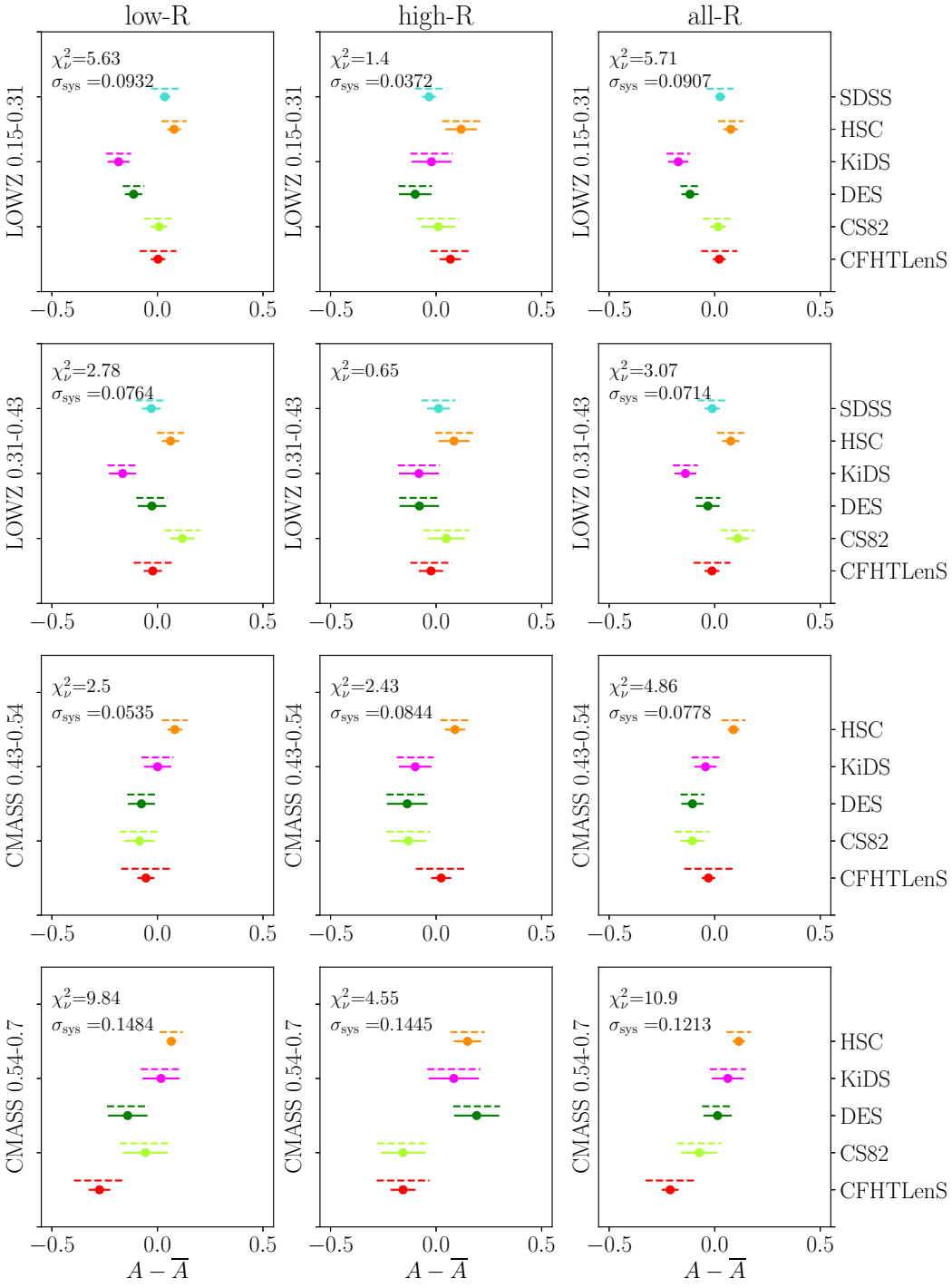

Figure 7. Amplitude fits to lensing data in three different radial ranges. Left columns show the amplitude fits for $r < 1$ Mpc. Middle columns show the results for $r > 1$ Mpc. Columns on the right correspond to the full radial range. The top rows show data for the two LOWZ samples and lower rows show the data for the two CMASS samples. We show $A - \overline{A}$ where A is the fitted amplitude and $\overline{A}$ is the mean amplitude for all surveys. Dashed lines show the sum in quadrature of the statistical and the reported systematic error. In each panel, χ^2_ν is the reduced chi square of the data points and σ_{sys} is the value of the estimated σ_{sys} that yields $\chi^2_\nu = 1$. When $\chi^2_\nu < 1$, upper limits are indicated for σ_{sys} in Fig. 14.

detected at 4σ. When removing the high n_{star} data point (KiDS) the trend is not detected. The L1 trend is therefore sensitive to the data point at high n_{star}. For C2, when removing the low n_{star} data point (CFHTLenS) the trend is not detected. When removing the high n_{star} data point (DES) we obtain positive slopes at 6σ for low-R, 3.4σ for high-R, and 6.5σ for all-R. The C2 results are therefore sensitive to the data point at low n_{star}.

For L1, given that both lenses and sources are bright, it seems difficult to imagine how n_{star} could impact the lensing amplitudes. This is, however, the redshift range where the lensing signals have

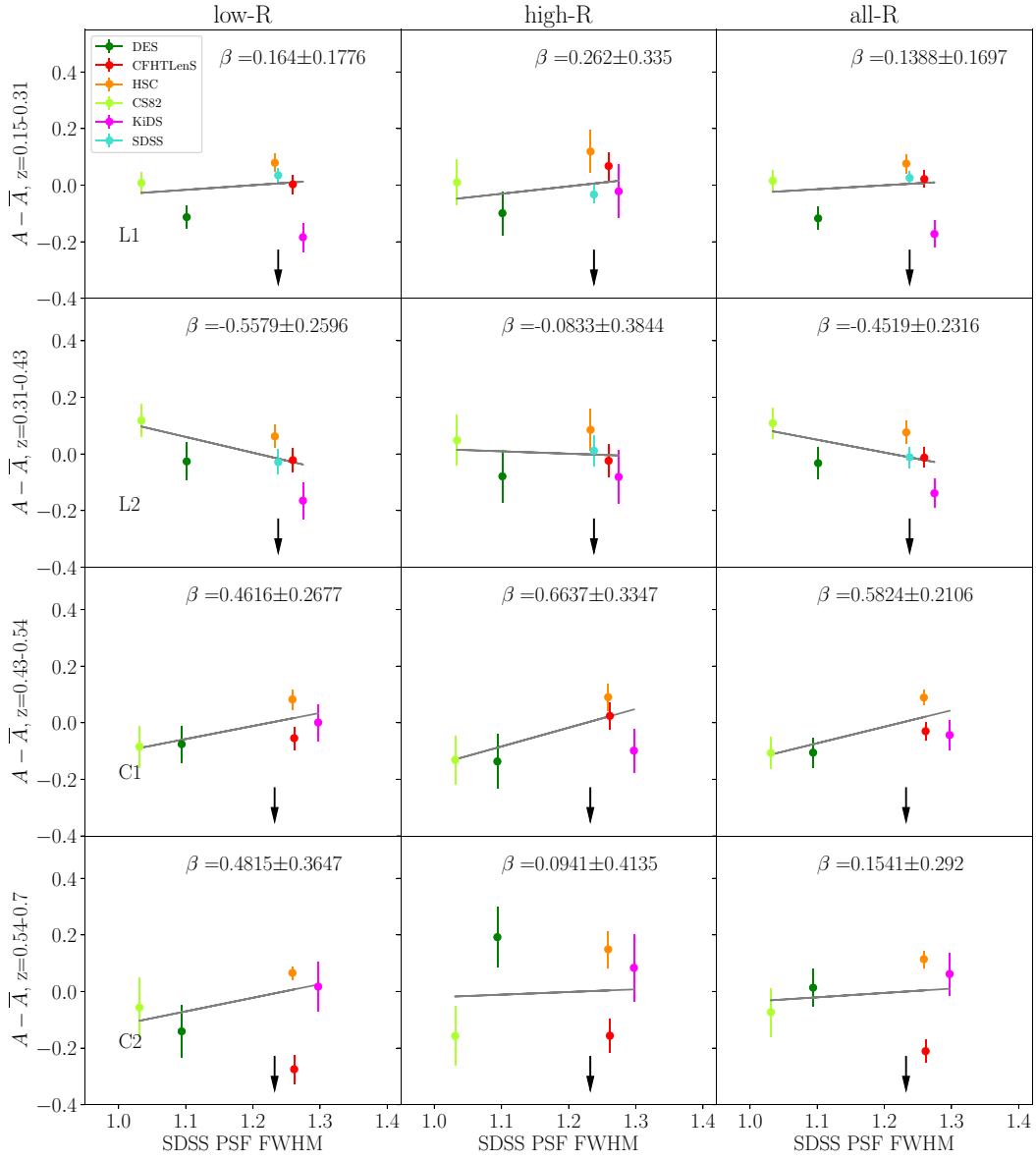

Figure 8. Amplitude of $\Delta\Sigma$ versus PSF of SDSS as given in the BOSS targeting catalogue. The two upper rows correspond to LOWZ and the two bottom rows correspond to CMASS. The vertical arrow indicates the average BOSS value for the SDSS PSF. We use the statistical errors to search for trends (Section 3.6 and Fig. 2) and Figs 8 through 13 display statistical errors. The results are fit with a linear relation with slope β. No trends are found between the amplitude of the lensing signal and the SDSS PSF.

the highest signal to noise (see Table 4). One possible explanation for the trend in L1 is that the statistical signal to noise of the lensing signals are slightly overestimated in this redshift regime. Unlike other trends found in this paper, this L1 trend disappears if we use the internal variance in the data rather than the reported statistical lensing errors. The trends drop below 3σ when the errors are inflated by 20 per cent. The statistical errors on the lensing signals could be underestimated or the cross-covariance between surveys (which we have neglected) could be playing a role. This aspect will be discussed further in Section 9.3.2. Underestimated statistical errors would result in an overestimate of the global systematic error. Since our primary goal is to rule out large systematic errors, ignoring the possibility of underestimated statistical lensing errors (or joint covariance) will lead to conservative conclusions. Finally, another possibility is that the significance of the trend is overestimated (see explanation in Fig. 2)

The C2 trend is of larger interest because the overall trend is stronger and this is also the regime where both the lenses and the source are fainter and trends with $n_{\rm star}$ are more plausible. There are four possible explanations for the C2 trend. The first explanation is that it could be an inhomogeneity in the lens sample that correlates with halo mass. This is plausible because we know from Ross et al. (2017) that the number density of CMASS galaxies varies with $n_{\rm star}$ with number density variations that are strongest in the C2 bin. The second explanation is that instead of being a variation in the halo masses of the lenses, this could be a systematic in the source sample that correlates with $n_{\rm star}$ (or another quantity that correlates with $n_{\rm star}$ such as galactic extinction) and leads to biased estimates of $\Delta\Sigma$. For example, one possibility could be errors in star-galaxy separation in

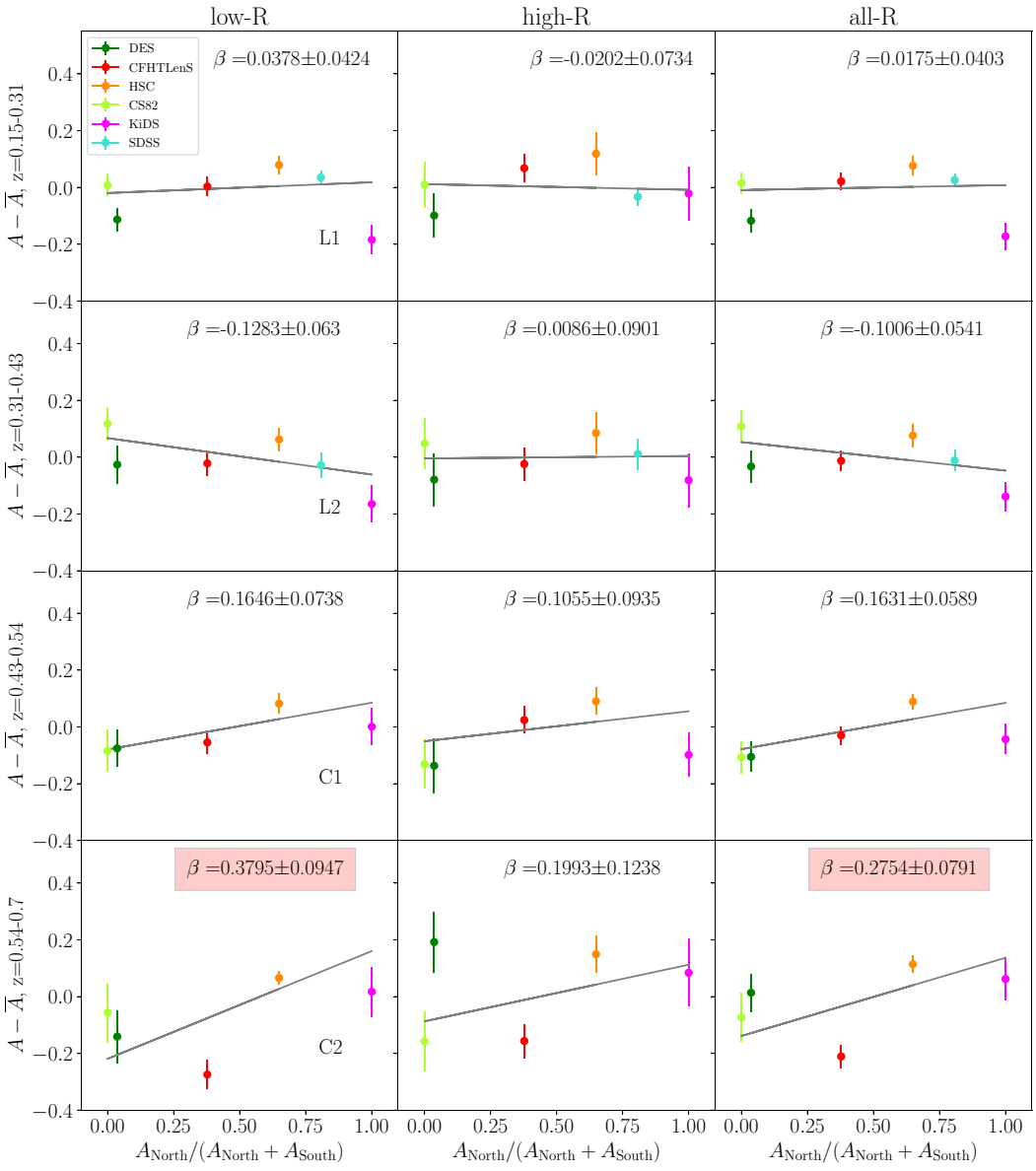

Figure 9. Amplitude of $\Delta\Sigma$ versus fraction of area in the Northern galactic cap. A trend is found in C2. However, the trend is less significant than the trend with $n_{\rm star}$ which is the driving trend in this redshift range.

source catalogues. The third possibility is the significance of the trend could be overestimated and that there is in fact no trend with $n_{\rm star}$ (see Fig. 2 for an explanation on why this might occur). The fourth possibility is that the statistical lensing errors are underestimated. However, unlike the trend found in the L1 bin, the C2 trend does not vanish when the internal variance between the data points are used rather than the reported errors. For this reason, underestimated statistical lensing errors seem unlikely to be the full story for C2.

In the first case, the variations of $\Delta\Sigma$ with $n_{\rm star}$ would not be counted as a source systematic because the variations would reflect true variations in halo mass. In the second case, the spread in signals would be interpreted as evidence for a lensing systematic that correlates with $n_{\rm star}$. In the third case, the spread would be interpreted as a source systematic, but one that does not correlate with $n_{\rm star}$.

After unblinding, we designed a test to attempt to differentiate between the first and the second explanation. Fig. 5 shows that HSC has a bimodal distribution in $n_{\rm star}$ (because of where the HSC fields are located with respect to the Milky Way). Taking advantage of this fact, together with the fact that HSC has the highest signal to noise at this redshift range, we divided the HSC field into two separate regions, one with low mean $n_{\rm star}$ and one with high mean $n_{\rm star}$. The resulting lensing amplitudes are shown in Fig. 11. When we use only HSC, interestingly, we do not find evidence for any trend with $n_{\rm star}$. This would disfavour the first explanation in favour of the other two possible explanations. However, follow-up work will be required to convincingly disentangle between the three scenarios described above because we cannot rule out with this work alone the possibility that a combination of lens inhomogeneity and source systematics simply happen to cancel out in the HSC data in Fig. 11. An interesting test to carry out would be to test the correlation separately for bright and faint stars. Further discussion on possible explanations for this trend is also presented in Appendix E.

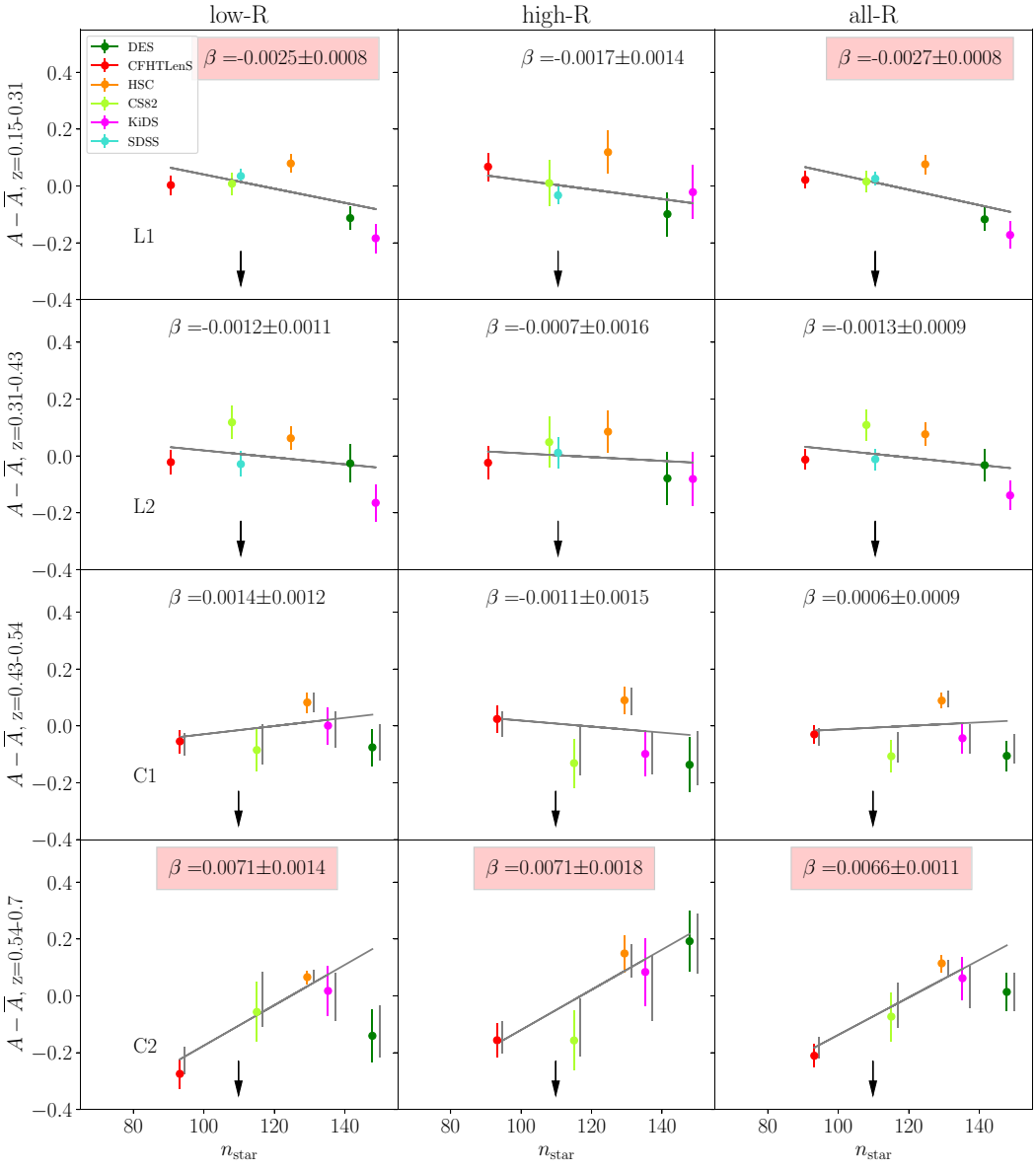

Figure 10. Amplitude of $\Delta\Sigma$ versus stellar density, $n_{\rm star}$. The vertical arrow indicates the average value for $n_{\rm star}$ for BOSS galaxies. Solid lines correspond to the amplitudes for signals computed with $w_{\rm tot}$. In the bottom panels, grey data points indicate the lensing amplitudes with no weighting applied (no $w_{\rm tot}$ weight). Grey data points are slightly offset along the x-axis for visual clarity. Applying the boss $w_{\rm tot}$ weights does not impact the amplitude of $\Delta\Sigma$. The trend is fit with a linear function with slope β. A red box indicates a value of beta detected to be non zero at greater than 3σ. No trends are found for L2 and C1, but a trend is found for L1 and C2. We interpret the L1 trend as possibly arising from underestimated lensing errors. For C2, the trend is 5σ for low-R, 3.9σ for high-R, and 6σ for all-R but the detection of this trend is sensitive to the low $n_{\rm star}$ data point.

8.4 Scaling of amplitudes with mean source redshift

Fig. 12 displays the lensing amplitudes as a function of the mean source redshift for each survey. No trends are found for $z_{\rm s}$ for L1 and for L2 but trends are found for C1 and C2. For C1, a trend with slope of $\beta \sim 0.8$ is detected at 2.9σ, 2.9σ, and 4.3σ, respectively for low-R, high-R, and all-R. For C2, a trend with similar slope values is detected at 4.2σ, 1.5σ, and 3.5σ, respectively for low-R, high-R, and all-R. We remind readers that we have used the statistical errors on $\Delta\Sigma$ to constrain this trend (see Section 3.6).

After unblinding, we decided to also test to determine how sensitive the trend is to data points at the extremities (at low $z_{\rm s}$ and high $z_{\rm s}$). For both C1 and C2, the significance of the detection of a positive slope remains similar even if we remove the low $z_{\rm s}$ data point (CS82). However, the trend is no longer detected in either C1 or C2 after the removal of the high $z_{\rm s}$ data point (HSC). We conclude that the correlation between the amplitude of the lensing signals is sensitive to the high $z_{\rm s}$ data point (HSC).

The absence of redshift trends in L1 and L2 is not inconsistent with the results for C1 and C2. The impact of effects such as blending, photo-z calibration, and shear calibration will become more apparent with fainter source galaxies and when the sources are more closely located behind the lens galaxies. The source galaxies that dominate the measurements for L1 and L2 will be brighter galaxies at lower redshift than for C1 and C2. The absence of a significant trend in

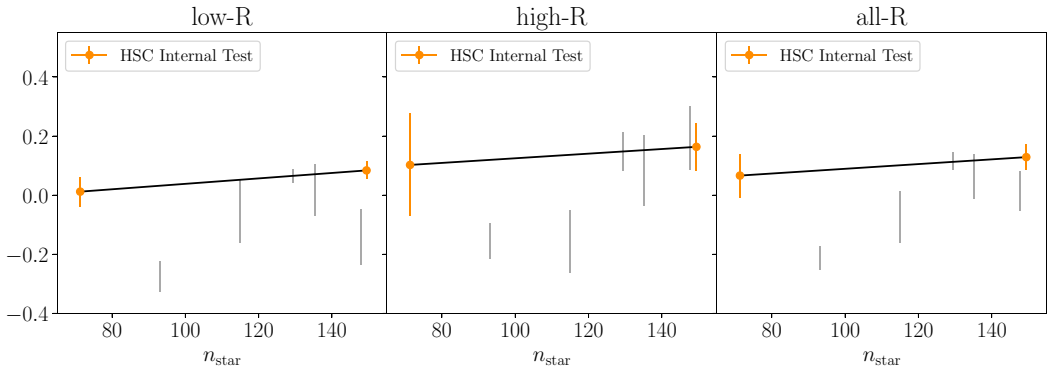

Figure 11. A test to determine if the correlation between the amplitudes of the lensing signals with $n_{\rm star}$ is due to lens inhomogeneity (and hence a genuine variation in halo mass) or due to a lensing source systematic. Grey data points show the amplitudes of the given surveys (same format as in Fig. 10). Left is low-R, middle is high-R, and right is all-R. The two orange data points show the lensing signal as measured internally using just HSC and dividing the survey into two separate areas with low and high $n_{\rm star}$. The internal trend with HSC is flat. This suggests that the trend observed in Fig. 10 is not an intrinsic variation in the halo masses of CMASS across the sky.

L1 and L2 points to effects that predominantly affect the fainter and higher redshift samples.

The C1 bin is high redshift for lensing with SDSS and so initially we had decided to not compute lensing with SDSS in this bin. However, after unblinding, we searched for an additional way to test the $z_{\rm s}$ trend in C1. We decided to compute the lensing amplitude with SDSS in the C1 bin. We have shown the SDSS data point in cyan in Fig. 12 but we do not use this SDSS data point in any of our fits.

HSC has the largest dynamic range in source redshifts and so it is interesting to consider testing for source redshift dependent trends by dividing the source sample. In fact this exact test was already carried out in Speagle et al. (2019) who performed tests with HSC that divided the source sample into a high redshift ($z_{\rm s} > 1$) and a low redshift ($z_{\rm s} < 1$) subset (section 7.5 in Speagle et al. 2019). They used the same CMASS and LOW galaxies as here, but in redshift bins of $0.2 < z < 0.4$ and $0.4 < z < 0.6$ and $0.6 < z < 0.8$. No significant trends (more than 3σ) were found. In the lens redshift bin $0.4 < z < 0.6$, which corresponds roughly to C1 and C2 here, there was a 2σ detection of the opposite trend than reported here (Speagle et al. 2019 found the signals from the low source redshift bin to be systematically higher than those from higher redshift). We conclude that this test is hard to carry out in a statistically significant manner yet with any single survey, underscoring the utility of harnessing the power of multiple surveys.

We use a Monte Carlo test to asses the likelihood of these trends arising both from our use of statistical errors in the fits and from the look elsewhere effect (e.g. Lyons 2008)[25]. We create a series of Monte Carlo tests in which there is no amplitude variation with redshift ($\beta = 0$). We perturb the amplitudes according to the sum in quadrature of each surveys statistical and systematic error. We fit the data points using the statistical errors only. We asses how often one trend will occur in the four low-R tests, and how often two trends will occur in the four all-R bins. We find a 1 per cent probability of having one trend occur for the low-R tests. For the high-R tests, the expected probability for two trends is much less than 1 per cent.

[25]The phenomenon where an apparently statistically significant observation may arise by chance because of the large size of the parameter space that is searched.

Prat et al. (2018) also perform a similar shear ratio test for DES Y1 (see fig. 12 in Prat et al. 2018). However, they did not use lens/source combinations whose redshift would have corresponded to a shear ratio test for C1 or C2 and so would not have seen the effects described here.

Giblin et al. (2021) performed the most recent shear ratio test for KiDS using CMASS galaxies (see their Fig. 11). Although the methods are not directly comparable because we use $\Delta\Sigma$ and Giblin et al. (2021) use $\gamma_{\rm t}$, we can compare to the redshift evolution of the KiDS galaxy–galaxy lensing signal for the C2 bin. Giblin et al. (2021) find a fully consistent galaxy–galaxy lensing signal for sources at a range of different photometric redshifts behind the BOSS lenses. They conclude, however, that this test is fairly insensitive to redshift and shear calibration errors once uncertainty in the amplitude of the intrinsic alignment of galaxies is included in the analysis (see Section 9.1).

Finally, after unblinding, we also decided to assess whether or not the $z_{\rm s}$ trend might arise from the different methodologies used to compute $\Delta\Sigma$. Fig. 13 shows the correlation between the lensing amplitudes and the $z_{\rm s}$ trend for those bins in which a trend was detected at more than 3σ. Data points are colour-coded according to different methodologies as detailed in Table 2. Fig. 13 does not reveal any obvious relationship between the lensing amplitudes and the methodology employed. This suggests instead that the trend is more likely to be caused by some intrinsic bias in the source sample that correlates with survey depth.

8.5 Trends including systematic errors

We have also investigated these trends using the sum in quadrature of the statistical and the reported systematic errors. In this case, all trends except the L1 all-R trend in Fig. 10 drop below 3σ. The L1 all-R trend drops from 3.4σ to 3.1σ (just above our predefined limit for claiming a detection). This is consistent with our interpretation that: (a) the L1 trend in Fig. 10 may result from underestimated statistical errors and from neglecting the covariance between the measurements (see Section 8.3.3), and (b) that the trends highlighted in the previous sections are globally consistent with the reported systematic error budgets (see also next Section). The trends could originate from the 'known knowns' types of systematic errors accounted for by existing error budgets.

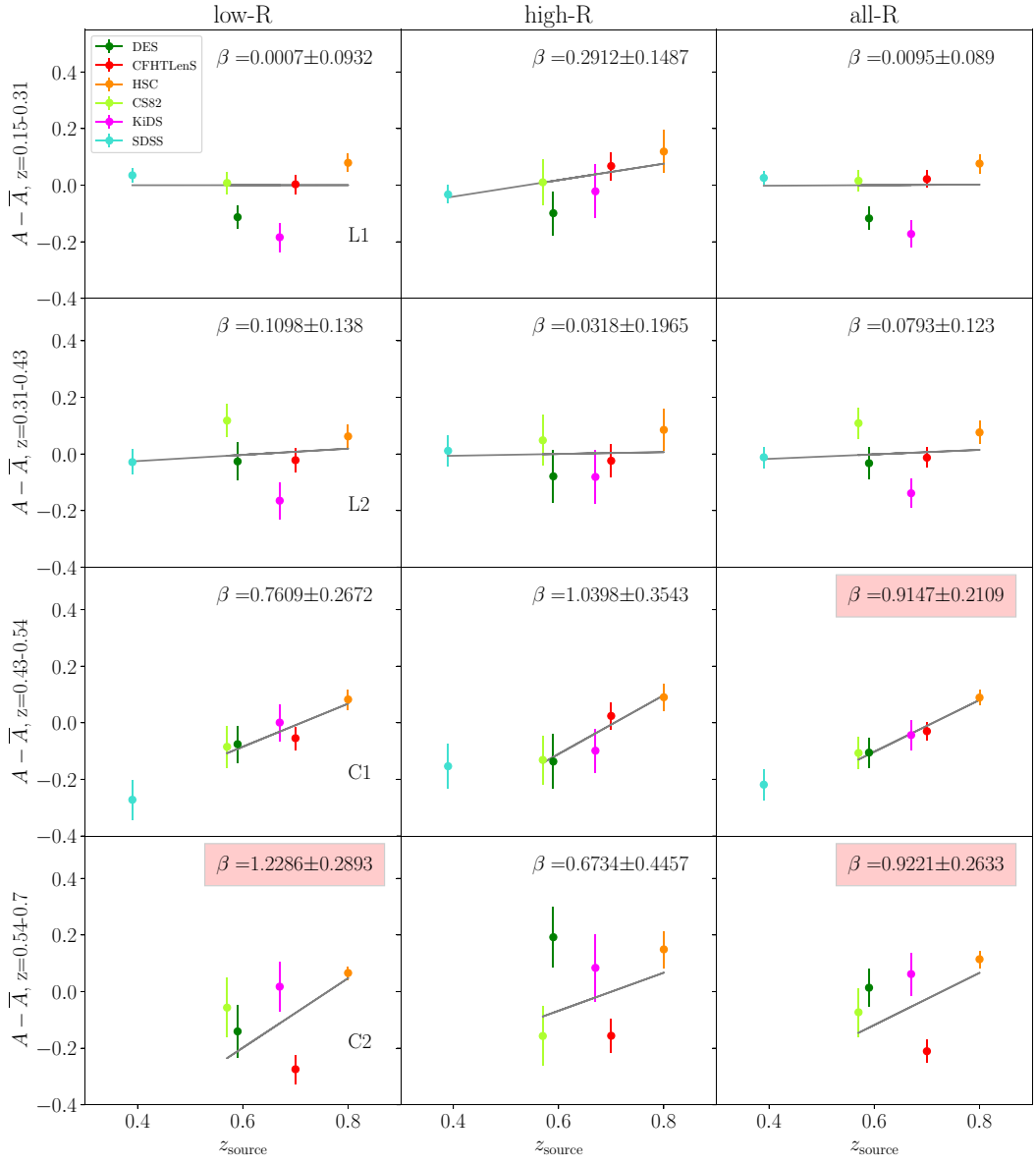

Figure 12. Amplitude of $\Delta\Sigma$ versus mean redshift of source galaxies (i.e. depth of lensing surveys). We find a correlation between the amplitude of the lensing signal and the mean source redshift of the lensing surveys for C1 and C2. The amplitude shift between the two extremities (CS82 and HSC) in the C1 bin is of order ∼25 per cent but is consistent with the systematic errors reported by each of the lensing surveys if we assume that the $z_{\rm s}$ trend dominates the error budget and that some surveys have experienced a 1.5σ upward or downward systematic shift (see Table 5 and discussion in Section 9.1). After unblinding, we decided to add a SDSS data point to the C1 bin. We show the SDSS result here but do not include it in the fits. The C2 bin is the noisiest because shallow lensing surveys (e.g. DES and CS82) have a limited number of source galaxies above the C2 lens redshift range. A tentative trend with $n_{\rm star}$ also adds scatter to the C2 bin.

8.6 Empirical estimates of the systematic error

We now present empirical systematic error estimates for the ensemble of lensing data following the methodology outined in Section 3.7. We will assume a single systematic error for all data and per test (one test corresponds to one panel in Fig. 7).

Table 5 and Figs 14 and 15 compare the values of the systematic errors reported by each of the surveys to our empirical estimates. As described in Section 3.7, Monte Carlo tests were used to show that the $\sigma_{\rm sys}$ value that we estimate is close to the mean systematic error among surveys. For all redshift and radial ranges, we find good agreement between our empirically estimated values and the reported systematic errors. Differences between the estimated and reported systematic errors are always less than 3σ (our pre-blinding determined criterion).

For L1, L2, and C1 (lenses at $z < 0.54$) we find excellent agreement between our empirical estimates and those reported by lensing surveys. No evidence is found for large unknown systematic errors. Using the values for $\sigma_{\rm sys}$ derived over the full radial range (all-R), we find estimated values of $\sigma_{\rm sys} = 0.09\pm^{0.06}_{0.02}$, $\sigma_{\rm sys} = 0.07\pm^{0.05}_{0.02}$, and $\sigma_{\rm sys} = 0.08\pm^{0.07}_{0.02}$ for L1, L2, and C1 respectively. These are in good agreement with the mean reported systematic errors which are 0.05, 0.05, and 0.056 for L1, L2, and C1 respectively.

For the C2 bin (lenses between $0.54 < z_{\rm L} < 0.7$ and sources at $z_{\rm s} \gtrapprox 0.7$) the estimated value is $\sigma_{\rm sys} = 0.12\pm^{0.09}_{0.03}$ which is higher than values found for L1, L2, and C1. However, the difference with the

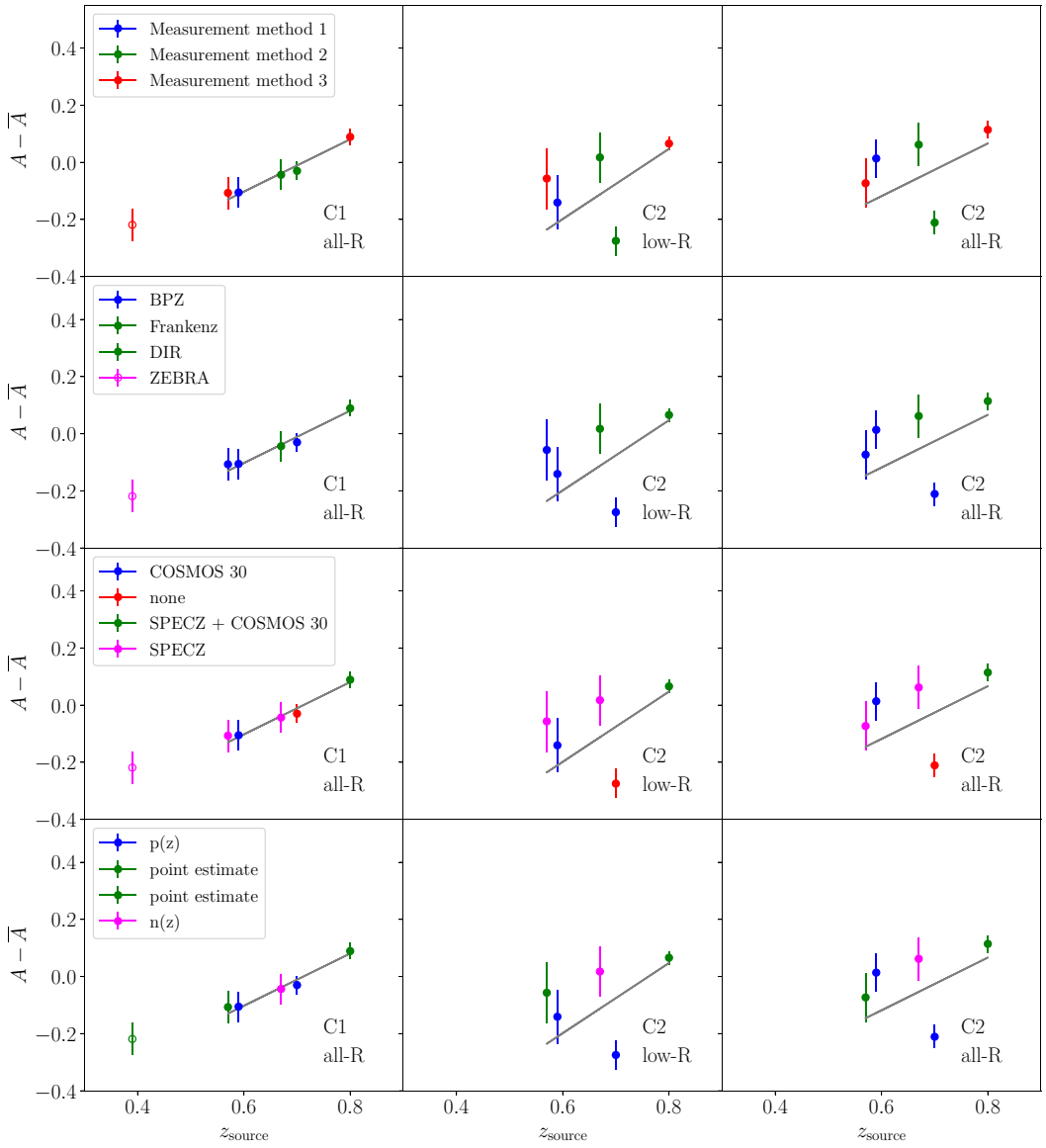

Figure 13. Amplitude of $\Delta\Sigma$ versus mean redshift of source galaxies for bins in which a trend greater than 3σ was detected in Fig. 12. Data points are colour coded using different choices for the computation of $\Delta\Sigma$ as summarized in Table 2. In the first row, data points are colour coded according to overall similarity between methods (see Section 6). In the second row, data points are colour coded according to the photo-z computation method. In the third row, data points are colour coded according to the calibration samples used to ensure unbiased photo-z's. In the fourth row, data points are colour coded according to the redshift adopted to compute Σ_c. We do not find evidence that any of these methodology or photo-z choices determine the trend. The SDSS data point was added after unblinding for C1, is shown using an open symbol, and was not used in the fit.

mean reported systematic error ($\overline{\sigma}_{\rm sys} = 0.058$) only has a significance of 2σ which does not meet out pre-defined 3σ threshold for claiming a detection. The interpretation of this redshift range is also rendered difficult given the possibility of a trend with $n_{\rm star}$ that could explain a large fraction of this spread. In Section 8.3.3, we discussed three possible explanations for the origin of the $n_{\rm star}$ trend. This trend could be due to intrinsic halo mass variations of the CMASS sample across the sky. If this is the case, it would be appropriate to subtract out this effect when reporting $\sigma_{\rm sys}$. If on the other hand, the variations with $n_{\rm star}$ are not related to intrinsic halo mass variations, the spread among data points should count as a source systematic. Since Fig. 11 seems to favour the second explanation, we quote raw values (without subtracting out the $n_{\rm star}$ trend). However, it is clear that more detailed follow-up work will be required to fully resolve the question of the $n_{\rm star}$ trend. Further discussion is presented in Appendix E.

For the C1 bin (lenses between $0.43 < z_{\rm L} < 0.54$ and sources at $z_{\rm s} \gtrsim 0.54$) we find that the amplitude of $\Delta\Sigma$ correlates with $z_{\rm s}$ with variations between surveys reaching a maximum difference of 23 per cent (between HSC and CS82). These variations are consistent with the reported systematic errors in this redshift range but only if we also assume that the $z_{\rm s}$ trend dominates the systematic error budget. The reported systematic errors for C1 are between 2 per cent and 11 per cent and the estimated values are between 5 per cent and 8 per cent (see $\sigma_{\rm sys}$ for C1 in Table 5). Assuming an ensemble value of $\sigma_{\rm sys} = 0.073$, Fig. 12 can be understood by assuming that surveys have experienced a 1.5σ upwards or downwards shift depending on the value of $z_{\rm s}$. This would imply that the correlation with $z_{\rm s}$ is the

Table 5. Comparison of reported and estimated systematic errors. The systematic that we consider is a multiplicative systematic on the amplitude of $\Delta\Sigma$ (see equation 8). We assume that this systematic error is Gaussian with width $\sigma_{\rm sys}$ and can take on a different value for each survey. Reported errors on the estimate of $\sigma_{\rm sys}$ are 68 per cent confidence. We also quote the 68 per cent and 95 per cent upper confidence limits on $\sigma_{\rm sys}$. Our global estimate of $\sigma_{\rm sys}$ correspond roughly to the mean systematic error among surveys. For C1 and C2, CFHTLenS reports the largest systematic error. For this reason, combined with the fact that the $n_{\rm star}$ trend is also found to depend on this data point, we also report the values of $\sigma_{\rm sys}$ without CFHTLenS. This does not have a large impact on C1 but it does reduce the difference seen in C2 and brings the estimated values into closer agreement with the mean reported value. For both C1 and C2, the 68 per cent and 95 per cent confidence limits are not as constraining as for L1 and L2 because of the higher lens redshift range and because five surveys are used to constrain $\sigma_{\rm sys}$ rather than six.

	LOWZ 0.15-0.31			LOWZ 0.31-0.43		
Survey	R1	R2	All R	R1	R2	All R
CFHTLenS	0.08	0.08	0.08	0.08	0.08	0.08
CS82	0.06	0.06	0.06	0.06	0.06	0.06
DES	0.03	0.02	0.02	0.03	0.02	0.02
HSC	0.05	0.05	0.05	0.05	0.05	0.05
KiDS	0.03	0.03	0.03	0.03	0.03	0.03
SDSS	0.06	0.06	0.06	0.06	0.06	0.06
Mean reported $\sigma_{\rm sys}$	0.052	0.050	0.050	0.052	0.050	0.050
			Ensemble Systematic Error			
Estimated	$0.09\pm^{0.06}_{0.02}$	$0.04\pm^{0.06}_{0.04}$	$0.09\pm^{0.06}_{0.02}$	$0.08\pm^{0.06}_{0.03}$	<0.06	$0.07\pm^{0.05}_{0.02}$
1σ upper confidence limit	0.16	0.1	0.15	0.14	0.06	0.13
2σ upper confidence limit	0.25	0.18	0.24	0.23	0.14	0.2
	CMASS 0.43-0.54			CMASS 0.54-0.7		
Survey	R1	R2	All R	R1	R2	All R
CFHTLenS	0.11	0.11	0.11	0.11	0.11	0.11
CS82	0.06	0.06	0.06	0.06	0.06	0.06
DES	0.02	0.02	0.02	0.03	0.03	0.03
HSC	0.05	0.05	0.05	0.05	0.05	0.05
KiDS	0.04	0.04	0.04	0.04	0.04	0.04
Mean reported $\sigma_{\rm sys}$	0.056	0.056	0.056	0.058	0.058	0.058
			Ensemble Systematic Error			
Estimated	$0.05\pm^{0.06}_{0.02}$	$0.08\pm^{0.09}_{0.04}$	$0.08\pm^{0.07}_{0.02}$	$0.15\pm^{0.11}_{0.04}$	$0.14\pm^{0.12}_{0.04}$	$0.12\pm^{0.09}_{0.03}$
1σ upper confidence limit	0.11	0.18	0.14	0.26	0.27	0.21
2σ upper confidence limit	0.2	0.32	0.26	0.46	0.48	0.37
			C1 and C2 without CFHTLenS			
Mean reported $\sigma_{\rm sys}$	0.043	0.043	0.043	0.045	0.045	0.045
Estimated	$0.07\pm^{0.09}_{0.03}$	$0.1\pm^{0.12}_{0.04}$	$0.1\pm^{0.1}_{0.03}$	$0.08\pm^{0.14}_{0.06}$	$0.12\pm^{0.16}_{0.06}$	$0.06\pm^{0.11}_{0.05}$
1σ upper confidence limit	0.16	0.22	0.2	0.2	0.28	0.17
2σ upper confidence limit	0.33	0.46	0.41	0.46	0.598	0.36

dominant term in the systematic error budget in this redshift range. In other terms, $\sigma_{\rm sys}$ and $z_{\rm s}$ are strongly correlated. Possible explanations for the origin of the $z_{\rm s}$ are discussed in Section 9.

For C1 and C2, CFHTLenS has the largest reported systematic errors (11 per cent). This fact, combined with the observation that the $n_{\rm star}$ trend is sensitive to the removal of this data point, led us to decide post-blinding to also quote values for C1 and C2 without the CFHTLenS data point. The removal of CFHTLenS has little impact on the results for for C1 but does reduce the estimated values of $\sigma_{\rm sys}$ for C2 and $\sigma_{\rm sys}$ drops from $\sigma_{\rm sys} = 0.12\pm^{0.09}_{0.03}$ to $\sigma_{\rm sys} = 0.06\pm^{0.11}_{0.05}$. The 68 and 95 per cent upper confidence levels are, however, relatively unchanged. Overall C2 remains less constraining than the other redshift bins because the lensing data are noisier in this redshift range.

In the context of the 'lensing is low' phenomenon in which models of the galaxy-halo connection applied to the clustering of CMASS and LOWZ over-predict $\Delta\Sigma$ by 20–30 per cent (Leauthaud et al. 2017; Lange et al. 2019, 2021; Singh et al. 2020), it is interesting to consider to what degree large unknown systematic errors are ruled out by these results. Overall, systematic errors greater than 15 per cent, 13 per cent, and 14 per cent are ruled out at 68 per cent confidence level for L1, L2, and C1 respectively. Systematic errors greater than 24 per cent, 20 per cent, and 26 per cent are ruled out at 95 per cent confidence level for L1, L2, and C1 respectively. These constraints could be made tighter by combining between lens bins – but we have not attempted this here. In summary, a 25 per cent systematic error is ruled out at 95 per cent confidence level in three fairly independent redshift bins. A more detailed discussion on this topic is presented in Section 9.

It is also interesting in Table 5 and Fig. 14 to consider whether or not there is a radial dependence to $\sigma_{\rm sys}$. When considering all lens samples and comparing the 'low-R' results and the 'high-R' results, we do not see evidence for a radial dependence in $\sigma_{\rm sys}$. The systematic that leads to the $z_{\rm s}$ trend identified in Fig. 12 (which was found to be radially independent) is likely to be a more dominant systematic than boost factor corrections.

The systematic errors estimates presented in this paper are conservative in the sense that we have included data from earlier lensing surveys. Lensing methods have been evolving rapidly and the most recent lensing surveys probably have errors that are smaller than the

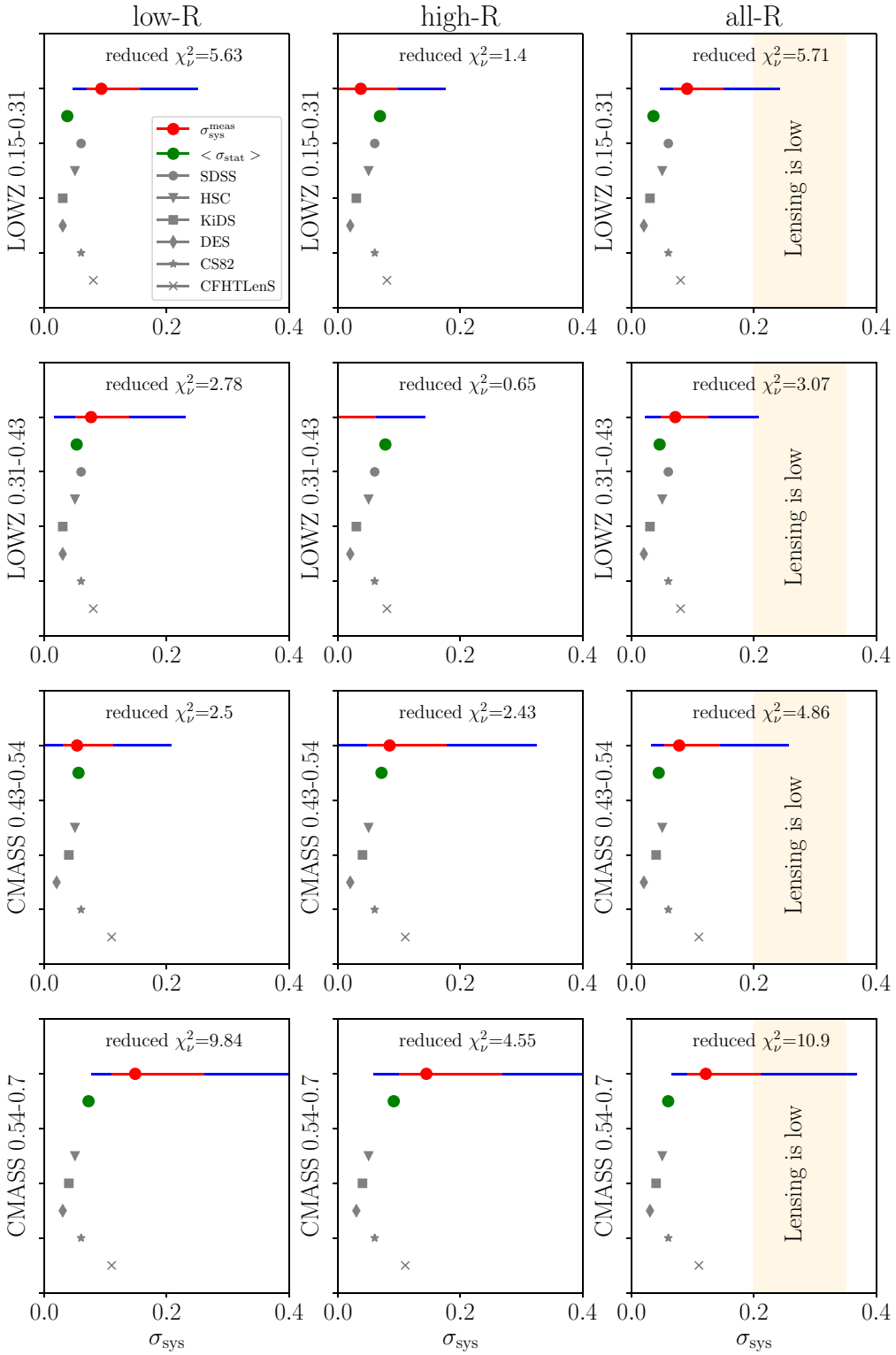

Figure 14. Comparison between reported systematic errors and our empirically estimated values. Red lines correspond to 68 per cent confidence errors on $\sigma_{\rm sys}$ and blue lines to 95 per cent confidence. For L1, L2, and C1 we find good agreement between our empirical estimates and those reported by lensing surveys. Generally speaking, systematic errors greater than 15 per cent are ruled out at 68 per cent confidence. At $z_{\rm L} < 0.54$, systematic errors greater than 30 per cent are ruled out in most bins at 95 per cent confidence. For C2 (lenses between $0.54 < z_{\rm L} < 0.7$ and sources at $z_{\rm s} \gtrsim 0.7$) $\sigma_{\rm sys}$ is larger than reported values. A large fraction of this scatter could be explained by a correlation between the lensing amplitudes and $n_{\rm star}$. The C2 bin is also the noisiest because there are fewer sources galaxies with which to perform the measurements in this lens range and the upper limits on $\sigma_{\rm sys}$ are therefore not very constraining. The 'lensing is low effect' corresponds to a 20 per cent to 35 per cent shift in amplitudes and is highlighted by the orange shaded region. We do not find strong evidence for boost factors impacting $\sigma_{\rm sys}$ – the signature of this would be larger $\sigma_{\rm sys}$ values at low-R. Rather, $\sigma_{\rm sys}$ is found to be fairly independent of radial scales (albeit with large uncertainties on $\sigma_{\rm sys}$). The green data point shows the mean statistical error across all surveys. Galaxy–galaxy lensing measurements are entering an era in which systematic errors will dominate error budgets.

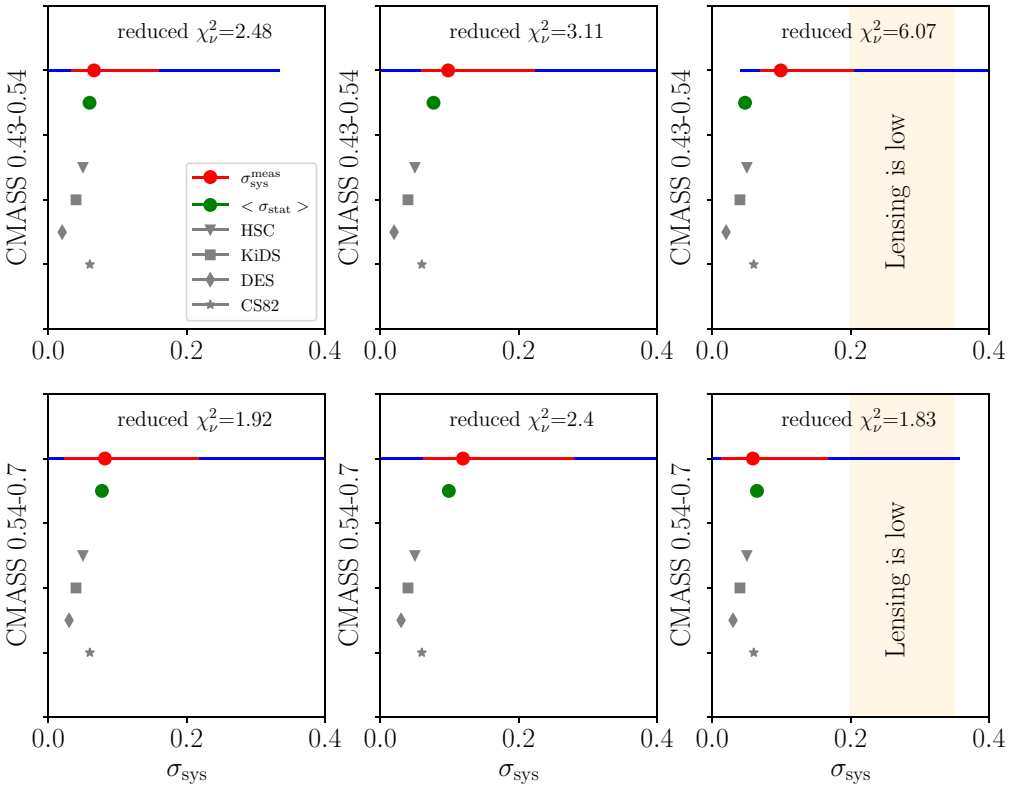

Figure 15. Same as Fig. 14 for C1 and C2 but without the CFHTLenS data points (which have the largest reported systematic errors in this redshift range). There is some evidence that removing the CFHTLenS data points reduces the variance in the C2 bin with χ^2_ν dropping from 9.84 to 1.92 for low-R, from 4.55 to 2.4 for high-R, and from 10.9 to 1.83 for all-R. However, the 68 and 95 per cent upper limits are similar to those reported in Fig. 14 and are less constraining than for L1 and L2. The 'lensing is low effect' corresponds to a 20–35 per cent shift in amplitudes and is highlighted by the orange shaded region. When the CFHTLenS data points are removed, the upper limits on $\sigma_{\rm sys}$ for C1 and C2 are not constraining enough to rule out 20 per cent amplitude shifts at more than 1σ because only four surveys are used to constrain $\sigma_{\rm sys}$ instead of 5.

ensemble values reported here. Again, one of our goals in the paper is to rule out large unknown sources of systematic error. For this, our conservative ensemble estimate is sufficient.

These results are the first empirically derived tests on the systematic errors on $\Delta\Sigma$ and present an important sanity check on the validity of reported $\Delta\Sigma$ lensing signals. However, the uncertainties on the reported $\sigma_{\rm sys}$ are still large compared to the values of $\sigma_{\rm sys}$ reported by the lensing surveys. In Section 9, we discuss how $\sigma_{\rm sys}$ could be better constrained in future implementations of these tests.

8.7 Comparison with statistical error

Finally, we present a comparison between the current statistical errors on $\Delta\Sigma$ and the systematic errors. Green data points in Fig. 14 indicate the mean statistical error on $\Delta\Sigma$ (averaged over all lensing surveys). When considering the full radial range, the mean statistical error is 3.6 per cent, 4.6 per cent, 4.4 per cent, and 6 per cent respectively for L1, L2, C1, and C2. Individual surveys have different constraining power with signal-to-noise ratios given in Table 4. Generally speaking, as can be seen by comparing the values of $\overline{\sigma}_{\rm sys}$ with the systematic errors in Fig. 14, galaxy–galaxy lensing measurements are entering an era in which systematic errors will dominate error budgets and identifying the origin of the $z_{\rm s}$ trend found in Fig. 12 is a top priority.

9 DISCUSSION

We discuss variations in $\Delta\Sigma$ with $z_{\rm s}$, outline a number of considerations for future implementations of these tests, discuss the connection with cosmic shear, and implications for the 'lensing is low' effect.

9.1 Amplitude of $\Delta\Sigma$ versus source redshift

Fig. 12 finds the amplitude of $\Delta\Sigma$ to correlate with the mean source redshift of the lensing survey for the C1 and C2 bins when considering statistical errors (see Section 3.6). The trend is such that the amplitude shift between the two extremities (CS82 and HSC) is of order ∼23 per cent and is independent of radial scales. While we see a trend that is detected with high significance, the spread caused by this trend (∼23 per cent between CS82 and HSC) is consistent with the systematic errors reported by each of the lensing surveys (see test in Section 8.5) if we assume that the $z_{\rm s}$ trend is the leading term in the systematic error budget and some surveys have experienced a 1.5σ upwards/downwards shift. This effect could be attributed to photometric redshift calibration or other effects that correlate with the depth of lensing surveys. We now outline several plausible explanations for the origin of this trend. We have ranked this list from top to bottom beginning with what we consider to be the most plausible explanation and ending with the least likely.

(i) The impact of unrecognized (and uncorrected) blends is a possible important source of systematic error. The impact of unrecognized blends could correlate with the depth of lensing surveys. The trends discussed here could therefore be a manifestation of errors in the multiplicative shear calibration bias or redshift distributions due to blending. The lensing data used here include varying levels of sophistication to account for galaxy blends, and the interplay between galaxy blends and redshift distributions. In DES, MacCrann et al. (2022) find that blending has the largest impact on high redshift source samples and is one of the dominant systematic errors for DES Y3. It is therefore quite plausible that the trend we find here is caused by unaccounted for unrecognied blends. It is not clear how unrecognized blends would impact different surveys. On the one hand, surveys with higher mean source redshifts (e.g. HSC) could be more impacted by blends. On the other hand, HSC also has the best seeing and should, in principle, have a better ability to disentangle galaxy blends. In the future, it will be interesting to consider a joint analysis of both z_s and survey specific PSF to see which has the dominant impact.

(ii) Both HSC and DES rely on the COSMOS 30-band catalogue (Laigle et al. 2016) in calibrating their photometric redshifts. It is possible that there are biases in the COSMOS 30-band catalogue (Hildebrandt et al. 2020; Joudaki et al. 2020; Myles et al. 2021). Similarly, CFHTLenS redshifts have been found to be biased (Choi et al. 2016). HSC, CS82, and KiDS also use similar spectroscopic galaxy catalogues (e.g. DEEP2, VVDS, Primus, etc.) to calibrate redshifts and/or to perform $f_{\rm bias}$ corrections. Unidentified selection effects in these spectroscopic catalogues that cannot be corrected for by re-weighting schemes could also be at play (Masters et al. 2015; Gruen & Brimioulle 2017; Wright et al. 2020). However, Fig. 13 did not reveal any pattern suggesting that photo-z calibration choices determine the trend.

(iii) With imperfect photometric redshifts, gg-lensing can be impacted by the so-called 'gI' intrinsic alignments (this is the correlation between the intrinsic shape of a source galaxy (I) and the position of the lens galaxy (g), see Joachimi et al. 2015 and references therein). Intrinsic alignments could explain the z_s trend because deeper surveys (where source galaxies are on average further behind lens galaxies and therefore are less likely to be physically associated with the lens) will have reduced contamination from gI intrinsic alignment. However, Blazek et al. (2012) used BOSS galaxies to quantify the level of gI contamination for gg-lensing and concluded that with stringent photo-z cuts, the expected gI contamination was at the 1–2 per cent per cent level which would be too low to explain the trend observed here.

(iv) The magnification of galaxies can impact the amplitude of $\Delta\Sigma$. In particular, magnification due to the lens sample can have an important effect (Unruh et al. 2020; von Wietersheim-Kramsta et al. 2021). This effect can be calculated analytically given the slope of the galaxy luminosity function at the cut that determines the lens selection (for example, see equation 13 in Unruh et al. 2020). For CMASS, the $i_{\rm cmod}$ cut limits the CMASS sample in this redshift range. Appendix F shows that CMASS has a slope, α, that is steeper than 1. Because $\alpha > 1$, lens magnification would induce a positive bias in the average tangential shear γ_t that does not strongly depend on z_s. Because the critical surface density for a given lens sample is anti-correlated with z_s, the lens magnification impact on $\Delta\Sigma = \gamma_t\Sigma_{\rm crit}$ should also decrease with z_s. Since the net effect of lens magnification is positive, this means the measured $\Delta\Sigma$ (without correction for lens magnification) should decrease with z_s. This is the opposite of what we find.

(v) Amplitude offsets could be caused by dilution effects and should be investigated further. However, when estimates of the dilution factor are made using spectroscopic calibration samples, they are generally found to be relatively small (a few per cent).

The assumption of an incorrect cosmology does bias measurements of $\Delta\Sigma$, but this bias is only a very weak function of z_s and we have checked that this cannot explain the trends seen in Fig. 12. Boost factor corrections would cause a scale-dependent trend, unlike the one found.

Overall, unrecognized blends and redshift calibration seem to be the most plausible explanations. These effects represent two of the biggest challenges for weak-lensing analyses. The trend identified in Fig. 12 warrants further investigation as a leading term in the systematic error budget for galaxy–galaxy lensing measurements for $z_s > 0.43$.

9.2 On the possibility of underestimated statistical errors

Some of the trends found in this paper could be the result of underestimated statistical errors. This is a likely explanation for the correlation between the lensing amplitude and $n_{\rm star}$ for the L1 bin. Indeed, we have neglected the covariance between lensing surveys (this is discussed further in Section 9.3.2). An underestimation of the lensing statistical errors would result in an overestimation of $\sigma_{\rm sys}$. Since the main goal of this paper is to place upper limits on σ_{sys}, underestimated lensing errors would lead to conservative conclusions.

9.3 Considerations for future implementations of these tests

The types of empirically motivated tests presented here will become more precise as lensing surveys expand and the overlap between surveys grows. Furthermore, DESI will both provide a larger number of galaxies to use for lenses and DESI lens samples will also extend in redshift to $z > 0.7$. Samples from 4MOST (4-metre Multi-Object Spectroscopic Telescope, de Jong et al. 2019) will also provide additional lens samples with spectroscopic redshifts. Here we discuss a number of considerations for future implementations of this work.

9.3.1 Inhomogeneity with angular clustering

In future work, a deeper understanding of inhomogeneity effects could also be gained by investigating the angular clustering, $w(\theta)$ in the regions of overlap with lensing surveys, as well as studying the properties of BOSS or DESI galaxies (e.g. stellar masses and sizes) as a function of various observing conditions (SDSS PSF, stellar density, etc.) using a deeper catalogue than SDSS (e.g. the HSC catalogue).

9.3.2 Joint covariance

In the present analysis, we have assumed that the errors on $\Delta\Sigma$ for measurements by different surveys are uncorrelated. In reality, there is likely to be a small positive correlation between the signals measured by different surveys as discussed in more detail below. The assumption of uncorrelated errors is thus conservative in terms of avoiding false positive detections of discrepancies. Future implementations of this program will need to account for the cross correlations between surveys, especially as the overlap regions between surveys increases. There will be two sources of correlation of statistical errors. One is correlated large-scale structure in the regions of overlap. The second will be correlated shape noise for sources that

are in common between surveys. Differences in depth and binning will reduce the number of common sources in overlap regions and will have to be modelled correctly. For the correlated large-scale structure component, it will be possible to use mock catalogues that are lensed by the same large-scale structure which have the respective *dn/dz* for each survey. For the correlated shape noise component, a data-driven approach will be possible. The correlated shape noise component can be estimated by taking galaxies in common between two given surveys, rotating them by the same angle, and measuring the correlated shape noise.

9.3.3 Tightening constraint on σ_{sys}

The constraints on $\sigma_{\rm sys}$ from this paper are still relatively weak because of the limited overlap between lensing surveys and BOSS. In future tests, it will be desirable to tighten the constraints on $\sigma_{\rm sys}$. One way to tighten constraints on $\sigma_{\rm sys}$ will be to combine across redshift bins to make a more stringent test. However, this will require modelling the cross-covariance between lens bins (due to the overlap in the source populations). Also, if present, inhomogeneity effects must first be characterized and removed (or ruled out) before attempting to combine across redshift bins.

9.3.4 Disentangling biases

The approach presented here provides guidance on the source of any differences that are found (e.g. boost factors versus photo-z bias) but unavoidable degeneracies will also be present (e.g. calibration bias versus photo-zs and methodology). As such, tests on simulations will also be important to disentangle factors that contribute to $\sigma_{\rm sys}$. Such tests are already underway within the context of the DESI collaboration (Lange et al in preparation).

Future implementation of these tests should also consider how to optimally make use of the overlap regions between surveys. For example, using areas of common overlap avoids the issue of inhomogeneous lens samples. However, using the full area will always yield tighter constraints on $\sigma_{\rm sys}$. On the other hand, DESI will be much more homogeneous (Kitanidis et al. 2020) and inhomogeneity may not be an issue for DESI. It will also be interesting to consider joint fits for different effects using radially dependent functional form with the expected scaling for various effects (e.g. one could assume a specific radially dependent functional form for the impact of boost factors).

9.4 Connection with cosmic shear

The ensemble systematic error estimates presented here are relevant for cosmic shear measurements in the sense that the same source galaxies used to measure $\Delta\Sigma$ are also used in cosmic shear measurements. Systematic errors in the source sample that correlate with survey depth will affect both $\Delta\Sigma$, as well as cosmic shear. However, it is not trivial to directly translate the numbers found here into systematic errors for a given cosmic shear tomographic bin. For example, $\Delta\Sigma$ will have a different sensitivity to photo-z errors than cosmic shear. Follow-up work is warranted to study the connection between the numbers reported here for $\Delta\Sigma$ and cosmic shear.

It is also important to note that trends found in this paper could also originate from different methodologies used to compute $\Delta\Sigma$. While teams have performed numerous tests of their methods, it is still possible that trends could originate from different ways in which photo-z's are used for example (although Fig. 13 did not reveal anything obvious). To rule out this possibility, the methodology employed by SDSS, HSC, and CS82 (as implemented in the `dsigma` pipeline) will undergo extensive testing in the DESI Lensing Mock Challenge (Lange et al in preparation) and the next implementation of LWBs will employ a single measurement pipeline. Ruling out methodology differences will also be important in the effort to build the connection between the numbers presented here and cosmic shear.

9.5 Implication for the 'lensing is low' effect

It has been observed that models of the galaxy-halo connection applied to the BOSS CMASS and LOWZ samples over-predict $\Delta\Sigma$ by 20–30 per cent (Leauthaud et al. 2017; Lange et al. 2019, 2021; Singh et al. 2020). What are the implications of the finding of this paper with regards to 'lensing is low'?

Leauthaud et al. (2017) studied the CMASS sample over the full redshift range $0.45 < z < 0.7$. For the range $0.45 < z < 0.54$, we find estimated values for the ensemble systematic error to be of order 5–8 per cent in good agreement with those reported by the lensing surveys. For the redshift range $0.54 < z < 0.7$, we find values that are larger: 12–15 per cent but also with large uncertainties on the reported values. Our results in the range $0.54 < z < 0.7$ also depend on the interpretation of the amplitude-$n_{\rm star}$ trend (because if the $n_{\rm star}$ trend is due to lens inhomogeneity then the variance of the amplitudes in this redshift range should not be attributed to source systematics). This redshift range will therefore require further investigation and constraints on $\sigma_{\rm sys}$ are not tight enough to draw any conclusions at $z_{\rm L} > 0.54$.

Lange et al. (2019) find the 'lensing is low' effect in the LOWZ redshift range using CFHTLenS. Lange et al. (2021) also find the same effect using LOWZ in stellar mass bins using lensing from SDSS. In this redshift range, we find σ_{sys} values that are in good agreement with those reported by lensing surveys.

Using a fully empirical method, we do not find evidence for large (20–30 per cent level) unknown systematic errors. Using the values for $\sigma_{\rm sys}$ derived over the full radial range, systematic errors greater than 15 per cent, 13 per cent, and 14 per cent are ruled out at 68 per cent confidence level for L1, L2, and C1 respectively. Systematic errors greater than 24 per cent, 20 per cent, and 26 per cent are ruled out at 95 per cent confidence level for L1, L2, and C1 respectively. Given that the observed 'lensing is low' effect is at the 20 per cent to 35 per cent level (see e.g. fig. 5 in Lange et al. 2019), this suggests that it is difficult to explain the 'lensing is low' effect via lensing systematic errors alone (e.g. see Fig. 14). A follow-up paper will explore the 'lensing is low' effect with BOSS clustering data and updated lensing data (Amon et al. in preparation).

10 SUMMARY AND CONCLUSIONS

LWBs is a a cross-survey collaboration created to assess the consistency of lensing signals computed with different data sets and to perform empirically motivated tests of lensing systematic errors. Our main tests are based on the premise that the gg-lensing signal ($\Delta\Sigma$) around BOSS galaxies is a physical quantity. The amplitude of $\Delta\Sigma$ should be independent of the lensing data used to perform the measurement. The excess spread (above the expected statistical uncertainties) in the amplitude of $\Delta\Sigma$ among lensing surveys can be used to estimate an ensemble systematic error, $\sigma_{\rm sys}$. We estimate $\sigma_{\rm sys}$ via a blind comparison of the amplitude of $\Delta\Sigma$ using four lens sample from BOSS and using the sources catalogues and methodologies from six distinct lensing surveys (SDSS, CS82, CFHTLenS,

DES, HSC, and KiDS). Our main results are summarized below.

(i) For all redshift and radial ranges, we find good agreement between our empirically estimated values and the reported systematic errors (see Table 5 and Fig. 14). Differences between the estimated and reported systematic errors are always less than 2.3σ.

(ii) Estimated values for $\sigma_{\rm sys}$ are largest in the C2 bin with values that are $\sim 2\sigma$ larger than reported values. But differences remain below 3σ (our pre-blinding determined criterion for claiming a detection). There are also other effects (see below) that complicate interpretation in this redshift range.

(iii) For lenses with $0.43 < z_{\rm L} < 0.54$ (source galaxies with $z_{\rm s} \gtrapprox 0.5$), we detect a correlation between the $\Delta\Sigma$ amplitudes and the depth of lensing surveys (detected at 3–4σ using statistical errors). This correlation explains most of the scatter between the lensing measurements. Section 9.1 presents several explanations for this trend. Two likely candidates are unrecognized blends and photometric redshift calibration. Investigating the origin of this trend is key as it will be a leading term in the systematic error budget for gg-lensing measurements at $z_{\rm s} \gtrapprox 0.5$.

(iv) For lenses between $0.54 < z_{\rm L} < 0.7$ (source galaxies with $z_{\rm s} \gtrapprox 0.7$), we find a correlation between the amplitude of $\Delta\Sigma$ and the foreground stellar density, $n_{\rm star}$ (4–6σ using statistical errors). We raise three possible explanations for this trend (see Appendix E). We perform a test using HSC data (Fig. 11) that leads us to favour an explanation in which the trend originates from a background source systematic (blending, shear calibration, etc.), or that the trend is a statistical fluke, however, the picture could also be more complicated (e.g. see Singh et al. 2020). Further work will be required to draw conclusive statements about the origin of the dispersion among lensing measurements in this redshift range.

(v) The combined effects of the $n_{\rm star}$ and the $z_{\rm s}$ (survey depth) trend explain a majority of the observed scatter in the lensing amplitudes at $z > 0.43$. Investigating the origin of both the $n_{\rm star}$ and the $z_{\rm s}$ trends will be key for gg-lensing measurements in the redshift range $z_{\rm L} > 0.43$.

(vi) All trends except the L1 all-R trend in Fig. 10 drop below 3σ when the estimated systematic errors are summed in quadrature with the statistical errors. This is consistent with our assessment that trends with $z_{\rm s}$ and $n_{\rm star}$ are within the reported systematic error budgets (see Appendix 8.7 and Fig. 2). These trends could originate from 'known known' (and thus accounted for) sources of systematic error.

(vii) Our systematic errors estimates do not appear to be strongly radially dependent. This suggests that boost factors are not a dominant cause of spread between the measurements.

(viii) We compare our empirical estimates of the systematic error to current statistical constraints on the amplitude of gg-lensing (Fig. 14). We find that current measurements have $\sigma_{\rm stat} \sim \sigma_{\rm sys}$. This is a statement about the ensemble of data – individual surveys may have lower systematics than this. Understanding the origin of systematic errors, and reducing uncertainty in our corrections for these errors, will be the next key challenge facing gg-lensing measurements.

(ix) Using a fully empirical method, we do not find evidence for large (20–30 per cent level) unknown systematic errors. Using the values for $\sigma_{\rm sys}$ derived over the full radial range, systematic errors greater than 15 per cent, 13 per cent, and 14 per cent are ruled out at 68 per cent confidence level for L1, L2, and C1, respectively. Systematic errors greater than 24 per cent, 20 per cent, and 26 per cent are ruled out at 95 per cent confidence level for L1, L2, and C1, respectively. These constraints could be made tighter by combining between lens bins – but we have not attempted this here. The overlap between lensing surveys and BOSS limits the constraining power of our tests, but the methods developed here will become more powerful as DESI comes online, and as the coverage of lensing surveys continue to grow.

We have provided the first direct and empirically motivated test of the consistency of the gg-lensing amplitude across lensing surveys and developed a framework for such comparisons. We do not find evidence for large unknown systematic errors in these lensing data. However, systematic errors that are common between different lensing surveys cannot be tested with our methodology.

Our results are relevant for cosmic shear measurements because the same source galaxies used to measure $\Delta\Sigma$ are also used in cosmic shear measurements. However, cosmic shear and galaxy–galaxy lensing are different measurements (for example they are sensitive to photo-z's in different ways) and so the numbers reported here cannot be directly applied to cosmic shear.

For lenses in the range $z_{\rm L} < 0.54$ we find $\sigma_{\rm sys}$ of order 4–9 per cent in good agreement with reported values and systematic errors of 25 per cent are ruled out at 2σ in three different lens bins. In this same redshift range, the 'lensing is low' effect is at 20 per cent–35 per cent (e.g. fig. 5 in Lange et al. 2021). We conclude that for lenses below $z_{\rm L} < 0.54$, it is difficult to explain the 'lensing is low' mis-match with clustering (Leauthaud et al. 2017; Lange et al. 2019, 2021; Singh et al. 2020) via lensing systematic errors alone (Fig. 14). Constraints on $\sigma_{\rm sys}$ are not tight enough to draw any conclusions at $z_{\rm L} > 0.54$. The 'lensing is low' effect in relation to clustering will be explored in greater detail in a companion paper (Amon et al in preparation).

ACKNOWLEDGEMENTS

Author contributions: The authorship list reflects the two lead authors (AL, AA), those who measured the lensing signals and contributed key ideas to the paper (SS, DG, UL, SH, NR, TV, YL, CH, and CB) followed by an alphabetical group.

This paper has gone through internal review by the DES, HSC, and KiDS collaborations.

We acknowledge use of the lux supercomputer at UC Santa Cruz, funded by NSF MRI grant AST 1828315. This material is based on work supported by the U.D Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC0019301. AL acknowledges support from the David and Lucille Packard foundation, and from the Alfred.P Sloan foundation.

CH acknowledges support from the European Research Council under grant number 647112, and support from the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research. KK acknowledges support from the Royal Society and Imperial College. HH is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1), as well as an ERC Consolidator Grant (no. 770935). AHW is supported by an European Research Council Consolidator Grant (no. 770935).

Funding for the DES Projects has been provided by the U.S. Department of Energy; the U.S. National Science Foundation; the Ministry of Science and Education of Spain; the Science and Technology Facilities Council of the United Kingdom; the Higher Education Funding Council for England; the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign; the Kavli Institute of Cosmological Physics at the

University of Chicago; the Center for Cosmology and Astro-Particle Physics at the Ohio State University; the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University; Financiadora de Estudos e Projetos; Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro; Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência; Tecnologia e Inovação; the Deutsche Forschungsgemeinschaft; and the Collaborating Institutions in the Dark Energy Survey.

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

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

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

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

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organisation (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the *Large Synoptic Survey Telescope*. We thank the LSST Project for making their code available as free software at http://dm.lsst.org

This paper is based [in part] on data collected at the *Subaru Telescope* and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant no. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018, and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA, and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples).

This work was supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515 and as a part of the Panofsky Fellowship awarded to DG. MM is partially funded by FAPERJ, CNPq, and CONICET. BM acknowledges support from the Brazilian funding agency FAPERJ.

DATA AVAILABILITY

Our lensing signals and code used to make the main figures are available at https://github.com/alexieleauthaud/lensingwithoutborders.

REFERENCES

Abazajian K. et al., 2004, AJ, 128, 502
Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Ahn C. P. et al., 2014, ApJS, 211, 17
Aihara H. et al., 2011, ApJS, 193, 29
Aihara H. et al., 2018a, Publ. Astron. Soc. Japan, 70, S4
Aihara H. et al., 2018b, Publ. Astron. Soc. Japan, 70, S8
Alam S. et al., 2015, ApJS, 219, 12
Alam S. et al., 2017, MNRAS, 470, 2617
Amon A. et al., 2018a, MNRAS, 477, 4285
Amon A. et al., 2018b, MNRAS, 479, 3422
Annis J. et al., 2014, ApJ, 794, 120
Applegate D. E. et al., 2014, MNRAS, 439, 48
Asgari M. et al., 2019, A&A, 624, A134
Asgari M. et al., 2021, A&A, 645, A104
Benítez N., 2000, ApJ, 536, 571
Blake C. et al., 2016, MNRAS, 456, 2806

Blake C. et al., 2020, A&A, 642, A158
Blazek J., Mandelbaum R., Seljak U., Nakajima R., 2012, J. Cosmol. Astropart. Phys., 2012, 041
Bolton A. S. et al., 2012, AJ, 144, 144
Bosch J. et al., 2018, Publ. Astron. Soc. Japan, 70, S5
Boulade O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham, p. 72
Bridle S. et al., 2009, Ann. Appl. Stat., 3, 6
Buchs R. et al., 2019, MNRAS, 489, 820
Bundy K. et al., 2015, ApJS, 221, 15
Cacciato M., van den Bosch F. C., More S., Mo H., Yang X., 2013, MNRAS, 430, 767
Carlstrom J. E. et al., 2011, PASP, 123, 568
Choi A. et al., 2016, MNRAS, 463, 3737
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
Davis C. et al., 2017, preprint (arXiv:1710.02517)
Dawson K. S. et al., 2013, AJ, 145, 10
de Jong R. S. et al., 2019, The Messenger, 175, 3
DES Collaboration, 2021, preprint (arXiv:2105.13549)
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Drlica-Wagner A. et al., 2018, ApJS, 235, 33
Dunkley J. et al., 2009, ApJS, 180, 306
Dvornik A. et al., 2018, MNRAS, 479, 1240
Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, The Messenger, 154, 32
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Eisenstein D. J. et al., 2011, AJ, 142, 72
Erben T. et al., 2005, Astron. Nachr., 326, 432
Erben T. et al., 2009, A&A, 493, 1197
Erben T. et al., 2013, MNRAS, 433, 2545
Feldmann R. et al., 2006, MNRAS, 372, 565
Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, MNRAS, 467, 1627
Finkbeiner D. P. et al., 2016, ApJ, 822, 66
Flaugher B. et al., 2015, AJ, 150, 150
Ford J. et al., 2015, MNRAS, 447, 1304
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Gatti M. et al., 2018, MNRAS, 477, 1664
Giblin B. et al., 2021, A&A, 645, A105
Górski K. M., Hivon E., Banday A. J., Wand elt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gruen D. et al., 2014, MNRAS, 442, 1507
Gruen D., Brimioulle F., 2017, MNRAS, 468, 769
Gunn J. E. et al., 1998, AJ, 116, 3040
Gunn J. E. et al., 2006, AJ, 131, 2332
Heymans C. et al., 2006, MNRAS, 368, 1323
Heymans C. et al., 2012, MNRAS, 427, 146
Heymans C. et al., 2021, A&A, 646, A140
Hikage C. et al., 2019, Publ. Astron. Soc. Japan, 71, 43
Hildebrandt H. et al., 2010, A&A, 523, A31
Hildebrandt H. et al., 2012, MNRAS, 421, 2355
Hildebrandt H. et al., 2016, MNRAS, 463, 635
Hildebrandt H. et al., 2020, A&A, 633, A69
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hoyle B. et al., 2018, MNRAS, 478, 592
Huang S. et al., 2018a, Publ. Astron. Soc. Japan, 70, S6
Huang S., Leauthaud A., Greene J. E., Bundy K., Lin Y.-T., Tanaka M., Miyazaki S., Komiyama Y., 2018b, MNRAS, 475, 3348
Huang H.-J., Eifler T., Mandelbaum R., Dodelson S., 2019, MNRAS, 488, 1652
Huff E., Mandelbaum R., 2017, preprint (arXiv:1702.02600)
Jee M. J., Tyson J. A., Schneider M. D., Wittman D., Schmidt S., Hilbert S., 2013, ApJ, 765, 74
Joachimi B. et al., 2015, Space Sci. Rev., 193, 1
Joudaki S. et al., 2020, A&A, 638, L1
Kannawadi A. et al., 2019, A&A, 624, A92
Kilbinger M. et al., 2017, MNRAS, 472, 2126
Kitanidis E. et al., 2020, MNRAS, 496, 2262
Kneib J.-P. et al., 2003, ApJ, 598, 804
Komiyama Y. et al., 2018, Publ. Astron. Soc. Japan, 70, S2
Krause E. et al., 2017, preprint (arXiv:1706.09359)
Kuijken K. et al., 2015, MNRAS, 454, 3500
Laigle C. et al., 2016, ApJS, 224, 24
Lange J. U., Yang X., Guo H., Luo W., van den Bosch F. C., 2019, MNRAS, 488, 5771
Lange J. U., Leauthaud A., Singh S., Guo H., Zhou R., Smith T. L., Cyr-Racine F.-Y., 2021, MNRAS, 502, 2074
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Leauthaud A. et al., 2012, ApJ, 744, 159
Leauthaud A. et al., 2017, MNRAS, 467, 3024
Lee S. et al., 2019, MNRAS, 489, 2887
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Luis Bernal J., Peacock J. A., 2018, J. Cosmol. Astropart. Phys., 2018, 002–002
Lyons L., 2008, Ann. Appl. Stat., 2
MacCrann N. et al., 2022, MNRAS 509 3371
Mandelbaum R., Tasitsiomi A., Seljak U., Kravtsov A. V., Wechsler R. H., 2005, MNRAS, 362, 1451
Mandelbaum R., Seljak U., Cool R. J., Blanton M., Hirata C. M., Brinkmann J., 2006, MNRAS, 372, 758
Mandelbaum R., Hirata C. M., Leauthaud A., Massey R. J., Rhodes J., 2012, MNRAS, 420, 1518
Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, MNRAS, 432, 1544
Mandelbaum R. et al., 2014, ApJS, 212, 5
Mandelbaum R. et al., 2018a, Publ. Astron. Soc. Japan, 70, S25
Mandelbaum R. et al., 2018b, MNRAS, 481, 3170
Mandelbaum R. et al., 2018c, MNRAS, 481, 3170
Massey R. et al., 2007, MNRAS, 376, 13
Masters D. et al., 2015, ApJ, 813, 53
McClintock T. et al., 2019, MNRAS, 482, 1352
Melchior P. et al., 2015, MNRAS, 449, 2219
Miller L. et al., 2013, MNRAS, 429, 2858
Miralda-Escude J., 1991, ApJ, 370, 1
Miyazaki S. et al., 2018, Publ. Astron. Soc. Japan, 70, S1
Myles J. et al., 2021, MNRAS, 505, 4249
Nakajima R., Mandelbaum R., Seljak U., Cohn J. D., Reyes R., Cool R., 2012, MNRAS, 420, 3240
Newman J. A. et al., 2013, ApJS, 208, 5
Planck Collaboration, 2020, A&A, 641, A1
Prat J. et al., 2018, Phys. Rev. D, 98, 042005
Reid B. et al., 2016, MNRAS, 455, 1553
Reyes R., Mandelbaum R., Gunn J. E., Nakajima R., Seljak U., Hirata C. M., 2012, MNRAS, 425, 2610
Ross A. J. et al., 2011, MNRAS, 417, 1350
Ross A. J. et al., 2012, MNRAS, 424, 564
Ross A. J. et al., 2017, MNRAS, 464, 1168
Schirmer M., Erben T., Schneider P., Wolf C., Meisenheimer K., 2004, A&A, 420, 75
Schlafly E. F. et al., 2010, ApJ, 725, 1175
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Sheldon E. S. et al., 2004, AJ, 127, 2544
Simet M., Mandelbaum R., 2015, MNRAS, 449, 1259
Simet M., McClintock T., Mandelbaum R., Rozo E., Rykoff E., Sheldon E., Wechsler R. H., 2016, preprint (arXiv:1603.06953)
Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, MNRAS, 471, 3827

Singh S., Mandelbaum R., Seljak U., Rodríguez-Torres S., Slosar A., 2018, preprint (arXiv:1811.06499)
Singh S., Mandelbaum R., Seljak U., Rodríguez-Torres S., Slosar A., 2020, MNRAS, 491, 51
Smee S. A. et al., 2013, AJ, 146, 32
Speagle J. S. et al., 2019, MNRAS, 490, 5658
Spergel D. et al., 2015, preprint (arXiv:1503.03757)
Stoughton C. et al., 2002, AJ, 123, 485
Tanaka M. et al., 2018, Publ. Astron. Soc. Japan, 70, S9
The Dark Energy Survey Collaboration, 2015, preprint (arXiv:1507.05552)
Tojeiro R. et al., 2014, MNRAS, 440, 2222
Troxel M. A. et al., 2018a, Phys. Rev. D, 98, 043528
Troxel M. A. et al., 2018b, MNRAS, 479, 4998
Unruh S., Schneider P., Hilbert S., Simon P., Martin S., Puertas J. C., 2020, A&A, 638, A96
Varga T. N. et al., 2019, MNRAS, 489, 2511
von Wietersheim-Kramsta M. et al., 2021, MNRAS, 504, 1452
Wilson G., Kaiser N., Luppino G. A., Cowie L. L., 2001, ApJ, 555, 572
Wright A. H. et al., 2019, A&A, 632, A34
Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., 2020, A&A, 637, A100
Xia Q. et al., 2020, A&A, 633, A89
York D. G. et al., 2000, AJ, 120, 1579
Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, MNRAS, 434, 1604
Zuntz J. et al., 2018, MNRAS, 481, 1149

APPENDIX A: IMPACT OF ONE PERCENT FLUX CALIBRATION ON VARIATIONS IN $\Delta\Sigma$

Modern day surveys control flux calibration to about 1 per cent or better. This means that $\delta f/f \sim 0.01$. Stellar mass is proportional to flux. Ignoring the fact that we need to use colours to estimate M^* and considering instead only the scaling of M^* with total luminosity, this means the impact of flux uncertainties on M^* is also of order 1 per cent. Assuming the stellar-to-halo mass relation of Leauthaud et al. (2012) and the redshift and mass range of CMASS, this corresponds to a halo mass shift of 0.0043 dex. The lensing observable $\Delta\Sigma$ in the one-halo regime scales as $\Delta\Sigma \propto (M_{\rm halo})^{2/3}$ so this imparts a 0.6 per cent shift on $\Delta\Sigma$. This level of difference is not detectable with current surveys. Other effects, such as the ability to accurately model and measure the total luminosities of massive galaxies, including their outer envelopes (e.g. Huang et al. 2018b, 2019), or dust corrections, are likely to be more important.

APPENDIX B: DERIVATION OF $F_{\rm BIAS}$ CORRECTION FACTOR

Using photo-z point source estimates for the source redshift in the calculation of $\Delta\Sigma_{\rm crit}$ can bias estimates of $\Delta\Sigma$. However, this bias, which we call $f_{\rm bias}$, can be estimated from a calibration sample. This calibration sample can be formed from galaxies with known spectroscopic redshifts or from galaxies with higher quality photometric redshifts, such as the COSMOS 30-band catalogue (e.g. Speagle et al. 2019). Here, we derive the estimator for $f_{\rm bias}$ if such a calibration sample is available.

$f_{\rm bias}$ is the expected ratio of the true $\Delta\Sigma_{\rm T}$ versus the estimate using photometric redshifts $\Delta\Sigma_{\rm P}$:

$$f_{\rm bias}^{-1} = \frac{\Delta\Sigma_{\rm P}}{\Delta\Sigma_{\rm T}} = \frac{1}{\Delta\Sigma_{\rm T}} \frac{\sum_{\rm Ls} w_{\rm sys} w_{\rm Ls}\ \gamma_{\rm t,Ls}\ \Sigma_{\rm c,Ls,P}}{\sum_{\rm Ls} w_{\rm sys} w_{\rm Ls}}. \tag{B1}$$

In the above equation, $\Delta\Sigma_{\rm T}$ is defined analogously to $\Delta\Sigma_{\rm P}$ but with $\Sigma_{\rm c,Ls,T}$ instead of $\Sigma_{\rm c,Ls,P}$. Here, $\Sigma_{\rm c,Ls,T}$ and $\Sigma_{\rm c,Ls,P}$ correspond to the critical surface density derived from high-quality and photometric redshifts, respectively.

In principle, the above equation can only be solved if high-quality redshifts for all sources are known. However, we can get an estimate for $f_{\rm bias}$ using a representative sample of sources with high-quality redshifts. The high-quality sample does not need to be perfectly representative of the full source sample. Instead, we can account for systematic differences with respect to colour using calibration weights ($w_{\rm calib,s}$) applied to the high-quality sample.

Furthermore, we note that the expectation value for the tangential shear $\gamma_{\rm t,Ls}$ is $\Delta\Sigma_{\rm T}/\Sigma_{\rm c,Ls,T}$. Thus, we can approximate:

$$f_{\rm bias}^{-1} \approx \frac{\sum_{\rm Ls} w_{\rm sys}\ w_{\rm calib,s} w_{\rm Ls}\ \Sigma_{\rm c,Ls,P} \Sigma_{\rm c,Ls,T}^{-1}}{\sum_{\rm Ls} w_{\rm sys} w_{\rm calib,s}\ w_{\rm Ls}} \tag{B2}$$

We note that the above expression does not depend anymore on the actual value of the excess surface density, $\Delta\Sigma$. Thus, unless the systematic lens weights depend systematically on redshift, they

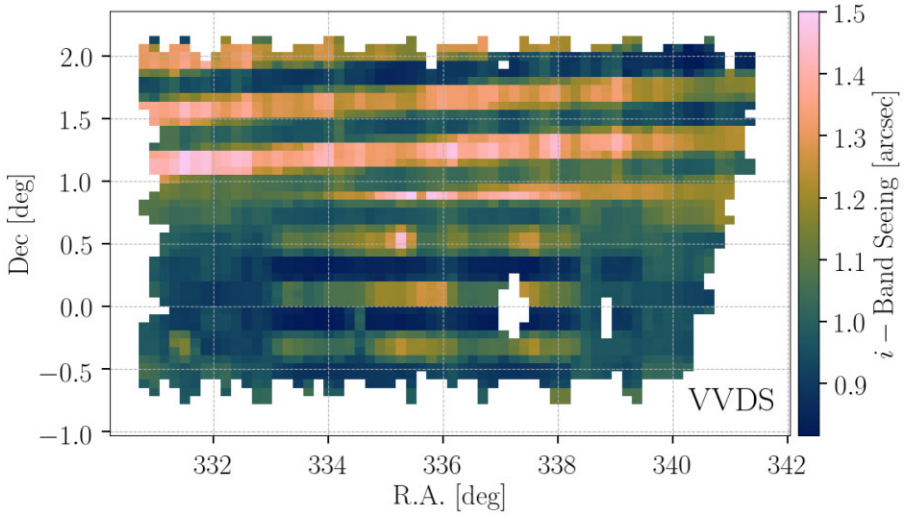

Figure C1. SDSS PSF in the HSC regions. Strong spatial variations in the PSF mean that it is difficult to directly estimate inhomogeneity effects from the data because one cannot use jack-knife errors when separating regions with low and high PSF.

can safely be neglected from the above equation. Finally, substituting equation (27) for the lens-source weight $w_{\rm Ls}$ into the above expression leads to equation (29). This form for $f_{\rm bias}$ includes the dilution factor correction. Note that $f_{\rm bias}$ is only meant to correct for photometric redshift errors of sources physically uncorrelated with the lens. Biases in $\Delta\Sigma$ due to sources physically associated with the lens are incorporated in the boost factor.

APPENDIX C: SPATIAL MAPS OF HOMOGENEITY

Fig. C1 shows spatial variations of the SDSS PSF in HSC regions. Strong spatial variations in the SDSS PSF mean that errors cannot be estimated from jack-knife when separating the data into separate regions of low and high SDSS PSF.

APPENDIX D: IMPACT OF MISSING REDSHIFTS

Our fiducial signals include the BOSS weight, w_z, to account for missing redshifts. None the less, we have also studied the impact of missing redshifts for the LOWZ sample using SDSS. In Fig. D1, we show the ratio of impact of redshift weights (weights accounting for missing redshifts due to close pairs and redshift failures). Using the redshift weights increases the $\Delta\Sigma$ by $\sim (2 \pm 1)$ per cent per cent in the North, and $\sim (5 \pm 2)$ per cent in the South (see Table D1). For our lensing signals, the missing redshift weight is applied as a part of the total systematic lens weight $w_{\rm tot}$. If the w_z weight is not applied, we can expect differences related to missing redshift to be small, less than 5 per cent.

APPENDIX E: TREND WITH $n_{\rm STAR}$

Here, we present further discussion on possible origins for the $n_{\rm star}$ trend.

E1 Lens inhomogeneity

One explanation is that there may be intrinsic variations in the halo masses of CMASS that correlate with $n_{\rm star}$. This is plausible because we know from Ross et al. (2017) that the number density of CMASS galaxies varies with $n_{\rm star}$ with number density variations that depend on galaxy surface brightness. These effects are strongest in the C2 bin. In addition, Singh et al. (2020) recently found that higher $n_{\rm star}$ on average leads to lower observed surface brightness of galaxies and that surface brightness was observed to be negatively correlated with the galaxy bias. On the other hand, our internal test using HSC displayed in Fig. 11 using HSC data alone does not favour this hypothesis. The question of lens inhomogeneity therefore remains a puzzle.

If lens inhomogeneity is indeed the correct explanation, there are a number of interesting consequences to consider. First, the impact of inhomogeneity in lens samples on combined probe analyses of lensing and clustering has typically not been well studied. The trend identified in Fig. 10 could impact combined probe studies because (with the exception of SDSS) most lensing surveys cover smaller areas than the BOSS footprint and have different $n_{\rm star}$ distributions compared to BOSS. Secondly, it is also interesting to note that the BOSS systematic weights do not correct for the effects. This is because $w_{\rm sys}$ is simply designed to homogenize number densities across the BOSS footprint but knows nothing about correlations between galaxies that are lost near bright stars and halo mass. In order to design a $w_{\rm sys}$ that would correct for this effect, one would first need to understand and map the physical origin of the $n_{\rm star}$ which will necessitate (a) a physical understanding of which galaxies are being lost near bright stars, and (b) connecting this with knowledge about the high mass end of the stellar-to-halo mass relation. Thirdly, lens inhomogeneity could be of importance for the derivation of covariance matrices for both clustering and for lensing (halo mass variations across the sky are typically not considered).

In future work, it will be important to devise additional ways of testing whether or not lens inhomogeneity is present in the CMASS data. However, BOSS will also soon be superseded by data from DESI. DESI lens samples are expected to be more homogeneous than those of BOSS (Kitanidis et al. 2020) but the precision of joint probe studies with DESI will also be greater compared to BOSS and so the requirement of homogeneous lens samples will also be more stringent for DESI compared to BOSS.

E2 Source systematic that correlates with $n_{\rm star}$

The second explanation is that instead of being a variation in the halo masses of the lenses, this could be a systematic in the source sample that correlates with $n_{\rm star}$ and leads to biased estimates of $\Delta\Sigma$. It is possible that the same stars that cause number density variations in CMASS are also responsible for some lensing systematic. For example, $n_{\rm star}$ could also impose a surface brightness selection on source galaxies. The presence of a higher background around bright stars could also impact deblending algorithms. Or the correlation could be due to a different parameter that correlates with $n_{\rm star}$ such as galactic extinction.

If this explanation is correct, then it has important consequences. Indeed, the correlation with $n_{\rm star}$ is the leading cause of scatter in C2. This could therefore be the leading term in the systematic error budget for source galaxies at $z_{\rm s} > 0.7$. Investigating the origin of this correlation could therefore help to reduce systematic errors for source galaxies at high redshifts.

E3 No trend

The third possibility is that the trend is a statistical fluctuation and that there is in fact no trend with $n_{\rm star}$ (e.g. see Section 3.6). Indeed, the trend is sensitive to the data point at low $n_{\rm star}$. In this case, the observed spread between the data would be related to source systematics, but these systematics would not correlate with $n_{\rm star}$. It is clear that further work will be required to fully elucidate this question. An interesting avenue to also pursue would be to investigate the correlation between the lensing amplitudes and other quantities such as galactic extinction.

APPENDIX F: LENS MAGNIFICATION

For CMASS, the *i*-band CMODEL flux is the primary cut that limits the number of CMASS galaxies in the range $z > 0.54$. Fig. F1 shows the number of galaxies above a given *i*-band flux divided by the *i*-band flux limit. The flux distribution is fairly steep with slope $\alpha \sim 3$ (dotted line) and α is steeper than 1 (dashed line).

Fig. F2 displays an estimate of the lens magnification effect for CMASS assuming $z_{\rm L} = 0.6$, $z_{\rm s} = 0.8$, and $\alpha = 3$ (see also

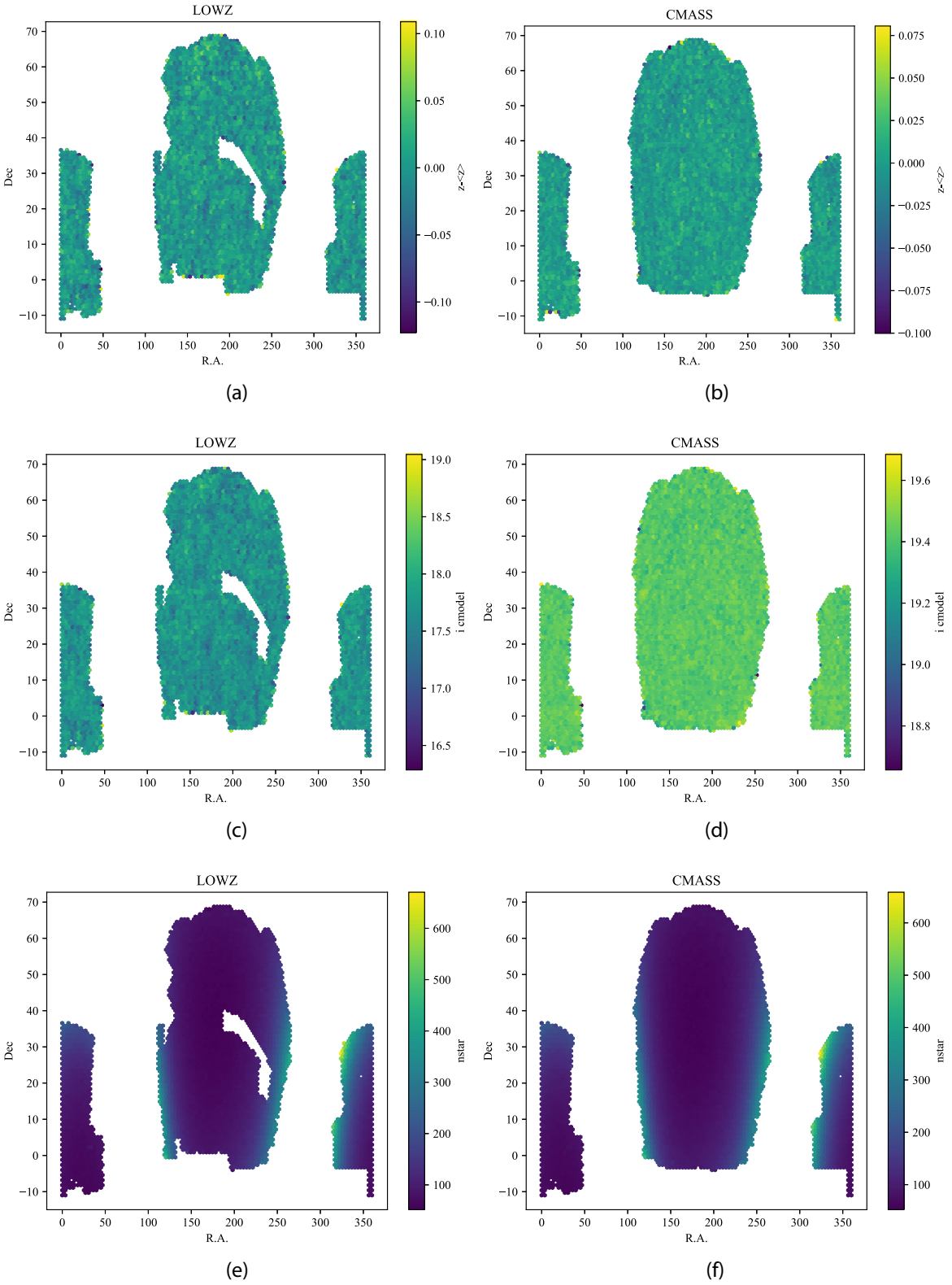

Figure C2. Spatial variation of $z - \overline{z}$, $i_{\rm cmod}$, and $n_{\rm star}$ for LOWZ and CMASS DR14 data.

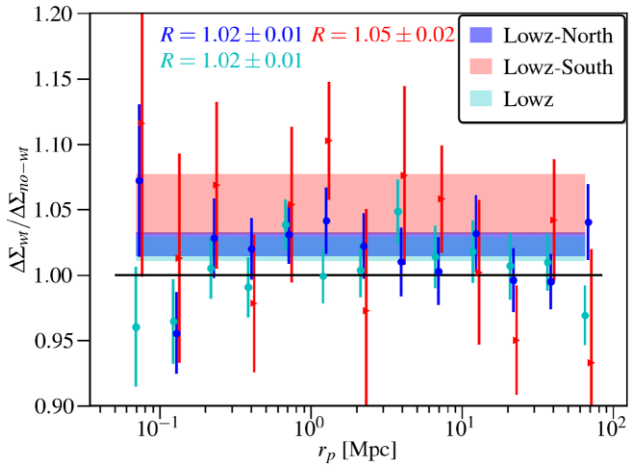

Figure D1. The ratio, R, of $\Delta\Sigma$ obtained using z weights and no weights. Since z-weights tend to up weight the regions with more galaxies, $\Delta\Sigma$ obtained using weights is slightly larger. The bands and the numbers quoted in the text are the mean values of R obtained by fitting a constant to values between $0.5 < r_p < 15$ Mpc.

Table D1. Results for differences between North and South and the impact of missing redshifts. This was computed using SDSS lensing and the LOWZ sample.

Effect	All R
North v.s. South	0.95 ± 0.08
Missing redshifts	1.02 ± 0.01
Missing redshifts – North	1.02 ± 0.01
Missing redshifts – South	1.05 ± 0.02

von Wietersheim-Kramsta et al. 2021). Fig. F2 shows the spurious additive signal, $\Delta\Sigma_{\rm LM}$, caused by lens magnification.

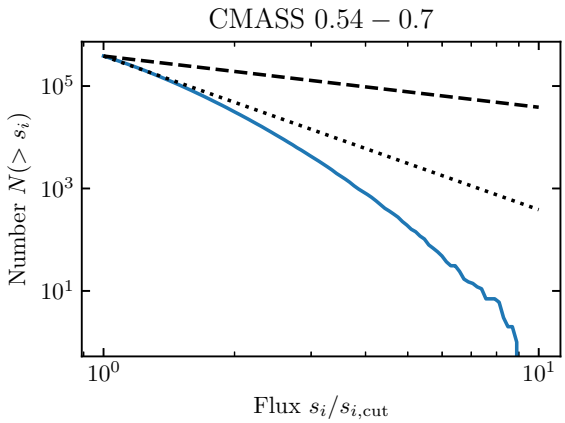

Figure F1. Number of galaxies above a given i-band flux divided by the i-band flux limit for CMASS. CMASS has a slope of $\alpha \sim 3$ (dotted line). The slope is steeper than 1 (dashed line).

[1]*Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064 USA*
[2]*Kavli Institute for Particle Astrophysics & Cosmology, P.O. Box 2450, Stanford University, Stanford, CA 94305, USA*
[3]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[4]*Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA*
[5]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*
[6]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[7]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[8]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[9]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[10]*Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, D-44780 Bochum, Germany*
[11]*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn VIC 3122, Australia*
[12]*Laboratoire d'Annecy De Physique Des Particules (LAPP), 9 Chemin de Bellevue, F-74940 Annecy, France*
[13]*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro RJ-20921-400, Brazil*
[14]*Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA*
[15]*Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*
[16]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[17]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[18]*CEA Irfu, DAp, AIM, Universitè Paris-Saclay, Universitè de Paris, CNRS, F-91191 Gif-sur-Yvette, France*
[19]*Department of Physics, Northeastern University, Boston, MA 02115, USA*
[20]*Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[21]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[22]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[23]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[24]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[25]*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain*

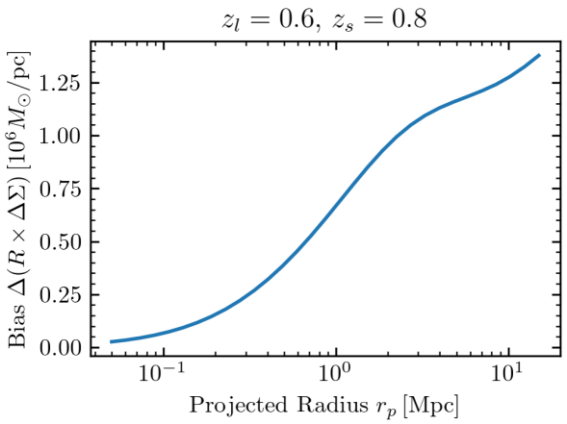

Figure F2. Impact of lens magnification on $\Delta\Sigma$. This is the additive spurious signal caused by lens magnification.

[26]*Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*
[27]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[28]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[29]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[30]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[31]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390, USA*
[32]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[33]*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
[34]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[35]*Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[36]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ-20921-400, Brazil*
[37]*Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029 Hamburg, Germany*
[38]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain*
[39]*Physics Departement, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[40]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[41]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[42]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[43]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[44]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[45]*Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway*
[46]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[47]*Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland*
[48]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[49]*ASTRAVEO LLC, P.O. Box 1668, MA 01931, USA*
[50]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[51]*Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA*
[52]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[53]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[54]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[55]*Leiden Observatory, Leiden University, P.O.Box 9513, NL-2300 RA Leiden, the Netherlands*
[56]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil*
[57]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
[58]*International Center for Advanced Studies & ICIFI, CONICET, ECyT-UNSAM, 1650, Buenos Aires, Argentina*
[59]*Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, CEP 22290-180, Rio de Janeiro, RJ, Brazil*
[60]*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*
[61]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[62]*Kobayashi-Maskawa Institute for the Origin of Particles and the Universe (KMI), Nagoya University, Nagoya 464-8602, Japan*
[63]*Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-8583, Japan*
[64]*Instituto de Fisica, Universidade Federal do Rio de Janeiro, 21941-972, Rio de Janeiro, RJ, Brazil*
[65]*The Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India*
[66]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[67]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[68]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[69]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[70]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA*
[71]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[72]*Physics Department, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720 USA*
[73]*Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile*
[74]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[75]*Department of Statistical Sciences, University of Toronto, Toronto, M5S 3G3, Canada*
[76]*David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, M5S 3H4, Canada*
[77]*Dunlap Institute for Astronomy & Astrophysics, University of Toronto, Toronto, M5S 3H4, Canada*
[78]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[79]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[80]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10012, USA*
[81]*Department of Physics, Duke University, Durham, NC 27708, USA*
[82]*Department of Astronomy and Astrophysics, University of British Columbia, 6224 Agricultural Road, V6T 1Z1, Vancouver, Canada*

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