Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **498,** 2948–2956 (2020)
Advance Access publication 2020 August 21

doi:10.1093/mnras/staa2537

Compressing combined probes: redshift weights for joint lensing and clustering analyses

Rossana Ruggeri ★ and Chris Blake

Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia

Accepted 2020 August 17. Received 2020 August 3; in original form 2020 June 19

ABSTRACT

Combining different observational probes, such as galaxy clustering and weak lensing, is a promising technique for unveiling the physics of the Universe with upcoming dark energy experiments. Whilst this strategy significantly improves parameter constraints, decreasing the degeneracies of individual analyses and controlling the systematics, processing data from tens of millions of galaxies is not a trivial task. In this work, we derive and test a new compressed statistic for joint clustering and lensing data analysis, maximizing the scientific return and decreasing the computational cost. Our approach compresses the data by up-weighting the components most sensitive to the parameters of interest, with no loss of information, taking into account information from the cross-correlation between the two probes. We derive optimal redshift weights which may be applied to individual galaxies when testing a given statistic and cosmological model.

Key words: gravitational lensing: weak – methods: statistical – large-scale structure of Universe.

1 INTRODUCTION

Combining different observational probes is a promising technique to unveil the physics of the Universe with upcoming dark energy experiments. First, any tensions or inconsistencies between different probes can indicate new physics or help us correct for systematic errors not controlled in an individual analysis. Secondly, a joint analysis significantly improves measurements of the parameters of interest, decreasing the degeneracies of an individual analysis (Bernstein 2009; Joachimi & Bridle 2010; Yoo & Seljak 2012).

The potential of these tests will be greatly enhanced by current and future cosmological surveys such as the Kilo-Degree Survey (de Jong et al. 2013), Dark Energy Survey (Abbott et al. 2018), Hyper-Suprime-Cam (HSC) lensing survey (Aihara et al. 2018), Large Synoptic Survey Telescope (Ivezić et al. 2019) and *Euclid* satellite for gravitational lensing (Laureijs et al. 2011), and the Dark Energy Spectroscopic Instrument (Levi et al. 2019) and 4-metre Multi-Object Spectroscopic Telescope for galaxy clustering (de Jong et al. 2012). Whilst this large volume of data represents a unique opportunity to understand the Universe, processing tens of millions of galaxies to detect the subtle signatures of new physics is not a trivial task. Developing new algorithms and strategies to analyse this data is critical to maximize the outcome of these investments.

Further, these unprecedented data volumes create another key challenge: how do we combine information from galaxies at different epochs in the evolution of the Universe? Past analyses dealt with this evolution in the data by binning galaxies in different sub-samples by epoch. However, this technique is inefficient for several reasons: it assumes no evolution within each bin, it neglects the cross-correlation between sub-samples, and it is time-consuming because we are required to repeat the same analysis for each sub-sample of galaxies. Moreover, systematic error may be imprinted by redshift evolution, if the same galaxy carries different weights towards different statistics in the joint analysis.

Rather than breaking the sample into multiple subsets, optimal weighting of the data is an alternative to this traditional approach which instead compresses the data, maintaining sensitivity to evolution in the sample. Strategies for how to compress data have gained increasing attention as a powerful method to handle 'big data', compared to brute-force data analysis (Tegmark, Taylor & Heavens 1997; Heavens, Jimenez & Lahav 2000). As discussed in Tegmark et al. (1997), optimal weighting based on the Karhunen–Loéve approach can compress a data set with no loss of information if the mean is known, obtaining results with close-to-maximal accuracy. In simple words, the optimal weights identify those aspects of the data that are most sensitive to the physics we care about, and amplify them with respect to other aspects of the data, which contribute mostly to the noise. Similar to a principal component analysis, these weighted modes are constructed to be an optimal estimate of the cosmological parameters of interest through the Fisher Information Matrix.

Tegmark et al. (1997) discussed the need for data-compression when analysing the cosmic microwave background (CMB) with $>10^7$ pixel all-sky maps, where a direct numerical inversion of the covariance matrix is clearly unfeasible. More recently, Mootoovaloo et al. (2020) discussed the application of a data-compression algorithm such as MOPED (Heavens et al. 2000) to weak-lensing measurements.

Previous studies have also developed optimal weighting schemes for data compression with focus on measuring the growth rate of structure (Ruggeri et al. 2017, 2019a, b; Zhao et al. 2019), angular diameter distance (Zhu et al. 2018), primordial non-Gaussianity (Castorina et al. 2019) and cosmic shear (Bellini et al. 2019). These

★ E-mail: rruggeri@swin.edu.au

studies explored optimal weighting for measurements with individual probes, demonstrating how an optimal weighting scheme applied to a data set gives unbiased results and is efficient in decreasing the computational costs. Our current study extends the MOPED algorithm and the 'redshift weights' scheme developed by Ruggeri et al. (2017) to galaxy–galaxy lensing statistics, and the combination of lensing and clustering measurements. The MOPED algorithm, presented in Heavens et al. (2000), allows for a dramatic reduction in the data set size with little or no loss of information.

When combining multiple probes, the weights for the individual probes, e.g. for clustering or lensing only, 'lose their optimality' if we neglect the cross-correlation between the different probes, which contains important information on the parameter space we are exploring. In this work, we derive and test a new compressed statistic for combining galaxy–galaxy lensing and galaxy clustering based on their covariance. Our optimal data compression presents various advantages with respect to a more standard approach of tomographic redshift binning: by compressing the information along the redshift direction it allows for a time-efficient analysis and drastically reduces the computational time and covariance requirements, enabling us to perform data analysis over a wide redshift bin. The weights depend on the specific cosmological statistic and fiducial model and may not be optimal for other cosmological models; however, in this case the resulting fitted parameters will remain unbiased.

The paper is organized as follows. In Section 2.1, we briefly describe the model for the galaxy clustering and galaxy–galaxy lensing (cross-)power spectra and covariance. In Section 3, we derive the optimal weights to be applied to the lenses to optimize the statistical error of the combined probes fit. In Section 4, we test parameter fits based on our compressed statistic using Gaussian realizations and compare the results with uncompressed analyses. In particular, we verify that the derived weights produce a lossless compression of the data and unbiased results. Section 5, we discuss a scenario in which optimal weights reduce systematic biases in fitted parameters by tracing the redshift evolution of the galaxy bias. In Section 6, we conclude by discussing future applications and extensions of this method.

2 MODELS AND COVARIANCE

2.1 Angular power spectra for combined probes

We model the angular (cross-)power spectra between two different fields δ_a, δ_b of redshift samples i, j, as a function of projected Fourier mode, ℓ, as

$$C^{ij}_{ab}(\ell) = \int \mathrm{d}\chi \, \frac{q^i_a(\chi) q^j_b(\chi)}{\chi^2} P_{ab}(\ell/\chi, z(\chi)), \tag{1}$$

where $P_{ab}(k, z)$ is the 3D (cross-)power spectrum of the fields at wavenumber k and redshift z, and $\chi(z)$ is the comoving distance (Hu & Jain 2004; Krause & Eifler 2017). The weight function $q_{a,b}(\chi)$ depends on the field considered: we focus here on auto- and cross-correlations between gravitational lensing and galaxy large-scale structure.

For the galaxy density field δ_g, $q_g(\chi)$ is proportional to the redshift distribution of galaxies in each bin

$$q^i_g(\chi) = \frac{n^i_{\rm lens}(z)}{\bar{n}^i_{\rm lens}} \frac{\mathrm{d}z}{\mathrm{d}\chi}, \tag{2}$$

where $n^i_{\rm lens}(z)$ is the lens redshift distribution of sample i, with z the redshift corresponding to χ, and $\bar{n}^i_{\rm lens}$ is the average lens density.

For the convergence field δ_κ, $q_\kappa(\chi)$ is given by the lensing efficiency

$$q^i_\kappa(\chi) = \frac{3H_0^2\Omega_{\rm m}}{2c^2} \frac{\chi}{a(\chi)} \int_\chi^{\chi_{\rm max}} \mathrm{d}\chi' \frac{n^i_{\rm source}(z)}{\bar{n}^i_{\rm source}} \frac{\mathrm{d}z}{\mathrm{d}\chi'} \frac{(\chi' - \chi)}{\chi'}, \tag{3}$$

where $\Omega_{\rm m}$ and H_0 are the values of the present-day matter density and Hubble parameter, $\chi_{\rm max}$ is the maximum comoving distance of the source distribution, and $n^i_{\rm source}(z)$ and $\bar{n}^i_{\rm source}$ are the source redshift distribution and average density of sources in sample i. We note that equation (1) is derived assuming the Limber and flat-sky approximations (Lemos, Challinor & Efstathiou 2017).

2.2 Covariance matrix

The Gaussian covariance matrix between two angular power spectra $C^{ij}_{ab}(\ell_1)$, $C^{kl}_{cd}(\ell_2)$, for samples (i, j, k, l) is given by Hu & Jain (2004) and Krause & Eifler (2017)

$$\begin{aligned}\mathbf{C} = &\frac{4\pi\delta_{\ell_1\ell_2}}{\Omega_s(2\ell_1+1)\Delta\ell_1} \\ &\times \left[\left(C^{ik}_{ac}(\ell_1) + \delta_{ik}\delta_{ac}N^i_a\right)\left(C^{jl}_{bd}(\ell_2) + \delta_{jl}\delta_{bd}N^j_b\right)\right. \\ &\left. + \left(C^{il}_{ad}(\ell_1) + \delta_{il}\delta_{ad}N^i_a\right)\left(C^{jk}_{bc}(\ell_2) + \delta_{jk}\delta_{bc}N^j_b\right)\right],\end{aligned} \tag{4}$$

where Ω_s is the angular area of the overlapping sample in steradians. For galaxy–galaxy lensing, the covariance of the angular power spectrum $C_{g\kappa}$ depends on the C_{gg}, $C_{g\kappa}$, and $C_{\kappa\kappa}$ terms. For these probes the noises terms are

$$\begin{aligned}N_{gg} &= 1/\bar{n}_{\rm lens}, \\ N_{\kappa\kappa} &= \sigma_e^2/\bar{n}_{\rm source},\end{aligned} \tag{5}$$

where σ_e is the shape noise.

2.3 Fiducial cosmology

We adopt a fiducial cosmological model with matter density $\Omega_{\rm m} = 0.3$, baryon density $\Omega_{\rm b} = 0.044$, Hubble parameter $h = 0.7$, amplitude of matter clustering $\sigma_8 = 0.8$ and spectral index $n_s = 0.95$. For the galaxy bias model we choose a simple redshift-dependent relation

$$b(z) = b_{\rm piv} \frac{D(z_{\rm piv})}{D(z)}, \tag{6}$$

where $D(z)$ is the linear growth rate and we selected $b_{\rm piv} = 2$ as the value of the galaxy bias at the pivot redshift $z_{\rm piv} = 0.45$. This relation is approximately correct for the clustering amplitude of magnitude-selected galaxy samples (Eisenstein et al. 2001). To model the galaxy–galaxy and galaxy-convergence power spectra P_{gg} and $P_{g\kappa}$, we assume a linear bias relation where $P_{gg} \propto b^2\sigma_8^2$ and $P_{g\kappa} \propto b\sigma_8^2$. The power spectrum of the matter on non-linear scales is computed from CAMB (Lewis & Bridle 2002).

3 OPTIMAL WEIGHTS METHODOLOGY

We are interested in defining optimal redshift weights which average measurements from samples at different redshifts into a single final data set containing the same information, i.e. which perform lossless data compression. In this section, we briefly introduce the optimal weights formalism and derive weights to combine galaxy clustering and galaxy–galaxy lensing measurements, $w_{gg+g\kappa}$, comparing them with individual-probe weights w_{gg} and $w_{g\kappa}$.

3.1 Derivation

3.1.1 Optimal weights for a single parameter

Consider a data set $\mathbf{x}$ containing n values, Gaussian-distributed with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$. A linear compression transforms this data set into a single number y:

$$y = \mathbf{w}^T\mathbf{x}, \tag{7}$$

where $\mathbf{w}$ is a vector of weights of length n. The compressed measurement y has mean $\mathbf{w}^T\boldsymbol{\mu}$ and variance $\mathbf{w}^T\mathbf{C}\mathbf{w}$ (Tegmark et al. 1997).

In order to obtain lossless compression we need to select weights $\mathbf{w}$ which preserve the information of the original data set $\mathbf{x}$ in the new value y. More formally, such weights would conserve the Fisher information of $\mathbf{x}$. Considering a single parameter of interest, e.g. θ_i, we can express the Fisher information of θ_i in terms of the statistics of y as

$$F_{ii} = \frac{1}{2}\left(\frac{\mathbf{w}^T\mathbf{C}_{,i}\mathbf{w}}{\mathbf{w}^T\mathbf{C}\mathbf{w}}\right)^2 + \frac{\left(\mathbf{w}^T\mu_{,i}\right)^2}{\mathbf{w}^T\mathbf{C}\mathbf{w}}, \tag{8}$$

where the index, i denotes $\partial/\partial\theta_i$. We note that the normalization of the weights is arbitrary (cancels in equation 8).

We select $\mathbf{w}$ that maximizes F_{ii} in equation (8). A general procedure to achieve this is discussed in Tegmark et al. (1997) and Heavens et al. (2000). As is common practice, we perform our analysis for a fixed fiducial covariance matrix (e.g. evaluated from mock catalogues), independent of the model parameters, and therefore assume $\mathbf{C}_{,i} = 0$ and that the information on θ_i is coming only from the second term $\propto \mu_{,i}$. In this case, the unique solution for the weights $\mathbf{w}$ in equation (8) is given by

$$\mathbf{w}^T = \mathbf{C}^{-1}\mu_{,i}. \tag{9}$$

Substituting equation (9) into equation (7), we obtain the relation

$$y = \mathbf{C}^{-1}\mu_{,i}\mathbf{x}. \tag{10}$$

By substituting equation (10) in equation (8), we can see that the Fisher matrix is invariant with respect to $\mathbf{w}$, thus y contains as much information as $\mathbf{x}$ about θ_i (Tegmark et al. 1997).

3.1.2 Optimal weights for multiple parameters

In order to determine multiple parameters from a data set, we need to compress the data set into multiple values to retain the information about the parameters. We specify two equivalent approaches, following Heavens et al. (2000) and Zhao et al. (2019), which lead to the same results.

First, following Heavens et al. (2000), we search for a second number y' that contains the same information as $\mathbf{x}$ about the second parameter θ_j

$$y' = \mathbf{w'}^{\mathrm{T}}\mathbf{x}. \tag{11}$$

If we require y' to be uncorrelated with y, i.e.

$$\mathbf{w'}^{\mathrm{T}}\mathbf{C}\mathbf{w}^{\mathrm{T}} = 0 \tag{12}$$

then, substituting equation (12) into equation (8), we find the solution for $\mathbf{w}'$ to be

$$\mathbf{w}' = \frac{\mathbf{C}^{-1}\mu_{,\mathbf{j}} - (\mu^{\mathrm{T}}_{,j}\mathbf{w})\mathbf{w}}{\sqrt{\mu_{,j}\mathbf{C}^{-1}\mu_{,\mathbf{j}} - (\mu_{,j}\mathbf{w})^2}}. \tag{13}$$

An equivalent formulation of this approach without the Gram–Schmidt orthogonalization is described by Zhao et al. (2019), which defines a derivative matrix

$$\mathbf{D} = \left(\frac{\partial\mu}{\partial\theta_i}, \frac{\partial\mu}{\partial\theta_j}\right) \tag{14}$$

and the multiparameter weights are derived as

$$\mathbf{W} = \mathbf{C}^{-\mathbf{1}}\mathbf{D}, \tag{15}$$

which generalizes equation (9). Both approaches provide lossless compression, leaving the Fisher matrix of the compressed sample equal to the Fisher matrix of the original data set. We compute and test the weights from both methods, confirming that they lead to identical results. Solutions for more than two parameters are also described by Heavens et al. (2000).

3.2 Optimal weights for σ_8

As shown above, the optimal weighting scheme depends on both the parameters of interest and the statistics used in the analysis. As a proof-of-concept, we consider determining the single parameter σ_8 from galaxy clustering and galaxy–galaxy lensing statistics individually, and from their combination.

3.2.1 C_{gg} or $\mathrm{C}_{g\kappa}$ only

We first consider the case of optimal weights for averaging a single statistic at given ℓ over redshift. For C_{gg} the uncompressed data set $\mathbf{x}$

$$\mathbf{x} = \begin{pmatrix} C_{\mathrm{gg}}(\ell, z_1) \\ \vdots \\ C_{\mathrm{gg}}(\ell, z_n) \end{pmatrix} \tag{16}$$

across n redshift bins, is compressed into a new data set y following equation (7). From equation (9) the optimal weights for C_{gg} have the form

$$w_{\mathrm{gg}} = \mathbf{C}^{-1}\partial C_{\mathrm{gg}}/\partial\sigma_8, \tag{17}$$

where $\mathbf{C}$ is the covariance corresponding to $\mathbf{x}$, i.e. between $C_{\mathrm{gg}}(\ell, z_i)$ and $C_{\mathrm{gg}}(\ell, z_j)$, which is a diagonal matrix in the Limber approximation, and

$$\frac{\partial C_{\mathrm{gg}}}{\partial\sigma_8} = \int \mathrm{d}\chi \frac{q^i_{\mathrm{g}}(\chi)q^j_{\mathrm{g}}(\chi)}{\chi^2}\frac{\partial P_{\mathrm{gg}}(\ell/\chi, z(\chi))}{\partial\sigma_8}. \tag{18}$$

Similarly, for a data compression of the $C_{g\kappa}$ power spectrum, we have

$$w_{g\kappa} = \mathbf{C}^{-1}\partial C_{g\kappa}/\partial\sigma_8, \tag{19}$$

where $\mathbf{C}$ is the covariance between $C_{g\kappa}(\ell, z_i)$ and $C_{g\kappa}(\ell, z_j)$, and

$$\frac{\partial C_{g\kappa}}{\partial\sigma_8} = \int \mathrm{d}\chi \frac{q^i_{\kappa}(\chi)q^j_{\mathrm{g}}(\chi)}{\chi^2}\frac{\partial P_{g\kappa}(\ell/\chi, z(\chi))}{\partial\sigma_8}. \tag{20}$$

3.2.2 C_{gg} and $\mathrm{C}_{g\kappa}$ combined

The weights determined in Section 3.2.1 are optimal for individual measurements of C_{gg} or $C_{g\kappa}$ only. Since C_{gg} and $C_{g\kappa}$ are correlated, these weights would not be optimal for data compression of the combined statistics $C_{\mathrm{gg}} + C_{g\kappa}$. In this section, we derive the optimal weights $\mathbf{w}_{\mathrm{gg}+g\kappa}$ when compressing both C_{gg} and $C_{g\kappa}$.

We construct a data vector $\mathbf{x}$ of $2N$ measurements of $C_{\rm gg}(z_i)$ and $C_{\rm g\kappa}(z_i)$, with $i = 1\cdots N$,

$$\mathbf{x} = \begin{pmatrix} C_{\rm gg}(\ell, z_1) \\ \vdots \\ C_{\rm gg}(\ell, z_n) \\ C_{\rm g\kappa}(\ell, z_1) \\ \vdots \\ C_{\rm g\kappa}(\ell, z_n) \end{pmatrix}, \tag{21}$$

and compress this data vector into a number y,

$$y = \mathbf{w}^{\rm T}_{\rm gg+g\kappa}\mathbf{x}. \tag{22}$$

We can derive the optimal weights used in equation (22) following equation (9)

$$\mathbf{w}_{\rm gg+g\kappa} = \mathbf{D}\cdot\mathbf{C}^{-1} \tag{23}$$

with

$$\mathbf{D} = \begin{pmatrix} \partial_{\sigma_8} C_{\rm gg}(\ell, z_1) \\ \vdots \\ \partial_{\sigma_8} C_{\rm gg}(\ell, z_n) \\ \partial_{\sigma_8} C_{\rm g\kappa}(\ell, z_1) \\ \vdots \\ \partial_{\sigma_8} C_{\rm g\kappa}(\ell, z_n) \end{pmatrix} \tag{24}$$

and

$$\mathbf{C} = \begin{pmatrix} \langle C_{\rm gg}(\ell, z_1) C_{\rm gg}(\ell, z_1)\rangle \ldots \langle C_{\rm gg}(\ell, z_1) C_{\rm g\kappa}(\ell, z_n)\rangle \\ \vdots \\ \langle C_{\rm g\kappa}(\ell, z_n)\rangle C_{\rm gg}(\ell, z_1) \ldots \langle C_{\rm g\kappa}(\ell, z_n) C_{\rm g\kappa}(\ell, z_n)\rangle \end{pmatrix}, \tag{25}$$

where these covariance matrix elements may be evaluated using equation (4).

3.3 Optimal weights for multiple parameters (σ_8 and $b_{\rm piv}$)

Combined-probe statistics are valuable for breaking degeneracies between model parameters. In this study, we consider the proof-of-concept of using $C_{\rm gg}$ and $C_{\rm g\kappa}$ to break the degeneracy between the galaxy bias $b_{\rm piv}$ and σ_8, since $C_{\rm gg} \propto b^2_{\rm piv}\sigma^2_8$ while $C_{\rm g\kappa} \propto b_{\rm piv}\sigma^2_8$. Here, we derive the optimal weighting scheme to be applied in this case, following the method described in Section 3.1.2. For simplicity we present only the derivation using the method of Zhao et al. (2019).

We consider the data vector of measurements in Section 3.2.2, of length $2N$, and the covariance matrix (of dimension $2N \times 2N$) from equation (25). We generalize equation (24) to the multiparameter case, by constructing a $(N \times 2)$ matrix of the derivatives $\mathbf{D}$ of the model in each redshift bin with respect to σ_8 and $b_{\rm piv}$,

$$\mathbf{D} = \begin{pmatrix} \partial_{\sigma_8} C_{\rm gg}(\ell, z_1) & \partial_b C_{\rm gg}(\ell, z_1) \\ \vdots \\ \partial_{\sigma_8} C_{\rm gg}(\ell, z_n) & \partial_b C_{\rm gg}(\ell, z_n) \\ \partial_{\sigma_8} C_{\rm g\kappa}(\ell, z_1) & \partial_b C_{\rm g\kappa}(\ell, z_1) \\ \vdots \\ \partial_{\sigma_8} C_{\rm g\kappa}(\ell, z_n) & \partial_b C_{\rm g\kappa}(\ell, z_n) \end{pmatrix} \tag{26}$$

The optimal weight matrix (of dimension $N \times 2$) is calculated from

$$\mathbf{W}_{\rm gg+g\kappa} = \mathbf{C}^{-1}\mathbf{D} \tag{27}$$

using equation (4), and the compressed data set $\mathbf{y}$ now has dimension 2×1,

$$\mathbf{y} = \mathbf{W}^T_{\rm gg+g\kappa}\mathbf{x}. \tag{28}$$

3.4 Individual galaxy weights

In Section 3.2, we derived weights to be applied to the power spectra measured in different redshift bins, compressing them into a single mode containing the same information as the original. These weights can be equivalently applied to individual galaxies, which can be convenient for some analyses (e.g. enabling statistics to be measured across wider redshift intervals).

Once we have determined power spectrum weights $w_{\rm gg}$ and $w_{\rm g\kappa}$ (or the corresponding sections of the total weight vector $w_{\rm gg\,+\,g\kappa}$) for a particular parameter and scale, we can assign these to individual galaxies as $w_{\rm g} = \sqrt{w_{\rm gg}}$ for a clustering measurement and $w_{\rm g} = w_{\rm g\kappa}$ for a galaxy–galaxy lensing measurement. Hence, a galaxy catalogue may contain multiple weights per galaxy, where different weights are used for the measurement of different statistics. This is expected as the optimal weights will always depend on the statistic under consideration. This recipe for applying the weights to individual galaxies has been applied in survey data analysis by e.g. Ruggeri et al. (2019a). Since the weights are expected to vary slowly on the scales of interest for clustering (Zhao et al. 2019), we can choose a single effective scale instead of computing weight for every scale, which would be impractical.

In configuration space, if combining e.g. the shear-galaxy correlation function $\gamma_t(\theta)$ and the galaxy–galaxy angular correlation $w(\theta)$, we can also apply weights to a pair directly, instead of to an individual galaxy (Zhu et al. 2018).

4 RESULTS

4.1 Survey configuration

In this section, we apply the data compression framework derived in Section 3 to joint measurements of galaxy–galaxy lensing and galaxy clustering. We demonstrate that an optimal weighting scheme allows for loss-less compression of the data set, and recovers unbiased parameter constraints.

For demonstration purposes, we construct as test data a set of Gaussian realizations (see Section 4.2) representative of current lensing and clustering surveys (we do not employ N-body simulations as we are interested in a proof-of-concept where data are precisely drawn from models). For the lenses, we assume a homogeneous galaxy sample with a constant number density distribution

$$n_{\rm lens}(z) = 10^{-4}\,h^3\,{\rm Mpc}^{-3} \quad 0.2 < z < 0.7, \tag{29}$$

representative of a Luminous Red Galaxy sample (Eisenstein et al. 2001). We model the redshift distribution of the sources as

$$n_{\rm source}(z)/\overline{n}_{\rm source}(z) \propto z^2\exp\left(-z/z_0\right) \quad 0.1 < z < 3.5, \tag{30}$$

with $z_0 = 1/3$, which is representative of the HSC photometric lensing catalogue (Oguri & Takada 2011). These redshift probability distributions are displayed in Fig. 1. The shape noise for each ellipticity component $\sigma_{\rm e}$, source density, and angular area used for the test data are defined in Table 1. The values chosen are consistent with the HSC data set.

We explored alternative survey configurations, varying the density of the lenses or their redshift range, to investigate the behaviour of the weights for different signal-to-noise ratios. All tests performed

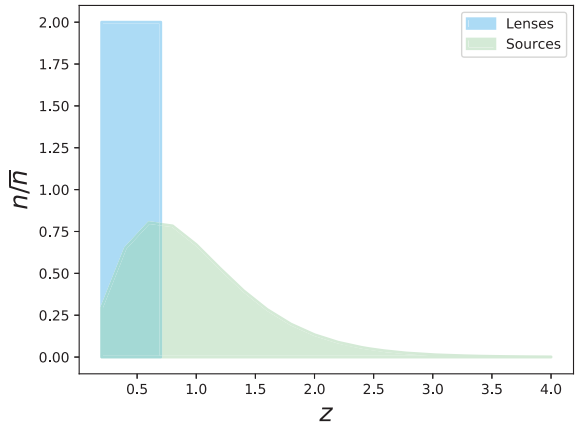

Figure 1. The source and lens redshift probability distribution of our model survey configuration.

Table 1. The survey configuration adopted for our test lensing data set.

Shape noise per component	σ_e	0.28
Source density	$\bar{n}_{\rm source}$	17 arcmin^{-2}
Survey area	Ω_s	1000 deg^2

led to equivalent conclusions, and therefore we limit our discussion to the single survey configuration described here.

4.2 Gaussian realizations

For the test data we use a set of Gaussian realizations of the angular power spectra $C_{\rm gg}$ and $C_{\rm g\kappa}$. We consider one redshift bin for the sources, $0.0 < z < 3.5$, and $N = 5$ redshift bins for the lenses of width $\Delta z = 0.1$ in the range $0.2 < z < 0.7$. For illustrative purposes, we select modes in the range $0 < \ell < 1000$ in bins with $\Delta\ell = 10$ ($\ell = 1000$ corresponds to $k \approx 1.7\,h$ Mpc^{-1} at $z = 0.2$ and $k = 0.6\,h$ Mpc^{-1} at $z = 0.7$).

We assume measurements in different multipole bins to be independent, and for each bin ℓ we compute the $2N \times 2N$ covariance matrix between $C_{\rm gg}(z_i,\ \ell)$ and $C_{\rm g\kappa}(z_j,\ \ell)$ (equation 4), thereby including the correlations between different lens redshift slices. To generate each Gaussian realization we Cholesky-decompose the covariance matrix $\mathbf{C}$ as

$$\mathbf{L}\mathbf{L}^{*} = \mathbf{C}, \tag{31}$$

where $\mathbf{L}$ is a lower triangular matrix with real and positive diagonal entries, and '$*$' denotes the conjugate transpose. The noisy data vector $\mathbf{x}$ of each Gaussian realization for $C_{\rm gg}$ and $C_{\rm g\kappa}$ is then given by

$$\mathbf{x} = \mathbf{L}\mathbf{v} + \mu, \tag{32}$$

where $\mu^{\rm T} = [C_{\rm gg}(z_1) \ldots C_{\rm g\kappa}(z_N)]$, and $\mathbf{v}$ is a random vector of length $2N$, drawn from a normal distribution with mean 0 and unit standard deviation.

Once the test data are created, we fit each realization for either or both of the amplitude parameters (σ_8, b), fixing the other cosmological parameters (the fiducial cosmology considered is listed in Section 2.3). We use a normal chi-squared likelihood method to perform the fit, comparing each data vector $\mathbf{x}$ with the model described in Section 2.1, generated in redshift slices and weighted in the same way as the data. We quantify the errors in the fits using the standard deviation of the best-fitting parameters across 1000 different Gaussian realizations, and use these parameter errors to test the Fisher matrix predictions, the effectiveness of the loss-less data compression and the systematic errors described in Section 5.

4.3 Single parameter fit for σ_8

In this section, we consider results fitting only σ_8, and fixing the other parameters to their fiducial values. In the following section, we will consider joint fits to σ_8 and b. We repeat the parameter fits using three different approaches:

i) weighted analysis,
ii) uncompressed sample analysis,
iii) wide redshift bin analysis.

Method (i) uses the optimal weighting scheme to compress the information in the redshift direction into a single measurement, and method (ii) corresponds to an uncompressed analysis in which the multiple redshift slices are retained and jointly analysed. Method (iii) instead utilizes initial measurements in a single wide redshift bin ($0.2 < z < 0.7$ in this case), without maintaining sensitivity to the redshift evolution across the sample, which we expect to lose information.

For each method (i)–(iii), we consider fitting the amplitude parameters using

a) $C_{\rm gg}$ only,
b) $C_{\rm g\kappa}$ only,
c) the combination of $C_{\rm gg}\ + C_{\rm g\kappa}$,

to investigate how the optimal weights and parameter errors depend on the statistic(s) analysed.

(i) Weighted analysis. We compress the data from each realization by applying the optimal weighting scheme presented in Section 3. We apply the weights to each angular power spectrum $C_{\rm gg}(z_1, \ell)..C_{\rm g\kappa}(z_n, \ell)$ as a function of ℓ, obtaining a single mode $C_{\rm gg\,+\,g\kappa}$ for each ℓ considered. If the compression is lossless, $C_{\rm gg\,+\,g\kappa}$ is expected to carry the same information as the uncompressed statistics $C_{\rm gg}(z_1, \ell)..C_{\rm g\kappa}(z_n, \ell)$. We derive different weights for the different choices of statistics (a), (b), and (c) listed above. As discussed in Section 3, the optimal weights depend on the mean and the covariance matrix of the statistic(s) employed.

ii) Uncompressed analysis. We consider the angular power spectra of all N redshifts slices $C_{\rm gg}(z_i,\ \ell)$, $C_{\rm g\kappa}(z_j,\ \ell)$ and the full covariance between them. No compression of the data or optimal weighting is applied in this approach, and the χ^2 function for the likelihood fitting is given by

$$\chi^2(\ell) = \mathbf{d}^T\mathbf{C}^{-1}\mathbf{d}, \tag{33}$$

for each ℓ, where $d = [C^{\rm D}_{\rm gg}(z_1, \ell) - C^{\rm M}_{\rm gg}(z_1, \ell) \cdots C^{\rm D}_{\rm g\kappa}(z_N, \ell) - C^{\rm M}_{\rm g\kappa}(z_N, \ell)]$, where the superscripts D and M indicate the data and models, respectively. The models for $C_{\rm gg}$ and $C_{\rm g\kappa}$ are described in Section 2.1. We keep the covariance matrix fixed at the fiducial model.

iii) Wide redshift bin analysis. Here, we analyse the data considering a single wide redshift bin for the lens distribution, generating the model and data at a fixed redshift, which we take to be the mean lens redshift $z_c = 0.45$. Otherwise, we perform fits using the same χ^2 likelihood method as described above. Comparing results from this approach with method (i) demonstrates the benefit of optimal (loss-less) compression of the tomographic samples.

Fig. 2 compares the angular power spectra for the analyses (i)–(iii) in the range $1 < \ell < 1000$, and Fig. 3 displays the weights employed as a function of redshift when compressing statistics (a), (b), and (c).

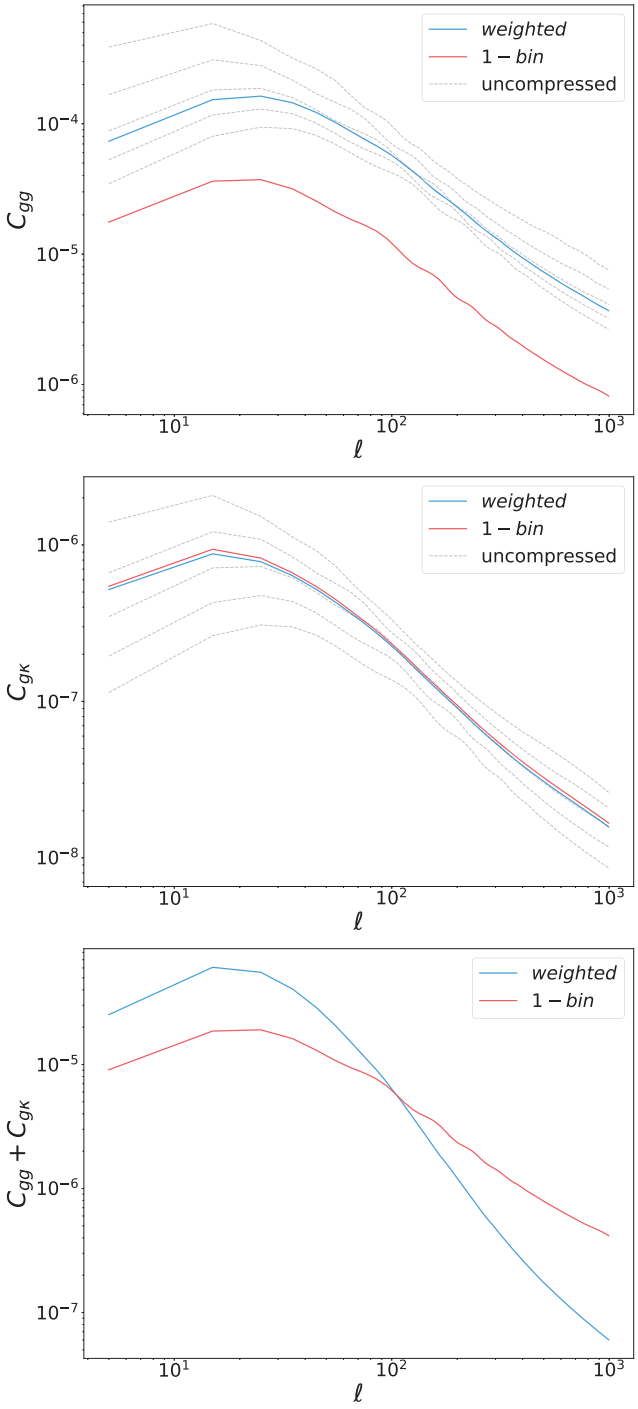

Figure 2. Model galaxy clustering and galaxy–galaxy lensing angular power spectra used in this analysis. Top panel: the galaxy clustering power spectra, C_{gg}. The blue line indicates the weighted, compressed C_{gg}, the red line displays C_{gg} for one wide redshift bin, and the dotted grey lines indicate the values of C_{gg} for each uncompressed redshift slice. Middle panel: the galaxy–galaxy lensing power spectra $C_{g\kappa}$, with the same lines and colours as the top panel. Bottom panel: the summed $C_{gg} + C_{g\kappa}$ power spectra, for the weighted case (blue line) and the wide redshift bin case (red line).

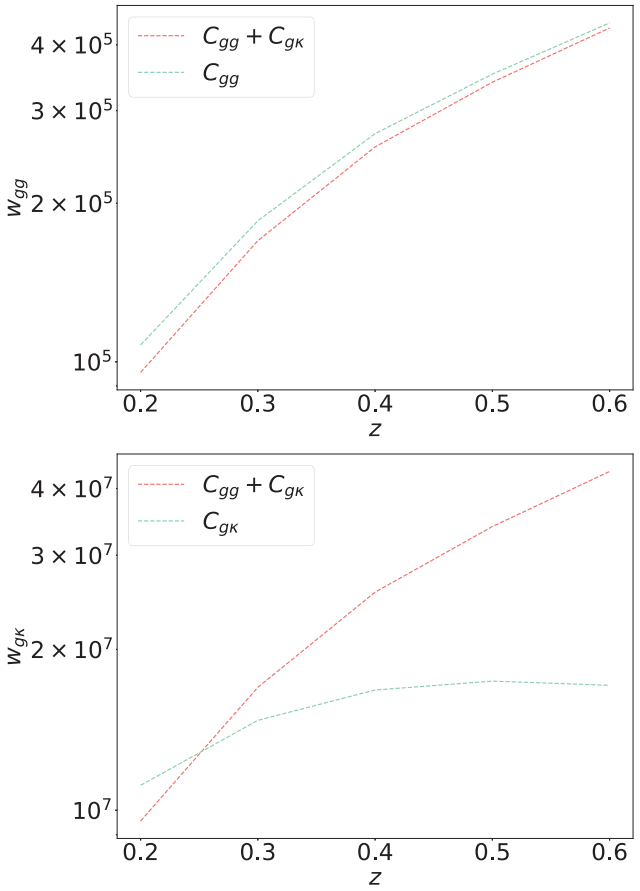

Figure 3. The optimal weights for C_{gg} (top panel) and $C_{g\kappa}$ (bottom panel) for $\ell = 200$, as a function of lens redshift. The red lines indicate the optimal weights when the two power spectra are combined in a joint analysis. The green dashed lines indicate the weights for C_{gg} (top panel) and $C_{g\kappa}$ (bottom panel) when these statistics are considered individually.

We note that the redshift weights applied for each statistic will be different if that statistic is analysed individually, or in combination. We display these weights for $\ell = 200$, which corresponds to the rough location of the linear to non-linear transition at the mean redshift of the lens sample, although the weights show a similar redshift dependence for different ℓ. The weighting scheme does not depend on the normalization (as seen in Section 3), thus a convenient normalization is set for the comparison.

For the C_{gg} statistic, the redshift weights do not vary significantly between cases (a) and (c). For $C_{g\kappa}$, we notice a stronger redshift dependence of the weights when moving from case (b) to (c). This is due to the C_{gg} terms in the covariance matrix in case (c), enhancing the redshift sensitivity.

Fig. 4 presents the errors in σ_8 obtained for the analyses: (i) weighted data and (ii) uncompressed data, for the three different choices of statistics (cases a–c). For each of these cases we plot the mean and standard deviation of the best-fitting σ_8 parameters for each of the 1000 Gaussian realization. All cases considered provide an unbiased estimation of σ_8.

The analyses of the compressed data sets provide parameter errors that are comparable to those obtained in the corresponding uncompressed analyses for all cases (a)–(c), confirming that the compression is loss-less.

4.4 Multiparameter fit

We now consider jointly fitting σ_8 and the bias parameter $b_{\rm piv}$ defined in equation (6) to our data sets. We again compare results using the

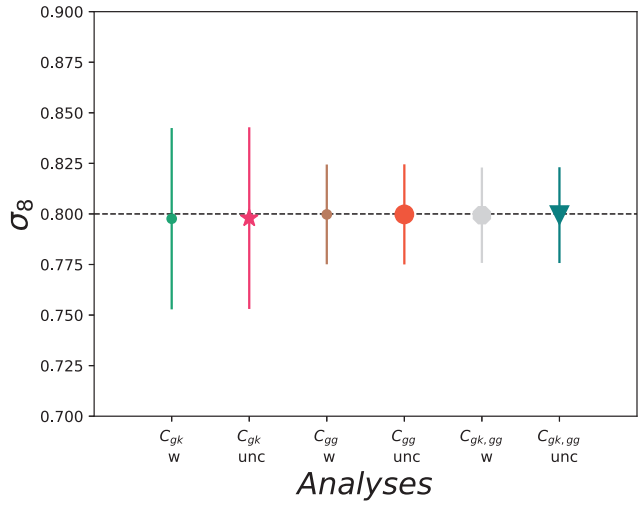

Figure 4. Comparison of the mean and standard deviation of the best fits of a single parameter σ_8 to the individual or jointly analysed galaxy clustering and galaxy–galaxy lensing power spectrum, C_{gg} and $C_{g\kappa}$, for analyses of weighted and compressed data ('w') and uncompressed data ('unc').

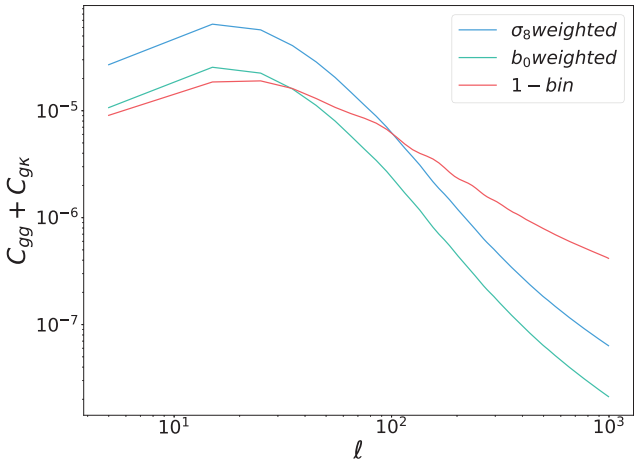

Figure 5. The sum of the C_{gg} and $C_{g\kappa}$ power spectra, comparing cases applying optimal weighting for σ_8 (blue line) and bias (green line) and a single wide redshift bin without optimal weighting (red line).

weighted, compressed data sets, the uncompressed measurements, and wide redshift bin analysis, similarly to Section 4.3.

i) **Optimal weight compression for two parameters.** We use the optimal weighting scheme discussed in Section 3.3 to define weights corresponding to σ_8 and $b_{\rm piv}$ and hence compress the data and models of $C_{gg}(z_i)$, $C_{g\kappa}(z_j)$ into two power spectra, which we jointly analyse.

ii) **Uncompressed analysis.** We again consider the angular power spectra for all N redshift slices and their covariance when computing the χ^2 statistic, adding $b_{\rm piv}$ as a free parameter in the fits.

iii) **Wide redshift bin.** We fit the data in one wide redshift bin, constructing our models at fixed $z = z_c$ as in Section 4.3. We fit for $\sigma_8(z_c)$ and $b_{\rm piv}(z_c)$, considering our model constant over redshift while the data are constructed from a model containing an evolving bias. In this way, we are constructing a test for the systematic error associated with a discrepancy between the assumed and fiducial bias evolution (see Section 5 for more details).

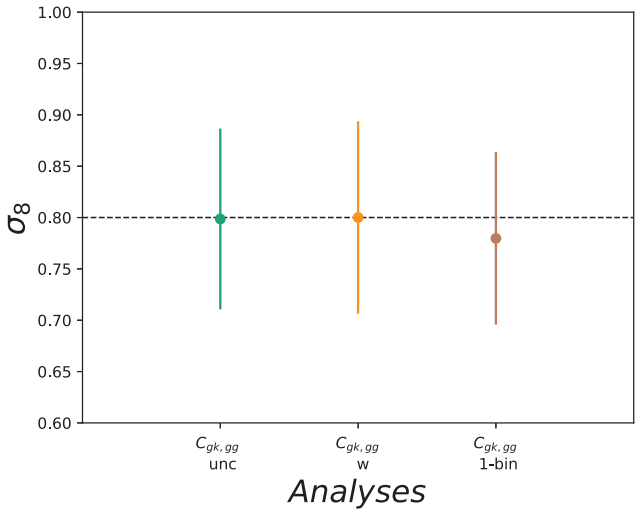

Figure 6. The mean and standard deviation of the best fits of σ_8 for a joint fit of σ_8 and the bias parameter to the combined C_{gg} and $C_{g\kappa}$ data set. We compare analyses of weighted and compressed data ('w'), uncompressed data ('unc'), and in a wide redshift bin ('1-bin').

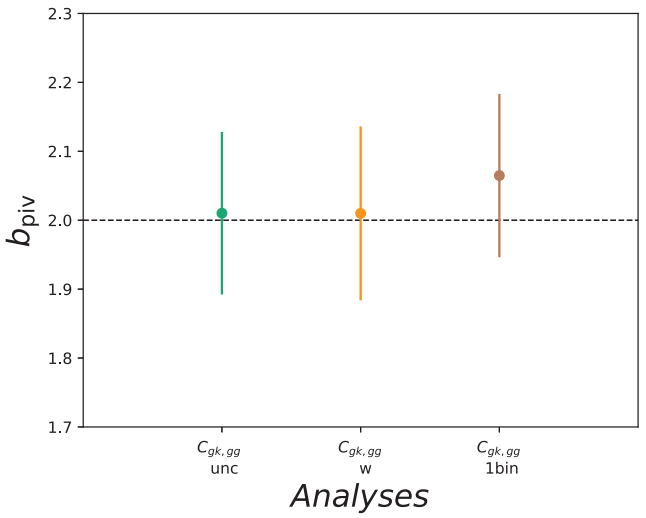

Figure 7. The mean and standard deviation of the best fits for the bias parameter for a joint fit of σ_8 and the bias to the combined C_{gg} and $C_{g\kappa}$ data set. We compare analyses of weighted and compressed data ('w'), uncompressed data ('unc'), and in a wide redshift bin ('1-bin').

Fig. 5 compares the angular power spectra for the analyses (i) and (iii) in the range $1 < \ell < 1000$. We compare the two weighted power spectra: w_{σ_8} (blue line) and $w_{\rm b}$ (green line) with the wide-bin angular power spectrum (red line). The σ_8 and $b_{\rm piv}$ weighting schemes produce similar results with different amplitude.

Figs 6 and 7 present the comparison between σ_8 and bias parameter fits for methods (i), (ii), and (iii), for the multiparameter fit. As for the single-parameter fits, the weighting and uncompressed analysis recover unbiased estimates of σ_8. When a wide redshift bin is used, we find a systematic error in the recovered parameters due to the discrepancy in the assumed bias evolution (see Section 5 for more details). Fig. 8 shows the distribution of the best-fitting values for σ_8 and $b_{\rm piv}$ across the 1000 realisations for optimally weighted (green contour) and uncompressed (red contour) analyses across the 1000 Gaussian realizations. As expected the two analyses show a very close agreement.

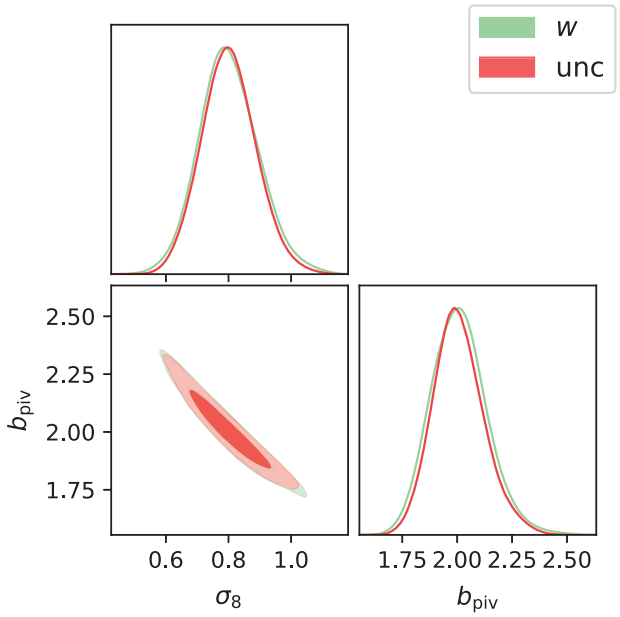

Figure 8. Distribution of the best-fitting values for σ_8 and $b_{\rm piv}$ for optimally weighted (green contour) and uncompressed (red contour) analyses across the 1000 Gaussian realizations.

5 AMPLITUDE SYSTEMATICS

The specification of the optimal redshift weights in Section 3 depends on the redshift evolution of the amplitudes of $C_{\rm gg}$ and $C_{\rm g\kappa}$, which may not be known in advance. Further, this redshift evolution can imprint systematic errors into parameter fits if not correctly modelled, given that it influences the signal contributed by each lens galaxy.

This issue arises due to the fact that $C_{\rm gg}$ and $C_{\rm g\kappa}$ (equation 1) have different redshift kernels $q_{\rm g}^2$ and $q_{\rm g}q_\kappa$. Therefore, a bias factor that evolves with redshift affects the amplitude of $C_{\rm gg}$ and $C_{\rm g\kappa}$ differently at the same redshift. As a simple illustration of the effect, suppose we consider a wide redshift bin and, neglecting the redshift evolution within, generate the power spectra in equation (1) at a single effective redshift $z_{\rm eff}$,

$$C_{\rm gg}(\ell) = b^2 P_{\rm mm}(\ell/\chi_{\rm eff}, z_{\rm eff}) \int {\rm d}z\, q_{\rm g}^2(z), \tag{34}$$

$$C_{\rm g\kappa}(\ell) = b P_{\rm mm}(\ell/\chi_{\rm eff}, z_{\rm eff}) \int {\rm d}z\, q_{\rm g}(z)\, q_{\rm k}(z). \tag{35}$$

where $P_{\rm mm}(k, z)$ is the matter power spectrum. Here, we assumed a single bias parameter for which $P_{\rm gg} = b^2\, P_{\rm mm}$ and $P_{\rm g\kappa} = b\, P_{\rm mm}$. If we now compare equation (34) with equation (1) we find that the galaxy bias amplitude from $C_{\rm gg}$ is given by

$$b^2(C_{\rm gg}) = \frac{\int {\rm d}z\, q_{\rm g}^2(z)\, P_{\rm mm}(\ell/\chi, z)}{P_{\rm mm}(\ell/\chi_{\rm eff}, z_{\rm eff}) \int {\rm d}z\, q_{\rm g}^2(z)}. \tag{36}$$

Similarly for $C_{\rm g\kappa}$ we have

$$b(C_{\rm g\kappa}) = \frac{\int {\rm d}z\, q_{\rm g}(z)\, q_\kappa(z)\, P_{\rm mm}(\ell/\chi, z)}{P_{\rm mm}(\ell/\chi_{\rm eff}, z_{\rm eff}) \int {\rm d}z\, q_{\rm g}(z)\, q_\kappa(z)}. \tag{37}$$

In general $b(C_{\rm gg}) \neq b(C_{\rm g\kappa})$, therefore fitting the galaxy bias factor from the combination of $C_{\rm g\kappa}$ and $C_{\rm gg}$ would produce a systematic multiplicative error in the amplitude.

We quantify this systematic error by comparing fits to the uncompressed and weighted/compressed data with a wide-redshift bin analysis where the bias evolution is neglected. These results are presented in Figs 6 and 7, demonstrating how a redshift weights analysis gives unbiased results, consistent with the uncompressed sample analysis.

To mitigate this systematic and compress the data we first need to obtain a model for the evolution of the bias. One approach is to divide the sample into narrow redshift bins and fit for $b(z)$, then compress the narrow bins into a single measurement using redshift weights. An alternative strategy, following Ruggeri et al. (2019a), is to introduce a free functional form for the galaxy bias (e.g. a Taylor expansion). We can then set up an iterative process, computing the first set of weights for fiducial bias parameters, fitting the parameters, and then re-generating the weights. We note here that incorrect weights are expected to cause sub-optimality, but not bias, in the resulting parameter fits.

6 CONCLUSIONS

As cosmology transits from a data-starved science to a data-driven discipline, developing new strategies to handle the upcoming big-data volumes is a key requirement. In this work, we presented a proof-of-concept of an efficient approach for combining galaxy–galaxy lensing and galaxy clustering probes across a wide redshift range in an optimal way, compressing the data set with no loss of information. We considered just the amplitude parameters in this study, but the work could be extended to other cosmological parameters. We derived a set of weights to constrain the galaxy bias and σ_8, to be applied to the angular power spectra $C_{\rm gg}$ and $C_{\rm g\kappa}$ (and which may alternatively be applied to individual galaxies). We test the weights on a set of Gaussian realizations mimicking the lens and source distributions of representative surveys. We compared the weighted analysis with the uncompressed data sets to demonstrate that the weights carry the same information as the original data set. Finally, we discussed how to handle potential systematic errors associated with evolution in redshift of the galaxy bias. The next step in this work is to apply the methodology to full mock catalogues and survey data samples.

ACKNOWLEDGEMENTS

This research was funded by the Australian Government through Australian Research Council Discovery Project DP160102705. We thank Prof. Alexie Leauthaud and the University of California Santa Cruz for valuable discussions and hospitality during the completion of this work. We also thank Prof. Alan Heavens for helpful suggestions.

DATA AVAILABILITY

The clustering and lensing measurements underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abbott T. M. C. et al., 2018, ApJS, 239, 18
Aihara H. et al., 2018, PASJ, 70, S4
Bellini E., van Waerbeke L., Joudaki S., Alonso D., 2019, Open J. Astrophys., 2, E11
Bernstein G. M., 2009, ApJ, 695, 652
Castorina E. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 010
de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, Exp. Astron., 35, 25
de Jong R. S. et al., 2012, McLean I. S., Ramsay S. K., Takami H., Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460T

Eisenstein D. J. et al., 2001, AJ, 122, 2267
Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965
Hu W., Jain B., 2004, Phys. Rev. D, 70, 043009
Ivezić Ž. et al., 2019, ApJ, 873, 111
Joachimi B., Bridle S. L., 2010, A&A, 523, A1
Krause E., Eifler T., 2017, MNRAS, 470, 2100
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lemos P., Challinor A., Efstathiou G., 2017, J. Cosmol. Astropart. Phys., 2017, 014
Levi M. et al., 2019, BAAS, 51, 57
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Mootoovaloo A., Heavens A. F., Jaffe A. H., Leclercq F., 2020, MNRAS, 497, 2213
Oguri M., Takada M., 2011, Phys. Rev. D, 83, 023008
Ruggeri R., Percival W. J., Gil-Marín H., Zhu F., Zhao G.-B., Wang Y., 2017, MNRAS, 464, 2698
Ruggeri R. et al., 2019a, MNRAS, 483, 3878
Ruggeri R., Percival W. J., Mueller E.-M., Gil-Marín H., Zhu F., Padmanabhan N., Zhao G.-B., 2019b, MNRAS, 484, 4100
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Yoo J., Seljak U. c. v., 2012, Phys. Rev. D, 86, 083504
Zhao G.-B. et al., 2019, MNRAS, 482, 3497
Zhu F. et al., 2018, MNRAS, 480, 1096

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