



**HAL**  
open science

## Efficient Computation of Super-Sample Covariance for Stage IV Galaxy Surveys

Fabien Lacasa, Marie Aubert, Philippe Baratta, Julien Carron, Adélie Gorce, Sylvain Gouyou Beauchamps, Louis Legrand, Azadeh Moradinezhad Dizgah, Isaac Tutusaus

► **To cite this version:**

Fabien Lacasa, Marie Aubert, Philippe Baratta, Julien Carron, Adélie Gorce, et al.. Efficient Computation of Super-Sample Covariance for Stage IV Galaxy Surveys. *Astronomy and Astrophysics - A&A*, 2023, 671, pp.A115. 10.1051/0004-6361/202245148 . hal-03815294

**HAL Id: hal-03815294**

**<https://hal.science/hal-03815294>**

Submitted on 9 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Efficient computation of the super-sample covariance for stage IV galaxy surveys

Fabien Lacasa<sup>1,2,3,4</sup>, Marie Aubert<sup>5</sup>, Philippe Baratta<sup>6</sup>, Julien Carron<sup>3</sup>, Adélie Gorce<sup>7</sup>,  
Sylvain Gouyou Beauchamps<sup>6</sup>, Louis Legrand<sup>3</sup>, Azadeh Moradinezhad Dizgah<sup>3</sup>, and Isaac Tutusaus<sup>8,3</sup>

<sup>1</sup> Dipartimento di Fisica e Astronomia “Galileo Galilei”, Università degli Studi di Padova, via Marzolo 8, 35131 Padova, Italy

<sup>2</sup> INFN, Sezione di Padova, via Marzolo 8, 35131 Padova, Italy  
e-mail: [fabien.lacasa@pd.infn.it](mailto:fabien.lacasa@pd.infn.it)

<sup>3</sup> Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, 24 quai Ernest Ansermet, 1211 Geneva, Switzerland

<sup>4</sup> Université Paris-Saclay, CNRS, Institut d’Astrophysique Spatiale, 91405 Orsay, France

<sup>5</sup> University of Lyon, UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, France

<sup>6</sup> Aix-Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France

<sup>7</sup> Department of Physics and Trotter Space Institute, McGill University, Montreal, QC H3A2T8, Canada

<sup>8</sup> Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France

Received 5 October 2022 / Accepted 11 January 2023

## ABSTRACT

Super-sample covariance (SSC) is an important effect for cosmological analyses that use the deep structure of the cosmic web; it may, however, be nontrivial to include it practically in a pipeline. We solve this difficulty by presenting a formula for the precision (inverse covariance) matrix and show applications to update likelihood or Fisher forecast pipelines. The formula has several advantages in terms of speed, reliability, stability, and ease of implementation. We present an analytical application to show the formal equivalence between three approaches to SSC: (i) at the usual covariance level, (ii) at the likelihood level, and (iii) with a quadratic estimator. We then present an application of this computationally efficient framework for studying the impact of inaccurate modelling of SSC responses for cosmological constraints from stage IV surveys. We find that a weak-lensing-only analysis is very sensitive to inaccurate modelling of the scale dependence of the response, which needs to be calibrated at the  $\sim 15\%$  level. The sensitivity to this scale dependence is less severe for the joint weak-lensing and galaxy clustering analysis (also known as  $3\times 2$ pt). Nevertheless, we find that both the amplitude and scale-dependence of the responses have to be calibrated at better than 30%.

**Key words.** large-scale structure of Universe – methods: analytical – methods: statistical – surveys

## 1. Introduction

Super-sample covariance (SSC) (Hu & Kravtsov 2003; Takada & Hu 2013), also known as sample variance for cluster counts, is an important part of statistical errors or cosmic variance for the current survey of the large-scale structure (LSS) of the Universe (e.g., Hildebrandt et al. 2017), and even more for near future cosmic surveys that aim to deepen our understanding of dark energy and gravity. Specifically, it is expected to be the dominant non-Gaussian contribution to the covariance of weak-lensing statistics (Barreira et al. 2018a), although other covariance terms are also important for galaxy clustering (Lacasa 2018, 2020). It has been shown that for a stage IV survey, the dark energy figure of merit can be degenerate by a factor of two (Gouyou Beauchamps et al. 2022) when the SSC is accounted for in the joint analysis of weak-lensing (WL), photometric galaxy clustering (GC). This joint analysis is also called  $3\times 2$ pt analysis. However, it is difficult to estimate the SSC because it cannot be calibrated reliably with jackknife or bootstrap methods or simulations with sizes comparable to the survey size (Lacasa & Kunz 2017). Instead, significant progress has recently been made in the analytical modelling of the SSC (e.g., Takada & Spergel

2014; Takahashi et al. 2014; Li et al. 2018; Chan et al. 2018; Lacasa et al. 2018; Wadekar & Scoccimarro 2020) as well as in semi-analytical predictions of the SSC using measurements of the simulated reaction of observables to super-survey modes (e.g., Barreira & Schmidt 2017; Barreira et al. 2018b). An alternative approach that includes the effect of super-survey modes is to account for them with additional nuisance parameters that are marginalised over, that is, to account for them at the likelihood level and not in the covariance (Li et al. 2014; Lacasa & Grain 2019). In this article, we examine the two approaches to the modelling of the impact of super-survey modes (see Sect. 2).

We address four challenges in implementing the SSC in an analysis pipeline: change, speed, reliability, and stability. The first challenge, change, is the difficulty of changing an existing pipeline that may use assumptions that are broken by the SSC (e.g., that the covariance matrix is diagonal) and comparing new results with the previous pipeline version. The second challenge is that the covariance pipeline may be slowed down by the time necessary to predict the SSC and by the inversion of the total covariance matrix (if large), which is no longer diagonal. This challenge is not relevant when the covariance is fixed, but when it is varied with parameters and/or when the robustness of the results is tested with respect to changes in the covariance. The

third challenge is about the stability of the results with respect to variations in the theoretical ingredients of the SSC prediction because the SSC literature is quite technical and evolves rapidly. The fourth challenge is a practical one found by the authors: the SSC gives an ill-conditioned covariance matrix so that the inversion becomes unstable when the other covariance term becomes negligible (e.g., when an analysis is pushed in the deep non-linear regime or in a combined probe case with large matrices). In the first section of this article, we propose a framework that solves these four problems. We then develop theoretical and numerical applications of the framework.

In detail, the article is organised as follows. In Sect. 2 we show that SSC can be computed as an update to the precision (i.e. inverse covariance) matrix, present examples for different probes, and show the application in likelihood or Fisher pipelines. In Sect. 3 we use the previous formulae to show the equivalence of three approaches to SSC: at the covariance level, at the likelihood level, and with a quadratic estimator. In Sect. 4 we apply the framework to a mock stage IV photometric galaxy survey, for WL alone or in combination with GC. We conclude in Sect. 5.

## 2. SSC as an update to the precision matrix

In this section, we show that the inverse covariance matrix can be computed as an update to the noSSC case in a numerically efficient manner. We then show applications to compute the likelihood or the Fisher matrix of an LSS observable.

To do this, we use the Woodbury identity for matrix inversion (Woodbury 1950). This identity states that if  $A$  is an invertible  $(n, n)$  matrix of known inverse that is updated with a matrix of the form  $UCV$ , with  $U, C$ , and  $V$  the matrices of the respective shapes  $(n, m)$ ,  $(m, m)$ , and  $(m, n)$ , and  $C$  is invertible, then the inverse of the updated matrix is given by

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \quad (1)$$

More generally, even if  $C$  is not invertible, we have

$$(A + UCV)^{-1} = A^{-1} - A^{-1}UC(\mathbb{1} + VA^{-1}UC)^{-1}VA^{-1}. \quad (2)$$

We use the general form of the Woodbury identity given by Eq. (2) for the case that  $A$  and  $C$  are symmetric matrices and  $V = U^T$ . The original and updated matrices are therefore both symmetric.

It is not immediately obvious that the right-hand side of Eq. (2) is also symmetric, as it should be. However, this becomes clearer when  $C$  is invertible, where we can use the equivalent form in Eq. (1), which shows that the right-hand side is symmetric. Then we can note that matrix symmetry is a continuous property, and invertible matrices are dense in matrix space. The property therefore holds by continuity even if  $C$  is not invertible.

### 2.1. General formula

Let  $O(i, \alpha)$  be an abstract LSS probe, where  $\alpha$  is the index (multi-index) for the redshift bin (redshift tuple), and  $i$  is the index for all other quantities on which the observable depends (e.g., Fourier or angular wavevector, distance or angular separation, mass, size, or peak height bin). This observable is vectorised through an index  $i_o$ , with a mapping  $i_o \rightarrow (i, \alpha)$  that is considered implicit hereafter.

We call  $n_z$  be the number of indices  $\alpha$  and  $n_o$  the number of indices  $i_o$ , that is, the size of the observable. For instance, if there

is the same number  $n_i$  of quantities for each redshift bin, then we have  $n_o = n_i \times n_z$ . This situation can, however, be broken by analysis choices employing a redshift-dependent range for the  $i$  quantities, for example a cut  $k_{\max}(z)$  or  $\ell_{\max}(z)$ .

We use the Sij approximation to SSC (e.g., Hu & Kravtsov 2003 for cluster counts, Lacasa & Grain 2019 for power spectra and other observables). This is a mean-field approximation for the redshift evolution, which is expected valid for observables with compact support in redshift and has been validated for use for cosmological analysis with clusters (Hu & Kravtsov 2003; Agueña & Lima 2018) or galaxies (Lacasa & Grain 2019). In this approximation, the SSC of the probe is given by

$$\text{Cov}_{\text{SSC}}[O(i, \alpha), O(j, \beta)] = \partial_{\delta_b} O(i, \alpha) \partial_{\delta_b} O(j, \beta) S_{\alpha, \beta}, \quad (3)$$

where  $\delta_b = (\rho_{\text{survey}} - \bar{\rho})/\bar{\rho}$  is the change in matter density in the survey with respect to the global mean. This is also known as the super-survey mode. The  $S$  matrix is the covariance matrix of these  $\delta_b$  in each redshift bin, due to the clustering of matter.  $\partial_{\delta_b} O$  is the reaction of the probe  $\delta_b$ , also known as the response. In practice, this response is computed either from a nonlinear theory (e.g., Takada & Hu 2013; Lacasa & Rosenfeld 2016), from an ansatz (Lacasa & Grain 2019), or from simulations (Barreira et al. 2018b).

In the following we often shorten  $C_{\text{SSC}} \equiv \text{Cov}_{\text{SSC}}$ , that is,

$$\begin{aligned} (C_{\text{SSC}})_{i_o, j_o} &= \text{Cov}_{\text{SSC}}[O(i_o), O(j_o)] \\ &= \text{Cov}_{\text{SSC}}[O(i, \alpha), O(j, \beta)], \end{aligned} \quad (4)$$

where  $i_o \rightarrow (i, \alpha)$  and  $j_o \rightarrow (j, \beta)$  through the implicit mapping mentioned previously.

We now remark that this SSC matrix can be cast in the following form:

$$C_{\text{SSC}} = USU^T, \quad (5)$$

where  $S$  is a square matrix of shape  $(n_z, n_z)$  (making it evident that  $C_{\text{SSC}}$  has rank  $\leq n_z$ ), and  $U$  is a rectangular matrix of shape  $(n_o, n_z)$  given by

$$U_{i_o, \beta} = \partial_{\delta_b} O(i, \alpha) \delta_{\alpha, \beta}, \quad (6)$$

where the final  $\delta$  on the right-hand side is a Kronecker symbol.

Therefore, using the Woodbury identity in Eq. (2), the inverse of the total covariance matrix  $C_{\text{tot}} = C_{\text{noSSC}} + C_{\text{SSC}}$  is given by

$$C_{\text{tot}}^{-1} = C_{\text{noSSC}}^{-1} - C_{\text{noSSC}}^{-1}US(\mathbb{1} + I(\delta_b)S)^{-1}U^T C_{\text{noSSC}}^{-1}, \quad (7)$$

where we introduced the notation (that we justify below)

$$I(\delta_b) \equiv U^T C_{\text{noSSC}}^{-1}U. \quad (8)$$

Formula Eq. (7) thus gives the total precision matrix as an update to the noSSC case. It is particularly useful numerically when the SSC dominates the covariance matrix elements: the SSC matrix indeed has a low rank and will thus destabilise the covariance inversion. This was, for instance, noted previously by one of the authors in numerical investigations pushing angular power spectra analyses at high multipoles in the SSC-dominated regime (Lacasa 2020).

### 2.2. Examples

We now apply the general formula described in the previous subsection to specific class of LSS observables. The case of one-point functions is described in Sect. 2.2.1, the case of two-point functions is set out in Sect. 2.2.2, and the case of other statistics is presented in Sect. 2.2.3.

### 2.2.1. Counts or one-point functions

The first class of observables includes counts of clusters as a function of (bins of) mass proxy (and redshift), of voids as a function of size, or of WL peaks as a function of peak height. In all cases, the observable can be noted  $N_{\text{obj}}(i, \alpha)$ : the number of objects as a function of the index  $i$  of quantity (mass, size, and height) and redshift index  $\alpha$ . For clusters and voids, and except for factors of confusion in bin assignment (e.g., mass errors or photo- $z$  errors), the noSSC (=standard) covariance is Poissonian,

$$\text{Cov}_{\text{noSSC}} [N_{\text{obj}}(i, \alpha), N_{\text{obj}}(j, \beta)] = N_{\text{obj}}(i, \alpha) \delta_{i,j} \delta_{\alpha,\beta}, \quad (9)$$

where the final  $\delta$  on the right-hand side are Kronecker symbols, that is, the noSSC matrix is diagonal. This diagonality somewhat simplifies Eq. (7),

$$\begin{aligned} C_{\text{tot}}^{-1} [N_{\text{obj}}(i, \alpha), N_{\text{obj}}(j, \beta)] &= N_{\text{obj}}^{-1}(i, \alpha) \delta_{i,j} \delta_{\alpha,\beta} \\ &- \sum_{\gamma} [\partial_{\delta_b} N_{\text{obj}}(i, \alpha)] N_{\text{obj}}^{-1}(i, \alpha) \times S_{\alpha,\gamma} \times (\mathbb{1} + I(\delta_b) S)_{\gamma,\beta}^{-1} \\ &\times [\partial_{\delta_b} N_{\text{obj}}(j, \beta)] N_{\text{obj}}^{-1}(j, \beta), \end{aligned} \quad (10)$$

which can be written in a more compact form as

$$\begin{aligned} C_{\text{tot}}^{-1} [(i, \alpha), (j, \beta)] &= \frac{\delta_{i,j} \delta_{\alpha,\beta}}{N_{\text{obj}}(i, \alpha)} \\ &- \partial_{\delta_b} \ln N_{\text{obj}}(i, \alpha) (S (\mathbb{1} + I(\delta_b) S)^{-1})_{\alpha,\beta} \partial_{\delta_b} \ln N_{\text{obj}}(j, \beta). \end{aligned} \quad (11)$$

Here, the matrix  $I(\delta_b)$  has the shape  $(n_z, n_z)$ , and is given by

$$I(\delta_b)_{\alpha,\beta} = \left( \sum_i [\partial_{\delta_b} N_{\text{obj}}(i, \alpha)]^2 N_{\text{obj}}^{-1}(i, \alpha) \right) \delta_{\alpha,\beta}. \quad (12)$$

We note that  $\mathbb{1} + I(\delta_b) S$  is not symmetric (but  $S \cdot (\mathbb{1} + I(\delta_b) S)^{-1}$  is) and must be inverted numerically. This is not a problem because (i) it is regular (even if  $S$  is not invertible, the total is regularised by the presence of the identity matrix), and (ii) it has a small size: the number of redshift bins.

### 2.2.2. Power spectra or two-point functions

The second class of observable includes the (angular) power spectra or the (angular) correlation function of galaxies, weak-lensing, clusters, cosmic microwave background lensing, HI intensity mapping, and so on as a function of Fourier (harmonic) wave number or (angular) separation and of redshift bin. If we only consider auto-spectra (i.e. correlate only in the same redshift bin), then  $\alpha$  is the index of the redshift bin. If we also consider cross-spectra, then  $\alpha$  is the index of the redshift pair. Even in the noSSC auto-spectrum case, there is often cross-talk between redshift bins, for example because of photo- $z$  errors, so that the covariance is not diagonal in this direction. However, the noSSC covariance is diagonal in terms of Fourier or harmonic wave number under two conditions: (i) it may be argued that the Gaussian covariance is the dominant term beyond SSC (Barreira et al. 2018a showed this is the case for WL, although Lacasa 2020 reported that other non-Gaussian terms are important for constraints on  $n_s$  with GC), and (ii) the wave numbers are binned sufficiently to erase the mode-coupling due to the survey mask (see e.g., Hivon et al. 2002). As for counts, this slightly simplifies Eq. (7), although less so because we still have redshift correlations. For example, for an angular power spectrum  $C_\ell(\alpha)$ ,

the Gaussian covariance is a bloc diagonal multipole by multipole. We call  $\Sigma_\ell^2$  the multipole bloc with the shape  $(n_z, n_z)$ ,

$$\Sigma_\ell^2(\alpha, \beta) = \text{Cov}(C_\ell(\alpha), C_\ell(\beta)), \quad (13)$$

and denote its inverse with  $\Sigma_\ell^{-2}$ . We thus have

$$\begin{aligned} C_{\text{tot}}^{-1} [C_\ell(\alpha), C_{\ell'}(\beta)] &= \Sigma_\ell^{-2}(\alpha, \beta) \delta_{\ell,\ell'} \\ &- \sum_{\gamma, \epsilon, \zeta} \partial_{\delta_b} C_\ell(\gamma) \Sigma_\ell^{-2}(\gamma, \alpha) S_{\gamma,\epsilon} (\mathbb{1} + I(\delta_b) S)_{\epsilon,\zeta}^{-1} \\ &\times \partial_{\delta_b} C_{\ell'}(\zeta) \Sigma_{\ell'}^{-2}(\zeta, \beta), \end{aligned} \quad (14)$$

which can be written in a more compact form as

$$C_{\text{tot}}^{-1} = C_{\text{Gauss}}^{-1} - N_\ell^T S (\mathbb{1} + I(\delta_b) S)^{-1} N_\ell, \quad (15)$$

where the matrix  $N_\ell$  has the shape  $(n_z, n_z)$  and we have

$$N_\ell(\alpha, \beta) = \partial_{\delta_b} C_\ell(\alpha) \Sigma_\ell^{-2}(\alpha, \beta). \quad (16)$$

Here, the matrix  $I(\delta_b)$  has the shape  $(n_z, n_z)$  and is given by

$$I(\delta_b)_{\alpha,\beta} = \sum_\ell \partial_{\delta_b} C_\ell(\alpha) \Sigma_\ell^{-2}(\alpha, \beta) \partial_{\delta_b} C_\ell(\beta). \quad (17)$$

Equation (15) greatly decreases the numerical cost of inverting  $C_{\text{tot}}$ . For the covariance matrix of shape  $(n_\ell \times n_z, n_\ell \times n_z)$ , the cost of brute force computation is  $\mathcal{O}((n_\ell \times n_z)^3)$ , while through Eq. (15), we invert only (many)  $(n_z, n_z)$  matrices, reducing the cost to  $\mathcal{O}(n_\ell \times n_z^2)$ . Therefore, we achieve a speedup factor of  $n_\ell^2$ . For a typical analysis with tens of multipole bins (e.g., we use 40 bins in Sect. 4), this results in an appreciable speedup factor in the hundreds. This speedup means that a larger number of multipole bins can be used instead of binning aggressively to extract as much information as possible from the measurement, for instance, to hunt for localised features.

### 2.2.3. Other statistics

The formulae can easily be generalised to other statistics. For instance, a combination of angular power spectrum measurements of different probes amounts to generalising the  $\alpha$  multi-index to index both probes and redshift tuples. We use this in Sect. 4 for the so-called 3×2pt analysis, which combines WL, GC, and their cross correlation (XC). Higher-order statistics are also possible. For instance, for a bispectrum analysis,  $i$  would become a multi-index for triplets of wavenumbers, and  $\alpha$  becomes a multi-index for triplets of redshift bins.

In all cases, the formula becomes particularly useful when the noSSC covariance matrix is diagonal in some direction, as it can leverage this to lower the numerical cost of the inversion. For instance, for a combination of power spectra, the Gaussian covariance only correlates with identical wavenumbers, and for bispectra, it only correlates with identical wavenumber triplets.

## 2.3. Applications

### 2.3.1. Log likelihood

For a standard Gaussian likelihood with fixed covariance, we have

$$-2 \ln \mathcal{L} = (d - m)^T C_{\text{tot}}^{-1} (d - m), \quad (18)$$

where  $d$  is the data vector and  $m$  is the corresponding model.

Adding SSC can thus be considered as an update to the log-likelihood:  $\ln \mathcal{L} = \ln \mathcal{L}_{\text{noSSC}} + \delta \ln \mathcal{L}$  with

$$-2\delta \ln \mathcal{L} = X^T S (\mathbb{1} + I(\delta_b) S)^{-1} X, \quad (19)$$

where  $X = U^T C_{\text{noSSC}}^{-1} (d - m)$ . This makes it easy and numerically efficient to update a noSSC pipeline to account for SSC. In the noSSC pipeline, the covariance matrix  $C_{\text{noSSC}}$  should already have been inverted, so no numerical cost is added there. Then we have a matrix operation to downsize the problem into multiplication and inversion of small  $(n_z, n_z)$  matrices, which is cheap with respect to all the other operations. Hence the computation of  $\delta \ln \mathcal{L}$  takes a time comparable to that of  $\ln \mathcal{L}_{\text{noSSC}}$  and is not a bottleneck of the pipeline.

If the covariance is allowed to vary with model parameters, there is an additional determinant term in the likelihood. The impact of SSC can be computed through the determinant lemma,

$$\det(A + UVV^T) = \det(\mathbb{1} + V^T A^{-1} UV) \times \det(A), \quad (20)$$

which gives

$$\ln \det C_{\text{tot}} = \ln \det C_{\text{noSSC}} + \ln \det(\mathbb{1} + U^T C_{\text{noSSC}}^{-1} US). \quad (21)$$

We thus keep the property  $\ln \mathcal{L} = \ln \mathcal{L}_{\text{noSSC}} + \delta \ln \mathcal{L}$ , and the additive determinant term is numerically cheap since it is the determinant of a  $(n_z, n_z)$  matrix.

### 2.3.2. Fisher matrix

We consider the observable  $\mathcal{O}$  as a vector. The vector of the model parameters is  $\theta$  with a length  $n_\theta$ , and the parameters are indexed with Greek indices  $\mu, \nu, \dots$ . With these notations, the Fisher matrix elements are given by

$$F_{\mu,\nu} = \partial_\mu \mathcal{O}^T C_{\text{tot}}^{-1} \partial_\nu \mathcal{O}, \quad (22)$$

where  $\partial_\mu \mathcal{O}$  is the partial derivative of the observable with respect to the parameter  $\mu$ , considered as a vector. If we plug in Eq. (7), this yields

$$F_{\mu,\nu} = F_{\mu,\nu}^{\text{noSSC}} - Y_\mu^T S (\mathbb{1} + I(\delta_b) S)^{-1} Y_\nu, \quad (23)$$

where  $Y_\mu = U^T C_{\text{noSSC}}^{-1} \partial_\mu \mathcal{O}$  is a vector with the size  $n_z$ . We can thus consider  $Y$  as a  $(n_z, n_\theta)$  matrix, which allows rewriting Eq. (23) in matrix form,

$$F = F_{\text{noSSC}} - Y^T S (\mathbb{1} + I(\delta_b) S)^{-1} Y. \quad (24)$$

In particular, for a single redshift bin ( $n_z = 1$ ), we obtain

$$F = F_{\text{noSSC}} - \frac{Y^T Y S}{1 + I(\delta_b) S}, \quad (25)$$

where  $Y$  is a covector and  $I(\delta_b) = \partial_{\delta_b} \mathcal{O}^T C_{\text{noSSC}}^{-1} \partial_{\delta_b} \mathcal{O}$  is a scalar that can be interpreted as the Fisher information on  $\delta_b$  as an additional parameter. This interpretation will become clearer in Sect. 3. We note that Eq. (25) recovers a result from Lacasa & Grain (2019), specifically, their Eq. (15) in matrix form, with the mapping  $Y_\mu \leftrightarrow f_\mu^{\text{SSC}}$ ,  $I(\delta_b) \leftrightarrow V^T C_{\text{noSSC}}^{-1} V$ , and  $S \leftrightarrow S_{i,i}$ . We note, however, that the latter result is limited to a single redshift bin (which is admittedly of little practical interest), while the general result in Eq. (23) applies to any observable(s) in any number of bins.

In the multi-bin case,  $I(\delta_b)$  is a  $(n_z, n_z)$  matrix that represents the Fisher information on a vector of additional parameters  $\delta_b(\alpha)$ , indeed we have

$$I(\delta_b)_{\alpha,\beta} = \sum_{i,j} \partial_{\delta_b} \mathcal{O}(i, \alpha) C_{\text{noSSC}}^{-1}(i, \alpha; j, \beta) \partial_{\delta_b} \mathcal{O}(j, \beta). \quad (26)$$

### 2.3.3. Inverse Fisher matrix

The inverse of the Fisher matrix is required to forecast marginalised errors on the parameters. Fortunately, the form of Eq. (24) means that the Woodbury inversion formula can be used again. After some algebra, this yields

$$F^{-1} = F_{\text{noSSC}}^{-1} + \delta F^{-1}, \quad (27)$$

$$\delta F^{-1} = F_{\text{noSSC}}^{-1} Y^T S (\mathbb{1} + (I(\delta_b) - \delta I) S)^{-1} Y F_{\text{noSSC}}^{-1}, \quad (28)$$

with

$$\delta I = Y F_{\text{noSSC}}^{-1} Y^T. \quad (29)$$

This can be rewritten in the form that is used throughout this article:

$$F^{-1} = F_{\text{noSSC}}^{-1} + Z^T \Sigma^2(\delta_b) Z, \quad (30)$$

with

$$Z \equiv Y F_{\text{noSSC}}^{-1}, \quad (31)$$

$$\Sigma^2(\delta_b) \equiv S (\mathbb{1} + (I(\delta_b) - \delta I) S)^{-1}, \quad (32)$$

whose physical interpretations will become clearer later.

Equation (30) allows the impact of SSC on the parameter constraints to be forecast as an update to the noSSC case. This has advantages in terms of speed for two reasons: (i) The noSSC pipeline does not need to be rewritten nor rerun, its results can be reused, and (ii) if an assumption of diagonality of the noSSC covariance is used, it can also be used for the SSC Fisher elements: it will speed up the computation of  $Y$ , and from then on, only cheap matrix multiplications are necessary (matrix sizes:  $n_z$  or  $n_\theta$ ).

## 3. Equivalence of SSC approaches

In this section, we show that three approaches to SSC are equivalent under the conditions detailed in the respective subsections. The first is the classical approach used in the previous section, where SSC is included as an additional covariance term. In the second approach, SSC is instead accounted for by promoting  $\delta_b$  to be a nuisance parameter (single bin case; Li et al. 2014) or vector of parameters (one per bin, multi-bin case; Lacasa & Grain 2019) that is marginalised over with a prior  $\mathcal{N}(0, S)$ . The third approach is new. It estimates  $\delta_b$  with a quadratic estimator (QE) to undo the SSC effect.

### 3.1. Equivalence between the covariance and likelihood approaches in the single bin case

In the case of a single redshift bin, the inverse Fisher matrix in the covariance approach is a special case of Eq. (30),

$$F^{-1} = F_{\text{noSSC}}^{-1} + \frac{Z^T Z S}{1 + (I(\delta_b) - \delta I) S}. \quad (33)$$

In the likelihood approach, the covariance of the data is the noSSC covariance, and the parameter vector has been extended to include  $\delta_b$ :  $\theta' = (\theta, \delta_b)$ . This gives a Fisher matrix,

$$F_{\mu,\nu}(\text{free}\delta_b) = \partial_\mu \mathcal{O}^T \cdot C_{\text{noSSC}}^{-1} \cdot \partial_\nu \mathcal{O}, \quad (34)$$

where  $\mu$  and  $\nu$  are the indices of the model parameters, running from 1 to  $n_\theta$ .

On top of this, a prior  $\mathcal{N}(0, S)$  is imposed on  $\delta_b$ . Because the Fisher matrix is the inverse covariance of the parameters and because inverse covariances of independent Gaussians add up, imposing this prior is equivalent to adding a Fisher matrix,

$$F_{\mu,\nu}(\text{prior}) = \frac{1}{S} \delta_{\delta_b,\mu} \delta_{\delta_b,\nu}, \quad (35)$$

where the  $\delta$  on the right-hand side are Kronecker symbols.

When we introduce the notations of Sect. 2.3.2, the total Fisher matrix  $F' = F(\text{free}\delta_b) + F(\text{prior})$  then takes the form

$$F' = \begin{pmatrix} F^{\text{noSSC}} & Y^T \\ Y & I(\delta_b) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{S} \end{pmatrix}. \quad (36)$$

Then some algebra that is detailed in Appendix B for transparency shows that the inverse Fisher matrix can be written as

$$F'^{-1} = \begin{pmatrix} F^{\text{noSSC}-1} & 0 \\ 0 & 0 \end{pmatrix} + \frac{S}{1 + (I - \delta I)S} \begin{pmatrix} Z^T Z & Z^T \\ Z & 1 \end{pmatrix}. \quad (37)$$

In particular, for the lower right element, we have

$$F'^{-1}_{\delta_b,\delta_b} = \frac{S}{1 + (I - \delta I)S}. \quad (38)$$

The upper left block for the parameters  $\theta$  reads

$$F'^{-1} = F^{\text{noSSC}-1} + \frac{S}{1 + IS - \delta IS} Z^T Z. \quad (39)$$

It is identical to Eq. (33), which shows that the covariance and likelihood approaches give the same constraints on the model parameters, that is, they are equivalent.

It seems natural that the equivalence should extend to the case of several bins, although we did not attempt this calculation.

### 3.2. Equivalence between the QE and covariance approaches for the power spectrum in the same multipole range

In this section, we denote estimators with an upper hat. For example,  $\hat{C}_\ell$  is the angular power spectrum estimator. The expectation value is noted with an upper bar, for instance  $\bar{C}_\ell$ , and can be predicted with a cosmological code as a function of  $\theta$  the vector of the model parameters. We assume the theoretical predictions to be unbiased.

A QE can be built to estimate the large-scale modes of the density field using their coupling to better measure small-scale modes (see e.g., Li et al. 2020; Darwish et al. 2021). The natural idea is to use such an estimator to measure  $\delta_b$ , which is the zero mode of the survey, and then combine this with the traditional power spectrum measurement. However, for this zero mode, the QE collapses to a linear combination of power spectrum measurements in full sky. For instance, in terms of angular observables and using notations from the QE literature, Appendix A shows that in the full sky, the  $\delta_b$  estimator is

$$\hat{\delta}_b = \frac{1}{N} \sum_{\ell} g_{\ell,\ell}(0) (\hat{C}_\ell - \bar{C}_\ell(\theta)), \quad (40)$$

with

$$N = \sum_{\ell,\ell'} \partial_{\delta_b} C_{\ell} C_{\ell'}^{-1} \partial_{\delta_b} C_{\ell'}, \quad (41)$$

and

$$g_{\ell,\ell}(0) = \sum_{\ell'} C_{\text{noSSC};\ell,\ell'}^{-1} \partial_{\delta_b} C_{\ell'}. \quad (42)$$

This is consistent with comparable equations from Li et al. (2014), who worked on 3D observables and found an estimator written in terms of  $P(k)$  measurements.

If the QE and the power spectrum are estimated on the same field, and if the multipole range used in the QE is included in the multipole range in which the power spectrum is measured, then the QE measurement is entirely determined by the power spectrum measurement. The joint probability is accordingly

$$P(\hat{C}, \hat{\delta}_b) = P(\hat{C}) \times \delta_D \left( \hat{\delta}_b - \frac{1}{N} \sum_{\ell} g_{\ell,\ell}(0) (\hat{C}_\ell - \bar{C}_\ell) \right), \quad (43)$$

where  $\bar{C}_\ell$  is fixed at its value in the fiducial cosmology needed to build the QE. As in Sect. 3.1, we note  $\theta' = (\theta, \delta_b)$  the vector of parameters. The likelihood then reduces to

$$-\ln \mathcal{L}(\hat{C}, \hat{\delta}_b | \theta') = -\ln \mathcal{L}(\hat{C} | \theta') - \ln \delta_D \left( \hat{\delta}_b - \frac{1}{N} \sum_{\ell} g_{\ell,\ell}(0) (\hat{C}_\ell - \bar{C}_\ell) \right), \quad (44)$$

where the first term is the usual likelihood from the power spectrum alone. In the second term,  $\delta_D$  is a Dirac distribution. This second term has no dependence on the model parameters (e.g., for the weights  $g_{\ell,\ell}(0)$ , all cosmological parameters are fixed at the fiducial value used to build the estimator). The Fisher matrix is given in terms of the likelihood by

$$F_{\alpha,\beta} = \left\langle - \frac{\partial^2 \ln \mathcal{L}}{\partial \theta'_\alpha \partial \theta'_\beta} \middle| \theta' \right\rangle, \quad (45)$$

where  $\langle | \theta' \rangle$  denotes the conditional expectation given the values of  $\theta'$ .

If we call  $F''$  the Fisher matrix of this QE + power spectrum approach, we obtain

$$F''_{\alpha,\beta} = F'_{\alpha,\beta} + \left\langle - \frac{\partial^2 \ln \delta_D(\hat{\delta}_b - \dots)}{\partial \theta'_\alpha \partial \theta'_\beta} \middle| \theta' \right\rangle, \\ = F'_{\alpha,\beta}, \quad (46)$$

where the second term vanished because the derivatives with respect to model parameters vanish. Thus we have the same Fisher constraints as the likelihood of the power spectrum, as well as the traditional covariance approach for constraints on  $\theta$ .

### 3.3. Physical interpretation and conclusion

In this section, we use the two previous sections to give a physical interpretation to some of the quantities introduced in Sects. 2.3.2 and 2.3.3:  $I(\delta_b)$ ,  $\Sigma^2(\delta_b)$ ,  $\delta I$ , and  $Z$ .

First, Eq. (36) shows that  $I(\delta_b)$  is the direct Fisher information on  $\delta_b$ . Furthermore, from the QE normalisation in Eq. (41), we can recognise that  $N = I(\delta_b)$ . Computing the variance of the QE gives

$$\langle \hat{\delta}_b^2 \rangle = \frac{1}{N} = \frac{1}{I(\delta_b)}. \quad (47)$$

Hence  $I(\delta_b)$  is the inverse variance of the maximum likelihood estimator for  $\delta_b$ , which consolidates its interpretation as the raw information on  $\delta_b$  when all other model parameters are fixed.

Second,  $\Sigma^2(\delta_b)$  is defined in Eqs. (32), and (38) (in the single bin case) shows that it is the marginalised Fisher constraint on  $\delta_b$ . In the single bin case, it can be cast in the form

$$\Sigma^2(\delta_b) = \frac{1}{I(\delta_b) - \delta I + \frac{1}{S}}, \quad (48)$$

which makes it clear that there is a difference with the unmarginalised QE error due to (i) the presence of a prior  $\frac{1}{S}$ , which adds information, (ii) the negative term  $-\delta I$ , which removes information. In the limit of no prior, we have  $S \rightarrow \infty$  and  $\Sigma^2(\delta_b) \rightarrow \frac{1}{I(\delta_b) - \delta I}$ . We therefore conclude that  $\delta I$  quantifies the loss of information on  $\delta_b$  due to the degeneracy with the other model parameters in  $\theta$ .

Third, the off-diagonal matrix elements of Eq. (37) show that the covector  $Z$  quantifies the degeneracy between  $\delta_b$  and each model parameter in  $\theta$ . Furthermore, when we write Eq. (39) on the diagonal, we have

$$F_{\mu,\mu}^{-1} = F_{\text{noSSC},\mu,\mu}^{-1} + \Sigma^2(\delta_b) (Z_\mu)^2, \quad (49)$$

which shows that  $Z_\mu$  is the (square root of the) increase in variance on parameter  $\mu$  as a function of the constraint on  $\delta_b$ . To reflect this, we could have noted this quantity  $\sqrt{\frac{\partial \sigma^2(\mu)}{\partial \sigma^2(\delta_b)}}$  or perhaps  $\delta \sigma_{\text{rms}}^{\text{SSC}}(\mu)$ , but this would have made the notations of all previous equations in Sect. 3 much heavier and harder to read. We therefore chose to keep a simple  $Z$ .

Finally, as a summary of this section, we have shown that accounting for SSC at the likelihood level is formally equivalent, in terms of parameter constraints, to accounting for it traditionally at the covariance level. Combining a QE and a power spectrum does not bring better parameter constraints than the power spectrum alone as long as both estimators act upon the same field and the same multipole range.

The  $\hat{\delta}_b$  estimator might be used mainly (i) when some observational effects can be dealt with more efficiently (e.g., survey mask or systematics) in the QE than in the power spectrum analysis, (ii) if it allows adding information from another field (e.g., estimate  $\delta_b$  with galaxies to improve a WL analysis), from another statistic (e.g., a linear, cubic, or log-normal estimator that uses information from the one-point or three-point function or from the log field), or from another survey (see e.g., Digman et al. 2019) when a full combined probe analysis would be too costly.

#### 4. Application to a stage IV weak-lensing and galaxy clustering analysis

In this section, we apply the previous results to the case of a future stage IV photometric galaxy survey that constrains cosmology through WL alone and in combination with GC and XC. We describe the experimental and modelling setup in Sect. 4.1. In Sect. 4.2 we examine how well the survey can constrain the super-sample modes. In Sect. 4.3 we ask whether any theoretical prior on  $\delta_b$  can be lifted. Finally, in Sect. 4.4 we study the precision to which the SSC responses should be modelled in order to pass certain precision requirements on the reported cosmological errors.

##### 4.1. Setup

For the cosmological modelling, we worked within the  $w_0 w_a$  CDM model, and in the Fisher forecast, we varied each of its

parameters:  $\Omega_m, \Omega_b, w_0, w_a, h, n_s$ , and  $\sigma_8$ . In practice, we mostly focused on the parameters that are most affected by SSC:  $w_0, w_a, \sigma_8$ , and  $\Omega_m$ . For the prediction of all cosmological quantities, we used the software *cosmosis* by Zuntz et al. (2015), and for the prediction of the matter power spectrum  $P(k)$  on non-linear scales, we used the prescription by Takahashi et al. (2012) and Bird et al. (2012).

In terms of experimental specifications, we followed all the specifications of Euclid Collaboration (2020) for a 15 000 deg<sup>2</sup> photometric galaxy survey to analyse WL, GC and XC. In slightly more detail, the galaxy redshift distribution is

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \quad (50)$$

with  $z_0 = 0.9/\sqrt{2}$  and a total density of 30 galaxies per square arcmin. Then we defined ten tomographic redshift bins such that they are equi-populated, yielding the boundaries  $z_i = (0.0010, 0.42, 0.56, 0.68, 0.79, 0.90, 1.02, 1.15, 1.32, 1.58, 2.50)$ . The redshifts were then perturbed by photometric errors with the main dispersion and a population of outliers, following the specific details given in Euclid Collaboration (2020). These galaxies are both the source and lens sample, that is, their shape was used for the WL measurement and their position for the GC measurement.

In terms of probe modelling, we also followed Euclid Collaboration (2020). The GC was modelled with a scale-independent bias with a fiducial value  $b(z) = \sqrt{1+z}$ . For cosmological constraints, the galaxy bias value in each tomographic bin is a nuisance parameter that needs to be marginalised over. The WL is modelled classically as a tracer of matter, its power is hampered on small scales by the shape noise with standard deviation  $\sigma_\epsilon = 0.3$ . Another issue for WL is the contribution from the intrinsic alignments (IA) of galaxies. To model this effect, we followed Euclid Collaboration (2020): the IA power spectrum and cross-spectrum with matter are given by  $P_{\text{II}} = [-A(z)]^2 P_{\text{mm}}$  and  $P_{\text{ml}} = -A(z) P_{\text{mm}}$ , where

$$A(z) = \mathcal{A}_{\text{IA}} \frac{C_{\text{IA}} \Omega_m \mathcal{F}_{\text{IA}}(z)}{D(z)}, \quad (51)$$

with  $C_{\text{IA}} = 0.0134$  being a normalisation constant,  $D(z)$  the growth factor, and  $\mathcal{A}_{\text{IA}}$  a nuisance parameter that controls the amplitude of the IA contribution. The redshift dependence is given by

$$\mathcal{F}_{\text{IA}}(z) = (1+z)^{\eta_{\text{IA}}} \left[ \frac{\langle L \rangle(z)}{L_*(z)} \right]^{\beta_{\text{IA}}}, \quad (52)$$

where  $\langle L \rangle(z)/L_*(z)$  is the ratio of the mean source luminosity and the characteristic scale of the luminosity function (Hirata et al. 2007; Bridle & King 2007). Additionally, we have two nuisance parameters,  $\eta_{\text{IA}}$  and  $\beta_{\text{IA}}$ , making a total of three nuisance parameters for intrinsic alignments, with fiducial values (Euclid Collaboration 2020)  $(\mathcal{A}_{\text{IA}}, \eta_{\text{IA}}, \beta_{\text{IA}}) = (1.72, -0.41, 2.17)$ .

For the statistical observable, we used the angular power spectrum for all probes: the WL auto-spectrum, the GC auto-spectrum, and the XC cross-spectrum. For simplicity, we adopted a common range of multipoles from  $\ell_{\text{min}} = 10$  to  $\ell_{\text{max}} = 3000$  for all probes. The covariance of these power spectra is described by a classical Gaussian term (including shape noise for WL and shot-noise for GC), and the super-sample covariance term is accounted for through the developments of Sect. 2. Specifically for the SSC, we computed the

$S$  matrices using PySSC<sup>1</sup> (Lacasa & Grain 2019), and for the responses, we followed Gouyou Beauchamps et al. (2022) and took  $\partial_{\delta_b} C_\ell = R \times C_\ell$  with a probe-independent dimensionless response  $R = 4$ . Following Barreira et al. (2018a), other non-Gaussian covariance terms have a negligible impact on cosmological constraints from a WL survey, hence we neglected these terms here.

#### 4.2. Information on $\delta_b$

In this section, we compare the information on  $\delta_b$  brought by the survey, either raw information with  $I(\delta_b)$  defined by Eq. (8) or marginalised information with  $I - \delta I$  defined by Eq. (29), to the prior information given by the  $S$  matrix for the considered redshift bins.

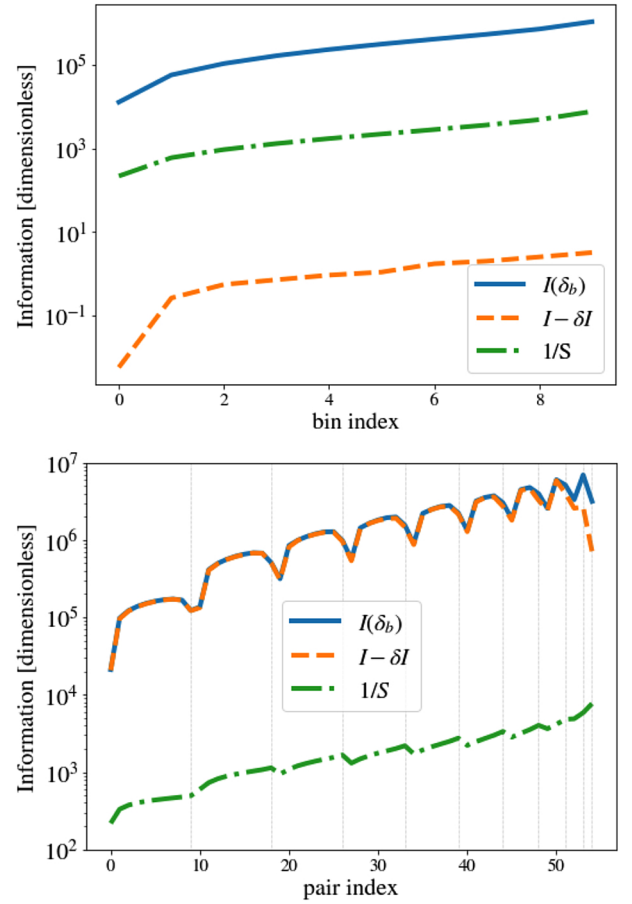
We first consider the WL-only analysis. The first case of interest is to try to constrain  $\delta_b$  in a redshift bin by only using the information in this bin. The result is shown in the top panel of Fig. 1, where we plot  $I(\delta_b)$ ,  $I - \delta I$ , and  $1/S$  in each of the ten redshift bins. The raw information (solid blue) is an order of magnitude larger than the prior (dash-dotted green), which appears to mean that the data can constrain  $\delta_b$  very well. However, the marginalised information (dashed orange) falls by  $\sim 3$  orders of magnitude compared to the unmarginalised information, making it lower than the prior by  $\sim 2$  orders of magnitude. This comes from a high cancellation of  $I(\delta_b)$  and  $\delta I$  due to high degeneracy between  $\delta_b$  and cosmological parameters. We found that this is, in fact, not peculiar to  $\delta_b$ ; with a single redshift bin, most cosmological parameters are near perfectly degenerate, and the Fisher matrix has very poor conditioning.

This fact motivated us to pursue the single bin case no longer and instead analyse the tomographic case of a joint analysis of all bins and further pairs of bins together. The result is shown in the bottom panel of Fig. 1, where we plot  $I(\delta_b)$ ,  $I - \delta I$ , and  $1/S$  in each of the 55 redshift pairs. Explicitly, each of these quantities is now a  $(55, 55)$  matrix, and the figure shows the diagonal of these matrices.

In this tomographic case, the raw information is larger than the prior by an even larger amount,  $\sim 2$  orders of magnitude, than in the single bin case. The parameter degeneracies are now all lifted, both for cosmological parameters and for nuisance parameters such as  $\delta_b$ . We attribute this to the fact that the tomography provides information on the history of expansion and growth. Hence, here the marginalised information  $I - \delta I$  sits almost perfectly on top of the unmarginalised information  $I(\delta_b)$ , with only slight loss in the high-redshift pairs.

Next, we studied the  $3 \times 2$ pt analysis, which combines the WL auto-spectra, the GC auto-spectra, and the XC cross-spectra. There are now 210 bin pairs: 55 for WL, 55 for GC, and 100 for XC because the latter is not symmetric. Figure 2 shows  $I(\delta_b)$ ,  $I - \delta I$ , and  $1/S$  in each of the 210 redshift pairs. We chose to order the pairs with WL first, then XC, and then GC; the corresponding intervals are indicated by vertical dashed lines in the figure.

The marginalised information is oftentimes indistinguishable from the unmarginalised one and is larger than the prior. Deviations from this behaviour are visible as breaks of the lines in the XC and GC pairs. We cut the y-axis of the plot to zoom in on the interesting region. Otherwise, the blue and orange lines fall extremely low, making the plot unreadable. For these pairs, the two redshift kernels overlap extremely little, so that the signal is practically absent. This does not happen for WL because the WL kernels are broad in redshift, integrating everything between the source and the observer so that all kernels overlap strongly. How-



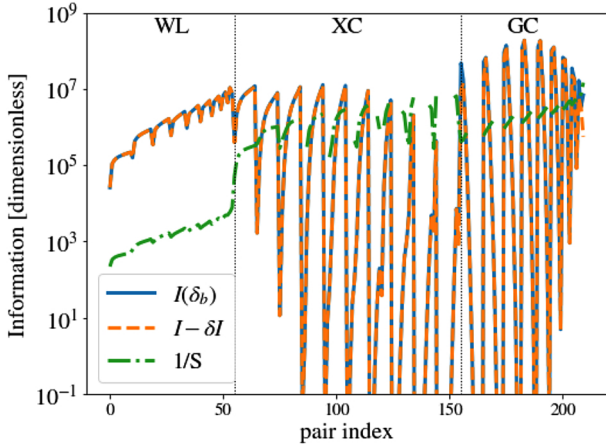
**Fig. 1.** Information on  $\delta_b$ : Raw ( $I$ ), marginalised ( $I - \delta I$ ), and prior ( $1/S$ ). *Top*: information in each bin of redshift for WL analyses of each bin separately. *Bottom*: information in each pair of redshift bins for a joint tomographic WL analysis. The 55 pairs are ordered as  $(0, 0)(0, 1) \dots (0, 9)(1, 1) \dots (1, 9) \dots (8, 8)(8, 9)(9, 9)$ . The light vertical lines indicate the jump of the first redshift bin, i.e., the leftmost line is at  $(0, 9)$  before the jump to  $(1, 1)$ .

ever, the GC kernels are much more localised, and there is, for instance, practically no overlap between the first and last redshift bins. The cross-correlation XC will also vanish when measured between a low or median WL bin and a high GC bin. We note that the prior (green line) also breaks in these pair cases because PySSC sets the  $S$  matrix element to 0 when the overlap of kernels is too small for a reliable computation. In practice, however, this is not an issue because in these pairs, the signal is too low to contain any cosmological information. Essentially, Fig. 2 shows that  $\delta_b$  can be well constrained for all relevant portions of the data.

A theoretical concern that could come to mind, already for the tomographic WL analysis and even more for the  $3 \times 2$ pt analysis, is the physical significance of these numerous  $\delta_b$  parameters: one for each bin pair. Through Eq. (6), they change only the observable for that bin pair; physically, they are the survey matter density contrast averaged over redshift with the product of the two kernels of the pair. The concern is that all these pairs probe a similar redshift range, and that therefore the parameters should be strongly correlated. The answer is that this is indeed the case, and in practice, the correlation of these parameters is quantified by the  $S$  matrix, which has high off-diagonal correlation coefficients. When the formulae developed in Sect. 2 for the likelihood or Fisher forecast are used, the  $\delta_b$  parameters do not appear in

<sup>1</sup> <https://pyssc.readthedocs.io/>





**Fig. 2.** Information on  $\delta_b$  in each redshift bin pair in a tomographic  $3\times 2$ pt analysis (WL + GC + XC). There are 210 pairs: 55 for WL, 100 for XC, and 55 for GC. The pairs are ordered as in Fig. 1. The only difference is that XC has more pairs because there is no symmetry  $(i, j) \leftrightarrow (j, i)$ .

the computation, and all the formulae are well behaved numerically. In a different context, for a different kind of analysis that explicitly has  $\delta_b$  nuisance parameters, there could be numerical problems with their number and high correlation. In that case, Appendix C devises an analysis with fewer less strongly correlated  $\delta_b$  parameters, at the expense of needing more observable predictions.

### 4.3. Asymptotic errors

In this section, we determine the change in the Fisher constraints when an infinitely thin or wide prior  $\mathcal{N}(0, S)$  is assumed. In the single bin case, this corresponds to the respective cases  $S \rightarrow 0$  and  $S \rightarrow \infty$ .

The case of an infinitely thin prior is simple. From Eq. (25) (single bin case) or Eq. (24) (multi-bin case), the Fisher constraint asymptote to the noSSC one. This happens because the prior fixes  $\delta_b$  to 0 in each bin, so that there is no SSC effect.

The case of an infinitely wide prior is more interesting. It corresponds to  $\delta_b$  being completely free,  $S \rightarrow \infty$ . It might be motivated by the observation of Sect. 4.2 that the data appear to constrain  $\delta_b$  much better than the prior, so that the latter seems not to be needed.

The Fisher matrix was already given in Eq. (34), but for the parameter vector  $\theta' = (\theta, \delta_b)$ . After marginalisation over  $\delta_b$ , the inverse Fisher is

$$F^{-1} = F_{\text{noSSC}}^{-1} + Z^T (I(\delta_b) - \delta I)^{-1} Z. \quad (53)$$

The corresponding marginalised error bars can be compared to the fiducial case in which  $S$  is given by the theoretical prediction from PySSC. The ratios of these two cases for cosmological parameters (after marginalisation over all the others, as well as the nuisance parameters) are given in Table 1. This conservative no-prior analysis would increase the cosmological errors by unacceptably large factors between 1.1 ( $\Omega_b$ , WL-only) and 5 ( $w_a$ , WL-only), with an average of 2.8 for WL-only and 2.5 for  $3\times 2$ pt.

At first, this may seem at odds with the result of Sect. 4.2 that the data appear to constrain  $\delta_b$  much better than the prior. However, this happens because the data mostly constrain each  $\delta_b$  independently, while, as discussed at the end of Sect. 4.2, the

**Table 1.** Ratio of the error bar for each cosmological parameter between the case of no prior on  $\delta_b$  and the case of the fiducial prior.

|             | $\Omega_m$ | $\Omega_b$ | $w_0$ | $w_a$ | $h$  | $n_s$ | $\sigma_8$ |
|-------------|------------|------------|-------|-------|------|-------|------------|
| WL          | 3.12       | 1.12       | 3.87  | 5.00  | 1.80 | 2.11  | 2.55       |
| $3\times 2$ | 3.49       | 1.27       | 3.13  | 3.56  | 1.71 | 1.41  | 2.96       |

**Notes.** Second line: Case of a tomographic WL analysis. Third line: Case of a  $3\times 2$ pt analysis.

prior has strong off-diagonal correlations, which encode the fact that the  $\delta_b$  probe similar redshift ranges. In the no-prior case, we therefore assign much freedom to the data, which blows up the constraints.

### 4.4. Requirements for reaching robust cosmic errors

The results of Sect. 4.3 pose the more general question of how precisely we need to model the super-sample covariance in order for cosmological constraints to be robust. This question is better posed at the covariance level than at the likelihood level used in Sect. 4.3. We focus on the impact on the error bars, although, as noted by Sellentin & Starck (2019) using an incorrect covariance matrix is also equivalent to analysing a biased data set.

To answer this question, we let the SSC responses vary both in amplitude and scale dependence and forecast the variation in cosmological constraints with respect to the case with the fiducial responses.

#### 4.4.1. Weak-lensing-only case

For the WL case, we introduced two parameters,  $A_0$  and  $A_1$ , which change the response as

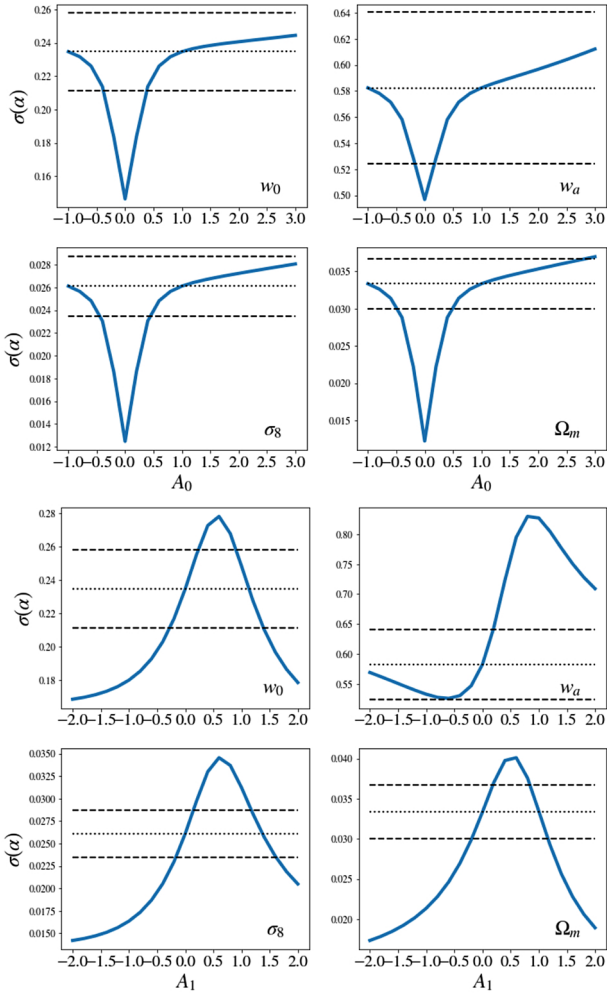
$$\partial_{\delta_b} C_\ell \rightarrow A_0 \left( 1 + A_1 \frac{\ell - \ell_0}{2\ell_0} \right) \times \partial_{\delta_b} C_\ell, \quad (54)$$

with  $\ell_0$  a pivot scale chosen to be the centre of the multipole range so that the fraction varies by 1 between  $\ell_{\min} \sim 0$  and  $\ell_{\max} = 2\ell_0$ .

Based on the developments of Sect. 2.3, we can compute a Fisher forecast without building and inverting a new covariance matrix for each value of  $(A_0, A_1)$  we wish to test. Instead, we first computed the Fisher forecast for the Gaussian covariance matrix and then used Eq. (30) to determine the inverse Fisher matrix. We found that the bottleneck is the computation of  $I(\delta_b)$ , whose brute-force computation takes  $\sim 3$  s on a laptop with i7 CPU. By comparison, a covariance inversion takes  $\sim 120$  s on the same laptop, which would make the forecast painful to run when many values for  $A_0$  and  $A_1$  need to be tested. This speedup was achieved without using the fact that the Gaussian covariance is diagonal in terms of multipoles.

We can then study how the marginalised errors on cosmological parameters change with  $A_0$  and  $A_1$ . The results are shown in Fig. 3 for the most affected parameters, which are in fact the usual targets of WL experiments: the dark energy equation of state,  $\sigma_8$  and  $\Omega_m$ .

We let  $A_0$  run between  $-1$  and  $3$ , that is, between predicting the response with a sign mistake and misestimating the response by a factor of 3. We let  $A_1$  run between  $-2$  and  $2$ , which is a scale dependence that is strong enough to cause the response to vanish at the minimum or maximum bound of the multipole range. The plots also contain dashed horizontal lines to indicate



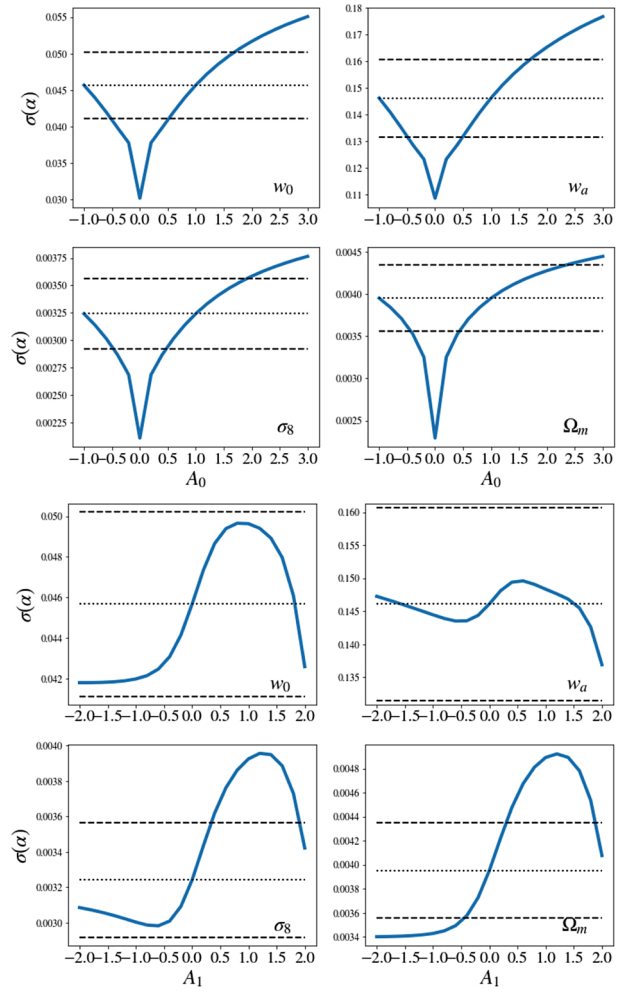
**Fig. 3.** Forecasted errors on the dark energy equation of state,  $\sigma_8$  and  $\Omega_m$  for a tomographic WL analysis. *Top two rows:* constraints as a function of  $A_0$  (which changes the amplitude of the SSC response). *Bottom two rows:* constraints as a function of  $A_1$  (which changes the scale dependence of the SSC response). Dotted lines show the value for the fiducial covariance,  $(A_0, A_1) = (1, 0)$ , and the dashed lines show  $\pm 10\%$  around the dotted line.

when the errors deviate from the fiducial case by more than 10%, in line with the requirements of stage IV missions (e.g., *Euclid*, Laureijs et al. 2011).

The results show that the branch in which  $A_0$  is negative is symmetric to the positive branch. This happens because  $A_0$  is squared in the covariance equation, so that an overall sign mistake has no impact. As  $A_0$  increases (in absolute value), so do the cosmological errors. The requirement that errors should have a precision of 10% constrains the allowed value of  $A_0$ . The strongest constraints come from  $\Omega_m$ , giving  $A_0 \in [0.47, 2.82]$ . We note that this is a quite large interval: response predictions can be under- or over-estimated by a factor larger than 2.

For  $A_1$  the curves are no longer symmetric and have a steep slope near the fiducial case  $A_1 = 0$ , so that they easily exceed the 10% requirements. Therefore, satisfying the required precision places strong constraints on  $A_1$ ; the strongest constraint comes from  $\sigma_8$ , giving  $A_1 \in [-0.16, 0.15]$ .

In summary, WL constraints are rather robust to a mismodelling of the amplitude of its response. However, the constraints are more sensitive to the scale dependence of the response, which must be modelled at the 15% level.



**Fig. 4.** Forecasted errors on the dark energy equation of states,  $\sigma_8$ , and  $\Omega_m$  for a  $3 \times 2$ pt analysis. *Top two rows:* constraints as a function of  $A_0$ , which changes the amplitude of the SSC response. *Bottom two rows:* constraints as a function of  $A_1$ , which changes the scale dependence of the SSC response. Dotted lines show the value for the fiducial covariance, and the dashed lines are  $\pm 10\%$  around the dotted line.

#### 4.4.2. $3 \times 2$ pt case

For the  $3 \times 2$ pt case, that is, when WL is combined with GC, including XC, we can first use the same parametrisation Eq. (54), with  $A_0$  and  $A_1$  being the same for all three probes. We can then observe the impact of varying these parameters, which change the amplitude and scale dependence of all probes together. The results are shown in Fig. 4 for the parameters that are usually most affected: dark energy,  $\sigma_8$ , and  $\Omega_m$ .

The forecasts are generally more robust to these mismodelling parameters. For  $A_0$ , the stringent constraints are set by  $w_0$ , which requires  $A_0 \in [0.53, 1.70]$ . This is slightly stricter than the WL case, but still quite large. For  $A_1$ , the stringent constraints come from  $\Omega_m$ , which requires  $A_1 \in [-0.44, 0.29]$ . The scale dependence must therefore be modelled at  $\sim 30\%$  precision; this constraint is twice looser than for the WL-only case.

These results hold for a common mismodelling of all responses of all probes, for example in a separate universe approach, if we have a systematic common to all probes in the simulations, or in a theoretical approach, if we were missing a term in our equations. To study the case when the physics of a single probe is modelled incorrectly, we defined a new model

for the responses with new parameters  $\epsilon_p$  for the amplitude of the response of each probe independently,

$$\partial_{\delta_b} C_\ell^p \rightarrow \epsilon_p \left( 1 + A_1 \frac{\ell - \ell_0}{2\ell_0} \right) \times \partial_{\delta_b} C_\ell^p, \quad (55)$$

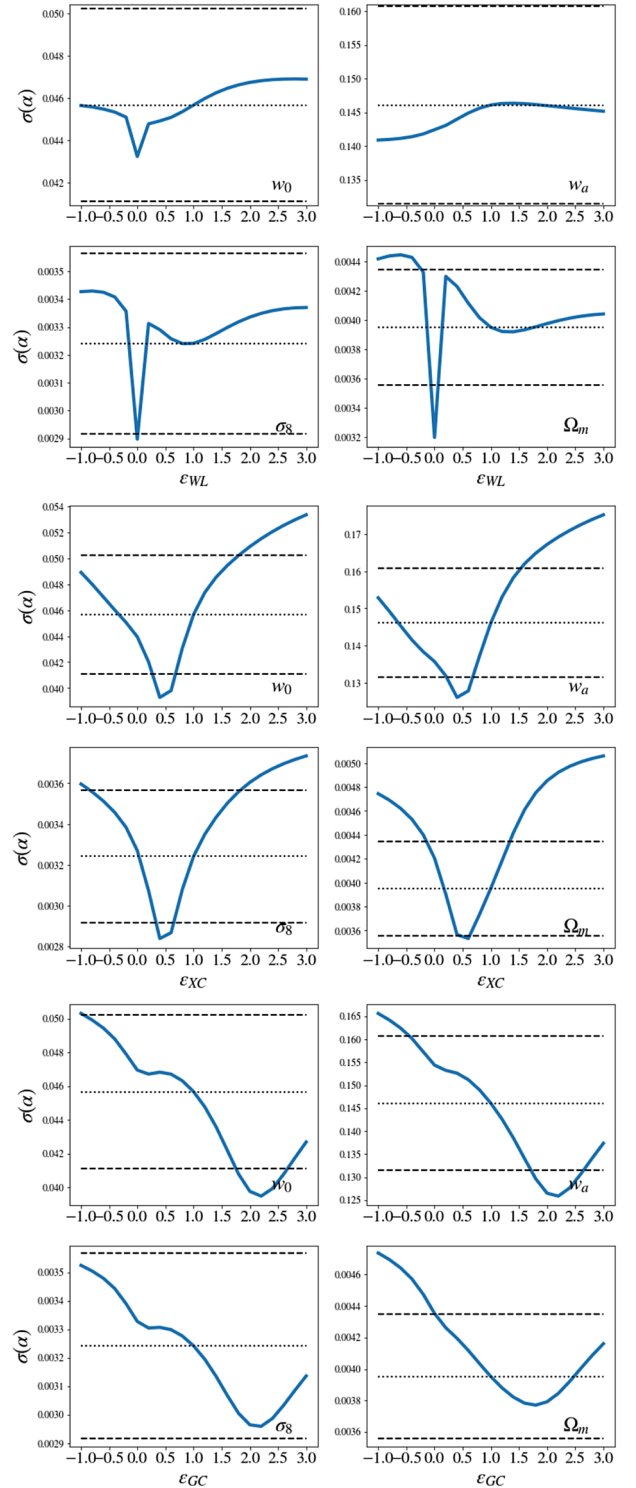
where  $(\epsilon_{\text{WL}}, \epsilon_{\text{GC}}, \epsilon_{\text{XC}})$  modulate the response of the power spectrum of WL, GC, and XC respectively. Figure 5 shows the impact of these parameters on the parameters that are usually most affected (dark energy equation of state parameters,  $\sigma_8$ , and  $\Omega_m$ ),  $\epsilon_{\text{WL}}$  has little impact on the constraints, which easily stay within the requirements for most of the explored range of values. Only  $\Omega_m$  discards some values. In particular, the constraints on dark energy are very robust throughout all explored ranges. The error bars are also quite robust to  $\epsilon_{\text{GC}}$ , although to a lesser extent than for  $\epsilon_{\text{WL}}$ . For instance, constraints on  $\sigma_8$  stay within the requirements in all explored ranges. Lower and upper values on  $\epsilon_{\text{GC}}$  are set by different cosmological parameters: to meet the 10% requirement,  $\Omega_m$  requires  $\epsilon_{\text{GC}} \geq 0.01$  and  $w_a$  requires  $\epsilon_{\text{GC}} \leq 1.71$ . The combined allowed range is thus  $\epsilon_{\text{GC}} \in [0.01, 1.71]$ , which is quite large. By contrast, in the case of  $\epsilon_{\text{XC}}$ , the errors exceed the requirements much more easily. The lower constraint is  $\epsilon_{\text{XC}} \geq 0.69$  due to  $w_0$ , and the upper constraint is  $\epsilon_{\text{XC}} \leq 1.34$  due to  $\Omega_m$ . The lower constraint could be relaxed by adopting a slightly looser requirement, as the curves drop just below the  $-10\%$  line, although we do not enter into these considerations. In summary, the amplitude of this response must be modelled at a precision of  $\sim 30\%$ .

It is expected that variations in the XC response have the most important effect because it has been shown that this cross correlation is of the utmost importance to control systematics and unlock the full power of the  $3 \times 2$ pt analysis (Tutusaus et al. 2020).

In summary,  $3 \times 2$ pt constraints are more robust to SSC mis-modelling than WL-only constraints. The most important quantities to model correctly are the amplitude of the XC response and the scale dependence of all responses. Both need to be calibrated to the  $\sim 30\%$  level.

## 5. Conclusion

Within the framework of the Sij approximation to SSC, we have presented a new formula in which SSC is an update to the precision (inverse covariance) matrix. This update can be computed numerically in a time comparable to that of the noSSC precision matrix and can take advantage of any sparsity in the noSSC covariance. We showed practical examples for two types of cosmological probes: number counts, and power spectrum. We presented the application of this formula to the computation of the likelihood or a Fisher forecast as an update to the noSSC case. The Fisher matrix (and inverse Fisher matrix) were then used to prove the formal equivalence of three approaches to super-sample covariance: (i) classically at the covariance level, (ii) at the likelihood level through dedicated nuisance parameters, and (iii) with a QE to estimate the super-sample modes. Examination of these three approaches allowed us to give a physical interpretation of the various quantities entering the formula. Although these approaches are formally equivalent, we argue that the second approach, at the likelihood level, is conceptually more interesting for two reasons: (i) the  $\delta_b$  nuisance parameters can be examined to track systematics, which would cause them to deviate significantly from their prior (e.g., a systematic inducing an excess of power in some bins would push  $\delta_b > 0$  for these bins), and (ii) it provides a natural framework to determine whether the



**Fig. 5.** For a  $3 \times 2$ pt analysis, forecast errors on dark energy,  $\sigma_8$ , and  $\Omega_m$ . *Top:* as a function of  $\epsilon_{\text{WL}}$ , which changes the WL response alone. *Centre:* as a function of  $\epsilon_{\text{XC}}$ , which changes the cross-correlation response alone. *Bottom:* as a function of  $\epsilon_{\text{GC}}$ , which changes the GC response alone. Dotted line: Value for the fiducial covariance. Dashed lines:  $\pm 10\%$  around the dotted line.

data prefer low density ( $\delta_b < 0$ ) at low redshift. This potential explanation has been put forward in the past for acceleration or dark energy or the  $H_0$  tension.

Furthermore, we presented a numerical application of this framework to a mock stage IV weak-lensing survey. Namely,

we considered a WL-only and a WL + GC + XC (also known as 3×2pt) case. We showed that although the  $\delta_b$  parameters are well constrained, we cannot recommend to remove their theoretical prior because it would strongly degrade the cosmological constraints. We also examined the change induced by mismodelling on the cosmological errors, and especially how much mismodelling can be allowed to still satisfy the requirement of a precision of 10% for stage IV surveys. This is of interest as there are always theoretical or simulation uncertainties in the prediction of these responses. We found that large mispredictions of the response amplitude are allowed for the WL-only case, but the scale dependence must be predicted more precisely at the  $\sim 15\%$  level. For the 3×2pt case, large amplitude mispredictions are possible if it is common to all probes, but a misprediction specific to the cross-spectrum signal (XC = GC – WL power spectrum) is more problematic. This amplitude needs to be calibrated to a precision of  $\sim 30\%$ . The scale dependence of the response must also be predicted precisely, at the  $\sim 30\%$  level, although the requirement is less strict than for the WL-only case. In practice, this is the target of convergence that should be aimed for in the comparison of the response models or simulations.

*Acknowledgements.* We thank Stéphane Ilić for fruitful exchanges and Joachim Harnois-Déraps for proofreading the manuscript. F.L. acknowledges support from the Swiss National Science Foundation through grant number IZSEZO\_207059, and from Progetto di Eccellenza 2022 of the Physics and Astronomy department at the University of Padova “Physics of the Universe”, within the grant “CosmoGraLSS – Cosmology with Gravitational waves and Large Scale Structure”. JC and LL acknowledge support from a SNSF Eccellenza Professorial Fellowship (No. 186879). AMD acknowledges support from Tommala Foundation for research in gravity as well as Boninchi Foundation. SGB was supported by CNES, and focused on the *Euclid* mission. The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille University - A\*MIDEX, a French “Investissements d’Avenir” program (AMX-19-IET-008 -IPhU). AG’s work is supported by the McGill Astrophysics Fellowship funded by the Trottier Chair in Astrophysics, as well as the Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholars program and the Canada 150 Programme. The calculations in this article made use of `class` (Blas et al. 2011), `cosmosis` by Zuntz et al. (2015), `astropy` (Astropy Collaboration 2013, 2018), `matplotlib` (Hunter 2007), `scipy` (Jones et al. 2001), and `numpy` (Oliphant 2006).

## References

- Agüena, M., & Lima, M. 2018, *Phys. Rev. D*, **98**, 123529
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Barreira, A., & Schmidt, F. 2017, *JCAP*, **2017**, 053
- Barreira, A., Krause, E., & Schmidt, F. 2018a, *JCAP*, **10**, 053
- Barreira, A., Krause, E., & Schmidt, F. 2018b, *JCAP*, **6**, 015
- Bird, S., Viel, M., & Haehnelt, M. G. 2012, *MNRAS*, **420**, 2551
- Blas, D., Lesgourgues, J., & Tram, T. 2011, *JCAP*, **2011**, 034
- Bridle, S., & King, L. 2007, *New J. Phys.*, **9**, 444
- Chan, K. C., Moradinezhad Dizgah, A., & Noreña, J. 2018, *Phys. Rev. D*, **97**, 043532
- Darwish, O., Foreman, S., Abidi, M. M., et al. 2021, *Phys. Rev. D*, **104**, 123520
- Digman, M. C., McEwen, J. E., & Hirata, C. M. 2019, *JCAP*, **2019**, 004
- Euclid Collaboration (Blanchard, A., et al.) 2020, *A&A*, **642**, A191
- Gouyou Beauchamps, S., Lacasa, F., Tutusaus, I., et al. 2022, *A&A*, **659**, A128
- Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, *MNRAS*, **465**, 1454
- Hirata, C. M., Mandelbaum, R., Ishak, M., et al. 2007, *MNRAS*, **381**, 1197
- Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, *ApJ*, **567**, 2
- Hu, W., & Kravtsov, A. V. 2003, *ApJ*, **584**, 702
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open source scientific tools for Python*
- Lacasa, F. 2018, *A&A*, **615**, A1
- Lacasa, F. 2020, *A&A*, **634**, A74
- Lacasa, F., & Grain, J. 2019, *A&A*, **624**, A61
- Lacasa, F., & Kunz, M. 2017, *A&A*, **604**, A104
- Lacasa, F., & Rosenfeld, R. 2016, *JCAP*, **8**, 005
- Lacasa, F., Lima, M., & Agüena, M. 2018, *A&A*, **611**, A83
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *ArXiv e-prints* [arXiv:1110.3193]
- Li, Y., Hu, W., & Takada, M. 2014, *Phys. Rev. D*, **90**, 103530
- Li, Y., Schmittfull, M., & Seljak, U. 2018, *JCAP*, **2**, 022
- Li, P., Dodelson, S., & Croft, R. A. C. 2020, *Phys. Rev. D*, **101**, 083510
- Okamoto, T., & Hu, W. 2003, *Phys. Rev. D*, **67**, 083002
- Oliphant, T. 2006, *NumPy: A guide to NumPy* (USA: Trelgol Publishing)
- Sellentin, E., & Starck, J.-L. 2019, *JCAP*, **2019**, 021
- Takada, M., & Hu, W. 2013, *Phys. Rev. D*, **87**, 123504
- Takada, M., & Spergel, D. N. 2014, *MNRAS*, **441**, 2456
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, *ApJ*, **761**, 152
- Takahashi, R., Soma, S., Takada, M., & Kayo, I. 2014, *MNRAS*, **444**, 3473
- Tutusaus, I., Martinelli, M., Cardone, V. F., et al. 2020, *A&A*, **643**, A70
- Wadekar, D., & Scoccimarro, R. 2020, *Phys. Rev. D*, **102**, 123517
- Woodbury, M. A. 1950, Memorandum Rept. 42, Statistical Research Group (Princeton, NJ : Princeton University)
- Zuntz, J., Paterno, M., Jennings, E., et al. 2015, *Astron. Comput.*, **12**, 45

## Appendix A: Quadratic estimator for $\delta_b$

We wish to estimate the background density over the full sky,  $\delta_b(z) = \int \frac{d^2\hat{n}}{4\pi} \delta_m(r\hat{n}, z)$ , which is proportional to the monopole of the matter density field,  $\delta_b(z) = a_{00}(\delta_m)/\sqrt{4\pi}$ .

We use notations from the seminal article of [Okamoto & Hu \(2003\)](#) with slight variations to avoid problems at  $L = 0$ , and consider a single pair of fields. The estimator for the mode  $(L, M)$  has the general form

$$d_L^M = A_L \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} (-1)^M \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & -M \end{pmatrix} \times g_{\ell_1, \ell_2}(L) a_{\ell_1, m_1} a_{\ell_2, m_2}. \quad (\text{A.1})$$

We apply this to  $L = M = 0$ , which leads to an important collapse. The Wigner 3J symbol enforces the triangular conditions  $|\ell_1 - \ell_2| \leq L = 0$  and  $m_1 + m_2 - M = 0$ , so that we have  $\ell_1 = \ell_2$  and  $m_2 = -m_1$ . We obtain

$$d_0^0 = A_0 \sum_{\ell, m} g_{\ell, \ell}(0) a_{\ell, m} a_{\ell, m}^*, \quad (\text{A.2})$$

where we absorbed multiplicative factors into  $g_{\ell, \ell}(0)$  and used  $a_{\ell, -m} = (-1)^m a_{\ell, m}^*$ . Then we use the definition of the power spectrum estimator in full sky  $\hat{C}_\ell = \frac{1}{2\ell+1} \sum_m a_{\ell, m} a_{\ell, m}^*$ , absorb the multiplicative factors again, and obtain

$$d_0^0 = A_0 \sum_{\ell} g_{\ell, \ell}(0) \hat{C}_\ell. \quad (\text{A.3})$$

The normalisation  $A_0$  has to be chosen such that the estimator is unbiased. In our case,  $\langle \hat{C}_\ell \rangle_{\delta_b \text{ fixed}} = \bar{C}_\ell(\theta) + \delta_b \partial_{\delta_b} C_\ell$ , so

$$\langle d_0^0 \rangle = A_0 \sum_{\ell} g_{\ell, \ell}(0) \bar{C}_\ell(\theta) + \delta_b A_0 \sum_{\ell} g_{\ell, \ell}(0) \partial_{\delta_b} C_\ell. \quad (\text{A.4})$$

Our estimator therefore is

$$\hat{\delta}_b = d_0^0 - A_0 \sum_{\ell} g_{\ell, \ell}(0) \bar{C}_\ell(\theta) \quad (\text{A.5})$$

$$= A_0 \sum_{\ell} g_{\ell, \ell}(0) (\hat{C}_\ell - \bar{C}_\ell(\theta)), \quad (\text{A.6})$$

where the normalisation follows

$$A_0 = \left( \sum_{\ell} g_{\ell, \ell}(0) \partial_{\delta_b} C_\ell \right)^{-1}. \quad (\text{A.7})$$

We are left with optimising the weight  $g_{\ell, \ell}(0)$  (noting that a constant multiplicative factor in its definition is unimportant as it would cancel out in the ratio in Eq. A.6). This optimisation is performed by minimising the variance of the estimator,

$$\text{Var}(d_0^0) = \langle d_0^0 d_0^0 \rangle - \langle d_0^0 \rangle^2 = A_0^2 \sum_{\ell, \ell'} g_{\ell, \ell}(0) g_{\ell', \ell'}(0) C_{\ell, \ell'}, \quad (\text{A.8})$$

where  $C_{\ell, \ell'}$  is the covariance of  $\hat{C}_\ell$ . A significant detail here is that all the brackets mean average at  $\delta_b$  fixed,

$$C_{\ell, \ell'} = \langle \hat{C}_\ell \hat{C}_{\ell'} \rangle_{\delta_b \text{ fixed}} - \langle \hat{C}_\ell \rangle_{\delta_b \text{ fixed}} \langle \hat{C}_{\ell'} \rangle_{\delta_b \text{ fixed}}. \quad (\text{A.9})$$

Hence  $C_{\ell, \ell'}$  is the noSSC part of the covariance.

After some algebra, we find that the minimisation yields

$$\sum_{\ell'} C_{\text{noSSC}; \ell, \ell'} g_{\ell', \ell'}(0) = \partial_{\delta_b} C_\ell \times \text{cst}. \quad (\text{A.10})$$

We can therefore take

$$g_{\ell, \ell}(0) = \sum_{\ell'} C_{\text{noSSC}; \ell, \ell'}^{-1} \partial_{\delta_b} C_{\ell'}. \quad (\text{A.11})$$

This also gives the normalisation of the estimator,

$$A_0 = \left( \sum_{\ell, \ell'} \partial_{\delta_b} C_\ell C_{\text{noSSC}; \ell, \ell'}^{-1} \partial_{\delta_b} C_{\ell'} \right)^{-1}. \quad (\text{A.12})$$

In summary, the quadratic estimator takes the form

$$\hat{\delta}_b = \frac{1}{N} \sum_{\ell} g_{\ell, \ell}(0) (\hat{C}_\ell - \bar{C}_\ell(\theta)), \quad (\text{A.13})$$

with  $N = A_0^{-1}$  given by Eq. A.12 and  $g_{\ell, \ell}(0)$  given by Eq. A.11

## Appendix B: Inverse Fisher matrix in the likelihood approach

In the likelihood approach, the total Fisher matrix  $F' = F(\text{free}\delta_b) + F(\text{prior})$  takes the form

$$F' = \begin{pmatrix} F^{\text{noSSC}} & Y^T \\ Y & I(\delta_b) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{S} \end{pmatrix}. \quad (\text{B.1})$$

For readability, we now write  $I \equiv I(\delta_b)$ . In order to obtain the forecast constraints on  $\theta$  marginalised over  $\delta_b$ , we now need to invert  $F'$  and extract the upper left  $(n_\theta, n_\theta)$  block. For this, we shuffle terms around so that we can write  $F'$  in the form  $F' = F'_{(1)} + F'_{(2)}$  with

$$F'_{(1)} = \begin{pmatrix} F^{\text{noSSC}} & 0 \\ 0 & \frac{1}{S} + I \end{pmatrix}, \quad F'_{(2)} = \begin{pmatrix} 0 & Y^T \\ Y & 0 \end{pmatrix}. \quad (\text{B.2})$$

This makes  $F'_{(1)}$  easy to invert as it is block-diagonal. For the second term, we remark that

$$F'_{(2)} = \begin{pmatrix} Y & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} Y^T & 0 \\ 0 & 1 \end{pmatrix}, \quad (\text{B.3})$$

which puts  $F'$  in a form amenable to using the Woodbury identity one more time. After some algebra, we find

$$F'^{-1} = F'_{(1)}^{-1} + \frac{S}{1 + (I - \delta I)S} \begin{pmatrix} Z^T Z & Z^T \\ Z & \frac{\delta I S}{1 + \delta I S} \end{pmatrix}. \quad (\text{B.4})$$

Shuffling terms around, we obtain the cleaner form

$$F'^{-1} = \begin{pmatrix} F^{\text{noSSC}} & 0 \\ 0 & 0 \end{pmatrix} + \frac{S}{1 + (I - \delta I)S} \begin{pmatrix} Z^T Z & Z^T \\ Z & 1 \end{pmatrix}. \quad (\text{B.5})$$

## Appendix C: Extension of the Sij approximation to probe-independent $\delta_b$ s

A LSS observable  $O(i, \alpha)$  can generally be cast in the form of a light-cone integral of some density  $\nu(i, \alpha)$ ,

$$O(i, \alpha) = \int dV \nu(i, \alpha). \quad (\text{C.1})$$

(We note that for correlation functions of order  $n \geq 2$ , this requires the use of Limber's approximation.)

In this case, the equation for super-sample covariance is given by (e.g. [Lacasa & Rosenfeld 2016](#))

$$\text{Cov}_{\text{SSC}} [\mathcal{O}(i, \alpha), \mathcal{O}(j, \beta)] = \int dV_{12} \frac{\partial \mathcal{O}(i, \alpha)}{\partial \delta_b} (z_1) \frac{\partial \mathcal{O}(j, \beta)}{\partial \delta_b} (z_2) \times \sigma^2(z_1, z_2), \quad (\text{C.2})$$

where  $\sigma^2(z_1, z_2) = \text{Cov} [\delta_b(z_1), \delta_b(z_2)]$  is the covariance of the background shifts of infinitesimal redshift slices. Starting from this equation, [Lacasa & Grain \(2019\)](#) derived the general Sij approximation Eq. 3 by assuming that the responses vary slowly with redshift compared to  $\sigma^2(z_1, z_2)$ . Now this approximation can be refined by cutting the integrals into small pieces and performing the approximation in each piece. Concretely, we start by defining a series of functions  $\Phi_n(z)$  that sum to unity over a redshift range suitable for the integrals for all considered observables,

$$\sum_{n=1}^{n_\Phi} \Phi_n(z) = 1 \quad \forall z \in [z_{\min}, z_{\max}]. \quad (\text{C.3})$$

An example of such  $\Phi_n$  is series of non-intersecting top-hat functions that cover  $[z_{\min}, z_{\max}]$ . Gaussian-like functions are another possibility. They have the advantage of avoiding ringing effects in later integrals.

The LSS observable then takes the form

$$\mathcal{O}(i, \alpha) = \sum_n \mathcal{O}(i, \alpha)_n \text{ with } \mathcal{O}(i, \alpha)_n = \int dV \Phi_n(z) \mathcal{O}(i, \alpha), \quad (\text{C.4})$$

and its super-sample covariance is

$$\text{Cov}_{\text{SSC}} [\mathcal{O}(i, \alpha), \mathcal{O}(j, \beta)] = \sum_{m,n} \text{Cov}_{\text{SSC}} [\mathcal{O}(i, \alpha)_m, \mathcal{O}(j, \beta)_n],$$

where we can apply the Sij approximation to  $\text{Cov}_{\text{SSC}} [\mathcal{O}(i, \alpha)_m, \mathcal{O}(j, \beta)_n]$ :

$$\begin{aligned} \text{Cov}_{\text{SSC}} [\mathcal{O}(i, \alpha)_m, \mathcal{O}(j, \beta)_n] &= \int dV_{12} \Phi_m(z_1) \Phi_n(z_2) \\ &\times \frac{\partial \mathcal{O}(i, \alpha)}{\partial \delta_b} (z_1) \frac{\partial \mathcal{O}(j, \beta)}{\partial \delta_b} (z_2) \\ &\times \sigma^2(z_1, z_2) \quad (\text{C.5}) \\ &\approx \partial_{\delta_b} \mathcal{O}(i, \alpha)_m \partial_{\delta_b} \mathcal{O}(j, \beta)_n S_{m,n}, \quad (\text{C.6}) \end{aligned}$$

with

$$S_{m,n} = \int dV_1 dV_2 \frac{\Phi_m(z_1)}{\Delta V_m} \frac{\Phi_n(z_2)}{\Delta V_n} \sigma^2(z_1, z_2), \quad (\text{C.7})$$

and

$$\Delta V_m = \int dV \Phi_m(z). \quad (\text{C.8})$$

where we assumed that the  $\Phi_n$  functions have sufficiently narrow support for probe-specific kernels to be considered constant. In practical tests, we found that a width  $\Delta z \sim 0.1$  is small enough. The SSC matrix thus has the same form as Eq. 5,

$$C_{\text{SSC}} = U S U^T, \quad (\text{C.9})$$

where now  $U$  is a dense rectangular matrix with the shape  $(n_o, n_\Phi)$ ,

$$U_{i_o, n} = \partial_{\delta_b} \mathcal{O}(i, \alpha)_n. \quad (\text{C.10})$$

Thus all the machinery defined in the main text of the article can be used here, as well; Eq. 7 for the inverse covariance, Eq. 19 for the likelihood, and Eq. 24 and Eq. 30 for the Fisher and inverse Fisher matrices.

The key difference is that the  $S_{m,n}$  matrix is now independent of the considered physical probe; indeed, it is the covariance matrix of probe-independent  $\delta_b$  parameters,

$$S_{m,n} = \text{Cov} [\delta_{b,m}, \delta_{b,n}] \text{ with } \delta_{b,n} = \int dV \Phi_n(z) \delta_b(z) \quad (\text{C.11})$$

In the likelihood approach to SSC, this  $\delta_{b,n}$  is the vector of nuisance parameters over which is to be marginalised. The difference with the main text of the article is that this set of parameters is now independent of the physical probe and does not grow when probes are combined. The number of these parameters is defined by the width of the  $\Phi_n$  functions, and optimisation can be studied to keep this width as large as possible while maintaining the validity of the Sij approximation.

According to the results of [Lacasa & Grain \(2019\)](#), a width  $\Delta z = 0.1$  is already small enough to maintain the validity of the approximation. For a stage IV galaxy survey out to  $z_{\max} = 2$ , 20 parameters or even fewer are therefore enough to analyse all physical probes. This compares favourably to the 55  $\delta_b$  parameters of the WL-only case in the main text or to the 210 parameters of the 3x2pt case. Moreover, this number is small enough not to be a limiting factor in an MCMC analysis, especially in the context of a combined probe analysis where the number of other nuisance parameters is already high. The only downside to this approach is that the pipelines need to be modified to output all the  $\partial_{\delta_b} \mathcal{O}(i, \alpha)_m$ , an output with a size 20 times the size of the observable prediction. We note, however, that this needs to be done in general only for the fiducial cosmology for which the covariance is to be computed.