Optimized angular power spectra for spectroscopic galaxy surveys

Stefano Camera,[1,2,3,4,5]★ José Fonseca,[4] Roy Maartens[4,6] and Mário G. Santos[4,7]

[1]*Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy*
[2]*INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy*
[3]*INAF – Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Strada Osservatorio 20, 10025 Pino Torinese, Italy*
[4]*Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa*
[5]*Jodrell Bank Centre for Astrophysics, The University of Manchester, Alan Turing Building, Oxford Road, Manchester M13 9PL, UK*
[6]*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*
[7]*SKA South Africa, The Park, Cape Town 7405, South Africa*

ABSTRACT

The angular power spectrum $C_\ell(z_i, z_j)$ is a gauge-independent observable that is in principle the natural tool for analysing galaxy number counts. In practice, the problem is that the computational requirements for next-generation spectroscopic surveys such as *Euclid* and the Square Kilometre Array are currently unfeasible. We propose a new method to save computational time for spectroscopic angular power spectra. This hybrid method is modelled on the Fourier power spectrum approach of treating relatively thick redshift bins ($\Delta z \sim 0.1$) as separate surveys. In the hybrid method, each thick bin is further subdivided into thin bins ($\delta z \sim 0.01$); all the correlations within each thick bin are computed, whilst cross-bin correlations beyond the thick bins are neglected. Constraints on cosmological parameters from the hybrid method are comparable to those from the standard galaxy $P_g(\boldsymbol{k}, z)$ analysis – but they have the advantage that cosmic evolution, wide-angle, and lensing effects are naturally included, whilst no Alcock–Paczynski correction is needed. The hybrid method delivers much tighter constraints than a 2D tomographic approach that is typical for photometric surveys, which considers only thick bins and the correlations between them. Furthermore, for standard cosmological parameters our method is not biased by neglecting the effects of lensing on number counts, whilst the tomographic method is strongly biased.

Key words: *(cosmology:)* cosmological parameters – cosmology: observations – cosmology: theory – *(cosmology:)* large-scale structure of Universe.

1 INTRODUCTION

The concordance cosmological model Λ cold dark matter (ΛCDM) represents the current best fit to a number of very different observations across a wide range of physical scales and cosmic time. Temperature anisotropies of the cosmic microwave background measured by the *Planck* satellite (Ade et al. 2015), combined with many other data sets, have shown that present-day data do not favour any extension to ΛCDM. Nevertheless, key questions remain to be answered, like what drives the late-time cosmic accelerated expansion, or whether there is non-Gaussianity in the primordial density fluctuations. Furthermore, some tensions among data sets – in particular, between high- and low-redshift observables – have been claimed to herald breaches in the adamantine ΛCDM model (Battye, Charnock & Moss 2015; Spergel, Flauger & Hlozek 2015; Addison et al. 2016; Joudaki et al. 2017a,b; Pourtsidou & Tram

2016; Raveri 2016; An, Feng & Wang 2017; Camera, Martinelli & Bertacca 2017a; Charnock, Battye & Moss 2017). All of these calls for a better understanding of the late-time Universe and the growth and clustering of cosmic structures, from ultra-large to mildly non-linear scales.

The main envisaged methods to probe the large-scale structure are measurements of cosmic shear and the creation of huge galaxy catalogues with which we can reconstruct the clustering of dark matter haloes over a wide range of scales and redshifts. In this paper, we focus on the latter method, namely on the measurement of the power spectrum of galaxy number counts, although we will comment on cosmic shear in the conclusions (Section 5). Up to now, power spectrum measurements have mainly relied on two approaches: reconstructing baryon acoustic oscillation (BAO) peaks and redshift-space distortions (RSD) from the three-dimensional (3D) galaxy positions in a spectroscopic redshift survey, or estimating the two-dimensional (2D) clustering of galaxies in redshift slices from broad-band photometric measurements. Despite the fact that photometric galaxy catalogues are significantly more populated than

★ E-mail: stefano.camera@unito.it

spectroscopic ones, the former method is more informative, since it gives access to the cosmological information encoded in the fully 3D cosmic web. However, it relies on a number of assumptions that require various add-on techniques in order to exploit the data quality and sky and redshift coverage envisaged by next-generation experiments, as we discuss in Section 3.1.

In this paper, we introduce a new method to optimize angular power spectrum computations for spectroscopic galaxy surveys. This method is in some sense a combination of the galaxy Fourier power spectrum $P_g(\boldsymbol{k}, z)$ approach (discussed in Section 3.1) and a 2D tomographic approach (outlined in Section 3.2). For this reason, we refer to the new method as 'hybrid'. In Section 4, we present our main results, showing how our hybrid approach yields constraints on cosmological parameters that are comparable to those from a standard $P_g(\boldsymbol{k}, z)$ analysis, and are more than twice as tight as those obtained with a standard tomographic analysis. This is further supported by a study of the 'information gain' earned by going from the standard tomography to the hybrid method.

We also demonstrate that our hybrid approach is robust with respect to not including the corrections to galaxy number counts due to gravitational lensing. Specifically, we show that best-fit values of cosmological parameters estimated via density fluctuations and RSD alone – namely, neglecting lensing – are biased by less than 20 per cent of the 1σ error on the parameter.[1] This is important, since the numerical computation of the lensing correction, being an integrated effect, is significantly slower. Hence, our method allows for faster computation of the data likelihood and is then more suitable for implementation in Markov chain Monte Carlo pipelines, relevant in view of upcoming spectroscopic galaxy surveys, such as the European Space Agency *Euclid* satellite (Laureijs et al. 2011; Amendola et al. 2013, 2018) or the Square Kilometre Array (SKA; Abdalla et al. 2015; Maartens et al. 2015). Since our method follows one key aspect of the $P_g(\boldsymbol{k}, z)$ approach, i.e. treating thick bins as separate surveys, this suggests that the Fourier power spectrum method might not be biased by excluding lensing effects.

2 METHODOLOGY

2.1 Statistical tools

We consider a six-dimensional cosmological parameter set $\boldsymbol{\vartheta} = \{\Omega_b, \Omega_{DM}, n_s, \mathcal{A}_s, H_0\}$, with Ω_b and Ω_{DM} the baryonic and dark matter density fractions, n_s and A_s the slope and the amplitude of the primordial power spectrum measured at some pivot scale k_0, with $\mathcal{A}_s = \ln(10^{10} A_s)$, and $H_0 = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ the Hubble constant. Fiducial values for the parameters are $\overline{\boldsymbol{\vartheta}} = \{0.05, 0.26, 0.9667, 3.06, 67.74\}$.

We work in a Bayesian statistics framework, and make use of Fisher matrices to forecast the capabilities of future surveys (e.g. Trotta 2008). All relevant formulas can be found in Appendix A1, together with a description of the tests we performed to ensure the stability of the matrices and the robustness of our results. We use 'relative errors' to denote marginal errors $\sigma(\vartheta_\alpha)$ divided by parameter fiducial values $\overline{\vartheta}_\alpha$, and 'relative biases' on parameters to mean biases on cosmological parameters in units of the measurement precision, i.e. $b(\vartheta_\alpha)/\sigma(\vartheta_\alpha)$.

[1] Note that this does not apply to the non-Gaussianity parameter f_{NL}, whose best fit is biased by neglecting lensing (see Section 4.4).

2.2 Observational assumptions

To compare results from our optimized hybrid method to the standard 2D tomographic approach, we apply both methods to the same observational set-up, namely a Stage IV Dark Energy Task Force (DETF) cosmological experiment (see Albrecht et al. 2006). To compute the redshift distribution of sources and magnification bias, we use model 3 of Pozzetti et al. (2016). For this work, we assume a flux cut threshold for the survey $\mathcal{F}_* = 3 \times 10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ in the redshift range $0.6 \leq z \leq 2$. Using this assumption, we find the following fits for the spectroscopic distribution of sources per solid angle per unit redshift:

$$\mathrm{d}N_g/\mathrm{d}z = z^{1.281}$$
$$\times \exp\left(9.976 - 2.317z - 0.617z^2 + 0.265z^3 - 0.030z^4\right), \tag{1}$$

and for the magnification bias

$$\mathcal{Q} = -0.66 - 2.95z + 1.59z^2 - 0.40z^3 + 0.04z^4, \tag{2}$$

where sometimes the magnification bias is referred to as s, and $\mathcal{Q} = 5s/2$ holds. Note that these fits are only valid for $z \in [0.6, 3]$ and do not reproduce well the results out of this range. We model the galaxy bias as $b_g(z) = \sqrt{1+z}$ (Amendola et al. 2013, 2018), and assume a sky area of $15,000\,\mathrm{deg}^2$. All angular power spectra are computed with the public code CAMB_sources (Challinor & Lewis 2011) using the distribution of sources given by equation (1).

When we quote results for standard tomography (see Section 3.2), we divide the source redshift distribution into 20 equi-populated, top-hat redshift bins, with edges smeared by a Gaussian window with $\sigma_z = 0.002$. For our hybrid method, we follow the recipe outlined in Section 3.3.

As a final remark, we employ the linear matter power spectrum. Depending on the maximum angular wavenumber considered, this may not be correct at low redshifts, since nonlinear scales could already be contributing to the angular power spectrum. This can be seen via the Limber approximation, valid for small angles, $k = \ell/\chi$. At low redshift, $\ell_{max} = 800$ already exceeds the nonlinear threshold $k_{nl} \simeq 0.2\,h\,\mathrm{Mpc}^{-1}$. Nonetheless, our aim is not to provide specific forecasts for a given experiment, but rather to compare our method with a standard tomographic analysis, with both methods using the linear matter power spectrum and the same survey specifications. The inclusion of nonlinearities does not add relevant information and can be disregarded without loss of generality.

3 GALAXY NUMBER COUNTS

First, we briefly describe two of the main approaches to galaxy clustering data: the Fourier power spectrum, usually employed with spectroscopic data; and the 2D tomographic angular power spectrum, usually used in photometric surveys. Secondly, we introduce the new hybrid approach towards optimal angular power spectra for spectroscopic galaxy surveys.

3.1 Fourier power spectrum

Neglecting RSD and the Alcock–Paczynski (AP) effect for simplicity, the galaxy number density at real-space position $\boldsymbol{x}$ is

$$n_g(\boldsymbol{x}, z) = \bar{n}_g(z)\left[1 + \delta_g(\boldsymbol{x}, z)\right], \tag{3}$$

where $\bar{n}_g(z)$ is the mean galaxy number density at redshift z, the fluctuations δ_g have zero mean and same-time power spectrum

$$\left\langle \delta_{\mathrm{g}}(\boldsymbol{k}, z)\delta_{\mathrm{g}}^*(\boldsymbol{k}', z)\right\rangle = (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k} - \boldsymbol{k}')P_{\mathrm{g}}(k, z), \quad \text{where} \quad P_{\mathrm{g}}(k, z) =$$
$b_{\mathrm{g}}^2(z)P_\delta(k, z)$, $b_{\mathrm{g}}(z)$ is the galaxy bias, and $P_\delta(k,z)$ the linear matter power spectrum, which is related to the (dimensionless) primordial power spectrum of curvature fluctuations, $\Delta_\zeta^2(k) = A_{\mathrm{s}}(k/k_0)^{n_{\mathrm{s}}-1}$, through

$$P_\delta(k, z) = \frac{8\pi^2}{25}\frac{k_0 A_{\mathrm{s}}}{H_0^4\Omega_{\mathrm{m}}^2 g_\infty^2}D_+^2(z)T_{\mathrm{m}}^2(k)\left(\frac{k}{k_0}\right)^{n_{\mathrm{s}}}. \tag{4}$$

Here, $T_{\mathrm{m}}(k)$ is the matter transfer function (normalized so that $T_{\mathrm{m}} \to 1$ for $k \to 0$), $D_+(z)$ is the growth factor of matter density fluctuations (normalized so that $D_+ \to 1$ for $z \to 0$), and $g_\infty = \lim_{z \to \infty}(1 + z)D_+(z) \approx 1.27$. The usual way of assessing the constraining power of a spectroscopic galaxy redshift survey is via the Fisher matrix, defined e.g. in Seo & Eisenstein (2003) as

$$\mathbf{F}\left(\vartheta_\alpha, \vartheta_\beta\right) = \sum_{i=1}^{N_{\mathrm{b}}}\frac{1}{4\pi^2}V(z_i)f_{\mathrm{sky}}$$
$$\times \int_{k_{\min}}^{k_{\max}}\mathrm{d}k\, k^2\frac{\partial P_{\mathrm{g}}(k, z_i)}{\partial\vartheta_\alpha}\frac{\partial P_{\mathrm{g}}(k, z_i)}{\partial\vartheta_\beta}\left[P_{\mathrm{g}}(k, z_i) + \frac{1}{\bar{n}_i}\right]^{-2}. \tag{5}$$

In this equation: i labels the redshift bin centred in z_i; N_{b} is the number of redshift bins considered; $V(z_i)$ is the cosmological volume in bin i; f_{sky} is the fraction of sky surveyed; and $\bar{n}_i$ is the *volumetric* number density of galaxies in the ith bin. In practice, in each redshift bin we count all the k-modes contained within the volume $V(z_i)f_{\mathrm{sky}}$, with an uncertainty accounting for both cosmic variance and the effect of discrete Poisson sampling.

The method relies on a number of assumptions, the most important being

(i) all quantities are assumed constant within the redshift bin;
(ii) no correlation among redshift bins is considered;
(iii) it is valid only in the flat-sky limit;
(iv) it does not include the effect of lensing.

A number of works in the recent literature have tried to overcome some of these limitations. For instance, Ruggeri et al. (2017) apply a set of weights to extract RSD measurements as a function of redshift, acknowledging that future surveys covering a broad redshift range can no longer ignore cosmic evolution. Bailoni, Mancini & Amendola (2017) improve upon the standard 3D $P_{\mathrm{g}}(\boldsymbol{k}, z)$ Fisher method by taking into account three effects: the finite window function, the correlation between redshift bins and the uncertainty on redshift estimation. Gil-Marín et al. (2016) use a line-of-sight dependent power spectrum to deal with the large sky coverage of the Sloan Digital Sky Survey (SDSS) DR12 catalogue. Similarly, Blake, Carter & Koda (2018) study power spectrum multipoles in a curved sky for the 6-degree Field Galaxy Survey (6dFGS), and indicate how these techniques may be extended to studies of overlapping galaxy populations via multipole cross-power spectra. However, to our knowledge none of these methods includes lensing and none simultaneously addresses all of the three other issues.

Tansella et al. (2018) derive an exact expression for the galaxy correlation function in redshift shells – including lensing – thus addressing all four issues listed above. Their approach is an alternative to ours and is based on the correlation function. However, they do not give forecasts to compare their method with previous work.

3.2 Tomographic angular power spectrum

Uncertainties in photometric measurements prevent us from mapping angles $\hat{\boldsymbol{n}}$ and redshifts to 3D positions $r[\chi(z), \hat{\boldsymbol{n}}]$, which we

would, in turn, translate into $\boldsymbol{k}_\perp$ and $k_{\|}$. Instead, we have to rely on angular power spectra, where the radial information is averaged over a redshift bin, taking into account probability density functions of photometric redshifts. This leads to a tomographic matrix with entries

$$C_\ell^{ij} = 4\pi\int\mathrm{d}\ln k\, \mathcal{W}_\ell^i(k)\mathcal{W}_\ell^j(k)\Delta_\zeta^2(k), \tag{6}$$

where $\mathcal{W}_\ell^i(k)$ is the transfer function of the ith bin, containing weights for all the terms included, i.e. Newtonian density fluctuations, RSD, lensing and possibly other relativistic contributions (see e.g. Camera, Santos & Maartens 2015a). Weak lensing convergence, modulated by the magnification bias, gives the strongest relativistic correction (Fonseca et al. 2015; Raccanelli et al. 2016). We thus limit our analysis to this effect only, beyond density fluctuations and RSD.

The angular power spectrum transfer functions of equation (6) can be rewritten as

$$\mathcal{W}_\ell^i(k) = \int\mathrm{d}\chi\,\frac{\mathrm{d}N_{\mathrm{g}}^i}{\mathrm{d}\chi}\mathcal{W}_\ell(k, \chi), \tag{7}$$

where we remind the reader that $(\mathrm{d}N_{\mathrm{g}}^i/\mathrm{d}\chi)\mathrm{d}\chi = (\mathrm{d}N_{\mathrm{g}}^i/\mathrm{d}z)\mathrm{d}z$. In longitudinal gauge,

$$\mathcal{W}_\ell(k, \chi) = b_{\mathrm{g}}(\chi)\delta_k(\chi)j_\ell(k\chi) + \frac{k}{\mathcal{H}(\chi)}v_k(\chi)j_\ell''(k\chi)$$
$$+ 2\left[\mathcal{Q}(\chi) - 1\right]\ell(\ell + 1)\kappa(\chi). \tag{8}$$

The first line is the main contribution, from density perturbations (δ is the comoving matter density contrast, necessary for a physical model of bias[2]) and RSD (v is the peculiar velocity), and the second line is the lensing contribution, proportional to the weak lensing convergence

$$\kappa(\chi) = \frac{1}{2}\int_0^\chi\mathrm{d}\tilde{\chi}\,\frac{(\chi - \tilde{\chi})}{\chi\tilde{\chi}}\left[\Phi_k(\tilde{\chi}) + \Psi_k(\tilde{\chi})\right]j_\ell(k\tilde{\chi}), \tag{9}$$

with Φ and Ψ the two gauge-invariant Bardeen potentials. A symbol with index k represents the Fourier transform of the corresponding configuration-space quantity.

For the tomographic case, the Fisher matrix takes the form (summation over same indexes is assumed)

$$\mathbf{F}\left(\vartheta_\alpha, \vartheta_\beta\right) = \sum_{\ell=\ell_{\min}}^{\ell_{\max}}\frac{2\ell + 1}{2}f_{\mathrm{sky}}\frac{\partial C_\ell^{ij}}{\partial\vartheta_\alpha}\left[\widetilde{\mathbf{C}}_\ell^{-1}\right]_{jm}\frac{\partial C_\ell^{mn}}{\partial\vartheta_\beta}\left[\widetilde{\mathbf{C}}_\ell^{-1}\right]_{ni}, \tag{10}$$

where $\widetilde{\mathbf{C}}_\ell^{-1}$ denotes the inverse of the observed tomographic matrix, $\widetilde{\mathbf{C}}_\ell$, whose j-m entry reads

$$\widetilde{C}_\ell^{jm} = C_\ell^{jm} + \delta^{jm}\frac{1}{\tilde{n}_j}, \tag{11}$$

with $\tilde{n}_j$ is the *angular* number density of galaxies per steradian in bin j.

3.3 Hybrid angular power spectrum

In principle, the correct way to analyse spectroscopic galaxy clustering is via the angular power spectra C_ℓ^{ij}, including all cross-bin correlations. These power spectra are a physical and gauge-independent

---

[2]The correction term from using the comoving δ is omitted since it is negligible on sub-Hubble scales.
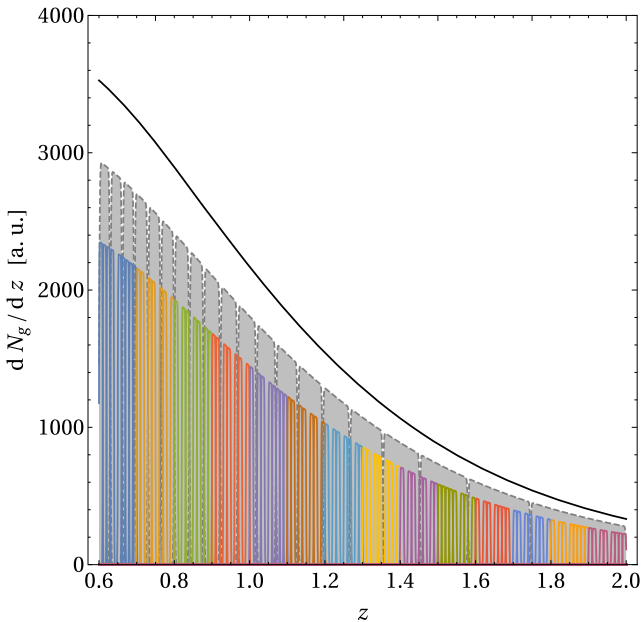
Figure 1. Estimated redshift distribution of spectroscopically selected H α galaxies with flux $> 3 \times 10^{-16}$ erg cm^{-2} s^{-1} (from Pozzetti et al. 2016). The top black curve shows the unbinned dN_g/dz, as in equation (1). The shaded grey regions bounded by dashed lines are the 20 equi-populated bins of a standard tomographic analysis (Section 3.2). The bundles of coloured curves depict the 14 equi-spaced coarse redshift bins with width $\Delta z = 0.1$, in turn subdivided into 10 thinner top-hat bins of width $\delta z = 0.01$, finally convolved with a Gaussian ($\sigma_z = 0.001$), as in Section 3.3. (Note that all distributions have been rescaled by an arbitrary factor to enhance readability.)

representation of correlations in galaxy number counts, whilst the Fourier power spectrum is not an observable in itself, since it is gauge dependent (e.g. Bonvin & Durrer 2011). In practice, however, this is computationally unfeasible, especially since many realizations are required for simulated data. Our new method for computing angular power spectra for spectroscopically selected galaxies does not rely on the computation of an extremely large number of auto- and cross-correlations between redshift bins, but is still capable of constraining cosmological parameters with an accuracy comparable to that of a Fourier $P_g(k, z)$ analysis (see Section 4.1).

For the $P_g(k, z)$ approach in Section 3.1 (including RSD and AP corrections), one usually takes a rather thick redshift bin – with width $\sim\Delta z = 0.1$ – and then counts all 3D k modes within the bin, disregarding correlations among all other redshift bins. This effectively renders the covariance matrix diagonal in redshift. On the other hand, if we were to do the same with the 2D tomographic approach of Section 3.2, we would lose substantial information by squashing all the galaxies contained within the Δz bin on to a single redshift slice.

This motivates a hybrid approach: divide the redshift distribution into $\Delta z = 0.1$ bins as in the $P_g(k, z)$ approach; subdivide each bin into 10 top-hat bins with width $\delta z = 0.01$; and eventually convolve the thin bins with a Gaussian with spread $\sigma_z = 0.001$ to account for the small but non-negligible errors in the spectroscopic redshift estimation.

The set-up is illustrated in Fig. 1, where

(i) we start from the redshift distribution in equation (1) of galaxies expected to be detected by a Stage IV cosmological survey with

H α flux $> 3 \times 10^{-16}$ erg cm^{-2} s^{-1} (upper, black curve, as in Model 3 of Pozzetti et al. 2016);

(ii) we first bin galaxies into 14 equi-spaced, coarse redshift bins (groups of curves with the same colour);

(iii) we subdivide the coarse bins into 10 thinner top-hat bins;

(iv) we finally convolve the thinner bins with the $\sigma_z = 0.001$ Gaussian (a total of 140 distinct, slightly overlapping curves).

We effectively consider each thick bin as a separate survey, using the full spectroscopic angular power spectra in each thick bin, by applying equation (10) with respect to its sub-bins. Then, as in equation (5), we sum the 14 Fisher matrices thus obtained. This hybrid approach stems from the more physically motivated angular power spectra, but resembles what is done with the $P_g(k, z)$ analysis, since it preserves radial information within the thick redshift bins by considering all the cross-correlations between the thin sub-bins.

The hybrid method is far less computationally onerous than the full 3D angular power spectra, which involve application of equation (10) to a spectroscopic tomographic matrix with bin width 0.01. For $0.6 \leq z \leq 2$, this involves a total of 9870 spectra, between auto- and cross-bin correlations. By contrast, our hybrid method only requires 770 of them, more than one order of magnitude fewer.

In principle, by considering each thick bin as an independent survey, we could neglect valuable information encoded in cross-correlations among distant redshift bins, or between thin bins at the edges of the thick bins. This is what is usually done in observational $P_g(k, z)$ analyses, but it is not clear how one can assess the loss of information in that case (for a possible solution, see e.g. Bailoni et al. 2017). In our case, we can assess the effects of neglecting such cross-bin correlations by looking at the correlation coefficient between bins. If we consider $C_\ell(z_1, z_2)$, the angular cross-power spectrum of galaxy number counts between redshifts z_1 and z_2, integrated over a thin slice of width $\delta z = 0.01$, we can define the correlation coefficient as

$$r_\ell(z_1, z_2) \equiv \frac{C_\ell(z_1, z_2)}{\sqrt{C_\ell(z_1, z_1) C_\ell(z_2, z_2)}}, \tag{12}$$

where we make explicit the dependence on the angular scale, ℓ. Alternatively, we can redefine $r_\ell(\bar{z}, \Delta z)$ for two bins centered in $\bar{z}$ and separated by Δz, with z_1 and z_2 of equation (12) thus becoming $\bar{z} - \Delta z/2$ and $\bar{z} + \Delta z/2$, respectively. We emphasize that here Δz does not refer to the width of the thick bins introduced above, but we nevertheless employ this slight abuse of notation for a reason that will become clear when looking at Fig. 2. There, we plot $r_\ell(\bar{z}, \Delta z)$ as a function of the redshift separation between bins of width $\delta z = 0.01$, centred in redshifts $\bar{z} = 0.8, 1.2, 1.6$, and 2.0, for angular scales $\ell = 5$ and 100. It is clear that the correlation coefficient quickly falls as a function of Δz, thus indicating that most of the information comes from correlations at small radial separation.

The take-home message is as follows. We have assumed that all correlations carry similar information about the parameters we are interested in, and treated each thick $\Delta z = 0.1$ slice as an independent survey. Fig. 2 shows that this a reasonable assumption. Nevertheless, we shall come back later to this for further checks (see Section 4.5).

4 RESULTS AND DISCUSSION

4.1 Constraints on cosmological parameters

First of all, we focus on the constraining power of the hybrid approach on ΛCDM cosmological parameters. As explained above, we limit our analysis to linear scales and set the maximum multipole
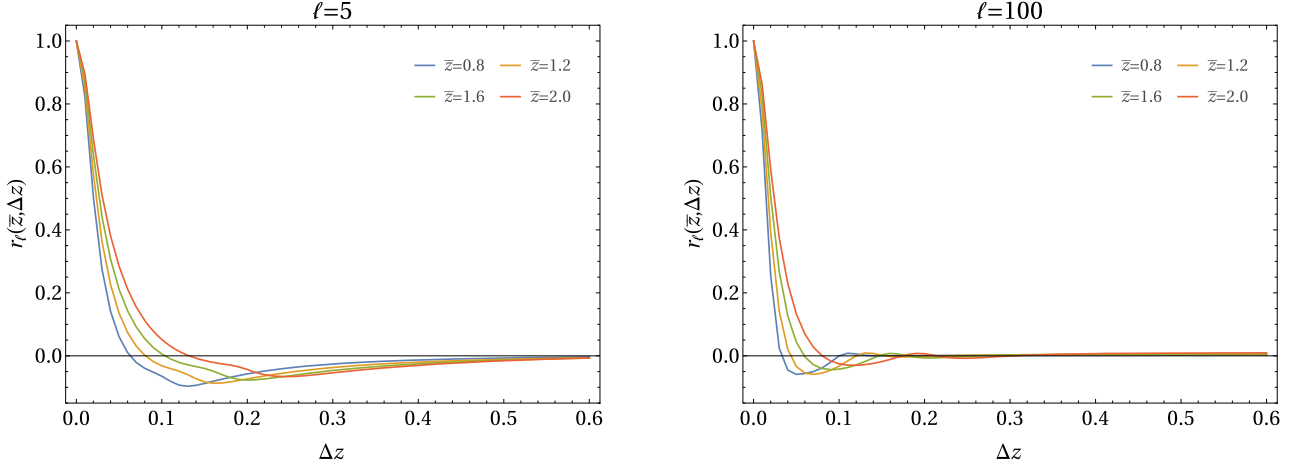
Figure 2. Correlation coefficient $r_\ell(\bar{z}, \Delta z)$ as a function of redshift separation between bins centred in $\bar{z} = 0.8$, 1.2, 1.6, and 2.0, for two reference multipoles, $\ell = 5$ and $\ell = 100$.
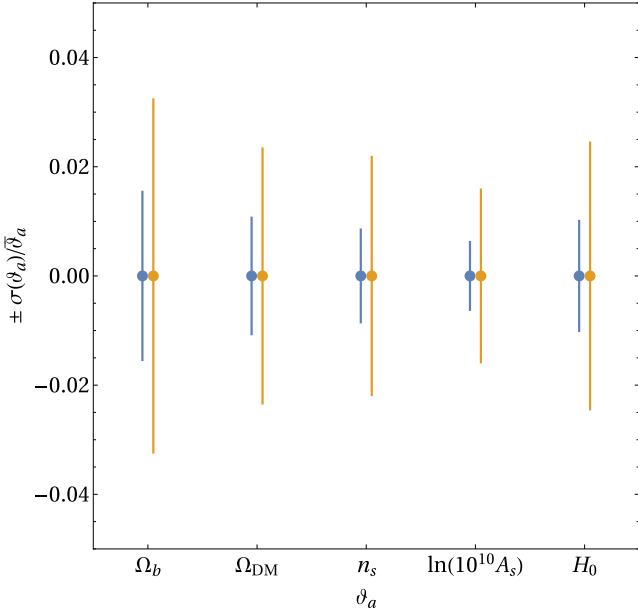

Figure 3. Relative errors on cosmological parameters obtained with our hybrid binning (blue) compared to standard tomography (yellow).

for C_ℓ^{ij} to

$$\ell_{\max}(z_i, z_j) = \min[\chi(z_i), \chi(z_j)] k_{\mathrm{nl}}. \tag{13}$$

The maximum angular scale is determined by $4\pi f_{\mathrm{sky}}$, i.e. 15, 000 deg^2. Thus, $\ell_{\min} = 2$.

In Fig. 3, we compare the constraining power of our hybrid approach (blue error bars) to that of a standard tomographic analysis (yellow error bars). Each cosmological parameter ϑ_α is rescaled by its fiducial value $\bar{\vartheta}_\alpha$ to focus more easily on the relative tightness of the forecast 1σ constraints. It is clear that the finer binning of our hybrid method retrieves more information from the spectroscopic galaxy clustering data even on strictly linear scales, where for instance the largest angular multipole allowed is $\ell_{\max}(z_{140}, z_{140}) = 718$.

So far, we have focused on the cosmological parameters, disregarding any uncertainty on survey specifics such as the galaxy bias. Now we introduce nuisance parameters by adding one parameter per redshift bin, thus allowing for a freely-varying bias amplitude in each bin. We emphasize that this is a conservative approach, as it does not make use of any prior knowledge on the bias. As a result, the standard 2D tomographic approach has 20 nuisance parameters, which are in turn marginalized over (see Section A1). For the hybrid approach, the situation is slightly different. As each coarse redshift bin is further subdivided into 10 thinner bins, a total of 140 nuisance parameters is included. However, each $\Delta z = 0.1$ thick bin is effectively treated as independent from the others, as in the $P_{\mathrm{g}}(k, z)$ case of equation (5). Thus, marginalization over nuisance parameters is performed separately for each Fisher matrix pertaining to a specific thick bin, and only at the end are all the marginal Fisher matrices summed up to give the final cosmological parameter Fisher matrix. For simplicity, we fix the smallest angular scale to a nominal $\ell_{\max} = 800$.

Table 1 shows the resulting 1σ relative marginal errors on ΛCDM cosmological parameters. We find that our constraints are comparable to those from a standard $P_{\mathrm{g}}(k, z)$ analysis (including AP corrections): this can be seen by comparing the last row of Table 1 with the fourth column of Table 2 in Bailoni et al. (2017), who use *Euclid* specifications that are the same as those in Section 2.2, with the same number of thick bins. The first row of Table 1 shows constraints from standard broad-bin tomography. The hybrid approach is approximately twice as constraining as a standard tomographic binning, despite the larger number of nuisance parameters.

4.2 Bias on parameter estimation from neglecting lensing

As discussed in Section 3.1, future spectroscopic galaxy surveys will cover an unprecedented redshift range, so that the effect of weak lensing convergence on number counts needs to be assessed. This is straightforward to do with angular power spectra, which naturally include the weak lensing contribution. This contribution is computationally intensive, and here we evaluate the consequences of ignoring it in the interests of speeding up computations.

Following Camera et al. (2017b) and Fonseca et al. (2015), we introduce a 'fudge' factor parametrizing the *theoretical systematic effect* represented by neglecting the lensing contribution in galaxy count angular power spectra. Then equation (8) becomes

$$\mathcal{W}_\ell = b_{\mathrm{g}} \delta_k j_\ell + \frac{k}{\mathcal{H}} v_k j_\ell'' + 2 A_\kappa (\mathcal{Q} - 1) \ell(\ell + 1) \kappa, \tag{14}$$

Table 1. Forecast 1σ relative errors on cosmological parameters, after marginalizing over nuisance bias parameters, for the two approaches.

$(\ell_{\min}, \ell_{\max}) = (2, 800)$	$\sigma(\Omega_b)/\overline{\Omega}_b$ (per cent)	$\sigma(\Omega_{DM})/\overline{\Omega}_{DM}$ (per cent)	$\sigma(n_s)/\overline{n}_s$ (per cent)	$\sigma(\mathcal{A}_s)/\overline{\mathcal{A}}_s$ (per cent)	$\sigma(H_0)/\overline{H}_0$ (per cent)
Standard tomography (Section 3.2)	4.8	3.3	2.5	2.2	3.1
Hybrid tomography (Section 3.3)	2.5	1.7	1.1	0.9	1.6

Table 2. Forecast relative biases on cosmological parameters, after marginalizing over nuisance bias parameters, for the two approaches.

$(\ell_{\min}, \ell_{\max}) = (2, 800)$	$b(\Omega_b)/\sigma(\Omega_b)$	$b(\Omega_{DM})/\sigma(\Omega_{DM})$	$b(n_s)/\sigma(n_s)$	$b(\mathcal{A}_s)/\sigma(\mathcal{A}_s)$	$b(H_0)/\sigma(H_0)$
Standard tomography (Section 3.2)	8.4	11.1	−8.9	0.2	6.0
Hybrid tomography (Section 3.3)	0.11	−0.22	0.08	0.05	−0.10

where the parameter A_κ has fiducial value $\overline{A}_\kappa = 1$ when lensing is correctly included in the analysis and is 0 when lensing is neglected. Note that intermediate values of A_κ have a physical meaning if we wish to account for an uncertainty in the magnification bias term $(\mathcal{Q} - 1)$, which modulates the contribution from lensing convergence.

The Fisher matrix technique for computing bias in parameters is outlined in the Appendix A2. In the present case, the vector of shifts $\delta\varphi_a$ reduces to $\delta A_\kappa \equiv (\overline{A}_\kappa - 0) = 1$. Table 2 shows the bias on cosmological parameters in units of standard deviations for a standard tomographic analysis (first row) and when implementing our hybrid approach (second row). As expected, standard tomography is very sensitive to lensing effects, because its wider redshift bins acquire a significant contribution from the integrated effect of lensing. On the other hand, the thinner the bin, the less important is lensing (see also Villa, Di Dio & Lepori 2018). Thus, our hybrid approach, with its thin sub-bins, is almost insensitive to it, and we can therefore safely ignore the lensing contribution in the hybrid approach. A major advantage is that this allows for a significant speed-up of computations.

4.3 Information gain

An alternative way to assess the enhancement in constraining power delivered by the hybrid approach over standard 2D tomography is a statistical tool called 'information gain'. When we compare the parameter posterior distributions from two experiments, $p_1(\vartheta)$ and $p_2(\vartheta)$, the information gain in going from p_1 to p_2 is $D(p_2||p_1)$. Information gain (also known as 'relative entropy') was originally motivated by information theory, but it can be used to compute the information gained by Bayesian updates in units of bits (see Appendix A3).

One of the most useful properties of information gain is that it is invariant under invertible transformations in the random variable ϑ. In other words, it represents a more agnostic way to compare two experimental set-ups, because re-parametrizations of the parameter set $\vartheta \equiv \{\vartheta_\alpha\}$ do not affect the information gain. This frees us from misleading interpretations of forecast parameter constraints, which can look tighter or looser depending on the choice of the parameter basis.

For two alternative ways to analyse spectroscopic galaxy surveys, the correct way to use information gain is to compare the gain the two methods (i.e. 2D tomography and our hybrid approach) have over a common prior information on cosmological parameters

(Grandis et al. 2016). Namely, if $p_\Pi(\vartheta)$ is the prior of the parameter set, we compute $D(p_{2D}||p_\Pi)$ and $D(p_{hyb}||p_\Pi)$ separately, and we then compare the incremental gain.

We adopt Gaussian priors on the cosmological parameter set described by a Fisher matrix $\mathbf{F}_\Pi$, whose entries are $\sigma^{-2}(\vartheta)$. Following Raveri et al. (2016), we take $\sigma^{-2}(\{\Omega_b h^2, \Omega_{DM} h^2, h, \mathcal{A}_s, n_s, \tau\}) = \{100, 1, 6.25, 0.25, 0.1\}$. By using equation (A7) for both standard tomography and our hybrid approach on the same prior p_Π, we obtain an increment in information gain of 5.25 bits in favour of the hybrid method. According to Grandis et al. (2016, and references therein), this is similar to the increment on *WMAP*9 represented by a compilation of BAO data from the 6dFGS, BOSS in SDSS III, WiggleZ, and SDSS DR7. As a comparison, the increment brought by *Planck* data are 10 bits, whereas weak lensing from CFHTLenS gives 1.7 bits.

4.4 The case of primordial non-Gaussianity

Detection of non-zero values of the primordial non-Gaussianity parameter f_{NL} is one of the main goals of forthcoming galaxy surveys (e.g. Amendola et al. 2013, 2018; Alonso et al. 2015; Camera et al. 2015c). It is a difficult measurement, since the strongest signal arises from scale-dependent galaxy bias on ultra-large scales. This is in sharp contrast to the standard cosmological parameters, which rely on the highest possible number of available modes – thus implying a strong preference towards small scales. Our hybrid approach is geared towards the standard parameters and we do not expect it to perform optimally on f_{NL}.

To quantify this, we now include f_{NL} in the set of cosmological parameters, restricting the remaining parameters to those that are most degenerate with it, namely the amplitude of primordial fluctuations, $\mathcal{A}_s$, and the dark matter fraction, Ω_{DM}. As before, we include a free amplitude in each redshift bin as a nuisance parameter, then marginalize over all parameters but f_{NL}.

We find that the forecast marginal errors on the f_{NL} increase by a factor of $\sim$2 when moving from standard 2D tomography to our hybrid approach. On the other hand, the bias on the estimation of f_{NL} induced by neglecting the lensing term in equation (8) is smaller with our hybrid method, as summarized in Table 3.

The fact that both biases are $>1\sigma$ confirms the findings of Camera, Maartens & Santos (2015b) that it is necessary to include lensing – as well as other relativistic large-scale corrections – when trying to measure ultra-large scale effects such as primordial non-Gaussianity.

Table 3. Forecast relative bias on the non-Gaussianity parameter, f_{NL}, after marginalizing over nuisance bias parameters and Ω_{DM} and $\mathcal{A}_s$, for the two approaches.

$(\ell_{min}, \ell_{max}) = (2, 800)$	$b(f_{NL})/\sigma(f_{NL})$
Standard tomography (Section 3.2)	3.0
Hybrid tomography (Section 3.3)	1.2

4.5 Validity of the working assumptions

To assess how much the neglect of correlations among distant thin bins impacts our results, we focus first on the thick bins. For each thick bin, we compute 10 Fisher matrices, starting from the full angular power spectrum tomographic matrix used so far, $\mathbf{C}_\ell$, and then removing off-diagonal elements C_ℓ^{ij} as follows: the second Fisher matrix is computed from all C_ℓ^{ij}'s except for the two entries at the extrema of the anti-diagonal, which are set to zero; and so on until we compute the 10th Fisher matrix, where C_ℓ^{ij} is non-zero only for $i = j$.

The inset in Fig. 4 illustrates this procedure, with shading going from black (only diagonal elements considered) to the lightest of grey (all the cross-correlations between thin bins considered). The larger figure shows the structure of the full tomographic matrix of our hybrid method for this test of the importance of off-diagonal correlations. In this case, white means no correlation included.

Now, we consider two extrema: the first and the last of the thick bins. Referring to the inset of Fig. 4, for both of these thick bins, we start from the full angular power spectrum tomographic matrix and at each step we remove one off-diagonal, from the lightest grey until we remain only with the black, main diagonal. In Fig. 5, we illustrate the result of this test, by plotting the absolute bias on cosmological parameters induced by neglecting lensing magnification (normalized to the reference values of Table 2), as a function of how many cross-correlations between thin bins we include. Apart from the first point(s), for which the bias may be lower, but just because the Fisher matrices are very noisy, it appears clear that the behaviour is very flat. As expected, the trend is flatter at low redshift in the first thick bin.

Fig. 6 shows the bias on cosmological parameters due to neglecting lensing, as a function of how many off-diagonal spectra are considered in the full angular power spectrum, i.e. the larger matrix in Fig. 4. On the horizontal axis, 1 means black elements in the larger matrix, all the way to 10, representing the inclusion of all elements up to the lightest grey shade. Each point is normalized to its previous value, e.g. the seventh blue bullet point tells us that the bias on Ω_b will increase by ~20 per cent if we include up to the seventh off-diagonal, with respect to what we would guess by considering only up to the sixth. We notice a net trend: the bias on cosmological parameter best-fit values induced by neglecting lensing may be significant if we considered only the main diagonal or the first few off-diagonals, but then it does not change any longer. This means that in our hybrid method the information encoded in the lensing term $\propto (\mathcal{Q} - 1)\kappa$ flattens out more rapidly. This is mainly due to the thin slicing adopted for the sub-bins, in which lensing does not contribute substantially, as also noted by Villa et al. (2018).

Finally, we perform a last test. After having checked that cross-correlations at separations larger that our chosen $\Delta z = 0.1$ fall out rapidly (see Fig. 2) and that the bias on cosmological parameters due to neglecting lensing stabilizes well before all the distant–bin correlations have been included (this section), we now investigate what happens to those thin redshift bins at the edges of the thick bins – where the blocks along the diagonal meet, in Fig. 4. Those thin bins are nominally separated by less than Δz but are not included in the analysis because they pertain to different thick bins, which are considered as separate surveys. To address this issue, we proceed as follows:

(i) We recompute the Fisher matrix for the hybrid approach starting from the full tomographic matrix as in Fig. 4, rather than summing up all the Fisher matrices for each thick bin;

(ii) We then add cross-correlations at the edges of the thick bins, basically as if we were overlapping more blocks on the main diagonal.

The first computation (*i*), by definition, has to give the same results as have been presented so far. Nonetheless, this is an important test for our code to pass, given the possible numerical instabilities in the inversion of a noisy 140×140 matrix. Then, we illustrate the second computation (*ii*) in Fig. 7, where we show the observed angular power spectrum of equation (11) – that is to say, signal plus noise – at $\ell = 2$ for the first five thick bins, for the sake of clarity. As expected, some small non-zero correlation is present, although we emphasize that this is the very largest angular scale probed, and from Fig. 2 we know that these correlations not only disappear rapidly as the separation in redshift increases, but also at larger multipoles.

Once we have constructed this upgraded tomographic matrix for the observed signal, we insert it into equation (10) and compute a new Fisher matrix, which now includes additional noise due to the off-diagonal correlations between thin bins at the edges of the thick bins. We find that forecast marginal errors change by less than ~2 per cent for all cosmological parameters. Let us emphasize that, by not including nuisance parameters, in the present case the sensitivity to cosmological parameters is the highest. Therefore, in a more realistic case, the contribution due to those correlations will be even more suppressed. This last check further confirms the validity of our assumption of using the thick bins as separate surveys.

5 CONCLUSIONS

In this paper, we have presented a novel approach to optimize the use of angular power spectra with spectroscopic galaxy survey data. Our work is part of a community-wide effort aiming at crafting the best tools to exploit the oncoming wealth of cosmological data from large-scale structure experiments, such as the SKA, the European Space Agency's *Euclid* satellite, or the Large Synoptic Survey Telescope. We emphasize that the approach outlined here is meant to be complementary to others (see e.g. Asorey et al. 2012; Bailoni et al. 2017; Tansella et al. 2018).

Our work is motivated by two facts. On the one hand, future galaxy surveys will reach unprecedented depths and sky coverages, even with spectroscopy. On the other hand, the techniques usually employed to estimate the 3D galaxy power spectrum through spectroscopic measurements rely on a number of assumptions (see list in Section 3.1) that will hold no longer if we analyse such new catalogues as a whole. A more correct approach is that of 2D tomography, which, however, either does not fully exploit the exquisite radial resolution of spectroscopic redshifts, or becomes computationally untreatable.

Thus, the method we have introduced here for the first time draws inspiration from the standard 3D $P_g(\mathbf{k}, z)$ approach, but implements it within the consistent, gauge-independent formalism of the 2D tomographic matrix $\mathbf{C}_\ell$. For this reason, we refer to it as 'hybrid'. Note that our method intentionally neglects lensing, in order to
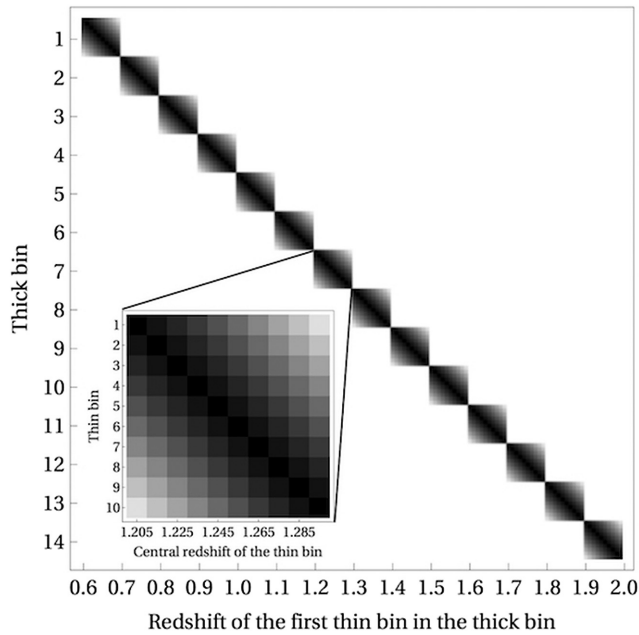
Figure 4. General structure of the full angular power spectrum tomographic matrix, $\mathbf{C}_\ell$, in our hybrid approach. Block diagonal matrices represent the thick bins, which we consider as independent. The inset illustrates the structure of one of those thick bins, with the thin bins shown. Shades of grey refer to the tests described in Section 4.5.

Figure 6. The same as in Fig. 5 but for the full angular power spectrum tomographic matrix of the hybrid method, namely the larger matrix in Fig. 4. Here, each point is normalized to the previous one.

resemble more closely the standard method in observations. Our main results can be summarized as follows:

(i) As presented in Section 4.1, forecasts show that hybrid constraints on cosmological parameters are comparable to those from a $P_g(\mathbf{k}, z)$ analysis, and more than twice as tight as those obtained via standard tomography, despite the larger number of nuisance parameters included. This trend is even more pronounced when limiting the analysis to strictly linear scales. We have also shown this by means of the increment of information gain, which roughly

corresponds to that achieved by BAO measurements over *WMAP*9 data.

(ii) Our hybrid approach is more robust than 2D tomography with respect to the neglect of corrections to galaxy number count fluctuations beyond density and RSD. In Section 4.2, we have shown that if we do not include lensing magnification (the most important remaining contribution), best fits on cosmological parameters stay within a few tens of per cent of a standard deviation from the true values, whereas with a standard tomographic approach they get biased by more than 1σ.
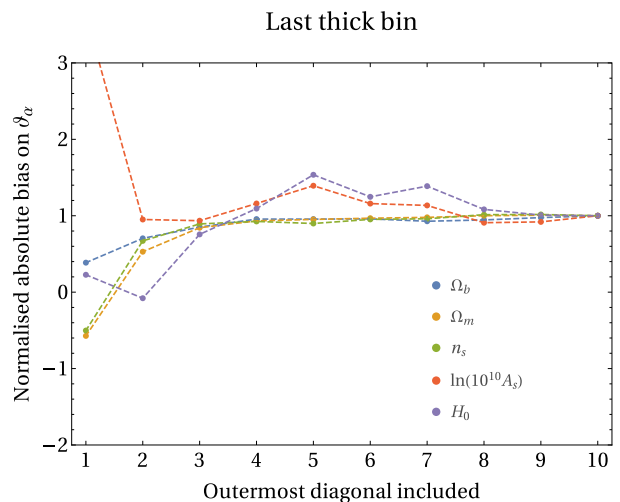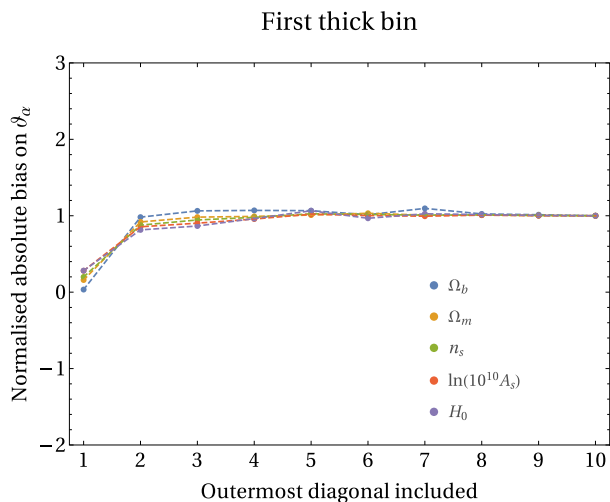

Figure 5. Absolute bias on cosmological parameters induced by neglecting lensing magnification, as function a of off-diagonal elements included, for an analysis carried out either in the first or in the last of the thick redshift bins only, normalized to what is obtained by including all cross-correlations between thin bins in the given thick bin.
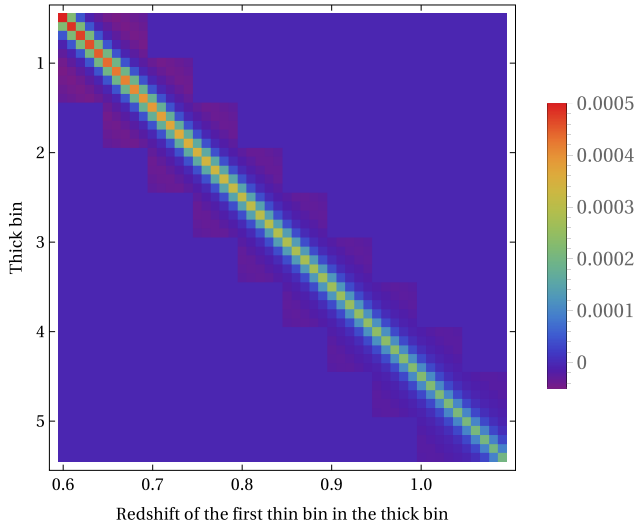
Figure 7. Observed tomographic angular power spectrum, $\widetilde{\mathbf{C}}_\ell$, at $\ell = 2$ for the first five thick redshift bins.

Summarizing, our method is able to match the $P_{\mathrm{g}}(\boldsymbol{k}, z)$ constraints on ΛCDM cosmological parameters, whilst easily incorporating cosmic evolution, wide-angle effects and lensing effects. It does not require AP corrections, since the angular power spectrum is an observable and does not need one to assume a fiducial cosmology. It improves significantly on broad-bin 2D tomography. And it is not biased by neglecting lensing effects, i.e. even when implementing only the two main contributions to galaxy number count fluctuations, namely density perturbations and RSD. This last fact leads to a further, major advantage compared to broad-bin tomography: it allows for faster code implementations, since the inclusion of integrated lensing terms significantly slows computation of angular power spectra. In a forthcoming work, we plan to estimate how the thickness of the redshift slicing affects the robustness of our method, namely what is the border between when it is safe to neglect lensing and when one has to include it.

ACKNOWLEDGEMENTS

REFERENCES

Abdalla F. B. et al., 2015, PoS, AASKA14, 017
Addison G. E., Huang Y., Watts D. J., Bennett C. L., Halpern M., Hinshaw G., Weiland J. L., 2016, ApJ, 818, 132
Ade P. et al., 2015, A&A, 594, A13
Albrecht A. J. et al., 2006, preprint (astro-ph/0609591)
Alonso D., Bull P., Ferreira P. G., Maartens R., Santos M. G., 2015 , ApJ, 814, 145
Amendola L. et al., 2013, Living Rev. Rel., 16, 6
Amendola L. et al., 2018, Living Rev. Rel., 21
An R., Feng C., Wang B., 2018, JCAP, 1802, 038
Asorey J., Crocce M., Gaztanaga E., Lewis A., 2012, MNRAS, 427, 1891
Bailoni A., Mancini A. S., Amendola L., 2017, MNRAS, 470, 688
Battye R. A., Charnock T., Moss A., 2015, Phys. Rev. D, 91, 103508
Blake C., Carter P., Koda J., 2018, MNRAS, 479, 5168
Bonvin C., Durrer R., 2011, Phys. Rev. D, 84, 063505
Camera S., Santos M. G., Bacon D. J., Jarvis M. J., McAlpine K., Norris R. P., Raccanelli A., Röttgering H., 2012, MNRAS, 427, 2079
Camera S., Santos M. G., Maartens R., 2015a, MNRAS, 448, 1035
Camera S., Maartens R., Santos M. G., 2015b, MNRAS, 451, L80
Camera S. et al., 2015c, PoS, AASKA14, 025
Camera S., Martinelli M., Bertacca D., 2017a, preprint (1704.06277)
Camera S., Harrison I., Bonaldi A., Brown M. L., 2017b, MNRAS, 464, 4747
Challinor A., Lewis A., 2011, Phys. Rev. D, 84, 043516
Charnock T., Battye R. A., Moss A., 2017, PRD, 95, 123535
Fonseca J., Camera S., Santos M., Maartens R., 2015, ApJ, 812, L22
Gil-Marín H. et al., 2016, MNRAS, 460, 4188
Grandis S., Seehars S., Refregier A., Amara A., Nicola A., 2016, JCAP, 1605, 034
Heavens A. F., Kitching T. D., Verde L., 2007, MNRAS, 380, 1029
Joudaki S. et al., 2017a, MNRAS, 471, 1259
Joudaki S. et al., 2017b, MNRAS, 465, 2033
Kullback S., Leibler R. A., 1951, Ann. Math. Stat., 22, 79
Laureijs R. et al., 2011, ESA-SRE, 12
Maartens R., Abdalla F. B., Jarvis M., Santos M. G., 2015, PoS, AASKA14, 016
Pourtsidou A., Tram T., 2016, Phys. Rev. D, 94, 043518
Pozzetti L. et al., 2016, A&A, 590, A3
Raccanelli A., Montanari F., Bertacca D., Dore O., Durrer R., 2016, JCAP, 1605, 009
Raveri M., 2016, Phys. Rev. D, 93, 043522
Raveri M., Martinelli M., Zhao G., Wang Y., 2016, preprint (1606.06273)
Ruggeri R., Percival W. J., Mueller E.-M., Gil-Marin H., Zhu F., Padmanabhan N., Zhao G.-B., 2017, preprint (1712.03997)
Sellentin E., Quartin M., Amendola L., 2014, MNRAS, 441, 1831
Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720
Spergel D. N., Flauger R., Hlozek R., 2015, Phys. Rev. D, 91, 023518
Tansella V., Bonvin C., Durrer R., Ghosh B., Sellentin E., 2018, JCAP, 1803, 019
Trotta R., 2008, Contemp. Phys., 49, 71
Villa E., Di Dio E., Lepori F., 2018, JCAP, 1804, 033

APPENDIX: FISHER MATRIX FORMALISM AND TOOLS

We start from the assumption that a given observable, for which we have a prediction based on some theoretical model, corresponds to a function of a set of parameters $\boldsymbol{\vartheta}$. In the Fisher information matrix formalism, the observed outcome is the mean value of the observable assumed as the null hypothesis. Thanks to this, we can estimate errors on the parameters, given errors in observable quantities.

In the frequentist approach, the Fisher matrix $\mathbf{F}$ is defined as the expectation value of the Hessian of the log-likelihood function $\mathcal{L} = -\ln L$, i.e.

$$\mathbf{F}\left(\vartheta_\alpha, \vartheta_\beta\right) = \left\langle \frac{\partial^2 \mathcal{L}}{\partial \vartheta_\alpha \partial \vartheta_\beta} \right\rangle, \tag{A1}$$

whilst in the Bayesian approach the data is no longer represented by random variables, and no averaging takes place. Hence, the Fisher

Table A1. Maximum discrepancy between the inverse of the Fisher matrix, $\mathbf{F}^{-1}$, and its Moore–Penrose pseudo-inverse, $\mathbf{F}_{\mathrm{MP}}^{-1}$, or its inverse via eigen-decomposition, $\mathbf{F}_{\mathrm{ED}}^{-1}$.

	$\max\left\{\left\|\mathbf{F}_{\mathrm{MP}}^{-1}/\mathbf{F}^{-1} - 1\right\|\right\}$	$\max\left\{\left\|\mathbf{F}_{\mathrm{ED}}^{-1}/\mathbf{F}^{-1} - 1\right\|\right\}$
2D without A_κ	1.3×10^{-10}	4.6×10^{-7}
2D with A_κ	8.9×10^{-10}	1.4×10^{-6}
hybrid without A_κ	5.0×10^{-10}	1.5×10^{-4}
hybrid with A_κ	4.3×10^{-11}	5.7×10^{-6}

matrix is simply evaluated at the parameter maximum-likelihood best fit, namely

$$\mathbf{F}\left(\vartheta_\alpha, \vartheta_\beta\right) = \left.\frac{\partial^2 \mathcal{L}}{\partial \vartheta_\alpha \partial \vartheta_\beta}\right|_{\boldsymbol{\vartheta} = \bar{\boldsymbol{\vartheta}}}. \tag{A2}$$

The two definitions coincide if the data are Gaussian and the parameters enter the mean and the variance in a linear way, or in the case of forecasting (Sellentin, Quartin & Amendola 2014).

The Cramer–Rao inequality states that a model parameter ϑ_α cannot have a variance smaller than $\delta(\vartheta_\alpha) = 1/\sqrt{\mathbf{F}(\vartheta_\alpha, \vartheta_\alpha)}$, when all other parameters are fixed (these are called 'conditional errors'), or be measured to a precision better than $\sigma(\vartheta_\alpha) = [\mathbf{F}^{-1}(\vartheta_\alpha, \vartheta_\alpha)]^{1/2}$, when all other parameters are marginalized over ('marginal errors'). Here, $\mathbf{F}^{-1}$ denotes the inverse of the Fisher matrix.

A1 Stability tests on Fisher matrices

To assess the validity of our results, we proceed as follows. We note that, particularly in the presence of nuisance parameters, the dimensionality of the Fisher matrices is high and their eigenvalues span a wide dynamic range. This may lead to numerical instabilities in the inversion of the Fisher matrix, and thus to spurious constraints on the cosmological parameters of interest. Therefore, to reduce the dynamic range, we first change the Fisher matrix basis via

$$\mathbf{F}\left(\vartheta_{\alpha'}, \vartheta_{\beta'}\right) = \mathbf{J}_{\alpha'\alpha}^T \mathbf{F}\left(\vartheta_\alpha, \vartheta_\beta\right) \mathbf{J}_{\beta'\beta}, \tag{A3}$$

where $\mathbf{J}_{\alpha'\alpha}$ is the Jacobian of the transformation from ϑ_α to ϑ'_α. As a new basis we choose $\vartheta'_\alpha = \left\{\ln \Omega_b, \ln \Omega_{\mathrm{DM}}, \ln n_s, \mathcal{A}_s, \ln h\right\}$, thus reducing the spread between the maximum and minimum eigenvalues by a factor of 4.3×10^4.

We marginalize over the bias nuisance parameters, 20 for the standard binning approach and 10×14 for our new hybrid approach. The marginalization involves a first inversion of the Fisher matrix, whose rows and columns referring to the bias nuisance parameters are then dropped. At this point, the inverse of the Fisher matrix (i.e. the parameter covariance matrix) only contains entries relative to the cosmological parameters. Then we can invert it and obtain the marginalized Fisher matrix in the new basis. Lastly, we perform the inverse change of variable of equation (A3), thus obtaining the final, marginalized Fisher matrix for the initial cosmological parameter set.

Despite the reduction of the dynamic range, matrix inversions may still propagate numerical instabilities into the resulting Fisher matrix. Therefore, we check the stability of our inverse matrices $\mathbf{F}^{-1}$ (computed via one-step row reduction) by comparing them to the Moore–Penrose pseudo-inverse and to the inverse obtained via eigen decomposition. This last approach helps in removing degeneracies from the Fisher matrix. Indeed, when marginalizing over the set of nuisance parameters, if one or more eigenvalues (nearly) vanish, then this degeneracy does not propagate into the cosmological parameters of interest (Albrecht et al. 2006; Camera et al.

2012). For a well-defined square matrix, the three matrix inversions must coincide. If we denote the Moore–Penrose pseudo-inverse of $\mathbf{F}$ by $\mathbf{F}_{\mathrm{MP}}^{-1}$ and the eigen-decomposed inverse by $\mathbf{F}_{\mathrm{ED}}^{-1}$, we can quantify the reliability of the marginal Fisher inversion by computing $\left|\mathbf{F}_X^{-1}/\mathbf{F}^{-1} - 1\right|$, with $X = \mathrm{MP}$, ED. Table A1 shows the maximum values of $\left|\mathbf{F}_X^{-1}/\mathbf{F}^{-1} - 1\right|$ for the marginal Fisher matrices employed in this work. It is easy to see that the discrepancy between the various methods of matrix inversion is at the very most $\mathcal{O}(0.01 \text{ per cent})$. We conclude that our marginal Fisher matrices are robust under inversion, and so are our cosmological parameter constraints.

A2 Estimating the bias on cosmological parameters

Given a set of cosmological, nuisance, and systematic-effect parameters, we can estimate the amount of biasing we will incur if we neglect some systematic effect within the Fisher information matrix framework. If

$$\mathcal{F} = \left[\begin{array}{c|c} \mathbf{F}\left(\vartheta_\alpha, \vartheta_\beta\right) & \mathbf{F}\left(\vartheta_\alpha, \varphi_b\right) \\ \hline \mathbf{F}\left(\varphi_a, \vartheta_\beta\right) & \mathbf{F}\left(\varphi_a, \varphi_b\right) \end{array}\right] \tag{A4}$$

is the full Fisher matrix, marginalized over nuisance parameters, that includes both cosmological parameters, $\boldsymbol{\vartheta}$, and systematic-effect parameters, $\boldsymbol{\varphi}$, we can compute the bias as (Heavens, Kitching & Verde 2007; see also Camera et al. 2017b, the Appendix)

$$b(\vartheta_\alpha) = \mathcal{F}^{-1}\left(\vartheta_\alpha, \vartheta_\beta\right) \mathcal{F}\left(\vartheta_\beta, \varphi_a\right) \delta\varphi_a, \tag{A5}$$

where $\mathcal{F}^{-1}(\vartheta_\alpha, \vartheta_\beta) \neq [\mathcal{F}(\vartheta_\alpha, \vartheta_\beta)]^{-1}$ means that one has first to invert $\mathcal{F}$ and then consider only the rows and columns of the *full* Fisher matrix relative to the cosmological parameters $\boldsymbol{\vartheta}$, whilst $\mathcal{F}(\vartheta_\beta, \varphi_a) \equiv \mathbf{F}(\vartheta_\beta, \varphi_a)$. Here, $\delta\varphi_a$ are the shifts from the true values of φ_a to the values incorrectly assumed in the analysis.

A3 Information gain

As a figure of merit to compare the new method we outlined in Section 3.3, we introduce here the concept of the Kullback–Leibler divergence, also known as 'relative entropy' or 'information gain' (Kullback & Leibler 1951). This is a tool for multivariate posterior distributions to compare the constraining power from different data sets or survey implementations. If X is a continuous, d-dimensional random variable with probability density functions $p_1(X)$ and $p_2(X)$, the information gain is

$$D(p_2||p_1) = \frac{1}{2\ln 2} \int \mathrm{d}^d X \, p_2(X) \ln \frac{p_2(X)}{p_1(X)} \text{ bits}. \tag{A6}$$

It represents the information gain (in bits) obtained by updating the distribution describing X from p_1 to p_2. Although it is not symmetric in p_2 and p_1, $D(p_2||p_1)$ is often interpreted as a distance between the two distributions. Indeed, it has remarkable properties, such as being non-negative, $D(p_2||p_1) \geq 0$, and zero if and only if $p_2 = p_1$.

In the case of multivariate normal (Gaussian) posterior distributions $\mathcal{N}(\boldsymbol{\vartheta})$ for the cosmological parameters, the information gain from a prior knowledge $\mathcal{N}_1$ to the posterior obtained from a new

experiment $\mathcal{N}_2$ is

$$D(\mathcal{N}_2||\mathcal{N}_1) = \frac{1}{2 \ln 2} \left\{ \ln \frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}_2} + \text{tr} \left[\boldsymbol{\Sigma}_1^{-1} \left(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1 \right) \right] \right\} \text{ bits} ,$$
(A7)

where $\boldsymbol{\Sigma}_i$ denotes the covariance matrix of the multivariate Gaussian $\mathcal{N}_i(\boldsymbol{\vartheta})$. Note that we consider the prior and posterior distributions as having the same mean. Even though this is not usually the case

with real data, it is a standard ansatz in forecasts. $\mathcal{N}_2$ can either be interpreted as the posterior of the data from a completely new experiment/analysis with respect to a previous experiment/analysis $\mathcal{N}_1$, or an updated posterior over a prior represented by $\mathcal{N}_1$.

This paper has been typeset from a TeX/LaTeX file prepared by the author.