Ultralarge-scale approximations and galaxy clustering: Debiasing constraints on cosmological parameters

Matteo Martinelli [ID],[1]⋆ Roohi Dalal [ID],[2] Fereshteh Majidi,[3,4] Yashar Akrami [ID],[5,6] Stefano Camera [ID][7,8,9,10] and Elena Sellentin[11]

[1]*Instituto de Física Teórica, Universidad Autónoma de Madrid, Campus de Cantoblanco, E-28049 Madrid, Spain*
[2]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[3]*Department of Physics and Chemistry, Alzahra University, Vanak Village Street, Tehran, 1993893973, Iran*
[4]*Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada*
[5]*Laboratoire de Physique de l'École Normale Supérieure, Universite PSL, CNRS, Sorbonne Université, F-75005 Paris, France*
[6]*Observatoire de Paris, Université PSL, Sorbonne Université, LERMA, F-75014 Paris, France*
[7]*Dipartimento di Fisica, Università degli Studi di Torino, via P. Giuria 1, I-10125 Torino, Italy*
[8]*INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Torino, via P. Giuria 1, I-10125 Torino, Italy*
[9]*INAF – Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, strada Osservatorio 20, I-10025 Pino Torinese, Italy*
[10]*Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa*
[11]*Leiden Observatory, Leiden University, Huygens Laboratory, Niels Bohrweg 2, NL-2333 CA, Leiden, the Netherlands*

ABSTRACT

Upcoming galaxy surveys will allow us to probe the growth of the cosmic large-scale structure with improved sensitivity compared to current missions, and will also map larger areas of the sky. This means that in addition to the increased precision in observations, future surveys will also access the ultralarge-scale regime, where commonly neglected effects such as lensing, redshift-space distortions, and relativistic corrections become important for calculating correlation functions of galaxy positions. At the same time, several approximations usually made in these calculations such as the Limber approximation break down at those scales. The need to abandon these approximations and simplifying assumptions at large scales creates severe issues for parameter estimation methods. On the one hand, exact calculations of theoretical angular power spectra become computationally expensive, and the need to perform them thousands of times to reconstruct posterior probability distributions for cosmological parameters makes the approach unfeasible. On the other hand, neglecting relativistic effects and relying on approximations may significantly bias the estimates of cosmological parameters. In this work, we quantify this bias and investigate how an incomplete modelling of various effects on ultralarge scales could lead to false detections of new physics beyond the standard ΛCDM model. Furthermore, we propose a simple debiasing method that allows us to recover true cosmologies without running the full parameter estimation pipeline with exact theoretical calculations. This method can therefore provide a fast way of obtaining accurate values of cosmological parameters and estimates of exact posterior probability distributions from ultralarge-scale observations.

Key words: methods: statistical – cosmology: observations – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

In recent years, the development of cosmic microwave background observations, led by surveys such as the Wilkinson Microwave Anisotropy Probe (WMAP; Hinshaw et al. 2013), *Planck* (Planck Collaboration I, VI 2020a,b), the South Pole Telescope (SPT; Carlstrom et al. 2011), and the Atacama Cosmology Telescope (ACT; Aiola et al. 2020), has brought cosmology into the precision era. The new frontier for cosmological observations is to now reach a similar precision in surveys of the cosmic large-scale structure. Observations of the large-scale structure can provide information on the matter distribution in the Universe and on the growth of primordial perturbations with time. This is achieved for example by observing the lensing effect of intervening matter on background

galaxies (cosmic shear) or by measuring the correlation function of the positions of galaxies (galaxy clustering). The former has been the main focus of the Kilo-Degree Survey (KiDS) collaboration, which has provided constraints on cosmological parameters both for the standard ΛCDM model and for some extensions (Köhlinger et al. 2017). The latter has been explored to exquisite precision by several observational collaborations, such as the two-degree Field Galaxy Redshift Survey (Cole et al. 2005), the six-degree Field Galaxy Survey (Beutler et al. 2011), WiggleZ (Blake et al. 2011; Parkinson et al. 2012), and the Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2005; Percival et al. 2010; Anderson et al. 2012; Alam et al. 2017). Experiments like the Dark Energy Survey (DES) have recently provided state-of-the-art measurements of cosmological parameters using both shear and clustering from photometric measurements (DES Collaboration 2021).

In the near future, observations of the large-scale structure will be further improved by new missions, either space-borne such as

⋆ E-mail: matteo.martinelli@inaf.it

Euclid (Laureijs et al. 2011; Amendola et al. 2013, 2018; Euclid Collaboration 2020), the *Roman* Space Telescope (Spergel et al. 2015), and the Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) (Doré et al. 2014, 2018), or ground-based such as the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016a,b), the Rubin Observatory Legacy Survey of Space and Time (LSST; LSST Science Collaboration 2009; LSST Dark Energy Science Collaboration 2018; Ivezić et al. 2019), and the SKA Observatory (SKAO; Abdalla et al. 2015; Brown et al. 2015; Bull et al. 2015; Camera et al. 2015a; Raccanelli et al. 2015; Santos et al. 2015; SKA Cosmology Science Working Group 2020). These future surveys will indeed improve the sensitivity of the measurements, and, in addition, will make it possible to perform observations on large volumes of the sky. With such observations, it will be possible to access, for the first time, ultralarge scales when measuring the correlation function of galaxy positions and shear. While this ability to access such large scales will allow us to better constrain cosmological models and test fundamental theories such as general relativity (Baker & Bull 2015; CANTATA Collaboration 2021), it will also pose new challenges to our ability to theoretically model the observables involved.

In particular, the galaxy correlation function at very large scales receives contributions from lensing, redshift-space distortions (RSD), and relativistic effects (Yoo 2010; Bonvin & Durrer 2011; Challinor & Lewis 2011; Bertacca, Maartens & Clarkson 2014), which are mostly negligible for the scales probed by current surveys (see e.g. Alonso et al. 2015; Fonseca et al. 2015; Yoo & Seljak 2015). The modelling problem presented by such contributions is not as severe as the one of modelling non-linear effects at small scales, where one needs to rely on model-dependent numerical simulations (see e.g. Bose et al. 2021; Chartier et al. 2021; Chartier & Wandelt 2021; Martinelli et al. 2021; Safi & Farhang 2021). However, in order to simplify the modelling of large-scale effects, several approximations are commonly made in computing theoretical predictions for galaxy number counts, such as the Limber (LoVerde & Afshordi 2008) and the flat-sky (Matthewson & Durrer 2021) approximations. Such simplifications hold for the scales probed by current surveys (Kilbinger et al. 2017), but they may fail when larger scales will be accessed by future surveys.

Calculations that include large-scale effects and do not rely on approximations are feasible, and codes commonly used to compute theoretical predictions, such as CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) and CLASS (Blas, Lesgourgues & Tram 2011), allow us to obtain 'exact' galaxy clustering power spectra. However, the computational time required for such exact calculations is significantly longer, causing parameter estimation pipelines to become unfeasible, as they require calculating tens of thousands of spectra to reconstruct posterior probability distributions for cosmological parameters.

Several attempts have been made to overcome this problem. For instance, fast Fourier transform (FFT) or logarithmic FFT (FFTLog) methods can be exploited to accelerate the computation of the theoretical predictions (Assassi, Simonović & Zaldarriaga 2017; Campagne, Neveu & Plaszczynski 2017; Grasshorn Gebhardt & Jeong 2018). Alternatively, approximations can be made to reduce the dimensionality of the integration, namely either assuming that the observed patch of sky is flat, and thus performing a two-dimensional Fourier transform on the sky (Datta, Choudhury & Bharadwaj 2007; White & Padmanabhan 2017; Jalilvand et al. 2020; Matthewson & Durrer 2021), or exploiting the behaviour of spherical Bessel functions at large angular multipoles (Limber 1953, 1954; Kaiser 1992).

In this work, we investigate how applying these commonly used approximations and neglecting lensing, RSD and relativistic

contributions at large scales can bias the estimation of cosmological parameters, and possibly lead to false detections of non-standard cosmological models. Such an analysis has been of interest for some time (see e.g. Camera et al. 2015b; Camera, Maartens & Santos 2015d; Villa, Di Dio & Lepori 2018; Thiele, Duncan & Alonso 2020), but we investigate it here considering all the large-scale effects and approximations at the same time, while relying on a full Markov chain Monte Carlo (MCMC) pipeline for parameter estimation, rather than using Fisher matrices. Note that other studies (e.g. Cardona et al. 2016; Tanidis & Camera 2019; Tanidis, Camera & Parkinson 2020) did approach the problem from the MCMC point of view, but they all, in one way or another, had to simplify the problem in a way that either made them differ from a benchmark analysis, or assumed some of the aforementioned approximations.

Additionally, we propose a simple debiasing method to recover the true values of cosmological parameters without the need for exact calculations of the power spectra. Such a method will allow us to analyse future data sets in a manner that avoids computational problems, but ensures that we accurately obtain the correct best-fitting values of cosmological parameters and estimates of their posterior distributions.

The paper is structured as follows. We review in Section 2 the theoretical modelling of galaxy number count correlations, presenting both the exact computation and the approximated one. In Section 3, the experimental setup used throughout the paper is presented, while in Section 4 we describe the cosmological models considered in this paper and their impacts on galaxy number counts. In Section 5, we present our analysis pipeline and introduce a debiasing method able to significantly reduce the bias on cosmological parameters introduced by incorrect modelling of the observables. We present our results in Section 6 and draw our conclusions in Section 7.

2 GALAXY NUMBER COUNTS AND HARMONIC-SPACE CORRELATION FUNCTIONS

Observed fluctuations in galaxy number counts are primarily caused by underlying inhomogeneities in the matter density field on cosmological scales and, for galaxies, are a biased tracer of the cosmic large-scale structure. However, there is a score of secondary effects that also contribute to the observed signal (Yoo 2010; Bonvin & Durrer 2011; Challinor & Lewis 2011). The most important of them are the well-known redshift-space distortions, which represent the dominant term on sub-Hubble scales, and weak-lensing magnification, important for deep surveys and wide redshift bins. Additionally, there is a more complicated set of relativistic terms that arise from radial and transverse perturbations along the photon path from the source to the observer.

Thus, we can write the observed galaxy number count fluctuation field in real space and up to first order in cosmological perturbation theory as (see e.g. Ghosh, Durrer & Sellentin 2018)[1]

$$\Delta_\mathrm{g} = b_\mathrm{lin}\,\delta - \frac{1}{\mathcal{H}}\partial_\parallel^2 V - b_\mathrm{mag}\,\kappa + \Delta_\mathrm{loc} + \Delta_\mathrm{int}. \tag{1}$$

---

[1]Note that several different symbols are used in the literature to denote the magnification bias and – as we shall see later on – the evolution bias, e.g. α, $\mathcal{Q}$, and s for the former, and b_e and f_evo for the latter (see also Maartens et al. 2021). Here, however, we adopt a more uniform notation, with b_lin, b_mag, and b_evo, respectively, denoting the linear galaxy bias, the magnification bias, and the evolution bias. For the first two, the rationale behind our notation is that they respectively are what modulates the matter density fluctuations and lensing convergence.

(Note that hereafter we shall use units such that $c = 1$.) To understand better what the expression above means, we shall now break it up in all its terms:

(i) The first term in equation (1) sees the linear galaxy bias, b_{lin}, multiplying matter density fluctuations in the comoving-synchronous gauge, δ.

(ii) The second term is linear RSD, with $\partial_\parallel$ the spatial derivative along the line-of-sight direction, $\hat{r}$, and V the peculiar velocity potential.

(iii) The third term is the lensing magnification contribution, sourced by the integrated matter density along the line of sight, i.e. the weak-lensing convergence κ, modulated by the so-called magnification bias, b_{mag}, which respectively take the forms

$$\kappa(\mathbf{r}) = \int_0^r \mathrm{d}x \ (r-x) \frac{x}{r} \nabla_\perp^2 \Upsilon(\hat{r}, r = x), \tag{2}$$

$$b_{mag}(z) = 2 \left[1 - \left. \frac{\partial \ln \bar{n}_g(z; F > F_{cut})}{\partial \ln F} \right|_{F_{cut}} \right], \tag{3}$$

with $r(z)$ the radial comoving distance to redshift z such that $\mathrm{d}r = \mathrm{d}z/H(z)$ and $H(z) = (1+z)\mathcal{H}(z)$, $\nabla_\perp^2$ the Laplacian on the transverse screen space, $\Upsilon = (\Phi + \Psi)/2$ the Weyl potential, where Φ and Ψ are the two Bardeen potentials of the perturbed metric, and $\bar{n}_g$ the mean redshift-space comoving number density of galaxies, which is a function of redshift and flux F (equivalently luminosity, or magnitude). Here, F_{cut} represents the flux value that a galaxy should have in order to be detected by the adopted instrument.

(iv) The penultimate term in equation (1) gathers all the local contributions at the source, such as Sachs–Wolfe and Doppler terms, and reads

$$\Delta_{loc} = (3 - b_{evo})\mathcal{H} V + A \partial_\parallel V - b_{mag}\Phi + (1-A)\Psi + \frac{\Phi'}{\mathcal{H}}, \tag{4}$$

with

$$b_{evo}(z) = -\frac{\partial \ln \bar{n}_g(z)}{\partial \ln(1+z)}, \tag{5}$$

usually referred to as the evolution bias,[1]

$$A \equiv b_{evo} + b_{mag} - 2 - \frac{\mathcal{H}'}{\mathcal{H}^2} - \frac{b_{mag}}{\mathcal{H}r}, \tag{6}$$

and a prime denoting derivation with respect to conformal time.

(v) The last term, on the other hand, collects all non-local contributions such as time-delay and integrated Sachs–Wolfe type terms and reads

$$\Delta_{int} = 2\frac{b_{mag}}{r} \int_0^r \mathrm{d}x \ \Upsilon - 2A \int_0^r \mathrm{d}x \ \Upsilon'. \tag{7}$$

2.1 The exact expression

The exact linear harmonic-space angular power spectrum of the observed galaxy number count fluctuations between two (infinitesimally thin) redshift slices at z and z', $C_\ell^{Ex}(z, z')$, is then obtained by expanding equation (1) in spherical harmonics, and taking the ensemble average

$$\left\langle \Delta_{g,\ell m}(z)\Delta_{g,\ell' m'}^*(z') \right\rangle \equiv \delta_{\ell\ell'}^K \delta_{mm'}^K C_\ell^{Ex}(z, z'), \tag{8}$$

with δ^K the Kronecker delta symbol. This leads to the expression ('Ex' meaning 'exact')

$$C_\ell^{Ex}(z, z') = 4\pi \int \mathrm{d}\ln k \ \mathcal{W}_\ell^g(k; z) \mathcal{W}_\ell^g(k; z') \mathcal{P}_\zeta(k), \tag{9}$$

with $\mathcal{W}_\ell^g$ the kernel of galaxy clustering, encompassing contributions from all terms present in equation (1), and $\mathcal{P}_\zeta(k) = A_s k^{n_s - 1}$ the power spectrum of primordial curvature perturbations, A_s and n_s, respectively, being its amplitude and spectral index.

For a full expression for $\mathcal{W}_\ell^g$, we can write

$$\mathcal{W}_\ell^g = \mathcal{W}_\ell^{g,den} + \mathcal{W}_\ell^{g,vel} + \mathcal{W}_\ell^{g,len} + \mathcal{W}_\ell^{g,rel}, \tag{10}$$

with $\mathcal{W}_\ell^{g,vel} = \mathcal{W}_\ell^{g,RSD} + \mathcal{W}_\ell^{g,Dop}$ the term related to galaxies' velocities, where (see e.g. Di Dio et al. 2013)

$$\mathcal{W}_\ell^{g,den}(k; z) = b_{lin}(k, z) \, T_\delta(k, z) \, j_\ell \left[kr(z) \right], \tag{11}$$

$$\mathcal{W}_\ell^{g,RSD}(k; z) = \frac{k}{\mathcal{H}(z)} T_V(k, z) \, j_\ell'' \left[kr(z) \right], \tag{12}$$

$$\mathcal{W}_\ell^{g,Dop}(k; z) = \left\{ [b_{evo}(z) - 3] \frac{\mathcal{H}(z)}{k} j_\ell \left[kr(z) \right] \right.$$
$$\left. - A(z) j_\ell' \left[kr(z) \right] \right\} T_V(k, z), \tag{13}$$

$$\mathcal{W}_\ell^{g,len}(k; z) = \ell(\ell+1) b_{mag}(z)$$
$$\times \int_0^{r(z)} \mathrm{d}x \, \frac{r(z) - x}{r(z)x} T_\Upsilon(k, r = x) \, j_\ell(kx), \tag{14}$$

$$\mathcal{W}_\ell^{g,rel}(k; z) = \left\{ [1 - A(z)] T_\Psi(k, z) - 2 b_{mag}(z) T_\Phi(k, z) \right.$$
$$\left. + \frac{1}{\mathcal{H}(z)} T_{\Phi'}(k, z) \right\} j_\ell \left[kr(z) \right]$$
$$+ 2 \frac{b_{mag}(z)}{r(z)} \int_0^{r(z)} \mathrm{d}x \, T_\Upsilon(k, r = x) j_\ell(kx)$$
$$- 2 A(z) \int_0^{r(z)} \mathrm{d}x \, T_{\Upsilon'}(k, r = x) j_\ell(kx). \tag{15}$$

In the equations above, T_X denotes the transfer function describing the evolution of the random variable X and $T_X(k, z) \equiv T_X[k, r(z)]$. Note that, with a slight abuse of notation, a prime applied to a spherical Bessel function denotes a derivative with respect to its argument.

In harmonic-space analyses, it is customary to subdivide the observed source population into redshift bins. This is done, for instance, to reduce the dimensionality of the data vector – and consequently the covariance matrix – with the aim of reducing, in turn, the computational complexity of the problem. Otherwise, redshift information for the observed galaxies might be too poor to allow us to pin them down in the radial direction, as is the case with photometric redshift estimation. In this case, galaxies are usually binned into $\mathcal{O}(1) - \mathcal{O}(10)$ bins spanning the observed redshift range. Whatever the reason, in practice this corresponds to having

$$C_{ij\ell}^{Ex} = 4\pi \int \mathrm{d}\ln k \ \mathcal{W}_{i\ell}^g(k) \mathcal{W}_{j\ell}^g(k) \mathcal{P}_\zeta(k), \tag{16}$$

where

$$\mathcal{W}_{i\ell}^g(k) = \int \mathrm{d}z \ \mathcal{W}_\ell^g(k; z) \, n_i(z), \tag{17}$$

with $n_i(z)$ the galaxy redshift distribution in the ith redshift bin, normalized to unit area.

2.2 Widely used approximations

The computation of harmonic-space power spectra has to be performed following the triple integral of equation (16) and the equations giving the kernel $\mathcal{W}_{i\ell}^g$. However, such an integration is numerically cumbersome, especially because of the presence of spherical Bessel

functions – highly oscillatory functions whose amplitude and period vary significantly with the argument of the function. As a consequence, numerical integration has to be performed with highly adaptive methods at the cost of computation speed. Over the years, various algorithms have been proposed with the aim of speeding up the computation of harmonic-space power spectra. Mostly, they rely on FFT/FFTLog methods (see e.g. Assassi et al. 2017; Campagne et al. 2017; Grasshorn Gebhardt & Jeong 2018).

On the other hand, the full computation is not always necessary, and approximations can be made to speed up the numerical evaluation, e.g. by applying the Limber or the flat-sky approximations (often erroneously thought to be the same, see e.g. Matthewson & Durrer 2021). Here, we shall focus on the former, which is by far the most widely employed. It relies on the following property of spherical Bessel functions,

$$j_\ell(x) \xrightarrow[\ell \gg 1]{} \sqrt{\frac{\pi}{2\ell + 1}} \delta_D \left(\ell + \frac{1}{2} - x \right), \qquad (18)$$

where δ_D is a Dirac delta.[2] By performing the substitution of equation (18) into equation (16), which contains j_ℓ through the $\mathcal{W}_{i\ell}^g(k)$, we can effectively get rid of two integrations, thus boosting significantly the speed of the computation.

Moreover, the relative importance of the different terms in equation (10) depends on various survey-dependent factors. For instance, RSD are mostly washed out for broad redshift bins, whereas, on the contrary, lensing magnification favours them. Similarly, the Doppler contribution decays quickly as the redshift of the shell grows, whilst integrated terms like lensing gain in weight. Lastly, the importance of the various effects also varies with the scales of interest, as can be seen by the $\mathcal{H}/k$ factors in equation (11) to equation (15). Moreover, note that at first order in cosmological perturbation theory, the Einstein equations fix $V \sim \delta/k$ and $\Phi \sim \Psi \sim \delta/k^2$. All combined, this makes $\mathcal{W}_\ell^{g,\mathrm{rel}}$ important only on very large scales.

For these reasons, galaxy clustering in harmonic space is customarily restricted to Newtonian density fluctuations alone, leading to the well-known expression for the approximated angular spectra ('Ap' standing for 'approximated')

$$C_{ij\ell}^{\mathrm{Ap}} = \int \mathrm{d}z \, \frac{[H(z) \, b_{\mathrm{lin}}(z)]^2 \, n^i(z) \, n^j(z)}{r^2(z)} \, P_{\mathrm{lin}} \left[\frac{\ell + 1/2}{r(z)}, z \right], \qquad (19)$$

where $P_{\mathrm{lin}}(k, z)$ is the linear matter power spectrum, and for now we have assumed that linear galaxy bias is only redshift-dependent. Let us emphasize that this approximation, and in particular the neglection of RSD, is oft-times common in harmonic-space analyses of galaxy clustering (see e.g. Granett et al. 2012; van Uitert et al. 2018; DES Collaboration et al. 2021), albeit with noticeable exceptions (Padmanabhan et al. 2007; Loureiro et al. 2019; Joachimi et al. 2021; Tanidis & Camera 2021). Oppositely, real- and Fourier-space analyses do customarily account for RSD.

The actual accuracy of such an approximation, however, cannot be estimated a priori, since it strongly depends on the integrand of equation (19). In particular, equation (19) is known to agree well with the exact expression of equation (16) if the kernel of the integral is broad in redshift. Moreover, the Limber approximation works better at low redshift than at high redshift, because the higher the redshift, the larger the scale subtended by a given angular separation; in other
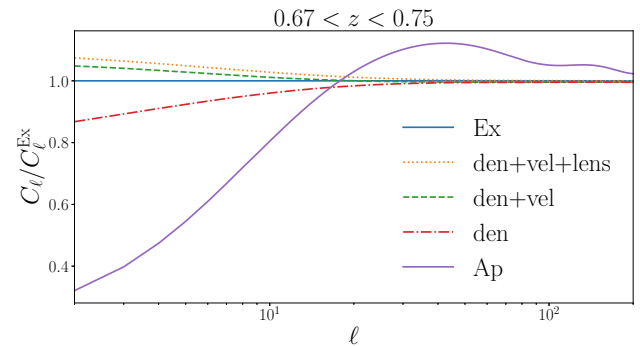
²Note that the $+1/2$ term comes from the relation between a spherical Bessel function of order ℓ, j_ℓ, and the ordinary Bessel function of order $L = \ell + 1/2$, J_L.

Figure 1. Ratio of the approximated C_ℓ to the exact C_ℓ^{Ex} of equation (16). The labels of the different curves correspond to the contributions that enter the window function in equation (10). None of the spectra considered here use the Limber approximation, except the 'Ap' spectrum, which corresponds to the fully approximated C_ℓ^{Ap} of equation (19). The spectra shown here refer to the autocorrelation in a redshift bin with $0.67 < z < 0.75$, using the survey specifications discussed in Section 3.

words, the minimum multipole for which the Limber approximation agrees well with the exact solution increases with redshift.

In Fig. 1, we highlight the contributions of the different terms to the final spectra by showing the ratio of approximated spectra to the exact ones. We consider here the autocorrelation spectra in a redshift bin with $0.67 < z < 0.75$, using the survey specifications we later discuss in Section 3. None of the spectra shown in the figure use the Limber approximation, except the 'Ap' spectrum, which corresponds to equation (19). We notice how removing different terms makes the theoretical prediction move away from the exact one, although only at very large scales and not in a dramatic way, even when only the density term of equation (11) is kept. However, once the Limber approximation is used, the predictions significantly depart from the exact spectrum over a wide range of multipoles.

3 SURVEY SPECIFICATIONS

In the coming decade, several planned surveys of the cosmic large-scale structure will provide us with observations of the galaxy distribution with unprecedented sensitivity at very large scales. It is therefore crucial to assess how the common approximations described in Section 2 will impact the accuracy of the results we will be able to obtain. Therefore, in this paper we adopt the specifications of a very deep and wide galaxy clustering survey with high redshift accuracy. We emphasize that we are not interested in forecasts for a specific experiment, but rather in assessing whether and how much various approximations affect the final science output. For this reason, we shall focus on an idealized survey, loosely inspired by the envisaged future construction phase of the SKAO. Specifically, we consider an HI-galaxy redshift survey, assuming that the instrument will be able to provide us with spectroscopic measurements of the galaxies' redshifts through the detection of the HI emission line in the galaxy spectra. Therefore, for the purposes of the harmonic-space tomographic studies we focus on in this paper, we shall consider the error on such redshift measurements to be negligible.

Here, we follow the prescription and fitting functions of Yahya et al. (2015) to characterize the source galaxy distribution as a function of both redshift and flux limit. The latter will be particularly important in determining the magnification bias of the sample. Calculations in Yahya et al. (2015) were based on the S^3-SAX simulations by Obreschkow & Rawlings (2009) and assumed that

Table 1. Survey specifications and fiducial cosmology used in the present work to obtain the mock data set and experimental noise.

Survey specifications								
N_{gal}	f_{sky}	z_{min}	z_{max}	c_1	c_2	c_3	c_4	c_5
9.4×10^8	0.7	0.001	1.1	6.32	1.74	5.42	0.55	0.78

Fiducial cosmology							
ω_b	ω_c	h	$A_s \times 10^9$	n_s	$\sum m_\nu$ [eV]	w	f_{NL}
0.022 45	0.120 56	0.67	2.126	0.96	0.06	-1	0
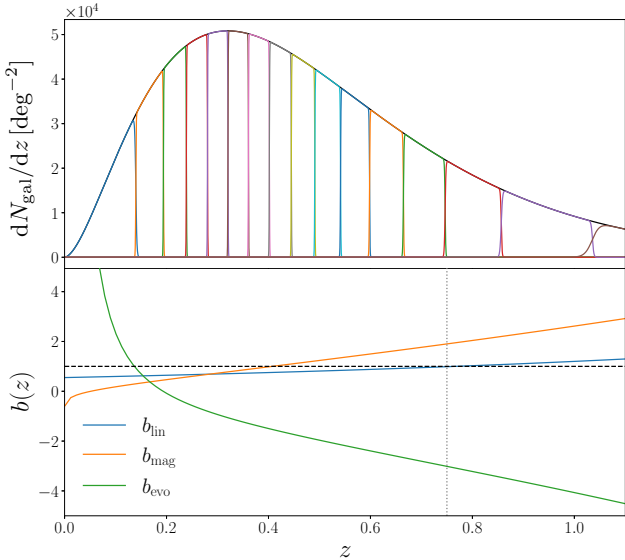
Figure 2. *Upper panel:* Galaxy distribution as described by equation (20; in black) with the limits of the equipopulated redshift bins considered in the present paper (in colour). *Lower panel:* trends in redshift for the linear galaxy bias of equation (21; blue curve), the magnification bias of equation (3; orange curve), and the evolution bias of equation (5; green curve). The intersection between the horizontal dashed and vertical dotted black lines shows where the linear galaxy bias crosses unity.

any galaxy with an integrated line flux above a given signal-to-noise ratio threshold would be detected. The fitting formulae adopted here are

$$\frac{dN_{gal}}{dz} = 10^{c_1} z^{c_2} \exp(-c_3 z) \, \text{deg}^{-2}, \quad (20)$$

$$b_{lin}(z) = c_4 \exp(c_5 z), \quad (21)$$

where N_{gal} is the total number of galaxies in the entire redshift range of the survey, and parameters c_i can be found in Yahya et al. (2015) for a wide range of flux thresholds, from 0 to 200 μJy. We show in Table 1 values of c_i used in the present work, corresponding to those used in Sprenger et al. (2019) and obtained in Bull (2016) as a result of fitting these functions to the expected galaxy number density, given the survey design.

Given the galaxy distribution of equation (20), we focus on the redshift range $0.001 < z < 1.1$ with N_{gal} given in Table 1, and divide it into $N_{bin} = 15$ redshift bins assuming that each one contains the same number of galaxies (see the upper panel of Fig. 2). In the lower panel of Fig. 2, we show the redshift evolution of the linear galaxy bias, the magnification bias and the evolution bias given, respectively, by equations (21), (3), and (5), for the survey under consideration.

Using these survey specifications, we create a simulated data set for galaxy clustering observations; we calculate the exact angular power

spectra C_ℓ^{Ex}, described in Section 2, in a fiducial cosmology and we add to these the noise computed using the survey specifications. For the rest of this paper we use the calculations of the exact and approximated power spectra as implemented in CAMB[3] (Lewis et al. 2000; Howlett et al. 2012). We assume a ΛCDM cosmology with fiducial values of parameters given in Table 1, where ω_b and ω_c are the baryon and cold dark matter physical energy densities, respectively, h is the reduced present-day Hubble expansion rate, A_s and n_s are, respectively, the amplitude and spectral index of the primordial curvature power spectrum, and $\sum m_\nu$ is the sum of the neutrino masses.

4 CASE STUDIES

We study four representative cosmological models in order to demonstrate how the approximations of Section 2.2 can bias the estimation of cosmological parameters using a next-generation survey able to access ultralarge scales, as described in Section 3, and how the method we present in this paper debiases the constraints while keeping the computational cost of the parameter estimation procedure significantly lower than that of an exact analysis. These four models are the standard ΛCDM model and three of its minimal extensions, where either the dark energy equation of state w or the sum of the neutrino masses $\sum m_\nu$ or the local primordial non-Gaussianity (PNG) parameter f_{NL} is allowed to vary as an additional free parameter. We denote these extensions by wCDM, ΛCDM+m_ν, and ΛCDM+f_{NL}, respectively.

4.1 Standard model and its simple extensions

We specify the standard ΛCDM model by the five free parameters ω_b, ω_c, h, A_s, and n_s.[4] Following Planck Collaboration VI (2020b), we fix the value of $\sum m_\nu$ to 0.06 eV for ΛCDM. The parameters $\{\omega_b, \omega_c, h, A_s, n_s\}$ affect the angular power spectra of the observed galaxy number count fluctuations differently and on different angular scales. Here we are interested in ultralarge scales, which are expected to be particularly sensitive to the parameters that quantify cosmic initial conditions, i.e. A_s and n_s.

In order to illustrate the large-scale effects of the parameters, we show, as an example, in the upper left hand panel of Fig. 3 the impact of varying the scalar spectral index n_s on the power spectrum at angular scales larger than $\ell = 400$ computed at redshift bins 5, 10, and 15 (corresponding to redshift ranges $0.28 < z < 0.32$, $0.49 < z < 0.54$, and $0.86 < z < 1.04$, respectively) as given in the upper panel of Fig. 2. For each redshift bin, the corresponding galaxy number count

---

[3]https://github.com/cmbant/CAMB

[4]Note that ΛCDM also requires the reionization optical depth τ as a free parameter. However, we do not vary τ in our analysis as we do not expect the large-scale observables to constrain this quantity.
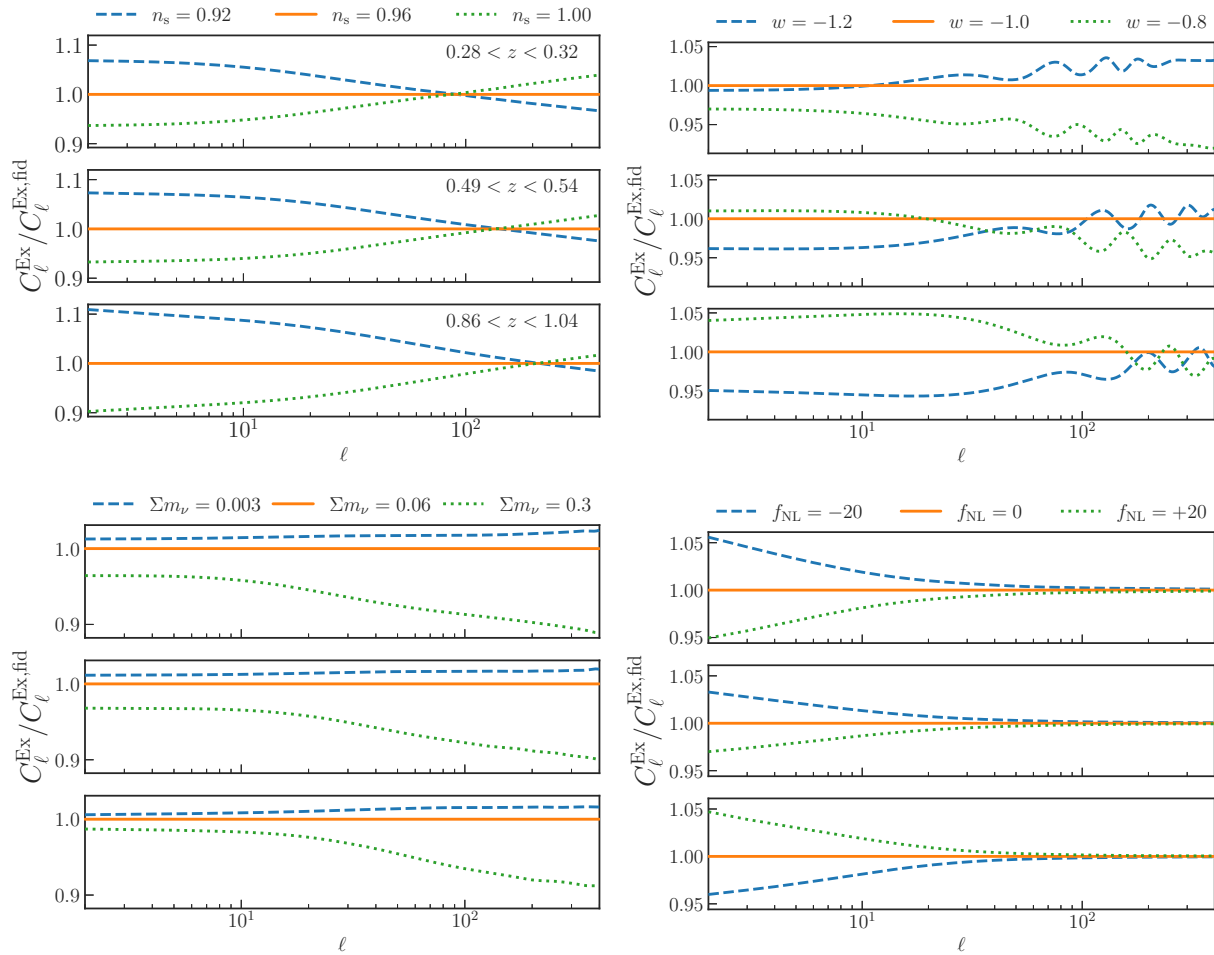
Figure 3. Effects of cosmological parameters on the angular power spectrum of observed galaxy number count fluctuations, C_ℓ, on large scales. The four panels depict the effects of: the primordial scalar spectral index n_s in ΛCDM (upper left panel); the dark energy equation of state w in wCDM (upper right panel); the sum of the neutrino masses $\sum m_\nu$ in ΛCDM+m_ν, with values in eV (lower left panel); and the local primordial non-Gaussianity f_{NL} in ΛCDM+f_{NL} (lower right panel). All the power spectra are exact, i.e. no approximations are made in their computations, and they are shown in comparison with the fiducial ΛCDM spectra with $n_s = 0.96$, $w = -1$, $\sum m_\nu = 0.06$ eV, and $f_{NL} = 0$. Each panel contains three sets of spectra computed for the three redshift bins 5, 10, and 15, corresponding to low, medium, and high redshifts (from top to bottom in each panel). The redshift range of each bin is indicated in the respective plot in the upper left panel.

power spectra for three values of $n_s = 0.92$, $n_s = 0.96$, and $n_s = 1$ are shown relative to the spectrum for $n_s = 0.96$, which we use as our fiducial value in the rest of this paper. Note that these spectra are all exact, i.e. they are computed without making any approximations. As the figure shows, in all the redshift bins, the lower the value of n_s, the more enhanced the power spectra at ultralarge scales, namely scales with $\ell \lesssim \mathcal{O}(100)$, while we see the opposite effect at smaller scales. This is because the smaller the value of n_s, the steeper (or more 'red-tilted') the primordial power spectrum, resulting in larger amplitudes of fluctuations at extremely large scales. This steeper spectrum will then lead to suppression of amplitudes at scales smaller than some 'pivot' scale. Note, however, that the enhancement or suppression on large scales is not physical, as it depends on the scale used as a pivot – namely, fixing either A_s or σ_8 (amplitude of the linear power spectrum on the scale of $8\,h^{-1}$ Mpc) as a fundamental parameter.

The figure for n_s already shows the importance of correctly computing the angular power spectra for accurately estimating the cosmological parameters using ultralarge-scale information. As the figure shows, even changing n_s to the extreme values of 0.92 and 1,

both of which having already been ruled out by the current constraint $n_s \approx 0.965 \pm 0.0042$ (Planck Collaboration VI 2020b), changes the power spectra by $\mathcal{O}(10$ per cent). On the other hand, as we will see in Section 6, the approximations of Section 2.2 may easily result in $> \mathcal{O}(10$ per cent) errors in the computation of the spectra on large scales, which will then lead to inaccurate, or biased, estimates of parameters like n_s.

An inaccurate estimation of a cosmological parameter can also result in a false detection of new physics when there is none, or in no detection when there is. In order to demonstrate this problem, we present in the upper right and lower left panels of Fig. 3 the effects of the two important non-standard cosmological parameters, w and $\sum m_\nu$, on the power spectrum at large scales for wCDM and ΛCDM+m_ν, the two simple extensions of ΛCDM that we introduced earlier. The panels again depict the spectra for the three redshift bins 5, 10, and 15, with the additional parameters w and $\sum m_\nu$ of the two extensions set to $\{-1.2, -1, -0.8\}$ and $\{0.003, 0.06, 0.3\}$, respectively. Note that throughout this paper, we always use $w = -1$ and $\sum m_\nu = 0.06$ as the fiducial values for these parameters.

We notice that changing the value of w has a few large-scale effects. First of all, setting w to a value smaller or larger than -1 does not affect the spectra similarly in different redshift bins. Focusing first on the $w = -1.2$ case, which corresponds to a phantom dark energy, we see that the spectra are all suppressed at ultralarge scales compared to the standard $w = -1$ case, and by increasing the bin's redshift, not only does the range of the suppressed power extend to smaller scales, but also the higher the redshift, the more suppressed the spectra (on all scales). The effect is the opposite for the $w = -0.8$ case, and increasing the bin's redshift results in more enhanced spectra compared to the baseline $w = -1$. The $w \neq -1$ enhancement or suppression of power and its redshift dependence can be explained for smaller scales by the fact that the linear growth rate of the large-scale structure, f, is significantly affected by w, especially at low redshifts, where dark energy becomes more important (see e.g. Amendola & Tsujikawa 2010). At any given redshift z, a larger w makes the dark energy component more important compared to the matter component, and since the growth rate $f(z)$ increases by increasing the dark matter component, it decreases by increasing w. This is exactly what we see in Fig. 3 for the three values of $w = -1.2$, $w = -1$, and $w = -0.8$. We also see that, as expected, the differences between the three spectra at smaller scales are significantly reduced when we increase the bin's redshift. The dependence of the power spectrum on the value of w is, however, much more involved for very large scales, as the spectrum on those scales is determined by a combination of different w-dependent effects, such as the integrated Sachs–Wolfe effect. Finally, in all the three bins of the upper right panel of Fig. 3, the oscillatory features in the ratios $C_\ell/C_\ell^{\text{fid}}$ are a consequence of the fact that the baryon acoustic oscillations shift towards smaller scales with increasing redshift for both $w = -1.2$ and $w = -0.8$.

When considering the sum of the neutrino masses, we see that increasing $\sum m_\nu$ results in the suppression of power on all scales and in all redshift bins, although this suppression is significantly stronger at smaller scales (or higher multipoles). There are several reasons for the small-scale reduction of the power spectra in the presence of massive neutrinos (see e.g. Lesgourgues & Pastor 2012), the most important of which is the absence of neutrino perturbations in the total power spectrum and a slower growth rate of matter perturbations at late times. On extremely large scales, however, neutrino free-streaming can be ignored (see e.g. Lesgourgues & Pastor 2012) and neutrino perturbations are therefore indistinguishable from matter perturbations. The power spectra then depend only on the total matter + neutrino density fraction today and on the primordial power spectrum. The small suppression of the angular power spectra at ultralarge scales, as seen in Fig. 3, is therefore because of the contribution of massive neutrinos to the total density parameter Ω_m.

4.2 Primordial non-Gaussianity and scale-dependent bias

An important extension of the standard ΛCDM model for our studies of ultralarge scales is ΛCDM+f_{NL}, where the parameter f_{NL} is added to the model in order to capture the effects of a non-zero local primordial non-Gaussianity. It has been shown (Dalal et al. 2008; Matarrese & Verde 2008; Slosar et al. 2008) that a local PNG modifies the Gaussian bias by contributing a scale-dependent piece of the form

$$\Delta b(z, k) = 3[b_{\text{lin}}(z) - 1]\frac{\delta_c \, \Omega_m \, H_0^2}{k^2 \, T(k) \, D(z)} f_{\text{NL}}, \tag{22}$$

where Ω_m is the present-day matter density parameter, H_0 is the value of the Hubble expansion rate today, $T(k)$ is the matter transfer function (with $T \to 1$ as $k \to 0$), $D(z)$ is the linear growth factor normalized

to $(1 + z)^{-1}$ in the matter-dominated Universe, and $\delta_c \sim 1.68$ is the (linear) critical matter density threshold for spherical collapse. The appearance of the k^2 factor in the denominator of equation (22) immediately tells us that ultralarge scales are the natural choice for placing constraints on f_{NL} using this scale-dependent bias, as the signal becomes stronger when $k \to 0$.

The lower right panel of Fig. 3 shows the effects of non-zero values of f_{NL} on the power spectrum at large scales – note that similar to the previous cases, the spectra are exact, i.e. no approximations are made in their computations. We first notice that, as expected, a non-zero f_{NL} only affects the ultralarge scales substantially, by enhancing or suppressing the power spectra, and that this happens in all the redshift bins shown in the figure. This again emphasizes the importance of accurately and precisely measuring the power spectra at ultralarge scales, as even the unrealistically large values of $f_{\text{NL}} = \pm 20$ (see Planck Collaboration VI 2020c for the current observational constraints on f_{NL}) shown in the figure affect the spectra by only < 5 per cent.

The figure also shows that a negative (positive) f_{NL} enhances (suppresses) the spectra for the two low-redshift bins 5 and 10, while the effect is the opposite for the high-redshift bin 15. Here, we explain the reason for this surprising but important feature. For that, let us investigate the redshift dependence of equation (22) for the full bias. The quantity b_{lin} is redshift-dependent and is given by equation (21) for the survey we consider in this paper. As can be seen in the lower panel of Fig. 2, the quantity $b_{\text{lin}} - 1$ is negative for $z \lesssim 0.75$ and positive for $z \gtrsim 0.75$, which means that a negative (positive) f_{NL} enhances the bias at low (high) redshifts and suppresses it at high (low) redshifts. Now looking at the upper panel of Fig. 2, we see that the two upper bins of the lower right panel of Fig. 3 (bins 5 and 10) contain redshifts that are lower than 0.75, while the lower bin (bin 15) includes redshifts higher than 0.75.

It is, however, important to note that this is the case only if one assumes a $b_{\text{lin}} - 1$ factor in equation (22), whose validity has been questioned in the literature (see e.g. Barreira 2020 and references therein). For this reason, we modify equation (22) as

$$\Delta b(z, k) = 3 f_{\text{NL}}[b_{\text{lin}}(z) - p]\delta_c \frac{\Omega_m H_0^2}{k^2 T(k) D(z)}, \tag{23}$$

where p is now a free parameter to be determined by cosmological simulations. It is argued by Barreira (2020) that $p = 1$ for gravity-only dynamics and when universality of the halo mass function is assumed, while other values of p provide better descriptions of observed galaxies where both of these assumptions are violated. Depending on the specific analysis and modelling, different values of p have been obtained, e.g. Slosar et al. (2008) and Pillepich, Porciani & Hahn (2010) showed that $p = 1.6$ provides a better description of host halo mergers, while Barreira et al. (2020) showed that $p = 0.55$ better describes IllustrisTNG-simulated stellar-mass-selected galaxies.

5 PARAMETER ESTIMATION METHODOLOGY

In this paper, we are interested in estimating the impact of large-scale effects and approximations on the estimation of cosmological parameters. In order to do so, we fit the mock data set obtained by the exact C_ℓ^{Ex} spectra as described in Section 3 using the C_ℓ^{Ap} spectra which make use of the several common approximations discussed in Section 2.2.

Throughout this work we rely on CAMB (Lewis et al. 2000; Howlett et al. 2012) to compute the exact and approximated power spectra. We use a modified version of the code, following Camera, Santos &

Maartens (2015c), when we consider the primordial non-Gaussianity parameter, f_{NL}. We implement in the public code COBAYA[5] (Torrado & Lewis 2020) a new likelihood module which enables us to obtain from CAMB the approximated spectra C_ℓ^{Ap} and compare them with the mock data set. Such an analysis matches the approach commonly used for parameter estimation with galaxy number count data, where C_ℓ^{Ap} is computed at each step in the MCMC rather than C_ℓ^{Ex}, as the computation of the latter is extremely time consuming and therefore unfeasible to repeat $\mathcal{O}(10^4)$ times.

For each point $\boldsymbol{\Theta}$ in the sampled parameter space, we compute the χ^2 using the approach presented in Audren et al. (2013), i.e.

$$\Delta\chi^2(\boldsymbol{\Theta}) = \sum_\ell (2\ell+1) f_{sky} \left(\frac{d_\ell^{mix}(\boldsymbol{\Theta})}{d_\ell^{th}(\boldsymbol{\Theta})} + \ln \frac{d_\ell^{th}(\boldsymbol{\Theta})}{d_\ell^{obs}} - N_{bin} \right), \quad (24)$$

where N_{bin} is the number of bins, and

$$d_\ell^{th}(\boldsymbol{\Theta}) = \det\left[\tilde{C}_{ij\ell}^{Ap}(\boldsymbol{\Theta}) \right], \quad (25)$$

$$d_\ell^{obs} = \det\left[\tilde{C}_{ij\ell}^{Ex}(\boldsymbol{\Theta}^{fid}) \right]. \quad (26)$$

The tilde indicates that the used spectra contain an observational noise $N_{ij\ell} = \delta_{ij}^K/n_i$, with n_i the number of galaxies in the ith bin and δ_{ij}^K the Kronecker delta, i.e. $\tilde{C}_{ij\ell} = C_{ij\ell} + N_{ij\ell}$. The quantity $d_\ell^{mix}(\boldsymbol{\Theta})$ in equation (24) is constructed from $d_\ell^{th}(\boldsymbol{\Theta})$ by replacing, one after each other, the theoretical spectra with the corresponding observational ones (for details, see Audren et al. 2013).

Note that equation (24) allows one to compute the difference between the χ^2 at each point in the parameter space and its minimum value, with $\Delta\chi^2$ vanishing when computed using the fiducial values of our free parameters. This is the quantity that we compute at each step of our MCMC, i.e. a constant rescaling of the χ^2 by an offset, which therefore correctly samples both the peak and the shape of the posterior, as it does not change the dependency of the χ^2 on the free parameters.

For currently available observations, which are not able to survey extremely large volumes of the Universe and therefore do not explore the ultralarge-scale regime, the approximated spectra generally mimic the true power spectrum. Thus, the approximations made do not significantly affect the results. However, we expect future surveys such as the HI-galaxy redshift survey for which we generated the mock data set in Section 3, to provide data at scales where lensing, RSD, relativistic effects, and the Limber approximation significantly impact the power spectra. Consequently, using the different approximations presented in Section 2.2 in fitting the models to the data will likely lead to shifts in the inferred cosmological parameters with respect to the fiducial values used to generate the data set. In this paper, we quantify the significance of these shifts, in units of σ, as

$$S(\Theta) = \frac{|\Theta - \Theta^{fid}|}{\sigma_\Theta}, \quad (27)$$

where Θ is a generic parameter of the full set $\boldsymbol{\Theta}$ estimated in our analysis, σ_Θ is the Gaussian error we obtain on Θ, and Θ^{fid} is the fiducial value of Θ used to generate the mock data set.

We apply this pipeline to the models described in Section 4, with the baseline ΛCDM model described by the set of five free parameters $\boldsymbol{\Theta} = \{\omega_b, \omega_c, h, A_s, n_s\}$. When analysing an extended model, we add one extra free parameter to this set: the dark energy equation of state w, the sum of the neutrino masses $\sum m_\nu$, or the local primordial non-Gaussianity parameter f_{NL}. We adopt flat priors on all these parameters.
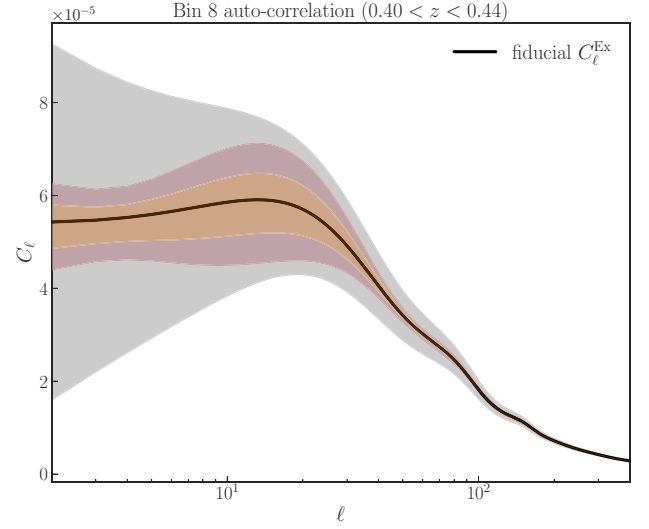
Figure 4. Effect of debiasing when different $\boldsymbol{\Theta}_0$ points in the parameter space are used to compute the debiasing term $\alpha(\boldsymbol{\Theta}^0)$. Here the autocorrelation in the eighth redshift bin is shown as a typical example. $\alpha(\boldsymbol{\Theta}^0)$ is computed at the fiducial set of parameters, $\boldsymbol{\Theta}^{fid}$, and at 500 other points in the parameter space, randomly sampled from a Gaussian distribution centred at $\boldsymbol{\Theta}^{fid}$ with a variance of 10 per cent for each parameter. The black solid curve shows the fiducial $C_\ell^{Ex}(\boldsymbol{\Theta}^{fid})$ spectrum, while the grey band shows the errors corresponding to the experimental setup considered in the paper. For each of the 500 computed $\alpha(\boldsymbol{\Theta}^0)$, the debiasing term is applied to the $C_\ell^{Ap}(\boldsymbol{\Theta}^{fid})$ spectrum. Assuming that the resulting spectra also follow a Gaussian distribution around the $C_\ell^{Ap}(\boldsymbol{\Theta}^{fid}) + \alpha(\boldsymbol{\Theta}^{fid})$ spectra, the orange and red areas show the 1σ and 2σ uncertainty regions, respectively.

Note that here we consider an optimistic setting in which the linear galaxy bias $b_{lin}(z)$ is perfectly known. Adding nuisance parameters accounting for the uncertainty on this function and marginalizing over them would enlarge the errors on cosmological parameters and reduce the statistical significance of the shifts we find, but would not qualitatively change the effects we are interested in. Moreover, as we are interested in the largest scales, in our analysis we only consider the data up to the multipole $\ell = 400$. Adding smaller scales to the analysis could reduce the significance of the shifts, but would not change our results qualitatively.

5.1 Debiasing constraints on cosmological parameters

As we will show in Section 6, using approximated spectra, C_ℓ^{Ap}, in the MCMC analysis results in significant shifts on cosmological parameters. To mitigate this, we propose a method for debiasing the parameter estimates while still allowing for the use of the quickly computed C_ℓ^{Ap}. This method is based on adding a correction to the C_ℓ's used in the likelihood evaluation as

$$C_\ell^{Ap}(\boldsymbol{\Theta}) \rightarrow C_\ell^{Ap}(\boldsymbol{\Theta}) + \left[C_\ell^{Ex}(\boldsymbol{\Theta}^0) - C_\ell^{Ap}(\boldsymbol{\Theta}^0) \right], \quad (28)$$

where $\boldsymbol{\Theta}^0$ refers to a specific set of the cosmological parameters. We define the debiasing term $\alpha(\boldsymbol{\Theta}^0)$ as

$$\alpha(\boldsymbol{\Theta}^0) \equiv C_\ell^{Ex}(\boldsymbol{\Theta}^0) - C_\ell^{Ap}(\boldsymbol{\Theta}^0). \quad (29)$$

We use $\boldsymbol{\Theta}^0 = \boldsymbol{\Theta}^{fid}$ for most of the results shown below, but discuss in Section 5.2 how we can use a maximum likelihood estimate of $\boldsymbol{\Theta}^0$ when working with actual data for which $\boldsymbol{\Theta}^{fid}$ does not exist. In Fig. 4 we show the dependence of this debiasing method on the choice of $\boldsymbol{\Theta}^0$; we compute the debiasing term at $\boldsymbol{\Theta}^{fid}$ and at 500 other points of

Table 2. Maximum likelihood (ML) parameter set obtained by minimizing the χ^2 when C_ℓ^{Ex} is used to fit the data set described in Section 3. The values are obtained through the `BOBYQA` minimization algorithm implemented in `Cobaya`.

Θ	Fiducial value	ML (or peak) value
ω_b	0.022 445	0.022 485
ω_c	0.1206	0.1209
h	0.67	0.67
$A_s \times 10^{-9}$	2.126 05	2.11
n_s	0.96	0.96

the parameter space, randomly sampled from a Gaussian distribution centred at Θ^{fid} with a variance on each parameter corresponding to 10 per cent of its fiducial value. These debiasing terms are then applied to $C_\ell^{\text{Ap}}(\Theta^{\text{fid}})$. Assuming that the resulting spectra also follow a Gaussian distribution around the $C_\ell^{\text{Ap}} + \alpha(\Theta^{\text{fid}})$ spectra, we show the corresponding 1σ and 2σ uncertainty regions. The figure shows that although the results we present below are based on computing $\alpha(\Theta^0)$ using Θ^{fid}, which would not be known in the case of actual data, our results would also hold for other choices of Θ^0 if they were reasonably close to Θ^{fid}. This method of debiasing cosmological parameter estimates works precisely because the debiasing term $\alpha(\Theta^0)$ is not strongly dependent on the choice of Θ^0 and can therefore account for the differences between the exact and approximated spectra over the full range of parameter space that the MCMC explores. Since $\alpha(\Theta^0)$ only needs to be computed at a single set of parameter values, rather than each step in the MCMC, it allows one to obtain unbiased results without being computationally expensive, unlike using C_ℓ^{Ex} which makes the analysis unfeasible.

We therefore use, at each sampled point Θ, the χ^2 expression of equation (24), but with the substitution

$$\tilde{C}_{ij\ell}^{\text{Ap}}(\Theta) \to \tilde{C}_{ij\ell}^{\text{Ap}}(\Theta) + \alpha(\Theta^0). \tag{30}$$

5.2 Debiasing with maximum likelihood

While in this paper we work with mock data sets, and therefore Θ^{fid} is known, this will not be the case when analysing real data. In order to use our approach in a realistic setting, we need to find a point in the parameter space that approximates the fiducial cosmology, which corresponds to the peak of the multivariate posterior probability distribution for the parameters. This can be achieved by analysing the mock data set built with C_ℓ^{Ex} using the correct theoretical predictions, but without attempting to reconstruct the full shape of the posterior distribution. One can use maximization methods to find the peak of the distribution, and since these methods only aim to find the maximum likelihood (or best-fitting) point in the parameter space, they require a significantly smaller number of iterations with respect to MCMC methods.

Here, we use the maximization pipeline of `Cobaya`, which relies on the `BOBYQA` algorithm (Cartis et al. 2018a; Cartis, Roberts & Sheridan-Methven 2018b), to fit the C_ℓ^{Ex} spectra to our mock data set, and we find the maximum likelihood parameter set presented in Table 2. The maximum likelihood point (Θ^{peak}) found with this method is very close to the actual fiducial point used to generate the data set and would therefore be suitable for computing the debiasing term α (see Section 5.1). Although we use the fiducial parameter set Θ^{fid} to compute the debiasing term in the rest of this paper, we have verified that there would be no significant changes in our results if Θ^{peak} were chosen instead (see Section 6.1).

We find that the maximization approach is much less computationally expensive compared to running a full MCMC with the exact spectra. A single iteration of our likelihood code using the exact spectra takes ∼150 s (compared to ∼5 s with the approximated spectra). The number of accepted iterations before reaching convergence[6] with the approximated spectra is 37 500. If we assume this to be the minimum number of iterations needed, the MCMC with the exact spectra would take at least 65 d (and likely much longer when taking into account the rejected steps). In contrast, the likelihood maximization took under 1 month running on a workstation with many other background processes, and the MCMC run that followed (with the approximated spectra) took only 2 d, demonstrating the computational feasibility of our approach.

We want to stress, however, that this minimization approach might fail should the posterior distribution be complicated; the presence of multiple peaks or very flat posteriors might bias the estimate of the maximum likelihood point in the parameter space and therefore possibly hinder the feasibility of this approach.

6 RESULTS AND DISCUSSION

In this section, we present the results of our analysis, highlighting how neglecting effects that are relevant at very large scales can result in significant biases in the estimation of cosmological parameters, potentially leading to false detections of non-standard physics. We split our results in two sub-sections, the first focusing on ΛCDM and its simple extensions ΛCDM$+m_\nu$ and wCDM, and the second discussing the results obtained when a scale-dependent bias generated by primordial non-Gaussianity is included in the analysis. All the MCMC samples obtained using the methodology described in the previous section are analysed using the public code `GetDist`[7] (Lewis 2019).

6.1 ΛCDM and its simple extensions

In Table 3, we present the results obtained by analysing our mock data set, generated with C_ℓ^{Ex} spectra for a ΛCDM fiducial cosmology, using C_ℓ^{Ap} spectra for the three assumed cosmologies ΛCDM, ΛCDM$+m_\nu$, and wCDM. In the first case, we find that the obtained constraints on cosmological parameters are significantly shifted with respect to their fiducial values, despite using for the analysis the same cosmological model as the one assumed in generating the mock data set. With the exception of A_s, which affects the amplitude of the spectra, the other parameters are all shifted by more than 2σ, with n_s being the most affected parameter ($S = 4.9\sigma$), as a result of using approximations to achieve a reasonable computation time for the MCMC analysis. When we allow for simple extensions of ΛCDM, we see that such an effect leads to significant false detections of departures from the standard model. With the sum of the neutrino masses $\sum m_\nu$ added as an extra free parameter, we indeed find a significant detection of a non-vanishing value, where $\sum m_\nu = 0$ eV is excluded with more than 6σ significance and the estimated value is shifted from the fiducial minimal value $\sum m_\nu = 0.06$ eV by $S = 6.2\sigma$; this implies that an analysis of data sensitive to large-scale effects would provide a false detection of the neutrino masses if one used the approximations considered here. The same effect can be seen if one allows for dark energy with an equation-of-state parameter

---

[6]The convergence criteria used by `Cobaya` are that the Gelman–Rubin R-1 on the means be <0.01 and that on the standard deviations be <0.2.
[7]https://github.com/cmbant/getdist

Ultralarge-scale approximations

Table 3. *Top table:* marginalized constraints on the sampled parameters Θ and values of the shift estimator $S(\Theta)$ obtained by analysing the fiducial data set with the approximated C_ℓ^{Ap} spectra for the standard ΛCDM model and its simple extensions, ΛCDM+m_ν, and wCDM, considered in the present work. The last column of the table shows the minimum $\Delta\chi^2$ values obtained for the different cosmologies, which, given equation (24), should vanish for an unbiased analysis. *Bottom table:* same as the top table, but applying the debiasing term $\alpha(\Theta^{\mathrm{fid}})$ to the theoretical predictions that are then compared with the data.

			Cosmological parameters							$\Delta\chi^2_{\min}$
			ω_{b}	ω_{c}	h	$A_{\mathrm{s}} \times 10^9$	n_{s}	$\sum m_\nu$ [eV]	w	
		Θ^{fid}	0.22445	0.1206	0.67	2.12605	0.96	0.06	-1	
Biased results	ΛCDM	Θ	$0.0163^{+0.0016}_{-0.0018}$	0.1098 ± 0.0046	0.616 ± 0.018	$2.176^{+0.068}_{-0.081}$	0.9948 ± 0.0071	–	–	4972
		$S(\Theta)[\sigma]$	3.6	2.3	3.0	0.7	4.9	–	–	
	$+m_\nu$	Θ	$0.0160^{+0.0015}_{-0.0018}$	$0.1143^{+0.0044}_{-0.0052}$	$0.615^{+0.017}_{-0.019}$	2.290 ± 0.082	0.9872 ± 0.0074	0.327 ± 0.043	–	4925
		$S(\Theta)[\sigma]$	3.8	1.3	3.0	2.0	3.7	6.2	–	
	wCDM	Θ	$0.0156^{+0.0013}_{-0.0016}$	0.1014 ± 0.0039	0.600 ± 0.015	$2.398^{+0.079}_{-0.088}$	1.0494 ± 0.0096	–	-0.886 ± 0.013	4912
		$S(\Theta)[\sigma]$	4.7	4.9	4.7	3.2	9.3	–	8.7	
Debiased results	ΛCDM	Θ	$0.0233^{+0.0021}_{-0.0031}$	$0.1227^{+0.0053}_{-0.0074}$	$0.677^{+0.019}_{-0.026}$	2.104 ± 0.078	$0.9573^{+0.0088}_{-0.0074}$	–	–	4.02
		$S(\Theta)[\sigma]$	0.31	0.33	0.30	0.28	0.34	–	–	
	$+m_\nu$	Θ	$0.0235^{+0.0022}_{-0.0032}$	$0.1238^{+0.0055}_{-0.0078}$	$0.679^{+0.020}_{-0.027}$	2.113 ± 0.091	$0.9563^{+0.0094}_{-0.0080}$	<0.115	–	4.11
		$S(\Theta)[\sigma]$	0.39	0.46	0.39	0.15	0.41	0.50	–	
	wCDM	Θ	$0.0233^{+0.0021}_{-0.0032}$	$0.1228^{+0.0054}_{-0.0078}$	$0.677^{+0.019}_{-0.027}$	2.103 ± 0.087	$0.957^{+0.012}_{-0.010}$	–	$-1.001^{+0.015}_{-0.013}$	0.07
		$S(\Theta)[\sigma]$	0.30	0.31	0.29	0.27	0.28	–	0.07	

that deviates from the cosmological constant value ($w = -1$). In this case, the free parameter w is shifted from the fiducial value by $S = 8.7\sigma$, resulting in a significant detection of a non-standard behaviour, which is driven only by the use of the approximated C_ℓ^{Ap} in the parameter estimation pipeline. Also in these extended cases, the estimated values of the standard cosmological parameters are shifted with respect to the fiducial ones. This highlights how these simple extensions alone are not able to mimic the C_ℓ^{Ex} spectra, as shifts in the values of the standard parameters are also necessary for fitting the C_ℓ^{Ex} to the data when C_ℓ^{Ap} are being used. The new degeneracies introduced by extensions of the ΛCDM model explain the changes in values of S with respect to the standard model.

We also show in Table 3 the minimum value of the $\Delta\chi^2$ found by analysing the posterior distribution reconstructed with the MCMC ($\Delta\chi^2_{\min}$). The decrease in the values of $\Delta\chi^2_{\min}$ for the extended models with respect to ΛCDM shows that a false detection of the extensions allows the approximated spectra to be in better agreement with the data. However, given that we expect from equation (24) to obtain a $\Delta\chi^2_{\min}$ close to zero if the theoretical spectra match the data, the values shown in Table 3 highlight how even these significant shifts in the C_ℓ^{Ap} spectra are not able to reproduce the cosmology used to generate the data. Notice that here we are not suggesting that the reduction in $\Delta\chi^2_{\min}$ is pointing towards a statistical preference for one model over the other; such a comparison would require using Bayesian model comparison techniques also accounting for the number of free parameters of a given model. Moreover, the $\Delta\chi^2_{\min}$ values are estimates that might be slightly different from the real minimum value, as it is not guaranteed that the MCMC is able to perfectly sample the peak point in the parameter space. Thus, with such small differences between different models, a more accurate computation of $\Delta\chi^2_{\min}$ would be needed if one wanted to perform model comparison.

In Fig. 5, we show the 68 per cent and 95 per cent confidence level contours on a few representative parameters for the cases described above. The colour-filled contours show the results of the analysis performed with C_ℓ^{Ap}, highlighting the deviation of the estimated values of the parameters from the fiducial values (shown with black dashed lines). The empty contours instead show the results obtained when the debiasing term described in Section 5.1 is added to the spectra, which are then compared to the mock data set. These results

show how the method we propose is able to debias the results and how it allows us to recover the correct values for the parameters, for both the standard ΛCDM cosmology and its extensions, thus avoiding false detections of non-standard cosmologies and improving the goodness of fit with a χ^2 now of $\mathcal{O}(1)$.

In order to see in more detail the biasing effect of the approximations included in the C_ℓ^{Ap}, we show in Fig. 6 the impact of the biases on the angular power spectra for a representative redshift bin autocorrelation, highlighting how the approximated C_ℓ^{Ap} spectrum (green) significantly departs from the expected C_ℓ^{Ex} spectrum (black) when the fiducial values of the cosmological parameters are used to obtain both. We also include, with a red dashed curve, the C_ℓ^{Ap} spectrum obtained using the biased values of the cosmological parameters reported in Table 3, showing how in this case the C_ℓ^{Ap} at the shifted best-fitting cosmology are better able to reproduce the fiducial C_ℓ^{Ex}, thus producing a better fit to the data.

While in Fig. 5 we only show a sub-set of the free parameters of our models, the debiasing procedure is effective for all cosmological parameters. In Fig. 7, we show the constraints obtained on all the free parameters of our ΛCDM analysis, obtained by both comparing the C_ℓ^{Ap} to the mock data set (red, filled contours) and applying the debiasing method of Section 5.1, using the debiasing term computed at both the fiducial values, $\alpha(\Theta^{\mathrm{fid}})$ (yellow, filled contours), and the peak values found in Section 5.2, $\alpha(\Theta^{\mathrm{peak}})$ (purple, empty contours). We notice how in the first case all the parameters are shifted with respect to their expected values, with the most significant shifts on n_s, ω_{b}, and h, while when we apply the debiasing approach the fiducial values are recovered for all the parameters, with no significant differences between the two cases of $\alpha(\Theta^{\mathrm{fid}})$ and $\alpha(\Theta^{\mathrm{peak}})$. Even though the results shown in Fig. 7 correspond to the ΛCDM model, they are qualitatively similar for all the considered cosmologies.

The posterior probability distributions of the parameters recovered after debiasing the MCMC results do not necessarily coincide with those that would be obtained by a full analysis. We can, however, consider these as reasonable estimates, as Fig. 4 shows that the debiasing term does not depend strongly on the Θ^0 point at which it is computed, as long as $\alpha(\Theta^0)$ is sufficiently close to Θ^{fid}. Thus, rather than computing $\alpha(\Theta^0)$ at each point in the parameter space, we can approximate $\alpha(\Theta^0)$ with $\alpha(\Theta^{\mathrm{peak}})$ (or $\alpha(\Theta^{\mathrm{fid}})$ in the case of the results shown here). This applies only in the vicinity of the peak
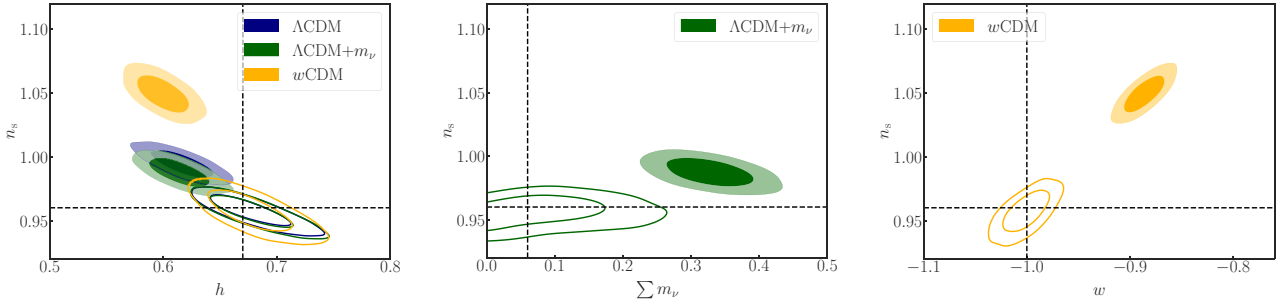
Figure 5. 68 per cent and 95 per cent confidence level contours obtained by fitting the approximated C_ℓ^{Ap} spectra to the data set built using the exact C_ℓ^{Ex} spectra (colour-filled contours). The violet contours show the result when a ΛCDM model is assumed, while the green and orange contours correspond instead to ΛCDM+m_ν and wCDM cosmologies, respectively. The empty contours show the results of the analysis when the debiasing term described in Section 5.1 is included. The black dashed lines show the fiducial values of the cosmological parameters.
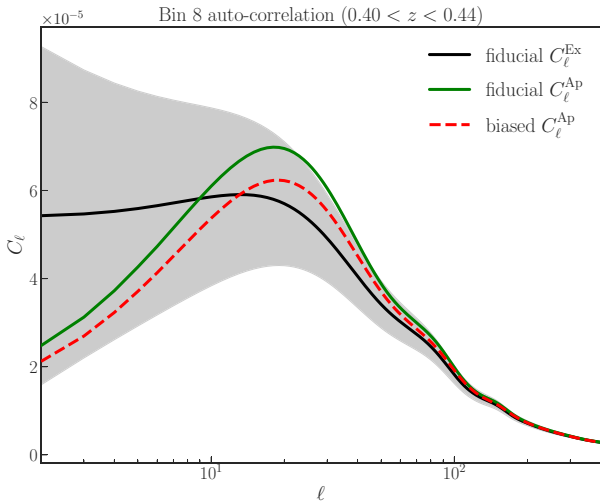
Figure 6. Angular power spectra for the eighth redshift bin autocorrelation in a ΛCDM cosmology using the exact C_ℓ^{Ex} (black solid curve) and the approximated C_ℓ^{Ap} (green solid curve) obtained assuming the fiducial values for the cosmological parameters. The red dashed curve shows the C_ℓ^{Ap} obtained for the biased parameter estimation of Table 3. The grey area shows the errors corresponding to the experimental setup used throughout the paper.

of the distribution, and the estimation of the tails suffers from an error that propagates into the confidence intervals shown in Figs 5 and 7. We leave a quantification of this error for future work.

6.2 Primordial non-Gaussianity

In this sub-section, we focus on the results when f_{NL} is included as a free parameter, thus allowing for a non-vanishing local primordial non-Gaussianity; this affects the galaxy clustering spectra through the scale-dependent bias as described in Section 4.2. As a first case, we use the same experimental setup we used in Section 6.1, and use the standard expression of equation (22) for our theoretical predictions for the scale-dependent bias. In this case, which we refer to as 'baseline', when we analyse the mock data set using the approximated C_ℓ^{Ap} we find results that are similar to the ΛCDM case of Section 6.1, with approximately the same shifts for the standard parameters and no bias for f_{NL} (see Table 4). This may seem to be a surprising result, as the impact of f_{NL} on the theoretical predictions is significant at very large scales (see Fig. 3), where the approximations included in C_ℓ^{Ap} fail. One would therefore expect that a biased value
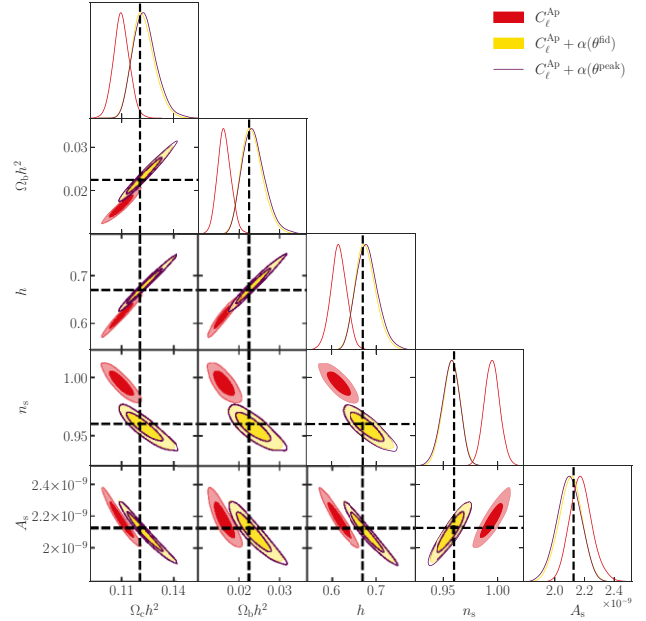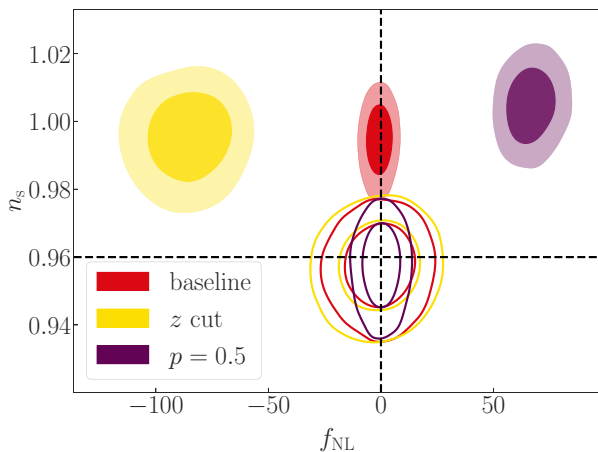

Figure 7. 68 per cent and 95 per cent confidence level contours, as well as one-dimensional marginalized posterior probability distribution functions, obtained by fitting the ΛCDM model to the mock data set. The red, filled contours correspond to the analysis where the theoretically predicted approximated spectra C_ℓ^{Ap} are used to fit the model to the mock data set. The yellow, filled contours show the results obtained when the debiasing term $\alpha(\Theta^{fid})$ is included, and the purple, empty contours correspond to the debiasing term $\alpha(\Theta^{peak})$ computed at the estimated maximum likelihood point Θ^{peak}.

for this parameter would help with fitting the C_ℓ^{Ex} spectra of the mock data set, and that a false non-vanishing f_{NL} would be detected. However, given equation (22) that we rely upon, the scale-dependent bias depends not only on f_{NL}, but also on the $b_{lin} - 1$ factor. As shown in Fig. 2 and discussed in Section 4.2, our choice of the linear galaxy bias implies that $b_{lin} - 1$ changes sign at $z \approx 0.75$; the impact of f_{NL} on the C_ℓ^{Ap} spectra is therefore the opposite for the redshift bins beyond this redshift threshold with respect to the lower redshift ones. Such an effect leads to a cancellation of the impact of the primordial non-Gaussianity on the goodness of fit, and therefore the standard case of $f_{NL} = 0$ is still preferred.

In order to ensure that this indeed is the reason for the lack of shift in the recovered f_{NL} value, we run our parameter estimation pipeline

Table 4. *Top table:* marginalized constraints on the sampled parameters Θ and values of the shift estimator $S(\Theta)$ obtained by analysing the fiducial data set with the approximated C_ℓ^{Ap} spectra for the three cases of $\Lambda CDM+f_{NL}$ considered in the present work. *Bottom table:* same as the top one, but applying the debiasing term $\alpha(\Theta^{fid})$ to the theoretical predictions that are then compared with the data.

			Cosmological parameters						$\Delta\chi^2_{min}$
			ω_b	ω_c	h	$A_s \times 10^9$	n_s	f_{NL}	
		Θ^{fid}	0.224 45	0.1206	0.67	2.12605	0.96	0	
Biased results	baseline	Θ	$0.0164^{+0.0016}_{-0.0019}$	$0.1101^{+0.0044}_{-0.0050}$	0.617 ± 0.018	2.170 ± 0.077	0.9945 ± 0.0070	-0.8 ± 3.9	4985
		$S(\Theta)\,[\sigma]$	3.4	2.2	2.9	0.6	4.9	0.2	
	z cut	Θ	$0.0219^{+0.0024}_{-0.0035}$	$0.1300^{+0.0064}_{-0.0089}$	$0.679^{+0.023}_{-0.031}$	1.839 ± 0.084	$0.9955^{+0.0093}_{-0.0081}$	-85^{+13}_{-12}	1175
		$S(\Theta)\,[\sigma]$	0.2	1.2	0.3	3.4	4.0	6.9	
	$p = 0.5$	Θ	$0.0169^{+0.0017}_{-0.0021}$	$0.1143^{+0.0048}_{-0.0056}$	0.620 ± 0.020	2.069 ± 0.077	1.0047 ± 0.0075	66.5 ± 7.2	4755
		$S(\Theta)\,[\sigma]$	2.9	1.2	2.6	0.7	6.0	9.2	
Debiased results	baseline	Θ	$0.0235^{+0.0022}_{-0.0033}$	$0.1230^{+0.0054}_{-0.0079}$	$0.678^{+0.020}_{-0.028}$	$2.101^{+0.087}_{-0.077}$	$0.9573^{+0.0086}_{-0.0076}$	0 ± 10	17.8
		$S(\Theta)\,[\sigma]$	0.34	0.35	0.33	0.30	0.33	0.04	
	z cut	Θ	$0.0235^{+0.0025}_{-0.0038}$	$0.1232^{+0.0063}_{-0.0091}$	$0.679^{+0.023}_{-0.032}$	2.101 ± 0.096	$0.9573^{+0.0094}_{-0.0083}$	-1 ± 12	17.9
		$S(\Theta)\,[\sigma]$	0.31	0.32	0.30	0.26	0.30	0.11	
	$p = 0.5$	Θ	$0.0234^{+0.0022}_{-0.0033}$	$0.1229^{+0.0055}_{-0.0080}$	$0.678^{+0.020}_{-0.028}$	$2.103^{+0.088}_{-0.079}$	$0.9575^{+0.0088}_{-0.0077}$	0.1 ± 5.5	17.9
		$S(\Theta)\,[\sigma]$	0.31	0.32	0.30	0.27	0.30	0.03	

Figure 8. 68 per cent and 95 per cent confidence level contours obtained by fitting the approximated C_ℓ^{Ap} spectra, with a free f_{NL} parameter, to the data set built using the exact C_ℓ^{Ex} spectra (colour-filled contours). The red and yellow contours show the results obtained with the scale-dependent bias of equation (22), with our baseline settings and with removing the last two redshift bins, respectively. The violet contour shows instead the case where the scale-dependent bias is computed following equation (23) with $p = 0.5$. The empty contours show the results of the analysis when the debiasing term described in Section 5.1 is included. The black dashed lines show the values of the fiducial cosmological parameters.

by removing the redshift bins above $z \approx 0.75$. We refer to this case as 'z cut'. The results are shown in Table 4, where it can be seen how removing the higher redshift bins eliminates the cancellation effect described above; now we find significant biases on f_{NL} and A_s, with $S(f_{NL}) = 6.9\sigma$ and $S(A_s) = 3.4\sigma$, respectively, for the shifts with respect to the fiducial values. The shifts on the other free parameters are reduced with respect to the baseline case. The combined effect of f_{NL} and A_s allows the C_ℓ^{Ap} to fit the mock data set, as the global effect is boosting the power spectra at large scales.

On the other hand, as we discussed in Section 4.2, the modulating factor $b_{lin}(z) - 1$ in equation (22) is not the only possibility for describing the scale-dependent bias. We have repeated our analysis,

following the more general equation (23), by setting $p = 0.5$, which ensures that the $b_{lin}(z) - p$ factor does not change sign in our redshift range, given our choice of the linear galaxy bias. In the last two columns of Table 4 we report the results we find in this case, where we see again a significant false detection of a non-vanishing f_{NL}, with $S(f_{NL}) = 9.2\sigma$, while the other parameters are less shifted from their fiducial values compared to the baseline case, with the exception of n_s. In Fig. 8, we also notice how the shift on f_{NL} has an opposite sign in this $p = 0.5$ case with respect to the z cut case, where the analysis prefers a negative value of f_{NL}. This is due to the fact that the $b_{lin} - p$ factor is now always positive, and one needs an $f_{NL} > 0$ in order to achieve the boost in the C_ℓ^{Ap} needed to fit the model to the mock data set.

Finally, we apply the debiasing procedure of Section 5.1 to the three cases described and show the results in Fig. 8. As the figure shows, applying the debiasing correction allows us to recover a vanishing f_{NL}. The debiased contours are different from each other here, which was not the case in Section 6.1; this is due to the different strategies applied to account for the effects of f_{NL} in our analysis.

7 CONCLUSIONS

The continual improvement in galaxy surveys will soon unlock the largest scales in the sky for cosmological studies. While the expected angular correlations at smaller scales are well understood and efficiently modelled (up to the non-linear regime), calculations of power spectra commonly make use of approximations aimed at reducing the computational efforts needed to obtain theoretical predictions of the spectra. This is a necessary requirement for such calculations if one wants to exploit MCMC methods for performing parameter estimation analyses. Such approximations, however, break down at very large scales, where effects including lensing, galaxy velocities, and relativistic corrections become relevant.

In this paper, we have investigated the impact of approximations that neglect such large-scale effects on a parameter estimation analysis. We have produced a mock data set for a next-generation survey, with specifications based on those envisaged for the SKAO, which will be able to explore the angular correlation of galaxies at very large scales through the full treatment described in Section 2.1.

We have then analysed this data set by applying the commonly used approximations described in Section 2.2, where the large-scale corrections due to lensing, velocities, and relativistic effects have been neglected, and the Limber approximation has been employed. We have found that this analysis produces significantly biased results, with parameter estimates being shifted up to $\sim 5\sigma$ when assuming a minimal 5-parameter ΛCDM cosmology, and with false detections of non-standard cosmologies when simple extensions of the standard model are considered.

We have also explored the impact of the approximations on a more complex extension of the ΛCDM model, where we have allowed for a non-vanishing local primordial non-Gaussianity by including f_{NL} as a free parameter in our analysis. This contributes to a scale-dependent term to the galaxy bias which is relevant at large scales. We expected estimates of this parameter to be significantly biased, as a non-zero f_{NL} would help the approximated spectra to mimic those used in creating the data set. However, we have found that in our baseline setting, such an effect cannot be seen due to a cancellation between the low- and high-redshift bins. Given our choice of the linear galaxy bias (see Section 3), the commonly used scale-dependent term changes sign at $z \approx 0.75$ and therefore the effect of a non-vanishing f_{NL} on the overall goodness of fit cancels out between low- and high-redshift bins. We have confirmed this explanation by cutting out all bins at $z > 0.75$, and we have found, with this setting, a significant false detection of a non-vanishing and negative f_{NL}. We have also performed our analysis for a case where the scale-dependent piece of the bias depends differently on the linear bias term (equation 23). We have found in this case a 9.2σ shift in the estimated value of f_{NL}, opposite in sign with respect to the previous case, highlighting how different modellings of the scale-dependent term can affect the final results.

In this work, not only have we assessed the impact of the approximations on the estimation of cosmological parameters, but we have also proposed a simple method to obtain debiased results that can approximate those that one would obtain by taking into account all the effects. We have described this method in Section 5.1 and pointed out how the computation of the debiasing term $\alpha(\Theta^0)$ does not depend strongly on the choice of the parameter set Θ^0 where the computation is performed, as long as it is close to the true cosmology. Indeed, our advantage in using this method relies on the fact that, in our forecasts, the fiducial cosmology has been known. However, we have pointed out that in a realistic setting, with an unknown fiducial cosmology, one could rely on minimization algorithms to identify the best-fitting point in the parameter space. Such a minimization would be significantly less computationally expensive than a full parameter estimation pipeline and could therefore be performed using the exact spectra. We have tested the feasibility of such an approach, and we found in Section 5.2 that the debiased cosmological parameter constraints found using an estimate for the peak of the multivariate distribution are almost exactly the same as those found using the fiducial point. Thus, this method can be applied to real data, where the fiducial point is unknown.

We have applied the debiasing method to all the cases we have investigated, and we have found that it indeed allows us to recover the expected values for the free parameters of our analyses. This method could therefore be used in real data analysis when unexpected detections of non-standard behaviour are seen. Additionally, while not providing a fully correct parameter estimation, our method allows one to obtain accurate values for cosmological parameters and estimates of their corresponding posterior probability distributions. While the recovered distributions are reasonable estimates of the ones

obtained through a full analysis, we leave a quantitative assessment of the errors on their shapes for future work.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abdalla F. B. et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 17, preprint (arXiv:1501.04035)
Aiola S., et al., 2020, J. Cosmol. Astropart. Phys., 12, 047
Alam S., et al., 2017, MNRAS, 470, 2617
Alonso D., Bull P., Ferreira P. G., Maartens R., Santos M. G., 2015, ApJ, 814, 145
Amendola L., Tsujikawa S., 2010, Dark Energy: Theory and Observations. Cambridge Univ. press, Cambridge
Amendola L. et al., 2013, Liv. Rev. Relativ., 16, 6
Amendola L. et al., 2018, Liv. Rev. Relativ., 21, 2
Anderson L. et al., 2012, MNRAS, 427, 3435
Assassi V., Simonović M., Zaldarriaga M., 2017, J. Cosmol. Astropart. Phys., 2017, 054
Audren B., Lesgourgues J., Bird S., Haehnelt M. G., Viel M., 2013, J. Cosmol. Astropart. Phys., 2013, 026
Baker T., Bull P., 2015, ApJ, 811, 116
Barreira A., 2020, J. Cosmol. Astropart. Phys., 12, 031
Barreira A., Cabass G., Schmidt F., Pillepich A., Nelson D., 2020, J. Cosmol. Astropart. Phys., 12, 013
Bertacca D., Maartens R., Clarkson C., 2014, J. Cosmol. Astropart. Phys., 2014, 037
Beutler F. et al., 2011, MNRAS, 416, 3017
Blake C. et al., 2011, MNRAS, 418, 1707
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 2011, 034
Bonvin C., Durrer R., 2011, Phys. Rev. D, 84, 063505
Bose B. et al., 2021, MNRAS, 508, 2479
Brown M. et al., 2015, in PoS,Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 23, preprint (arXiv:1501.03828)

Bull P., 2016, ApJ, 817, 26

Bull P., Camera S., Raccanelli A., Blake C., Ferreira P., Santos M., Schwarz D. J., 2015, in Pos, Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 24, preprint (arXiv:1501.04088)

Camera S. et al., 2015a, in Pos, Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 25, preprint (arXiv:1501.03851)

Camera S., Carbone C., Fedeli C., Moscardini L., 2015b, Phys. Rev. D, 91, 043533

Camera S., Santos M. G., Maartens R., 2015c, MNRAS, 448, 1035

Camera S., Maartens R., Santos M. G., 2015d, MNRAS, 451, L80

Campagne J. E., Neveu J., Plaszczynski S., 2017, A&A, 602, A72

CANTATA Collaboration, 2021, preprint (arXiv:2105.12582)

Cardona W., Durrer R., Kunz M., Montanari F., 2016, Phys. Rev. D, 94, 043007

Carlstrom J. E., et al., 2011, PASP, 123, 568

Cartis C., Fiala J., Marteau B., Roberts L., 2018a, preprint (arXiv:1804.00154)

Cartis C., Roberts L., Sheridan-Methven O., 2018b, preprint (arXiv:1812.11343)

Challinor A., Lewis A., 2011, Phys. Rev. D, 84, 043516

Chartier N., Wandelt B. D., 2021, MNRAS, 509, 2220

Chartier N., Wandelt B., Akrami Y., Villaescusa-Navarro F., 2021, MNRAS, 503, 1897

Cole S., et al., 2005, MNRAS, 362, 505

Dalal N., Dore O., Huterer D., Shirokov A., 2008, Phys. Rev., D77, 123514

Datta K. K., Choudhury T. R., Bharadwaj S., 2007, MNRAS, 378, 119

DES Collaboration, 2021, preprint (arXiv:2105.13549)

DES Collaboration, 2021, preprint (arXiv:2105.13549)

DESI Collaboration, 2016a, preprint (arXiv:1611.00036)

DESI Collaboration, 2016b, preprint (arXiv:1611.00037)

Di Dio E., Montanari F., Lesgourgues J., Durrer R., 2013, J. Cosmol. Astropart. Phys., 11, 044

Doré O. et al., 2014, preprint (arXiv:1412.4872)

Doré O. et al., 2018, preprint (arXiv:1805.05489)

Eisenstein D. J., et al., 2005, ApJ, 633, 560

Euclid Collaboration, 2020, A&A, 642, A191

Fonseca J., Camera S., Santos M., Maartens R., 2015, ApJ, 812, L22

Ghosh B., Durrer R., Sellentin E., 2018, J. Cosmol. Astropart. Phys., 2018, 008

Granett B. R. et al., 2012, MNRAS, 421, 251

Grasshorn Gebhardt H. S., Jeong D., 2018, Phys. Rev. D, 97, 023504

Hinshaw G. et al., 2013, ApJS, 208, 19

Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmol. Astropart. Phys., 1204, 027

Ivezić v., et al., 2019, ApJ, 873, 111

Jalilvand M., Ghosh B., Majerotto E., Bose B., Durrer R., Kunz M., 2020, Phys. Rev. D, 101, 043530

Joachimi B. et al., 2021, A&A, 646, A129

Kaiser N., 1992, ApJ, 388, 272

Kilbinger M., et al., 2017, MNRAS, 472, 2126

Köhlinger F., et al., 2017, MNRAS, 471, 4412

Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

Lesgourgues J., Pastor S., 2012, Adv. High Energy Phys., 2012, 608515

Lewis A., 2019, preprint (arXiv:1910.13970)

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Limber D. N., 1953, ApJ, 117, 134

Limber D. N., 1954, ApJ, 119, 655

Loureiro A. et al., 2019, MNRAS, 485, 326

LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506

LSST Dark Energy Science Collaboration, 2018, preprint (arXiv:1809.01669)

LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)

Maartens R., Fonseca J., Camera S., Jolicoeur S., Viljoen J.-A., Clarkson C., 2021, JCAP, 12, 009

Martinelli M., et al., 2021, A&A, 649, A100

Matarrese S., Verde L., 2008, ApJ, 677, L77

Matthewson W. L., Durrer R., 2021, J. Cosmol. Astropart. Phys., 02, 027

Obreschkow D., Rawlings S., 2009, ApJ, 703, 1890

Padmanabhan N. et al., 2007, MNRAS, 378, 852

Parkinson D. et al., 2012, Phys. Rev. D, 86, 103518

Percival W. J. et al., 2010, MNRAS, 401, 2148

Pillepich A., Porciani C., Hahn O., 2010, MNRAS, 402, 191

Planck Collaboration, 2020a, A&A, 641, A1 I

Planck Collaboration, 2020b, A&A, 641, A6 VI

Planck Collaboration, 2020c, A&A, 641, A9 IX

Raccanelli A. et al., 2015, in Pos, Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 31, preprint (arXiv:1501.03821)

Safi S., Farhang M., 2021, ApJ, 914, 65

Santos M. et al., 2015, in PoS, Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 19, preprint (arXiv:1501.03989)

SKA Cosmology Science Working Group, 2020, PASA, 37, e007

Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, J. Cosmol. Astropart. Phys., 08, 031

Spergel D. et al., 2015, preprint (arXiv:1503.03757)

Sprenger T., Archidiacono M., Brinckmann T., Clesse S., Lesgourgues J., 2019, JCAP, 02, 047

Tanidis K., Camera S., 2019, MNRAS, 489, 3385

Tanidis K., Camera S., 2021, preprint (arXiv:2107.00026)

Tanidis K., Camera S., Parkinson D., 2020, MNRAS, 491, 4869

Thiele L., Duncan C. A. J., Alonso D., 2020, MNRAS, 491, 1746

Torrado J., Lewis A., 2020, JCAP, 05, 057

van Uitert E. et al., 2018, MNRAS, 476, 4662

Villa E., Di Dio E., Lepori F., 2018, J. Cosmol. Astropart. Phys., 04, 033

White M., Padmanabhan N., 2017, MNRAS, 471, 1167

Yahya S., Bull P., Santos M. G., Silva M., Maartens R., Okouma P., Bassett B., 2015, MNRAS, 450, 2251

Yoo J., 2010, Phys. Rev. D, 82, 083508

Yoo J., Seljak U., 2015, MNRAS, 447, 1789

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