

Galaxy bias in the era of LSST: perturbative bias expansions

The LSST Dark Energy Science collaboration

Andrina Nicola,^{a,*} Boryana Hadzhiyska,^{b,c} Nathan Findlay,^d Carlos García-García,^e
David Alonso,^e Anže Slosar,^f Zhiyuan Guo,^g Nickolas Kokron,^{h,i} Raúl Angulo,^{j,k}
Alejandro Aviles,^{l,m} Jonathan Blazek,ⁿ Jo Dunkley,^{o,p} Bhuvnesh Jain,^q Marcos Pellejero,^j
James Sullivan,^{r,s} Christopher W. Walter^g and Matteo Zennaro^e

^aArgelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

^bMiller Institute for Basic Research in Science, University of California, Berkeley, CA, 94720, U.S.A.

^cPhysics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, U.S.A.

^dInstitute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth, PO1 3FX, United Kingdom

^eDepartment of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom

^fBrookhaven National Laboratory, Physics Department, Upton, NY 11973, U.S.A.

^gDepartment of Physics, Duke University, Durham NC 27708, U.S.A.

^hKavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA, U.S.A.

ⁱKavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, Menlo Park, CA, U.S.A.

^jDonostia International Physics Center (DIPC), Paseo Manuel de Lardizabal, 4, 20018 Donostia-San Sebastián, Spain

^kIKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain

^lInstituto de Ciencias Físicas, Universidad Nacional Autónoma de México, 62210, Cuernavaca, Morelos, Mexico

^mConahcyt — Departamento de Física, Instituto Nacional de Investigaciones Nucleares, Apartado Postal 18-1027, Col. Escandón, Ciudad de México, 11801, Mexico

ⁿDepartment of Physics, Northeastern University, Boston, MA, 02115, U.S.A.

^oDepartment of Physics, Jadwin Hall, Princeton University, Princeton, NJ 08544, U.S.A.

^pDepartment of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, U.S.A.

*Corresponding author.

^q*Center for Particle Cosmology, Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.*

^r*Department of Astronomy, University of California, Berkeley, CA 94720, U.S.A.*

^s*Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, U.S.A.*

E-mail: anicola@uni-bonn.de

ABSTRACT: Upcoming imaging surveys will allow for high signal-to-noise measurements of galaxy clustering at small scales. In this work, we present the results of the Rubin Observatory Legacy Survey of Space and Time (LSST) bias challenge, the goal of which is to compare the performance of different nonlinear galaxy bias models in the context of LSST Year 10 (Y10) data. Specifically, we compare two perturbative approaches, Lagrangian perturbation theory (LPT) and Eulerian perturbation theory (EPT) to two variants of Hybrid Effective Field Theory (HEFT), with our fiducial implementation of these models including terms up to second order in the bias expansion as well as nonlocal bias and deviations from Poissonian stochasticity. We consider a variety of different simulated galaxy samples and test the performance of the bias models in a tomographic joint analysis of LSST-Y10-like galaxy clustering, galaxy-galaxy-lensing and cosmic shear. We find both HEFT methods as well as LPT and EPT combined with non-perturbative predictions for the matter power spectrum to yield unbiased constraints on cosmological parameters up to at least a maximal scale of $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ for all samples considered, even in the presence of assembly bias. While we find that we can reduce the complexity of the bias model for HEFT without compromising fit accuracy, this is not generally the case for the perturbative models. We find significant detections of non-Poissonian stochasticity in all cases considered, and our analysis shows evidence that small-scale galaxy clustering predominantly improves constraints on galaxy bias rather than cosmological parameters. These results therefore suggest that the systematic uncertainties associated with current nonlinear bias models are likely to be subdominant compared to other sources of error for tomographic analyses of upcoming photometric surveys, which bodes well for future galaxy clustering analyses using these high signal-to-noise data.

KEYWORDS: cosmological parameters from LSS, galaxy clustering, galaxy surveys, power spectrum

ARXIV EPRINT: [2307.03226](https://arxiv.org/abs/2307.03226)

Contents

1	Introduction	1
2	Perturbative bias models	3
2.1	Eulerian perturbation theory	4
2.2	Lagrangian perturbation theory	5
2.3	HEFT	6
2.4	Relations between bias models	7
2.5	Implementation choices	8
3	Observables	9
4	Simulations	11
5	Methods	13
5.1	Generating smooth spectra	13
5.2	Computing simulated data vectors	13
5.3	Deriving best-fit parameters and associated uncertainties	16
5.4	Assessing model performance	17
6	Results	19
6.1	HEFT	19
6.2	LPT/EPT	20
6.3	Bias consistency relations	22
6.4	Alternative samples	23
6.4.1	Magnitude-limited sample	24
6.4.2	Assembly bias	25
6.4.3	Including CMB lensing convergence	25
6.4.4	No HALOFIT rescaling	26
6.5	Stochasticity	26
6.6	Minimal bias model	28
6.7	Error as a function of maximal wavenumber	30
7	Conclusions	32
A	Description of methodology to create smoothed power spectra	35
B	Detailed description of goodness-of-fit tests	37
B.1	HEFT	37
B.2	LPT/EPT	38
C	Consistency of bias parameter values from different models	39
D	Comparison to MCMC results	40

E	Fitting data from UNIT	41
F	Three-dimensional power spectrum analysis	42
G	Consistency of results for different data sets	43

1 Introduction

The last three decades have seen the emergence of the Λ CDM cosmological model as the concordance model preferred by a number of different cosmological probes. Current and future surveys across all wavelengths will allow us to put this model to its most stringent test to date. In the optical wavelength range, these include the Dark Energy Survey (DES),¹ the Hyper-Suprime Cam Survey (HSC),² the Kilo Degree Survey (KiDS),³ the Dark Energy Camera Legacy Survey (DECaLS),⁴ the Baryon Oscillation Spectroscopic Survey (BOSS),⁵ the Dark Energy Spectroscopic Instrument (DESI),⁶ the Rubin Observatory Legacy Survey of Space and Time (LSST),⁷ Euclid,⁸ and the Roman Space Telescope.⁹

Alongside weak gravitational lensing, galaxy clustering is one of the main probes observable with these surveys. This powerful cosmological probe offers the promise to deliver tight constraints on modifications of Λ CDM, such as neutrino masses [1] and primordial non-Gaussianity [2], as well as the physics of galaxy formation. Theoretical modeling of galaxy clustering on small scales is hampered for two main reasons: first, on small scales the clustering of Dark Matter (DM) becomes non-linear. Second, on these scales, the relation between galaxy tracers and the DM field also becomes non-linear and mildly non-local. In the absence of baryons, the clustering of dark matter in the mildly nonlinear regime can be modeled either using analytic perturbative approaches or using N-body simulations (see e.g. refs. [3–5]). The relation between galaxies and dark matter on small scales depends on the physics of galaxy formation, which involves a variety of different processes and spans several orders of magnitude in scale. These processes are impossible to model ab-initio on a cosmological scale, even using the highest-accuracy hydrodynamic simulations (see e.g. ref. [6]). The relation between galaxies, or any biased tracer, and the underlying DM distribution therefore presents the largest theoretical systematic uncertainty in galaxy clustering analyses. Several approaches have been developed to model these effects, and they can be subdivided in two different categories: (i) perturbative models, and (ii) phenomenological models. The former models use a perturbative expansion to jointly model the non-linear evolution of the DM distribution and its relation to the distribution of biased tracers. This expansion can be either performed in the initial conditions or at late times, leading to two distinct

¹<https://www.darkenergysurvey.org/>.

²<https://hsc.mtk.nao.ac.jp/ssp/0>.

³<https://kids.strw.leidenuniv.nl/>.

⁴<https://www.legacysurvey.org/decamls/>.

⁵<https://www.sdss4.org/surveys/boos/>.

⁶<https://www.desi.lbl.gov/>.

⁷<https://www.lsst.org/>.

⁸<https://www.euclid-ec.org/>.

⁹<https://roman.gsfc.nasa.gov/>.

frameworks within which to model tracer bias, Lagrangian or Eulerian perturbation theory (PT) (see e.g. refs. [7–12]). The effective field theory of Large-Scale Structure (EFToLSS, hereafter we use EFT for short) presents a closely related approach that treats cosmological fields at mildly non-linear scales as effective fields emerging from a more complete theory describing small-scale structure formation [13, 14]. A number of studies have analyzed the reach of these methods and have found them to be accurate up to maximal wave numbers of $k_{\text{max}} \sim 0.1 - 0.3 h \text{ Mpc}^{-1}$ at redshift $z = 0$ (see e.g. refs. [15–20]). Recently, ref. [16] proposed a new method aimed to improve upon this reach by developing a hybrid bias model that combines the accuracy of N-body simulations with the theoretical underpinning of Lagrangian perturbation theory, called Hybrid Effective Field Theory (HEFT) hereafter. In ref. [16] it was shown that this model allows for an accurate fit to N-body data up to $k_{\text{max}} \sim 0.6 h \text{ Mpc}^{-1}$ at redshift $z = 0$. In contrast to these methods, phenomenological models of galaxy bias typically rely on the Halo Model [21–23] coupled with a Halo Occupation Distribution (HOD). These models are built on the assumption that all matter in the Universe exists in the form of halos and that galaxies populate these halos with statistics solely determined by halo mass. The advantage of these models is that they are, despite their conceptual simplicity, surprisingly successful at explaining clustering with decent accuracy well into the non-linear regime, where perturbative approaches falter. Their main disadvantage is that they rely on a number of heuristic assumptions on a qualitative level, and thus cannot strictly be shown to provide a complete description of clustering. Additionally, predictions based on the Halo Model typically show inaccuracies in the transition region between the 1- and 2-halo terms and do not account for smearing of the Baryonic Acoustic Oscillation feature (see e.g. refs. [24, 25]).

All of the methods outlined above have successfully been applied to data (see e.g. refs. [19, 26–30]), but they tend to be computationally intensive, thus making parameter inference an expensive part of cosmological analyses. Therefore, several recent works have developed hybrid methods that couple these bias models with machine learning methods to build emulators that can generate fast predictions for statistics involving biased tracers (see e.g. refs. [31–37]).

Combined in a so-called “3×2pt” analysis, galaxy clustering and weak lensing form a key component of the cosmological analysis planned by upcoming photometric galaxy redshift surveys such as Euclid and LSST. A large amount of cosmological and astrophysical information will be contained in galaxy clustering at small spatial scales. In order to ensure robust and unbiased constraints from these data, it is crucial to assess the performance of different bias models in this high signal-to-noise regime.

In this work, we aim to perform a consistent comparison of non-linear galaxy bias models and assess their performance in 3×2pt analyses including high-precision galaxy clustering data from Stage IV surveys, using as an example the LSST survey. To this end, we use the ABACUSSUMMIT simulations and generate simulated data vectors and corresponding covariances, loosely matching LSST Year 10 (Y10) data [38]. We then analyze these data using a number of current non-linear galaxy bias models. Specifically, we employ Eulerian (or Standard) Perturbation Theory, Lagrangian Perturbation Theory, and two implementations of HEFT, **anzu** and **BACCO**. We fit all of these models to the simulated data, and assess the performance of each model based on the accuracy of the returned cosmological parameter

constraints and the goodness-of-fit. Our results have implications for photometric surveys beyond LSST, but do not directly apply to spectroscopic surveys such as DESI, as we do not model a number of effects important in this regime, such as redshift space distortions (RSDs).

There are three over-arching questions that we would like to answer in this work. Which model and approach offers the most robust and accurate constraints on fundamental cosmological parameters given the high-precision of forthcoming photometric Stage IV surveys? How deep into the non-linear regime can we go using the best-performing method? How much do constraints on cosmological parameters improve as we push to increasingly smaller scales?

The last question is particularly interesting, because it will guide us in further theoretical developments. While linear scales retain the most amount of memory regarding the primordial fields and their subsequent evolution, we expect a relative loss of cosmological sensitivity on non-linear scales.¹⁰ Despite this fact, the number of observable modes increases significantly at smaller scales, and it is thus interesting to investigate the impact of these competing effects. Previous galaxy clustering analyses using HODs have found significant improvements in cosmological constraining power when increasing the minimal scale included in the analysis (see e.g. refs. [40–42]), while analyses based on PT have found smaller effects, tied to degeneracies between cosmological and bias parameters (see e.g. ref. [43]). Here, we aim to investigate these questions also in the light of upcoming Stage IV photometric surveys.

This manuscript is structured as follows. In section 2, we introduce perturbative bias models, and in section 3 we describe the observables considered. Section 4 gives an overview of the simulations employed, while section 5 describes the methodology used in our analysis. We present our results in section 6 and conclude in section 7. Implementation details are deferred to the appendices.

2 Perturbative bias models

The basic premise behind perturbative approaches to describe galaxy biasing is acknowledging the presence of complex physical, non-gravitational processes behind the formation of galaxies. These processes are non-local, and in general involve all the matter in a region around each galaxy of size R_g (e.g. the Lagrangian size of the parent halo¹¹). On scales larger than R_g however, galaxy formation can be described as an effectively local process, thus removing the need to describe these physical processes in detail. In this limit, one can invoke the Equivalence Principle which, in its non-relativistic limit, implies that the only measurable gravitational local quantities in a freely-falling frame are the second derivatives of the gravitational potential $\partial_i\partial_j\Phi$ (see e.g. ref. [8]). In other words, on these scales the overdensity of galaxies $\delta_g(\mathbf{x})$ can be described by a general function F of these second derivatives:

$$1 + \delta_g = F[\partial_i\partial_j\Phi]. \quad (2.1)$$

Perturbative bias models then proceed by expanding F in powers of $\partial_i\partial_j\Phi$. Since $1 + \delta_g$ is a scalar quantity, each order in this expansion can only involve scalar combinations of $\partial_i\partial_j\Phi$,

¹⁰Ref. [39] for example, show that the halo- and matter fields start to decorrelate once one-loop corrections to the power spectrum become significant.

¹¹The Lagrangian radius depends on halo mass; as an example for $M_h = 10^{13}M_\odot$, it is roughly given by $R_g \simeq 4$ Mpc.

which further limits the number of possible unique terms at each order in the expansion. Up to second order in powers of $\partial^2\Phi$, only three terms are allowed: the matter overdensity δ_m (proportional to the trace $\nabla^2\Phi$), the squared overdensity δ_m^2 , and the trace squared of the tidal tensor $s^2 \equiv s_{ij}s^{ij}$, where $s_{ij} \equiv \partial_i\partial_j\Phi - \delta_{ij}\nabla^2\Phi/3$.

The approximation of local bias is expected to break down on scales close to or smaller than R_g [9], and the leading correction to eq. (2.1) due to non-local processes is given by the Laplacian of the matter density field, i.e. $R_g^2\nabla^2\delta_m$ [8, 9]. In Fourier space, we have $\nabla^2\delta_m(\mathbf{k}) = -k^2\delta_m(\mathbf{k})$, which is what we include in our model given recent detections of non-local bias for halos (see e.g. [44]). We note that a similar expansion of eq. (2.1) at higher orders in $\partial^2\Phi$ and higher-order derivative operators can be derived when including non-local terms, but we will limit our discussion to the lowest-order contribution described here.

Finally, the details of galaxy formation are sensitive to fluctuations in the initial conditions on scales smaller than R_g . This leads to stochasticity in the galaxy bias relation which, at lowest order, can be captured by an additional stochastic field ε that is uncorrelated on large scales (i.e. it is assumed to have a white power spectrum on $k \ll R_g^{-1}$) [45–47], and does not correlate with any of the perturbative terms described above.

Under the above assumptions, in this work, we describe the galaxy overdensity δ_g perturbatively up to second order as

$$1 + \delta_g = 1 + b_1\delta_m + \frac{b_2}{2!}(\delta_m^2 - \langle\delta_m^2\rangle) + \frac{b_{s^2}}{2!}(s^2 - \langle s^2\rangle) + \frac{b_{\nabla^2}}{2!}\nabla^2\delta_m + \varepsilon, \quad (2.2)$$

where we have employed the Eulerian bias picture (though one could analogously expand the galaxy field in the Lagrangian picture). In eq. (2.2), we have removed the variance of the quadratic fields to ensure a mean zero galaxy overdensity. Furthermore, the quantity b_1 denotes the linear bias, b_2 is the quadratic bias, b_{s^2} is the tidal bias, b_{∇^2} denotes the non-local bias, and finally ε is the stochastic contribution. Different perturbative bias models perform this expansion at different points in time, and can thus be subdivided into Eulerian and Lagrangian approaches. In Eulerian Perturbation Theory (EPT), the perturbative expansion is performed locally at the time corresponding to the galaxy redshift. In Lagrangian Perturbation Theory (LPT) the bias parameters are defined with respect to the initial density field, and galaxy positions are then traced forward in time following their expected trajectories under gravity. If complete to a given order, Eulerian and Lagrangian bias expansions are equivalent [48–52]. In the following, we briefly discuss the traditional implementations of Eulerian and Lagrangian perturbation theory as well as an extension of LPT, named Hybrid Effective Field Theory, which aims to track the galaxy Lagrangian trajectories non-perturbatively. For a more detailed description of galaxy bias, the reader is referred to ref. [9].

2.1 Eulerian perturbation theory

In Eulerian perturbation theory, the equations of structure formation are solved by focusing on a particular point in space and following the fluid’s movement through this point in time [7]. Keeping all terms up to second order, and allowing for non-local Eulerian bias (in both space and time), the galaxy field at any given redshift z can thus be expressed by eq. (2.2), with all quantities (δ_g , δ_m , s^2) evaluated at the current galaxy position and time [8, 53].

The galaxy-galaxy and galaxy-matter power spectra in the Eulerian bias framework are thus given by

$$P_{gg}(k, z) = \sum_{i,j} b_i b_j P_{ij}(k, z) + P_{\text{SN}}, \quad P_{gm}(k, z) = \sum_i b_i P_{i\delta_m}(k, z), \quad (2.3)$$

where $i, j \in \{\delta_m, \delta_m^2, s^2, \nabla^2 \delta_m\}$, and the set $\mathbf{b} = \{b_1, b_2, b_{s^2}, b_{\nabla^2}\}$ denotes the corresponding bias parameters. As a technical subtlety, we note that in purely perturbative approaches, those bias parameters are the renormalized version of the “bare” bias parameters in Equation (2.2). In full generality, the power spectrum of first and third order bias terms (proportional to $b_1 b_{3\text{NL}}$) gives rise to terms of the same order as those generated by the auto-correlation of second order fields [52]. These power spectra however, are strongly degenerate with other terms and can thus be absorbed into lower-order bias coefficients. In this work, we therefore do not consider bias terms beyond second order but note that these are crucial for fitting e.g. higher-order statistics. P_{SN} is the power spectrum of the stochastic term ε in eq. (2.2), which is assumed to be scale-independent on the scales considered in this analysis. At the lowest order, this term can be thought of as the Poisson noise associated with the discrete nature of galaxy tracers, but it also incorporates a variety of other effects. Physically, these phenomena are described as halo exclusion, but in perturbation theory, they naturally arise as a renormalization of the terms that result in white power spectra on large scales [54, 55]. As pure Poisson noise gives rise to a stochastic power spectrum $P_{\text{SN}} = \bar{n}_g^{-1}$, we expect this term to be of the same order, but not exactly equal.

The power spectra between the different terms in the bias expansion (P_{ij} in eq. (2.3)) can be computed using Eulerian perturbation theory (although see section 2.5), and to do so, we use the FAST PT¹² package [56, 57].

As described in more detail below, in this work we fit the set of bias parameters $\mathbf{b} = \{b_1, b_2, b_{s^2}, b_{\nabla^2}\}$ and the shot noise parameter P_{SN} to simulated data from ABACUS-SUMMIT.

2.2 Lagrangian perturbation theory

In the Lagrangian bias picture [10, 12, 58, 59], the perturbative expansion of eq. (2.2) is applied to the proto-galaxy field from which galaxies form in the initial conditions (i.e. at high redshifts during matter domination). In this case, $\partial_i \partial_j \Phi$ is the Hessian of the linear gravitational potential, δ_m is the linear Lagrangian density etc., and all fields are evaluated at the initial Lagrangian coordinates $\mathbf{q}$, and denoted by the subscript L (see e.g. ref. [12]).

Once the initial proto-galaxy overdensity field is established, its evolution is determined by the Lagrangian trajectories of the galaxies under gravity. Thus at late times, when galaxies are actually observed, the galaxy overdensity δ_g at Eulerian coordinates $\mathbf{x}$ is

$$1 + \delta_g(\mathbf{x}, z) = \int d^3 \mathbf{q} \, \delta^3(\mathbf{x} - \mathbf{q} - \mathbf{\Psi}(\mathbf{q}, z)) (1 + \delta_g^L(\mathbf{q})), \quad (2.4)$$

where δ^3 is the 3-dimensional Dirac delta function, δ_L^g is the Lagrangian-space galaxy overdensity, given by eq. (2.2) in terms of the different bias expansion operators in the initial

¹²The code can be found at <https://github.com/JoeMcEwen/FAST-PT>.

conditions, and Ψ is the Lagrangian displacement field. In the usual parlance of Lagrangian perturbation theory, the final galaxy overdensity is found by “advecting” the Lagrangian bias overdensity to the final Eulerian coordinates $\mathbf{x}$.

As in the Eulerian bias expansion, the galaxy-galaxy and galaxy-matter power spectra are given by eq. (2.3), with the exception that in this case the indices i, j run over an extended set of operators $\{1, \delta_L, \delta_L^2, s_L^2, \nabla^2 \delta_L\}$, with corresponding bias parameters $\mathbf{b} = \{b_0, b_1^L, b_2^L, b_{s^2}^L, b_{\nabla^2}^L\}$. Here, $b_0 \equiv 1$ is not a free parameter, and P_{11} denotes the non-linear matter power spectrum. Finally, P_{SN} denotes a stochastic term as discussed in the previous section. The reason for the additional term is the fact that, in the Lagrangian picture, the advection of completely homogeneous density field simply yields the inhomogeneous matter density in Eulerian space. This does not change the number of free parameters of the model (since b_0 is fixed), but it implies that the Lagrangian and Eulerian bias parameters are not identical to one another (hence the L superscripts above).

To calculate the displacement field Ψ and thus the power spectra of the advected operators (i.e. the $P_{ij}(k)$ in eq. (2.3)) we can use Lagrangian perturbation theory. The details of this calculation can be found in e.g. ref. [60]. In this work, we compute these LPT power spectra using **velocileptors**,¹³ which is described in ref. [61].

2.3 HEFT

The physical processes underlying galaxy formation are complex, and thus formulating a non-perturbative bias model based on first principles is commensurately difficult. However, non-linear evolution under gravity is a simpler problem that can be solved to high accuracy numerically via N -body simulations. This fact may be used to formulate a hybrid bias expansion, where the relation between galaxy overdensity and $\partial_i \partial_j \Phi$ is given perturbatively at early times, and the subsequent evolution under gravity (i.e. solving for the Lagrangian displacement Ψ) is carried out numerically via simulations. First proposed in ref. [16] and subsequently explored by refs. [31, 62, 63], this approach allows for higher accuracy predictions while keeping the physical intuition of LPT.

Building on ref. [16], two separate works [31, 32] employed suites of N -body simulations to build emulators for the different power spectra in eq. (2.3), corresponding to the advected fields $\{1, \delta_L, \delta_L^2, s_L^2, \nabla^2 \delta_L\}$. These emulators thus allow for the computation of HEFT predictions for a wide range of cosmological models. We describe these briefly below, but refer the reader to refs. [31, 32] for further details.

- **anzu**: in ref. [31], the authors use the **Aemulus** suite of N -body simulations to compute predictions for the LPT basis spectra. This simulation suite has been designed for emulating cosmological quantities, and covers the parameter space $\vartheta = \{\Omega_b h^2, \Omega_c h^2, \sigma_8, H_0, n_s, N_{\text{eff}}, w\}$ within priors set by a combination of current observational constraints. Using the base power spectra, $P_{ij}(k, z)$, computed from **Aemulus** for a broad range of cosmological models, **anzu** employs polynomial chaos expansions [64] to emulate these quantities, and is publicly available on **github**¹⁴ (see ref. [35] for an updated version of this emulator).

¹³The code can be found at <https://github.com/sfschen/velocileptors>.

¹⁴The code can be found at <https://github.com/kokron/anzu>.

- **BACCO**: the analysis presented in ref. [32] uses a similar approach: the authors use the BACCO suite of numerical simulations coupled with cosmology rescaling [65] to create a library of LPT base power spectra that cover the parameter space $\vartheta = \{\Omega_m, \Omega_b, \sigma_8, n_s, h, M_\nu, w_0, w_a\}$. Using these power spectra, the authors construct an emulator using a simple Neural Network.¹⁵

Various works [16, 31, 32, 63] have found that this hybrid approach is able to reproduce the galaxy-galaxy and galaxy-matter power spectra in real space down to significantly smaller scales than the purely perturbative approaches. In general, precision of a few per cent can be achieved up to $k_{\text{max}} \sim 0.6 \, h\text{Mpc}^{-1}$ for redshifts $0 \lesssim z \lesssim 1$.

2.4 Relations between bias models

As galaxies found in the evolved, Eulerian field can be traced back to proto-galaxies in Lagrangian space at early times, we can relate Lagrangian bias parameters to their Eulerian counterparts [9]. Assuming coevolution of the galaxy and the matter distribution, which is equivalent to galaxy number conservation and vanishing velocity bias, allows us to derive simple relations between bias parameters defined at early times to those defined at all later times. Physically, this is due to the fact that the density ratio of conserved, comoving fluids is unchanged under gravity owing to the equivalence principle. Expanding the galaxy overdensity in Lagrangian space up to second order neglecting non-local terms, and using the continuity equation, leads to a relation between Lagrangian and Eulerian bias parameters given by [9]

$$\begin{aligned} b_1 &= 1 + b_1^L, \\ b_2 &= \frac{8}{21} b_1^L + b_2^L, \\ b_{s^2} &= -\frac{4}{7} b_1^L + b_{s^2}^L. \end{aligned} \tag{2.5}$$

Note that we have adjusted the prefactors to match the bias definition in eq. (2.2). A popular toy model for galaxy bias is the so-called local-in-matter-density (LIMD) Lagrangian bias. In this model, we assume the galaxy overdensity at early times to be solely a function of the local matter density, which amounts to setting $b_{s^2}^L = b_{\nabla^2}^L = 0$ in eq. (2.2). In this special case eq. (2.5) implies a relation between Eulerian bias parameters given by $b_{s^2} = -\frac{4}{7}(b_1 - 1)$ [52]. This shows that gravitational evolution leads to a bias with respect to the squared tidal field at late times, even in absence of such a bias at early times, i.e. LIMD at early times is inconsistent with LIMD at late times. In general, we do not expect these relations to hold exactly, as galaxy evolution is a complex process determined by forces other than gravity such as momentum transfer due to baryonic feedback and radiation pressure [9]. In addition, several works have investigated empirical relations between bias parameters, beyond coevolution. Refs. [66, 67] for example have found strong correlations between bias parameters for halos, and refs. [68, 69] have found similar results for galaxies, albeit with a larger intrinsic scatter and slightly different relations between bias parameters. In section 6.3 we will test the validity of a subset of these relations for the galaxy samples considered in this work.

¹⁵The code can be found at <https://bacco.dipc.org/emulator.html>.

Finally, we note that gravitational co-evolution will generate third-order bias terms in Eulerian space from a second-order bias expansion in Lagrangian space. The models described in sections 2.1 and 2.2 are therefore not fully equivalent. However, in this work we have chosen to sacrifice full theoretical consistency for consistency in the number of bias parameters used for each model. We will further discuss this choice and how it might affect our results in section 6.3.

2.5 Implementation choices

When quantifying the validity of the different bias expansions introduced above, we will make use of a few implementation-specific choices.

Bias evolution. The perturbative bias expansions explored here, in their different incarnations, determine the scale dependence of the different terms contributing to the final galaxy power spectra, but do not impose any restrictions on the redshift dependence of the bias parameters. As described in more detail below, our analysis is based on angular cross-correlations between galaxies in different tomographic redshift bins. Since each tomographic bin in principle corresponds to a distinct galaxy sample, we assume different and independent bias parameters in each bin. We consider relatively broad redshift bins, and since we expect galaxy properties and thus galaxy bias to evolve with redshift, our model must account for this. In our fiducial implementation of all perturbation theory models, we therefore allow for a redshift evolution in the lowest-order (linear) bias parameters, b_1 and b_1^L . We assume a linear bias evolution with redshift of the form (e.g. for Eulerian bias):

$$b_1(z) = b_1 + b_{1,p}(z - \bar{z}), \quad (2.6)$$

where $\bar{z}$ denotes the mean redshift of each bin, and we use a similar expression for b_1^L .¹⁶ By default all other bias parameters are assumed constant within each redshift bin, although we will also investigate the impact of allowing for a redshift-evolution of b_2 and b_2^L in section 6.

Decorated PT. As stated in sections 2.1 and 2.2, in the EPT and LPT frameworks, the power spectra of the bias expansion terms (P_{ij} in eq. (2.3)), may be computed to a given order in Eulerian or Lagrangian perturbation theory. This perturbative approach can be potentially improved by replacing the $P_{\delta_m \delta_m}(k, z)$ terms by its non-perturbative prediction calculated e.g. via the HALOFIT fitting function [4, 70], effectively re-summing all PT terms contributing to it. This selective resummation approach has been found to improve the quality of the fit while remaining largely unbiased [19, 71, 72]. When studying the EPT and LPT frameworks, we therefore consider two different approaches:

PTPk approach: We use Eulerian or Lagrangian perturbation theory at next-to-leading order to compute all terms in eq. (2.3). The power spectrum of the non-local term and the matter density is approximated as

$$P_{\delta_m, \nabla^2 \delta_m}(k, z) \rightarrow -k^2 P_{1\text{-loop}}(k, z), \quad (2.7)$$

¹⁶We consider a linear relation between bias and redshift as this corresponds to the lowest-order Taylor expansion of a possibly more complex relation, but can be generalized to higher orders.

where $P_{1\text{-loop}}$ is the 1-loop matter power spectrum and we have used that in Fourier space $\nabla^2 \delta_m(\mathbf{k}) = -k^2 \delta_m(\mathbf{k})$.¹⁷ All other terms in the expansion of P_{gm} and P_{gg} will be calculated using EPT or LPT at next-to-leading order.¹⁸

NLPk approach: In a second approach, we use the full non-linear matter power spectrum $P_{\text{NL}}(k, z)$ from HALOFIT to compute the term multiplying b_1 in P_{gm} and b_1^2 in P_{gg} in the EPT case. For LPT, we make the following substitutions in the case of P_{gm} and P_{gg} respectively:

$$P_{11}(k, z) + b_1^L P_{1\delta_m}(k, z) \rightarrow (1 + b_1^L) P_{\text{NL}}(k, z) \quad (2.8)$$

$$P_{11}(k, z) + 2b_1^L P_{1\delta_m}(k, z) + (b_1^L)^2 P_{\delta_m \delta_m}(k, z) \rightarrow (1 + b_1^L)^2 P_{\text{NL}}(k, z). \quad (2.9)$$

We calculate all other terms in the expansion using EPT or LPT at next-to-leading order. As above, the cross-power spectrum of the non-local term and the matter density is approximated as

$$P_{\delta_m, \nabla^2 \delta_m}(k, z) \rightarrow -k^2 P_{\text{NL}}(k, z). \quad (2.10)$$

As opposed to the other bias models considered, where we keep all auto- and cross-correlations involving $\nabla^2 \delta_m$, in this case we only keep the power spectra multiplying the matter density in order to not mix HALOFIT and PT predictions for the nonlocal bias terms. In the EPT case, we thus only keep the term involving b_1 and $\nabla^2 \delta_m$, while for LPT we make the substitution $P_{1, \nabla^2 \delta_m}(k, z) + b_1^L P_{\delta_m, \nabla^2 \delta_m}(k, z) \rightarrow -(1 + b_1^L) k^2 P_{\text{NL}}(k, z)$.

3 Observables

Joint analyses of galaxy clustering, galaxy-galaxy lensing and cosmic shear in a so-called 3×2 pt analysis will provide key cosmological constraints for LSST and other future photometric surveys. In this work, we assess the performance of different galaxy bias models in a Fourier-space-based 3×2 pt analysis, closely matching the specifications for LSST Y10. In full generality, the spherical harmonic power spectrum between probes a and b , and redshift bins i, j can be computed using the Limber approximation¹⁹ [74–76] as:

$$C_\ell^{a_i b_j} = \int dz \frac{c}{H(z)} \frac{q^{a_i}(\chi(z)) q^{b_j}(\chi(z))}{\chi^2(z)} P_{ab}\left(z, k = \frac{\ell + 1/2}{\chi(z)}\right), \quad (3.1)$$

where c is the speed of light, $H(z)$ denotes the Hubble parameter, and $\chi(z)$ is the comoving distance. In addition, $q^{a_i}(\chi(z))$ denotes the window function for probe a and bin i , and $P_{ab}(z, k)$ is the three-dimensional power spectrum between probes a and b . For galaxy tracers, the window function is given by

$$q^{\delta_{g,i}}(\chi(z)) = \frac{H(z)}{c} p^i(z), \quad (3.2)$$

¹⁷We note that the functional form of this expression is equivalent to the counterterms present in EFT approaches (see e.g. refs. [13, 14]). We will further discuss this similarity and its impact on our results in appendix C.

¹⁸In particular, the matter power spectrum is modeled as $P_{mm} = P_{1\text{-loop}}$.

¹⁹We note that the Limber approximation has been shown to lead to biased results at low multipoles (see e.g. ref. [73]), but we do not expect this to affect our conclusions as these scales are still well within the linear regime and we choose a conservative minimal multipole of $\ell_{\text{min}} = 32$.

where $p^i(z)$ denotes the normalized redshift distribution of the considered sample. The window function for lensing tracers is

$$q^{\gamma_i}(\chi(z)) = \frac{3}{2} \frac{\Omega_m H_0^2}{c^2} \frac{\chi(z)}{a} \int_{\chi(z)}^{\chi_h} dz' p^i(z') \frac{\chi(z') - \chi(z)}{\chi(z')}, \quad (3.3)$$

where Ω_m denotes the fractional matter density today, H_0 is the current expansion rate, a denotes the scale factor, and χ_h is the comoving distance to the horizon. Furthermore, we have $P_{\gamma\gamma}(z, k) = P_{\delta_m \delta_m}(z, k)$.

Additionally, we also consider a combination of the 3×2 pt data vector with CMB lensing, and the window function in this case is given by:

$$q^{\kappa_{\text{CMB}}}(\chi(z)) = \frac{3}{2} \frac{\Omega_m H_0^2}{c^2} \frac{\chi(z)}{a} \frac{\chi(z_*) - \chi(z)}{\chi(z_*)}, \quad (3.4)$$

where z_* denotes the redshift to the surface of last scattering.

In this work, we model the uncertainties associated to this data vector analytically. The 3×2 pt covariance matrix generally consists of three different parts: a Gaussian contribution, a non-Gaussian contribution and a contribution due to super-sample covariance (SSC) (see e.g. ref. [40]). Here we make the simplifying assumption that the data covariance is Gaussian, and thus neglect non-Gaussian and SSC contributions. Including those contributions will have two main effects on the covariance: (i) increasing the size of the error bars, and (ii) correlating different ℓ -modes. While we expect the first effect to lead to our analysis being conservative since our smaller error bars lead to tighter requirements, the second might affect goodness-of-fit tests in a nonlinear way. We defer an investigation including non-Gaussian covariances to future work. Under the assumption of Gaussianity, we compute the covariance between the spherical harmonic power spectra C_ℓ^{ij} and $C_{\ell'}^{i'j'}$ using

$$\text{Cov}_G(C_\ell^{ij}, C_{\ell'}^{i'j'}) = \langle \Delta C_\ell^{ij} \Delta C_{\ell'}^{i'j'} \rangle = \frac{\delta_{\ell\ell'}}{(2\ell+1)\Delta\ell f_{\text{sky}}} \left[(C_\ell^{ii'} + N_\ell^{ii'})(C_\ell^{jj'} + N_\ell^{jj'}) + (C_\ell^{ij'} + N_\ell^{ij'})(C_\ell^{i'j} + N_\ell^{i'j}) \right], \quad (3.5)$$

where C_ℓ^{ij} denotes the signal part of the spherical harmonic power spectrum, N_ℓ^{ij} is the corresponding noise, f_{sky} denotes the fraction of sky covered by the experiment, and $\Delta\ell$ is the width of the ℓ -bin used for power spectrum estimation. Usually, we set $N_\ell^{ij} = \delta_{ij} N_\ell^{ii}$, but here we keep a more generic expression in order to cater for potentially scale-dependent noise correlated across different probes.

In addition, in refs. [31, 32], it was shown that the three-dimensional power spectra obtained from the HEFT emulators exhibit relative errors of around 1% when compared to the power spectra measured directly from the N -body simulations. Both works account for this systematic uncertainty through an additional term in their covariance matrix. We follow these analyses assuming full correlation of theoretical errors of any two power spectra, and add a systematic error floor to our covariance matrix given by

$$\text{Cov}_{\text{sys}}(C_\ell^{ij}, C_{\ell'}^{i'j'}) = f_{\text{sys}}^2 C_\ell^{ij} C_{\ell'}^{i'j'} \delta_{\ell\ell'}, \quad (3.6)$$

where f_{sys} is the fractional error of the theoretical model, which we set to $f_{\text{sys}} = 0.01$ as per ref. [31]. The total covariance matrix is thus given by

$$\text{Cov}(C_{\ell}^{ij}, C_{\ell'}^{i'j'}) = \text{Cov}_G(C_{\ell}^{ij}, C_{\ell'}^{i'j'}) + \text{Cov}_{\text{sys}}(C_{\ell}^{ij}, C_{\ell'}^{i'j'}). \quad (3.7)$$

The systematic error floor becomes significant at the smallest scales considered in our analysis where the combined error can amount to almost five times the contribution from statistical uncertainties alone. However, for the majority of scales considered, the ratio between statistical and full errors is above 0.65.

4 Simulations

We use the ABACUSSUMMIT suite of high-performance cosmological N -body simulations [77] to create the simulated galaxy samples used in this analysis. The ABACUSSUMMIT suite was designed to meet the simulation requirements of the Dark Energy Spectroscopic Instrument (DESI) survey and was run with the high-accuracy cosmological code ABACUS [78, 79]. We utilize one of the “base”-resolution ABACUSSUMMIT boxes, `AbacusSummit_base_c000_ph006`, at the fiducial cosmology: $\Omega_b h^2 = 0.02237$, $\Omega_c h^2 = 0.12$, $h = 0.6736$, $10^9 A_s = 2.0830$, $n_s = 0.9649$. Its box size is 2000 Mpc/ h , and it contains 6912^3 particles with mass $M_{\text{part}} = 2.1 \times 10^9 M_{\odot}/h$.

To construct the mock galaxy samples from the N -body outputs, we utilize the 10% particle subsample (i.e., including both the particle A and B subsamples, which constitute 3% and 7% of the particles, respectively), which is selected randomly and is consistent across redshift, and the on-the-fly halo catalogues, which are generated using the halo-finding algorithm COMPASO [80]. Specifically, we apply the ABACUSHOD model [81] to the halo catalogue outputs at $z = 0.1, 0.3, 0.5, 0.8, 1.1, 1.4, 1.7, 2.0, 2.5, 3.0$. The ABACUSHOD model builds upon the baseline halo occupation distribution (HOD) model by incorporating various generalizations pertaining to halo-scale physics and assembly bias.

For the fiducial galaxy samples considered in this analysis, we assume that they are well-approximated by the “baseline HOD” model with no decorations. The model is akin to the 5-parameter model of [82], which gives the mean expected number of central and satellite galaxies per halo given halo mass M :

$$\bar{N}_{\text{cent}}(M) = \frac{1}{2} \text{erfc} \left[\frac{\log_{10}(M_{\text{min}}/M)}{\sqrt{2}\sigma_M} \right], \quad (4.1)$$

$$\bar{N}_{\text{sat}}(M) = \left[\frac{M - \kappa M_{\text{min}}}{M_1} \right]^{\alpha} \bar{N}_{\text{cent}}(M). \quad (4.2)$$

Here, M_{min} characterizes the minimum halo mass to host a central galaxy, M_1 is the typical halo mass that hosts one satellite galaxy, σ_M describes the steepness of the transition from 0 to 1 in the number of central galaxies, α is the power law index on the number of satellite galaxies, and κM_{min} gives the minimum halo mass to host a satellite galaxy. Central galaxies are placed at the halo centre, whereas satellite galaxies are “painted” on random particles.

In this work, we consider two main galaxy samples, which are designed to mimic two distinct galaxy sample choices for clustering analyses: (i) a homogeneous sample of Luminous

Parameter	red	maglim
$\mu_{\min}$	12.95	11.88
$\mu_{\min,p}$	−2.0	−0.5
μ_0	12.3	11.88
$\mu_{0,p}$	0.0	−0.5
μ_1	14.0	13.08
$\mu_{1,p}$	−1.5	0.9
α	1.32	1.0
$\sigma_{\log M}$	0.25	0.4

Table 1. Summary of HOD parameters for the two galaxy samples considered in this work. The parametrization of the redshift-dependent HOD follows ref. [84].

Red Galaxies (LRGs) with a moderate number density (called ‘red’ hereafter), which will constitute our fiducial sample throughout this work, and (ii) a magnitude-limited, high-number density sample (called ‘maglim’ hereafter).²⁰ We model the HODs of these samples based on observational constraints using Equations (4.1) and (4.2). In order to account for evolution of both samples, we assume a redshift dependence of the three HOD masses ($M_{\min}$, M_1 , and $M_0 \equiv \kappa M$) of the form [84]

$$\log_{10} M_x/M_{\odot} = \mu_x + \mu_{x,p} \left[\frac{1}{1+z} - \frac{1}{1+z_p} \right], \quad (4.3)$$

with $z_p = 0.65$. Using this parameterization, we model the HOD of the maglim sample using the results derived for the HSC sample studied in ref. [84].²¹ For the red sample on the other hand, we assume redshift-dependent HOD masses consistent with the DESI LRG selection described in ref. [30].²² The adopted HOD parameters for both samples are given in table 1.

We complement these samples with an additional data set for which we also model assembly bias effects. A number of studies have investigated the dependence of clustering statistics on quantities other than halo mass, such as halo concentration, environment or spin (see e.g. refs. [87–90]), and some of these works have resulted in significant detections of this so-called assembly bias effect. The nonlinear galaxy bias models considered in this work in principle offer the flexibility to also describe tracers affected by these processes. In order to test this, we generate an additional galaxy sample from ABACUSSUMMIT via the fast and efficient decorated HOD model, AbacusHOD [91], incorporating assembly bias effects due to halo concentration and environment as described and defined in refs. [81, 92]. In

²⁰Similar samples have been considered in the DES Y3 analyses (see e.g. ref. [83]), and we expect such samples to also be used for clustering analyses of LSST data.

²¹Specifically, for $M_{\min}$ and M_1 we use the posterior means derived for the fiducial case in ref. [84] as given in their table 4. As M_0 is unconstrained in this analysis, we follow refs. [85, 86], and set $M_0 = M_{\min}$. Finally, we set the remaining parameters to the fiducial values assumed in ref. [84].

²²In practice, we fit the results given in their table 4 as a function of redshift to derive values for the HOD parameters.

particular, we adopt non-zero values for the concentration and environment assembly bias parameters, $A_{c,s}$ and $B_{c,s}$, for both the central and satellite populations (subscript c referring to centrals and s to satellites), given by $A_c = -0.73$, $A_s = -0.24$, $B_c = -0.0093$, and $B_s = 0.0037$. These parameters modify the central and satellite halo occupations by reranking the haloes at fixed mass based on their intrinsic concentration and environment, and their values are to be close to the best-fit for CMASS data [93] and modify the central and satellite halo occupations (hence, the subscripts) by re-ranking the haloes at fixed mass by their concentration (A_c and A_s) and their environment (B_c and B_s). For full definitions of the parameters and how they are used, see ref. [91].

We would like to stress that all galaxy samples considered in this work are generated using a simplified HOD model, which includes some extensions beyond the “vanilla” model described in ref. [94]. However, for maximal realism, one could imagine incorporating other effects such as redshift dependence of the selection functions, deviations of the satellite occupations from a Poisson distribution, and a further dependence of the central occupations on assembly bias. These additional effects might significantly affect the clustering and stochasticity of the samples, and thus potentially also our results. We leave an investigation of more complex HOD models to future work.

In addition in this work, we do not consider the additional uncertainties in realistic samples due to galaxy selection effects and photo- z estimation. These uncertainties are specific to photometric surveys; recent analyses from Stage III surveys, e.g. ref. [71], have found that imperfect correction of “survey properties” appear to dominate statistical uncertainties for red samples in DES. Given LSST’s statistical precision, these issues may require significant effort. In addition, the sample selection and redshift binning interplays with bias evolution, magnification and other effects not considered here. Finally, we note that the Euclid survey has planned an ambitious tomographic analysis comprising about a dozen redshift bins. The opportunities and challenges posed by such an analysis require separate investigation.

5 Methods

5.1 Generating smooth spectra

We use the simulated galaxy samples from ABACUSSUMMIT to compute three-dimensional galaxy-galaxy auto- and galaxy-matter cross-power spectra for all the redshifts covered by the simulations. Sample variance in the power spectra extracted from ABACUSSUMMIT is significant on large scales compared to the expected error bars, which could lead to biases in our analysis. As described in detail in appendix A, we suppress this noise by combining the spectra measured from the simulations with theoretical predictions for the linear power spectrum on large scales. This procedure allows us to model the measured power spectra to an accuracy below the sample variance uncertainties of the ABACUSSUMMIT simulation box. The reader is referred to appendix A for a more detailed description of the methodology.

5.2 Computing simulated data vectors

In order to compute spherical harmonic power spectra using eq. (3.1), we need to specify redshift distributions for the source galaxies as well as the two clustering samples considered in

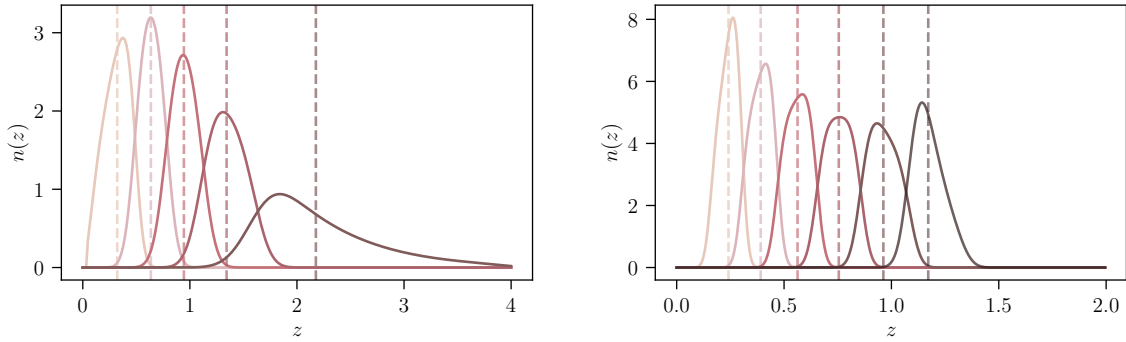


Figure 1. Assumed redshift distributions for the samples considered in this work. *Left panel:* tomographic redshift bins for both the source sample and the maglim clustering sample. *Right panel:* tomographic redshift bins for the red clustering sample. The dashed lines denote the mean redshift of each bin. We note that all redshift bins are normalized to unity.

this work. For the source sample, we assume a number density and overall redshift distribution based on the LSST Y10 sample as specified in the LSST DESC Science Requirements document [38]. The sample is divided into 5 redshift bins, enforcing an equal number of galaxies in each bin, and assuming Gaussian photometric redshift uncertainties with standard deviation $\sigma_z = 0.05(1 + z)$. The resulting redshift distributions are shown in the left panel of figure 1, with the inter-bin tails caused by the photo- z uncertainties. For the clustering on the other hand, we assume different distributions for the maglim and the red sample: for the maglim sample we make the assumption that we use the same set of galaxies for lensing and clustering measurements, and thus set the clustering redshift distributions to those assumed for the source sample. For the red sample on the other hand, we assume smaller photometric redshift uncertainties and a lower maximal redshift. The number density and overall redshift distribution is calculated from the luminosity function of red galaxies, following the procedure detailed in ref. [95]. The sample is then divided into 6 redshift bins, evenly spaced in photometric redshift, and assuming a photo- z scatter $\sigma_z = 0.02(1 + z)$. The resulting tomographic redshift bins are shown in the right panel of figure 1.

Using these redshift distributions, we obtain spherical harmonic power spectra for an LSST-like 3×2 pt analysis. We use the computed three-dimensional power spectra from ABACUSSUMMIT and interpolate those in redshift to evaluate eq. (3.1). We verified that the number of samples in redshift used in this interpolation was large enough to produce accurate angular power spectra. In a next step, we compute all possible auto- and cross-correlations between probes and tomographic bins, except that we do not include cross-correlations between clustering bins in our analysis. The latter choice is driven by the method employed to compute three-dimensional power spectra from ABACUSSUMMIT in the presence of stochasticity (see description below). In general however, we expect the considered bias models to also be applicable for galaxy clustering cross-correlations, and it would be interesting to extend our analysis to these observables given their additional information content. In practice, we perform all projections into angular power spectra using the DESC Core Cosmology Library

(CCL²³) [96]. This leaves us with a set of 51 or 45 spherical harmonic power spectra for the red or maglim clustering samples respectively, which we evaluate for 22 bandpowers with edges $\ell = \{32, 42, 52, 66, 83, 105, 132, 167, 210, 265, 333, 420, 528, 665, 838, 1054, 1328, 1672, 2104, 2650, 3336, 4200, 5287\}$. When applying a scale cut based on a maximal wavenumber k_{max} for power spectra involving clustering, we only keep bandpowers for which the mean of the ℓ -bin satisfies $\ell < k_{\text{max}}\chi(\bar{z})$, where $\chi(\bar{z})$ denotes the comoving distance to the mean redshift of the respective bin. For the cosmic shear power spectra on the other hand, we fix $\ell_{\text{max}} = 2000$ for all cases.

As the aim of this work is to assess the performance of different galaxy bias models, we need to prevent biasing our results due to inaccuracies in modeling the matter power spectrum used for cosmic shear predictions. We therefore rescale the ABACUSUMMIT matter power spectrum to match the HALOFIT [4, 70] prediction (see e.g. ref. [20] for a similar approach).²⁴ In section 6.4.4, we explore the effect of not rescaling cosmic shear power spectra. We note that all other power spectra are left unchanged and taken from the ABACUSUMMIT simulations, which means that matter power spectrum mis-modeling might affect our results for $P_{\delta_g\delta_g}$ and $P_{\delta_g\delta_m}$. As described in section 2.5, we test the sensitivity of LPT and EPT to implementation choices for modeling the matter power spectrum, but leave a more extensive analysis to future work.

We generate covariances for these data as outlined in section 3, modeling LSST Y10 data based loosely on the LSST DESC Science Requirements Document (SRD) [38]. We set the fraction of sky covered to $f_{\text{sky}} = 0.4$. We model the noise for cosmic shear following ref. [99], setting the effective number density of galaxies to $n_{\text{eff}} = 27 \text{ arcmin}^{-2}$, and the intrinsic ellipticity to $\sigma_e = 0.28$. For galaxy clustering on the other hand, we follow an alternative approach: we find the considered samples to exhibit a significant level of non-Poissonian stochasticity. Subtracting an estimate for the Poissonian shot noise of the sample from the simulated three-dimensional power spectra would therefore lead to biased results. We circumvent this issue by considering clustering noise levels consistent with the stochasticity determined by the HODs of the simulated galaxy samples. In practice this means that we do not subtract any Poisson shot noise estimate from the simulated power spectra, but account for it in the theoretical modeling as described in sections 2.1 and 2.2.²⁵

Finally we consider an extension of our LSST-like 3×2 pt analysis: we additionally model a joint analysis of our fiducial red galaxy sample from LSST with CMB lensing from CMB S4. Practically, we thus extend our fiducial 3×2 pt data vector with all possible cross-correlations with the CMB lensing convergence leading to a set of 62 spherical harmonic power spectra. To model CMB S4 lensing and its associated covariance through Equations (3.1) and (3.7),

²³<https://github.com/LSSTDESC/CCL>.

²⁴As shown in ref. [97], the matter power spectrum from ABACUSUMMIT agrees with that from HALOFIT to a level of 2% for $k < 0.5h \text{ Mpc}^{-1}$ for the fiducial ABACUSUMMIT cosmology. For other cosmological models however, the discrepancies increase; an example is shown in ref. [35], who find differences of up to 4% between the *Aemulus* ν simulations and *HMCode2020* [98] at $k = 0.2h \text{ Mpc}^{-1}$. We expect these to be comparable to the differences between ABACUS and HALOFIT for a range of cosmological models.

²⁵We note that this approach amounts to assuming that the noise levels determined by the assumed HODs are consistent with the clustering noise levels expected for LSST Y10 data. Given that we have modeled our samples according to observational results, we believe this to be an acceptable approximation.

we follow the specifications given in the CMB S4 wiki²⁶ (private communication Toshiya Namikawa and Colin Hill), and use the CMB lensing reconstruction noise computed using the `cmb lensplus`²⁷ code. We assume a common sky coverage of LSST and CMB S4 given by $f_{\text{sky}} = 0.4$.

To fully simulate the expected data vectors, one should draw a noise realization from the covariance matrix and add it to the noiseless prediction for the observables. We choose not to do this: for an ideal model, our result should thus center on the true values of cosmological parameters and the corresponding χ^2 should satisfy $\chi^2 = 0$. Therefore, any deviation in χ^2 and parameter values indicates systematic errors in the theory modeling (see also section 5.4).

5.3 Deriving best-fit parameters and associated uncertainties

For all bias models and galaxy samples analyzed in this work, we assume a Gaussian likelihood with simulated data and covariance matrix as described above. In order to compare the performance of different bias models and test if the fits to the simulated data return unbiased constraints on cosmological parameters, we would ideally sample the likelihood in an Monte Carlo Markov Chain (MCMC) and thus derive the full posterior of our model parameters. As we are considering a large number of bias models and different implementations, this is computationally expensive (although see ref. [100] for potential ways to accelerate this process). We therefore resort to a simplified approach: for all models considered, we run an optimization algorithm to maximize the likelihood and determine the corresponding values of cosmological and bias parameters. Specifically, we use the BOBYQA²⁸ algorithm [101–103] through the `cobaya`²⁹ package as well as Powell’s optimization method as implemented in `scipy` [104]. We choose the latter method as the parameter space considered in this analysis exhibits significant degeneracies, and we find that directional optimization as implemented in Powell’s method outperforms BOBYQA in most cases. We therefore use Powell’s method for most of our fiducial results but always compare to the results obtained using BOBYQA to test the robustness of our conclusions. In addition, we perform a number of further tests to ensure the stability of our results: we rerun each case multiple times to test stability against changing the initial conditions, we also vary the convergence criterion for both methods as well as the number of starting points for BOBYQA, finding consistent results for reasonable settings of these parameters.

In order to characterize biases on cosmological parameters, we additionally need to estimate parameter uncertainties. In this work, we use a Fisher matrix (FM) formalism to forecast uncertainties on cosmological and nonlinear bias parameters. The Fisher matrix allows for propagation of uncertainties on observables (in our case, the spherical harmonic power spectra) to corresponding uncertainties on model parameters. Under the assumption that the dependence of the data covariance matrix $\mathbf{C}$ on the parameters of interest θ_α can be neglected (which is a very good approximation [105]), the Fisher matrix F is given by

²⁶The description can be found at https://cmb-s4.uchicago.edu/wiki/index.php/Survey_Performance_Expectations.

²⁷The package can be found at <https://github.com/toshiyan/cmb lensplus>.

²⁸The code can be found at <https://numericalalgorithmgroup.github.io/pybobyqa/build/html/index.html>.

²⁹The code can be found at <https://cobaya.readthedocs.io/en/latest/>.

(see e.g. refs. [106–108])

$$F_{\alpha\beta} = \frac{\partial \mathbf{D}}{\partial \theta_\alpha} \mathbf{C}^{-1} \frac{\partial \mathbf{D}}{\partial \theta_\beta}, \quad (5.1)$$

where $\mathbf{D}$ denotes the data vector of a given experiment (in our case, a list containing all the power spectra used in the analysis). The Cramér-Rao bound then states that the uncertainty on θ_α , marginalized over all other θ_β satisfies

$$\Delta\theta_\alpha \geq \sqrt{(F^{-1})_{\alpha\alpha}}. \quad (5.2)$$

In this analysis, we consider a fiducial model characterized by two cosmological and a set of six nonlinear bias parameters per clustering redshift bin, i.e. $\boldsymbol{\theta} = \{\sigma_8, \Omega_c, b_1^{p,i}, b_{1p}^{p,i}, b_2^{p,i}, b_{s2}^{p,i}, b_{\nabla^2}^{p,i}, P_{\text{SN}}^{p,i}\}$, where i denotes the redshift bin (i.e. $i = 1 \dots 5/6$ in our analysis), and $p = L$ in the case of LPT and HEFT. Computing the Fisher matrix requires the assumption of fiducial values for the model parameters, which we set to the best-fit values derived from the minimizer. We compute derivatives of the observables with respect to the model parameters numerically using a five-point stencil with step size $\epsilon = 0.01\theta$, where θ denotes any parameter considered in our analysis.³⁰ We test the stability of our results by varying the parameter ϵ and find our results to be largely insensitive to this choice.

Fisher matrix analyses are prone to numerical instabilities (see e.g. refs. [109–111]). An additional complication in our case is that the posterior is likely to exhibit non-Gaussian features due to degeneracies between model parameters. We therefore test the robustness of our results by comparing our FM constraints to those derived using an MCMC for selected cases, finding reasonable agreement as described in detail in appendix D.

5.4 Assessing model performance

For all the bias models considered in this analysis, we assess model performance as a function of the smallest scale (largest wavenumber) k_{max} considered for galaxy clustering and galaxy-galaxy lensing data. As we are working with angular power spectra, we convert this quantity to a maximal angular multipole for each redshift bin using the approximate relation $\ell_{\text{max}} = k_{\text{max}}\chi(\bar{z})$, where $\bar{z}$ denotes the mean redshift of each bin.

In this work, we assess model performance in several different ways. First, we test if the values of the cosmological parameters recovered from our fits are consistent with their fiducial values within statistical uncertainties as derived from the Fisher matrix analysis. This test does not guarantee good model performance, as even models that do not fit the data can potentially yield unbiased parameter constraints. We therefore additionally assess goodness-of-fit (GOF) of all models considered. As described in section 5.2, we work with noiseless data vectors. We therefore do not expect the χ^2 of the fit to follow a χ^2 -distribution with mean given by the number of degrees of freedom in the data, but rather be significantly smaller and dominated by the model performance and numerical errors. Based on this, we devise a set of two χ^2 -tests which we use to validate our results:

³⁰For parameters with fiducial value of zero, we set $\epsilon = 0.01$.

Noiseless GOF: First, we determine the minimal χ^2 -value of the fit to the noiseless data vector, which we will call χ_{theory}^2 . We need to make sure that this quantity is significantly smaller than the $\Delta\chi^2$ allowed by statistical uncertainties and the number of degrees-of-freedom of our model. In the presence of noise, we expect the minimal χ^2 of the fit to roughly follow a χ^2 -distribution with number of degrees-of-freedom $dof = n_{\text{data}} - n_{\text{param}}$, where n_{data} denotes the number of data points, and n_{param} is the number of parameters in the model.³¹ We call this latter quantity χ_{noise}^2 , which in our case is given by $\chi_{\text{noise}}^2 = n_{C_\ell} - n_{\text{param}} - n_{C_\ell^{\gamma\gamma}}$, where n_{C_ℓ} denotes the total number of spherical harmonics C_ℓ , and $n_{C_\ell^{\gamma\gamma}}$ denotes the number of shear-only data points. As described above, we set the matter power spectrum of the simulations to its HALOFIT prediction, and we therefore do not count the shear power spectra as degrees-of-freedom of the model.

In our first test we thus assess that the sum of the expected χ_{noise}^2 and χ_{theory}^2 is reasonably likely given a χ^2 distribution with mean χ_{noise}^2 . This test ensures that the differences due to systematic uncertainties in the theory are subdominant to statistical uncertainties. For the noiseless data vector, we also examine the size of the fit residuals.

Noisy GOF: Secondly, we use our synthetic data vector alongside the covariance matrix to create a noisy realization of the data. We then fit this data using all the bias models discussed and test that the best-fit yields a χ^2 consistent with $\chi_{\text{noise}}^2 + \chi_{\text{theory}}^2$. In contrast to earlier, we set $\chi_{\text{noise}}^2 = n_{C_\ell} - n_{\text{param}}$, as we do sample the cosmic shear data as well. We additionally assert that the histogram of the fit residuals is consistent with a Gaussian.

In principle, we require the bias models to pass both these tests in order to be deemed acceptable. Given the large number of cases considered in our analysis, and the computational cost of performing the ‘Noisy GOF’ test for each case, in practice we only perform the second test for our fiducial case. As we find largely consistent results between the two tests in this case, we resort to only the ‘Noiseless GOF’ test for the remainder of this work. As is customary, we set a threshold of a p -value larger than $p = 0.05$ for a model to pass a specific test.³²

³¹Strictly, this is only valid for linear models with independent basis functions (see e.g. ref. [112]). We will nevertheless use this criterion throughout this analysis, as we are mainly interested in deriving a threshold and not accurate probabilities-to-exceed.

³²We note that this criterion might appear overly generous as it could e.g. complicate a direct assessment of the GOF when analyzing observational data. Nevertheless we believe it to be adequate for this analysis given our simplistic treatment of systematic uncertainties: as described in section 3, we follow ref. [31] and add a 1% relative error floor to the covariance matrix, which amounts to half of the level assumed in ref. [32]. When applying these methods to data, we envision performing a more in-depth investigation of systematic uncertainties. Specifically, we could e.g. compare the two HEFT implementations and derive systematic error bounds from their residuals, specifically for spherical harmonic power spectra. We could include additional degrees-of-freedom in our fit using PCA components derived from these residuals or employ them to derive a more accurate estimate of the error floor. We expect these types of approaches to reduce the values for χ_{theory}^2 to levels that would lead to interpretable χ^2 values when applying the methods to data, but leave a detailed investigation to future work. In addition, all comparisons between the results for the noiseless GOF are consistent with the noisy GOF criterion, which provides further validation of our choice of fiducial GOF.

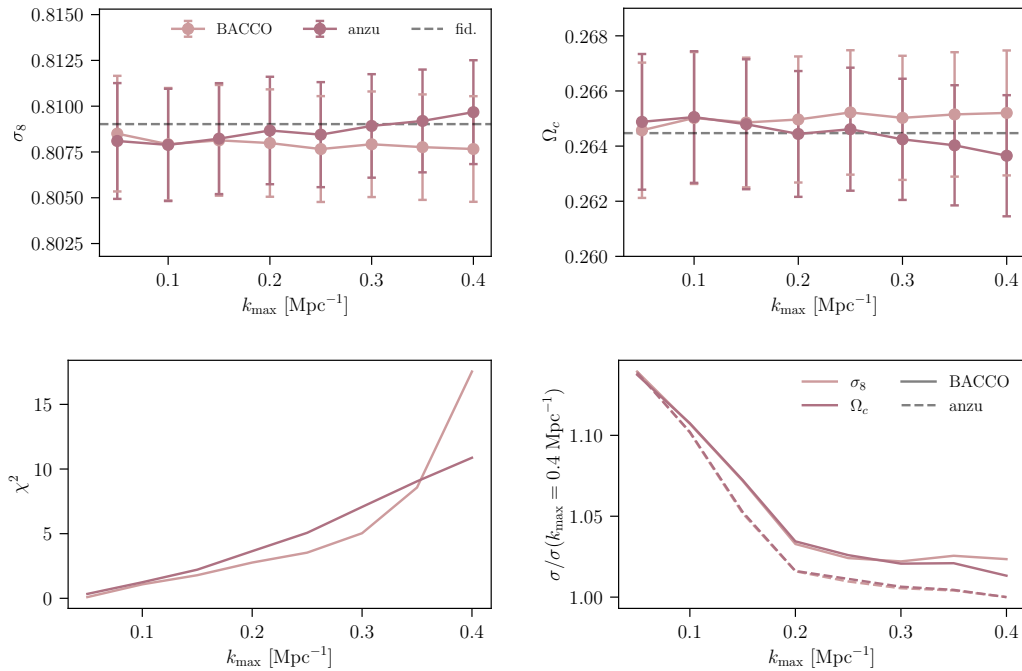


Figure 2. Results of fitting the red galaxy sample with the HEFT implementations considered in this work, **anzu** and **BACCO**. *Top panels:* recovered values of σ_8 and Ω_c as a function of maximal wavenumber, k_{max} . *Bottom panels:* minimum χ^2 -value obtained from the fit, and relative uncertainties on σ_8 and Ω_c , as a function of k_{max} . The uncertainties are normalized with respect to those obtained for **anzu** at $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$.

6 Results

In the following sections, we present the results of our analysis. Unless stated otherwise, all results will be presented for our fiducial, red sample. At the end of the section, we also give a summary of the results obtained for the alternative samples considered.

6.1 HEFT

We first investigate the performance of the two HEFT implementations, **anzu** and **BACCO**. As discussed in section 5.3, we minimize the likelihood with respect to two cosmological parameters $\{\Omega_c, \sigma_8\}$, as well as six bias parameters per redshift bin, $\{b_1, b_{1p}, b_2, b_{s^2}, b_{\nabla^2}, P_{\text{SN}}\}$. We keep the remaining cosmological parameters fixed at their fiducial values adopted in ABACUSSUMMIT.

The upper panels of figure 2 show the values of the cosmological parameters σ_8 and Ω_c recovered from fitting our fiducial red galaxy sample with **anzu** and **BACCO**. As can be seen, both HEFT methods are able to recover the true parameter value within 1σ uncertainty for all values of k_{max} considered.³³ $3\times 2\text{pt}$ analyses are generally most sensitive to a combination of σ_8 and Ω_c (or Ω_m), as these two parameters have degenerate effects on the power spectra

³³For **BACCO** as well as the PT-based methods discussed in section 6.2, we see a sign that the recovered values of σ_8 and Ω_c are systematically under- or over-estimated. These effects are not significant at the precision we are working in, but might be a sign for biases that could become important at higher signal-to-noise. However,

(cf. S_8 in weak lensing). It is therefore interesting to additionally investigate the resulting bias on the best-constrained combination of these two parameters. Without explicitly showing the results, we find that the differences between recovered and fiducial values of σ_8 and Ω_c result in shifts along the degeneracy direction depicted in the left-hand panel of figure 17 and thus do not result in a bias on the best-constrained parameter combination. The bottom left panel of figure 2 shows the minimal χ^2 obtained for both models, and we can use these values to assess the goodness-of-fit in all cases. The most stringent test for each model is obtained at the maximal k_{max} considered in our analysis, $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ (corresponding to $0.6 h \text{ Mpc}^{-1}$), and in the following we only discuss this case. The results of our goodness-of-fit tests are described in detail in appendix B.1, and here we only give a summary of our main findings. For the noiseless data vector (noiseless GOF), we find that both models, **anzu** and **BACCO**, pass our χ^2 tests.³⁴ In addition, most obtained fit residuals lie within 1σ of the synthetic data and the relative differences are generally of order 1%. We additionally consider two noisy realizations of the data vector (noisy GOF), finding all tests to pass with the exception of the analysis of one of the noisy realizations with **BACCO**, which we consider to be a statistical fluctuation. In all cases considered, we further find the distribution of fit residuals to be consistent with a Gaussian.

In order to test for possible systematic modeling uncertainties, we compare the consistency of the bias parameter values obtained for the two HEFT implementations. While a detailed description is deferred to appendix C, we generally find consistency between bias parameters derived using **anzu** or **BACCO**. The only exception being b_{∇^2} for which we find consistently higher values for **BACCO** than we do for **anzu**. We attribute these differences to slightly different implementations of the nonlocal power spectra in the two HEFT models, as well as known sensitivities of b_{∇^2} to small-scale implementation details such as different smoothing scales used when deriving template spectra (see e.g. ref. [9]).

Finally, we expect that systematics in the modeling might manifest themselves as dependencies of the derived bias parameter values on the maximal wavenumber k_{max} used in the analysis. In order to test for this, we thus compare the bias parameter values obtained for **anzu** when varying k_{max} from $k_{\text{max}} = 0.1 \text{ Mpc}^{-1}$ to $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$, finding largely consistent results.

Based on these results, the two HEFT implementations considered in this work thus appear promising for analyzing high-precision data as expected from LSST.

6.2 LPT/EPT

In addition to the HEFT implementations discussed above, we also assess the performance of two PT-based models, Lagrangian and Eulerian perturbation theory. As described in section 2.5, we consider two different ways to model the matter power spectrum $P_{\delta_m \delta_m}(z, k)$ as well as nonlocal contributions in $P_{\delta_g \delta_g}$ and $P_{\delta_g \delta_m}$. These are denoted **NLPk** for the **HALOFIT**-based model, and **PTPk** for the fully perturbative case. We remind the reader that this is

this could also be a feature of the particular simulation realization as the bias parameters derived as a function of k_{max} are highly correlated.

³⁴We note that for both HEFT methods as well as for our fiducial implementations of LPT and EPT, we find consistent results when we restrict our data vector to only include galaxy clustering and galaxy-galaxy lensing in a so-called 2x2pt analysis.

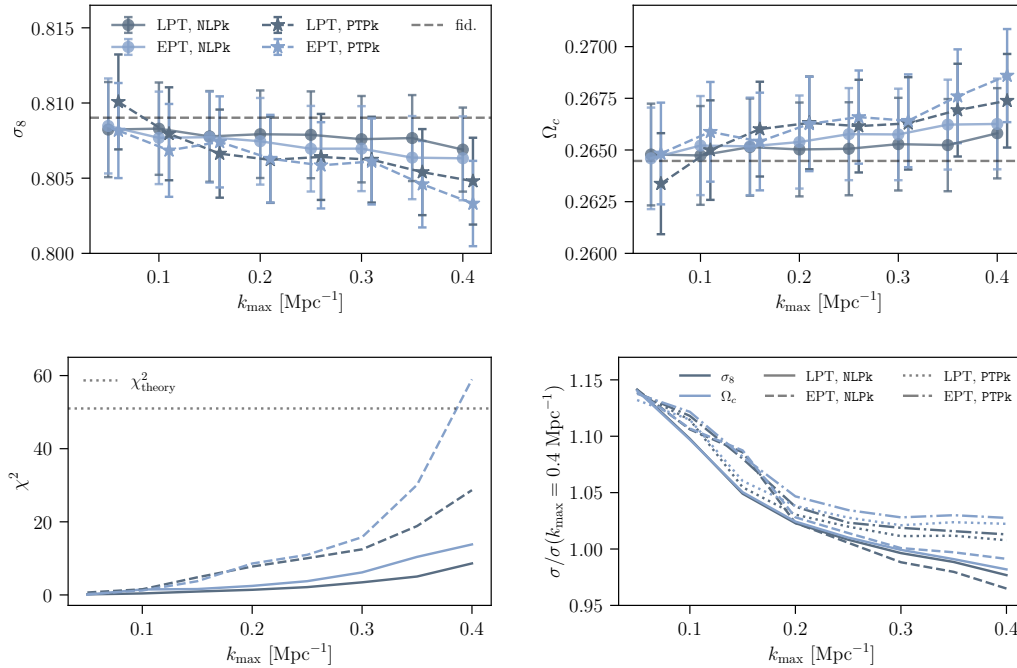


Figure 3. Results of fitting the red sample with the NLPk and PTPk implementations of LPT and EPT. *Top panels:* recovered values of σ_8 and Ω_c as a function of maximal wavenumber, k_{max} . The data points for the PTPk case have been displaced by $\Delta k_{\text{max}} = 0.01 \text{ Mpc}^{-1}$ for clarity. *Bottom panels:* minimum χ^2 -value obtained from the fit, and relative uncertainties on σ_8 and Ω_c , as a function of k_{max} . The uncertainties are normalized with respect to those obtained for **anzu** at $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$.

separate from our rescaling of the matter power spectrum for modeling cosmic shear (as described in section 5.2), which is performed for both NLPk and PTPk.

The results obtained from fitting the red galaxy sample with the LPT and EPT models are shown in figure 3. Similarly to the results for the two HEFT implementations, we find that for the NLPk models, the recovered values for σ_8 and Ω_c agree with their fiducial values within 1σ in all cases. For the PTPk models on the other hand, we find the recovered values for the cosmological parameters to start showing biases of around 1σ for $k_{\text{max}} \gtrsim 0.3 \text{ Mpc}^{-1}$, increasing to 1.5 to 2σ for $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$. This is borne out by our goodness-of-fit tests, as described in appendix B.2. We find that both NLPk methods pass our tests on the noiseless data. Despite both PTPk models returning biased constraints on cosmological parameters, we find that the fully perturbative implementation of LPT still passes our goodness-of-fit test, while EPT yields a χ^2 larger than χ^2_{theory} . These results suggest that predicting the galaxy power spectra considered in this analysis up to a maximal wavenumber of $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ requires more accurate modeling of the matter power spectrum and nonlocal terms, which we achieve here through our use of the HALOFIT fitting function. These findings are consistent with previous studies, such as e.g. refs. [19, 20, 71], which also found significantly improved performance of EPT bias models when using non-perturbative predictions for the matter power spectrum.

Nevertheless, comparing to previous analyses that roughly find LPT and EPT bias models to only reach up to $k_{\text{max}} \sim 0.3 h \text{ Mpc}^{-1}$ (see e.g. ref. [18]), these results are somewhat

surprising. We believe this to be due both to differences in the considered data vector as well as the criterion chosen to assess goodness-of-fit: first, in this work we focus on spherical harmonic power spectra, which constitute a line-of-sight projection of the underlying three-dimensional power spectrum usually studied in the literature. The line-of-sight projection results in the smoothing of power spectrum features and might thus increase the reach of perturbative nonlinear bias models. In contrast to other works, we additionally do not include redshift space distortions (RSDs) in our analysis, as their impact on projected statistics is expected to be small (although potentially not negligible [95]), which might also lead to better performance of PT-based models. Finally, our results could also be affected by our pragmatic approach to assess the goodness-of-fit of a given model: we only require the fit to return unbiased constraints on Ω_c and σ_8 , as well as returning a χ^2 -value passing the criteria discussed in section 5.4. In particular, we do not make any requirements on model residuals as was done in previous analyses, and we include a theoretical relative error floor of 1% in all our covariances. We believe it is for these reasons that our results find perturbative bias models to be applicable up to slightly smaller scales than previous analyses. In appendix E we further investigate this by performing an analogous analysis of three-dimensional power spectrum data in real space from the UNIT simulation [113].

As above, in a final test we compare the bias parameter values obtained using our fiducial NLPk implementation of LPT and EPT. As discussed in appendix C, we generally find good agreement between both methods, although the EPT approach seems to prefer significantly larger values for b_{∇^2} and P_{SN} compared to LPT. In addition, the recovered bias parameters for the two NLPk models are generally consistent with those obtained from the HEFT methods. The only exception being b_{∇^2} , for which we obtain significantly lower values for the HEFT implementations than we do for the perturbation-theory-based models, particularly at high redshift. As further discussed in appendix C, this could be a possible sign for larger model inaccuracies in the PT-based methods as compared to HEFT.

6.3 Bias consistency relations

As discussed in section 2.4, we expect Eulerian and Lagrangian bias parameters to exhibit consistency relations under purely gravitational evolution. While we do not anticipate these relations to hold exactly due to non-gravitational processes involved in galaxy formation, they present a complementary means for validating our results, and our bias parameter fitting procedure in particular. We therefore investigate the validity of the coevolution relations given in eq. (2.5). The results using the NLPk models with $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ are illustrated in figure 4 both for our fiducial, red sample as well as the maglim sample, which we discuss in more detail below: in the upper panels we show the value of b_2 and b_{s^2} as a function of b_1^L, b_2^L and $b_1^L, b_{s^2}^L$, respectively, for all tomographic redshift bins. The lower panel illustrates the relation between b_{s^2} and b_1 . In these figures, we also show the uncertainties associated with the various bias parameters. However, as discussed in more detail in sections C and D, we find the Fisher matrix-derived uncertainties on bias parameters to be rather unstable for the full model, while they match their MCMC counterparts if we fix either b_{s^2} or b_{∇^2} . Where possible we thus always show error bars obtained setting $b_{s^2} = 0$ (even though this has no effect on cosmological parameter uncertainties). As this is not possible in this case, we

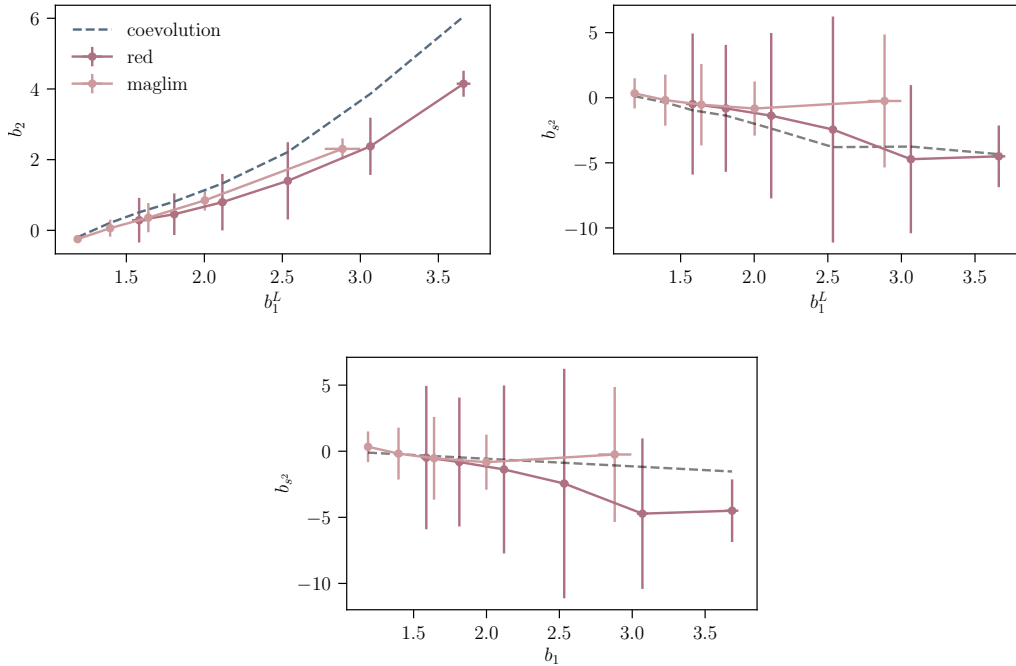


Figure 4. Comparison of the theoretical bias consistency relations given in eq. (2.5) to the relations obtained from fitting the ABACUSSUMMIT data with the MLPk implementations of EPT and LPT. Each data point denotes a different tomographic redshift bin for either the maglim or the red sample. All cases show results for $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$. *Top panels:* consistency relations between LPT and EPT biases. *Bottom panel:* consistency relation for EPT biases in case of local-in-matter-density (LIMD) Lagrangian bias.

caution the reader to keep these instabilities in mind when interpreting the uncertainties. As can be seen from the figure, the obtained bias values largely follow the theoretically expected trends. However, we do see signs of significant deviations, particularly in the consistency between b_2 and b_1^L, b_2^L . Further investigating this, we also compare our results to the empirical relations between Lagrangian bias parameters found in ref. [69] and repeat the analysis for the UNIT simulations, finding similar results in both cases. This suggests that these findings are not driven by our use of ABACUSSUMMIT, and that empirical bias relations do not provide a better fit to our data than those derived from coevolution. Another possible reason for these results might be our choice to not consider third-order bias parameters in EPT, which breaks the full correspondence between EPT and LPT (as discussed in section 2.4), and might lead to some of these differences being absorbed by lower-order bias parameters. We leave a further investigation to future work.

6.4 Alternative samples

We investigate the generality of the results presented so far by applying the nonlinear bias models considered in this work to the analysis of three alternative samples, as described in sections 4 and 5: a magnitude-limited galaxy sample, a galaxy sample featuring assembly bias, and a joint analysis of our fiducial data vector with CMB lensing data from a CMB

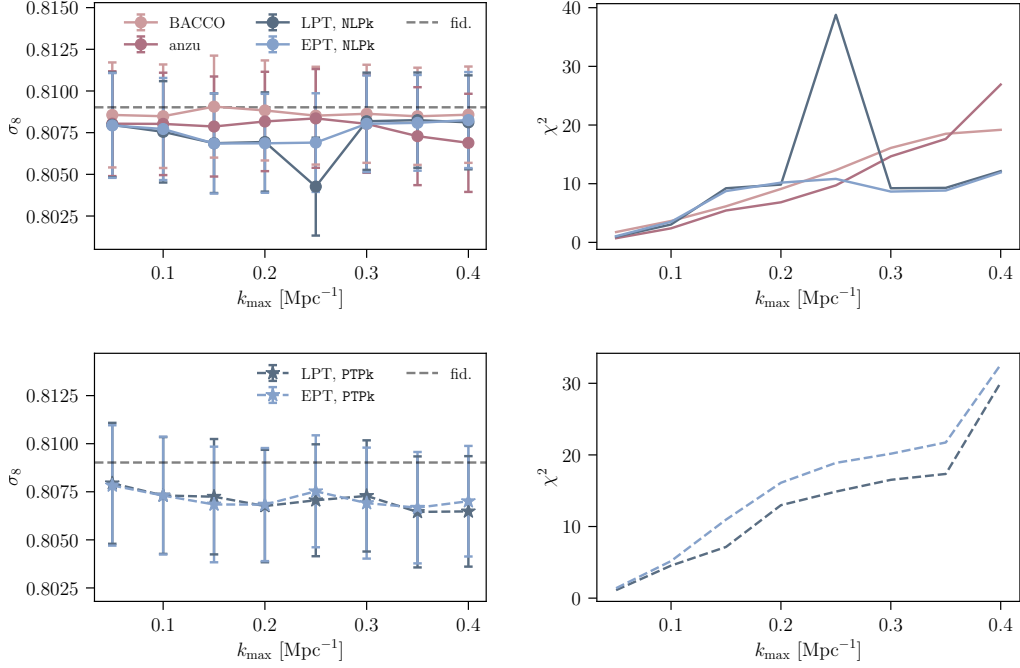


Figure 5. Results of fitting the maglim sample using the nonlinear bias models considered in this work, LPT, EPT, **anzu** and BACCO. *Top panels:* recovered values of σ_8 and minimal χ^2 obtained from the fit as a function of maximal wavenumber, $k_{\max}$ for HEFT and the NLPk implementation of EPT and LPT. *Bottom panels:* same as above but using the PTPk implementation of LPT and EPT.

S4-like experiment. For all the following cases, we only show the results for σ_8 as the results for Ω_c are qualitatively similar.

6.4.1 Magnitude-limited sample

We repeat our analysis using the maglim sample discussed in section 4, using the same redshift bins for galaxy clustering and weak gravitational lensing. In its fiducial implementation BACCO only covers redshifts $z \leq 1.5$, and we thus need to extend the emulator using LPT at high redshift in order to model the maglim sample. The results obtained for all bias models after implementing this extension are shown in figure 5. As can be seen from the upper panels we generally recover unbiased results for all our four fiducial bias models.³⁵ For the maglim sample we only consider five clustering bins, thus leading to an effective number of degrees-of-freedom of $dof = 436$. Following section 5.4, we thus require $\chi^2_{\text{theory,max}} \leq 49$ to pass our goodness-of-fit test. From figure 5 we can see that all minimal χ^2 -values derived using our fiducial bias models satisfy this criterion. In the lower panels of figure 5, we additionally show the recovered parameter values for the PTPk variants of LPT and EPT. As opposed to the results obtained for the red sample, we recover unbiased constraints on

³⁵The only exception is the fit using LPT at $k_{\max} = 0.25 \text{ Mpc}^{-1}$, which is an outlier as compared to the other cases. We suspect this to be due to a parameter degeneracy specific to this case preventing the minimizer to converge to the global minimum. This hypothesis is strengthened by the fact that evaluating the likelihood at $k_{\max} = 0.25 \text{ Mpc}^{-1}$ with the best-fit parameter values derived using data up to $k_{\max} = 0.4 \text{ Mpc}^{-1}$ leads to a significantly lower χ^2 -value.

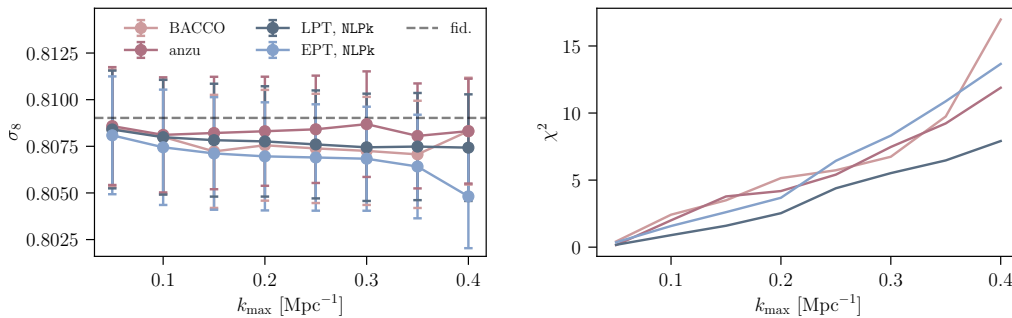


Figure 6. Results of fitting the galaxy sample with assembly bias using the main nonlinear bias models considered in this work, LPT, EPT, *anzu* and BACCO. Recovered values of σ_8 and minimal χ^2 obtained from the fit as a function of maximal wavenumber, k_{max} .

σ_8 and Ω_c also for those models. This suggests a higher reach of fully-perturbative bias models for the maglim sample, likely related to the fact that its associated linear bias is significantly smaller than that of the red sample.

6.4.2 Assembly bias

Given the current uncertainties on the dependence of galaxy clustering on quantities beyond halo mass, it is important to assess the performance of nonlinear bias models in the presence of these effects. To this end we apply our four fiducial models to the galaxy sample with assembly bias described in section 4, jointly fitting the combination of galaxy clustering, galaxy-galaxy lensing and cosmic shear. The results for both HEFT methods as well as the NLPk implementation of EPT and LPT are shown in figure 6.³⁶ As can be seen, all four models yield unbiased constraints on cosmological parameters, thus confirming theoretical expectations that these models offer the flexibility to capture the effects of assembly bias as implemented in our simulated data for LSST Y10-like precision (see ref. [31] for similar results).

6.4.3 Including CMB lensing convergence

The combination of LSS surveys with CMB measurements has been shown to be a powerful way to constrain deviations from Λ CDM (see e.g. refs. [114–118]), and thus constitutes a key priority for current and future surveys. We therefore additionally test the applicability of current nonlinear bias modeling techniques to the combination of galaxy clustering, galaxy-galaxy lensing, cosmic shear and their cross-correlations with the CMB lensing potential. Specifically, we use the *anzu* implementation of HEFT to analyze the simulated joint LSST and CMB S4 data vector described in section 5.2. The inclusion of CMB lensing leads to reduced uncertainties on cosmological parameters by roughly 10% to 30% as can be seen from figure 7. Nevertheless, we find unbiased recovery of σ_8 and Ω_c for all maximal wavenumbers considered even for this extended data vector, which suggests that these methods meet the accuracy-requirements for future joint analyses of LSST with CMB S4 or SO.

³⁶Given the results found for the red sample in section 6.1, we do not consider the PTPk implementation of EPT and LPT here.

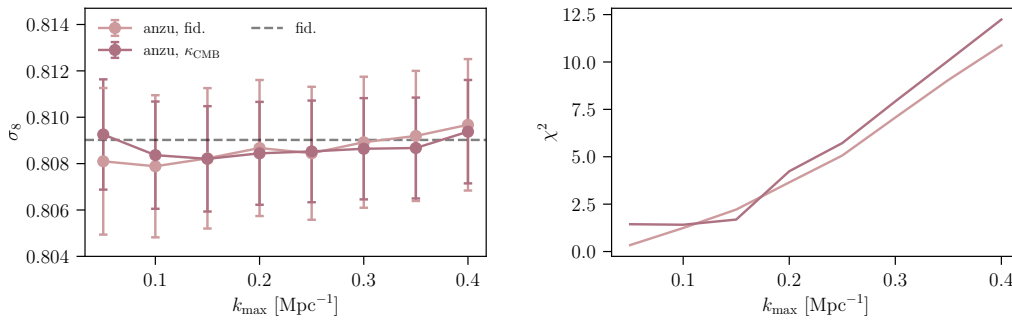


Figure 7. Results of fitting the red galaxy sample combined with CMB S4-like CMB lensing using the four nonlinear bias models considered in this work, as compared to our fiducial results. Recovered values of σ_8 and minimal χ^2 obtained from the fit as a function of maximal wavenumber, k_{max} .

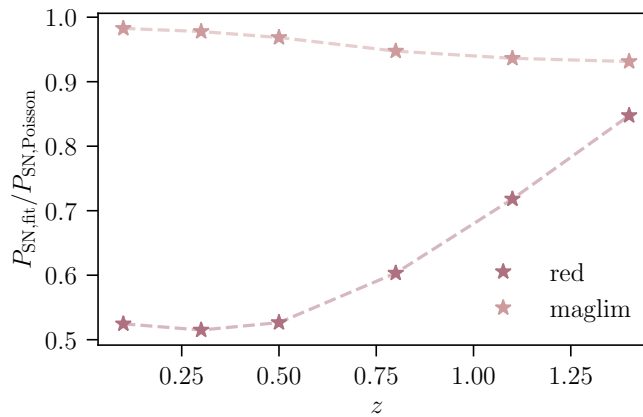


Figure 8. Ratio of recovered value of stochasticity power spectrum and its Poissonian expectation as obtained from fitting three-dimensional power spectrum data with the **anzu** model at discrete redshifts. All results are shown for $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$.

6.4.4 No halofit rescaling

In a final test, we investigate the impact of not rescaling the matter power spectrum to match the HALOFIT prediction. While we do not explicitly show the results, in this case we recover significantly biased constraints on cosmological parameters for all bias models considered. These results suggest that at the precision of LSST Y10, we are sensitive to systematic differences between the HALOFIT fitting function and matter power spectra measured from simulations. This is not related to the problem of characterizing galaxy bias, which is the focus of this paper, and instead is limited to the shear-shear component of the 3×2 pt data vector. A thorough study of the precision with which the matter power spectrum needs to be modeled for LSST (including the impact of baryons) will be the focus of future work.

6.5 Stochasticity

A number of studies have found evidence for non-Poissonian stochasticity in simulated galaxy samples (see e.g. refs. [54, 55]). It is therefore interesting to compare the levels

of stochasticity obtained in this work with their Poissonian expectations. The results presented so far have been derived using spherical harmonic power spectra, which constitute line-of-sight projections of the associated three-dimensional power spectra. It is therefore difficult to relate the redshift-averaged stochasticities obtained in our analysis to their Poisson counterparts, and we thus use measurements of three-dimensional power spectra to investigate stochasticity. As described in detail in appendix F, we consider a data vector consisting of $\mathbf{d} = \{P_{gg}(z, k), P_{gm}(z, k), P_{mm}(z, k)\}$ at discrete redshifts, and determine the best-fitting bias parameters.

The ratio of the derived values of P_{SN} to their Poissonian expectations for $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ are shown in figure 8 as a function of redshift. As can be seen, for the red sample we detect significantly sub-Poissonian stochasticity, with the largest discrepancies at low redshift. As shown in ref. [55], non-Poissonian stochasticity is due to two different effects: halo exclusion and HOD stochasticity. Halo exclusion denotes the effect that large halos are not a Poisson realization of an underlying density field as the halos cannot overlap [54]. This effectively decreases the volume available to the halos and thus reduces stochasticity. HOD stochasticity on the other hand denotes the additional stochasticity in galaxy samples caused by the variance in the galaxy-halo-distribution. This effect is always positive and thus acts to increase P_{SN} . The stochasticity of a given galaxy sample is thus determined by the combined effects of halo exclusion and HOD stochasticity. Our results therefore suggest that the host halos of the red galaxy sample exhibit significant halo exclusion, which is qualitatively and quantitatively comparable to the findings presented in ref. [55] for DESI LRGs. However, our results suggest lower values for P_{SN} , especially at low and high redshift. We believe these changes to be due to either a difference in methodology to constrain stochasticity, or differences in the adopted HODs for the DESI LRG sample, or a combination of both.³⁷

Performing an analogous analysis for the maglim sample, we find a significantly lower level of sub-Poissonian noise than what is detected for the red sample. Specifically we find a 7% reduction as compared to the Poissonian expectation at most. Comparing the mean halo masses of both samples at redshift $z = 0.65$, we find $\bar{M}_h = 4.1 \times 10^{13} M_{\odot}$ for the red sample, while we find $1.3 \times 10^{13} M_{\odot}$ for the maglim sample. In addition, we find significantly higher satellite fractions for the maglim sample. These higher satellite fractions for the maglim sample can be interpreted using the results of e.g. ref. [62] that finds the satellite fraction to increase steeply with decreasing luminosity but also with increasing redness of the sample. The higher satellite fraction of the maglim sample thus suggests that its lower luminosity dominates over its reduced fraction of red galaxies. Combined, these two properties of the maglim sample indicate that it exhibits stochasticity closer to its Poissonian expectation because its host halos show lower halo exclusion and the larger satellite fractions give rise to higher HOD stochasticity.

As shown in appendix G, the analysis in terms of spherical harmonic power spectra leads to similar conclusions, as we find good agreement between bias parameter constraints derived using two- and three-dimensional power spectra.

³⁷The differences in HODs might be caused by our assumption of a redshift-evolving HOD model, and comparing secondary halo properties of our sample to those reported in ref. [30], we find the largest differences at low and high redshift, which is also where we see the largest discrepancies with ref. [55].

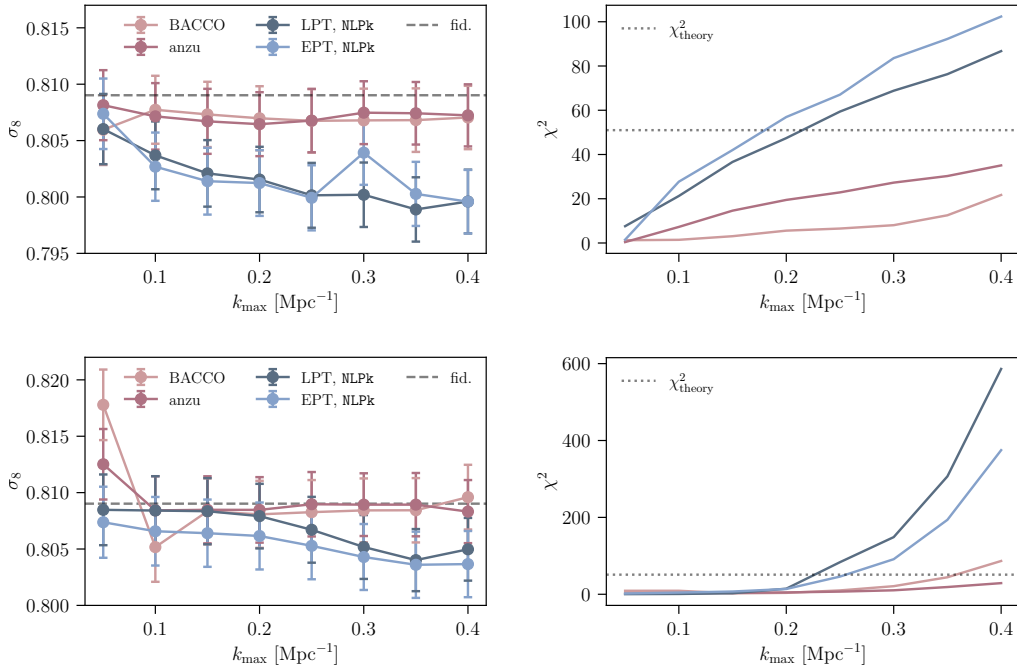


Figure 9. Results of fitting the red galaxy sample with reduced bias models. *Top panels:* recovered values of σ_8 and minimal χ^2 obtained from the fit as a function of maximal wavenumber, k_{max} when fixing $b_{s^2} = 0$. *Bottom panels:* the same as in the upper panels but fixing $b_{\nabla^2} = 0$.

Finally, it is interesting to investigate if the stochasticity shows evidence of scale dependence. To this end, we focus on the analysis of the red sample with **anzu** and compare the stochasticity derived as a function of k_{max} for all tomographic redshift bins used in the spherical harmonic analysis (cf. section 6.1). While not showing the results explicitly, we find consistent results for the stochastic bias parameters derived as a function of k_{max} , and thus do not find any evidence for scale-dependence from this simple test.

6.6 Minimal bias model

The results presented so far have been derived considering galaxy bias terms up to quadratic order, resulting in a model with six bias parameters per redshift bin. As shown in sections 6.1, 6.2 and 6.4, this model allows us to fit the ABACUSSUMMIT data reasonably well, but it is interesting to ask if we can reduce the complexity of the model while maintaining its performance. To this end, we first focus on the red sample and repeat our analysis, setting a number of bias parameters to zero. Specifically, we consider the cases $b_{s^2} = 0$, $b_{\nabla^2} = 0$ and $b_{1,p} = 0$. This choice corresponds to the most drastic approach to model reduction, and we note that one could also explore alternative methods such as placing tight priors on bias parameters. The values recovered for σ_8 and their corresponding minimal χ^2 -values when setting $b_{s^2} = 0$ and $b_{\nabla^2} = 0$ respectively are shown in figure 9 for the HEFT methods as well as the NLPk implementation of EPT and LPT. As can be seen, we find that the HEFT methods return unbiased constraints on σ_8 even for these reduced models, while the perturbation-theory-based methods show significant biases, most pronounced at high

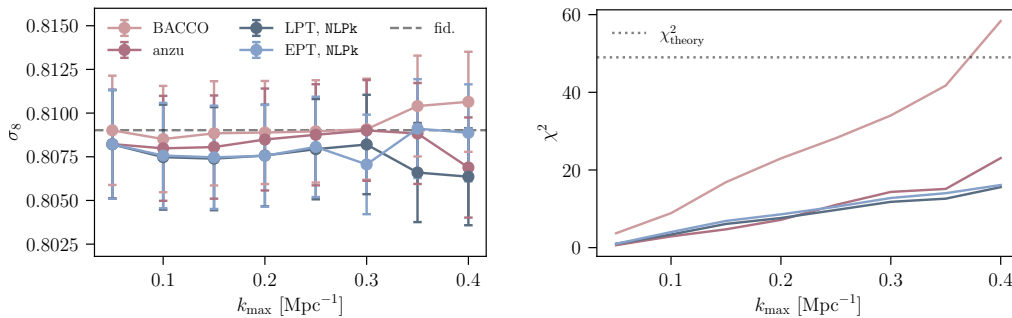


Figure 10. Results of fitting the maglim sample using the nonlinear bias models considered in this work, LPT, EPT, **anzu**, BACCO, and setting $b_{s2} = 0$ in all cases. The plots show the recovered values of σ_8 and minimal χ^2 obtained from the fit as a function of maximal wavenumber, k_{max} .

k_{max} . This is borne out by looking at the χ^2_{min} -values: for $b_{s2} = 0$ both HEFT methods pass our goodness-of-fit test at $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$, while the PT methods do not. For $b_{\nabla^2} = 0$ on the other hand, only **anzu** recovers a χ^2 low enough to pass our test at the largest wavenumber considered. It is interesting to note that while the biases in the PT methods are more pronounced when setting $b_{s2} = 0$, the χ^2_{min} -values are significantly worse for the model with $b_{\nabla^2} = 0$. Additionally, while we do not show the results, we find that all models yield biased constraints on cosmological parameters when we set $b_{1,p} = 0$, highlighting the need to account for the evolution of at least the linear bias across each redshift bin at the level of precision achieved by LSST Y10.

We test the stability of these results by comparing the values of the common bias parameters recovered for the three levels of model complexity considered ($\{b_1, b_{1,p}, b_2, b_{s2}, b_{\nabla^2}\}$, $\{b_1, b_{1,p}, b_2, b_{\nabla^2}\}$, and $\{b_1, b_{1,p}, b_2, b_{s2}\}$) using the **anzu** implementation of HEFT. Without explicitly showing the results, we find all recovered values to be consistent within uncertainties. This suggests that the values recovered for the linear and quadratic bias, and, perhaps more importantly, for the stochasticity P_{SN} , are not driven by inaccuracies in the bias model, but rather represent a genuine physical feature of the sample under consideration.

In order to investigate the sample-dependence of these findings, we repeat our analysis for the maglim and assembly bias samples. The results obtained when setting the tidal bias to zero for the magnitude-limited sample, i.e. $b_{s2} = 0$, are shown in figure 10 for the HEFT methods as well as the NLPk implementation of EPT and LPT. As can be seen, all bias models recover unbiased constraints on cosmological parameters, but the BACCO fits lead to χ^2 -values significantly higher than our threshold.³⁸ We find similar results when setting the non-local bias to zero, i.e. $b_{\nabla^2} = 0$, and thus do not show the results explicitly. The only exception is that in this case all models, including BACCO, pass our goodness-of-fit tests. Investigating this further, we repeat our analysis for $b_{s2} = 0$ excluding the highest redshift bin for the clustering statistics. In this case, we find unbiased constraints on cosmological parameters as well as minimal χ^2 -values passing our test criteria for all bias models. The model residuals for BACCO thus appear predominantly driven by the highest redshift bin, and we suspect that the low

³⁸We note that in contrast to the results presented in section 6.4.1, the PTPk models without tidal bias, $b_{s2} = 0$, give biased constraints on cosmological parameters.

model performance observed might be caused by inaccuracies in our extension of BACCO to high redshift, but we leave a more detailed analysis to future work. For the assembly bias sample, we find results very similar to those obtained for the red sample, i.e. EPT and LPT yield biased constraints in all cases, while both HEFT implementations show good performance.

These results confirm the expectation that the minimal bias model depends on the data set as well as the actual model. For the red and the assembly bias samples, we find that all PT-based methods (NLPk and PTPk) lead to biases when setting $b_{s2} = 0$ or $b_{\nabla^2} = 0$, while the HEFT methods perform generally well. We attribute this to the fact that a less accurate model will need more parameters to fit a given data set in order to compensate for inaccuracies in the template power spectra. It is interesting that a vanishing tidal bias appears to be more consistent with the data, while we start seeing biases even for the HEFT methods when setting $b_{\nabla^2} = 0$. In addition, it is worth noting that the scale-dependent modification induced by $b_{\nabla^2} \neq 0$ is qualitatively different in real and harmonic space. Thus, these results may be sensitive to our analysis choices, in which scale cuts are imposed in harmonic space. Finally, we find both the HEFT and NLPk methods to yield unbiased results for the maglim sample, thus suggesting that the inaccuracies of PT-based models are most pronounced for more highly biased galaxy samples as compared to their less biased counterparts, and thus their modeling requires a smaller number of bias parameters.

It is worth noting that an alternative approach to finding a minimal bias model would be to make use of the rather tight relations between bias parameters found in the literature (see section 2.4 or e.g. ref. [69]) and use them to fix the value of certain bias parameters rather than set them to zero. We leave an investigation thereof to future work.

6.7 Error as a function of maximal wavenumber

In order to determine optimal scale cuts for a given analysis involving galaxy clustering data, it is essential to investigate to what extent parameter constraints are tightened by the inclusion of additional small-scale information. In particular it is interesting to investigate if in the setup considered in this analysis, additional small-scale information results in tighter constraints on cosmological parameters, or rather serves to constrain bias parameters more tightly. To this end we compare the uncertainties on cosmological parameters as well as the four bias parameters b_1 , b_2 , b_{∇^2} and P_{SN} obtained as a function of k_{max} .³⁹ The relative uncertainties for σ_8 and Ω_c are shown in the lower right panel of figure 2 while figure 11 illustrates the corresponding relations for the bias parameters. All uncertainties are normalized with respect to those obtained for **anzu** at $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$, and as can be seen from figure 2, the marginalized 1σ errors on cosmological parameters decrease only by roughly 15% as we increase k_{max} from 0.05 Mpc^{-1} to 0.4 Mpc^{-1} . These gains are small given the significant increase in the amount of small-scale information included in the fits. Looking at figure 11 on the other hand, we see that the bias parameter constraints tighten significantly, by up to 2 orders of magnitude, for the same increase in k_{max} , in particular for b_{∇^2} and P_{SN} . These results suggest that in a joint analysis of galaxy clustering, galaxy-galaxy lensing and weak lensing with LSST Y10-like specifications, the information contained in

³⁹We focus on these, as we find b_{1p} to be only weakly constrained in our analysis. We further note that we compute uncertainties setting $b_{s2} = 0$, as discussed in section 6.1.

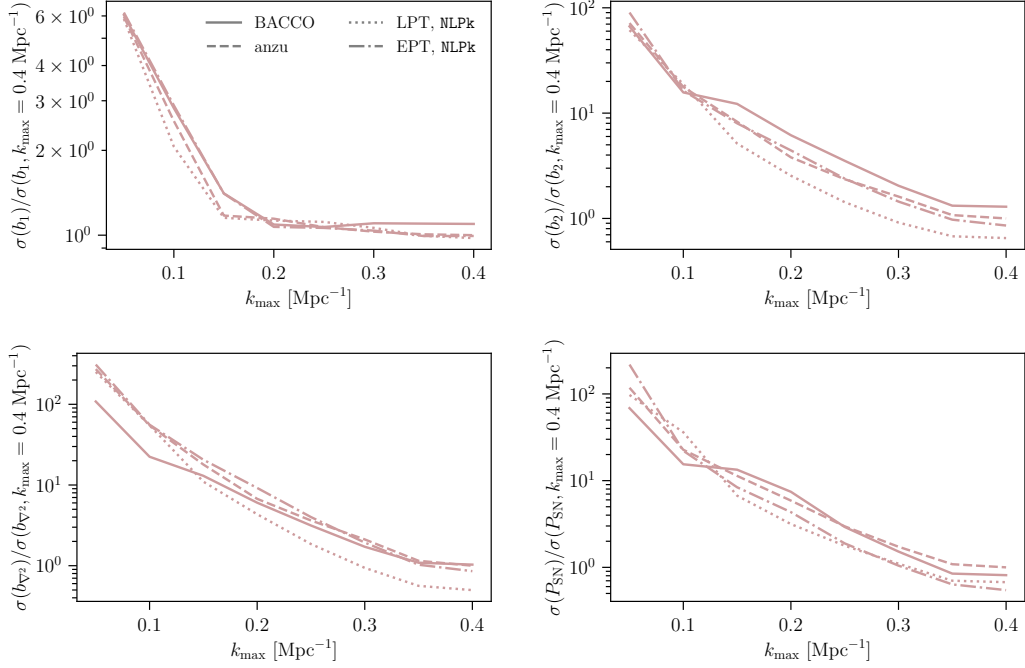


Figure 11. Relative uncertainties on the bias parameters b_1 , b_2 and b_{sn} for the fourth redshift bin as a function of maximal wavenumber, $k_{\max}$. All uncertainties are normalized with respect to those obtained for *anzu* at $k_{\max} = 0.4 \text{ Mpc}^{-1}$.

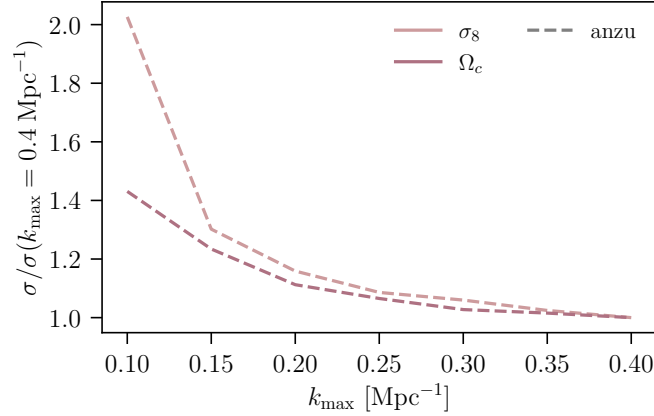


Figure 12. Relative uncertainties on σ_8 and Ω_c as a function of maximal wavenumber, $k_{\max}$, obtained when jointly fitting galaxy clustering and galaxy-galaxy lensing from our fiducial sample with *anzu*.

small-scale galaxy clustering mainly serves to constrain galaxy bias parameters as compared to cosmological parameters.

The lack of improvement in cosmological constraining power with increasing $k_{\max}$ appears to be driven by the cosmic shear data, as we find that the errors on σ_8 and Ω_c decrease by roughly a factor of $1.5 - 2$ when we increase $k_{\max}$ from 0.1 Mpc^{-1} to 0.4 Mpc^{-1} and only

consider galaxy-galaxy and galaxy-shear correlations, as can be seen from figure 12.⁴⁰ To further test this hypothesis, we rerun our fiducial analysis, freeing up the Hubble parameter h and the scalar spectral index n_s in addition to σ_8 and Ω_c . While the improvements in uncertainties for σ_8 and Ω_c are similar, we find larger gains for h and n_s of up to 30%. In particular, the gains do not seem to saturate at high k_{max} as they do in our fiducial analysis.

While the relatively modest gains in cosmological constraining power seem to be partly driven by our inclusion of weak lensing data, which yields tight constraints on σ_8 and Ω_c , and thus limits the gains from small-scale clustering, we do see a general trend that small-scale clustering improves bias parameter constraints more significantly than constraints on their cosmological counterparts. However, we caution the reader that these conclusions might change when constraining a larger number of cosmological parameters, including systematic uncertainties, or considering data vectors different from the ones we are investigating in this work (e.g. including higher-order correlators, or redshift-space distortions in spectroscopic survey analyses). As an example, galaxy-galaxy lensing at small scales has been shown to help constrain intrinsic galaxy alignments and photometric redshift systematics, which would otherwise significantly degrade the constraining power of cosmic shear (see e.g. ref. [119]).

7 Conclusions

In this work, we compare the performance of a number of nonlinear galaxy bias models when applied to an LSST Y10-like tomographic joint analysis of galaxy clustering, galaxy-galaxy lensing and cosmic shear (3×2pt analysis). Specifically, we compare two perturbative approaches, Lagrangian perturbation theory (LPT) [10] and Eulerian perturbation theory (EPT) to two implementations of Hybrid Effective Field Theory (HEFT), **anzu** and **BACCO**, which combines a perturbative bias expansion in Lagrangian space with an exact treatment of the gravitational evolution via cosmological simulations [16, 31, 32]. We test all the methods using simulated data vectors computed from the ABACUSSUMMIT [78, 79] cosmological simulation, considering several different galaxy samples: a DESI-like red sample, a magnitude-limited sample based loosely on HSC DR1, and a galaxy sample with assembly bias. We fit these simulated data using all bias models considered, keeping terms up to second order, and account for nonlocal bias as well as deviations from Poissonian stochasticity. In a final step, we compare their performance based on the accuracy and precision of the constraints obtained for the cosmological parameters σ_8 and Ω_c as well as the goodness-of-fit.

For our fiducial, red galaxy sample we find that the two HEFT implementations allow us to jointly model galaxy clustering, galaxy-galaxy lensing and cosmic shear with LSST Y10-like precision up to at least a maximal wavenumber of $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$. This is also true for LPT and EPT when we combine these methods with non-perturbative predictions for the matter power spectrum entering some of the terms in the expansion (NLPk). In contrast, when we use the predictions from perturbation theory for these terms (PTPk), the LPT and EPT implementations lead to biased constraints on cosmological parameters for $k \gtrsim 0.2 \text{ Mpc}^{-1}$. We find comparable results when analyzing the galaxy sample with assembly bias. For the magnitude-limited sample on the other hand, we find good performance for all

⁴⁰While the figure only shows **anzu**, we find similar results for the other bias models.

bias models, including EPT and LPT with a perturbative prediction for the matter power spectrum. We further consider an extension of our fiducial galaxy sample with CMB lensing cross-correlations loosely matching the specifications for CMB S4, finding unbiased constraints on cosmological parameters as well as good minimal χ^2 -values. In all these analyses, we find significant detections of non-Poissonian stochasticity in the galaxy clustering auto-correlations.

We further investigate the effect of reducing bias model complexity by setting the tidal and nonlocal bias to zero respectively, finding sample- and model-dependent results. We find that the HEFT approaches are able to obtain unbiased constraints and provide a good fit to the data with this reduced models in most cases, the only exception being BACCO for vanishing nonlocal bias and $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$. In turn, while LPT and EPT perform well on the magnitude-limited sample, they lead to biases on cosmological parameters when applied to the more highly-biased red sample (with or without assembly bias) within these reduced parameterizations. This is the case regardless of the prescription used to model the matter power spectrum (NLPk or PTPk).

Investigating the constraints on cosmological and bias parameters obtained as a function of maximal wavenumber k_{max} , we find the uncertainties on cosmological parameters to decrease only by around 15% as we increase k_{max} from 0.05 Mpc^{-1} to 0.4 Mpc^{-1} . The bias parameter uncertainties on the other hand decrease significantly, in some cases by more than an order of magnitude. Removing the weak lensing auto-correlations from our data vector yields larger relative improvements on cosmological parameter uncertainties, of up to a factor of 2. This is a useful case to consider as the weak lensing auto-correlations can have separate systematic uncertainties (e.g. PSF induced additive shear contributions) while not being sensitive to galaxy bias, so separating out the cosmology inferred from those is a powerful consistency check. Nevertheless, the qualitative result is that pushing towards smaller scales seems to lead to significantly larger improvements in bias parameters than in cosmological parameters. These results are subject to a number of caveats: most importantly, we only consider constraints on galaxy bias as well as two cosmological parameters, σ_8 and Ω_c , and we do not account for systematics such as photometric redshift uncertainties or calibration biases in cosmic shear. This might artificially increase the constraining power of weak lensing, thus reducing the gain obtained from small-scale clustering. We have also not considered uncertainties due to magnification, intrinsic alignments and their interplay with photo-z uncertainties — these effects can modify our findings on the improvements from small scale information (likely in the direction of greater gain). We leave an investigation of these effects to future work.

In summary, our results confirm the performance of HEFT approaches found in previous work, while suggesting a potentially higher reach for EPT and LPT than previously found. This is likely in part due to our focus on projected clustering statistics (as opposed to three-dimensional clustering in redshift space), and to the specific metrics used to quantify goodness of fit (based on the expected performance of an LSST-like experiment, rather than on ad-hoc precision requirements).

With regards to LSST, the results of our analysis suggest that current nonlinear bias models appear promising for the analysis of tomographic galaxy clustering, galaxy-galaxy lensing and weak lensing from LSST Y10 data. Amongst the methods investigated, we find the recently developed HEFT methods to show particular promise. This bodes well for future

tomographic galaxy clustering and 3×2 pt analyses using small-scale information for both current and future photometric surveys such as LSST and Euclid. Nevertheless, more work is necessary in order to be able to fully exploit these methods in a robust and reliable manner. First, application of these methods to future data will require investigating the impact of observational effects such as photometric redshift uncertainties or large-scale galaxy clustering systematics and associated scale cuts. In addition, as we have shown, a good characterization of the theoretical uncertainties associated with the non-linear scheme used is vital to obtain unbiased constraints with adequate errors. A more precise treatment and modeling of these uncertainties than that used in our analysis will therefore be needed before the bias models studied here can be applied to future data sets. Furthermore our results have highlighted the susceptibility of perturbative bias prescriptions to modeling of the matter power spectrum, thus requiring higher-accuracy models. Although our analysis has covered the most likely target samples for galaxy clustering (LRG-like and magnitude-limited samples), including some amount of assembly bias, the results found here should be validated against a wider variety of galaxy samples, incorporating different physical effects such as satellite segregation and baryonic effects in the dark matter distribution [120]. Furthermore, as the internal consistency relations between different bias coefficients provide an avenue to reduce the freedom allowed to the bias model, studying the applicability of these relations to these samples in the context of LSST would be a useful exercise. Finally, higher-order statistics have the potential to unlock a significant amount of untapped non-Gaussian information in the galaxy distribution, and thus studying the ability of the bias models explored here to describe these observables is of high priority.

Acknowledgments

This paper has undergone internal review by the LSST Dark Energy Science Collaboration. We kindly thank the internal reviewers Simone Ferraro, Andrew Hearin and Shivam Pandey for providing helpful comments, which helped us improve the quality and clarity of the paper.

We are very happy to thank Toshiya Namikawa and Colin Hill for sharing the CMB S4 lensing noise curves and for help with their usage. With pleasure we would also like to thank Martin White for many very helpful comments and suggestions.

We would further like to thank the anonymous referee for carefully reading our manuscript and many helpful suggestions and comments.

The contributions from the primary authors are as follows: DA: Co-designed the project, constructed simulated data vectors, contributed to likelihood software. NF: Performed initial fits with LPT and EPT models to determine the minimum scales to which unbiased results could be recovered. CGG: Implemented CMB lensing in likelihood, data and covariances. ZG: Implemented and tested Fisher matrix for uncertainty estimation. BH: Generated mocks from AbacusSummit and contributed to some iteration of the likelihood. NK: Implemented the hybrid EFT code anzu into the CCL-based likelihood used to compare models and data in the challenge. AN: Co-designed the project, lead the analysis and writing of the paper. AS: Contributed to the design of the experiment, help with interpretation of results, contributed to the paper text.

CGG acknowledges support from the European Research Council Grant No: 693024 and the Beecroft Trust. DA acknowledges support from the Beecroft Trust, and from the John O'Connor Research Fund, at St. Peter's College, Oxford. ZG and CW were supported by the Office of Science of the U.S. Department of Energy, grant DE-SC0010007. JB is partially supported by NSF grant AST-2206563.

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3-Lyon/Villeurbanne — France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by U.K. BEIS National E-infrastructure capital grants; and the U.K. particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515.

The author(s) are pleased to acknowledge that the work reported on in this paper was substantially performed using the Princeton Research Computing resources at Princeton University which is consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing.

This work made use of the following software packages: `astropy`,⁴¹ `matplotlib`,⁴² `numpy`⁴³ and `scipy`.⁴⁴

A Description of methodology to create smoothed power spectra

To eliminate the noise in the measured power spectra due to the finite size of the ABACUS-SUMMIT simulation boxes, we start by modeling the overdensity of a given tracer x as

$$\delta_x(\mathbf{k}) = b_x(k) \delta_{\text{IC}}(\mathbf{k}) + n_x(\mathbf{k}), \quad (\text{A.1})$$

where δ_{IC} denotes the linear matter overdensity in the initial conditions, $b_x(k)$ is a deterministic function of k and, by definition, $n(\mathbf{k})$ is the small-scale component of the overdensity field, δ_x , that does not correlate with δ_{IC} .

Then at each snapshot, we compute the following power spectra from the simulation: $P_{gg}(k)$, $P_{gm}(k)$, $P_{mm}(k)$, $P_{g,\text{IC}}(k)$, $P_{m,\text{IC}}(k)$, and $P_{\text{IC},\text{IC}}(k)$, where m and g represent the overdensities of matter and of the target sample of galaxies. We also have a theoretical prediction for the linear power spectrum of the initial conditions, $\bar{P}_{\text{IC},\text{IC}}(k)$. From the model in eq. (A.1), we compute a first estimate of the bias functions $b_g(k)$ and $b_m(k)$ as

$$\hat{b}_x(k) \equiv \frac{P_{x,\text{IC}}(k)}{P_{\text{IC},\text{IC}}(k)}. \quad (\text{A.2})$$

⁴¹<https://www.astropy.org/>.

⁴²<https://matplotlib.org/>.

⁴³<https://numpy.org/>.

⁴⁴<https://scipy.org/>.

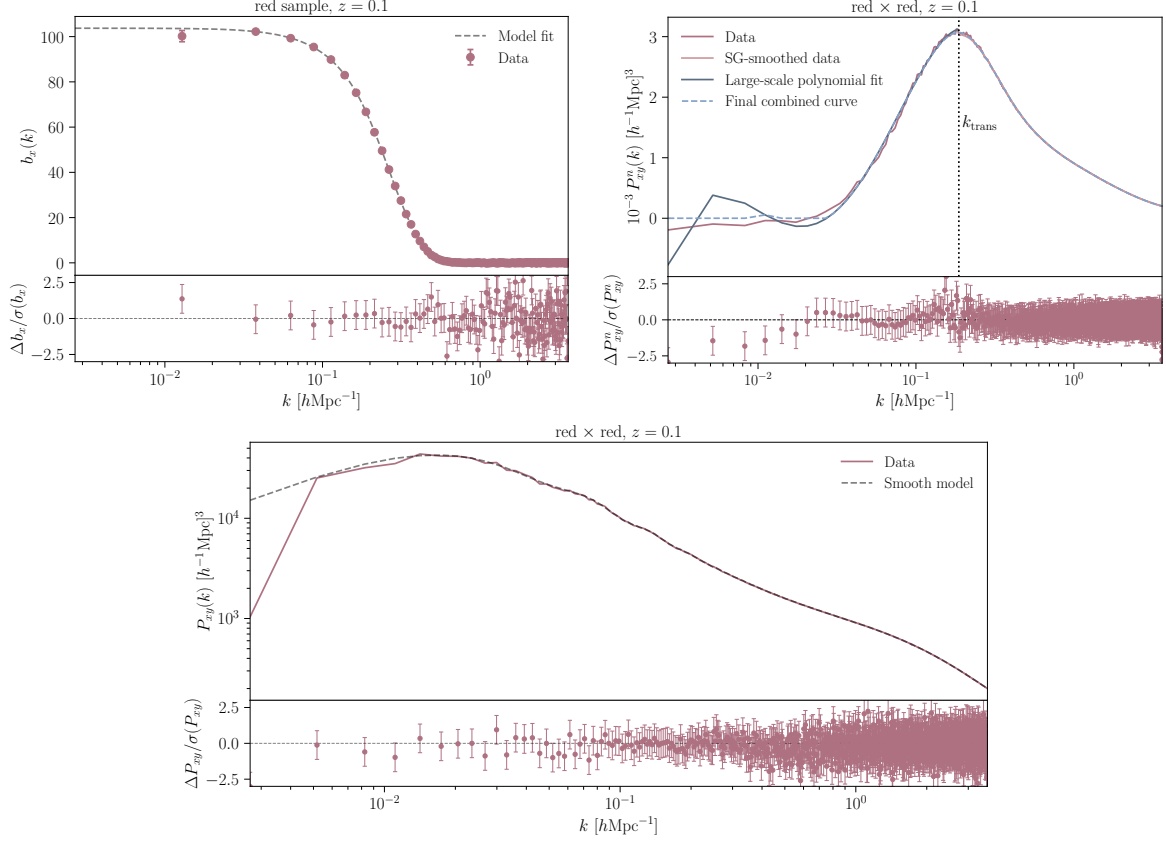


Figure 13. Steps involved in producing smooth power spectra from the measurements made on the ABACUSUMMIT simulation. Results shown for the red galaxy sample in the $z = 0.1$ snapshot. The top left panel shows the measurement of the large-scale bias function $b_x(k)$, together with its fit to eq. (A.2). The top right panel shows the measured power spectrum of the small-scale component, together with the different contributions to the final smooth fit (described in the text). The bottom panel shows the global fit to the galaxy power spectrum constructed by combining the results in the two top panels. The bottom sub-panels in each plot show the relative deviation between data and model fits as a fraction of the statistical uncertainties. The error bars were computed from the scatter between each group of 8 adjacent data points.

Since both $P_{x,\text{IC}}$ and $P_{\text{IC},\text{IC}}$ come from the same realization, the resulting bias function is reasonably smooth on large scales. To obtain a fully smooth function that is defined on the full continuum of k , we fit the resulting measured $\hat{b}_x$ to a smooth function of the form:

$$b_x(k) = b_0 e^{-(k/k_0)^\alpha} \left[1 + c e^{-\left(\frac{k-k_1}{0.1}\right)^2} \right], \quad (\text{A.3})$$

with b_0 , k_0 , α , c and k_1 as free parameters. We found this functional form to provide a good fit in all cases explored (see top left panel of figure 13).

After determining the b_x , we estimate the power spectrum between the small-scale components n_x and n_y as

$$P_{xy}^n(k) = P_{xy}(k) - \frac{P_{x,\text{IC}}(k)P_{y,\text{IC}}(k)}{P_{\text{IC},\text{IC}}(k)}. \quad (\text{A.4})$$

The resulting curve is fairly smooth on large k , but exhibits residual noise for small and intermediate wavenumbers ($k \lesssim 0.2 h \text{ Mpc}^{-1}$), which we correct for as follows (for illustration, see the top right panel of figure 13):

1. $P_{xy}^n(k)$ reaches a maximum at a transition scale k_{trans} around $\sim 0.1 h \text{ Mpc}^{-1}$, which provides a subdivision between the large-scale and small-scale components of $P_{xy}^n(k)$. In a first step, we determine this transition scale as the value at which the $P_{xy}^n(k)$ estimated from the simulation reaches its maximum value.
2. We fit the data on scales below k_{trans} using a 4th-order polynomial. Since this polynomial can take negative values, which are purely driven by large-scale noise, we apply a positivity prior on the resulting function. We note that on the largest scales, where this occurs, the final power spectrum is dominated by the large-scale correlated part from before, so these choices have a negligible impact on the final result.
3. On scales above k_{trans} , we find the measured spectra to exhibit small noise-like oscillations, which we reduce by smoothing the spectra with a Savitzky-Golay (SG) filter of order 1 and window size 25.
4. Finally, we combine the large-scale polynomial $P_{xy}^{n,L}(k)$ and the SG-smoothed small-scale component $P_{xy}^{n,S}(k)$ by smoothing the transition between the two regimes such as:

$$P_{xy}^n(k) = e^{-(k/k_{\text{trans}})^5} P_{xy}^{n,L}(k) + \left[1 - e^{-(k/k_{\text{trans}})^5}\right] P_{xy}^{n,S}(k). \quad (\text{A.5})$$

This procedure thus provides us with a smooth set of tabulated measurements of P_{xy}^n , which we then interpolate linearly in $\log(k)$. In a last step, we compute the final power spectrum from the theoretical linear power spectrum of the initial conditions combined with the models for $b_x(k)$ and $P_{xy}^n(k)$ as:

$$P_{xy}(k) = b_x(k)b_y(k)\bar{P}_{\text{IC,IC}}(k) + P_{xy}^n(k). \quad (\text{A.6})$$

To go beyond the smallest scale measured by ABACUSSUMMIT ($k_{\text{max}} \simeq 2.7 \text{ Mpc}^{-1}$) we use a power law extrapolation with a logarithmic slope calculated from the last 50 points in the measurement. As our results will be mostly based on scales $k \lesssim 0.4 \text{ Mpc}^{-1}$, the effects of this extrapolation are largely irrelevant.

Note that other approaches have been recently proposed in the literature [121, 122] to reduce the impact of noise on simulation-based measurements of summary statistics. We verified that the scheme described above is able to provide a good description of the measured power spectra, well below the statistical uncertainties of the ABACUSSUMMIT simulation box (see bottom panel of figure 13), and leave a comparison to other approaches to future work.

B Detailed description of goodness-of-fit tests

B.1 HEFT

In the following, we describe the goodness-of-fit tests employed to test the performance of the two HEFT implementations considered.

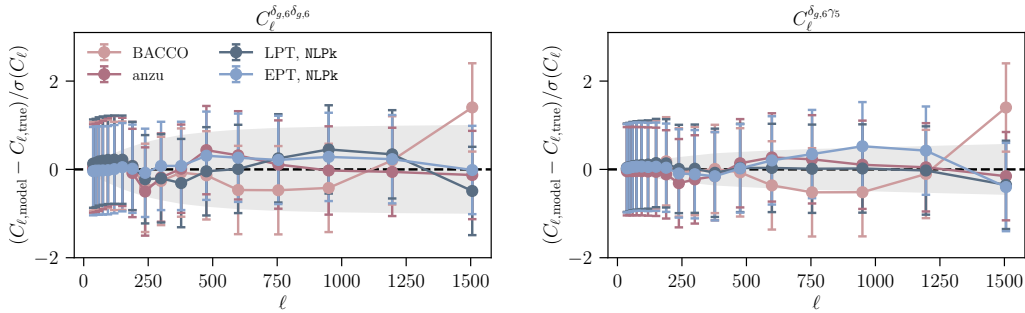


Figure 14. Normalized residuals between the ABACUSSUMMIT data and the predictions for the red galaxy sample obtained using the bias models considered in this work. *Left panel:* results for the auto-correlation of the highest clustering redshift bin. *Right panel:* results for the cross-correlation between the highest clustering bin and the highest weak lensing bin. In both figures, the gray area denotes relative deviation of 1% with respect to the theoretical prediction.

The number of degrees-of-freedom for our fiducial red sample is given by $\chi^2_{\text{noise}} = 478$, where we have used $n_{C_{\ell}} = 786$, $n_{C_{\ell}^{\gamma}} = 270$, and $n_p = 38$. Therefore, the maximal χ^2_{theory} allowed by our p -value criterion is given by $\chi^2_{\text{theory, max}} = 51$. The recovered χ^2 -values for both HEFT models pass this test, with $\chi^2_{\text{theory}} \simeq 11$ for **anzu**, and $\chi^2_{\text{theory}} \simeq 17$ for **BACCO**.⁴⁵ The associated p -values are $p = 0.36$ and $p = 0.28$, respectively. In figure 14, we also show the normalized fit residuals for the auto-correlation of the highest clustering redshift bin as well as its cross-correlation with the highest weak lensing bin.⁴⁶ As can be seen, we find most of the residuals to be within 1σ , and the relative differences generally lie within 1%.

Additionally, we consider two noisy realizations of the synthetic data, finding χ^2 p -values of $p = 0.096$ ($p = 0.63$) and $p = 0.012$ ($p = 0.11$) for **anzu** and **BACCO** respectively. The distributions of galaxy clustering and galaxy-galaxy lensing residuals are consistent with Gaussians, as can be seen in figure 15 for one of the realizations, and a Kolmogorov-Smirnov (KS) test yields p -values of $p = 0.38$ ($p = 0.93$) for **BACCO** and $p = 0.21$ ($p = 0.89$) for **anzu**. Given these results, we conclude that the two HEFT implementations we consider in this work are both suited to the analysis of high-precision data as expected from LSST, and we regard the failed p -value test for one of the noisy realizations for **BACCO** as a statistical fluctuation.

B.2 LPT/EPT

For the NLPk implementations of LPT and EPT, we obtain minimal χ^2 -values of $\chi^2_{\text{theory}} = 8.6$ for LPT, and $\chi^2_{\text{theory}} = 13.8$ for EPT when using the noiseless data vector. Furthermore, as can be seen from figure 14, we find the residuals between model and data to mostly lie within their 1σ uncertainties, with relative differences largely smaller than 1% as we found for the HEFT models. For the PTPk models on the other hand, we find χ^2 -values of $\chi^2_{\text{theory}} = 28.7$ for LPT, and $\chi^2_{\text{theory}} = 59.0$ for EPT at our minimal cutoff scale ($k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$).

⁴⁵To illustrate the effect of the systematic error floor, we note that χ^2_{theory} increases to $\chi^2_{\text{theory}} \simeq 16$ for **anzu** when we only consider statistical uncertainties.

⁴⁶We choose these particular two combinations as they display some of the largest differences between the simulated data and the models, and thus serve as illustration of the worst model performance.

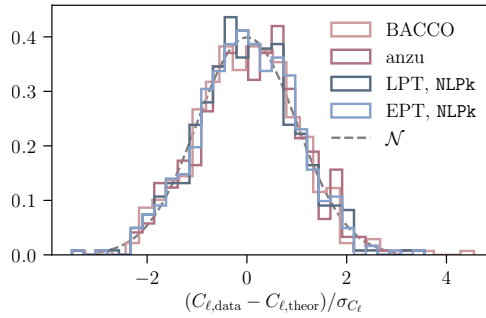


Figure 15. Distribution of galaxy clustering and galaxy-galaxy lensing C_ℓ residuals for a fit to noisified data, and using a maximal wavenumber $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$.

Finally, we also analyze the same two noisy realizations of the data vector as used in the HEFT case with our fiducial NLPk implementation of LPT and EPT, finding χ^2 p -values of $p = 0.16$ ($p = 0.77$) and $p = 0.064$ ($p = 0.83$) for LPT and EPT respectively. As shown in figure 15 for one of the realizations, we also find the distribution of fit residuals to be consistent with a Gaussian (with corresponding p -values of $p = 0.99$ for LPT and $p = 0.98$ for EPT).

C Consistency of bias parameter values from different models

In order to test for possible systematic modeling uncertainties, we compare the consistency of the bias parameter values obtained for the two HEFT implementations. As discussed in more detail in appendix D below, we find the Fisher matrix uncertainties on the bias parameters to be rather unstable when varying both b_{s2} and b_{∇^2} . The constraints obtained for the cosmological parameters on the other hand, are stable in all cases considered. Fixing one of the two parameters, b_{s2} or b_{∇^2} , yields stable error bars in all cases, which are additionally consistent with their MCMC analogs (see appendix D). We therefore compare the recovered bias parameters for **anzu** and **BACCO** fixing b_{s2} to its recovered value, and the results are shown in figure 16 for all parameters except b_{s2} .⁴⁷ As can be seen, these are generally consistent, the only exception being b_{∇^2} for which we find consistently higher values for **BACCO** than we do for **anzu**. These differences might be due to the fact that the nonlocal bias parameter b_{∇^2} depends on small-scale properties of the fields considered and is thus affected by implementation details such as different smoothing scales used when deriving template spectra (see e.g. ref. [9]). In addition, **anzu** and **BACCO** employ different methods to model power spectra involving higher derivative terms: while in **BACCO** these terms are determined from the simulations, **anzu** uses the approximation $\langle X, \nabla^2 \delta_L \rangle = -k^2 \langle X, 1 \rangle$, where X denotes one of the fields described in section 2.

We find similar results when comparing the bias parameter values obtained using our fiducial NLPk implementation of LPT and EPT. As can be seen from figure 16, we generally find good agreement between both methods, although the EPT approach seems to prefer significantly larger values for b_{∇^2} and P_{SN} compared to LPT. It is also interesting to note that

⁴⁷We find the recovered values for b_{s2} to be generally consistent between all four models considered here, and thus do not show the plots as we are unable to provide error bars for the measurements.

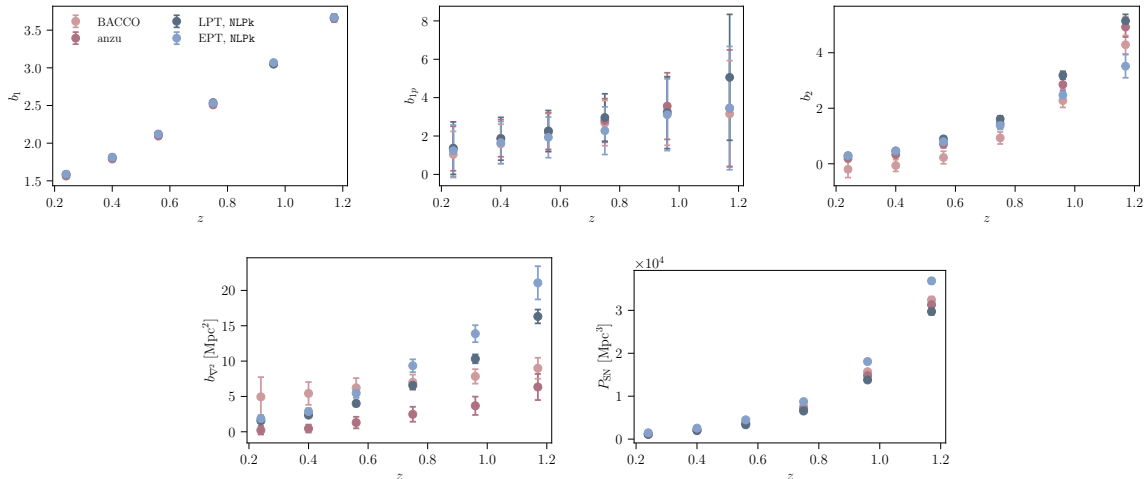


Figure 16. Comparison of bias parameter values obtained from fitting our fiducial red galaxy sample with all nonlinear bias models considered in this work.

the recovered bias parameters for the two PT-based models are generally consistent with those obtained from the HEFT methods. As above, the notable exception is b_{∇^2} , for which we obtain significantly lower values for the HEFT implementations than we do for the perturbation-theory-based models, particularly at high redshift. We see two possible explanations for these findings: (i) As discussed above, our results could be another consequence of different bias parameter normalizations due to implementation details such as smoothing. (ii) Nonlocal bias terms parameterized by b_{∇^2} have the same functional form as EFT counter-terms, which absorb corrections to the bias model due to small-scale physics. These corrections are partly accounted for in the HEFT approach, while they are not present in PT-based models. Without providing rigorous confirmation, these results suggest that the larger nonlocal bias values obtained for LPT and EPT are due to larger contributions from EFT counterterms for these models (see ref. [19] for similar results).

D Comparison to MCMC results

Fisher matrix analyses are prone to numerical instabilities (see e.g. refs. [109–111]), and it is therefore essential to validate our results by comparing our FM constraints to those derived using a Monte Carlo Markov Chain (MCMC). Here we focus on our fiducial case using **anzu** with $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$, and perform two separate MCMC analyses using the same specifications as used for the Fisher matrix computation: in our first analysis we allow for variations in all cosmological and bias parameters, while in the second case we fix $b_{s2} = 0$. The comparison of the MCMC constraints on cosmological parameters obtained in the first case and their FM counterparts are shown in the left hand panel of figure 17. As can be seen, we find the two approaches to yield consistent constraints on the $\Omega_c - \sigma_8$ plane. This is not the case for the constraints on bias parameters as discussed above and we thus repeat this analysis setting $b_{s2} = 0$ in both cases. The ratio of the $1-\sigma$ uncertainties obtained from the FM and MCMC analyses respectively are shown in the right hand panel of figure 17 for

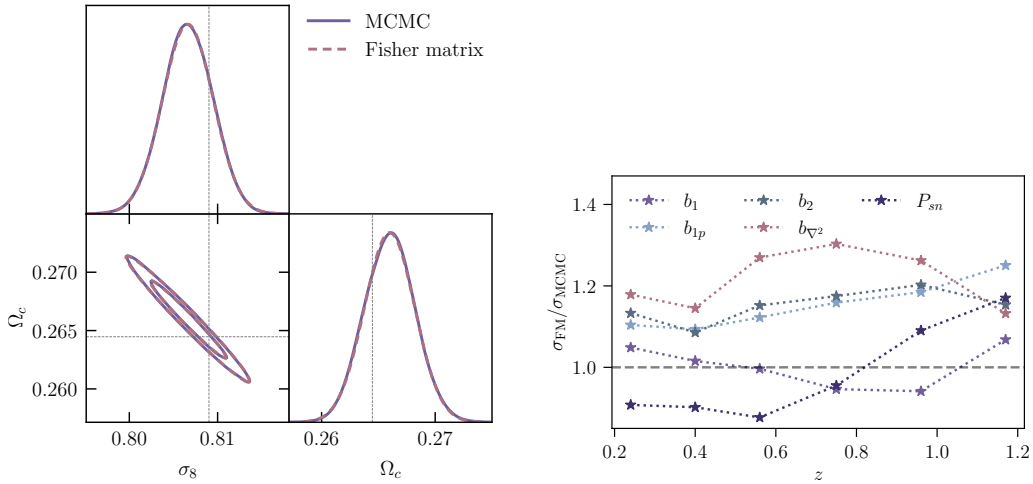


Figure 17. Comparison of parameter constraints obtained from **anzu** for $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ using an MCMC or a Fisher matrix approximation. *Left panel:* constraints on Ω_c and σ_8 allowing all parameters to vary. *Right panel:* ratio of FM and MCMC 1- σ uncertainties on bias parameters obtained when setting $b_{s^2} = 0$.

all bias parameters considered. Similarly to before, we find those to be consistent within roughly 30%. This is a reasonable level of agreement, given the approximations and numerical instabilities involved in Fisher matrix analyses.

E Fitting data from UNIT

As an additional consistency test of the results presented in sections 6.1 and 6.2, we repeat our analysis using the **redmagic** sample from the UNIT simulations [113] employed in ref. [31]. Specifically, we work with the three-dimensional power spectrum and fit the combination of $\mathbf{d} = \{P_{gg}(z, k), P_{gm}(z, k)\}$ using the covariance matrix derived in ref. [31]. The results for all bias models considered in this work are shown in figure 18. As can be seen, we find results very similar to those obtained for ABACUSSUMMIT data, i.e. both HEFT and NLPk implementations of EPT and LPT yield unbiased constraints on σ_8 and Ω_c , while the PTPk methods give rise to biases in the recovered cosmological parameters for large maximal wavenumbers k_{max} . Compared to the results shown in figure 3, we find the PTPk models to break down at smaller wavenumbers. This is particularly true for EPT, which yields significantly biased constraints on cosmological parameters for $k_{\text{max}} \gtrsim 0.2 \text{ Mpc}^{-1}$. These results suggest that spherical harmonic power spectra might indeed be less susceptible to nonlinear bias modeling systematics than their three-dimensional counterparts, and provide confirmation that the findings reported in sections 6.1 and 6.2 are not driven by our usage of data from the ABACUSSUMMIT simulations.

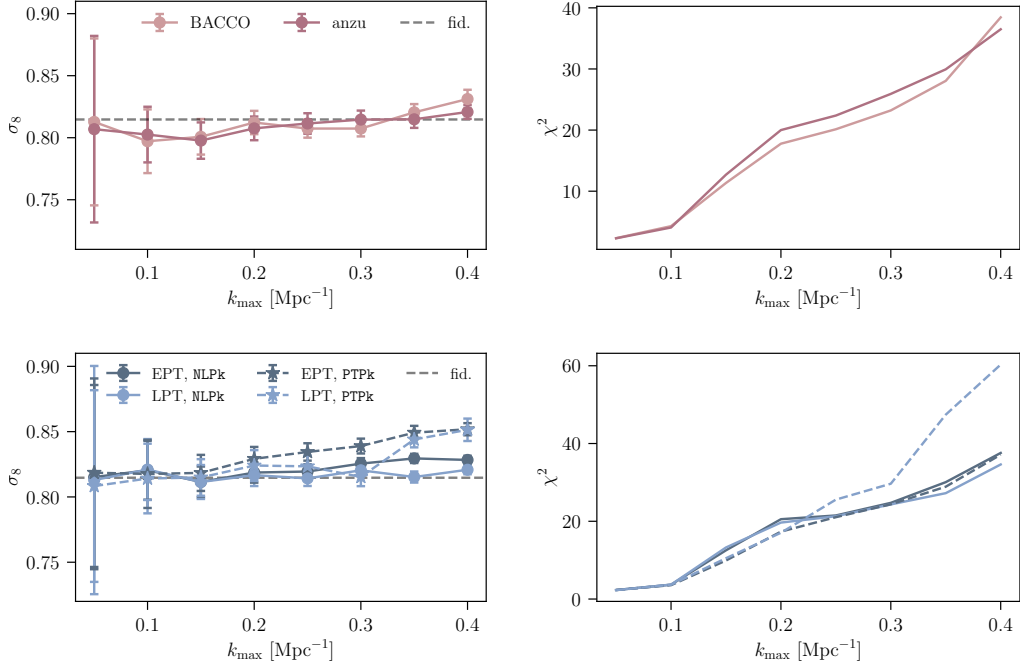


Figure 18. Results of fitting the UNIT data with the main nonlinear bias models considered in this work, LPT, EPT, **anzu** and **BACCO**. *Top panels:* recovered values of σ_8 and Ω_c as a function of maximal wavenumber, $k_{\max}$ for the two HEFT implementations. *Bottom panels:* recovered values of σ_8 and Ω_c as a function of maximal wavenumber, $k_{\max}$ for NLPk and PTPk implementations of LPT and EPT.

F Three-dimensional power spectrum analysis

In order to investigate stochasticity in the red and maglim galaxy samples, we analyze three-dimensional power spectrum data. Specifically, we use power spectra at six discrete redshifts covered by the ABACUSUMMIT simulations, i.e. $z = \{0.1, 0.3, 0.5, 0.8, 1.1, 1.4\}$. We consider a data vector $\mathbf{d} = \{P_{gg}(z, k), P_{gm}(z, k), P_{mm}(z, k)\}$ and assume a Gaussian likelihood with Gaussian covariance matrix given by (see e.g. [31, 123])

$$\text{Cov}(P_{ij}(k), P_{ln}(k')) = \kappa \frac{2\pi^2 \delta_{kk'}}{k^2 \Delta k V} \begin{cases} 2P_{ii}^2(k), & \text{if } i = j = l = n \\ 2P_{gg}(k)P_{gm}(k), & \text{if } i, j = g \text{ and } l = g, n = m \\ [P_{gg}(k)P_{mm}(k) + P_{gm}^2(k)], & \text{if } i, l = g, \text{ and } j, n = m, \end{cases} \quad (\text{F.1})$$

where V denotes the volume of the survey and Δk is the width of each k -bin. For consistency with our previous analysis, we aim to ensure a similar signal-to-noise ratio for the $P(k)$ measurements. We therefore follow a rather crude approach and introduce a scaling factor κ that scales the relative $P(k)$ errors to be equal to those for C_ℓ data obtained for bins with matching mean redshift.⁴⁸ We then use this likelihood and fit the data with the

⁴⁸We note that we repeat this analysis without the scaling factor and recover very similar values for the stochasticity.

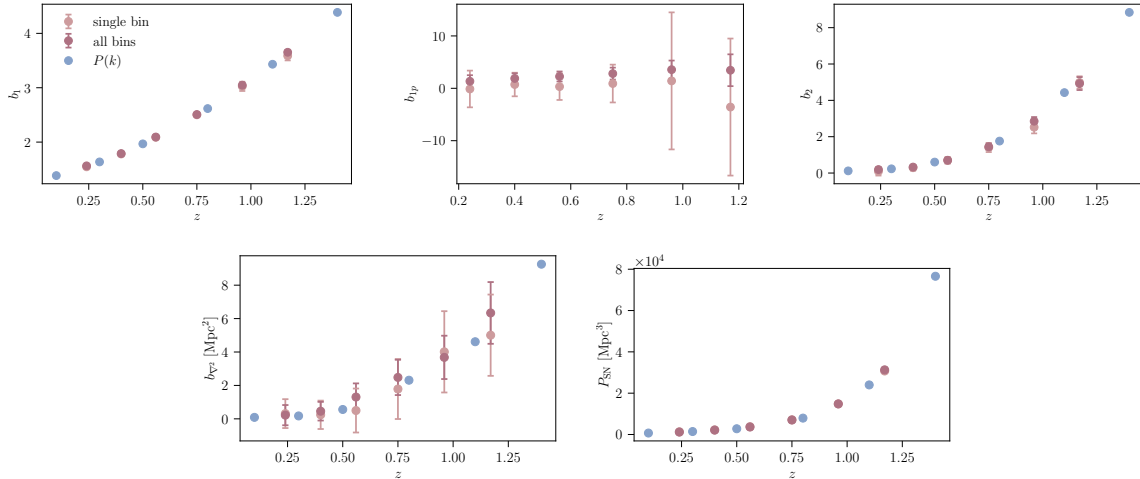


Figure 19. Comparison of bias parameters obtained by fitting three different data sets using `anzu`: (i) fit each redshift bin separately, (ii) fit all redshift bins simultaneously, (iii) fit three-dimensional power spectra instead of spherical harmonic power spectra.

`anzu` implementation of HEFT, keeping the cosmological parameters fixed at their fiducial ABACUSSUMMIT values.⁴⁹

G Consistency of results for different data sets

Extending the analysis described in section 6.5 and as a further consistency test of our results, we compare the bias values obtained using different combinations of our simulated data. Specifically, we compare our fiducial constraints obtained using spherical harmonics and fitting all redshift bins simultaneously to those obtained by fitting each bin separately both using spherical harmonics and three-dimensional power spectra. For the $P(k)$ data we proceed as described in section 6.5, while for the single-bin spherical harmonic fits, we choose to only fit the galaxy auto-correlation and the cross-correlation between the galaxies and DM while keeping the cosmological parameters fixed at their fiducial values as above. As discussed in section 6.1, we find that while our fiducial analysis fits for all bias parameters simultaneously, the Fisher matrix constraints obtained in this setup are numerically unstable. In order to compare the constraints obtained from different data sets, we therefore follow the approach described in section 6.1, and consider the FM errors obtained at the best-fit values when setting $b_{s,2} = 0$. In addition, given the ad-hoc procedure to rescale the covariance matrix used for analyzing $P(k)$ data, we do not consider any error bars for these measurements. The results obtained for `anzu` are shown in figure 19, and as can be seen we find a very good agreement between the bias values derived in all cases. This suggests that the fits to the spherical harmonic power spectra show internal consistency, and that the redshift-averaged bias parameters obtained from these are consistent with the single-redshift fits obtained from the $P(k)$ analysis.

⁴⁹We find this to be necessary, as the single bin fits exhibit a strong degeneracy between cosmological and bias parameters, and thus the minimization and error bars become unstable when fitting all parameters jointly.

References

- [1] M. Lattanzi and M. Gerbino, *Status of neutrino properties and future prospects — cosmological and astrophysical constraints*, *Front. in Phys.* **5** (2018) 70 [[arXiv:1712.07109](#)] [[INSPIRE](#)].
- [2] N. Dalal, O. Doré, D. Huterer and A. Shirokov, *The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev. D* **77** (2008) 123514 [[arXiv:0710.4560](#)] [[INSPIRE](#)].
- [3] J.J.M. Carrasco, S. Foreman, D. Green and L. Senatore, *The effective field theory of large scale structures at two loops*, *JCAP* **07** (2014) 057 [[arXiv:1310.0464](#)] [[INSPIRE](#)].
- [4] VIRGO CONSORTIUM collaboration, *Stable clustering, the halo model and nonlinear cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1311 [[astro-ph/0207664](#)] [[INSPIRE](#)].
- [5] EUCLID collaboration, *Euclid preparation: II. The EuclidEmulator — a tool to compute the cosmology dependence of the nonlinear matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 5509 [[arXiv:1809.04695](#)] [[INSPIRE](#)].
- [6] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [[arXiv:1707.03397](#)] [[INSPIRE](#)].
- [7] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [[astro-ph/0112551](#)] [[INSPIRE](#)].
- [8] P. McDonald and A. Roy, *Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS*, *JCAP* **08** (2009) 020 [[arXiv:0902.0991](#)] [[INSPIRE](#)].
- [9] V. Desjacques, D. Jeong and F. Schmidt, *Large-scale galaxy bias*, *Phys. Rept.* **733** (2018) 1 [[arXiv:1611.09787](#)] [[INSPIRE](#)].
- [10] T. Matsubara, *Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture*, *Phys. Rev. D* **78** (2008) 083519 [Erratum *ibid.* **78** (2008) 109901] [[arXiv:0807.1733](#)] [[INSPIRE](#)].
- [11] J. Carlson, B. Reid and M. White, *Convolution Lagrangian perturbation theory for biased tracers*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 1674 [[arXiv:1209.0780](#)] [[INSPIRE](#)].
- [12] Z. Vlah, E. Castorina and M. White, *The Gaussian streaming model and convolution Lagrangian effective field theory*, *JCAP* **12** (2016) 007 [[arXiv:1609.02908](#)] [[INSPIRE](#)].
- [13] J.J.M. Carrasco, M.P. Hertzberg and L. Senatore, *The effective field theory of cosmological large scale structures*, *JHEP* **09** (2012) 082 [[arXiv:1206.2926](#)] [[INSPIRE](#)].
- [14] D. Baumann, A. Nicolis, L. Senatore and M. Zaldarriaga, *Cosmological non-linearities as an effective fluid*, *JCAP* **07** (2012) 051 [[arXiv:1004.2488](#)] [[INSPIRE](#)].
- [15] E. Munari et al., *Improving fast generation of halo catalogues with higher order Lagrangian perturbation theory*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 4658 [[arXiv:1605.04788](#)] [[INSPIRE](#)].
- [16] C. Modi, S.-F. Chen and M. White, *Simulations and symmetries*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 5754 [[arXiv:1910.07097](#)] [[INSPIRE](#)].
- [17] A. Eggemeier et al., *Testing one-loop galaxy bias: power spectrum*, *Phys. Rev. D* **102** (2020) 103530 [[arXiv:2006.09729](#)] [[INSPIRE](#)].

- [18] L. Fonseca de la Bella, D. Regan, D. Seery and D. Parkinson, *Impact of bias and redshift-space modelling for the halo power spectrum: testing the effective field theory of large-scale structure*, *JCAP* **07** (2020) 011 [[arXiv:1805.12394](#)] [[INSPIRE](#)].
- [19] DES collaboration, *Perturbation theory for modeling galaxy bias: validation with simulations of the Dark Energy Survey*, *Phys. Rev. D* **102** (2020) 123522 [[arXiv:2008.05991](#)] [[INSPIRE](#)].
- [20] LSST DARK ENERGY SCIENCE collaboration, *Perturbation theory models for LSST-era galaxy clustering: tests with subpercent mock catalog measurements in Fourier and configuration space*, *Phys. Rev. D* **105** (2022) 123518 [[arXiv:2111.00501](#)] [[INSPIRE](#)].
- [21] C.-P. Ma and J.N. Fry, *Deriving the nonlinear cosmological power spectrum and bispectrum from analytic dark matter halo profiles and mass functions*, *Astrophys. J.* **543** (2000) 503 [[astro-ph/0003343](#)] [[INSPIRE](#)].
- [22] J.A. Peacock and R.E. Smith, *Halo occupation numbers and galaxy bias*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 1144 [[astro-ph/0005010](#)] [[INSPIRE](#)].
- [23] U. Seljak, *Analytic model for galaxy and dark matter clustering*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 203 [[astro-ph/0001493](#)] [[INSPIRE](#)].
- [24] A. Mead et al., *An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 1958 [[arXiv:1505.07833](#)] [[INSPIRE](#)].
- [25] J. Chaves-Montero, R.E. Angulo and S. Contreras, *The galaxy formation origin of the lensing is low problem*, *Mon. Not. Roy. Astron. Soc.* **521** (2023) 937 [[arXiv:2211.01744](#)] [[INSPIRE](#)].
- [26] BOSS collaboration, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier-space*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 2242 [[arXiv:1607.03150](#)] [[INSPIRE](#)].
- [27] G. D’Amico et al., *The cosmological analysis of the SDSS/BOSS data from the effective field theory of large-scale structure*, *JCAP* **05** (2020) 005 [[arXiv:1909.05271](#)] [[INSPIRE](#)].
- [28] M.M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological parameters from the BOSS galaxy power spectrum*, *JCAP* **05** (2020) 042 [[arXiv:1909.05277](#)] [[INSPIRE](#)].
- [29] M. White et al., *Cosmological constraints from the tomographic cross-correlation of DESI luminous red galaxies and Planck CMB lensing*, *JCAP* **02** (2022) 007 [[arXiv:2111.09898](#)] [[INSPIRE](#)].
- [30] R. Zhou et al., *The clustering of DESI-like luminous red galaxies using photometric redshifts*, *Mon. Not. Roy. Astron. Soc.* **501** (2021) 3309 [[arXiv:2001.06018](#)] [[INSPIRE](#)].
- [31] N. Kokron et al., *The cosmology dependence of galaxy clustering and lensing from a hybrid N-body-perturbation theory model*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 1422 [[arXiv:2101.11014](#)] [[INSPIRE](#)].
- [32] M. Zennaro et al., *The BACCO simulation project: biased tracers in real space*, *Mon. Not. Roy. Astron. Soc.* **524** (2023) 2407 [[arXiv:2101.12187](#)] [[INSPIRE](#)].
- [33] A.P. Hearin, N. Ramachandra, M.R. Becker and J. DeRose, *Differentiable predictions for large scale structure with SHAMNet*, *Open J. Astrophys.* **5** (2022) 3 [[arXiv:2112.08423](#)] [[INSPIRE](#)].
- [34] A. Eggemeier et al., *COMET: clustering observables modelled by emulated perturbation theory*, *Mon. Not. Roy. Astron. Soc.* **519** (2022) 2962 [[arXiv:2208.01070](#)] [[INSPIRE](#)].
- [35] J. DeRose et al., *Aemulus ν : precise predictions for matter and biased tracer power spectra in the presence of neutrinos*, *JCAP* **07** (2023) 054 [[arXiv:2303.09762](#)] [[INSPIRE](#)].

- [36] M. Pellejero-Ibanez et al., *Modelling galaxy clustering in redshift space with a Lagrangian bias formalism and N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 3993 [[arXiv:2109.08699](#)] [[INSPIRE](#)].
- [37] M. Pellejero-Ibanez et al., *The BACCO simulation project: BACCO hybrid Lagrangian bias expansion model in redshift space*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 3725 [[arXiv:2207.06437](#)] [[INSPIRE](#)].
- [38] LSST DARK ENERGY SCIENCE collaboration, *The LSST Dark Energy Science Collaboration (DESC) science requirements document*, [arXiv:1809.01669](#) [[INSPIRE](#)].
- [39] C. Modi, M. White and Z. Vlah, *Modeling CMB lensing cross correlations with CLEFT*, *JCAP* **08** (2017) 009 [[arXiv:1706.03173](#)] [[INSPIRE](#)].
- [40] E. Krause and T. Eifler, *cosmolike — cosmological likelihood analyses for photometric galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **470** (2017) 2100 [[arXiv:1601.05779](#)] [[INSPIRE](#)].
- [41] Z. Zhai et al., *The Aemulus project III: emulation of the galaxy correlation function*, *Astrophys. J.* **874** (2019) 95 [[arXiv:1804.05867](#)] [[INSPIRE](#)].
- [42] J.U. Lange et al., *Five per cent measurements of the growth rate from simulation-based modelling of redshift-space clustering in BOSS LOWZ*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 1779 [[arXiv:2101.12261](#)] [[INSPIRE](#)].
- [43] S.-F. Chen, Z. Vlah and M. White, *A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO*, *JCAP* **02** (2022) 008 [[arXiv:2110.05530](#)] [[INSPIRE](#)].
- [44] T. Lazeyras and F. Schmidt, *A robust measurement of the first higher-derivative bias of dark matter halos*, *JCAP* **11** (2019) 041 [[arXiv:1904.11294](#)] [[INSPIRE](#)].
- [45] T. Matsubara, *Stochasticity of bias and nonlocality of galaxy formation: linear scales*, *Astrophys. J.* **525** (1999) 543 [[astro-ph/9906029](#)] [[INSPIRE](#)].
- [46] A. Taruya and J. Soda, *Stochastic biasing and galaxy mass density relation in the weakly nonlinear regime*, *Astrophys. J.* **522** (1999) 46 [[astro-ph/9809204](#)] [[INSPIRE](#)].
- [47] A. Dekel and O. Lahav, *Stochastic nonlinear galaxy biasing*, *Astrophys. J.* **520** (1999) 24 [[astro-ph/9806193](#)] [[INSPIRE](#)].
- [48] J.N. Fry, *The evolution of bias*, *Astrophys. J. Lett.* **461** (1996) L65 [[INSPIRE](#)].
- [49] T. Baldauf, U. Seljak, V. Desjacques and P. McDonald, *Evidence for quadratic tidal tensor bias from the halo bispectrum*, *Phys. Rev. D* **86** (2012) 083540 [[arXiv:1201.4827](#)] [[INSPIRE](#)].
- [50] K.C. Chan, R. Scoccimarro and R.K. Sheth, *Gravity and large-scale non-local bias*, *Phys. Rev. D* **85** (2012) 083509 [[arXiv:1201.3614](#)] [[INSPIRE](#)].
- [51] T. Matsubara, *Integrated perturbation theory and one-loop power spectra of biased tracers*, *Phys. Rev. D* **90** (2014) 043537 [[arXiv:1304.4226](#)] [[INSPIRE](#)].
- [52] S. Saito et al., *Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum*, *Phys. Rev. D* **90** (2014) 123522 [[arXiv:1405.1447](#)] [[INSPIRE](#)].
- [53] M.M. Abidi and T. Baldauf, *Cubic halo bias in Eulerian and Lagrangian space*, *JCAP* **07** (2018) 029 [[arXiv:1802.07622](#)] [[INSPIRE](#)].
- [54] T. Baldauf et al., *Halo stochasticity from exclusion and nonlinear clustering*, *Phys. Rev. D* **88** (2013) 083507 [[arXiv:1305.2917](#)] [[INSPIRE](#)].

- [55] N. Kokron et al., *Priors on red galaxy stochasticity from hybrid effective field theory*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 2198 [[arXiv:2112.00012](#)] [[INSPIRE](#)].
- [56] J.E. McEwen, X. Fang, C.M. Hirata and J.A. Blazek, *FAST-PT: a novel algorithm to calculate convolution integrals in cosmological perturbation theory*, *JCAP* **09** (2016) 015 [[arXiv:1603.04826](#)] [[INSPIRE](#)].
- [57] X. Fang, J.A. Blazek, J.E. McEwen and C.M. Hirata, *FAST-PT II: an algorithm to calculate convolution integrals of general tensor quantities in cosmological perturbation theory*, *JCAP* **02** (2017) 030 [[arXiv:1609.05978](#)] [[INSPIRE](#)].
- [58] L. Senatore, *Bias in the effective field theory of large scale structures*, *JCAP* **11** (2015) 007 [[arXiv:1406.7843](#)] [[INSPIRE](#)].
- [59] R.A. Porto, L. Senatore and M. Zaldarriaga, *The Lagrangian-space effective field theory of large scale structures*, *JCAP* **05** (2014) 022 [[arXiv:1311.2168](#)] [[INSPIRE](#)].
- [60] T. Matsubara, *Resumming cosmological perturbations via the Lagrangian picture: one-loop results in real space and in redshift space*, *Phys. Rev. D* **77** (2008) 063530 [[arXiv:0711.2521](#)] [[INSPIRE](#)].
- [61] S.-F. Chen, Z. Vlah and M. White, *Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory*, *JCAP* **07** (2020) 062 [[arXiv:2005.00523](#)] [[INSPIRE](#)].
- [62] SDSS collaboration, *Galaxy clustering in the completed SDSS redshift survey: the dependence on color and luminosity*, *Astrophys. J.* **736** (2011) 59 [[arXiv:1005.2413](#)] [[INSPIRE](#)].
- [63] B. Hadzhiyska et al., *Hefty enhancement of cosmological constraints from the DES Y1 data using a hybrid effective field theory approach to galaxy bias*, *JCAP* **09** (2021) 020 [[arXiv:2103.09820](#)] [[INSPIRE](#)].
- [64] N. Wiener, *The homogeneous chaos*, *Amer. J. Math.* **60** (1938) 897.
- [65] R.E. Angulo et al., *The BACCO simulation project: exploiting the full power of large-scale structure for cosmology*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 5869 [[arXiv:2004.06245](#)] [[INSPIRE](#)].
- [66] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, *Precision measurement of the local bias of dark matter halos*, *JCAP* **02** (2016) 018 [[arXiv:1511.01096](#)] [[INSPIRE](#)].
- [67] T. Lazeyras and F. Schmidt, *Beyond LIMD bias: a measurement of the complete set of third-order halo bias parameters*, *JCAP* **09** (2018) 008 [[arXiv:1712.07531](#)] [[INSPIRE](#)].
- [68] A. Barreira, T. Lazeyras and F. Schmidt, *Galaxy bias from forward models: linear and second-order bias of IllustrisTNG galaxies*, *JCAP* **08** (2021) 029 [[arXiv:2105.02876](#)] [[INSPIRE](#)].
- [69] M. Zennaro et al., *Priors on Lagrangian bias parameters from galaxy formation modelling*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 5443 [[arXiv:2110.05408](#)] [[INSPIRE](#)].
- [70] R. Takahashi et al., *Revising the halofit model for the nonlinear matter power spectrum*, *Astrophys. J.* **761** (2012) 152 [[arXiv:1208.2701](#)] [[INSPIRE](#)].
- [71] DES collaboration, *Dark Energy Survey year 3 results: constraints on cosmological parameters and galaxy-bias models from galaxy clustering and galaxy-galaxy lensing using the redMaGiC sample*, *Phys. Rev. D* **106** (2022) 043520 [[arXiv:2105.13545](#)] [[INSPIRE](#)].
- [72] A. Krolewski, S. Ferraro and M. White, *Cosmological constraints from unWISE and Planck CMB lensing tomography*, *JCAP* **12** (2021) 028 [[arXiv:2105.03421](#)] [[INSPIRE](#)].

- [73] X. Fang, E. Krause, T. Eifler and N. MacCrann, *Beyond Limber: efficient computation of angular power spectra for galaxy clustering and weak lensing*, *JCAP* **05** (2020) 010 [[arXiv:1911.11947](#)] [[INSPIRE](#)].
- [74] D.N. Limber, *The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field*, *Astrophys. J.* **117** (1953) 134.
- [75] N. Kaiser, *Weak gravitational lensing of distant galaxies*, *Astrophys. J.* **388** (1992) 272 [[INSPIRE](#)].
- [76] N. Kaiser, *Weak lensing and cosmology*, *Astrophys. J.* **498** (1998) 26 [[astro-ph/9610120](#)] [[INSPIRE](#)].
- [77] N.A. Maksimova et al., *AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **508** (2021) 4017 [[arXiv:2110.11398](#)] [[INSPIRE](#)].
- [78] L.H. Garrison, D.J. Eisenstein and P.A. Pinto, *A high-fidelity realization of the Euclid code comparison N-body simulation with Abacus*, *Mon. Not. Roy. Astron. Soc.* **485** (2019) 3370 [[arXiv:1810.02916](#)] [[INSPIRE](#)].
- [79] L.H. Garrison et al., *The Abacus cosmological N-body code*, *Mon. Not. Roy. Astron. Soc.* **508** (2021) 575 [[arXiv:2110.11392](#)] [[INSPIRE](#)].
- [80] B. Hadzhiyska et al., *compaso: a new halo finder for competitive assignment to spherical overdensities*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 501 [[arXiv:2110.11408](#)] [[INSPIRE](#)].
- [81] S. Yuan et al., *Evidence for galaxy assembly bias in BOSS CMASS redshift-space galaxy correlation function*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 3582 [[arXiv:2010.04182](#)] [[INSPIRE](#)].
- [82] Z. Zheng, A.L. Coil and I. Zehavi, *Galaxy evolution from halo occupation distribution modeling of DEEP2 and SDSS galaxy clustering*, *Astrophys. J.* **667** (2007) 760 [[astro-ph/0703457](#)] [[INSPIRE](#)].
- [83] DES collaboration, *Dark Energy Survey year 3 results: galaxy clustering and systematics treatment for lens galaxy samples*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2665 [[arXiv:2105.13540](#)] [[INSPIRE](#)].
- [84] LSST collaboration, *Tomographic galaxy clustering with the Subaru Hyper Suprime-Cam first year public data release*, *JCAP* **03** (2020) 044 [[arXiv:1912.08209](#)] [[INSPIRE](#)].
- [85] S. Ando, A. Benoit-Lévy and E. Komatsu, *Angular power spectrum of galaxies in the 2MASS redshift survey*, *Mon. Not. Roy. Astron. Soc.* **473** (2018) 4318 [[arXiv:1706.05422](#)] [[INSPIRE](#)].
- [86] N. Koukoufilippas, D. Alonso, M. Bilicki and J.A. Peacock, *Tomographic measurement of the intergalactic gas pressure through galaxy — tSZ cross-correlations*, *Mon. Not. Roy. Astron. Soc.* **491** (2020) 5464 [[arXiv:1909.09102](#)] [[INSPIRE](#)].
- [87] R.H. Wechsler et al., *Concentrations of dark halos from their assembly histories*, *Astrophys. J.* **568** (2002) 52 [[astro-ph/0108151](#)] [[INSPIRE](#)].
- [88] L. Gao, V. Springel and S.D.M. White, *The age dependence of halo clustering*, *Mon. Not. Roy. Astron. Soc.* **363** (2005) L66 [[astro-ph/0506510](#)] [[INSPIRE](#)].
- [89] R.H. Wechsler, A.R. Zentner, J.S. Bullock and A.V. Kravtsov, *The dependence of halo clustering on halo formation history, concentration, and occupation*, *Astrophys. J.* **652** (2006) 71 [[astro-ph/0512416](#)] [[INSPIRE](#)].
- [90] N. Dalal, M. White, J.R. Bond and A. Shirokov, *Halo assembly bias in hierarchical structure formation*, *Astrophys. J.* **687** (2008) 12 [[arXiv:0803.3453](#)] [[INSPIRE](#)].

- [91] S. Yuan et al., *AbacusHOD: a highly efficient extended multitracer HOD framework and its application to BOSS and eBOSS data*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3301 [[arXiv:2110.11412](#)] [[INSPIRE](#)].
- [92] B. Hadzhiyska et al., *Limitations to the ‘basic’ HOD model and beyond*, *Mon. Not. Roy. Astron. Soc.* **493** (2020) 5506 [[arXiv:1911.02610](#)] [[INSPIRE](#)].
- [93] B. Reid et al., *SDSS-III Baryon Oscillation Spectroscopic Survey data release 12: galaxy target selection and large scale structure catalogues*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 1553 [[arXiv:1509.06529](#)] [[INSPIRE](#)].
- [94] Z. Zheng et al., *Theoretical models of the halo occupation distribution: separating central and satellite galaxies*, *Astrophys. J.* **633** (2005) 791 [[astro-ph/0408564](#)] [[INSPIRE](#)].
- [95] D. Alonso et al., *Ultra large-scale cosmology in next-generation experiments with single tracers*, *Astrophys. J.* **814** (2015) 145 [[arXiv:1505.07596](#)] [[INSPIRE](#)].
- [96] LSST DARK ENERGY SCIENCE collaboration, *Core cosmology library: precision cosmological predictions for LSST*, *Astrophys. J. Suppl.* **242** (2019) 2 [[arXiv:1812.05995](#)] [[INSPIRE](#)].
- [97] B. Hadzhiyska et al., *Synthetic light cone catalogues of modern redshift and weak lensing surveys with AbacusSummit*, [arXiv:2305.11935](#) [[INSPIRE](#)].
- [98] A. Mead, S. Brieden, T. Tröster and C. Heymans, *HMcode-2020: improved modelling of non-linear cosmological power spectra with baryonic feedback*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 1401 [[arXiv:2009.01858](#)] [[INSPIRE](#)].
- [99] C. Chang et al., *The effective number density of galaxies for weak lensing measurements in the LSST project*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) 2121 [[arXiv:1305.0793](#)] [[INSPIRE](#)].
- [100] B. Hadzhiyska et al., *Cosmology with 6 parameters in the stage-IV era: efficient marginalisation over nuisance parameters*, *Open J. Astrophys.* **6** (2023) [[arXiv:2301.11895](#)] [[INSPIRE](#)].
- [101] M. Powell, *The BOBYQA algorithm for bound constrained optimization without derivatives*, technical report, https://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, U.K. (2009).
- [102] C. Cartis, J. Fiala, B. Marteau and L. Roberts, *Improving the flexibility and robustness of model-based derivative-free optimization solvers*, [arXiv:1804.00154](#) [[INSPIRE](#)].
- [103] C. Cartis, L. Roberts and O. Sheridan-Methven, *Escaping local minima with local derivative-free methods: a numerical investigation*, *Optimization* **71** (2021) 2343 [[arXiv:1812.11343](#)] [[INSPIRE](#)].
- [104] M.J.D. Powell, *An efficient method for finding the minimum of a function of several variables without calculating derivatives*, *Comput. J.* **7** (1964) 155 [[INSPIRE](#)].
- [105] D. Kodwani, D. Alonso and P. Ferreira, *The effect on cosmological parameter estimation of a parameter dependent covariance matrix*, *Open J. Astrophys.* **2** (2019) 3 [[arXiv:1811.11584](#)] [[INSPIRE](#)].
- [106] R.A. Fisher, *The logic of inductive inference*, *J. Roy. Statist. Soc.* **98** (1935) 39.
- [107] M. Kendall and A. Stuart, *The advanced theory of statistics. Volume 2: inference and relationship*, (1979).
- [108] M. Tegmark, A. Taylor and A. Heavens, *Karhunen-Loeve eigenvalue problems in cosmology: how should we tackle large data sets?*, *Astrophys. J.* **480** (1997) 22 [[astro-ph/9603021](#)] [[INSPIRE](#)].

- [109] EUCLID collaboration, *Euclid preparation: VII. Forecast validation for Euclid cosmological probes*, *Astron. Astrophys.* **642** (2020) A191 [[arXiv:1910.09273](#)] [[INSPIRE](#)].
- [110] N. Bhandari, C.D. Leonard, M.M. Rau and R. Mandelbaum, *Fisher matrix stability*, [arXiv:2101.00298](#) [[INSPIRE](#)].
- [111] S. Yahia-Cherif et al., *Validating the Fisher approach for stage IV spectroscopic surveys*, *Astron. Astrophys.* **649** (2021) A52 [[arXiv:2007.01812](#)] [[INSPIRE](#)].
- [112] R. Andrae, T. Schulze-Hartung and P. Melchior, *Dos and don'ts of reduced chi-squared*, [arXiv:1012.3754](#) [[INSPIRE](#)].
- [113] C.-H. Chuang et al., *UNIT project: universe N-body simulations for the investigation of theoretical models from galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **487** (2019) 48 [[arXiv:1811.02111](#)] [[INSPIRE](#)].
- [114] S. Mishra-Sharma, D. Alonso and J. Dunkley, *Neutrino masses and beyond- Λ CDM cosmology with LSST and future CMB experiments*, *Phys. Rev. D* **97** (2018) 123544 [[arXiv:1803.07561](#)] [[INSPIRE](#)].
- [115] E. Giusarma et al., *Scale-dependent galaxy bias, CMB lensing-galaxy cross-correlation, and neutrino masses*, *Phys. Rev. D* **98** (2018) 123526 [[arXiv:1802.08694](#)] [[INSPIRE](#)].
- [116] M. Schmittfull and U. Seljak, *Parameter constraints from cross-correlation of CMB lensing with galaxy clustering*, *Phys. Rev. D* **97** (2018) 123540 [[arXiv:1710.09465](#)] [[INSPIRE](#)].
- [117] EUCLID collaboration, *Euclid preparation — XV. Forecasting cosmological constraints for the Euclid and CMB joint analysis*, *Astron. Astrophys.* **657** (2022) A91 [[arXiv:2106.08346](#)] [[INSPIRE](#)].
- [118] L. Wenzl et al., *Cosmology with the Roman Space Telescope — synergies with CMB lensing*, *Mon. Not. Roy. Astron. Soc.* **512** (2022) 5311 [[arXiv:2112.07681](#)] [[INSPIRE](#)].
- [119] DES collaboration, *Dark Energy Survey year 3 results: exploiting small-scale information with lensing shear ratios*, *Phys. Rev. D* **105** (2022) 083529 [[arXiv:2105.13542](#)] [[INSPIRE](#)].
- [120] J. Chaves-Montero, R.E. Angulo and S. Contreras, *The galaxy formation origin of the lensing is low problem*, *Mon. Not. Roy. Astron. Soc.* **521** (2023) 937 [[arXiv:2211.01744](#)] [[INSPIRE](#)].
- [121] N. Kokron et al., *Accurate predictions from small boxes: variance suppression via the Zel'dovich approximation*, *JCAP* **09** (2022) 059 [[arXiv:2205.15327](#)] [[INSPIRE](#)].
- [122] J. DeRose, S.-F. Chen, N. Kokron and M. White, *Precision redshift-space galaxy power spectra using Zel'dovich control variates*, *JCAP* **02** (2023) 008 [[arXiv:2210.14239](#)] [[INSPIRE](#)].
- [123] R. Scoccimarro, M. Zaldarriaga and L. Hui, *Power spectrum correlations induced by nonlinear clustering*, *Astrophys. J.* **527** (1999) 1 [[astro-ph/9901099](#)] [[INSPIRE](#)].