Hefty enhancement of cosmological constraints from the DES Y1 data using a Hybrid Effective Field Theory approach to galaxy bias

Boryana Hadzhiyska,[a,1] **Carlos García-García,**[b] **David Alonso,**[b] **Andrina Nicola,**[c] **Anže Slosar**[d]

[a]Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA

[b]Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom

[c]Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

[d]Brookhaven National Laboratory, Physics Department, Upton, NY 11973, USA

E-mail: boryana.hadzhiyska@cfa.harvard.edu

Abstract. We present a re-analysis of cosmic shear and galaxy clustering from first-year Dark Energy Survey data (DES Y1), making use of a Hybrid Effective Field Theory (HEFT) approach to model the galaxy-matter relation on weakly non-linear scales, initially proposed in [1]. This allows us to explore the enhancement in cosmological constraining power enabled by extending the galaxy clustering scale range typically used in projected large-scale structure analyses. Our analysis is based on a recomputed harmonic-space data vector and covariance matrix, carefully accounting for all sources of mode-coupling, non-Gaussianity and shot noise, which allows us to provide robust goodness-of-fit measures. We use the ABACUSSUMMIT suite of simulations to build an emulator for the HEFT model predictions. We find that this model can explain the galaxy clustering and shear data up to wavenumbers $k_{\max} \sim 0.6 \, \mathrm{Mpc}^{-1}$. We constrain $(S_8, \Omega_m) = (0.786 \pm 0.020, 0.273^{+0.030}_{-0.036})$ at the fiducial $k_{\max} \sim 0.3 \, \mathrm{Mpc}^{-1}$, improving to $(S_8, \Omega_m) = (0.786^{+0.015}_{-0.018}, 0.266^{+0.024}_{-0.027})$ at $k_{\max} \sim 0.5 \, \mathrm{Mpc}^{-1}$. This represents a $\sim 10\%$ and $\sim 35\%$ improvement on the constraints derived respectively on both parameters using a linear bias relation on a reduced scale range ($k_{\max} \lesssim 0.15 \, \mathrm{Mpc}^{-1}$), in spite of the 15 additional parameters involved in the HEFT model. We investigate whether HEFT can be used to constrain the Hubble parameter and find $H_0 = 70.7^{+3.0}_{-3.5} \, \mathrm{km} \, s^{-1} \, \mathrm{Mpc}^{-1}$. Our constraints are investigative and subject to certain caveats discussed in the text.

Contents

1 Introduction

We are entering the prime time era of photometric cosmological surveys. A number of these surveys have reported results over the past few years, including the Dark Energy Survey (DES)[1] [2], the Hyper Suprime-Cam survey (HSC)[2] [3] and the Kilo-Degree Survey (KiDS)[3] [4], and several experiments are currently in construction, most notably the Rubin Observatory Legacy Survey of Space and Time (LSST)[4], Euclid[5] and the Roman Telescope[6].

    The analysis of photometric survey data is challenging, both in terms of avoiding biases in cosmological parameters from observational systematics, as well as in terms of theoretical modeling of the expected signal. In order to mitigate modeling uncertainties, the analysis of the DES Year 1 galaxy clustering data, alone [5] or in combination with cosmic shear [2], was limited to large scales, where a linear bias relation could be safely used. In this model, it is assumed that the contrast in the number densities of galaxies is proportional to that of the dark matter. While deceptively simple, this model can be shown to be exact in the limit of infinitely large scales under rather general assumptions, as long as the galaxy formation process is local on some finite scale. It is therefore sufficient on large enough scales, but requires that the information on the best measured intermediate and small scales be

---

[1]https://www.darkenergysurvey.org.
[2]https://hsc.mtk.nao.ac.jp/ssp.
[3]http://kids.strw.leidenuniv.nl.
[4]https://www.lsst.org.
[5]https://www.euclid-ec.org.
[6]https://roman.gsfc.nasa.gov.

discarded. The purpose of this paper is to investigate the implications of a re-analysis of the DES Y1 data by employing a more sophisticated bias model.

There are two main difficulties in modeling data on smaller scales. On one hand, the dark matter clustering becomes non-linear, and on the other hand, the manner in which galaxies trace the underlying dark matter fluctuations (which are assumed to be the dynamically dominant component) is also non-linear. The dark matter clustering can be modelled through high-accuracy dark matter N-body simulations; this modeling would be exact if it were not for the hydrodynamical effects of baryons, which are luckily confined to scales $k \gtrsim 1$ $h \, \mathrm{Mpc}^{-1}$, and whose residual effects on large scales can be emulated with numerous approximate schemes [6–10]. The tracing of dark matter by galaxies, however, is a more difficult problem. It is hopeless to model the richness of galaxy formation physics completely ab-initio in cosmologically relevant volumes, and even the most precise small-scale hydro simulations need to model subgrid physics phenomenologically [11]. Nevertheless, there are numerous phenomenological models, which can be surprisingly successful at explaining galaxy clustering to deeply non-linear scales. Many of these models mostly rely on the notion of halo occupation distribution (HOD) [12, 13], namely the idea that galaxies occupy dark matter halos with statistics that mostly depend on halo mass. The issue is that, while these are very successful in explaining the observed clustering at 10-percent level, it is hard to set up models that are precise at the percent level without at least some a-posteriori tuning of the model to fit the data. The basic difficulty lies in the fact that these models are under little theoretical control, and are instead based on heuristic models of halo occupation, rather than fundamental principles. Thus there are no strict rules on when to turn left and when to step forward in this dance.

On the other end of the spectrum of modeling approaches lie purely analytical models that start with a linear-biasing scheme and use perturbative expansion to model both the non-linear dark matter structure growth and the galaxy tracing, integrating any residual small-scale effects into renormalization of the large-scale quantities. This is known as the effective theory of large-scale structure (see e.g. [14, 15]). While this approach is exact and under full theoretical control, it has the downside that it leads to only a modest gain in the smallest scales that one can still model.

In this paper we employ a method first proposed in [1] that combines the accuracy of N-body simulations with the theoretical robustness of analytical bias expansions. In this approach N-body simulations provide accurate information about statistics of dark matter clustering that are required to calculate the galaxy clustering for an analytical bias model in Lagrangian space. In particular, the N-body simulations are used to calculate 15 "basis" power spectra which multiply the bias coefficients when modeling the galaxy signal. The beauty of this approach is that it maintains the near-exactness of N-body simulations for dark matter clustering with the theoretical control of an analytical bias expansion. We refer to this model as the Hybrid EFT (HEFT) model.

The HEFT method is most naturally expressed in Fourier space, because any analytic Taylor expansion will break down on sufficiently small Fourier scales. Therefore, on the data side, the main difference with the DES Y1 analysis is that we perform a power spectrum rather than a correlation function analysis. We re-measure the auto- and cross-power spectra and covariance matrix, but otherwise leave the DES analysis largely unchanged, we use the same input catalogs, tomographic bins, and systematic models.

This paper is structured as follows. In Section 2 we discuss how we measure auto- and cross-power spectra of the DES Y1 data, including estimation of the covariance matrix.

This section relies heavily on our previous works using galaxy and shear data from HSC and DES [16, 17]. In Section 3 we discuss the theoretical underpinning of the HEFT model, and method used to build an emulator of the HEFT basis power spectra from the ABACUSSUMMIT simulations. We then proceed to analyze the DES Y1 data using the HEFT model in Section 4, where we also spend some time addressing variations of the parameter inference to investigate the sensitivity to various analysis choices. We conclude in Section 5 with a recap of our main results and list of potential caveats.

2 Data and power spectrum measurement

2.1 Data

We make use of the Dark Energy Survey's first-year (Y1) public data release [18]. In particular, we use the publicly available key Y1KP catalogs[7], used to derive cosmological constraints from the joint analysis of galaxy clustering and cosmic shear data (the so-called "3×2-point" analysis), presented in [2] (DY1 hereon). The resulting clustering and shear samples cover an area of $\sim 1320 \deg^2$.

As a galaxy clustering sample we use the REDMAGIC catalog described in [5]. The REDMAGIC algorithm selects luminous red galaxies in a way that minimizes photometric redshift uncertainties, producing samples with small redshift distribution tails, well-suited for cosmological galaxy clustering studies. We divide the sample into the same 5 redshift bins used in DY1, and use the same galaxy weights to correct for sky systematics. In the analysis of these data, we make use of the fiducial redshift distributions published with the Y1 dataset, as well as the same model to marginalize over photo-z uncertainties in them (see Section 4.1). Further details can be found in [5].

The DES Y1 analysis was carried out using two different shape-measurement algorithms: METACALIBRATION [19, 20] and IM3SHAPE [21]. We restrict the current analysis to the METACALIBRATION sample, applying the same cuts used in [21], including its division into four tomographic reshift bins (see also [17]). We use the redshift distributions provided with the Y1 release for these bins, details for which can be found in [22]. Details regarding the calibration of galaxy ellipticities, including mean subtraction and the inclusion of selection effects in the METACALIBRATION response tensor, can be found in [17], and follow the same prescriptions used by DY1.

Fig. 1 shows one of the angular window functions and redshift distributions of the DES Y1 samples used here.

2.2 Power spectra and covariances

Starting from the official Y1 catalogs described in the previous section, we compute a set of angular auto- and cross-power spectra between the clustering and shear samples. Thus, although our analysis is based on the same samples as DY1, we construct an independent data vector, instead of using the summary statistics (real-space angular correlation functions). We do this in order to take advantage of recent improvements in the algorithms used for the estimation of angular power spectra of LSS data and their covariance [17, 23, 24], which allow us to produce robust goodness-of-fit tests, vital for this analysis (see Section 4.2). The use of angular power spectra also allows us to impose simple scale cuts that are more directly connected with a comoving Fourier scale $k_{\max}$. Finally, our analysis can also validate the

---

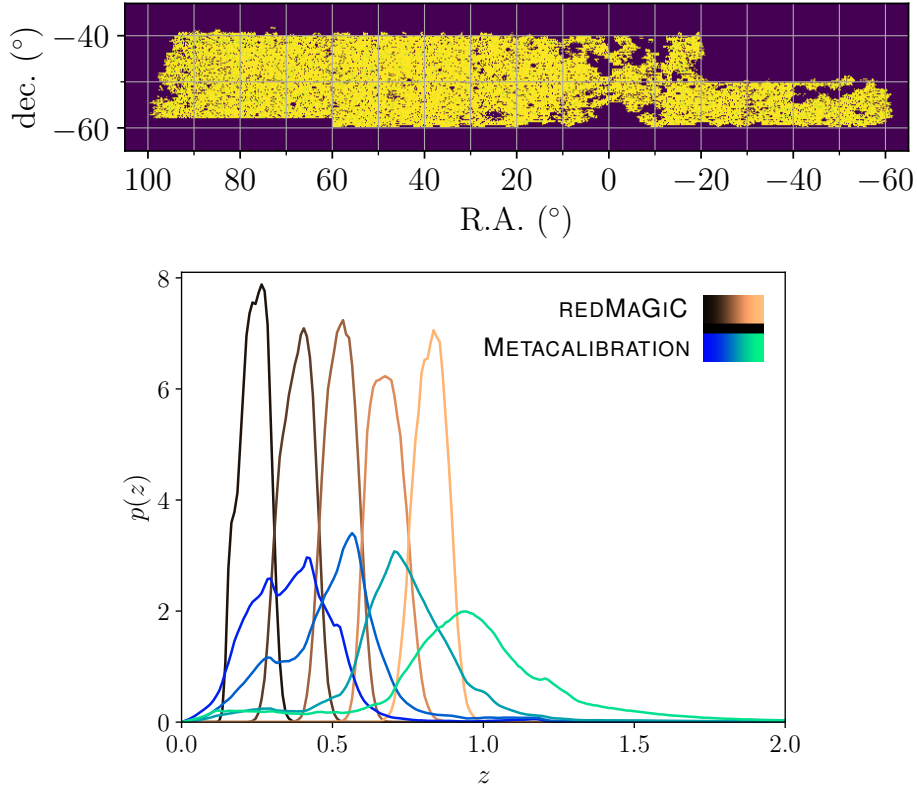[7]https://des.ncsa.illinois.edu/releases/y1a1/key-catalogs

Figure 1: *Top:* one of the angular window functions for the DES Y1 data analysed here. *Bottom:* normalized redshift distribution of the galaxy clustering (REDMAGIC) and cosmic shear (METACALIBRATION) samples.

results found by the DES Y1 analisis in real space with an independent harmonic-space pipeline (see also [25]).

Power spectra were calculated using a pseudo-C_ℓ approach [26] as implemented in NA-MASTER. The details of the pseudo-C_ℓ algorithm are described in detail in [23], and we provide only a succinct summary here. Pseudo-C_ℓ estimators are a fast implementation of the optimal quadratic estimator [27]. The algorithm is based on assuming a diagonal map-level covariance matrix, such that inverse-variance weighting is equivalent to a simple multiplication of the sky map by a "weights map" or "mask" $w(\hat{\mathbf{n}})$. Calculating the power spectrum of two weighted maps is then reduced to averaging the product of their spherical harmonic coefficients over the multipole order m, and the time-consuming mode-coupling matrix $M_{\ell\ell'}$ can be computed analytically using fast algorithms to calculate Wigner 3-j symbols. The method works for fields of arbitrary spin, including the galaxy overdensity (spin-0) and cosmic shear (spin-2). We compute all auto- and cross-correlations between the 5 REDMAGIC and 4 METACALIBRATION redshift bins. Note that, as done in DY1, we discard cross-correlations between different galaxy clustering bins in the likelihood analysis, given their sensitivity to photometric redshift systematics. All maps were generated using `HEALPix`[8] [28] with resolution parameter $N_{\rm side} = 4096$, corresponding to pixels of size $\delta\theta_{\rm pix} \sim 1'$. This is small enough that, as described in [17], all pixelization effects can be ignored on the scales used in this

---

[8]http://healpix.sourceforge.net.

analysis (see Section 4.1).

The galaxy overdensity maps are estimated using the same method described in [16]. The overdensity δ_p at pixel p is computed as $\delta_p = N_p/(w_p \bar{N}) - 1$, where N_p is the weighted number of objects in the pixel, $\bar{N}$ is the mean weighted number of objects per pixel, and w_p is the unmasked fractional area of pixel p. The mean is computed as $\bar{N} = \sum_p N_p / \sum_p w_p$. The fractional area map w_p, which we also use as the weights map for the galaxy overdensity in the pseudo-C_ℓ estimator, is provided in the Y1 data release. The overdensity is set to zero in all fully masked pixels. Furthermore, to avoid noise in strongly masked pixels, we set the weights map to zero for pixels where $w_p < 0.5$. Finally, the galaxy noise power spectrum is estimated as described and validated in [16], with an extra correction factor [29] to account for effect of galaxy weights. This is then subtracted from all auto-correlations.

The auto- and cross-correlations involving shear maps only were computed using the procedure described in [17]. The weights map associated with a shear redshift bin is simply proportional to the sum of shape-measurement weights of all galaxies in the pixel, and the associated shear map is the weighted mean calibrated ellipticity in each pixel. The resulting shear power spectra were presented and validated in [17]. The clustering-shear cross-correlations were simply estimated as the pseudo-C_ℓ between the corresponding galaxy overdensity and shear maps. As a sanity check, we verify that all cross-correlation involving shear B-modes are compatible with zero by examining the probability-to-exceed (PTE) of their χ^2 with respect to the null hypothesis.

All power spectra were calculated for a set of ℓ-bins (bandpowers) covering the range $\ell \in [0, 12288)$. We use a linear spacing with $\Delta\ell = 30$ up to $\ell = 240$, and a logarithmic spacing thereafter with $\Delta \log_{10} \ell = 0.055$. Even after correcting for the effects of survey geometry by inverting the binned mode-coupling matrix, residual mode-coupling remains as a result of binning. We account for this exactly by convolving the theory prediction with the bandpower window functions as described in [23].

The covariance matrix of these measured power spectra is computed using analytical methods. As discussed in [30, 31], the main contributions to the covariance of power spectra of large-scale structure tracers can be written as:

$$\mathrm{Cov}_{\ell\ell'} = \mathrm{Cov}_{\ell\ell'}^{\mathrm{G}} + \mathrm{Cov}_{\ell\ell'}^{\mathrm{cNG}} + \mathrm{Cov}_{\ell\ell'}^{\mathrm{SSC}}, \tag{2.1}$$

where $\mathrm{Cov}_{\ell\ell'}^{\mathrm{G}}$ is the "Gaussian" covariance matrix, associated with the fields' disconnected trispectrum, and $\mathrm{Cov}_{\ell\ell'}^{\mathrm{cNG}}$ and $\mathrm{Cov}_{\ell\ell'}^{\mathrm{SSC}}$ are non-Gaussian terms, sourced by the non-linear evolution of the matter overdensities under gravity.

$\mathrm{Cov}_{\ell\ell'}^{\mathrm{G}}$ dominates the error budget, and therefore must be carefully calculated, accounting for the effects of survey geometry in the form of mode-coupling. To do so, we follow the approximate methods of [17, 24, 32]. The exact calculation of the covariance matrix scales as $O(\ell_{\mathrm{max}}^6)$, and is therefore unfeasible for the range of scales used in this work. The calculation can be reduced to $O(\ell_{\mathrm{max}}^3)$ under the approximation of a narrow mode-coupling kernel and a sufficiently flat underlying power spectrum. While this is a good approximation for galaxy clustering, the large inhomogeneity of the weak lensing mask (effectively proportional to the galaxy density) breaks these assumptions and can lead to $O(1)$ errors in the cosmic shear covariance. To remedy this, we make use of the improved narrow-kernel approximation presented in [17], which is able to accurately recover the true power spectrum uncertainties up to a few percent on the scales used here, including the different noise and signal contributions.

We estimate the SSC and cNG contributions to the total covariance, following the halo model based approach of Ref. [33], as was done in Refs. [16, 17, 34]. For a more detailed

description, we refer the reader to these works and the references therein. As the total covariance matrix given in Eq. 2.1 is dominated by the Gaussian part, we model finite sky effects for the non-Gaussian corrections using the approximations given in Ref. [33] and do not fully account for mode-coupling as we do for the Gaussian part. Note that, within the range of scales used here, the effect of the non-Gaussian terms on the χ^2 is smaller than 2%, and therefore this should be a good approximation as long as survey geometry effects are accurately accounted for in the Gaussian part.

3 Modeling the signal

3.1 Projected statistics

We will extract constraints on cosmological parameters from the two-point statistics of two fields projected on the celestial sphere: the galaxy overdensity $\delta_g^\alpha(\hat{\mathbf{n}})$ and the weak lensing shear $\gamma^\alpha(\hat{\mathbf{n}})$ for galaxies in redshift bin α. These are related to the three-dimensional fluctuations in the galaxy number density $\Delta_g(\mathbf{x})$ and the matter density $\Delta_m(\mathbf{x})$ via [35, 36]

$$\delta_g^\alpha(\hat{\mathbf{n}}) = \int_0^{\chi_H} d\chi\, q_g^\alpha(\chi)\, \Delta_g(\chi(z)\hat{\mathbf{n}}, z), \quad \gamma^\alpha(\hat{\mathbf{n}}) = \int_0^{\chi_H} d\chi\, q_\gamma^\alpha(\chi) \left[-\chi^{-2} \eth\eth\nabla^{-2}\Delta_m(\chi\hat{\mathbf{n}}, z) \right],$$

$$q_g^\alpha(\chi) \equiv \frac{H(z)}{c} p_\alpha(z), \quad q_\gamma^\alpha(\chi) \equiv \frac{3}{2} H_0^2 \Omega_m \frac{\chi}{a(\chi)} \int_{z(\chi)}^\infty dz'\, p_\alpha(z') \frac{\chi(z') - \chi}{\chi(z')}, \tag{3.1}$$

where c is the speed of light, $\hat{\mathbf{n}}$ is the sky direction, χ is the comoving radial distance at redshift z, χ_H is the distance to the horizon, $H(z)$ is the Hubble expansion rate, $H_0 \equiv H(z=0)$, Ω_m is the matter density parameter today, $p_\alpha(z)$ is the redshift distribution in bin α, and $\eth$ is the spin-raising differential operator, acting on a spin-s quantity as:

$$\eth_s f(\theta, \varphi) = -(\sin\theta)^s \left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta} \frac{\partial}{\partial\varphi} \right) (\sin\theta)^{-s}\, {}_s f \tag{3.2}$$

and turning it into a spin-$(s+1)$ quantity.

The power spectrum between quantities X and Y (δ_g or γ) in bins α and β respectively, $C_\ell^{(X,\alpha)(Y,\beta)}$ is the covariance of the spherical harmonic coefficients of both fields, and can be related to the power spectrum of the three-dimensional quantities associated with X and Y (Δ_g or Δ_M) $P_{XY}(k, z)$ via:

$$C_\ell^{(X,\alpha),(Y,\beta)} = \int \frac{d\chi}{\chi^2}\, q_X^\alpha(\chi)\, q_Y^\beta(\chi)\, P_{XY}\left(k = \frac{\ell + 1/2}{\chi}, z(\chi) \right). \tag{3.3}$$

Equation 3.3 uses Limber's approximation [37, 38], valid for the wide radial kernels considered here. In order to account for the difference between angular and three-dimensional derivatives in Eq. 3.1 (i.e. $\chi^2 \eth^2 \nabla^{-2} \not\equiv 1$), the lensing kernel must be multiplied by an ℓ-dependent prefactor

$$G_\ell \equiv \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \frac{1}{(\ell+1/2)^2}, \tag{3.4}$$

which becomes irrelevant (sub-percent) for $\ell > 11$ [39].

Thus, given a set of cosmological parameters, and the redshift distributions of all bins considered, all that remains to specify is the 3-dimensional power spectra between Δ_m and Δ_g. For the matter power spectrum $P_{mm}(k, z)$, we use the non-linear prediction from HALOFIT [40, 41], as was done in DY1. We describe the model used to describe the galaxy-matter connection in the next section.

3.2 Hybrid EFT model for galaxy clustering

We follow a perturbative effective field theory (EFT) approach to galaxy biasing in Lagrangian space, coupled with the non-linear dynamical evolution of N-body simulations as prescribed in [1]. We describe the logic behind this Hybrid EFT method here, and refer readers to [1, 15, 42] for a detailed description of galaxy bias.

The complex physical processes that govern the formation and evolution of galaxies necessarily imply a complex relationship between their distribution and that of the matter inhomogeneities they trace. In general, this relation should be non-linear, non-local and, without exact knowledge of the small-scale physics, stochastic. Non-local and stochastic effects are sourced by the dependence of the galaxy abundance at a given point in space on the small-scale physical processes in a region around it, and therefore should become negligible on scales larger than the size of this region (e.g. the Lagrangian radius of a typical dark matter halo).

On scales where galaxy formation can be modelled as a local process, it is then possible to invoke the equivalence principle, in terms of which the leading gravitational effects are associated with the Hessian of the gravitational potential $\partial_i\partial_j\Phi$. This can be split into its scalar trace, proportional to the matter overdensity δ, and the traceless tidal tensor $s_{ij} \equiv (\partial_i\partial_j\nabla^{-2} - 1/3)\delta$. On these scales, we can therefore describe the number overdensity of galaxies found at redshift z in the Lagrangian initial conditions as a general functional of the local δ and s_{ij}. Expanding this functional up to second order:

$$1 + \Delta_{g,L} = 1 + b_1\delta_L + b_2(\delta_L^2 - \langle\delta_L^2\rangle) + b_s(s_L^2 - \langle s_L^2\rangle) + b_\nabla\nabla^2\delta_L, \tag{3.5}$$

where we have only kept scalar combinations of s_{ij}, $s^2 \equiv s_{ij}s^{ij}$, and we have included a leading-order non-local contribution $\propto \nabla^2\delta_L$. The subscript L is a reminder that all quantities are evaluated at the initial Lagrangian coordinates $\mathbf{q}$. One can then evolve $\Delta_{g,L}$ to its observed redshift by advecting the Lagrangian Δ_g to the final Eulerian coordinates:

$$1 + \Delta_g(\mathbf{x}) = \int d^3\mathbf{q}\,[1 + \Delta_{g,L}(\mathbf{q})]\,\delta^D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q})), \tag{3.6}$$

where Ψ is the Lagrangian displacement vector.

This calculation can be done using Lagrangian perturbation theory (e.g. [43]) or, as proposed in [1], solving the full non-linear evolution in an N-body simulation. Substituting Eq. 3.5 into 3.6, the final galaxy overdensity is a linear combination of the individual operators in Eq. 3.5 (δ_L, δ_L^2, s_L^2 and $\nabla^2\delta_L$) advected to the Eulerian positions. These fields can be calculated in the simulation by weighting each matter particle in a given snapshot by the value of the corresponding operator in the initial conditions at the original Lagrangian coordinates. The cross-power spectrum between the galaxy and matter overdensities, as well as the galaxy-galaxy power spectrum are then given by:

$$P_{gm}(k) = \sum_{\alpha\in\mathcal{O}} b_\alpha P_{1\alpha}(k), \quad P_{gg}(k) = \sum_{\alpha\in\mathcal{O}}\sum_{\beta\in\mathcal{O}} b_\alpha b_\beta P_{\alpha\beta}(k) \tag{3.7}$$

where $\mathcal{O} \equiv \{1, \delta_L, \delta_L^2, s_L^2, \nabla^2\delta_L\}$ is the full set of second-order operators in Eq. 3.5, $P_{\alpha\beta}(k)$ is the power spectrum of the advected fields α and β, and b_α are the corresponding bias coefficients. In this formalism $P_{11}(k)$ is the non-linear matter power spectrum, and its corresponding bias parameter (called b_0 here) is $b_0 = 1$. Our bias model then reduces to computing

Sim. name	$\Omega_b h^2$	$\Omega_c h^2$	n_s	σ_8
c000	0.02237	0.1200	0.9649	0.811355
c100/c101	0.02282/0.02193	–	–	–
c102/c103	–	0.1240/0.1161	–	–
c104/c105	–	–	0.9749/0.9549	–
c112/c113	–	–	–	0.8276/0.7954

Table 1: Simulations from the ABACUSSUMMIT suite used to generate the power spectrum templates used in this analysis.

a set of 15 different power spectrum templates $P_{\alpha\beta}(k)$ for the five different operators. We construct these from the ABACUSSUMMIT suite of simulations, described in the next section.

We note that while formally Lagrangian and Eulerian models are both valid and complete descriptions, their predictions do not match at any given order. Therefore, even the simplest linear analysis will give different results in Eulerian and Lagrangian spaces for a finite maximum wavenumber $k_{\max}$.

Finally, it is also common to include an additive stochastic term ϵ in the bias expansion (Eq. 3.6), to account for the impact of small-scale density fluctuations on galaxy formation. In the simplest case, this effect can be modelled by treating galaxies as a Poisson sampling of an underlying smooth galaxy density field, in which case the contribution to the galaxy auto-correlation is $P_{\epsilon\epsilon} = 1/\bar{n}_g$, where $\bar{n}_g$ is the comoving number density of galaxies. In our analysis, we will also consider the effect of departures from this pure shot-noise contribution, of the form $P_{\epsilon\epsilon} = A_\epsilon/\bar{n}_g$, with $A_\epsilon \neq 1$ (see Section 4.1 and Section 4.3).

3.3 Abacus simulations

To build a model for the different power spectrum templates $P_{\alpha\beta}(k)$ in Eq. 3.7 we use the ABACUSSUMMIT suite of N-body simulations [44–46].

ABACUSSUMMIT was designed to meet the cosmological simulation requirements of the Dark Energy Spectroscopic Instrument (DESI) survey and run on the Summit supercomputer at the Oak Ridge Leadership Computing Facility. The simulations are run with the highly accurate ABACUS cosmological N-body simulation code [47], optimized for GPU architectures and large-volume, moderately clustered simulations. The ABACUS code is extremely fast, performing 70 million particle updates per second on each of the Summit nodes, and also extremely accurate, with typical force accuracy below 10^{-5}. The output halo catalogs and particle subsamples (A and B, amounting to 10% of the total particle population) are organized into 12 primary redshift snapshots ($z = 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.1, 1.4, 1.7, 2.0, 2.5$, and 3.0). The data products have been designed with the aim of supporting mock catalogs to be constructed using halo occupation distributions, as well as efficient access to measurements of the density fields. In this work, we use the 7 redshift snapshots at $z \leq 1.1$, $z = 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.1$, covering the redshift range of the DES Y1 clustering sample, and the particle subsample A, which contains 3% of the total particle content.

The ABACUSSUMMIT suite contains simulations for various cosmologies. Here, we employ the fiducial simulation `AbacusSummit_base_c000_ph000` at base resolution (6912^3 dark

matter particles in a box of length 2000 Mpc/h), and some of the "linear derivatives" simulations at the same resolution. These are used to account for the parameter dependence of the power spectrum templates as described below. The simulations used are listed in Table 1. The parameters that define the fiducial cosmology are the baryon energy density ($\Omega_b h^2 = 0.02237$), the cold dark matter energy density ($\Omega_c h^2 = 0.1200$), the primordial tilt ($n_s = 0.9649$), the amplitude of matter fluctuations ($\sigma_8 = 0.811355$), and the distance to last scatter ($100\theta_* = 1.041533$). The "linear derivative" simulations vary each of these in turn, as indicated by the columns in the table.

For each simulation and redshift bin we produce 15 power spectrum templates, corresponding to the auto- and cross-correlation between the 5 operators of the Lagrangian bias expansion (Eq. 3.6) following the method outlined in [1]:

1. We use the initial conditions to calculate 4 fields: δ_L, δ_L^2, s_L^2 and $\nabla^2 \delta_L$. These fields are computed on a cubic grid of size 2304^3. Following [1, 48] we do not apply any smoothing and so our results must depend weakly on cell size.

2. At each snapshot, we evolve the initial condition fields to the corresponding redshift assuming a linear growth factor. Then, for each of the 5 bias operators (1, δ_L, δ_L^2, s_L^2, $\nabla^2 \delta_L$), we compute their advected version by assigning each dark matter particle a weight given by the value of the corresponding operator at the particle's position in the initial conditions.

3. Finally, we compute and store the auto- and cross-power spectra between all advected fields.

The power spectrum templates thus calculated are noisy on large scales due to cosmic variance. To avoid this, we combine the simulated power spectra with theoretical predictions from Lagrangian perturbation theory (LPT) using the `velocileptors` code [43, 49] on large scales. We use the following combination of both predictions to enforce a smooth transition between them at $k \sim k_{\rm pivot}$:

$$P_{\alpha\beta}(k) = (1 - w(k))P_{\alpha\beta}^{\rm sim}(k) + w(k)\,P_{\alpha\beta}^{\rm LPT}(k), \tag{3.8}$$

where $P_{\alpha\beta}^{\rm sim}$ and $P_{\alpha\beta}^{\rm LPT}$ are the predictions from ABACUSSUMMIT and LPT respectively, and weighting function

$$w(k) \equiv \frac{1}{2}\left[1 - \tanh\left(\frac{k - k_{\rm pivot}}{\Delta k_w}\right)\right], \tag{3.9}$$

which ensures smooth interpolation between the two limits. We use $\Delta k_w = 0.01\,h{\rm Mpc}^{-1}$, but manually fine-tune values of $k_{\rm pivot}$ for the different operator combinations, based on their large-scale behavior:

$$k_{\rm pivot}^{0,s} = k_{\rm pivot}^{1,s} = k_{\rm pivot}^{\nabla,s} = 0.2\,h{\rm Mpc}^{-1}, \quad k_{\rm pivot}^{2,s} = k_{\rm pivot}^{s,s} = 0.03\,h{\rm Mpc}^{-1}, \tag{3.10}$$

$$k_{\rm pivot}^{2,\nabla} = 0.3\,h{\rm Mpc}^{-1}, \quad k_{\rm pivot}^{1,\nabla} = 0.07\,h{\rm Mpc}^{-1}, \quad k_{\rm pivot}^{\rm other} = 0.09\,h{\rm Mpc}^{-1}. \tag{3.11}$$

The power spectrum templates thus produced for the fiducial c000 simulation at $z = 0.5$ are shown in Fig. 2, which also illustrates the procedure outlined above. The code used to generate these templates is available at https://github.com/boryanah/hybrid_eft_nbody.

In order to account for the dependence of the power spectrum templates on cosmological parameters, we use a linear Taylor expansion around the fiducial ABACUSSUMMIT cosmology
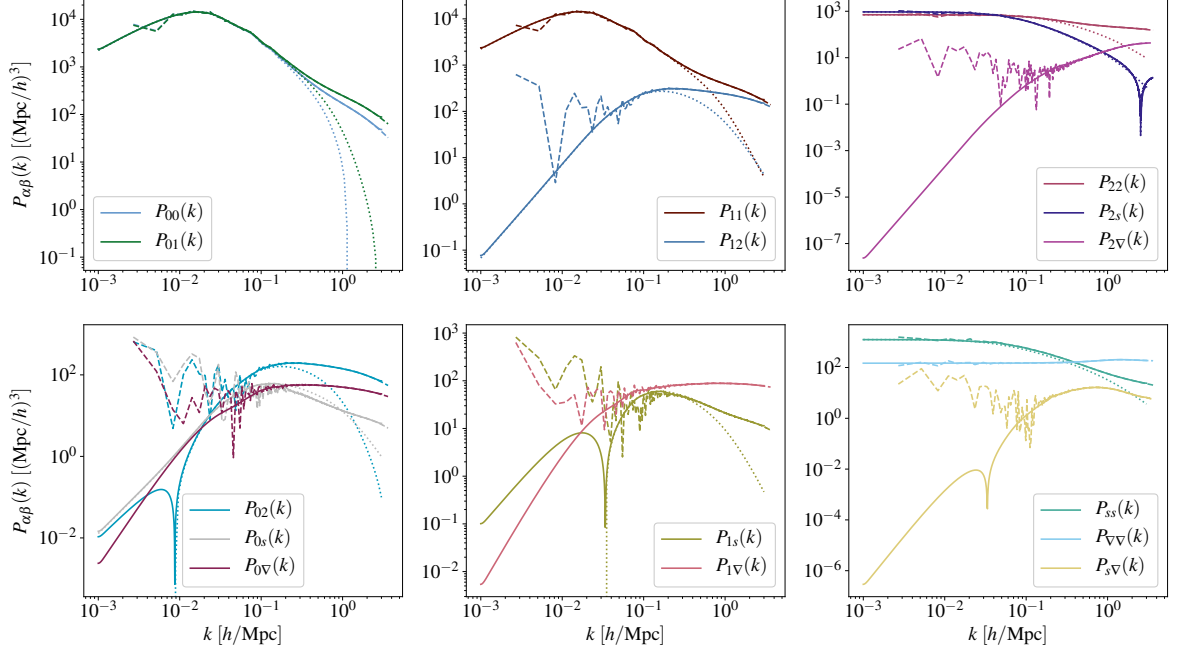
Figure 2: Power spectrum templates at $z = 0.5$ for the fiducial ABACUSSUMMIT simulation (`c000`). The dotted lines correspond to the 1-loop LPT power spectra computed using `velocileptors` [49], while the dashed lines are derived from ABACUSSUMMIT. In solid lines, we show the combined power spectra, obtained by smoothly interpolating between both. As expected, the 1-loop LPT theory lacks small-scale power compared with the N-body result. We note that 1-loop LPT makes no prediction for the templates involving cross-correlations with the $\nabla^2\delta_L$ field, for which we assume $P_{X\nabla}(k) \approx k^2 P_{X\delta}$ on large scales and the numerical values on small scales. We model the $P_{\nabla\nabla}(k)$ as approximately constant on large scales.

(first row of Table 1). We first compute an estimate of the derivative of the power spectrum templates with respect to the four cosmological parameters $(\Omega_b h^2, \Omega_c h^2, \sigma_8, n_s)$ via numerical differentiation of the templates found for the fiducial `c000` simulation and the linear derivative simulations. I.e. for a given parameter θ:

$$\partial_\theta P_{\alpha\beta}(k) = \frac{P_{\alpha\beta}(k; \theta_F + \delta\theta) - P_{\alpha\beta}(k; \theta_F - \delta\theta)}{2\delta\theta}, \tag{3.12}$$

where $P_{\alpha\beta}(k; \theta_F + \delta\theta)$ are the templates calculated in the two linear-derivative simulations for which this parameter is varied (by an amount $\delta\theta$), and θ_F is the value of the parameter in the fiducial ABACUSSUMMIT cosmology. The power spectrum template at a set of cosmological parameters $\vec{\theta}$ is then given by:

$$P_{\alpha\beta}(k; \vec{\theta}) = P_{\alpha\beta}(k; \vec{\theta}_*) + (\vec{\theta} - \vec{\theta}_*) \cdot \nabla_\theta P_{\alpha\beta}(k), \tag{3.13}$$

where $\vec{\theta}_*$ is the fiducial ABACUSSUMMIT cosmology.

Note that the initial conditions of all simulations we use are produced for the same phase (`ph000`). Thus, when taking finite differences, we eliminate some of the cosmic variance noise. We verified the validity of the linear Taylor expansion by comparing the prediction from Eq. 3.13 with power spectrum templates calculated directly from the extended grid of "linear

derivative" ABACUSSUMMIT simulations at redshifts $0.5 \leq z \leq 1.1$ (labelled `c117-120`, `c119,` `c120`). The prediction is found to be accurate at the $\sim 1\%$ level on all scales of interest $k \leq 1$ $h \, \mathrm{Mpc}^{-1}$.

For parameter values sufficiently far away from the fiducial ABACUSSUMMIT cosmology, we expect the linear Taylor expansion to break down. Thus, in order to recover some of the true parameter dependence, we combine ABACUSSUMMIT predictions for the *ratio* between $P_{\alpha\beta}(k)$ and the matter power spectrum $P_{mm}(k) \equiv P_{11}(k)$, with the `HALOFIT` prediction for the latter as follows:

$$P_{\alpha\beta}(k; \vec{\theta}) = \frac{P_{mm}^{\mathtt{HF}}(k; \vec{\theta})}{P_{11}^{\mathtt{AB}}(k; \vec{\theta})} P_{\alpha\beta}^{\mathtt{AB}}(k; \vec{\theta}), \tag{3.14}$$

where $P^{\mathtt{HF}}$ and $P^{\mathtt{AB}}$ are the `HALOFIT` and ABACUSSUMMIT predictions respectively. By taking the ratio of the ABACUSSUMMIT predictions we thus mitigate the impact of residual noise on the templates, and some of the error made in assuming a linear dependence on cosmological parameters around the ABACUSSUMMIT fiducial model. In particular, this approach recovers the `HALOFIT` matter power spectrum exactly.

An important caveat must be noted. The linear derivative simulations available in the ABACUSSUMMIT suite use a value of the Hubble parameter h determined by holding the comoving angular diameter distance to the last-scattering surface, θ_*, constant, and equal to the value inferred by *Planck* [50]. Effectively, this means that our parametrization of the cosmological dependence of $P_{\alpha\beta}$ is "missing" the Hubble parameter, which is assumed to reproduce the position the peak position in the *Planck* CMB power spectrum. This is one of the best and most robustly measured quantities in cosmology; however, it complicates the direct comparison between our results and those of DY1. Although we find that our results are not strongly sensitive to this (see Section 4), a robust implementation of this method should allow for variation of all basic cosmological parameters.

As stated above, our ability to account for the dependendence of the power spectrum templates on cosmological parameters accurately is limited by the range of cosmologies covered by the ABACUSSUMMIT suite. Although combining the simulation data with `HALOFIT` should allow us to capture some of the dependence beyond the linear Taylor expansion, the current model could clearly be improved by building a more complete power spectrum emulator covering a wider region of the cosmological parameter space. This has recently been done in [51] and [52]. Nevertheless, as we show in Section 4 by switching off the cosmological dependence of the power spectrum templates completely, our final constraints on (S_8, Ω_m) are insensitive to the inaccuracy of the linear expansion for the data used here.

4 Results

In this section, we present the cosmological constraints extracted from the DES Y1 galaxy clustering and cosmic shear data for different choices of bias parametrization and scale cuts.

4.1 Likelihood analysis

In order to derive constraints on cosmological and bias parameters we use a Gaussian likelihood of the form:

$$\log p(\vec{\theta}|\mathbf{d}) = -\frac{1}{2}[\mathbf{d} - \mathbf{t}(\vec{\theta})]^T \mathsf{C}^{-1}[\mathbf{d} - \mathbf{t}(\vec{\theta})] + \log p_p(\vec{\theta}) + K \tag{4.1}$$
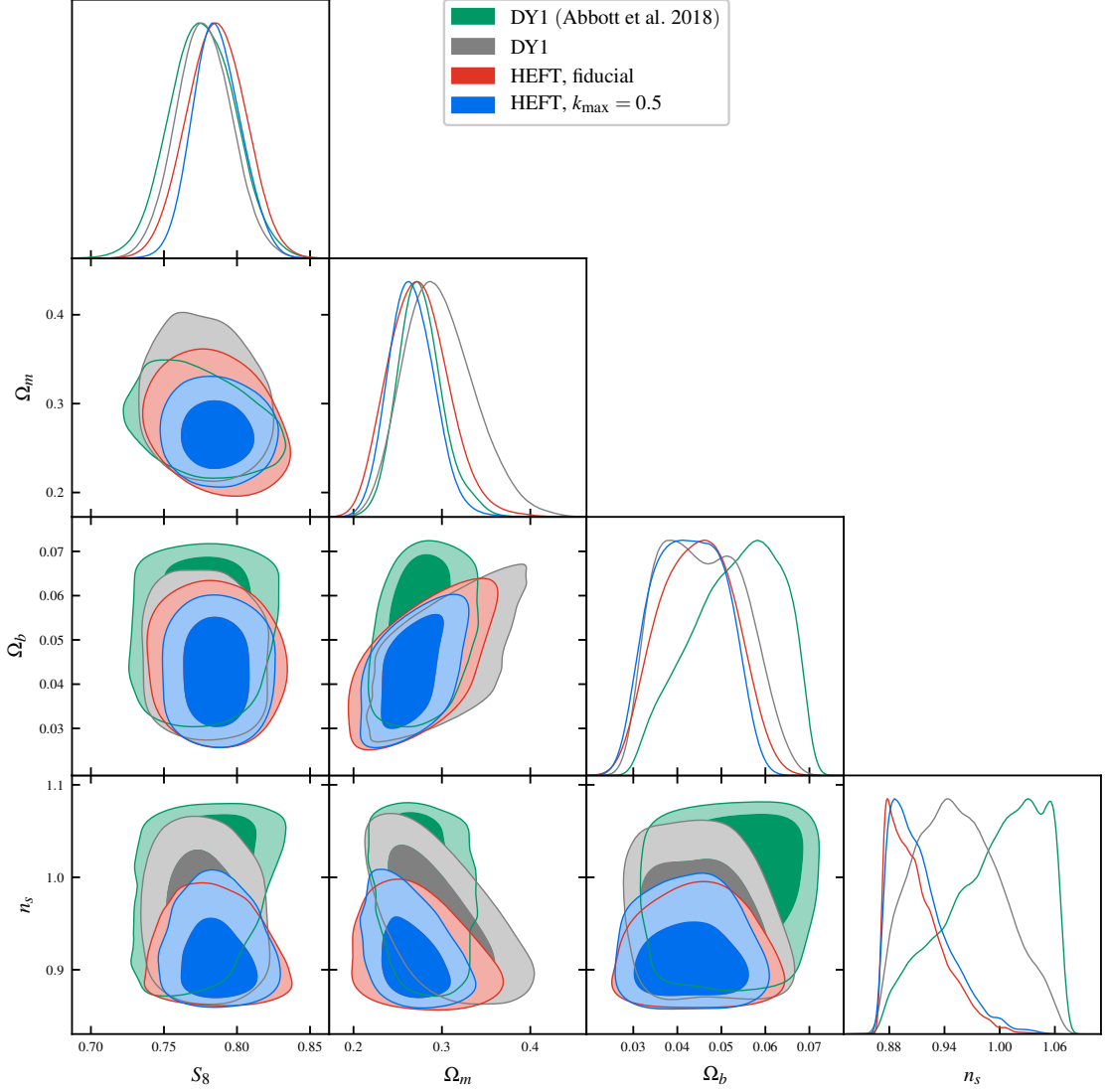
Figure 3: Contour plots showing the constraints on cosmological parameters using a linear bias model, also adopted in the DES Y1 analysis, in gray and the hybrid model presented in this work with $k_{\max} = 0.3$ and 0.5 Mpc^{-1} in red and blue, respectively. In green, we also show the parameter constraints obtained by the DES collaboration in their 3×2-point analysis [2]. We see that the hybrid approach is able to place tighter constraints on Ω_m and S_8, while n_s and Ω_b remain largely unconstrained, in agreement with [2] (see Section 4.3 for a more detailed discussion about n_s). The DY1 linear bias constraints on Ω_m and S_8 are consistent with each other and exhibit certain differences attributable to the following: our analysis is performed in harmonic space; it uses a different cosmological parametrization, and additionally does not vary massive neutrino energy density. For a more quantitative assessment, see Table 3.

where $\vec{\theta}$ is the set of parameters to be constrained, $\mathbf{d}$ is a data vector of power spectra, C is its covariance matrix, $\mathbf{t}(\vec{\theta})$ is the theory prediction, $p_p(\vec{\theta})$ is the prior distribution, and K is

Parameter	Prior	Parameter	Prior
$\Omega_c h^2$	[0.08, 0.16]	A_ϵ	$\mathcal{N}(1, 0.1)$
$\Omega_b h^2$	[0.013, 0.031]	$A_{\rm IA}$	[-5, 5]
σ_8	Free	$\eta_{\rm IA}$	[-5, 5]
n_s	[0.87, 1.07]	m_i	Table I in [2]
$100\theta_*$	1.041533	Δz_g^i	Table I in [2]
$b_0^i, b_1^i, b_2^i, b_s^i, b_\nabla^i$	[-5, 5]	Δz_s^i	Table I in [2]

Table 2: Model parameters and priors. An index i denotes parameters with independent copies in each galaxy clustering or shear redshift bin. The definition of the nuisance parameters $(A_{\rm IA}, \eta_{\rm IA}, m_i, \Delta z_g^i, \Delta_s^i)$ can be found in DY1 [2]. Square brackets denote a flat prior, whereas $\mathcal{N}$ denotes a Gaussian prior. In addition to A_ϵ, the latter is adopted also for $m_i, \Delta z_g^i, \Delta_s^i$. The comoving angular distance to last scatter, θ_*, is held fixed in the fiducial case and loosened in one of our tests (see Table 3).

a normalization constant. We sample this likelihood with a modified version of the MCMC sampler `MontePython`[9] [53], using the Core Cosmology Library [54] to calculate all angular power spectra contributing to $\mathbf{t}(\vec{\theta})$.

In our fiducial case, $\mathbf{d}$ contains all galaxy auto-correlations, and all galaxy-shear and shear-shear correlations. We impose the following scale cuts: in all cases, shear-shear power spectra are used on scales $\ell < 2000$, to avoid the impact of baryonic effects [3]. Galaxy-galaxy and galaxy-shear correlations, on the other hand, are cut on $\ell < k_{\max}\bar\chi$, where $\bar\chi$ is the distance to the mean redshift of the corresponding galaxy clustering sample, and $k_{\max}$ is a comoving cutoff scale. The fiducial value of $k_{\max}$ is 0.3 Mpc^{-1}, but we will consider values in the range $[0.15, 0.6]$ Mpc^{-1}.

Our model is described by a number of cosmological, bias and nuisance parameters. We consider variations in five cosmological parameters: the cold dark matter and baryon densities $(\Omega_c h^2, \Omega_b h^2)$, the primordial tilt n_s, the amplitude of density fluctuations σ_8, and the distance to the surface of last scattering θ_*. In the fiducial case we fix the latter to the value used in the AbacusSummit $100\,\theta_* = 1.041533$, but we also explore the impact of freeing this parameter.

Our fiducial bias model is defined by the set of EFT bias parameters $(b_1, b_2, b_s, b_\nabla)$. We assign different bias parameters for the 5 different galaxy redshift bins, for a total of 20 free parameters. In order to reproduce the linear bias model used in DY1, we also consider a single free bias parameter b_0 per redshift bin, while keeping all other bias parameter fixed to zero (note that $b_0 = 1$ in the HEFT model). When reproducing the DY1 analysis we used a scale cut $k_{\max} = 0.15$ Mpc^{-1}, roughly corresponding to the inverse of the minimum comoving separation used in DY1. As an extension to the 4-parameter HEFT model, we also marginalized over a stochastic bias parameter with a flat power spectrum. We do so by scaling the shot noise power spectrum with a free amplitude parameter A_ϵ with a 10% Gaussian prior centered on $A_\epsilon = 1$. The width of the prior was determined by comparing the noise power spectrum estimated as described in Section 2.2 with the galaxy power spectrum at the high-ℓ, noise-dominated regime.

Apart from the bias and cosmological parameters, we also vary 15 nuisance parameters,

---

describing other sources of systematic uncertainty in the DES Y1 data. These include shifts in the mean of the redshift distribution in each tomographic bin, multiplicative shear bias parameters, and a two-parameter intrinsic alignment model. The details are described in the DY1 paper [2], and we use the same priors specified there. The full set of model parameters and priors used are listed in Table 2.

4.2 Comparison between the linear and HEFT bias models

One of our main goals is comparing the final constraints achievable through a linear bias model in a reduced range of scales (as done e.g. in DY1), with those found using the HEFT model using data on smaller, mildly non-linear scales. For concreteness, the parameter space in both cases is defined as follows:

- DY1: we vary a single linear bias parameter b_0 in each galaxy clustering redshift bin. We also vary four cosmological parameters $(\Omega_c h^2, \Omega_b h^2, n_s, \sigma_8)$, and 15 nuisance parameters, for a total of 24 free parameters.

- HEFT: we vary four bias parameters $(b_1, b_2, b_s, b_\nabla)$ in each redshift bin, while keeping $b_0 = 1$, leading to a 39-dimensional parameter space when combined with the cosmological and nuisance parameters.

The linear bias model will be restricted to comoving scales $k_{\mathrm{max}} \lesssim 0.15$ Mpc^{-1}, while we will present results for the HEFT case as a function of k_{max}. Note that, although we label it "DY1", the linear bias model is slightly different from that used by [2] since, while the Hubble constant was a free parameter in their analysis, in the fiducial case we determine it by holding θ_* fixed.

We will quantify the goodness of fit of a given bias model in terms of the probability-to-exceed (PTE) of the model's minimum χ^2 value. The PTE depends on the total number of degrees of freedom ν, which in turn depends on the number of free parameters in the model. For a linear model with N_θ linearly independent free parameters with unconstrained priors, ν would be simply $N_{\mathrm{data}} - N_\theta$. In the presence of non-linear parameters and tight priors, the definition of ν is less clear. Here we use an effective number of degrees of freedom ν_{eff} determined as follows. We generate a synthetic data vector drawn from a multi-variate Gaussian distribution with a mean given by the theoretical prediction for a set of fiducial parameters θ_{fid}, and the power spectrum covariance described in Section 2.2. We then run a χ^2 minimizer varying the cosmological, bias and nuisance parameters, to find the best fit parameters θ_{bf}, and compute the difference in χ^2 between both sets of parameters $\Delta\chi^2 \equiv \chi^2(\theta_{\mathrm{fid}}) - \chi^2(\theta_{\mathrm{bf}})$. ν_{eff} is then given by the median of $\Delta\chi^2$ for several realizations of the synthetic data vector. Through this method we find that ν_{eff} is well approximated by $\nu_{\mathrm{eff}} \simeq N_{\mathrm{data}} - (N_b + 3)$, where N_b is the number of bias parameters in the model and N_{data} is the number of data points. Each bias parameter is effectively an independent parameter with a broad flat prior affecting the model at the linear level, and thus should add $+1$ to the total ν_{eff}. The contribution from all other parameters is effectively $\Delta\nu_{\mathrm{eff}} = 3$, either due to their tight priors (in the case of calibrated nuisance parameters), or their limited or correlated impact on the predicted cosmic shear and galaxy clustering observables.

The main result of this analysis is shown in Fig. 3, which presents the constraints on the four cosmological parameters Ω_b, Ω_c, $S_8 \equiv \sigma_8(\Omega_m/0.3)^{1/2}$, and n_s, for both bias models, with the HEFT results shown for $k_{\mathrm{max}} = 0.3$ Mpc^{-1}(fiducial case) and $k_{\mathrm{max}} = 0.5$ Mpc^{-1}. In the (Ω_m, S_8) projection, we observe a notable improvement in the constraints on Ω_m. The 68%

confidence interval shrinks from $\Omega_m = 0.298^{+0.033}_{-0.045}$ in the linear bias case to $\Omega_m = 0.273^{+0.030}_{-0.036}$ (15% improvement from DY1) and $\Omega_m = 0.266^{+0.024}_{-0.027}$ (35% improvement from DY1) using HEFT with $k_{\max} = 0.3$ Mpc^{-1} and 0.5 Mpc^{-1} respectively. The improvement on S_8 is less striking (about 10%). The figure also shows that the scalar spectral index n_s is pushed significantly towards its lower prior bound. As we show in Section 4.3, this is most likely due to the incorrect parameter dependence of the power spectrum templates used, although this does not alter the results found for (Ω_m, S_8). These results are also summarized in Table 3, together with the constraints founds for all other data and model configurations explored here.

It is important to note that, although our results are based on an independent reanalysis of the DY1 dataset in Fourier space, our constraints using a linear bias model are in excellent agreement with those reported by DES [2]. The estimated power spectra and best-fit predictions for the linear bias and HEFT model are shown in Fig. 4. Results are shown for the galaxy auto-correlations (labelled gg here) and the cross-correlations with cosmic shear (labelled gs). Overall the agreement on large scales is good. The HEFT bias model is able to describe the data on small scales, while the linear model does not capture the small-scale clustering. In the case of the gs we find specific cross-correlations that the model has difficulty fitting, particularly in the lower-left part of the figure, corresponding to cross-correlations where a significant fraction of the lens sample lies behind the source bin. This is likely due to residual systematics in the characterization of the source and lens redshift distributions that are not well captured by the nuisance parameters. Nevertheless, the overall goodness of fit is acceptable, with PTEs above 4.5% (see Section 4.3). The upper panel of Fig. 4 shows, in gray, the shot-noise contribution to the clustering auto-correlations. As we go to higher redshifts, the extended scale range used here lies partially within the noise-dominated regime, limiting the amount of information that can be extracted from the small-scale regime. The use of denser samples, at the cost of broader photomeric redshift uncertainties will likely benefit photometric clustering analyses making use of mildly non-linear scales [55].

Fig. 5 shows the power spectrum residuals as a fraction of the 1σ uncertainties in the same cases. The linear bias analysis assumes $k_{\max} = 0.15$ Mpc^{-1}, while the HEFT model has $k_{\max} = 0.5$ Mpc^{-1}. In the majority of cases, the data lie within 1-2σ of the theory in their respective scale ranges, with a small number of exceptions that do not spoil the overall goodness of fit of the HEFT and linear bias models. Importantly, the linear DY1 model shows a significant deviation from the gg data on small scales, which HEFT is able to capture adequately.

We next explore our results as a function of the scale cut $k_{\max}$. The recently developed HEFT N-body emulators [48, 52] was found to fit the halo power spectrum to sub-percent accuracy down to scales $k_{\max} \sim 0.6~h\,\mathrm{Mpc}^{-1}$, drilling significantly deeper into the non-linear regime, where various assembly bias effects, related to halo concentration, occupation, local environment, and spin, are known to affect the clustering properties [56–59]. In particular, the hybrid approach has been shown to be effective in describing the clustering of more complex tracer populations (see Section 6.3 of [48]), whereas more traditional methods lead to errors larger than 1% at $k \approx 0.2$ Mpc^{-1}. The improvement in final constraints due to the additional information gained from these modes is offset by the extra freedom allowed by the HEFT model, and eventually limited by the ability of the model to describe the clustering of galaxies. Therefore it is interesting to explore the evolution of the parameter uncertainties and goodness of fit with $k_{\max}$.

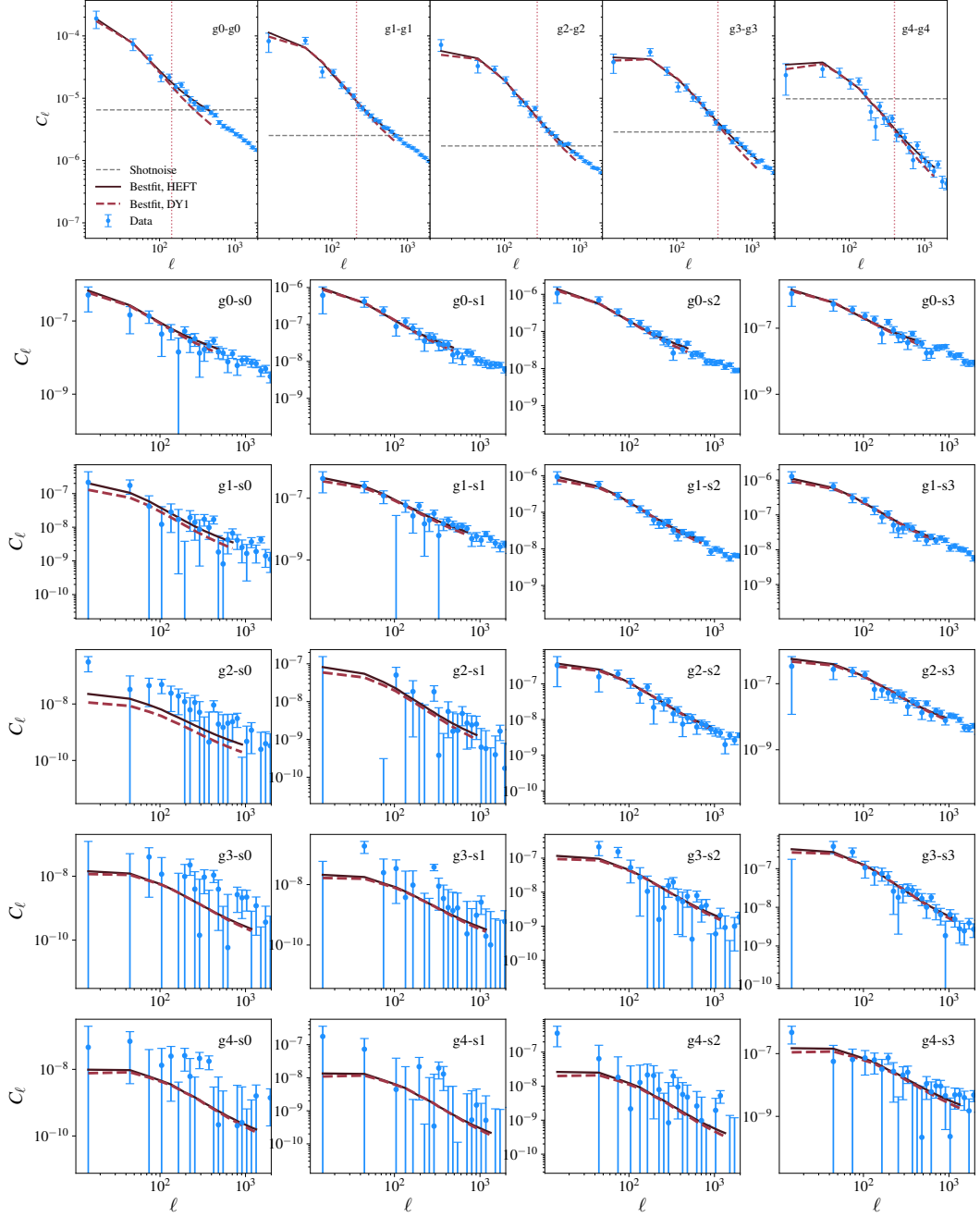In Fig. 6, we present the joint constraints on Ω_m and S_8 for the HEFT model with $k_{\max}$

Figure 4: Measured galaxy-galaxy ("gi-gj", top panel) and galaxy-shear ("gi-sj", bottom panel) power spectra (blue). Here, i and j denote the different tomographic bins (see Fig. 1). The solid black and dashed red lines show the theoretical predictions adopting the Hybrid EFT (HEFT) and the DY1 linear bias models respectively. Although we show these predictions up to scales corresponding to $k_{max} = 0.5$ Mpc^{-1}, the red dotted vertical line shows the scale cut $k_{max} = 0.15$ Mpc^{-1} used for the linear bias analysis. The gray dashed horizontal line indicates the Poisson shot-noise of the autocorrelations.

in the range $[0.3, 0.6]$ Mpc^{-1}, compared with the constraints found with the linear bias model
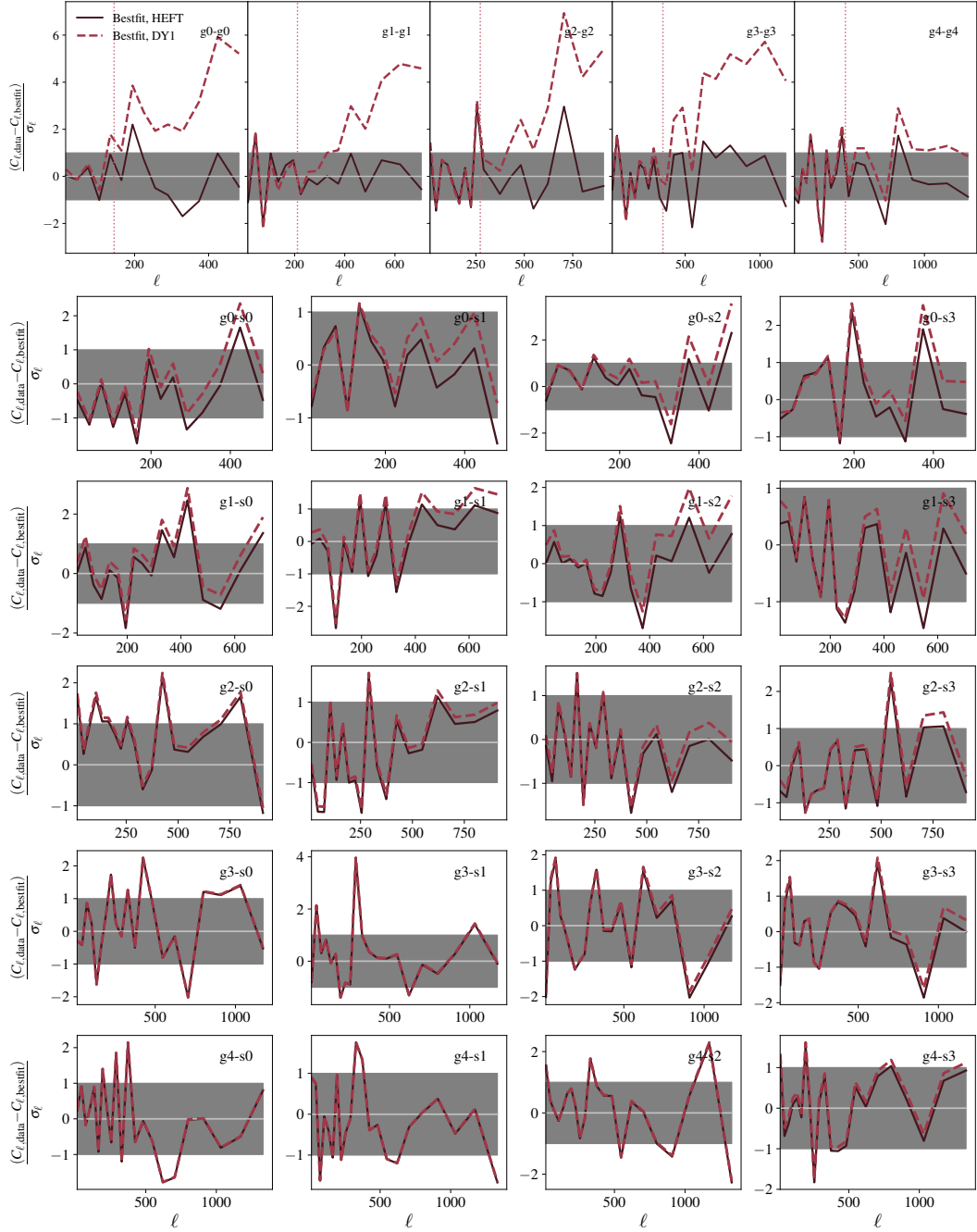
Figure 5: As in Fig. 4, we show the power spectrum residuals, for the galaxy-galaxy and shear-shear power spectra normalized by their 68% uncertainties. In solid black and dashed red, we show the residuals for the Hybrid EFT (HEFT) and the DY1 linear bias models respectively. Both agree on large scales, but the linear model is not able to fit the clustering data on scales smaller than the 0.15 Mpc^{-1} scale cut (marked by the vertical dotted red line).

up to $k_{\max} = 0.15$ Mpc^{-1}. As found before, the uncertainties, particularly on Ω_m, shrink steadily as $k_{\max}$ is increased. This improvement, however, seems to asymptote on scales $k_{\max} = 0.6$ Mpc^{-1}, where we recover constraints essentially equivalent to the $k_{\max} = 0.5$

Model	$\chi^2/\nu_{\rm eff}$	S_8	Ω_m	n_s	H_0
DY1	470.9/467	0.778 ± 0.019	$0.298^{+0.033}_{-0.045}$	$0.956^{+0.042}_{-0.055}$	–
HEFT, fiducial	583.5/577	0.786 ± 0.020	$0.273^{+0.030}_{-0.036}$	$0.910^{+0.012}_{-0.038}$	–
HEFT, $k_{\rm max} = 0.4$	650.0/632	0.781 ± 0.017	$0.279^{+0.025}_{-0.032}$	$0.913^{+0.013}_{-0.041}$	–
HEFT, $k_{\rm max} = 0.5$	702.8/682	$0.786^{+0.015}_{-0.018}$	$0.266^{+0.024}_{-0.027}$	$0.914^{+0.014}_{-0.040}$	–
HEFT, $k_{\rm max} = 0.6$	733.2/717	$0.790^{+0.016}_{-0.018}$	$0.261^{+0.021}_{-0.032}$	$0.914^{+0.017}_{-0.040}$	–
HEFT, $k_{\rm max} = 0.15$	458.8/452	0.785 ± 0.020	$0.294^{+0.037}_{-0.050}$	$0.916^{+0.017}_{-0.042}$	–
HEFT, fixed $P_{ij}(k)$	585.4/577	0.788 ± 0.021	$0.265^{+0.023}_{-0.026}$	$0.920^{+0.024}_{-0.039}$	–
HEFT, marg. stochastic	585.0/577	0.784 ± 0.019	$0.279^{+0.028}_{-0.039}$	$0.913^{+0.015}_{-0.039}$	–
HEFT, b_1, b_2, b_s	584.4/582	0.782 ± 0.018	$0.270^{+0.026}_{-0.034}$	$0.925^{+0.021}_{-0.049}$	–
HEFT, b_1, b_2	589.2/587	0.775 ± 0.017	$0.267^{+0.023}_{-0.033}$	$0.945^{+0.034}_{-0.053}$	–
DY1, gg, gs	258.3/227	0.781 ± 0.042	$0.279^{+0.031}_{-0.061}$	$0.995^{+0.074}_{-0.026}$	–
HEFT, gg, gs	382.9/337	$0.777^{+0.032}_{-0.038}$	$0.299^{+0.038}_{-0.045}$	$0.9132^{+0.0099}_{-0.043}$	–
DY1, H_0	470.3/467	0.777 ± 0.019	$0.299^{+0.036}_{-0.057}$	$0.960^{+0.043}_{-0.061}$	68.6 ± 6.6
HEFT, $H_0, k_{\rm max} = 0.5$	703.6/682	0.785 ± 0.017	$0.264^{+0.025}_{-0.032}$	$0.913^{+0.015}_{-0.040}$	$70.7^{+3.0}_{-3.5}$

Table 3: Constraints (68% confidence level) on the cosmological parameters S_8, Ω_m, and n_s for the different data and model configurations considered in this study. We also list the best-fit χ^2 values and effective degrees of freedom $\nu_{\rm eff}$ for each case as a measure of goodness of fit. The definition of $\nu_{\rm eff}$ is discussed in the main text. Rows marked "DY1" use a linear bias parametrization, while the others use the HEFT model. The main result is an improvement of $\sim (35\%, 10\%)$ in the parameter uncertainties for (Ω_m, S_8) when using the HEFT model on an extended range of scales. The quoted values of the scale cut $k_{\rm max}$ are in units of $\rm Mpc^{-1}$.

$\rm Mpc^{-1}$ case, in spite of adding 35 additional data points. These results are also summarized in Fig. 7, which shows the marginalized constraints on cosmological and bias parameters as a function of bias model and $k_{\rm max}$. The constraints found for different models and scale cuts are broadly consistent with each other. Other than Ω_m and S_8, all other parameters do not benefit significantly from the extended scale range. The bias parameters, however, particularly b_2 and b_∇, are significantly better constrained by the small-scale modes.

The bottom panel of Figure 7 shows the χ^2 PTE for the different cases explored here. Based on the results of [48] we do not expect the HEFT model to be valid far beyond $k_{\rm max} \simeq 0.6 \, h \, \rm Mpc^{-1}$. Nevertheless, we find that the HEFT model is able to describe the data down to the smallest scale explored, which expressed in the *little-h* units of [48] corresponds to a bone shaking $k_{\rm max} \simeq 0.86 \, h \, \rm Mpc^{-1}$. Of course, this statement depends on the statistical power of the data used, and will likely change with future more precise datasets.

4.3 Tests and validation

To validate the results presented in the previous section we have carried out a number of additional tests, exploring the contents of the HEFT bias model, the parameter dependence of the power spectrum templates used, and the impact of stochastic bias terms.
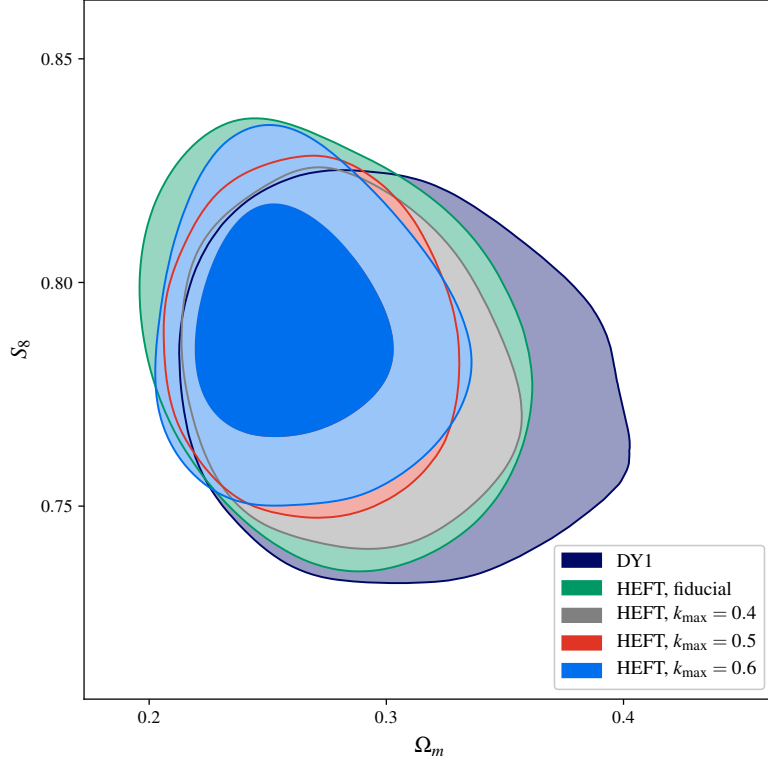
Figure 6: Constraints on the matter fraction Ω_m, and the amplitude of matter fluctuations, parametrized by $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$. Results are shown using the linear bias model with $k_{\max} = 0.15$ Mpc^{-1} (dark blue), and the HEFT model with $k_{\max} = 0.3$, 0.4, 0.5, 0.6 Mpc^{-1}, (green, gray, red and blue respectively). As we go to higher $k_{\max}$, the hybrid approach is able to extract additional cosmological information, especially on Ω_m.

4.3.1 The HEFT ingredients

The HEFT model adds 3 additional free parameters compared with the linear bias parametrization. However, some of these additional degrees of freedom may not be necessary to describe the data on a given range of scales, and it is therefore interesting to explore the possibility of simplifying the model by excluding some of these parameters. In particular, the tidal bias and non-local bias terms (b_s and b_∇) describe the impact of the local tidal forces, and of physical processes on scales smaller than the characteristic scale for galaxy formation (e.g. the Lagrangian halo size), both of which are arguably subdominant to the impact of the local environmental density [60, 61].

To explore this, we repeated our analysis removing b_∇ alone and together with b_s in our fiducial case with $k_{\max} = 0.3$ Mpc^{-1}. The results from this exercise are summarized in Table 3. The full-bias model does not provide any significant improvement in the fit ($\Delta\chi^2 = (0.9, 5.7)$ for 5 and 10 additional parameters respectively). Nevertheless, we find compatible constraints in all cases, which reassures that the additional complexity of the full HEFT model does not degrade the final cosmological constraints significantly. This is to be expected, since non-local effects are likely subdominant these scales.
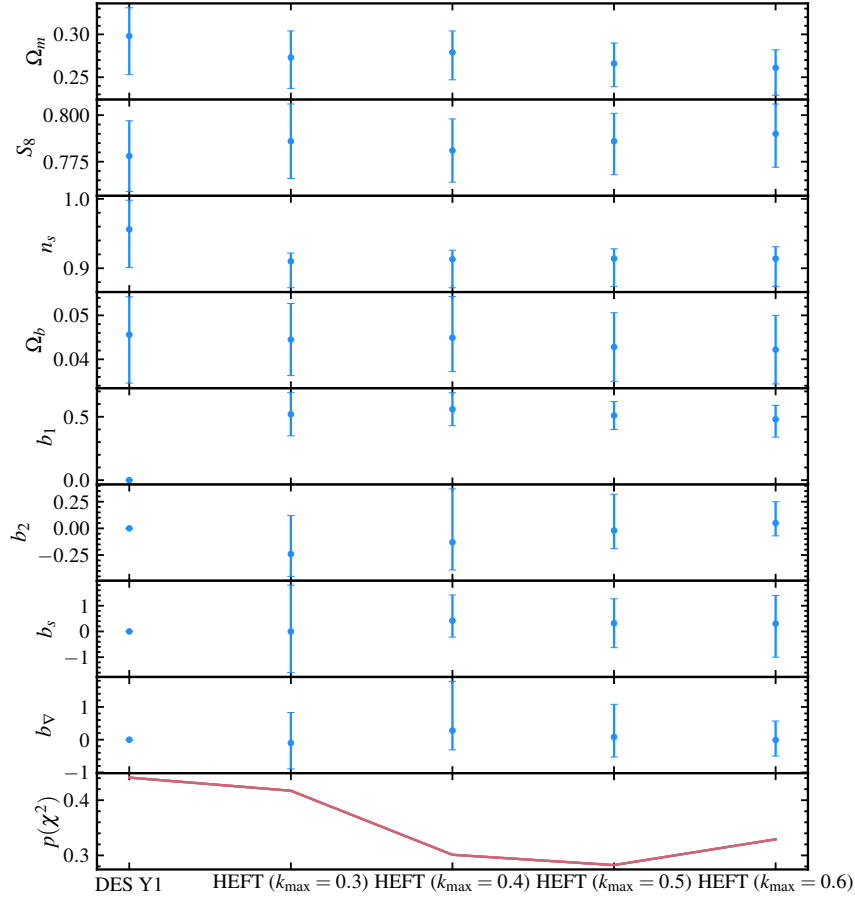
Figure 7: 68% confidence level constraints on four cosmological parameters (Ω_m, S_8, n_s, Ω_b) and four bias parameters (b_1, b_2, b_s, b_∇) for the median tomographic redshift bin of the DES Y1 galaxy sources (see Fig. 1). The lowermost plot shows the best-fit χ^2 PTEs for the 5 models considered: the linear bias model of the DY1 analysis [2] and the hybrid model (HEFT) with $k_{\max} = 0.3$, 0.4, 0.5, and 0.6 Mpc^{-1}. Although the PTEs decrease, as we go to higher $k_{\max}$, they remain above 30% throughout. Overall the parameter constraints get tighter, as we go from left to right, implying that the hybrid approach is able to place stricter constraints on both the cosmological parameters (most noticeably Ω_m) and the bias parameters. Note that the Lagrangian bias parameters are set to zero in the DES Y1 linear model.

4.3.2 The parameter dependence of HEFT templates

One of the caveats of the analysis carried out here is the simplified method used to characterize the dependence of the HEFT power spectrum templates on cosmological parameters. As described in Section 3.3, the ratios of the templates with respect to the matter power spectrum are only allowed to vary linearly with respect to the cosmological parameters around the

fiducial cosmology of the ABACUSSUMMIT suite.

To quantify the impact of this approximation we repeat our analysis for $k_{\max} = 0.3$ Mpc^{-1} removing this linear parameter dependence altogether, and assuming constant power spectrum template ratios. In this cruder approximation, all parameter dependence of the power spectrum templates is captured by the matter power spectrum. The results, listed in Table 3, show that the constraints on the primary large-scale structure parameters S_8 and Ω_m are in good agreement with those found in the fiducial analysis, and that the parameter dependence of the template ratios does not improve them significantly. The reduction in uncertainties found before is therefore not artificially caused by incorrectly accounting for this parameter dependence.

However, we find that the constraints on n_s in both cases, although still compatible at 1σ, show significantly different uncertainties. We believe this effect is indeed caused by an inaccurate modeling of the n_s dependence, where the linear approximation assumed here breaks down more readily. To confirm that the improved uncertainties on n_s are indeed artificial, and not caused by the additional small-scale information enabled by the HEFT model, we repeat our analysis in the reduced scale range $k < 0.15\ \mathrm{Mpc}^{-1}$ used with the linear bias model. The results of this test are also listed in Table 3, and show that, even in this reduced range, we obtain tighter constraints on n_s than those found with the simpler linear bias model. This confirms that the dependence of the power spectrum template ratios on n_s is not correctly captured with our simplified setup. All future analyses using the HEFT model should therefore rely on full emulators where this dependence is correctly captured. We emphasize, however, that the constraints found in the (Ω_m, S_8) plane are unaffected by this approximation, and indeed the HEFT results in the reduced scale range agree rather well with those found for the linear bias model for these parameters.

Another aspect of the cosmological dependence not captured by our current implementation is the dependence on the local expansion rate H_0. Since θ_* is the same for all the ABACUS-SUMMIT simulations used here, H_0 was treated as a derived parameter, determined in terms of the other cosmological parameters by fixing θ_* to the ABACUSSUMMIT value. Although we do not expect this to impact our results significantly, since the DY1 analysis was not able to constrain H_0, we have repeated our analysis treating H_0 as a free parameter. This was done for the DY1 linear bias model with $k_{\max} = 0.15\ \mathrm{Mpc}^{-1}$, and for the HEFT model with $k_{\max} = 0.5$ Mpc^{-1}. Due to the limitation we just described, we are not able to account for this parameter dependence in the HEFT power spectrum ratios, and therefore the dependence is only included in the `HALOFIT` matter power spectrum. The results, listed in Table 3, show that the constraints on (Ω_m, S_8) are not affected by the additional parameter freedom. Interestingly, we find that we are able to recover tighter constraints on H_0 ($\sigma(H_0) \simeq 3.3\,\mathrm{km}\,s^{-1}\,\mathrm{Mpc}^{-1}$) than those found in the linear analysis ($\sigma(H_0) \simeq 6.6\,\mathrm{km}\,s^{-1}\,\mathrm{Mpc}^{-1}$). Since the dependence on H_0 may be imperfectly captured by our implementation, we cannot make a claim regarding this result. However, the possibility of constraining H_0 from the joint analysis of galaxy clustering and weak lensing on the smaller scales enabled by the HEFT bias model, should be explored in the future.

4.3.3 Stochasticity

As described in Section 3.2, we expect the galaxy overdensity to receive a stochastic contribution sourced by the small-scale physics governing galaxy formation and evolution. This contribution should dominate the galaxy power spectrum on small scales and, as shown in Fig. 1, dominates the deterministic signal on angular scales $\ell \gtrsim 1000$. By inspecting the mea-

sured power spectra on these scales, we find that the stochastic contribution is well described by a Poisson-like process, with $C_\ell^{\rm SN} = 1/\bar{n}_\Omega$ after accounting for the effects of inhomogeneous sky coverage, where $\bar{n}_\Omega$ is the number of galaxies per steradian. Nevertheless, to allow for possible residual stochastic contributions, we repeat our analysis allowing for an additional free amplitude parameter A_ϵ per redshift bin multiplying this shot-noise contribution, and marginalize over it with a conservative 10% Gaussian prior centered on $A_\epsilon = 1$. Since A_ϵ is a linear parameter, this marginalization is done analytically by simply modifying the covariance matrix as:

$$\text{Cov}_{\ell\ell'} \longrightarrow \text{Cov}_{\ell\ell'} + \sigma_A^2 C_\ell^{\rm SN} C_{\ell'}^{\rm SN}, \tag{4.2}$$

where $\sigma_A = 0.1$.

The result of this test, listed in Table 3, shows that the effect of residual stochastic contributions is negligible, both in the best-fit parameters and their uncertainty.

4.3.4 Excluding shear

The three main components of the data vector used here are the galaxy-galaxy, galaxy-shear, and shear-shear correlations (*gg*, *gs* and *ss* respectively). Of these, *ss* is not sensitive to any modeling improvement brought about by the use of the HEFT model. In order to better isolate the parameter constraint enhancement, and the goodness of fit associated with the HEFT model, we repeated our analysis excluding the *ss* component from the data. This was done both for the DY1 linear bias model with $k < 0.15$ Mpc^{-1}, and for the HEFT model with $k_{\rm max} = 0.3$ Mpc^{-1}. The results are shown in Table 3.

When excluding the shear-shear data, the uncertainty on S_8 grows by $\sim 60\%$ for both bias models, while the constraints on Ω_m grow by about $\sim 20\%$. This shows that the inclusion of galaxy clustering information plays a vital role in constraining Ω_m [2]. The relative improvement in constraints associated with the HEFT model is similar in both cases. Excluding the effects of intrinsic alignments, cosmic shear is an unbiased tracer of the matter fluctuations. Thus, in spite of its significantly lower signal-to-noise ratio compared with galaxy clustering, the shear power spectrum is able to obtain comparatively tighter constraints.

It is worth noting that the goodness of fit of the *gg-gs* sector is notably lower for both bias models than that of the full data vector, with a minimum PTE of 4.5%. Although the statistical significance is not high, this may be a sign of unmodelled systematics in the data. A visual inspection of the data suggests that it is impossible for the model to fit some of the galaxy-shear power spectra corresponding to lens samples partially behind the sources (e.g. "g2-s0" or "g3-s0" in Fig. 1). This is a feature of both the linear and HEFT models, and could be ascribed to an imperfect model of the source or lens redshift distributions.

5 Discussion & Conclusions

This paper presents the first application of the Hybrid EFT (HEFT) bias model of [1] to observational data. In particular, we re-analyzed the combination of lensing shear and projected galaxy clustering from the DES Y1 data release. This methods holds promise to be the method of choice for next generation surveys, since it efficiently combines the reach of N-body simulations deep into the non-linear regime with the analytical exactness and theoretical control of bias expansions. We note that photometric surveys are especially well suited for the first application of this method, since we do not require a precise model for redshift-space distortions.

We calibrated our HEFT basis power spectra from the ABACUSSUMMIT N-body simulations, smoothly interpolating into large-scale analytical solutions and using a linear expansion around the fiducial model to account for the cosmology dependence. In this preliminary work we have focused on the goodness-of-fit as basic quantity by which to choose whether a model produces a reliable fit.

Our results can be summarized as follows:

- HEFT models offer a good fit to the data. Nominally, the χ^2 is good all the way to $k_{\max} = 0.6 \ \mathrm{Mpc}^{-1}$

- HEFT models offer a stable fit to the data. While the uncertainties decrease, the inferred values of all parameters remain consistent as we push to higher $k_{\max}$;

- Given the enormous increase in signal-to-noise, as we increase the $k_{\max}$, the improvement on the cosmological parameters of interest is rather modest. The standard interpretation is that all the new information is going mostly into determining the bias parameters rather than the cosmology. In some sense this method offers "graceful transition to ignorance", where the information in the 2-point function is being exhausted and the results do not depend strongly on the choice of scale cut. This observation also implies that a 3-point function over the same scales could break further degeneracies and bring concrete improvements in our determination of cosmological parameters.

- While the constraint on S_8 improves modestly (about $\sim 10\%$) for $k_{\max} = 0.5 \ \mathrm{Mpc}^{-1}$, we do notice a significant improvement in the uncertainty on Ω_m, which shrinks by about 35% to $\Omega_m = 0.266^{+0.024}_{-0.027}$. This implies that HEFT starts to break the degeneracy in the less constraining direction of the "weak-lensing banana" contour on the $\sigma_8 - \Omega_m$-plane. Interestingly, our results are $\sim 2\sigma$ away from the central value found by *Planck*, $\Omega_m = 0.315 \pm 0.007$ [50].

- In our fits, we have left Ω_b, n_s and H_0 (with fixed angular diameter distance to the last scattering) as free parameters. While it is not expected that the combination of cosmic shear and galaxy clustering at the level of two-point functions can measure these with significant precision, it is nevertheless re-assuring that the parameters were bound by the data, rather than the prior and that Planck-determined values lie well withing the posteriors spanned by them.

- While analyzing the data, we noticed that the χ^2 values for the gg and gs power spectra are somewhat high. However, the statistical significance is small when accounting for the effective number of fitted degrees of freedom.

We reiterate that the work presented here is fundamentally exploratory in nature and our constraints should not be taken to be as robust as those supported by extensive testing on realistic mock datasets. There are numerous steps where our method can be improved.

Most importantly, the exact $k_{\max}$ limits should be obtained based on mock galaxy catalogs populated with realistic distribution of galaxies, rather than relying solely on the goodness of fit. Ideally, such mock catalogs should be based on an independent suite of N-body simulations spanning plausible cosmologies and recipes for populating halos with galaxies. Methodologically, the way templates are created leaves several open questions in terms of how basis spectra are generated and interpolated. At the moment we have chosen the pivot points for interpolation between analytic and N-body solutions essentially "by hand". Basis

spectra involving δ_L^2 and $\nabla\delta_L$ are noisy and may show a weak dependence on the smoothing scale (not used in this paper). The sample variance in the basis spectra is poorly understood. As we were building our prediction scheme, independent emulators were developed by [48] and [52]. Repeating our analysis by replacing our simple linear parametrization (see Eq. 3.13 with such an emulator would be a very interesting cross-check that would quantify the systematic errors in our theory model. We leave this, and the other caveats listed in this paragraph, for future work. As the power of the HEFT approach becomes apparent, we expect that the coming years will bring about more sophisticated methods to measure and interpolate the relevant quantities from N-body simulations.

Despite these caveats, HEFT worked on DES Y1 data "out of the box", producing a good and stable fit without any tweaking. Our results confirm the robustness of DES Y1 data and demonstrate the promise of the upcoming generation of photometric galaxy surveys when analyzed with state-of-the-art theory prediction tools.

Acknowledgements

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

References

[1] C. Modi, S.-F. Chen and M. White, *Simulations and symmetries*, MNRAS **492** (2020) 5754 [1910.07097].

[2] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen et al., *Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing*, Phys. Rev. D **98** (2018) 043526 [1708.01530].

[3] C. Hikage, M. Oguri, T. Hamana, S. More, R. Mandelbaum, M. Takada et al., *Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data*, PASJ **71** (2019) 43 [1809.09148].

[4] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi et al., *KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, arXiv e-prints (2020) arXiv:2007.15632 [2007.15632].

[5] J. Elvin-Poole, M. Crocce, A. J. Ross, T. Giannantonio, E. Rozo, E. S. Rykoff et al., *Dark Energy Survey year 1 results: Galaxy clustering for combined probes*, Phys. Rev. D **98** (2018) 042006 [1708.01536].

[6] M. P. van Daalen, J. Schaye, C. M. Booth and C. Dalla Vecchia, *The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology*, MNRAS **415** (2011) 3649 [1104.1174].

[7] E. Semboloni, H. Hoekstra, J. Schaye, M. P. van Daalen and I. G. McCarthy, *Quantifying the effect of baryon physics on weak lensing tomography*, MNRAS **417** (2011) 2020 [1105.1075].

[8] T. Eifler, E. Krause, S. Dodelson, A. R. Zentner, A. P. Hearin and N. Y. Gnedin, *Accounting for baryonic effects in cosmic shear tomography: determining a minimal set of nuisance parameters using PCA*, MNRAS **454** (2015) 2451 [1405.7423].

[9] H.-J. Huang, T. Eifler, R. Mandelbaum and S. Dodelson, *Modelling baryonic physics in future weak lensing surveys*, MNRAS **488** (2019) 1652 [1809.01146].

[10] A. Schneider, R. Teyssier, J. Stadel, N. E. Chisari, A. M. C. Le Brun, A. Amara et al., *Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation*, J. Cosmology Astropart. Phys. **2019** (2019) 020 [1810.08629].

[11] N. E. Chisari, M. L. A. Richardson, J. Devriendt, Y. Dubois, A. Schneider, A. M. C. Le Brun et al., *The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation*, MNRAS **480** (2018) 3962 [1801.08559].

[12] A. A. Berlind and D. H. Weinberg, *The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass*, ApJ **575** (2002) 587 [astro-ph/0109001].

[13] J. A. Peacock and R. E. Smith, *Halo occupation numbers and galaxy bias*, MNRAS **318** (2000) 1144 [astro-ph/0005010].

[14] L. Senatore, *Bias in the effective field theory of large scale structures*, J. Cosmology Astropart. Phys. **2015** (2015) 007 [1406.7843].

[15] V. Desjacques, D. Jeong and F. Schmidt, *Large-scale galaxy bias*, Phys. Rep. **733** (2018) 1 [1611.09787].

[16] A. Nicola, D. Alonso, J. Sánchez, A. Slosar, H. Awan, A. Broussard et al., *Tomographic galaxy clustering with the Subaru Hyper Suprime-Cam first year public data release*, J. Cosmology Astropart. Phys. **2020** (2020) 044 [1912.08209].

[17] A. Nicola, C. García-García, D. Alonso, J. Dunkley, P. G. Ferreira, A. Slosar et al., *Cosmic shear power spectra in practice*, arXiv e-prints (2020) arXiv:2010.09717 [2010.09717].

[18] T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, J. Asorey et al., *The Dark Energy Survey: Data Release 1*, ApJS **239** (2018) 18 [1801.03181].

[19] E. Huff and R. Mandelbaum, *Metacalibration: Direct Self-Calibration of Biases in Shear Measurement*, arXiv e-prints (2017) arXiv:1702.02600 [1702.02600].

[20] E. S. Sheldon and E. M. Huff, *Practical Weak-lensing Shear Measurement with Metacalibration*, ApJ **841** (2017) 24 [1702.02601].

[21] J. Zuntz, E. Sheldon, S. Samuroff, M. A. Troxel, M. Jarvis, N. MacCrann et al., *Dark Energy Survey Year 1 results: weak lensing shape catalogues*, MNRAS **481** (2018) 1149 [1708.01533].

[22] B. Hoyle, D. Gruen, G. M. Bernstein, M. M. Rau, J. De Vicente, W. G. Hartley et al., *Dark Energy Survey Year 1 Results: redshift distributions of the weak-lensing source galaxies*, MNRAS **478** (2018) 592 [1708.01532].

[23] D. Alonso, J. Sanchez, A. Slosar and LSST Dark Energy Science Collaboration, *A unified pseudo-C_ℓ framework*, MNRAS **484** (2019) 4127 [1809.09603].

[24] C. García-García, D. Alonso and E. Bellini, *Disconnected pseudo-C_l covariances for projected large-scale structure data*, J. Cosmology Astropart. Phys. **2019** (2019) 043 [1906.11765].

[25] C. Doux, C. Chang, B. Jain, J. Blazek, H. Camacho, X. Fang et al., *Consistency of cosmic shear analyses in harmonic and real space*, MNRAS (2021) [2011.06469].

[26] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet and F. Hansen, *MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets*, ApJ **567** (2002) 2 [astro-ph/0105302].

[27] M. Tegmark, *How to measure CMB power spectra without losing information*, Phys. Rev. D **55** (1997) 5895 [astro-ph/9611174].

[28] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wand elt, F. K. Hansen, M. Reinecke et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, ApJ **622** (2005) 759 [astro-ph/0409513].

[29] J. N. Grieb, A. G. Sánchez, S. Salazar-Albornoz, R. Scoccimarro, M. Crocce, C. Dalla Vecchia et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic*

Survey: Cosmological implications of the Fourier space wedges of the final sample, MNRAS **467** (2017) 2085 [1607.03143].

[30] M. Takada and B. Jain, *The impact of non-Gaussian errors on weak lensing surveys*, MNRAS **395** (2009) 2065 [0810.4170].

[31] M. Takada and W. Hu, *Power spectrum super-sample covariance*, Phys. Rev. D **87** (2013) 123504 [1302.6994].

[32] G. Efstathiou, *Myths and truths concerning estimation of power spectra: the case for a hybrid estimator*, MNRAS **349** (2004) 603 [astro-ph/0307515].

[33] E. Krause and T. Eifler, *cosmolike - cosmological likelihood analyses for photometric galaxy surveys*, MNRAS **470** (2017) 2100 [1601.05779].

[34] A. Nicola, J. Dunkley and D. N. Spergel, *Joint cosmology and mass calibration from thermal Sunyaev-Zel'dovich cluster counts and cosmic shear*, Phys. Rev. D **102** (2020) 083505 [2006.00008].

[35] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, Phys. Rep. **340** (2001) 291 [astro-ph/9912508].

[36] E. Krause, T. F. Eifler, J. Zuntz, O. Friedrich, M. A. Troxel, S. Dodelson et al., *Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses*, arXiv e-prints (2017) arXiv:1706.09359 [1706.09359].

[37] D. N. Limber, *The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.*, ApJ **117** (1953) 134.

[38] N. Afshordi, Y.-S. Loh and M. A. Strauss, *Cross-correlation of the cosmic microwave background with the 2MASS galaxy survey: Signatures of dark energy, hot gas, and point sources*, Phys. Rev. D **69** (2004) 083524 [astro-ph/0308260].

[39] M. Kilbinger, C. Heymans, M. Asgari, S. Joudaki, P. Schneider, P. Simon et al., *Precision calculations of the cosmic shear power spectrum projection*, MNRAS **472** (2017) 2126 [1702.05301].

[40] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce et al., *Stable clustering, the halo model and non-linear cosmological power spectra*, MNRAS **341** (2003) 1311 [astro-ph/0207664].

[41] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, *Revising the Halofit Model for the Nonlinear Matter Power Spectrum*, ApJ **761** (2012) 152 [1208.2701].

[42] T. Matsubara, *Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture*, Phys. Rev. D **78** (2008) 083519 [0807.1733].

[43] S.-F. Chen, Z. Vlah, E. Castorina and M. White, *Redshift-Space Distortions in Lagrangian Perturbation Theory*, arXiv e-prints (2020) arXiv:2012.04636 [2012.04636].

[44] L. H. Garrison, D. J. Eisenstein, D. Ferrer, J. L. Tinker, P. A. Pinto and D. H. Weinberg, *The Abacus Cosmos: A Suite of Cosmological N-body Simulations*, ApJS **236** (2018) 43 [1712.05768].

[45] N. Maksimova, L. Garrison, D. Eisenstein, B. Hadzhiyska and S. Bose, ABACUSSUMMIT*: A Massive Set of High-Accuracy, High-Resolution N-Body Simulations*, In preparation [2104.XXXXX].

[46] L. Garrison, D. Eisenstein, N. Maksimova, D. Ferrer, B. Hadzhiyska, M. V. Metchnik et al., *The Abacus Cosmological N-body Code*, In preparation [2104.XXXXX].

[47] L. H. Garrison, D. J. Eisenstein and P. A. Pinto, *A high-fidelity realization of the Euclid code comparison N-body simulation with ABACUS*, MNRAS **485** (2019) 3370 [1810.02916].

[48] N. Kokron, J. DeRose, S.-F. Chen, M. White and R. H. Wechsler, *The cosmology dependence of galaxy clustering and lensing from a hybrid N-body-perturbation theory model*, arXiv e-prints (2021) arXiv:2101.11014 [2101.11014].

[49] S.-F. Chen, Z. Vlah and M. White, *Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory*, J. Cosmology Astropart. Phys. **2020** (2020) 062 [2005.00523].

[50] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, A&A **641** (2020) A6 [1807.06209].

[51] N. Kokron, J. DeRose, S.-F. Chen, M. White and R. H. Wechsler, *The cosmology dependence of galaxy clustering and lensing from a hybrid N-body-perturbation theory model*, arXiv e-prints (2021) arXiv:2101.11014 [2101.11014].

[52] M. Zennaro, R. E. Angulo, M. Pellejero-Ibáñez, J. Stücker, S. Contreras and G. Aricò, *The BACCO simulation project: biased tracers in real space*, arXiv e-prints (2021) arXiv:2101.12187 [2101.12187].

[53] T. Brinckmann and J. Lesgourgues, *MontePython 3: Boosted MCMC sampler and other features*, *Physics of the Dark Universe* **24** (2019) 100260 [1804.07261].

[54] N. E. Chisari, D. Alonso, E. Krause, C. D. Leonard, P. Bull, J. Neveu et al., *Core Cosmology Library: Precision Cosmological Predictions for LSST*, ApJS **242** (2019) 2 [1812.05995].

[55] A. Porredon, M. Crocce, P. Fosalba, J. Elvin-Poole, A. Carnero Rosell, R. Cawthon et al., *Dark Energy Survey Year 3 results: Optimizing the lens sample in a combined galaxy clustering and galaxy-galaxy lensing analysis*, Phys. Rev. D **103** (2021) 043503 [2011.03411].

[56] R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov and A. Dekel, *Concentrations of Dark Halos from Their Assembly Histories*, ApJ **568** (2002) 52 [astro-ph/0108151].

[57] L. Gao, V. Springel and S. D. M. White, *The age dependence of halo clustering*, MNRAS **363** (2005) L66 [astro-ph/0506510].

[58] R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov and B. Allgood, *The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation*, ApJ **652** (2006) 71 [astro-ph/0512416].

[59] B. Hadzhiyska, S. Bose, D. Eisenstein, L. Hernquist and D. N. Spergel, *Limitations to the 'basic' HOD model and beyond*, MNRAS **493** (2020) 5506 [1911.02610].

[60] D. Alonso, E. Eardley and J. A. Peacock, *Halo abundances within the cosmic web*, MNRAS **447** (2015) 2683 [1406.4159].

[61] S. Alam, Y. Zu, J. A. Peacock and R. Mandelbaum, *Cosmic web dependence of galaxy clustering and quenching in SDSS*, MNRAS **483** (2019) 4501 [1801.04878].

[62] T. E. Oliphant, *A guide to NumPy*, vol. 1. Trelgol Publishing USA, 2006.

[63] S. Van Der Walt, S. C. Colbert and G. Varoquaux, *The numpy array: a structure for efficient numerical computation*, *Computing in Science & Engineering* **13** (2011) 22.

[64] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17** (2020) 261.

[65] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray et al., *Astropy: A community Python package for astronomy*, A&A **558** (2013) A33 [1307.6212].

[66] Astropy Collaboration, A. M. Price-Whelan, B. M. SipHocz, H. M. G"unther, P. L. Lim, S. M. Crawford et al., *The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*, aj **156** (2018) 123 [1801.02634].

[67] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python*, *Journal of Open Source Software* **4** (2019) 1298.

[68] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak et al., *nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure*, AJ **156** (2018) 160 [1712.05834].

[69] A. Lewis, *GetDist: a Python package for analysing Monte Carlo samples*, *arXiv e-prints* (2019) arXiv:1910.13970 [1910.13970].

[70] J. D. Hunter, *Matplotlib: A 2d graphics environment*, *Computing in Science & Engineering* **9** (2007) 90.