The cosmology dependence of galaxy clustering and lensing from a hybrid N-body–perturbation theory model

Nickolas Kokron [1,2]★ Joseph DeRose,[3,4] Shi-Fan Chen [3] Martin White [3,5] and Risa H. Wechsler[1,2]

[1]*Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[2]*Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*
[3]*Department of Physics, University of California, Berkeley, 366 LeConte Hall, Berkeley, CA 94720, USA*
[4]*Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, CA 95064, USA*
[5]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 93720, USA*

ABSTRACT

We implement a model for the two-point statistics of biased tracers that combines dark matter dynamics from N-body simulations with an analytic Lagrangian bias expansion. Using Aemulus, a suite of N-body simulations built for emulation of cosmological observables, we emulate the cosmology dependence of these non-linear spectra from redshifts $z = 0$ to $z = 2$. We quantify the accuracy of our emulation procedure, which is sub-per cent at $k = 1\,h\,\mathrm{Mpc}^{-1}$ for the redshifts probed by upcoming surveys and improves at higher redshifts. We demonstrate its ability to describe the statistics of complex tracer samples, including those with assembly bias and baryonic effects, reliably fitting the clustering and lensing statistics of such samples at redshift $z \simeq$ 0.4 to scales of $k_{\mathrm{max}} \approx 0.6\,h\,\mathrm{Mpc}^{-1}$. We show that the emulator can be used for unbiased cosmological parameter inference in simulated joint clustering and galaxy–galaxy lensing analyses with data drawn from an independent N-body simulation. These results indicate that our emulator is a promising tool that can be readily applied to the analysis of current and upcoming data sets from galaxy surveys.

Key words: methods: statistical – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

We are entering a golden era for studying the large-scale structure of the Universe. Over the next decade, ambitious imaging surveys will map out large swathes of the sky to unprecedented depths, imaging billions of galaxies and their shapes (Laureijs et al. 2011; Doré et al. 2015, 2019; Ivezić et al. 2019), enabling studies of weak gravitational lensing by the intervening distribution of matter (Bartelmann & Schneider 2001; Mandelbaum 2018). Weak lensing has only recently begun to contribute competitive cosmological constraints on dark matter and dark energy (Abbott et al. 2018; Heymans et al. 2020), but is one of the most promising future directions to pursue. Meanwhile, spectroscopic surveys will observe tens of millions of radial positions of galaxies (Takada et al. 2014; Aghamousa et al. 2016), enabling unparalleled understanding of the spatial distribution of galaxies in our Universe. The cross-correlation between positions and lensing, galaxy–galaxy lensing, is and will continue to be a key driver of cosmological constraints from galaxy surveys.

The quality and quantity of these upcoming data sets imposes a significant challenge in their analysis. Even now, models for summary statistics such as correlation functions and power spectra are inadequate across the full range of scales probed by such surveys (Krause et al. 2017; Nishimichi et al. 2020). Either a large amount of the data must be discarded, or mitigation schemes must be developed to prevent contamination from scales where

the models are insufficiently calibrated or constrained (MacCrann et al. 2020; Park, Rozo & Krause 2020). Models for clustering and lensing must be substantially improved if we are to extract the maximal information about the Universe we live in, from surveys that are already ongoing or planned. To date, two separate approaches have been developed to build models for the observables of cosmic surveys: analytically, through perturbative techniques, or numerically, using non-linear N-body simulations.

Perturbation theory provides a systematic, analytic way to compute N-point summary statistics to systematically higher precision and smaller scales (Bernardeau et al. 2002). Below the non-linear scale the effects of these non-linearities can be tamed and parametrized within the framework of effective theories (Baumann et al. 2012; Carrasco, Hertzberg & Senatore 2012; Vlah, White & Aviles 2015). This increased precision, however, comes at the cost of very large inaccuracies beyond the non-linear scale at which the self-gravitating dark matter fluid ceases to be perturbative (Blas, Garny & Konstandin 2014; McQuinn & White 2016). In addition, perturbative frameworks provide a rigorous, first-principles approach to include physics beyond the standard lambda cold dark matter (ΛCDM) model in large-scale structure observables such as neutrinos, baryonic effects and more exotic early-universe scenarios (Lewandowski, Perko & Senatore 2015; Senatore & Zaldarriaga 2017; Aviles & Banerjee 2020; Aviles et al. 2020; Chen, Vlah & White 2020c; D'Amico et al. 2020; Ivanov et al. 2020; Laguë et al. 2020). Understanding the domain of applicability of perturbation theory is still an active field of research (Baldauf, Schaan & Zaldarriaga 2016b; Chen et al. 2020a; Nishimichi et al. 2020).

★ E-mail: kokron@stanford.edu

The other approach, simulation-based modelling, involves numerically solving the equations of motion for an initial distribution of matter (Hockney & Eastwood 1988; Bagla 2005; Kuhlen, Vogelsberger & Angulo 2012). The resulting catalogues can be analysed in a way analogous to data to obtain predictions of cosmological observables across a wide range of scales at the cosmological parameters of the simulation.

However, a limiting factor in simulation-based analyses is that *N*-body simulations require significant computational resources for a single realization. Thus, standard inference procedures such as Markov chain Monte Carlo (MCMC) become prohibitively expensive when using models derived from simulations.

In order to ameliorate the issues with simulation-based inference, recent developments in statistical learning have popularized so-called emulators as models (Heitmann et al. 2010, 2009; Lawrence et al. 2010). Emulators combine a set of simulations that representatively sample cosmological parameter space with sophisticated regression techniques to 'fill in the blanks' across parameter space. Once trained, an emulator provides rapid evaluations of a model which can be seamlessly integrated in analysis pipelines. For example, recent emulators for the non-linear matter power spectrum (Knabenhans et al. 2019) have runtimes with negligible overhead compared to the underlying Boltzmann codes used for linear predictions.

While galaxy surveys observe luminous tracers of the underlying dark matter density distribution, most suites of *N*-body simulations used to construct emulators deal only with the dark matter component. Thus, emulators for galaxy survey observables are presented with the additional challenge of capturing the relationship between the galaxy distribution and the underlying dark matter. Understanding the details of this relationship, known as the galaxy–halo connection, is an active field of research (see e.g. Wechsler & Tinker 2018, for a recent review). Even for well-studied samples of galaxies, there are no consensus models to describe this relationship. For any given model of the galaxy–halo connection, an entirely new emulator has to be trained (Kwan et al. 2015; Wibking et al. 2019; Zhai et al. 2019; McLaughlin et al. 2021). Emulation of models with a large number of free parameters is also a challenging task, with techniques such as Gaussian processes (GPs) scaling as $\mathcal{O}(N^3)$ with *N* training points and a substantially larger set of training data being required as one increases the dimensionality of the model. The simplest forms of galaxy–halo connections such as halo occupation distributions (HODs) have five free parameters (Zheng et al. 2005), and it is expected that for more complex selections of galaxy samples the number will grow considerably (Yuan, Eisenstein & Garrison 2018; Guo et al. 2019; Favole et al. 2020; Zu 2020).

In comparison, modern perturbation theory approaches to galaxy clustering operate at the field level via so-called bias expansions, which encode the response of small-scale galaxy physics (e.g. the galaxy–halo connection) to large-scale structure via a series of bias coefficients (see e.g. Desjacques, Jeong & Schmidt 2018 for a recent review). A key advantage of bias models is that while their dependence on parameters is simple and analytic, they should describe the statistics of a broad range of galaxy (and halo) samples as long as they are formed by processes that respect the symmetries of the underlying processes of structure and galaxy formation, namely rotational and Galilean invariance and the equivalence principle. Indeed, it was recently shown that the bias expansion can be directly derived by generating all possible dynamical terms and eliminating combinations not allowed by these symmetries (Fujita & Vlah 2020).

The challenges in using bias models come, instead, from the aforementioned limitations of perturbation theory models themselves. Similarly to perturbation theories for the clustering of dark matter,

bias models are not expected to hold across all scales. Instead, they are expected to be valid at scales larger than or comparable to the Lagrangian size of haloes. This regime is where one is insensitive to the internal structure of haloes (McDonald & Roy 2009; Vlah, Castorina & White 2016; Lazeyras & Schmidt 2019; Fujita et al. 2020). It is worth noting, however, that the non-linear and halo scales are not identical and scale differently with redshift – at higher redshifts perturbative models may be more limited by the larger Lagrangian radii of (typically more luminous or massive) samples than dynamical non-linearities, and vice versa at lower redshifts. This distinction is particularly apparent in the Lagrangian basis (Matsubara 2008; Vlah et al. 2016), in which galaxy clustering due to dynamics and biasing are explicitly disentangled. Recently Modi, Chen & White (2020) suggested a way to combine the generality of bias expansion-based models with *N*-body simulations in a manner that is particularly suited for emulation, particularly in the regime where dynamics become non-linear on scales larger than the halo scales of interest. Since higher-order Lagrangian biases have been found in simulations to be small for low- and intermediate-mass haloes (Abidi & Baldauf 2018; Lazeyras & Schmidt 2018), this scheme keeps the dynamical non-linearities from *N*-body simulations to all orders while including Lagrangian bias only up to second order.

In the remainder of this work, we concern ourselves with the construction of an emulator for the halo–halo and halo–matter correlations with analytic dependence on bias parameters, extending the method presented in Modi et al. (2020) to a generic cosmological parameter dependence which can then be readily used for cosmological clustering analyses. The structure is as follows: in Section 2 we briefly review the Lagrangian description of galaxy bias. In Section 3 we describe the hybrid technique which combines displacements obtained from *N*-body simulations with Lagrangian bias. Section 4 describes the Aemulus suite of simulations (DeRose et al. 2019b), which we use to build the training data for the emulator. The measurements of the 'basis spectra' of the hybrid Lagrangian bias model, and their emulation, are outlined in Section 5. Section 6 concerns itself with assessing the performance of the emulator. Specifically, Section 6.1 addresses the scale and redshift-dependent error for each of the ten basis functions that span the model. Section 6.2 assesses how well the model describes the statistics of complicated galaxy samples, including those possessing concentration and spin secondary biases, as well as the effect of baryons at small scales. Our final test, Section 6.3, pits the emulator against a series of increasingly complex simulated likelihood analyses, in order to assess potential biases in inferred cosmological parameters using our emulator and their origin.

2 LAGRANGIAN BIAS EXPANSION

In the Lagrangian approach to bias formulated in Matsubara (2008), the observed clustering of galaxies is obtained through first weighting fluid elements by a local functional $F[\delta(\boldsymbol{q})]$ at their initial (Lagrangian) positions $\boldsymbol{q}$ and then advecting these weights to their observed positions via fluid trajectories $\boldsymbol{x} = \boldsymbol{q} + \Psi$, where $\Psi(\boldsymbol{q}, t)$ is the Lagrangian displacement. As discussed in the introduction, the bias functional F is obtained by summing up all scalar terms allowed by Galilean invariance and the equivalence principle up to a given order in the initial conditions; up to quadratic order we have (Vlah et al. 2016)

$$F(\boldsymbol{q}) \approx 1 + b_1 \delta_L(\boldsymbol{q}) + \frac{b_2}{2!}\left(\delta_L^2(\boldsymbol{q}) - \left\langle \delta_L^2 \right\rangle\right)$$
$$+ b_{s^2}\left(s_L^2(\boldsymbol{q}) - \left\langle s_L^2 \right\rangle\right) + b_{\nabla^2}\nabla^2 \delta_L(\boldsymbol{q}) + \epsilon(\boldsymbol{q}), \quad (1)$$

where $s^2 = s_{ij}s_{ij}$ is the tidal shear tensor. The bias expansion is local above the halo scale and the initial fields in the above functional are to be interpreted as smoothed; any 'non-local' effects as we approach this scale, as well as dependences on smoothing, are parametrized to lowest order by the derivative bias b_{∇^2}. Modes below the halo scale, uncorrelated with the large scales of interest, are represented by the stochastic noise ϵ.

From the weighting $F(\boldsymbol{q})$, the observed clustering is given via number conservation to be

$$1 + \delta_\alpha(\boldsymbol{x}, z) = \int \mathrm{d}^3 q \, \delta^D(\boldsymbol{x} - \boldsymbol{q} - \Psi(\boldsymbol{q}, z))F(\boldsymbol{q}), \tag{2}$$

where the Lagrangian displacement Ψ denotes the movement of the fluid element relative to its initial position. At any given order, the Lagrangian galaxy overdensity above can be mapped on to e.g. the Eulerian basis of McDonald & Roy (2009) by Taylor expanding Ψ. However, keeping the non-linear mapping in the integral above will generate a tower of Eulerian bias parameters even if only a few of the Lagrangian bias parameters are non-zero (see e.g. Abidi & Baldauf 2018). We will treat the bias values, b_α, as free parameters. Ab initio predictions of the b_α for general tracer populations is a harder problem, and a current active area of research.

3 LAGRANGIAN BIAS AND SIMULATIONS

Recently, it has been proposed that one can combine the fully resolved dark matter dynamics of an *N*-body simulation with the analytic perturbative bias techniques we outlined in the previous section (Modi et al. 2020). The use of dynamics from an *N*-body simulation means this hybrid model circumvents the need for perturbative calculations related to the equations of motion of the dark matter fluid itself. Additionally, *N*-body simulations are relatively inexpensive (compared to hydrodynamical simulations) and well-controlled, well-defined limits for observables exist so that convergence of measured quantities can be assessed systematically (e.g. Power et al. 2016; Joyce, Garrison & Eisenstein 2020; Mansfield & Avestruz 2020). As such, this hybrid model combines two techniques with solid theoretical foundations, ensuring robustness of its predictions. We will briefly describe the technique and how one implements it below, but refer the reader to Modi et al. (2020) for a more complete discussion.

When creating initial conditions of an *N*-body simulation, one starts from a noiseless linear cosmological density field, $\delta_L(\boldsymbol{x})$. Traditionally, this density is only used to sample initial displacements that impart a cosmological signal on a set of *pre*-initial conditions. First-order displacements using the Zeldovich approximation,

$$\Psi(\boldsymbol{q}) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} e^{i\boldsymbol{k}\cdot\boldsymbol{q}} \frac{i\boldsymbol{k}}{k^2} \delta_L(\boldsymbol{k}), \tag{3}$$

result in so-called 1LPT initial conditions. However, higher order initial conditions (Crocce, Pueblas & Scoccimarro 2006; Garrison et al. 2016; Michaux et al. 2020) are now ubiquitous in modern simulations.

The noiseless initial density field can also be used to construct the different component fields of the Lagrangian bias expansion of the initial conditions:

$$O_L \supset 1, \delta_L, \delta_L^2, s_L^2, \nabla^2\delta_L, \cdots, \tag{4}$$

where the subscript L indicates these are the Lagrangian fields. Advecting *N*-body particles weighted by $\mathcal{O}_L$ to a specific snapshot

results in bias-weighted fields,

$$\delta_{\mathcal{O}_L}(\boldsymbol{x}) \equiv \int \mathrm{d}^3 q \, \mathcal{O}_L(\boldsymbol{q}) \, \delta_D(\boldsymbol{x} - \boldsymbol{q} - \Psi(\boldsymbol{q})), \tag{5}$$

which trace the non-linear dark matter distribution. In Fig. 1 (middle panel), we show an example of the different bias-weighted fields produced by this procedure. These fields are similar to the 'Eulerian-shifted' operator basis of Schmittfull et al. (2019). A notable difference is that in our case the displacements are fully resummed, while the Eulerian-shifted basis of Schmittfull et al. (2019) only resums the Zeldovich displacement (1LPT). Higher order displacements (*n*LPT) are Taylor-expanded up to third order as part of their bias expansion. The difference is because our aim in this paper is to attempt to model scales beyond the reach of standard one-loop perturbation theory, whereas the goal of Schmittfull et al. (2019) was to validate one-loop perturbation theory at the field level (see also Taruya, Nishimichi & Jeong 2018).

The power spectrum of any combination of tracers can then generically be written as ($X, Y \equiv \delta_{\mathcal{O}_L}$)

$$P^{ab}(k) = \sum_{X,Y} b_X^a b_Y^b P_{XY}(k) + P_{\mathrm{SN}}, \tag{6}$$

where P_{XY} is the cross-power spectrum at a fixed cosmology between the different fields at a given redshift. For example, the unweighted spectrum, P_{11}, is the non-linear matter power spectrum.

This Lagrangian bias model can handle cross-correlations of arbitrary tracers. However, we also note that given a set of bias parameters for a single tracer sample α, $\{b_X^\alpha, X \in \mathcal{O}_L\}$, one can also self-consistently predict the tracer–matter cross-correlation by taking the second sample to have $b_Y^m = 0$ except for $Y = 1$. In this case, there are only P_{X1} terms. The tracer–matter cross-correlation is the primary cosmic contribution to the signal of galaxy–galaxy lensing, one of the key cosmological observables of current and upcoming galaxy surveys (Yoo et al. 2006; Mandelbaum 2018; Prat et al. 2018; Wibking et al. 2020). The tracer–matter cross-correlation is also the primary contribution to the cross-correlation between galaxy positions and lensing of the cosmic microwave background (CMB), one of the most powerful and complementary statistics that is measured between galaxy and CMB surveys (Bianchini et al. 2015; Pullen et al. 2016; DiPompeo et al. 2017; Peacock & Bilicki 2018; Omori et al. 2019; Singh et al. 2019; Krolewski et al. 2020). For notational convenience, throughout the remainder of this paper we will refer to the tracer–tracer correlation as $P^{hh}(k)$ and the tracer–matter correlation as $P^{hm}(k)$.

This hybrid approach of combining *N*-body simulations with Lagrangian bias can fit the power spectrum of tracers to significantly smaller scales than standard Lagrangian perturbation theory (Modi et al. 2020). While the dependence on the Lagrangian bias parameters b_X is analytic in this model, one still requires an *N*-body simulation to measure the basis spectra. An *N*-body simulation at a given point of cosmological parameter space then provides a measurement of the basis spectra at that point. With *N*-body simulations that sufficiently sample parameter space, one can estimate the cosmological dependence of these basis functions across the entire space. This is precisely the goal of this work.

4 THE Aemulus SIMULATIONS

In order to properly emulate the cosmology dependence of the basis spectra $P_{XY}(k)$, the underlying suite of *N*-body simulations used for measurements of observables must be constructed carefully.
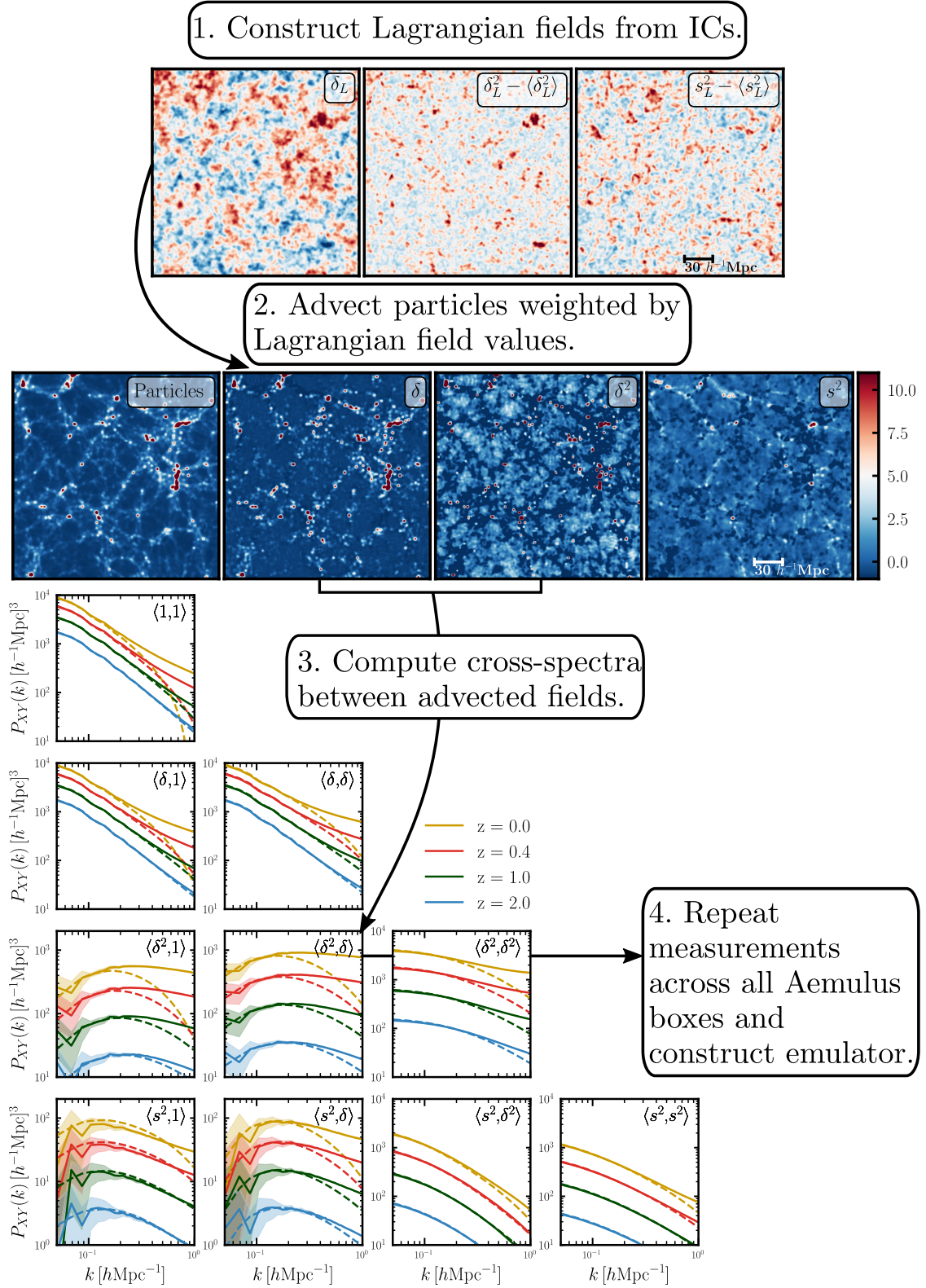
Figure 1. Visualization of the methodology implemented in this paper, from the advection process to the measurements of the basis spectra. Our emulation scheme approximates the cosmology and redshift dependence of each spectrum in the 10 panels in the lower part of the figure. The top panel has each Lagrangian field scaled to have equal variance, in order to highlight the qualitative differences between the fields. The middle panel shows the bias weighted-fields that result from the advection process. Different weights highlight qualitatively different aspect of the matter density. The cross-spectra of these fields give the spectra shown in the lower panel.

The `Aemulus` suite of *N*-body simulations (DeRose et al. 2019b) has been purpose-built for precise emulation of cosmological observables measured in galaxy surveys. The suite is composed of a set of 75 simulations that span 47 points in the *w*CDM parameter space allowed by a combination of modern CMB, BAO, and type Ia supernova experiments.

Each `Aemulus` box has a size $L_{box} = 1050\,h^{-1}$ Mpc with $N = 1400^3$ particles, corresponding to a mass resolution of $3.51 \times 10^{10} \left(\frac{\Omega_m}{0.3}\right) h^{-1}$ M$_\odot$. The `Aemulus` simulations have undergone rigorous convergence and validation tests for several observables. There are 10 particle snapshots ranging from $0 < z < 3$, allowing for measurements the redshift-dependence of the non-linear basis spectra.

`Aemulus`' halo mass function emulator has sufficient accuracy to remain valid, for the defined cosmological parameter space, through the Rubin Observatory's Y1 LSST survey (McClintock et al. 2019), while the galaxy correlation function can predict the clustering of massive galaxy samples, such as those observed by DESI, to within 1 per cent down to scales of $r \approx 1\,h^{-1}$ Mpc (Zhai et al. 2019).

Thus, `Aemulus` represents an appropriate setting to construct an emulator for the Lagrangian bias basis spectra described in Section 3. The only missing component is that the initial conditions code used in `Aemulus`, `2LPTIC` (Crocce, Pueblas & Scoccimarro 2012), does not output the noiseless linear density fields. We patched the code to read out this field and re-generated the initial conditions.

5 EMULATING THE BASIS SPECTRA

5.1 Measuring basis spectra

We now describe in detail our implementation of the hybrid Lagrangian biasing scheme described in Section 3. Schematically, the process of obtaining measurements of the basis spectra from an *N*-body box can be broken down into four steps:

(i) Compute the Lagrangian bias fields: given the noiseless density field δ_L one constructs the other weight fields $\mathcal{O}_L$ by applying the appropriate transformations.

(ii) Advect particles to a given snapshot: every particle ID can be associated with a grid cell $\{i, j, k\}$ in the fields $\mathcal{O}_L$. Every particle in a snapshot receives a weight $\left(\frac{D(z)}{D(z_0)}\right)^n \times \mathcal{O}_L[i, j, k]$, where $\left(\frac{D(z)}{D(z_0)}\right)$ is the ratio of growth factors between the snapshot and initial conditions, and n is the number of powers in the linear density field that make up $\mathcal{O}_L$.

(iii) Paint the weighted particles to a grid, to form the late-time bias fields.

(iv) Measure the basis spectra: the painted bias fields are cross-correlated with each other to measure the basis spectra P_{XY} for that given cosmology and redshift.

This procedure imposes some additional storage requirements. While a particle catalogue normally has seven entries for every particle, $(ID, \boldsymbol{x}, \boldsymbol{v})$, each bias field weight will add an additional entry. Naively saving component weights at every snapshot will lead to a 57 per cent increase in catalogue size. However, the time evolution of the weights is determined entirely by the linear growth function and can be determined on the fly. Thus, the fractional increase in catalogue size will only be of order $\sim (1/7)(N_b/N_z)$, where N_b is the number of bias-weighted fields computed and N_z is the number of snapshots used. For the second order basis of $\mathcal{O} = \{1, \delta_L, \delta_L^2, s_L^2, \nabla^2\delta_L\}$ this represents a fractional increase in catalogue size of 6 per cent. Even if the weights are not stored,

all of the steps outlined above can be carried out on the fly when needed.

In Figs. 1 and 2, we show a comparison between the predictions of one-loop Lagrangian perturbation theory and the basis spectra averaged across five `Aemulus` boxes with the same cosmology, from the test suite. For all basis spectra, we recover the LPT result at large scales to within a few per cent. While one would expect the agreement at large scales to be exact, it is well known that *N*-body simulations struggle to correctly recover linear growth at large scales (Heitmann et al. 2010; Garrison et al. 2016; Schneider et al. 2016) due to transients from the grid that particles are initialized on, and the discrete nature of the kick-drift-kick operators used in time-stepping. This discrepancy is also present in `Aemulus`, as can be seen in fig. 13 of DeRose et al. (2019b). The `Aemulus` simulations have a 1 per cent mismatch in growth at large scales, which is redshift independent at the largest scales. Differences in growth between linear theory and the simulations would then be amplified for the basis spectra built from multiple fields. In Appendix C, we explore the $k \to 0$ differences between LPT and our emulator, present prescriptions for enforcing consistency and discuss the small impact they have on parameter inference.

At small scales, we see that non-linear structure formation imbues significant differences between LPT and the simulations. At the highest redshift shown, $z = 2$, the agreement for the three spectra that dominate the signal ($\langle 1, 1 \rangle$, $\langle 1, \delta \rangle$, and $\langle \delta, \delta \rangle$) is close throughout all scales probed in our simulation. Thus, for the scales under consideration, we find no need to extend the emulator to $z > 2$.

Above $z = 1$, the `Aemulus` simulations only have snapshots at $z = 2$ and $z = 3$, and thus any attempt to emulate redshift evolution between these snapshots is too poorly sampled for the emulator to achieve our desired performance. For $z \geq 2$, the emulator reverts to predictions from `velocileptors` (Chen, Vlah & White 2020b), a public code to predict LPT power spectra and correlation functions to one loop order. This agrees quite well with most basis spectra given Fig. 1. When reverting to LPT at $z > 2$, our implementation includes an additional free parameter. This parameter corresponds to the k^2 counterterm for matter that takes into account the effects of small-scale physics not captured by perturbation theory (Vlah et al. 2015). We note that there are no specific impediments to measuring basis spectra, or the emulation scheme adopted, at higher redshifts. Given simulations that are sufficiently well sampled in time, out to the furthest bin one wishes to include, the techniques described here should apply.

The LPT predictions shown in Fig. 1 are a limit of a more complete theory that includes redshift-space distortions (Chen et al. 2020a, b). The agreement between *N*-body simulations and this subset of LPT at large scales implies the bias parameters in the full theory and our hybrid model are equivalent; a set of bias parameters obtained from fitting the emulator to a sample can then be used in tandem with RSD (redshift-space distortion) measurements analysed purely with perturbation theory at a slightly more restrictive k_{max}. Since the RSD measurements are done in 3D, rather than projection, one can achieve small measurement errors at more restrictive k_{max} making this combination an efficient one, e.g. for testing general relativity (Alam et al. 2017; Zhang et al. 2020).

We note that we omit results for the basis spectra $\langle X, \nabla^2\delta \rangle$. The initial weight field $\nabla^2\delta_L$ has a large amount of power at very small scales, making its Fourier transform unwieldy due to the presence of an explicit smoothing scale of $k \sim L_{grid}^{-1}$. As a result, we find the basis spectra as measured through the advection procedure have a cosmology-dependent amplitude mismatch when compared to LPT predictions at large scales. Therefore we adopt the approximation $\langle X,$

$\nabla^2 \delta \rangle \approx -k^2 \langle X, 1 \rangle$ in the actual emulation scheme. Since these higher derivative bias contributions most closely correspond to the effects of baryonic physics and finite-size effects for haloes, we check that the approximation performs similarly in Section 6.2. Specifically, in Fig. 8 we explicitly show the differences between the measured $P_{1\nabla^2}$ and the approximation employed. We also note the approximation lowers the complexity of the emulation scheme, reducing the full set of basis functions at second order to be emulated from 15 to 10.

5.2 Principal components of non-linear spectra

Once the basis spectra have been measured across all boxes, the emulator is built by adopting a suitable interpolation scheme between the different spectra. While other emulators using the Aemulus simulations have been constructed using GPs, we adopt a different approach here, similar to that used in the Euclid emulator (Knabenhans et al. 2019), using a combination of principal component analysis and polynomial chaos expansions (PCEs) (Xiu 2010).

We prefer PCE to GP emulation for a few practical reasons. GPs are more difficult to train, requiring explicit choices for kernels and tuning of real valued hyper-parameters. Additionally, the run time for evaluating a trained GP scales with the amount of data used for training, while the runtime of a PCE model evaluation scales only with the order of the PCE. Furthermore, the polynomial nature of PCEs means that they have fast, analytic gradients, making them easy to integrate with sampling techniques such as Hamiltonian Monte Carlo (Hoffman & Gelman 2011), although we have not done so in this work. GPs may still be preferred when the model being emulated is highly complex, but, as we show in the following sections, we are able to attain a nearly optimal emulator performance with the simpler and faster PCE scheme.

To begin, we compute one-loop LPT predictions for each basis spectrum at every cosmology and redshift in the Aemulus training design, which we will refer to as $P_{XY}^{\text{LPT}}(k, \Omega)$, where Ω denotes the cosmology and redshift in question. To do this, we make use of the VELOCILEPTORS code (Chen et al. 2020b).

We then compute the ratio between the LPT predictions and the measured basis spectra, $P_{XY}^{\text{NL}}(k, \Omega)$, from each snapshot. These ratios are thus consistent with unity at small wavenumbers, and while they deviate significantly from unity at high k they have significantly less dynamic range than the basis spectra. In order to de-noise these ratios, we apply a Savitsky–Golay (Savitsky & Golay 1964) filter of order three using an 11-point window in k. Doing so dramatically reduces the amount of noise in the spectra, and is a simple alternative to reduce noise at high k, where techniques such as fixed amplitude, paired phase simulations do little to reduce variance (Angulo & Pontzen 2016; Villaescusa-Navarro et al. 2018; Chuang et al. 2019). As a final pre-processing step, we also take the base-10 logarithm of these smoothed ratios in order to further decrease the dynamic range. This yields the quantity that we emulate, which we call $\Gamma^{XY}(k, \Omega)$,

$$\Gamma^{XY}(k, \Omega) \equiv \log_{10}\left(\frac{P_{XY}^{\text{NL}}(k, \Omega)}{P_{XY}^{\text{LPT}}(k, \Omega)}\right). \tag{7}$$

After these pre-processing steps, we proceed by constructing a principal component basis for these spectra. At this point we restrict ourselves to $0.1 < k < 1$ and $0 < z < 2$, however we note that in principle there are no issues extending to broader scales and redshifts if simulations allow for it.

Let X_{XY} be the $N \times M$ array containing Γ^{XY}, where $N = N_{\text{cosmo}} \times N_z$, N_{cosmo} is the number of cosmologies in our training set, N_z is the number of redshift outputs per cosmology in our training set, and M is

the number of k values under consideration. Then a basis of principal components can be constructed by computing the eigenvectors of the covariance matrix of X_{XY}:

$$C_{XY} = X_{XY}^{\text{T}} X_{XY},$$
$$= W_{XY} \Lambda_{XY} W_{XY}^{\text{T}}, \tag{8}$$

where the rows of W_{XY} are the eigenvectors, i.e. the principal components, in question and Λ_{XY} is a diagonal matrix of the eigenvalues, which are equal to the variance of the data described by each eigenvector. In all cases, greater than 99 per cent of the variance in each basis spectrum is described by the first two principal components, shown in Fig. 3. We thus disregard all other principal components for the duration of this work. Given the results discussed in Section 6, we deem this to be sufficient. Having computed the principal components, we then determine the projection of them on to each measured Γ^{XY} via:

$$A_{XY} = X_{XY} W_{XY}, \tag{9}$$

where A_{XY} is an $N \times 2$ matrix containing the principle component coefficients $\alpha_i^{XY}(\Omega)$ for each cosmology and redshift in our training set. It is the dependence of these coefficients on cosmology and redshift that we build a surrogate model for using PCEs (Wiener 1938). In Fig. 4 we show an example of this dependence for the coefficient associated to the first principal component and three of the dominant basis spectra.

5.3 Emulating cosmology dependence with polynomial chaos

With our principal components in hand, every point in cosmological parameter space sampled by the training set has coefficients for the approximation

$$\Gamma^{XY}(k, \Omega) \approx \sum_i \alpha_i^{XY}(\Omega) \text{PC}_i^{XY}(k). \tag{10}$$

The problem of emulating the cosmology dependence of the Γ^{XY} functions is now reduced to that of figuring out the cosmology dependence of the PC coefficients $\alpha_i(\Omega)$. A PCE (of order N) of this dependence is the decomposition of the α_i on to a basis of products of orthogonal polynomials $\Phi_{\mathbf{i}}(\Omega)$ organized by a multi-index $\mathbf{i}$ (Xiu 2010):

$$\alpha(\Omega) = \sum_{|\mathbf{i}| \leq N} c_{\mathbf{i}} \Phi_{\mathbf{i}}(\Omega). \tag{11}$$

Each component of the multi-index $\mathbf{i} = (i_1, \cdots, i_d)$, denotes the order of the polynomial for that cosmological parameter, e.g.

$$\Phi_{\mathbf{i}}(\Omega) = \phi_{i_1}(\Omega_1) \cdots \phi_{i_d}(\Omega_d), \tag{12}$$

and so $\phi_{i_d}(\Omega_d)$ is a univariate orthogonal polynomial of order i_d.

While this is in principle a decomposition into a combinatorially large space of coefficients $c_{\mathbf{i}}$, it is known to be a sparse representation (Blatman & Sudret 2008, 2011), and there exist many algorithms (and numerical libraries) optimized to perform regression over this space and obtain values for the coefficients. We use the package Chaospy (Feinberg & Langtangen 2015; Feinberg, Eck & Langtangen 2018) to perform the decomposition and subsequent regression. Note that since the parameter dependence of the principal components is given by a combination of polynomials, our model in principle has an analytic dependence on cosmology, redshift, and bias. Since the coefficients are determined via regression, a PCE emulator does not recover the input data exactly. However, the tests conducted in Section 6 indicate that this drawback is not an issue.
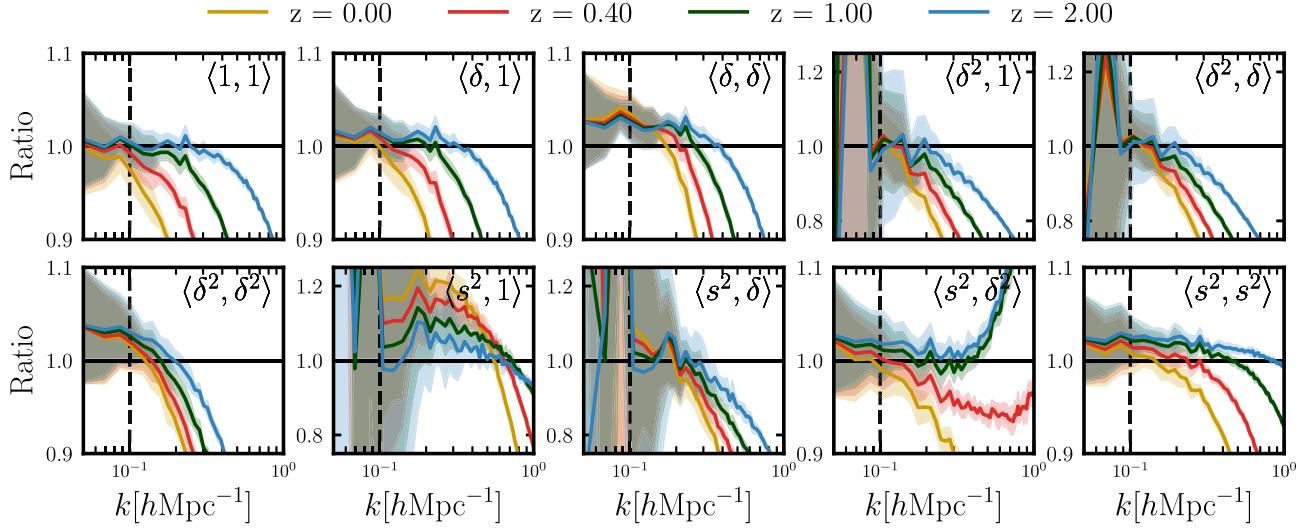
Figure 2. Ratio of the measured basis spectra compared to LPT predictions for one of the cosmologies in the Aemulus test set. The mean of the five independent boxes in the test set is shown, and the shaded band represents one standard deviation as inferred from the boxes. The dashed vertical line at $k = 0.1 h\,\mathrm{Mpc}^{-1}$ shows the point where we revert to predictions of LPT. As discussed in the text, we find some small multiplicative differences at large scales for most basis spectra, that are larger for the basis spectra built from higher powers of the density field. This is most likely due to discrepancies in growth factors obtained between linear theory and N-body simulations.
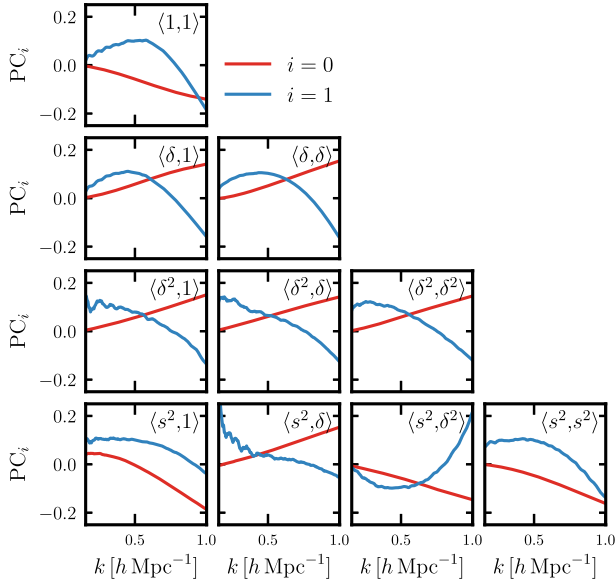
Figure 3. The first two principal components of the log ratios between N-body and LPT spectra, Γ^{XY}, for each basis spectrum. The principal components are very smooth compared to the raw basis spectrum measurements from the simulations. Two principal components are sufficient to explain greater than 99 per cent of the variance in all spectra as a function of redshift and cosmology.

In total, the hyperparameters in the model are as follows:

(i) The number of principal components used, N_{PC}.

(ii) The maximum order of the multi-index $|\boldsymbol{i}|$. In practice we separately optimize over the maximum polynomial order of each individual parameter i_d, with $i_d \leq 4$.

As mentioned previously, we restrict ourselves to $N_{\mathrm{PC}} = 2$, as this is sufficient to capture over 99 per cent of the variance in each basis spectrum. To optimize over the polynomial orders i_d, we run a simple
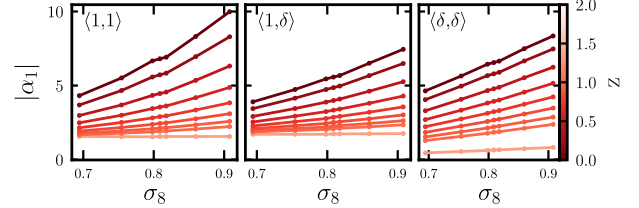

Figure 4. Coefficients of $\mathrm{PC}_1^{XY}(k)$ for the first three basis spectra as a function of σ_8, coloured by redshift. The coefficients vary smoothly for all redshifts as σ_8 is varied. It is the dependence of these coefficients that we emulate via PCE as a function of cosmology and redshift. The panels look similar for the remaining basis spectra.

grid search across the aforementioned values for the seven wCDM parameters $\Omega = (\Omega_b h^2, \Omega_c h^2, \sigma_8, H_0, n_s, N_{\mathrm{eff}}, w)$ and evaluate our results on the Aemulus test suite. We select the set of orders that minimizes global error across all test boxes and snapshots. We describe the tests of this optimized emulator below.

6 RESULTS

6.1 Analysis of emulator residuals

A crucial step in producing viable emulators of cosmological observables is characterizing the accuracy of the emulation scheme. We use the Aemulus set of test boxes to assess the performance of the scheme described in the previous section. The test boxes span seven points in cosmological parameter space, each with five independent realizations of that cosmology. We use the average of five basis spectra at each test cosmology as reference quantities to understand the errors induced in the emulation procedure as a function of scale, across parameter space.

We report the accuracy of our optimized PCE emulator for the basis spectra over the range $0.1 h\,\mathrm{Mpc}^{-1} \leq k \leq 1.0 h\,\mathrm{Mpc}^{-1}$ in the lower left panel of Fig. 5. Across most redshift bins in the test suite and for most basis spectra, we achieve better than 1 per cent accuracy
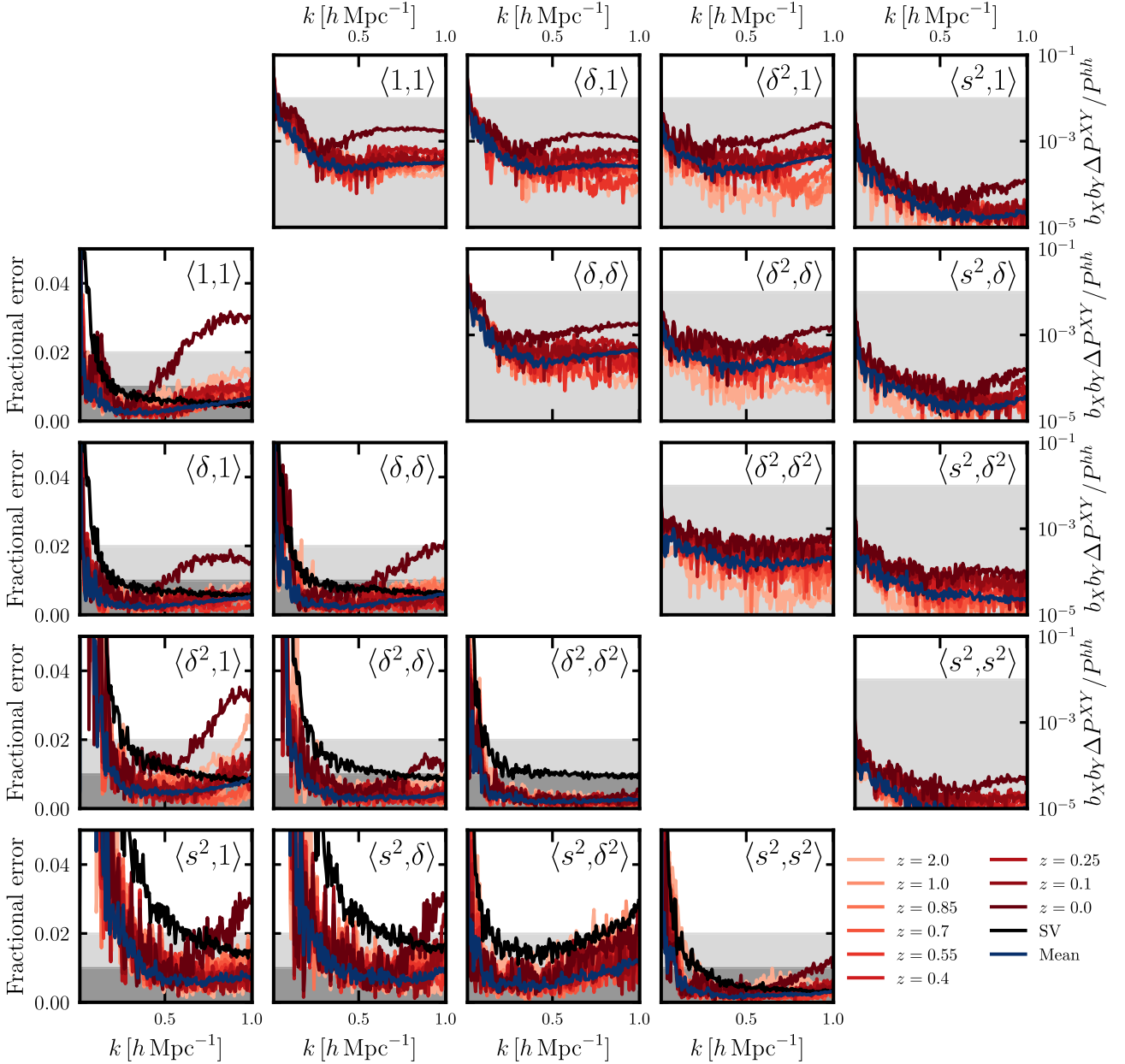
Figure 5. Emulation residuals for basis spectra. *Lower left triangle:* the fractional error obtained for each basis spectrum when compared to the measurements averaged from each set of boxes in the test suite. *Upper right triangle:* the relative size of the emulator residuals compared to the total halo–halo spectrum measured for a fiducial halo sample. In each panel, the dark blue curves are the mean residuals across all redshifts and test boxes, the red curves report the median residual error across the test suite as a function of redshift, and the black curves report the expected sample variance at the volume of an Aemulus training box.

in the test set. At $z = 0$, we observe worse performance, however this can be attributed to numerical difficulties in computing the LPT spectra at $z = 0$ at small scales, as can be seen in Fig. 1. As there is little cosmological information in the very low redshift universe, we do not consider this to be a significant issue. Indeed, our additional validation tests support that the model has sufficient accuracy to analyse current survey data.

Adopting a fiducial set of bias parameters corresponding to a halo sample of $12.5 \leq \log_{10}\left(\frac{M}{h^{-1}\,M_\odot}\right) \leq 13$, we compute the emulator residuals for each basis spectrum relative to the *total* $P^{hh}(k)$. The results are shown in the upper right triangle of Fig. 5. The individual basis spectrum error rarely exceeds a permille of the total power. This

implies that the slightly larger errors for cubic basis spectra shown in Fig. 5 are subleading relative to the total signal we expect to model.

6.2 Fitting assembly bias and baryons

Beyond samples of fixed halo mass, the general bias expansion in equation (1) should also be able to describe the clustering statistics of more complex tracer populations. It is well known that haloes of a fixed mass bin exhibit different clustering properties depending on whether they are subselected on certain properties. This effect, originally discovered in the context of assembly history, and generally known as assembly bias or secondary bias, has been observed for selections on concentration, occupation, local environment, spin, and
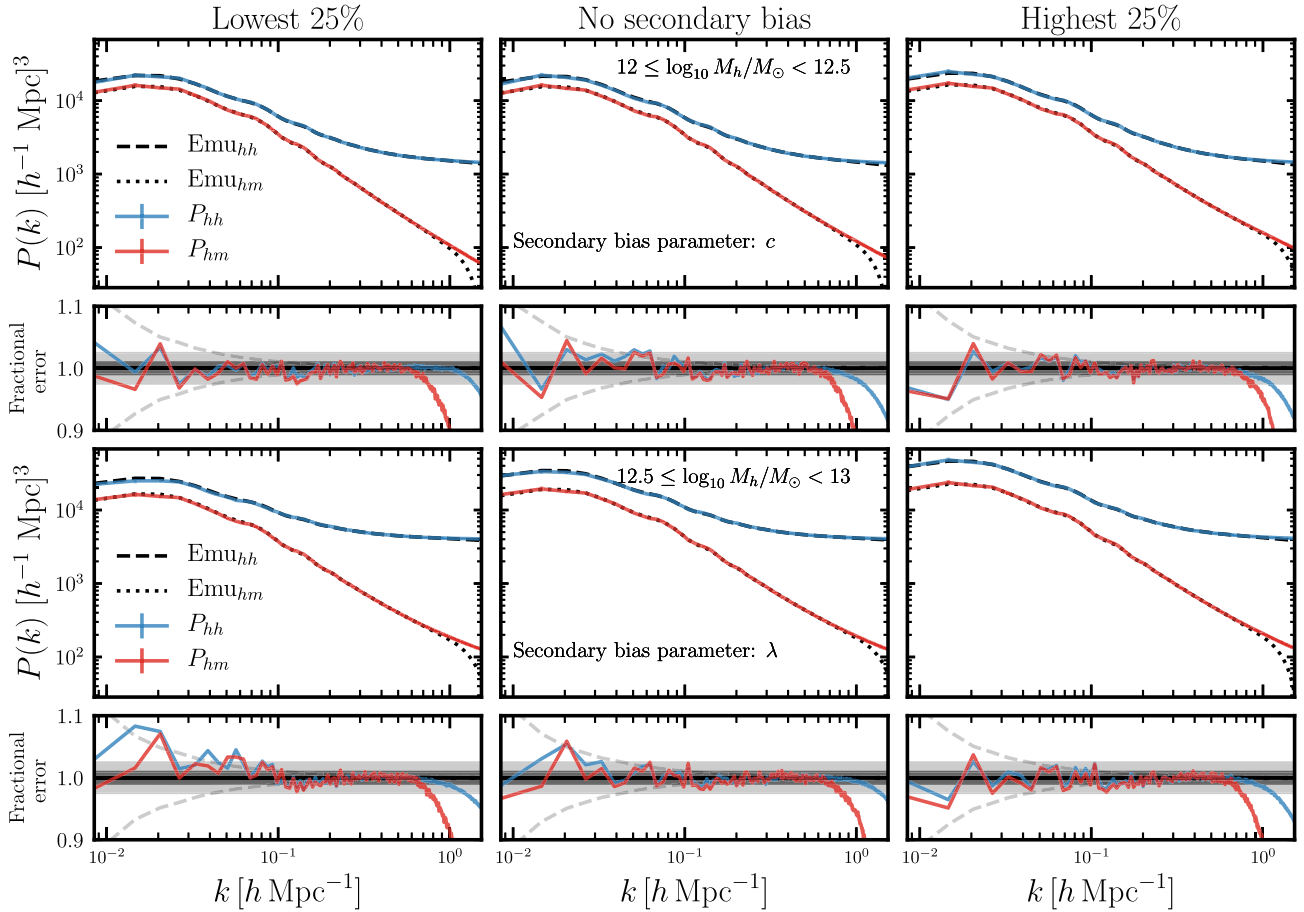
Figure 6. Emulator predictions at fixed cosmology for halo samples exhibiting concentration (top panels) and spin (bottom panels) assembly bias. Central panels show the signal from the halo sample with no selection on a secondary parameter. The left-hand and right-hand panels show samples split on the lowest and highest quartiles of the relevant secondary bias parameter, respectively. Shaded bands show the regions where residuals are within 2 per cent and 1 per cent, respectively, while the dashed envelope shows the expected cosmic variance for a sample with $V \approx 5.8 \, (h^{-1}\mathrm{Gpc})^3$. The spectra are measured at $z = 0.7$ and the fit is performed with the data vector out to $k_{\mathrm{max}} = 0.6 \, h \, \mathrm{Mpc}^{-1}$.

other secondary halo properties (Wechsler et al. 2002; Gao, Springel & White 2005; Wechsler et al. 2006; Dalal et al. 2008; Mao, Zentner & Wechsler 2018; Salcedo et al. 2018; Mansfield & Kravtsov 2020).

As a test of our model, we construct halo catalogues with different amounts of concentration and spin secondary bias, splitting the sample by quartile. The magnitude of the effect varies differently as a function of mass for each secondary bias parameter. Thus, we adopt separate halo mass bins for each parameter, in a regime where we have both reliable estimates of the secondary quantities and know that the secondary bias effect is not drastic, following fig. 4 of Sato-Polito et al. (2019). The mass range $12 \le \log_{10}\left(\frac{M}{h^{-1}\mathrm{M}_\odot}\right) \le 12.5$ was used to build samples contaminated with concentration bias, and $12.5 \le \log_{10}\left(\frac{M}{h^{-1}\mathrm{M}_\odot}\right) \le 13$ for spin bias. We consider the highest and lowest quartile samples in both concentration and spin, as well as a sample with no secondary bias, subsampled to the same number density as the samples contaminated with secondary bias. We additionally do not subtract the shot-noise contribution from measured spectra, and opt instead to include it in our covariance matrix as detailed in equation (14).

Using the emulator for the basis spectra evaluated at the cosmology of these test boxes, we jointly fit the halo–halo and halo–matter spectra $\{P_{hh}, P_{hm}\}$ with five parameters: $b_i = \{b_1, b_2, b_{s^2}, b_{\nabla^2}, \bar{n}^{-1}\}$. We minimize the χ^2 between the mean of five simulations assuming

a disconnected covariance for the observables as described in equation (14), with $V = 5 \times (1.05 \, h^{-1}\mathrm{Gpc})^3$ each. The resulting fits are shown in Fig. 6.

We fit the spectra to a maximum scale of $k_{\mathrm{max}} = 0.6 \, h \, \mathrm{Mpc}^{-1}$. For most panels, we see that the hybrid *N*-body/Lagrangian bias model can jointly describe the clustering and lensing spectra to within 1 per cent down to scales even smaller than employed for the model fit. At large scales, the lowest spin assembly bias bin seems to be systematically higher by at most 10 per cent. Changing the k_{max} of the fit down to $0.2 \, h \, \mathrm{Mpc}^{-1}$ does not qualitatively alleviate the large-scale discrepancies. We observe similar behaviour if the average of the basis spectra from this cosmology are used instead of the emulator, implying this is not an issue of the emulator and could perhaps be attributed to large-scale noise. Another possibility is that a second-order Lagrangian bias model is unable to fully capture the effects of spin secondary bias, but we leave this investigation to future work.

In Fig. 7, we show the reduced χ^2 for the fits to the samples split on concentration. We see that the goodness of fit degrades significantly past $k \simeq 0.6 \, h \, \mathrm{Mpc}^{-1}$ for some subsamples. The fits to smaller k_{max} have $\chi^2/\mathrm{d.o.f.} \lesssim 1.5$. Note that in these tests we use the emulator at a volume that is significantly larger than the boxes it was trained on, and the covariance matrices do not have any contributions due
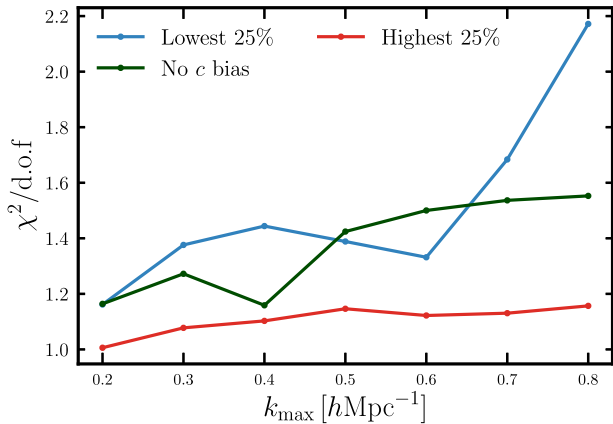
Figure 7. The goodness of fit χ^2/d.o.f. from increasing k_{max} for the halo sample selected on concentration quartiles, using the emulator as a model. Note the significant degradation of the goodness of fit for the subsample split on the lowest quartile after $k_{max} = 0.6$.

to the emulator uncertainty. If we instead use the mean basis spectra the χ^2/d.o.f. cross the χ^2/d.o.f. ~ 1 threshold at $k_{max} \sim 0.6 h\,\mathrm{Mpc}^{-1}$ and grow significantly afterwards, signalling a potential breakdown of the applicability of this Lagrangian bias model to these samples.

Baryonic physics is known to impact the statistics of biased tracers at the scales we are considering (White 2004; Zhan & Knox 2004; Chisari et al. 2019; van Daalen, McCarthy & Schaye 2020). In our model, the $\langle 1, \nabla^2\delta \rangle$ basis spectrum should have the scale dependence required to capture the first-order impacts of baryons (Lewandowski et al. 2015). In order to test this, we produce mock 'baryonified' spectra using the fitting function of van Daalen et al. (2020), which is obtained from analysis of a comprehensive suite of hydrodynamic simulations. We compare the fitting function to two parametrizations for the impact of baryons:

(i) Including terms that scale as the basis functions $b_{\nabla^2}\langle 1, \nabla^2\delta \rangle$ and $b_1 b_{\nabla^2}\langle \delta, \nabla^2\delta \rangle$.

(ii) Same as above, but substituting the basis functions with the approximation $\langle X, \nabla^2\delta \rangle \simeq -k^2\langle X, 1\rangle$.

The results of this test are shown in Fig. 8. While the baryonic suppression factors presented by the two parametrizations differ, in the bottom panel we see that both capture the effects of baryons to within 1 per cent out to $k \approx 0.8\,h\,\mathrm{Mpc}^{-1}$, whereas not including the contributions leads to errors larger than 1 per cent at $k \approx 0.2\,h\,\mathrm{Mpc}^{-1}$.

Additionally, our framework can simultaneously treat the effects of finite halo size and baryonic physics. As both are captured by the same basis spectra, this corresponds to treating the halo tracer as having one set of b_{∇}^2 and the matter tracer in the P^{hm} correlation as having a separate higher derivative coefficient b'_{∇^2}, while keeping all other bias parameters equal to zero.

6.3 Recovering input cosmology

In this section, we present an increasingly complex series of tests to ensure our emulator can be used for cosmological inference, i.e. to demonstrate that it can recover input cosmological parameters in an unbiased way. The general structure of the analyses we run is as follows.

The input data vectors will be the joint halo–halo and halo–matter power spectra $\boldsymbol{d} = \{P_{hh}(k), P_{hm}(k)\}$. We assume a Gaussian
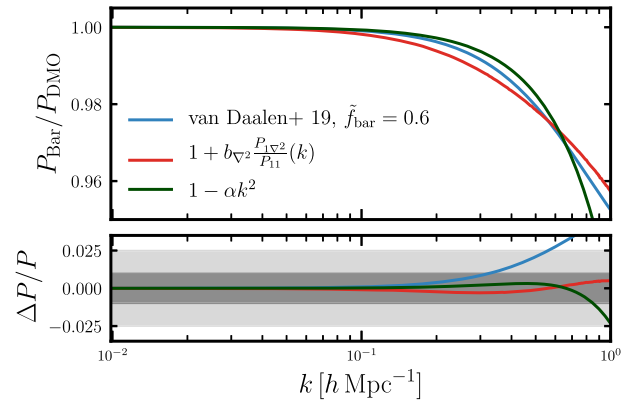

Figure 8. Higher derivative bias terms and their comparison to the baryonic physics fitting function of van Daalen et al. (2020). The top panel shows the fitting function, the basis spectrum as measured in the *N*-body simulations and the approximation we employ in the text. In the lower panel, we show residuals between the different treatments and the fitting function. The blue curve in the lower panel shows the difference between the unprocessed dark matter power spectrum and the fitting function. The green curve is the approximation to the higher derivative fitting functions that is implemented in our analyses.

likelihood in the residuals between $\boldsymbol{d}$ and the emulator prediction at a cosmology $\boldsymbol{x}(\Omega)$:

$$\log \mathcal{L}(d|\Omega) \propto -(\boldsymbol{d} - \boldsymbol{x}(\Omega))^T \boldsymbol{C}^{-1}(\boldsymbol{d} - \boldsymbol{x}(\Omega)). \quad (13)$$

We adopt a baseline covariance matrix that includes only dependence on the two-point functions of the tracer density field, known as the disconnected contribution (Li et al. 2019). The result is a block-diagonal matrix with format

$$\boldsymbol{C}(k, k') \equiv \frac{2\pi^2 \delta_{k,k'}}{k^2 \Delta k V} \times \begin{cases} 2P_{hh}^2(k), & \text{for } hh \times hh \\ 2P_{hh}(k)P_{hm}(k), & \text{for } hh \times hm \\ \left[P_{hh}(k)P_{mm}(k) & \text{for } hm \times hm \right. \\ \left. + P_{hm}^2(k) \right], & \end{cases} \quad (14)$$

for each subblock. We use non-linear power spectra and P_{hh} includes the shot-noise contribution. At the smaller scales, we probe in the resulting analyses, the purely disconnected approximation is known to fail and off-diagonal (connected) components become increasingly important (Meiksin & White 1999; Scoccimarro, Zaldarriaga & Hui 1999; Cooray & Hu 2001; Mohammed, Seljak & Vlah 2016; Lacasa 2018). The intent of this paper is not to conclusively quantify the information content available at small scales. Rather, we would like to ensure that the emulator is an unbiased model when pushing to such small scales. Therefore, we consider the form of the covariance in equation (14) to be a sufficient baseline to carry out our analyses. We assess its performance in more detail in Appendix A. As we will discuss in more detail in Section 6.3.2, the approximation of taking only the disconnected contribution neglects two forms of error: that arising from the connected contribution, and model error from the emulator itself. We discuss the contribution to the covariance from emulator error in Appendix A, and find that in the regime under which our tests are carried out, its inclusion is important in achieving unbiased constraints.

We sample the posterior distributions of the model parameters via Markov chain Monte Carlo (MCMC), using EMCEE (Goodman & Weare 2010; Foreman-Mackey et al. 2013). Chains are run with either $N = 64$ or $N = 128$ walkers across 8000 (4000) steps, respectively. We

Table 1. Boundaries of the cosmological parameters of simulations spanned by the Aemulus training suite. These are the values used as flat priors for cosmological parameters.

Parameter	Range
$\Omega_b h^2$	[0.0207, 0.0237]
$\Omega_c h^2$	[0.101, 0.132]
w_0	[−1.399, −0.566]
n_s	[0.928, 0.997]
σ_8	[0.575, 0.964]
H_0	[61.69, 74.77]
$N_{\rm eff}$	[2.62, 4.28]

checked that these values ensure converged chains for the simulated likelihood analyses we run; the posteriors are not altered significantly by doubling the length or number of walkers. We adopt wide uniform priors on the bias parameters,

$$b_i \sim U(-5, 5), \tag{15}$$

and uniform priors surrounding the boundaries of the Aemulus training suite, specified in Table 1.

6.3.1 Synthetic data

As a first test of the emulator, we perform a simulated likelihood analysis on a noiseless data vector drawn from the emulator itself. We fit the basis spectra to a halo sample of mass $12 \leq \log_{10} M_h/M_\odot \leq 12.5$ from one of Aemulus' test boxes. The cosmology and best-fitting bias values are used as inputs to the emulator to produce a mock noiseless data vector. As the data in this test are not a random draw from a distribution, the exact format of the covariance matrix does not matter. However, we use the block-diagonal disconnected covariance of equation (14) with $V = (1050\,h^{-1}{\rm Mpc})^3$ so as to replicate an analysis on an individual Aemulus test box.

The results of this first mock analysis are shown in in Fig. 9. The three-parameter analysis constrains all cosmological and bias parameters in an unbiased fashion, indicating that there are no issues in fitting the emulator to itself at this volume. We also conduct seven-parameter analysis for wCDM parameters. The results returns unbiased posteriors relative to the true input values, however it is hard to constrain all wCDM parameters using a single halo sample at the volume of a single Aemulus box. For this reason, several of the cosmological parameters simply saturate the priors and remain unconstrained.

6.3.2 Halo samples from the test suite

A subsequent test we perform is inference on halo catalogues drawn from the Aemulus test suite. We refer to Aemulus I (DeRose et al. 2019b) for details on how the halo finding procedure was done. This fiducial halo sample contains the mass bin $13 \leq \log_{10}\left(\frac{M_h}{h^{-1}{\rm M}_\odot}\right) \leq 13.5$ at $z = 0.4$.

We run a suite of chains for this halo sample, to assess emulator performance in terms of inferring cosmological parameters. We measure the halo–halo and halo–matter power spectra for each independent test box across the seven different test cosmologies. The data vector is averaged over the five independent realizations from the test suite. This set of chains allows us to assess the emulator biases such that they are less susceptible to projection effects. We
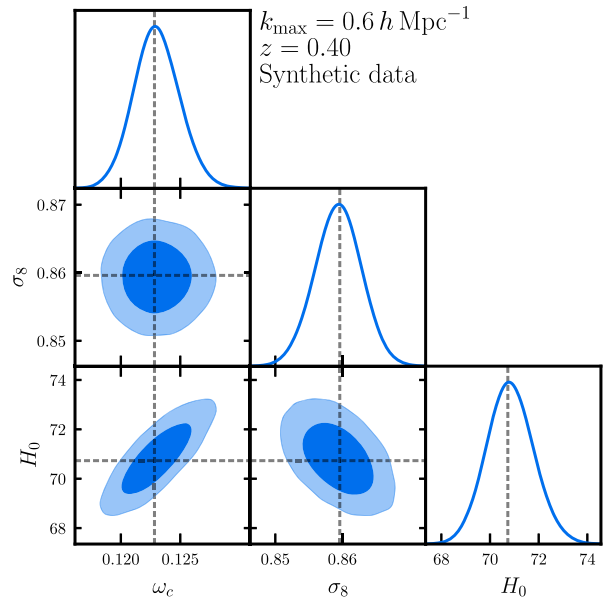
Figure 9. Cosmological parameter inference using the emulator where the data are a noiseless draw from itself. We vary the subset of parameters ω_c, σ_8, and H_0, using a Gaussian likelihood and purely disconnected covariance with volume $V = (1.05)^3 (h^{-1}{\rm Gpc})^3$. The fiducial values used to generate the data vector are shown in the dashed lines. The bias parameter posteriors are equally unbiased and Gaussian, but omitted from the figure for aesthetic purposes.

can also study the cosmology dependence of the emulator in this way and the interplay between the bias coefficients of our model and cosmological parameters.

Using solely the purely disconnected covariance matrix in equation (14) leads to strong biases in inferred cosmological parameters, despite all residuals being smaller than 1 per cent as a function of scale. This can be understood by the fact that sample variance at small scales will eventually become smaller than the 1–2 per cent emulator error observed in Fig. 5 (see also Fig. A1). However, the aforementioned figure allows us to estimate the emulator uncertainty as a function of scale. This can then be included as a separate contribution to the covariance matrix. We detail how this is done in Appendix A.[1]

The result of the test is shown in Fig. 10, with the full set of contours shown in Fig. C1. The cosmological parameters inferred scatter around the best fits for ω_c and σ_8, whereas they recover H_0 to within one standard deviation for most cosmologies but biased slightly high. However, we note these tests are conservative, as they neglect the contribution to the covariance matrix arising from shape noise, the lensing equivalent of shot-noise that would contribute to the $hm \times hm$ term of the covariance matrix. Given the conservative nature of this test we deem the emulator performance to be sufficient and continue with the final and most stringent test we consider in this work.

6.3.3 A redMaGiC sample from an independent simulation

So far, we have reported tests performed on samples that originate either from the emulator itself or from the same suite of simulations used to construct it. It is also important that the model is useful

---

[1]All contour plots shown from this point forward will include the effects of emulator error unless stated otherwise.
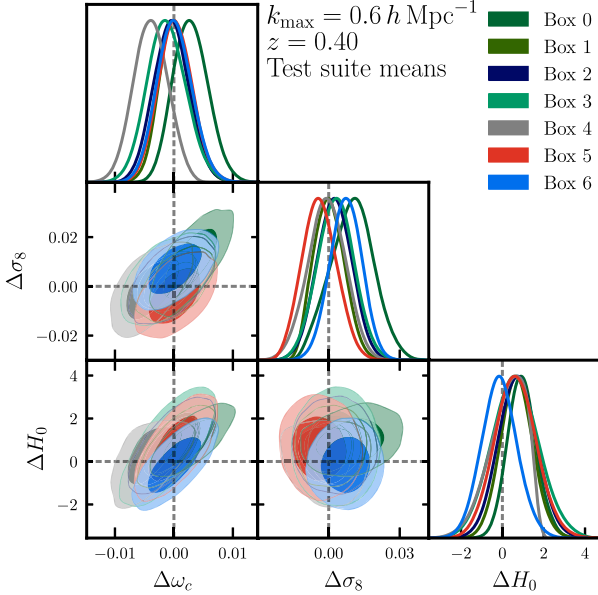
Figure 10. Cosmological parameter inference using the emulator fit to the mean of five realizations at the seven `Aemulus` test cosmologies. We vary the cosmological parameters ω_c, σ_8, and H_0, using a disconnected covariance with volume $V = (1.05)^3 (h^{-1}\text{Gpc})^3$, including a contribution arising from correlated emulator residuals. The contours are shown in the space of differences relative to the true cosmology of each box.

Table 2. HOD parameters used to populate the `redMaGiC` sample described in Section 6.3.3.

$\log M_{\min}$	$\sigma_{\log M}$	f_c	$\log M_0$	$\log M_1'$	α
12.1	0.4	0.13	11.45	13.73	1.48

for inference on spectra measured from tracer samples generated by independent methods, both in how halo samples are defined and the underlying N-body simulation used. For example, in Modi et al. (2020) it was shown that this hybrid Lagrangian bias model can successfully fit galaxy power spectra produced from an HOD (see e.g. Zheng et al. 2005).

We perform a final test: a simulated likelihood analysis with spectra produced from populating an independent N-body simulation with an HOD that matches the density and clustering properties of `redMaGiC` galaxies (Rozo et al. 2016). `redMaGiC` galaxies are the primary photometric luminous red galaxy sample used in current and future weak-lensing surveys (Elvin-Poole et al. 2018).

The HOD parametrization we adopt is an extension of the model presented in Zheng, Coil & Zehavi (2007), allowing for the central occupation at high mass to be less than unity

$$\langle N_{\text{cen}}(M) \rangle = \frac{f_c}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right], \quad (16)$$

$$\langle N_{\text{sat}}(M) \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right] \left(\frac{M - M_0}{M_1'} \right)^\alpha \quad (17)$$

The HOD parameters corresponding to the `redMaGiC` samples used can be found in Table 2, and are derived from a `redMaGiC` sample selected from simulations similar to those presented in DeRose et al. (2019a).

We paint `redMaGiC` galaxies on to halo catalogues measured from the `UNIT` simulations (Chuang et al. 2019) at $z \approx 0.59$, a redshift different from the `Aemulus` snapshots. A `UNIT` realization boasts a comparable volume to `Aemulus` of $V = 1 (h^{-1}\text{Gpc})^3$ at a

significantly higher number of particles, $N = (4096)^3$. Every `UNIT` simulation has two realizations with opposite phases and fixed amplitudes. Averaging two-point statistics measured from these paired–fixed realizations leads to very high sample variance suppression at large scales, comparable to averaging $\sim$150 simulations of the same volume.

The cosmological parameter constraints corresponding to this test are shown in Fig. 11. The emulator recovers the input cosmology of `UNIT` within its 68 per cent contours. Although this test is idealized, the constraints inferred are promising if they translate even moderately well to a realistic analysis: a 2.5 per cent constraint on ω_c, a 0.5 per cent constraint on σ_8, and a 1.6 per cent constraint of H_0. In a realistic lensing analysis, one would expect these quantities to be degraded due to the inclusion of shape noise and only having access to two-dimensional lensing maps instead of the 3D matter field. Nevertheless, even a 100 per cent degradation of these constraints due to the aforementioned complications would still result in highly competitive measurements of these parameters. Note we adopt no priors beyond the (moderately informative) priors set by the boundaries of the `Aemulus` suite.

The simulated likelihood analysis performed on this sample additionally allow us to quantify both model and emulator errors in a space that is closer to observations that will be carried out in the near future. As `redMaGiC` galaxies are commonly used as lens samples in galaxy–galaxy lensing analyses, we can translate the P^{hh}, P^{hm} residuals to those in the observables C_ℓ^{gg}, $C_\ell^{g\kappa}$. We assume a redshift distribution $n(z)$ for `redMaGiC` galaxies consistent with data (Elvin-Poole et al. 2018) and fiducial parametrizations for the source sample that are consistent with those that will be achieved in future imaging surveys (Mandelbaum et al. 2018). For a `redMaGiC` sample spanning $z = [0.45, 0.6]$, we present the results in Fig. 12. The harmonic space observables are calculated assuming the Limber approximation, with the additional approximation that the residuals between 3D power spectra do not evolve as a function of redshift. The residuals stay within 1 per cent out to $\ell \approx 1000$. If we instead use residuals from fitting the emulator at fixed cosmology to the same sample out to $k_{\max} = 1.0 \, h \, \text{Mpc}^{-1}$, the residuals remain within 10 per cent out to $\ell_{\max} = 2000$, at the cost of worse performance at large scales. This indicates that the combined emulator and model error remain well under control for the analysis of current galaxy–galaxy lensing data sets.

7 CONCLUSIONS

In this work, we have built an emulator to study the cosmology dependence of the model of Modi et al. (2020) for the two-point statistics of biased tracers. The model combines N-body simulations with a symmetries-based bias expansion to provide accurate predictions beyond the regime of validity of standard perturbative approaches.

Specifically, we built an emulator for the cosmology and redshift dependence of the 10 non-linear basis functions that span this model. We use measurements from the `Aemulus` suite of simulations, which has been designed to enable the construction of emulators that satisfy the modelling requirements of upcoming cosmic surveys. The model and emulation techniques used are general; there are no limitations to extending the range of validity given the availability of an improved suite of simulations.

We find that:

(i) The emulator recovers each basis spectrum to $\lesssim$1 per cent accuracy across a wide range of scales, $0.1 < k/(h^{-1} \text{Mpc}) \leq 1.0$, and redshifts, $0 \leq z \leq 2$.
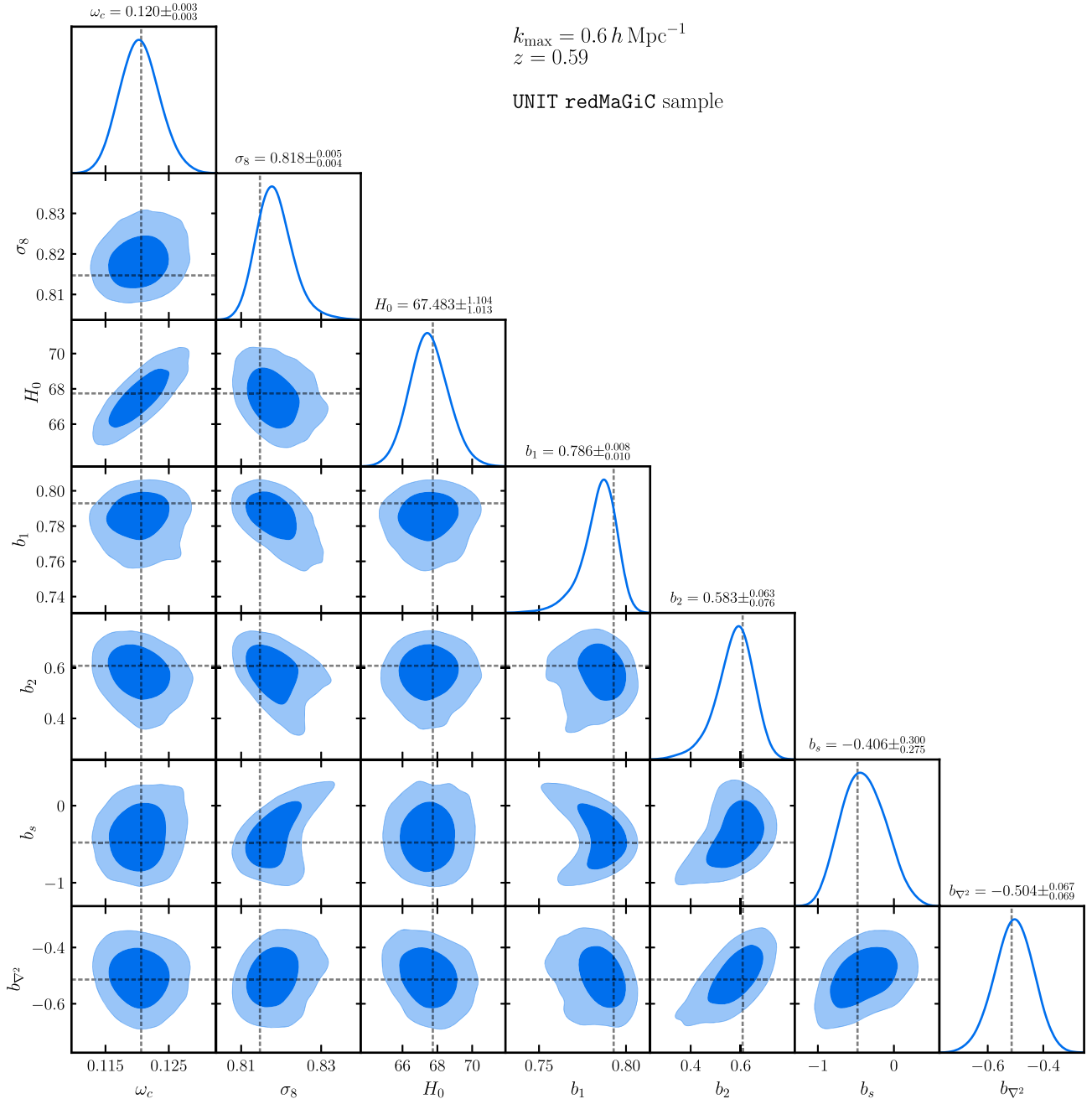
Figure 11. Cosmological parameter constraints from the `redMaGiC` sample constructed from the `UNIT` simulations. The true cosmological parameters and the best-fitting bias parameters assuming the true cosmology are shown in the dashed lines. All parameters are recovered to well within the 1σ errors.

(ii) The Lagrangian bias model is capable of capturing the clustering and lensing statistics of samples imbued with non-trivial amounts of secondary bias and contamination from baryonic physics.

(iii) The test set used to validate the emulator can also be used to calibrate its 'theoretical uncertainty'. This allows us to include contributions to the covariance matrix of an analysis related to model error, which cannot be neglected when pushing to small scales.

(iv) The emulator, as constructed, can recover unbiased cosmological parameters from realistic simulated likelihood analyses.

These findings indicate that our emulator is a robust tool that can be readily applied to analyses of current and even upcoming data sets. The code will be made publicly available and can be integrated

with modern sampling packages such as COBAYA (Torrado & Lewis 2020). We also point out a few further directions to be investigated as a result of this work.

First, while the simulations used here are sufficient to obtain per cent level emulator accuracy, improved simulations will be important for maximizing the applicability of this model. The biggest immediate limitation of this emulator is the extent of the cosmological parameter space that it is trained on. We plan on running simulations over a broader parameter space, including massive neutrinos, in the near future. Another limiting factor in the current emulator construction is our ability to match the basis spectra measured from our simulations to their perturbation theory analogs at low k. Running larger simulation volumes, or implementing a method
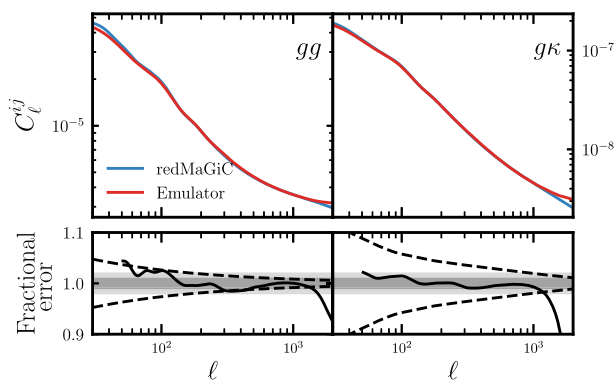
Figure 12. Residuals of the emulator fit to the `redMaGiC` sample in the space of a projected analysis. Residuals are shown for the range $\ell \in [50, 2000]$. The redshift distributions of this analysis are consistent with those of current and upcoming surveys. The dashed envelope corresponds to the sample variance contribution in the absence of noise with sky coverage consistent with upcoming surveys and angular binning of $\Delta\ell = 50$. That is, shot/shape noise will only increase the size of this envelope. The light grey and dark grey bands correspond to 2 and 1 per cent error bands, respectively.

for sample variance mitigation such as that presented in Chartier et al. (2020), would ameliorate this issue by reducing noise in the *N*-body measurements. This will allow them to be matched more easily to the perturbation theory predictions at scales that are still safely within perturbative reach. Mismatches in the linear growth predictions from *N*-body simulations also limit the accuracy of the large-scale matching. Simulations with more stringent time-stepping criteria would reduce these inaccuracies, at the cost of increased run-time. For this reason, methods that explicitly enforce linear growth on large scales may be worth exploring in the future (Howlett, Manera & Percival 2015; Feng et al. 2016). Finally, the accuracy of the model for redshift evolution of the basis spectra in the current emulator is limited by the number of snapshots saved in the `Aemulus` suite. For this reason, saving snapshots with finer redshift resolution out to higher redshifts will be a priority when running future simulations to upgrade the current emulator.

While in this paper we have restricted ourselves to predictions of survey observables in Fourier space, one could use this same field-level approach to measure configuration-space correlation statistics instead. The model employed should also be able to describe the statistics of biased tracers at the field level, beyond two-point statistics. This includes both field-level characterizations of the Lagrangian bias model similarly to what was investigated in Schmittfull et al. (2019) and higher order functions such as the bispectrum or the collapsed tri-spectra that form the connected component of covariance matrices.

The field-level approach to bias modelling described in Schmittfull et al. (2019) was recently extended to redshift space (Schmittfull et al. 2020). For our emulator to be used to describe the statistics of 3D galaxy clustering in spectroscopic galaxy surveys, it would need to be extended to redshift space in a similar manner. Alternatively, we note that the bias parameters in this model are equivalent to those of the Lagrangian perturbation theory of Chen et al. (2020a, b). This suggests one could perform a joint analysis that combines perturbation theory for describing the redshift-space clustering, where the 3D nature of the measurements allow tight constraints even on quasi-linear scales, and an emulator for describing projected statistics, which need to extend to smaller scales in order to beat down sample variance. In addition to providing a large dynamic range and sensitivity to both metric potentials, the combination of

measurements would help to break bias parameter degeneracies and thus improve cosmological constraints.

The second release of the `Aemulus` suite, `Aemulus-ν`, will include two-fluid simulations that capture the effects of massive neutrinos on the matter density field. The techniques described in this paper can be translated to this new set of simulations to construct an emulator that can be used to constrain the sum of neutrino masses, one of the key science drivers of ongoing and future cosmological surveys.

We leave these extensions to future work.

DATA AVAILABILITY

The data underlying this article are available in the `Aemulus` Project's website.

REFERENCES

Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Abidi M. M., Baldauf T., 2018, J. Cosmol. Astropart. Phys., 07, 029
Aghamousa A. et al., 2016, preprint (arXiv e-prints)
Alam S., Miyatake H., More S., Ho S., Mandelbaum R., 2017, MNRAS, 465, 4853
Angulo R. E., Pontzen A., 2016, MNRAS, 462, L1
Aviles A., Banerjee A., 2020, J. Cosmol. Astropart. Phys., 2020, 034
Aviles A., Valogiannis G., Rodriguez-Meza M. A., Cervantes-Cota J. L., Li B., Bean R., 2020, preprint (arXiv:2012.05077)
Bagla J. S., 2005, Curr. Sci., 88, 1088
Baldauf T., Mirbabayi M., Simonović M., Zaldarriaga M., 2016a, preprint (arXiv e-prints)
Baldauf T., Schaan E., Zaldarriaga M., 2016b, J. Cosmol. Astropart. Phys., 03, 007
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Baumann D., Nicolis A., Senatore L., Zaldarriaga M., 2012, J. Cosmol. Astropart. Phys., 2012, 051
Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bianchini F. et al., 2015, ApJ, 802, 64
Blas D., Garny M., Konstandin T., 2014, J. Cosmol. Astropart. Phys., 2014, 010

Blatman G., Sudret B., 2008, Comptes Rendus Mecanique, 336, 518

Blatman G., Sudret B., 2011, J. Comput. Phys., 230, 2345

Carrasco J. J. M., Hertzberg M. P., Senatore L., 2012, J. High Energy Phys., 09, 082

Chartier N., Wandelt B., Akrami Y., Villaescusa-Navarro F., 2020, preprint (arXiv:2009.08970)

Chen S.-F., Vlah Z., Castorina E., White M., 2020a, Redshift-Space Distortions in Lagrangian Perturbation Theory, preprint (arXiv:2012.04636)

Chen S.-F., Vlah Z., White M., 2020b, J. Cosmol. Astropart. Phys., 07, 062

Chen S.-F., Vlah Z., White M., 2020c, J. Cosmol. Astropart. Phys., 11, 035

Chisari N. E. et al., 2019, Open J. Astrophys., 2, 4

Chuang C.-H. et al., 2019, MNRAS, 487, 48

Chudaykin A., Ivanov M. M., Simonović M., 2020, preprint (arXiv e-prints)

Cooray A., Hu W., 2001, ApJ, 554, 56

Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369

Crocce M., Pueblas S., Scoccimarro R., 2012, Astrophysics Source Code Library, record ascl:1201.005

D'Amico G., Senatore L., Zhang P., Zheng H., 2020, preprint (arXiv:2006.12420)

Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12

DeRose J. et al., 2019a, preprint (arXiv:1901.02401)

DeRose J. et al., 2019b, ApJ, 875, 69

Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1

DiPompeo M. A., Hickox R. C., Eftekharzadeh S., Myers A. D., 2017, MNRAS, 469, 4630

Doré O. et al., 2015, Cosmology with the SPHEREX All-Sky Spectral Survey, preprint (arXiv:1412.4872)

Doré O. et al., 2019, WFIRST: The Essential Cosmology Space Observatory for the Coming Decade, preprint (arXiv:1904.01174)

Elvin-Poole J. et al., 2018, Phys. Rev. D, 98, 042006

Favole G. et al., 2020, MNRAS, 497, 5432

Feinberg J., Langtangen H. P., 2015, J. Comput. Sci., 11, 46

Feinberg J., Eck V. G., Langtangen H. P., 2018, SIAM J. Sci. Comput., 40, A199

Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Fujita T., Vlah Z., 2020, J. Cosmol. Astropart. Phys., 2020, 059

Fujita T., Mauerhofer V., Senatore L., Vlah Z., Angulo R., 2020, J. Cosmol. Astropart. Phys., 01, 009

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Garrison L. H., Eisenstein D. J., Ferrer D., Metchnik M. V., Pinto P. A., 2016, MNRAS, 461, 4125

Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65

Guo H. et al., 2019, ApJ, 871, 147

Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, ApJ, 156, 160

Harris C. R. et al., 2020, Nature, 585, 357–362

Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, ApJ, 705, 156

Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104

Heymans C. et al., 2020, preprint (arXiv:2007.15632)

Hockney R., , Eastwood J., 1988, Computer Simulation Using Particles. CRC Press

Hoffman M. D., Gelman A., 2011, preprint (arXiv:1111.4246)

Howlett C., Manera M., Percival W. J., 2015, Astron. Comput., 12, 109

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Ivanov M. M., McDonough E., Hill J. C., Simonović M., Toomey M. W., Alexander S., Zaldarriaga M., 2020, Phys. Rev. D, 102, 103502

Ivezić v. et al., 2019, ApJ, 873, 111

Joyce M., Garrison L., Eisenstein D., 2020, MNRAS, 501, 5051

Knabenhans M. et al., 2019, MNRAS, 484, 5509

Krause E. et al., 2017, preprint (arXiv e-prints)

Krolewski A., Ferraro S., Schlafly E. F., White M., 2020, J. Cosmol. Astropart. Phys., 2020, 047

Kuhlen M., Vogelsberger M., Angulo R., 2012, Phys. Dark Univ., 1, 50

Kwan J., Heitmann K., Habib S., Padmanabhan N., Lawrence E., Finkel H., Frontiere N., Pope A., 2015, ApJ, 810, 35

Lacasa F., 2018, A&A, 615, A1

Laguë A., Bond J. R., Hložek R., Marsh D. J., Söding L., 2020, preprint (arXiv e-prints)

Laureijs R. et al., 2011, Euclid Definition Study Report, preprint (arXiv:1110.3193)

Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, ApJ, 713, 1322

Lazeyras T., Schmidt F., 2018, J. Cosmol. Astropart. Phys., 2018, 008

Lazeyras T., Schmidt F., 2019, J. Cosmol. Astropart. Phys., 11, 041

Lewandowski M., Perko A., Senatore L., 2015, J. Cosmol. Astropart. Phys., 05, 019

Lewis A., 2019, GetDist: a Python package for analysing Monte Carlo samples, preprint (arXiv:1910.13970)

Li Y., Singh S., Yu B., Feng Y., Seljak U., 2019, J. Cosmol. Astropart. Phys., 2019, 016

MacCrann N., Blazek J., Jain B., Krause E., 2020, MNRAS, 491, 5498

Mandelbaum R. et al., 2018, preprint (arXiv e-prints)

Mandelbaum R., 2018, ARA&A, 56, 393

Mansfield P., Avestruz C., 2020, MNRAS, 500, 3309

Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, MNRAS, 474, 5143

Matsubara T., 2008, Phys. Rev. D, 78, 083519

McClintock T. et al., 2019, ApJ, 872, 53

McDonald P., Roy A., 2009, J. Cosmol. Astropart. Phys., 2009, 20

McLaughlin S., Wechsler R. H., Banerjee A., DeRose J., Mao Y.-Y., Tinker J. L., Zhai Z., 2021, preprint (arXiv e-prints)

McQuinn M., White M., 2016, J. Cosmol. Astropart. Phys., 2016, 043

Meiksin A., White M., 1999, MNRAS, 308, 1179

Michaux M., Hahn O., Rampf C., Angulo R. E., 2020, MNRAS, 500, 663

Modi C., White M., Vlah Z., 2017, J. Cosmol. Astropart. Phys., 2017, 009

Modi C., Chen S.-F., White M., 2020, MNRAS, 492, 5754

Mohammed I., Seljak U., Vlah Z., 2016, MNRAS, 466, 780

Nishimichi T., D'Amico G., Ivanov M. M., Senatore L., Simonović M., Takada M., Zaldarriaga M., Zhang P., 2020, preprint (arXiv e-prints)

Omori Y. et al., 2019, Phys. Rev. D, 100

Park Y., Rozo E., Krause E., 2020, preprint (arXiv e-prints)

Peacock J. A., Bilicki M., 2018, MNRAS, 481, 1133

Power C., Robotham A. S. G., Obreschkow D., Hobbs A., Lewis G. F., 2016, MNRAS, 462, 474

Prat J. et al., 2018, Phys. Rev. D, 98 , 42005

Pullen A. R., Alam S., He S., Ho S., 2016, MNRAS, 460, 4098

Rozo E. et al., 2016, MNRAS, 461, 1431

Salcedo A. N., Maller A. H., Berlind A. A., Sinha M., McBride C. K., Behroozi P. S., Wechsler R. H., Weinberg D. H., 2018, MNRAS, 475, 4411

Sato-Polito G., Montero-Dorta A. D., Abramo L. R., Prada F., Klypin A., 2019, MNRAS, 487, 1570

Savitzky A., Golay M. J. E., 1964, Anal. Chem., 36, 1627

Schmittfull M., Simonović M., Assassi V., Zaldarriaga M., 2019, Phys. Rev. D, 100 , 43514

Schmittfull M., Simonovic M., Ivanov M. M., Philcox O. H. E., Zaldarriaga M., 2020, Modeling Galaxies in Redshift Space at the Field Level, preprint (arXiv:2012.03334)

Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 2016, 047

Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1

Senatore L., Zaldarriaga M., 2017, preprint (arXiv:1707.04698)

Singh S., Mandelbaum R., Seljak U., Rodríguez-Torres S., Slosar A., 2019, MNRAS, 491, 51

Takada M. et al., 2014, PASJ, 66, R1

Taruya A., Nishimichi T., Jeong D., 2018, Phys. Rev. D, 98, 103532

Torrado J., Lewis A., 2020, Cobaya: Code for Bayesian Analysis of hierarchical physical models, preprint (arXiv:2005.05290)

van Daalen M. P., McCarthy I. G., Schaye J., 2020, MNRAS, 491, 2424

Villaescusa-Navarro F. et al., 2018, ApJ, 867, 137

Virtanen P. et al., 2020, Nat. Methods, 17, 261

Vlah Z., White M., Aviles A., 2015, J. Cosmol. Astropart. Phys., 2015, 014

Vlah Z., Castorina E., White M., 2016, J. Cosmol. Astropart. Phys., 2016, 007

Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435

Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

White M. J., 2004, Astropart. Phys., 22, 211

Wibking B. D. et al., 2019, MNRAS, 484, 989

Wibking B. D., Weinberg D. H., Salcedo A. N., Wu H.-Y., Singh S., Rodríguez-Torres S., Garrison L. H., Eisenstein D. J., 2020, MNRAS, 492, 2872

Wiener N., 1938, Am. J. Math., 60, 897

Xiu D., 2010, Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton Univ. Press, Princeton

Yoo J., Tinker J. L., Weinberg D. H., Zheng Z., Katz N., Dave R., 2006, ApJ, 652, 26

Yuan S., Eisenstein D. J., Garrison L. H., 2018, MNRAS, 478, 2019

Zhai Z. et al., 2019, ApJ, 874, 95

Zhan H., Knox L., 2004, ApJ, 616, L75

Zhang Y. et al., 2020, 501, 1013, MNRAS

Zheng Z. et al., 2005, ApJ, 633, 791

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

Zu Y., 2020, preprint (arXiv e-prints)

APPENDIX A: INCLUDING EMULATOR ERROR IN THE COVARIANCE MATRIX

As seen in Fig. 5, there is a scale-dependent error associated with our emulation scheme. This error is small, on the order of ~ 1 per cent, and within the accuracy requirements for the next generation of surveys. However, at the smallest scales we would like to test this model, $k \simeq 0.6\,h\,\mathrm{Mpc}^{-1}$, it will often be larger than the combined cosmic variance and shot noise (and absence of shape noise) of our tests. In this regime, the combination of using the average of only five boxes as our data, the approximate disconnected form of the covariance in equation (14) and failing to include model uncertainty in an analysis could then lead to biased inference on cosmological parameters (Baldauf et al. 2016a; Chudaykin, Ivanov & Simonović 2020).

Since the Aemulus test suite is composed of 35 simulations, at seven distinct points of cosmological parameter space, we can use the emulator residuals at these points to construct a model for the theoretical uncertainty. In this appendix, we discuss our procedure to construct this model and study its impact when employed in inference.

Let $\bar{P}_{XY}(k, \Omega_i)$ be the mean basis spectrum measured from five Aemulus boxes at the cosmology Ω_i. For a given box, we define
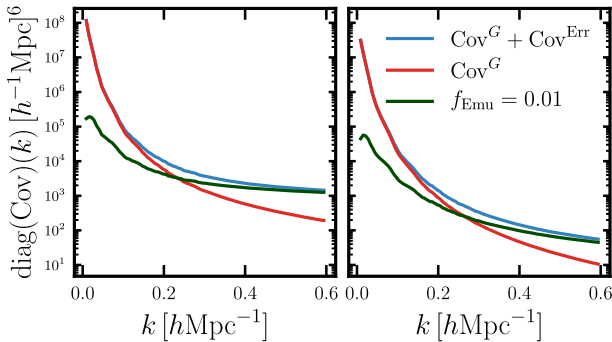

Figure A1. Comparison of our model for emulator uncertainty compared to the disconnected component of the covariance matrix. The left-hand panel corresponds to $P_{hh}P_{hh}$ contribution and the right-hand panel to $P_{hm}P_{hm}$.
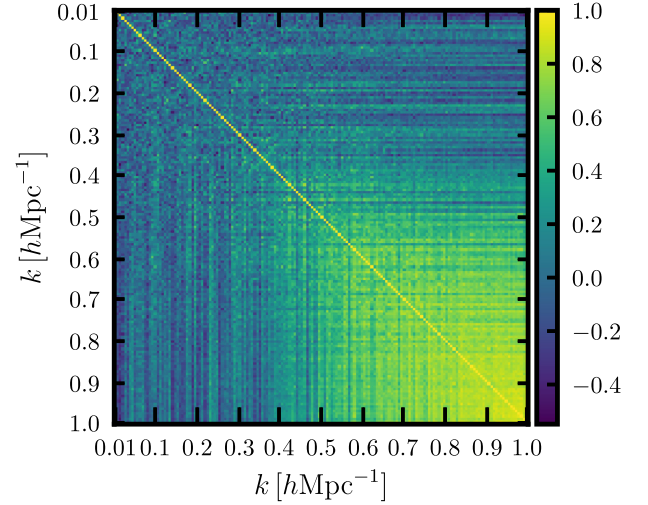

Figure A2. Correlation matrix of emulator residuals described in Appendix A. We see at small scales, past $k \simeq 0.4\,h\,\mathrm{Mpc}^{-1}$, the emulator residuals are significantly correlated.

normalized emulator residuals as

$$\hat{r}^{XY}(k) = \frac{\hat{P}_{XY}(k, \Omega_i) - P^{\mathrm{Emu}}_{XY}(k, \Omega_i))}{\bar{P}_{XY}(k, \Omega_i)}, \tag{A1}$$

where P^{Emu}_{XY} is the emulator prediction at the same cosmology. Normalized this way, we assume the residuals are cosmology independent. At each redshift, we have 35 sets of residuals. With these measurements, we can build an estimate of the residual correlation matrix

$$\mathrm{Corr}^{\mathrm{Emu}}(k, k') = \frac{\mathrm{Cov}[\hat{r}^{XY}(k), \hat{r}^{XY}(k')]}{\sqrt{\mathrm{Cov}(k, k)\mathrm{Cov}(k', k')}}, \tag{A2}$$

which captures how correlated the emulator residuals are across the test set as a function of scale. The quantities in the numerator and denominator of equation (A2) are the same, but we apply the shorthand $\mathrm{Cov}(k, k) \equiv \mathrm{Cov}[\hat{r}^{XY}(k), \hat{r}^{XY}(k)]$ to not overload the expression. This residual correlation matrix is shown in Fig. A2. We proceed to define an emulator floor, f_{Emu}, specifying what fraction of the signal is of the order emulator error. From Fig. 5, the dominant source of uncertainty will come from the error in the P_{11} spectrum. This implies $f_{\mathrm{Emu}} \simeq 0.01$ at small scales for redshifts $z > 0$. We then estimate that the emulator error will scale as

$$\mathrm{Cov}^{\mathrm{Err}}(k, k') = (f_{\mathrm{Emu}} P_{hh,hm}(k))^2 \times \mathrm{Corr}^{\mathrm{Emu}}(k, k'), \tag{A3}$$

where $P_{hh,hm}$ is used depending on whether we are including this contribution to the block corresponding to the halo–halo correlation or the halo–matter correlation. We then add this contribution in quadrature to equation (14)

$$\mathrm{Cov}(k, k') = \mathrm{Cov}^{G}(k, k') + \mathrm{Cov}^{\mathrm{Err}}(k, k'). \tag{A4}$$

We run chains with the covariance in equation (A4), as well as chains including only the diagonal contribution due to uncertainty, which we will call the 'floor' covariance. The contours for cosmological parameters are shown in Fig. A3. While this is clearly an approximate treatment, we observe that including this contribution helps prevent significant biases in cosmological parameter inference due to the noisy input data and very low noise assumed in the fit.
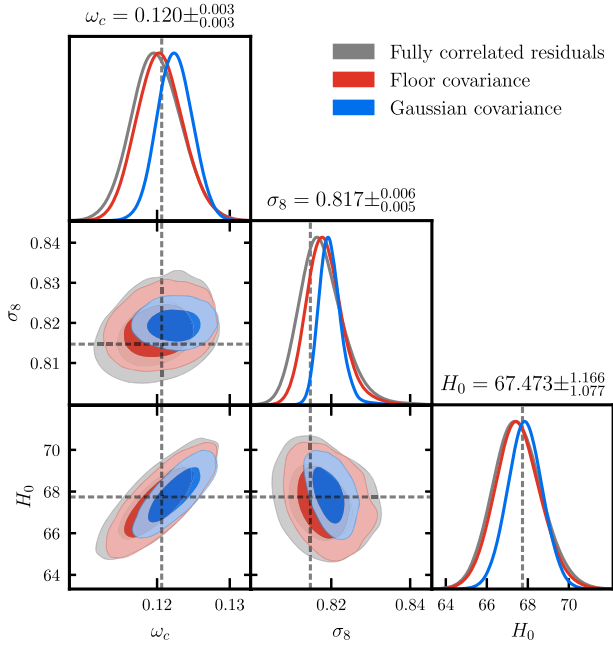
Figure A3. UNIT contours with the different covariance forms discussed in Appendix A. Chains are run with the standard scale cuts of $k_{max} = 0.6\,h^{-1}\mathrm{Mpc}$.

APPENDIX B: SUBSETS OF THE BIAS MODEL

A common critique of EFT-based models is that they are over-parametrized, and can fit to any signal due to the large number of free parameters. For perturbative Lagrangian bias models, this question has been previously explored in the context of CMB lensing cross-correlations. In Modi, White & Vlah (2017), it was shown that significant biases are obtained in σ_8 in these analyses if one uses a simplified model with linear galaxy bias and non-linear matter power spectra. To address whether this holds for our model, we run a series of tests of the emulator, with differing subsets of the bias parameters set to zero. The full set we adopt is

(i) 'All b_i's ', the full bias parametrization.
(ii) 'b_1 only', where $b_2 = b_{s^2} = b_{\nabla^2} = 0$.
(iii) 'b_1, b_{∇^2}', where $b_2 = b_{s^2} = 0$.
(iv) 'No b_{s^2}', where $b_{s^2} = 0$.
(v) 'No b_2', where $b_2 = 0$.

A contribution due to shot-noise is included in all of the chains. All chains in Fig. B1 are run with the same data vector and covariance matrices, and the k_{max} cuts highlighted in each row. We observe significant biases for every subset of bias parameters, except for the complete parametrization which recovers the input cosmological parameters as previously discussed in Section 6.3.3. This implies, at least in this simplified analysis, that the full set of bias parameters is required to achieve unbiased inference with this model.

To check the scale dependence of the importance of the full parametrization, the second row of Fig. B1 repeats this test limiting ourselves to $k_{max} = 0.4h\,\mathrm{Mpc}^{-1}$. The full bias model and the subset including only linear, quadratic and higher derivative biases perform comparatively well.
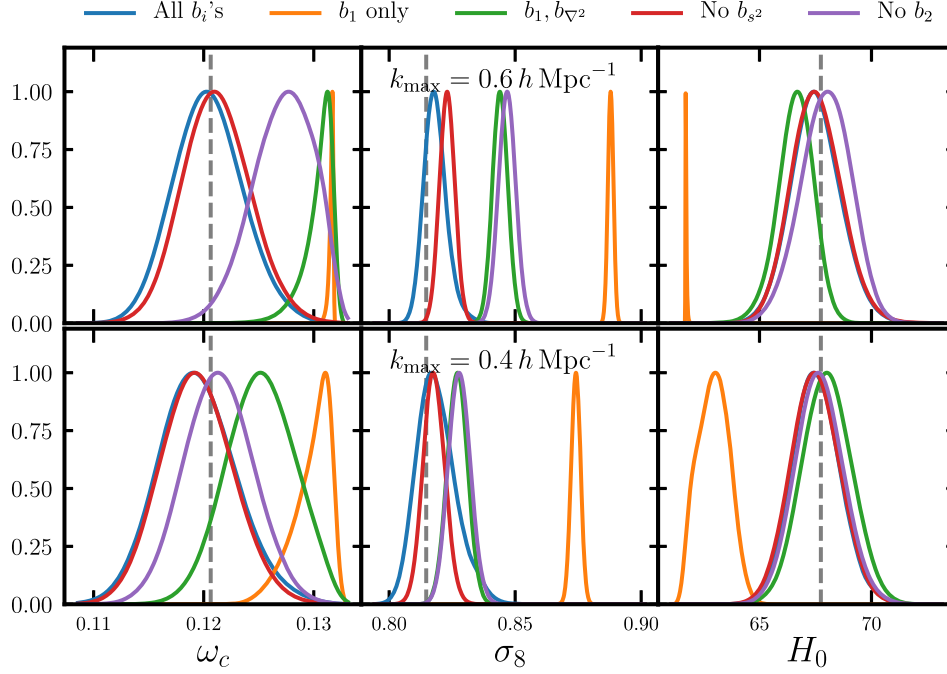
Figure B1. Posteriors for varying subsets of the bias model in equation (1), for two different scale cut configurations.

APPENDIX C: THE $k \to 0$ LIMIT OF THE EMULATOR

In this appendix, we investigate the impact of not correctly recovering large-scale linear growth in *N*-body simulations on the emulator, as highlighted in Section 1. We implement two different forms of enforcing consistency with linear theory at large scales:

(i) Strictly reverting to LPT at $k < k_{min}$. This introduces a 'kink' in the basis spectra predicted by the emulator.

(ii) Extrapolating the principal component predictions out to $k < k_{min}$, but with a filter to enforce linear growth.

The filter is applied to the $\Gamma^{XY}(k)$ that we use to build the emulator,

$$\Gamma^{XY}(k, \Omega) \to F(k)\Gamma^{XY}(k, \Omega). \tag{C1}$$

With this filtering approach, we recover LPT at large scales by construction without the discontinuity introduced by simply forcing LPT after some transition. The functional form we adopted for $F(k)$

is

$$F(k) = \frac{1}{2} \left[1 + \tanh\left(\alpha \frac{k - k_*}{k_*} \right) \right]. \tag{C2}$$

This quantity asymptotes to 0 at large scales, ensuring the Γ^{XY} are 0, and thus the ratios are consistent with unity. Fiducial values adopted are $k_* = 0.125$ and $\alpha = 2.5$ but the impact is similar for other values.

Since the samples we use to test the emulator are also derived from boxes with incorrect growth, for all figures in this paper we adopt a 'fiducial model' where we use the Γ^{XY} with no corrections at large scales. The emulator then has large-scale growth compatible with the boxes.

If we perform a simulated likelihood analysis with the other variants that enforce LPT at large scales, we see small shifts in some cosmological parameters away from their true values. The shifts in parameters are all less than one σ, and one must keep in mind that the noise levels in our analysis are quite stringent (for example, we have no shape noise in the simulated lensing constraint). When phrased in terms of the uncertainties in parameters obtained by recent analyses (Heymans et al. 2020), these shifts are less than $(1/4)\sigma$.
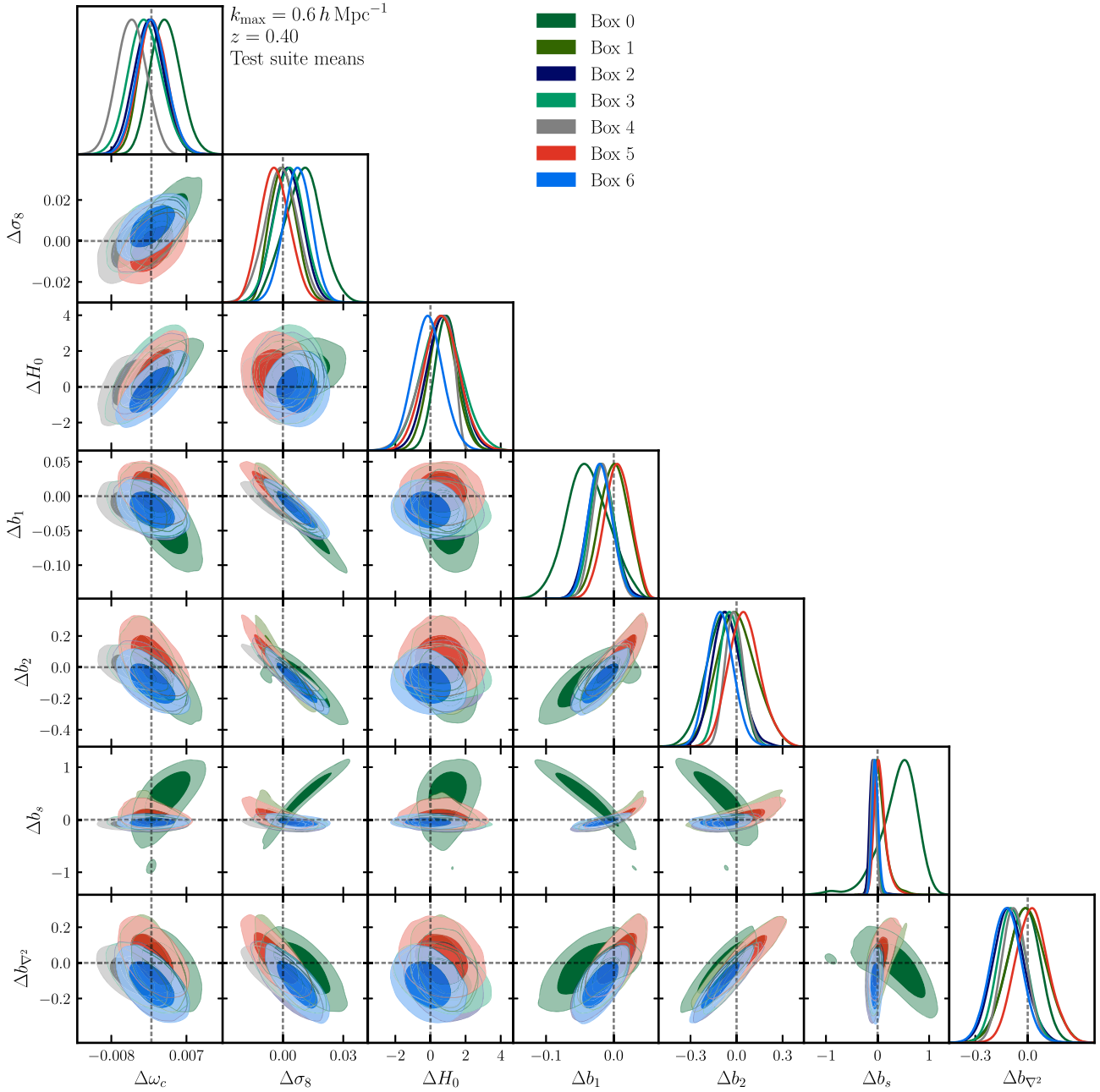
Figure C1. The same chains as Fig. 10 but showing all parameters varied.

This paper has been typeset from a TEX/LATEX file prepared by the author.