

PAPER

Aemulus v: precise predictions for matter and biased tracer power spectra in the presence of neutrinos

To cite this article: Joseph DeRose *et al* JCAP07(2023)054

View the [article online](#) for updates and enhancements.

You may also like

- [The Aemulus Project. III. Emulation of the Galaxy Correlation Function](#)

Zhongxu Zhai, Jeremy L. Tinker, Matthew R. Becker et al.

- [Revisiting Vainshtein screening for fast \$N\$ -body simulations](#)

Guilherme Brando, Kazuya Koyama and Hans A. Winther

- [The Mira-Titan Universe. III. Emulation of the Halo Mass Function](#)

Sebastian Bocquet, Katrin Heitmann, Salman Habib et al.

Aemulus ν : precise predictions for matter and biased tracer power spectra in the presence of neutrinos

**Joseph DeRose,^a Nickolas Kokron,^{b,c} Arka Banerjee,^d
Shi-Fan Chen,^e Martin White,^{a,f,g} Risa Wechsler,^{b,c}
Kate Storey-Fisher,^h Jeremy Tinker^h and Zhongxu Zhai^{i,j,k,l}**

^aPhysics Division, Lawrence Berkeley National Laboratory,
1 Cyclotron Rd, Berkeley, CA, U.S.A.

^bDepartment of Physics, Stanford University,
382 Via Pueblo Mall, Stanford, CA 94305, U.S.A.

^cKavli Institute for Particle Astrophysics and Cosmology,
SLAC National Accelerator Laboratory,
2575 Sand Hill Road, Menlo Park, CA 94025, U.S.A.

^dDepartment of Physics, Indian Institute of Science Education and Research,
Homi Bhabha Road, Pashan, Pune 411008, India

^eInstitute for Advanced Study,
1 Einstein Drive, Princeton, NJ 08540, U.S.A.

^fBerkeley Center for Cosmological Physics, Department of Physics, U.C. Berkeley,
Campbell Hall 341, CA 94720, U.S.A.

^gDepartment of Physics, University of California,
366 Physics North MC 7300, Berkeley, CA 94720, U.S.A.

^hCenter for Cosmology and Particle Physics, Department of Physics, New York University,
726 Broadway, New York, NY 10003, U.S.A.

ⁱDepartment of Astronomy, School of Physics and Astronomy,
Shanghai Jiao Tong University,
Shanghai 200240, China

^jShanghai Key Laboratory for Particle Physics and Cosmology,
Shanghai 200240, China

^kWaterloo Center for Astrophysics, University of Waterloo,
Waterloo, ON N2L 3G1, Canada

^lDepartment of Physics and Astronomy, University of Waterloo,
Waterloo, ON N2L 3G1, Canada
E-mail: jderose@lbl.gov, kokron@stanford.edu, arka@iiserpune.ac.in,
sfschen@ias.edu, mwhite@berkeley.edu, rwechsler@stanford.edu, k.sf@nyu.edu,
tinker@nyu.edu, zhongxuzhai@sjtu.edu.cn

Received March 28, 2023

Accepted June 22, 2023

Published July 25, 2023

Abstract. We present the **Aemulus** ν simulations: a suite of 150 $(1.05h^{-1}\text{Gpc})^3$ N -body simulations with a mass resolution of $3.51 \times 10^{10} \frac{\Omega_{cb}}{0.3} h^{-1} M_{\odot}$ in a $w\nu\text{CDM}$ cosmological parameter space. The simulations have been explicitly designed to span a broad range in σ_8 to facilitate investigations of tension between large scale structure and cosmic microwave background cosmological probes. Neutrinos are treated as a second particle species to ensure accuracy to 0.5 eV, the maximum neutrino mass that we have simulated. By employing Zel'dovich control variates, we increase the effective volume of our simulations by factors of $10 - 10^5$ depending on the statistic in question. As a first application of these simulations, we build new hybrid effective field theory and matter power spectrum surrogate models, demonstrating that they achieve $\leq 1\%$ accuracy for $k \leq 1 \text{hMpc}^{-1}$ and $0 \leq z \leq 3$, and $\leq 2\%$ accuracy for $k \leq 4 \text{hMpc}^{-1}$ for the matter power spectrum. We publicly release the trained surrogate models, and estimates of the surrogate model errors in the hope that they will be broadly applicable to a range of cosmological analyses for many years to come.

Keywords: cosmological neutrinos, cosmological parameters from LSS, cosmological simulations, dark matter simulations

ArXiv ePrint: [2303.09762](https://arxiv.org/abs/2303.09762)

Contents

1	Introduction	1
2	Parameter space design	4
3	Initial conditions and N-body solver	7
3.1	Convergence tests	10
4	Lagrangian perturbation theory and hybrid effective field theory	13
5	Zel'dovich control variates	16
6	Surrogate model construction and performance	18
6.1	Surrogate model methodology	18
6.2	Surrogate model performance	22
6.3	Error modeling	23
7	Summary	25
A	HMcode2020 surrogate model error	29

1 Introduction

After recombination and on sufficiently large scales, the dynamics of the Universe can be described by the collisionless Boltzmann equation in an expanding background. Much of the theoretical effort in cosmology over the last few decades has been devoted to increasing the accuracy, flexibility, and speed of methods for solving these equations, with techniques generally falling into two camps: perturbation theory and simulation-based methods. In this work, we combine aspects of both of these methodologies in order to make accurate non-linear predictions for real-space power spectra in cold dark matter (CDM) cosmologies, including the effects of massive neutrinos and non-cosmological constant dark energy (DE) equations of state.

Perturbative methods for predicting matter density and velocity statistics have matured over the past few decades to the point that they are now commonly used to confront contemporary observations [1–6]. Methods for incorporating the effects of dark energy [7], and massive neutrinos have been developed [8–10], although near Λ CDM cosmologies the Einstein de-Sitter approximation is typically accurate enough for current levels of observational errors [11]. While fast, flexible, and accurate over the range of scales where perturbation is valid, these methods inevitably break down in the non-linear regime of structure formation. With the incorporation of effective field theory (EFT) techniques, the reach of perturbation theory has been pushed to $k \sim 0.3 \, h\text{Mpc}^{-1}$ for real-space power spectra [12] and $k \sim 0.2 \, h\text{Mpc}^{-1}$ for redshift-space power spectra [13].

Methodologies for running N -body simulations more rapidly and in larger volumes have also seen a great deal of development in the last few decades [14–17]. These simulations use a variety of methods to solve a discretized version of the non-relativistic, collisionless Boltzmann equation. They produce fully non-linear solutions, but great care must be taken to ensure

that they are converged with respect to various choices that are made when running them, such as mass and time resolution, volume, and initial conditions.

A variety of methods for including the effects of massive neutrinos in N -body simulations have been developed. For sufficiently small neutrino masses, e.g. ~ 0.3 eV and below, treating the neutrino component linearly while solving for the full non-linear evolution of the CDM distribution is sufficient [18–22]. At larger masses, a non-linear treatment of the neutrino component is necessary to achieve sub-percent accuracy above $k \sim 1 h\text{Mpc}^{-1}$ [22–26]. Including neutrinos as a separate particle species in N -body simulations has been shown to be an accurate route for such a treatment [27, 28]. While the most widely used implementation of this method incurs significant biases in the neutrino distribution itself due to the impact of shot noise at early times when the neutrino auto-power spectrum is small, this complication is insignificant, as all currently relevant observables either depend on the total matter field, or the CDM and baryon fields, which are accurately recovered, albeit with slightly increased noise, using such a technique for realistic neutrino masses.

Although N -body simulations are able to accurately solve for non-linear CDM and neutrino dynamics, they are fundamentally limited in the scales that they can describe, as neglected processes such as radiative cooling, star formation and subsequent supernovae, as well as Mpc-scale outflows from supermassive black holes become non-negligible on the scales of galaxies and galaxy clusters. Furthermore, because N -body simulations do not include the actual objects that are observed in contemporary surveys, i.e. galaxies and other luminous tracers of the matter distribution, there will always be uncertainty associated with the connection between N -body simulations and galaxy surveys that must be appropriately accounted for in order to use these simulations to confront observations.

Traditional methods for connecting N -body simulations to galaxy survey observations involve statistical models for populating galaxies in simulated dark matter halos. Most common among these techniques is the halo occupation distribution (HOD) formalism [29–31], which assumes functional forms for the distribution of the number of galaxies that occupy halos of a given mass, as well as the phase-space distribution of those galaxies within a halo. The end product of this technique is a catalog of simulated galaxies, from which various summary statistics can be computed and compared to observational data. N -body simulations paired with the HOD formalism have recently been used to extract cosmological constraints from a range of data sets, including redshift-space clustering [32–35], galaxy-galaxy lensing [36, 37], and higher-order statistics of the galaxy field [38, 39].

There are two major limitations of the HOD formalism, which also apply to many other methods for statistically populating simulated dark matter halos with galaxies. The first is that such methods are significantly restricted by their reliance on halo-finding algorithms. In order to model galaxy samples that populate low mass halos, the simulations must resolve these halos, placing stringent requirements on the resolution of the simulations and thus increasing the expense of running these simulations. Historically, this restriction has been one of the main inhibitors in using simulations to analyze large scale structure data. Relatedly, the halo definition employed in data analysis matters (e.g. [40]). Different choices of halo definition can lead to very different resolution requirements [41, 42], and different conclusions about the necessity of parameters beyond halo mass in the HOD parameterization [43, 44].

The second limitation is more philosophical, namely that the functional form parameterizations used in HOD models draw motivation from hydrodynamical simulations. These simulations are able to model the formation of galaxies, and thus directly measure the relevant quantities required for HOD models. Although such simulations have made significant

progress in terms of reproducing observations over the last decade [45–48], they still rely heavily on sub-grid physics models with tens of parameters that are hand tuned to match observations. Thus the extent to which these simulations are predictive is limited, and the accuracy of their predictions seldom meets the stringent requirements of cosmological parameter estimation. Furthermore, because hydrodynamical simulations are relied on to help determine key ingredients of HOD models, there is no controlled series of HOD terms that one can systematically check sensitivity to when performing data analysis. Nevertheless, if these uncertainties can be satisfactorily controlled, then HOD and related methods paired with high resolution N -body simulations have the potential to extract information from significantly smaller scales than perturbation theory.

On the other hand, the bias expansions used to connect matter statistics to observed tracer statistics in perturbation theory are well controlled: at a given order there are a finite number of terms that obey the symmetries of the problem in question, and analyses can be systematically checked for sensitivity to these terms. Furthermore, these expansions are flexible enough to model any biased tracer. Because of these advantages, significant effort has been invested in the last few years in using the principles of perturbative bias expansions to connect N -body simulations to observables, a methodology that has become known as hybrid effective field theory (HEFT). HEFT was first introduced in [49], where it was demonstrated that appropriately post-processed outputs of N -body simulations can replace the perturbation theory basis spectra that are used in the bias expansion. This work showed that such a technique enables one to fit real-space halo power spectra to $k = 0.6 \, h\text{Mpc}^{-1}$, a factor of three smaller in scale than what is possible with Lagrangian perturbation theory (LPT). Ref. [50] demonstrated that HEFT could also accurately model summary statistics sensitive to higher-order auto- and cross-clustering of tracers and matter [51, 52] on quasi-linear scales ($\gtrsim 15 h^{-1}\text{Mpc}$).

One sacrifice that must be made when exchanging perturbatively predicted basis spectra for their analogous simulation-based predictions is that the cosmology dependence of these spectra must then be predicted by running simulations at many different cosmologies. Running a new simulation for each cosmology sampled in a Monte Carlo Markov Chain (MCMC) analysis would be a prohibitively expensive endeavor, but significant progress has been made in the last decade on so-called “emulation” or “surrogate modeling” techniques that circumvent the need for this. These techniques interpolate between measurements made from small suites of simulations run at a few different cosmologies in order to obtain accurate predictions over the entire range of cosmologies spanned by the simulations. Constructing such surrogate models has become relatively commonplace, with simulation-based models now existing for the matter power spectrum [53–55], the halo mass function [56, 57], linear halo bias [58, 59], galaxy clustering and lensing statistics [33, 60–62], higher-order statistics [38, 39], as well as HEFT models for a number of statistics [63–67].

Another sacrifice that must be made when using simulation-based predictions is that sample variance is introduced due to the finite volumes that simulations are run in. A number of methods have been introduced to mitigate the effect of sample variance on simulated measurements. The most commonly used method is called fixed amplitude simulations, where instead of initializing the amplitude of each Fourier mode of the simulation with a Rayleigh distributed random number, the amplitude of each mode is fixed to its expectation value [68, 69]. In combination with this, a second simulation is often run where each Fourier mode is taken to be 180° out of phase with the first simulation [69, 70]. Together these methods are called “paired and fixed” simulations, and have been used to reduce the variance of measurements from a number of suites of simulations [54, 71, 72]. These methods come

with a cost though, as they require one to run twice as many simulations. They also forego Gaussian initial conditions and thus statistics measured from them must be painstakingly examined for biases [73–75]. Finally, the improvements in variance obtained from paired-and-fixed simulations degrade significantly with non-linear and higher-order statistics [74, 75].

The method of control variates [76] offers an alternative to paired and fixed simulations that does not suffer from these drawbacks. Control variates allow for the reduction in variance of a random variable in the presence of a correlated random variable with known mean. This technique was introduced to the cosmology literature under the name ‘Convergence Acceleration by Regression and Pooling’ (CARPool) [77–79] in order to reduce the variance of measurements made from N -body simulations, where the control variate was taken to be measurements from approximate simulations such as COLA [80] or FASTPM [81]. Ref. [82] and [83] demonstrated that comparable variance reduction is possible at significantly less expense by using the Zel’dovich approximation (ZA) as the control variate to reduce the variance of real and redshift-space power spectra measured from N -body simulations, respectively.

In this work, we present a new suite of 150 simulations, run in a $w\nu$ CDM cosmological parameter space, simulating neutrinos as an extra particle species to ensure accuracy of our predictions to 0.5 eV. Furthermore, we run these simulations in the broadest parameter space ever used in a single suite of simulations in order to ensure that they are accurate over the full range of $w\nu$ CDM cosmologies allowed by current data. We perform convergence tests of these simulations to ensure their accuracy and we take advantage of the Zel’dovich control variate (ZCV) method to reduce the variance of the HEFT spectra that we measure from our simulations. We then build a surrogate model for these HEFT spectra, including the matter power spectrum.

This rest of this work is organized as follows. In section 2, we describe the parameter space that the simulations are run in, focusing on how we optimized it to deliver both breadth and accurate surrogate models. In section 3 we describe the settings used for our N -body simulations, emphasizing the improved accuracy provided by initializing our simulations at relatively low redshift using 3rd-order LPT. Section 4 introduces the basics behind LPT and HEFT and section 5 describes our ZCV methodology and resultant improvements in the precision of our HEFT measurements. Section 6 describes our surrogate modeling methodology, quantifies the accuracy of the final HEFT models, and provides comparisons to previous matter power spectrum surrogate models. Finally, in section 7 we summarize our results, detail our data release plans, and discuss future directions of inquiry.

2 Parameter space design

The design of surrogate model parameter space is crucial for ensuring reliable analysis results, and it requires balancing two factors: parameter space breadth and surrogate model accuracy. Covering a broad parameter space is complicated by “tensions” that have arisen in contemporary cosmological constraints [84–87]. To avoid being biased by these tensions, simulations must span the range of parameter values that both sets of experiments prefer. To achieve this goal, the simulations’ parameter space should be as broad as possible without exceeding a fixed threshold for model accuracy. We set a goal of 1% accuracy for two main reasons. First, N -body codes agree to $\sim 0.5\%$ at $k \sim 1 \text{ hMpc}^{-1}$ for the $z = 0$ matter power spectrum [14], and so this sets a hard lower limit on how accurate a simulation based matter power spectrum model can be. Secondly, ongoing and upcoming surveys such as the Dark Energy Spectroscopic Instrument [88], Rubin Observatory [89], Simons Observatory [90] will measure angular

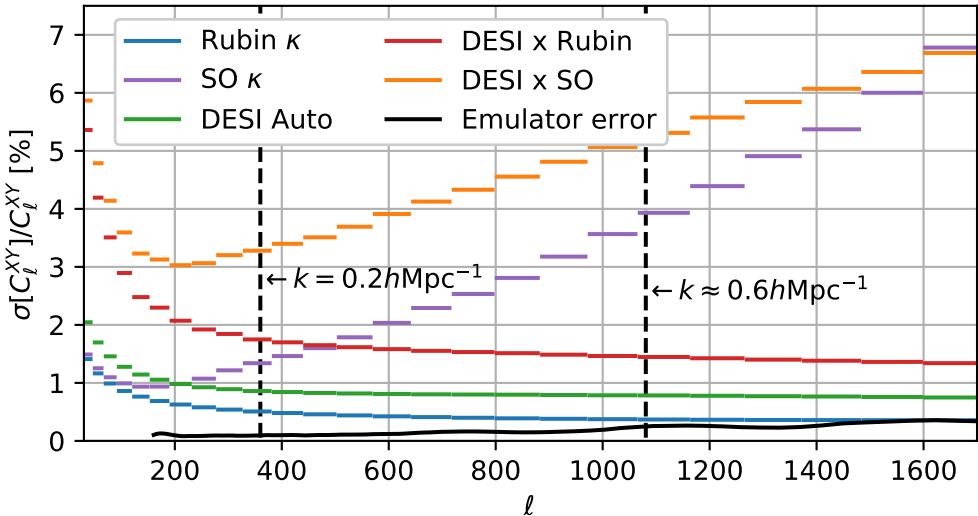


Figure 1. Forecasted fractional uncertainties on angular galaxy and CMB lensing, and galaxy clustering auto- and cross-power spectra from ongoing and upcoming surveys. The assumptions we have made in computing these are described in the main text. The dashed vertical lines represent the approximate angular scales that correspond to $k = 0.2 \text{ hMpc}^{-1}$, the k_{max} assumed in most analyses that use pure perturbation theory models, and $k = 0.6 \text{ hMpc}^{-1}$, the approximate k_{max} to which HEFT is unbiased [63], for an effective redshift of $z = 0.7$ where the DESI sample is assumed to be centered. We also plot our measured emulator error for the DESI auto spectrum at this redshift, computed as described in section 6.3, to illustrate that the accuracy we achieve in this work is significantly below the statistical errors of the measurements considered here.

weak lensing and galaxy clustering spectra at roughly 1% precision, as demonstrated in figure 1. We have made the forecasts in this figure assuming a disconnected covariance approximation, which should place a lower bound on the actual errors on these spectra. For DESI we have assumed $\bar{n} = 6 \times 10^{-4} h^3 \text{Mpc}^{-3}$ and biases as measured in [91], consistent with the DESI luminous red galaxy number density at $z \sim 0.7$ [92], in a bin of roughly constant density between $z = [0.6, 0.8]$. For Rubin, we take $\sigma_e = 0.26$ and use a number density that is consistent with a quarter of the LSST gold sample [93], $\bar{n} = 10.8 \text{ arcmin}^{-2}$, and a source distribution that matches the gold sample at $z \geq 1$ convolved with photometric redshift uncertainties. SO noise curves are taken from [94], and for all three surveys we assume $f_{\text{sky}} = 0.4$. The angular band powers shown have a width of roughly $\Delta\ell \sim 3\sqrt{\ell}$. We have also plotted the error on the galaxy angular auto-power spectrum that the model in this work achieves, computed as described in section 6.3 to illustrate that the final accuracy of our model is sufficient for upcoming data, although see discussion in that section related to error correlations between different scales. This is a conservative bound compared to the emulator error achieved on the galaxy-matter angular power spectrum or the matter-matter power spectrum for these redshift bins.

Additionally, it is important to have confidence that the chosen parameter space sampling will deliver the desired level of accuracy before running any simulations. To address this concern, we produce mock matter power spectrum measurements using `HMcode2020` [95], assuming a fixed number of simulations, while varying the bounds of the parameter space. We then build surrogate models for these spectra and test their accuracy against a dense sampling of `HMcode2020` predictions that were not included in the surrogate model training.

Here we aim to span a range of parameters in the $w\nu$ CDM cosmological model, including the matter density ω_m , the baryon density ω_b , the dark energy equation of state parameter w , the Hubble parameter H_0 , the scalar spectral index n_s , the normalization of the power spectrum A_s , and the sum of neutrino masses $\sum m_\nu$. The parameters ω_b and n_s are tightly constrained by CMB data, and their impact on LSS observables is minor [54, 58, 88]. We set their bounds to $0.0173 \leq \omega_b \leq 0.0272$, and $0.93 \leq n_s \leq 1.01$, broader than is allowed by current CMB constraints; the small impact these parameters have on LSS observables means that our surrogate model errors are not significantly affected by this breadth. We set broader limits on the other parameters, as they have a more significant impact on LSS observables. We use a prior on the logarithm of the sum of the neutrino masses $\sum m_\nu$, using a range that is constrained by combinations of CMB and LSS data [88, 96] and by laboratory-based experiments [97]: $0.01, \text{eV} \leq \sum m_\nu \leq 0.5, \text{eV}$. The lower edge of the range is set below the current minimum allowed value of 0.06eV [98, 99] to avoid modeling errors at the minimum allowed mass.

We also investigate the impact of varying the widths of our parameter space in $\log 10^{10} A_s$, w , ω_c and H_0 on the accuracy of the resulting surrogate models. In particular, we generate experimental designs in a four-dimensional grid, such that the bounds of each design are

$$\begin{aligned} 10^9 A_s &\in [2.1 - 0.33i, 2.1 + 0.33i] \\ w &\in [-1 - 0.28j, -1 + 0.28j] \\ \omega_c &\in [0.12 - 0.02k, 0.12 + 0.02k] \\ H_0 &\in [67 - 7.5l, 67 + 7.5l], \end{aligned}$$

and $i, j, k, l \in \{1, 2, 3\}$.

We sample each of this set of 81 parameter bounds with 100 points in a Latin hypercube, maximizing the minimum distance between pairs of points in two dimensions. At each point in parameter space, we produce non-linear matter power spectrum predictions using `HMCode2020` [95], $P_{m,m}^{\text{HM}}(k, z)$, and 1-loop LPT predictions of the matter power spectrum $P_{m,m}^{\text{1-loop}}(k, z)$, as will be described in more detail in section 4. We evaluate these models at 100 points, logarithmically spaced between $k = 10^{-1} h\text{Mpc}^{-1}$ and $k = 1 h\text{Mpc}^{-1}$, at the same redshifts that we output snapshots at, described in section 3.

We build surrogate models for the logarithm of the ratio of these predictions:

$$\Gamma = \log_{10}(P_{m,m}^{\text{HM}}(k, z)/P_{m,m}^{\text{1-loop}}(k, z)), \quad (2.1)$$

reproducing the surrogate modeling methodology in [63] using a combination of principal component analysis (PCA) and polynomial chaos expansions (PCE). We performed a hyper-parameter optimization, using the same methodology as described in section 6, on the design with $i, j, k, l = 2$. We have tested sensitivity to which design this optimization is performed on and found negligible impact to our conclusions. One notable difference to the procedure described in section 6, is that here we have used redshift as our time variable, rather than using $\sigma_8(z)$ as we do in section 6. This is because before running our simulations, we did not consider this option and our design choices were made using redshift as a time variable, so we have not altered this after the fact.

In order to test the surrogate models trained on these 81 parameter spaces, we produce a test set of 10,000 models over the broadest parameter space considered here, i.e. $i, j, k, l = 3$.

	Tier 1 min.	Tier 1 max.	Tier 2 min.	Tier 2 max.
$10^9 A_s$	1.10	3.10	1.77	2.43
n_s	0.93	1.01	0.93	1.01
H_0	52.0	82.0	59.5	74.5
w	-1.56	-0.44	-1.28	-0.72
ω_b	0.0173	0.0272	0.0198	0.0248
ω_c	0.08	0.16	0.11	0.13
$\sum m_\nu$ (eV)	0.01	0.50	0.01	0.50

Table 1. *Aemulus* ν parameter space boundaries.

When measuring errors for each choice of parameter limits, we require that the test points lie within the hypercube defined by the minimum and maximum value of each parameter in the design under consideration, such that the number of test points varies from design to design.

In figure 2, we show the 68th percentile error for a sub-selection of parameter space designs. It is clear that as we expand the boundaries of our parameter space, the surrogate model errors become larger. The main notable feature is that surrogate model performance is very sensitive to the chosen bounds on the dark energy equation of state parameter, w . Only designs with $j = 1$ achieved our goal of sub-percent 68th percentile error over the entire range of scales considered. Given the current constraints on w from individual LSS probes, we determined that this choice would be too restrictive.

Because of this, we chose employ a two-tiered parameter space design, with the first tier of 100 simulations, sampled using a Latin hypercube, spanning as broad a parameter space as possible while not exceeding 2% 68th percentile error, and a more restricted second tier of 50 simulations, sampled with a Sobol sequence [100], where the 68th percentile error is less than 1%. We have used a Sobol sequence for this second tier so that it is straightforward to add additional simulations in the future by using the next element of the Sobol sequence. Figure 3 shows the parameter space sampling resulting from this optimization compared with that used in *Aemulus* α [101], as well as the cosmological constraints presented in [87]. Table 1 lists the boundaries for the Tier 1 and Tier 2 parameter spaces. Note that these are the bounds of the Latin hypercube and Sobol sequence respectively, and not the actual minimum and maximum values of simulated cosmologies, although the difference between these is negligible. In section 6 and appendix A we demonstrate that the accuracy that we achieve with our N -body based surrogate models is consistent with our forecast error, and explore the HMCode2020 surrogate model error as a function of cosmological parameters for the final design used for the *Aemulus* ν simulations.

3 Initial conditions and N -body solver

Appropriate initialization of simulations is as important for obtaining converged N -body predictions as any setting in the N -body solver itself. In particular, it has been shown that early initialization of simulations with low order LPT can lead to appreciable transient errors in force calculations [102, 103]. These transients are sourced by sparse sampling of modes around the Nyquist frequency, k_{nyq} , of the initial particle grid, leading to an incorrect growing mode in the simulations until gravitational collapse leads to denser sampling on these scales [104]. The impact of this effect becomes more appreciable for higher-order statistics,

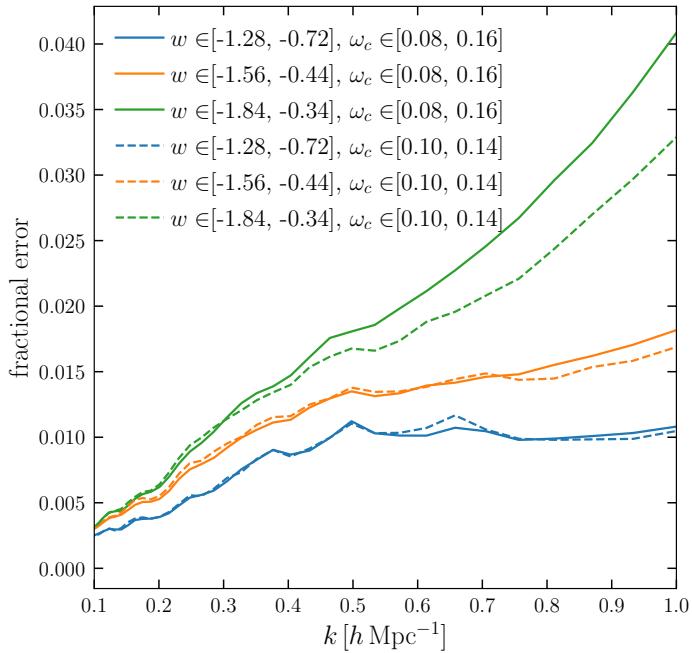


Figure 2. 68th percentile fractional error for `HMCode2020` surrogate models as a function of k for a sub-selection of the 81 parameter space bounds considered. The main feature of note is that the surrogate model error depends significantly on the range of dark energy equation of state parameters that we include in the training domain. Because of this, we have opted to use two tiers of simulations, in order to allow for broad exploration of DE parameter space, while still maintaining very high accuracy near Λ CDM.

where anisotropy in the particle distribution due to the imprint of the grid causes an even larger impact. Ref. [103] showed that this transient effect can be alleviated by starting simulations at significantly later cosmic times than previously used, enabled by using third-order LPT (3LPT) to initialize the particle distributions, as implemented in the `monofonic` code. Ref. [105] then extended `monofonic` to treat massive neutrinos by implementing an approximate three-fluid 3LPT. In this work, we make use of this extended version of `monofonic` to generate initial CDM and neutrino particle distributions at $z = 12$ using 3LPT.

The presence of massive neutrinos imparts a scale-dependent growth, $D(k, z)$, to the matter distribution that disallows the typical practice of back-scaling a $z = 0$ linear matter power spectrum using a scale-independent growth factor to initialize simulations. Instead, we compute the linear CDM and baryon (cb) power spectrum, $P_{cb, \text{lin}}(k, z = 0)$, using `CLASS` [106], and back-scale to our starting redshift z_{ini} using a scale dependent growth factor, $D_{cb}(k, z_{\text{ini}})$, that is computed using a first-order Newtonian fluid approximation as implemented in `zwindstroom` [105]. Conveniently, this also accounts for the lack of a radiation component in our simulations, as well as the fact that Newtonian mechanics breaks down at the highest redshifts and very largest scales that we simulate [107].

The initial cb power spectrum is then given by

$$P_{cb}(k, z_{\text{ini}}) = D_{cb}(k, z_{\text{ini}})^2 P_{cb, \text{lin}}(k, z = 0). \quad (3.1)$$

This is the quantity that is used to generate our initial Gaussian density field, from which we compute displacements in order to initialize the cb particle distribution with 3LPT as

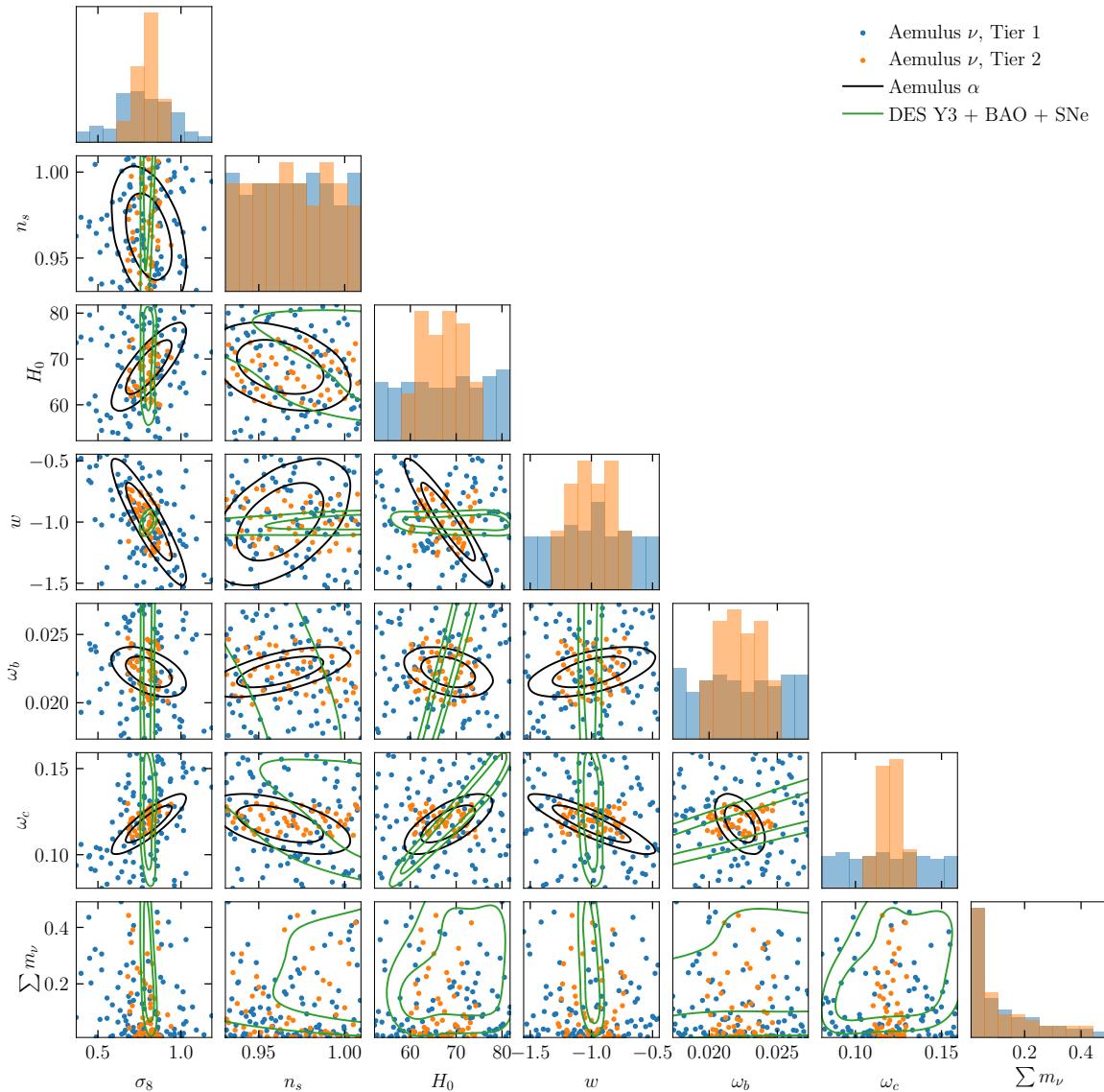


Figure 3. Comparison of the **Aemulus** ν parameter space sampling to that used in **Aemulus** α . Blue and orange points represent our tier 1 and 2 simulations respectively, while black points are cosmologies used for the **Aemulus** α simulations. For reference, we also plot constraints from the combination of DES Y3 weak lensing and galaxy clustering combined with BAO and type Ia supernovae [87, 96] (green), as well as Planck 2018 constraints combined with the same BAO and type Ia supernovae data (red) [88, 96].

implemented by `monofonic`. This re-scaling only works exactly at linear order at $z = 0$, but it has been shown to achieve $\sim 0.1\%$ accuracy at redshifts relevant for LSS studies [107]. As discussed above, using 3LPT allows us to initialize our simulations significantly later than would otherwise be possible with lower-order LPT models. We note that beyond-linear LPT treatment of neutrinos is not particularly important, as the neutrino overdensities remain small to very low redshifts due to the effect of free streaming. Using 3LPT for δ_{cb} is, however, essential in order to start at $z_{\text{ini}} = 12$ [103].

N_{sim}	L_{box} [$h^{-1}\text{Mpc}$]	N_{cb}	N_{ν}	$m_{\text{part},cb}$ [$h^{-1}\text{M}_{\odot}$]	ϵ [$h^{-1}\text{kpc}$]	$\Delta \ln a_{\text{max}}$	z_{ini}
150	1050	1400^3	1400^3	$3.30 \times 10^{10} \left(\frac{\Omega_{cb}}{0.3}\right)$	20	0.01	12

Table 2. Summary of the settings used to run the `Aemulus` ν simulations. Columns are the total number of simulations, simulation side length, number of cb and neutrino particles, cb particle mass, Plummer equivalent force softening, maximum time step, and starting redshift.

We generate neutrino particle initial conditions using `fastDF` [108], which produces initial particle displacements and velocities at $z = 12$ by integrating neutrino particles along geodesics starting from $z = 10^9$ using the linear metric perturbations output from `CLASS`. We set the `fastDF` time stepping parameter to $\Delta \log a = 0.05$, with a mesh size of $M = 384$, as these settings are shown to lead to converged results in [108]. We assume that all three neutrinos have equal masses, usually referred to as the degenerate mass approximation. This has been shown to be a good approximation to both inverted and normal hierarchy scenarios, and is significantly more accurate than assuming one massive and two massless neutrinos [27].

We make use of modified version of `Gadget-3` in order to evolve our cb and neutrino particle distributions from $z_{\text{ini}} = 12$ to $z = 0$. We use 1400^3 cb and neutrino particles respectively with a box size of $L_{\text{box}} = 1050 h^{-1}\text{Mpc}$, yielding a cb particle mass of $3.51 \times 10^{10} \frac{\Omega_{cb}}{0.3} h^{-1}\text{M}_{\odot}$. We use a Plummer-equivalent force softening of $\epsilon_{\text{plummer}} = 20 h^{-1}\text{kpc}$, a mesh size of $N = 2100$ for large-scale force computations and a maximum time step of $\max[\Delta \ln a] = 0.01$. In order to facilitate arbitrary background evolution models, we have modified `Gadget-3` to read in a tabulated $H(z)$ as output by `zwindstrom`, which accounts for relativistic corrections to $H(z)$ and allows us to simulate non-cosmological constant dark energy models. We have also modified `Gadget-3` to only compute forces from neutrinos at the particle mesh level. This is an extremely accurate approximation, as neutrinos do not cluster significantly on scales smaller than the mesh resolution $L_{\text{box}}/N = 0.5 h^{-1}\text{Mpc}$, as visually illustrated in 4. This allows us to run simulations with neutrino particles that take only approximately 10% longer than the same simulation without neutrinos. These settings are summarized in table 2.

We output fixed time snapshots at 30 epochs logarithmically spaced in scale factor between $z = 3$ and $z = 0$. Halo finding is performed on each snapshot with `Rockstar`, using strict spherical overdensity masses with $\Delta = 200_m$, where strict refers to the inclusion of unbound particles in the mass estimates of halos. Halos finding is performed with only cb particles. All results relating to halos in this work use only host halos, i.e. those halos that are not enclosed by halos with a higher maximum circular velocity.

3.1 Convergence tests

We made two changes to the accuracy settings of the `Aemulus` ν simulations compared to those used in [101]. First, we decreased the maximum allowed time step to $\max[\Delta \ln a] = 0.01$ from $\max[\Delta \ln a] = 0.025$. This change was made to ensure accurate recovery of linear growth on large scales, which was only achieved at $\sim 1\%$ accuracy in the `Aemulus` α simulations. We did not perform additional convergence tests of this change, because it is more conservative than our previous choice. However, we note that this change is one of the reasons that we are able to recover linear growth accurately, as discussed in section 6.

Second, we lowered the starting redshift of our simulations from $z_{\text{ini}} = 49$ to $z_{\text{ini}} = 12$. While [103] tested the reliability of this choice, their simulations used slightly different settings, including a larger particle mass, than those used here. The accuracy of a particular starting time and LPT order depends on the resolution of the simulation: higher resolution

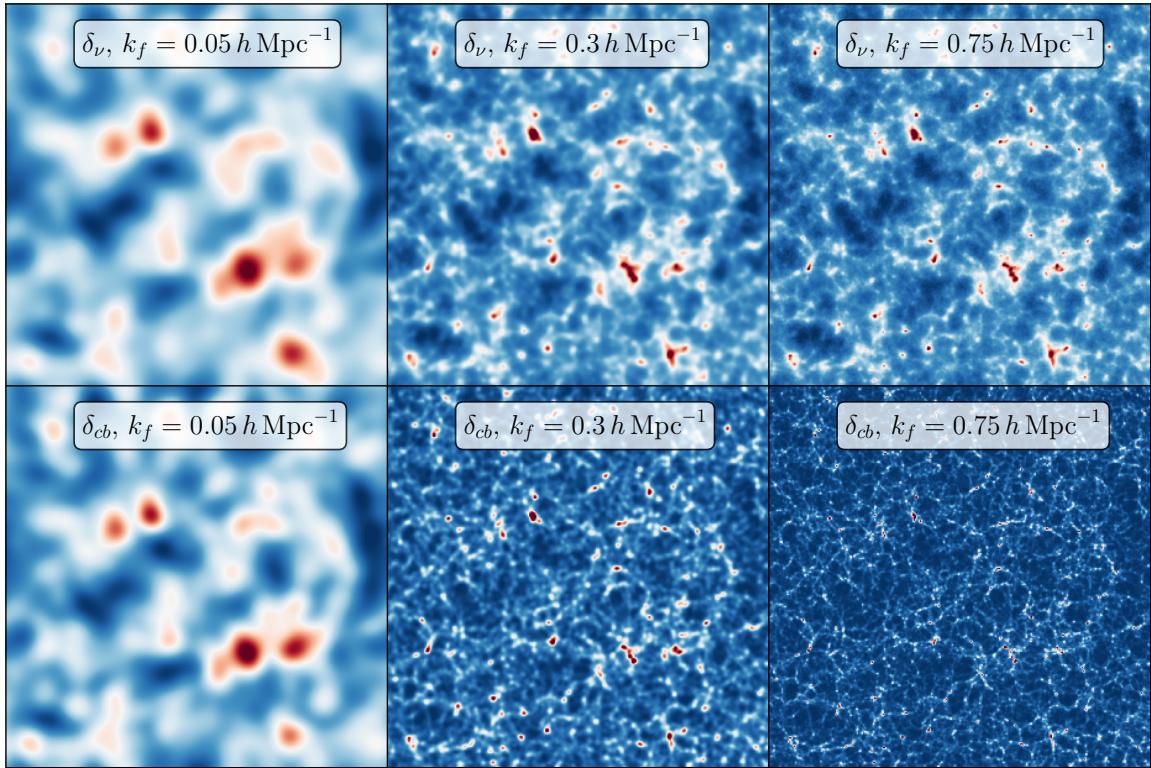


Figure 4. The top row shows the neutrino overdensity field, $\delta_\nu(z = 0)$, smoothed with a Gaussian filter of width $k_f = 0.05 h \text{Mpc}^{-1}$ (left), $k_f = 0.3 h \text{Mpc}^{-1}$ (middle), and $k_f = 0.75 h \text{Mpc}^{-1}$ (right) for the simulation in our suite with the largest neutrino mass. The bottom row shows the same but for the CDM field, i.e. $\delta_{cb}(z = 0)$. For the largest smoothing scale, which is larger than the free streaming length of the neutrinos, the cb and neutrino fields are almost identical. For progressively smaller smoothing scales, the cb field continuously exhibits more structure, while the neutrino field remains very similar between the middle and right most columns. This visually illustrates that the neutrino component does not cluster on scales close to the grid scale.

simulations require earlier starting times at fixed LPT order because they resolve smaller scales that become non-linear earlier. Therefore, we need to make sure that the findings in [103] hold for the exact settings we used for `Aemulus ν`. We will now present a series of tests to show that starting at $z_{\text{ini}} = 12$ produces more converged results compared to starting our simulations at a higher redshift.

To perform our convergence tests, we have run three simulations, all at the Planck 2018 best fit Λ CDM cosmology [88]. The first simulation, which we call T0, uses identical settings to our fiducial simulations, but is run in a reduced volume of $(525 h^{-1} \text{Mpc})^3$, evolving 700^3 cb particles and 700^3 neutrino particles, using a particle mesh size of 1050^3 . The second simulation, which we call T1, is identical to T0, except we use $z_{\text{ini}} = 24$. The final simulation, T2, is identical to T1, but evolves 4×700^3 cb particles and 4×700^3 neutrino particles, sampling the exact same modes as T0, using the “face-centered-cubic” lattice mode in `monofonic`. Because particle discreteness effects decrease with the number of particles used in the simulation, we can start the T2 simulation earlier and thus it provides a test of whether 3LPT is sufficiently accurate at $z = 12$. We have run these simulations in a reduced volume because we only wish to compare relative differences between them, and so can use the same initial seed in order to remove sample variance from our comparisons.

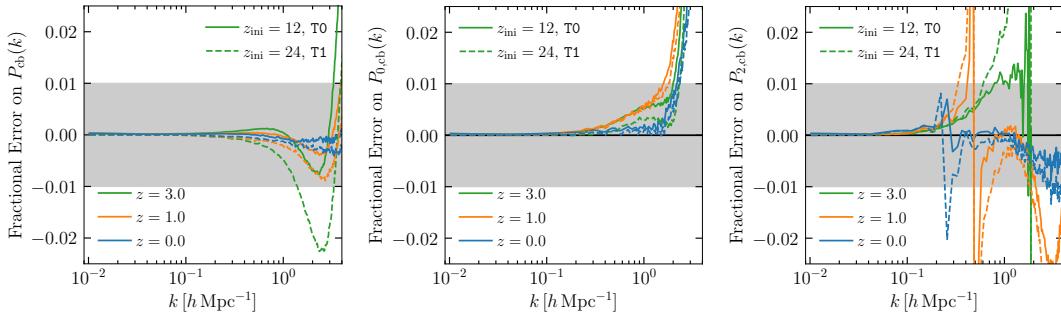


Figure 5. Comparison of real-space power spectra (left), monopole (middle) and quadrupole (right) of redshift space cb power spectra between the T0 simulation started at $z = 12$ (solid) and the T1 simulation started at $z = 24$ (dashed) to the T2 simulation initialized at $z = 24$ with four times the number of particles, but sampling the exact same modes. Initialization at $z = 12$ with 3LPT yields the best results compared to this higher resolution simulation.

Figures 5 and 6 show comparisons of these three test simulations, where all figures show fractional errors comparing T0 and T1 to T2. Figure 5 shows comparisons of real- and redshift-space cb power spectra between the three simulations. The left hand panel depicts the fractional error on the real-space cb power spectrum as a function of redshift. We see that starting at $z_{ini} = 24$ leads to significantly larger errors at fixed k compared to $z_{ini} = 12$. This effect is amplified at high redshift, mostly due to the fact that the amplitude of the power spectrum is lower, so a similar absolute error translates into a larger fractional error. The redshift-space monopole, $P_{0,cb}$ does not exhibit the same trends, likely because the finger-of-god effect has washed out these subtle issues on small scales. The redshift-space quadrupole, $P_{2,cb}$, is recovered slightly better in T0 than T1, but the effect is again difficult to interpret due to the large role that virial velocities play at high- k in redshift-space statistics. Nevertheless, we see that our fiducial settings are converged at the $\leq 1\%$ level to $k \sim 1 \text{ hMpc}^{-1}$ compared to the higher resolution T2 simulation.

Figure 6 shows similar comparisons to figure 5, but focuses on halo statistics. The left side shows fractional errors on the spherical overdensity halo mass function for the T0 and T1 simulations, again compared to the higher resolution T2. Error bars are estimated via jackknife resampling, using 128 jackknife regions. We see that for $z = 1$ and $z = 0$ both simulations are converged to at the $\leq 1\%$ level until $\sim 10^{13} h^{-1} M_\odot$, after which the T1 simulation begins to diverge. The T0 simulation remains converged at the $\leq 1\%$ until $\sim 2 \times 10^{12} h^{-1} M_\odot$. Compared to our findings in the **Aemulus** α suite of simulations, starting our simulations at $z_{ini} = 12$ yields halo masses that are converged at a factor of two lower in mass, with the same mass resolution and force softening. At $z = 2$, both T0 and T1 are converged at the $\sim 2\%$ level until $\sim 4 \times 10^{12} h^{-1} M_\odot$. At $z = 3$ both the T0 and T1 simulations diverge significantly from the T2 simulation at all masses measureable in these simulations and so we do not plot statistics for this redshift.

The right hand side of figure 6 shows halo redshift-space monopole measurements for the same redshift outputs, where we use halos in a bin of mass from $12.5 \leq \lg M \leq 13$. We observe a similar trend in convergence as for the mass function, where for $z < 2$ both simulations are converged with respect to the higher resolution simulation. At high k , the T1 simulation deviates by $\sim 2\%$, although when we select samples by cumulative abundance rather than mass this discrepancy disappears, suggesting that it is simply a difference in

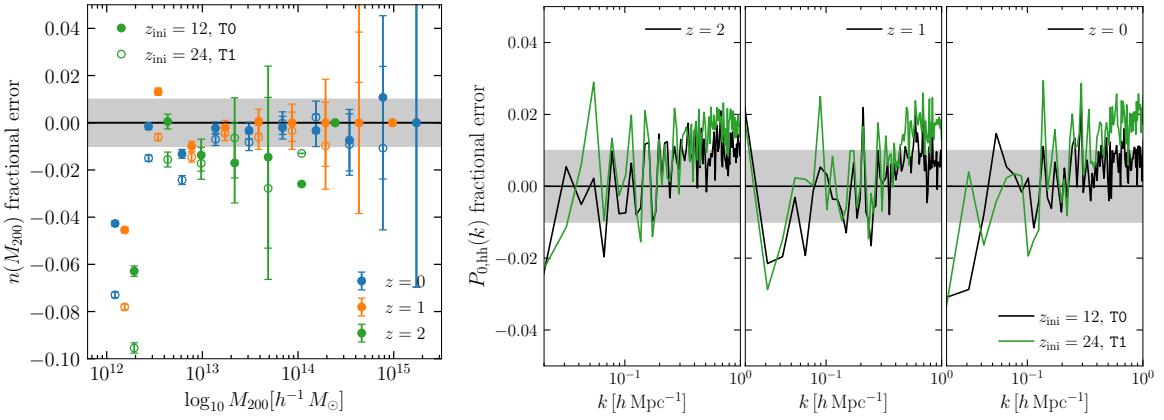


Figure 6. (Left) Same as figure 5, but comparing halo mass functions. Again we see that the T0 simulation started at $z = 12$ slightly outperforms the T1 $z = 24$ simulation in comparison to the higher resolution T2 reference simulation started at $z = 24$. The only exception to this is for $z = 2$ where T0 and T1 perform comparably. At $z = 3$ the mass functions are not converged at the $\geq 5\%$ level for any masses measurable in these simulations, and so we do not plot them and caution against the use of the $z = 3$ outputs from these simulations in general. At the two lower redshifts shown, the error on the halo mass function in the $z_{\text{ini}} = 12$ simulation remains below 1% until $\sim 2 \times 10^{12} h^{-1} M_\odot$, whereas the later start deviates by more than a percent by $\sim 5 \times 10^{12} h^{-1} M_\odot$. (Right) Same as figure 5, but comparing the monopole of the halo redshift-space power spectrum, measured in a halo mass bin of $10^{12.5} h^{-1} M_\odot \leq M_{200b} < 10^{13} h^{-1} M_\odot$. Again we see that at $z \leq 2$ the agreement between the T0 and T2 simulations is at the $\leq 1\%$ level. The T1 simulation and T2 agree until $k \sim 0.5 h\text{Mpc}^{-1}$, after which they begin to disagree by more than one percent. A similar trend is seen for the real-space halo power spectrum as well, although we do not display it here. At $z = 3$, the halo (real and redshift space) power spectra for both T0 and T1 simulations disagree with the T2 by $\sim 2\%$ at all k .

shot noise due to the slightly different number densities of the mass selected samples in this simulation. The $z = 3$ outputs from T0 and T1 exhibit a significant deviation from T2 and so we do not plot them. Higher multipoles are too noisy to be of use in these comparisons. In general this shows that masses and two-point clustering statistics of halos with $\lg M > 12.5$ and $z < 2$ are converged at the $\sim 1 - 2\%$ level, and halos from the $z = 3$ outputs should not be used in these simulations.

4 Lagrangian perturbation theory and hybrid effective field theory

The utility of LPT extends beyond the simple use case of initializing N -body simulations. In particular, it has proven to be a highly effective model for the density fields of biased tracers in CDM cosmologies [3, 109]. In this work, we make use of LPT in two ways. First, we use the ZA as a control variate to reduce the cosmic variance of our simulated measurements. We also construct a surrogate model for HEFT measurements from our simulations, which can be seen as a non-perturbative extension of LPT. In this section, we introduce some basic LPT notation to clarify our presentation of these two aspects later in this work.

In LPT, we can express the combined CDM and baryon density field as

$$1 + \delta_{cb}(\mathbf{x}, a) = \int d^3 \mathbf{q} \delta^D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q}, a)), \quad (4.1)$$

where

$$\mathbf{x} = \mathbf{q} + \Psi(\mathbf{q}, a), \quad (4.2)$$

and $\Psi(\mathbf{q}, a)$ is computed perturbatively, such that

$$\Psi(\mathbf{q}, a) = \sum_{n=1}^{\infty} \Psi^{(n)}(\mathbf{q}, a). \quad (4.3)$$

Truncating this sum at $n = 1$ yields the well known ZA, with

$$\Psi^{(1)}(\mathbf{q}, a) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{q}} \frac{i\mathbf{k}}{k^2} \delta_{cb,L}(\mathbf{k}, a) \quad (4.4)$$

$$= \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{q}} \frac{i\mathbf{k}}{k^2} D(k, a) \delta_{cb,L}(\mathbf{k}, 1), \quad (4.5)$$

where $\delta_{cb,L}(\mathbf{k}, a)$ is the Fourier transform of the linear density field $\delta_{cb}(\mathbf{q}, a)$, and $D(a, k)$ is the linear growth factor, which is scale dependent in the presence of massive neutrinos, normalized to 1 at $a = 1$. Higher-order displacement terms are non-trivial to compute exactly in LPT in the presence of dark energy and massive neutrinos, but computing these terms with kernels derived for Einstein-de Sitter (EdS) cosmologies is quite accurate for cosmologies close to Λ CDM [8, 9, 110–112]. In this work, we use **ZeNBu**¹ and **velocileptors**² to make ZA and higher-order LPT predictions, respectively, following the neutrino approximations described in appendix A of [85].

It has been shown that galaxies and halos are best treated as biased tracers of the δ_{cb} field rather than total matter field, δ_m [20, 28]. In this case, in LPT we can express the density field of a biased tracer, $\delta_t(k)$, as

$$1 + \delta_t(\mathbf{x}, a) = \int d^3\mathbf{q} F[\delta_{cb,L}(\mathbf{q})] \delta^D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q}, a)), \quad (4.6)$$

where $F[\delta_{cb,L}(\mathbf{q})]$ is a functional of the linear field, $\delta_{cb,L}(\mathbf{q})$, specifying the relationship between the tracer density and matter field at early times. In this work, we will consider an expansion of $F[\delta_{cb,L}(\mathbf{q})]$ to second order:

$$\begin{aligned} F[\delta(\mathbf{q})] \approx 1 &+ b_1 \delta(\mathbf{q}) + b_2 (\delta^2(\mathbf{q}) - \langle \delta^2 \rangle) \\ &+ b_s (s^2(\mathbf{q}) - \langle s^2 \rangle) + b_{\nabla^2} \nabla^2 \delta(\mathbf{q}) + \dots \end{aligned} \quad (4.7)$$

where $s^2(\mathbf{q}) = s_{ij}(\mathbf{q}) s^{ij}(\mathbf{q})$ and

$$s_{ij}(\mathbf{q}) = \left(\frac{\partial_i \partial_j}{\partial^2} - \frac{\delta_{ij}}{3} \right) \delta(\mathbf{q}), \quad (4.8)$$

where we have dropped the cb subscripts for brevity since there is no ambiguity regarding whether we are dealing with the cb or total matter field when modeling biased tracers. The bias coefficients in front of each term serve to parameterize our ignorance of the galaxy formation physics that sets the dependence of the tracer density field on these linear fields.

¹<https://github.com/sfschen/ZeNBu>.

²<https://github.com/sfschen/velocileptors>.

We can rewrite eq. (4.6) to explicitly emphasize that $\delta_t(\mathbf{x}, a)$ is given by a sum over advected linear fields

$$1 + \delta_t(\mathbf{x}, a) = \sum_{\substack{\{\mathcal{O}_i \in cb, \delta, \\ \delta^2, s^2, \dots\}}} b_{\mathcal{O}_i} \mathcal{O}_i(\mathbf{x}, a), \quad (4.9)$$

where the advected operators, \mathcal{O}_i , are given in both configuration and Fourier space by

$$\begin{aligned} \mathcal{O}_i(\mathbf{x}, a) &= \int d^3\mathbf{q} \mathcal{O}_i(\mathbf{q}) \delta_D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q}, a)) \\ \mathcal{O}_i(\mathbf{k}, a) &= \int d^3\mathbf{q} e^{-i\mathbf{k}\cdot(\mathbf{q} + \Psi(\mathbf{q}))} \mathcal{O}_i(\mathbf{q}), \end{aligned} \quad (4.10)$$

and setting $b_{cb} = 1$ for notational convenience. With this notation in hand, we can express cross-spectrum of two biased tracer fields, δ_a and δ_b , as

$$P_{ab}(k) = \sum_{\mathcal{O}_i, \mathcal{O}_j} b_{\mathcal{O}_i}^a b_{\mathcal{O}_j}^b P_{ij}(k) \quad (4.11)$$

where we have defined the *basis spectra*

$$P_{ij}(k)(2\pi)^3 \delta_D(k + k') = \langle \mathcal{O}_i(\mathbf{k}) \mathcal{O}_j(\mathbf{k}') \rangle, \quad (4.12)$$

and the bias coefficients for the two tracer fields are independent of each other. As gravitational lensing is sensitive to the total matter field, δ_m , we will also be interested in its auto-power spectrum, $P_{m,m}(k)$, and cross-power spectrum with the δ_{cb} and biased tracer fields:

$$P_{m,a}(k) = \sum_{\mathcal{O}_i} b_{\mathcal{O}_i}^a P_{m,i}(k) \quad (4.13)$$

where

$$P_{m,i}(k)(2\pi)^3 \delta_D(k + k') = \langle \delta_m(\mathbf{k}) \mathcal{O}_i(\mathbf{k}') \rangle. \quad (4.14)$$

Note that none of the expressions we have written for biased tracer spectra depend on how we have computed the displacements, Ψ . In real space for sufficiently low-bias tracers, where eq. (4.8) holds, it is the perturbative calculation of Ψ in LPT that limits the range of scales that can be modeled. On the other hand, N -body simulations solve discretized versions of the same equations of motion that form the basis of LPT. Furthermore, all of the ingredients that are used in the above expressions can be directly computed from N -body simulations: the displacements are simply the difference between each particle's position and the grid point it began at, and the linear fields used in eq. (4.8) can be directly computed from the Gaussian initial conditions used to seed the simulation. Thus, we can do away with perturbative computations of Ψ and the EdS approximation, and use N -body simulations to compute the basis spectra, P_{ij} . This model has become known as HEFT.

More explicitly, after we have run a simulation to $z = 0$, we compute the HEFT $P_{ij}(k, a)$ in each snapshot by performing the following algorithm:

1. Re-scale the Gaussian $\delta_{cb,L}(k, a_{\text{ini}})$ field used to initialize the simulations by the ratio of scale dependent growth factors $\frac{D(k, a)}{D(k, a_{\text{ini}})}$.
2. Compute the Lagrangian fields $\mathcal{O}_i(\mathbf{q})$ from $\delta_{cb,L}(\mathbf{k}, a)$.

3. Deposit cb particles to a mesh weighted by the values of $\mathcal{O}_i(\mathbf{q}, a)$ to compute $\mathcal{O}_i(\mathbf{x}, a)$.
4. Measure the cross-power spectrum of $\mathcal{O}_i(\mathbf{x}, a)$ and $\mathcal{O}_j(\mathbf{x}, a)$ to give $P_{ij}(k, a)$.

We use a mesh of size $N = 1400^3$ with a cloud-in-cell mass deposition scheme and interleave the $\mathcal{O}_j(\mathbf{x}, a)$ fields to partially de-alias them [113]. $D(k, a)$ is computed using CLASS. This algorithm is implemented in the `anzu`³ package.

5 Zel'dovich control variates

In order to reduce the variance on many of the measurements that we present in this work, we will make use of Zel'dovich control variates (ZCV). In this section, we briefly summarize the ZCV method, and refer readers to more detailed presentations in [82, 83] for further details.

Control variates [76] are well studied in the statistics literature as a method for reducing the variance on the estimate of the mean of a random variable, X , when a correlated random variable, or control variate, C is available. In this case, we can construct the following quantity

$$Y = X - \beta(C - \mu_c), \quad (5.1)$$

where μ_c is the mean of C , and β can be optimized to minimize $\text{var}[Y]$. Doing so yields

$$\beta^* = \frac{\text{Cov}[X, C]}{\text{Var}[C]}. \quad (5.2)$$

It can then be shown that

$$\frac{\text{Var}[Y]}{\text{Var}[X]} = 1 - \frac{\text{Cov}[X, C]}{\text{Var}[X]\text{Var}[C]} = 1 - \rho_{xc}^2 \quad (5.3)$$

i.e. the variance of Y is reduced with respect to X by an amount proportional to the covariance of X and C , assuming μ_c is known to arbitrary precision. The method of control variates was first applied to cosmology in [77], where COLA [114] simulations were used as control variates for full N -body simulations. Ref. [82] pointed out that one of the main limitations of the control variate technique, namely the Monte-Carlo estimation of μ_c and additional computational expense incurred by this process, can be avoided by using the Zel'dovich approximation (ZA) [115] as a control variate, where μ_c can be calculated analytically to arbitrary precision. More explicitly, in our case, X will be measurements from an N -body simulation, and C will be the analogous measurements made from a Zel'dovich approximation realization with the same initial $\delta_{cb,L}(\mathbf{q})$ as the N -body simulation.

The ZA is known to be highly correlated with full N -body simulations, even when it fails to reproduce their means. Furthermore, in the ZA we can make analytic predictions for μ_c for real- and redshift-space power spectra that are accurate out to high- k . Thus, given the linear density field used to initialize a simulation, we can significantly reduce the variance on measurements of two-point functions made from that simulation.

In this work we will be interested in reducing the variance of HEFT basis spectra measured from our simulations. Doing so will make the surrogate model that we describe in the following sections more accurate, and will allow us to seamlessly transition to pure LPT

³<https://github.com/kokron/anzu>.

predictions for these basis spectra on very large scales. For each HEFT basis spectrum, $\hat{P}_{ij}^{NN}(k)$, we can measure the analogous ZA spectrum, $\hat{P}_{ij}^{ZZ}(k)$, and the cross-spectrum between the HEFT and ZA fields, $\hat{P}_{ij}^{NZ}(k)$, where the hats denote measured quantities and $P_{ij}^{NZ} = \langle \mathcal{O}_i^N \mathcal{O}_j^{ZA} \rangle$. For spectra involving the total matter field, we use the analogous cb field spectra for our ZA control variates.

We compute $\mathcal{O}_i^{ZA}(\mathbf{k}, a)$ using the same algorithm described in section 4, with two key differences. First, we do not rescale the Gaussian initial conditions by $\frac{D(k, a)}{D(k, a_{\text{ini}})}$ for each snapshot, instead opting to use the Lagrangian fields, $\mathcal{O}_i^{ZA}(\mathbf{q})$ computed from $\delta_{cb, L}(\mathbf{k}, a_{\text{ini}})$ and re-scaled by the ratio of scale-independent growth factors $\frac{D(a)}{D(a_{\text{ini}})}$ for all snapshots. This allows us to avoid recomputing $\mathcal{O}_j^{ZA}(\mathbf{q})$ for each snapshot, which would dominate the runtime of the ZCV algorithm. This incurs a negligible cost in the performance of the algorithm. Second, instead of using N -body particles to compute displacements, we use the displacement field produced by `monofonic`. Furthermore, following [82], we smooth $\delta_{cb, L}(\mathbf{k}, a_{\text{ini}})$ and $\Psi(\mathbf{k})$ at the Nyquist frequency of the mesh that the initial conditions are generated on, $k = \frac{\pi N_{cb}}{L_{\text{box}}} = 4.19 \text{ hMpc}^{-1}$ using a Gaussian kernel.

Given these spectra, we can compute β^* assuming a disconnected covariance approximation:

$$\begin{aligned} \beta^* &= \frac{\text{Cov}[\hat{P}_{ij}^{NN}, \hat{P}_{ij}^{ZZ}]}{\text{Var}[\hat{P}_{ij}^{ZZ}]} \\ &= \frac{\hat{P}_{ii}^{NZ}(k) \hat{P}_{jj}^{NZ}(k) + \hat{P}_{ij}^{NZ}(k) \hat{P}_{ji}^{NZ}(k)}{\hat{P}_{ii}^{ZZ}(k) \hat{P}_{jj}^{ZZ}(k) + (\hat{P}_{ij}^{ZZ}(k))^2}. \end{aligned}$$

On small enough scales, where the fields being correlated become highly non-Gaussian, this disconnected approximation inevitably fails. In order to avoid adding extra noise to our measured HEFT spectra on these scales, we damp β^* to zero using a tanh function, with the same damping parameters, $k_0 = 0.618 \text{ hMpc}^{-1}$ and $\delta k = 0.167 \text{ hMpc}^{-1}$, as described in [82]. We have investigated whether a redshift dependent damping of β^* is warranted, but found that the required damping parameters are relatively constant with redshift in the range probed by our simulations, and so we have opted to keep them fixed.

Figure 7 demonstrates the effective volume of our simulations as a function of wavenumber when employing ZCV to reduce the variance of HEFT spectra. We take $V_{\text{eff}} = \frac{V_{\text{fid}}}{1 - \rho_{xc}^2}$, where ρ_{xc} is the cross correlation coefficient between the HEFT and ZA basis spectra in question, again estimated using a disconnected covariance approximation and applying the same damping factor to ρ_{xc} as to β^* . The results, reiterating those shown in [82], are quite impressive, with improvements of factors of $10 - 10^5$ in effective volume depending on the spectrum and scale in question. We also observe a slight decrease in ρ_{xc} , and thus V_{eff} , as a function of redshift at fixed k . This is to be expected as beyond-linear displacements become appreciable at lower wave-numbers as a function of redshift. At $k > k_0$, the effective volume asymptotes to V_{fid} as a consequence of our damping of β^* . V_{eff} turns over on very large scales for spectra that include the δ field, a behavior not seen in [82]. We attribute this to our decision to not recompute $\mathcal{O}_i(\mathbf{q})$ at each redshift for the ZA fields, thus causing a slight de-correlation due to different scale dependent growth in the N -body and ZA fields.

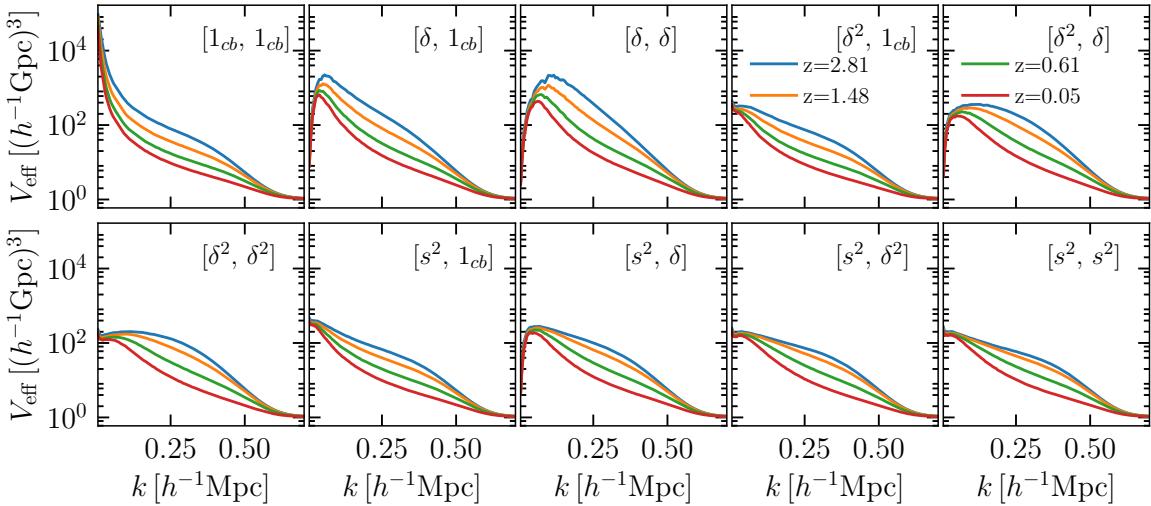


Figure 7. Effective volume for our HEFT spectrum measurements after applying ZCV. The asymptote to $(1.05 h^{-1} \text{Gpc})^3$ at $k > 0.6$ is due to the damping that we apply to β^* . We do not show spectra computed with the total matter field, as the results are nearly identical to the analogous spectra using the δ_{cb} field.

6 Surrogate model construction and performance

6.1 Surrogate model methodology

Having measured HEFT spectra for all of our simulations and reduced their variance using ZCV, we now proceed to construct surrogate models for them. We generally follow the surrogate modeling methodology described in [63], with a few notable changes which we highlight when described. The quantities that we emulate are not the basis spectra, P_{ij} , but rather the logarithm of the ratio of these spectra to their 1-loop LPT counterparts:

$$\Gamma^{ij}(k, \Omega) = \lg \left[\frac{P_{ij}^{\text{HEFT}}(k, \Omega)}{P_{ij}^{\text{1-loop}}(k, \Omega)} \right] \quad (6.1)$$

where Ω is a set of $w\nu$ CDM cosmological parameters and $\sigma_8(z)$. We have empirically found that using $\sigma_8(z)$ instead of redshift or scale factor as a time variable significantly improves the performance of our surrogate models. Before taking the logarithm of the ratio of the HEFT and 1-loop LPT spectra, we apply a third-order Savitsky-Golay filter to the ratio with a window length of 11 in order to remove additional variance. We have tested that this smoothing procedure does not lead to any appreciable bias.

Importantly, and differently from [63], we do not use standard convolutional Lagrangian effective field theory (CLEFT) [116] predictions for $P_{ij}^{\text{1-loop}}(k, \ln a, \Omega)$, but rather infrared (IR) re-summed “k-expanded” CLEFT (KECLEFT), where the long-wavelength displacement correlators A_{ij} are expanded in a Taylor series as described in appendix E of [117]. Because A_{ij} is Taylor expanded rather than exponentiated directly, we are able to avoid damping the linear power spectrum that goes into these terms, leading to KECLEFT predicting additional power at small scales compared to CLEFT. This additional power in KECLEFT predictions leads to more stable ratios at high- k in eq. (6.1) than those derived using

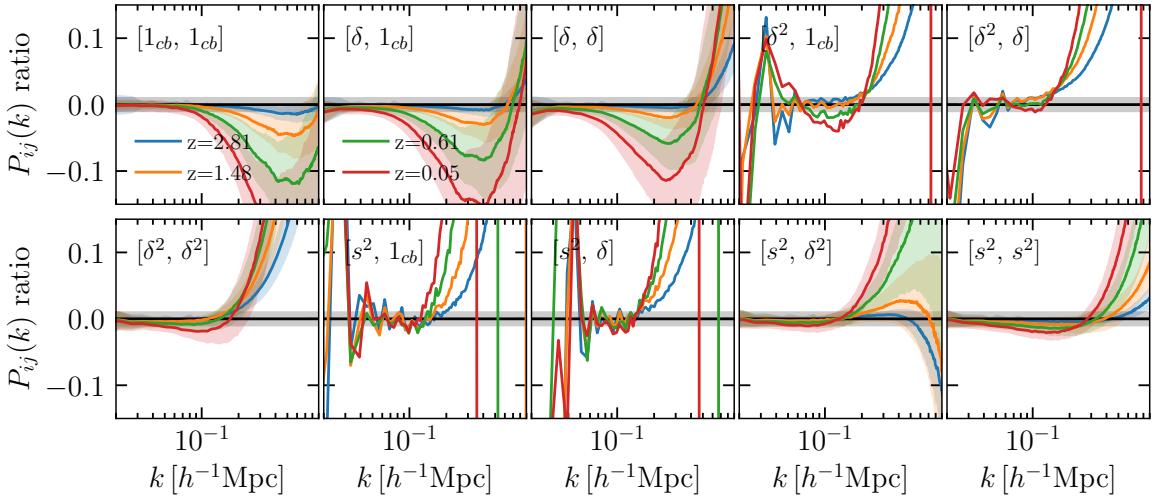


Figure 8. Fractional deviation of HEFT basis spectra from 1-loop LPT predictions, i.e. $P_{ij}^{\text{HEFT}}/P_{ij}^{\text{1-loop}} - 1$. Solid lines represent the means, and shaded regions are one sigma errors, while different colors represent different redshifts. On large scales, the spectra agree nearly perfectly. For spectra depending on cubic combinations of fields, e.g. $P_{s^2,cb}$, we have included an additional term proportional to $k^2 P_{cb,cb}(k)$ in order to bring HEFT and 1-loop LPT spectra into agreement on large scales. The need to do this stems from the different smoothing conventions used in our simulations and 1-loop LPT predictions.

CLEFT, thus reducing the dynamic range in Γ^{ij} that we must emulate. Using CLEFT instead of KECLEFT, while leaving all other choices the same leads to factors of two or more degradation in surrogate model errors.

Figure 8 shows ratios of the HEFT basis spectra to their 1-loop LPT counterparts. On large scales, the spectra converge to one another, indicating that the HEFT spectra are equivalent to 1-loop LPT predictions on these scales. In particular, $P_{cb,cb}$ matches the 1-loop LPT prediction nearly perfectly on the largest scales plotted here, indicating that our simulations recover linear growth to significantly sub-percent accuracy. Notably, HEFT spectra involving cubic combinations of fields, such as $P_{s^2,cb}$, do not converge to their LPT counterparts over the range of scales that are measurable in our simulations. The residuals between HEFT and LPT spectra in these cases take the form of $\alpha k^2 P_{cb,cb}$, where α is a free coefficient. This is the same form as the derivative bias spectra, $P_{\nabla^2,cb}$, indicating that the differences in these spectra stem from differences in the smoothing conventions employed in our simulated measurements and the LPT predictions, due to our choice to implicitly smooth our predictions at the scale of the initial conditions grid. In figure 8 we have included this extra $\alpha k^2 P_{cb,cb}$ in the cubic 1-loop LPT predictions. Because we always include these derivative bias operators in our predictions when fitting observational data, the differences between the large scale behavior of these spectra are unimportant when interpreting data.

After measuring Γ^{ij} for each snapshot, we perform a principal component analysis (PCA) decomposition. We employ PCA as we wish to reduce the dimensionality of our surrogate modeling problem. We construct the $N \times M$ matrix, \mathbf{X}^{ij} , where

$$X_{\alpha\beta}^{ij} = \Gamma^{ij}(k_\beta, \Omega_\alpha) - \langle \Gamma^{ij}(k_\beta) \rangle, \quad (6.2)$$

where $N = N_{\text{sim}} \times N_{\text{snap}}$, where N_{sim} is the number of simulations, 150, and N_{snap} is the

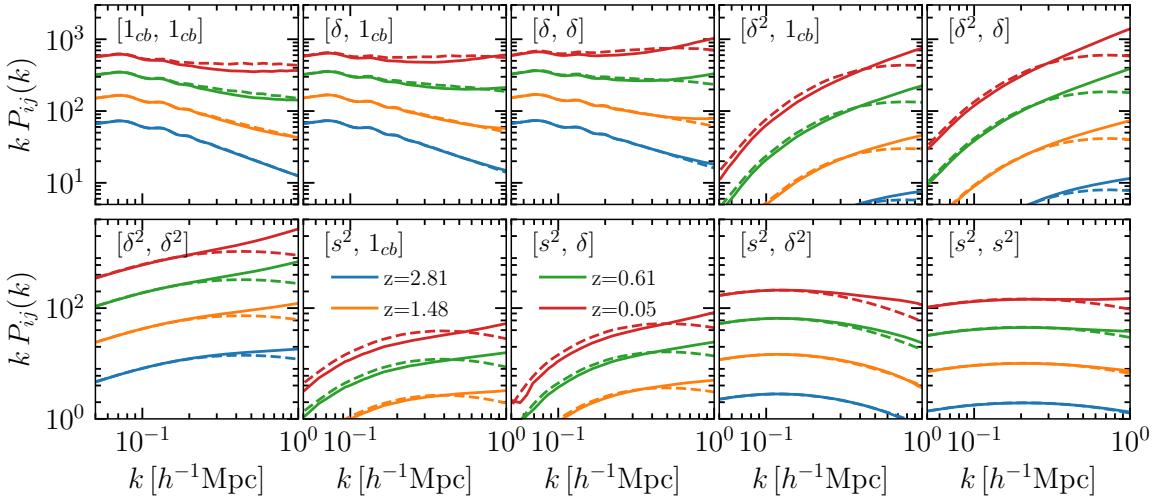


Figure 9. Predictions from our HEFT surrogate model (solid), compared with their analogous 1-loop LPT predictions (dashed). The combination of our ZCV procedure and the use of PCA in our surrogate model leads to nearly noiseless model predictions.

number of snapshots per simulation, 30, and M is the number of wavenumbers in our measurements of Γ^{ij} . Ω_α is a vector containing the cosmology and $\sigma_8(z)$ value for one of our simulation snapshots. The average in the last term on the right hand side of eq. (6.2) is taken over all cosmologies and redshifts in our training set. For $P_{m,m}$, $P_{m,cb}$ and $P_{cb,cb}$ we perform a PCA on wavenumbers between $k_{\min} = 0.05 \text{ hMpc}^{-1}$ and $k_{\max} = 4 \text{ hMpc}^{-1}$, giving $M = 659$, while for the rest of the spectra we take $k_{\max} = 1 \text{ hMpc}^{-1}$, where $M = 159$. $P_{m,m}$, $P_{m,cb}$ and $P_{cb,cb}$ are relevant for cosmic shear and intrinsic alignment predictions, and so predictions between $k = 1 \text{ hMpc}^{-1}$ and $k = 4 \text{ hMpc}^{-1}$ are useful, while the rest of the spectra are only used in the Lagrangian bias expansion, which we expect to break down significantly before $k = 1 \text{ hMpc}^{-1}$ and so there is no reason to emulate beyond this. Additionally, we have found that Γ^{ij} becomes numerically unstable for $k > 1 \text{ hMpc}^{-1}$, further motivating our choice to set $k_{\max} = 1 \text{ hMpc}^{-1}$ for everything other than $P_{m,m}$, $P_{m,cb}$ and $P_{cb,cb}$. The principal components (PCs) are then given by

$$\begin{aligned} \mathbf{C}^{ij} &= (\mathbf{X}^{ij})^T \mathbf{X}^{ij} \\ &= (\mathbf{W}^{ij})^T \mathbf{\Lambda}^{ij} \mathbf{W}^{ij} \end{aligned}$$

where \mathbf{W}^{ij} is an $M \times M$ array whose rows are the eigenvectors that we will use, i.e. $W_{\alpha\beta}^{ij} = \text{PC}_\alpha(k_\beta)$, and $\mathbf{\Lambda}^{ij}$ are the corresponding eigenvalues. By keeping a subset of N_{PC} of these PCs, we can further reduce the noise on our final predictions without decreasing their accuracy. We have found that the error incurred by truncating at $N_{\text{PC}} = 20$ is less than 0.1%, and so we use $N_{\text{PC}} = 20$ for the duration of this work. This is notably higher than what was required in [63], where we used 2 PCs. The need for more PCs in this work is driven both by the larger range in scales that we emulate here, as well as the significantly broader range in cosmology and redshift in the **Aemulus** ν simulations. We then project all Γ^{ij} onto these PCs:

$$\mathbf{A}^{ij} = \mathbf{X}^{ij} \mathbf{W}^{ij}, \quad (6.3)$$

such that

$$\Gamma^{ij}(k_\beta, \Omega_\alpha) = A_{\alpha\gamma}^{ij} W_{\gamma\beta}^{ij} + \langle \Gamma^{ij}(k_\beta) \rangle.$$

The remaining task is to construct surrogate models, \tilde{A}_γ^{ij} , for the cosmology and redshift dependence of the first N_{PC} PC coefficients, $A_{\alpha\gamma}^{ij}$, such that

$$A_{\alpha\gamma}^{ij} \simeq \tilde{A}_\gamma^{ij}(\Omega_\alpha). \quad (6.4)$$

We will use polynomial chaos expansions (PCEs) [118] as our model for \tilde{A}_γ^{ij} . A PCE of order p is an expansion in orthonormal polynomials:

$$\tilde{A}(\Omega) = \sum_{|\beta| \leq p} \eta_\beta \Psi_\beta(\Omega) \quad (6.5)$$

where $\beta \in \mathbb{N}^D$, and D is the dimensionality of the domain of \tilde{A} . In our case $D = 7 + 1$, as we have $\Omega = \{\omega_b, \omega_c, w, n_s, 10^9 A_s, H_0, \sum m_\nu, \sigma_8(z)\}$. The independent variables in our problem are uncorrelated and uniformly distributed (other than $\sum m_\nu$), so we can use the Stieltjes three-term recurrence relation [119] to construct orthonormal polynomials, $\psi_{\beta_i}(\Omega_i)$, of order β_i in each variable independently. In practice we scale each parameter Ω_i such that $\Omega_i \in [-1, 1]$, and so this recurrence relation yields functions that are proportional to the Legendre polynomials, i.e. $\psi_{\beta_i}(\Omega_i) \propto P_{\beta_i}(\Omega_i)$, where P_{β_i} is the order β_i Legendre polynomial. Ψ_β is then given by

$$\Psi_\beta(\Omega) = \prod_{i=1}^D \psi_{\beta_i}(\Omega_i), \quad (6.6)$$

and we obtain the final expression for our surrogate models:

$$\tilde{\Gamma}^{ij}(k_\alpha, \Omega) = \tilde{A}_\gamma^{ij}(\Omega) W_{\gamma\alpha}^{ij} + \langle \Gamma^{ij}(k_\alpha) \rangle \quad (6.7)$$

$$= \eta_{\gamma\beta}^{ij} \Psi_\beta(\Omega) W_{\gamma\alpha}^{ij} + \langle \Gamma^{ij}(k_\alpha) \rangle. \quad (6.8)$$

We then use least-squares regression to fit for $\eta_{\gamma\beta}^{ij}$ given a chosen maximum order p . We perform these fits using the `chaospy` [120] python package.

In order to optimize our choice of p , we minimize a suitably defined measurement of the generalization error of our surrogate models. For this work we use

$$\epsilon_{ij}^{\max} = \text{Var} \left[\max_{k,z} \frac{P_{ij}(k, z, \Omega) - \tilde{P}_{ij}(k, z, \Omega)}{P_{ij}(k, z, \Omega)} \right] \quad (6.9)$$

i.e. the variance of the maximum error taken over all k and z of our surrogate model \tilde{P}_{ij} . Since we do not have a separate suite of simulations on which to evaluate this error, we instead use a cross-validation approach, leaving one simulation out of our training set at a time, computing the error on the simulation that has been left out, and using that to evaluate eq. (6.9). We then perform a grid search over $1 \leq p_i \leq 7$, independently for each input parameter. The resulting best fit polynomial order is $\mathbf{p} = \{2, 3, 3, 2, 2, 3, 2, 6\}$.

Figure 9 shows the predictions from our surrogate model compared to the corresponding 1-loop LPT predictions as a function of redshift. The main notable features are that the HEFT predictions asymptote to the 1-loop LPT predictions at low k , and that there is no discernible residual noise on our simulation predictions.

6.2 Surrogate model performance

Figures 10 and 11 summarize the performance of our surrogate models after performing the optimization procedure described in the previous section. Figure 10 shows the one-sigma leave-one-out error evaluated for the more restrictive parameter space Tier 2 simulations as a function of k and z for each basis spectrum. We do not plot spectra that include the δ_m field, as they exhibit nearly identical behavior to their δ_{cb} counterparts. Note, that we still train on the full set of Tier 1 and Tier 2 simulations in this case. The bottom triangle depicts the fractional error for each basis spectrum, and the top triangle shows the contribution of that fractional error to a tracer auto-power spectrum with $b_1 = 1$, $b_2 = 0.2$, $b_s = 0.2$, and $\bar{n} = 5 \times 10^{-4} h^3 \text{Mpc}^{-3}$. The errors are significantly below one percent for most redshifts, other than for spectra including the s^2 field. For reasonable choices of b_s , these terms are typically small, and so 2% errors in these spectra contribute at the $< 0.01\%$ level to tracer auto-power predictions, as can be seen in the upper triangle. Figure 11 shows the same thing, but evaluated over both the Tier 1 and Tier 2 simulations. The trends remain the same, but with approximately 50% worse performance in the broader parameter space. We emphasize that much of this parameter space is already significantly ruled out, and we include it when constructing surrogate models in order to stabilize their performance around the edges of the parameter space allowed by current data.

Figure 12 shows the performance of our surrogate model for $P_{m,m}$ compared to three other state-of-the-art models: `HMCode2020` [95], `CosmicEmu` [55] and `EuclidEmu2` [54]. The top row of the figure depicts the 68th percentile error of each model compared to the `Aemulus` ν Tier 2 simulations where lines are color coded by their redshift, while the bottom panel shows errors for individual simulations at $z = 0.53$. This redshift was chosen to be close to the peak of the galaxy-lensing kernel for upcoming galaxy-weak-lensing surveys. The performance of our $P_{m,m}$ and $P_{m,cb}$ surrogate models are quite comparable to the performance of the $P_{cb,cb}$ surrogate in figure 10, with slightly degraded performance above $k = 1 h\text{Mpc}^{-1}$.

We see that the `HMCode2020` model has a 68th percentile error for $z \leq 2$ close to the 2.5% error quoted in [95]. `EuclidEmu2` and `CosmicEmu` have a relatively smaller 68th percentile error, peaking at about 2% at $k = 1 h\text{Mpc}^{-1}$ and staying relatively constant with redshift. The error increases significantly at the highest k and z shown in this figure, likely due to insufficient resolution in our simulations at these high redshifts and wavenumbers. At low wavenumbers, the `CosmicEmu` predictions disagree with our simulations, and linear theory at the $\sim 1 - 1.5\%$ level. There is also residual sample variance at the 1% level discernible between $0.1 h\text{Mpc}^{-1} \leq k \leq 0.5 h\text{Mpc}^{-1}$ at low redshifts in the `EuclidEmu2` model, where the paired and fixed amplitude simulations they employ remove less sample variance from their measurements than our ZCV technique. Although not plotted here, we find that `EuclidEmu2` gives a 50th percentile error of about 1% at $k \sim 1 h\text{Mpc}^{-1}$ and above, roughly consistent with the statistics quoted in [54]. We note that for `CosmicEmu` and `EuclidEmu2` only 18 and 27 of our tier 2 simulations are within their training domain, respectively, and `CosmicEmu` does not make predictions for $z > 2.02$. `HMCode2020` fares significantly worse than `CosmicEmu`, `EuclidEmu2` and our surrogate model in the quasi-linear regime, due to well studied issues in halo models at k ranges that bridge the one- and two-halo regimes [121].

The black solid and dashed lines show the change in the matter power spectrum for a change of $\delta \sum m_\nu = 0.03 \text{ eV}$ and $\delta w = 0.025$ away from the Planck 2018 best fit cosmology. Given the errors on our surrogate model, we would be able to distinguish these changes, whereas the other models considered would not be able to tell these apart from surrogate modeling errors. The $\delta \sum m_\nu = 0.03 \text{ eV}$ is particularly relevant, because cosmological con-

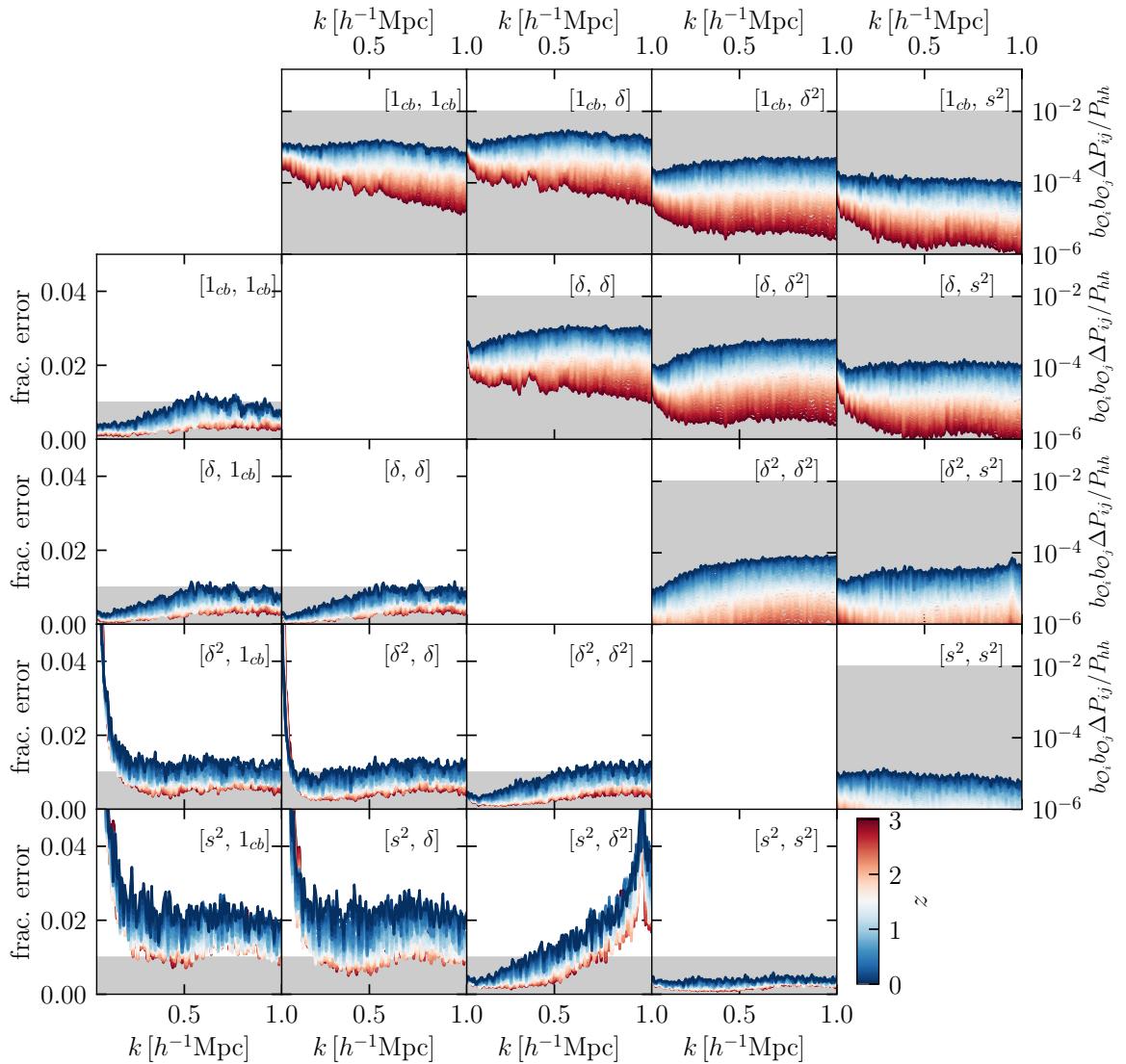


Figure 10. (lower triangle) Leave-one-out 68th percentile fractional error of our surrogate model for each basis spectrum evaluated over the tier 2 simulations. Lines are color coded by redshift, going from red to blue from $z = 3$ to $z = 0$. The shaded grey region shows 1% errors; most of the statistics and redshifts fall below this error threshold. We note that this cannot be directly compared to figure 5 of [63], as there we quoted errors in terms of median absolute deviation (MAD) from our simulation measurements. The surrogate model presented in this work is more accurate than that of [63] in terms of MAD. (upper triangle) Same as for the lower triangle, but now quoted in terms of the fractional error contribution to P_{hh} , assuming $b_1 = 1$, $b_2 = 0.2$, $b_s = 0.2$, and $\bar{n} = 5 \times 10^{-4} h^3 \text{Mpc}^{-3}$.

straints must achieve $\sigma(\sum m_\nu) = 0.03 \text{ eV}$ in order to conclusively distinguish the normal and inverted mass hierarchies (e.g. [122, 123]).

6.3 Error modeling

Although this performance satisfies our goals of $\leq 2\%$ 68th percentile error over our full parameter space, and $\leq 1\%$ 68th percentile error in our tier 2 parameter space, the errors on our surrogates may not be entirely negligible for all analyses. In particular, as shown in

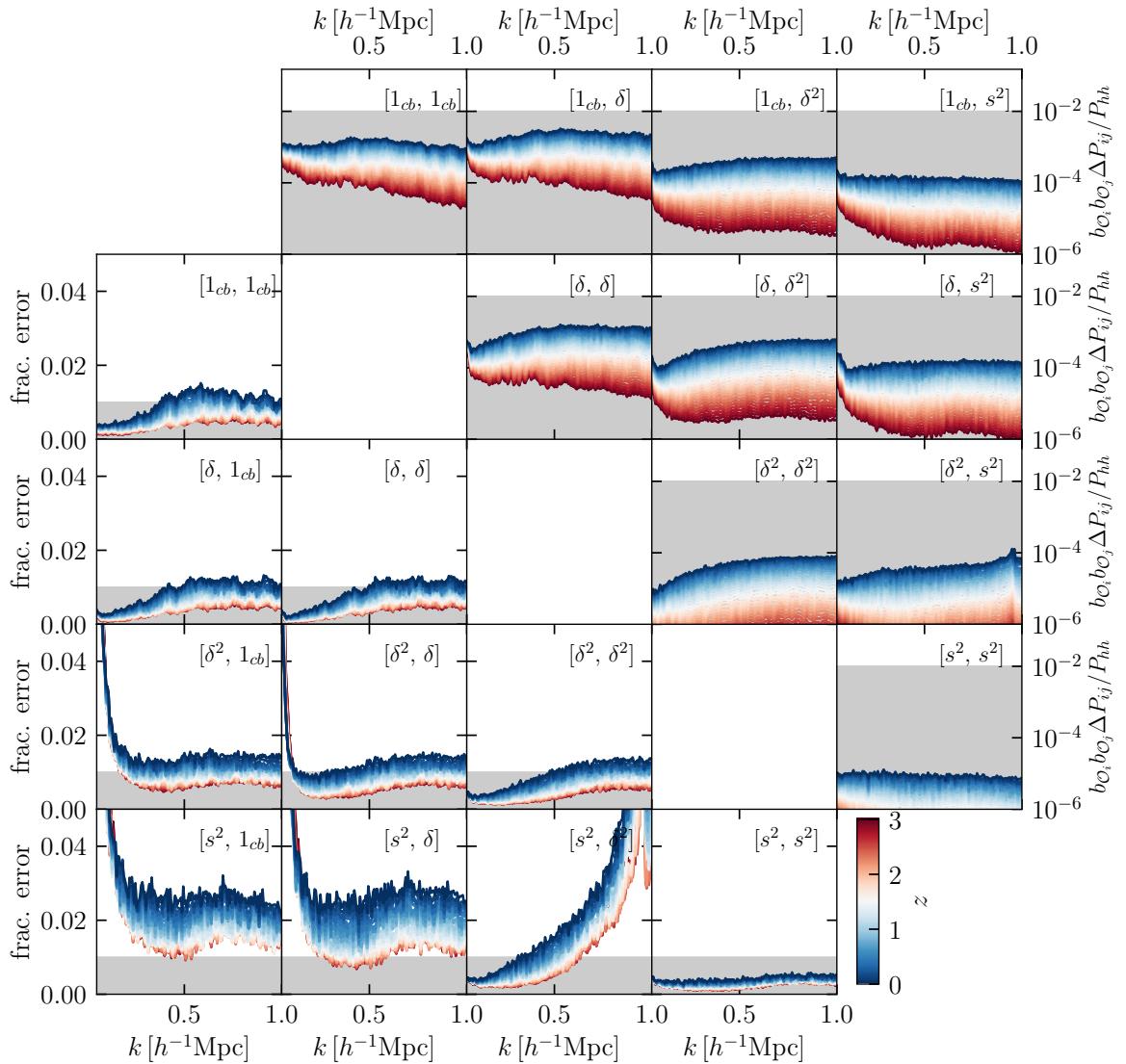


Figure 11. Same as figure 10, but now evaluated over the entire simulation suite. The broader range in cosmologies results in approximately 50% larger error.

figure 13, the errors have correlations as a function of scale that may contribute significantly to otherwise small off-diagonal elements of covariance matrices. To facilitate the incorporation of these errors into analyses, we provide additional functionality to produce covariance matrices of these fractional errors for each basis spectrum along with our trained surrogate models. In particular, we measure the following quantity:

$$\text{Cov} [\epsilon_{i,j}(k), \epsilon_{l,m}(k')] = \sum_{n=0}^{N_{\text{sim}}} (\epsilon_{i,j}^n(k) - \bar{\epsilon}_{i,j}(k)) (\epsilon_{l,m}^n(k') - \bar{\epsilon}_{l,m}(k')) , \quad (6.10)$$

where

$$\epsilon_{i,j}^n(k) = \frac{P_{ij}(k, z, \Omega^n) - \tilde{P}_{ij}^n(k, z, \Omega^n)}{P_{ij}(k, z, \Omega^n)} \quad (6.11)$$

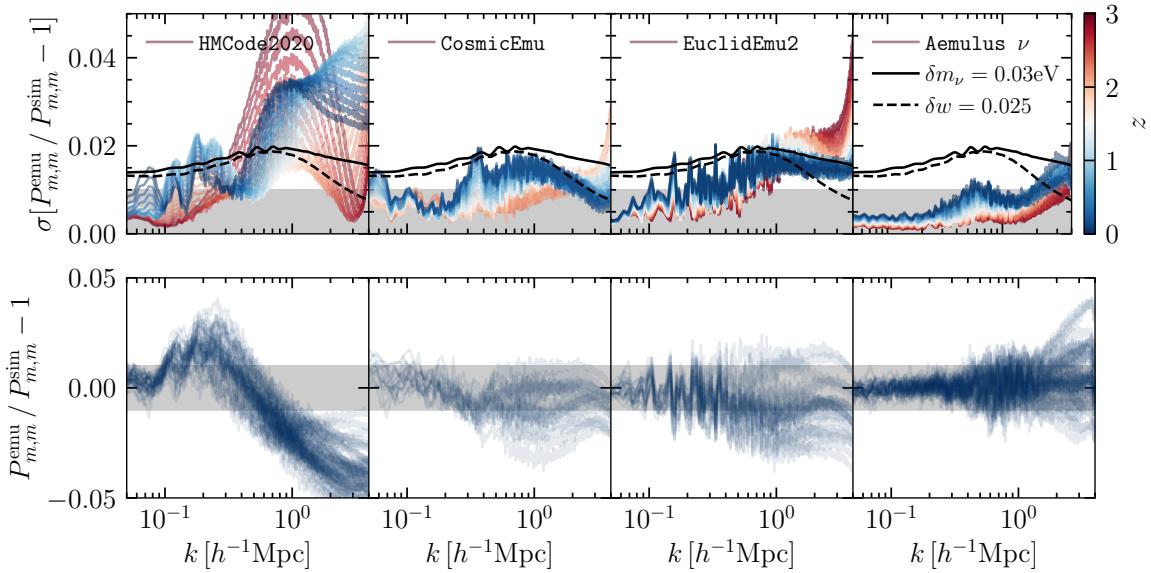


Figure 12. (Top) 68th percentile fractional error of `HMCode2020` (left) `CosmicEmu` (middle left) `EuclidEmu2` (middle right) and the matter power spectrum surrogate model presented in this work (right) to our tier 2 simulations. For `Aemulus ν` we use leave-one-out errors. The black solid and dashed lines show the change in the matter power spectrum for a change of $\delta \sum m_\nu = 0.03 \text{ eV}$ and $\delta w = 0.025$ away from the Planck 2018 best fit cosmology. (Bottom) Fractional error at $z = 0.53$ for each of our tier 2 simulations. For `CosmicEmu` and `EuclidEmu2` we only compare to 18 and 27 of our tier 2 simulations that are in their domains, respectively. None of the models exhibit significant outliers in their performance compared to the 68th percentile errors shown in the top panels. The performance of `Aemulus ν` becomes slightly more unstable above $k = 1 \text{ hMpc}^{-1}$, with a few 2–3% outliers apparent.

is the fractional error for simulation n in our suite and $\tilde{P}_{ij}^n(k, z, \Omega^n)$ is the emulator prediction trained on all simulations except for the n th one. We can then compute errors on P_{gg} assuming bias parameters b_i as

$$\text{Cov}^{\text{emu}} [P_{gg}(k), P_{gg}(k')] = \sum_{i,j,l,m} b_i b_j b_l b_m P_{i,j}(k) P_{l,m}(k') \text{Cov} [\epsilon_{i,j}(k), \epsilon_{l,m}(k')] \quad (6.12)$$

and this covariance can then be further integrated along the line of sight, via e.g. the Limber approximation, to produce emulator error contributions to the covariances of angular power spectra. One such prediction for a DESI-like sample is used in figure 1.

7 Summary

In this work, we have introduced the `Aemulus ν` suite, a set of 150 N -body simulations in a $w\nu$ CDM parameter space that evolve massive neutrinos as an additional particle species. The $w\nu$ CDM cosmological parameter space of these simulations is sufficiently broad to make them useful for investigating tensions between cosmological constraints coming from large scale structure and cosmic microwave background experiments. Along with these simulations, we present new hybrid effective field theory (HEFT) and matter power spectrum surrogate models that represent significant improvements over the current state of the art.

In section 2 we describe how the sampling of $w\nu$ CDM parameter space was optimized in order to simultaneously maximize parameter space breadth while maintaining surrogate

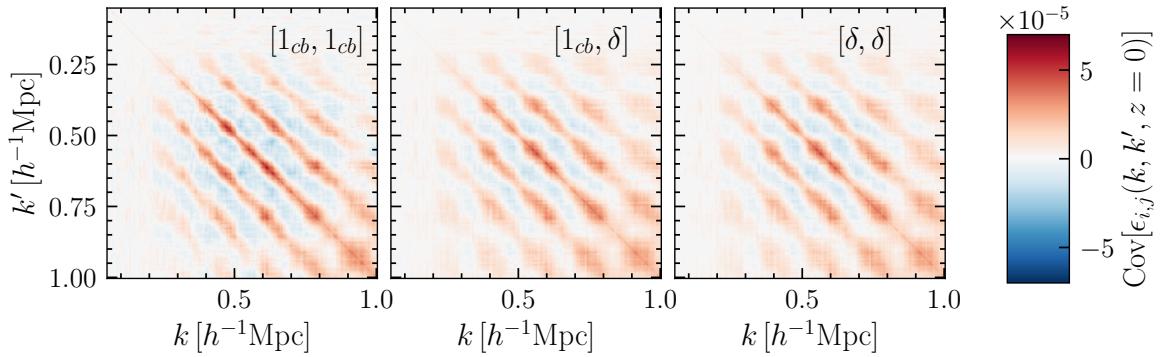


Figure 13. Covariance of the fractional error of our $P_{cb,cb}$, $P_{cb,\delta}$ and $P_{\delta,\delta}$ surrogate models at $z = 0$. Significant off-diagonal structure is present, although the amplitude of the overall errors are small. Our surrogate modeling code provides an interface to these covariances so that they can be appropriately accounted for when using the models presented in this work in data analysis.

model accuracy. To do so, we constructed surrogate models for `HMCode2020` matter power spectrum predictions, varying the allowed boundaries of our parameter space. Using these surrogate models, we found that a two-tiered design, with 100 simulations in a broad $w\Lambda$ CDM parameter space, and 50 simulations in a parameter space with bounds restricted to be closer to currently preferred constraints, yielded an accuracy that met our requirements.

In section 3, we describe the settings that were used to run our simulations, the simulation convergence tests that we performed. We use a modified version of `Gadget3` to evolve 1400^3 CDM and neutrino particles (for a total of 2×1400^3 particles) in a $(1.05h^{-1}\text{Gpc})^3$ volume, yielding a cb particle mass of $3.51 \times 10^{10} \frac{\Omega_{cb}}{0.3} h^{-1} M_\odot$. In order to mitigate systematic errors associated with early simulation starting times, we initialize our simulations at $z = 12$ using third order Lagrangian perturbation theory (3LPT) as implemented in a modified version of `monofonic`. As this is a significant change from previous simulations run as part of the `Aemulus` project, we perform convergence tests demonstrating that starting at $z_{\text{ini}} = 12$ is converged at the $\leq 1\%$ level with respect to a simulation run with four times the number of particles, starting at $z_{\text{ini}} = 24$.

Section 4 introduces relevant perturbation theory notation, and describes how we measure HEFT basis spectra from our simulations. Then, in section 5, we describe our Zel'dovich control variate (ZCV) methodology, showing that it reduces the sample variance on HEFT spectra to the equivalent errors that we would obtain by running $10 - 10^5$ times larger simulations, depending on the basis spectrum and scale in question. Doing so leads to extremely smooth predictions from our surrogate models, outperforming the improvements obtained from paired and fixed simulations, and allowing us to run more simulations than we would have otherwise been able to while still meeting our accuracy requirements. We believe the ZCV technique will play an important role in similar applications in the future.

We described the implementation and optimization of our combined principal component analysis (PCA) and polynomial chaos expansion (PCE) surrogate models in section 6. There, we demonstrated that our HEFT surrogate model achieves $\leq 1\%$ 68th percentile error for most basis spectra over our Tier 2 parameter space and $\leq 2\%$ 68th percentile error over our full parameter space for $k \leq 1 \text{hMpc}^{-1}$ and $0 \leq z \leq 3$. The basis spectra that exceed these error thresholds are sub-dominant, and thus their slightly worse accuracy is not of great concern. Our matter power spectrum surrogate model achieves $\leq 1\%$ 68th percentile

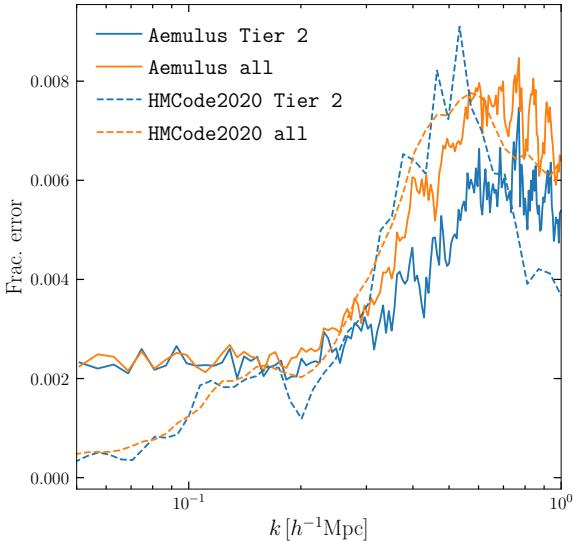


Figure 14. Comparison between 68th percentile fractional residuals, averaged over all redshifts, of our HMCode2020 $P_{m,m}$ surrogate model (dashed) and **Aemulus** ν surrogate model (solid), when both are trained using our fiducial parameter space design. For the HMCode2020 model, errors are evaluated using a set of 10,000 cosmologies generated over the Tier 1 parameter space, while errors for the **Aemulus** ν model are evaluated using the leave-one-out methodology described in section 6. Blue lines are errors measured in the Tier 2 parameter space, while orange lines are evaluated over the full Tier 1 parameter space. The errors are quite comparable between the model built on HMCode2020 data compared to that built from our actual **Aemulus** ν simulations, except at $k \leq 0.2 h\text{Mpc}^{-1}$, where residual sample variance may be contributing additional variance to the **Aemulus** ν based model.

error over our tier 2 simulations out to $k = 1 h\text{Mpc}^{-1}$ and $\leq 2\%$ 68th percentile error for $1 h\text{Mpc}^{-1} < k \leq 4 h\text{Mpc}^{-1}$. We compare our matter power spectrum model to HMCode2020, CosmicEmu and EuclidEmu2, finding that our model outperforms them when evaluated on the tier 2 **Aemulus** ν simulations. We also provide estimates of surrogate model error covariance as a function of redshift along with our trained surrogate models so that it may be incorporated as a theory error in analyses using our models.

We anticipate that the HEFT model presented here will be useful for analyses of projected galaxy clustering and weak lensing for many future surveys. This model can straightforwardly replace 1-loop LPT models for these statistics, extending their reach in scale by a factor of two to three. HEFT can also serve as an immediate upgrade in terms of model flexibility and accuracy for analyses that currently use non-linear matter power spectra and constant linear bias. The matter power spectrum model presented here is the most accurate to date, and can play a vital role in stage III and stage IV analyses of weak lensing auto-spectra. We have made these models available at https://github.com/AemulusProject/aemulus_heft. We will also make halo catalogs and downsampled particle catalogs available at https://github.com/AemulusProject/aemulus_nu_public upon publication of this work, and full snapshot data will be made available upon request to the authors. In future work we will also release new halo mass function and halo bias models based on the simulations presented here.

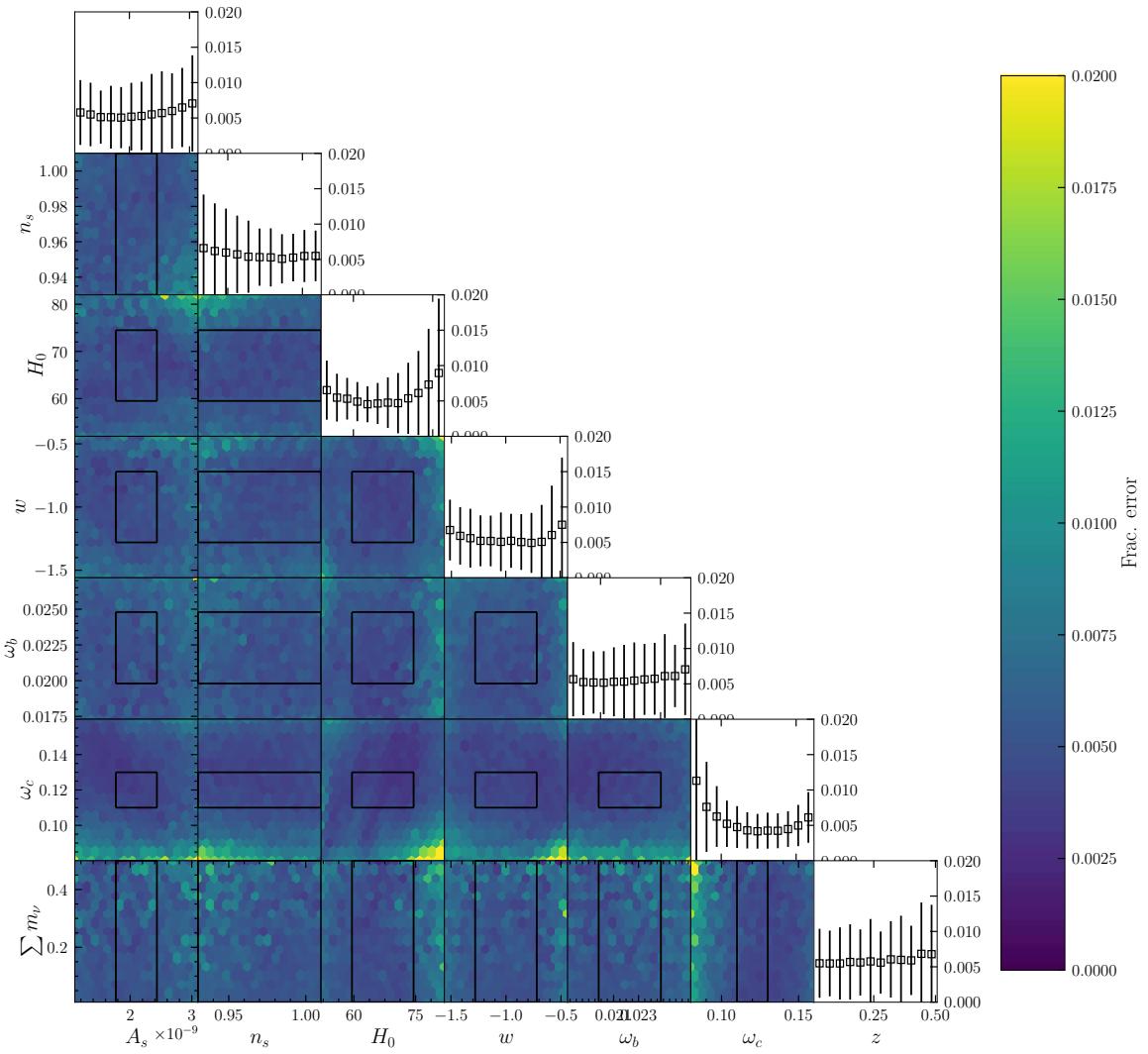


Figure 15. 68th percentile fractional error at $k = 1 \text{ hMpc}^{-1}$ evaluated for the `HMCode2020` $P_{m,m}$ surrogate model trained on our fiducial parameter space design, and tested on a set of 10,000 cosmologies in the Tier 1 parameter space drawn from a Sobol sequence. The black boxes represent the boundaries of the Tier 2 parameter space. The points with error bars are the one-dimensional marginal mean fractional errors with 1-sigma errors.

Acknowledgments

The authors thank Michaël Michaux, Oliver Hahn and Willem Elbers for making `monofonic` publicly available. J.D. is supported by the Lawrence Berkeley National Laboratory Chamberlain Fellowship. S.C. is supported by the Bezos Membership at the Institute for Advanced Study. M.W. is supported by the DOE. K.S.F. is supported by the NASA FINESST program under award number 80NSSC20K1545. This research has made use of NASA’s Astrophysics Data System and the arXiv preprint server. This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Calculations and figures

in this work have been made using the SciPy Stack [124–126] and `chaospy` [120]. Power spectrum measurements were made with `pypower`. This work used `Stampede2` at the Texas Advanced Computing Center and `Bridges2` at the Pittsburgh Supercomputing Center through allocation PHY200083 from the Extreme Science and Engineering Discovery Environment (XSEDE) [127], which was supported by National Science Foundation grant number 1548562.

A HMcode2020 surrogate model error

In this appendix we discuss the accuracy of the HMCode2020 surrogate model trained on the same cosmologies used to run the `Aemulus` ν simulations. In order to train our HMCode2020 surrogate model we use the same optimization procedure described in section 6. In figure 14 we compare the resulting HMCode2020 68th percentile surrogate model error taken over all cosmologies and redshifts to that obtained using the `Aemulus` ν simulations. The errors for the HMCode2020 model are evaluated on the independent set of 10,000 HMCode2020 predictions sampled using a Sobol sequence [100] over the Tier 1 cosmology space, while the `Aemulus` ν errors are computed using the leave-one-out cross-validation procedure described in section 6. For $k > 0.3 \text{ hMpc}^{-1}$ the errors are quite comparable to each other both for the Tier 2 cosmologies (blue) and the full set of cosmologies (orange). At lower k values than this, the `Aemulus` ν errors are notably larger, likely due to residual sample variance in our simulation measurements.

Because we have a dense set of HMCode2020 test models, we can accurately measure the cosmology dependence of our surrogate model error. Given that the cosmology averaged 68th percentile errors are comparable between our HMCode2020 and `Aemulus` ν models, we can hope that the cosmology dependence of the errors are also similar. Figure 15 shows the 68th percentile fractional error, taken over cosmology and redshift, at $k = 1 \text{ hMpc}^{-1}$ as a function of cosmology for our HMCode2020 surrogate model. The black lines represent the bounds of the Tier 2 parameter space. Within this region the error is very stable and almost always below 1%. Beyond this, the error monotonically increases, with some excursions beyond 2% at the edges of our parameter space in w , ω_c and H_0 .

References

- [1] G. D’Amico et al., *The cosmological analysis of the SDSS/BOSS data from the effective field theory of large-scale structure*, *JCAP* **05** (2020) 005 [[arXiv:1909.05271](https://arxiv.org/abs/1909.05271)] [[INSPIRE](#)].
- [2] M.M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological parameters from the BOSS galaxy power spectrum*, *JCAP* **05** (2020) 042 [[arXiv:1909.05277](https://arxiv.org/abs/1909.05277)] [[INSPIRE](#)].
- [3] S.-F. Chen, Z. Vlah, E. Castorina and M. White, *Redshift-space distortions in Lagrangian perturbation theory*, *JCAP* **03** (2021) 100 [[arXiv:2012.04636](https://arxiv.org/abs/2012.04636)] [[INSPIRE](#)].
- [4] O.H.E. Philcox et al., *Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order*, *Phys. Rev. D* **106** (2022) 043530 [[arXiv:2206.02800](https://arxiv.org/abs/2206.02800)] [[INSPIRE](#)].
- [5] G. D’Amico et al., *The BOSS bispectrum analysis at one loop from the effective field theory of large-scale structure*, [arXiv:2206.08327](https://arxiv.org/abs/2206.08327) [[INSPIRE](#)].
- [6] G. D’Amico et al., *The one-loop bispectrum of galaxies in redshift space from the effective field theory of large-scale structure*, [arXiv:2211.17130](https://arxiv.org/abs/2211.17130) [[INSPIRE](#)].
- [7] M. Lewandowski, A. Maleknejad and L. Senatore, *An effective description of dark matter and dark energy in the mildly non-linear regime*, *JCAP* **05** (2017) 038 [[arXiv:1611.07966](https://arxiv.org/abs/1611.07966)] [[INSPIRE](#)].

- [8] L. Senatore and M. Zaldarriaga, *The effective field theory of large-scale structure in the presence of massive neutrinos*, [arXiv:1707.04698](https://arxiv.org/abs/1707.04698) [INSPIRE].
- [9] A. Aviles and A. Banerjee, *A lagrangian perturbation theory in the presence of massive neutrinos*, *JCAP* **10** (2020) 034 [[arXiv:2007.06508](https://arxiv.org/abs/2007.06508)] [INSPIRE].
- [10] A. Aviles et al., *Redshift space power spectrum beyond Einstein-de Sitter kernels*, *JCAP* **04** (2021) 039 [[arXiv:2012.05077](https://arxiv.org/abs/2012.05077)] [INSPIRE].
- [11] S.-F. Chen, Z. Vlah and M. White, *A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO*, *JCAP* **02** (2022) 008 [[arXiv:2110.05530](https://arxiv.org/abs/2110.05530)] [INSPIRE].
- [12] S. Foreman, H. Perrier and L. Senatore, *Precision comparison of the power spectrum in the EFTofLSS with simulations*, *JCAP* **05** (2016) 027 [[arXiv:1507.05326](https://arxiv.org/abs/1507.05326)] [INSPIRE].
- [13] T. Nishimichi et al., *Blinded challenge for precision cosmology with large-scale structure: results from effective field theory for the redshift-space galaxy power spectrum*, *Phys. Rev. D* **102** (2020) 123541 [[arXiv:2003.08277](https://arxiv.org/abs/2003.08277)] [INSPIRE].
- [14] V. Springel, R. Pakmor, O. Zier and M. Reinecke, *Simulating cosmic structure formation with the gadget-4 code*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 2871 [[arXiv:2010.03567](https://arxiv.org/abs/2010.03567)] [INSPIRE].
- [15] L.H. Garrison et al., *The abacus cosmological N-body code*, *Mon. Not. Roy. Astron. Soc.* **508** (2021) 575 [[arXiv:2110.11392](https://arxiv.org/abs/2110.11392)] [INSPIRE].
- [16] D. Potter, J. Stadel and R. Teyssier, *PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys*, [arXiv:1609.08621](https://arxiv.org/abs/1609.08621) [INSPIRE].
- [17] S. Habib et al., *HACC: simulating sky surveys on state-of-the-art supercomputing architectures*, *New Astron.* **42** (2016) 49 [[arXiv:1410.2805](https://arxiv.org/abs/1410.2805)] [INSPIRE].
- [18] J. Brandbyge and S. Hannestad, *Grid based linear neutrino perturbations in cosmological N-body simulations*, *JCAP* **05** (2009) 002 [[arXiv:0812.3149](https://arxiv.org/abs/0812.3149)] [INSPIRE].
- [19] Y. Ali-Haïmoud and S. Bird, *An efficient implementation of massive neutrinos in non-linear structure formation simulations*, *Mon. Not. Roy. Astron. Soc.* **428** (2012) 3375 [[arXiv:1209.0461](https://arxiv.org/abs/1209.0461)] [INSPIRE].
- [20] E. Castorina et al., *DEMNUni: the clustering of large-scale structures in the presence of massive neutrinos*, *JCAP* **07** (2015) 043 [[arXiv:1505.07148](https://arxiv.org/abs/1505.07148)] [INSPIRE].
- [21] A. Upadhye et al., *Redshift-space distortions in massive neutrino and evolving dark energy cosmologies*, *Phys. Rev. D* **93** (2016) 063515 [[arXiv:1506.07526](https://arxiv.org/abs/1506.07526)] [INSPIRE].
- [22] EUCLID collaboration, *Euclid: modelling massive neutrinos in cosmology — a code comparison*, *JCAP* **06** (2023) 035 [[arXiv:2211.12457](https://arxiv.org/abs/2211.12457)] [INSPIRE].
- [23] M. Viel, M.G. Haehnelt and V. Springel, *The effect of neutrinos on the matter distribution as probed by the intergalactic medium*, *JCAP* **06** (2010) 015 [[arXiv:1003.2422](https://arxiv.org/abs/1003.2422)] [INSPIRE].
- [24] A. Banerjee and N. Dalal, *Simulating nonlinear cosmological structure formation with massive neutrinos*, *JCAP* **11** (2016) 015 [[arXiv:1606.06167](https://arxiv.org/abs/1606.06167)] [INSPIRE].
- [25] S. Bird, Y. Ali-Haïmoud, Y. Feng and J. Liu, *An efficient and accurate hybrid method for simulating non-linear neutrino structure*, *Mon. Not. Roy. Astron. Soc.* **481** (2018) 1486 [[arXiv:1803.09854](https://arxiv.org/abs/1803.09854)] [INSPIRE].
- [26] J.M. Sullivan, J.D. Emberson, S. Habib and N. Frontiere, *Improving initialization and evolution accuracy of cosmological neutrino simulations*, *JCAP* **06** (2023) 003 [[arXiv:2302.09134](https://arxiv.org/abs/2302.09134)] [INSPIRE].
- [27] A. Banerjee, D. Powell, T. Abel and F. Villaescusa-Navarro, *Reducing noise in cosmological N-body simulations with neutrinos*, *JCAP* **09** (2018) 028 [[arXiv:1801.03906](https://arxiv.org/abs/1801.03906)] [INSPIRE].

[28] A.E. Bayer, A. Banerjee and Y. Feng, *A fast particle-mesh simulation of non-linear cosmological structure formation with massive neutrinos*, *JCAP* **01** (2021) 016 [[arXiv:2007.13394](https://arxiv.org/abs/2007.13394)] [[INSPIRE](#)].

[29] U. Seljak, *Analytic model for galaxy and dark matter clustering*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 203 [[astro-ph/0001493](https://arxiv.org/abs/astro-ph/0001493)] [[INSPIRE](#)].

[30] A.A. Berlind and D.H. Weinberg, *The halo occupation distribution: towards an empirical determination of the relation between galaxies and mass*, *Astrophys. J.* **575** (2002) 587 [[astro-ph/0109001](https://arxiv.org/abs/astro-ph/0109001)] [[INSPIRE](#)].

[31] J.S. Bullock, R.H. Wechsler and R.S. Somerville, *Galaxy halo occupation at high redshift*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 246 [[astro-ph/0106293](https://arxiv.org/abs/astro-ph/0106293)] [[INSPIRE](#)].

[32] B.A. Reid et al., *A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 476 [[arXiv:1404.3742](https://arxiv.org/abs/1404.3742)] [[INSPIRE](#)].

[33] J.U. Lange et al., *Five percent measurements of the growth rate from simulation-based modelling of redshift-space clustering in BOSS LOWZ*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 1779 [[arXiv:2101.12261](https://arxiv.org/abs/2101.12261)] [[INSPIRE](#)].

[34] Z. Zhai et al., *The Aemulus project. V. Cosmological constraint from small-scale clustering of BOSS galaxies*, *Astrophys. J.* **948** (2023) 99 [[arXiv:2203.08999](https://arxiv.org/abs/2203.08999)] [[INSPIRE](#)].

[35] S. Yuan et al., *AbacusHOD: a highly efficient extended multitracer HOD framework and its application to BOSS and eBOSS data*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3301 [[arXiv:2110.11412](https://arxiv.org/abs/2110.11412)] [[INSPIRE](#)].

[36] B.D. Wibking et al., *Cosmology with galaxy-galaxy lensing on non-perturbative scales: emulation method and application to BOSS LOWZ*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 2872 [[arXiv:1907.06293](https://arxiv.org/abs/1907.06293)] [[INSPIRE](#)].

[37] H. Miyatake et al., *Cosmological inference from an emulator based halo model. II. Joint analysis of galaxy-galaxy weak lensing and galaxy clustering from HSC-Y1 and SDSS*, *Phys. Rev. D* **106** (2022) 083520 [[arXiv:2111.02419](https://arxiv.org/abs/2111.02419)] [[INSPIRE](#)].

[38] K. Storey-Fisher et al., *The Aemulus project VI: emulation of beyond-standard galaxy clustering statistics to improve cosmological constraints*, [arXiv:2210.03203](https://arxiv.org/abs/2210.03203) [[INSPIRE](#)].

[39] G. Valogiannis and C. Dvorkin, *Going beyond the galaxy power spectrum: an analysis of BOSS data with wavelet scattering transforms*, *Phys. Rev. D* **106** (2022) 103509 [[arXiv:2204.13717](https://arxiv.org/abs/2204.13717)] [[INSPIRE](#)].

[40] R. García and E. Rozo, *Halo exclusion criteria impacts halo statistics*, *Mon. Not. Roy. Astron. Soc.* **489** (2019) 4170 [[arXiv:1903.01709](https://arxiv.org/abs/1903.01709)] [[INSPIRE](#)].

[41] J.L. Tinker et al., *Toward a halo mass function for precision cosmology: the limits of universality*, *Astrophys. J.* **688** (2008) 709 [[arXiv:0803.2706](https://arxiv.org/abs/0803.2706)] [[INSPIRE](#)].

[42] B. Dai, Y. Feng, U. Seljak and S. Singh, *High mass and halo resolution from fast low resolution simulations*, *JCAP* **04** (2020) 002 [[arXiv:1908.05276](https://arxiv.org/abs/1908.05276)] [[INSPIRE](#)].

[43] A.S. Villarreal et al., *The immitigable nature of assembly bias: the impact of halo definition on assembly bias*, *Mon. Not. Roy. Astron. Soc.* **472** (2017) 1088 [[arXiv:1705.04327](https://arxiv.org/abs/1705.04327)] [[INSPIRE](#)].

[44] P. Mansfield and A.V. Kravtsov, *The three causes of low-mass assembly bias*, *Mon. Not. Roy. Astron. Soc.* **493** (2020) 4763 [[arXiv:1902.00030](https://arxiv.org/abs/1902.00030)] [[INSPIRE](#)].

[45] D. Nelson et al., *The IllustrisTNG simulations: public data release*, [arXiv:1812.05609](https://arxiv.org/abs/1812.05609) [[INSPIRE](#)].

[46] J. Schaye et al., *The EAGLE project: simulating the evolution and assembly of galaxies and their environments*, *Mon. Not. Roy. Astron. Soc.* **446** (2015) 521 [[arXiv:1407.7040](https://arxiv.org/abs/1407.7040)] [[INSPIRE](#)].

[47] I.G. McCarthy, J. Schaye, S. Bird and A.M.C. Le Brun, *The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2936 [[arXiv:1603.02702](https://arxiv.org/abs/1603.02702)] [[INSPIRE](#)].

[48] P.F. Hopkins et al., *FIRE-2 simulations: physics versus numerics in galaxy formation*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 800 [[arXiv:1702.06148](https://arxiv.org/abs/1702.06148)] [[INSPIRE](#)].

[49] C. Modi, S.-F. Chen and M. White, *Simulations and symmetries*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 5754 [[arXiv:1910.07097](https://arxiv.org/abs/1910.07097)] [[INSPIRE](#)].

[50] A. Banerjee, N. Kokron and T. Abel, *Modelling nearest neighbour distributions of biased tracers using hybrid effective field theory*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2765 [[arXiv:2107.10287](https://arxiv.org/abs/2107.10287)] [[INSPIRE](#)].

[51] A. Banerjee and T. Abel, *Nearest neighbour distributions: new statistical measures for cosmological clustering*, *Mon. Not. Roy. Astron. Soc.* **500** (2020) 5479 [[arXiv:2007.13342](https://arxiv.org/abs/2007.13342)] [[INSPIRE](#)].

[52] A. Banerjee and T. Abel, *Cosmological cross-correlations and nearest neighbour distributions*, *Mon. Not. Roy. Astron. Soc.* **504** (2021) 2911 [[arXiv:2102.01184](https://arxiv.org/abs/2102.01184)] [[INSPIRE](#)].

[53] K. Heitmann et al., *The Mira-Titan universe: precision predictions for dark energy surveys*, *Astrophys. J.* **820** (2016) 108 [[arXiv:1508.02654](https://arxiv.org/abs/1508.02654)] [[INSPIRE](#)].

[54] EUCLID collaboration, *Euclid preparation: IX. EuclidEmulator2 — power spectrum emulation with massive neutrinos and self-consistent dark energy perturbations*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 2840 [[arXiv:2010.11288](https://arxiv.org/abs/2010.11288)] [[INSPIRE](#)].

[55] K.R. Moran et al., *The Mira-Titan universe — IV. High-precision power spectrum emulation*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 3443 [[arXiv:2207.12345](https://arxiv.org/abs/2207.12345)] [[INSPIRE](#)].

[56] T. McClintock et al., *The Aemulus project II: emulating the halo mass function*, *Astrophys. J.* **872** (2019) 53 [[arXiv:1804.05866](https://arxiv.org/abs/1804.05866)] [[INSPIRE](#)].

[57] S. Bocquet et al., *The Mira-Titan universe. III. Emulation of the halo mass function*, *Astrophys. J.* **901** (2020) 5 [[arXiv:2003.12116](https://arxiv.org/abs/2003.12116)] [[INSPIRE](#)].

[58] T. McClintock et al., *The Aemulus project IV: emulating halo bias*, [arXiv:1907.13167](https://arxiv.org/abs/1907.13167) [[INSPIRE](#)].

[59] T. Nishimichi et al., *Dark quest. I. Fast and accurate emulation of halo clustering statistics and its application to galaxy clustering*, *Astrophys. J.* **884** (2019) 29 [[arXiv:1811.09504](https://arxiv.org/abs/1811.09504)] [[INSPIRE](#)].

[60] B.D. Wibking et al., *Emulating galaxy clustering and galaxy-galaxy lensing into the deeply non-linear regime: methodology, information, and forecasts*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 989 [[arXiv:1709.07099](https://arxiv.org/abs/1709.07099)] [[INSPIRE](#)].

[61] A.N. Salcedo et al., *Spatial clustering of dark matter haloes: secondary bias, neighbour bias, and the influence of massive neighbours on halo properties*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 4411 [[arXiv:1708.08451](https://arxiv.org/abs/1708.08451)] [[INSPIRE](#)].

[62] Z. Zhai et al., *The Aemulus project III: emulation of the galaxy correlation function*, *Astrophys. J.* **874** (2019) 95 [[arXiv:1804.05867](https://arxiv.org/abs/1804.05867)] [[INSPIRE](#)].

[63] N. Kokron et al., *The cosmology dependence of galaxy clustering and lensing from a hybrid N-body — perturbation theory model*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 1422 [[arXiv:2101.11014](https://arxiv.org/abs/2101.11014)] [[INSPIRE](#)].

- [64] M. Zennaro et al., *The BACCO simulation project: biased tracers in real space*, [arXiv:2101.12187](https://arxiv.org/abs/2101.12187) [INSPIRE].
- [65] M. Pellejero-Ibanez et al., *Modelling galaxy clustering in redshift space with a Lagrangian bias formalism and N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 3993 [[arXiv:2109.08699](https://arxiv.org/abs/2109.08699)] [INSPIRE].
- [66] M. Pellejero-Ibanez et al., *The Bacco simulation project: Bacco hybrid Lagrangian bias expansion model in redshift space*, *Mon. Not. Roy. Astron. Soc.* **520** (2023) 3725 [[arXiv:2207.06437](https://arxiv.org/abs/2207.06437)] [INSPIRE].
- [67] B. Hadzhiyska et al., *Hefty enhancement of cosmological constraints from the DES Y1 data using a hybrid effective field theory approach to galaxy bias*, *JCAP* **09** (2021) 020 [[arXiv:2103.09820](https://arxiv.org/abs/2103.09820)] [INSPIRE].
- [68] J.A. Peacock and S.J. Dodds, *Nonlinear evolution of cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **280** (1996) L19 [[astro-ph/9603031](https://arxiv.org/abs/astro-ph/9603031)] [INSPIRE].
- [69] R.E. Angulo and A. Pontzen, *Cosmological N-body simulations with suppressed variance*, *Mon. Not. Roy. Astron. Soc.* **462** (2016) L1 [[arXiv:1603.05253](https://arxiv.org/abs/1603.05253)] [INSPIRE].
- [70] A. Pontzen, A. Slosar, N. Roth and H.V. Peiris, *Inverted initial conditions: exploring the growth of cosmic structure and voids*, *Phys. Rev. D* **93** (2016) 103519 [[arXiv:1511.04090](https://arxiv.org/abs/1511.04090)] [INSPIRE].
- [71] EUCLID collaboration, *Euclid preparation: II. The EuclidEmulator — a tool to compute the cosmology dependence of the nonlinear matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 5509 [[arXiv:1809.04695](https://arxiv.org/abs/1809.04695)] [INSPIRE].
- [72] R.E. Angulo et al., *The BACCO simulation project: exploiting the full power of large-scale structure for cosmology*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 5869 [[arXiv:2004.06245](https://arxiv.org/abs/2004.06245)] [INSPIRE].
- [73] F. Villaescusa-Navarro et al., *Statistical properties of paired fixed fields*, *Astrophys. J.* **867** (2018) 137 [[arXiv:1806.01871](https://arxiv.org/abs/1806.01871)] [INSPIRE].
- [74] C.-H. Chuang et al., *UNIT project: universe N-body simulations for the investigation of theoretical models from galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **487** (2019) 48 [[arXiv:1811.02111](https://arxiv.org/abs/1811.02111)] [INSPIRE].
- [75] F. Maion, R.E. Angulo and M. Zennaro, *Statistics of biased tracers in variance-suppressed simulations*, *JCAP* **10** (2022) 036 [[arXiv:2204.03868](https://arxiv.org/abs/2204.03868)] [INSPIRE].
- [76] A.B. Owen, *Monte Carlo theory, methods and examples*, <https://artowen.su.domains/mc/> (2013).
- [77] N. Chartier, B. Wandelt, Y. Akrami and F. Villaescusa-Navarro, *CARPool: fast, accurate computation of large-scale structure statistics by pairing costly and cheap cosmological simulations*, *Mon. Not. Roy. Astron. Soc.* **503** (2021) 1897 [[arXiv:2009.08970](https://arxiv.org/abs/2009.08970)] [INSPIRE].
- [78] N. Chartier and B.D. Wandelt, *CARPool covariance: fast, unbiased covariance estimation for large-scale structure observables*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 2220 [[arXiv:2106.11718](https://arxiv.org/abs/2106.11718)] [INSPIRE].
- [79] N. Chartier and B.D. Wandelt, *Bayesian control variates for optimal covariance estimation with pairs of simulations and surrogates*, *Mon. Not. Roy. Astron. Soc.* **515** (2022) 1296 [[arXiv:2204.03070](https://arxiv.org/abs/2204.03070)] [INSPIRE].
- [80] S. Tassev, D.J. Eisenstein, B.D. Wandelt and M. Zaldarriaga, *sCOLA: the N-body COLA method extended to the spatial domain*, [arXiv:1502.07751](https://arxiv.org/abs/1502.07751) [INSPIRE].

[81] Y. Feng, M.-Y. Chu, U. Seljak and P. McDonald, *FastPM: a new scheme for fast simulations of dark matter and haloes*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 2273 [[arXiv:1603.00476](https://arxiv.org/abs/1603.00476)] [[INSPIRE](#)].

[82] N. Kokron et al., *Accurate predictions from small boxes: variance suppression via the Zel'dovich approximation*, *JCAP* **09** (2022) 059 [[arXiv:2205.15327](https://arxiv.org/abs/2205.15327)] [[INSPIRE](#)].

[83] J. DeRose, S.-F. Chen, N. Kokron and M. White, *Precision redshift-space galaxy power spectra using Zel'dovich control variates*, *JCAP* **02** (2023) 008 [[arXiv:2210.14239](https://arxiv.org/abs/2210.14239)] [[INSPIRE](#)].

[84] M. White et al., *Cosmological constraints from the tomographic cross-correlation of DESI luminous red galaxies and Planck CMB lensing*, *JCAP* **02** (2022) 007 [[arXiv:2111.09898](https://arxiv.org/abs/2111.09898)] [[INSPIRE](#)].

[85] S.-F. Chen, M. White, J. DeRose and N. Kokron, *Cosmological analysis of three-dimensional BOSS galaxy clustering and Planck CMB lensing cross correlations via Lagrangian perturbation theory*, *JCAP* **07** (2022) 041 [[arXiv:2204.10392](https://arxiv.org/abs/2204.10392)] [[INSPIRE](#)].

[86] C. Heymans et al., *KiDS-1000 cosmology: multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, *Astron. Astrophys.* **646** (2021) A140 [[arXiv:2007.15632](https://arxiv.org/abs/2007.15632)] [[INSPIRE](#)].

[87] DES collaboration, *Dark Energy Survey year 3 results: cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **105** (2022) 023520 [[arXiv:2105.13549](https://arxiv.org/abs/2105.13549)] [[INSPIRE](#)].

[88] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](https://arxiv.org/abs/1807.06209)] [[INSPIRE](#)].

[89] LSST collaboration, *LSST: from science drivers to reference design and anticipated data products*, *Astrophys. J.* **873** (2019) 111 [[arXiv:0805.2366](https://arxiv.org/abs/0805.2366)] [[INSPIRE](#)].

[90] SIMONS OBSERVATORY collaboration, *The Simons Observatory: Astro2020 decadal project whitepaper*, *Bull. Am. Astron. Soc.* **51** (2019) 147 [[arXiv:1907.08284](https://arxiv.org/abs/1907.08284)] [[INSPIRE](#)].

[91] N. Kokron et al., *Priors on red galaxy stochasticity from hybrid effective field theory*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 2198 [[arXiv:2112.00012](https://arxiv.org/abs/2112.00012)] [[INSPIRE](#)].

[92] DESI collaboration, *Target selection and validation of DESI luminous red galaxies*, *Astron. J.* **165** (2023) 58 [[arXiv:2208.08515](https://arxiv.org/abs/2208.08515)] [[INSPIRE](#)].

[93] LSST DARK ENERGY SCIENCE collaboration, *The LSST Dark Energy Science Collaboration (DESC) science requirements document*, [arXiv:1809.01669](https://arxiv.org/abs/1809.01669) [[INSPIRE](#)].

[94] N. Sailer, E. Castorina, S. Ferraro and M. White, *Cosmology at high redshift — a probe of fundamental physics*, *JCAP* **12** (2021) 049 [[arXiv:2106.09713](https://arxiv.org/abs/2106.09713)] [[INSPIRE](#)].

[95] A.J. Mead et al., *A hydrodynamical halo model for weak-lensing cross correlations*, *Astron. Astrophys.* **641** (2020) A130 [[arXiv:2005.00009](https://arxiv.org/abs/2005.00009)] [[INSPIRE](#)].

[96] EBOSS collaboration, *Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, *Phys. Rev. D* **103** (2021) 083533 [[arXiv:2007.08991](https://arxiv.org/abs/2007.08991)] [[INSPIRE](#)].

[97] KATRIN collaboration, *Direct neutrino-mass measurement with sub-electronvolt sensitivity*, *Nature Phys.* **18** (2022) 160 [[arXiv:2105.08533](https://arxiv.org/abs/2105.08533)] [[INSPIRE](#)].

[98] SUPER-KAMIOKANDE collaboration, *Search for neutrinos in Super-Kamiokande associated with the GW170817 neutron-star merger*, *Astrophys. J. Lett.* **857** (2018) L4 [[arXiv:1802.04379](https://arxiv.org/abs/1802.04379)] [[INSPIRE](#)].

[99] P.F. de Salas, S. Gariazzo, J. Lesgourges and S. Pastor, *Calculation of the local density of relic neutrinos*, *JCAP* **09** (2017) 034 [[arXiv:1706.09850](https://arxiv.org/abs/1706.09850)] [[INSPIRE](#)].

[100] W.J. Morokoff and R.E. Caflisch, *Quasi-random sequences and their discrepancies*, *SIAM J. Sci. Comput.* **15** (1994) 1251.

- [101] J. DeRose et al., *The Aemulus project I: numerical simulations for precision cosmology*, *Astrophys. J.* **875** (2019) 69 [[arXiv:1804.05865](https://arxiv.org/abs/1804.05865)] [[INSPIRE](#)].
- [102] L.H. Garrison et al., *Improving initial conditions for cosmological N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **461** (2016) 4125 [[arXiv:1605.02333](https://arxiv.org/abs/1605.02333)] [[INSPIRE](#)].
- [103] M. Michaux, O. Hahn, C. Rampf and R.E. Angulo, *Accurate initial conditions for cosmological N-body simulations: minimizing truncation and discreteness errors*, *Mon. Not. Roy. Astron. Soc.* **500** (2020) 663 [[arXiv:2008.09588](https://arxiv.org/abs/2008.09588)] [[INSPIRE](#)].
- [104] B. Marcos et al., *Linear perturbative theory of the discrete cosmological N-body problem*, *Phys. Rev. D* **73** (2006) 103507 [[astro-ph/0601479](https://arxiv.org/abs/astro-ph/0601479)] [[INSPIRE](#)].
- [105] W. Elbers et al., *An optimal non-linear method for simulating relic neutrinos*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 2614 [[arXiv:2010.07321](https://arxiv.org/abs/2010.07321)] [[INSPIRE](#)].
- [106] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: overview*, [arXiv:1104.2932](https://arxiv.org/abs/1104.2932) [[INSPIRE](#)].
- [107] M. Zennaro et al., *Initial conditions for accurate N-body simulations of massive neutrino cosmologies*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 3244 [[arXiv:1605.05283](https://arxiv.org/abs/1605.05283)] [[INSPIRE](#)].
- [108] W. Elbers et al., *Higher order initial conditions with massive neutrinos*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 3821 [[arXiv:2202.00670](https://arxiv.org/abs/2202.00670)] [[INSPIRE](#)].
- [109] M. White, *The Zel'dovich approximation*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 3630 [[arXiv:1401.5466](https://arxiv.org/abs/1401.5466)] [[INSPIRE](#)].
- [110] R. Takahashi, *Third order density perturbation and one-loop power spectrum in a dark energy dominated universe*, *Prog. Theor. Phys.* **120** (2008) 549 [[arXiv:0806.1437](https://arxiv.org/abs/0806.1437)] [[INSPIRE](#)].
- [111] M. Fasiello and Z. Vlah, *Nonlinear fields in generalized cosmologies*, *Phys. Rev. D* **94** (2016) 063516 [[arXiv:1604.04612](https://arxiv.org/abs/1604.04612)] [[INSPIRE](#)].
- [112] Y. Donath and L. Senatore, *Biased tracers in redshift space in the EFTofLSS with exact time dependence*, *JCAP* **10** (2020) 039 [[arXiv:2005.04805](https://arxiv.org/abs/2005.04805)] [[INSPIRE](#)].
- [113] E. Sefusatti, M. Crocce, R. Scoccimarro and H. Couchman, *Accurate estimators of correlation functions in Fourier space*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 3624 [[arXiv:1512.07295](https://arxiv.org/abs/1512.07295)] [[INSPIRE](#)].
- [114] C. Howlett, M. Manera and W.J. Percival, *L-PICOLA: a parallel code for fast dark matter simulation*, *Astron. Comput.* **12** (2015) 109 [[arXiv:1506.03737](https://arxiv.org/abs/1506.03737)] [[INSPIRE](#)].
- [115] Y.B. Zel'Dovich, *Gravitational instability: an approximate theory for large density perturbations*, *Astron. Astrophys.* **500** (1970) 13.
- [116] Z. Vlah, E. Castorina and M. White, *The Gaussian streaming model and convolution Lagrangian effective field theory*, *JCAP* **12** (2016) 007 [[arXiv:1609.02908](https://arxiv.org/abs/1609.02908)] [[INSPIRE](#)].
- [117] S.-F. Chen, Z. Vlah and M. White, *Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory*, *JCAP* **07** (2020) 062 [[arXiv:2005.00523](https://arxiv.org/abs/2005.00523)] [[INSPIRE](#)].
- [118] D. Xiu, *Numerical methods for stochastic computations: a spectral method approach*, Princeton University Press, Princeton, NJ, U.S.A. (2010).
- [119] W. Gautschi, *Orthogonal polynomials — constructive theory and applications*, *J. Comput. Appl. Math.* **12-13** (1985) 61.
- [120] J. Feinberg and H.P. Langtangen, *Chaospy: an open source tool for designing methods of uncertainty quantification*, *J. Comput. Sci.* **11** (2015) 46.

- [121] F. van den Bosch et al., *Cosmological constraints from a combination of galaxy clustering and lensing — I. Theoretical framework*, *Mon. Not. Roy. Astron. Soc.* **430** (2013) 725 [[arXiv:1206.6890](https://arxiv.org/abs/1206.6890)] [[INSPIRE](#)].
- [122] R. Jimenez, T. Kitching, C. Pena-Garay and L. Verde, *Can we measure the neutrino mass hierarchy in the sky?*, *JCAP* **05** (2010) 035 [[arXiv:1003.5918](https://arxiv.org/abs/1003.5918)] [[INSPIRE](#)].
- [123] J. Lesgourgues and S. Pastor, *Neutrino mass from cosmology*, *Adv. High Energy Phys.* **2012** (2012) 608515 [[arXiv:1212.6154](https://arxiv.org/abs/1212.6154)] [[INSPIRE](#)].
- [124] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [[arXiv:2006.10256](https://arxiv.org/abs/2006.10256)] [[INSPIRE](#)].
- [125] P. Virtanen et al., *SciPy 1.0 — fundamental algorithms for scientific computing in python*, *Nature Meth.* **17** (2020) 261 [[arXiv:1907.10121](https://arxiv.org/abs/1907.10121)] [[INSPIRE](#)].
- [126] J.D. Hunter, *Matplotlib: a 2D graphics environment*, *Comput. Sci. Eng.* **9** (2007) 90 [[INSPIRE](#)].
- [127] J. Towns et al., *XSEDE: accelerating scientific discovery*, *Comput. Sci. Eng.* **16** (2014) 62 [[INSPIRE](#)].