UC Berkeley
UC Berkeley Previously Published Works

Title

The DESI N-body Simulation Project — II. Suppressing sample variance with fast simulations

Permalink

Journal

ISSN

Authors

Ding, Zhejie
Chuang, Chia-Hsun
Yu, Yu
et al.

Publication Date

DOI

The DESI *N*-body Simulation Project – II. Suppressing sample variance with fast simulations

Zhejie Ding,[1,2]★ Chia-Hsun Chuang,[3,4]† Yu Yu,[1,2]‡ Lehman H. Garrison,[5] Adrian E. Bayer,[6,7] Yu Feng,[7] Chirag Modi,[8,5] Daniel J. Eisenstein,[9] Martin White,[6,7,10,11] Andrei Variu,[12] Cheng Zhao,[12] Hanyu Zhang,[13] Jennifer Meneses Rizo,[14] David Brooks,[15] Kyle Dawson,[4] Peter Doel,[15] Enrique Gaztanaga,[16,17] Robert Kehoe,[18] Alex Krolewski,[19,20] Martin Landriau,[11] Nathalie Palanque-Delabrouille,[11,21] Claire Poppett[22]

[1]*Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, China*
[2]*Key Laboratory for Particle Astrophysics and Cosmology(MOE)/Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai, 200240, China*
[3]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[4]*Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA*
[5]*Center for Computational Astrophysics, Flatiron Institute Simons Foundation, 162 Fifth Ave. New York, NY 10010, USA*
[6]*Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA*
[7]*Department of Physics, University of California, Berkeley, CA 94720, USA*
[8]*Center for Computational Mathematics, Flatiron Institute, New York, NY 10010, USA*
[9]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA*
[10]*Department of Astronomy, University of California, Berkeley, CA 94720, USA*
[11]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[12]*Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[13]*Department of Physics, Kansas State University, Manhattan, KS 66506, USA*
[14]*Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, Ciudad de México, México*
[15]*Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK*
[16]*Institute of Space Sciences (ICE, CSIC), 08193 Barcelona, Spain*
[17]*Institut d´ Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain*
[18]*Department of Physics, Southern Methodist University, Dallas, TX 75275, USA*
[19]*AMTD Fellow, Waterloo Centre for Astrophysics, University of Waterloo, Waterloo ON N2L 3G1, Canada*
[20]*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON NL2 2Y5, Canada*
[21]*IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, F-91191, France*
[22]*Space Sciences Laboratory (SSL), UC Berkeley, 7 Gauss Way, Berkeley, CA 94720, USA*

ABSTRACT

Dark Energy Spectroscopic Instrument (DESI) will construct a large and precise three-dimensional map of our Universe. The survey effective volume reaches $\sim 20\,h^{-3}\mathrm{Gpc}^3$. It is a great challenge to prepare high-resolution simulations with a much larger volume for validating the DESI analysis pipelines. AbacusSummit is a suite of high-resolution dark-matter-only simulations designed for this purpose, with $200\,h^{-3}\mathrm{Gpc}^3$ (10 times DESI volume) for the base cosmology. However, further efforts need to be done to provide a more precise analysis of the data and to cover also other cosmologies. Recently, the CARPool method was proposed to use paired accurate and approximate simulations to achieve high statistical precision with a limited number of high-resolution simulations. Relying on this technique, we propose to use fast quasi-*N*-body solvers combined with accurate simulations to produce accurate summary statistics. This enables us to obtain 100 times smaller variance than the expected DESI statistical variance at the scales we are interested in, e.g. $k < 0.3\,h\,\mathrm{Mpc}^{-1}$ for the halo power spectrum. In addition, it can significantly suppress the sample variance of the halo bispectrum. We further generalize the method for other cosmologies with only one realization in AbacusSummit suite to extend the effective volume ~ 20 times. In summary, our proposed strategy of combining high-fidelity simulations with fast approximate gravity solvers and a series of variance suppression techniques sets the path for a robust cosmological analysis of galaxy survey data.

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

1 INTRODUCTION

★ E-mail: zhejied@sjtu.edu.cn
† E-mail: albert.chuang@utah.edu
‡ E-mail: yuyu22@sjtu.edu.cn

Dark Energy Spectroscopic Instrument (DESI) will collect more than 30 million galaxy spectra within 5 yr (DESI Collaboration et al.

[2016](#)), constructing four main target catalogues, including bright galaxy sample, luminous red galaxies (LRGs), emission line galaxies (ELGs), and quasi-stellar objects. We will use them to determine the distances of the galaxies and construct a three-dimensional map of the large-scale structure of our Universe. A wealth of valuable information about our Universe is hidden in the statistics of the matter distribution, for example: the nature of dark energy (Copeland et al. 2006), modified theories of gravity (Huterer et al. 2015; Alam et al. 2021), and neutrino mass (Lesgourgues & Pastor 2006; Font-Ribera et al. 2014; Allison et al. 2015; Hahn & Villaescusa-Navarro 2021; Kreisch et al. 2021; Massara et al. 2021; Bayer et al. 2021a,b). To extract the cosmological information from observation, we need to build theoretical models that allow comparison to data. While the non-linear effect is difficult to model analytically, one alternative approach is running N-body simulations and thus building the models numerically. To this end, we want to generate simulations with huge volumes and high resolutions that are unfortunately limited by computing resources.

Grove et al. (2021) (the first paper of the DESI N-body simulation projects) compared multiple N-body codes and found good agreement which showed their robustness. Also, they compared the simulations with different mass resolutions and identified that the errors introduced by simulations with particle mass of $2.11 \times 10^9 \, h^{-1}M_\odot$ are well below the DESI statistical uncertainties. However, it is a challenge to generate simulations with sufficient volume for DESI with such a mass resolution. The physical survey volume of DESI is about $60 \, h^{-3}\mathrm{Gpc}^3$ but the effective volume is about $20 \, h^{-3}\mathrm{Gpc}^3$ if we take into account the observed galaxy densities at different redshifts (DESI Collaboration et al. 2016). Thus, for the given mass resolution and volume, the simulation will need more than 800 billion particles. If we want to have the theoretical uncertainty below 1/10 of DESI statistical error, the simulation volume should be 100 times larger, thus requiring 80 trillion particles. While it is not an impossible mission, we do not consider that it is a practical approach since it will cost a tremendous amount of computing resources.

The initial conditions (ICs) of cosmological simulations are constructed based on the Gaussian realizations that naturally introduce noise into the simulations. Although the introduced noises are physically motivated (i.e. due to inflation), we want to minimize them to provide a noiseless theoretical prediction to compare with observed data. Otherwise, we will waste the constraining power from the observation. A brutal way to reduce the noise would be generating simulations with larger volumes until the noise is well below the statistical error of the survey. However, this strategy will not be practical when the survey volume increases dramatically as DESI. Some techniques have been developed to reduce the noise without running massive volume simulations. Angulo & Pontzen (2016) proposed the fixed-amplitude technique to suppress the variance at large scales by modifying the ICs of simulations. In addition, it can further suppress sample variance using pairs of the fixed-amplitude simulations with initial phases differed by π rad (Pontzen et al. 2016). The so-called paired-and-fixed method has been studied by a series of work (e.g. Villaescusa-Navarro et al. 2018; Chuang et al. 2019; Klypin et al. 2020; Avila & Gutierrez Adame 2022; Maion et al. 2022).

Recently, the CARPool method, proposed by Chartier et al. (2021), takes a different approach. It reduces the noise by learning the calibrations from a large set of quasi-N-body simulations. While the fixed-amplitude method has only small improvement or no improvement in the precision at small scales, e.g. $k > 0.2 \, h\,\mathrm{Mpc}^{-1}$, the CARPool method still has a significant gain at even smaller scales.

In this work, we apply the CARPool method to the AbacusSummit simulations, an extensive simulation suite generated on the Summit

supercomputer[1]. To do so, we prepare a set of FastPM simulations matching the configuration, including the ICs of AbacusSummit simulations. While the CARPool method has been validated for the statistics of dark matter particles (Chartier et al. 2021), we focus on dark matter haloes that are expected to host DESI-like galaxy samples, e.g. ELGs (Gonzalez-Perez et al. 2018; Avila et al. 2020) and LRGs (Hernández-Aguayo et al. 2021; Zhou et al. 2021). The AbacusSummit suite includes various cosmology models, but only the base cosmology has the largest volume which is $200 \, h^{-3}\mathrm{Gpc}^3$ (i.e. 10 times DESI volume). In this work, we also extend the CARPool method to increase the effective volume of simulations other than the base cosmology as well. Our work paves the way for providing the most precise and accurate galaxy clustering predictions based on N-body simulations for DESI or future surveys.

This paper is organized as follows. In Section 2 we describe the simulations used in this study. We briefly describe the CARPool method in Section 3. We show the CARPool performance for the halo two-point and three-point clustering statistics in Section 4. We show the results of extending the method to different cosmologies in Section 5. Finally, in Section 6, we present the conclusions and discussions.

2 SIMULATIONS

We describe the N-body simulations and quasi-N-body simulations we use or prepare for this study.

2.1 AbacusSummit Simulations

AbacusSummit[2] (Maksimova et al. 2021) is a suite of high-fidelity N-body simulations based on the Abacus N-body code (Metchnik 2009; Garrison et al. 2016, 2018, 2019, 2021). They were generated on the Summit supercomputer at the Oak Ridge Leadership Computing Facility for the scientific goals of DESI survey. AbacusSummit consists of simulations that span different cosmologies, box sizes and mass resolutions. The base cosmology, denoted as c000, is the flat Λ cold dark matter (ΛCDM) model constrained from Planck 2018 (Planck Collaboration et al. 2020). In c000, there are 25 base boxes, each of which is a $2 \, h^{-1}\mathrm{Gpc}$ box with 6912^3 particles and particle mass resolution $2.11 \times 10^9 \, h^{-1}M_\odot$. There are four secondary cosmologies (c001–c004) and each cosmology has six base boxes that share the ICs of the first six boxes (ph000–ph005) in c000. We show the parameters of cosmologies c000, c002, and c004 in Table 1.

AbacusSummit uses the highly efficient on-the-fly *Competitive Assignment to Spherical Overdensities* (CompaSO) group finder (Hadzhiyska et al. 2022) and outputs halo catalogues at 12 primary redshifts, $z = 3.0, 2.5, 2.0, 1.7, 1.4, 1.1, 0.8, 0.5, 0.4, 0.3, 0.2,$ and 0.1, as well as at 21 secondary redshifts. Due to the large amount of data, we only focus on the halo catalogues at $z = 1.1$ which is the median redshift where the primary targets (ELGs) of DESI will be observed ($z = 0.6$–1.6). We do not use the "cleaned" version of the CompaSO catalogues, which are not available when we started this work. The details of the cleaning method are described in Bose et al. (2022); Hadzhiyska et al. (2022). In summary, based on halo merger trees, the cleaning method can remove unphysical haloes identified by CompaSO. The misidentification is mainly caused by two sides. One

---

Table 1. Parameters of the AbacusSummit cosmologies. c000 is the flat ΛCDM based on Planck 2018; c002 and c004 are two of the secondary cosmologies with some parameters different from c000's. c002 is a thawing dark energy model with $w_0 = -0.7$ and $w_a = -0.5$. c004 has lower clustering amplitude, i.e. smaller A_s and σ_8 than those of c000. More details of the AbacusSummit cosmological models can be found in Maksimova et al. (2021).

Cosmology	$\Omega_b h^2$	$\Omega_{\rm cdm} h^2$	h	A_s	n_s	α_s	$N_{\rm ur}$	$N_{\rm ncdm}$	$\Omega_{\rm ncdm} h^2$	$w_{0,\,\rm fld}$	$w_{a,\,\rm fld}$	σ_{8_m}	$\sigma_{8_{c\cdot b}}$
c000	0.02237	0.1200	0.6736	2.0830e-9	0.9649	0.0	2.0328	1	0.00064420	-1.0	0.0	0.807952	0.811355
c002	0.02237	0.1200	0.6278	2.3140e-9	0.9649	0.0	2.0328	1	0.00064420	-0.7	-0.5	0.808189	0.811577
c004	0.02237	0.1200	0.6736	1.7949e-9	0.9649	0.0	2.0328	1	0.00064420	-1.0	0.0	0.749999	0.753159

is due to the halo dynamical processes such as fly-bys, partial mergers, and splits, and the other is from the strict spherical overdensity criterion that can overly deblend single haloes into two or more components. After cleaning, the number of haloes will be decreased by a few per cent and mainly for low-mass haloes ($M_{\rm halo} \sim 10^{11}\,h^{-1}{\rm M_\odot}$). We believe that whether using the cleaned or uncleaned version of the halo catalogues should not affect our main conclusions, though the cleaned version will cause some difference on halo correlation function or power spectrum around the scale of one-halo to two-halo transition and small difference on the overall halo bias. For example, the halo bias from the cleaned catalogue with $M_{\rm halo} \sim 10^{11.5}\,h^{-1}{\rm M_\odot}$ will be ~ 6 per cent lower (fig. 10 in Bose et al. 2022) compared with that of the uncleaned one. In this work, we analyse the AbacusSummit halo statistics with the assistance of the package abacusutils[3].

2.2 DESI-FastPM Simulation

We choose FastPM[4] (Feng et al. 2016) as the surrogate to pair with AbacusSummit. FastPM is a fast simulation method to approximate clustering from N-body solvers. It implements the particle mesh (PM) scheme (Quinn et al. 1997) with modified kick and drift factors to guarantee the accuracy of the linear displacement at large scales. The accuracy of FastPM is mainly determined by the particle mass resolution m_0, the initial redshift z_0, the number of time-steps T, and the force resolution which is parametrized as the ratio of the force mesh size over the number of particles along one axis of the simulation box, denoted as B.

2.2.1 Configuration of DESI-FastPM Simulations

Apart from the accuracy, we need to consider the computational cost since a large number of FastPM simulations are required to construct covariance matrices and to do cosmological analysis. Therefore, we first need to determine the configuration parameters to balance its accuracy and computational cost. In order to pair FastPM with AbacusSummit, we set the FastPM box size the same as the AbacusSummit base runs, i.e. $L_{\rm box} = 2\,h^{-1}{\rm Gpc}$. Given a FastPM simulation with some certain configuration parameters, we compare its matter power spectrum with that of AbacusSummit at redshift 0.2. We set the number of particles in FastPM as $N_p = 5184$ per side; hence, the particle mass is $m_0 = 5 \times 10^9\,h^{-1}{\rm M_\odot}$, about 2.4 times larger than that of AbacusSummit for the base cosmology.

While the default version of FastPM uses an extra particle species to simulate massive neutrinos (Bayer et al. 2021c), which are labelled as ncdm (not-cold dark matter), we use a modified version to treat ncdm the same as that in AbacusSummit, i.e. the effect of massive

neutrinos only contributes to the Hubble expansion rate but not to the gravitational forces from clustering.

In the end, we find a reasonable set of configuration parameters for the massive production of FastPM simulations. We run FastPM from the initial redshift $z_0 = 19$ with the second-order Lagrangian perturbation theory (2LPT) IC to the final redshift $z = 0.1$, with 40 time-steps linearly separated in scale a. We set the PM size parameter $B = 2$. We validate such choice of the configuration parameters in Appendix A.

2.2.2 ICs and Cosmologies

Using the AbacusSummit base cosmology (c000), we have produced a bunch of FastPM simulations including

- 25 boxes with the AbacusSummit ICs;
- 201 boxes with independent ICs;
- 237 boxes with the fixed-amplitude ICs.

The technique of fixed amplitude of the initial density field (Angulo & Pontzen 2016) is a method to effectively suppress the sample variance, and hence the number of realizations is greatly reduced to reach a certain precision at large scales. Different from the Rayleigh distributed amplitude in the Gaussian density field, the amplitude of the density fluctuation is fixed to be the square root of the input linear power spectrum, i.e.

$$\delta^L(\boldsymbol{k}) = \sqrt{P(k)}\exp(i\theta(\boldsymbol{k})), \tag{1}$$

and $\theta(\boldsymbol{k})$ is the phase uniformly distributed in $(0, 2\pi)$. There have been multiple studies (e.g. Villaescusa-Navarro et al. 2018; Chuang et al. 2019) showing the unbiasedness of several statistics from the fixed-amplitude method, compared with the results simulated from the Gaussian initial density field.

In terms of AbacusSummit secondary cosmologies, we have produced

- 25 boxes for both c002 and c004 (50 boxes in total) with the same white noises as the base cosmology c000.

2.2.3 Running DESI-FastPM on NERSC

As a project in the DESI collaboration, we run all the FastPM simulations on Cori supercomputer at the National Energy Research Scientific Computing Center (NERSC)[5]. NERSC is one of the largest facilities in the world devoted to basic scientific research. We use KNL computing nodes for the computation. Each KNL node has 68 CPU cores and 96 GB memory. We assign 1152 nodes and 36 MPI tasks per node for each FastPM simulation. Each simulation takes

---

[3] https://abacusutils.readthedocs.io/en/latest/
[4] https://github.com/fastpm/fastpm

[5] https://www.nersc.gov/

about 50 min of wall-clock time. We down-sample dark matter particles and compress particle information into integers. Such process uses the same amount of nodes and about 7 min of wall-clock time. The total cost of the simulations is about 24 million NERSC hours. For each simulation, we utilize about 55000 GB temporary space from the Burst Buffer[6] to store the output. Once the simulation is finished, the output is transported from the Burst Buffer to the disk automatically in the backend.

2.2.4 Products and Storage

We output 12 snapshots of FastPM dark matter catalogues and halo catalogues at the same primary redshifts of AbacusSummit. Finding haloes from the dark matter field, we use the FastPM internal Friends-of-Friends algorithm with the linking length equal to 0.2 times of the mean separation of particles. We store haloes with mass larger than $M_{halo} = 5 \times 10^{10} \, h^{-1} M_\odot$. Since we need to run hundreds of simulations, the total output data will take too much storage. To save disk space, we down-sample dark matter particles by 1/27 and store their positions and velocities in 1-byte integer for future usage, e.g. matter density field and weak lensing light-cone construction. Specifically, the position is stored in the form of displacement from Lagrangian lattice. The float displacement and velocity are converted into bits by the error function $erf(sx)$. We minimize the loss in the conversion by choosing the optimal scaling factor s in the error function for each redshift. We store information of halo catalogues, including positions, velocities, masses, inertial tensors, velocity dispersion and angular momenta, using 4-byte floating points. For each simulation, the total data size is about 810 GB with 350 GB for dark matter particles and 460 GB for haloes.

3 CARPOOL METHOD

Based on the principle of control variates (Rubinstein & Marcus 1985; Avramidis & Wilson 1993; Porta Nova & Wilson 1993), Chartier et al. (2021) applied it to construct variance-reduced observables of large-scale structure clustering based on simulations. The approach is named as *Convergence Acceleration by Regression and Pooling*, short for CARPool. It pairs a few N-body simulations and surrogates which approximate N-body simulations and share the ICs from the N-body simulations. The method of CARPool can be summarized by the equation as

$$x = y - \beta(c - \mu_c),\tag{2}$$

where x is constructed as a representative of y which is some observable from an N-body simulation, e.g. AbacusSummit in our case, c is the same observable from the paired surrogate, e.g. FastPM, β is the control variate, and μ_c is the mean of c. If μ_c is unknown, we can estimate it from a separate set of surrogates that do not share the ICs from the N-body simulations, i.e.

$$\hat{\mu}_c = \frac{1}{M} \sum_{j=1}^{M} c_j,\tag{3}$$

where M is the number of surrogates. In our case, we have 201 independent normal boxes and 237 boxes using the fixed-amplitude ICs (see Sec. 2.2.2), both of which can be used to estimate μ_c. The computational cost of CARPool can be much cheaper than the

traditional method which usually needs to run a large number of N-body simulations.

First, by design, x is unbiased relative to y, as the ensemble average of equation (2) gives the expectation of x equal to that of y, i.e. $\bar{x} = \bar{y}$. In the following we use overbars to denote means. Secondly, we can find the best β to minimize the variance of x. If y is a vector, e.g. halo power spectrum, we have

$$\beta^\star = \Sigma_{yc} \Sigma_{cc}^{-1},\tag{4}$$

where Σ_{yc} is the cross-covariance matrix between y and c, and Σ_{cc} is the covariance matrix of c, i.e.

$$\Sigma_{yc} = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})(c_i - \bar{c})^T,\tag{5}$$

$$\Sigma_{cc} = \frac{1}{N-1} \sum_{i=1}^{N} (c_i - \bar{c})(c_i - \bar{c})^T,\tag{6}$$

where N is the number of paired simulations, and $\bar{y}$ and $\bar{c}$ denote the mean of the paired N-body simulations and surrogates, respectively.

As suggested by Chartier et al. (2021), if there are not many paired simulations, which is our case, it is better to choose the diagonal form of $\beta^\star$ to give better performance of CARPool, i.e.

$$\beta^{diag} = Diag(\beta^\star) = \frac{\sigma_{yc}^2}{\sigma_c^2},\tag{7}$$

where σ_{yc}^2 is the cross-correlation between y and c from the same bins, and σ_c^2 is the variance of c. Our following results are all based on the diagonal form of β; hence, we ignore the superscript of β^{diag} hereafter.

The variance of the calibrated variable x is

$$\sigma_x^2 = \sigma_y^2 - 2\beta\sigma_{yc}^2 + \beta^2\sigma_c^2 + \beta^2\sigma_{\hat{\mu}_c}^2,\tag{8}$$

where we account for the variance of $\hat{\mu}_c$ estimated from surrogates,

$$\sigma_{\hat{\mu}_c}^2 = \frac{1}{M(M-1)} \sum_{j=1}^{M} (c_j - \hat{\mu}_c)^2.\tag{9}$$

Note that there is no cross-correlation between y and $\hat{\mu}_c$ in equation (8), since they are generated from different ICs. Once substituting β in equation (8), we obtain

$$\sigma_x^2 = \left(1 - \rho_{yc}^2\right)\sigma_y^2 + \frac{\sigma_{yc}^4}{\sigma_c^4}\sigma_{\hat{\mu}_c}^2,\tag{10}$$

where ρ_{yc} is the Pearson correlation coefficient between y and c, i.e. $\rho_{yc} = \frac{\sigma_{yc}^2}{\sigma_y \sigma_c}$. One can see that σ_x could be very small if y and c are highly correlated (ρ_{yc} close to 1.0), and if $\sigma_{\hat{\mu}_c}$ is small too. In addition, we can derive the variance of $\bar{x}$ by

$$\sigma_{\bar{x}}^2 = \frac{1}{N}\left(1 - \rho_{yc}^2\right)\sigma_y^2 + \frac{\sigma_{yc}^4}{\sigma_c^4}\sigma_{\hat{\mu}_c}^2.\tag{11}$$

where the scaling factor $1/N$ should not be applied to the term with $\sigma_{\hat{\mu}_c}^2$. This is because we use the same set of surrogates to estimate $\hat{\mu}_c$ when we calculate x for each paired simulation. Taking the mean of the CARPool result x can only suppress the sample variances of the N-body simulations, but not that of the surrogate mean $\hat{\mu}_c$.

One can also derive the effective volume by comparing $\sigma_{\bar{x}}$ and $\sigma_{\bar{y}}$, where $\sigma_{\bar{y}}^2 = \sigma_y^2/N$ and $N = 25$ in this study. Note that the combination of 25 AbacusSummit base simulations has a total volume of $200 \, h^{-3}Gpc^3$ which is about 10 times the effective volume of

---

[6] It uses flash or SSD (solid-state drive) array to achieve high speed on I/O.

DESI survey. Thus, the effective volume gained from CARPool can be given by

$$V_{\text{eff}} = \frac{\sigma_{\overline{y}}^2}{\sigma_{\overline{x}}^2} 200 \, h^{-3} \text{Gpc}^3 = \frac{\sigma_{\overline{y}}^2}{\sigma_{\overline{x}}^2} 10 \, V_{\text{DESI}} \qquad (12)$$

4 APPLICATION ON HALO CLUSTERING

Chartier et al. (2021) have applied the CARPool method on the clustering statistics of dark matter in real space. In our study, we extend their analysis to halo clustering with two-point statistics in redshift space and three-point statistics in real space.

4.1 Halo power spectrum

We study halo catalogues at redshift $z = 1.1$ which is a typical redshift since DESI will observe 10 million ELGs at redshift $0.6 < z < 1.6$. We apply a mass cut and select haloes with mass larger than $10^{11} \, h^{-1} M_\odot$ which is the expected minimum halo mass hosting ELGs. Apart from the differences on the dark matter simulations, the halo finders from AbacusSummit and FastPM are different as well. Therefore, we do expect some difference on the halo clustering even on large scales. We first study the halo power spectrum $P(\boldsymbol{k})$ defined as

$$\langle \delta(\boldsymbol{k}) \delta(\boldsymbol{k}') \rangle \equiv (2\pi)^3 \delta_D(\boldsymbol{k} + \boldsymbol{k}') P(\boldsymbol{k}), \qquad (13)$$

where $\delta(\boldsymbol{k})$ is the halo number density fluctuation as a function of wavevector $\boldsymbol{k}$ in Fourier space and δ_D is the Dirac delta function. In redshift space, power spectrum is not isotropic due to the peculiar velocity along the line of sight, hence, it can be decomposed into multipoles

$$P_\ell(k) = (2\ell + 1) \int_0^1 P(k, \mu) L_\ell(\mu) d\mu, \qquad (14)$$

where $L_\ell(\mu)$ is the Legendre polynomial of order ℓ, and μ is the cosine of the angle between $\boldsymbol{k}$ and the line of sight, i.e.

$$\mu = k_\parallel / k, \quad k = \sqrt{k_\perp^2 + k_\parallel^2}, \qquad (15)$$

with $k_\perp$ and $k_\parallel$ being the components of $\boldsymbol{k}$ perpendicular and parallel to the line of sight, respectively. In our study, we present the results of monopole and quadrupole that are widely analysed in galaxy surveys.

We calculate the halo power spectrum multipoles via NBODYKIT[7] (Hand et al. 2018). For the calculation, we paint haloes in a mesh with 1024^3 cells using the triangular-shaped cloud mass assignment window. We eliminate the aliasing effect with the interlacing technique which can compensate the window function effect. We set 120 linear μ bins in the range $[0, 1]$ and integrate the anisotropic power spectrum over μ (based on equation 14) to obtain the power spectrum multipoles. We study the multipoles in the k range from 0.0038 to $0.8025 \, h \, \text{Mpc}^{-1}$ which is half of the Nyquist frequency. The total number of k bins is 161 with the interval $0.005 \, h \, \text{Mpc}^{-1}$. To be compact, for the following, we show only the CARPool results from the power spectrum monopole ($\ell = 0$) and quadrupole ($\ell = 2$) while ignoring the hexadecapole which has larger statistical noise.

Fig. 1 shows the mean halo power spectrum multipoles calculated from 25 pairs of AbacusSummit and FastPM catalogues, respectively. In the upper panels, we compare the overall shapes of monopoles (left-hand panel) and quadrupoles (right-hand panel)

[7] https://nbodykit.readthedocs.io

with the standard deviations of the mean from the paired simulations. In the lower panels, we show the ratio of the mean from AbacusSummit and FastPM. For the monopole, there is a constant bias about 4.2% at large scales. Although there is noticeable difference on the halo number densities from the two simulations[8], we have checked that using abundance matching can only reduce the bias to 3.7%. Instead, if we use the cleaned AbacusSummit halo catalogues, the bias decreases to 1.2%. Thus, we argue that the constant bias at large scales is mainly due to different halo finders in the two simulations. As a supplement, we show the power spectrum of the cleaned haloes in Appendix B. At small scales, the power spectrum of AbacusSummit is smaller than that of FastPM. This is caused by the underestimation of the small-scale damping of redshift-space distortions (RSDs; i.e. the Fingers-of-God effect) in FastPM, since it is not able to trace the velocity field of particles precisely at small scales.

4.1.1 Performance of CARPool method

Based on the power spectrum multipoles from the paired AbacusSummit and FastPM halo catalogues, we can calculate their cross-correlation, as well as the variance from each simulation. From equation (7), we obtain β^{diag} for the monopole and quadrupole, as shown in Fig. 2. We find that the overall shapes of β^{diag} for the monopole and quadrupole are similar, i.e. it is close to 1.0 on large scales (small k), which is due to the high cross-correlation between the multipoles from the paired AbacusSummit and FastPM simulations. This is guaranteed since the paired simulations share the same ICs and have similar information of the large scale structures. As the scale becomes smaller, the cross-correlation decreases, due to the differences from the non-linear evolution, the halo finders, etc., between the two simulations. The noises in the curves are due to the fact that we are using only 25 pairs of simulations to determine the variances.

We compare the influence on β^{diag} from mass cut, abundance matching, as well as the halo cleaning of AbacusSummit in Appendix B. There is negligible improvement from abundance matching, whereas using the cleaned AbacusSummit catalogues improves β^{diag} closer to 1.0 by a few per cent in the range $k < 0.2 \, h \, \text{Mpc}^{-1}$. We believe that neither using abundance matching nor cleaned haloes of AbacusSummit will influence our results much.

For given paired power spectrum multipoles from AbacusSummit and FastPM, we construct the new power spectrum multipoles $P_{x, \ell}$ via CARPool,

$$P_{x, \ell} = P_{\text{AbacusSummit}, \ell} - \beta(P_{\text{FastPM}, \ell} - \overline{P}'_{\text{FastPM}, \ell}), \qquad (16)$$

where $\overline{P}'_{\text{FastPM}, \ell}$ with an overline denotes the mean halo power spectrum multipoles calculated from a separate set of FastPM simulations, e.g. 201 FastPM simulations with random ICs. As discussed in Section 3, the mean $P_{x, \ell}$ should be unbiased compared with the mean $P_{\text{AbacusSummit}, \ell}$, and the statistical error of $P_{x, \ell}$ should be smaller than that of $P_{\text{AbacusSummit}, \ell}$. We check these by calculating the mean of $P_{x, \ell}$ over 25 realizations and the standard deviation,

$$\overline{P}_{x, \ell} = \frac{1}{N} \sum_i^N P_{x, \ell}^i, \qquad (17)$$

$$\sigma_{P_{x, \ell}}^2 = \frac{1}{N - 1} \sum_i^N (P_{x, \ell}^i - \overline{P}_{x, \ell})^2. \qquad (18)$$

[8] With mass cut $10^{11} \, h^{-1} M_\odot$, AbacusSummit has halo number density $0.036 \, h^3 \text{Mpc}^{-3}$, 36% higher than that of FastPM.
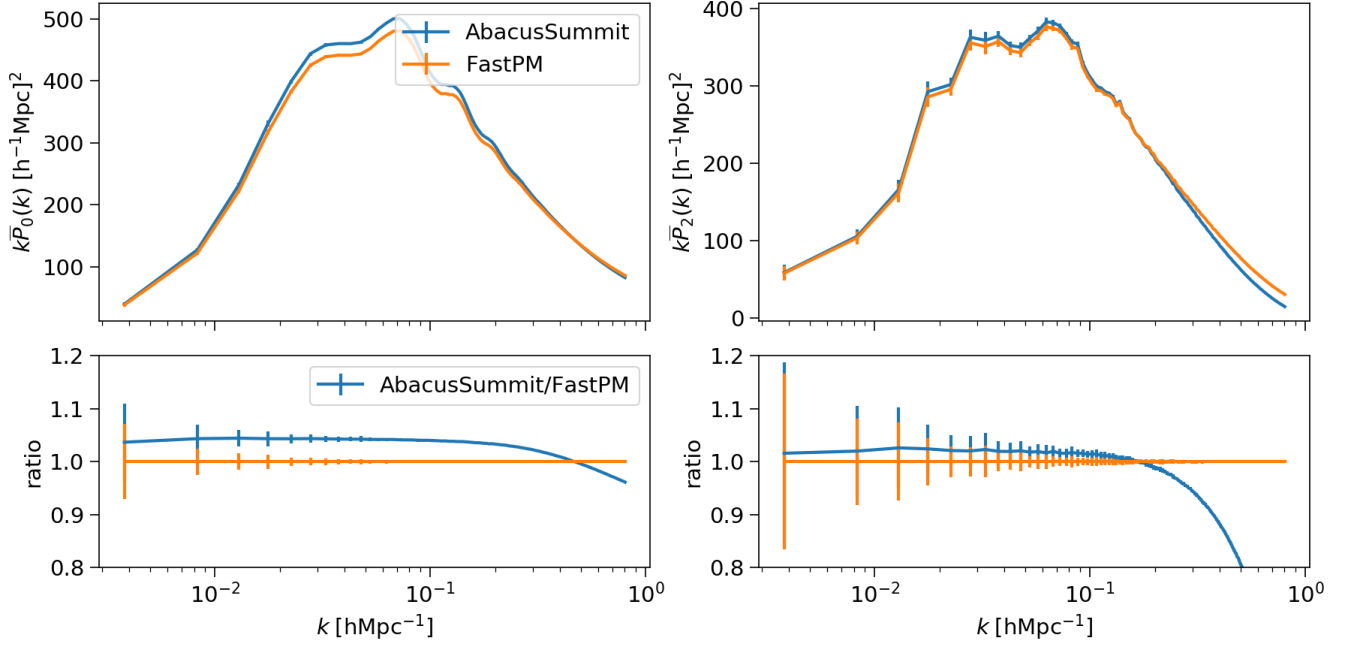
Figure 1. *Upper panels*: the mean halo power spectrum monopole (left-hand panel) and quadrupole (right-hand panel) from 25 halo catalogues of the paired AbacusSummit and FastPM simulations. The halo catalogues are at redshift 1.1 and contain haloes with mass larger than $10^{11}\ h^{-1}M_\odot$. The blue lines are for AbacusSummit and the orange lines are for FastPM. *Lower panels*: the ratio of the mean between AbacusSummit and FastPM. The error bars represent the standard deviations of the mean multipoles.
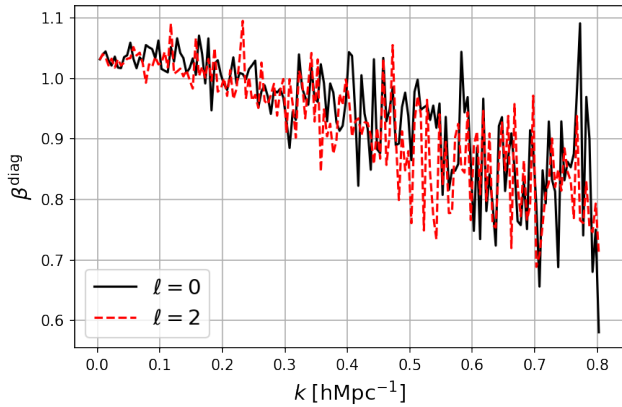

Figure 2. β^{diag} of the halo power spectrum multipoles, i.e. monopole $\ell = 0$ and quadrupole $\ell = 2$, from the paired AbacusSummit and FastPM halo catalogues with mass cutoff $10^{11}\ h^{-1}M_\odot$. The cross-correlation between the multipoles of AbacusSummit and FastPM is high on large scales, as β^{diag} is close to 1.0, and it decreases as the scale becomes smaller. The shape of β^{diag} is similar between the monopole and quadrupole.

Since we estimate the mean of FastPM multipoles from a limited number of realizations, the standard deviation of $P_{x,\ell}$ (equation 18) is underestimated. Considering the error of $\overline{P}'_{\mathrm{FastPM},\ell}$, we have

$$\sigma^2_{P'_{x,\ell}} = \sigma^2_{P_{x,\ell}} + \beta^2\sigma^2_{\overline{P}'_{\mathrm{FastPM},\ell}}. \tag{19}$$

And for the standard deviation of the mean of $P'_{x,\ell}$, it is

$$\sigma^2_{\overline{P}'_{x,\ell}} = \frac{1}{N}\sigma^2_{P_{x,\ell}} + \beta^2\sigma^2_{\overline{P}'_{\mathrm{FastPM},\ell}}, \tag{20}$$

From now on, we ignore the subscript ℓ of multipoles for simplicity.

In the upper panels of Fig. 3, we show the mean multipoles with the standard deviations from CARPool as the orange lines. To avoid crowdedness of data points at small scales, we redo k binning for the power spectrum multipoles using a larger k interval. The left-hand panel is for the monopole and the right-hand panel is for the quadrupole. In the standard deviation of the CARPool mean, we take account of the error of the FastPM mean, which is calculated from 201 sets of regular FastPM simulations. We compare the results with the mean multipoles from 25 AbacusSummit halo catalogues shown as the blue points. We see that they agree well with each other in 1σ error over all the scales except for some points at large scales, which is just due to cosmic variance. In the lower panels, we show the ratio of the mean between AbacusSummit and CARPool, shown as the blue points. The orange shaded region denotes the noise-to-signal ratio, i.e. $\sigma_{\overline{P}_x}/\overline{P}_x$ from CARPool. The blue points fluctuate near unity over all the scales, illustrating the unbiasedness of P_x. In addition, the standard deviation of $\overline{P}_x$ from CARPool is much smaller than the original standard deviation of $\overline{P}_{\mathrm{AbacusSummit}}$.

To quantify the reduction on the sample variance from CARPool, we compute the ratio of the standard deviation of P_x over that of $P_{\mathrm{AbacusSummit}}$, and show the results in Fig. 4. The left-hand panel is for the monopole and the right-hand panel is for the quadrupole. In each panel, the blue line is the optimal case where we assume that the surrogate mean, i.e. $\overline{P}'_{\mathrm{FastPM}}$, is known. The red dashed line is a realistic case once we consider the standard deviation of $\overline{P}'_{\mathrm{FastPM}}$ from 201 regular FastPM simulations. On large scales, the reduction of statistical error from CARPool is very significant, even after in-
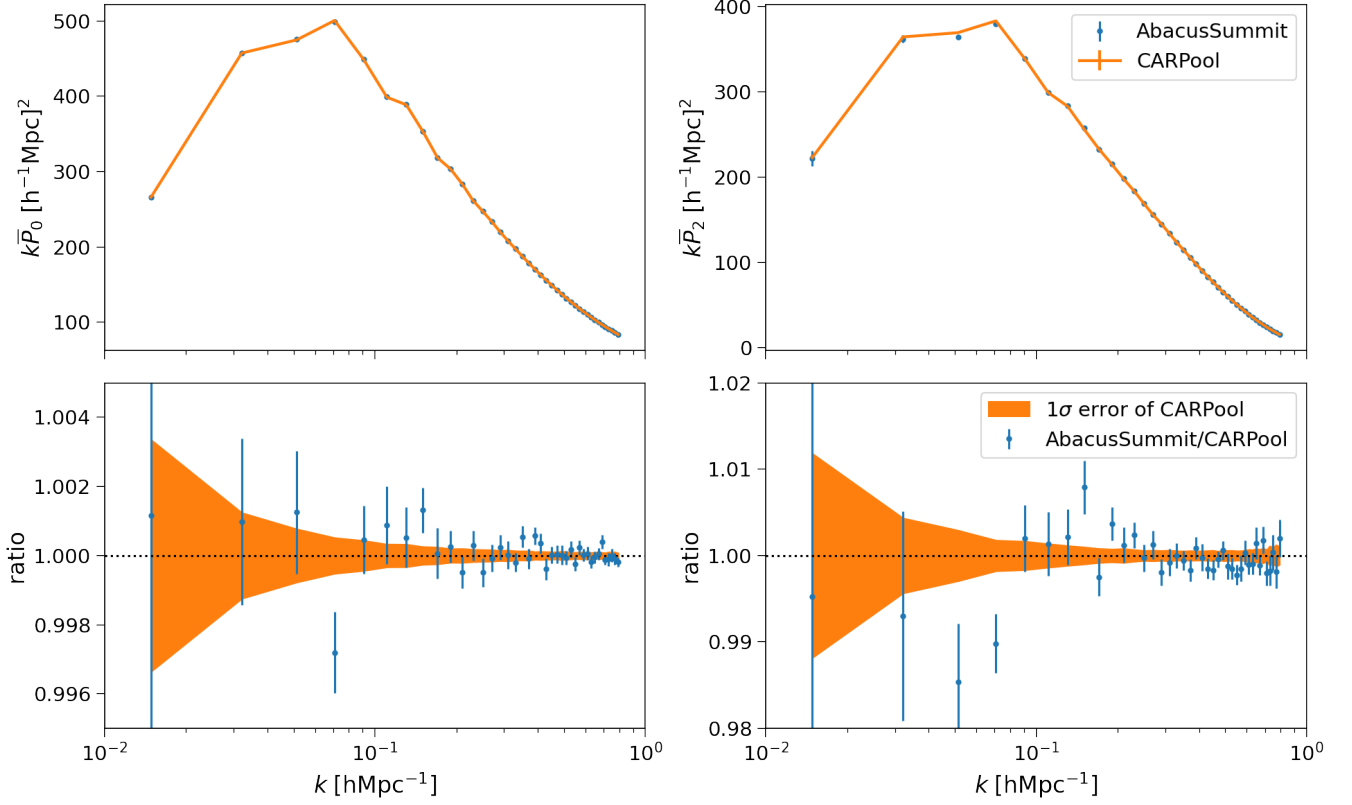
Figure 3. *Upper panels*: the mean of the halo power spectrum monopole (left-hand panel) and quadrupole (right-hand panel) from 25 AbacusSummit simulations compared with the mean of P_x from CARPool. Error bars have been scaled to represent the standard deviation of the mean. *Lower panels*: the ratio of the mean from AbacusSummit and CARPool, shown as the blue points, compared with the 1σ error of the mean P_x, shown as the orange shades. There is no bias on the mean power spectrum multipoles from CARPool, and the sample variance is reduced significantly.
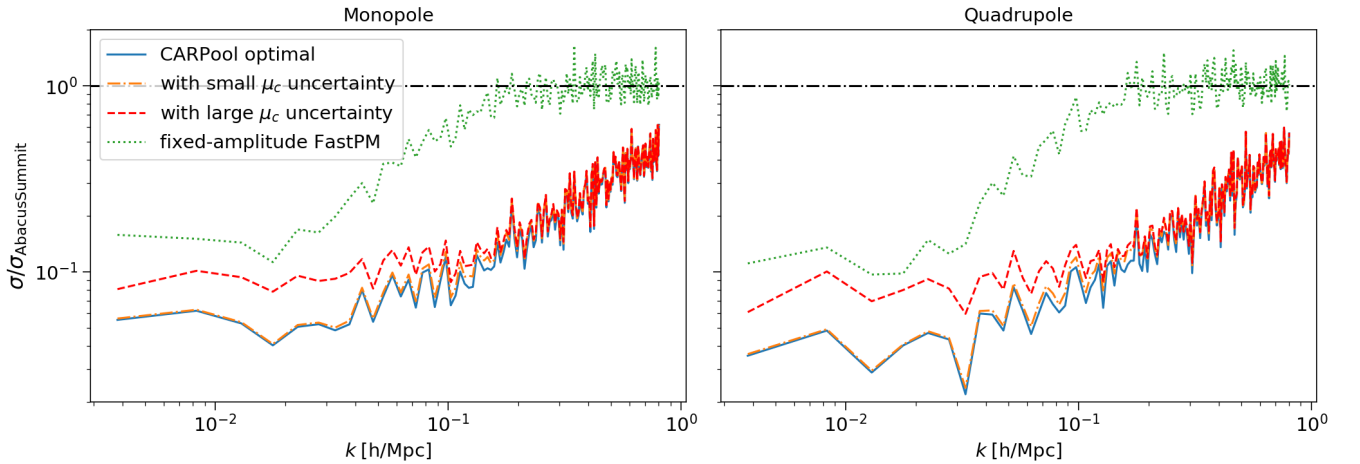
Figure 4. The reduction of the sample variance on the halo power spectrum multipoles from CARPool and the method of the fixed-amplitude IC. The left-hand panel is for the monopole and the right-hand panel is for the quadrupole. In each panel, the blue solid line shows the optimal gain expected from CARPool if the mean of the FastPM multipole in (16) is known. The red dashed line is one realistic case considering the sample variance of the surrogate mean estimated from 201 nonfixed-amplitude FastPM realizations. The fixed-amplitude method can also effectively reduce the sample variance, especially on large scales. We calculate the standard deviation of 237 fixed-amplitude FastPM realizations and compare it with that of the non-fixed-amplitude ones, shown as the green dotted lines. Using CARPool with the surrogate mean from the fixed-amplitude FastPM, we can further increase the error reduction from the red dashed lines to the orange dot–dashed lines which almost overlap with the blue lines. The performance of CARPool on the monopole and quadrupole is similar, which is expected since the cross-correlation between AbacusSummit and FastPM is similar for the monopole and quadrupole, shown as β^{diag} in Fig. 2.

cluding the error of the surrogate mean. In our conservative case (red dashed line), CARPool can reduce the standard deviation to be 10 times smaller, for both monopole and quadrupole on large scales. As the scale goes smaller ($k > 0.1\,h\,\mathrm{Mpc}^{-1}$), the gain from CARPool gradually decreases; however, it still has 50% error suppression up to $k = 0.8\,h\,\mathrm{Mpc}^{-1}$.

Since the fixed-amplitude technique can also reduce statistical errors especially on large scales, it would be interesting to compare the performance with CARPool. In our project, we have run 237 realizations of FastPM with the fixed-amplitude ICs. We compare the standard deviation of the fixed-amplitude FastPM simulations with that of the FastPM simulations using the AbacusSummit ICs. The results are shown as the green dotted lines in Fig. 4. For the power spectrum multipoles, the error reduction from the fixed-amplitude method has a similar trend as the CARPool result, but is less significant than that of CARPool, and it quickly reduces to zero on smaller scales, in our case, $k > 0.2\,h\,\mathrm{Mpc}^{-1}$. Because the error reduction from CARPool is limited by the error of surrogate mean, we confirm that if we replace the surrogate mean by the mean from the fixed-amplitude FastPM simulations, we can further reduce the error of the CARPool result. As a result, we obtain the gain close to the optimal case, shown as the orange dot–dashed lines, which nearly overlap with the blue lines.

The reduction on the sample variance corresponds to the increase of the effective volume. Based on equation (12), we estimate the effective volume from the combination of 25 AbacusSummit boxes with CARPool and compare it with the total DESI 5 yr survey volume. We summarize our results in Fig. 5 with the left-hand panel for the monopole and the right-hand panel for the quadrupole. In each panel, the blue line is the optimal case from CARPool assuming that the surrogate mean is known. The red dashed line is a realistic case considering the error of $\overline{P}'_{\mathrm{FastPM}}$ from the non-fixed-amplitude FastPM, where we can increase the effective volume about 100 times over all the k range. Different from the optimal case that has a larger increase of the effective volume on larger scales, in the realistic case the error of the surrogate mean suppresses such increase as the error of the surrogate mean dominates in equation (20). The green dotted line represents the effective volume from 25 fixed-amplitude FastPM boxes. On large scales, it is higher than the red dashed line but drops quickly as the scale goes smaller and turns to no gain on scales of k larger than $0.2\,h\,\mathrm{Mpc}^{-1}$. If we replace the surrogate mean from the non-fixed-amplitude FastPM by that of the fixed-amplitude ones, the error of the surrogate mean is dramatically suppressed. We obtain the result shown as the orange dot–dashed line which can be understood as the combination of the red and green lines. With CARPool and the fixed-amplitude FastPM simulations, we can extend the effective volume of 25 AbacusSummit simulations to $10^2 \sim 10^3$ times the DESI volume with the dependence on scales.

We further study the CARPool performance on the halo catalogues with higher mass cut $10^{13}\,h^{-1}\mathrm{M}_\odot$, which is about the mass boundary of host haloes of LRGs. Such catalogues are just the subsamples with high-mass haloes from the halo catalogues with mass cut $10^{11}\,h^{-1}\mathrm{M}_\odot$. Fig. 6 shows the gain of effective volumes from CARPool for the monopole and quadrupole. Compared with the results from the halo catalogues with lower mass cut, the performance of CARPool is worse based on the increased effective volume. We think that it can be caused by shot noise, as the number of haloes decreases dramatically with higher mass cut. At lower redshifts ($z < 1.1$), the number of LRG-host haloes is larger. For DESI, the number density of LRGs peaks around $z = 0.6$–0.8. We check the CARPool performance at $z = 0.8$ in Appendix C. Indeed, when the number density increases by 40% from $z = 1.1$ to 0.8, the effective volume

obtained from CARPool increases by 35%. Apart from that, the differences on the halo populations from the different halo finders, the non-linear structure growth and halo bias may affect the CARPool performance as well. We compare the Pearson correlation coefficients of the halo power spectrum multipoles from AbacusSummit and FastPM in Appendix D. The diagonal terms of the Pearson coefficients are closer to 1.0 for the lower halo mass cut, which indicates better performance from CARPool.

4.2 Halo correlation function

Since a correlation function is just the Fourier transform of power spectrum, we expect that the performance of CARPool on the halo correlation function is similar to the power spectrum. Chartier et al. (2021) mainly studied the performance of CARPool on the clustering statistics in Fourier space, hence, it is worth investigating on the improvement in configuration space. For a cubic box, we can calculate the halo correlation function based on Peebles & Hauser (1974), i.e.

$$\xi(s, \mu) = \frac{DD(s, \mu)}{RR(s, \mu)} - 1, \tag{21}$$

where $DD(s, \mu)$ and $RR(s, \mu)$ are respectively the normalized number of pairs of haloes and random points as a function of the separation distance $s = \sqrt{s_\perp^2 + s_\parallel^2}$ and the cosine angle between the separation vector s and the line of sight, i.e. $\mu = s_\parallel/s$. We calculate the number of halo pairs using Corrfunc[9] (Sinha & Garrison 2019, 2020) and FCFC[10] (Zhao et al. 2021), and have checked that the results from the two codes are consistent. For a cubic box, the number of random pairs can be predicted theoretically. In our calculation, we set 40 linear radial bins in the range $5 \le s \le 200\,h^{-1}\mathrm{Mpc}$ and 60 linear μ bins in the range $0 \le \mu \le 1.0$. Similar to the power spectrum multipoles, we calculate the correlation function multipoles from the anisotropic correlation function, i.e.

$$\xi_\ell(s) = (2\ell + 1) \sum_{i=1}^{N_\mu} \xi(s, \mu_i) L_\ell(\mu_i) \Delta\mu, \tag{22}$$

where N_μ is the total number of μ bins.

Fig. 7 shows the mean of the halo correlation function multipoles from 25 paired AbacusSummit and FastPM simulations, respectively. Same as Fig. 1, we study the halo catalogues at $z = 1.1$ with mass cutoff $10^{11}\,h^{-1}\mathrm{M}_\odot$. For the monopole, the difference from the two simulations is within 5% over all the scales. For the quadrupole, the difference is less than that of the monopole, except for scales smaller than $10\,h^{-1}\mathrm{Mpc}$.

4.2.1 Performance of CARPool method

We show the halo correlation function multipoles from CARPool in Fig. 8. Same as Fig. 3, we compare the mean of the monopoles (in the left-hand panels) and quadrupoles (in the right-hand panels) from AbacusSummit and CARPool. The upper panels show the overall shapes. The lower panels show the ratio of the mean multipoles of AbacusSummit over those of CARPool, given by the blue points. The shaded regions denote the standard deviations of the mean multipoles from CARPool. For the monopole, the relative difference of the mean is within 1σ error of AbacusSummit, while for the quadrupole,

---

[9] https://github.com/manodeep/Corrfunc
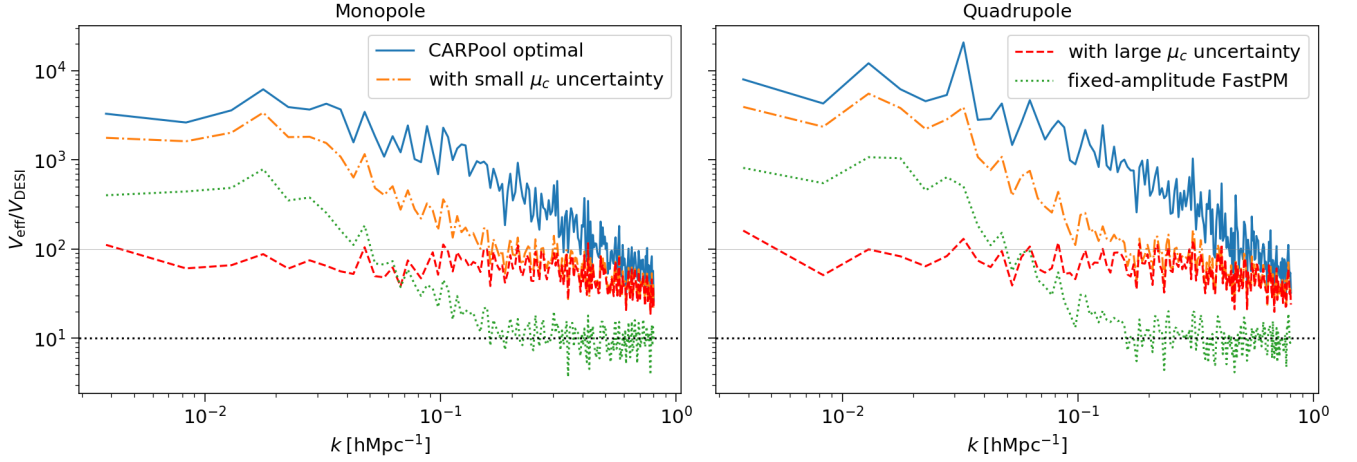[10] https://github.com/cheng-zhao/FCFC

Figure 5. The increase of the effective volume from the combination of 25 ABACUSSUMMIT simulations with CARPool compared with the result from the fixed-amplitude method. The left-hand (right-hand) panel is for the monopole (quadrupole). In each panel, the blue line represents the optimal case assuming no sample variance from the surrogate mean. The red dashed line takes account of the sample variance of the surrogate mean based on the non-fixed-amplitude FASTPM catalogues. The orange dot–dashed line is the case if we reduce the sample variance of the surrogate mean using the fixed-amplitude FASTPM catalogues. For comparison, we also show the effective volume from 25 fixed-amplitude FASTPM catalogues as the green dotted line.
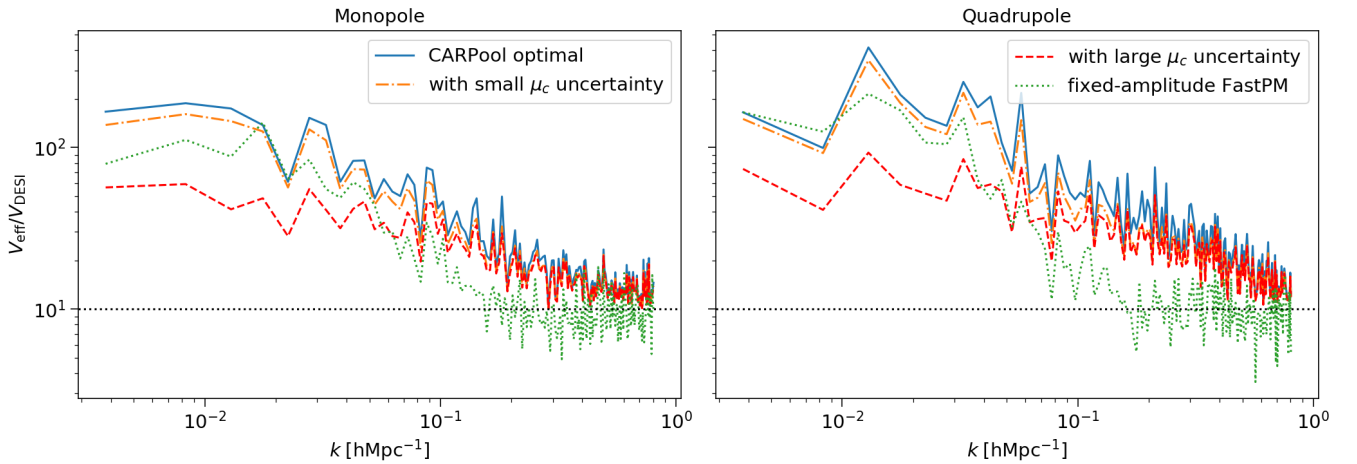
Figure 6. Same as Fig. 5 but for the halo catalogues with haloes massive than $10^{13} \, h^{-1} M_\odot$.

around scales $50 \, h^{-1}$Mpc, the difference is about 3σ error of ABA-CUSSUMMIT. However, it does not indicate that the CARPool result has bias; instead, it is just due to the sample variance from the paired FASTPM catalogues, which we demonstrate in Appendix E.

Fig. 9 shows the reduction of the sample variance of the correlation function multipoles from CARPool. Similar to Fig. 4, we compare the cases whether the surrogate mean in CARPool is assumed as known or not. As shown in equation (11), the uncertainty of the surrogate mean will propagate to that of CARPool result. For the red lines, we include the uncertainty of the mean correlation function multipoles that are calculated from the non-fixed-amplitude FASTPM catalogues. Comparing Fig. 9 and 4, we see that the reduction of the sample variance from CARPool is consistent for the halo power spectrum and correlation function at large scales.

We also show the gain of the effective volume from CARPool based on the suppressed variance of the correlation function multipoles in

Fig. 10. The meaning of each line is the same as that with the same line type in Fig. 5. The upper and lower panels show the results from the halo catalogues with mass cut 10^{11} and $10^{13} \, h^{-1} M_\odot$, respectively. We do not see the increase of V_{eff} from small scales to large scales as what we observe in the case of the power spectrum multipoles. It is mainly due to the high cross-correlation between correlation function bins.

4.3 Halo Bispectrum

Chartier et al. (2021) have studied the performance of CARPool on matter bispectrum in real space and found similar significance as that of matter power spectrum. Chuang et al. (2019) studied the sample variance reduction from the fixed-amplitude method on real-space halo bispectrum and found no improvement. Therefore, it is interesting to check the CARPool performance on higher-order halo
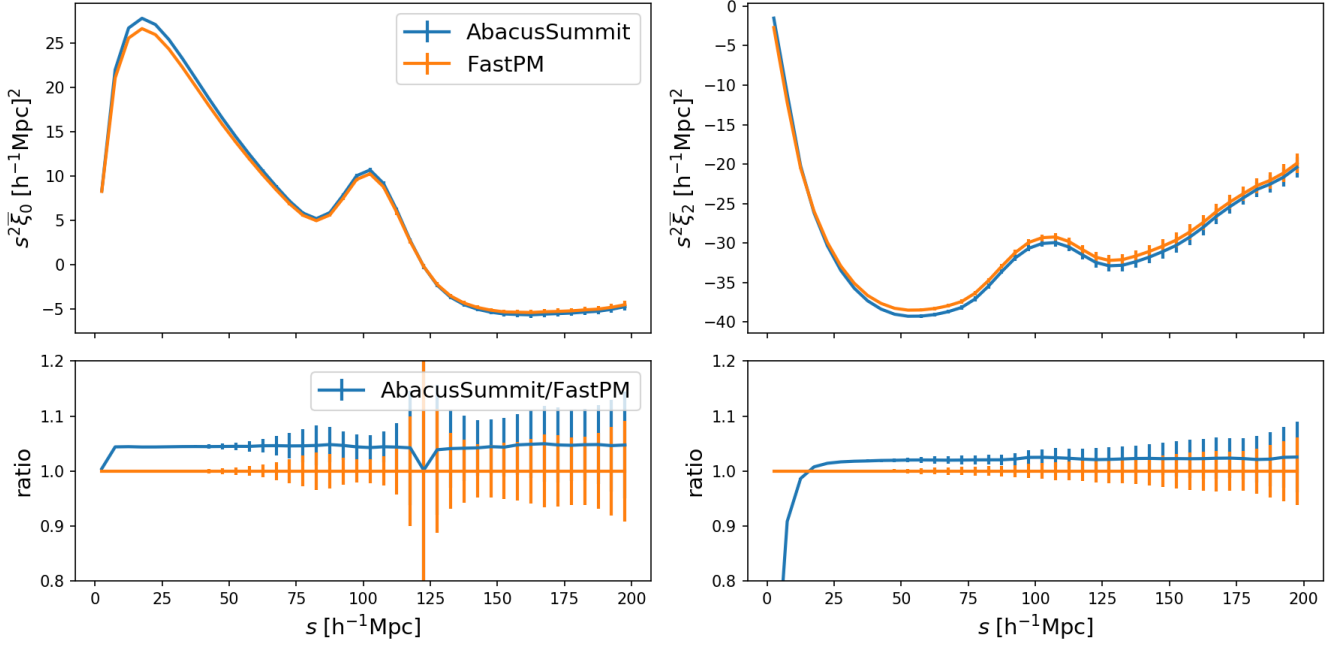
Figure 7. *Upper panels*: the mean of the halo correlation function monopole (left-hand panel) and quadrupole (right-hand panel) from 25 paired AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ and FᴀꜱᴛPM catalogues with halo mass larger than $10^{11}\,h^{-1}M_\odot$ and at redshift 1.1. The error bars have been scaled for the mean. The blue lines are the results of AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ and the orange lines are from FᴀꜱᴛPM. *Lower panels*: the ratio of the mean between AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ and FᴀꜱᴛPM. The orange horizontal lines denote the results of FᴀꜱᴛPM. For the monopole ratio, the error bars have a sudden peak around $125\,h^{-1}$Mpc, which is simply due to the zero crossing of the denominator.

clustering statistics. We study the halo bispectrum defined as

$$\langle \delta(\boldsymbol{k_1})\delta(\boldsymbol{k_2})\delta(\boldsymbol{k_3})\rangle \equiv \delta_D(\boldsymbol{k_1} + \boldsymbol{k_2} + \boldsymbol{k_3})B(k_1, k_2, k_3), \qquad (23)$$

where $\delta(\boldsymbol{k})$ is the halo number density contrast in Fourier space and δ_D is the Dirac delta function which ensures that the wavevectors $\boldsymbol{k_1}$, $\boldsymbol{k_2}$ and $\boldsymbol{k_3}$ form a closed triangle.

We use Pʏʟɪᴀɴꜱ3[11] (Villaescusa-Navarro et al. 2020) to calculate the bispectra for the halo catalogues with mass cut $10^{11}\,h^{-1}M_\odot$. The package implements a fast Fourier transform-based estimator (Sefusatti & Scoccimarro 2005; Scoccimarro 2015; Sefusatti et al. 2016) for the bispectrum calculation. The output bispectrum has been subtracted by the Poisson shot noise. Same as in Chuang et al. (2019), we choose the triangle configuration of $k_1 = 0.1\,h\,\mathrm{Mpc}^{-1}$ and $k_2 = 0.2\,h\,\mathrm{Mpc}^{-1}$, which are the scales related to the baryon acoustic oscillations (BAO) and RSD analysis. We study the reduced bispectrum defined as

$$Q(k_1, k_2, k_3) = \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2) + P(k_1)P(k_3) + P(k_2)P(k_3)}, \quad (24)$$

where $P(k)$ is the halo power spectrum. Given the amplitude of $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$, by varying $\boldsymbol{k_3}$, the reduced bispectrum is a function of θ, which is the angle between $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$. We use $Q(\theta)$ to represent it.

4.3.1 Performance of CARPool method

To compare the results in Chuang et al. (2019), we show our results in real space, but we have checked that the results in redshift space are similar too. Same as the analysis in halo power spectrum, we first

[11] https://pylians3.readthedocs.io/en/master/index.html

compare the mean of the reduced bispectra from 25 paired AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ and FᴀꜱᴛPM halo catalogues, shown in the first row of Fig. 11. We see that the bispectra from the two sets of halo catalogues agree relatively well with each other. The difference is within 10%. We also check the β^{diag} which is based on equation (7) and find it close to 1.0, shown as the middle row. We calculate the mean of the bispectra from the non-fixed-amplitude FᴀꜱᴛPM catalogues. We implement all the necessary elements into the equation of CARPool and obtain the final result. We compare the mean of CARPool and that of AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ and show the ratio between the two in the bottom row of the figure. The error bars denote the standard deviation of the mean of AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ, and the shaded region is the standard deviation of the CARPool mean with the consideration of the error of the surrogate mean from the non-fixed-amplitude FᴀꜱᴛPM catalogues. The agreement between the mean $Q(\theta)$ from CARPool and AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ is within 3σ error of AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ.

Comparing the standard deviation of the reduced bispectrum from CARPool and AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ, we can obtain the increased effective volume due to CARPool. Similar to Fig. 5, we show the effective volume from 25 AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ boxes before and after CARPool in Fig. 12. Again, the volume of 25 AᴮᴀᴄᴜꜱSᴜᴍᴍɪᴛ base boxes corresponds to 10 times the DESI volume, shown as the horizontal dotted line. CARPool can also significantly increase the effective volume for the halo bispectrum. The blue line shows the optimal case without considering the error of the surrogate mean. The red dashed line denotes a realistic case considering the error of the surrogate mean from the non-fixed-amplitude FᴀꜱᴛPM catalogues. It is about five times larger than the default volume.

We also check the improvement from the fixed-amplitude method, based on the standard deviation of the reduced bispectrum from the fixed-amplitude FᴀꜱᴛPM catalogues. The result is shown as the
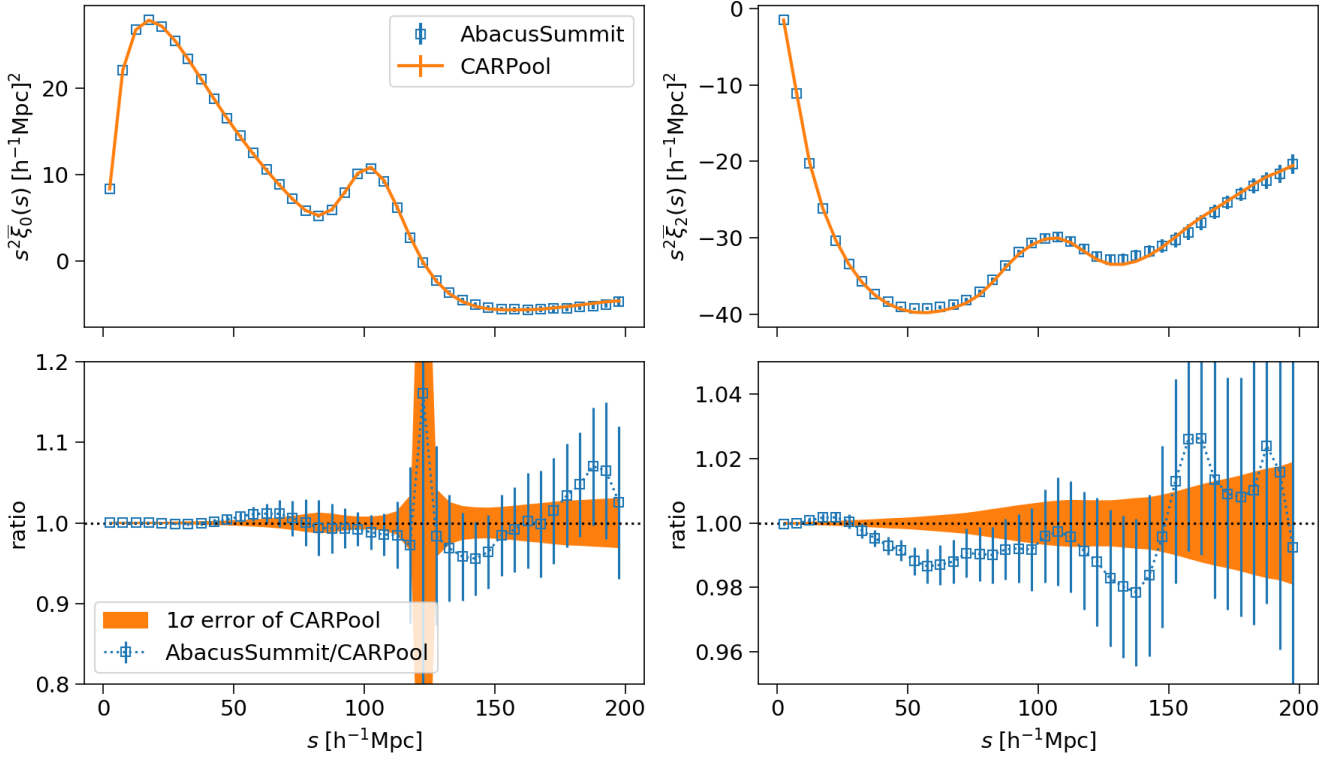
Figure 8. *Upper panels*: the mean of the halo correlation function multipoles from 25 AbacusSummit simulations compared with the mean of ξ_x from CARPool. The error bars have been scaled for the standard deviation of the mean multipoles. *Lower panels*: the ratio of the mean before and after CARPool, shown as the blue points, along with 1σ error of the mean ξ_x, shown as the orange shaded regions. For the monopole ratio, the error bars have a sudden peak around $125\,h^{-1}$Mpc, which is simply due to the zero crossing of the denominator. For the quadrupole, the discrepancy between the AbacusSummit and CARPool mean on scales around $50\,h^{-1}$Mpc is due to the statistical fluctuation of the mean from the paired FastPM, which we demonstrate in Fig. E1.
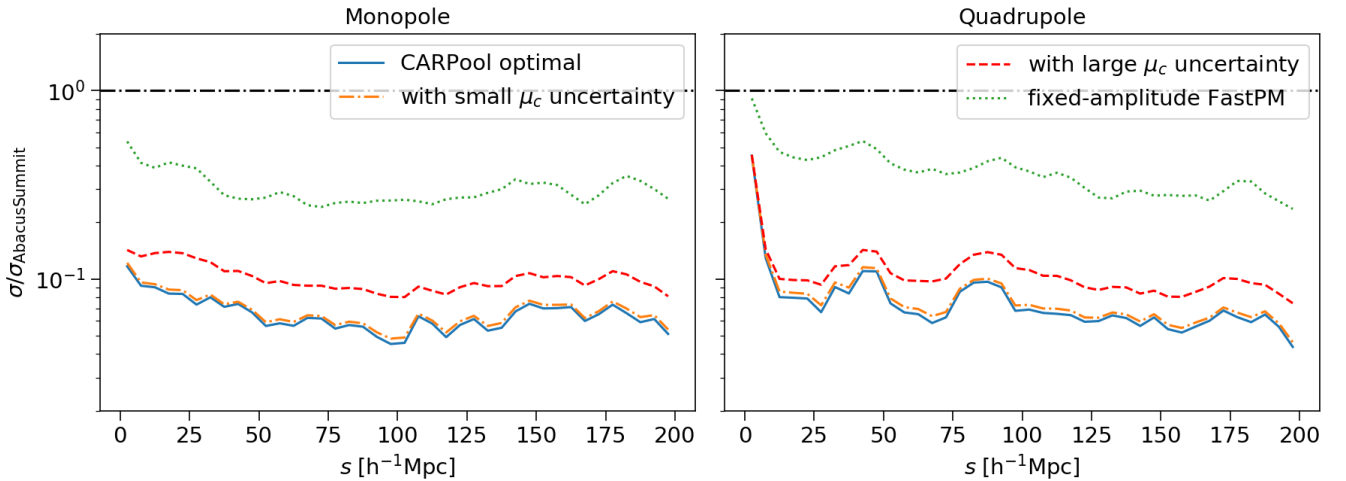
Figure 9. Similar to Fig. 4 but for the correlation function multipoles. The blue lines stand for the optimal gains from CARPool assuming that the surrogate mean is known. The red dashed (orange dot–dashed) lines are the cases while considering the sample variance of the surrogate mean which are estimated from the non-fixed (fixed)-amplitude FastPM, respectively. The green dotted lines are obtained from the ratio of the standard deviation of the fixed-amplitude FastPM over that of AbacusSummit.
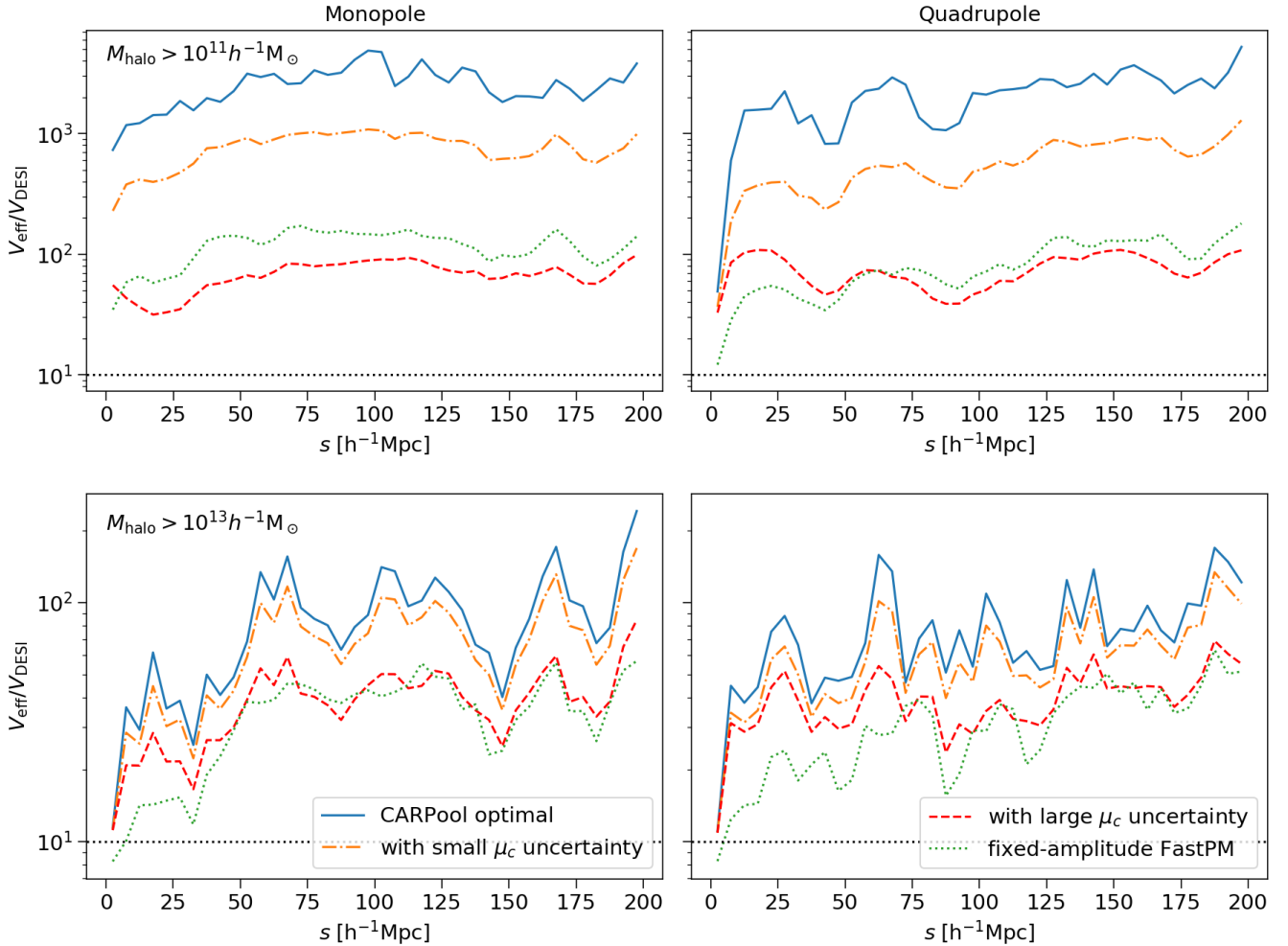
Figure 10. The effective volume derived from the halo correlation function multipoles using CARPool. We consider the effective volume from the combination of 25 ABACUSSUMMIT boxes. The upper (lower) panels show the results from the halo catalogues with halo mass larger than $10^{11}\,h^{-1}M_\odot$ ($10^{13}\,h^{-1}M_\odot$). The blue lines denote the optimal cases assuming that the surrogate mean is known, i.e. without considering the uncertainty of the surrogate mean. The red dashed (orange dot–dashed) lines take account of the statistical error of the surrogate mean estimated from the non-fixed (fixed)-amplitude FASTPM catalogues. We also include the results if we only use 25 fixed-amplitude FASTPM boxes, shown as the green dotted lines.

green dotted line, matching with the black dotted line, indicating no improvement, which is consistent with the finding of Chuang et al. (2019). Therefore, even if we use the surrogate mean from the fixed-amplitude catalogues instead of the non-fixed-amplitude ones in CARPool, there will be little improvement, shown as the orange dot–dashed line. Angulo & Pontzen (2016) found that the paired-and-fixed method can suppress the sample variance of the matter bispectrum with the triangle configuration $k_1 = 0.02\,h\,\mathrm{Mpc}^{-1}$ and $k_2 = 0.04\,h\,\mathrm{Mpc}^{-1}$ which is at a larger scale than the one we study here. But the variance suppression mainly comes from pairing instead of fixing. Klypin et al. (2020) further checked that as the scale becomes smaller, around the scales of BAO, the reduction from the paired-and-fixed method quickly vanishes. We believe that it should be true for the halo bispectrum too. At least, we have checked that the fixed-amplitude method does not reduce the sample variance of the halo bispectrum at the configuration of $k_1 = 0.02\,h\,\mathrm{Mpc}^{-1}$ and $k_2 = 0.04\,h\,\mathrm{Mpc}^{-1}$.

The failure of variance suppression from the fixed-amplitude method can be understood from the theoretical investigations on

the bispectrum in literature. Matsubara (2007) and Qin et al. (2022) found that the phases of the non-Gaussian density field are vital to the bispectrum. The late-time evolution induced phase autocorrelation and phase-modulus cross-correlation contribute almost equally to the bispectrum, whereas the modulus autocorrelation contributes little. Therefore, fixing the amplitude (modulus) of the initial density field should not affect the sample variance of the bispectrum. Adding pairing with the inverse phases may help reducing the sample variance from the phase-modulus cross-correlation at very large scales, but the reduction quickly vanishes as the scale becomes smaller. On the contrary, CARPool shows significant advantage in reducing the sample variance of bispectrum. For more detailed study on the bispectrum from the fixed-amplitude method (with pairing) and CARPool, we leave it for future work.
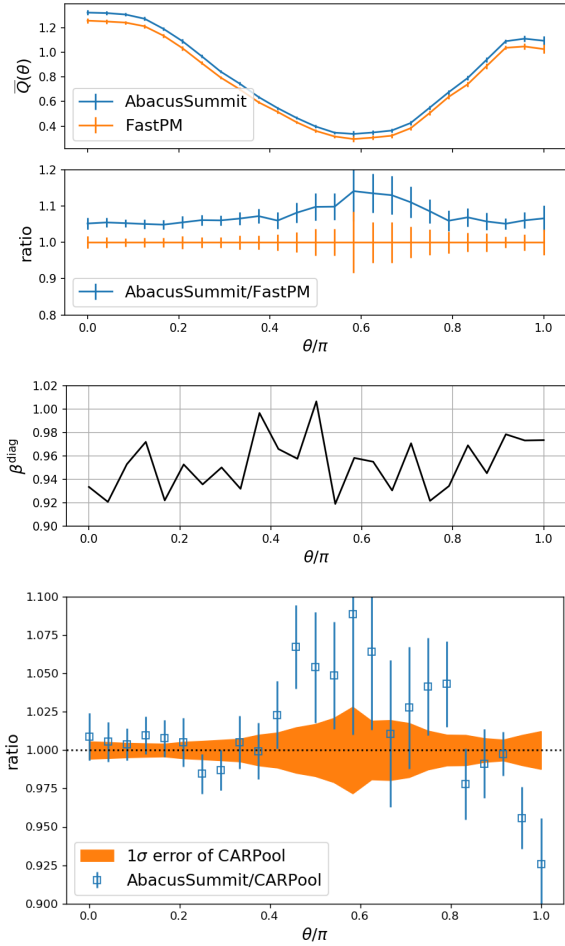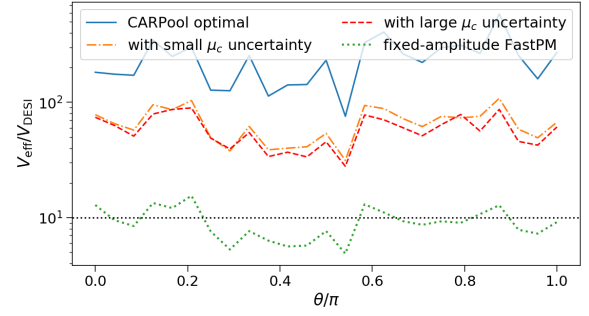
Figure 12. Similar to Fig. 5, we show the increased effective volume for the reduced bispectrum from CARPool. The blue line is the optimal case. The red dashed line and the orange dot–dashed line represent the cases taking account of the uncertainty of the mean bispectra estimated from the non-fixed-amplitude and the fixed-amplitude FASTPM catalogues, respectively. The green dotted line represents the result if we only use the fixed-amplitude FASTPM catalogues.

a larger amount of efforts as well. To resolve these, we modify the CARPool method by using the products in the base cosmology.

Since for the base cosmology we have the initial white noise of the 25 ABACUSSUMMIT boxes, we can use them to generate the same number of FASTPM simulations for a given secondary cosmology. We can pair these FASTPM simulations with the ABACUSSUMMIT ones in the base cosmology. Since they share the same initial white noise, there is non-zero cross-correlation between the halo statistics from the two sets of simulations even though with different cosmologies. We calculate the cross-correlation between the ABACUSSUMMIT from the base cosmology and the FASTPM from a second cosmology. We assume that it is close to the cross-correlation between the ABACUSSUMMIT from the secondary cosmology and the FASTPM from the base cosmology, i.e.

$$\text{Cov}(A_{c002}, F_{c000}) \approx \text{Cov}(A_{c000}, F_{c002}). \quad (25)$$

where A and F are short for ABACUSSUMMIT and FASTPM, respectively. The subscript c002 represents a secondary cosmology as an example. We roughly examine the viability of the above relation in Appendix F, but a better validation can be conducted if we have 25 sets of ABACUSSUMMIT simulations with the secondary cosmology c002. In the end, we construct the CARPool result for a secondary cosmology via

$$X_{c002} = A_{c002} - \beta(F_{c000} - \overline{F}'_{c000}), \quad (26)$$

where

$$\beta = \text{Diag}[\text{Cov}(A_{c000}, F_{c002})]/\text{Var}(F_{c000}). \quad (27)$$

Note that in the denominator of β we still use the variance of the FASTPM statistics from c000 instead of c002, so that we only modify the cross-correlation term (numerator) compared with the exact definition of β. Another approach we are developing to estimate the cross-correlation is using the jackknife method (see Zhang et al., in preparation).

We apply our method for the secondary cosmology c002. We first compare the mean of the halo power spectrum multipoles from the ABACUSSUMMIT halo catalogues with c002 and from the FASTPM halo catalogues with c000 in Fig. 13. The halo catalogues are at $z = 1.1$ with halo mass larger than $10^{11} \, h^{-1} M_\odot$. The left-hand panels are for the monopoles and the right-hand panels are for the quadrupoles. For the ABACUSSUMMIT with c002, there are only 6 boxes, while for the FASTPM we include all the 25 boxes with the same white
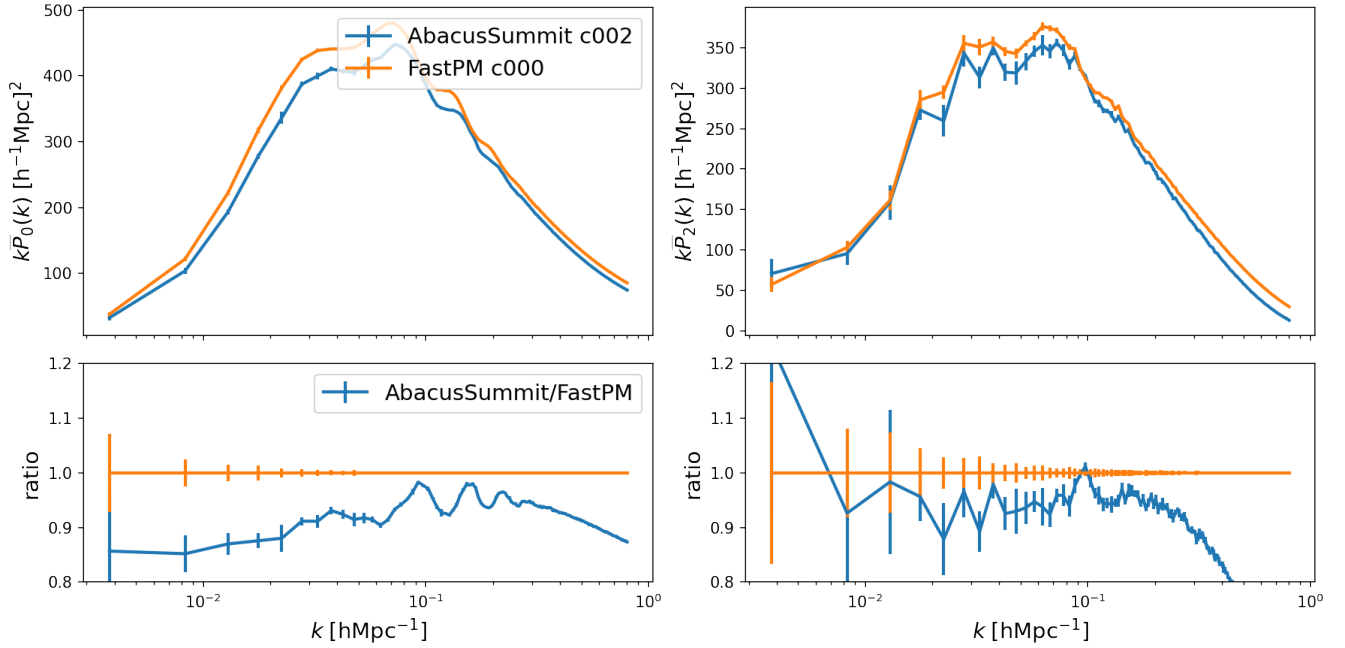

Figure 11. *Top row*: the comparison of the real-space halo reduced bispectra from the ABACUSSUMMIT and FASTPM halo catalogues with mass cut $10^{11} \, h^{-1} M_\odot$. The triangle configuration of the bispectrum is chosen with $k_1 = 0.1 \, h \, \text{Mpc}^{-1}$ and $k_2 = 0.2 \, h \, \text{Mpc}^{-1}$. The upper panel shows the mean of the monopoles averaged over 25 paired ABACUSSUMMIT and FASTPM simulations, respectively. The lower panel shows the ratio between the two. *Middle row*: the diagonal β for the reduced bispectra calculated from equation (7). It is close to 1.0, indicating strong correlation between the bispectra from the ABACUSSUMMIT and FASTPM catalogues. *Bottom row*: the ratio between the mean of the reduced bispectra from ABACUSSUMMIT and CARPool, shown as the blue points. The shaded region shows the standard deviation of the CARPool result.

5 APPLICATION ON SIMULATIONS WITH DIFFERENT COSMOLOGIES

We extend the CARPool method to different cosmologies, even if there is only one ABACUSSUMMIT simulation for a given cosmology. For secondary cosmologies, ABACUSSUMMIT has much fewer number of simulations compared with that of the base cosmology. For the base box size, there are only six simulations for some of the secondary cosmologies (i.e. c00[2−4]), and only one simulation for each of the other secondary cosmologies. If we use the same CARPool method as that in the base cosmology, there are two bottlenecks. One is that we do not have enough ABACUSSUMMIT simulations to pair with FASTPM in order to calculate the cross-correlation between the two. The other is that we need to generate a large number of FASTPM simulations to estimate the mean of clustering statistics, which takes

Figure 13. The comparison of the mean power spectrum multipoles from the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ with the secondary cosmology c002 and from the FᴀsᴛPM with the base cosmology c000. For the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ with c002, there are only 6 boxes, while for the FᴀsᴛPM with c000, there are 25 boxes. The standard deviations of the FᴀsᴛPM multipoles are smaller than those of the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ.
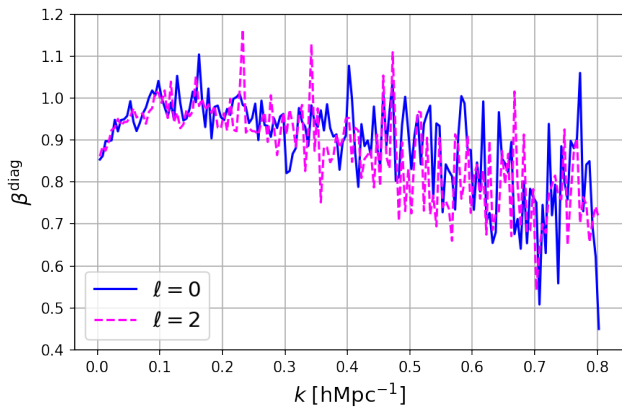

Figure 14. Similar as Fig. 2, but the β^{diag} is calculated from the cross-correlation between the halo power spectrum multipoles from the paired AʙᴀᴄᴜsSᴜᴍᴍɪᴛ in c000 and the FᴀsᴛPM in c002.

noise from the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ in c000, so that the statistical noise of the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ mean is larger than that of the FᴀsᴛPM. In the left-hand panels, we show the ratio of the mean multipoles from the two simulations, denoted as the blue lines. For the monopole, there is BAO residual (with wiggles) caused by the shifting of k coordinates due to the difference of the dark energy parameters w_0 and w_a between the two cosmologies.

We show the diagonal terms of β for the halo power spectrum monopole and quadrupole calculated from equation (27) in Fig. 14. The general shapes are similar to the ones in Fig. 2. We show the results of the multipoles from CARPool in Fig. 15. Same as Fig. 3, the blue points are the mean from six AʙᴀᴄᴜsSᴜᴍᴍɪᴛ halo catalogues.

The orange lines and shaded regions represent the CARPool results. From the ratio between the mean, on the one hand, we do not see any bias of the CARPool result compared with the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ; on the other hand the variance reduction from CARPool is also significant. Therefore, the method of CARPool also applies for the clustering from a different cosmology.

To see the gain of the effective volume from CARPool compared with the default volume of one AʙᴀᴄᴜsSᴜᴍᴍɪᴛ box, we study the situations from one box and from the combination of six boxes, respectively, as shown in Fig. 16. In each panel, the blue line represents the effective volume of one AʙᴀᴄᴜsSᴜᴍᴍɪᴛ box from CARPool with the assumption that the surrogate mean is known. For the realistic case where we estimate the surrogate mean from the non-fixed-amplitude FᴀsᴛPM catalogues with c000, the effective volume decreases by about half on large scales shown as the orange dashed line. On scales up to $k = 0.3\,h\,\mathrm{Mpc}^{-1}$, the gain of the effective volume is above 20 times. For the effective volume from the fixed-amplitude method shown as the green dotted line, it has similar but lower gain compared with the CARPool results, and it quickly decreases to zero at smaller scales $k > 0.1\,h\,\mathrm{Mpc}^{-1}$. We also check the effective volume from the combination of six boxes with CARPool shown as the red dashed line, which has noticeable increase over all the scales compared to the one box case.

We perform the same analysis for the secondary cosmology c004 and find similar results. We show the results in Appendix G.

6 CONCLUSIONS AND DISCUSSIONS

We have prepared a set of FᴀsᴛPM simulations including 25 boxes with the AʙᴀᴄᴜsSᴜᴍᴍɪᴛ ICs, >200 boxes with independent ICs, 237 boxes with the fixed-amplitude ICs, and 2 sets of 25 boxes with different cosmologies.
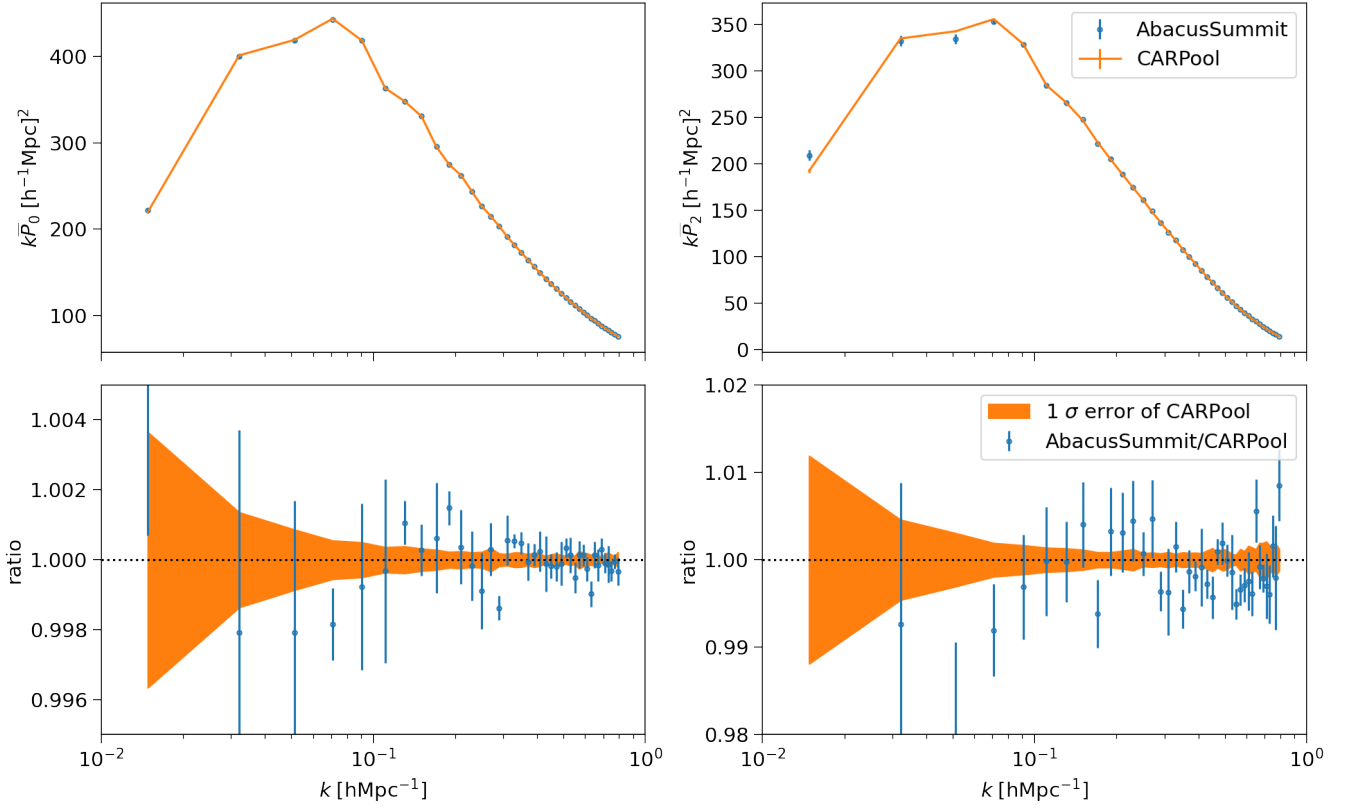
Figure 15. Same as Fig. 3 but for the halo power spectrum multipoles from the secondary cosmology c002. We have also rebinned k to have better clarity for the data points.
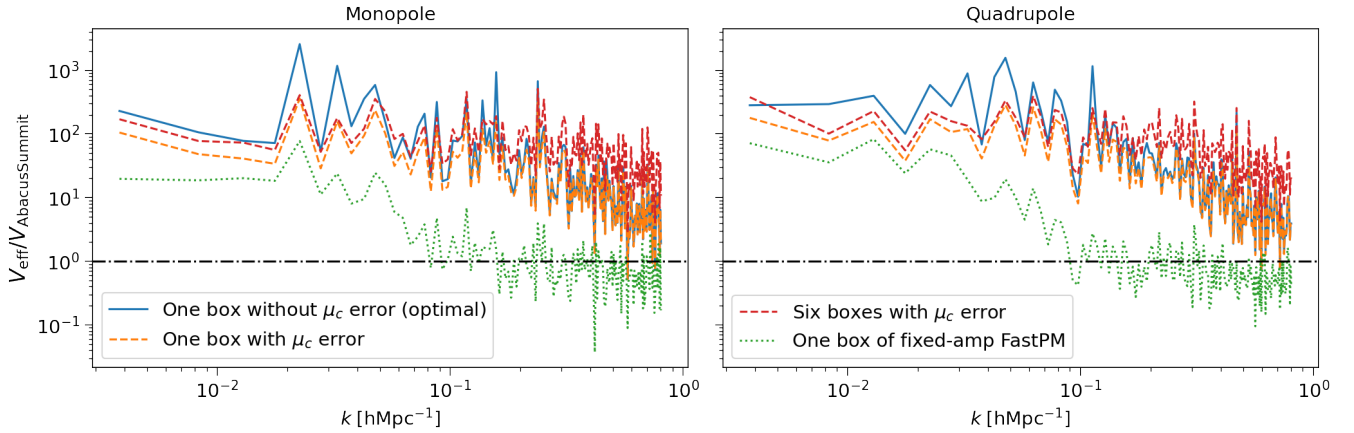

Figure 16. The gain on the effective volume for the halo power spectrum multipoles with the secondary cosmology c002 from CARPool. The left-hand panel is for the monopole and the right-hand panel is for the quadrupole. We show the increase of effective volume in terms of the volume of one AbacusSummit box. In each panel, the blue line denotes the effective volume of one AbacusSummit box with CARPool assuming that the surrogate mean is known. The orange dashed line is the same as the blue line but includes the statistical error of the surrogate mean estimated from the non-fixed-amplitude FastPM catalogues. Similar to the orange dashed line, the red dashed line is the effective volume if we combine six AbacusSummit boxes with CARPool. The green dotted line denotes the effective volume from one fixed-amplitude FastPM box.

We use the FASTPM simulations to improve the precision of ABACUSSUMMIT simulations by adopting the CARPool method. We study the clustering of two halo catalogues at $z = 1.1$, (i) halo mass $> 10^{11} \, h^{-1} M_\odot$ (i.e. hosts of DESI ELGs) and (ii) halo mass $> 10^{13} \, h^{-1} M_\odot$ (i.e. hosts of DESI LRGs). We present the sample variance reduction and the increased effective volume from CARPool. For the ELG-host-halo catalogues, the effective volume is larger than 100 times DESI volume, i.e. 2000 $h^{-3} \mathrm{Gpc}^3$ at $k < 0.3 \, h \, \mathrm{Mpc}^{-1}$ for the power spectrum measurements. We confirm that the fixed-amplitude variance suppressing method can be very effective at smaller k (e.g. $k < 0.1 \, h \, \mathrm{Mpc}^{-1}$) but the performance drops quickly while k increases (e.g. no improvement at $k > 0.2 \, h \, \mathrm{Mpc}^{-1}$). CARPool performs better than the fixed-amplitude method at smaller scales. We also check the effective volume for the correlation function. It is larger than 30 times DESI volume at $r > 10 \, h^{-1} \mathrm{Mpc}$ from CARPool. Similarly, for the LRG-host-halo catalogues, the effective volume is larger than 20 times DESI volume (i.e. 400 $h^{-3} \mathrm{Gpc}^3$) at $k < 0.2 \, h \, \mathrm{Mpc}^{-1}$ and larger than 20 times DESI volume at $r > 10 \, h^{-1} \mathrm{Mpc}$.

Furthermore, CARPool can effectively suppress the sample variance of the halo bispectrum at the region of BAO scales, whereas the fixed-amplitude method cannot, which is consistent with the findings in Chuang et al. (2019); Klypin et al. (2020). Using CARPool, we can obtain larger than 50 times DESI volume for the ELG-host-halo bispectrum with the triangle configuration $k_1 = 0.1 \, h \, \mathrm{Mpc}^{-1}$ and $k_2 = 0.2 \, h \, \mathrm{Mpc}^{-1}$. The trivial gain from the fixed-amplitude method is probably due to the fact that the bispectrum is not mainly contributed from the modulus correlation of the density field, but the phase correlation and phase-modulus cross-correlation (Matsubara 2007; Qin et al. 2022).

We further generalize the method to increase the effective volume of simulations with different cosmologies. These simulations have only one or a few simulations for each cosmology while there are 25 simulations for the primary cosmology. Even with one simulation, we find that the effective volume can be increased by more than 20 times. This finding can be useful when we want to understand how our data analysis pipeline responds to different cosmologies. We can then avoid the systematic bias introduced by a fixed fiducial cosmology model. On the other hand, a tricky step for the generalized method is to estimate the cross-covariance between an ABACUSSUMMIT simulation and the paired FASTPM simulation with different cosmologies. We have used an approximation in this paper but we are developing another approach based on an internal sample evaluation (Zhang et al., in preparation).

The performance of this technique depends on the cross-correlation between the paired N-body and approximate simulations. There are many methods producing approximate simulations, e.g. see Chuang et al. (2015) and Lippich et al. (2019) for various methods, but we do not expect that the other methods can perform better than quasi-N-body codes, e.g. FASTPM, COLA (Tassev et al. 2013), etc. On the other hand, the computing time for generating quasi-N-body simulations is still not negligible; thus, it could be interesting to try other more efficient methods generating approximate simulations.

We have shown that massive approximate gravity calculations combined with a limited number of accurate N-body simulations can be exploited with the CARPool and variance suppression techniques to obtain accurate error estimates on the two-point and three-point statistics of halo clustering. For future work, we can extend CARPool to ABACUSSUMMIT galaxy mocks constructed for DESI observables. CARPool can help to improve the constraints on galaxy clustering models which are related with cosmological parameters and galaxy–halo connection. Since we do not expect a good one-to-one match

between the low-mass haloes in ABACUSSUMMIT and FASTPM runs, the galaxy assignment on FASTPM side may not provide much gains on the cross-correlation. We are expecting that the stochasticity in the galaxy–halo connection might reduce the performance of CARPool at small scales. On the other hand, we expect that CARPool is likely to perform better on the constraints of galaxy–halo connection than the fixed-amplitude method (with pairing), since the latter one fails to reduce the sample variance of halo power spectrum at $k > 0.2 \, h \, \mathrm{Mpc}^{-1}$ (also see Avila & Gutierrez Adame 2022) and halo bispectrum. All in all, it is worth investigating on CARPool with galaxy clustering in the next step. This work sets the path for a robust cosmological analysis of galaxy surveys.

DATA AVAILABILITY

The ABACUSSUMMIT simulations used in this study are publicly available (https://abacusnbody.org/). We plan to make the FASTPM simulations available along with one of the DESI data releases. The halo clustering statistics used in this study, including correlation functions,

power spectra and bispectra from the AbacusSummit and FastPM simulations, is available at https://doi.org/10.5281/zenodo.5993283.

REFERENCES

Alam S., et al., 2021, J. Cosmology Astropart. Phys., 2021, 050
Allison R., Caucal P., Calabrese E., Dunkley J., Louis T., 2015, Phys. Rev. D, 92, 123535
Angulo R. E., Pontzen A., 2016, MNRAS, 462, L1
Avila S., Gutierrez Adame A., 2022, arXiv e-prints, p. arXiv:2204.11103
Avila S., et al., 2020, MNRAS, 499, 5486
Avramidis A. N., Wilson J. R., 1993, Operations Research Letters, 14, 187
Bayer A. E., Banerjee A., Seljak U., 2021a, arXiv e-prints, p. arXiv:2108.04215
Bayer A. E., et al., 2021b, ApJ, 919, 24
Bayer A. E., Banerjee A., Feng Y., 2021c, J. Cosmology Astropart. Phys., 2021, 016
Bose S., Eisenstein D. J., Hadzhiyska B., Garrison L. H., Yuan S., 2022, MNRAS, 512, 837
Chartier N., Wandelt B., Akrami Y., Villaescusa-Navarro F., 2021, MNRAS, 503, 1897
Chuang C.-H., et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 686–700
Chuang C.-H., et al., 2019, MNRAS, 487, 48
Copeland E. J., Sami M., Tsujikawa S., 2006, International Journal of Modern Physics D, 15, 1753
DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036
Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273
Font-Ribera A., McDonald P., Mostek N., Reid B. A., Seo H.-J., Slosar A., 2014, Journal of Cosmology and Astroparticle Physics, 5, 023
Garrison L. H., Eisenstein D. J., Ferrer D., Metchnik M. V., Pinto P. A., 2016, MNRAS, 461, 4125
Garrison L. H., Eisenstein D. J., Ferrer D., Tinker J. L., Pinto P. A., Weinberg D. H., 2018, ApJS, 236, 43
Garrison L. H., Eisenstein D. J., Pinto P. A., 2019, MNRAS, 485, 3370
Garrison L. H., Eisenstein D. J., Ferrer D., Maksimova N. A., Pinto P. A., 2021, MNRAS, 508, 575
Gonzalez-Perez V., et al., 2018, MNRAS, 474, 4024
Grove C., et al., 2021, arXiv e-prints, p. arXiv:2112.09138
Hadzhiyska B., Eisenstein D., Bose S., Garrison L. H., Maksimova N., 2022, MNRAS, 509, 501
Hahn C., Villaescusa-Navarro F., 2021, J. Cosmology Astropart. Phys., 2021, 029
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Hernández-Aguayo C., Prada F., Baugh C. M., Klypin A., 2021, MNRAS, 503, 2318
Huterer D., et al., 2015, Astroparticle Physics, 63, 23
Klypin A., Prada F., Byun J., 2020, MNRAS, 496, 3862
Kreisch C. D., Pisani A., Villaescusa-Navarro F., Spergel D. N., Wandelt B. D., Hamaus N., Bayer A. E., 2021, arXiv e-prints, p. arXiv:2107.02304
Lesgourgues J., Pastor S., 2006, Physics Reports, 429, 307–379
Lippich M., et al., 2019, Monthly Notices of the Royal Astronomical Society, 482, 1786
Maion F., Angulo R. E., Zennaro M., 2022, arXiv e-prints, p. arXiv:2204.03868
Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, MNRAS, 508, 4017
Massara E., Villaescusa-Navarro F., Ho S., Dalal N., Spergel D. N., 2021, Phys. Rev. Lett., 126, 011301
Matsubara T., 2007, ApJS, 170, 1
Metchnik M. V. L., 2009, PhD thesis, The University of Arizona
Peebles P. J. E., Hauser M. G., 1974, ApJS, 28, 19
Planck Collaboration et al., 2020, A&A, 641, A6
Pontzen A., Slosar A., Roth N., Peiris H. V., 2016, Phys. Rev. D, 93, 103519
Porta Nova A. M., Wilson J. R., 1993, European Journal of Operational Research, 71, 80
Qin J., Pan J., Yu Y., Zhang P., 2022, MNRAS, 514, 1548
Quinn T., Katz N., Stadel J., Lake G., 1997, arXiv e-prints, pp astro–ph/9710043
Rubinstein R. Y., Marcus R., 1985, Operations Research, 33, 661
Scoccimarro R., 2015, Phys. Rev. D, 92, 083532
Sefusatti E., Scoccimarro R., 2005, Phys. Rev. D, 71, 063001
Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624
Sinha M., Garrison L. H., 2019, in Software Challenges to Exascale Computing. Second Workshop. pp 3–20 (arXiv:1911.08275), doi:10.1007/978-981-13-7729-7_1
Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmology Astropart. Phys., 2013, 036
Villaescusa-Navarro F., et al., 2018, ApJ, 867, 137
Villaescusa-Navarro F., et al., 2020, ApJS, 250, 2
Zhao C., et al., 2021, MNRAS, 503, 1149
Zhou R., et al., 2021, MNRAS, 501, 3309

APPENDIX A: CONFIGURATION PARAMETERS OF THE FASTPM SIMULATION

Before the massive production of the FastPM simulations, we run some tests on the main configuration parameters, i.e. the force resolution parameter B, the time-steps T, as well as the redshift of the IC. We would like to find a set of configuration parameters that can make the FastPM simulation as close to the AbacusSummit simulation as possible, based on their matter power spectra. Meanwhile, such configuration should not be expensive to run, since even for FastPM, running a $2\,h^{-1}$Gpc box with mass resolution $5 \times 10^9\,h^{-1}\mathrm{M}_\odot$ still takes considerable computational resources.

In fact, AbacusSummit has generated many simulations with box size $500\,h^{-1}$Mpc, which can be used to check the FastPM performance. Using the IC from one such AbacusSummit box, we generate different FastPM simulations by varying the configuration parameters. To reduce the sample variance, we fix the amplitude of the IC in the FastPM simulations. We set the same particle mass resolution as that of the FastPM base box ($2\,h^{-1}$Gpc). Thanks to the small box size, it is relatively cheap and convenient to generate many realizations for the test purpose. We believe that a set of reasonable configuration parameters we find for a small box still holds for a larger box with the same particle mass resolution. To determine the FastPM performance from a given set of configuration parameters, we compare the real-space matter power spectrum with that of AbacusSummit at redshift $z = 0.2$. We show the comparison in Fig. A1–A3. In each figure, we vary one parameter and fix the other two.

In Fig. A1, we show the impact from B while fixing $T = 40$ and the initial redshift $z_0 = 19$, i.e. the initial scale $a_0 = 0.05$. Note that in our case T is always counted as the number of time-steps from a given initial redshift to the finial redshift 0.1 with the linear step size in scale a. Different coloured lines represent the results from different Bs. The black dotted line denotes the result from AbacusSummit, which we take as the standard. In the upper panel, we show the matter power spectra at $z = 0.2$ from AbacusSummit as well as from FastPM with different Bs. In the lower panel, we show the ratio between the power spectra from FastPM and AbacusSummit. Since the embedded routines for calculating matter power spectra in FastPM and AbacusSummit are not exactly the same, the number of modes from some k bins varies a bit due to different counting schemes for the k modes on the bin boundaries, hence, it can cause some fluctuation at very large scales where the number of modes is small. We can ignore such fluctuation from numerical issues. Based
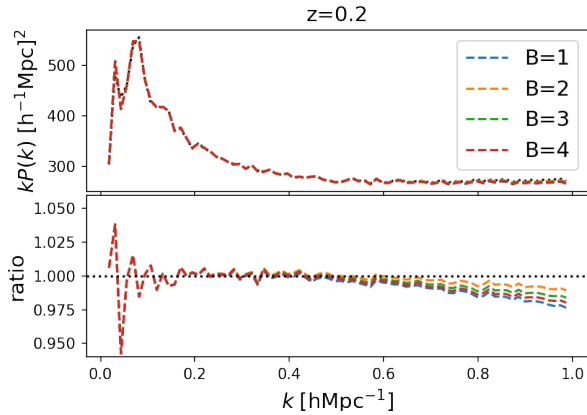
Figure A1. The comparison between the matter power spectra at $z = 0.2$ from the AbacusSummit (black dotted line) and the FastPM (coloured lines) with different force resolution parameters Bs but the same $T = 40$ and $a_0 = 0.05$. The upper panel shows the overall shapes up to $k = 1.0\,h\,\mathrm{Mpc}^{-1}$. The lower panel shows the ratio between the FastPM and the AbacusSummit which we take as the standard. Different Bs cause the deviations of FastPM on small scales differently. In this case, $B = 2$ performs the best.

on the ratio, we see the deviation of the FastPM power spectra from the AbacusSummit on small scales. The deviation from $B = 1$ is the largest, about 2.5%, as it has the lowest force resolution. Interestingly, $B = 2$ gives the lowest deviation (about 1% at $k = 1.0\,h\,\mathrm{Mpc}^{-1}$) than the cases with $B = 3$ or 4. The reason is beyond our knowledge, which we leave it for future study.

In Fig. A2, we vary $T = 20, 40, 80$, but fix $B = 2$ and $a_0 = 0.05$. The performance from $T = 40$ is comparable with that from $T = 80$, but with about half of the computational time of $T = 80$. $T = 20$ gives relatively large deviation at scales smaller than $k = 0.5\,h\,\mathrm{Mpc}^{-1}$. In Fig. A3, we vary the initial redshift from 99, 19 and 9, respectively, while fixing $B = 2$ and $T = 40$. A larger initial redshift gives a more accurate IC from 2LPT, but results in a larger step size given a fixed number of time-steps. It turns out that $a_0 = 0.05$ gives better performance compared with the other two.

Based on the above tests, we choose our configuration parameters as $B = 2$, $T = 40$, and $a_0 = 0.05$. In fact, given a set of parameters, the performance would vary depending on redshifts. It is not guaranteed that the parameters we choose give better performance at some low redshifts than other choices. However, we do not strictly examine which set of parameters are the best, but try to find a reasonable one based on our goals.

APPENDIX B: THE EFFECT FROM MASS CUT, ABUNDANCE MATCHING, AND HALO CLEANING

Given a halo mass cut, e.g. $10^{11}\,h^{-1}\mathrm{M}_\odot$, the halo number density of FastPM is about 36% lower than that of AbacusSummit (uncleaned). To check whether the difference of number densities will reduce the CARPool performance or not, we select high-mass haloes from the paired FastPM catalogues to match with the number of haloes from AbacusSummit with mass cut. We compare the control variant β^{diag} of the power spectrum monopoles from mass cut and abundance matching, and show the result in Fig. B1. The improved performance of abundance matching (orange dashed line) is not very significant compared with that of mass cut (blue solid line). We have checked the conclusion is true for the quadrupole too. Furthermore,
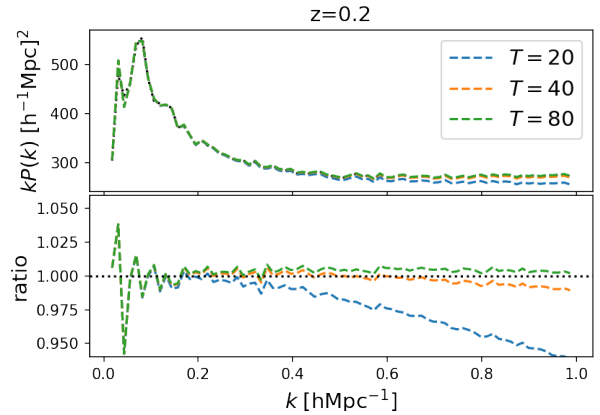
Figure A2. Similar to Fig. A1 but for the test of the number of time-steps T. We fix $B = 2$ and $a_0 = 0.05$. Considering the accuracy and computation time, $T = 40$ performs better than $T = 20$ and 80.
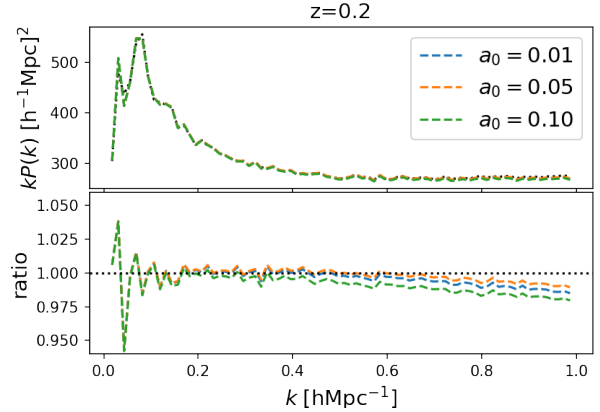

Figure A3. Similar to Fig. A1 but for the test of the initial scale a_0. We fix $B = 2$ and $T = 40$. The initial scale $a_0 = 0.05$, i.e. $z_0 = 19$, performs better than the other two.

in the case of mass cut, we replace the uncleaned AbacusSummit halo catalogues by the cleaned ones and obtain the black dotted line. On large scales $k < 0.2\,h^{-1}\mathrm{Mpc}$, β^{diag} seems performing better than that from the uncleaned haloes as it is closer to 1.0. We expect that using the cleaned AbacusSummit haloes will benefit the CARPool performance on large scales.

In Fig. B2, we show the ratio of the power spectrum multipoles between the cleaned AbacusSummit and FastPM catalogues with mass cut $10^{11}\,h^{-1}\mathrm{M}_\odot$. Compared with Fig. 1, the difference of halo biases at large scales between AbacusSummit and FastPM reduces significantly from 4.2% to 1.2% after the halo cleaning in AbacusSummit.

APPENDIX C: CARPOOL PERFORMANCE OF LRG-HOST HALOES AT REDSHIFT 0.8

For DESI LRGs, the number density peaks around redshift 0.6–0.8 (Zhou et al. 2021). We simply check the performance of CARPool for the LRG-host haloes (with mass larger than $10^{13}\,h^{-1}\mathrm{M}_\odot$) at $z = 0.8$
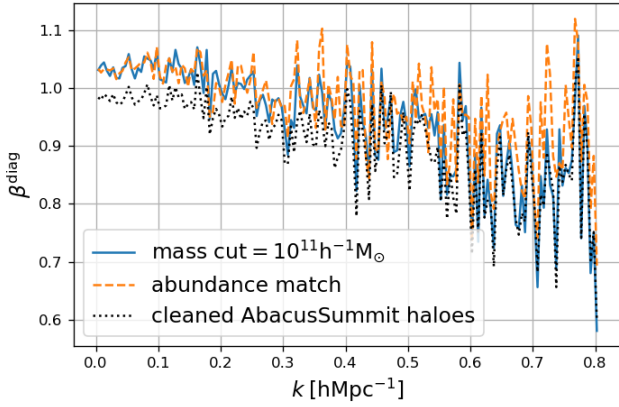
Figure B1. The comparison of β^{diag} of the power spectrum monopoles calculated from the case of mass cut equal to $10^{11}\ h^{-1}\mathrm{M}_\odot$ versus the case of abundance matching. For the abundance matching, we match the number of FastPM haloes close to that of AbacusSummit with mass cut. In addition, we overplot the result (black dotted line) from the case if we use the cleaned version of AbacusSummit halo catalogues.
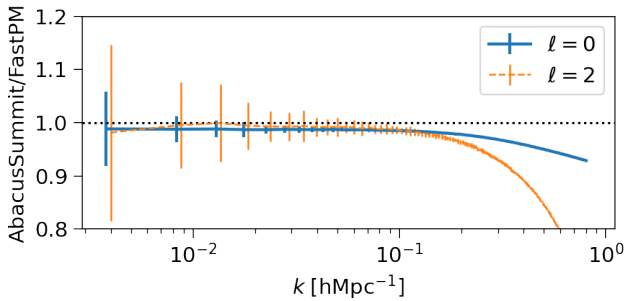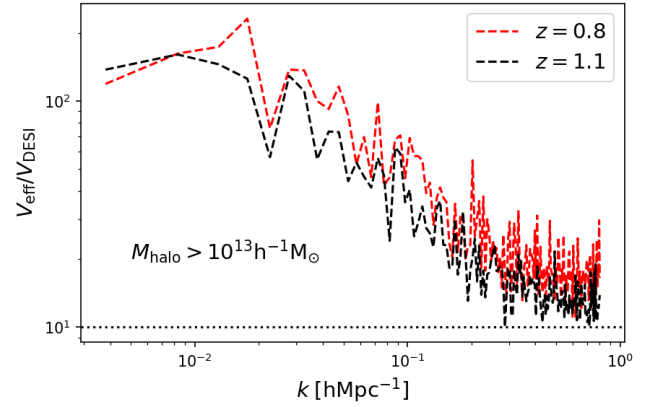

Figure C1. The comparison of the effective volumes obtained from CARPool for the halo power spectrum monopoles at redshifts 0.8 and 1.1. The haloes are massive than $10^{13}\ h^{-1}\mathrm{M}_\odot$.

Figure B2. The ratio of the mean power spectrum monopoles ($\ell = 0$) and quadrupoles ($\ell = 2$) between the cleaned AbacusSummit and FastPM catalogues with mass cut $10^{11}\ h^{-1}\mathrm{M}_\odot$. We slightly shift the k coordinates of the quadrupole for clarity.

spectrum multipoles and the covariance matrices for AbacusSummit and FastPM with the matched AbacusSummit ICs. We also calculate the cross-covariance matrix of the multipoles from the two simulations, and obtain the Pearson correlation coefficients via

$$\rho_{\mathrm{A,F}}(k_i, k_j) = \frac{\mathrm{Cov}_{\mathrm{A,F}}(k_i, k_j)}{\sqrt{\mathrm{Cov}_{\mathrm{A}}(k_i, k_i)\mathrm{Cov}_{\mathrm{F}}(k_j, k_j)}}, \qquad (\mathrm{D1})$$

where the subscripts A and F denote AbacusSummit and FastPM, respectively. In Fig. D1, we show the diagonal terms of ρ_{ij} for the monopole (left-hand panel) and quadrupole (right-hand panel) with two mass cuts. For both monopole and quadrupole, the Pearson coefficients are closer to 1.0 from the halo mass cut $10^{11}\ h^{-1}\mathrm{M}_\odot$, which indicates better performance from CARPool.

APPENDIX E: STATISTIC NOISE OF THE CORRELATION FUNCTION QUADRUPOLE

We investigate the reason why there is about $2 \sim 3\sigma$ bias between the mean of the correlation function quadrupoles from the AbacusSummit halo catalogues and that from CARPool around $s = 50\ h^{-1}\mathrm{Mpc}$ in Fig. 8.

We find that it is due to the statistical noise of the mean of the paired FastPM catalogues. In the upper panel of Fig. E1, we show the mean of the quadrupoles from 200 FastPM halo catalogues with random ICs as the black dotted line. We divide these 200 catalogues into 8 groups, each of which has 25 catalogues. We calculate the mean for each group and plot it as a gray line. We also plot the mean of the paired FastPM catalogues with the AbacusSummit ICs as the red dashed line with error bars. In the lower panel, we divide all the results by the mean of 200 catalogues. We see that the fluctuation amplitude of the red dashed line is comparable with that of the gray lines. The deviation of the red dashed line from the black dotted line around $s = 50\ h^{-1}\mathrm{Mpc}$ is about $2 \sim 3\sigma$. Based on CARPool, we have the ratio of the mean between the AbacusSummit catalogues and that from CARPool as

$$\frac{\overline{y}}{\overline{x}} = 1 + \beta\frac{\overline{c} - \hat{\mu}_c}{\overline{x}}, \qquad (\mathrm{E1})$$

where $\overline{c}$ and $\hat{\mu}_c$ correspond to the red dashed line and black dotted line, respectively. Since we have checked that β is close to 1.0 for

compared with that from $z = 1.1$. The number density of AbacusSummit haloes increases from 1.7×10^{-4} to $2.4 \times 10^{-4}\ h^3\mathrm{Mpc}^{-3}$ from $z = 1.1$ to 0.8. We compare the effective volumes from CARPool at the two redshifts in Fig. C1. We show the results with the surrogate mean from the fixed-amplitude FastPM catalogues. Overall there is an average of $\sim 35\%$ increase on V_{eff} at $z = 0.8$, thanks to the increase of the halo number density. Here we use the surrogate mean from the fixed-amplitude FastPM catalogues. The increase is similar for the case using the surrogate mean from the non-fixed-amplitude catalogues.

APPENDIX D: THE PEARSON CORRELATION COEFFICIENTS OF THE HALO POWER SPECTRUM MULTIPOLES BETWEEN ABACUSSUMMIT AND FASTPM

To compare the performance of CARPool for the halo power spectrum multipoles with different halo mass cuts (10^{11} and $10^{13}\ h^{-1}\mathrm{M}_\odot$) at $z = 1.1$, we check the Pearson correlation coefficients between AbacusSummit and FastPM. For each mass cut, we calculate the power
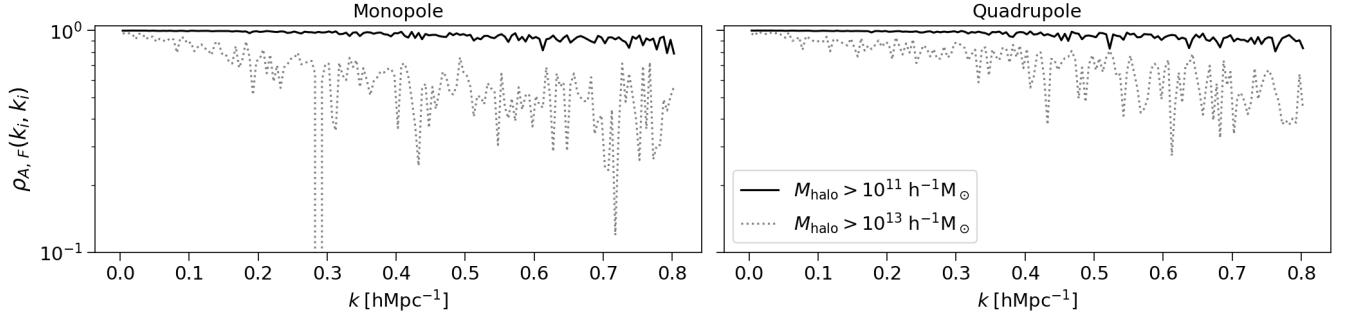
Figure D1. The diagonal terms of the Pearson correlation coefficients of the halo power spectrum multipoles between AbacusSummit and FastPM. The left-hand (right-hand) panel is for the monopole (quadrupole). The solid (dotted) lines are from the halo mass cut $10^{11}\ h^{-1}M_\odot$ ($10^{13}\ h^{-1}M_\odot$).
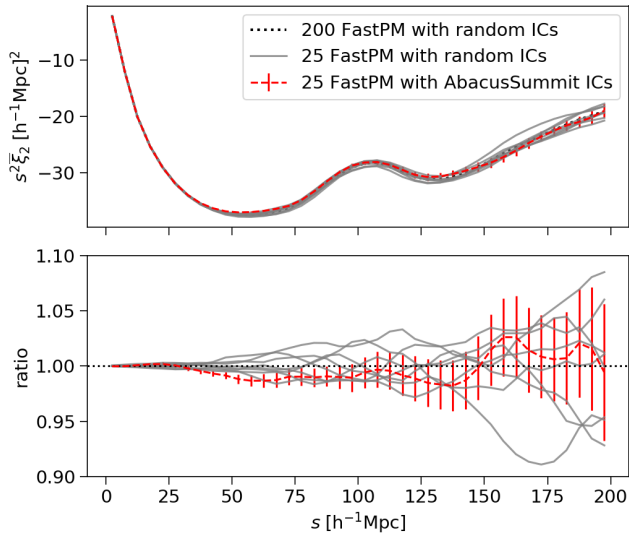

Figure E1. *Upper panel*: we compare the correlation function quadrupoles from the FastPM halo catalogues with different cases. The dotted line shows the mean of the quadrupoles from 200 FastPM catalogues with random ICs. Each gray line represents the mean of every 25 out of 200 catalogues. The red dashed line is the mean of 25 FastPM catalogues with the AbacusSummit ICs. *Lower panel*: we show the ratio of the result from each case by the mean of 200 FastPM catalogues. The red dashed line fluctuates around the horizontal dotted line and the fluctuation amplitude is similar to that of gray lines. At around 50 h^{-1}Mpc, there is about $2 \sim 3\ \sigma$ deviation from 1.0, which is just due to the statistical noise.

$s > 20\ h^{-1}$Mpc, the statistical bias between $\bar{c}$ and $\hat{\mu}_c$ directly relates to the bias between the mean of AbacusSummit and CARPool.

APPENDIX F: CROSS-CORRELATION BETWEEN THE CLUSTERING FROM DIFFERENT COSMOLOGIES

To test the validity of equation (25), we use the halo catalogues of 25 paired FastPM simulations which are from the two cosmologies c000 and c002. Each paired simulation shares the same IC. We use the halo catalogues at $z = 1.1$ with mass larger than $10^{11}\ h^{-1}M_\odot$. We divide haloes into three groups based on halo mass, i.e. $10^{11} < M_{\text{halo}} < 10^{12}\ h^{-1}M_\odot$, $10^{12} < M_{\text{halo}} < 10^{13}\ h^{-1}M_\odot$, and $M_{\text{halo}} > 10^{13}\ h^{-1}M_\odot$. We use cat1, cat2 and cat3 to represent them. As is

known that the number of haloes decreases as the halo mass increases. We conduct subsampling for cat1 and cat2 with the percentage 2.85% and 28%, respectively. The number of haloes after subsampling is about 7.2 million for cat2. At the end, we reach two goals: one is that after subsampling, the number of the haloes from the combined cat1 and cat3, denoted as cat13, is close to that of cat2; the other is that the halo clustering statistics is similar between cat2 and cat13. Doing such process, we can mimic cat2 as a catalogue from an N-body simulation and cat13 as a catalogue from a paired surrogate, since cat13 is constructed to mimic cat2 in terms of the shot noise and two-point clustering signal and it shares the same IC with cat2. We obtain cat2 and cat13 from 25 halo catalogues in each cosmology. We calculate the halo power spectrum multipoles from cat2 and cat13, and compare β_1 and β_2 calculated from the cross-correlation and variance of the multipoles over 25 realizations, i.e.

$$\beta_1 = \frac{\text{Diag}[\text{Cov}(\text{cat2}_{c002}, \text{cat13}_{c000})]}{\text{Var}(\text{cat13}_{c000})}, \tag{F1}$$

$$\beta_2 = \frac{\text{Diag}[\text{Cov}(\text{cat2}_{c000}, \text{cat13}_{c002})]}{\text{Var}(\text{cat13}_{c000})}. \tag{F2}$$

We show the results in Fig. F1, in which the blue lines denote β_1 and the orange dotted lines denote β_2. They have a similar shape for both monopole and quadrupole.

APPENDIX G: CARPOOL RESULTS OF THE SECONDARY COSMOLOGY C004

Similar to Section 5, we apply CARPool on the halo catalogues with the secondary cosmology c004. We compare the halo power spectrum multipoles from c004 and c000 in Fig. G2. Compared with Fig. 13, we can see that the difference of the power spectrum monopoles between c004 and c000 is smaller than that between c002 and c000, which is expected as c004 is different from c000 only on σ_8. We show the β in Fig. G1, and the effective volume increased from CARPool in Fig. G3.

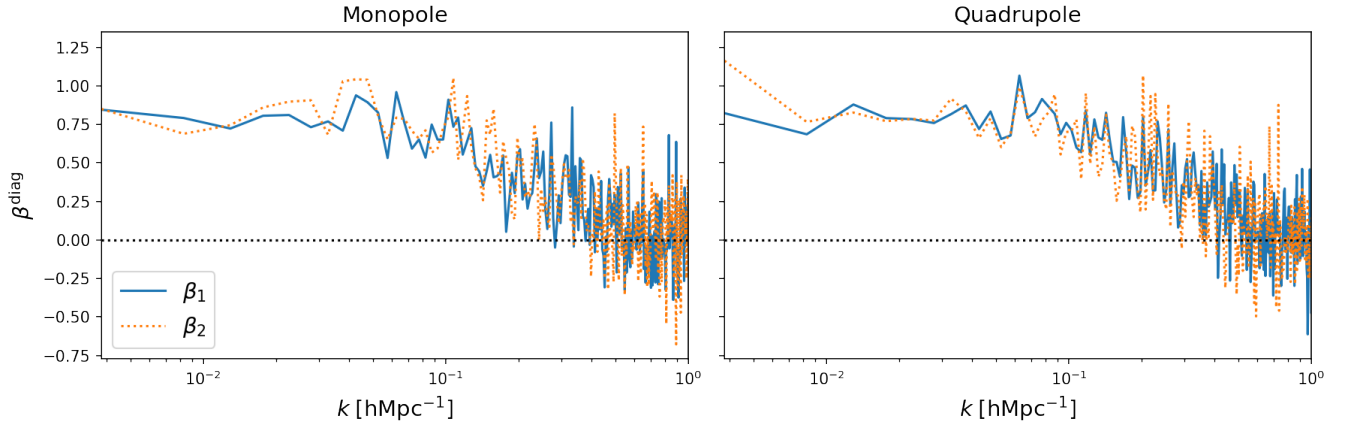This paper has been typeset from a TEX/LATEX file prepared by the author.

Figure F1. The shape of β^{diag} which is calculated from the cross-correlation of the halo power spectrum multipoles between the cat2 and cat13 sharing the same ICs but with different cosmologies. β_1 is calculated from the cross-correlation between the cat2 with the cosmology c002 and the cat13 with the cosmology c000, while β_2 is calculated from the cat2 with c000 and the cat13 with c002. They agree well on the general shape for both monopole and quadrupole.
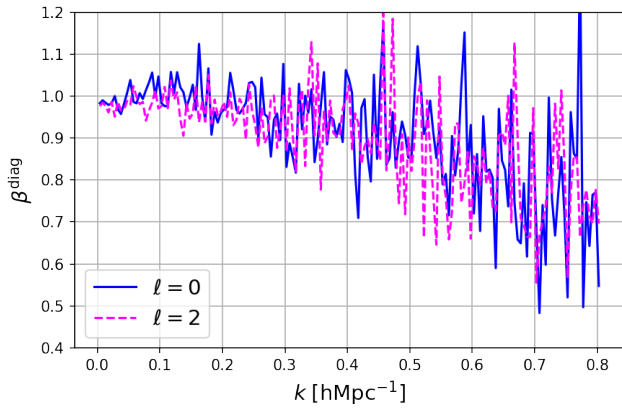

Figure G1. β^{diag} is calculated from the cross-correlation between the halo power spectrum multipoles from the paired AbacusSummit in c000 and the FastPM in c004. $\ell = 0$ (2) is for the monopole (quadrupole).
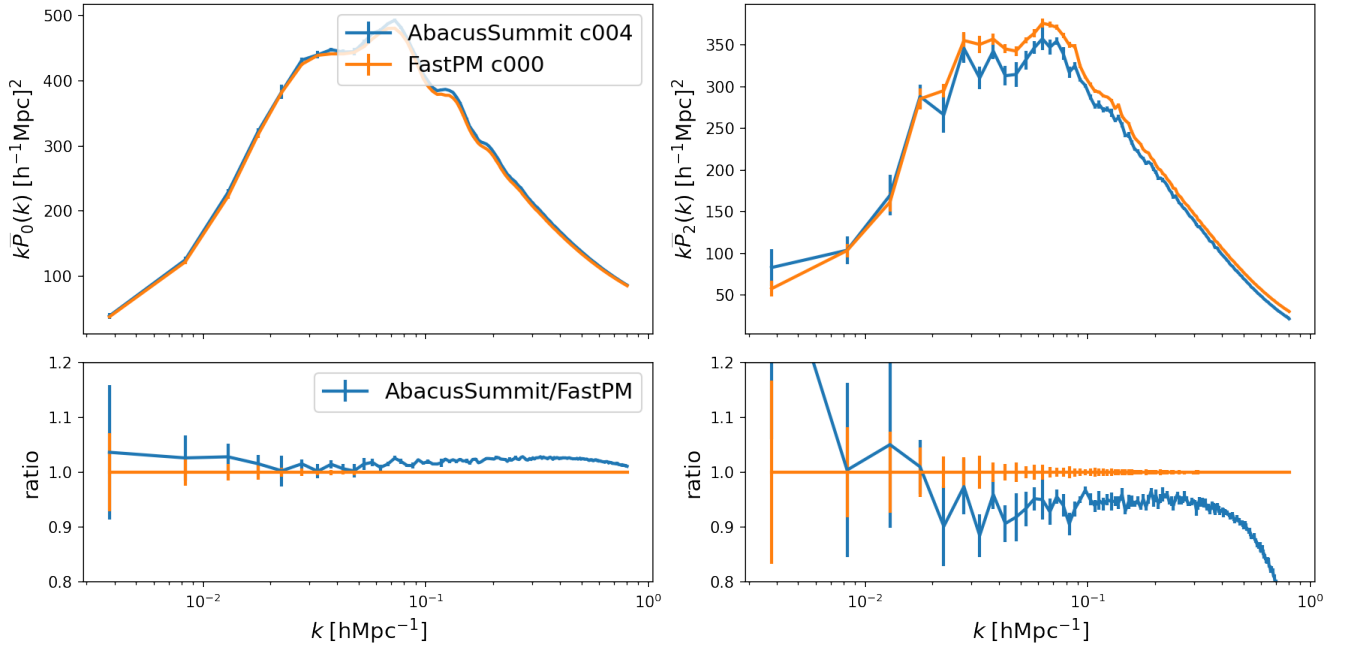
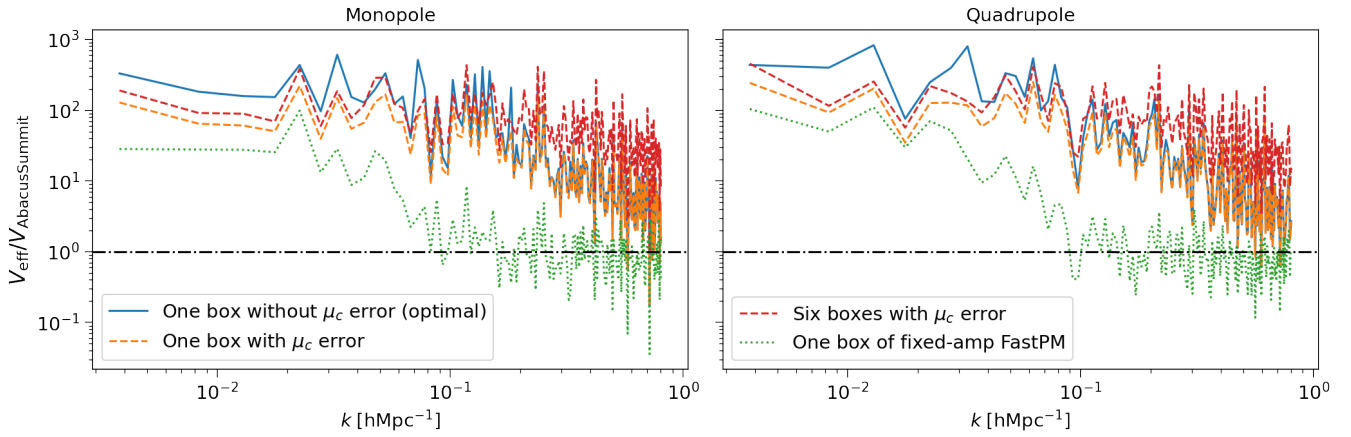**Figure G2.** Same as Fig. 13 but for the secondary cosmology c004.

Figure G3. Same as Fig. 16 but for the secondary cosmology c004.