Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **485,** 2806–2824 (2019)
Advance Access publication 2019 February 19

doi:10.1093/mnras/stz507

Comparing approximate methods for mock catalogues and covariance matrices II: power spectrum multipoles

Linda Blot [1,2]★ Martin Crocce [1,2] Emiliano Sefusatti [3,4] Martha Lippich,[5,6] Ariel G. Sánchez,[5] Manuel Colavincenzo,[7,8,9] Pierluigi Monaco,[3,4,9] Marcelo A. Alvarez,[10] Aniket Agrawal,[11] Santiago Avila [12] Andrés Balaguera-Antolínez,[13,14] Richard Bond,[15] Sandrine Codis,[15,16] Claudio Dalla Vecchia,[13,14] Antonio Dorta,[13,14] Pablo Fosalba,[1,2] Albert Izard,[1,2,17,18] Francisco-Shu Kitaura,[13,14] Marcos Pellejero-Ibanez,[13,14] George Stein,[15] Mohammadjavad Vakili [19] and Gustavo Yepes[20,21]

Affiliations are listed at the end of the paper

Accepted 2019 February 14. Received 2019 February 14; in original form 2018 June 28

ABSTRACT

We study the accuracy of several approximate methods for gravitational dynamics in terms of halo power spectrum multipoles and their estimated covariance matrix. We propagate the differences in covariances into parameter constraints related to growth rate of structure, Alcock–Paczynski distortions, and biasing. We consider seven methods in three broad categories: algorithms that solve for halo density evolution deterministically using Lagrangian trajectories (ICE–COLA, PINOCCHIO, and PEAKPATCH), methods that rely on halo assignment schemes on to dark matter overdensities calibrated with a target N-body run (HALOGEN, PATCHY), and two standard assumptions about the full density probability distribution function (Gaussian and lognormal). We benchmark their performance against a set of three hundred N-body simulations, running similar sets of approximate simulations with matched initial conditions, for each method. We find that most methods reproduce the monopole to within 5 per cent, while residuals for the quadrupole are sometimes larger and scale dependent. The variance of the multipoles is typically reproduced within 10 per cent. Overall, we find that covariances built from approximate simulations yield errors on model parameters within 10 per cent of those from the N-body-based covariance.

Key words: methods: data analysis – methods: numerical – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

The study of the large-scale structure of the Universe has seen a major step up with the completion of the cosmological analysis of large galaxy redshift surveys such as SDSS (Tegmark et al. 2004; Eisenstein et al. 2005), 2dFGRS (Percival et al. 2001; Cole et al. 2005), and more recently BOSS (Alam et al. 2017) and VIPERS (Pezzotta et al. 2017). We have entered an era in which the accuracy on cosmological parameters from the analysis of low-redshift tracers becomes comparable to the one reached by cosmic microwave background experiments, allowing tight constraints on parameters related to the late-time cosmic acceleration. This observational effort continues nowadays with the analysis of state-of-the-art imaging surveys for weak lensing measurements: DES (Troxel et al. 2017), KiDS (Hildebrandt et al. 2017), and HSC (Mandelbaum et al. 2018), as well as campaigns reaching high-redshift tracers such as the Lyman α forest and Quasars in eBOSS (Dawson et al. 2016; Ata et al. 2018). Even more excitingly, the near future will see larger observational campaigns such as *Euclid* (Laureijs et al. 2011), DESI (DESI Collaboration 2016), or LSST (Ivezic et al. 2008; LSST Science Collaboration 2009) that will enable a better understanding of several open questions, in addition to the nature of cosmic acceleration, such as the physics of the primordial Universe and the neutrino mass scale. In parallel, the requirements on the accuracy in the evaluation of systematics and statistical errors have grown to become an important component of the error budget. In that sense, ensembles of galaxy mock catalogues

★ E-mail: blot@ice.cat

© 2019 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

are a vital component for the analysis of galaxy surveys, not only for their internal use but also for assessing the level of agreement among different data sets. Mock catalogues are useful in at least three ways: (1) to study the impact of observational systematic effects, (2) to test science pipelines and analysis methodologies (including the recovery of the input cosmology of the mocks), and (3) to provide covariance matrices among observables (e.g. accounting for cosmic variance, noise, masking, and correlated survey property fluctuations).

N-body simulations coupled with semi-analytical models or halo occupation distribution (HOD) schemes for galaxy assignment are perhaps the best possible route nowadays to construct realistic galaxy catalogues (Carretero et al. 2015; Smith et al. 2017; Takahashi et al. 2017). However, running ensembles of high-resolution N-body simulations of cosmological size is computationally very expensive (Angulo et al. 2012; Fosalba et al. 2015; Heitmann et al. 2015; Potter, Stadel & Teyssier 2017), in particular for future surveys such as *Euclid* or DESI, which will cover large volumes and will need high particle mass resolution to reach the expected galaxy number density in the observations.

One alternative route is to run fast algorithms (also known as 'approximate methods', see Monaco 2016 for an overview) that reproduce, to a variable extent, the large-scale statistics of N-body simulations. This is done at the expense of losing accuracy in the small-scale physics since these methods are not able to resolve halo substructures. These approximate methods have been widely used over the past few years to help bring forward galaxy clustering analysis in a more comprehensive way (Manera et al. 2013, 2015; Howlett et al. 2015; Kitaura et al. 2016; Koda et al. 2016; Avila et al. 2017). None the less, an assessment of how those fast methods reproduce the covariance from N-body simulations, or their impact on derived cosmology, is still missing in the literature. Generally speaking, the requirements for future missions, such as *Euclid*, are that systematic errors in the estimation of covariance matrices should not bias the estimation of errors in cosmological parameters by more than $\sim$10 per cent.

Hence, in this article we perform a robust and thorough comparison of clustering measurements in ensembles of mock catalogues produced from several state-of-the-art algorithms, which span basically all the various types of approaches available in the literature. We use as benchmark an ensemble of 300 large N-body simulations, short named Minerva (Grieb et al. 2016). As noted previously, it is the first time this kind of study is performed. On the one hand, we concentrate on how well the observables are reproduced with high precision because of the large number of realizations (see a previous comparison work by Chuang et al. 2015b using one N-body simulation). On the other hand, because we extend the comparison to how well these methods reproduce the full covariance matrix of power spectrum multipoles, and how those inaccuracies propagate into constraints on the parameters of interest for galaxy redshift surveys. Two companion articles present similar studies in terms of two-point correlation function (Lippich et al. 2019) and bispectrum (Colavincenzo et al. 2019).

We span different types of approaches: one representing the combination of multistep second-order Lagrangian perturbation theory (2LPT) to solve the large-scale displacements with a fast particle mesh (PM) solver to speed up the intermediate-scale force computation, ICE–COLA (Izard, Crocce & Fosalba 2016; Izard, Fosalba & Crocce 2018). We then consider one group of methods based on halo finding in Lagrangian space coupled with LPT for the evolution equations of the collapsed regions. In this category, we consider PINOCCHIO (Monaco et al. 2002) and PEAKPATCH (Bond & Myers 1996a,b,c; Stein, Alvarez & Bond 2018). Another group of methods uses LPT to evolve the matter field and then applies a biasing scheme to produce a halo sample. For this class, we consider HALOGEN (Avila et al. 2015) and PATCHY (Kitaura, Yepes & Prada 2014). The latter methods need to be calibrated on one N-body simulation before running the desired set of mocks.[1] Finally, we consider approaches that make assumptions on the probability distribution function (PDF) of the density field: a Gaussian model for the covariance (Grieb et al. 2016) and a set of 1000 lognormal mocks (Agrawal et al. 2017). Both of them use as input the actual clustering signal and number density of haloes as measured in the benchmark N-body simulation. There are several other methods in the literature (e.g. QUICKPM by White, Tinker & McBride 2014, FASTPM by Feng et al. 2016, EZMOCKS by Chuang et al. 2015a), but they can all roughly fit into one of the above categories. Hence, we expect our study to lead to conclusions that are of general applicability in the field.

This article is organized as follows: in Section 2, we describe all the approximate methods that we consider and how they compare in terms of computational cost. In Section 3, we introduce the halo samples over which we do the comparison. Section 4 describes our methodology. Section 5 contains and discusses our results, which are summarized in Section 6. Lastly, we include two appendices, one describing an alternative way of defining the halo samples, and another discussing the statistical uncertainty in the comparison of cosmological parameter errors as shown in our results.

2 COMPARED METHODS

We now briefly summarize the main features of the methods and mocks used in this work. For a comprehensive description, we refer the reader to the first article of this series (Lippich et al. 2019), while for a general review we refer to Monaco (2016). For the purpose of presenting results, we will classify the compared methods in three categories: *predictive methods*, which do not require recalibration against a parent N-body for each specific sample or cosmological model (ICE–COLA, PINOCCHIO, and PEAKPATCH); *calibrated methods*, which need prior information about the sample to simulate (HALOGEN and PATCHY); and *analytical methods*, which predict the covariance by making assumptions on the shape of the PDF of the density field (lognormal and Gaussian). A summary is provided in Table 1, where we report the computing requirements (per single mock) for each of the methods used in this work, together with some general considerations and references. The methods are presented in decreasing computing time order.

Our comparison will focus on halo samples defined in comoving outputs at redshift $z = 1$.

2.1 Reference N-body: Minerva

Our benchmark for the comparison is the set of N-body simulations called Minerva, first described in Grieb et al. (2016). For this project, 200 additional realizations were run, for a total of 300 realizations. These are performed using the code GADGET-3 (Springel 2005), with $N_{\rm p} = 1000^3$ particles in a box of linear size $L_{\rm box} = 1500\,h^{-1}$ Mpc and a starting redshift of $z_{\rm ini} = 63$. The mass resolution is therefore $2.67 \times 10^{11}\,h^{-1}\,{\rm M}_{\odot}$. The initial conditions are given by the 2LPTIC

[1]Typically, they require the halo abundance and bias as a function of mass as measured in the parent simulation.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

Table 1. Name of the methods, type of algorithm, halo definition, computing requirements, and references for the compared methods. All computing times are given in cpu-hours per run and memory requirements are per run. We do not include the generation of initial conditions. The computational resources for halo finding in the N-body and ICE–COLA mocks are included in the requirements. The computing time refers to runs down to redshift 1 except for the N-body where we report the time down to redshift 0 (we estimate an overhead of ∼50 per cent between $z = 1$ and $z = 0$). Since every code was run in a different machine, the computing times reported here are only indicative. We include the information needed for calibration/prediction of the covariance where relevant. (∗) In order to resolve the lower mass haloes of the first sample, a higher resolution version of PEAKPATCH should be run, requiring more computational resources than quoted here.

Method	Algorithm	Computational requirements	Reference
Minerva	**N-body**	CPU Time: 4500 h	Grieb et al. (2016)
	Gadget-2	Memory allocation: 660 Gb	*https://wwwmpa.mpa-garching.mpg.de/gadget/*
	Haloes : SUBFIND	–	
ICE–COLA	Predictive	CPU Time: 33 h	Izard et al. (2016)
	2LPT + PM solver	Memory allocation: 340 Gb	Modified version of
	Haloes : FOF(0.2)	–	*ttps://github.com/junkoda/cola_halo*
PINOCCHIO	Predictive	CPU Time: 6.4 h	Monaco et al. (2013); Munari et al. (2017b)
	3LPT + ellipsoidal collapse	Memory allocation: 265 Gb	*https://github.com/pigimonaco/Pinocchio*
	Haloes : ellipsoidal collapse	–	–
PEAKPATCH	Predictive	CPU Time: 1.72 h∗	Bond & Myers (1996a,b,c)
	2LPT + ellipsoidal collapse	Memory allocation: 75 Gb∗	Not public
	Haloes : Spherical patches	–	–
	over initial overdensities	–	–
HALOGEN	Calibrated	CPU Time: 0.6 h	Avila et al. (2015).
	2LPT + biasing scheme	Memory allocation: 44 Gb	*https://github.com/savila/halogen*
	Haloes : exponential bias	Input: $\bar{n}$, 2-pt correlation function,	–
	–	halo masses, and velocity field	–
PATCHY	Calibrated	CPU Time: 0.2 h	Kitaura et al. (2014)
	ALPT + biasing scheme	Memory allocation: 15 Gb	Not public
	Haloes : non-linear, stochastic,	Input: $\bar{n}$, halo masses, and	–
	and scale-dependent bias	environment	–
Lognormal	Calibrated	CPU Time: 0.1 h	Agrawal et al. (2017)
	lognormal density field	Memory allocation: 5.6 Gb	*https://bitbucket.org/komatsu5147/lognormal_galaxies*
	Haloes : Poisson sampled points	Input: $\bar{n}$, 2-pt correlation function	
Gaussian	Theoretical	CPU Time: n/a	Grieb et al. (2016)
	Gaussian density field	Memory allocation: n/a	–
	Haloes : n/a	Input: $P(k)$ and $\bar{n}$	–

code,[2] and the cosmological parameters are fixed to their best-fitting value from the WMAP + BOSS DR9 analysis (Sánchez et al. 2013). The halo catalogues are obtained with the SUBFIND algorithm (Springel et al. 2001).

2.2 ICE–COLA

COLA is a fast N-body method, which solves for particle trajectories using a combination of 2LPT and a PM algorithm, where 2LPT is used for integrating the large-scale dynamics, while the PM part solves for the small-scale dynamics (Tassev, Zaldarriaga & Eisenstein 2013). This allows to drastically reduce the number of time-steps needed by the solver to recover the large-scale clustering of a full N-body method within a few per cent. In this article, we use the ICE–COLA version of the code, as optimized in Izard et al. (2016). Thus, the mocks used in this work use 30 time-steps linearly spaced in the expansion factor a from the starting redshift at $z_{\rm ini} = 19$ to $z = 0$[3] and a PM grid-size of $3 \times N_{\rm p}{}^{(1/3)}$ in each spatial dimension. They share the same particle load and box size as Minerva. Note that this configuration is optimized for reproducing dark matter clustering and weak lensing observables (Izard et al. 2018), and it is somewhat overdemanding in terms of computational cost for halo clustering alone (Izard et al. 2016), as seen in Table 1. Haloes are found on-the-fly using a Friends-of-Friends (FOF) algorithm with a linking length of $b = 0.2$ times the mean interparticle distance. This method does not require calibration, so we consider it a predictive method in the following.

2.3 PINOCCHIO

In this work, we use the latest version of PINOCCHIO, presented in Munari et al. (2017a). The algorithm can be summarized as follows: first, a linear density field is generated, using 1000^3 particles, and it is smoothed on different scales. For each particle, the collapse time at all smoothing scales is computed using the ellipsoidal collapse in the 3LPT approximation and the earliest is assigned as collapse time of the particle. The process of halo formation is implemented through an algorithm that mimics the hierarchical assembly of dark matter haloes through accretion of matter and merging with other haloes. Finally, haloes are displaced to their final position using 3LPT. This method only needs to be calibrated once and no re-calibration was required for this work, so we consider it predictive in what follows.

2.4 PEAKPATCH

In this work, we use a new massively parallel version of the PEAKPATCH algorithm (Stein et al. 2018; Alvarez et al., in preparation), originally developed by Bond & Myers (1996a,b,c). This approach

[2] http://cosmo.nyu.edu/roman/2LPT/

[3] Only half of these steps are completed by $z = 1$.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

is a Lagrangian space halo finder that associates haloes with regions that have just collapsed by a given time. The algorithm used in this work features four main steps: First, a density field is generated with 1000^3 particles using the 2LPTIC code. Collapsed regions are then identified using the homogeneous ellipsoidal collapse approximation and once exclusion and merging are imposed in Lagrangian space the regions are displaced to their final position using 2LPT. This method does not require calibration, so we will consider it a predictive method in the following.

2.5 HALOGEN

The HALOGEN method relies on the generation of a 2LPT matter density field at the redshift of interest and a subsequent assignment of haloes according to an exponential bias prescription. The algorithm used in this work can be summarized as follows: First, a 2LPT matter density field is generated at $z = 1$ using 768^3 particles, downsampling the initial conditions of the Minerva simulations. The particles are then assigned to a 300^3 grid. Haloes are assigned to grid cells with a probability proportional to $\rho_{\rm cell}^{\alpha(M_{\rm h})}$, where $M_{\rm h}$ is a halo mass sampled from the average mass function of the Minerva simulations. The position of the halo is drawn randomly from particle positions within the grid cell, while the velocity is given by the particle velocity multiplied by a velocity bias factor $f_{\rm vel}(M_{\rm h})$. The function $\alpha(M_{\rm h})$ is calibrated using the average Minerva two-point correlation function, while the velocity bias one, $f_{\rm vel}(M_{\rm h})$, is calibrated using the variance of halo velocities in Minerva, for several bins in halo mass. Therefore, we consider HALOGEN as a calibrated method in the following.

2.6 PATCHY

In this work, we employ the PATCHY version described in Kitaura et al. (2015), which uses augmented LPT (ALPT) to evolve the initial density field to the redshift of interest and then assigns haloes using a non-linear, scale-dependent, and stochastic biasing prescription. The algorithm used in this work can be summarized as follows: First, a linear matter density field is generated using 500^3 particles in a 500^3 grid, downsampling the initial conditions of the Minerva simulations, and then evolved to $z = 1$ using ALPT. The number density of haloes is determined using a deterministic bias prescription and haloes are assigned to grid cells by sampling a negative binomial distribution, with a stochastic bias parametrization to model deviations from Poissonianity. Halo velocities are given by modelling the coherent flow with ALPT and adding a Gaussian dispersion, while masses are drawn from the Minerva simulations and assigned according to halo environment properties (local density and cosmic web type) as determined by the HADRON code (Zhao et al. 2015). Two separate sets of mocks were produced for this work in order to reproduce the two samples considered (see Section 3). PATCHY will be considered a calibrated method in the following.

2.7 Lognormal

The lognormal mocks used in this work are produced with the code described in Agrawal et al. (2017). The algorithm used in this work can be summarized as follows: a Gaussian field $G(\vec{x})$ is generated on a 256^3 grid for matter and haloes. For haloes, we first measure the average real-space two-point correlation function ξ in Minerva, which is 'gaussianized' through the relation $\xi_{\rm G} = \ln(1 + \xi(r))$. We then get the power spectrum $P_{\rm G}$ such that its Fourier transform matches $\xi_{\rm G}$. This $P_{\rm G}$ is used to generate the Gaussian density field in each point of the Fourier Space grid (where modes are independent). This field is then transformed back to configuration space as $G(\vec{x})$. In turn, the matter density field is generated directly in Fourier space using the average power spectrum of Minerva and is only used to estimate the matter velocity field (needed to account for redshift space distortions) by solving the linear continuity equation. Both fields are then transformed into lognormal fields using the measured variance of the Gaussian fields, as $\delta_{\rm lg}(\vec{x}) = \exp[-(1/2)\sigma_{\rm G}^2 + G(\vec{x})] - 1$, and added. The number of haloes in a given cell is sampled from a Poisson distribution with mean $\bar{n}_{\rm h}[1 + \delta_{\rm h}(\vec{x})]V_{\rm cell}$, where $\bar{n}_{\rm h}$ is the mean number density of haloes in the Minerva simulations, $\delta_{\rm h}(\vec{x})$ is the halo density field, and $V_{\rm cell}$ is the volume of the cell. Finally, haloes are randomly positioned within the cell and their velocity is given by the velocity of the cell as computed by solving the linear continuity equation for matter. We will consider the lognormal mocks as an analytical prediction and use 1000 mocks to reduce the noise. Note that none of these are matched to the Minerva initial conditions.

2.8 Gaussian covariance

Here, we use the Gaussian covariance model described in Grieb et al. (2016). In this model, the contribution of the trispectrum and of the supersample covariance is neglected, so that the covariance is diagonal. The binned power spectrum multipole covariance in the thin shell approximation is given by

$$\mathbf{C}_{\ell_1,\ell_2}(k_i, k_j) = \delta_{ij}\frac{(2\ell_1+1)(2\ell_2+1)}{2N_{k_i}} \times \int_{-1}^{1}\left[P(k_i,\mu)+\frac{1}{\bar{n}}\right]^2 \mathcal{L}_{\ell_1}(\mu)\mathcal{L}_{\ell_2}(\mu)\mathrm{d}\mu, \quad (1)$$

where δ_{ij} is the Kroneker delta function, N_{k_i} is the number of independent Fourier modes in the bin, $P(k, \mu)$ is the anisotropic halo power spectrum, $\mu = \cos\theta$ where θ is the angle between the line of sight (LOS) and the separation vector of a galaxy pair, and $\mathcal{L}_\ell$ denotes the Legendre polynomial of order ℓ. In this work, $P(k, \mu)$ is obtained by fitting the average halo power spectrum of the N-body simulations with the model given in Section 4.2. The evaluation of the covariance takes around 2 s on a regular computer.

3 HALO SAMPLES

We define two samples in the N-body simulations by cutting in mass at two thresholds: 1.1×10^{13} and $2.7 \times 10^{13}\,{\rm M}_\odot\,h^{-1}$ that correspond to 42 and 100 particles, respectively. This yields samples with number densities of 2.13×10^{-4} and $5.44 \times 10^{-5}\,h^3\,{\rm Mpc}^{-3}$, respectively. We only consider redshift $z = 1$.

Results for the higher mass limit are more robust because the haloes are sampled with a larger number of particles making the estimation of their mass and position more reliable. Nevertheless, future surveys will observe galaxies that reside in lower mass haloes and that have a higher number density, thus probing a different shot-noise regime. That is why we consider a sample at lower mass that is closer to the target population of such surveys (e.g. for *Euclid* we expect a number density of $9 \times 10^{-4}\,h^3\,{\rm Mpc}^{-3}$ at this redshift). In Fig. 1, we show the shot-noise subtracted average halo power spectrum in real space, with the corresponding level of shot noise indicated with a dashed line. While Sample 1 is always signal dominated in the range of scales that we consider, Sample 2 is shot-noise dominated for $k > 0.14\,h\,{\rm Mpc}^{-1}$.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

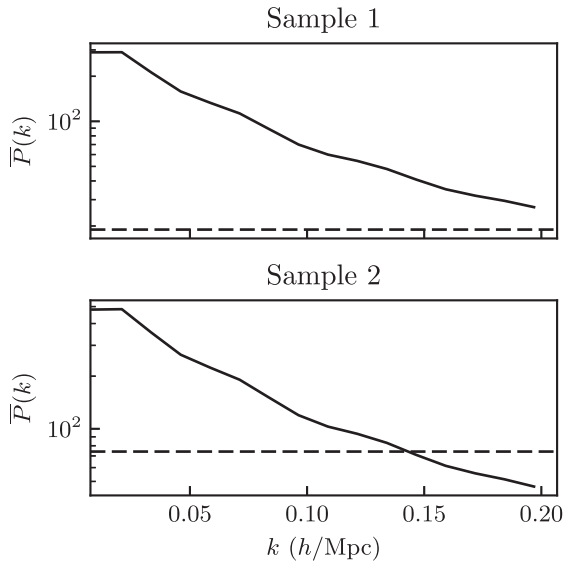

Figure 1. Average real space halo power spectrum from the 300 N-body simulations (continuous line) and the corresponding Poisson shot-noise level (dashed line), for sample 1 (top) and sample 2 (bottom) as defined in the top entries of Table 2, at redshift 1.

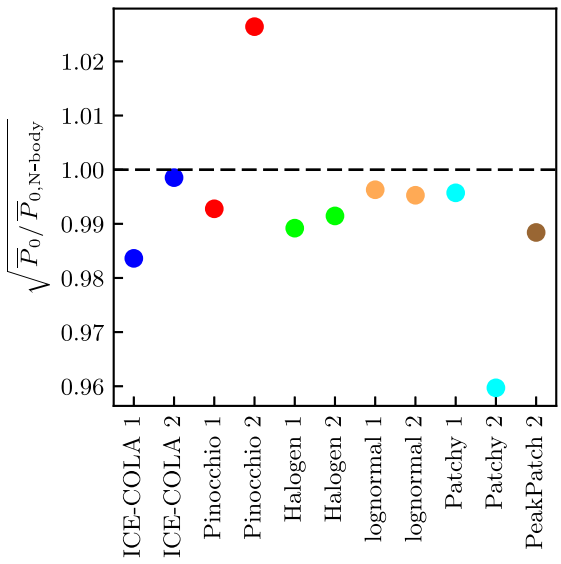

Figure 2. Square root of the average clustering amplitude of the approximate mocks divided by the average of the N-body simulations in the range $0.008 < k(h\,\mathrm{Mpc}^{-1}) < 0.096$ for the two samples, as indicated in the labels.

Table 2. Characteristics of the halo samples considered. For Sample 1 (top), we define in each approximate method an equivalent sample matched by abundance to the density of haloes more massive than $1.12 \times 10^{13}\,h^{-1}\,\mathrm{M}_\odot$ in the N-body simulation. For Sample 2 we do equivalently, but matching abundance for haloes with $M_h \geq 2.67 \times 10^{13}\,h^{-1}\,\mathrm{M}_\odot$ in the N-body. The second column displays the resulting abundance of those samples, which are not exactly equal to the N-body due to mass function discretization in steps equivalent to the particle mass. Samples are defined at $z = 1$. The significant difference in the mass cut of PEAKPATCH is due to the different mass definition criterium, as explained in Appendix A.

Method	$\bar{n}_{\mathrm{haloes}}$ ($h^3\,\mathrm{Mpc}^{-3}$)	M_{min} ($h^{-1}\,\mathrm{M}_\odot$)
	Sample 1	
N-body	2.130×10^{-4}	1.121×10^{13}
ICE–COLA	2.123×10^{-4}	1.086×10^{13}
PINOCCHIO	2.148×10^{-4}	1.044×10^{13}
HALOGEN	2.138×10^{-4}	1.121×10^{13}
Lognormal	2.131×10^{-4}	1.121×10^{13}
PATCHY	2.129×10^{-4}	1.121×10^{13}
	Sample 2	
N-body	5.441×10^{-5}	2.670×10^{13}
ICE–COLA	5.455×10^{-5}	2.767×10^{13}
PINOCCHIO	5.478×10^{-5}	2.631×10^{13}
HALOGEN	5.393×10^{-5}	2.670×10^{13}
Lognormal	5.441×10^{-5}	2.670×10^{13}
PATCHY	5.440×10^{-5}	2.670×10^{13}
PEAKPATCH	5.439×10^{-5}	2.355×10^{13}

We define equivalent samples in the approximate mocks by matching the abundance of haloes to the average abundance in the N-body simulations. The calibrated methods have the same abundance of haloes at these thresholds by construction, while in the case of the predictive methods we match the abundance by changing the mass threshold. The resulting samples are listed in Table 2. The abundance-matching procedure does not yield exactly the same number of haloes because the mass function is discretized in steps corresponding to the particle mass, nevertheless the recovered abundances agree within 1 per cent. In the case of PINOCCHIO, since the abundance could not be matched to high precision, halo masses were made continuous by randomly distributing the masses of N-particle haloes between $N \times M_\mathrm{p}$ and $(N + 1) \times M_\mathrm{p}$, where M_p is the particle mass of the Minerva simulations. The PEAKPATCH mocks, as implemented here, do not resolve the lower mass haloes, hence results are presented only for the second sample. In Fig. 2, we show the average clustering amplitude in the range $0.008 < k\,(h\,\mathrm{Mpc}^{-1}) < 0.096$ for the approximate mocks divided by the average of the N-body simulations in the same range, from which we can infer that the linear bias is recovered within a few per cent in all the samples.

This procedure to define samples resembles what is done in the analysis of real data, where mocks are produced by matching the number density and bias of the observed sample. Moreover, since we expect shot noise to have an important role in the covariance of these samples, matching the abundance will yield similar shot-noise levels in the compared covariances and will enhance discrepancies arising from other factors, notably from the approximations introduced by the different methods. Finally, since every approximate method defines haloes differently it is not expected that cutting at the same mass threshold as the N-body simulations would yield the same halo samples, we therefore believe that samples matched by abundance will result in a fairer comparison. In Appendix A, we compare results for the predictive methods for sample defined by cutting at the same mass threshold and samples matched by abundance. The abundance-matching procedure is especially helpful for the PEAKPATCH mocks in which haloes are identified as Lagrangian spherical overdensities, contrary to the other methods that use FOF or are calibrated to reproduce FOF mass functions.

4 METHODOLOGY

We compare 300 realizations for each approximate method with initial conditions matching the Minerva N-body simulations, so that the comparison is not affected by sample variance.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

4.1 Clustering measurements

We compute the redshift space power spectrum multipoles in the distant-observer approximation. We choose the direction of the LOS parallel to the three axes of the simulation box and average the results for the three directions to further reduce the noise in the measurements. We compare results at redshift $z = 1$ and in the range of scales $0.008 < k\,(h\,\mathrm{Mpc}^{-1}) < 0.197$, unless otherwise stated. We measure the power spectrum multipoles P_ℓ ($\ell = 0, 2, 4$) using the public code POWERI4,[4] which employs the interlacing technique to reduce aliasing contribution (Sefusatti et al. 2016). We use the fourth-order mass interpolation scheme (piecewise cubic spline) on a 256^3 grid. We bin the power spectrum multipoles in intervals of $\Delta k = 3k_\mathrm{f}$, where $k_\mathrm{f} = 2\pi/L_\mathrm{box}$ is the fundamental frequency of the box, for a total of 16 bins per multipole.

4.2 Theory modelling and parameter space

We compare the performance of different covariance matrices in terms of the errors and contours in a series of model parameters that are generally standard in galaxy clustering analysis. In particular, we adopt a model very close to the one in the anisotropic clustering study of BOSS DR12 data presented in Grieb et al. (2017) and Sánchez et al. (2017). Such a model includes (1) non-linear matter clustering through the matter and velocity divergence correlations (gRPT, Crocce et al. in preparation), (2) a biasing scheme to one-loop in the initial power spectra, and (3) the non-linear equivalent of the Kaiser effect for redshift space distortions through the correlations between densities and large-scale velocities. The main difference with respect to a galaxy analysis is that we do not consider parameters related to small-scale pair-wise velocity dispersion (the so-called Fingers-of-God effect) because we are considering haloes and these are not affected by this effect. Broadly speaking, the theory model is given by

$$P^s_\mathrm{hh}(k,\mu) = P_\mathrm{hh}(k) + 2f\mu^2 P_\mathrm{h\theta}(k) + f^2\mu^4 P_{\theta\theta}(k) + P^{(2)}_\mathrm{novir}(k,\mu) + P^{(3)}_\mathrm{novir}(k,\mu), \quad (2)$$

where P_hh, $P_\mathrm{h\theta}$, and $P_{\theta\theta}$ stand for halo density h and velocity divergence θ auto and cross-power spectra, and f is the growth rate of structure (where $\theta \equiv \nabla\boldsymbol{u} \equiv -\nabla\boldsymbol{v}/aHf$). The first three terms are the non-linear 'Kaiser' effect, with P_hh and $P_\mathrm{h\theta}$ written to one loop in the biasing scheme. Explicit expressions can be found in Sánchez et al. (2017) and they depend on b_1 (linear bias), b_2 (non-linear, second order in density fluctuations), γ_2 (non-local, second order), and γ_3^- (non-local, third order in perturbations), following the notation and bias basis of Chan, Scoccimarro & Sheth (2012). Higher order terms, such as b_3, are 'renormalized' into the above parameters (at one-loop), see McDonald (2006). The last two terms in equation (2) also arise from the transformation of real to redshift space coordinates. Defining $D_\mathrm{s} \equiv \delta_h + f\nabla_z u_z$, we have

$$P^{(2)}_\mathrm{novir}(k,\mu) = \int \frac{q_z}{q^2}[B_{\theta D_\mathrm{s} D_\mathrm{s}}(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}, -\boldsymbol{k}) + B_{\theta D_\mathrm{s} D_\mathrm{s}}(\boldsymbol{q}, -\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{q})]\mathrm{d}^3q, \quad (3)$$

which depends on the cross-bispectrum B between the velocity divergence θ and D_s, which can be evaluated at tree-order (thus, depends only on b_1, b_2, and γ_2), while

$$P^{(3)}_\mathrm{novir}(k,\mu) = \int \frac{q_z}{q^2}\frac{(k_z - q_z)}{(\boldsymbol{k}-\boldsymbol{q})^2}\left(b_1 + f\mu_q^2\right)\left(b_1 + f\mu^2_{k-q}\right) \times P_{\delta\theta}(k-q)P_{\delta\theta}(q)\mathrm{d}^3q, \quad (4)$$

which is already second order in $P(k)$ and hence only depends on linear bias b_1. Equations (3) and (4) are basically the equivalent of the $A(k,\mu)$ and $B(k,\mu)$ terms, respectively, in equation (18) of Taruya, Nishimichi & Saito (2010) extended to biased tracers (see also Nishimichi & Taruya 2011). We have checked that the sensitivity of our halo samples clustering to γ_2 is minor, hence we keep terms proportional to γ_2, but we fix γ_2 to its local Lagrangian relation to the linear bias, i.e. $\gamma_2 = -(2/7)(b_1 - 1)$.

The final component of the model is the Alcock–Paczynski (AP) effect. Anisotropic clustering analysis needs to assume a 'fiducial' cosmological model in order to transform observables (redshifts and angular positions) into co-moving coordinates. A mismatch between this 'fiducial' model and the 'true' cosmology (i.e. the one being assumed in the likelihood evaluation) leads to distortions, known as AP effect (Alcock & Paczynski 1979), which can be used to place cosmological constraints (e.g. Samushia, Percival & Raccanelli 2012). Such distortions are characterized by two parameters that transform comoving coordinates in the 'true' cosmology, denoted $(k_\perp, k_{\shortparallel})$,[5] to coordinates in the 'fiducial' cosmology, denoted $(k'_\perp, k'_\parallel)$, as

$$k'_\perp = \alpha_\perp k_\perp \quad k'_\parallel = \alpha_\parallel k_\parallel, \quad (5)$$

with

$$\alpha_\perp = \frac{D_\mathrm{A}(z)r'_\mathrm{d}}{D'_\mathrm{A}(z)r_\mathrm{d}} \quad \alpha_\parallel = \frac{H'(z)r'_\mathrm{d}}{H(z)r_\mathrm{d}}, \quad (6)$$

where D_A is the angular diameter distance to the sample redshift, $H(z)$ the Hubble parameter, and r_d is the sound horizon at the drag redshift. In terms of wave-vector amplitude and angle to the LOS, this transformation is (Ballinger, Peacock & Heavens 1996)

$$k(\mu', k') = k'\left[\alpha_\parallel^{-2}\mu'^2 + \alpha_\perp^{-2}(1-\mu'^2)\right]^{1/2} \quad (7)$$

$$\mu(\mu') = \mu'\alpha_\parallel^{-1}\left[\alpha_\parallel^{-2}\mu'^2 + \alpha_\perp^{-2}(1-\mu'^2)\right]^{-1/2}. \quad (8)$$

Therefore, after evaluating the power spectrum model at the cosmology being assumed by the likelihood step, i.e. $P(k,\mu)$, we need to transform back to the basis of the 'fiducial' cosmology, together with the multipole decomposition, as

$$P_\ell(k') = \frac{2\ell+1}{2\alpha_\perp^2\alpha_\parallel}\int \mathcal{L}_\ell(\mu')P(k(k',\mu'),\mu(\mu'))\mathrm{d}\mu'. \quad (9)$$

In summary, our parameter space has six free parameters: b_1, b_2, γ_3^- for galaxy physics, the growth rate of structure $f\sigma_8$, and the two AP parameters $\alpha_\perp$, $\alpha_{\shortparallel}$.

4.3 Covariance matrix estimation

The covariances are computed using the sample covariance estimator:

$$\mathbf{C}_\ell(k_i, k_j) = \frac{1}{N_\mathrm{s}-1}\sum_{n=1}^{N_s}(P^n_\ell(k_i) - \bar{P}_\ell(k_i))(P^n_\ell(k_j) - \bar{P}_\ell(k_j)), \quad (10)$$

[4] https://github.com/sefusatti/PowerI4

[5] Here, $\perp$ and $\shortparallel$ refer to wave-vectors parallel and perpendicular to the LOS, respectively.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

Table 3. Best-fitting parameter values to the N-body power spectrum multipoles for the two samples.

Sample	b_1	b_2	γ_3^-
1	2.637	−1.660	0.693
2	3.434	−1.067	1.601

where $N_s = 300$ is the number of simulations, $\ell = 0$, 2, 4 is the multipole considered, and $\bar{P}_\ell = 1/N_s \sum_{n=1}^{N_s} P_\ell^n$ is the average power spectrum. Analogously, the cross-covariances between different multipoles are computed as

$$\mathbf{C}_{\ell_1,\ell_2}(k_i, k_j) = \frac{1}{N_s - 1} \sum_{n=1}^{N_s} \left(P^n_{\ell_1}(k_i) - \bar{P}_{\ell_1}(k_i)\right) \times \left(P^n_{\ell_2}(k_j) - \bar{P}_{\ell_2}(k_j)\right). \quad (11)$$

In what follows, we compute covariance matrices in each of the three LOS directions independently and then average them.

4.4 Fitting procedure

The main use for covariances in the context of galaxy survey data analysis is to perform a fit to the measured observable given a model in the likelihood analysis framework. In order to understand if the approximations in the methods presented here systematically increase/decrease the errors on the recovered parameters, we run MCMC chains based on the assumption of a Gaussian likelihood:

$$-2\log\mathcal{L} = \sum_{\ell_1\ell_2}\sum_{ij} \left(\hat{P}_{\ell_1}(k_i) - P_{\mathrm{m},\ell_1}(k_i)\right)^T \times \mathbf{\Psi}_{\ell_1\ell_2}(k_i, k_j)\left(\hat{P}_{\ell_2}(k_j) - P_{\mathrm{m},\ell_2}(k_j)\right), \quad (12)$$

where $\mathbf{\Psi} = \mathbf{C}^{-1}$ is the precision matrix and P_m is a model of the power spectrum multipoles that depends on the cosmological and nuisance parameters.

Since we are only interested in the performance of the covariance, we use as $\hat{P}$ a fit to the N-body average power spectrum multipoles with the model described in Section 4.2. To do this fit, we fix the cosmological parameters to their true values and only vary the bias parameters b_1, b_2, and γ_3^-. The best-fitting parameter values that we obtain for the two samples are given in Table 3.

We then run MCMC chains by changing the covariance with the ones estimated with the approximate mocks. We vary the AP parameters ($\alpha_\parallel$, $\alpha_\perp$) and $f\sigma_8$ together with the nuisance bias parameters b_1, b_2, and γ_3^-. This matches what is done in article I for the configuration space analysis. For each method, we first run a short chain to have an estimate of the parameter covariance and then run a full chain that uses that parameter covariance as input. All the full chains have 2×10^5 points.

5 RESULTS

5.1 Mean of the power spectrum multipoles

Fig. 3 shows the relative difference of the average power spectrum multipoles as measured from the approximate mocks with respect to the corresponding N-body measurements, except for the case of the hexadecapole where we use the theoretical estimate as reference because the large noise makes the N-body measurements cross 0 and can lead to diverging ratios. The shaded regions show the standard deviation of the 300 N-body simulations. In general, the monopole is better reproduced than the higher order multipoles by all methods and the agreement with the N-body results is better in the first sample than in the second.

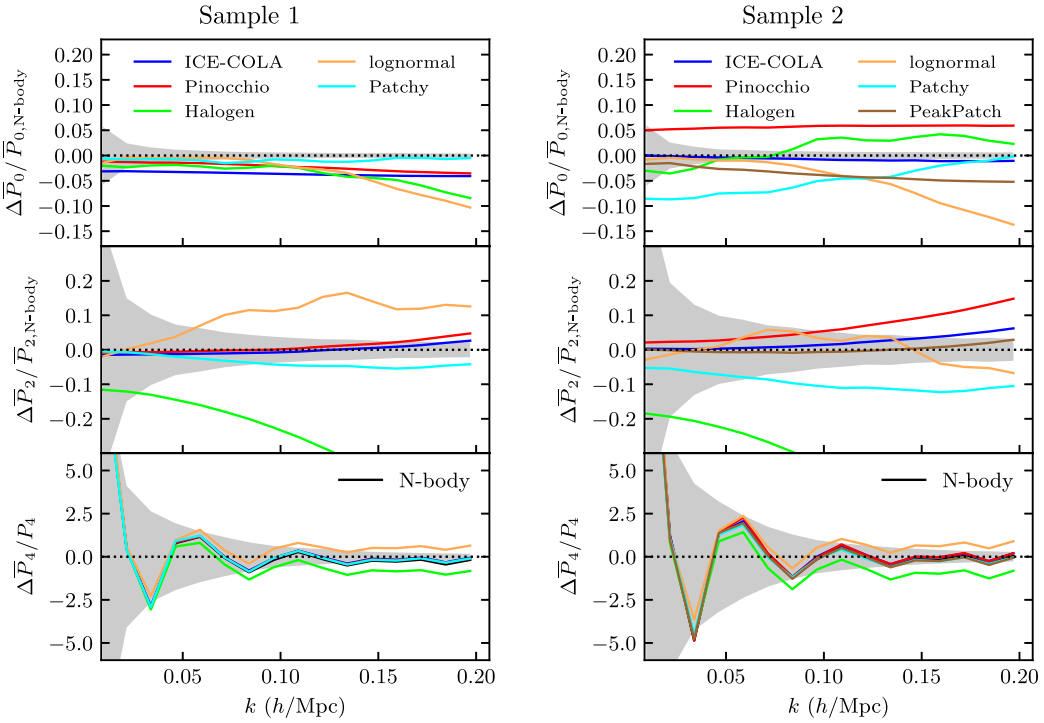

Figure 3. Relative difference of the power spectrum monopole and quadrupole (hexadecapole) with respect to the N-body (model) ones for sample 1 (left-hand plot) and sample 2 (right-hand plot). In each plot, we show the monopole (top panel), quadrupole (middle panel), and hexadecapole (bottom panel). The grey shaded region shows the standard deviation of the N-body measurements.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

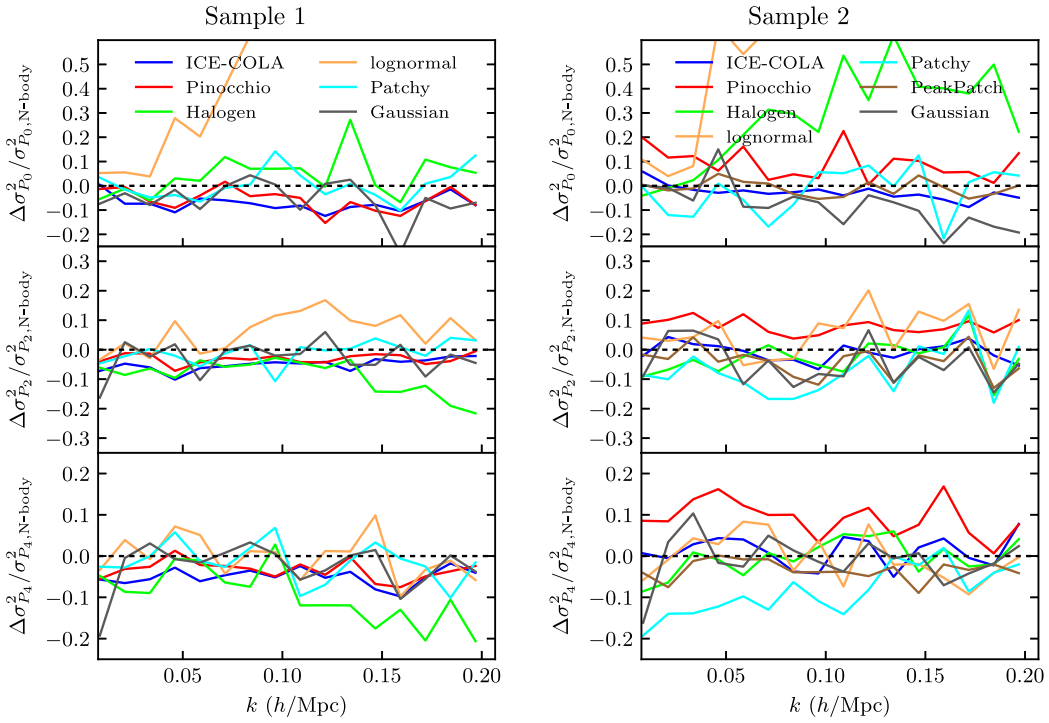

Figure 4. Relative difference of the variance of the power spectrum multipoles with respect to the N-body ones for sample 1 (top plot) and sample 2 (bottom plot). In each plot, we show the monopole (top panel), quadrupole (middle panel), and hexadecapole (bottom panel). Notice that the Gaussian is a noiseless estimate so that differences are dominated by the noise in the N-body variance estimate.

The monopole is reproduced to within ∼5 per cent by all methods at large scales in both samples, with the exception of PATCHY in the second sample. ICE–COLA and PINOCCHIO perform the best in terms of scale dependence showing a very flat difference, as well as PATCHY in the first sample. In the second sample, the calibrated methods show a scale dependence and oscillations in the ratio, indicating that the BAO features are not completely reproduced by these methods in this high-bias sample. PATCHY in the first sample and ICE–COLA in the second have the best performances; they reproduce the monopole to within 1 per cent on the entire range of scales considered here. Notice that flat residuals in the monopole indicate that there is a mismatch in the linear bias, which is considered a nuisance parameter in cosmological analysis.

All methods show some degree of scale-dependent difference in the quadrupole, with HALOGEN performing notably bad even at large scales. The other methods show a good agreement at large scales that deteriorates at small scales. ICE–COLA and PINOCCHIO have very good performances in the first sample, reproducing the quadrupole to within 1 per cent in the whole range of scales considered. The same is true for PEAKPATCH in the second sample, where ICE–COLA has a residual deviation of ∼5 per cent by $k = 0.2\,h^{-1}$ Mpc and PINOCCHIO ∼14 per cent. PATCHY shows 5–10 per cent deviations in the quadrupole. This performance can be potentially improved with an explicit fit to the velocity dispersion, modelling virialized motions, in the N-body. Lastly, we note that lognormal shows residual oscillations at BAO scales also in the quadrupole for both samples.

In the case of the hexadecapole, it is more difficult to evaluate per cent differences because of noise amplification due to the very small values of the denominator when taking the ratio, but these are on average of the order of 10–15 per cent for ICE–COLA, PINOCCHIO, PEAKPATCH, and PATCHY, while they are > 40 per cent for HALOGEN and lognormal.

5.2 Variance of the power spectrum multipoles

In Fig. 4, we show the relative difference of the variance of the power spectrum multipoles with respect to the N-body variance. We also include the Gaussian prediction described in Section 2.8. Note that the variance of the power spectrum in the Gaussian model is by construction noiseless hence the scatter in the plotted points is due to noise in the reference N-body variance (the denominator of the ratio). We notice that the variance of Sample 2 has a larger scatter, pointing to a larger contribution of shot noise in the covariance error for this sample.

In Sample 1, all the methods agree with the N-body variance within 10 per cent on large scales, with the exception of lognormal for reasons that we address next. HALOGEN reaches a 20 per cent difference beyond $k \sim 0.1\,h^{-1}$ Mpc. In Sample 2, ICE–COLA and PEAKPATCH do particularly well with an average of ∼3 per cent deviations from the N-body results in all multipoles, while the other methods show a more diverse behaviour. In the monopole variance, PINOCCHIO has an average deviation of ∼9 per cent and HALOGEN of 30 per cent. The quadrupole variance is reproduced at the level of 5–10 per cent, while the hexadecapole variance is within 5 per cent for all methods, with the exception of PINOCCHIO and PATCHY (9 per cent). Overall, the Gaussian model performs well at the level of the mock-based variances.

One noteworthy result is that lognormal mocks show a significantly larger variance than the N-body in the monopole and

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

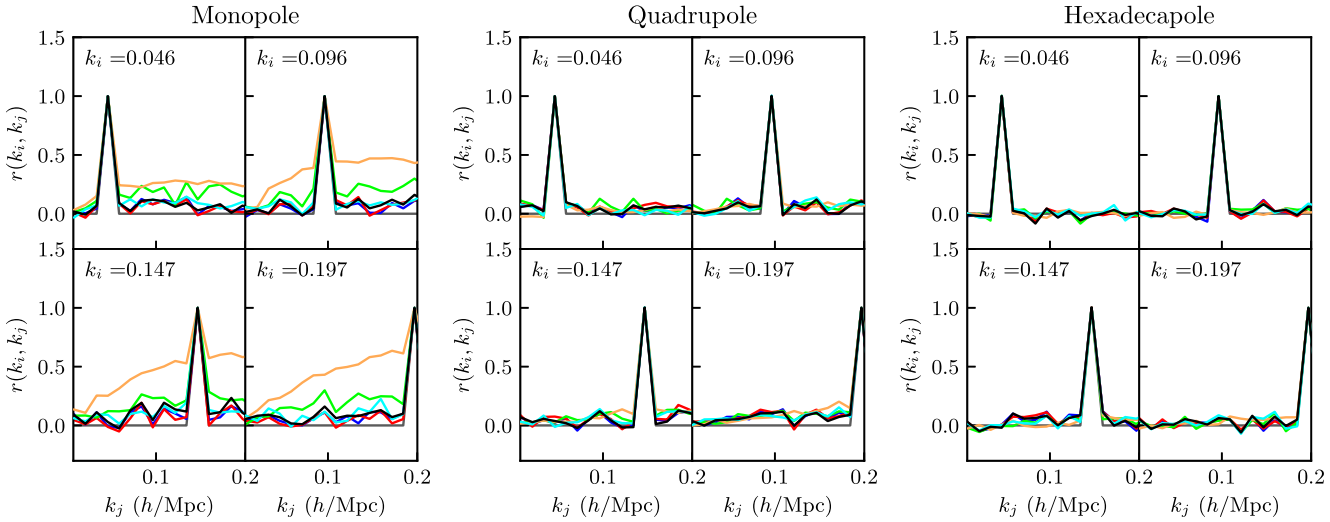

Figure 5. Cut through the correlation coefficient of the monopole (left-hand plot), quadrupole (middle plot), and hexadecapole (right-hand plot) for sample 1 at four different values of k as indicated in the panels. We show the results for N-body (black), ICE–COLA (blue), PINOCCHIO (red), HALOGEN (green), PATCHY (cyan), lognormal (orange), and Gaussian (grey).

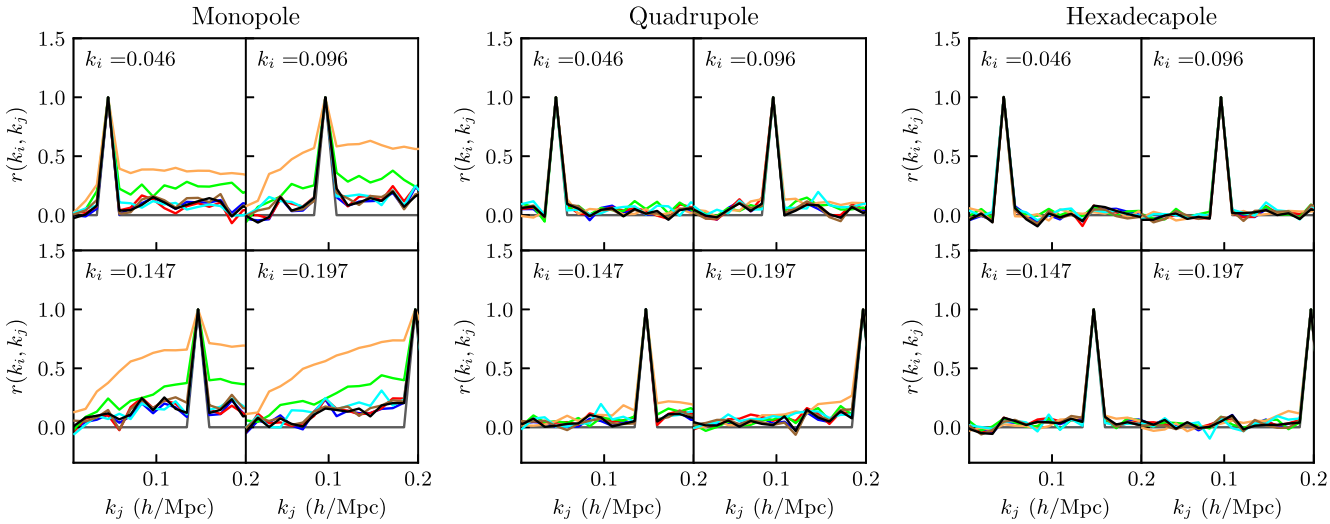

Figure 6. Cut through the correlation coefficient of the monopole (left-hand plot), quadrupole (middle plot), and hexadecapole (right-hand plot) for sample 2 at four different values of k as indicated in the panels. We show the results for N-body (black), ICE–COLA (blue), PINOCCHIO (red), PEAKPATCH (brown), HALOGEN (green), PATCHY (cyan), lognormal (orange), and Gaussian (grey) (color online).

that the discrepancy has a strong dependence on binning, showing increasing deviations with larger bins. We attribute this behaviour to the fact that these mocks show larger non-Gaussian contributions than the N-body simulations. In fact, the Gaussian part of the variance decreases with increasing bin widths, while the non-Gaussian part does not depend on the binning (see e.g. Scoccimarro, Zaldarriaga & Hui 1999). Since here we are showing the relative difference, the relative magnitude of the non-Gaussian part increases with larger binning, inducing larger deviations from the N-body result. In addition, we found a trend of increasing deviations from the N-body with increasing bias that will be investigated in future work. All the other methods do not show this pronounced behaviour with binning, indicating that the relative contributions of the Gaussian and non-Gaussian parts to the covariance are better reproduced.

5.3 Correlation coefficient of the power spectrum multipoles

To show the off-diagonal elements of the covariance matrix, we compute the correlation coefficient for each method as

$$\mathbf{r}_{ij} = \frac{\mathbf{C}_\ell(k_i, k_j)}{\sqrt{\mathbf{C}_\ell(k_i, k_i)\mathbf{C}_\ell(k_j, k_j)}}. \tag{13}$$

In Figs 5 and 6, we show a cut through the correlation coefficient $\mathbf{r}$ for the power spectrum multipoles for four different k_i values, as indicated in the plot panels. The colour coding is the same as the previous plots with the addition of the N-body in black. The first consideration to make is that the level of correlation for the power spectrum multipoles at these scales and with the chosen binning is very low. The monopole covariance shows the largest correlations, on average ∼7 per cent with a maximum of ∼20 per cent, while

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

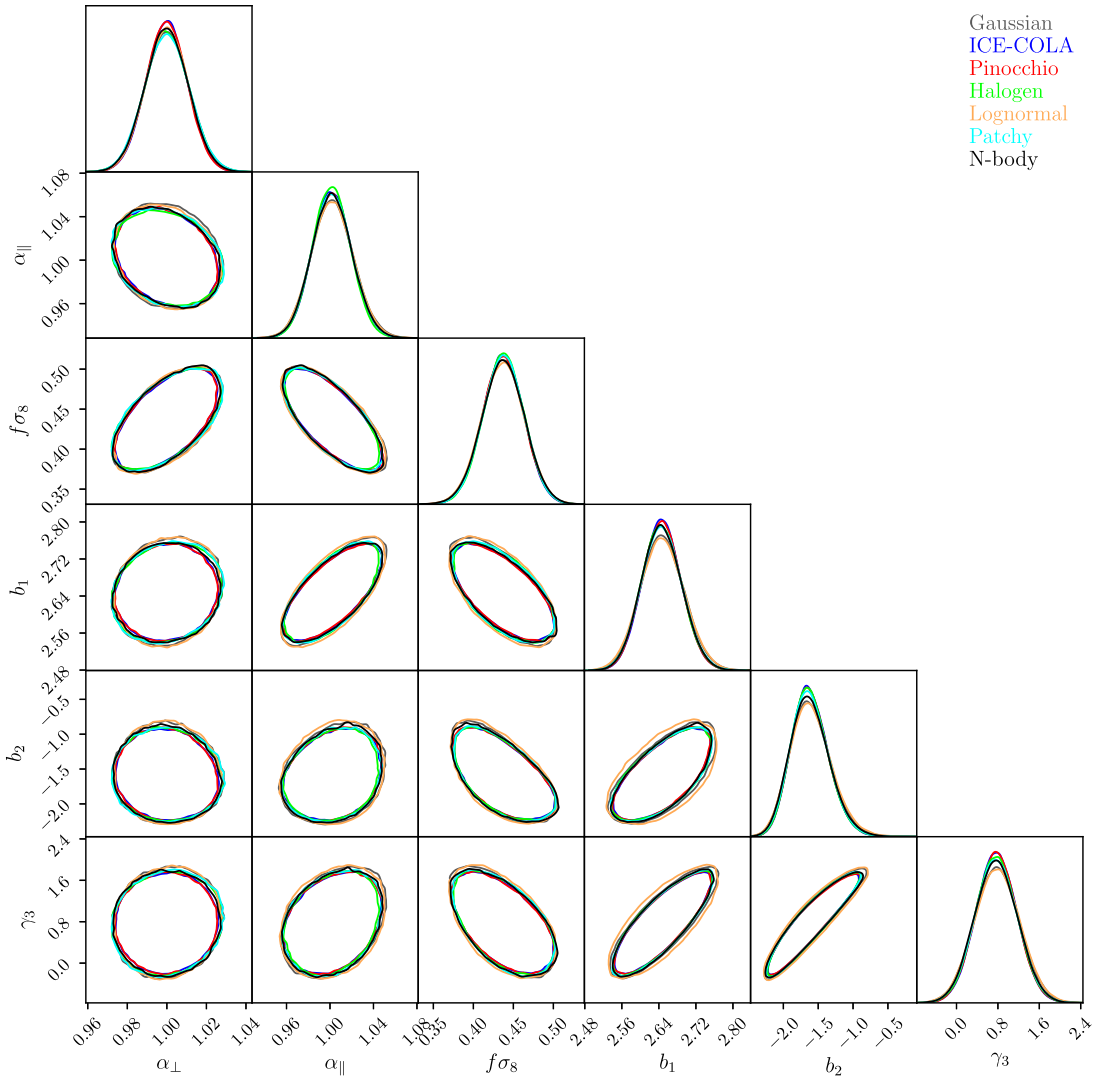

Figure 7. Marginalized 2σ contours for cosmological and nuisance parameters for sample 1.

for the quadrupole they are on average of the order of ∼4 per cent and for the hexadecapole ∼1 per cent. The second consideration is that because of the matching Initial conditions all the approximate methods show noise properties that are highly correlated to the noise in the N-body covariance. Finally, all methods seem to qualitatively reproduce the level of correlations of the N-body spectra with the exception of the monopole for HALOGEN, which predicts correlations two to three times higher than the N-body, and lognormal for the reason explained in the previous section.

5.4 Cosmological parameter constraints

To gauge the accuracy of the approximate methods in a realistic context, we propagate errors on the covariance all the way to cosmological parameter errors using a likelihood analysis, as described in Section 4.4. In Figs 7 and 8, we show the resulting 2σ parameter contours for the cosmological and nuisance parameters. Overall, the position and direction of degeneracies of the contours is well reproduced by the approximate mocks. Recovered best-fitting parameter values,[6] for both samples, agree with the input ones to $\lesssim$1 per cent for all cosmological parameters and b_1, but show larger deviations for the higher order halo bias ones even when using the N-body covariance. We attribute this to the fact that the latter parameters are poorly constrained and sometimes show asymmetric posterior distributions (see Figs 7 and 8). Similar results are obtained if the best-fitting values are defined as the mean of the marginalized posterior distribution. In any case, these best-fitting parameter biases are always marginally negligible if defined with respect to their corresponding 1σ error.

To quantify differences between the recovered parameter errors, we plot the ratio of the marginalized error on cosmological and nuisance parameters with respect to the N-body results in Fig. 9. Before further discussion, we point out that despite matching the Initial conditions there is a residual scatter on parameter error ratios

[6]Defined as the maximum of the posterior distribution for each parameter after all others have been marginalized over.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

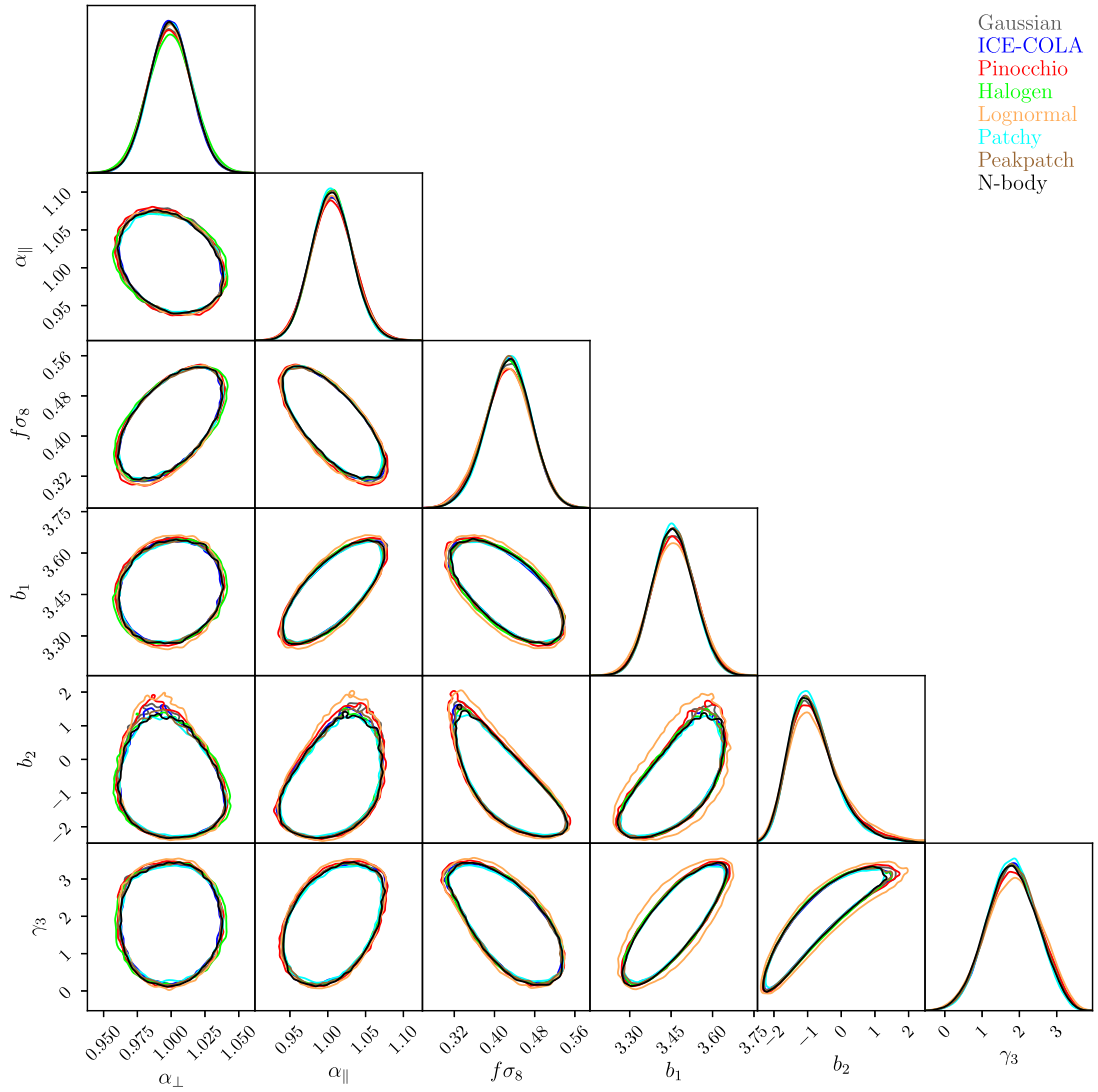

Figure 8. Marginalized 2σ contours for cosmological and nuisance parameters for sample 2.

that we quantify in Appendix B to be at the level of 4–5 per cent, (indicated in Fig. 9 as a grey shaded region). One should bear this in mind when evaluating differences found. Moreover, for lognormal and Gaussian the error exceeds this value because they do not match the initial conditions to the N-body simulations and thus have a different realization of noise in the covariance. This gets propagated to the cosmological parameter errors. This means that we expect results for these two methods to be more strongly affected by the noise in the N-body covariance. We estimate uncertainties for these two methods to be at the level of 10 per cent.

Overall, all the methods reproduce the N-body errors within the statistical uncertainty, with few exceptions. PINOCCHIO slightly exceeds the 5 per cent limit in the nuisance parameters b_2 and γ_3^- in Sample 1 and in the cosmological parameters and b_1 in Sample 2, while ICE–COLA has similar issues for the nuisance parameters b_2 and γ_3^- only in Sample 1. HALOGEN shows deviations larger than the uncertainty in one of the AP parameters in both samples, while PATCHY only shows a slight underestimation of the error on b_2 in Sample 2. In Sample 2, lognormal shows deviations for the nuisance parameters that are larger than the expected 10 per cent statistical uncertainty.

As a further test, we compute the volume of the 3D ellipsoid for the cosmological parameters as

$$V = \sqrt{\det \mathbf{cov}(\alpha_\parallel, \alpha_\perp, f\sigma_8)}, \tag{14}$$

where $\mathbf{cov}(\alpha_\parallel, \alpha_\perp, f\sigma_8)$ is the parameter covariance that we recover from the full MCMC chains. We show the ratio of V to the N-body result in Fig. 10. Deviations larger than 10 per cent are found for ICE–COLA and lognormal in Sample 1, and for PINOCCHIO in Sample 2. The single parameter errors shown in Fig. 9 for these cases have smaller deviations from the N-body result but the off-diagonal elements of the parameter covariance drive the deviations in terms of volume to higher values.

5.4.1 *Dimensionality of parameter space and* k_{max}

We finish by studying the stability of our results against changes of the dimensionality of our parameter space and of the non-linear scale cut-off. This is motivated in part because, despite the bad performance of lognormal on the monopole covariance, the cosmological parameter errors are reproduced quite well. At

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

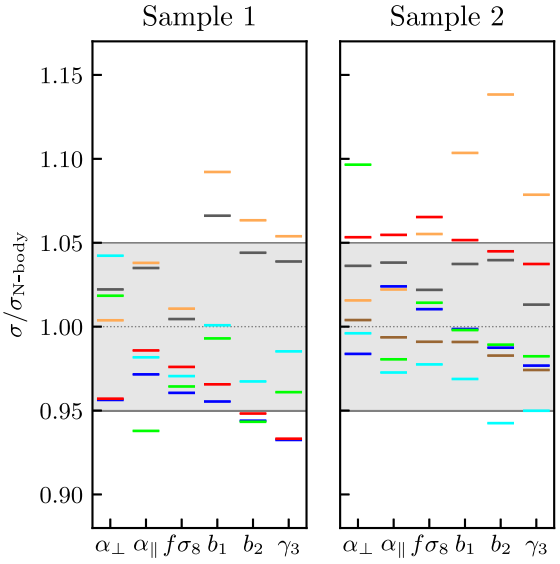

Figure 9. Ratio of the error on cosmological and nuisance parameters with respect to N-body for sample 1 (left) and sample 2 (right). The grey shaded area indicates 5 per cent differences. We show the results for ICE–COLA (blue), PINOCCHIO (red), PEAKPATCH (brown), HALOGEN (green), PATCHY (cyan), lognormal (orange), and Gaussian (grey) (color online).

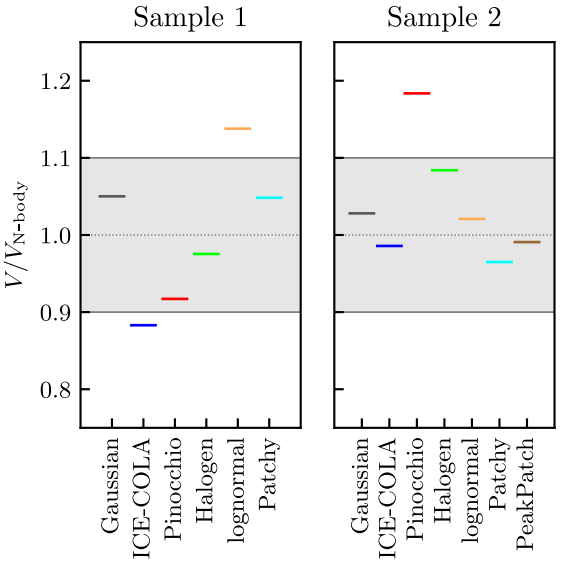

Figure 10. Ratio of the volume of the 3D ellipsoid of cosmological parameters with respect to N-body for sample 1 (left) and sample 2 (right). The grey shaded area indicates 10 per cent differences.

the same time, the nuisance parameters b_2 and γ_3 are poorly constrained, especially in Sample 2, so that some of the differences between methods can be concealed. In order to test this, we run chains by fixing γ_3^- to its true value reported in Table 3 for two different values of $k_{\rm max}$ (one of them our baseline). The results are shown in Fig. 11 where we can see that in fact the discrepancy of the error obtained with the lognormal method increases to 10 per cent for $f\sigma_8$ and can be as large as 50 per cent for the nuisance parameter b_1 and b_2, with a clear trend of increasing discrepancy with higher $k_{\rm max}$. We also notice that the performances of the other methods are not affected by fixing γ_3^- or increasing the $k_{\rm max}$ to $0.3\,h\,{\rm Mpc}^{-1}$,

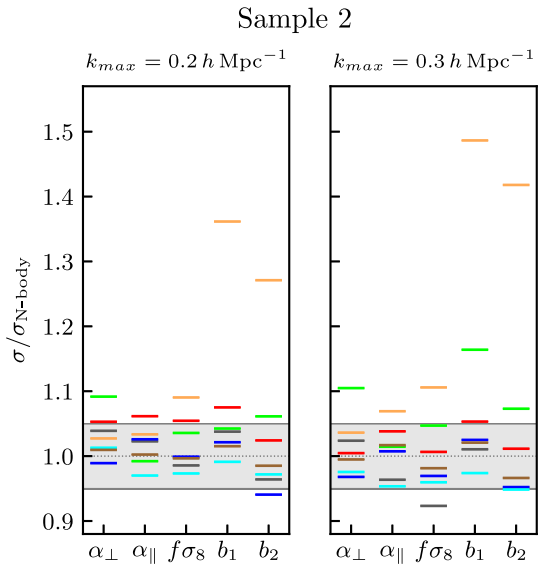

Figure 11. *Dependence with dimensionality of parameter space and* $k_{\rm max}$: Ratio of the error on cosmological for a reduced set of nuisance parameters (γ_3 fixed), with respect to the N-body ones, for sample 2. Left-hand panel corresponds to fits with $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ (our baseline value) and right-hand panel with $k_{\rm max} = 0.3\,h\,{\rm Mpc}^{-1}$. The grey shaded area indicates 5 per cent differences. We show the results for ICE–COLA (blue), PINOCCHIO (red), PEAKPATCH (brown), HALOGEN (green), PATCHY (cyan), lognormal (orange), and Gaussian (grey). Our results are stable against these changes, except for lognormal, whose performance decreases rapidly, and, to a less extent, Halogen (color online).

with the exception of Halogen that shows increased ratios on all the parameters and exceeds the 10 per cent level for $\alpha_\perp$ and b_1 when $k_{\rm max}$ is increased.

6 DISCUSSION AND CONCLUSIONS

The cosmological analysis of galaxy redshift surveys requires an estimation of the covariance of the observables to be able to put constraints on cosmological parameters in a Gaussian likelihood analysis framework. One way of doing this is to simulate the galaxy sample of interest a large number of times and computing the covariance using the sample covariance estimator. Since full N-body simulations are computationally very expensive, galaxy mock catalogues used for covariance estimation are produced using so-called 'approximate methods', which speed up the calculation by introducing approximations in the dynamics or the statistics of the galaxy density field.

In this work, we studied for the first time the accuracy of these methods in reproducing the N-body halo power spectrum multipoles, their covariances, and the recovered parameter errors on the set of cosmological and biasing parameters $\{\alpha_\parallel, \alpha_\perp, f\sigma_8, b_1, b_2, \gamma_3^-\}$. We did this at $z = 1$ by comparing ensembles of simulations from approximate methods with a reference set of 300 N-body simulations. To minimize the impact of sample variance errors and noise due to the limited number of simulations, we used the same initial conditions as the N-body runs for the approximate mocks.

For completeness, we considered codes that introduce different types of approximations, trying to span the broad range of algorithms proposed so far in the literature. These can be summarized as follows: fast PM methods, which introduce approximations in

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

the force computation in a PM code, represented by ICE–COLA; Lagrangian methods, which identify haloes in Lagrangian space and displace them to the final redshift using LPT, represented by PINOCCHIO and PEAKPATCH; bias-based methods, which use LPT to generate a matter field at the redshift of interest and apply a biasing scheme to define haloes, represented by PATCHY and HALOGEN; and methods that rely on assumptions on the shape of the PDF of the density field, represented by lognormal mocks and Gaussian covariances. In this order, the grouping transitions from higher resolution and complexity, and hence computational cost, to simpler and faster methods. The methods used in this article, spanning the algorithms above, are presented in Table 1.

We then compared performances for two halo samples: The first has a linear bias of $\sim$2.6 and a number density of $\sim 2 \times 10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$, while the second has a linear bias of $\sim$3.4 and a number density of $\sim 5.5 \times 10^{-5}\,h^3\,\mathrm{Mpc}^{-3}$. Provided with this, we defined halo samples in each approximate method by abundance matching to those two samples above. This was mainly motivated by the fact that different methods define haloes in different ways.

Our conclusions can be summarized as follows:

(i) *Power spectrum multipole covariance:* The variance of the power spectrum multipoles is in general reproduced within 10 per cent by all methods in both samples, with the exception of the monopole for HALOGEN and lognormal. In this regard, a special mention is due to lognormal, and to a less extent HALOGEN, which shows very large variance and covariance for the monopole power spectrum that only reaches the Gaussian expectation on very large scales. For lognormal, this might be due to the fact that our samples are highly biased, in a regime where the lognormal and Gaussian fields are not a good approximation to each other. Further work is needed to explain this behaviour, which translates into an overestimation of error bars in the bias parameters.

(ii) *Recovered model parameters:* We use the covariance obtained from each method to run a likelihood analysis using a known theory data vector. From the MCMC chains, we recover best-fitting model parameters related to biasing, AP, and growth rate of structure and their errors. Although not shown, we find that different covariances do not bias the recovery of cosmological parameters, which are always within 1 per cent of their input values, while we find larger (few per cent) biases for poorly constrained nuisance parameters.

(iii) *Errors on model parameters:* Reaching conclusions from the translation of differences in the covariances into differences at the level of errors in model parameter σ is not trivial because of the finite number of mocks available, which is none the less 300 mocks for each method in this work. To partially overcome this limitation, we match the initial conditions of the approximated mocks with those of the N-body runs. This reduces the expected error on the 'error ratio' ($\sigma/\sigma_{\text{N-body}}$) by a factor of $\lesssim 2$ to about 4–5 per cent. Although a rough estimate, this should be regarded as an indication of the statistical limit of our analysis. Bearing this in mind, we find that all the methods recover the error on cosmological parameters for both samples at the 5 per cent level or better, with the exception of HALOGEN, which has larger deviations for one of the AP parameters in both samples, and PINOCCHIO in the second sample. The errors on nuisance parameters are within the statistical uncertainty only for PATCHY in the first sample and PEAKPATCH, ICE–COLA, HALOGEN, and Gaussian in the second sample. In turn, the volume of the 3D ellipsoid defined by the 1σ contours in the parameter space $\{\alpha_\parallel, \alpha_\perp, f\,\sigma_8\}$ is reproduced within 10 per cent by all methods with the exception of ICE–COLA and lognormal in Sample 1 and PINOCCHIO in Sample 2. Remarkably, the Gaussian prediction (albeit using a non-linear power spectrum) performs well presumably because it matches the large-scale bias and the shot noise independently. In our set-up of periodic co-moving boxes, the two-point Gaussian term seems dominant against the missing contribution from a four-point connected term. Further exploration of analytical prescriptions is well motivated, e.g. to understand whether the same is true for more realistic situations (see next). In summary, if the requirement for approximate method covariances is to reproduce N-body derived parameter errors at the 5 per cent level, then this is already achieved by several methods on the most important parameters, but not all. A lesser requirement of 10 per cent is reached by all methods on all parameters, with the only exception of lognormal.

(iv) Lastly, we find that in general approximate mocks have slightly better performances in the less massive sample (Sample 1) than in Sample 2, especially when looking at the multipoles and their variance. This might be due to a larger non-linear contributions in the higher mass sample since we observe a similar behaviour in the two-point correlation function and bispectrum analysis (Colavincenzo et al. 2019; Lippich et al. 2019). There may also be a component of non-Poissonian shot noise from halo exclusion affecting the Fourier space analyses. A detailed analysis of this is beyond the scope of these articles, and we leave it for further work.

The analysis presented in this series of articles can be regarded as of general applicability for the next generation of galaxy surveys, but it has been developed with particular attention to scientific requirements of the *Euclid* mission. Despite the simplified setting used in this first analysis, these results confirm that usage of approximate methods for estimating the covariance matrix of power spectra is allowed at the cost of a small systematic error on the recovered parameter uncertainty (generally within 10 per cent, i.e. complying with *Euclid* requirements). Even though great care was put into minimizing the impact of noise in the comparison, the strength of our conclusions is somewhat limited by the number of mocks that we are using. We thus plan to perform follow-up studies with larger number of approximate mocks and to study more interesting regimes of mass resolution, which would allow us to reach number densities that are closer to those in future galaxy surveys. In addition, we have considered a rather simplistic scenario of halo snapshots in co-moving outputs with periodic boundary conditions. Further work will incorporate masking effects, addition of galaxies into haloes (e.g. in particular haloes populated using an HOD prescription for galaxies matching the H_α emitters that *Euclid* will observe), and inclusion of supersurvey modes. We expect that the inclusion of these effects will enhance covariances through the coupling between long and short modes, possibly making differences among methods more evident. Moreover, the modelling of such effects is not trivial in an analytical framework, such as the Gaussian approximation presented in this article, thus it will be interesting to test the performance of this method in more realistic scenarios.

ACKNOWLEDGEMENTS

This article and companion articles have benefitted from discussions and the stimulating environment of the *Euclid* Consortium, which is warmly acknowledged.

LB acknowledges support from the Spanish Ministerio de Economía y Competitividad (MINECO) grant ESP2015-66861.

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

MCr acknowledges support from the Spanish Ramón y Cajal MICINN program. MCr has been funded by AYA2015-71825.

MCo is supported by the *Departments of Excellence 2018 – 2022* Grant awarded by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) (L. 232/2016), by the research grant *The Anisotropic Dark Universe* Number CSTO161409, funded under the program CSP-UNITO *Research for the Territory 2016* by Compagnia di Sanpaolo and University of Torino; and the research grant TAsP (Theoretical Astroparticle Physics) funded by the Istituto Nazionale di Fisica Nucleare (INFN). PM and ES acknowledge support from grant MIUR PRIN 2015 *Cosmology and Fundamental Physics: illuminating the Dark Universe with Euclid* and from Consorzio per la Fisica di Trieste; they are part of the INFN InDark research group.

ML and AGS acknowledge support from the Transregional Collaborative Research Centre TR33 *The Dark Universe* of the German Research Foundation (DFG).

CDV acknowledges support from the MINECO through grants AYA2013-46886, AYA2014-58308, and RYC-2015-18078. SA acknowledges support from the UK Space Agency through grant ST/K00283X/1. AB acknowledges financial support from MINECO under the Severo Ochoa program SEV-2015-0548. MP acknowledges support from MINECO under the grant AYA2012-39702-C02-01. PF acknowledges support from MINECO through grant ESP2015-66861-C3-1-R and Generalitat de Catalunya through grant 2017-SGR-885. AI was supported in part by Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. He was also supported in part by NASA ROSES 13-ATP13-0019, NASA ROSES 14-MIRO-PROs-0064, NASA ROSES 12-EUCLID12-0004, and acknowledges support from the JAE program grant from the Spanish National Science Council (CSIC). RB, SC, and GS are supported by the Canadian Natural Sciences and Engineering Research Council (NSERC). GY acknowledges financial support from MINECO/FEDER (Spain) under research grant AYA2015-63810-P.

The Minerva simulations have been performed and analysed on the Hydra cluster and on the computing cluster for the *Euclid* project at the Max Planck Computing and Data Facility (MPCDF) in Garching.

ICE–COLA simulations were run at the MareNostrum supercomputer – Barcelona Supercomputing Center (BSC-CNS, http://www.bsc.es), through the grant AECT-2016- 3-0015.

PINOCCHIO mocks were run on the GALILEO cluster at CINECA because of an agreement with the University of Trieste.

PEAKPATCH simulations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario; Ontario Research Fund – Research Excellence, and the University of Toronto.

Numerical computations with HALOGEN were done on the Sciama High Performance Compute (HPC) cluster, which is supported by the ICG, SEPNet, and the University of Portsmouth.

PATCHY mocks have been computed in part at the MareNostrum supercomputer of the Barcelona Supercomputing Center because of a grant from the Red Española de Supercomputación (RES), and in part at the Teide High-Performance Computing facilities provided by the Instituto Tecnológico y de Energías Renovables (ITER, S.A.).

This work has used the NUMPY, MATPLOTLIB, and CHAINCONSUMER PYTHON packages.

REFERENCES

Agrawal A., Makiya R., Chiang C.-T., Jeong D., Saito S., Komatsu E., 2017, J. Cosmol. Astropart. Phys., 2017, 003
Alam S. et al., 2017, MNRAS, 470, 2617
Alcock C., Paczynski B., 1979, Nature, 281, 358
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Ata M. et al., 2018, MNRAS, 473, 4773
Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, MNRAS, 450, 1856
Avila S. et al., 2017, MNRAS, 479, 94
Ballinger W. E., Peacock J. A., Heavens A. F., 1996, MNRAS, 282, 877
Bond J. R., Myers S. T., 1996a, ApJS, 103, 1
Bond J. R., Myers S. T., 1996b, ApJS, 103, 41
Bond J. R., Myers S. T., 1996c, ApJS, 103, 63
Carretero J., Castander F. J., Gaztañaga E., Crocce M., Fosalba P., 2015, MNRAS, 447, 646
Chan K. C., Scoccimarro R., Sheth R. K., 2012, Phys. Rev. D, 85, 083509
Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015a, MNRAS, 446, 2621
Chuang C.-H. et al., 2015b, MNRAS, 452, 686
Colavincenzo M. et al., 2019, MNRAS, 482, 4883
Cole S. et al., 2005, MNRAS, 362, 505
Dawson K. S. et al., 2016, AJ, 151, 44
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273
Fosalba P., Crocce M., Gaztañaga E., Castander F. J., 2015, MNRAS, 448, 2987
Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, MNRAS, 457, 1577
Grieb J. N. et al., 2017, MNRAS, 467, 2085
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Heitmann K. et al., 2015, ApJS, 219, 34
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Howlett C., Ross A. J., Samushia L., Percival W. J., Manera M., 2015, MNRAS, 449, 848
Ivezić Ž. et al., 2008, preprint (arXiv:0805.2366)
Izard A., Crocce M., Fosalba P., 2016, MNRAS, 459, 2327
Izard A., Fosalba P., Crocce M., 2018, MNRAS, 473, 3051
Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, MNRAS, 450, 1836
Kitaura F.-S. et al., 2016, MNRAS, 456, 4156
Koda J., Blake C., Beutler F., Kazin E., Marin F., 2016, MNRAS, 459, 2118
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lippich M. et al., 2019, MNRAS, 482, 1786
LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
Mandelbaum R. et al., 2018, PASJ, 70, S25
Manera M. et al., 2013, MNRAS, 428, 1036
Manera M. et al., 2015, MNRAS, 447, 437
McDonald P., 2006, Phys. Rev. D, 74, 103512
Monaco P., 2016, Galaxies, 4, 53
Monaco P., Theuns T., Taffoni G., Governato F., Quinn T., Stadel J., 2002, ApJ, 564, 8
Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, MNRAS, 433, 2389
Munari E., Monaco P., Koda J., Kitaura F.-S., Sefusatti E., Borgani S., 2017a, JCAP, 7, 050
Munari E., Monaco P., Sefusatti E., Castorina E., Mohammad F. G., Anselmi S., Borgani S., 2017b, MNRAS, 465, 4658
Nishimichi T., Taruya A., 2011, Phys. Rev. D, 84, 043526
Percival W. J. et al., 2001, MNRAS, 327, 1297
Percival W. J. et al., 2014, MNRAS, 439, 2531
Pezzotta A. et al., 2017, A&A, 604, A33

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophys. Cosmol., 4, 2
Samushia L., Percival W. J., Raccanelli A., 2012, MNRAS, 420, 2102
Sánchez A. G. et al., 2013, MNRAS, 433, 1202
Sánchez A. G. et al., 2017, MNRAS, 464, 1640
Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1
Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624
Smith A., Cole S., Baugh C., Zheng Z., Angulo R., Norberg P., Zehavi I., 2017, MNRAS, 470, 4646
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Stein G., Alvarez M. A., Bond J. R., 2018, MNRAS, 483, 2236
Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shiroyama K., 2017, ApJ, 850, 24
Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 2013, 036
Taylor A., Joachimi B., Kitching T., 2013, MNRAS, 432, 1928
Tegmark M. et al., 2004, Phys. Rev. D, 69, 103501
Troxel M. A. et al., 2017, Phys. Rev. D, 98, 043528
White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594
Zhao C., Kitaura F.-S., Chuang C.-H., Prada F., Yepes G., Tao C., 2015, MNRAS, 451, 4266

APPENDIX A: SAMPLE DEFINITION MATCHING HALO MASS OR HALO ABUNDANCE

Here, we show the difference in the results for samples defined by cutting at the same halo mass threshold as opposed to matching the abundance of N-body haloes, as done in the main body of the article. We do this only for those methods that produce halo samples independently of the N-body ('predictive') as the remaining methods work by reproducing abundance and mass distribution of the N-body samples, so selecting by mass thresholds or abundance matching is equivalent. In Table A1, we list the minimum halo mass and the average number of haloes for these samples, for ICE–COLA, PINOCCHIO, and PEAKPATCH.

In Figs A1 and A2, we show the average power spectrum multipoles and their variance, respectively. In Fig. A3, we show how these differences translate into parameter errors, depicting with

Table A1. Characteristics of the mass threshold samples used in Appendix A, together with the abundance matched ones (marked with ∗) used in the main body of the article.

Method	Selection type	$M_{\rm min}$ ($h^{-1}\,{\rm M}_\odot$)	$\bar{n}_{\rm haloes}$ ($h^3\,{\rm Mpc}^{-3}$)
	Sample 1		
N-body	Mass threshold	1.121×10^{13}	2.130×10^{-4}
ICE–COLA	Mass threshold	1.121×10^{13}	2.057×10^{-4}
PINOCCHIO	Mass threshold	1.121×10^{13}	1.948×10^{-4}
ICE–COLA	abund. match∗	1.086×10^{13}	2.122×10^{-4}
PINOCCHIO	abund. match∗	1.044×10^{13}	2.147×10^{-4}
	Sample 2		
N-body	Mass threshold	2.670×10^{13}	5.441×10^{-5}
ICE–COLA	Mass threshold	2.670×10^{13}	5.719×10^{-5}
PINOCCHIO	Mass threshold	2.670×10^{13}	5.258×10^{-5}
PEAKPATCH	Mass threshold	2.670×10^{13}	4.450×10^{-5}
ICE–COLA	abund. match∗	2.766×10^{13}	5.455×10^{-5}
PINOCCHIO	abund. match∗	2.631×10^{13}	5.478×10^{-5}
PEAKPATCH	abund. match∗	2.355×10^{13}	5.439×10^{-5}

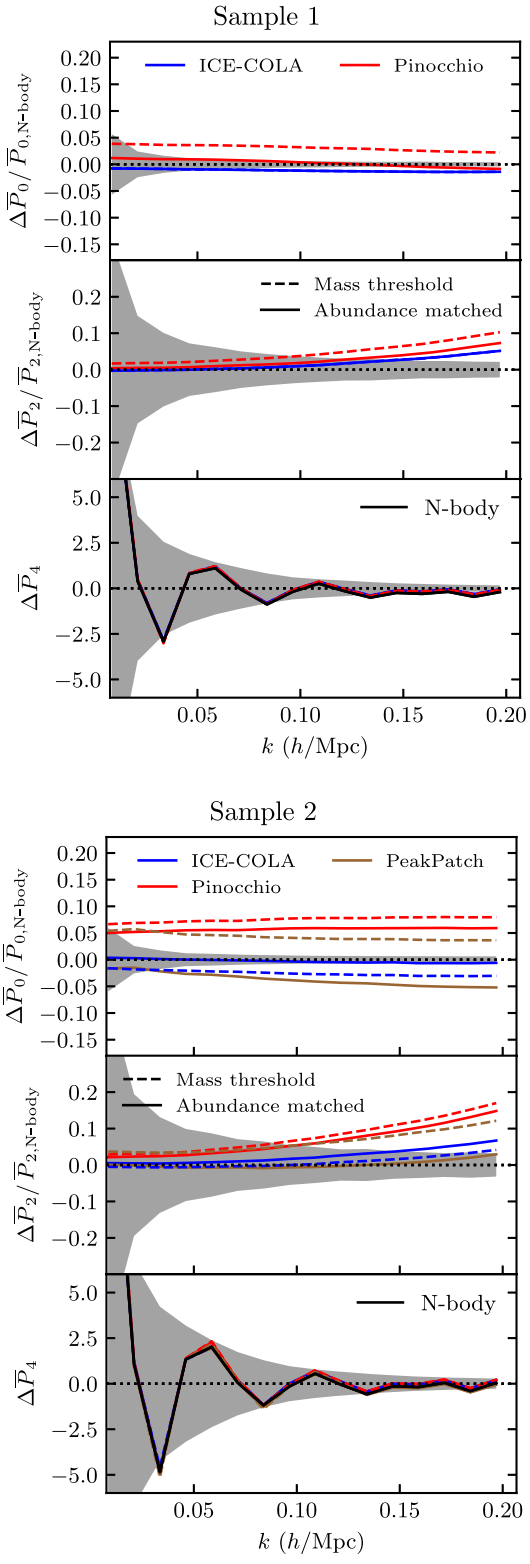

Figure A1. Relative difference of the power spectrum monopole and quadrupole (hexadecapole) with respect to N-body (model) for the sample 1 (top plot) and sample 2 (bottom plot). Each plot shows the monopole (top panel), quadrupole (middle panel), and hexadecapole (bottom panel) for mass samples (dashed lines) and density samples (solid lines).

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

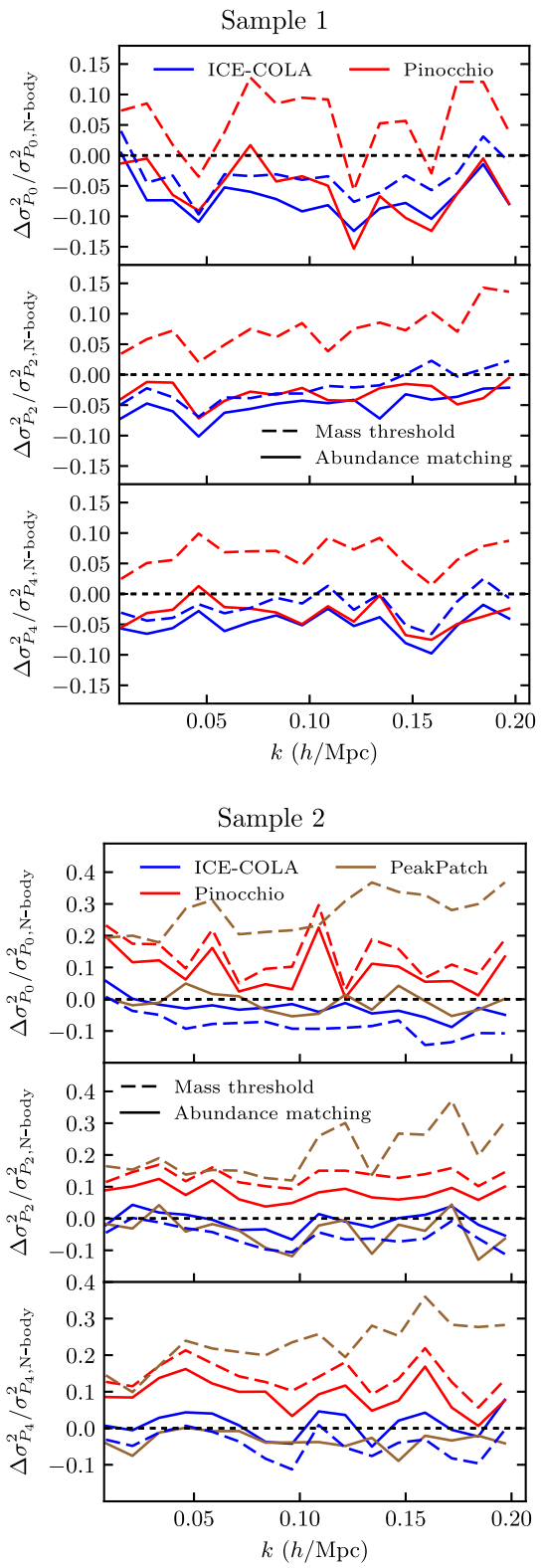

Figure A2. Relative difference of the variance of power spectrum multipoles with respect to N-body for the sample 1 (top plot) and sample 2 (bottom plot). Each plot shows monopole (top panel), quadrupole (middle panel), and hexadecapole (bottom panel) for mass samples (dashed lines) and density samples (solid lines).

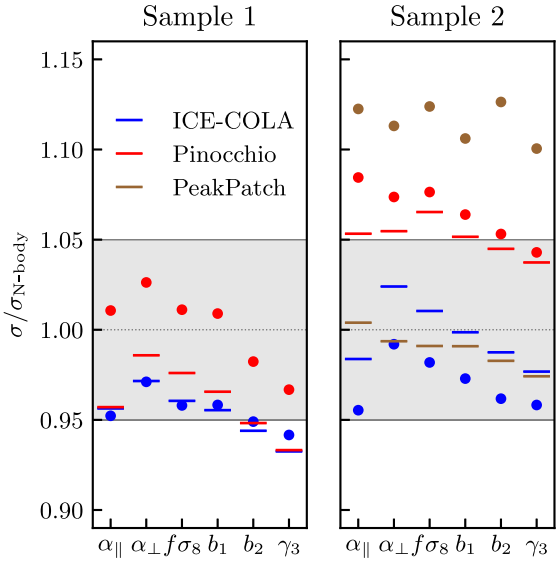

Figure A3. Ratio of the error on cosmological and nuisance parameters with respect to N-body for sample 1 (left) and sample 2 (right) for mass samples (dots) and density samples (lines). The grey shaded area indicates 5 per cent differences (color online).

bars the results from abundance matching (as shown previously in Fig. 9) and with dots the ones using mass thresholds.

In the case of ICE–COLA, the abundance matching works noticeably better in the most massive sample, while it is only marginally worse than the direct selection by mass at the low-mass sample. This is compatible with the findings of Izard et al. (2016), where the following trends were identified: (i) halo masses are underestimated by ∼2–3 per cent, and (ii) at low masses the halo catalogues are incomplete (and the size of this effect will depend, for example, on the PM grid size or the number of time-steps). At high masses (Sample 2), one can simply correct halo masses by doing abundance matching (or clustering matching). This effectively selects more massive haloes, thus increasing the clustering amplitude and bringing the shot-noise level closer to its right value. Otherwise, shot noise is underestimated and so are the errors on parameters (since in this regime shot noise is not negligible in the covariance). This is the behaviour seen in the right-hand panel of Fig. A3. In turn, closer to the halo mass resolution of the catalogues (Sample 1), the effect is the opposite. Selecting haloes by the same mass threshold as the N-body already shows slightly less clustering and abundance than the N-body. Imposing abundance matching shifts the mass scale and lowers the clustering amplitude even more (see top panel of Fig. A1). None the less, in this mass regime the effect is small and the impact on parameter error is negligible. Both mass and abundance matching yield equivalent results in parameter errors for Sample 1, see left-hand panel of Fig. A3.

For PINOCCHIO, doing abundance matching always implies selecting effectively less massive haloes and lowering the clustering amplitude. Since the raw clustering amplitudes are overpredicted at both Samples 1 and 2, the abundance matched helps in both cases. The performance of the abundance-matched selection is better in Sample 2 only in terms of parameter errors, see Fig. A3.

PEAKPATCH shows a similar behaviour in Sample 2 but this is much more expected. Haloes in PEAKPATCH resemble by construction a more spherical structure closer perhaps to spherical

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

overdensity samples, it is not calibrated to reproduce an FOF mass function. Hence, in this case, mass thresholds yield lower abundance than the N-body SUBFIND / FOF haloes, and hence higher clustering. The low abundance translates into higher shot noise, and very overestimated error bars (see second panel of Fig. A3). This is solved by abundance matching that brings σ and $\sigma_{\text{N-body}}$ to within few per cent.

Overall, we find that abundance matching improves the agreement between samples in the approximate methods and the ones in N-body, at least in terms of parameter error bars. In article I of this series (Lippich et al. 2019), this comparison is extended to include samples defined by matching clustering amplitudes.

APPENDIX B: ERRORS ASSOCIATED WITH VARIANCE CANCELLATION

In this appendix, we estimate the error in our results due to the finite number of N-body simulations that we are using. To do this, we use a larger set of 10 000 PINOCCHIO realizations with the same simulation and cosmological parameters.

We expect our results to depend to some extent on the number of the N-body realizations we are using as benchmark for the approximate methods (Hartlap, Simon & Schneider 2007; Dodelson & Schneider 2013; Taylor, Joachimi & Kitching 2013; Percival et al. 2014). The central figure in our comparison is the ratio between the parameter uncertainties obtained from the approximate mocks and from the N-body, so in this appendix we will attempt to estimate the error on such quantity as well as the advantages provided by matching the seeds of the 300 realizations. We will also consider more extensively how our results vary as a function of the largest wavenumber included in the analysis, k_{max}.

In order to do this, we will consider a simpler model depending on just two parameters that allows a fully analytical analysis of the likelihood function. We write the halo power spectrum as

$$P_{\text{model}}(k) = b_1^2\, P_{\text{L}}(k) + P_{\text{SN}}\,, \tag{B1}$$

where b_1 is the linear bias parameter and P_{SN} is a constant accounting for the shot-noise component, while we assume the linear matter power spectrum $P_{\text{L}}(k)$ to be known. The likelihood function is therefore given by

$$\ln \mathcal{L}_P = -\frac{1}{2}\sum_{ij}^{k_{\text{max}}} \delta P(k_i)\, \left[\mathbf{C}\right]^{-1}_{ij}\, \delta P(k_j)\,, \tag{B2}$$

where $\delta P \equiv P_{\text{data}} - P_{\text{model}}$, while $\mathbf{C}$ is the covariance matrix. Such model can be rewritten as

$$P_{\text{model}} = \sum_{\alpha=1}^{2} p_\alpha\, \mathcal{P}_\alpha, \tag{B3}$$

where $\{p_\alpha\} = \{b_1^2, P_{SN}\}$ and $\{\mathcal{P}_\alpha\} = \{P_L(k), 1\}$. Adding $p_0 = -1$ and $\mathcal{P}_0 = P_{\text{data}}$, we can also write

$$-\delta P = P_{\text{model}} - P_{\text{data}} = \sum_{\alpha=0}^{2} p_\alpha\, \mathcal{P}_\alpha \tag{B4}$$

and therefore it is easy to see that we can rewrite the likelihood as

$$\ln \mathcal{L}_P = -\frac{1}{2}\sum_{\alpha,\beta}^{2} p_\alpha\, p_\beta\, \mathcal{D}_{\alpha\beta}\,, \tag{B5}$$

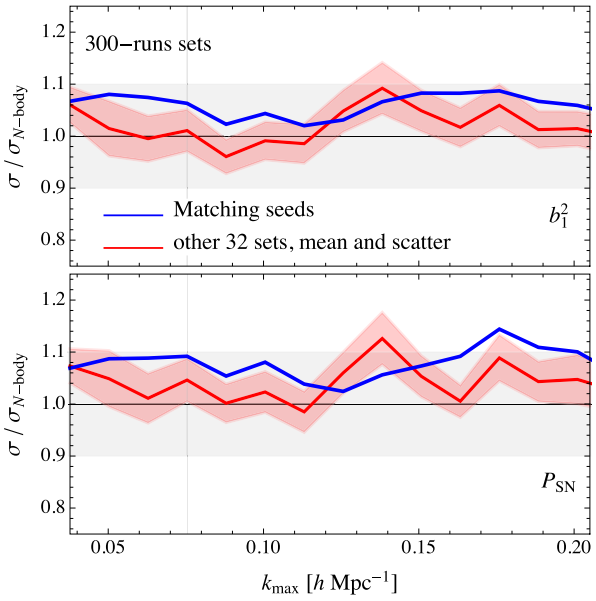

Figure B1. Comparison between the ratio of marginalized uncertainties $\sigma/\sigma_{\text{N-body}}$ obtained from the 300 PINOCCHIO and N-body realizations with matching seeds (blue curve) and the same quantity obtained from 32 distinct sets of 300 PINOCCHIO realizations and the same 300 N-body runs (red curve). The shaded area corresponds to the scatter across the 32 sets. The power spectrum model here depends on only two parameters, b_1^2 and P_{SN}. See the text for an explanation of the assumed likelihood function and power spectrum model (color online).

where

$$\mathcal{D}_{\alpha\beta} \equiv \sum_{i,j}^{k_{\text{max}}} \mathcal{P}_\alpha(k_i)\, \left[\mathbf{C}\right]^{-1}_{ij}\, \mathcal{P}_\beta(k_j)\,. \tag{B6}$$

In this way, $\mathcal{L}_P$ is explicitly written as an exact, multivariate Gaussian distribution in the parameters p_α. Clearly, once the quantities $\mathcal{D}_{\alpha\beta}$ are computed, we can evaluate any marginalization analytically. This allows us to easily derive results for any value of k_{max} and present them as a function of such quantity.

As a first test, we compare the ratio of marginalized uncertainties $\sigma/\sigma_{\text{N-body}}$ for 300 PINOCCHIO runs with matching seeds to the same quantity derived from 32 distinct sets of 300 PINOCCHIO runs with independent initial conditions. We should notice that this will not provide a full estimate of the error on the ratio $\sigma/\sigma_{\text{N-body}}$ since the 32 sets are compared always with the same, single set of N-body realizations. Still, we can naïvely expect the actual relative error on $\sigma/\sigma_{\text{N-body}}$ to be $\sqrt{2}$ larger than the one we derive in this way. The results are shown in Fig. B1. The blue curve shows the result for matching seeds as a function of k_{max}, top and bottom panels corresponding to the ratio of the marginalized uncertainties on, respectively, b_1^2 and P_{SN}. The red curve shows the mean across the 32 estimates of uncertainties ratio with non-matching seeds, while the shaded area shows its scatter.

In the first place, we notice how the ratio can vary significantly as a function of k_{max}, even when we assume matching initial conditions. One should bear this in mind when discussing the recovered values of $\sigma/\sigma_{\text{N-body}}$ presented in the main text. For non-matching seeds, the scatter across different values of k_{max} is larger, as we can expect. None the less, in both cases the scatter is within the overall

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

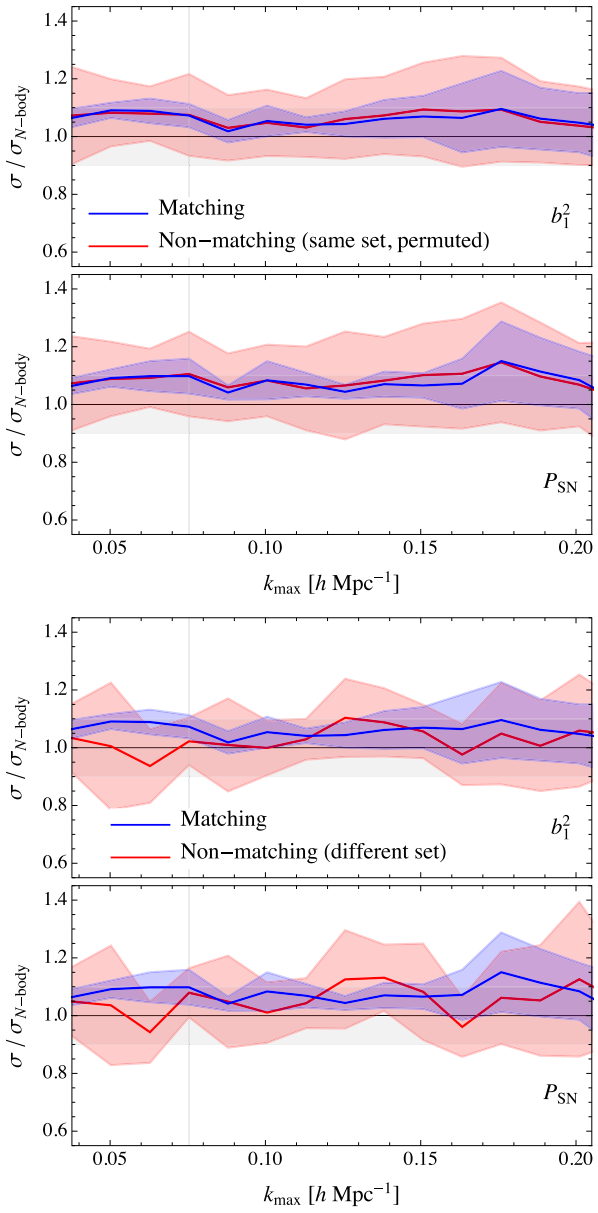

Figure B2. The top panels show the mean and standard deviation (shaded area) of the ratios of the marginalized errors on the two parameters b_1^2 and P_{SN} obtained from the PINOCCHIO covariance to the same errors obtained from the Minerva covariance. The blue line corresponds to the case of six sets of 50 realizations, where we match the seed of the two methods. The red line is obtained from the same sets but permuted in order to illustrate the case of non-matching seeds. The bottom panels show the same results for the matching seeds case, while the other case considers an independent set of PINOCCHIO runs (color online).

uncertainty on the ratio that we estimate to be ∼4–5 per cent (see next).

The comparison of the mean over the independent sets and the matching seeds results shows in some cases quite large differences, but we should keep in mind that the former assumes always the same N-body set and therefore can be affected by large fluctuations in these realizations. Such fluctuations, on the other hand, can be compensated in the matching seed case. From this figure, we see that the scatter in the recovered value of σ is ∼4–5 per cent. Remarkably, a direct application of the formulae in Taylor et al. (2013; their equations 55 and 56) results in an estimate for this error, assuming $N_S = 300$ runs and $N_D = 3 \times 16$ data points, of ∼4.5 per cent, which is in very good agreement. Lastly, if we estimate the full error on the ratio to be $\sqrt{2}$ times the error estimated here from the scatter in the numerator of the ratio alone we recover a 7–8 per cent uncertainty. Differences between the solid blue and red lines in Fig. B1 are well within these values.

As an additional test, we subdivide the 300 mocks in six sets of 50 realizations. This number should ensure a marginally reliable estimation of the power spectrum covariance matrix, at least for small values of k_{max}. In this way, we can evaluate the mean and the scatter of the ratio $\sigma/\sigma_{\text{N-body}}$ between PINOCCHIO and the N-body results across six pairs, and these can be computed for six matching pairs as well as six non-matching pairs obtained by permutations of the subsets. The results are shown in the upper panels of Fig. B2. This exercise shows that matching the initial conditions reduces the scatter on the uncertainties ratio by roughly a factor of 2 at large scales ($k \lesssim 0.1\,h\,\text{Mpc}^{-1}$) with respect to the case of independent runs. Clearly, such reduction is less significant at smaller scales, as one can expect. In the lower panels of Fig. B2, for the comparison among non-matching realizations shown by the red curve, we replace the six subsets of PINOCCHIO runs with other six, independent runs in order to avoid the permutation. Clearly, the blue curve showing the mean ratio from matching seeds is the same as before. It may appear that in this more general comparison between non-matching realizations the mean value of the independent set is closer to one, indicating, apparently, a better performance of PINOCCHIO with respect to our main results. In fact, this is simply reflecting an overall fluctuation of the first 300 N-body runs. Such fluctuation is captured to a large extent by the PINOCCHIO realizations with matching seeds allowing to trust the few per cent overestimate of the parameters uncertainties provided by the approximate method.

To summarize, we estimate an uncertainty of about 7–8 per cent coming from the fact that our comparisons are limited to 300 N-body and approximate method runs, which is, however, reduced to about 4–5 per cent from adopting the same initial conditions in both cases.

[1]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[2]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[3]*Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34143 Trieste, Italy*
[4]*Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Via Valerio, 2, I-34127 Trieste, Italy*
[5]*Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Giessenbach str, D-85741 Garching, Germany*
[6]*Universitäts-Sternwarte München, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, D-81679 Munich, Germany*
[7]*Dipartimento di Fisica, Università di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[8]*Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[9]*Dipartimento di Fisica, Sezione di Astronomia, Università di Trieste, via Tiepolo 11, I-34143 Trieste, Italy*
[10]*Berkeley Center for Cosmological Physics, Campbell Hall 341, University of California, Berkeley, CA 94720, USA*
[11]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85741 Garching, Germany*
[12]*Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK*
[13]*Instituto de Astrofísica de Canarias, C/Vía Láctea, s/n, E-38200 La Laguna, Tenerife, Spain*

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022

[14] *Departamento Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain*
[15] *Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St George Street, Toronto, ON M5S 3H8, Canada*
[16] *Institut d'Astrophysique de Paris, CNRS & Sorbonne Université, UMR 7095, 98 bis boulevard Arago, F-75014 Paris, France*
[17] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[18] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[19] *Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands*
[20] *Departamento de Física Teórica, Módulo 15, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*
[21] *Centro de Investigación Avanzada en Física Fundamental (CIAFF), Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

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

Downloaded from https://academic.oup.com/mnras/article/485/2/2806/5342086 by inaf user on 17 March 2022