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

Blot, L.; Crocce, M.; Sefusatti, E.; Lippich, M.; Sánchez, A.G.; Colavincenzo, M.; … ; Yepes, G.

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[9,3,4],
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[17,18,1,2], Francisco-Shu Kitaura[13,14], Marcos Pellejero-Ibanez[13,14],
George Stein[15], Mohammadjavad Vakili[19], Gustavo Yepes[20,21].

[1] *Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain*
[2] *Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain*
[3] *Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Trieste, via Tiepolo 11, 34143 Trieste, Italy*
[4] *Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Via Valerio, 2, 34127 Trieste, Italy*
[5] *Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Giessenbachstr., 85741 Garching, Germany*
[6] *Universitäts-Sternwarte München, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 Munich, Germany*
[7] *Dipartimento di Fisica, Università di Torino, Via P. Giuria 1, 10125 Torino, Italy*
[8] *Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy*
[9] *Dipartimento di Fisica, Sezione di Astronomia, Università di Trieste, via Tiepolo 11, 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, 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, 38200, La Laguna, Tenerife, Spain*
[14] *Departamento Astrofísica, Universidad de La Laguna, 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, 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, 2300 RA, Leiden, The Netherlands*
[20] *Departamento de Física Teórica, Módulo 15, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[21] *Centro de Investigación Avanzada en Física Fundamental (CIAFF), Universidad Autónoma de Madrid, 28049, Madrid, Spain*

*E-mail:blot@ice.cat

11 November 2019

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 constrains 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 onto dark-matter overdensities calibrated with a target N-body run (HALOGEN, PATCHY) and two standard assumptions about the full density PDF (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%, while residuals for the quadrupole are sometimes larger and scale dependent. The variance of the multipoles is typically reproduced within 10%. Overall, we find that covariances built from ICE-COLA, PINOCCHIO, PEAKPATCH, PATCHY and the Gaussian approximation yield errors on model parameters within 5% of those from the N-body based covariance, while for HALOGEN and lognormal this agreement degrades to $\sim 10\%$.

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 constrains 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 like the Lyman-alpha 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 et al. 2016) or LSST (Ivezic et al. 2008; LSST Science Collaboration et al. 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 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 datasets. 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 (SAM) 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 et al. 2017), in particular for future surveys such as *Euclid* or DESI, that 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 sub-structures. 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; Manera et al. 2015; Howlett et al. 2015; Kitaura et al. 2016; Koda et al. 2016; Avila et al. 2017). Nonetheless 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*, is that systematic errors in the estimation of covariance matrices should not bias the estimation of errors in cosmological parameters by more than ∼ 10%.

Hence, in this paper we perform a robust and thorough comparison of clustering measurements in ensembles of mock catalogues produced from several state-of-the-art algorithms, that 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 because we concentrate on how well the observables are reproduced with high precision, thanks to the large number of realisations (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 papers present similar studies in terms of two-point correlation function (Lippich et al. 2018) and bispectrum (Colavincenzo et al. 2018).

We span different types of approaches: one representing the combination of multi-step second order Lagrangian Perturbation Theory (LPT) to solve the large-scale displacements with a fast Particle Mesh (PM) solver to speed up the intermediate scale force computation, ICE-COLA (Izard et al. 2016, 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; Alvarez et al. 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 et al. 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 halos as measured in the benchmark N-body simulation. There are several other methods in the literature (e.g. QUICKPM by White et al. (2014), FASTPM by Feng et al. (2016), EZ-MOCKS 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 paper is organised as follows: in Sec. 2 we describe all the approximate methods that we consider and how they compare in terms of computational cost. In Sec. 3 we introduce the halo samples over which we do the compari-

---

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

son. Sec. 4 describes our methodology. Section 5 contains and discusses our results, which are summarised in Sec. 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 summarise the main features of the methods and mocks used in this work. For a comprehensive description we refer the reader to the first paper of this series (Lippich et al. 2018), 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*, that do not require re-calibration against a parent N-body for each specific sample or cosmological model (ICE-COLA, Pinocchio and PeakPatch), *calibrated methods*, that need prior information about the sample to simulate (Halogen and Patchy) and *analytical methods*, that 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 realisations were run, for a total of 300 realisations. These are performed using the code Gadget-3 (Springel 2005), with $N_p = 1000^3$ particles in a box of linear size $L_{box} = 1500\,h^{-1}$ Mpc and a starting redshift of $z_{ini} = 63$. The mass resolution is therefore $2.67 \times 10^{11}\,h^{-1}\,M_\odot$. The initial conditions are given by the 2LPTic code[2] and the cosmological parameters are fixed to their best fit 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, that solves for particle trajectories using a combination of second order Lagrangian Perturbation Theory (2LPT) and a Particle-Mesh (PM) algorithm, where 2LPT is used for integrating the large scale dynamics while the PM part solves for the small scale dynamics (Tassev et al. 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 paper we use the ICE-COLA version code, as optimised 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 $z_{ini} = 19$ to $z = 0$[3] and a PM grid-size of $3 \times N_p^{(1/3)}$ in each spatial dimension. They share the same particle load and box-size as Minerva. Note that this configuration is optimised for reproducing dark-matter clustering and weak lensing observables (Izard et al. 2018) and it is somewhat over-demanding in terms of computational cost for halo clustering alone (Izard et al. 2016), as seen in Table 1. Halos are found on-the-fly using a Friends of Friends (FoF) algorithm with a linking length of $b = 0.2$ the mean inter-particle 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 summarised 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 halos through accretion of matter and merging with other halos. Finally, halos 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 (Alvarez et al. 2018), originally developed by Bond & Myers (1996a,b,c). This approach is a Lagrangian space halo finder that associates halos to 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 halos according to an exponential bias prescription. The algorithm used in this work can be summarised 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 grid of linear size 300. Halos are assigned to grid-cells with a probability proportional to $\rho_{cell}^{\alpha(M_h)}$, where M_h is a halo mass sampled from the average mass function of the Minerva simulations. The position of the halo is drawn

---

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

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

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 $\sim 50\%$ 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 halos of the first sample a higher resolution version of PEAKPATCH should be run, requiring more computational resources than quoted here.

randomly from particle positions within the grid cell, while the velocity is given by the particle velocity multiplied by a velocity bias factor $f_{vel}(M_h)$. The function $\alpha(M_h)$ is calibrated using the average Minerva 2-point correlation function while the velocity bias one, $f_{vel}(M_h)$, 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), that uses augmented LPT (ALPT) to evolve the initial density field to the redshift of interest and then assigns halos using a non-linear scale-dependent and stochastic biasing prescription. The algorithm used in this work can be summarised 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 num-

ber density of halos is determined by using a deterministic bias prescription and halos are assigned to grid cells by sampling a negative binomial distribution, with a stochastic bias parametrisation to model deviations from Poissonity. 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 set 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), which models the matter and halo density fields as lognormal fields. The algorithm used in this work can be summarised as follows: a

Gaussian field $G(\vec{x})$ is generated on a grid of linear size 256 for matter and halos. The generation of these Gaussian fields requires as input the matter and halo two-point correlation function measured in Minerva. Both fields are then transformed into lognormal fields using the measured variance of the Gaussian fields, as $\delta_{lg}(\vec{x}) = \exp\left[-\sigma_G^2 + G(\vec{x})\right] - 1$. The velocity field is computed by solving the linear continuity equation for matter. The number of halos in a given cell is then determined $\bar{n}_h[1 + \delta_h(\vec{x})]V_{cell}$, where $\bar{n}_h$ is the mean number density of halos in the Minerva simulations, $\delta_h(\vec{x})$ is the halo density field and V_{cell} is the volume of the cell. Finally, halos are randomly positioned within the cell and their velocity is given by the velocity of the cell. 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 ICs.

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 super-sample covariance are neglected, so that the covariance is diagonal. The binned power spectrum multipole covariance in the thin shell approximation is given by:

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

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.

3 HALO SAMPLES

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

In Figure 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. Results for the higher mass limit are more robust because the halos 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 halos 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.

We define equivalent samples in the approximate mocks by matching the abundance of halos to the average abundance in the N-body simulations. The calibrated methods
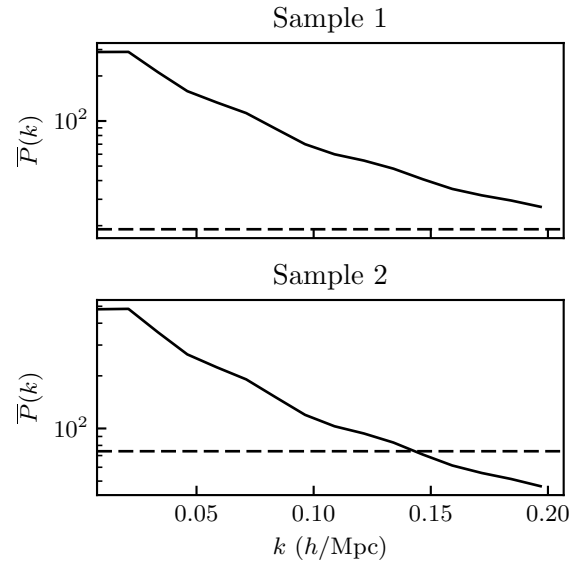

Figure 1. Average real space halo power spectrum from the 300 N-body simulations (continuous line) and the corresponding Poisson shot noise level (dashed line), for sample 1 (top) and sample 2 (bottom) as defined in the top entries of Table 2, at redshift 1.

have the same abundance of halos 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 halos because the mass function is discretised in steps corresponding to the particle mass, nevertheless the recovered abundances agree within 1%. 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 halos between $N \times M_p$ and $(N+1) \times M_p$, where M_p is the particle mass of the Minerva simulations. The PEAKPATCH mocks, as implemented here, do not resolve the lower mass halos, hence results are presented only for the second sample. In Figure 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 percent 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, we expect shot-noise to have an important role in the covariance of these samples, so 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 halos 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
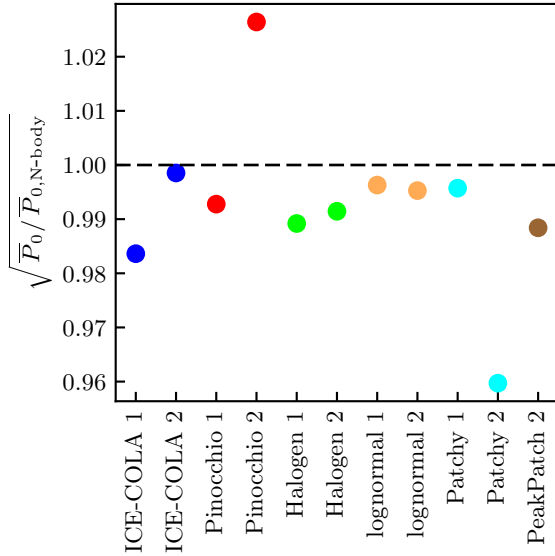
Figure 2. Square root of the average clustering amplitude of the approximate mocks divided by the average of the N-body simulations in the range $0.008 < k(h\,\mathrm{Mpc}^{-1}) < 0.096$ for the two samples, as indicated in the labels.

Method	$\bar{n}_{halos}$ $(h^3\mathrm{Mpc}^{-3})$	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}

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 halos 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 halos with $M_h \geqslant 2.67 \times 10^{13}\,h^{-1}\,\mathrm{M}_{\odot}$ in the N-body. Second column displays the resulting abundance of those samples, which are not exactly equal to the N-body due to mass function discretisation in steps equivalent to the particle mass. Samples are defined at $z = 1$.

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 halos are identified as lagrangian spherical overdensities, contrary to the other methods that use friends-of-friends (FoF) or are calibrated to reproduce FoF mass functions.

4   METHODOLOGY

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

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 $0.008 < k(h\,\mathrm{Mpc}^{-1}) < 0.197$. We measure the power spectrum multipoles P_ℓ ($\ell = 0, 2, 4$) using the public code PowerI4[4], that employs the interlacing technique to reduce aliasing contribution, and we use the 4th order mass interpolation scheme on a grid of linear size of 256. We bin the power spectrum multipoles in intervals of $\Delta k = 3k_f$, where $k_f = 2\pi/L_{box}$ 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 which 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 model includes: (1) nonlinear matter clustering through the matter and velocity divergence correlations (gRPT, Crocce et al. in prep.), (2) a biasing scheme to one-loop in the initial power spectra (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 halos and these are not affected by this effect. Broadly speaking, the theory model is given by:

$$\begin{aligned}
P_{hh}^s(k,\mu) &= P_{hh}(k) + 2f\mu^2 P_{h\theta}(k) + f^2\mu^4 P_{\theta\theta}(k) \\
&+ P_{novir}^{(2)}(k,\mu) + P_{novir}^{(3)}(k,\mu)
\end{aligned} \tag{2}$$

where P_{hh}, $P_{h\theta}$ and $P_{\theta\theta}$ stand for halo density h and velocity divergence θ auto and cross correlations, and f is the growth rate of structure. The first three terms are the nonlinear "Kaiser" effect, with P_{hh} and $P_{h\theta}$ 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 et al. (2012). The last two terms in Eq. (2) also arise from

---

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

the transformation of real to redshift space coordinates,

$$P_{\text{novir}}^{(2)}(k,\mu) = \int \frac{q_z}{q^2} \Big[B_{\theta D_s D_s}(\mathbf{q}, \mathbf{k}-\mathbf{q}, -\mathbf{k})$$
$$+ B_{\theta D_s D_s}(\mathbf{q}, -\mathbf{k}, \mathbf{k}-\mathbf{q}) \Big] d^3 q, \qquad (3)$$

which can be evaluated at tree-order (thus, depends only on b_1, b_2 and γ_2), while

$$P_{\text{novir}}^{(3)}(k,\mu) = \int \frac{q_z}{q^2} \frac{(k_z - q_z)}{(\mathbf{k}-\mathbf{q})^2} (b_1 + f\mu_q^2)(b_1 + f\mu_{k-q}^2)$$
$$\times P_{\delta\theta}(k-q) P_{\delta\theta}(q) d^3 q, \quad (4)$$

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 in Eq. (18) of Taruya et al. (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 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 Alcock-Paczynski (AP) effect (Alcock & Paczynski 1979), that can be used to place cosmological constrains (e.g. Samushia et al. (2012)). Such distortions are characterised by two parameters that transforms comoving coordinates in the "true" cosmology, denoted $(k_\perp, k_\parallel)$[5], to coordinates in the "fiducial" cosmology, denoted $(k'_\perp, k'_\parallel)$, as

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

with

$$\alpha_\perp = \frac{D_A(z) r'_d}{D'_A(z) r_d} \quad \alpha_\parallel = \frac{H'(z) r'_d}{H(z) r_d}, \qquad (6)$$

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 line-of-sight this transformation is (Ballinger et al. 1996),

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

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

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

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

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 $\alpha_\perp$, $\alpha_\parallel$.

---

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

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

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

4.3 Covariance Matrix Estimation

The covariances are computed using the sample covariance estimator:

$$C_\ell(k_i, k_j) = \frac{1}{N_s - 1} \sum_{n=1}^{N_s} (P_\ell^n(k_i) - \bar{P}_\ell(k_i))(P_\ell^n(k_j) - \bar{P}_\ell(k_j)), \qquad (10)$$

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:

$$C_{\ell_1, \ell_2}(k_i, k_j) = \frac{1}{N_s - 1} \sum_{n=1}^{N_s} (P_{\ell_1}^n(k_i) - \bar{P}_{\ell_1}(k_i))(P_{\ell_2}^n(k_j) - \bar{P}_{\ell_2}(k_j)). \qquad (11)$$

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 induce a bias on the recovered parameters, or if they systematically increase/decrease the errors on such parameters we run MCMC chains based on the assumption of a Gaussian likelihood:

$$-2\log\mathcal{L} = \sum_{ij} (\hat{P}(k_i) - P_m(k_i))^T \Psi(k_i, k_j)(\hat{P}(k_j) - P_m(k_j)), \qquad (12)$$

where $\Psi = C^{-1}$ 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. We fix the cosmological parameters to their true values and only vary the bias parameters b_1, b_2 and γ_3^-. The best fit 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 Alcock-Paczynski 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 Paper 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 and we consider a burn-in phase of 8×10^4 points. We have checked that our results do not depend on the burn-in assumption.

5   RESULTS

5.1   Mean of the power spectrum multipoles

Figure 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 show the absolute difference since the values are very close to 0. The shaded regions show the error on the average 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.

The monopole is reproduced to within $\sim 5\%$ 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% 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% 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 $\sim 5\%$ by $k = 0.2\,h^{-1}\,\mathrm{Mpc}$ and PINOCCHIO $\sim 14\%$. PATCHY shows $5\% - 10\%$ deviations in the quadrupole. This performance can be potentially improved with an explicit fit to the velocity dispersion, modelling virialised 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 when taking the ratio, but these are on average of the order of $10 - 15\%$ for ICE-COLA, PINOCCHIO, PEAKPATCH and PATCHY while they are $> 40\%$ for HALOGEN and Lognormal.

5.2   Variance of the power spectrum multipoles

In Figure 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). Overall the Gaussian model performs well, at the level of the mock based variances.

In Sample 1 all the methods agree with the N-body variance within 10% on large scales, with the exception of Lognormal for reasons that we address below. HALOGEN reaches a 20% difference beyond $k \sim 0.1\,h^{-1}\,\mathrm{Mpc}$.

In Sample 2 ICE-COLA and PEAKPATCH do particularly well with an average of $\sim 3\%$ 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 $\sim 9\%$ and HALOGEN of 30%. The quadrupole variance is reproduced at the level of $5 - 10\%$ while the hexadecapole variance is within 5% for all methods, with the exception of PINOCCHIO and PATCHY (9%).

One noteworthy result is that Lognormal mocks show a significantly larger variance than the N-body in the monopole and 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 et al. 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. 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. In addition, we found a trend of increasing deviations from the N-body with increasing bias that will be investigated in future work.

### 5.3   Correlation coefficient of the power spectrum multipoles

To show the off-diagonal elements we compute the correlation coefficient for each method as

$$r_{ij} = \frac{C_\ell(k_i, k_j)}{\sqrt{C_\ell(k_i, k_i) C_\ell(k_j, k_j)}}. \tag{13}$$
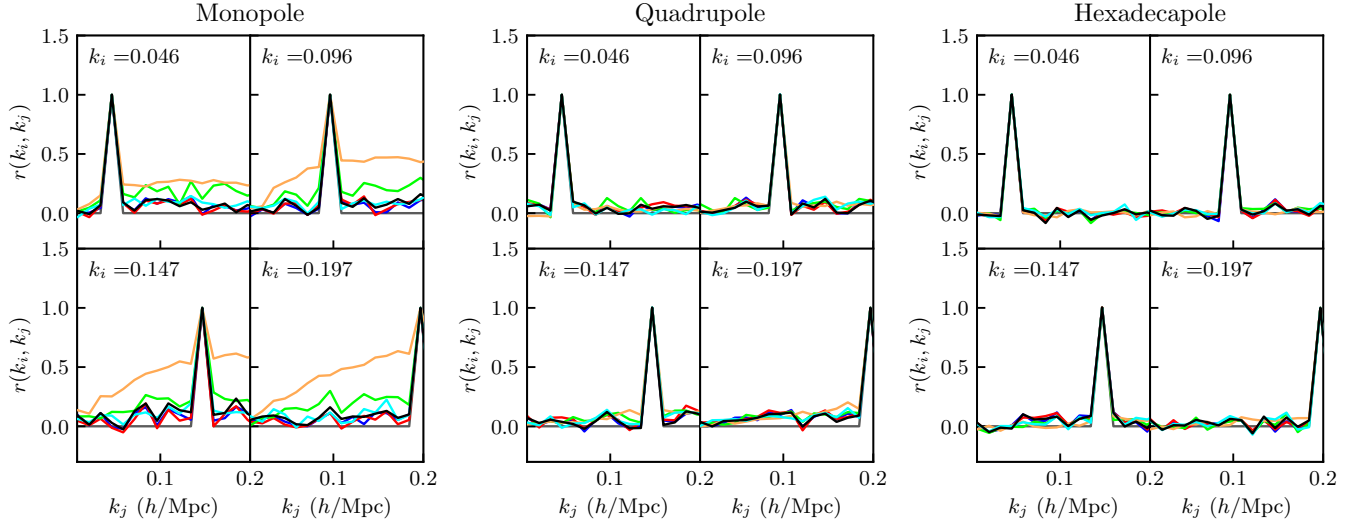
In Figures 5-6 we show a cut through the correlation coefficient r for the power spectrum multipoles for 4 different k_i values, as indicated in the plot panels. The colour coding is the same as the previous plots with the addition of the N-body in black. The first consideration to make is that the level 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 $\sim 7\%$ with a maximum of $\sim 20\%$, while for the quadrupole they are on average of the order of $\sim 4\%$ and for the hexadecapole $\sim 1\%$. The second consideration is that thanks to the matching ICs all the approximate method 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, that predicts correlations 2 to 3 times higher than the N-body, and Lognormal for the reason explained in the previous section.
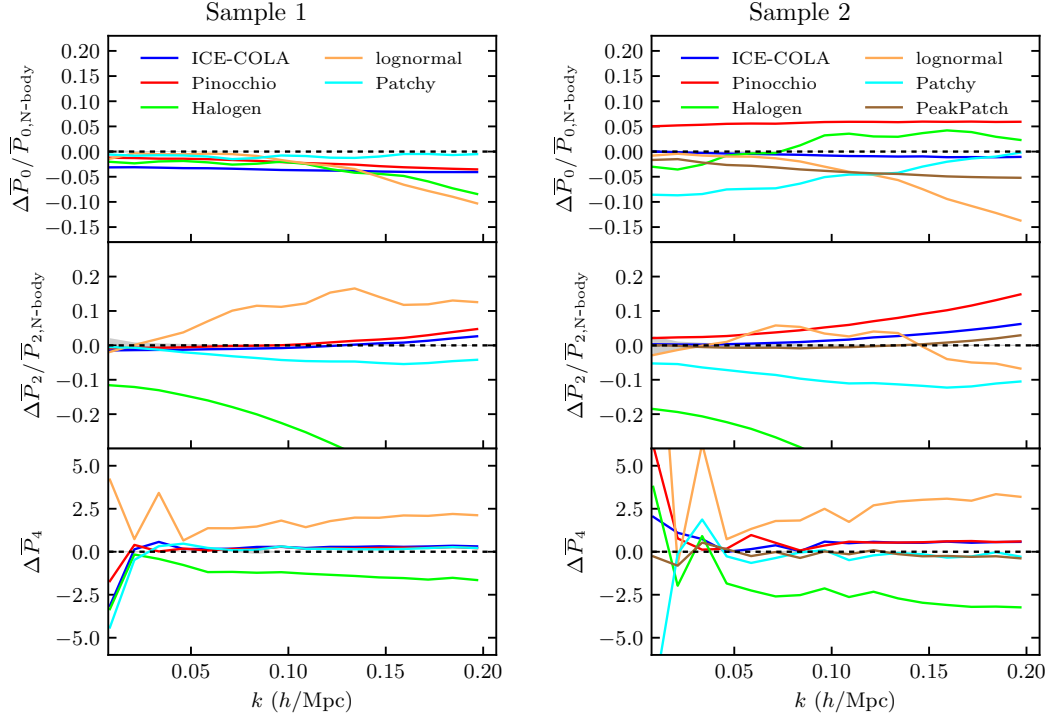
Figure 3. Relative (absolute) difference of the power spectrum monopole and quadrupole (hexadecapole) with respect to the N-body ones for sample 1 (left plot) and sample 2 (right plot). In each plot we show the monopole (top panel), quadrupole (middle panel) and hexadecapole (bottom panel). The grey shaded region shows the error on the mean of the N-body measurements.
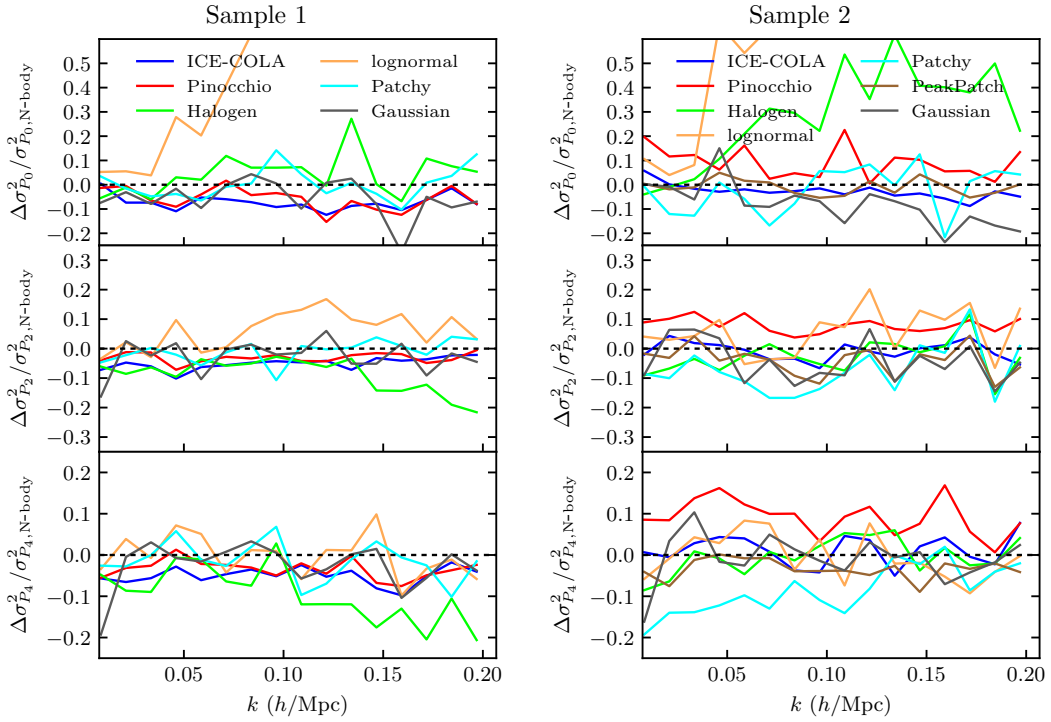
Figure 4. Relative difference of the variance of the power spectrum multipoles with respect to the N-body ones for sample 1 (top plot) and sample 2 (bottom plot). In each plot we show the monopole (top panel), quadrupole (middle panel) and hexadecapole (bottom panel). Notice that the Gaussian is a noiseless estimate so that differences are dominated by the noise in the N-body variance estimate.

Figure 5. Cut through the correlation coefficient of the monopole (left plot), quadrupole (middle plot) and hexadecapole (right 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).
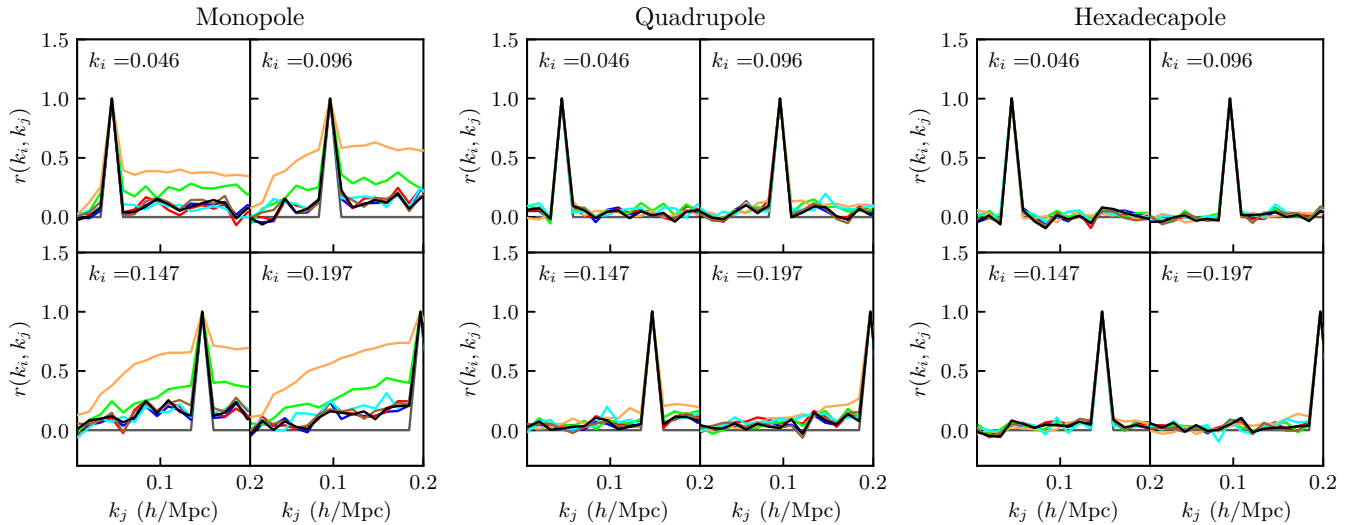
Figure 6. Cut through the correlation coefficient of the monopole (left plot), quadrupole (middle plot) and hexadecapole (right plot) for sample 2 at four different values of k as indicated in the panels. We show the results for N-body (black), ICE-COLA (blue), Pinocchio (red), PeakPatch (brown), Halogen (green), Patchy (cyan), Lognormal (orange) and Gaussian (grey).

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 Figures 7-8 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, with the exception of Lognormal, that shows clear deviations from the N-body results especially

in the bias parameters and in some cases (e.g. b_2 in Sample 2) the Gaussian approximation. The parameter values are recovered by the use of approximate mock covariances to better than 1% for all parameters except γ_3^-, which is less constrained.

To quantify differences between the recovered parameter errors we plot the ratio of the marginalised error on cosmological and nuisance parameters with respect to the N-body results in Figure 9. Before further discussion we point out that despite matching the ICs there is a resid-

ual scatter on parameter error ratios that we quantify in Appendix B to be at the level of $4\% - 5\%$. 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 ICs to the N-body simulations and thus have a different realisation 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.

In the first sample all methods except HALOGEN and Lognormal recover the error on cosmological parameters and b_1 within 5%, while for higher order bias parameters there is more variability but still within the 10% level. HALOGEN is underestimating all errors except b_1, with deviations between 5 and 10%.

In the second sample the predictive methods and PATCHY recover the N-body errors within 5% for all parameters. The Gaussian approximation show deviations $\gtrsim 5\%$ while HALOGEN is generally overestimating the errors of $\sim 10\%$ and shows a 17% overestimation of the error on $f\sigma_8$. Lognormal shows large deviations in the errors on the bias parameters but stays within the 10% level for cosmological parameters in both samples. This indicates that the large differences in the monopole covariance that we observed in the previous sections have been absorbed mainly into the nuisance parameter errors, affecting to a lower level the cosmological parameters.

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

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

where $cov(\alpha_\parallel, \alpha_\perp, f\sigma_8)$ is the parameter covariance that we recover from the full MCMC chains. We show the ratio of V to the N-body result in Figure 10. All methods except Lognormal and HALOGEN in the second sample are within 10% of the N-body result.

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", that 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 minimise the impact of sample variance errors and noise due to the limited

number of simulations, we used the same ICs 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 summarised as follows: fast PM methods, that introduces approximations in the force computation in a PM code (represented by ICE-COLA), Lagrangian methods, that identify halos in Lagrangian space and displace them to the final redshift using LPT (represented by PINOCCHIO and PEAKPATCH), bias-based methods that use LPT to generate a matter field at the redshift of interest and applies a biasing scheme to define halos (PATCHY and HALOGEN) and methods that rely on assumptions on the shape of the PDF of the density field (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 paper, spanning the algorithms above, are presented in Table 1.

We then compared performances for two halo samples: the first has (in the N-body runs) a linear bias of ~ 2.6 and a number density of $\sim 2 \times 10^{-4}$ while the second has a linear bias of ~ 3.4 and a number density of $\sim 5.5 \times 10^{-5}$. 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 halos in different ways.

Our conclusions can be summarised as follows:

(i) *Power Spectrum Multipole covariance:* The variance of the power spectrum multipoles is in general reproduced within 10% 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 show very large variance and covariance for the monopole power spectrum. Further work is needed to explain this behaviour, which translates into large over-estimation of error bars in linear bias and growth rate.

(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 fit model parameters (related to biasing, AP and growth rate of structure) and parameter errors. Although not shown we find that different covariances do not bias the recovery of model parameters, which are always within 1% of their input values.

(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 (denoted σ) is not trivial because of the finite number of mocks available (which is nonetheless 300 mocks for each method in this paper). To partially overcome this 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" ($\sigma/\sigma_{\mathrm{N-body}}$) by a factor of $\lesssim 2$ to about $4\% - 5\%$. 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 several methods recover the error on all parameters for both samples at the 5% level or better (ICE-COLA, PINOCCHIO, PEAKPATCH, PATCHY). HALOGEN and
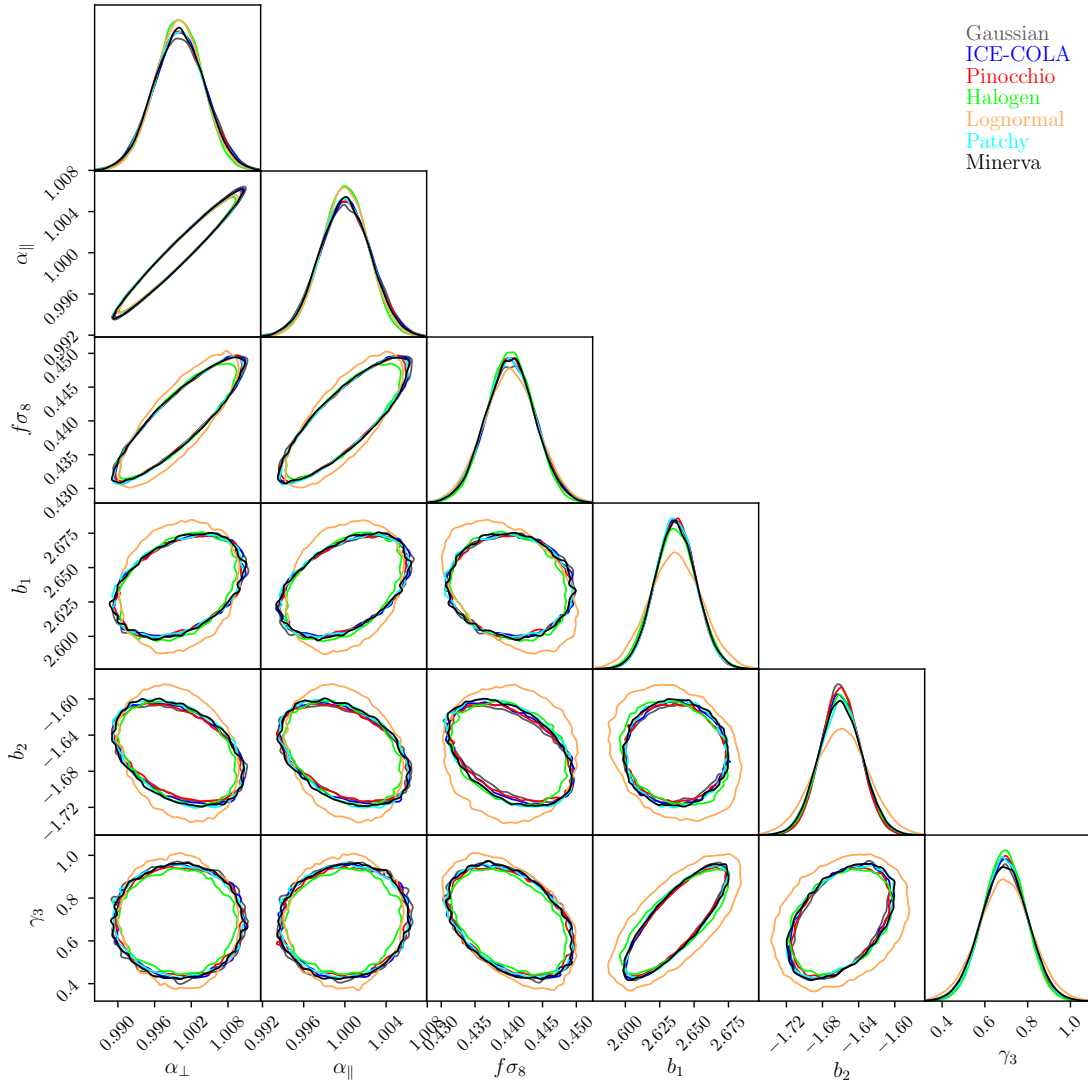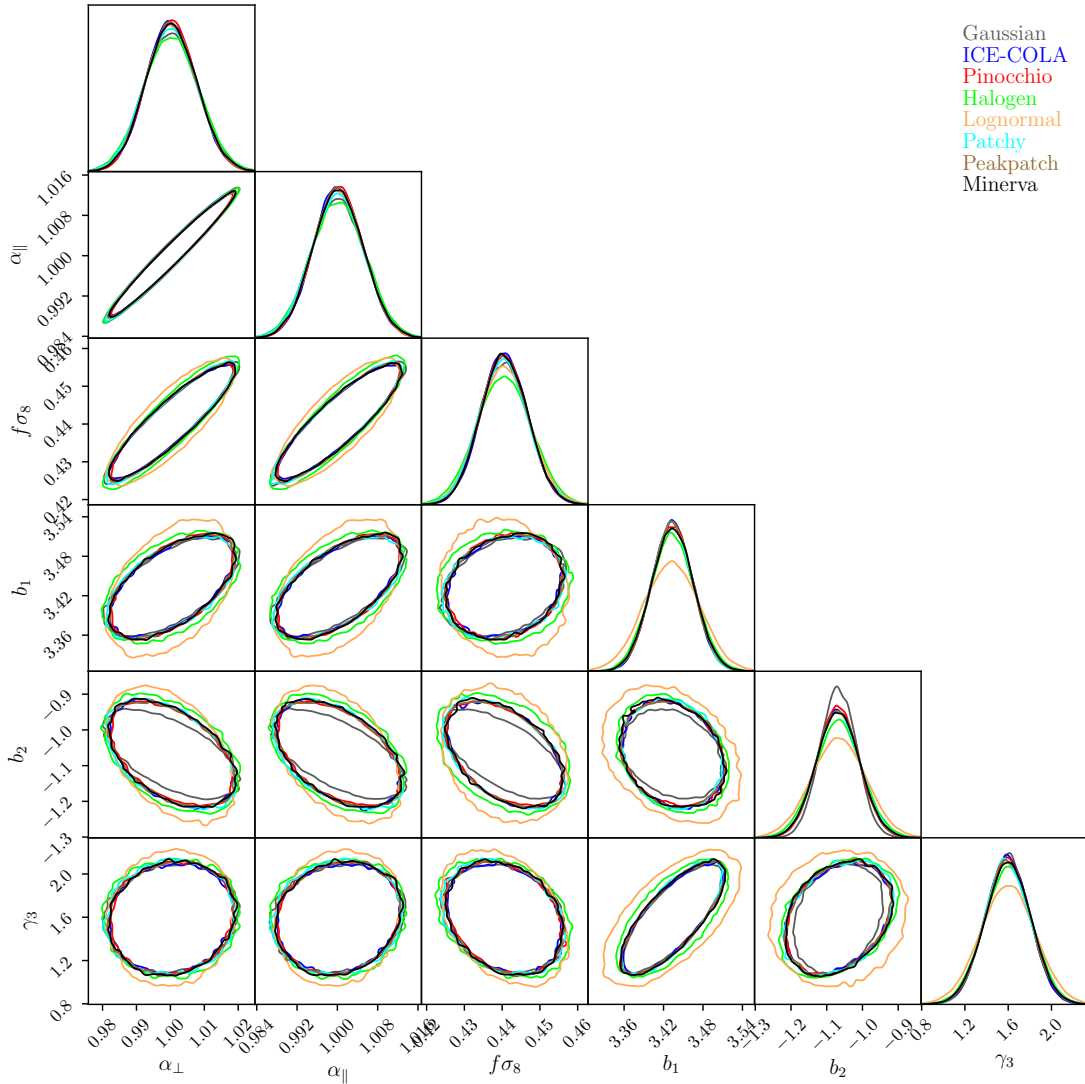
Figure 7. Marginalised 2σ contours for cosmological and nuisance parameters for sample 1.

the Gaussian analytical estimate are typically at the 10% level. Lognormal is also at the 10% level but only for cosmological parameter errors, for biasing ones differences are much larger, presumably because of the large monopole variance discussed above. 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% by all methods with the exception of HALOGEN in the second sample and Lognormal in both samples. Remarkably the Gaussian prediction (albeit using a nonlinear power spectrum) performs well, presumably because it matches the large-scale bias and the shot-noise independently. In our set up (periodic co-moving boxes) the two-point Gaussian term seems dominant against the missing contribution from a four-point connected term (tri-spectrum). Further exploration of analytical prescriptions is well motivated, e.g. to understand whether the same is true for more realistic situations (see below). In summary: if the requirement for approximate method covariances is to reproduce N-body derived parameter errors at the 5% level, then this is already achieved by several methods on the most important parameters, but not all. A looser requirement of

10% 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 (Lippich et al. 2018; Colavincenzo et al. 2018). 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 papers, and we leave it for further work.

The analysis presented in this series of papers 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 of 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 al-

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

lowed at the cost of a small systematic error on the recovered parameter uncertainty (generally within 10%, 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 are 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, that 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 comoving outputs with periodic boundary conditions. Further work will incorporate masking effects, addition of galaxies into halos (e.g. in particular halos populated using an HOD prescription for galaxies matching the H_α emitters that *Euclid* will observe) and inclusion of super-survey 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 paper, thus it will be interesting to test the performance of this method in more realistic scenarios.

ACKNOWLEDGMENTS

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 indicate 10% differences. We show the results for ICE-COLA (blue), Pinocchio (red), PeakPatch (brown), Halogen (green), Patchy (cyan), Lognormal (orange) and Gaussian (grey).
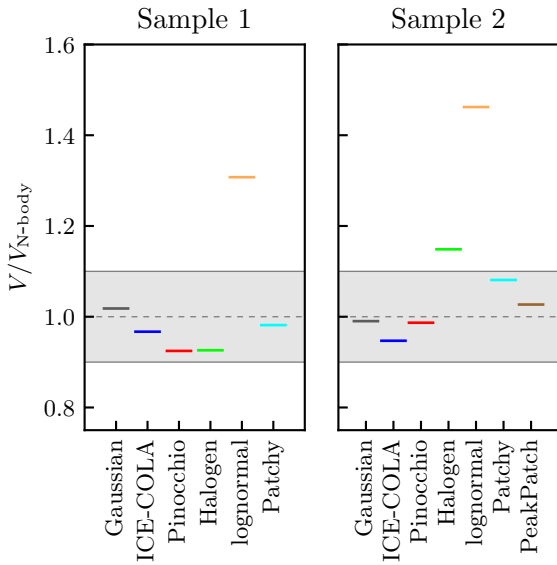
Figure 10. Ratio of the volume of the 3D ellipsoid of cosmological parameters with respect to N-body for sample 1 (left) and sample 2 (right). The grey shaded area indicate 10% differences.

REFERENCES

Agrawal A., Makiya R., Chiang C.-T., Jeong D., Saito S., Komatsu E., 2017, JCAP, 10, 003
Alam S., et al., 2017, MNRAS, 470, 2617
Alcock C., Paczynski B., 1979, Nature, 281, 358
Alvarez M., et al., 2018, in prep.
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, preprint, (arXiv:1712.06232)

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, PRD, 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., 2018, in prep.

Cole S., et al., 2005, MNRAS, 362, 505

DESI Collaboration et al., 2016, preprint, (arXiv:1611.00036)

Dawson K. S., et al., 2016, AJ, 151, 44

Dodelson S., Schneider M. D., 2013, PRD, 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

Ivezic Z., 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

LSST Science Collaboration et al., 2009, preprint, (arXiv:0912.0201)

Laureijs R., et al., 2011, ArXiv: 1110.3193,

Lippich M., et al., 2018, in prep.

Mandelbaum R., et al., 2018, PASJ, 70, S25

Manera M., et al., 2013, MNRAS, 428, 1036

Manera M., et al., 2015, MNRAS, 447, 437

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, PRD, 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

Potter D., Stadel J., Teyssier R., 2017, Computational Astrophysics and Cosmology, 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

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

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, PRD, 82, 063522

Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, JCAP, 6, 036

Taylor A. N., Joachimi B., Kitching T. D., 2013, MNRAS, 432, 1928

Tegmark M., et al., 2004, PRD, 69, 103501

Troxel M. A., et al., 2017, preprint, (arXiv:1708.01538)

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, Mon. Not. Roy. Astron. Soc., 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 halos, as done in the main body of the paper. We do this only for those methods that produce halo samples independently from the N-body ("predictive") as the remaining methods (e.g. HALOGEN or PATCHY) 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 halos for these samples, for ICE-COLA, PINOCCHIO and PEAK-PATCH

In Figures A1-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 bars the results from abundance match (shown also previously in Fig. 9) and with dots the new ones using mass thresholds.

In the case of ICE-COLA abundance match 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). In PM methods one can identify the following trends: i) halo masses are underestimated by $\sim 2\% - 3\%$, and ii) at low masses the halo catalogs are incomplete (and the size of this effect will depend for example on the PM grid-size of number of time-steps).

For ICE-COLA at high masses (Sample 2) one can simply correct halo masses by doing abundance matching (or clustering matching). This is achieved by effectively selecting more massive halos in COLA. This increases the clustering amplitude and brings 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 panel of Fig. A3. In turn, closer to the halo mass resolution of the catalogues (Sample 1) the effect is the opposite. Selecting halos 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

Method	Selection Type	M_{min} ($h^{-1}\,\mathrm{M_\odot}$)	$\bar{n}_{halos}$ ($h^3\mathrm{Mpc}^{-3}$)
	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}

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 paper.

mass scale and lowers the clustering amplitude even more (see top panel of Fig A1). Nonetheless 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 panel of Fig A.

For Pinocchio doing abundance matching always implies selecting effectively less massive halos and lowering the clustering amplitude. Since the raw clustering amplitudes are over-predicted at both Sample 1 and 2, the abundance matched helps in both cases. Overall the performance of the abundance matched selection is better in both mass regimes in terms of parameter errors, see Fig. A3.

PeakPatch shows a similar behaviour in Sample 2 but this is much more expected. Halos in PeakPatch resemble by construction a more spherical structure closer perhaps to spherical overdensity samples, it is not calibrated to reproduce a FoF mass function. Hence, in this case, mass thresholds yield lower abundance than the N-body subFind / FoF halos, and hence higher clustering. The low abundance translates into higher shot-noise, and very over-estimated error bars (see second panel of Fig A3). This is solved by abundance matching which brings σ and σ_{N-body} to within few percent.

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 paper I of this series (Lippich et al. 2018) 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 make use of a larger set of 10,000 Pinocchio realisations with the same simulation and cosmological parameters.
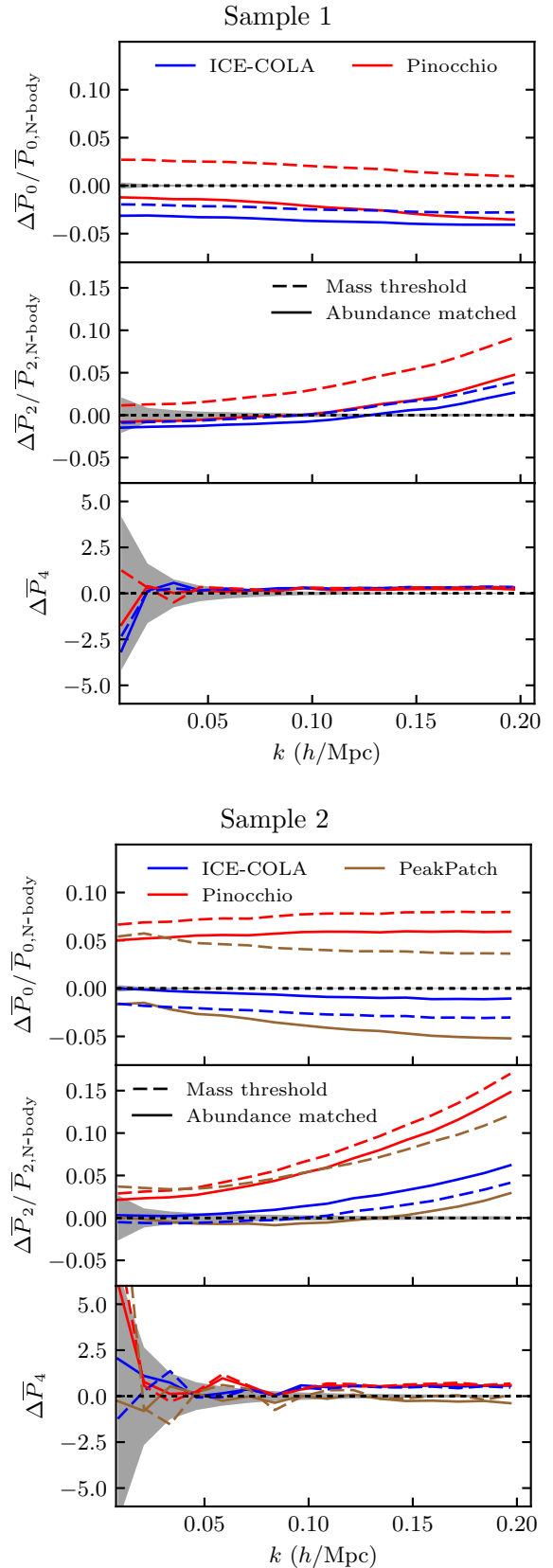
Figure A1. Relative difference of the power spectrum multipoles with respect to N-body for the sample 1 (top plot) and sample 2 (bottom plot). Each plot shows the monopole (top panel), quadrupole (middle panel) and hexadecapole (bottom panel) for mass samples (dashed lines) and density samples (solid lines).
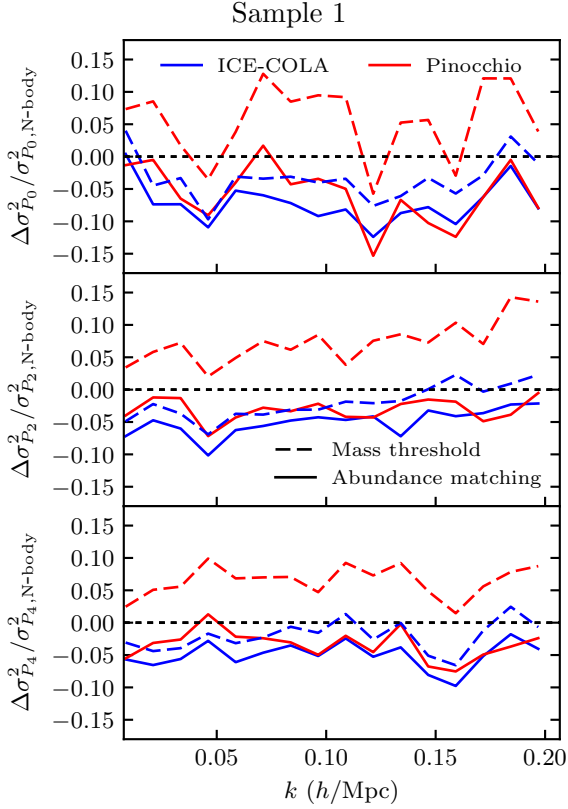
Sample 1

Sample 2

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).
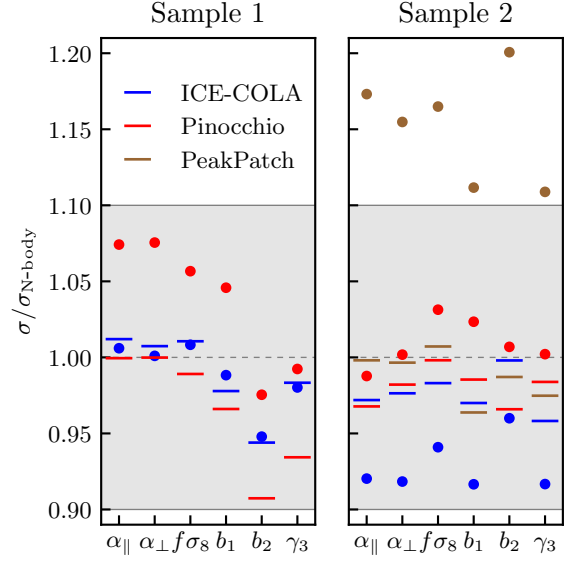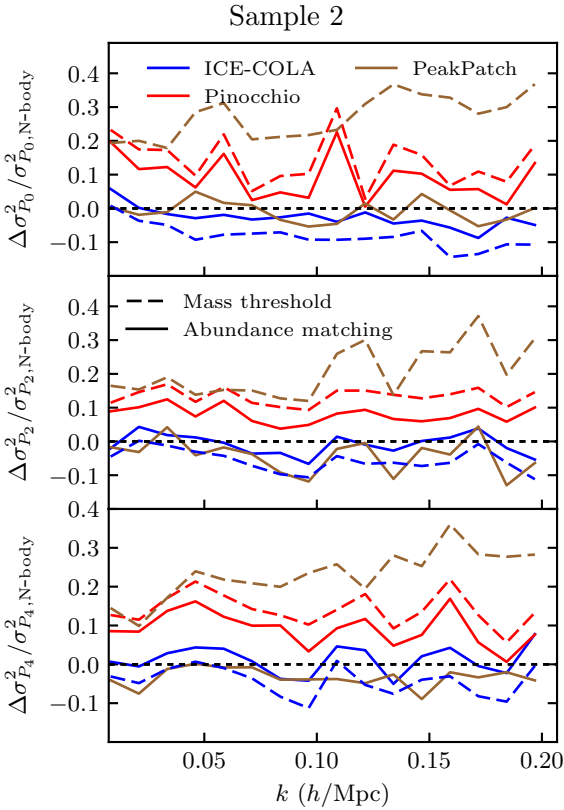
Sample 1   Sample 2

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 indicate 10% differences.

We expect our results to depend to some extent on the number of the N-body realisations we are using as benchmark for the approximate methods (Hartlap et al. 2007; Taylor et al. 2013; Dodelson & Schneider 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 realisations. 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

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

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 $P_L(k)$ to be known. The likelihood function is therefore given by

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

where $\mathrm{d}P \equiv P_{data} - P_{model}$ while C_{ij} is the covariance matrix. Such model can be rewritten as

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

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

$$-\delta P = P_{model} - P_{measured} = \sum_{\alpha=0}^{2} p_\alpha \, \mathcal{P}_\alpha \qquad \text{(B4)}$$

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

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

where

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

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 marginalisation 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 marginalised uncertainties $\sigma/\sigma_{\mathrm{N-body}}$ 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 $\sigma/\sigma_{\mathrm{N-body}}$ since the 36 sets are compared always with the same, single set of N-body realisations. Still, we can naïvely expect the actual relative error on $\sigma/\sigma_{\mathrm{N-body}}$ to be $\sqrt{2}$ larger than the one we derive in this way. The results are shown in figure B1. The blue curve shows the result for matching seeds as a function of k_{max}, top and bottom panel corresponding to the ratio of the marginalised 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 $\sigma/\sigma_{\mathrm{N-body}}$ presented in the main text. For non matching seeds the scatter across different values of k_{max} is larger, as we can expect. Nonetheless in both cases the scatter is within the overall uncertainty on the ratio which we estimate to be $\sim 4\% - 5\%$ (see below).

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 realisations. 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 $\sim 4\% - 5\%$. Remarkably, a direct application of the formulae in Taylor et al. (2013) (their Eqs. 55 & 56) results in an estimate for this error, assuming $N_S = 300$ runs and $N_D = 3 \times 16$ data-points, of $\sim 4.5\%$, 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 $7\% - 8\%$ uncertainty. Differences between solid blue and red lines in Fig. B1 are well within these values.

As an additional test we subdivide the 300 mocks in 6 sets of 50 realisations. This number should ensure a marginally reliable estimation of the power spectrum covariance matrix, at least for large values of k_{max}. In this way we can evaluate the mean and the scatter of the ratio $\sigma/\sigma_{\mathrm{N-body}}$ between Pinocchio and the N-body results
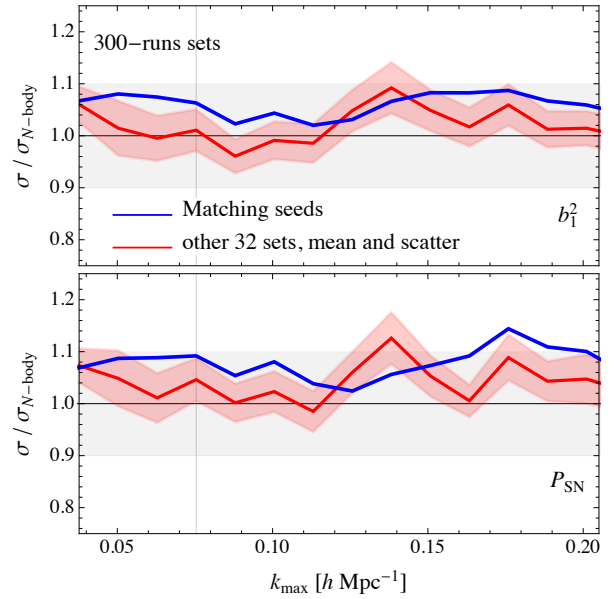
Figure B1. Comparison between the ratio of marginalised uncertainties $\sigma/\sigma_{\mathrm{N-body}}$ obtained from the 300 Pinocchio and N-body realisations with matching seeds (blue curve) and the same quantity obtained from 32 distinct sets of 300 Pinocchio realisations and the same 300 N-body runs (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.

across 6 pairs and these can be computed for six matching pairs as well as 6 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 two at large scales ($k \lesssim 0.1 \, h \, \mathrm{Mpc}^{-1}$) with respect to the case of independent runs. Clearly such reduction is less significant at smaller scales, as one can expect. We repeat the same test using six different subsets of Pinocchio realisations (not from a permutation of the six considered before). The results are shown in the lower panels of fig. B2, where the blue curve shows the same mean ratio from matching phases as before. In this case, the fact that the mean value of the independent set is closer to one is simply reflecting an overall fluctuation of the first 300 N-body runs, rather than a better performance of Pinocchio. Such fluctuation, instead, is captured by the Pinocchio realisation with matching seeds allowing to trust the few percent overestimate of the parameters uncertainties provided by the approximate method. By $k \sim 0.2 \, h \, \mathrm{Mpc}^{-1}$ the difference in the scatter is roughly a factor $\lesssim 2$.

To summarise, we estimate an uncertainty of about $7\% - 8\%$ coming from the fact that our comparisons are limited to 300 N-body and approximate methods runs which is, however, reduced to about $4\% - 5\%$ from adopting the same initial conditions in both cases.
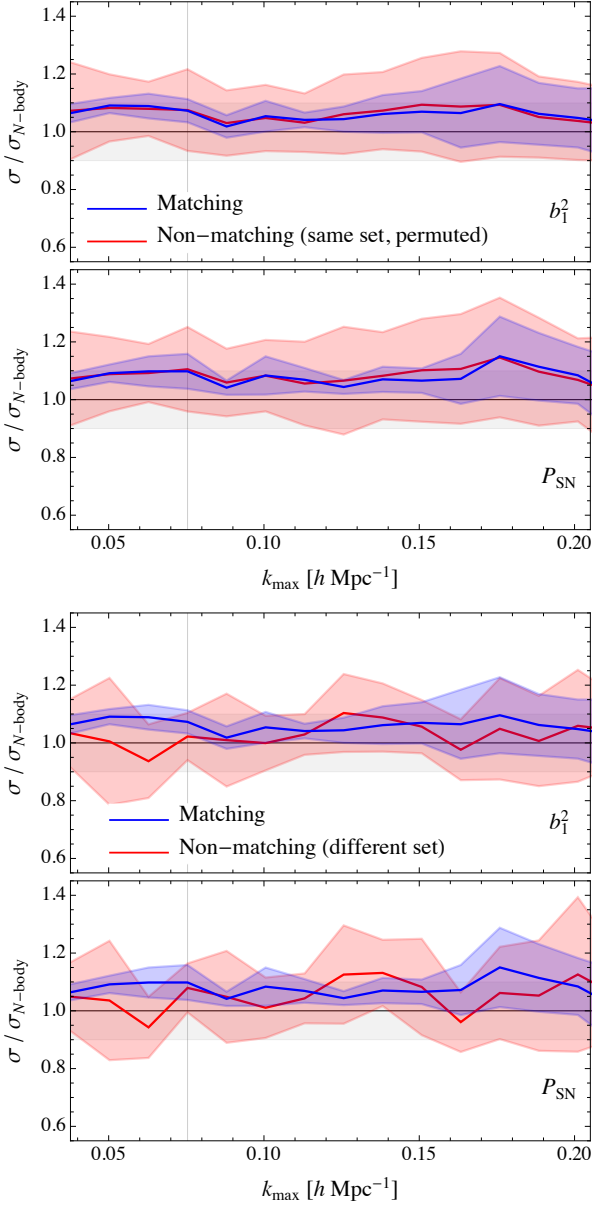
Figure B2. The top panels show the mean and standard deviation (shaded area) of the ratios of the marginalised errors on the two parameters b_1^2 and P_{SN} obtained from the PINOCCHIO covariance to the same errors obtained from the Minerva covariance. The blue line corresponds to the case of 6 sets of 50 realisations where we match the seed of the two methods. The red line to 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.