Comparing approximate methods for mock catalogues and covariance matrices – I. Correlation function

Martha Lippich,[1,2]★ Ariel G. Sánchez,[2] Manuel Colavincenzo,[3,4,5]
Emiliano Sefusatti ◉,[6,7] Pierluigi Monaco,[5,6,7] Linda Blot ◉,[8,9] Martin Crocce ◉,[8,9]
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,[8,9] Albert Izard,[8,9,17,18]
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

ABSTRACT

This paper is the first in a set that analyses the covariance matrices of clustering statistics obtained from several approximate methods for gravitational structure formation. We focus here on the covariance matrices of anisotropic two-point correlation function measurements. Our comparison includes seven approximate methods, which can be divided into three categories: predictive methods that follow the evolution of the linear density field deterministically (ICE-COLA, PEAK PATCH, and PINOCCHIO), methods that require a calibration with N-body simulations (PATCHY and HALOGEN), and simpler recipes based on assumptions regarding the shape of the probability distribution function (PDF) of density fluctuations (lognormal and Gaussian density fields). We analyse the impact of using covariance estimates obtained from these approximate methods on cosmological analyses of galaxy clustering measurements, using as a reference the covariances inferred from a set of full N-body simulations. We find that all approximate methods can accurately recover the mean parameter values inferred using the N-body covariances. The obtained parameter uncertainties typically agree with the corresponding N-body results within 5 per cent for our lower mass threshold and 10 per cent for our higher mass threshold. Furthermore, we find that the constraints for some methods can differ by up to 20 per cent depending on whether the halo samples used to define the covariance matrices are defined by matching the mass, number density, or clustering amplitude of the parent N-body samples. The results of our configuration-space analysis indicate that most approximate methods provide similar results, with no single method clearly outperforming the others.

Key words: cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

The statistical analysis of the large-scale structure (LSS) of the Universe is one of the primary tools of observational cosmology. The analysis of the signature of baryon acoustic oscillations (BAO) and redshift-space distortions (RSD) on anisotropic two-point clustering measurements can be used to infer constraints on the expansion history of the Universe (Blake & Glazebrook 2003; Linder 2003) and the redshift evolution of the growth rate of cosmic structures (Guzzo et al. 2008). Thanks to this information, LSS observations have shaped our current understanding of some of the most challenging open problems in cosmology, such as the nature of dark energy, the behaviour of gravity on large scales, and the physics of inflation (e.g. Efstathiou et al. 2002; Cole et al. 2005; Eisenstein et al. 2005; Sánchez et al. 2006, 2012; Anderson et al. 2012, 2014a,b; Alam et al. 2017).

★ E-mail: mlippich@mpe.mpg.de

Future galaxy surveys such as *Euclid* (Laureijs et al. 2011) or the Dark Energy Spectroscopic Instrument (DESI) Survey (DESI Collaboration 2016) will contain millions of galaxies covering large cosmological volumes. The small statistical uncertainties associated with clustering measurements based on these samples will push the precision of our tests of the standard Λ cold dark matter scenario even further. In this context, it is essential to identify all components of the systematic error budget affecting cosmological analyses based on these measurements, as well as to define strategies to control or mitigate them.

A key ingredient to extract cosmological information out of anisotropic clustering statistics is robust estimates of their covariance matrices. In most analyses, covariance matrices are computed from a set of mock catalogues designed to reproduce the properties of a given survey. Ideally, these mock catalogues should be based on *N*-body simulations, which can reproduce the impact of non-linear structure formation on the clustering properties of a sample with high accuracy. Due to the finite number of mock catalogues, the estimation of the covariance matrix is affected by statistical errors and the resulting noise must be propagated into the final cosmological constraints (Dodelson & Schneider 2013; Taylor, Joachimi & Kitching 2013; Percival et al. 2014; Sellentin & Heavens 2016). Reaching the level of statistical precision needed for future surveys might require the generation of several thousands of mock catalogues. As *N*-body simulations are expensive in terms of runtime and memory, the construction of a large number of mock catalogues might be infeasible. The required number of realizations can be reduced by means of methods such as resampling the phases of *N*-body simulations (Hamilton, Rimes & Scoccimarro 2006; Schneider et al. 2011), shrinkage (Pope & Szapudi 2008), calibrating the non-Gaussian contributions of an empirical model against *N*-body simulations (O'Connell et al. 2016), or covariance tapering (Paz & Sánchez 2015). However, even after applying such methods, the generation of multiple *N*-body simulations with the required number density and volume to be used for the clustering analysis of future surveys would be extremely demanding.

During the last decades, several approximate methods for gravitational structure formation and evolution have been developed, which allow for a faster generation of mock catalogues; see Monaco (2016) for a review. The accuracy with which these methods reproduce the covariance matrices estimated from *N*-body simulations must be thoroughly tested to avoid introducing systematic errors or biases on the parameter constraints derived from LSS measurements.

The nIFTy comparison project by Chuang et al. (2015) presented a detailed comparison of major approximate methods regarding their ability to reproduce clustering statistics (two-point correlation function, power spectrum and bispectrum) of halo samples drawn out of *N*-body simulations. Here, we take the comparison of different approximate methods one step further. We compare the covariance matrices inferred from halo samples obtained from different approximate methods to the corresponding ones derived from full *N*-body simulations. Furthermore, we also test the performance of the different covariance matrices at reproducing parameter constraints obtained using *N*-body simulations. We include seven approximate methods, which can be divided into three classes: predictive methods that evolve the linear density field deterministically on Lagrangian trajectories, including ICE-COLA (Tassev, Zaldarriaga & Eisenstein 2013; Izard, Crocce & Fosalba 2016), PEAK PATCH (Bond & Myers 1996), and PINOCCHIO (Monaco, Theuns & Taffoni 2002; Munari et al. 2017), methods that require higher calibration with *N*-body simulations, such as HALOGEN (Avila et al. 2015) and PATCHY (Kitaura, Yepes & Prada 2014), and two simpler

recipes based on models of the probability distribution function (PDF) of the density fluctuations, the Gaussian recipes of Grieb et al. (2016) and realizations of lognormal density fields constructed using the code of Agrawal et al. (2017). For the predictive and calibrated methods, we generate the same number of halo catalogues as the reference *N*-body simulations using identical initial conditions. We focus here on the comparison of the covariance matrices of two-point anisotropic clustering measurements in configuration space, considering Legendre multipoles (Padmanabhan & White 2008) and clustering wedges (Kazin, Sánchez & Blanton 2012). Our companion papers, Blot et al. (2018) and Colavincenzo et al. (2018), perform an analogous comparison based on power spectrum and bispectrum measurements.

The structure of the paper is as follows. Section 2 presents a brief description of the reference *N*-body simulations and the different approximate methods and recipes included in our comparison. In Section 3, we summarize the methodology used in this analysis, including a description of the halo samples that we consider (Section 3.1), our clustering measurements (Section 3.2), the estimation of the corresponding covariance matrices (Section 3.3), and the modelling for the correlation function used to assess the impact of the different methods when estimating parameter constraints (Section 3.4). We present a comparison of the clustering properties of the different halo samples in Section 4.1 and their corresponding covariance matrices in Section 4.2. In Section 4.3, we compare the performance of the different covariance matrices by analysing parameter constraints obtained from representative fits, using as a reference the ones obtained when the analysis is based on *N*-body simulations. We discuss the results from this comparison in Section 5. Finally, Section 6 presents our main conclusions.

2 APPROXIMATE METHODS FOR COVARIANCE MATRIX ESTIMATES

2.1 Methods included in the comparison

In this comparison project, we included covariance matrices inferred from different approximate methods and recipes, which we compared to the estimates obtained from a set of reference *N*-body simulations. Approximate methods have recently been revived by high-precision cosmology, due to the need of producing a large number of realizations to compute covariance matrices of clustering measurements. This topic has been reviewed by Monaco (2016), where methods have been roughly divided into two broad classes. 'Lagrangian' methods, as *N*-body simulations, are applied to a grid of particles subject to a perturbation field. They reconstruct the Lagrangian patches that collapse into dark matter (DM) haloes, and then displace them to their Eulerian positions at the output redshift, typically with Lagrangian Perturbation Theory (hereafter LPT). ICE-COLA, PEAK PATCH, and PINOCCHIO fall in this class. These methods are predictive, in the sense that, after some cosmology-independent calibration of their free parameters [that can be thought at the same level as the linking length of friends-of-friends (FoF) halo finders], they give their best reproduction of halo masses and clustering without any further tuning. This approach can be demanding in terms of computing resources and can have high memory requirements. In particular, ICE-COLA belongs to the class of Particle-Mesh codes; these are in fact *N*-body codes that converge to the true solution (at least on large scales) for sufficiently small time-steps. As such, Particle-Mesh codes are expected to be more accurate than other approximate methods, at the expense of higher computational costs.

The second class of 'bias-based' methods is based on the idea of creating a mildly non-linear density field using some version of LPT, and then populate the density field with haloes that follow a given mass function and a specified bias model. The parameters of the bias model must be calibrated on a simulation, so as to reproduce halo clustering as accurately as possible. The point of strength of these methods is their very low computational cost and memory requirement, that makes it possible to generate thousands of realizations in a simple workstation, and to push the mass limit to very low masses. This is however achieved at the cost of lower predictivity, and need of recalibration when the sample selection changes. HALOGEN and PATCHY fall in this category.

In the following, we will refer to the two classes as 'predictive' and 'calibrated' models. All approximate methods used here have been applied to the same set of 300 initial conditions (ICs) of the reference *N*-body simulations, so as to be subject to the same sample variance; as a consequence, the comparison, though limited to a relatively small number of realizations, is not affected by sample variance.

Additionally, we included in the comparison two simple recipes for the shape of the PDF of the density fluctuations, a Gaussian analytic model that is only valid in linear theory and a lognormal model. The latter was implemented by generating 1000 catalogues of 'haloes' that Poisson-sample a lognormal density field; in this test case, we do not match the ICs with the reference simulations, and use a higher number of realizations to lower sample variance.

2.2 Reference *N*-body halo catalogue: Minerva

Our reference catalogues for the comparison of the different approximate methods is derived from a set of 300 *N*-body simulations called Minerva, which were performed using GADGET-3 (last described in Springel 2005). To the first set of 100 realizations, which is described in more detail in Grieb et al. (2016) and was used in the recent BOSS analyses by Sánchez et al. (2017) and Grieb et al. (2017), 200 new independent realizations were added, which were generated with the same set-up as the first simulations. The ICs were derived from second-order Lagrangian perturbation theory (2LPT) and use the cosmological parameters that match the best-fitting results of the *WMAP* + BOSS DR9 analysis by Sánchez et al. (2013) at a starting redshift $z_{ini} = 63$. Each realization is a cubic box of side length $L_{box} = 1.5\,h^{-1}$ Gpc with 1000^3 DM particles and periodic boundary conditions. For the approximate methods described in the following sections we use the same box size and exactly the same ICs for each realization as in the Minerva simulations. Haloes were identified with a standard FoF algorithm at a snapshot of the simulations at $z = 1.0$. FoF haloes were then subject to the unbinding procedure provided by the SUBFIND code (Springel et al. 2001), where particles with positive total energy are removed and haloes that were artificially linked by FoF are separated. Given the particle mass resolution of the Minerva simulations, the minimum halo mass is $2.667 \times 10^{12}\,h^{-1}\,M_{\odot}$.

2.3 Predictive methods

2.3.1 ICE-COLA

COLA (Tassev et al. 2013) is a method to speed up *N*-body simulations by incorporating a theoretical modelling of the dynamics into the *N*-body solver and using a low-resolution numerical integration. It starts by computing the ICs using 2LPT (see Crocce, Pueblas & Scoccimarro 2006). Then, it evolves particles along their

2LPT trajectories and adds a residual displacement with respect to the 2LPT path, which is integrated numerically using the *N*-body solver. Mathematically, the displacement field x is decomposed into the LPT component x_{LPT} and the residual displacement x_{res} as

$$x_{res}(t) \equiv x(t) - x_{LPT}(t). \tag{1}$$

In a DM-only simulation, the equation of motion relates the acceleration to the Newtonian potential Φ, and omitting some constants it can be written as $\partial_t^2 x(t) = -\nabla\Phi(t)$. Using equation (1), the equation of motion reads

$$\partial_t^2 x_{res}(t) = -\nabla\Phi(t) - \partial_t^2 x_{LPT}(t). \tag{2}$$

COLA uses a Particle-Mesh method to compute the gradient of the potential at the position x (first term of the right-hand side), it subtracts the acceleration corresponding to the LPT trajectory and finally the time derivatives on the left-hand side are discretized and integrated numerically using few time-steps. The 2LPT ensures convergence of the dynamics at large scales, where its solution is exact, and the numerical integration solves the dynamics at small non-linear scales. Haloes can be correctly identified running a FoF algorithm (Davis et al. 1985) on the DM density field, and halo masses, positions, and velocities are recovered with accuracy enough to build mock halo catalogues.

ICE-COLA (Izard et al. 2016; Izard, Fosalba & Crocce 2018) is a modification of the parallel version of COLA developed in Koda et al. (2016) that produces all-sky light-cone catalogues on-the-fly. Izard et al. (2016) presented an optimal configuration for the production of accurate mock halo catalogues and Izard et al. (2018) explains the light-cone production and the modelling of weak lensing observables.

Mock halo catalogues were produced with ICE-COLA placing 30 time-steps between an initial redshift of $z_i = 19$ and $z = 0$[1] and forces were computed in a grid with a cell size three times smaller than the mean inter-particle separation distance. For the FoF algorithm, a linking length of $b = 0.2$ was used. Each simulation reached redshift 0 and used 200 cores for 20 min in the MareNostrum3 supercomputer at the Barcelona Supercomputing Center,[2] consuming a total of 20 CPU kh for the 300 realizations.

2.3.2 PEAK PATCH

From each of the 300 initial density field maps of the Minerva suite, we generate halo catalogues following the peak patch approach initially introduced by Bond & Myers (1996). In particular, we use a new massively parallel implementation of the peak patch algorithm to create efficient and accurate realizations of the positions and peculiar velocities of DM haloes (Stein, Alvarez & Bond 2018). The peak patch approach is essentially a Lagrangian space halo finder that associates haloes with the largest regions that have just collapsed by a given time. The pipeline can be separated into four subprocesses: (1) the generation of a random linear density field with the same phases and power spectrum as the Minerva simulations; (2) identification of collapsed regions using the homogeneous ellipsoidal collapse approximation; (3) exclusion and merging of the collapsed regions in Lagrangian space; and (4) assignment of displacements to these haloes using 2LPT.

The identification of collapsed regions is a key step of the algorithm. The determination of whether any given region will have

---

[1] The time-steps were linearly distributed with the scale factor.
[2] http://www.bsc.es.

collapsed or not is made by approximating it as an homogeneous ellipsoid, the fate of which is determined completely by the principal axes of the deformation tensor of the linear displacement field (i.e. the strain) averaged over the region. In principle, the process of finding these local mass peaks would involve measuring the strain at every point in space, smoothed on every scale. However, experimentation has shown that equivalent results can be obtained by measuring the strain around density peaks found on a range of scales.[3] This is done by smoothing the field on a series of logarithmically spaced scales with a top-hat kernel, from a minimum radius of $R_{f,\mathrm{min}} = 2a_{\mathrm{latt}}$, where a_{latt} is the lattice spacing, to a maximum radius of $R_{f,\mathrm{max}} = 40\,\mathrm{Mpc}$, with a ratio of 1.2. For each candidate peak, we then find the largest radius for which a homogeneous ellipsoid with the measured mean strain would collapse by the redshift of interest. If a candidate peak has no radius for which a homogeneous ellipsoid with the measured strain would have collapsed, then that point is thrown out. Each candidate point is then stored as a peak patch at its location with its radius. We then proceed down through the filter bank to all scales and repeat this procedure for each scale, resulting in a list of peak patches which we refer to as the unmerged catalogue.

The next step is to account for exclusion, an essential step to avoid double counting of matter, since distinct haloes should not overlap, by definition. We choose here to use binary exclusion (Bond & Myers 1996). Binary exclusion starts from a ranked list of candidate peak patches sorted by mass or, equivalently, Lagrangian peak patch radius. For each patch we consider every other less massive patch that overlaps it. If the smaller patch is outside of the larger one, then the radius of the two patches is reduced until they are just touching. If the centre of the smaller patch is inside the large one, then that patch is removed from the list. This process is repeated until the least massive remaining patch is reached.

Finally, we move haloes according to 2LPT using displacements computed at the scale of the halo.

This method is very fast: each realization ran typically in 97 s on 64 cores of the GPC supercomputer at the SciNet HPC Consortium in Toronto (1.72 h in total). It allows to get accurate – and fast – halo catalogues without any calibration, achieving high precision on the mass function typically for masses above a few $10^{13}\,\mathrm{M}_\odot$.

2.3.3 PINOCCHIO

The PINpointing Orbit Crossing Collapsed HIerarchical Objects (PINOCCHIO) code (Monaco et al. 2002) is based on the following algorithm.

A linear density contrast field is generated in Fourier space, in a way similar to *N*-body simulations. As a matter of fact, the code version used here implements the same loop in *k*-space as the initial condition generator (N-GENIC) used for the simulations, so the same realization is produced just by providing the code with the same random seed. The density is then smoothed using several smoothing radii. For each smoothing radius, the code computes the time at which each grid point (particle) is expected to get to the highly

non-linear regime. The dynamics of grid points, as mass elements, is treated as the collapse of a homogeneous ellipsoid, whose tidal tensor is given by the Hessian of the potential at that point. Collapse is defined as the time at which the ellipsoid collapses on the first axis, going through orbit crossing and into the highly non-linear regime; this is a difference with respect to PEAK PATCH, where the collapse of extended structures is modelled. The equations for ellipsoidal collapse are solved using third-order Lagrangian perturbation theory (3LPT). Following the ideas behind excursion-sets theory, for each particle we consider the earliest collapse time as obtained by varying the smoothing radius.

Collapsed particles are then grouped together using an algorithm that mimics the hierarchical assembly of haloes: particles are addressed in chronological order of collapse time; when a particle collapses the six nearest neighbours in the Lagrangian space are checked, if none has collapsed yet then the particle is a peak of the inverse collapse time (defined as $F = 1/D_c$, where $D_c = D(t_c)$ is the growth rate at the collapse time) and it becomes a new halo of one particle. If the collapsed particle is touching (in the Lagrangian space) a halo, then both the particle and the halo are displaced using LPT, and if they get 'near enough' the particle is accreted to the halo, otherwise it is considered as a 'filament' particle, belonging to the filamentary network of particles that have suffered orbit crossing but do not belong to haloes. If a particle touches two haloes, then their merging is decided by moving them and checking whether they get again 'near enough'. Here, 'near enough' implies a parametrization that is well explained in the original papers (see Munari et al. 2017, for the latest calibration). This results in the construction of haloes together with their merger histories, obtained with continuous time sampling. Haloes are then moved to the final position using 3LPT. The so-produced haloes have discrete masses, proportional to the particle mass M_p, as the haloes found in *N*-body simulations. To ease the procedure of number density matching described below in Section 3, halo masses were made continuous using the following procedure. It is assumed that a halo of N particles has a mass that is distributed between $N \times M_p$ and $(N + 1) \times M_p$, and the distribution is obtained by interpolating the mass function as a power law between two values computed in successive bins of width M_p. This procedure guarantees that the cumulative mass function of haloes of mass $> N \times M_p$ does not change, but it does affect the differential mass function.

We use the latest code version presented in Munari et al. (2017), where the advantage of using 3LPT is demonstrated. No further calibration was required before starting the runs. That paper presents scaling tests of the massively parallel version V4.1 and timings. The 300 runs were produced in the GALILEO@CINECA Tier-1 facility, each run required about 8 min on 48 cores.

2.4 Calibrated methods

2.4.1 HALOGEN

HALOGEN (Avila et al. 2015) is an approximate method designed to generate halo catalogues with the correct two-point correlation function as a function of mass. It constructs the catalogues following four simple steps:

(i) Generate a 2LPT DM field, and distribute their particles on a grid with cell size l_{cell}.

(ii) Draw halo masses M_h from an input halo mass function (HMF).

---

[3]This is not to say that a halo found on a given scale corresponds to a peak in the density smoothed on that scale, however, which is only the case when the strain is isotropic and the collapse is spherical. Thus, the use of density peaks as centres for strain measurements and ellipsoidal collapse calculations in the algorithm is only an optimization, to avoid wasting computations measuring the properties of regions of Lagrangian space that are unlikely to collapse in the first place.

(iii) Place the halo masses (from top to bottom) in the cells with a probability that depends on the cell density and the halo mass $P \propto \rho_{\mathrm{cell}}^{\alpha(M_{\mathrm{h}})}$. Within cells we choose random particles to assign the halo position. We further ensure mass conservation within cells and avoid halo overlap.

(iv) Assign halo velocities from the particle velocities, with a velocity bias factor: $\boldsymbol{v}_{\mathrm{halo}} = f_{\mathrm{vel}}(M_{\mathrm{h}}) \cdot \boldsymbol{v}_{\mathrm{part}}$

Following the study in Avila et al. (2015), we fix the cell size at $l_{\mathrm{cell}} = 5 \, h^{-1}$ Mpc. In this paper, we take the input HMF from the mean of the 300 Minerva simulations, but in other studies analytical HMF have been used. The parameter $\alpha(M_{\mathrm{h}})$ controls the clustering as a function of halo mass and has been calibrated using the two-point function from the Minerva simulations in logarithmic mass bins ($M_{\mathrm{h}} = 1.06 \times 10^{13}$, 2.0×10^{13}, 4.0×10^{13}, 8.0×10^{13}, $1.6 \times 10^{14} \, h^{-1} \, \mathrm{M}_{\odot}$). The factor $f(M_{\mathrm{h}})$ is also tuned to match the variance of the halo velocities from the N-body simulations.

HALOGEN is a code that advocates for the simplicity and low needs of computing resources. The fact that it does not resolve haloes (i.e. using a halo finder), allows to probe low halo masses while keeping low the computing resources. This has the disadvantage of needing to introduce free parameters. However, HALOGEN only needs one clustering parameter α and one velocity parameter f_{vel}, making the fitting procedure simple.

2.4.2 PATCHY

The PATCHY code (Kitaura et al. 2014, 2015) relies on modelling the large-scale density field with an efficient approximate gravity solver, which is populated with the halo density field using a non-linear, scale dependent, and stochastic biasing description. Although it can be applied to directly paint the galaxy distribution on the density mesh (see Kitaura et al. 2016).

The gravity solver used in this work is based on Augmented Lagrangian Perturbation Theory (ALPT; Kitaura & Heß 2013), fed with the same ICs as those implemented in the Minerva simulations. In the ALPT model, 2LPT is modified by employing a spherical collapse model on small comoving scales, splitting the displacement field into a long- and a short-range component. Better results can in principle be obtained using a Particle-Mesh gravity solver at a higher computational cost (see Vakili et al. 2017).

Once the DM density field is computed, a deterministic bias relating it to the expected number density of haloes is applied. This deterministic bias model consists of a threshold, an exponential cut-off, and a power-law bias relation. The number density is fixed by construction using the appropriate normalization of the bias expression.

The PATCHY code then associates the number of haloes in each cell by sampling from a negative binomial distribution modelling the deviation from Poissonity with an additional stochastic bias parameter.

In order to provide peculiar velocities, these are split into a coherent and a quasi-virialized component. The coherent flow is obtained from ALPT and the dispersion term is sampled from a Gaussian distribution assuming a power law with the local density.

The masses are associated with the haloes by means of the HADRON code (Zhao et al. 2015). In this approach, the masses coming from the N-body simulation are classified in different density bins and in different cosmic web types (knots, filaments, sheets, and voids) and their distribution information is extracted. Then HADRON uses this information to assign masses to haloes belonging to mock

catalogues. This information is independent of ICs, meaning it will be the same for each of the 300 Minerva realizations.

We used the MCMC python wrapper published by Vakili et al. (2017) to infer the values of the bias parameters from Minerva simulations using one of the 300 random realizations. Once these parameters are fixed one can produce all of the other mock catalogues without further fitting. The PATCHY mocks were produced using a down-sampled white noise of 500^3 instead of the 1000^3 original Minerva ones with an effective cell side resolution of $3 \, h^{-1}$ Mpc to produce the DM field.

2.5 Models of the density PDF

2.5.1 Lognormal distribution

The lognormal mocks were produced using the public code presented in Agrawal et al. (2017), which models the matter and halo density fields as lognormal fields, and generates the velocity field from the matter density field, using the linear continuity equation.

To generate a lognormal field $\delta(\boldsymbol{x})$, a Gaussian field $G(\boldsymbol{x})$ is first generated, which is related to the lognormal field as $\delta(\boldsymbol{x}) = \mathrm{e}^{-\sigma_G^2 + G(\boldsymbol{x})} - 1$ (Coles & Jones 1991). The pre-factor with the variance σ_G^2 of the Gaussian field $G(\boldsymbol{x})$ ensures that the mean of $\delta(\boldsymbol{x})$ vanishes. Because different Fourier modes of a Gaussian field are uncorrelated, the Gaussian field $G(\boldsymbol{x})$ is generated in Fourier space. The power spectrum of $G(\boldsymbol{x})$ is found by Fourier transforming its correlation function $\xi^G(r)$, which is related to the correlation function $\xi(r)$ of the lognormal field $\delta(\boldsymbol{x})$ as $\xi^G(r) = \ln[1 + \xi(r)]$ (Coles & Jones 1991). Having generated the Gaussian field $G(\boldsymbol{x})$, the code transforms it to the lognormal field $\delta(\boldsymbol{x})$ using the variance σ_G^2 measured from $G(\boldsymbol{x})$ in all cells.

In practice, we use the measured real-space matter power spectrum from Minerva and Fourier transform it to get the matter correlation function. For haloes we use the measured real-space correlation function. We then generate the Gaussian matter and halo fields with the same phases, so that the Gaussian fields are perfectly correlated with each other. Note however, that we use random realizations for these mocks, and so, these phases are not equal to those of the Minerva ICs. We then exponentiate the Gaussian fields, to get matter ($\delta_m(\boldsymbol{x})$) and halo ($\delta_g(\boldsymbol{x})$) density fields, following a lognormal distribution.

The expected number of haloes in a cell is given as $N_g(\boldsymbol{x}) = \bar{n}_g[1 + \delta_g(\boldsymbol{x})]V_{\mathrm{cell}}$, where $\bar{n}_g$ is the mean number density of the halo sample from Minerva, $\delta_g(\boldsymbol{x})$ is the halo density at position $\boldsymbol{x}$, and V_{cell} is the volume of the cell. However, this is not an integer. So, to obtain an integer number of haloes from the halo density field, we draw a random number from a Poisson distribution with mean $N_g(\boldsymbol{x})$, and populate haloes randomly within the cell. The lognormal matter field is then used to generate the velocity field using the linear continuity equation. Each halo in a cell is assigned the three-dimensional velocity of that cell.

Since the lognormal mocks use random phases, we generate 1000 realizations for each mass bin, with the real-space clustering and mean number density measured from Minerva as inputs. Also note that because haloes in this prescription correspond to just discrete points, we do not assign any mass to them. An effective bias relation can still be established using the cross-correlation between the halo and matter fields, or using the input clustering statistics (Agrawal et al. 2017).

The key advantage of using this method is its speed. Once we had the target power spectrum of the matter and halo Gaussian fields, each realization of a 256^3 grid, as in Minerva, was produced in 20 s

using 16 cores at the RZG in Garching. The resulting catalogues agree perfectly with the Minerva realizations in their real-space clustering as expected. Because we use linear velocities, they also agree with the redshift-space predictions on large scales (Agrawal et al. 2017).

2.5.2 *Gaussian distribution*

A different approach to generating 'mock' halo catalogues with fast approximate methods is to model the covariance matrix theoretically. This has the advantage that the resulting estimate is free of noise. In this comparison project, we included a simple theoretical model for the linear covariance of anisotropic galaxy clustering that is described in Grieb et al. (2016). Based on the assumption that the two-dimensional power spectrum $P(k, \mu)$ follows a Gaussian distribution and that the contributions from the trispectrum and super-sample covariance can be neglected, Grieb et al. (2016) derived the explicit formulae for the covariance of anisotropic clustering measurements in configuration and Fourier space. In particular, they obtain that the covariance between two Legendre multipoles of the correlation function of order ℓ and ℓ' (see Section 3.2) evaluated at the pair separations s_i and s_j, respectively, is given by

$$C_{\ell,\ell'}(s_i, s_j) = \frac{i^{\ell+\ell'}}{2\pi^2} \int_0^\infty k^2 \sigma_{\ell\ell'}^2(k) \bar{j}_\ell(ks_i) \bar{j}_{\ell'}(ks_j) \, \mathrm{d}k, \tag{3}$$

where $\bar{j}_\ell(ks_i)$ is the bin-averaged spherical Bessel function as defined in equation (A19) of Grieb et al. (2016), and

$$\sigma_{\ell\ell'}^2(k) \equiv \frac{(2\ell+1)(2\ell'+1)}{V_s} \tag{4}$$

$$\times \int_{-1}^1 \left[P(k, \mu) + \frac{1}{\bar{n}} \right]^2 L_\ell(\mu) L_{\ell'}(\mu) \, \mathrm{d}\mu.$$

Here, $P(k, \mu)$ represents the two-dimensional power spectrum of the sample, V_s is its volume, and $\bar{n}$ corresponds to its mean number density.

Analogously, the covariance between two configuration-space clustering wedges μ and μ' (see Section 3.2) is given by

$$C_{\mu,\mu'}(s_i, s_j) = \sum_{\ell_1\ell_2} \frac{i^{\ell_1+\ell_2}}{2\pi^2} \bar{L}_{\ell_1,\mu} \bar{L}_{\ell_2,\mu'}$$

$$\times \int_0^\infty k^2 \sigma_{\ell_1\ell_2}^2(k) j_{\ell_1}(ks_i) j_{\ell_2}(ks_j) \, \mathrm{d}k, \tag{5}$$

where $\bar{L}_{\ell_1,\mu}$ represents the average of the Legendre polynomial of order ℓ within the corresponding μ-range of the clustering wedge. The covariance matrices derived from the Gaussian model have been tested against N-body simulations with periodic boundary conditions by Grieb et al. (2016), showing good agreement within the range of scales typically included in the analysis of galaxy redshift surveys ($s > 20 \, h^{-1}$ Mpc).

3 METHODOLOGY

3.1 Halo samples

In this section, we describe the criteria used to construct the halo samples on which we base our covariance matrix comparison.

We define two parent halo samples from the Minerva simulations by selecting haloes with masses $M \geq 1.12 \times 10^{13} \, h^{-1} \, \mathrm{M}_\odot$ and $M \geq 2.667 \times 10^{13} \, h^{-1} \, \mathrm{M}_\odot$, corresponding to 42 and 100 DM

particles, respectively. We apply the same mass cuts to the catalogues produced by the approximated methods included in our comparison. We refer to the resulting samples as 'mass1' and 'mass2'.

Note that the PATCHY and lognormal catalogues do not have mass information for individual objects and match the number density and bias of the parent samples from Minerva by construction. The Gaussian model predictions are also computed for the specific clustering amplitude and number density as the mass1 and mass2 samples. For the other approximate methods, the samples obtained by applying these mass thresholds do not reproduce the clustering and the shot noise of the corresponding samples from Minerva. These differences are in part caused by the different applied methods for identifying or assigning haloes, e.g. PEAK PATCH uses spherical overdensities in Lagrangian space to define halo masses while most other methods are closer to FoF masses, as described in Section 2. Therefore, for the ICE-COLA, HALOGEN, PEAK PATCH, and PINOCCHIO catalogues we also define samples by matching number density and clustering amplitude of the halo samples from Minerva. For the number-density-matched samples, we find the mass cuts where the total number of haloes in the samples drawn from each approximate method best matches that of the two parent Minerva samples. We refer to these samples as 'dens1' and 'dens2'. Analogously, we define bias-matched samples by identifying the mass thresholds for which the clustering amplitude in the catalogues drawn from the approximate methods best agrees with that of the mass1 and mass2 samples from Minerva. More concretely, we define the clustering-amplitude-matched samples by selecting the mass thresholds that minimize the difference between the mean correlation function measurements from the catalogues drawn from the approximate methods and the Minerva parent samples on scales $40 \, h^{-1}$ Mpc $< s < 80 \, h^{-1}$ Mpc. We refer to these samples as 'bias1' and 'bias2'.

The mass thresholds defining the different samples, the number of particles corresponding to these limits, their halo number densities, and bias ratios with respect to the Minerva parent samples are listed in Table 1. Note that, as the halo masses of the PINOCCHIO and PEAK PATCH catalogues are made continuous for this analysis, the mass cuts defining the density- and bias-matched samples do not correspond to an integer number of particles. Also note for the calibrated methods that the HALOGEN catalogue was calibrated using the input HMF from the mean of the 300 Minerva simulations in logarithmic mass bins for this analysis, whereas the PATCHY mass samples were calibrated for each mass cut individually. For the case of the HALOGEN catalogue, the selected high-mass threshold lies nearly half way (in logarithmic scale) between two of the mass thresholds of the logarithmic input HMF. This explains why whereas for the first mass cut, bias and number density are matched by construction, that is not the case for the second mass cut. This has the effect that the bias2 sample of the HALOGEN catalogue has 15 per cent fewer haloes than the corresponding Minerva sample. Comparisons of the ratios of the number densities and bias of the different samples drawn from the approximate methods to the corresponding ones from Minerva are shown in Fig. 1. Since the catalogues drawn from lognormal and PATCHY match the number density and bias of the Minerva parent samples by construction, they are not included in the tables and figures.

In the following, we refer to all samples corresponding to the first mass limit, mass1, dens1, and bias1 as 'sample1', and the samples corresponding to the second mass limit, mass2, dens2, and bias2 as 'sample2'.

Table 1. Overview of the different samples, including the mass limits, $M_{\rm lim}$, expressed in units of $h^{-1}\,{\rm M_\odot}$, the corresponding number of particles, $N_{\rm p}$, the mean number density, $\bar{n}$, and the bias ratio to the corresponding Minerva parent sample, $\langle(\xi_{\rm app}/\xi_{\rm Min})^{1/2}\rangle$. The sample names 'mass', 'dens', and 'bias', indicate if the samples were constructed by matching the mass threshold, number density, or clustering amplitude of the parent halo samples from Minerva.

Code	Sample name	$M_{\rm lim}\,\left(h^{-1}\,{\rm M_\odot}\right)$	$N_{\rm p}$	$\bar{n}\,\left(h^3\,{\rm Mpc}^{-3}\right)$	Bias ratio
Minerva	mass1	1.12×10^{13}	42	2.12×10^{-4}	1.00
Minerva	mass2	2.67×10^{13}	100	5.42×10^{-5}	1.00
ICE-COLA	mass1	1.12×10^{13}	42	2.06×10^{-4}	0.99
ICE-COLA	dens1	1.09×10^{13}	41	2.12×10^{-4}	0.98
ICE-COLA	bias1	1.17×10^{13}	44	1.93×10^{-4}	1.00
ICE-COLA	mass2	2.67×10^{13}	100	5.81×10^{-5}	0.99
ICE-COLA	dens2, bias2	2.77×10^{13}	104	5.45×10^{-5}	1.00
HALOGEN	mass1, dens1, bias1	1.12×10^{13}	42	2.14×10^{-4}	1.00
HALOGEN	mass2, dens2	2.67×10^{13}	100	5.40×10^{-5}	0.98
HALOGEN	bias2	2.91×10^{13}	109	4.61×10^{-5}	1.00
PEAK PATCH[a]	mass2	2.67×10^{13}	100	4.45×10^{-5}	1.04
PEAK PATCH	dens2, bias2	2.35×10^{13}	88.3	5.44×10^{-5}	1.00
PINOCCHIO	mass1	1.12×10^{13}	42	1.95×10^{-4}	1.02
PINOCCHIO	dens1	1.04×10^{13}	39.1	2.15×10^{-4}	1.00
PINOCCHIO	bias1	1.06×10^{13}	39.9	2.09×10^{-4}	1.00
PINOCCHIO	mass2	2.67×10^{13}	100	5.35×10^{-5}	1.03
PINOCCHIO	dens2	2.63×10^{13}	98.6	5.48×10^{-5}	1.03
PINOCCHIO	bias2	2.42×10^{13}	90.7	6.27×10^{-5}	1.00

Note. [a]As the halo masses corresponding to our low-mass threshold are not correctly resolved in the PEAK PATCH catalogues, only the high-mass threshold (mass2) is considered in this case.
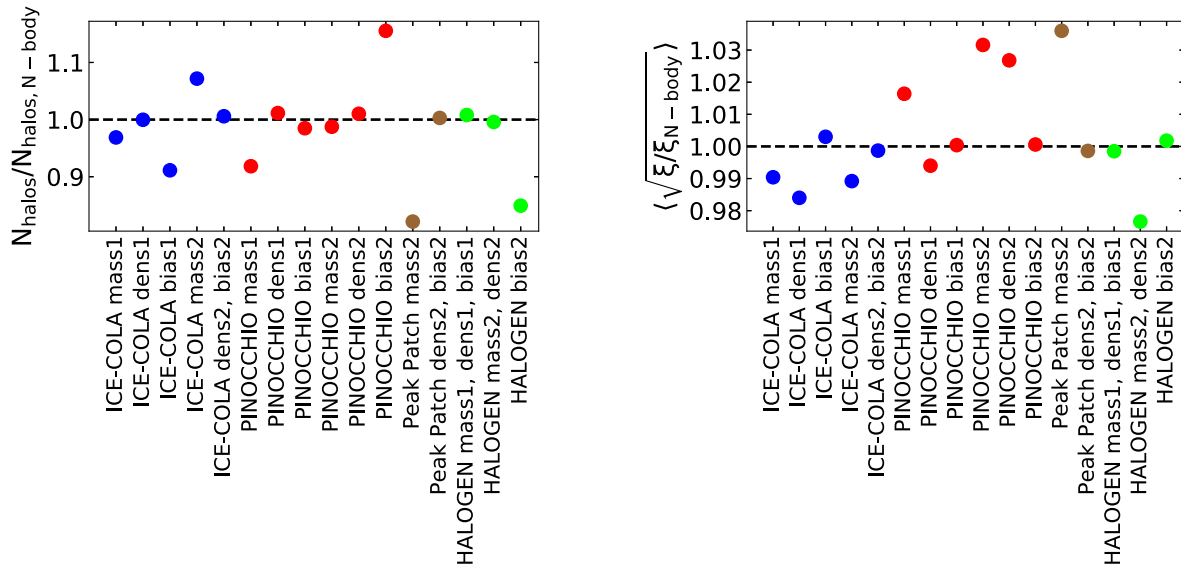
Figure 1. Ratios of the total halo number (left-hand panel) and the clustering amplitude (right-hand panel) of samples drawn from the approximate methods to the corresponding quantity in the Minerva parent samples. By definition, for the dens samples the halo number is matched to the corresponding *N*-body samples and therefore the corresponding ratio is close to one in the left-hand panel, while the ratios from the bias samples are meant to be close to one in the right-hand panel. For some cases, two or three samples are represented with the same symbol, e.g. ICE-COLA dens2, bias2 which means that the ICE-COLA dens2 sample is the same as the ICE-COLA bias2 sample.

3.2 Clustering measurements in configuration space

Most cosmological analyses of galaxy redshift surveys are based on two-point clustering statistics. In this paper, we focus on configuration-space analyses and study the estimation of the covariance matrix of correlation function measurements. The information of the full two-dimensional correlation function, $\xi(s, \mu)$, where μ is the cosine of the angle between the separation vector s and the

line of sight, can be compressed into a small number of functions such as the Legendre multipoles, $\xi_\ell(s)$, given by

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^{1} L_\ell(\mu)\xi(\mu, s)\,{\rm d}\mu, \tag{6}$$

where $L_\ell(\mu)$ denotes the Legendre polynomial of order ℓ. Typically, only multipoles with $\ell \leq 4$ are considered. An alternative tool is the

clustering wedges statistic (Kazin et al. 2012), which corresponds to the average of the full two-dimensional correlation function over wide bins in μ, that is

$$\xi_{w,i}(s) = \frac{1}{\Delta \mu} \int_{(i-1)/n}^{i/n} \xi(\mu, s)\,\mathrm{d}\mu, \tag{7}$$

where $\xi_{w,i}$ denotes each individual clustering wedge and n represents the total number of wedges. We follow the recent analysis of Sánchez et al. (2017) and divide the μ range from 0 to 1 into three equal-width intervals, $i = 1, 2, 3$.

We compute the Legendre multipoles and clustering wedges of the halo samples defined in Section 3.1. As these measurements are based on simulation boxes with periodic boundary conditions, the full $\xi(s, \mu)$ can be computed using the natural estimator, namely

$$\xi(s, \mu) = \frac{DD(s, \mu)}{RR(s, \mu)} - 1, \tag{8}$$

where $DD(s, \mu)$ are the normalized data pair counts and $RR(s, \mu)$ the normalized random pair counts, which can be computed as the ratio of the volume of a shell $\mathrm{d}V$ and the total box volume V_s, $RR = \mathrm{d}V/V_s$. The obtained $\xi(s, \mu)$ can be used to estimate Legendre multipoles and clustering wedges using equations (6) and (7), respectively. We consider scales in the range $20\,h^{-1}\,\mathrm{Mpc} \leq s \leq 160\,h^{-1}\,\mathrm{Mpc}$ for all our measurements and implement a binning scheme with $\mathrm{d}s = 10\,h^{-1}\,\mathrm{Mpc}$ for the following analysis. For illustration purposes, we also use a binning of $\mathrm{d}s = 5\,h^{-1}\,\mathrm{Mpc}$ for the figures showing correlation function measurements. Considering Legendre multipoles with $\ell \leq 4$ and three μ wedges, the dimension of the total data vector, $\boldsymbol{\xi}$, containing all the measured statistics is the same in both cases ($N_b = 42$ and $N_b = 84$ for the cases of $\mathrm{d}s = 10\,h^{-1}\,\mathrm{Mpc}$ and $\mathrm{d}s = 5\,h^{-1}\,\mathrm{Mpc}$, respectively).

3.3 Covariance matrix estimation

It is commonly assumed that the likelihood function of the measured two-point correlation function is Gaussian in form,

$$-2\ln\mathcal{L}(\boldsymbol{\xi}|\boldsymbol{\theta}) = (\boldsymbol{\xi} - \boldsymbol{\xi}_{\mathbf{theo}}(\boldsymbol{\theta}))^t \boldsymbol{\Psi}(\boldsymbol{\xi} - \boldsymbol{\xi}_{\mathbf{theo}}(\boldsymbol{\theta})), \tag{9}$$

where $\boldsymbol{\xi}_{\mathbf{theo}}$ represents the theoretical model of the measured statistics, which here correspond to the Legendre multipoles or clustering wedges, for the parameters $\boldsymbol{\theta}$, and $\boldsymbol{\Psi}$ is the precision matrix, given by the inverse of the covariance matrix, $\boldsymbol{\Psi} = \mathbf{C}^{-1}$.

The covariance matrix, $\mathbf{C}$ is usually estimated from a large set of N_s mock catalogues as

$$C_{ij} = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} (\xi_i^k - \bar{\xi}_i)(\xi_j^k - \bar{\xi}_j), \tag{10}$$

where $\bar{\xi}_i = \frac{1}{N_s}\sum_k \xi_i^k$ is the mean value of the measurements at the ith bin and ξ_i^k is the corresponding measurement from the kth mock. This estimator has the advantage over other techniques such as jackknife estimates from the data or theoretical modelling, that it tends to be less affected by biases than estimates from the data and does not require any assumptions regarding the properties of the true covariance matrix. However, the noise in $\mathbf{C}$ due to the finite number of realizations leads to an additional uncertainty, which must be propagated into the final parameter constraints (Dodelson & Schneider 2013; Taylor et al. 2013; Percival et al. 2014; Sellentin & Heavens 2016). Depending on the analysis configuration, the control of this additional error might require a large number of realizations, with N_s in the range of a few thousands. For the new generation of large-volume surveys such as *Euclid*, the construction of a large

number of mock catalogues might be extremely demanding and will need to rely, at least partially, on approximate N-body methods. The goal of our analysis is to test the impact on the obtained parameter constraints of using estimates of $\mathbf{C}$ based on different approximate methods.

We use equation (10) to compute the covariance matrices associated with the measurements of the multipoles and clustering wedges of the halo samples defined in Section 3.1. In order to reduce the noise in these measurements due to the limited number of realizations, we obtain three separate estimates of $\mathbf{C}$ from each sample by treating each axis of the simulation boxes as the line-of-sight direction when computing $\xi(s, \mu)$. Our final estimates correspond to the average of the covariance matrices measured on the different lines of sight. The Gaussian theoretical covariance matrices were computed for the specific number density and clustering of the halo samples from Minerva. We used as input the model of the two-dimensional power spectrum described in Section 3.4, whose parameters were fitted to reproduce the clustering of parent halo samples.

3.4 Testing the impact of approximate methods for covariance matrix estimates

The cosmological information recovered from full-shape fits to anisotropic clustering measurements is often expressed in terms of the BAO shift parameters

$$\alpha_\perp = \frac{D_A(z)\,r_d'}{D_A'(z)\,r_d}, \tag{11}$$

$$\alpha_\parallel = \frac{H'(z)\,r_d'}{H(z)\,r_d}, \tag{12}$$

where $H(z)$ is the Hubble parameter at redshift z, $D_A(z)$ is the corresponding angular diameter distance, r_d is the sound horizon at the drag redshift, and the primes denote quantities in the fiducial cosmology; and the RSD parameter combination $f\sigma_8(z)$, where $f(z)$ represents the logarithmic growth rate of density fluctuations and $\sigma_8(z)$ is the linear rms mass fluctuation in spheres of radius $8\,h^{-1}\,\mathrm{Mpc}$.

The constraints on these parameters are sensitive to details in the definition of the likelihood function, such as the way in which the covariance matrix of the measurements is estimated. In order to assess the impact of using approximate methods to estimate $\mathbf{C}$, we perform full-shape fits of anisotropic clustering measurements in configuration space to obtain constraints on $\alpha_\perp$, $\alpha_\parallel$, and $f\sigma_8(z)$ assuming the Gaussian likelihood function of equation (9). We compare the constraints obtained when $\mathbf{C}$ is estimated from a set of full N-body simulations with the results inferred from the same set of measurements when the covariance matrix is computed using the approximate methods described in Section 2.

Our fits are based on the same model of the full two-dimensional correlation function $\xi(\mu, s)$ as in the analyses of the final BOSS galaxy samples (Grieb et al. 2017; Salazar-Albornoz et al. 2017; Sánchez et al. 2017) and the eBOSS DR12 catalogue (Hou et al. 2018). This model includes the effects of the non-linear evolution of density fluctuations based on gRPT (Crocce, Scoccimarro & Blas, in preparation) bias (Chan & Scoccimarro 2012), and RSD (Scoccimarro, in preparation). The only difference between the model implemented in these studies and the one used here is that, since we analyse halo samples instead of galaxies, we do not include the so-called fingers-of-God factor, $W_\infty(k, \mu)$ (see equation 18 in Sánchez et al. 2017). In total, our parameter space contains six free parameters, the BAO and RSD parameters $\alpha_\parallel$, $\alpha_\perp$, and $f\sigma_8$, and the
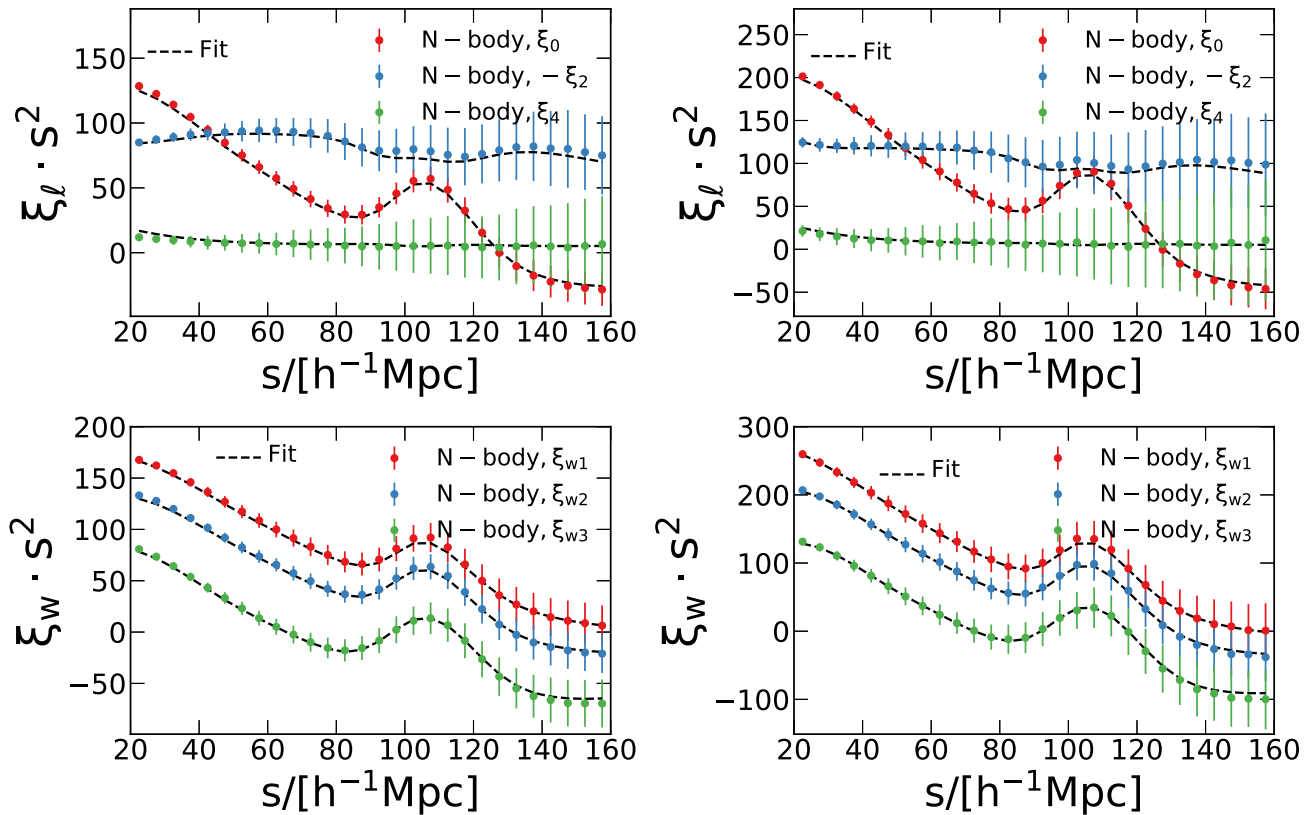
Figure 2. Comparison of the mean correlation function multipoles (upper panels) and clustering wedges (lower panels) of the mass1 and mass2 samples (left-hand and right-hand panels, respectively) drawn from our Minerva *N*-body simulations, and the model described in Section 3.4. The points with error bars show to the simulation results and the dashed lines correspond to the fit to these measurements. The error bars on the measurements correspond to the dispersion inferred from the 300 Minerva realizations. In all cases, the model predictions show good agreement with the *N*-body measurements.

nuisance parameters associated with the linear and quadratic local bias, b_1 and b_2, and the non-local bias γ_3^-. We explore this parameter space by means of the Monte Carlo Markov Chain (MCMC) technique. This analysis set-up matches that of the covariance matrix comparison in Fourier space of our companion paper Blot et al. (2018).

In order to ensure that the model used for the fits has no impact on the covariance matrix comparison, we do not fit the measurements of the Legendre multipoles and wedges obtained from the *N*-body simulations. Instead, we use our baseline model to construct synthetic clustering measurements, which we then use for our fits. For this, we first fit the mean Legendre multipoles measured from the parent Minerva halo samples using our model and the *N*-body covariance matrices. We fix all cosmological parameters to their true values and only vary the bias parameters b_1, b_2, and γ_3^-. We then use the mean values of the parameters inferred from the fits, together with the true values of the cosmological parameters, to generate multipoles and clustering wedges of the correlation function using our baseline model. Fig. 2 shows the mean multipoles and clustering wedges measured from the Minerva halo sample for both mass cuts and the resulting fits. In all cases, our model gives a good description of the simulation results. The parameter values recovered from these fits were also used to compute the input power spectra when computing the Gaussian predictions of **C**. As these synthetic data are perfectly described by our baseline model by construction, their analysis should recover the true values of the BAO parame-

ters $\alpha_\parallel = \alpha_\perp = 1.0$, and the growth-rate parameter $f\sigma_8 = 0.4402$. The comparison of the parameter values and their uncertainties recovered using different covariance matrices allows us to test the ability of the approximate methods described in Section 2 to reproduce the results obtained when **C** is inferred from full *N*-body simulations.

4 RESULTS

In this section, we present a detailed comparison of the covariance matrix measurements in configuration space obtained from the approximate methods described in Section 2 and their performance at recovering the correct parameter estimates.

4.1 Two-point correlation function measurements

In order to estimate the covariance matrices from all the samples introduced in Section 3.1, we first measure configuration-space Legendre multipoles and clustering wedges for each sample and in each realization as described in Section 3.2.

As an illustration of the agreement between the clustering measurements obtained from the approximate methods and the Minerva simulations we focus here on two cases: (i) the multipoles of the density-matched samples for the first mass cut (dens1 samples) and (ii) the clustering wedges of the bias-matched samples for the second mass cut (bias2 samples). As described in Section 3.1, for PATCHY
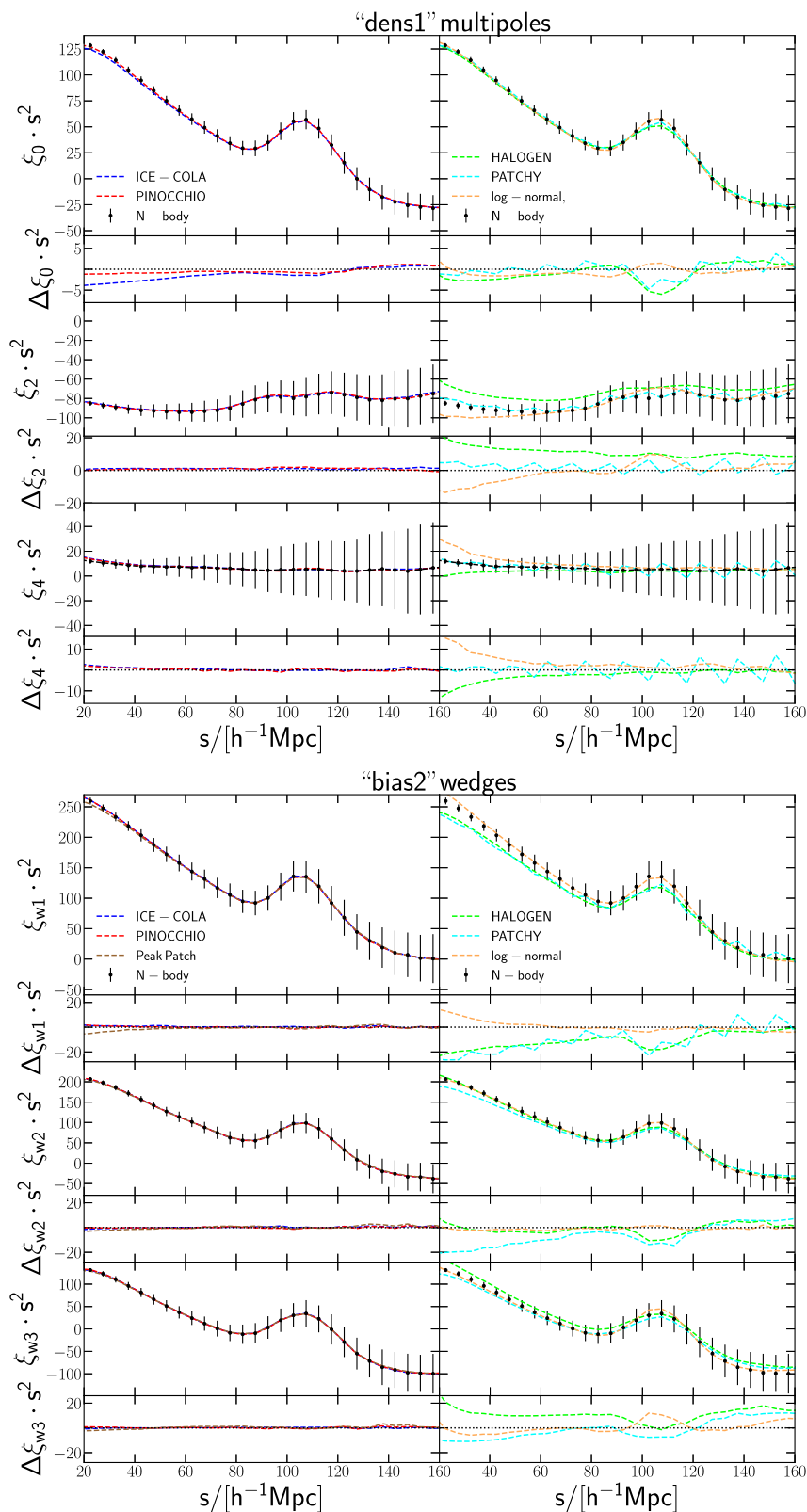
Figure 3. *Upper panel*: Measurements of the mean multipoles for the density matched samples for the first mass cut (dens1 samples). The first, third, and fifth row show the monople, quadrupole and hexadecapole, respectively. *Lower panel*: measurements of the mean clustering wedges for the bias matched samples for the second mass cut (bias2 samples). The first, third and fifth row show the transverse, intermediate, and parallel wedge, respectively. Comparison of the measurements drawn from the results of the predictive methods ICE-COLA and PINOCCHIO (*left-hand panels*) and the calibrated methods HALOGEN and PATCHY and the lognormal model (*right panels*) to the corresponding *N*-body parent sample. The error bars correspond to the dispersion of the results inferred from the 300 *N*-body catalogues. The remaining rows show the difference of the mean measurements drawn from the results of the approximate methods to the corresponding *N*-body measurement.

and the lognormal realizations, the density- and bias-matched samples are identical to the mass-matched samples by construction.

The upper panel of Fig. 3 shows the mean multipole measurements from all realizations for the dens1 samples obtained from the predictive methods ICE-COLA and PINOCCHIO (left-hand panels) and the calibrated methods HALOGEN, PATCHY, and the lognormal recipe (right-hand panels). The predictive methods are in excellent agreement with the measurements from the Minerva parent sample, showing only differences of less than 3 per cent for the ICE-COLA monopole measurements on scales $< 40\,h^{-1}$ Mpc. The monopole measurements obtained from the calibrated methods and the lognormal model are also in good agreement with the results from Minerva. However, the quadrupole and hexadecapole measurements obtained from HALOGEN and the lognormal samples exhibit deviations of more than 20 per cent on scales $< 60\,h^{-1}$ Mpc.

The lower panel of Fig. 3 shows the mean wedges measurements from all realizations for the bias2 samples obtained from the predictive methods ICE-COLA, PINOCCHIO, and PEAK PATCH (left-hand panels), and for the corresponding samples obtained from calibrated methods HALOGEN, PATCHY, and the lognormal recipe (right-hand panels). Here, we find that the measurements obtained from the predictive methods and the lognormal model agree well within the error bars with the corresponding Minerva measurements. We notice that the strongest deviations are present in the measurements of the transverse and parallel wedge from the HALOGEN samples, of up to 6 per cent and 20 per cent, respectively, on scales $< 60\,h^{-1}$ Mpc. The measurements recovered from PATCHY show deviations ranging between 5 per cent and 10 per cent on small scales.

4.2 Covariance matrix measurements

In this section, we focus on the comparison of the covariance matrix estimates obtained from the different approximate methods, which we computed as described in Section 3.3.

The structure of the off-diagonal elements of **C** of Legendre multipoles and clustering wedges measurements can be more clearly seen in the correlation matrix, defined as

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \tag{13}$$

Fig. 4 shows the correlation matrices of the multipoles inferred from the mass1 halo samples from Minerva (upper panel) and the wedges of the mass2 samples (lower panel).

The estimates of **R** obtained from the approximate methods are indistinguishable by eye from the ones inferred from the Minerva parent samples and therefore not shown here. Instead, we compare the variances and cuts through the correlation function matrices derived from the different samples. Fig. 5 shows the ratios of the variances drawn from the approximate methods with respect to those of the corresponding Minerva parent catalogues. We focus here on the same example cases as in Section 4.1: the multipoles measured from the dens1 samples, and the clustering wedges measured from the bias2 samples. We notice that in both cases the predictive methods perform better than the calibrated schemes and the PDF-based recipes. On average, the variance from Minerva is recovered within 10 per cent, with a maximum difference of 20 per cent for the variance of the monopole inferred from the PINOCCHIO dens1 sample at scales around $80\,h^{-1}$ Mpc. The variances recovered from the other methods show larger deviations, in some cases up to 40 per cent.

Fig. 6 shows cuts through the correlation matrix at $s_j = 105\,h^{-1}$ Mpc for the same two example cases. The error bars for
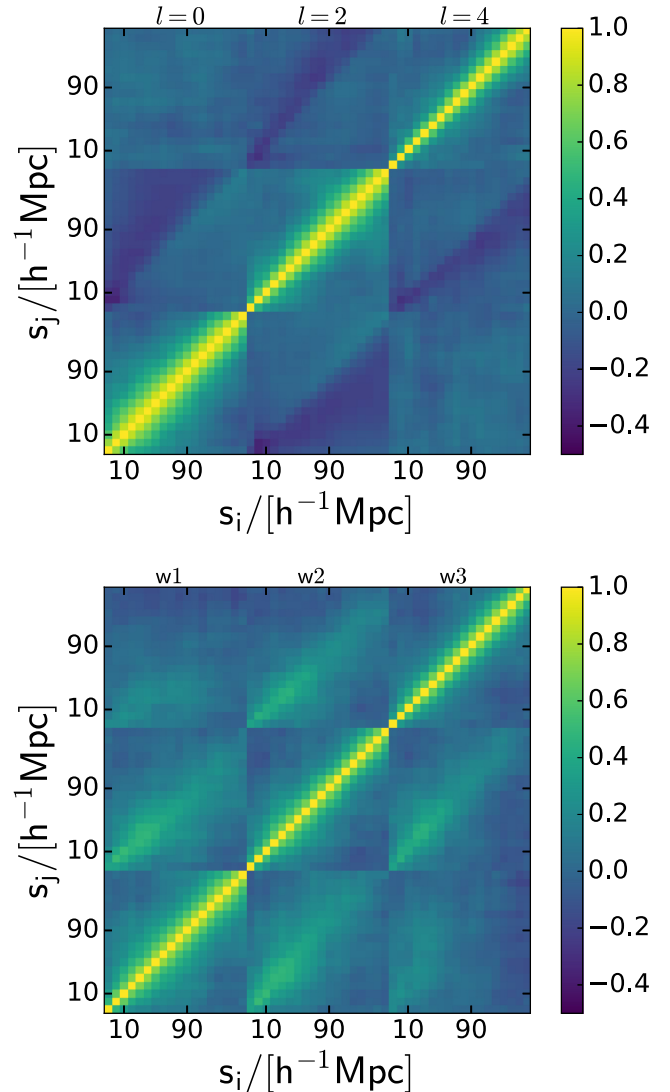

Figure 4. The full correlation matrix inferred from the multipoles of the *N*-body parent sample for the low-mass cut (mass1, upper panel) and from the clustering wedges of the mass2 *N*-body parent sample (lower panel).

the measurements of the corresponding Minerva parent samples are obtained from a jackknife estimate using the 300 Minerva mocks,

$$(\Delta M_{ij})^2 = \frac{N_S - 1}{N_S} \sum_s \left(M_{ij}^{(s)} - M_{ij} \right)^2, \tag{14}$$

where **M** is the covariance matrix **C** or the correlation matrix **R** (for Fig. 6 we use **R**). $M^{(s)}$ is the covariance or correlation matrix which is obtained when leaving out the *s*th realization,

$$M_{ij}^{(s)} = \frac{1}{N_S - 1} \sum_{r \neq s} \left(\xi_i^{(r)} - \bar{\xi}_i \right) \left(\xi_j^{(r)} - \bar{\xi}_j \right). \tag{15}$$

For the comparison of the cuts through the correlation matrices, all methods agree well the corresponding *N*-body measurements with only very small differences. In order to quantify the discrepancies between the covariance and correlation matrices drawn from the approximate methods to the corresponding *N*-body measurements,
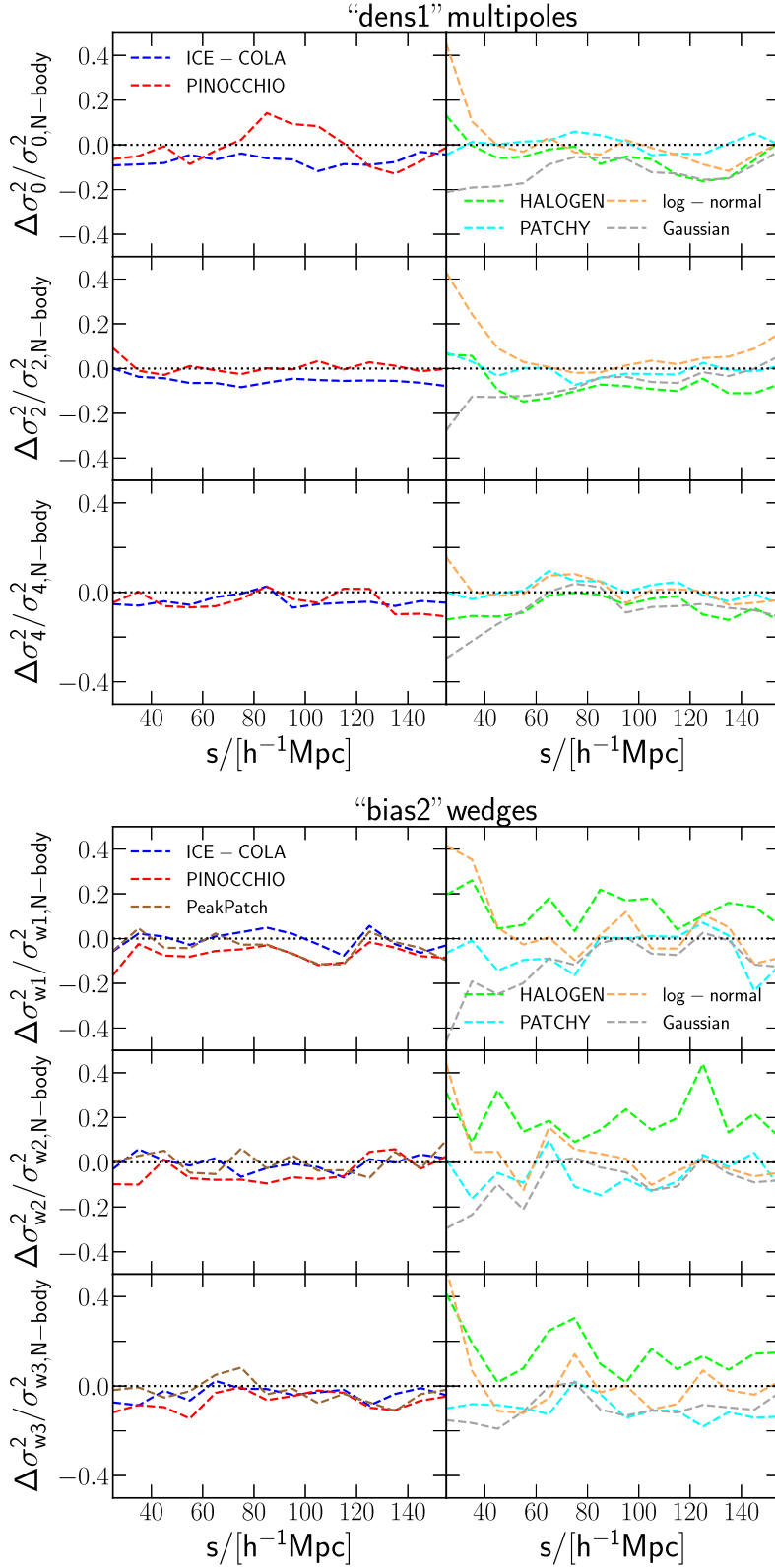
Figure 5. *Upper panel*: Relative variance of the multipoles of the correlation function measurements from the density matched samples for the first mass cut (dens1 samples).The first, third, and fifth row show the measurements for monopole, quadrupole, and hexadecapole, respectively. *Lower panel*: Relative variance of the clustering wedges of the two-point correlation function for the bias matched samples for the second mass cut (bias2 samples). The first, third, and fifth row show the measurements for transverse, intermediate, and parallel wedge, respectively. Comparison of the relative variance drawn from the results of the predictive methods ICE-COLA, PINOCCHIO, PEAK PATCH (*left-hand panel*) and HALOGEN, PATCHY, and the lognormal model (*right-hand panel*) to the corresponding *N*-body parent sample.
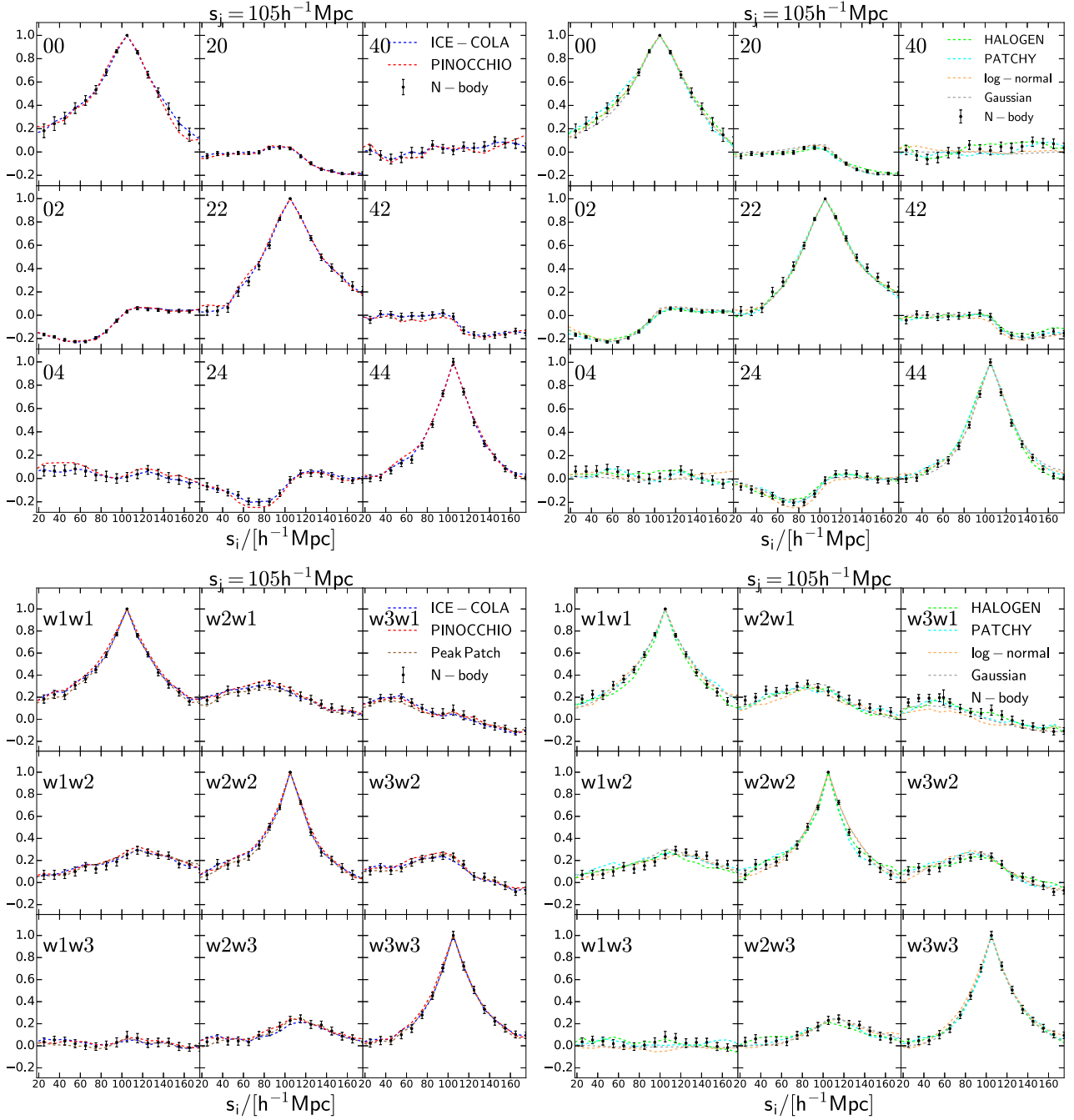
Figure 6. Cuts at $s_j = 105\,h^{-1}$ Mpc through the correlation matrices for the two example cases drawn from the results of the approximate methods and the corresponding *N*-body parent sample. *Upper, left panel*: Correlation matrices measured from the multipoles of the correlation function drawn from dens1 samples from the predictive methods ICE-COLA and PINOCCHIO. *Upper, right panel*: Correlation matrices measured from the multipoles of the correlation function drawn from dens1 samples from the calibrated methods HALOGEN and PATCHY and the Gaussian and lognormal recipes. *Lower, left panel*: Correlation matrices measured from the clustering wedges of the correlation function drawn from the bias2 samples from the predictive methods ICE-COLA, PINOCCHIO, and PEAK PATCH. *Lower, right panel*: Correlation matrices measured from the clustering wedges of the correlation function drawn from bias2 samples from the calibrated methods HALOGEN and PATCHY and the Gaussian and lognormal recipes. The error bars are obtained from a jackknife estimate using the 300 Minerva realizations.

we use an χ^2 approach. Concretely, we compute χ^2 as

$$\chi^2 = \sum_i \sum_{j \geq i} \frac{(C_{ij,\mathrm{approx}} - C_{ij,\mathrm{Minerva}})^2}{\Delta C_{ij,\mathrm{Minerva}}^2} \qquad (16)$$

and

$$\chi^2 = \sum_i \sum_{j > i} \frac{(R_{ij,\mathrm{approx}} - R_{ij,\mathrm{Minerva}})^2}{\Delta R_{ij,\mathrm{Minerva}}^2}, \qquad (17)$$

where the indices i and j run over the bins corresponding to the range of interest of 20–$160\,h^{-1}$ Mpc and $\Delta \mathbf{C}_{\mathrm{Minerva}}$ and $\Delta \mathbf{R}_{\mathrm{Minerva}}$

Table 2. Values of the relative χ^2 for the covariance matrices **C** (equation 16), correlation matrices **R** (equation 17), and values for the Kullback–Leibler divergence D_{KL} (equation 19) obtained from the approximate methods.

Code	Sample	χ^2_{rel} for **C** from ξ_{024}	χ^2_{rel} for **C** from ξ_w	χ^2_{rel} for **R** from ξ_{024}	χ^2_{rel} for **R** from ξ_w	D_{KL} for ξ_{024}	D_{KL} for ξ_w
ICE-COLA	mass1	0.19	0.21	0.17	0.16	0.24	0.24
ICE-COLA	dens1	0.31	0.42	0.17	0.15	0.28	0.27
ICE-COLA	bias1	0.20	0.11	0.19	0.19	0.27	0.27
PINOCCHIO	mass1	0.48	0.51	0.27	0.26	0.33	0.33
PINOCCHIO	dens1	0.76	0.67	0.78	0.70	0.77	0.77
PINOCCHIO	bias1	0.23	0.20	0.24	0.22	0.28	0.29
HALOGEN	mass1	1.22	0.90	1.09	0.77	1.28	1.14
PATCHY	mass1	0.67	0.40	0.73	0.44	0.82	0.79
Gaussian	mass1	2.50	2.20	2.04	0.91	0.82	1.08
Lognormal	mass1	1.76	1.09	1.31	0.97	0.96	0.98
ICE-COLA	mass2	0.40	0.36	0.38	0.33	0.43	0.45
ICE-COLA	dens2	0.36	0.23	0.35	0.27	0.28	0.28
PINOCCHIO	mass2	1.03	1.20	0.44	0.41	0.46	0.44
PINOCCHIO	dens2	0.81	0.83	0.44	0.40	0.41	0.40
PINOCCHIO	bias2	0.70	0.31	0.42	0.54	0.41	0.73
PEAK PATCH	mass2	1.84	2.02	0.69	0.69	1.05	1.03
PEAK PATCH	dens2	0.48	0.47	0.48	0.45	0.46	0.48
HALOGEN	mass2	1.77	1.32	1.70	1.29	1.07	1.07
HALOGEN	bias2	2.24	1.76	2.06	1.59	1.28	1.32
PATCHY	mass2	1.41	1.26	1.21	0.97	0.99	1.01
Gaussian	mass2	2.02	1.77	1.75	1.03	0.78	1.14
Lognormal	mass2	2.27	2.57	1.64	1.88	1.02	1.07

are the estimated errors from equation (14). If the approximate methods perfectly reproduce the expected covariances from the N-body simulations, the χ^2 obtained from the approximate methods should be $\chi^2 \approx 0$ for the predictive and calibrated methods. This is due to the fact that the simulation boxes of the predictive and calibrated methods match the ICs of Minerva and therefore the properties of the noise in the estimates of **C** should be very similar. For the covariance and correlation matrices obtained from the PDF-based predictions, we expect $\chi^2 \approx N(N-1)/2$ where N is the number of bins of the covariance or correlation matrix, since these predictions do not correspond to the same ICs. In Table 2, we list the obtained relative χ^2-values,

$$\chi^2_{rel} = \frac{\chi^2}{N(N-1)/2}, \tag{18}$$

where $N = 42$, for all considered samples and clustering statistics. We notice that the χ^2-values are in most cases smaller for the predictive than the calibrated methods. Furthermore, the χ^2-values from the wedges measurements are overall smaller than the corresponding ones from the multipole measurements. Also, in most cases the χ^2-values obtained from the covariance matrices are slightly larger than the corresponding ones from the correlation matrices, indicating discrepancies in the variances obtained from the approximated methods.

The computed χ^2-values do not take the covariance between the different entries of **C** into account. In order to provide a more complete picture of how far the multipole and wedges distributions characterized by the different covariance matrices are, we also compute the Kullback–Leibler divergence (Kullback & Leibler 1951; O'Connell et al. 2016). In our case (two multivariate normal distributions with the same means), the Kullback–Leibler divergence is given as

$$D_{KL}(\mathbf{C}_{Minerva} \parallel \mathbf{C}_{approx}) = \frac{1}{2} \left(\mathrm{tr}(\mathbf{C}^{-1}_{approx} \mathbf{C}_{Minerva}) + \ln \left(\frac{\det \mathbf{C}_{approx}}{\det \mathbf{C}_{Minerva}} \right) - N \right). \tag{19}$$

If the approximate methods perfectly reproduce the expected distributions from the N-body simulations, including the same noise, we expect $D_{KL} \approx 0$. In Table 2, we list the obtained D_{KL} values. We find that the values for D_{KL} are closer to zero for the predictive than for the other approximate methods. For the calibrated methods and for the distributions with different noise, obtained from the Gaussian and lognormal models, we find values $D_{KL} \approx 1$.

4.3 Performance of the covariance matrices

For the final validation of the covariance matrices inferred from the different approximate methods, we analyse their performance on cosmological parameter constraints. We perform fits to the synthetic clustering measurements described in Section 3.4, using the estimates of **C** obtained from the different halo samples and approximate methods. We focus on the constraints on the BAO shift parameters $\alpha_\parallel$, $\alpha_\perp$, and the growth rate $f\sigma_8$.

Fig. 7 shows the two-dimensional marginalized constraints in the $\alpha_\perp$–$f\sigma_8$ plane for the analysis of our two examples cases, the Legendre multipoles measured from the dens1 samples (upper panels), and the clustering wedges recovered from the bias2 samples (lower panel).

In general, the allowed regions for these parameters obtained using the estimates of **C** inferred from the different approximate methods (shown by the solid lines) agree well with those obtained using the covariance matrices from Minerva (indicated by the dotted lines in all panels). However, most cases exhibit small deviations, either slightly under- or overestimating the statistical uncertainties.
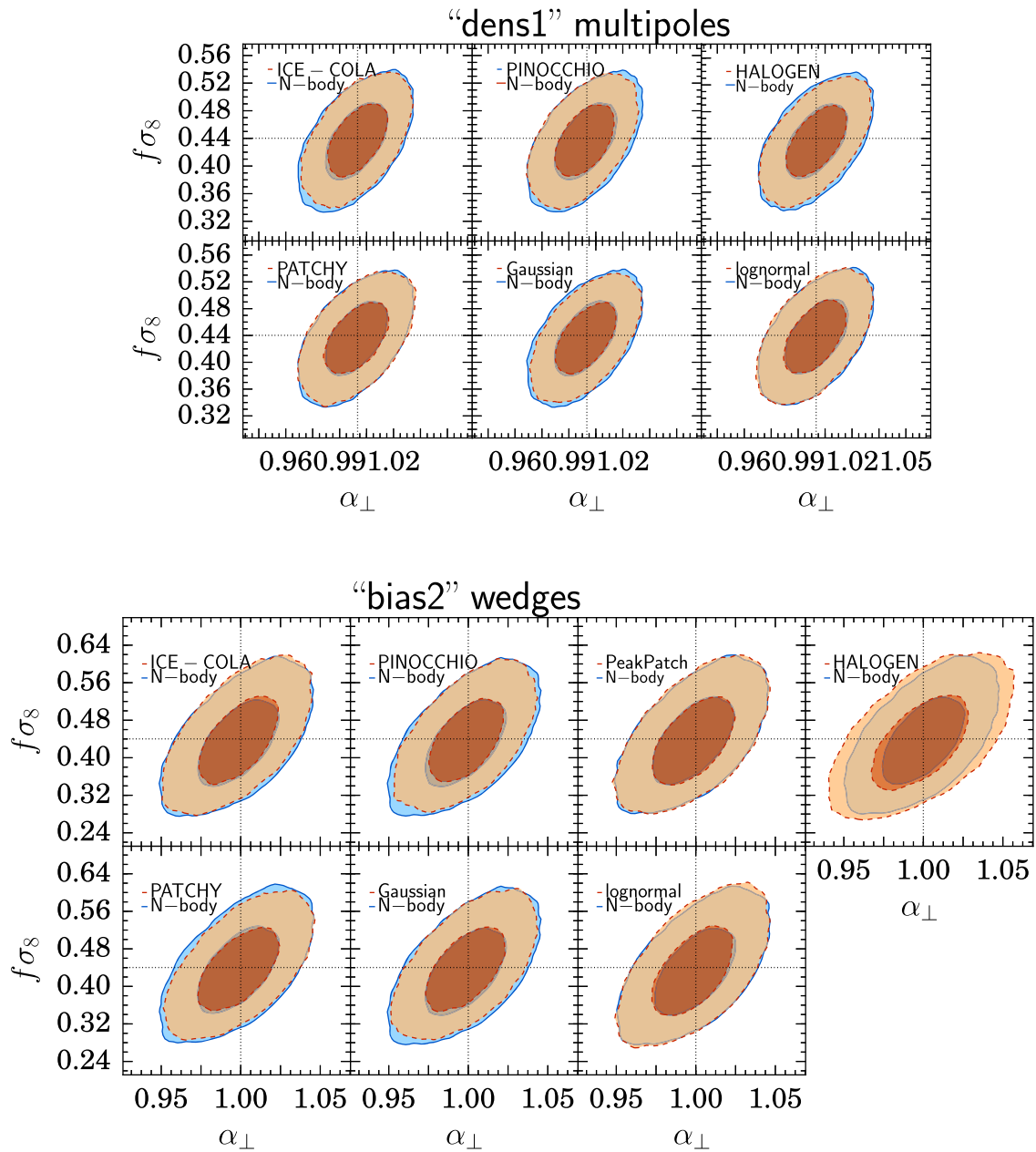
Figure 7. Comparison of the marginalized two-dimensional constraints in the $\alpha_\perp$–$f\sigma_8$ plane for the analysis of samples from the approximate methods with the corresponding constraints obtained from analysis of the parent Minerva sample. The contours correspond to the 68 per cent and 95 per cent confidence levels.*Upper panel*: Results from the analysis of the multipoles measured from the dens1 samples. *Lower panel*: Results from the analysis of the clustering wedges measured from the bias2 samples.

We find that, for all samples and clustering statistics, the mean parameter values inferred using approximate methods are in excellent agreement with the ones from the corresponding *N*-body analysis, showing differences that are much smaller than their associated statistical errors. The parameter uncertainties recovered using covariances from the approximate methods show differences with respect to the *N*-body constraints ranging between 0.3 per cent and 8 per cent for the low-mass samples, while most of the results agree within 5 per cent with the *N*-body results, and between 0.1 per cent and 20 per cent for the high-mass cut, while most of the results agree within 10 per cent with the *N*-body results. For the comparison of the obtained parameter uncertainties it is important to point out that in our companion paper, Blot et al. (2018)

estimate that the statistical limit of our parameter estimation is about 4–5 per cent. Fig. 8 shows the ratios of the marginalized parameter errors drawn from the analysis with the different approximate methods with respect to the *N*-body results. We observe that for the samples corresponding to the first mass cut, all methods reproduce the *N*-body errors within 10 per cent for all parameters, and in most cases within 5 per cent corresponding to the statistical limit of our analysis. For the samples corresponding to the second mass cut also most methods reproduce the *N*-body errors within 10 per cent with exception of the PEAK PATCH mass-matched and the HALOGEN bias-matched samples. This might be due to the fact that these two samples have 15–20 per cent less haloes than the corresponding *N*-body sample.
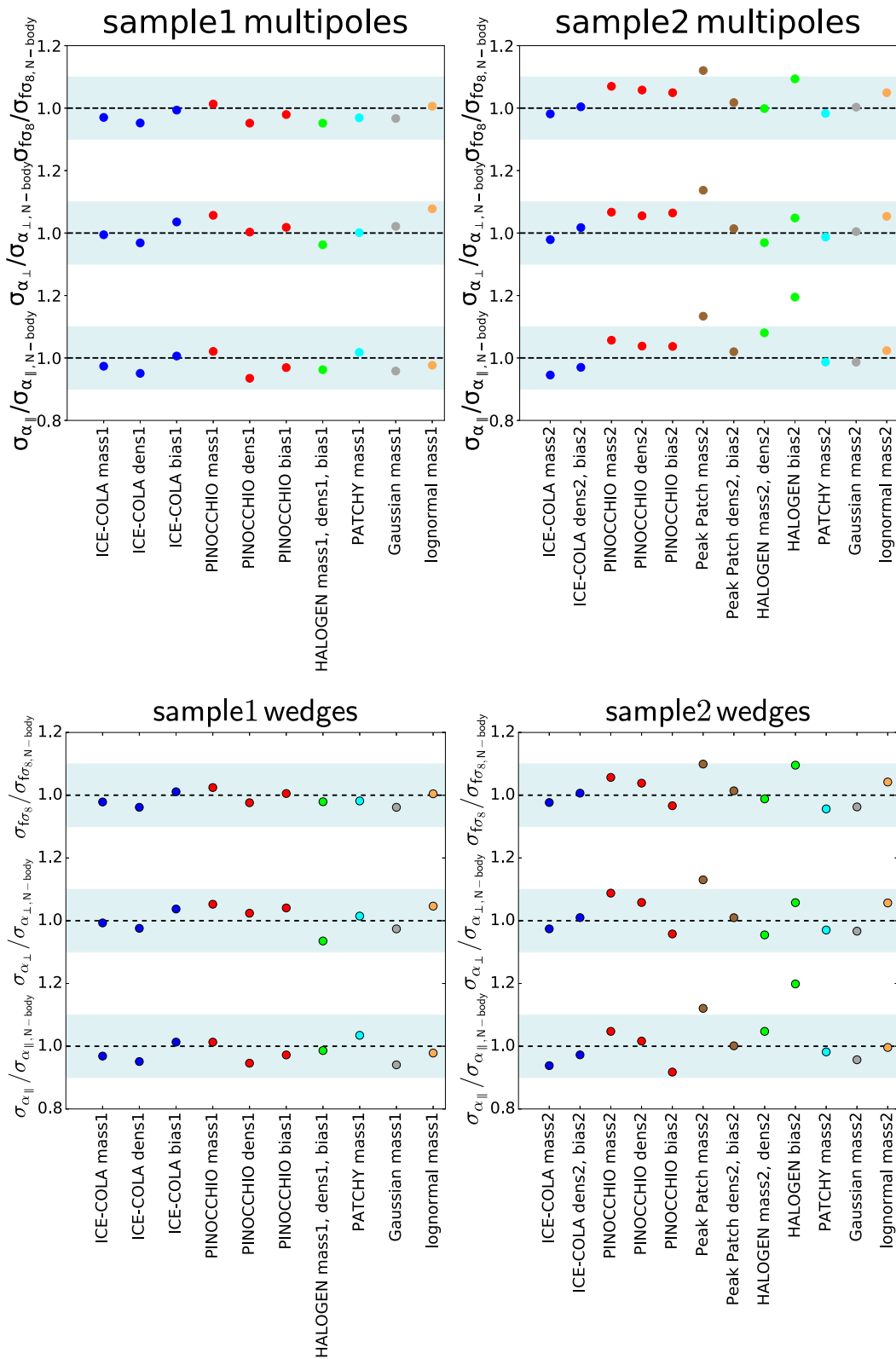
Figure 8. Comparison of the marginalized error on the parameters $\alpha_{\parallel}$, $\alpha_{\perp}$, and $f\sigma_8$ which are obtained from the analysis using the covariance matrices from the approximate methods to the corresponding ones from the *N*-body catalogues. The light grey band indicates a range of ±10 per cent deviation from a ratio equal to 1. The different panels show the results obtained from the analysis of *upper, left panel*: the multipoles drawn from the samples corresponding to the first *N*-body parent sample with the lower mass cut, *upper, right panel*: the multipoles drawn from the samples corresponding to the second *N*-body parent sample with the higher mass cut, *lower, left panel*: the wedges drawn from the samples corresponding to the first *N*-body sample, *lower, right panel*: the wedges drawn from the samples corresponding to the second *N*-body sample.

In order to evaluate the parameter errors, we use the volume of the allowed region in the three-dimensional parameter space of $\alpha_\parallel$, $\alpha_\perp$, and $f\sigma_8$, which can be estimated as

$$V = \sqrt{\det \mathrm{Cov}(\alpha_\parallel, \alpha_\perp, f\sigma_8)}, \qquad (20)$$

where $\det \mathrm{Cov}(\alpha_\parallel, \alpha_\perp, f\sigma_8)$ is the determinant of the parameter covariance matrix. For a Gaussian posterior distribution, the allowed volume is proportional to the volume enclosed by the three-dimensional 68 per cent C.L. contour. This definition is similar to the two-dimensional Dark Energy Task Force figure of merit of the dark-energy equation-of-state parameters w_0–w_a (Albrecht et al. 2006; Wang 2008), but without taking the inverse of the allowed volume. The ratios of the allowed volumes obtained from the analysis with the different approximate methods and the N-body results are shown in Fig. 9. Here, the differences in the performance of the methods become clearer. For the first mass cut, we notice that most approximate methods can reproduce the N-body volume at a 10 per cent level, with the exception of HALOGEN and the Gaussian and lognormal models, which lead to slightly worse results and show 10–15 per cent agreement. For the second mass cut, we find overall larger differences between the samples. The results from the majority of the samples agree within 10 per cent with the N-body results, the rest shows differences of 10–15 per cent, and for the PEAK PATCH mass2 and HALOGEN bias2 samples differences of up to 40 per cent. For both mass cuts, we find significant differences in the performances of samples drawn from the same approximate method but using different selection criteria.

5 DISCUSSION

In this section, we discuss our results on the allowed parameter space volumes obtained in Section 4.3. Fig. 9 clearly shows that there are significant differences in the volume ratios between samples drawn from the same approximate method when applying different selection criteria to define the halo catalogues. Matching the parent samples from Minerva by mass limit, number density or bias can lead to differences of up to 20 per cent on the obtained results.

For each approximate method, mass limit, and clustering statistic, we identified the best selection criteria for matching to the N-body parent samples. As discussed in Section 3.1, for PATCHY, lognormal, and the Gaussian model we only have samples characterized by the same mass cuts as the N-body catalogues. The best cases in decreasing order of the accuracy with which the results of the N-body covariances are reproduced are as follows:

(i) Lower mass cut, Legendre multipoles: PATCHY ($V/V_{\mathrm{Min}} = 1.02$), PINOCCHIO bias matched ($V/V_{\mathrm{Min}} = 0.97$), ICE-COLA mass matched ($V/V_{\mathrm{Min}} = 0.96$), lognormal ($V/V_{\mathrm{Min}} = 1.11$), Gaussian ($V/V_{\mathrm{Min}} = 0.88$), HALOGEN mass, density, bias matched ($V/V_{\mathrm{Min}} = 0.85$).

(ii) Lower mass cut, clustering wedges: PINOCCHIO bias matched ($V/V_{\mathrm{Min}} = 1.01$), ICE-COLA mass matched ($V/V_{\mathrm{Min}} = 0.96$), PATCHY ($V/V_{\mathrm{Min}} = 1.07$), lognormal ($V/V_{\mathrm{Min}} = 1.09$), HALOGEN mass, density, bias matched ($V/V_{\mathrm{Min}} = 0.87$), Gaussian ($V/V_{\mathrm{Min}} = 0.87$).

(iii) Higher mass cut, Legendre multipoles: ICE-COLA density matched ($V/V_{\mathrm{Min}} = 0.96$), HALOGEN mass, density matched ($V/V_{\mathrm{Min}} = 1.04$), PEAK PATCH density, biased matched ($V/V_{\mathrm{Min}} = 1.06$), PATCHY ($V/V_{\mathrm{Min}} = 0.94$), Gaussian ($V/V_{\mathrm{Min}} = 0.92$), PINOCCHIO density matched ($V/V_{\mathrm{Min}} = 1.11$), lognormal ($V/V_{\mathrm{Min}} = 1.16$).

(iv) Higher mass cut, clustering wedges: HALOGEN mass, density matched ($V/V_{\mathrm{Min}} = 1.02$), ICE-COLA density matched ($V/V_{\mathrm{Min}} = $ 0.97), PEAK PATCH density, biased matched ($V/V_{\mathrm{Min}} = 1.03$), PATCHY ($V/V_{\mathrm{Min}} = 0.91$), lognormal ($V/V_{\mathrm{Min}} = 1.09$), PINOCCHIO density matched ($V/V_{\mathrm{Min}} = 1.1$), Gaussian ($V/V_{\mathrm{Min}} = 0.87$).

For a better illustration, Fig. 10 shows the two-dimensional marginalized constraints on $\alpha_\perp$ and $f\sigma_8$ obtained from the Legendre multipoles for the low (upper panels) and high (lower panels) mass limits. The different panels show the results obtained from the different approximate methods when the best selection criteria for each case is implemented. The overall agreement with the results derived from the N-body covariances is better in this case than when the same definition is applied to all methods.

The best strategy to define the halo samples for a given approximate method is often different for our two mass limits. For example, considering the results from PINOCCHIO, while for our first mass limit the bias-matched halo samples lead to the best agreement with the constraints inferred from the N-body covariances, for the second mass threshold the density-matched samples provide a better performance. Focusing on the results from the multipole analysis, we observe that for the first mass limit PATCHY, ICE-COLA, and PINOCCHIO perform slightly better than the other methods. These methods reproduce the statistical volume of the allowed parameter regions obtained using the N-body covariances within 5 per cent while the other methods only reach a 10–15 per cent agreement. For the second mass limit ICE-COLA, HALOGEN, and PEAK PATCH can reproduce the N-body results within 5 per cent, PATCHY and the Gaussian model within 10 per cent, and PINOCCHIO and the lognormal model within 15 per cent. It is also interesting to note that the order of performance of the methods is slightly different for the multipole and the wedges analysis. For example, the multipole analysis using the PATCHY covariance matrix leads to a better than 2 per cent agreement with the N-body results, whereas the wedge analysis only reaches 7 per cent.

Our analysis is part of a general comparison project of approximate methods involving also the covariances of power spectrum and bispectrum measurements (Blot et al. 2018; Colavincenzo et al. 2018). The power spectrum analysis of Blot et al. (2018) is more closely related to the one presented here, as it is based on the same baseline model of the two-dimensional power spectrum and explore constraints on the same nuisance and cosmological parameters. The bispectrum covariance analysis of Colavincenzo et al. (2018) is different in terms of the model and the parameter constraints included in the comparison. Both of our companion papers consider the same approximate methods and mass cuts used here, but focus on the abundance-matched samples. A comparison of the results of the three studies shows that the differences between the predictive, calibrated and PDF-based approximate methods are less evident for the correlation function analysis than for the power spectrum and bispectrum. This can be clearly seen by comparing the variations of the statistically allowed volumes recovered from the different approximate methods when applied to the correlation function, power spectrum and bispectrum covariances. Since our companion papers focus on the density-matched samples, we also show the allowed volumes only for the 'dens' samples in Fig. 11. The differences between the approximate methods are less evident in configuration space, become more evident for the power spectrum and are strongest for the bispectrum analysis.

In summary, our results and those of our companion papers indicate that approximate methods can provide robust covariance matrix estimates for cosmological parameter constraints. However, the differences seen between the various recipes, statistics, and selection criteria considered here highlight the importance of performing de-
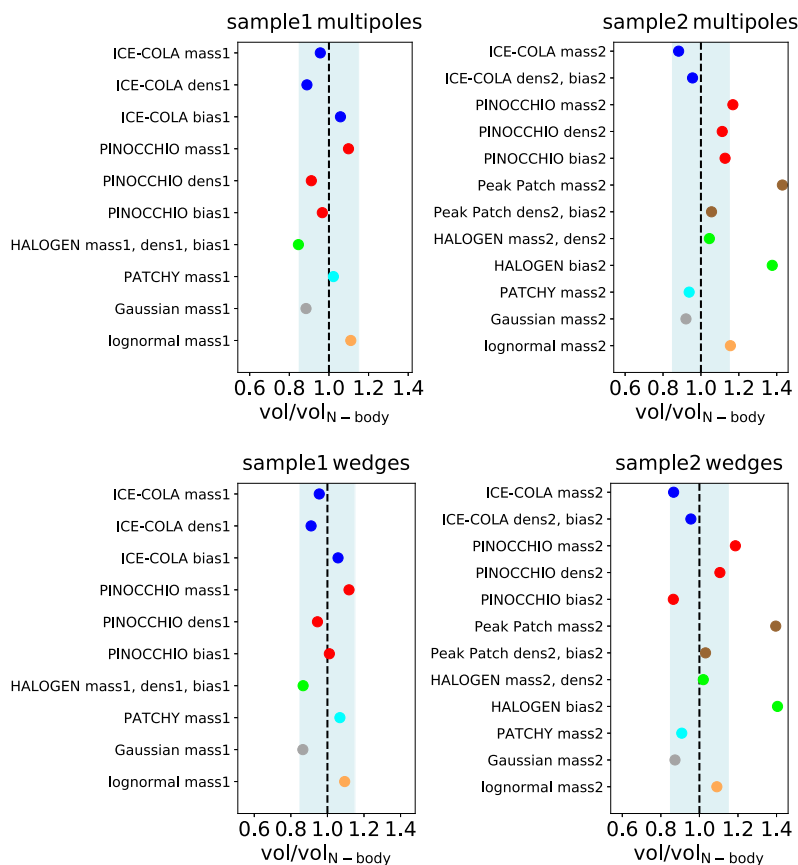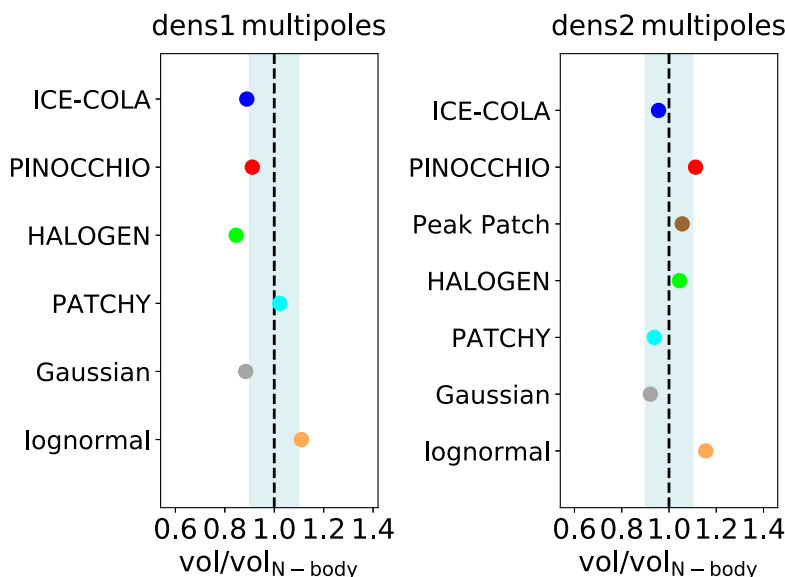
Figure 9. Comparison of the volume ratios between the allowed statistical volumes obtained from the analysis using the covariance matrices from the approximate methods to the corresponding ones from the *N*-body catalogues. The light grey band indicates a range of ± 10 per cent deviation from a ratio equal to 1. The different panels show the results obtained from the analysis of *upper, left panel*: the multipoles drawn from the samples corresponding to the first *N*-body parent sample with the lower mass cut, *upper, right panel*: the multipoles drawn from the samples corresponding to the second *N*-body parent sample with the higher mass cut, *lower, left panel*: the wedges drawn from the samples corresponding to the first *N*-body sample, *lower, right panel*: the wedges drawn from the samples corresponding to the second *N*-body sample.

tailed tests to find the best strategy to draw halo samples from any given approximate method.

6 SUMMARY AND CONCLUSIONS

We have analysed the performance of several approximate methods at providing estimates of the covariance matrices of anisotropic two-point clustering measurements in configuration space. Our analysis is part of a comparison project, including also detailed studies of the covariance matrices of power spectrum and bispectrum measurements, which are summarized in our companion papers Blot et al. (2018) and Colavincenzo et al. (2018), respectively.

Our comparison included seven approximate methods, which we divided into three categories: predictive methods (ICE-COLA, PEAK PATCH, and PINOCCHIO), methods that require calibration with *N*-body simulations (HALOGEN and PATCHY), and recipes based on assumptions regarding the shape of the density PDF (lognormal and Gaussian density fields). We compared these methods against the results obtained from the Minerva simulations. We generated sets of 300 halo catalogues using the predictive and calibrated methods, matching the ICs of the reference *N*-body simulations. For the lognormal predictions we generated a set of 1000 catalogues designed to match the number density and mean correlation function measured from the *N*-body simulations.

We defined two halo samples from the Minerva simulations by applying mass thresholds corresponding to 42 and 100 DM particles. We then selected different halo samples from the approximate methods by matching the mass threshold, number density and clustering amplitude of the parent samples from the *N*-body simulations. We estimated the covariance matrices of the Legendre multipoles and clustering wedges corresponding to all halo samples and compared the results with the corresponding ones from the parent catalogues.

Our main comparison was focused on the accuracy with which the covariance matrices inferred from the approximate methods reproduce the cosmological parameter constraints obtained from the *N*-body results. For this, we first used a model of the two-dimensional power spectrum applied in recent LSS analyses (Grieb et al. 2017; Salazar-Albornoz et al. 2017; Sánchez et al. 2017; Hou et al. 2018) to construct synthetic clustering measurements, and then fitted these data with the same baseline model, using the covariances from the different methods and assuming a Gaussian likelihood function. We analysed the obtained parameter constraints on $\alpha_{\parallel}$, $\alpha_{\perp}$ and $f\sigma_8$. The mean values obtained from the fits agree perfectly with the *N*-body results for all the samples. Most methods recover the marginalized *N*-body parameter errors within 5 per cent for the lower mass cut, which corresponds also to the statistical limit of our analysis, and 10 per cent for the higher mass cut. The comparison of the statisti-
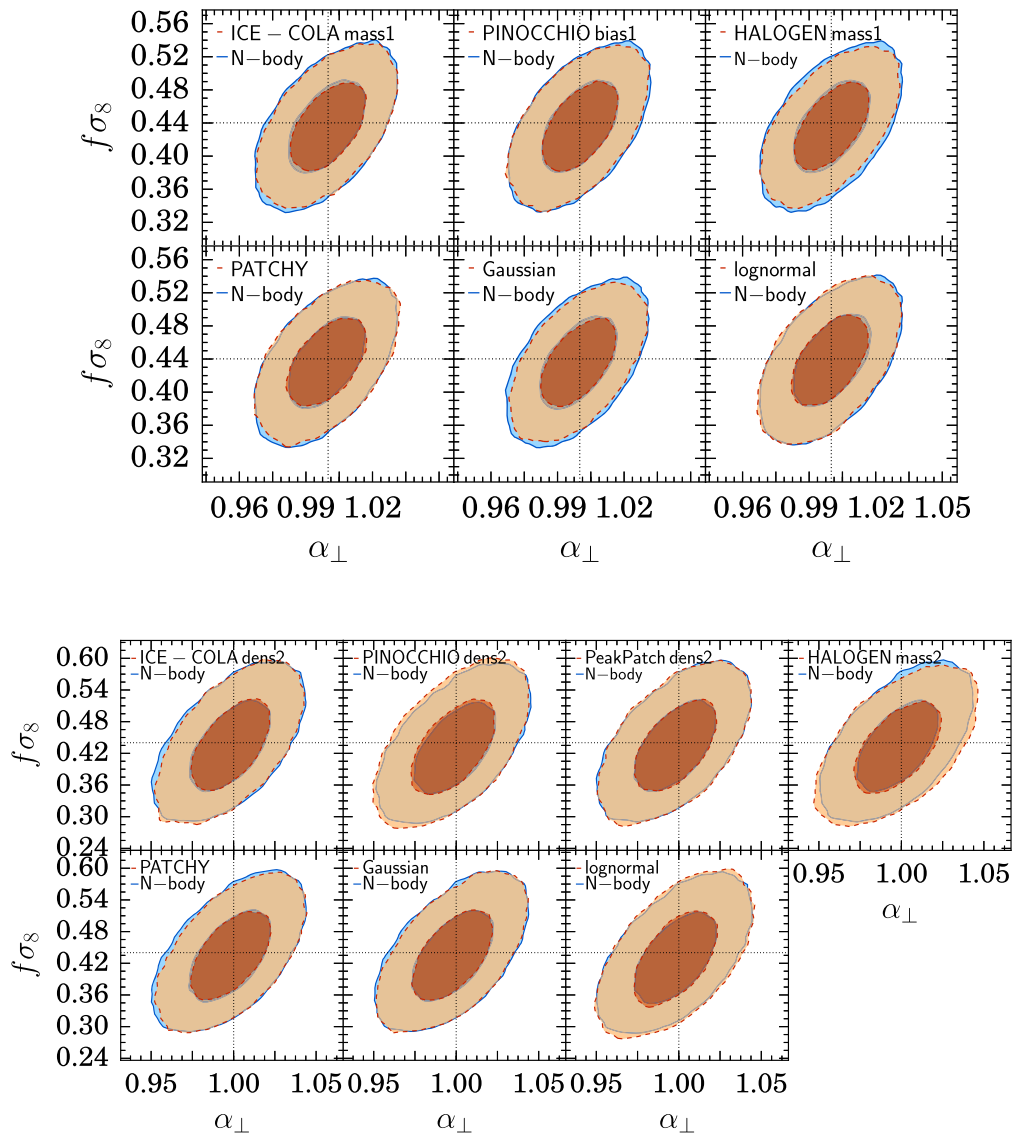
Figure 10. Comparison of the marginalized two-dimensional constraints in the $\alpha_\perp$–$f\sigma_8$ plane for the multipole analysis using the best choice of matching for each approximate method individually to the corresponding constraints obtained from the *N*-body analysis. The contours correspond to the 68 per cent and 95 per cent confidence levels. *Upper panel*: Results for the samples corresponding to the first mass cut. *Lower panel*: Results for the samples corresponding to the second mass cut.

cally allowed volumes in the three-dimensional parameter space of $\alpha_\parallel$, $\alpha_\perp$, and $f\sigma_8$ shows that the results obtained from any given approximate method by implementing different selection criteria, i.e. by matching the mass, number density, or bias of the parent *N*-body samples, can differ by up to 20 per cent. Therefore, for each approximate method and mass limit we identified the selection scheme that provided the closest agreement with the results obtained using the estimates of **C** from the *N*-body simulations. For the first mass cut, we found that the methods ICE-COLA, PINOCCHIO, and PATCHY reproduce the *N*-body results slightly better than the other methods, with differences of less than 10 per cent in the allowed volumes. The remaining methods show a 10–15 per cent agreement with the *N*-body results. For the second mass cut, ICE-COLA, HALOGEN, and PEAK PATCH perform the best, recovering the *N*-body allowed volumes within 5 per cent. The fits using the other methods

lead to a 5–15 per cent agreement. It is noteworthy that the simple Gaussian prediction performs similar to the other approximate methods.

We conclude that, with respect to the covariance matrices of configuration-space clustering measurements, there is no clear preference for one of the approximate methods. The predictive methods ICE-COLA, PEAK PATCH, and PINOCCHIO do not outperform the calibrated methods and simpler recipes significantly. The advantage of using the calibrated methods is that they are computationally less expensive. However, the calibration using full *N*-body simulations can also be challenging and time-consuming. In future studies, we will include additional effects, such as the impact of survey geometry, that will allow us to extend our analysis to assess the impact of applying approximate methods to the analysis of real galaxy surveys.

Figure 11. Volume ratios between the allowed statistical volumes obtained from the analysis using the covariance matrices from the approximate methods to the corresponding ones from the *N*-body catalogues for the density matched samples. The light grey band indicates a range of ± 10 per cent deviation from a ratio equal to 1.

REFERENCES

Agrawal A., Makiya R., Chiang C.-T., Jeong D., Saito S., Komatsu E., 2017, J. Cosmol. Astropart. Phys., 10, 003
Alam S. et al., 2017, MNRAS, 470, 2617
Albrecht A. et al., 2006, preprint (arXiv:astro-ph/0609591)
Anderson L. et al., 2012, MNRAS, 427, 3435
Anderson L. et al., 2014a, MNRAS, 439, 83
Anderson L. et al., 2014b, MNRAS, 441, 24
Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, MNRAS, 450, 1856
Blake C., Glazebrook K., 2003, ApJ, 594, 665
Blot L.et al., 2018, preprint (arXiv:1806.09497)
Bond J. R., Myers S. T., 1996, ApJS, 103, 1
Chan K. C., Scoccimarro R., 2012, Phys. Rev. D, 86, 103519
Chuang C.-H. et al., 2015, MNRAS, 452, 686
Colavincenzo M.et al., 2018, preprint (arXiv:1806.09499)
Cole S. et al., 2005, MNRAS, 362, 505
Coles P., Jones B., 1991, MNRAS, 248, 1
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
Efstathiou G. et al., 2002, MNRAS, 330, L29
Eisenstein D. J. et al., 2005, ApJ, 633, 560
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
Guzzo L. et al., 2008, Nature, 451, 541
Hamilton A. J. S., Rimes C. D., Scoccimarro R., 2006, MNRAS, 371, 1188
Hou J. et al., 2018, MNRAS, 480, 2521
Izard A., Crocce M., Fosalba P., 2016, MNRAS, 459, 2327
Izard A., Fosalba P., Crocce M., 2018, MNRAS, 473, 3051
Kazin E. A., Sánchez A. G., Blanton M. R., 2012, MNRAS, 419, 3223
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., Heß S., 2013, MNRAS, 435, L78
Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
Kitaura F.-S. et al., 2016, MNRAS, 456, 4156
Koda J., Blake C., Beutler F., Kazin E., Marin F., 2016, MNRAS, 459, 2118
Kullback S., Leibler R. A., 1951, Ann. Math. Stat., 22, 79
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Linder E. V., 2003, Phys. Rev. Lett., 90, 091301
Monaco P., 2016, Galaxies, 4, 53
Monaco P., Theuns T., Taffoni G., 2002, MNRAS, 331, 587
Munari E., Monaco P., Koda J., Kitaura F.-S., Sefusatti E., Borgani S., 2017, J. Cosmol. Astropart. Phys., 7, 050
O'Connell R., Eisenstein D., Vargas M., Ho S., Padmanabhan N., 2016, MNRAS, 462, 2681
Padmanabhan N., White M., 2008, Phys. Rev. D, 77, 123540
Paz D. J., Sánchez A. G., 2015, MNRAS, 454, 4326
Percival W. J. et al., 2014, MNRAS, 439, 2531
Pope A. C., Szapudi I., 2008, MNRAS, 389, 766
Salazar-Albornoz S. et al., 2017, MNRAS, 468, 2938
Schneider M. D., Cole S., Frenk C. S., Szapudi I., 2011, ApJ, 737
Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
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 R. J., 2018, preprint (arXiv:1810.07727)
Sánchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 189
Sánchez A. G. et al., 2012, MNRAS, 425, 415
Sánchez A. G. et al., 2013, MNRAS, 433, 1202
Sánchez A. G. et al., 2017, MNRAS, 464, 1640
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 6, 036
Taylor A., Joachimi B., Kitching T., 2013, MNRAS, 432, 1928
Vakili M., Kitaura F.-S., Feng Y., Yepes G., Zhao C., Chuang C.-H., Hahn C., 2017, MNRAS, 472, 4144
Wang Y., 2008, Phys. Rev. D, 77, 123525
Zhao C., Kitaura F.-S., Chuang C.-H., Prada F., Yepes G., Tao C., 2015, MNRAS, 451, 4266

[1] *Universitäts-Sternwarte München, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, D-81679 Munich, Germany*
[2] *Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Giessenbachstr, D-85741 Garching, Germany*
[3] *Dipartimento di Fisica, Università di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[4] *Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[5] *Dipartimento di Fisica, Sezione di Astronomia, Università di Trieste, via Tiepolo 11, I-34143 Trieste, Italy*
[6] *Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34143 Trieste, Italy*
[7] *Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy*
[8] *Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[9] *Institut d'Estudis Espacials de Catalunya (IEEC), E-08193 Barcelona, Spain*
[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*
[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, PO 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.