Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527**, 11539–11558 (2024) https://doi.org/10.1093/mnras/stad3742
Advance Access publication 2023 December 4

DESI mock challenge: constructing DESI galaxy catalogues based on FastPM simulations

Andrei Variu,[1★] Shadab Alam,[2,3★] Cheng Zhao,[1] Chia-Hsun Chuang,[4,5] Yu Yu,[6,7] Daniel Forero-Sánchez,[1] Zhejie Ding,[6,7] Jean-Paul Kneib,[1] Jessica Nicole Aguilar,[8] Steven Ahlen,[9] David Brooks,[10] Todd Claybaugh,[8] Shaun Cole,[11] Kyle Dawson,[4] Axel de la Macorra,[12] Peter Doel,[10] Jaime E. Forero-Romero,[13,14] Satya Gontcho A Gontcho,[8] Klaus Honscheid,[15,16] Martin Landriau,[8] Marc Manera,[17,18] Ramon Miquel,[18,19] Jundan Nie,[20] Will Percival,[21,22,23] Claire Poppett,[8,24] Mehdi Rezaie,[25] Graziano Rossi,[26] Eusebio Sanchez,[27] Michael Schubnell,[28] Hee-Jong Seo,[29] Gregory Tarlé,[28] Mariana Vargas Magana[12] and Zhimin Zhou[20]

Affiliations are listed at the end of the paper

Accepted 2023 November 27. Received 2023 November 14; in original form 2023 July 26

ABSTRACT

Together with larger spectroscopic surveys such as the Dark Energy Spectroscopic Instrument (DESI), the precision of large-scale structure studies and thus the constraints on the cosmological parameters are rapidly improving. Therefore, one must build realistic simulations and robust covariance matrices. We build galaxy catalogues by applying a halo occupation distribution (HOD) model upon the FastPM simulations, such that the resulting galaxy clustering reproduces high-resolution N-body simulations. While the resolution and halo finder are different from the reference simulations, we reproduce the reference galaxy two-point clustering measurements – monopole and quadrupole – to a precision required by the DESI Year 1 emission line galaxy sample down to non-linear scales, i.e. $k < 0.5\,h\,\mathrm{Mpc}^{-1}$ or $s > 10\,\mathrm{Mpc}\,h^{-1}$. Furthermore, we compute covariance matrices based on the resulting FastPM galaxy clustering – monopole and quadrupole. We study for the first time the effect of fitting on Fourier conjugate (e.g. power spectrum) on the covariance matrix of the Fourier counterpart (e.g. correlation function). We estimate the uncertainties of the two parameters of a simple clustering model and observe a maximum variation of 20 per cent for the different covariance matrices. Nevertheless, for most studied scales the scatter is between 2 and 10 per cent. Consequently, using the current pipeline we can precisely reproduce the clustering of N-body simulations and the resulting covariance matrices provide robust uncertainty estimations against HOD fitting scenarios. We expect our methodology will be useful for the coming DESI data analyses and their extension for other studies.

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

1 INTRODUCTION

The study of the large-scale structure of the Universe has significantly improved in the last two decades leading to Baryon Oscillation Spectroscopic Survey (BOSS; Alam et al. 2017) and extended-BOSS (eBOSS; Alam et al. 2021a) surveys. They have published the largest 3D map of over 2 millions galaxies and quasars (Alam et al. 2021a). This has allowed the measurement of cosmological parameters to a per cent-level precision studying baryonic acoustic oscillations (BAOs) and redshift–space distortions (RSDs).

Currently, the Dark Energy Spectroscopic Instrument (DESI; Levi et al. 2013; DESI Collaboration 2022) is a 5 yr long spectroscopic survey that will outperform previous surveys by a an order of magnitude (DESI Collaboration 2016a), aiming to constrain the cosmological parameters with precision at a sub-per cent level. With its 5000 robotically controlled optical fibres (DESI Collaboration 2016b; Miller et al. 2023; Silber et al. 2023), DESI will scan a third of the sky to map 40 millions galaxies (Lan et al. 2023) and quasars (Alexander et al. 2023). Only after the 5-month survey validation (DESI Collaboration 2023a), DESI has measured the spectra of more than one million galaxies leading to the recent Early Data Release (DESI Collaboration 2023b).

Based on the DESI Legacy Imaging Surveys (Zou et al. 2017; Dey et al. 2019; Schlegel et al. in preparation), there are five types of targets that are selected (Myers et al. 2023) on which optical fibres are assigned (Raichoor et al. in preparation) to measure and analyse their spectra (Brodzeller et al. 2023; Guy et al. 2023; Bailey et al. in preparation): Milky Way Stars (Allende Prieto et al. 2020; Cooper et al. 2023), bright galaxies (Ruiz-Macias et al. 2020; Hahn et al. 2023), luminous red galaxies (Zhou et al. 2020, 2023), emission-line galaxies (ELGs; Raichoor et al. 2020, 2023b), quasars (Yèche

★ E-mail: andrei.variu@epfl.ch (AV); shadab.alam@tifr.res.in (SA)

et al. 2020; Chaussidon et al. 2023). Such a complex system requires pipelines to optimize the observations (Schlafly et al. 2023; Kirkby et al. in preparation).

The sub-per cent precision measurements expected from ongoing and future surveys require careful analyses of the systematic effects. To this end, the DESI Mock Challenge was launched as a series of studies and projects to build and validate the methodology for the cosmological analysis. In particular, one must find the most robust way to estimate the uncertainty of the measurements (Chuang et al. in preparation). To achieve this goal, one needs to create multiple realistic simulations of the large-scale structure, which is required to lower the noise on covariance matrix and to describe accurately the non-linear scales.

On the one hand, the full N-body simulations – e.g. [Scinet LIght-Cone Simulations (SLICS); Harnois-Déraps et al. 2018], (UNIT; Chuang et al. 2019), and (ABACUSSUMMIT; Maksimova et al. 2021) – are accurate, but they are computationally expensive. Since it becomes impractical with the increase of mapped volume to have enough full N-body realizations to compute and test covariance matrices, they are mainly used in testing models and systematic effects. Consequently, faster but less accurate techniques have been developed – e.g. (EZMOCKS; Chuang et al. 2015; Zarrouk et al. 2021; Zhao et al. 2021), (PATCHY; Kitaura, Yepes & Prada 2013), and (BAM; Balaguera-Antolínez et al. 2019, 2020; Pellejero-Ibañez et al. 2020) – to be run multiple times and estimate robustly the uncertainty.

In this study, we investigate the possibility to tune FASTPM catalogues to reproduce the clustering of SLICS reference with the final goal of estimating the covariance matrix. In contrast to the other fast methods, FASTPM uses accelerated particle-mesh (PM) solvers to evolve the dark-matter field, which should provide a higher accuracy of the large-scale structure. The additional accuracy provided by FASTPM can be important given the unprecedented statistical power of the DESI survey. Therefore, the final FASTPM covariance matrix together with other methods – i.e. BAM, EZMOCK, Jackknife (Zhang et al. in preparation), analytical models (Xu et al. 2013; Wadekar & Scoccimarro 2020; Wadekar, Ivanov & Scoccimarro 2020) – are compared to the SLICS reference covariance matrix, in a parallel DESI Mock Challenge paper (Chuang et al. in preparation).

Fundamentally similar to full N-body simulations, FASTPM evolves the dark matter field into the cosmic web, the skeleton of the large-scale structure in the Universe (e.g. Mo, van den Bosch & White 2010; Wechsler & Tinker 2018). After the dark matter haloes are selected, one must implement galaxy–halo connection models (Wechsler & Tinker 2018) to assign galaxies. There are more empirically inspired models such as the halo occupation distribution (HOD; e.g. Benson et al. 2000; Peacock & Smith 2000; Seljak 2000; White, Hernquist & Springel 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002) and subhalo abundance matching (e.g. Kravtsov et al. 2004; Tasitsiomi et al. 2004; Vale & Ostriker 2004) and more physically inspired ones such as full hydrodynamical simulations (e.g. Schaye et al. 2010, 2015; Dubois et al. 2014; McCarthy et al. 2017; Pillepich et al. 2018; Davé et al. 2019) or semi-analytical models (e.g. Guo et al. 2011; Gonzalez-Perez et al. 2014). In this case, we adopt an HOD model as it is one the most efficient ways to create mock galaxy catalogues.

Comparisons between the matter power spectra of FASTPM and full N-body references have shown deviations by ≈0.5 per cent at $k = 0.3\,h\,\mathrm{Mpc}^{-1}$ (Feng et al. 2016) and by ≈20 per cent at $k = 1\,h\,\mathrm{Mpc}^{-1}$ (Grove et al. 2022). In addition, Feng et al. (2016) have observed that the FASTPM halo power spectrum can deviate up to 5 per cent around $k \simeq 0.5\,h\,\mathrm{Mpc}^{-1}$. Lastly, they have noticed that the FASTPM less massive haloes are not as well matched to the reference as the higher mass haloes. As a consequence, the purpose of this paper is to show that using an HOD model to assign galaxies on FASTPM halo catalogues, one can overcome the differences between the FASTPM and the full N-body simulations and thus reproduce the reference SLICS galaxy clustering. We note that the vanilla HOD model used in this work may not be sufficient to capture the physics of galaxy formation, as shown in Chaves-Montero, Angulo & Contreras (2023) and Alam, Paranjape & Peacock (2024). Nevertheless, given the aim of this article, we are only concerned with the ability to produce the covariance matrix for two-point galaxy clustering using FASTPM. We defer the study of other observables such as galaxy–galaxy lensing, count-in-cell statistics, and the Voronoi volume function (Paranjape & Alam 2020) to future work.

Furthermore, we compare the impact of different clustering statistics and examine the effects of various scales on the HOD fitting. Finally, we calculate covariance matrices for all the studied scenarios and perform a comparison to understand the influence of the HOD modelling on the parameter uncertainty estimation.

In Section 2, we present the SLICS and FASTPM simulations. The methodology that we follow is detailed in Section 3. We describe our results on the HOD fitting performance and the covariance matrix comparison in Section 4. In the end, Section 5 concludes the article.

2 SIMULATIONS

2.1 Scinet LIght-Cone Simulations

The SLICS (Harnois-Déraps & van Waerbeke 2015; Harnois-Déraps et al. 2018) consist of over 900 N-body mocks based on noise-independent initial conditions. The large number of realizations is exploited to estimate the covariance matrices for weak lensing data (Hildebrandt et al. 2017; Joudaki et al. 2017; Martinet et al. 2018; Harnois-Déraps et al. 2023) and for combinations of weak lensing and foreground clustering data (Brouwer et al. 2018; van Uitert et al. 2018).

The cubic mocks – with $L_{\mathrm{box}} = 505\,\mathrm{Mpc}\,h^{-1}$ – simulate a flat Λ cold dark matter cosmology, described by the cosmology of the WMAP9 + SN + BAO, i.e. $(\Omega_\mathrm{m}, \sigma_8, \Omega_\mathrm{b}, w_0, h, n_\mathrm{s}) = (0.2905, 0.826, 0.0447, -1.0, 0.6898, 0.969)$. They are obtained by running the non-linear double-mesh Poisson solver CUBEP^{3}M (Harnois-Déraps et al. 2013) to gravitationally evolve 1536^3 particles – with a particle mass $m_\mathrm{p} = 2.88 \times 10^9\,M_\odot\,h^{-1}$ – on a 3072^3 grid from $z = 99.0$ up to $z = 0$.

The dark matter haloes have been selected by applying a spherical overdensity halo-finder (Harnois-Déraps et al. 2013). Their mass function follows precisely the Sheth, Mo & Tormen (2001) fitting function, as shown in fig. 2 of Harnois-Déraps et al. (2018). The redshift of the halo catalogues included in this study is $z = 1.041$. Lastly, given that some halo catalogues have been corrupted at the run time, we are limited to only 139 independent mocks.

This study, together with the BAM (Balaguera-Antolínez et al. 2023) and the DESI covariance matrix comparison paper (Chuang et al. in preparation), is focused on the DESI ELGs sample. Therefore, the SLICS haloes have been populated using the HMQ HOD model presented in Alam et al. (2020), capable of describing the ELG galaxies – consult (Alam et al. 2021b) for a comparison between different HOD models for ELGs. In addition, the parameters of the HOD model have been set to match the expected linear bias of the DESI ELG sample. The final product is a set of 139 galaxy catalogues that are used as reference in all the studies mentioned before. More details about the SLICS galaxy catalogues production can be found in the DESI covariance matrix comparison paper (Chuang et al. in preparation).

2.2 FAST PM

Accelerated PM solvers – such as the FASTPM software (Feng et al. 2016) – are able to produce accurate halo populations with respect to the full N-body simulations. Thus, they are suitable to accurately simulate large volumes.

FASTPM uses a pencil domain-decomposition Poisson solver and Fourier-space four-point differential kernel to compute the force. Additionally, the vanilla leap-frog scheme for the time integration is adjusted to account for the acceleration of velocity during a step, allowing for the accurate tracking of the linear growth of large-scale modes regardless of the number of time-steps.

For the current analysis. we have run FASTPM with two resolutions, resulting in one set of 778 Low Resolution boxes (LR; 1296^3 particles) and one set of 141 High Resolution (HR; 1536^3 particles) catalogues. Both sets output snapshots at the same redshift ($z = 1.041$), and have the same box side length ($L_{\rm box} = 505\,{\rm Mpc}\,h^{-1}$) and cosmology as the SLICS simulations. The particle mass of the HR simulations is $2.86444 \times 10^9\ M_\odot\,h^{-1}$, while the one of LR is $4.77 \times 10^9\ M_\odot\,h^{-1}$. The resolution of the force mesh is boosted by a factor of $B = 2$ compared to the number of particles per side, for both LR and HR. Lastly, 40 linear steps have been used to evolve the density field from $a = 0.05$ to $a = 0.96$.

Due to the small number of SLICS galaxy realizations, for 123 runs of the FASTPM (LR and HR likewise), we use the SLICS initial conditions. This plays an important role to reduce the effect of the cosmic variance in the clustering statistics and thus in the HOD fitting. SLICS initial conditions have been built on a 1536^3 regular grid using the Zel'dovich approximation (Zel'dovich 1970) to displace the particles from the grid positions. Lastly, the initial conditions have been downgraded to the LR by cutting in Fourier space the high-frequency modes larger than the Nyquist frequency corresponding to the LR field.

The haloes have been selected from the dark matter field with the Friends-of-Friends halo finder in NBODYKIT (Hand et al. 2018). During the galaxy assignment process – Section 3.3 – we only use haloes with a minimum mass of $5.72 \times 10^{10}\ M_\odot\,h^{-1}$.

Finally, in Section 3, when we mention FASTPM, we imply for simplicity both HR and LR. We only make the distinction in the results section, i.e. Section 4.

3 METHODOLOGY

3.1 Clustering computation

3.1.1 Two-point correlation function

Mathematically, the two-point correlation function (2PCF) is a continuous function that can describe the clustering of galaxies. However, given the discrete nature of the galaxy distribution in the Universe, the 2PCF is measured using discrete estimators. In the case of cubic mocks, one can implement the natural estimator (Peebles & Hauser 1974):

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

where $DD(s, \mu)$ and $RR(s, \mu)$ are the data and the random pair counts, respectively, as functions of the radial distance

$$s = \sqrt{s_\perp^2 + s_\parallel^2}, \tag{2}$$

and the cosine of the angle between $\mathbf{s}$ and the line of sight

$$\mu = \frac{s_\parallel}{s}. \tag{3}$$

In the previous equations, $s_\perp$ and $s_\parallel$ are the perpendicular ($\perp$) and parallel ($\parallel$) to the line-of-sight components of $\mathbf{s}$, respectively. While the DD term is evaluated directly on the data catalogue, RR is calculated theoretically.

In the present analysis, we run PYFCFC[1] the PYTHON wrapper of the Fast Correlation Function Calculator[2] (FCFC Zhao 2023) to estimate the 2PCF. Lastly, we decompose the 2D 2PCF ($\xi(s, \mu)$) into 1D multipoles ($\xi_\ell(s)$) with the help of the Legendre polynomials $L_\ell(\mu)$ of order ℓ, as follows:

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^{1} \xi(s, \mu) L_\ell(\mu) {\rm d}\mu. \tag{4}$$

3.1.2 Power spectrum

From the mathematical point of view, the power spectrum $P(\mathbf{k})$ is the Fourier Transform of the 2PCF. However, the limited volume of a survey or a simulation creates mode coupling and makes the two clustering measurements not completely equivalent. Consequently, we compute the power spectrum multipoles $P_\ell(k)$ starting from the density field in Fourier space $\delta(\mathbf{k})$, as follows (de Mattia et al. 2021):

$$P_\ell(k) = \frac{2\ell + 1}{V} \int \frac{{\rm d}\Omega}{4\pi} \delta(\mathbf{k})\delta(-\mathbf{k}) L_\ell(\hat{\mathbf{k}} \cdot \hat{\eta}) - P_\ell^{\rm noise}, \tag{5}$$

where the unit vector $\hat{\mathbf{k}}$ represents the direction of $\mathbf{k}$ and $\hat{\eta}$ is the global line-of-sight unit vector, chosen as the z-axis of the cubic simulations. Finally, given the finite galaxy number density $\bar{n}_{\rm g}$, the monopole shot-noise is computed as

$$P_0^{\rm noise} = \frac{1}{\bar{n}_{\rm g}}, \tag{6}$$

while for the higher order multipoles, the shot-noise is zero.

In practice, we harness the versatility of POWSPEC[3] described in Zhao et al. (2021) through its PYTHON wrapper[4] to calculate the power spectra and their multipoles starting from the galaxy catalogues. We estimate the density field on a grid of size 512^3, by applying the cloud-in-cell (Sefusatti et al. 2016) particle assignment scheme on the catalogues of galaxies. Lastly, we exploit the grid interlacing technique (Sefusatti et al. 2016) to reduce the alias effects at small scales.

In the current analysis, we show the monopole ($\ell = 0$), quadrupole ($\ell = 2$), and hexadecapole ($\ell = 4$) for both the 2PCF and the power spectrum.

3.1.3 Bi-spectrum

The power spectrum and the 2PCF are two-point clustering statistics, but higher order statistics are necessary to characterize more precisely

[1] https://github.com/dforero0896/pyfcfc
[2] https://github.com/cheng-zhao/FCFC
[3] https://github.com/cheng-zhao/powspec
[4] https://github.com/dforero0896/pypowspec

the galaxy distributions. In this study, we also look at the three-point clustering statistics, namely the bi-spectrum $B(\mathbf{k_1}, \mathbf{k_2}, \mathbf{k_3})$, the Fourier pair of the three-point correlation function (e.g. Bernardeau et al. 2002):

$$\delta^D(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) B(\mathbf{k_1}, \mathbf{k_2}, \mathbf{k_3}) = \langle \delta(\mathbf{k_1}) \delta(\mathbf{k_2}) \delta(\mathbf{k_3}) \rangle. \quad (7)$$

The three vectors $\mathbf{k_1}$, $\mathbf{k_2}$, $\mathbf{k_3}$ are chosen to form a triangle whose two of the three sides are fixed ($k_1 = 0.1 \pm 0.05$ and $k_2 = 0.2 \pm 0.05$), but the angle θ_{12} between $\mathbf{k_1}$ and $\mathbf{k_2}$ is varied from 0 to π. In practice, we compute the monopole of the bi-spectrum on a grid of size 512^3.

3.2 FASTPM HOD model

The galaxy population and its associated clustering covariance matrix can potentially be influenced by halo properties beyond just mass, as shown in Alam et al. (2024). None the less, such effects are expected to be small for large volume surveys such as DESI and hence we plan to address them in future work. Additionally, the FASTPM haloes are less accurate than the ones from an N-body simulation, thus we do not expect that the final HOD model and parameters maintain the same physical interpretation.

As a consequence, we can adopt the simple five-parameter HOD model described in Zheng et al. (2005) to assign galaxies to the FASTPM halo catalogues, as long as the resulting clustering and covariance matrix match the reference. Nevertheless, in future work one can study more complex and more adapted models for the studied ELG sample.

The current model assumes that each halo can host at most one central galaxy with a probability $\mathcal{B}(1) = \langle N_{\rm cen} \rangle (M_{\rm h})$ dependent on the halo mass $M_{\rm h}$, where $\mathcal{B}(x)$ denotes the Bernoulli distribution and

$$\langle N_{\rm cen} \rangle (M_{\rm h}) = \frac{1}{2} \left[1 + {\rm erf} \left(\frac{\log M_{\rm h} - \log M_{\rm min}}{\sigma_{\log M}} \right) \right] \quad (8)$$

with erf the error function:

$$ {\rm erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x {\rm e}^{-u^2} {\rm d}u. \quad (9)$$

$\log M_{\rm min}$ is the halo mass at which the probability to host a central galaxy is one-half and $\sigma_{\log M}$ controls the steepness of the transition from a probability of one to zero. Lastly, the positions and velocities of the central galaxies are precisely the values of their parent haloes.

In contrast, the number of satellite galaxies $n_{\rm sat}$ per halo is sampled from a Poisson distribution $\mathcal{P}(n_{\rm sat} | \langle N_{\rm sat} \rangle (M_{\rm h}))$ with the mean

$$\langle N_{\rm sat} \rangle (M_{\rm h}) = \left(\frac{M_{\rm h} - M_0}{M_1} \right)^{\alpha}, \quad (10)$$

where M_0 is a minimum halo mass threshold below which haloes cannot host satellite galaxies; M_0 and M_1 indicate the halo mass at which one halo hosts on average one satellite galaxy, and α is the power-law index. Furthermore, the positions and velocities of the satellite galaxies follow the Navarro–Frenk–White (NFW; Navarro, Frenk & White 1996) density profile.

In the interest of adjusting the smaller scales and the quadrupole, we introduce a velocity dispersion factor ($v_{\rm disp}$) for the velocity parallel ($\|$) to the line of sight (i.e **oZ** in the current case) of the satellite galaxies, in addition to the five HOD parameters:

$$v_{\|}^{\rm sat,new} = \left(v_{\|}^{\rm sat,old} - v_{\|}^{\rm halo} \right) \times v_{\rm disp} + v_{\|}^{\rm halo}, \quad (11)$$

where $v_{\|}^{\rm halo}$ is the velocity parallel to the line of sight of the satellites' parent halo. Finally, the six free parameters are fitted so that the resulting FASTPM clustering matches the SLICS one.

Table 1. A summary of some of the most important and possibly confusing notations and their meaning.

Notation	Meaning
$N_{\rm mocks}^{\rm cov} = 123$	The number of FASTPM and SLICS pairs that share the same initial conditions. These catalogues have been used to compute $\mathbf{C}_s$, equation (19), part of $\Sigma_{\rm diff}$.
$N_{\rm mocks}^{\rm fit} = 20$	The number of FASTPM and SLICS pairs for which we have computed the clustering during the HOD fitting described in Section 3.3.1 and Section 3.3.2.
$\Sigma_{\rm diag}$	Equation (16): Diagonal matrix used during the first step of the HOD fitting, see Section 3.3.1.
$\sigma_{n_{\rm g}}$	Estimation of the galaxy number density noise used in $\Sigma_{\rm diag}$. Standard deviation of 139 SLICS mocks, divided by $\sqrt{139}$.
$\Sigma_{\rm diff}$	Equation (20): Difference covariance matrix used during the second step of the HOD fitting, see Section 3.3.2.
$\sigma'_{n_{\rm g}}$	Estimation of the galaxy number density noise used in $\Sigma_{\rm diff}$. Standard deviation of 139 SLICS mocks, divided by $\sqrt{N_{\rm mocks}^{\rm fit}}$.
Σ_{χ}	Equation (22): The covariance matrix used to compute the χ_ν^2, equation (21). It is not used for fitting.

Table 2. The limits of the uniform prior distributions included in the HOD fitting. Note that M_0 from equation (10) is $M_0 \equiv \kappa \times M_{\rm min}$. $M_{\odot}$ denotes the solar mass.

Name	$\log \frac{M_{\rm min}}{M_{\odot}}$	$\sigma_{\log M}$	$\log \frac{M_1}{M_{\odot}}$	κ	α	$v_{\rm disp}$
Min	11.6	0.01	9	0	0	0.7
Max	13.6	4.01	14	20	1.3	1.5

3.3 HOD fitting

We would like to draw the attention of the reader to Table 1. It contains a summary of important symbols related to the HOD fitting.

With the aim of finding the best-fitting FASTPM clustering, we run an HOD Optimization Routine (HODOR[5]). It uses the HALOTOOLS (Hearin et al. 2017) package to define and apply the HOD model and PYMULTINEST (Buchner et al. 2014), the PYTHON wrapper of MULTINEST (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Feroz et al. 2019), to sample the six HOD parameters.

MULTINEST is a sampler based on Bayes' theorem that provides the maximum likelihood (best-fitting) parameters, as well as the posterior probability distribution of parameters alongside the Bayesian evidence. Bayes' theorem combines prior knowledge about the Θ parameters of a model M with information from the data D to calculate the posterior probability density of the Θ parameters:

$$p(\Theta | D, M) = \frac{p(D | \Theta, M) p(\Theta | M)}{p(D | M)}, \quad (12)$$

where $p(\Theta|M)$ is the prior distribution of Θ of the model M, $p(D|\Theta, M)$ is the likelihood, and $p(D|M)$ is a normalizing factor called Bayesian evidence.

The uniform prior distributions that we impose on all six parameters are shown in Table 2. Furthermore, we approximate the

[5] https://github.com/Andrei-EPFL/HODOR

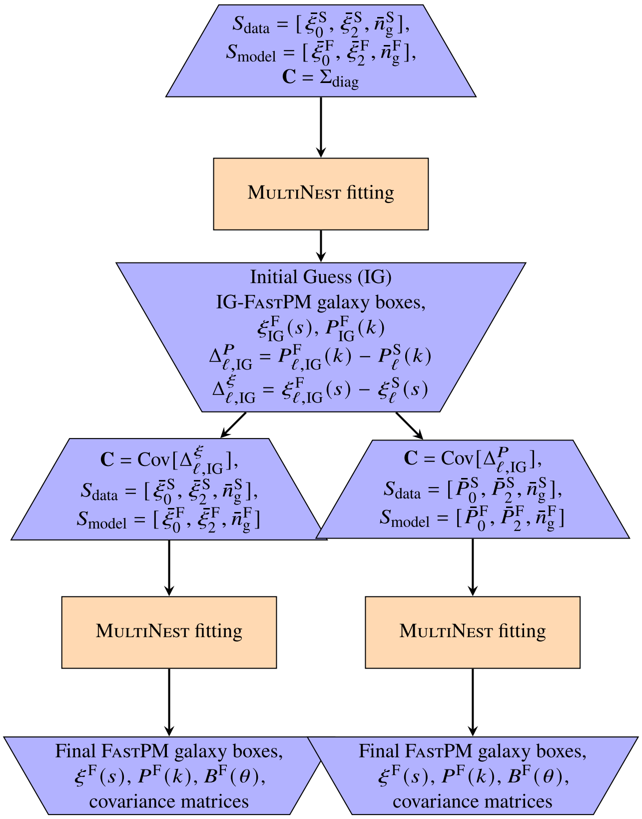

Figure 1. The two-step HOD fitting process that is detailed in Section 3.3.

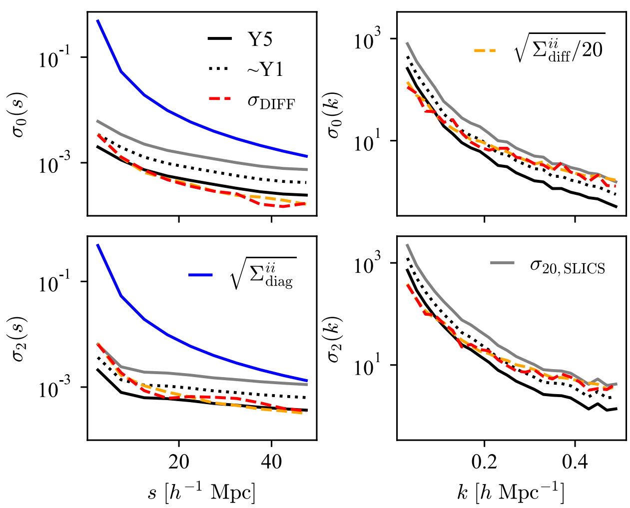

Figure 2. Monopole and quadrupole error bars: left panels – 2PCF; right panels – power spectrum. Black – estimated uncertainty for the entire DESI survey (Y5); dotted black – estimated uncertainty for the Year 1 (Y1) DESI survey; blue – square root of the $\Sigma_{\rm diag}$'s terms; dashed orange – square root of the $\Sigma_{\rm diff}$'s diagonal terms, divided by $\sqrt{20}$, $N^{\rm fit}_{\rm mocks} = 20$; dashed red – standard deviation of the differences between the best-fitting FASTPM clustering (from the second HOD fitting step, one HOD fitting scenario) and SLICS ($N^{\rm fit}_{\rm mocks}$ realizations), further divided by $\sqrt{N^{\rm fit}_{\rm mocks}}$; grey – standard deviation of $N^{\rm fit}_{\rm mocks}$ SLICS clustering realizations, further divided by $\sqrt{N^{\rm fit}_{\rm mocks}}$. We emphasize that the Y5 and Y1 errors are estimated by scaling the SLICS covariance to the effective volumes of 24 and 8 Gpc^3 h^{-3}, respectively.

likelihood by a multivariate Gaussian:

$$p(D|\Theta, M) = \mathcal{L}(\Theta) \sim {\rm e}^{-\chi^2(\Theta)/2}, \tag{13}$$

with the chi-squared:

$$\chi^2(\Theta) = \mathbf{v}^{\rm T}\mathbf{C}^{-1}\mathbf{v}, \tag{14}$$

where $\mathbf{v}$ is the difference between the data and model vectors $\mathbf{v} = S_{\rm data} - S_{\rm model}(\Theta)$, and $\mathbf{C}$ is the covariance matrix.

The purpose of a covariance matrix $\mathbf{C}$ is to estimate the noise in the data, in the context of a noise-free model. Nevertheless, the peculiarity of this study is that both the model ($S_{\rm model}(\Theta)$, FASTPM) and the data ($S_{\rm data}$, SLICS) are affected by noise. Due to the small volume of the SLICS and FASTPM boxes, the cosmic variance component of the noise would be larger than the expected precision of ongoing surveys such as DESI. Therefore, the simulations have been run with matching initial conditions. In this case, the relevant noise component is no longer the cosmic variance but rather the accumulated noise due to gravitational evolution while starting with exactly the same initial conditions. Thus, the mock covariance estimated by SLICS or FASTPM substantially overestimates the error for our fittings, see Fig. 1.

Mathematically, since one computes the difference vector $\mathbf{v} = S_{\rm data} - S_{\rm model}(\Theta)$ to estimate the χ^2, one needs to estimate the covariance matrix of $\mathbf{v}$, i.e. $\mathbf{C}(\mathbf{v})$:

$$\mathbf{C}(\mathbf{v}) = \mathbf{C}(S_{\rm data}) + \mathbf{C}(S_{\rm model}) - \mathbf{C}^\times(S_{\rm data}, S_{\rm model}) - \mathbf{C}^\times(S_{\rm model}, S_{\rm data}), \tag{15}$$

where $\mathbf{C}^\times$ represents the cross covariance matrix (Gubner 2006) between the data and the model vectors. In the more common case of independent data and model vectors and noise-free model, one obtains $\mathbf{C}(\mathbf{v}) = \mathbf{C}(S_{\rm data})$. None the less, in our case, the FASTPM clustering model has positively correlated noise with the SLICS clustering – given the matching initial conditions – hence $\mathbf{C}(S_{\rm model}) \neq 0$ and $\mathbf{C}^\times \neq 0$.

Consequently, in order to more appropriately estimate the noise – i.e. an estimation of $\mathbf{C}(\mathbf{v})$ – we perform a two-step HOD fitting as schematically shown in Fig. 2:

(i) we fit the monopole and quadrupole of the 2PCF [ξ_0, ξ_2] and the galaxy number density $n_{\rm g}$ using a diagonal covariance matrix ($\Sigma_{\rm diag}$) and thus obtain an initial-guess (IG) best-fitting FASTPM galaxy catalogues (IG-FASTPM), see Section 3.3.1;

(ii) we compute the differences [Δ_0, Δ_2] between the clustering (monopole, quadrupole) of the IG best-fitting FASTPM and the SLICS galaxy catalogues; we use these differences to calculate a new covariance matrix ($\Sigma_{\rm diff}$) with which we perform again the fitting, see Section 3.3.2.

In both cases, we use 20 FASTPM (F) and 20 SLICS (S) halo boxes ($N^{\rm fit}_{\rm mocks} = 20$) – sharing the same initial conditions – for the purpose of additionally decreasing the noise. None the less, the average $\bar{n}^{\rm S}_{\rm g}$ is computed using 139 realizations, while the average $\bar{n}^{\rm F}_{\rm g}$ is calculated using the 20 realizations included in the HOD fitting. There are three main reasons behind this discrepancy: first, it quickly becomes expensive to apply galaxies using HOD to more than 20 FASTPM simulations; secondly, the number of SLICS reference simulations has to be the same as for FASTPM, so that the cosmic variance is reduced in the clustering by the shared initial conditions; thirdly, the noise in the galaxy number density is not reduced by the shared initial conditions, thus one needs more realizations to estimate a (practically) noiseless SLICS reference galaxy number density. The galaxy number density is an important constraint as it governs the shot-noise which has a significant role in the covariance matrix.

3.3.1 The first step

Initially, we perform the HOD fitting on the monopole and the quadrupole of the 2PCF, together with the galaxy number density. Hence, the data vector $S_{\rm data}$ is formed by concatenating their respective averages for the SLICS (S) mocks: $S_{\rm data} = [\bar{\xi}_0^{\rm S}, \bar{\xi}_2^{\rm S}, \bar{n}_{\rm g}^{\rm S}]$. Similarly, the model vector $S_{\rm model}$ is determined from the FASTPM (F) boxes: $S_{\rm model} = [\bar{\xi}_0^{\rm F}, \bar{\xi}_2^{\rm F}, \bar{n}_{\rm g}^{\rm F}]$.

Considering that the computing time of clustering measurements scales with the maximum separation, we need a large enough upper limit to constrain relevant parameters, but small enough to keep a reasonable execution time for model evaluation during the HOD fitting. Additionally, since we are interested in capturing the non-linear effects, the lower limit is set to 0. Consequently, the monopole and the quadrupole of the 2PCF are evaluated for $s \in [0, 50]\,{\rm Mpc}\,h^{-1}$, with a bin size of $5\,{\rm Mpc}\,h^{-1}$. Thus, s is an array containing 10 elements $(s_1, \ldots, s_{10})$.

As previously argued, in the first step, there is no appropriate noise estimation. Therefore, we can use an approximate covariance matrix that enables us to proceed to the second step and calculate a more suitable one. In this regard, we create a diagonal covariance matrix:

$$\Sigma_{\rm diag} = \begin{pmatrix} \sigma_1^2 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_{10}^2 & & & & \\ & & & \sigma_1^2 & & & \\ & & & & \ddots & & \\ & & & & & \sigma_{10}^2 & \\ & & & & & & \sigma_{n_{\rm g}}^2 \end{pmatrix}, \quad (16)$$

where the first 20 elements are defined as follows:

$$\sigma_i = \frac{3}{s_i^2},\ i = 1, \ldots, 10. \quad (17)$$

This selection of the diagonal covariance matrix is based on an examination of the $s^2\sigma_{\rm SLICS}(s)$ values, where $\sigma_{\rm SLICS}(s)$ represents the standard deviation of the SLICS 2PCF. Notably, the highest value is approximately three; hence, we initially approximate all values as three for simplicity.

The last element $\sigma_{n_{\rm g}}$ is computed as the standard deviation of 139 SLICS galaxy number densities, divided by $\sqrt{139}$, so that it estimates the uncertainty corresponding to the average of 139 realizations. The strong constraint on the n_g improves the fitting time, as HODOR initially evaluates the goodness-of-fit based only on the $\bar{n}_{\rm g}^{\rm F}$ and $\bar{n}_{\rm g}^{\rm S}$, and does not compute the clustering if $\bar{n}_{\rm g}^{\rm F}$ is 10σ away from the reference. Additionally, the lack of covariance terms in the covariance matrix should, as well, decrease the convergence time.

Finally, we apply the best-fitting HOD model to all $N_{\rm mocks}^{\rm cov} = 123$ FASTPM halo boxes that share the initial conditions with the SLICS mocks to obtain the IG-FASTPM.

3.3.2 The second step

To examine the influence of smaller scales on the HOD fitting, we compute the following for both SLICS and FASTPM:

(i) the power spectrum for $k \in [0.02,\ k_{\rm max}]\,h\,{\rm Mpc}^{-1}$, with a bin size of $0.02\,h\,{\rm Mpc}^{-1}$,

(ii) the 2PCF for $s \in [s_{\rm min},\ 50]\,{\rm Mpc}\,h^{-1}$, with a bin size of $5\,{\rm Mpc}\,h^{-1}$,

Table 3. The fitting ranges for the HOD fitting process described in Section 3.3.2: $k \in [0.02,\ k_{\rm max}]\,h\,{\rm Mpc}^{-1}$ and $s \in [s_{\rm min},\ 50]\,{\rm Mpc}\,h^{-1}$. $N_{\rm bins}^{\ell}$ is the number of bins per multipole ℓ.

Name	Large	Medium	Small
$k_{\rm max}$ [$h\,{\rm Mpc}^{-1}$]	0.5	0.4	0.3
$N_{\rm bins}^{\ell}$	24	19	14
$s_{\rm min}$ [${\rm Mpc}\,h^{-1}$]	0	5	10
$N_{\rm bins}^{\ell}$	10	9	8

where the values of $k_{\rm max}$ and $s_{\rm min}$ are presented in Table 3. Consequently, we create the data and model vectors as follows:

(i) $S_{\rm data} = [\bar{P}_0^{\rm S}, \bar{P}_2^{\rm S}, \bar{n}_{\rm g}^{\rm S}]$ and $S_{\rm model} = [\bar{P}_0^{\rm F}, \bar{P}_2^{\rm F}, \bar{n}_{\rm g}^{\rm F}]$;
(ii) $S_{\rm data} = [\bar{\xi}_0^{\rm S}, \bar{\xi}_2^{\rm S}, \bar{n}_{\rm g}^{\rm S}]$ and $S_{\rm model} = [\bar{\xi}_0^{\rm F}, \bar{\xi}_2^{\rm F}, \bar{n}_{\rm g}^{\rm F}]$.

In order to estimate the noise in the context of shared initial conditions between SLICS and FASTPM, we use the $N_{\rm mocks}^{\rm cov}$ galaxy boxes of both SLICS and IG-FASTPM, along with their corresponding clustering measurements (power spectrum or 2PCF). Furthermore, we introduce $\Delta_{\ell,{\rm IG}}^{P} = P_{\ell,{\rm IG}}^{\rm F}(k) - P_{\ell}^{\rm S}(k)$ and $\Delta_{\ell,{\rm IG}}^{\xi} = \xi_{\ell,{\rm IG}}^{\rm F}(s) - \xi_{\ell}^{\rm S}(s)$ as well as the generic vector $\Delta^{\rm IG}(x) = [\Delta_{0,\,{\rm IG}},\ \Delta_{2,\,{\rm IG}}]$ to express the difference between the SLICS and the IG-FASTPM galaxy clustering that share the initial conditions. Here, the variable x represents either k or s.

Taking advantage of the previous definitions, we further define a matrix $\mathbf{M}$ with the following elements:

$$\mathbf{M}_{ij} = \Delta_i^{\rm IG}(x_j) - \bar{\Delta}^{\rm IG}(x_j),\ i = 1, 2, ..., N_{\rm mocks}^{\rm cov},\ x_j \in [x_{\rm min}, x_{\rm max}], \quad (18)$$

where $\Delta_i^{\rm IG}$ denotes the vector corresponding to the i-th (SLICS, IG-FASTPM) pair, $\bar{\Delta}^{\rm IG}$ represents the mean vector over all (SLICS, IG-FASTPM) pairs, and $[x_{\rm min},\ x_{\rm max}]$ defines the interval of points involved in the fitting, see Table 3. Starting from this matrix and its transpose, we calculate the sample covariance matrix $\mathbf{C}_s$ as follows:

$$\mathbf{C}_s = \frac{1}{N_{\rm mocks}^{\rm cov} - 1}\mathbf{M}^{\rm T}\mathbf{M}. \quad (19)$$

Lastly, we calculate the $\sigma'_{n_{\rm g}}$ as the standard deviation of 139 SLICS galaxy number densities, divided by $\sqrt{N_{\rm mocks}^{\rm fit}}$ – so that it estimates the uncertainty corresponding to the average of $N_{\rm mocks}^{\rm fit}$ realizations – and we attach it to the $\mathbf{C}_s$ to obtain the final covariance matrix used in the HOD fitting:

$$\Sigma_{\rm diff} \equiv \begin{pmatrix} \mathbf{C}_s & 0 \\ 0 & \sigma_{n_{\rm g}}^{\prime 2} \end{pmatrix}. \quad (20)$$

Note that while the error estimate for the clustering is based on the difference in clustering due to matched initial condition, the error of the number density is directly computed from the SLICS realizations, as we aim to constrain the absolute number density, which has strong effect on the final clustering covariance.

3.3.3 Goodness-of-fit

In this section, we define a reduced χ^2 – χ_ν^2 – that expresses the goodness-of-fit for the average of $N_{\rm mocks}^{\rm fit}$ FASTPM galaxy clustering realizations with respect the SLICS reference, i.e. the $n_{\rm g}$ is not

Table 4. The fitting ranges – $k \in [0.02, \mathcal{K}]\,h\,\mathrm{Mpc}^{-1}$ and $s \in [\mathcal{S}, 200]\,\mathrm{Mpc}\,h^{-1}$ used in the clustering fitting described in Section 3.4.

$\mathcal{K}\,[h\,\mathrm{Mpc}^{-1}]$	0.1	0.15	0.2	0.25
$\mathcal{S}\,[\mathrm{Mpc}\,h^{-1}]$	15	20	25	30

included:

$$\chi^2_\nu = N^{\mathrm{fit}}_{\mathrm{mocks}} \times \frac{\Delta^{\mathrm{T}}\Sigma^{-1}_\chi\Delta}{\nu}, \tag{21}$$

where $\mathbf{\Delta}$ denotes the difference between FastPM and SLICS clustering – monopole and quadrupole – and $\nu = N_{\mathrm{bins}} - N_{\mathrm{params}}$, with

(i) $N_{\mathrm{params}} = 6$ – the number of free parameters;
(ii) $N_{\mathrm{bins}} = 2 \times N^{\ell}_{\mathrm{bins}}$ – the length of the $\Delta^{\mathrm{IG}}(x)$ vector, see Table 3.

The Σ^{-1}_χ is the unbiased estimate of the inverse covariance matrix (Hartlap, Simon & Schneider 2007):

$$\Sigma^{-1}_\chi = \mathbf{C}^{-1}_s \frac{N^{\mathrm{cov}}_{\mathrm{mocks}} - N_{\mathrm{bins}} - 2}{N^{\mathrm{cov}}_{\mathrm{mocks}} - 1}, \tag{22}$$

where $\mathbf{C}_s$ is defined in equation (19). Sellentin & Heavens (2016) and Percival et al. (2022) have shown that this correction may not be the optimal choice for accurately determining the uncertainty of the parameters. However, since our main focus is on obtaining the best-fitting clustering and assessing its goodness-of-fit, it remains a reasonable correction.

Finally, as we fit the average of $N^{\mathrm{fit}}_{\mathrm{mocks}}$ realizations, we must scale the covariance matrix $\mathbf{C}_s$ by a factor of $1/N^{\mathrm{fit}}_{\mathrm{mocks}}$. As a consequence, the $N^{\mathrm{fit}}_{\mathrm{mocks}}$ factor appears in equation (21).

3.4 Covariance matrix comparison

Given that the main goal is to have a robust estimation of the uncertainty on the cosmological parameters, we want to compare the constraining power of the covariance matrices. To this end, we fit the 123 individual SLICS clustering (monopole and quadrupole) with the following models:

$$P^{\ell}_{\mathrm{model}}(k) = b_\ell \times \bar{P}^{\ell}_{123,\mathrm{SLICS}}(k) \tag{23}$$

and

$$\xi^{\ell}_{\mathrm{model}}(s) = b_\ell \times \bar{\xi}^{\ell}_{123,\mathrm{SLICS}}(s), \tag{24}$$

where $\bar{P}^{\ell}_{123,\mathrm{SLICS}}(k)$ and $\bar{\xi}^{\ell}_{123,\mathrm{SLICS}}(s)$ are averages of the 123 realizations and b_ℓ denotes the two free parameters.

Moreover, the covariance matrices are computed similarly to the equation (22), but using 778 LR FastPM realizations. The fitting is performed using PyMultiNest, for different fitting ranges ($k \in [0.02, \mathcal{K}]\,h\,\mathrm{Mpc}^{-1}$ and $s \in [\mathcal{S}, 200]\,\mathrm{Mpc}\,h^{-1}$, see Table 4) for the purpose of comparing the effect of the covariance matrices at different scales. The largest fitting intervals are chosen so that they cover the nominal scales included in the BAO and RSD analyses, i.e. $\mathcal{K} \approx 0.2\,h\,\mathrm{Mpc}^{-1}$ and $\mathcal{S} \approx 20\,\mathrm{Mpc}\,h^{-1}$ (e.g. Tamone et al. 2020; de Mattia et al. 2021). Finally, the shown values are the average (b_ℓ) and standard deviation (σ_{b_ℓ}) of the marginalized posterior $p(b_\ell)$ and covariance ($\mathcal{R}[b_0, b_2]$) of the posterior distribution of b_0 and b_2, $p(b_0, b_2)$. By construction, the values of b_ℓ should be one.

The main reason why we perform such a simplified test is to avoid the systematic errors that can arise due to the modelling. Consequently, the comparison between the quoted σ_{b_ℓ} and $\mathcal{R}[b_0, b_2]$ should be directly related to the differences in FastPM covariance matrices. We, nevertheless, reckon that these comparisons do not show how the errors on the parameters of a realistic BAO/RSD model would behave.

4 RESULTS

One of the challenges of HOD fitting is addressing the high precision imposed by large volume surveys such as DESI because it requires prohibitively many large volume simulations. Fig. 1 illustrates this issue as a comparison between $\sigma_{20,\,\mathrm{SLICS}}$,[6] the noise corresponding to the average of $N^{\mathrm{fit}}_{\mathrm{mocks}} = 20$ SLICS clustering realizations, and the expected DESI Y5[7] and Y1[8] errors of the ELG sample. It is obvious that $N^{\mathrm{fit}}_{\mathrm{mocks}}$ SLICS realizations do not reach the required precision.[9]

In order to overcome this issue, we employ the novel matched initial conditions simulations (SLICS and FastPM). In this case, the effect of the cosmic variance on the clustering difference is mostly removed. Therefore, as discussed in Section 3.3, the relevant error estimate is given by the covariance matrix of the clustering difference between the two simulations. Given the fact that we use $N^{\mathrm{fit}}_{\mathrm{mocks}}$ pairs to perform the HOD fitting, the covariance matrix must be rescaled by $N^{\mathrm{fit}}_{\mathrm{mocks}}$. The square root of the diagonal of the resulting covariance matrix is illustrated with an dashed orange line in Fig. 1. One can observe that the matched initial conditions significantly reduce the noise to values below $\sigma_{20,\,\mathrm{SLICS}}$.

Furthermore, we would like to highlight that the precision depicted by the dashed orange line is either better than or equal to DESI Y1 precision up to $k \approx 0.25\,h\,\mathrm{Mpc}^{-1}$. Consequently, the results presented in this paper are precise enough with respect to the requirements of further DESI Y1 analyses. None the less, it might be necessary to readdress this study for the full DESI sample, to account for even lower noise levels. For this, one could use the 1800 AbacusSummit (Maksimova et al. 2021) N-body $0.5\,\mathrm{Gpc}\,h^{-1}$ cubic boxes.

In addition, Fig. 1 illustrates the comparison between σ_{DIFF} and the square root of the diagonal elements of Σ_{diff}. In this context, σ_{DIFF} represents the standard deviation of the differences between the best-fitting FastPM (obtained from the second HOD fitting step) and SLICS clustering, further divided by $\sqrt{N^{\mathrm{fit}}_{\mathrm{mocks}}}$. Ideally, an iterative HOD fitting process should be performed to ensure a robust Σ_{diff}, but the close agreement between σ_{DIFF} and the diagonal elements of Σ_{diff} suggests that Σ_{diff} has approximately converged after a single iteration. A more detailed argument in support of the convergence of Σ_{diff} is presented in Section A.

As pointed out in Section 3.3, it is important that the FastPM galaxy catalogues reproduce the SLICS shot-noise. Examining the FastPM galaxy number densities of all HOD fitting cases, we observed that the largest deviation, $|\bar{n}^{\mathrm{S}}_{\mathrm{g}} - \bar{n}^{\mathrm{F}}_{\mathrm{g}}|/\sigma'_{n_{\mathrm{g}}}$, is approximately $0.5\,\sigma$, but most values are below $0.2\,\sigma$. This strongly supports that the galaxy number density is well constrained and that it is safe to define a χ^2_ν without including n_{g} – see equation (21).

Furthermore, the values of the χ^2_ν are subject to uncertainties due to the finite number of realizations used to estimate the covariance matrix and the limited number of HOD realizations per

[6]This would be the noise level in a hypothetical case where SLICS and FastPM would not share the initial conditions.
[7]The DESI Year 5 error is estimated by rescaling $\sigma_{20,\,\mathrm{SLICS}}$ to match the Y5 ELG sample volume, which is assumed to be $24\,\mathrm{Gpc}^3\,h^{-3}$.
[8]The DESI Year 1 error is estimated by rescaling $\sigma_{20,\,\mathrm{SLICS}}$ to match the Y1 ELG sample volume, which is assumed to be one-third of the Y5 volume.
[9]A simple calculation reveals that one would need 192 SLICS realizations to meet the DESI Y5 precision requirements.

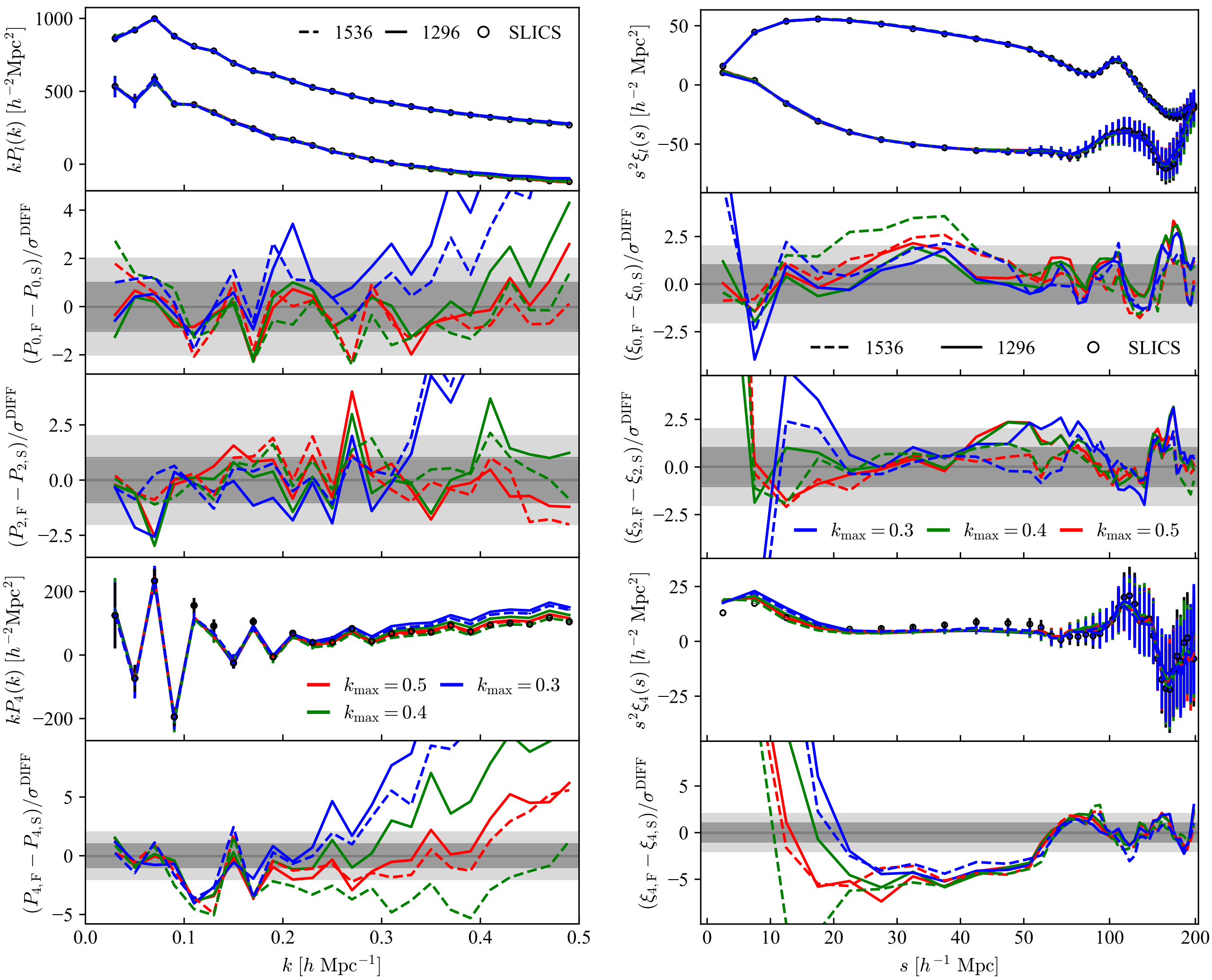

Figure 3. The average of 20 SLICS (reference-black) and 20 FASTPM (model-colours) clustering realizations and the tension (σ_{DIFF} is shown in Fig. 1) between them: left: power spectrum and right: 2PCF. FASTPM mocks share the white-noise through the initial conditions with the SLICS ones. The fitting has been performed: (1) on the monopole and quadrupole of the power spectrum; (2) for three different fitting ranges, see Table 3; (3) using HR (dashed) and LR (continuous) FASTPM realizations. The Ox-axis of the 2PCF panels has a linear scale from 0 to 50 Mpc h^{-1} and a logarithmic scale above this limit.

halo catalogue. The most significant uncertainty, ≈27 per cent, arises from the limited number of HOD realizations. The remaining values are below 20 per cent, see Section B for more details. The χ_ν^2 is simply used as a metric to evaluate the goodness-of-fit. For this reason it is important to consider that it is affected by a large uncertainty when comparing its magnitude to the expected value of one.

The primary focus of this paper is to investigate the limits of the FASTPM capabilities to model the non-linear scales captured by N-body simulations. Furthermore, we study the effect of fitting to successively more non-linear scales and either Fourier or configuration space statistics on the FASTPM covariance matrix.

4.1 Power spectrum fitting

Fig. 3 shows the results of the HOD fitting performed on the power spectrum for three different k intervals, defined in Table 3. The second, third, and fifth rows display the difference in the clustering scaled by the difference error. We remind the reader that this error is smaller than the expected one for the given volume, due to the matched initial conditions between the two simulations, see Fig. 1.

The best-fitting monopoles and quadrupoles are within $\pm 1\sigma$ for most scales. Moreover, the results for the HR FASTPM – presented with dashed line – are only marginally better than the ones for LR FASTPM. Given the modest difference between the performances of the two resolutions, we believe that the LR FASTPM is precise enough to describe the two-point clustering to non-linear scales for a DESI ELG-like galaxy sample, within the estimated DESI Y1 error bars, see Fig. 1.

Considering that we only fit the first two even multipoles, there is no guarantee that the third one would match the reference. Nevertheless, the fifth row of Fig. 3 illustrates that fitting the monopole and quadrupole to smaller scales improves the agreement of the hexadecapole. For instance, fitting on the large interval pushes the $\ell = 4$ multipole within $\pm 2\sigma$ for $k < 0.4\,h\,\mathrm{Mpc}^{-1}$, whereas for medium and small intervals, the hexadecapole is placed within $\pm 2\sigma$ only for $k < 0.3\,h\,\mathrm{Mpc}^{-1}$ or $k < 0.2\,h\,\mathrm{Mpc}^{-1}$, respectively.

Due to the fact that the power spectrum is affected by the window function, it is not obvious that a good matching in Fourier space

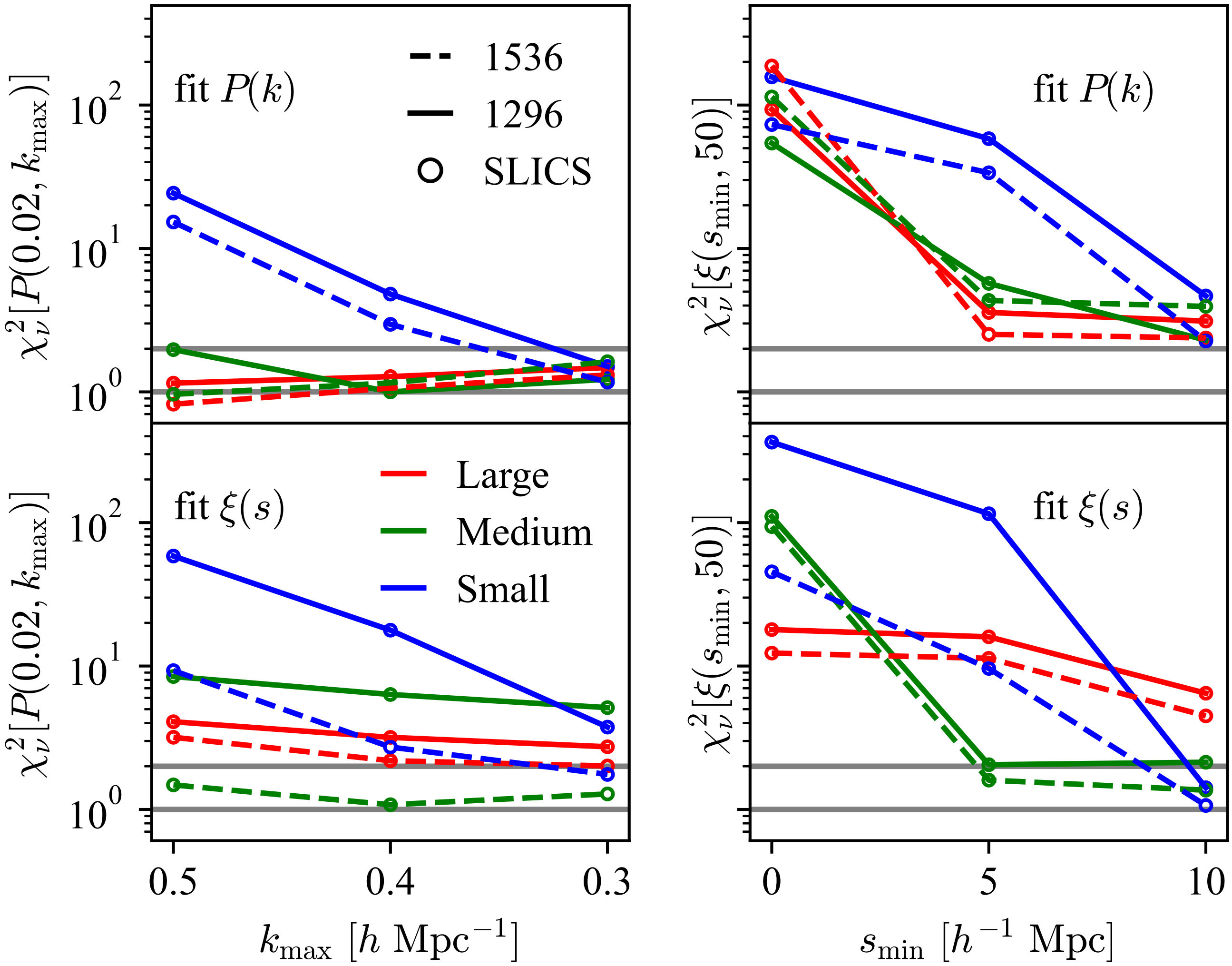

Figure 4. The χ^2_ν as defined in Section 3.3.3. We compute χ^2_ν: (1) for different intervals (see Oy and Ox axes) of the clustering statistics (left panels: power spectrum; right panels: 2PCF); (2) for different fitted clustering (upper panels: power spectrum, see Section 4.1; lower panels: 2PCF, see Section 4.2); (3) for different fitting ranges (see Table 3).

translates as a good matching in Configuration space. Thus, we compute and display the corresponding 2PCF in the right-hand side of Fig. 3. Most monopoles and quadrupoles agree within $\pm 2\sigma$ with SLICS for separations larger than $20\,\mathrm{Mpc}\,h^{-1}$. This suggests that it is possible to obtain a reasonable 2PCF above a certain minimum separation, even when performing HOD fitting on the power spectrum. However, fitting on the medium and large intervals, the 2σ matching goes down to a separation of $10\,\mathrm{Mpc}\,h^{-1}$.

In contrast, for separations smaller than $5\,\mathrm{Mpc}\,h^{-1}$, the non-linear effects become dominant, making it difficult to replicate the velocity field. This is why increasing the fitting range up to $k_{\max} = 0.5$ can improve the monopole but not the quadrupole. Lastly, the 2PCF hexadecapole exhibits a bias of over 3σ for $s < 50\,\mathrm{Mpc}\,h^{-1}$ in all six cases.

After a more qualitative description of the results, we present the χ^2_ν values in the upper panels of Fig. 4. Generally, the HR FASTPM produces lower χ^2_ν values than the LR, as expected from Fig. 3. However, $\chi^2_\nu[P(0.02, k_{\max})] \simeq 1$, which reiterates that by fitting the monopole and quadrupole of the power spectrum up to the three $k_{\max}$ values, one can achieve a good match with the SLICS reference, within the DESI Y1 precision even with LR. In addition, $\chi^2_\nu[\xi(20, 50)] \simeq 2$ for the small fitting interval of the LR power spectrum, reinforcing the fact that one can get a reasonable 2PCF above a certain minimum separation threshold when the fitting is performed on the power spectrum.

Additionally, we can observe the behaviour of χ^2_ν when it is estimated on different intervals than those used for the fitting. When the fitting is performed on the large interval, the $\chi^2_\nu \simeq 1$ for all smaller intervals, regardless of the resolution. However, fitting on the medium interval shows that the difference between HR and LR becomes more significant for $k > 0.4\,h\,\mathrm{Mpc}^{-1}$ (see also Fig. 3): the $\chi^2_\nu \simeq 2$ for LR, while for HR, it is close to one. These findings imply that fitting up to $k \leq 0.4\,h\,\mathrm{Mpc}^{-1}$ is satisfactory for HR FASTPM, whereas smaller scales play a more significant role in LR.

Furthermore, fitting on the small interval shows that although $\chi^2_\nu[P(0.02, 0.3)] \simeq 1$, it is much larger for $k > 0.3\,h\,\mathrm{Mpc}^{-1}$, indicating strong clustering divergence beyond that value (see Fig. 3). Therefore, both LR and HR benefit from considering the clustering information contained in smaller scales $k > 0.3\,h\,\mathrm{Mpc}^{-1}$.

4.2 2PCF fitting

When the HOD fitting is performed on the power spectrum, the minimum 2PCF χ^2_ν is $\chi^2_\nu[\xi(10, 50)] \approx 2$. While this translates to

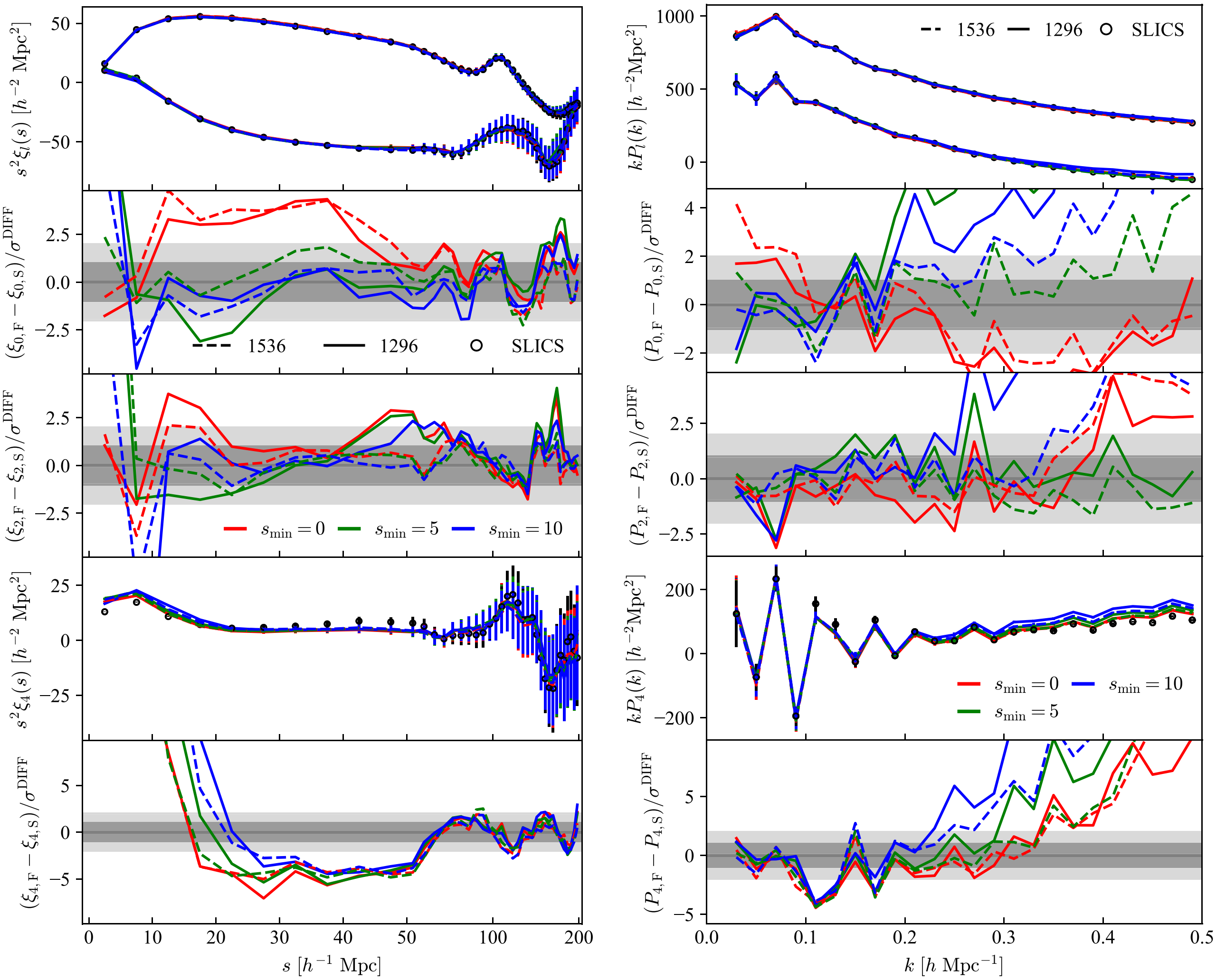

Figure 5. Same as Fig. 3, but the fitting is done on the monopole and quadrupole of the 2PCF.

a 2σ agreement down to the separation of $10\,\mathrm{Mpc}\,h^{-1}$ between FastPM and SLICS 2PCF, we test whether fitting directly the 2PCF can improve the results. Therefore, in this section, we analyse the outcomes of the HOD fitting performed on the 2PCF monopole and quadrupole, for $s \in [s_{\min}, 50]\,\mathrm{Mpc}\,h^{-1}$, see Table 3.

Fig. 5 presents the monopole, quadrupole, and hexadecapole of the 2PCF computed for $s \in [0, 200]\,\mathrm{Mpc}\,h^{-1}$ as well as the tensions between the FastPM and SLICS. The FastPM clustering typically falls within 2σ of the reference for scales larger than $50\,\mathrm{Mpc}\,h^{-1}$ and is largely unaffected by the fitting scenario. However, the HR monopoles are consistently closer to the reference than LR monopoles by approximately 0.5σ at scales larger than $\approx 150\,\mathrm{Mpc}\,h^{-1}$.

Including the smallest scales (large interval) in the HOD fitting, we observe a 1σ to 2σ agreement with the reference for $s < 10\,\mathrm{Mpc}\,h^{-1}$ in both the monopole and quadrupole. However, at intermediate scales $s \in [10, 50]\,\mathrm{Mpc}\,h^{-1}$, the monopole is significantly biased, exhibiting a deviation of 3σ. In contrast, for the medium and small scenarios, we notice that the tensions for the monopole and quadrupole at intermediate scales drop to 1σ, while the smallest scales can get biased by more than 3σ. Nevertheless, they match better the reference than the power spectrum HOD fitting case. Lastly, the hexadecapole does not depend on the resolution or the fitting range and is strongly biased for $s < 60\,\mathrm{Mpc}\,h^{-1}$, showing no improvement compared to the power spectrum fitting.

As in the previous subsection, we test the clustering statistics of the best-fitting FastPM boxes that were not included in the HOD fitting, i.e. the power spectrum in the Fig. 5. The first observation is that these FastPM power spectra do not fit as well the reference as the ones from Fig. 3. On the one hand, for the HR case and medium and small fitting intervals a $\pm 1\sigma$ matching is possible up to $k = 0.4\,h\,\mathrm{Mpc}^{-1}$ and $k = 0.3\,h\,\mathrm{Mpc}^{-1}$, respectively. On the other hand, the LR FastPM allows a good matching up to $k \approx 0.2\,h\,\mathrm{Mpc}^{-1}$ for the same fitting intervals. While the $s_{\min} = 0$ case has a good matching quadrupole up to $k \approx 0.4\,h\,\mathrm{Mpc}^{-1}$, its monopole follows similar trend to the 2PCF monopole, i.e. the intermediate scales $k \in [0.25, 0.4]\,h\,\mathrm{Mpc}^{-1}$ are biased and the rest are mostly within 2σ deviation. Lastly, the hexadecapole is within $\pm 2\sigma$ up to $k \approx 0.3\,h\,\mathrm{Mpc}^{-1}$ for the large fitting interval and up to $k \approx 0.2\,h\,\mathrm{Mpc}^{-1}$ for the other cases.

A quantitative evidence that directly fitting the 2PCF yields superior matching of the 2PCF compared to fitting the power spectrum is displayed in Fig. 4. The majority of the χ^2_ν values in the lower right panel are lower compared to those in the upper right panel. Furthermore, fitting on the small interval ($s_{\min} = 10$), the $\chi^2_\nu \approx 1$, indicating that the 2PCF is in good agreement with the

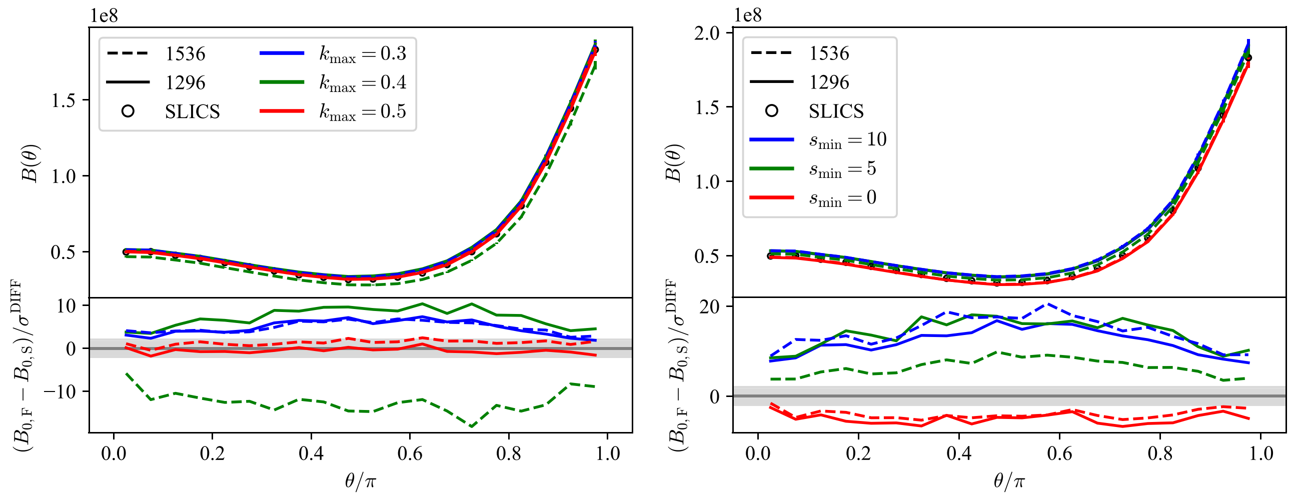

Figure 6. A comparison between the average SLICS bi-spectrum and the average FASTPM bi-spectrum. The averages are computed from the 20 realizations used during the HOD fitting. The left panel shows the results from fitting the power spectrum, as in Fig. 3. The right panel displays the results from fitting the 2PCF, as in Fig. 5. The shaded area denotes $\pm 2\sigma$ deviation. The bi-spectra are computed for $k_1 = 0.1 \pm 0.05\,h\,\mathrm{Mpc}^{-1}$ and $k_2 = 0.2 \pm 0.05\,h\,\mathrm{Mpc}^{-1}$, with the θ angle between $\mathbf{k_1}$ and $\mathbf{k_2}$ varying from 0 to π.

SLICS reference above a certain minimum separation. The almost constant χ^2_ν for the large fitting interval in the lower right panel of Fig. 4 is explained by the discrepancy at the intermediate scales of the monopole for the large fitting interval in Fig. 5. Lastly, as in the previous fitting scenario, the HR FASTPM generally provides a lower χ^2_ν than the LR. In contrast, only the HR simulations can provide a $\chi^2_\nu < 2$ to both the 2PCF and the power spectra, and only when fitting with the medium and small intervals to 2PCF. Although not shown in the aforementioned figure, it is important to note that $\chi^2_\nu[P(0.02, 0.2)] = 2.4$ for the 2PCF small interval LR case.

4.3 Bi-spectrum comparison

The two-point clustering covariance matrix is directly related to the tri-spectrum, i.e. the four-point clustering, however this is difficult to tune. None the less, Baumgarten & Chuang (2018) have shown that for similar two-point clustering, the corresponding covariance matrices – at scales of $s < 40\,\mathrm{Mpc}\,h^{-1}$ – are sensitive to changes in the bi-spectrum. Consequently, even though we do not include the bi-spectrum into the HOD fitting, we aim to understand its behaviour when incorporating various scales of the two-point clustering in the HOD fitting.

Fig. 6 compares the average bi-spectrum of the 20 best-fitting FASTPM boxes with the one computed on the corresponding SLICS boxes, for $2\,k_1 = k_2 = 0.2\,h\,\mathrm{Mpc}^{-1}$ configuration as done by Baumgarten & Chuang (2018). These scales are chosen because they are sensitive to BAO and RSD, the primary scientific case of DESI. It is evident that by increasing the fitting range to include smaller scales, the FASTPM bi-spectrum changes to the extent that for $k_{\max} = 0.5$, the tension ranges from 1σ to 2σ. In contrast, when fitting the 2PCF the resulting bi-spectrum is more biased, i.e. the lowest deviation is $\approx 5\sigma$, for $s_{\min} = 0$ case. In addition, we have checked multiple configurations of the bi-spectrum with $k_i \in [0.01, 0.21]\,h\,\mathrm{Mpc}^{-1}$ and $i = 1$ or 2 and observed that for $k_{\max} = 0.5$ and $s_{\min} = 0$, the overall agreement between the FASTPM and SLICS is around 1 and 5 per cent, respectively. Lastly, there is no significant improvement in terms of the goodness-of-fit between the HR and LR.

In the previous sections, we compare the HR and LR FASTPM with SLICS using the two-point clustering of the 20 cubic mocks included

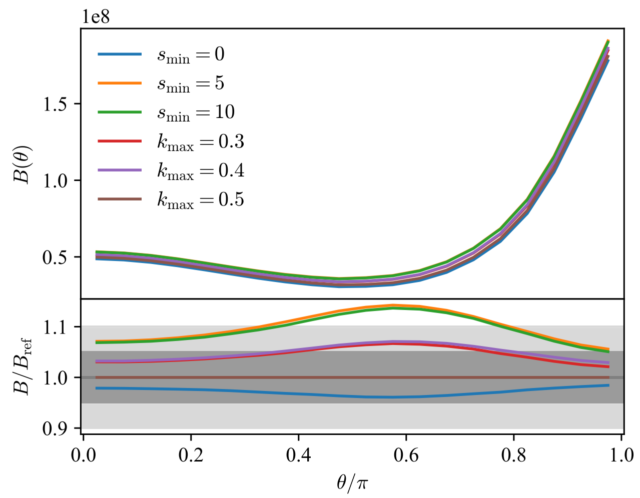

Figure 7. Average bi-spectra computed using 778 LR FASTPM realizations for different HOD fitting cases, see Table 3. The reference for the bi-spectra ratios is the average bi-spectrum computed for the $k_{\max} = 0.5$ case. The bi-spectra are computed for $k_1 = 0.1 \pm 0.05\,h\,\mathrm{Mpc}^{-1}$ and $k_2 = 0.2 \pm 0.05\,h\,\mathrm{Mpc}^{-1}$, with the θ angle between $\mathbf{k_1}$ and $\mathbf{k_2}$ varying from 0 to π.

in the HOD fitting. The HR simulations perform better than LR to model the extremely non-linear scales, such as $k \approx 0.5\,h\,\mathrm{Mpc}^{-1}$ and $s \approx 0\,\mathrm{Mpc}\,h^{-1}$. In contrast, at mildly non-linear scales ($k \approx 0.3\,h\,\mathrm{Mpc}^{-1}$, $s \approx 10\,\mathrm{Mpc}\,h^{-1}$) that are more relevant to BAO and RSD analyses (e.g. Tamone et al. 2020; de Mattia et al. 2021), LR and HR show similar performance. Moreover, Fig. 6 suggests that the bi-spectrum does not depend strongly on the resolution. Nevertheless, the computing cost of HR is significantly higher than for LR and given the small difference in precision, we argue it is optimal to use LR FASTPM for further analyses.

Furthermore, in Fig. 7, we compare the average bi-spectra – computed from 778 LR FASTPM realizations – corresponding to the six HOD fitting cases, see Table 3. In this and the next figures, we choose the $k_{\max} = 0.5$ case as reference because

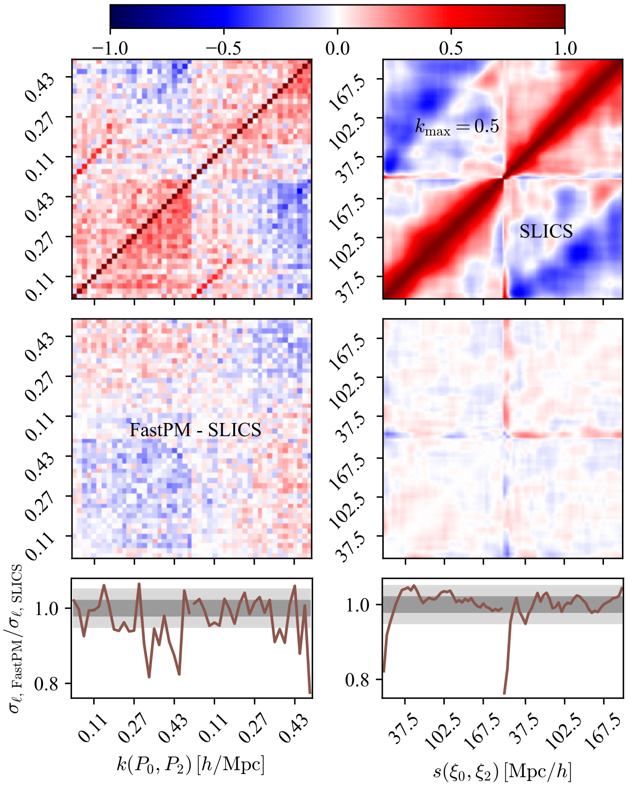

Figure 8. The correlation matrices and the standard deviations computed using 123 realizations of the monopole and quadrupole of the power spectrum (left) and 2PCF (right). In the first row, the upper triangular matrices display the correlation matrices of the FastPM $k_{\max} = 0.5$ case, while the lower triangular ones show the SLICS correlation matrices. The second row of panels contains the differences between the SLICS and FastPM correlation matrices. The third-row panels illustrate the ratios between the standard deviations. The shaded regions denote 2 and 5 per cent limits.

(i) Fig. 4 shows that the best-fitting power spectrum provides $\chi^2_\nu \approx 1$;

(ii) Fig. 6 implies that the corresponding bi-spectrum is the closest to the SLICS reference.

One can notice that $s_{\min} = 0$ bi-spectrum is at most 5 per cent different than the reference, while the rest can reach 15 per cent discrepancies. The $k_{\max} = 0.3$ and $k_{\max} = 0.4$ cases are 1 to 2 per cent different from each other and similarly for $s_{\min} = 5$ and $s_{\min} = 10$.

4.4 Covariance matrix comparison

Having studied the behaviour of the bi-spectra with respect to the scales introduced into the HOD fitting, we now want to understand the resulting impact on the two-point clustering covariance matrices. First, Fig. 8 compares the $k_{\max} = 0.5$ and SLICS correlation matrices and standard deviations. Given the limited number of realizations used to estimate the correlation matrices, it is difficult to assess the importance of the differences. Nevertheless, there seems to be more significant differences at very small scales for the 2PCF and important differences at almost all scales for the power spectrum. Numerically, the standard deviations deviate by more than 7 to 10 per cent from the SLICS case for $s < 10\,\mathrm{Mpc}\,h^{-1}$ and $k > 0.35\,h\,\mathrm{Mpc}^{-1}$. A more detailed comparison is performed by Chuang et al. (in preparation).

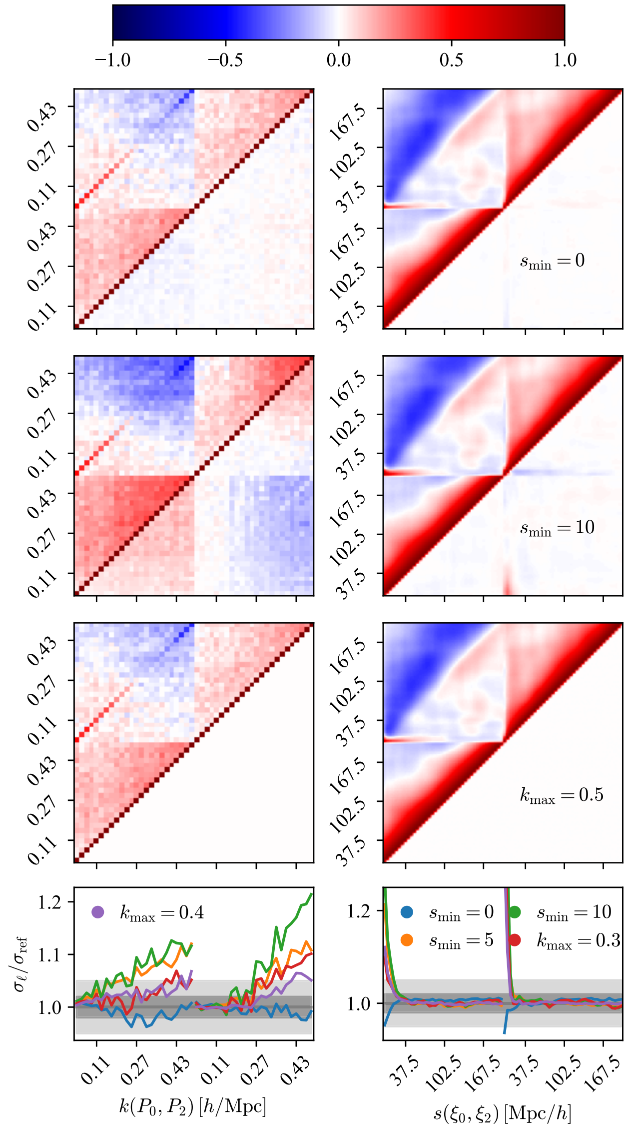

Figure 9. The correlation matrices and the standard deviations computed using 778 realizations of the monopole and quadrupole of the power spectrum (left) and 2PCF (right). Given the similarity between some correlation matrices, only three out of six different HOD fitting cases – see Table 3 – are presented here. However, all cases can be found in Appendix C. The reference case corresponds to $k_{\max} = 0.5$. The upper triangular matrices display the correlation matrices, while the lower triangular ones show the differences between the correlation matrices and the reference one. The bottom panels illustrate the ratios between the standard deviations. The shaded regions denote 2 and 5 per cent limits.

Secondly, we compare the correlation matrices and standard deviations obtained from 778 FastPM realizations, of the six HOD fitting cases, where $k_{\max} = 0.5$ is the reference.

4.4.1 *Power spectrum covariance matrix*

Fig. 9 presents the correlation matrices and the corresponding standard deviations σ_ℓ for the monopole and quadrupole of the power spectrum. The following pairs ($k_{\max} = 0.4$, $k_{\max} = 0.3$), ($s_{\min} = 5$, $s_{\min} = 10$), and ($s_{\min} = 0$, $k_{\max} = 0.5$) have very similar correlation matrices, thus we only show three cases. However, we introduce all of them in Appendix C.

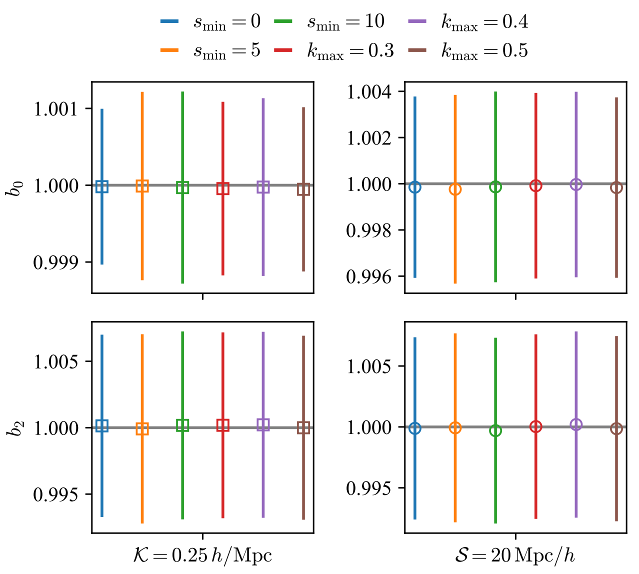

Figure 10. The average of the 123 fitting parameters (b_ℓ) obtained from 123 SLICS clustering realizations, as described in Section 3.4. The measurements performed on the power spectra with $k \in [0.02, \mathcal{K}]\,h\,\mathrm{Mpc}^{-1}$ are shown on the left, while those based on the 2PCF with $s \in [\mathcal{S}, 200]\,\mathrm{Mpc}\,h^{-1}$ are depicted on the right. The error bars are computed as the average of 123 σ_{b_ℓ}, divided by $\sqrt{123}$. Here, σ_{b_ℓ} represents the standard deviation of the b_ℓ posterior distribution. The colours correspond to the different FASTPM covariance matrices illustrated in Fig. 9. Due to the similar results, only one value for $\mathcal{K}$ and $\mathcal{S}$ are shown here. However, all tested values are presented in Appendix C.

The similarity to the reference correlation matrix diminishes in the following order: $s_{\rm min} = 0$, $k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$, $s_{\rm min} = 5$, and $s_{\rm min} = 10$. However, for the largest scales of the quadrupole ($k < 0.15\,h\,\mathrm{Mpc}^{-1}$), the correlation coefficients are practically the same for all cases.

The standard deviations in the lowest panels show similar trends. The $s_{\rm min} = 0$ case is within 2 per cent of the reference case. The $k_{\rm max} = 0.4$ and $k_{\rm max} = 0.3$ cases overestimate the $\sigma_\ell(k)$ by approximately 2 per cent for $k < 0.27\,h\,\mathrm{Mpc}^{-1}$ and by $\approx$5 per cent for smaller scales. Nevertheless, these two cases are consistent with each other within 1 to 2 per cent. In contrast, $s_{\rm min} = 5$ and $s_{\rm min} = 10$ can overestimate the $\sigma_\ell(k)$ by $\approx$2 to 5 per cent for $k < 0.27\,h\,\mathrm{Mpc}^{-1}$ and by 10 to 20 per cent for smaller scales. These two cases are also consistent with each other for most scales, except for the quadrupole $k > 0.3\,h\,\mathrm{Mpc}^{-1}$. These findings are in agreement with the trends observed in the bi-spectrum comparison in Fig. 7.

In order to quantify the differences between the covariance matrices we adopt the method described in Section 3.4 and thus obtain the results displayed in Figs 10 and 11. The first one reveals that none of the six covariance matrices bias the two fitting parameters, regardless of the fitting range. We only present here the results of one fitting range; however all cases can be found in Appendix C.

Examining the uncertainty on b_0 in Fig. 11, we observe that including the smaller scales the discrepancy between the error estimates of the six covariance matrices increases, as we expect from Fig. 9, reaching a maximum of $\approx$20 per cent larger error estimation for the $s_{\rm min} = 10$ covariance at $\mathcal{K} = 0.25\,h\,\mathrm{Mpc}^{-1}$. Moreover, each of the pairs ($k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$), ($s_{\rm min} = 5$, $s_{\rm min} = 10$), and ($s_{\rm min} = 0$, $k_{\rm max} = 0.5$) provide coherent estimations of the uncertainty, which is consistent with the observations on the correlation matrices and standard deviations. Lastly, a 5 per cent consensus between all six covariance matrices is achieved when we fit the power spectra on the $k \in [0.02, 0.1]\,h\,\mathrm{Mpc}^{-1}$.

The agreement between covariance matrices on σ_{b_2} is much better than σ_{b_0}. Given the error bars, the six methods estimate the uncertainty with a 2 per cent tolerance with each other for all $\mathcal{K}$ values.

Finally, all six covariance matrices provide values of $\mathcal{R}[b_0, b_2]$ that are consistent at the level of 5 per cent, given the error bars and up to $\mathcal{K} = 0.2\,h\,\mathrm{Mpc}^{-1}$. For $\mathcal{K} = 0.25\,h\,\mathrm{Mpc}^{-1}$, the largest discrepancy is shown by $s_{\rm min} = 10$ case which underestimates the value of $\mathcal{R}[b_0, b_2]$ by almost 50 per cent. The other cases underestimate $\mathcal{R}[b_0, b_2]$ by 10 to 20 per cent.

4.4.2 2PCF covariance matrix

Comparing the correlation matrices obtained from 778 2PCF in Fig. 9, one can observe that the largest differences occur at the smallest scales $s < 30\,\mathrm{Mpc}\,h^{-1}$. Similarly to the power spectrum correlation matrices, the same pairs of cases show resembling behaviours at all scales. Equivalent qualitative comments can be made about the ratios of the standard deviations. None the less, all cases are within $\approx$2 per cent from each other for $s > 30\,\mathrm{Mpc}\,h^{-1}$, while at smaller scales, the differences can get larger than $\approx$20 per cent.

Following the method described in Section 3.4, we obtain the results shown in Figs 10 and 12. The first figure proves that all six covariance matrices provide unbiased measurements of b_ℓ parameters.

Resembling the power spectrum fitting case, the six estimations of σ_{b_0} in Fig. 12 are in better agreement when the smallest scales are not included in the 2PCF fitting; however for $\mathcal{S} \geq 20\,\mathrm{Mpc}\,h^{-1}$ they are all within $\approx$5 per cent from each other. The largest discrepancy is around 10 per cent and occurs between $s_{\rm min} = 0$ and $s_{\rm min} = 10$ for $\mathcal{S} = 15\,\mathrm{Mpc}\,h^{-1}$. The values of σ_{b_2} are all consistent within $\approx$2 per cent, given the error bars. Interestingly, including the smaller scales, the $\mathcal{R}[b_0, b_2]$ values are more coherent, such that all discrepancies are within 5 per cent, given the error bars and for $\mathcal{S} < 30\,\mathrm{Mpc}\,h^{-1}$. In contrast, when $\mathcal{S} = 30\,\mathrm{Mpc}\,h^{-1}$ the $s_{\rm min} = 10$ and $s_{\rm min} = 5$ provide values $\mathcal{R}[b_0, b_2]$ that are approximately 10 per cent larger than the reference, but nevertheless consistent within the error bars.

The covariance matrix obtained through fitting the power spectrum or the 2PCF can be different on highly non-linear scales, therefore an analysis using such scales needs to be more careful with the covariance matrix estimation. In contrast, in the quasi-linear regime ($s \gtrsim 20\,\mathrm{Mpc}\,h^{-1}$; $k \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$) the differences are not statistically significant (see e.g. $s_{\rm min} = 0$ and $k_{\rm max} = 0.5$ cases), thus one can use either space to obtain the covariance matrices and apply them to both analyses.

5 CONCLUSIONS

We have implemented an HOD model to assign galaxies on the FASTPM halo cubic mocks, such that the resulting clustering – monopole and quadrupole – matches the SLICS reference one. In order to remove the cosmic variance, we have used 20 SLICS galaxy catalogues and 20 halo FASTPM mocks (low-resolution or high-resolution) that share the initial conditions with the SLICS simulations. Given the shared white noise, the standard covariance matrix is obsolete, thus we have performed a two-step HOD fitting:

(i) use a simple diagonal covariance matrix to get IG best-fitting FASTPM galaxy mocks;

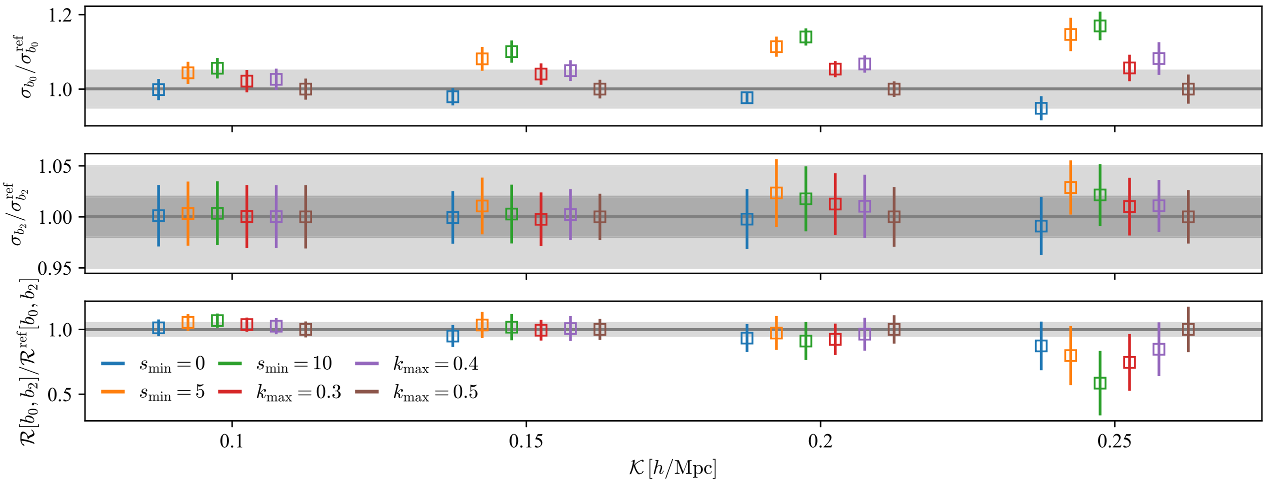

Figure 11. The averages of 123 σ_{b_ℓ} and 123 $\mathcal{R}[b_0, b_2]$ – the standard deviation of the b_ℓ posterior distribution and the covariance between b_0 and b_2, respectively, detailed in Section 3.4 – obtained from 123 SLICS power spectra fitted on $k \in [0.02, \mathcal{K}]\,h\,{\rm Mpc}^{-1}$. In order to estimate the error bars, we split the 778 FastPM realizations in six distinct sets of 123 realizations and compute for each set u a covariance matrix $\Sigma^u_{123,\rm FASTPM}$ with which we fit the 123 SLICS clustering realizations. Having obtained 123 values of $\sigma^u_{b_\ell}$ per set, we compute their average $\bar{\sigma}^u$. Finally, the error bars are the standard deviation of the six $\bar{\sigma}^u$ divided by $\sqrt{6}$. The different colours stand for the different FastPM covariance matrices exhibited in Fig. 9. The $k_{\rm max} = 0.5$ represents the reference, i.e. all values (σ_{b_ℓ} and its error bars) are scaled by the σ_{b_ℓ} corresponding to $k_{\rm max} = 0.5$ case. This is why all brown squares have the value of one. The shaded areas delineate the 2 and 5 per cent regions with respect to the reference.

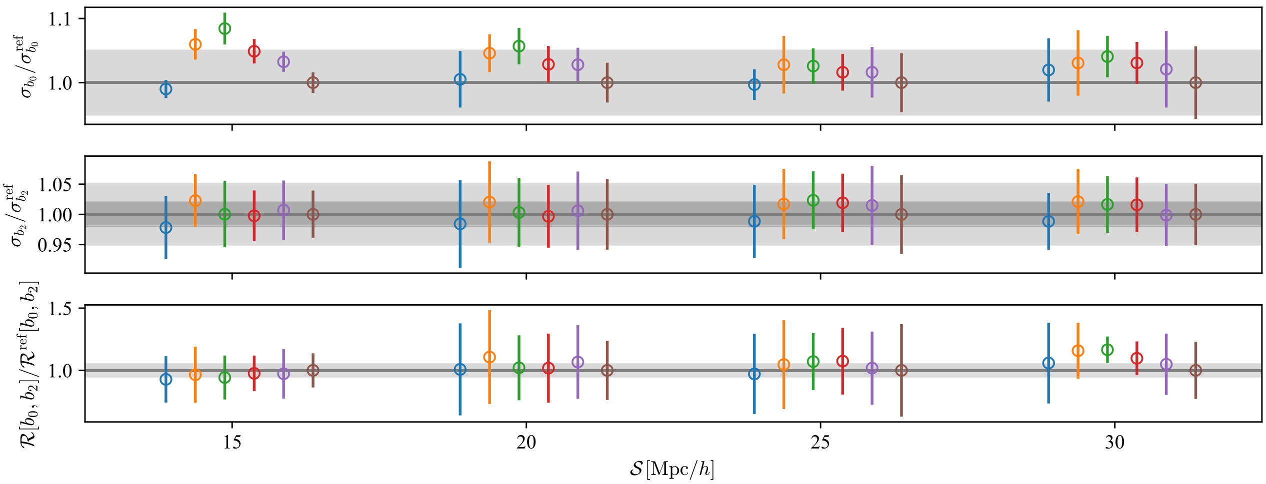

Figure 12. Same as Fig. 11, but the fitting is performed on 123 SLICS 2PCF and $s \in [\mathcal{S}, 200]\,{\rm Mpc}\,h^{-1}$.

(ii) compute the covariance matrix of the 123 realizations of the difference between the IG-FastPM and the SLICS clustering, and use it to perform the final HOD fitting.

The final HOD fitting has been performed on three different fitting ranges for both power spectrum $k \in [0.02, k_{\rm max}]\,h\,{\rm Mpc}^{-1}$ and 2PCF $s \in [s_{\rm min}, 50]\,{\rm Mpc}\,h^{-1}$: ($k_{\rm max} = 0.5$, $k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$) and ($s_{\rm min} = 0$, $s_{\rm min} = 5$, $s_{\rm min} = 10$), respectively.

On the one hand, the HR FastPM generally performs better than the LR at modelling the SLICS clustering. On the other hand, LR is also able to provide a $\chi^2_\nu \approx 1$ for $k_{\rm max} = 0.5$, $k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$, and $s_{\rm min} = 10$. The $k_{\rm max} = 0.5$ case is one of the most valuable as it additionally offers 2σ matching:

(i) power spectrum hexadecapole for $k < 0.4\,h\,{\rm Mpc}^{-1}$;
(ii) 2PCF monopole and quadrupole for $s > 10\,{\rm Mpc}\,h^{-1}$;
(iii) bi-spectrum.

Nevertheless, fitting the 2PCF with $s_{\rm min} = 10$ produces a 1σ matching power spectrum monopole and quadrupole for $k \lesssim 0.2$, but a strongly biased bi-spectrum. In a similar way as the power spectrum, one must include the smallest scales to better reproduce the SLICS bi-spectrum, i.e. for $s_{\rm min} = 0$ the bi-spectrum tension drops from 20σ to 5σ. As a general remark, the power spectrum hexadecapoles can be slightly tuned by changing the values of $k_{\rm max}$ or $s_{\rm min}$, but the 2PCF hexadecapole is practically independent on the fitting range.

The fact that there is no HOD fitting case that provides both power spectrum and 2PCF 1σ matching with the reference might raise the question whether one needs two sets of catalogues for the DESI analysis. Therefore, it could be interesting for future studies to perform a joint fitting of both Fourier and Configuration clustering statistics to test for possible improvements in modelling non-linear scales. Future studies could test also the necessary level of agreement between the FastPM and the full N-body reference in order to have a similar behaviour when observational systematic effects are included. None the less, for the purpose of this article, in the second part of the study, we have exposed and compared the resulting covariance matrices.

First, we have briefly shown with the 123 FastPM realizations – which share the initial conditions with SLICS – that the standard deviations of the $k_{\rm max} = 0.5$ case agree within 5 per cent with the SLICS reference case for $s > 10\,{\rm Mpc}\,h^{-1}$ and $k < 0.35\,h\,{\rm Mpc}^{-1}$. Secondly, we have focused on the 778 LR FastPM realizations corresponding to the six best-fitting cases, where $k_{\rm max} = 0.5$ is considered the reference.

Since Baumgarten & Chuang (2018) have shown that the bi-spectrum computed with a configuration that includes the BAO and RSD effects – i.e. $2 \times k_1 = k_2 = 0.2\,h\,{\rm Mpc}^{-1}$ – affects the covariance matrix, we have compared the six bi-spectra and checked how this comparison reflects into the two-point clustering covariance matrices. Lastly, we have examined the constraining power of these covariance matrices using a simplified clustering model with two scaling parameters, i.e. b_0 and b_2 for the monopole and quadrupole. We have focused on fitting intervals similar to the ones used in standard BAO and RSD analyses i.e. $\mathcal{K} \lesssim 0.20\,h\,{\rm Mpc}^{-1}$ and $\mathcal{S} \gtrsim 20\,{\rm Mpc}\,h^{-1}$ (e.g. Tamone et al. 2020; de Mattia et al. 2021).

The $s_{\rm min} = 0$ bi-spectrum is at most 5 per cent different than the $k_{\rm max} = 0.5$, while the other cases can reach a discrepancy of 15 per cent. However, each of the pairs ($k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$) and ($s_{\rm min} = 5$, $s_{\rm min} = 10$) yields similar bi-spectra. These observations are in a good agreement with a qualitative description of the shown correlation matrices and the standard deviations.

Quantitatively, the power spectrum standard deviations of $s_{\rm min} = 0$ and ($k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$) are within 2 per cent from the reference for $k < 0.5\,h\,{\rm Mpc}^{-1}$ and $k < 0.27\,h\,{\rm Mpc}^{-1}$, respectively. Furthermore, the 2PCF standard deviations of all cases are within 2 per cent from the reference for $s > 30\,{\rm Mpc}\,h^{-1}$.

Using the simplified clustering model, b_0 and b_2 are measured accurately for both power spectrum and 2PCF using all six covariance matrices. The six estimations of σ_{b_0} from the power spectrum fitting up to $\mathcal{K} = 0.20\,h\,{\rm Mpc}^{-1}$ are scattered within at most 20 per cent from the reference, whereas the values of σ_{b_2} are within 2 per cent agreement, given the error bars. Lastly, the covariances between b_0 and b_2 are scattered within 5 per cent from the reference.

In contrast, the estimations of σ_{b_0} from the 2PCF fitting down to $\mathcal{S} = 20\,{\rm Mpc}\,h^{-1}$ are found within 5 per cent from each other. Similarly to the power spectrum case, the σ_{b_2} values agree at the level of 2 per cent. Given the error bars, the covariances between b_0 and b_2 are consistent at the level of 5 per cent.

Lastly, analysing individual cases in the following pairs of HOD fitting cases: ($s_{\rm min} = 0$, $k_{\rm max} = 0.5$), ($k_{\rm max} = 0.4$, $k_{\rm max} = 0.3$), ($s_{\rm min} = 5$, $s_{\rm min} = 10$), one can observe that each pair provides similar uncertainty estimation σ_{b_ℓ}, irrespective of the level of matching between the FastPM and SLICS two-point clustering. In order words, it seems that the estimations of the uncertainties are closer when the bi-spectra – computed for $2 \times k_1 = k_2 = 0.2\,h\,{\rm Mpc}^{-1}$ – are more similar. The most striking case is the 2PCF for $s_{\rm min} = 0$, which is more than 2σ different than the $k_{\rm max} = 0.5$ 2PCF, for $s < 40\,{\rm Mpc}\,h^{-1}$, but the 2PCF standard deviations and the σ_{b_ℓ} estimations are practically identical.

In conclusion, one can use an HOD model on the low-resolution FastPM halo catalogues to tune the galaxy clustering such that it matches the SLICS reference down to certain minimum scales. Additionally, the HOD fitting intervals can have an impact on the final FastPM-based covariances. This influence is observed as a scatter in the uncertainty estimation of up to 20 per cent for power spectrum and 5 per cent for 2PCF at the scales interesting for BAO and RSD analyses. Nevertheless, more accurate analyses could be performed in the future using actual BAO and RSD models and a larger sample of mocks, such as AbacusSummit.

ACKNOWLEDGEMENTS

AV, CZ, and DFS acknowledge support from the Swiss National Science Foundation (SNF) ‘Cosmology with 3D Maps of the Universe’ research grants 200020_175751 and 200020_207379. SA acknowledges support of the Department of Atomic Energy, Government of india, under project no. 12-R&D-TFR-5.02-0200. SA is partially supported by the European Research Council through the COSFORM Research Grant (#670193) and STFC consolidated grant no. RA5496. The authors are thankful for the constructive comments of Joseph DeRose that have helped improve the clarity and quality of our article.

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF’s National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation, the U.S. Department of Energy, or any of the listed funding agencies.

The authors are honoured to be permitted to conduct scientific research on Iolkam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

DATA AVAILABILITY

The SLICS and FastPM boxes used in this study can be provided upon reasonable request to authors. The clustering measurements can be found on Zenodo https://doi.org/10.5281/zenodo.8185822.

REFERENCES

Alam S. et al., 2017, MNRAS, 470, 2617
Alam S., Peacock J. A., Kraljic K., Ross A. J., Comparat J., 2020, MNRAS, 497, 581
Alam S. et al., 2021a, Phys. Rev. D, 103, 083533
Alam S. et al., 2021b, MNRAS, 504, 4667
Alam S., Paranjape A., Peacock J. A., 2024, MNRAS, 527, 3771
Alexander D. M. et al., 2023, AJ, 165, 124
Allende Prieto C. et al., 2020, Res. Notes Am. Astron. Soc., 4, 188
Balaguera-Antolínez A., Kitaura F.-S., Pellejero-Ibáñez M., Zhao C., Abel T., 2019, MNRAS, 483, L58
Balaguera-Antolínez A. et al., 2020, MNRAS, 491, 2565
Balaguera-Antolínez A. et al., 2023, A&A, 673, A130
Baumgarten F., Chuang C.-H., 2018, MNRAS, 480, 2535

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Brodzeller A. et al., 2023, AJ, 166, 66
Brouwer M. M. et al., 2018, MNRAS, 481, 5189
Buchner J. et al., 2014, A&A, 564, A125
Chaussidon E. et al., 2023, ApJ, 944, 107
Chaves-Montero J., Angulo R. E., Contreras S., 2023, MNRAS, 521, 937
Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, MNRAS, 446, 2621
Chuang C.-H. et al., 2019, MNRAS, 487, 48
Cooper A. P. et al., 2023, ApJ, 947, 37
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
de Mattia A. et al., 2021, MNRAS, 501, 5616
DESI Collaboration, 2016a, preprint (arXiv:1611.00036)
DESI Collaboration, 2016b, preprint (arXiv:1611.00037)
DESI Collaboration, 2022, AJ, 164, 207
DESI Collaboration, 2023a, preprint (arXiv:2306.06307)
DESI Collaboration, 2023b, preprint (arXiv:2306.06308)
Dey A. et al., 2019, AJ, 157, 168
Dubois Y. et al., 2014, MNRAS, 444, 1453
Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2019, Open J. Astrophys., 2, 10
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Grove C. et al., 2022, MNRAS, 515, 1854
Gubner J. A., 2006, Probability and Random Processes for Electrical and Computer Engineers. Cambridge Univ. Press, Cambridge
Guo Q. et al., 2011, MNRAS, 413, 101
Guy J. et al., 2023, AJ, 165, 144
Hahn C. et al., 2023, AJ, 165, 253
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Harnois-Déraps J., van Waerbeke L., 2015, MNRAS, 450, 2857
Harnois-Déraps J., Pen U.-L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013, MNRAS, 436, 540
Harnois-Déraps J. et al., 2018, MNRAS, 481, 1337
Harnois-Déraps J., Hernandez-Aguayo C., Cuesta-Lazaro C., Arnold C., Li B., Davies C. T., Cai Y.-C., 2023, MNRAS, 525, 6336
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Hearin A. P. et al., 2017, AJ, 154, 190
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Joudaki S. et al., 2017, MNRAS, 465, 2033
Kitaura F.-S., Yepes G., Prada F., 2013, MNRAS, 439, L21
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B. o., Primack J. R., 2004, ApJ, 609, 35
Lan T.-W. et al., 2023, ApJ, 943, 68
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, MNRAS, 508, 4017
Martinet N. et al., 2018, MNRAS, 474, 712
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
Miller T. N. et al., 2023, preprint (arXiv:2306.06310)
Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution. Cambridge Univ. Press, Cambridge
Myers A. D. et al., 2023, AJ, 165, 50
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Paranjape A., Alam S., 2020, MNRAS, 495, 3233
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peebles P. J. E., Hauser M. G., 1974, ApJS, 28, 19
Pellejero-Ibañez M. et al., 2020, MNRAS, 493, 586
Percival W. J., Friedrich O., Sellentin E., Heavens A., 2022, MNRAS, 510, 3207
Pillepich A. et al., 2018, MNRAS, 473, 4077
Raichoor A. et al., 2020, Res. Notes Am. Astron. Soc., 4, 180
Raichoor A. et al., 2023b, AJ, 165, 126
Ruiz-Macias O. et al., 2020, Res. Notes Am. Astron. Soc., 4, 187
Schaye J. et al., 2010, MNRAS, 402, 1536
Schaye J. et al., 2015, MNRAS, 446, 521
Schlafly E. F. et al., 2023, AJ, 166, 259
Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624
Seljak U., 2000, MNRAS, 318, 203
Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Silber J. H. et al., 2023, AJ, 165, 9
Tamone A. et al., 2020, MNRAS, 499, 5527
Tasitsiomi A., Kravtsov A. V., Wechsler R. H., Primack J. R., 2004, ApJ, 614, 533
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van Uitert E. et al., 2018, MNRAS, 476, 4662
Wadekar D., Scoccimarro R., 2020, Phys. Rev. D, 102, 123517
Wadekar D., Ivanov M. M., Scoccimarro R., 2020, Phys. Rev. D, 102, 123521
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
White M., Hernquist L., Springel V., 2001, ApJ, 550, L129
Xu X., Cuesta A. J., Padmanabhan N., Eisenstein D. J., McBride C. K., 2013, MNRAS, 431, 2834
Yèche C. et al., 2020, Res. Notes Am. Astron. Soc., 4, 179
Zarrouk P. et al., 2021, MNRAS, 503, 2562
Zel'dovich Y. B., 1970, A&A, 5, 84
Zhao C., 2023, A&A, 672, A83
Zhao C. et al., 2021, MNRAS, 503, 1149
Zheng Z. et al., 2005, ApJ, 633, 791
Zhou R. et al., 2020, Res. Notes Am. Astron. Soc., 4, 181
Zhou R. et al., 2023, AJ, 165, 58
Zou H. et al., 2017, PASP, 129, 064101

APPENDIX A: CONVERGENCE TESTS

Given the two-step approach to fit the HOD model, there is an important question of convergence that has to be answered. Thus, we study whether the $\Sigma_{\rm diff}$ estimates well the noise in our HOD fitting process.

Fig. A13 illustrates that the magnitude of the errors estimated after the second HOD fitting step for all six HOD fitting scenarios is consistent between themselves within at most 10 to 20 per cent. However, there seems to be a slight divergence between the error estimations when one approaches the lowest scales.

Finally, compared to the square root of the diagonal of $\Sigma_{\rm diff}$, the six standard deviations for both 2PCF and power spectrum are found within at most 20 per cent deviation. In order to quantify these discrepancies, we have built the difference covariance matrix using each of the six best-fitting FastPM (of the second HOD fitting step, see Table 3) set of galaxies. Furthermore, we have computed the χ^2_ν for the same best-fitting FastPM of the second HOD fitting step, but using these six new difference covariance matrices. The results are summarized in Fig. A14.

Even though for most of the fitting cases, the six new difference covariance matrices seem to provide coherent biased χ^2_ν values compared to the $\Sigma_{\rm diff}$, the biases do not share the same sign between the fitting cases. In addition, most χ^2_ν values are within the error bars shown in Table B1 with respect to the reference. This suggests that given the error bars, the hypothetical best fits obtained using the six new covariance matrices would be indistinguishable from the best-fitting FastPM of the second HOD step. Consequently, we argue that

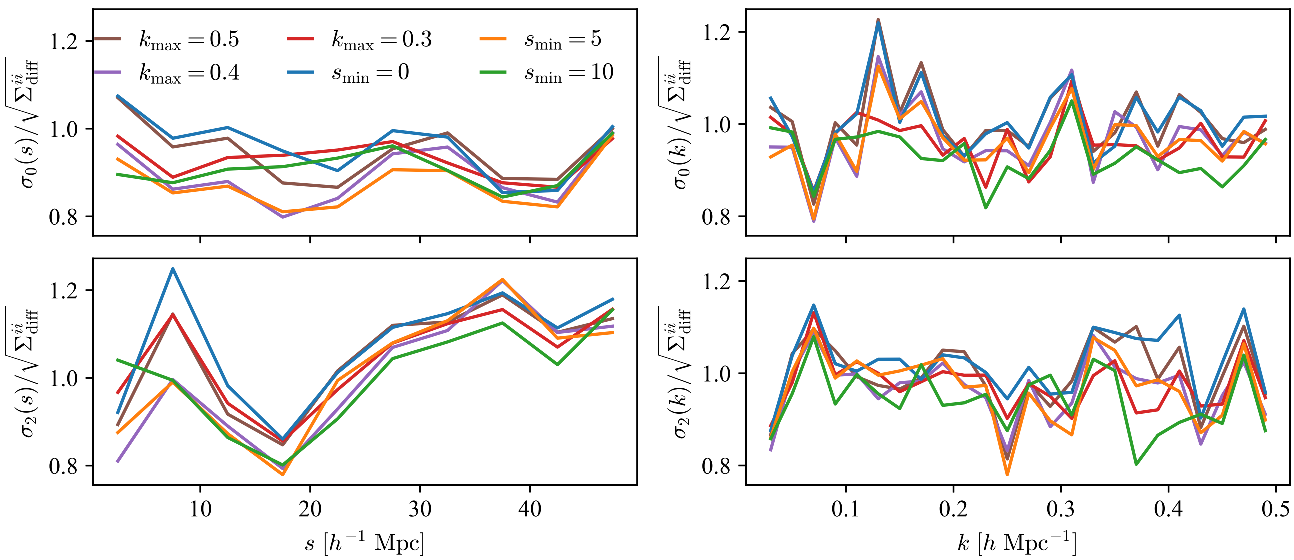

Figure A13. The ratios between the standard deviations computed on the differences of $N^{\rm cov}_{\rm mocks} = 123$ (LR FASTPM, SLICS) clustering pairs and the square root of the diagonal of $\Sigma_{\rm diff}$, i.e. $\Sigma^{ii}_{\rm diff}$. The colours denote the HOD fitting scenarios in the second HOD fitting step, see Table 3. While the left panels include monopole and quadrupole of the 2PCF, the right ones display the power spectrum.

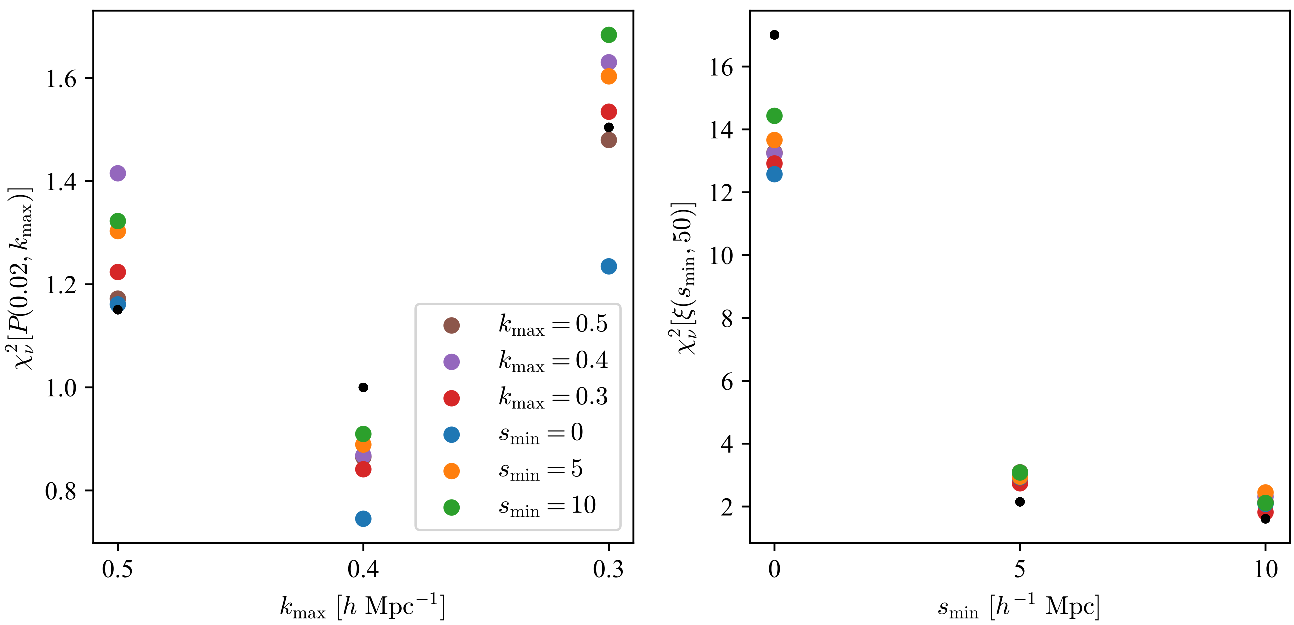

Figure A14. The χ^2_ν as defined in Section 3.3.3, but using different covariance matrices. We compute χ^2_ν: (1) for the six best-fitting FASTPM cases, three cases for the power spectrum in the left panel (see Section 4.1), and three cases for the 2PCF in the right panel (see Section 4.2); (2) with the six difference covariance matrices obtained after the second HOD fitting step (the coloured dots). The black points show the best-fitting χ^2_ν for the six cases that also appear in Fig. 4.

the $\Sigma_{\rm diff}$ is a good approximation of the noise in the difference of the (FASTPM, SLICS) clustering, thus a hypothetical third step HOD would not drastically change the best-fitting FASTPM compared to the ones after the second step.

APPENDIX B: UNCERTAINTY OF THE GOODNESS-OF-FIT

In this section, we are studying the uncertainty introduced by the covariance matrix and the finite number of HOD realizations per FASTPM halo catalogue in the values of χ^2_ν, as defined in equation (21). The results are summarized in Table B1.

B1 Covariance matrix induced uncertainty

Due to the fact that we have only $N^{\rm cov}_{\rm mocks} = 123$ SLICS and FASTPM realizations that share the same initial conditions, we are bound to use the JackKnife (JK) method to estimate the uncertainty introduced by the covariance matrix. Additionally, the HOD fitting is computationally expensive (≈6000 CPU-hours), thus we are not able to perform hundreds of HOD fittings with different covariance matrices. Consequently, after obtaining one set of best-fitting HOD parameters, we computed the χ^2_ν with the same best-fitting FASTPM clustering, but with $N^{\rm cov}_{\rm mocks}$ different covariance matrices.

Table B1. The values of the χ^2_ν and their uncertainties introduced by the covariance matrix (Eqs. B1 and B2) and the finite number of HOD realizations per FASTPM halo catalogue (Eqs. B3 and B4). The estimations have been performed on the LR (1296^3) FASTPM galaxy catalogues and on both the 2PCF and the power spectrum for the three specific fitting ranges defined in Table 3.

	$P(k)$	$\xi(s)$
Large		
χ^2_ν from Fig. 4	1.15	16.94
$\bar{\chi}^2_{\nu,\mathrm{HOD}} \pm \sigma_{\chi,\mathrm{HOD}}$	1.38 ± 0.26	17.03 ± 1.65
$\bar{\chi}^2_{\nu,\mathrm{JK}} \pm \sigma_{\chi,\mathrm{JK}}$	1.16 ± 0.31	16.97 ± 2.68
Medium		
	1.00	2.16
	1.13 ± 0.23	2.19 ± 0.42
	1.00 ± 0.19	2.16 ± 0.32
Small		
	1.50	1.59
	1.50 ± 0.29	1.68 ± 0.41
	1.51 ± 0.25	1.59 ± 0.26

The covariance matrices – Σ^i_χ, with i from 1 to $N^{\mathrm{cov}}_{\mathrm{mocks}}$ – are estimated using equation (22), but with only $N^{\mathrm{cov}}_{\mathrm{mocks}} - 1$ clustering realizations. Furthermore, we compute $\chi^{2,i}_{\nu,\mathrm{JK}}$ for each Σ^i_χ, as defined in equation (21) and we calculate the mean $\bar{\chi}^2_{\nu,\mathrm{JK}}$ and the variance $\sigma^2_{\chi,\mathrm{JK}}$:

$$\bar{\chi}^2_{\nu,\mathrm{JK}} = \frac{1}{N^{\mathrm{cov}}_{\mathrm{mocks}}} \sum_{i=1}^{N^{\mathrm{cov}}_{\mathrm{mocks}}} \chi^{2,i}_{\nu,\mathrm{JK}}, \tag{B1}$$

$$\sigma^2_{\chi,\mathrm{JK}} = \left[\frac{N^{\mathrm{cov}}_{\mathrm{mocks}} - 1}{N^{\mathrm{cov}}_{\mathrm{mocks}}} \sum_{i=1}^{N^{\mathrm{cov}}_{\mathrm{mocks}}} \left(\chi^{2,i}_{\nu,\mathrm{JK}} - \bar{\chi}^2_{\nu,\mathrm{JK}} \right)^2 \right]. \tag{B2}$$

B2 HOD induced uncertainty

During the HOD fitting, for each FASTPM halo catalogue we create a single galaxy realization, in order to reduce the optimization time. As a consequence, we introduce additional noise in the HOD fitting process, which is not considered in the covariance matrix.

With the aim of estimating the effect of this noise on the χ^2_ν, we compute 100 galaxy realizations for a given set of best-fitting HOD parameters and per FASTPM halo catalogue. Furthermore, using the 20 galaxy realizations corresponding to the halo catalogues used in the HOD fitting process and the same covariance matrix, we compute $\chi^{2,i}_{\nu,\mathrm{HOD}}$ as in equation (21), where $i = 1,..., 100$. Finally, we calculate the mean and the standard deviation of the 100 $\chi^{2,i}_{\nu,\mathrm{HOD}}$ values:

$$\bar{\chi}^2_{\nu,\mathrm{HOD}} = \frac{1}{100} \sum_{i=1}^{100} \chi^{2,i}_{\nu,\mathrm{HOD}}, \tag{B3}$$

$$\sigma^2_{\chi,\mathrm{HOD}} = \left[\frac{1}{100 - 1} \sum_{i=1}^{100} \left(\chi^{2,i}_{\nu,\mathrm{HOD}} - \bar{\chi}^2_{\nu,\mathrm{HOD}} \right)^2 \right]. \tag{B4}$$

APPENDIX C: COVARIANCE MATRIX COMPARISON

Analysing Figs C15 and C16 one can observe that the ($s_{\mathrm{min}} = 5$, $s_{\mathrm{min}} = 10$) and ($k_{\mathrm{max}} = 0.3$, $k_{\mathrm{max}} = 0.4$) pairs have very similar correlation matrices. Consequently, we only show $k_{\mathrm{max}} = 0.3$ and $s_{\mathrm{min}} = 10$ in the main text.

Figs C17 and C18 show the results of fitting the clustering with the simplified model detailed in Section 3.4. Since most results are consistent with the expected value of one, we only display the values for $\mathcal{K} = 0.25\,h\,\mathrm{Mpc}^{-1}$ and $\mathcal{S} = 20\,\mathrm{Mpc}\,h^{-1}$ in the main text.

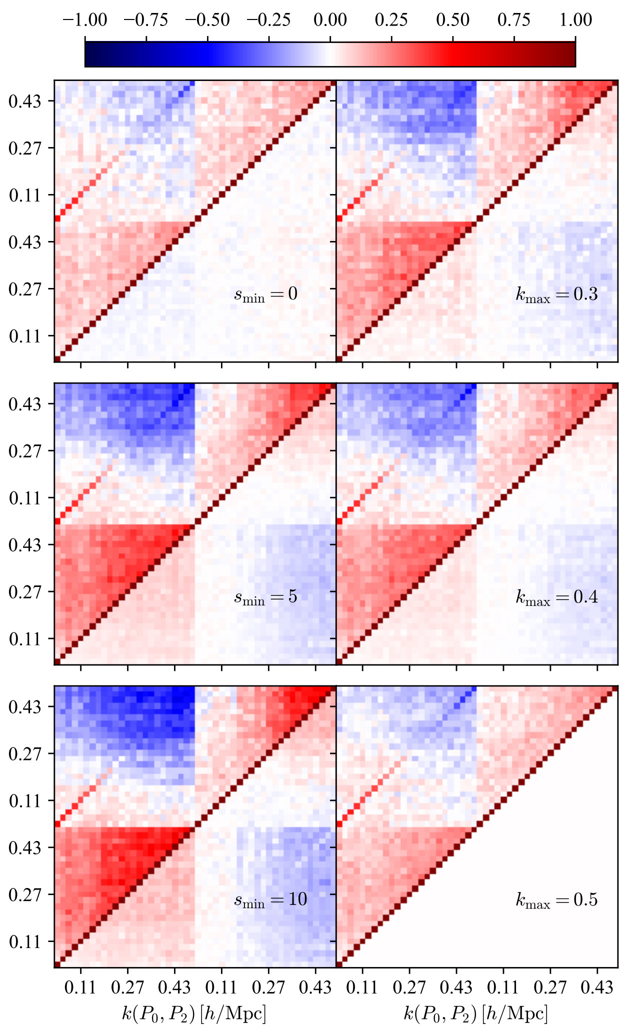

Figure C15. Upper triangular matrices: correlation matrices of the power spectrum monopole and quadrupole, for the fitting cases defined in Table 3. Lower triangular matrices: the difference between the shown correlation matrix and the reference one, i.e. $k_{\mathrm{max}} = 0.5$.

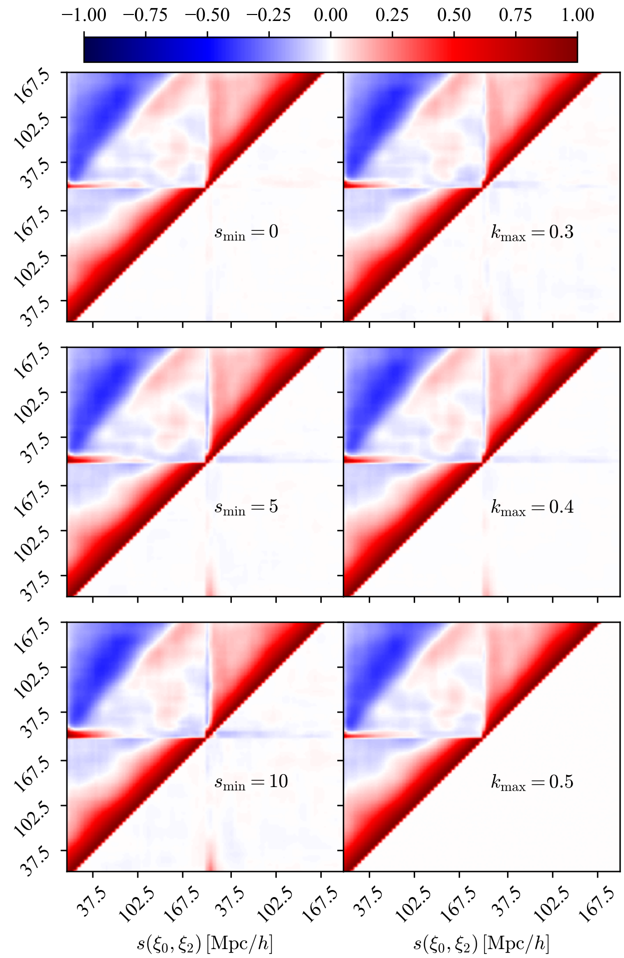

Figure C16. Same as Fig. C15, but for 2PCF.

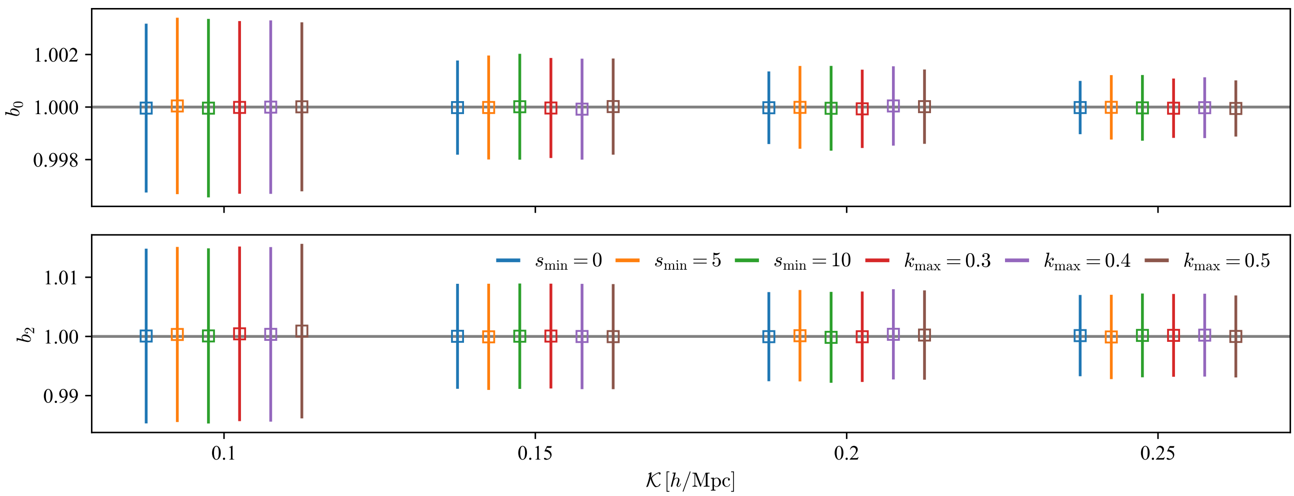

Figure C17. The average of the 123 fitting parameters (b_ℓ) detailed in Section 3.4, obtained from 123 SLICS power spectra fitted on $k \in [0.02, \mathcal{K}]\, h\,\mathrm{Mpc}^{-1}$. The error bars are computed as the average of 123 σ_{b_ℓ}, divided by $\sqrt{123}$, where σ_{b_ℓ} is the standard deviation of the b_ℓ posterior distribution. The different colours stand for the different FASTPM covariance matrices exhibited in Fig. C15.

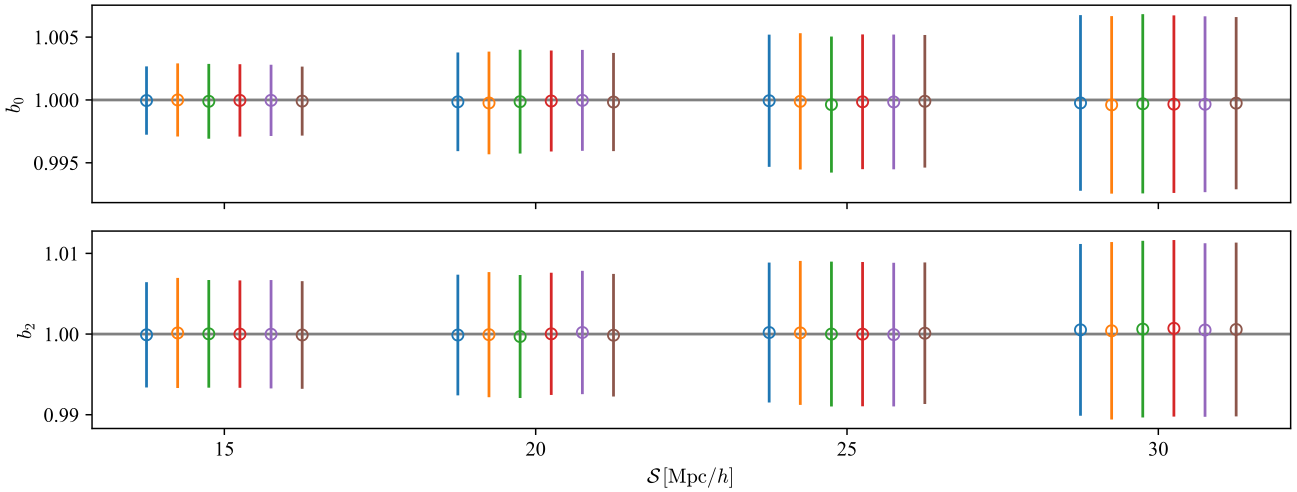

Figure C18. Same as Fig. C17, but the fitting is performed on 123 SLICS 2PCF and $s \in [\mathcal{S}, 200]$ using the covariance matrices exhibited in Fig. C16.

[1] *Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[2] *Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[3] *Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[4] *Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA*
[5] *Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[6] *Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China*
[7] *Key Laboratory for Particle Astrophysics and Cosmology(MOE)/Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai, 200240, China*
[8] *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[9] *Physics Department, Boston University, 590 Commonwealth Avenue, Boston, MA 02215, USA*
[10] *Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[11] *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[12] *Instituto de Física, Universidad Nacional Autónoma de México, Cd. de México C.P. 04510, México*
[13] *Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711 Bogotá, Colombia*
[14] *Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia*
[15] *Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[16] *Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[17] *Departament de Física, Serra Húnter, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain*
[18] *Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[19] *Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, E-08010 Barcelona, Spain*
[20] *National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Rd, Chaoyang District, Beijing 100012, P.R. China*
[21] *Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[22] *Perimeter Institute for Theoretical Physics, 31 Caroline St North, Waterloo, ON N2L 2Y5, Canada*
[23] *Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[24] *Space Sciences Laboratory, University of California, Berkeley, 7 Gauss Way, Berkeley, CA 94720, USA*
[25] *Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*
[26] *Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea*
[27] *CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain*
[28] *Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[29] *Department of Physics & Astronomy, Ohio University, Athens, OH 45701, USA*

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