The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: exploring the halo occupation distribution model for emission line galaxies

S. Avila [ORCID],[1,2,3]★ V. Gonzalez-Perez [ORCID],[3,4]★ F. G. Mohammad,[5,6] A. de Mattia,[7] C. Zhao,[8] A. Raichoor,[8] A. Tamone,[8] S. Alam,[9] J. Bautista [ORCID],[3] D. Bianchi,[10,11] E. Burtin,[7] M. J. Chapman,[5,6] C.-H. Chuang [ORCID],[12] J. Comparat,[13] K. Dawson,[14] T. Divers,[3] H. du Mas des Bourboux,[14] H. Gil-Marin [ORCID],[10,11] E. M. Mueller,[15] S. Habib,[16] K. Heitmann,[16] V. Ruhlmann-Kleider,[7] N. Padilla,[17] W. J. Percival,[5,6,18] A. J. Ross [ORCID],[19] H. J. Seo,[19] D. P. Schneider[20] and G. Zhao[3,21]

Affiliations are listed at the end of the paper

ABSTRACT

We study the modelling of the halo occupation distribution (HOD) for the eBOSS DR16 emission line galaxies (ELGs). Motivated by previous theoretical and observational studies, we consider different physical effects that can change how ELGs populate haloes. We explore the shape of the average HOD, the fraction of satellite galaxies, their probability distribution function (PDF), and their density and velocity profiles. Our baseline HOD shape was fitted to a semi-analytical model of galaxy formation and evolution, with a decaying occupation of central ELGs at high halo masses. We consider Poisson and sub/super-Poissonian PDFs for satellite assignment. We model both Navarro–Frenk–White and particle profiles for satellite positions, also allowing for decreased concentrations. We model velocities with the virial theorem and particle velocity distributions. Additionally, we introduce a velocity bias and a net infall velocity. We study how these choices impact the clustering statistics while keeping the number density and bias fixed to that from eBOSS ELGs. The projected correlation function, w_p, captures most of the effects from the PDF and satellites profile. The quadrupole, ξ_2, captures most of the effects coming from the velocity profile. We find that the impact of the mean HOD shape is subdominant relative to the rest of choices. We fit the clustering of the eBOSS DR16 ELG data under different combinations of the above assumptions. The catalogues presented here have been analysed in companion papers, showing that eBOSS RSD+BAO measurements are insensitive to the details of galaxy physics considered here. These catalogues are made publicly available.

Key words: methods: analytical – methods: numerical – galaxies: evolution – galaxies: haloes – cosmology: theory – large-scale structure of the Universe.

1 INTRODUCTION

The large-scale structure of the Universe contains a wealth of information on cosmology. The spectroscopic galaxy surveys via studies of galaxy clustering have measured the scale of baryonic acoustic oscillations (BAO), redshift space distortions (RSD), and constrained Primordial Non-Gaussianities among other cosmological probes. Although previous wide angle surveys such as BOSS have targeted luminous red galaxies (LRG) at low redshift ($z < 0.6$) achieving unprecedented constrains on cosmology from the large-scale structure (Alam et al. 2017), this type of galaxies becomes harder to target at higher redshifts. That is why recently the focus has turned to new targets such as the star-forming emission line galaxies (ELGs) that can be targeted at higher redshifts, such as eBOSS $0.6 < z < 1.1$ (Dawson et al. 2016) and DESI $0.6 < z < 1.6$ (DESI Collaboration 2016), or observed via slit-less spectroscopy (Euclid, $0.9 < z < 1.8$ Laureijs et al. 2011).

The eBOSS survey, part of the SDSS-IV program (Blanton et al. 2017), has created the largest spectroscopic sample of star-forming ELGs to date with a final sample of 173,736 ELGs in the redshift range $0.6 < z < 1.1$ (Raichoord et al. 2020). This was achieved after measuring the spectra of ELG targets selected from the DECaLS photometric survey. Additionally, eBOSS has surveyed close to half a million of LRGs in the range $0.6 < z < 1.0$ and $\sim 330\,000$ quasars in the range $0.8 < z < 2$ (Lyke et al. 2020; Ross et al. 2020).

N-body simulations play an important role in the large-scale structure analysis in order to validate theoretical tools used for data analysis. One unknown is the way a certain type of galaxies relates to the underlying dark matter distribution. One way to explore this relation is with semi-analytical models (SAMs) of galaxy formation and evolution. However, these models require dark matter halo merger trees and high-mass resolution, rarely available in simulations able to probe beyond the (Gpc/h)^3-scale volume. For the very large-scale simulations, typically, dark matter haloes are populated with halo occupation distribution (HOD) models (e.g. Seljak 2000;

Berlind & Weinberg 2002; Cooray & Sheth 2002; Zehavi et al. 2005; Zheng et al. 2005) or, alternatively, with (sub) halo abundance matching techniques (e.g. Favole et al. 2016). In the original HOD models, galaxy properties are determined from the halo mass of the host. More sophisticated HOD models try to encapsulate the assembly bias, i.e. dependence of halo clustering on properties other than halo mass by introducing secondary parameters (e.g. Hearin et al. 2016; Zehavi et al. 2018). We defer an exploration of the assembly bias for ELGs for future studies.

In this paper, we applied a series of ELG HOD models motivated from previous theoretical studies to produce galaxy catalogues based on the OUTER RIM simulation dark matter-only only simulation. We then compare and fit the clustering statistics of these mocks to ELG data from eBOSS.

The produced mocks are used in companion papers (Alam et al. 2020; de Mattia et al. 2020; Raichoord et al. 2020; Tamone et al. 2020) to test the robustness of theoretical models of galaxy anisotropic clustering against variations in the HOD model, finding that those models can be trusted at least (with a conservative budget computation) to within 1.8 per cent, 1.5 per cent, 3.3 per cent for, respectively, $\{\alpha_\|,\ \alpha_\perp,\ f\sigma_8\}$ (Alcock–Paczynski and growth rate parameters), well below the statistical errors for eBOSS.

Our baseline model takes the shape of the average HOD ($\langle N(M) \rangle$) presented in Gonzalez-Perez et al. (2018) for ELGs, which we approximate by a stepwise Gaussian plus a decaying power law for central galaxies, whereas we model satellites following a power law above a certain halo mass. We complete the baseline model with the following usual assumptions in the generation of galaxy mock catalogues regarding the assignment of satellites (e.g. Carretero et al. 2015; Hearin et al. 2017): the number of satellites N_{sat} is drawn from a Poisson distribution, their spatial distribution follows a Navarro–Frenk–White (NFW) profile, and the velocity profiles can be inferred from the virial theorem.

We then create alternative models by varying each of the baseline assumptions of the HOD and studying their effect on clustering via monopole, quadrupole, and projected correlation functions (ξ_0, ξ_2, w_p). We explore the choice of a Gaussian or a smooth step function as an alternative for the shape of the mean HOD for central galaxies, $\langle N_{cen}(M) \rangle$, options explored in other HOD studies (Zehavi et al. 2005; Favole et al. 2016; Guo et al. 2019). Motivated by the study in Jiménez et al. (2019), we also study non-Poissonian probability distribution functions (PDF, $P(N|\langle N \rangle)$) for populating haloes with satellite galaxies: the nearest integer and the negative binomial distribution.

As an alternative to NFW profiles, we also use the particle distribution within haloes and allowed for a rescaling of the halo concentrations. The latter is motivated by some studies predicting ELGs should be in the outskirts of haloes (e.g. Orsi & Angulo 2018). With respect to satellite velocities, in the case of NFW profiles, velocities follow a Gaussian distribution with a dispersion predicted by the virial theorem, whereas in the case of using particles, satellites take the particle velocity. We also include a velocity bias parameter that modulates the dispersion of satellite velocities with respect to the halo velocity. Another ingredient for our alternative models consists on adding a net infall velocity as motivated by Orsi & Angulo (2018).

This study is part of a coordinated release of the final eBOSS measurements of BAO and RSD in the clustering of not only ELGs (de Mattia et al. 2020; Raichoord et al. 2020; Tamone et al. 2020), but also LRG (Gil-Marin et al. 2020; Bautista et al. 2020), and quasars (Hou et al. 2020; Neveux et al. 2020). An essential component of these studies is the construction of data catalogues (Lyke et al. 2020; Ross et al. 2020), approximated mock catalogues (Lin et al. 2020; Zhao et al. 2020), and N-body simulations based mock catalogues for

Table 1. The OUTER RIM simulation cosmological and setup parameters. The cosmological parameters (Komatsu et al. 2011): Ω_{cdm}, Ω_b and Ω_Λ are the average densities of cold dark matter, baryonic matter, and vacuum energy in units of the critical density today, H_0 is the Hubble parameter, σ_8 is the rms of the matter fluctuations at $8\ h^{-1}$Mpc, and n_s is the spectral index of the primordial power spectrum. Simulation parameters: redshift of the snapshot used in this paper, volume of the simulation box, number of particles in the simulation, and particle mass resolution.

Ω_{cdm}	0.220
Ω_b	0.0448
Ω_Λ	0.7352
$h \equiv H_0/(100\ \mathrm{km\,s^{-1}\,Mpc^{-1}})$	0.71
σ_8	0.8
n_s	0.963
z_{snap}	0.865
Volume	$(3000\ h^{-1}\ \mathrm{Mpc})^3$
N	10240^3
m_p	$1.85 \times 10^9\ h^{-1}\ \mathrm{M_\odot}$

assessing systematic errors (Alam et al. 2020; Rossi et al. 2020; Smith et al. 2020), as the ones presented here. At the highest redshifts ($z > 2.1$), the coordinated release of final eBOSS measurements includes measurements of BAO in the Ly α forest (du Mas des Bourboux et al. 2020). The cosmological interpretation of these results in combination with the final BOSS results and other probes is found in Bautista et al. (2020).[1]

The plan of this paper is as follows. In Section 2, we introduce the OUTER RIM simulation. In Section 3, we describe the input observational data: ELG catalogue, summary statistics, and correlation functions. In Section 4, we describe the different halo occupation models we use to generate mock catalogues varying the mean HOD for central and satellite galaxies (Section 4.1), the PDF (Section 4.2), the radial (Section 4.3), and velocity (Section 4.4) distributions of satellite galaxies. In Section 5, we present mocks that best fit the data under different assumptions. Finally, we conclude in Section 6 and discuss our results and future prospects.

2 THE OUTER RIM SIMULATION

The OUTER RIM simulation (Heitmann et al. 2019) was run assuming a cosmology consistent with the 7th year release from WMAP (Komatsu et al. 2011), as summarized in Table 1, which will be our fiducial cosmology throughout the paper. The OUTER RIM simulation has outputs at 99 redshifts (34 between $1 < z < 3.5$). Haloes with at least 20 particle members, were identified at $0 < z < 10$ using the friends-of-friends (FoF) algorithm (Davis et al. 1985) with a linking length of $b = 0.168$. Technical properties of the OUTER RIM simulation are summarized in the second part of Table 1, some of these are similar to those from its predecessor, the Q Continuum simulation (Heitmann et al. 2015).

For this study, we use a snapshot at fixed $z = 0.865$, close to the effective redshift of the distribution of eBOSS/ELGs ($z_{eff} = 0.845$, Raichoord et al. 2020). Two key functions for the HOD model are the halo mass function and the bias function of the OUTER RIM simulation shown in Fig. 1. We will see in Section 4 how their integrals are used to put basic constraints on the HOD parameters. When referring to the bias in this study, we will always refer to the linear local bias, shown to describe accurately the clustering of haloes/galaxies at large scales.

---

[1] A description of eBOSS and a link to its associated publications can be found in https://www.sdss.org/surveys/eboss/.
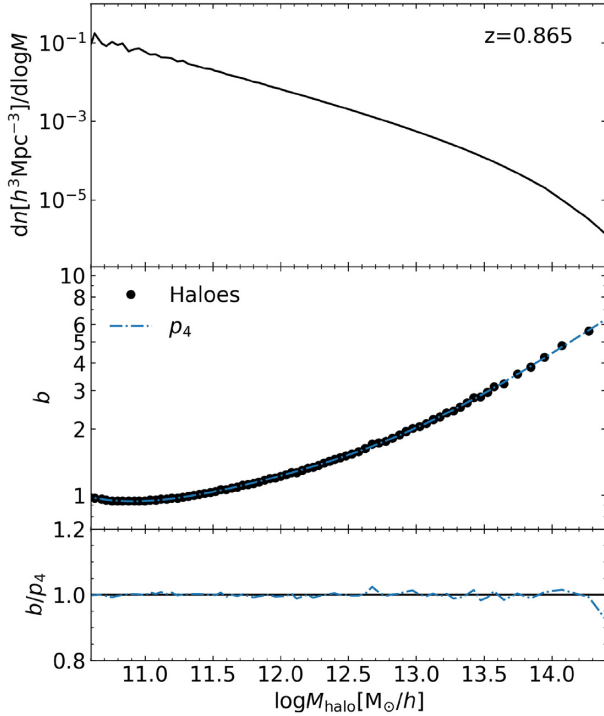
Figure 1. *Top:* The halo mass function from the OUTER RIM simulation output at $z = 0.865$. *Middle:* The halo bias function, fitted for separations $20 \leq r(h^{-1}\,\mathrm{Mpc}) \leq 80$ for the OUTER RIM simulation dark matter haloes as a function of the halo mass (see text), the black circles. A fourth order polynomial fit, p_4, to the bias function, $b(M)$ is also shown. *Bottom:* Ratio of the bias, $b(M)$ to the polynomial fit, p_4.

In order to compute the OUTER RIM simulation bias function, we split the simulation box in 27 cubes of $l = 1\mathrm{Gpc}\,h^{-1}$ side. We further split the halo catalogues by mass in logarithmic mass bins with $\Delta(\log M) = 0.1$[2] for $\log M < 12.5$, and making larger intervals at higher masses in order to decrease the shot noise. We compute the correlation function (details in Section 2.1) in real space $\xi_{M_i}(r)$ for each mass bin and subbox. Then, we compute the mean $\bar{\xi}_{M_i}$, and corresponding standard deviation, $\sigma(\xi_{M_i})$ for the 27 subboxes. We find the bias b_i that minimizes the χ^2 defined as

$$\chi^2(b_i) = \sum_r \left(\frac{\xi_{\mathrm{lin}}(r) \cdot b_i^2 - \bar{\xi}_{M_i}(r)}{\sigma(\xi_{M_i})(r)} \right)^2. \tag{1}$$

The summation above is done in the range $20 \leq r \leq 80\,\mathrm{Mpc}\,h^{-1}$ (with this choice we avoid non-linearities affecting small scales and the BAO feature). We found (here, and also in Section 3.2, equation 11) that the fits are more stable against noise and scale cuts, when using logarithmic binning for the data/mock correlation functions and most stable in the selected range of scales. We compute $\xi_{\mathrm{lin}}(r)$ by Hankel transforming the linear power spectrum obtained from CAMB (Lewis & Bridle 2002):

$$\xi_{\mathrm{lin}}(r) = \frac{1}{2\pi^2} \int_0^\infty P_{\mathrm{lin}}(k) j_0(kr) k^2 \mathrm{d}k. \tag{2}$$

The bias as a function of halo mass, $b(\log M)$, is then fitted to polynomials of orders from 2 to 5. We discard the points beyond $\log M = 14.4$ as they yield a poor fit. At those masses, the binning

---

[2]Throughout this paper, we use *log* for the decimal logarithm and take its argument in units of $M_\odot\,h^{-1}$.

becomes too coarse if we want to keep low contribution from shot noise and the definition of the bin centre becomes ambiguous as the halo mass function decays exponentially within the bin. We find a good fit with a polynomial of order $k = 4$ or larger. For the rest of the work, we use the fourth-order polynomial fit, p_4, to approximate the bias function. This fit is shown in Fig. 1.

2.1 Correlation functions from OUTER RIM mocks

When computing 2-point correlation functions (2PCF) in the simulation, we will always assume periodic conditions for the subboxes. This will introduce a small error in the boundaries, but we expect it to be negligible, given that our subbox size is much larger than the maximum scale used here.

A way to correct for the boundary conditions would be to introduce a random catalogue in order to account for the geometry of the subbox, as we will do for the survey geometry (see Section 3.4). However, this process would significantly slow down the 2PCF calculations as random catalogues are typically required to have at least 10 times more objects than the data catalogue in order to avoid introducing extra noise.

In the cases we compute correlations in redshift space, we use

$$\boldsymbol{s} = \boldsymbol{r} + \frac{1+z}{H(z)} \frac{\boldsymbol{v} \cdot \boldsymbol{r}}{|\boldsymbol{r}|} \frac{\boldsymbol{r}}{|\boldsymbol{r}|}, \tag{3}$$

with $\boldsymbol{s}$ representing the halo/galaxy position in redshift space, $\boldsymbol{r}$ in real space, $\boldsymbol{v}$ its comoving velocity, and $H(z)$ the Hubble parameter at redshift z. We adopt the plane-parallel approximation and assume the Z-axis as the line of sight.

Given the assumed boundary conditions, $\xi(r, \mu)$ is calculated using simply the natural estimator: $1 + \xi(r, \mu) = DD(r, \mu)/(n\Delta V)$, where DD is the number of galaxy/halo pairs with separation between r and $r + \Delta r$ and an orientation between μ and $\mu + \Delta \mu$ (where μ is the cosine of the angle with respect to the line of sight) and the denominator is the average number of galaxies found in the volume ΔV of the spherical shell of radius r and thickness dr. This was computed with a modified version of the code CUTE (Alonso 2012).[3]

We then compute the multipoles, integrating over the Legendre polynomials L_ℓ:

$$\xi_\ell(s) = (2\ell + 1) \int_0^1 \xi(s, \mu) L_\ell(\mu) \mathrm{d}\mu. \tag{4}$$

The projected two-point correlation function, $w_p(r_p)$, is obtained with the publicly available PYTHON package CORRFUNC.[4] This correlation function removes most of the effect of peculiar velocities by integrating along the line of sight:

$$w_p(r_p) = 2 \int_0^{\pi_{\max}} \xi(r_p, \pi) \mathrm{d}\pi, \tag{5}$$

where we use $\pi_{\max} = 80\,\mathrm{Mpc}\,h^{-1}$. In this case, ξ is also computed using the natural estimator, but counting pairs in bins of comoving distance parallel (π) and perpendicular (r_p) to the line of sight.

3 THE EBOSS ELG DATA

The eBOSS survey (Dawson et al. 2016) has observed a spectroscopic sample of star-forming ELGs (Raichoor et al. 2016, 2017) in fields both in the North and South Galactic Caps (NGC and SGC). The

---

[3]https://github.com/damonge/CUTE
[4]https://github.com/manodeep/Corrfunc (Sinha & Garrison 2020).

redshift of these galaxies has been identified using the [O II] doublet, with rest-frame wavelengths of $\lambda = 3727, 3729$ Å.

We aim at reproducing the eBOSS ELG number density and linear bias with the mock catalogues we generate in this work. We describe below how these quantities are measured from the data. We also describe the correlation functions of the data used as an input for Section 5.

3.1 eBOSS LSS ELG catalogues

Optical and near-infrared cosmological surveys are targeting star-forming ELGs at $0.5 < z < 2$, as these galaxies can provide a high enough effective volume to measure the BAO with high precision (e.g. Comparat et al. 2016). Star-forming ELGs present strong spectral emission lines that allow for a robust determination of their redshifts in a small observing time, maximizing the volume covered by the survey (e.g. Okada et al. 2016). Strong spectral lines can also be produced by galaxies with nuclear activity, AGNs. These are expected to be hosted by different average dark matter haloes than star-forming galaxies. No broad-band lines have been found among the eBOSS ELG sample and only a small fraction of eBOSS ELGs are expected to be AGNs in the redshift range under study, based on previous studies (Comparat et al. 2013).

Here, we use the data from the DR16 ELG clustering catalogue, described in Raichoord et al. (2020), which only includes ELGs with a good redshift determination. There are 173 736 eBOSS ELGs, within the redshift range $0.6 < z < 1.1$ and with an effective redshift of $z_{\rm eff} = 0.845$. The effective areas in the North and South fields are $A_{\rm NGC,eff} = 369\,{\rm deg}^2$, $A_{\rm SGC,eff} = 358\,{\rm deg}^2$. The catalogue also contains the weights to correct individual galaxies for systematic errors:

$$w_{\rm ELG} = w_{\rm sys}\, w_{\rm no\,z}\, w_{\rm CP}, \qquad (6)$$

due to the photometric target selection, $w_{\rm sys}$; redshift failures, $w_{\rm no\,z}$; and $w_{\rm CP}$ includes the 'close pairs' fiber collision correction adopted in eBOSS cosmological analysis (Raichoord et al. 2020; Ross et al. 2020; de Mattia et al. 2020). We will not use the $w_{\rm CP}$ weights in this study as the fiber collision effect will be accounted for by the pairwise-inverse-probability (PIP) + angular up-weighting (ANG) weights, which are more accurate, specially at small scales (see Section 3.4). Additionally, the standard inverse variance weights $w_{\rm FKP}$ are applied, in order to improve signal-to-noise ratio (Feldman, Kaiser & Peacock 1994).

3.2 Number density of the eBOSS ELGs

We aim to reproduce the abundance and clustering of the eBOSS ELGs in the OUTER RIM simulation. The number density and bias derived from the data will need to rely in the assumed cosmology, which in this case is that of the OUTER RIM simulation (Table 1).[5]

First, we compute the number density of ELGs for the NGC, the SGC, and the combination of both:

$$\bar{n} = \frac{N_{\rm eff}}{V_{\rm eff}} \qquad \bar{n}_{\rm eBOSS} = 2.187 \cdot 10^{-4}({\rm Mpc}/h)^{-3}$$

$$\bar{n}_{\rm SGC} = 2.267 \cdot 10^{-4}({\rm Mpc}/h)^{-3},\ \bar{n}_{\rm NGC} = 2.110 \cdot 10^{-4}({\rm Mpc}/h)^{-3}, \qquad (7)$$

[5]Note that the latter is different from the cosmology used for the main data analysis (de Mattia et al. 2020, Tamone et al. 2020, with namely $\Omega_M = 0.31$).

with

$$V = \frac{1}{3}\left(\chi(z_{\rm max})^3 - \chi(z_{\rm min})^3\right) \cdot A_{\rm eff}\left(\frac{\pi}{180\,{\rm deg}}\right)^2 \qquad (8)$$

giving a volume of $V_{\rm SGC,eff} = 0.410({\rm Gpc}/h)^3$, $V_{\rm NGC,eff} = 0.424({\rm Gpc}/h)^3$, $V_{\rm tot,eff} = 0.834({\rm Gpc}/h)^3$. Note that this $V_{\rm eff}$ refers purely to the observational volume taken into account, given the effective area and redshift range. In other studies, $V_{\rm eff}$ may refer to the equivalent volume of a cosmic variance limited sample with the same power spectrum variance (i.e. the shot noise being interpreted as a reduction of the effective volume).

The eBOSS number density $\bar{n}_{\rm eBOSS}$ will be used as a reference throughout the paper. In some occasions, we will use a factor 7 or 10 higher density in order to measure more accurately the clustering of the mocks.

3.3 Linear bias of eBOSS ELGs

The second quantity that we want to measure from the data is the large-scale bias b. A simple way is to compute the bias from the monopole of the data, and fit it using the Kaiser factor (Kaiser 1987) together with linear perturbation theory:

$$\xi_{0,\rm lin}(s) = \left(b^2 + \frac{2}{3}bf + \frac{1}{5}f^2\right)\xi_{\rm lin}(s), \qquad (9)$$

with f the log-derivative of the growth factor. Within the standard assumption of General Relativity, $d\log D/d\log a \approx \Omega_m(a)^{0.545}$ (Peebles 1980; Linder 2005). We fix f using this approximation for the OUTER RIM cosmology.

In earlier versions of the (data and mock) catalogues, we used the approach described above. For the mocks presented in this paper, we decided to fit the bias of the data by rescaling the shape of $\xi_0(s)$ of a mock with similar amplitude as that of the data (derived in earlier versions of the mocks):

$$\xi_0(b^2, s) = \frac{\left(b + \frac{2}{3}bf + \frac{1}{5}f^2\right)}{\left(b_{\rm mock} + \frac{2}{3}b_{\rm mock}f + \frac{1}{5}f^2\right)}\xi_{0,\rm mock}(s). \qquad (10)$$

This method encapsulates better the non-linearities, as they are present in the simulation and are expected in the data.

To measure the large-scale linear bias, we follow a similar approach as in Section 2 and split the simulation in 27 subboxes of size $l = 1\,{\rm Gpc}\,h^{-1}$ and comparable volume $V_{\rm mock} = l^3$ to the total eBOSS ELG effective volume. The $b_{\rm mock}$ is computed as explained in Section 4.1 (equation 15) and validated using the same approach as in Section 2 (equation 1) for the haloes. We compute the monopole in each subbox and the global mean $\xi_{0,\rm mock}$, to be input to equation (10), and standard deviation $\sigma_{\rm mock}$, so that we can compute

$$\chi^2(b) = \sum_s \left(\frac{\xi_0(b, s) - \xi_{0,\rm data}(s)}{\sigma_{\rm mock}(s)\sqrt{V_{\rm mock}/V_{\rm eBOSS}}}\right)^2, \qquad (11)$$

and minimize χ^2 to get the best fit bias b with a 1σ confidence interval corresponding to $\Delta\chi^2 = 1$. The fit to the data bias is more stable when using logarithmic binning for both the analytical approximation and the mocks. We use the range of scales $20 < s < 55\,{\rm Mpc}\,h^{-1}$, where the p-values are good for all cases (combining linear versus logarithmic scale, using the linear theory or mocks for the fits, and using either or both galactic caps). Following this procedure, we obtain

$$b_{\rm eBOSS} = 1.320 \pm 0.014,$$

$$b_{\rm SGC} = 1.310 \pm 0.020,\ b_{\rm NGC} = 1.330 \pm 0.020. \qquad (12)$$

As we find that the NGC and SGC have compatible biases, in the remainder of this paper we use the combined data, unless otherwise specified. Note again that this differs from bias values found in complementary studies, where the assumed cosmology was different.

3.4 Weighted correlation functions

eBOSS measured spectra using optical fibres positioned in pre-drilled plates at the 2.5 m Sloan Telescope (Gunn et al. 2006; Smee et al. 2013). The plates have a field of view of $\sim 7 \mathrm{deg}^2$ and can hold up to 1000 fibres. Typically, 100 fibres are used for calibration and 900 for science targets. Each fibre plus its ferrule has a diameter of 62 arcsec. The fibre collision scale imposes a minimum separation between objects that can be observed simultaneously. The eBOSS survey repeats observations of the same regions of the sky, allowing to observe some of the target objects, missed at previous passes, due to fibre collision. However, not all the targets will be (spectroscopically) observed once the survey is finished.

In this work, we find the best HOD models by fitting different clustering statistics measured from the model catalogues to the observed ones (details of this procedure can be found in Section 5). The data derived from observations has been corrected for the effect of missing spectra for photometric targets using the PIP weighting and the ANG techniques (Bianchi & Percival 2017). When pairs of galaxies are counted for calculating the correlation functions, these techniques modify the standard fibre collision correction, w_{CP} (see equation 3.1).

The PIP weight of a given pair of galaxies is defined as the inverse of the probability of this pair being assigned a fibre within the ensemble set from which the survey undertaken is considered to be randomly drawn (Bianchi & Percival 2017; Mohammad et al. 2020). This weighting scheme does not take into account the fraction of colliding pairs of galaxies that fall in single pass regions. The small-scale clustering, affected by fibre collisions, can be recovered using the ANG scheme proposed by Percival & Bianchi (2017). This scheme assumes that the set of un-observed pairs is statistically equivalent to the observed one. The ANG scheme is applied to the counts of pairs both of observed galaxies and observed-random ones. This method gives a statistically unbiased estimator for the clustering even at scales below the fibre collision.

Additionally, for large scales $s > 25 \mathrm{\,Mpc}\,h^{-1}$, there are some photometric angular systematics in the quadrupole that are not accounted for by any of the schemes described above. This is why in Tamone et al. (2020) they have decided to remove the $0.6 < z < 0.7$ data and use a *modified* correlation function that removes angular power. This is also the reason for our work to use the quadrupole data only for $s < 25$, see Section 5. Alternatively, we could used power spectrum multipoles, where these angular systematics are nulled with a pixelization scheme. We leave this for a future study, as comparing data and mocks in Fourier Space requires a careful window function treatment of the mocks. Moreover, the information is spread differently in Fourier space, with 1-halo and 2-halo terms more entangled (see Appendix A).

For the data we compute the correlation function from data–data, data–random, and random–random pairs using the Landy–Szalay estimator (Landy & Szalay 1993).

$$\xi(x, y) = \frac{DD(x, y) - 2DR(x, y) + RR(x, y)}{RR(x, y)}. \tag{13}$$

This gives us $\xi(s, \mu)$ and $\xi(r_p, \pi)$ in order to obtain, respectively, the multipoles (equation 4) and w_{p} (equation 5). We also use this

approach when computing correlations of the EZMOCKS in Section 5, as they also include the survey geometry.

4 THE MOCK CATALOGUES

Within the HOD Model framework, we assume that a galaxy mock catalogue can be constructed directly from a halo catalogue containing just the halo positions, velocities, and masses. Only in some specified examples below (Sections 4.3 and 4.4), we will use additional information from particles within haloes.

The HOD models used here have contributions from two galaxy populations: centrals and satellites, with $\langle N_{\mathrm{cen}}(M)\rangle$ and $\langle N_{\mathrm{sat}}(M)\rangle$ their expected number of galaxies per halo of mass M. The number density of the total galaxy sample in the model catalogues is calculated as follows:

$$\bar{n}_{\mathrm{gal}} = \int \frac{\mathrm{d}n(M)}{\mathrm{d}M} \big[\langle N_{\mathrm{cen}}(M)\rangle + \langle N_{\mathrm{sat}}(M)\rangle \big] \mathrm{d}M, \tag{14}$$

with $\frac{\mathrm{d}n(M)}{\mathrm{d}M}$ the differential halo mass function.

The clustering of the resulting model galaxy sample has two contributions: one coming from galaxies on the same halo, the 1-halo term, and one coming from correlations of galaxies hosted by different haloes, the 2-halo term. The clustering at large scales, which is dominated by the 2-halo term, can be described almost completely by the linear bias, which depends only on $\langle N_{\mathrm{tot}}(M)\rangle = \langle N_{\mathrm{cen}}(M)\rangle + \langle N_{\mathrm{sat}}(M)\rangle$ (e.g. Berlind & Weinberg 2002):

$$b_{\mathrm{gal}} = \frac{1}{\bar{n}_{\mathrm{gal}}} \int \frac{\mathrm{d}n(M)}{\mathrm{d}M} \cdot b(M) \big[\langle N_{\mathrm{cen}}(M)\rangle + \langle N_{\mathrm{sat}}(M)\rangle \big] \mathrm{d}M. \tag{15}$$

The clustering at small scales is dominated by the 1-halo term, which is affected by a range of properties beyond the linear bias of the sample. Below, we list the modelling of a set of properties that can have a strong impact on the 1-halo term of the clustering:

(i) The split between satellite and central galaxies, and the specific HOD shape $\langle N_{\mathrm{cen}}(M)\rangle$, $\langle N_{\mathrm{sat}}(M)\rangle$ (Section 4.1).
(ii) The PDF $P(N|\langle N\rangle)$ (Section 4.2).
(iii) The radial profile of satellites $\rho_{\mathrm{sat}}(r)$ (Section 4.3).
(iv) The velocity profile of the satellites $\phi(v_{\mathrm{r}})$ (Section 4.4).

In the following subsections, we describe the choices we make about those properties, and how we vary them to explore their influence on the clustering.

4.1 Mean halo occupation distribution for centrals and satellites

The mean HOD, $\langle N_i(M)\rangle$, encapsulates the average distribution of a given type of galaxy hosted per halo of a certain mass M. The analytical description of the mean HODs has been derived from either semi-analytical or hydro-dynamical simulations for galaxy formation and evolution (e.g. Berlind et al. 2003; Zheng et al. 2005). The shape of the model mean HOD depends on the properties of the selected galaxies. When galaxies are selected by their magnitude or stellar mass, the $\langle N_{\mathrm{cen}}\rangle$ can be described as a smoothed step function $(\mathrm{erf}(x))$ and a power law for satellite galaxies. This is the most commonly used shape for the mean HOD. Here, we label this shape as HOD-1:

HOD-1 centrals:

$$\langle N_{\mathrm{cen}}(M)\rangle = \frac{1}{2} A_c \left(1 + \mathrm{erf}\left(\frac{\log(\mathrm{M}) - \mu}{\sigma} \right) \right). \tag{16}$$

Satellites (all HODs):

$$\langle N_{\rm sat}(M)\rangle = A_s \left(\frac{M-M_0}{M_1}\right)^{\alpha}. \tag{17}$$

This shape has been shown to describe well the abundance of galaxies for a magnitude-limited sample (e.g. Zehavi et al. 2011). In a complete sample, we would have $A_c = 1$ and the number of central galaxies would transition from 0 to 1 at $\log M \sim \mu$ with a smoothing scale of σ. This means that for high enough masses all haloes are expected to have a central galaxy. For samples that are quite incomplete in mass, such as ELGs, QSO, or colour-selected samples, one could have $A_c < 1.0$ (e.g. Geach et al. 2012; Smith et al. 2020). Note that even if the mean HOD of model LRGs does not follow HOD-1 exactly (Hernández-Aguayo et al. 2020), these samples have been shown to be well described with such a parametrization (e.g. Gil-Marín et al. 2020). Here, we define the completeness as the ratio between the number of galaxies in a given sample and the total number of galaxies. In a more general case, A_c could vary with mass, changing the shape of the mean HOD. In the literature μ is usually denoted as $\log M_{\rm min}$, but we choose this nomenclature for consistency with models HOD-2 and HOD-3 (see below).

In equation (17), the satellite occupation follows an increasing power law, implying that the more massive the halo, the more satellite galaxies we expect to find, with α controlling the mass–richness relation. For $A_s = 1.0$ and $M_0 \ll M_1$, M_1 represents the mass at which we expect 1 satellite per halo. We note that A_s is completely degenerate with M_1, but we keep both parameters to separate their physical meaning and interpret A_s as the completeness of the satellites.

In equation (16), central galaxies have a constant probability to be found in haloes above a certain mass. However, this is at odds with the results derived for observed star-forming ELGs (Geach et al. 2012; Cochrane et al. 2017; Guo et al. 2019) and model ones (Cochrane & Best 2018; Gonzalez-Perez et al. 2018; Favole et al. 2019). More generally, the soft step function shape is not representative of samples of model galaxies selected by their age or their star formation rate (Zheng et al. 2005; Contreras et al. 2013). The star formation of galaxies depends in a non-trivial way on their stellar mass and environment. Massive galaxies tend to have lower star formation rates per stellar mass unit (e.g. Davies et al. 2019). ELGs are on average less massive than samples such as LRGs, and are mostly found in filaments (e.g. Darvish et al. 2014; Gonzalez-Perez et al. 2020). Galaxy formation processes affecting star-forming galaxies impact the expected shape of their HOD.

Fig. 2 shows the HOD for eBOSS ELGs from the SAM of galaxy formation and evolution presented by Gonzalez-Perez et al. (2018). Here, we fit the shape for centrals with two mean HOD models:

HOD-2 centrals:

$$\langle N_{\rm cen}(M)\rangle = \frac{A_c}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(\log M-\mu)^2}{2\sigma^2}} \tag{18}$$

HOD-3 centrals (default):

$$\langle N_{\rm cen}(M)\rangle = \begin{cases} \frac{A_c}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(\log M-\mu)^2}{2\sigma^2}} & \log M \leq \mu \\ \frac{A_c}{\sqrt{2\pi}\sigma}\cdot \left(\frac{M}{10^{\mu}}\right)^{\gamma} & \log M \geq \mu \end{cases}. \tag{19}$$

The HOD-2 has a simpler expression for centrals, being a Gaussian with amplitude A_c, mean μ, and variance σ^2. However, it fails to describe the asymmetry at $\log M > \mu$. That is why HOD-3 introduces a decaying power law for $\log M > \mu$. Note that all three HODs have the same functional shape for the satellites, although their parameters
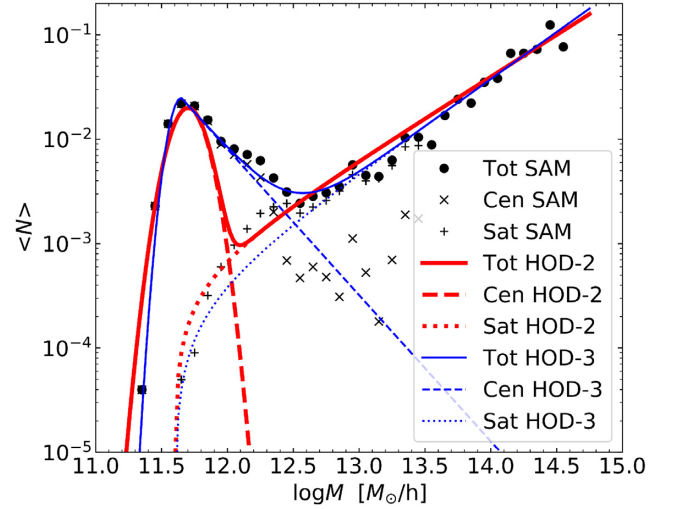

Figure 2. The mean eBOSS ELG HOD from a semi-analytical model (SAM) of galaxy formation (Gonzalez-Perez et al. 2018) for central ($\times$), satellites ($+$), and all galaxies (circles). The SAM HOD has been fitted using HOD-2, (equation 18, the red thick lines) and HOD-3 (equation 19, the blue thin lines). The solid lines show the fit to the total mean HOD, the dashed lines the contribution from centrals, and the dotted lines that from satellite galaxies.

may be different. The functional form that best describes the HOD from the SAM is the HOD-3, which we will consider as our default model in this work. The other two HOD shapes will be considered as variations of our model in which we increase (HOD-1) or decrease (HOD-2) the central galaxy occupation on the high halo mass end.

For every HOD model, we first apply the constraints of $\bar{n}_{\rm gal} = n_{\rm eBOSS}$ (or a multiple of it) and $b_{\rm gal} = b_{\rm eBOSS}$, using equations (14) and (15). Then, we set the fraction of satellites with

$$f_{\rm sat} = \frac{1}{\bar{n}_{\rm gal}}\int \frac{{\rm d}n(M)}{{\rm d}M}\langle N_{\rm sat}(M)\rangle {\rm d}M. \tag{20}$$

We fix the offset between μ and the mass parameters, $\log M_0$ and $\log M_1$. These fixed offsets and the parameters σ, α, and γ are derived from fitting the HOD-2 and HOD-3 equations to the ELG mean HOD derived in Gonzalez-Perez et al. (2018) and shown in Fig. 2. For the case of HOD-1 those values are taken from Zehavi et al. (2005). All the choices made for the different HOD parameters are shown in Table 2. For a given HOD (of the three described above) and the fixed choices of parameters just described, any given choice of $\{A_c, A_s, \mu\}$ yields a a set of $\{\bar{n}_{\rm gal}, f_{\rm sat}, b_{\rm gal}\}$, and vice versa.

On top left of Fig. 3, we present together the HOD shape of HOD-1, HOD-2, and HOD-3 for $\bar{n}_{\rm gal} = n_{\rm eBOSS}$, $b_{\rm gal} = b_{\rm eBOSS}$, and $f_{\rm sat} = 0.30$, as well as HOD-3 with $f_{\rm sat} = 0, 0.15, 0.45$ (and the same $\bar{n}_{\rm gal}$, $b_{\rm gal}$) and the original fit to the SAM with HOD-3, for comparison. We also show the corresponding monopole $\xi_0(s)$, quadrupole $\xi_2(s)$, and projected correlations $w_{\rm p}(r_{\rm p})$.

In order to compute more accurately the correlation functions, we increase the number density by a factor of 10 (7 in the case of $f_{\rm sat} = 0$ in order to avoid hitting the $\langle N_{\rm cen}\rangle = 1$ limit). The shaded area represents the one $\sigma_{\rm eBOSS}$ region expected for eBOSS computed as $\sigma_{\rm eBOSS} = \sqrt{\frac{1({\rm Gpc}/h)^3}{V_{\rm eBOSS}}}\sigma_{\rm mock}$, with $\sigma_{\rm mock}$ the standard deviation over the 27 $l = 1 Gpc\, h^{-1}$ subboxes of a mock realization with the eBOSS number density. We will follow this approach for all other figures in this section.

We find that the differences on the shown scales for the monopole are negligible. This is expected as we fixed the bias and we are looking

Table 2. List of parameters used in the three HOD models described in equations (16), (18), and (19); A_c, A_s, μ are free parameters that mostly control the quantity indicated in brackets, although one has to simultaneously fit the three free parameters to obtain the target $\{\bar{n}_{\rm gal}, f_{\rm sat}, b_{\rm gal}\}$. The values shown correspond to $\{\bar{n}_{\rm gal} = n_{\rm eBOSS}, f_{\rm sat} = 0.30, b_{\rm gal} = b_{\rm eBOSS}\}$. The rest of mass parameters ($\log M_0, \log M_1$) have a fixed offset with respect to μ and the scaling parameters (σ, α, and γ) are fixed. Both the offsets and scaling parameters are derived from the fits shown in Fig. 2 for HOD-2 and HOD-3, whereas for HOD-1 we take the values from Zehavi et al. (2005).

	A_c	A_s	μ	$\log M_0$	$\log M_1$	σ	α	γ
HOD-1	0.00723 ($\bar{n}_{\rm gal}$)	0.01294 ($f_{\rm sat}$)	11.110 ($b_{\rm gal}$)	μ	$\mu + 1.3$	0.15	1.0	–
HOD-2	0.01185 ($\bar{n}_{\rm gal}$)	0.009008 ($f_{\rm sat}$)	11.707 ($b_{\rm gal}$)	$\mu - 0.1$	$\mu + 0.3$	0.12	0.8	–
HOD-3	0.00537 ($\bar{n}_{\rm gal}$)	0.005301 ($f_{\rm sat}$)	11.515 ($b_{\rm gal}$)	$\mu - 0.05$	$\mu + 0.35$	0.08	0.9	-1.4
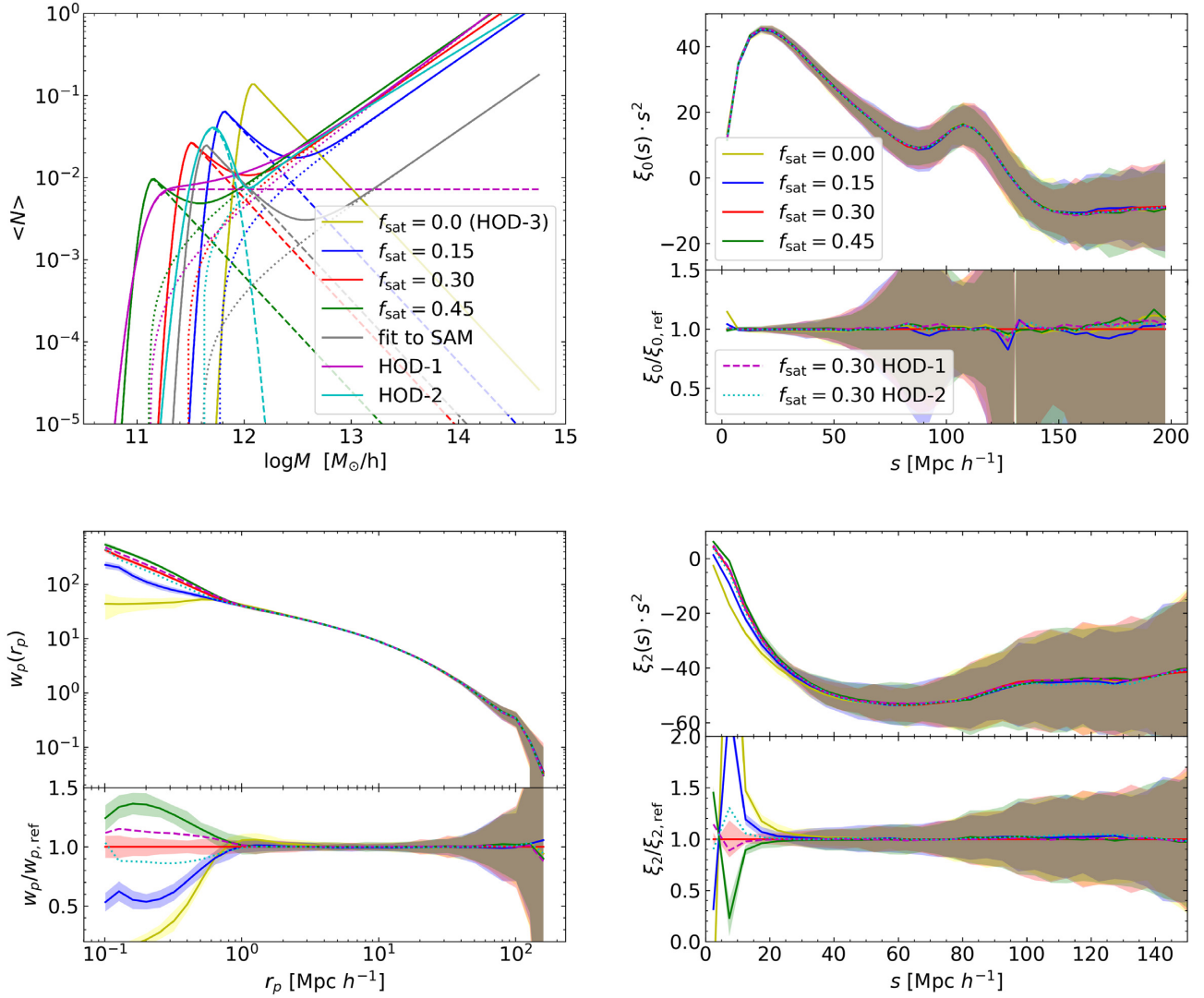
Figure 3. *Top left*: Mean halo occupation distribution for different models (equations 16, 18, and 19) and different fractions of satellites as indicated in the label. All HODs have been fitted to have the same number density and bias as the data, except for the grey curve (labelled *SAM*) that was fitted in Fig. 2 to the SAM, and is shown here for comparison. We use the same line styles as in Fig. 2: dashed for centrals, dotted for satellites, and solid for all. *Top right:* Monopole of the 2PCF for the HODs shown in the top left subfigure. *Bottom left:* Projected correlation function of the different HODs. *Bottom right:* Quadrupole of the 2PCF of the different HODs. The mock catalogues shown here were constructed assuming that satellites follow a Poisson distribution ($\beta = 0$, see Section 4.2), were distributed in haloes following NFW ($K = 1$, see Section 4.3) and that their velocities follow the virial theorem ($\alpha = 1$, see Section 4.4). These choices, together with HOD-3 and $f_{\rm sat} = 0.3$ are the default choices, and will be used unless otherwise specified. The reference model (used for the ratios and labelled as *ref*) has all the default choices and is the same across Figs 3, 4, 5, and 6. For the clustering subfigures, we have used an enhanced number density whereas the shaded area corresponds to the error expected for eBOSS data (see text for details). We use the same approach for Figs 4, 5, and 6. The legend is consistent across the different clustering subfigures.

at linear or quasi-linear scales, so we will not show the monopole in the following subsections. We would also find differences if we explored lower scales on the monopole using logarithmic binning. However, we find more illustrative to study the projected correlation function and the quadrupole, which offer complementary information of the effects of positions and velocities of satellites (as we will see along this section), respectively, whereas for the monopole on logarithmic binning those effects appear entangled (as it happens in Fourier space, see Appendix A).

The projected correlation shows the expected trend: a higher signal at small scales as we increase f_{sat}, increasing the contribution of the 1-halo term. For the quadrupole, we find differences already at $s \sim 25 \, \mathrm{Mpc} \, h^{-1}$ with lower (closer to zero, we will use this terminology in the remainder) signal for larger fractions of satellites, due to an increase of the finger-of-god effect (Jackson 1972; Peebles 1980). For the lowest point, some of the mocks invert their quadrupole's sign.

It is remarkable that the differences introduced by the choice of HOD shape are much less significant than the value of the satellite fraction, or the choices detailed in the subsections below. We also note that the clustering of the HOD-3 is approximately half way between that of HOD-1 and that of HOD-2, confirming our interpretation of HOD-3 being bracketed by models HOD-1 and HOD-2.

4.2 Probability distribution function

In Section 4.1, we have studied the mean HOD of satellites and centrals. Here, we study how we go from the mean value $\langle N \rangle$ to a given realization N, of the (integer) number of galaxies in a halo. This is given by the PDF, $P(N|\langle N \rangle)$.

By definition, for central galaxies N can only be 0 or 1. If satellite galaxy formation were a random uncorrelated process, they would follow a Poisson distribution. However, galaxy formation could affect their PDF, increasing or decreasing the scatter (Jiménez et al. 2019). The three PDFs for satellites that we consider are

(i) **Poisson distribution** (default)

$$P(N|\lambda) = \frac{e^{-N}\lambda^N}{N!},\tag{21}$$

with $\lambda \equiv \langle N \rangle \; (= \langle P(N|\lambda) \rangle)$ and $\sigma = \sqrt{\langle N \rangle}$

The Poisson distribution will be used for assigning satellite galaxies to haloes, unless otherwise stated.

(ii) **Nearest integer distribution**.

This function only allows two possible values for N, which are the two closest integers to $\lambda = \langle N \rangle$:

$$P(N|\lambda) = \begin{cases} 1 - (\lambda - \mathrm{INT}(\lambda)) & N = \mathrm{INT}(\lambda) \\ \lambda - \mathrm{INT}(\lambda) & N = \mathrm{INT}(\lambda) + 1. \\ 0 & \text{else} \end{cases}\tag{22}$$

The function $\mathrm{INT}(x)$ represents the truncation of x to the nearest lower integer. This distribution is always used for the centrals (for which only the 0 and 1 values are allowed) and it will be used for the satellites only when specified. This function has a lower scatter than the Poisson distribution: $\sigma = \sqrt{\Delta(1-\Delta)}$, with $\Delta = \lambda - \mathrm{INT}(\lambda)$.

(iii) **Negative binomial distribution**.

This function allows for a larger scatter than the Poisson distribution. The parameter β represents the relative increment of the standard deviation with respect to the Poisson distribution with $\lambda = \sqrt{N}$ and $\sigma = \lambda(1 + \beta)$. Here, we follow a similar notation

as in Jiménez et al. (2019).

$$P(N|r, p) = \frac{\Gamma(N+r)}{\Gamma(r)\Gamma(N+1)} p^r (1-p)^N \text{ with}$$

$$p = \frac{1}{(1+\beta)^2}, \; r = \frac{\lambda}{\beta(1+2\beta)}.\tag{23}$$

The Poisson distribution is recovered in the limit $\beta \to 0$, after solving a few indeterminations.

In Fig. 4, we show how the different PDFs introduced here affect the overall distribution N_{sat} (top subfigure) and how they affect the galaxy clustering. Changing the PDF has a small impact on the quadrupole, except for small scales below $10 \, \mathrm{Mpc} \, h^{-1}$. However, the effect on the projected correlation function is large at scales below $1 \, \mathrm{Mpc} \, h^{-1}$, increasing the signal as we increase the scatter. This is expected as when the scatter is increased, the probability of having a pair (or more) of satellites in the same halo increases.

4.3 Spatial distribution of satellite galaxies

We now study how to spatially distribute galaxies within a halo. The central galaxies are always placed at the position of the host halo. However, here we consider three ways to place satellite galaxies within their host haloes:

(i) **NFW profile** (default)

We place satellite galaxies following an NFW profile (Navarro, Frenk & White 1997):

$$\rho(x) \propto \frac{1}{x \cdot (1+x)^2} \quad \text{with} \quad x = c\frac{r}{r_{vir}},\tag{24}$$

where c is the concentration of the halo, which we take from the values tabulated in Klypin et al. (2016): $c(M) = c_{kly}(M)$. The virial radius from equation (24), r_{vir}, is computed following a common approach (e.g. Carretero et al. 2015; Avila et al. 2018) based on the spherical collapse model (Lacey & Cole 1993):

$$r_{vir} = \left(\frac{3}{4\rho_{crit}\Delta_{vir}M} \right)^{1/3},\tag{25}$$

and

$$\Delta_{vir} = 18\pi^2 + 82(1 - \Omega_M(z)) - 39(1 - \Omega_M(z))^2.\tag{26}$$

(ii) **Modified NFW**.

Observations find star-forming galaxies preferentially in the outskirts of filaments (e.g. Chen et al. 2017; Kraljic et al. 2018). SAMs of galaxy formation and evolution have found that star-forming galaxies tend to be found in the outskirts of haloes (Orsi & Angulo 2018). This suggests that star-forming ELGs will also be preferably located in the outskirts of haloes. We model this effect by placing satellite ELGs following a less concentrated profile. In this case, we modify the halo concentration from equation (24) by a factor $K < 1$:

$$c(M) = K \cdot c_{kly}(M).\tag{27}$$

(iii) **Particles**.

We pick a random particle within the halo and assign that position to the satellite galaxy, we will denote this choice as PART. This is computationally expensive. Additionally, since we only transferred 1 per cent of the halo particles randomly selected, with a minimum of five particles per halo, we find a few cases in which we run out of particles. The number of cases is really small, always fewer than 50/217 000 cases for n_{eBOSS} number density in extreme parameter choices. When that happens, we assign an additional satellite to the
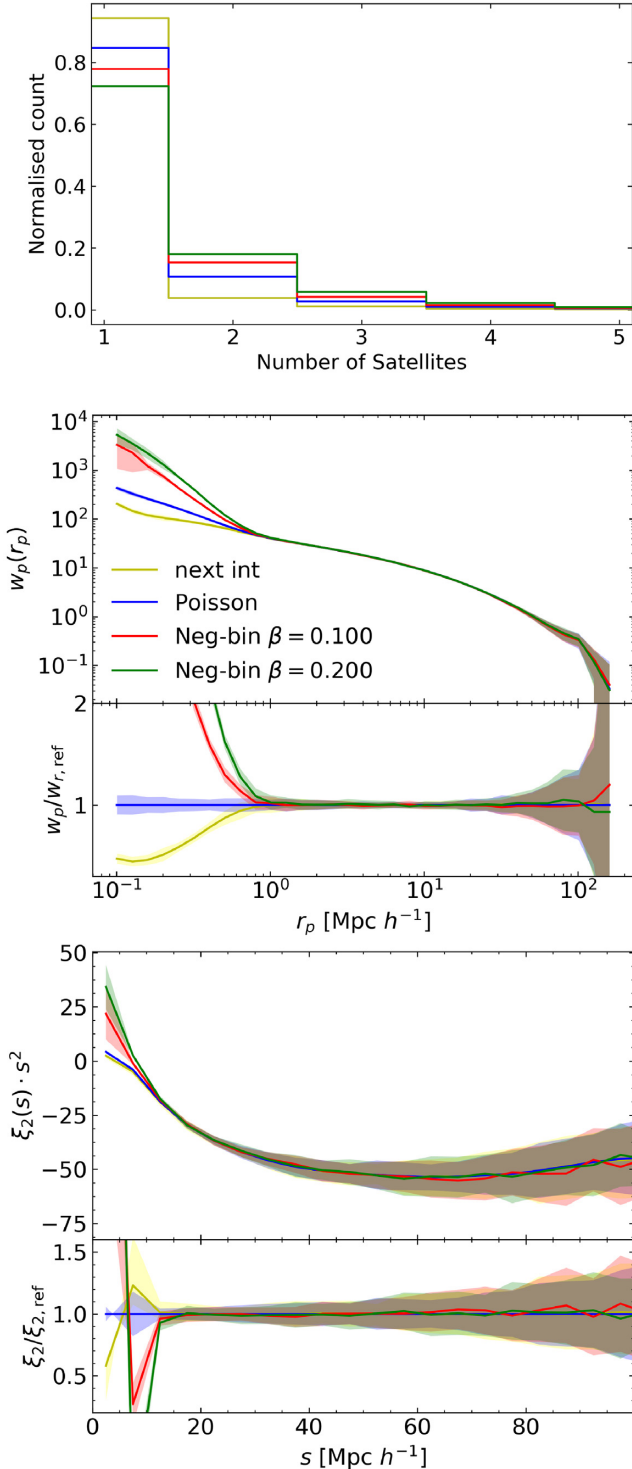
Figure 4. Effect of PDF of satellite assignment, considering: Poisson (equation 21), nearest integer (equation 22), negative binomial with $\beta = 0.1$ and $\beta = 0.2$ (equation 23). *Top:* Counts of haloes occupied by a given number of satellite galaxies $N \geq 1$ in the full ELG mock sample (with contributions from all halo masses and their corresponding mean occupations $\langle N_{\mathrm{sat}}(M)\rangle$, equation 17), divided by the total number of counts. *Middle:* Projected correlation function w_{p} for the same mocks, as indicated in the legend, and the ratios with respect to the mocks with the Poisson distribution. *Bottom:* Quadrupole and ratios. Besides the PDF specified, we take the default choices: $f_{\mathrm{sat}} = 0.30$, NFW profile ($K = 1.0$, see Section 4.3) and the virial theorem for the velocities ($\alpha = 1.0$, see Section 4.4).

next halo. Since we apply the HOD to haloes ordered by decreasing mass, the mass of the next halo is expected to be similar to that of the previous one. In this way, the original number of galaxies is maintained at the expense of modifying very slightly the PDF and HOD. Since the numbers are very small, we do not expect this choice to impact our results.

(iv) **Particles with a modified profile**. In this case, we use particles positions, but also model the ELGs preference to be in the outskirts of haloes. To accomplish this, once the satellite positions are assigned to random particles, these are perturbed following:

$$\boldsymbol{r}_{\mathrm{sat}} = \boldsymbol{r}_{\mathrm{h}} + \frac{1}{K}(\boldsymbol{r}_{\mathrm{DM}} - \boldsymbol{r}_{\mathrm{h}}), \tag{28}$$

with $\boldsymbol{r}_{\mathrm{sat}}$, $\boldsymbol{r}_{\mathrm{h}}$, $\boldsymbol{r}_{\mathrm{DM}}$ the position of the satellite galaxy, the halo, and the dark mater particle, respectively. This prescription is equivalent to the rescaling of concentrations for the NFW case, in both cases we are rescaling the profiles by $1/K$.

In the top panel of Fig. 5, we show the profile of satellite galaxies for mock catalogues with different K values for both NFW profiles and particle position assignments. There are clear differences between the profiles from mocks constructed assuming either NFW profiles or using the particle information. We can explain these differences first, because only relaxed dark matter have profiles that can be described analytically (e.g. Wang et al. 2019), and at $z \sim 1$, only half of the dark matter haloes in the OUTER RIM simulation are relaxed (Child et al. 2018). Secondly, relaxed dark matter haloes are better described by Einasto (1965) profiles, rather than an NFW one (e.g. Gao et al. 2008; Child et al. 2018). Thirdly, due to reduced access to the information, we derive the halo concentration from their FOF mass using Klypin et al. (2016). In general, in order to compute accurately the concentration of a halo one would fit the distribution of particles with a given profile (for the NFW case, equation 24).

Given that assuming NFW galaxy profiles is a common practice in mock catalogue generation and that the concentrations are defined in a more straightforward way in this case (e.g. Klypin et al. 2016), we continue to use the NFW galaxy assignment and compare their results to using particle profiles. In fact, despite the differences seen in the top panel of Fig. 5, the differences in the projected clustering (w_{p}, middle panel) are much smaller. At very small scales $r \sim 0.1\,\mathrm{Mpc}\,h^{-1}$, the particle profiles flattens and becomes lower than the NFW profile, giving also a smaller correlation at those scales. However, at $r \sim 1.0\,\mathrm{Mpc}\,h^{-1}$ the situation is inverted, the NFW profile ($K = 1.0$) has nearly reached the tail of r_{vir}, with $r_{\mathrm{vir}} = 0.78\,\mathrm{Mpc}\,h^{-1}$ for $\log M = 14$, having only 2.6 per cent of satellites beyond that mass.

Finally, we note that the impact of changing the density profiles is negligible for the quadrupole. The differences found between the PART and NFW profiles, are mostly due to having different velocities, as described below. This confirms our initial claim that these two 2PCF statistics (w_{p} and ξ_2) have very complementary information.

4.4 Velocity distributions

The remaining choice to make is the assignment of velocities to galaxies. For the central galaxies, we simply assume the same velocity as the halo, whereas for the satellites we consider several options:

(i) **Virial theorem**. (default) When using NFW for the position profiles, we follow Bryan & Norman (1998) for the veolicity profiles,
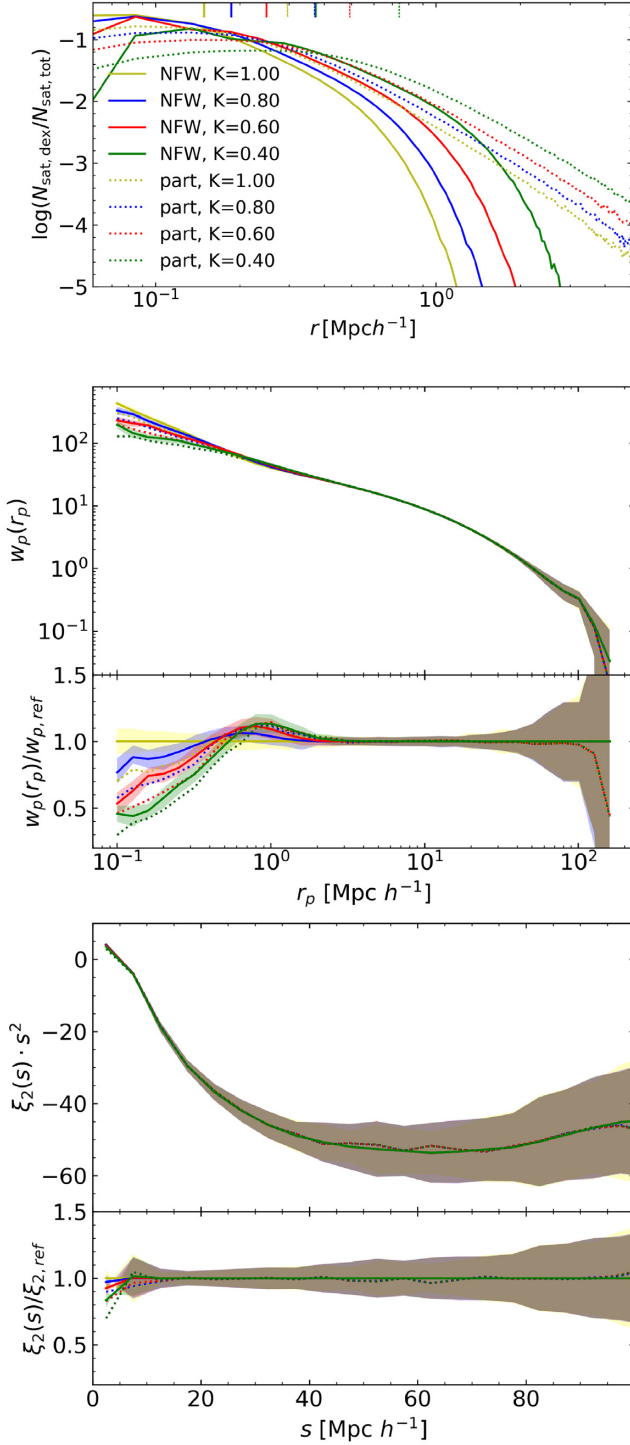
Figure 5. Similar to Fig. 4 but for the effect of the density profile of the satellites. We consider a Navarro–Frenk–White profile, profiles following the distribution of particles (PART), and both of them with modified concentration profiles ($K \neq 1$, equation 27). *Top:* Normalized mock ELG satellite count as a function of the distance from the halo centres r for the mocks indicated in the legend. The short vertical lines indicate the mean values of r. *Middle:* Projected two-point correlation function and ratios to the NFW with $K = 1$ case. *Bottom:* Quadrupole and ratios. Note the legends are consistent across subfigures.

also used in Carretero et al. (2015) and Avila et al. (2018):

$$\sigma_{\mathrm{vir}} = 476 \cdot 0.9 [\Delta_{\mathrm{vir}} E^2(z)]^{1/6} \left(\frac{M}{10^{15} M_\odot h^{-1}} \right)^{1/3} \mathrm{km\,s^{-1}}, \qquad (29)$$

with $E(z) = H(z)/H_0$. Note that this scaling is already predicted by the virial theorem. And, we assign

$$v_i^{\mathrm{gal}} \curvearrowleft \mathcal{N}(v_i^{\mathrm{h}}, \sigma_{\mathrm{vir}}) \quad \text{for } i = x, y, z, \qquad (30)$$

with $\mathcal{N}(\mu, \sigma)$ a normal distribution with mean μ and variance σ^2 and with $\boldsymbol{v}_h$ representing the velocity of the halo.

(ii) **Particle velocity**

When using particles for the satellite positions (Section 4.3), we also assign the velocity of the particles to the satellite galaxies.

(iii) **Velocity bias**

The dispersion of velocities of dark matter particles is a priori expected to be different than that of galaxies and subhaloes. For this reason, we include a velocity bias α_v in the velocity assignment:

$$\boldsymbol{v}_{\mathrm{sat}} = \boldsymbol{v}_{\mathrm{h}} + \alpha_v (\boldsymbol{v}_{\mathrm{DM}} - \boldsymbol{v}_{\mathrm{h}}). \qquad (31)$$

This equation is directly applicable when using particles. When using the virial theorem, the velocity bias may be understood as a rescaling of the velocity dispersions:

$$v_i^{\mathrm{gal}} \curvearrowleft \mathcal{N}(v_i^{\mathrm{h}}, \alpha_v \cdot \sigma_{\mathrm{vir}}). \qquad (32)$$

(iv) **Infall velocity**

We can split the galaxy velocity with respect to the halo in two components, radial v_r (defined along the line between the halo centre and the galaxy position) and angular v_ϕ:

$$\boldsymbol{v}_r = (\boldsymbol{v}^{\mathrm{gal}} - \boldsymbol{v}^{\mathrm{h}}) \cdot \boldsymbol{u}_r, \qquad (33)$$

$$\boldsymbol{v}_\phi = (\boldsymbol{v}^{\mathrm{gal}} - \boldsymbol{v}^{\mathrm{h}}) - \boldsymbol{v}_r, \qquad (34)$$

with

$$\boldsymbol{u}_r = \frac{\boldsymbol{r}_{\mathrm{sat}} - \boldsymbol{r}_{\mathrm{h}}}{|\boldsymbol{r}_{\mathrm{sat}} - \boldsymbol{r}_{\mathrm{h}}|}. \qquad (35)$$

Orsi & Angulo (2018) studied the velocity distributions of star-forming galaxies from a SAM of galaxy formation and evolution. They found that the radial component of the velocity, v_r, of star-forming galaxies has two contributions: a Gaussian centred at 0 equivalent to that described by equation (30) ($v_r \curvearrowleft \mathcal{N}(0, \sigma_{\mathrm{vir}})$) and a net infall velocity towards the centre of the halo, that follow approximately:

$$v^{\mathrm{infall}} \curvearrowleft \mathcal{N}(-500 \, \mathrm{km\,s^{-1}}, 200 \, \mathrm{km\,s^{-1}}). \qquad (36)$$

We include this contribution in the mock generation (when specified) by adding a new component to the velocity:

$$\boldsymbol{v}_{\mathrm{tot}}^{\mathrm{gal}} = \boldsymbol{v}^{\mathrm{gal}} + v^{\mathrm{infall}} \cdot \boldsymbol{u}_r. \qquad (37)$$

In the top panel of Fig. 6, we show the radial velocity distribution of the model ELGs generated with the velocity models detailed above. We see on one hand how the velocity bias α_v affects the width of the velocity distribution and on the other hand how adding the infall velocity (v^{infall}) shifts the centre of the distribution from $v_r \sim 0$ to $v_r \sim -500 \, \mathrm{km\,s^{-1}}$. This is a more extreme case than that described in Orsi & Angulo (2018), where they found a combination of the two peaks (at $v_r \sim 0$ and $v_r \sim -500 \, \mathrm{km\,s^{-1}}$). We expect that for a case similar to that described in Orsi & Angulo (2018), the results on galaxy clustering will be in between the two cases presented here (with and without v^{infall}).

The distribution of the particle velocities follows a distribution with a width similar that derived from the virial theorem, although
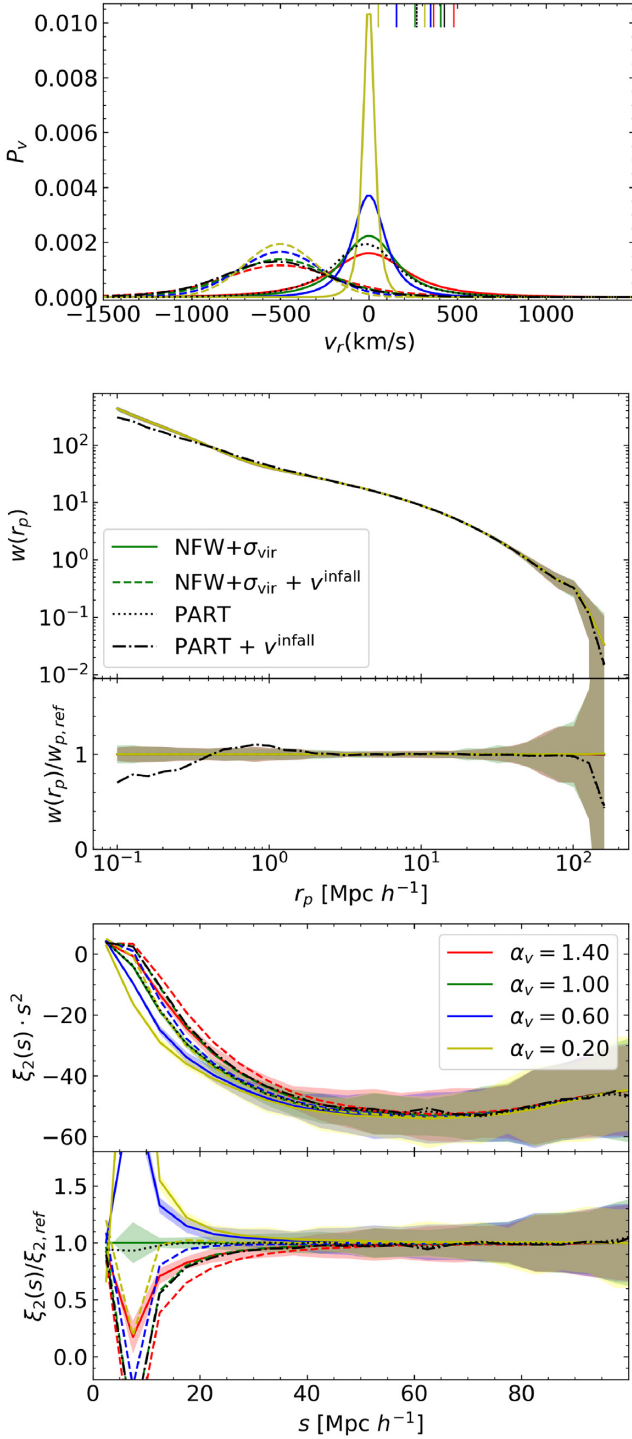
Figure 6. Similar to Fig. 4, but for the effect of the velocity profile of the satellites. We consider a distribution given by the virial theorem (NFW + $\sigma_{\rm vir}$) and velocities as given by the particles (PART), adding also different values of the velocity bias α_v (only for $\sigma_{\rm vir}$, we keep $\alpha_v = 1$ for PART in this figure) and a net infall velocity of $v_r = -500 \pm 200\,{\rm km\,s^{-1}}$ (for both NFW and PART). *Top:* Radial (from the halo centre to the satellite) velocity profiles. The area under the curves are normalized to unity. The upper vertical lines indicate the corresponding dispersion of satellite velocities around haloes along the Z-axis, $\sqrt{\langle (v_{\rm gal,z}^{\rm tot} - v_z^{\rm h})^2\rangle}$. *Middle:* Projected correlation function and ratios with respect to the default mocks (NFW + $\sigma_{\rm vir}$, $\alpha_v = 1$). Note that most curves line up in this subfigure, see text. *Bottom:* Quadrupole and ratios. Note the legends are consistent across subfigures.

slightly skewed towards negative values. This is expected for the particles, as not all of them are virialized and some are expected to currently being accreted.

The effect of the velocity distributions on the projected correlation function (middle of Fig. 6) is negligible. The only appreciable differences are actually due to the differences reported for the positions of satellite galaxies for NFW and particle profiles (see Section 4.3).

As expected, it is in the quadrupole where the main differences appear, due to variations in the velocity profiles. We find that the quadrupole is more suppressed (closer to zero) for higher velocity dispersion (higher α_v), as expected by the finger-of-god effect. The effect of adding a net infall velocity also suppresses the quadrupole power. This is expected, as this additional peculiar velocity uncorrelated with the large-scale structure also erases clustering along the line of sight.

We added in the top subfigure of Fig. 6 some vertical lines indicating the dispersion of velocities of satellite galaxies around haloes along the Z-axis (arbitrarily chosen as the line of sight) $\sqrt{\langle (v_{\rm gal,z}^{\rm tot} - v_z^{\rm h})^2\rangle}$. This allows us to quantify to first order the finger-of-god effect, and the ordering if these lines can be identified with the ordering of the quadrupoles at the mildly non-linear scales.

We remark that this dispersion is different to $\sqrt{\langle v_r\rangle}$, we checked the ordering would be quite different in that case and unrelated to the quadrupoles. When taking an arbitrary line of sight (e.g. the Z-axis), the dispersion due to the addition of $v^{\rm infall}$, appears a factor 1/3 smaller due to projection effects. This does not occur for the viral velocities whose dispersion occur in all the three dimensions.

The quadrupole from models using the particle information are within 1σ of that from assuming NFW profiles, and their ratios are very close to 1. This is remarkable, since the velocity assignment follows different procedures, and their velocities profile (top panel) showed some differences.

On the other hand, paying attention to the exact shape of the quadrupole, one can find subtle but statistically significant differences in the shape induced by the effect of velocity bias and the effect of infall velocities. For example, the $\alpha_v = 0.2$ and $v^{\rm infall}$case (yellow dashed) follows closely the line of the standard case ($\alpha_v = 1$ and $v_r = 0$, green solid) for $s > 12\,{\rm Mpc}\,h^{-1}$, where the effect of the overall velocity dispersion is already apparent. These curves, however, diverge at smaller scales. Many other subtleties could be found exploring in more detail the small scales, however, it is unclear that we could gain any further intuition from basic principles.

We also analysed the hexadecapole $\xi_4(s)$, and found some differences at small scales due to the finger-of-god effect. The differences among mocks were qualitatively similar to the results found in the quadrupole, but statistically less significant, which is why we omit the hexadecapole in the figures.

Additionally, in Appendix A we show the clustering in Fourier space of all the mocks presented in Sections 4.1, 4.2, 4.3, and 4.4. When exploring the effect of velocity profile choices, we also show the hexadecapole of the power spectrum.

5 FITTING THE EBOSS DATA

In this section, we describe the optimization procedure followed to find the HOD models that produce the mock catalogues with clustering closest to the eBOSS ELG data.

5.1 Optimization

In the previous section, we have presented the wide variety of HOD mocks generated for this work. Here, we explore the parameter space of the HOD models and constrain it with the observational data presented in Section 3. We construct our data vector, θ, as a combination of the monopole, the quadrupole, and the projected correlation function, at different scales (see also Section 3):

$$\theta_{0,2,r_p} = \{\xi_0(s_0),\ \xi_2(s_2),\ w_p(r_p),\ \}$$

$$\forall\ \{15\ <\ s_0\ <\ 40;$$

$$10\ <\ s_2\ <\ 25;\ 0.02 \leq r_p \leq 4.5\}\ [\text{Mpc}\,h^{-1}]. \quad (38)$$

We choose these scales so that the information in each statistics are complementary. The range of scales considered are enclosed by the two dotted vertical lines in Fig. 7, which is described in detail in Section 5.7. The w_p is cut at $r_{p,\text{max}} = 4.5\,\text{Mpc}\,h^{-1}$ so that it mostly contains information from the 1-halo term (see Figs 3, 4, and 5) given that the amplitude of the 2-halo term is set by the bias, which is fixed. As shown in Fig. 7, below $r_{p,\text{min}} = 0.02\,\text{Mpc}\,h^{-1}$ there is sudden change in the behaviour of $w_p(r_p)$ for the observational data. This might be due to some not fully accounted for systematic errors. As we have covered a wide range of the 1-halo term at these small scales, which are well below the fibre collision diameter scale (Section 3.1), we do not to include points below $r_p \leq r_{p,\text{min}} = 0.02\,\text{Mpc}\,h^{-1}$.

For the monopole, we only use quasi-linear scales ($15 < s < 40$), which for most cases will not be affected by the choice of HOD parameters (once b is fixed), but it does help ruling out some extreme models. For the quadrupole, we also use quasi-linear scales. In this case, those scales enter mildly in the 1-halo velocity term, as we saw in Section 4.4 that these scales are already affected by choices of the velocity profiles. As we explained in Section 3, there are some systematics that have not been removed in this study because the way they were eliminated in Tamone et al. (2020) would imply a big change for the definition and interpretation of the quadrupole. For this reason, we use $s_{2,\text{max}} = 25\,\text{Mpc}\,h^{-1}$, which is the scale at which the effect starts to appear.

We note that the scale choices previously mentioned can affect the results that we find in the subsections below, so these have to be interpreted carefully. In Appendix C, Supporting Information, we look at changes in the scale cuts, finding qualitatively similar results to the fiducial choices. In the absence of systematic errors one would not need to worry about choosing certain scales, and any redundant information would be accounted for by the covariance matrix. Hence, the study in Appendix C, Supporting Information must be understood as a consistency check in order to search for possible systematics not accounted for. We also prefer to choose shorter data vectors because with larger and more correlated data vectors the inversion of the covariance matrix becomes numerically noisier or even unfeasible by standard methods.

We now explore the HOD models by creating 27 $(1\text{Gpc}\,h^{-1})^3$ mock catalogues at each parameter space point, following a grid in parameter space that is refined at later iterations. We compute the different 2PCF of these mocks. Their mean will be our theory vector θ_{th} at a given point in the explored parameter space, for which we compute the χ^2 against the data vector θ_{data}:

$$\chi^2 = (\theta_{\text{th}} - \theta_{\text{data}})^T C^{-1} (\theta_{\text{th}} - \theta_{\text{data}}), \quad (39)$$

with C the covariance matrix of θ. Our best fit is then defined as the point in parameter space that minimizes χ^2.

In order to compute C, we use the 1000 EZMOCKS presented in Zhao et al. (2020). We use a version of the mocks that include the
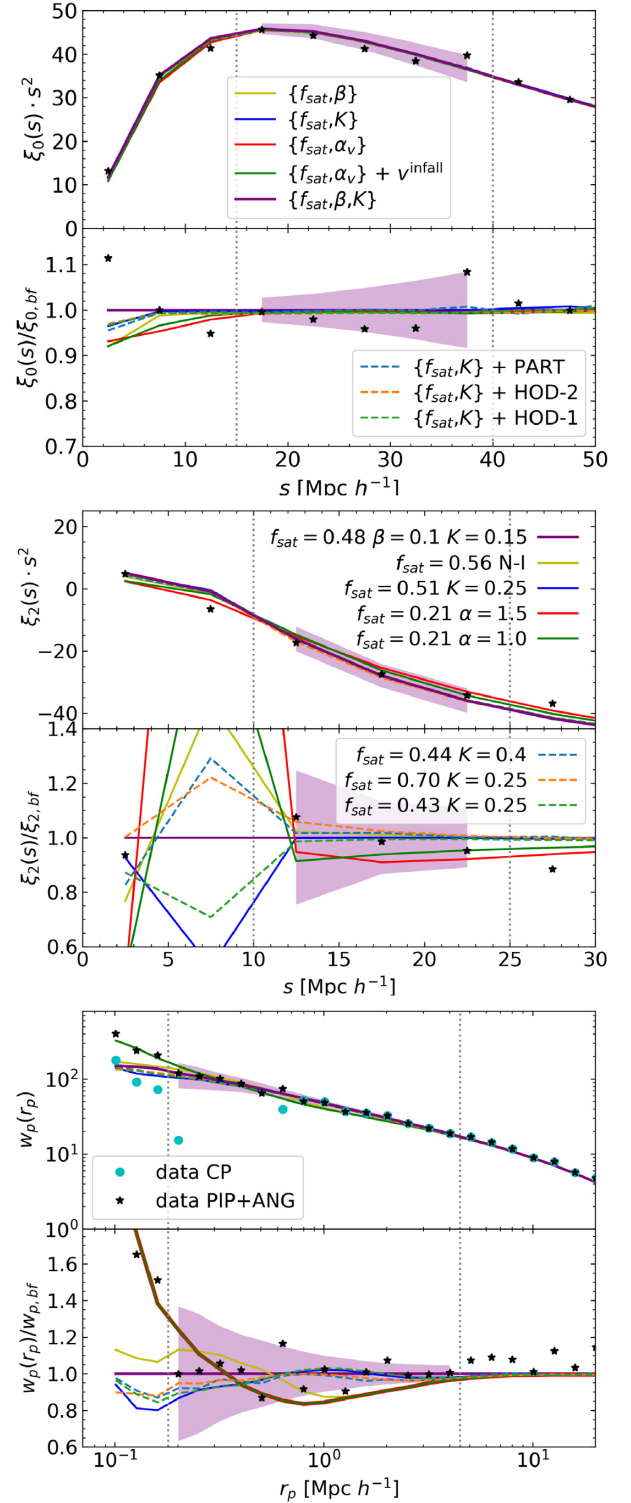

Figure 7. 2-point correlation functions (ξ_0, ξ_2, w_p) of the mocks that best fit the eBOSS data, under different assumptions mocks 1, 2, 3, 4, 6, 13, 16, and 11 of Table 3, as explained in Sections 5.2, 5.3, 5.4, and 5.5. The stars show the clustering of eBOSS data corrected with the PIP + ANG weights (Section 3.4). Points show the eBOSS data with the CP corrections instead of the ANG + PIP for w_p (the only case where differences are noticeable). We use as reference model the mock 11 for the ratios and the shaded area represents the error bars derived from the diagonal of the covariance matrix in equation (40). The vertical lines represent the interval of scales considered for the fits.

Table 3. List of best-fitting mocks under different assumptions. For each best fit, we write from left to right: a number in order to label it (Mock), the mean HOD choice (HOD), the fraction of satellites ($f_{\rm sat}$), the model parameters β, K, α_v, the profile choice, the choice of infall velocity, and the χ^2 of the mock with respect to the data for the projected correlation function, the quadrupole, the monopole, and the combined χ^2. Variables that are set free are in **bold**. For the details on the modelling see Section 4, and for a description of the fits see Section 5.

Mock	HOD	$f_{\rm sat}$	β	K	α_v	Profile	$v^{\rm infall}$	χ^2_{wp}	χ^2_2	χ^2_0	$\chi^2_{\rm tot}$ (bins: 14+3+5)
0	HOD-3	**0.22**	0	1	1	NFW	0	21.7	1.1	5.0	31.3
1	HOD-3	**0.56**	**N-I**	1	1	NFW	0	17	0.3	4.7	24
2	HOD-3	**0.51**	0	**0.25**	1	NFW	0	7.2	0.3	5	12.7
3	HOD-3	**0.21**	0	1	**1.5**	NFW	0	22	0.5	5.3	28.3
4	HOD-3	**0.21**	0	1	**1.0**	NFW	−500	22	0.5	5.0	28
5	HOD-3	**0.36**	**0.0**	1	1	PART	0	15.5	0.7	4.7	23
6	HOD-3	**0.44**	0	**0.4**	1	PART	0	8	0.3	4.6	13.5
7	HOD-3	**0.26**	0	1	**1.2**	PART	0	15	0.2	4.6	21.4
8	HOD-3	**0.26**	0	1	**0.8**	PART	−500	16	0.9	4.3	21.2
9	HOD-3	**0.48**	**0.10**	**0.15**	1	NFW	0	6	0.3	4.9	10.9
10	HOD-3	**0.21**	**0.0**	1	**1.5**	NFW	0	22	0.5	5.3	28.3
11	HOD-3	**0.51**	0	**0.25**	**1.0**	NFW	0	7.2	0.3	5	12.7
12	HOD-1	**0.40**	**N-I**	1	1	NFW	0	17.9	0.3	4.7	25
13	HOD-1	**0.43**	0	**0.25**	1	NFW	0	7	0.3	5.0	12.4
14	HOD-1	**0.18**	0	1	**1.6**	NFW	0	22	0.3	5.5	28.6
15	HOD-2	**0.70**	**N-I**	1	1	NFW	0	21	0.3	4.9	28.4
16	HOD-2	**0.70**	0	**0.25**	1	NFW	0	8.1	0.3	4.8	13.8
17	HOD-2	**0.22**	0	1	**1.5**	NFW	0	22	0.2	5.4	29.1

eBOSS geometry, but no observational systematics (in particular no fibre assignment). The effect of the observational systematics is expected to be minor compared to the need of rescaling explained below. We reanalyse these mocks with the OUTER RIM fiducial cosmology in order to follow the procedure used with the observational data. This changes the amplitude of the clustering statistics with respect to using the true cosmology of the EZMOCKS. There is also a miss-match in the amplitude of $w_{\rm p}$ at scales below the fibre collision scale regardless of the choice of cosmology. This is expected as the EZMOCKS were not tuned to match the clustering at highly non-linear scales. For this reason, we rescale the covariance matrix by

$$C_{ij} = C^{\rm EZ}_{ij} \frac{\theta^{\rm OR}_i \cdot \theta^{\rm OR}_j}{\theta^{\rm EZ}_i \cdot \theta^{\rm EZ}_j}, \qquad (40)$$

where $C^{\rm EZ}$ is the raw covariance matrix from the EZMOCKS, $\theta^{\rm EZ}$ the mean of the correlation function of the EZMOCKS, and $\theta^{\rm OR}$ the correlation function of one of our best-fitting OUTER RIM mocks (mock 9 in Table 3, with the lowest χ^2). This rescaling is derived from the assumption that the correlation matrix of the EZMOCKS, their relative uncertainty and the amplitude of the clustering of the OUTER RIM mock are all correct. We do not expect this choice of covariance matrix to affect strongly the best-fitting values. In fact, we also did the same analysis with a diagonal covariance inferred directly from the standard deviation of the 1Gpc h^{-1}-subboxes of mock 9 in Table 3, rescaled to the eBOSS volume (as done in Section 3.2 for the errorbars). For this case, we found results for the best fits similar to the ones shown here, but with artificially small χ^2 contours, given that no correlation between points was taken into account.

5.2 Baseline model

Our baseline model consists of mock catalogues with the HOD-3 shape (Section 4.1), satellite galaxies drawn from a Poisson distribution ($\beta = 0$, Section 4.2) following an NFW profile (with

$K = 1.0$, Section 4.3) and with virial velocities ($\alpha_v = 1$ and $v^{\rm infall} = 0$, Section 4.4). These are the default choices taken in Section 4, except that now we have not fixed $f_{\rm sat}$, which is set to vary following equation (20).

For all the models, we keep constant the linear bias, $b = b_{\rm eBOSS}$, and set the number density to $n = 7 \times n_{\rm eBOSS}$. This number density is the maximum before reaching $\langle N_{\rm cen} \rangle = 1.0$ for $f_{\rm sat} = 0$, where the model breaks down (i.e. when the number of central galaxies becomes larger than 1). Having a larger number density reduces the noise in our theoretical model, θ, hence reducing the noise in the inferred best fit and χ^2 contours.

The best fit from this baseline model is the first entry in Table 3, mock 0. It has $f_{\rm sat} = 0.22^{+0.02}_{-0.03}$, with error bars representing the $\Delta\chi^2 = 1$ interval. The best value found for $f_{\rm sat}$ is close to that found by Guo et al. (2019), 13–17 per cent, where an earlier version of the eBOSS ELG sample was analysed with the incomplete conditional stellar mass function model presented in Guo, Yang & Lu (2018).

The baseline model gives a poor fit to the observational data, so we explore other alternatives below.

5.3 Baseline + 1 parameter model

In this subsection, we relax the baseline model by allowing an extra degree of freedom. The change is introduced, one at a time, in one of the three following aspects: (i) a non-Poissonian PDFs for satellite galaxies, (ii) rescaling the satellite density profile by a factor K, and (iii) modifying the velocity dispersion by a factor α_v. Additionally, for the α_v model we also consider a separate case with a net infall velocity, $v^{\rm infall}$. The results are summarized in the second tier of Table 3, mocks 1–4, and in Fig. 8. This figure shows the χ^2 as a function of $f_{\rm sat}$ and another variable (β, K, or α_v) for $w_{\rm p}$, ξ_2 and the combination of $\{w_{\rm p}, \xi_2, \xi_0\}$.

In the case of modifying the PDF of satellite galaxies, the nearest integer is represented with a negative β in Fig. 8 (with an arbitrary value of $\beta = -0.1$), the Poisson distribution with $\beta = 0$ and negative
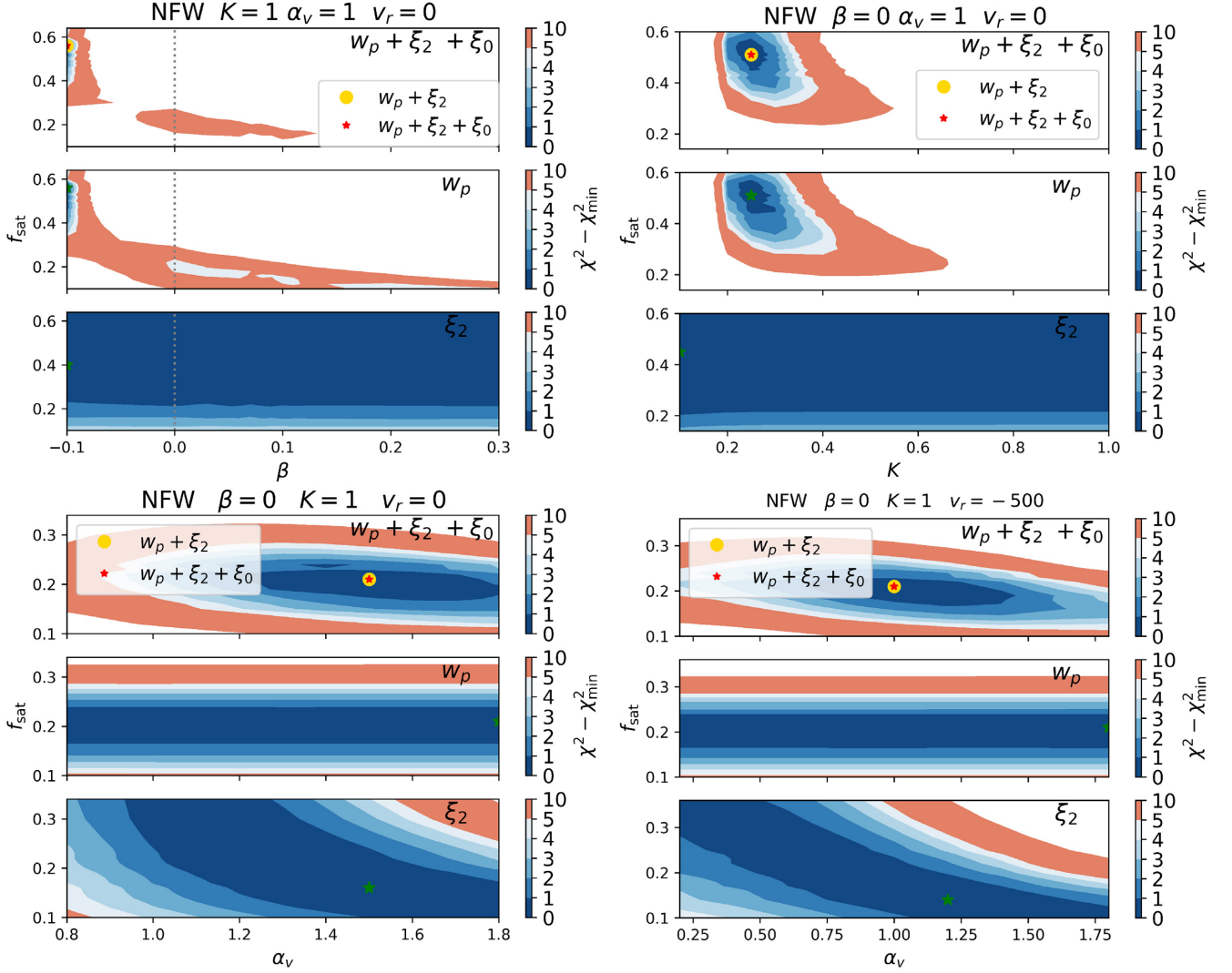
Figure 8. χ^2 contours for the baseline + one free parameter (different in each subfigure) model. *Top left:* β is set free, with $\beta = 0$ representing the Poisson distribution, $\beta > 0$ a negative binomial distribution, and $\beta = -0.1$ (arbitrary choice for the representation) the nearest integer distribution (see Section 4.2). *Top right:* The concentration of the satellite profiles are rescaled by a free parameter K, $c \to K \cdot c$ (see Section 4.3). *Bottom left:* The velocity dispersion is allowed to vary, $\sigma_{\rm vir} \to \alpha_v \sigma_{\rm vir}$ (see Section 4.4). *Bottom right:* An infall velocity component is added, $v^{\rm infall} = -500 \pm 200 \, {\rm km \, s}^{-1}$, while still letting α_v free. Within each subfigure, we show the χ^2 component of the quadrupole (bottom panel), projected correlation function (middle panel), and the combination full fit to the monopole, quadrupole, and $w_{\rm p}$ (top panel, scales are defined in equation 38). Except for the variations specified in each subfigure, we assume the baseline model described in Section 5.2: HOD-3, NFW profile, $\beta = 0$, $K = 1$, $\alpha_v = 1$, and $v^{\rm infall} = 0$. The models shown here assume, $n = 7 \times n_{\rm eBOSS}$. The filled stars show the minimum χ^2 for the statistics used in each subfigure. In the top panels, we also include a filled circle for the minimum χ^2 if we exclude the monopole of the fit.

binomial distributions with $\beta > 0$. Remarkably, the best fit shows a preference for low scatter, with the nearest integer PDF, and a large satellite fraction ($f_{\rm sat} = 0.56$). We highlight that there is not a smooth transition between a Poisson and a nearest integer PDF in terms of scatter, resulting in the *break* that appears at $\beta = 0$ in the top left subfigure of Fig. 8.[6] In line with the effects found in Section 4.2, ξ_2 does not constrain β, but does set a lower limit for $f_{\rm sat}$ (with a very slight degeneracy with β). $w_{\rm p}$ constrains β to be small, with a notable degeneracy with $f_{\rm sat}$. A preference for a sub-Poissonian distribution, $\beta < 0$, is at odds with the results from SAMs (Jiménez

et al. 2019). However, as we will show in the following subsections, our best-fitting values for β depend on other model assumptions.

When we vary the profile concentration, setting K free (see Section 4.3), we find a similar effect, with ξ_2 only constraining $f_{\rm sat}$ and $w_{\rm p}$ driving the main constraints. We find the best fit for $K = 0.25$ (with a step of 0.05 in the parameter space grid) and $f_{\rm sat} = 0.52$, clearly favouring profiles less concentrated than NFW, in line with previous studies (see references in Section 4.3).

In the bottom left of Fig. 8, we show what happens when allowing for a velocity bias α_v. In this case, $w_{\rm p}$ is the quantity that is insensitive to the choice of α_v (in lines with Fig. 6), constraining only $f_{\rm sat}$. ξ_2 shows a strong degeneracy between $f_{\rm sat}$ and α_v. When combining both, we find the best fit at $f_{\rm sat} = 0.21$ and $\alpha_v = 1.5$. For the NFW profiles, α_v represents a deviation from the galaxy velocity dispersion found in Bryan & Norman (1998). Hence, the observational data prefers an enhanced velocity dispersion within this model.

---

[6]We use the CONTOURF function from the PYTHON library MATPLOTLIB for these figures. This function interpolates the colour between discrete values. This is why the *break* appears below $\beta = 0$.

Building upon the preference for a larger velocity dispersion, we also include a net infall velocity of $v^{\text{infall}} \sim -500 \, \text{km s}^{-1}$ (see Section 4.4), letting α_v to also vary. The bottom right-hand panel of Fig. 8 shows similar results to the previous case, but with a preference for lower α_v values. The constraints from w_p remain the same and those from ξ_2 shift and get distorted. In this case, we get a best fit of $\alpha_v = 1.0$ and the same fraction of satellites $f_{\text{sat}} = 0.21$. This suggest that including $v^{\text{infall}} \sim -500 \, \text{km s}^{-1}$ enhances the finger-of-god effect, in a similar way as with $\sim 0.5\sigma_{\text{vir}}$. We note again that the modelling for v^{infall} adopted here is more extreme than that of Orsi & Angulo (2018).

Of the four extensions to the baseline model considered here, the modification of the satellite density profile (with the concentration controlled by K) yields the best fit to the data, with a reduced χ^2 below unity (we do not show explicitly the reduced χ^2 as it can be derived from the data already provided in Table 3). We consider combinations of these extensions in Section 5.5.

5.4 Particles + 1 parameter model

In this section, we repeat the variations described in Section 5.3, but using the particle position and velocity profiles, PART, instead of the NFW profile and virial theorem velocities. A lower density $n = 1 \times n_{\text{eBOSS}}$ is used, as the computation is much more demanding in this case. With this choice, we also minimize the cases for which we run out of particles. However, this choice increases the noise in the modelling of θ, giving, in turn, noisier contours.

The results are summarized in Fig. 9 and the third tier of Table 3, mocks 5–8. The best fits for these models, follow roughly the results described in Section 5.3, with some differences:

(i) In the $\{f_{\text{sat}}, \beta\}$ plane, this time the data prefer the Poisson PDF ($\beta = 0$), together with a lower fraction of satellites, following a similar degeneracy as that seen in Section 5.3.

(ii) Results in the other planes ($\{f_{\text{sat}}, K\}$, $\{f_{\text{sat}}, \alpha_v\}$, $\{f_{\text{sat}}, \alpha_v + v^{\text{infall}}\}$) show qualitatively similar results to the NFW case.

(iii) The best PART fit give $K = 0.4$, which, although closer to the baseline model ($K = 1$) than the NFW case, is clearly preferring less concentrated profiles, $K < 1$. The differences in the profiles, seen in the top panel of Fig. 5, change the preferred value of K. The satellite fraction is also somewhat lower than for the NFW case.

(iv) The best fit α_v gets shifted by $\Delta\alpha_v \sim -0.3$ (-0.2 for the v^{infall} case) when using PART profiles.

(v) For the PART models, α_v corresponds to the definition of galaxy velocity bias: the ratio of the velocity dispersion of galaxies to the one of dark matter particles. The value found here, $\alpha_v = 1.2$, is compatible with that found in subhaloes in simulations (Diemand, Moore & Stadel 2004).

(vi) For those models with α_v set as a free parameter (with or without v^{infall}), the PART models give better fits than for the NFW ones. For the other cases, the χ^2 has similar values.

As for both NFW, with PART profiles we find the effect of v^{infall} to be mostly equivalent to a shift in α_v. Thus, we will not explore further the $v^{\text{infall}} = -500$ case in the following subsections or in Appendix C, Supporting Information.

5.5 Baseline + 2 parameter mode

We keep increasing the level of complexity of the model by setting free f_{sat} and two other parameters, while fixing the rest to the default choices, including an NFW profile. We set free at once the following groups of parameters: $\{f_{\text{sat}}, \beta, K\}$, $\{f_{\text{sat}}, \beta, \alpha_v\}$, and $\{f_{\text{sat}}, \alpha_v, K\}$. The

results from these fits are summarized in Fig. 10 and the fourth tier of Table 3, mocks 9–11. We highlight some of the main results:

(i) When both β and K are free, the data prefer positive β and K gets even smaller ($K = 0.15$, see Table 3).

(ii) $\{f_{\text{sat}} = 0.48, K = 0.15, \beta = 0.1\}$ gives the best fit of all models considered.

(iii) When both β and α_v are set free, the data prefer a Poisson distribution (unlike when only β was free, Section 5.2) and we recover the $\alpha_v = 1.5$ found in Section 5.2 (for α_V free).

(iv) When both K and α_v are varied, we recover the $K = 0.25$ case (like in Section 5.2) and the data prefer a model with the fiducial virial theorem velocity, $\alpha_v = 1$.

(v) A general result is that a low K (a less concentrated profile than NFW) is necessary to obtain a good fit to the data.

5.6 HOD variations + 1 parameter

So far in this section, we have used the HOD-3 model in all cases. In this subsection, we use the HOD-1 and HOD-2 functions (equations 16, 18, and Table 2) for the mean HODs, and explore the parameter space by leaving f_{sat} and one parameter (β, K, or α_v) free. The results are summarized in the last 2 tiers of Table 3, mocks 12–7, and Fig. 11.

We find that the choice of mean HOD changes the preferred fraction of satellites. In the case of the HOD-1, f_{sat} is always found lower than for models assuming the HOD-3. This can easily be explained by the behaviour of the w_p presented in Fig. 3. For the same value of $f_{\text{sat}} = 0.30$, the w_p 1-halo term is larger for HOD-1 and lower for HOD-2. Hence, with respect to HOD-3, HOD-1 models will need a lower fraction of satellite, and HOD-2 a higher one, to match the w_p from the data.

Remarkably, the best-fitting value of the rest of parameters (β, K, or α_v) and the overall shape of the χ^2-contours remain unchanged. This implies that the HOD shape is not degenerated with any parameter other than f_{sat}. Additionally, the χ^2 shown in Table 3 mildly disfavours HOD-2.

5.7 Best fits

Finally, we compare the correlation functions of the data to the best fit mocks in several of the most representative scenarios assumed in past subsections. Results are shown in Fig. 7 with the lowest χ^2 mock (9 in Table 3) as a reference.

All the best fits shown in Fig. 7 reproduce similarly well the monopole of the data over the range of selected scales. We also find that the quadrupoles from all the mocks shown agree with the data, within the error bars. Nevertheless, we find a differentiated shape for the quadrupole for the two fits (with and without v^{infall}) with $f_{\text{sat}} = 0.21$ and fitted α_v.

When studying the projected correlation function, the differences between mocks are clearer partially because for this statistics we explore a larger range of scales. Here, we see clearly that only models with $K < 1$, i.e. with profiles less concentrated than the NFW/PART default one, can fit well small scale data. We can see the differentiated shape induced in the w_p for the different ingredients for the HOD (e.g. β versus K versus α_v). We note that the HOD shape does not change much the w_p once the f_{sat} is free to compensate for the excess or deficit of small scale clustering. We also see the importance of using the PIP + ANG weights at scales $r_p < 1 \, \text{Mpc} \, h^{-1}$, with the traditional CP weights
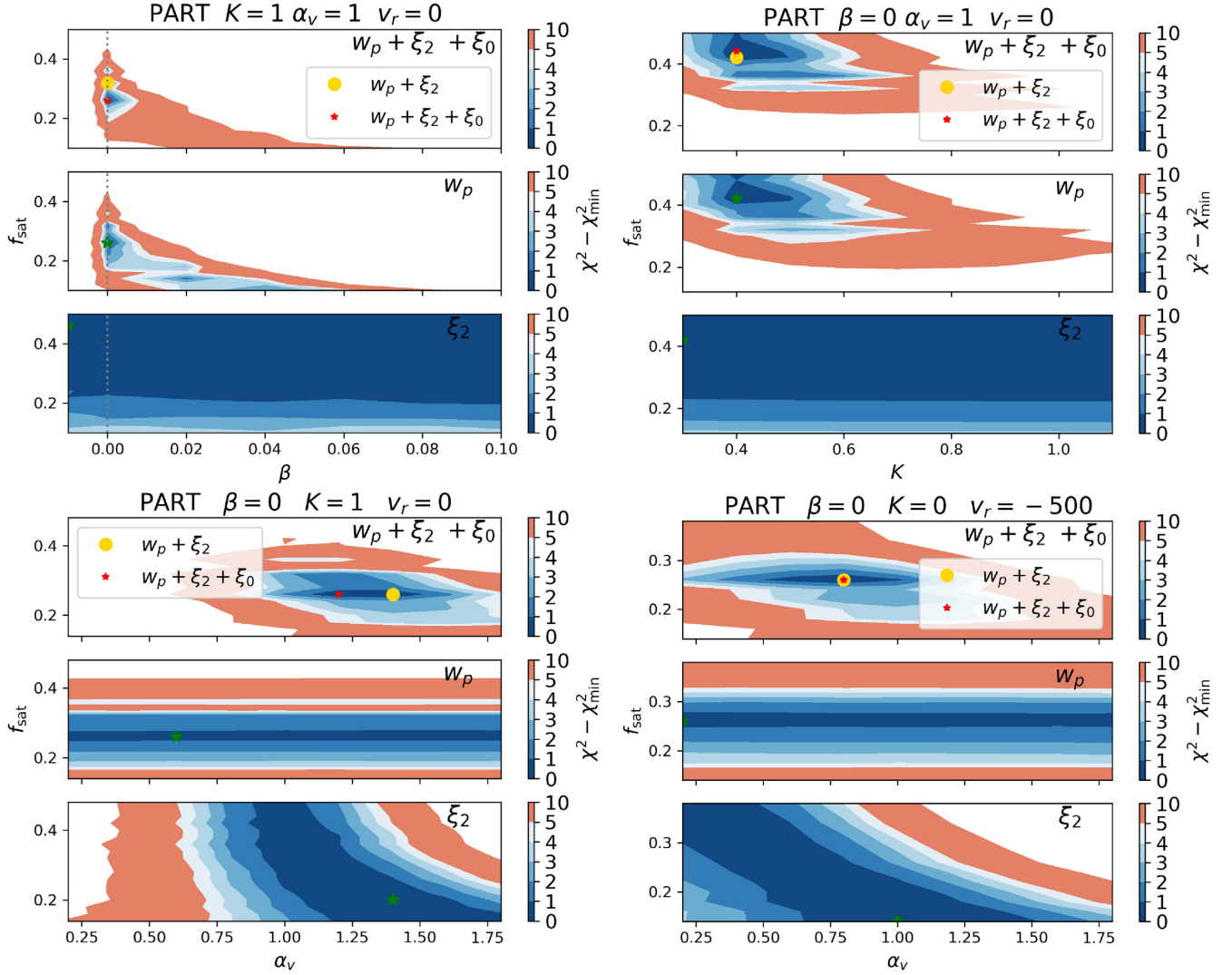
Figure 9. χ^2 contours for the particle $+ f_{sat} +$ one free parameter model. We follow the same structure as in Fig. 8, but using the particles profiles and velocities as a starting point for the models. See Sections 4.3 and 4.4 for the details of the differences in the modelling. Note that this figure is made from mocks with lower number density, $n = 1 \times n_{eBOSS}$, yielding noisier contours.

resulting in a lower clustering (more details in Mohammad et al. 2020).

6 CONCLUSIONS

In this paper, we study a series of HOD models for star-forming ELGs motivated by theoretical and observational studies in the literature. We create mock ELG catalogues using OUTER RIM simulation haloes at $z = 0.865$. This is one of the largest ($L = 3$ Gpc h^{-1}) existing dark matter-only N-Body simulations within its mass resolution range, $m_p \sim 2 \cdot 10^9 M_\odot \, h^{-1}$. Throughout this study, we fix the galaxy bias of the mock ELGs to that observed in the eBOSS data catalogue and we take its number density, $n = n_{eBOSS}$, as a reference. We use $n = n_{eBOSS}$ for error estimations or reporting HOD parameters, and $\times 7$ to 10 higher density when computing clustering signal at a given point of the parameter space. We make the mock catalogues available so they may be used for model testing in preparation for future surveys.[7]

We revisit the HOD model for the case of ELGs, reconsidering most of the assumptions that go into it. We consider different shapes of the mean HOD for central galaxies, $\langle N_{cen}(M) \rangle$: from the classical smoothed step function (HOD-1), through a model with decaying occupation for higher masses (HOD-3) up to a model with no occupation at large masses (HOD-2). We set our default choice to HOD-3 (Section 4.1) with a piecewise Gaussian plus a power law that best fits the results from the SAM of galaxy formation and evolution presented in Gonzalez-Perez et al. (2018). For the mean HOD of satellite galaxies, $\langle N_{sat}(M) \rangle$, we always use a power law, as is typically done in the literature. We allow three HOD parameters to vary in order to control the number density n, the bias b, and the fraction of satellites f_{sat}, while fixing the scaling relations for the other parameters (see Table 2). We highlight that one basic need to match the number density and the large-scale bias of a ELG sample is to account for the incompleteness in stellar mass of the ELG central galaxy sample (see also Favole et al. 2016). This is common to other samples clearly incomplete in stellar mass such as QSOs (Lyke et al. 2020).

We do not focus exclusively on the functions and parameters that control the shape of the mean HOD, but also on the other choices that need to be made when populating haloes with satellite galaxies.
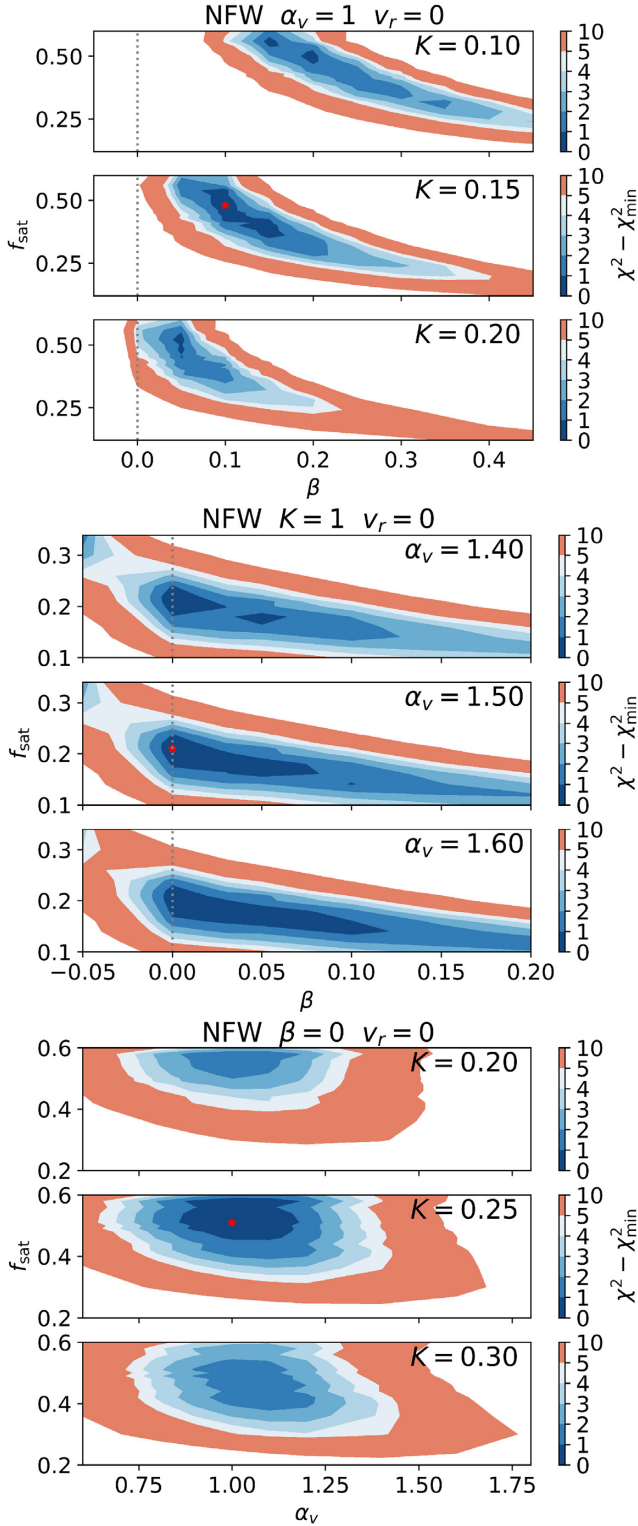
Figure 10. χ^2 contours for the baseline + 2 parameters model. These two parameters are $\{\beta \,\& \, K\}$, $\{\beta \,\& \, \alpha_v\}$, and $\{K \,\& \, \alpha_v\}$ for the top, middle, and bottom subfigures, respectively. In this figure, we only show the combined $(\{\xi_0, \xi_2, w_p\})$ χ^2. Within subfigures, panels are used to represent an extra dimension, which is set constant within the panel, and varied across panels. The star represents the minimum of the χ^2.

The first of these choices is the PDF $P(N_{\mathrm{sat}}|\langle N_{\mathrm{sat}}\rangle)$. We consider a Poisson distribution as our default ($\beta = 0$), but we also consider a negative binomial with greater scatter ($\beta > 0$) and a nearest integer function ($\beta < 0$).

We place satellite galaxies either following an NFW profile or the particle distribution within haloes. We allow for a rescaling of the profiles as we expect ELGs to follow different profiles than dark matter. We assign velocities either using the virial theorem for the NFW profiles or simply the particle velocities. On top of that we allow for a velocity bias and for the inclusion of a net infall velocity.

With different combinations of the above choices, we construct a range of HOD models. We study how each of these models affect the clustering of mock galaxies, mainly via the projected correlation function w_p and the quadrupole ξ_2. We find that these statistics help separating different effects. Although the fraction of satellites affects both statistics, the PDF, and the position assignment affect mostly the projected correlation. The velocity assignment mostly affects the quadrupole. We also studied the monopole, but it shows nearly no variations on the linear scales because we fixed the bias to that of the data.

In Section 5, we fit to the eBOSS ELG data, mocks produced with different HOD models. Some general findings are as follows:

(i) In all the analysed scenarios, the observational data prefers dispersed profiles for ELGs ($K = 0.15 - 0.4$).

(ii) We also find a mild preference (lower χ^2) for particle profiles as opposed to NFW ones. However, this preference goes away if we let K vary.

(iii) We find some preference for a positive velocity bias ($\alpha_v > 1$), i.e. a larger velocity dispersion of satellite galaxies around the halo velocity, although once K is set free we recover the $\alpha_v = 1.0$ scenario (no velocity bias).

(iv) The PDF preference depends a lot on the rest of assumptions. We find that negative, positive, and vanishing β are preferred in different scenarios.

(v) The shape of the mean HOD does have some effect on the clustering but can be mostly compensated by increasing or decreasing the fraction of satellites. After that change in f_{sat}, the effect of the HOD shape is found subdominant. The data slightly disfavour HOD-2.

(vi) For HOD-3, the fraction of satellites found to match the clustering vary from $f_{\mathrm{sat}} \sim 0.21$, for the cases where both K and β are fixed to their default values (1 and 0, respectively), to $f_{\mathrm{sat}} \sim 0.50$ when either of those parameters are let vary. In the latter case, f_{sat} rises to 0.70 when assuming HOD-2 and it goes down to ~ 0.40 for HOD-1. The change of profile (to PART) also affects f_{sat}.

(vii) The key ingredient to match the data seems to be the profile rescaling with a factor K, $c \to K \cdot c$ (see Section 4.3).

We find that small-scale clustering strongly depends on the details of how we place satellite galaxies within haloes. These details may be more relevant than the shape of the mean HOD, which is the quantity many studies in the literature put the focus on.

The general results obtained here, such as ELGs distributing in more disperse profiles than NFW, are expected to also be applicable to star-forming galaxies at the studied redshifts, independently of their particular selection. Thus, this work is relevant for DESI, that will select ELGs also based on their [O II] flux, but also Euclid and other surveys targeting star-forming galaxies at $z \sim 1$ in different ways from eBOSS.

This study shows what scenarios of the ELG–dark matter relation are preferred or ruled out by the observational data. These findings have implications for the modelling of physical processes that shape the formation and evolution of ELGs. Studies like this one, can give
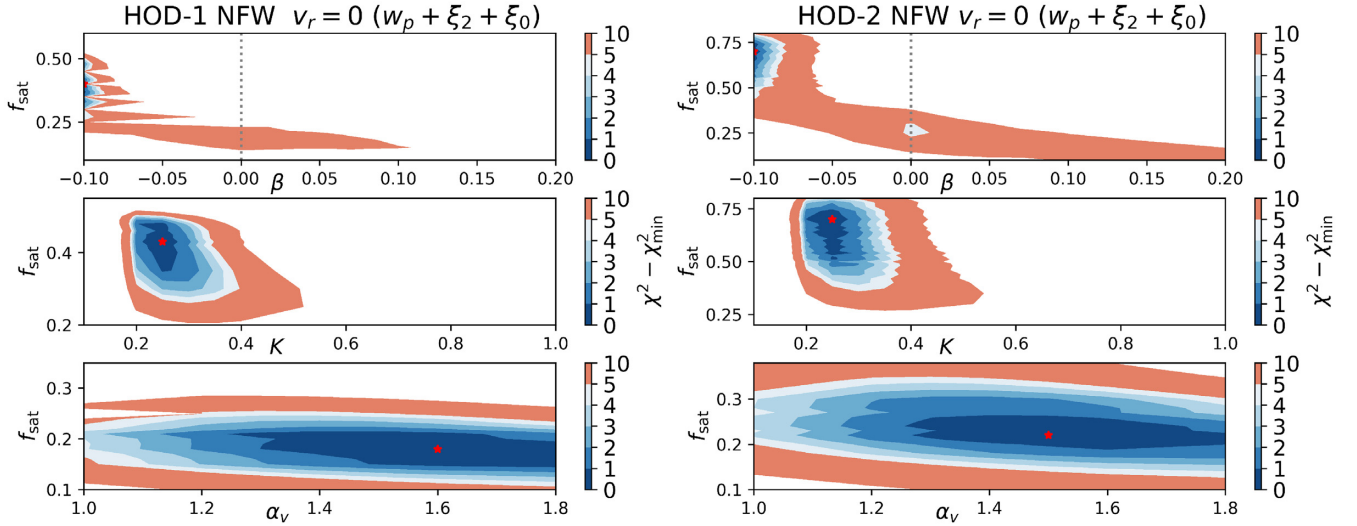
Figure 11. χ^2 contours of the HOD-1 (left subfigure) and HOD-2 (right subfigure) mocks, constructed varying f_{sat} and one other parameter (β, K, or α_v as indicated in each panel). The rest of default choices are the same as in Fig. 8. All the panels in this Figure show results for the full data vector $w_p + \xi_2 + \xi_0$.

us a unique insight of the physics of galaxy formation and evolution of ELGs. Such study could also provide information on other samples that will be available with current and future cosmological surveys.

In this study, we did not include galaxy assembly bias, i.e. the dependence of the galaxy clustering on properties other than the halo mass. This is an effect widely seen in model galaxy (e.g. Zehavi et al. 2018). Although several observational studies have concluded that galaxy assembly bias is not a strong source of systematic uncertainty (Tinker, Wetzel & Conroy 2011; Walsh & Tinker 2019), others have found evidence of galaxy assembly bias (Obuljen, Percival & Dalal 2020). Exploring such effect is beyond the scope of this work partly because we only had access to limited information of the FoF halo catalogue. Additionally, other approaches, such as subhalo abundance matching, might be more adequate for such purpose (e.g. Contreras, Angulo & Zennaro 2020). Another effect that has not been considered in this study is the conditional probability of N_{sat} on whether or not the central galaxy is an ELG. This is known as galactic conformity and has been studied elsewhere (e.g. Kauffmann et al. 2013; Hearin, Watson & van den Bosch 2015; Lacerna et al. 2018; Alam et al. 2019).

In this study, we have not explored the change of the selection function and galaxy evolution within the redshift range of this ELGs sample. The results from Guo et al. (2019) indicate that the variation in number density at different redshifts has the largest effect on the derived eBOSS ELG mean HODs. This suggests that the shape and, likely, distribution of satellite galaxies, could be maintained, while adjusting the target number density. Similar results are obtained for model ELGs (Gonzalez-Perez et al. 2018). We defer to the future a more in-depth study of the evolution of ELG samples.

Our results could be sensitive to the choice of fiducial cosmology. In order to assess that, we would need other OUTER RIM-size simulations at different cosmologies. This is something beyond the scope of this project, but stage-IV surveys already in preparation might need to consider this.

The eBOSS ELG program, with the largest ELG sample to date serves as a bridge from stage-III to stage-IV experiments, where ELGs will be crucial. ELGs probe, on average, lower halo masses compared to LRGs and have a more complicated selection function. This posed the question whether ELGs would have in turn a complicated relation to dark matter that could have implications when

interpreting their anisotropic clustering to understand cosmology. This study probes a very wide variety of plausible scenarios within our current knowledge of ELG formation and evolution. The mocks presented here have been analysed following the same procedure used to derived the eBOSS ELG BAO + RSD measurements (de Mattia et al. 2020; Raichoord et al. 2020; Tamone et al. 2020; see Appendix B and Alam et al. 2020 for more details), finding no evidence of any bias on the derived parameters within the statistical errors provided by the set-up of OUTER RIM simulation, which is much lower than the eBOSS uncertainties.

If we want to extract the full cosmological potential of future surveys, we will need to consider smaller and smaller scales in the analysis. Studies like this one will need to be carried out in order to validate the correct extraction of cosmological information and to test ways to disentangle cosmology from baryon physics when interpreting galaxy clustering.

DATA AVAILABILITY

A selection of mocks, including those used for Figs 3, 4, 5, and 6, is currently available here: http://popia.ft.uam.es/eBOSS_ELG_OR_mocks. This data base was compiled by the corresponding authors and Alexander Knebe. Other mocks may be provided upon request.

REFERENCES

Alam S. et al., 2017, MNRAS, 470, 2617
Alam S., Peacock J. A., Kraljic K., Ross A. J., Comparat J., 2019, MNRAS, 497, 581
Alam S. et al., 2020, MNRAS, submitted
Alonso D., 2012, MNRAS, 497, 581
Avila S. et al., 2018, MNRAS, 479, 94
Bautista J. E. et al., 2020, MNRAS, tmp 2651B (staa2800)
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Berlind A. A. et al., 2003, ApJ, 593, 1
Bianchi D., Percival W. J., 2017, MNRAS, 472, 1106
Blanton M. R. et al., 2017, AJ, 154, 28
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Carretero J., Castander F. J., Gaztañaga E., Crocce M., Fosalba P., 2015, MNRAS, 447, 646
Chen Y.-C. et al., 2017, MNRAS, 466, 1880
Child H. L., Habib S., Heitmann K., Frontiere N., Finkel H., Pope A., Morozov V., 2018, ApJ, 859, 55
Cochrane R. K., Best P. N., 2018, MNRAS, 480, 864
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
Comparat J. et al., 2013, MNRAS, 428, 1498
Comparat J. et al., 2016, A&A, 592, A121
Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
Contreras S., Angulo R., Zennaro M., 2020, preprint (arXiv:2005.03672)
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Darvish B., Sobral D., Mobasher B., Scoville N. Z., Best P., Sales L. V., Smail I., 2014, ApJ, 796, 51
Davies L. J. M. et al., 2019, MNRAS, 483, 1881

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dawson K. S. et al., 2016, AJ, 151, 44
de Mattia A. et al., 2020, MNRAS, submitted
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Diemand J., Moore B., Stadel J., 2004, MNRAS, 352, 535
du Mas des Bourboux H. et al., 2020, ApJ, 901, 153
Einasto J., 1965, Tr. Astrofiz. Inst. Alma-Ata, 5, 87
Favole G. et al., 2016, MNRAS, 461, 3421
Favole G. et al., 2019, MNRAS, 497, 5432
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, MNRAS, 387, 536
Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, MNRAS, 426, 679
Gil-Marin H. et al., 2020, MNRAS, 498, 2492
Gonzalez-Perez V. et al., 2018, MNRAS, 474, 4024
Gonzalez-Perez V. et al., 2020, MNRAS, 498, 1852
Gunn J. E. et al., 2006, AJ, 131, 2332
Guo H., Yang X., Lu Y., 2018, ApJ, 858, 30
Guo H. et al., 2019, ApJ, 871, 147
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Hearin A. P., Watson D. F., van den Bosch F. C., 2015, MNRAS, 452, 1958
Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, MNRAS, 460, 2552
Hearin A. P. et al., 2017, AJ, 154, 190
Heitmann K. et al., 2015, ApJS, 219, 34
Heitmann K. et al., 2019, ApJS, 245, 16
Hernández-Aguayo C., Prada F., Baugh C. M., Klypin A., 2020, MNRAS, preprint (arXiv:2006.00612)
Hou J. et al., 2020, preprint (arXiv:2007.08998)
Jackson J. C., 1972, MNRAS, 156, 1P
Jiménez E., Contreras S., Padilla N., Zehavi I., Baugh C. M., Gonzalez-Perez V., 2019, MNRAS, 490, 3532
Kaiser N., 1987, MNRAS, 227, 1
Kauffmann G., Li C., Zhang W., Weinmann S., 2013, MNRAS, 430, 1447
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Komatsu E. et al., 2011, ApJS, 192, 18
Kraljic K. et al., 2018, MNRAS, 474, 547
Lacerna I., Contreras S., González R. E., Padilla N., Gonzalez-Perez V., 2018, MNRAS, 475, 1177
Lacey C., Cole S., 1993, MNRAS, 262, 627
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lin S. et al., 2020, MNRAS, 498, 5251
Linder E. V., 2005, Phys. Rev. D, 72, 043529
Lyke B. W. et al., 2020, ApJS, 250, 9
Mohammad F. G. et al., 2020, MNRAS, 498, 128
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neveux R. et al., 2020, MNRAS, 499, 210
Obuljen A., Percival W. J., Dalal N., 2020, MNRAS, 498, 128
Okada H. et al., 2016, PASJ, 68, 47
Orsi Á. A., Angulo R. E., 2018, MNRAS, 475, 2530
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe, Princeton University Press
Percival W. J., Bianchi D., 2017, MNRAS, 472, L40
Raichoor A. et al., 2016, A&A, 585, A50
Raichoor A. et al., 2017, MNRAS, 471, 3955
Raichoor A. et al., 2020, MNRAS, submitted
Ross A. J. et al., 2020, MNRAS, 498, 2354
Rossi G. et al., 2020, MNRAS, preprint (arXiv:2007.09002)
Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624
Seljak U., 2000, MNRAS, 318, 203
Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022
Smee S. A. et al., 2013, AJ, 146, 32
Smith A. M. J. et al., 2020, MNRAS, 499, 269

Tamone A. et al., 2020, MNRAS, tmp, 2857

Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522

Taruya A., Bernardeau F., Nishimichi T., Codis S., 2012, Phys. Rev. D, 86, 103528

Tinker J., Wetzel A., Conroy C., 2011, preprint (arXiv:1107.5046)

Walsh K., Tinker J., 2019, MNRAS, 488, 470

Wang J., Bose S., Frenk C. S., Gao L., Jenkins A., Springel V., White S. D. M., 2019, Nature, 585, 39

Zehavi I. et al., 2005, ApJ, 630, 1

Zehavi I. et al., 2011, ApJ, 736, 59

Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, ApJ, 853, 84

Zhao C. et al., 2020, submitted

Zheng Z. et al., 2005, ApJ, 633, 791

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Appendix_C.pdf

Please note: Oxford University Press is not responsible fobr the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: CLUSTERING IN FOURIER SPACE

For completeness, in this Appendix, we analyse the clustering of the mock catalogues presented in Section 4 (Figs 3, 4, 5, and 6) in Fourier space (Figs A1, A2, A3, and A4). The power spectrum
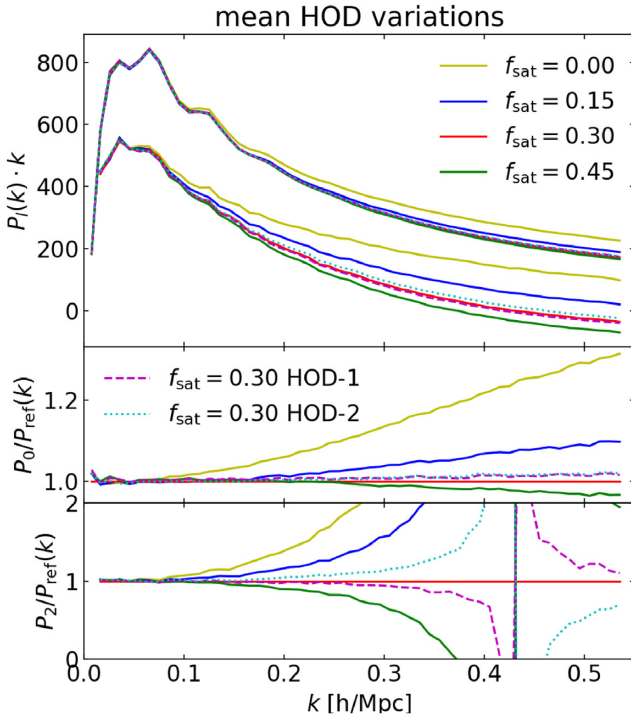

Figure A1. Power Spectrum multipoles of mean occupation halo variations: HOD shapes (HOD-1, HOD-2, HOD-3), and fraction of satellites (f_{sat}). These correspond to the same mocks as shown in Fig. 3. In the top panel, we present both the monopole (upper set of lines) and quadrupole (lower set of lines). In the middle/lower panel, we show the ratio of the monopole/quadrupole to the reference model: {HOD-3, $f_{sat}=0.30$, NFW, $K=1$, $\beta=0$, $\alpha_v=1$}.
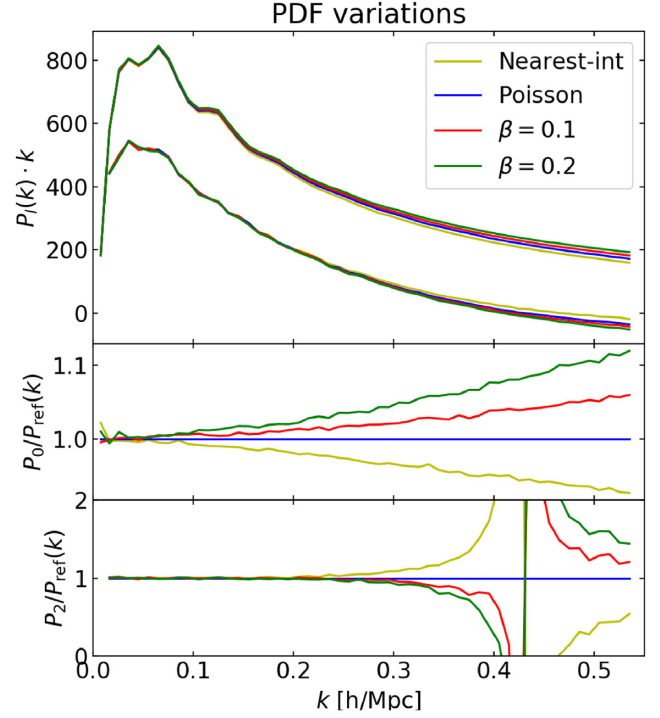

Figure A2. Power Spectrum multipoles of mocks with different Point Distribution Function Variations: nearest integer, Poisson ($\beta=0$) or Negative Binomial ($0 < \beta < 1$)). These correspond to the mocks shown in Fig. 4. In the top panel, we present both the monopole (upper set of lines) and quadrupole (lower set of lines). In the middle/lower panel, we show the ratio of the monopole/quadrupole to the reference model: {HOD-3, $f_{sat}=0.30$, NFW, $K=1$, $\beta=0$, $\alpha_v=1$}.
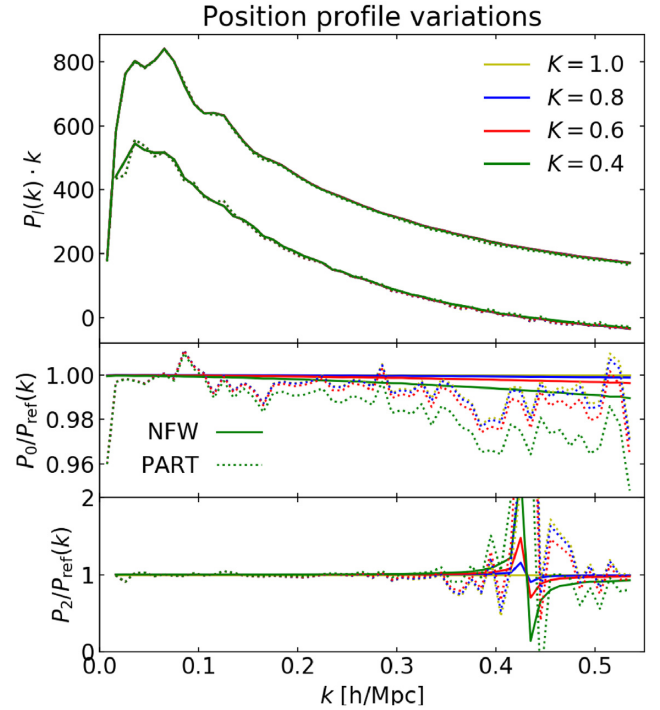

Figure A3. Power Spectrum multipoles of mocks with different position assignment (NFW versus PART and different values of K). These correspond to the same mocks as shown in Fig. 5. In the top panel, we present both the monopole (upper set of lines) and quadrupole (lower set of lines). In the middle/lower panel, we show the ratio of the monopole/quadrupole to the reference model: {HOD-3, $f_{sat}=0.30$, NFW, $K=1$, $\beta=0$, $\alpha_v=1$}.
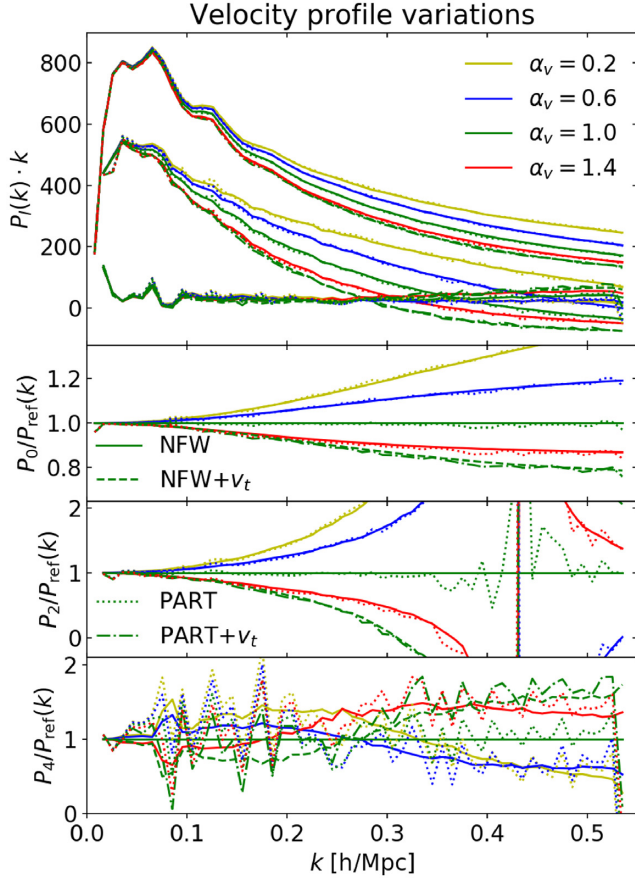
Velocity profile variations

Figure A4. Power Spectrum multipoles of mocks with different velocity assignment (NFW versus PART, different values of α_v and optionally $v_t = v^{\text{infall}}$) as indicated in the legend. Due to visualization purposes, we show slightly different combinations of choices with respect to Fig. 6. In the top panel, we represent the monopole (upper set of lines), quadrupole (middle set of lines), and hexadecapole (lower set of lines). In the lower panels, we show the ratio of the monopole (P_0), quadrupole (P_2) and hexadecapole (P_4) to the reference model: {HOD-3, f_{sat}=0.30, NFW, $K = 1$, $\beta = 0$, $\alpha_v = 1$}.

multipoles are computed in the $L = 3$ Gpc h^{-1} box using periodic conditions and redshift space distortions along the Z-axis. We use the NBODYKIT code (Hand et al. 2018) to compute the power spectrum multipoles $P_\ell(k)$ using a grid size of 512^3 and Triangular-Shape-Cloud mass assignment together with interlacing (Sefusatti et al. 2016). This configuration gives a Nyquist frequency of $k_{\text{Ny}} = 0.54h\,\text{Mpc}^{-1}$. We refer to de Mattia et al. (2020) for further details.

Qualitatively, the results are similar to those found in Section 4, but in some cases the information is spread differently in k space. We find that the effects that did not change the quadrupole in configuration space, ξ_2, have a very small effect in P_0 and P_2 at the scales shown here. This is very clear in Fig. A3, where profile variations barely change the power spectrum multipoles. In those lines, in Fig. A2, we find a mild effect of the PDF on to the multipoles. On the other hand, the effects that did change the ξ_2 have a strong effect on P_2 but are also relevant for P_0. This is clearly seen when varying the

fraction of satellites (Fig. A1) or the velocity profiles (Fig. A4). The above findings can be summarized by saying that power spectrum multipoles, within the explored scales, are affected only by the Finger-of-God effect from the 1-halo term.

In Fig. A4, for completeness, we also show the hexadecapole $P_4(k)$. We find the the dependence on the satellite velocity assignment scheme is relatively low. We note that de Mattia et al. (2020) finds that, for the eBOSS uncertainty, the hexadecapole is compatible with zero.

A remarkable result is that, whereas for the analysis in configuration space we could clearly split the 1-halo contributions from the large-scale signal used for cosmology, in $P_\ell(k)$ the 1-halo does affect modes with $k \sim 0.1h\,Mpc^{-1}$, which are scales that are also used BAO and RSD analysis. Despite this, we show in Appendix B that this has a negligible effect on the derived cosmological constraints, as the effects are absorbed by the nuisance parameters.

APPENDIX B: COSMOLOGICAL CONSTRAINS

In this Appendix, we present the results of fitting a RSD and anisotropic Alcock–Paczynski model (with $f\sigma_8$, $\alpha_\parallel$, $\alpha_\perp$ as free parameters) to the mock catalogues presented in Section 4 based on their power spectrum multipoles [$P_l(k)$, Appendix A]. These mocks have been analysed following the methodology used for the eBOSS data in de Mattia et al. (2020). The model considered here combines regularized perturbation theory with the Taruya, Nishimichi, and Saito RSD model (Taruya, Nishimichi & Saito 2010; Taruya et al. 2012).

A representative subset of the mocks presented in this work is also shown in Alam et al. (2020), together with other complementary N-Body mocks. In that paper, we not only show the results from fitting those mocks in Fourier space, but also in configuration space ξ_ℓ, following the methodology used for the data in Tamone et al. (2020). We refer to Alam et al. (2020) for details on the way the fits presented here were performed, with the only difference being that here we only report results using the Z-axis as the line of sight.

Alam et al. (2020) also reports a systematic error budget due to possible theory uncertainties of ({ 3.3 per cent, 1.8 per cent, 1.5 per cent} for {$f\sigma_8$, $\alpha_\parallel$, $\alpha_\perp$}). These are conservative systematic error budgets, corresponding to $2\times \sigma_{\text{stat}}$, as no significant deviation was found at the $2\sigma_{\text{stat}}$ level, with σ_{stat} being the statistical uncertainty on the OUTER RIM simulation.

For completeness, here we present the fits in Fourier space of a wider range of mocks sweeping the parameter space of the HOD considered as done in the main body of this paper. We consider this analysis more necessary for the Fourier space case, as the cosmological and HOD scales mix more than in configuration space.

Fig. B1 shows the results all the fits, following the same notation and line-styles as in Figs A1, A2, A3, and A4. We find that the fits are consistent with the OUTER RIM cosmology (the dashed horizontal line) within the systematic error budget, marked by the grey bands in Fig. B1.

From Fig. B1, it is clear that the cosmological constraints are robust against the different details of the HOD modelling, and the same result is found in Alam et al. (2020). We also note that the theoretical systematic error reported is an order of magnitude lower than the statistical error of eBOSS.
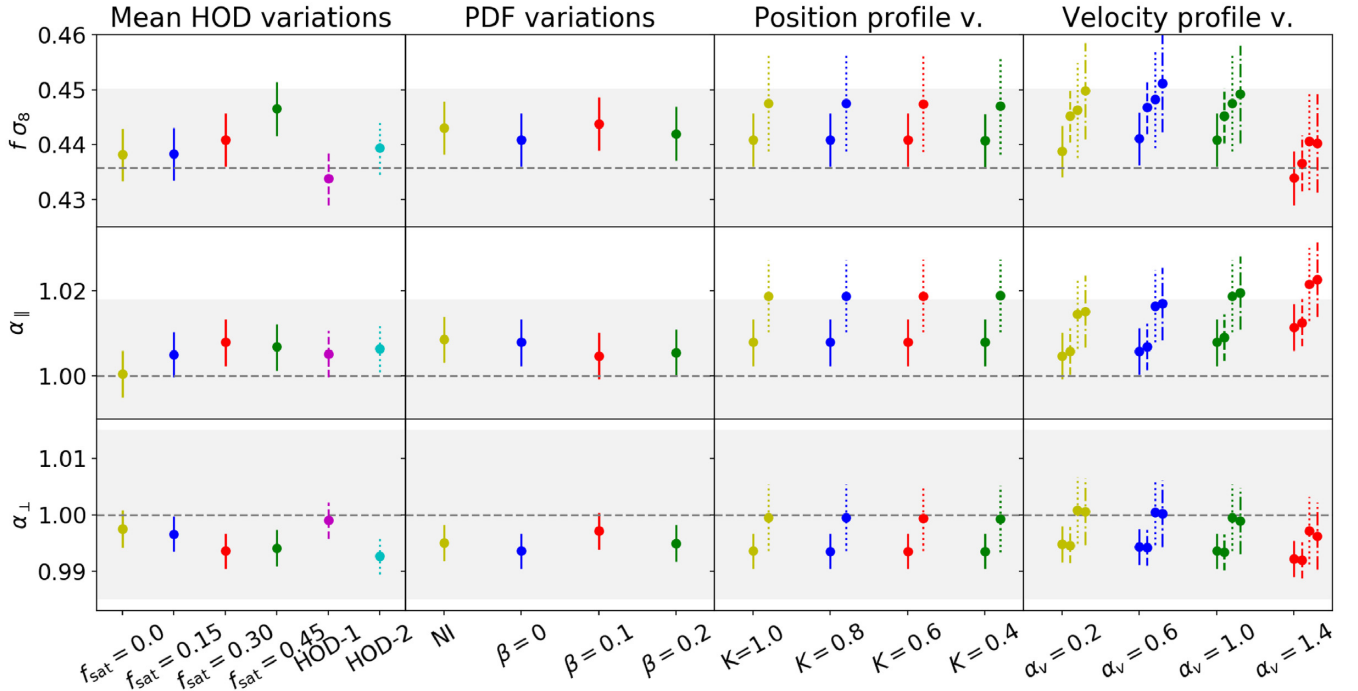
Figure B1. Cosmological parameters fits for different mocks presented in this paper. On the top row, we show the product of the growth rate and the normalization of the power spectrum $f\sigma_8$, on the middle row of panels we show the Alcock–Paczynski distortion along the line of sight $\alpha_{\parallel}$ and on the bottom row we show the Alcock–Paczynski parameter on the angular direction $\alpha_{\perp}$. From left to right, we show the same effects as in Figs A1, A2, A3, and A4, respectively, following the same line styles. The error bars show the statistical uncertainty of the best fit. The horizontal grey-dashed line represents the true value for OUTER RIM simulation, and the grey band the systematic error reported in Alam et al. (2020).

[1]*Departamento de Física Teórica, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Cantoblanco, Madrid, Spain*
[2]*Instituto de Física Teorica UAM-CSIC, Universidad Autónoma de Madrid, E-28049 Cantoblanco, Madrid, Spain*
[3]*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*
[4]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[5]*Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[6]*Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[7]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[8]*Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Federale de Lausanne (EPFL), CH-1290 Versoix, Switzerland*
[9]*Institute for Astronomy, University of Edinburgh, Royal Observatory, EH9 3HJ Edinburgh, UK*
[10]*ICC, University of Barcelona, IEEC-UB, Martí i Franquès, 1, E-08028 Barcelona, Spain*
[11]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*

[12]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[13]*Max-Planck Institut für extraterrestrische Physik, Postfach 1312, D-85741 Garching bei München, Germany*
[14]*Department of Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA*
[15]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK*
[16]*HEP Division, Argonne National Laboratory, Lemont, IL 60439, USA*
[17]*Instituto de Astrofísica, Pontificia Universidad Católica de Chiless, Vicuña Mackenna 4860, Santiago, Chile*
[18]*Perimeter Institute for Theoretical Physics, 31 Caroline St North, Waterloo, ON N2L 2Y5, Canada*
[19]*Center for Cosmology and AstroParticle Physics, The Ohio State University, Columbus, OH 43212, USA*
[20]*The Pennsylvania State University, University Park, PA 16802, USA*
[21]*National Astronomy Observatories, Chinese Academy of Science, Beijing 100012, P. R. China*

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