The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: *N*-body mock challenge for the quasar sample

Alex Smith,[1]★ Etienne Burtin,[1] Jiamin Hou,[2] Richard Neveux,[1] Ashley J. Ross ⬡,[3] Shadab Alam,[4] Jonathan Brinkmann,[5] Kyle S. Dawson,[6] Salman Habib,[7,8] Katrin Heitmann,[7] Jean-Paul Kneib,[9] Brad W. Lyke,[10] Hélion du Mas des Bourboux,[6] Eva-Maria Mueller,[11] Adam D. Myers,[10] Will J. Percival,[12,13,14] Graziano Rossi,[15] Donald P. Schneider,[16,17] Pauline Zarrouk[1,18] and Gong-Bo Zhao[19,20]

[1]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[2]*Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstr., D-85748 Garching bei München, Germany*
[3]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[4]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[5]*Apache Point Observatory and New Mexico State University, P.O. Box 59, Sunspot, NM 88349, USA*
[6]*Department Physics and Astronomy, University of Utah, 115 S 1400 E, Salt Lake City, UT 84112, USA*
[7]*High Energy Physics Division, Argonne National Laboratory, Lemont, IL 60439, USA*
[8]*Computational Science Division, Argonne National Laboratory, Lemont, IL 60439, USA*
[9]*Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[10]*Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA*
[11]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[12]*Waterloo Centre for Astrophysics, Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[13]*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[14]*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*
[15]*Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea*
[16]*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*
[17]*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*
[18]*Institute for Computational Cosmology, Dept. of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[19]*National Astronomy Observatories, Chinese Academy of Science, Beijing, 100101, P.R. China*
[20]*Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK*

ABSTRACT

The growth rate and expansion history of the Universe can be measured from large galaxy redshift surveys using the Alcock–Paczynski effect. We validate the Redshift Space Distortion models used in the final analysis of the Sloan Digital Sky Survey (SDSS) extended Baryon Oscillation Spectroscopic Survey (eBOSS) Data Release 16 quasar clustering sample, in configuration and Fourier space, using a series of halo occupation distribution mock catalogues generated using the OuterRim *N*-body simulation. We test three models on a series of non-blind mocks, in the OuterRim cosmology, and blind mocks, which have been rescaled to new cosmologies, and investigate the effects of redshift smearing and catastrophic redshifts. We find that for the non-blind mocks, the models are able to recover $f\sigma_8$ to within 3 per cent and $\alpha_\parallel$ and $\alpha_\perp$ to within 1 per cent. The scatter in the measurements is larger for the blind mocks, due to the assumption of an incorrect fiducial cosmology. From this mock challenge, we find that all three models perform well, with similar systematic errors on $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ at the level of $\sigma_{f\sigma_8} = 0.013$, $\sigma_{\alpha_\parallel} = 0.012$, and $\sigma_{\alpha_\perp} = 0.008$. The systematic error on the combined consensus is $\sigma_{f\sigma_8} = 0.011$, $\sigma_{\alpha_\parallel} = 0.008$, and $\sigma_{\alpha_\perp} = 0.005$, which is used in the final DR16 analysis. For baryon acoustic oscillation fits in configuration and Fourier space, we take conservative systematic errors of $\sigma_{\alpha_\parallel} = 0.010$ and $\sigma_{\alpha_\perp} = 0.007$.

Key words: methods: data analysis – catalogues – large-scale structure of Universe.

1 INTRODUCTION

The ΛCDM model of cosmology has been extremely successful in describing the expansion history and formation of structure in our Universe. Measurements covering a wide range of redshifts, from the temperature fluctuations in the cosmic microwave background at $z \sim 1100$ (Planck Collaboration VI 2020) to galaxy clustering measurements at late times (e.g. Eisenstein et al. 2005; Alam et al. 2017; Abbott et al. 2018), are all remarkably consistent with a flat, ΛCDM Universe. In the standard ΛCDM model, gravity follows general relativity, dark matter is composed of a collisionless, cold

★ E-mail: alexander.smith@cea.fr

dark matter (CDM), and dark energy is described by the cosmological constant, Λ. However, this dark energy, which makes up the largest component of the energy density (≈ 70 per cent) of the Universe (Planck Collaboration VI 2020) and is responsible for driving the present-day accelerated expansion, is poorly understood. Constraints can be placed on theories of dark energy and modifications to general relativity by measuring the expansion history and growth rate of structures from surveys of galaxies in large cosmological volumes. Reaching the high-precision measurements that are needed to tighten current constraints requires surveys covering even larger volumes.

One of the primary measurements that is taken to constrain the nature of dark energy is of the baryon acoustic oscillation (BAO; e.g. Cole et al. 2005; Eisenstein et al. 2005). In the early Universe, acoustic waves in the hot plasma propagate outwards from overdense regions until they are frozen out at recombination, when the baryons and photons decouple. These small enhancements in the density field can be seen in the distribution of galaxies at later times. A BAO feature can be seen in measurements of the two-point statistics of galaxies, at a characteristic comoving length scale of $\sim 100\,h^{-1}\mathrm{Mpc}$, which can be used as a standard ruler for measuring distances throughout cosmic time.

Redshift space distortions (RSD) can be used to measure the growth of structure. Since we measure the redshift for each galaxy, the apparent position along the line of sight depends on its peculiar velocity. The infall of galaxies towards overdense regions, and the motion of galaxies within virialized haloes, leads to anisotropic distortions in the spatial distribution of galaxies in redshift space (Kaiser 1987). RSD measurements can be used to measure the linear growth rate, f, which in the ΛCDM model is related to the matter content of the Universe through $f \simeq \Omega_m^\gamma$. In general relativity, $\gamma \simeq 0.55$, and, therefore, measurements of the growth rate allow constraints to be placed on theories of modified gravity (e.g. Guzzo et al. 2008).

The extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) is a survey of luminous red galaxies (LRGs), emission line galaxies (ELGs), and quasars, which is an extension of the previous BOSS survey (Dawson et al. 2013). Between $0.8 < z < 2.2$, quasars are ideal tracers of the matter density field, since they are the brightest objects, and can be observed more easily than other types of galaxies. This was first done in Ata et al. (2018), filling in this redshift range.

Two-point statistics, which are commonly used in cosmology, provide a way to quantify the distribution of galaxies in a given volume. They measure the excess probability of finding pairs of galaxies, as a function of spatial separation, compared to a random distribution (e.g. Landy & Szalay 1993). Measurements of the two-point statistics of galaxies in the survey can be used to probe the growth rate and expansion history (e.g. Blake et al. 2011b; Beutler et al. 2012; Alam et al. 2017). However, calculating cosmological distances from measurements of redshift requires a fiducial cosmology to be chosen. If this choice is incorrect, it will produce incorrect distance measurements, resulting in a distortion of the BAO peak, which is known as the Alcock–Paczynski effect (Alcock & Paczynski 1979). By measuring the distortion of the BAO feature, which is quantified by the parameters $\alpha_\parallel$ and $\alpha_\perp$ (parallel and perpendicular to the line of sight, respectively), constraints can be placed on the transverse comoving distance, $D_M(z)/r_{\mathrm{drag}}$, and the Hubble distance $D_H(z)/r_{\mathrm{drag}}$, where r_{drag} is the sound horizon at the drag epoch.

In order to extract cosmological information from the survey, we need to be able to model the two-point statistics of quasars, also known as Quasi-Stellar Objects (QSOs). Our models, which are based on using perturbation theory to calculate cosmological power

spectra or correlation functions (e.g. Crocce & Scoccimarro 2006; Taruya et al. 2012; Carlson, Reid & White 2013), are for dark matter and neglect the formation and evolution mechanisms of quasars. On large scales, this complicated non-linear, baryonic and strong gravity physics is expected to decouple and be modelled using a few nuisance parameters (Hou et al. 2018; Zarrouk et al. 2018). It is therefore crucial to test the limits of such models, with realistic simulated quasar catalogues.

To validate the models used, we run our analysis on a set of *N*-body mock catalogues in which the true cosmology is known. Measuring the scatter in the best-fitting parameters obtained with mocks produced using different halo occupation distributions (HODs), cosmologies, and including the effects of redshift smearing and catastrophic redshifts, allows us to estimate the systematic uncertainties in these measurements. Previously, quasar mock catalogues for the first year of eBOSS analysis were created in Rodríguez-Torres et al. (2017), and multitracer mocks were created in Alam et al. (2020a). The aim of this mock challenge is to firstly validate the RSD models on a set of non-blind *N*-body mocks, with a wide variety of HOD models, where the true cosmology is known, and also on a set of mocks where the cosmology is blinded. Secondly, we use these mocks to measure the systematic uncertainties on $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$.

The mock challenge presented in this paper is similar to what was previously done for the BOSS survey, where *N*-body mocks, generated using a range of HOD models and also in different cosmologies, were analysed blindly (Beutler et al. 2017). For the WiggleZ survey (Drinkwater et al. 2010), the RSD analysis was performed using a range of existing models without explicitly testing them on mocks (Blake et al. 2011a). In the cosmological analysis of the Dark Energy Survey Year 1 data (Abbott et al. 2018), the parameter measurements were validated using *N*-body mocks that simulated galaxy clustering and lensing observables (MacCrann et al. 2018). For future surveys, such as DESI (DESI Collaboration 2016a, b), LSST (Ivezić et al. 2019), and Euclid (Laureijs et al. 2011), much effort will be required in producing many accurate mock catalogues to ensure that the model systematics can be reduced to the required levels.

The quasar mock challenge is part of the final release of BAO and RSD measurements from eBOSS. The construction of the data catalogues is described in Ross et al. (2020) and Lyke et al. (2020), and the configuration-space and Fourier-space analysis of the quasar sample is presented in Hou et al. (2020) and Neveux et al. (2020), respectively. In addition, cosmological measurements are made using the sample of luminous red galaxies (LRGs; Bautista et al. 2020; Gil-Marín et al. 2020) and emission line galaxies (ELGs; Raichoor et al. 2017; de Mattia et al. 2020; Tamone et al. 2020), with mock catalogues and mock challenges described in Alam et al. (2020b), Avila et al. (2020), Lin et al. (2020), Rossi et al. (2020), and Zhao et al. (2020). At high redshifts, the BAO analysis of Ly α forest measurements is found in des Mas du Bourboux et al. (2020). The final cosmological implications of these results are presented in eBOSS Collaboration (2020).[1]

This paper is outlined as follows. In Section 2, we give an overview of the eBOSS QSO clustering sample and observational effects. In Section 3, we describe the RSD models used in the clustering

---

[1] A summary of all SDSS, BAO, and RSD measurements with accompanying legacy figures can be found at https://sdss.org/science/final-bao-and-rsd-measurements/. The full cosmological interpretation of these measurements can be found at: https://sdss.org/science/cosmology-results-from-eboss/.

analysis. Section 4 describes the HOD models and methodology for creating the mocks. Results are discussed in Section 5 for the non-blind mocks and in Section 6 for the blinded mocks. Our conclusions are summarized in Section 7.

2 QUASAR SAMPLE

The eBOSS survey was performed as part of SDSS-IV (Blanton et al. 2017), the fourth phase of the Sloan Digital Sky Survey (SDSS; York et al. 2000). The survey, which began in 2014 July and completed observations in 2019 March, was carried out using the 2.5-m Sloan Foundation Telescope at Apache Point Observatory in New Mexico (Gunn et al. 2006). The survey is the successor of the previous BOSS survey (Dawson et al. 2013, which was part of SDSS-III) and utilizes the two BOSS spectrographs (Smee et al. 2013), which are able to measure a total of 1000 spectra per observation.

In order to make cosmological measurements covering a wide range of redshifts, the eBOSS survey targeted several different types of tracers. These tracers are luminous red galaxies (LRGs; Prakash et al. 2016), covering the redshift range of $0.6 < z < 1.0$, emission line galaxies (ELGs; Raichoor et al. 2017), over the redshift range of $0.6 < z < 1.1$ and quasars (QSOs; Myers et al. 2015). The quasars can be split into a lower redshift sample ($0.8 < z < 2.2$), which can be used as direct tracers of the matter field, and a higher redshift sample, which can be utilized for measurements of the Ly α forest ($z > 2$). In this work, we focus on the sample of quasars in the redshift range of $0.8 < z < 2.2$, which can be used as direct tracers.

In the total sample, there are ∼330 000 quasars, with ∼60 per cent in the northern galactic cap, and ∼40 per cent in the south. The effective redshift of the quasar sample is $z_{eff} = 1.48$, covering a total area of 4700 deg^2. See Ross et al. (2020) for details of the large-scale structure catalogue, and Lyke et al. (2020) for details of the quasar catalogue, including the procedure for determining redshifts. The number of quasars in the DR16 sample is approximately doubled from the earlier DR14 sample. For the previous DR14 QSO analyses, see Gil-Marín et al. (2018), Hou et al. (2018), and Zarrouk et al. (2018).

Cosmological information can be extracted from the two-point clustering of the quasar redshift catalogue. However, it is difficult to measure a precise redshift for quasars. There is some uncertainty in the redshift measurements, due to astrophysical outflows. In addition, a small fraction of quasars have 'catastrophic' redshift measurements, where the redshift estimate is very far from the true redshift. It is essential that our models can deal with the impact this has on the clustering measurements in order to obtain unbiased cosmological measurements. We outline these effects below.

2.1 Redshift smearing

There is a wide range of behaviour in the width of the emission lines in the spectra of quasars. Often, the optically selected quasars show broad emission lines, which is due to the fast rotation of hot gas close to the central black hole. The gas is also affected by radiation-driven winds, which leads to offsets in the position of the emission lines in the spectra. This results in systematic uncertainties in the measured quasar redshifts.

Several methods are used to estimate the redshift of quasars in the eBOSS survey. The pipeline redshifts are the result of fitting four eigenspectra to each spectrum. The redshift can also be estimated from the MgII line, from a principle component analysis of the full quasar spectrum, or from visual inspection. The uncertainties in the measured redshifts, i.e. redshift smearing, can be measured using the

distribution of Δz, where Δz is the difference between two of these different redshift estimates (for details of the redshift estimates, see Lyke et al. 2020; Ross et al. 2020).

The quantity Δz can equivalently be thought of as a velocity difference, Δv, where $\Delta v = c\Delta z/(1 + z)$. The distribution of Δv is approximately a Gaussian but with wide tails extending to large velocities (see e.g. fig. 4 of Zarrouk et al. 2018). In the survey requirements document, the distribution of Δv is a Gaussian with mean $\langle \Delta v \rangle = 0$, and rms given by

$$\begin{aligned} \sigma_v(z) &= 300 \, \text{km s}^{-1} & z < 1.5 \\ \sigma_v(z) &= 450(z - 1.5) + 300 \, \text{km s}^{-1} & z > 1.5. \end{aligned} \quad (1)$$

The requirement of 300 km s^{-1} is relaxed for $z > 1.5$, since it is more difficult to obtain accurate redshifts. We use this distribution as an initial model for the redshift uncertainties that are applied in the mocks (see Section 4.2) and then improve the modelling of redshift smearing by looking at the distributions seen in the data.

While the distribution of Δv is approximately Gaussian, a Gaussian distribution is unable to model the wide tails that extend to high velocities. We investigated the distribution of Δv by looking at the set of quasars that have duplicate observations and found that the distribution can be better modelled by the sum of two Gaussians (see fig. 4 of Lyke et al. 2020). This double-Gaussian distribution can be written as

$$\frac{dN}{d(\Delta v)} = \frac{1}{\sqrt{2\pi}(1 + F)} \left[\frac{F}{\sigma_1} \exp\left(\frac{-\Delta v^2}{2\sigma_1^2} \right) + \frac{1}{\sigma_2} \exp\left(\frac{-\Delta v^2}{2\sigma_2^2} \right) \right], \quad (2)$$

where both Gaussians are centred on zero, with rms σ_1 and σ_2, and F sets the fraction of the two Gaussians. For objects that have been re-observed, the distribution of Δv is fit well by a double Gaussian probability distribution with $\sigma_1 = 150$ km s^{-1}, $\sigma_2 = 1000$ km s^{-1}, and $F = 4.478$ (Lyke et al. 2020).

2.2 Catastrophic redshifts

Measurements of quasar clustering are also affected by catastrophic redshifts, where an incorrect redshift is measured for a small fraction of objects, e.g. due to line confusion or contamination from sky lines. The catastrophic redshift failure rate is low; for DR16, this is about 1.5 per cent. This is estimated from a set of 10 000 spectra that are randomly chosen for visual inspection. A catastrophic redshift is defined as having a pipeline redshift that differs from the visual inspection redshift by $\Delta v > 3000$ km s^{-1}. For more details, see Lyke et al. (2020). The inclusion of catastrophic redshifts will impact the measurements of $f\sigma_8$, since the clustering signal is being diluted.

2.3 Effective redshift

There is also some uncertainty on the effective redshift of the QSO sample, as there is some ambiguity in the definition of z_{eff} (see appendix A of Hou et al. 2020). Changing the definition shifts the effective redshift of the QSOs from $z_{eff} = 1.48$ to $z_{eff} = 1.52$. As shown in Hou et al. (2020), the impact this has on the cosmological measurements is small compared to the statistical error. The effect of redshift smearing and catastrophic redshifts of z_{eff} is negligible.

3 MODELS FOR TWO-POINT STATISTICS

In this section, we describe the RSD and BAO models that we test in this mock challenge. Section 3.1 gives an overview of two-point

statistics in redshift space, and Section 3.2 describes the Alcock–Paczynski effect. The full-shape RSD models in Fourier space that we consider are described in Section 3.3, with configuration-space models in Section 3.4. Finally, the BAO-only models, in both Fourier and configuration space, are described in Section 3.5.

3.1 Two-point statistics

The galaxies and quasars observed in large cosmological surveys are biased tracers of the underlying matter density field. The density contrast of tracers, $\delta_t(\boldsymbol{x})$, at position $\boldsymbol{x}$ is related to the matter overdensity, $\delta_m(\boldsymbol{x})$, by the linear bias, b_1,

$$\delta_t(\boldsymbol{x}) = b_1 \delta_m(\boldsymbol{x}). \tag{3}$$

The two-point correlation function, $\xi(\boldsymbol{r})$ is defined as

$$\xi(\boldsymbol{r}) = \langle \delta(\boldsymbol{x}) \delta(\boldsymbol{x} + \boldsymbol{r}) \rangle, \tag{4}$$

where the angled brackets indicate an ensemble average, and its Fourier transform is the power spectrum, $P(\boldsymbol{k})$.

Galaxies and quasars are observed in redshift space (Sargent & Turner 1977), and their apparent position along the line of sight is shifted by the peculiar velocity,

$$\boldsymbol{s} = \boldsymbol{x} + u_z \hat{\boldsymbol{z}}, \tag{5}$$

where u_z is the component of velocity along the line of sight of the observer, expressed in units of the Hubble velocity, $u_z = v_z/aH$. The effect of velocity is that it imparts anisotropies on the redshift-space two-point correlation function, due to two effects. The coherent infall of galaxies towards large overdensities leads to a flattening in the clustering measurements on large scales, which is the Kaiser effect (Kaiser 1987), while the motion of galaxies within virialized haloes leads to Finger-of-God distortions (Jackson 1972). In redshift space, the 2D correlation function, $\xi(s, \mu)$, can be decomposed into Legendre multipoles,

$$\xi(s, \mu) = \sum_{\ell} \xi_{\ell}(s) \mathcal{L}_{\ell}(\mu), \tag{6}$$

where $\mathcal{L}_{\ell}(\mu)$ is the ℓth-order Legendre polynomial, and μ is the cosine of the angle between the line of sight and the pair separation vector. The multipoles $\xi_{\ell}(s)$ are evaluated through

$$\xi_{\ell}(s) = \frac{2\ell + 1}{2} \int_{-1}^{1} \xi(s, \mu) \mathcal{L}_{\ell}(\mu) d\mu. \tag{7}$$

In linear theory, only the monopole, $\xi_0(s)$, quadrupole, $\xi_2(s)$, and hexadecapole, $\xi_4(s)$, are non-zero.

3.2 Alcock–Paczynski effect

On large scales, a BAO peak can be seen in the correlation function of galaxies (or alternatively oscillatory features in the power spectrum), which is shifted if an incorrect fiducial cosmology is assumed (Alcock & Paczynski 1979). The BAO peak position is scaled by $s'_{\parallel} = \alpha_{\parallel} s_{\parallel}$ and $s'_{\perp} = \alpha_{\perp} s_{\perp}$ parallel and perpendicular to the line of sight, respectively, where

$$\alpha_{\parallel} = \frac{H^{\text{fid}}(z) r_{\text{drag}}^{\text{fid}}}{H(z) r_{\text{drag}}} \text{ and } \alpha_{\perp} = \frac{D_{\text{M}}(z) r_{\text{drag}}^{\text{fid}}}{D_{\text{M}}^{\text{fid}}(z) r_{\text{drag}}}. \tag{8}$$

$H(z)$ is the Hubble parameter, $D_{\text{M}}(z)$ is the transverse comoving distance, and r_{drag} is the sound horizon at the drag epoch. Quantities labelled 'fid' are in the fiducial cosmology, while the true quantities that we aim to measure have no label. Therefore, constraining the $\alpha_{\parallel}$

and $\alpha_{\perp}$ parameters from the two-point clustering measurements can be used to place constraints on $D_{\text{M}}(z)/r_{\text{drag}}$ and $D_{\text{H}}(z)/r_{\text{drag}}$, where the Hubble distance $D_{\text{H}}(z) = c/H(z)$.

3.3 Fourier-space RSD models

In the linear approximation, the power spectrum in redshift space can be described using the well-known Kaiser formula,

$$P_t^{\text{lin}}(k, \mu) = \left(b_1^2 + 2f b_1 \mu^2 + f^2 \mu^4\right) P_m(k), \tag{9}$$

where f is the linear growth rate, which is defined as

$$f(a) = \frac{d \ln D(a)}{d \ln a}. \tag{10}$$

D is the linear growth function, and a is the expansion factor (Kaiser 1987). This is valid only in the linear regime, where the density perturbations are small ($\delta_m \ll 1$), and non-linear terms in the fluid equations describing the density and velocity of the matter field are negligible. This corresponds to large physical scales ($\gtrsim 50\ h^{-1}$Mpc), where the density of the galaxies is related to matter density by a constant linear bias (equation 3), and there is a linear coupling between the matter velocity and density fields ($\theta_m = f\delta_m$). Haloes and galaxies form in high-density regions on physical scales much smaller than $50\ h^{-1}$Mpc, where the non-linear terms in the fluid equations are important, and the assumptions of linear theory break down. To predict the two-point statistics on small scales, it is therefore important to model RSD in the non-linear regime.

The non-linear matter power spectrum for biased tracers, $P_t(k, \mu)$, is given by

$$\begin{aligned} P_t(k, \mu) = {} & P_{\delta\delta,t}(k) + 2f\mu^2 P_{\delta\theta,t}(k) + f^2 \mu^4 P_{\theta\theta}(k) \\ & + b_1^3 A(k, \mu) + b_1^4 B(k, \mu), \end{aligned} \tag{11}$$

where $P_{\delta\delta,t}(k)$, $P_{\delta\theta,t}(k)$, and $P_{\theta\theta}(k)$ are the density, density–velocity, and velocity–velocity power spectra, respectively, for tracers, and $\theta = \nabla \cdot \boldsymbol{u}$ is the divergence of the velocity field (Taruya, Nishimichi & Saito 2010). The first three terms in this expression are the non-linear Kaiser formula, b_1 is the linear bias, and $A(k, \mu)$ and $B(k, \mu)$ are correction terms. These correction terms arise from the non-linear coupling between the velocity and density fields and depend on the cross-bispectrum, $P_{\delta\theta,t}(k)$, and $P_{\theta\theta}(k)$. The power spectrum and correction terms in equation (11) can be calculated using perturbation theory models.

On even smaller scales ($\lesssim 10\ h^{-1}$Mpc), the motion of galaxies within haloes becomes important, giving rise to Finger-of-God distortions. This can be modelled as

$$P_t^s(k, \mu) = F_{\text{FoG}}(k, \mu) P_t(k, \mu) \exp\left[-(k\mu\sigma_{\text{zerr}})^2\right]. \tag{12}$$

The first term, $F_{\text{FoG}}(k, \mu)$, is a damping function that models the Finger-of-God effect and usually takes the form of a Gaussian or Lorentzian function. We use a function of the form

$$F_{\text{FoG}}(k, \mu) = \frac{1}{\sqrt{1 + \mu^2 k^2 a_{\text{vir}}^2}} \exp\left(\frac{-\mu^2 k^2 \sigma_v^2}{1 + \mu^2 k^2 a_{\text{vir}}^2}\right), \tag{13}$$

where a_{vir} is the kurtosis of the velocity distribution, and σ_v is the velocity dispersion (Grieb et al. 2017; Sánchez et al. 2017b; Hou et al. 2018).

The final term in equation (12) is an exponential function that models the effect of redshift uncertainty, where σ_{zerr} is the redshift error (Hou et al. 2018). In the case of the double-Gaussian redshift smearing of equation (2), the parameters σ_v, a_{vir} and σ_{zerr} are able to pick out the different Gaussian components (see Section 5.3.2).

Table 1. Parameters of the RegPT (Section 3.3.1), RESPRESSO (Section 3.3.2), and CLPT (Section 3.4.1) models. Prior ranges are indicated for the free parameters.

RegPT		RESPRESSO		CLPT	
b_1	$(0,5)$	b_1	$(0.25,6)$	b	$(0,3)$
b_2	$(-8,8)$	b_2	$(-2,3)$	F''	$(0,10)$
b_{s2}	Fixed	γ_2	Fixed		
b_{3nl}	Fixed	γ_3^-	$(-2,2)$		
A_g	$(-1,5)$				
σ_v	$(0,15)$	σ_v	Fixed	σ	$(0,20)$
a_{vir}	$(0,15)$	a_{vir}	$(0.2,10)$		
		σ_{zerr}	$(0,6)$		

In the next subsections, we briefly outline the specific Fourier-space models we use.

3.3.1 RegPT

In the first model that we consider, the power spectra $P_{\delta\delta}(k)$, $P_{\delta\theta}(k)$, and $P_{\theta\theta}(k)$, and the correction terms $A(k, \mu)$ and $B(k, \mu)$ of equation (11), are calculated using Regularized Perturbation Theory (RegPT) at two-loop order (Taruya et al. 2012).

The effect of redshift errors is treated as a velocity dispersion. The parameter σ_v in the Finger-of-God term is a free parameter to take into account the effect of redshift errors, while the additional parameter σ_{zerr} is set to zero. This choice implies that there is no additional damping due to redshift errors.

In the bias expansion, b_1 and b_2 are the linear and second-order bias, and the local biases b_{s2} and b_{3nl} are kept fixed assuming local Lagrangian bias. There is also an additional shot noise parameter, A_g (Neveux et al. 2020). In addition to $\alpha_\parallel$, $\alpha_\perp$, and $f\sigma_8$, there are five free parameters (b_1, b_2, A_g, a_{vir}, and σ_v), and their priors are given in Table 1.

The model we have described here has previously been used in the analysis of BOSS LRGs (Beutler et al. 2014, 2017). At $z = 1$, the power spectrum prediction using RegPT is within per cent-level agreement of N-body simulations up to $k \sim 0.23\ h\mathrm{Mpc}^{-1}$ (Taruya et al. 2012).

In the final eBOSS DR16 quasar analysis, the same model is fit to measurements of the power spectrum multipoles in Fourier space (Neveux et al. 2020), and they are therefore able to use the systematic uncertainties quoted in this study. This model is additionally used in the Fourier-space analysis of the eBOSS ELGs (de Mattia et al. 2020). Throughout this paper, we will refer to this model, which combines RegPT with the specific FoG prescription described above, as simply 'RegPT'.

3.3.2 RESPRESSO + Fitting formula

In the second model we consider, the power spectrum $P_{\delta\delta,t}(k)$ is computed using the code RESPRESSO (Nishimichi, Bernardeau & Taruya 2017). RESPRESSO is based on the response function, which characterizes how the non-linear power spectrum varies in response to small perturbations of the initial power spectrum. $P_{\delta\theta,t}(k)$ and $P_{\theta\theta,t}(k)$ are calculated from the fitting functions of Bel et al. (2019), which are based on measurements from N-body simulations.

In the Finger-of-God term of equation (13), σ_v is kept fixed to the linear theory prediction. However, σ_v and a_{vir} are defined differently than in the RegPT model described in Section 3.3.1, differing by

a factor of f in their normalization. The effect of redshift errors is modelled by the parameter σ_{zerr}, which is kept free.

The bias expansion of Chan, Scoccimarro & Sheth (2012) is used (see equation 22 of Hou et al. 2020), with first- and second-order bias parameters b_1 and b_2, and non-local bias parameters γ_2 and γ_3^-. γ_2 is fixed assuming local Lagrangian bias. In total, there are five free parameters (b_1, b_2, γ_3^-, a_{vir}, and σ_{zerr}), with priors in Table 1.

While this model gives a prediction of the redshift-space power spectrum, the two-point correlation function can be calculated from its Fourier transform. A similar model to this has previously been used in the DR14 QSO analysis in configuration space of Hou et al. (2018) (and also in the analysis of the BOSS wedges Grieb et al. 2017; Sánchez et al. 2017b), but this differs in that the power spectra were calculated using Galilean-invariant Renormalized Perturbation Theory (gRPT). The RESPRESSO prediction of the matter power spectrum is in good agreement with N-body simulations, within 2 per cent up to $k = 0.3\ h\mathrm{Mpc}^{-1}$, and shows a slight improvement compared to gRPT (see fig. 4 of Hou et al. 2020). In addition, Hou et al. (2018) showed that cosmological parameter estimates are less biased with the parameter z_{err} included.

The full model as presented here, which we measure the systematic errors of, is used in the configuration space DR16 quasar analysis of Hou et al. (2020). We will refer to this model as 'RESPRESSO'.

3.4 Configuration-space RSD models

An alternative approach is to model the two-point correlation function in configuration space.

3.4.1 CLPT

In the Lagrangian approach, particles (or tracers) are moved from their initial Lagrangian coordinates, $\boldsymbol{q}$, to their final coordinates $\boldsymbol{x}$, by a displacement field, $\boldsymbol{\psi}$,

$$\boldsymbol{x}(\boldsymbol{q}, t) = \boldsymbol{q} + \boldsymbol{\psi}(\boldsymbol{q}, t). \tag{14}$$

The displacement field can be written as a perturbative expansion, $\boldsymbol{\psi} = \boldsymbol{\psi}^{(1)} + \boldsymbol{\psi}^{(2)} + \boldsymbol{\psi}^{(3)} + \cdots$, where the first-order term is the Zel'dovich approximation. In the Convolution Lagrangian Perturbation Theory (CLPT) model, more terms are summed together in this expansion, which provides a good description for the real space correlation function (Matsubara 2008; Carlson et al. 2013).

The real space correlation function can then be transformed into redshift space using the Gaussian streaming model,

$$1 + \xi(s_\perp, s_\parallel) = \int dr_\parallel (1 + \xi(r)) \mathcal{G}(s_\parallel - r_\parallel, v_{12}, \sigma_{12}), \tag{15}$$

where $\mathcal{G}$ is a Gaussian function centred on μv_{12}, which describes the probability that a pair of galaxies with separation $r_\parallel$ in real space have a separation $s_\parallel$ in redshift space,

$$\mathcal{G}(s_\parallel - r_\parallel, v_{12}, \sigma_{12}) = \frac{1}{\sqrt{2\pi\sigma_{12}^2(r, \mu)}} \exp\left(\frac{(s_\parallel - r_\parallel - \mu v_{12})^2}{2\sigma_{12}^2(r, \mu)}\right). \tag{16}$$

This probability depends on the velocities of galaxies; v_{12} is the pairwise infall velocity, and σ_{12} is the pairwise velocity dispersion.

In total, there are three nuisance parameters, which are the linear and second-order bias, b and F'', respectively, and σ, which takes into account the Finger-of-God effect and redshift smearing. Their priors are given in Table 1.

The CLPT prediction for the redshift-space correlation function is fit to measurements of the correlation function in real space. Previously, this model was used in the eBOSS DR14 quasar analysis of Zarrouk et al. (2018) in configuration space and also in Icaza-Lizaola et al. (2019) for the LRGs. CLPT was also used in the BOSS LRG analysis of Satpathy et al. (2017). At the redshift of the QSOs, the model provides accurate correlation function predictions, which agree with *N*-body simulations down to small scales of $\sim 20\,h^{-1}$Mpc (Zarrouk et al. 2018).

The CLPT model is used in the eBOSS DR16 LRG analysis of Bautista et al. (2020) but is not used in the final analysis of the QSOs. For the analysis of the DR16 quasar sample, the RegPT model (Section 3.3.1) and the RESPRESSO model (Section 3.3.2) are used in Fourier and configuration space, respectively. As we show in this paper, the three models work equally well, and there is no reason why the CLPT model should be chosen to be used in the final analysis over any of the other models that we consider. In this work, we use the CLPT model as an additional check to verify the rescaling of the blinded mocks (Section 4.3) independently of the other models.

3.5 BAO models

In addition to the RSD models described in the sections above, as part of the mock challenge, we also verify the BAO fitting procedure, which provides another way to measure $\alpha_{\scriptscriptstyle\parallel}$ and $\alpha_{\perp}$ from the two-point statistics of the data. BAO modelling aims to isolate cosmological information from the BAO only, which differs from RSD modelling, which predicts the two-point statistics over a wide range of scales. In Neveux et al. (2020), BAO fits are performed in Fourier space on measurements of the power spectrum, while in Hou et al. (2020), the fits are done to the correlation function measurements in configuration space. We describe the BAO models below.

3.5.1 Fourier-space analysis

The linear power spectrum can be decomposed into two parts (e.g. Kirkby et al. 2013): a 'smooth' (or 'no-wiggles') broad-band term, $P_{\rm sm}(k)$, and 'peak' term, $P_{\rm peak}(k)$, which isolates the oscillations of the BAO feature,

$$P_{\rm lin}(k) = P_{\rm sm}(k) + P_{\rm peak}(k). \quad (17)$$

In the first BAO model we consider, the redshift-space power spectrum, which models non-linear effects, is given by

$$P(k, \mu) = \frac{b^2(1 + \beta\mu^2)^2}{1 + (k\mu\Sigma_{\rm s})^2/2} \left[P_{\rm sm}(k, \mu) + P_{\rm peak}(k, \mu)e^{-k^2\Sigma_{\rm nl}^2} \right], \quad (18)$$

where *b* is the bias, and $\beta = f/b$ is the RSD parameter (Bautista et al. 2018). The factor of $(b^2(1 + \beta\mu^2)^2)/(1 + (k\mu\Sigma_{\rm s})^2/2)$ is a Lorentzian, which models the Finger-of-God effect and redshift smearing on small scales, with the damping parameter $\Sigma_{\rm s}$. The factor of $b^2(1 + \beta\mu^2)^2$ comes from the Kaiser formula (equation 9). In equation (18), $P_{\rm peak}(k, \mu)$ is multiplied by an exponential term, which models the anisotropic, non-linear damping of the BAO feature (Eisenstein, Seo & White 2007). This anisotropic damping parameter, $\Sigma_{\rm nl}$, is defined as

$$\Sigma_{\rm nl}^2 = (1 - \mu^2)\Sigma_{\perp}^2/2 + \mu^2\Sigma_{\parallel}^2/2. \quad (19)$$

The mock catalogues are used to obtain values of the damping parameters. For mocks that do not contain redshift smearing, the damping parameters are kept fixed in the fitting procedure to the

values $\Sigma_{\rm s} = 1\,h^{-1}$Mpc, $\Sigma_{\perp} = 3\,h^{-1}$Mpc, and $\Sigma_{\scriptscriptstyle\parallel} = 8\,h^{-1}$Mpc. These values are obtained from fits to the mocks with no redshift smearing (see Section 4.2). For mocks that do contain redshift smearing, the value of $\Sigma_{\rm s}$ is increased to $\Sigma_{\rm s} = 4\,h^{-1}$Mpc. This value of $\Sigma_{\rm s}$ is obtained from mocks with realistic smearing, with the values of $\Sigma_{\scriptscriptstyle\parallel}$ and $\Sigma_{\perp}$ fixed.

The broad-band component of the power spectrum multipoles, which does not contain BAO information, is fit by a polynomial, $P_{\ell}^f(k) = P_{\ell}(k) + a_{0,\ell}/k + a_{1,\ell} + a_{2,\ell}k$, where $a_{i,\ell}$ are the polynomial coefficients (Bautista et al. 2018). In total, there are 10 nuisance parameters in this BAO model (the bias, and three broad-band terms for each multipole).

The BAO model described in this section has previously been used in the analysis of the eBOSS DR14 LRGs (Bautista et al. 2018). In Neveux et al. (2020), this model is fit in Fourier space to measurements of the power spectrum of the DR16 QSOs.

3.5.2 Configuration-space analysis

In the second BAO model we use, the power spectrum in redshift space is given by

$$P(k, \mu) = \left(\frac{1 + \beta\mu^2}{1 + (k\mu\Sigma_{\rm s})^2/2} \right)^2 \left[P_{\rm sm}(k, \mu) + P_{\rm peak}(k, \mu)e^{-k^2\Sigma_{\rm nl}^2} \right], \quad (20)$$

which differs from equation (18) by the Lorentzian function, which models the Fingers-of-God distortions, being squared (Ross et al. 2017). The anisotropic damping parameter, $\Sigma_{\rm nl}$, is defined the same way (equation 19).

$P_{\rm lin}$ is calculated using CAMB (Lewis, Challinor & Lasenby 2000), while the smooth 'no-wiggles' power spectrum is calculated from the fitting formulae of Eisenstein & Hu (1998). The damping parameters $\Sigma_{\rm s}$, $\Sigma_{\perp}$ and $\Sigma_{\scriptscriptstyle\parallel}$. are kept fixed to the same values as are used for the Fourier-space BAO model.

The broad-band parts of the correlation function monopole and quadrupole are each modelled as a cubic polynomial, with two additional parameters that adjust the BAO feature (for details, see Section 5 of Hou et al. 2020). In total, there are eight nuisance parameters.

The model $P(k, \mu)$ is Fourier transformed to obtain the correlation function. BAO fits using this model in configuration space were previously done in the analysis of the BOSS LRGs (Ross et al. 2017). The same model is used in the DR16 QSO correlation function BAO fits of Hou et al. (2020).

4 MOCK CATALOGUES

In this work, we construct the mock catalogues using the OuterRim *N*-body simulation (Habib et al. 2016; Heitmann et al. 2019a, b), which contains $10,240^3$ particles of mass $m_p = 1.85 \times 10^9\,h^{-1}M_\odot$ in a box of side length $3000\,h^{-1}$Mpc. Haloes are identified using a friends-of-friends (FOF) algorithm (Davis et al. 1985) with linking length $b = 0.168$. The OuterRim simulation uses a flat ΛCDM cosmology with $\Omega_{\rm cdm}h^2 = 0.1109$, $\Omega_{\rm b}h^2 = 0.02258$, $h = 0.71$, $\sigma_8 = 0.8$, and $n_{\rm s} = 0.963$, which is consistent with the WMAP7 measurements (Komatsu et al. 2011). Mock catalogues are constructed from the simulation snapshot at $z = 1.433$. This snapshot is chosen as it is closest in redshift to the effective redshift of the eBOSS quasar sample ($z_{\rm eff} = 1.48$). Using a single snapshot allows the accuracy of the models to be tested at a single redshift, without evolution. However, in the

mocks we construct, we also add in evolution in the velocities, testing our models with f that evolves with redshift.

Quasar clustering measurements can be interpreted using the halo model, in which quasars reside within dark matter haloes. The clustering on large scales is described by a two-halo term, from pairs of quasars that occupy different haloes, while the one-halo term on small scales is from pairs residing within the same halo. This link between quasars and their host haloes can be modelled using an HOD. Recent work has aimed to measure the HOD and satellite fraction of quasars, from both observations and simulations (e.g. Leauthaud et al. 2015; Powell et al. 2018; Georgakakis et al. 2019; Oogi et al. 2020). Precise measurements of the small-scale quasar clustering (Mohammad et al. 2020), which is dominated by the 1-halo term, will help to further constrain the models. In this mock challenge, in order to place a conservative upper bound on the systematic error in the measurements from the RSD models, we create mock catalogues covering a wide variety of different HODs.

We also aim to test the models on a range of different cosmologies, since the true cosmology is also not known. The OuterRim simulation is a single N-body simulation, which was run in a cosmology that is consistent with the WMAP7 cosmological parameters. To generate mocks in different cosmologies, it is not feasible to run many N-body simulations with the volume and resolution of OuterRim. Instead, the halo catalogue of the original OuterRim simulation can be modified in order to mimic a catalogue of a different cosmology.

In this section, we describe the HOD models used and methodology for creating mocks in different cosmologies.

4.1 HOD modelling

The HOD describes the average number of central and satellite quasars residing in haloes as a function of the halo mass, M. The total number of quasars per halo is given by the sum of central and satellite quasars,

$$\langle N_{\text{tot}}(M)\rangle = \langle N_{\text{cen}}(M)\rangle + \langle N_{\text{sat}}(M)\rangle. \tag{21}$$

Since quasars are rare, the probability that more than one quasar resides within the same dark matter halo is low.

We construct mock catalogues using five different functional forms for the HODs. This allows us to explore a wide range of HODs and test the impact of the HOD on the model fits.

4.1.1 Smooth step and power law

The first HOD model we consider is the same as was used for the quasar mocks in Zarrouk et al. (2018) and uses a $5 + 1$ parameter HOD (e.g. Tinker et al. 2012), which is motivated by a monotonic relation between quasar luminosity and host halo mass, with the brightest quasars residing in the most massive haloes. The probability that a halo contains a central quasar is given by the smooth step function

$$\langle N_{\text{cen}}(M)\rangle = \tau \frac{1}{2}\left[1 + \text{erf}\left(\frac{\log M - \log M_{\text{cen}}}{\log \sigma_{\text{m}}}\right)\right], \tag{22}$$

where the position of this step is set by M_{cen}. The probability that a halo with mass $M \ll M_{\text{cen}}$ hosts a central quasar is zero, which transitions to a probability of τ for $M \gg M_{\text{cen}}$. τ is the quasar duty cycle, which takes into account that not all central black holes are active. This is defined as the fraction of haloes that host an active central galaxy and sets the height of the step function. In our HOD models, the duty cycle is assumed to be constant, while observations show a dependence on redshift and stellar mass of the host galaxy (e.g. Bongiorno et al. 2016; Georgakakis et al. 2017). The width

of the transition is set by the parameter $\log \sigma_{\text{m}}$. This softening of the step function accounts for scatter in the relation between quasar luminosity and halo mass.

The number of satellite quasars in each halo is Poisson distributed, with mean given by a power law,

$$\langle N_{\text{sat}}(M)\rangle = \left(\frac{M}{M_{\text{sat}}}\right)^{\alpha_{\text{sat}}} \exp\left(-\frac{M_{\text{cut}}}{M}\right), \tag{23}$$

where α_{sat} is the slope of the power law, M_{sat} sets the normalization, and M_{cut} is a cutoff at low masses.

4.1.2 Gaussian

This HOD model is the same as was used in Kayo & Oguri (2012) and Eftekharzadeh, Myers & Kourkchi (2019), in which the total HOD for all quasars is given by a Gaussian

$$\langle N_{\text{tot}}(M)\rangle = \frac{\tau}{\sqrt{2\pi}\log\sigma_{\text{m}}} \exp\left[-\frac{(\log M - \log M_{\text{cen}})^2}{2(\log\sigma_{\text{m}})^2}\right]. \tag{24}$$

This model is motivated by haloes in a narrow mass range hosting a wide range of quasar luminosities. The parameter that sets the mean of the Gaussian, $\log M_{\text{cen}}$, is the halo mass that is most likely to host quasars, and $\log \sigma_{\text{m}}$ sets the width of the Gaussian. As with the previous HOD model, τ is the quasar duty cycle. The parameters M_{m}, f_{N}, and Δ_{m} in equation (6) of Eftekharzadeh et al. (2019) are related to the parameters in equation (24): $M_{\text{m}} = M_{\text{cen}}$, $f_{\text{N}} = \sqrt{\ln(10)}\tau$, and $\Delta_{\text{m}} = \ln(10)\log\sigma_{\text{m}}$.

The satellite fraction, f_{sat}, is an additional parameter used to split the total occupation distribution into the central and satellite HODs. For central quasars, $\langle N_{\text{cen}}(M)\rangle = (1 - f_{\text{sat}})\langle N_{\text{tot}}(M)\rangle$, while for satellites, $\langle N_{\text{sat}}(M)\rangle = f_{\text{sat}}\langle N_{\text{tot}}(M)\rangle$.

4.1.3 Gaussian and power law

We also consider an HOD model where the occupation function for central quasars is given by a Gaussian but with a power law for satellites. The central HOD has the same form as equation (24), and the satellite occupation function is given by equation (23). This model and the remaining HOD models we consider are not physically motivated but are included in order to test our models over a wide range of different HODs.

4.1.4 Sharp step and power law

In this HOD model, the central occupation function is given by a sharp step function,

$$\langle N_{\text{cen}}(M)\rangle = \tau\Theta(M - M_{\text{cen}}), \tag{25}$$

where $\Theta(x)$ is the Heaviside step function, M_{cen} sets the position of the step, and τ is the quasar duty cycle. For satellites, the occupation function is the same power-law function as in equation (23).

4.1.5 Top hat and power law

The final HOD uses a top-hat function for the central quasars, given by

$$\langle N_{\text{cen}}(M)\rangle = \tau\Theta(\log\sigma_{\text{m}} - |M - M_{\text{cen}}|), \tag{26}$$

where $\Theta(x)$ is the Heaviside step function, M_{cen} sets the central position of the top hat, $\log \sigma_{\text{m}}$ is the width, and τ is the duty cycle.

Table 2. HOD parameters used for each of the mock catalogues. The shape of the HOD for central quasars is a smooth step function (Sm), Gaussian (G), top-hat (TH) function, or a sharp step (Sh), while satellites follow either a power law (PL) or a Gaussian (G) distribution. Satellites are also positioned in the halo using dark matter particles from the simulation (Par), or following an NFW profile. The parameter $f_{\rm sat}$ is the satellite fraction, and τ is the quasar duty cycle. $\log M_{\rm cen}$ is the central position of the HOD function for central quasars. $\log \sigma_M$ sets the width of smooth step function, or Gaussian HODs, while $\Delta \log M$ is the width of the top-hat HOD. $M_{\rm sat}$ is the low mass cutoff in the power law for satellites, and $\alpha_{\rm sat}$ is the power-law slope. 'Sat. cut' indicates whether the central HOD was used to add the low mass cutoff in the HOD of satellites. The final column, n, is the number density of quasars in units $10^{-5}(h^{-1}{\rm Mpc})^{-3}$.

HOD	Cen. HOD	Sat. HOD	Sat. pos.	$f_{\rm sat}$	τ	$\log M_{\rm cen}$	$\log \sigma_M$	$\Delta \log M$	$M_{\rm sat}$	$M_{\rm cut}$	$\alpha_{\rm sat}$	Sat. cut	$n/10^{-5}$ $(h^{-1}{\rm Mpc})^{-3}$
HOD0	Sm	PL	Par	0.19	0.012	12.13	0.2	–	15.29	11.61	1.0	No	2.185
HOD1	G	G	NFW	0.07	0.074	12.75	2.0	–	–	–	–	–	2.186
HOD2	TH	PL	NFW	0.60	0.022	12.80	–	0.6	15.03	10.57	1.0	No	2.183
HOD3	G	PL	Par	0.21	0.020	12.80	0.3	–	15.47	10.57	1.0	No	2.183
HOD4	Sm	PL	NFW	0.08	0.014	12.13	0.2	–	15.64	11.61	1.0	No	2.195
HOD5	Sh	PL	NFW	0.17	0.016	12.20	–	–	15.57	10.57	1.0	No	2.184
HOD6	G	PL	NFW	0.56	0.021	13.00	0.3	–	15.05	10.57	1.0	No	2.201
HOD7	TH	PL	Par	0.24	0.023	12.60	–	0.6	15.41	10.57	1.0	No	2.202
HOD8	G	G	NFW	1.00	0.074	12.75	2.0	–	–	–	–	–	2.186
HOD9	Sh	PL	NFW	0.42	0.015	12.30	–	–	15.18	10.57	1.0	No	2.197
HOD10	Sh	PL	Par	0.002	0.017	12.15	–	–	15.36	10.57	1.0	Yes	2.181
HOD11	G	G	NFW	0.10	0.060	12.60	1.6	–	–	–	–	–	2.184
HOD12	G	PL	NFW	0.05	0.014	12.55	0.2	–	15.61	10.57	1.0	Yes	2.188
HOD13	Sm	PL	NFW	0.73	0.016	12.80	0.4	–	14.89	10.57	1.0	No	2.183
HOD14	TH	PL	Par	0.17	0.031	12.40	–	0.3	15.09	10.57	1.0	Yes	2.174
HOD15	G	G	Par	0.50	0.060	12.60	1.6	–	–	–	–	–	2.185
HOD16	TH	PL	NFW	0.12	0.033	12.40	–	0.3	15.23	10.57	1.0	Yes	2.186
HOD17	Sh	PL	NFW	0.04	0.014	12.10	–	–	13.93	10.57	1.0	Yes	2.198
HOD18	Sm	PL	NFW	0.36	0.018	12.50	0.4	–	15.12	10.57	1.0	No	2.182
HOD19	G	PL	Par	0.07	0.012	12.50	0.2	–	15.49	10.57	1.0	Yes	2.187

As with the previous HODs, the satellite occupation function is given by the power law of equation (23).

HOD parameters for the 20 HOD models we use are given in Table 2. When using the HODs to populate haloes, we assume that the probability that a halo contains a central or satellite quasar is independent, making it possible for a halo to contain satellite quasars with no central. In the HODs described above, the power law for satellites can extend to low masses, below any cutoff in the centrals. For some of the HOD models we use, the power law is continued to low masses. For others, we multiply the satellite HOD by the central HOD, so the satellites have the same cutoff as the centrals. The various HOD models described above are illustrated in Fig. 1. For each of the five HOD models, four sets of HOD parameters are used to give a total of 20 HODs.

We have considered a wide range of HOD models, some of which are not physically motivated (the models with sharp cuts). While this could potentially inflate the final uncertainties that we measure for the different RSD models, we find that there is no strong dependence in our results on the particular HOD model (see Section 5.3).

The 20 chosen sets of HOD parameters are tuned to produce approximately the same large-scale clustering and quasar number density. This tuning of the HOD parameters is done so that the clustering on scales $s > 20\,h^{-1}{\rm Mpc}$ is in agreement with the DR16 clustering measurements, and the number density of quasars in all mocks is approximately the same, which is close to $2 \times 10^{-5}(h^{-1}{\rm Mpc})^{-3}$.

In the following sections, we describe the methodology for creating the non-blind and blind mocks for this mock challenge. Table 3 summarizes all the mocks that have been produced in this work.

4.2 Non-blind mocks

The mock catalogues for the non-blind part of the mock challenge are constructed by populating the OuterRim simulation snapshot at $z = 1.433$ with quasars using the 20 HOD models outlined in the previous section. The effects of redshift smearing and catastrophic redshifts are then applied to the mocks. These observational effects are described in Section 4.2.1, and the full methodology for creating the mocks is given in Section 4.2.2.

4.2.1 Observational effects

To test that the models are robust against the effects of redshift smearing (Section 2.1) and catastrophic redshifts (Section 2.2), we consider four different cases of mocks.

The 'no smearing' mocks do not contain the effects of redshift smearing or catastrophic redshifts.

The 'Gaussian smearing' mocks include the effect of redshift smearing, where the redshift of each quasar is shifted by a random Δz, which corresponds to a velocity shift, Δv (Section 2.1). Δv is drawn from the redshift-dependent Gaussian distribution (equation 1), which is defined in the survey requirements document.

We also create mocks with 'realistic smearing', where for each quasar, Δv is randomly drawn from a double-Gaussian distribution (equation 2). This distribution includes the wide tails, extending to high velocities, which are seen in the data. We keep the shape of this distribution fixed with redshift. Including this double-Gaussian redshift smearing in the mocks will test whether the models can recover the expected parameters with a realistic redshift smearing distribution. The mocks with Gaussian redshift smearing will test the models for the effect of the degradation of redshift accuracy at high redshifts.

The 'catastrophic redshifts' mocks are identical to the 'realistic smearing' mocks but, in addition, include the effect of catastrophic redshifts. A random 1.5 per cent of objects (which matched the catastrophic redshift failure rate of the data) are assigned a new redshift, which is drawn from a uniform distribution in z, such that the object remains inside the cubic box.
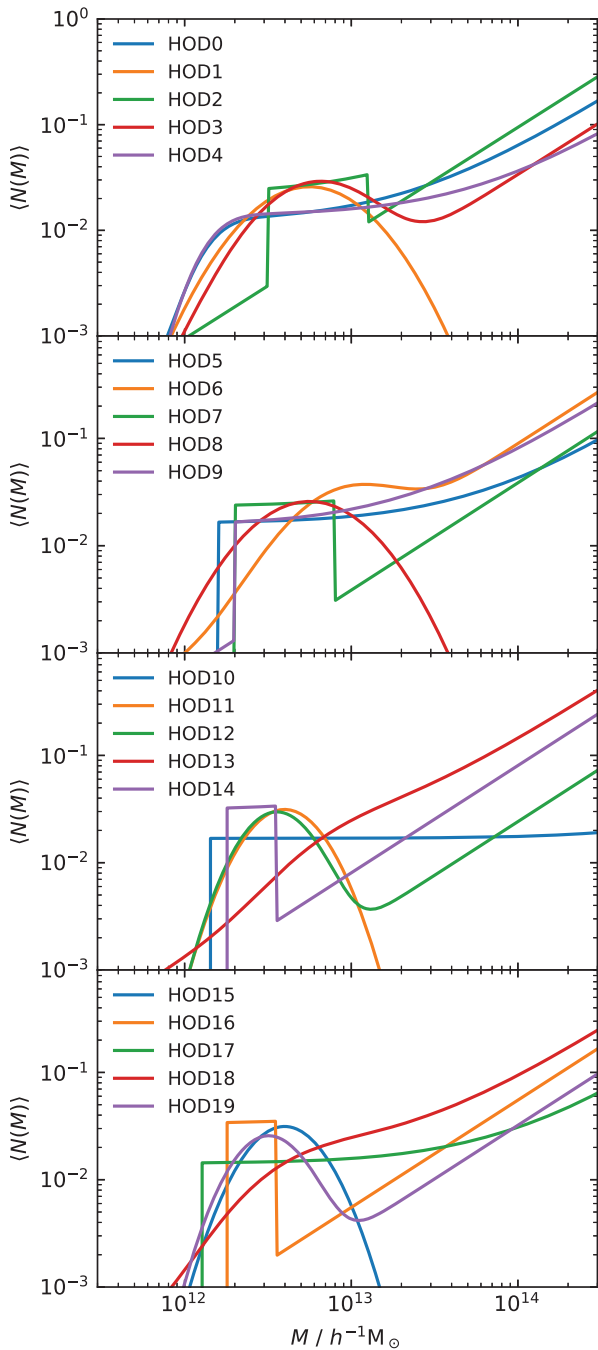
independent of each other, so it is possible for a halo to contain a satellite quasar with no central.

Central quasars are positioned at the centre of each halo and are assigned the same velocity. Satellite quasars are positioned around the centre of the halo using one of two methods, depending on the HOD (see Table 2). For some of the HOD models, the satellites are assigned the position and velocity of dark matter particles, which are randomly chosen from the FOF group belonging to that halo. For other HOD models, the haloes are positioned randomly, following a Navarro, Frenk & White (1997) (NFW) density profile, using a concentration–mass relation from Ludlow et al. (2014). The satellites are assigned a random virial velocity, where the component in each direction is drawn from a Gaussian distribution with variance $\sigma^2(M) = GM_{\rm vir}/2R_{\rm vir}$. The virial radius is determined from the halo mass using a relation measured from the MDPL2 simulation (Klypin et al. 2016).

The method used to position satellite quasars affects the clustering measurements on very small scales (on the order of a few h^{-1}Mpc), which are smaller than the scales used in the analysis.

For each of the 20 HOD models, we generate 100 independent random realizations. Since the quasar duty cycle is of the order of 1 per cent, it is expected that each halo would appear once in the full ensemble (on average). Since the same N-body simulation is used, the 100 realizations of each HOD are not independent. However, we have measured the cross-correlation between pairs of mocks and found that each pair of individual realizations is uncorrelated to within the statistical precision, due to the low duty cycle.

To convert to redshift space, the velocity of each halo is projected along the line of sight of the observer. Before converting to redshift space, the periodic box is first replicated, which takes into account any objects that move in or out of the box. The box is then cut back to the original volume.

The observer is positioned at a distance of 2800 h^{-1}Mpc from the centre of the box in the positive or negative direction along one of the simulation axes. This places the observer so that it is facing one of the six box sides. Since the observer is not at infinity, the lines of sight are not parallel. The choice of observer affects the clustering measurements and particularly impacts the quadrupole and hence the best-fitting values of $f\sigma_8$ (see Section 5.3.1). This is a statistical effect, which is due to cosmic variance in the finite simulation box. In order to mitigate the impact of observer position on the $f\sigma_8$ measurements, we alternate between which of the six box sides the observer is placed it. As is shown in Smith et al. (2020), there is an anticorrelation between measurements of the quadrupole with different lines of sight. Therefore, the reduction in the error when averaging over all the box sides is greater than what would be gained by tripling the simulation volume.

The comoving position to each QSO is then converted to a cosmological redshift, $z_{\rm cos}$. The observed redshift, which takes the line-of-sight velocity, $v_{\rm los}$, into account, is calculated from $(1 + z_{\rm obs}) = (1 + z_{\rm cos})(1 + v_{\rm los}/c)$, which adds evolution of velocity (and hence f) to the mocks. For each mock, to test the effects of redshift smearing and catastrophic redshifts, we create a 'no smearing', 'Gaussian smearing', 'realistic smearing', and 'catastrophic redshift' version, as described in Section 4.2.1.

4.3 Blind mocks

In addition to the non-blind mocks, we also create mocks in rescaled cosmologies, which are analysed blindly, using the OuterRim cosmology as the fiducial cosmology. We describe the method for rescaling the cosmology of the simulation in Sections 4.3.1–4.3.3,

Figure 1. The total halo occupation function for all 20 HOD models used to construct the non-blind quasar mock catalogues. For clarity, the models are split into four panels, where the HOD model is indicated by the colour, given in the legend.

4.2.2 Methodology for creating non-blind mocks

To construct mocks from the OuterRim simulation, the $z = 1.433$ snapshot is populated with quasars using our HOD models. For each halo, with mass M, a uniform random number $0 < x < 1$ is drawn, and the halo is chosen to contain a central quasar if $x < \langle N_{\rm cen}(M)\rangle$. The number of satellites in each halo is drawn from a Poisson distribution with mean given by $\langle N_{\rm sat}(M)\rangle$. The central and satellite HODs are

Table 3. Summary of all the mock catalogues constructed for the quasar mock challenge. This includes the non-blind mocks, the blind mocks with fixed HOD, and the blind mocks where the HOD is varied. For each, N_{tot} is the total number of sets of 100 mocks, N_{HOD} is the number of unique HODs used, and N_{cosmo} is the number of unique cosmologies. We also indicate whether observational effects (redshift smearing and catastrophic redshifts) are included, the location of tables where more information about the individual mocks can be found, and the sections that describe the construction of the mocks and discuss the results.

Mocks	N_{tot}	N_{HOD}	N_{cosmo}	Obs. effects	Table	Construction	Results
Non-blind	20	20	1	Yes	Table 2	Section 4.2.2	Section 5.3
Blind (fixed HOD)	8	1	8	No	Table 4	Section 4.3.5	Section 6.2.1
Blind	24	7	8	No	Table 5	Section 4.3.5	Section 6.2.2

which is validated in Section 4.3.4. In Section 4.3.5, we apply this method to the OuterRim simulation to construct the blind mocks.

4.3.1 Changing the cosmology

In order to rescale the halo catalogues, we use the methodology of Mead & Peacock (2014a, 2014b). This is an extension of the method of Angulo & White (2010), where the rescaling was applied to the dark matter particles. Scaling the halo catalogue makes it easier to modify the cosmology of large simulations, such as OuterRim, since the halo catalogue is much smaller in size, and particle data are not always available. Recently, an alternative method of warping the simulation cosmology was outlined in Garrison & Eisenstein (2019), which requires a set of *N*-body simulations in different cosmologies. However, since there is only one OuterRim simulation, in one cosmology, we use the method of Mead & Peacock (2014a). In this section, we give a brief overview of the rescaling procedure. Our PYTHON implementation is publicly available.[2]

The aim of the rescaling procedure is to take a snapshot of the OuterRim simulation at redshift z and scale the halo properties to a new 'target' cosmology at redshift z'. Quantities in the target cosmology are denoted with a prime, while quantities without a prime are in the original OuterRim cosmology. To create our blind mocks, we rescale the simulation to a target cosmology $z' = 1.433$, which is the same redshift as the unblind mocks, and close to the effective redshift of the quasar sample.

4.3.2 Global rescaling of simulation units

The first stage aims to modify the halo mass function for the new cosmology. This is done with a global rescaling of the units of the simulation. Comoving position vectors, $\mathbf{x}$, in units h^{-1}Mpc are scaled by a factor s,

$$\mathbf{x}' = s\mathbf{x}, \tag{27}$$

which as a result also scales the box size to $L' = sL$. For masses, M, in units $h^{-1}M_\odot$, the scaling is

$$M' = s_m M \equiv s^3 \frac{\Omega'_m}{\Omega_m} M, \tag{28}$$

where Ω_m is the matter density parameter. Finally, velocities $\mathbf{v}$, in proper km s^{-1}, are scaled as

$$\mathbf{v}' = s_v \mathbf{v} \equiv s \frac{H'(z')f'(z')}{1+z'} \frac{1+z}{H(z)f(z)} \mathbf{v}, \tag{29}$$

where $H(z)$ is the Hubble parameter and $f(z)$ is the growth rate.

The values of s and z are found such that the rms linear density fluctuation in the rescaled cosmology, $\sigma(R/s, z)$, matches that of the target cosmology, $\sigma'(R, z')$ [by minimizing equation (2) of Mead & Peacock 2014a]. Simple models of the halo mass function depend only on $\sigma(R, z)$ (e.g. Press & Schechter 1974; Sheth & Tormen 1999), so if $\sigma(R, z)$ is correctly reproduced, the halo mass function will also be correct. This is not strictly true for the mass function of an *N*-body simulation, but deviations in the mass function are expected to be of the order of a few per cent (Mead & Peacock 2014a).

For many potential target cosmologies, the value of z found will not correspond to a simulation snapshot. We therefore choose only cosmologies where the input redshift, z, matches the redshift of one of the OuterRim snapshots. This places a constraint on the combinations of cosmological parameters that are allowed. Despite this, it is still possible to rescale the simulation to a wide variety of new cosmologies.

4.3.3 Modifying the power spectrum

The scaling of comoving positions by a factor s also shifts the BAO scale to a new position sr_{bao}. However, this is not necessarily the correct BAO scale, r'_{bao}, for the target cosmology. This can be seen in Fourier space as residual wiggles when taking the ratio of the rescaled power spectrum to the target power spectrum. In addition, the overall shape of the scaled power spectrum, while close to the target cosmology, is not necessarily correct. To correct the shape of the power spectrum, small displacements are applied to the rescaled positions and velocities of each halo, using the Zel'dovich approximation.

The Lagrangian displacement field, $\boldsymbol{\psi}$ (see equation 14), is related to the matter density field, δ. In Fourier space, this can be written as

$$\boldsymbol{\psi}_k = -i \frac{\delta_k}{k^2} \mathbf{k}. \tag{30}$$

Therefore, the displacement field can be obtained from the Fourier transform of the density field. The density field can be determined from the halo catalogue, after the first stage of rescaling, by computing the overdensity of haloes on a grid and debiasing using the effective bias of the sample of haloes. We use a grid with 750^3 cells, which corresponds to a cell size of $\sim 4h^{-1}$Mpc (the exact cell size depends on the size of the rescaled box).

The differential displacement, due to the difference between the original and target cosmologies is given by

$$\delta\boldsymbol{\psi}'_{k'} = \left[\sqrt{\frac{\Delta'^2_{lin}(k', z')}{\Delta^2_{lin}(sk', z)}} - 1\right]\boldsymbol{\psi}'_{k'}, \tag{31}$$

where

$$\Delta^2_{lin}(k, z) = 4\pi \left(\frac{k}{2\pi}\right)^3 P_{lin}(k, z) \tag{32}$$

is the dimensionless linear power spectrum. Each halo is then displaced so that the final position vector is

$$x'' = x' + b'(M')\delta\psi',\tag{33}$$

where $b'(M')$ is the halo bias. The displacement is multiplied by the halo bias to ensure that the final rescaled snapshot has the correct mass-dependent bias. A similar differential displacement is also applied to the velocity field (see equation 32 of Mead & Peacock 2014a).

The full halo catalogue is used to compute the displacement field, so that the displacement field is not noisy. To ensure that the mass-dependent bias is correct, the halo bias $b(M)$ is used when displacing the halo positions. This means that different haloes are displaced by different amounts, which is unphysical, e.g. a massive halo would be displaced a greater distance than a smaller satellite halo, and therefore the large halo could 'overtake' the satellite. Since the typical halo displacement is very small,[3] any effect on the halo clustering would be on scales much smaller than are used in our analysis.

In the Mead & Peacock (2014a) method, the halo bias is calculated using the peak background split. However, since haloes in the Outer-Rim simulation are identified as FOF groups with linking length $b = 0.168$, and not the more standard $b = 0.2$, the Sheth & Tormen (1999) halo bias differs with the halo bias measured from the simulation. We therefore measure the halo bias directly from the simulation, from the clustering of haloes in the original OuterRim simulation in five mass bins at four different redshifts, and modify the parameters of the Sheth–Tormen halo bias to match the measurements of the bias.

4.3.4 Validation

To validate the rescaling procedure, we rescale one of the snapshots of the Multidark Planck 2 (MDPL2) simulation (Klypin et al. 2016) to the cosmology of the Millennium-XXL (MXXL) simulation (Angulo et al. 2012). This allows us to compare the halo-clustering measurements of the rescaled snapshot to the MXXL simulation, which was run in the target cosmology. These two simulations are chosen since they both have the same halo mass definition (FOF groups with linking length $b = 0.2$), enabling a direct comparison. This is different to the OuterRim simulation, which uses $b = 0.168$.

The MDPL2 simulation has a box size of $1\ h^{-1}$Gpc in a Planck cosmology with $\Omega_m = 0.3071$, $\Omega_b = 0.0482$, $h = 0.6777$, $\sigma_8 = 0.8228$, and $n_s = 0.96$ (Planck Collaboration XVI 2014). The MXXL simulation is a $3\ h^{-1}$Gpc box, in a WMAP1 cosmology with $\Omega_m = 0.25$, $\Omega_b = 0.045$, $h = 0.73$, $\sigma_8 = 0.9$, and $n_s = 1$ (Spergel et al. 2003). The MDPL2 snapshot at $z = 1.425$ is rescaled to $z = 1.67$,[4] which is very close to the redshift of one of the MXXL simulation snapshots ($z = 1.63$). Comoving positions, halo masses, and velocities are scaled by the factors $s = 1.049$, $s_m = 0.940$, and $s_v = 0.993$, respectively.

The mass function of the MDPL2 snapshot, after rescaling, is shown in Fig. 2, in comparison with the target mass function of the MXXL simulation. The scaling of halo masses is able to reproduce the target mass function to a level of a few per cent, which is in agreement with Mead & Peacock (2014a).

The clustering of the MDPL2 snapshot, measured at each stage of the rescaling procedure, is shown in Fig. 3, in comparison with
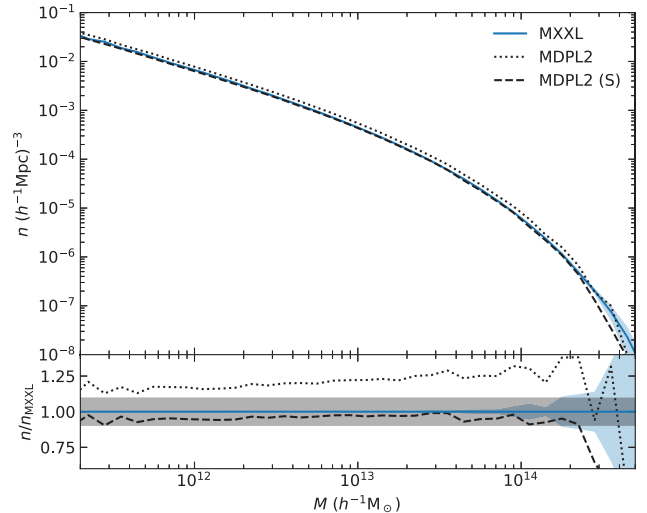
---

[3]The distribution of halo displacements peaks at ~0.2 h^{-1}Mpc, with only ~0.04 per cent of haloes being displaced by greater than $1\ h^{-1}$Mpc.
[4]The rescaled redshift is not arbitrary and is determined by the initial and target cosmologies.

Figure 2. Top panel: Halo mass function of the MDPL2 simulation snapshot at $z = 1.425$ before (black dotted curve) and after (black dashed curve) rescaling the simulation to the MXXL cosmology at $z = 1.67$ The blue curve shows the mass function of the MXXL simulation, where the shaded region indicates the 1σ scatter between 8 subvolumes with the same volume as the rescaled MDPL2 snapshot. Bottom panel: Ratio to the MXXL mass function. The grey shaded region indicates 10 per cent.

the MXXL simulation, where both catalogues have a mass cut of $M > 10^{12}\ h^{-1}M_\odot$ applied. For the MDPL2 snapshot, this mass cut is applied to the rescaled mass, M'. The shaded regions indicate the cosmic variance, estimated from the 1σ scatter between eight sub-cubes of MXXL, which have the same volume as the rescaled MDPL2 snapshot. Clustering measurements from MDPL2 are the average of six measurements, with the observer placed at each of the six box sides (see Section 5.3.1). In the MDPL2 simulation, the BAO length scale is smaller than for the MXXL simulation, due to the differences in cosmology. The first stage of the rescaling procedure shifts the MDPL2 BAO peak to larger scales, by the factor s, but this is not enough to reproduce the monopole of the MXXL simulation. The second step of displacing the halo positions is necessary and brings the BAO position into excellent agreement with the measurements from the MXXL simulation. The shape of the monopole on smaller scales is also in good agreement, within the expected cosmic variance. The amplitude of the quadrupole is also shifted during the rescaling procedure. Scaling velocities by the factor s_v has the largest effect on the quadrupole, with a smaller shift when the velocities are displaced. After the scaling and displacements, the quadrupole is in good agreement with the MXXL simulation and is within the expected cosmic variance. This validates that the rescaling of halo positions is working as expected. For velocities, this is less clear, since the cosmic variance is large compared to the shift in the quadrupole when displacing the velocities. As an additional check, after creating the blind OuterRim mocks, we fit the clustering measurements using the CLPT model in the true rescaled cosmology of the mock (Section 6.1). Recovering $\alpha_\parallel = \alpha_\perp = 1$ and the value of $f\sigma_8$ expected for that cosmology will validate the position and velocity scaling.

4.3.5 Methodology for creating blind mocks

Blind mocks are constructed by applying the rescaling procedure outlined in Section 4.3.1 to snapshots of the OuterRim simulation.
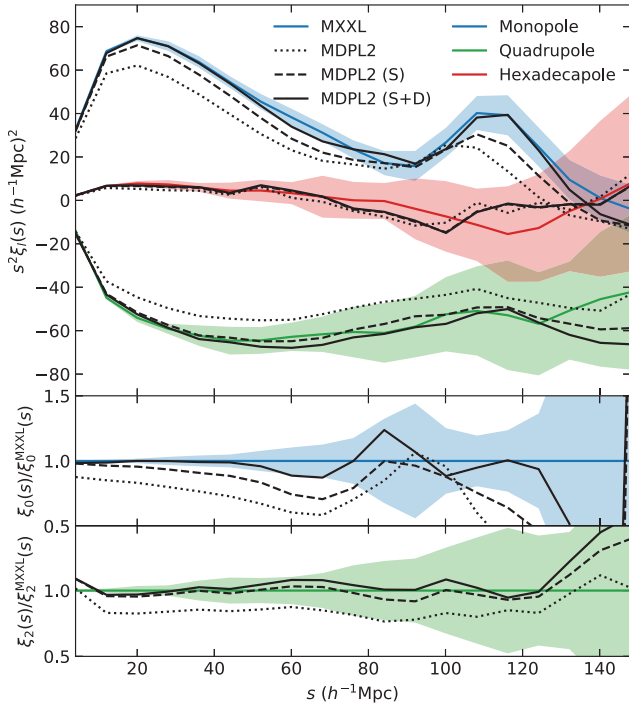
Figure 3. Top panel: Correlation function monopole (blue), quadrupole (green), and hexadecapole (red) of the MXXL simulation snapshot at $z = 1.63$. Black curves show the clustering of the MDPL2 simulation snapshot at $z = 1.425$, which has been rescaled to the Millennium cosmology at $z = 1.67$, at each stage of the rescaling procedure. The dotted curves indicate the clustering in the original MDPL2 snapshot before scaling. The dashed curves show the clustering after the global scaling of positions, masses, and velocities (S). Solid curves indicate the clustering after scaling and additional position and velocity displacements (S + D). For the MXXL simulation, the shaded regions indicate the scatter between eight subvolumes with the same volume as the rescaled MDPL2 simulation. For both simulations, a mass cut of $10^{12} \, h^{-1} \mathrm{M_\odot}$ is applied. Lower panels: ratio of the monopole and the quadrupole to MXXL.

The snapshots at $z = 1.494$ and $z = 1.376$ are rescaled to eight new cosmologies at $z' = 1.433$, which is the same redshift as the snapshot used to create the non-blind mocks. The new cosmologies are chosen by randomly modifying the fiducial OuterRim values of Ω_b, Ω_cdm, h, and σ_8 by multiplying by a random number generated from a Gaussian distribution. The parameter n_s is then fixed by requiring the redshift to match the snapshot redshift when rescaling. Cosmological parameters are shown in Table 4.

The dimensionless linear power spectrum, $\Delta^2(k)$, defined in equation (32) is shown in Fig. 4 for each of the eight new cosmologies, compared to the original OuterRim cosmology, at $z = 1.433$. The second panel shows the ratio between each of these cosmologies and OuterRim, highlighting the differences between them. The bottom panel shows the ratio of each of the power spectra, after the initial rescaling of positions, masses, and velocities by the factors s, s_m, and s_v, respectively, to the target cosmology. While this scaling brings the power spectra close to the target cosmology, there are still differences in the shape, and residual BAO wiggles can be seen. This is corrected for in the second part of the rescaling procedure, where the halo positions and velocities are displaced using the Zel'dovich approximation.

The methodology for populating the blind mocks with quasars is the same as for the unblind mocks. Quasars are added to each halo

by re-using some of the HODs, unchanged, from Section 4.1. While the used HODs were tuned to give the same number density for the non-blind mocks, this is no longer true for the blind mocks, since the underlying halo mass function changes when the cosmology is changed. The largest change in number density is by a factor of ~20 per cent. As before, central quasars are positioned at the centre of each halo, but since particle data are not available for these snapshots, satellites must be positioned following an NFW profile. Since the cosmology has been modified, the halo concentrations should also be modified to reflect this change in cosmology. The positioning of satellites within their host haloes affects only the clustering on very small scales, which are smaller than the scales we consider in the RSD analysis. This is shown with the non-blind mocks, where the recovered parameters are unaffected by the method used to position the satellites. Therefore, we use the same concentration–mass relation as was used for the non-blind mocks.

For the first part of the blind mock challenge, we create eight sets of mocks, where each of the eight rescaled OuterRim boxes is populated using the exact same HOD (HOD0 from the non-blind challenge). As before, 100 realizations are generated for each of the cosmologies. Since the HOD is the same, this can be used to assess the effect of changing the cosmology on the recovered parameters.

For the second part of the mock challenge, we create 24 sets of mocks. These mocks use the same eight rescaled snapshots, which are each populated using three different HODs. In some of these mocks, an additional scaling is applied to the velocities, where all the velocities are increased or decreased by a few per cent. This has the effect of modifying the expected value of $f\sigma_8$ by the same factor. The cosmology, choice of HOD, and velocity scaling factors are given in Table 5.

5 NON-BLIND CHALLENGE

In this section, we present the analysis and results of the non-blind part of the mock challenge. These mocks have been constructed using the HOD models and methodology described in Section 4.2.

5.1 Clustering measurements

The correlation function and power spectrum multipoles are measured for each of the individual mock catalogues. Correlation functions are measured using the publicly available correlation function code CUTE (Alonso 2012), in evenly spaced bins of width $8 \, h^{-1}\mathrm{Mpc}$, up to a maximum separation of $168 \, h^{-1}\mathrm{Mpc}$. A random catalogue is used, which contains 20 times of the number objects that are in the mocks. While the mocks are generated in a cubic box where $\bar{n}$ is known, we use a random catalogue. This is because the lines of sight are not parallel for our choice of observer position, so periodic boundary conditions cannot be used. For each object in the random catalogue, a uniform random value is drawn for its x, y, and z coordinate in the cubic box. In all correlation function calculations, an OuterRim fiducial cosmology is used. Since the volume of the OuterRim box is much larger than the effective volume of the quasar sample ($V_\mathrm{eff} = 1.7 \, h^{-3}\mathrm{Gpc}^3$), and there are 100 realizations for each HOD, precise measurements can be made of the clustering and hence the systematic offsets in the cosmological parameter measurements. Uncertainties in the measurements from each HOD model can be estimated from the scatter between realizations.

The average correlation function multipoles of the 20 sets of mocks, without redshift smearing, are shown in Fig. 5, in comparison with the DR16 measurements. This shows that the clustering of the mocks is in good agreement with the data. There is a small amount

Table 4. Cosmologies of the rescaled OuterRim snapshots, which are used to construct the blinded mocks. z_{orig} is the redshift of the original snapshot, which is rescaled to the new cosmology at $z = 1.433$. s, s_m, and s_v are the factors used to scale positions, masses, and velocities, respectively. Values of $\alpha_{\parallel}$ and $\alpha_{\perp}$ are calculated using a fiducial cosmology, which is the same as the cosmology of the original OuterRim simulation. $f\sigma_8$ is evaluated in each of the cosmologies at the target redshift of $z = 1.433$.

	Ω_b	Ω_{cdm}	h	σ_8	n_s	z_{orig}	s	s_m	s_v	$f\sigma_8$	$\alpha_{\parallel}$	$\alpha_{\perp}$
OR	0.0448	0.2200	0.7100	0.8000	0.9630	1.433	1.0000	1.0000	1.0000	0.3820	1.0000	1.0000
cosmo0	0.0461	0.2205	0.7228	0.7742	0.9628	1.494	0.9827	0.9555	0.9734	0.3694	0.9988	1.0000
cosmo1	0.0426	0.2360	0.6967	0.7981	0.9384	1.494	1.0175	1.1085	1.0308	0.3795	0.9911	1.0002
cosmo2	0.0410	0.2331	0.7405	0.7380	0.9594	1.494	0.9087	0.7766	0.9129	0.3514	0.9745	0.9805
cosmo3	0.0447	0.2408	0.6882	0.7903	0.9436	1.494	0.9950	1.0624	1.0208	0.3750	0.9966	1.0102
cosmo4	0.0467	0.2202	0.6964	0.7991	0.9815	1.376	0.9628	0.8997	0.9788	0.3812	1.0090	1.0105
cosmo5	0.0382	0.1973	0.7197	0.8526	0.9606	1.376	1.1049	1.1992	1.0535	0.4103	0.9926	0.9719
cosmo6	0.0541	0.2295	0.7275	0.7910	0.9671	1.376	0.9327	0.8688	0.9779	0.3756	1.0113	1.0238
cosmo7	0.0475	0.1844	0.7239	0.7783	1.0280	1.376	0.9603	0.7756	0.9086	0.3749	1.0220	0.9981
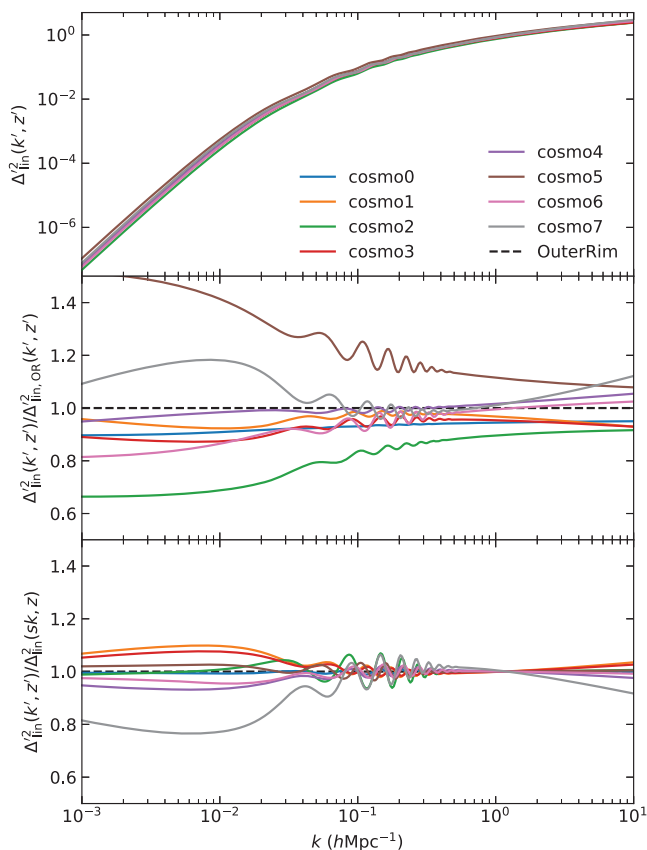
Figure 4. Top panel: Dimensionless linear power spectra of the eight new cosmologies at the target redshift $z' = 1.433$ (coloured curves, as indicated in the legend), compared to OuterRim (black). Middle panel: Ratio of the power spectra to the OuterRim power spectrum. Bottom panel: Ratio of the power spectrum, rescaled by the factor s, to the target power spectrum.

of scatter between the mocks, but this is much smaller than the DR16 error bars, which are estimated using the EZmocks (Zhao et al. 2020).

Power spectra are computed using the NBODYKIT package (Hand et al. 2018), which uses the Yamamoto estimator (Yamamoto et al. 2006). A random catalogue is also used when measuring the power spectrum, since the distance to the observer is not infinite, and periodic boundary conditions cannot be assumed. The computation of the power spectrum is corrected for the effects of the window function (Wilson et al. 2017), using the implementation of de Mattia et al. (2020).

Table 5. Table summarizing the second set of blind mocks. Cosmological parameters are summarized in Table 4, and HOD parameters in Table 2. The velocity scaling is an additional scaling applied to velocities, in which all velocities are scaled by this parameter.

Mock	Cosmology	HOD	Velocity scaling
mockb0	cosmo1	HOD17	$+5\%$
mockb1	cosmo4	HOD11	
mockb2	cosmo0	HOD11	$+5\%$
mockb3	cosmo1	HOD11	
mockb4	cosmo0	HOD17	
mockb5	cosmo4	HOD2	
mockb6	cosmo1	HOD2	-5%
mockb7	cosmo0	HOD2	
mockb8	cosmo4	HOD17	
mockb9	cosmo3	HOD9	$+2\%$
mockb10	cosmo7	HOD16	
mockb11	cosmo5	HOD3	
mockb12	cosmo7	HOD9	$+3\%$
mockb13	cosmo6	HOD9	-8%
mockb14	cosmo3	HOD16	
mockb15	cosmo2	HOD16	-5%
mockb16	cosmo2	HOD3	$+3\%$
mockb17	cosmo5	HOD9	$+8\%$
mockb18	cosmo6	HOD3	
mockb19	cosmo6	HOD13	-2%
mockb20	cosmo3	HOD13	-3%
mockb21	cosmo5	HOD13	
mockb22	cosmo2	HOD13	
mockb23	cosmo7	HOD3	

5.2 Fitting the models

The RSD models described in Section 3 are fit to the correlation function or power spectrum multipoles measured from each of the non-blind OuterRim mocks. The RegPT model is fit in Fourier space to the power spectrum multipoles in the range of $0.02 < k < 0.3$ hMpc^{-1} The covariance matrix used is estimated from the EZmocks (Zhao et al. 2020). Fits are done using the MINUIT algorithm (James & Roos 1975) to minimize χ^2.

The RESPRESSO model is fit in configuration space in 19 evenly spaced bins of separation, in the range of $16 < s < 168$ h^{-1}Mpc. An analytical Gaussian covariance matrix is used, following the prescription of Grieb et al. (2016). Best-fitting values of each parameter are found from the MCMC chains by taking the median.

CLPT is also fit to the correlation function multipoles in 16 evenly spaced bins in s between $24 < s < 152$ h^{-1}Mpc. The same binning was used as for the RESPRESSO model but with a different minimum
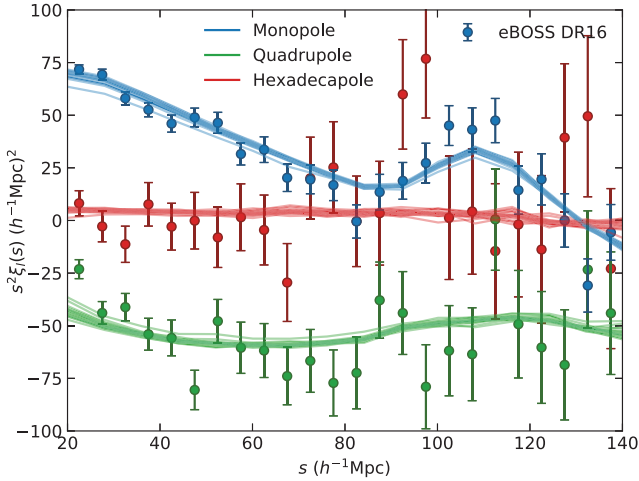
Figure 5. The average clustering of each of the 20 sets of non-blind mocks, showing the monopole (blue), quadrupole (green), and hexadecapole (red), with no redshift smearing. Points with error bars show the eBOSS DR16 quasar clustering measurements, assuming an OuterRim cosmology, where the errors are estimated from the EZmocks.

and maximum scale. Fits were done to the mean clustering of each set of mocks, using MINUIT, with Gaussian covariance matrices.

BAO fits in Fourier space are done in the k range of $0.02 < k < 0.23$ $h\mathrm{Mpc}^{-1}$, using the covariance matrix estimated from the EZmocks. The fits are done using MINUIT. BAO fits in configuration space are done with the monopole and quadrupole, using the same correlation function binning as the RESPRESSO model, and a Gaussian covariance matrix.

5.3 Results

Model fits are shown in Fig. 6 to the power spectrum and correlation function measurements of Mock3 (i.e. the mocks constructed using HOD3). For the RegPT and RESPRESSO models, the fits are done with each of the different cases of redshift smearing. For the CLPT model, the fit is done only with no redshift smearing. This figure highlights the impact of redshift smearing on the clustering measurements.

The impact of redshift smearing on the correlation function is most noticeable in the shape of the quadrupole on small scales, but it also has the effect of increasing the amplitude of the monopole on small scales by a small amount, and also affects the shape of the hexadecapole. The difference between the two models of redshift smearing is small, with the most noticeable difference being in the hexadecapole on small scales. Including catastrophic redshifts reduces the amplitude of the clustering by a small fraction.

Differences in the power spectrum are much more apparent. Including redshift smearing lowers the amplitude of the monopole and the quadrupole at large k, and a larger difference is seen in the shape of the monopole and the quadrupole between the two kinds of redshift smearing. The two-point clustering statistics measured from the mock are well fit by all models, with and without redshift smearing.

The results for all 20 sets of mocks for all redshift smearing cases are shown in Fig. 7. The points in the plot depict the mean values of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ measured from each of the 20 sets of 100 mocks, and the error bars indicate the error on the mean from the 100 individual mocks. The shaded areas indicate 3 per cent in $f\sigma_8$ and 1 per cent in $\alpha_\parallel$ and $\alpha_\perp$. This range represents ~ 30 per cent of the statistical error, which we consider as the limit that would be acceptable for the systematic error.

For the case in which there is no redshift smearing (indicated by the blue shading in Fig. 7), the models are able to recover values of $f\sigma_8$ within 3 per cent and $\alpha_\parallel$ and $\alpha_\perp$ to within 1 per cent for almost all of the mocks. The results obtained using the RegPT and CLPT models are mostly in agreement with each other. However, there is a small offset between these and the results from RESPRESSO. The RESPRESSO model, on average, measures values of $f\sigma_8$ and $\alpha_\perp$, which are systematically smaller, by ~ 2 per cent and ~ 0.5 per cent, respectively, and the values of $\alpha_\parallel$ are ~ 0.5 per cent larger. This difference is explained by the parameter σ_{zerr} in the FoG prescription used in the RESPRESSO model, which models the redshift error. This is a free parameter, but since these mocks do not contain redshift smearing, fitting the value σ_{zerr} leads to a small bias in the measurements of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$. We tested the effect of fixing σ_{zerr} to $\sigma_{\mathrm{zerr}} = 0$, and this has the effect of increasing the best-fitting values of $f\sigma_8$ and $\alpha_\perp$, and reducing $\alpha_\parallel$, bringing the results of all three models into agreement.

The results for the case of Gaussian redshift smearing are indicated by the yellow shading in Fig. 7. Again, all the results are within the target ranges for each of the parameters. With Gaussian smearing, the best-fitting $\alpha_\parallel$ and $\alpha_\perp$ parameters from the RegPT and RESPRESSO models are mostly in close agreement. For $f\sigma_8$, there is a small offset between the two, with the RESPRESSO model measuring values that are systematically slightly higher by ~ 2 per cent. Since these mocks contain redshift smearing, which is modelled in the RESPRESSO model by the parameter σ_{zerr}, the best-fitting parameters measured by the RESPRESSO model are closer to the true values, compared to the no smearing case.

With the redshift smearing modelled as a more realistic double-Gaussian, the results obtained are shown by the green shading in Fig. 7. Again, the results are within 3 per cent for $f\sigma_8$ and 1 per cent for $\alpha_\parallel$ and $\alpha_\perp$ for all models. However, now the $f\sigma_8$ is in agreement between the models, with a small difference of ~ 0.5 per cent in $\alpha_\parallel$. Compared to the case of Gaussian smearing, there is very little change in the results using RESPRESSO. The $f\sigma_8$ from the RegPT model is, on average, ~ 1 per cent larger, with a ~ 0.5 per cent lower value for $\alpha_\parallel$.

Finally, the effect of including catastrophic redshifts is shown by the red shading in Fig. 7. The mocks here are the same as the mocks with realistic redshift smearing, but 1.5 per cent of objects are assigned a random redshift, from a uniform distribution. This has only a small effect on $\alpha_\parallel$ and $\alpha_\perp$ but leads to values of $f\sigma_8$ that are systematically lower by ~ 3 per cent. This effect on $f\sigma_8$ is expected, since assigning objects random redshifts will dilute the clustering, reducing the amplitude of the correlation function quadrupole, resulting in smaller values of $f\sigma_8$. Despite this shift in $f\sigma_8$, the best-fitting values are still within 3 per cent of the expected fiducial value.

In addition to the full-shape analyses, the BAO-only fits in configuration and Fourier space are also shown in Fig. 7 for the different redshift smearing cases. For most mocks, the best-fitting values of $\alpha_\parallel$ and $\alpha_\perp$ are within 1 per cent. The best-fitting values of $\alpha_\perp$ from the two BAO models are in good agreement, while there is an offset of ~ 0.5 per cent in $\alpha_\parallel$, with the Fourier-space fits measuring values that are systematically larger. The inclusion of redshift smearing or catastrophic redshifts does not strongly affect the BAO results.

5.3.1 Observer position

To mitigate the effect of observer position on the average results, we alternate between which of the six box sides the observer is
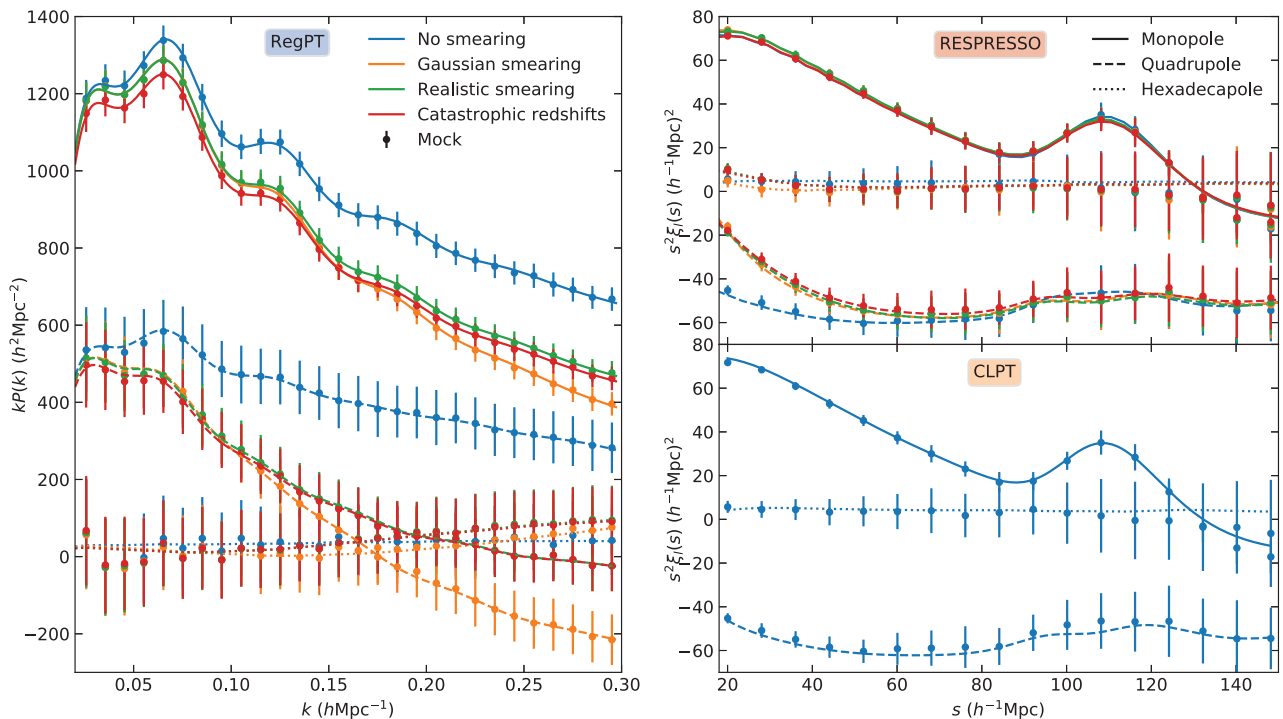
Figure 6. Left-hand panel: Power spectrum multipoles measured from Mock3, and the best-fitting RegPT model, for the cases of no smearing (blue), Gaussian smearing (yellow), realistic smearing (green), and catastrophic redshifts (red). Points indicate the average measurement from the 100 mocks, where the error bar is the 1σ scatter. The best-fitting model is shown by the curves, where the solid, dashed, and dotted curves are the monopole, quadrupole, and hexadecapole, respectively. Right-hand panel: Correlation function multipoles and the best-fitting RESPRESSO model (top) and CLPT model (bottom).

positioned at. If the observer is kept fixed, this can lead to offsets in the measurements of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$.

The impact of observer position is shown in Fig. 8 for Mock0, which shows the offset in the best-fitting parameters for the subset of mocks with observer at each box side, compared to the mean of all mocks. For the α parameters, the shifts are small and within 1 per cent. Since the choice of observer position has only a small impact on the monopole, it is expected that the shifts in the α parameters will be small. However, for $f\sigma_8$, the shifts can be larger than 5 per cent for certain observer positions.

The impact of varying the observer position on the quadrupole, and how this propagates through to the $f\sigma_8$ measurements, is investigated in Smith et al. (2020). The offsets seen in the $f\sigma_8$ measurements depend on the box size and also on the bias of the tracer. For the QSOs, which are highly biased, large offsets of this magnitude are expected. Averaging together the different lines of sight mitigates this effect, reducing the variance by a factor better than 1/3.

5.3.2 Stability of the models

The results of the non-blind mock challenge show that all models are able to recover the expected value of $f\sigma_8$ to within 3 per cent, and $\alpha_\parallel$ and $\alpha_\perp$ to within 1 per cent, even for mocks that are constructed using extreme HOD models, which are not motivated by quasar physics. RSD in the mock catalogues are impacted by the choice of HOD and also the prescription of redshift smearing. However, the nuisance parameters that enter our models are able to absorb these effects in terms of the redshift uncertainty. In this section, we give examples of how the FoG parameters in the models respond to the HOD (in terms of satellite fraction) and redshift smearing. We note that the other nuisance parameters also show some sensitivity to these effects.

In the model that we refer to as 'RegPT' (which combines RegPT with the specific RSD prescription described in Section 3.3.1), the free parameters are $a_{\rm vir}$ and σ_v, while for the 'RESPRESSO' model (Section 3.3.2), the parameters are $a_{\rm vir}$ and $\sigma_{\rm zerr}$.

The different HOD models we consider cover a wide range of satellite fractions, some of which are very high (e.g. HOD13 and HOD8, where the satellite fractions are 73 per cent and 100 per cent, respectively). The parameters that model the RSDs all show trends with the satellite fraction. In the upper panel of Fig. 9, we show the best-fitting values of $\sigma_{\rm zerr}$ as a function of satellite fraction, measured by the RESPRESSO model, for the mocks with Gaussian smearing. The values of $\sigma_{\rm zerr}$ follow a linear trend, which increases with the satellite fraction. Satellite quasars are assigned a random virial velocity (Section 4.2.2), which is similar to satellites having a larger redshift uncertainty. However, the satellite fraction and redshift smearing impact the clustering measurements on different scales. The quadrupole is impacted by the satellite fraction on scales $\sim 40\,h^{-1}{\rm Mpc}$, while redshift smearing has an impact on larger scales, up to $\sim 90\,h^{-1}{\rm Mpc}$.

The bottom panel of Fig. 9 shows the best-fitting values of $a_{\rm vir}$ and σ_v from the RSD prescription of the RegPT model, and also $a_{\rm vir}$ and $\sigma_{\rm zerr}$ of the RESPRESSO model, for different cases of redshift smearing. The treatment of the FoG factor is very similar between the two models, but the RESPRESSO model contains a velocity dispersion 'offset'.[5] The parameters respond to the redshift smearing schemes, but their behaviours differ between the redshift smearing implementations. The velocity dispersion offset in the RESPRESSO

---

[5]In the FoG prescription used in the RESPRESSO model, σ_v is kept fixed to a non-zero value. This results in a k and μ dependence, which does not vanish when $a_{\rm vir} = \sigma_{\rm zerr} = 0$.
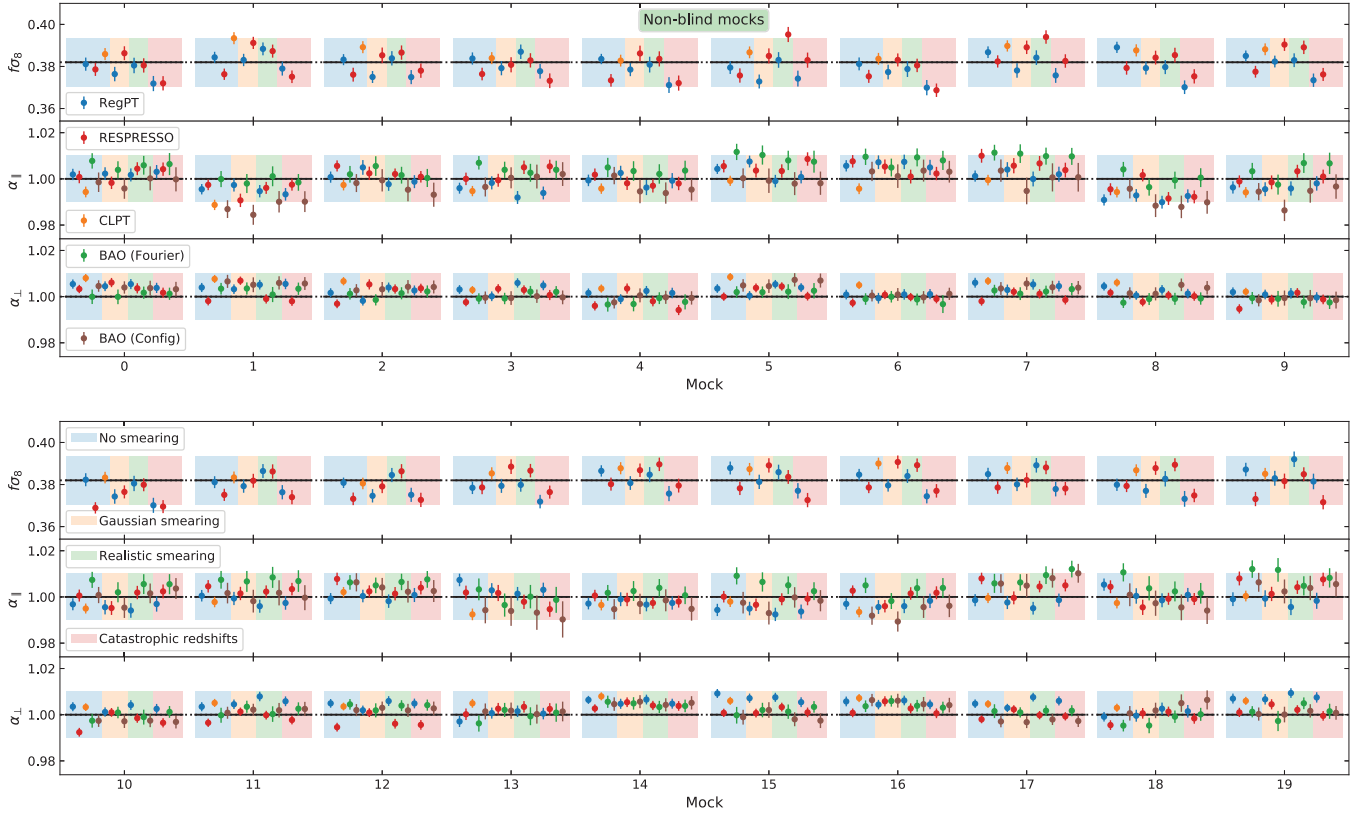
Figure 7. Best-fitting values of $f\sigma_8$ (upper panels), $\alpha_\parallel$ (middle panels), and $\alpha_\perp$ (lower panels), measured from the 20 sets of 100 non-blind mocks, for all redshift smearing prescriptions. The results from the RegPT model are shown in blue, RESPRESSO in red, and CLPT in yellow. The BAO fits in Fourier and configuration space are shown in green and brown, respectively. Points show the mean of the best-fitting parameters of the 100 mocks, and the error bars show the standard error on the mean. Shaded regions indicate 3 per cent for $f\sigma_8$, and 1 per cent for $\alpha_\parallel$ and $\alpha_\perp$. The colour of the shaded region corresponds to the redshift smearing included in the mocks. No smearing is indicated in blue, Gaussian smearing in yellow, realistic smearing in green, and realistic smearing with 1.5 per cent catastrophic redshift errors in red.
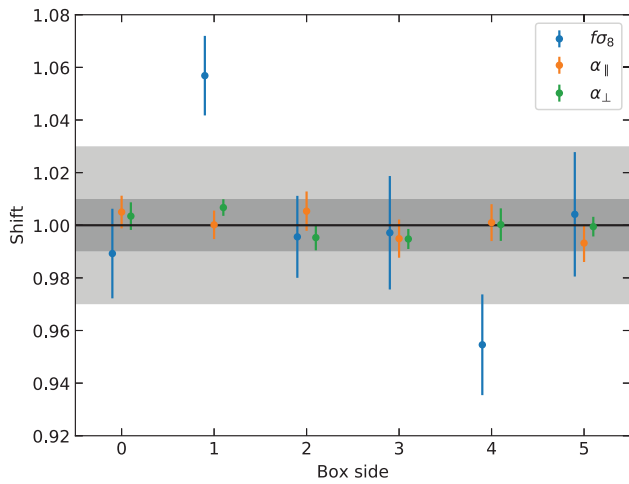
Figure 8. Shift in the best-fitting values of $f\sigma_8$ (blue), $\alpha_\parallel$ (yellow), and $\alpha_\perp$ (green), measured from Mock0 using the CLPT model, with an observer at each of the six box sides, compared to the average value. Error bars indicate the error on the mean. Grey shaded regions indicate 1 per cent and 3 per cent.

model lowers the values of σ_{zerr}. In both models, the parameter a_{vir} never vanishes, which demonstrates that the FoG factor cannot be modelled by a pure Gaussian function. Using a pure Gaussian or a pure Lorentzian FoG factor leads to a biased measurement of $f\sigma_8$.
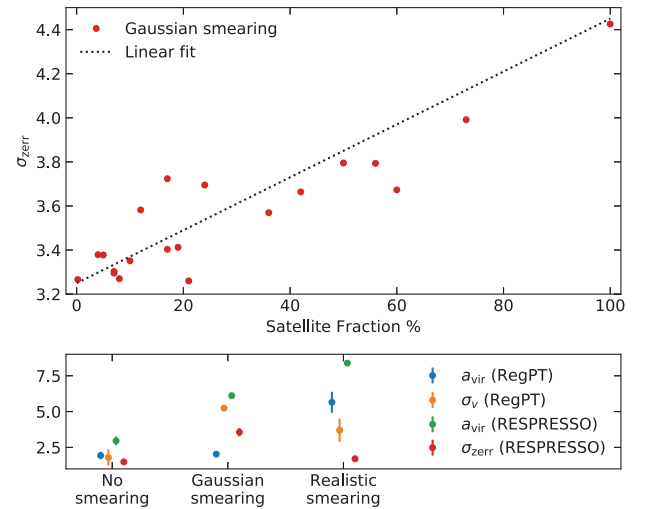
Figure 9. Top panel: Best-fitting values of σ_{zerr} from the RSD prescription used in the RESPRESSO model, plotted as a function of satellite fraction, for the 20 sets of non-blind mocks with Gaussian redshift smearing. The dotted line shows a linear fit. Bottom panel: Average values of the a_{vir} (blue) and σ_v (yellow) parameters, from the RSD prescription of the RegPT model, and a_{vir} (green) and σ_{zerr} (red) of RESPRESSO for the cases of no smearing, Gaussian smearing, and realistic smearing. Error bars indicate the 1σ scatter between mocks.
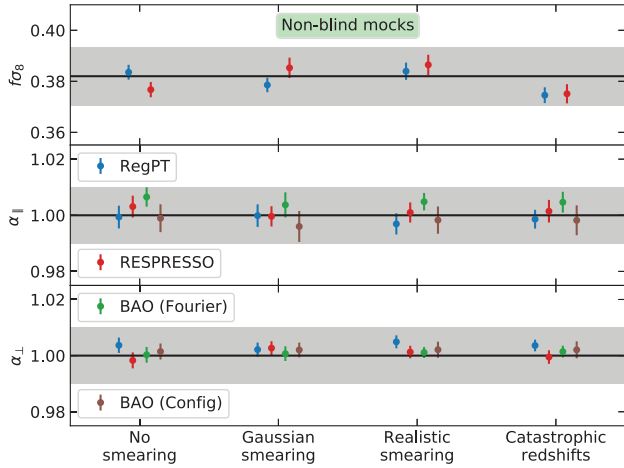
Figure 10. Results from the non-blind mock challenge, averaged over all 20 sets of mock catalogues, for the cases of no redshift smearing, Gaussian smearing, realistic double-Gaussian smearing, and 1.5 per cent catastrophic redshifts. Points show the average values measured using RegPT (blue), RESPRESSO (red), and BAO fits in Fourier space (green) and configuration space (brown), where error bars show the 1σ scatter between all mocks. Grey shaded regions indicate 3 per cent for $f\sigma_8$ and 1 per cent for $\alpha_\parallel$ and $\alpha_\perp$.

We also checked that our results are stable for different choices of binning in the two-point clustering measurements. For the RESPRESSO model, for five sets of mocks, we re-measure the correlation function multipoles in bins of separation $5\,h^{-1}$Mpc and fit the model to these correlation function measurements. We find that reducing the bin separation from $8\,h^{-1}$Mpc to $5\,h^{-1}$Mpc has a very small effect on our results and does not change our picture of the modelling systematics.

5.3.3 Non-blind systematics

The results from all the non-blind mocks are summarized in Fig. 10. The points are the average of all 20 sets of mocks, for each of the different cases of redshift smearing, where the error bars show the standard deviation. In order to calculate a systematic error, we also calculate the rms, defined as

$$\text{rms} = \sqrt{\sum_i^N \frac{(x_i - x_{\text{true},i})^2}{N}}, \qquad (34)$$

where x_i is the value of a parameter measured from mock i, $x_{\text{true, i}}$ is the true value, and N is the number of mocks. In the case of the non-blind mocks, $x_{\text{true, i}}$ is constant, and the rms is equivalent to adding the offset from the true value, and the standard deviation, in quadrature. The average, standard deviation, and rms of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ for the non-blind mocks are given in Table 6 for all models.

6 BLIND CHALLENGE

6.1 Validation of blind mocks

Before analysing the blind mocks in the OuterRim fiducial cosmology, we first analyse the clustering using CLPT in the target cosmology to ensure that the rescaling of the cosmology is correct. This is done with the initial set of eight mocks, which were constructed using the same HOD. Since we use the target cosmology as the fiducial cosmology, the expected value of $\alpha_\parallel$ and $\alpha_\perp$ is 1, while $f\sigma_8$ varies between the different cosmologies. These results

Table 6. Results from the non-blind mock challenge using the RegPT, RESPRESSO, and BAO models, averaged over all 20 sets of mocks, for the cases of no smearing, Gaussian smearing, realistic smearing, and catastrophic redshifts. $\langle f\sigma_8 \rangle$ is the mean of the 20 values of $f\sigma_8$, $\sigma(f\sigma_8)$ is 1 standard deviation, and rms$(f\sigma_8)$ is defined in equation (34). The same values are provided for $\alpha_\parallel$ and $\alpha_\perp$.

RegPT	No smearing	Gaussian smearing	Realistic smearing	Catastrophic redshifts
$\langle f\sigma_8 \rangle$	0.3836	0.3786	0.3840	0.3746
$\sigma(f\sigma_8)$	0.0029	0.0028	0.0034	0.0031
rms$(f\sigma_8)$	0.0033	0.0044	0.0039	0.0081
$\langle \alpha_\parallel \rangle$	0.9994	0.9999	0.9969	0.9986
$\sigma(\alpha_\parallel)$	0.0040	0.0040	0.0038	0.0033
rms$(\alpha_\parallel)$	0.0041	0.0040	0.0049	0.0036
$\langle \alpha_\perp \rangle$	1.0037	1.0021	1.0049	1.0036
$\sigma(\alpha_\perp)$	0.0027	0.0025	0.0023	0.0021
rms$(\alpha_\perp)$	0.0046	0.0033	0.0054	0.0042
RESPRESSO				
$\langle f\sigma_8 \rangle$	0.3767	0.3853	0.3864	0.3751
$\sigma(f\sigma_8)$	0.0030	0.0040	0.0040	0.0038
rms$(f\sigma_8)$	0.0060	0.0052	0.0060	0.0078
$\langle \alpha_\parallel \rangle$	1.0031	0.9996	1.0010	1.0015
$\sigma(\alpha_\parallel)$	0.0039	0.0036	0.0036	0.0040
rms$(\alpha_\parallel)$	0.0050	0.0036	0.0037	0.0043
$\langle \alpha_\perp \rangle$	0.9983	1.0027	1.0013	0.9995
$\sigma(\alpha_\perp)$	0.0029	0.0025	0.0022	0.0024
rms$(\alpha_\perp)$	0.0033	0.0036	0.0026	0.0025
BAO (Fourier)				
$\langle \alpha_\parallel \rangle$	1.0065	1.0037	1.0048	1.0046
$\sigma(\alpha_\parallel)$	0.0034	0.0045	0.0031	0.0037
rms$(\alpha_\parallel)$	0.0073	0.0058	0.0058	0.0059
$\langle \alpha_\perp \rangle$	1.0003	1.0007	1.0011	1.0014
$\sigma(\alpha_\perp)$	0.0028	0.0026	0.0020	0.0021
rms$(\alpha_\perp)$	0.0028	0.0027	0.0023	0.0025
BAO (Config.)				
$\langle \alpha_\parallel \rangle$	0.9990	0.9960	0.9983	0.9982
$\sigma(\alpha_\parallel)$	0.0050	0.0055	0.0049	0.0053
rms$(\alpha_\parallel)$	0.0051	0.0068	0.0052	0.0056
$\langle \alpha_\perp \rangle$	1.0015	1.0020	1.0021	1.0021
$\sigma(\alpha_\perp)$	0.0028	0.0026	0.0029	0.0030
rms$(\alpha_\perp)$	0.0032	0.0033	0.0036	0.0037

are shown in Fig. 11. For all the mocks, the measured values of $f\sigma_8$ are within 3 per cent, and $\alpha_\parallel$ and $\alpha_\perp$ are within 1 per cent. On average, the measured values of $f\sigma_8$ and $\alpha_\perp$ are higher than the true values by ~ 1 per cent and ~ 0.5 per cent, respectively, while $\alpha_\parallel$ is ~ 0.5 per cent low. The same offsets are seen when analysing the non-blind mocks using CLPT (Fig. 7). This validates that the rescaling method is working as expected, and we now go on to analyse the mocks blind, using the RegPT and RESPRESSO models, with OuterRim as the fiducial cosmology.

6.2 Results

6.2.1 Blind mocks with fixed HOD

The results, when analysing the mocks with an OuterRim fiducial cosmology, are shown in Fig 12. Since the fiducial cosmology no longer matches the true cosmology of the simulation, the expected values of the α parameters vary from 1 by up to 3 per cent, but the values of $f\sigma_8$ are unaffected. The best-fitting parameters are now in
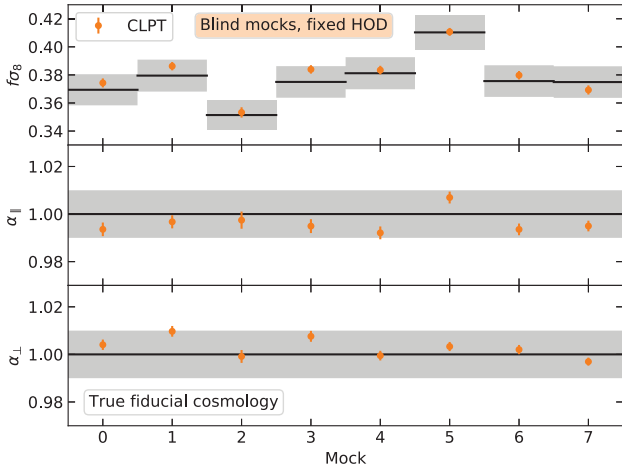
Figure 11. Results from analysing the eight sets of blind mocks, with fixed HOD, with the CLPT model, using the target cosmology as the fiducial cosmology. Points with error bars are the results from CLPT, while the true values are shown in black. The grey shaded regions indicate 3 per cent for $f\sigma_8$ and 1 per cent for $\alpha_{\parallel}$ and $\alpha_{\perp}$.
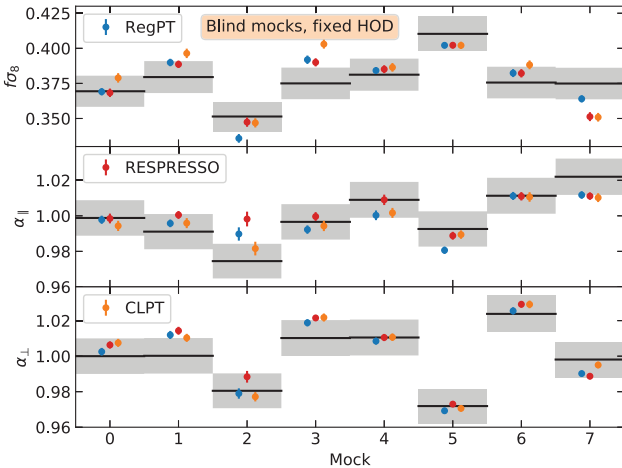

Figure 12. Results from the blind mocks, with fixed HOD, using an OuterRim WMAP7 fiducial cosmology. Points show the results from the RegPT model (blue), RESPRESSO model (red), and CLPT (yellow), with error bars showing 1σ. The expected values are shown by the black lines, with shaded regions indicating 3 per cent for $f\sigma_8$ and 1 per cent for $\alpha_{\parallel}$ and $\alpha_{\perp}$.

less good agreement with the expected results than when the fiducial cosmology was the true cosmology of the rescaled simulation. For many of the mocks, the measured values are outside the ranges of 3 per cent for $f\sigma_8$ and 1 per cent for the α parameters. For example, the measurements of $f\sigma_8$ are approximately 5 per cent too high for mock 3 and 5 per cent too low for mock 7. The variation between the measured and expected values is similar for all of the models.

The origin of these differences is not due to the models themselves, since they are able to recover the expected parameters when the correct fiducial cosmology is used, and we have validated that there is no systematic due to the rescaling procedure. The larger scatter in the results is due to the fiducial cosmology used.

When the galaxy clustering is measured using an incorrect fiducial cosmology, this has the effect of shifting the BAO position by the factors $\alpha_{\parallel}$ and $\alpha_{\perp}$, parallel and perpendicular to the line of sight, respectively. It has been shown that BAO fitting can recover unbiased
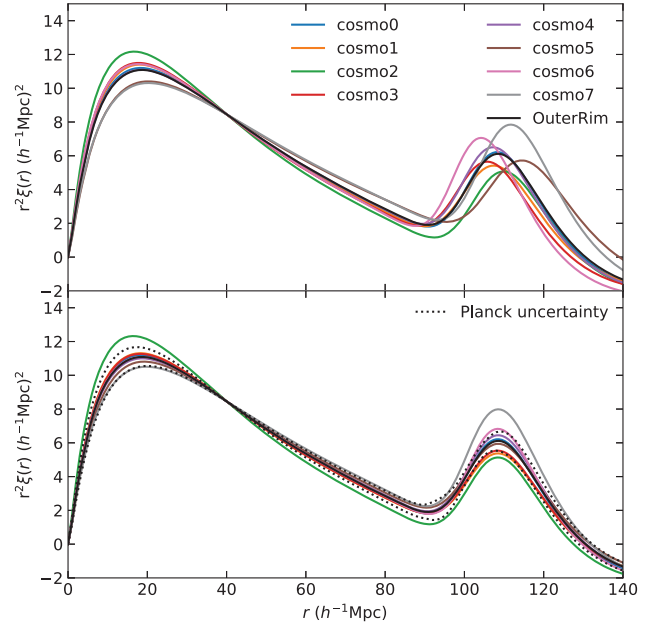
Figure 13. Top panel: correlation function calculated from the linear power spectrum of the OuterRim simulation (black dashed curve) and for the eight new cosmologies (solid coloured curves, as indicated in the legend). Correlation functions have been shifted vertically so that they match OuterRim at $r = 40\ h^{-1}$Mpc. Bottom panel: same as the top panel, but with a rescaling of r so that the BAO peaks align. Black dotted lines indicate the maximum change in the OuterRim $\xi(r)$ when varying the cosmological parameters within the 1σ errors from Planck.

measurements of $\alpha_{\parallel}$ and $\alpha_{\perp}$, over a range of different choices of fiducial cosmology (e.g. Vargas-Magaña et al. 2018; Carter et al. 2020). However, in order to also measure $f\sigma_8$, it is necessary to perform full shape fits to the correlation function and power spectrum multipoles. While the position of the BAO peak in the data will be shifted by the expected α parameters, compared to the template, which is calculated in the fiducial cosmology, the shape of the correlation function on smaller scales will not necessarily match the template, leading to biased measurements of $f\sigma_8$, $\alpha_{\parallel}$, and $\alpha_{\perp}$. The overall shape of the measured correlation function does not change when a different fiducial cosmology is used, but the shape of the template does change.

Correlation functions are shown in the top panel of Fig. 13 in the cosmologies of the eight blind mocks, compared to the OuterRim cosmology. These curves are calculated from the linear power spectrum and have been normalized to have the same amplitude as OuterRim at $r = 40\ h^{-1}$Mpc to emphasize the differences in shape. The templates used in the model fitting also show the same differences in shape. In the lower panel of Fig. 13, the distances have been rescaled to align the position of the BAO peaks, further emphasizing the differences. For most of the cosmologies, the change in shape, compared to OuterRim, is within the range expected when modifying cosmological parameters within the 1σ errors of Planck. The most extreme cosmologies are cosmology 7, which has a high amplitude at the BAO scale, and cosmology 2, where the slope on small scales is much steeper than OuterRim. If there is a large mismatch in the slope on small scales, this leads to biased measurements of $\alpha_{\parallel}$ and $\alpha_{\perp}$. For example, it is not possible to rescale the black OuterRim correlation function curve in Fig. 13 to match both the slope on small scales and the BAO peak position,
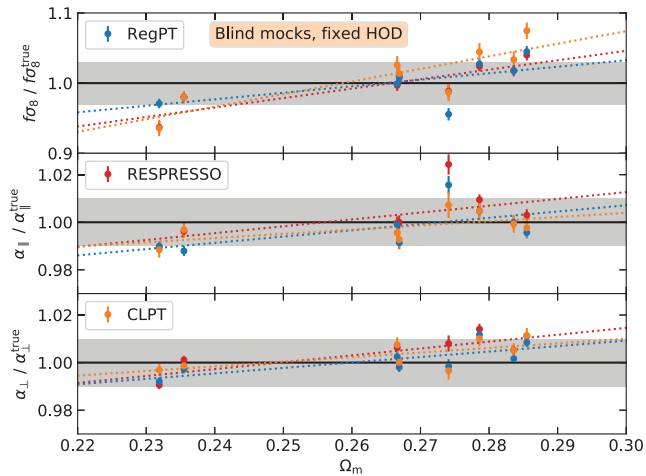
Figure 14. Ratio of the measured values of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ from the blind mocks with fixed HOD to the expected value, as a function of Ω_m. Points are the results using the RegPT model (blue), RESPRESSO model (red), and CLPT (yellow). Dotted lines show a linear fit to the results from each model, as a function of Ω_m. Grey shaded regions indicate 3 per cent for $f\sigma_8$ and 1 per cent for the α parameters.

of cosmology 2, so in the best-fitting model, there will be a systematic offset in the BAO position.

The precise shape of the correlation function is affected by each of the cosmological parameters chosen (except σ_8, which changes only the normalization). The parameter that has the largest effect on the shape is Ω_m. Fig. 14 shows the ratio between the measured and expected values, as a function of Ω_m, for the three RSD models. This reveals a correlation with Ω_m: for cosmologies with low values of Ω_m compared to the fiducial cosmology, the best-fitting values of all the parameters are systematically low, while for the mocks with larger values of Ω_m, the measured parameters tend to be higher than the expected values. For each of the parameters, the dependence on Ω_m is approximately linear for the changes in cosmology considered here, as indicated by the linear fits in Fig. 14. The best-fitting values of $f\sigma_8$ are within 3 per cent of these linear fits, and the α parameters are within 1 per cent, with the exception being the mock with cosmology 2. This is the cosmology that has the largest difference in the shape of the template, with a steep slope on small scales (the green curve in Fig. 13).

6.2.2 Blind mocks with varying HOD

Clustering measurement from the set of 24 blind mocks (with varying HOD and additional velocity scaling) is fit using the RegPT and RESPRESSO models. Results are shown in Fig. 15, with the ratios to the true values shown in Fig. 16. For most of the mocks, the measured value is within 3 per cent for $f\sigma_8$ and 1 per cent for $\alpha_\parallel$, but there are some outliers, which come from the mocks with the largest changes in cosmology (cosmology 2 and 7, see Tables 4 and 5), and hence have the largest difference in the shape of the template (Fig. 13).

The mean, standard deviation, and rms from this set of blind mocks are shown in Table 7 for all models. There is a small offset in the best-fitting values from the RESPRESSO model: $f\sigma_8$ is systematically ∼2 per cent low, and values of $\alpha_\parallel$ are 0.7 per cent high, when compared to the true values for each of the cosmologies. This is because these mocks do not contain redshift smearing, while the parameter σ_{zerr}, which models the redshift smearing, is a free parameter in the RESPRESSO model. As was shown in the case of the

non-blind mocks with no smearing, this leads to a systematic offset in the measurements of these parameters.

The BAO fits are less affected by the assumption of fiducial cosmology. Compared to the full-shape fits, the scatter in the ratios of $\alpha_\parallel$ between the 24 mocks is smaller, leading to smaller values of the rms in Table 7. For $\alpha_\perp$, the rms is comparable to the full-shape fits.

The distributions of the measurements of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ from the blind mocks in Fig. 16 are shown in Fig. 17. The mean, standard deviation, and rms are indicated by the vertical lines (for $f\sigma_8$, the rms is divided by the average true value of $f\sigma_8$ from all the blind mocks). For models where the mean value is close to 1, the standard deviation and rms are very close. The value of the rms is primarily driven by the standard deviation but is slightly increased for models with the largest offsets.

6.3 Systematic errors

The aim of the mock challenge is to estimate the systematic error in the measured values of $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$ in the analysis of the eBOSS QSO clustering sample, using the RSD models of RESPRESSO and RegPT. This systematic error needs to encompass the effects of different HODs, redshift smearing, catastrophic redshifts, and an incorrect assumption for the fiducial cosmology. To calculate the total systematic error, we therefore combine the results of the set of 20 catastrophic redshift non-blind mocks, with the set of 24 blind mocks.

For the blind mocks, in order to estimate the systematic from different cosmologies, we calculate the rms of the 24 sets of mocks (defined in equation 34). However, the RESPRESSO model is affected by an additional systematic from having the parameter σ_{zerr} free when the mocks do not contain redshift smearing.[6] To remove this systematic offset, and to measure only the systematic due to the fiducial cosmology, we use the results from the non-blind mocks with no redshift smearing to apply a correction. For $f\sigma_8$, the corrected value is

$$f\sigma_8^{corr} = \left(\frac{f\sigma_8^{nb,true}}{f\sigma_8^{nb}}\right) f\sigma_8, \qquad (35)$$

where $f\sigma_8^{nb}$ is the average value of $f\sigma_8$ measured from the non-blind mocks (given in Table 6), and $f\sigma_8^{nb,true} = 0.382$ is the true value of $f\sigma_8$ in the non-blind mocks. A similar correction is made to $\alpha_\parallel$ and $\alpha_\perp$. The RegPT model is unaffected, by this, so no correction is needed. The results, after the correction, are given in the third column of Table 7. The values of rms are slightly larger for RESPRESSO than the RegPT model, but both are at a level of ∼0.01 for $f\sigma_8$ (which is ∼2.5 per cent for a value of $f\sigma_8 = 0.382$), ∼1 per cent for $\alpha_\parallel$ and ∼0.6 per cent for $\alpha_\perp$. Since the BAO fits are affected less by the fiducial cosmology, the rms values of $\alpha_\parallel$ and $\alpha_\perp$ are smaller than for the full-shape fits.

To measure the effect of different HODs, redshift smearing, and catastrophic redshifts, we calculate the rms from the 'catastrophic redshift' set of non-blind mocks, which include all these effects. These values are given in the final column of Table 6 for each model. Compared to the blind mocks, the rms values from the non-blind mocks are all smaller.

---

[6]To test the flexibility of the model, σ_{zerr} is left as a free parameter during the blind analysis. Afterwards, we compared the difference between varying and fixing σ_{zerr}, and we found systematic shifts of ∼1 per cent, ∼0.3 per cent, and ∼0.2 per cent in $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$, respectively.
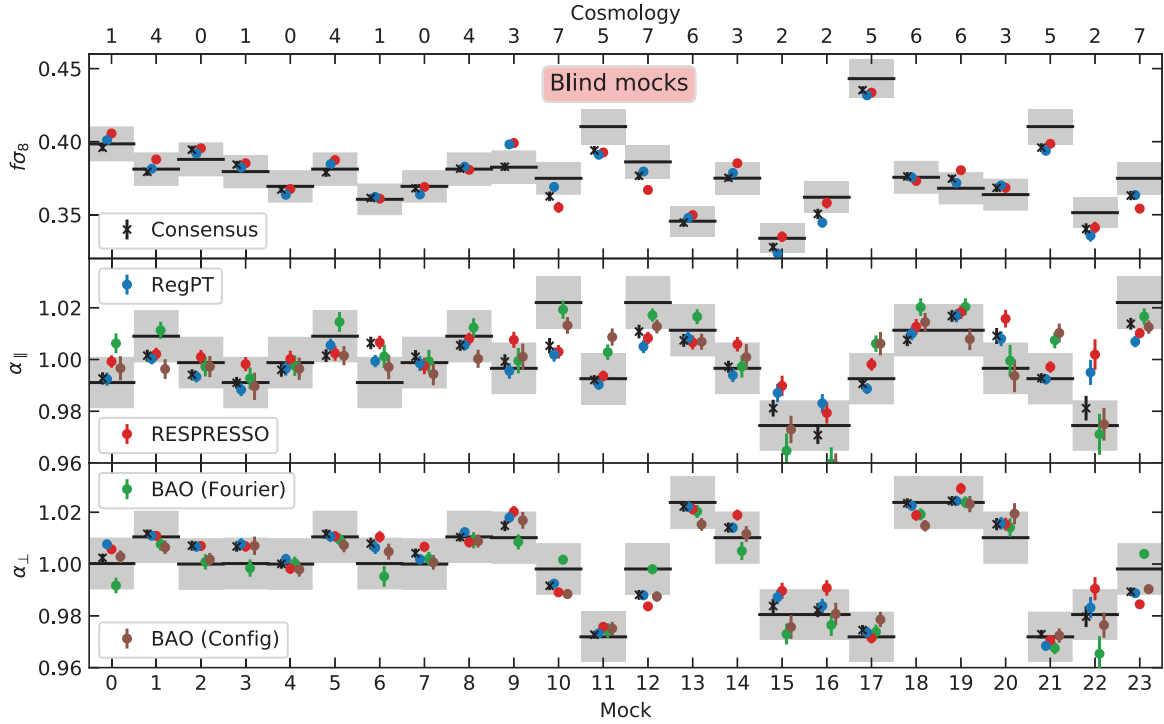
Figure 15. Results from the blind mock challenge, for the 24 sets of blind mocks in which the HOD was varied, and additional velocity scalings were also applied. Points show the average results from the RegPT model (blue), RESPRESSO model (red), combined consensus (black), and BAO fits in Fourier space (green), and configuration space (brown) of $f\sigma_8$, $\alpha_{\parallel}$, and $\alpha_{\perp}$. Black lines indicate the true values, with the shaded regions showing 3 per cent for $f\sigma_8$ and 1 per cent for the α parameters.
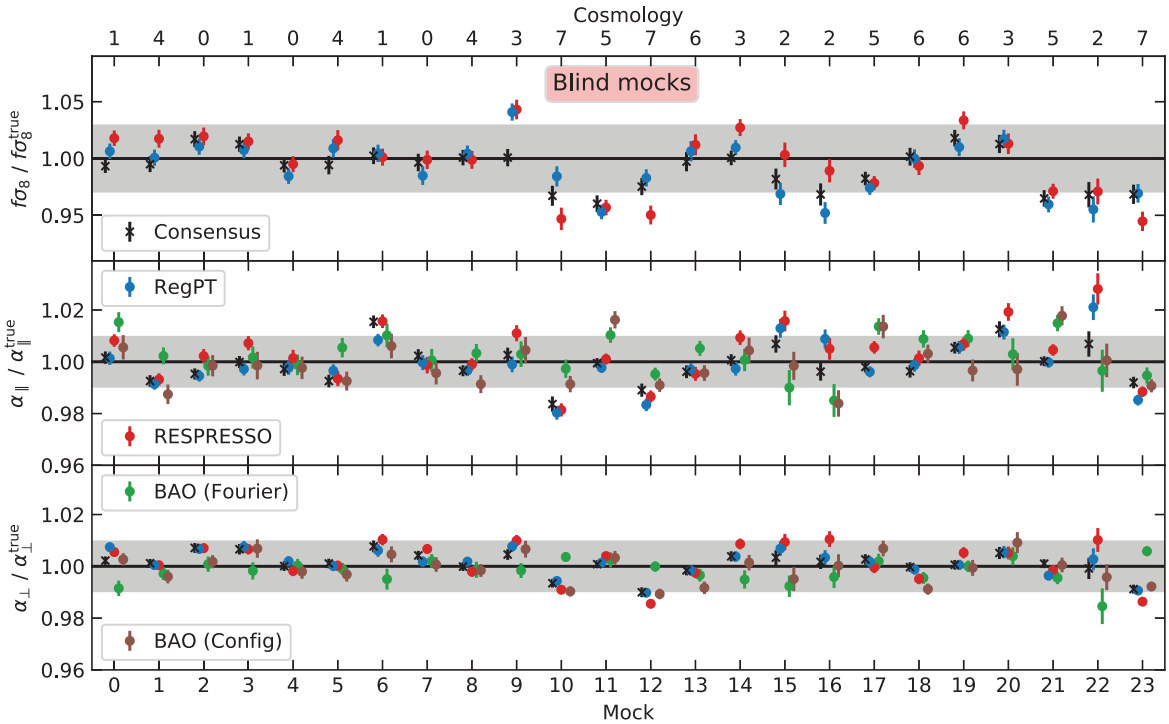
Figure 16. Same as Fig. 15, but showing the ratio to the true values.

The total error is calculated by adding the rms from the blind and non-blind mocks in quadrature. These are given in Table 8. These are conservative estimates for the total systematic errors. First, adding the errors in quadrature assumes that they are uncorrelated, which is not necessarily true, and results in a slight overestimate of the total error. Also, the blind mocks cover a wide range of different cosmologies. For the non-blind mocks, the scatter between the many HODs has the largest effect on the measured rms, while for the blind

Table 7. Results from the blind mock challenge for the RegPT and RESPRESSO models, and BAO fits in Fourier and configuration space, averaged over the 24 blind mocks. For each of the parameters $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$, we calculate the mean, standard deviation, and rms of the ratios to the true values. In the third column, the results of the non-blind mocks have been used to correct the systematic offsets in the RESPRESSO model due to having the parameter σ_{zerr} free.

	RegPT	RESPRESSO	RESPRESSO (corrected)	BAO (Fourier)	BAO (Config.)
$\langle f\sigma_8/f\sigma_8^{\text{true}}\rangle$	0.9904	0.9820	0.9957	–	–
$\sigma(f\sigma_8/f\sigma_8^{\text{true}})$	0.0230	0.0265	0.0269	–	–
rms($f\sigma_8$)	0.0093	0.0123	0.0105	–	–
$\langle \alpha_\parallel/\alpha_\parallel^{\text{true}}\rangle$	0.9992	1.0066	1.0035	1.0029	0.9992
$\sigma(\alpha_\parallel/\alpha_\parallel^{\text{true}})$	0.0089	0.0104	0.0103	0.0074	0.0083
rms($\alpha_\parallel$)	0.0091	0.0122	0.0109	0.0079	0.0085
$\langle \alpha_\perp/\alpha_\perp^{\text{true}}\rangle$	1.0017	1.0006	1.0022	0.9980	0.9995
$\sigma(\alpha_\perp/\alpha_\perp^{\text{true}})$	0.0048	0.0071	0.0071	0.0045	0.0057
rms($\alpha_\perp$)	0.0051	0.0072	0.0072	0.0049	0.0056
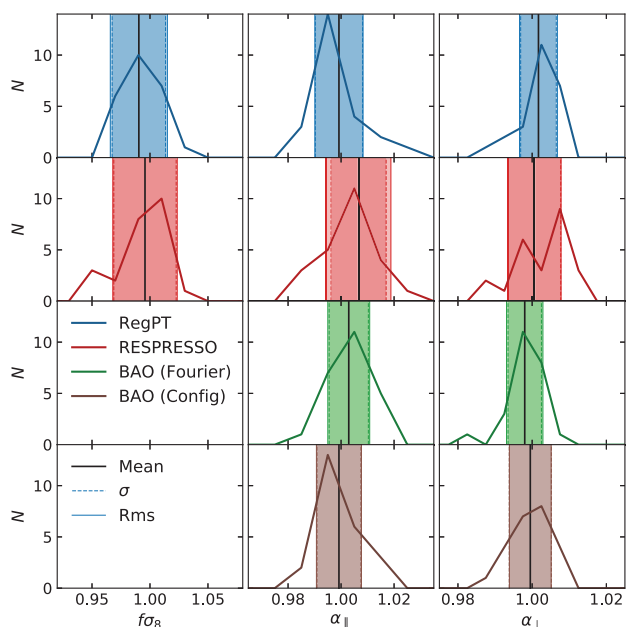
Figure 17. Distributions of the measured values of $f\sigma_8$ (left column), $\alpha_\parallel$ (middle column), and $\alpha_\perp$ (right column) from the 20 sets of blind mocks for the RegPT model (blue), RESPRESSO model (red), and the BAO fits in real space (green) and Fourier space (brown). The vertical lines indicate the mean (black solid line), σ (coloured dotted line), and rms (coloured solid line).

Table 8. Final systematic errors in $f\sigma_8$, $\alpha_\parallel$, and $\alpha_\perp$, from the RESPRESSO and RegPT models, combined consensus, and BAO fits.

	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$
RegPT	0.0123	0.0098	0.0066
RESPRESSO	0.0131	0.0117	0.0078
Consensus	0.0106	0.0079	0.0048
BAO (Fourier)	–	0.0098	0.0055
BAO (Configuration)	–	0.0102	0.0067

mocks, the main effect is from the different cosmologies. However, there is also an additional effect in the blind mocks from changing the HODs.

The RegPT and RESPRESSO models both perform well, with similar systematic errors. For both RSD models, we take conservative

systematic errors of $\sigma_{f\sigma_8} = 0.013$, $\sigma_{\alpha_\parallel} = 0.012$, and $\sigma_{\alpha_\perp} = 0.008$. These errors are small compared to the errors in the data, and are, at most, expected to be 30 per cent of the statistical error. The BAO fits in configuration and Fourier space also both perform well, and we take conservative systematic errors of $\sigma_{\alpha_\parallel} = 0.010$ and $\sigma_{\alpha_\perp} = 0.007$.

In addition to calculating a systematic error for the individual models, the results using the RegPT and RESPRESSO models are combined into a single consensus result. The results from the two models are combined using the method of Sánchez et al. (2017a). Despite the results of the two analyses being highly correlated, by combining the results, further constraints are placed on the allowed parameter space, resulting in the combined results having a smaller uncertainty than each individual analysis. The systematic error in the consensus result is shown in Table 8. The systematic errors of the combined consensus result are reduced to $\sigma_{f\sigma_8} = 0.011$, $\sigma_{\alpha_\parallel} = 0.008$, and $\sigma_{\alpha_\perp} = 0.005$. These systematic errors are included as part of the quoted errors in the final eBOSS DR16 consensus results.

7 CONCLUSIONS

The eBOSS quasar sample, in the redshift range of $0.8 < z < 2.2$, can be used as a direct tracer of the matter density field, enabling measurements to be taken of the growth rate, Hubble distance, and transverse comoving distance at this redshift. These measurements can be used to place constraints on theories of dark energy and models of modified gravity. The aim of the quasar mock challenge is to validate the RSD and BAO models used in the analysis of Hou et al. (2020) and Neveux et al. (2020), in configuration and Fourier space, and estimate the systematic uncertainties.

Mock catalogues are generated using the OuterRim N-body simulation, which has a box size of $3\ h^{-1}$Gpc and was run in a WMAP7 cosmology. The box is populated with quasars using 20 different HODs, which are tuned to give approximately the same clustering measurements as the eBOSS DR16 quasar sample and the same number density. We populate the snapshot at $z = 1.433$, which is closest to the effective redshift of the data, $z_{\text{eff}} = 1.48$.

For the non-blind part of the mock challenge, 100 individual mocks are generated from the OuterRim box for each of the 20 HODs, which are analysed using the OuterRim cosmology as the fiducial cosmology. For each mock, we create a version with no redshift smearing, with Gaussian redshift smearing, and a more realistic double Gaussian redshift smearing. We also create a set of mocks that

contains realistic smearing and 1.5 per cent catastrophic redshifts, which are generated from a uniform redshift distribution.

For the non-blind mocks with no smearing, the RegPT, RESPRESSO, and CLPT models all perform equally well, and there is no reason to choose any particular model over any other. In the Fourier-space analysis of Neveux et al. (2020), the RegPT model is used, while RESPRESSO is used for the configuration-space analysis of Hou et al. (2020). These models are able to measure the parameters $f\sigma_8$ within 3 per cent, and $\alpha_{\parallel}$ and $\alpha_{\perp}$ to within 1 per cent, with and without redshift smearing. The inclusion of catastrophic redshifts dilutes the clustering signal, leading to the measured $f\sigma_8$ being biased towards smaller values, but this is still within the limit of $\sim$3 per cent.

Blind mocks are generated from the OuterRim simulation, using the method of Mead & Peacock (2014a) to rescale the halo positions, velocities, and masses, in order to mimic a simulation that was run in a different target cosmology. We rescale the OuterRim simulation to eight new cosmologies, and generate eight sets of 100 mocks, using the same HOD for each. The CLPT model is used to validate the rescaled cosmologies by recovering the expected values of $f\sigma_8$, $\alpha_{\parallel}$, and $\alpha_{\perp}$. When the OuterRim cosmology is used as the fiducial cosmology, the scatter in the measured values of the parameters is larger for all models, due to the incorrect assumption of the fiducial cosmology. While the BAO peak position is scaled by the expected α parameters, the incorrect assumption of the fiducial cosmology distorts the clustering on small scales, leading to clustering measurements that do not match the shape of the template. This results in a biased measurement of $f\sigma_8$, $\alpha_{\parallel}$, and $\alpha_{\perp}$, which is correlated with Ω_m. We also generate a set of 24 blind mocks, using the same rescaled OuterRim snapshots, but with the HOD varied, and an additional scaling applied to velocities.

To calculate the systematic error, we combine the results from the 24 blind mocks, with the 20 non-blind mocks with catastrophic redshifts, to take into account the effects of the incorrect assumption of the fiducial cosmology, different HODs, redshift smearing, and catastrophic redshifts. We calculate the rms of the blind and non-blind mocks, which are added in quadrature to estimate the total error. We use the non-blind mocks with no smearing to correct the results from the RESPRESSO model for having the parameter σ_{zerr} free, which biases the results when the mocks do not contain redshift smearing.

The RegPT and RESPRESSO models perform equally well, and we take systematic errors for both models of $\sigma_{f\sigma_8} = 0.013$, $\sigma_{\alpha_{\parallel}} = 0.012$, and $\sigma_{\alpha_{\perp}} = 0.008$. For the BAO fits, we take systematic errors of $\sigma_{\alpha_{\parallel}} = 0.010$ and $\sigma_{\alpha_{\perp}} = 0.007$. These results are a conservative upper bound, since they assume that the errors in the blind and non-blind mocks are uncorrelated, and that the blind mocks cover a wide range of cosmological parameters. The results of the RegPT and RESPRESSO models are combined into a single consensus result, which tightens the constraints on the parameters, compared to the individual analyses. The consensus systematic errors are $\sigma_{f\sigma_8} = 0.011$, $\sigma_{\alpha_{\parallel}} = 0.08$, and $\sigma_{\alpha_{\perp}} = 0.005$. These errors are combined with the errors in the final consensus results from the analysis of the eBOSS DR16 quasar sample. The tests that we have performed in this work show that the measurement of the BAO scale is very robust, despite the effects of redshift uncertainty and the wide range of HOD models we have used. We can therefore be confident that using such methods will enable accurate cosmological distance measurements in future surveys.

The mock challenge presented in this paper is similar to what has been done previously in the BOSS survey. Future surveys, such as DESI, LSST, and Euclid, will cover much larger volumes and aim to very precise, sub-per cent-level cosmological measurements. In order to achieve this precision, it will be essential that systematics in the models can be reduced as much as possible. Using mock catalogues based on a single *N*-body simulation, as was done in this work, will not be good enough, and assessing the systematics will require large numbers of *N*-body simulations, with different cosmological parameters, both within ΛCDM and also extending to models beyond ΛCDM. For the DESI survey, many of the simulations and mock catalogues required have already been produced, and work on the mock challenge is currently under way.

DATA AVAILABILITY

The non-blind mock catalogues underlying this article will be shared on reasonable request to the corresponding author. The blind mocks

will be shared on reasonable request to the corresponding author with permission of Katrin Heitmann and Salman Habib. Correlation function and power spectrum measurements can be found at https://svn.sdss.org/public/data/eboss/mcmc/trunk/.

REFERENCES

Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Alam S. et al., 2017, MNRAS, 470, 2617
Alam S., Peacock J. A., Kraljic K., Ross A. J., Comparat J., 2020a, MNRAS, 497, 581
Alam S. et al., 2020b, preprint (arXiv:2007.09004)
Alcock C., Paczynski B., 1979, Nature, 281, 358
Alonso D., 2012, preprint (arXiv:1210.1833)
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Ata M. et al., 2018, MNRAS, 473, 4773
Avila S. et al., 2020, MNRAS, in press
Bautista J. E. et al., 2018, ApJ, 863, 110
Bautista J. E. et al., 2020, MNRAS, in press
Bel J., Pezzotta A., Carbone C., Sefusatti E., Guzzo L., 2019, A&A, 622, A109
Beutler F. et al., 2012, MNRAS, 423, 3430
Beutler F. et al., 2014, MNRAS, 443, 1065
Beutler F. et al., 2017, MNRAS, 466, 2242
Blake C. et al., 2011a, MNRAS, 415, 2876
Blake C. et al., 2011b, MNRAS, 418, 1707
Blanton M. R. et al., 2017, AJ, 154, 28
Bongiorno A. et al., 2016, A&A, 588, A78
Carlson J., Reid B., White M., 2013, MNRAS, 429, 1674
Carter P., Beutler F., Percival W. J., DeRose J., Wechsler R. H., Zhao C., 2020, MNRAS, 494, 2076
Chan K. C., Scoccimarro R., Sheth R. K., 2012, Phys. Rev. D, 85, 083509
Cole S. et al., 2005, MNRAS, 362, 505
Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 063519
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
de Mattia A. et al., 2020, preprint (arXiv:2007.09008)
du Mas des Bourboux H. et al., 2020, preprint (arXiv:2007.08995)
DESI Collaboration, 2016a, preprint (arXiv:1611.00036)
DESI Collaboration, 2016b, preprint (arXiv:1611.00037)
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
eBOSS Collaboration, 2020, preprint (arXiv:2007.08991)
Eftekharzadeh S., Myers A. D., Kourkchi E., 2019, MNRAS, 486, 274
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J., Seo H.-J., White M., 2007, ApJ, 664, 660
Garrison L. H., Eisenstein D. J., 2019, MNRAS, 485, 2407
Georgakakis A., Aird J., Schulze A., Dwelly T., Salvato M., Nandra K., Merloni A., Schneider D. P., 2017, MNRAS, 471, 1976
Georgakakis A., Comparat J., Merloni A., Ciesla L., Aird J., Finoguenov A., 2019, MNRAS, 487, 275
Gil-Marín H. et al., 2018, MNRAS, 477, 1604
Gil-Marín H. et al., 2020, MNRAS, 498, 2492
Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, MNRAS, 457, 1577
Grieb J. N. et al., 2017, MNRAS, 467, 2085
Gunn J. E. et al., 2006, AJ, 131, 2332
Guzzo L. et al., 2008, Nature, 451, 541
Habib S. et al., 2016, New A, 42, 49
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Heitmann K. et al., 2019a, ApJS, 245, 16
Heitmann K. et al., 2019b, ApJS, 244, 17
Hou J. et al., 2018, MNRAS, 480, 2521

Hou J. et al., 2020, preprint (arXiv:2007.08998)
Icaza-Lizaola M. et al., 2019, MNRAS, 492, 4189
Ivezić, Ž. et al., 2019, ApJ, 873, 111
Jackson J. C., 1972, MNRAS, 156, 1P
James F., Roos M., 1975, Comput. Phys. Commun., 10, 343
Kaiser N., 1987, MNRAS, 227, 1
Kayo I., Oguri M., 2012, MNRAS, 424, 1363
Kirkby D. et al., 2013, J. Cosmology Astropart. Phys., 2013, 024
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Komatsu E. et al., 2011, ApJS, 192, 18
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Leauthaud A. et al., 2015, MNRAS, 446, 1874
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Lin S. et al., 2020, MNRAS, 498, 5251
Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, MNRAS, 441, 378
Lyke B. W. et al., 2020, ApJS, 250, 8
MacCrann N. et al., 2018, MNRAS, 480, 4614
Matsubara T., 2008, Phys. Rev. D, 77, 063530
Mead A. J., Peacock J. A., 2014a, MNRAS, 440, 1233
Mead A. J., Peacock J. A., 2014b, MNRAS, 445, 3453
Mohammad F. G. et al., MNRAS, 2020, 498, 128
Myers A. D. et al., 2015, ApJS, 221, 27
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neveux R. et al., 2020, MNRAS, preprint (arXiv:2007.08999)
Nishimichi T., Bernardeau F., Taruya A., 2017, Phys. Rev. D, 96, 123515
Oogi T., Shirakata H., Nagashima M., Nishimichi T., Kawaguchi T., Okamoto T., Ishiyama T., Enoki M., 2020, MNRAS, 497, 1
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration VI, 2020, A&A, 641, A6
Powell M. C. et al., 2018, ApJ, 858, 110
Prakash A. et al., 2016, ApJS, 224, 34
Press W. H., Schechter P., 1974, ApJ, 187, 425
Raichoor A. et al., 2017, MNRAS, 471, 3955
Rodríguez-Torres S. A. et al., 2017, MNRAS, 468, 728
Ross A. J. et al., 2017, MNRAS, 464, 1168
Ross A. J. et al., 2020, MNRAS, 498, 2354

/mnras/staa2416
Rossi G. et al., 2020, preprint (arXiv:2007.09002)
Sánchez A. G. et al., 2017a, MNRAS, 464, 1493
Sánchez A. G. et al., 2017b, MNRAS, 464, 1640
Sargent W. L. W., Turner E. L., 1977, ApJ, 212, L3
Satpathy S. et al., 2017, MNRAS, 469, 1369
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Smee S. A. et al., 2013, AJ, 146, 32
Smith A., de Mattia A., Burtin E., Chuang C.-H., Zhao C., 2020, preprint (arXiv:2007.11417)
Spergel D. N. et al., 2003, ApJS, 148, 175
Tamone A. et al., 2020, preprint (arXiv:2007.09009)
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. L. et al., 2012, ApJ, 745, 16
Vargas-Magaña M. et al., 2018, MNRAS, 477, 1153
Wilson M. J., Peacock J. A., Taylor A. N., de la Torre S., 2017, MNRAS, 464, 3121
Yamamoto K., Nakamichi M., Kamino A., Bassett B. A., Nishioka H., 2006, PASJ, 58, 93
York D. G. et al., 2000, AJ, 120, 1579
Zarrouk P. et al., 2018, MNRAS, 477, 1639
Zhao C. et al., 2020, preprint (arXiv:2007.08997)

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