MNRAS **468,** 728–740 (2017)
Advance Access publication 2017 February 24
doi:10.1093/mnras/stx454

Clustering of quasars in the first year of the SDSS-IV eBOSS survey: interpretation and halo occupation distribution

Sergio A. Rodríguez-Torres,[1,2,3]★† Johan Comparat,[1,3]‡ Francisco Prada,[1,2,4] Gustavo Yepes,[3] Etienne Burtin,[5] Pauline Zarrouk,[5] Pierre Laurent,[5] ChangHoon Hahn,[6] Peter Behroozi,[7] Anatoly Klypin,[8,9] Ashley Ross,[10,11] Rita Tojeiro[12] and Gong-Bo Zhao[11,13]

[1]*Instituto de Física Teórica, (UAM/CSIC), Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[2]*Campus of International Excellence UAM+CSIC, Cantoblanco, E-28049 Madrid, Spain*
[3]*Departamento de Física Teórica M8, Universidad Autonóma de Madrid (UAM), Cantoblanco, E-28049, Madrid, Spain*
[4]*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain*
[5]*CEA, Centre de Saclay, IRFU/SPP, F-91191 Gif-sur-Yvette, France*
[6]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA*
[7]*Theoretical Astrophysics Center, Astronomy and Physics Departments, University of California, Berkeley, Berkeley CA 94720, USA*
[8]*Astronomy Department, New Mexico State University, Las Cruces, NM-88003, USA*
[9]*Severo Ochoa Associate Researcher at the Instituto de Física Teórica (UAM/CSIC), Cantoblanco, E-28049, Madrid, Spain*
[10]*Center for Cosmology and AstroParticle Physics, The Ohio State University, Columbus, OH 43210, USA*
[11]*Institute of Cosmology and Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK*
[12]*School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*
[13]*National Astronomy Observatories, Chinese Academy of Science, Beijing 100012, P.R. China*

Accepted 2017 February 20. Received 2017 February 17; in original form 2016 September 17

ABSTRACT

In current and future surveys, quasars play a key role. The new data will extend our knowledge of the Universe as it will be used to better constrain the cosmological model at redshift $z > 1$ via baryon acoustic oscillation and redshift space distortion measurements. Here, we present the first clustering study of quasars observed by the extended Baryon Oscillation Spectroscopic Survey. We measure the clustering of ∼70 000 quasars located in the redshift range $0.9 < z < 2.2$ that cover 1168 deg^2. We model the clustering and produce high-fidelity quasar mock catalogues based on the BigMultiDark Planck simulation. Thus, we use a modified (sub)halo abundance matching model to account for the specificities of the halo population hosting quasars. We find that quasars are hosted by haloes with masses $\sim 10^{12.7}$ M$_\odot$ and their bias evolves from 1.54 ($z = 1.06$) to 3.15 ($z = 1.98$). Using the current extended Baryon Oscillation Spectroscopic Survey data, we cannot distinguish between models with different fractions of satellites. The high-fidelity mock light-cones, including properties of haloes hosting quasars, are made publicly available.

Key words: quasars: general – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

How quasars (QSO) populate the large-scale structure is a puzzle in modern cosmology. It is known that these objects trace the dark matter density field. Therefore, using measurements of the baryon acoustic oscillations (BAO) or redshift space distortions (RSD) from quasars, one can infer information of the cosmological model. However, for these studies or to increase the knowledge of the evolution of quasars, we require a good estimation of their distribution at all scales. Thus, spectroscopic surveys and high-fidelity galaxy mocks from simulations are a great help when solving many riddles concerning quasars.

Large galaxy spectroscopic surveys are an excellent tool to construct a precise 3D map of our Universe. They allow us to study the distribution of different populations in the Universe and constrain cosmological information via BAO scale or RSD measurements. The Sloan Digital Sky Survey (SDSS; York et al. 2000) and the two degree field galaxy redshift survey (Norberg et al. 2001) first measured the BAO scale in the local universe (Cole et al. 2005;

★ E-mail: sergio.rodriguez@uam.es
† Campus de Excelencia Internacional UAM/CSIC Scholar.
‡ Severo Ochoa Fellow.

Eisenstein et al. 2005). The Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), included in the SDSS III program (Eisenstein et al. 2011), recently provided accurate redshifts for 1.5 million galaxies as faint as $i = 19.1$, that cover the redshift range $0.2 < z < 0.75$ on 10 000 deg^2. In combination with SDSS-I/II (York et al. 2000), it provided a subpercent level measurement of the position of the BAO peak at redshift $z = 0.57$ (Alam et al. 2016). SDSS is an example of how spectroscopic surveys can provide strong constraints on our knowledge of the Universe.

Bright quasars constitute the best targets to sample the matter field at high redshift with a small exposure time. Indeed, quasars bear an active galactic nucleus (AGN) that generates light which outshines the entire host galaxy. SDSS I/II published a sample of ∼100 000 confirmed quasars (Schneider et al. 2010) and SDSS-III observed ∼170 000 quasars with redshift $2.1 < z < 3.5$ as faint as $g = 22$ (Pâris et al. 2014). Using both samples, the BAO feature was measured to a few per cent in the Lyman α (Ly α) forest (Font-Ribera et al. 2014; Delubac et al. 2015). Despite the large sample of quasars observed by the SDSS programs, there is still a large region in redshift ($1 < z < 2.1$) that ought to be studied by targeting quasars fainter than $i = 19.1$ in the SDSS imaging. Recent data from other experiments (Wright et al. 2010, e.g. *WISE*) provides additional information to best target quasars. A cutting-edge target selection algorithm was implemented in Myers et al. (2015) and is being observed by the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016), part of the SDSS-IV program. It will increase the number of quasars found by SDSS I/II in the redshift range $0.9 < z < 2.2$ by a factor of 5. This new sample will cover ∼7500 deg^2, increasing both the volume and the low number density of the previous samples. It is designed to measure the BAO scale with quasars as tracers of the matter field. In this study, we consider the eBOSS First Year QSO data (hereafter Y1Q). For more details, please see Section 2.1.

Different models have been used to analyse the clustering of quasars. In the literature, many studies focus on the linear regime (large scales). At these scales, correlation function can be described by a power law (e.g. Chehade et al. 2016), mostly due to the intrinsic low density of quasars. A more sophisticated method used to model the galaxy clustering and generate mock catalogues is the halo occupation distribution (HOD; Jing, Mo & Börner 1998; Peacock & Smith 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002; Zheng et al. 2005). The HOD model recovers the quasar clustering, but its parameters are largely degenerate, producing poor constraints on the host halo masses and satellite fraction (Richardson et al. 2012; Shen et al. 2013). Galaxy samples have also been studied with another method, namely halo abundance matching (HAM), which reproduces the clustering of complete galaxy samples with a reasonable agreement (e.g. Kravtsov et al. 2004; Conroy, Wechsler & Kravtsov 2006; Behroozi, Conroy & Wechsler 2010; Guo et al. 2010; Trujillo-Gomez et al. 2011; Nuza et al. 2013; Reddick et al. 2013). By including the stellar mass distribution (or luminosity distribution), the HAM also accounts for incomplete samples (e.g. Rodríguez-Torres et al. 2016). HAM requires knowledge of the stellar mass function, the scatter in the stellar mass to halo mass relation and the incompleteness of the sample. In the case of quasars, obtaining such information is not an easy task. However, modifications of the standard method can be implemented to describe the quasar population.

In this study, we generate light-cones based on the BigMultiDark Planck simulation (BigMDPL; Klypin et al. 2016), using a modified HAM technique to reproduce the Y1Q clustering properties. The BigMDPL is an N-body simulation with box size 2.5 h^{-1} Gpc and 3840^3 particles, which yields a volume large enough to encompass Y1Q. A variety of mocks, which model different populations of galaxies, has already been constructed using the BigMDPL simulation. They predict, with a good agreement, the observed two-point and three-point statistics (Guo et al. 2015; Favole et al. 2016; Rodríguez-Torres et al. 2016).

This paper is structured as follows. In Section 2, we describe the data used in our analysis. Section 3 presents the different steps to construct the BigMDPL eBOSS quasar mocks, including how we populate dark matter haloes using a modified HAM algorithm. A set of predictions from our model is shown in Section 4. Subsequently, we discuss and summarize the most relevant results in Sections 5 and 6. In this paper, we assume a fiducial Λ cold dark matter (ΛCDM) cosmology with the PLANCK-I parameters $\Omega_{\rm m} = 0.307$, $\Omega_B = 0.048$, $\Omega_\Lambda = 0.693$ (Planck Collaboration XVI 2014).

2 DATA

2.1 eBOSS QSO survey and clustering

The eBOSS (Dawson et al. 2016) is part of a six year SDSS-IV programme (fall 2014 to spring 2020). It combines the potential of SDSS-III/BOSS and new photometric information to optimize target selection and extend BAO studies to higher redshift. eBOSS uses the 2.5-m Sloan Foundation Telescope at Apache Point Observatory (Gunn et al. 2006) and the same fibre-fed optical spectrograph as BOSS, where each fibre subtends a 2 arcsec diameter of the sky (Smee et al. 2013). This survey will provide redshifts for 300 000 luminous red galaxies (LRG) in the redshift range $0.6 < z < 1.0$, a new sample of ∼200 000 emission line galaxies (ELG) at redshift $z > 0.6$, more than 500 000 spectroscopically confirmed quasars at $0.9 < z < 2.2$ and ∼120 000 new Ly α forest quasars at redshift $z > 2.1$.

eBOSS dedicates 1800 plates to cover an area of 9000 deg^2: 1500 plates to measure LRG and QSO redshifts on 7500 deg^2 and 300 plates to measure ELG redshifts on 1000 deg^2. The first two years, observations were dedicated to the QSO and LRG samples. In order to maximize the tiling completeness and fibre efficiency in the LRG/QSO sample, a tiered priority is adopted (Dawson et al. 2016), where the QSO targets have maximal priority and are assigned to fibres first.

eBOSS has adopted two approaches to target quasars for redshift >0.9 (Myers et al. 2015). In the first approach, 'Clustering' quasar targets (QSO_CORE) are used as a direct tracer of the large-scale structure in the redshift range $0.9 < z < 2.2$. The second approach consists in detecting quasars at $z > 2.1$ to map the large-scale structure via absorption of the Ly α forest (Palanque-Delabrouille et al. 2016).

(i) The CORE quasar sample is constructed combining optical selection in *ugriz* using a likelihood-based routine called XDQSOz (Bovy et al. 2011), with a mid-IR–optical colour cut. eBOSS CORE selection (to $g < 22$ or $r < 22$) should obtains ∼70 quasars deg^{-2} at redshifts $0.9 < z < 2.2$ and about 7 quasars deg^{-2} at $z > 2.2$.

(ii) The Ly α quasar selection is based on variability in multi-epoch imaging from the Palomar Transient Factory (Palanque-Delabrouille et al. 2016). It recovers an additional 3 or 4 quasars deg^{-2} at $z > 2.2$ to $g < 22.5$. A linear model of how imaging systematics affect target density recovers the angular distribution of eBOSS CORE quasars over 96.7 per cent (76.7 per cent) of the SDSS North (South) Galactic Cap area (Myers et al. 2015).

Table 1. Distribution of the Y1Q sample in four redshift bins. $\bar{n}$ represents the comoving number density of QSO, N is the number of QSO and V is the comoving volume of the redshift bin subtended by 1168 deg^2. The last line shows the values for the complete sample.

Redshift	$\bar{n}$ (10^{-5} Mpc^{-3} h^3)	N	V (10^9 h^{-3} Mpc^3)
$0.9 < z < 1.2$	1.36	13 484	0.99
$1.2 < z < 1.5$	1.48	17 578	1.19
$1.5 < z < 1.8$	1.36	17 778	1.31
$1.8 < z < 2.2$	1.05	19 429	1.84
$0.9 < z < 2.2$	1.28	68 269	5.34

Busca et al. (2013) measure the BAO scale using Ly α quasars from the BOSS data. Font-Ribera et al. (2014) also give measurements of this scale using the cross-correlation between visually confirmed quasars with the Ly α forest absorption. One of the goals of eBOSS is to provide a first detection of the BAO scale using only the CORE quasar sample.

In this context, we focus our study on the spectroscopically confirmed QSO using the Y1Q data which includes 68 269 objects that cover 1168 deg^2 of the sky. Table 1 shows the abundance of CORE QSO at different redshift ranges.

2.2 Redshift error and statistical weights

eBOSS expects a redshift precision better than 300 s^{-1}km rms for the QSO CORE at $z < 1.5$ and better than [300+400(z −1.5)] km s^{-1} at $z > 1.5$ (Myers et al. 2015). It corresponds to redshift errors of the order of 1×10^{-3} for $z < 1.5$ and $\sim 5 \times 10^{-3}$ for larger redshift. These errors have an important impact on scales smaller than 10 h^{-1}Mpc (see Appendix B). For this reason, we add redshift errors to the mock catalogues using these upper limits. In addition, less than 1 per cent of the sample is expected to have catastrophic redshift errors.

In order to include the observed redshift precision in the light-cones, we model redshift errors using a Gaussian distribution with mean value $z_{\rm true}$ and width Δz,

$$z = z_{\rm true} + \Delta z \mathcal{N}(0, 1), \tag{1}$$

where $\mathcal{N}(0, 1)$ is a random number coming from a Gaussian distribution with mean 0 and standard deviation 1 and

$$\Delta z = \begin{cases} 300 \text{ km s}^{-1}c^{-1} & \text{if } z<1.5 \\ [300 + 400(z - 1.5)] \text{ km s}^{-1}c^{-1} & \text{if } z \geq 1.5, \end{cases} \tag{2}$$

c represents the speed of light. We also include 1 per cent of catastrophic redshift errors, which introduces a reduction in the amplitude of the correlation function of $\sim$1 per cent at all scales (Appendix B). In order to include these errors, we randomly select 1 per cent of the mock galaxies and replace their redshift by a random value within the range of the catalogue.

A correct estimation of redshift errors is important in order to understand the behaviour of the clustering at small scales. The monopole of the correlation function is affected by over 50 per cent at scales below 10 h^{-1} Mpc. The impact is larger on the quadrupole, where the effects are detected at scales below 40 h^{-1} Mpc (Reid & White 2011). In Appendix B, we explore with more detail the impact of these errors on clustering measurements. Nevertheless, even if we model the redshift errors, this is still an approximation that can introduce unphysical effects. This can result in a wrong estimation of the model's parameters if scales affected by errors are included in the fitting procedure. For this reason, we fix the parameters using the monopole of the correlation function between 10 and 40 h^{-1} Mpc, where the impact of redshift errors decreases and the effects of the cosmic variance and shot noise become smaller (Appendix B).

In addition to redshift measurement, the 5σ detection limit for point sources (also called depth) of the SDSS photometric survey varies across the footprint and differs for each band. The amplitude of the variations implies that faint targets end up very close to the detection limit. These targets are then more likely to be missed by the target selection algorithm. eBOSS corrects this effect by applying a depth-dependent weight, called 'systematics weight' $w_{\rm sys}$ to each quasar (see Laurent et al., in preparation for a detailed description).

Finally, eBOSS takes fibre collisions and redshift failures into account by using weights for each, w_{cp} and w_{zf}, respectively. Those quantities are initialized to one for all objects. Then, if a quasar has a nearest neighbour with a redshift failure or its redshift was not obtained because it was in a close pair, w_{zf} or w_{cp} are increased by one (Ross et al. 2012). Including all these effects, the total weight for each quasar in the observed data is given by

$$w_{\rm Q} = w_{\rm FKP} w_{\rm sys}(w_{cp} + w_{zf} - 1), \tag{3}$$

where $w_{\rm FKP}$ is the density weight applied for an optimal estimation of the two-point function and is defined by the expression (Feldman, Kaiser & Peacock 1994)

$$w_{\rm FKP} = \frac{1}{1 + n(z)P_{\rm FKP}}, \tag{4}$$

where $n(z)$ is the number density at redshift z and $P_{\rm FKP} = 6000\ h^{-3}$ Mpc^3.

Corrections for fibre collisions using close pair weights do not provide an accurate clustering signal at small scales (Guo, Zehavi & Zheng 2012; Hahn et al. 2017). However, in the quasar sample the distribution of objects is disperse and the number of collided pairs is very small. Additionally, our analysis does not use scales below 10 h^{-1}Mpc, so the close pair correction is good enough for our purpose. In the case of the simulated quasars, we include FKP weights but do not simulate the effects that require any of the additional weights applied to the data sample.

2.3 The eBOSS BigMultiDark light-cone

The suite of MultiDark[1] Planck (MDPL) simulations adopts a flat ΛCDM model with PLANCK-I cosmological parameters (Planck Collaboration XVI 2014): $\Omega_{\rm m} = 0.307$, $\Omega_B = 0.048$, $\Omega_\Lambda = 0.693$, $\sigma_8 = 0.829$, $n_s = 0.96$ and a dimensionless Hubble parameter $h = 0.678$. We only use two of the N-body simulations described in Klypin et al. (2016). The BigMultiDark (BigMDPL) has a box length of 2.5 h^{-1} Gpc with 3840^3 particles of mass 2.4 $\times$ 10^{10} h^{-1}M$_\odot$ and the MDPL has a box length of 1.0 h^{-1} Gpc with 3840^3 particles with a mass of 1.5×10^9 h^{-1} M$_\odot$. Both were built with GADGET-2 (Springel 2005) using initial Gaussian fluctuations generated with the Zel'dovich approximation at redshift 100.

From the dark matter catalogues of the simulation, haloes are defined with the Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement halo finder (ROCKSTAR; Behroozi, Wechsler & Wu 2013). Spherical dark matter haloes and subhaloes are identified using an approach based on adaptive hierarchical refinement of friends-of-friends groups in six-phase space dimensions

[1] http://www.multidark.org/

Table 2. Deviation from the mass function at redshift 0 for the MDPL and the BigMDPL simulations. The masses and the maximum circular velocities are the threshold above which the completeness in this box relative to the mass function is higher than the percentage given in the header (see equation 6). The corresponding number of particles is provided in brackets.

	$\log(M_{200c}(z)/M_\odot)$				V_{max}			
Fraction	80 per cent	90 per cent	95 per cent	97 per cent	80 per cent	90 per cent	95 per cent	97 per cent
				Central haloes				
MDPL	11.04 (71)	11.10 (82)	11.26 (119)	11.61 (266)	57.3	68.6	98.3	121.9
BigMD	12.22 (69)	12.28 (79)	12.32 (87)	12.36 (98)	131.0	145.9	201.6	299.3

and one time dimension. ROCKSTAR computes halo mass using spherical overdensities of a virial structure (Bryan & Norman 1998). Before calculating halo masses and circular velocities, the halo finder performs a procedure that removes unbound particles from the final mass of the halo.[2] We include observational effects and construct a catalogue with similar volume to the eBOSS sample, by making light-cones based on different snapshots of the BigMDPL simulation.

We perform the modified HAM by using the maximum circular velocity of the halo (V_{max}) in order to link dark matter haloes and quasars. The maximum circular velocity is one of the best candidates for matching dark matter haloes and galaxies (Reddick et al. 2013). V_{max} can be related to the virial mass of the halo through a power law given by

$$V_{max} = \beta(z)[M_{vir}E(z)/(10^{12}h^{-1}M_\odot)]^{\alpha(z)} \quad (5)$$

where, $E(z) = \sqrt{\Omega_{\Lambda,0} + \Omega_{m,0}(1+z)^3}$, $\log_{10}\beta(z) = 2.209 + 0.060a - 0.021a^2$ and $\alpha(z) = 0.346 - 0.059a + 0.025a^2$, with $a = 1/(1+z)$ the scalefactor (see Rodriguez-Puebla et al. 2016). There are better candidates to perform the matching between dark matter haloes and galaxies, such as, the maximum circular velocity along the whole history of the halo (V_{peak}). However, the BigMDPL simulation has a small number of snapshots (4) in the quasar redshift range thus preventing a good estimation of quantities that are computed by tracing haloes between snapshots. For this reason, we use V_{max} to implement our model. Differences between V_{peak} and V_{max} become important in case of substructures, while the selection of host haloes is similar with both quantities. Reddick et al. (2013) show a significantly larger amount of subhaloes when V_{peak} is used rather than other quantities. However, in our model the impact of choosing V_{max} can be compensated by using the fraction of satellites as a free parameter. Furthermore, the poor information of the one halo term in the quasar sample and the large errors in observations will not allow us to distinguish which quantity performs the matching better.

Table 2 presents the deviation of each simulation from a model of the complete mass function (Comparat et al. 2017), which is obtained by fitting a data set that contains the complete part of each of the MultiDark Planck simulation (SMDPL, MDPL, BigMDPL, HMDPL). Masses in Table 2 fulfil the condition given by

$$N_{sim}(M_{200} > M_i)/N_{mod}(M_{200} > M_i) < \text{percentage}, \quad (6)$$

where N_{sim} is the number of objects in the simulation with M_{200} smaller than the threshold mass M_i and N_{mod} is the corresponding number of haloes in the model. Previous works showed that quasars live in haloes with masses of the order of $\log(M/M_\odot) \sim 12.5$ (Shen et al. 2013; Chehade et al. 2016). Both simulations mentioned above are complete for this mass as is shown in Table 2. But depending on the dispersion of the distribution of haloes hosting QSO, a small fraction of haloes coming from the incomplete part of the simulation enter in the final mock. We quantify the effect of the resolution in our catalogues with the MDPL, where this effect is negligible thanks to its higher resolution. MDPL has enough resolution to cover the halo mass range for the QSO population. However, its volume is smaller than the one covered by eBOSS, so one cannot construct a complete light-cone without box replications. Furthermore, the shot noise from a mock using this volume is very large, due to the low number density of the observed sample. In Appendix A, we show this effect by comparing the mocks generated from both simulations.

We include the redshift evolution in the number density and of the clustering when constructing light-cones from the BigMDPL simulation. These light-cones cover the redshift range $0.9 < z < 2.2$ and 1,481.75 deg^2 of the sky, which is comparable with the area of Y1Q. The mocks are built with the SUrvey GenerAtoR code (SUGAR; Rodríguez-Torres et al. 2016). In this procedure, we use all available snapshots from the BigMDPL simulation, $z = 2.145$, 1.445,1, 0.8868. In order to analyse the effects of the incompleteness, we select only the closest snapshots from the MultiDark simulation ($z = 1.425$, 0.987, see Appendix A). We present results from three different light-cones, the first one uses a single set of parameters to describe the Y1Q (BigMDPL-QSO). The second one is obtained by fitting the clustering in four redshift bins with a different set of parameters (BigMDPL-QSOZ). The last light-cone uses a single set of parameters, but only host haloes are included (the fraction of substructures is equal to zero, BigMDPL-QSO-NSAT).

2.4 Galaxy mocks for QSO (GLAM)

In order to estimate the uncertainties in the clustering measurements, we use the GaLAxy Mocks (GLAM) scheme for the eBOSS quasar sample. For this application, GLAM implements a new parallel particle mesh method (Klypin & Prada 2017) to construct the dark matter density field and an optimization to populate the simulation with quasars (Comparat et al., in preparation). We run the SUGAR code to construct light-cones (Rodríguez-Torres et al. 2016). Errors are extracted from the covariance matrix of 1000 GLAM-QSO mocks which cover the same area as the data. They are computed using the diagonal terms, $\sigma_i(x_i) = \sqrt{C_{ii}}$, thus these errors correspond to one standard deviation (1σ) away from the mean value of the mocks. We use the covariance matrix estimator given by

$$C_{ij} = \frac{1}{n_s - 1}\sum_{k=1}^{n_s}\left(x_i^k - \mu_i\right)\left(x_j^k - \mu_j\right), \quad (7)$$

where n_s is the total number of mocks and the mean of each measurement is

$$\mu_i = \frac{1}{n_s}\sum_{k=0}^{n_s} x_i^k. \quad (8)$$

[2] http://www.cosmosim.org/

Using the covariance matrix from these mocks we perform the fitting with the χ^2 statistics,

$$\chi^2 = \sum_{ij} \left[x_i^d - x_i^m\right] C_{ij}^{-1} \left[x_j^d - x_j^m\right], \quad (9)$$

where x_i^m and x_i^d are the measurements from the model and the data in the bin i, respectively. χ^2 values presented in this work are computed from the monopole of the correlation function.

3 CLUSTERING MODEL

One of the best ways to study the observed clustering of a survey is to simulate not only the effect of the gravity on the dark matter but also on the baryonic matter. In this case, stellar physics should be included to provide a direct prediction of the relation between dark matter haloes and the galaxies and their evolution in time. This approach is undertaken by hydrodynamical simulations, that include galaxy formation processes, stellar physics and AGN feedback. EAGLE (Rahmati et al. 2015) and ILLUSTRIS (Sijacki et al. 2015) are two of the most recent realizations which predict a realistic distribution of galaxies and quasar populations. However, these simulations are constructed in rather small boxes of $\sim 75\,h^{-1}$ Mpc and this impedes studies of the large-scale structure. The large amount of computational resources required for a hydrodynamic simulation is prohibitive and the computation of volumes comparable to observations nearly infeasible.

An alternative approach, cheaper in computational time, is to use the dark matter only simulations and add galaxies in a statistical way. There are two widely used models based on these statistical relations. The first one is the HOD (e.g. Guo et al. 2014), which gives the probability, $P(N|M_{\rm h})$, that a halo of mass $M_{\rm h}$ hosts N galaxies. This probability is described by a fitting formula, which is fixed using the clustering measurements from the observational data. The second method to populate the dark matter haloes is the HAM (e.g. Reddick et al. 2013). This model assumes that the most massive galaxies populate the most massive haloes.

3.1 The modified SHAM model

Favole et al. (2016) introduced a modified (sub)halo abundance matching (SHAM), designed to reproduce the clustering of the BOSS ELG sample. They select haloes from the simulation using a probability function which is the sum of two terms corresponding to host and satellite haloes. This probability is a Gaussian function described by three parameters: the mean mass, the width of the distribution and the satellite fraction. This method is useful to describe incomplete samples, such as the Y1Q, which is not complete in halo mass or stellar mass whatsoever. In this paper, we use a similar model to study the clustering of quasars. Favole et al. (2016) use the virial mass of haloes to implement their method. Instead of that, we use $V_{\rm max}$ and assume that the distribution of haloes hosting quasars has a Gaussian shape. The most general model is split in central and satellite haloes as done in Favole et al. (2016). When a QSO is located in the centre of a host halo, it is denoted as a central QSO. The satellite fraction refers to the fraction of QSO living in a subhalo. This fraction does not represent systems of binary quasars. The central halo which is the counterpart of a satellite QSO can host another kind of galaxy.

In the case of quasars, we do not use the luminosity or the stellar mass of the observed sample. Our model only uses the $V_{\rm max}$ distribution of haloes, as done by Nuza et al. (2013). Rodríguez-Torres et al. (2016) extend the HAM technique implemented by Nuza et al. (2013) using the stellar mass function and modelling the incompleteness of the sample. In that study, galaxies are assigned to haloes via a standard HAM and then they are downsampled to obtain the observed stellar mass distribution. Here, we assume that the intrinsic scatter between quasars and dark matter haloes, plus the incompleteness of the sample will produce a $V_{\rm max}$ distribution with a Gaussian shape. Then, the model orders haloes by $V_{\rm max}$ and downsamples objects as done by Rodríguez-Torres et al. (2016).

3.2 Implementation

Assuming that the final $V_{\rm max}$ distribution of the simulated quasar catalogue is Gaussian, we need to construct a probability distribution function that selects haloes from the complete simulation based on this condition. In a general case, the $V_{\rm max}$ distribution of the final catalogue will be

$$\begin{aligned}\phi_{\rm QSO}(V_{\rm max}) &= \phi^s_{\rm QSO} + \phi^c_{\rm QSO} \\ &= P_s(V_{\rm max})\phi^s_{\rm sim}(V_{\rm max}) + P_c(V_{\rm max})\phi^c_{\rm sim}(V_{\rm max}) \\ &= \mathcal{G}_s(V_{\rm max}) + \mathcal{G}_c(V_{\rm max}),\end{aligned}$$

where $\phi^c_{\rm sim}$ and $\phi^s_{\rm sim}$ represent the $V_{\rm max}$ distribution of host haloes and subhaloes, respectively, $\mathcal{G}_c$ and $\mathcal{G}_s$ are Gaussian functions with mean $V_{\rm mean}$, standard deviation $\sigma_{\rm max}$ and each one is normalized using

$$\int \mathcal{G}_s(V_{\rm max}, z){\rm d}V_{\rm max} = N_{\rm tot}(z) f_{\rm sat}$$

$$\int \mathcal{G}_c(V_{\rm max}, z){\rm d}V_{\rm max} = N_{\rm tot}(z)(1 - f_{\rm sat}),$$

where $N_{\rm tot}(z)$ is the total number of quasars per redshift bin given by the observed number density.

In order to construct the probability distribution, we sort all haloes in the simulation and compute the maximum circular velocity function ($V_{\rm max}$) for subhaloes and host haloes separately. Using the fraction of satellites as a free parameter and the observed number density, we normalize the Gaussian distribution for central and satellite haloes. We split all haloes of the simulation in bins of $V_{\rm max}$ and compute the probability of assigning a quasar to a dark matter halo (central or satellite) per bin as

$$P_{s/c}(V_{\rm max}) = \frac{N^{\rm gaus}_{s/c}}{N^{\rm tot}_{\rm sub/host}}, \quad (10)$$

where $N^{\rm tot}_{\rm sub/host}$ is the total number of subhaloes/host haloes in the range $[V_{\rm max} - \Delta V_{\rm max}/2, V_{\rm max} + \Delta V_{\rm max}/2]$ and $N^{\rm gaus}_{s/c}$ is the number of satellite/central quasars necessary to produce the final Gaussian shape. Using equation (10), we downsample all haloes in the simulation to obtain the QSO mock catalogue.

Our model consists of five different parameters, the mean and standard deviation values for satellite and central distributions and the fraction of satellites. However, we assume the same mean and standard deviation for central and satellite quasars thus decreasing the number of parameters. In addition, the current data do not provide enough information at small scales ($<1.0\ h^{-1}$ Mpc) to extract precise information about the standard deviation of the distribution and the satellite fraction of the eBOSS QSO sample. For these reasons, our unique parameter to fit the clustering is the mean value of the distribution ($V_{\rm mean}$).

3.3 Parameters

The most general model is defined by three parameters. However, due to the poor information at small scales, we only use one free parameter ($V_{\rm mean}$) to describe the Y1Q sample. Fig. 2 presents the χ^2 maps we obtain for different combinations of the three parameters $V_{\rm mean}$, $\sigma_{\rm max}$ and $f_{\rm sat}$. We find the satellite fraction, $f_{\rm sat}$, to be degenerate with $V_{\rm mean}$ (left-hand panel of Fig. 2) and this degeneracy could be broken only with information from the one halo term. However, the current Y1Q data do not allow going to those scales. For this reason, we do not fix the number of satellites in two of the three mocks presented, which means that host haloes and subhaloes are not distinguished when the selection is implemented. In addition, just as Favole et al. (2016), we do not find a dependency of the clustering with the width of the Gaussian distribution ($\sigma_{\rm max}$). $\sigma_{\rm max}$ cannot be constrained with the current data as is shown in the right-hand panel of Fig. 2. In the mass regime where QSOs live, $\sigma_{\rm max}$ impacts the clustering at small scales ($<0.5\ h^{-1}$ Mpc), so it is not possible to constrain this parameter.

In the case of quasars, at scales larger than $1.0\ h^{-1}$ Mpc, the clustering amplitude only depends on $V_{\rm mean}$. In order to fix $\sigma_{\rm max}$, we use previous results in the literature. The model shown in Chehade et al. (2016) is consistent with a width in $V_{\rm max}$ of $\sigma_{\rm max} = 45\,{\rm km\ s^{-1}}$. However, due to the resolution of BigMultiDark, we decrease this value to $\sigma_{\rm max} = 30\,{\rm km\ s^{-1}}$. If we use larger values of $\sigma_{\rm max}$, we will include a larger fraction of haloes from the incomplete mass region of the simulation. Fixing $\sigma_{\rm max} = 30\,{\rm km\ s^{-1}}$, we ensure that the BigMDP light-cones have only ∼2 per cent of haloes selected from regions where the incompleteness is greater than 10. Thus, we avoid including any unphysical effects coming from the low resolution of the simulation.

Thus, our model describes the quasar sample with a single parameter which is fixed by minimizing the χ^2 distribution. As mentioned previously, we use the monopole of the correlation function between 10 and 40 h^{-1} Mpc (10 data points shown in Fig. 1), thereby avoiding systematic effects that influence the clustering measurements at small scales. Varying $V_{\rm max}$, we find that the χ^2 distribution is well described by a quadratic function. This is used to find the parameter that best represents the data.

4 RESULTS

We compare the Y1Q 2-point correlation function (2PCF) with that of the mocks using the χ^2 statistics with 9 degrees of freedom (10 data points and 1 parameter). In order to compute the 2PCF, we use a modified version of the Correlation Utilities and Two-point Estimation code (CUTE; Alonso 2012). We first analyse the complete sample, using the clustering measurements in the redshift range $0.9 < z < 2.2$. We find the best value for the parameter $V_{\rm mean} = 341.2\,{\rm km\ s^{-1}}$, which corresponds to a sample of mock QSO with mean mass $\log[M_{200}/{\rm M}_\odot] = 12.66 \pm 0.16$. Fig. 1 presents the clustering measurements (2PCF and power spectrum) along with the prediction of the best-fitting mock light-cone. We find an excellent agreement between the data and the model for the studied scales.

When fitting is performed using the clustering of the complete redshift range, the evolution of the mass distribution is not taken into account. In order to investigate this effect, we divide the sample in four redshift bins and find the best parameter to match the clustering in each individual redshift range. It slightly improves the quality of the fits, presented in Table 3 which gives the best-fitting values of $V_{\rm mean}$ and their corresponding reduced χ^2.

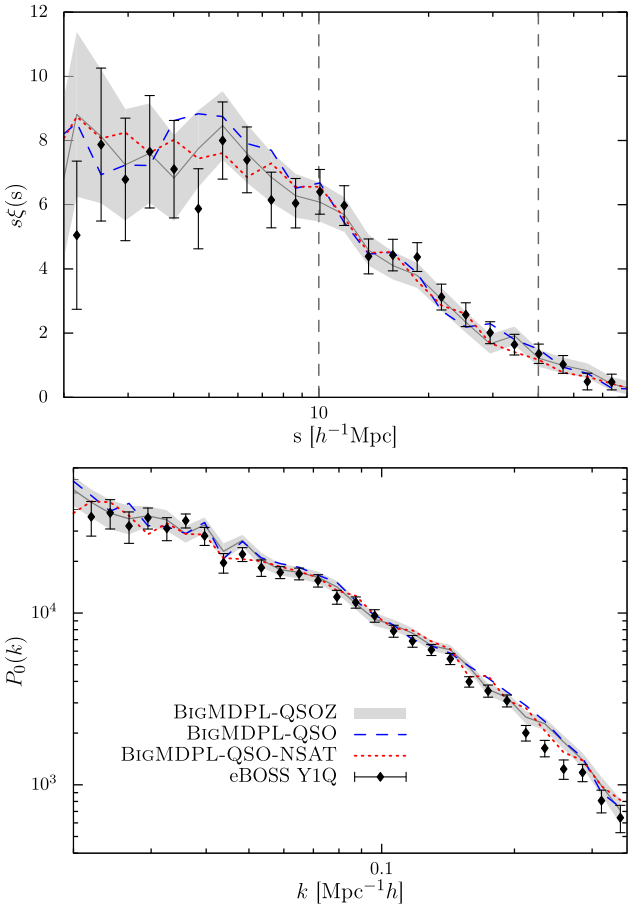

Figure 1. Top panel: monopole of the correlation function in configuration space of Y1Q (points with error bars). The shaded area represents the BigMDPL-QSOZ light-cone fitted in four different redshift bins. The dashed line represents the BigMDPL-QSO light-cone fitted on a single redshift bin and the dotted line is the BigMDPL-QSO-NSAT. The vertical lines represent the limit values used for fitting the parameters. Bottom panel: monopole of power spectrum of the Y1Q (points with error bars) and the three BigMDPL light-cone. The agreement between the best model and the data is remarkable. Error bars and dashed areas are computed using 1000 GLAM catalogues and correspond to 1σ deviation from the mean value. Differences at high k are due to redshift errors.

Table 3. Results of the fit per redshift bin. A gives the area in deg^2 subtended by the mock light-cone. z bin gives the lower and upper boundary of the redshift bin. $V_{\rm mean}$ is the best-fitting parameter found. $\log_{10}(M_{200}/{\rm M}_\odot)$ is the corresponding mean ± standard deviation of the halo mass of the population selected. χ^2_r is the reduced χ^2 per 9 degrees of freedom. We fixed $\sigma_{\rm max} = 30\,{\rm km\ s^{-1}}$ and $f_{\rm sat}$ is percentage of satellites in the catalogue.

A (deg^2)	z bin	$V_{\rm mean}$ (s^{-1}km)	$\log_{10}\frac{M_{200}}{{\rm M}_\odot}$	χ^2_r	$f_{\rm sat}$
		BigMDPL-QSO			
1481.75	0.9–2.2	341.2 ± 30.0	12.66 ± 0.16	1.78	5.3
		BigMDPL-QSOZ			
3275.06	0.9–1.2	282.8 ± 30.2	12.53 ± 0.17	1.47	9.0
2371.81	1.2–1.5	324.1 ± 30.1	12.63 ± 0.14	1.85	5.0
1879.13	1.5–1.8	339.5 ± 29.9	12.69 ± 0.14	1.70	4.3
1481.75	1.8–2.2	353.5 ± 29.7	12.60 ± 0.13	2.24	3.3
		BigMDPL-QSO-NSAT			
1481.75	0.9–2.2	349.5 ± 30.3	12.70 ± 0.16	1.52	0.0

Table 4. Mean halo mass and satellite fraction prediction from the BigMDPL light-cones.

Light-cone	$V_{\rm mean}$ (s^{-1}Km)	$\log_{10}[M_{200}/M_\odot]$	$f_{\rm sat}$
BigMDPL-QSOZ	326.9	12.61	0.048
BigMDPL-QSO	341.2	12.66	0.053
BigMDPL-QSO-NSAT	349.5	12.70	0.0

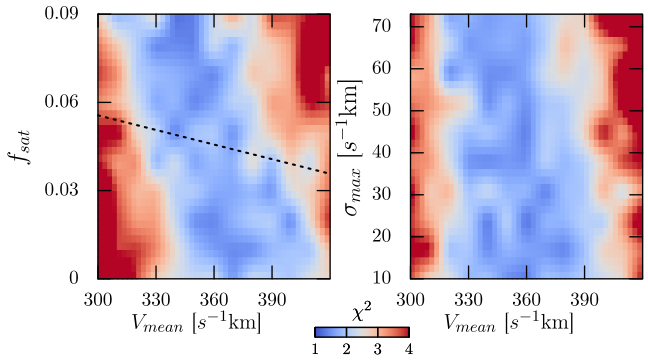

Figure 2. χ^2 maps for the three parameters of the model implemented on the BigMDPL-QSO. The left-hand panel shows the satellite fraction versus $V_{\rm mean}$. It is possible to note a degeneracy between both parameters. This is why we use the $f_{\rm sat}$ given by the simulation. The dashed line shows the satellite fraction given by the simulation for different values of $V_{\rm mean}$. The right-hand panel presents $\sigma_{\rm max}$ versus $V_{\rm mean}$. $\sigma_{\rm max}$ cannot be constrained using the current data.

Comparing the values of M_{200} presented in Table 3 with those of Table 2, we infer that the best-fitting mocks have less than 1 per cent of objects taken from a bin where the completeness is lower than 90 per cent. The effect of the resolution on the clustering is discussed in more detail in Appendix A.

Table 3 shows the values of satellite fractions of the BigMDPL light-cones. As we explained in Section 3.3, we do not use $f_{\rm sat}$ as a parameter so the fraction of satellites in the mock has the same dependency with $V_{\rm max}$ as the complete simulation. The third light-cone is the only catalogue where we fix $f_{\rm sat} = 0$. We include it to show the impact of removing all substructures from our analysis. The second parameter of the model, $\sigma_{\rm max}$ is also not constrained (see Fig. 2). A similar problem was found by Shen et al. (2013), their HOD parameters are largely degenerate and the fraction of satellites is not well constrained. For these reasons, we only vary the mean value of the Gaussian distribution ($V_{\rm mean}$) to fix the clustering of the model.

4.1 Trends of the QSO clustering with redshift

The signal of the quasar clustering does not have an important evolution, as shown in Fig. 3. The monopole varies mildly in the linear regime in all four redshift bins. If we assume a constant distribution of $V_{\rm max}$ for the whole redshift range, the evolution of the dark matter field will produce a non-constant signal of clustering in the different redshifts. In order to reproduce the observed evolution and predict a most realistic linear bias, we divide the complete redshift range into four regions, fitting the clustering of the light-cone in each bin. Table 3 presents the redshift range and the best-fitting parameters found to match the observed data. We use different areas for each redshift bin to maximize the volume

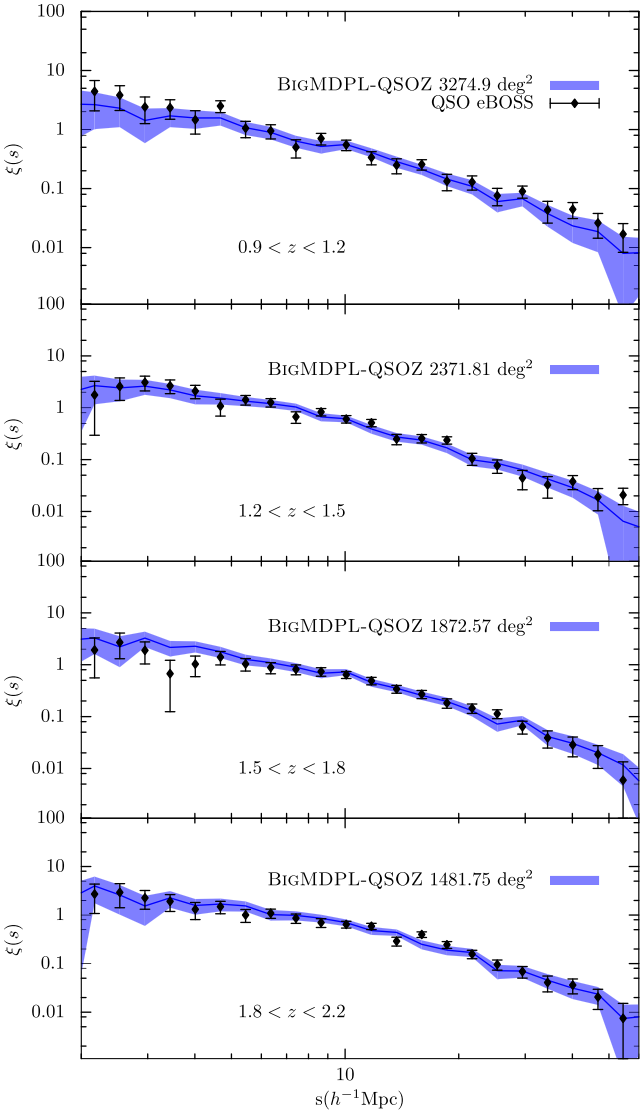

Figure 3. Monopole 2PCF versus redshift. We show the Y1Q (points) and the best-fitting mock (shaded area) of the BigMDPL-QSOZ light-cone (see Table 3). Each panel corresponds to a different redshift bin. Error bars and dashed areas are computed using 1000 GLAM catalogues and correspond to 1σ deviation from the mean value.

used from the simulation. These larger areas increase the statistics and reduce the shot noise in the 2PCF of the mocks as seen in Table 3.

Fig. 1 shows the monopole of the correlation function and the power spectrum of the three different mocks (BigMDPL-QSO/QSOZ/QSO-NSAT) compared to the observed data for the whole redshift range. All light-cones can reproduce the eBOSS data with a good agreement. We underline that the BigMDPL light-cones have shot noise and cosmic variance similar to the data. Due to these large errors in the model and the data, it is difficult to distinguish which light-cone reproduces the data better in the complete redshift range. However, if the model reproduces the clustering at different redshifts, we can estimate the evolution of the bias with better accuracy.

In order to quantify the difference between two models, we compare them using the Bayes factor. We can compute it with the

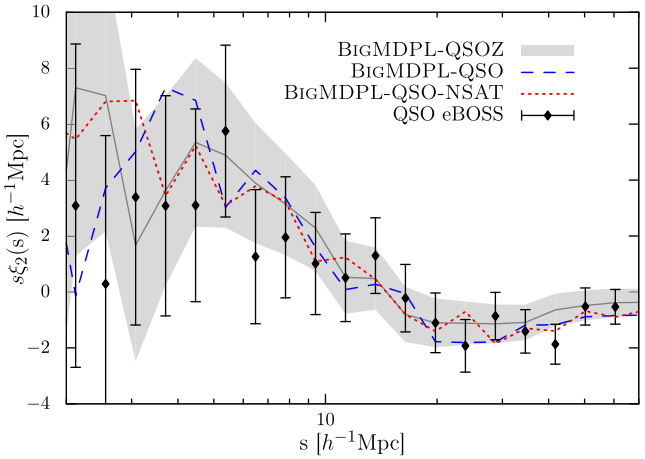

Figure 4. Quadrupole versus comoving scale in redshift space predicted by the BigMDPL-QSOZ (shaded region), BigMDPL-QSO (dashed line) and BigMDPL-QSO-NSAT (dotted lines) compared to the Y1Q (black points). All mocks are in agreement with observations. Error bars and shaded areas are computed using 1000 GLAM catalogues and correspond to 1σ deviation from the mean value.

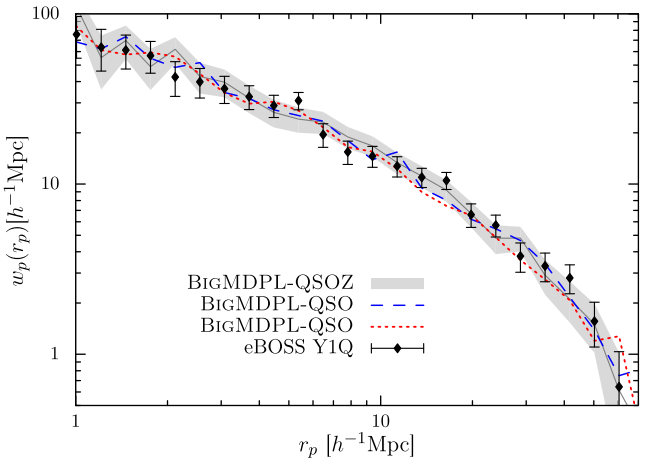

Figure 5. Projected correlation function predicted by the BigMDPL-QSOZ (shaded region), BigMDPL-QSO (dashed line) and BigMDPL-QSO-NSAT (dotted line) compared to the Y1Q (black points). The width of the shaded area represents 1σ errors computed with 1000 GLAM catalogues and correspond to 1σ deviation from the mean value. Our model reproduces the clustering for all relevant scales.

maximum likelihood

$$P(\mathbf{x}|\mathbf{p}) = \frac{|\tilde{\mathbf{C}}^{-1}|}{(2\pi)^p} \exp\left[-\frac{1}{2}\sum_{ij}\left(x_i^d - x_i(\mathrm{p})\right)\tilde{C}_{ij}^{-1}\left(x_j^d - x_j(\mathrm{p})\right)\right] \tag{11}$$

where x^d represents the data and x(p) the model. We estimate the inverse covariance matrix using equation (7) and correcting for bias using the Hartlap factor (Hartlap, Simon & Schneider 2007)

$$\tilde{C}_{ij}^{-1} = \frac{N_{\mathrm{mock}} - N_{\mathrm{p}} - 2}{N_{\mathrm{mock}} - 1} C_{ij}^{-1}, \tag{12}$$

where N_{p} represents the number of data points used. The Bayes factor between the BigMDPL-QSO and the BigMDPL-QSOZ model is

$$K = \frac{P(\xi_{\mathrm{data}}|\xi_{\mathrm{QSOZ}})}{P(\xi_{\mathrm{data}}|\xi_{\mathrm{QSO}})} = 5.45. \tag{13}$$

This result suggests that BigMDPL-QSOZ model is more substantially supported by the data than BigMDPL-QSO. The Bayes factor between the BigMDPL-QSOZ and the BigMDPL-QSO-NSAT is $K = 1.67$. In this case, we cannot conclude which model better reproduces the data. Furthermore, the BigMDPL light-cones have an important variability between realizations when the random seed is changed and it is not possible to construct a sufficient number of independent light-cones to make a definitive statement about the two models. In terms of χ^2 both light-cones are in agreement with the current data, though including a model with more parameters will improve the fitting of the data.

4.2 Checking $\xi_2(s)$ and $w_p(r_p)$

The quadrupole is very sensitive to processes affecting the small scales. Effects due to fibre collisions have an important impact at scales beyond the fibre size. However, the effect of fibre collisions is very small in the QSO sample. The most important observational effect is due to redshift errors, as shown in Appendix A. Fig. 4 shows the quadrupole of the BigMDPL-QSO, BigMDPL-QSOZ and BigMDPL-QSO-NSAT light-cones compared to the observations. All light-cones reproduce the data within 1σ error. This agreement suggests that we are using a reasonable model to account for redshift errors. We note that the BigMDPL-QSOZ light-cone reproduces the quadrupole better than the other two light-cones.

We compared the projected correlation function for the three light-cones and the observed data, finding a good agreement shown in Fig. 5.

The clustering predicted by the best-fitting model, which is mainly determined by the V_{mean}, reproduces with good agreement the two-point statistics of the observed data. We do not find significant differences between the three light-cones presented, all of them can reproduce the two-point statistics of the complete Y1Q sample with good agreement.

4.3 Bias

The Y1Q data allows for accurate measurements of the correlation function $\xi(r)$ and of the quasar bias b_{Q}, within the redshift range $0.9 < z < 2.2$. Laurent et al. (in preparation) obtain $b_{\mathrm{Q}} = 2.45 \pm 0.05$, when averaged over separations between 10 and 90 h^{-1} Mpc. This value is compatible with previous SDSS measurements, $b_{\mathrm{Q}}(z = 1.58) = 2.42 \pm 0.40$, by Ross et al. (2009).

We estimate the bias using the dark matter counterpart of the QSO mock light-cone. Using the autocorrelation of the dark matter sample, and the correlation function of the QSO mock in real space, we estimate the bias using

$$b(r)^2 = \frac{\xi(r)}{\xi_{\mathrm{DM}}(r)}. \tag{14}$$

Fig. 6 presents the bias of the BigMDPL-QSOZ and the BigMDPL-QSO compared to previous studies.

The bias measurements presented in Fig. 6 come from spectroscopically confirmed quasars in the two degree field (Porciani & Norberg 2006) at $0.8 < z < 2.1$, SDSS-I/II (Ross et al. 2009) at $z < 2.2$, the Quasar Dark Energy Survey pilot (2QDESp; Chehade et al. 2016) for redshift between 0.8 and 2.5 and the BOSS sample (Eftekharzadeh et al. 2015) at $2.2 < z < 2.8$. All these studies parametrize the real space correlation function by a power law,

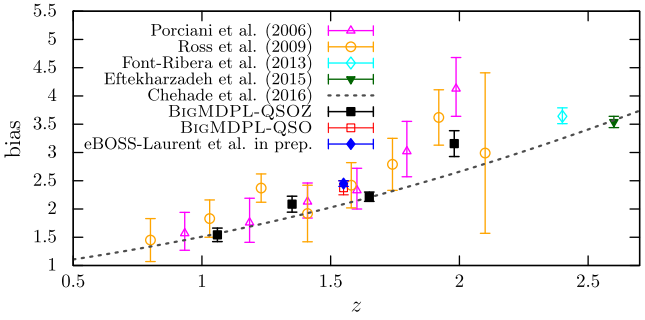

Figure 6. QSO bias as a function of redshift. The bias is computed using BigMDPL-QSOZ and BigMDPL-QSO light-cones. We include results from Chehade et al. (2016), Eftekharzadeh et al. (2015), Font-Ribera et al. (2014), Ross et al. (2009) and Porciani & Norberg (2006). eBOSS bias measurements are in agreement with previous results and about 10 times more precise. Results of eBOSS from Laurent et al. (in preparation) are also included.

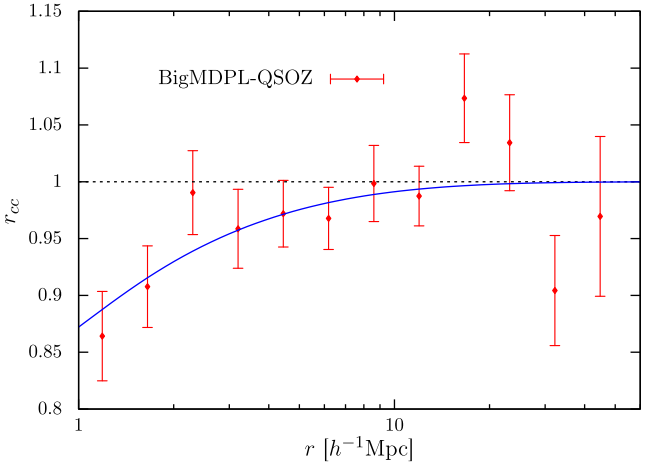

Figure 7. Cross-correlation coefficient between the dark matter field and the BigMDPL-QSOZ light-cone. The best model from (17) is shown with a solid line.

$\xi(r) = (r/r_0)^{\gamma}$, which can be related with the observed correlation function (redshift space) by

$$\xi(s) = \left(b_{\mathrm{Q}}^2 + \frac{2}{3}b_{\mathrm{Q}}f + \frac{f^2}{5}\right)\xi(r), \tag{15}$$

where $f = [\Omega_{\mathrm{m}}(z)]^{0.56}$ is the gravitational growth factor. In addition, we include measurements of quasars via Lyman α absorption at redshift 2.4 from the BOSS sample (Font-Ribera et al. 2014). Eftekharzadeh et al. (2015) also show a comparison between different estimations of the bias. At the redshifts studied, the bias measurements obtained in our study are in good agreement (see Fig. 6) and they are a factor 5 to 10 times more precise than previous studies.

4.4 Cross-correlation coefficients

The linear bias provides a good description of the relationship between dark matter and QSO mock in the linear regime. However, a single parameter b_{Q} is not enough to understand the link between galaxies and dark matter at all scales. To parametrize this relationship, we use the second-order bias, which is related to scales smaller than 10 h^{-1} Mpc. The second-order bias is inferred from the cross-correlation coefficient. It gives an estimation of the correlation between the positions of quasars and the dark matter field (Dekel & Lahav 1999). The cross-correlation, denoted r_{cc}, between quasars and the dark matter field is defined as

$$r_{\mathrm{cc}}(r) = \frac{\xi_{qm}(r)}{\sqrt{\xi_{qq}(r)\xi_{mm}(r)}}, \tag{16}$$

where q denotes the quasar sample and m the dark matter. r_{cc} is sensitive to the non-linear stochastic bias of the sample. Fig. 7 shows the cross-correlation coefficient between BigMDPL-QSOZ and the dark matter field. For scales larger than 10 h^{-1} Mpc, the cross-correlation function is consistent with 1. As expected, in this regime, we have $\xi_{gm} = b_{\mathrm{Q}}\xi_{mm}$ and $\xi_{gg} = b_Q^2\xi_{mm}$. At smaller separations, r_{cc} becomes smaller than one. This tendency is described in perturbation theory (Baldauf et al. 2010), where r_{cc} is described with the second-order bias by

$$r_{\mathrm{cc}}(r) \approx 1 - b_2^2\frac{\xi_{\mathrm{lin}}(r)}{4}, \tag{17}$$

where b_2 is the second-order bias and ξ_{lin} is the linear correlation function. The cross-correlation coefficient fit directly to the clustering by $b_2 = 0.314 \pm 0.030$. This relation is sufficient for the scales studied ($1 < rh^{-1}$ Mpc < 10), see the solid line in Fig. 7.

4.5 Halo occupation distribution

Table 5 shows the mean mass of haloes hosting quasars, the satellite fraction characterizes how quasars populate dark matter haloes and the mean value of V_{max} for all light-cones built in this study.

If the satellite fraction is not fixed (no distinction between haloes and subhaloes), we obtain a non-negligible fraction of satellites, ∼5 per cent. This value is consistent with Shen et al. (2013) which finds a satellite fraction of 6.8 per cent. However, due to the degeneracy between V_{mean} and f_{sat}, our model could also match the clustering with a negligible fraction (Fig. 2), as presented in Richardson et al. (2012).

Another way to formulate how QSO populate the density field is the probability of finding N quasars in a halo of mass M ($\langle N(M)\rangle$), namely the HOD model. This method describes how quasars would statistically populate haloes using a set of parameters fitted directly on the clustering. In SHAM models, $\langle N(M)\rangle$ is given by the halo catalogue by counting the total number of host haloes and the number of QSO per bin of mass. Fig. 8 shows the HOD predicted by the BigMDPL-QSO light-cone. We use this light-cone rather than the other as it has a negligible fraction of objects from the incomplete

Table 5. Mass prediction of haloes hosting quasars for different samples. It is presented with the name of the method used to analyse the sample and the used redshift range.

Sample	N_{QSO}	z	Method	$\log_{10}(M_h/\mathrm{M}_\odot)$
eBOSS[a]	68 269	0.9–2.2	HAM	12.5–12.82
SDSS-I/II[b]	8198	0.3–0.9	Power-law fit	12.75
SDSS-I/II[c]	48 000	0.4–2.5	HOD	12.70–12.77
BOSS[d]	27 129	2.2–2.8	Power-law fit	12.59–11.65
BOSS[e]	55 826	2.2–2.8	Power-law fit	11.63-12.63
2QDESp[f]	10 000	0.8–2.5	Power-law fit	12.17–12.64

Notes. [a]This work; [b]Shen et al. (2013); [c]Richardson et al. (2012); [d]White et al. (2012); [e]Eftekharzadeh et al. (2015); [f]Chehade et al. (2016).

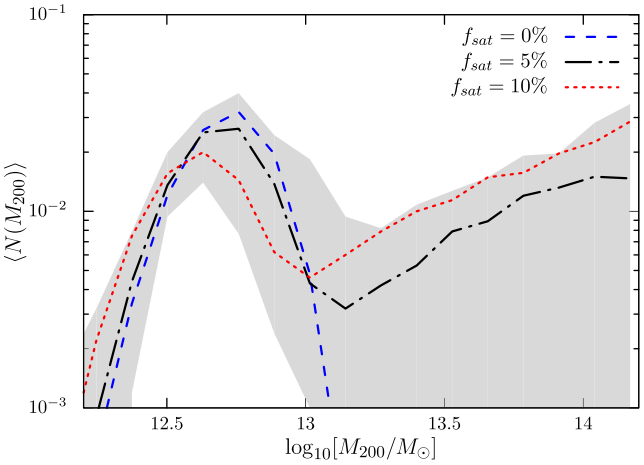

Figure 8. HOD for central plus satellites predicted from the BigMDPL-QSO light-cone. We present three light-cones using different fraction of satellites. The shaded area is computed adding 1σ error in the $V_{\rm mean}$ parameter for each light-cone. In addition, we vary the width of the distribution from 10 to 60 s^{-1} km to see the impact of this parameter in the HOD. $f_{\rm sat}$ is also changed from 0 to 0.12

part of the BigMDPL simulation. It also allows $\sigma_{\rm max}$ and $f_{\rm sat}$ to vary in a wide range, letting us show the dependency of $\langle N(M) \rangle$ on these parameters reflected in the different lines of Fig. 8.

Additionally, we construct light-cones with different $V_{\rm mean}$ including variations of 1σ from the best fit. We also vary the width of the distribution between 10 and 60 s^{-1} km. We do not use a larger $\sigma_{\rm peak}$, because we do not want to include a large fraction of objects coming from the incomplete part of the simulation. $f_{\rm sat}$ also varies between 0 and 10 per cent. The shaded area in Fig. 8 represents all HODs encompassed by these parameter variations.

Compared to previous HOD results (Shen et al. 2013), our model puts new constraints for masses below 10^{13} $M_\odot$. We find a distribution dominated by the mean halo mass of the sample. However, $\langle N(M) \rangle$ has a strong dependency with the other two parameters of the model, which we cannot constrain with the current data. An improvement on small scales of the QSO clustering or the cross-correlation between ELG and QSO in future surveys would constrain $\sigma_{\rm peak}$ and $f_{\rm sat}$ and therefore provide better HOD predictions.

5 DISCUSSION

Previous HOD analysis of the SDSS QSO sample combined different data sets to get more information about the distribution of QSOs inside haloes. However, due to large uncertainties in the data, the parameters of the HOD remain degenerate. eBOSS will greatly increase the statistical size of quasar samples, giving an excellent opportunity to learn more about this population and its connection with the dark matter. What we do here is to present the first study of the Y1Q clustering introducing a modified HAM that allows us to predict the HOD, masses of the dark matter haloes and the bias of the sample.

Several studies have provided information about quasars at different redshifts using their clustering measurements. Richardson et al. (2012) study the clustering of the 48 000 QSO from the SDSS sample in the redshift range $0.4 < z < 2.5$. They interpret the measurements of the projected correlation function at redshift 1.4. In addition, 4426 spectroscopically identified quasars in the redshift interval $2.9 < z < 5.4$ (Shen et al. 2007) are used to study the small-scale clustering. However, they use a regular HOD without including a duty cycle. For this reason, their parameters reproduce the clustering, but most of them are unphysical. Shen et al. (2013) study the two-point cross-correlation function of 8198 SDSS QSO and 349 608 BOSS CMASS galaxies in the redshift range $0.3 < z < 0.9$. They provide predictions of the HOD from quasars. However, the large degeneracies of the parameters make it impossible to have a well-constrained HOD. The BOSS sample provides a set of CORE QSO which is studied by Eftekharzadeh et al. (2015). They extend the analysis of the projected correlation function of the BOSS sample done by White et al. (2012). In that analysis, ∼70 000 quasars in the redshift range 2.2–3.4 are studied. In a more recent study, Chehade et al. (2016) combine the optical photometry of the 2dF Quasar Dark Energy Survey pilot (2QDESp) and the bands of the *Wide-field Infrared Survey Explorer* (*WISE*) to provide a sample of ∼10 000 QSO in the redshift range 0.8–2.5. Our study uses a larger and wider QSO sample than in previous works. It allows us to have a good estimation of the clustering in the redshift range $0.9 < z < 2.2$.

The mean mass of haloes hosting quasars has been measured by different methods finding a reasonable agreement between their results. However, the range of masses cover by quasars is still not well constrained. Richardson et al. (2012) predict a mean halo mass for central haloes $M_{\rm cen} \sim 10^{12.77}$ $M_\odot$ with a small fraction of QSO satellites, 7.4×10^{-4}. This result is in agreement with the BigMDPL-QSO-NSAT, which provides host halo masses for quasars of $10^{12.7 \pm 0.16}$ $M_\odot$. Shen et al. (2013) model the cross-correlation between CMASS galaxies and QSO by a power law, $\xi_{\rm QG} = (r/r_0)^\gamma$, with $r_0 = 6.61 \pm 0.25\ h^{-1}$ Mpc and $\gamma = 1.69 \pm 0.07$ for scales r = 2–25 h^{-1} Mpc. They find a characteristic mean halo mass of $10^{12.8}$ $M_\odot$. In contrast to Richardson et al. (2012), a non-negligible satellite fraction is predicted by Shen et al. (2013). They find that 6.8 per cent of QSO are hosted by subhaloes. This result is in better agreement with our mocks without fixing the fraction of satellites, which predict ∼5 per cent of quasars living in subhaloes. The halo masses predicted by this HOD are also in agreement within 1σ errors with our measurements. Nevertheless, they have larger degeneracies between their parameters. From the BOSS sample, White et al. (2012) find the quasar halo masses covering a wide mass range between $10^{11.59}$ and $10^{12.65}$ $M_\odot$. Just as in the previous cases, these values of masses are still in agreement with our results shown in Table 3. The Chehade et al. (2016) results are compared with other surveys (SDSS, 2QZ and 2SLAQ). As in previous works, they find no evidence of a dependency between the clustering and the luminosity of the QSO. In addition, they show that quasar clustering depends on redshift, in particular, when BOSS data are included. They describe the clustering of the sample using a power law, where $r_0 = 7.3 \pm 0.1\ h^{-1}$ Mpc at redshift 2.4, while the correlation scale for the whole redshift range is $r_0 = 6.1 \pm 0.1\ h^{-1}$. Their measurements are consistent with host haloes masses of $\sim 10^{12.46}$. Future observations will allow cross-correlations between ELGs and quasars, which will enable a better understanding of the distribution of quasars within the dark matter halo. These measurements could fix the satellite fraction of quasars. However, the width of the distribution is more difficult to constrain. In the similar case of ELG, Favole et al. (2016) faced an equivalent problem to describe their clustering. They use constraints from lensing measurements to understand the clustering on the smallest scales. Unfortunately, such measurements are not available for quasars.

Using our model, the signal of the clustering in the linear regime is dominated by the mean halo mass of the distribution. This is clear in the HOD (Fig. 8), where the distribution has a strong peak near the mean halo mass of the sample. We find a more constrained HOD region for quasars than Shen et al. (2013). However, more information from small scales is needed to have better constraints in the satellite fraction and width of the distribution in order to provide more realistic uncertainties. We find a bias equal to 2.37 ± 0.12 for the redshift range $0.9 < z < 2.2$, which is in good agreement with previous analysis and with eBOSS data from Laurent et al. (in preparation, Fig. 6). We provide measurements for the evolution of the bias using the BigMDPL-QSOZ light-cone, finding that the eBOSS quasars are in agreement with b_Q=1.54, 2.08, 2.21, 3.15 for redshift 1.06, 1.35, 1.65, 1.98. Furthermore, to give a complete parametrization of the scales studied in this work, we calculate the second-order bias from the cross-correlation coefficients, finding $b_2 = 0.314 \pm 0.030$. Table 5 presents a comparison of the halo mass predictions of previous studies and our result.

6 SUMMARY

We modelled the clustering of ∼70 000 optical quasars from the eBOSS Y1Q CORE sample in the redshift range $0.9 < z < 2.2$. We used a modified HAM that takes into account the incompleteness of the QSO sample and the intrinsic scatter between QSOs and dark matter haloes. This model was implemented in a light-cone constructed from a 2.5 h^{-1} Gpc simulation, covering an area comparable to the eBOSS Y1Q sample.

Our main results can be summarized as follows.

(i) We assume that the $V_{\rm max}$ distribution of haloes hosting QSOs is described by a Gaussian function which is defined by its mean and width plus one parameter for the satellite fraction. The current observations do not bear information on small-scale clustering. For this reason, we cannot constrain the fraction of satellites. Hence, we do not distinguish between host and subhaloes when the selection is done. The final mock thus has the same fraction of satellites as the complete simulation in the mass range used.

(ii) We model the clustering of the Y1Q using a single free parameter ($V_{\rm mean}$). The width of the Gaussian distribution is fixed to 30 $\rm s^{-1}$ km and we only impose a value to the satellite fraction in the BigMDPL-QSO-NSAT light-cone, for the other light-cones we do not fix this parameter.

(iii) The prediction of our model is in a good agreement with the 2PCF and the monopole of the power spectrum of the Y1Q data. The light-cone is constructed assuming Gaussian redshift errors given by Dawson et al. (2016). Their modelling improves the agreement between our model and the data. It provides a good description of the observed clustering on small scales, which is very sensitive to variations caused by these errors.

(iv) We construct three kinds of light-cones: one including the evolution of the parameters with redshift (BigMDPL-QSOZ), another describing the whole redshift range with a single parameter (BigMDPL-QSO) and a third one fixing the satellite fraction to zero (BigMDPL-QSO-NSAT). The mean halo masses are $10^{12.61}$, $10^{12.66}$ and $10^{12.70}$ $\rm M_\odot$, respectively.

(v) Using the Bayes factor, we find a strong evidence that the BigMDPL-QSOZ (four parameters) reproduces the data better than the BigMDPL-QSO (one parameter). However, we cannot make the same conclusion with the model without satellites, which reproduces the data with a similar agreement to the BigMDPL-QSOZ model.

(vi) We find a mean bias of the Y1Q sample equal to 2.37 ± 0.12 and a second-order bias $b_2 = 0.314 \pm 0.030$, which both describe the relation between the dark matter and the QSO mock for the studied scales.

BigMDPL-QSOs and GLAM-QSO eBOSS mocks are publicly available through the *Skies and Universes* website.[3]

ACKNOWLEDGEMENTS

SRT is grateful for support from the Campus de Excelencia Internacional UAM/CSIC.

SRT, JC, FP acknowledge support from the Spanish MICINN Consolider-Ingenio 2010 Programme under grant MultiDark CSD2009-00064 MINECO Severo Ochoa Award SEV-2012-0249 and grant AYA2014-60641-C2-1-P.

GY acknowledges financial support from MINECO/FEDER (Spain) under research grants AYA2012-31101 and AYA2015-63810-P.

The BIGMULTIDARK simulations have been performed on the SuperMUC supercomputer at the Leibniz-Rechenzentrum (LRZ) in Munich, using the computing resources awarded to the PRACE project number 2012060963. The authors want to thank V. Springel for providing us with the optimised version of GADGET-2.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University and Yale University.

SRT thanks Sylvie Adenis for her help improving the grammar and the style of this paper.

REFERENCES

Alam S. et al., 2016, MNRAS, preprint (arXiv:1607.03155)
Alonso D., 2012, preprint (arXiv:1210.1833)
Baldauf T., Smith R. E., Seljak U., Mandelbaum R., 2010, Phys. Rev. D, 81, 063531

[3] http://projects.ift.uam-csic.es/skies-universes

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bovy J. et al., 2011, ApJ, 729, 141
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Busca N. G. et al., 2013, A&A, 552, A96
Chehade B. et al., 2016, MNRAS, 459, 1179
Cole S. et al., 2005, MNRAS, 362, 505
Comparat J., Prada F., Yepes G., Klypin A., 2017, preprint (arXiv: 1702.01628)
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
Dekel A., Lahav O., 1999, ApJ, 520, 24
Delubac T. et al., 2015, A&A, 574, A59
Eftekharzadeh S. et al., 2015, MNRAS, 453, 2779
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J. et al., 2011, AJ, 142, 72
Favole G. et al., 2016, MNRAS, 461, 3421
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Font-Ribera A. et al., 2014, J. Cosmology Astropart. Phys., 5, 027
Gunn J. E. et al., 2006, AJ, 131, 2332
Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111
Guo H., Zehavi I., Zheng Z., 2012, ApJ, 756, 127
Guo H. et al., 2014, MNRAS, 441, 2398
Guo H. et al., 2015, MNRAS, 453, 4368
Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodriguez-Torres S., 2017, MNRAS, 467, 1940
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, 1
Klypin A., Prada F., 2017, preprint (arXiv:1701.05690)
Klypin A., Yepes G., Gottlober S., Prada F., Hess S., 2016, MNRAS, 457, 4340
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Myers A. D. et al., 2015, ApJS, 221, 27
Norberg P. et al., 2001, MNRAS, 328, 64
Nuza S. E. et al., 2013, MNRAS, 432, 743
Palanque-Delabrouille N. et al., 2016, A&A, 587, A41
Pâris I. et al., 2014, A&A, 563, A54
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Planck Collaboration XVI, 2014, A&A, 571, A16
Porciani C., Norberg P., 2006, MNRAS, 371, 1824
Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, MNRAS, 452, 2034
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Reid B. A., White M., 2011, MNRAS, 417, 1913
Richardson J., Zheng Z., Chatterjee S., Nagai D., Shen Y., 2012, ApJ, 755, 30
Rodriguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893
Rodríguez-Torres S. A. et al., 2016, MNRAS, 460, 1173
Ross N. P. et al., 2009, ApJ, 697, 1634
Ross A. J. et al., 2012, MNRAS, 424, 564
Schneider D. P. et al., 2010, AJ, 139, 2360
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Shen Y. et al., 2007, AJ, 133, 2222
Shen Y. et al., 2013, ApJ, 778, 98
Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575
Smee S. A. et al., 2013, AJ, 146, 32
Springel V., 2005, MNRAS, 364, 1105
Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16
White M. et al., 2012, MNRAS, 424, 933
Wright E. L. et al., 2010, AJ, 140, 1868
York D. G. et al., 2000, AJ, 120, 1579
Zheng Z. et al., 2005, ApJ, 633, 791

APPENDIX A: SIMULATION RESOLUTION

In order to reproduce the observed clustering of QSO or ELG samples, simulations with large volume and a high resolution are needed to resolve haloes of masses $\sim 10^{12.5}\,M_{\odot}$. The Y1Q sample covers $\sim$1100 deg^2 of the sky. This area is comparable to the BigMDPL-QSO light-cone. However, a small part of the halo mass range occupied by quasars can be in the incomplete part of the simulation.

We use the 1 h^{-1} Gpc MDPL simulation to quantify the effect of incompleteness of the BigMDPL light-cone. We select two snapshots from each simulation with similar redshift (Table A1). We apply the model using the parameters of Table 3. Table A1 presents a comparison between both simulations. In terms of halo mass, mocks constructed with both simulations provide consistent mean halo masses. Similar results are found for the satellite fraction.

In terms of clustering, both simulations give coherent results with differences of the order of 3 per cent. Fig. A1 shows the difference on the monopole between both simulations. These discrepancies are not a problem for our analysis, where errors from the data are of the order of 15 per cent.

In addition to the large errors in the data, discrepancies between both boxes seem reasonable if we notice the other sources of error.

(i) Both simulations have different initial conditions, this includes variations due to the cosmic variance between simulations.

Table A1. Comparison of the halo mass of mocks constructed with the BigMDPL and MDPL simulations. For comparison, all snapshots of the BigMDPL simulation in the redshift range $0.9 < z < 2.2$ were used. We select snapshots with the nearest redshift from the MDPL simulation.

Box	z	$\log_{10}[M/M_{\odot}]$	$V_{\rm mean}$	$f_{\rm sat}$
MDPL	0.987	12.41	284.25	0.08
	1.425	12.54	325.95	0.07
BigMDPL	1.000	12.40	284.25	0.11
	1.445	12.55	325.95	0.07

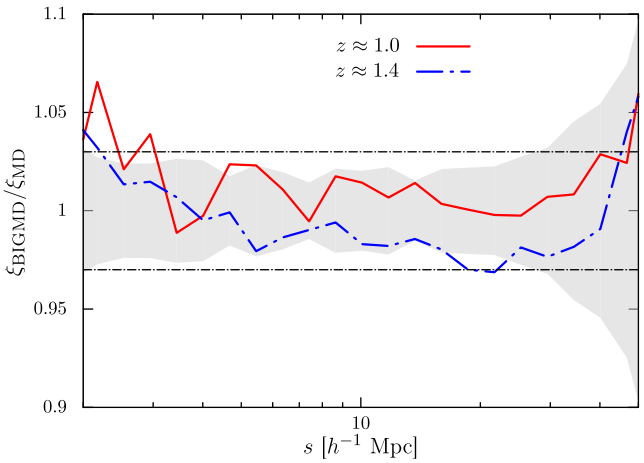

Figure A1. Ratio between BigMDPL and MDPL mocks of the monopole of the correlation function in configuration space. The horizontal lines represent 3 per cent differences. The shaded area shows 1σ dispersion due to the random selection in the MDPL boxes. We use 15 realizations to compute the shaded area.

(ii) The shot noise in the correlation function is larger in the MDPL simulation due to the smaller volume.

(iii) The random selection of our model is another source of errors. The shaded area in Fig. A1 represents the 1σ dispersion of 15 mocks produced with different seeds.

(iv) The BigMDPL simulation includes long waves that are not included in the 1 h^{-1} Gpc box size.

APPENDIX B: EFFECTS OF OBSERVATIONAL ERRORS ON THE CLUSTERING

The model presented in this work includes two observational errors: catastrophic redshift errors and redshift errors. The first errors cause a constant reduction in the clustering amplitude at all the scales. Fig. B1 shows the effect of applying 1 per cent of catastrophic redshifts. We find a reduction of ∼1 per cent in all scales of the correlation function in configuration space.

Redshift errors have the strongest impact on the clustering. The selection of QSO implies fixing maximum width (precision) to identify the emission/absorption features of the spectra. We introduce the effect of this tolerance using Gaussian errors with a width given by Dawson et al. (2016). Redshift errors have an important impact at scales $<10\,h^{-1}$ Mpc. In Fig. B1, it is possible to see a disagreement larger than 40 per cent, which cannot be explained by statistical errors of the sample (shaded area in Fig. B1).

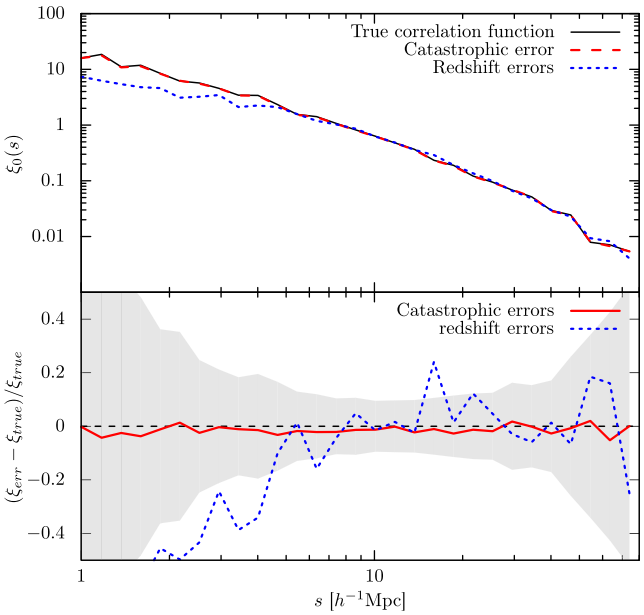

Figure B1. Top panel: impact of catastrophic redshift errors and redshift errors on the monopole of the correlation function. A light-cone reproducing the Y1Q 1-point and two-point statistics is used for this comparison. Bottom panel: normalized differences between mocks including redshift errors (blue dotted line) and catastrophic redshift errors (red line) with a model without errors. The shaded area represents the statistical errors in the light-cone computed from 1000 GLAM catalogues. Differences due to catastrophic redshift errors are ∼1 per cent. Redshift errors have an important impact at small scales which cannot be explained by uncertainties from mocks.

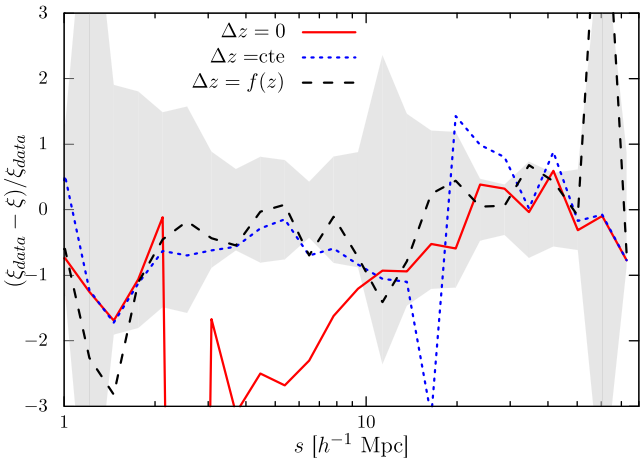

Figure B2. Impact of redshift errors in the quadrupole of the correlation function in configuration space. Lines show the normalized difference between observed data and model without redshift errors (red solid line), constant redshift error $\Delta z = 0.005$ (blue dotted line) and including redshift errors given by equation (2) (black dashed line). Shaded area represent 1σ error computed with 1000 GLAM catalogues for one light-cone.

The impact of redshift error is very important in the monopole of the correlation function. However, the effects on the quadrupole are larger. Fig. B2 shows the ratio of quadrupole from the observed data and the different mocks. The model introduced in this work describes the very large difference found between our mock and the observed data.

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