Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **460,** 1173–1187 (2016)
Advance Access publication 2016 April 29

doi:10.1093/mnras/stw1014

The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release

Sergio A. Rodríguez-Torres,[1,2,3★†] Chia-Hsun Chuang,[1,4‡] Francisco Prada,[1,2,5,6] Hong Guo,[7,8] Anatoly Klypin,[9,10] Peter Behroozi,[11] Chang Hoon Hahn,[12] Johan Comparat,[1,3] Gustavo Yepes,[3] Antonio D. Montero-Dorta,[8] Joel R. Brownstein,[8] Claudia Maraston,[13] Cameron K. McBride,[14] Jeremy Tinker,[12] Stefan Gottlöber,[4] Ginevra Favole,[1,2] Yiping Shu,[8] Francisco-Shu Kitaura,[4] Adam Bolton,[8] Román Scoccimarro,[12] Lado Samushia,[13,15,16] David Schlegel,[5] Donald P. Schneider[17,18] and Daniel Thomas[13]

Affiliations are listed at the end of the paper

Accepted 2016 April 27. Received 2016 April 27; in original form 2015 September 29

ABSTRACT

We present a study of the clustering and halo occupation distribution of Baryon Oscillation Spectroscopic Survey (BOSS) CMASS galaxies in the redshift range $0.43 < z < 0.7$ drawn from the Final SDSS-III Data Release. We compare the BOSS results with the predictions of a halo abundance matching (HAM) clustering model that assigns galaxies to dark matter haloes selected from the large BigMultiDark *N*-body simulation of a flat Λ cold dark matter Planck cosmology. We compare the observational data with the simulated ones on a light cone constructed from 20 subsequent outputs of the simulation. Observational effects such as incompleteness, geometry, veto masks and fibre collisions are included in the model, which reproduces within 1σ errors the observed monopole of the two-point correlation function at all relevant scales: from the smallest scales, $0.5\,h^{-1}$ Mpc, up to scales beyond the baryon acoustic oscillation feature. This model also agrees remarkably well with the BOSS galaxy power spectrum (up to $k \sim 1\ h\ \mathrm{Mpc}^{-1}$), and the three-point correlation function. The quadrupole of the correlation function presents some tensions with observations. We discuss possible causes that can explain this disagreement, including target selection effects. Overall, the standard HAM model describes remarkably well the clustering statistics of the CMASS sample. We compare the stellar-to-halo mass relation for the CMASS sample measured using weak lensing in the Canada–France–Hawaii Telescope Stripe 82 Survey with the prediction of our clustering model, and find a good agreement within 1σ. The BigMD-BOSS light cone including properties of BOSS galaxies and halo properties is made publicly available.

Key words: methods: numerical – galaxies: abundances – galaxies: haloes – large-scale structure of Universe.

1 INTRODUCTION

One of the major goals in cosmology is to explain the formation of the large-scale structure (LSS) of the Universe. However, the main ingredient that drives this evolution – the dark matter – can only be probed using the distribution of galaxies, and galaxies are biased tracers of the matter field. This makes this study challenging. In the last 20 years, vast amounts of observational data have been obtained, improving each time the precision of the LSS measurements and demanding ever more accurate theoretical models. In fact, one of the strongest arguments that we understand how the LSS forms and evolves is our ability to reproduce the galaxy clustering through cosmic time, starting from the primordial Gaussian perturbations.

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

During the last decade, surveys such as the Sloan Digital Sky Survey (SDSS-I/II/III; York et al. 2000; Eisenstein et al. 2011) have made it possible to determine the clustering of galaxy populations at scales out to tens of Mpc and beyond with reasonable accuracy.

The Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) Data Release 12 (DR12; Alam et al. 2015) provides redshift of 1.5 million massive galaxies in 10 000 deg^2 area of the sky and for redshifts in the range 0.15–0.75. BOSS DR12 has an effective volume seven times larger than that of the SDSS-I/II project. These data provide us with a sufficiently statistical sample to examine our theoretical predictions over a range of scales.

In order to compare the Λ cold dark matter (ΛCDM) model and the observational data, it is necessary to link the galaxy and the dark matter distributions. There are a number of methods to assign galaxies to the dark matter. State-of-the-art hydrodynamical simulations, which include detailed galaxy formation descriptions, are computationally unaffordable for the volumes considered here (e.g. Vogelsberger et al. 2014; Schaye et al. 2015), and indeed, there are no large samples of simulated galaxies that can be used to match BOSS. Semi-analytic models are less computationally consuming methods to populate dark matter haloes with galaxies (e.g. Knebe et al. 2015). These models incorporate some physics of galaxy formation.

The most popular models are based on the statistical relations between galaxies and dark matter haloes. One of the most used models is the halo occupation distribution (HOD; e.g. Jing, Mo & Börner 1998; Peacock & Smith 2000; Berlind & Weinberg 2002; Zheng et al. 2005; Leauthaud et al. 2012; Guo et al. 2014). The main component of the HOD is the probability, $P(N|M_{\rm halo})$, that a halo of virial mass $M_{\rm halo}$ hosts N galaxies with some specified properties. These models have several parameters which allow one to match the observed clustering.

The model known as the halo abundance matching (HAM; 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) connects observed galaxies to simulated dark matter haloes and subhaloes by requiring a correspondence between the luminosity or stellar mass and a halo property. The assumption of this model is that more luminous (massive) galaxies are hosted by more massive haloes. However, this relation is not a one-to-one relation because there is a physically motivated scatter between galaxies and dark matter haloes (e.g. Shu et al. 2012). By construction, the method reproduces the observed luminosity function, LF (or stellar mass function, SMF). HAM relates the LF (SMF) of an observed sample with the distribution of haloes in an N-body simulation. The implemented assignment requires that one works with complete samples in luminosity (stellar mass) or have precise knowledge of the incompleteness as a function of the luminosity (stellar mass) of the galaxy sample. Luminous red galaxies (LRGs) are the most massive galaxies in the Universe, and they represent the high-mass end of the SMF. This feature makes this population of galaxies an excellent group to be reproduced with the abundance matching.

In this paper, we compare the clustering of the BOSS CMASS DR12 sample with predictions from N-body simulations. We use an abundance matching to populate the dark matter haloes of the BigMultiDark Planck simulation (BigMDPL; Klypin et al. 2016). In order to include systematic effects from the survey, as well as the proper evolution of the clustering, we construct light cones which reproduce the angular selection function, the radial selection function and the clustering of the monopole in configuration space. To generate these catalogues, we developed the SUrvey GenerAtoR (SUGAR) code. Once the HAM and the light cone are applied, we compute the predictions of our model for two-point statistics and the three-point correlation function (3PCF). We also present the prediction of the stellar-to-halo mass relation and its intrinsic scatter compared to lensing measurements. The HAM, the BigMDPL and the methodology to produce light cone played a key role in the construction of the MultiDark PATCHY BOSS DR12 mocks (MD-PATCHY mocks; Kitaura et al. 2016, companion paper).

In order to have a good estimation of the uncertainties in this work, we use 100 MD-PATCHY mocks. These mocks are produced using five boxes at different redshifts that are created with the PATCHY code (Kitaura, Yepes & Prada 2014). The PATCHY code can be decomposed into two parts: (1) computing approximate dark matter density field and (2) populating galaxies from dark matter density field with the biasing model. The dark matter density field is estimated using augmented Lagrangian perturbation theory (Kitaura & Heß 2013) which combines the second-order perturbation theory (see e.g. Buchert 1994; Bouchet et al. 1995; Catelan 1995) and spherical collapse approximation (see Bernardeau 1994; Mohayaee et al. 2006; Neyrinck 2013). The biasing model includes deterministic bias and stochastic bias (for details see Kitaura et al. 2014). The velocity field is constructed based on the displacement field of dark matter particles. The modelling of finger-of-god has also been taken into account statistically. The MD-PATCHY mocks are constructed based on the BigMD simulation with the same cosmology used in this work. The mocks match the clustering of the galaxy catalogues for each redshift bin (see Kitaura et al. 2016, companion paper, for details). The BigMultiDark light-cone catalogues of BOSS CMASS galaxies in the Final DR12 (hereafter BigMD-BOSS light cone) presented in this work are publicly available.

This paper is structured as follows: Sections 2 and 3 describe the SDSS-III/BOSS CMASS galaxy sample and the BigMDPL N-body cosmological simulations used in this work. In Section 4, we provide details on different observational effects and briefly describe the SUGAR code. Section 4.1 presents the main ingredients of the HAM modelling of the CMASS galaxy clustering. A comparison of our results to observation is shown in Section 5. Subsequently, we discuss the principal results in Section 6. Finally, in Section 7, we present a summary of our work. For all results in this work, we use the cosmological parameters $\Omega_{\rm m} = 0.307$, $\Omega_{\rm B} = 0.048$, $\Omega_\Lambda = 0.693$.

2 SDSS-III/BOSS CMASS SAMPLE

The Baryon Oscillation Spectroscopic Survey[1] (BOSS; Bolton et al. 2012; Dawson et al. 2013) is part of the SDSS-III programme (Eisenstein et al. 2011). The project used the 2.5 m aperture Sloan Foundation Telescope at Apache Point Observatory (Gunn et al. 2006). The telescope used a drift-scanning mosaic CCD camera (Gunn et al. 1998) with five colour bands, u, g, r, i, z (Fukugita et al. 1996). Spectra are obtained using the double-armed BOSS spectrographs, which are significantly upgraded from those used by SDSS I/II, covering the wavelength range 3600–10 000 Å with a resolving power of 1500–2600 (Smee et al. 2013). BOSS provides redshift for 1.5 million galaxies in 10 000 deg^2 divided into two samples: LOWZ and CMASS. The LOWZ galaxies are selected to be the brightest and reddest of the low-redshift galaxy population ($z \lesssim 0.4$), extending the SDSS I/II LRGs. The CMASS target

[1] http://skyserver.sdss.org/dr12/en/home.aspx

selection is designed to isolate galaxies at higher redshift ($z \gtrsim 0.4$), most of them being also LRGs.

In the present paper, we focus on the CMASS DR12 North Galactic Cap (NGC) sample. Galaxies are selected from SDSS DR8 imaging (Aihara et al. 2011) according to a series of colour cuts designed to obtain a sample with approximately 'constant stellar mass' (Reid et al. 2016). The following photometric cuts are applied:

$$17.5 < i_{\rm cmod} < 19.9 \quad (1)$$

$$r_{\rm mod} - i_{\rm mod} < 2 \quad (2)$$

$$d_{\perp} > 0.55 \quad (3)$$

$$i_{\rm fib2} < 21.5 \quad (4)$$

$$i_{\rm cmod} < 19.86 + 1.6(d_{\perp} - 0.8), \quad (5)$$

where i and r indicate magnitudes, and $i_{\rm fib2}$ is the i-band magnitude within a 2 arcsec aperture. All magnitudes are corrected for Galactic extinction [via the Schlegel et al. (1998) dust maps]. The subscript 'mod' denotes the 'model' magnitudes and the subscript 'cmod' refers to the 'cmodel' magnitudes. The model magnitudes represent the best fit of the DeVaucouleurs and exponential profile in the r band (Stoughton et al. 2002) and the cmodel magnitudes denote the best-fitting linear combination of the exponential and DeVaucouleurs models (Abazajian et al. 2004). $d_{\perp}$ is defined as

$$d_{\perp} = r_{\rm mod} - i_{\rm mod} - (g_{\rm mod} - r_{\rm mod})/8.0. \quad (6)$$

Star–galaxy separation is performed on the CMASS targets via

$$i_{\rm psf} - i_{\rm mod} > 0.2 + 0.2(20.0 - i_{\rm mod}) \quad (7)$$

$$z_{\rm psf} - z_{\rm mod} > 9.125 - 0.46 z_{\rm mod}. \quad (8)$$

The subscript 'psf' refers to point spread function magnitudes. CMASS sample contains galaxies with redshift $z > 0.4$, having the peak of the number density at $z \approx 0.5$. We will concentrate our analysis in the redshift range $0.43 < z < 0.7$ for this sample.

BOSS sample is corrected for redshift failures and fibre collisions. In the following sections, we will use the same weights given in Anderson et al. (2014) in order to correct the clustering signal affected by these systematics (Ross et al. 2012). The total weight for a galaxy is given by

$$w_{\rm g} = w_{\rm star} w_{\rm see} (w_{\rm zf} + w_{\rm cp} - 1). \quad (9)$$

In this equation, $w_{\rm zf}$ denotes the redshift failure weight and $w_{\rm cp}$ represents the close pair weight. Both quantities start with unit weight. If a galaxy has a nearest neighbour (of the same target class) with a redshift failure ($w_{\rm zf}$) or its redshift was not obtained because it was in a close pair ($w_{\rm cp}$), we increase $w_{\rm zf}$ or $w_{\rm cp}$ by one. As found in Ross et al. (2012), the impact of this effect is very small for the CMASS sample; for this reason, we do not model the redshift failures in this study. For CMASS, additional weights are applied to account for the observed systematic relationships between the number density of observed galaxies and stellar density and seeing (weights $w_{\rm star}$ and $w_{\rm see}$, respectively).

3 BIGMULTIDARK SIMULATION

The BigMDPL is one of the MultiDark[2] N-body simulation described in Klypin et al. (2016). The BigMDPL was performed with GADGET-2 code (Springel 2005). This simulation was created in a box of 2.5 h^{-1} Gpc on a side, with 3840^3 dark matter particles. The mass resolution is $2.4 \times 10^{10}\ h^{-1}\ {\rm M}_{\odot}$. The initial conditions, based on initial Gaussian fluctuations, are generated with Zeldovich approximation at $z_{\rm init} = 100$. The suite of BigMultiDark is constituted of four simulations with different sets of cosmological parameters. In this study, we adopt a flat ΛCDM model with the Planck cosmological parameters: $\Omega_{\rm m} = 0.307$, $\Omega_{\rm B} = 0.048$, $\Omega_{\Lambda} = 0.693$, $\sigma_8 = 0.829$, $n_{\rm s} = 0.96$ and a dimensionless Hubble parameter $h = 0.678$ (Klypin et al. 2016). The simulation provides 20 redshift outputs (snapshots) within the redshift range $0.43 < z < 0.7$.

For the present analysis, we use the ROCKSTAR (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement) halo finder (Behroozi, Wechsler & Wu 2013a). 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 and one time dimension. ROCKSTAR computes halo mass using spherical overdensities of a virial structure. Before calculating halo masses and circular velocities, the halo finder performs a procedure which removes unbound particles from the final mass of the halo. ROCKSTAR creates particle-based merger trees. The merger trees algorithm (Behroozi et al. 2013b) was used to estimate the peak circular velocity over the history of the halo, $V_{\rm peak}$, which we use to perform the abundance matching.

4 METHODOLOGY: THE SUGAR CODE

We construct light-cone catalogues from the BigMDPL simulation which reproduce the clustering measured in the monopole of the redshift-space correlation function from the BOSS CMASS DR12 sample. For this purpose, we developed the SUGAR code which implements the HAM technique to generate galaxy catalogues from a dark matter simulation. The code can apply the geometric features of the survey and selection effects, including stellar mass incompleteness and fibre collision effects. All the available outputs (snapshots) of the BigMDPL simulation are used, so that the light cone has the proper evolution of the clustering.

In the following subsections, we present the ingredients used to produce the BigMD-BOSS light cone, which is shown in Figs 1 and 2. We present the HAM method and the SMF adopted in this work. The light-cone production, the fibre collision assignment and the modelling of the stellar mass incompleteness are also shown.

4.1 HAM procedure

We use a HAM technique to populate dark matter haloes with galaxies (see e.g. Nuza et al. 2013). This physically motivated method produces mock galaxy catalogues that in the past gave good representations of large galaxy samples (see for SDSS, e.g. Trujillo-Gomez et al. 2011; Reddick et al. 2013). The basic assumption of this method is that massive haloes host massive galaxies. This allows one to generate a rank-ordered relation between dark matter haloes and galaxies. However, observations show that this assignment cannot be a one-to-one relation (Shu et al. 2012). In order to create a more realistic approach, it is necessary to include scatter in

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

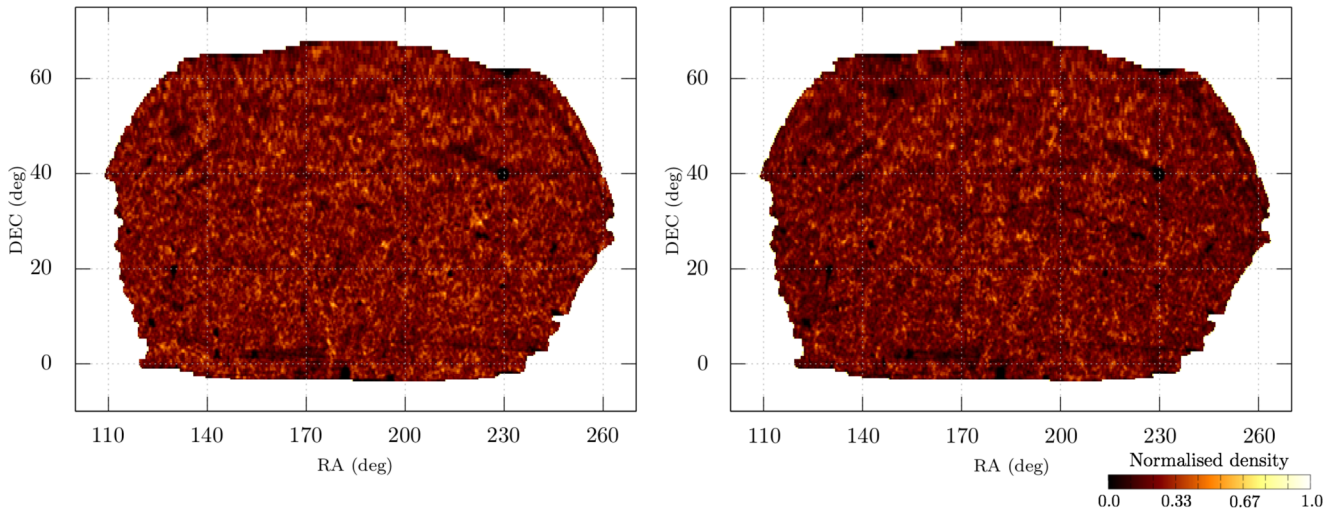

Figure 1. Left-hand panel: sky area covered by the BigMD-BOSS light cone. This region includes the BOSS CMASS DR12 geometry and veto masks. Right-hand panel: sky area covered by the BOSS CMASS DR12 sample. Colours indicate the angular number density, which is normalized by the most dense pixel. Each pixel has an angular area of 1 deg^2. BigMD-BOSS light cone uses the same mask as the BOSS CMASS DR12, including angular completeness and veto masks.

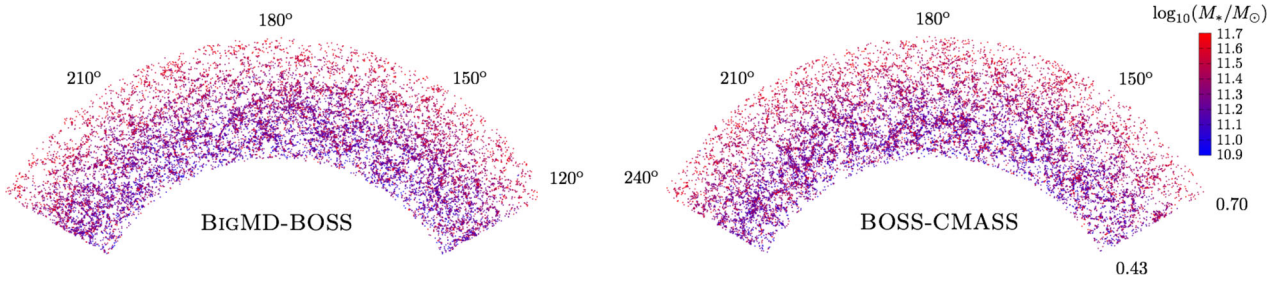

Figure 2. Pie plot of the BigMD-BOSS light cone (left-hand panel) and the BOSS CMASS DR12 data (right-hand panel). Both figures were made with 2 deg of thickness (Dec. coordinate).

this matching. The HAM can relate galaxy luminosities or stellar mass from galaxies to a halo property. In this paper, we use the peak value of the circular velocity over the history of the halo ($V_{\rm peak}$), which has advantages compared to the halo mass ($M_{\rm halo}$). $M_{\rm halo}$ is well defined for host haloes, but its definition becomes ambiguous for subhaloes. The subhalo mass also depends on the halo finder used (Trujillo-Gomez et al. 2011; Reddick et al. 2013). In addition to $M_{\rm halo}$ and $V_{\rm peak}$, HAM can be performed using other quantities such as the maximum circular velocity of the halo ($V_{\rm max}$), the maximum circular velocity of the halo at time of accretion ($V_{\rm acc}$) or the halo mass at time of accretion ($M_{\rm acc}$). Other studies present the effect of the halo property in the HAM (e.g. Reddick et al. 2013; Guo et al. 2015a).

We adopt a modified version of the scatter proposed in Nuza et al. (2013). Our implementation of the abundance matching can be briefly summarized in the following steps.

(i) For the dark matter haloes, we define a scattered $V_{\rm peak}$, which is used only to assign stellar mass to the haloes. This scattered quantity is defined by

$$V_{\rm peak}^{\rm scat} = (1 + \mathcal{N}(0, \sigma_{\rm HAM}))V_{\rm peak}, \tag{10}$$

where $\mathcal{N}$ is a random number, produced from a Gaussian distribution with mean 0 and standard deviation $\sigma_{\rm HAM}(V_{\rm peak}|M_*)$.

(ii) Sort the catalogue by $V_{\rm peak}^{\rm scat}$, starting from the object with the largest velocity and continuing down until reaching all the available objects. Use this catalogue to construct the cumulative number density of the haloes as a function of $V_{\rm peak}^{\rm scat}$.

(iii) Compute the cumulative number density of galaxies as a function of the stellar mass using the adopted SMF (see Section 4.2).

(iv) Finally, construct a monotonic relation between the cumulative number density functions from steps (ii) and (iii) such as

$$n_{\rm gal}(> M_*^i) = n_{\rm halo}(> V_{{\rm peak},i}^{\rm scat}). \tag{11}$$

This relation implies that a halo with $V_{{\rm peak},i}^{\rm scat}$ will contain a galaxy with stellar mass M_*^i.

This assignment is monotonic between $V_{\rm peak}^{\rm scat}$ and M_*, but not between $V_{\rm peak}$ and M_*. The relation of these two quantities is mediated by the scatter parameter, $\sigma_{\rm HAM}(V_{\rm peak}|M_*)$.

4.2 Stellar mass function

We employ the Portsmouth SED-fit DR12 stellar mass catalogue (Maraston et al. 2013) with the Kroupa initial mass function (Kroupa 2001) to estimate the SMF. The CMASS LSS catalogue does not include the stellar mass information. For that reason, we matched the BOSS and the LEGACY stellar mass catalogues with the LSS BOSS CMASS catalogue. In order to identify an SDSS spectrum in the different catalogues, there are three numbers that determine each galaxy: PLATE, MJD and FIBERID. We use these three quantities

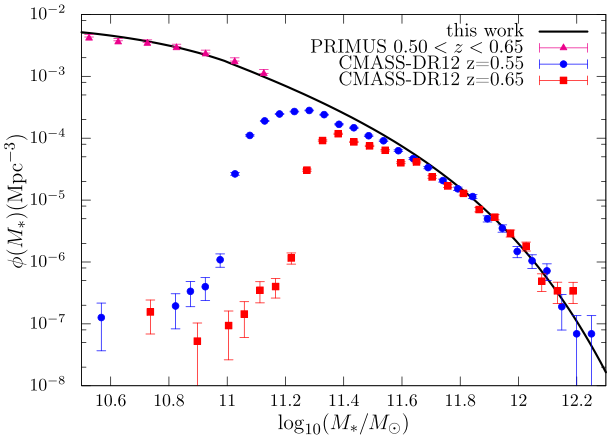

Figure 3. SMF from BOSS CMASS DR12 sample. Circles and squares show the stellar mass distribution for two redshift bins from the Portsmouth DR12 catalogue. Poissonian errors are included. The solid line shows the estimate of the SMF for this work, which is constructed combining the high-mass end of the BOSS sample and Guo et al. (2010) for the low-mass range ($\log_{10} M_* < 11.0$). In order to compare with a complete sample in the redshift range 0.5–0.65, we include the PRIMUS SMF (triangles) in the low-mass regime.

to match the stellar mass catalogues (LEGACY and BOSS) and the LSS BOSS CMASS catalogue. Once the stellar masses of the observed sample are assigned, we need to construct an SMF which describes the mass distribution.

The Portsmouth DR12 catalogue has the SMF that is different from SMF of previous surveys (Maraston et al. 2013). Fig. 3 shows the mass distribution of the CMASS DR12 for two different redshift regions. A detailed study of the Portsmouth catalogues and other stellar mass catalogues was reported by Maraston et al. (2013).

Due to the selection function in the BOSS data, we do not have the information on the shape of the SMF at low masses. There are different ways of handling this problem. For example, Leauthaud et al. (2016) use the stripe 82 massive galaxy catalogue to compute the SMF of the BOSS data. We use a different approach: for the high-mass end, we use the Portsmouth stellar masses and we combine them with Guo et al. (2010) results to describe the low-mass regime. Specifically, to compute the SMF for masses larger than $3.2 \times 10^{10}\,\mathrm{M}_\odot$ (which is the mass range used in the CMASS sample).

In order to construct the SMF, we select galaxies in the redshift range $0.55 < z < 0.65$, because this is the most complete range for the CMASS sample (see Montero-Dorta et al. 2014). We combine the CMASS sample for masses larger than $2.5 \times 10^{11}\,\mathrm{M}_\odot$ and the SMF from Guo et al. (2010) for low masses. We fit both results using a double Press–Schechter mass function (Press & Schechter 1974) with the parameters given in Table 1.

Fig. 3 presents the SMF used in this work. We also add in Fig. 3 the PRIMUS SMF (Moustakas et al. 2013) in the redshift range $0.5 < z < 0.65$ with the purpose of comparing the low-mass range of our SMF with a complete sample in the same redshift and mass ranges. A detailed comparison of the Portsmouth catalogues and other stellar mass catalogues is presented in Maraston et al. (2013).

In our analysis, we do not include redshift evolution of the SMF. This approximation agrees with results of the PRIMUS survey (Moustakas et al. 2013), which is a complete survey in the redshift range we study. Moustakas et al. (2013) show that there is only a small evolution of the SMF in the CMASS redshift range.

Table 1. Parameters of the double Press–Schechter SMF for this work.

Mass range ($\mathrm{M}_\odot$)	ϕ_* ($\mathrm{Mpc}^3 \log_{10} \mathrm{M}_\odot^{-1}$)	α	$\log_{10} M_*$ ($\mathrm{M}_\odot$)
$\log_{10} M_* \leq 11.00$	4.002×10^{-3}	−0.938	10.76
$\log_{10} M_* > 11.00$	2.663×10^{-4}	−2.447	11.42

4.3 Production of light-cones

We implement a method to generate light cones from snapshots of cosmological simulations. This method has been implemented previously (see e.g. Blaizot et al. 2005; Kitzbichler & White 2007). The SUGAR code works with cubic boxes using positions and velocities of dark matter haloes as inputs. We will now describe the procedure which we use to construct mocks for the CMASS sample.

BigMD-BOSS light cones are constructed from the BigMDPL simulation which is large enough ($2.5\,h^{-1}$ Gpc) to map the CMASS NGC. We use the periodic boundary conditions to maximize the use of the volume (Manera et al. 2013), but we do not reuse any region of the box. So there are no duplicated structures in our light cone.

The first step in the construction of the light cone is to locate the observer ($z = 0$) and transform from comoving Cartesian coordinates to equatorial coordinates (RA,Dec.) and redshift. To include the effects of galaxy peculiar velocities in the redshift measurements, we transform the coordinates of the haloes to redshift space using

$$\boldsymbol{s} = \boldsymbol{r}_\mathrm{c} + \frac{\boldsymbol{v} \cdot \hat{\boldsymbol{r}}}{aH(z_\mathrm{real})}, \tag{12}$$

where $\boldsymbol{r}_\mathrm{c}$ is the comoving distance in real space, $\boldsymbol{v}$ is the velocity of the object with respect to Hubble flow, $\hat{\boldsymbol{r}}$ is the line-of-sight direction, a is the scale factor and H is the Hubble constant at z_real, which is the redshift corresponding to r_c, and is computed from

$$r_\mathrm{c}(z_\mathrm{real}) = \int_0^{z_\mathrm{real}} \frac{c\,\mathrm{d}z}{H_0\sqrt{\Omega_\mathrm{m}(1+z)^3 + \Omega_\Lambda}}, \tag{13}$$

where c is light speed and H_0 is the Hubble constant in s^{-1} Mpc^{-1} km. Using equations (12) and (13), it is possible to compute $s(z_\mathrm{obs})$, where z_obs is the observed redshift. The next step is to select objects from each snapshot to construct shells for the light cone. Thus, an object with redshift z_obs, which comes from a snapshot at $z = z_i$, will be selected if $(z_i + z_{i-1})/2 < z_\mathrm{obs} \leq (z_i + z_{i+1})/2$. We repeat this process for all objects in snapshots between $z = 0.43$ and 0.7. We fix the number density in each shell following the radial selection function of the BOSS CMASS sample. Fig. 4 shows the comparison between the radial selection function of the observed data and the one obtained on the BigMD-BOSS light cone.

Finally, we apply the angular CMASS NGC mask to match the area of the observed sample. The angular completeness is taken into account by downsampling the regions where it is smaller than one. As was done in the BOSS CMASS catalogue, we select regions in the sky with completeness weight larger than 0.7. Due to the presence of random numbers in the selection process, the observed radial selection function can have variations of ∼4 per cent. Fig. 4 presents the standard deviation from 100 MD-PATCHY mocks to examine the effect of different seeds on the random generator.

Fig. 1 shows the angular distribution of the BigMD-BOSS light cone. In order to reproduce the angular distribution, we applied the BOSS CMASS DR12 NGC geometry, and, in addition, we applied veto mask to exclude exactly the same regions removed in the observed data. Fig. 2 presents a 2D comparison of the spatial

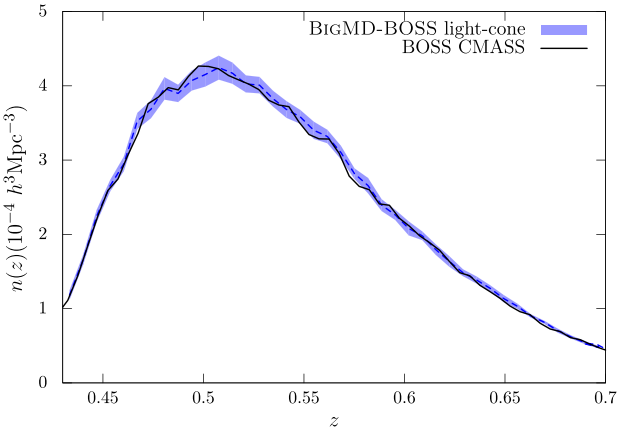

Figure 4. The comoving number density of BOSS CMASS DR12 NGC (black line) compared to the comoving number density of the BigMD-BOSS light cone (dashed line). Shaded area comes from 100 MD-PATCHY mocks.

galaxy distribution between the BigMD-BOSS light cone and the BOSS CMASS data.

4.4 Stellar mass incompleteness

This paper focuses in the production of mocks which can describe the full CMASS DR12 sample. Instead of extracting a subsample which has better completeness in terms of stellar mass, we 'model' the observed stellar mass incompleteness. This model not only accounts for the incompleteness at small masses (presented across the complete redshift range), but also incompleteness in the high-mass end, which is important for $z \lesssim 0.45$. Fig. 5 compares the results of our modelling in the BigMD-BOSS light cone to the observed data for three different redshifts.

In order to reproduce the observed stellar mass distribution, we construct a continuous function by interpolation. Once the abundance matching is applied and galaxies are assigned to dark matter haloes, we select galaxies by downsampling based on the observed stellar mass distribution. This process is repeated for 20 different redshifts (corresponding to the snapshots of the simulation). Then, in order to construct the observed stellar mass distribution corresponding to snapshot at $z = z_i$, a galaxy with redshift z_g in the stellar mass catalogue will be selected if $(z_i + z_{i-1})/2 < z_g \leq (z_i + z_{i+1})/2$. This model has an important impact on the scatter applied to the abundance matching. Since bias is as a function of stellar mass, incompleteness that varies as a function of stellar mass will affect the overall bias as well. This effect reduces the amplitude of the clustering, which implies that a smaller scatter is required to reproduce the signal of the observed clustering. If we ignore the incompleteness effect, we can still reproduce the clustering in the two-point correlation function (2PCF). However, this scatter is not the intrinsic one, and the final stellar mass distribution will not match the observed sample. Favole et al. (2015a) show a similar model to reproduce the incompleteness of the Emission Line Galaxies population from the BOSS sample.

Most galaxies in the CMASS sample are red galaxies. However, there is also a fraction of blue galaxies in the data. In addition, the blue sample is less complete than the red one (Montero-Dorta et al. 2014).The random downsampling of galaxies in the BigMD-BOSS light cone does not distinguish between both populations, which can produce potential systematics due to the different completeness of both samples. In this study, we reproduce the observed stellar mass distribution by downsampling galaxies from a no-evolving SMF. However, SMF evolves with redshift, which can produce underestimation of the incompleteness for some ranges of stellar mass and overestimation for other ranges.

4.5 Fibre collisions

A feature of the BOSS fibre-fed spectrograph is that the finite size of the fibre housing makes impossible to place fibres within 62 arcsec of each other in the same plate. This causes a number of galaxies to not have a fibre assigned and hence, there is no measurement of their redshift. We model the effect of fibre collisions as follows. A total of 5 per cent of the CMASS targets could not been observed due to the fibre collisions. These objects have an important effect at scales $\lesssim 10\ h^{-1}$ Mpc. In this paper, we model the fibre collision effect by adopting the method described in Guo, Zehavi & Zheng (2012).

The first step is to find the maximum number of galaxies that could be assigned fibres. This decollided sample (D_1) is a set of galaxies which are not angularly collided with other galaxies in this subsample. The second population (D_2) are the potentially collided galaxies. Each galaxy in this subsample is within the fibre collision scale of a galaxy in population 1. We must determine from the observed sample the fraction of collided galaxies (D_2'') in the D_2

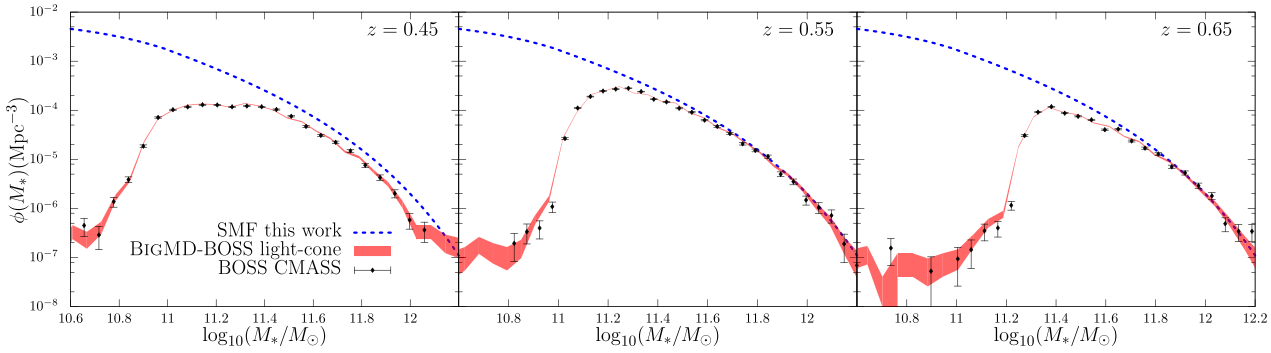

Figure 5. Incompleteness modelling for three different redshift bins. Shaded area shows the BigMD-BOSS light cone; dots are the measurements from the CMASS Portsmouth catalogue. In both cases, Poissonian errors are used. Dashed line represents the SMF adopted in this work. We select three bins as an example to show the results of the incompleteness modelling implemented in this work. Stellar mass distribution in the BigMD-BOSS light cone is produced by downsampling galaxies from the SMF adopted. Left-hand panel shows the incompleteness at low redshift in the high mass of the SMF.

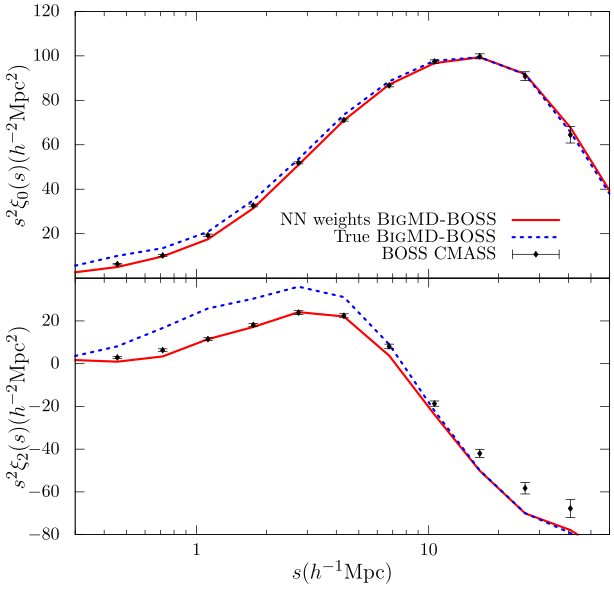

Figure 6. Monopole (top panel) and quadrupole (bottom panel) of the redshift-space correlation function for the BigMD-BOSS light cone before and after applying fibre collisions. Fibre collisions are corrected using nearest-neighbour (NN) weights. The effects of the fibre collisions are stronger in the quadrupole, with important differences for scales $s \lesssim 7\ h^{-1}$ Mpc. The impact on the monopole is smaller. The fibre collision assignment is an approximative method which can introduce systematic effects. In order to avoid these effects, we select the range 2–30 h^{-1} Mpc to fit the monopole with the scatter parameter, $\sigma_{\mathrm{HAM}}(V_{\mathrm{peak}}|M_*)$.

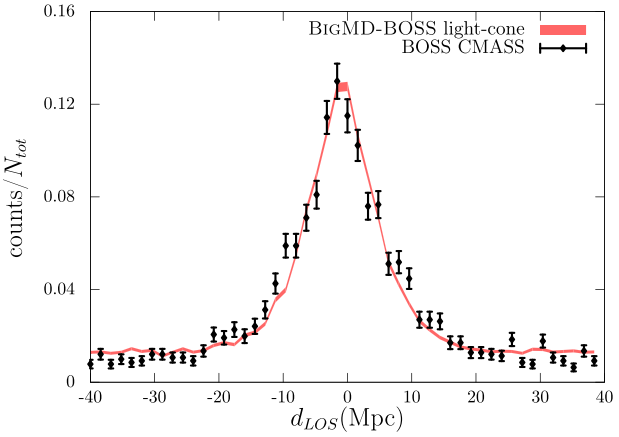

Figure 7. Line-of-sight displacement of a collided galaxy due to the fibre collision. The figure shows the number of counts per bin divided by the total number of collided galaxies. Uncertainties were computed using Poissonian errors.

group (i.e. D_2''/D_2) for sectors covered by different numbers of tiles. Finally, we randomly select the fraction D_2''/D_2 to the D_2 galaxies in the mocks to be collided galaxies.

Fig. 6 displays the impact of the fibre collisions on the correlation function in redshift space. The effect in the monopole becomes very important for scales smaller than 1 h^{-1} Mpc. However, the quadrupole is more sensitive to this effect, with big impact for scales smaller than 10 h^{-1} Mpc. The assignment of fibre collisions has an important impact on the fraction of satellites. Before fibre collisions the satellite fraction of the light cone is 11.8 per cent, and after the assignment is equal to 10.5 per cent. This effect reduces the central–satellite pairs, which have a strong impact on the quadrupole.

Unlike Guo et al. (2012), we only use nearest-neighbour weights for both samples. Our goal is to compare the results of the abundance matching with data, so that we implement the same fibre collision correction to our light cone as observed data.

When nearest-neighbour weights are applied, a collided galaxy will be 'moved' from its original coordinates to the position of its nearest neighbour. Fig. 7 presents the line-of-sight displacement of those collided galaxies from their original positions.

The displacement for the simulation shown in Fig. 7 is computed using the old and new positions of the collided galaxies. In CMASS data, the displacement is calculated using the overlapping tiled regions of the survey where the spectroscopic redshifts of both galaxies within the fibre collision angular scale are resolved. Fig. 7 demonstrates an excellent agreement between our model and the observed data, suggesting that the combination between the clustering at small scales of the simulation and the fibre collision model used in the mock has a reasonable agreement with observations.

5 MODELLING BOSS CMASS CLUSTERING

The clustering signal in the abundance matching is determined by two quantities: the number density and the scatter in the M_*-V_{peak} relation. The number density is fixed by the radial selection function of the observed sample. In order to find a scatter value that reproduces the clustering of the CMASS sample, we fit the monopole of the correlation function in redshift space. The following sections present the results of this monopole fitting, and the prediction of our model of the quadrupole in redshift space, projected correlation function, monopole in Fourier space and the 3PCF.

BigMD-BOSS light cone covers the same volume as CMASS sample between redshift $z = 0.43$ and 0.7. In order to have a good estimation of the uncertainties in our measurements, we use 100 MD-PATCHY mocks (Kitaura et al. 2016, companion paper). These mocks are produced using five boxes at different redshifts that are created with the PATCHY code (Kitaura et al. 2014). This code matches the clustering of the galaxy catalogues for each redshift bin. The MD-PATCHY mocks are based on the BigMDPL simulation, and they are produced with the same cosmology used in this work. To compute errors, we use the square root of the diagonal terms of the covariance matrix defined as

$$C_{ii} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2, \tag{14}$$

where N is the number of mock catalogues and X is the statistical quantity measured.

5.1 Two-point clustering: result from model and observations

In order to compute the correlation function for our light cone and the observed data, we use a Landy & Szalay estimator (Landy & Szalay 1993). The correlation function is defined by

$$\xi(r) = \frac{DD - 2DR + RR}{RR} \tag{15}$$

where DD, DR and RR represent the normalized data–data, data–random and random–random pair counts, respectively, for the distance range $[r - \Delta r/2, r + \Delta r/2]$.

In this paper, we use random catalogues 20 times larger than the data catalogues. In order to estimate the projected correlation function and the multipoles of the correlation function, we use the

2D correlation function, $\xi(r_\mathrm{p}, \pi)$, where $s = \sqrt{r_\mathrm{p}^2 + \pi^2}$, r_p is the perpendicular component to the line of sight and π represents the parallel component. The correlation function of the BigMD-BOSS light cone is computed using close pair weights and FKP weights (Feldman, Kaiser & Peacock 1994),

$$w_\mathrm{FKP} = \frac{1}{1 + n(z)P_\mathrm{FKP}}, \tag{16}$$

where $n(z)$ is the number density at redshift z and $P_\mathrm{FKP} = 20\,000\ h^{-3}$ Mpc^3. We use the FKP weights to optimally weight regions with different number densities. In the case of the BOSS CMASS sample, we use the galaxy weights given in equation (9) and in addition the FKP weights. The total weights for the data used in our analysis are $w_\mathrm{tot} = w_\mathrm{FKP} w_\mathrm{g}$.

Note that P_FKP is chosen to minimize the variance of power spectrum measurements. For the correlation function measurements, one should use the optimal weight from Hamilton (1993),

$$w_\mathrm{H} = 1/(1 + n(z)J_w), \tag{17}$$

where

$$J_w = \int_0^r \xi(r)\,\mathrm{d}V. \tag{18}$$

However, since we are fixing w_FKP or w_H to be a constant to simplify the computation, we expect that w_H should be similar to w_FKP. In any case, the choice of optimal weight will not bias the measurements.

5.1.1 Redshift-space correlation function

Previous works demonstrated the impact of the scatter in the clustering signal of a mock generated with the abundance matching (e.g. Reddick et al. 2013). In this study, we search for a scatter parameter ($\sigma_\mathrm{HAM}(V_\mathrm{peak}|M_*)$) which reproduces the monopole of the correlation function and provides the prediction for other quantities. The multipoles of the 2PCF, in redshift space, are defined by

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^{1} \xi(r_\mathrm{p}, \pi) P_l(\mu)\,\mathrm{d}\mu, \tag{19}$$

where

$$\mu = \frac{\pi}{\sqrt{r_\mathrm{p}^2 + \pi^2}} \tag{20}$$

and $P_l(\mu)$ is the Legendre polynomial. We will present results for the monopole ($l = 0$) and the quadrupole ($l = 2$).

To find the best value, we fit the clustering using the monopole in the redshift space for the range 2–30 h^{-1} Mpc. The top panel in Fig. 8 shows the results of the fitting compared to the CMASS DR12 data. Errors in Figs 8 and in 9 are computed using 100 MD-PATCHY mocks (Kitaura et al. 2016, companion paper). The parameter that best reproduces the clustering in the monopole is $\sigma_\mathrm{HAM}(V_\mathrm{peak}|M_*) = 0.31$. This result is in agreement with previous works on abundance matching (Trujillo-Gomez et al. 2011; Nuza et al. 2013; Reddick et al. 2013).

The simulation provides a good agreement with data in the monopole for scales smaller than 50 h^{-1} Mpc. However, the bottom panel in Fig. 8 shows a disagreement in the quadrupole for scales smaller than 0.7 h^{-1} Mpc, which can be due to the method used to assign the fibre collisions in the BigMD-BOSS light cone; for this reason, we do not analyse these scales. An additional disagreement is found at scales larger than 6 h^{-1} Mpc, which will be commented in the last section of this work. Nuza et al. (2013) use

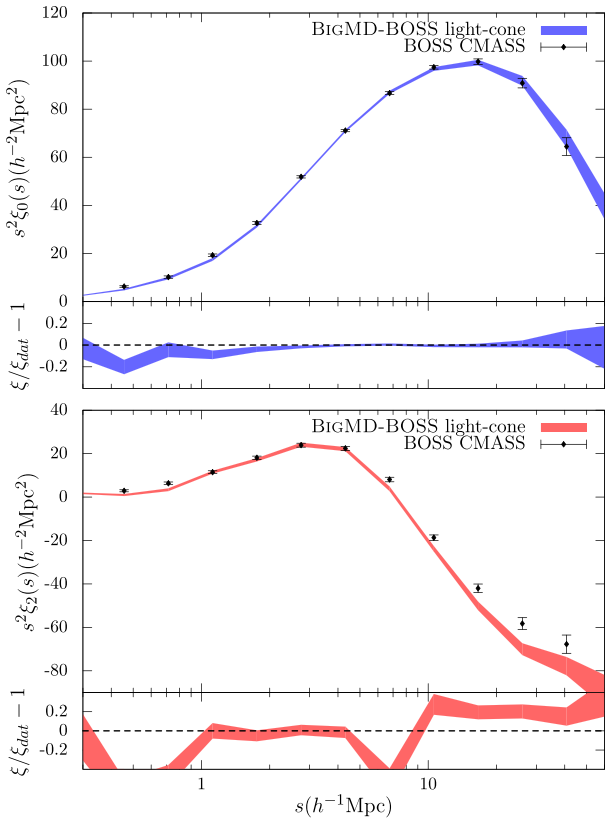

Figure 8. Top panel: monopole in redshift space from CMASS DR12 sample (black points). The shaded area represents the modelling of the monopole using the BigMD-BOSS light cone. Bottom panel: quadrupole in redshift space from CMASS DR12 sample compared with the theoretical prediction from the BigMD-BOSS light cone. Error bars were computed using MD-PATCHY mocks. Small panels show the ratio between the model and the observed data. Fitting of the monopole is performed between 2 and 30 h^{-1} Mpc. The observed monopole is in good agreement with our model for scales larger than 2 h^{-1} Mpc. However, the quadrupole shows tensions with observations for scales $<1\ h^{-1}$ Mpc and $5 > h^{-1}$ Mpc.

the MultiDark simulation with $\Omega_\mathrm{m} = 0.27$. Comparing their results for the monopole, we obtain a better agreement for scales larger than 10 h^{-1} Mpc, mainly due to the difference in cosmologies used in this work.

Fig. 9 shows the prediction of the monopole and quadrupole for large scales compared to the observed data. Discrepancies for some values between the model and the data at scales larger than 60 h^{-1} Mpc could not be due only to the cosmic variance. Differences at the baryon acoustic oscillation (BAO) scales are of the order of 1σ errors while for large scales differences can be of the order of 2σ or 3σ. In Fig. 9, we can see that the BOSS CMASS correlation function at large scales is systematically shifted. This excess of power in the correlation function monopole could be due to the potential photometric calibration systematics which only affect very large scales. Huterer, Cunha & Fang (2013) make a detailed study about the photometric calibration errors and their implication in the measurements of clustering and demonstrate that calibration uncertainties generically lead to large-scale power.

5.1.2 Projected correlation function

The projected correlation function is a quantity which is insensitive to the impact of the redshift-space distortion and provides an

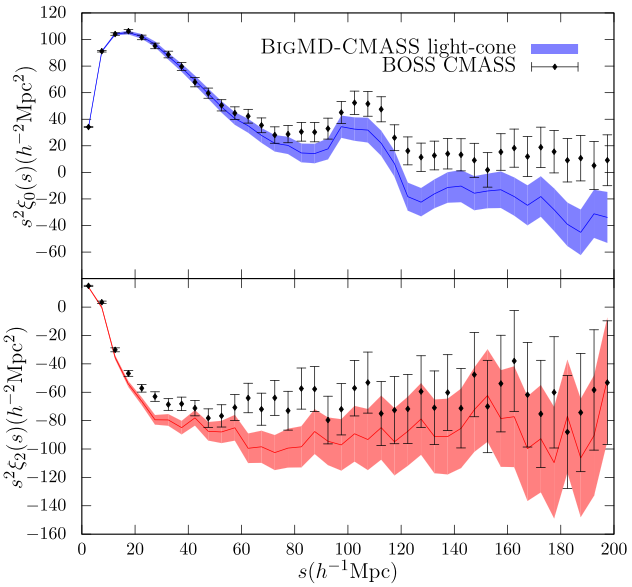

Figure 9. Monopole (top panel) and quadrupole (bottom panel) of the redshift-space correlation function. The shaded areas are the model predictions for large scales using a single light cone. Error bars were computed using MD-PATCHY mocks. Differences in the quadrupole are the same shown in Fig. 8. The monopole has a good agreement up to 100 h^{-1} Mpc. However, large scales present significant difference, but this can be due to the cosmic variance and remaining systematics in the data. These differences are within 2σ errors.

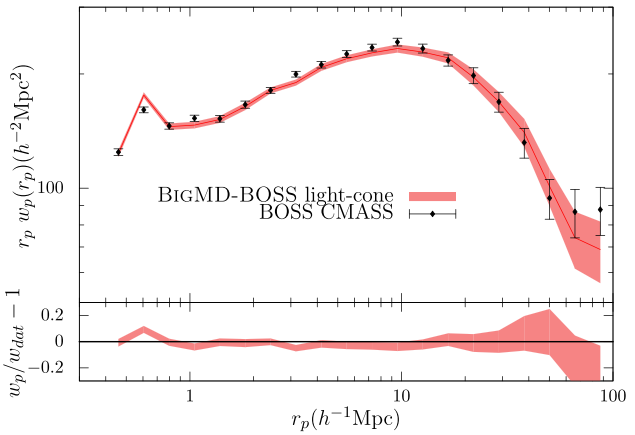

Figure 10. Projected correlation function prediction from the BigMD-BOSS light cone (shaded region) compared to the BOSS CMASS sample. The width of the shaded area represents 1σ errors, computed using MD-PATCHY mocks. Our model reproduces the clustering for all relevant scales. Scales $<0.6\ h^{-1}$ Mpc are dominated by fibre collision effects.

approximation to the real-space correlation function (Davis & Peebles 1983). The projected correlation function is defined as the integral of the 2D correlation function, $\xi(r_\mathrm{p}, \pi)$, over the line of sight:

$$w_\mathrm{p}(r_\mathrm{p}) = 2\int_0^\infty \xi(r_\mathrm{p}, \pi)\,\mathrm{d}\pi. \tag{21}$$

In order to compute $w_\mathrm{p}(r_\mathrm{p})$ from the discrete correlation function (equation 15), we use the estimator

$$w_\mathrm{p}(r_\mathrm{p}) = 2\sum_i^{\pi_\mathrm{max}} \xi(r_\mathrm{p}, \pi_i)\Delta\pi_i. \tag{22}$$

We adopt a linear binning in the light-of-sight direction, $\Delta\pi_i = \Delta\pi = 5\ h^{-1}$ Mpc. We selected $\pi_\mathrm{max} = 100\ h^{-1}$ Mpc. Nuza et al. (2013) find convergence of the projected correlation for this scale. Fig. 10 shows the results found for the BigMD-BOSS light cone compared to the CMASS data. Error bars were computed using 100 MD-PATCHY mocks.

Fig. 10 reveals a discrepancy at scales $\approx 3\ h^{-1}$ Mpc. However, results are in agreement at the 2σ level, so we can consider the data consistent with the prediction of our model. Scales below $0.5\ h^{-1}$ Mpc are dominated by fibre collision. Due to this effect, the clustering declines rapidly.

5.1.3 *Fourier space*

The power spectra for the BOSS CMASS sample with nearest angular neighbour upweighted weights and the BigMDPL are computed using the Feldman et al. (1994) power spectrum estimator modified to account for the systematic weights of the galaxies. In BOSS CMASS, each galaxy is assigned a systematics weight (equation 9), which is accounted for in the estimator. For the BigMD-BOSS light cone, we set $w_\mathrm{g} = w_\mathrm{cp}$, for the power spectrum using nearest-neighbour upweighted fibre collisions weights, and $w_\mathrm{g} = 1$ for the true power spectrum.

The power spectrum for the BOSS CMASS sample is computed using the method described in Hahn et al. (in preparation) in order to correct the effects of fibre collisions on smaller scales. The fibre collision correction method reconstructs the clustering of fibre-collided pairs by modelling the distribution of the line-of-sight displacements between them using pairs with measured redshifts. In addition, the method corrects fibre collisions in the shot-noise correction term of the power spectrum estimator. In simulated mock catalogues, the correction method successfully reproduces the true power spectrum with residuals $\lesssim$1 per cent at $k \sim 0.3\ h$ Mpc^{-1} and $<$10 per cent at $k \sim 0.9\ h$ Mpc^{-1}. The top panel of Fig. 11 compares the fibre collision and systematics corrected BOSS CMASS power spectrum to the true power spectrum of BigMD-BOSS light cone, showing remarkably good agreement between data and model. Figs 8 and 11 confirm that the standard HAM is accurate in the modelling of the clustering not only at large scales, but also in the one-halo term.

Monopoles from our model and the BOSS CMASS data using fibre collision weights are shown in the bottom panel of Fig. 11. Both power spectra agree for k smaller than 1 h Mpc^{-1}. The BigMD-BOSS light cone and the observed data have a remarkably good agreement in the BAO region (inset panel Fig. 11), which is not seen in the correlation function (Fig. 9). This difference can be due to remaining systematics that have a bigger impact on the correlation function than in the power spectrum. The agreement between our model and the observed data, for the true power spectrum and the nearest-neighbour corrected power spectrum, demonstrates that the method used to assign fibre collisions in the BigMD-BOSS light cone is a good approach to simulate this effect.

As we discussed in Section 5.1.1, the disagreement between the model and the data in the correlation function monopole could be due to potential photometric calibration systematics. The effect on the power spectrum will be limited to very small k, so that it has less impact on the BAO scales. However, this excess of power does not have impact on BAO measurements from correlation functions when we marginalize the overall shape (see Chuang et al. 2013; Ross et al., in preparation).

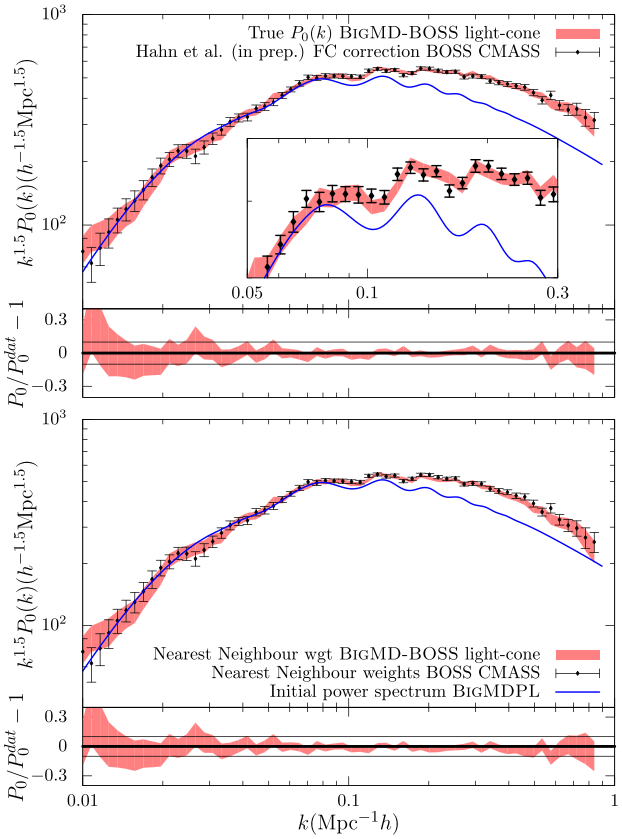

Figure 11. Monopole of power spectrum from the BigMD-BOSS light cone and the CMASS DR12 sample. Top panel: the true power spectrum for our light cone compared to the CMASS DR12 data corrected by fibre collisions using Hahn et al. (in preparation) method. Solid curve shows the initial matter power spectra of the BigMDPL simulation scaled to match the amplitude of fluctuations at long waves. A remarkable agreement between the data and the model is found for scales $k \lesssim 1\,h\,\mathrm{Mpc}^{-1}$. Bottom panel: the comparison between simulation and observed data using nearest-neighbour weights ($w_{\rm cp}$). In addition to $w_{\rm cp}$, observed measurements include systematics weights: $w_{\rm star}$, $w_{\rm zf}$ and $w_{\rm see}$. The agreement between the data and the model, in both panels, shows the good performance of the fibre collision assignment in the light cone. In bottom subpanels, dashed lines represent an accuracy level of 10 per cent.

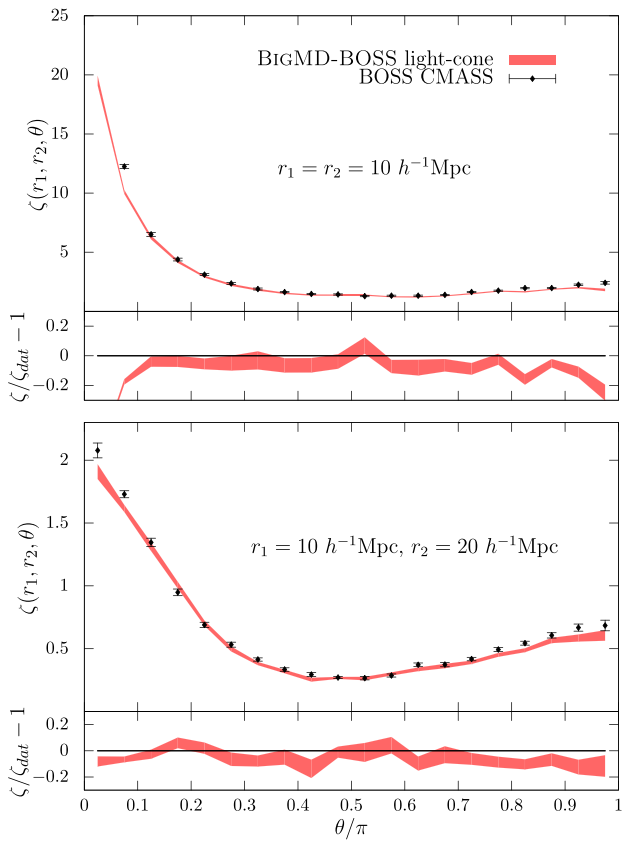

Figure 12. Top panel: BOSS CMASS DR12 3PCF compared with the model prediction of this work. Shaded area shows 1σ uncertainties, with limits $r_1 = 10\ h^{-1}$ Mpc and $r_2 = 20\ h^{-1}$ Mpc. Bottom panel: 3PCF for limits $r_1 = r_2 = 10\ h^{-1}$ Mpc. The BigMD-BOSS light cone can reproduce almost all scales between 2σ errors.

5.2 Three-point correlation function

We are also interested in comparing the prediction of the 3PCF using the HAM on the BigMDPL simulation with the observed data. The 3PCF provides a description of the probability of finding three objects in three different volumes. In the same manner as the 2PCF, the 3PCF is defined as

$$\zeta(r_{12}, r_{23}, r_{31}) = \langle \delta(r_1)\delta(r_2)\delta(r_3)\rangle, \tag{23}$$

where $\delta(r)$ is the dimensionless overdensity at the position r and $r_{ij} = r_i - r_j$. We use the Szapudi & Szalay estimator (Szapudi & Szalay 1998)

$$\zeta = \frac{DDD - 3DDR + 3DRR - RRR}{RRR}. \tag{24}$$

Fig. 12 displays our prediction compared with the BOSS CMASS data. We see the results for two kinds of triangles: $r_1 = r_2 = 10\,h^{-1}$ Mpc and $r_1 = 10\,h^{-1}$ Mpc, $r_2 = 20\,h^{-1}$ Mpc, where θ is the angle between r_1 and r_2.

A good agreement in the shape of the 3PCF is seen in Fig. 12 between our prediction and the data. Most of the points are in agreement within 2σ errors for both configurations represented in Fig. 12. However, the BigMD-BOSS light cone is underestimating the 3PCF for $\theta \sim 0$ and $\theta \sim \pi$. Guo et al. (2015b) find similar discrepancies for those scales, which can be produced by velocity effects and can be corrected including a velocity bias. Therefore, the disagreement in the 3PCF and in the quadrupole of the correlation function can be caused by the same kind of effects.

5.3 Stellar-to-halo mass relation

The stellar-to-halo mass ratio (SHMR) is an important quantity to evaluate if the simulated light cone is providing a realistic halo occupation. In this way, we use results from weak lensing, which is one of the most powerful mechanisms to know the observational SHMR. Fig. 13 shows the SHMR predicted by the BigMD-BOSS light cone and measurements in the Canada–France–Hawaii Telescope (CFHT) Stripe 82 Survey (Shan et al. 2015). In order to ensure the convergence of the haloes in our prediction, we select haloes with masses larger than $5.2 \times 10^{12}\,\mathrm{M}_\odot$. This limit is 150 dark matter particles which give convergence for subhaloes (Klypin et al. 2015).

Predictions of the abundance matching are in agreement with the weak lensing data. In Fig. 13, shaded blue area shows the intrinsic scatter measured. The dependence between scatter and stellar mass

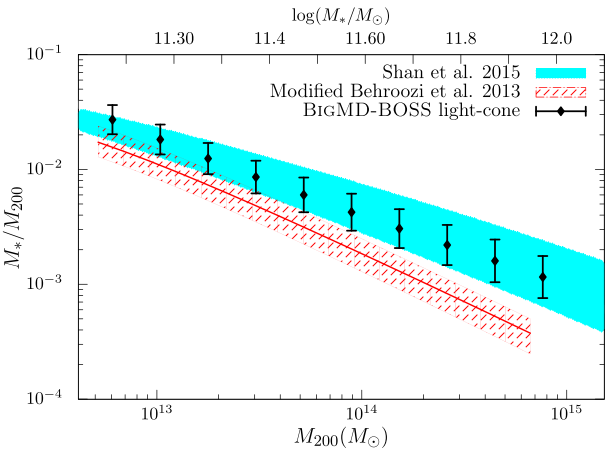

Figure 13. Stellar-to-halo mass ratio. The shaded blue area represents the best fit of the stellar-to-halo mass relation measured using weak lensing in the CFHT Stripe 82 Survey (Shan et al. 2015). The red area represents previous HAM result from Behroozi, Wechsler & Conroy (2013c). The analysis in Behroozi et al. (2013c) was modified using the Planck cosmology parameters and changing the definition of the halo mass. Black dots are the prediction from the HAM-BigMD-BOSS light cone. Differences between our model and Behroozi et al. (2013c) are mainly due to the SMF adopted in both works. Scatter between M_{200} and M_* is similar between the data and our model. We adopted constant scatter while observed data suggest a dependence of the scatter with the stellar mass.

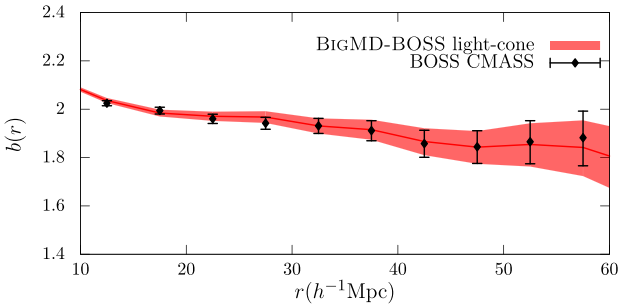

Figure 14. Scale-dependent galaxy bias from the model presented in this work. We measure the bias with respect to the correlation function of dark matter in the BigMDPL light cone for the data and the model. There is an excellent agreement between the CMASS observations and the predictions of the HAM-BigMD-BOSS light cone.

is clear. It is also shown in the abundance matching (e.g. Trujillo-Gomez et al. 2011; Reddick et al. 2013). However, our HAM model uses a constant scatter to reproduce the clustering. This approximation can generate the disagreement in the scatter between data and mock. The red area in Fig. 13 indicates the results from Behroozi et al. (2013c). We modify Behroozi et al. (2013c) in order to use the same definition of halo mass and implement the Planck cosmology in the analysis. The SMF assumptions can be one of the origins for the disagreement between both predictions. While we use the BOSS DR12 stellar mass catalogues to estimate the SMF, Behroozi et al. (2013c) use the PRIMUS SMF (Moustakas et al. 2013). The difference in how the stellar mass catalogues handle profile fitting produces a variation in the high-mass end of both SMFs. This effect causes important difference at large stellar mass between both predictions.

Shankar et al. (2014) present the stellar-to-halo mass relation assuming different mass functions and compare their results with recent models. They find differences between Behroozi et al. (2013c) and Maraston et al. (2013) similar to the one shown in our Fig. 13. Shankar et al. (2014) also find that an intrinsic scatter in stellar mass at fixed halo mass of 0.15 dex is needed to reproduce the BOSS clustering. This result is in agreement with our model, which predicts an intrinsic scatter in stellar mass of 0.14 dex at a fixed halo mass.

5.4 Bias prediction

Using the HAM-BigMD-BOSS light cone and its corresponding dark matter light cone, we can estimate the real-space bias, $b(r)$, solving the equation (Kaiser 1987; Hamilton 1992)

$$\xi(s) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) b(r)^2 \xi_{\rm DM}(r), \qquad (25)$$

where $\beta \approx f/b$ is the redshift-space parameter and $f(z = 0.55) = 0.77$ (Planck cosmology).

Fig. 14 shows the linear bias, which is in agreement with previous papers that reproduced the CMASS clustering (see Nuza et al. 2013). For the data and the model, we use the dark matter correlation function from the BigMD simulation. For the scales shown, the scale-dependent bias factor is in the range 1.8–2. We use the BigMD dark matter light cone to estimate the relative bias of the CMASS sample to this catalogue.

6 DISCUSSION

The BigMD-BOSS light cone is designed to reproduce the full BOSS CMASS sample between redshift 0.43 and 0.7, including observational effects. In order to recover the information at small scales, similar papers (e.g. Nuza et al. 2013; Guo et al. 2014, 2015c) correct the observed data by fibre collision (see Guo et al. 2012; Hahn et al., in preparation). In this work, we assign fibre collisions to galaxies in the light cone, and we use nearest-neighbour weights in the data and in the model. Our model can be useful to test methods that recover the clustering in the fibre collision region (Guo et al. 2012) or in the production of mocks for covariance matrices (Kitaura et al. 2016, companion paper). The fibre collision assignment adopted in this work can reproduce in a good way this observational effect (Fig. 11). However, this approach can introduce small systematics that we do not include in our modelling.

White et al. (2011) model the full CMASS clustering. They find a good fit of the HOD parameters to reproduce the observed data. However, they cannot describe the small scales because they only include close pair weights in the data measurements, which cannot recover the small-scale clustering (Guo et al. 2012). Nuza et al. (2013) also reproduce with a good agreement the CMASS data using a standard HAM model; they correct by fibre collision using the method explained in Guo et al. (2012). Our paper continues the work presented in Nuza et al. (2013), including light-cone effects, redshift evolution, radial selection function, etc. All these papers can reproduce the clustering of the full CMASS sample.

Recent papers show tensions between models and observed data when a most careful selection is done. Guo et al. (2015c) study a volume-limited LRG sample in the redshift range of $0.48 < z < 0.55$ of the CMASS sample. They need a galaxy velocity bias to describe the clustering of the most massive galaxies ($\sim 10^{13}$–10^{14} h^{-1} $\mathrm{M}_\odot$) using HOD. Saito et al. (2015) show an extension of the HAM to describe the colour dependence of the clustering for the CMASS sample. Guo et al. (2015a) present a comparison

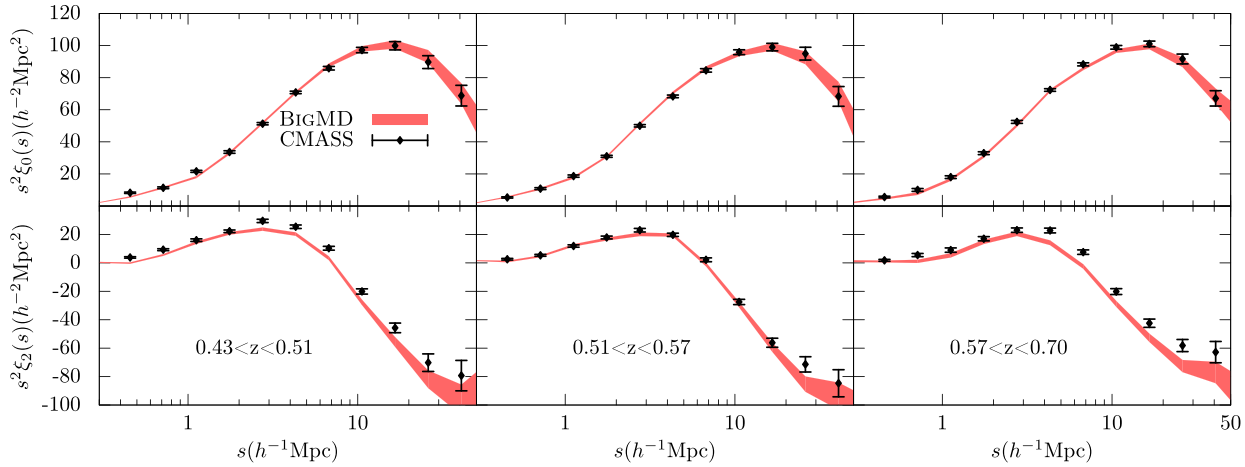

Figure 15. Monopole and quadrupole of the redshift-space correlation function of the CMASS DR12 sample compared to the HAM-BigMD-BOSS light cone for three redshift bins. The monopole is fitted for all redshift ranges. The middle bin is the most complete range in the CMASS sample, and also the best reproduced quadrupole. We perform a HAM with three different scatter parameters to fit each of the redshift bins. Differences at low and high redshift can be due to target selection effects we do not include in this study. Another source of discrepancy can be the relation between the scatter and the more massive galaxies (Saito et al. 2015).

between HOD and HAM models; they also modify the standard HAM model in order to reproduce clustering at different luminosity cuts. Favole et al. (2015b) present a study of the blue population properties compared to the red galaxies. They present a modified HOD which allows them to include both samples in the same mock catalogue. The clustering dependence on stellar mass (luminosity) is not implemented in our model, and we do not distinguish between blue and red galaxies. Our implementation of the HAM and stellar mass incompleteness is capable of reproducing the full CMASS sample, including a big amount of data in our analysis. Zu & Mandelbaum (2015) present a modified HOD in order to include the stellar mass incompleteness (iHOD). This model combines galaxy cluster and galaxy–galaxy lensing and allows one to increase ∼80 per cent the number of modelled galaxies than the traditional HOD models.

We find the largest discrepancy between our model and the data in the quadrupole measurements (Fig. 8). For scales larger than 10 h^{-1} Mpc, this difference is within the 3σ errors. The disagreement for $s < 1\ h^{-1}$ Mpc is larger than 20 per cent. However, this can be due to the uncertainties introduced by the fibre collisions at those scales and effects of the resolution of the simulation. Therefore, we will focus our attention at scales larger than 5 h^{-1} Mpc where the impact of fibre collision is smaller.

In order to study the clustering in different redshift bins using the HAM implemented in this work, we divide the full range into three bins. We select approximately the same number of galaxies in each redshift bin in order to have similar statistics in all of them. We perform an abundance matching (different scatter values that vary from 0.05 to 0.5) for each range to fit the monopole. Fig. 15 shows the monopole and quadrupole for the three different redshift bins. The discrepancy in the quadrupole can be due to one or more of the approaches used in this work. Possible causes of this discrepancy are enumerated below.

(i) Guo et al. (2015c) find similar discrepancies in the quadrupole in configuration space for scales >5 h^{-1} Mpc. They argue that the underestimation of the quadrupole on large scales is possible due to the correlated neighbouring bins in the covariance matrix. They obtain a reasonable χ^2, even with this feature of the predicted quadrupole.

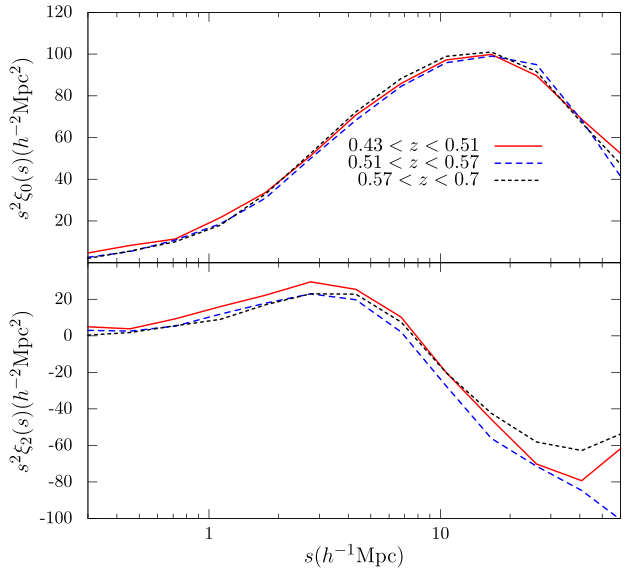

Figure 16. Correlation function for the CMASS sample in three redshift bins. Top panel: monopole with small variations in time. Bottom panel: the quadrupole for the selected ranges. In contrast with the monopole, the quadrupole shows larger variations for the different redshifts.

(ii) Montero-Dorta et al. (2014) show that the intermediate-redshift bin ($0.51 < z < 0.57$) is the most complete region in the CMASS sample. The standard HAM can reproduce monopole and quadrupole for this redshift bin (see Fig. 15), but cannot reproduce the quadrupole for the other two bins. The CMASS DR12 sample has small variations in the monopole. However, quadrupole changes and it becomes similar for the two redshift ranges where the incompleteness of the sample is larger (Fig. 16).

(iii) The values of scatter used to fit the monopole of the correlation function in the different redshift bins vary in a wide range. This can be due to the evolution of the number density in the CMASS sample and some approximations used in this work. Leauthaud et al. (2016) show a non-negligible evolution of the SMF at low redshift compared with the complete redshift range (0.43–0.7). Our

approximation of non-evolving SMF could overestimate the incompleteness in the low-redshift range (Fig. 5, left-hand panel), and then the necessary scatter to reproduce the observed correlation function will be smaller. We also assume a constant mean scatter, but indeed scatter depends on the stellar mass; it increases with the mass of the galaxies (Trujillo-Gomez et al. 2011; Reddick et al. 2013). This dependence can explain why the scatter needed to reproduce the clustering of the low-redshift range is smaller than the one used in the intermediate redshift. At low redshift, the number density is equal to $3.466 \times 10^{-4}\ h^3\ \mathrm{Mpc}^{-3}$, which is smaller than $3.942 \times 10^{-4}\ h^3\ \mathrm{Mpc}^{-3}$ for the middle redshift. If both samples were complete, we will expect a larger scatter in the first range. However, due to the large incompleteness in the high-mass end at low redshift, the mean mass of this sample is $1.86 \times 10^{11}\ \mathrm{M}_{\odot}$ compared to $2.04 \times 10^{11}\ \mathrm{M}_{\odot}$ for the second redshift range. For this reason, the scatter needed to reproduce the clustering is smaller in the low-redshift range. In the high-redshift bin, we can only see very massive galaxies (see Fig. 5, right-hand panel) compared to the whole population of galaxies in the CMASS sample. This range is complete in the high-mass end and, compared to the other two redshift ranges, has a number density very small ($1.534 \times 10^{-4}\ h^3\ \mathrm{Mpc}^{-3}$), which implies larger mean mass ($2.63 \times 10^{11}\ \mathrm{M}_{\odot}$) and scatter than for the other samples.

(iv) We have added a simple model for the stellar mass incompleteness in the CMASS sample. However, there can be other effects of the incompleteness in the target selection that cannot be modelled in this simple way. Although the selection is performed to select LRG, an incomplete blue cloud is in the sample and its fraction compared to the red sequence evolves with redshift (e.g. Guo et al. 2013; Montero-Dorta et al. 2014). Those two populations can live in different kinds of haloes, and therefore they should be described by different scatter values. The errors introduced by this effect can increase with redshift, because the fraction of blue galaxies increases as well. As opposed to the low-redshift bin, the high-redshift bin is complete in the high-mass end (Fig. 5, $z = 0.65$), but the fraction of blue galaxies is larger than the middle bin, which can affect the prediction of the quadrupole. The presence of a small fraction of the so-called 'green valley' can also introduce small errors in our modelling.

(v) The number density in the high-redshift bin ($0.57 < z < 0.70$) is very small compared to the middle redshift range. In this region, the fraction of small galaxies decreases and the impact of the most massive objects in the clustering becomes stronger. Guo et al. (2015a) and Saito et al. (2015) need modification of the HAM model when colour cuts are applied. In addition, Guo et al. (2015c) show the necessity to introduce a velocity bias in the HOD to reproduce the most massive galaxies. If the standard HAM does not describe the clustering of the most massive galaxies, HAM mocks, which model samples as the CMASS in the redshift range $0.57 < z < 0.70$, will not reproduce accurately the clustering of the observed data.

(vi) In addition, recent papers report results for LRG samples where the number of significant miscentral galaxies in haloes is larger than expected (e.g. Hoshino et al. 2015) or the presence of off-centring for central galaxies (e.g. Hikage et al. 2013). The implementation of these results in the construction of mocks reproducing LRG samples could also modify the quadrupole.

7 SUMMARY

We investigated the galaxy clustering of the BOSS CMASS DR12 sample using light cones constructed from the BigMDPL simulation. We perform a HAM to populate the dark matter haloes with galaxies using the Portsmouth DR12 stellar mass catalogue. In addition, the stellar mass distribution is modelled to take into account the incompleteness in stellar mass of the CMASS sample. Our study included features such as the survey geometry, veto masks and fibre collision. The combination of HAM and the BigMDPL simulation provides results in good agreement with the observed data. Our results show that the HAM is a method extremely useful in the study of the relation between dark matter haloes and galaxies, and can be very helpful in the production of mock catalogues (Kitaura et al. 2016, companion paper).

Our main results can be summarized as follows.

(i) We model the observed monopole in configuration space using HAM. Assuming a complete sample, the scatter parameter is very large compared to previous studies. The modelling of stellar mass incompleteness significantly decreases the value of scatter to $\sigma_{\mathrm{HAM}}(V_{\mathrm{peak}}|M_*) = 0.31$. Our model reproduces the observed monopole for nearly every scale.

(ii) The prediction of the quadrupole in configuration space appears to be in disagreement with the observed data. We present possible explanations of this disagreement. In future works, we will concentrate on reducing the possible systematics, in order to understand better the limits of our model.

(iii) We compute the projected correlation function and the 3PCF, finding good agreement between the model and the observed data within 1σ errors for most of the scales. For scales ~ 0 and $\sim \pi$, the differences are of the order of 2σ errors, which can be related to the same factors of the disagreement in the quadrupole. The monopole in k-space of the BigMD-BOSS light cone is in remarkable agreement with the measurement from the CMASS sample corrected by fibre collisions ($\sim$10 per cent of difference at $k = 0.9$). The same agreement is found when we use nearest-neighbour weights, which shows that the assignment of fibre collision in the light cone can reproduce the observed data.

(iv) We compare our prediction of the stellar-to-halo mass relation with lensing measurements. The results are in good agreement with the observed data. Our assumption of a constant scatter is reflected in the differences with observations. Lensing measurements suggest the need to include the stellar mass dependence in the scatter of the HAM.

The BigMD-BOSS light cone is publicly available. It can be found in the SDSS SkyServer.[3] The current version includes angular coordinates (RA, Dec.), redshift in real space and redshift space, peculiar velocity in the line of sight, M_{200}, V_{peak} and M_*. Properties of galaxies such as effective radius (R_{eff}), velocity dispersion (σ_{v}) and mass-to-light ratio (M/L) will be included in future updates.

ACKNOWLEDGEMENTS

SRT is grateful for support from the Campus de Excelencia Internacional UAM/CSIC. SRT also thanks Fernando Campos del Pozo for useful discussions and help while developing the SUGAR code.

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 them with the optimized version of GADGET-2.

[3] http://skyserver.sdss.org/dr12/en/home.aspx

SRT, CC, FP, AK, FSK, GF and SG acknowledge support from the Spanish MICINNs Consolider-Ingenio 2010 Programme under grant MultiDark CSD2009-00064, MINECO Centro de Excelencia Severo Ochoa Programme under grant SEV-2012-0249 and grant AYA2014-60641-C2-1-P. GY acknowledges support from MINECO (Spain) under research grants AYA2012-31101 and FPA2012-34694 and Consolider Ingenio SyeC CSD2007-0050. FP wishes to thank the Lawrence Berkeley National Laboratory for the hospitality during the development of this work. FP also acknowledges the Spanish MEC 'Salvador de Madariaga' programme, Ref. PRX14/00444.

CH also wants to thank the Instituto de Física Teórica UAM/CSIC for the hospitality during his summer visit, where part of this work was completed. GF acknowledges financial support from the Ministerio de Educación y Ciencia of the Spanish Government through FPI grant AYA2010-2131-C02-01. FSK acknowledges the support of the Karl-Schwarzschild Program from the Leibniz Society.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation and the US Department of Energy Office of Science. The SDSS-III website is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington and Yale University.

REFERENCES

Abazajian K. et al., 2004, AJ, 128, 502
Aihara H. et al., 2011, ApJS, 193, 29
Alam S. et al., 2015, ApJS, 219, 12
Anderson L. et al., 2014, MNRAS, 441, 24
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18
Behroozi P. S., Wechsler R. H., Conroy C., 2013c, ApJ, 770, 57
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bernardeau F., 1994, ApJ, 427, 51
Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
Bolton A. S. et al., 2012, AJ, 144, 144
Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575
Buchert T., 1994, MNRAS, 267, 811
Catelan P., 1995, MNRAS, 276, 115
Chuang C.-H. et al., 2013, preprint (arXiv:1312.4889)
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Dawson K. S. et al., 2013, AJ, 145, 10
Eisenstein D. J. et al., 2011, AJ, 142, 72
Favole G. et al., 2015a, preprint (arXiv:1507.04356)
Favole G., McBride C. K., Eisenstein D. J., Prada F., Swanson M. E., Chuang C.-H., Schneider D. P., 2015b, preprint (arXiv:1506.02044)
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Gunn J. E. et al., 1998, AJ, 116, 3040
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., 2013, ApJ, 767, 122
Guo H. et al., 2014, MNRAS, 441, 2398
Guo H. et al., 2015a, MNRAS, 449, L95
Guo H. et al., 2015b, MNRAS, 449, L95
Guo H. et al., 2015c, MNRAS, 453, 4368
Hamilton A. J. S., 1992, ApJ, 385, L5
Hamilton A. J. S., 1993, ApJ, 417, 19
Hikage C., Mandelbaum R., Takada M., Spergel D. N., 2013, MNRAS, 435, 2345
Hoshino H. et al., 2015, MNRAS, 452, 998
Huterer D., Cunha C. E., Fang W., 2013, MNRAS, 432, 2945
Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, 1
Kaiser N., 1987, MNRAS, 227, 1
Kitaura F.-S., Heß S., 2013, MNRAS, 435, L78
Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
Kitaura F.-S. et al., 2016, MNRAS, 456, 4156
Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2
Klypin A., Yepes G., Gottlober S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Klypin A., Prada F., Yepes G., Heß S., Gottlöber S., 2015, MNRAS, 447, 3693
Knebe A. et al., 2015, MNRAS, 451, 4029
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kroupa P., 2001, MNRAS, 322, 231
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Leauthaud A. et al., 2012, ApJ, 744, 159
Leauthaud A. et al., 2016, MNRAS, 457, 4021
Manera M. et al., 2013, MNRAS, 428, 1036
Maraston C. et al., 2013, MNRAS, 435, 2764
Mohayaee R., Mathis H., Colombi S., Silk J., 2006, MNRAS, 365, 939
Montero-Dorta A. D. et al., 2014, preprint (arXiv:1410.5854)
Moustakas J. et al., 2013, ApJ, 767, 50
Neyrinck M. C., 2013, MNRAS, 428, 141
Nuza S. E. et al., 2013, MNRAS, 432, 743
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Press W. H., Schechter P., 1974, ApJ, 187, 425
Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30
Reid B. et al., 2016, MNRAS, 455, 1553
Ross A. J. et al., 2012, MNRAS, 424, 564
Saito S. et al., 2015, preprint (arXiv:1509.00482)
Schaye J. et al., 2015, MNRAS, 446, 521
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Shan H. et al., 2015, preprint, (arXiv:1502.00313)
Shankar F. et al., 2014, ApJ, 797, L27
Shu Y., Bolton A. S., Schlegel D. J., Dawson K. S., Wake D. A., Brownstein J. R., Brinkmann J., Weaver B. A., 2012, AJ, 143, 90
Smee S. A. et al., 2013, AJ, 146, 32
Springel V., 2005, MNRAS, 364, 1105
Stoughton C. et al., 2002, in Tyson J. A., Wolff S., eds, Proc. SPIE Conf. Ser. Vol. 4836, Survey and Other Telescope Technologies and Discoveries. SPIE, Bellingham, p. 339
Szapudi I., Szalay A. S., 1998, ApJ, 494, L41
Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
White M. et al., 2011, ApJ, 728, 126
York D. G. et al., 2000, AJ, 120, 1579
Zheng Z. et al., 2005, ApJ, 633, 791
Zu Y., Mandelbaum R., 2015, MNRAS, 454, 1161

[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 Autonoma de Madrid (UAM), Cantoblanco, E-28049 Madrid, Spain*
[4]*Leibniz-Institut für Astrophysik Potsdam (AIP), D-14482 Potsdam, Germany*
[5]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[6]*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain*
[7]*Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 20030, China*
[8]*Department of Physics and Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA*
[9]*Astronomy Department, New Mexico State University, Las Cruces, NM 88003, USA*
[10]*Severo Ochoa Associate Researcher at the Instituto de Física Teórica (UAM/CSIC), E-28049 Madrid, Spain*
[11]*Space Telescope Science Institute, Baltimore, MD 21218, USA*
[12]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA*
[13]*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*
[14]*Center for Astrophysics, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA*
[15]*Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*
[16]*National Abastumani Astrophysical Observatory, Ilia State University, 2A Kazbegi Ave., GE-1060 Tbilisi, Georgia*
[17]*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*
[18]*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*

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