

Building a better understanding of the massive high-redshift BOSS CMASS galaxies as tools for cosmology

Ginevra Favole,^{1,2★} Cameron K. McBride,^{3★} Daniel J. Eisenstein,³
Francisco Prada,^{1,2,4} Molly E. Swanson,³ Chia-Hsun Chuang^{1,2†}
and Donald P. Schneider^{5,6}

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*Center for Astrophysics, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA*

⁴*Instituto de Astrofísica de Andalucía (CSIC), Granada E-18008, Spain*

⁵*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*

⁶*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*

Accepted 2016 July 20. Received 2016 July 20; in original form 2015 June 5

ABSTRACT

We explore the massive bluer star-forming population of the Sloan Digital Sky Survey (SDSS) III/BOSS CMASS DR11 galaxies at $z > 0.55$ to quantify their differences, in terms of redshift-space distortions and large-scale bias, with respect to the luminous red galaxy sample. We perform a qualitative analysis to understand the significance of these differences and whether we can model and reproduce them in mock catalogues. Specifically, we measure galaxy clustering in CMASS on small and intermediate scales ($0.1 \lesssim r \lesssim 50 h^{-1}$ Mpc) by computing the two-point correlation function – both projected and redshift-space – of these galaxies, and a new statistic, $\Sigma(\pi)$, able to separate the coherent and dispersed redshift-space distortion contributions and the large-scale bias. We interpret our clustering measurements by adopting a Halo Occupation Distribution (HOD) scheme that maps them on to high-resolution N -body cosmological simulations to produce suitable mock galaxy catalogues. The traditional HOD prescription can be applied to the red and the blue samples, independently, but this approach is unphysical since it allows the same mock galaxies to be either red or blue. To overcome this ambiguity, we modify the standard formulation and infer the red and the blue models by splitting the full mock catalogue into two complementary and non-overlapping submocks. This separation is performed by constraining the HOD with the observed CMASS red and blue galaxy fractions and produces reliable and accurate models.

Key words: galaxies: distances and redshifts – galaxies: haloes – galaxies: statistics – cosmology: observations – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

In the last decade, an enormous effort has been spent to explore the formation and evolution of the large-scale structure of our Universe. The standard cold dark matter (Λ CDM) model with cosmological constant, together with the theory of cosmic inflation, has become the leading theoretical picture in which structures can form, providing a clear prediction for their initial conditions and hierarchical growth through gravitational instability (e.g. Primack 1997). Testing this model requires one to combine large N -body simulations with measurements from last generation large-volume photometric

and spectroscopic galaxy surveys, as the Sloan Digital Sky Survey (SDSS), (York et al. 2000; Gunn et al. 2006; Smee et al. 2013) and the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013). In particular, BOSS has been able to measure the Baryon Acoustic Oscillation (BAO) feature in the clustering of galaxies and Lyman α forest with unprecedented accuracy, by collecting spectra of 1.5 million galaxies up to $z = 0.7$ (Anderson et al. 2014), over a 10 000 deg² area of sky, and about 160 000 Lyman α forest spectra of quasars in the redshift range $2.2 < z < 3$ (Slosar et al. 2011).

The Λ CDM paradigm claims that galaxies form at the centre of dark matter haloes, thus estimating the clustering features of such complex structures, is currently one of the main targets of modern cosmology (Kravtsov & Borgani 2012). Despite the recent dramatic improvement in the observational data, what primarily prevents us

*E-mail: g.favole@csic.es (GF); cmcbride@cfa.harvard.edu (CKM)

† MultiDark Fellow.

from achieving this goal immediately is the theoretical uncertainty of galaxy bias i.e. the difference between the distribution of galaxies and that of the matter. Galaxies are treated as biased tracers of the underlying matter distribution, and observations of their clustering properties are used to infer those cosmological parameters that govern the matter content of the Universe. In this context, the Halo Occupation Distribution (HOD; Berlind & Weinberg 2002; Kravtsov et al. 2004; Zheng et al. 2005; Zheng, Coil & Zehavi 2007) framework has emerged as a powerful tool to bridge the gap between galaxies and dark matter haloes, providing a theoretical framework able to characterize their mutual relation in terms of the probability, $P(N|M)$, that a halo of virial mass M contains N galaxies of a given type. At the same time, it provides a robust prediction of the relative spatial and velocity distributions of galaxies and dark matter within haloes. In this approach, the use of large-volume N -body cosmological simulations is crucial to produce reliable maps of the dark matter sky distribution.

In this work, we explore the red/blue colour bimodality observed in the BOSS CMASS DR11 (Alam et al. 2015) galaxy sample. In order to quantify and model the differences between these two galaxy populations, we measure their clustering signal on small and intermediate scales, from $r \sim 0.1 h^{-1}$ Mpc up to $r \sim 50 h^{-1}$ Mpc. Specifically, we compute the two-point correlation function (2PCF) – both projected and in redshift-space – of these galaxies, and a new metric, $\Sigma(\pi)$, designed to extract the maximum amount of information about the small-scale non-linear redshift-space distortions. We map our results to the MultiDark cosmological simulation (Prada et al. 2011; Riebe et al. 2011) using an HOD approach (Zheng et al. 2007; White et al. 2011), to generate reliable mock galaxy catalogues. In this context, we investigate whether we can find an HOD parametrization able to model both the blue and red observed clustering amplitudes, with small variations in its parameters. As an alternative to HOD models, one can interpret clustering observations with an Halo Abundance Matching (HAM) prescription (e.g. Trujillo-Gomez et al. 2011; Nuza et al. 2013) with the advantage of avoiding free parameters, only assuming that more luminous galaxies are associated with more massive haloes, monotonically, through their number densities. HAM is a straightforward technique that provides accurate predictions for clustering measurements; nevertheless, we choose to model our CMASS clustering measurements using a five-parameter HOD scheme because it is a general method, based on a halo mass parametrization, and does not require a specific luminosity (stellar mass) function (Montero-Dorta et al. 2014) to reproduce the observations.

The traditional HOD modelling reproduces well the clustering signal observed in CMASS, but it provides unphysical predictions when applied to the red and blue subsamples, independently. In fact, in the process of populating a halo with central and satellite galaxies, it allows the same galaxy to be either red or blue i.e. to be placed in haloes with different masses. To overcome this ambiguity, we propose an alternative prescription, that recovers the red and the blue models by splitting the full mock catalogue into two non-overlapping submocks. The separation is performed in a ‘natural’, way by reproducing the observed CMASS red and blue galaxy fractions, as a function of the central halo mass. The resulting mocks are no longer independent – they are based on the same HOD parameter set – and the total number of degrees of freedom is reduced from 15 (three independent models, with five parameters each) to 5 (full HOD) plus 2 (galaxy fraction constraint).

We investigate the impact of redshift-space distortions on the clustering signal, both on small (one-halo term) and intermediate (two-halo level) scales. Our new metrics, $\Sigma(\pi)$, allows us to separate

and quantify both the non-linear elongation seen in the two-point correlation function below $2 h^{-1}$ Mpc, and the Kaiser compression at scales beyond $10 h^{-1}$ Mpc. We model these effects in terms of two parameters, v_σ and β , respectively encoding the galaxy velocity dispersion with respect to the surrounding Hubble flow, and the linear large-scale bias. In agreement with several previous works (see, for instance, Zehavi et al. 2005b; Wang et al. 2007; Swanson et al. 2008), we find that red galaxies are more clustered (i.e. higher peculiar velocity contribution) and biased, compared to their blue star-forming companions. Similar red/blue comparisons in terms of redshift-space distortions and linear galaxy bias have been performed in previous studies (e.g. Masters et al. 2011; Ross et al. 2014). So far, however, most results for blue galaxies are for much less massive samples than CMASS. In addition, CMASS is a very large data set, and this provides a good motivation for being quantitatively exact in estimating its large-scale bias and small-scale peculiar velocities, even if the qualitative behaviour is standard.

The paper is organized as follows. In Section 2 we introduce the methodology used to measure and model galaxy clustering in the BOSS CMASS DR11 sample: we define the metrics used, the correlation function and the covariance estimators. We then give an overview of the MultiDark simulation, discuss the HOD formalism adopted to create mock galaxy catalogues, and introduce the analytic tools used to model both finger-of-god (FoG) and Kaiser effects. In Section 3 we present the CMASS DR11 sample and the specific red/blue colour selection used in the analysis, we illustrate how to weight the data to account for fibre collision and redshift failure effects, and outline the procedure adopted to generate randoms. Section 4 describes how we model our full CMASS clustering measurements building reliable mock galaxy catalogues that take into account the contribution of redshift-space distortions, and present the first results for the three metrics of interest: $\xi(s)$, $w_p(r_p)$, $\Sigma(\pi)$. In Section 5 we apply the traditional HOD formulation individually to the full, the red and the blue CMASS galaxy samples to create their own independent mock catalogues. Then, we present an alternative method to recover the red and blue submocks from the full one, by splitting it using, as a constraint, the observed CMASS red/blue galaxy fractions. Our data versus mock $\Sigma(\pi)$ results, compared to the v_σ , β analytic models are shown in Section 6. Section 7 reports our main conclusions.

2 METHODS

2.1 Clustering measurements

We quantify the clustering of galaxies by computing the two-point correlation function i.e. the excess probability over random to find a pair of galaxies typically parametrized as a function of their co-moving separation (see e.g. Peebles 1980). The galaxy correlation function is well known to approximate a power law across a wide range of scales,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \quad (1)$$

where r_0 is the correlation length, and γ is the power-law slope or spectral index. However, improved models (see review at Cooray & Sheth 2002) have been shown to better match the data (Zehavi et al. 2004).

The redshift-space correlation function differs from the real-space one due to the distortion effects caused by our inability to separate the peculiar velocities of galaxies from their recession velocity when we estimate distances from the redshift. These distortions

introduce anisotropies in the 2PCF in two different ways. On large scales, where the linear regime holds, galaxies experience a slow in-fall towards an over-dense region, and the peculiar velocities make structures appear squashed in the line-of-sight direction, an effect commonly known as ‘Kaiser compression’ (Kaiser 1987; Hamilton 1998). At smaller scales, non-linear gravitational collapse creates virialized systems and thereby relatively large velocity differences arise between close neighbours resulting in structures appearing significantly stretched along the line of sight (Jackson 1972). This effect is commonly referred to as the ‘FoG’.

We are interested in using three related two-point clustering metrics: the redshift-space monopole, $\xi(s)$, the projected correlation function, $w_p(r_p)$, and a new line of sight focused measurement to capture small-scale redshift-space distortion effects, $\Sigma(\pi)$, which we define below. In our formalism, s represents the redshift-space pair separation, while r_p and π are the perpendicular and parallel components with respect to the line of sight such that $s = \sqrt{r_p^2 + \pi^2}$. We can parametrize the redshift-space correlation function as a function of redshift-space separation s or, equivalently, in terms of r_p and π . We can mitigate the impact of redshift-distortions by integrating along the line of sight to approximate the real-space clustering (Davis & Peebles 1983) in the projected correlation function,

$$w_p(r_p) = 2 \int_0^\infty \xi(r_p, \pi) d\pi. \quad (2)$$

This integration is performed over a finite line-of-sight distance as a discrete sum,

$$w_p(r_p) = 2 \sum_i^{\pi_{\max}} \xi(r_p, \pi) \Delta\pi_i, \quad (3)$$

where π_i is the i th bin of the line-of-sight separation, and $\Delta\pi_i$ is the corresponding bin size. We use $\pi_{\max} = 80 h^{-1}$ Mpc and $\Delta\pi = 10 h^{-1}$ Mpc.

Since $w_p(r_p)$ is not affected by redshift-space distortions, the best-fitting power law is equivalent to a real-space measurement. One can therefore quantify the deviation of the redshift-space $\xi(r_p, \pi)$ correlation function from the real-space behaviour by measuring the ratio,

$$\Sigma(\pi) = \frac{\xi(\bar{r}_p, \pi)}{\xi(\pi)}, \quad (4)$$

where $\xi(\pi)$ is the best-fitting power law to $w_p(r_p)$, evaluated at the π scale, and \bar{r}_p indicates that we perform a spherical average in the range $0.5 \leq r_p \leq 2 h^{-1}$ Mpc. This statistic illuminates the non-linear FoG effects by normalizing out the expected real-space clustering along the line-of-sight direction. We are interested in the behaviour of pairs that potentially occupy the same halo, hence our focus at small r_p values. In the attempt to interpret the small-scale non-linear redshift-space distortions, $\Sigma(\pi)$ is preferable to measuring the quadrupole-to-monopole ratio, $\xi_2(s)/\xi_0(s)$ (Hamilton 1992, 1998; Peacock et al. 2001), because it is a compressed representation of $\xi(r_p, \pi)$ which allows us to disentangle the contribution of the distortions along the line of sight from the effects across it. In Appendix A, we provide a comparison between the $\Sigma(\pi)$ and $\xi_2(s)/\xi_0(s)$ statistics as a function of the physical scale.

2.2 Correlation function estimation

For our clustering statistics, we use the estimator of Landy & Szalay (1993):

$$\xi(s) = \frac{\text{DD}(s) - 2\text{DR}(s) + \text{RR}(s)}{\text{RR}(s)} \quad (5)$$

where DD, DR and RR are the data–data, data–random and random–random weighted pair counts computed from a data sample of N galaxies and a random catalogue of N_R points. These pair counts are normalized by the number of all possible pairs, typically by dividing by $N(N-1)/2$, NN_R and $N_R(N_R-1)/2$, respectively, and weighted by Ross et al. (2012)

$$\text{DD}(r_p, \pi) = \sum_i \sum_j w_{\text{tot},i} w_{\text{tot},j} \Theta_{ij}(r_p, \pi) \quad (6)$$

with w_{tot} given by equation (31), and $\Theta_{ij}(r_p, \pi)$ represents a step-function which is 1 if r_p belongs to the i th and π to the j th bin, and 0 otherwise. These weights correct the galaxy densities to provide a more isotropic selection, therefore they should not be applied to the random catalogue, which is based on an isotropic distribution. For randoms $w_{\text{tot},i} = w_{\text{tot},j} = 1$, therefore

$$\text{DR}(r_p, \pi) = \sum_i \sum_j w_{\text{tot},i} \Theta_{ij}(r_p, \pi), \quad (7)$$

$$\text{RR}(r_p, \pi) = \sum_i \sum_j \Theta_{ij}(r_p, \pi). \quad (8)$$

To evaluate the correlation function, we create a random catalogue that has the same selection as the BOSS CMASS galaxy data matching both the redshift distribution and sky footprint (see e.g. Anderson et al. 2014). The method of random catalogue construction is almost identical to that described in Anderson et al. (2014), but constructed to be 10 times as dense as the galaxy data. We down-sample random points based on sky completeness, and ‘shuffle’ the observed galaxy redshifts assigning them to random sky positions so as to exactly reproduce the observed redshift distribution.

2.3 Covariance estimation

To estimate the uncertainties in our clustering measurements, we utilize the jackknife re-sampling technique (Quenouille 1956; Turkey 1958; Miller 1974; Norberg et al. 2009, 2011). There are known limitations to this type of error estimation (see e.g. Norberg et al. 2009), but they have proven sufficient in analyses on scales similar to our analysis (Zehavi et al. 2002, 2005a, 2011; Anderson et al. 2012; Guo, Zehavi & Zheng 2012; Ross et al. 2012). The jackknife covariance matrix for N_{res} re-samplings is computed by

$$\mathbf{C}_{ij} = \frac{N_{\text{res}} - 1}{N_{\text{res}}} \sum_{a=1}^{N_{\text{res}}} (\xi_i^a - \bar{\xi}_i) (\xi_j^a - \bar{\xi}_j), \quad (9)$$

where $\bar{\xi}_i$ is the mean jackknife correlation function estimate in the specific i th bin,

$$\bar{\xi}_i = \sum_{a=1}^{N_{\text{res}}} \xi_i^a / N_{\text{res}}. \quad (10)$$

The overall factor in equation (9) takes into account the lack of independence between the N_{res} jackknife configurations: from one copy to the next, only two subvolumes are different or, equivalently, $N_{\text{res}} - 2$ subvolumes are the same (Norberg et al. 2011).

2.4 The MultiDark simulation

MultiDark (Prada et al. 2011) is a N -body cosmological simulation with 2048^3 dark matter particles in a periodic box of $L_{\text{box}} = 1 h^{-1}$ Gpc on a side. The first run, MDR1, was performed in 2010, with an initial redshift of $z = 65$, and a mass resolution of $8.721 \times 10^9 h^{-1} M_{\odot}$. It is based on the *WMAP5* cosmology (Komatsu et al. 2009), with parameters: $\Omega_m = 0.27$, $\Omega_b = 0.0469$, $\Omega_{\Lambda} = 0.73$, $n_s = 0.95$ and $\sigma_8 = 0.82$. Here Ω is the present day contribution of each component to the matter-energy density of the Universe; n_s is the spectral index of the primordial density fluctuations, and σ_8 is the linear RMS mass fluctuation in spheres of $8 h^{-1}$ Mpc at $z = 0$.

MultiDark includes both the Bound Density Maxima (BDM; Klypin & Holtzman 1997; Riebe et al. 2011), and the Friends-of-Friends (FOF; Davis et al. 1985) halo-finders. For the current analysis, we use only BMD haloes that are identified as local density maxima truncated at some spherical cut-off radius, from which unbound particles (i.e. those particles whose velocity exceeds the escape velocity) are removed. According to the overdensity limit adopted, two different BDM halo catalogues are produced: (i) BDMV – haloes extend up to $\Delta_{\text{vir}} \times \rho_{\text{back}}$, where $\Delta_{\text{vir}} = 360$ is the virial overdensity threshold, $\rho_{\text{back}} = \Omega_m \times \rho_c$ is the background or average matter density, and ρ_c is the critical density of the Universe. (ii) BDMW – the maximum halo density is $\Delta_{200} \times \rho_c$, where $\Delta_{200} = 200$, which implies that BDMW haloes are smaller than BDMV ones. The bound density maxima algorithm treats haloes and subhaloes (those substructures whose virial radius lies inside a larger halo) in the same way, with no distinction. In this work we use the BDMW halo catalogues, since they resolve better the distribution of substructures in distinct haloes, leading to a clearer small-scale clustering signal.

2.5 HOD model using subhaloes

The halo model (reviewed in Cooray & Sheth 2002) is a powerful tool to understand the clustering of galaxies. The HOD (Berlind & Weinberg 2002) is a commonly used method of mapping galaxies to dark matter haloes, which characterizes the bias between galaxies and the underlying dark matter distribution. The HOD is based on the conditional probability, $P(N|M)$, that a halo with mass M contains N galaxies of a given type. In our analysis, we apply the five-parameter HOD formalism presented in Zheng et al. (2007) using the MDR1 simulation at $z = 0.53$. First, we populate distinct haloes with central galaxies whose mean is given by the function form of:

$$\langle N_{\text{cen}}(M) \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\text{min}}}{\sigma_{\log M}} \right) \right], \quad (11)$$

where erf is the error function, $\text{erf}(x) = 2 \int_0^x e^{-t^2} dt / \sqrt{\pi}$.

The free parameters are M_{min} , the minimum mass scale of haloes that can host a central galaxy, and $\sigma_{\log M}$, the width of the cut-off profile. At a halo mass of M_{min} , 50 per cent of haloes host a central galaxy, which in terms of probability means that $P(1) = 1 - P(0)$. If the relation between galaxy luminosity and halo mass had no scatter, $\langle N_{\text{cen}}(M) \rangle$ would be modelled by a hard step function. In reality, this relation must possess some scatter, resulting in a gradual transition from $N_{\text{cen}} \simeq 0$ to $N_{\text{cen}} \simeq 1$. The width of this transition is $\sigma_{\log M}$. In order to place the satellite galaxies, we assume their number in haloes of a given mass follows a Poisson distribution, which is consistent with theoretical predictions (Berlind & Weinberg 2002; Kravtsov et al. 2004; Zheng et al. 2005). We approximate the mean

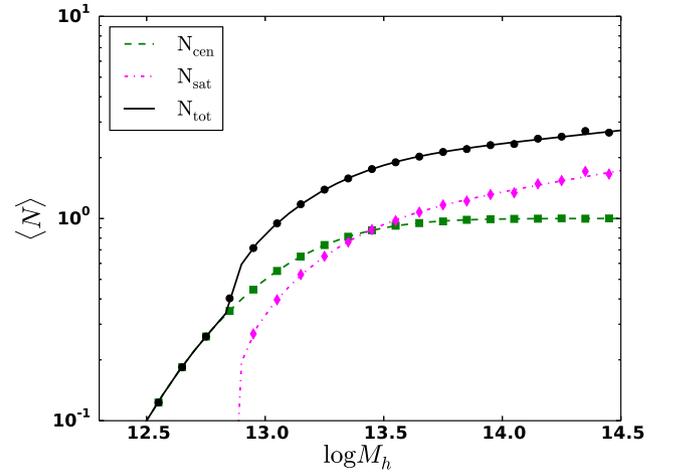


Figure 1. Five-parameter HOD model for MDR1, at $z = 0.53$. The parametrization is from Zheng et al. (2007), and the input values from White et al. (2011). The total (solid line) population of galaxies is the sum of two contributions: central (dashed) and satellite (dot-dashed) galaxies.

number of satellite galaxies per halo with a power law truncated at a threshold mass of M_0

$$\langle N_{\text{sat}} \rangle = \langle N_{\text{cen}}(M) \rangle \left(\frac{M - M_0}{M'_1} \right)^{\alpha'}. \quad (12)$$

The parameter M'_1 corresponds to the halo mass where $N_{\text{sat}} \simeq 1$, when (as in our case) $M'_1 > M_0$ and $M'_1 > M_{\text{min}}$. When $\alpha' = 1$ and $M > M_0$, the mean number of satellites per halo is proportional to the halo mass. To populate with satellite galaxies, we randomly extract from each host halo a certain number of its subhaloes, following a Poisson distribution with mean given by equation (12). The coordinates of these subhaloes become the locations for satellites. This approach, explored in previous works as Kravtsov et al. (2004), White et al. (2011), is intrinsically different from the more commonly used procedure, in which satellites are assigned by randomly assigning the positions of dark-matter particles (see e.g. Reid & Spergel 2009). In our case, satellite galaxies are assigned by reflecting the original halo structure made of one central halo plus none, one, or many subhaloes.

Fig. 1 shows our HOD model built from MultiDark BDMW at $z = 0.53$, for the full CMASS sample: central galaxies are represented by the dashed curve; satellites are the dot-dashed line and the total contribution is the solid curve. As input parameters, we adopt the values consistent with the BOSS CMASS HOD modelling in White et al. (2011).

2.6 Analytic models

Kaiser (1987) demonstrated that on large scales, where the linear regime holds, the redshift-space correlation function can be factorized in terms of its real space version, $\xi(r)$, as

$$\xi(s) = \xi(r) \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2 \right), \quad (13)$$

where β is the Kaiser factor encoding the compression effect (Section 2.1) seen in the clustering signal and b is the linear bias between galaxies and the underlying matter distribution. These two quantities can be related (e.g. Peebles 1980; Linder 2005) through the approximate expression

$$\beta \simeq \Omega_m^{0.6} / b, \quad (14)$$

where the numerator in the right-hand side represents a parametrization of the growth rate of structures, which has been shown (see e.g. Linder 2005; di Porto, Amendola & Branchini 2012; Hudson & Turnbull 2012) that holds for the Λ CDM model. In general, one can decompose the redshift-space separation s into its parallel and transverse components to the line of sight and approximate $\xi(r)$ with the power law in equation (1) to produce (Matsubara & Suto 1996):

$$\xi(r_p, \pi) = \xi(r) \left\{ 1 + \frac{2(1 - \gamma\mu^2)}{3 - \gamma} \beta + \frac{3 - 6\gamma\mu^2 + \gamma(2 + \gamma)\mu^4}{(3 - \gamma)(5 - \gamma)} \beta^2 \right\}. \quad (15)$$

Here γ is the power-law spectral index and μ is the cosine of the angle between the separation and the line-of-sight direction. We include the small-scale non-linear FoG by convolving with a pairwise velocity distribution (Fisher et al. 1994; Hamilton 1998; Croom et al. 2005), which can be modelled as an exponential,

$$f_{\text{exp}}(w) = \frac{1}{\sqrt{2}v_\sigma} \exp\left(-\sqrt{2}\frac{|w|}{v_\sigma}\right), \quad (16)$$

or a Gaussian form,

$$f_{\text{norm}}(w) = \frac{1}{\sqrt{2\pi}v_\sigma} \exp\left(-\frac{w^2}{2v_\sigma^2}\right), \quad (17)$$

where v_σ is the pairwise velocity dispersion. The full model then becomes

$$\xi(r_p, \pi) = \int_{-\infty}^{+\infty} \xi(r_p, r_z(w)) f(w) dw, \quad (18)$$

with $\xi(r_p, r_z(w))$ given by equation (15). The quantity $r_z(w) \equiv (\pi - w)/(aH(z))$ is the line-of-sight component of the real-space distance r , $a = (1 + z)^{-1}$ is the scale factor, and $H(z)$ is the Hubble parameter evaluated at redshift z . The full $\Sigma(\pi)$ analytic model, as a function of v_σ and β , is obtained by averaging equation (18) in the range $0.5 \leq r_p \leq 2 h^{-1}$ Mpc and integrating the result in π bins, as explained in Section 2.1.

Combining these definitions and matching the binning in Δr_p and $\Delta\pi$, we have:

$$\Sigma(\pi) = \frac{\int \frac{dZ}{\Delta\pi} \int \frac{dR}{\Delta r_p} \int \xi\left(R, \frac{Z-w}{aH(z)}\right) f(w) dw}{\int \frac{dZ}{\Delta\pi} \int \frac{dR}{\Delta r_p} \left(\frac{r_p^2}{R^2 + Z^2}\right)^{\gamma/2}} \quad (19)$$

The FoG and Kaiser effects could be overlapping and, as fit parameters in a model, they are correlated. The importance of our modelling is not to isolate their value, but to differentiate between models and data with subpopulations of galaxies. Fig. 2 shows how both effect contribute to modulate our $\Sigma(\pi)$ model. There is a degeneracy between the parameter values, in the sense that both increasing v_σ or reducing β produces an enhancement in the $\Sigma(\pi)$ peak. This dependence prevents us from interpreting the β parameter as the only one responsible of the $\Sigma(\pi)$ amplitude.

2.7 Fitting $w_p(r_p)$

To implement the integral in equation (2), to estimate the projected correlation function $w_p(r_p)$, we need to truncate it at some upper value, π_{max} , above which the contribution to correlation function becomes negligible. If one includes very large scales, the measurement will be affected by noise; inversely, if we consider only very small scales, the clustering amplitude will be underestimated. In our

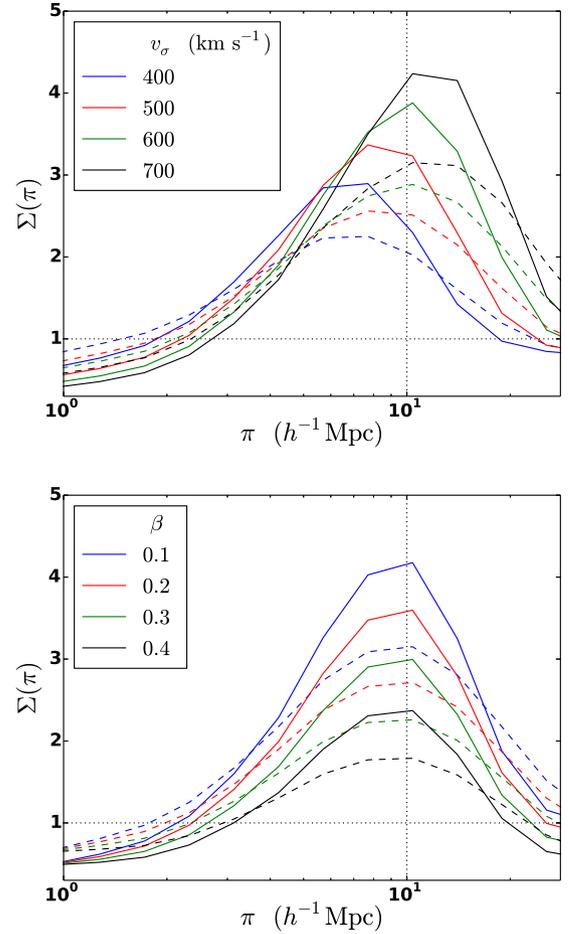


Figure 2. $\Sigma(\pi)$ analytic model as a function of the pairwise velocity dispersion, v_σ (top panel) and the Kaiser parameter, β (bottom panel). Solid lines represent the Gaussian model given in equation (17); dashed curves are the exponential functions in equation (16). We choose to model our $\Sigma(\pi)$ measurements using the normal functional form only, since it reproduces more accurately the small-scale feature provoked by the FoG distortions and peak at larger scales.

case, CMASS results are not sensitive to $\pi \geq 80 h^{-1}$ Mpc, therefore we adopt this value as our π_{max} limit. The projected auto-correlation function is related to the real-space one by (Davis & Peebles 1983)

$$w_p(r_p) = 2 \int_{r_p}^{\pi_{\text{max}}} \frac{r \xi(r)}{\sqrt{r^2 - r_p^2}} dr. \quad (20)$$

Zehavi et al. (2005b) demonstrates that for a generic power law, $\xi(r) = (r/r_0)^\gamma$, the equation above can be written in terms of the Euler's Gamma function as

$$w_p(r_p) = r_p \left(\frac{r_p}{r_0}\right)^\gamma \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{\gamma-1}{2}\right) / \Gamma\left(\frac{\gamma}{2}\right). \quad (21)$$

allowing one to infer the best-fitting power law for $\xi(r)$ from $w_p(r_p)$, corresponding to the full CMASS galaxy sample, blue and red subsamples. Fig. 3 presents the power-law fits to the full, red and blue CMASS projected correlation functions, and the resulting (r_0, γ) optimal values.

3 BOSS CMASS DATA

BOSS target galaxies primarily lie within two main samples: CMASS, with $0.43 < z < 0.7$ and LOWZ, with $z < 0.43$ (Anderson

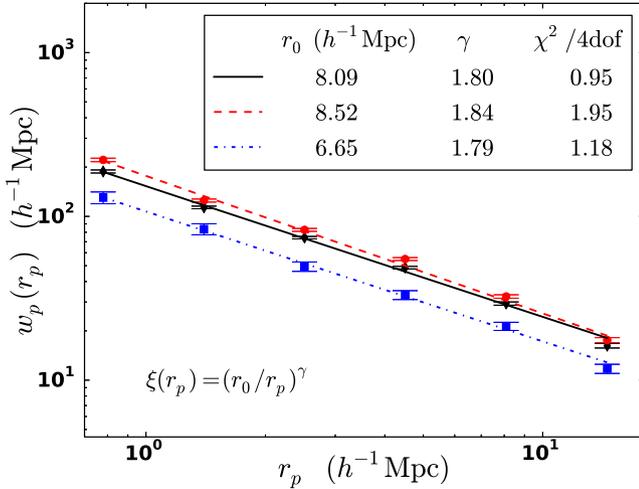


Figure 3. Power-law fits to the CMASS full, red and blue projected correlation functions, which define the denominator in equation (19). The r_0 and γ values we find are consistent with Zehavi et al. (2005a), and show that red galaxies cluster more than blue star-forming ones. The error bars correspond the 1σ uncertainties estimated using 200 jackknife resamplings (Section 2.3).

et al. 2012; Bolton et al. 2012; Ross et al. 2012). These samples are selected on the basis of photometric observations done with the dedicated 2.5-m Sloan Telescope (Gunn et al. 2006), located at Apache Point Observatory in New Mexico, using a drift-scanning mosaic CCD camera with five colour-bands, *ugriz* (Fukugita et al. 1996; Gunn et al. 1998). Spectra of the LOWZ and CMASS samples 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). Spectroscopic redshifts are then measured using the minimum- χ^2 template-fitting procedure described in Aihara et al. (2011), with templates and methods updated for BOSS data as described in Bolton et al. (2012).

We select galaxies from CMASS DR11 (Alam et al. 2015) – North plus South Galactic caps – which is defined by a series of colour cuts designed to obtain a galaxy sample with approximately constant stellar mass. Specifically, these cuts are:

$$17.5 < i_{\text{cmod}} < 19.9 \quad (22)$$

$$r_{\text{mod}} - i_{\text{mod}} < 2 \quad (23)$$

$$d_{\perp} > 0.55 \quad (24)$$

$$i_{\text{fib2}} < 21.5 \quad (25)$$

$$i_{\text{cmod}} < 19.86 + 1.6(d_{\perp} - 0.8), \quad (26)$$

where i_{cmod} is the *i*-band cmodel magnitude. The quantities i_{mod} and r_{mod} are model magnitudes, i_{fib2} is the *i*-band magnitude within a 2 arcsec aperture and d_{\perp} is defined as

$$d_{\perp} = r_{\text{mod}} - i_{\text{mod}} - (g_{\text{mod}} - r_{\text{mod}})/8.0. \quad (27)$$

All the magnitudes are corrected for Galactic extinction using the dust maps from Schlegel, Finkbeiner & Davis (1998). In addition to the above colour cuts, CMASS objects must also pass two star-galaxy separation constraints:

$$i_{\text{psf}} - i_{\text{mod}} > 0.2 + 0.2(20.0 - i_{\text{mod}}) \quad (28)$$

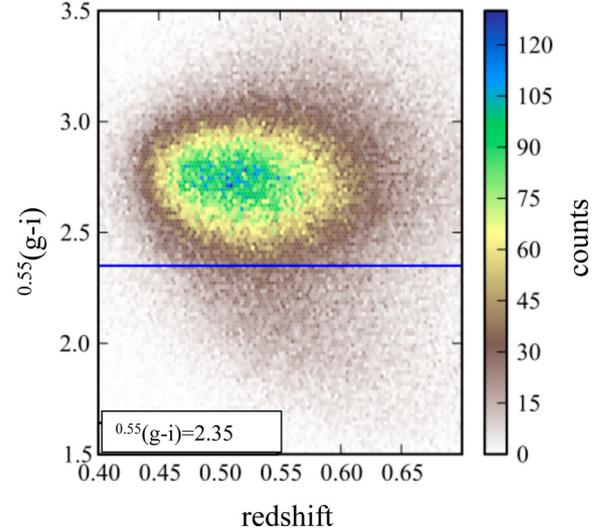


Figure 4. BOSS CMASS DR11 colour selection: the $(g - i)$ colour cut divides the full sample into a red dense population (above the blue horizontal line) and a sparse blue tail (below the line).

$$z_{\text{psf}} - z_{\text{mod}} > 9.125 - 0.46z_{\text{mod}}, \quad (29)$$

unless the objects also pass the LOWZ criteria. Therefore, to distinguish CMASS from LOWZ candidates, it is necessary to select them by redshift.

3.1 Colour selection

The CMASS sample is mainly composed of massive, luminous, red galaxies, which are favourite subjects to study galaxy clustering. Among them, however, there is an intrinsic bluer, star-forming population of massive galaxies (Guo et al. 2012; Ross et al. 2012;), of which little is known. In the attempt to explore this bluer component to understand its contribution in the clustering properties, we split the CMASS sample into its blue and red components by applying the colour cut

$${}^{0.55}(g - i) = 2.35 \quad (30)$$

constant in redshift and K -corrected to the $z = 0.55$ rest-frame using the code by Blanton & Roweis (2007). Masters et al. (2011) applied this same colour cut, with no K -corrections, to the BOSS CMASS DR8 sample to study the morphology of the LRG population; Ross et al. (2014) used a similar selection, ${}^{0.55}(r - i) = 0.95$, to measure galaxy clustering at the BAO scale in CMASS DR10. Fig. 4 presents our CMASS colour selection, splitting the full sample into a red denser population (above the blue horizontal line) and a sparse blue tail (below the line), whose completeness dramatically increases when we move towards high-redshift values ($z > 0.55$). For our analysis, we focus on the high-redshift tail of the CMASS sample, selecting only galaxies with redshift beyond $z > 0.55$.

3.2 Weights

Due to its structural features, a survey inevitably introduces some kind of spatial variation in its measurements. To avoid these distortions, we weight our pair counts by defining a linear combination of four different weights (Anderson et al. 2012; Ross et al. 2012; Sánchez et al. 2012):

$$w_{\text{tot}} = w_{\text{FKP}} w_{\text{sys}}(w_{\text{fc}} + w_{\text{zf}} - 1), \quad (31)$$

each one correcting for a different effect. In the expression above, w_{zf} accounts for targets with missing or corrupted redshift (z failure); w_{fc} corrects for fibre collision, compensating the fact that fibres cannot be placed closer than 62 arcsec on the survey plates. This limitation prevents obtaining spectra of all galaxies with neighbours closer than this angular distance in a single observation. The default value of w_{zf} and w_{fc} is set to unity for all galaxies. When a fibre collision is detected, we increment by one the value of w_{fc} for the first neighbour closer than 62 arcsec. In the same way, for the neighbour we increase by one the value of w_{zf} of the nearest galaxy with a good redshift. To minimize the error in the measured clustering signal, we also require a correction based on the redshift distribution of our sample, namely the w_{FKP} factor Feldman, Kaiser & Peacock (1994), that weights galaxies according to their number density, $n(z)$. It is defined as

$$w_{FKP} = \frac{1}{1 + n(z)P_{FKP}}, \quad (32)$$

where P_{FKP} is a constant that roughly corresponds to the amplitude of the CMASS power spectrum $P(k)$, at $k = 0.1 h \text{ Mpc}^{-1}$. We assume $P_{FKP} = 2 \times 10^4 h^3 \text{ Mpc}^{-3}$, in Anderson et al. (2012). The last weight, w_{sys} , accounts for a number of further systematic effects that could cause spurious angular fluctuations in the galaxy target density. These effects are treated in detail in Ross et al. (2012), but we do not include them in this analysis, since they are not relevant at the scales considered in this work. Therefore we set in $w_{sys} = 1$ in the following analysis.

4 MODELLING FULL CMASS SAMPLE

4.1 Full CMASS clustering

We construct an HOD model using MultiDark haloes and subhaloes (see model description in Section 2.5), and produce a mock galaxy catalogue which we compare to the full CMASS DR11 population. This mock is built by varying the HOD parameters to match $\xi(s)$, populating the MD simulation in each step, and using the peculiar velocities in the simulation to model redshift-space distortions. The intention is that changing the HOD will constrain the overall galaxy bias, hence we fit only one statistic. We then evaluate and further investigate these fits over the three clustering metrics: $\xi(s)$, $w_p(r_p)$ and $\Sigma(\pi)$.

However, since implementing a formal fit to determine the optimal HOD parameters is beyond the scope of this work, we improve the matching empirically, changing the input values until we find a suitable ($\log M_{\min}$, M_0 , M'_1 , α' , $\sigma_{\log M}$) set that reproduces the observed $\xi(s)$ amplitude. We fit only M_{\min} (the minimum halo mass), M'_1 (the mass scale of the satellite cut-off profile) and α' (the satellite slope). The remaining parameters are fixed to their default values given by White et al. (2011): $\log M_0 = 12.8633$, $\sigma_{\log M} = 0.5528$. The specific choice of these three parameters arises from their connection to two physical quantities we want to measure: (i) the satellite fraction, f_{sat} , that controls the slope of the one-halo term at small scales, where substructures of the same halo dominate; (ii) the galaxy number density, $n(z)$, affecting the two-halo term at larger scales, where correlations between substructures of different hosts become appreciable. Fig. B1 in the appendix illustrates how a change in M_{\min} , M'_1 and α' affects the projected correlation function.

Fig. 5 displays the redshift-space monopole corresponding to our empirical best fit ($\chi^2 = 11.08/7$ dof including the full covariance

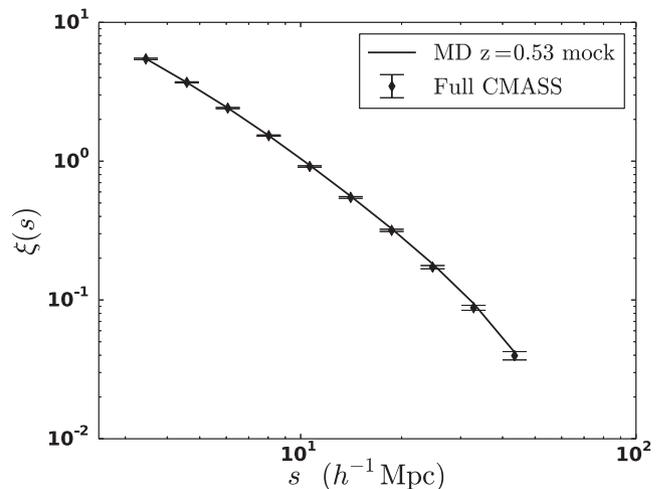


Figure 5. Redshift-space monopole correlation functions of our $z = 0.53$ MultiDark full mock galaxy catalogue (solid line) compared to BOSS CMASS DR11 measurements. Error bars are estimated using 200 jackknife regions.

Table 1. Our best empirical estimates of the HOD parameters for the total, red and blue independent models of the CMASS populations. We obtain these values only by fitting $\xi(s)$ with a three-dimensional grid in $\log M_{\min}$, $\log M'_1$ and α' . The resulting χ^2 values are: 11.08/7dof (full CMASS), 13.54/7dof (red) and 14.91/7dof (blue).

	Total	Red	Blue
$\log M_{\min}$	13.00	13.10	12.50
$\log M'_1$	13.30	13.02	13.85
α'	0.20	0.22	0.15
f_{sat}	0.27	0.33	0.11
$\langle \log M_h \rangle$	12.75	13.00	12.50

matrix computed with jackknife; the HOD parameters are given in Table 1) mock galaxy catalogue from the MultiDark simulation. The projected correlation function, $w_p(r_p)$, and the line-of-sight statistic, $\Sigma(\pi)$, corresponding to this model are shown in Fig. 6. In agreement with many previous works (Zehavi et al. 2004, 2005b; Guo et al. 2012), we find that CMASS galaxies are more highly clustered at small scales (one-halo regime); then, as the spatial separation between the pairs increases, the clustering strength drops (two-halo term). Compared to White et al. (2011), our best-fitting mock has a much lower satellite slope, α' , and M'_1 , resulting in a higher satellite fraction (about 27 per cent); however, our mean satellite occupation function is compatible with results from Guo et al. (2015). Overall, the amplitude of our model galaxies is in good agreement with observations. Error bars are estimated using 200 jackknife regions gridded in right ascension and declination as follows: 10 RA \times 15 DEC cells for the CMASS North Galactic Cap ($N_{\text{res}} = 150$), plus 5 RA \times 10 DEC regions for the South Galactic Cap, ($N_{\text{res}} = 50$). This approach produces 200 equal areas of about 100 deg^2 each.

To compute $\Sigma(\pi)$ through equation (4) for the full CMASS galaxy sample and MD mock, we use the best-fitting power law to their projected correlation functions, $w_p(r_p)$. The relative r_0 and γ estimates are shown in Fig. 3. Beyond 8–10 $h^{-1} \text{ Mpc}$, where the Kaiser squashing becomes predominant, the jackknife uncertainties on $\Sigma(\pi)$ are wider. This measurement reveals that the deviation

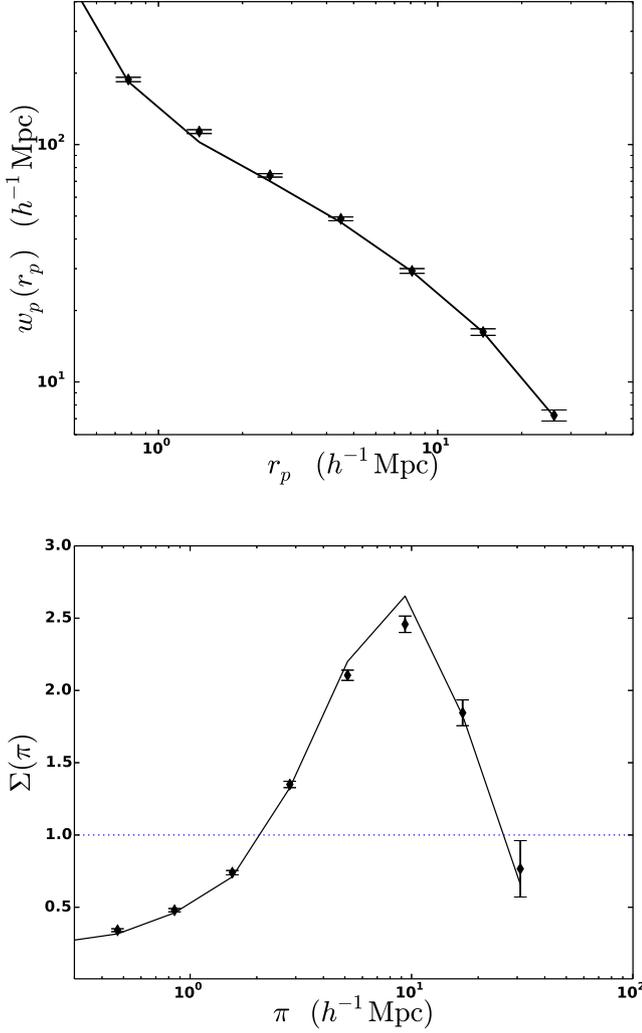


Figure 6. Projected correlation function (top) and $\Sigma(\pi)$ (bottom) for the $z = 0.53$ MultiDark full mock galaxy catalogue (solid line), compared to BOSS CMASS DR11 measurements. Error bars are estimated using 200 jackknife regions containing the same number of randoms.

of $\xi(\bar{r}_p, \pi)$ from the real-space behaviour dramatically changes according to the scale of the problem: at very small redshift separations i.e. $\pi \leq 2 h^{-1}$ Mpc, where the FoG dominates, the contribution of peculiar velocities pushes $\Sigma(\pi)$ below unity. Above $3 h^{-1}$ Mpc, $\Sigma(\pi)$ increases sharply and peaks around $8 h^{-1}$ Mpc. On larger scales, the correlation between pairs of galaxies is compressed along the line of sight since the Kaiser infall dominates and $\Sigma(\pi)$ drops. The $\Sigma(\pi)$ measurement shows very different and characteristic features according to the scale of interest, therefore it is a valuable tool to quantify both small and large-scale clustering effects.

4.2 Modelling redshift-space distortions and galaxy bias

In redshift-space, two different distortion features are observed: the FoG effect which dominates below $2 h^{-1}$ Mpc, and the Kaiser flattening, which becomes important beyond $10\text{--}15 h^{-1}$ Mpc. These phenomena preferentially manifest themselves on different scales, but a certain degree of entanglement is unavoidable in both regimes. In order to better separate the two effects, we examine $\Sigma(\pi)$ in our

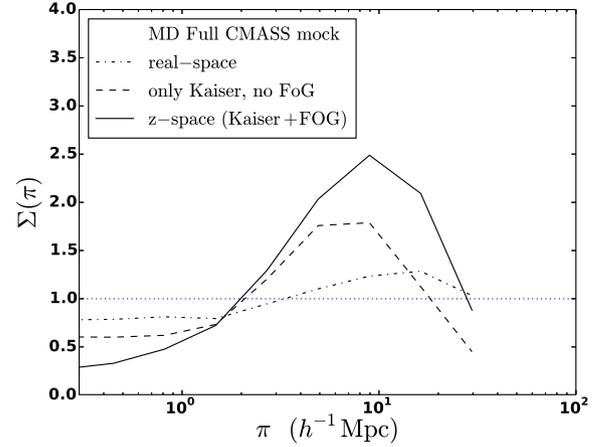


Figure 7. $\Sigma(\pi)$ in real-space (dot-dashed line), redshift-space with only Kaiser contribution (dashed) and Kaiser plus FoG (solid). As expected, the real-space behaviour is close to unity at all scales.

MultiDark full mock catalogue in three different configurations: real-space, redshift-space with only Kaiser effect and full redshift-space (FoG+Kaiser), as shown in Fig. 7. The real-space $\Sigma(\pi)$ is defined in equation (4), omitting the peculiar velocities both in the numerator and in $w_p(r_p)$ to which we fit the power law at the denominator. Since $\Sigma(\pi)$ is the ratio between two spherically averaged power laws, we expect it to be close to unity at all scales. Hence, the dot-dashed line in Fig. 7 is compatible with expectations. The redshift-space case with only Kaiser contribution (dashed line) is computed by assigning satellite galaxies their parental v_{pec} value. In this way, each satellite shares the coherent motion of its parent, but it does not show any random motion with respect to it. The last case considered is the full redshift-space $\Sigma(\pi)$ (solid line), in which satellite galaxies have their own peculiar velocity, which is independent from their parents.

We are now able to provide a full description of our $\Sigma(\pi)$ results by modelling them through equation (19), in terms of four parameters: the power-law correlation length, r_0 , its slope γ , the pairwise velocity dispersion, v_σ and the β parameter, which is inversely proportional to the linear galaxy bias, b , through equation (14).

The linear galaxy bias is scale dependent and has been computed (e.g. Nuza et al. 2013) as the ratio between the galaxy and matter correlation functions,

$$b(s) = \sqrt{\frac{\xi(s)}{\xi_m(s)}}. \quad (33)$$

Our goal is to provide an estimate of both the peculiar velocity field causing the distortions we observe in redshift-space in our clustering measurements and the large-scale bias, using the v_σ , β values we find from our full, red and blue CMASS and MultiDark $\Sigma(\pi)$ modelling. To this purpose, we do not compute the bias as Nuza et al. (2013), through equation (33), but we estimate it from equation (14).

Fig. 8 displays the v_σ , β models (dashed curves) for our CMASS measurements (left-hand panel, squares) and full MultiDark mock catalogue (right-hand panel, crosses). All the model fits are performed including the full covariance matrix, estimated by using 200 jackknife re-samplings (Section 4.1). For the MultiDark mock, we assume the same scatter of the CMASS data.

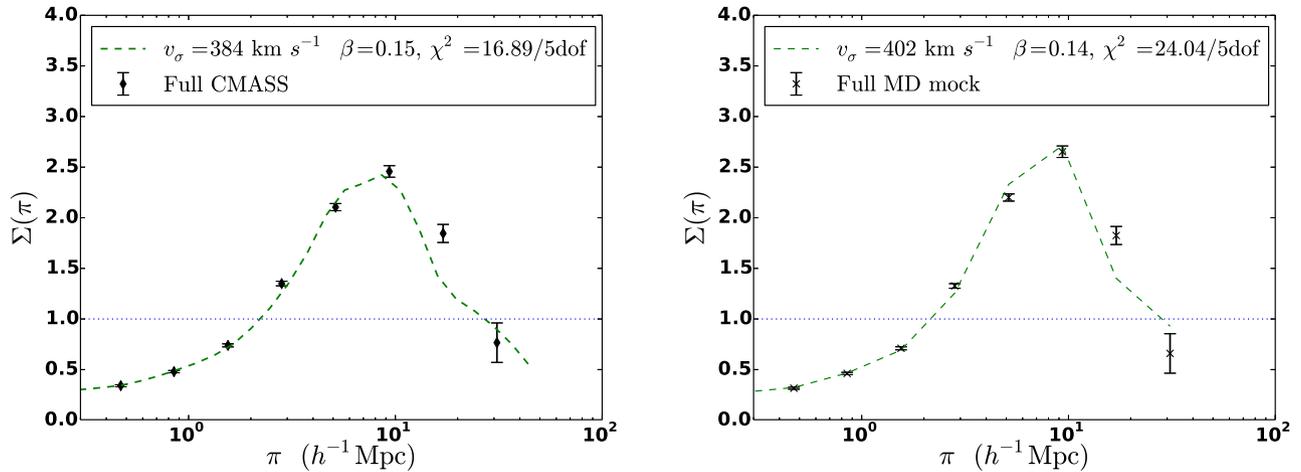


Figure 8. $\Sigma(\pi)$ full CMASS DR11 measurement (left-hand panel) and our MultiDark $z = 0.53$ mock (right-hand panel), versus their v_σ, β analytic model (dashed lines). For both data and mock sets we assume the errors are given by our jackknife estimate, computed using 200 resamplings. The fits are performed by using the full covariance matrix. These plots reveal that the full CMASS sample and the MultiDark model galaxies share almost the same large-scale bias value, while the peculiar velocity contribution is higher in the mocks.

Table 2. Best-fitting values of the v_σ, β parameters that model $\Sigma(\pi)$ in both full, red, blue CMASS measurements and MultiDark mocks. All the fits are fully covariant. The bias is computed using the approximation given in equation (14).

	v_σ (km s $^{-1}$)	β	b	χ^2
Full CMASS	384 ± 6	0.15 ± 0.01	~ 3	16.89/5dof
Full mock	402_{-6}^{+9}	$0.14_{-0.02}^{+0.01}$	~ 3	24.04/5dof
Red CMASS	402_{-9}^{+8}	$0.15_{-0.02}^{+0.01}$	~ 3	24.00/5dof
Red mock	432_{-8}^{+10}	0.13 ± 0.01	~ 3.5	27.21/5dof
Blue CMASS	364_{-39}^{+47}	$0.21_{-0.04}^{+0.05}$	~ 2	8.14/5dof
Blue mock	268 ± 35	$0.16_{-0.09}^{+0.07}$	~ 2.8	2.61/8dof

Adopting a normal function (equation 17) to mimic the contribution of peculiar velocities (see Table 2), results in the MD model galaxies that have slightly higher bias – which means a lower β value – than the full CMASS population and higher peculiar velocity contribution – higher v_σ value. This result is in agreement with the bottom panel of Fig. 6: CMASS data points (diamonds) experience a stronger Kaiser squashing at $\sim 10 h^{-1}$ Mpc i.e. they have a smaller large-scale bias, compared to the MultiDark model galaxies (solid line). From these v_σ, β values, we conclude that our full MD mock catalogue can be considered a reliable representation of the full CMASS sample.

The reduced χ^2 values we derive from the full CMASS and MultiDark $\Sigma(\pi)$ model fits are less stringent compared to the estimates for $\xi(s)$ reported in the caption of Table 1. The main reason for this result resides in how we build the v_σ, β model (see equation 19) to reproduce the $\Sigma(\pi)$ feature, which is the ratio of a 2PCF, spherically averaged in the range $0.5 \leq r_p < 2 h^{-1}$ Mpc, over a real-space term. To mimic this average in our model, we convolve (numerator in equation 19) a real-space correlation function with a peculiar velocity term, $f(w)$, and integrate the result to eliminate the dependence on r_p . Such an integration is performed numerically in (r_p, π) bins, by stacking r_p into a single average value per bin. The denominator in equation (19) is a real-space term, given by best-fitting power law to $w_p(r_p)$, spherically averaged in the same way as the numerator.

Thus, the v_σ, β model reproduces the $\Sigma(\pi)$ measurement numerically in bins of (r_p, π) and not analytically in each point. The approximations adopted to define our $\Sigma(\pi)$ model are justified by the fact that our goal is to provide a qualitative prediction of the linear bias and redshift-space distortions in the full, red and blue CMASS samples. For this reason, we do not heavily focus on the goodness of our model fits, but instead stress the importance of a cross-comparison in terms of v_σ, β values.

From the full CMASS model, we find a bias of $b \sim 3$, which is relatively high compared to the estimate reported in Nuza et al. (2013), $b \sim 2$. This discrepancy is due to the fact that we select only the massive bright high-redshift tail (i.e. $z > 0.55$) of the CMASS sample; for these specific galaxies the bias is expected to be higher than in Nuza et al. (2013).

4.3 Full CMASS covariance

We compute the full CMASS jackknife covariance matrix for the three metrics of interest using equation (9), in which ξ is either $\xi(s)$, $w_p(r_p)$, or $\Sigma(\pi)$. We estimate the goodness of our model fits to the CMASS measurements by computing the relative χ^2 values as

$$\chi^2 = \mathbf{A}^T \mathbf{C}_*^{-1} \mathbf{A}, \quad (34)$$

where $\mathbf{A} = (\xi_{\text{data}}^i - \xi_{\text{model}}^i)$ is a vector with $i = 1, \dots, n_b$ components and \mathbf{C}_*^{-1} is an unbiased estimate of the inverse covariance matrix (Hartlap, Simon & Schneider 2007; Percival et al. 2014),

$$\mathbf{C}_*^{-1} = (1 - D)\mathbf{C}^{-1}, \quad D = \frac{n_b + 1}{N_{\text{res}} - 1}. \quad (35)$$

In the equation above, n_b is the number of observations and N_{res} the number of jackknife re-samplings. For the full CMASS population, the correction factor $(1 - D)$ represents a 8 per cent effect on the final χ^2 value.

In Appendix D, we test our jackknife error estimates using a set of 100 Quick Particle Mesh (QPM; White, Tinker & McBride 2014) galaxy mock catalogues.

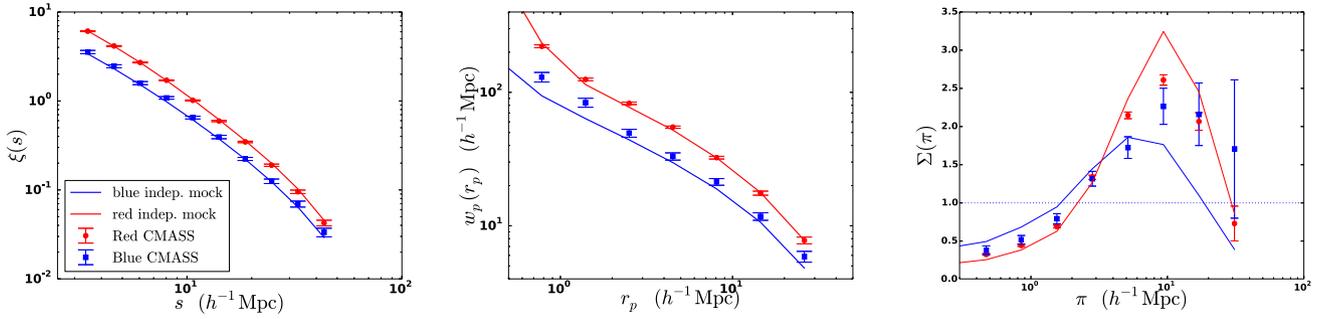


Figure 9. Independent mock catalogues designed to model CMASS DR11 red and blue $\xi(s)$, $w_p(r_p)$ and $\Sigma(\pi)$ measurements (points and squares). The error bars are the 1σ regions estimated using 200 jackknife re-samplings of the data. Despite we fit only $\xi(s)$, we find good agreement between data and mocks in all our three statistics. As expected, red galaxies show a higher clustering amplitude compared to the blue population.

5 MODELLING COLOUR SUBSAMPLES

We repeat the same analysis described in Section 4 on the red and blue colour subsamples. We first use $\xi(s)$ to fit an HOD and match the overall clustering, then use our analytic model to obtain fits for v_σ and β . There remains a question on how to model the subpopulations in the mocks; we explore two methods.

5.1 Independent red and blue models

For simplicity, our first attempt at the colour subsamples is to individually model the red and the blue CMASS populations. That is, we assume the clustering comes from a complete sample and we generate an HOD populating haloes independently of whether a galaxy is red or blue. By definition, there is no connection in the overlap and the same halo or subhalo could host either a red and blue galaxy in the corresponding mocks. This is an over-simplified view, as clearly a galaxy can be either red or blue and not both. However, it is an assumption that is embedded within several related analyses (Zehavi et al. 2004, 2005b; Guo et al. 2012, 2014).

Fig. 9 shows the agreement between the CMASS monopole, projected 2PCF and $\Sigma(\pi)$ measurements and our independent red and blue model galaxies. Our empirical best-fitting HOD parameter values are reported in Table 1, together with the satellite fraction; the fraction is higher for red than for blue galaxies, confirming that luminous red galaxies tend to live in a denser environment (Zehavi et al. 2005b; Wang et al. 2007; Swanson et al. 2008). We conclude that we are able to fit correctly all our red and blue CMASS clustering results, by means of the same HOD technique, with small variations in its input parameters. However, these red and blue independent models are non-physical, because they allow the same galaxy to be either red or blue. In other words, they place both red and blue galaxies in the same hosting haloes, which is not the case.

To overcome this problem, we propose an alternative HOD approach (see next section) in which the red and the blue models are obtained by splitting the full mock catalogue into subpopulations that match the observed red/blue CMASS galaxy fractions. In this way, the red and blue model galaxies are no longer independent and, by construction, they cannot occupy the same positions in a given halo.

5.2 Splitting colour samples using galaxy fractions

Inspired by the result in the previous section, we develop a more physically motivated model of red/blue colour separation. In line

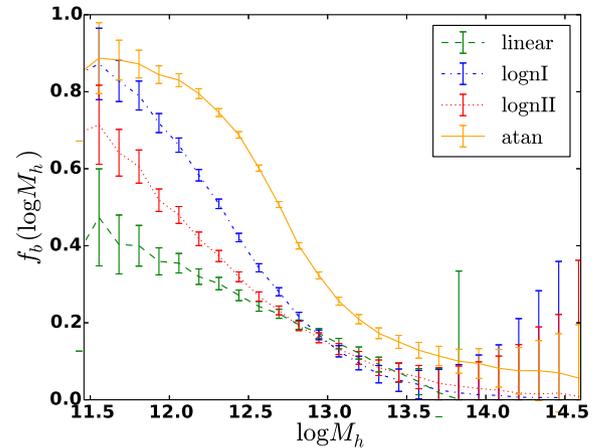


Figure 10. Blue galaxy fraction models, f_b , and the corresponding Poisson error, as a function of the central halo mass: linear (dashed line), log-normal I (dot-dashed), log-normal II (dotted), inverse tangent (solid). The red galaxy fractions are recovered by $f_r = 1 - f_b$.

with the standard halo model, we explore a splitting method based entirely on host halo mass, with each of them matching the corresponding observed CMASS galaxy fraction. By modelling these red/blue fractions, $f_{b,r}$, as a function of the central halo mass, we are able to correlate the red and the blue mock catalogues to the full one, reducing the number of free parameters from 15 (5 for each independent HOD) to 5 (full HOD) plus 2 (constraint on galaxy fractions). Our galaxy fraction model must verify two conditions: (i) to obtain reliable results, the models must reproduce the overall $f_{b,r}$ values observed in our CMASS red/blue selection; this is done by requiring that

$$\sum_{i=1}^N f_b(\log M_h(i)) / N = 0.25,$$

$$f_r(\log M_h) = 1 - f_b(\log M_h) = 0.75 \quad (36)$$

where we allow 20 per cent of scatter, and (ii) the red (blue) fraction must approach zero at low (high) mass scales. We build our theory as a function of the central halo mass only, omitting the dependence on satellite masses. Despite this simplifying assumption, the resulting red and blue mocks match correctly the observed clustering amplitude. To mimic the red/blue split, we test different functional forms of $f_{b,r}$, starting with a basic linear one (Fig. 10, dashed line) and two different log-normal models (dot-dashed and dotted curves) with three degrees of freedom each; they are treated in detail in Appendix C. In order to produce a clear separation between the two

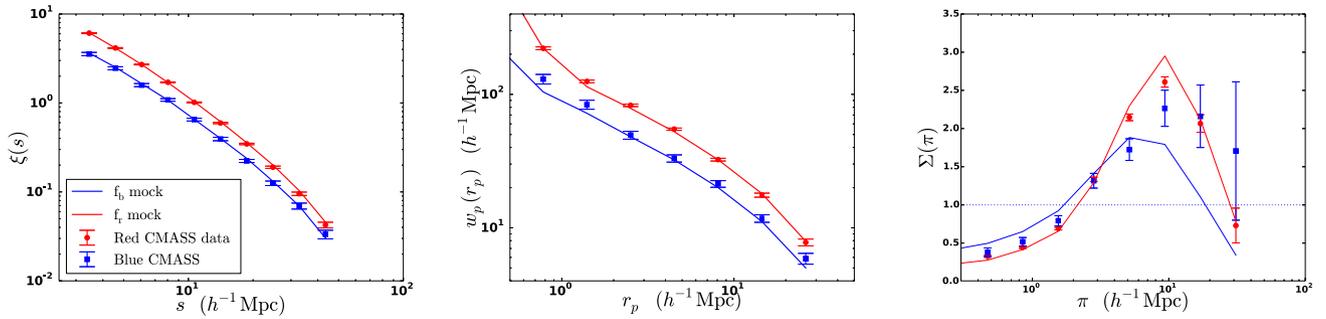


Figure 11. CMASS DR11 red and blue clustering measurements versus mocks. The models are obtained by splitting the full MultiDark mock into its red and blue components, matching the observed CMASS red/blue galaxy fraction, $f_{b,r}$. In this way, we prevent the same mock galaxy to be either red or blue, and guarantee the reliability of the model. We find good agreement between the CMASS measurements and our MultiDark mocks, and confirm that red galaxies reside in more dense environments compared to the blue population.

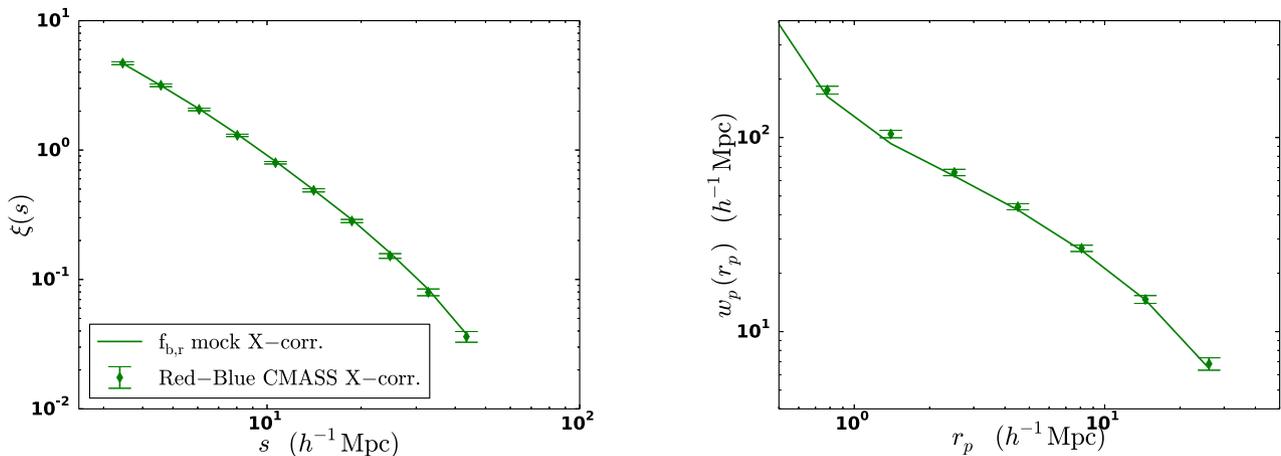


Figure 12. Red–blue CMASS DR11 (diamonds) versus inverse tangent mock (lines) cross-correlation functions. These plots are useful to check the mutual behaviour of the red and the blue CMASS samples. In fact, as expected, we find that the cross-correlation of these galaxies lies in between their auto-correlation functions, and the size of the error bars (computed with 200 jackknife resamplings) is consistent with the uncertainties on their individual clustering measurements.

populations, the best compromise is an inverse tangent-like function (solid line), with only two free parameters. The resulting functional form, as a function of the central halo mass, is

$$f_b(\log M_h) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left[\frac{\log M_h - D}{10^C} \right],$$

$$f_r(\log M_h) = 1 - f_b(\log M_h) \quad (37)$$

where the parameter C determines how rapidly the blue fraction drops and D establishes the half-width of the curve. Applying equations (36), (37) to the full CMASS mock catalogue, we select the (C, D) combination that globally best fits the observed red and blue redshift-space auto-correlation functions, $\xi(s)$. The best-fitting values are $C = -0.50$, $D = 12.50$, with $\chi_{\text{red}}^2 = 15.43/5\text{dof}$, $\chi_{\text{blue}}^2 = 6.20/5\text{dof}$ and $\chi_{\text{tot}}^2 = 10.82/10\text{dof}$. We use these red and blue inverse tangent mocks to match the other two statistics, $w_p(r_p)$ and $\Sigma(\pi)$, which are shown in Fig. 11 and the cross-correlation functions in Fig. 12. The $\xi(s)$ fit is performed using the full covariance matrix and the uncertainties are estimated via jackknife resampling (Section 2.3). The red and blue MultiDark models for the monopole 2PCF – both the independent (left-hand panel, Fig. 9) and the $f_{b,r}$ mocks (left-hand panel, Fig. 11) – show a remarkable agreement with the CMASS measurements. In the projected correlation function (central panels, same figures), we find

overall good agreement between the red mocks and the CMASS data points. The blue models show, instead, some deviations in $w_p(r_p)$: the blue independent model under predicts the clustering, and the deviation is larger towards smaller scales ($r_p \lesssim 2 h^{-1}$ Mpc); the blue $f_{b,r}$ model reveals good agreement with the CMASS results at $r_p \gtrsim 2.5 h^{-1}$ Mpc, and slightly deviates on smaller scales. For what concerns $\Sigma(\pi)$, in general we find less agreement with the observations, compared to the previous measurements. As already discussed in Section 4.2, this is mostly due to the fact that $\Sigma(\pi)$ is a quantity built from two-point correlation functions using several approximations i.e. it is a ‘derived’ measurement. The $f_{b,r}$ red and blue $\Sigma(\pi)$ models (Fig. 11, right-hand panel) match better the CMASS data than the independent mocks (Fig. 9, right plot), especially in the range of the peak, $2 \lesssim \pi \lesssim 20 h^{-1}$ Mpc. The $\Sigma(\pi)$ model can, of course, be refined to improve the agreement with the observations; consider, however, that we are fitting our mocks only to the monopole CF. The models for $w_p(r_p)$ and $\Sigma(\pi)$ do not involve any fitting routine, and correspond to the monopole best-fitting mocks. The agreement we find between models and data even without a fit is therefore quite impressive. The main purpose of this paper is to provide a straightforward model to characterize the full, red and blue CMASS clustering both in terms of redshift-space distortions and linear bias, thus we do not focus on improving the $\Sigma(\pi)$ model construction.

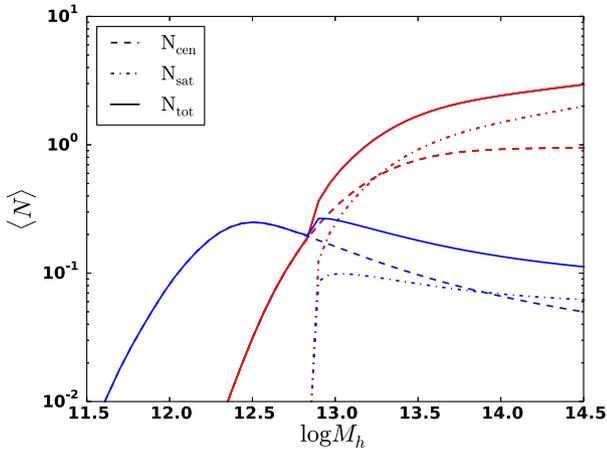
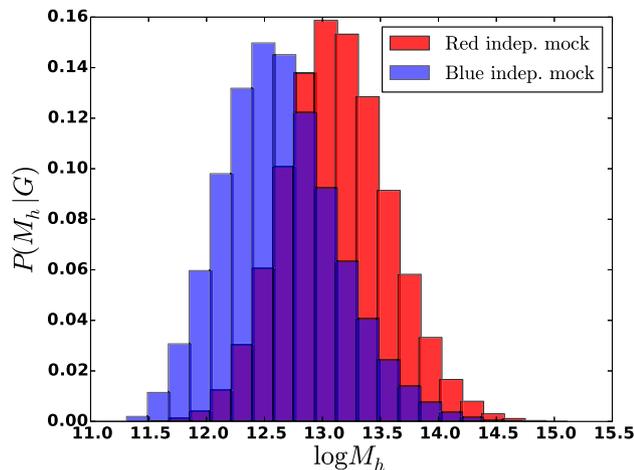


Figure 13. Red and blue HOD models obtained by applying the galaxy red/blue fraction condition to the MultiDark mock catalogue for the full CMASS population. The lines are the predictions computed by normalizing $\langle N_c \rangle$, $\langle N_s \rangle$, $\langle N_t \rangle$ by $f_{b,r}$. For red galaxies, the HOD shape is consistent with Fig. 1, confirming that the red/blue galaxy separation we are imposing with the satellite fraction constraint is reliable for the red population. For blue mocks, the expected average number of galaxies per halo mass is about 10 times less than for red ones at $\log M_h = 13.5$, and drops almost linearly as the halo mass increases. This reveals that blue star-forming galaxies preferentially populate low-mass haloes.

The cross-correlations between red and blue CMASS galaxies (see Fig. 12) behave similarly to the auto-correlation functions: they are stronger on small scales and weaker when the pair separation increases. These functions represent a consistency check of our red/blue fitting scheme and they provide robust information about red and blue galaxy bias: the younger and more star-forming is the galaxy, the lower are its clustering amplitude and bias.

Fig. 13 displays the red and blue HOD models inferred by splitting the full MultiDark mock using the observed CMASS red/blue galaxy fraction. The lines are the predictions computed normalizing $\langle N_c \rangle$, $\langle N_s \rangle$, $\langle N_t \rangle$ by $f_{b,r}$. For red galaxies the HOD shape is compatible with the model shown in Fig. 1, confirming that the red/blue separation we imposed with the galaxy fraction constraint is reli-



able for the red population. For blue mocks, the average number of galaxies per halo mass is ~ 10 times less compared to the red $\langle N_{\text{cen}} \rangle$, at $M_h = 10^{13.5} h^{-1} M_\odot$ and drops almost linearly (3 per cent factor) as the halo mass increases. Such a trend reflects the preference of blue star-forming galaxies to populate low-mass haloes. From this analysis, we estimate the conditional probability, $P(M_h | G)$, that a galaxy β with a specific colour is hosted by a central halo having mass M_h ; see Fig. 14. As expected, the result demonstrates that CMASS early-type redder galaxies are associated with more massive haloes ($M_h \sim 10^{13.1} h^{-1} M_\odot$), compared to the late-type bluer ($M_h \sim 10^{12.7} h^{-1} M_\odot$) companions.

6 RESULTS

6.1 Red and blue v_σ , β models

We apply the same v_σ , β modelling performed in Section 4.2 on the full CMASS sample and the MultiDark full mock galaxy catalogue to the red and blue data samples and $f_{b,r}$ mocks, to quantify how significant their differences are at the level of large-scale bias and redshift-space distortions. Our results are presented in Fig. 15: the top row displays the red and blue $\Sigma(\pi)$ CMASS measurements (points and squares), versus the analytic models (dashed lines); in the bottom row are the results for the red and blue MD mocks (solid lines), versus their models (dashed curves). For both CMASS data and MD mocks we assume the errors are given by our jackknife estimate, done using 200 resamplings. All the model fits are fully covariant and our best estimate of the v_σ , β parameters are reported in Table 2. As expected, the blue CMASS galaxies are less biased and have lower peculiar velocity contribution (i.e. smaller FoG elongation effect) compared to the red sample. A similar behaviour is seen in a comparison of the red and the blue MultiDark model galaxies, suggesting that we are correctly modelling our results in terms of redshift-space distortions and large-scale bias. Our relatively high bias values are due to the specific high-redshift CMASS selection we are considering. In fact, for CMASS galaxies at $z > 0.55$, the bias is expected to be higher than the typical value reported by Nuza et al. (2013), $b \sim 2$. As discussed in Section 4.2, the relatively high χ^2 values we find from our model fits are due to the numerical limitations in the $\Sigma(\pi)$ definition. However, since the goal of this work is a qualitative comparison of the full, red and

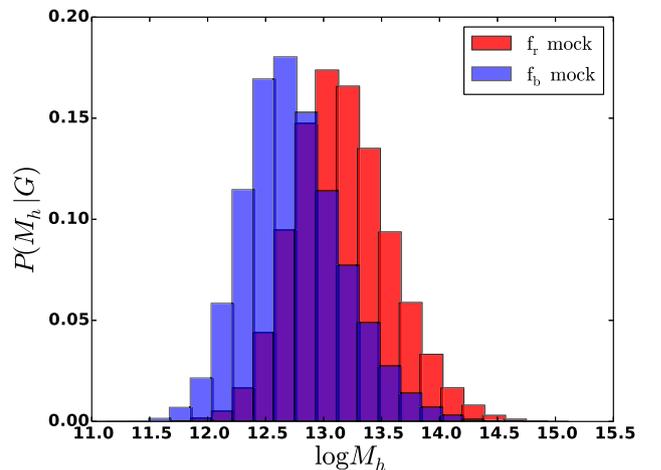


Figure 14. Conditional probability that a given galaxy β with a specific colour is hosted by a central halo with mass M_h obtained from our red and blue independent mock catalogues (left) and applying the galaxy fraction constraint (right). In both cases, as expected, we find that red galaxies live in more massive haloes compared to the blue ones.

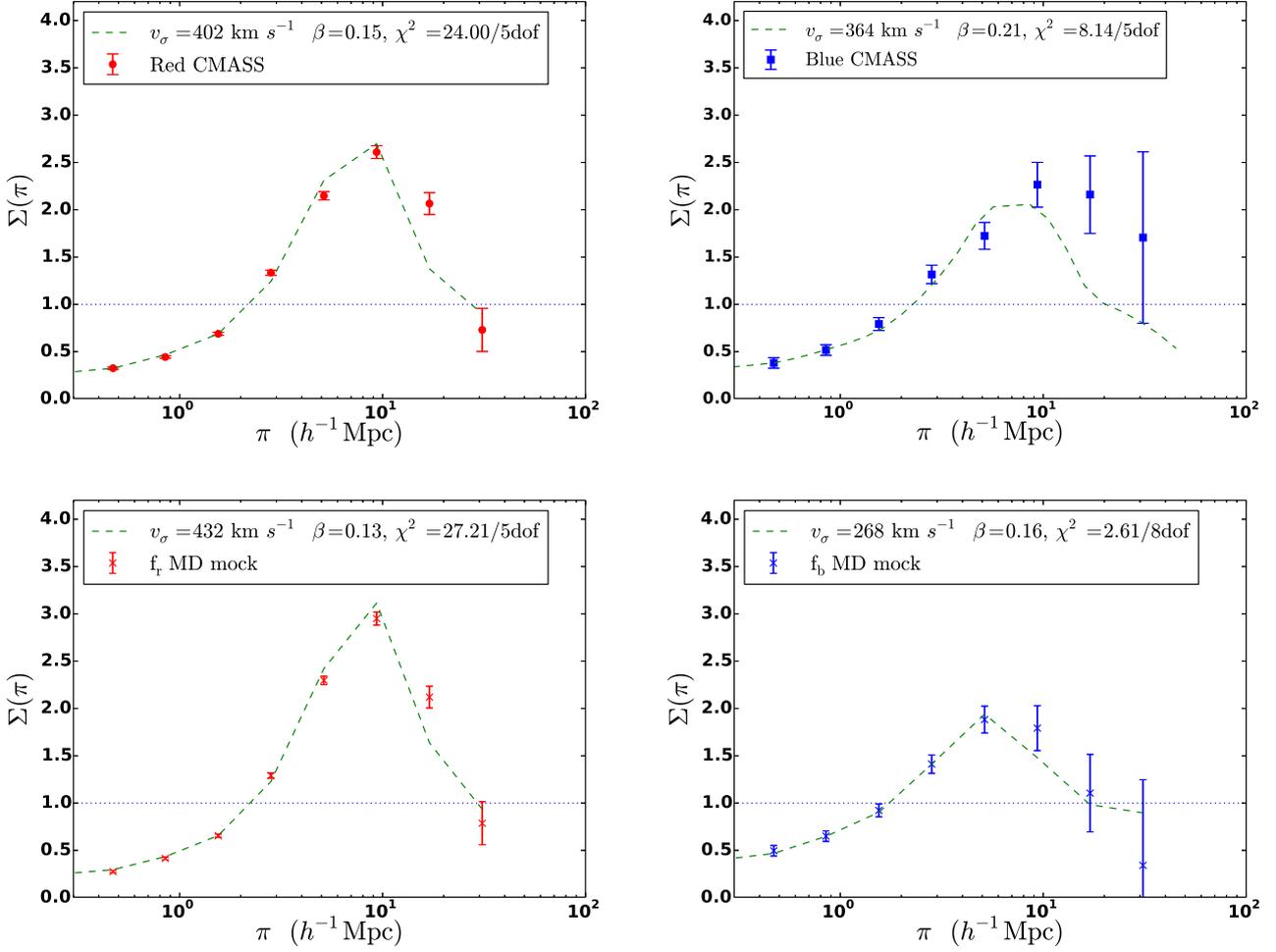


Figure 15. Top row: CMASS DR11 $\Sigma(\pi)$ red (left) and blue (right) measurements and the v_σ , β analytic models (dashed lines). Bottom row: $f_{b,r}$ MultiDark mocks (solid curves) and their models (dashed lines). For the mocks we adopt the jackknife errors estimated for the blue CMASS data doing jackknife. These fits are fully covariant. From these plots we conclude that blue CMASS galaxies are less biased and show a lower peculiar velocity contribution compared to the red population.

blue CMASS redshift-space clustering and bias features, we do not heavily focus on the goodness of the fits and give priority to the qualitative interpretation.

Fig. 16 presents the 68 per cent and 95 per cent covariant confidence regions of the v_σ , β models for the CMASS measurements. The 1σ blue region is spread out: due to their larger uncertainties, blue galaxies have less power to constrain the v_σ , β values compared to the red and full CMASS populations. The dots indicate the position of the best-fitting models for the three samples. As seen in Fig. 15, red CMASS galaxies possess higher velocity dispersion and large-scale bias compared to the blue sample.

6.2 Large-scale bias

The linear bias factor b , defined in equation (33), is related to the red–blue cross-correlation, $\xi_{\times}(s)$, by

$$b_r(s)b_b(s) = \frac{\xi_{\times}(s)}{\xi_m(s)}. \quad (38)$$

where the subscripts r, b indicate, respectively, red and blue galaxies, and $\xi_m(s)$ is the dark matter correlation function. We then expect that the ratio $\xi_{\times}(s)/\sqrt{\xi_r(s)\xi_b(s)}$ – where each term in the denominator is given by equation (33) – is close to unity. Fig. 17 shows that our

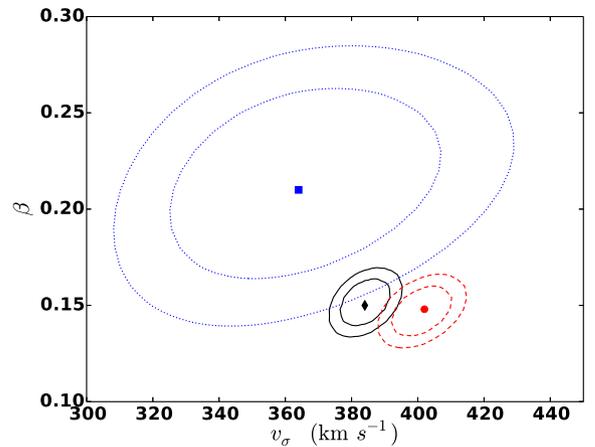


Figure 16. 68 per cent and 95 per cent confidence levels of the full (solid), red (dashed) and blue (dotted) $\Sigma(\pi)$ CMASS measurements shown in Figs 8 (left-hand panel) and 15 (top row). All the contours include covariances. Consistently with the size of the error bars in Fig. 15, the blue contours are much less tight than the red and full ones. The blue CMASS galaxies are less biased and have lower velocity dispersion than the red and full populations.

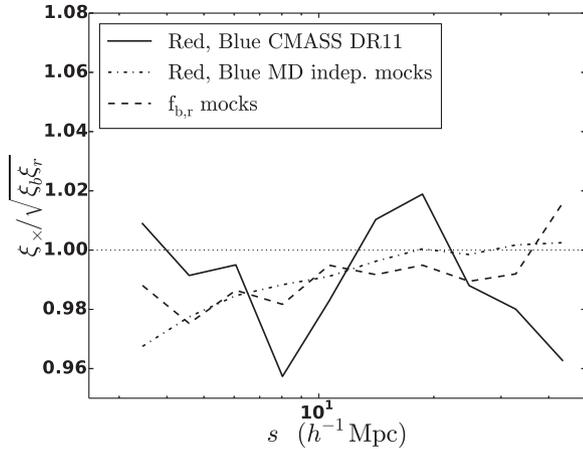


Figure 17. Ratio of the quantity $b_b b_r$ computed using the red-blue cross-correlation function, over the same quantity computed using the red and blue auto-correlation measurements. CMASS data (solid) versus independent (dot-dashed) and inverse tangent (dashed) mocks. Compatibly with expectations, the result is consistent with unity within 5 percent and the fluctuations are Poisson noise.

analysis produces a result that is consistent with expectations within 5 per cent.

7 DISCUSSION AND CONCLUSIONS

We have presented a qualitative analysis, as a function of colour, of the clustering signal in the high-redshift tail (i.e. $z > 0.55$) of the BOSS CMASS DR11 massive galaxy sample. Applying the colour cut defined in equation (30), we have divided the full CMASS sample into a redder and a bluer population of galaxies, and there we have computed the redshift-space and projected correlation functions at small and intermediate scales ($0.1 \leq r \leq 50 h^{-1}$ Mpc). Our measurements (see Section 4) are consistent with previous results by Wang et al. (2007), Zehavi et al. (2005b), Swanson et al. (2008) and show that blue star-forming galaxies preferentially populate less dense environments, compared to the red ones. Besides the 2PCF results, we have defined and measured a new quantity, $\Sigma(\pi)$ (equation 4), which is able to disentangle the non-linear small-scale redshift-space distortions (i.e. FoG) from the large-scale Kaiser flattening, respectively along and across the line of sight.

We have then mapped these results on to the MultiDark cosmological simulation (Section 2.4) using a five-parameter HOD model (Section 2.5), to generate reliable mock galaxy catalogues that reproduce the observed clustering signal in all the CMASS subsamples considered. First, using a traditional HOD approach, we have separately fit $\xi(s)$ of the full, red and blue CMASS populations building three independent mock catalogues (three different HOD models, with independent parameters). Instead of performing a formal fit, we have empirically tuned the HOD input parameters until we found suitable values that reproduce the observed clustering amplitude. To simplify the task, we chose to vary only three parameters, specifically those values related to physical quantities we want to measure: M_{\min} , the minimum host halo mass, which is connected to the galaxy number density, M'_1 , governing the satellite fraction, and α' , the slope of the satellite contribution. Our best empirical estimates for the independent HODs are reported in Table 1, and confirm that red galaxies preferentially populate more clustered environments, where the satellite fraction is higher than for blue-star forming galaxies. From this results we conclude that

we are able to individually match the clustering of the full, the red and the blue CMASS samples with small variations in the input parameters. Using these independent mocks, we have calculated the probability, $P(M_h|G)$, that a specific galaxy β is hosted by a halo with central mass M_h (left-hand panel of Fig. 14), and estimated the mean central halo masses of our red and blue model galaxies. We found $M_h \sim 10^{13.1}, 10^{12.5} M_\odot h^{-1}$, respectively for star-forming bluer and late-type redder galaxies, which again confirms that red galaxies live in more massive haloes.

The traditional HOD formulation reproduces both red and blue CMASS clustering; however, it is based on a non-physical assumption: being independent, the red and blue models share a certain number of mock galaxies. This means that the same galaxy can be either red or blue, whatever its mass is. To overcome this failure, we have modified the standard HOD assignment to infer both red and blue models from the full one, in such a way they are complementary and do not overlap. We have split (see Section 5.2) the full mock catalogue into a red and a blue submock by constraining it with an appropriate condition that mimics the observed CMASS red/blue galaxy fraction, $f_{b,r}$ (equation 36). We have tested four different functional forms of $f_{b,r}$ (see Appendix C for details), depending on a different number of parameters, and concluded that the best functional $f_{b,r}$ form is an inverse-tangent-like function (equation 37). It only depends on two free parameters, C and D , which respectively determine how fast the blue (red) fraction drops (grows) as the halo mass increases and the position of the half-width point of the curve. Our results, presented in Fig. 11, show good agreement between the MultiDark model galaxies and the CMASS observations.

We have then quantified the differences in the blue and red CMASS subpopulations from the point of view of the redshift-space distortions and large-scale bias (Section 6). Two regimes are interesting to this purpose: on large scales, the gravitational infall of galaxies to density inhomogeneities compresses the two-point correlation function along the line-of-sight direction; on small scales, the 2PCF experiences an elongation effect due to the non-linear peculiar velocities of galaxies, with respect to the Hubble flow (see Section 2.1). In order to separate the two contributions and isolate the small-scale elongation effect, we have built the new metric $\Sigma(\pi)$, defined in equation (4) as the ratio between a 2PCF – spherically averaged in the range $0.5 \leq r_p < 2 h^{-1}$ Mpc to maximize the FoG effect – and the best-fitting power law (spherically averaged) to the projected correlation function. This quantity is preferable than the quadrupole-to-monopole ratio in the attempt to maximize the FoG contribution on small scales, because it permits to separate the redshift-space features along and across the line of sight (see Section A). We have then modelled $\Sigma(\pi)$ by convolving the real-space best-fitting power law to $w_p(r_p)$, with a peculiar velocity term, assumed to be a normal function (equation 17) and the Kaiser factor (equation 15). The resulting model only depends on two parameters: β , that measures the Kaiser compression and is proportional throughout equation (14) to the inverse of the linear bias, b , and v_σ , that is the pairwise velocity dispersion, which quantifies the FoG elongation effect.

Fitting this v_σ, β parametrization to our full, red, blue $\Sigma(\pi)$ BOSS CMASS DR11 and MultiDark mock results, we found (see Table 2) that blue galaxies are less biased than red ones and have a lower peculiar velocity contribution, which leads to a smaller clustering amplitude.

In conclusion, we have performed a qualitative clustering analysis as a function of colour in a specific massive galaxy sample, the BOSS CMASS DR11, selecting only galaxies at $z > 0.55$. We have divided the sample in a redder and a bluer subpopulations, and here

we have measured the monopole, the projected 2PCF and a new quantity, $\Sigma(\pi)$, which is a compression of $\xi(r_p, \pi)$, specifically designed to study the FoG distortions and the linear bias. We have proposed and tested a straightforward model for $\Sigma(\pi)$, depending only on two parameters, that allows us to disentangle and constrain both the large-scale bias and the galaxy peculiar velocities, and provides a more exhaustive vision of the red/blue galaxy bimodality.

ACKNOWLEDGEMENTS

GF is supported by the Ministerio de Educación y Ciencia of the Spanish Government through FPI grant AYA2010-2131-C02-01 and wishes to thank the Smithsonian Center for Astrophysics and the Harvard University, Astronomy Department, for the hospitality during the creation of this work. GF acknowledges support from the Spanish Government through EEBB-I-13-07167 and EEBB-I-12-05220 grants. GF, FP, and CC acknowledge support from the Spanish MICINN's Consolider-Ingenio 2010 Programme under grant MultiDark CSD2009-00064, MINECO Centro de Excelencia Severo Ochoa Programme under grant SEV-2012-0249, and MINECO grant AYA2014-60641-C2-1-P.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation and the U.S. Department of Energy.

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, University of Cambridge, 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

Aihara H. et al., 2011, *ApJS*, 193, 29
 Alam S. et al., 2015, *ApJS*, 219, 12
 Anderson L. et al., 2012, *MNRAS*, 427, 3435
 Anderson L. et al., 2014, *MNRAS*, 441, 24
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
 Blanton M. R., Roweis S., 2007, *AJ*, 133, 734
 Bolton A. S. et al., 2012, *AJ*, 144, 144
 Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
 Croom S. M. et al., 2005, *MNRAS*, 356, 415
 Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
 Dawson K. S. et al., 2013, *AJ*, 145, 10
 di Porto C., Amendola L., Branchini E., 2012, *MNRAS*, 419, 985
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
 Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J. P., 1994, *MNRAS*, 267, 927
 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 H., Zehavi I., Zheng Z., 2012, *ApJ*, 756, 127

Guo H. et al., 2014, *MNRAS*, 441, 2398
 Guo H. et al., 2015, *MNRAS*, 449, L95
 Hamilton A. J. S., 1992, *ApJ*, 385, L5
 Hamilton A. J. S., 1998, in Hamilton D., ed., *Astrophysics and Space Science Library*, Vol. 231, The Evolving Universe. Kluwer, Dordrecht, p. 185
 Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
 Hudson M. J., Turnbull S. J., 2012, *ApJ*, 751, L30
 Jackson J. C., 1972, *MNRAS*, 156, 1P
 Kaiser N., 1987, *MNRAS*, 227, 1
 Klypin A., Holtzman J., 1997, preprint ([arXiv:astro-ph/9712217](https://arxiv.org/abs/astro-ph/9712217))
 Komatsu E. et al., 2009, *ApJS*, 180, 330
 Kravtsov A. V., Borgani S., 2012, *ARA&A*, 50, 353
 Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
 Linder E. V., 2005, *Phys. Rev. D*, 72, 043529
 Masters K. L. et al., 2011, *MNRAS*, 418, 1055
 Matsubara T., Suto Y., 1996, *ApJ*, 470, L1
 Miller R. G., 1974, *Biometrika*, 61, 1
 Montero-Dorta A. D. et al., 2014, preprint ([arXiv:1410.5854](https://arxiv.org/abs/1410.5854))
 Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
 Norberg P., Gaztañaga E., Baugh C. M., Croton D. J., 2011, *MNRAS*, 418, 2435
 Nuza S. E. et al., 2013, *MNRAS*, 432, 743
 Peacock J. A. et al., 2001, *Nature*, 410, 169
 Peebles P. J. E., ed., 1980, *The Large-scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ, p. 435
 Percival W. J. et al., 2014, *MNRAS*, 439, 2531
 Prada F., Klypin A., Yepes G., Nuza S. E., Gottloeber S., 2011, preprint ([arXiv:1111.2889](https://arxiv.org/abs/1111.2889))
 Primack J. R., 1997, preprint ([arXiv:astro-ph/9707285](https://arxiv.org/abs/astro-ph/9707285))
 Quenouille M. H., 1956, *Biometrika*, 43, 353
 Reid B. A., Spergel D. N., 2009, *ApJ*, 698, 143
 Riebe K. et al., 2011, preprint ([arXiv:1109.0003](https://arxiv.org/abs/1109.0003))
 Ross A. J. et al., 2012, *MNRAS*, 424, 564
 Ross A. J. et al., 2014, *MNRAS*, 437, 1109
 Sánchez A. G. et al., 2012, *MNRAS*, 425, 415
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
 Slosar A. et al., 2011, *J. Cosmol. Astropart. Phys.*, 9, 1
 Smee S. A. et al., 2013, *AJ*, 146, 32
 Swanson M. E. C., Tegmark M., Blanton M., Zehavi I., 2008, *MNRAS*, 385, 1635
 Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, *ApJ*, 742, 16
 Turkey J., 1958, *Ann. Math. Stat.*, 29, 1
 Wang Y., Yang X., Mo H. J., van den Bosch F. C., 2007, *ApJ*, 664, 608
 White M. et al., 2011, *ApJ*, 728, 126
 White M., Tinker J. L., McBride C. K., 2014, *MNRAS*, 437, 2594
 York D. G. et al., 2000, *AJ*, 120, 1579
 Zehavi I. et al., 2002, *ApJ*, 571, 172
 Zehavi I. et al., 2004, *ApJ*, 608, 16
 Zehavi I. et al., 2005a, *ApJ*, 621, 22
 Zehavi I. et al., 2005b, *ApJ*, 630, 1
 Zehavi I. et al., 2011, *ApJ*, 736, 59
 Zheng Z. et al., 2005, *ApJ*, 633, 791
 Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760

APPENDIX A: QUADRUPOLE-TO-MONOPOLE RATIO VERSUS $\Sigma(\pi)$ STATISTICS

The novelty of our $\Sigma(\pi)$ statistics is that it allows one to extract the maximum contribution of small-scale redshift-space distortions separating the effects along the line of sight (i.e. FoG) from the effects (i.e. Kaiser squashing) across it. In fact, $\Sigma(\pi)$ is defined (see equation 4) by normalizing out the real-space contribution from the redshift-space 2PCF, spherically averaged in the range

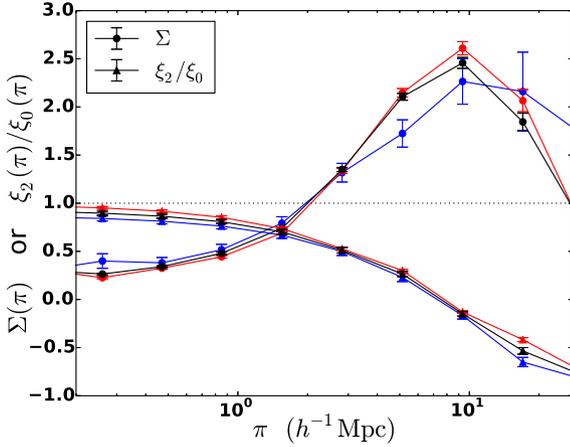


Figure A1. $\Sigma(\pi)$ (points) versus $\xi_2(s=\pi)/\xi_0(s=\pi)$ (triangles) measurements for the full (black), red and blue CMASS samples. The advantage of measuring $\Sigma(\pi)$ is that, by construction, it allows us to disentangle the small-scale non-linear redshift-space distortion effects from the Kaiser squashing on larger scales.

$0.5 \leq r_p \leq 2 h^{-1}$ Mpc to maximize the FoG effect. In alternative to $\Sigma(\pi)$, one could measure the quadrupole-to-monopole ratio, $\xi_2(s)/\xi_0(s)$, to extract information about the redshift-space clustering features. However, this ratio is computed as a function of the redshift-space distance $s = \sqrt{r_p^2 + \pi^2}$, and does not permit to disentangle the FoG elongation from the Kaiser flattening. By modelling $\Sigma(\pi)$ in a straightforward way only as a function of two parameters v_σ , β (see Section 4.2), we are able to separate the small-scale non-linear FoG regime, where the peculiar velocities (quantified by v_σ) dominate, from the large-scale linear regime, where the Kaiser compression becomes important, and the linear bias (quantified by β) can be estimated. In Fig. A1 we show a comparison of the full, red and blue CMASS DR11 $\Sigma(\pi)$ measurements and quadrupole-to-monopole ratios, these latter evaluated at $s = \pi$. The feature of the two metrics is not comparable, nor the information they carry. The advantage of using $\xi_2(s)/\xi_0(s)$ is the smaller size of the error bars, which would lead to tighter constraints in the analysis. On the other hand, the advantage of using $\Sigma(\pi)$ is, as explained above, that it permits to quantify both linear galaxy bias and FoG contribution in a straightforward way, disentangling the effects as a function of the physical scale. In Fig. A2 we display

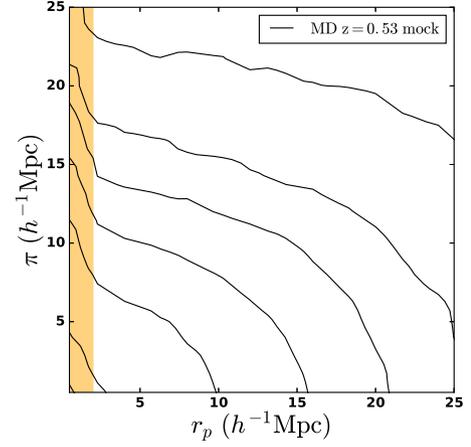


Figure A2. Two-point correlation function of the full CMASS MultiDark mock galaxy catalogue. The orange shaded area represents the $\Sigma(\pi)$ domain where the FoG effect is maximized.

the 2PCF of the MultiDark model galaxies for the full CMASS sample, given as a function of the parallel (i.e. π) and perpendicular (i.e. r_p) components to the line of sight. The orange shaded region highlights the $\Sigma(\pi)$ domain, where the FoG effect is maximized.

APPENDIX B: CLUSTERING SENSITIVITY ON HOD PARAMETERS

The left column in Fig. B1 presents our HOD model (see Section 2.5) as a function of three parameters: M_{\min} (top row), M'_1 (middle), and α' (bottom); the remaining two parameters are fixed to the default values given by White et al. (2011): $\log M_0 = 12.8633$, $\sigma_{\log M} = 0.5528$. The projected correlation functions based on these mocks are shown in the right column. Increasing the value of M_{\min} (top row, from lighter to darker solid lines) globally enhances the clustering amplitude, with a strong contribution from substructures belonging to different hosts (two-halo term). On the other side, the interaction between satellites belonging to the same central halo (one-halo term) weakens as M'_1 increases (bottom row, from lighter to darker solid lines), resulting in a smoother slope at scales $r_p \leq 1 h^{-1}$ Mpc. The extreme case is achieved when $\log M'_1 = 16.00$, where the satellite contribution becomes almost negligible, and $f_{\text{sat}} = 5.45 \times 10^{-4} \simeq 0$.

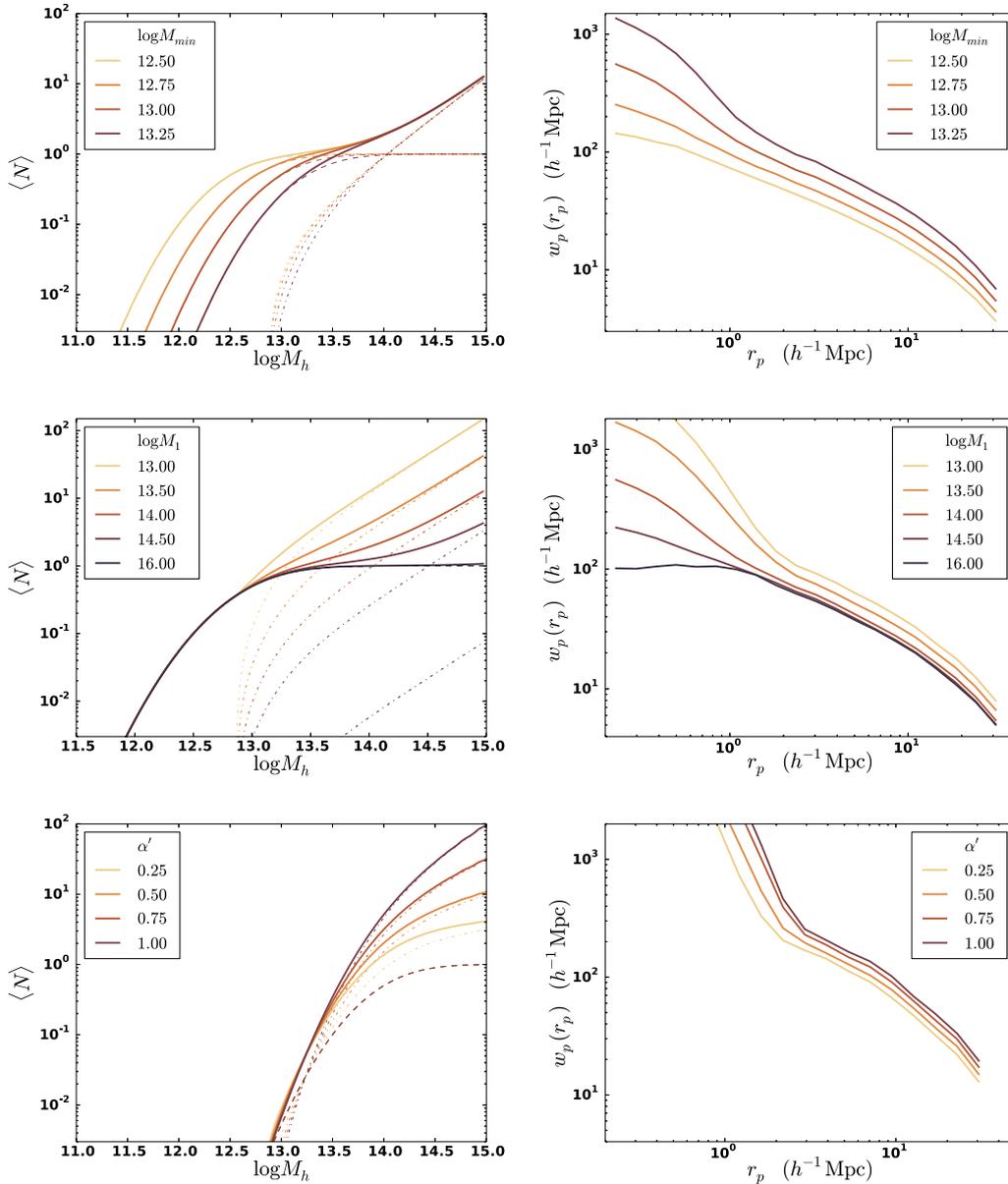


Figure B1. Implication of a change in the HOD input parameters (left column) on the projected correlation function (right column). We allow only one parameter to vary at a time: M_{\min} , in the top row, especially affects the two-halo term; M'_1 ($\log M_{\min} = 13.00$) and α' , respectively in the middle and bottom row, have a strong effect on the one-halo term. The resulting correlation functions are degenerate with respect to the variation of these three parameters. The remaining two parameters are fixed at the default values given by White et al. (2011): $\log M_0 = 12.8633$, $\sigma_{\log M} = 0.5528$.

APPENDIX C: RED AND BLUE GALAXY FRACTION MODELS

In addition to the inverse tangent fraction model defined in equation (37), to mimic the red and blue galaxy fractions as a function of the central halo mass, we test also a linear model

$$f_b(\log M_h) = -M \log M_h + N, \quad (\text{C1})$$

and two log-normal like functions, with three degrees of freedom each. The first one (Logn I) is given by

$$f_b(\log M_h) = \frac{P_b}{P_b + P_r}, \quad (\text{C2})$$

where

$$P_{b,r} = \exp\left(-\frac{(\log M_h - \mu_{b,r})^2}{2\sigma^2}\right) \quad (\text{C3})$$

is a density function. The parameters $\mu_{b,r}$ are the blue and red, mean galaxy masses, respectively, and σ is the log-normal width. The second version (Logn II) has fixed amplitude σ , and a new parameter k , that controls the mutual heights of the red and blue peaks. We have

$$f_b(\log M_h) = \frac{P_b}{P_b + kP_r}, \quad (\text{C4})$$

where $P_{b,r}$ is given by equation (C3). After applying these constraints to the full MultiDark mock catalogue, we split it into its red

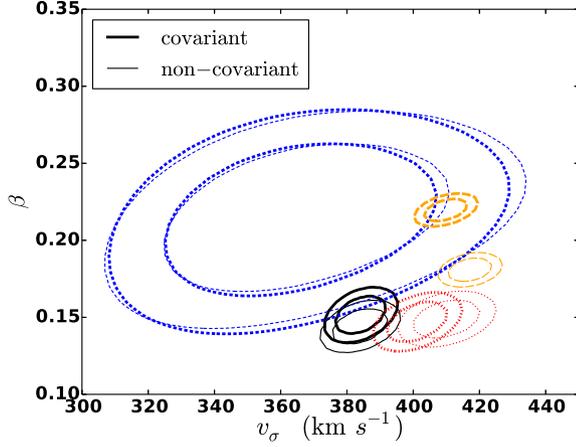


Figure D1. Covariant (thick contours) versus non-covariant (weak lines) 68 per cent and 95 per cent confidence levels of the v_σ, β models for the $\Sigma(\pi)$ full (black solid), red (red dotted) and blue (blue dashed) CMASS measurements versus QPM mocks (orange long-dashed). QPMs have slightly different cosmology: $\Omega_m = 0.29$. The inclusion of covariances is almost negligible for the blue population, and weakly appreciable in the full case. Inversely, in the red population, covariances slightly move the fit towards higher velocity values; for QPMs, this shift is significant and drives the contours towards smaller bias values and slightly higher velocities.

and blue components. We then fit the clustering amplitudes of our model galaxies to the CMASS red and blue samples.

APPENDIX D: TESTING THE ERRORS – JACKKNIFE VERSUS QPM MOCKS

We test our full CMASS jackknife error estimates by computing the $\xi(s)$, $w_p(r_p)$, and $\Sigma(\pi)$ covariance matrices from a set of 100 Quick Particle Mesh (QPM; White et al. 2014) mock catalogues, with slightly different cosmology: $\Omega_m = 0.29$. Since these mocks

are all independent of each other, we can compute their covariance as

$$C_{kl}^{\text{QPM}} = \frac{1}{N_{\text{QPM}} - 1} \sum_{b=1}^{N_{\text{QPM}}} (\xi_k^b - \bar{\xi}_k)(\xi_l^b - \bar{\xi}_l), \quad (\text{D1})$$

where $N_{\text{QPM}} = 100$, and $\bar{\xi}_k$ is the mean QPM correlation function in the k th bin,

$$\bar{\xi}_k = \sum_{b=1}^{N_{\text{QPM}}} \xi_k^b / N_{\text{QPM}}. \quad (\text{D2})$$

Fig. D1 shows the covariant (thick lines) and the non-covariant (weak) v_σ, β contours of the full, red and blue CMASS $\Sigma(\pi)$ models versus QPM mocks (orange long-dashed). The inclusion of covariances is almost negligible for the blue CMASS model, while it moves the full and red models towards smaller bias values and higher velocity dispersion values, respectively. QPM contours are narrow, analogously to the full CMASS sample, and the inclusion of covariances in this case significantly moves the fit towards lower bias values and slightly higher velocities. Fig. D2 compares the normalized $\xi(s)$, $w_p(r_p)$, and $\Sigma(\pi)$ (respectively from left to right column) covariance matrices estimated using the QPM mocks (top row) and the jackknife re-samplings of the full, red and CMASS galaxy samples, to test the correlation between our observations at different scales. Overall, the QPM mocks show stronger covariances than jackknife in all three metrics. $\Sigma(\pi)$ is less correlated than the redshift-space and projected correlation functions; this is due to its definition, see equation (4). Since $\Sigma(\pi)$ is the ratio of two clustering measurements, both errors propagate in it, resulting in a smoother correlation at all scales. The red CMASS sample includes the majority of the CMASS galaxies, thus it is reasonable that its covariance matrices behave similarly to the ones of the full sample. The blue case is slightly different: errors are larger and covariances are almost negligible in all the three measurements, especially in $\Sigma(\pi)$.

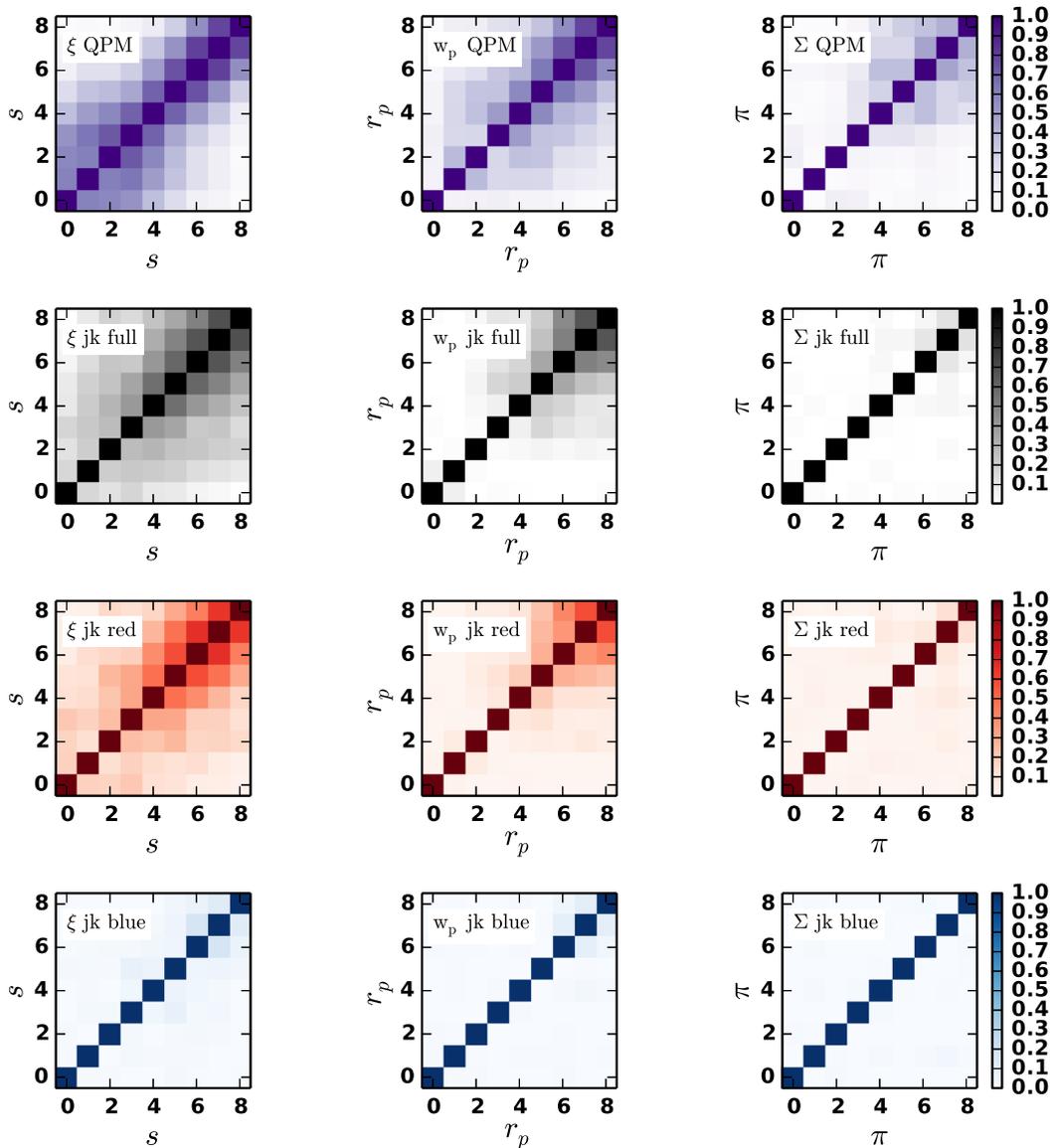


Figure D2. Normalized QPM (first row from the top) versus full (second row), red (third row) and blue (bottom row) CMASS jackknife covariance matrices for $\xi(s)$ (left column), $w_p(r_p)$ (central), and $\Sigma(\pi)$ (right), as a function of the s , r_p and π bins, respectively. We adopted a 10-step logarithmic binning scheme in the range $3\text{--}50 h^{-1}$ Mpc for s , $0.1\text{--}35 h^{-1}$ Mpc for r_p , and $0.1\text{--}40 h^{-1}$ Mpc for π . Overall, QPM mocks show higher covariances compared to the full, red, and blue CMASS samples, confirming the result shown in Fig. D1. The left column reveals that covariances become appreciable in the red and full redshift-space 2PCFs at intermediate scales (i.e. $s \geq 8 h^{-1}$ Mpc), while they are almost negligible in the blue population. The red and full CMASS projected 2PCF (central column) are covariant at $r_p \geq 2 h^{-1}$ Mpc, while the blue case is almost covariance-free at all scales. The $\Sigma(\pi)$ measurements (right column) are significantly less covariant than the other two clustering statistics: QPM mocks show appreciable covariances only above $\pi \sim 3 h^{-1}$ Mpc, while the three CMASS samples are substantially covariance-free.

This paper has been typeset from a \LaTeX file prepared by the author.