Halo mass distribution reconstruction across the cosmic web

Cheng Zhao[1][*], Francisco-Shu Kitaura[2][†], Chia-Hsun Chuang[3][‡], Francisco Prada[3,4,5], Gustavo Yepes[6], Charling Tao[1,7]

[1] *Tsinghua Center of Astrophysics & Department of Physics, Tsinghua University, Beijing 100084, China.*
[2] *Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[3] *Instituto de Física Teórica, (UAM/CSIC), Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[4] *Campus of International Excellence UAM+CSIC, Cantoblanco, E-28049 Madrid, Spain*
[5] *Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, E-18080 Granada, Spain*
[6] *Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, 28049, Madrid, Spain*
[7] *CPPM, Université Aix-Marseille, CNRS/IN2P3, Case 907, 13288 Marseille Cedex 9, France*

23 January 2015

ABSTRACT

We study the relation between halo mass and its environment from a probabilistic perspective. We find that halo mass depends not only on local dark matter density, but also on non-local quantities such as the cosmic web environment and the halo-exclusion effect. Given these accurate relations, we have developed the HADRON-code (Halo mAss Distribution ReconstructiON), a technique which permits us to assign halo masses to a distribution of haloes in three-dimensional space. This can be applied to the fast production of mock galaxy catalogues, by assigning halo masses, and reproducing accurately the bias for different mass cuts. The resulting clustering of the halo populations agree well with that drawn from the BigMultiDark N-body simulation: the power spectra are within 1-σ up to scales of $k = 0.2\,h\,\mathrm{Mpc}^{-1}$, when using augmented Lagrangian perturbation theory based mock catalogues. Only the most massive haloes show a larger deviation. For these, we find evidence of the halo-exclusion effect. A clear improvement is achieved when assigning the highest masses to haloes with a minimum distance separation. We also compute the 2- and 3-point correlation functions, and find an excellent agreement with N-body results. Our work represents a quantitative application of the cosmic web classification. It can have further interesting applications in the multi-tracer analysis of the large-scale structure for future galaxy surveys.

Key words: (cosmology:) large-scale structure of Universe – galaxies: clusters: general – catalogues – galaxies: statistics

1 INTRODUCTION

Hierarchical structure formation in the Cold Dark Matter theory predicts the production of gravitationally bound compact objects called haloes (White & Rees 1978; Fry & Peebles 1978). They host the galaxies we observe in our Universe according to the standard cosmological paradigm. Nevertheless, their biased relationship with respect to the underlying dark matter distribution remains still a matter of study. In spite of the great progress made during the last decades some questions have not been fully answered yet

(for a review, cf. Cooray & Sheth 2002). We certainly know now that bias is nonlinear and scale-dependent (e.g. Nuza et al. 2013). Recent studies have shown that non-local (e.g. Saito et al. 2014) and stochastic contributions (e.g. Kitaura et al. 2014a) are also relevant in the three-point clustering statistics. In fact, parametrized bias expressions are degenerate in the two-point clustering statistics (Kitaura et al. 2014b).

A proper bias weighting or mass weighting can reduce the variance of the clustering measurements (Percival, Verde, & Peacock 2004; Seljak et al. 2009). These methods can be applied to the galaxy surveys (e.g., SDSS-III/BOSS[1]

[*] E-mail: zhao-c13@mails.tsinghua.edu.cn
[†] E-mail: kitaura@aip.de, Karl-Schwarzschild fellow
[‡] E-mail: chia-hsun.chuang@uam.es, MultiDark fellow

[1] `http://www.sdss.org`

(Eisenstein et al. 2011; Dawson et al. 2013)), of which one can estimate the biases or masses of the galaxy sample. The next generation of galaxy surveys (e.g. SDSS-IV/eBOSS[2], DESI[3] (Schlegel et al. 2011), DES[4] (Frieman et al. 2013), LSST[5] (Abate et al. 2012), J-PAS[6] (Benitez et al. 2014), 4MOST[7] (de Jong et al. 2012) or Euclid[8] (Laureijs et al. 2009)) will exploit the multi-tracer analysis to constrain dark energy, the growth rate of the Universe, and hence gravity models (e.g. McDonald & Seljak 2009; Blake et al. 2013). In such a multi-tracer approach, different population of tracers of the cosmic density field are treated as independent measurements, which weighted by their distinct bias (cf. McDonald & Seljak 2009; Seljak et al. 2009; Hamaus et al. 2010), will yield much tighter cosmological constraints (cf. also Abramo & Leonard 2013). In this context, it is fundamental to have a deep understanding of the bias for different population of tracers.

We aim at answering several questions in this study, such as: how are haloes distributed in the cosmic web, and which properties determine the bias of different halo populations? In particular, we investigate in this work the relation between halo mass and environment. As a practical application, we want to understand how to statistically assign halo mass to a mixed population of haloes. We will present in a subsequent publication how to extend this work to galaxy stellar masses (Kitaura et al., in prep.; Rodríguez-Torres et al., in prep.).

The technique presented in this work has a direct application to the fast generation of mock galaxy or halo catalogues, and could be applied for the reconstruction of halo masses and density field (cf. e.g. Wang et al. 2009; Muñoz-Cuartas, Müller, & Forero-Romero 2011; Wang et al. 2012; Muñoz-Cuartas & Müller 2012; Kitaura 2013), and to the multi-tracer analysis from galaxy redshift surveys.

This paper is structured as follows: In section 2, we first present the theoretical approach. Then, we show in section 4 our numerical experiments based on the BigMulti-Dark (BigMD) *N*-body simulations[9] (Klypin et al. 2014), which we described in section 3, followed by the application to the generation of mock galaxy/halo catalogues based on perturbation theory in section 5. Finally, we present our conclusions in section 6.

2   THEORETICAL APPROACH

The aim of this study is to examine the properties of the large-scale structures which statistically determine the mass of haloes. The starting point is given by the mass function, which predicts the number of compact objects (haloes) of a certain mass (cf. pioneering works of Press & Schechter (1974); Bardeen et al. (1986), and the later seminal works by

Mo & White (1996); Sheth & Tormen (2004a)). The question is which additional quantities $\{q\}$ have a significant impact on the mass of haloes from a statistical perspective. In particular, we want to answer: what determines the conditional probability distribution function of the halo mass M_{h}^i of an object at position $\boldsymbol{r}_{\mathrm{h}}^i$, given a distribution of haloes in three-dimensional space $\{\boldsymbol{r}_{\mathrm{h}}\}$, and at redshift z with cosmological parameters $\{p_{\mathrm{c}}\}$, i.e.

$$M_{\mathrm{h}}^i \curvearrowleft P(M_{\mathrm{h}}(\boldsymbol{r}_{\mathrm{h}}^i)|\{\boldsymbol{r}_{\mathrm{h}}\}, \{q\}, \{p_{\mathrm{c}}\}, z)\,. \qquad (1)$$

The accuracy of this mass assignment will have an impact on various statistical measures, such as the two- and the three-point clustering statistics. Hence, it controls the bias, which is the ultimate goal of this work. Here, we follow a hierarchical approach in which we include increasingly more information in the conditional probability distribution function, evaluating at each stage the precision of the resulting bias. From theoretical considerations based on the literature, we need to examine the impact of nonlinear local, non-local, and stochastic components of the bias (e.g. Press & Schechter 1974; Peacock & Heavens 1985; Bardeen et al. 1986; Fry & Gaztanaga 1993; Mo & White 1996; Pen 1998; Dekel & Lahav 1999; Sheth & Lemson 1999; Seljak 2000; Berlind & Weinberg 2002; Smith et al. 2007; McDonald & Roy 2009; Desjacques et al. 2010; Beltrán Jiménez & Durrer 2011; Valageas & Nishimichi 2011; Elia et al. 2012; Chan et al. 2012; Baldauf et al. 2012, 2013; Angulo et al. 2014; Baldauf et al. 2014):

(i) We start with the simplest assumption based solely on the mass function neglecting any other information: $P(M_{\mathrm{h}}^i|\{\boldsymbol{r}_{\mathrm{h}}\}, \{p_{\mathrm{c}}\}, z)$.

(ii) The peak background split picture models the formation of haloes of different masses in density peaks above corresponding density thresholds (Kaiser 1984; Bardeen et al. 1986; Sheth & Tormen 2004b). This theory indicates that we should consider as a next order approximation the dependence on the local density field ρ_{M}: $P(M_{\mathrm{h}}^i|\{\boldsymbol{r}_{\mathrm{h}}\}, \rho_{\mathrm{M}}, \{p_{\mathrm{c}}\}, z)$.

(iii) Recent studies have shown that non-local effects are relevant in the three-point clustering statistics (e.g. Saito et al. 2014). In particular, we are interested in investigating the importance of the cosmic environment in which different haloes reside. We will study this through the eigenvalues of the tidal field tensor T, which is a non-local measure: $P(M_{\mathrm{h}}^i|\{\boldsymbol{r}_{\mathrm{h}}\}, \rho_{\mathrm{M}}, T, \{p_{\mathrm{c}}\}, z)$.

(iv) Finally, we aim at investigating stochastic biasing (e.g. Kitaura et al. 2014a,b). This component encodes in an effective way the non-local bias contributions (in this case beyond the tidal field tensor). We will consider in particular the deviation from Poissonity. A larger dispersion than Poisson corresponds to over-dispersion, a smaller one to under-dispersion, which are due to the positive or negative correlation on sub-grid scales, respectively (e.g. Peebles 1980; Somerville et al. 2001; Casas-Miranda et al. 2002; Baldauf et al. 2013; Kitaura et al. 2014a). We will focus in this study on the minimum separation between haloes $\Delta r_{\mathrm{min}}^M$: $P(M_{\mathrm{h}}^i|\{\boldsymbol{r}_{\mathrm{h}}\}, \rho_{\mathrm{M}}, T, \Delta r_{\mathrm{min}}^M, \{p_{\mathrm{c}}\}, z)$.

In the next section we will investigate the relevance of the different bias components $\{q\} = \{\rho_{\mathrm{M}}, T, \Delta r_{\mathrm{min}}^M\}$, based on the analysis of a large *N*-body cosmological simulation.

3 REFERENCE N-BODY SIMULATION AND HALO CATALOGUES

In particular, we employ the dark matter particle and halo catalogues at redshift $z = 0.5618$, extracted from one of the BigMD simulations, which was performed using the TreePM N-body code GADGET-2 (Springel 2005) with 3840^3 particles in a volume of $(2.5\,h^{-1}\text{Gpc})^3$, within the framework of Planck ΛCDM cosmology with $\{\Omega_m = 0.307115, \Omega_b = 0.048206, \sigma_8 = 0.8288, n_s = 0.96\}$, and the Hubble parameter $(H_0 \equiv 100\,h\,\text{km s}^{-1}\text{Mpc}^{-1})$ given by $h = 0.6777$.

We have two sets of halo catalogues constructed by using two different halo finders, the spherical overdensity Bound Density Maxima (BDM) (Klypin & Holtzman 1997; Gottloeber 1998; Riebe et al. 2011) and the Friends-of-Friends (FoF) (Gottloeber 1998; Riebe et al. 2011) halo finders with linking length $l = 0.17$ times the mean interparticle distance ($0.11\ h^{-1}$Mpc). We select the BDM haloes and subhaloes by V_{max} (i.e., maximum circular velocity), and use mass for FoF haloes, to construct complete samples from both halo catalogues with number density $3.5 \times 10^{-4}\,h^3\,\text{Mpc}^{-3}$, as that for typical Luminous Red Galaxies in large-scale surveys.

We follow Klypin, Trujillo-Gomez, & Primack (2011) for the definition of V_{max}:

$$V_{\text{max}} = \sqrt{\left.\frac{GM(<r)}{r}\right|_{\text{max}}}. \qquad (2)$$

We choose this quantity as a proxy for halo mass, since (cf. Prada et al. 2012), 1) it is a more reliable quantity than the mass defined at a given over-density, and 2) it is better for the characterization of haloes when relating them to the galaxies inside. It has a more direct relation with observational quantities, such as luminosity or stellar mass, that are used for defining galaxy catalogues with Halo Abundance Matching (HAM) modelling (Trujillo-Gomez et al. 2011). We will focus in this work on BDM (sub)haloes and show some results using FoF haloes in Appendix A.

4 THE HALO MASS ENVIRONMENTAL DEPENDENCE

Haloes are generally identified in an N-body dark matter density field using the so-called halo-finder algorithms. This is essentially an estimate of the halo bias, which encodes a certain relation between halo masses and the dark matter density field.

Therefore, let us start by studying the bias from N-body cosmological simulations that provide both dark matter particles and the corresponding halo catalogues. To follow the analysis proposed in the previous section we will investigate the clustering statistics of different populations of haloes conditioned on different degrees of information.

To evaluate the accuracy of the bias, we will start with the two-point statistics in Fourier space. In particular, we compute power spectra of different populations of haloes, whose amplitudes are essentially a direct estimate of the bias factor (e.g. Cen & Ostriker 1992), and their shapes show the scale-dependency. In this work, we adopt the cloud-in-cells particle assignment scheme (CIC) for haloes with grid size of 512^3 to perform the fast Fourier transform, and then
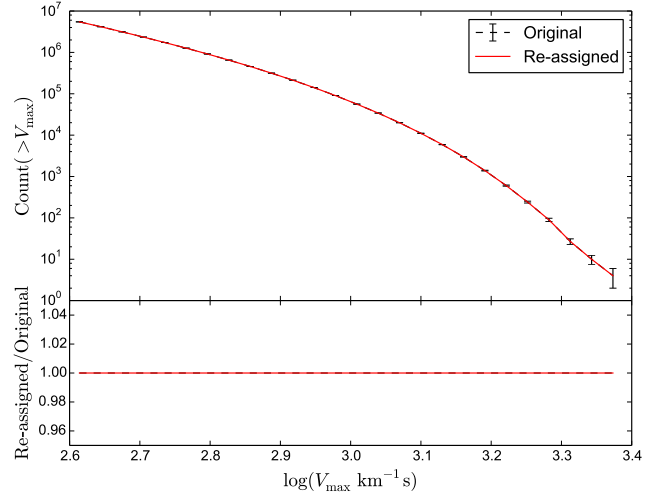
Figure 1. Halo mass (V_{max}) functions of the original data drawn from the BigMD simulation and re-assigned halo catalogue. They agree with each other by construction. The error bars show Poisson errors.

compute the power spectra with aliasing and shot noise corrections taken into account (cf. Jing 2005).

Our aim is to reduce systematic deviations on the power spectra to a few percent on scales relevant to baryon acoustic oscillations, i.e. $k \lesssim 0.2\,h\,\text{Mpc}^{-1}$. We study in this section different mass assignment procedures, and show how this goal can be achieved, provided we take into account density–mass relation, cosmic web environment, and halo-exclusion. Let us now perform our numerical analysis trying to recover the masses of a given three-dimensional distribution of haloes going through the steps outlined in §2.

4.1 Mass function

We start by studying the halo V_{max} reconstruction considering only the mass (V_{max}) function, i.e. the conditional probability function $P(M_{\text{h}}^i | \{\boldsymbol{r}_{\text{h}}\}, \{p_{\text{c}}\}, z)$. In particular, each halo gets a V_{max} assigned following the function shown in Fig. 1 regardless of its location. In this case, the V_{max} cumulative function is reproduced by construction.

To verify the performance of the halo V_{max} reconstruction, we divide both the original BDM halo catalogue and the re-assigned one, into 8 sub-samples of halo V_{max} and compare the clustering for each sub-sample. In particular, we cut the catalogues in V_{max} bins of $\{[<410), [410\text{-}427), [427\text{-}447), [447\text{-}472), [472\text{-}501), [501\text{-}550), [550\text{-}640), [\geqslant 640]\}$ km s^{-1} to have similar number of haloes in each bin. We can see in Fig. 2 how the power spectra of the different sub-samples have amplitudes which disagree with the true ones. The degree of the systematic deviation depends on the mass of the halo population. Low-mass haloes ($V_{\text{max}} \sim 500$ km s^{-1}) yield an overestimation of the bias (the ratio of the reconstructed and the true power spectrum is greater than one). Only haloes in the range around $V_{\text{max}} \sim 500 - 550$ km s^{-1} are closely unbiased, and more massive haloes lead to an underestimation of the bias (the ratio of the reconstructed and the true power spectrum is lower than one). Nevertheless, the deviations of up to $\sim 40\%$ throughout the full k-range in the power spectra hint that we need to con-
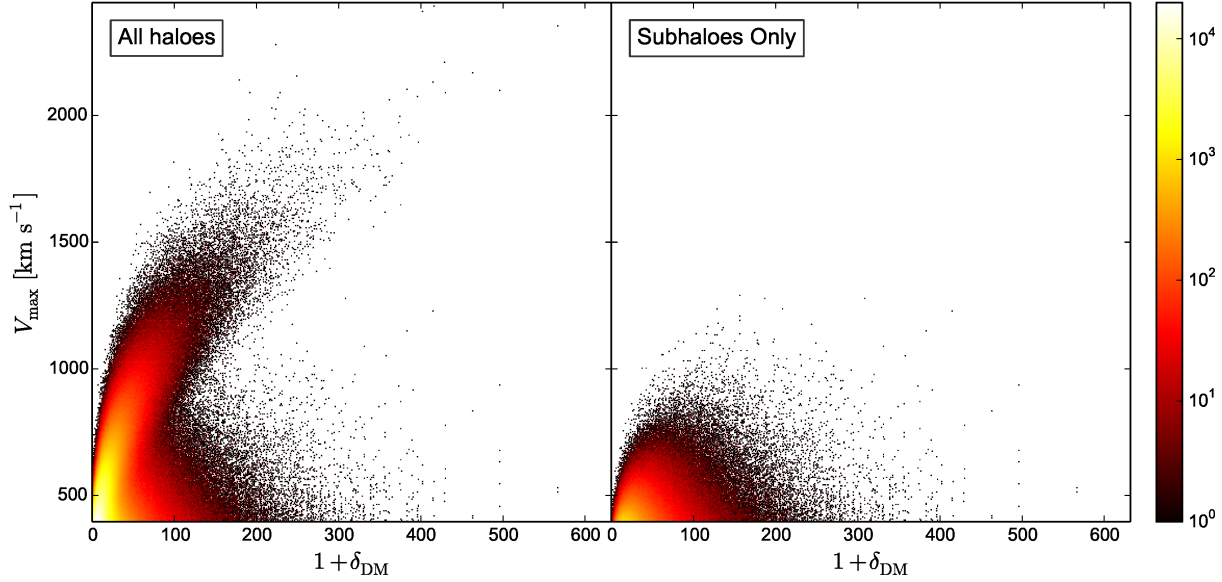
Figure 3. Number of (sub)haloes in the BigMD simulation with certain V_{max} and local DM density.
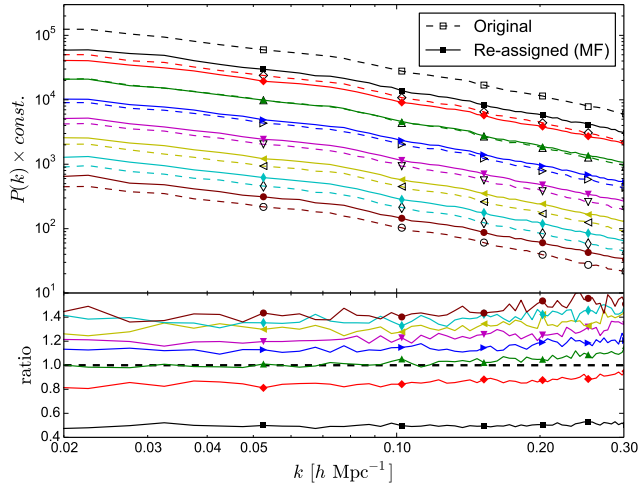

Figure 2. Power spectra of the halo sub-samples discussed in the text, with different halo V_{max}. Dashed lines indicate the power spectra for the sub-samples drawn from the BigMD BDM catalogue, while solid lines correspond to that after applying the mass assignment procedure. MF stands for the mass (V_{max}) function used in this case, see Fig. 1. The different colour codes correspond to the different V_{max} bins of the sub-samples increasing from the bottom to the top lines ($\{[<410), [410\text{-}427), [427\text{-}447), [447\text{-}472), [472\text{-}501), [501\text{-}550), [550\text{-}640), [\geqslant 640]\}$ km s^{-1}). For visualisation purposes the power spectra corresponding to different V_{max} bins have been multiplied with different constants in the upper panel to enlarge the differences

.

sider additional environmental indicators to make a precise mass assignment.

4.2 Density–halo mass relation

Let us now investigate the relation between the halo mass (or equivalently V_{max}) and the local dark matter density. We employ the CIC scheme for dark matter particles in the
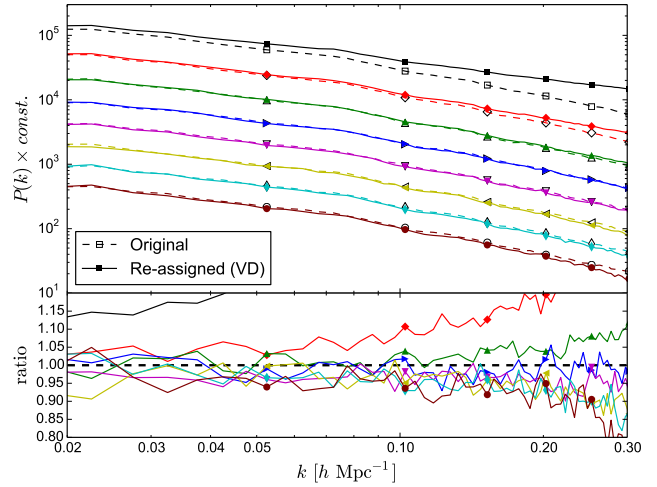
Figure 4. Power spectra of the sub-samples with different halo V_{max}. Same convention as in Fig. 2. VD stands for the V_{max}–density relation used in this case.

BigMD simulation with a grid size of 960^3 to obtain the dark matter density field, and then distribute haloes to the same mesh using the nearest-grid-point scheme (NGP), since as our application, the mocks employ CIC while computing the dark matter density field, and we need to use integers for the mass assignment procedure. For each halo, the local dark matter density is defined by the number density contrast of dark matter particles in the corresponding cell, i.e., $1 + \delta_{DM} = \rho/\bar{\rho}$, with δ_{DM} being the density fluctuations, ρ the density, and $\bar{\rho}$ the mean density.

The left panel in Fig. 3 shows the distribution of all haloes, and subhaloes only in the V_{max} vs. local dark matter density plane, where we adopt 500 bins for both V_{max} and dark matter density. Massive haloes tend to be located in dense environments, obeying a power law. Interestingly, there is a halo mass range suppressed at high densities, indicating that these moderately massive haloes are merged to
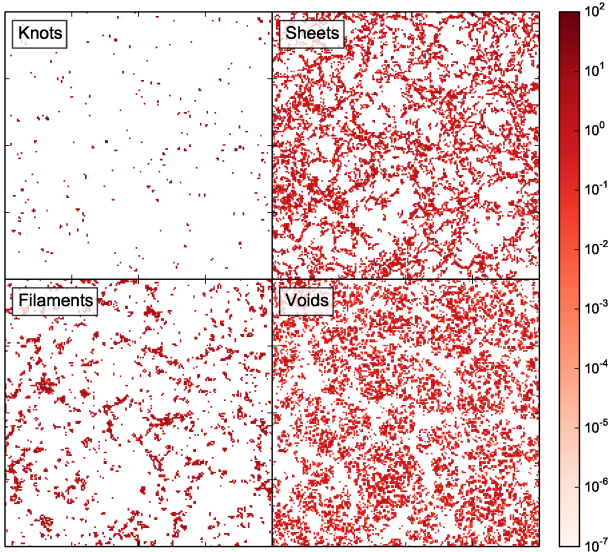
Figure 5. Dark matter density field in the BigMD simulation corresponding to different cosmic web structures, as indicated in the legend of the panels.

Figure 6. V_{max}–density relation of different cosmic web structure classes. The mass of knots increase from class 1 to class 3.

more massive ones in such environments. Another remarkable feature is that there are two branches in the high density regions, indicating two different groups of haloes residing in the same dark matter environment. The haloes in the BDM catalogue can be divided into distinct and subhaloes. Thus, we find that haloes in the low-V_{max} region are predominantly sub-structures of those with higher V_{max} (see right panel of Fig. 3). This is why the haloes in very dense regions (e.g. $1 + \delta_{DM} > 400$) are located in discrete bins of constant $1 + \delta_{DM}$ in Fig. 3, as can be more clearly seen in the outliers at high $1 + \delta_{DM}$ values. This relation shows the probability of finding a halo with a given V_{max} in a certain dark matter density environment.

We then keep the local dark matter density of each halo, and assign V_{max} to the haloes according to the extracted probability, ignoring their original V_{max}. The new catalogue has the same V_{max}–density relation as the original one, given the same V_{max} and density bins. The re-assignment procedure is equivalent to a shuffling of the halo V_{max} in each dark matter density bin.

A clear improvement is found with respect to the previous results given in §4.1, as can be seen in the power spectra shown in Fig. 4. Nevertheless, the lowest and the next to largest V_{max} bins still show systematic deviations of about 10 and 15% up to $k \sim 0.15\,h\,\mathrm{Mpc}^{-1}$, respectively (and increasingly larger towards larger k). The largest V_{max} bin shows even a deviation exceeding 30% at $k \sim 0.15\,h\,\mathrm{Mpc}^{-1}$. This indicates that we still need to include more information to reach the desired accuracy of $\sim 10\%$.

4.3 Cosmic web environment

As the next step, we now include non-local indicators. The tidal field tensor includes second order non-local information, and its eigenvalues have been used to make a cosmic web classification (e.g. Hahn et al. (2007); Forero-Romero et al. (2009); Aragón-Calvo et al. (2010); Hoffman et al. (2012)). Since different types of cosmic web structures can
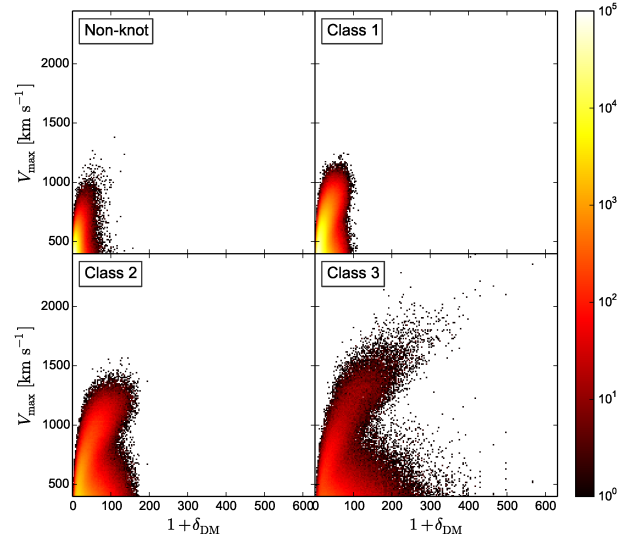
have the same local matter density, our analysis in the previous section mixed haloes residing in different structures and thus with different biases. Therefore, we split the haloes according, not only to their local dark matter density, but also to the type of cosmic web structures they live in.

In particular, we classify cosmic web structures following Hahn et al. (2007) and Forero-Romero et al. (2009). From the dark matter density field, we obtain the gravitational potential ϕ from the Poisson equation, and construct the tidal field tensor

$$T_{ij} = \frac{\partial^2 \phi}{\partial x_i \partial x_j}. \tag{3}$$

If all the three eigenvalues of T_{ij} are above (below) a certain threshold (λ_{th}), then the local structure is collapsing (expanding) in all directions, forming a knot (void). When one (two) eigenvalue is above the threshold, we have filament-like (sheet-like) structures.

A slice of the dark matter density field for $\lambda_{th} = 0$ is shown in Fig. 5, for different types of cosmic web structures. Although voids, sheets, and filaments occupy most of the volume, the majority of the haloes are located in knots. To be more precise, we further classify knots by their total enclosed mass into several classes. To compute the mass of each knot, we adopt a simple FoF algorithm resolving single knots from the dark matter density field. The mesh cells of knots are marked during the cosmic web classification, and subsequently all marked cells that are next to each other are merged to construct a single knot. The mass of the knot is then proportional to the sum of dark matter mass of all the connected cells. Our analysis shows that, for the halo masses considered in this study, the distinction between non-knots structures does not add any information. Therefore, we combine the rest of structures into a single specie. This situation may however be different when considering lower mass haloes.

With the different types of cosmic web structures resolved from the dark matter density field, we can then extract the relation between the halo V_{max} and the dark matter density for each class, as shown in Fig. 6. In this plot,
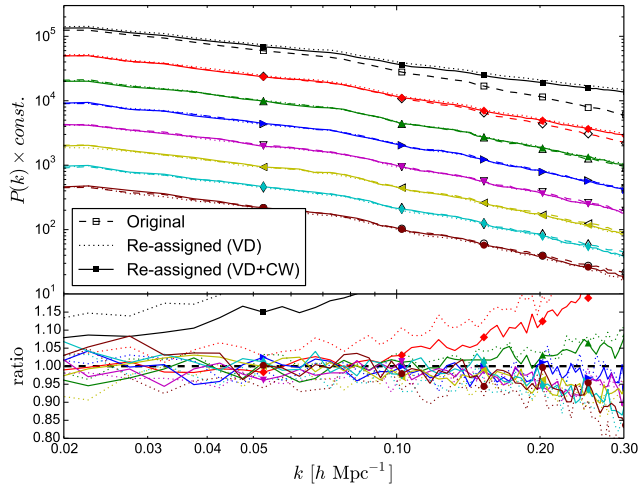
Figure 7. Power spectra of the sub-samples with different halo $V_{\max}$. Same convention as in Fig. 2. CW stands for the cosmic web classification additionally used in this case. The dotted lines indicate the results of §4.2.

Figure 8. Power spectra of the sub-samples with different halo $V_{\max}$. Same convention as in Fig. 2. EX stands for the halo-exclusion additionally used in this case. The dotted lines indicate the results in §4.3

only three classes of knots are shown to illustrate the various $V_{\max}$–density relations as a function of the mass of knots. Nevertheless, the total number of classes in the re-assignment procedure exceeds 500.

We build the same number of classes for the catalogue without halo mass, and then separately assign masses to haloes in different classes. This procedure leads to a clear improvement, as can be seen in Fig. 7. The power spectra are now compatible with the true ones up to $k \sim 0.15\,h\,\mathrm{Mpc}^{-1}$ within 5%. Nevertheless, the more massive haloes with the largest $V_{\max}$ still show an increasing systematic deviation towards high k. The next to the largest mass bin deviates by ~10% at $k = 0.2\,h\,\mathrm{Mpc}^{-1}$, and more than 20% at $k = 0.25\,h\,\mathrm{Mpc}^{-1}$, while the largest mass bin deviates already more than 20% at $k = 0.1\,h\,\mathrm{Mpc}^{-1}$. Let us therefore focus in the next section on the most massive haloes.

4.4 Halo-exclusion

According to the results of the previous section, the information of the mass function, local density, and cosmic web environment is not enough to accurately determine the mass of the most massive haloes. Still, a clear systematic deviation is present in the power spectra with a tendency to overestimate the clustering of massive objects (larger power towards high k).

At this stage, we should note that the mass assignment we have conducted, although depending on the environment, followed a random procedure, hereby ignoring deviations from the Poisson distribution beyond the ones present in the actual three-dimensional distribution of haloes. In this sense, we did not distinguish between distinct haloes and subhaloes. A distinct halo, and its subhaloes sharing the same dark matter environment, could get the same assigned mass with equal probability. Thus the re-assigned catalogue, does not prevent two massive haloes from being arbitrarily close to each other, and hence leads to a higher power spectrum. But actually, the halo-exclusion effect affecting massive haloes yields under-dispersed (dispersion smaller than
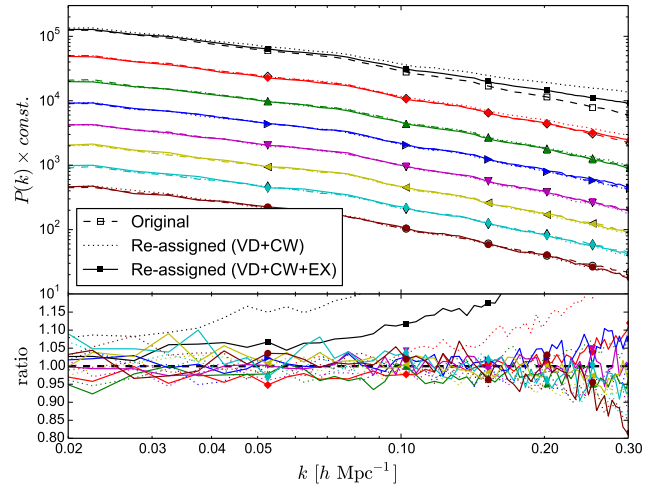
Poisson) distributions (e.g. Somerville et al. 2001; Casas-Miranda et al. 2002; Baldauf et al. 2013).

We therefore consider now the minimum separation between massive haloes $\Delta r_{\min}^M$. In particular, to separate massive haloes in the $V_{\max}$ assignment procedure, we perform an additional operation by setting a $V_{\max}$ threshold. In order to distribute haloes with $V_{\max}$ above the threshold to different cells as far as possible, we follow a top-down procedure, beginning with the highest $V_{\max}$ and continuing towards lower values, randomly selecting un-occupied cells to ensure that there is only one such halo in each cell. However, for a relatively low $V_{\max}$ threshold, there might not be available cells for all the haloes. In this case, we assign two or even more haloes to one cell.

This procedure can be refined using the distribution of separation between haloes for different mass bins extracted from N-body simulations and applied in a stochastic way, as we do with the $V_{\max}$–density relation. We leave such a study for future work. Nevertheless, the procedure described above serves to test our hypothesis and leads to already clear improvements. In particular, we find that most of the mass ($V_{\max}$) bins show power spectra which are compatible with the true ones within about 5%, up to $k \sim 0.2\,h\,\mathrm{Mpc}^{-1}$. Only the most massive bin still shows a clear systematic deviation which has been reduced to less than 15% up to $k \sim 0.15\,h\,\mathrm{Mpc}^{-1}$, indicating that we still need to increase the halo-exclusion effect for this mass bin (see Fig. 8).

As a further demonstration of the halo-exclusion correction performed in this section, we investigate the one-dimensional probability distribution function (PDF) of haloes with $V_{\max}$ above a threshold of $500\,\mathrm{km\,s}^{-1}$, as shown in Fig. 9. Here we can see that the PDF matches the true one only after applying the correction. We know from previous studies that an accurate PDF is especially important regarding the higher order statistics (e.g. Kitaura et al. 2014b). An analogous analysis has been conducted with FoF haloes and the results are shown in Appendix A1.

Let us now have a look at applications of this method and its performance in terms of 2- and 3-point statistics in
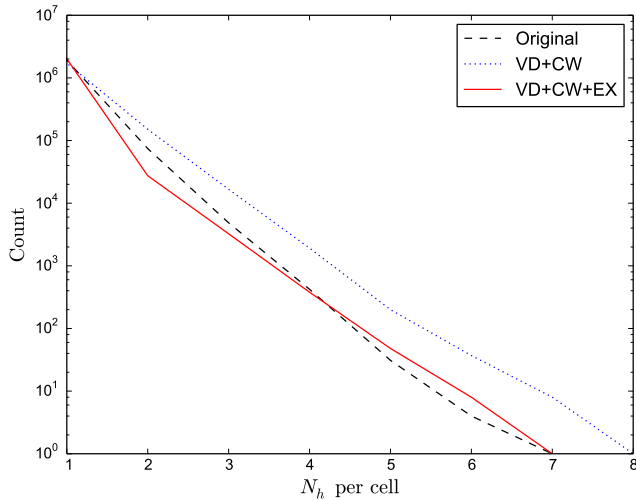
Figure 9. Probability distribution function of haloes with $V_{\rm max}$ above $500\,{\rm km\,s}^{-1}$.

Figure 10. Comparison of power spectra for the mock catalogues before $V_{\rm max}$ assignment.

5 APPLICATION: MOCK CATALOGUES

In the previous section, we have developed a prescription which allows us to accurately (within 15% up to $k \sim 0.15\,h\,{\rm Mpc}^{-1}$) describe the complex halo mass–environment dependence. We dub this method the Halo mAss Distribution ReconstructiON code (HADRON). As an application of our method, we study in this section the assignment of masses to mock halo catalogues constructed with perturbation theory. In particular, we consider two methods, the PerturbAtion Theory Catalogue generator of Halo and galaxY distributions (PATCHY) (Kitaura et al. 2014a,b) and the Effective Zel'dovich approximation mocks (EZMOCKS) (Chuang et al. 2015). Both methods provide mock halo catalogues calibrated with the BigMD BDM and FoF halo catalogues, as well as the dark matter particle distributions. While PATCHY includes an explicit Eulerian nonlinear and stochastic bias description, EZMOCKS uses effective modifications of the initial conditions and bias modelling to reproduce the bias of objects in the final catalogue. The dark matter density field used in EZMOCKS is given by the Zel'dovich approximation (Zel'dovich 1970), while PATCHY uses ALPT (Kitaura & Heß 2013). The different approximations have an impact on the accuracy of the bias, as we will show below.

5.1 The HADRON-code

We outline below the steps included in the HADRON-code to assign masses to haloes constructed with approximate gravity solver based mock generators.

(i) First, we compute the density field and cosmic web structures (knots, sheets, filaments, and voids) according to the dark matter particles from a reference N-body simulation. Then, we further classify the knots into different classes
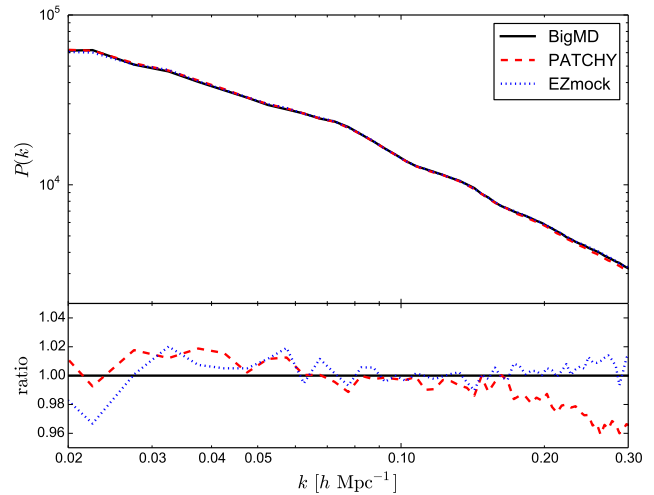
according to their enclosed mass (see details in §4.3). As a result, we obtain the density ($\rho_{\rm DM}$) and cosmic web classification type ($t_{\rm CW}$) for each cell.

(ii) Second, we compute the number of haloes in each density and cosmic web classification type bin according to the halo catalogue from the reference N-body simulation.

(iii) Third, we take the dark matter particles according to the approximate catalogue mock generator, and compute the density and cosmic web type in an analogous way to step (i). Since these quantities are different for simulations and mocks, we rank order those from mocks, in order to have an equivalent population of haloes in each $\rho_{\rm DM}$ and $t_{\rm CW}$ bin.

(iv) Fourth, we apply halo-exclusion to the halo catalogue from the mock generator, i.e. we assign mass above the threshold to haloes (see details in §4.4).

(v) Finally, we assign the mass to the rest of the haloes. For each mock halo without mass (some of them have already acquired a mass in step (iv)), we find the local density $\rho_{\rm DM}$ and cosmic web classification type $t_{\rm CW}$. From step (ii) we have the distribution of halo masses for a given $\rho_{\rm DM}$ and $t_{\rm CW}$, we then choose the mass of this halo with the probability from the corresponding distribution.

5.2 Mass assignment in perturbation theory based mocks

We apply the HADRON-code to PATCHY and EZMOCK based catalogues which have been generated using initial conditions based on the ones corresponding to the Planck BigMD simulation but re-sampled to a lower resolution of 960^3 cells. This reduces the cosmic variance, however, the population of haloes based on the dark matter field has still a random component (cf. Kitaura et al. 2014a; Chuang et al. 2015). As in section 4, we focus on the BDM halo catalogue from BigMD selected by $V_{\rm max}$.

Fig. 10 shows the power spectra of the entire PATCHY and EZMOCK mock halo catalogues compared to the BDM halo catalogue from BigMD, which serves as the reference for calibration of the mock catalogues. An analogous analysis has been conducted with FoF haloes and the results are
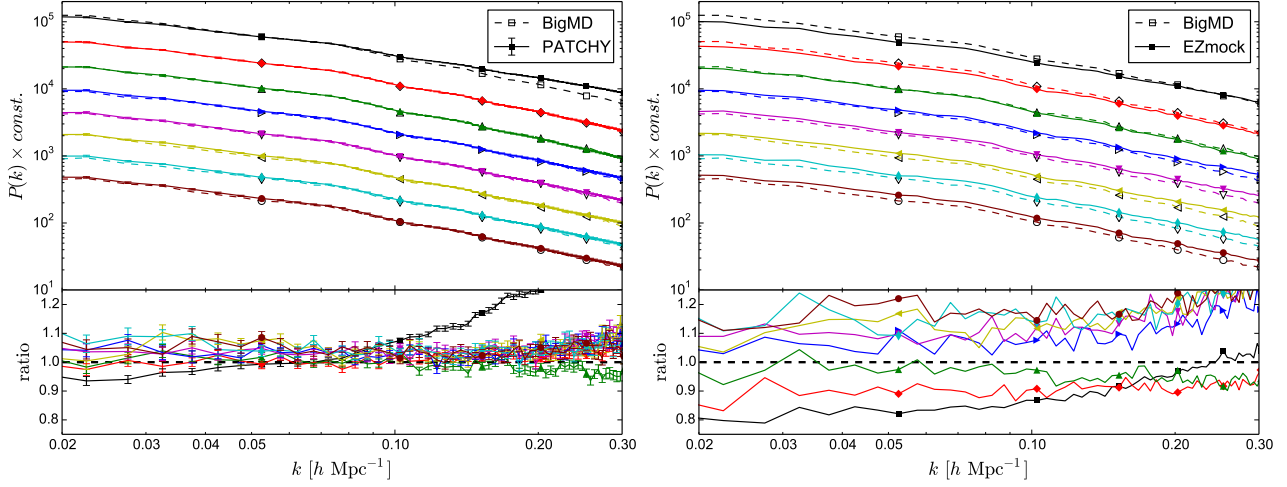
Figure 11. Power spectra of (left:) PATCHY and (right:) EZMOCK BDM mock catalogues in different V_{max} bins after V_{max} assignment.

shown in Appendix A2. EZMOCKS uses a clouds-in-cell based population of haloes, whereas PATCHY uses counts-in-cell in the population step according to the negative binomial distribution modelling over-dispersion, i.e., the deviation from Poissonity (Kitaura et al. 2014a). This results in slightly better agreements in terms of the power spectra for EZ-MOCKS based catalogues when using CIC estimators, as we do here and can be seen in Fig. 10. We assign V_{max} to these two mock catalogues using the procedure described in §5.1. However, as we have mentioned, we have still some degrees of freedom within this procedure, in the definition of the cosmic web structures (λ_{th}) and the V_{max} threshold. We calibrate these degrees of freedom according to the performance (power spectra of sub-samples) of the catalogues after the V_{max} assignment. In this case, we employ $\lambda_{th} = -0.25$, and the V_{max} threshold $V_{th} = 500 \, \text{km s}^{-1}$. Moreover, we generate 100 PATCHY realizations with the same initial conditions, but changing the random seeds constructing the halo catalogues. This permits us to estimate the error bars for the 2- and 3-point statistics sharing the same large scale cosmic variance for a consistent comparison to the N-body simulation. The comparison of power spectra for different V_{max} bins is shown on the left and right panels of Fig. 11 for PATCHY and EZMOCK mock catalogues, respectively.

We find that PATCHY reaches a significantly higher accuracy in the biased tracers, due to the more precise dark matter field used within the approach (ALPT vs Zel'dovich). In particular, the precision reached with PATCHY equals the one reached with the mass re-assignment tests using the N-body simulation (see §4.4), whereas the systematic deviations with EZMOCK grow to 20% within $k < 0.2 \, h \, \text{Mpc}^{-1}$. This means that the remarkable performance of EZMOCK in terms of the global 2- and 3-point statistics (on large scales) as shown in several works (Chuang et al. 2015, 2014) suffers from its crude dark matter density approximation (on small scales) with the mass assignment scheme presented in this work. Let us therefore continue our analysis focused on PATCHY. Fig. 13 shows the great performance of the 2- and 3-point correlation functions for different V_{max} bins using PATCHY. We see that the deviation found in the power spectrum for the most massive bin is also apparent on small scales in the 2-point correlation function (see black lines in
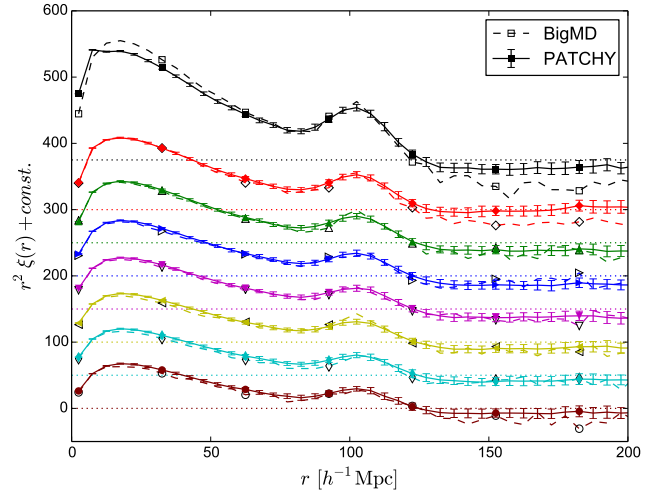

Figure 12. 2-point correlation functions of PATCHY BDM mock catalogues in different V_{max} bins after V_{max} assignment.

Fig. 13). Nevertheless, the BAO peak is matched within 1-σ for all mass bins, showing an excellent agreement down to the smallest scales ($\sim 5 \, h^{-1} \text{Mpc}$), but for the most massive bin. The 3-point correlation function is essentially compatible with the N-body simulation for the different mass bins. It is interesting to observe how the anisotropy increases towards higher masses, showing that the tracers with larger mass are less homogeneously distributed across the cosmic web.

6   CONCLUSIONS

In this work, we have studied the probabilistic dependence of the halo mass distribution as a function of local and non-local indicators, such as the local density, the cosmic web environment, and the halo-exclusion effect. We have found complex non-linear relations between the halo mass and the local density field, showing a degeneracy between parent haloes and subhaloes in certain density environments, as expected. Furthermore, we have used the non-local cosmic web
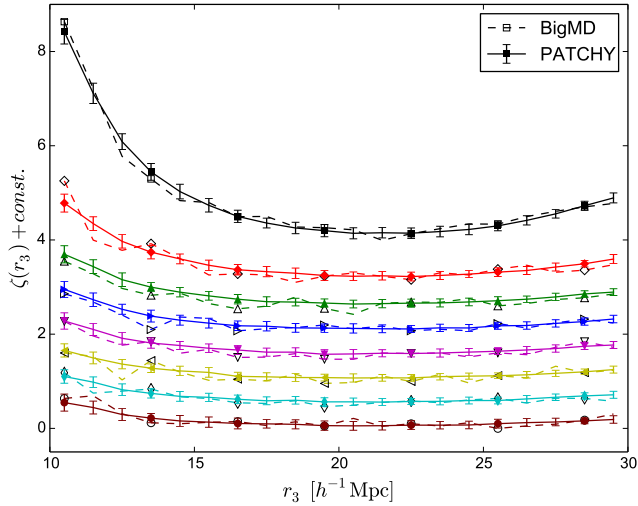
Figure 13. 3-point correlation functions of PATCHY BDM mock catalogues in different V_{max} bins after V_{max} assignment.

environment information according to the eigenvalues of the tidal field tensor. This permits us to find accurate statistical relations between the halo mass and the density and cosmic web environment. Such relations can be used to assign masses to a distribution of haloes. We dubbed the implementation of our method the HADRON (Halo mAss Distribution ReconstructiON) code. We first have tested this on the halo distribution of the Planck BigMultiDark simulation by ignoring the actual information of their masses to reconstruct them using the statistical relations found in this work. We furthermore tested this method on a halo distribution produced by perturbation theory based codes, such as PATCHY and EZMOCKS. Our results show that accurate perturbation theory models are required to properly model the halo mass to density relation. In particular, augmented Lagrangian perturbation theory (ALPT), as opposed to the Zel'dovich approximation, permits us to dramatically reduce the errors. We find that the resulting populations (classified into different mass bins) of haloes using ALPT agree in terms of power spectra within 1σ up to scales of $k = 0.2$ for different mass cuts, demonstrating that we recover the correct scale dependent bias on those scales. Only the most massive haloes ($V_{max} \gtrsim 550$ km s^{-1}) show a larger deviation. For these, we find evidence of the halo-exclusion effect, as a clear improvement is achieved when assigning those high masses with a minimum separation. Furthermore, we have computed the two- and three-point correlation functions finding an excellent agreement for arbitrary mass cuts.

This method can be applied for efficient massive production of mock halo or galaxy catalogues. Our work represents a quantitative application of the cosmic web classification. It can have further interesting applications in the multi-tracer analysis of the large-scale structure for future galaxy surveys.

REFERENCES

Abate et al. (LSST Dark Energy Science Collaboration), 2012, arXiv:1211.0310
Abramo L. R., Leonard K. E., 2013, MNRAS, 432, 318
Angulo R. E., Foreman S., Schmittfull M., Senatore L., 2014, arXiv:1406.4143
Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., 2010, MNRAS, 408, 2163
Baldauf T., Mercolli L., Mirbabayi M., Pajer E., 2014, arXiv:1406.4135
Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Phys.Rev.D, 86, 083540
Baldauf T., Seljak U., Smith R. E., Hamaus N., Desjacques V., 2013, Phys.Rev.D, 88, 083507
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Beltrán Jiménez J., Durrer R., 2011, Phys.Rev.D, 83, 103509
Benitez N. et al. (J-PAS Collaboration), 2014, arXiv:1403.5237
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Blake C. et al., 2013, MNRAS, 436, 3089
Casas-Miranda R., Mo H. J., Sheth R. K., Boerner G., 2002, MNRAS, 333, 730
Cen R., Ostriker J. P., 1992, ApJ, 399, L113
Chan K. C., Scoccimarro R., Sheth R. K., 2012, Phys.Rev.D, 85, 083509
Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, MNRAS, 446, 2621
Chuang C.-H., Zhao C., Prada F., Munari E., Avila S., Izard A., Kitaura F.-S., Manera M., et al. 2014, arXiv:1412.7729
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Dawson K. S. et al. (SDSS3 Collaboration), 2013, AJ, 145, 10
de Jong R. S. et al. (4MOST Collaboration), 2012, SPIE, 8446, 84460T
Dekel A., Lahav O., 1999, ApJ, 520, 24
Desjacques V., Crocce M., Scoccimarro R., Sheth R. K., 2010, Phys.Rev.D, 82, 103529

Eisenstein D. J. et al. (SDSS3 Collaboration), 2011, AJ, 142, 72

Elia A., Ludlow A. D., Porciani C., 2012, MNRAS, 421, 3472

Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815

Frieman J. et al. (Dark Energy Survey Collaboration), 2013, AAS, 221, #335.01

Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447

Fry J. N., Peebles P. J. E., 1978, ApJ, 221, 19

Gardner J. P., Connolly A., McBride C., 2007, arXiv:0709.1967

Gottloeber S., 1998, in Mueller V., Gottloeber S., Muecket J. P., Wambsganss J., eds, Large Scale Structure: Tracks and Traces Galaxy Tracers in Cosmological N-Body Simulations. pp 43–46

Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489

Hamaus N., Seljak U., Desjacques V., Smith R. E., Baldauf T., 2010, Phys.Rev.D, 82, 043515

Hoffman Y., Metuki O., Yepes G., Gottlöber S., Forero-Romero J. E., Libeskind N. I., Knebe A., 2012, MNRAS, 425, 2049

Jing Y. P., 2005, ApJ, 620, 559

Kaiser N., 1984, ApJ, 284, L9

Kitaura F.-S., 2013, MNRAS, 429, L84

Kitaura F.-S., Heß S., 2013, MNRAS, 435, L78

Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21

Kitaura F.-S., Gil-Marín H., Scoccola C., Chuang C.-H., Müller V., Yepes G., Prada F., 2014, arXiv:1407.1236

Klypin A., Holtzman J., 1997, astro-ph/9712217

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Klypin A., Yepes G., Gottlober S., Prada F., Hess S., 2014, arXiv:1411.4001

Laureijs R. et al. (EUCLID Collaboration), 2009, arXiv:0912.0914

McBride C. K., Connolly A. J., Gardner J. P., Scranton R., Newman J. A., Scoccimarro R., Zehavi I., Schneider D. P., 2011, ApJ, 726, 13

McDonald P., Roy A., 2009, J.Cosm.Astr.Phys., 8, 20

McDonald P., Seljak U., 2009, Journal of Cosmology and Astroparticle Physics, 10, 007

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

Muñoz-Cuartas J. C., Müller V., Forero-Romero J. E., 2011, MNRAS, 417, 1303

Muñoz-Cuartas J. C., Müller V., 2012, MNRAS, 423, 1583

Nuza S. E., Sánchez A. G., Prada F., Klypin A., Schlegel D. J., Gottlöber S., Montero-Dorta A. D., et al 2013, MNRAS, 432, 743

Peacock J. A., Heavens A. F., 1985, MNRAS, 217, 805

Peebles P. J. E., 1980, The large-scale structure of the universe

Pen U.-L., 1998, ApJ, 504, 601

Percival W. J., Verde L., Peacock J. A., 2004, MNRAS, 347, 645

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

Press W. H., Schechter P., 1974, ApJ, 187, 425

Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottloeber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2011, arXiv:1109.0003
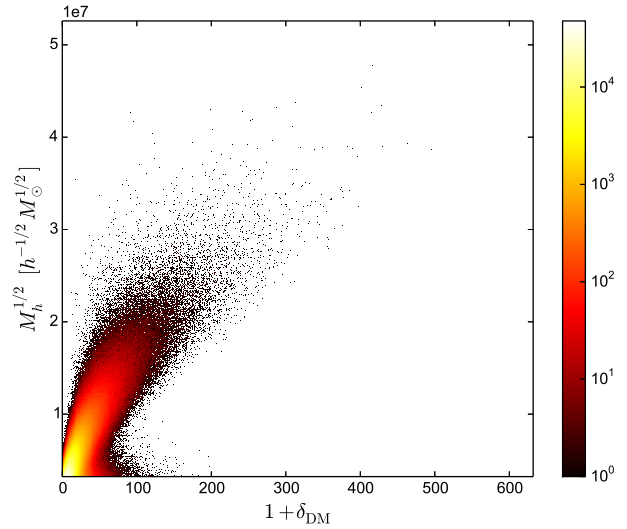
Figure A1. Number of FoF haloes in BigMD with certain mass and local DM density.

Saito S., Baldauf T., Vlah Z., Seljak U., Okumura T., McDonald P., 2014, Phys.Rev.D, 90, 123522

Schlegel D. et al. (DESI Collaboration), 2011, arXiv:1106.1706

Seljak U., 2000, MNRAS, 318, 203

Seljak U., Hamaus N., Desjacques V., 2009, Physical Review Letters, 103, 091303

Sheth R. K., Lemson G., 1999, MNRAS, 304, 767

Sheth R. K., Tormen G., 2004a, MNRAS, 349, 1464

Sheth R. K., Tormen G., 2004b, MNRAS, 350, 1385

Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys.Rev.D, 75, 063512

Somerville R. S., Lemson G., Sigad Y., Dekel A., Kauffmann G., White S. D. M., 2001, MNRAS, 320, 289

Springel V., 2005, MNRAS, 364, 1105

Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16

Valageas P., Nishimichi T., 2011, A&A, 527, A87

Wang H., Mo H. J., Jing Y. P., Guo Y., van den Bosch F. C., Yang X., 2009, MNRAS, 394, 398

Wang H., Mo H. J., Yang X., van den Bosch F. C., 2012, MNRAS, 420, 1809

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Zel'dovich Y. B., 1970, A&A, 5, 84

APPENDIX A: FOF HALO CATALOGUES SELECTED BY MASS

This appendix shows the analysis performed using FoF haloes in an analogous way to the BDM halo analysis shown in the main text (see §4 and 5.2).

A1 *N*-body simulation based catalogues

When applying the HADRON method to FoF catalogues, we have to adopt mass assignment instead of V_{max} assignment, since V_{max} is not available for FoF haloes. In this case, the extracted mass–density relation is shown in Fig. A1.
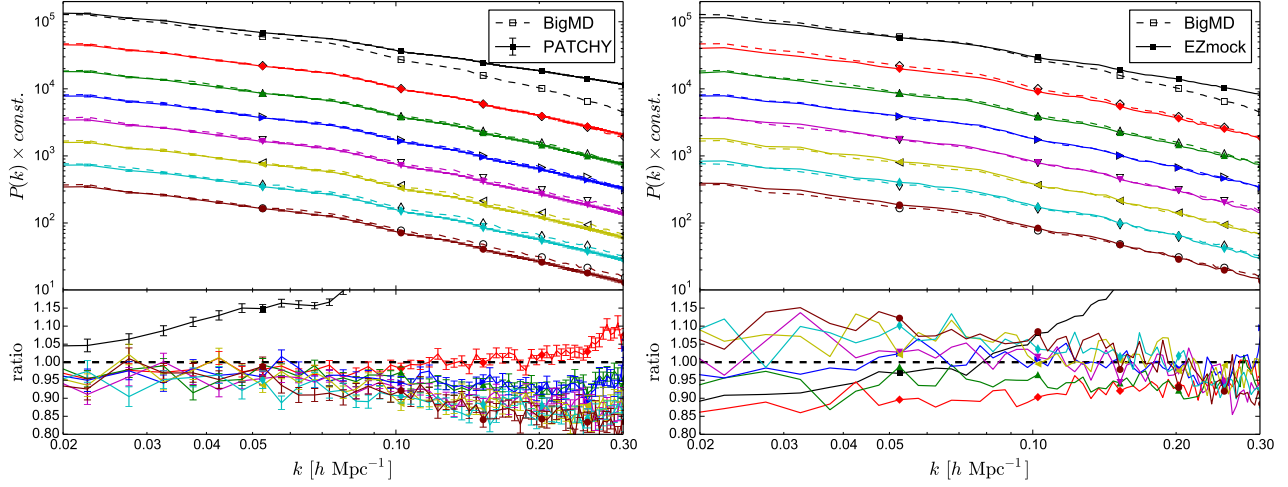
Figure A4. Power spectra of (left:) PATCHY and (right:) EZMOCK FoF mock catalogues in different mass bins after mass assignment.
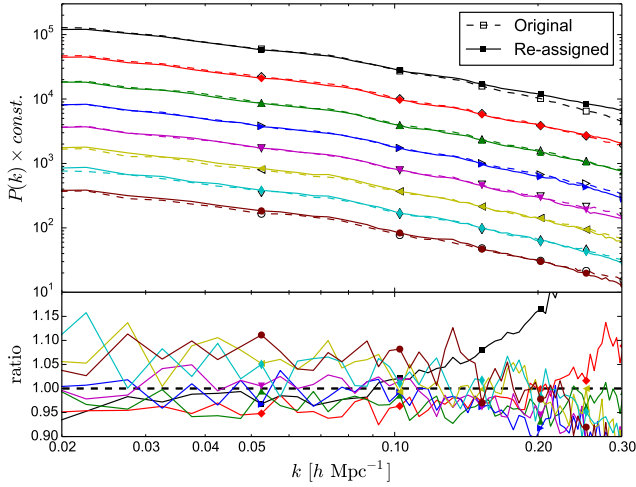

Figure A2. Power spectra of re-assigned sub-catalogues with different halo masses, in comparison to the original FoF mock catalogues.
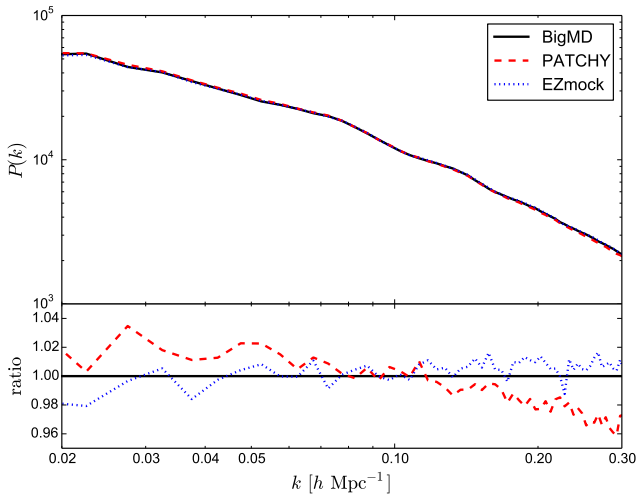

Figure A3. Comparison of power spectra for the catalogues before mass assignment.

We then follow the same steps as for the $V_{\max}$ assignment (see §4), i.e. classify the cosmic web structures and employ a mass threshold. To justify the re-assigned result, we also divide the catalogues into 8 sub-samples of different halo masses, i.e. $\{[<1.17),\ [1.17\text{-}1.33),\ [1.33\text{-}1.54),\ [1.54\text{-}1.84),\ [1.84\text{-}2.28),\ [2.28\text{-}3.08),\ [3.08\text{-}4.92),\ [\geqslant 4.92]\} \times 10^{13}\,h^{-1}M_\odot$. Fig. A2 displays the power spectra of the sub-samples drawn from the original BIGMD FoF halo catalogue and that after mass re-assignment. The mass threshold for the exclusion operation are $2.51 \times 10^{13}\,h^{-1}M_\odot$ for both catalogues.

We note that the performance for FoF is worse than for the BDM catalogue, reaching deviations of about 10% already at $k \sim 0.15\,h\,\mathrm{Mpc}^{-1}$.

A2 Perturbation theory based catalogues

For the FoF samples, the performance of PATCHY and EZ-MOCK is shown in Fig. A3. With $\lambda_{\mathrm{th}} = -0.25$, and mass threshold of $2.51 \times 10^{13}\,h^{-1}M_\odot$ for the mass assignment procedure, the power spectra for different mass bins is shown on the left and right panels of Fig. A4 for PATCHY and EZ-MOCK mock catalogues, respectively. The performance for FoF with PATCHY is worse than for the BDM catalogues. We have seen in A1, that this is also true for the N-body simulation. Nevertheless, another reason is that PATCHY has been designed to model over-dispersion, which is present in the BDM catalogues when taking the full population of haloes, including subhaloes, but not true for the FoF catalogue.