The Astrophysical Journal, 835:69 (9pp), 2017 January 20 doi:10.3847/1538-4357/835/1/69

LOG-NORMAL DISTRIBUTION OF COSMIC VOIDS IN SIMULATIONS AND MOCKS

E. Russell[1] and J.-R. Pycke[1,2]
[1] Division of Science and Mathematics, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE; er111@nyu.edu
[2] Laboratoire de Mathématiques et Modélisation d'Évry, Université d'Évry, P.O. Box 91037, Évry Cedex, France; jrp15@nyu.edu
Received 2016 October 27; revised 2016 December 7; accepted 2016 December 7; published 2017 January 19

ABSTRACT

Following up on previous studies, we complete here a full analysis of the void size distributions of the Cosmic Void Catalog based on three different simulation and mock catalogs: dark matter (DM), haloes, and galaxies. Based on this analysis, we attempt to answer two questions: Is a three-parameter log-normal distribution a good candidate to satisfy the void size distributions obtained from different types of environments? Is there a direct relation between the shape parameters of the void size distribution and the environmental effects? In an attempt to answer these questions, we find here that all void size distributions of these data samples satisfy the three-parameter log-normal distribution whether the environment is dominated by DM, haloes, or galaxies. In addition, the shape parameters of the three-parameter log-normal void size distribution seem highly affected by environment, particularly existing substructures. Therefore, we show two quantitative relations given by linear equations between the skewness and the maximum tree depth, and between the variance of the void size distribution and the maximum tree depth, directly from the simulated data. In addition to this, we find that the percentage of voids with nonzero central density in the data sets has a critical importance. If the number of voids with nonzero central density reaches $\geqslant 3.84\%$ in a simulation/mock sample, then a second population is observed in the void size distributions. This second population emerges as a second peak in the log-normal void size distribution at larger radius.

Key words: catalogs – galaxies: clusters: intracluster medium – large-scale structure of universe – methods: numerical – methods: statistical

1. INTRODUCTION

The large-scale structure of the present-day universe has been intricately formed by an interplay between random Gaussian fluctuations and gravitational instability. When gravitational instabilities start to dominate the dynamical evolution of the matter content of the universe, the formation of structure evolves from a linear to a highly nonlinear regime. In this framework, voids are formed in minima and haloes are formed in maxima of the same primordial Gaussian field, and later on these features present different types of dynamical evolution due to their initial conditions in the nonlinear regime. This fact has been known since early studies showed that voids are integral features of the universe (Chincarini & Rood 1975; Gregory & Thompson 1978; Einasto et al. 1980; van de Weygaert & Platen 2011). Sheth & van de Weygaert (2004) and Russell (2013, 2014) show that the distribution of voids can be affected by their environments. As a result, the void size distribution may play a crucial role in understanding the dynamical processes affecting the formation of structure in the universe (Goldberg & Vogeley 2004; Croton et al. 2005; Hoyle et al. 2005).

The early statistical models of void probability functions (VPFs) (Fry 1986; Elizalde & Gaztanaga 1992) are based on the counts in randomly placed cells following the prescription of White (1979). Apart from VPFs, the number density of voids is another key statistic necessary to obtain the void distribution. Recently Pycke & Russell (2016) showed that void size distributions obtained from the Cosmic Void Catalog (CVC) satisfy a three-parameter log-normal probability function. This is particularly interesting, because observations and theoretical models based on numerical simulations of galaxy distributions (Hamilton 1985; Coles & Jones 1991; Bernardeau 1992, 1994; Bouchet et al. 1993; Kofman et al. 1994; Taylor & Watts 2000; Kayo et al. 2001) show that the mass distribution of the Galaxy satisfies a log-normal function rather than a Gaussian. Taking into account that voids are integral features of the universe, one may expect to obtain a similar distribution profile for voids. Apart from this, Pycke & Russell (2016) discuss a possible quantitative relation between the shape parameters of the void size distribution and the environmental effects.

Following up on the study of Pycke & Russell (2016), we here extend their analysis of void size distributions to all simulated and mock samples of CVC of Sutter et al. (2012). The three main catalogs under study are dark matter (DM), halo, and galaxy catalogs. Therefore, we confirm that the system of a three-parameter log-normal distribution obtained by Pycke & Russell (2016) provides a fairly satisfactory model of the size distribution of voids. In addition to this, we obtain equations that satisfy linear relations between maximum tree depth and the shape parameters of the void size distribution as proposed by Pycke & Russell (2016).

2. VOID CATALOG: SIMULATIONS AND MOCK DATA

Extending the study by Pycke & Russell (2016), we here fully investigate the void size distribution function statistically in simulations and mocks catalogs of the public CVC of Sutter et al. (2012). It is useful to note that all the data of CVC used here are generated from a Λ cold dark matter (ΛCDM) N-body simulation by using an adaptive treecode 2HOT (Sutter et al. 2014a, 2014b). In addition, in all data sets voids are identified with the modified version of the parameter-free void finder ZOBOV (Neyrinck 2008; Lavaux & Wandelt 2012; Sutter et al. 2012). The data sets of CVC we use here can be categorized into three main groups:

1. DM simulations are DM Full, DM Dense, and DM Sparse. Although these DM simulations have the same cosmological parameters from the *Wilkinson Microwave Anisotropy Probe* seven-yeardata release (DR7, Komatsu et al. 2011) as well as the same snapshot at $z = 0$, they have different tracer densities of 10^{-2}, 4×10^{-3}, and 3×10^{-4} particles $(\mathrm{Mpc}/h)^{-3}$, which are respectively DM Full, DM Dense, and DM Sparse. Also the minimum effective void radii $R_{\mathrm{eff,min}} = 5$, 7, and 14 Mpc/h are obtained from the simulations for DM Full, DM Dense, and DM Sparse respectively.
2. The halo catalog in which two halo populations are generated: Haloes Dense and Haloes Sparse. In the halo catalog the halo positions are used as tracers to find voids. The minimum resolvable halo mass of Haloes Dense is $1.47 \times 10^{12}\, M_{\odot}/h$ while that of Haloes Sparse is $1.2 \times 10^{13}\, M_{\odot}/h$. In addition, the minimum effective void radii of Haloes Dense and Sparse are $R_{\mathrm{eff,min}} = 7$ and 14 Mpc/h respectively. The main reason to construct these halo populations with different minimum resolvable halo masses is to compare the voids in haloes to voids in relatively dense galaxy environments; see Sutter et al. (2014a) for more details.
3. Galaxy catalogs: there are two galaxy mock catalogs, which are produced from the above halo catalog by using the Halo Occupation Distribution (HOD) code of Tinker et al. (2006) and the HOD model by Zheng et al. (2007). These galaxy mock catalogs are called HOD Dense and HOD Sparse (Sutter et al. 2014a). The HOD Dense catalog has 9503 voids with effective minimum radii $R_{\mathrm{eff,min}} = 7\ \mathrm{Mpc}/h$ and includes relatively high-resolution galaxy samples with density 4×10^{-3} DM particles $(\mathrm{Mpc}/h)^{-3}$, matching the main sample of the Sloan Digital Sky Survey (SDSS) DR7 (Strauss 2002) using one set of parameters found by Zehavi et al. (2011) ($\sigma_{\log M} = 0.21$, $M_0 = 6.7 \times 10^{11} h^{-1} M_{\odot}$, $M_1' = 2.8 \times 10^{13} h^{-1} M_{\odot}$, $\alpha = 1.12$). The HOD Sparse mock catalog consists of 1422 voids with effective minimum radii $R_{\mathrm{eff,min}} = 14\ \mathrm{Mpc}/h$, and this void catalog represents a relatively low-resolution galaxy sample with density 3×10^{-4} particles $(\mathrm{Mpc}/h)^{-3}$, matching the number density and clustering of the SDSS DR9 galaxy sample (Dawson et al. 2013) using the parameters found by Manera et al. (2013) ($\sigma_{\log M} = 0.596$, $M_0 = 1.2 \times 10^{13} h^{-1} M_{\odot}$, $M_1' = 10^{14} h^{-1} M_{\odot}$, $\alpha = 1.0127$, and M_{min} chosen to fit the mean number density). In addition to this another mock galaxy catalog is used here: the *N*-body Mock catalog, which is a single HOD Mock galaxy catalog in real space at $z = 0.53$, generated by a DM simulation of 4096^3particles (with a particle mass resolution $7.36 \times 10^{10} h^{-1} M_{\odot}$) in a 4 Gpc/h box; it is tuned to SDSS DR9 in full cubic volume by using the HOD parameters found in Manera et al. (2013) and it consists of 155,196 voids (Sutter et al. 2014b). Although the *N*-body Mock catalog is processed slightly differently than HOD Sparse and HOD Dense, it is a HOD mock catalog and it uses *Planck* first-year cosmological parameters (Planck Collaboration 2014).

In the following section, we examine the above data sets from a statistical perspective, such as histograms, parameters of location (range, mean, median), mode or dispersion (standard deviation), and shape (skewness, kurtosis) by following the previous study of Pycke & Russell (2016). From a statistical perspective, we also investigate the connection between the distribution and the environment of void populations.

3. STATISTICAL PROPERTIES OF VOID DISTRIBUTIONS

As a first step, the raw data plots of void size distributions are obtained for DM Full, DM Dense, DM Sparse, Haloes Dense, and Haloes Sparse. Note that the void size distributions for HOD Dense, HOD Sparse, and *N*-body mock data sets are discussed in great detail in Pycke & Russell (2016). In the raw void size distributions, an unexpected local peak is observed around the value 20 Mpc/h in the DM Full sample and around 27 Mpc/h in the DM Dense sample; see upper and lower left panels in Figure 1. A similar behavior is observed by Pycke & Russell (2016) in the *N*-body Mock sample around the value 50 Mpc/h. It is crucial to mention that the samples DM Full and DM Dense are high-resolution data sets, unlike the samples we investigate here—DM Sparse, Haloes Dense, and Haloes Sparse. We find that the low-resolution samples DM Sparse and Haloes Dense show single populations (see Figure 2). In the previous study, Pycke & Russell (2016) show that HOD Sparse and HOD Dense samples also present single-population void size distributions. Note that our goal is not to compare the invariant distributions in Figure 2, but to find the reason for the emergence of the second population (Figure 1), in contrast to the single population in the void size distributions of CVC. After analyzing the components of the data sets, we realize that the second peak is caused by a second void population in the samples. We find that in both cases the second void population has a higher central density than the dominant voids of the distributions, which have zero central densities. Following the same strategy as Pycke & Russell (2016) to investigate these samples in terms of statistical properties, the second peaks in DM Full and DM Dense are excluded by excluding voids with nonzero central densities from the data sets. This procedure yields new subsamples of DM Full and DM Dense. As is seen in the upper and lower right panels in Figure 1, the data samples present void size distributions obtained by single void populations. Although we eliminate the second peak to investigate these data sets in a statistical framework, there are two questions left to answer to understand the occurrence of more than one void population in the void size distributions: Why can we not observe a second peak in the remaining data sets—DM Sparse, Haloes Dense, and Haloes Sparse (see Figure 2) although all samples include some fraction of voids with nonzero central densities? Is it possible to find a criterion for observing two different void populations in the void size distributions due to their central densities, particularly in the simulation and mock data?

Here we attempt to address these questions. Taking into account that *N*-body Mock, DM Full, and DM Dense are high-resolution data sets, we expect to observe more substructures in the simulations, which leads to the formation of two different void populations, particularly in terms of central densities, indicating two different void environments. We find that the emergence of multiple populations in the void size distributions is directly correlated with the number of voids with nonzero central densities, and especially with the highest central densities $\rho_{\mathrm{cent}} \in [0.2, 0.09]$. Table 1 presents the total number of voids in the samples and the percentage of voids with central densities in the range [0.2–0.09], which includes the densest voids in each sample. As is seen in Table 1, the contribution of the larger central densities is low compared to that of voids with zero

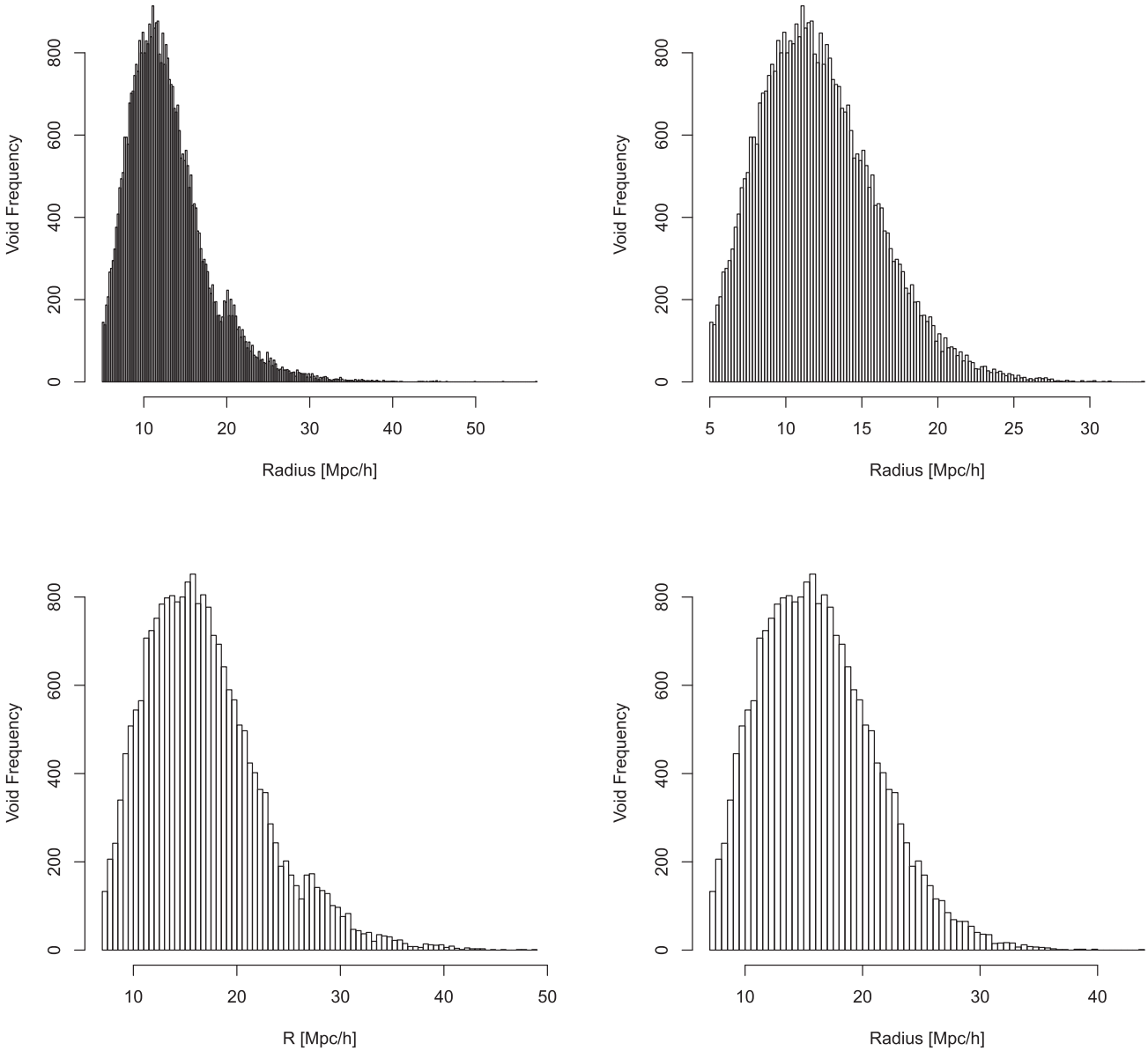

Figure 1. Void size distributions of DM Full (upper panels) and DM Dense (lower panels) data sets. The upper and lower left panels present the peak formations due to the presence of the second void population, whereas the right panels show their respective subsamples consisting of the void populations with only zero central density.

central densities, <3.84% of all samples that present a single population. What we observe is that the second population forms a peak in the void size distribution as long as the percentage of voids with nonzero density is at least 3.84% of the overall data set (see Table 1). It seems that the contribution of voids with large central densities to the overall void population holds an answer to the question of whether a second or even a further population is formed in the void size distributions.

After reducing the DM Full and DM Dense samples to a single population, we can now investigate the statistical properties of the void size distributions:

1. Mean ($\bar{r}$):

$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{1}$$

where N is the total number of voids and r_i the radius/size of each void in a given sample.

2. Centered Moments:

$$m_k = \frac{1}{N}\sum_{i=1}^{N}(r_i - \bar{r})^k,\ k = 2, 3, \ldots \tag{2}$$

For instance m_3 and m_4 are related to skewness and kurtosis, respectively, but they are influenced by the unit of measurement.

3. Variance m_2, a measurement of the dispersion of the data. The skewness and kurtosis are formulated using the variance and the higher moments m_3 and m_4.
   (a) Skewness b_1: a measurement of the degree to which a distribution is asymmetrical.

$$b_1 = \frac{m_3^2}{m_2^3} \tag{3}$$

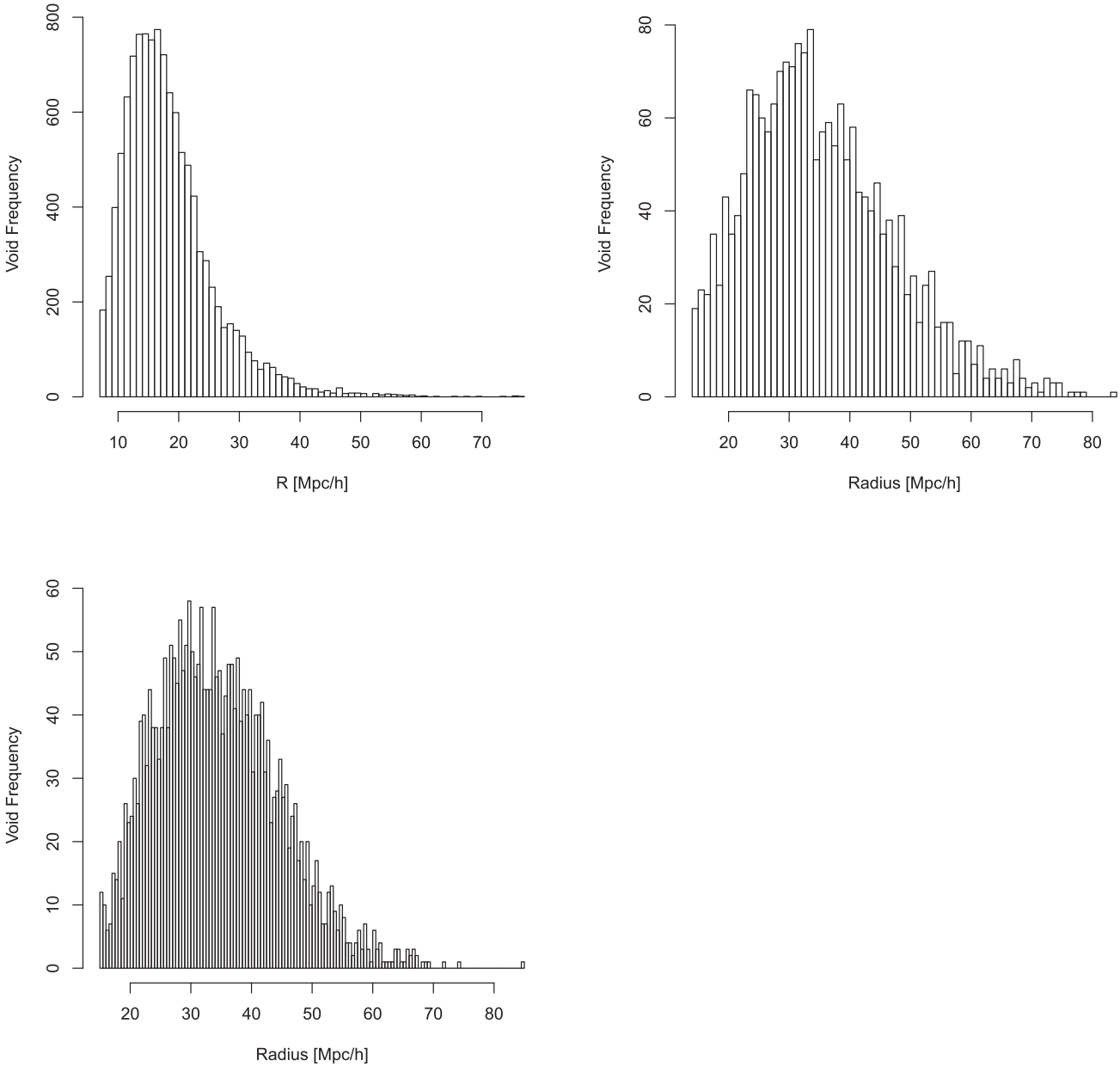

Figure 2. Void size distributions of Haloes Dense (upper right), Haloes Sparse (upper left), and DM Sparse (lower left).

(b) Kurtosis b_2: a measurement of the degree of peakedness.

$$b_2 = \frac{m_4}{m_2^2} \tag{4}$$

(see formulas (1.235)–(1.236) in Johnson et al. (1994, p. 51) and formulas (3.85)–(3.86) in Kendall & Stuart (1977, p. 85)). The shape parameters of the void size distributions for each sample are presented in Table 2. According to Table 2, all the void size distributions share the property of being significantly positively skewed ($m_3 > 0$), while the values of the kurtosis indicate a leptokurtic ($b_2 > 3$) behavior. Note that a leptokurtic type of distribution is characterized by a high degree of peakedness (see Sheskin 2011, pp. 16–30 for more details). The same result of positively skewed and leptokurtic void size distributions is also pointed out by Pycke & Russell (2016) for three samples of CVC.

Table 1
Percentages of Number of Voids with Central Densities ρ_{cent} in the Interval [0.2, 0.09] and the Total Number of Voids, $N_{\text{Tot-voids}}$, in each Sample

Sample	Percentages of N_{voids} ($\rho_{\text{cent}} \in [0.2, 0.09]$)	$N_{\text{Tot-voids}}$
HOD Dense	2.66	9503
HOD Sparse	2.65	1422
DM Full	5.32	42948
DM Dense	3.84	21865
DM Sparse	≈0.54	2611
Haloes Dense	3.03	11384
Haloes Sparse	2.65	2073
N-body Mock	≈5	155196

We would like to extend and modify here the discussion from Pycke & Russell (2016) in which they propose the maximum tree depth as an environmental indicator. Note that Sutter et al. (2014a) define the maximum tree depth as the

Table 2
First Moments, Skewness b_1, Kurtosis b_2, and Maximum Tree Depth (from Sutter et al. 2014a) of the Sample Distributions

Sample	HOD Sparse	HOD Dense	*N*-body Mock	Haloes Dense
$\bar{r}$	40.4	16.7	32.0	18.3
m_2	236	40.1	96.0	52.5
m_3	4420	386	899	567
m_4	307000	11100	38100	19400
b_1	1.49	2.32	0.912	2.23
b_2	5.52	6.93	4.13	7.07
Maximum Tree Depth	4	10	⋯	7
Sample	**DM Dense**	**DM Full**	**Haloes Sparse**	**DM Sparse**
$\bar{r}$	16.3	12.2	36.4	34.4
m_2	24.5	24.5	178	102
m_3	73.3	32.5	2190	596
m_4	1970	736	127000	34600
b_1	0.365	0.434	0.841	0.328
b_2	3.29	3.49	3.96	3.28
Maximum Tree Depth	3	⋯	2	0

Note. Note that the moments of the *N*-body Mock, DM Full and DM Dense distributions are computed for a single population.

length from root to tip of the tallest tree in the hierarchy, and it indicates the number of substructures in the most complex void in the sample. Taking one step beyond the point of Pycke & Russell (2016) and using the values of maximum tree depth from Sutter et al. (2014a), we provide here a simple linear relation indicating that the shape of the void size distribution is strongly correlated with the maximum tree depth of the simulated void catalogs. Table 2 also presents the maximum tree depth as well as the parameters of the void size distribution of the samples: HOD Dense, HOD Sparse, Haloes Dense, Haloes Sparse, DM Dense, and DM Sparse. As is seen, in our computations the samples *N*-body Mock and DM Full are not taken into account since the shape parameters of their void size distributions are obtained after excluding the second populations in the samples in order to provide the correlations without any interference. Based on the maximum tree depth and the skewness parameters of the samples from Table 2 we obtain a linear function,

$$b_1 = 0.2104 \pm 0.0324\ \mathrm{MTD} + 0.4735 \pm 0.1881,\quad R^2 = 0.9337, \tag{5}$$

where MTD refers to the maximum tree depth and $R^2 = 0.9337$ is the regression of the data. According to Equation (5), when the number of substructures is high (high MTD value), the void size distributions tend to be more positively skewed. Also, if one obtains the maximum tree depth of a sample, then it is possible to have a fairly good estimation of the amount of skewness of the void size distribution. As is seen in Figure 3, the higher the value of MTD, the more skewed the distribution is. This may lead to the fact that the skewness of a void size distribution can be a good indicator of the number of substructures in a sample. On the other hand, we must take into account a crucial parameter that can affect the relation between skewness and MTD. This parameter is the minimum radius-cut in the data sets. Basically there two main

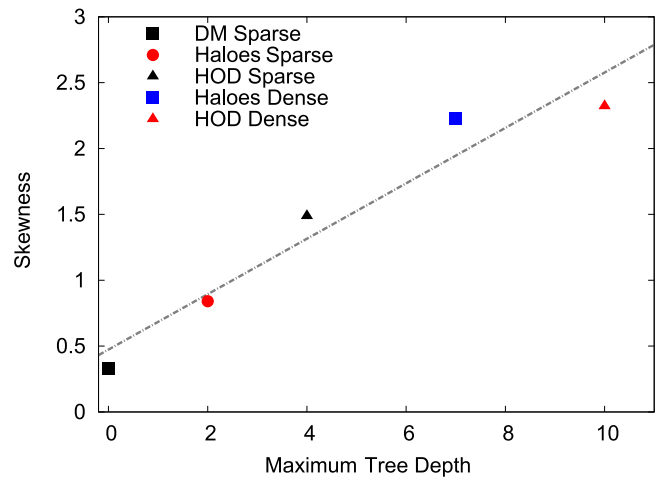

Figure 3. Relation between skewness and maximum tree depth for DM Sparse, Haloes Dense, Haloes Sparse, HOD Sparse, and HOD Dense.

density-based criteria that are imposed at different stages of the data production of CVC; the first threshold-cut comes from ZOBOV, in which voids only include as members Voronoi cells with density <0.2 times the mean particle density (Sutter et al. 2014a). In the second density criterion only voids with mean central densities <0.2 times the mean particle density are included. This is a particularly important criterion since it gives an insight about the radius-cut. Sutter et al. (2014a) give the central density within a sphere with radius

$$R = \frac{1}{4} R_{\mathrm{eff}}, \tag{6}$$

in which the effective radius R_{eff} is obtained from the following equation (Sutter et al. 2014a):

$$R_{\mathrm{eff}} = \left(\frac{3V}{4\pi}\right)^{1/3}. \tag{7}$$

Here V stands for total void volume. Sutter et al. (2014a) point out that they ignore voids with R_{eff} below the mean particle spacing of the tracer population. As a result, this constraint on the minimum radius-cut imposed by the density criteria can affect the skewness–MTD relation. Apart from the linear skewness–MTD relation, we investigate the correlation between variance/dispersion and MTD. Hence, It is found that the correlation between MTD and variance shows a distinction between Sparse and Dense samples; see Figure 4. Because sparse data show high dispersion by their nature, unlike dense data, it is an expected result to observe two main dispersions. As is seen in Figure 4, while sparse data show a fairly good (with high regression, $R^2 = 0.9938$) linear relation between MTD and variance (black line in Figure 4),

$$\mathrm{variance}(m_2)_{\mathrm{Sparse}} = 33.438(\pm 2.634)\mathrm{MTD} + 105.041(\pm 6.801),\ R^2 = 0.9938, \tag{8}$$

the variance of the dense data sets does not give enough information about the relation between MTD and dispersion (red line in Figure 4) although fitting to the dense data points

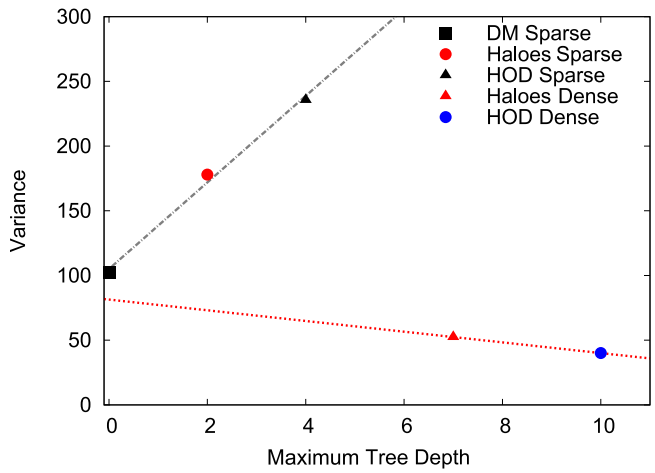

Figure 4. Relation between MTD and variance of the sparse and dense data sets.

gives a linear fit without an error.

$$\text{variance}(m_2)_{\text{Dense}} = -4.134\text{MTD} + 81.392,\ R^2 = 1. \quad (9)$$

This result is mathematically inconsistent since the variance of a distribution is always a positive number by definition. Thus the negative value in Equation (9) cannot be accepted. It seems that we need more data points to show a direct relation between MTD and the variance. Therefore we cannot conclude here that there is a direct relation between the variance and the MTD of the dense data sets.

We also extend the data points in the skewness and kurtosis plane (b_1, b_2) given by Pycke & Russell (2016), shown as the log-normal line in Figure 5. Figure 5 shows the void size distributions generated from the simulations and mock samples of CVC, and these distributions can be considered to behave as log-normal distributions with respect to their skewness and kurtosis.

Extending the data sets from the previous paper of Pycke & Russell (2016), we confirm here that the three-parameter log-normal distribution with random variable $\boldsymbol{R} = \boldsymbol{LN}(\theta, \zeta, \sigma)$,

$$p_{\theta,\zeta,\sigma}(R) = \frac{e^{-\frac{\{\log(R-\theta)-\zeta\}^2}{2\sigma^2}}}{(R-\theta)\sigma\sqrt{2\pi}}, \quad R > \theta, \quad (10)$$

fits the void size distributions obtained from the simulated and mock data sets of CVC. It is important to note that in this study, following up on the previous paper by Pycke & Russell (2016), we use the moment method to obtain the fit. This method is one of the standard methods in the field of estimation. We prefer this method to the maximum likelihood method because of the technical uncertainties and difficulties related to the latter according to Johnson et al. (1994, see p. 228). In addition to this, we do not provide here the standard goodness-of-fit tests, such as Anderson–Darling, Cramér–von Mises or Kolmogorov, to the data samples of CVC for the following reasons. Sheskin (2011) cites the studies of Conover (1980a, 1980b), which point out that if one employs a large enough sample size, almost any goodness-of-fit test will result in rejection of the null hypothesis. Conover (1980a, 1980b) also states that if the sample data are reasonably close to the hypothesized

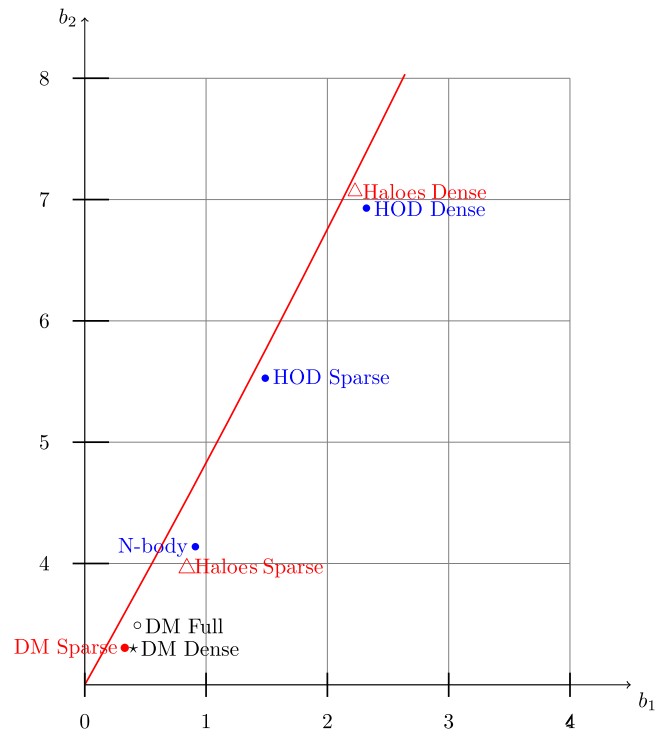

Figure 5. Kurtosis b_2 vs. skewness b_1 of the sample data in which the solid red line represents the standard three-parameter log-normal distribution. Note that here the parameters of the *N*-body (Pycke & Russell 2016), DM Full, and DM Dense samples are obtained after excluding the second populations to obtain statistical properties.

distribution, one can probably operate on the assumption that the sample data provide an adequate fit for the hypothesized distribution. Taking into account the size of the simulated and mock samples as well as the uncertainties of the construction of CVC (for example, minimum radius-cuts), the moment method provides a straightforward tool to obtain the parameters of the distribution (see the formulas of the first three sample moments in Johnson et al. 1994, p. 228).

In the above distribution formula (10), a random variable $\boldsymbol{R}$ is defined by $\boldsymbol{LN}(\theta, \zeta, \sigma)$ if $\log(\boldsymbol{R} - \theta)$ follows a Gaussian distribution with mean ζ and variance σ^2 given by Johnson et al. (1994). Johnson et al. (1994) describe the characteristics of a random variable $\boldsymbol{LN}(\theta, \zeta, \sigma)$ as

1. range: (θ, ∞),
2. mode: $\theta + e^{\zeta-\sigma^2}$,
3. median: $\theta + e^{\zeta}$,
4. mean: $\theta + e^{\zeta+\sigma^2/2}$.

Some characteristics of a log-normal random variable as well the relations between the shape parameters and estimators to fit the data samples are discussed in great detail by Johnson et al. (1994). One can obtain the estimators by using the above characteristics with the shape parameters as indicated by Johnson et al. (1994). We also provide here the estimators of the log-normal void size distributions of the Haloes Dense, DM Dense, and DM Full data sets in Table 3. The estimates of the three-parameter log-normal void size distributions of HOD Dense, HOD Sparse, and *N*-body Full are given by Pycke & Russell (2016). As mentioned above, the samples Haloes Sparse and DM Sparse are not taken into account in the further

Table 3
Estimates of the Three-parameter Log-normal Distributions for the Samples Haloes Dense, DM Dense, and DM Ful

Sample	Haloes Dense	DM Dense	DM Full (single population)
$\hat{\theta}$	2.67	−8.61	−5.43
$\hat{\zeta}$	2.65	3.20	2.85
$\hat{\sigma}$	0.441	0.196	0.213

analysis due to their highly fluctuating distributions; see Figure 2. As is seen in Table 3, θ parameters of DM Dense and DM Full samples accept negative values. On the other hand, these negative θ values do not cause any inconsistency in the distributions as long as $R > \theta$. Therefore $\log(\boldsymbol{R} - \theta)$ is always defined for the samples even with negative θ values.

It seems that the three-parameter log-normal distribution is a natural candidate to fit the size distributions of the void samples. Pycke & Russell (2016) have already shown that HOD Dense, HOD Sparse, and *N*-body Mock also satisfy this three-parameter log-normal void size distribution. The goodness-of-fit of our model is illustrated in Figure 6, which displays the sample histograms with the curves of the log-normal densities whose parameters are the estimates computed from the samples.

4. CONCLUSIONS AND DISCUSSION

Here, extending our previous study (Pycke & Russell 2016) to attempt to find a universal void size distribution, we investigate the statistical properties of the void size distribution such as the shape parameters and their relations to the void environment of CVC by using the moment method following Johnson et al. (1994). As mentioned above, the moment method is easy to apply. Therefore, we confirm our previous result on the size distributions of voids, which states that the three-parameter log-normal distribution gives a satisfactory model of the size distribution of voids, which is obtained from simulation and mock catalogs of CVC; *N*-body Mock, DM Full, DM Dense, DM Sparse, Haloes Dense, Haloes Sparse, HOD Sparse, and HOD Dense (see Figure 6, also Figure 3 in Pycke & Russell 2016).

On the other hand, we should keep in mind that all the data sets of CVC are generated by a single *N*-body simulation that operates by counting scales as $N \log N$. Therefore the nature of these data sets may force us to obtain such a unique void size distribution. At this point it is essential to be critical before stating that there is a universal void size distribution satisfying the three-parameter log-normal. As a result, a thorough investigation of the void size distribution by using other catalogs of voids to unveil the truth beyond the relation between the shape of the void size distribution and the void environment is of great importance. In particular, one should take into account that Nadathur & Hotchkiss (2014) have pointed out some problems and inconsistencies in CVC, such as the identification of some overdense regions as voids in the Galaxy data of the SDSS DR7 (Abazajian 2009). Proceding from the problems of CVC, Nadathur & Hotchkiss (2014) provide an alternative public catalog of voids, obtained by using an improved version of the same watershed transform algorithm. Therefore, it is essential to extend our analysis of void size distributions to the catalog given by Nadathur & Hotchkiss (2014). Again, this is particularly important in order to confirm whether the three-parameter log-normal void size distribution is valid in a different void catalog. If the three-parameter log-normal distribution fits another simulated/mock void catalog, then this may indicate that voids have universal (redshift-independent) size distributions given by the log-normal probability function.

Apart from this, Hamaus et al. (2014) show that the average density profile of voids can be represented by an empirical function in ΛCDM *N*-body simulations by using ZOBOV. This function is universal across void size and redshift. Following this, Nadathur et al. (2015) investigate the density profiles of voids that are identified by again using the ZOBOV in mock catalogs of luminous red galaxies (LRGs) from the Jubilee simulation, and in void catalogs constructed from the SDSS LRG and Main Galaxy samples. As a result, Nadathur et al. (2015) show that the scaled density profiles of real voids show a universal behavior over a wide range of galaxy luminosities, number densities, and redshifts. Proceding from these results, there is a possibility that the three-parameter log-normal void size distribution may be a universal distribution for voids in simulated as well as real data samples. That is why it is critical to extend our analysis to other simulated as well as real data sets.

We also observe that the numbers of nonzero and zero void central densities in the samples have important effects on the shape of the three-parameter log-normal void size distributions. As is seen in Table 1 and Figure 1, if the percentage of voids with nonzero central densities reaches 3.84% in a simulated or mock sample in CVC, then a second population emerges in the void size distribution. This second population presents itself as a second peak in the log-normal size distribution, at larger radius.

Also, we obtain here a linear relation between the maximum tree depth and the skewness of the samples, and this relation is given by Equation (5) (see Figure 3). This linear relation indicates that if there is a void in a simulated/mock sample with a large maximum tree depth, then we expect a more skewed log-normal distribution. Therefore, there is a direct correlation between the void substructure and the skewness of the void size distribution. The possibility of this relation is mentioned by Pycke & Russell (2016). Therefore, we confirm here that the skewness of a void size distribution is a good indicator of void substructures in a simulated/mock sample. As mentioned above, the minimum radius-cut of CVC samples defined by two density-based criteria given by Sutter et al. (2014a) can affect the relation between skewness and MTD. The minimum radius-cut of CVC is particularly important because it may affect not only the resulting skewness of the data sets but also other shape parameters of the void size distributions, which can violate the confirmation of the log-normal distribution of the samples. That is why it is important to study raw samples to understand the effect of the minimum radius-cut.

In addition to the linear relation between skewness and MTD, another linear correlation is obtained between the maximum tree depth and the variance of the sparse samples of CVC (see Equation (8)). As is seen from Figure 4, sparse samples with large maximum tree depth tend to be more dispersed than a sparse sample with lower maximum tree depth. On the other hand, although we obtain a linear relation between the maximum tree depth and the variance of the dense samples (see Equation (9)), this relation does not provide enough information to define a relation between these

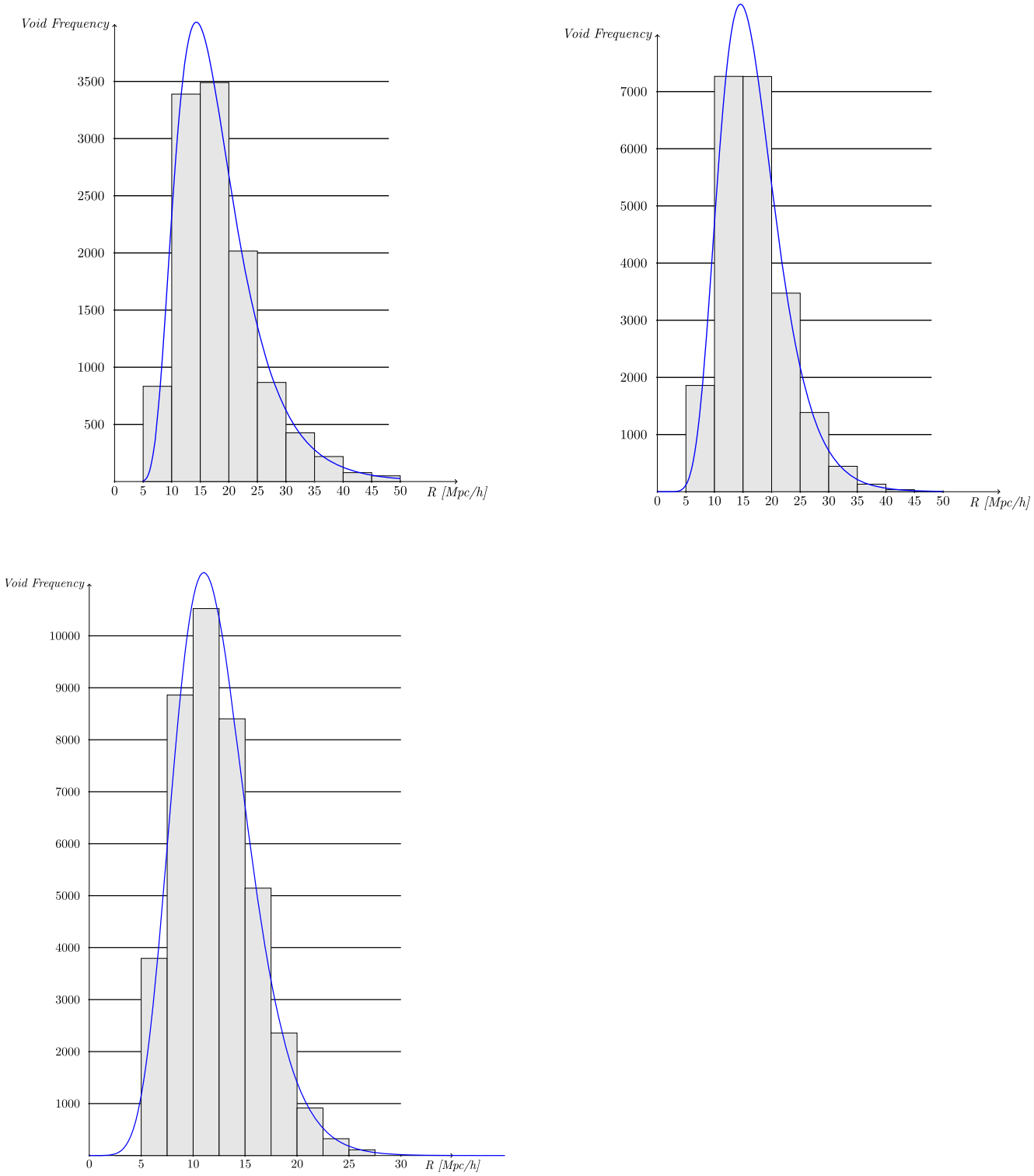

Figure 6. The three-parameter log-normal void size distributions of Haloes Dense (upper left), DM Dense (upper right), and DM Full (lower left).

parameters due to the lack of dense samples with maximum tree depth (see Figure 4, red dotted line). But it is obvious that the relation between maximum tree depth and variance shows two different behaviors for the sparse and the dense samples. This is an expected result since variance is the indicator of dispersion by definition. While sparse samples are highly dispersed with high variance values, the dense samples are expected to show lower variance/dispersion. These relations indicate that there is a direct correlation between the shape parameters of the void size distribution such as skewness, variance, and the void substructures. Our next goal is to address the following questions: Is it possible to relate the shape parameters of the void size distribution to the environment in real data samples? Do the shape parameters change in time,

indicating the dynamical evolution of the void size distribution? Is the three-parameter log-normal void size distribution universal?

The authors would like to thank Paul Sutter and his team for constructing and sharing the Cosmic Void Catalog. All void catalogs used here can be found in the folder void_catalog_2014.06.08 at http://www.cosmicvoids.net.

REFERENCES

Abazajian, K. N. 2009, ApJS, 182, 543
Bernardeau, F. 1992, ApJ, 392, 1
Bernardeau, F. 1994, A&A, 291, 697
Bouchet, F. R., Strauss, M. A., Davis, M., et al. 1993, ApJ, 417, 36
Chincarini, G., & Rood, H. J. 1975, Natur, 257, 294
Coles, P., & Jones, B. 1991, MNRAS, 248, 1
Conover, W. J. 1980a, Practical Nonparametric Statistics (2nd ed.; New York: Wiley)
Conover, W. J. 1980b, Practical Nonparametric Statistics (3rd ed.; New York: Wiley)
Croton, D. J., Farrar, G. R., Norberg, P., et al. 2005, MNRAS, 356, 1155
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
Einasto, J., Joeveer, M., & Saar, E. 1980, Natur, 283, 47
Elizalde, E., & Gaztanaga, E. 1992, MNRAS, 254, 247
Fry, J. N. 1986, ApJ, 306, 358
Goldberg, D. M., & Vogeley, M. S. 2004, ApJ, 605, 1
Gregory, S. A., & Thompson, L. A. 1978, ApJ, 222, 784
Hamaus, N., Sutter, P. M., & Wandelt, B. D. 2014, PhRvD, 112, 25
Hamilton, A. J. S. 1985, ApJL, 292, L35
Hoyle, F., Rojas, R. R., Vogeley, M. S., et al. 2005, ApJ, 620, 618
Johnson, N. L., Kotz, S., & Balakrishnan, N. 1994, Continuous Univariate Distributions, Vol. 1 (2nd ed.; New York: Wiley)
Kayo, I., Taruya, A., & Suto, Y. 2001, ApJ, 561, 22
Kendall, M., & Stuart, A. 1977, The Advanced Theory of Statistics. Vol 1: Distribution Theory (2nd ed.; New York: Macmillan)
Kofman, L., Bertschinger, E., Gelb, J. M., Nusser, A., & Dekel, A. 1994, ApJ, 420, 44
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Lavaux, G., & Wandelt, B. D. 2012, ApJ, 754, 109
Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428 1036
Nadathur, S. 2015, MNRAS, 449, 3997
Nadathur, S., & Hotchkiss, S. 2014, MNRAS, 440, 1248
Neyrinck, M. C. 2008, MNRAS, 386, 2101
Planck Collaboration 2014, A&A, 571, A19
Pycke, J.-R., & Russell, E. 2016, ApJ, 821, 110
Russell, E. 2013, MNRAS, 436, 3525
Russell, E. 2014, MNRAS, 438, 1630
Sheskin, D. J. 2011, Handbook of Parametric and Nonparametric Statistical Procedures (Boca Raton, FL: CRC Press)
Sheth, R. K., & van de Weygaert, R. 2004, MNRAS, 350, 517
Strauss, M. A. 2002, AJ, 124, 1810
Sutter, P. M., Lavaux, G., Hamaus, N., et al. 2014, MNRAS, 442, 462
Sutter, P. M., Lavaux, G., Wandelt, B. D., & Weinberg, D. H. 2012, ApJ, 761, 44
Sutter, P. M., Lavaux, G., Wandelt, B. D., et al. 2014, MNRAS, 442, 3127
Taylor, A. N., & Watts, P. I. R. 2000, MNRAS, 314, 92
Tinker, J. L., Weinberg, D. H., & Zheng, Z. 2006, MNRAS, 368, 85
van de Weygaert, R., & Platen, E. 2011, IJMPS, 1, 41
White, S. D. M. 1979, MNRAS, 186, 145
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760