Cosmic voids uncovered – first-order statistics of depressions in the biased density field

T. Ronconi,[1,2]★ S. Contarini,[3,4] F. Marulli,[3,4,5] M. Baldi [ID][3,4,5] and L. Moscardini[3,4,5]

[1]*SISSA - International School for Advanced Studies, via Bonomea 265, I-34136 Trieste, Italy*
[2]*INFN - Sezione di Trieste, via Valerio 2, I-34127 Trieste, Italy*
[3]*Dipartimento di Fisica e Astronomia - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, I-40129 Bologna, Italy*
[4]*INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, I-40129 Bologna, Italy*
[5]*INFN - Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy*

ABSTRACT

Cosmic voids occupy most of the volume of the Universe, and their statistical properties can be exploited for constraining dark energy, as well as for testing theories of gravity. Nevertheless, in spite of their growing popularity as cosmological probes, a gap of knowledge between cosmic void observations and theory still persists. In particular, the void size function models proposed in literature have been proven unsuccessful in reproducing the results obtained from cosmological simulations in which cosmic voids are detected from biased tracers of the density field. The goal of this work is to cover this gap. In particular, we make use of the findings of a previous work in which we have improved the void selection procedure, presenting an algorithm that redefines the void ridges and, consequently, their radii. By applying this algorithm, we validate the volume conserving model of the void size function on a set of unbiased simulated density field tracers. We highlight the difference in the internal structure between voids selected in this way and those identified by the popular VIDE void finder. We also extend the validation of the model to the case of biased tracers. We find that a relation exists between the tracer used to sample the underlying dark matter density field and its unbiased counterpart. Moreover, we demonstrate that, as long as this relation is accounted for, the size function is a viable approach for studying cosmology with cosmic voids.

Key words: methods: numerical – methods: statistical – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

Cosmic voids are large underdense regions where matter is evacuated and then squeezed in between their boundaries. While galaxy clusters encapsulate most of the mass in the Universe, voids dominate it in terms of volume. Voids are only mildly non-linear objects since, during their evolution, they encounter a crucial physical limit: matter density cannot be less than zero, that is the minimum possible density contrast is $\delta = -1$. Conversely to what happens to overdensities, the void evolution tends to reduce the possible non-sphericity of the initial density perturbations (Icke 1984). These properties suggest that the employment of a simplified spherical expansion model may be more accurate in describing the void formation and evolution than it is in the description of overdensities (Blumenthal et al. 1992).

In the last decade, cosmic voids have gathered a large popularity as cosmological tools. With typical sizes over tens of megaparsecs, they are by far the largest observable structures in the Universe. For this reason they are particularly suited to provide information about several hot topics. They are indeed extreme objects, meaning that they generate from the longest wavelengths of the matter perturbation power spectrum. Moreover, their underdense nature makes them suitable for a wide range of dark energy measurements. Indeed, the super-Hubble velocity field within voids causes a suppression in structure growth, making them a pristine environment in which to study structure formation (Benson et al. 2003; Van de Weygaert & Platen 2011), and a test bench for constraining dark matter (DM), dark energy, and theories of gravity (Lavaux & Wandelt 2012; Li, Zhao & Koyama 2012; Sutter et al. 2014; Hamaus et al. 2015, 2016; Massara et al. 2015; Pollina et al. 2016; Kreisch et al. 2018).

The excursion-set model has been applied to the study of underdensities, in particular to predict the size function of cosmic voids detected in the DM field (Sheth & van de Weygaert 2004, hereafter SvdW). It is based on the assumption that cosmic voids

are nearly spherical regions which have gone through shell-crossing. The latter represents the time when the inner, more underdense, matter shells inside voids have reached the outer, as a result of the differential outwards acceleration they experience. Jennings, Li & Hu (2013) then proposed a modification of this model that makes it in better agreement with the actual size distribution of cosmic voids measured in both simulated and real void catalogues.

The first goal of our work is to further validate the Jennings et al. (2013) size function model on simulated void catalogues detected from unbiased density field tracers, at different redshifts and numerical resolutions. Then, we extend the model to account for the more realistic case of biased tracers, covering the existing gap between the theoretical and observational sides of the problem.

Our work is organized as follows. In Section 2, the existing void size function theoretical models are presented. We focus in particular on their realm of applicability, underlining the existing issues in modelling void statistics in biased matter distributions. We then propose a modification of the Jennings et al. (2013) void size function model to account for the tracer bias. Section 3 is dedicated to the validation of our fiducial model. We investigate its reliability using void catalogues extracted from a set of cosmological *N*-body simulations with different selections. Specifically, while in Section 3.1 we make use of DM particle simulations, meaning that voids are detected from unbiased tracers of the density field, in Section 3.2 we extend our study to voids identified in the distribution of biased DM haloes. Finally, in Section 4 we draw our conclusions.

2 METHODOLOGY

Whether it is possible to exploit cosmic voids as cosmological probes is a matter of how reliable is our capability to model their properties. The most straightforward measure we can compute from whatever survey of extragalactic sources is the abundance of sources as a function of a particular feature. In the case of cosmic voids, this is the size function.

In Section 2.1, we first present the existing theoretical models for the size function of cosmic voids extracted from unbiased tracers of the DM density field. Then, we introduce our new parametrization to describe the void size function in the more realistic case of void catalogues extracted from biased tracer distributions. In Section 2.2, we introduce the detailed procedure that we will use later on to identify the cosmic voids in a set of DM-only *N*-body simulations, and to clean the catalogue.

The simulations employed in this work have been performed with the C-GADGET module (Baldi et al. 2010), a modified version of the (non-public) GADGET3 *N*-body code (Springel 2005), while the software exploited for the post-processing and the data analysis[1] has been presented in Ronconi & Marulli (2017, RM17 hereafter), where we also provided a first description of our new theoretical size function model and cleaning method.

2.1 Size function

The development of any theoretical model of void size function relies on the definition of cosmic voids. Let us stick with the

common assumption that an isolated void evolves from an initial spherical top-hat depression in the density field. This is analogous to the assumption that an isolated DM halo results from the spherical collapse of a top-hat peak in the density field. The evolution of both overdensities (peaks) and underdensities (depressions) can be described via the classical spherical evolution model (Blumenthal et al. 1992). In the overdensity case, a halo is said to have formed at the moment its density contrast reaches a level corresponding to either the virialization of the spherical perturbation (the critical linear overdensity $\delta_c \approx 1.69$, for an EdS universe, see e.g. Bond et al. 1991; Paranjape, Lam & Sheth 2012) or, with a milder assumption, when the perturbation reaches the turnaround ($\delta_t \approx 1.06$, for an EdS universe[2]), and detaches from the overall expansion of the Universe.

On the other hand, cosmic voids do not ever detach from the overall expansion, but instead they expand with a super-*Hubble* flow. The expansion rate is inversely proportional to the embedded density, therefore shells centred around the same point are expected to expand faster as closer as they are to the centre. This eventually leads to the inner shells reaching the outer, the so-called shell-crossing, which is typically assumed to be the moment when a void can be said to have formed. In linear theory, the shell-crossing occurs at a fixed level of density contrast ($\delta_v^L \approx -2.71$, for an EdS universe[3]). This threshold can be used in an excursion-set framework to predict the typical distribution of voids, analogously to the case of halo formation by spherical collapse (Bond et al. 1991; Zentner 2007). With these assumptions, SvdW developed a model of the void size function that, furthermore, considers the void-in-cloud side effect, which accounts for the squeezing of voids that happen to evolve within larger scale overdensities.

In this framework, the probability distribution of voids of a certain size is given by

$$f_{\ln\sigma}(\sigma) = 2\sum_{j=1}^{\infty} j\pi x^2 \sin(j\pi\Delta)\exp\left[-\frac{(j\pi x)^2}{2}\right], \quad (1)$$

where

$$x = \frac{\Delta}{|\delta_v^L|}\sigma \quad (2)$$

and

$$\Delta = \frac{|\delta_v^L|}{\delta_c + |\delta_v^L|}. \quad (3)$$

The size-dependence of equation (1) is given by the square root of the variance,

$$\sigma^2(r) = \frac{1}{2\pi}\int k^2 P(k)W^2(k,r)\,\mathrm{d}k, \quad (4)$$

where $P(k)$ is the matter power spectrum, $W(k,r)$ the window function, and r is the smoothing scale, that is the radius of the spherical underdense region we define as void. In this work, we apply the approximation proposed by Jennings et al. (2013), which

---

is accurate at the 0.2 per cent level:

$$f_{\ln\sigma}(\sigma) = \begin{cases} \sqrt{\dfrac{2}{\pi}} \dfrac{|\delta_v^{\mathrm{L}}|}{\sigma} \exp\left(-\dfrac{\delta_v^{\mathrm{L}2}}{2\sigma^2}\right) & x \leq 0.276 \\[2ex] 2\displaystyle\sum_{j=1}^{4} j\pi x^2 \sin(j\pi\Delta)\exp\left[-\dfrac{(j\pi x)^2}{2}\right] & x > 0.276. \end{cases}$$

(5)

With the kernel probability distribution of equation (1), it is straightforward to obtain the size-abundance distribution of cosmic voids by applying:

$$\left.\frac{\mathrm{d}n}{\mathrm{d}\ln r}\right|_{\mathrm{linear}} = \frac{f_{\ln\sigma}(\sigma)}{V(r)}\frac{\mathrm{d}\ln\sigma^{-1}}{\mathrm{d}\ln r}.$$

(6)

In order to derive the void size function in the non-linear regime, a conservation criterion has to be applied. The SvdW size function relies on the assumption that the total number of voids is conserved when going from linear to non-linear regimes.[4] While reaching shell-crossing, underdensities are expected to have expanded by a factor $a \propto (1+\delta_v^{\mathrm{NL}})^{-1/3}$, thus a correction in radius by this factor is required:

$$\left.\frac{\mathrm{d}n}{\mathrm{d}\ln r}\right|_{\mathrm{SvdW}} = \left.\frac{\mathrm{d}n}{\mathrm{d}\ln(a\,r)}\right|_{\mathrm{linear}}.$$

(7)

Jennings et al. (2013) argued that such a prescription is unphysical, since this leads to a volume fraction occupied by voids which is larger than the total volume of the Universe. They thus introduced a volume conserving model (hereafter Vdn model) in which it is instead assumed that the total volume occupied by cosmic voids is conserved in the transition to non-linearity. None the less, when shell-crossing is reached, voids are thought to recover the overall expansion rate, and continue growing with the Hubble flow (Blumenthal et al. 1992; Sheth & van de Weygaert 2004). The conservation of volume is achieved by applying:

$$\left.\frac{\mathrm{d}n}{\mathrm{d}\ln r}\right|_{\mathrm{Vdn}} = \left.\frac{\mathrm{d}n}{\mathrm{d}\ln r}\right|_{\mathrm{linear}} \frac{V(r^{\mathrm{L}})}{V(r)}\frac{\mathrm{d}\ln r^{\mathrm{L}}}{\mathrm{d}\ln r},$$

(8)

where r^{L} indicates the radius predicted by linear theory (i.e. not accounting for the conversion factor a).

Several authors have tested the SvdW model on both simulated DM density fields and in mock halo catalogues, finding out that it systematically underpredicts the void comoving number density (see e.g. Colberg et al. 2005; Sutter et al. 2012; Nadathur & Hotchkiss 2015; Pisani et al. 2015). To overcome this mismatch, the underdensity threshold δ_v^{L} is commonly left as a free parameter, tuned on simulated halo catalogues. This severely affects the possibility of using the void size function as a cosmological probe. Jennings et al. (2013) have shown that the Vdn model does not require such a fine-tuning, as long as the void catalogue is properly cleaned from spurious voids. However, their results are limited to the case of cosmic voids detected from simulated DM distributions. We extend their study to the case of biased samples, such as mock DM halo catalogues, which are more representative of the real case.

Several authors have underlined that the tracer bias has to be taken into account in order to extract unbiased cosmological information from the number counts of cosmic voids detected in galaxy redshift surveys (see e.g. Pollina et al. 2018). In fact, the comoving number

density and sizes of voids traced by the DM distribution are different from the ones of the voids traced by a biased DM halo density field.

In RM17, we introduced a simple modification of the void size function model that accounts for this effect. It follows from the results by Pollina et al. (2017), that showed that the DM density field within voids, $\delta_{v,\mathrm{DM}}^{\mathrm{NL}}$, is linearly related to the density field traced by some biased tracers, $\delta_{v,\mathrm{tr}}^{\mathrm{NL}}$:

$$\delta_{v,\mathrm{tr}}^{\mathrm{NL}} = b\,\delta_{v,\mathrm{DM}}^{\mathrm{NL}}.$$

(9)

Therefore, the threshold density to be used in the size function model of cosmic voids detected from biased tracers has to be corrected taking into account equation (9). Moreover, to recover the linear extrapolated value $\delta_{v,\mathrm{tr}}^{\mathrm{L}}$, we use the fitting formula provided by Bernardeau (1994), namely

$$\delta^{\mathrm{NL}} = (1 - \delta^{\mathrm{L}}/\mathcal{C})^{-\mathcal{C}},$$

(10)

with $\mathcal{C} = 1.594$. equation (10) provides a good fit, with errors below 0.2 per cent, for standard cosmological models with any values of Ω_{M} and Ω_{Λ}, especially in the case of underdense regions ($\delta^{\mathrm{L}} < 0$), and it is exact in the limit $\Omega_{\mathrm{M}} \to 0$, for $\Omega_{\Lambda} = 0$. Combining equation (10) with equation (9), it gives

$$\delta_{v,\mathrm{tr}}^{\mathrm{L}} = \mathcal{C}\left[1 - (1 + b\,\delta_{v,\mathrm{DM}}^{\mathrm{NL}})^{-1/\mathcal{C}}\right].$$

(11)

Our recipe for a void size function that works regardless of whether the void catalogue is affected or not by a tracer bias is to use the value of δ_v^{L} found with equation (11) in the probability function given by equation (1).

2.2 Void finding and data reduction

There is not a general concordance in the definition of voids. Indeed, many different techniques have been proposed and applied over the years (see e.g. Colberg et al. 2008; Elyiv et al. 2015). A significant part of our work has been dedicated to cover this gap of knowledge, through the development of a procedure to standardize the outcome of void finders of different types (see also RM17). In this work, we use the public void finder VIDE[5] (Sutter et al. 2015, Void IDentification and Examination Toolkit), which implements an enhanced version of ZOBOV (Neyrinck 2008, ZOnes Bordering On Voidness) to construct voids with a watershed algorithm, whatever the nature of the tracers and independently of the presence of a biasing factor. The VIDE catalogues obtained are then cleaned with the pipeline developed and presented in RM17. This allows us to align the objects included in the void catalogue with the theoretical definition of cosmic voids used to derive the void size function (Section 2.1).

We summarize here the main features of the cleaning pipeline, referring to RM17 for a detailed description. The procedure is divided into three main steps: (i) a preliminary trimming of the catalogue based on density and reliability criteria, (ii) a rescaling of the effective radius of all the voids in the catalogue and, finally, (iii) the catalogue is trimmed again to erase overlapping between voids, selecting the voids with the smaller central density or, equivalently, with the larger density contrast (see e.g. Neyrinck 2008).[6] As

---

[4]This is the same assumption which is implicitly made to derive the halo-mass function.

[5]http://www.cosmicvoids.net

[6]The overlap-check acts on the rescaled voids by comparing either the central density or the internal density ratio (i.e. the ratio between the overall density contrast measured inside the void and its central density contrast) of two overlapping voids, discarding one of the two. Since in step (ii) of the algorithm we impose the overall density embedded by the voids to be equal

Table 1. Fiducial cosmological parameters of the *N*-body simulations used in this work.

h_0	Ω_{CDM}	Ω_b	Ω_Λ	$\mathcal{A}_s$	n_s
0.7	0.226	0.0451	0.7289	$2.194 \cdot 10^{-9}$	0.96

underlined also in RM17, this cleaning method is necessary to match the distribution of voids measured in simulated catalogues with the predictions of the Vdn model. The rescaling part of the procedure [step (ii)] is based on the requirement that cosmic voids have to embed a density matching the shell-crossing prescription defined by the model. This is crucial in order to have a reliable description of their distribution in terms of spherical expansion. Specifically, spheres are grown around the centre of voids selected by the adopted void finder algorithm up to the scale inside which an a priori chosen value of underdensity is reached. The latter value is typically chosen as the non-linear counterpart of the shell-crossing threshold, namely $\delta_v^{NL} \approx -0.795$ at redshift $z = 0$ for an EdS universe.

Even though it is reasonable to select this particular value as the critical threshold in both the model and cleaning method, there are no restrictions in this sense and whatever value is acceptable, as long as it is used in both measuring (that is when applying the cleaning algorithm) and modelling the distribution (that is when building the modified model described in Section 2.1). This is a key issue: the choice of the underdensity threshold, that has to be embedded by the voids identified in whatever tracer distribution, must match the threshold δ_v^L in the model (equation 1) used to predict their distribution. Given that when measuring the size distribution of cosmic voids we are dealing with a non-linear universe, while the theoretical model is derived from linear theory, a prescription to convert the threshold back and forth (such as the fitting relation from Bernardeau 1994) is necessary.

3 RESULTS

We test the procedure outlined in Section 2 with both a set of DM-only *N*-body simulation snapshots, and a set of mock halo catalogues extracted by means of a friends-of-friends (FoF) algorithm and applying different mass selection cuts. Specifically, we consider a suite of *N*-body simulations of the standard ΛCDM cosmology performed with the C-GADGET module (Baldi et al. 2010) for different box sizes and particle numbers. Our largest box corresponds to the ΛCDM run of the L-CoDECS simulations (Baldi 2012), with a volume of $(1\,\mathrm{Gpc}\,h^{-1})^3$ and a total number of 1024^3 particles, with a mass resolution of $6 \times 10^{10}\,\mathrm{M}_\odot\,h^{-1}$. The smaller simulations have been run with the same code specifically

to a fixed value, these two measures are strictly related for our rescaled voids: voids with smaller central density have a larger internal density ratio and vice versa. In the original algorithm presented in RM17, choosing whether to discard the void with the larger central density (smaller internal density ratio) or the one with the smaller central density (larger internal density ratio) was left to the user. Selecting among these two choices reduces to deciding whether to keep the larger voids (which typically have a smaller central density, thus, a larger internal density ratio) or the smaller ones. Given that, as also stated in the original ZOBOV paper (Neyrinck 2008), voids with a larger internal density ratio are statistically more significant, we decided to remove this free parameter in the newer version of our algorithm which now automatically applies the criterion we dubbed larger favoured in the previous versions.

Table 2. Our set of cosmological simulations with the corresponding relevant physical quantities: L_{box} is the simulation box-side length in units of (Mpc h^{-1}); N_p is the total number of particles; m.i.s. is the mean interparticle separation in the box in units of (Mpc h^{-1}); in the last column, the particle mass in units of ($10^{10}\,\mathrm{M}_\odot\,h^{-1}$) is reported.

L_{box}	N_p	m.i.s.	Particle mass
(Mpc h^{-1})		(Mpc h^{-1})	($10^{10}\,\mathrm{M}_\odot\,h^{-1}$)
1000	1024^3	≈ 1.00	5.84
500	256^3	≈ 2.00	56.06
256	256^3	1.00	7.52
128	256^3	0.50	0.94
64	256^3	0.25	0.12

for this work, and share with the largest box the same fiducial cosmological parameters (consistent with WMAP7 constraints, see Table 1; Komatsu et al. 2011), spanning a range of lower, equal, and higher particle mass and spatial resolutions, with the aim of testing the dependence of our analysis on these numerical parameters. The main properties of the different simulations are reported in Table 2.

As a first step, in Section 3.1 we validate our procedure on the four *N*-body simulations with 256^3 particles listed in Table 2. These simulations vary in mass resolution and mean interparticle separation (m.i.s.). We gathered snapshots at four different redshifts ($z = 0, 0.5, 1, 1.5$) for each of these four simulations.

Afterwards, in Section 3.2 we extend our analysis to the case of biased tracers. The set of halo catalogues is extracted from a snapshot at redshift $z = 0$ of the simulation with box-side length $L_{box} = 1000\,\mathrm{Mpc}\,h^{-1}$. We could not use the same simulations used in the validation part of the work because of the extremely low number of voids identified by VIDE in those small boxes.

3.1 Unbiased tracers

In order to validate our procedure, we first run the VIDE void finder on top of each of the 16 simulation snapshots described before with varying resolution and redshift. The resulting void catalogues are then cleaned with our pipeline. We evolve the shell-crossing threshold, $\delta_v^L = -2.71$, as

$$\delta_v^L(z) = \delta_v^L \times \frac{\mathcal{D}(z)}{\mathcal{D}(0)}, \tag{12}$$

by means of the growth factor $\mathcal{D}(z)$ at redshift z. Applying the fitting formula given in equation (10), we obtain its non-linear extrapolation, $\delta_v^{NL}(z)$, and use this value as a threshold in the cleaning pipeline. Doing so, all the voids in our catalogues embed a fixed density

$$\rho_v = \langle \rho \rangle \left[1 + \delta_v^{NL}(z) \right], \tag{13}$$

where $\langle \rho \rangle$ is the average matter density of the snapshot.

First, we investigate the difference between the void density profiles before and after the application of our cleaning algorithm. It is generally believed that, in the standard ΛCDM cosmology, voids have a self-similar internal structure, which does not depend on the tracers used to define them (see e.g. Ricciardelli, Quilis & Planelles 2013; Hamaus, Sutter & Wandelt 2014; Ricciardelli, Quilis & Varela 2014). The self-similarity assures that the internal structure of voids can be characterized by a single parameter, the void effective radius, r_{eff}, which is the radius of a sphere embedding the same volume embedded by a given cosmic void. Despite the wide range of values this parameter covers, it is possible to average the internal

density distribution to recover a common behaviour for voids with approximately the same r_{eff}.

It has been noticed that, to recover self-similarity, some criteria have to be applied to the selection of objects to be classified as voids (Nadathur & Hotchkiss [2015]). Nevertheless, the most crucial role seems to be played by the cut on the minimum density of the candidates, ρ_{min}, which has to be sufficiently low (that is $\rho_{\text{min}} \lesssim 0.3 \langle \rho \rangle$). All voids undergoing our cleaning procedure with our chosen value of δ_v^{NL} satisfy this requirement by construction. We are therefore confident that the objects we select should behave in a self-similar manner.

We select all the cosmic voids in the range of radii $1\,\text{Mpc}\,h^{-1} < r_{\text{eff}} < 1.5\,\text{Mpc}\,h^{-1}$, from the catalogue with box side length $L_{\text{box}} = 128\,\text{Mpc}\,h^{-1}$, at four different redshifts. This is a compromise between having a large box (thus having more objects) and having a high resolution. None the less, the results we are going to present are similar for the other simulations used in this section. The radius window has been chosen to have a sufficiently high number of objects with radii far enough from the lower limit imposed by the spatial resolution of the box that, for this catalogue, is $0.5\,\text{Mpc}\,h^{-1}$ (see Table [2]).

In Fig. [1], we show the stacked void profiles before and after cleaning the void catalogues with our algorithm. The stacked voids are all centred in the VIDE centres. We measure the cumulative number density of particles as a function of the distance from the centre, given in units of r_{eff}, for each of the selected voids, particle-per-particle from $0.3\,r_{\text{eff}}$ to $3.5\,r_{\text{eff}}$. We divide this range of radii in 30 equally spaced bins and average the profiles. This procedure is repeated before and after applying the cleaning algorithm.

From Fig. [1] it is possible to appreciate how differently we define the ridges of a void: the orange dashed profiles show the profiles of voids identified with VIDE, while the blue solid ones are for the same voids after cleaning. The vertical dotted lines mark the void radius in both cases. We also show the background density level (solid horizontal black line) and the underdensity threshold of equation (12) computed for the different redshifts (dashed horizontal black line) as a reference. It can be noticed that, while VIDE void limits reside close to the compensation wall and embed a density that approaches the background, our voids are, by construction, defined in much deeper regions of the density field.

These observations have to be kept in mind to understand how the algorithm we apply to clean the cosmic void catalogues affects the void structure. Thanks to this modification, we can be confident that the voids we are using to measure the size function are as close as possible to the objects for which we are able to construct the theoretical model.

Fig. [2] compares the measured and modelled void size functions, defined as the comoving number density of voids in logarithmic bins of effective radii, divided by the logarithmic extent of the bins. The four panels show the results at different redshifts. The different symbols in the upper part of each panel show the void size distributions measured in the simulation snapshots at different resolutions, while the gray shaded regions represent the model predictions. The model depends on two thresholds, the one of the shell-crossing, which is fixed at the value $\delta_v^{\text{L}} = -2.71$, corresponding to our choice of threshold in the void catalogue cleaning, and the overdensity one, δ_c, used to account for the void-in-cloud effect. There is a persisting uncertainty on the definition of the latter, since it is not clear if this effect becomes relevant at turn-around (corresponding to a density contrast of $\delta_c \approx 1.06$) or when the overdensity collapses (at a density contrast $\delta_c \approx 1.69$). Since in realistic occurrences we do not expect the spatial resolution of tracers to be high enough to inspect the

radii range in which the void size function becomes sensitive to the void-in-cloud effect, we leave the overdensity threshold free to vary in the range $1.06 \leq \delta_c \leq 1.69$, thus the shaded region of Fig. [2].
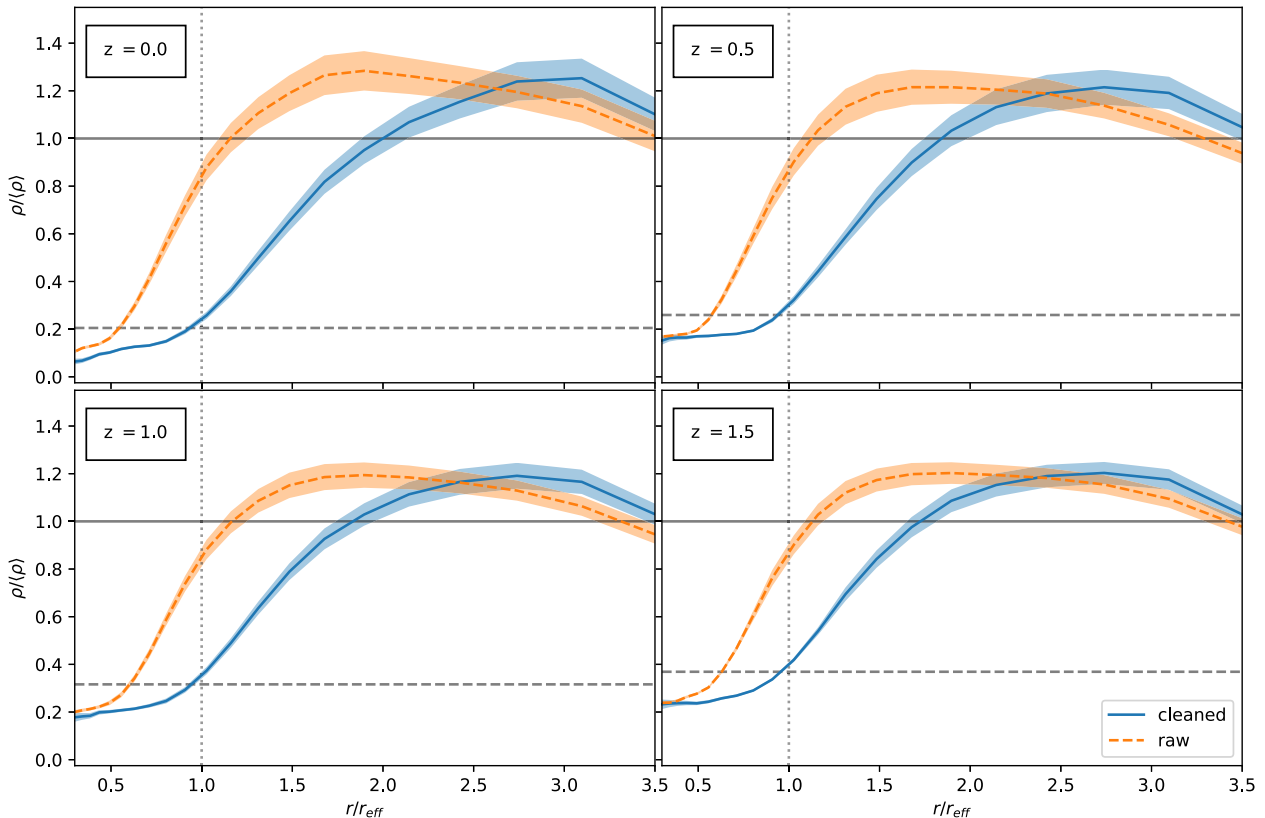
We cut the distribution measured in each of the simulations at twice the m.i.s. of the given simulation box (see Table [2]). To highlight the difference between the theoretical model predictions and the measured distributions, we show in the lower part of each panel the logarithmic difference between the simulation data and the model. Since voids are wide underdense regions, their measured properties and number density can be significantly affected by the spatial resolution of the sample. At scales comparable to the spatial resolution of the simulation, voids are not well represented by the mass tracers of the underlying density field. This causes a loss of power in the number counts, which can be noticed at the smallest scales in each panel of Fig. [2], in spite of the applied cut. At large scales, instead, the measurements are limited by the simulation box extension. Since our cleaning algorithm does not consider the periodic boundary conditions of the N-body simulations yet, the largest peripheral objects of the box may exceed the box boundaries. As a consequence, we cannot trust voids close to the boundaries of the simulation box, and therefore we choose to reject them from our analysis. Contrary to the limits at small radii, which cannot be overcome without increasing the resolution of the sample, this inaccuracy in the void counts at large radii could be faced with an upgrade of the void finding algorithm to account for the periodicity of the box. None the less, since our final goal is to provide a framework to exploit the distributions of voids observed in real catalogues, this issue is not addressed in this work and we simply cut away these objects and reduce the total volume of the box consistently.

Fig. [2] shows that the Vdn model predictions are in overall good agreement with the results from N-body ΛCDM simulations, when cosmic voids are selected from unbiased matter tracers. We notice that the model predictions are closer to the simulation measurements at higher redshifts, in the full range of investigated radii. This might be due to the larger value of the underdensity threshold used for cleaning the catalogue: cosmic voids are less evolved at more distant cosmic epochs and have had less time to evacuate their interiors. A higher threshold means a higher density of tracers, thus, a higher resolution. Concerning the time evolution of the void size function, the density of the larger voids decreases with increasing redshifts. On the other hand, the smaller scales are less affected. The Vdn model well predicts this behaviour.

3.2 Biased tracers

Having established the reliability of the Vdn model in predicting the number counts of cosmic voids detected in a simulated distribution of DM particles, we now extend our analysis to the case of biased tracers of the underlying DM density field. Our ultimate goal is to implement a model capable of predicting the void size distribution for observable tracers of the underlying density field, such as galaxies or clusters of galaxies. In this paper, we start focusing on the simplest case of biased tracers, namely DM haloes.

It has been extensively demonstrated that voids in biased tracers of the underlying DM distribution are systematically larger than those predicted by the void size function models (Colberg et al. [2005]; Sutter et al. [2012]; Nadathur & Hotchkiss [2015]; Pisani et al. [2015]). Specifically, the typical void sizes increase with the minimum mass of the tracers, that is with their effective bias (Furlanetto & Piran [2006]).

Figure 1. Comparison between the density profile within voids obtained with VIDE before (dashed orange line) and after (solid blue line) applying our cleaning algorithm. Each panel shows the result of averaging all the density profiles in a narrow bin of radii (with $1 < r_{\mathrm{eff}}[\,\mathrm{Mpc}\,h^{-1}] < 1.5$) and the 2σ confidence as a shaded region around the mean, at four different redshifts. The two horizontal lines show the position of the shell-crossing threshold (dashed line) and of the background density (solid line). The vertical dotted line highlights the position of the effective radius.

To infer cosmological constraints from the number count statistics of cosmic voids, the development of a reliable model independent of the tracers used to detect them is crucial. In this section, we make use of the set of halo catalogues obtained by means of a FoF algorithm applied to the N-body simulation snapshot at redshift $z = 0$ with volume $1\,(\mathrm{Gpc}\,h^{-1})^3$, whose properties are reported in Table 2.

These catalogues are obtained by cutting the original FoF samples at 3 different minimum masses: $M_{\mathrm{min}} = \{2 \times 10^{12}\ \mathrm{M}_\odot\ h^{-1}, 5 \times 10^{12}\ \mathrm{M}_\odot\ h^{-1}, 10^{13}\ \mathrm{M}_\odot\ h^{-1}\}$ (see Table 3). These catalogues differ for the biasing factor with respect to the underlying DM distribution.

Our hypothesis is that the existing theoretical size function models fail in predicting the abundances of voids in the observed distributions of tracers because they implicitly assume that voids in biased tracers embed the same level of underdensity of the underlying DM distribution. This is far from being true and can be shown by a simple analysis of the density profiles around the void centres.

We obtain the VIDE void catalogues from each of the DM halo catalogues reported in Table 3, and clean them following the same procedure considered in Section 3.1 for the unbiased DM distributions. In particular, we keep the cleaning underdensity threshold parameter unchanged, meaning that we set it to $\delta_v^{\mathrm{NL}} = -0.795$, as we have done in Section 3.1 for the unbiased catalogues at $z = 0$. This choice is dictated by the need of selecting sufficiently underdense regions, to be distinguishable from noise, but at the same time sufficiently dense in tracers not to incur in resolution

issues. None the less, we checked our final conclusions are robust with respect to the adopted threshold (Contarini et al. 2019).

We then divide our catalogue of cosmic voids into a set of logarithmically equally spaced bins within the radii ranges reported in Table 3, and compute the stacked density profile in each bin. Fig. 3 compares the stacked density profiles of the voids selected in DM halo catalogues with three different mass cuts, measured in the halo and DM distributions. The profiles stacked in each bin of radii varies from a minimum of 5 objects to a maximum of 300 objects and depends on the total number of voids found by the algorithm in the considered range. The number decreases going from smaller to larger radii. In order to have a reasonable convergence of the mean profile, we only consider the stacked profiles for which we have at least 30 objects to average. Fig. 3 also shows the background density and the shell-crossing underdensity level used to clean the catalogues which, by definition, crosses the profiles at $r = r_{\mathrm{eff}}$. The void density profiles measured from the distribution of DM haloes are significantly different with respect to those measured in the background matter density field, except at radii larger than about $2\,r_{\mathrm{eff}}$, approximately where the background density level is recovered. Specifically, the density measured within voids in the DM halo distribution is way deeper than that measured in the underlying DM density field.

The clear discrepancy between the halo-void density profiles and the underlying DM density profiles is the reason why we cannot use the Vdn model directly to predict void abundances in biased tracers: theoretical models of the void size function are based on

Figure 2. Upper part of each panel: Void size function prediction (gray shaded region) at four different redshifts (different panels) compared to the distribution of cosmic voids after applying the cleaning method described in Section 2.2 (different markers correspond to different simulations, the values of L_{box} in legend corresponding to those of Table 2). Lower part of each panel: Logarithmic differences between the measured distribution and the Vdn model prediction. For the largest simulation ($L_{box} = 1000$ Mpc h^{-1}) we only have snapshots at $z = 0$ and $z = 1$, thus the corresponding abundances are missing from the $z = 0.5$ and $z = 1.5$ panels.

Table 3. Characteristic quantities for our set of halo catalogues. We report the minimum halo mass M_{min}, the total number of haloes N_H, the corresponding mean numerical density $\langle n \rangle$, the m.i.s., the radii range used for profiling, the coefficients of the linear fit of equation (14) (b_{slope} and c_{offset}) with the value of the residuals (res.), and the ratio between the profile measured in the underlying unbiased DM distribution and in the DM halo distribution [$(\delta_{halo}/\delta_{DM})(r_{eff, h})$]. All the catalogues have been obtained by applying different lower mass cuts (reported in the first column) to the FoF halo catalogue extracted from the simulation with 1024^3 particles (see Table 2).

M_{min}	N_H	$\langle n \rangle$	m.i.s.	Radii range	b_{slope}	c_{offset}	Res.	$(\delta_{halo}/\delta_{DM})(r_{eff, h})$
$(M_\odot\, h^{-1})$	10^5	$(\text{Mpc}\, h^{-1})^{-3}$	$(\text{Mpc}\, h^{-1})$	$(\text{Mpc}\, h^{-1})$				
2×10^{12}	15.8	1.58×10^{-3}	8.58	$18 \leq r_{eff} \leq 40$	1.408	0.022	0.005	1.406 ± 0.001
5×10^{12}	7.6	7.62×10^{-4}	10.95	$22 \leq r_{eff} \leq 45$	1.488	0.013	0.002	1.522 ± 0.001
10^{13}	3.9	3.94×10^{-4}	13.64	$27 \leq r_{eff} \leq 50$	1.657	0.026	0.008	1.668 ± 0.003

the evolution of underdensities in the DM density field and their shape severely depends on the void density contrast. Ideally, we would like to recover the tracer density contrast, which corresponds to the shell-crossing value. However, since tracers are typically sparse, this is not practically viable: such a density contrast would be too low to be traced with enough statistics. What we can do instead is to fix the density threshold in the tracer distribution and recover the corresponding density contrast in the underlying DM distribution. Given the hypothesis that voids in tracers are centred in the same position of their DM counterparts, we search for the density contrast that voids identified in the tracer distribution reach in the underlying DM distribution. By taking the ratio between the two profiles, as suggested by Pollina et al. (2017), one can infer the relation between the density measured from biased tracers and the underlying unbiased DM density.

Fig. 4 shows the ratio between all the stacked profiles obtained for each catalogue, markers with error bars representing the uncertainty on the mean in the bin. Given the high level of uncertainty in the inner regions of the profiles, due to the sparsity of tracers, we exclude all the values with $r \leq 0.5\, r_{eff}$. In agreement with the results by Pollina et al. (2017), we find that the densities measured with the two different tracers, that is DM particles and haloes, are linearly related:

$$\delta_{halo} = b_{slope} \times \delta_{DM} + c_{offset}. \tag{14}$$

The best-fitting values of the two free parameters, b_{slope} and c_{offset}, are reported in Table 3. The relation between the best-fitting values of b_{slope} and c_{offset} to the large-scale effective bias of the tracer samples, b_{eff}, is addressed in Contarini et al. (2019).

The existing theoretical models predict the comoving number density of underdense regions characterized by a given embedded density contrast in the DM density field. Thus, to extend the models to the case of voids identified in the halo density field, we have to assess the density contrast in the underlying DM distribution. This can be obtained by taking the ratio between the density profiles measured in the two different tracers. Estimating it at exactly one effective radius allows us to associate the precise DM density contrast, required by the size function model for voids detected in distributions of biased tracers.

Fig. 5 shows the ratio between the density contrast we have requested in the cleaning algorithm for halo-voids ($\delta_{\rm halo}^{\rm NL} = -0.795$) and the value of $\delta_{\rm DM}^{\rm NL}(r_{\rm eff})$ measured from the three different catalogues. Of all the stacked profiles we have obtained, we use for this measurement only those which counted at least 30 objects in the radial bins to avoid issues with the convergence of the mean profile. The mean value of this ratio is therefore the conversion factor between the underdensity thresholds used in the detection and cleaning of cosmic voids in our halo catalogues and the non-linear counterpart of the DM underdensity threshold required by the void size function theoretical model (the values of the mean and standard deviations are reported in Table 3). In the considered range of effective radii, the mean value of this ratio is consistent with what can be inferred from a linear fit, as also shown in Fig. 5.

Using the Vdn size function model, we can describe the distribution of voids in the DM distribution underlying the distribution of tracers. If a cosmic void with radius $r_{\rm eff}$ embeds a density contrast $\delta_{v,\,\rm DM}^{\rm NL}$ in the DM density field, then the same radius $r_{\rm eff}$ will embed a density contrast $\delta_{v,\,\rm tr}^{\rm NL}$ in the tracer distribution, with $\delta_{v,\,\rm tr}^{\rm NL}$ and $\delta_{v,\,\rm DM}^{\rm NL}$ given by equation (9). Therefore, if $\delta_{v,\,\rm tr}^{\rm NL}$ is the threshold we use when cleaning the original void catalogue (Section 2.2), then the resulting void size distribution has to be modelled by a size function with an underdensity threshold given by:

$$\delta_{v,\,\rm DM}^{\rm L} = \mathcal{C}\left[1 - \left(1 + \delta_{v,\,\rm DM}^{\rm NL}\right)^{-1/\mathcal{C}}\right], \quad (15)$$

where $\delta_{v,\,\rm DM}^{\rm NL}$ is the value shown in Fig. 5.

Fig. 6 shows the measured size functions of the voids detected in the three different DM halo catalogues, compared to the size function model described above. The shaded band in each plot shows the 2σ confidence region around the average value of $\delta_{v,\,\rm DM}^{\rm NL}$, obtained from the analysis of the void density profiles.

We find an excellent agreement between the measured and predicted void size functions in all the considered catalogues (almost all the points are within the 2σ confidence region), except at the smallest radii, where the spatial resolution of our catalogues does not allow us to have enough statistics of voids with $r_{\rm eff} \lesssim 2.5$ m.i.s., thus causing a loss of power. On the other hand, at large radii the size of the simulation box limits the possibility to obtain a reliable counting of large voids. To investigate the impact of the offset parameter, $c_{\rm offset}$, in Fig. 6 we also show the size function model obtained by considering only $b = b_{\rm slope}$ as the correction parameter. This model systematically underestimates the measured void size function, though the mismatch is within 2σ, considering
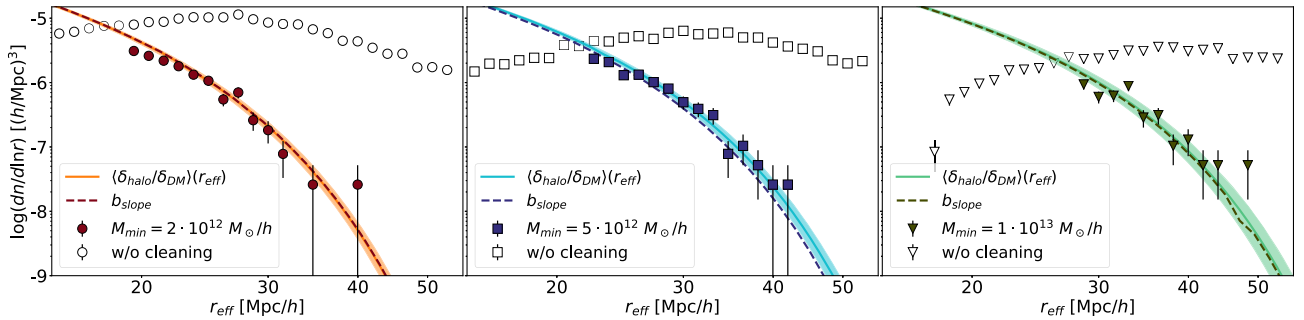

Figure 3. Stacked density profiles in different radii bins, for the three catalogues considered ($M_{\rm min} = 2 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1}, 5 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1}, 10^{13}\,\mathrm{M}_\odot\,h^{-1}$, from left to right). Each panel shows the void stacked density profile. The dashed lines show the profiles measured in the DM halo distribution, while the solid lines represent the profile measured in the underlying DM distribution. The shaded bands around the mean profiles mark the 2σ confidence region. These density profiles have been measured using the cleaned VIDE void catalogues. The set of parameters used for the cleaning is the same one used for the unbiased DM voids in Section 3.1.

Figure 4. Density contrast in the unbiased tracer distribution, $\delta_{\rm DM}$, versus density contrast in the biased tracer distribution, $\delta_{\rm halo}$. Points are obtained gathering the measures of all the stacked profiles; the error bars represent a 1σ uncertainty on the mean values. The solid lines show the linear fit (the fitting parameters $b_{\rm slope}$ and $c_{\rm offset}$ are reported in Table 3), while the narrow shaded regions show the uncertainty on the best-fitting values given in terms of the residuals (also reported in Table 3). Each panel shows the results for one of the halo catalogues of Table 3, from left to right: $M_{\rm min} = 2 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1}, 5 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1}, 10^{13}\,\mathrm{M}_\odot\,h^{-1}$.

Figure 5. The ratio between δ_{halo} and δ_{DM} measured at $r = r_{\mathrm{eff}}$ for voids with different sizes. Each point is obtained by measuring $\delta_{\mathrm{DM}}(r_{\mathrm{eff}})$ in different effective radius bins. The dashed lines are the values of b_{slope} measured from the linear fit of Fig. 4, while the solid lines are the average values, with the corresponding shaded regions delimiting 2σ deviations from the mean. Each panel shows the results for one of the halo catalogues of Table 3, from left to right: $M_{\mathrm{min}} = 2 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1},\ 5 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1},\ 10^{13}\,\mathrm{M}_\odot\,h^{-1}$.

Figure 6. Size distribution of VIDE voids after the application of our cleaning prescriptions, measured in three different halo catalogues (markers). The shaded bands represent the 2σ confidence region around the mean values of the Vdn model modified to account for the different levels of DM underdensity enclosed by halo voids (solid lines). The dashed line shows the modified Vdn model in which, instead, we have used a threshold modified with the value of b_{slope}. Each panel shows the results for one of the halo catalogues of Table 3, from left to right: $M_{\mathrm{min}} = 2 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1},\ 5 \times 10^{12}\,\mathrm{M}_\odot\,h^{-1},\ 10^{13}\,\mathrm{M}_\odot\,h^{-1}$. For comparison, in each plot we show the size-abundance distribution of cosmic voids measured in the VIDE void catalogue before the cleaning is applied (empty markers).

the estimated uncertainties. This suggests that, even though it might not be the case when working with cosmic void density profiles, as stated in Pollina et al. (2017), the value of the c_{offset} coefficient of the relation cannot be ignored when computing the size function of cosmic voids.

The shell-crossing threshold provides a reasonable value to define a cosmic void. However, the size function model is not forced to be constructed using this threshold. The model is in fact potentially capable of predicting the first-order statistics of density fluctuations whatever the threshold. In particular, in this section we choose to use the shell-crossing threshold (δ_v^{NL}) to clean the catalogues, selecting and rescaling the voids to the radius at which they reach this specific density contrast. Then, in the theoretical size function model (equation 1), we substitute the threshold δ_v^{L} with the value that we measured from the stacked profiles (δ_{DM}), converted to its linear extrapolated counterpart by means of equation (10). On the one hand, this is required for being representative of the underdensities embedded by the voids traced by biased distributions. On the other hand, it also demonstrates that the use of the shell-crossing threshold is strictly required. The detection of voids is prone to the nature of the tracer used for sampling the underlying DM distribution. Our work highlights that modelling their statistics cannot be done trivially without accounting for this nature.

The results presented in this section rely on the existence of a linear relation between the density contrast embedded by voids in the tracer distribution and the density contrast of the underlying DM distribution. This relation, given by equation (14), has been

calibrated on DM-only N-body simulations of the standard ΛCDM model (see also Pollina et al. 2017; Contarini et al. 2019). A larger set of simulations in different cosmological frameworks, possibly including also baryonic effects, are required to better calibrate this relation in a more general context. This goes beyond the scope of this paper, and will be addressed in a forthcoming work.

Our overall conclusion is that the number count statistics of cosmic voids is completely determined by the cosmological model and by the relation between the density contrast of void tracers and DM inside the voids. This represents a key step towards the use of the void size function as a cosmological probe.

4 SUMMARY

The main goal of this work was to develop a theoretical model of the size function of cosmic voids detected from the distribution of biased tracers, extending the Vdn model, up to now validated for the DM distribution only.

The main steps of our analysis can be summarized as follows:

(i) We have developed a method to clean cosmic void catalogues, whose numerical implementation was presented in RM17. In particular, we have searched for the largest spherical regions embedding a mean density contrast of $0.2\,\overline{\rho}$, with $\overline{\rho}$ being the average density of the considered sample, ensuring that the identified regions do not overlap. The condition on the mean density ensures that the selected underdensities have actually passed through shell-crossing, while the second condition guarantees the volume conservation.