The nature of voids: I. Watershed void finders and their connection with theoretical models

S. Nadathur& S. Hotchkiss

1Department of Physics, University of Helsinki and Helsinki Institute of Physics, P.O. Box 64, FIN-00014, University of Helsinki, Finland
2Department of Physics and Astronomy, University of Sussex, Falmer, Brighton, BN1 9QH, UK

27 April 2015

ABSTRACT
The statistical study of voids in the matter distribution promises to be an important tool for precision cosmology, but there are known discrepancies between theoretical models of voids and the voids actually found in large simulations or galaxy surveys. The empirical properties of observed voids are also not well understood. In this paper we study voids in an $N$-body simulation, using the ZOBOV watershed algorithm. As in other studies, we use sets of subsampled dark matter particles as tracers to identify voids, but we use the full-resolution simulation output to measure dark matter densities at the identified locations. Voids span a wide range of sizes and densities, but there is a clear trend towards larger voids containing deeper density minima, a trend which is expected for all watershed void finders. We also find that the tracer density at void locations is smaller than the true density, and that this relationship depends on the sampling density of tracers. We show that fitting functions given in the literature fail to match the density profiles of voids either quantitatively or qualitatively. The average enclosed density contrast within watershed voids varies widely with both the size of the void and the minimum density within it, but is always far from the shell-crossing threshold expected from theoretical models. Voids with deeper density minima also show much broader density profiles. We discuss the implications of these results for the excursion set approach to modelling such voids.

Key words: cosmology: observations – large-scale structure of Universe – methods: numerical – methods: data analysis

1 INTRODUCTION
The study of large underdense voids in the large-scale matter distribution of the Universe has become increasingly important in recent years, with the creation of a number of public catalogues of voids in galaxy survey data (Pan et al. 2012; Sutter et al. 2012; Nadathur & Hotchkiss 2014) and a wide variety of statistical analyses based on them.

Voids are interesting primarily because of the cosmological information they may contain. Various studies have suggested that they could be used to constrain the expansion history of the Universe and the equation of state of dark energy (e.g. Ryden 1995; Lee & Park 2009; Bos et al. 2012; Hamaus et al. 2014), to test modified theories of gravity (Li et al. 2012; Clampitt et al. 2013; Cai et al. 2014; Zivick et al. 2014), to calibrate measurements of galaxy bias (Hamaus et al. 2014; Chan et al. 2014), to constrain initial conditions of structure formation (Kamionkowski et al. 2009), or to probe more exotic theories such as coupled dark energy (Sutter et al. 2014). The primary void observables used in such studies are their abundances and size distributions, the distortion of their shapes in redshift space (Alcock & Paczynski 1979), their dark matter density profiles, and the void-galaxy or void-void position correlations. Given the exciting potential applications, a rigorous comparison of theoretical predictions of these properties and those seen for voids in $N$-body simulations and galaxy surveys is very important.

However, this aim is complicated by the degree of ambiguity surrounding a very fundamental question: what exactly is a ‘void’? From a theoretical perspective, there is a clear answer, provided by the spherical evolution model of Sheth & van de Weygaert (2004), who identify voids as those non-linear underdense regions which have evolved to reach shell-crossing. This identification is convenient, as voids can then be modelled analogously to collapsed overdense halos using the excursion set formalism (Press & Schechter 1974).

* seshadri.nadathur@helsinki.fi

© 0000 RAS
In practical terms, however, the definition of a void is not so clear. When dealing with either N-body simulations or galaxy survey data, an algorithmic approach is required to identify regions as voids, which is complicated by the fact that voids are naturally poorly sampled by observable tracers, making shot noise a serious issue. A number of different void finders have been proposed (see Colberg et al. (2008) for a review of methods), which unfortunately do not always agree with each other. Watershed void finders (e.g. Platen et al. 2007, Neyrinck 2008, Sousbie 2011, Cautun et al. 2013) form an interesting class of algorithms. They use tessellation techniques (Schaap & van de Weygaert 2000) to reconstruct the density field from discrete data points, and the watershed algorithm for creating a void hierarchy. They present a number of advantages for practical studies, as they are more resilient to shot noise in the density reconstruction (though see also, e.g. Elyiv et al. 2015, Shandarin & Medvedev 2014, for other interesting proposals), and do not make prior assumptions about void geometries. They are also the most commonly used. Watershed voids may therefore be considered a reasonable practical definition of a void.

However, watershed algorithms make no reference to shell-crossing, which is the defining characteristic of theoretical models. The obvious question is therefore how, or whether, these two void definitions are related to each other. The answer is important to the practical use of watershed voids in cosmology, as well as to the development of further theoretical predictions. A number of studies (e.g. Achitouv et al. 2013, Sutter et al. 2013, Chan et al. 2014, Pisani et al. 2014, Chongchitnan 2015) which apply the Sheth & van de Weygaert (2004) formalism to describe watershed voids assume that the two approaches describe the same or closely related objects. There is known to be a significant disagreement between model predictions for void abundances as a function of their size and results obtained for watershed voids in dark matter simulations. This can be partially resolved (at least at large void sizes) by the ad hoc assumption that shell-crossing and void formation occur at less extreme densities than predicted by the spherical model, but as we will also show, such an approach lacks self-consistency. A more direct comparison of void properties with the model is therefore desirable.

At the same time, a number of properties of watershed voids remain imperfectly understood. A generic property of watershed void finders is that voids containing the deepest density minima should have the largest sizes. Yet the fitting form provided by Hamaus et al. (2014) to describe density profiles about void centres appears to suggest the opposite behaviour (though note that the applicability of this fitting form is not universally accepted, e.g. Nadathur et al. 2013). Perhaps related to this problem is the question of how to define the ‘centre’ of a void — which is also important for void correlation studies. The standard procedure assigns the centre to a weighted average of the positions of the tracers of mass within a void (e.g. Lavaux & Wandelt 2012, Sutter et al. 2012, Nadathur & Hotchkiss 2014, Sutter et al. 2014), but it is not clear that this will correctly identify the region with the greatest absence of mass. Another interesting question is how densities reconstructed from discrete tracer distributions relate to the true underlying density field. In studies of voids from simulations, the full simulation output is typically randomly down-sampled to provide a set of tracers, which can have a dramatic effect on the recovered void properties (Sutter et al. 2014).

Our goals in this paper are two-fold. We wish to understand the relationship between watershed voids and theoretical models. To do this we move beyond the fitting of void number functions alone and identify other important characteristics of voids which can be used to test the assumption of shell-crossing more broadly. We also want to empirically examine the properties of watershed voids in simulations in order to understand the working of the algorithm and clarify some of the issues above.

To do so we make use of the popular ZOBOV watershed algorithm (Neyrinck 2008). To enable comparison of our results with others in the literature, we will mostly use the options for ZOBOV implemented in the VIDE toolkit (Sutter et al. 2015). We analyse voids identified using randomly subsampled dark matter particles as tracers, and relate them to true densities determined from the full resolution simulation output. We propose a new definition of the void centre, which is designed to better identify the true location of the underdensity within the void. The details of the N-body simulation, the watershed algorithm and the methods for identifying void centres and measuring density profiles are described in Section 2.

Section 3 provides a summary of the spherical model for void evolution, which we use to extract general identifying characteristics of shell-crossed voids for comparison with simulation results. This comparison is performed in Section 4, where we also outline the general properties of voids in our simulation. We show that larger voids do correspond to deeper density minima, as expected for the watershed algorithm. We examine the viability of the fitting formula of Hamaus et al. (2014) to describe the density profiles of voids, and show that it does not provide a good quantitative or qualitative description of the variation of the average profile within the void population. In addition, subsampled tracers overestimate the density contrast in voids. All of these results have practical implications for future studies that use watershed void finders. We compare these results from simulated voids to theory and argue that there is no evidence that watershed void finders in general, and VIDE and ZOBOV in particular, satisfy the primary defining criteria of the Sheth & van de Weygaert (2004) model. This leads us to reassess the viability of describing watershed voids using existing theoretical techniques. We summarise and conclude in Section 5.

2 NUMERICAL METHODS

2.1 Simulation

In this paper we use N-body simulation data from the MultiDark Run1 (MDR1) release of the MultiDark project (Prada et al. 2012) \footnote{Publicly available from \url{www.cosmosim.org}.} MDR1 uses an Adaptive-Refinement-Tree (ART) code, based on adaptive mesh refinement, to simulate 2048$^3$ dark matter particles within
a cubic volume of $1 \left(h^{-1}\text{Gpc}\right)^3$, in a $\Lambda$CDM cosmological model with parameters $(\Omega_m, \Omega_\Lambda, \Omega_b, h, n_s, \sigma_8) = (0.27, 0.73, 0.0469, 0.7, 0.95, 0.82)$. The simulation has mass resolution $m_p = 8.721 \times 10^9 \ h^{-1} M_\odot$ and force resolution $7 \ h^{-1}$kpc. Initial conditions were set using the Zeldovich approximation at redshift $z = 65$.

From the full particle output at redshift 0 we randomly subsample the dark matter particles down to a number density of $n = 3.2 \times 10^{-3} \ h^3 \text{Mpc}^{-3}$, similar to that of typical galaxy samples (e.g. Zehavi et al. 2011). This corresponds to a mean nearest-neighbour separation of $\bar{n}^{-1/3} \sim 7 \ h^{-1}$Mpc. We refer to the resulting sample as the Main sample, and use these particles as tracers for the void finding. In addition, we have used a control sample with a higher tracer density $8.7 \times 10^{-3} \ h^3 \text{Mpc}^{-3}$, which we refer to as the Dense sample. However, our primary conclusions regarding the properties of watershed voids do not depend strongly on the tracer number density. Therefore unless otherwise stated, all results presented in this paper refer to voids from the Main sample.

Note that a random subsampling of tracers introduces shot noise but does not change the fundamental clustering properties of the dark matter field. Therefore despite having the same tracer number density, the properties of voids in such subsampled tracers and those in the galaxy distribution would not be expected to be (and are not) the same, since galaxies are biased tracers of the matter density. However for our purposes of understanding the general properties of watershed voids in this work, a random subsampling is sufficient. We consider the effects of galaxy bias separately in a companion paper (Nadathur & Hotchkiss 2015).

Although our tracers are themselves dark matter particles, as the subsampling procedure increases shot noise we will distinguish between the tracer number density, and the underlying dark matter density. The dark matter density in MDR1 is determined from the full resolution particle output of the simulation at redshift 0, by using a cloud-in-cell interpolation on a $1024^3$ grid, followed by smoothing with a Gaussian kernel with width equal to one grid cell. The sub-Mpc resolution of this grid is much smaller than the typical void size scales, so that this procedure in effect provides a continuous underlying dark matter density, of which the subsampled tracer particles are an approximately Poisson realization.

In the following, we will reserve the symbols $\rho$ and $\Delta$ for dark matter densities determined using this gridded smoothed density field, and use the symbols $n$ and $\Delta_n$ for the equivalent quantities determined from the tracer number densities.

### 2.2 Void finding

To identify voids in the dark matter particle distribution, we make use of the ZOBOV watershed void finder (Neyrinck 2008), with the options implemented in the VIDE toolkit (Sutter et al. 2015). Although there are known issues with the application of VIDE to galaxy survey data with irregular survey volumes and masks (Nadathur & Hotchkiss 2014), when dealing with a simulation cube with periodic boundary conditions these do not present a problem. However, note that in some cases, especially the definition of the void centre described below, we use our own modification of the ZOBOV algorithm.

ZOBOV works by reconstructing the density field based on a Voronoi tessellation of the simulation cube around the discrete distribution of tracer particles. The tessellation associates each particle with a Voronoi cell consisting of the region of space closer to it than to any other particle. The volume of the Voronoi cell $i$ relative to the mean volume is then used to estimate the local tracer number density $n_i$ at the particle location. Such a reconstruction is naturally scale-adaptive and thus far more resilient against shot noise effects than naive counts-in-cells measurements.

After reconstructing the density field, the algorithm identifies all local minima of the reconstructed density field and determines the “catchment basins” around each minimum, known as zones. Zones are then merged to form a nested hierarchy of voids according to the watershed principle (Platen et al. 2007), such that the zone with the smallest minimum density $n_{\text{min}}$ then acquires neighbouring higher-density zones as sub-voids, in increasing order of the minimum density on the watershed ridge separating them. For each void thus formed, we define an effective void radius $R_v$ as the radius of a sphere with equivalent volume $V$,

$$R_v = \left( \frac{3}{4\pi} V \right)^{1/3}. \quad (1)$$

Even in the absence of any merging, deeper density minima typically correspond to zones of larger volume and thus larger $R_v$. However, the watershed merging procedure also ensures that voids with deepest density minima contain greater numbers of merged sub-voids and therefore have the largest sizes. This correlation of minimum density and void size is a common property of all watershed void-finders and is not unique to the ZOBOV algorithm.

To avoid excessive merging leading to essentially infinite void sizes, VIDE imposes a restriction preventing the merger of two zones unless the minimum link density along the watershed ridge separating them satisfies $n_{\text{link}} < n_{\text{max}}$, where $n_{\text{max}} = 0.27$. This condition applies only to the lowest density point on such a ridge, and does not prevent voids from containing regions of much higher densities. The value of 0.2 therefore has no theoretical motivation, and $n_{\text{max}}$ should be considered an arbitrary free parameter. Indeed alternative values of $n_{\text{max}}$ have been considered in other works (Achitouv et al. 2013; Nadathur & Hotchkiss 2014; Hotchkiss et al. 2015; Nadathur et al. 2015), and properties such as the abundance of root-level voids, the distribution of void sizes and void density profiles will all depend on the value chosen. However, for ease of comparison with previous results we shall restrict ourselves to the default value hard-coded in VIDE.

Further selection cuts might be desirable at this stage, since ZOBOV simply reports all local density minima as potential voids, without regard to the value of the minimum density within them or any reference to shell-crossing. VIDE requires that the tracer number density within a defined central region of the void, as measured by naive number counting, be less than 0.27. However, as pointed out by Nadathur & Hotchkiss (2013), such a density measure is extremely badly affected by shot noise. We find that it is essentially uncorrelated with the reconstructed $n_{\text{min}}$, and in any case only excludes a small fraction of final voids, render-
ing it irrelevant. We also do not apply the much more conservative cuts on \( n_{\text{min}} \) suggested by Nadathur & Hotchkiss (2014), Hotchkiss et al. (2015), Nadathur et al. (2015).

A selection cut based on the void radius has sometimes been advocated in the literature, to remove voids with \( R_v < \pi^{-1/3} \), which are claimed to be below the resolution limit. In fact the adaptive nature of the tessellation means that ZOBOV automatically excludes small voids below its resolution limit, as we show in Section 4. Therefore no further cut on \( R_v \) is necessary.

By applying these criteria we find in total 27,450 voids in the Main tracer sample. Of these, 26,919 are root-level voids in the hierarchy, i.e. they are not sub-voids of any parent voids and their volumes do not overlap each other.

### 2.3 Void centres

Since voids obtained from from the watershed algorithm have arbitrary shapes, different prescriptions may be used to define the location of the void ‘centre’. The most commonly used definition, which is also the definition implemented in VIDE, is the volume-weighted barycentre of the void member particles

\[
X_{vbc} = \frac{1}{\sum_i V_i} \sum_i V_i x_i,
\]

where \( x_i \) is the position of the \( i \)th particle, \( V_i \) is the volume of it’s corresponding Voronoi cell, and the sum runs over all member particles of void \( v \).

In the low-density interior of a void, Voronoi cells in the tessellation are typically greatly elongated, and the particles contained within them lie far from their geometrical centres. This means that the position \( x_i \) corresponding to each cell is an imprecise measure of the location of the cell. In addition, watershed voids contain a great number of member particles — the median number for our void sample is 82, and many voids contain several hundreds — the vast majority of which reside in the overdense walls and filaments on the outskirts of voids. A combination of these two factors means that although the barycentre position defined by eq. 2 is roughly symmetrically located with respect to the overdense void walls, it is typically very far from the position of minimum density. This is because the barycentre definition is fundamentally based on the locations where tracers are present, rather than locations from where they are absent.

For most purposes, it would appear more logical to define the void centre to coincide with the location of the minimum density within it. To do this, we adopt the following procedure. We identify the core particle of the void as the particle with the largest Voronoi cell (i.e., corresponding to the minimum tracer density \( n_{\text{min}} \)), and examine the tessellation output to identify all Voronoi cells adjacent to it. From this set we select the lowest density neighbouring particle, and then, in order of increasing density, two other particles that are adjacent to both the core particle and the previous selections. This provides us with the four lowest density mutually adjacent Voronoi cells in the void; we now define the void centre to lie at the point of intersection of these four cells, which is also the circumcentre of the tetrahedron formed by the four tracer particles. This point represents the location within the void that is maximally distant from all tracers.

We shall refer to this alternative definition of the void centre as the circumcentre and denote its location by \( X_{vcc} \). In Section 3 we show that both the number density of tracers and the underlying dark matter density are indeed significantly lower at the circumcentre than the barycentre.

### 2.4 Density profile determination

A fundamental quantity of interest is the average distribution of tracers and dark matter about the void centre, and the variation of this distribution with void properties. We study this behaviour by constructing stacked density profiles for subsets of voids satisfying different criteria. To do so we rescale distances within each qualifying void in units of the void radius \( R_v \), and then estimate the average density in the stack in concentric spherical shells about the void centre.

Estimating the tracer number density in this way is complicated by shot noise effects, since the interiors of voids by definition contain very few tracer particles which can be used for number density measurements. Nadathur et al. (2015) showed that an unbiased estimate accounting for Poisson noise can be obtained using the volume-weighted estimator for the average number density in the \( j \)th radial shell,

\[
\bar{n}_j = \frac{\sum_{i=1}^{N_j} N_j V_i}{\sum_{i=1}^{N_j} V_i},
\]

where the \( j \)th shell has width \( \Delta \bar{r} \) in units of the rescaled radial distance \( \bar{r} \) for each void, \( V_i \) is the true volume of the \( i \)th shell of the \( j \)th void and \( N_j \) is the number of tracer particles contained within it, and the sum over \( i \) runs over all voids included in the stack. Note that the \( N_j \) in this formula includes all tracer particles within the shell, not just those that are identified as members of the void by the watershed algorithm. Under the assumption that the individual numbers \( N_j \) are Poisson realizations of the true underlying density, the error in eq. 3 can then be estimated at any desired confidence level directly from the definition of the Poisson distribution. In this paper plotted errorbars indicate the 68% confidence limits on \( n \).

As the resolution of the dark matter density field is much finer than the typical void size, no such complications are required when estimating \( \rho \) over the stack of voids. We simply sample the dark matter density at all grid points contained within the radial shell and calculate the mean and standard deviation of the values obtained. We present our results for the stacked dark matter densities in terms of the average total enclosed density within a radius \( r, 1 + \Delta r \), since this allows a more direct contact with the theory described in Section 3.

All density profiles are calculated out to three times the void radius, and are measured in radial bin steps of 0.1 times the radius.

### 3 EXCURSION SET MODELS OF VOIDS

Most existing theoretical descriptions of voids are derived from the framework presented by Sheth & van de Weygaert (2004). This in turn derives from the original excursion set approach of Press & Schechter (1974). Epstein (1983) and Bond
et al. (1991) and is based on the model of spherical evolution of mass shells (Gunn & Gott 1972; Lilje & Lahav 1991). In this Section we briefly summarize such models in order to highlight the key areas of comparison with the results of watershed void finders.

In this picture the evolution of a spherical mass shell of radius \( r \) is determined by the total enclosed density contrast within the radius of the shell at time \( t \), \( 1 + \Delta(r, t) \), where

\[
\Delta(r, t) = \frac{3}{r^2} \int_0^r \left( \frac{\rho(y, t)}{\bar{\rho}(t)} - 1 \right) y^2 dy,
\]

and by the time evolution of the cosmological density parameter \( \Omega(t) \). Underdense spherical regions contain a density deficit (i.e., \( \Delta(r, t) < 0 \)) which causes shells to expand outwards. This deficit is stronger for inner shells, which therefore expand faster than outer shells, and mass evacuated from the centre of the underdensity begins to pile up at its edges. For a steep enough starting density profile, at some point in the evolution inner shells catch up with shells which were initially further out from them, in an event known as shell-crossing. The moment of shell-crossing marks a transition in the evolution of the underdensity, as it subsequently expands outwards self-similarly (Suto et al. 1984; Fillmore & Goldreich 1984; Bertshinger 1985).

Within the spherical model, it can be shown that shell-crossing occurs when the average density enclosed within the void is

\[
\rho_{\text{enc}}/\bar{\rho} = 1 + \Delta(r, t) \simeq 0.2.
\]

This corresponds to a linearly extrapolated average density contrast of

\[
\Delta_{\text{lin}} = \delta_{\text{c}} \simeq -2.81,
\]

at the epoch of shell-crossing, with this value independent of radius \( r \) and largely independent of the cosmological parameters governing the background evolution. This is analogous to the case of spherical collapse of clusters, which occurs above a linear overdensity threshold of \( \Delta_{\text{lin}} = \delta_{\text{c}} \simeq 1.69 \).

Following Blumenthal et al. (1992); Dubinski et al. (1993); Sheth & van de Weygaert (2004) then identify the population of voids with only those evolved underdensities that have reached the stage of shell-crossing. If the initial Gaussian density fluctuation field is smoothed on a range of different smoothing scales \( R \), this physical picture identifies fluctuations which exceed the density threshold \( \delta_{\text{c}} \) on smoothing scale \( R \) with potential voids of radius \( R \) today. These fluctuations can be characterized by their depth in units of the rms fluctuation of the density field on scale \( R \)

\[
\nu \equiv \frac{\delta_{\text{c}}^2}{\sigma_0^2(R)},
\]

where \( \sigma_0(R) \) is one of the set of spectral moments

\[
\sigma_j^2(R) \equiv \int \frac{k^{2+2j}}{2\pi^2} W^2(kR) P(k) dk,
\]

with \( P(k) \) the power spectrum of the unsmoothed density fluctuation field and \( W(kR) \) the smoothing filter.

At this point, the abundance and size distribution of voids can be predicted by a number of models of varying degrees of sophistication. Sheth & van de Weygaert (2004) use the excursion set approach (e.g., Bond et al. 1991; Sheth 1998) to account for fluctuations which cross the \( \delta_{\text{c}} \) threshold on some small scale but are overdense with \( \Delta_{\text{lin}} > \delta_{\text{c}} \) on some larger scale. Such underdensities would be crushed by the collapse of the surrounding cluster and so would not be visible as voids today (the void-in-cloud effect). This amounts to a two-barrier problem. According to this model, assuming void number density is conserved on evolving from Lagrangian to Eulerian space, this number density can be expressed as a function of the Eulerian void radius \( R_v \) (e.g., Jennings et al. 2013; Chan et al. 2014) as

\[
\frac{dN}{dR_v} = \left( \frac{3}{4\pi R_{L}^2} \right) f(\nu) \frac{d\nu}{dR_v},
\]

where

\[
f(\nu) \simeq \sqrt{\frac{1}{2\nu^2}} \exp\left(-\frac{\nu}{2}\right) \exp\left(-\frac{|\delta_c| D^2}{4\nu} - 2\frac{D^4}{\nu^2}\right),
\]

and

\[
D \equiv \frac{|\delta_c|}{\delta_{\text{c}} + |\delta_c|}.
\]

Here the Lagrangian radius \( R_L = 0.58 R_v \), a relationship determined by the shell-crossing condition above.

However, it is well known that eq. (9) does not provide a good fit to the distribution of voids found by watershed algorithms, since it predicts a sharp cutoff in void sizes above \( \sim 5 \, h^{-1}\text{Mpc} \), much smaller than observed for watershed voids. A number of studies (Jennings et al. 2013; Sutter et al. 2014; Chan et al. 2014; Pisani et al. 2015) have attempted to improve fits by relaxing the shell-crossing condition \( \delta_{\text{c}} \sim -2.81 \) and treating \( \delta_{\text{c}} \) as a free parameter instead. Note that simply allowing \( \delta_{\text{c}} \) to vary freely is not even a theoretically self-consistent procedure unless the relationship \( R_L = 0.58 R_v \) is also correspondingly altered, though this is not usually done. Despite this, eq. (9) fails to describe the distribution of small voids, and the fit values of \( \delta_{\text{c}} \) for large voids vary widely. Chan et al. (2014) obtain \( \delta_{\text{c}} \simeq -1 \) when fitting to voids of radius \( > 20 \, h^{-1}\text{Mpc} \) only, and find little redshift dependence of this value, contrary to theoretical expectation. On the other hand, Sutter et al. (2014) find a range of different \( \delta_{\text{c}} \) values for voids from different samples, ranging from \(-0.26 \) to \(-0.5 \). Pisani et al. (2015) quote \( \delta_{\text{c}} \simeq -0.45 \). It is hard to construct a theoretical explanation for such low values of \( \delta_{\text{c}} \).

Another feature of such fits is that insofar as such models with free \( \delta_{\text{c}} \) describe the distribution of the largest voids, they do so simply by replicating an exponential cutoff in the distribution at large \( R_v \). To demonstrate this, in Figure 1 we show the distribution of void sizes obtained from both our simulation samples, together with the form of eq. (9) with value \( \delta_{\text{c}} \simeq -1.12 \) obtained from fitting to the distribution of voids with \( R_v > 25 \, h^{-1}\text{Mpc} \). The fit value of \( \delta_{\text{c}} \) is highly sensitive to the choice of this radius cut, but the quality of

\[ ^3 \text{Jennings et al. (2013) also propose an alternative adaptation of this model, but this cuts off the distribution at even smaller } R_v, \text{ so would make the discrepancy worse.}
\]

\[ ^4 \text{To present a fair comparison with other studies, in obtaining}
\]
Therefore no longer strictly self-consistent. In this fit we also do not vary $R_v$ when changing $\delta_v$. The model is therefore no longer strictly self-consistent.

A related property of shell-crossed voids is that if the enclosed density contrast is to be the same at all void radii, eq. 9 requires that larger voids must correspond to more extreme fluctuations of the parameter $\nu$. It can be shown that this in turn means that larger voids should on average correspond to shallower but broader initial density profiles $\delta(r)$, while smaller voids correspond to deeper and steeper profiles. That is, smaller shell-crossed voids should contain deeper density minima than large voids.

The analogous situation for collapsing halos is that the most massive halos should be the least centrally concentrated, which is indeed the case (e.g. Navarro et al. 1996, 1997). More generally, voids with the deepest density minima should have the steepest density profiles, and vice versa.

These qualitative properties provide clear tests of the assumption that watershed voids have undergone shell-crossing. However, as we show in the next Section, neither of them hold true for the voids obtained using VIDE and ZOBOV, nor should one expect them to hold for other watershed void finders.

4 PROPERTIES OF WATERSHED VOIDS

4.1 Sizes and densities

Figure 2 shows the distribution of void sizes and minimum tracer number densities for all voids in our Main tracer sample. It is immediately obvious that lower minimum number densities are correlated with larger void sizes, as we argued would always be the case for watershed void finders. This means that the selection criterion $r_v > r_N$ advocated by some studies (e.g. Sutter et al. 2012, Sutter et al. 2014) has no practical effect, whereas a tighter criterion $R_v > 2r_N$ (Hamaus et al. 2014) is unnecessarily conservative.

Also shown are contours showing the 95% and 99% confidence limit contours for the distribution of spurious ‘voids’ and may introduce a scatter in the values of $\Delta$ over the void population. Nevertheless, strong variation of the enclosed density with properties of watershed voids would be a clear sign that they do not correspond to similar shell-crossed objects.

![Figure 1. The differential number density of voids in simulation as a function of their size, for both simulation samples. Error bars are calculated assuming the void numbers in each bin are Poisson distributed. The dashed line shows the best fit of the Sheth & van de Weygaert (2004) model to the $R_v > 25 h^{-1} \text{Mpc}$ data, with $\delta_v = -1.12$. The solid line shows an exponential cutoff model which describes the same data better.](image-url)
Figure 2. The distribution of the minimum tracer number densities within voids and void sizes in the Main sample. There is a clear trend towards increasing void size as the minimum density decreases. The dotted lines show the contours enclosing 95% and 99% of all ‘voids’ identified in a random uniform distribution of points with the same number density and in the same volume. The arrow indicates the value $R_v = \pi^{-1/3}$, roughly the mean inter-particle separation, which has sometimes been suggested as minimum size cut. The dashed line shows the true minimum achievable void size resolution as a function $n_{\text{min}}$: most voids automatically lie well away from this limit.

Figure 3. Binned average values of the dark matter density at the location of the void centre, as a function of the void size. The two sets of points refer to the same voids in the Main sample, but to two different definitions of the void centres, as described in the text. Bins are chosen to contain equal numbers of voids. The dark matter density is determined from the full resolution simulation output.

The nature of voids I

identified by the same algorithm in a random uniform distribution of points with the same volume and same number density as the Main sample. Clearly, although there is considerable overlap, the correlations between the dark matter particle positions produce a rather different distribution.

Another noteworthy aspect of Figure 2 is the apparent saturation of the minimum densities within voids with increasing $R_v$. This is a consequence of the finite tracer number density, and the saturation value is dependent on the mean density $\bar{n}$. Subsampling tracer particles lowers $\bar{n}$ and thus reduces the apparent tracer density contrast in voids. Conversely, Nadathur & Hotchkiss (2015) show that at the same mean tracer density, more highly biased tracers result in much lower values of $n_{\text{min}}$ within voids.

Given the uncertainties associated with the tracer number density discussed below, the true dark matter density at void locations is perhaps a more informative quantity. To measure this we make use of the dark matter density field described in Section 2.1 and simply measure its value $\rho$ in the grid cell corresponding to the position of the void centre. Unsurprisingly, there is a considerable scatter in these values. Figure 3 shows the binned average values as a function of the void radius $R_v$, for the same set of voids but for both definitions of the void centre described in Section 2.3. Bins were chosen to contain equal numbers of voids, and the error bars represent the 1σ uncertainty in the mean. The circumcentre is clearly better at locating the regions of low density within the void, especially so for smaller voids. However for both centre definitions $\rho$ clearly decreases with increasing void size.

Figure 4 shows the distribution of the average tracer density $n_{\text{avg}}$ within the void, calculated from the number of void member particles and the void volume, and $R_v$. As pointed out by Achitouv et al. (2013), Nadathur & Hotchkiss (2014), $n_{\text{avg}}$ is typically $\gtrsim 1$ and much larger than $n_{\text{min}}$, simply because the watershed definition means that voids always extend to include high density regions on the separating ridges. This feature of ZOBOV and VIDE indirectly
demonstrates that most tracers in identified voids reside in overdensities, which explains why the barycentre is a poor locator of the minimum underdensity, and also suggests that the void radius can be significantly overestimated when \( n_{\text{avg}} > 1 \).

It is worth noting that a selection cut on the minimum void radius alone is a sub-optimal way of excluding voids that are on average overdense, since it would eliminate many with the lowest \( n_{\text{avg}} \) values as well.

### 4.2 Tracer density vs. dark matter density

The relationship between the tracer number density and the true underlying dark matter density in the simulation is also of interest. Even though the tracers in our case are subsampled dark matter particles, we find that these two quantities are in general not the same. Subsampling of the dark matter particles has two effects on their relationship. Firstly, there is inherent shot noise in tracer densities arising from the fact that the tracers are a discrete realization of the underlying continuous density field, which is enhanced by subsampling to lower \( \pi \). This means that, particularly in void regions, tracer densities are not a precise reflection of the true dark matter density. This problem is to a large extent mitigated by the self-adaptive nature of the tessellation density reconstruction, but cannot be completely removed. Secondly, decreasing \( \pi \) through subsampling brings the void minimum tracer density \( n_{\text{min}} \) closer to the mean.

These effects are illustrated in Figure 5 which shows the relationship between the dark matter density \( \rho \) at the position of the void centre and the minimum tracer number density within the void as determined from the tessellation, for both definitions of the void centre, and for voids from both the Main and Dense tracer samples. That \( \rho \) exceeds \( n_{\text{min}} \) at the barycentre for both samples is to be expected, since the barycentre typically lies quite far from the location of the tracer density minimum. But even at the circumcentre, which is guaranteed to lie in the region of minimum tracer number density, although there is a clear linear relationship between \( \rho \) and \( n_{\text{min}} \), the tracer number density does not accurately reflect the dark matter density, particularly in the most underdense voids. Voids in the Dense sample tend both to have lower \( n_{\text{min}} \) than Main voids, and to locate the true dark matter underdensities slightly better.
4.3 Density profiles

We now turn to the distribution of tracer particles and dark matter around void centres. Anticipating that the form of the density profiles will depend on the void size, we first examine the average profiles for stacks of voids within different ranges of $R_v$, chosen such that each stack contains an equal number of voids.

The results for the tracer density profile obtained from eq. 3 are shown in Figure 6. Profiles in the left panel are for voids stacked around their barycentres, and can be directly compared with the results of Hamaus et al. (2014) (although note that due to the improved statistical stability of our density estimator, we are able to calculate the density at all distances from the void centres).

We find a number of noteworthy trends. The overdensities in the surrounding walls at $r \sim R_v$ are much higher for small voids than for large ones. Tracer density profiles also transition from undercompensated to overcompensated as the average void radius in the stack decreases, and for the smallest voids the profile does not return to the mean even at large distances from the centre. These findings are qualitatively in agreement with those of Hamaus et al. (2014). However, in the void interiors we find a strong trend towards decreasing central density with increasing void size. This is entirely consistent with the expected generic behaviour of watershed void finders and the distribution shown in Figure 2 but contradicts the results of Hamaus et al. (2014). The fitting formula provided by those authors therefore fails to describe the profiles we obtain. We also find no evidence for the self-similarity of tracer density profiles seen by Nadathur et al. (2015), but in this case differences in method-
Figure 8. Stacked profiles of the total enclosed dark matter density, for voids within the same size range $15 < R_v < 20 \, h^{-1} \text{Mpc}$, but with different minimum tracer densities. Profiles in the left panel are stacked about the barycentres, and in the right panel about the circumcentres.

The effect of the change in definition of the void centre can be seen by comparison with the right panel of Figure 3, which shows the stacked density profiles for the same voids, but based around the void circumcentre. Unsurprisingly, voids of all sizes show extremely low tracer number densities close to the circumcentre. The general asymmetry of the circumcentre location with respect to particles in the void walls is also apparent in the fact that the stacked profiles about this location are less able to resolve the high densities in these walls. This is the essential tradeoff between the two centre definitions: the barycentre has a greater degree of symmetry with respect to the surrounding overdensities, whereas the circumcentre identifies the true location of the underdensity.

Close to the circumcentre, the tracer densities estimated by number counts using eq. 3 are very close to zero for all void sizes, although $n_{\text{min}}$ values are never so small and vary with void size. This is because the circumcentre is a special point. In a discrete distribution of tracers it is always possible to choose a point such that a sufficiently small volume (smaller than the mean volume per particle) around it contains no tracers at all. The Voronoi tessellation avoids this issue because of its self-adaptive resolution. It can be seen from Figure 3 that the Voronoi reconstructed $n_{\text{min}}$ is a much better predictor of the true dark matter density than number counts — which is why it is preferable to reconstruct the density field from the tessellation in the first place. For this reason, such stacked number density profiles should not be relied upon for quantitative analysis without calibration.

For this purpose we instead make use of the full dark matter density field at high resolution. Profiles of the average enclosed dark matter density are shown in Figure 8 for the same void stacks as before. These confirm some properties of watershed voids which are of significance for the attempts to model them theoretically. Firstly, as already seen in Figures 2 and 3, larger voids contain deeper density minima. Secondly, the enclosed density contrast within these voids is $\Delta(r) > -0.8$, for all void sizes and at all distances $r$. The condition for shell-crossing to occur is thus not satisfied at any point within the average void.

So far, following earlier works (Hamaus et al. 2014; Nadathur et al. 2015), we have only considered the variation in the mean profile with the size of the voids included in the stack, but it is clear that this cannot be the only important variable. In fact, as shown in Figure 3, voids of similar sizes but different minimum densities $n_{\text{min}}$ have very different density profiles. Voids with different $n_{\text{min}}$ clearly do not enclose the same density contrasts, and deeper density minima do not correspond to steeper density profiles. The enclosed density contrast $\Delta$ clearly varies widely over the void population, providing further evidence that the population of watershed voids does not satisfy the foundational assumption of the excursion set model.

It is also clear that a more complete description of the density profiles around voids is obtained by accounting for the extent of variation in both dimensions of the $(n_{\text{min}}, R_v)$ plane. Fitting formulae such as those provided by Hamaus et al. (2014) or Nadathur et al. (2015), which account only for variation with void radius, will be unable to describe the full variety of watershed voids.

However, the distribution of highly biased galaxies will trace dark matter underdensities rather differently than the randomly subsampled dark matter particles we have used in this work, and it is the dark matter profiles of galaxy voids which are of greater practical interest in cosmology.
Therefore although the profiles obtained here provide genuine qualitative insights, we do not attempt to extract numerical fits to the data.

5 CONCLUSIONS

Our aim in this paper was to provide an empirical investigation into the properties of watershed voids in order to better understand the operation of void finding algorithms such as VIDE and ZOBOV and the relation to theory. Several previous studies have focussed on the distribution of void sizes alone, and have attempted to fit this using modifications of the spherical evolution model. Such an approach however misses the important relationship between void size and density: larger voids correspond to deeper density minima. This is a fundamental feature of ZOBOV that holds irrespective of whether the tracers used for void identification are simulation dark matter particles, halos or galaxies. It is also a more general property that should apply to any watershed void finder.

The conclusion that follows from this relationship — and which we also demonstrate directly through stacked density profiles around void centres — is that watershed voids cannot correspond to a population of objects which all enclose the same density contrast, which is the principal starting assumption of theoretical descriptions deriving from the model of Sheth & van de Weygaert (2004). It has long been known that the void number function prediction of this model fails to match that of watershed voids by many orders of magnitude. It has sometimes been argued without proof (Sutter et al. 2014; Chongchitnan 2015) that the void formation threshold δc might differ from the shell-crossing value in the spherical model due to the generally aspherical nature of watershed voids. This assumption has led several authors to treat δc as a free parameter but without altering the basic model. We have already argued that in practice this approach has not been self-consistently applied in obtaining the fit to δc. More importantly, given the range in enclosed density contrasts Δ over the watershed void population, a single value of δc for all voids does not seem tenable. Even more suggestive is the fact that for no subset of these voids does the average enclosed density contrast satisfy the criterion for shell-crossing, Δ ≳ −0.8, at any radial distance from the centre, let alone at the void radius Rv.

The simplest interpretation of this evidence is that watershed voids simply do not correspond to objects that have undergone shell-crossing. With hindsight this should not seem surprising — ZOBOV uses only information on the local topology of the density field, and makes no reference to shell-crossing. Furthermore, neither VIDE nor ZOBOV apply any meaningful conditions even on the minimum tracer density nmin, within voids, instead reporting all local density minima. Attempts to explain how the shell-crossing density criterion may be altered in such voids therefore seem to be misguided. A simpler starting proposition would be to give up the enforced assumption of shell-crossing and to describe watershed voids simply as what they are: regions of density minima.

We should stress that breaking this link to theoretical models of shell-crossed voids does not necessarily make the results obtained from watershed void finders less useful for practical cosmological studies. For instance, these voids can still be used to identify large-scale underdense environments. Some of them (though not all) will also correspond to maxima of the gravitational potential, and so they can still be used for studies of lensing (Melchior et al. 2014) or the ISW effect (Cai et al. 2014; Hotchkiss et al. 2015; Planck Collaboration et al. 2015). Equally, we do not intend to claim that the Sheth & van de Weygaert (2004) model does not correctly describe shell-crossed underdensities on much smaller scales. It is simply that this and related models do not match simulation or observational data because the word ‘void’ has a different meaning in the two contexts.

Another interesting feature of our results is the relationship between the underdensity in voids measured using subsampled tracers and using the full resolution dark matter density. We show that values of n and ρ do not completely agree, and apparent tracer underdensities in deep voids are deeper than the true dark matter minima. The relationship between n and ρ depends on the mean sampling density of tracers; it will certainly also change if biased tracers are used. This does not affect the basic operation of ZOBOV, which only uses relative tracer densities to identify minima, but it argues against the use of absolute values of the central tracer density in applying selection cuts, as has sometimes been suggested (e.g. Sutter et al. 2012; Jennings et al. 2013; Sutter et al. 2014; Nadathur & Hotchkiss 2014). In other words, selecting a region which apparently satisfies the shell-crossing criteria in terms of the tracer number density does not ensure that it does so in the true matter density. This was already pointed out by Furlanetto & Piran (2006) for the case when the tracers are galaxies; our results show that it applies even if the tracers are a subset of dark matter particles in the simulation.

6 ACKNOWLEDGEMENTS

We thank Ravi Sheth for stimulating correspondence and Alexis Finoguenov for helpful discussions. SH acknowledges support from the Science and Technology Facilities Council [grant number ST/L000652/1].

The MultiDark Database used in this paper and the web application providing online access to it were constructed as part of the activities of the German Astrophysical Virtual Observatory as result of a collaboration between the Leibniz-Institute for Astrophysics Potsdam (AIP) and the Spanish MultiDark Consolider Project CSD2009-00064. The MultiDark simulations were run on the NASA’s Pleiades supercomputer at the NASA Ames Research Center.

REFERENCES

Achitouv I., Neyrinck M., Paranjape A., 2013, ArXiv e-prints, 1309.3799
Alcock C., Paczynski B., 1979, Nature, 281, 358
Bardeen J. M., Bond J., Kaiser N., Szalay A., 1986, ApJ, 304, 15
Bertschinger E., 1985, ApJS, 58, 1
Blumenthal G. R., da Costa L. N., Goldwirth D. S., Lecar M., Piran T., 1992, ApJ, 388, 234
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bos E. G. P., van de Weygaert R., Dolag K., Pettorino V., 2012, MNRAS, 426, 440
Cai Y.-C., Neyrinck M. C., Szapudi I., Cole S., Frenk C. S., 2014, ApJ, 786, 110
Cai Y.-C., Padilla N., Li B., 2014, ArXiv e-prints, 1410.1510
Cautun M., van de Weygaert R., Jones B. J. T., 2013, MNRAS, 429, 1286
Chan K. C., Hamaus N., Desjacques V., 2014, Phys.Rev.D, 90, 103521
Chongchitnan S., 2015, ArXiv e-prints, 1502.07705
Colberg J. M. et al., 2008, MNRAS, 387, 933
Dubinski J., Nicolaci da Costa L., Goldwirth D., Lecar M., Piran T., 1993, ApJ, 410, 458
Elyiv A., Marulli F., Pollina G., Baldi M., Branchini E., Cimatti A., Moscardini L., 2015, MNRAS, 448, 642
Epstein R. I., 1983, MNRAS, 205, 207
Fillmore J., Goldreich P., 1984, ApJ, 281, 1
Flender S., Hotchkiss S., Nadathur S., 2013, JCAP, 1302, 013
Furlanetto S., Piran T., 2006, MNRAS, 366, 467
Gunn J. E., Gott III J. R., 1972, ApJ, 176, 1
Hamaus N., Sutter P. M., Lavaux G., Wandelt B. D., 2014, JCAP, 12, 13
Hamaus N., Sutter P. M., Wandelt B. D., 2014, Physical Review Letters, 112, 251302
Hamaus N., Wandelt B. D., Sutter P. M., Lavaux G., Warren M. S., 2014, Physical Review Letters, 112, 041304
Hotchkiss S., Nadathur S., Gottlöber S., Iliev I. T., Knöbel A., Watson W. A., Yepes G., 2015, MNRAS, 446, 1321
Jennings E., Li Y., Hu W., 2013, MNRAS, 434, 2167
Kamionkowski M., Verde L., Jimenez R., 2009, JCAP, 1, 10
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lavaux G., Wandelt B. D., 2012, ApJ, 754, 109
Lee J., Park D., 2009, ApJ, 696, L10
Li B., Zhao G.-B., Koyama K., 2012, MNRAS, 421, 3481
Lilje P. B., Lahav O., 1991, ApJ, 374, 29
Melchior P., Sutter P. M., Sheldon E. S., Krause E., Wandelt B. D., 2014, MNRAS, 440, 2922
Musso M., Sheth R. K., 2012, MNRAS, 423, L102
Nadathur S., Hotchkiss S., 2013, ArXiv e-prints, 1310.6911
Nadathur S., Hotchkiss S., 2014, MNRAS, 440, 1248
Nadathur S., Hotchkiss S., 2015, in preparation
Nadathur S., Hotchkiss S., Diego J. M., Iliev I. T., Gottlöber S., Watson W. A., Yepes G., 2015, MNRAS, 449, 3997
Nadathur S., Lavinto M., Hotchkiss S., Rääsänen S., 2014, Phys.Rev.D, 90, 103510
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neyrinck M. C., 2008, MNRAS, 386, 2101
Pan D. C., Vogeley M. S., Hoyle F., Choi Y.-Y., Park C., 2012, MNRAS, 421, 926
Paranjape A., Lam T. Y., Sheth R. K., 2012, MNRAS, 420, 1648
Paranjape A., Sheth R. K., 2012, MNRAS, 426, 2789
Peebles P., 1993, Principles of Physical Cosmology. Princeton Univ. Press, Princeton, NJ
Pisani A., Lavaux G., Sutter P. M., Wandelt B. D., 2014, MNRAS, 443, 3238
Pisani A., Sutter P. M., Hamaus N., Alizadeh E., Biswas R., Wandelt B. D., Hirata C. M., 2015, ArXiv e-prints, 1503.07690
Planck Collaboration et al., 2015, ArXiv e-prints, 1502.01595
Platen E., van de Weygaert R., Jones B. J., 2007, MNRAS, 380, 551
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
Ryden B. S., 1995, ApJ, 452, 25
Schaap W. E., van de Weygaert R., 2000, A&A, 363, L29
Shandarin S. F., Medvedev M. V., 2014, ArXiv e-prints, 1409.7634
Sheth R. K., 1998, MNRAS, 300, 1057
Sheth R. K., van de Weygaert R., 2004, MNRAS, 350, 517
Sousbie T., 2011, MNRAS, 414, 350
Suto Y., Sato K., Sato H., 1984, Progress of Theoretical Physics, 71, 938
Sutter P., Lavaux G., Wandelt B. D., Weinberg D. H., 2012, ApJ, 761, 44
Sutter P. M. et al., 2015, Astronomy and Computing, 9, 1
Sutter P. M., Lavaux G., Hamaus N., Wandelt B. D., Weinberg D. H., Warren M. S., 2014, MNRAS, 442, 462
Sutter P. M., Pisani A., Wandelt B. D., Weinberg D. H., 2014, MNRAS, 443, 2983
Zehavi I. et al., 2011, ApJ, 736, 59
Zivick P., Sutter P. M., Wandelt B. D., Li B., Lam T. Y., 2014, ArXiv e-prints, 1411.5694