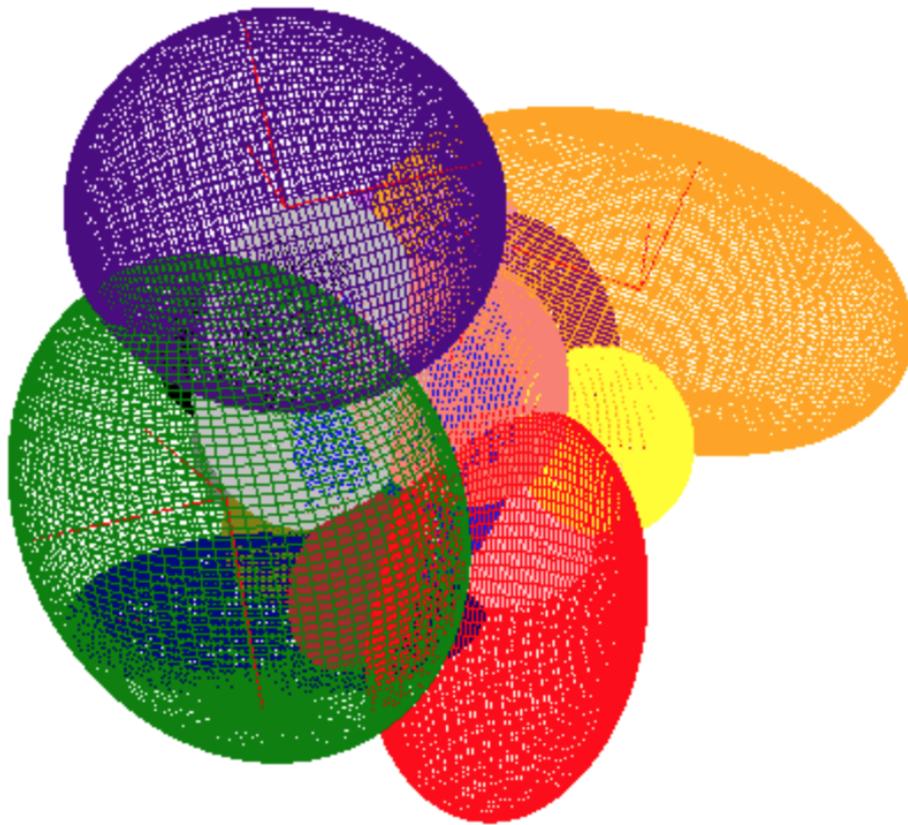

ON THE STRUCTURE, SHAPES, AND ALIGNMENTS OF VOIDS

BACHELOR PROJECT ASTRONOMY 2019



WRITTEN BY

TIMON BEEFTINK

SUPERVISED BY

RIEN VAN DE WEIJGAERT

*University of Groningen
Kapteyn Institute*

If man had no eternal consciousness, if, at the bottom of everything, there were merely a wild, seething force producing everything, both large and trifling, in the storm of dark passions, if the bottomless void that nothing can fill underlay all things, what would life be but despair?

- Søren Kierkegaard

Abstract

On its largest scales, the Universe exhibits a particular structure, known as the cosmic web, in which clusters of galaxies are separated by tenuous voids. Due to their tenuousness, these voids contain valuable information on galaxy formation, the reionisation transition in the early Universe, the nature of dark matter and energy, and the spatial organisation of the cosmic web. In order to model voids, and to subsequently extract cosmological parameters governing processes in the Universe on its largest scales, an Alcock-Paczynski (AP) test is sometimes used, in which - under the assumption that they are not preferentially aligned - voids are identified and bundled in a large volume survey, such that a spherical total void is obtained. Several studies, however, have shown that voids in the large scale structure show alignments originating from tidal forces induced by the inhomogeneous cosmic matter distribution, questioning the validity of the application of the AP test on these scales. This study uses the Watershed Void Finder (WVF) to identify and map voids for an N -body Λ CDM-simulation, both in two and three dimensions, together with their density profiles, such that the shapes, possible alignments, and sphericity of voids can be investigated. Voids are identified and fitted by ellipses (two dimensions) and ellipsoids (three dimensions), together with their neighbouring voids, in order to find correlations in orientation, volume, and shape of neighbouring voids with respect to the central void. It is found that (three-dimensional) voids are prolate ellipsoids, only differing in the significance of their prolateness, that the principal semi-axes of voids are not oriented at random, i.e. exhibit a significant alignment, and that more voluminous voids tend to be aligned more substantially. These results imply that it is dubitably to assume that, on the largest scales, voids do not show any preferential alignments, i.e. to use the AP test - on a bundle of voids in a large volume survey - to describe voids properly, since they exhibit a significant (volume-dependent) alignment and tend to be prolate spheroids. This study asks for redefining the assumptions on which most studies in the field of voids are based on and shows that it is probably unjustified to simply use the AP test and spherical approximation of voids within the cosmic web.

Contents

1	Introduction	3
2	Identifying Voids: The Watershed Void Finder	6
2.1	Λ CDM-simulation: The Data Cube	7
2.2	Density Cube and Grid	7
2.3	Void Identification	7
3	Void Identification and Visualisation	10
3.1	Voids: 2D	10
3.1.1	Ellipse Fitting of Voids	10
3.1.2	Identify Neighbouring Voids in 2D	13
3.2	Voids: 3D	14
3.2.1	Ellipsoidal Fitting of Voids	15
3.2.2	Identify Neighbouring Voids in 3D	17
3.3	Discontinuities and Periodicities	19
3.3.1	Discontinuous Voids	19
3.3.2	Periodic Voids	19
4	Shapes and Alignments of Voids	21
4.1	Shapes of Voids	21
4.2	Volumes of Neighbouring Voids	23
4.3	Alignments of Neighbouring Voids	25
4.3.1	Centres of Neighbouring Voids	25
4.3.2	Orientations of Principal Semi-axes	27
4.4	Intertwined Correlations	29
5	Density Profiles of Voids	30
5.1	Circular Density Profiles	30
5.1.1	Identify Voids and Corresponding Densities	30
5.1.2	Obtain Density Profile	32
5.2	Shape-adapted Density Profiles	33
5.2.1	Identify Voids and Corresponding Densities	33
5.2.2	Obtain Density Profile	34
5.3	Preliminary Results	35
6	Conclusion and Discussion	38
A	The Covariance Matrix	43

B	Periodic Boundary Conditions	45
C	Random Selections and CDFs	48
D	Alpha shapes and Polygons	49

Introduction

On its largest scales, as is indicated by both observations and N -body simulations, the Universe exhibits a particular structure, known as the cosmic web, in which sheets and filaments of galaxies are separated by tenuous voids, the nodes of these sheets and filaments appearing as superclusters. (Chincarini & Rood, 1975; Gregory & Thompson, 1978; Einasto, J  eveer, & Saar, 1980; De Lapparent, Geller, & Huchra, 1986; Kirshner, Oemler Jr, Schechter, & Sackett, 1987; Colless et al., 2001; Tegmark et al., 2004; Guzzo et al., 2013) In the galaxy survey shown in Figure 1.1, these structures are evident as blue concentrations (galaxies) and white planes (voids).

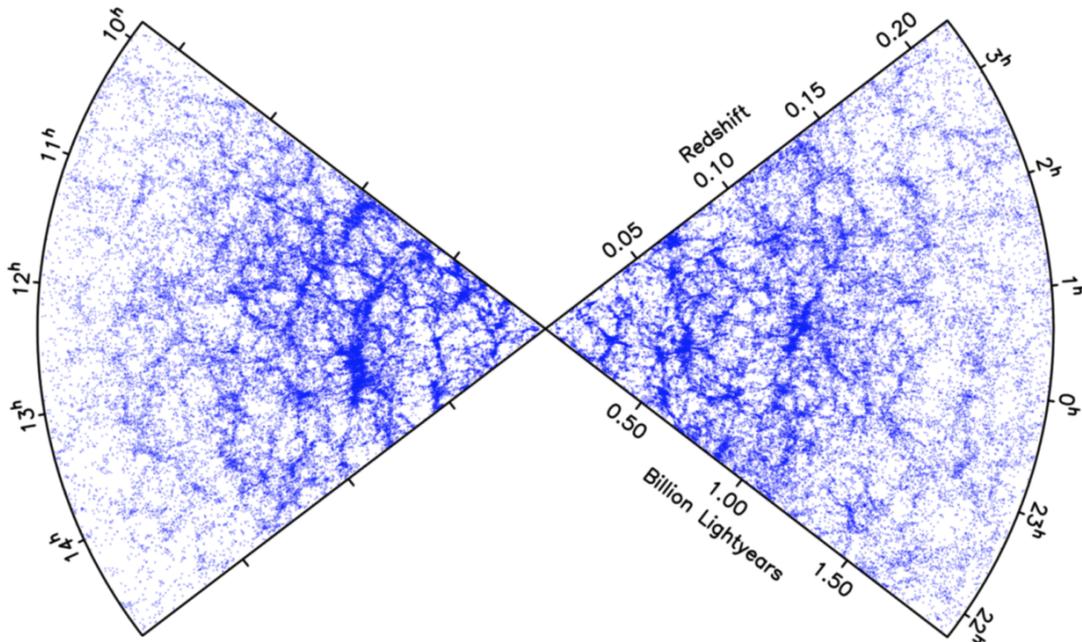


Figure 1.1: Map showing the distribution of nearly 250 000 galaxies at different redshifts as obtained by the 2dF Galaxy Redshift Survey. (Colless et al., 2001) The white, tenuous regions are referred to as voids.

Voids - having sizes of $20 - 50h^{-1}$ Mpc, thus, are a prominent aspect of the cosmic web, encompassing the largest parts of its volume and shaping its mass distribution and structure by means of tidal forces induced by the inhomogeneous cosmic matter

distribution on large scales, and therefore contain valuable information on the formation and evolution of the structure of the Universe on its largest scales. (Bond, Kofman, & Pogosyan, 1996; Park & Lee, 2007; Platen, van de Weygaert, & Jones, 2008; Chan, Hamaus, & Desjacques, 2014) Voids can therefore be used as probes for cosmological parameters, since it is realised that their structure, morphology, and dynamics reflects the nature of dark matter and energy, as well as the possibly non-Gaussian nature of the primordial perturbation field. (van de Weygaert, 2014) In addition, due to their emptiness, voids form the principal regions along which the ionising radiation produced by the first stars in the Universe propagates, marking the significance of voids for describing the Universe on its largest scales. Such a void-based description, used to model the spatial matter distribution at a specific cosmic epoch, illustrates the cosmic structure formation reaching a non-linear stage of evolution, resulting in an all-pervasive network. (Icke, 1984; Sheth & van de Weygaert, 2004; Aragón-Calvo, Platen, van de Weygaert, & Szalay, 2010) Hence, voids can be used to provide information for numerous cosmological questions. Here, we summarise four. (van de Weygaert, 2014)

- Voids form a prominent aspect of the Universe on its largest scales and are crucial for understanding the spatial organisation of the cosmic web. (Icke, Weygaert, et al., 1987; Bond et al., 1996; Sheth & van de Weygaert, 2004)
- Voids are one of the cleanest measures of dark matter and energy, and can therefore be used to test possible modifications of gravity as described by General Relativity. (Suto, Sato, & Sato, 1984; Bolejko, Krasinski, & Hellaby, 2005; Mota, Shaw, & Silk, 2008; February, Larena, Smith, & Clarkson, 2010)
- Voids, due to their tenuousness, are ideal probes for studying galaxy formation and the influence of cosmic environment on the formation of galaxies. (Furlanetto & Piran, 2006; van de Weygaert et al., 2011)
- Voids play a significant role in the key reionisation transition in the early Universe. (Furlanetto, Oh, & Briggs, 2006; Morales & Wyithe, 2010)

These characteristics underline the importance of voids when studying the Megaparsec Universe, since their influences cannot be ignored in order to model the cosmic matter distribution properly. It is, however, difficult to obtain a general model describing voids, due to their diversity in shapes, orientations, sizes, and matter content. A commonly used method to endeavour to probe physical processes governing the formation and evolution of voids, and subsequently extract cosmological parameters describing the Universe in general, is the Alcock-Paczynski (AP) test, which finds the relation between the angular size and radial extent in redshift space of an 'object' expanding with cosmological flow. (Alcock & Paczyński, 1979) Doing so, the voids' total contents can be investigated by stacking voids in a large volume survey, resulting in a spherical total void. (El-Ad & Piran, 1997; Lavaux & Wandelt, 2012; Pisani, Lavaux, Sutter, & Wandelt, 2014) The AP test, however, assumes that objects are not preferentially aligned either along or perpendicular to the line-of-sight, whereas several studies have shown that a notable aspect of voids within the cosmic web is their alignment on reasonably large scales, for which there appears to be a correlation between the orientation of the voids, as well as the sizes of their principal semi-axes (which deviate slightly from spherical), originating from the cosmic tidal field, marking the dubitability of the use of the AP test to describe (the large scale structure of) voids within the cosmic web. (Bond et al., 1996; Platen et al., 2008)

In order to check the validity of the AP test for voids within a large N -body Λ CDM-simulation, the shapes and alignments of voids within such a survey should be investigated. The structure of voids is mainly modelled using computer models, which are generally based on the topological concept of the watershed transform. The analysis of voids is complicated by the fact that there does not exist a proper definition of a void, causing the algorithms to handle different definitions and therefore yielding different results. Examples of commonly used methods are the WVF (Watershed Void Finder) (Platen, van de Weygaert, & Jones, 2007), ZOBOV (ZOnes Bordering On Voidness) (Neyrinck, 2008), and VIDE (Void IDentification and Examination) (Sutter et al., 2015). In order to investigate the alignments of shapes and orientations of voids, as well as their density fields and profiles, such that the justifiability of the assumption of sphericity for voids can be investigated, the WVF is used, since this identification method does not rely on any predefined shape of the voids within a provided data set. (Platen et al., 2007) The WVF thus can be used to identify voids within a data set (Chapter 2), such that the results can be analysed and used for investigating the shapes and orientations of the voids (Chapter 3 & 4), as well as their density profiles (Chapter 5). Using these shapes, orientations, and profiles of voids, possible alignments and correlations can be investigated and discussed, and the validity of the use of the AP test to describe voids on the largest scales can be investigated (Chapter 6).

Identifying Voids: The Watershed Void Finder

The Watershed Void Finder is based on the Watershed Transform (WST), which is commonly used for imaging processing purposes, and, as its name already suggests, is based on an analogue with landscapes, in which maxima, i.e. mountaintops, mark a boundary between the distribution of water in the valley, i.e. a watershed. (Platen et al., 2007) The WST algorithm starts at the minima of the data points within a provided set, which, in case of a density file, are defined as the lowest-density points. A point is referred to as the lowest-density point when it is solely surrounded by points with higher densities. By flooding the profile from its minimum, a basin starts to develop, which grows by subsequently raising the water level, eventually meeting another basin at the ridges of the field, corresponding to the saddle points in the density field. These points mark the boundary between the basins. In the context of the cosmic web, these basins represent the voids, whereas the ridges represent the clusters of galaxies. (Platen et al., 2007) The procedure is visualised below.

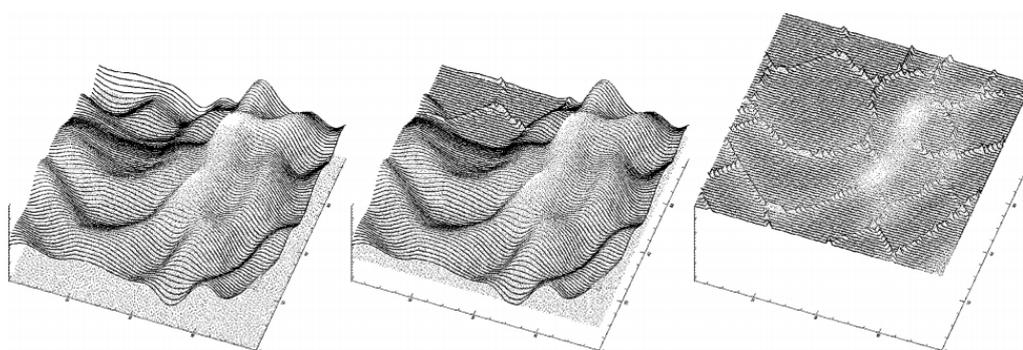


Figure 2.1: Procedure of the WVF, starting at a landscape with water on a ground level (left), which is flooded according to the minima of the landscape (middle), eventually resulting in a field containing ridges and basins (right). (Platen et al., 2007)

2.1 Λ CDM-simulation: The Data Cube

In order to use the WVF-procedure, a data file must be supplied. In general, models and simulations within the research field of cosmology use a Λ CDM-model - commonly referred to as the standard model, which parametrises a Universe containing three major components: (1) ordinary matter, (2) a cosmological constant, and (3) cold dark matter. The model is a product of decades of intensive research and reproduces prominent elements of the Universe, such as the existence and structure of the Cosmic Microwave Background (CMB), the distribution of galaxies, and the abundances of elements, reasonably well. (Smoot et al., 1992; Spergel et al., 2003; Ade et al., 2014) In order to investigate the presence and prominence of voids in the cosmic web, i.e. in the Universe on its largest scales, the Λ CDM-model acts as a basis for simulating a data set in which each point represents the density-value at a specific location in the simulated Universe. In the case of this project, a data file originating from a Λ CDM-simulation, with dimensions 512 x 512 x 512, was taken. The data cube contains unique density values at every spatial coordinate, specified by two spatial (x and y) and one 'redshift'-component (z), i.e. the data cube consists of the two-dimensional distribution of density values for a specific region of the sky, which are referred to as the x - and y -coordinates of the data point, and a one-dimensional distribution at different timescales, which depend on the redshift of the component, and is referred to as the z -coordinate, together constituting a three-dimensional data set. This data cube forms the basis for all results obtained in this research.

2.2 Density Cube and Grid

The Λ CDM data set can easily be used to obtain a density cube, in which all density values at every unique spatial coordinate are given. In order to get a grasp of the shape and content of the density cube, it is visualised in a (two-dimensional) grid by slicing the data set into 512 parts, such that the distribution of density values at a specific redshift can be found. By taking $z = 0$, the density profile shown in Figure 2.2 is obtained. This figure shows different colours, highlighting the different density values of the points within the two-dimensional data grid. As can be observed, the grid shows (greenish) high-density regions, corresponding to clustered galaxies in the cosmic web, as well as (blueish) low-density regions, corresponding to the voids within the cosmic web. The shape of the density map highly mimics the distribution of galaxies at different redshifts shown in Figure 1.1, i.e. concentrations of points are separated by tenuous regions, implicating that the simulated Λ CDM profile matches the actual structure and morphology of the Universe quite vigorously.

2.3 Void Identification

The density field as shown in Figure 2.2 acts as a starting point for the WVF, which, as explained in the introduction of this chapter, takes the lowest-density points as minima, and subsequently floods the 'landscape' in order to obtain a field containing ridges and basins, i.e. borders and voids. Technically, the WVF algorithm uses the

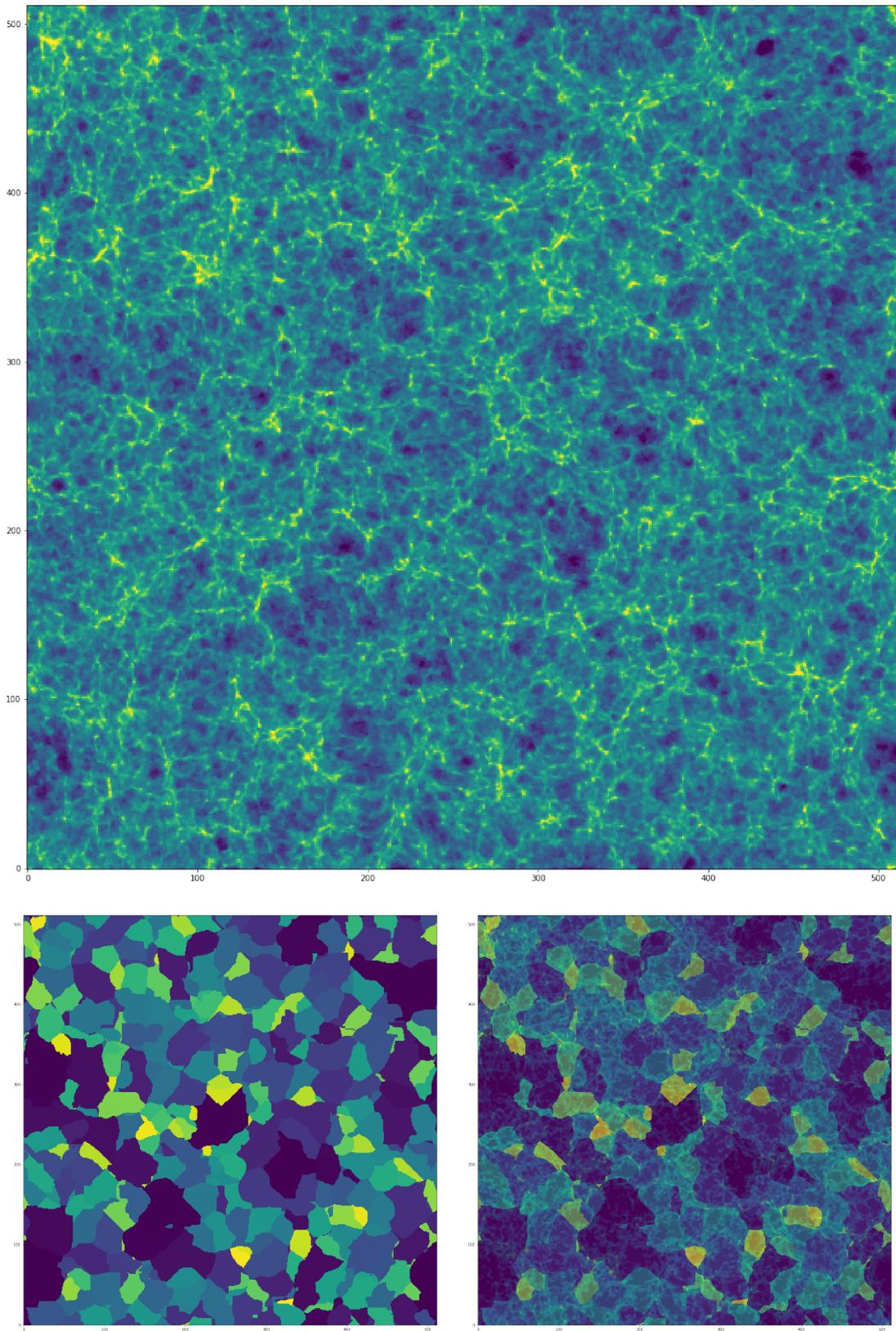


Figure 2.2: Two dimensional grids of density values (top) and corresponding voids (left), as well as their overlapping grids (right), as obtained by the WVF, shown for a slice in a specific z -direction ($z = 0$).

following steps in order to identify voids within a provided density field. (Platen et al., 2007)

1. Delaunay Triangulation Field Estimator (DFTE) is used to define a continuous density field throughout the sample volume, which is smoothed using a Gaussian kernel and discretised into a finite set of density contour levels.
 2. The local minima, i.e. those pixels that are solely surrounded by pixels with a higher density value, are found.
 3. Starting from the local minima, the density contour levels are continuously increased, for which pixels with a density value lower than the flooding threshold are added to the basin of a particular minimum. As soon as a pixel is reached by two different basins, the pixel is identified as belonging to the boundary between the corresponding basins.
-

By applying the WVF on the sliced density file, the two-dimensional grid showing the different voids within the data set in Figure 2.2 is obtained. The file resulting from the WVF is a 512 x 512 x 512 data cube, in which every pixel is assigned a specific void ID, which simply is a number ranging from 1 to - depending on the size of the data set, as well as the density-cutoff - about 3000. By finding all pixels that have the same number, i.e. correspond to the same void, all voids identified within the data set can be found and visualised. The voids are distinguished visually by using different colours¹, such that their shapes and positions can be distinguished. As can be observed, the (two-dimensional) voids differ significantly in shapes and orientations.

This two-dimensional grid, showing the positions and structure of voids within the data set, is used as a starting point for investigating the shapes, orientations, and possible alignments of voids, since it enables to isolate voids and determine their properties, as well as the properties of their neighbouring voids. This is done in the following sections.

¹The different colours of the voids shown in Figure 2.2 do not reflect any physical quantity, but depend only on their void ID, and are merely used to distinguish between the voids within the data set. Please note that some neighbouring voids have similar colours, impeding the process of distinguishing voids visually.

Void Identification and Visualisation

As is shown in the preceding chapter, and in particular in Figure 2.2, the WVF can be used to identify, separate, and visualise voids within a density data set. Whereas the visualisation was limited to two dimensions, it can be extended to three dimensions as well. In this chapter, the groundwork for visualising, identifying, and analysing (neighbouring) voids is laid, such that the general shapes and alignments of the voids can eventually be found in Chapter 4. In order to describe the voids' shapes, (two-dimensional) ellipse fitting (Section 3.1) and (three-dimensional) ellipsoid fitting procedures (Section 3.2) are established, whereas the neighbours are identified by using an N -dimensional 'Neighbouring Void Finder', discussed in both Section 3.1 and 3.2. The discontinuities and periodicities occurring within the data cubes and grids are discussed and analysed in Section 3.3. Using the obtained algorithms and results, the characteristics of both two- and three-dimensional voids can be obtained and analysed, such that their shapes and alignments can be investigated, and the validity of the AP test on large volume surveys can be explored.

3.1 Voids: 2D

By using the Python-module *imshow*, the data file as obtained by the WVF, containing the void IDs of all pixels within the 512 x 512 x 512 data cube, can be visualised, as is done in Figure 2.2. Here, a particular slice of redshift $z = 0$ is taken, whereas, actually, voids are three-dimensional objects encompassing a considerable chunk of space. The configuration for such a z -slice only reflects the shapes of voids for a particular moment in time, without providing information on the entire shape of the void, which generally deviates for different slices in the z -direction significantly, as is discussed in Section 3.2 and Figure 3.4. A two-dimensional void grid, however, is an ideal start for investigating the algorithms used to describe and fit voids (Section 3.1.1), and to find the neighbouring voids for a specific central void (Section 3.1.2).

3.1.1 Ellipse Fitting of Voids

The first step in describing the behaviour of voids is to approximate their shapes as ellipses, such that (in two dimensions) their semi-minor and -major axis, orientation,

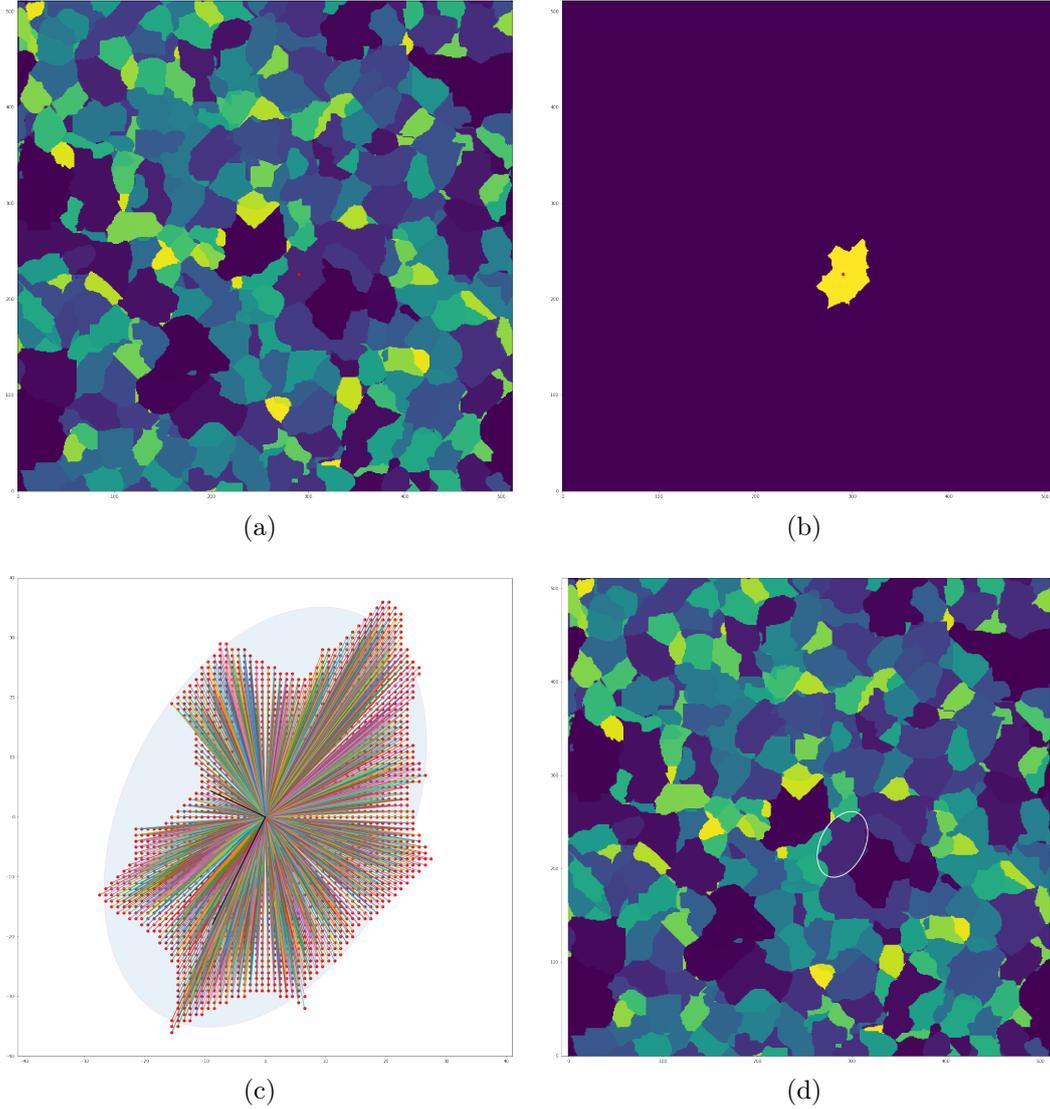


Figure 3.1: Example of ellipse fitting for void 333, in which a) the centre of the void is identified, b) the void is isolated, c) the void is relocated and fitted by an ellipse, and d) the ellipse is plotted in the void profile.

shape, and area can be distinguished and compared. In order to fit the voids within the data set by ellipses, the following steps are taken:

1. Identify all points corresponding to a given void by taking a specific void ID and determine its spatial centre by taking the mean of the pixels in the x - and y -directions.
2. Isolate the void from its surroundings and relocate to the origin.
3. Plot vector positions to all points within the void (with respect to the void's centre) and subsequently fit a covariance ellipse¹, in which the ellipse's area is corrected by taking the actual area of the void into account, such that the area of the covariance ellipse matches the actual area of the void.

¹For more information on the covariance matrix and ellipse, please see Appendix A.

4. Determine the semi-major (a) and semi-minor axis (b) of the ellipse by using the eigenvalues of the covariance matrix. The corresponding eigenvectors are used to define the orientation of the ellipse.
5. Plot the resulting ellipse in the original void profile.

An example of the fit-procedure described above is given in Figure 3.1, in which a well-defined void (with void ID 333), i.e. a void without discontinuities due to boundary conditions or spatial gaps, is fitted by an ellipse. This procedure can, by taking periodic boundary conditions into account, which prevent the algorithm to define the voids on the edges of the data set incorrectly, be generalised in order to fit all voids by ellipses, such that the area and orientation of voids, as well as their structural correlations, can be quantised and compared.² By repeating the procedure for different well-defined voids, i.e. voids without discontinuities due to boundaries or gaps, within the data set, the profile shown in Figure 3.2 is obtained, displaying the shapes and orientations of single voids.

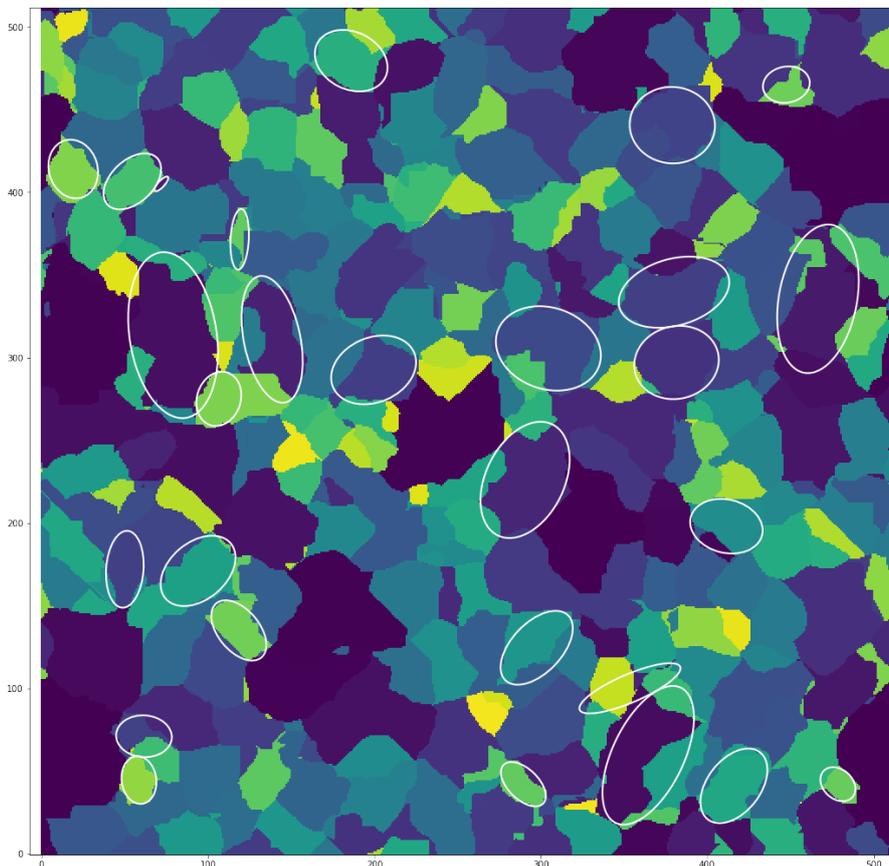


Figure 3.2: Void profile resulting from ellipse fitting for voids on different spatial positions, showing their shapes and orientations for a specific slice ($z = 0$) of the Λ CDM data file.³

²More information on periodic boundary conditions - for both two- and three-dimensional data sets - can be found in Section 3.3 and Appendix B.

³Sometimes, it just looks like that an ellipse does not encompass the entire void. This can have two causes: 1) a neighbouring void has a very similar colour, or 2) the void contains irregularities or discontinuities.

3.1.2 Identify Neighbouring Voids in 2D

The neighbouring voids of a given central void can be identified in order to deduce whether significant preferential alignments and correlations occur within the simulated cosmic web, and whether it is justifiable to use the AP test to describe their behaviour. By implementing a 'Neighbouring Void Finder' (NVF) in Python, the voids at the border of a specific central void can be found. ([user3717023, 2016](#))

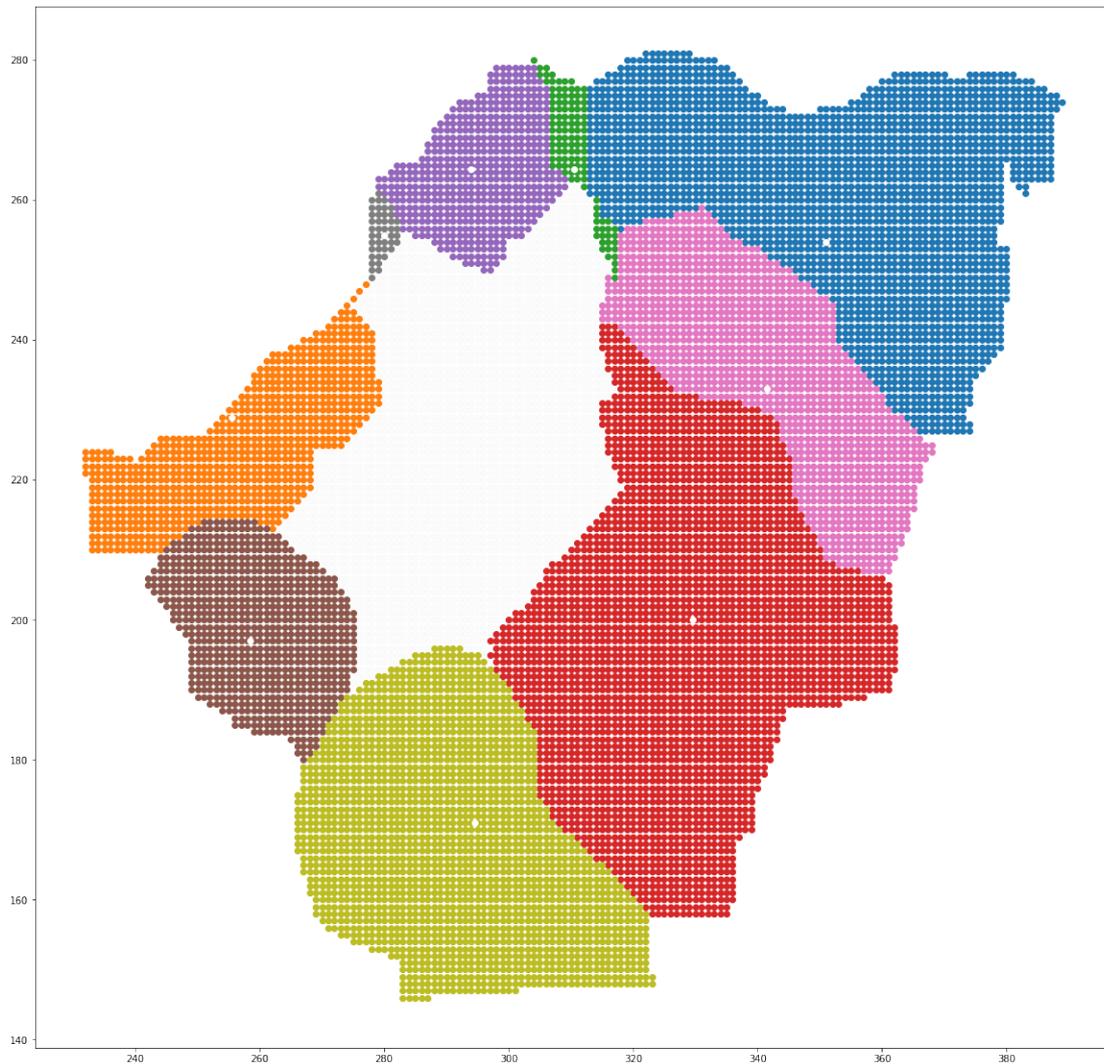


Figure 3.3: Example plot showing the neighbours (in different colours) - and their spatial centres (white dots) - of a specific central void with ID 333 (white).

The procedure follows the following steps:

1. Slice a 512 x 512 two-dimensional data list from the three-dimensional data cube, and identify all pixels corresponding to a specific void by taking all points matching the ID of this void.
2. Identify the centre of the specific void and use the centre as a starting point for processing the list containing all void IDs by moving up, down, to the left and right simultaneously, and identify the first encountered IDs that differ from the specific (central) void ID.

3. Find all points corresponding to these first encountered IDs, which are the neighbouring voids of the central void, and append these points to a new list.
4. Plot the new list, encompassing all neighbouring voids, together with their spatial centres, and the central void in a single figure.

The result of this procedure for a specific void (with ID 333) is shown in Figure 3.3. As the neighbouring voids are identified, the ellipse fitting procedure can be applied in order to obtain the angular orientation and area of the voids, which subsequently can be compared to the orientation and area of the centring void, such that possible correlations can be established and investigated. However, since voids are actually three-dimensional objects, whereas in the procedures above only two-dimensional data was taken into account, one has to shift to three-dimensional data in order to be able to investigate these correlations properly.

3.2 Voids: 3D

In order to investigate voids properly, their three-dimensional shape must be investigated, rather than the two-dimensional data grids shown in the preceding section. By taking the same data set, and by reducing its size to 256 x 256 x 256 (such that the processing time is reduced significantly), the voids can be identified in three dimensions, termed x , y , and z .

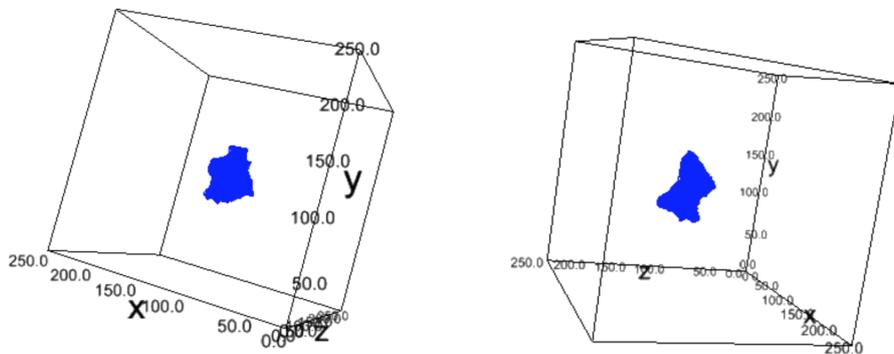


Figure 3.4: Three-dimensional representation of a single void (with ID 1111), using *ipyvolume*, shown at different viewing angles.

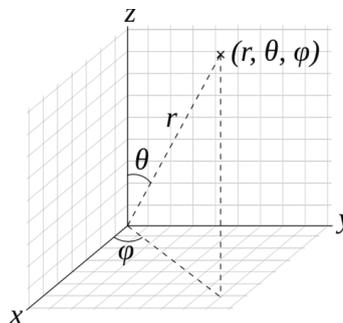
In general, it is rather complicated to visualise three-dimensional arrays in Python, especially when using *mplot3d* as provided by the *matplotlib* package. Luckily, a relatively simple module, called *ipyvolume*, was developed in order to visualise three-dimensional data in interactive plots, such that the data points and orientation of voids can easily be investigated. (Breddels, 2017) An example of a void (with ID 1111) in three dimensions is shown in Figure 3.4. As can be seen from (the right-hand side of) Figure 3.4, the void's shape differs significantly for slices in the z -direction, which were used in Section 3.1, marking that the data used and the ellipses subsequently fitted in Figure 3.2 were merely an investigation of the data, but did not represent useful, quantitative information for investigating the area

(in two dimensions) and volume (in three dimensions), angular orientation, and alignments of the voids, highlighting the necessity of identifying and analysing the voids in three dimensions by using an ellipsoidal fit procedure (Section 3.2.1) and the NVF (Section 3.2.2).

3.2.1 Ellipsoidal Fitting of Voids

In three dimensions, a void can, analogous to the two-dimensional case, be fitted by a (covariance) ellipsoid, which can be used as a measure of its volume, orientation, and shape. The ellipsoid's orientation is characterised using the spherical coordinates (r, θ, φ) , whereas its shape is defined by its principal semi-axes a , b , and c , forced to satisfy the condition $a > b > c$, such that the voids' shapes can be compared easily. The relation between spherical and cartesian coordinates, as is used in processing the data in Python, is shown below.⁴ (Tatum, 2019)

$$\begin{aligned}x &= r \sin \theta \cos \varphi \\y &= r \sin \theta \sin \varphi \\z &= r \cos \theta\end{aligned}$$



The procedure used for fitting an ellipsoid to a three-dimensional void is as follows:

1. Identify all points corresponding to a given void by taking a specific void ID and determine its spatial centre by taking the mean of the pixels in the x -, y -, and z -directions.
2. Isolate the void from its surroundings and relocate to the origin.
3. Compute the covariance matrix of data points and isolate eigenvectors and eigenvalues, and subsequently obtain the orientation and principal semi-axes of the ellipsoid. (Armstrong, 2018)
4. Correct the volume of the ellipsoid by the actual volume of the void and adapt the magnitudes of the principal semi-axes such that the ellipsoid's volume matches the volume of the void.
5. Parametrise the ellipsoid by computing its cartesian coordinates, and subsequently flatten its coordinates, such that the ellipsoid's rotation can be vectorised and visualised. (Armstrong, 2018)
6. Relocate the void to its original position by adding the mean values found in step 1 and plot the resulting wireframe using *ipyvolume*.

⁴In order to map all points, the ranges of θ and ϕ are often taken to be $[0, \pi]$ and $[0, 2\pi]$, respectively.

The result of this procedure for a specific void (with ID 1111) is shown in Figure 3.5, in which the void in Figure 3.4 (shown in cyan) is fitted by an ellipsoid (shown in blue). Since a confidence interval of 95% is taken for fitting the covariance ellipsoid, not all points of the void are encompassed by the ellipsoid, as can be observed from the figure. However, the ellipsoid matches the shape and orientation of the void reasonably well, and can therefore easily be used for quantifying its characteristics.

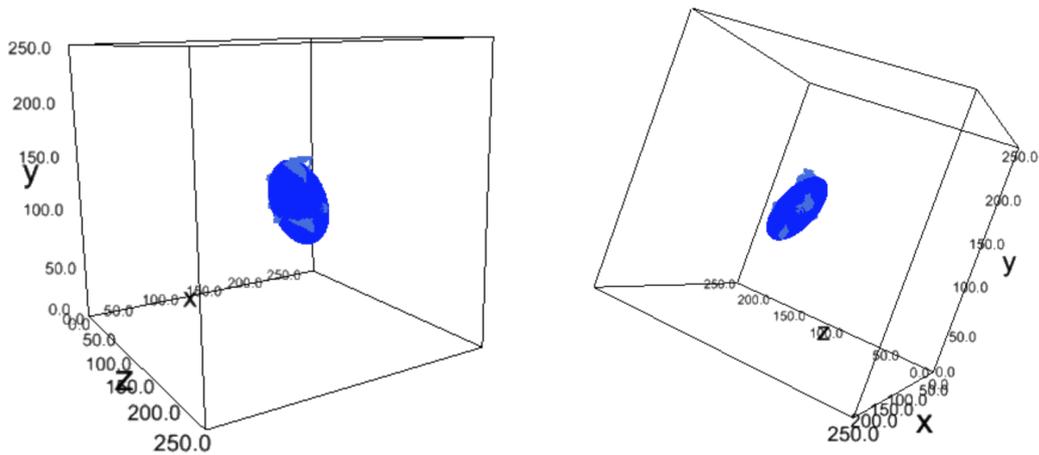


Figure 3.5: Ellipsoid fit procedure (blue) for a specific void 1111 (cyan), shown at different viewing angles.

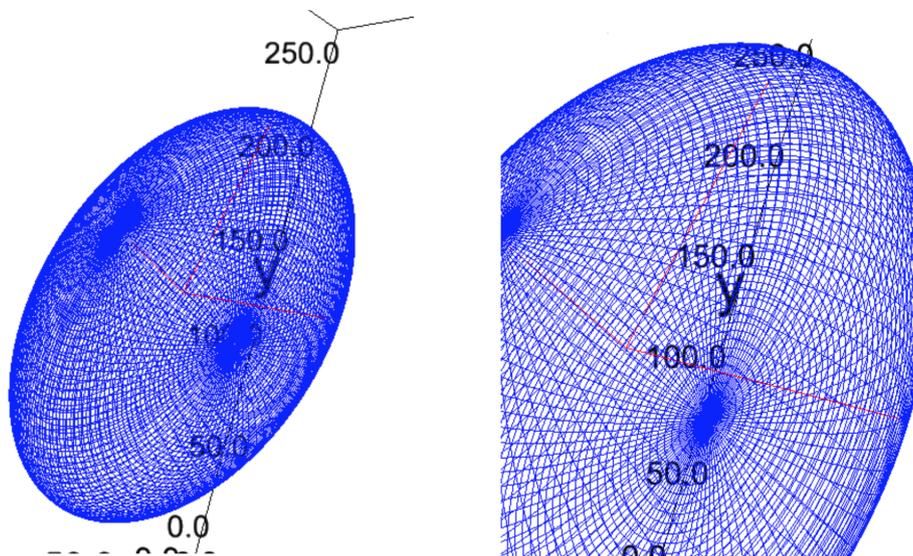


Figure 3.6: Close-up of fitted covariance ellipsoid shown in Figure 3.5, in which the principal semi-axes a , b , and c are marked red.

In a close-up, the principal semi-axes, as defined by the eigenvectors (orientation) and eigenvalues (magnitude), can be shown as well. The result is shown in Figure 3.6. These principal semi-axes actually define the position, orientation, and size of the void, and can therefore easily be used to compare voids in terms of shapes and alignments.

3.2.2 Identify Neighbouring Voids in 3D

Since the NVF, as described in section 3.1.2, can be used for any N -dimensional data set, it can be used in three dimensions as well, such that identifying and plotting the neighbouring voids of a three-dimensional central void is easily implemented. The result for the void shown in Figure 3.4, i.e. void 1111, is visualised in Figure 3.7, again by using *ipyvolume*.

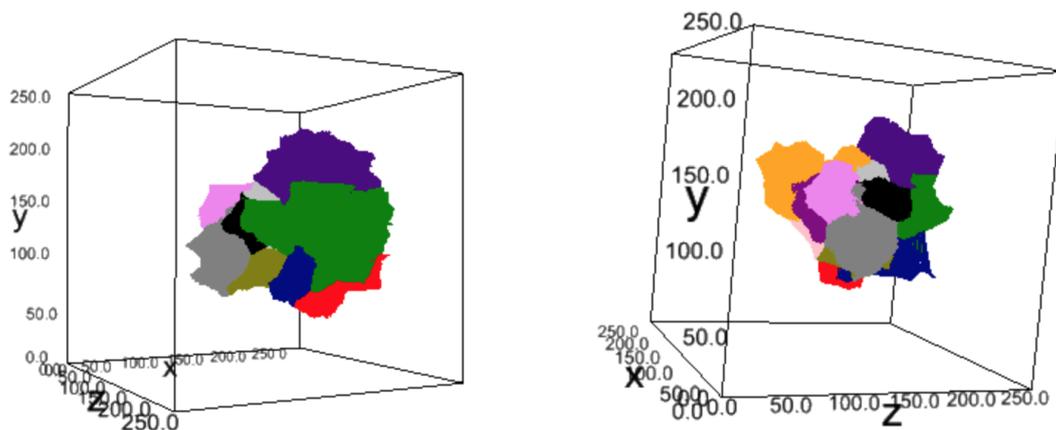


Figure 3.7: Neighbouring voids - all marked by a specific, random colour - of a central void (with ID 1111), shown in three dimensions at different viewing angles.

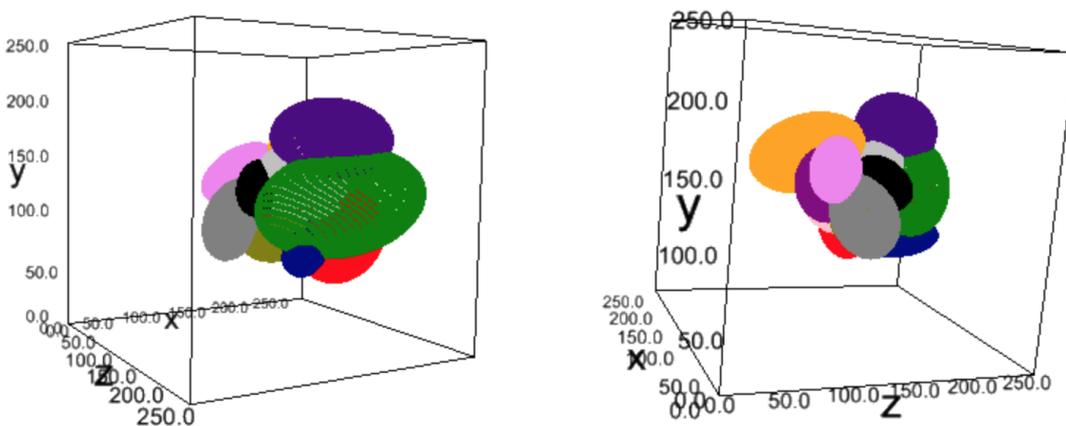


Figure 3.8: Ellipsoid fit procedure for the neighbouring voids of a specific central void (with ID 1111), shown in three dimensions at different viewing angles. The random colours, again, distinguish between the different voids.

Subsequently, the ellipsoid fit procedure, established in section 3.2.1, can be used to fit the neighbouring voids as well, such that both the central and neighbouring voids can be represented by ellipsoids, showing an idealised configuration in which the volumes, orientations, and axes of the different voids can be compared and correlated. The result of the ellipsoid fit procedure for all neighbouring voids (shown in Figure 3.7) for void 1111 is shown in Figure 3.8. As can be observed from Figure 3.7 and 3.8, the shapes and orientations of the ellipsoids match the shape

and orientation of the corresponding voids excellently. Similar to Figure 3.6, the neighbouring and central ellipsoidal voids can be shown in close-up as well, as is done in Figure 3.9, showing the ellipsoids fitted to the neighbouring voids (in different colours) and their corresponding principal semi-axes (in red).

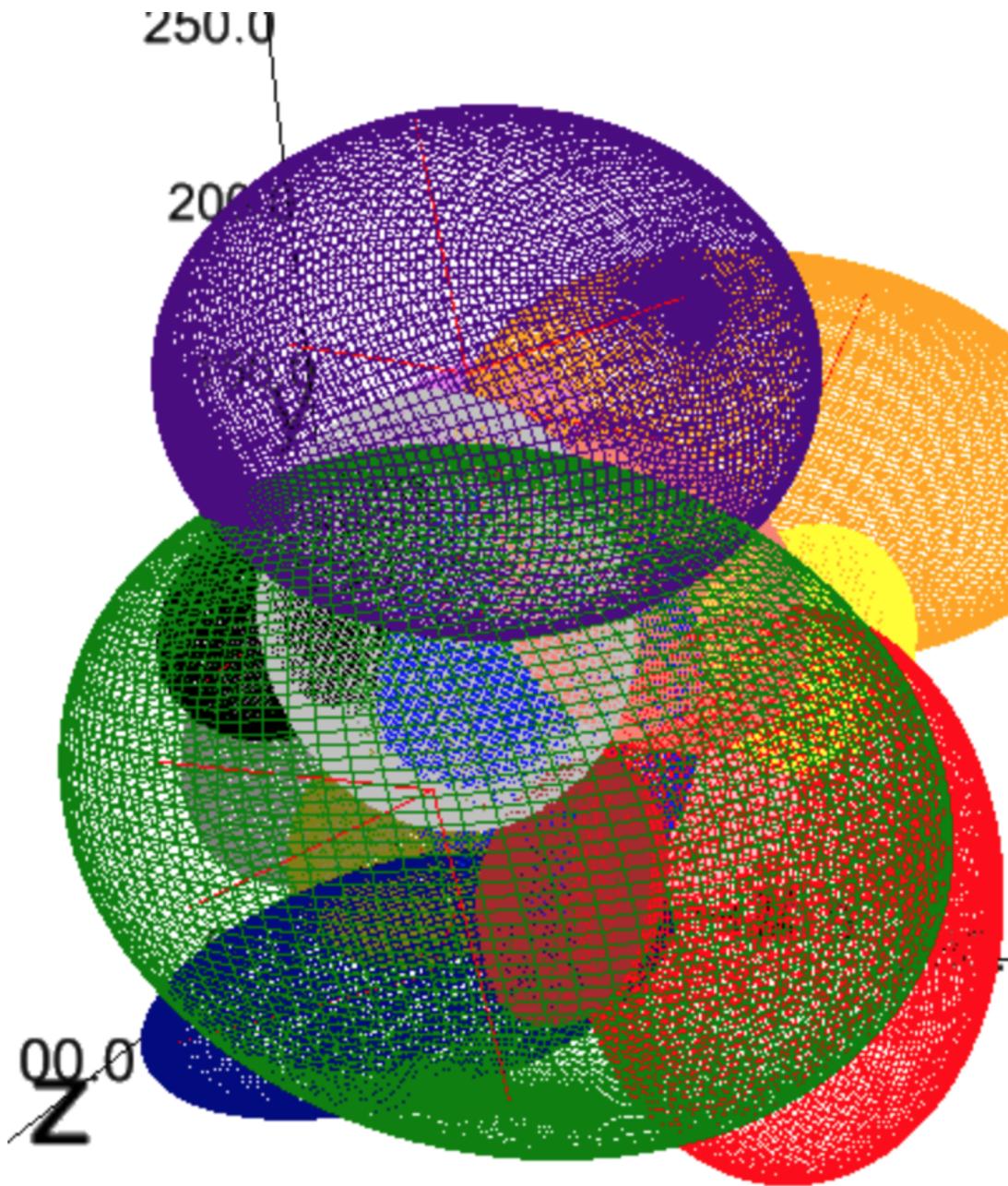


Figure 3.9: Close-up of fitted covariance ellipsoids shown in Figure 3.8, in which the principal semi-axes are marked red.

As can be observed, these ellipsoids all surround (and touch on) the central void, which is still shown in (light)blue. Using the characteristics of these neighbouring voids, with respect to the characteristics of the central void, i.e. their volumes, principal semi-axes, and orientations, possible correlations can be investigated, as is done in Chapter 4.

3.3 Discontinuities and Periodicities

Unfortunately, not all voids within the data set are well-defined, forcing one to be careful when adopting statistics in order to describe the behaviour of voids within the cosmic web. Generally, one can distinguish three types of voids within the void profile:

1. Well-defined void: a void composed of a continuous distribution of points, without any gaps or periodicities. This type of void is used in the analysis up to now.
2. Discontinuous void: a void containing one or more discontinuities, in which the points are separated by at least one gap, primarily occurring in two dimensions.
3. Periodic void: a void at the boundaries of the data grid (or cube), which - due to the periodicity of the data set - continue at the other side of the set.

These types appear both in two and three dimensions and must be taken into account in order to quantify the shapes and alignments of voids properly. Examples of these three cases are shown in Figure 3.10 on the next page, in which, again, a two-dimensional data slice (for $z = 0$) is used.

3.3.1 Discontinuous Voids

Discontinuous voids originate from the WVF, for which a user-supplied density threshold is used to mark the maxima of the basins, i.e. voids, in the cosmic web. The natural threshold value is the typical void underdensity $\Delta = -0.8$. (Platen et al., 2007) This threshold, however, causes some discontinuities within the identified voids, allocating the same void ID to points at different locations. This defect is largely eliminated by taking a larger density threshold, which, however, affects the size and shapes of the voids drastically, whereas a smaller threshold causes even more discontinuities. Therefore, the natural threshold value is taken as a golden mean, for which the number of discontinuous voids is minimised, and the representativeness of voids is maximised, and the small number of discontinuous voids and following systematic errors are taken for granted, as a defect inherent to the void identification procedure. However, as one can imagine, the three-dimensional voids are significantly less discontinuous, since data slices in the z -direction cause the void to be split up at a specific redshift, whereas, in three dimensions, the void is actually well-defined, causing the effect of discontinuous voids on three-dimensional data to be relatively small, and is therefore taken for granted while analysing three-dimensional voids, as is done in Section 4 when finding the shapes and alignments of voids.

3.3.2 Periodic Voids

The $512 \times 512 \times 512$ data cube exhibits periodicity, as can easily be observed from, for example, Figure 3.2, in which voids located at the boundaries of the data set proceed to the left, right, bottom, and/or top. These so-called periodic voids require periodic boundary conditions to be implemented in order to be able to quantify the shapes and orientations of these (three-dimensional) voids. This implementation is

explained and visualised in Appendix B. In the following chapters, these conditions are used to find the shapes and alignments of all voids within the data cube.

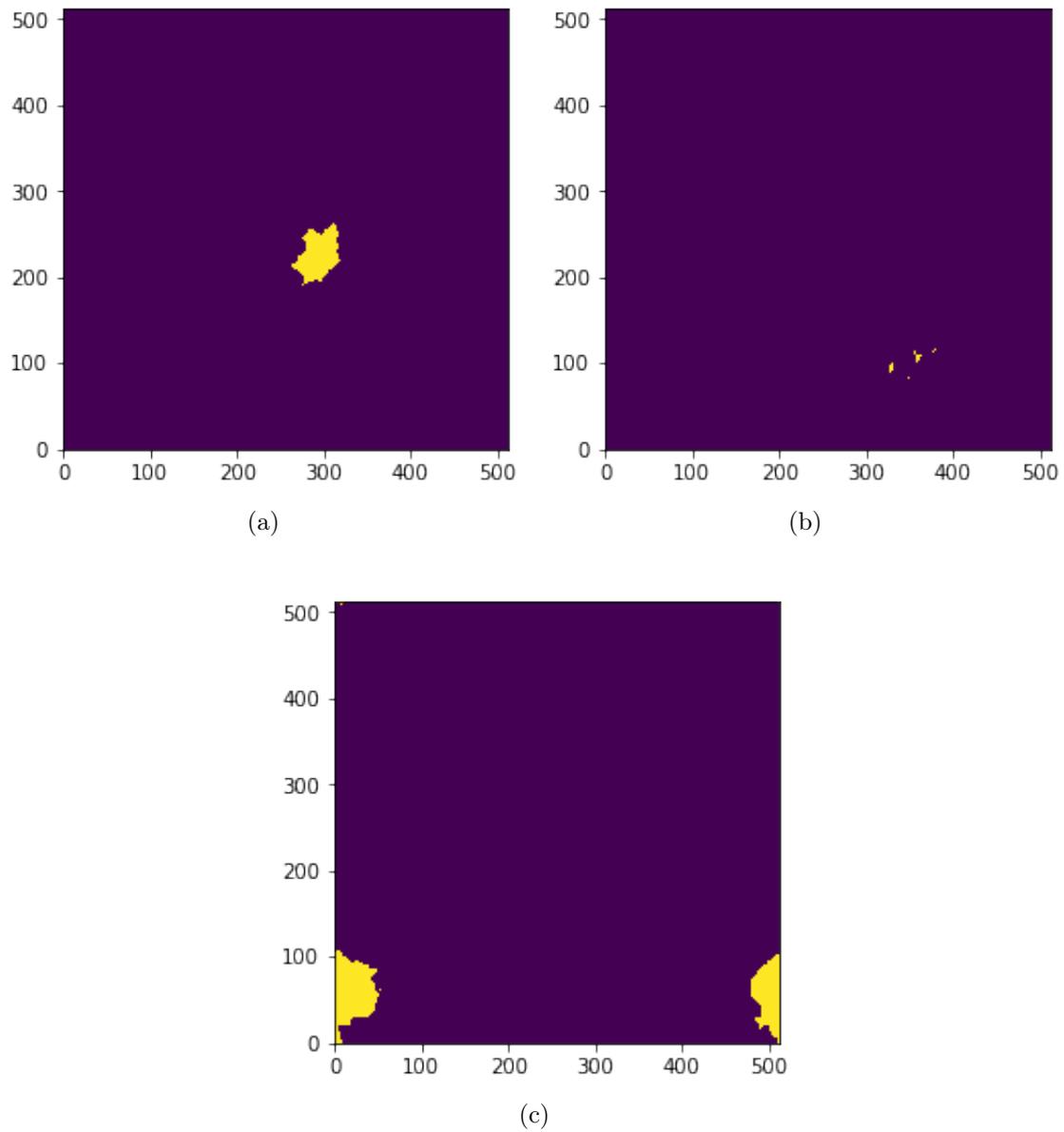


Figure 3.10: Example plot showing a case 1) (left), case 2) (right), and case 3) (bottom) void in a two-dimensional ($z = 0$) data grid.

Shapes and Alignments of Voids

In the preceding chapters, the foundation was laid to interpret, analyse, and visualise - both in two and three dimensions - the data file containing the density values of a $512 \times 512 \times 512$ cube, obtained by an Λ CDM-simulation. These foundations, in particular the (neighbouring) void identification and ellipsoid-fitting procedures, can be used to study the shapes (Section 4.1), volumes (Section 4.2), and alignments of voids (Section 4.3), i.e. the correlations of neighbouring voids in terms of orientations (Section 4.3.1) and principal semi-axes (Section 4.3.2). As stated in Chapter 1, tidal forces, induced by the inhomogeneous cosmic matter distribution on large scales, show an intimate link with the orientation of voids. This orientation-tide alignment remains significant on scales exceeding twice the typical void size, which shows that the long-range external field is responsible for the alignment of voids, indicating that the tidal force field is the main agent for the large-scale spatial organisation of the cosmic web. (Bond et al., 1996; Platen et al., 2008) One thus expects to find significant preferential alignments within the large volume Λ CDM survey used in the preceding sections, which may imply that the assumptions underlying the AP test, and the assumption of a resulting spherical total void, may not be warranted. On the contrary, the absence of any correlation would enable stacking up voids, drastically simplifying the analysis of voids within the cosmic web. Moreover, the presence (or absence) of these correlations forms a basis for the evolution of voids as the Universe expands, and indicates the strength of the orientation-tide alignment within the cosmic web, marking the significance of investigating the shapes and alignments of voids.

4.1 Shapes of Voids

First of all, the general shape of voids within the Λ CDM data cube is investigated, in order to check whether the fitted ellipsoids tend to assume specific forms. In order to do so, the Λ CDM-data file was reduced to a size of $256 \times 256 \times 256$, such that the time required to analyse the results is reduced, as explained in section 3.2.¹ The voids within this data cube, while applying the periodic boundary conditions

¹In general, it took about one hour to compute each individual graph shown in this section. Imagine the time required for a $512 \times 512 \times 512$ data cube...

- see Appendix B², are all fitted using the ellipsoid fit procedure, such that the characteristic shapes of the voids can be obtained and compared. These shapes can be retrieved from their principal semi-axes: a , b , and c , characterising the shape of an ellipsoid. By forcing the condition $a > b > c$, the axes of all ellipsoids can be compared, such that one can investigate whether voids tend to be spherical, oblate, or prolate. Graphically, the shapes of the ellipsoids can be investigated by plotting the oblateness versus the prolateness of the ellipsoid, denoted as p and q , respectively, and given by:

$$p = \frac{b}{a} \quad (4.1)$$

$$q = \frac{c}{a} \quad (4.2)$$

The result for the voids in the 256 x 256 x 256 data cube is shown in Figure 4.1, in which each blue dot represents a single void within the data set.

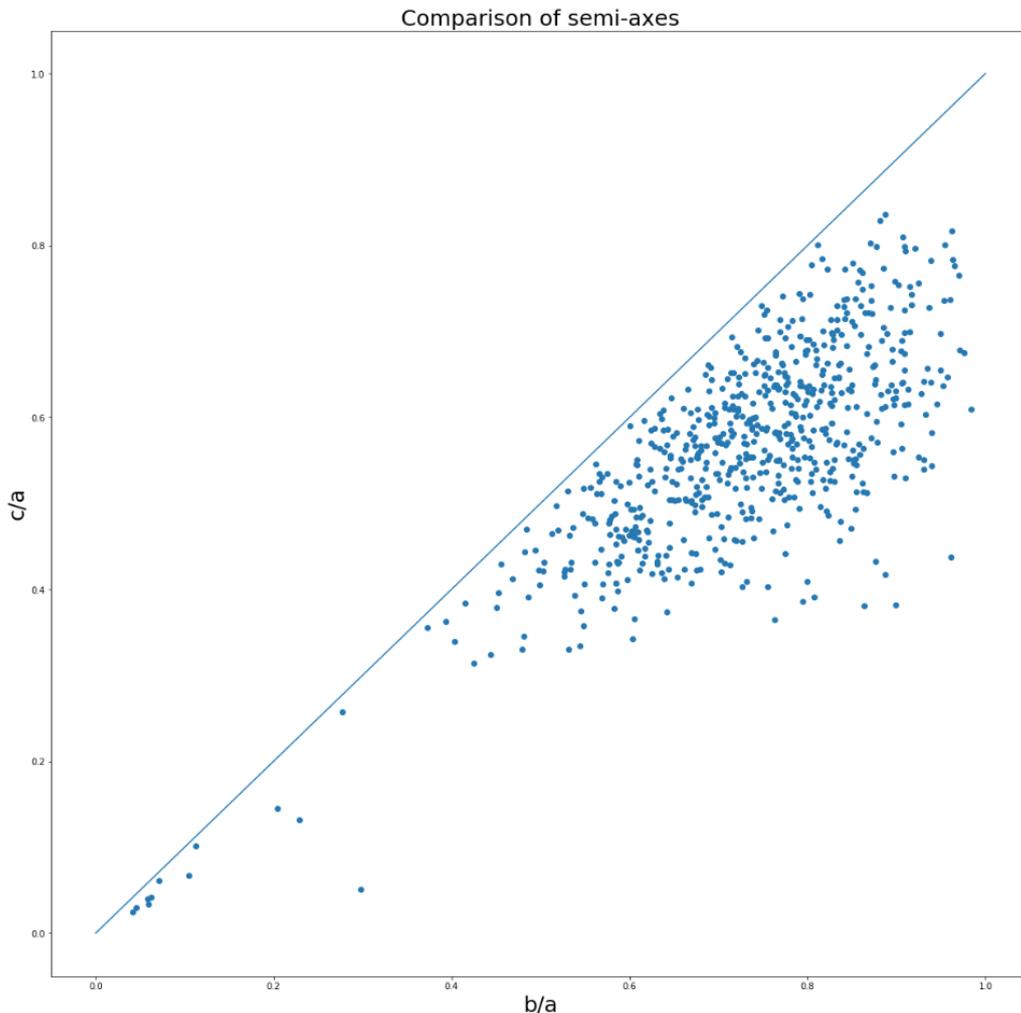


Figure 4.1: Plot of $\frac{b}{a}$ versus $\frac{c}{a}$ for all fitted ellipsoidal voids within the 256 x 256 x 256 data cube.

²The voids within the data cube were fitted and analysed with and without boundary conditions, leading to an interesting result in which the graphs deviate significantly, and from which the necessity of periodic boundary conditions became evident. For more information, please see the appendix.

This plot contains some indications regarding the general shape of voids. As one can infer, the voids tend to be quite spherical, since most points are located near $(\frac{b}{a}, \frac{c}{a}) = (1, 1)$, which represents a sphere. However, it is clear that all data points are more or less located below the spheroid line, indicating that all voids satisfy the condition $\frac{b}{a} < \frac{c}{a}$, i.e. that all voids show some prolateness. (Platen et al., 2008) The only difference is the significance of this prolateness, since some voids are relatively far away from the solid line, indicating that these voids are quite prolate, whereas others are relatively close to the solid line, indicating that these voids are quite spheroidal. Despite the higher concentrations of voids near the solid line and near the coordinate of sphericity, i.e. (1,1), the spread of points indicates a significant (prolate) deviation from spheroidality and sphericity for the voids within the 256 x 256 x 256 data set.

4.2 Volumes of Neighbouring Voids

Besides the general shapes of voids, the correlations in terms of volumes of central and neighbouring voids can be investigated as well by using the NVF. As assumed by the AP test, the voids within a large volume survey can be treated independently, showing no internal relation between their volumes. Several studies, however, indicate that voids show a significant hierarchy, in which a larger void is made up by the mergers of smaller voids, and voids embedded within a larger-scale overdensity get squeezed out of existence as the overdensity collapses around them, suggesting that voids cannot be simply treated separately. (van de Weygaert & van Kampen, 1993; Sheth & van de Weygaert, 2004) By testing whether voids show any preferred volume composition, i.e. whether the volumes of voids are correlated, an indication for whether the application of the AP test on large volume surveys is valid can be obtained.

By using the NVF, and subsequently fit all voids by ellipsoids, the volume of the central and neighbouring voids can be found and related, resulting in a scatter plot representing the volume of the central void (on the x -axis) and the neighbouring void (on the y -axis) for N voids. However, since voids encompass a significant chunk of space, attention must be paid when selecting the voids to be processed by the NVF, since neighbouring voids show systematic correlations; they touch mostly on the same voids, causing the dots in the scatter plot to coincidence on the y -axis, resulting in a non-random distribution of volumes. To prevent these undesired correlations as much as possible, a completely random slice of fifty voids - while taking periodic boundary conditions into account - was taken from the data file containing the void IDs of the voids within the 256 x 256 x 256 data cube. This random selection probably limits the number of neighbouring voids within the selection and acts as a fair balance between time-consuming algorithms and representativeness of the results. Doing so, results in the graph shown in Figure 4.2, showing the volume of the central (x -axis) and neighbouring voids (y -axis), and a solid straight line representing the values for which the volume of the central void is equal to the volume of the neighbouring void, i.e. neighbouring voids above this line are more voluminous than their central void, whereas neighbouring voids below this line have smaller volumes than their central void. All neighbours of a specific central void are marked with a specific colour, such that the different central voids can easily be distinguished.

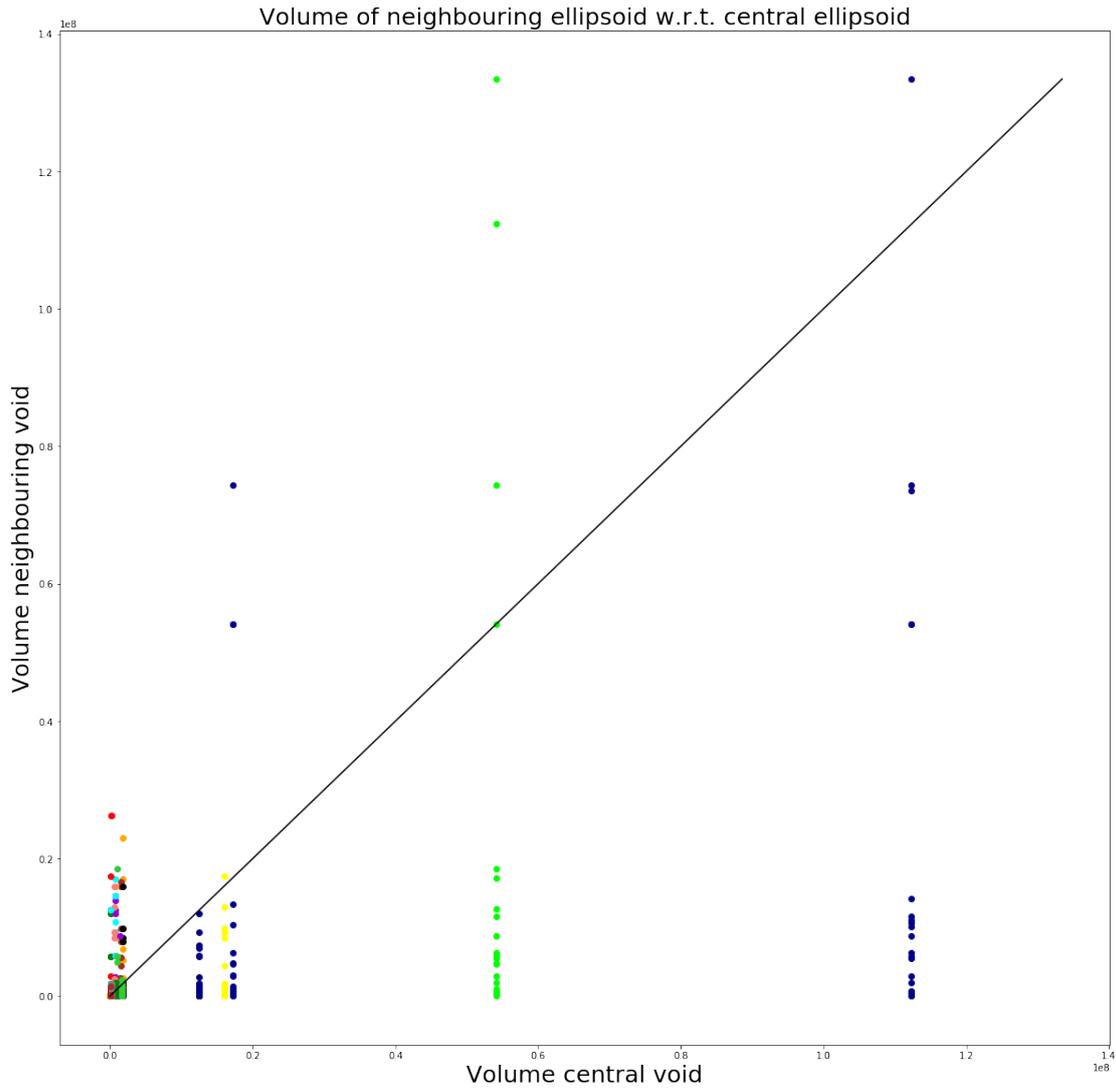


Figure 4.2: Distribution plot of the volumes of neighbouring voids with respect to the specific central void for a random selection of fifty voids within the $256 \times 256 \times 256$ data cube. The straight, solid line represents equal volumes for neighbouring and central voids.

As can be observed from the distribution, the smaller voids, i.e. the voids located on the left-hand side of the scatter plot, are generally surrounded by relatively more voluminous voids - since most neighbours surpass the solid line, whereas the larger voids are generally surrounded by smaller voids. The significance of the contribution of voluminous voids to the distribution of data points, however, is not evident, since voluminous voids are present for almost all central voids, implying that most voids are surrounded by, at least, one voluminous void. This seems rather obvious, since voluminous voids encompass more space, and therefore touch on a larger set of voids than small voids. A voluminous void, however, is generally merely surrounded by smaller voids, as can be seen from the right-hand side of the graph. This allows a possible correlation between the volumes of neighbouring voids, since, as voids grow, they may tend to compress their neighbouring voids, causing a non-uniform distribution of the volumes of voids within the data cube, conforming the hypothesis of the hierarchical structure of voids. This suggests that the assumption of uncorrelated voids for large volume surveys is invalid, and that the AP test does not describe the behaviour of voids properly.

4.3 Alignments of Neighbouring Voids

Thus far, only the shapes of voids were considered, without addressing their spatial orientations, which, as pointed out by several studies, show significant alignments. (Bond et al., 1996; Platen et al., 2008) Again, the NVF can be used to find the characteristics of ellipsoidal neighbouring voids with respect to a central void, such that, by using the same random selection of fifty voids discussed in Section 4.2, their relative positions (Section 4.3.1) and orientations can be compared and investigated (Section 4.3.2), allowing to find any significant alignments for the orientations of voids.

4.3.1 Centres of Neighbouring Voids

Among other things, the NVF can be used to identify the position of the centre of a specific neighbouring void. These positions can be processed in order to investigate whether the distribution of (neighbouring) voids is anisotropic. The procedure used to investigate the positions of neighbouring voids is as follows:

1. Use the 'Neighbouring Void Finder' to identify and locate all voids at the border of a particular central void, and subsequently identify their centres.
2. Draw a vector from the centre of the central void to the centre of each neighbouring void and rescale this vector such that its length reduces to 1.
3. Obtain the orientation of the vectors in terms of spherical coordinates and find the points of the orientation of the vectors on a unit sphere.
4. Plot the resulting unit sphere, vectors, and data points in three dimensions using *ipyvolume*.

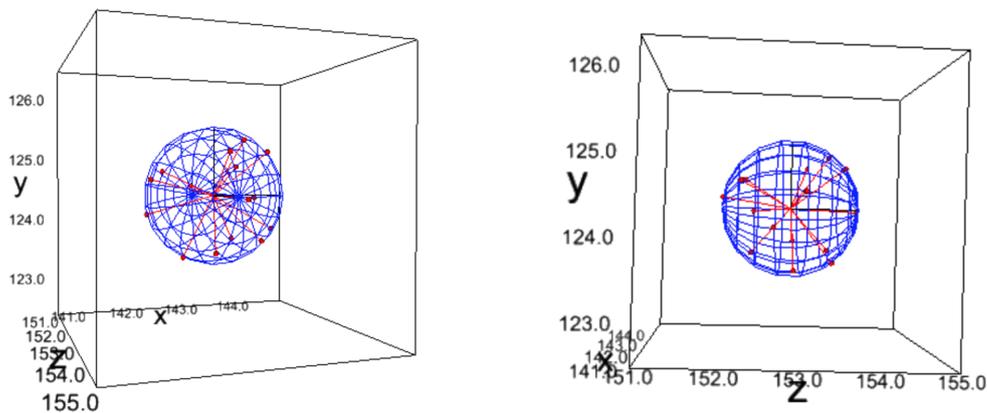


Figure 4.3: Example unit sphere showing the orientation of the centres of neighbouring voids of void 1111, which is also shown in Figure 3.4, 3.5, and 3.8.

An example of this procedure (for void 1111) is shown in Figure 4.3. It is, however, impossible to deduce - from a single plot - whether the distribution of centres of

neighbouring voids for the complete data set is isotropic, since the signal of an individual void is too weak to apply statistical analysis. Therefore, the procedure should preferentially be generalised and applied to all voids within the data cube. This, however, as already explain in Section 4.2, consumes a lot of computational time, such that it is more convenient to take the same random selection of fifty voids within the $256 \times 256 \times 256$ data cube. By applying periodic boundary conditions, and using the procedure described above, the orientation of centres of neighbouring voids can be found and plotted using a Cumulative Distribution Function (CDF), in which the probability - ranging from 0 to 1 - of a random variable X of taking a value less than or equal to x , is given in a single plot. In the case of a random, non-correlated distribution, one expects a straight line, for which all probabilities of X taking a specific value x are equal. In the statistics of spherical coordinates, generally two distributions are investigated: $\cos \theta$ and φ , since an isotropic distribution of θ shows a linear behaviour in $\cos \theta$, i.e. would match a straight line when plotted in a CDF, whereas φ behaves linearly by definition. The CDF for the angular orientations of the centres of neighbouring voids with respect to the centre of the central void is obtained as follows:

-
1. Take a random selection of the void IDs and find the neighbouring voids for each void using the 'Neighbouring Void Finder' algorithm, and subsequently find the correlation sphere (similar to the one shown in Figure 4.2).
 2. Find for each vector, obtained in the procedure for plotting the correlation sphere, the polar and azimuth angle, i.e. θ and φ . Repeat the process for all voids in the random selection.
 3. Create two sorted lists: one containing all values of φ for the (neighbouring) voids, and one containing all values of $\cos \theta$.
 4. Find the total number of values for $\cos \theta$ and φ , and assume equal weights for all values, and find the probability of X taking a specific value x , i.e. a specific value of $\cos \theta$ (between -1 and 1)³ or φ (between 0 and 2π).
 5. Add up the probabilities and plot the resulting CDF.
-

The resulting CDFs for fifty randomly selected voids within the data cube are shown in Figure 4.4, in which the spherical coordinates θ and φ correspond to the spherical coordinate system explained in section 3.2.1. Both CDFs, in particular of $\cos \theta$, show a significant deviation from uniformity, i.e. do not follow the straight line completely. These deviations can in particular be found at $-0.75 \leq \cos \theta \leq 0.0$ and $\pi \leq \varphi \leq 2\pi$, in which the distribution of φ shows a minute periodic behaviour. The divergence from uniformity indicates that some orientations of neighbouring voids with respect to their central void are preferred, hinting on a preferential alignment of voids within the data cube and an exclusion of the AP test to describe the voids properly. To check whether these behaviours are common to all random selections, i.e. whether some directions are preferred within the whole data cube, a larger number of voids should be investigated, as is done in Appendix C, resulting in the observation that for $\cos \theta$, a deviation from uniformity occurs systematically, implying some preferred directions within the data cube, as can be seen from Figure C.1. The distribution of

³Since θ ranges from 0 to π , $\cos \theta$ ranges from -1 to 1 .

φ , on the contrary, does not mimic the periodic behaviour continually, which makes it plausible that the distribution of φ is randomly uniform.

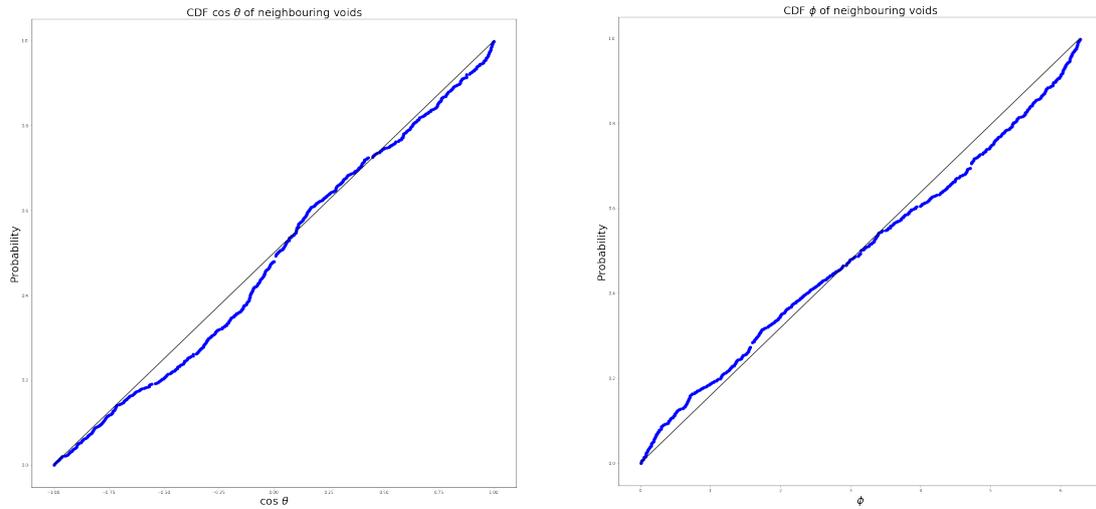


Figure 4.4: Cumulative Distribution Function (CDF) of $\cos \theta$ (left) and φ (right) for the orientation of (the centres of) neighbouring voids of a random selection of fifty central voids within the $256 \times 256 \times 256$ data cube.

4.3.2 Orientations of Principal Semi-axes

Besides the orientation of centres of neighbouring voids with respect to the central void, the principal semi-axes of the neighbouring voids can be compared to the axes of the centring void as well in order to find possible alignments within the data cube. Again, by using the condition $a > b > c$, the different principal semi-axes can be identified. By using the algorithm for fitting covariance ellipsoids, these axes can be found and subsequently plotted for the neighbouring voids by using the NVF. The result for void 1111 is shown below.

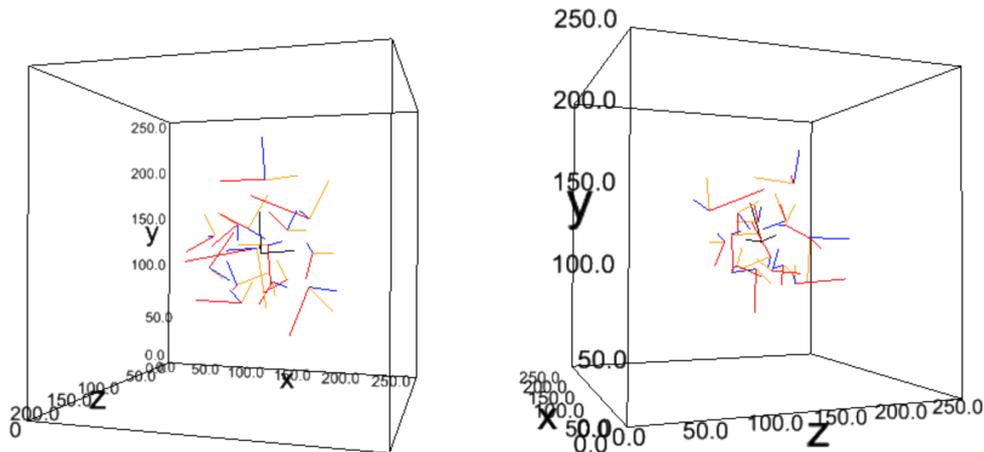


Figure 4.5: Three-dimensional plot showing the principal semi-axes a (red), b (orange), and c (blue) for the neighbouring voids of the central void 1111 with semi-axes $a_{central}$, $b_{central}$, and $c_{central}$ (all black).

The orientation of these axes can be used to find out whether neighbouring voids are oriented in a specific direction with respect to their central void. In order to quantify these angles, just as is done in Section 4.2.1, a Cumulative Distribution Function (CDF) is produced by using the following steps:

-
1. Take a random selection of the void IDs and find the neighbouring voids for each void using the 'Neighbouring Void Finder' algorithm, and subsequently find the covariance ellipsoid for all (neighbouring) voids, from which the principal semi-axes a , b , and c can be found and plotted.
 2. Find the orientation of the a , b , and c axes for the central void, and for all neighbouring voids, and find the angle θ between these axes, i.e. the angle between the central a -axis and a -axis of the neighbouring voids, the central b -axis and b -axes of the neighbouring voids, and the central c -axis and c -axis of the neighbouring voids, using the dot-product between both vectors as follows:

$$\cos \theta_a = \frac{\vec{a}_{central} \cdot \vec{a}_{neighbour}}{|\vec{a}_{central}| |\vec{a}_{neighbour}|} \quad (4.3)$$

$$\cos \theta_b = \frac{\vec{b}_{central} \cdot \vec{b}_{neighbour}}{|\vec{b}_{central}| |\vec{b}_{neighbour}|} \quad (4.4)$$

$$\cos \theta_c = \frac{\vec{c}_{central} \cdot \vec{c}_{neighbour}}{|\vec{c}_{central}| |\vec{c}_{neighbour}|} \quad (4.5)$$

3. Create two sorted lists, containing all values of $\cos \theta$ for the a -, b -, and c -axes, i.e. containing all values of $\cos \theta_a$, $\cos \theta_b$, and $\cos \theta_c$.
 4. Find the total number of values for $\cos \theta$, and assume equal weights for all values, and find the probability of X taking a specific value of x , i.e. a specific value of $\cos \theta$ (between -1 and 1) for a , b , and c .
 5. Add up the probabilities and plot the resulting CDF.
-

Similar to the CDFs obtained in Figure 4.4, for statistical convenience and computational time, the same random selection of fifty voids was used in order to obtain the CDF for the orientation of the a -, b -, and c -axes of neighbouring voids with respect to their central void. The resulting CDFs for these three axes are shown in Figure 4.6, marked by the same colours as in Figure 4.5. Again, a completely random distribution would follow the straight line shown in all CDF-plots. However, as one can observe, the distributions for $\cos \theta_a$ and $\cos \theta_b$ deviate significantly from the straight line, indicating that the distribution of orientations of axes for neighbouring voids with respect to their central voids within the cosmic web is not random, nor isotropic, but exhibits some correlations. The plots indicate that the correlation of principal semi-axes is strongest for a , subsequently for b , and finally for c , implying that the largest axes are correlated the strongest. Due to these significant alignments of the semi-principal axes of neighbouring voids with respect to their central void, at least for the principal semi-axes a and b , it may be unwarranted to use the AP test on a large volume survey and to assume a spherical total void resulting from the method in order to statistically describe voids. The results, on the contrary,

favour a model predicting alignments of voids, such as the cosmic tidal force field induced by the inhomogeneous cosmic matter distribution within the cosmic web, as discussed in Chapter 1.

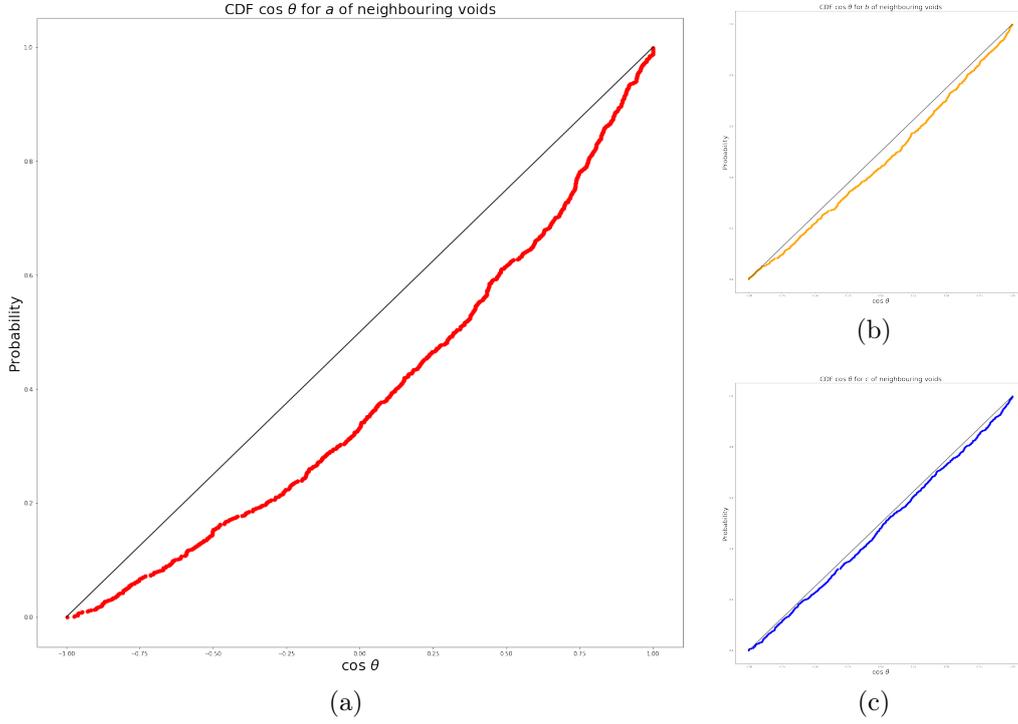


Figure 4.6: Cumulative Distribution Function (CDF) of $\cos \theta_a$ (a), $\cos \theta_b$ (b), and $\cos \theta_c$ (c) for the orientation of the a -, b -, and c -axes, respectively, of the neighbouring voids with respect to their central voids for a random selection of fifty voids within the $256 \times 256 \times 256$ data cube.

4.4 Intertwined Correlations

Besides the separate correlations of neighbouring voids with respect to a central void in terms of positions, principal semi-axes, and volumes, the correlations between these results for different random selections within the $256 \times 256 \times 256$ data cube can be investigated as well, as is done in Appendix C. It can be observed visually that an increasing amount of voluminous centring voids, causes a larger deviation from uniformity for the principal semi-axis a and the $\cos \theta$ orientation of centres. This suggests that the size of the central void affects the correlations between central and neighbouring voids, i.e. as the volume of the central void increases, the correlations of the positions and principal semi-axes of neighbouring voids, at least in terms of $\cos \theta$ and $\cos \theta_a$, are more substantial. This implies that the (correlation in) orientation of voids within the cosmic web is strongly determined by the orientation of the most voluminous voids. The question whether these alignments of voluminous central voids with respect to their neighbouring voids originate from the cosmic tidal field, however, remains unanswered.⁴

⁴To find out the origin of these alignments, a more thorough analysis should be executed, in which the correlations between volumes and orientations are analysed quantitatively. The results obtained thus far are merely an indication, but not decisive.

Density Profiles of Voids

Thus far, only the shapes and orientations of voids were considered. However, in the Universe, the points contained within a specific void, differ by - among other things - their density values. In order to investigate the shapes and alignments of voids, these characteristics should be taken into account as well. By again applying the Watershed Void Finder (WVF) to the same file obtained from a Λ CDM-simulation as is used in the preceding chapters, a $512 \times 512 \times 512$ data cube, containing all density values of the voids, can be obtained. These density values are used to obtain both (two-dimensional) circular (Section 5.1) and shape-adapted (Section 5.2) density profiles, such that it can be tested whether spherical averaging of voids is valid for large volume surveys.

5.1 Circular Density Profiles

As explained in Section 2.2, the density file can easily be obtained using the WVF, resulting (in two dimensions) in a density grid, containing the density values of all pixels within the set, which - in a graph - differ by their colour-intensity, as can be seen from Figure 2.2, which shows the density, void, and combined (density and void) grid for the first slice, i.e. $z = 0$, of the $512 \times 512 \times 512$ Λ CDM data set. Using the combined grid, the density values corresponding to points within a specific void can be obtained and analysed, such that a density profile can be obtained. As is expected, the density profile of a void increases from the centre to its boundaries, implying that matter moves out faster than surrounding material closer to the boundary of the void, eventually accumulating in a steep ridge around the void. (Icke, 1984; van de Weygaert & van Kampen, 1993; Sheth & van de Weygaert, 2004) One thus expects a peak in the density profiles at radii sufficiently far away from the centre of the void, representing a ridge of accumulated matter, for which the density value decreases afterwards and eventually equals the density value of the surrounding medium.

5.1.1 Identify Voids and Corresponding Densities

Similar to the procedure used in Chapter 3, the voids within the data grid can be identified and subsequently isolated. By simultaneously appending the density

values of the pixels within the void, the density profile can eventually be obtained. As a starting point, the following procedure is used to identify the void and its centre:

1. Use the files containing the density values and void IDs for all pixels within the data cube, and slice the data in a particular z -direction.¹
2. Find all pixels corresponding to a void with a specific ID, and append the values to three lists: one containing all x -values, another containing all y -values, and the last containing all corresponding density values.
3. Find the centre of the void by computing the mean in both the x - and y -direction and subsequently plot the result within the data grids.

An example of this procedure is shown in Figure 5.1, for a void with ID 333.

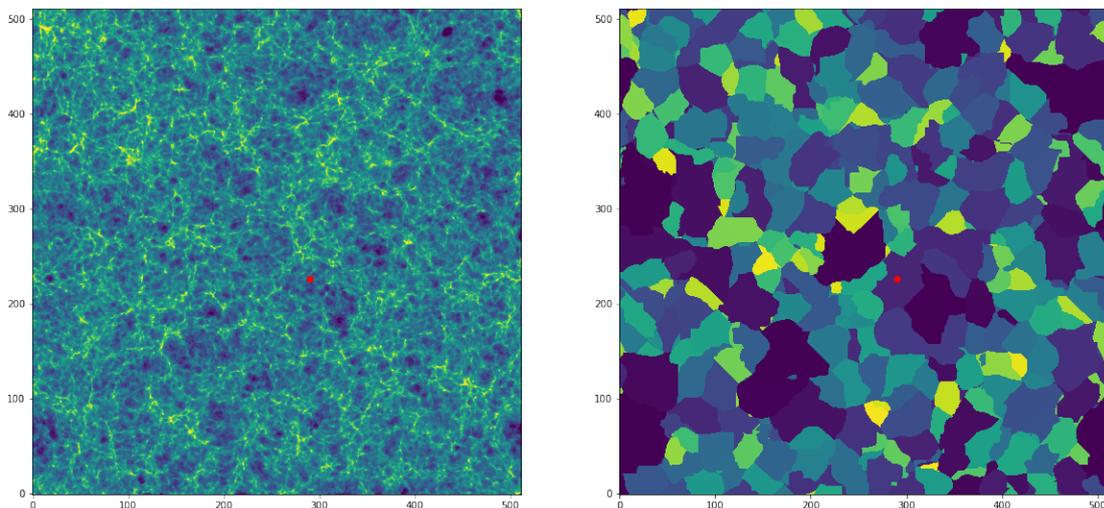


Figure 5.1: Example two-dimensional plot of finding the centre (shown in red) of a specific void (with ID 333) in both the density and void grids.

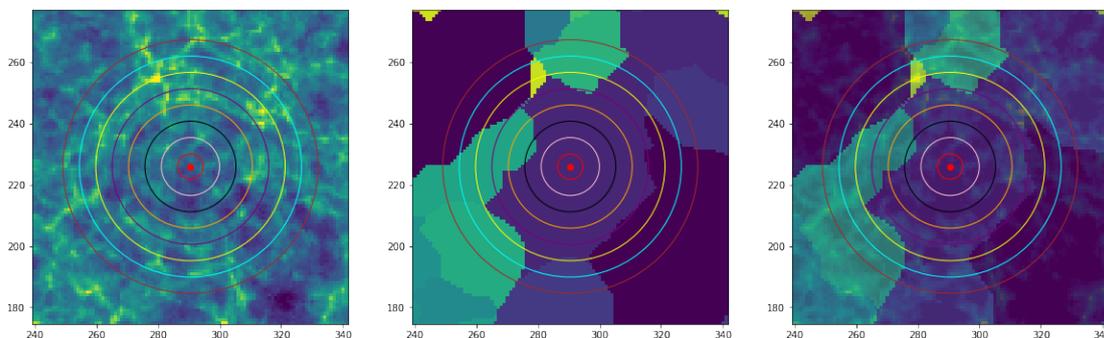


Figure 5.2: Example two-dimensional plot of the centre (red) and circles surrounding a specific void (with ID 333), shown in density grid (left), void grid (middle), and combined grid (right).

¹In this chapter, this direction is always taken to be $z = 0$.

The centre of the void is used as a starting point for the determination of its circular density profile. For each void, eight circles with equal spacing are used, with an example shown in Figure 5.2. By using circles with different radii, and by identifying all points within the given circle and consequently find their density values, two types of profiles can be obtained: 1) a density profile δ , in which the density of a shell at a distance r is found, and 2) a total density profile Δ , in which the total density within a shell at distance R is found. These profiles are, in two dimensions, related by:

$$\Delta(R) = \frac{1}{\pi R^2} \int_0^R 2\pi r \delta(r) dr \quad (5.1)$$

5.1.2 Obtain Density Profile

Using the circles drawn around the void, the circular density profile can be obtained by using the following procedure:

-
1. Find the data sets for both the density values and positions (in x - and y -direction) of a specific void and append these values to a new list.
 2. Use the maximum values within the lists containing the positions of the pixels and use these values to create a list containing eight equal-spaced radii of circles, with the outer circle (of radius r_{max}) encompassing the total void, i.e.

$$r_{max} = \sqrt{x_{max}^2 + y_{max}^2} \quad (5.2)$$

3. Use the condition

$$(x - x_{mean})^2 + (y - y_{mean})^2 < r^2 \quad (5.3)$$

to find all points (x, y) within a circle of radius r , and append these values to a new list. Plot the data points and corresponding circles for all radii in a figure containing eight subplots.

4. Add up the densities corresponding to all points within the circles, such that eight values are obtained, divide by the area of the circle, and plot the resulting cumulative circular density profile Δ . By taking the derivative, the density profile δ is obtained.

The procedure described above is used to obtain a circular density profile of void 333, for which the result of step 3 is shown in Figure 5.3, showing eight equal-spaced circles encompassing all points within a certain radius r , gradually showing the shape of the void as the radius increases. This procedure leads to a cumulative circular density profile, in which all density values are added to obtain a profile. However, in order to obtain the actual density profile, i.e. a profile in which the density is a function of radius, and for which the value at a specific radius r represents the actual density at that specific location, rather than the total density encompassed

within that radius, the derivative of the total circular density profile should be taken. These density profiles can be used to study the containment of voids, and whether cosmic voids have a characteristic matter content.²

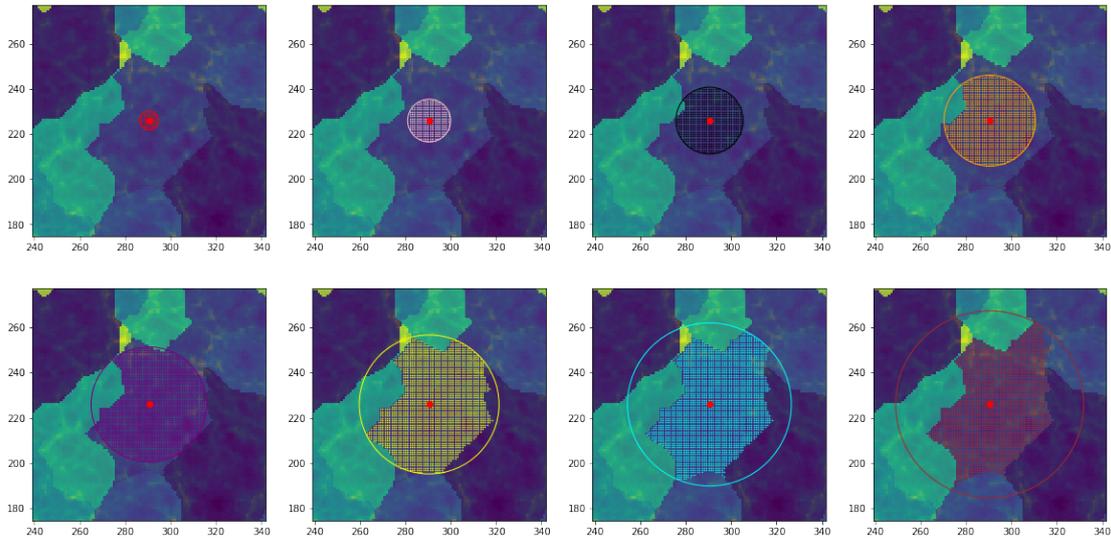


Figure 5.3: Example plot showing step 3 of the algorithm, applied to void 333, in which for different radii the points within the circle are identified and marked by different colours.

5.2 Shape-adapted Density Profiles

The circular selection of data points becomes insufficient for large radii, since for irregularly shaped voids the circles do not mimic the actual shape of the void properly, causing a substantial deviation from the physical contents of the voids. Modelling irregularly shaped voids by means of circular profiles, therefore, appears to be inadequate, asking for a shape-adapted profile in which the actual shape of the (two-dimensional) void is taken into account. By comparing the density profiles obtained via both procedures, it can statistically be checked whether voids are adequately modelled by circular profiles, and thus whether voids, in large volume surveys, can be assumed to be circular.³

5.2.1 Identify Voids and Corresponding Densities

In order to produce a shape-adapted density profile of voids, the method for finding the spatial coordinates and density values corresponding to a void with a specific void ID is used in the first place, such that all points corresponding to the void are identified. The same procedure as explained in Section 5.1.1 is used.

²The density profiles obtained thus far are not properly normalised, and are presented as a preliminary result in Section 5.3.

³The results obtained in such a case are merely an indication for the validity of spherical averaging of voids in large volume surveys, since actually their three-dimensional shape should be taken into account.

5.2.2 Obtain Density Profile

The shape-adapted density profile is obtained using the following algorithm:

1. Find the data sets for both the density values and positions (in x - and y -direction) of a specific void and append these values to a new list.
2. Use the total data set containing all spatial coordinates of the void and compute the alpha shape of the void.⁴ (Edelsbrunner, Kirkpatrick, & Seidel, 1983; Harniel, 2014) This alpha shape marks the boundary of the void.
3. Create a new list, containing all boundary points of the alpha shape, and draw vectors to these points. By multiplying these vectors by specific magnitudes (between 0.1 and 1, where 1 is their original size), the alpha shape for the void at different 'radii' is found.
4. Find all points in the original density file within the alpha shape, and append these values to a new list. Plot the data points and corresponding alpha shapes for all 'radii' in a figure containing eight subplots. (Dwyer, 2014)
5. Add up the densities corresponding to all points within the alpha shapes, such that eight values are obtained, divide by the area of the shape, and plot the resulting cumulative shape-adapted density profile. By taking the derivative, the general shape-adapted density profile can be obtained.

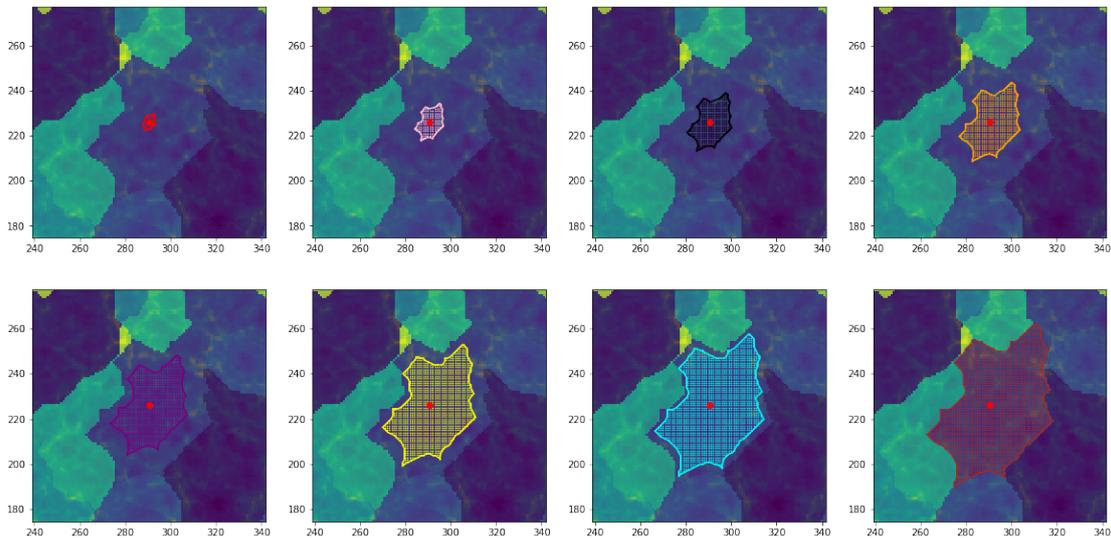


Figure 5.4: Example plot showing step 4 of the shape-adapted algorithm, applied to void 333, in which for different radii the points within the alpha shape are identified and marked by different colours.

This algorithm is, again, used for void 333, in which the result of the shape-adapted algorithm is shown in Figure 5.4. As can be observed, the shape of the boundary line perfectly matches the shape of the void, and the points selected in the last step

⁴An alpha shape is a family of piecewise linear simple curves associated with the shape of a finite set of points. For more information, please see Appendix D.

(shown in red) coincide with the same points in Figure 5.3. By comparing the shape-adapted and circular algorithms (see Figure 5.5), one finds that as the radius of the circle increases, the difference between the circular and shape-adapted procedures starts to increase, suggesting that the density profiles of both algorithms will start to deviate significantly for larger values of r . To check whether this is actually the case, the density profiles of the shape-adapted algorithm should be obtained and compared to the circular ones.⁵

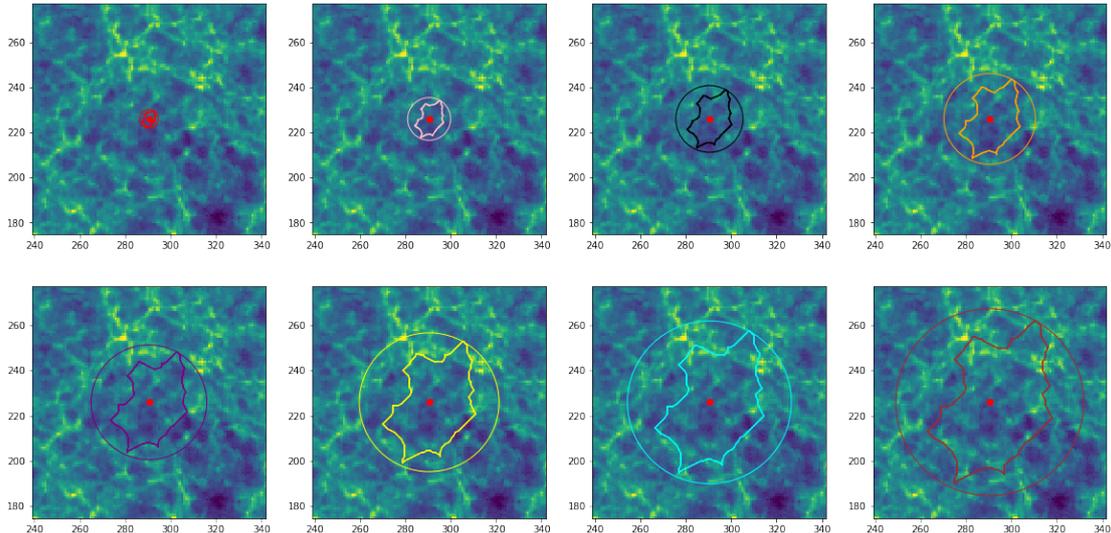


Figure 5.5: Example plot comparing step 3 of the circular algorithm, and step 4 of the shape-adapted algorithm, for void 333, in which the boundaries at different radii are marked with different colours.

5.3 Preliminary Results

The algorithms described in Section 5.1 and 5.2, i.e. the circular and shape-adapted algorithms used to obtain the density profiles of two-dimensional voids, were used to obtain the density profiles of a selection of ten well-defined voids, i.e. type 1 voids, as discussed in Section 3.3, within the 512 x 512 ($z = 0$) data grid. The circular density profiles, however, do not show the proper normalisation factors, whereas the shape-adapted ones show unexpected, nonphysical behaviour, i.e. the density value increases behind the steep ridge surrounding the void, in contrast to the behaviour expected, as discussed in Section 5.1. However, some conclusions may still be drawn from the general shape of the density profiles, making it worth to discuss the preliminary results obtained thus far.

The circular and shape-adapted density profiles of a selection of ten voids are shown in Figure 5.6 and 5.7, showing the cumulative density profile Δ on the left-hand side, and the density profile δ on the right-hand side. Both profiles start at low-density values, confirming the emptiness of the central region of voids, and both increase up to a certain 'peak', representing the predicted ridge at which matter is accumulated, marking the boundary of the void. (Icke, 1984; van de Weygaert & van Kampen,

⁵The shape-adapted density profiles obtained thus far do not behave as expected, suggesting an error within the analysis of density values. Therefore, the profile is presented as a preliminary result in Section 5.3.

1993; Sheth & van de Weygaert, 2004) As can be noted, the narrowness of the peak strongly depends on the procedure used, i.e. the shape-adapted method produces a sharp, narrow peak, whereas the circular method produces a broad peak. This may be due to the fact that the circular procedure disregards the shape of the void, deviating from the actual area of the void significantly for larger radii, smoothing out the steep ridge located near the boundary of the void. The shape of this ridge naturally mimics the shape of the actual void, and thus is much more apparent in a shape-adapted profile, indicating that disregarding the shape of the void, and simply apply a circular algorithm to obtain its density profile, yields a non-representative density profile of the void.

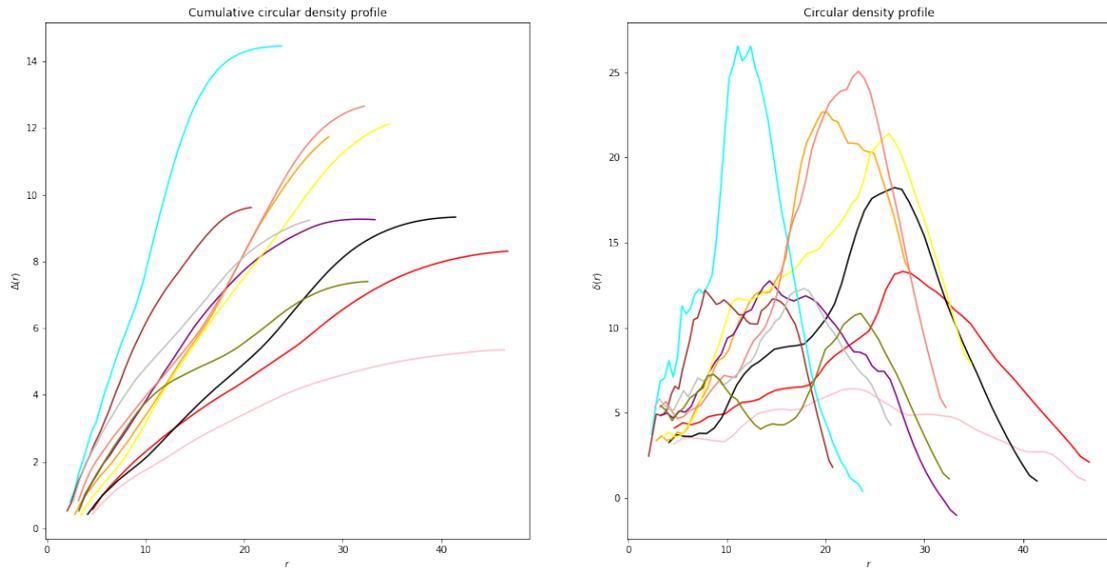


Figure 5.6: Combined density profiles, resulting from the circular algorithm, of a random selection of ten voids within the 512 x 512 ($z = 0$) data grid. Each void is shown in a different colour.

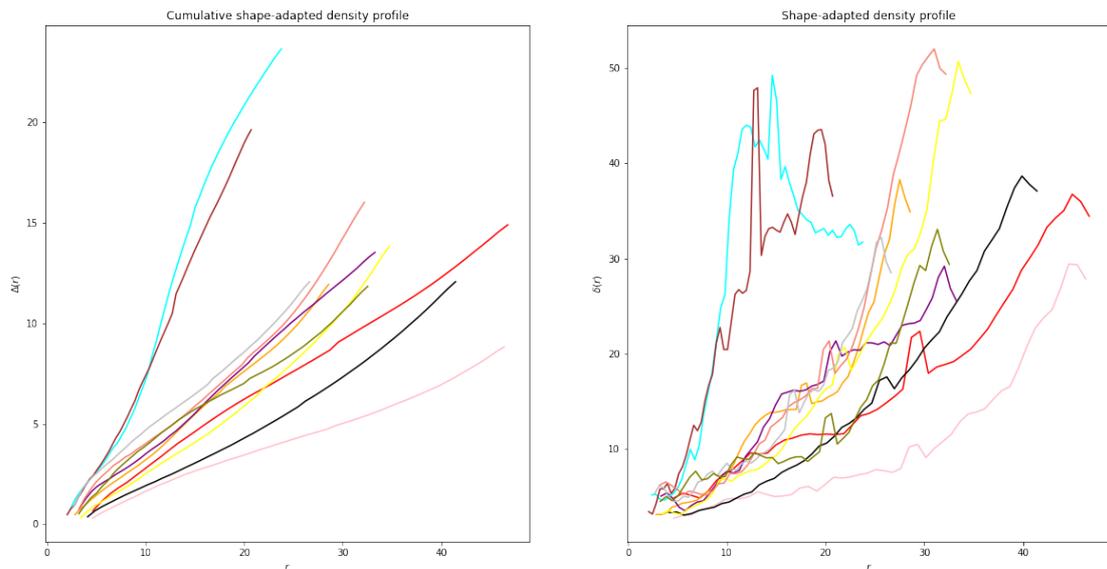


Figure 5.7: Combined density profiles, resulting from the shape-adapted algorithm, of a random selection of ten voids within the 512 x 512 ($z = 0$) data grid. Each void is shown in a different colour.

However, using these preliminary results, one cannot arrive at a conclusive remark regarding the validity of the circular approximation of voids within a large volume survey. In order to be able to do so, the flaws within the code must be fixed, and a larger number of voids should statistically be analysed. Additionally, to check the validity of the AP test and the subsequent spherical averaging of total voids, as is the main object of this research, the density profiles should be analysed in three dimensions, i.e. a spherical and shape-adapted algorithm should be implemented and used to obtain three-dimensional density profiles, describing the contents of voids far more sophisticated. The results presented in this section are, thus, merely a small hint regarding the containment of two-dimensional voids, and the validity of a circular approximation of irregularly shaped voids, but in no way conclusive.

Conclusion and Discussion

The Universe exhibits a particular structure, in which sheets and filaments of galaxies are separated by tenuous voids, known as the cosmic web. These voids encompass the largest parts of the volume of the Universe, and shape its mass distribution by means of tidal forces, therefore containing valuable information on the evolution of the structure of the Universe. Voids can thus be used as probes for cosmological parameters, such that, in a void-based description of the Universe, a general model describing the diversity in shapes, orientations, sizes, and matter contents of voids, and subsequently the cosmic web, can be obtained. In order to extract the required cosmological information from voids, sometimes an Alcock-Paczynski (AP) test is used, which assumes no preferential alignments within the cosmic web, and a extensive number of voids in a large volume survey is amassed into a spherical total void. (Alcock & Paczyński, 1979; El-Ad & Piran, 1997; Lavaux & Wandelt, 2012; Pisani et al., 2014) Several studies, however, indicate that voids show significant alignments on reasonably large scales due to the cosmic tidal field, marking the dubitability of the application of the AP test and the subsequent spherical averaging of voids within the cosmic web. (Bond et al., 1996; Platen et al., 2008) To check the validity of the assumptions underlying the AP test and spherical averaging, a $512 \times 512 \times 512$ Λ CDM data set was analysed, using the WVF, in terms of shapes, volumes, alignments, and density profiles, both for individual and neighbouring voids. It was found that voids generally tend to be prolate, rather than spherical, and that more voluminous voids are surrounded by a relatively larger amount of small voids, confirming the hierarchical structure of voids. (van de Weygaert & van Kampen, 1993; Sheth & van de Weygaert, 2004) In addition, it was found that both the locations of the centres, and the principal semi-axes, in particular a , of neighbouring voids with respect to their central voids, are distributed non-uniformly, and show - when plotted in a CDF - a significant correlation, in particular for more voluminous central voids. The density profiles obtained by the circular and shape-adapted algorithms indicate that it is unwarranted to approximate (the containment of) two-dimensional voids by circles in order to obtain a density profile, but no decisive remark can be made in this respect. Altogether, the results presented in this project indicate that it may be unwarranted to use the AP test and subsequent spherical averaging to describe voids on large scales properly, since these voids show significant alignments.

However, to arrive at a conclusive remark regarding the validity of the application of the AP test and the spherical averaging of voids within a large volume survey, a more extensive research should be executed, and the results presented in this re-

search should be excavated. First of all, a larger amount of Λ CDM data sets should be used, in order to apply the analysis to several distributions of matter and corresponding density values, such that a higher confidence level of the significance of the alignments within the cosmic web can be found. To rule out the possibility of alignments to be merely a defect of the Λ CDM simulations, other cosmological models should be investigated as well. Moreover, to check the significance of alignments for voids at larger distances, the algorithms used should be generalised in order to deal with neighbouring voids at different spatial distances, such that it can be analysed whether alignments tend to cease at larger distances, or whether the alignments are common to all voids within the selected data cube. The investigations considering the density profiles of voids deserves a more thorough analysis, such that one can arrive at a conclusive remark on the validity of the circular algorithm with respect to the shape-adapted one, and the procedures should be generalised to three dimensions in order to obtain representative profiles. Additionally, analysing the velocity fields of voids as well, an aspect not discussed thus far, but containing valuable information on the spherical behaviour of voids, would indicate whether the motion of matter within voids tends to be radially outward, as is expected for spherical voids. (Icke, 1984)

These aspects should be analysed when investigating the shapes and alignments of voids, and are the objectives for a succeeding study concerning the validity of the AP test and the spherical averaging of voids in large volume surveys. Up to now, it is found that it may not be warranted to assume that voids do not exhibit preferential alignments on large scales, and to subsequently use the AP test and spherical averaging to analyse their structures, shapes, and containments.

References

- Ade, P. A., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., ... others (2014). Planck 2013 results. xvi. cosmological parameters. *Astronomy & Astrophysics*, 571, A16.
- Alcock, C., & Paczyński, B. (1979). An evolution free test for non-zero cosmological constant. *Nature*, 281(5730), 358.
- Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., & Szalay, A. S. (2010). The spine of the cosmic web. *The Astrophysical Journal*, 723(1), 364.
- Armstrong, S. (2018). *Visualises an ellipsoid in 3d given a mean and covariance matrix from a dataset*. Retrieved from <https://github.com/CircusMonkey/covariance-ellipsoid>
- Bolejko, K., Krasinski, A., & Hellaby, C. (2005). Formation of voids in the universe within the lemaître–tolman model. *Monthly Notices of the Royal Astronomical Society*, 362(1), 213–228.
- Bond, J. R., Kofman, L., & Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. *Nature*, 380(6575), 603.
- Breddels, M. A. (2017). *IPyvolume*. Retrieved from <https://ipyvolume.readthedocs.io/en/latest/#>
- Chan, K. C., Hamaus, N., & Desjacques, V. (2014). Large-scale clustering of cosmic voids. *Physical Review D*, 90(10), 103521.
- Chincarini, G., & Rood, H. J. (1975). Size of the coma cluster. *Nature*, 257(5524), 294.
- Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., ... others (2001). The 2df galaxy redshift survey: spectra and redshifts. *Monthly Notices of the Royal Astronomical Society*, 328(4), 1039–1063.
- Da, T. K. F. (2016). *CGAL 4.14 - 2D Alpha Shapes*. Retrieved from https://doc.cgal.org/latest/Alpha_shapes_2/
- De Lapparent, V., Geller, M. J., & Huchra, J. P. (1986). A slice of the universe. *The Astrophysical Journal*, 302, L1–L5.
- Dwyer, K. (2014). *Concave hulls using shapely and scipy*. Retrieved from <https://gist.github.com/dwyerk/10561690>
- Edelsbrunner, H., Kirkpatrick, D., & Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on information theory*, 29(4), 551–559.
- Einasto, J., Jõeveer, M., & Saar, E. (1980). Structure of superclusters and supercluster formation. *Monthly Notices of the Royal Astronomical Society*, 193(2), 353–375.
- El-Ad, H., & Piran, T. (1997). Voids in the large-scale structure. *The Astrophysical Journal*, 491(2), 421.
- February, S., Larena, J., Smith, M., & Clarkson, C. (2010). Rendering dark energy void. *Monthly Notices of the Royal Astronomical Society*, 405(4), 2231–2242.
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. (2006). Cosmology at low frequencies: The 21 cm transition and the high-redshift universe. *Physics reports*, 433(4-6), 181–301.
- Furlanetto, S. R., & Piran, T. (2006). The evidence of absence: galaxy voids in the excursion set formalism. *Monthly Notices of the Royal Astronomical Society*, 366(2), 467–479.
- Gregory, S. A., & Thompson, L. A. (1978). The coma/a1367 supercluster and its environs.

- Guzzo, L., Abbas, U., Adami, C., Arnouts, S., Bel, J., Bolzonella, M., ... others (2013). Vipers: An unprecedented view of galaxies and large-scale structure halfway back in the life of the universe. *arXiv preprint arXiv:1303.3930*.
- Hanniel, I. (2014). *Boundary enclosing a given set of points*. Retrieved from <https://stackoverflow.com/questions/50549128/boundary-enclosing-a-given-set-of-points>
- Icke, V. (1984). Voids and filaments. *Monthly Notices of the Royal Astronomical Society*, 206(1), 1P–3P.
- Icke, V., Weygaert, R., et al. (1987). Fragmenting the universe. *Astronomy and Astrophysics*, 184, 16.
- Kirshner, R. P., Oemler Jr, A., Schechter, P. L., & Shectman, S. A. (1987). A survey of the bootes void. *The Astrophysical Journal*, 314, 493–506.
- Lavaux, G., & Wandelt, B. D. (2012). Precision cosmography with stacked voids. *The Astrophysical Journal*, 754(2), 109.
- Morales, M. F., & Wyithe, J. S. B. (2010). Reionization and cosmology with 21-cm fluctuations. *Annual review of astronomy and astrophysics*, 48, 127–171.
- Mota, D. F., Shaw, D. J., & Silk, J. (2008). On the magnitude of dark energy voids and overdensities. *The Astrophysical Journal*, 675(1), 29.
- Neyrinck, M. C. (2008). Zobov: a parameter-free void-finding algorithm. *Monthly Notices of the Royal Astronomical Society*, 386(4), 2101–2109.
- Park, D., & Lee, J. (2007). Void ellipticity distribution as a probe of cosmology. *Physical Review Letters*, 98(8), 081301.
- Pisani, A., Lavaux, G., Sutter, P. M., & Wandelt, B. D. (2014). Real-space density profile reconstruction of stacked voids. *Monthly Notices of the Royal Astronomical Society*, 443(4), 3238–3250.
- Platen, E., van de Weygaert, R., & Jones, B. J. (2007). A cosmic watershed: the wvf void detection technique. *Monthly notices of the royal astronomical society*, 380(2), 551–570.
- Platen, E., van de Weygaert, R., & Jones, B. J. (2008). Alignment of voids in the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 387(1), 128–136.
- Sheth, R. K., & van de Weygaert, R. (2004). A hierarchy of voids: much ado about nothing. *Monthly Notices of the Royal Astronomical Society*, 350(2), 517–538.
- Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E., Aymon, J., Boggess, N., ... others (1992). Structure in the coBE differential microwave radiometer first-year maps. *The Astrophysical Journal*, 396, L1–L5.
- Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M., Bennett, C., ... others (2003). First-year wilkinson microwave anisotropy probe (wmap)* observations: determination of cosmological parameters. *The Astrophysical Journal Supplement Series*, 148(1), 175.
- Spruyt, V. (2014). *A geometric interpretation of the covariance matrix*. Retrieved from <https://www.visiondummy.com/2014/04/geometric-interpretation-covariance-matrix/>
- Suto, Y., Sato, K., & Sato, H. (1984). Expansion of voids in a matter-dominated universe. *Progress of Theoretical Physics*, 71(5), 938–945.
- Sutter, P., Lavaux, G., Hamaus, N., Pisani, A., Wandelt, B. D., Warren, M., ... Thompson, B. B. (2015). Vide: the void identification and examination toolkit. *Astronomy and Computing*, 9, 1–9.
- Tatum, J. (2019). 3.3: *Cylindrical and Spherical Coordinates*. Retrieved from <https://phys.libretexts.org/Bookshelves/Astronomy>

[_and_Cosmology_TextMaps/Map%3A_Celestial_Mechanics_\(Tatum\)/3%3A_Plane_and_Spherical_Trigonometry/3.3%3A_Cylindrical_and_Spherical_Coordinates](#)

- Tegmark, M., Blanton, M. R., Strauss, M. A., Hoyle, F., Schlegel, D., Scoccimarro, R., ... others (2004). The three-dimensional power spectrum of galaxies from the sloan digital sky survey. *The Astrophysical Journal*, 606(2), 702.
- user3717023. (2016). *Re: Determine adjacent regions in numpy array [Online discussion group]*. Retrieved from <https://stackoverflow.com/questions/38073433/determine-adjacent-regions-in-numpy-array>
- van de Weygaert, R. (2014). Voids and the cosmic web: cosmic depression & spatial complexity. *Proceedings of the International Astronomical Union*, 11(S308), 493–523.
- van de Weygaert, R., Kreckel, K., Platen, E., Beygu, B., Van Gorkom, J., van der Hulst, J., ... others (2011). The void galaxy survey. In *Environment and the formation of galaxies: 30 years later* (pp. 17–24). Springer.
- van de Weygaert, R., & van Kampen, E. (1993). Voids in gravitational instability scenarios–i. global density and velocity fields in an einstein–de sitter universe. *Monthly Notices of the Royal Astronomical Society*, 263(2), 481–526.

Appendix A

The Covariance Matrix

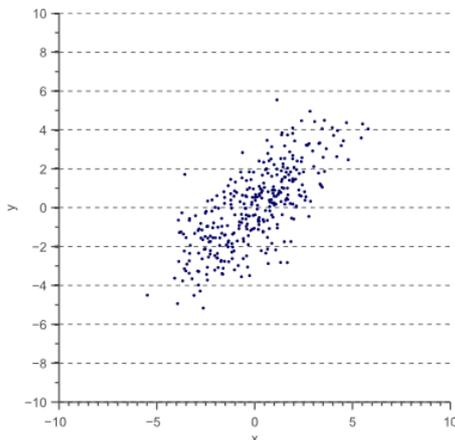
In general, a matrix is a rectangular array of numbers, arranged in m rows and n columns, and denoted as an $(m \times n)$ -matrix. When the dimensions of rows and columns are equal. i.e. $m = n$, the matrix is denoted as an $(n \times n)$ -matrix, and is referred to as a square matrix. In statistics, an often used type of matrix is the $(n \times n)$ -covariance matrix, which describes the spread of data in n dimensions, based on the (co)variance of the data. Variance is the expectation of the squared deviation (σ) of a random variable (x) from its mean (μ) and is given by

$$\sigma_x^2 = \sigma(x, x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2 \quad (\text{A.1})$$

with N as the sample size. Similarly, the variance of a random variable y is given by:

$$\sigma_y^2 = \sigma(y, y) = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu)^2 \quad (\text{A.2})$$

However, when dealing with two-dimensional data, the spread is not only captured by the variance in the x - and y -direction separately; the data can be correlated diagonally, as is shown for example in the figure below.



$$\Sigma = \begin{bmatrix} \sigma(x, x) & \sigma(x, y) \\ \sigma(y, x) & \sigma(y, y) \end{bmatrix}$$

The correlation can be captured by introducing the 'covariance' of the data, denoted as $\sigma(x, y)$ and $\sigma(y, x)$, and yielding the covariance matrix shown above. The covariance of two jointly distributed random variables is given by

$$\sigma_{xy} = \sigma(x, y) = E[(X - E[X])(Y - E[Y])] \quad (\text{A.3})$$

where $E[X]$ denotes the expected value of random variable X , also known as the mean of X , and $E[Y]$ the expected value of random variable Y . The covariance matrix thus defines both the spread (variance) and the orientation (covariance) of the data. (Spruyt, 2014) By finding the eigenvectors (\vec{v}) and -values (λ) of the covariance matrix, i.e. the values and vectors satisfying

$$\Sigma \vec{v} = \lambda \vec{v} \quad (\text{A.4})$$

one can find the orientation and size of the covariance ellipse describing the data, defined by the eigenvectors, which act as the elementary axes of the principal axes, and eigenvalues, which describe the size of the semi-minor and -major axis. In two dimensions, this can be implemented in Python as follows:

```
import numpy as np

# Find covariance ellipse using vector positions of points
def covariance_ellipse(x_vector, y_vector):

    coords = np.vstack([x_vector], [y_vector]) # Create combined list of coordinates
    cov = np.cov(coords) # Find covariance matrix of datapoints
    eigval, eigvec = np.linalg.eig(cov) # Find eigenvalues and -vectors

    sort_in = np.argsort(eigval)[::-1] # Sort eigenvalues from high to low
    a = np.sqrt(eigval[0]) # Compute semi-major axis
    b = np.sqrt(eigval[1]) # Compute semi-minor axis

    return a, b
```

Using this code, one can find a covariance ellipse for all well-defined voids identified within the data set, provided that the vector positions of all points with respect to the void's centre are known. This procedure can be generalised to three dimensions, yielding a covariance ellipsoid, rather than a covariance ellipse, which is based on a (3×3) -covariance matrix, and yields the principal semi-axes a , b , and c . Since probability theory acts as a foundation for finding the covariance ellipse (in two dimensions) or ellipsoid (in three dimensions), a confidence interval must be chosen in order to estimate the region that contains the true data. In general, and in this project, this interval is taken to be 95%.

Appendix B

Periodic Boundary Conditions

The 512 x 512 x 512 data cube used for this project shows periodic behaviour, in which the voids located at the boundaries of the set continue to the opposite side of the set, as can - for example - be seen from Figure 2.2 (left). By applying the algorithms described in Section 3 and 4 naively, i.e. by not taking into account the periodicity of the voids near boundaries, the results obtained do not reflect the actual data. One evident example of this behaviour is shown in Figure B.1, in which the principal semi-axes of all ellipsoids in the 256 x 256 x 256 data cube are compared before (left) and after applying periodic boundary conditions (right).

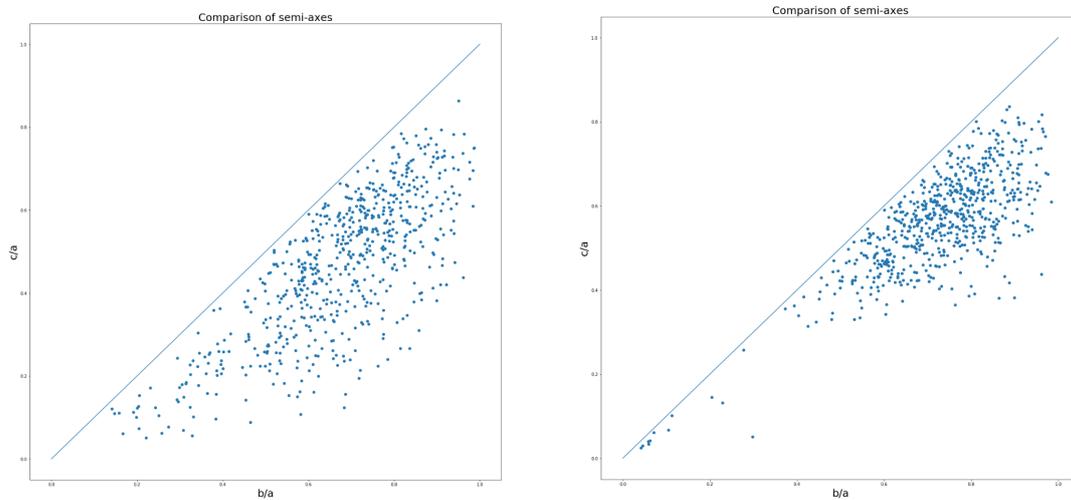


Figure B.1: Comparison of the principal semi-axes, $\frac{b}{a}$ vs. $\frac{c}{a}$, for all voids within the 256 x 256 x 256 data cube before (left) and after applying periodic boundary conditions (right).

As can be observed, the spread in points is significantly reduced after applying periodic boundary conditions, and most points are relocated to the topright-part of the plot, implying that without applying periodic boundary conditions, the shape of voids is parametrised incorrectly, since an ellipsoid is fitted to two or more sets of data points at different sides of the data cube, causing an ellipsoid with high prolateness to be produced, explaining the spread of points on the left-handed plot. While implementing periodic boundary conditions, at least two requirements need to be satisfied: 1) the data must remain unaltered, i.e. the relative positions of the

voids must remain unchanged, and 2) the code must remain substantially fast, since one deals with large data files of, for a 256 x 256 x 256 data cube, 256 MB. The structure of the 256 x 256 x 256 sliced data file investigated thus far is as follows:

- (a) Voids touching (at least) one axis, i.e. for which the data points can be found (at least) at $x = 0$, $y = 0$, or $z = 0$, continue at the other side of the total (unsliced) data cube, for which $x = 512$, $y = 512$, or $z = 512$.
- (b) Voids touching the boundaries of the (sliced) 256 x 256 x 256 data cube, i.e. at $x = 256$, $y = 256$, or $z = 256$, continue just behind these points in the total (unsliced) 512 x 512 x 512 data cube, i.e. up to all points for which $x > 256$, $y > 256$, or $z > 256$.
- (c) Voids located in between the axes and boundaries are well-defined, and can easily be plotted and investigated, without the need of applying periodic boundary conditions.

The algorithm must thus return a well-defined void, i.e. a void without discontinuities or gaps, for the three cases explained above. Several algorithms were programmed and used in Python, but most of them appeared to be way too slow, and thus required a new type of algorithm. Eventually, an algorithm based on the following steps was used for the 256 x 256 x 256 data cube:

-
- 1. Identify all unique void IDs within the given 256 x 256 x 256 data cube and create a list encapsulating these IDs.
 - 2. Find the positions of the voids within the data cube, and find the case-type of the void, and subsequently apply the following step (depending on the case):
 - (a) Identify whether the void is found at $x = 0$, $y = 0$, and/or $z = 0$, and subsequently roll this axis (using the Numpy-function *roll*) by 256 points, such that the void is relocated to the centre of the data cube.
 - (b) Use the data from the unsliced 512 x 512 x 512 data cube, such that all points corresponding to the void can be found.
 - (c) Leave the data set unaltered; the void is already well-defined.
 - 3. Return a new list containing the (shifted) data points for the voids within the data cube.
-

Since the void IDs of the 256 x 256 x 256 data cube are used, instead of all void IDs within the 512 x 512 x 512 data cube, the computation time of the algorithm is significantly reduced.¹ The new list as obtained by the algorithm, can again be used to find the covariance ellipsoid of the void, as described in Section 3.2.1, as well as the neighbouring voids by using the 'Neighbouring Void Finder', as described in Section 3.2.2. The results of the algorithm, as well as the ellipsoidal voids, are shown in Figure B.2 and B.3 for (case b-)void 12, showing the void as identified before and after applying periodic boundary conditions (Figure B.2), and the result

¹Please note that only the void IDs of the voids in the 256 x 256 x 256 data cube are used, but that, for the 'Neighbouring Void Finder', all voids within the dataset are used, such that all neighbours of the void can be found, without loss of generality.

of the ellipsoidal fit procedure and NVF in the original and total data cube (Figure B.3).

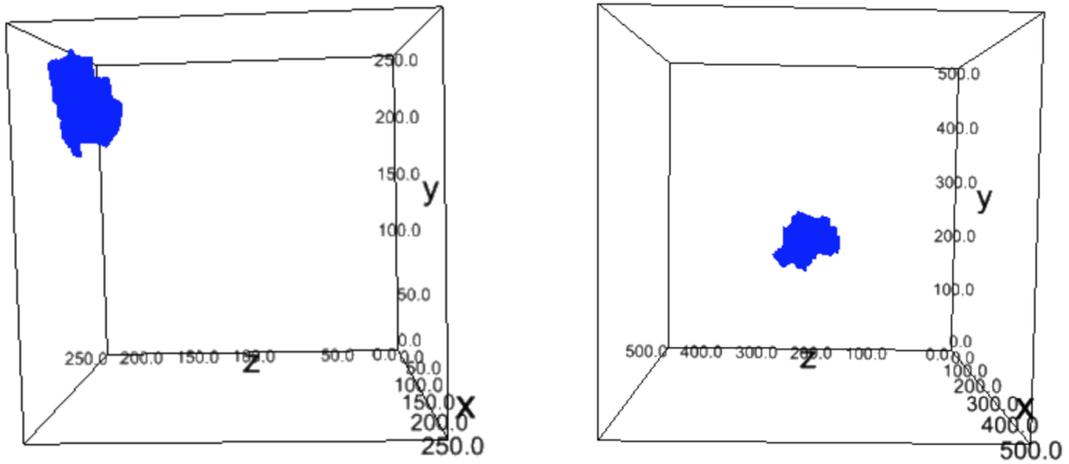


Figure B.2: Three-dimensional plot of (case b-)void 12 before (left) and after applying the periodic boundary conditions algorithm (right).

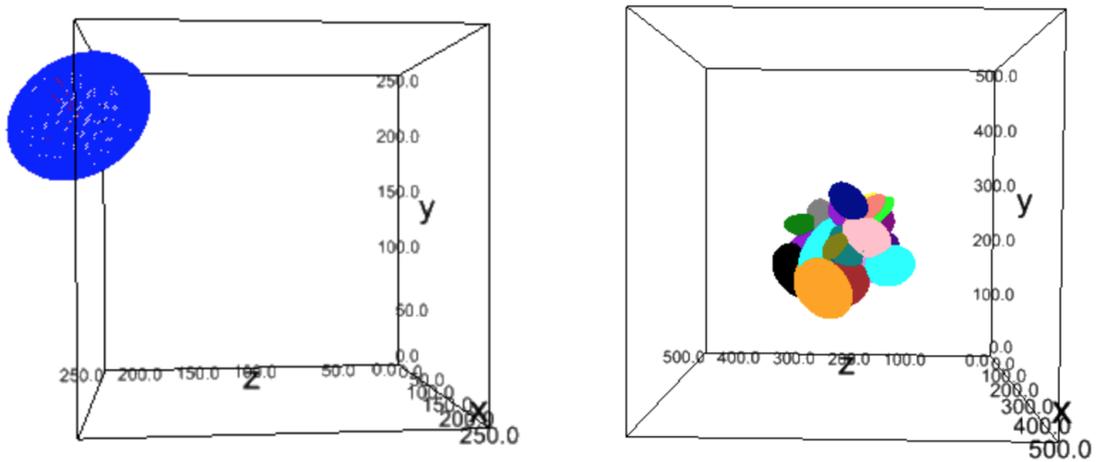


Figure B.3: Ellipsoid fit procedure for (case b-)void 12 in the original 256 x 256 x 256 data cube (left), and the neighbouring voids obtained by using the 'Neighbouring Void Finder' for the 512 x 512 x 512 data cube (right).

Random Selections and CDFs

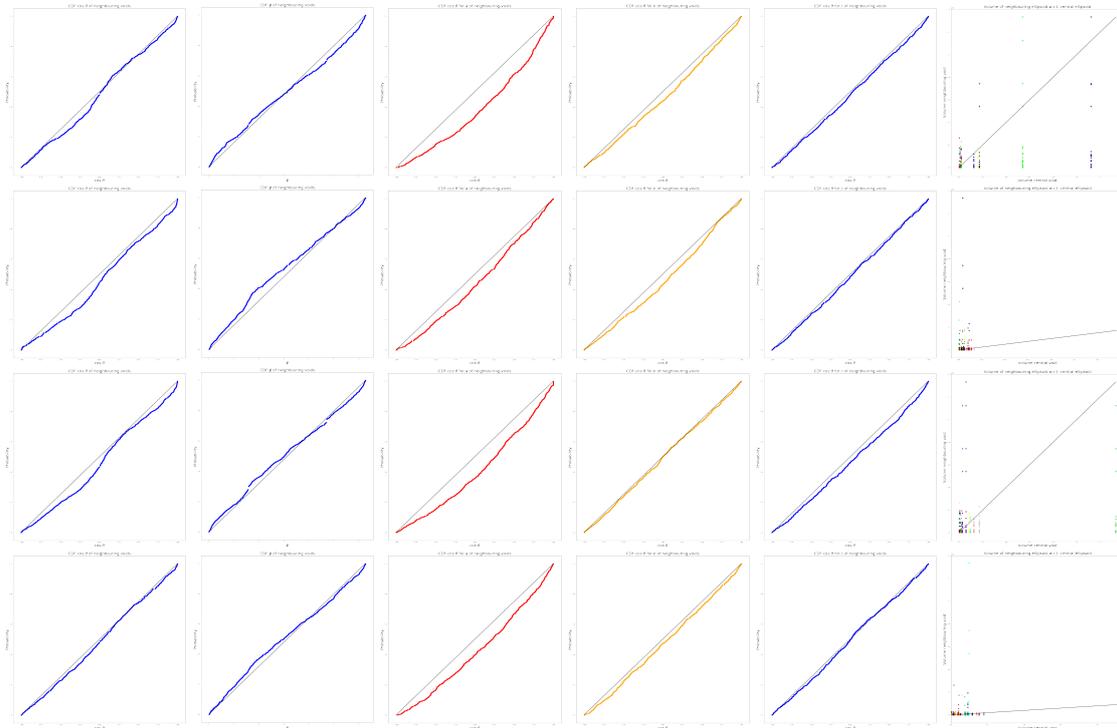


Figure C.1: Combined plot showing the CDFs of $\cos \theta$ and ϕ for the positions of neighbouring voids with respect to a central void, and of $\cos \theta_a$, $\cos \theta_b$, and $\cos \theta_c$ showing their orientations of the principal semi-axes with respect to a central void, and the volumes of the neighbouring voids with respect to the central void, for different random selections of the $256 \times 256 \times 256$ data cube. The different selections are shown in different columns, in which each row represents the distribution of 1) $\cos \theta$, 2) ϕ , 3) $\cos \theta_a$, 4) $\cos \theta_b$, 5) $\cos \theta_c$, and 6) the volumes.

The Cumulative Distribution Functions (CDFs) for the orientations of central and neighbouring voids is obtained using a random selection of fifty voids within the $256 \times 256 \times 256$ data cube. It appears that, for different selections, the results differ substantially. The results of four selections are shown above, and discussed in Chapter 4. Three general observations can be made for all data selections: 1) the $\cos \theta$ distribution continually shows a certain, non-uniform behaviour, 2) the principal semi-axis a shows a strong correlation, and 3) the relative number of small neighbouring voids increase with an increasing volume of the central void.

Alpha shapes and Polygons

In order to find the boundaries of irregularly distributed scatter points within a two-dimensional data grid, an alpha shape can be used. Intuitively, one can think of an α -shape as the following.

Imagine a huge mass of ice-cream making up the space \mathbb{R}^3 and containing the points as 'hard' chocolate pieces. Using one of these sphere-formed ice-cream spoons, we carve out all parts of the ice-cream block we can reach without bumping into chocolate pieces, thereby even carving out holes in the inside (e.g. parts not reachable by simply moving the spoon from the outside). We will eventually end up with a (not necessarily convex) object bounded by caps, arcs, and points. If we now straighten all 'round' faces to triangles and line segments, we have an intuitive description of what is called the α -shape of S . (Edelsbrunner et al., 1983; Da, 2016)

This procedure is visualised in Figure D.1.

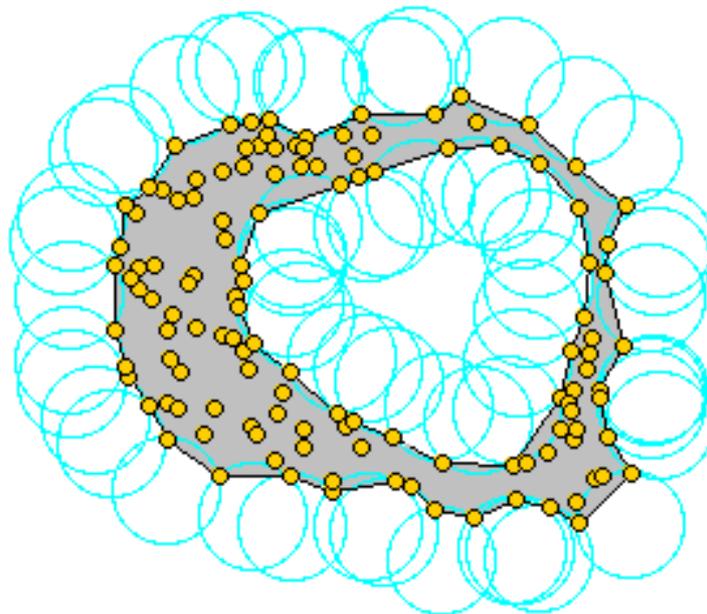


Figure D.1: Two-dimensional example of the α -shape procedure, in which the ice-cream spoon is simply a circle of radius r .(Da, 2016)

The shapes as obtained by the procedure depend on the parameter α , which, in case of the ice-cream analogy, corresponds to the square radius of the carving spoon. This indicates that small values of α allow us to eat up all the ice-cream except the chocolate points themselves, implying that as $\alpha \rightarrow 0$, the α -shape degenerates to the point-set S . On the contrary, as $\alpha \rightarrow \infty$, the spoon is too large to move even between two points, and the shape reduces to a convex hull of S . The value of α depends on the data set used, and can often be customarily defined. Since the two-dimensional voids show irregular shapes, α -shapes are ideal probes for marking the boundary points of these voids, and subsequently identify all points within this given shape. In this project, α is taken to be 2, preventing possible discontinuities to occur, whereas the accuracy of the result is maximised. The boundary lines as identified by the α -shape algorithm, are subsequently used to define a polygon encompassing the entire void, since the Python-module *shapely* contains many predefined functions allowing to analyse polygons, e.g. to find their area and the points contained within the shape, leading to the results shown in Figure 5.4. (Dwyer, 2014; Hanniel, 2014)