MCCXXII

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Master Degree in Physics

Final Dissertation

Modeling the void size function and its cosmology dependence in galaxy surveys

Thesis supervisor	Candidate
Prof. Sabino Matarrese	Anna Pugno
Thesis co-supervisor	
Dr. Nico Hamaus	

Academic Year 2019/2020

Acknowledgments

I would like to wholeheartedly thank Prof. Sabino Matarrese for the support and the advices he gave me during all my course of study. His guidance helped me grow as a student and a physicist.

I owe a great thank to Dr. Nico Hamaus, who supervised me during my Erasmus traineeship and with his patience and support taught me so much about how research is done. Without him this work would have never been possible.

Furthermore I would like to thank Prof. Jochen Weller and all the Cosmology Group of USM. Despite the difficult situation due to SARS-CoV-2, they made me feel part of the group and they gave me valid advices for the progress of my work.

Last but not least, I must thank all the people who supported me during the long years of university. Thanks to my family, who always believed in me and in my capabilities more than I do. Thanks to my friends, who shared with me the long study hours, the home, or just valuable time. A special thanks to Benedetta, the queen of bureaucracy, for her help in both scientific and administrative matters. And thanks to Geremia with all my heart, because he never left my side despite how annoying I can be when I panic for an exam, and his faith in me gave me the strength to face every challenge on my path.

Abstract

In the field of large scale structure, cosmic voids are gaining more and more importance as useful probes of cosmology and fundamental physics. This is partly due to the more linear gravitational evolution as compared to their overdense halo counterparts. However, the models and parameters used to describe cosmic voids are still not univocal, and their analysis is often limited to the largest sizes. This ignores the so-called void-in-cloud effect, describing smaller voids embedded in overdense surroundings.

The aim of this thesis is to model the void size function including the void-in-cloud effect, and to study its cosmology dependence. In order to do so, we generate different void catalogs from the PATCHY mocks of SDSS BOSS DR12 using the void finder VIDE, each assuming a different value of the matter density parameter $\Omega_{\rm m}$ for the transformation from redshifts to distances. We adopt a multivariate analysis to remove spurious voids by comparing their properties with voids from a random catalog. We fit different models for the void size function proposed in literature to the void abundance of the cleaned catalog. Each fit is performed using a MCMC with three free parameters: the linear threshold for void formation δ_V, the linear threshold for collapse δ_c, and the rescaling factor r_s. We find that the various models exhibit different levels of agreement with the data, and a different behaviour of the parameters as a function of $\Omega_{\rm m}$. The Vdn model shows an anomaly in its parameters corresponding to the true value of $\Omega_{\rm m}$ used in the construction of the mock galaxy catalog, indicating a way to recognise it among the different cosmologies. After the application an Alcock-Paczyński correction to the catalogs, the Vdn model shows a residual dependence of the fitting parameters upon the value of $\Omega_{\rm m}$. The results of this work underline the importance of studying the void-in-cloud effect in order to understand the distribution of cosmic voids in galaxy surveys.

Contents

Introduction

Observations and simulations of the universe indicate that at large scales matter is distributed in a web-like structure, the cosmic web (Bond & Myers (1996), Bond et al. (1996)). The cosmic web is composed by haloes (virialized high density regions of matter), filaments, walls and cosmic voids. The latter are underdense regions in the distribution of galaxies, and fill the majority of the volume of the Universe. Due to their low density character, cosmic voids are useful probes to investigate many aspects of the cosmology of the large-scale universe (Pisani et al. (2019), van de Weygaert & Platen (2011)). Redshift-space distortions (RSDs), distortions of the spatial distribution of galaxies due to their peculiar velocities, can lead to important information about the growth factor of structures. Cosmic voids allow the analysis of RSD at small scales, where the two-point statistics of galaxies is characterised by non-linearity and virialization, and provide information about the formation and evolution of structures (Hamaus et al., 2015). Being by definition devoid of matter, cosmic voids are dark-energy dominated objects. Their evolution is ruled only by the gravitational attraction and the expansion of the universe. Via the study of the expansion of voids during cosmic history it is possible to gain information about the nature of dark energy (Bos et al., 2012). Voids are also good laboratories to investigate Einstein's General Relativity and its possible extensions. Modified gravity leads to a difference in the matter density profile compared with General Relativity that can be measured via gravitational lensing around voids (Baker et al., 2018), and it causes a faster expansion of voids that can be captured by RSDs around voids (Hamaus et al., 2016). Voids are particularly sensitive to diffuse components, and they are therefore becoming a topic of interest for studying neutrinos and their properties. Additionally, the typical sizes of voids span the range of the free-streaming range of neutrinos (Massara et al., 2015).

We have just listed a few examples of the usefulness of voids as probes to investigate cosmology, but there is still not a univocal theory to describe them. Different aspects of voids has been studied and modelled in the literature, such as their density and velocity profiles (Hamaus et al., 2014). The number density of voids as a function of the void radius, the void size function, is the analogous of the halo mass function to estimate the abundance of voids and compare it to the observations. The first model proposed by Sheth & van de Weygaert (2004) to describe the void size function was based on the conservation of the number density of voids during their evolution. To take into account the fact that voids can merge together during their evolution, Jennings et al. (2013)

proposed a different model, the Vdn model, which considers as conserved quantity the cumulative volume fraction of void.

It has to be noted that the majority of the theories proposed so far have been developed in the dark matter haloes and voids context. Little attention has been devoted to the study of voids from galaxy surveys, and the models that well describe dark matter voids might not be equally valid for galaxy voids. Aim of this thesis work is to investigate the void size function of voids identified from galaxy surveys, studying the different models proposed in the literature. Via the parameters of the models obtained from the fits to the void abundance of catalogs constructed using different values of the matter density parameter in the transform from redshift to distances, we want to analyse the dependence of the different models of the void size function upon the chosen cosmology.

In Chapter 1 we will see the theory underlying this work, from the cosmological background to the derivation of the evolution of underdensities of the matter field which leads to voids. Adopting excursion set theory we will derive the expressions of the different models of the void size function. In Chapter 2 a description of the MultiDark PATCHY mock galaxy catalog (Kitaura et al. (2016), Rodríguez-Torres et al. (2016)) used to generate the void catalogs is presented, as well as a brief overview about how it was produced. In Chapter 3 we will look at the methods used: the void finder VIDE (Sutter et al., 2015) to identify voids from the galaxy catalog, and the Least Square and Markov Chain Monte Carlo methods to search for the best fitting parameters. In Chapter 4 the analysis of the work is presented. We will generate different catalogs using different values of the matter density parameter in the transform from redshifts to distances. Adopting a Markov Chain Monte Carlo method we will fit the different void size function models to the void abundance as a function of radius. Via the agreement of the data to the different models and the best fitting parameters we will study the dependence of the void size function upon the chosen model. Looking at the behaviour of the parameters obtained from the fits we will study the cosmology dependence of the different models of the void size function. Finally, we will apply an Alcock-Paczyński correction (Alcock & Paczyński, 1979) to the void catalogs and look for a residual cosmological dependence of the parameters. In Chapter 5 we will discuss the results obtained in Chapter 4 and we will suggest future developments of this work.

Chapter 1

Theoretical background

The structures that we observe today are the evolution of small perturbations in the matter density field of the universe. Hierarchical scenarios of structure formation successfully explain the formation histories of gravitationally bound virialized haloes. The excursion set approach first proposed by Press & Schecher (1974) and then developed by Epstein (1983) and Bond et al. (1991), provides an analytical description of the collapse and virialization of overdense dark matter haloes.

The universe at Megaparsec scales, however, shows filamentary and sheetlike structures that are not well described by virialized spherical structures. Excursion set theory can nevertheless be applied to the large empty regions between galaxy clusters and filaments, cosmic voids. While galaxy clusters contain the majority of the mass of the universe, cosmic voids are the dominant component of the total volume of the universe at Megaparsec scales. They are underdensities of the matter field of the universe, and they can be studied as spherical evolution of negative density perturbation (Blumenthal et al., 1992). In a void-based description of structure formation the matter is squeezed between expanding voids, forming filaments and sheets at the intersections between void walls (Icke, 1984).

In section (1.1) we will briefly see the elements of the cosmology we will be dealing with, in section (1.2) we will see the spherical evolution of overdensities and underdensities of the matter density field, and in section (1.3) we will study excursion set theory applied to voids, introducing the different models of first-crossing distribution function and void size function. In the development of this chapter we will mainly follow Hamaus (2017).

1.1 Cosmological Background

In this section we will first give a look to the cosmological framework we will be working in, focusing on the relevant aspects for the subsequent analysis. We will mainly follow Coles and Lucchin cosmology book (Coles & Lucchin, 1995). We are interested in systems at very large scales, where the only relevant force is the gravitational one. Einstein's General Relativity describes gravity as a distortion of the spacetime itself, and Einstein's equations relate the metric of the spacetime to the matter distribution of the universe:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci tensor constructed form the Riemann curvature tensor, $g_{\mu\nu}$ is the metric tensor, R is the Ricci scalar defined as the trace of the Ricci tensor $R = g^{\mu\nu}R_{\mu\nu}$, Λ is the cosmological constant, G is the gravitational constant, c is the speed of light and $T_{\mu\nu}$ is the stress-energy tensor. The cosmological constant was first introduced by Einstein himself in order to ensure static cosmological solutions, but today it is incorporated in the energy-momentum tensor as dark energy, the non-zero intrinsic energy density of the vacuum. The stress-energy tensor describes the matter distribution, and for a perfect fluid it takes the form

$$T_{\mu\nu} = (p + \rho c^2)u_\mu u_\nu - p g_{\mu\nu} \tag{1.2}$$

where u_μ is the fluid 4-velocity, and ρ and p are the energy density and the pressure of the fluid, respectively.

The cosmological principle states that on very large scales the universe is homogeneous and isotropic. Applying these symmetries, the most general solution of Einstein's equation takes the name of Friedmann-Lemaître-Robertson-Walker (FLRW) metric. Defining the infinitesimal line element as $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$, the FLRW metric takes the form

$$\mathrm{d}s^2 = (c\ \mathrm{d}t)^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1 - kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)\right] \tag{1.3}$$

where r, θ and ϕ are the comoving coordinates, t is the proper time, $a(t)$ is the scale factor which describes the expansion of the universe, and k is the curvature parameter. Depending on the sign of the curvature parameter the universe is open ($k > 0$), flat ($k = 0$) or closed ($k < 0$). From the scale factor we can define the Hubble parameter as

$$H \equiv \frac{\dot{a}}{a} \tag{1.4}$$

where the dot indicates the derivative with respect to the proper time t. Edwin Hubble introduced the parameter that bears his name looking at the redshifts and the distances of different galaxies, which led him to the discovery of the expansion of the universe. We can define the redshift parameter as

$$z = \frac{\lambda_{\mathrm{obs}} - \lambda_{\mathrm{em}}}{\lambda_{\mathrm{em}}} \tag{1.5}$$

where λ_{em} and λ_{obs} are the wavelengths emitted by the source and observed on the Earth, respectively. The Doppler effect generates a shift in the wavelength depending on the velocity of the source, and it permits us to relate the redshift parameter to the velocity.

$$\frac{v}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1} \tag{1.6}$$

where v the radial speed of recession. Eq. (1.6), for non-relativistic velocities becomes

$$\frac{v}{c} \approx z\,. \tag{1.7}$$

Hubble found a relation of proportionality between the velocity and the distance of the objects observed, and the proportionality constant is the Hubble parameter:

$$\frac{v}{c} = H \cdot d \tag{1.8}$$

where d is the non-comoving distance (Carroll & Ostlie, 2007). When Hubble first proposed his law he didn't know that H is a time dependent parameter, and considered instead a constant H_0 which can be expressed in the form

$$H_0 = 100\ h\ \mathrm{km\ s^{-1}Mpc^{-1}} \tag{1.9}$$

which is nothing but the value of the Hubble parameter evaluated today $H_0 = H(t_0)$. The parameter h is a dimensionless parameter which incorporates the tension that regards the value of the Hubble parameter in the literature.

In the FLRW metric Einstein's equations take the form of the Friedmann equations:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2}\,, \tag{1.10}$$

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\left(\rho + 3\frac{p}{c^2}\right). \tag{1.11}$$

From Eq. (1.10) we can define the critical energy density ρ_c as the energy density that the universe would assume if it were flat ($k = 0$):

$$\rho_\mathrm{c} = \frac{3H^2}{8\pi G} \tag{1.12}$$

and from it we can define the density parameter

$$\Omega \equiv \frac{\rho}{\rho_\mathrm{c}} = \frac{8\pi G\rho}{3H^2}\,. \tag{1.13}$$

The conservation of the energy-momentum tensor in the FLRW metric for a perfect fluid becomes

$$\dot{\rho} = -3H\left(\rho + \frac{p}{c^2}\right). \tag{1.14}$$

In many cases we can express the equation of state of the fluid as $p = w\rho c^2$, where the parameter w depends on the fluid and lies in the range $0 < w < 1$.

The universe during its evolution has crossed different epochs, each one characterised by the domination of one component (radiation, matter, dark energy). In each epoch we can consider as a good approximation a mono-component universe, with $\Omega_i = 1$ where i indicates the specific constituent. Each component can be expressed as a perfect fluid with a different value of w: $w_\mathrm{m} = 0$ for matter, $w_\mathrm{r} = \frac{1}{3}$ for radiation and $w_\Lambda = -1$ for dark energy.

1.1.1 Einstein-de Sitter universe

Let us now consider the simplified case of a flat universe $k = 0$ with no cosmological constant $\Lambda = 0$, and composed only by matter ($w = 0$, $\Omega_{\mathrm{m}} = 1$). This solution is known as Einstein-de Sitter universe. In such a universe the solution of Eq. (1.14) is

$$\rho \propto a^{-3} \tag{1.15}$$

and from Eq. (1.10) we obtain

$$a \propto t^{\frac{2}{3}} \tag{1.16}$$

from which

$$H = \frac{2}{3t} \,. \tag{1.17}$$

These solutions will be useful later on, and they show that an Einstein-de Sitter universe will keep expanding forever, with a constant deceleration parameter $q \equiv -\frac{\ddot{a}a}{\dot{a}^2} = \frac{1}{2}$.

1.1.2 Standard model of cosmology

Mono-component universes as the Einstein-de Sitter one can be good approximations, but our universe is more complicated than that. The most accepted model that describes the universe is the flat-ΛCDM model, also called standard model of cosmology. The name itself tells us that we are considering a universe with zero curvature $k = 0$, and a non-null cosmological constant $\Lambda \neq 0$.

The standard model of cosmology takes into account the composition of the universe at $z = 0$, mainly dark energy and matter. The major component is dark energy, with $\Omega^0_\Lambda \approx 0.7$. The cosmological constant Λ is transposed to the right hand side of Eq. (1.1) and can be considered as a perfect fluid with energy density

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} \tag{1.18}$$

and pressure $p_\Lambda = w_\Lambda \rho_\Lambda c^2 = -\rho_\Lambda c^2$. We can see from Eq. (1.14) that the energy density of dark energy is constant in time. The matter component of the present universe has a density parameter of $\Omega^0_{\mathrm{m}} \approx 0.3$, of which only $\Omega^0_{\mathrm{b}} \approx 0.05$ is baryonic matter. Almost all matter of the universe is non-baryonic matter which interacts only gravitationally, called dark matter. Finally, the universe has a small component of radiation, consisting mainly of photons and neutrinos, with density parameter $\Omega^0_{\mathrm{r}} \approx 10^{-5}$. In this work we will neglect the radiation component, and consider

$$\Omega_\Lambda + \Omega_{\mathrm{m}} = 1 \,. \tag{1.19}$$

In such a composed universe we can write the Hubble parameter H as a function of the quantity of matter in the universe:

$$H = H_0 \sqrt{\Omega^0_{\mathrm{m}} \left(\frac{a}{a_0} \right)^{-3} + 1 - \Omega^0_{\mathrm{m}}} \,. \tag{1.20}$$

We can observe that the Hubble law (1.8), which with Eq. (1.7) relates redshifts and distances, depends on the parameter H, and therefore on the value of Ω_{m}.

1.2 Spherical evolution

The cosmological principle states that the universe is homogeneous. However this is valid only at very large scales, and the universe has inhomogeneities that lead to structure formation. All structures observed today, such as galaxies, cluster of galaxies and voids, have grown from small density perturbations in the matter field of the universe. Let us define the density contrast

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t)}{\bar{\rho}(t)} - 1 \tag{1.21}$$

where $\rho(\mathbf{x}, t)$ is the density field at time t and position $\mathbf{x}$, and $\bar{\rho}$ is the mean density of the universe. At early times, when perturbations are small, they can be described by linear perturbation theory. However, physical structures evolve a highly non-linear regime which cannot be described by linear theory and it requires a different analysis. To deal with non-linear gravitational dynamics analytically we need to make assumptions on the density distribution, and the simplest way to do so is to consider a spherically symmetric gravitational system. We need to remember that this model is not physically accurate, but it allows us to understand the processes and the dynamics involved. Considering a spherically symmetric density fluctuation $\delta(t, r)$ we can define the average density contrast

$$\Delta(r, t) = \frac{3}{r^3} \int_0^r \delta(r', t) r'^2 \mathrm{d}r' \, . \tag{1.22}$$

The total mass inside a sphere of radius r is then

$$\begin{aligned} M(r) &= \int_0^r \rho(r', t) 4\pi r'^2 \mathrm{d}r' \\ &= \int_0^r \bar{\rho}(t)(1 + \delta(r', t)) 4\pi r'^2 \mathrm{d}r' \\ &= \frac{4\pi}{3} r^3 \bar{\rho}[1 + \Delta(r)] \, . \end{aligned} \tag{1.23}$$

In linear theory, when the density contrast is small, we can factorize it into its temporal and spatial dependence as

$$\delta(\mathbf{x}, t) = D(t)\delta(\mathbf{x}) \tag{1.24}$$

where $D(t)$ is the linear growth factor, from which we can define the linear growth rate

$$f(t) \equiv \frac{\mathrm{d} \ln D(t)}{\mathrm{d} \ln a(t)} \, . \tag{1.25}$$

To a good approximation $f(t)$ can be related to Ω_m as $f(t) \approx \Omega_\mathrm{m}^{0.6}$ (Peebles, 1980). To linear order in the density contrast the conservation of mass implies

$$\begin{aligned}\frac{\partial\delta(\mathbf{x},t)}{\partial t}+\frac{1}{a(t)}\nabla\cdot\mathbf{v}(\mathbf{x},t)&=0\\&=\frac{dD(t)}{dt}\delta(\mathbf{x})+\frac{1}{a(t)}\nabla\cdot\mathbf{v}(\mathbf{x},t)\\&=\frac{\mathrm{d}\ln D(t)}{\mathrm{d}\ln a(t)}\frac{\mathrm{d}\ln a(t)}{\mathrm{d}t}\delta(\mathbf{x},t)+\frac{1}{a(t)}\nabla\cdot\mathbf{v}(\mathbf{x},t)\\&=f(t)H(t)\delta(\mathbf{x},t)+\frac{1}{a(t)}\nabla\cdot\mathbf{v}(\mathbf{x},t)\end{aligned}\tag{1.26}$$

where $\mathbf{v}(\mathbf{x},t)$ the peculiar velocity field. In the second row we have applied the factorization of Eq. (1.24), and in the third row the definition of the linear growth rate (1.25). Integrating over some volume, applying the divergence theorem and using the definition of the average density contrast (1.22) we can write the radial velocity in linear theory as

$$v(r,t)=-\frac{1}{3}f(t)a(t)H(t)r\Delta(r,t)\tag{1.27}$$

As a first analysis we will consider a matter dominated Einstein-de Sitter universe with no cosmological constant $\Lambda=0$ and $\Omega_\mathrm{m}=1$. In section (1.2.3) we will see the generalization to a ΛCDM cosmology. Let us now consider a test particle at a distance r from the center of a spherically symmetric density distribution. Birkhoff's theorem states that a spherically symmetric body affects other objects as if all its mass were concentrated at a point at its center. Our test particle then is affected only by the mass inside the sphere of radius r, $M(r)$. Newtonian gravitational theory leads to the equation of motion

$$\ddot{r}=-\frac{GM(r)}{r^2}\,.\tag{1.28}$$

Integrating this equation over time we obtain the energy equation

$$\frac{1}{2}\dot{r}^2-\frac{GM(r)}{r}=K\tag{1.29}$$

where the first term corresponds to kinetic energy, the second term to the potential gravitational energy and K is the integration constant. Replacing the mass $M(r)$ with the expression obtained in Eq. (1.23) we obtain

$$\dot{r}^2-\frac{8\pi G}{3}\bar{\rho}r^2[1+\Delta(r)]=2K\,.\tag{1.30}$$

Looking at the definition of Ω in Eq. (1.13) we can define the background matter density parameter $\Omega_\mathrm{m}=\frac{\bar{\rho}}{\rho_\mathrm{c}}$ and write Eq. (1.30) as

$$\dot{r}^2-\Omega_\mathrm{m}H^2r^2[1+\Delta(r)]=2K\,.\tag{1.31}$$

We can set the value of K from the initial conditions. The initial perturbation is small and evolves following linear theory, so we can use Eq. (1.27) to determine the initial velocity:

$$\dot{r}_i \simeq r_i H_i - \frac{1}{3} f_i H_i r_i \Delta_i(r_i) \tag{1.32}$$

where all quantities are calculated at the initial time and are characterised by the subscript i. Eq. (1.32) leads to the expression of the integration constant

$$2K = (r_i H_i)^2 \left[\left(1 - \frac{\Delta_i}{3}\right)^2 - (1 + \Delta_i) \right] \simeq -\frac{5}{3} (r_i H_i)^2 \Delta_i \,. \tag{1.33}$$

Eq. (1.23) allows us to relate the average density contrasts at general and initial time through mass conservation

$$1 + \Delta = (1 + \Delta_i) \frac{r_i^3 \bar{\rho}_i}{r^3 \bar{\rho}} \,. \tag{1.34}$$

From the definition of Ω_m we can write

$$\frac{\bar{\rho}_i}{\bar{\rho}} = \left(\frac{H_i}{H}\right)^2 . \tag{1.35}$$

From Eq. (1.31) and using Eq. (1.33), (1.34) and (1.35) we can finally obtain the dynamical equation of a spherically symmetric density perturbation

$$\left(\frac{\dot{r}}{r}\right)^2 = H_i^2 \left[(1 + \Delta_i) \left(\frac{r}{r_i}\right)^{-3} - \left(\frac{5}{3}\Delta_i\right) \left(\frac{r}{r_i}\right)^{-2} \right] . \tag{1.36}$$

We will now consider separately the overdensity and underdensity cases, defined by $\Delta_i > 0$ and $\Delta_i < 0$ respectively.

1.2.1 Overdensity

For $\Delta_i > 0$ we can solve Eq. (1.36) in parametric form

$$\frac{r}{r_i} = \frac{1}{2} \left(\frac{5}{3}\Delta_i\right)^{-1} (1 - \cos\eta) \tag{1.37}$$

$$H_i t = \frac{1}{2} \left(\frac{5}{3}\Delta_i\right)^{-\frac{3}{2}} (\eta - \sin\eta) \tag{1.38}$$

where the parameter η is defined as

$$\mathrm{d}\eta = \frac{r_i}{r} \sqrt{\frac{5}{3}\Delta_i} H_i \mathrm{d}t \,. \tag{1.39}$$

At $\eta_\mathrm{ta} = \pi$, when $\dot{r} = 0$, the radius of the overdensity reaches its maximum, and after that moment, called moment of turn-around, it decouples from the Hubble flow and it starts to shrink. We can derive the time of turn-around t_ta from Eq. (1.38).

Mass conservation of Eq. (1.34) in a matter dominated Einstein-de Sitter universe leads to

$$\begin{aligned} 1+\Delta &= (1+\Delta_i)\left(\frac{r}{r_i}\right)^{-3}\left(\frac{a}{a_i}\right)^3 \\ &= (1+\Delta_i)\frac{\left(\frac{3}{2}H_i t\right)^2}{\left(\frac{r}{r_i}\right)^3} \\ &= (1+\Delta_i)\frac{9(\eta-\sin\eta)^2}{2(1-\cos\eta)^3} \end{aligned} \tag{1.40}$$

where the first row derives from Eq. (1.34) applying Eq. (1.15) of an Einstein de Sitter universe, and in the second row we have applied properties (1.16) and (1.17). In the last row we have replaced the solutions of Eq. (1.37) and Eq. (1.38). Expanding the solution equation of the time Eq. (1.38) to first order we get

$$H_i t \simeq \frac{1}{2}\left(\frac{5}{3}\Delta_i\right)^{-\frac{3}{2}}\frac{\eta^3}{6} \tag{1.41}$$

and similarly the equation of mass conservation (1.40) leads to

$$\begin{aligned} 1+\Delta &\simeq (1+\Delta_i)\left(1+\frac{3}{20}\eta^2\right) \\ &\simeq (1+\Delta_i)\left[1+\Delta_i\left(\frac{3}{2}H_i t\right)^{\frac{2}{3}}\right] \end{aligned} \tag{1.42}$$

where we used the linearized solution Eq. (1.41). After the turn-around point the overdensity shrinks until it reaches the collapse, where all matter is concentrated in a single point. We can see from Eq. (1.37) that this happens at $\eta_{\rm c} = 2\pi$. At the moment of collapse, $t_{\rm c} = 2t_{\rm ta}$, the linear average density contrast will be

$$1+\Delta_c \simeq (1+\Delta_i)\left[1+\frac{3}{5}\left(\frac{6\pi}{4}\right)^{\frac{2}{3}}\right] \simeq 1+1.686\,. \tag{1.43}$$

We can then define $\delta_c = 1.686$ as the linear density threshold for collapse. Remember that this is the value of the average density contrast in linear theory at the moment of collapse, for a spherically symmetric density distribution.

1.2.2 Underdensity

In the case $\Delta_i < 0$ we have an underdensity of the matter density field. In this case the parametric solution of Eq. (1.36) is

$$\frac{r}{r_i} = \frac{1}{2}\left(\frac{5}{3}|\Delta_i|\right)^{-1}(\cosh\eta - 1) \tag{1.44}$$

$$H_i t = \frac{1}{2}\left(\frac{5}{3}|\Delta_i|\right)^{-\frac{3}{2}}(\sinh\eta - \eta) \tag{1.45}$$

where the parameter η is now defined as

$$\mathrm{d}\eta = \frac{r_i}{r} i \sqrt{\frac{5}{3}|\Delta_i|} H_i \mathrm{d}t \tag{1.46}$$

and it is imaginary. In this case the perturbation never reaches a point of turn-around and continues to expand. When two shells with different initial radius r_i cross each other, they reach a stage of non-linearity and we can call that moment as the formation of the void. From Eq. (1.44) and (1.45) we can derive the infinitesimal distance between shells and the infinitesimal time

$$\mathrm{d}r = \frac{1}{2}\left(\frac{5}{3}|\Delta_i|\right)^{-1}\left[(\cosh\eta - 1)\left(\mathrm{d}r_i - r_i\frac{\mathrm{d}\Delta_i}{\Delta_i}\right) + r_i \sinh\eta \ \mathrm{d}\eta\right] \tag{1.47}$$

$$\mathrm{d}t = \frac{1}{2H_i}\left(\frac{5}{3}|\Delta_i|\right)^{-\frac{3}{2}}\left[(\cosh\eta - 1)\mathrm{d}\eta - \frac{3}{2}(\sinh\eta - \eta)\frac{\mathrm{d}\Delta_i}{\Delta_i}\right]. \tag{1.48}$$

At the moment of shell-crossing the infinitesimal distance between shells vanishes at one moment in time $\mathrm{d}r = 0$ $\mathrm{d}t = 0$, which from Eq. (1.47) and (1.48) lead to

$$\frac{\mathrm{d}\ln\Delta_i}{\mathrm{d}\ln r_i} = \left[1 - \frac{3}{2}\frac{\sinh\eta(\sinh\eta - \eta)}{(\cosh\eta - 1)^2}\right]^{-1}. \tag{1.49}$$

The shell-crossing condition depends on the particular density distribution that we choose. From the definition of the average density contrast (1.22) we can derive

$$\frac{\mathrm{d}\ln\Delta}{\mathrm{d}\ln r} = 3\left(\frac{\delta(r)}{\Delta(r)} - 1\right). \tag{1.50}$$

Let us now assume an inverted top-hat density distribution, described by

$$\delta_i(r_i) = \begin{cases} \delta_0 & r_i < r_0 \\ 0 & r_i \geq r_0 \end{cases} \tag{1.51}$$

from which from Eq. (1.22) we get

$$\Delta_i(r_i) = \begin{cases} \delta_0 & r_i < r_0 \\ \delta_0\left(\frac{r_0}{r_i}\right)^3 & r_i \geq r_0 \end{cases} \tag{1.52}$$

which leads to

$$\frac{\mathrm{d}\ln\Delta_i}{\mathrm{d}\ln r_i} = \begin{cases} 0 & r_i < r_0 \\ -3 & r_i \geq r_0\,. \end{cases} \tag{1.53}$$

In the case of a spherical inverted top-hat distribution the shell-crossing condition (1.49) becomes

$$\frac{\sinh\eta(\sinh\eta-\eta)}{(\cosh\eta-1)^2}=\begin{cases}0 & r_i<r_0\\ \frac{8}{9} & r_i\geq r_0\,.\end{cases} \tag{1.54}$$

At $r_i<0$ the only solution is the trivial one $\eta=0$, while in the $r_i\geq r_0$ case the solution is $\eta_{\rm sc}\simeq 3.488$, and the condition is first satisfied at the boundary shell at $r_i=r_0$. As we did in the overdensity case, from the equation of mass conservation (1.34) we can derive the average density contrast as a function of the initial density contrast and the parameter η

$$1+\Delta=(1+\Delta_i)\frac{9(\sinh\eta-\eta)^2}{2(\cosh\eta-1)^3}\,. \tag{1.55}$$

At the shell-crossing point $\eta_{\rm sc}=3.488$ the average density contrast is then

$$1+\Delta_{\rm sc}\simeq 0.2047\simeq 1-0.8 \tag{1.56}$$

from which we can define the non-linear density threshold for void formation $\Delta_{\rm sc}\simeq -0.8$. At the moment of shell-crossing the comoving radius of the underdensity has expanded by a factor $(1+\Delta_{\rm sc})^{-1/3}\simeq 1.7$. Expanding to linear order Eq. (1.55) leads to the linear average density contrast

$$1+\Delta\simeq(1+\Delta_i)\left[1-|\Delta_i|\left(\frac{3}{2}H_it\right)^{\frac{2}{3}}\right]=(1+\Delta_i)\left[1-\frac{3}{20}6^{\frac{2}{3}}(\sinh\eta-\eta)^{\frac{2}{3}}\right] \tag{1.57}$$

which, calculated at the shell-crossing point $\eta_{\rm sc}\simeq 3.488$ results in

$$1+\Delta_{\rm sc}\simeq 1-2.717\,. \tag{1.58}$$

We can define the linear density contrast for void formation δ_V as the density contrast at the moment of shell-crossing in linear theory, and its value in the case of a spherical inverted top-hat density distribution is $\delta_V=-2.717$.

1.2.3 ΛCDM generalization

In a general ΛCDM cosmology the motion of the test particle in a spherically symmetric density distribution is given by

$$\ddot{r}=-\frac{GM}{r^2}+\frac{\Lambda}{3}r \tag{1.59}$$

from which, following Mo et al. (2010), we can derive the dependence of the linear density threshold for collapse upon $\Omega_{\rm m}$:

$$\delta_c\approx 1.686\ \Omega_{\rm m}^{0.0055}\,. \tag{1.60}$$

In the case of the standard model for cosmology, considering $\Omega_{\rm m}=0.3$, we obtain $\delta_c=1.674$. We can see that there is a small change in the value of δ_c going from an

Einstein-de Sitter to a ΛCDM universe. Similarly, the value of δ_V goes from -2.717 in an Einstein-de Sitter universe to $\delta_V = -2.731$ in a ΛCDM universe (Jennings et al., 2013). In the following calculations and analysis we will consider the parameters corresponding to the standard model of cosmology.

1.3 Excursion set theory

In the previous section we considered perturbations of the density field of the universe, that can fluctuate on arbitrarily small scales. Excursion set theory allow us to smooth the density perturbations at a cutoff scale R_S making use of a window function. The smoothed density contrast will be

$$\delta_{R_\mathrm{S}}(\mathbf{x}) = \int \delta(\mathbf{x}')W_{R_\mathrm{S}}(\mathbf{x}-\mathbf{x}')\mathrm{d}^3x' \tag{1.61}$$

where W_{R_S} is the window function. Different choices are possible, such as the Gaussian window in real or Fourier space. We will use a top-hat window function in real space:

$$W_{R_\mathrm{S}}(\mathbf{x}) = \begin{cases} \frac{3}{4\pi R_\mathrm{S}^3} & |\mathbf{x}| < R_\mathrm{S} \\ 0 & |\mathbf{x}| \geq R_\mathrm{S} \end{cases} \tag{1.62}$$

normalized such that $\int W_{R_\mathrm{S}}(\mathbf{x})\mathrm{d}^3x = 1$. To simplify the following calculation we will analyse it in Fourier space. The Fourier transform of the density contrast is defined as

$$\delta(\mathbf{k}) = \int \delta(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}\mathrm{d}^3x\,. \tag{1.63}$$

The convolution of Eq. (1.61) then becomes a simple multiplication

$$\delta_{R_\mathrm{S}}(\mathbf{k}) = \delta(\mathbf{k})W_{R_\mathrm{S}}(\mathbf{k}) \tag{1.64}$$

where $W_{R_\mathrm{S}}(\mathbf{k})$ is the Fourier transform of the top-hat window function:

$$W_{R_\mathrm{S}}(\mathrm{k}) = \frac{3}{(kR_\mathrm{S})^3}[\sin(kR_\mathrm{S}) - kR_\mathrm{S}\cos(kR_\mathrm{S})]\,. \tag{1.65}$$

Let us define the two-point correlation function in real space as the joint ensamble average of the density contrast at two points separated by a distance $\mathbf{r}$:

$$\xi(r) \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \tag{1.66}$$

which depends only on the magnitude of the distance, due to the properties of homogeneity and isotropy. The analogous in Fourier space will be

$$\begin{aligned}\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle &= \int\int \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}}e^{-i\mathbf{k}'\cdot\mathbf{r}}\mathrm{d}^3x\ \mathrm{d}^3r \\ &= \int \xi(r)e^{-i\mathbf{k}'\cdot\mathbf{x}}\mathrm{d}^3r \int e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}}\mathrm{d}^3x \\ &\equiv P(\mathbf{k}')(2\pi)^3\delta_D(\mathbf{k}+\mathbf{k}')\end{aligned} \tag{1.67}$$

where $\delta_D(\mathbf{k}+\mathbf{k}')$ is the Dirac delta function, and in the last row we have defined the power spectrum as the Fourier transform of the two-point correlation function

$$P(k) = \int \xi(r) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3 r \tag{1.68}$$

that, like the two-point correlation function, depends only on the magnitude of $\mathbf{k}$. We can now define the variance of the density field as the two-point correlation function of the smoothed density contrast calculated at a distance $\mathbf{r} = 0$

$$\begin{aligned} \sigma^2_{R_\mathrm{S}} &\equiv \langle \delta^2_{R_\mathrm{S}}(\mathbf{x}) \rangle \\ &= \xi_{R_\mathrm{S}}(r=0) \\ &= \frac{1}{(2\pi)^3} \int P_{R_\mathrm{S}}(k) \mathrm{d}^3 k \\ &= \frac{1}{2\pi^2} \int P(k) W^2_{R_\mathrm{S}}(k) k^2 \mathrm{d}k \,. \end{aligned} \tag{1.69}$$

From now on we will consider the variance as the quantity that identifies the smoothing scale R_S.

1.3.1 First-crossing distribution function

Let us consider a sequence in the initial Gaussian random field δ_{R_S} where R_S is decreased stepwise from a large initial value. Fig. (1.1) shows in the left-most panel of each set the smoothed density contrast as a function of the variance at a given cutoff scale σ_{R_S}. We can see that the density contrast performs random walks (Bond et al., 1991), and each set shows a different path. The two other panels of each set show the structure evolution of each configuration.

The first set represents the cloud-in-cloud process. The density contrast crosses the linear threshold for collapse δ_c at two different scales: in the central panel we can see different clouds of matter, which will collapse in one bigger cloud (last panel).

The second set shows the cloud-in-void process. The density contrast crosses the linear threshold for collapse at a small scale, while it crosses the linear threshold for void formation δ_V at larger scales. This represents the presence of a cloud inside a larger void.

The third set shows the void-in-void process. The density contrast crosses twice the threshold δ_V. In the central panel we can see a hierarchy of voids at different scales, which will end up merging into one large void (similarly to the cloud-in-cloud effect).

The last set shows the void-in-cloud process. Here the density contrast crosses the threshold for void formation δ_V at a smaller scale than the one in which it crosses the threshold for collapse δ_c. This sets represents the presence of a void inside a matter halo, as we can see in the central panel. While the larger matter cloud collapses the void is squeezed at smaller and smaller sizes. When the halo has collapsed the void no longer exists (Sheth & van de Weygaert, 2004).

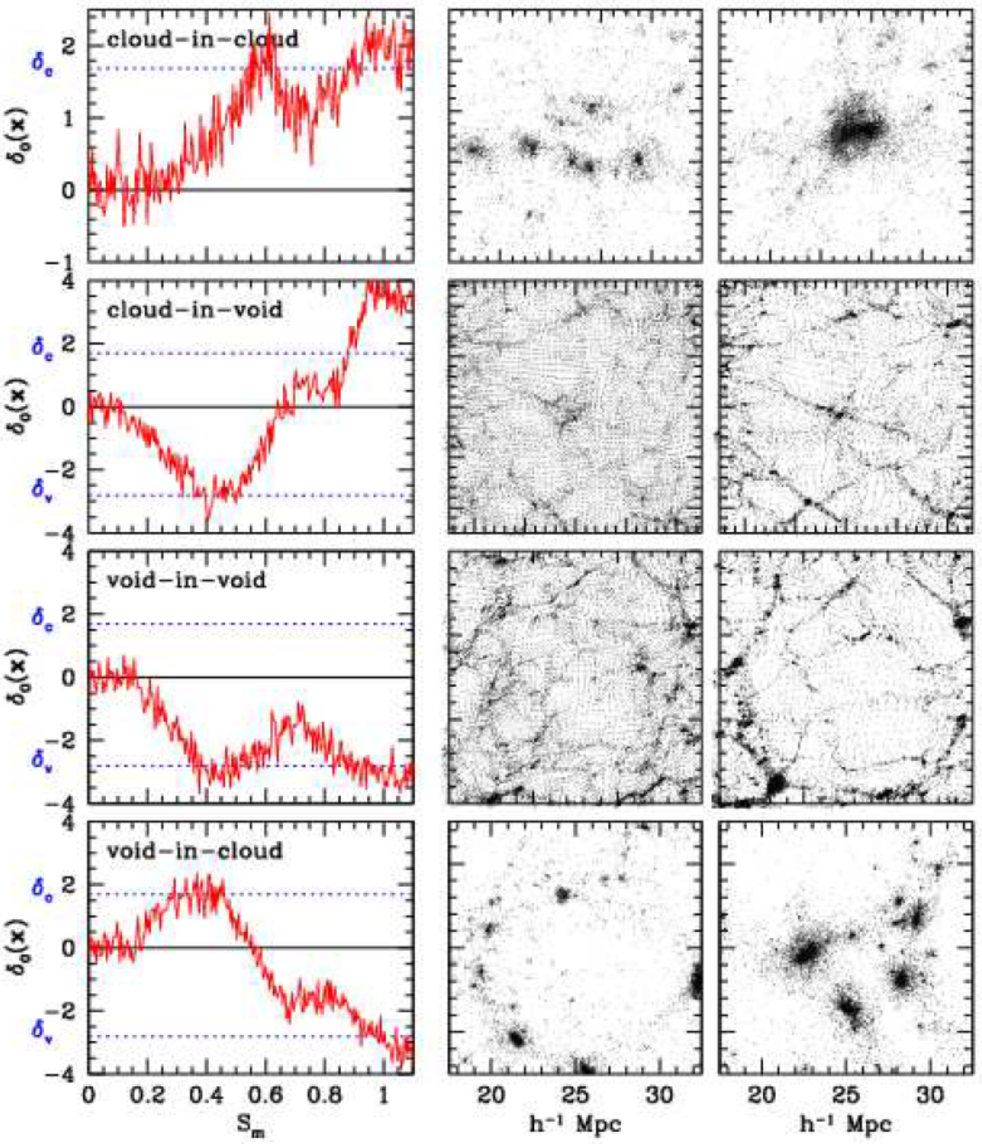


Figure 1.1: Random walks of the density contrast and corresponding structures and evolution in four cases: cloud-in-cloud, cloud-in-void, void-in-void and void-in-cloud processes. *Left panels*: Random walks of the density contrast δ_{R_S} as a function of the variance $S_\mathrm{m} = \sigma^2_{R_\mathrm{S}}$. *Central panels*: Structure corresponding to the path of the random walk in the corresponding left panel. *Right panels*: Evolution of the structure of the corresponding central panel. Adopted from (Sheth & van de Weygaert, 2004).

We can quantify the number of voids defining the first-crossing distribution function as the fraction of random walks which cross the threshold for void formation δ_V and had not crossed the threshold for collapse at larger scales:

$$f(\sigma_{R_\mathrm{S}}, \delta_V, \delta_c) = f(\sigma_{R_\mathrm{S}}, \delta_V) - \int_0^{\sigma^2_{R_\mathrm{S}}} f(\sigma_{R'_\mathrm{S}}, \delta_c) f(\sigma_{R_\mathrm{S}}, \delta_V | \sigma_{R'_\mathrm{S}}, \delta_c) \mathrm{d}\sigma^2_{R'_\mathrm{S}} \,. \tag{1.70}$$

The second term on the right hand side represents the void-in-cloud process, which has been subtracted to the fraction of random walks that cross δ_V. Defining the void-and-cloud parameter as

$$D \equiv \frac{|\delta_V|}{\delta_c + |\delta_V|} \tag{1.71}$$

and the variable

$$\nu \equiv \frac{|\delta_V|}{\sigma_{R_\mathrm{S}}} \tag{1.72}$$

the solution of Eq. (1.70) is found using Laplace transforms (Sheth & van de Weygaert, 2004) and can be expressed in terms of these new quantities as:

$$\nu f(\nu) = \sum_{j=1}^{\infty} \frac{2j\pi D^2}{\nu^3} \sin(j\pi D) e^{-\frac{(j\pi D)^2}{2\nu^2}} \tag{1.73}$$

where we have dropped the explicit dependence on the parameters δ_V and δ_c. Sheth and van de Weygaert (Sheth & van de Weygaert, 2004) proposed an approximation form of Eq. (1.73), valid for $\frac{\delta_c}{|\delta_V|} \gtrsim \frac{1}{4}$:

$$\nu f(\nu) \approx \sqrt{\frac{2}{\pi}} \nu e^{-\frac{\nu^2}{2}} e^{\left[-\frac{|\delta_V|}{\delta_c}\left(\frac{D}{2\nu}\right)^2 - 2\left(\frac{D}{\nu}\right)^4\right]} \tag{1.74}$$

where the second exponential represents the void-in-cloud term. In fig. (1.2) we can see the comparison between the two forms of the first-crossing distribution function, i.e. the general form (blue line) and Sheth-van de Weygaert approximation (orange line) as a function of ν. We can see the comparison of the two functions in the case of four different sets of parameters. Note that the two forms are quite similar to each other when the parameters sets are inside the regime of validity of the approximation, while in the upper-left plot, where the parameters are outside that regime, the two functions differ both in magnitude and in the value of ν that corresponds to the peak.

It will be useful to define also an "hybrid" model (Jennings et al., 2013):

$$\nu f(\nu) = \begin{cases} \sum_{j=1}^{\infty} \frac{2j\pi D^2}{\nu^3} \sin(j\pi D) e^{-\frac{(j\pi D)^2}{2\nu^2}} & x > 0.276 \\ \sqrt{\frac{2}{\pi}} \nu e^{-\frac{\nu^2}{2}} e^{\left[-\frac{|\delta_v|}{\delta_c}\left(\frac{D}{2\nu}\right)^2 - 2\left(\frac{D}{\nu}\right)^4\right]} & x \leq 0.276 \end{cases} \tag{1.75}$$

where $x = \frac{D}{\nu}$. The reason of the utility of this function will be clarified in Chapter 4.

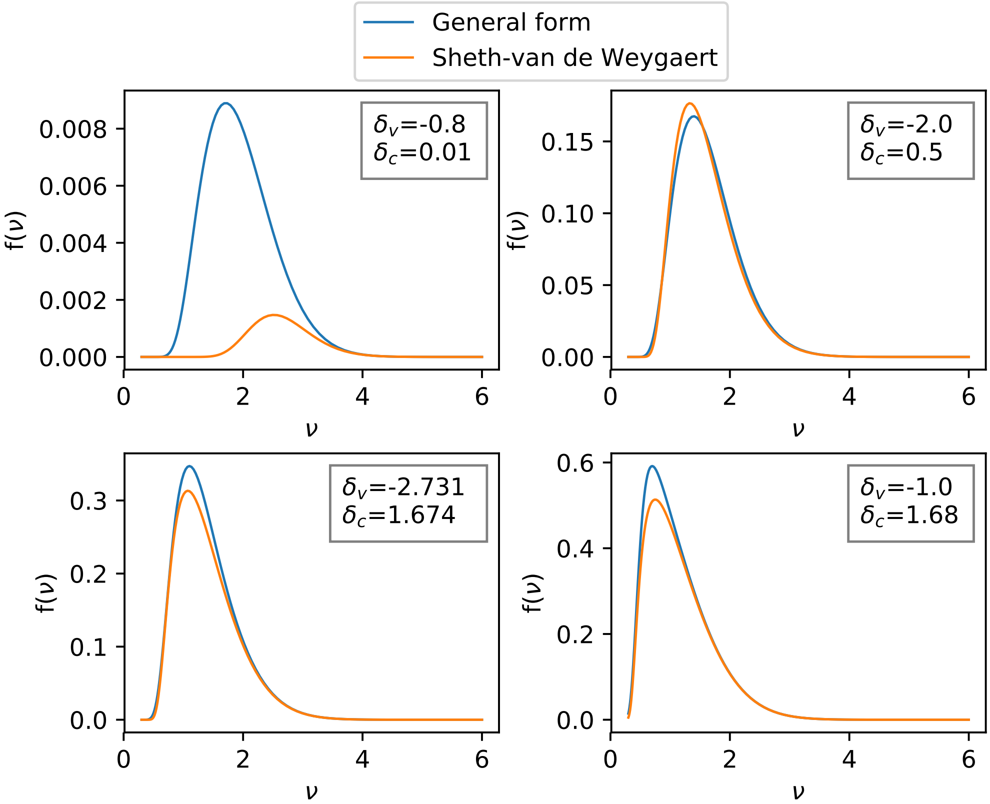


Figure 1.2: Comparison between the general solution of the first-crossing distribution function (blue line) and the Sheth-van de Weygaert approximation (orange line) in the case of four different sets of parameters δ_V and δ_c.

1.3.2 Void size function

A better quantity to express the abundance of voids that can be compared with data is the void size function, which expresses the number density of voids per logarithmic bin in the void radius $\frac{\mathrm{d}n}{\mathrm{d}\ln R}$. We have seen that during spherical evolution a void expands by a factor independent of its initial size. We can then define a Lagrangian void radius

$$R_L = r_s \cdot R \tag{1.76}$$

as the Eulerian one R multiplied by a rescaling constant r_s. We have seen in section (1.2.2) that when it reaches the shell-crossing condition the void has expanded by a factor $(1+\Delta_{\mathrm{sc}})^{-1/3} \simeq 1.7$, which implies a rescaling factor $r_s \simeq 0.58$ in the case of a spherical inverted top-hat distribution (Chan et al., 2014).

The first model of the void size function was proposed by Sheth and van de Weygaert (Sheth & van de Weygaert, 2004), and it is based on the conservation of the comoving number density of voids during their evolution and the transition from linear to non-linear theory

$$n = n_L \tag{1.77}$$

which, along with the definition of the Lagrangian void radius (1.76), implies

$$\frac{\mathrm{d}n}{\mathrm{d}\ln R} = \frac{\mathrm{d}n_L}{\mathrm{d}\ln R_L}\,. \tag{1.78}$$

The void size function in the Sheth-van de Weygaert model can be written in the form

$$\begin{aligned}\frac{\mathrm{d}n}{\mathrm{d}\ln R} &= \frac{1}{V}\left(\frac{R}{R_L}\right)^3 \nu f(\nu)\frac{\mathrm{d}\ln\nu}{\mathrm{d}\ln R}\\ &= \frac{1}{V}\frac{1}{r_s^3}\nu f(\nu)\frac{\mathrm{d}\ln\nu}{\mathrm{d}\ln R}\end{aligned} \tag{1.79}$$

where V is the void volume

$$V(R) = \frac{4\pi}{3}R^3\,. \tag{1.80}$$

Sheth-van de Weygaert model however presents the issue that the cumulative volume fraction of voids larger than a value R_V

$$\mathcal{F}(R_V) = \int_{R_V}^{\infty}\frac{\mathrm{d}r}{r}V(r)\frac{\mathrm{d}n}{\mathrm{d}\ln r} \tag{1.81}$$

exceeds unity. In order to overcome this problem Jennings, Li and Hu (Jennings et al., 2013) proposed a new model of the void size function that takes into account the fact that during their evolution voids can merge together. The conserved quantity, then, is not the number density of voids, but the total volume of voids:

$$\mathcal{F}(R) = \mathcal{F}(R_L) \tag{1.82}$$

that, from the definition of $\mathcal{F}$ Eq. (1.81), implies

$$V(R)\mathrm{d}n(R) = V(R_L)\mathrm{d}n(R_L)\,. \tag{1.83}$$

This model is the so-called Vdn model, and it modifies the void size function of Eq. (1.79) as

$$\begin{aligned} \frac{\mathrm{d}n}{\mathrm{d}\ln R} &= \frac{\mathrm{d}n_L}{\mathrm{d}\ln R_L}\left(\frac{V_L}{V}\right) \\ &= \frac{\mathrm{d}n_L}{\mathrm{d}\ln R_L} r_s^3 \\ &= \frac{1}{V}\nu f(\nu)\frac{\mathrm{d}\ln\nu}{\mathrm{d}\ln R}\,. \end{aligned} \tag{1.84}$$

In Fig. (1.3) we can see the comparison between the Vdn model and Sheth-van de Weygaert model. Each column shows a different set of parameters, while each row shows the models calculated using the two different forms of the first-crossing distribution function: the first row considers the general form, while the second row adopts the Sheth-van de Weygaert approximation.

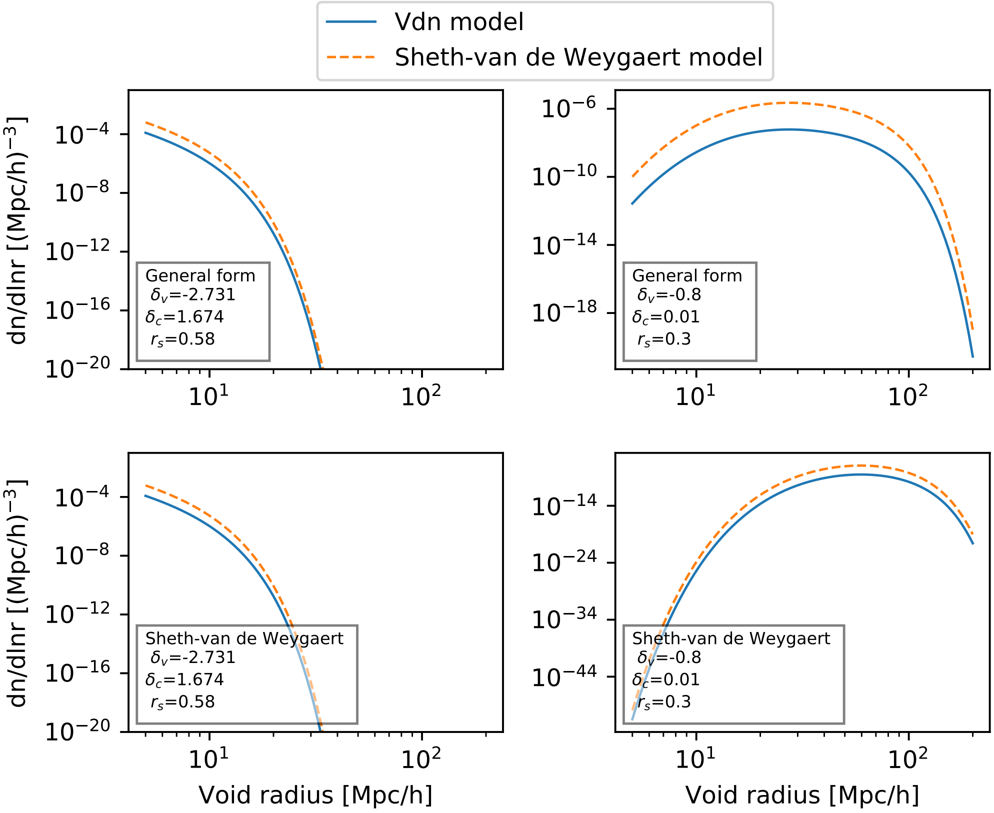


Figure 1.3: Comparison between Vdn model (blue continuous line) and the Sheth-van de Weygaert model (orange dashed line) in the case of two different sets of parameters δ_V and δ_c and the two different first-crossing distribution functions, i.e. the general form and Sheth-van de Weygaert approximation.

Chapter 2

Data set

In this work we will use a mock catalog of the twelfth release of the Baryon Oscillation Spectroscopic Survey (BOSS)(Dawson et al., 2013), a branch of the Sloan Digital Sky Survey (SDSS). The Sloan Digital Sky Survey has carried out imaging and spectroscopy over one third of the Celestial Sphere. It started routing operations in 2000, April, using a dedicated 2.5-meter wide-field telescope instrumented with a sequence of sophisticated imagers and spectrographs. We are interested in SDSS-III (Alam et al., 2015), which started operations in Fall 2008, completing in Summer 2014.

BOSS is a survey designed to measure the universe using baryon acoustic oscillations (BAO), a feature imprinted on the clustering of matter by acoustic waves that propagate in the pre-recombination universe. Sound waves that propagate in the early Universe, like spreading ripples in a pond, imprint a characteristic scale on cosmic microwave background fluctuations. These fluctuations have evolved into today's distribution of walls and voids of galaxies, meaning this baryon acoustic oscillation (BAO) scale (about 150 Mpc) is visible among galaxies today. BOSS project used the 2.5-meter aperture Sloan Foundation Telescope at Apache Point Observatory. The telescope used a drift-scanning mosaic CCD-camera with five color-bands. Spectra are obtained using the double-armed BOSS spectrographs, covering the wavelength range 3600-10000 Å with a resolving power of 1500 to 2600. BOSS provides redshifts for 1.5 million galaxies divided into two samples: LOWZ, which selected the brightest and reddest galaxies of the low-redshift galaxy population ($z \lesssim 0.4$), and CMASS, which isolates galaxies at higher redshift ($z \gtrsim 0.4$) (Rodríguez-Torres et al., 2016).

The mock galaxy catalog that we will use in this work is the MultiDark Patchy mock catalog, and it has been constructed using the code PATCHY (Kitaura et al. (2016), Rodríguez-Torres et al. (2016)). In the next section we will briefly see its construction.

2.1 MultiDark-Patchy mock galaxy catalog

The Baryon Oscillation Spectroscopic Survey scanned the sky with great accuracy, but the analysis of such survey arises a question: if the universe is comparable with an unique huge experiment, how can we determine the uncertainties in the measurements

Parameter	Value
$\Omega_{\rm m}$	0.307115
Ω_Λ	0.692885
$\Omega_{\rm b}$	0.048
σ_8	0.8288
h	0.6777

Table 2.1: Cosmological parameters adopted in the construction of MultiDark-Patchy mocks.

derived from observing it? To solve this issue, during the past decades there has been an effort to encode our knowledge of structure formation in computational algorithms, and compare the models to the observations.

The PATCHY method, that has been adopted in the construction of the catalog we will use, relies on the large-scale density field obtained from approximate gravity solvers and uses a biasing prescription to populate it with mock galaxies. To construct high-fidelity mock light cones for interpreting the BOSS DR11 &DR12 galaxy clustering, an iterative training procedure has been applied, in which a reference catalog is statistically reproduced with approximate gravity solvers and analytical-statistical biasing models. The five steps that summarise the catalog construction are reported in the following subsections. The cosmological parameters used in the construction of the catalog are show in Table (2.1).

2.1.1 First step: generation of a reference catalog

The first step consists in the generation of an accurate reference catalog. In this case the reference catalog is extracted from the BigMDPL simulation, one of the MultiDark N-body simulations described by Kyplin et al. (2016). The BigMDPL is performed with GADGET-2 code (Springel, 2005), with 3840^3 particles on a volume of $(2.5h^{-1}{\rm Mpc})^3$, assuming a ΛCDM cosmology with the cosmological parameters of Tab. (2.1). Haloes are defined based on the Bound Density Maximum halo finder (Kyplin & Holzman, 1997).

To populate the haloes with galaxies the Halo Abundance Matching (HAM) technique (see e.g. Nuza et al. (2013)) has been used. The basic assumption of the HAM method is that massive haloes host massive galaxies. Since observations show that this cannot be a one-to-one relation, it is necessary to include scatters. Therefore the maximum circular velocity $V_{\rm peak}$ of each object has been modified adding a gaussian noise:

$$V_{\rm peak}^{\rm scat} = (1 + \mathcal{N}(0, \sigma_{\rm HAM}))V_{\rm peak} \tag{2.1}$$

where $\mathcal{N}$ is a random number produced with a gaussian distribution with zero mean and standard deviation $\sigma_{\rm HAM}$. Then, the catalog has been sorted by $V_{\rm peak}^{\rm scat}$ starting from the object with the largest velocity. The catalog has been used to construct the cumulative number density of the haloes as a function of $V_{\text{peak}}^{\text{scat}}$. The cumulative number density of galaxies has been concurrently computed from the Stellar Mass Function (Rodríguez-Torres et al., 2016). Finally, a monotonic relation between the two cumulative number densities has been constructed:

$$n_{\text{gal}}(> M_*^i) = n_{\text{halo}}(> V_{\text{peak},i}^{\text{scat}}) \tag{2.2}$$

which implies that a halo with $V_{\text{peak},i}^{\text{scat}}$ will contain a galaxy with stellar mass M_*^i.

2.1.2 Second step: PATCHY code

In the second step the PATCHY code (**P**erturb**A**tion **T**heory **C**atalog generator of **H**alo and galax**Y** distributions) is trained to match the two- and three-point clustering of the full mock galaxy catalogues for each redshift bin. To simulate structure formation augmented Lagrangian Perturbation Theory (ALPT) has been adopted. Let us first define the displacement field $\mathbf{\Psi}(\mathbf{q}, z)$, which maps a distribution of dark matter particles at initial Lagrangian position $\mathbf{q}$ to the final Eulerian position $\mathbf{x}(z)$ at redshift z:

$$\mathbf{x}(z) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, z)\,. \tag{2.3}$$

In this approximation the displacement field is split into a short-range component $\mathbf{\Psi}_{\text{S}}(\mathbf{q}, z)$, calculated using spherical collapse approximation, and a long-range component $\mathbf{\Psi}_{\text{L}}(\mathbf{q}, z)$. The latter is calculated relying on second order Lagrangian Perturbation Theory (2LPT), and then filtered with a Gaussian kernel $\mathcal{K}(\mathbf{q}, r_{\text{S}}) = \exp(-|\mathbf{q}|^2/(2r_{\text{S}}^2))$, where r_{S} is the smoothing radius. The ALPT displacement is then

$$\mathbf{\Psi}_{\text{ALPT}}(\mathbf{q}, z) = \mathcal{K}(\mathbf{q}, r_{\text{S}}) \circ \mathbf{\Psi}_{\text{2LPT}} + (1 - \mathcal{K}(\mathbf{q}, r_{\text{S}})) \circ \mathbf{\Psi}_S \tag{2.4}$$

and it is used to move a set of homogeneously distributed particles from Lagrangian initial conditions to the Eulerian final ones. Then, using a clouds-in-cell scheme, the particles have been gridded to produce a smooth density field δ_{ALPT}.

The galaxies are then populated on the mesh according to the bias model, which is composed by different contributions:

- *Deterministic bias*:

  It relates the expected number count of haloes or galaxies $\rho_{\text{g}} \equiv \langle N_{\text{g}} \rangle_{\partial V}$ at a given finite volume to the underlying dark matter field ρ_{M}, where $\langle ... \rangle_{\partial V}$ is the ensemble average over the differential volume element ∂V.

  The bias relation can be arbitrarily complex $\rho_{\text{g}} = B(\rho_{\text{M}})$. The one used in the construction of MultiDark Patchy mocks is composed by a threshold bias, the local bias expansion, and a bias which compensates for the missing power of PT based methods.

- *Stochastic bias*:

  The halo distribution is a discrete sample of the continuous underlying dark matter distribution. To account for shot noise one could do Poissonian realizations of the halo density field, but the excess probability of finding haloes in high density regions generates over-dispersion.

The strategy has been to generate a mock catalog which reproduces the clustering of the whole population of galaxies at a given redshift. Mixing massive and low mass galaxies the over-dispersion has a dominant role, and can be modelled with a negative binomial probability distribution function, including an additional parameter to account for over-dispersion (Kitaura et al., 2016).

2.1.3 Third step: HADRON code

Once we have a spatial distribution of objects which accurately reproduce the clustering of the whole galaxy sample at a given redshift, we assign the halo/stellar masses to each object according to the statistical information extracted from the BigMultiDark simulation using the **H**alo m**A**ss **D**istribution **R**econstructi**ON** (HADRON) code (Zhao et al., 2015). The main steps of the HADRON code are:

- Compute the density field and cosmic web structures according to the dark matter particles from the reference N-body simulation. Then, classify the knots into different classes according to their enclosed mass, obtaining the density ρ_{DM} and cosmic web classification type t_{CW} for each cell.

- Compute the number of haloes in each density and cosmic web classification type bin according to the halo catalog from the reference N-body simulation.

- Taking the dark matter particles according to the approximate catalog mock generator, compute the density and cosmic web type.

- Apply halo exclusion to the halo catalog from the mock generator, assigning mass above the threshold to haloes.

- Assign mass to the rest of the haloes. Find the local density ρ_{DM} and cosmic web classification type t_{CW} for each mock halo, and assign the mass with the probability from the corresponding distribution of halo masses with density ρ_{DM} and type t_{CW}.

2.1.4 Fourth step: SUGAR code

The fourth step consists in the construction of light cones applying the **SU**rvey **G**ener**A**to**R** (SUGAR) code (Rodríguez-Torres et al., 2016). The SUGAR code works with cubic boxes using positions and velocities of dark matter haloes as inputs. The first step is to locate the observer ($z = 0$) and transform from comoving cartesian coordinates to equatorial ones (right ascension and declination) and redshift. In order to include the effects of galaxy peculiar velocities the transform of halo coordinates into redshift space is

$$\mathbf{s} = \mathbf{r}_{\mathrm{c}} + \frac{\mathbf{v} \cdot \hat{\mathbf{r}}}{aH(z_{\mathrm{real}})} \tag{2.5}$$

where $\mathbf{r}_\mathrm{c}$ is the comoving distance in real space, $\mathbf{v}$ is the velocity of the object with respect to the Hubble flow, $\hat{\mathbf{r}}$ is the line of sight direction, a is the scale factor and $H(z_\mathrm{real})$ the Hubble parameter at the redshift corresponding to $\mathbf{r}_\mathrm{c}$, computed from

$$r_\mathrm{c}(z_\mathrm{real}) = \int_0^{z_\mathrm{real}} \frac{c\mathrm{d}z}{H_0\sqrt{\Omega_\mathrm{m}(1+z)^3+\Omega_\Lambda}} \tag{2.6}$$

where c is the speed of light and H_0 the Hubble constant. From Eq. (2.5) and Eq. (2.6) it is possible to compute $s(z_\mathrm{obs})$, where z_obs is the observed redshift. In order to construct light-cones from snapshots, an object with redshift z_obs from a snapshot with redshift z_i will be selected if

$$\frac{z_i+z_{i-1}}{2} < z_\mathrm{obs} < \frac{z_i+z_{i+1}}{2}\,. \tag{2.7}$$

Finally, in order to complete the mock catalog, and angular mask is applied to match the area of the selected sample.

2.1.5 Fifth step: comparison with observations

Once the mock catalogs are complete they are compared to observations. First, in configuration space the two- and three-point correlation functions has been calculated, followed by the monopole and quadrupole. Then, the monopole and quadrupole have been analyzed in Fourier space. Finally, the cosmic evolution modelled in the MultiDark PATCHY mocks has been compared to the one of the observations. All the process is iterated until the desired accuracy for the statistical measures is reached.

In figure (2.1) a visualisation of the BOSS DR12 and one MultiDark PATCHY mock is shown. We can see that both the data and the mocks follow the same selection criteria, including the survey mask. The empty regions seem to be similarly distributed in the two cases. The color code stands for the stellar mass (Kitaura et al., 2016).

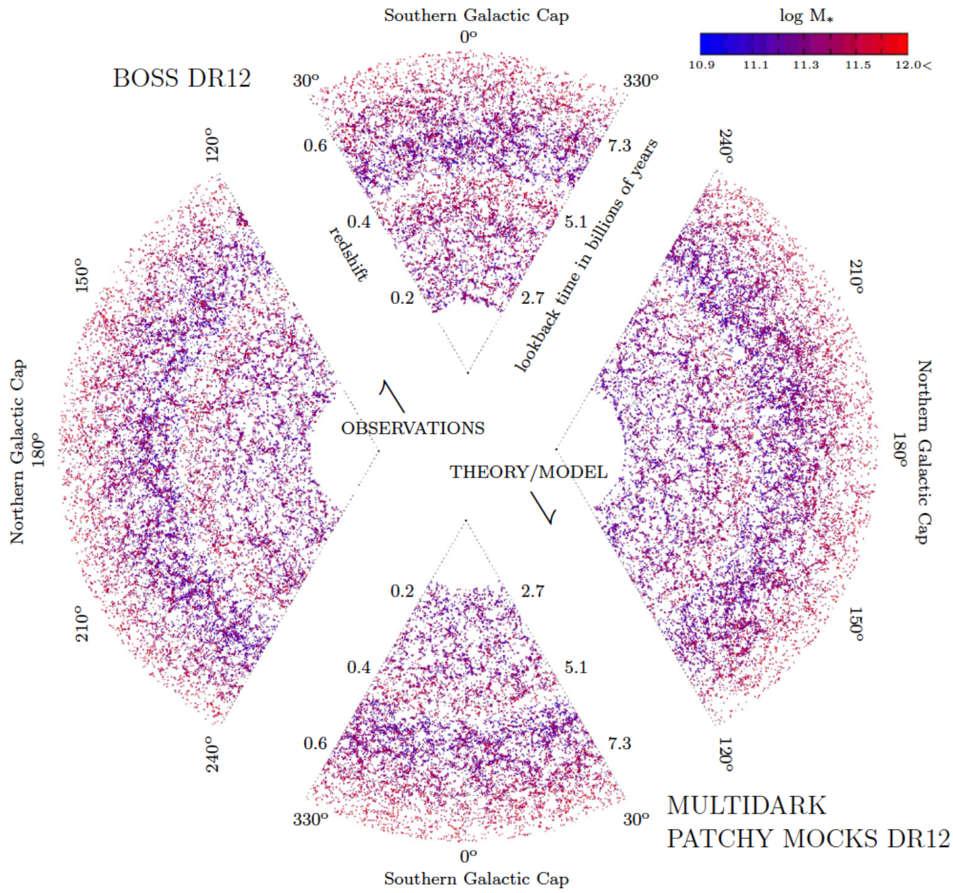


Figure 2.1: Pie plot of the BOSS DR12 observations (upper left region), and one MultiDark patchy mock realization (lower right region), from Kitaura et al. (2016)

Chapter 3

Method

In order to step from the theory we have seen in Chapter 1 to the results of the analysis of Chapter 4, we need to take some time to understand the methods that we will use in order to obtain valid results. In this chapter, therefore, we will see the methods that will be used later on. To find cosmic voids from the MultiDark PATCHY mock galaxy catalog of the BOSS SDSS DR12 (see Chapter 2) we will use the VIDE code (Sutter et al., 2015). VIDE code is described in Section 3.1. In order to perform fits and search for the optimal parameters of the model which describe the data we will first apply the Least Squares method (Section 3.2). As we will see in Section 4.2 another approach will be needed, and the fits will be performed using a Markov Chain Monte Carlo method (Section 3.3). The code that will be used is the emcee code (Foreman-Mackey et al., 2013), whose main steps are shown in Section 3.3.3.

3.1 VIDE: the Void IDentification and Examination toolkit

To identify cosmic voids from the MultiDark PATCHY mock galaxy catalog we are using the publicly-available code **V**oid **ID**entification and **E**xamination toolkit (VIDE) (Sutter et al., 2015), an enhanced version on the watershed public code **ZO**nes **B**ordering **O**n **V**oidness (ZOBOV) (Neyrinck, 2008).

In subsection 3.1.1 we will see the types of input supported by VIDE. The core of the void finder is ZOBOV, which is an inversion of the dark-matter-halo finder VOBOZ (**VO**ronoi **BO**und **Z**ones) (Neyrinck et al., 2005). The major change in ZOBOV is that it looks for density minima instead of maxima. In Fig. (3.1) we can follow the void finding mechanism main steps, explained in subsections 3.1.2, 3.1.3 and 3.1.4. The results and outputs of VIDE are shown in subsection 3.1.5.

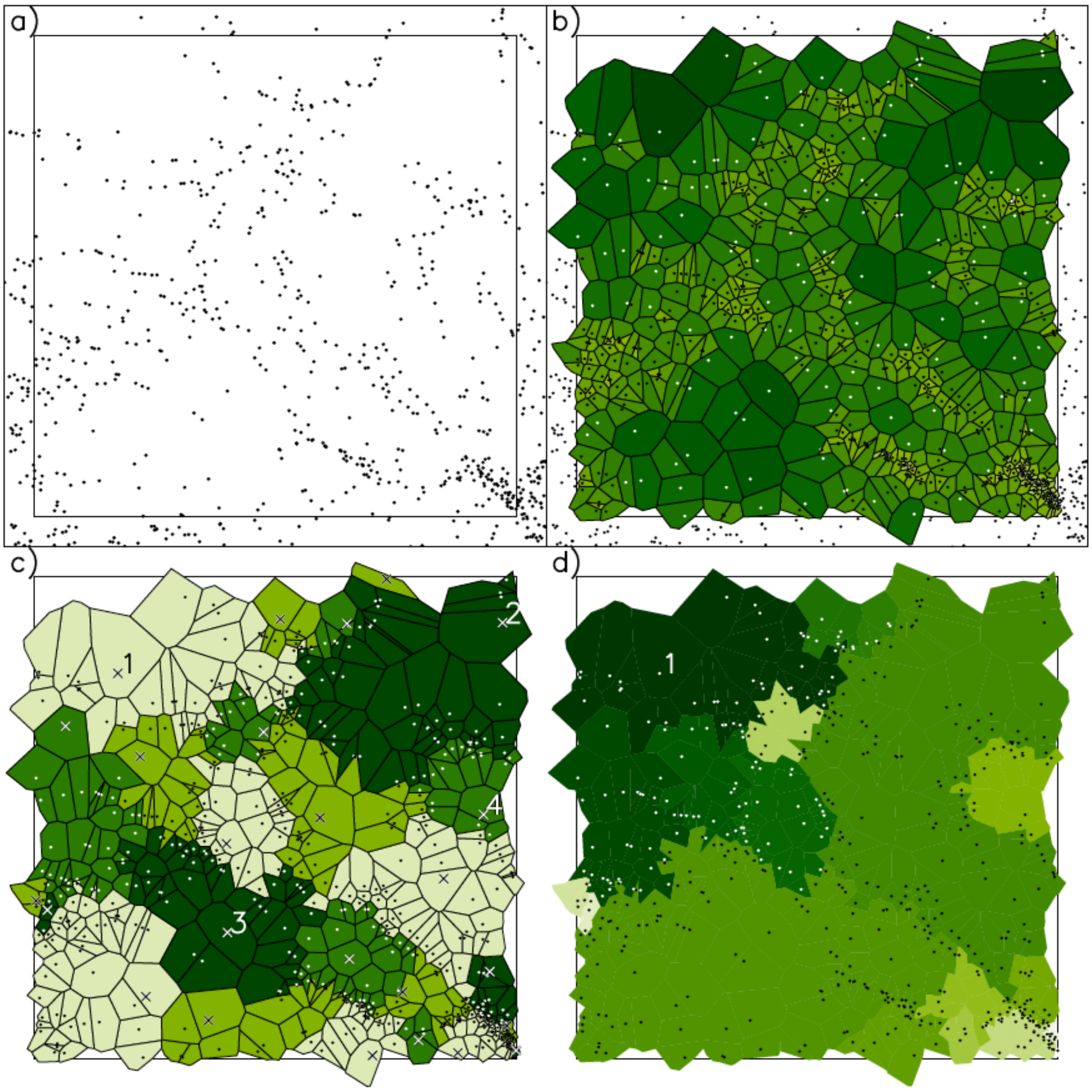


Figure 3.1: Void finding mechanism (from Neyrinck (2008)). **(a)** Galaxies from a $40 \times 40 \times 5(h^{-1}\mathrm{Mpc})^3$ slice of the AAVFCP region of the Millenium simulation. **(b)** 2D Voronoi tessellation, with each particle's Voronoi cell shaded according to its area. **(c)** Zones of galaxies. The cores of each zone, i.e. the density minima, are shown with crosses. Different colours demarcate different zones. **(d)** Watershed transform. Void 1, the deepest void of the sample, is represented with the darkest colour. Colours from dark to light indicate the stages at which each zone is added to the void.

3.1.1 Inputs

VIDE is extended to accept different inputs:

- **Simulations**:

  To identify voids in N-body dark matter simulations VIDE is able to read Gadget (Springel, 2005), FLASH (Dubey et al., 2008) and RAMSES (Teyssier, 2002) simulation outputs, files in the Self-Describing Format and generic ASCII files listing positions and velocities. It can also find voids in halo populations, constructing a mock galaxy population using a Halo Occupation Distribution (HOD) formalism (Berlind & Weinberg, 2002).

- **Observations**:

  To identify voids in observations the user has to provide an ASCII file listing galaxy right ascension, declination and redshift, as well as a pixelization of the survey mask using HEALPIX (Górski et al., 2005). HEALPIX can easily determine which pixels lie on the boundary between the survey and any masked region, an essential feature for VIDE to constrain voids to the survey volume. VIDE also provides a utility for constructing a rudimentary mask from the galaxy positions themselves.

3.1.2 Voronoi tessellation

The first step in the void finding mechanism is the density estimation of each particle in the survey, using the Voronoi tessellation field estimator, the dual of the Delaunay tessellation field estimator (Schaap, 2007). The Voronoi tessellation divides the space into cells around each particle, with the cell around particle i defined as the region of space closer to the particle i than to any other particle. If $V(i)$ is the volume of the Voronoi cell of particle i, we can define the density estimation as $1/V(i)$. Voronoi tessellation defines also a set of neighbours for each particle: the particles whose cells are adjacent to the Voronoi cell of particle i. The dual process, Delaunay triangulation, connects neighbouring particles. In Fig. (3.2) we can see Voronoi (on the left) and Delaunay (on the right) tessellations (Schaap, 2007).
Fig. (3.1a) shows a set of particles corresponding to galaxies in a slice of the Millenium simulation. Fig (3.1b) depicts the Voronoi tessellation of these particles.

3.1.3 Zoning

The next step of the void finder is the partition of the Voronoi cells into zones around each density minimum. A minimum is a particle with lower density than any of its Voronoi neighbours. The void finder sends each particle to its neighbour with higher density, until no more higher-density neighbours can be found. This way it creates local catchment basins, whose cores are the minimum density particles of each zone. Fig. (3.1c) shows the partition of the Voronoi cells into zones.

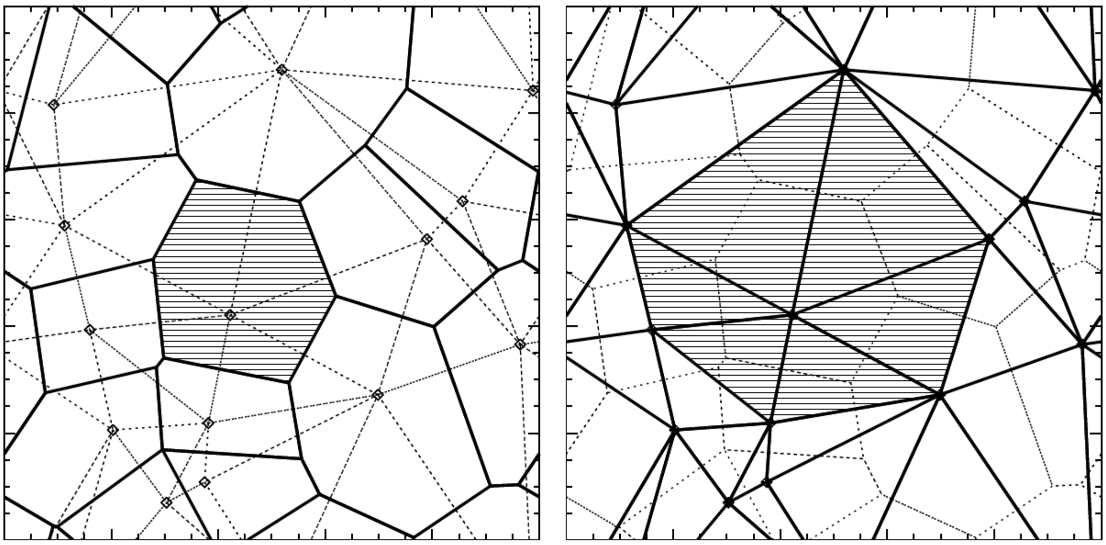


Figure 3.2: A set of 20 points with their Voronoi (left-hand frame: solid lines) and Delaunay (right-hand frame: solid lines) tessellation. *Left*: the shaded region indicates the Voronoi cell corresponding to the point located just below the centre. *Right*: the shaded region is the contiguous Delaunay cell of the same point as in the left-hand frame (from (Schaap, 2007))

3.1.4 Watershed transform

Finally, zones are joined into voids using a watershed transform. The word *watershed* refers to the analogy of a landscape being flooded by a rising level of water. In Fig. (3.3) we can follow the idea: on the left image we can see a surface in the shape of a landscape. As the level of the water rises gradually, more and more landscape will be flooded, until basins meet at the ridges corresponding to saddle-points of the surface. In the central image of Fig. (3.3) we can see this step. In the end, the landscape will be completely flooded, showing different cells separated by the ridge dams, as we can see in the last image of Fig. (3.3) (Platen et al., 2007).

In the case of ZOBOV and VIDE the quantity to be raised is the density. Starting from the absolute minimum value of the survey and then rising gradually its level, the adjacent zones are joined together when the level of the density field overcomes the value of the density of the wall between the basins. Fig. (3.1d) shows the growth of the density from the deepest basin, represented by the darkest colour on the upper left angle of the image. Successively lighter colours shade zones added when the density level reaches successively higher values. The void finder imposes a density-based threshold, where adjacent zones are only added to a void if the density of the wall between them is less than $0.2\bar{n}$, where $\bar{n}$ is the mean particle density of the survey. This way it is possible to define a nested hierarchy of voids, with multiple "parent" voids. A parent void is defined as a void which contains all the zones of a sub-void plus at least one more (Lavaux

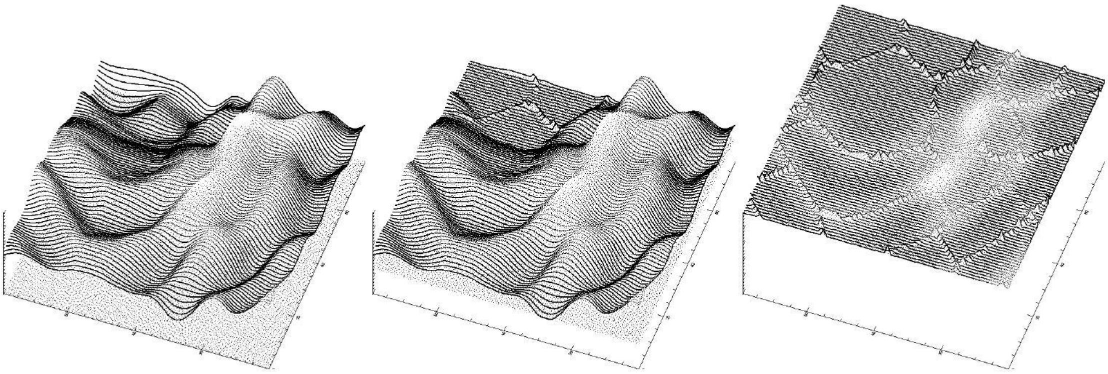


Figure 3.3: Principle of the watershed transform (form Platen et al. (2007)). *Left*: The surface to be studied. *Center*: Rising the level of water from the local minima different basins are flooded until they meet up near a ridge of the surface. *Right*: The entire surface is flooded, leaving a network of dams which delineates the web of the original surface.

& Wandelt, 2012). In our void catalogs we have set the threshold to zero, so we only consider lowest level voids without parents.

3.1.5 Outputs

Fig. (3.4) shows an example of a final void found by VIDE. The galaxies are shown as red dots, where dot sizes are proportional to the distance from the point of view. The different purple regions are the different Voronoi cells that form the void.

VIDE gives the user access to different void properties for each void, and in the following we will list the ones that will be useful later on in the analysis:

- Volume V.
- Radius: effective radius of the void, calculated from the volume of the void as

$$R_{\text{eff}} \equiv \left(\frac{3}{4\pi}V\right)^{\frac{1}{3}} \tag{3.1}$$

- Redshift z.
- Core density: density of the largest Voronoi cell of the void.
- Density contrast: the ratio d of the minimum density of the ridge separating the void from adjacent zones to the core density of the void.
- Number of member particles: number of tracers that define the void.

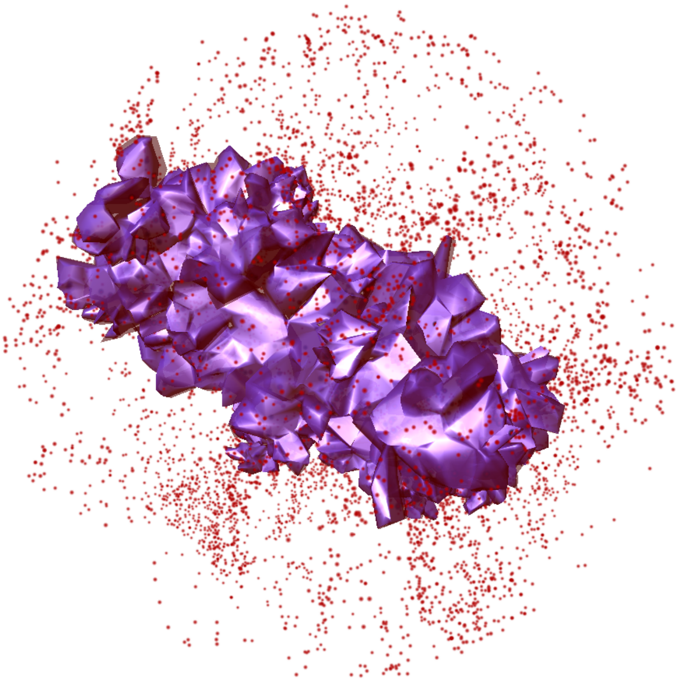


Figure 3.4: Example Voronoi-based void. The Voronoi cells that define the void are shown in purple with galaxies in red. The void has an effective radius of $20h^{-1}$ Mpc within a spherical region of radius $50h^{-1}$Mpc. Galaxy point sizes are proportional to their distance from the point of view. Galaxies interior to the void are shaded dark red (Sutter et al., 2012).

3.2 Least squares method

The method of Least Squares is a procedure to determine the best fit model to data. It was advanced early in nineteenth century by Gauss and Legendre for estimation in problems of astronomical measurement (Bickel & Doksum, 1977). The basic problem regards linear regression, where, given two sets of n observations $\{x_i\}$ and $\{y_i\}$, the aim is to find the best fit straight line $y = ax + b$. The method easily generalizes to finding the best fit of the form

$$y = c_1 f_1(x) + ... + c_k f_k(x) \,. \tag{3.2}$$

It is not necessary for the functions f_k to be linear in x, all that is necessary is for y to be a linear combination of these functions (Miller, 2006).

The idea is thus to find the curve as close as possible to all points in the plane xy. To quantify this concept we introduce the residual sum of squares RSS

$$RSS = \sum_{i=1}^{n} (y_i - (ax_i + b))^2 \tag{3.3}$$

in the case of the linear fit. The RSS measures the vertical distance from each data point to the line $ax + b$ and then sums the squares of these distances. The least squares estimates a^* and b^* are defined to be the values of the parameters a and b such that the RSS is minimum (Casella & Berger, 1990). In the general case of the fit $y = f(x)$, the least square method can be expressed as

$$\min \left(\sum_{i=1}^{n} (y_i - f(c_k, x_i))^2 \right) = \sum_{i=1}^{n} (y_i - f(c_k^*, x_i))^2 \tag{3.4}$$

where c_k^* is the set of parameters that minimizes the value of RSS, therefore the set of parameters that defines the best fit.

3.3 Markov Chain Monte Carlo

As we will see in section 4.2, we will need a different approach from the least squares method to obtain the fitting parameters. The choice has naturally fallen on a Markov Chain Monte Carlo (MCMC) algorithm, in particular the emcee code (Foreman-Mackey et al., 2013). MCMC is a numerical method for approximate inference, based on Bayesian inference. In the next sections we will see what Bayesian inference is (Section 3.3.1), the main steps to construct a MCMC (Section 3.3.2), and the emcee algorithm (Section 3.3.3).

3.3.1 Bayesian inference

Bayesian inference allows us to make inferences about the world around us, interpreting some data $\mathbf{D}$ in light of an underlying model M as a function of some parameters $\mathbf{\Theta}_M$ of the model. We are interested in finding the parameters which best describe the data $\mathbf{D}$, so we want to infer the probability $P(\mathbf{\Theta}_M|\mathbf{D}, M)$ that the parameters are actually $\mathbf{\Theta}_M$ given our data $\mathbf{D}$ and assuming the model M. Using factoring of probability and rearranging the terms we obtain Bayes' Theorem

$$P(\mathbf{\Theta}_M|\mathbf{D}, M) = \frac{P(\mathbf{D}|\mathbf{\Theta}_M, M)P(\mathbf{\Theta}_M|M)}{P(\mathbf{D}|M)} \tag{3.5}$$

where:

- $P(\mathbf{D}|\mathbf{\Theta}_M, M)$ is the *likelihood*, the probability of observing the data $\mathbf{D}$ assuming a specific choice of the parameter $\mathbf{\Theta}_M$ of the model M;
- $P(\mathbf{\Theta}_M|M)$ is the *prior*, the probability of having a particular set of parameters $\mathbf{\Theta}_M$ for our model M before conditioning upon the data;
- $P(\mathbf{D}|M)$ is the *evidence*

  $$P(\mathbf{D}|M) = \int P(\mathbf{D}|\mathbf{\Theta}_M, M)P(\mathbf{\Theta}_M|M)\mathrm{d}\mathbf{\Theta}_M \tag{3.6}$$

  or marginal likelihood of the model M, marginalized over all possible parameters values $\mathbf{\Theta}_M$. It quantifies how well the model explains the data $\mathbf{D}$ after averaging over all possible values of the underlying parameters;
- $P(\mathbf{\Theta}_M|\mathbf{D}, M)$ is the *posterior*, the probability of the parameters $\mathbf{\Theta}_M$ after combining the prior intuition $P(\mathbf{\Theta}_M|M)$ with observations $P(\mathbf{D}|\mathbf{\Theta}_M, M)$, and normalizing by the evidence $P(\mathbf{D}|M)$.

The posterior is a very useful tool to make inferences about the world around us. We can use it to make guesses about the model, compare different models and generating predictions. We can define the expectation value of a generic parameter-dependent function $f(\mathbf{\Theta}_M)$ with respect to the posterior $P(\mathbf{\Theta}_M)$:

$$\mathbb{E}_P[f(\mathbf{\Theta}_M)] \equiv \frac{\int f(\mathbf{\Theta}_M)P(\mathbf{\Theta}_M)\mathrm{d}\mathbf{\Theta}_M}{\int P(\mathbf{\Theta}_M)\mathrm{d}\mathbf{\Theta}_M} = \int f(\mathbf{\Theta}_M)P(\mathbf{\Theta}_M)\mathrm{d}\mathbf{\Theta}_M \tag{3.7}$$

since $\int P(\mathbf{\Theta}_M) = 1$ by definition. This represents the weighted average of $f(\mathbf{\Theta}_M)$, where the weight of each value of $\mathbf{\Theta}_M$ is given by the chance we believe that value is correct (Speagle, 2019).

3.3.2 Towards Monte Carlo Markov Chain

In general the posterior can not be computed analytically. To estimate quantities such as Eq. (3.7) requires the use of numerical tools. Focusing on the case in which the integral over $\mathbf{\Theta}$ is 1D, and dropping the subscript M for simplicity of notation, we can approximate the integral using the Riemann sum over a discrete grid of points

$$\mathbb{E}_P[f(\mathbf{\Theta})] \approx \sum_{i=1}^{n} f(\mathbf{\Theta}_i)P(\mathbf{\Theta}_i)\Delta\mathbf{\Theta}_i \tag{3.8}$$

where $\Delta\boldsymbol{\Theta}_i = \boldsymbol{\Theta}_{j+1} - \boldsymbol{\Theta}_j$ is the spacing between the set $j = 1, ..., n+1$ points of the grid, and $\boldsymbol{\Theta}_i = \frac{\boldsymbol{\Theta}_{j+1}+\boldsymbol{\Theta}_j}{2}$. Substituting to the normalized posterior $P(\boldsymbol{\Theta})$ the unnormalized one $\tilde{P}(\boldsymbol{\Theta})$, which is just $P(\boldsymbol{\Theta})$ multiplied by the evidence, the expectation value becomes

$$\mathbb{E}_P[f(\boldsymbol{\Theta})] \approx \frac{\sum_{i=1}^{n} f(\boldsymbol{\Theta}_i)\tilde{P}(\boldsymbol{\Theta}_i)\Delta\boldsymbol{\Theta}_i}{\sum_{i=1}^{n} \tilde{P}(\boldsymbol{\Theta}_i)\Delta\boldsymbol{\Theta}_i} \,. \tag{3.9}$$

Eq. (3.9) has the form of a weighted sample mean of $f(\boldsymbol{\Theta})$, where the value of the function in $\boldsymbol{\Theta}_i$, $f_i = f(\boldsymbol{\Theta}_i)$, is weighted by $w_i = \tilde{P}(\boldsymbol{\Theta}_i)\Delta\boldsymbol{\Theta}_i$. This resembles the fact that, according to the posterior function, some values of $f(\boldsymbol{\Theta})$ can be more important than others in the calculus of the expectation value. If we knew the shape of the posterior sufficiently well, for large n we should be able to adjust $\Delta\boldsymbol{\Theta}_i$ such that the weights w_i are uniform. When this happens

$$\Delta\boldsymbol{\Theta}_i \propto \frac{1}{\tilde{P}(\boldsymbol{\Theta}_i)} \tag{3.10}$$

for all i. As $n \to \infty$ we estimate the posterior using a larger and larger number of grid points whose spacing depend on $\boldsymbol{\Theta}$. We can then define the density of points based on the varying resolution $\Delta\boldsymbol{\Theta}(\boldsymbol{\Theta})$

$$Q(\boldsymbol{\Theta}) \propto \frac{1}{\Delta\boldsymbol{\Theta}(\boldsymbol{\Theta})} \,. \tag{3.11}$$

$Q(\boldsymbol{\Theta})$ is called proposal distribution. We can then rewrite Eq. (3.7) as

$$\mathbb{E}_P[f(\boldsymbol{\Theta})] = \frac{\int f(\boldsymbol{\Theta})\frac{\tilde{P}(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})}Q(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}}{\int \frac{\tilde{P}(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})}Q(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}} = \frac{\mathbb{E}_Q[f(\boldsymbol{\Theta})\frac{\tilde{P}(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})}]}{\mathbb{E}_Q[\frac{\tilde{P}(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})}]} \,. \tag{3.12}$$

Now we have two expectation values over $Q(\boldsymbol{\Theta})$ instead of $P(\boldsymbol{\Theta})$. This is useful because we can estimate the final expression using a series of randomly generated samples from $Q(\boldsymbol{\Theta})$. This method for estimating the expectation value is commonly referred to as *Monte Carlo* approach.

There are a lot of Monte Carlo sampling strategies, i.e. generating samples from the prior. Here we are interested in *Markov Chain Monte Carlo*, which generates samples in such a way that the importance weights associated with each sample are constant. MCMC hence seeks to generate samples that are proportional to the posterior, creating a chain of parameter values $\{\boldsymbol{\Theta}_1 \to ... \to \boldsymbol{\Theta}_n\}$ over n iterations such that the number of iterations $m(\boldsymbol{\Theta}_i)$ spent in a particular region centred on $\boldsymbol{\Theta}_i$ is proportional to the posterior density $P(\boldsymbol{\Theta}_i)$ contained within the region. In other words the sample density

$$\rho(\boldsymbol{\Theta}) \equiv \frac{m(\boldsymbol{\Theta})}{n} \tag{3.13}$$

tends towards the posterior distribution $\rho(\boldsymbol{\Theta}) \to P(\boldsymbol{\Theta})$ as $n \to \infty$. The expectation value $\mathbb{E}_P[f(\boldsymbol{\Theta})]$ therefore reduces to the sample mean (Speagle, 2019)

$$\mathbb{E}_P[f(\boldsymbol{\Theta})] = \frac{1}{n}\sum_{i=1}^{n} f_i \,. \tag{3.14}$$

3.3.3 Algorithm: emcee

The code we will use to run MCMCs is the emcee code (Foreman-Mackey et al., 2013), an open source which is a Python implementation of the affine-invariant ensemble sampler for Markov Chain Monte Carlo proposed by Goodman & Weare (2010). As we saw in section 3.3.2, MCMC is a procedure for generating random walks in the parameter space that, over time, draws a representative set of samples from the distribution. Each point in a Markov chain depends only on the position of the previous step. In the following we will briefly see the main steps of emcee algorithm.

Metropolis-Hastings algorithm

Metropolis-Hastings procedure is the simplest and most commonly used method to construct the MCMC. It is an iterative process schematised as follows:

1. Given a position $\mathbf{\Theta}(t)$ in the parameter space, sample a proposal position $\mathbf{\Psi}$ from the transition distribution $T(\mathbf{\Psi};\mathbf{\Theta}(t))$, which is a probability distribution for the transition from $\mathbf{\Theta}(t)$ to $\mathbf{\Psi}$.
2. Accept the proposal with probability

$$\min\left(1, \frac{P(\mathbf{\Psi}|\mathbf{D})}{P(\mathbf{\Theta}(t))|\mathbf{D})}\frac{T(\mathbf{\Theta}(t);\mathbf{\Psi})}{T(\mathbf{\Psi};\mathbf{\Theta}(t))}\right). \tag{3.15}$$

If this step is accepted the new position $\mathbf{\Theta}(t+1) = \mathbf{\Psi}$, otherwise the position $\mathbf{\Theta}(t)$ is repeated in the chain: $\mathbf{\Theta}(t+1) = \mathbf{\Theta}(t)$.

Stretch move

Metropolis-Hastings algorithm converges to a stationary set of sample from the distribution, but there are other algorithms with faster convergence. One of these is the algorithm proposed by Goodman & Weare (2010), an affine-invariant ensemble sampling called "stretch move". This method simultaneously evolves an ensemble of K walkers $S = \{\mathbf{\Theta}_k\}$, where the distribution of the walker k depends on the current positions of the other $K-1$ walkers. For the choice of the next position of the walker k, another walker j is randomly chosen between the remaining walkers, and the new position proposed is

$$\mathbf{\Theta}_k(t) \to \mathbf{\Psi} = \mathbf{\Theta}_j + Z[\mathbf{\Theta}_k(t) - \mathbf{\Theta}_j] \tag{3.16}$$

where Z is a random variable. The proposal is accepted with probability

$$\min\left(1, Z^{N-1}\frac{P(\mathbf{\Psi})}{P(\mathbf{\Theta}(t))}\right) \tag{3.17}$$

where N is the dimension of the parameter space. This procedure is then repeated in series for each walker.

Parallel stretch move

In order to parallelize the stretch move method we need to split the ensemble into two subsets $S^{(0)} = \{\mathbf{\Theta}_k \;\forall\, k = 1, ..., K/2\}$ and $S^{(1)} = \{\mathbf{\Theta}_k \;\forall\, k = K/2 + 1, ..., K\}$, and simultaneously update all the walkers of the set $S^{(0)}$ using the stretch move position, but considering only the positions of the walkers in the other set $S^{(1)}$. Then we can update the set $S^{(1)}$ using the new positions of $S^{(0)}$, and so on and so forth.

Chapter 4

Analysis

As we have seen in Section 1.3.2, the void size function quantifies the number of cosmic voids as function of their radius. Its mathematical form is not univocal, and both the Sheth-van de Weygaert model of Eq. (1.79) and the Vdn model of Eq. (1.84) depend on the same parameters: the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s.

The first aim of this thesis is to study the different models of the void size function through the investigation of the parameters of the models. This will be done searching for the best fit of the models to the data void abundance calculated from void catalogs constructed by VIDE using the mock galaxy catalog of the MultiDark PATCHY survey. The second goal of this thesis work is to study the cosmology dependence of the void size function. This will be achieved generating different void catalogs using VIDE with different values of the matter density parameter $\Omega_{\rm m}$ in the transform from redshifts to distances, and studying the fitting parameters of the void size function to the void abundances of the catalogs with different cosmologies. Model and cosmology dependence of the void size function will be analyzed in parallel through the investigation of the fitting parameters of the models.

To expose the work done we will follow the order in which the analysis has been carried out, in order to allow the reader to follow the choices made during the development of the thesis work, showing also the complications encountered during the analysis. In Section 4.1 we will explain how the void catalogs have been constructed and how the calculation of the data void abundance has been performed. Sections 4.2 and 4.3 will show the reader the reason why the multivariate analysis of Section 4.5 will be necessary, in order to clean our catalogs from spurious voids. In Section 4.6 Sheth-van de Weygaert and Vdn model of the void size function with general form and Sheth-van de Weygaert approximation of the first-crossing distribution function will be fitted to the void abundances of the different catalogs, and in Section 4.7 the cosmology dependence of the best fitting parameters will be examined. Section 4.8 will expose a study of the convergence of the series of Eq. (1.73), leading us to study model and cosmology dependence of the two void size function models with the hybrid function proposed by Jennings et al. (2013), in Section 4.9. In Section 4.10 we will question the choice of the most natural fitting

Ω_{m}
0.0
0.1
0.2
0.307115
0.4
0.5

Table 4.1: Values of the matter density parameter used in the construction of the void catalogs.

parameters considered. Alcock-Paczyński effect will be studied in Section 4.11, and after having applied the Alcock-Paczyński correction proposed by Correa et al. (2020) we will search for a residual cosmology dependence of the fitting parameters.

4.1 Generation of catalogs

The first step in our analysis is the generation of six different catalogs, each of them with a different value of the matter density parameter of the universe Ω_{m} in the transform between redshifts and distances. As we have seen in Eq. (1.20), the value of Ω_{m} affects the value of the Hubble parameter H, which influences the transform between redshift and distances, as we have seen in Eq. (1.8) and Eq. (1.7). Changing the value of Ω_{m} therefore changes the way in which VIDE reads the MultiDark PATCHY mock galaxy catalog, driving it to identify differently distorted structures from the same survey.

In Table (4.1) are listed the values of Ω_{m} used in the generation of the different catalogs. Note the presence of the value $\Omega_{\mathrm{m}} = 0.307115$, which is the value of the matter density parameter that has been used in the generation of the mock galaxy catalog (see Table (2.1)). From now on we will refer to this value of Ω_{m} as "true cosmology". In Fig. (4.1) are shown the abundances of the different catalogs as a function of the void radius $\frac{\mathrm{d}n}{\mathrm{d}\ln R}(\delta_V, \delta_c, r_s, R)$. Abundances are calculated as follows:

1. Load the void catalog and extract the void radius using VIDE;
2. Make an histogram of the void radius using a log-spacing binning;
3. Calculate the mean difference between bins $\mathrm{d}R$ and the mean value of each radius bin R;
4. Divide the histogram density by the survey volume;
5. In order to obtain the void abundance as a function of the logarithm of the radius divide the number density by $\mathrm{d}R/R$.

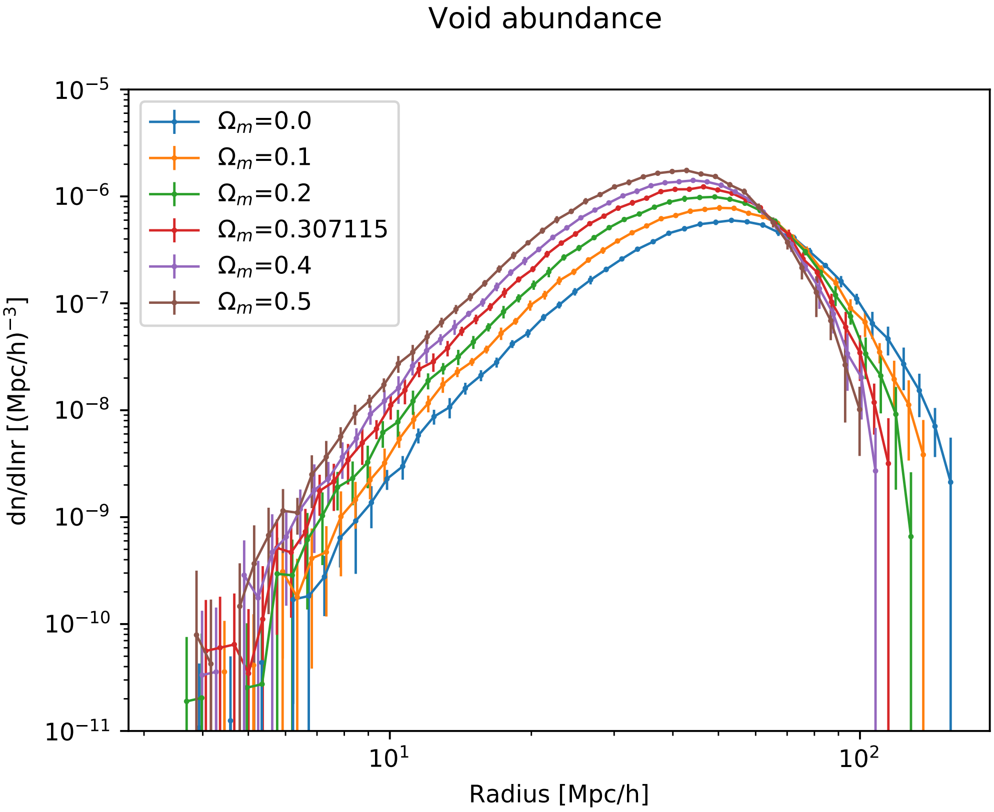


Figure 4.1: Void abundances as a function of the void radius of the six different catalogs generated using VIDE with different values of the matter density parameter Ω_{m} in the transform from redshifts to distances.

4.2 Least squares fits: ridge problem

Now that we have constructed the catalogs, we want to fit the abundances with the different models of the void size function. We remind here the four different models that we will consider:

- Sheth-van de Weygaert model (Eq. 1.79) using the general form of the first-crossing distribution function (Eq. 1.73). From now on we will call this model SvdW-general.

- Sheth-van de Weygaert model (Eq. 1.79) using Sheth-van de Weygaert approximation of the first-crossing distribution function (Eq. 1.74). From now on we will call this model SvdW-SvdW.

- Vdn model (Eq. 1.84) using the general form of the first-crossing distribution function (Eq. 1.73). From now on we will call this model Vdn-general.

- Vdn model (Eq. 1.84) using Sheth-van de Weygaert approximation of the first-crossing distribution function (Eq. 1.74). From now on we will call this model Vdn-SvdW.

In the calculation of the models we neglected the redshift evolution of the void size function, and instead we expressed it for the redshift average of all voids. The aim is now to fit these models to the different catalogs, and study the behaviour of the fitting parameters. In order to do so the first approach has been the Least Squares method (Section 3.2): for each void size function model $\frac{\mathrm{d}n^{\mathrm{model}}}{\mathrm{d}\ln R}(\delta_V, \delta_c, r_s, R)$ we want to minimize the residual sum of squares

$$RSS = \sum_{i=1}^{N} \left(\frac{\mathrm{d}n_i^{\mathrm{data}}}{\mathrm{d}\ln R} - \frac{\mathrm{d}n_i^{\mathrm{model}}}{\mathrm{d}\ln R}(\delta_V, \delta_c, r_s, R) \right)^2 \tag{4.1}$$

where $\frac{\mathrm{d}n^{\mathrm{data}}}{\mathrm{d}\ln R}$ is the void abundance calculated as in Section 4.1, i.e. our data, and N is the number of bins in which we are dividing our abundances. The optimal values of the parameters δ_V, δ_c and r_s are the ones that minimize the RSS of Eq. (4.1).

During this analysis, however, a problem has occurred: the optimal parameters found by the minimization were not univocal. For each model studied, changing the initial guess of parameters in the minimization gave different optimal results, with a different value of the RSS. In Fig. (4.2) we can see the behaviour of the RSS of the different minimizations as a function of the final parameters, in the case of the SvdW-SvdW model. The plot shows 194 points, each one corresponding to a different minimization. On the three axes we find the parameters δ_V, δ_c and r_s, and the value of the residual sum of squares RSS is indicated by the colour bar.

This problem is known as *ridge problem* (Russell & Norvig, 1995), and it is a well known issue in hill-climbing algorithms. In hill-climbing the aim is to find the global maximum of a function, and the ridge problem implies that, when going up the hill searching for the maximum, the algorithm gets stuck on a ridge and it is not able to find

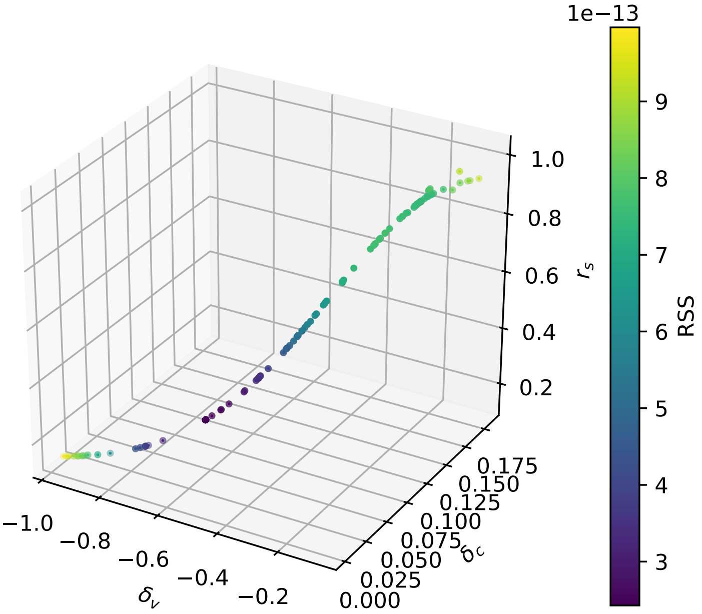


Figure 4.2: Residual sum of squares RSS as a function of the final parameters obtained in the least squares minimization fits for 194 different initial parameter guesses. The RSS color is darker for low values and lighter for higher ones, as shown in the color bar.

the direction to go further up. In our case we are searching for the global minimum, and the ridge problem consists in the fact that the minimization does not find a single value, but a curve composed by local minima. To produce the plot of Fig. (4.2) we have defined a grid of initial values and search for the absolute minimum, but as we can see we have found a ridge of local minima. With a rigid grid we are not able to gain the certainty of having found the global minimum, even if we can see that the values of RSS on the ridge are not constant but seem to indicate a trend towards the global minimum. However, the real absolute minimum could always be the result of a minimization that we have not performed yet. The least squares method, even iterated on a grid, is not the best approach to proceed with the analysis. A new method is needed, and from now on we will use a Markov Chain Monte Carlo in order to perform the fits (see Section 3.3).

4.3 MCMC fits: first attempt

In order to overcome the ridge problem we adopted a different method to fit the different models of the void size function to the data: Markov Chain Monte Carlo (MCMC). In Section 3.3 we explained what MCMCs are, and we briefly followed the steps of the emcee code (Foreman-Mackey et al., 2013) we have used to perform them.

We have then fitted the four different models of the void size function listed in Section 4.2 to the void abundances as a function of the radius of the six different catalogs. In Fig. (4.3) we can see the best fits of the four models to the data corresponding to the true cosmology (i.e. the catalog constructed with $\Omega_{\rm m} = 0.307115$). It is evident that none of the models fits to the data sufficiently well, especially on small radius scales, where the void-in-cloud effect is dominant. The ridge problem of the previous section has been overcome and the result of the MCMC fit is univocal, nevertheless the void size function models and the data abundance do not agree at all. The problem must lay either on the models or on the data.

4.4 Large radii limit

In the literature we often see only the right tail of the void abundance, that corresponds to the abundance of large voids (see e.g. Chan et al. (2014)). Small voids are more challenging to take into account, because at small scales the presence of the shot noise is more relevant, and the void-in-cloud effect becomes dominant. Therefore in this section we will consider only the large radii limit, i.e. only the voids with radius larger than the one corresponding to the peak of the data abundance, in order to investigate the reason of the poor agreement between the void size function models and the data abundances that we have seen in last section (see Fig. (4.3)).

Since we are now considering only large voids, we can neglect the void-in-cloud contribution. Therefore as first-crossing distribution function we will consider the Press-Schechter function (Press & Schecher, 1974):

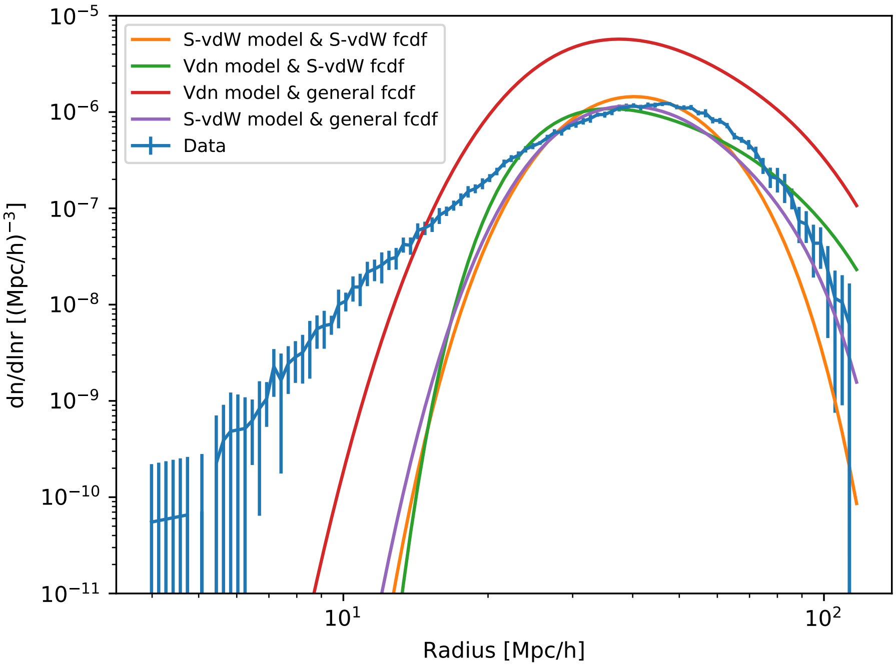


Figure 4.3: Best fits of the different void size function models to the void abundance of the catalog constructed using the true cosmology $\Omega_{\mathrm{m}} = 0.307115$. *Orange*: SvdW-SvdW model. *Green*: Vdn-SvdW model. *Red*: Vdn-general model. *Purple*: SvdW-general model.

$\Omega_{\rm m}$	δ_V	r_s
0.0	$-0.094^{+0.005}_{-0.005}$	$0.995^{+0.013}_{-0.013}$
0.1	$-0.129^{+0.007}_{-0.008}$	$0.916^{+0.019}_{-0.019}$
0.2	$-0.155^{+0.009}_{-0.009}$	$0.873^{+0.019}_{-0.021}$
0.307115	$-0.175^{+0.011}_{-0.012}$	$0.851^{+0.021}_{-0.024}$
0.4	$-0.189^{+0.014}_{-0.014}$	$0.839^{+0.024}_{-0.026}$
0.5	$-0.225^{+0.018}_{-0.017}$	$0.781^{+0.024}_{-0.028}$

Table 4.2: *Sheth-van de Weygaert* model with *Press-Schechter* first-crossing distribution function. Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter $\Omega_{\rm m}$. The fits have been performed only to voids with large radius.

$$\nu f(\nu) = \sqrt{\frac{2}{\pi}}\nu e^{-\frac{\nu^2}{2}} \tag{4.2}$$

Comparing this equation to Sheth-van de Weygaert first-crossing distribution function of Eq. (1.74), we can notice that the latter is formed by a first contribution identical to the Press-Schechter function, and a second contribution due to the void-in-cloud effect. Considering only Press-Schechter form of the first-crossing distribution function, we fit Sheth-van de Weygaert model (Eq. 1.79) and the Vdn model (Eq. 1.84) to the large voids abundance of all catalogs using MCMC method, with the linear density threshold for void formation δ_V and the rescaling factor r_s as free parameters. Fig. (4.4) shows the best fits of both models to the void catalogs abundances.

A great improvement from the fit to all voids of Fig. (4.3) is visible: the agreement between the models and the data is remarkable. The optimal parameters obtained from the fits of the Sheth-van de Weygaert model (Eq. 1.79) are shown in Table (4.2), while the parameters obtained in the fits of the Vdn model (Eq. (1.84)) are presented in Table (4.3). The Sheth-van de Weygaert model exhibits a linear threshold for void formation δ_V smaller, in absolute value, with respect to the one predicted by the spherical collapse theory (see Section 1.2.2). On the other hand the rescaling factor r_s is higher than expected, representing voids which have expanded less than the ones predicted by spherical collapse. The fitting parameters of the Vdn model are even farther from theory than the Sheth-van de Weygaert ones. The absolute values of the linear thresholds for void formation are two order of magnitude smaller than the one predicted by spherical collapse theory. The rescaling factors are instead larger than expected: with only one exception in the case of the catalog constructed with $\Omega_{\rm m} = 0.0$, all values are greater than 1, reaching $r_s \approx 6.5$ in the case of the catalog constructed with $\Omega_{\rm m} = 0.5$. A rescaling factor that exceeds unity represents voids which could shrink during their evolution, instead of expanding.

We are looking for a clue to recognise the true cosmology in the fitting parameters. Let us then investigate the dependence of the parameters upon the matter density pa-

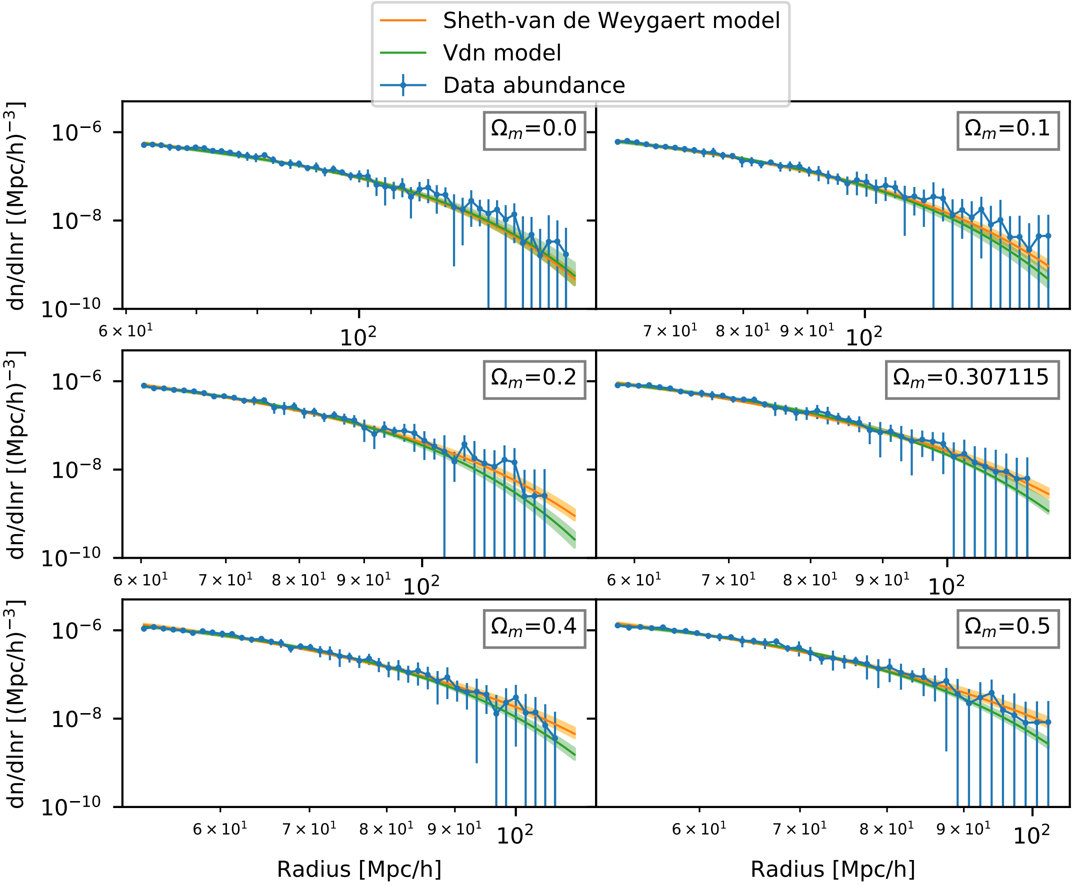


Figure 4.4: Best fits of *Sheth-van de Weygaert model* (in orange) and *Vdn model* (in green) using *Press-Schechter* first crossing distribution function, with as free parameters the linear threshold for void formation δ_V and the rescaling factor r_s. The data void abundances are represented in blue. Shaded regions show the error ranges of the parameters of each model. In the upper right of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

$\Omega_{\rm m}$	δ_V	r_s
0.0	$-0.096^{+0.013}_{-0.022}$	$0.971^{+0.130}_{-0.166}$
0.1	$-0.052^{+0.009}_{-0.010}$	$1.696^{+0.261}_{-0.208}$
0.2	$-0.035^{+0.008}_{-0.009}$	$2.374^{+0.424}_{-0.336}$
0.307115	$-0.027^{+0.008}_{-0.009}$	$2.950^{+0.731}_{-0.544}$
0.4	$-0.024^{+0.008}_{-0.010}$	$3.345^{+0.922}_{-0.665}$
0.5	$-0.009^{+0.003}_{-0.006}$	$6.510^{+2.040}_{-1.756}$

Table 4.3: *Vdn model with Press-Schechter first-crossing distribution function.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter $\Omega_{\rm m}$. The fits have been performed only to voids with large radius.

rameter $\Omega_{\rm m}$. In Fig. (4.5) and Fig. (4.6) are shown the cosmology dependences of the parameters obtained from the Sheth-van de Weygaert and Vdn model of the void size function, respectively. Sheth-van de Weygaert model parameters show an almost linear decreasing behaviour as $\Omega_{\rm m}$ increases. Our aim is to find a behaviour that can suggest a preference for the true cosmology, but none indicator of such preference is visible from these parameters. Fig. (4.6) displays instead an increasing behaviour of the fitting parameters as functions of $\Omega_{\rm m}$. Both trends look less linear than the Sheth-van de Weygaert case. A slight change of slope in the trends is visible corresponding to the $\Omega_{\rm m} = 0.4$ catalog, but it could be associated with an anomaly in the parameters of the $\Omega_{\rm m} = 0.5$ catalog. Remember that the values of the parameters in the Vdn fits result unphysical, and therefore their behaviour can not be considered very relevant.

We have fitted the two models to large radii voids in order to better understand the physics beyond the disagreement of Fig. (4.3). We have found a better agreement of the models to data, but the fitting parameters of the Vdn model result unlikely, representing voids which, instead of expand during their evolution, could shrink to smaller sizes.

4.5 Multivariate analysis

As we have briefly mentioned in last section, the abundance of voids can be altered by the presence of shot noise. This is due to the fact that voids identified by VIDE are simply density minima with depressions around them. This can lead to the identification of spurious voids due to the presence of discreteness noise, as first pointed out by Neyrinck (2008). This effect is especially relevant at small scales, but also the abundance of large voids can be influenced by shot noise. Neyrinck proposed a model to quantify the significance of voids: assuming that the discreteness noise is similar to that in a Poisson distribution, he converted the density contrast d to a probability. Defining the cumulative probability $P(d)$ as the fraction of voids in a Poisson particle distribution with density contrast greater than d, Neyrinck found

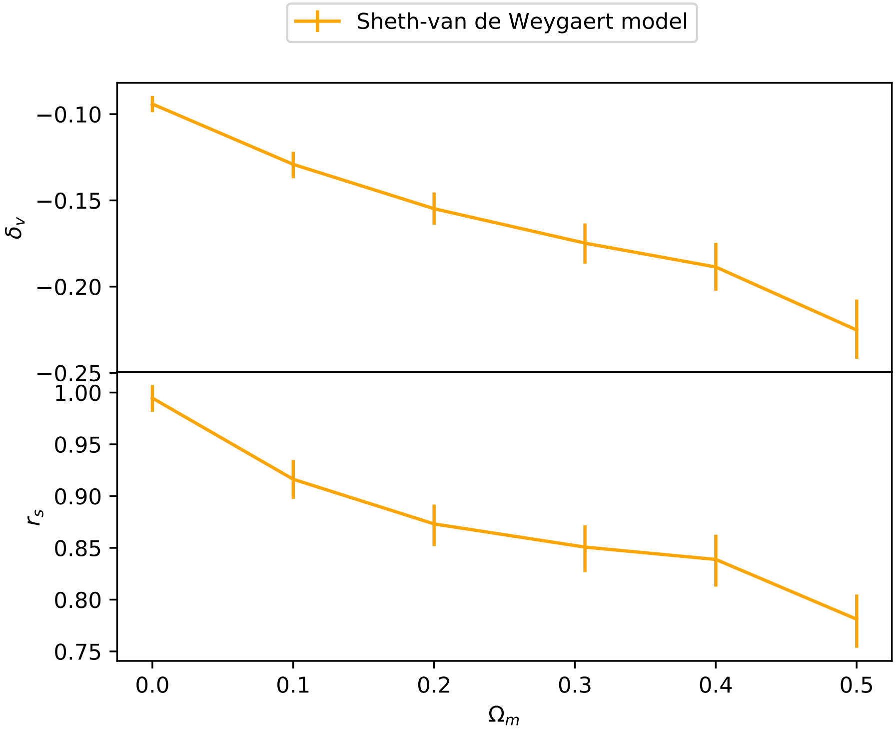


Figure 4.5: Best fits parameters as a function of $\Omega_{\rm m}$. The parameters were obtained fitting *Sheth-van de Weygaert model* using *Press-Schechter* first-crossing distribution function to the abundances as a function of void radius of the catalogs constructed with different $\Omega_{\rm m}$.

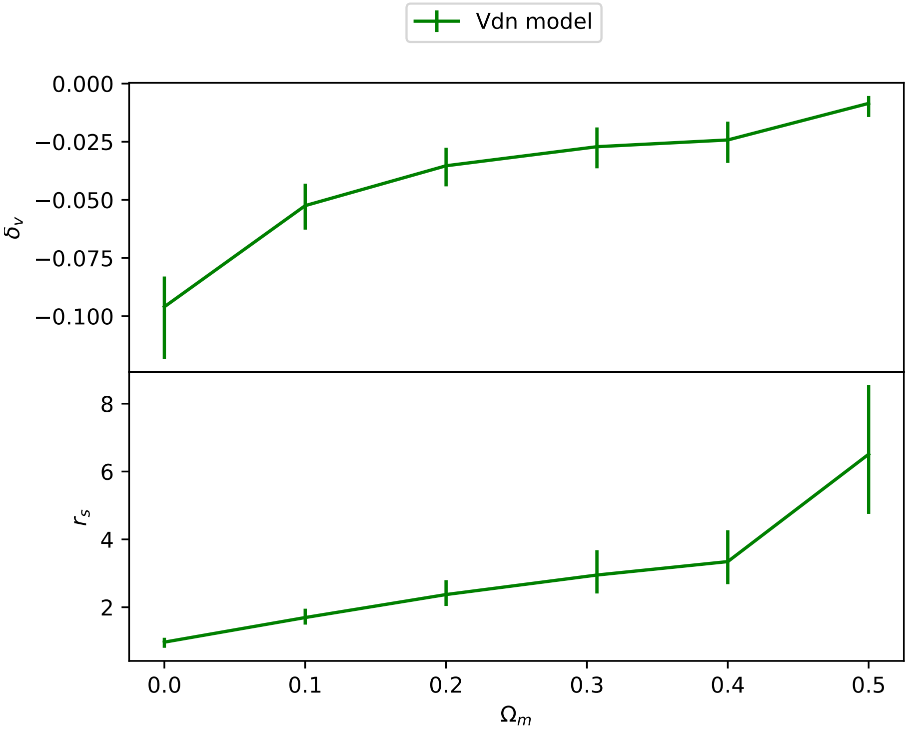


Figure 4.6: Best fits parameters as a function of Ω_{m}. The parameters were obtained fitting *Vdn model* using *Press-Schechter* first-crossing distribution function to the abundances as a function of void radius of the catalogs constructed with different Ω_{m}.

$$P(d) = \exp(-5.12(d-1) - 0.8(d-1)^{2.8}) \tag{4.3}$$

and used this probability to evaluate the significance of voids identified by the void finder. After Neyrinck's first proposal, other cuts have been used to trim the void catalog in an attempt to prune out spurious voids. Mao et al. (2017) used the density contrast as a measure of void significance. Nadathur & Hotchkiss (2015) suggested a criterion in which the void center, or circumcenter, is chosen as the point of intersection of the four lowest density mutually adjacent Voronoi cells in the void, and the void is then classified as spurious or genuine according to whether it is overdense or underdense compared with the mean density, respectively, at the circumcenter. The downside of this approach is that voids identified in a Poisson distribution are also underdense, and some genuine voids can have an overdense core. Hamaus et al. (2016) chose to exclude voids with radii below twice the mean particle separation. However, this approach strongly reduces the size of the sample without removing all random voids.

All these cuts, taken separately, seem to not be able to isolate spurious voids from real ones. We therefore choose to follow Cousinou et al. (2019) and perform a multivariate analysis. In order to distinguish between random voids and genuine ones , we constructed a void catalog using VIDE from the random catalog of the MultiDark PATCHY mocks, subsampled it in order to obtain the same statistics of the catalogs generated in Section 4.1, and compared its abundance as a function of radius to the one of the void catalog constructed using the true cosmology $\Omega_{\mathrm{m}} = 0.307115$. In Fig. (4.7) the abundance comparison is presented. In Fig. (4.7) we can already notice a difference between the two catalogs, as the random catalog shows a greater abundance at small voids with respect to the random catalog. The catalogs, however, are too close to suggest a simple cut in the radius of voids: we need to study more properties.

Looking at the comparisons of the different void properties that VIDE outputs (see Section 3.1.5), we can see an interesting difference regarding the core density, i.e. the density of the largest Voronoi cell in the void. In Fig. (4.8) this comparison is visible. We can see that the random catalog shows a peak at high values of the core density, while the peak of the galaxy catalog appears at lower values. Our aim is to cut out as many random voids as possible while keeping as many real voids as possible. Since the abundance of the galaxy catalog at core densities corresponding to the random peak is considerably lower than the random abundance, we choose to cut all voids with core density higher than the intersection between the two catalogs, excluding the random peak. The cut has been applied to core densities greater than 2.5. The abundance comparison of the two catalogs after the core density cut is shown in Fig. (4.9). Looking at the abundance comparison after the cut we can notice how the random catalog has been affected more than the galaxy catalog: the peak of the random catalog at about $30\,\mathrm{Mpc}/h$ in Fig. (4.7) is no more present, and the random abundance is everywhere lower in magnitude than the abundance of the galaxy catalog. We can also notice how the cut in the core density has excluded the voids with the smallest radii: in both catalogs voids with radius smaller than $10\mathrm{Mpc}/h$ have been cut out.

The cut in the core density has been able to distinguish between random and real

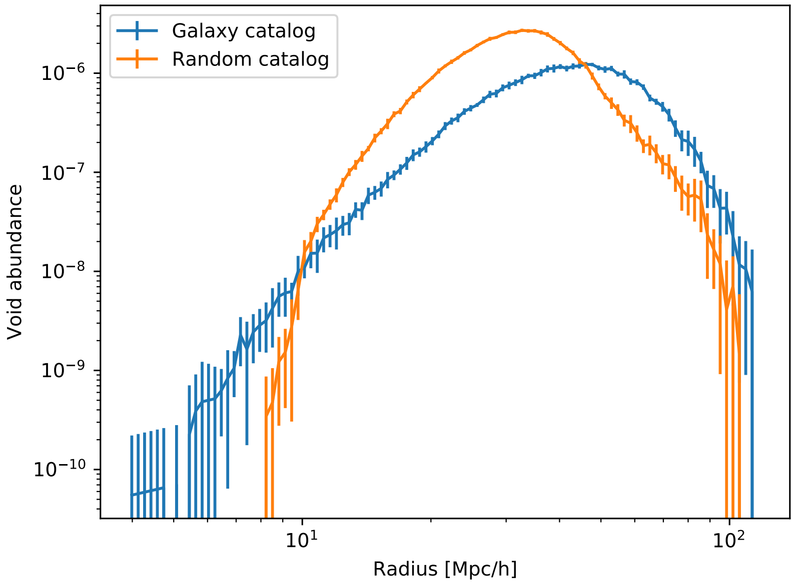


Figure 4.7: Void abundance comparison between the catalog constructed using $\Omega_{\mathrm{m}} = 0.307115$ (in blue) and the catalog constructed from the random catalog of the MultiDark PATCHY mocks (in orange).

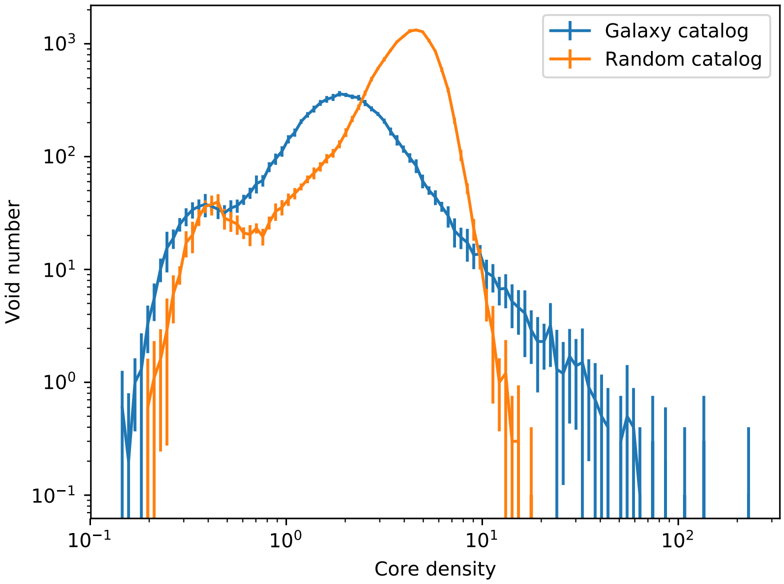


Figure 4.8: Core density comparison between the catalog constructed using $\Omega_{\mathrm{m}} = 0.307115$ (in blue) and the catalog constructed from the random catalog of the MultiDark PATCHY mocks (in orange).

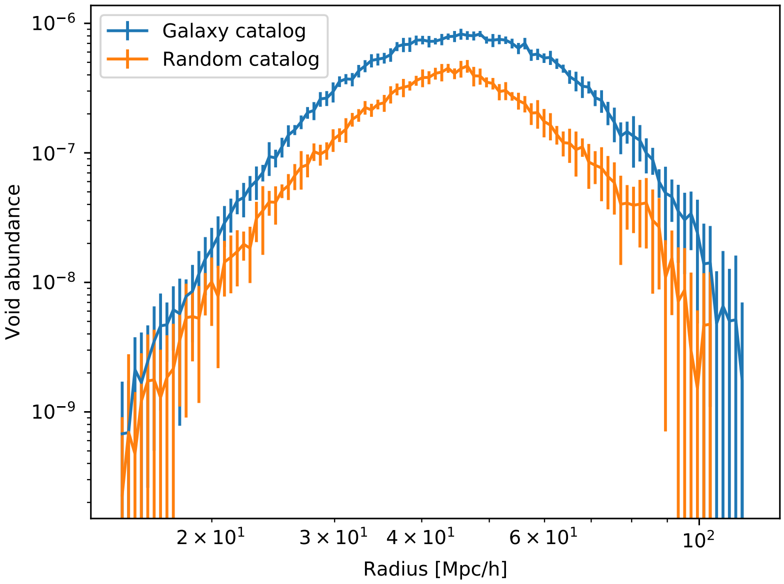


Figure 4.9: Void abundance comparison between the catalog constructed using $\Omega_\mathrm{m} = 0.307115$ (in blue) and the catalog constructed from the random catalog of the MultiDark PATCHY mocks (in orange), after having applied the core density cut, keeping only voids with core density smaller than 2.5.

voids, affecting random voids more, but the abundances of the two catalogs are still too close to each other to allow us to say that we have pruned out the majority of spurious voids. Therefore we focus now on another property of voids: compensation C, defined as

$$C = \frac{N_{\rm t}}{V} \tag{4.4}$$

where $N_{\rm t}$ is the number of tracers inside a void, and V is the void volume. In Fig. (4.10) the comparison between the compensation of the two catalogs after having applied the core density cut is shown.

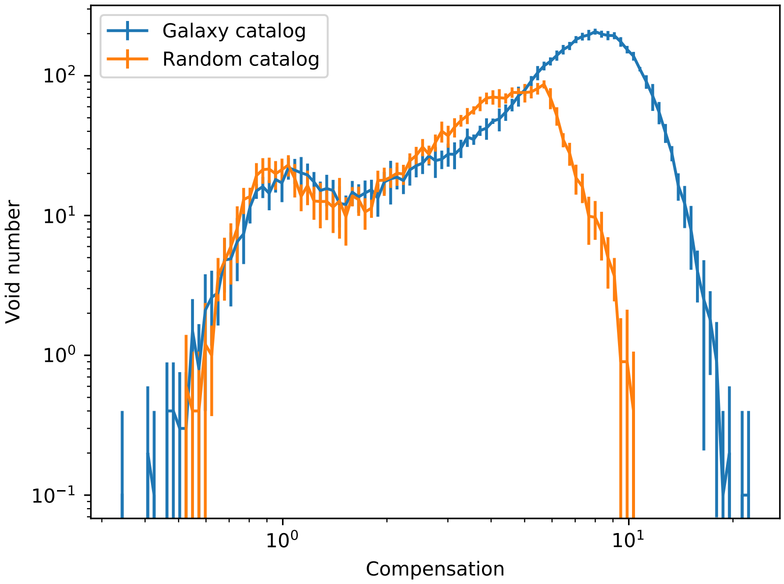


Figure 4.10: Compensation comparison between the catalog constructed using $\Omega_{\rm m} = 0.307115$ (in blue) and the catalog constructed from the random catalog of the MultiDark PATCHY mocks (in orange), after having applied the core density cut, keeping only voids with core density smaller than 2.5

In Fig. (4.10) the characteristics of random voids seem to be more evident. At low values of the compensation the abundances of the two catalogs follow almost the same curve, until at a value of the compensation of 5 the random catalog starts to decrease rapidly in abundance, while the galaxy catalog increases its abundance forming a peak. In order to exclude spurious voids in both catalogs, while keeping as many genuine voids as possible in the galaxy catalog, we apply a cut at the value of compensation of 5, keeping only voids with a compensation higher than that value. Once again we look at the void abundance as a function of void radius after both the core density and the compensation cuts, whose comparison is shown in Fig. (4.11).

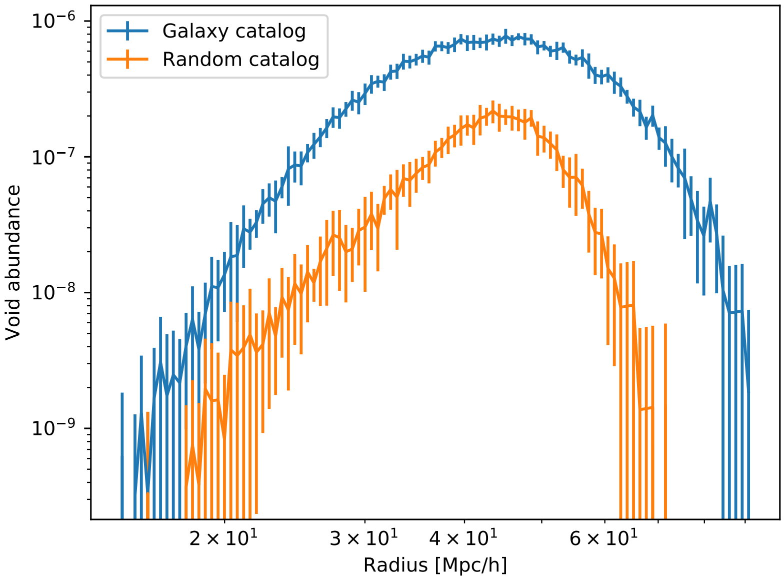


Figure 4.11: Void abundance comparison between the catalog constructed using $\Omega_\mathrm{m} = 0.307115$ (in blue) and the catalog constructed from the random catalog of the MultiDark PATCHY mocks (in orange), after having applied the core density and the compensation cuts, keeping only voids with core density smaller than 2.5 and compensation higher than 5.

Ω_m	δ_V	δ_c	r_s
0.0	$-0.182^{+0.007}_{-0.007}$	$0.035^{+0.002}_{-0.002}$	$0.888^{+0.035}_{-0.033}$
0.1	$-0.168^{+0.008}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$0.977^{+0.045}_{-0.040}$
0.2	$-0.154^{+0.007}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$1.083^{+0.051}_{-0.044}$
0.307115	$-0.136^{+0.009}_{-0.008}$	$0.040^{+0.003}_{-0.003}$	$1.244^{+0.078}_{-0.064}$
0.4	$-0.139^{+0.008}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$1.266^{+0.068}_{-0.058}$
0.5	$-0.127^{+0.009}_{-0.008}$	$0.037^{+0.003}_{-0.003}$	$1.398^{+0.100}_{-0.080}$

Table 4.4: *Vdn-general model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter Ω_m.

The abundance comparison of Fig. (4.11) shows a considerable difference between the two catalogs. The second cut had a little affect on the galaxy catalog abundance, while it had a massive effect on the random one. The abundance of the galaxy catalog is now almost one order of magnitude higher than the random catalog abundance. Having excluded the majority of voids associated to the random catalog we can presume to have cut out also the majority of spurious voids identified by VIDE due to shot noise in the galaxy catalog. We can continue with the analysis considering the six catalogs cleaned by the two cuts just described.

4.6 Model dependence of the void size function

In the last section we have analysed a way to get rid of spurious voids due to shot noise, which induces VIDE to identify voids that are not genuine. Therefore we apply to all six catalogs the two cuts already described, keeping only voids with:

- Core density < 2.5,
- Compensation > 5.

This way we can consider our catalogs "cleaned" of spurious voids, and search for the best fitting parameters of the four models of the void size function. The fits are performed using MCMC method. We will consider separately the fits of each model.

4.6.1 Vdn-general model

Let us first look at the fits to the catalogs abundances of the Vdn model (Eq. (1.84)) using the general form of the first-crossing distribution function (Eq. (1.73)). In Fig. (4.12) we can see the best fits to all catalogs. The agreement between the function and the data is remarkable, and as an estimate of the goodness of the fits we can look at the maximum posterior corresponding to the best fits parameters, whose mean value over all six catalogs is about -40.

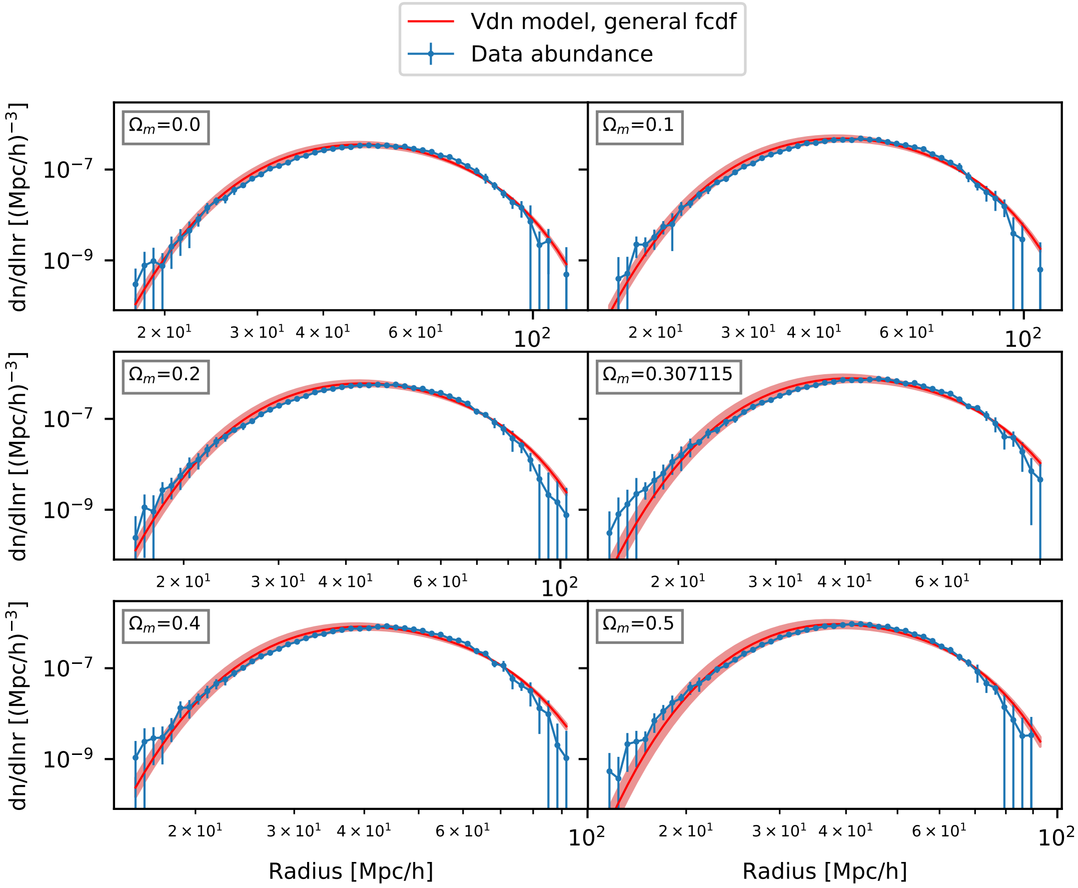


Figure 4.12: Best fits of *Vdn-general model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in red. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

$\Omega_{\rm m}$	δ_V	δ_c	r_s
0.0	$-0.287^{+0.010}_{-0.011}$	$0.050^{+0.002}_{-0.002}$	$0.583^{+0.023}_{-0.023}$
0.1	$-0.250^{+0.010}_{-0.011}$	$0.054^{+0.003}_{-0.003}$	$0.677^{+0.028}_{-0.028}$
0.2	$-0.217^{+0.008}_{-0.008}$	$0.055^{+0.003}_{-0.003}$	$0.783^{+0.029}_{-0.028}$
0.307115	$-0.197^{+0.009}_{-0.009}$	$0.057^{+0.003}_{-0.003}$	$0.879^{+0.038}_{-0.036}$
0.4	$-0.202^{+0.008}_{-0.009}$	$0.055^{+0.003}_{-0.003}$	$0.900^{+0.037}_{-0.036}$
0.5	$-0.188^{+0.008}_{-0.008}$	$0.053^{+0.003}_{-0.003}$	$0.980^{+0.040}_{-0.038}$

Table 4.5: *Vdn-SvdW model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter $\Omega_{\rm m}$.

In Tab. (4.4) the best fits parameters are presented. The absolute values of the linear thresholds for void formation δ_V are smaller than the one predicted by spherical evolution theory, as are the values of the linear thresholds for collapse δ_c, the latter being of two order of magnitude smaller than expected. The rescaling factors increase with $\Omega_{\rm m}$, and for the catalogs constructed with values of $\Omega_{\rm m}$ above 0.1, r_s is greater than 1.

4.6.2 Vdn-SvdW model

Fig. (4.13) shows the fits to the catalogs abundances of the Vdn model of the void size function (Eq. (1.84)) with Sheth-van de Weygaert approximation of the first-crossing distribution function (Eq. (1.74)).

The fits corresponding to this model seem quite good, but they do not perfectly agree with the data, especially in the range of small radii. The mean value of the maximum posterior of the fits is in fact about -80, smaller than the one of the Vdn-general model. The best fitting parameters are presented in Tab. (4.5). We can notice again how the linear thresholds are smaller, in absolute value, than expected. The linear thresholds for collapse δ_c are two order of magnitude far from the predicted value of 1.674. However, the rescaling factors r_s are smaller than 1, corresponding to expanding voids as we expect. In the case of the catalog constructed using $\Omega_{\rm m} = 0.0$ the value of r_s is the one predicted from spherical collapse $r_s = 0.58$.

4.6.3 Vdn models comparison

The two Vdn models we saw in the last two subsections deserve a closer look. We can see "by eye" from the fits of Fig. (4.12) and Fig. (4.13) that they are quite similar, but to compare the two models we better look at Fig. (4.14), where we can see the void abundance as a function of void radius of the catalog constructed with the true cosmology $\Omega_{\rm m} = 0.307115$, along with the two Vdn models with general form (in red) and Sheth-van de Weygaert approximation (in green) of the first-crossing distribution function. Looking at this figure we can notice how the two models fit similarly the data at large radii, while diverging more and more as the void radius becomes smaller and

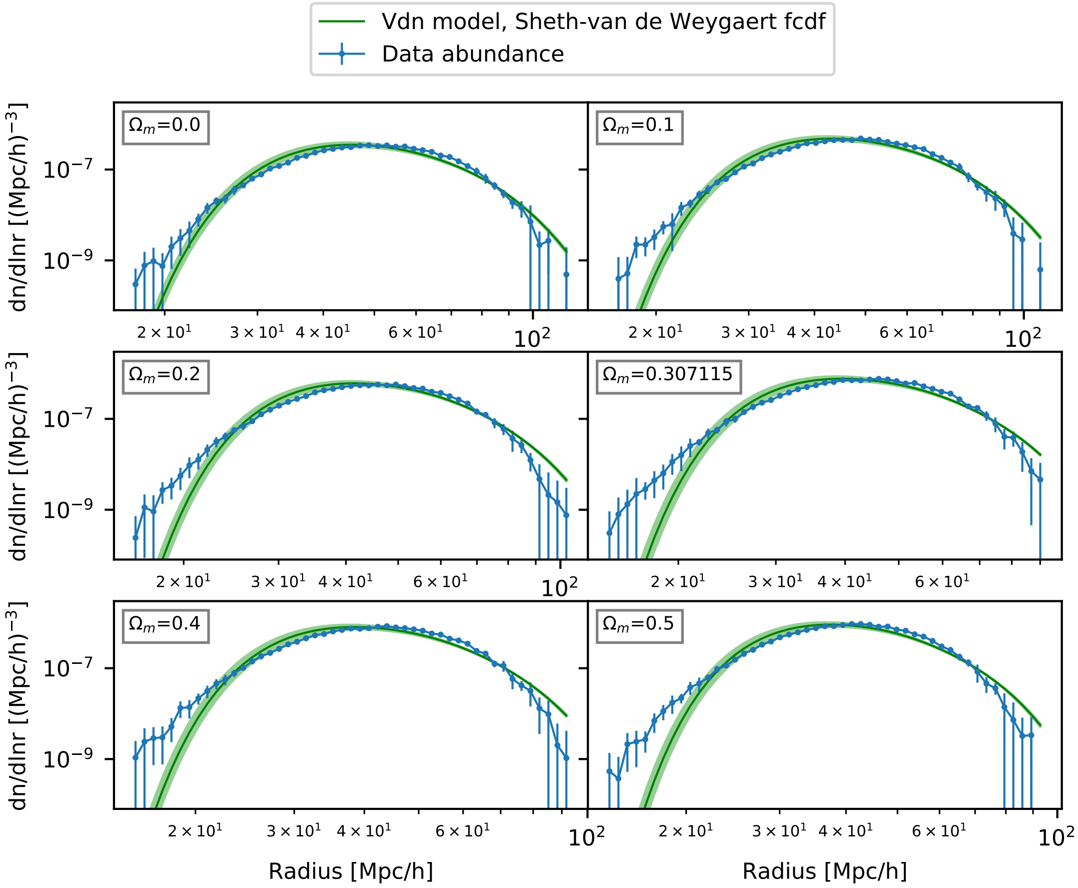


Figure 4.13: Best fits of *Vdn-SvdW model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in green. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

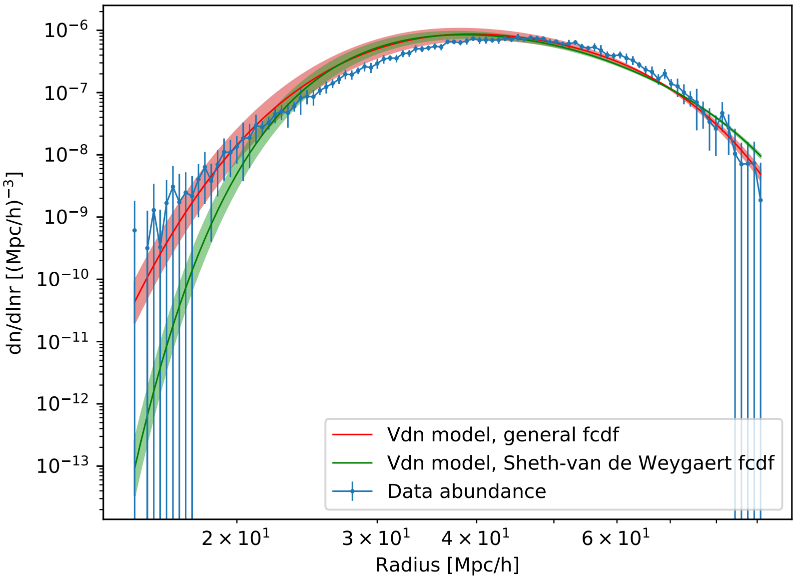


Figure 4.14: Best fits to the void abundance as a function of radius of the catalog constructed using the matter density parameter $\Omega_\mathrm{m} = 0.307115$ of the Vdn-general model (in red) and Vdn-SvdW model (in green).

$\Omega_{\rm m}$	$\left(\frac{\delta_c}{\lvert\delta_V\rvert}\right)^{\rm general}_{Vdn}$	$\left(\frac{\delta_c}{\lvert\delta_V\rvert}\right)^{\rm SvdW}_{Vdn}$
0.0	0.19	0.18
0.1	0.23	0.22
0.2	0.26	0.25
0.307115	0.29	0.29
0.4	0.28	0.27
0.5	0.29	0.28

Table 4.6: Ratio between the linear threshold for collapse δ_c and the absolute value of the linear threshold for void formation δ_V of the fitting parameters obtained for the *Vdn-general model* and *Vdn-SvdW model.*

smaller. Remember that the Sheth-van de Weyaert approximation of the first-crossing distribution function is valid for values of the parameter in the regime $\frac{\delta_c}{|\delta_V|} \gtrsim \frac{1}{4}$. In Tab. (4.6) the ratio of δ_c and the absolute value of δ_V is shown for each fit. The parameters are inside the regime of validity of the Sheth-van de Weygaert approximation for the catalogs constructed with $\Omega_{\rm m} \geq 0.2$ for both models. The $\Omega_{\rm m} = 0.0$ and $\Omega_{\rm m} = 0.1$ catalogs, however, for both models show a ratio of about 0.2, not far from the regime of validity of the approximation.

It is not surprising, therefore, that the two models are so similar to each other. The main difference at small radii reflects the difference we saw in Fig. (1.2), where a discrepancy is visible between the general function and Sheth-van de Weygaert function at small radii even in the regime of validity of the approximation. Fig. (4.14) compares, then, the two models in the regime of validity of the approximation, showing us how the approximation fits well the data at large radii, but still looks inaccurate at small radii, with respect to the general form of the first-crossing distribution function.

4.6.4 SvdW-general model

Let us look now at the best fits obtained with the Sheth-van de Weygaert model of the void size function (Eq. (1.79)) using the general form of the first-crossing distribution function (Eq. (1.73)), shown in Fig. (4.15).

The figure exhibits a poor agreement between the data and this model, with a mean value of the maximum posterior of about -700, greatly lower than the values of the models we have seen so far. The best fitting parameters are presented in Tab. (4.7). The values of the linear threshold for void formation δ_V are greater, in absolute value, than the value predicted by spherical evolution theory. The linear thresholds for collapse δ_c, unlike in the Vdn models, are much greater than the expected value. The rescaling factors, instead, are perfectly physical results. We do not have to forget, however, that the fits of this model present unsatisfactory results, reducing the relevance of looking at the fitting parameters.

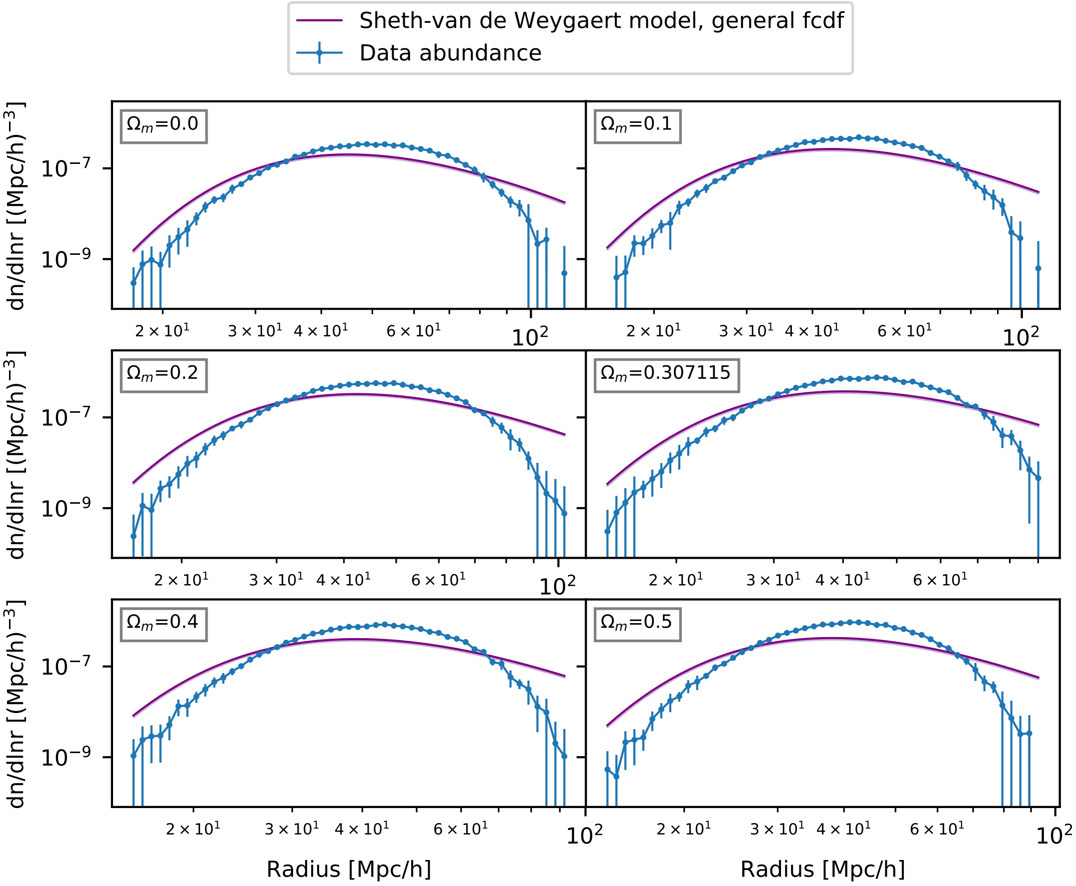


Figure 4.15: Best fits of *SvdW-general model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in purple. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

$\Omega_{\rm m}$	δ_V	δ_c	r_s
0.0	$-3.628^{+0.020}_{-0.019}$	$36.470^{+0.229}_{-0.215}$	$0.481^{+0.004}_{-0.004}$
0.1	$-3.971^{+0.022}_{-0.022}$	$39.921^{+0.251}_{-0.235}$	$0.446^{+0.004}_{-0.004}$
0.2	$-4.236^{+0.020}_{-0.019}$	$42.593^{+0.229}_{-0.213}$	$0.422^{+0.003}_{-0.003}$
0.307115	$-4.425^{+0.024}_{-0.024}$	$44.484^{+0.287}_{-0.268}$	$0.418^{+0.004}_{-0.004}$
0.4	$-4.516^{+0.023}_{-0.023}$	$45.404^{+0.278}_{-0.257}$	$0.420^{+0.004}_{-0.004}$
0.5	$-4.594^{+0.025}_{-0.024}$	$46.192^{+0.283}_{-0.266}$	$0.421^{+0.004}_{-0.004}$

Table 4.7: *SvdW-general model.* Optimal parameters obtained from the fits performed with MCMC method to the abundance of the catalogs constructed with different values of the matter density parameter $\Omega_{\rm m}$.

$\Omega_{\rm m}$	δ_V	δ_c	r_s
0.0	$-1.711^{+0.124}_{-0.121}$	$0.0147^{+0.0001}_{-0.0001}$	$0.146^{+0.013}_{-0.011}$
0.1	$-1.337^{+0.110}_{-0.115}$	$0.0154^{+0.0001}_{-0.0001}$	$0.196^{+0.019}_{-0.017}$
0.2	$-1.104^{+0.106}_{-0.110}$	$0.0161^{+0.0001}_{-0.0001}$	$0.244^{+0.026}_{-0.023}$
0.307115	$-0.870^{+0.097}_{-0.101}$	$0.0175^{+0.0003}_{-0.0002}$	$0.311^{+0.036}_{-0.031}$
0.4	$-0.806^{+0.088}_{-0.088}$	$0.0180^{+0.0004}_{-0.0003}$	$0.342^{+0.037}_{-0.031}$
0.5	$-0.729^{+0.074}_{-0.078}$	$0.0190^{+0.0005}_{-0.0004}$	$0.379^{+0.036}_{-0.033}$

Table 4.8: *SvdW-SvdW model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter $\Omega_{\rm m}$.

4.6.5 SvdW-SvdW model

Fig. (4.16) presents the best fits obtained fitting to the void catalogs abundances the Sheth-van de Weygaert model (Eq. (1.79)) with Sheth-van de Weygaert approximation of the first-crossing distribution function (Eq. (1.74)). The plots show an excellent agreement of the model to the data abundances, with a mean value of the maximum posterior of -11, the best of all models. The parameters obtained in the fits are exposed in Tab. (4.8). The values of δ_V shown in the table are the closest to the expected value of -2.731, and quite in agreement with the best value obtained also by Chan et al. (2014). The linear thresholds for collapse δ_c, on the other hand, are even smaller than the ones obtained in the Vdn models, two order of magnitude smaller than the predicted value. The rescaling factors are smaller than the predicted value of 0.58, but still physical, representing voids which have expanded more, during their evolution, than spherical ones.

Unlike the Vdn model, Sheth- van de Weygaert model shows very different results for the two different first-crossing distribution functions. The parameters of the model which uses the general form of the first-crossing distribution of Tab. (4.7) are well inside the regime of validity of Sheth- van de Weygaert approximation, with a mean value of

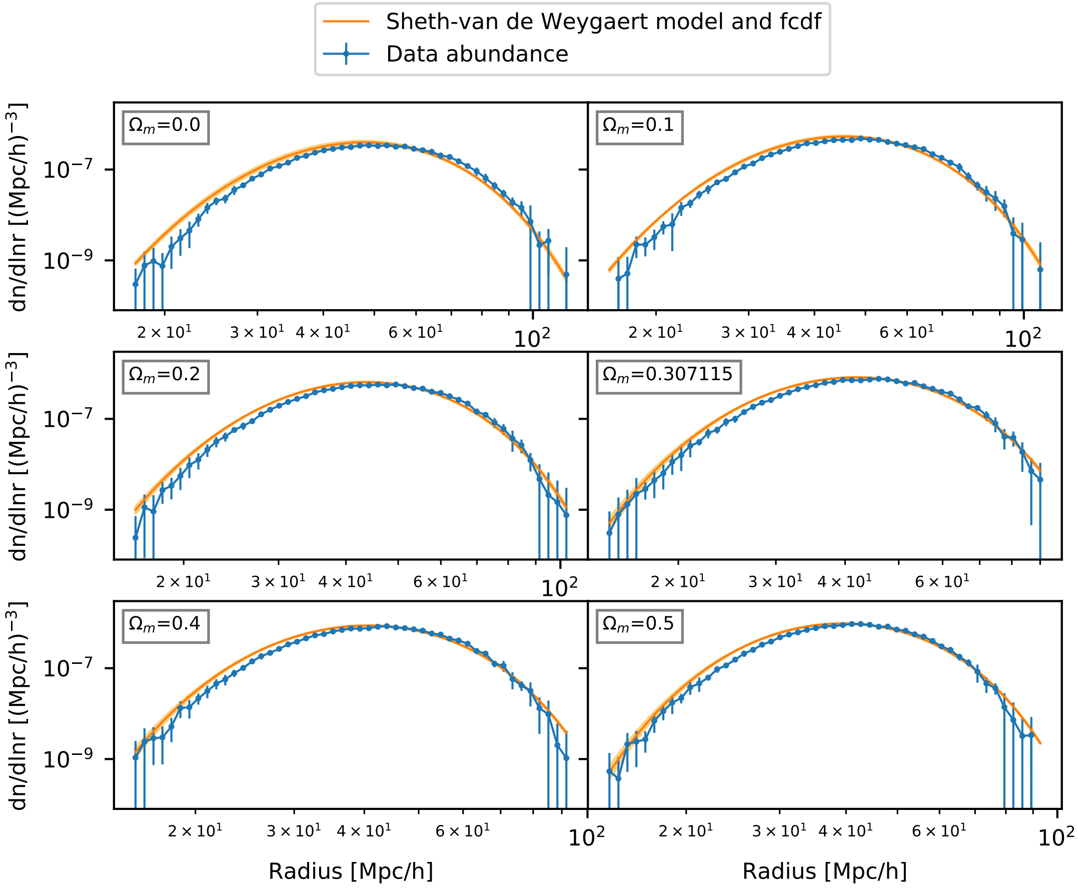


Figure 4.16: Best fits of *SvdW-SvdW model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in orange. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

$\frac{\delta_c}{|\delta_V|}$ of about 10, while the parameters of the model which uses the approximation form in Tab. (4.8) are well outside the $\frac{\delta_c}{|\delta_V|} \gtrsim \frac{1}{4}$ regime, with ratios of the order of 10^{-2}. With parameters outside the regime of validity of the Sheth-van de Weygaert approximation it seems reasonable to expect a better fit for the model which uses the general form of the first-crossing distribution function, but Fig. (4.15) and Fig. (4.16) show clearly that this is not the case. Looking at all four models we can see that the model which fits best the data abundances is the SvdW-SvdW model, with parameters outside the regime of validity of the approximation.

4.7 Cosmology dependence of the void size function

In last section we have seen the best fits of the four models to the data, finding the best fit for the SvdW-SvdW model, and a poor agreement for SvdW-general model. The two Vdn models, instead, are coherent to each other, with a better fit in the small radii range for the model which uses the general form of the first-crossing distribution function. Now we can look at the fitting parameters as a function of Ω_{m}, to study the dependence of the parameters upon the cosmology and look for an indication about the true cosmology.

Following the order of last section, let us first look at the Vdn model with general form of the first-crossing distribution function in Fig. (4.17). The linear thresholds for void formation δ_V show an increasing behaviour as Ω_{m} increases, almost linear except for a little raise at $\Omega_{\mathrm{m}} = 0.307115$. The linear thresholds for collapse δ_c compose a curve similar to a parabola with negative concavity at first approximation, with the maximum at $\Omega_{\mathrm{m}} = 0.307115$. The rescaling factors behave in a similar way to the one of δ_V, increasing almost linearly with a raise at the true cosmology value. We do not have to forget, however, that the values of r_s increases above 1, making the results unlikely. The analysis of these fitting parameters therefore has a limited relevance.

Fig. (4.18) exhibits the behaviour of the fitting parameters obtained for the Vdn-SvdW model. The trends of these parameters resemble the ones of the Vdn-general model of Fig. (4.17), but in this case we do not have unphysical results, making the analysis completely legitimate. The behaviour of δ_V and r_s is almost linearly increasing with the increasing of Ω_{m}. In this case the small raise at the value of Ω_{m} corresponding to the true cosmology is only a hint, especially in the trend of the rescaling factor. The linear thresholds for collapse δ_c show a distinct maximum corresponding to $\Omega_{\mathrm{m}} = 0.307115$, with values of δ_c that decreases for smaller and greater Ω_{m}.

Despite the poor agreement of the SvdW-general model to the data, we present the behaviour of the parameters in Fig. (4.19) for completeness of the analysis. The linear threshold for void formation shows a convex decreasing behaviour with the increasing of Ω_{m}, while the linear threshold for collapse exhibits an increasing concave trend. The rescaling factor rapidly decreases as Ω_{m} increases from 0.0 to 0.2, it shows a minimum at $\Omega_{\mathrm{m}} = 0.307115$ and then it slightly increases again as Ω_{m} increases from the true cosmology value to $\Omega_{\mathrm{m}} = 0.5$.

Fig. (4.20) presents the behaviour of the parameters obtained by the model which

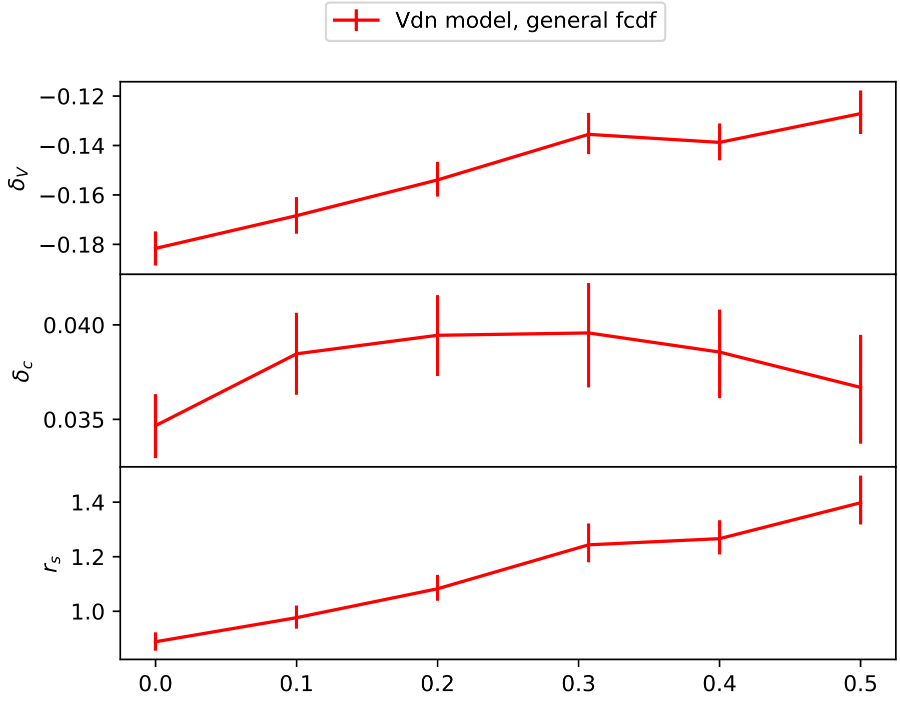


Figure 4.17: Best fit parameters as a function of $\Omega_{\rm m}$. The parameters were obtained fitting *Vdn-general model* to the abundances as a function of void radius of the catalogs constructed with different $\Omega_{\rm m}$.

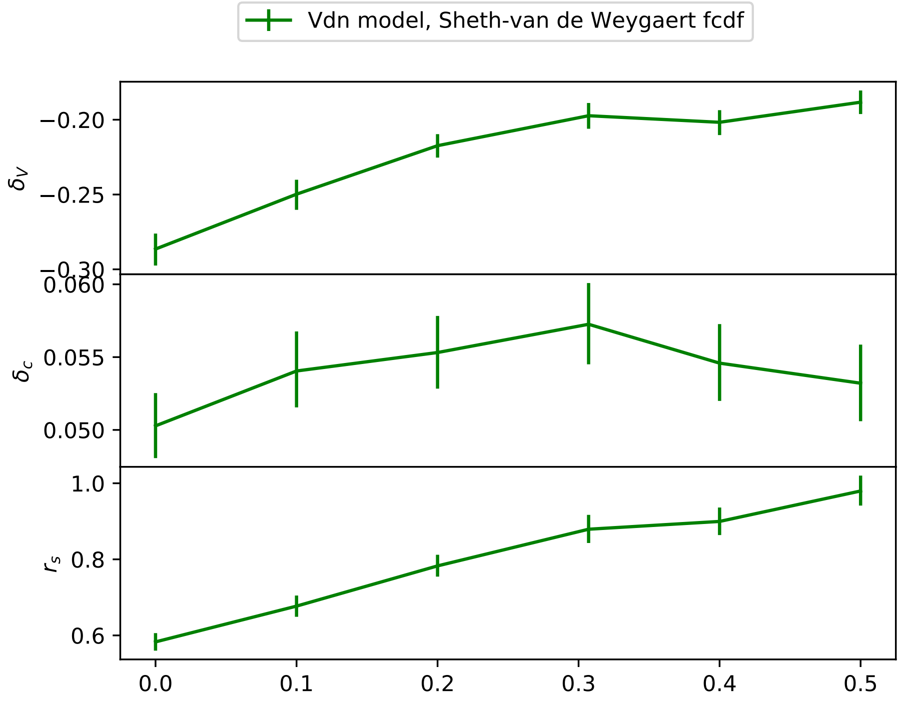


Figure 4.18: Best fit parameters as a function of Ω_m. The parameters were obtained fitting *Vdn-SvdW model* to the abundances as a function of void radius of the catalogs constructed with different Ω_m.

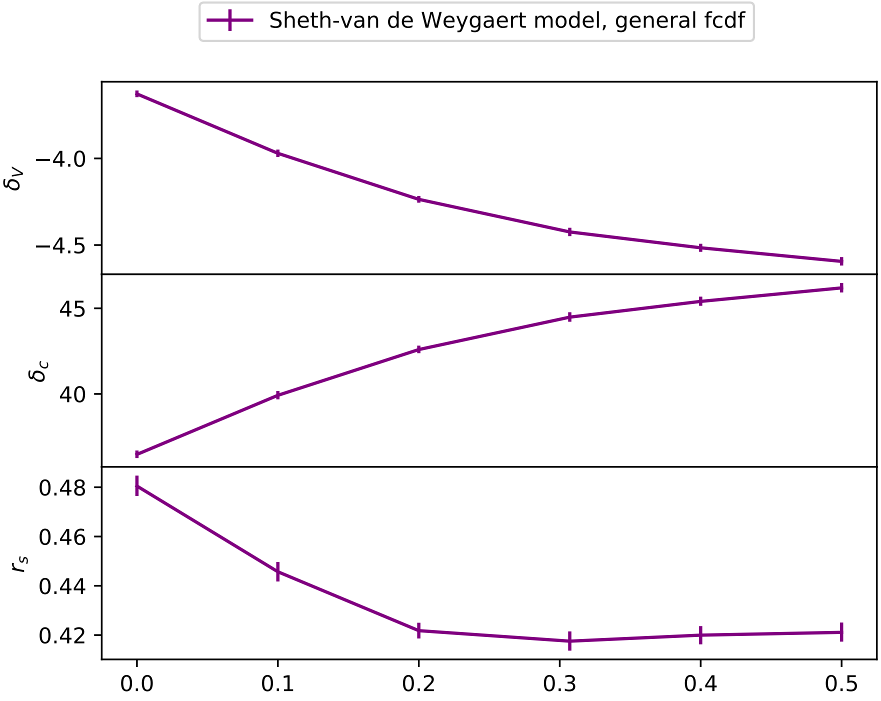


Figure 4.19: Best fit parameters as a function of Ω_{m}. The parameters were obtained fitting *SvdW-general model* to the abundances as a function of void radius of the catalogs constructed with different Ω_{m}.

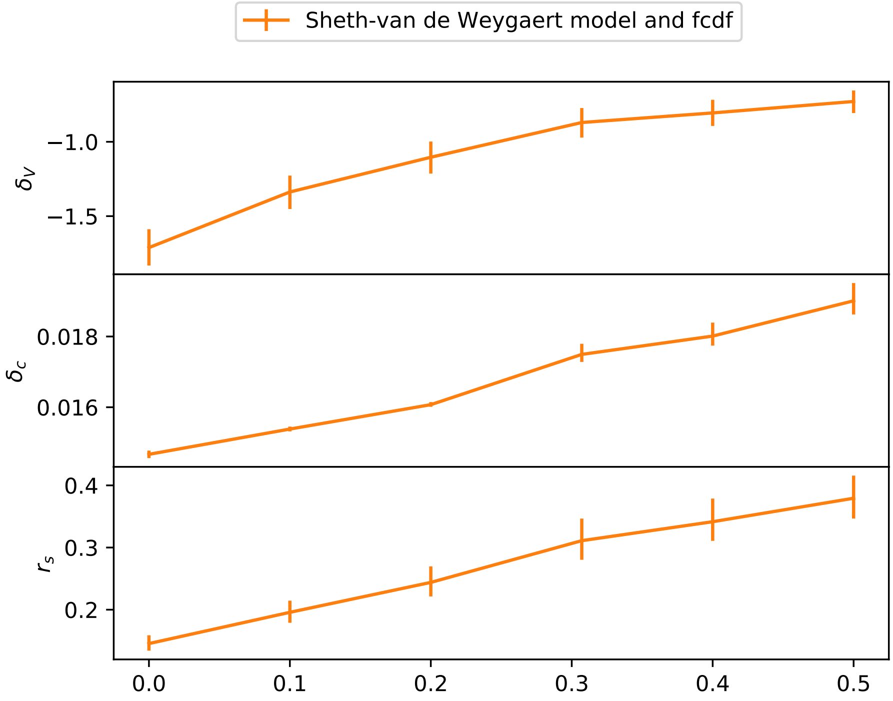


Figure 4.20: Best fit parameters as a function of Ω_m. The parameters were obtained fitting *SvdW-SvdW model* to the abundances as a function of void radius of the catalogs constructed with different Ω_m.

best fits the data abundances, i.e. SvdW-SvdW model. All parameters shows an increasing trend as Ω_{m} increases. The trend of the rescaling factor, considering the error-bars, is perfectly linear, while the small errors of the first three values of linear threshold for collapse allow us to recognize a little change of slope at $\Omega_{\mathrm{m}} = 0.307115$. The linear threshold for void formation increases in a slightly concave way, not showing any hint of the true cosmology.

The fitting parameters obtained in the analysis of the different models do not all behave in the same way: the model dependence extends to the trends of the parameters as well as their values. The parameters of the two Vdn models using the two different forms of the first-crossing distribution function show a similar behaviour as functions of Ω_{m}, suggesting a peculiarity of the true value of $\Omega_{\mathrm{m}} = 0.307115$ in the form of a small raise in the trends of the linear threshold for void formation δ_V and the rescaling factor r_s, and in the form of a maximum corresponding to the true cosmology in the case of the linear threshold for collapse δ_c. The fact that the parameters of the two Vdn models behave similarly is coherent with the agreement that the two models exhibit both in the fits and in the values of the parameters. The Sheth-van de Weygaert models, on the other hand, are very different also in this occasion. The model which uses Sheth-van de Weygaert approximation of the first-crossing distribution function, which is the best fitting model to the data abundances, barely suggests an indication of the true cosmology.

4.8 Convergence issue

In the analysis so far we have not questioned the convergence of the function of Eq. (1.73), which is an infinite series. At a first linear scale analysis the series converges rapidly after the first 20 terms, and in the fits performed in the analysis of the previous section the series has been performed using 100 terms. The poor fitting of the SvdW-general model, however, suggests the need of a closer look to this function. In Fig. (4.21) the function of Eq. (1.73) is drawn in the two cases of two sets of parameters close to the ones obtained in the fits of the Sheth-van de Weygaert model (see Tab. (4.7)) and Vdn model (see Tab. (4.4)). The different curves represent the series function calculated with different numbers of terms.

In both plots we can notice how the series does not well converge. In the upper plot we can see a shape that can resemble the void abundance of data in the range of ν between 10^{-1} and 10^{1}, but at higher values of ν the curves increase again. Increasing the number of terms shifts the second hill to higher values of ν, but even for the series function calculated using 10000 terms we can see at $\nu \approx 35$ a plateaux at $f(\nu) \approx 10^{-18}$. The difference between the two plots shows how much the function depends on the parameters δ_V and δ_c. In the lower plot the range of ν is wider, reaching values above 10^3. At small νs the tail of the main curve is visible. Notice how, depending on the number of terms, the tail of the curve suddenly drops to zero at different values of ν. In this plot we can see a second rising of the curve constructed using 1000 terms of the series, while the one constructed by 10000 terms performs the plateaux at $f(\nu) \approx 10^{-18}$

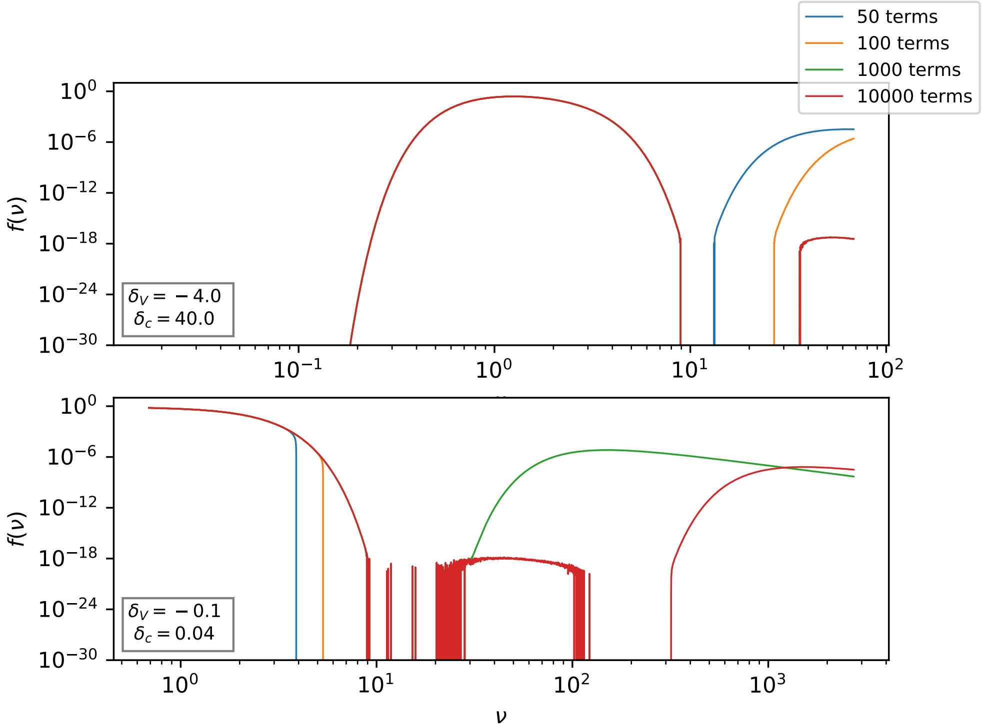


Figure 4.21: General form of the first-crossing distribution function of Eq. (1.73) calculated using 50 (blue line), 100 (orange line), 1000 (green line), and 10000 (red line) terms in the series. *Upper plot*: series function calculated using a set of parameters close to the ones obtained in the best fits of SvdW-general model of Tab. (4.7): $\delta_V = -4.0$, $\delta_c = 40.0$. *Lower plot*: series function calculated using a set of parameters close to the ones obtained in the best fits of Vdn-general model of Tab. (4.4): $\delta_V = -0.1$, $\delta_c = 0.04$.

at values of ν in the range $20 < \nu < 100$, and at $\nu \approx 300$ raises again in a curve similar to the ones with fewer terms. Even increasing the number of terms the series does not converge.

In order to better understand the behaviour of the series, in Fig. (4.22) are shown a few terms of the series (not summed), calculated using the same parameters used in Fig. (4.21), in linear scale. The two plots show the same terms calculated with a different set of parameters, and they demonstrate once again the strong dependence of the function on the parameters considered. In the upper plot we can see only negative curves, while in the lower plot the terms oscillate from the positive to the negative plane. This oscillatory behaviour could suggest an explanation for the ill convergence of the series: it is reasonable to assume that, not considering all series terms, some terms do not compensate each other, leading to an imbalance in the final calculation.

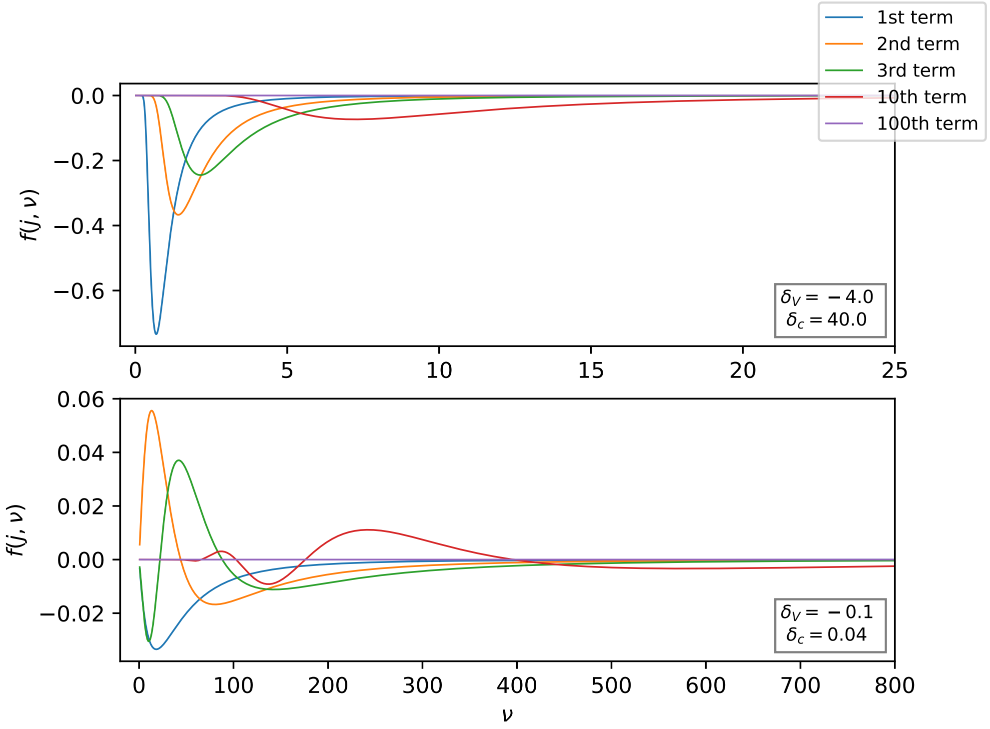


Figure 4.22: Series terms of the general form of the first-crossing distribution function of Eq. (1.73). Different lines represent the first (blue line), the second (orange line),the third (green line), the tenth (red line) and the hundredth (purple line) term of the series. *Upper plot*: series terms calculated using a set of parameters close to the ones obtained in the best fits of SvdW-general model of Tab. (4.7): $\delta_V = -4.0$, $\delta_c = 40.0$. *Lower plot*: series terms calculated using a set of parameters close to the ones obtained in the best fits of Vdn-general model of Tab. (4.4): $\delta_V = -0.1$, $\delta_c = 0.04$.

In Fig. (4.23) Sheth-van de Weygaert function (Eq. (1.74)) has been drawn over some of the curves of Fig. (4.21), i.e. the series function calculated considering 50 and 10000 terms. Sheth-van de Weygaert function, in light blue, follows the series function at $\nu > 1$, but instead of the latter it is well defined and does not increase again at higher values of ν. At small ν, however, Sheth-van de Weygaert function deviates from the general form of the first-crossing distribution function, as it is well visible in the upper plot. Let us now recall the hybrid function of Eq. (1.75), first proposed by Jennings et al. (2013):

$$\nu f(\nu) = \begin{cases} \sum_{j=1}^{\infty} \frac{2j\pi D^2}{\nu^3} \sin(j\pi D) e^{-\frac{(j\pi D)^2}{2\nu^2}} & x > 0.276 \\ \sqrt{\frac{2}{\pi}} \nu e^{-\frac{\nu^2}{2}} e^{\left[-\frac{|\delta_v|}{\delta_c}\left(\frac{D}{2\nu}\right)^2 - 2\left(\frac{D}{\nu}\right)^4\right]} & x \leq 0.276 \end{cases} \tag{4.5}$$

where $x = \frac{D}{\nu}$. This function is represented by the dash-dotted black line of Fig. (4.23)

Eq. (4.5) show a piecewise-defined equation, composed by the general form of the first crossing distribution function at $x > 0.276$, and by the Sheth-van de Weygaert approximation at $x \leq 0.276$. The point of intersection of the two functions depend both on the void in cloud parameter D and on ν. In the upper plot of Fig. (4.23) we can see that the function follows the general form of the first-crossing distribution function at small ν, then smoothly decreases following Sheth-van de Weygaert function at high values of ν, where the series function is ill defined. The hybrid function presents itself as the best first-crossing distribution model to fit our data, as it incorporates the accuracy of the general series function at $\nu < 1$ while solving the ill-definition problem at larger νs. In the next sections we will consider as first-crossing distribution function only this hybrid model.

4.9 Model and cosmology dependence of the void size function with hybrid first-crossing distribution function

In the last section we studied the convergence of the series of the general form of the first-crossing distribution function, and we concluded that the best first-crossing distribution function to fit to the abundances of the data catalogs is the hybrid function of Eq. (1.75), first proposed by Jennings et al. (2013). In this section therefore we will perform the MCMCs of the Vdn and Sheth-van de Weygaert models using this function, and we will look at the best fitting parameters obtained in the two cases. We will call the two models Vdn-hybrid and SvdW-hybrid, respectively.

4.9.1 Vdn-hybrid model

Fig. (4.24) shows the best fits obtained fitting the Vdn void size function of Eq. (1.84) using the hybrid function of Eq. (1.75) as first-crossing distribution function. The agreement of the model to the data is remarkable, quantified by a mean value of the maximum posterior of the best fits to the six catalogs data of -32. Comparing the best fits of this model to the ones we got fitting the Vdn-general and the Vdn-SvdW models of Fig. (4.12) and Fig. (4.13), respectively, we can notice the similarity of the

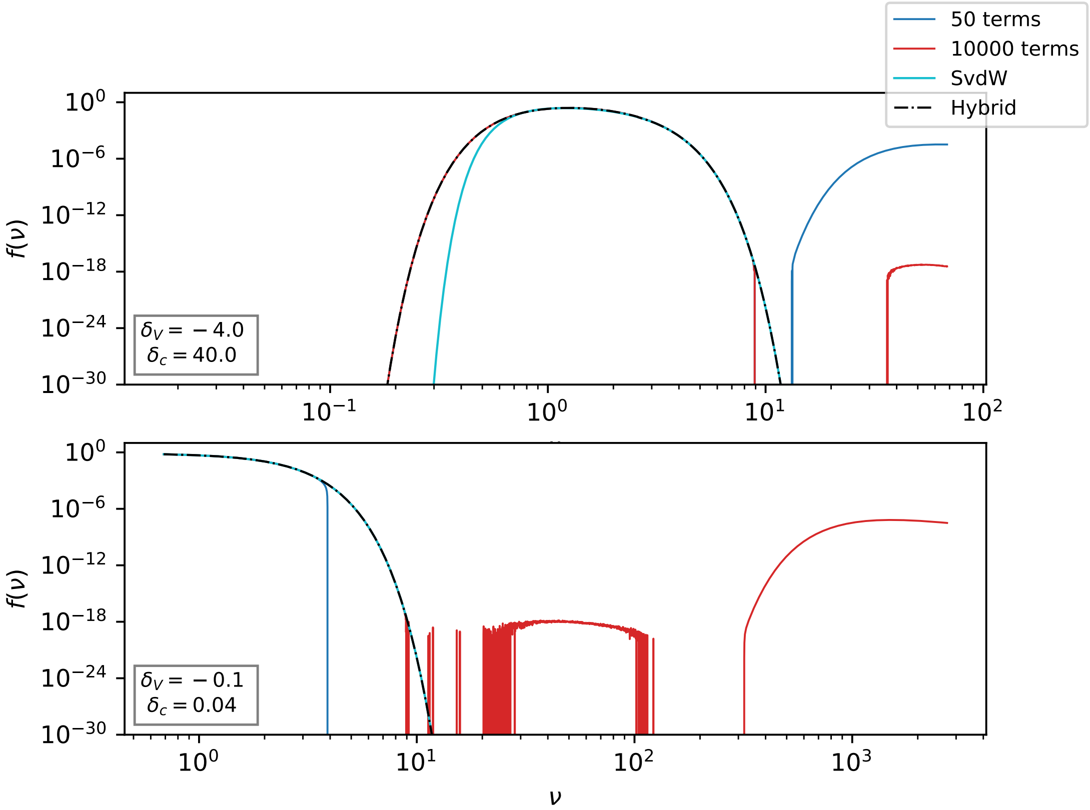


Figure 4.23: Comparison between the general form of the first-crossing distribution function of Eq. (1.73) calculated using 50 (blue line) and 10000 (red line) terms in the series, Sheth-van de Weygaert function of Eq. (1.74) (cyan line) and the hybrid function of Eq. (1.75) (dash-dotted black line). *Upper plot*: functions calculated using a set of parameters close to the ones obtained in the best fits of SvdW-general model of Tab. (4.7): $\delta_V = -4.0$, $\delta_c = 40.0$. *Lower plot*: functions calculated using a set of parameters close to the ones obtained in the best fits of Vdn-general model of Tab. (4.4): $\delta_V = -0.1$, $\delta_c = 0.04$.

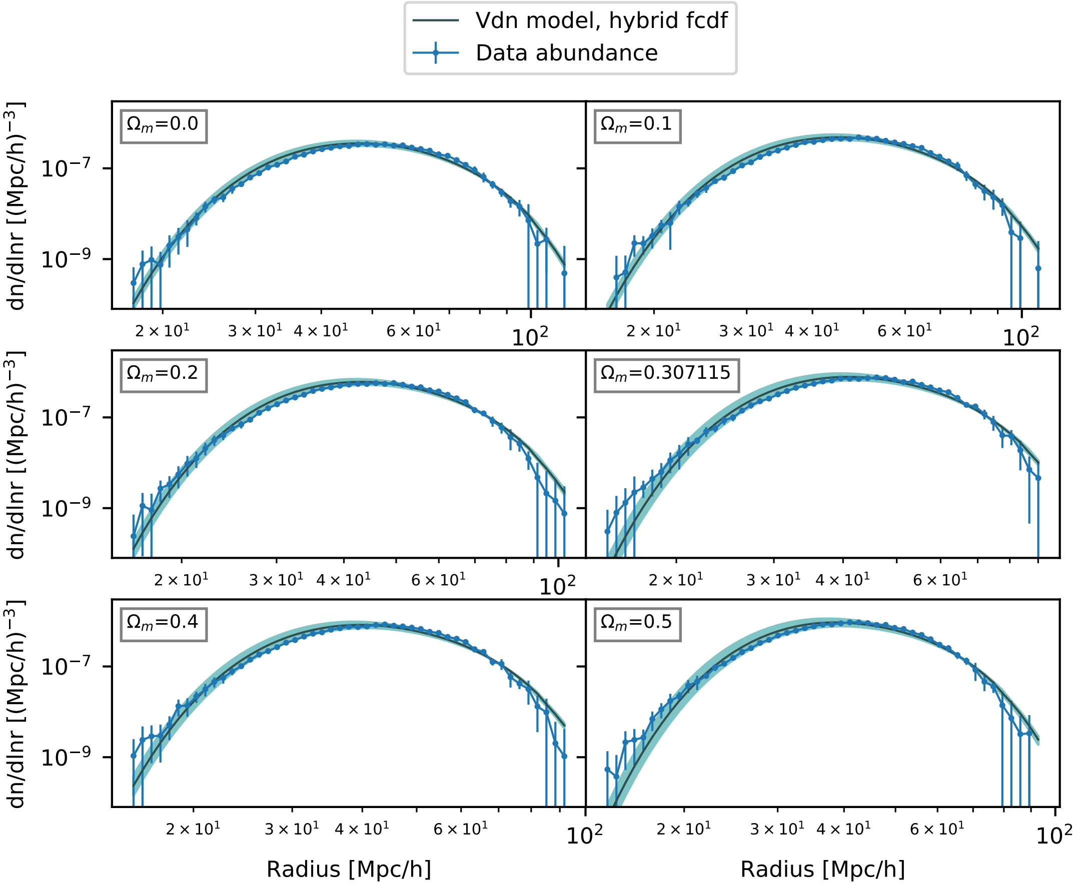


Figure 4.24: Best fits of *Vdn-hybrid model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in dark grey. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

$\Omega_{\rm m}$	δ_V	δ_c	r_s
0.0	$-0.181^{+0.007}_{-0.007}$	$0.035^{+0.002}_{-0.002}$	$0.890^{+0.037}_{-0.032}$
0.1	$-0.169^{+0.008}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$0.975^{+0.045}_{-0.038}$
0.2	$-0.154^{+0.007}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$1.085^{+0.049}_{-0.048}$
0.307115	$-0.136^{+0.008}_{-0.008}$	$0.040^{+0.003}_{-0.003}$	$1.239^{+0.074}_{-0.063}$
0.4	$-0.139^{+0.008}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$1.264^{+0.067}_{-0.059}$
0.5	$-0.128^{+0.009}_{-0.008}$	$0.037^{+0.003}_{-0.003}$	$1.387^{+0.096}_{-0.075}$

Table 4.9: *Vdn-hybrid model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter $\Omega_{\rm m}$.

fits of Fig. (4.24) to the ones obtained using the general expression. Tab. (4.9) presents the best fitting parameters of this model.

Consistently to the similarity of the best fits, also the parameters of Tab. (4.9) are close to the ones obtained in the best fits of the Vdn-general model of Tab. (4.4). The linear thresholds for void formation and for collapse are once again smaller than expected from the spherical collapse theory, and the rescaling factor becomes unphysically greater than 1 in the fits to the catalogs constructed with a matter density parameter $\Omega_{\rm m} \geq 0.2$. Fig. (4.25) shows the behaviour of the parameters as functions of $\Omega_{\rm m}$.

The trends of the parameters reflect the ones of the two Vdn models we have seen in Section 4.7. The linear threshold for void formation δ_V and the rescaling factor r_s increase almost linearly. A small raise with respect to the otherwise straight line is visible at the value of $\Omega_{\rm m} = 0.307115$, corresponding to the true cosmology. The linear threshold for collapse δ_c shows a hill-shaped behaviour, with a maximum corresponding to the true cosmology. The Vdn model with hybrid first-crossing distribution incorporates the characteristics of the two models of the Vdn void size function, i.e. the one which uses the general form of the first-crossing distribution function and the one which uses the Sheth-van de Weygaert approximation.

4.9.2 SvdW-hybrid model

In Fig. (4.26) we can see the best fits to the catalogs abundances of the Sheth-van de Weygaert void size function of Eq. (1.79) with hybrid form of the first-crossing distribution function of Eq. (1.75). This model shows a good agreement with the data abundances, with a mean value of the maximum posterior of -39. Notice how the fit to the $\Omega_{\rm m} = 0.0$ catalog appears almost perfect, while increasing the value of $\Omega_{\rm m}$ the subsequent fits agree less and less to the data, especially to voids with large radii. Tab. (4.10) shows the best fitting parameters corresponding to Fig. (4.26).

The parameters of this model are closer to the ones obtained in the Vdn models than the ones we got in the Sheth-van de Weygaert models we obtained in Tab. (4.7) and Tab. (4.8). The linear thresholds for void formation are smaller, in absolute value, than the ones of the other Sheth-van de Weygaert models, and more similar to the ones we

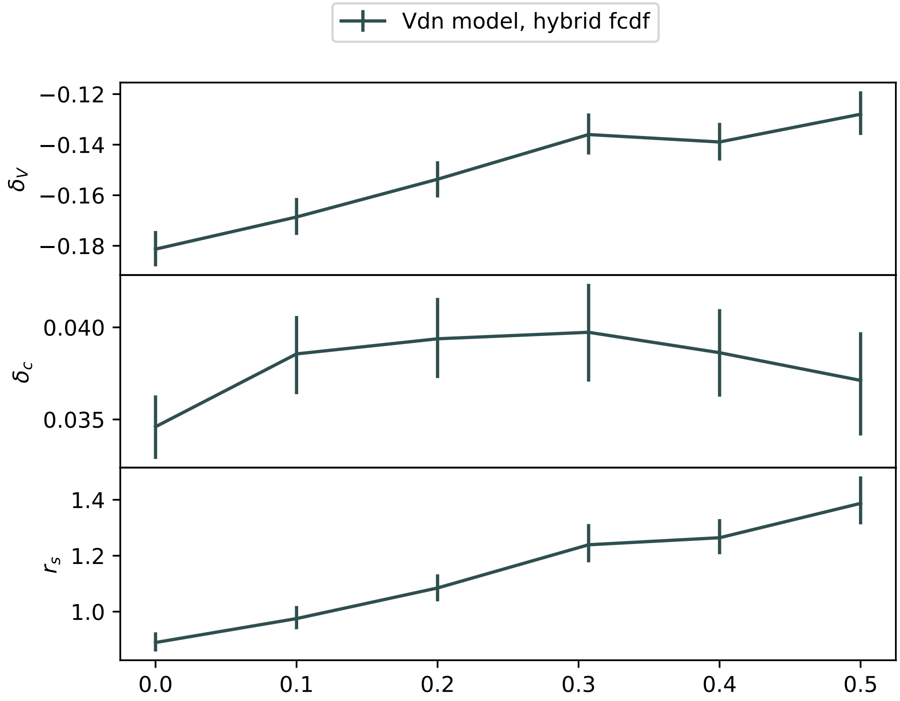


Figure 4.25: Best fit parameters as a function of Ω_m. The parameters were obtained fitting *Vdn-hybrid model* to the abundance as a function of void radius of the catalogs constructed with different Ω_m.

Ω_m	δ_V	δ_c	r_s
0.0	$-0.207^{+0.010}_{-0.010}$	$0.022^{+0.002}_{-0.001}$	$0.838^{+0.034}_{-0.030}$
0.1	$-0.194^{+0.012}_{-0.011}$	$0.031^{+0.003}_{-0.002}$	$0.897^{+0.040}_{-0.035}$
0.2	$-0.188^{+0.010}_{-0.010}$	$0.038^{+0.003}_{-0.002}$	$0.930^{+0.032}_{-0.031}$
0.307115	$-0.187^{+0.009}_{-0.009}$	$0.046^{+0.003}_{-0.003}$	$0.945^{+0.030}_{-0.028}$
0.4	$-0.184^{+0.008}_{-0.009}$	$0.048^{+0.003}_{-0.003}$	$0.976^{+0.025}_{-0.027}$
0.5	$-0.189^{+0.008}_{-0.008}$	$0.051^{+0.003}_{-0.003}$	$0.969^{+0.025}_{-0.025}$

Table 4.10: *SvdW-hybrid model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter Ω_m.

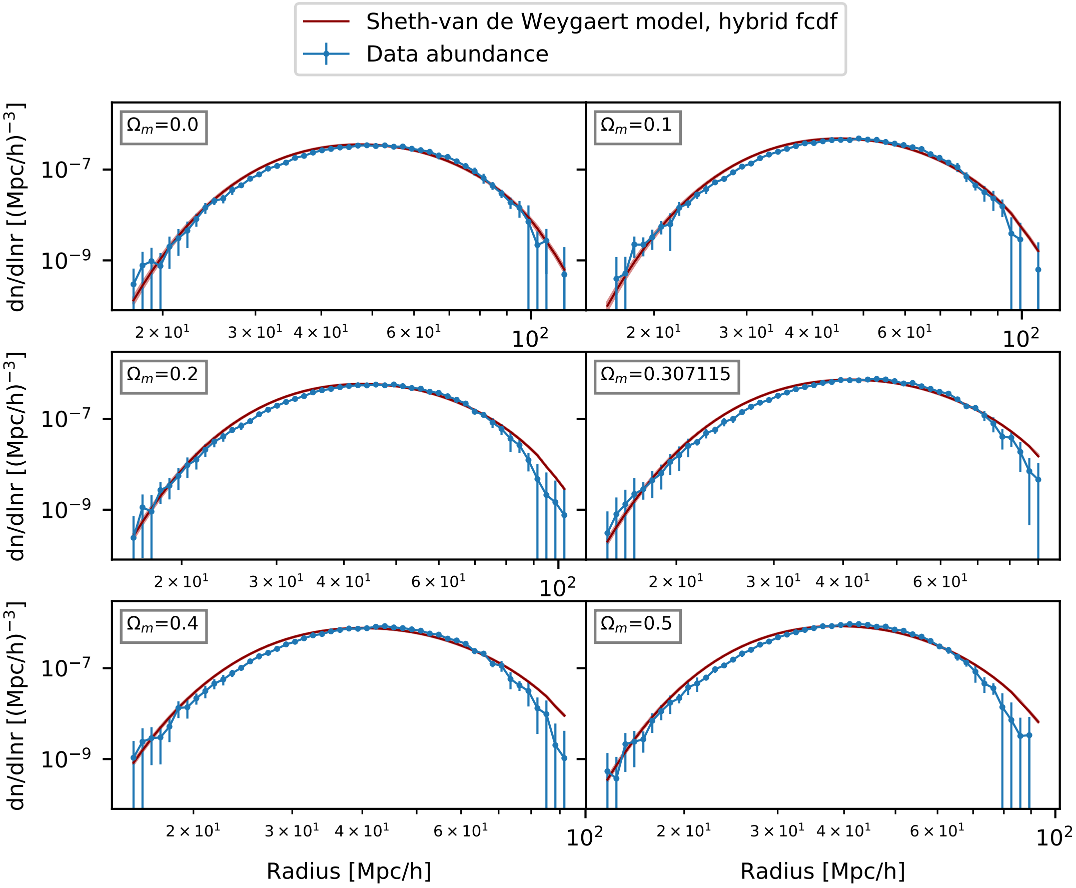


Figure 4.26: Best fits of *SvdW-hybrid model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in dark red. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

obtained in the Vdn models. The linear thresholds for void formation are very small in this case too, while the rescaling factors are greater than the value predicted in spherical evolution theory, but they don't exceed unity, keeping physical meaning. The behaviour of these parameters as functions of the matter density parameter $\Omega_{\rm m}$ is shown in Fig. (4.27).

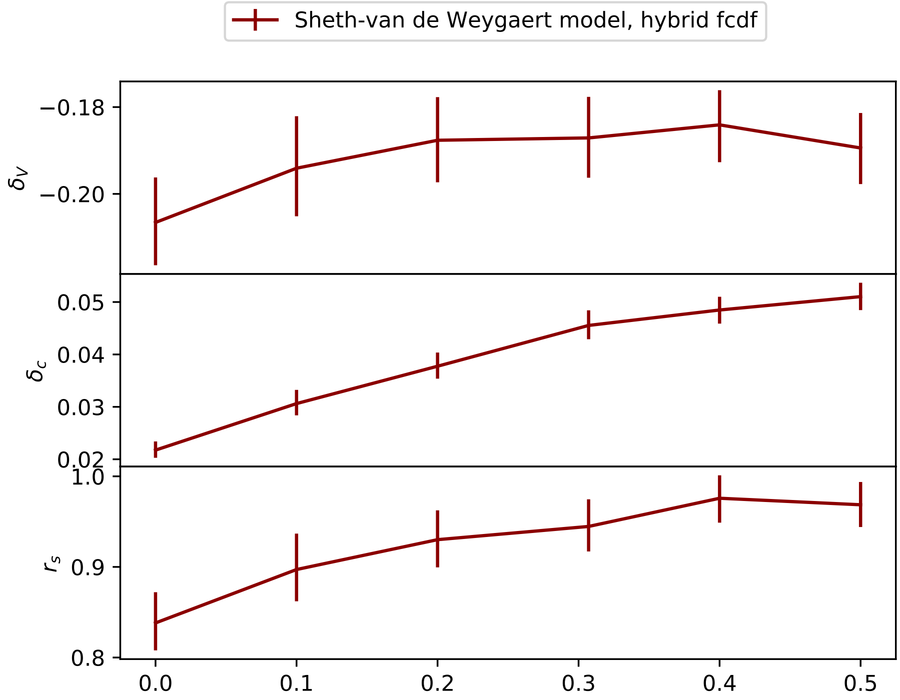


Figure 4.27: Best fit parameters as a function of $\Omega_{\rm m}$. The parameters were obtained fitting *SvdW-hybrid model* to the abundance as a function of void radius of the catalogs constructed with different $\Omega_{\rm m}$.

Comparing this figure to the cosmology dependence plots of Fig. (4.19) and Fig. (4.20) we see that the trends of these parameters are less defined than the ones of the models of Section 4.7. The linear threshold for void formation increases with the increasing of $\Omega_{\rm m}$ in the range $0.0 \leq \Omega_{\rm m} \leq 0.2$, but after $\Omega_{\rm m} = 0.2$ the value of δ_V remains almost constant. The linear threshold for collapse is characterised by smaller error-bars, allowing us to see a clear increasing trend with the increasing of $\Omega_{\rm m}$. A slight change of slope is visible at $\Omega_{\rm m} = 0.307115$. The rescaling factor values increases with the increasing of $\Omega_{\rm m}$, but the width of the error-bars prevent us to identify something definite in the behaviour of r_s at $\Omega_{\rm m} = 0.307115$.

4.10 Void-and-cloud parameter: a more natural fitting parameter?

So far we have analyzed the different models of the void size function fitting them to the catalogs void abundances as a function of void radius, considering as free parameters of the models the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s. However, looking at the form of function (1.75), one can argue that a more natural parameter of the model should be the void-and-cloud parameter D instead of the linear threshold for collapse δ_c. The linear threshold for void formation δ_V is implicitly present in the variable ν, and the rescaling factor is what differentiate the Sheth-van de Weygaert and Vdn models. The linear threshold for collapse, on the contrary, is explicitly present only once in the void-in-cloud term of Sheth-van de Weygaert first-crossing distribution function, while the void-and-cloud parameter appears more often in the first-crossing distribution functions.

It is worth noting, moreover, that the parameters of the models are not independent from one another. The rescaling factor r_s is defined by how much the voids have expanded from their initial size when they reach the stage of non-linearity defined by the non-linear threshold for void formation Δ_{sc}, as:

$$r_s = (1 + \Delta_{\mathrm{sc}})^{\frac{1}{3}} \tag{4.6}$$

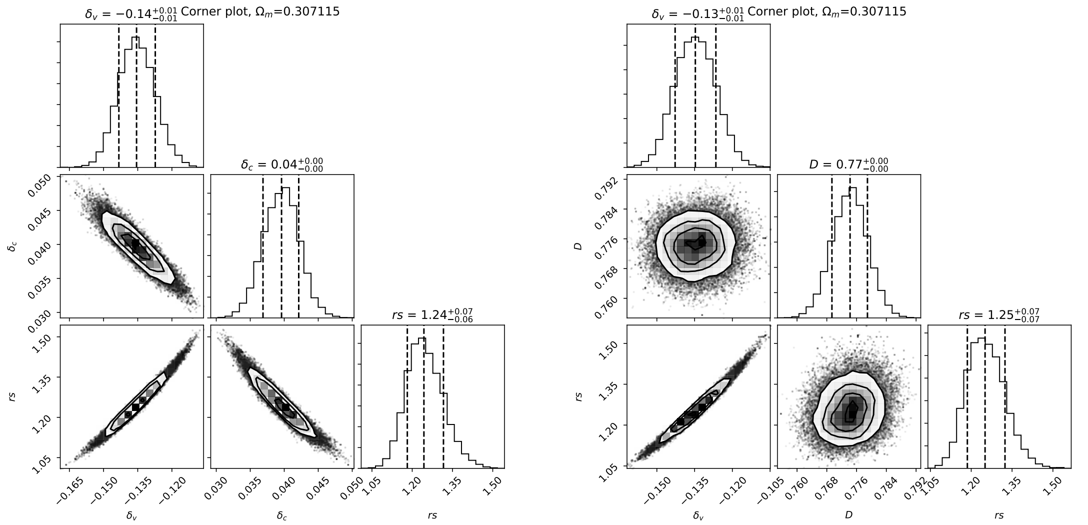
Assuming spherical collapse, studying the nonlinear evolution of density fluctuation in a random density field with Gaussian fluctuations Bernardeau (1994) proposed a relation between the non-linear and the linear thresholds for void formation:

$$\Delta_{\mathrm{sc}} = \left(1 - \frac{\delta_V}{c}\right)^{-c} - 1 \tag{4.7}$$

where $c \approx 1.594$. Assuming spherical collapse, then, Eq. (4.6) and Eq. (4.7) directly relates r_s and δ_V. The void-and-cloud parameter, as defined in Eq. (1.71), naturally depends on both the linear threshold for void formation δ_V and the linear threshold for collapse δ_c. In this section we will look at the correlation between the parameters, and we will repeat the analysis of Section 4.9 considering the void-and-cloud parameter instead of δ_c as free parameter of the models.

4.10.1 Vdn-hybrid model

In Fig. (4.28) we can see the comparison of the corner plots obtained from the MCMCs performed using Vdn void size function with hybrid first-crossing distribution function to the abundance of the catalog constructed using the true cosmology $\Omega_{\mathrm{m}} = 0.307115$. The MCMC corresponding to Fig. (4.28a) considers as free parameters δ_V, δ_c and r_s and it is the one performed in the analysis of Section 4.9, while the MCMC corresponding to Fig. (4.28b) considers as free parameters δ_V, the void-and-cloud parameter D and r_s.



(a) Free parameters: linear threshold for void formation δ_V, linear threshold for collapse δ_c, rescaling factor r_s.

(b) Free parameters: linear threshold for void formation δ_V, void-and-cloud parmaeter D, rescaling factor r_s.

Figure 4.28: Comparison between the corner plots obtained in the best fits of the *Vdn-hybrid model* with different free parameters to the catalog constructed using the matter density parameter $\Omega_\mathrm{m} = 0.307115$.

In both figures the strong correlation between the rescaling factor r_s and the linear threshold for void formation δ_V is visible. Fig. (4.28a) also shows a strong correlation of the linear threshold for collapse δ_c with both δ_V and r_s. The void-in-cloud parameter D, instead, exhibits an almost uncorrelated corner plot both with δ_V and r_s, as we can see in Fig. (4.28b). The void-and-cloud parameter D seems thus a more natural choice in the analysis of the Vdn model. Therefore we performed the MCMCs fitting the Vdn-hybrid model to the six catalogs constructed with different Ω_m in the transform from redshifts to distances, considering as free parameters δ_V, D and r_s. The best fits are shown in Fig. (4.29).

The agreement with the data is excellent, with a mean value of the maximum posterior of about -32, as in Fig. (4.24). The best fits of Fig. (4.29) are in fact almost identical to the ones of Fig. (4.24): the most relevant difference is the narrowness of the best fit curves of Fig. (4.29) with respect to the ones of Fig. (4.24), i.e. the shaded regions due to the errors of the parameters are wider in Fig. (4.24). The best fitting parameters are presented in Tab. (4.11). The linear thresholds for void formation and the rescaling factors are almost identical to the ones of Tab. (4.9), obtained from the fits which consider as free parameters δ_V, δ_c and r_s. The values of D of Tab. (4.11) are also consistent with the ones calculated form Tab. (4.9) using Eq. (1.71).

Let us now look at the cosmology dependence of the parameters of Tab. (4.11) in Fig. (4.30). The behaviours of the linear threshold for void formation δ_V and the

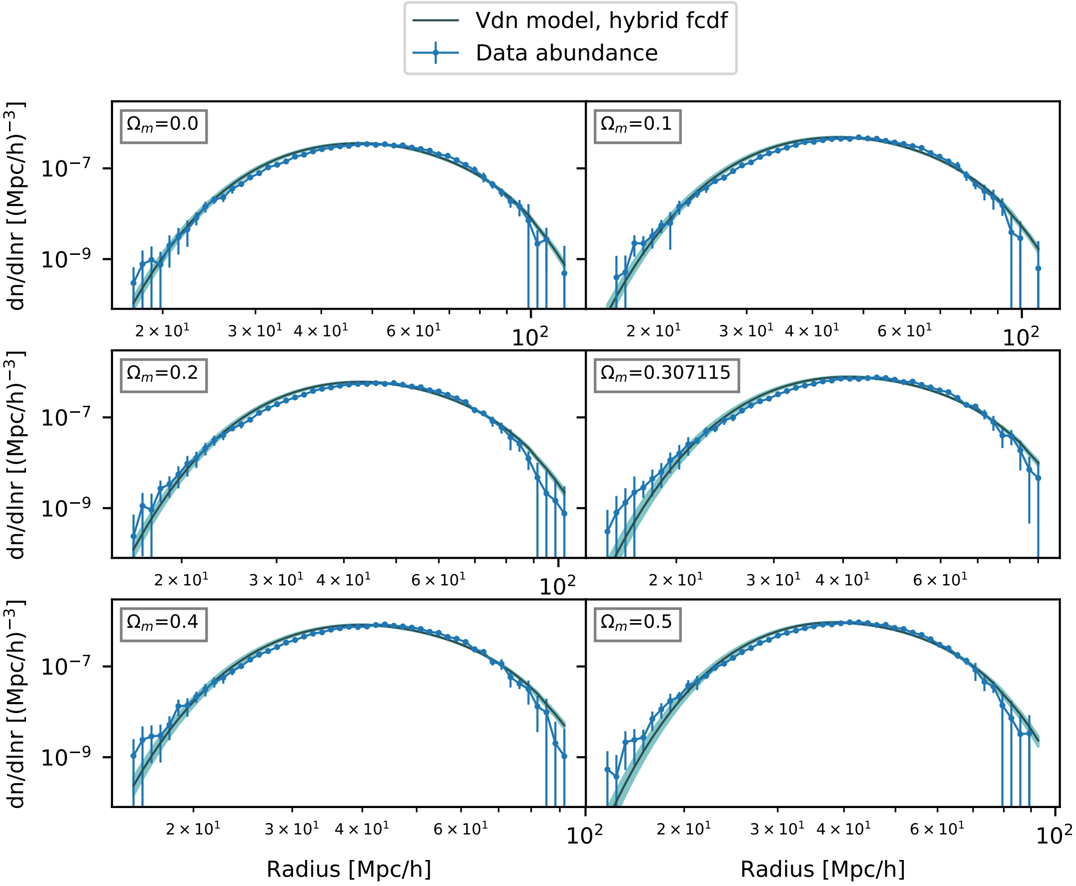


Figure 4.29: Best fits of *Vdn-hybrid model* with as free parameters the linear threshold for void formation δ_V, the void in cloud parameter D and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in dark grey. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

Ω_{m}	δ_V	D	r_s
0.0	$-0.181^{+0.007}_{-0.007}$	$0.840^{+0.004}_{-0.004}$	$0.891^{+0.035}_{-0.034}$
0.1	$-0.168^{+0.007}_{-0.007}$	$0.814^{+0.004}_{-0.004}$	$0.981^{+0.044}_{-0.039}$
0.2	$-0.153^{+0.007}_{-0.007}$	$0.796^{+0.004}_{-0.004}$	$1.091^{+0.049}_{-0.048}$
0.307115	$-0.135^{+0.008}_{-0.008}$	$0.774^{+0.005}_{-0.005}$	$1.251^{+0.073}_{-0.065}$
0.4	$-0.139^{+0.008}_{-0.007}$	$0.782^{+0.005}_{-0.005}$	$1.265^{+0.072}_{-0.060}$
0.5	$-0.127^{+0.010}_{-0.009}$	$0.776^{+0.005}_{-0.005}$	$1.399^{+0.112}_{-0.083}$

Table 4.11: *Vdn-hybrid model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter Ω_{m} with as free parameters the linear threshold for void formation δ_V, the void-and-cloud parameter D and the rescaling factor r_s.

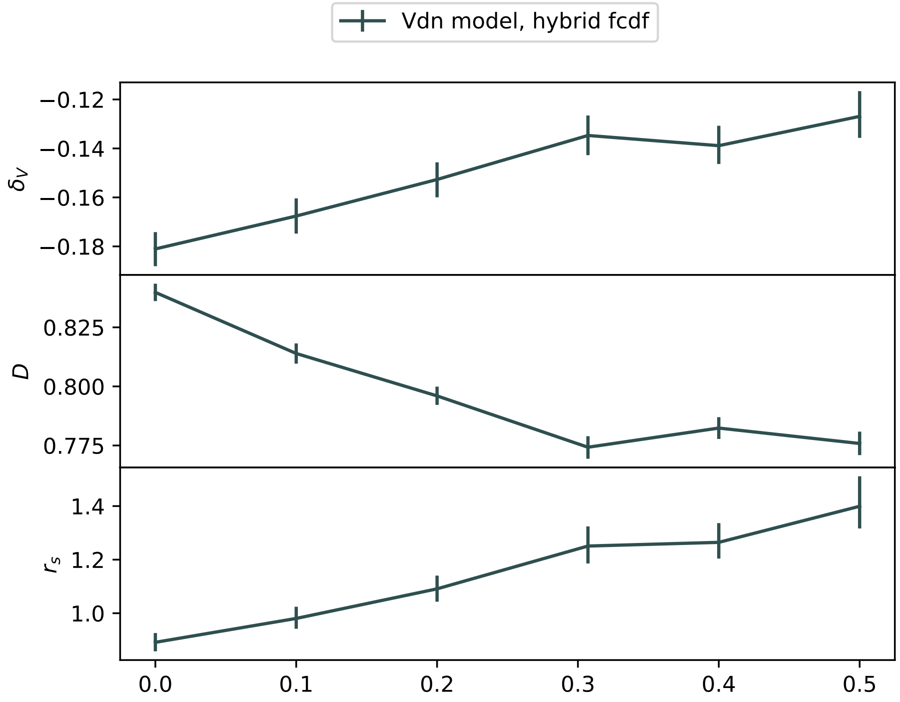
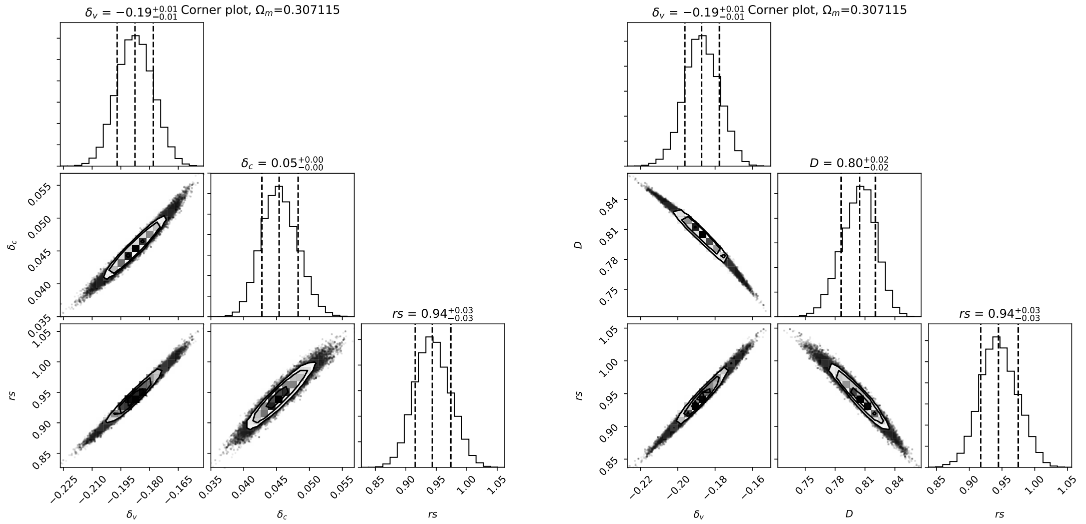


Figure 4.30: Best fit parameters as a function of Ω_{m}. The parameters were obtained fitting *Vdn-hybrid model* to the abundance as a function of void radius of the catalogs constructed with different Ω_{m}.



(a) Free parameters: linear threshold for void formation δ_V, linear threshold for collapse δ_c, rescaling factor r_s.

(b) Free parameters: linear threshold for void formation δ_V, void-and-cloud parmaeter D, rescaling factor r_s.

Figure 4.31: Comparison between the corner plots obtained in the best fits of the *SvdW-hybrid model* with different free parameters to the catalog constructed using the matter density parameter $\Omega_\mathrm{m} = 0.307115$.

rescaling factor r_s are the same of the ones of Fig. (4.25): both parameters increase with the increasing of Ω_m, showing a small raise at $\Omega_\mathrm{m} = 0.307115$. The void-and-cloud parameter D instead decreases with the increasing of Ω_m from 0.0 to 0.307115, where it shows a minimum, and then slightly increases again at $\Omega_\mathrm{m} = 0.4$, just to decrease again at $\Omega_\mathrm{m} = 0.5$. Corresponding to the true cosmology we clearly see a change of behaviour of the parameter, but due to the smallness of the fluctuations after that value and the scarceness of points we cannot say if $\Omega_\mathrm{m} = 0.307115$ corresponds to a real minimum.

4.10.2 SvdW-hybrid model

Let us now look at the comparison between the corner plots of the fits of Sheth-van de Weygaert model with hybrid first-crossing distribution function to the catalog constructed using $\Omega_\mathrm{m} = 0.307115$ in Fig. (4.31). Fig (4.31a) shows the corner plot performed during the analysis of Section 4.9, where we considered as free parameters δ_V, δ_c and r_s. All parameters are strongly correlated. Fig. (4.31b) instead corresponds to the fit which considers as free parameters δ_V, D and r_s. In the previous section we saw how the change of the choice of the free parameter from δ_c to D eased the correlation of the parameters in the Vdn model, showing in Fig. (4.28b) uncorrelated parameters. This is not the case for the Sheth-van de Weygaert void size function, where the change of the parameter considered as free merely changes the sign of the correlation: positive in the case of δ_c and negative for the void-and-cloud parameter D.

In Fig. (4.32) we can see the best fits obtained by the MCMCs considering D instead of δ_c as free parameter. Comparing this figure to Fig. (4.26) we can notice that the best

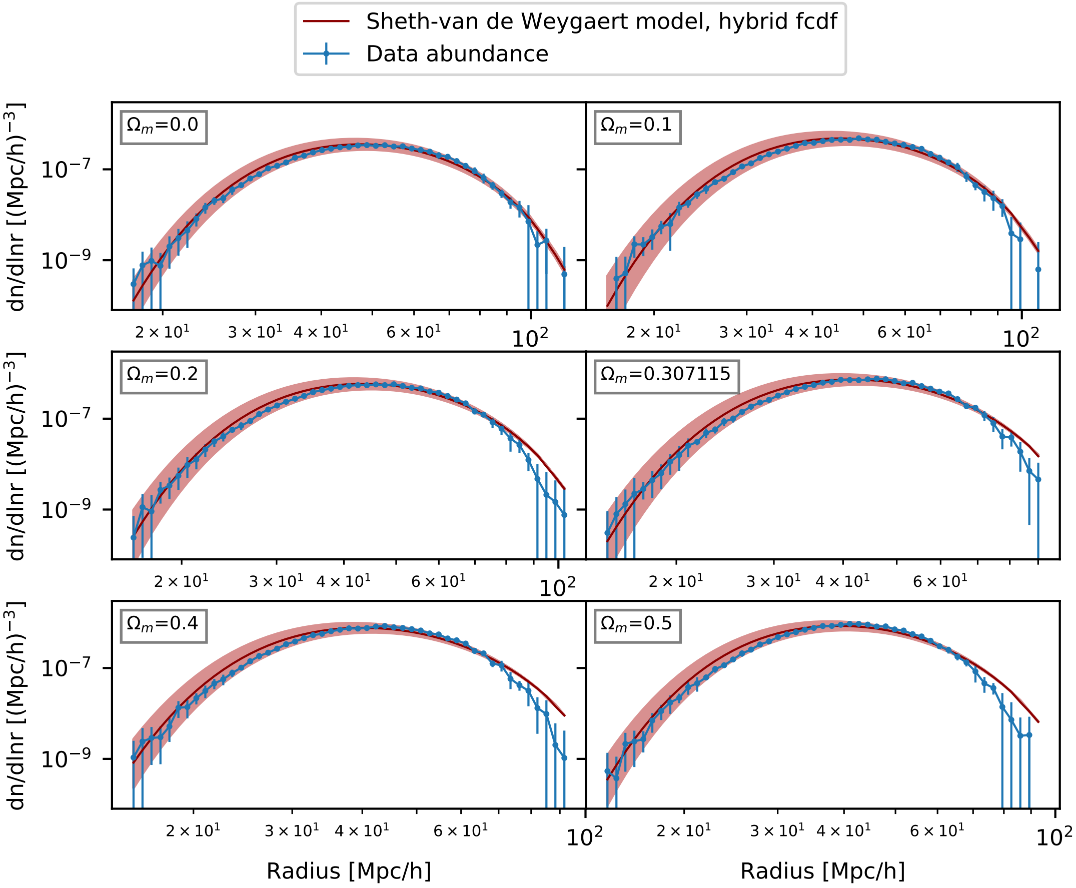


Figure 4.32: Best fits of *SvdW-hybrid model* with as free parameters the linear threshold for void formation δ_V, the void in cloud parameter D and the rescaling factor r_s. The data void abundances are represented in blue, the void size function in dark red. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

fits look identical. The mean value of the maximum posterior probability is also the same: -39. The only appreciable difference between the two figures lay in the shaded regions, which correspond to the errors of the best parameters sets. Fig. (4.32) exhibits in fact shaded regions noticeably wider than the ones of Fig. (4.26).

Tab. (4.12) shows the best fitting parameters corresponding to Fig. (4.32). Comparing these parameters to the ones of Tab. (4.10) we can see that the linear thresholds for void formation and the rescaling factors are identical in their error ranges. The values

Ω_{m}	δ_V	D	r_s
0.0	$-0.206^{+0.010}_{-0.010}$	$0.904^{+0.010}_{-0.012}$	$0.842^{+0.034}_{-0.032}$
0.1	$-0.194^{+0.012}_{-0.012}$	$0.863^{+0.016}_{-0.018}$	$0.899^{+0.040}_{-0.037}$
0.2	$-0.187^{+0.009}_{-0.010}$	$0.832^{+0.016}_{-0.017}$	$0.931^{+0.031}_{-0.031}$
0.307115	$-0.187^{+0.010}_{-0.009}$	$0.805^{+0.016}_{-0.019}$	$0.945^{+0.030}_{-0.027}$
0.4	$-0.184^{+0.008}_{-0.009}$	$0.791^{+0.016}_{-0.017}$	$0.976^{+0.026}_{-0.027}$
0.5	$-0.189^{+0.008}_{-0.008}$	$0.787^{+0.015}_{-0.015}$	$0.970^{+0.024}_{-0.024}$

Table 4.12: *SvdW-hybrid model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter Ω_{m} with as free parameters the linear threshold for void formation δ_V, the void-and-cloud parameter D and the rescaling factor r_s.

of D are also consistent with the ones calculated from δ_V and δ_c of Tab. (4.10). In Fig. (4.33) the behaviours of the parameters of Tab. (4.12) are shown. The trends of the linear threshold for void formation δ_V and the rescaling factor r_s are the same as the ones of Fig. (4.27). The void-and-cloud parameter exhibits a decreasing trend with the increasing of Ω_{m}, slightly convex.

Change the free parameter of the fits from δ_c to the void-and-cloud parameter seems to be a reasonable choice looking at the forms of the first-crossing distribution function Eq. (1.73), Eq. (1.74) and Eq. (1.75). This choice is justified also by the corner plot of Fig. (4.28b), which shows uncorrelated distributions of D both with δ_V and r_s. This improvement, however, is not shared by the Sheth-van de Weygaert model. The change of the fitting parameter also do not seem to bring any clarity in the behaviour of the parameters as functions of the cosmology.

4.11 Alcock-Paczyński correction

The analysis we have done so far is based on the construction of different catalogs with different values of Ω_{m} in the transform from redshifts to distances. Now let us focus on the meaning of this transform. In order to search for voids and analyze them the void finder needs a box in real space, but the only information available from observation catalogs are redshifts and angular positions, i.e. right ascension and declination. In order to transform them into Mpc/h scale we need to assume a cosmology, which means assume a fiducial value of Ω_{m}, that is what we have done in Section 4.1 to generate the catalogs. A discrepancy of the fiducial cosmology from the true one generates a distortion in the spatial distribution of the tracers in the survey, and it affects also the volume of voids. This is a manifestation of the Alcock-Paczynński effect (Alcock & Paczyński, 1979).

In order to better understand how the Alcock-Paczynński (AP) effect changes the volume of voids, and thus their radius, let us consider a spherical void defined by an angular radius $\Delta\theta$ and a redshift radius Δz. These radii can be converted into real space

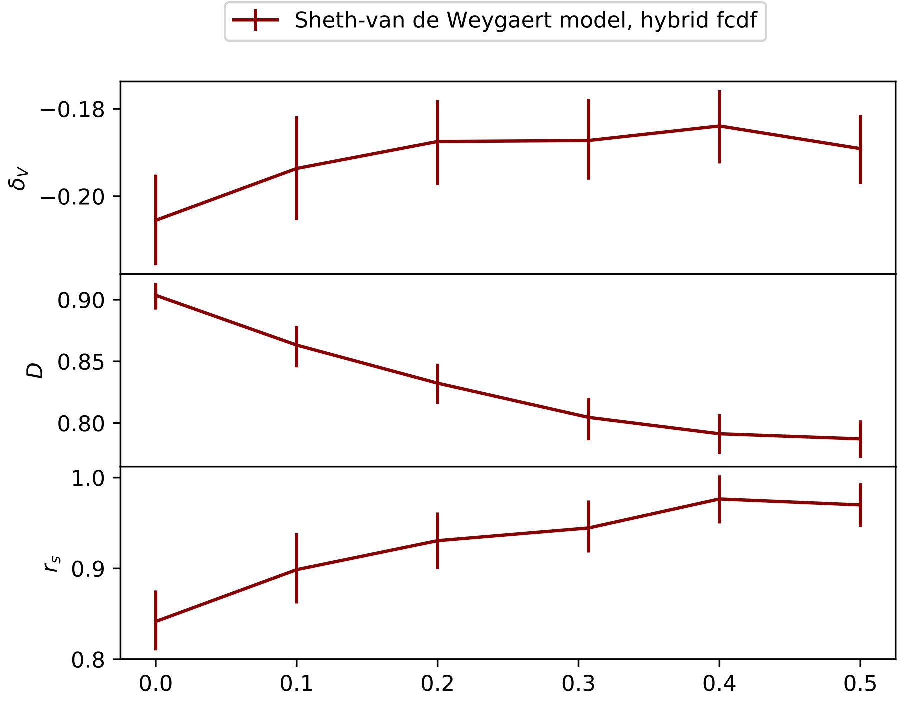


Figure 4.33: Best fit parameters as a function of Ω_m. The parameters were obtained fitting *SvdW-hybrid model* to the abundance as a function of void radius of the catalogs constructed with different Ω_m.

radii $R_\parallel$ and $R_\perp$ as

$$R_\perp = D_M(z, \Omega_\mathrm{m})\Delta z \tag{4.8}$$

$$R_\parallel = \frac{c}{H(z, \Omega_\mathrm{m})}\Delta z \tag{4.9}$$

where the Hubble parameter depends on the value of Ω_m as we have seen in Eq. (1.20), and D_M is the comoving angular diameter distance, defined as

$$D_M(z, \Omega_\mathrm{m}) = \int_0^z \frac{c}{H(z', \Omega_\mathrm{m})} = \int_0^z \frac{c}{H_0\sqrt{\Omega_\mathrm{m}(1+z)^3 + (1-\Omega_\mathrm{m})}}\,. \tag{4.10}$$

The value of the chosen Ω_m affects both radii in different ways, and if the fiducial cosmology is different from the true one a spherical void will appear as an ellipsoid, as we can see in Fig. (4.34). The semi-axis of the ellipsoid $s_\perp$ and $s_\parallel$ are defined by Eq. (4.8) and Eq. (4.9), respectively, using the fiducial value of the matter density parameter $\Omega_\mathrm{m}^\mathrm{fid}$. In Fig. (4.34) we can see the distortion of the spherical void in both directions. The real spherical void, represented by the blue circle, is distorted into the orange ellipsoid. The void finder identifies the void with the volume of the ellipsoid, and then defines an equivalent sphere of radius R^fid (red circle). The AP effect hence can lead to both an expansion or a contraction of the original real spherical void, depending on the value of the fiducial cosmology.

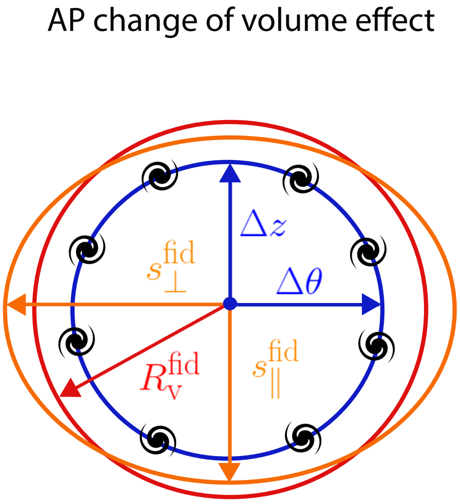


Figure 4.34: Alcock-Paczyński change of volume effect (from Correa et al. (2020)). The real spherical void is represented by a blue circle, the orange ellipsoid is the void distorted by the Alcock-Paczyński effect, and the red circle is the spherical void identified by the void finder, with same volume of the ellipsoid.

Correa et al. (2020) suggest a simple way to correct the AP effect, comparing the values of the semi-axis of the distorted ellipsoid to the real radii and thus defining the coefficients

$$q_{\perp}^{\mathrm{AP}} = \frac{s_{\perp}(\Omega_{\mathrm{m}}^{\mathrm{fid}})}{R_{\perp}^{\mathrm{true}}} = \frac{D_M(z, \Omega_{\mathrm{m}}^{\mathrm{fid}})}{D_M(z, \Omega_{\mathrm{m}}^{\mathrm{true}})} \tag{4.11}$$

$$q_{\parallel}^{\mathrm{AP}} = \frac{s_{\parallel}(\Omega_{\mathrm{m}}^{\mathrm{fid}})}{R_{\parallel}^{\mathrm{true}}} = \frac{\sqrt{\Omega_{\mathrm{m}}^{\mathrm{fid}}(1+z)^3 + (1-\Omega_{\mathrm{m}}^{\mathrm{fid}})}}{\sqrt{\Omega_{\mathrm{m}}^{\mathrm{true}}(1+z)^3 + (1-\Omega_{\mathrm{m}}^{\mathrm{true}})}}\,. \tag{4.12}$$

From $q_{\perp}^{\mathrm{AP}}$ and $q_{\parallel}^{\mathrm{AP}}$ they define the AP parameter

$$q^{\mathrm{AP}} = \left((q_{\perp}^{\mathrm{AP}})^2 q_{\parallel}^{\mathrm{AP}}\right)^{\frac{1}{3}}\,. \tag{4.13}$$

We can finally relate the radius R^{true} of the real spherical void to the radius R^{fid} identified by the void finder from the distorted void:

$$R^{\mathrm{fid}} = q^{\mathrm{AP}} R^{\mathrm{true}}\,. \tag{4.14}$$

Adopting this approach we corrected the radii of the voids of our different catalogs dividing each radius by the AP factor $q^{\mathrm{AP}}(z, \Omega_{\mathrm{m}})$. In Fig.(4.35) are shown the abundances of all voids of the catalogs after the AP correction. We can see that the abundances are almost superimposed on each other, verifying the validity of this correction. Note that Fig. (4.35) shows the complete catalogs, without any cuts.

In order to better quantify the precision of this method we performed the MCMCs fitting the Vdn and Sheth-van de Weygaert models of the void size function with hybrid first-crossing distribution function to the catalogs corrected as just described, and cutted as described in Section 4.5. In Fig. (4.36) and Fig. (4.37) the best fits to the six corrected catalogs are visible. Both figures show remarkable agreement of the models to the data of the corrected catalogs, with mean values of the maximum posterior probability of -35 for the Vdn-hybrid model and -43 for the SvdW-hybrid model. We can also note that Fig. (4.36) exhibits wider shaded regions with respect to Fig. (4.37). Tab. (4.13) and Tab (4.14) present the best fitting parameters of Vdn and Sheth-van de Weygaert void size function, respectively. The values of the linear thresholds for void formation δ_V and for collapse δ_c result greater, in absolute value, in the Sheth-van de Weygaert parameters of Tab. (4.14). On the other hand the rescaling factors r_s of Tab. (4.13) are all greater than the ones of the Sheth-van de Weygaert model, and they are all greater than 1. Fig. (4.38) presents the dependence of the fitting parameters of the Vdn-hybrid model upon Ω_{m}, while Fig. (4.39) shows the cosmology dependence of the parameters of SvdW-hybrid model.

Fig. (4.38) shows a residual dependence of the parameters upon the value of Ω_{m}, resulting in an increasing trend of both δ_V and r_s. The linear threshold for collapse δ_c instead exhibits a constant behaviour with a step between the values corresponding to the $\Omega_{\mathrm{m}} = 0.307115$ and the $\Omega_m = 0.4$ catalogs.

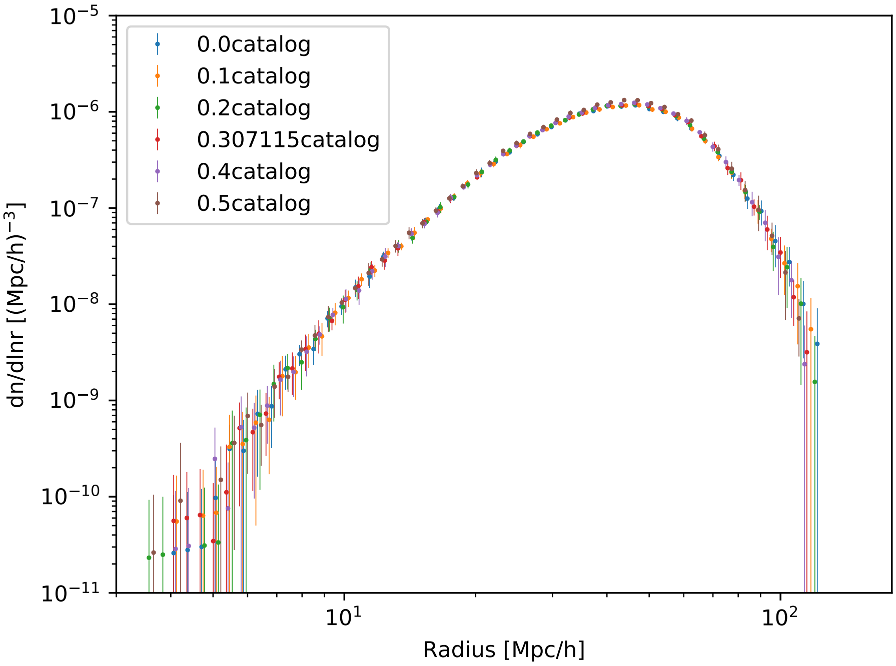


Figure 4.35: Comparison of the abundances as a function of void radius of the different catalogs constructed using different values of the matter density parmaeter Ω_m in the transform from redshifts to distances after having applied the Alcock-Paczyński correction to each void radius.

Ω_m	δ_V	δ_c	r_s
0.0	$-0.175^{+0.007}_{-0.007}$	$0.041^{+0.002}_{-0.002}$	$1.044^{+0.044}_{-0.038}$
0.1	$-0.165^{+0.007}_{-0.007}$	$0.040^{+0.002}_{-0.002}$	$1.093^{+0.045}_{-0.043}$
0.2	$-0.152^{+0.007}_{-0.007}$	$0.039^{+0.002}_{-0.002}$	$1.154^{+0.053}_{-0.047}$
0.307115	$-0.136^{+0.009}_{-0.008}$	$0.040^{+0.003}_{-0.003}$	$1.241^{+0.079}_{-0.065}$
0.4	$-0.125^{+0.010}_{-0.009}$	$0.034^{+0.003}_{-0.0038}$	$1.347^{+0.102}_{-0.084}$
0.5	$-0.122^{+0.010}_{-0.009}$	$0.035^{+0.003}_{-0.003}$	$1.347^{+0.106}_{-0.088}$

Table 4.13: *Vdn-hybrid model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter Ω_m after the Alcock-Paczyński correction, with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s.

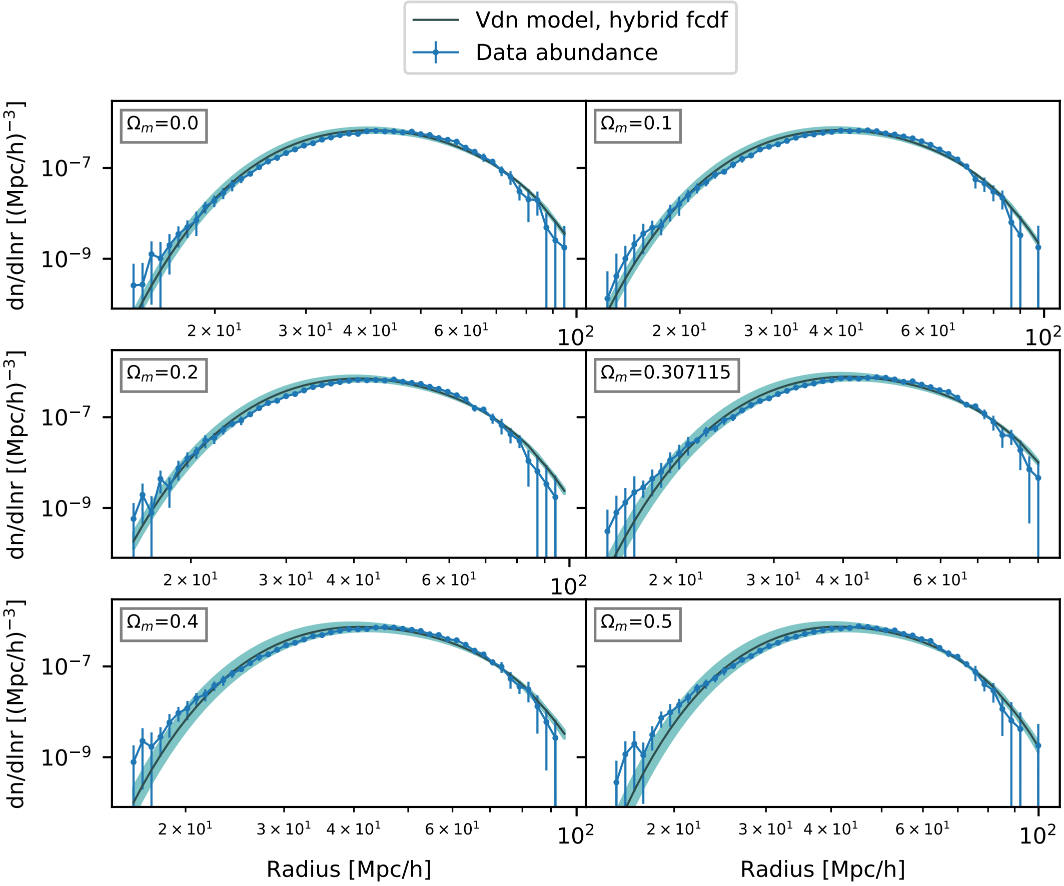


Figure 4.36: Best fits of *Vdn-hybrid model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s, to the void abundances of the catalogs after the Alcock-Paczyński correction. The data void abundances are represented in blue, the void size function in dark grey. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

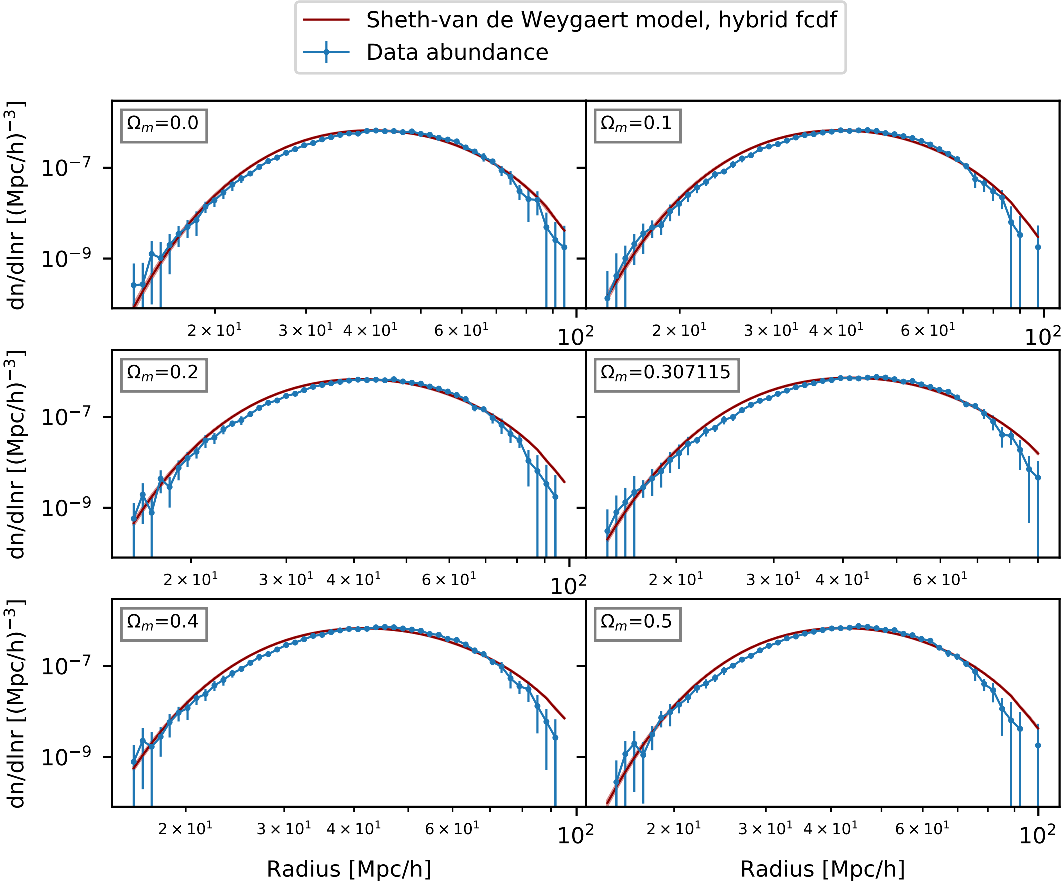


Figure 4.37: Best fits of *SvdW-hybrid model* with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s, to the void abundances of the catalogs after the Alcock-Paczyński correction. The data void abundances are represented in blue, the void size function in dark red. Shaded regions show the error ranges of the parameters of the model. In the upper left of each plot is shown the value of the matter density parameter Ω_m used in the construction of the corresponding catalog.

Ω_{m}	δ_V	δ_c	r_s
0.0	$-0.202^{+0.010}_{-0.011}$	$0.039^{+0.003}_{-0.003}$	$0.938^{+0.032}_{-0.033}$
0.1	$-0.193^{+0.010}_{-0.010}$	$0.041^{+0.003}_{-0.003}$	$0.958^{+0.032}_{-0.030}$
0.2	$-0.189^{+0.009}_{-0.010}$	$0.042^{+0.003}_{-0.003}$	$0.957^{+0.031}_{-0.029}$
0.307115	$-0.188^{+0.009}_{-0.009}$	$0.045^{+0.003}_{-0.003}$	$0.943^{+0.030}_{-0.027}$
0.4	$-0.181^{+0.008}_{-0.009}$	$0.046^{+0.003}_{-0.002}$	$0.966^{+0.027}_{-0.028}$
0.5	$-0.183^{+0.008}_{-0.009}$	$0.045^{+0.002}_{-0.002}$	$0.947^{+0.025}_{-0.027}$

Table 4.14: *SvdW-hybrid model.* Optimal parameters obtained from the fits performed with MCMC method to the abundances of the catalogs constructed with different values of the matter density parameter Ω_{m} after the Alcock-Paczyński correction, with as free parameters the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s.

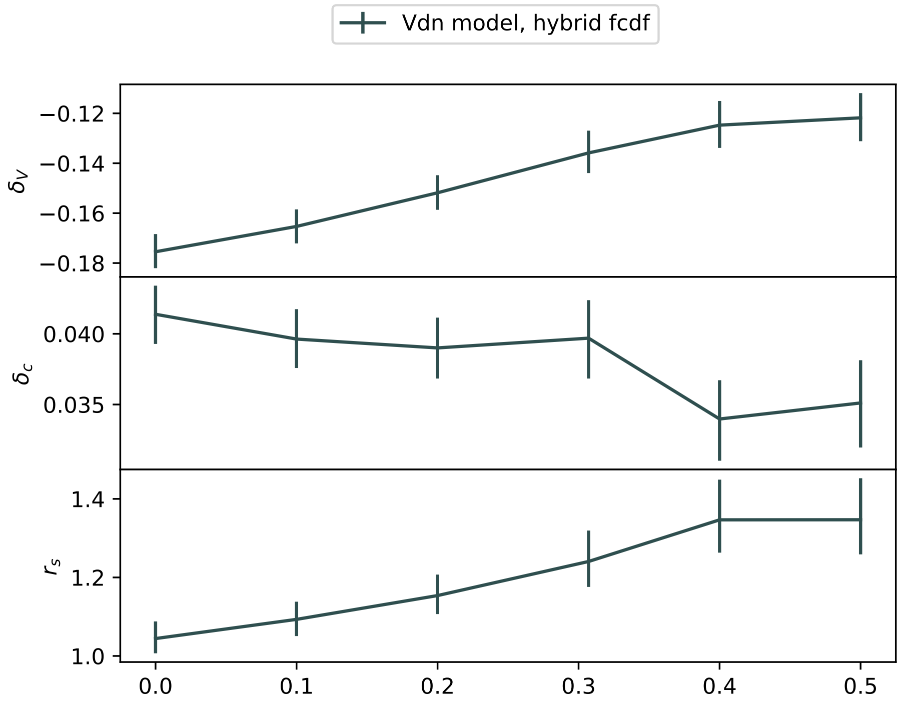


Figure 4.38: Best fit parameters as a function of Ω_{m}. The parameters were obtained fitting *Vdn-hybrid model* to the abundance as a function of void radius of the catalogs constructed with different Ω_{m} after the Alcock-Paczyński correction.

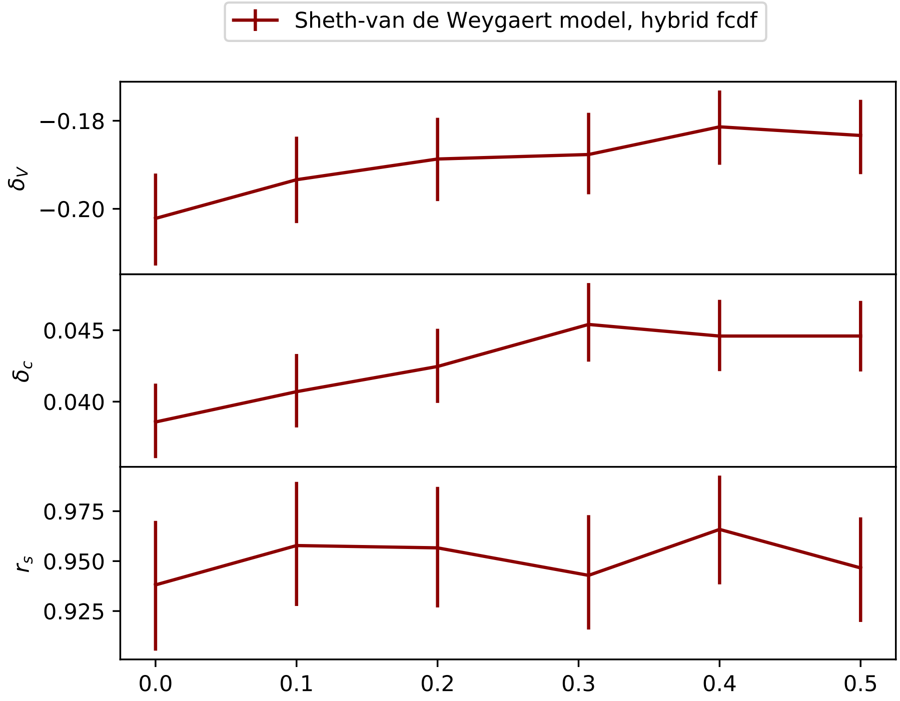


Figure 4.39: Best fit parameters as a function of Ω_m. The parameters were obtained fitting *SvdW-hybrid model* to the abundance as a function of void radius of the catalogs constructed with different Ω_m after the Alcock-Paczyński correction.

The fitting parameters of the Sheth-van de Weygaert model of Fig. (4.39) show a constant behaviour, with only a slight increase from $\Omega_{\mathrm{m}} = 0.0$ to $\Omega_{\mathrm{m}} = 0.307115$ in the δ_c case.

Chapter 5

Discussion and conclusion

In this conclusive chapter we will discuss the results obtained in the analysis in Section 5.1 and propose possible future developments of this work in Section 5.2. In the conclusion of Section 5.3 we will sum up the main results of this work.

5.1 Discussion

We have studied model and cosmology dependence of the void size function of the MultiDark PATCHY mock galaxy catalog. In order to do so we constructed six different catalogs using different values of the matter density parameter Ω_{m} in the transform from redshifts to distances.

As a first analysis we studied the Sheth-van de Weygaert model of Eq. (1.79) and the Vdn model of Eq. (1.84) of the void size function, using Press-Schechter first-crossing distribution function of Eq. (4.2), excluding voids with small radius and thus neglecting the void-in-cloud term. We found a good agreement of the models to the data abundances (see Fig. (4.4)), but the Vdn model exhibits extremely small absolute values of the linear threshold for void formation and unphysically large values of the rescaling factor (see Tab. (4.3)). In the behaviour of the parameters of both models it is not noticeable any specific trend.

The presence of shot noise leads the void finder VIDE (Sutter et al., 2015) to identify spurious voids in the survey; to clean the catalogs from these random voids and consider the void-in-cloud effect, dominant at small radii, we applied a multivariate analysis: comparing the void properties of the void catalog constructed using the true cosmology to the one constructed from the random PATCHY sample, we applied two cuts on the catalogs, in order to trim spurious voids while keeping as many real voids as possible. The first cut has been applied to the core density, i.e. the density of the largest Voronoi cell in a void, keeping only voids with core density smaller than 2.5. The second cut regarded the compensation, that is the ratio between the number of targets present in a void and the void volume. The choice has been that to keep only voids with a compensation greater than 5.

To analyze the void size function we fitted to the void abundances as a function of the void radius the two models of the void size function:

- Sheth van de Weygaert model of Eq. (1.79)
- Vdn model of Eq. (1.84)

with the two different models of the first-crossing distribution function:

- general expression of Eq. (1.73)
- Sheth-van de Weygaert approximation of Eq. (1.74).

After a first attempt using the Least Squares method, which led to a ridge problem in the optimization of the fit, the Markov Chain Monte Carlo method has been used to perform the fits, considering the linear threshold for void formation δ_V, the linear threshold for collapse δ_c and the rescaling factor r_s as free parameters. Since we are treating all three parameters as free in the fitting, we are not imposing any strong assumption on the evolution or on the properties of voids, such as the sphericity assumption often used in the literature (see e.g. Chan et al. (2014)). Nevertheless we compare the optimal parameters of the fits to the ones derived from spherical evolution theory, in order not to loose the physical meaning of the parameters.

The two Vdn void size functions with different first-crossing distribution functions coherently fit the data abundances of the different catalogs: Fig. (4.14) shows that the model which uses the general form of the first-crossing distribution function better agrees with the data at small radii, where the void-in-cloud effect is dominant. The values of the linear threshold for void formation and for collapse obtained in the best fits of the two models are also quite similar (see Tab. (4.4) and Tab. (4.5)), and inside or very near to the regime of validity of the Sheth-van de Weygaert approximation (see Tab. (4.6). It has to be noted, however, that the values of the rescaling factor we got from the Vdn-general model exceed unity in the fits to the abundances of catalogs constructed with $\Omega_{\mathrm{m}} \geq 0.2$. A rescaling factor greater than one represents voids which, instead of expanding during their evolution, can shrink to smaller sizes. In Section 1.2.2 we have seen that, according to spherical evolution, unlike overdensities of the matter field underdensities never reach a point of turn-around and continue to expand. A shrinking void is therefore theoretically legitimate only in the case of the void-in-cloud effect, where voids shrink due to the collapse of the cloud that surround them. The physical meaning of a rescaling factor which exceeds unity, therefore, is that all voids of the catalog are subject to the void-in-cloud effect, an unlikely result.

The Sheth-van de Weygaert void size function fits with different first-crossing distribution functions appear very different from each other, regarding both the agreement to the data (see Fig. (4.15) and Fig. (4.16)) and the best fitting values (see Tab. (4.7) and Tab. (4.8)). The SvdW-general model fits poorly the data void abundances, with optimal values of the linear threshold for collapse much greater than the value predicted from spherical collapse. The SvdW-SvdW model proves itself as the best model to fit the void abundances of the data catalogs. The parameters obtained in the MCMC fits, however, result outside the regime of validity of the approximation function, making the excellent agreement of this model to the data rather unexpected.

Studying the convergence of the general form of the first-crossing distribution function of Eq. (1.73), we found that neglect some terms of the series function may lead to unforeseen results at high values of $\nu = \delta_V/\sigma$, where σ is the variance at the length scale we are considering. In that regime, however, the Sheth-van de Weygaert function of Eq. (1.74) is well defined. Therefore we chose to consider as first-crossing distribution function the hybrid function of Eq. (1.75), first proposed by Jennings et al. (2013). Eq. (1.75) is a stepwise-defined function composed by the general form of Eq. (1.73) at high values of D/ν, and by the Sheth-van de Weygaert approximation at small values of D/ν, where D is the void-and-cloud parameter defined in Eq. (1.71). Both the Vdn and the Sheth-van de Weygaert void size function with the hybrid first-crossing distribution function show a remarkable agreement with the data (see Fig. (4.24 and Fig. (4.26)). The best fits and optimal fitting parameters of the Vdn model imitate the ones of the Vdn-general model (see Tab. (4.9)), including the unphysical values of the optimal rescaling factors. The best fitting parameters of the Sheth-van de Weygaert void size function, on the other hand, are dissimilar from the ones obtained in both Sheth-van de Weygaert models previously studied. The values of the linear threshold for void formation in Tab. (4.10), in particular, are one order of magnitude smaller than the ones of both Tab. (4.7) and Tab. (4.8).

Almost all models fitted so far produced values of the linear threshold for void formation smaller, in absolute value, than the predicted value of -2.731 from spherical evolution theory. From Fig. (1.1) we can see that a small absolute value of δ_V implies that a greater number of random walks will cross this threshold value, increasing the number of underdensities that, reaching the stage on non-linearity, become cosmic voids. Similarly, a small value of the linear threshold for collapse δ_c increases the number of random walks which cross that threshold, increasing therefore the abundance of collapsing objects. The consequence on the void size function is an increase in the importance of the void-in-cloud effect. Looking at the lower row of Fig. (1.1) we can notice how, as δ_c decreases, it becomes easier and easier to find this type of structure in the evolution of the matter density contrast. In all fitting parameters, with the exception of the case of SvdW-general model, the values of δ_c obtained are two order of magnitude smaller than the value predicted from spherical collapse, denoting an enhanced weight of the void-in-cloud effect with respect to the theories proposed so far. As we have already mentioned, the recurrent result of a rescaling factor greater than 1 corresponds to voids which, instead of expanding, could have shrunk to smaller sizes. The only way for a cosmic void to shrink without loosing physical meaning is to be inside a cloud. A rescaling factor that exceeds unity therefore implies that almost all voids of the catalog are subjected to the void-in-cloud effect. It is worth insisting on the fact that these unphysical values has been obtained in the study of the Vdn model both with general expression of the first-crossing distribution function and with its hybrid form of Eq. (1.75).

Via the parameters obtained from the fits we searched for a trend that could allow us to better understand the dependence of the void size function upon the fiducial chosen cosmology. In particular we would like to find a suggestion of the true cosmology in the behaviour of the parameters. Therefore we plotted the values of the parameters we got from the fits of each model as a function of the matter density parameter Ω_{m} used in the construction of the catalog to whom the fit was performed.

The behaviours of the fitting parameters of the Vdn-general and Vdn-SvdW models appear similar to each other (see Fig. (4.17) and Fig. (4.18)). The linear threshold for void formation δ_c shows an hill-shaped behaviour, with maximum corresponding to the true cosmology $\Omega_{\mathrm{m}} = 0.307115$. The curve presents itself as more regular in the case of the Vdn-general model, while in the Vdn-SvdW model the maximum at the true cosmology appears more evident. The trends of δ_V and r_s in both Vdn models are almost linearly increasing. A small anomaly in the otherwise straight lines is visible at $\Omega_{\mathrm{m}} = 0.307115$. The behaviour of the parameters obtained fitting to the data abundances the Vdn-hybrid model mirrors the one obtained for the Vdn void size functions just described (see Fig. (4.25)): δ_c shows an hill-shaped trend with maximum corresponding to the true cosmology matter density parameter, δ_V and r_s increase with the increasing of Ω_{m} linearly, with the exception of a small raise at $\Omega_{\mathrm{m}} = 0.307115$. We can thus reasonably conclude that a hint about the true cosmology is visible in the parameters of the Vdn model, even though two of the three models present values of the rescaling factor that exceed unity.

The behaviours of the parameters obtained from the fits of the Sheth-van de Weygaert void size function models with different first-crossing distribution functions are very different from each other, but due to the poor agreement of the SvdW-general model we will focus on the other models. The SvdW-SvdW model (see Fig. (4.20)) shows an increasing linear trend in the behaviour of all parameters. The anomaly in the linearity present in the Vdn model is only slightly hinted in the trend of the linear threshold for collapse. This behaviour of δ_c is shared by the SvdW-hybrid model (see Fig. (4.27)), in which case the other two parameters do not show a clear indication of the value of the true cosmology. Sheth-van de Weygaert model, therefore, does not allow us to identify the true cosmology value from the fitting parameters. Only small hints are present, not sufficient to give a definite outcome.

Looking at the correlation between the parameters, we analyzed the possibility to use the void-and-cloud parameter D instead of the linear threshold for collapse as free parameter of the fits. This choice seemed favourable in the study of the Vdn void size function, since it led to an uncorrelated corner plot between D and both δ_V and r_s (see Fig. (4.28)). On the other hand this improvement is not shared by the Sheth-van de Weygaert model (see Fig. (4.31)), and the change did not made the behaviour of the parameters clearer.

In the end we followed Correa et al. (2020) and we applied an Alcock-Paczyński correction to the void radii of all catalogs, rescaling each radius by a factor which depends on both the true cosmology and the fiducial cosmology used in the construction of the catalog to whom the void belongs. This correction seems to remove the discrepancies between the void abundances of the different catalogs constructed from the same MultiDark PATCHY galaxy catalog (see Fig. (4.35)). Nevertheless we fitted the Vdn and the

Sheth-van de Weygaert void size function models with hybrid first-crossing distribution function to the corrected catalog abundances, and found a residual cosmology dependence in the Vdn model (see Fig. (4.38)), represented by an increasing behaviour of δ_V and r_s as the value of $\Omega_{\rm m}$ increases.

The study of the void size function of the MultiDark PATCHY galaxy survey arose some unexpected results, such as the smallness of the values of the linear threshold for collapse obtained from the fits, that is worth a deeper study. Another outcome that could be further studied is the unphysical value of the rescaling factor recurrently obtained in the best fits of the Vdn model, that could represent shrinking cosmic voids.

Sheth-van de Weygaert model with Sheth-van de Weygaert first-crossing distribution function presents itself as the best model to fit the data abundances, but the parameters obtained from the fits are outside the regime of validity of the approximation. This result diversify the model from the one proposed, not by changing its expression but by being an excellent fitting model in a regime of parameters where the proposed function should not be valid.

The unexpected outcomes of this work could be related to the fact that, unlike many studies presented in the literature so far, we did not neglect the void-in-cloud effect that dominates the void size function at small radii. As we can notice comparing the results we obtained in the large radii limit to the ones we got considering all voids, the void-in-cloud effect has an important influence on the void size function. In order to trim spurious voids we made the choice of excluding voids with properties analogous to the ones of the voids of the random catalog. The void abundance has been influenced by these cuts, and undoubtedly alternative cuts would lead to different parameters of the models. We chose to consider high-compensation voids, that are more likely embedded in overdensities and might be more related to the void-in-cloud effect. The choice of the selection cuts to apply in order to consider only real voids is not obvious, and the selection done in this work is surely not the only one possible.

It has to be noted that all the models proposed were developed in the dark matter haloes and voids context, while we are considering a galaxy survey. Voids derived from galaxy distributions could be different from dark matter voids, and the theories proposed so far, such as Sheth-van de Weygaert and Vdn models of the void size function, could be inaccurate in this context. This feature could explain the deviation of the fitting parameters obtained from the ones predicted by the theory, including the smallness of the linear threshold for collapse and the fact that we often obtained a rescaling factor that exceeds unity.

5.2 Future development

This work takes a step forward in the understanding of the different models of the void size function and their cosmology dependence in galaxy surveys. However, the analysis here presented can be further extended and developed.

The theories proposed so far in the literature were developed in the study of voids identified from dark matter surveys. The clustering of galaxies do not exactly mirrors the clustering of the bulk of the dark matter distribution, leading Kaiser (1984) to introduce the galaxy bias to indicate that galaxies are biased tracers of the underlying matter density field (Pollina et al., 2017). A possible future improvement of the work could be the inclusion of the galaxy bias in the models, which quantifies the difference between the statistics of galaxies and dark matter haloes. By looking at the distribution of tracers at very large scales we can only observe the most luminous galaxies which are hosted by the most luminous haloes (Kaiser, 1984). Pollina et al. (2017) therefore studied the linear bias which, in terms of spatial correlation, can be written as

$$b = \frac{\xi_{\mathrm{tm}}}{\xi_{\mathrm{mm}}} \tag{5.1}$$

or

$$b = \sqrt{\frac{\xi_{\mathrm{tt}}}{\xi_{\mathrm{mm}}}} \tag{5.2}$$

where ξ_{tm} is the cross-correlation function between tracers (i.e. galaxies) and matter, ξ_{mm} is the matter auto-correlation function and ξ_{tt} is the galaxy auto-correlation function. The inclusion of this parameter in the analysis could lead to different values of the parameters, and hopefully a more accurate model to describe the void size function in galaxy surveys.

To better understand the dependence of the parameters upon the cosmology, the statistics could be implemented, adding some catalogs constructed with values of Ω_{m} in the transform from redshifts to distances closer to the true value used in the generation of the mock galaxy catalog. Increasing the number of catalogs could lead to a clearer trend in the behaviour of the parameters, and could allow a more detailed discussion of the subject.

Another possible future development could focus on the multivariate analysis and its choices, which allowed us to treat the void-in-cloud term of the void size function. The choices made in this analysis are certainly not the sole ones, and a deeper focus on this matter might open the possibility of an exhaustive study of the void size function models, including and modelling also the void-in-cloud effect.

5.3 Conclusion

In this thesis work we have studied the model and cosmology dependence of the void size function in galaxy surveys. In order to do so we constructed different catalogs from the MultiDark PATCHY mock galaxy catalog, using different values of the matter density parameter in the transform from redshifts to distances. In order to trim spurious voids identified by the void finder due to shot noise, we have performed a multivariate analysis, comparing the properties of the catalogs with the void properties of a void catalog constructed from the random PATCHY sample. We chose two selection cuts, on the core density and the compensation of voids. In this way we have been able to consider also the void-in-cloud term, dominant at small radii where the presence of shot noise is otherwise too important to allow an accurate analysis.

Using an MCMC method we fitted to the void abundances of the catalogs Sheth-van de Weygaert and Vdn models of the void size function, with the general and the Sheth-van de Weygaert approximation forms of the first-crossing distribution function. The fits has been performed using as free parameters the linear threshold for void formation, the linear threshold for collapse and the rescaling factor, without imposing any strong assumption on the voids. We found a coherent fit for the two Vdn models, with fitting parameters in the regime of validity of Sheth-van de Weygaert approximation. The best fit to the data corresponds to the Sheth-van de Weygaert model with Sheth-van de Weygaert first-crossing distribution function outside the regime of validity of the approximation, while the Sheth-van de Weygaert model with general form of the first-crossing distribution function shows a poor agreement with the catalogs abundances. Due to the ill-convergence of the series function in the expression of the general form of the first-crossing distribution function, we found best to use the hybrid function proposed by Jennings et al. (2013) as first-crossing distribution function. The fits of the Vdn model are again coherent with the ones obtained in the other two Vdn models, while Sheth-van de Weygaert model with hybrid first-crossing distribution function presents fitting parameters far from the values of both the other Sheth-van de Weygaert models studied. From the behaviour of the fitting parameters as functions of the matter density parameter Ω_{m} we studied the cosmology dependence of the void size function. All Vdn void size function models present an anomaly in the trend of their parameters at the value of Ω_{m} corresponding to the one used in the construction of the MultiDark PATCHY catalog, suggesting a particularity of that value with respect to the others. Sheth-van de Weygaert model does not share this trend, demonstrating once again the strong dependence of the parameters and their behaviour upon the chosen model.

Following Correa et al. (2020) we applied an Alcock-Paczyński correction to the radius of each void, and the comparison of the void abundances of the catalogs after the correction confirms its validity. However, a residual cosmology dependence of the fitting parameters is still visible in the case of the Vdn void size function with hybrid first-crossing distribution function.

This work analyzed the different models of the void size function proposed in the literature without neglecting the void-in-cloud effect, and via the parameters obtained from the fits to the catalogs constructed with different values of the matter density parameter in the transform from redshifts to distances, it studied the dependence of the different models upon the chosen cosmology.

Bibliography

Alam S., Albareti F. D., Allende P. et al (2015), *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*, The Astrophysical Journal Supplement Series, 219:12 (27 pp.).

Alcock C., Paczyński B. (1979), *An evolution free test for non-zero cosmological constant*, Nature 281: 358-359.

Baker T., Clampitt J., Jain B., Trodden M. (2018), *Void lensing as a test of gravity*, Physical Review D, 98, 023511.

Berind A. A., Weinberg D. H. (2002), *The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass*, The Astrophysical Journal, 575:587-616.

Bernardeau F. (1994), *The Nonlinear Evolution of Rare Events*, Astrophysical Journal, 427:51.

Bickel P. J., Doksum K. A. (1977), *Mathematical Statistic, Basic Ideas and Selected Topics, Volume I*, Prentice Hall.

Blumenthal G. R., Nicolaci da Costa L., Goldwirth D. S., Lecar M., Piran T. (1992), *The Large Possible Voids*, The Astrophysical Journal, 388:234-241,1992

Bond J. R., Cole S., Efstathiou G., Kaiser N. (1991), *Excursion set mass functionsfor hierarchical gaussian fluctuations*, The Astrophysical Journal 390: 440-460.

Bond J. R., Myers S. T. (1996), *The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms*, Astrophysical Journal Supplement, 103:1.

Bond J. R., Kofman L., Pogosyan D. (1996), *How filaments of galaxies are woven into the cosmic web*, Nature, 380:603-606.

Bos E. G. P., van de Weygaert R., Dolag K., Pettorino V. (2012), *The darkness that shaped the void: dark energy and cosmic voids*, Mon. Not. R. astr. Soc., 426:440-461.

Carroll B. W. & Ostlie D. A. (2007), *An Introduction to Modern Astrophysics*, 2nd Ed., Pearson Education.

Casella G., Berger R. L. (1990), *Statistical Inference*, Duxbury Advance Series.

Chan K. C., Hamaus N., Desjacques V. (2014). *Large-Scale Clustering of Cosmic Voids*, Phys. Rev. D 90, 103521.

Coles P., Lucchin F. (1995) *Cosmology-The Origin and Evolution of Cosmic Structure*, John Wiley & Sons, Ltd.

Correa C. M., Paz D. J., Sánchez A. G., Ruiz A. N., Padilla N. D., Angulo R. E. (2020), *Redshift-space effects in voids and their impact on cosmological tests. Part I: the void size function*, eprint arXiv:2007.12064.

Cousinou M. C., Pisani A., Tilquin A., Hamanus N., Hawken A. J., Escoffier S. (2019), *Multivariate analysis of cosmic void characteristics*, Astronomy and Computing, 27:53.

Dawson K. S., Schlegel D. J., Ahn C. P. at al. (2013), *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, The Astrophysical Journal, 145:10 (41 pp.).

Dubey A., Reid L. B., Fisher R. (2008), *Introduction to FLASH 3.0, with application to supersonic turbulence*, Physica Scripta, 132, 014046.

Epstein R. I.(1983), *Proto-galactic perturbations*, Mon. Not. R. astr. Soc. 205:207-229.

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J. (2013), *emcee: The MCMC Hammer*, Publications of the Astronomical Society of the Pacific,125:306.

Goodman J., Weare J. (2010), *Ensemble samplers with affine invariance*, Comm. App. Math. Comp. Sci., 5:65.

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M. (2005), *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, The Astrophysical Journal, 622:759-771.

Hamaus N., Sutter P. M., Wandelt B. D. (2014), *Universal Density Profile for Cosmic Voids*, Physical Review Letters, 112:25,251302.

Hamaus N., Sutter P. M., Lavaux G., Wandelt B. D. (2015), *Probing cosmology and gravity with redshift-space distortions around voids*, Journal of Cosmology and Astroparticle Physics, 2015:11,036.

Hamaus N., Pisani A., Sutter P. M., Lavaux G., Escoffier S., Wandelt B. D., Weller J. (2016), *Constraints on Cosmology and Gravity from the Dynamics of Voids*, Phys. Rev. Lett., 117, 091302.

Hamaus N. (2017) *Cosmic Voids*, Lecture notes for the *Lecture Series on Cosmology.*

Icke V. (1984), *Voids and filaments*, Mon. Not. R. astr. Soc. 206, Short Communication, 1P-3P.

Jennings E., Li Y., Hu W. (2013), *The abundance of voids and the excursion set formalism*, Mon. Not. Roy. Astron. Soc., 434:2167-2181.

Kaiser N. (1984), *On the spatial correlations of Abell clusters*, The Astrophysical Journal, 284:L9-L12.

Kitaura F.-S., Rodríguez-Torres S., Chuang C., Zhao C., Prada F., Gil-Marín H., Guo H., Yepes G., Klypin A., Scóccola C. G., Tinker J., McBride C., Reid B., Sánchez A. G., Salazar-Albornoz S., Grieb J. N., Vargas-Magana M., Cuesta A. J., Neyrinck M., Beutler F., Comparat J., Percival W. J., Ross A. (2016), *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release*, Mon. Not. Roy. Astron. Soc., 456:4156-4173.

Kyplin A., Holzman J. (1997), *Particle-Mesh code for cosmological simulations*, ArXiv Astrophysics e-prints.

Kyplin A., Yepes G., Gottlöber S., Prada F., Heß S. (2016), *MultiDark simulations: the story of dark matter halo concentrations and density profiles.*, Mon. Not. Roy. Astron. Soc., 457:4340-4359.

Lavaux G., Wandelt B. D. (2012), *Precision Cosmography with Stacked Voids*, The Astrophysical Journal, 754:109 (15 pp.).

Mao Q., Berlind A. A., Scherrer R. J., Neyrinck M. C., Scoccimarro R., Tinker J. L., McBride C. K. (2017), *Cosmic Voids in the SDSS DR12 BOSS Galaxy Sample: The Alcock-Paczynski test*, The Astrophysical Journal, 835:160 (10 pp.).

Massara E., Villaescusa-Navarro F., Vie M., Sutter P. M. (2015), *Voids in massive neutrino cosmologies*, Journal of Cosmology and Astroparticle Physics, 2015:11,018.

Miller S. J. (2006), *The Method of Least Squares*, Mathematics Department Brown University, Providence.

Mo H., van den Bosch F., White S. (2010), *Galaxy Formation and Evolution*, Cambridge University Press.

Nadathur S., Hotchkiss S. (2015), *The nature of voids – II. Tracing underdensities with biased galaxies*, Mon. Not. Roy. Astron. Soc., 454:889.

Neyrinck M. C., Gnedin N. Y., Hamilton A. J. S. (2005), *VOBOZ: an almost-parameter-free halo-finding algorithm*, Mon. Not. Roy. Astron. Soc., 356:1222-1232.

Neyrinck M. C. (2008), *ZOBOV: a parameter-free void-finding algorithm*, Mon. Not. Roy. Astron. Soc., 386:2101-2109.

Nuza S. E., Sánchez A., Prada F., Klypin A., Schlegel D. J., Gottlöber S., Montener-Dorta A. D., Manera M., McBride C. K., Ross A. J., Angulo R., Blanton M., Bolton

A., Favole G., Samushia L., Montesano F., Percival W. J., Padmanabhan N., Steinmetz M., Tinker J., Skibba R., Schneider D. P., Guo H., Zehavi I., Zheng Z., Bizyaev D., Malanushenko O., Malanushenko V., Oravetz A. E., Oravetz D. J., Shelden A. C. (2013), *The clustering of galaxies at* $z \approx 0.5$ *in the SDSS-III Data Release 9 BOSS-CMASS sample: a test for the* Λ*CDM cosmology*, Mon. Not. Roy. Astron. Soc., 432:743-760.

Peebles P. J. E. (1980). *The Large-Scale Structure of the Universe*, Princeton University Press.

Pisani A., Massara E., Spergel D. N., Alonso D., Baker T., Cai Y.-C., Cautun M., Davies C., Demchenko V., Doré O., Goulding A., Habouzit M., Hamaus N., Hawken A., Hirata C. M., Ho S., Jain B., Kreisch C. D., Marulli F., Padilla N., Pollina G., Sahlén M., Sheth R. K., Somerville R., Szapudi I., van de Weygaert R., Villaescusa-Navarro F., Wandelt B. D., Wang Y. (2019), *Cosmic voids: a novel probe to shed light on our Universe*, Bulletin of the American Astronomical Society, 51:40.

Platen E., van de Weygaert R., Bernard J. T. (2007), *A cosmic watershed: the WVF void detection technique*, Mon. Not. Roy. Astron. Soc., 380:551-570.

Pollina G., Hamaus N., Dolag K, Weller J., Baldi M., Moscardini L. (2017), *On the linearity of tracer bias around voids*, Mon. Not. Roy. Astron. Soc., 469:787-799.

Press W. H., Schechter P. (1974), *Formation of galaxies and clusters of galaxies by self-similar gravitational condensation*, The Astrophysical Journal 187:425-438.

Rodríguez-Torres S. A., Chuang C., Prada F., Guo H., Klypin A., Behroozi P., Hahn C. H., Comparat J., Yepes G., Montero-Dorta A. D., Brownstein J. R., Maraston C., McBride C. K., Tinker J., Gottlober S., Favole G., Shu Y., Kitaura F.-S., Bolton A., Scoccimarro R., Samushia L., Schlegel D., Schneider D. P., Thomas D. (2016), *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release*, Mon. Not. Roy. Astron. Soc., 460:1173-1187.

Russell S. J., Norvig P. (1995), *Artificial Intelligence, A Modern Approach*, Prentice Hall.

Schaap W. E. (2007), *DTFE: the Delaunay Tessellation Field Estimator*, Ph.D. thesis, University of Gröningen.

Sheth R. K., van de Weygaert R. (2004), *A hierarchy of voids: Much ado about nothing*, Mon. Not. Roy. Astron. Soc-. 350:517-538.

Speagle J. S. (2019), *A Conceptual Introduction to Markov Chain Monte Carlo Methods*, eprint arXiv:1909.12313.

Springel V. (2005), *The cosmological simulation code GADGET-2*, Mon. Not. Roy. Astron. Soc., 364:1105-1134.

Sutter P. M., Lavaux G.,Wandelt B. D., Weinberg D. H. (2012), *A Public Void Catalog from the SDSS DR7 Galaxy Redshift Surveys Based on the Watershed Transform*, The Astrophysical Journal, 761:44 (13 pp.).

Sutter P. M., Lavaux G., Hamaus N., Pisani A., Wandelt B. D., Warren M. S., Villaescusa-Navarro F., Zivick P., Mao Q., Thompson B. B. (2015), *VIDE:The Void IDentification and Examination toolkit*, Astronomy and Computing, 9:1-9.

Teyssier R. (2002), *Cosmological Hydrodynamics with Adaptive Mesh Refinement: a new high resolution code called RAMSES*, Astronomy and Astrophysics, 385:337-364.

van de Weygaert R., Platen E. (2011), *Cosmic Voids: Structure, Dynamics and Galaxies*, International Journal of Modern Physics: Conference Series, 1:41-66.

Zhao C., Kitaura F.-S., Chuang C.-H., Prada F., Yepes G., Tao C. (2015), *Halo mass distribution reconstruction across the cosmic web*, Mon. Not. Roy. Astron. Soc., 451:4266-4276.