UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Physics and Astronomy "Galileo Galilei"

Ph.D. COURSE IN: PHYSICS

COSMIC VOID STATISTICS AS A PROBE FOR COSMOLOGY

Thesis written with the financial contribution of Università degli Studi di Padova

Coordinator: Prof. Franco Simonetto

Supervisor: Prof. Chiara Sirignano

Co-Supervisor: Dr. Alessandro Renzi

Ph.D. student: Giovanni Alberto Verza

Introduction

We are experiencing a period of transition in cosmology. In the last 25 years, several cosmological experiments probing the cosmic microwave background (CMB) and the large-scale structure (LSS) of the Universe have led to the establishment of the standard model of cosmology, the ΛCDM model. This model fits with high accuracy a broad variety of cosmological data and observables. It describes the Universe as a flat and expanding space-time governed by the laws of General Relativity, the evolution of which is dictated by different constituents: the ordinary baryonic matter ($\sim 5\%$), the so-called cold dark matter ($\sim 27\%$), and the cosmological constant Λ ($\sim 68\%$). The ΛCDM model describes a Universe that is unknown at 95% in the nature of its constituents; as a matter of fact, modern cosmology is facing the exciting challenge of understanding the dark side of the Universe.

In spite of the preponderant dark sector that the ΛCDM model introduces to describe the evolution of our Universe, it provides an accurate description of almost all cosmological measurements. However, in recent years, the ΛCDM model started to show some tensions. Data analysis performed on the CMB and on the late Universe, are well fitted by ΛCDM, nevertheless, the derived value of some parameters is not in agreement between the two observables. In particular, there is a tension concerning the exact value of the expansion rate of the Universe today, i.e. the Hubble constant H_0, known as the H_0 tension. Measurements coming from the type Ia supernovae as standard candles infer a higher value of the Hubble constant with respect to the CMB analysis. At the same time, the estimation of the total matter density and the amplitude normalization of matter fluctuation inferred from weak lensing measurements of the late Universe do not agree with analysis of the CMB data assuming ΛCDM.

Beyond the tensions, modern cosmology is facing the so called dark energy problem. The cosmological constant Λ was introduced to describe the observed late-time acceleration of the Universe, and even if this property of the space-time is in good agreement with the data, its existence and especially its magnitude cannot be explained in a natural way. The discovery of the late-time acceleration is the greatest breakthrough in cosmology. Currently, the major effort of the cosmological scientific community is being spent to study its physical origin, methods to measure and characterize this accelerated expansion, and to develop experiments to investigate its effects.

In the next decades, an astonishing amount of data with unprecedented accuracy coming from several experiments will be available, possibly unveiling the nature of its dark components and new physics. In particular, measurements of CMB temperature and polarization, gravitational wave observatories, underground dark matter and neutrino detectors, and large-scale galaxy surveys are being conducted soon. One of the most relevant upcoming galaxy surveys is Euclid, a mission of the European Space Agency. Euclid will be one of the widest spectroscopic and photometric galaxy surveys, covering one-third of the sky area. It will measure galaxy spectra up to redshift $z \sim 2$, mapping in detail significant contiguous fractions

of the observable Universe over an unprecedented large cosmic volume. Moreover, Euclid will probe in detail the epoch at which the Universe started its accelerated expansion, possibly shedding new light on the dark energy problem. The two main cosmological probes that will be explored in the analysis of Euclid data are weak gravitational lensing and the baryonic acoustic oscillation of galaxy clustering. The joint analysis of these two cosmological probes will put unprecedented constraints on cosmological parameters.

Given the precision of upcoming cosmological analysis, the constraining power of the standard cosmological probes is close to reaching its maximum. For this reason, a way to go further in understanding the physics acting in our Universe is to look for new independent cosmological probes. The combination of different probes can help to reduce the uncertainty on cosmological parameters; moreover, the diversity between different methods may open new observational and theoretical windows. Among the new emerging cosmological probes, in this work we focus on cosmic voids that are particularly significant at the dawn of wide and deep spectroscopic galaxy surveys, such as Euclid. Cosmic voids are large underdense regions in the large-scale structure of the Universe, they span a large range of scales and constitute the largest observable objects in the Universe. Their size and underdense peculiarity make them a particularly suited probe in investigating the dark energy problem. Moreover, their sizes require surveys characterized by a very large volume to have significant statistics. To measure and model various cosmic void properties, a detailed map of galaxy positions with good spatial resolution is required. With these resources demanded, cosmic voids find in upcoming galaxy surveys the environment in which all their power can be released. The Euclid mission is the ideal galaxy survey for probing cosmology with cosmic voids. These cosmic void features and their requirements from the observational point of view make the cosmological exploitation of cosmic voids just at its beginning. Moreover, classical LSS clustering analyses focus on overdensities and collapsed regions, whereas cosmic voids probe underdensities, so they may provide a complementary point of view in studying the Universe, possibly shedding new light on the open questions in cosmology and physics, motivating effort in theoretical studies and in characterization for cosmological analysis.

In this thesis, we focus on the study of cosmic voids in the three-dimensional matter/galaxy distribution considering their statistical properties and their cosmological exploitation in spectroscopic galaxy surveys. We will provide an unprecedented theoretical treatment, a detailed study and modelization of their properties measured in N-body cosmological simulations, and we forecast the constraining power of cosmic voids in the Euclid survey.

The thesis is organized as follows. In Chapter 1 we introduce the cosmological framework in which the following research is developed. We review the dynamics and the properties of the background homogeneous Universe, and the matter-energy components driving its evolution; then we consider how to treat the evolution of inhomogeneities in the context of LSS, the formation of dark matter halos and galaxies from the initial perturbed density field, and their biased statistical properties with respect to matter fluctuations. We conclude by describing the standard cosmological model, its immediate extensions and criticalities, representing the starting point of the current research in cosmology.

In Chapter 2 we introduce our original work aimed at theoretically modeling the statistical properties of voids, halos, and their cross-correlations in the excursion-set framework. We consider how voids and halos form and how their formation can be implemented within the excursion-set framework, we introduce the equations describing proto-halos and proto-voids formation in the initial density field and their spatial correlations, we show numerical method-

ologies to solve them, and we present our original results.

In Chapter 3 we present our original analysis on the properties of cosmic voids measured in large N-body cosmological simulations. This work aims to explore some of the observable properties and statistics of cosmic voids as cosmological probes, investigating possible modelizations and evaluating their sensitivity to the dark energy equation of state and to the total neutrino mass. In particular, we introduce the N-body simulations and the void finding algorithm used for this study, then we focus on the abundance of voids and on the void density profile.

In Chapter 4 we present for the first time in the literature the forecast on cosmological parameters provided by the void abundance as a standalone probe in the Euclid spectroscopic survey. We introduce the Euclid mission, the simulations used, we widely discuss the methodology adopted, and we present our results.

Contents

Chapter 1

The cosmological framework

Cosmology is a branch of physics whose aim is the comprehension of the Universe as a whole. Modern cosmology relies on General Relativity, in which the evolution of space-time is related to the matter-energy component embedded in it. A description of the whole Universe at any scale is impossible from a practical point of view, so cosmology focuses on scales much larger than the typical galaxy size, usually considering scales $\gtrsim 1\,\mathrm{Mpc} = 3.086 \times 10^{22}$ m. Experiments in cosmology can be divided into two broad classes, Cosmic Microwave Background (CMB) experiments, i.e. the study of the light of the primordial plasma that traveled through the Universe toward us, and Large-Scale Structure (LSS) experiments, i.e. the study of the distribution of tracers, usually galaxies. This work focuses on cosmic voids as tracers of the large-scale structure of the Universe, so even if the current Standard Cosmological Model strongly depends on the progress performed in studying the CMB, we mainly focus on LSS.

In this chapter, we introduce the formalism and tools needed to describe the large-scale structure of the Universe and the research presented in the following chapters. In Sect. 1.1 we introduce the background dynamics of the Universe and the matter-energy components; in Sect. 1.2 the cosmological perturbation theory used to describe structure formation in the late Universe; in Sect. 1.3 we introduce the formalism to describe the relation between the statistical properties of collapsed object and of the underlying matter distribution; in Sect. 1.4 we discuss the status of the Standard Cosmological Model.

1.1 The smooth expanding Universe

1.1.1 The FLRW metric

Observations and ideas concerning cosmological physics can be formalized with General Relativity, which is, to date, the fiducial theory of gravity and space-time. In the context of General Relativity, space-time is a manifold on which we define a Lorentz four-dimensional metric $g_{\mu\nu}$. The invariant space-time length element is

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \,, \tag{1.1}$$

where the indices μ and ν range from 0 to 3. The first index is reserved for the time-like coordinate, $\mathrm{d}x^0 = \mathrm{d}t$, where we have considered natural units $c = 1$, and the last three for spatial coordinates. We will work with the metric signature $(-+++)$ so that a 4-vector v is time-like if $v_\mu v^\mu < 0$ and space-like if $v_\mu v^\mu > 0$. In a Lorentz frame, i.e. $g_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$, the scalar

product between two 4-vectors v and w thus is $(v, w) = g_{\mu\nu}v^\mu w^\nu = -v^0w^0 + v^1w^1 + v^2w^2 + v^3w^3$. Note also that the 4-velocity u of a (massive) particle satisfies $-(u^0)^2 + |\mathbf{u}|^2 = -1$ [1]. From the metric we can derive all the properties describing the space-time, in particular, the Christoffel symbols,

$$\Gamma^\lambda_{\ \mu\nu} = \frac{1}{2}g^{\lambda\sigma}\left(\partial_\nu g_{\sigma\mu} + \partial_\mu g_{\nu\sigma} - \partial_\sigma g_{\mu\nu}\right), \tag{1.2}$$

the Riemann tensor $R^\sigma_{\ \lambda\mu\nu}$, the Ricci tensor $R_{\mu\nu}$ and the scalar curvature R obtained by contractions of the Riemann tensor [2],

$$\begin{aligned} R^\sigma_{\ \lambda\mu\nu} &= \partial_\mu\Gamma^\sigma_{\ \lambda\nu} - \partial_\nu\Gamma^\sigma_{\ \lambda\mu} + \Gamma^\sigma_{\ \eta\mu}\Gamma^\eta_{\ \lambda\nu} - \Gamma^\sigma_{\ \eta\nu}\Gamma^\eta_{\ \lambda\mu} \\ R_{\mu\nu} &= R^\lambda_{\ \lambda\mu\nu} \\ R &= g^{\mu\nu}R_{\mu\nu}\,. \end{aligned} \tag{1.3}$$

The Standard Cosmological Model is built on the assumption of homogeneous and isotropic spatial sections [3]. A homogeneous space is one that is translation invariant, or the same at every point. An isotropic space is one that is rotationally invariant, or the same in every direction. A space which is everywhere isotropic is necessarily homogeneous, but the converse is not true. These assumptions are not verified on small scales; nevertheless observations of the CMB, counts of radio galaxies, galaxy distribution, and other experiments provide empirical evidence that the Universe is homogeneous on scales above ~ 100 Mpc [4]. More precisely, in a space-time with a homogeneous and isotropic space part, there is a preferred geodesic time coordinate t, called "cosmic time", such that the 3-space slices of constant t are maximally symmetric spaces or, equivalently, spaces of constant curvature. The maximally symmetric 3-space is characterized by a number $n(n+1)/2$ of isometries, represented in the three-dimensional space ($n = 3$), by 3 spatial translations and 3 spatial rotations, which, respectively, describe homogeneity and isotropy. The maximally symmetric 3-spaces have a constant curvature, since a position dependent curvature would break the isotropy and homogeneity. Therefore, the metric is of the form

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = -\mathrm{d}t^2 + a^2(t)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j\,, \tag{1.4}$$

where the Latin indices run over the spatial dimensions. The scale factor $a(t)$ is a function of time, it is not fixed by symmetries that refer to the spatial part only. The geometric properties of the spatial part depend on the constant curvature K, obtained from [3]

$$^{(3)}R_{ijkl} = K(\gamma_{ij}\gamma_{kl} - \gamma_{ik}\gamma_{jl}) \qquad \Rightarrow \qquad ^{(3)}R_{ij} = 2K\gamma_{ij}\,, \tag{1.5}$$

where $^{(3)}R_{ijkl}$ is the 3-space Riemann tensor and $^{(3)}R_{ij}$ is its contraction, i.e. the Ricci tensor. It is useful to express the spatial part in spherical coordinates,

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[f_K(r)\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right], \tag{1.6}$$

where $f_K(r)$ is a function of the radial coordinate r depending on the curvature. Calculating the Ricci tensor of this metric and comparing it with Eq. (1.5), we find that $f_K(r) = 1/(1 - Kr^2)$, therefore,

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1 - Kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right]. \tag{1.7}$$

This is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. The curvature K defines the geometrical properties of the 3-space isotropic and homogeneous metric, which can be the

Euclidean space ($K = 0$), the hypersphere ($K > 0$) and the hyperbolic space ($K < 0$). The curvature of the 3-spatial dimensions is used to classify the structure of FLRW universes, i.e. the Universe is spatially flat if a slice of constant t is isometric to the Euclidean space, closed if isometric to the hypersphere and open if isometric to the hyperbolic space. The geometrical properties of flat, close and open space-time become more explicit rewriting the spatial part of the metric as

$$\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j = \mathrm{d}\chi^2 + \chi^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\,, \tag{1.8}$$

where χ is the radial coordinate measuring the length at constant time, and satisfies

$$r = \begin{cases} K^{-1/2}\sin\left[\sqrt{K}\chi\right] & K > 0 \\ \chi & K = 0 \\ K^{-1/2}\sinh\left[\sqrt{-K}\chi\right] & K < 0\,. \end{cases} \tag{1.9}$$

It is often useful to consider the conformal FLRW metric. Two metric $g_{\mu\nu}$ and $\tilde{g}_{\mu\nu}$ are linked by a conformal transformation if they can be written as

$$\tilde{g}_{\mu\nu} = \Omega(x)g_{\mu\nu} \qquad \text{such that} \qquad \Omega(x) > 0\,. \tag{1.10}$$

A conformal transformation locally modifies spatial lengths and times, nevertheless, the angles and the causal structure are preserved. A simple conformal transformation can be performed by factorizing the scale factor in Eq. (1.8),

$$\mathrm{d}s^2 = a^2(t)\left[-\mathrm{d}\eta^2 + \mathrm{d}\chi^2 + \chi^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right] = a^2(t)\mathrm{d}\tilde{s}^2\,, \tag{1.11}$$

where η is the conformal time defined as

$$\eta(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')}\,. \tag{1.12}$$

In the conformal metric $\mathrm{d}\tilde{s}^2$ the spatial part is no longer multiplied by the scale factor $a(t)$, these space coordinates are called comoving coordinates. We recover the physical coordinates by multiplying the comoving coordinates by the scale factor $a(t)$. We will see later that the scale factor can be fixed up to a normalization, so is usually set $a(t_0) = 1$, where t_0 corresponds to the today epoch, in this way comoving and physical coordinates and lengths coincide in the present time.

1.1.2 Particle propagation and cosmological redshift

Once we have a metric, we can consider the propagation of particles in it, which in the absence of external forces follow the geodesic equations [3]. The geodesic equation can be easily derived by minimizing the Lagrangian that describes the system.

Let us consider a photon propagating radially as seen by an observer, so the relevant part of the metric Eq. (1.8) is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\chi^2 = 0\,, \tag{1.13}$$

where $\mathrm{d}s^2 = 0$ is because photons follow null geodesics. This is equivalent to say that the propagation of this photon is described by the Lagrangian

$$\mathcal{L} = \frac{1}{2}g_{\mu\nu}\dot{x}_\mu\dot{x}_\nu\,, \tag{1.14}$$

where the dots denote a derivative with respect to an affine parameter λ. Therefore, the Euler-Lagrange equation for χ reads

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial \mathcal{L}}{\partial \dot{\chi}} - \frac{\partial \mathcal{L}}{\partial \chi} = \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial \mathcal{L}}{\partial \dot{\chi}} = 0\,, \tag{1.15}$$

where we use the fact that $\mathcal{L}$ does not depend explicitly on χ, so $\partial\mathcal{L}/\partial\chi = 0$. From the above equations we can obtain the covariant momentum

$$p_\mu = \frac{\mathcal{L}}{\partial \dot{x}_\mu} \qquad \Rightarrow \qquad \begin{aligned} p_0 &= \frac{\mathcal{L}}{\partial \dot{t}} = -\dot{t} \\ p_\chi &= \frac{\mathcal{L}}{\partial \dot{\chi}} = a^2\dot{\chi}\,, \end{aligned} \tag{1.16}$$

that is conserved along the photon trajectory. The energy observed by the observer is

$$E_\mathrm{o} = -p_\mu u^\mu\,, \tag{1.17}$$

where $u^\mu = (-1, 0, 0, 0)$ is the 4-velocity of the comoving observer and $p_\mu = (p_0, p_\chi, 0, 0)$ is the photon covariant momentum. Moreover, from special relativity, $E_\mathrm{o} - p_\mathrm{o} = 0$, so

$$E_\mathrm{o} = p_\mathrm{o} = -p_0 u^0 = \dot{t} = a(t)\dot{\chi} = \frac{p_\chi}{a(t)}\,. \tag{1.18}$$

This equation shows that the observed photon energy E_o and the observed photon momentum p_o vary inversely with the scale factor $a(t)$. Since the photon energy is proportional to the inverse of the wavelength, $E_\mathrm{o} \propto \lambda^{-1}$, it follows that $\lambda \propto a(t)$. Let us consider E_e and p_e the energy and momentum of the photon observed by a comoving observer at some initial time, which can be the emitted time. It follows that

$$\frac{E_\mathrm{o}}{E_\mathrm{e}} = \frac{p_\mathrm{o}}{p_\mathrm{e}} = \frac{\lambda_\mathrm{e}}{\lambda_\mathrm{o}} = \frac{a(t_\mathrm{e})}{a(t_\mathrm{o})}\,. \tag{1.19}$$

The observed wavelength appears to be shifted by a factor

$$z = \frac{\lambda_\mathrm{o} - \lambda_\mathrm{e}}{\lambda_\mathrm{e}} = \frac{a(t_\mathrm{o}) - a(t_\mathrm{e})}{a(t_\mathrm{e})}\,. \tag{1.20}$$

If we consider that the observed time is the present, $t_\mathrm{o} = t_0$, it follows that $a(t_\mathrm{o}) = a(t_0) = 1$ and

$$z = \frac{1}{a} - 1 \qquad \Longleftrightarrow \qquad a = \frac{1}{1+z}\,, \tag{1.21}$$

where z is called redshift. As the Universe expands, the wavelength of a freely propagating photon increases as physical distances increase with the expansion.

Eq. (1.20) describes how the redshift is actually measured. For example, in the observed spectra of distant galaxies, we can recognize the emitting or absorbing lines of some components. The most relevant is the Hα line, corresponding to a wavelength of $\lambda_{\mathrm{H}\alpha} = 656.281$ nm. The observed wavelength would be larger due to the expansion and Eq. (1.20) provides us the corresponding redshift. Note that if the light we observe is emitted by comoving observers, the redshift is in a one-to-one map with the scale factor, epoch of the emission, time of the emission, and the distance of the emitter, as we will see later.

1.1.3 Dynamics of the homogeneous and isotropic Universe

Up to now, we have not considered General Relativity, since the properties of the FLRW metric can be inferred from spatial symmetries only. However, General Relativity is necessary to describe the dynamics of the Universe, described by the Einstein field equations,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \,, \tag{1.22}$$

where G is the Newtonian constant, $G_{\mu\nu}$ is the Einstein tensor, Λ is the cosmological constant, and $T_{\mu\nu}$ is the energy-momentum tensor. The left-hand side of this equation, i.e. the Einstein tensor, depends on the metric only and describes the space-time curvature and its evolution. The right-hand side describes the matter and energy components that source the space-time curvature and their evolution, sourced by the space-time curvature.

The energy-momentum tensor can be written in the generic form

$$T_{\mu\nu} = \rho u_\mu u_\nu + P(g_{\mu\nu} + u_\mu u_\nu) + \pi_{\mu\nu}, \tag{1.23}$$

where u^μ is the fluid-4 velocity with respect to the observer, ρ is the energy density, including both the rest mass and the internal energy, while P is the pressure, or isotropic stress. Note that both ρ and P are defined in the fluid rest frame. The $\pi_{\mu\nu}$ term is the shear stress or anisotropic stress, that is symmetric $\pi_{\mu\nu} = \pi_{\nu\mu}$, transverse, $u^\mu \pi_{\mu\nu} = 0$, and traceless, $\pi^\mu{}_\mu = 0$.

The isotropy and homogeneity of the Universe provide some constraints on the stress-energy tensor. Since there is no preferred direction, the fluid 4-velocity is $u_\mu = (-1, 0, 0, 0)$. This means that the fluid is at rest with the comoving observers. Then the isotropy condition requires that the anisotropic stress must be null. It follows that the general form of the stress-energy tensor with the FLRW symmetries has components

$$T_{00} = \rho, \qquad T_{0i} = T_{0i} = 0, \qquad T_{ij} = P a^2 \gamma_{ij} \,. \tag{1.24}$$

The T_{00} component is the total energy density, T_{0i} is the i-component of the energy flux, T_{i0} is the 3-momentum density, and T_{ij} is the flux of the fluid momentum p_i through a surface oriented in the j-direction.

To solve the Einstein equation, we compute the Einstein tensor $G_{\mu\nu}$ from the FLRW metric. The Einstein tensor is a symmetric two-index tensor, so it has *a priori* 10 degrees of freedom, nevertheless the symmetries involved greatly simplify the task. In particular, the components of the Ricci tensor and the scalar curvature are

$$R_{00} = 3\frac{\ddot{a}}{a}, \qquad R_{i0} = 0, \qquad R_{ij} = \left(\frac{\ddot{a}}{a} + 2\frac{K + \dot{a}^2}{a^2}\right) a^2 \delta_{ij}, \qquad R = 6\left(\frac{\ddot{a}}{a} + 2\frac{K + \dot{a}^2}{a^2}\right). \tag{1.25}$$

Substituting these quantities in Eq. (1.22) remains only two independent equations,

$$H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{K}{a^2} \,, \tag{1.26}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3} \,, \tag{1.27}$$

that are the first and second Friedmann equations, respectively. In these equations, we introduced the Hubble factor $H = \dot{a}/a$, which describes the expansion rate of the Universe at a given epoch.

Let us now consider the continuity equation $\nabla_\mu T^\mu_{\ \nu} = 0$. The spatial components identically vanish, $\nabla_\mu T^{\mu i} = 0$, due to the metricity of the connection in General Relativity, i.e. $\nabla_\mu g^\mu_{\ \nu} = 0$ [2, 3]. The time component is

$$\nabla_\mu T^{\mu 0} = \partial_\mu T^{\mu 0} + \Gamma^\mu_{\ \mu\nu} T^{\nu 0} + \Gamma^0_{\ \mu\nu} T^{\mu\nu} = 0\,, \tag{1.28}$$

where substituting Eq. (1.24) we obtain

$$\dot{\rho} + \Gamma^\mu_{\ \mu 0}\rho + \Gamma^0_{\ 00}\rho + \Gamma^0_{\ ij} T^{ij} = 0\,. \tag{1.29}$$

The only non-vanishing Christoffel symbol is $\Gamma^0_{\ ij} = H g_{ij}$, from which

$$\dot{\rho} = -3H(\rho + P)\,, \tag{1.30}$$

describing the evolution of the energy density. Note that this equation is not independent from Eq. (1.26) and Eq. (1.27) and therefore can also be obtained by combining them. The terms can be rearranged in order to express it in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3(t)\rho(t)\right] = P(t)\frac{\mathrm{d}a^3(t)}{\mathrm{d}t}\,, \tag{1.31}$$

that reads as the adiabatic equation for a perfect fluid: the change in the "internal energy" of a volume element equals the pressure times the change in the proper volume. This is the first law of thermodynamics in the cosmological context.

To describe the evolution of the energy density of the components of the Universe, we have to know the corresponding equation of state

$$P = P(\rho)\,, \tag{1.32}$$

allowing to solve both the equation of the evolution of the energy density Eq. (1.30) and the Friedmann equations Eqs. (1.26) and (1.27). Most of the fluids involved in cosmology can be described by a linear equation of state,

$$P = w\rho\,, \tag{1.33}$$

so that

$$\rho(t) = \rho_0 a^{-3} \exp\left[-3\int_1^a \frac{w(a')}{a'}\mathrm{d}a'\right]. \tag{1.34}$$

Moreover, most of the components of the Universe have a constant equation of state, i.e. $w(t) =$ const, so

$$\rho(t) = \rho_0 a^{-3(w+1)}\,. \tag{1.35}$$

1.1.4 The constituents of the Universe

To solve the Friedmann equations, we must consider the matter and energy components of the Universe and describe how they contribute to the total energy density budget.

- **Cold matter.** In this component, we consider all the non-relativistic massive elements, for which the momentum is negligible with respect to the mass, so they are pressureless, i.e. $P = 0 \Rightarrow w = 0$. Cold matter is mainly composed of ordinary matter, i.e. baryons, and cold dark matter. Additionally, non-relativistic massive neutrinos contribute to the cold matter. The energy density evolution is given by Eq. (1.30) with $w = 0$,

$$\rho(a) = \rho_0 a^{-3} \,. \tag{1.36}$$

  This equation describes the mass conservation: if the comoving particles number is conserved, the energy density is proportional to the particle number density, which scales as a^{-3}.

- **Radiation.** This component refers to relativistic elements, mainly represented by photons and relativistic neutrinos. The equation of state can be derived considering the energy-momentum tensor of the electromagnetic field, which is given by

$$T_{\mu\nu} = F_{\mu\lambda} F_{\nu}{}^{\lambda} - \frac{1}{4} g_{\mu\nu} F_{\alpha\beta} F^{\alpha\beta} \,, \tag{1.37}$$

  where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the Faraday tensor, in which A_μ is the 4-vector potential of the electromagnetic field [5]. Since $F_{\mu\nu}$ is antisymmetric, the corresponding stress energy tensor is traceless,

$$T^{\mu}_{\ \mu} = F^{\mu}_{\ \lambda} F_{\mu}{}^{\lambda} - \frac{1}{4} 4 F_{\mu\nu} F^{\mu\nu} = 0 \,, \tag{1.38}$$

  therefore, considering the components of the energy momentum tensor, Eq. (1.23), we obtain

$$P = \frac{\rho}{3} \,. \tag{1.39}$$

  From this equation of state follows

$$\rho(a) = \rho_0 a^{-4} \,. \tag{1.40}$$

  This equation can be physically interpreted considering that the number of photons is conserved, so their number density scales as $\propto a^{-3}$, while the observed energy of each photon is proportional to the wavelength, which scales proportionally to a^{-1}, therefore the photon energy density scales as $\propto a^{-4}$.

- **Curvature.** Even if the curvature is a property of the space-time and not a matter-energy field in the Universe, it can be treated as an energy component. In the first Friedmann equation, Eq. (1.26), we can factorize the $8\pi G/3$ term in such a way that we can consider

$$\rho_K = -\frac{3}{8\pi G} \frac{K}{a^2} \,. \tag{1.41}$$

  This means that the curvature acts like an energy component scaling as $\propto a^{-2}$, so the corresponding equation of state is $w = -1/3$, according to comparison with the second Friedmann equation Eq. (1.27).

- **Cosmological constant.** The cosmological constant is an intrinsic property of the space-space time, nevertheless, in the Friedmann equations can be treated as an energy component

$$\rho_\Lambda = -\frac{\Lambda}{8\pi G} \,. \tag{1.42}$$

It follows that we can consider that to the cosmological constant corresponds a constant energy density, for which the equation of state is $w = -1$, as can be seen by comparing the first and second Friedmann equations, Eq. (1.26) and Eq. (1.27). The fact that ρ_Λ does not change as the Universe expands has a remarkable consequence: since the energy density of the other components decreases with the scale factor, the cosmological constant becomes more dominant as time increases with respect to the other components. In addition, the fact that it corresponds to a constant energy density despite the increasing space volume, entails that the cosmological constant behaves like a vacuum energy.

The first Friedmann equation can now be written in the compact form

$$H^2 = \frac{8\pi G}{3} \sum_i \rho_i(a)\,, \tag{1.43}$$

where the index i labels the specie of matter-energy. We now introduce the critical density

$$\rho_\mathrm{c} = \frac{3H_0^2}{8\pi G}\,, \tag{1.44}$$

where H_0 is the Hubble rate at the present time. We can define the density parameters $\Omega(a) = \rho(a)/\rho_\mathrm{c}$, so that the first Friedmann reads

$$H^2(a) = H_0^2 \left[\Omega_\mathrm{r} a^{-4} + \Omega_\mathrm{m} a^{-3} + \Omega_K a^{-2} + \Omega_\Lambda\right]\,, \tag{1.45}$$

in which we consider the matter-energy components described previously: the subscript r stands for radiation, m for matter, K for curvature, Λ for the cosmological constant. We use the notation Ω_i to indicate a density parameter at present time, when the time dependence is implicit we will write $\Omega_i(a)$. Sometimes, when the notation would be ambiguous, we will add subscript 0, i.e. $\Omega_{i,0}$, to stress the fact that a density parameter is considered at the present time. From the definition of the density parameters follows

$$\sum_i^{i\neq K} \Omega_i + \Omega_K = 1\,. \tag{1.46}$$

Therefore, the Universe is open if $\Omega < 1$, flat if $\Omega = 1$ and closed if $\Omega > 1$, where $\Omega = \sum_i^{i\neq K} \Omega_i$ is the sum of all the density parameters except the curvature one.

1.1.5 Distances and horizons

Since the Universe is expanding, there is no unique way to define distances on our backward light cone. The first distance notion we consider is the comoving distance: this distance does not change with time for comoving observers, hence the name. The analytical expression of the comoving distance between us and a source is given by the conformal time spent or, equivalently, by the comoving distance traveled by a photon in the conformal space, described by Eq. (1.11),

$$\mathrm{d}\eta = \mathrm{d}\chi\,, \tag{1.47}$$

it follows

$$\chi = \int_t^{t_0} \frac{\mathrm{d}t'}{a(t')} = \int_a^1 \frac{\mathrm{d}a'}{a'^2 H(a')} = \int_0^z \frac{\mathrm{d}z'}{H(z')}\,. \tag{1.48}$$

Another way to determine distances in astronomy is to measure the angle θ subtended by an object of known physical size l. In the small angle limit, the distance to that object is

$$D_A = \frac{l}{\theta} . \tag{1.49}$$

To compute the angular diameter distance in an expanding universe, we first note that the comoving size of the object is $\chi = l/a$. The comoving distance is the distance in the conformal metric, and conformal transformations preserve angles, therefore the angle subtended is $\theta = (l/a)/\chi(a)$. Comparing with Eq. (1.48) we obtain

$$D_A = a\chi = \frac{\chi}{1+z} . \tag{1.50}$$

Note that the angular diameter distance converges to the comoving distance at low redshift, but decreases at higher redshift. This means that, at least in a flat universe, objects at higher redshift appear larger than they would at a lower redshift.

A third way to infer distances in cosmology is to measure the flux from an object of known luminosity. The observed flux F at distance D_L from a source of known luminosity L is

$$F = \frac{L}{4\pi D_L^2} , \tag{1.51}$$

where D_L is the luminous distance. The flux observed in comoving coordinate is

$$F = \frac{L(\chi)}{4\pi\chi^2} , \tag{1.52}$$

where $L(\chi)$ is the luminosity through a comoving spherical shell with radius $\chi(a)$. The luminosity is the the sum of photons passing through the comoving spherical shell per unit time multiplied by their energy. In a fixed time interval, photons travel farther on the comoving grid at early times than at late times, since the associated physical distance at early times is smaller. Therefore, the number of photons crossing a shell in the fixed time interval will be lower today than at emission by a factor of a. Similarly, the photons will be redshifted because of the expansion, so their energy is decreased by a factor a with respect to the emission. Therefore, the energy per unit time passing through a comoving shell at distance $\chi(a)$ from the source will be a factor of a^2 lower than the luminosity at the source L_e

$$F = \frac{L_\mathrm{e} a^2}{4\pi\chi^2} , \tag{1.53}$$

so

$$D_L = \frac{\chi}{a} = (1+z)\chi . \tag{1.54}$$

Let consider the comoving distance that light could have traveled since $t = 0$, this is defined by Eq. (1.48), where $t = 0$ corresponds to $a = 0$ and $z \to \infty$. Causal physical processes could not have occurred at distances greater than this distance, which is therefore called the "comoving horizon" or the "particle horizon". It is important to consider the causal horizon at various epochs, because this tells us on which scale causal physical processes occur, so the integration of Eq. (1.48) runs from $t = 0$ to the epoch considered, i.e. t_0 is substituted by a generic t. Another quantity related to causality is the Hubble radius, defined as

$$\lambda_H = \frac{1}{aH} . \tag{1.55}$$

This quantity defines the physical scale in which processes can be in causal relation in a time scale of H^{-1}. As the comoving horizon, the Hubble radius depends on the epoch considered.

1.2 The inhomogeneous Universe

Homogeneity and isotropy well describe our Universe at the largest scales; nevertheless, at smaller scales we see that the Universe is far from being homogeneous and isotropic. To describe the evolution of the Universe at different scales, in particular on the scales of inhomogeneity, it is possible to split the dynamics of the Universe into background and perturbations. Moreover, the Standard Cosmological Model provides that the initial condition, i.e. the post-inflationary scenario (see Sect. 1.4), is described by small amplitude perturbations. The amplitude remains small enough to allow for a linear order treatment at least up to the CMB epoch and further.

1.2.1 Perturbed metric

The most general perturbed FLRW metric can be written as

$$\mathrm{d}s^2 = a^2(t)\left\{-(1+2\Psi)\mathrm{d}\eta^2 + -2B_i\mathrm{d}\eta\mathrm{d}x^i + \left[(1-2\Phi)\gamma_{ij} + h_{ij}\right]\mathrm{d}x^i\mathrm{d}x^j\right\}, \tag{1.56}$$

where Ψ, Φ, B_i, and h_{ji} are all functions of time and space and include scalar, vector, and tensor type perturbations. The h_{ij} tensor is traceless, i.e. $\gamma^{ij}h_{ij} = 0$. In linear order in perturbations, we can treat the perturbation variables as 3-tensors and raise/lower their indices with γ_{ij}, and we raise/lower components of 4-vectors with the 4-metric $g_{\mu\nu}$. It is convenient to decompose the spatial-like vectors B_i into its longitudinal, i.e. curl-free, and transverse, i.e. divergence-free, pieces,

$$B_i = B_i^{\parallel} + B_i^{\perp} = \partial_i b + B_i^{\perp}, \tag{1.57}$$

where

$$\begin{aligned} \nabla \times \mathbf{B}^{\parallel} &= \epsilon^{ijk}\partial_j B_k^{\parallel} = \epsilon^{ijk}\partial_j\partial_k b = 0 \\ \nabla \cdot \mathbf{B}^{\perp} &= \partial^i B_i^{\perp} = 0\,, \end{aligned} \tag{1.58}$$

and ϵ^{ijk} is the Levi-Civita tensor. The potential b is a spin 0 perturbation, while $B_i^{\perp}$ is a spin 1 perturbation. The traceless spatial-like tensor h_{ij} can be decomposed into

$$h_{ij} = h_{ij}^{\top} + h_{ji}^{\parallel} + h_{ij}^{\perp}, \tag{1.59}$$

where $h_{ij}^{\top}$ is transverse, while the divergences of $h_{ij}^{\parallel}$ and $h_{ij}^{\perp}$ are longitudinal and transverse vectors, respectively, i.e.

$$\partial^i h_{ij}^{\top} = 0 \qquad \text{and} \qquad \gamma^{ij}h_{ij}^{\top} = 0\,, \tag{1.60}$$

$$\epsilon^{ijk}\partial_j\partial^l h_{kl}^{\parallel} = 0 \qquad \text{and} \qquad \gamma^{ij}h_{ij}^{\parallel} = 0\,, \tag{1.61}$$

$$\partial^i\partial^j h_{ij}^{\perp} = 0\,, \qquad \text{and} \qquad \gamma^{ij}h_{ij}^{\perp} = 0\,. \tag{1.62}$$

Note that $h_{ij}^{\parallel}$ is a scalar perturbation, while $h_{ij}^{\perp}$ is a vector perturbation, and $h_{ij}^{\top}$ corresponds to a genuine tensor perturbation. Its 2 independent degrees of freedom are the polarizations of gravitational waves, so it corresponds to a massless spin-2 field [6].

In General Relativity, the notion of perturbation is ambiguous. There is no unique decomposition of a variable, and it is always possible to perform arbitrary coordinate transformations.

To define perturbations, we have to compare the actual perturbed space-time $\mathcal{M}$ to a homogeneous and isotropic FLRW reference space-time $\overline{\mathcal{M}}$. After that, the density perturbation at the point $x^\mu \in \overline{\mathcal{M}}$ is

$$\delta\rho(x^\mu) = \rho(x^\mu)|_{\mathcal{M}} - \rho(x^\mu)|_{\overline{\mathcal{M}}} \,. \tag{1.63}$$

This introduces a gauge choice in the specific map between $\mathcal{M}$ and $\overline{\mathcal{M}}$. A gauge transformation changes the map between $\mathcal{M}$ and $\overline{\mathcal{M}}$ without changing the coordinate system, i.e. the chart. This induces a coordinate change on M known as a gauge transformation, which can be viewed as an infinitesimal coordinate change. Under this transformation, the coordinates of the point $x^\mu \in \overline{\mathcal{M}}$ change to $x'^\mu = x^\mu + \epsilon^\mu$. In the following, we will consider ϵ^μ as a small quantity, since from the point of view of General Relativity, the cosmological perturbations are always small quantities. It follows that a density perturbation after a gauge transformation can be expressed as

$$\begin{aligned}
\delta\rho'(x'^\mu) &= \rho'(x'^\mu) - \bar{\rho}'(x'^\mu) \\
&= \rho(x^\mu) - \bar{\rho}'(x'^\mu) \qquad (\rho \text{ transforms as a scalar}) \\
&= \rho(x^\mu) - \bar{\rho}(x'^\mu) \qquad (\bar{\rho} \text{ is the same on } \mathcal{M} \text{ and } \overline{\mathcal{M}}) \\
&= \rho(x'^\mu - \epsilon^\mu) - \bar{\rho}(x'^\mu) \\
&= \rho(x'^\mu) - \epsilon^\mu \partial_\mu \rho(x'^\mu) - \bar{\rho}(x'^\mu) \,.
\end{aligned} \tag{1.64}$$

Note that $\bar{\rho}$ depends on time only, so

$$\delta\rho' = \delta\rho - \epsilon^\mu \partial_\mu \rho = \delta\rho - \epsilon^0 \partial_0 \rho \,. \tag{1.65}$$

A gauge transformation involves $\epsilon_\mu = (\epsilon_0, \epsilon_i)$ that can be decomposed into two spin 0 perturbations, i.e. ϵ_0 and the spatial longitudinal part $\epsilon_i^{\parallel}$, and one spin 1 perturbation, i.e. the spatial transverse part $\epsilon_i^{\perp}$. It follows that a gauge transformation involves four functions, so the 10 degrees of freedom of the perturbed metric $g_{\mu\nu}$ are reduced to $10 - 4 = 6$ physical degrees of freedom. Moreover, we showed that the six physical degrees of freedom decompose into two spin 0, one spin 1 and one spin 2 perturbations. Since gauge transformations do not involve spin 2 quantities, the $h_{ij}^{\top}$ perturbation term, i.e. gravitational waves, is gauge invariant, so it remains the same under any redefinition of the background. All the other perturbation terms depend on the gauge choice.

The two most popular choices are the synchronous gauge, which corresponds to $\Psi = B_i = 0$, and the conformal Newtonian or longitudinal gauge, in which only scalar degrees of freedom are considered, therefore the corresponding metric is

$$\mathrm{d}s^2 = -(1 + 2\Psi)\mathrm{d}t^2 + a^2(t)(1 - 2\Phi)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j \,. \tag{1.66}$$

In the following, we will consider the Newtonian gauge only. The name of this gauge choice is due to the fact that in the non-relativistic limit, the perturbed time term takes the form of the Poisson equation with a Newtonian gravitational potential Ψ. The Φ term is the perturbation to the curvature of constant time hypersurfaces. In the absence of anisotropic stress $\Phi = \Psi$ [6]. After the radiation epoch, i.e. the cosmological epoch in which the radiation energy density was the dominant term in the Friedmann equations, the anisotropic stress became negligible, so $\Phi \simeq \Psi$. We will not consider tensor and vector perturbations because the scalar perturbations are the only ones that couple to matter perturbations and are the most important ones that couple to photon perturbations as well [6].

1.2.2 Newtonian approximation of structure formation

Let us consider the perturbed metric (Eq. (1.66)) explicitly recovering the light velocity c,

$$\mathrm{d}s^2 = -\left(1 + 2\frac{\Psi}{c^2}\right)c^2\mathrm{d}t^2 + a^2(t)\left(1 - 2\frac{\Phi}{c^2}\right)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j\,. \tag{1.67}$$

It can be seen that c simplifies in the Newtonian potential term Ψ, while it suppresses the curvature term Φ. The Newtonian approximation is valid until $\Phi/c^2 \ll 1$, so that this term can be neglected. Let us consider the Poisson equation

$$\nabla^2\Psi = 4\pi G\bar{\rho}\delta\,, \tag{1.68}$$

where delta is the density contrast defined as

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}\,. \tag{1.69}$$

If we assume the spherical approximation, the potential can be written as $\Psi = 4\pi G\rho\delta\lambda^2$, where λ is the linear comoving scale of the system. Using the first Friedmann equation, Eq. (1.26), we can write

$$\Psi \sim \lambda^2 H^2 a^2\delta \qquad \Longrightarrow \qquad \frac{\Psi}{c^2} \sim \frac{\lambda_{\mathrm{ph}}}{\lambda_H}\delta\,, \tag{1.70}$$

where $\lambda_{\mathrm{ph}} = a\lambda$ is the physical scale of the system and $\lambda_H = c/H$ is the comoving Hubble radius. At present time $\lambda_H \sim 10$ Gpc, and $\delta \ll 1$ on those scales. In LSS (non-relativistic) analysis, the maximum relative distances considered are $\sim 100h^{-1}$Mpc, it follows that the Newtonian approximation accurately describes the formation of structures.

Consider the dynamics of self-gravitating collisionless particles; each particle follows Newton's law

$$\mathbf{r} = a\mathbf{x} \qquad \ddot{\mathbf{r}} = -\nabla_{\mathbf{r}}\boldsymbol{\Psi} \tag{1.71}$$

with

$$\nabla^2\Psi = 4\pi G\rho\,. \tag{1.72}$$

With the symbol $\nabla_{\mathbf{r}}$ we indicate the gradient in the physical coordinates, i.e. $\nabla_{\mathbf{r}} = \partial/\partial\mathbf{r}$, while ∇ is the gradient in the comoving coordinates, $\nabla = \partial/\partial\mathbf{x}$. Note that the Poisson equation for the homogeneous background provides the second Friedmann equation, Eq. (1.27). The time derivative of the physical position $\mathbf{r}$ is

$$\dot{\mathbf{r}} = \dot{\mathbf{a}}\mathbf{x} + \mathbf{a}\dot{\mathbf{x}} = \mathbf{H}\mathbf{r} + \mathbf{v}\,, \tag{1.73}$$

where $\mathbf{v}$ is the physical velocity of the particle. We can now write the Lagrangian of the particle,

$$\mathcal{L}' = \frac{m}{2}(\dot{a}\mathbf{x} + a\dot{\mathbf{x}})^{\mathbf{2}} - \mathbf{m}\boldsymbol{\Psi}(\mathbf{x}, \mathbf{t})\,. \tag{1.74}$$

The equations of motion are defined from the Lagrangian unless to a total time derivative of a function [7]. Thus, we consider the canonical transformation

$$\mathcal{L} = \mathcal{L}' - \frac{\mathrm{d}\psi}{\mathrm{d}t} \qquad \text{where} \qquad \psi = \frac{m}{2}a\dot{a}x^2\,, \tag{1.75}$$

obtaining

$$\mathcal{L} = \frac{m}{2}a^2\dot{x}^2 - m\phi \qquad \text{where} \qquad \phi = \Psi - \Psi_{\rm bg} = \Psi + \frac{m}{2}a\ddot{a}x^2\,, \tag{1.76}$$

in which the footnote bg indicates background quantities. The Poisson equation in comoving coordinates is

$$\nabla^2\phi = 4\pi G a^2\rho + 4a\ddot{a} = 4\pi G a^2\bar{\rho}\delta. \tag{1.77}$$

From the Lagrangian we can derive the canonical momentum and force

$$\begin{aligned}\mathbf{p} &= \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}} = ma^2\dot{\mathbf{x}} = \mathbf{mav}\\ \dot{\mathbf{p}} &= \frac{\partial\mathcal{L}}{\partial\mathbf{x}} = -m\nabla_{\mathbf{x}}\phi = m\dot{a}\mathbf{v} + ma\dot{\mathbf{v}}\,,\end{aligned} \tag{1.78}$$

therefore

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + H\mathbf{v} = \frac{1}{a}\nabla\phi\,. \tag{1.79}$$

Note that this equation is the Euler fluid equation.

The Liouville theorem states that in the absence of collisions, the phase-space density $f(\mathbf{x},\mathbf{p},t)$ is conserved, i.e.

$$\mathrm{d}f(\mathbf{x},\mathbf{p},t) = 0 \qquad \Longrightarrow \qquad \frac{\partial f}{\partial t} + \dot{x}^i\frac{\partial f}{\partial x^i} + \dot{p}^i\frac{\partial f}{\partial p^i} = 0\,, \tag{1.80}$$

inserting the expression for velocity and momentum, we obtain the Vlasov equation [8]

$$\frac{\partial f}{\partial t} + \frac{p^i}{ma^2}\frac{\partial f}{\partial x^i} - m\partial^i\phi\frac{\partial f}{\partial p^i} = 0\,, \tag{1.81}$$

which must be solved together with the Poisson equation for ϕ. The Vlasov equation can be solved for any momentum of the phase-space distribution, in particular, the $0^{\rm th}$ order momentum [8]

$$\rho(\mathbf{x},t) = \frac{m}{a^3}\int \mathrm{d}^3p f(\mathbf{x},\mathbf{p},t) \tag{1.82}$$

is the mass density; the normalized $1^{\rm st}$ moment

$$v_i(\mathbf{x},t) = \frac{1}{ma}\frac{\int \mathrm{d}^3p\, p_i f(\mathbf{x},\mathbf{p},t)}{\int \mathrm{d}^3p f(\mathbf{x},\mathbf{p},t)} \tag{1.83}$$

is the stream velocity; the $2^{\rm st}$ moment provides the velocity dispersion tensor

$$\Pi_{ij}(\mathbf{x},t) = \langle v_iv_j\rangle - v_iv_j = \frac{1}{m^2a^2}\left[\frac{\int \mathrm{d}^3p\, p_ip_jf(\mathbf{x},\mathbf{p},t)}{\int \mathrm{d}^3p f(\mathbf{x},\mathbf{p},t)} - \frac{\int \mathrm{d}^3p\, p_if(\mathbf{x},\mathbf{p},t)\int \mathrm{d}^3p\, p_jf(\mathbf{x},\mathbf{p},t)}{\left(\int \mathrm{d}^3p f(\mathbf{x},\mathbf{p},t)\right)^2}\right]. \tag{1.84}$$

If we integrate the Vlasov equation (Eq. (1.81)) with respect to p, we obtain

$$\frac{\partial}{\partial t}\left(a^3\rho\right) + \frac{1}{a^2}\frac{\partial}{\partial x^i}\int \mathrm{d}^3p\, p_if(\mathbf{x},\mathbf{p},t) = 0\,, \tag{1.85}$$

that reduces to

$$\frac{\partial\rho}{\partial t} + 3\frac{\dot{a}}{a}\rho + \partial_i(\rho u^i) = 0\,, \tag{1.86}$$

where $u^i = \dot{x}^i$. This is the continuity equation. Now, if we split the density into the background part and the perturbation part, i.e. $\rho(\mathbf{x}, t) = \bar{\rho}(t)(1 + \delta(\mathbf{x}, t))$, and we consider the background evolution of $\bar{\rho}$ given by Eq. (1.30), we obtain

$$\frac{\partial \delta}{\partial t} + \partial_i \left[(1 + \delta(\mathbf{x})) u^i \right] = 0 \,, \tag{1.87}$$

Note that both δ and u_i are first-order terms in perturbation, so the first-order continuity equation is

$$\frac{\partial \delta}{\partial t} + \partial_i u^i = 0 \,. \tag{1.88}$$

Now we consider the first moment of the Vlasov equation, multiplying Eq. (1.81) by $\mathbf{p}$ and then integrating it. To perform the calculation, note that by integrating by parts we can simplify the third term as

$$\int \mathrm{d}^3 p p^i \frac{\partial f}{\partial p^j} = -\delta^i{}_j \int \mathrm{d}^3 p f = \delta^i{}_j \rho \frac{a^3}{m} \,, \tag{1.89}$$

where $\delta^i{}_j$ is the Kronecker delta. Therefore, the first moment of the Vlasov equation becomes

$$\frac{\partial}{\partial t} \int \mathrm{d}^3 p p^i f + \frac{1}{ma^2} \int \mathrm{d}^3 p f \partial_j \int \mathrm{d}^3 p p^i p^j f + a^3 \rho \partial^j \phi = 0 \,. \tag{1.90}$$

Then, substituting Eq. (1.82), Eq. (1.83), Eq. (1.84) and the continuity equation, we obtain [8]

$$\frac{\partial v^i}{\partial t} + \frac{\dot{a}}{a} v^i + \frac{1}{a} v_j \partial^j v^i = -\frac{1}{a} \partial^i \phi - \frac{1}{a\rho} \partial_j \left(\rho \Pi^{ij} \right) \,, \tag{1.91}$$

that is the Euler equation. In the single stream regime, the velocity dispersion Π^{ij} vanishes, meaning that there is only one stream for each position. This is true before structure formation; when a structure fully collapses, various streams cross each other, and the system reaches the so called "shell crossing" event. Before shell crossing occurs, we can consider $\Pi_{ij} = 0$, and if we only consider first-order terms, Eq. (1.91) becomes

$$\dot{u}^i + 2 \frac{\dot{a}}{a} u^i = -\frac{1}{a^2} \partial^i \phi \,. \tag{1.92}$$

If we combine this equation with the time derivative of the continuity equation Eq. (1.88), together with the Poisson equation, we obtain

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G \bar{\rho} \delta = 0 \,. \tag{1.93}$$

Note that, by Eq. (1.44), the last term can be rewritten as $3H_0^2 \Omega_\mathrm{m} a^3 / 2$. This equation describes the linear growth of perturbations, and it allows for analytical solution for matter dominated, radiation dominated, and ΛCDM universe. In the matter dominated universe, the solution of the first Friedmann equation Eq. (1.26), gives $a \propto t^{2/3}$, $H = 2/(3t)$, and $\bar{\rho} = 1/(6\pi G t^2)$. It follows that Eq. (1.93) becomes

$$\ddot{\delta} + \frac{4}{3t} \dot{\delta} - \frac{2}{3t^2} \delta = 0 \,. \tag{1.94}$$

It is useful to write $\delta = \delta(\mathbf{x}, t_i) D(t)$, since the linear growth equation shows that at the linear level the time evolution of the density contrast field is decoupled from the position. The above equation provides two solutions, the growing mode and the decaying mode,

$$D_+(t) \propto t^{3/2} \propto a \qquad D_-(t) \propto t^{-1} \propto H \,, \tag{1.95}$$

where $D_+(t)$ is commonly called "linear growth factor", or simply "growth factor". The growing mode of ΛCDM is [6]

$$D_+(a) = \frac{5\Omega_\mathrm{m}}{2}\frac{H(a)}{H_0}\int_0^a \mathrm{d}a' \left(\frac{H_0}{a'H(a')}\right)^3 , \tag{1.96}$$

while the solution for generic cosmology is numerical [9]. Note that the cosmological constant suppresses the growth of perturbations. In some types of analysis, such as measurements of redshift-space distortions or of the integrated Sachs-Wolfe effect, it is useful to use the logarithmic derivative of the growth factor, called "growth rate"

$$f = \frac{\ln D(a)}{\ln a} . \tag{1.97}$$

In ΛCDM this quantity is often effectively parameterized as [6]

$$f \simeq \Omega_\mathrm{m}^\gamma \tag{1.98}$$

with $\gamma \simeq 0.55$.

1.3 From dark matter to halos

The concordance cosmological model describes the Universe as dominated by dark components, i.e. dark matter and dark energy, generally speaking. Cosmological perturbation theory describes the statistical properties of the fluctuations of the matter density field. Nevertheless, observations measure the light emitted by galaxies, and cosmological analyses of galaxy surveys attempt to extract the global properties of the Universe via the measured statistical properties of galaxies. This is like inferring the shape of the continents on Earth (matter distribution on large scales) by looking at satellite pictures of city lights (the galaxies). Although a correlation between the two distributions is present, the mapping between galaxies and the underlying matter distribution is complex. This map is known as the galaxy bias.

Unfortunately, a complete theory of galaxy formation and evolution from first principles does not exist yet. The physics of galaxy formation is awfully complicated, and to date it can only be faced in an effective way. A first step towards understanding the mechanisms responsible for biasing is to ignore the effects of all forces but gravity and study how the distribution of (dark) matter clumps, within which galaxies should form, in relation to the underlying mass distribution. Therefore, in this treatment, matter is considered as a non-collisional gas at any scale, like cold dark matter, so we often exchange the terms of matter and dark matter, since they are treated in the same way. This simplified model of galaxy formation can be tackled both theoretically and via cosmological simulations that are particularly important in considering non-linearities at small scales. The objects we consider here are therefore dark matter halos instead of galaxies. A dark matter halo is defined as a fully collapsed region in the matter distribution of the Universe, forming a virialized gravitationally bound object. There are various methodologies to consider when an halo is formed, mainly based on thresholding: an halo is considered formed in a position $\mathbf{x}$, when the density contrast filtered on a given scale exceeds a threshold value. We postpone a detailed description of dark matter halo formation in Chapter 2, while in this section we introduce the relation between halos and matter distribution on larger scale. In particular we consider how bias emerges in the one-point (Sect. 1.3.1 and Sect. 1.3.3)

and two-point (Sect. 1.3.2) statistics of halos, how halos are considered in cosmological simulations (Sect. 1.3.5), and how the relation between galaxies and dark matter halo is treated in the literature (Sect. 1.3.6).

1.3.1 The halo mass function

The halo mass function is the differential number density of halos as a function of their mass. Press & Schechter [10] presented an analytical model to compute the number density of self-gravitating objects formed by gravitational instability in hierarchical scenarios of structure formation with Gaussian initial conditions. The key assumption is that all density fluctuations with linearly extrapolated density contrast above a critical threshold $\delta_{\rm c}$ of order one undergo gravitational collapse. More precisely, a region of volume $V = 4\pi R^3/3$ in the initial density field will collapse to form a halo of mass $M = \rho_{\rm m} V$, when its linearly extrapolated density contrast filtered on the scale R, δ_R, exceeds the critical threshold. In the treatment of halo bias we will often consider the initial density field linearly evolved to the epoch of interest, this is called Lagrangian space.

From the assumption of Gaussianity of the Lagrangian field, the fraction of fluctuations belonging to halos of mass greater than M is given by the complementary error function

$$p(\delta_R > \delta_{\rm c}) = \frac{1}{\sqrt{2\pi\sigma^2(R)}} \int_{\delta_{\rm c}}^{\infty} {\rm d}\delta \exp\left[-\frac{1}{2}\frac{\delta^2}{\sigma^2(R)}\right] = \frac{1}{2}{\rm erfc}\left[\frac{\delta_{\rm c}}{\sqrt{2}\sigma(R)}\right]. \tag{1.99}$$

Note that the halo mass is in one-to-one correspondence with the filtering radius. In this equation $\sigma(R)$ is the root-mean-square of linear smoothed density perturbations on the scale R,

$$\sigma^2(R) = \left\langle \delta_R^2 \right\rangle = \int \frac{{\rm d}k}{2\pi^2} k^2 P(k) W^2(Rk)\,, \tag{1.100}$$

where $W(Rk)$ is the filtering kernel in Fourier space. The error function $p(\delta_R > \delta_{\rm c})$ describes the Lagrangian volume fraction of the total mass enclosed in halos greater than M, which can be written as

$$F(>M) = \frac{1}{\rho_{\rm m}} \int_M^{\infty} {\rm d}M' M' \frac{{\rm d}n_{\rm h}(M')}{{\rm d}M'} = p(\delta_R > \delta_{\rm c})\,, \tag{1.101}$$

where $n_{\rm h}(M)$ is the number density of halos with mass M and ${\rm d}n_{\rm h}/{\rm d}M$ is the halo mass function. We can obtain an explicit expression of the halo mass function by taking the derivative with respect to the mass M of the above expression. The differential volume fraction of halos of mass M is therefore

$$f(M) = -\frac{{\rm d}F(>M)}{{\rm d}M} = -\frac{{\rm d}p(\delta_R > \delta_{\rm c})}{{\rm d}M}\,, \tag{1.102}$$

this quantity is usually called the "multiplicity function", and the corresponding halo mass function is

$$\frac{{\rm d}n_{\rm h}}{{\rm d}M} = \frac{\rho_{\rm m}}{M} f(M)\,. \tag{1.103}$$

Note that this mass function does not appear to be properly normalized, since an integration over all the mass included in halos only recovers half of the total mass,

$$\begin{aligned} \int_0^{\infty} {\rm d}M' M' \frac{{\rm d}n_{\rm h}(M')}{{\rm d}M'} &= -\rho_{\rm m} \int_M^{\infty} {\rm d}M' \frac{{\rm d}p(\delta_R > \delta_{\rm c})}{{\rm d}M'} \\ &= -\rho_{\rm m} \left[p(R=\infty) - p(R=0)\right] = \frac{1}{2}\rho_{\rm m}\,, \end{aligned} \tag{1.104}$$

where the last equality is obtained from Eq. (1.99) considering that $\sigma(R) \to 0$ for $R \to \infty$ and $\sigma(R) \to \infty$ for $R \to 0$. To resolve this issue, [10] introduced an *ad hoc* fudge factor of two which corrects the normalization and leads to the Press-Schechter mass function,

$$\frac{\mathrm{d}n_\mathrm{h}}{\mathrm{d}M} \to 2\frac{\mathrm{d}n_\mathrm{h}}{\mathrm{d}M}\,. \tag{1.105}$$

Let us now express the multiplicity function by taking the derivative with respect to the mass of Eq. (1.99) corrected by the fudge factor. Note that the multiplicity function defined in Eq. (1.102) is a differential quantity, so to properly change the variable, we must consider the differential form $f(X)\,\mathrm{d}X = f(M)\,\mathrm{d}M$, where X is the new generic variable. It follows that the Press-Schechter halo mass function can be expressed as

$$\frac{\mathrm{d}n_\mathrm{h}}{\mathrm{d}M}(\delta_\mathrm{c}) = \frac{\rho_\mathrm{m}}{M} f(\nu_\mathrm{c}) \left|\frac{\mathrm{d}\nu_\mathrm{c}}{\mathrm{d}M}\right| \tag{1.106}$$

with

$$f(\nu_\mathrm{c}) = \sqrt{\frac{2}{\pi}} \exp\left[-\frac{\nu_\mathrm{c}^2}{2}\right], \tag{1.107}$$

where $\nu_\mathrm{c} = \delta_\mathrm{c}/\sigma(R)$ is the significance of the critical density in terms of the standard deviation of matter fluctuations filtered on the scale R. This expression for the halo mass function has a clear interpretation: the ρ_m/M term is the number density of fluctuations with mass M, the multiplicity function is the fraction of the fluctuation of mass M that collapsed to form a halo, the last factor is the Jacobian of the transformation from the variable of the multiplicity function into the mass scale. Note that this formulation is general for any halo-mass function model, where, for each of them, corresponds a specific multiplicity function.

From a theoretical point of view, the Press-Schechter theory is the first attempt to predict the halo mass function from first principles, showing strength and weakness. The strength of the Press-Schechter model is that it introduces all the features that a halo mass function is expected to have. In particular, note that the halo mass function is fully determined by few ingredients, i.e. the power spectrum of linear perturbations, the expansion history of the Universe that determines the linear growth factor $D(z)$ hidden in σ, responsible of the linear extrapolation of the initial density field, the value of the threshold δ_c, the function form of the filter kernel in computing $\sigma(R)$, which fixes the exact relation between M and R. The weak point is that this model is not accurate enough to encapsulate the physics of halo formation, reflected by the fudge factor correcting the normalization, due to the so called *cloud-in-cloud* problem [11, 12]. This term was introduced to describe the fact that, in the Press-Schechter model, all the linearly extrapolated fluctuations exceeding the critical threshold would form a halo. However, if a region that exceeds the threshold on a scale R_1 is embedded in a forming halo overdensity on a larger scale $R_2 > R_1$, the larger fluctuation will collapse to form one single halo; consequently, the smaller fluctuation should not be taken into account in the halo mass function. The *cloud-in-cloud* problem, and consequently the wrong normalization, is solved within the excursion-set framework, that we widely discuss in Chapter 2.

Comparisons with simulations show that the Press-Schechter theory captures the general behavior of the halo mass function, but it does not accurately reproduce the measurements. In particular, the Press-Schechter halo mass function predicts a stronger high mass suppression with respect to simulations. The original idea of thresholding the initial density field was modified by Bond & Myers [13] considering the collapse of an ellipsoid and its impact in

describing halo formation and, consequently, in predicting the halo mass function. Sheth & Tormen [14] showed that the ellipsoidal collapse can be implemented effectively considering two extra degrees of freedom in the halo mass function, obtaining [14, 15]

$$f(\nu_c, p, q) = \sqrt{\frac{2}{\pi}}\left[1 + \frac{1}{2^p\sqrt{\pi}}\Gamma\left(\frac{1}{2} - p\right)\right]^{-1}\left[1 + (q\nu_c^2)^{-p}\right]\sqrt{q}e^{-q\nu_c^2/2}\,. \tag{1.108}$$

In this expression, known as the Sheth-Tormen multiplicity function, p and q are not *a priori* quantities, but they are fitted to the halo mass function measured in simulations; therefore, these parameters generally depend on the precise definition of halos used in the simulation analysis. In the original work, these values were fixed to $p = 0.3$ and $q = 0.707$ [14, 15] . Note that for $p = 0$ and $q = 1$ we recover the Press-Schechter multiplicity function.

1.3.2 Halos bias from Lagrangian clustering

From perturbation theory, we can derive the two-point statistics of matter density fluctuations. Considering the Press-Schechter arguments, we want to study the correlation of overdense regions above a threshold value in Lagrangian space, that we call proto-halos. As before, we consider that perturbations in Lagrangian space follow Gaussian statistics; therefore, the two-point correlation function, or equivalently the power spectrum, encapsulates all the statistical properties. The spatial distribution of the overdense regions is given by

$$p(\mathbf{q}) = \Theta(\delta_R(\mathbf{q}) - \delta_c)\,, \tag{1.109}$$

where $\mathbf{q}$ is the Lagrangian position, $\Theta(x)$ is the Heaviside step function, and the footnote R denotes that the density contrast field is filtered on a Lagrangian scale R. Note that this spatial density distribution is defined up to a normalization factor. The correlation of the overdense regions is given by

$$1 + \xi_h(r) = \frac{\langle\Theta(\delta_R(\mathbf{q}) - \delta_c)\,\Theta(\delta_R(\mathbf{q}+\mathbf{r}) - \delta_c)\rangle}{\langle\Theta(\delta_R(\mathbf{q}) - \delta_c)\rangle^2}\,. \tag{1.110}$$

the expected number density of overdense regions is the one computed in Press-Schechter theory, so

$$\langle\Theta(\delta_R(\mathbf{q}) - \delta_c)\rangle = \frac{1}{2}\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right), \tag{1.111}$$

while the numerator is

$$\langle\Theta(\delta_R(\mathbf{q}) - \delta_c)\,\Theta(\delta_R(\mathbf{q}+\mathbf{r}) - \delta_c)\rangle = \frac{\det(\mathbf{C}^{-1})}{2\pi}\int_{\delta_c}^{\infty}\int_{\delta_c}^{\infty} \mathrm{d}^2y \exp\left(-\frac{1}{2}\mathbf{y}^\top\mathbf{C}^{-1}\mathbf{y}\right) \tag{1.112}$$

where $\mathbf{y} = (y_1, y_2) = (\delta_R(\mathbf{q}), \delta_R(\mathbf{q}+\mathbf{r}))$ and $\mathbf{C}$ is the correlation matrix that we express as

$$\mathbf{C} = \sigma^2(R)\begin{pmatrix} 1 & w_R(r) \\ w_R(r) & 1 \end{pmatrix}, \qquad \mathbf{C}^{-1} = \frac{1}{\sigma^2(R)[1 - w_r^2(R)]}\begin{pmatrix} 1 & -w_R(r) \\ -w_R(r) & 1 \end{pmatrix}, \tag{1.113}$$

where $w_R(r) = \xi_R(r)/\sigma^2(R)$ and $\xi_R(r)$ is the two-point matter correlation function in Lagrangian space filtered on a scale R. Substituting $\nu = y/\sigma(R)$, we obtain

$$\begin{aligned}\langle \Theta(\delta_R(\mathbf{q}) - \delta_c)\, \Theta(\delta_R(\mathbf{q}+\mathbf{r}) - \delta_c)\rangle = \\ = \frac{1}{2\pi\sqrt{1-w_R^2(r)}} \int_{\nu c}^{\infty}\int_{\nu c}^{\infty} \mathrm{d}\nu_1 \mathrm{d}\nu_2 \exp\left[-\frac{\left(\nu_1 - w_R(r)\nu_2\right)^2}{2\left(1-w_R^2(r)\right)}\right] e^{-\nu^2/2} \\ = \frac{1}{2\pi}\int_{\nu c}^{\infty} \mathrm{d}\nu \exp\left[-\frac{\nu_c - w_R(r)\nu}{\sqrt{2\left(1-w_R^2(r)\right)}}\right] e^{-\nu^2/2}\,.\end{aligned} \tag{1.114}$$

The correlation of overdensities of Eq. (1.110) can now be expressed as

$$1 + \xi_h(r) = \sqrt{\frac{2}{\pi}}\left[\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-2} \int_{\nu c}^{\infty} \mathrm{d}\nu \exp\left[-\frac{\nu_c - w_R(r)\nu}{\sqrt{2\left(1-w_R^2(r)\right)}}\right] e^{-\nu^2/2}\,. \tag{1.115}$$

This computation was first performed by Kaiser [16] to explain the observed clustering properties of the Abell clusters. At large correlation distance $w_R(r) \ll 1$, and considering $\nu_c \gg 1$, i.e. large masses, the argument of the complementary error function in the integration sign can be approximated to $(\nu_c - w_R(r))/\sqrt{2}$, where we considered that only $\nu \sim \nu_c$ contributes to the integration, due to the exponential suppression of the $e^{-\nu^2/2}$ term. This approximation allows us to expand the complementary error function around the $\nu_c/\sqrt{2}$ value,

$$\mathrm{erfc}\left[\frac{1}{\sqrt{2}}\left(\nu_c - w(r)\right)\right] \simeq \mathrm{erfc}\left[\frac{\nu_c}{\sqrt{2}}\right] + \sqrt{\frac{2}{\pi}} w_R(r)\nu_c e^{-\nu_c^2/2}\,. \tag{1.116}$$

We then consider the Taylor expansion of the complementary error function in the limit for the argument tending to infinity,

$$\mathrm{erfc}(x) \simeq e^{-x^2}\left(\frac{1}{\sqrt{\pi}x} + \frac{1}{2\sqrt{\pi}x^3} + \frac{1}{4\sqrt{\pi}x^5} + \cdots\right) \qquad x \gg 1\,, \tag{1.117}$$

that allows us to write Eq. (1.115) as [16]

$$\xi_h(r) \simeq \frac{\nu_c^2}{\sigma^2(R)}\xi_R(r)\,. \tag{1.118}$$

This first result shows that the clustering of overdensity regions is not the same as that of the underlying matter distribution. Under the approximation used, i.e. r such that $\xi_R(r) \ll \sigma^2(R)$ and $\nu_c \gg 1$, the bias relation is linear. However, even in this simple thresholding model, the relation between $\xi_h(r)$ and $\xi_R(r)$ is generally non-linear.

Let us now consider the general N-point correlation, Eq. (1.112) can be generalized as

$$p_N(\mathbf{y}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det(\mathbf{C})}} \int_{\nu_c}^{\infty} \mathrm{d}^N y \exp\left(-\frac{1}{2}\mathbf{y}^\top \mathbf{C}^{-1}\mathbf{y}\right)\,, \tag{1.119}$$

where $\mathbf{y} = (y_1, \cdots, y_N)$ with $y_i = \nu(\mathrm{q}_i)$, therefore the covariance matrix can be written as $\mathbf{C} = \mathbb{I} + \mathbf{W}$, in components $C_{ij} = \delta_{ij} + w_{ij}$, where δ_{ij} is the Kronecker delta, $w_{ij} = 0$ for $i = j$

and $w_{ij} = w_{ji}$, since $\mathbf{C}$ is symmetric. Writing the N-dimensional Gaussian in Fourier space we obtain

$$p_N(\mathbf{y}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det(\mathbf{C})}} \int_{\nu_c}^{\infty} \mathrm{d}^N y \int_{-\infty}^{\infty} \frac{\mathrm{d}^N k}{(2\pi)^N} \exp\left(-\frac{1}{2}\mathbf{k}^\top \mathbf{C}\mathbf{k}\right) e^{-i\mathbf{k}\cdot\mathbf{y}} \,. \tag{1.120}$$

The Fourier transform allows to work with $\mathbf{C}$ instead of its inverse; consider now $\mathbf{k}^\top\mathbf{C}\mathbf{k}/2 = \mathbf{k}^2/2 + \mathbf{Q}(\mathbf{k})$, where $\mathbf{Q}(\mathbf{k}) = \mathbf{k}^\top\mathbf{W}\mathbf{k}/2$. This can be transformed as an operator acting on $e^{-i\mathbf{k}\cdot\mathbf{y}}$: $\mathbf{Q}(\mathbf{k}) \to \mathbf{Q}(i\partial/\partial\mathbf{y})$ so that it can be brought out of the k integral [17],

$$p_N(\mathbf{y}) = \frac{1}{(2\pi)^{N/2}} \int_{\nu_c}^{\infty} \mathrm{d}^N y \sum_{n=0}^{\infty} \frac{\mathbf{Q}^n\left(i\frac{\partial}{\partial\mathbf{y}}\right)}{n!} e^{-\mathbf{y}^2/2} \,. \tag{1.121}$$

Let go back to the case $N = 2$, where $w_{10} = w_{01} = w_R(r)$, so that $\mathbf{Q}(\mathbf{k}) = -w_R(r)k_1k_2$ and $\mathbf{Q}(i\partial/\partial\mathbf{y}) = w_R(r)\partial^2/\partial y_1\partial y_2$. The above equation reduces to

$$p_2(y_1, y_2) = \frac{1}{2\pi} \int_{\nu_c}^{\infty} \mathrm{d}y_1\mathrm{d}y_2 \sum_{n=0}^{\infty} \frac{1}{n!} \left(w_R(r)\frac{\partial}{\partial y_1}\frac{\partial}{\partial y_2}\right) e^{-\mathbf{y}^2/2} \,. \tag{1.122}$$

This expression shows that the proto-halo correlation function can be expressed as a series, where the $n = 0$ term is

$$\frac{1}{2\pi}\int_{\nu_c}^{\infty} \mathrm{d}y_1\mathrm{d}y_2 e^{-(y_1^2+y_2^2)/2} = \left[\frac{1}{2}\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^2 = \langle\Theta(\delta_R(\mathbf{q}) - \delta_c)\rangle^2 \,, \tag{1.123}$$

and the generic $n > 0$ is

$$\begin{aligned}\frac{1}{2\pi n!}\int_{\nu_c}^{\infty} \mathrm{d}y_1\mathrm{d}y_2 \left(w_R(r)\frac{\partial}{\partial y_1}\frac{\partial}{\partial y_2}\right)^n e^{-(y_1^2+y_2^2)/2} &= \frac{w_R^n(r)}{n!}\left(\left[\left(\frac{\partial}{\partial y}\right)^{n-1} e^{-y^2/2}\right]_{\nu_c}^{\infty}\right)^2 \\ &= \frac{w_R^n(r)}{n!}\left[2^{-\frac{n-1}{2}} e^{-\nu_c^2/2} H_{n-1}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^2 \,,\end{aligned} \tag{1.124}$$

where $H_n(x)$ is the Hermite polynomial of order n. The correlation function of overdensities can now be written as the series

$$\xi_{\mathrm{h}}(r) = \sum_{n=1}^{\infty} \frac{(b_n^{\mathrm{L}})^2}{n!} \xi_R^2(r) \,, \tag{1.125}$$

where

$$b_n^{\mathrm{L}} = \frac{2}{\sqrt{\pi}}\left[\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-1} \frac{2^{-n}e^{-\nu_c^2/2}}{\sigma^n(R)} H_{n-1}\left(\frac{\nu_c}{\sqrt{2}}\right) \,. \tag{1.126}$$

The b_n^{L} terms are the bias parameters, and the superscript L indicates that they refer to the Lagrangian density field. Since fluctuations in matter density are small on large scales, $\xi(r) \to 0$ as $r \to \infty$. Moreover, the expansion of Eq. (1.125) quickly converges, so at large scale the leading term accurately describe the correlation function of proto-halos,

$$\xi_{\mathrm{h}}(r) = (b_n^{\mathrm{L}})^2 \xi_R^2(r) \,. \tag{1.127}$$

This shows that on large scales the correlation function of overdense regions in Lagrangian space is directly proportional to the matter correlation function, and $\xi_{\rm h}(r)$ is enhanced with respect to the matter correlation function by the factor $(b_n^L)^2$. This bias term becomes large for rare high-mass overdensities, i.e. large Lagrangian patches that would collapse in massive halos, for which $\nu \gg 1$ and $\sigma(R) \to 0$. This simple thresholding model, as the Press-Schechter model, does not correctly describe low-mass objects due to the cloud-in-cloud problem, nevertheless, it captures the idea of halo bias and its general behavior.

Consider the density contrast of proto-halos $\delta_{\rm h}(\mathbf{q}) = n_{\mathbf{r}}(\mathbf{q})/\langle n_{\mathbf{h}}\rangle - 1$, note that if we express it as the series expansion in δ_R

$$\delta_{\rm h}(\mathbf{q}) = b_1^{\rm L}\delta_R^{\rm L}(\mathbf{q}) + \frac{1}{2}b_2^{\rm L}\left[(\delta_R^{\rm L}(\mathbf{q}))^2 - \sigma^2(R)\right] + \cdots , \tag{1.128}$$

we obtain Eq. (1.125) from the correlation $\langle\delta_{\rm h}(\mathbf{q})\delta_{\rm h}(\mathbf{q}+\mathbf{r})\rangle$. This series is known as the local bias expansion, since $\delta_{\rm h}(\mathbf{q})$ is a local function of the filtered matter density field $\delta_R^{\rm L}$, i.e. it depends only on the field $\delta_R^{\rm L}$ evaluated at the Lagrangian position q.

1.3.3 Peak-background split: from the halo mass function to halo bias

A model for the halo mass function of virialized structures automatically provides a way to predict their large-scale clustering properties. This can be shown considering the variation of the number density of halos with respect to the background matter density field. The basic idea of the peak-background split (PBS) approach is to decompose the density field into the sum of a background with a large coherence length and into a noisy component (peak) with a higher amplitude and a small coherence length [16, 18]. The PBS simply states that a long-wavelength density perturbation acts like a local modification of the background density, and it can be considered constant over the spatial scale within which halos form. In thresholding models for the halo mass function, the effect of the background can be reabsorbed in the critical threshold for halo formation. The PBS makes two assumptions in considering the halo number density: the abundance of halos depends on the amplitude of the matter power spectrum only through the variance of the matter density field $\sigma^2(M)$; the threshold of halo formation is unchanged by the presence of a long-wavelength density perturbation, which allows us to derive the bias parameters from the mass function [19]. This second assumption is exact for a spherical symmetric background, since it is a consequence of Birkhoff's theorem [3], and it is a good approximation for a background density field that varies on a length scale greater than the Lagrangian halo size, so that large-scale tidal effects can be ignored.

Let us consider the expected density contrast of proto-halos in Lagrangian space,

$$\delta_{\rm h}^{\rm L}(\delta_{\rm bg}^{\rm L}) = \frac{n_{\rm h}|_{\delta_{\rm bg}^{\rm L}}}{n_{\rm h}|_{\delta_{\rm bg}^{\rm L}=0}} - 1 . \tag{1.129}$$

An overdense density background increases the number of halos, since the halo formation threshold can be considered as $\delta_{\rm c} \to \delta_{\rm c} - \delta_{\rm bg}^{\rm L}$, so

$$\delta_{\rm h}^{\rm L}(\delta_{\rm bg}^{\rm L}) = \frac{n_{\rm h}(\delta_{\rm c} - \delta_{\rm bg}^{\rm L})}{n_{\rm h}(\delta_{\rm c})} - 1 . \tag{1.130}$$

From this expression, the peak-background split bias parameters in Lagrangian space can be obtained from the derivative of the mean number density of halos with respect to $\delta^{\rm L}_{\rm bg}$,

$$b^{\rm L}_N = \frac{1}{n_{\rm h}|_{\delta^{\rm L}_{\rm bg}=0}} \left.\frac{\partial^N n_{\rm h}}{\partial(\delta^{\rm L}_{\rm bg})^N}\right|_{\delta^{\rm L}_{\rm bg}=0} . \tag{1.131}$$

Let now consider the halo mass function defined in Eq. (1.106), the above equation becomes

$$b^{\rm L}_N(M) = \frac{(-1)^N}{\sigma^N(M)} \frac{1}{f(\nu_{\rm c})} \frac{{\rm d}^N f(\nu_{\rm c})}{{\rm d}^N \nu_{\rm c}} \tag{1.132}$$

Applying this formula to the Press-Schechter and Sheth-Tormen multiplicity function, i.e. Eq. (1.107) and Eq. (1.108) respectively, the linear and second-order Lagrangian bias parameters read

$$\begin{aligned} b^{\rm L}_{\rm PS,1} &= \frac{\nu_{\rm c}^2 - 1}{\delta_{\rm c}} \\ b^{\rm L}_{\rm PS,2} &= \frac{\nu_{\rm c}^2}{\delta_{\rm c}^2}(\delta_{\rm c}^2 - 3) , \end{aligned} \tag{1.133}$$

and

$$\begin{aligned} b^{\rm L}_{\rm ST,1} &= \frac{q\nu_{\rm c}^2 - 1}{\delta_{\rm c}} + \frac{1}{\delta_{\rm c}} \frac{2p}{1 + (q\nu_{\rm c})^p} \\ b^{\rm L}_{\rm ST,2} &= \frac{q\nu_{\rm c}}{\delta_{\rm c}^2}(q\delta_{\rm c}^2 - 3) + \frac{2p}{\delta_{\rm c}^2} \frac{(2p + 2q\nu_{\rm c}^2 - 1)}{1 + (q\nu_{\rm c}^2)^p} . \end{aligned} \tag{1.134}$$

The Sheth-Tormen PBS bias parameters provide a good fit with simulations.

To obtain these results, we introduced a background perturbation to the density contrast field. This method can be generalized to describe also other bias terms, e.g. with respect to the tidal field, higher derivative operators such as $\nabla^2\delta$ or $(\nabla\delta)^2$, and bias contributions induced by primordial non-Gaussianity [19].

1.3.4 Eulerian biasing

Up to now, we stressed the fact that the results presented refer to the Lagrangian space, i.e. the initial density field linearly evolved to the desired epoch. However, they do not account for large-scale flows that modify halo positions at the observation time with respect to the Lagrangian position. The fully non-linear evolved density field is called Eulerian space. A first approximation to the final distribution of halos can be provided by considering that large-scale regions in which halos are embedded will tend to collapse or expand depending on whether they are overdense or underdense. Assuming that the dynamics of the background field is in the single-stream regime, and considering that by definition we refer to proto-halos in the initial conditions as the exact progenitors of halos identified in the evolved field, the number of halos within a background patch of initial volume $V^{\rm L}$ and final volume V is conserved,

$$(1 + \delta^{\rm L}_{\rm h})V^{\rm L} = (1 + \delta_{\rm h})V , \tag{1.135}$$

while the mass conservation entails

$$(1 + \delta_{\rm bg})V = V^{\rm L}\,. \tag{1.136}$$

This last equality depends on the fact that the density contrast of the background patch in the initial condition is negligible by construction. Note that the quantities labeled with the superscript L refer to Lagrangian space, while the others refer to Eulerian space. Combining the two above equations, we obtain

$$1 + \delta_{\rm h} = (1 + \delta_{\rm h}^{\rm L})(1 + \delta_{\rm bg})\,. \tag{1.137}$$

If the background field is small, it does not undergo a full non-linear evolution, and its value can be interchanged with the corresponding one in linear evolution, i.e. its Lagrangian counterpart, $\delta_{\rm bg} \simeq \delta^L_{\rm gb}$, where $|\delta_{\rm gb}| \ll 1$. Eq. (1.137) becomes now

$$\delta_{\rm h} \simeq \delta_{\rm h}^{\rm L}|_{\delta^{\rm L}_{\rm bg}} + \delta^{\rm L}_{\rm bg} \simeq (1 + b_1^{\rm L})\,\delta^{\rm L}_{\rm bg}\,, \tag{1.138}$$

where we have truncated the Lagrangian bias expansion of $\delta_{\rm h}^{\rm L}|_{\delta^{\rm L}_{\rm bg}}$ to the first term because $|\delta^L_{\rm bg}| \ll 1$. It follows that the first term of Eulerian bias is

$$b_1^{\rm E} = 1 + b_1^{\rm L}\,. \tag{1.139}$$

Large-scale perturbations are small at any epoch of the Universe, therefore, the halo correlation function at $r \ll 1$ is well approximated considering the bias expansion truncated to the first term, as in Eq. (1.127). Nevertheless, this is not the case around the baryonic acoustic oscillation peak, i.e. $r \sim 100h^{-1}$Mpc, where non-linearities and non-locality become important [20, 21].

To extend the calculation of the Eulerian bias parameters to any order, we need the map between the Lagrangian and Eulerian background density field. As discussed at the beginning of this subsection, spherical symmetry provides an accurate approximation to describe the evolution of (non-extreme) background density fields, moreover the derivation of this map for the matter-dominated universe is analytical and its solution is almost cosmology independent (see Sect. 2.1). The spherical symmetry allows to expand the Lagrangian to Eulerian map and its inverse in the series [22–24]

$$\begin{aligned}
\delta &= \sum_{k=1}^{\infty} a_K \delta_{\rm L}^k \qquad \text{where} \qquad a_1 = 1, \quad a_2 = \frac{17}{21}, \quad a_3 = \frac{341}{567}, \quad \cdots \\
\delta_{\rm L} &= \sum_{k=1}^{\infty} a^I_K \delta^k \qquad \text{where} \qquad a^I_1 = 1, \quad a^I_2 = -\frac{17}{21}, \quad a^I_3 = \frac{2815}{3969}, \quad \cdots .
\end{aligned} \tag{1.140}$$

Inserting this map in Eq. (1.137) we obtain the Eulerian bias parameters [23, 25],

$$\begin{aligned}
b_1^{\rm E} &= 1 + b_1^{\rm L} \\
b_2^{\rm E} &= 2(1 + a_2^I) b_1^{+\rm L} + b_2^{\rm L} \\
b_3^{\rm E} &= 6(a_2^I + a_3^I) b_1^{\rm L} + 3(1 + 2a_2^I) b_2^{\rm L} + b_3^{\rm L} \\
&\vdots\,.
\end{aligned} \tag{1.141}$$

These bias parameters depend on the halo mass. For practical purposes, it is more convenient to work with the effective bias parameters, i.e. the bias values convolved with the halo mass function in a given mass range $I = (M_{\rm min}, M_{\rm max})$, usually $M_{\rm max} \to \infty$,

$$b_{{\rm eff},n} = \frac{\displaystyle\int_I {\rm d}M \frac{{\rm d}n_{\rm h}(M)}{{\rm d}M} b_n(M)}{\displaystyle\int_I {\rm d}M \frac{{\rm d}n_{\rm h}(M)}{{\rm d}M}}\,, \tag{1.142}$$

where $b_n(M)$ is the $n^{\rm th}$ bias parameter of the expansion, which can be Lagrangian or Eulerian.

1.3.5 Halos in N-body cosmological simulations

N-body simulations are numerical solvers of the non-collisional Vlasov equation given an initial phase space distribution. The continuous matter density field is discretized in particles with a specific mass, for which the initial position and velocity are known [see 26, for and extensive review]. Dark-matter halos coincide with the densest sites of the matter density field which do not have well-determined boundaries. The mass distribution surrounding their cores is continuous, for this reason there is always some degree of arbitrariness in defining the halo boundaries and its total mass. In numerical simulations, halos are identified using algorithms that implicitly also provide a definition of dark-matter halo. Historically, void finder algorithms can be divided into two classes [see 27, for a review].

The first class of halo finders shares the philosophy of the spherical overdensity (SO) algorithm [10, 28]. Procedurally, this class of algorithms is composed of two main steps. First, it identifies a density maximum after having smoothed the mass distribution, then it finds the radius R_Δ of the sphere centered in it which encloses a fixed density contrast $\Delta_{\rm SO}$. The halo is formed by the particles contained within this sphere. The density contrast is a free parameter and usually is fixed to $\sim$ 200 because this is the density contrast of a virialized object that undergoes to a quasi-spherical collapse in a matter dominated universe (see Sect. 2.1).

The second class of algorithms includes the ones that share the same philosophy of the friends-of-friends algorithm [FoF, 29–31]. This algorithm starts from a particle and finds all its neighbors within a fixed distance L, called friends. After that, it looks for friends of friends and keep iterating until no new friends are found. Each class of friends forms a halo. The length scale L is given in units of the mean particle separation (mps), i.e. $L = l \times$ mps. Thus, FoF halos correspond to regions for which the local density contrast is $\delta(\mathbf{x}) \geq l^{-3}$. Usually $l = 0.2$, corresponding to $\delta(\mathbf{x}) \geq 125$. Note that $\Delta_{\rm SO}$ is the density contrast integrated over the sphere with radius R_Δ, while in this case we consider the estimation of the local density contrast in each point of the FoF halo, i.e. $\delta(\mathbf{x})$ such that $\mathbf{x} \in$ halo.

These two algorithms remain the foundation of nearly every halo finder code: they often involve at least one phase where either particles are linked together, like in FoF, or shells are grown to collect particles, as in SO [27]. The FoF method, i.e. connecting and linking together particles that are close to each other, has also been extended in the 6D phase space [see e.g `ROCKSTAR`, 32], with an appropriate metric definition. These original halo finders have been improved at the expense of simplicity by combining them with algorithms that purge the halos by removing gravitationally unbound particles and algorithms that find substructures. Which halo finder is more appropriate to use depends on the particular application. In general, the FoF method better matches the mass distribution of halos with no symmetries, but tends to

spuriously link together multiple halos that are bridged by thin threads of particles. On the other hand, SO algorithms are expected to better correlate with galaxy clusters, especially for high $\Delta_{\rm SO}$ values, i.e. $\sim 400 - 500$. In any case, it is not obvious how to compare simulated halos with observations. Moreover, it is worth noting that the comparison between theoretical predictions of halo clustering and simulations assumes a halo finding algorithm with an implicit halo definition. In particular, the Sheth-Tormen multiplicity function is fitted to reproduce the FoF halos, while the Tinker et al. [33] model is calibrated to reproduce the SO halo mass function.

1.3.6 Galaxies in halos

As anticipated in the introduction to this section, the physics acting in galaxy formation is extremely complex, and a complete predictive theory of galaxy formation has not been established to date; nevertheless, it is possible to consider empirical connection between galaxies and halos, based on both observation and simulations. With the advent of wider and deeper redshift surveys, it became evident that there is no one-to-one connection between observed galaxies and halos described theoretically and with simulations. More specifically, the galaxy two-point correlation function at small scales is different in shape from the expected one of dark matter halos [34, 35]. This led to the development of the halo model paradigm, where an entire set of galaxies is associated with single dark-matter halos [36–41] [see 42, for a review]. Mathematically, multiple occupancy is described by the "halo occupation distribution" (HOD), which in its simplest version provides the probability $P(N_{\rm g}|M,z)$ of finding $N_{\rm g}$ galaxies in a halo of mass M at redshift z. Adopting the halo model paradigm [42], contributions to the clustering of galaxies can be split into two terms: the two-halo term, which considers correlations between galaxies belonging to different halos, and the one-halo term, considering galaxies of the same hosting halo. Moreover, in standard HOD models [41], galaxies are assumed to follow the matter distribution within their host dark matter halo, directly measured if we are dealing with cosmological simulation, or by using a universal mean mass profile, such as the Navarro-Frenk-White [NFW, 43]. Galaxies are generally divided into "central" and "satellite". Every halo hosting galaxies has one central galaxy, sitting in the halo barycenter or around it; all the other galaxies, if present, are satellite galaxies.

The galaxy number density can therefore be written as

$$n_{\rm g} = \int_M^\infty {\rm d}M \frac{{\rm d}n_{\rm h}(M)}{{\rm d}M} \langle N_{\rm c}(M)\rangle \left[1 + \langle N_{\rm s}(M)\rangle\right] , \tag{1.143}$$

where the term $\langle N_{\rm c}(M)\rangle \left[1 + \langle N_{\rm s}(M)\rangle\right]$ is a convenient way to express the expected number of galaxies hosted by a halo of mass M. Note that $N_{\rm c}$ can be 0 or 1, so if no central galaxy is present, satellite galaxies are not considered, whereas the number of hosted galaxies is $1 + N_{\rm s}$. It follows that the effective bias of galaxies can be written as

$$b_{\rm g,eff} = \frac{1}{n_{\rm g}} \int_M^\infty {\rm d}M \frac{{\rm d}n_{\rm h}(M)}{{\rm d}M} \langle N_{\rm c}(M)\rangle \left[1 + \langle N_{\rm s}(M)\rangle\right] b_{\rm h}(M) , \tag{1.144}$$

where $b_{\rm h}(M)$ is the halo bias.

Usually, the central galaxy term $N_{\rm c}$ is described to follow a step-like function, while $N_{\rm s}$ is considered to follow a Poisson statistic, for which the expectation value can be parameterized

by a power law with logarithmic slope [44],

$$\langle N_{\rm s}(M)\rangle = \begin{cases} \left(\dfrac{M - M_0}{M_1'}\right)^{\alpha} & M \geq M_0 \\ 0 & M < M_0\,, \end{cases} \tag{1.145}$$

where M_0 is the cut-off value of the halo mass, M_1' is the characteristic mass of halos hosting one satellite, and usually $\alpha \sim 1$. Then, to fix the model parameters, the mean number density and the projected correlation function of a HOD sample of N-body simulations are measured and compared with the data [45].

In recent years the HOD model has been expanded in the parameters and physical quantities on which galaxy distribution depends [see e.g. 46, and references therein]; moreover other methodologies has been explored to describe the connection between halos and galaxies, such as the sub-halo abundance matching techniques [SHAM, 47, 48].

1.4 The concordance model

Standard Cosmological Model

$$P_\Phi(k) = A_s \left(\frac{k}{k_0}\right)^{n_{\rm s}-1} k^3 \tag{1.146}$$

where

$$\langle \Phi(\mathbf{k})\Phi^*(\mathbf{k}')\rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k} - \mathbf{k}') P_\Phi(k)\,, \tag{1.147}$$

and $\delta^{(3)}(\mathbf{x})$ is the 3-dimensional Dirac delta. In analyses concerning the late Universe, $A_{\rm s}$ is exchanged with the σ_8 parameter, i.e. the root mean square of density fluctuations in spheres of 8 h^{-1}Mpc linearly evolved at present time. Both $A_{\rm s}$ and σ_8 are normalizations of the power spectrum, the first one fixes the amplitude of the initial conditions, the other fixes the amplitude of perturbation at present time, so they are in one-to-one correspondence. In the Standard Cosmological Model, inflation also causes a flat space-time, i.e. $\Omega_K = 0$, within a scale of the order of the comoving horizon before the inflation, which according to CMB measurements [49] is larger than the present comoving horizon.

Other parameters are the baryon energy density at present time $\Omega_{\rm b}$ and the total matter density at present time $\Omega_{\rm m}$, which corresponds to the sum on the energy density of baryons and cold dark matter, and more generally of each non-relativistic component. Another parameter is the Hubble constant at present time H_0. For historical reasons, H_0 is usually written as $H_0 = 100\,h\,{\rm Km\,s^{-1}Mpc^{-1}} = 3.236 \times 10^{-18} h\ {\rm s^{-1}} = 2.133 \times 10^{-33} h\ {\rm eV}\hbar$, and it is common to factorize the h parameter from observed quantities, such as masses and distances. The last parameter considered in this six-parameter model is the optical depth τ, due to CMB photons undergoing Compton scattering during their cosmic journey towards us. This last parameter does not explicitly enter into the ΛCDM model, but is needed to fit the ΛCDM model in CMB analyses [50]. Moreover, in principle, it can be derived by the other parameters, nevertheless, its physical origin depends on complex astrophysical processes difficult to model. Beyond these six parameters, the radiation density $\Omega_{\rm r}$ is crucial to predict the power spectrum of matter fluctuation, nevertheless its measure is so precise that it is considered fixed. All the parameters and quantities present in the Standard Cosmological Model can be derived from these 6+1

parameters. The term ΛCDM indicates the dark components considered in the model that are the cold, i.e. non-relativistic, dark matter and the cosmological constant Λ, that describes the late time acceleration of the Universe.

Although the Standard Model of Cosmology is quite successful, there are still several possible extensions, open questions, and discrepancies. An extension of ΛCDM widely explored in the literature is to consider the sum of neutrinos mass eigenvalues $\sum m_\nu$ as a free parameter [51–53]. To date, cosmology has provided the tightest constraints on the sum of neutrino masses [50].

The most important question that may extend the ΛCDM model concerns what causes the observed late time acceleration. In the Standard Cosmological Model this is described with the cosmological constant, i.e. an intrinsic property of the space-time, which value is fitted on data. However, there are other possible explanations. The first one is the vacuum energy of fundamental interactions, that can be described by an effective cosmological constant. Its value computed considering the known interactions is in tremendous disagreement with respect to observations; despite this discrepancy, a correct computation should consider all the particles and interactions existing in nature, which are unknown [see 54, for an extensive review]. Alternatively, the late time acceleration can be explored considering the presence of an exotic field minimally coupled with gravity, dubbed dark energy [see e.g 55, 56], or modifications of General Relativity, i.e. theories of modified gravity [57]. If the cosmological constant provides a time-independent equation of state $w = -1$, the equation of state of dark energy changes with time, as in inflationary cosmology. This equation of state can be effectively parameterized in order to measure various classes of dark energy models. The two most used parameterizations are the constant equation of state, with $w = \text{const} < -1/3$ and the Chevallier–Polarski–Linder (CPL) equation of state [58, 59],

$$w_{\rm CPL}(z) = w_0 + w_a(1-a) = w_0 + w_a \frac{z}{z+1}\,, \tag{1.148}$$

that considers a linear dependence with respect to the scale factor a.

The Standard Cosmological Model presents now a day also some discrepancies when it is used to measure early or late cosmological probes. The two main tensions refer to the H_0 and the $\Omega_{\rm m}$–σ_8 problems. The Hubble constant H_0 determined from early-Universe probes (CMB) appears to be in significant disagreement with respect to the estimates provided by late-Universe experiments. Even if many analyses addressed whether this might be due to some systematics hidden in either measurement, at the current status, this seems disfavored [60–70]. At the same time, differences also arise in the estimate of matter energy density $\Omega_{\rm m}$ and of σ_8 often summarized with the quantity $S_8 = \sigma_8 \sqrt{\Omega_{\rm m}/0.3}$ [71–83]. In the case these discrepancies are not attributable to some problems with the data, they may break the ΛCDM model.

Chapter 2

Cosmic voids in the excursion-set framework

The excursion-set mechanism offers a powerful framework for studying the statistical properties of halos and cosmic voids depending only on the linear statistical properties of matter. The fundamental idea of the excursion-set formalism relies on studying the statistical properties of halos or voids in the initial density field, which are consequently mapped in the fully nonlinear evolved density field, as introduced in Sect. 1.3. In the literature, the excursion-set framework is widely used to describe the clustering properties of dark-matter halos, whereas the treatment of cosmic voids is just at its beginning. In this chapter, based on unpublished work [84], we present an unprecedented modelization of void statistics in the excursion-set formalism, with the aim of describing from first principles the abundance of voids and their clustering properties, focusing on the void-void and void-halo correlation functions. Even if the treatment of voids in the excursion-set framework shows many analogies with respect to the halo case, it presents some important differences that we widely discuss. A theoretical description of void statistics from first principles translates into a predictive cosmological observable. This is particularly interesting considering that Large-Scale Structure (LSS) experiments are dominated by the analysis of the clustering properties of collapsed structures. In this picture, cosmic voids can provide a new point of view to tackle the open question of modern cosmology [85, 86]

This chapter is organized as follows. We present in detail void and halo formation (Sect. 2.1) and discuss the theoretical foundation of the excursion-set mechanism (Sect. 2.2); then we derive the stochastic field equations to describe halo and void formation and their statistical properties, providing some analytical solutions (Sect. 2.3); then we discuss numerical methods to solve them for the general case (Sect. 2.4.1); then we discuss the numerical results describing the clustering properties of voids and halos (Sect. 2.5).

2.1 The spherical model for halo and voids formation

The simplest dynamical model that approximately describes the formation of self-gravitating structures is the evolution of a spherical fluctuation in a homogeneous background, also known as the top-hat model. This toy model allows to give an insight on the gravitational evolution of a density perturbation from the linear to the strongly nonlinear regime. In order to study analytically the fully nonlinear gravitational evolution, spherical symmetry is assumed to reduce the degrees of freedom of the differential equations involved. Moreover, this simple model is

used to link the linear statistical properties of the matter density field with the formation of the LSS.

In General Relativity, Birkhoff's theorem [3, 87] states that a spherically symmetric region of space-time evolves independently of the surrounding space-time. This implies that for a generic spherically symmetric density perturbation with a monotonic density behavior, each concentric shell remains concentric and evolves like a closed ($\Delta > 0$) or open ($\Delta < 0$) FLRW universe, regardless of the background and outer shells. This allows us to treat the perturbation with the so called "separate-universe" approach [88–91]. Note that the Poission, continuity, and Euler equations introduced in Sect. 1.2.2, i.e. Eq. (1.77), Eq. (1.87), and Eq. (1.91), can be solved to describe the same problem. Nevertheless, the "separate-universe" approach is obtained with General Relativity without any approximation, so the result is more general and exact at any scale. We do not address possible extensions of this problem in modified theories of gravity, for which the treatment depends on the specific model assumed, or alternatively requires effective parameterizations [see e.g. 92–95]

Consider a spherically symmetric density perturbation with monotone density contrast with initial physical radius R_i and initial average density $\Delta_i = 3R_i^{-3}\int_0^{R_i}\delta(r)r^2\mathrm{d}r$ at some early time $t_i \ll 1/H_0$, such that $\Delta_i = \Delta(t_i) \ll 1$. The density enclosed within R_i is $\rho(t_i)(1+\Delta_i)$ and the mass is

$$M = \frac{4\pi R_i^3}{3}\rho(t_i)(1+\Delta_i)\,. \tag{2.1}$$

We assume a ΛCDM cosmology throughout; nevertheless, generalizations to include a dark energy component instead of a cosmological constant and a non-zero background curvature are straightforward [91]. The fluctuation evolves as a FLRW universe with cosmological parameters shifted with respect to the ones of the background, accordingly to $\Delta(t)$ quantity,

$$\tilde{\rho}_\mathrm{m}(t) = \rho_\mathrm{m}(1+\Delta(t))\,, \tag{2.2}$$

where t denotes the proper time of the comoving observer in the background universe. The other cosmological parameters are modified accordingly within the fluctuation

$$\tilde{\Omega}_\mathrm{m}\tilde{h}^2 = \Omega_\mathrm{m}h^2\,, \quad \tilde{\Omega}_\Lambda\tilde{h}^2 = \Omega_\Lambda h^2\,, \quad \tilde{\Lambda} = \Lambda\,, \tag{2.3}$$

where the tilde denotes the quantities within the fluctuation. Note that the cosmological constant does not depend on the environment, since according to the Einstein equations 1.22, it is a property of the space-time modifying the Einstein tensor, whereas it does not affect the stress-energy tensor. Note, however, that the density parameter associated with the cosmological constant varies within the fluctuation, according to the other quantities. The evolving radius of the spherical shell of the perturbation, i.e. $R(t)$, plays the role of the scale factor. It follows that, assuming mass conservation and taking the time derivative of Eq. (2.1), the expansion/contraction rate of the fluctuation reads

$$\frac{\dot{R}(t)}{R(t)} = \frac{\dot{a}(t)}{a(t)} - \frac{\dot{\Delta}(t)}{3[1+\Delta(t)]}\,, \tag{2.4}$$

while the corresponding first Friedmann equation is

$$\left(\frac{\dot{R}(t)}{R(t)}\right)^2 = \frac{8\pi G}{3}(1+\Delta(t))\rho_\mathrm{m}a(t)^{-3} + \frac{1}{3}\Lambda - \frac{\tilde{K}}{R(t)^2}\,. \tag{2.5}$$

The second Friedmann equation describing the dynamics of the fluctuation can be obtained from the time derivative of Equation (2.4), leading to the full nonlinear differential equation for $\Delta(t)$:

$$\ddot{\Delta} + 2H\dot{\Delta} - \frac{4}{3}\frac{\dot{\delta}^2}{1+\Delta} = 4\pi G\rho_{\rm m}(1+\Delta)\Delta\,. \tag{2.6}$$

To make the above equations to describe a valid FRLW dynamic, the curvature term $\tilde{K}$ has to be conserved [for demonstration and generalizations, see e.g. 91], so it is possible to evaluate it at an early time, when the cosmological constant contribute is negligible and the universe is matter dominated. In this regime $\Delta(t) \ll 1$ and the evolution is proportional to the scale factor (see Sect. 1.2.2), so

$$\begin{aligned} \Delta(t) &= \Delta_i \frac{a}{a_i} \\ \dot{\Delta} &= \Delta_i H \end{aligned} \qquad \text{for} \qquad t \ll 1/H_0 \tag{2.7}$$

Substituting these quantities in Equation (2.5) we get

$$\tilde{K} = \frac{5}{3}\Delta_i \frac{8\pi G}{3}\rho_{\rm m} a^{-3} R_i^2\,, \tag{2.8}$$

that can be rewritten as

$$\frac{\tilde{K}}{H_0^2} = \frac{5}{3}\Omega_{{\rm m},i}\Delta_i R_i^2\,, \tag{2.9}$$

where $\Omega_{{\rm m},i} = \Omega_{{\rm m},0}a_i^{-3}$. Alternatively, one can consider the fluctuation mass M Eq. (2.1) and obtain

$$\tilde{K} = \frac{5}{3}\frac{2GM}{R_i}\Delta_i\,, \tag{2.10}$$

with an explicit appearance of the Newtonian potential term. It is now possible to solve the differential equations that describe the evolution of a spherical perturbation. In the following, we will study the evolution of a top-hat perturbation in a flat matter dominated universe. This choice is for two reasons. First, in the matter dominated universe the above equations are analytically solvable, allowing to give an insight on the nonlinear dynamics. Second, the main quantity we are interested in, i.e. the map between linear and nonlinear density contrast evolution, is almost insensitive to cosmology [96, 97]. This map derived in the matter dominated universe accurately approximates the one for ΛCDM and its extension (see Sect. 1.4).

In the matter dominated universe Eq. (2.5), after having subtracted the contribution of the background, becomes

$$\left(\frac{\dot{R}(t)}{R(t)}\right)^2 = H_i^2\left[(1+\Delta_i)\left(\frac{R_i}{R(t)}\right)^3 - \frac{5}{3}\Delta_i\left(\frac{R_i}{R(t)}\right)^2\right], \tag{2.11}$$

where $H_i = H(t_i)$, which admits a parametric solution. To obtain this equation, we have considered mass conservation, expressed by the equivalence

$$1 + \Delta(t) = (1+\Delta_i)\left(\frac{R_i}{R(t)}\right)^3\,. \tag{2.12}$$

Let us consider the parameter η,

$$\mathrm{d}\eta = \frac{R_i}{R}\sqrt{\left|\frac{5}{3}\Delta_i\right|}H_i \mathrm{d}t\,, \tag{2.13}$$

that allows to rewrite Eq. (2.11) in the parametric form

$$\frac{R(\eta)}{R_i} = \begin{cases} \frac{1}{2}\left|\frac{5}{3}\Delta_i\right|^{-1} (\cosh\eta - 1) & \Delta_i < 0 \\ \frac{1}{2}\left|\frac{5}{3}\Delta_i\right|^{-1} (1 - \cos\eta) & \Delta_i > 0\,, \end{cases} \tag{2.14}$$

and

$$H_i t(\eta) = \begin{cases} \frac{1}{2}\left|\frac{5}{3}\Delta_i\right|^{-3/2} (\sinh\eta - \eta) & \Delta_i < 0 \\ \frac{1}{2}\left|\frac{5}{3}\Delta_i\right|^{-3/2} (\eta - \sin\eta) & \Delta_i > 0\,. \end{cases} \tag{2.15}$$

These are the well known solutions of the Friedmann equations for closed and open universes containing matter only [3]. The evolution of the density contrast can be obtained by considering

$$\tilde{\rho}_{\rm m}(t) = \rho_{\rm m}(t)(1 + \Delta(t)) = \frac{3}{4\pi}\frac{M}{R(t)^3} \tag{2.16}$$

and expressing the background density evolution as a function of time $\rho_{\rm m}(t) = 1/(6\pi G t^2)$ (see Sect. 1.2.2), obtaining

$$1 + \Delta(\eta) = \frac{9}{2}\frac{GMt^2}{R(t)^2} = \begin{cases} \frac{9}{2}\frac{(\sinh\eta - \eta)^2}{(\cosh\eta - 1)^3} & \Delta < 0 \\ \frac{9}{2}\frac{(\eta - \sin\eta)^2}{(1 - \cos\eta)^3} & \Delta > 0 \end{cases}. \tag{2.17}$$

This result can be generalized for the evolution of a spherical perturbation in a background universe with matter and non-vanishing curvature [see e.g. Appendix A in 98].

The peculiar velocity of the spherical shell is given by

$$v_{\rm p} = v - HR(t)\,, \tag{2.18}$$

where

$$v = \frac{{\rm d}R(t)}{{\rm d}t} = \frac{{\rm d}R(t)}{{\rm d}\eta}\frac{{\rm d}\eta}{{\rm d}t}\,. \tag{2.19}$$

Substituting Eq. (2.11) into this equation, we obtain

$$v_{\rm p} = HR\left[\frac{3}{2}g(\eta) - 1\right], \tag{2.20}$$

where

$$g(\eta) = \begin{cases} \frac{\sinh\eta\,(\sinh\eta - \eta)}{(\cosh\eta - 1)^2} & \Delta < 0 \\ \frac{\sin\eta\,(\eta - \sin\eta)}{(1 - \cos\eta)^2} & \Delta > 0\,. \end{cases} \tag{2.21}$$

The above equations describe the full nonlinear dynamics of a spherical fluctuation. Note that if we consider the corresponding linear dynamics, it is possible to map the nonlinear evolution into the corresponding linear one and *vice versa*. In the matter dominated universe

the linear growth of perturbations is proportional to the scale factor, $\Delta_{\rm L}(a) = \Delta_i a/a_i$, and $a/a_i = (\frac{3}{2}H_i t)^{2/3}$. Substituting the parametric expression for the time of Eq. (2.15), we obtain

$$\Delta_{\rm L}(\eta) = \begin{cases} \dfrac{3}{20}\left[6(\sinh\eta - \eta)\right]^{2/3} & \Delta_i < 0 \\ \dfrac{3}{20}\left[6(\eta - \sin\eta)\right]^{2/3} & \Delta_i > 0\,. \end{cases} \tag{2.22}$$

In this way, considering the evolution of the linear and nonlinear density contrast at the same η, it is possible to recover the linear density contrast value corresponding to the nonlinear one and *vice versa*. Therefore, we can map the nonlinear halo and void formation in linear theory.

2.1.1 The spherical collapse model for halo formation

Let us rewrite the nonlinear dynamics of overdensity to express it in a form that emphasizes the crucial events occurring in spherical halo formation. The second Friedmann equation for the spherical perturbation is

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\rho_{\rm m,i}(1+\Delta_i)\left(\frac{R_i}{R}\right)^3\,, \tag{2.23}$$

via Equation (2.1) it can be rewritten as

$$\frac{{\rm d}^2 R}{{\rm d}t^2} = -\frac{GM}{R^2}\,, \tag{2.24}$$

making an explicit link to Newtonian mechanics. Integrating once the equation of motion over time, we obtain the energy conservation equation

$$\frac{1}{2}\dot{R}^2 - \frac{GM}{R} = \text{const} = E\,. \tag{2.25}$$

Note that $GM = H_i^2 R_i^3(1+\Delta_i)/2$, therefore, considering Eq. (2.11), we can express $\dot{R}_i = H_i R_i(1-\Delta_i/3)$, and the conservation of energy allows to evaluate it at t_i

$$E = -\frac{3}{5}\frac{(H_i R_i)^2}{2}\Delta_i\,. \tag{2.26}$$

The evolution of the fluctuation's radius is given by Equation (2.11), which describes a cycloid. The maximum radius is reached at $\eta = \pi$, where $\dot{R}$ vanishes. It follows that

$$E = -\frac{GM}{R_{\rm max}} \qquad \Rightarrow \qquad R_{\rm max} = \left(1 + \frac{1}{\Delta_i}\right)R_i\,. \tag{2.27}$$

At $\eta = 2\pi$, the solution of Eqs. (2.11) and (2.17) formally collapses to a point, $R \to 0$ and $\Delta \to \infty$. In practice, however, a small deviation from sphericity is enough to break the symmetries, leading the perturbation to a non-collisional collapse and virialization [99]. The dimension and density contrast of the virialized object can be calculated using the virial theorem $E = T + W = W/2$, where T is the kinetic energy, while W is the potential energy. It follows that $R_{\rm vir} \simeq R_{\rm max}/2$. Assuming that $t_{\rm vir} = t(\eta = 2\pi)$, this implies that $t_{\rm vir} = 2t_{\rm max}$. Note that $R_{\rm vir}$ and

$R_{\rm max}$ are the physical radii of the fluctuation, the corresponding comoving radius is obtained by dividing these quantities by the scale factor a. It follows that the density at virialization is

$$1 + \Delta_{\rm vir} = \frac{9\pi^2}{16}\left(\frac{R_{\rm max}}{R_{\rm vir}}\right)^3\left(\frac{\rho_{\rm m}(t_{\rm max})}{\rho_{\rm m}(t_{\rm vir})}\right) = \frac{9\pi^2}{16} \times 8 \times 16 \simeq 178\,. \tag{2.28}$$

When an object starts to virialize, the fluid elements are no more in the single stream regime but various streams cross each other: the shell crossing condition is reached.

The above equations have several consequences in describing the spherical halo formation model. The result of Eq. (2.28) is the reason why the virialized region of a cluster is considered as the sphere with average density ~ 200 times the critical density of the universe, moreover this is why $\Delta \sim 200$ is the typical value used in spherical halo finders, described in Sect. 1.3.5. The second point concerns the radius. In all previous computations, R is the physical radius of the fluctuation, whereas Eq. (2.28) gives a simple method to link the comoving radius of an halo to the dimension of the primordial patch: $R_i^{\rm c} = R_{\rm halo}^{\rm c} \times 178^{1/3} \simeq R_{\rm halo}^{\rm c} \times 5.6$, where the superscript stands for comoving. The last important point is that we can map the nonlinear evolution and virialization in linear theory. In particular, the linear density contrast corresponding to full collapse can be obtained by Eq. (2.22) evaluated at the virialization time, $\eta = 2\pi$,

$$\delta_{\rm c} = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} = 1.68647\,. \tag{2.29}$$

Note that to recover the notation commonly used in the literature, we use $\delta_{\rm c}$ to indicate the linear collapsing threshold instead of $\Delta_{\rm c}^{\rm L}$, which in this section is used to describe the mean linear density contrast within a shell. This quantity is the critical threshold for spherical collapse, and it is almost cosmology independent. This threshold plays a central role in models that use the statistics of the initial density fluctuation field to describe the clustering of virialized objects, such as in the excursion-set framework. The power of this simple model is that the linear theory can state when spherical collapse will eventually occur. The initial density fluctuation can be linearly extrapolated at any initial density fluctuation $\Delta_i(t_i)$,

$$\Delta_{\rm L} = \Delta_i D(t) = \Delta_i \left(\frac{t}{t_i}\right)^{2/3}\,. \tag{2.30}$$

This region will collapse and virialize at a time $t_{\rm vir}$ such that $\Delta_{\rm L}(t_{\rm vir}) = \delta_c$. This expression can be generalized to any cosmological model substituting the corresponding expression for the linear growth factor $D(t)$, instead of the $(t/t_i)^{2/3}$ term, i.e. the linear growth factor of the matter dominated universe.

2.1.2 The spherical model for void formation

Void formation in the excursion-set framework, and more generally in frameworks that use statistics of the initial density fluctuation field to describe the statistical properties of evolved objects, greatly differs from halo formation. This is due to the fact that the evolution of a spherical underdensity differs from the overdense counterpart; consequently, the events characterizing their evolution are not the same. An overdense fluctuation starts evolving slower than the background, it reaches its maximal physical size and then begins to contract until it collapses into a virialized object. This evolution is characterized by two main events, turn around

and full collapse, to which correspond two specific linear density contrast values. In particular, the spherical halo formation model provides the linear threshold value corresponding to a full collapse, allowing to map virialized objects, i.e. regions that underwent shell crossing, in the initial linear density field. The evolution of a generic underdense spherical fluctuation is not characterized by any significant event. It starts to grow faster with respect to the background and it continues its outward expansion forever. Nevertheless, Sheth & van de Weygaert [98] showed that under particular conditions of the initial density profile, shell crossing can also occur for spherical underdensities.

Shell crossing in the evolution of spherical underdensities can be qualitatively described as follows. The outward peculiar acceleration of of a spherical shell is directly proportional to the integrated density deficit $\Delta(r)$ of the fluctuation. If $\Delta(r)$ increases monotonically, the inner shells feel a stronger deficit, and thus a stronger outward acceleration than the outer shells. It follows that inner shells grow faster than the outer ones. If the density profile is steep enough, shells start to accumulate in the outer part of the fluctuation. Shells that were initially close to the center will ultimately reach shells further outside until they eventually cross them: this marks the event of shell crossing. The corresponding gradual increase in density would turn into an infinitely dense ridge. This condition is satisfied for a step-function initial density profile near the edge of a top-hat void [96, 98].

Consider a family of trajectories $\Delta(R_i, \eta_{R_i})$ labeled by the initial radius R_i and the corresponding parameter η_{R_i} defined in Eq. (2.15). We consider the step function $\delta(R_i) = \delta_0 \Theta_{\rm H}(R_i - R_0)$ as the initial density profile, where $\Theta_{\rm H}(x)$ is the Heaviside step function. The corresponding integrated initial density profile is

$$\Delta_i(R_i) = \begin{cases} \delta_0 & R_i \leq R_0 \\ \delta_0 (R_0/R_i)^3 & R_i > R_0 \,. \end{cases} \tag{2.31}$$

To find where and when shell crossing first occurs, we differentiate the family of trajectories $\binom{R}{H_i t}$ described by Eqs. (2.14) and (2.15) with respect to the parameters R_i and η_{R_i} and we look for vanishing solutions. For $R_i > R_0$, Δ_i depends on R_i in such a way that $R_i \mathrm{d}\Delta_i/\mathrm{d}R_i = -3\Delta_i$, while its derivative is 0 for $R_i < R_0$. It follows that the shell crossing equation is

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathrm{d}R_i/R_i \\ \mathrm{d}\eta \end{bmatrix} = 0 \,, \tag{2.32}$$

where

$$\begin{aligned} A_{11} &= 2 \left| \frac{5}{3} \Delta_i \right|^{-1} (\cosh\eta - 1) \,, \\ A_{12} &= \frac{1}{2} \left| \frac{5}{3} \Delta_i \right|^{-1} \sinh\eta \,, \\ A_{21} &= \frac{9}{4} \left| \frac{5}{3} \Delta_i \right|^{-\frac{3}{2}} (\sinh\eta - \eta) \,, \\ A_{22} &= \frac{1}{2} \left| \frac{5}{3} \Delta_i \right|^{-\frac{3}{2}} (\cosh\eta - 1) \,, \end{aligned} \tag{2.33}$$

for $R_i > R_0$. To have a non-zero solution, we must have $\det A = 0$, from this condition follows

$$\frac{\sinh\eta\,(\sinh\eta - \eta)}{(\cosh\eta - 1)^2} = \frac{8}{9} \,. \tag{2.34}$$

Solving this equation, we find that shell crossing first occurs at $\eta_{sc} = 3.488$ for $R_i = R_0$. The density contrast corresponding to shell crossing is

$$\Delta_{\rm sc} = \frac{9}{2}\frac{(\sinh\eta_{\rm sc} - \eta_{\rm sc})^2}{(\cosh\eta_{\rm sc} - 1)^3} - 1 = -0.7953\,, \tag{2.35}$$

to which corresponds a liner density contrast of $\Delta^{\rm L}_{\rm sc} = -2.717$. Assuming mass conservation, when a spherical fluctuation reaches this level of underdensity, the comoving void radius is expanded by a factor of $(1 + \Delta_{\rm sc})^{-1/3} = 1.697$.

In the literature [96, 98] the shell crossing in void evolution was usually considered as the event characterizing void formation. Nevertheless, we state that there is no need for this parallelism between void and halo formation in the excursion-set framework. We now discuss that considering shell crossing in void formation and evolution is substantially irrelevant and practically useless. First of all, even if shell crossing can in principle occur in spherical void evolution, this event remains substantially different from the full collapse of overdensities. Contrary to halos, the shell crossing occurrence and the epoch at which it possibly happens strictly depend on the initial density profile of the fluctuation. Therefore, the linear density contrast corresponding to this event is not universal as the collapsing threshold for halo formation, and thus it has to be determined case by case. Moreover, shell crossing is unlikely to occur for more realistic density profiles in the initial density field. For the steepest possible profile, described by a step function, shell crossing occurs at the radius of the discontinuity of the initial density profile when the nonlinear density contrast reaches $\Delta \simeq -0.8$, corresponding to the linear value $\Delta_{\rm L} \simeq -2.7$. Hence, we can consider this value as the upper density contrast limit at which shell crossing eventually occurs in the general case. Moreover, it is interesting to note that the linear density contrast required to reach shell crossing in spherical underdensities is much higher in module with respect to the collapsing case, 1.686 against ≤ -2.717. This means that an initial underdensity (with a step density profile) requires much more time to reach shell crossing than overdensities with the same initial amplitude. In the matter dominated universe, the growth factor $D(a) \propto a \propto t^{2/3}$, therefore, the corresponding shell crossing epochs are related by $a_{\rm v}^{\rm sc} \simeq 1.6 \times a_{\rm h}^{\rm sc}$ and $t_{\rm v}^{\rm sc} \simeq 2 \times t_{\rm h}^{\rm sc}$. In addition, it is worth noting that these calculations are exact for spherical fluctuations in a homogeneous background; qualitative arguments show that in the collapsing case the breaking of symmetries leads to virialization, on the other hand, it has not been explored what this entails for void formation, but reasonably the shell crossing would require even harsher conditions to occur. More importantly, the formal derivation of the shell crossing condition in underdensities shows that this event strictly depends on the presence of a discontinuity in the initial density profile, which is unphysical for real fluctuations.

A second point in considering the shell crossing in void evolution as the event at which a void is considered formed concerns the detection and practical treatment of voids and halos for cosmological purposes and their link to the initial density field. Detecting halos means to find virialized objects. This gives a physical meaning to the linear collapsing threshold in the theoretical description of the statistical properties of halos, since this threshold maps the virialization event in linear theory: an halo is a virialized object, the density contrast for collapse identifies the regions in the initial density field that will virialize in a specific epoch. The void case is different, initial underdense regions reaching the linear threshold for void formation evolve in larger underdensities, i.e. voids are extended objects. A nonlinear density contrast of ~ -0.8, i.e. the upper limit at which shell crossing possibly occurs, would correspond to small regions in the core of the deepest underdensity basins in the matter distribution. In galaxy

surveys, a so negative threshold cannot be traced by galaxies, therefore it is not observable, neither in the galaxy distribution of the Universe nor in the halo distribution of simulations. In other words, the shell crossing event is useless in identifying voids, and a perfect parallelism between halo and void formation cannot be considered.

The main conclusion that we can infer from these arguments is the following: shell crossing is the event that breaks the map between linear and nonlinear theory; nevertheless cosmic voids do not experience shell crossing, at least at the scales where they are observable. This ensures that a map from the linear Lagrangian threshold δ_v[1] to the nonlinear Eulerian density contrast of voids always exists, at least for linear density contrast threshold $\delta_v \geq -2.717$. This condition ensures that the density profile of evolved voids (detected in various tracer distributions) can always be mapped in the linear theory. In this perspective, the linear void formation threshold δ_v can assume any negative value and does not correspond to any specific event in void evolution. The major difference between halo and void formation can be summarized as follows. A halo is formed when a region undergoes full collapse and virialization, this fully nonlinear event can be mapped to linear theory with the collapsing threshold δ_c. A void is a region that, in its evolution, has reached a fixed mean density contrast value. The corresponding linear threshold value does not refer to any specific event; nevertheless, since shell crossing has not occurred (at least for $\Delta(R) \geq -0.7953$), it is always possible to map the Eulerian density contrast of voids into the linear theory.

2.2 Excursion-set approach: building blocks

The fundamental quantity of the excursion-set formalism is the smoothed, linearly evolved initial density field at comoving coordinate $\mathbf{q}$

$$\begin{aligned}\delta(\mathbf{q}, R) &= \int \mathrm{d}^3 x W(|\mathbf{x}|, R)\delta(\mathbf{q}+\mathbf{x}) \\ &= \int \frac{\mathrm{d}^3 k}{(2\pi)^3} W(kR)\delta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{q}}\end{aligned} \tag{2.36}$$

where $W(|\mathbf{x}|, R)$ and $W(kR)$ are the smoothing filter in configuration and Fourier space, respectively. We recall that the linearly evolved initial density field is commonly called Lagrangian space, while the fully evolved system in comoving coordinates is called Eulerian space. The Lagrangian density contrast field $\delta(\mathbf{q})$ can be extrapolated at the desired redshift via the linear growth factor, in this way the redshift dependence can be factorized out via the identity $\delta(\mathbf{q}, z) = D(z)\delta(\mathbf{q})$, with normalization $D(z = 0) = 1$. The statistical properties of the smoothed density field $\delta(\mathbf{q}, R)$ are entirely specified by its correlation functions. In particular, we assume that the initial density contrast field follows Gaussian statistics, so that the linear matter power spectrum extrapolated to the present epoch $P(k)$ encodes the entire statistics. Note that according to the latest measurements of primordial non-Gaussianity [100], the assumption of a Gaussian initial density field is an accurate description of its actual statistical properties. The key assumption of the excursion-set formalism for halo and void statistics is that at a given redshift z, a Lagrangian point $\mathbf{q}$ belongs to a halo (void) of size R if R is the maximum smoothing scale at which the smoothed linear density contrast $\delta(\mathbf{q}, R)$ exceeds the linear extrapolated

[1] As for the halo case, to recover the notation commonly used in the literature, we use the notation δ_v to indicate the linear threshold of void formation instead of Δ_v^L used thorough this section to describe the mean linear density contrast within a shell.

Filter	Configuration	Fourier
Top-hat	$\frac{3}{4\pi R^3}\Theta_{\rm H}(1-x/R)$	$3\frac{j_1(kR)}{kR}$
Gaussian	$\frac{e^{-\frac{1}{2}x^2/R^2}}{(2\pi R^2)^{3/2}}$	$e^{-\frac{1}{2}x^2k^2}$
Sharp-k	$\frac{3}{4\pi R^3}\left[3\frac{j_1(x/R)}{x/R}\right]$	$\Theta_{\rm H}(1-kR)$

Table 2.1: The three most popular filtering kernels in configuration space and their Fourier transform.

critical density contrast $\delta_{\rm c}$ ($\delta_{\rm v}$). The voids case has the additional condition that the filtered field that reached $\delta_{\rm v}$ at a given smoothing length has not crossed the collapsing threshold at larger scales.

The detailed properties of $\delta(\mathbf{q}, R)$ depend on the choice of the specific filter function. Table 2.1 lists the most commonly used smoothing filters. The spherical top-hat filter is the one that provides the best possible link to the spherical model of halo and void formation discussed in Sect. 2.1. This filter is a top-hat in real space, so it provides a well defined boundary of the patch considered and consequently a well defined mass enclosed within the filter, i.e. in the fluctuation. The corresponding mass is $M_{\rm TH} = 4\pi R^3\rho_{\rm m}$ / 3, where both R and ρ are comoving quantities, i.e. $\rho_{\rm m} = \rho_{\rm m}(z = 0)$. The disadvantage of this filter is that its Fourier transform is numerically heavy to handle, since it is a slow decreasing oscillating function: $W_{\rm TH}(kR) = 3j_1(kR)/kR$, where $j_1(x) = (\sin x - x\cos x)/x^2$ is the spherical Bessel function of order 1. The Gaussian filter is commonly used too, because it is numerically more stable than the top-hat filter and it simplifies some analytical computations. On the other hand, the boundary of the fluctuation is not precise, the outer part weights less and less without a truncation. The mass is computed as the spherical integral of the Gaussian kernel multiplied by the mean comoving mass density, obtaining $M_{\rm G} = (2\pi)^{3/2}\rho_{\rm m}R^3$. The last filter we consider is the sharp-k filter, that is a top-hat in Fourier space. It became popular because it greatly simplifies many computations, as we will see later. On the other hand, its physical interpretation is not clear, since in configuration space the sharp-k filter is a slow decreasing oscillating function that takes both positive and negative values, therefore, the amount of mass enclosed in the filter is ill defined. One possibility to evaluate it is to compute the spherical integral of the filter and multiply it by the mean matter density, obtaining $M_{\rm sk} = 6\pi^2\rho_{\rm m}k^{-3}$ [101, 102]. An alternative procedure consists of defining $M_{\rm sk} = 4\pi R_{\rm TH}^3\rho_{\rm m}/3$, where $R_{\rm TH}$ is fixed requiting $\sigma^2_{\rm sk}(k) = \sigma^2_{\rm TH}(R_{\rm TH})$. Here $\sigma^2_{\rm sk}(k)$ indicates the variance of the density field smoothed with a sharp-k filter, and $\sigma^2_{\rm TH}(R_{\rm TH})$ is the same quantity computed using the top-hat filter. This definition introduced by Bardeen et al. [18], is implicitly used in the great majority of later works on the excursion-set and in all theoretical frameworks that use the statistics of the initial density fluctuation field to describe the statistical properties of evolved objects [see e.g. 15, 96, 98, 102].

The excursion-set formalism describes halo and void formation and their statistics from the trajectory in the (δ, R) plane given by the density contrast field $\delta(\mathbf{q}, R)$ filtered at decreasing R scales. We start from an infinitely large smoothing radius around a given point $\mathbf{q}$ in Lagrangian

space. Since the density averaged over this region is the mean density of the universe, the density contrast must vanish as $R \to \infty$. We then gradually decrease the smoothing radius and follow the trajectory $\delta(\mathbf{q}, R)$ in the (δ, R) plane. We consider that the Lagrangian point $\mathbf{q}$ belongs to a halo of mass $M(R_\times)$ if the trajectory of $\delta(\mathbf{q}, R)$ first crosses the halo formation threshold δ_c at $R_\times$. The behavior of the trajectory at $R < R_\times$ does not matter in halo formation, since any scale smaller than the first crossing scale will collapse inside the halo. So, even if the trajectory would decrease under the threshold value and then up-crosses the threshold again, we do not consider this second cross as the formation of another halo. In this way, the excursion-set mechanism solves the "cloud-in-cluod" problem [see Sect. 1.3.1, 11, 12]. Void formation is analogous, but involves two thresholds, the threshold of void formation δ_v and the threshold of collapse δ_c. We consider that the point $\mathbf{q}$ belongs to a void with Lagrangian radius $R_\times$, if the trajectory $\delta(\mathbf{q}, R)$ first crosses the void formation threshold at $R_\times$ without having crossed the collapsing threshold at any larger scale. The reason for this additional condition is that if the trajectory δ_c reaches the void threshold after having crossed the collapsing threshold, the underdense fluctuation would be located within a larger overdensity that will collapse, squeezing the initial underdensity in a virialized halo. This case is called "void-in-cloud", in analogy to the "cloud-in-cloud" [98]. Even for voids, the behavior of the trajectory for $R < R_\times$ does not have any impact on void formation in the excursion-set framework; it describes the density contrast within the void at smaller scales, centered in $\mathbf{q}$.

Formally, halo and void formation in the excursion-set framework is described by an unconditional and conditional first crossing problem, respectively, of the path $\delta(\mathbf{q}, R)$, for which the stochastic evolution depends only on its statistical properties.

2.3 Stochastic evolution of the Lagrangian density contrast field

The mechanism introduced qualitatively in the previous subsection is described formally by the Langevin equation [12, 101–103]. In this section we will introduce the Langevin equations that describe the evolution of the single field $\delta(\mathbf{q}, R)$, Sect. 2.3.1, and of the spatially correlated pair $\delta(\mathbf{q_1}, R_1)$ and $\delta(\mathbf{q_2}, R_2)$, Sect. 2.3.2.

2.3.1 The Langevin equation in the single field case

In studying the 0^{th} order statistics of the density contrast field, the exact position of the field is not important, as a consequence of the background homogeneity of the universe, so we will omit to write the explicit dependence and simply use $\delta(R) = \delta(\mathbf{q}, R)$. The effect of varying the smoothing scale R can be obtained by differentiating Eq. (2.36):

$$\frac{\partial \delta(R)}{\partial R} = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{\partial W(kR)}{\partial R} \delta(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{q}} = Q(R)\,. \tag{2.37}$$

This has the form of the Langevin equation, which shows how an infinitesimal change of the scale R affects the smoothed field as a function of the stochastic force $Q(R)$. The initial condition of this first-order stochastic differential equation is given by

$$R \to \infty\,, \qquad \delta(R) \to 0\,. \tag{2.38}$$

Since Eq. (2.37) is linear, the stochastic force is also a Gaussian random field with a vanishing expectation value, $\langle Q(R)\rangle = 0$, it is uniquely described by its correlation function, so the Langevin equation is fully described by the system [102]

$$\begin{cases} \dfrac{\partial\delta(R)}{\partial R} = \displaystyle\int \frac{\mathrm{d}^3k}{(2\pi)^3}\frac{\partial W(kR)}{\partial R}\delta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{q}} = Q(R) \\ \langle Q(R_1)Q(R_2)\rangle = \dfrac{1}{2\pi^2}\displaystyle\int_0^\infty \mathrm{d}k\, k^2 P(k)\frac{\partial W(kR_1)}{\partial R_1}\frac{\partial W^*(kR_2)}{\partial R_2}\,. \end{cases} \tag{2.39}$$

The coherence of each trajectory along R depends only on the form of the filter function. This coherence vanishes for the sharp-k filter. We now focus on this filter, since it provides simplifications that allow to solve the Langevin equation analytically, giving a useful insight on how the excursion-set mechanism works.

With sharp-k filter, the stochastic force at each step is independent of all the previous ones: decreasing the smoothing scale R a new set of Fourier modes of the unsmoothed distribution is added to the filtered field $\delta(R)$, consequently the stochastic evolution of the filtered field is Markovian

$$\frac{\partial W(kR)}{\partial R} = \frac{\partial\Theta(1-kR)}{\partial R} = -\frac{k}{R}\delta_\mathrm{D}(k-1/R)\,, \tag{2.40}$$

where $\delta_\mathrm{D}(x)$ is the Dirac delta. We can use the variance of the filtered density field as the time variable, since it is a monotonic function of R,

$$S = \sigma^2(R) = \langle\delta(R)^2\rangle = \frac{1}{2\pi^2}\int_0^{1/R}\mathrm{d}\,k^2P(k) \tag{2.41}$$

obtaining

$$\begin{cases} \dfrac{\partial\delta(S)}{\partial S} = Q(S) & \langle Q(S)\rangle = 0 \\ \langle Q(S_1)Q(S_2)\rangle = \delta_\mathrm{D}(S_1-S_2)\,. \end{cases} \tag{2.42}$$

The solution of the Langevin equation Eq. (2.42) at any arbitrary point of space, with the initial condition $\delta(S=0)=0$, is $\delta(S) = \int_0^S \mathrm{d}S'\,Q(S')$. The properties of the stochastic realization of this process are those of the Wiener processes [104, 105]:

$$\langle\delta(S)\rangle = 0 \qquad \langle(\delta(S_1)\delta(S_2)\rangle = \min(S_1,S_2) \tag{2.43}$$

i.e., the field $\delta(S)$ can be described as a Markovian random-walk, where S is the time variable.

We are now interested in solving the first crossing distribution of the threshold δ_c, due to its link to halo formation. In the sharp-k filter case, this distribution is analytical and was first found by Chandrasekhar [106]. The Langevin equation is always associated with a Fokker-Planck equation for the probability density $\Pi(\delta,S)$,

$$\frac{\partial\Pi(\delta,S)}{\partial S} = \frac{1}{2}\frac{\partial^2\Pi(\delta,S)}{\partial\delta^2}\,. \tag{2.44}$$

with the initial condition $\Pi(\delta,S=0)=\delta_\mathrm{D}(0)$ and the boundary condition $\Pi(\delta_\mathrm{c},S)=0$. The solution to this equation is [106]

$$\Pi(\delta,S) = \begin{cases} \dfrac{1}{\sqrt{2\pi S}}\left[e^{-\delta^2/2S} - e^{-(2\delta_\mathrm{c}-\delta)^2/2S}\right] & \delta<\delta_\mathrm{c} \\ 0 & \delta\geq\delta_\mathrm{c}\,. \end{cases} \tag{2.45}$$

This solution is known as the differential survival probability, since it describes the probability distribution of random walks (δ, S) that do not cross the barrier δ_c. This solution can be found using the image method. The density distribution of trajectories (δ, S) given by Equation (2.44) in the unconstrained case is a Gaussian with variance equal to the time variable S. To obtain the differential survival probability at time S, we must subtract all the paths that crossed the threshold δ_c at $S' < S$. To do that, consider a path that arrives in (δ, S) after having crossed the barrier δ_c at $S' < S$. We can construct a path that at time $> S'$ is the reflection of the original one with respect to the barrier δ_c, i.e. $(S, \delta_c - (\delta - \delta_c))$ and subtract it. This means that to take into account the absorbing barrier effect, we have to subtract a Gaussian centered in $2\delta_c$ with the time variable S as variance, which is exactly Eq. (2.44). Therefore, the survival probability is given by integrating Eq. (2.44), and the fraction of trajectories that hit the barrier at scales $< S$ is

$$F_c(< S) = 1 - \int_{-\infty}^{\delta_c} \mathrm{d}\delta \Pi(\delta, S) = \mathrm{erfc}\left[\frac{\delta_c}{\sqrt{2S}}\right]. \tag{2.46}$$

Deriving this quantity, we obtain the density probability distribution of the first crossing, i.e. the multiplicity function [103]:

$$f_c(S) = -\frac{\partial F_c(< S)}{\partial S} = \frac{\delta_c}{\sqrt{2\pi S^3}} e^{-\delta_c^2/2S}\,. \tag{2.47}$$

The multiplicity function is a differential quantity, so to change the variable, we consider

$$f_c(S)\mathrm{d}S = f_c(\sigma)\mathrm{d}\sigma = f_c(\nu)\mathrm{d}\nu = f_c(M)\mathrm{d}M = f_c(R)\mathrm{d}R\,. \tag{2.48}$$

Substituting S with $\nu_c = \delta_c/S$ we can express it in the more familiar form

$$f_c(\nu_c) = \sqrt{\frac{2}{\pi}} e^{-\nu_c^2/2}\,, \tag{2.49}$$

that is exactly what we found in Eq. (1.107) for the Press-Schechter model, without the need to correct the normalization: the "cloud-in-cloud" problem is solved by the excursion-set mechanism (see Sect. 1.3.1).

As we described in the previous subsection, void formation can be treated as a double barrier problem. The corresponding multiplicity function can be derived from the single barrier case [98], and can be written as

$$f(S, \delta_v, \delta_c) = f(S, \delta_v) - \int_0^S \mathrm{d}s f(s, \delta_c, \delta_v) f(S, \delta_v | s, \delta_c) \tag{2.50}$$

where the first term on the right-hand side is the first crossing distribution of the barrier δ_v, and the second term subtracts from it the subset of trajectories that have crossed δ_c before having reached δ_v. Note that since we are considering a Markovian process, this term can be rewritten as $f(S, \delta_v | s, \delta_c) = f(S - s, \delta_c - \delta_v)$. The Laplace transform of Eq. (2.50) is

$$\begin{aligned} L(t, \delta_v, \delta_c) &= L(t, \delta_v) - \int_0^\infty \mathrm{d}s f(s, \delta_c, \delta_v) e^{-ts} \int_{S-s}^\infty \mathrm{d}S\, f(S - s, \delta_c - \delta_v) e^{-t(S-s)} \\ &= L(t, \delta_v) - L(t, \delta_c, \delta_v) L(t, \delta_c - \delta_v)\,. \end{aligned} \tag{2.51}$$

By symmetries, we can derive $L(t, \delta_c, \delta_v)$, from which follows

$$L(t, \delta_v, \delta_c) = \frac{L(t, \delta_v) - L(t, \delta_c)L(t, \delta_c - \delta_v)}{1 - L(t, \delta_v - \delta_c)L(t, \delta_c - \delta_v)} = \frac{\sinh\left(\delta_c \sqrt{2t}\right)}{\sinh\left((\delta_c - \delta_v)\sqrt{2t}\right)} . \tag{2.52}$$

Now, inverting the Laplace transformation, we obtain [96, 98]

$$S f_v(S) = \sum_{j=1}^{\infty} e^{-\frac{(j\pi x)^2}{2}} j\pi x^2 \sin(j\pi\mathcal{D}) \qquad \text{with} \qquad \mathcal{D} = \frac{|\delta_v|}{\delta_c + |\delta_v|}, \quad x = \frac{\mathcal{D}}{|\delta_v|}\sqrt{S} , \tag{2.53}$$

that is the Sheth & van de Weygaert [98] multiplicity function for void formation.

Note that the above results refer to the implementation of the sharp-k filter, that allow to consider each step of the path $\delta(R)$ uncorrelated to all the previous ones. Even if this allows analytic solutions, the sharp-k has an obscure physical meaning. To solve the Langevin equation for the general case, we have to account for the effects of correlated steps, that we will address in Sect. 2.4.1.

In this subsection, we implemented the spherical formation of halo in the excursion-set formalism. This is done by studying the first crossing problem with the constant threshold $\delta_c = 1.686$. However, the actual general dynamic of collapse is not spherical. In the literature, it has been shown that the tidal shear K_{ij}, i.e. the traceless part of $\partial_i\partial_j\Phi$, plays a crucial role in the formation of nonlinear structures. A step further with respect to the spherical model in approximating the dynamics of halo formation is to consider the collapse of triaxial matter distribution [107–111]. In first approximation, the dynamics of ellipsoidal collapse can be effectively incorporated into the excursion-set approach through a scale-dependent threshold $B(S)$, usually called moving barrier. In this way, the critical density for non-spherical collapse also depends on $\partial_i\partial_j\Phi$, in particular it can be shown that it depends on two invariants of this tensor [19, 112–114]. Note that any additional variable adds a dimension to the first-crossing problem. Nevertheless, with good approximation to the exact result, the problem can be reduced to the a monodimensional one considering the barrier

$$B(S) = \sqrt{q}\delta_c \left[1 + \beta\left(\frac{q\sqrt{S}}{\delta_c}\right)^{-\alpha}\right] , \tag{2.54}$$

where α and β come from ellipsoidal dynamics [15]. In the case where a sharp-k case is adopted, this leads to the Sheth-Tormen multiplicity function, Eq. (1.108). From a practical point of view, it can be considered the moving barrier

$$B(S) = \delta_c \left[1 + \beta\sqrt{S}/\delta_c\right] , \tag{2.55}$$

with β fitted on cosmological simulations [115].

2.3.2 Lagrangian clustering in the excursion-set framework

Studying the correlated Langevin equations in the excursion-set framework, it is possible to describe the two-point, and more generally, the n-point statistics of halos and voids in Lagrangian space [102].

Let us consider two points in Lagrangian space $\mathbf{q}_1$ and $\mathbf{q}_2 = \mathbf{q}_1 + \mathbf{r}$; $\delta_1(R) = \delta(\mathbf{q}_1, R)$ and $\delta_2(R) = \delta(\mathbf{q}_2, R)$ are the density contrast fields filtered on scale R with a filter function. We

are interested in studying the joint probability distribution $P_{ij}(R_1, R_2; \mathbf{q}_1, \mathbf{r})$ of those pairs $\delta_1(R)$ and $\delta_2(R)$ that first cross the threshold respectively at R_1 and R_2, i.e the density probability that two points separated by r are contained within a halo or a voids. The indices i and j stand for halos or voids, so $P_{\rm hh}(R_1, R_2; \mathbf{q}_1, \mathbf{r})\mathrm{d}S_1\mathrm{d}S_2$ is the probability of finding the points $\mathbf{q}_1$ and $\mathbf{q}_1 + \mathbf{r}$ in two halos with Lagrangian radius in the range $(R_1, R_1 + \mathrm{d}R_1)$ and $(R_2, R_2 + \mathrm{d}R_2)$, while $P_{\rm vh}\mathrm{d}S_1\mathrm{d}S_2$ and $P_{\rm vv}\mathrm{d}S_1\mathrm{d}S_2$ are the probability to find the two points in a void and in a halo and in a pair of voids, respectively. Due to the underlying homogeneity and isotropy, the probability density cannot depend neither on the position $\mathbf{q}_1$ nor on the orientation of $\mathbf{r}$, so $P_{ij}(R_1, R_2; \mathbf{q}_1, \mathbf{r}) = P_{ij}(R_1, R_2, r)$.

To compute the correlation function in Lagrangian space for the i and j object, i.e. the halo and void auto-correlation function and the void-halo cross-correlation, we first select the scale of the object we want to correlate, $I = [R_{\rm min}, R_{\rm max}]$, where usually $R_{\rm max} \to \infty$. This corresponds to a mass cut in studying halos and to a minimum radius in selecting voids. The probability of determining two points separated by r contained within i and j objects is

$$\mathcal{P}_{ij}^{I_1 I_2}(r) = \int_{I_1}\int_{I_2} \mathrm{d}R_1 \mathrm{d}R_2 P_{ij}(R_1, R_2, r)\,. \tag{2.56}$$

Note that this is the probability of finding two mass elements in a pair of Lagrangian halos/voids of scale in the range I separated by distance r. To obtain the correlation function of the objects, we have to properly weigh the statistical contribution for each extended halo/void. Each mass element contributes to originate a halo or void in a statistical sense, with a weight of $1/V(R)$, where $V(R) = M(R)/\rho_{\rm m}$ is the volume associated with the filter used. Note that this term is the same that enters in the halo mass function to get the number density of halos from the multiplicity function (see Sect. 1.3.1). It follows that the number density of halos or voids is $n(R) = f(R)/V(R)$, i.e. the integration of Eq. (1.103), and the joint number density is $n_{ij}(R_1, R_2, r) = P_{ij}(R_1, R_2, r)/V(R_1)V(R_2)$, therefore, the correlation function of halos/voids can be expressed as

$$\xi_{ij}^{I_1 I_2}(r) = \frac{\int_{I_1}\int_{I_2} \mathrm{d}R_1 \mathrm{d}R_2\, n_{ij}(R_1, R_2, r)}{\int_{I_1} \mathrm{d}R\, n_i(R) \int_{I_2} \mathrm{d}R\, n_j(R)} - 1\,. \tag{2.57}$$

In analogy to the single field case, the joint density distribution can be obtained by integrating the system of correlated Langevin equations that describes the evolution of the spatially correlated processes $\delta_1(R)$ and $\delta_2(R)$. Let us now consider the R-evolution of two spatially correlated density contrast fields at distance r, the full system of Langevin equations is derived as in the single field case [102]

$$\begin{cases} \dfrac{\partial \delta_1(R)}{\partial R} = Q_1(R) \\ \dfrac{\partial \delta_2(R)}{\partial R} = Q_2(R) \\ \langle Q_1(R)\rangle = \langle Q_2(R)\rangle = 0 \\ \langle Q_1(R)Q_1(R')\rangle = \langle Q_2(R)Q_2(R')\rangle = \displaystyle\int_0^\infty \frac{\mathrm{d}k}{2\pi^2} k^2 P(k) \frac{\partial W(kR)}{\partial R}\frac{\partial W^*(kR')}{\partial R'} \\ \langle Q_1(R)Q_2(R')\rangle = \displaystyle\int_0^\infty \frac{\mathrm{d}k}{2\pi^2} k^2 P(k) j_0(kr) \frac{\partial W(kR)}{\partial R}\frac{\partial W^*(kR')}{\partial R'}\,, \end{cases} \tag{2.58}$$

where $j_0(x) = \sin x/x$ is the spherical Bessel function of order 0 and $\langle \delta_1(R) \rangle = \langle \delta_2(R) \rangle = 0$.

As in the single field case, the sharp-k filter allows to solve the correlated Langevin equations analytically. We change the time variable from R to the variance S as before, and in this way Eq. (2.58) simplifies as follows [102]

$$\begin{cases} \dfrac{\partial \delta_1(S)}{\partial S} = Q_1(S) & \langle \delta_1(S) \rangle = 0 \\ \dfrac{\partial \delta_2(S)}{\partial S} = Q_2(S) & \langle \delta_2(S) \rangle = 0 \\ \langle Q_1(S) \rangle = \langle Q_2(S) \rangle = 0 & \\ \langle Q_1(S) Q_1(S') \rangle = \langle Q_2(S) Q_2(S') \rangle = \delta_D(S - S') & \\ \langle Q_1(S) Q_2(S') \rangle = \dfrac{\partial \xi(r, S)}{\partial S} \delta_D(S - S') = j_0(k_f(S) r) \delta_{\rm D}(S - S') \,. & \end{cases} \tag{2.59}$$

The last equation is obtained in the following way. The quantity $\xi(r, S)$ is the linear two-point correlation function of matter filtered on the scale $k(S) = 1/R(S)$ with the sharp-k filter

$$\xi(r, S) = \frac{1}{2\pi^2} \int_0^{k(S)} {\rm d}k k^2 P(k) j_0(kr) \,, \tag{2.60}$$

and $S(k) = 1/2\pi^2 \int_0^k {\rm d}k k^2 P(k)$, it follows that

$$\frac{\partial \xi(r, S)}{\partial S} = j_0(kr) \,. \tag{2.61}$$

The associated Fokker-Planck equation for the unconstrained case is [102]

$$\frac{\partial \Pi_r(\delta_1, \delta_2, S)}{\partial S} = \frac{1}{2} \left[\frac{\partial^2}{\partial \delta_1^2} + \frac{\partial^2}{\partial \delta_2^2} + 2 \frac{\partial \xi(r, S)}{\partial S} \frac{\partial^2}{\partial \delta_1 \partial \delta_2} \right] \Pi_r(\delta_1, \delta_2, S) \,, \tag{2.62}$$

with initial conditions $\Pi_r(\delta_1, \delta_2, S = 0) = \delta_D(\delta_1) \delta_D(\delta_2)$ and boundary conditions vanishing at infinity. To find the first crossing joint distribution, we adopt the absorbing barrier approach. We consider the halo-halo case. The problem is to solve the Fokker–Planck equation with absorbing barriers at $\delta_1 = \delta_{\rm c}$ and $\delta_2 = \delta_{\rm c}$, thus finding the survival probability density for pairs that have never crossed the threshold. Nevertheless, the problem is now bi-dimensional and the absorbing barrier is formed by two perpendicular planes. We can address the problem by considering the probability current at each point [102]

$$\mathbf{J}(\delta_1, \delta_2, S) = -\frac{1}{2} \left[\frac{\partial \Pi_r}{\partial \delta_1} + \frac{\partial \xi(r, S)}{\partial S} \frac{\partial \Pi_r}{\partial \delta_2}, \frac{\partial \xi(r, S)}{\partial S} \frac{\partial \Pi_r}{\partial \delta_1} + \frac{\partial \Pi_r}{\partial \delta_2} \right] . \tag{2.63}$$

On the boundary $\delta_1 = \delta_{\rm c}$, $\Pi(\delta_{\rm c}, \delta_2, S) = 0$ implying $\partial \Pi_r / \partial \delta_2 = 0$, this reduces the current to

$$\mathbf{J}(\delta_{\rm c}, \delta_2, S) = -\frac{1}{2} \left[\left. \frac{\partial \Pi_r}{\partial \delta_1} \right|_{\delta_1 = \delta_{\rm c}}, \left. \frac{\partial \xi(r, S)}{\partial S} \frac{\partial \Pi_r}{\partial \delta_1} \right|_{\delta_1 = \delta_{\rm c}} \right] , \tag{2.64}$$

and analogously for $\delta_2 = \delta_{\rm c}$. The probability flux along the barrier $\delta_1 = \delta_{\rm c}$ is given by the scalar product $\mathbf{J} \cdot \mathbf{n}$, where $\mathbf{n} = (1, 0)$ is the normal of the absorbing plane,

$$\mathcal{F}_r(\delta_{\rm c}, \delta_2, S) = -\frac{1}{2} \left. \frac{\partial \Pi_r}{\delta_1} \right|_{\delta_1 = \delta_{\rm c}} . \tag{2.65}$$

This quantity describes the density probability of the pair (δ_1, δ_2) to hit the barrier at (δ_c, δ_2) at a given time S. After that δ_1 crosses the barrier at S_1, the problem reduces to the one-dimensional case, finding the survival probability of δ_2 with initial conditions $S = S_1$ and $\delta_2^{1\times} = \delta_2(S_1|\,\delta_1 = \delta_c)$. From Eq. (2.47) the survival probability is

$$P_1(S_2 - S_1, \delta_c - \delta_2^{1\times}) = \frac{\delta_c - \delta_2^{1\times}}{\sqrt{2\pi(S_2 - S_1)}} \exp\left[-\frac{(\delta_c - \delta_2^{1\times})^2}{2(S_2 - S_1)}\right], \tag{2.66}$$

and the joint probability distribution is the convolution

$$\begin{aligned} P_{\rm hh}(S_1, S_2, r) = &\int_{-\infty}^{\delta_c} {\rm d}\delta_2 \mathcal{F}_r(\delta_c, \delta_2, S_1) P_1(S_2 - S_1, \delta_c - \delta_2) + \\ &\int_{-\infty}^{\delta_c} {\rm d}\delta_1 \mathcal{F}_r(\delta_1, \delta_c, S_2) P_1(S_1 - S_2, \delta_c - \delta_1)\,. \end{aligned} \tag{2.67}$$

However, this expression requires to solve the probability density $\Pi_r(\delta_1, \delta_2, S)$, which can be done analytically for $r \to \infty$ and $r \to 0$, while for the general case there exists an approximate analytical solution [102].

At infinity, $r \to \infty$, the correlation of the two processes vanishes, so the solution of the Fokker-Planck equation is

$$\Pi_\infty(\delta_1, \delta_2, S) = \Pi(\delta_1, S)\Pi(\delta_2, S)\,, \tag{2.68}$$

where $\Pi(\delta, S)$ is the probability distribution for the one-dimensional process with an absorbing barrier described by Eq. (2.44), so

$$\Pi_\infty(\delta_1, \delta_2, S) = G(\delta_1, \delta_2, S) - G(\delta_1 - \delta_c, \delta_2, S) - G(\delta_1, \delta_2 - \delta_c, S) + G(\delta_1 - \delta_c, \delta_2 - \delta_c, S)\,, \tag{2.69}$$

where $G(\delta_1, \delta_2, S) = (2\pi S)^{-1} \exp[-(\delta_1^2 + \delta_2^2)/2S]$ is the solution of the Fokker-Planck equation without barriers. It is now straightforward to show that $P_{\rm hh}(S_1, S_2, r \to \infty) = f(S_1)f(S_2)$, from which follows $\xi_{ij}(r) \to 0$ as $r \to \infty$, as expected.

The limit $r \to 0$ is described by the perfectly correlated case [102]. Let us change the base of the stochastic process, choosing $\Sigma(S) = \delta_1(S) + \delta_2(S)$ and $\Delta(S) = \delta_1(S) - \delta_2(S)$, which are independent, i.e. $\langle \Sigma(S)\Delta(S)\rangle = 0$. The variances of the corresponding unconstrained probability distributions are $\sigma_\Sigma^2 = 2[S + \xi(r, S)]$ and $\sigma_\Delta^2 = 2[S - \xi(r, S)]$. For $r \ll R(S)$ we have $\xi(r, S) \simeq S$, therefore $\sigma_\Sigma^2 \simeq 4S$ and $\sigma_\Delta^2 \simeq 0$. This means that the distribution of $\Delta(S)$ can be approximated by a Dirac delta centered in 0. The solution of the Fokker-Planck equation without absorbing barrier in the $r \to 0$ case follows from Eq. (2.44),

$$\Pi_0(\delta_1, \delta_2, S) = \Pi(\delta_1 + \delta_2, 4S)\delta_{\rm D}(\delta_1 - \delta_2)\,. \tag{2.70}$$

From this equation, we obtain $P_{\rm hh}(S_1, S_2, r \to 0) = f(S_1)\delta_{\rm D}(S_1 - S_2)$, and consequently $\xi_{\rm hh}(r) \to 0$ as $r \to 0$. This is due to the fact that two mass elements separated by $r \to 0$, belong to the same halo or void.

The approximated general solution can be obtained with the ansatz [102]

$$\Pi_r(\delta_1, \delta_2, S) = G_r^+(\delta_1, \delta_2, S) - G_r^-(\delta_1 - \delta_c, \delta_2, S) - G_r^-(\delta_1, \delta_2 - \delta_c, S) + G_r^+(\delta_1 - \delta_c, \delta_2 - \delta_c, S)\,, \tag{2.71}$$

where

$$G_r^\pm(\delta_1, \delta_2, S) = \frac{1}{2\pi\sqrt{S^2 + \xi^2(r, S)}} \exp\left[\frac{S(\delta_1^2 + \delta_2^2) \mp 2\xi(r, S)\delta_1\delta_2}{2(S^2 - \xi^2(r, S))}\right]. \tag{2.72}$$

Note that only G_r^+ is a solution of the Fokker-Planck equation with an absorbing barrier. Substituting this ansatz into Eq. (2.62), we find that the approximation is valid for

$$2\frac{\partial\xi(r,S)}{\partial S}\frac{\partial^2}{\partial\delta_1\partial\delta_2}[G_r^-(\delta_1-\delta_c,\delta_2,S)+G_r^-(\delta_1,\delta_2-\delta_c,S)]\ll\frac{\partial\Pi_r(\delta_1,\delta_2,S)}{\partial S}\,, \tag{2.73}$$

and the approximated general joint distribution reads

$$\begin{aligned}P_{\rm hh}(S_1,S_2,r)=&\frac{\delta_c S_1 S_2+[S_1S_2-\delta_c(S_1+S_2)]\xi(r,S_{\rm min})+\delta_c^2\xi^2(r,S_{\rm min})+\xi^3(r,S_{\rm min})}{2\pi[S_1S_2\xi^2(r,S)]^{5/2}}\\&\times\exp\left[-\frac{\delta_c^2}{2}\frac{S_1+S_2-2\xi(r,S_{\rm min})}{S_1S_2-\xi^2(r,S_{\rm min})}\right].\end{aligned} \tag{2.74}$$

where $S_{\rm min}=\min(S_1,S_2)$. Using Eq. (2.57) it is now possible to obtain $\xi_{\rm hh}(r)$ [102].

2.4 Numerical solution of the Langevin equations

In the previous section, we introduced the Langevin equations that describe the evolution of the stochastic field $\delta(\mathbf{q},R)$ as a function of the filtering radius R. We showed that with the sharp-k filter, these equations admit analytical solutions. This is because the process they describe reduces to a Wiener one, so at each step the trajectory value $\delta(\mathbf{q},R)$ is uncorrelated from all the previous ones. Nevertheless, to describe properly the halo and void formation in the excursion-set framework, we should use a filter with a clear link to the physical quantity we want to describe, i.e. Lagrangian halos and voids. To analytically solve the Langevin equations in the excursion-set framework with correlated steps, approximate methods must be used [12, 115–124]. Moreover, the first crossing distribution for the single field can be written as a formal expansion in an infinite series [125–127], but summing this series requires approximations. Numerical methods are actually the only way to obtain an exact solution of the general Langevin equations for the single field, spatially correlated pairs, and more generally for N correlated fields in Lagrangian space.

2.4.1 Numerical implementation: the sharp-k filter case

In this subsection, we describe the numerical procedure used to solve the Langevin equations and to compute both the first-crossing distribution in the single field case and the two-point correlation functions using the sharp-k filter. Even if this filter allows to solve analytically the Langevin equations, it is important to explore its numerical implementation for at least two reasons. First, the numerical solution of this case is the basis to solve the Langevin equations with correlated time steps. Second, the sharp-k filter does not provide an exact analytical solution for the joint distribution in both the single and double barrier case.

Single field implementation: the first crossing distribution

The stochastic differential equation Eq. (2.42) is equivalent to the integral equation

$$\delta(S+\Delta S)=\delta(S)+\int_S^{S+\Delta S}\mathrm{d}\,S'Q(S')\,. \tag{2.75}$$

Given the fact that this is a Wiener process,

$$\delta(S + \Delta S) = \delta(S) + \sqrt{\Delta S}\, G(0,1)\,, \tag{2.76}$$

where $G(0,1)$ is a Gaussian variable that follows a Gaussian distribution with zero mean and unit variance. The evolution of the density contrast with respect to the filtering scale is a random walk that can be numerically implemented as

$$\delta(S) = \sum_i \sqrt{\Delta S_i}\, G_i(0,1)\,, \tag{2.77}$$

where G_is are independent Gaussian variables and $S = \sum_i \Delta S_i$. After several realizations of this stochastic process, the first-passage time distribution can be estimated by recording the scale at which the trajectory reaches the barrier for the first time. The resulting first crossing distribution can be obtained as

$$F(<S) = \frac{N_\times(<S)}{N_{\rm sample}} \to \int_0^S f_{\rm c}(S')\mathrm{d}S' \quad \text{for} \quad N_{\rm sample} \to \infty\,. \tag{2.78}$$

where $f_{\rm c}(S')$ is the multiplicity function. Nevertheless, the numerical implementation of this equation will produce a suppressed multiplicity function with respect to the analytical solution of Eqs. (2.49) and (2.53). This is due to the discretization of the time variable S. Note that the numerical implementation of the random walk can be seen as a discrete subsampling of a continuous random walk, increasing the step resolution, the random walk would continue to appear jagged. Fig. 2.1 visually represents the effect of increasing the resolution of the numerical realization of a random walk. Increasing the resolution, the same realization of a random walk (blue jagged line) may cross the barrier between two non-crossing steps of the worse resolved case (orange dots). This is known as inter-step crossing, and we have to account for it to find the exact first crossing distribution numerically. The probability of this event can be computed considering the first-crossing distribution of a Wiener process pinned at both $t = 0$ and $t = T$, i.e. a Brownian bridge, where t is the time variable of the process [104, 105].

Let $B(t)$ be a Brownian bridge, i.e. a Wiener process such that $B(0) = B_0$, $B(T) = B_T$, and $0 < t < T$. The random variables $B(t)$ and $B(T) - B(t)$ are independent and normally distributed according to the definition of Wiener processes. Thus, $X = B(t)$ and $Y = B(T) = B(t) + (B(T) - B(t))$ follow a bivariate normal distribution that can be written as

$$P_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}\frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}, \tag{2.79}$$

with $\mu_X = \mu_Y = B(0)$, $\sigma_X = \sqrt{t}$, and $\sigma_Y = \sqrt{T}$. The correlation ρ can easily be derived from

$$\begin{aligned}\langle X, Y\rangle = \sigma_X\sigma_Y\rho &= \langle B(t)(B(t) + [B(T) - B(t)]\,)\rangle \\ &= \langle B(t)^2\rangle + \cancel{\langle B(t)(B(T) - B(t))\rangle} \\ &= \sigma_X^2\,,\end{aligned} \tag{2.80}$$

it follows that

$$\rho = \frac{\sigma_X}{\sigma_Y} = \sqrt{\frac{t}{T}}\,. \tag{2.81}$$

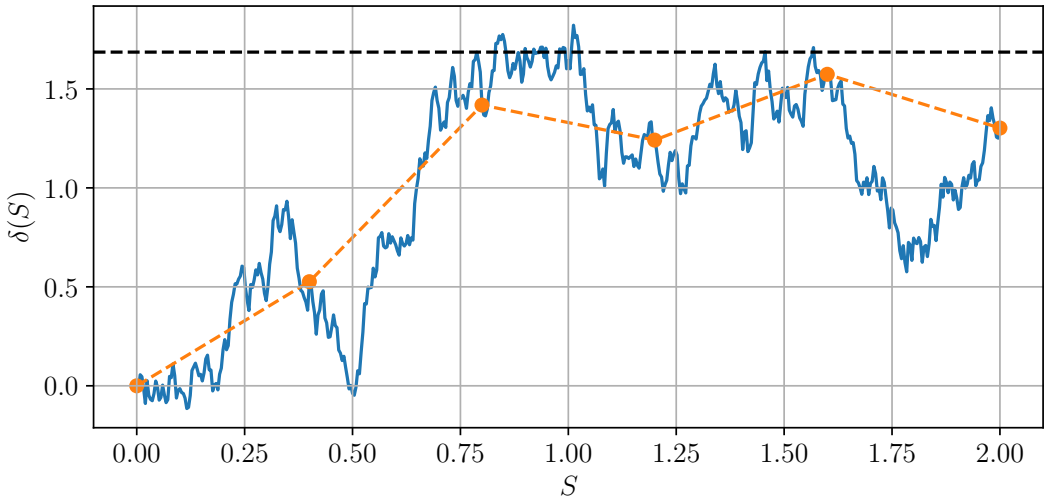

Figure 2.1: Inter-step crossing in Markovian random walk. The orange dots represent a low-resolution realization of the random walk, the blue jagged line is a realization of the Brownian bridge pinned at each point of the original random walk. It can be seen the effect of inter-step crossing due to higher resolution.

Fixing $y = B(T) = B_T$, the conditional distribution of X is the normal distribution

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P(y)} = \frac{1}{\sqrt{2\pi}\sigma_{X|Y}} \exp\left[-\frac{1}{2}\frac{(x-\mu_{X|Y})^2}{\sigma^2_{X|Y}}\right], \tag{2.82}$$

with

$$\mu_{X|Y} = \mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y-\mu_Y) = B_0 + \frac{t}{T}(B_T - B_0) \tag{2.83}$$

and

$$\sigma_{X|Y} = \sigma_X\sqrt{1-\rho^2} = \sqrt{t\left(1-\frac{t}{T}\right)}. \tag{2.84}$$

These equations show that the Brownian bridge is a Brownian motion with expectation value $\mu_{X|Y}$ and variance $\sigma^2_{X|Y}$. Note that, as expected, the variance vanishes at $t = 0$ and $t = T$ according to the constraints and reaches its maximum value at $t = T/2$, while the expectation value $\mu_{X|Y}$ is the linear interpolation of the initial and final points. The stochastic properties of the Brownian bridge allow to calculate the probability of inter-step crossing [104, 128, 129], which is given by

$$P_{\mathrm{BB}-\times}(B_0, B_T, B_b, T) = \exp\left[-2\frac{(B_b - B_0)(B_b - B_T)}{T}\right], \tag{2.85}$$

where $B_b = \delta_c$ is the boundary. The numerical solution of the Langevin equations with sharp-k filter can be obtained as follows. For each new i^{th} step of the random walk, we compute the inter-step crossing probability $P_{\mathrm{BB}-\times}(\delta_{i-1}, \delta_i, \delta_c, \Delta S)$. Then we perform a little Monte Carlo by generating a random number from a flat distribution over $(0, 1)$. If the generated number is smaller than $P_{\mathrm{BB}-\times}(\delta_{i-1}, \delta_i, \delta_c, \Delta S)$, we consider that an inter-step crossing has occurred, and so the random walk has first crossed the boundary in the time step (S_{i-1}, S_i). In this way,

the numerical implementation is exact and does not depend on the time-step resolution, so the numerical solution will converge to the analytical result as the number of stochastic realizations grows.

Spatially correlated fields and the joint distribution

Analogously to the single field case, the correlated Langevin equations with sharp-k filter, Eq. (2.59), can be numerically solved considering their integral equations

$$\begin{cases} \delta_1(S+\Delta S) = \delta_1(S) + \displaystyle\int_S^{S+\Delta S} \mathrm{d}S'\, Q_1(S') & \delta_1(0) = 0 \\ \delta_2(S+\Delta S) = \delta_2(S) + \displaystyle\int_S^{S+\Delta S} \mathrm{d}S'\, Q_2(S') & \delta_2(0) = 0\,, \end{cases} \tag{2.86}$$

where the stochastic forces $Q_1(S)$ and $Q_2(S)$ are correlated and their properties are given by Eq. (2.59). To solve this system, we want to decouple the integral of the first stochastic force from the second one. We can do that by exploiting the Gaussianity of the fields. Any M-dimensional random vector $\boldsymbol{\delta}$ with correlation matrix $\mathbf{C}$ of components $c_{ij} = \langle \delta_i \delta_j \rangle$, can be written as a lineal combination of another Gaussian random vector $\mathbf{X}$ with a diagonal correlation matrix, that can be the identity without any loss of generality [130]. Consider now the linear system $\boldsymbol{\delta} = \mathbf{A}\mathbf{X}$, where $\mathbf{A}$ is a $(M \times M)$ matrix with components α_{ij}. This system in components reads $\delta_i = \sum_j \alpha_{ij} X_j$. It follows that the correlation matrix can be rewritten as $\mathbf{C} = \mathbf{A}\mathbf{A}^T$, in components $c_{ij} = \sum_k \alpha_{ik}\alpha_{kj}$. We apply this decomposition to our problem that involves a 2-dimensional random field,

$$\delta_i = \sum_{j=1}^{2} \alpha_{ij} X_j \qquad c_{ij} = \sum_{k=1}^{2} \alpha_{ik}\alpha_{jk}\,. \tag{2.87}$$

The second equation fixes three of the four α_{ij}, so we have the freedom to fix one of them. Choosing $\alpha_{12} = 0$ we obtain:

$$\alpha_{11}^2 = c_{11} \qquad \alpha_{12} = 0 \tag{2.88}$$

$$\alpha_{11}\alpha_{21} = c_{21} = c_{12} \qquad \alpha_{22}^2 + \alpha_{21}^2 = c_{22}\,. \tag{2.89}$$

This way of triangulating the covariance matrix is also known as Cholesky decomposition (see Sect. 2.4.3). Exploiting this parameterization, Eq. (2.86) together with Eq. (2.58) can be rewritten as [102]

$$\begin{cases} \delta_1(S+\Delta S) = \delta_1(S) + \alpha_{11}(\Delta S, S)\, G_1(0,1) & \delta_1(0) = 0 \\ \delta_2(S+\Delta S) = \delta_2(S) + \alpha_{21}(\Delta S, S)\, G_1(0,1) + \alpha_{22}(\Delta S, S)\, G_2(0,1) & \delta_2(0) = 0 \\ \alpha_{11}(\Delta S, S)^2 = \alpha_{21}(\Delta S, S)^2 + \alpha_{22}(\Delta S, S)^2 = \Delta S & \\ \alpha_{11}(\Delta S, S)\,\alpha_{21}(\Delta S, S) = \xi(r; S+\Delta S) - \xi(r; S)\,, & \end{cases} \tag{2.90}$$

where $G_1(0,1)$ and $G_2(0,1)$ are two independent Gaussian variables with zero mean and unit variance. In the last equation, we can exploit the fact that

$$\xi(r; S+\Delta S) - \xi(r; S) = \int_S^{S+\Delta S} \frac{\partial \xi(r; S')}{\partial S'} \mathrm{d}S' = \int_S^{S+\Delta S} j_0(k_f(S')r)\mathrm{d}S'\,, \tag{2.91}$$

thus for $\Delta S \ll 1$, $\alpha_{11}(\Delta S, S)\,\alpha_{21}(\Delta S, S) \simeq j_0(k_f(S)r)\Delta S$.

The first equation of the system Eq. (2.90) is formally equal to Eq. (2.76). To numerically solve the second equation, we consider the corresponding inter-step crossing probability. This is analogous to Eq. (2.76), but with:

$$\sigma_X^2 = \alpha_{21}^2(t) + \alpha_{22}^2(t), \qquad \sigma_Y^2 = \alpha_{21}^2(T) + \alpha_{22}^2(T)\,. \tag{2.92}$$

This gives the inter-step expectation value and variance

$$\mu_{X|Y} = B_0 + \rho^2(t)(B_T - B_0), \qquad \sigma_{X|Y} = \sigma_X\sqrt{1-\rho^2(t)} \tag{2.93}$$

with

$$\rho(t) = \sqrt{\frac{\alpha_{21}^2(t) + \alpha_{22}^2(t)}{\alpha_{21}^2(T) + \alpha_{22}^2(T)}}\,. \tag{2.94}$$

Thus, the probability of inter-step crossing is

$$P_{\rm crossing} = \exp\left[-2\frac{(B_b - B_0)(B_b - B_T)}{\alpha_{21}^2(T) + \alpha_{22}^2(T)}\right], \tag{2.95}$$

where B_b is the boundary value.

2.4.2 General filter: standard implementation

In this subsection, we describe the classical method used to solve the Langevin equation in the general filter case for both the single field and spatially correlated fields.

The Langevin equation for the single field with a generic filter is given by Eq. (2.37). To solve it numerically, we discretize the time variable R, i.e. the smoothing scale, in a set $\{R_0 \dots R_i \dots R_N\}$, with $R_i > R_{i+1}$. Differently from the sharp-k case, the stochastic forces at different smoothing lengths are no longer independent to each other. The stochastic force $\mathcal{Q}(R_i)$ at a given smoothing length R_i depends on the entire set of previous steps $R_j > R_i$. To solve this equation by brute force, it is possible to draw a random Gaussian realization of $\mathcal{Q}(R_i)$ from the covariance matrix $\langle \mathcal{Q}(R_i)\mathcal{Q}(R_j)\rangle$ for each realization of the trajectory $\delta(R)$. A numerically more convenient alternative is to solve these equations by exploiting the Gaussianity of the unfiltered field $\delta(\mathbf{k})$, which implies that Fourier amplitudes of different wavenumbers are independent [19]. Let us consider the smoothed density field

$$\begin{aligned}\delta(\mathbf{q}, R) &= \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{-\mathbf{k}\cdot\mathbf{q}}\,\delta(\mathbf{k})W(kR) \\ &= \int_0^\infty dk W(kR)\left[\frac{1}{(2\pi)^3}k^2\int d\Omega_k e^{-\mathbf{k}\cdot\mathbf{q}}\,\delta(\mathbf{k})\right].\end{aligned} \tag{2.96}$$

The integral in the second line is rewritten in spherical coordinates, dividing the angular part from the radial part. Consider now the density contrast filtered with the sharp-k filter, i.e. the density contrast integrated over a sphere in k space,

$$\delta_{\rm sk}(\mathbf{q}, R) = \int_0^{k=1/R} \frac{dk'}{(2\pi)^3}k'^2\int d\Omega_{k'} e^{-\mathbf{k}'\cdot\mathbf{q}}\,\delta(\mathbf{k}')\,. \tag{2.97}$$

The stochastic force corresponding to a sharp-k filter, that we call $Q_{\rm sk}$, can be obtained by deriving the previous equation,

$$\frac{\partial \delta_{\rm sk}(\mathbf{q},R)}{\partial R} = \frac{\partial k}{\partial R}\frac{\partial}{\partial k}\delta_{\rm sk}(\mathbf{q},R=1/k) = \frac{\partial}{\partial k}\delta_{\rm sk}(\mathbf{q},R=1/k) = Q_{\rm sk}(k)\,, \tag{2.98}$$

where $Q_{\rm sk}(k) = Q_{\rm sk}(R = 1/k)$. Comparing this last expression with Eq. (2.97), it is clear that the sharp-k stochastic force $Q_{\rm sk}(k)$ is equal to the term in the square brackets of Eq. (2.96), that can be written as [19]

$$\delta(\mathbf{q},R) = \int_0^\infty dk W(kR) Q_{\rm sk}(k)\,. \tag{2.99}$$

This equation shows that a stochastic realization of the Langevin equation with a standard filter $W(kR)$, corresponds to a realization of the sharp-k case filtered at different filtering radii R with $W(kR)$. This methodology was exploited in the literature to compute the multiplicity function with various filters [115, 131, 132]. This stochastic integral can be solved numerically given a realization of a set of sharp-k stochastic force $Q_{\rm sk}(k_i)$. In particular, assuming small enough k steps, the stochastic force can be written as $Q_{\rm sk}(k_i) = G_i\sqrt{k_i^2 P(k_i)\Delta k/2\pi^2}$, where G_i is a random number drawn from a Gaussian distribution with zero mean and unit variance; this stochastic integral can be numerically solved as [19]

$$\delta(R) = \sum_i G_i W(k_i R)\sqrt{\frac{k_i^2 P(k_i)}{2\pi^2}\Delta k}\,. \tag{2.100}$$

This equation is exact in the limit $\Delta k = (k_{i+1} - k_i) \to 0$, but a finite Δk value becomes a source of biases in the final results, strongly dependent on the size of the steps. In the following, we propose a more stable numerical solution that is almost independent of the discretization of the k step. The integral in k of Eq. (2.99) can be split into many integrals with the integration interval given by the numerical steps k_i,

$$\delta(R) = \sum_i \int_{k_i}^{k_{i+1}} dk W(kR) Q_{\rm sk}(k)\,. \tag{2.101}$$

If $\Delta k/k \ll 1$, then $W(k_i R) \simeq W(k_{i+1}R)$. Since the value of the filter is slowly varying in the (k_i, k_{i+1}) interval, it can be evaluated at the mean k-value, $k = (k_{i+1} + k_i)/2$, and moved out of the integration sign. The numerical integration of the stochastic force $Q_{\rm sk}(k)$ in the interval (k_i, k_{i+1}), is better estimated changing variable from k to $S = \sigma^2(k)$, and the result is $G_i\sqrt{\Delta S}$, where $\Delta S = S(k_{i+1}) - S(k_i)$ and G_i is a random number drawn from a Gaussian distribution with zero mean and unit variance. This result is exact and does not depend on the integration extrema. For $\Delta k \ll 1$, $\Delta S = k_i^2 P(k)\Delta k/(2\pi^2)$ recovering Eq. 2.100. Finally, the numerical estimate of $\delta(R)$ implemented in this work is

$$\delta(R) = \sum_i G_i W\left(\frac{k_{i+1}+k_i}{2}R\right)\sqrt{\Delta S_i}\,. \tag{2.102}$$

The correlation between the different smoothing lengths changes the way to solve the first crossing distribution. The Markovian random walk solving the Langevin equation in the sharp-k case is jagged, and the inter-step crossing has to be taken into account to numerically find the

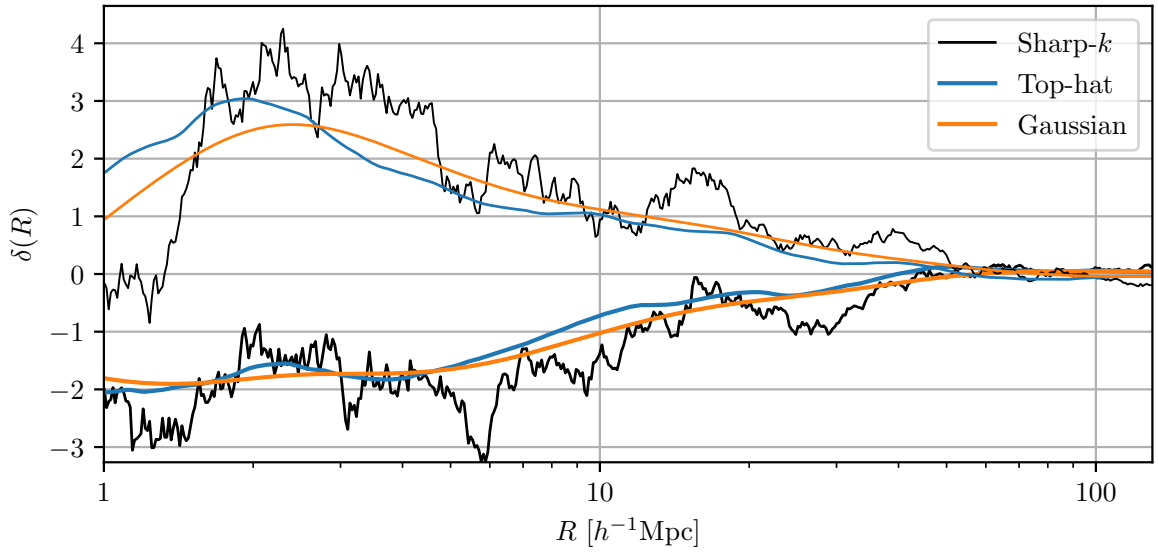

Figure 2.2: Two stochastic evolution of the Lagrangian filtered $\delta(R)$ field, represented by a different line width. The same stochastic realization is represented considering the sharp-k (black), top-hat (blue) and Gaussian (orange) filter. It can be noted that the filter impacts the scale of the first crossing occurrence.

first crossing distribution. In the general filter case, on the contrary, more the steps are close to each other and more the correlation is strong. It follows that the realizations of trajectories solving the Langevin equations are smooth, and to obtain the first crossing distribution is enough to follow the evolution of the field $\delta(R)$ at each R_i, finding where it crosses the threshold. Fig. 2.2 shows the impact of the filtering kernel on the same stochastic realization, note that the first crossing scale can be strongly impacted.

Let us now consider the R-evolution of two spatially correlated density contrast fields at distance r, $\delta_1(R_1)$ and $\delta_2(R_2)$, for which the stochastic evolution is described by the spatially correlated Langevin equations Equation (2.58). We can exploit Eq. (2.99) both the first and the second field, where we substitute $\sqrt{\Delta S_i}$ with the sharp-k realization defined by Eq. (2.90),

$$\begin{cases} \delta_1(R) = \sum_i W\Big(\langle k\rangle_i R\Big)\, \alpha_{11}\Big(\langle S\rangle_i, \Delta S_i\Big)\, G_i^{(1)} \\ \delta_1(R) = \sum_i W\Big(\langle k\rangle_i R\Big)\Big[\alpha_{21}\Big(\langle S\rangle_i, \Delta S_i\Big)\, G_i^{(1)} + \alpha_{22}\Big(\langle S\rangle_i, \Delta S_i\Big)\, G_i^{(2)}\Big], \end{cases} \tag{2.103}$$

where $\langle S\rangle_i = (S_{i+1} + S_i)/2$ and $\langle k\rangle_i = (k_{i+1} + k_i)/2$.

2.4.3 General filter: Cholesky decomposition

The use of Cholesky decomposition in solving the Langevin equations in the excursion-set framework was first proposed in [127]. This decomposition has the advantage of greatly speeding up the computation of the realization of stochastic trajectories $\delta(R)$.

Let us consider the correlation of the field $\delta(R)$ filtered at different smoothing lengths R,

$$\langle \delta(R_i)\delta(R_j)\rangle = C_{ij} = \int \frac{\mathrm{d}k}{2\pi^2} k^2 P(k) W(kR_i) W^*(kR_j)\,. \tag{2.104}$$

The matrix $\mathbf{C}$ is real, symmetric and positive-definite, so it has a unique decomposition $\mathbf{C} = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is a lower triangular matrix. It is now possible to construct a stochastic numerical realization of $\delta(R)$ as

$$\delta(R_i) = \sum_j L_{ij} G_j \,, \tag{2.105}$$

where G_j is the j^{th} element of a vector of Gaussian variables, following a distribution with 0 mean and unit variance. Note that since $\mathbf{L}$ is triangular, each δ_i requires a sum only over $j \leq i$. From the correlation of $\delta(R)$ it is easy to verify that this trajectory satisfies the required statistical properties,

$$\langle \delta(R_i)\delta(R_j)\rangle = \sum_{m,n} L_{im} L_{jn} \langle G_i G_j \rangle \,, \tag{2.106}$$

where $G_i G_j = \delta^{\text{K}}_{ij}$, and δ^{K}_{ij} is the Kronecker delta, so

$$\langle \delta(R_i)\delta(R_j)\rangle = \sum_m L_{im} L_{jm} = \mathbf{L}\mathbf{L}^T = \mathbf{C} \,. \tag{2.107}$$

The decomposition can be obtained from the following algorithm. We start with the elements on the diagonal, considering that $\mathbf{L}$ is a lower triangular matrix, we can write,

$$C_{jj} = \sum_{k=1}^{j} L_{jk}^2 = L_{jj}^2 + \sum_{k=1}^{j-1} L_{jk}^2 \qquad \Rightarrow \qquad L_{jj}^2 = C_{jj} - \sum_{k=1}^{j-1} L_{jk}^2 \,. \tag{2.108}$$

Let now consider $j < i$,

$$C_{ij} = \sum_{k=1}^{j} L_{ik} L_{jk} \,, \tag{2.109}$$

since for $j < k < i$, $L_{jk} = 0$, it follows

$$C_{ij} = L_{ij} L_{jj} + \sum_{k=1}^{j-1} L_{ik} L_{jk} \qquad \Rightarrow \qquad L_{ij} = \left[C_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} \right] L_{jj}^{-1} \,. \tag{2.110}$$

Starting from $i = 0$, Eqs. (2.108) and (2.110) provide the L_{ij} values for increasing i and $j \leq i$.

We extended the results of [127] showing that this decomposition can also be exploited for spatially correlated fields [84]. Let us consider the correlation of two fields,

$$\langle \delta_K(R_i)\delta_L(R_j)\rangle = \int \frac{\mathrm{d}k}{2\pi^2} k^2 P(k) W(kR_i) W^*(kR_j) j_0(kr_{LK}) \,. \tag{2.111}$$

In this equation, the labels K and L indicate the field, i.e. δ_1 or δ_2, and r_{KL} is the separation of the two fields, in such a way that

$$r_{KL} = \begin{cases} 0 & K = L \\ r & K \neq L \,. \end{cases} \tag{2.112}$$

The full correlation matrix is

$$\mathbf{C} = \begin{bmatrix} \langle \delta_1(R_i)\delta_1(R_j)\rangle & \langle \delta_1(R_i)\delta_2(R_j)\rangle \\ \langle \delta_2(R_i)\delta_1(R_j)\rangle & \langle \delta_2(R_i)\delta_2(R_j)\rangle \end{bmatrix} = \begin{bmatrix} \mathbf{C}^0 & \mathbf{C}^r \\ \mathbf{C}^r & \mathbf{C}^0 \end{bmatrix} , \tag{2.113}$$

where $\mathbf{C}^0$ denotes the covariance matrix of the field $\delta(R)$ at different smoothing lengths, while $\mathbf{C}^r$ is the covariance matrix of two fields separated by a distance r at various smoothing radii. The elements of this block matrix can be written as C_{IJ}. The upper case of the indices I and J stands for the fact that each dimension of C_{IJ} has twice the elements of each dimension of C^0_{ij} and C^r_{ij}. The indices run in the following way, $C_{IJ} = C^0_{ij}$ for both I and J running from 1 to n or both running from $n+1$ to $2n$, i.e. the diagonal blocks, where n is the number of smoothing radii R_i considered; $C_{IJ} = C^r_{ij}$ for I running from 1 to n and J running from $n+1$ to $2n$ or *vice versa*, i.e. the anti-diagonal blocks. Both C^0_{ij} and C^r_{ij} are real, symmetric, and positive definite matrices, therefore also C_{IJ} has the same properties, allowing the Cholesky decomposition

$$C_{IJ} = \sum_{K=1}^{2n} \mathcal{L}_{IK}\mathcal{L}_{JK} \,. \tag{2.114}$$

Note that also the $\mathcal{L}_{IJ}$ lower triangular matrix, it can be computed in analogy of Eqs. (2.108) and (2.110), and it can be expressed as a block matrix,

$$\mathcal{L}_{IJ} = \begin{bmatrix} \mathbf{L}^{(1)} & 0 \\ \mathbf{L}^{(2)} & \mathbf{L}^{(3)} \end{bmatrix}, \tag{2.115}$$

where the $\mathbf{L}^{(1)}$ and $\mathbf{L}^{(3)}$ blocks are lower triangular matrices; in particular the first one is exactly the one derived for the single field case. The upper left block is a zero matrix, while $\mathcal{L}^{(2)}$ is a non-zero $(n \times n)$ matrix. A numerical realization of the correlated paths can be obtained via a $2n$ dimensional vector of random Gaussian variable with zero mean and unitary variance, $G_I = (\mathbf{G}^{(1)}, \mathbf{G}^{(2)})$, where $G_{I=i} = G^{(1)}_i$ and $G_{I=n+i} = G^{(2)}_i$, where $i \leq n$,

$$\begin{cases} \delta_1(R_i) = \displaystyle\sum_{j=1}^{n} L^{(1)}_{ij} G^{(1)}_j \\ \delta_2(R_i) = \displaystyle\sum_{j=1}^{n} L^{(2)}_{ij} G^{(1)}_j + \sum_{j=1}^{n} L^{(3)}_{ij} G^{(2)}_j \,. \end{cases} \tag{2.116}$$

Note that the first equation is exactly Eq. (2.105), so even if the sum runs from 1 to n, it is non-zero only for $j \leq i$. The first term in the right-hand side of the second equation depends on the Gaussian realizations G^1_i, i.e. it is the term correlated to the evolution of the first field. Note that $L^{(2)}_{ij}$ is not a triangular matrix, so, for each R_i, the second field depends on the entire evolution of the first via all the n elements of the $G^{(1)}_j$ set. The second term is a sum over a lower triangular matrix as for the first field, but with different coefficients. Also in this case the sum is non-zero only for $j \leq i$. It is interesting to note the formal analogy between this system and the solutions of Equations (2.90) and (2.103). This is due to the fact that in both cases we decouple the first field from the second one using the Cholesky decomposition. In Equations (2.90) and (2.103), the correlation matrix reduces to a 2×2 matrix, so the Cholesky decomposition involves just one coefficient for the first field and two for the second one. On the other hand, by exploiting the decomposition also to decompose the correlation between different smoothing lengths, the problem is described by a $2n \times 2n$ block matrix of four $n \times n$ matrices.

Both the classical method and the implementation with Cholesky decomposition have their own advantages and disadvantages. In the top-hat filter case, the computation of the path

realization is faster using the Cholesky decomposition. This is due to the fact that the operations involved for each stochastic realization of the field $\delta(R)$ using the Cholesky are the generation of n Gaussian random numbers, multiplications and sums, described in Eq. (2.116). On the standard method side, for n filtering smoothing lengths, the realization of the sharp-k stochastic force generally requires $\gtrsim n\, k_i$ steps and a generation of a random number for each step. When the full vector of sharp-k stochastic force is generated, it is filtered at each R_i with the top-hat filter, that is a composition of trigonometric functions, much slower to compute with respect to summing matrix elements as Eq. (2.105). For spatially correlated fields, the computational resources saved with the Cholesky decomposition method are even larger with respect to the single field case. On the other hand, the Cholesky decomposition is a numerically delicate task. Eq. (2.108) involves the difference of numbers very close to each other, the results are positive for each j, but very close to 0. If the radius steps are small, numerical errors can be larger than the value to be computed, so the computation of L^2_{jj} may result to be negative, but L_{ij} is expected to be real. This requires high control in computing the integrals of Eq. (2.104) and Eq. (2.111), which involve two 1^{th} order and one 0^{th} order spherical Bessel functions, highly oscillating and slowly decreasing. Computing these integrals with a high enough accuracy is heavy, and in some cases the precision required may not be reachable, in particular for small smoothing radius steps.

2.5 Results

In this section, we discuss the statistical properties of halos and voids obtained by numerically solving the Langevin equations. In particular, we focus on the multiplicity function, on which the halo mass function and the void size function depend, and the two-point correlations, i.e. the halo-halo, void-void, and void-halo Lagrangian correlation functions. The statistics inferred from the solution of the Langevin equations are exact in Lagrangian space and describe the density and spatial correlations of spherically filtered patches that first reach the threshold value. Nevertheless, to compare the theoretical results with simulations and to use them in data analysis, there are two points to be aware of, the first concerning the relationship between the first crossing problem and the void or halo formation/identification, and the second regarding the map of the void and halo statistics from Lagrangian to Eulerian space.

Halo definition is a subtle point, since we have to consider how it is defined in the theoretical framework, i.e. the filtered density contrast field that reaches the collapsing threshold at a given smoothing scale, and how it is defined in observation/simulations. This means that, if we want to compare the theoretical results with simulations, we must be aware that the halo definition depends on the halo finder and, *vice versa*, the halo finder should be consistent with the theoretical definition of halo. From a practical point of view, halo formation in the theoretical framework and the halo definition implicitly contained in a halo finder can be effectively matched using a moving barrier on the theoretical side. In the void case, the treatment is similar but with a peculiarity, since they do not undergo to a full nonlinear collapse, the implementation of a moving barrier is not necessary. They can be defined both in the Lagrangian and in the Eulerian space as filtered fluctuations that reach a fixed threshold value. Void finding algorithms used to compare theory with data or simulations have to consider voids accordingly to this definition. Alternatively, effective modifications to the threshold can be performed, as in the halo case, but we do not consider this possibility here.

The second point concerns the map from Lagrangian to Eulerian space, which is crucial

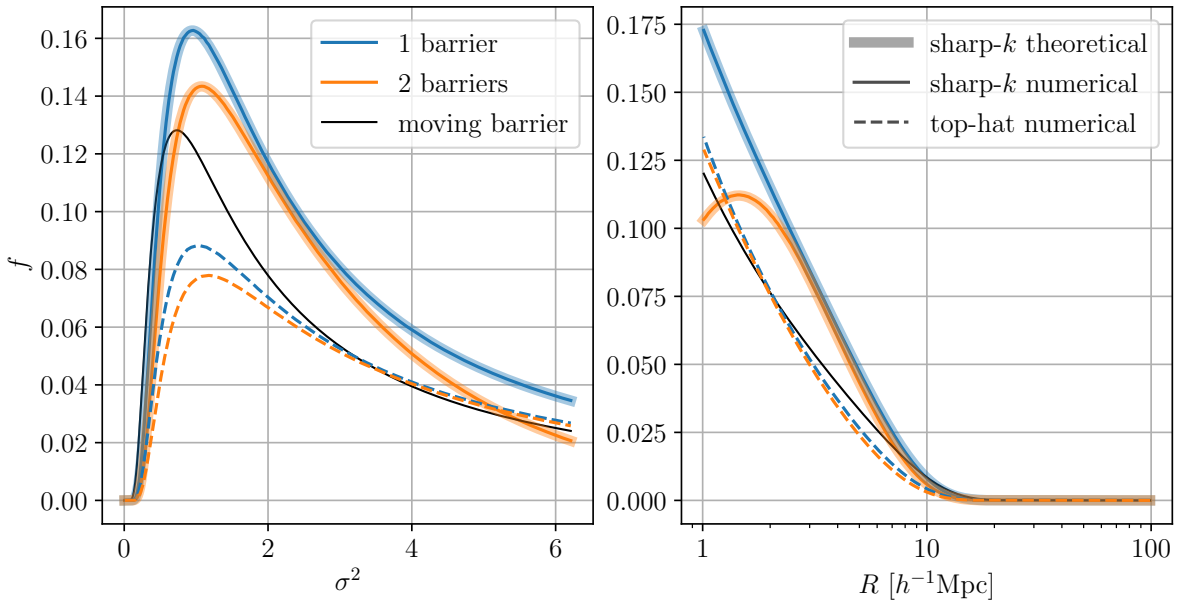

Figure 2.3: The multiplicity function from the first crossing distribution in the single (blue) and double barrier problem (orange) for theoretical prediction (shaded curve) and numerical computation (solid line) with the sharp-k filter, and the numerical result with the top-hat filter (dashed line). On the left panel, the multiplicity function is expressed as a function of the variance σ^2, on the right plot as a function of the smoothing length R. See the text for the conversions from σ^2 to R in the sharp-k case. The collapsing threshold is $\delta_c = 1.686$ and the void formation threshold considered is $\delta_v = -1.796$, corresponding to a nonlinear value of -0.7.

in the excursion-set framework. In particular, voids are greatly impacted by the map from Eulerian to Lagrangian space, since they expand and both the multiplicity function and the joint distribution of voids depend on the void radius, as discussed in Sect. 2.1.2. In the halo case, the Lagrangian to Eulerian map does not explicitly enter in the 0^{th} order statistics, i.e. the halo mass function and the multiplicity function; nevertheless, it is important in the two-points statistics, since the relative position between two halos moves in going from Lagrangian to Eulerian space. Consequently, in this section, aside from presenting the results coming from the numerical solution of the Langevin equations, we also discuss the impact of the map from Lagrangian to Eulerian space and the relationship between the first crossing and void/halo formation.

2.5.1 The multiplicity function

The multiplicity function is the first crossing probability between R and $R + \mathrm{d}R$, and can be numerically estimated as

$$f(\langle R_i \rangle)\Delta R_i = \frac{N_\times(R_i)}{N_{\text{sample}}}\,, \tag{2.117}$$

where $\langle R_i \rangle = (R_i + R_{i-1})/2$, $\Delta R_i = (R_i - R_{i-1})/2$, $N_\times(R_i)$ is the number of paths that at R_i have exceeded the threshold for the first time, and N_{sample} is the number of stochastic realizations performed. Fig. 2.3 shows the theoretical multiplicity function for the sharp-k filter and the numerical realization for the sharp-k and top-hat filter. The multiplicity function is expressed as a function of σ^2 (left panel) and of the smoothing length R (right panel). Note that the mul-

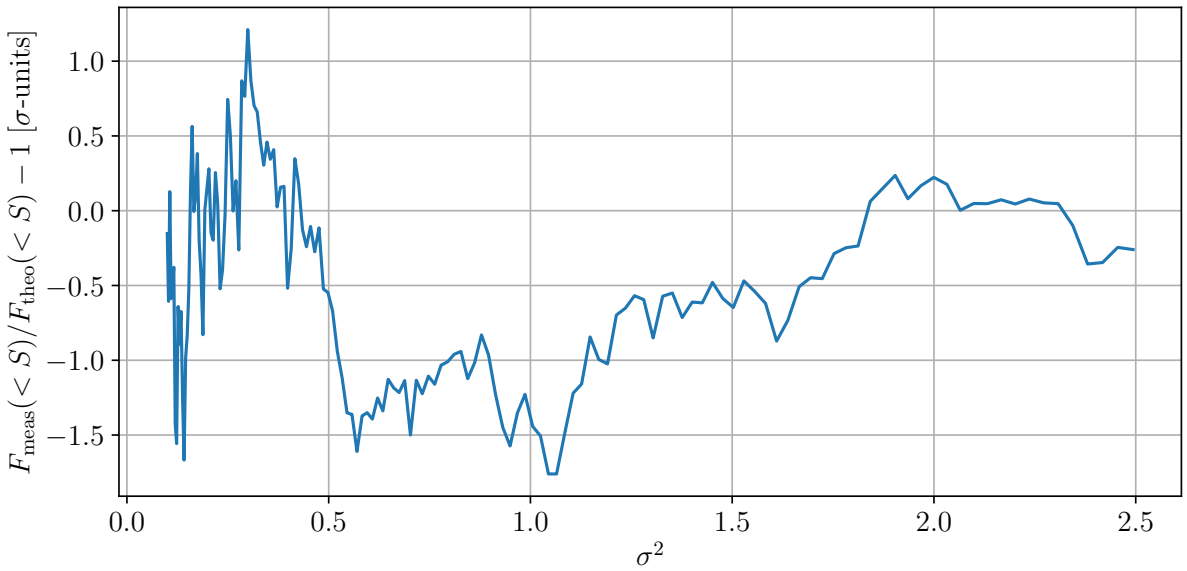

Figure 2.4: Sharp-k survival probability obtained with 1.2×10^7 stochastic realizations against the theoretical prediction in Poissonian-sigma units, defined as $1/\sqrt{F_{\rm theo}(<S)N_{\rm sample}}$, where $N_{\rm sample}$ is number of realization performed.

tiplicity function is a differential quantity, so we change the variable according to Eq. (2.48). In the sharp-k case of the right panel, the conversion from $\sigma(k)$ to R scale is obtained requiring $\sigma^2_{\rm sk}(k)=\sigma^2_{\rm th}(R)$ (see Sect. 2.2 for conversion from k to R or mass scale). Note that an increase in radius size corresponds to a decrease in σ^2, so large smoothing scales correspond to small σ^2. The numerical realization for the sharp-k case is exactly reproduced, thanks to having considered the inter-step crossing. Fig. 2.4 shows the difference between the numerical and the theoretical first crossing distribution. The results are expressed in units of the Poissonian uncertainty given the number of realizations, i.e. the square root of the inverse of the expected first crossing probability times the number of realizations performed; it can be seen that the deviation from the theoretical prediction is always of the magnitude of the Poissonian uncertainty. The top-hat multiplicity function is suppressed with respect to the sharp-k one at large scales (Fig. 2.3), but consider the arbitrary in the conversion from k to R for the sharp-k case, which translates into some freedom in fixing the x-axis scale.

Fig. 2.5 shows the halo mass function derived from the various multiplicity functions discussed, compared to the halo mass function measured in DEMNUni simulations for ΛCDM cosmology at $z=0$ (see Sect. 3.1). As can be seen, the halo mass function from the sharp-k filter (blue lines) overestimates the number of low mass halos and underestimates the number of the massive ones, for both the k to mass conversion adopted. The top-hat result (orange curve) is quite accurate at low masses but cannot reproduce massive halos, whereas the Sheth-Tormen multiplicity function accurately reproduces the measurements.

The void size function, or void abundance, is analogous to the halo mass function for voids, but it presents some relevant differences. The void size function is the number density of voids with respect to their radius; in the literature, this is often expressed as a function of the logarithm of the radius in h^{-1}Mpc units, $\mathrm{d}n/\mathrm{d}\ln R$. To express it theoretically, we must consider that mass conservation entails that the void radius evolves going from Lagrangian to Eulerian space. We define $R_{\rm L}$ the Lagrangian radius and R the corresponding Eulerian value. The void

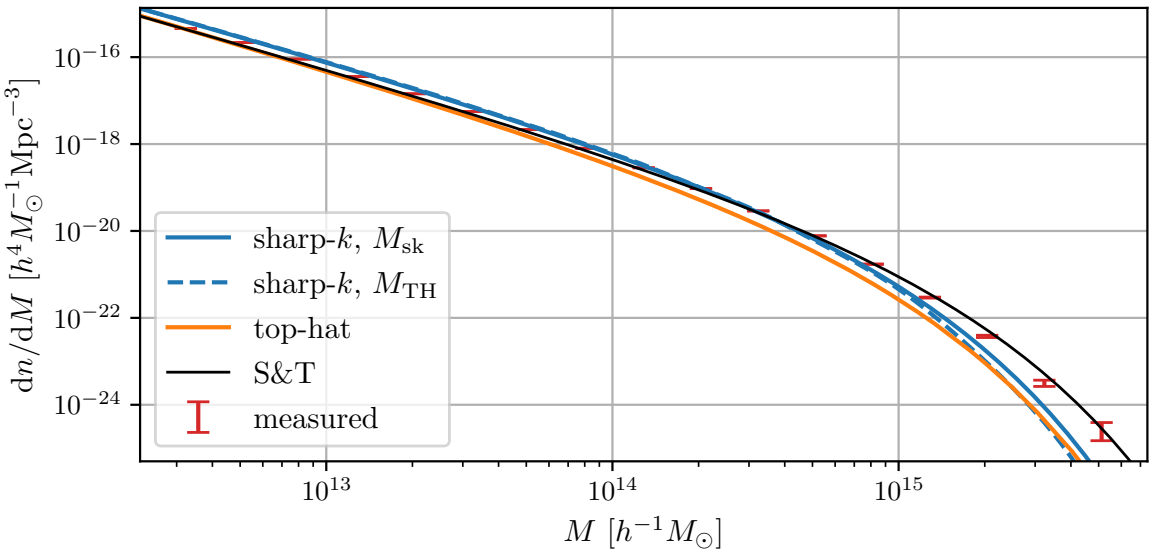

Figure 2.5: The halo mass function measured (red error-bars) in DEMNUni simulations for ΛCDM cosmology at $z = 0$ versus the prediction from the excursion-set model. The blue lines represent the predictions for the sharp-k filter where the two possible mass to k conversions are considered, $M_{\rm TH}$ (dashed line) is obtained requiring $\sigma_{\rm sk}(k) = \sigma_{TH}(R)$ and $M_{\rm sk}$ is the mass enclosed by the sharp-k filter, as described in Sect. 2.2. The orange and black curves represent the results for the top-hat filter and for the Sheth-Tormen multiplicity function, respectively.

size function in Lagrangian space is formally analogous to the universal halo mass function of Eq. (1.106),

$$\frac{\mathrm{d}n_{\rm L}}{\mathrm{d}M} = \frac{\rho}{M} f(\ln\sigma) \left|\frac{\mathrm{d}\ln\sigma}{\mathrm{d}M}\right| , \tag{2.118}$$

where M is the mass corresponding to a Lagrangian fluctuation of radius $R_{\rm L}$. We express the multiplicity function as a function of $\ln\sigma$ following the notation of Jennings et al. [96], which addressed the problem of the conversion from Lagrangian to Eulerian space for the first time. Using $\ln\sigma$ as variable is useful when dealing with the multiplicity function obtained analytically with the sharp-k filter; nevertheless, for multiplicity functions obtained numerically, there is no particular reason for this choice. Changing the variable from M to $\ln R_{\rm L}$, we obtain

$$\frac{\mathrm{d}n_{\rm L}}{\mathrm{d}\ln R_{\rm L}} = \frac{f(\ln\sigma)}{V(R_{\rm L})} \left|\frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln R_{\rm L}}\right| , \tag{2.119}$$

where $V(R_{\rm L}) = 4\pi R_{\rm L}^3/3$ is the volume of the sphere with radius $R_{\rm L}$. If the number of voids is conserved passing from Lagrangian to Eulerian space, the Eulerian void size function is

$$\frac{\mathrm{d}n}{\mathrm{d}\ln R} = \left.\frac{\mathrm{d}n_{\rm L}}{\mathrm{d}\ln R_{\rm L}}\right|_{R_{\rm L}=R_{\rm L}(R)} , \tag{2.120}$$

where $R_{\rm L}(R)$ is the Lagrangian radius corresponding to the Eulerian radius R. If we assume that voids are spherically symmetric, the mass conservation entails that this map reads

$$\frac{R}{R_{\rm L}} = (1+\delta_{\rm v}^{\rm NL})^{-1/3} , \tag{2.121}$$

where $\delta_{\rm v}^{\rm NL}$ is the nonlinear, i.e. the Eulerian, density contrast of void formation corresponding to the linear threshold $\delta_{\rm v}$. Note that voids in the excursion-set model are fluctuations filtered with a spherical filter; this definition does not say anything about the internal symmetry and mass distribution of voids, which can be any. This is particularly important in considering the map between Lagrangian and Eulerian space, since the internal structures of voids and the tidal field sourced by the external mass distribution impact the evolution of voids; therefore, the map can be more complex than Eq. (2.121) and the ones described in Sect. 2.1. Nevertheless, in the literature the spherical symmetric map is always assumed, since it should be reasonably accurate for large enough voids, at least as a first approximation.

Jennings et al. [96] noted that given Eqs. (2.118) and (2.121), the fraction of volume enclosed in voids with a radius larger than $R_{\rm min}$ is

$$\mathcal{F}(R_{\rm min}) = \int_{R_{\rm min}}^{\infty} \frac{{\rm d}R}{R} V(R) \frac{{\rm d}n}{{\rm d}\ln R} \,, \tag{2.122}$$

therefore, for the sharp-k multiplicity function of Eq. (2.53) the total volume fraction enclosed in voids is [96, 98]

$$\mathcal{F}(0) = \frac{1}{1+\delta_{\rm v}^{\rm NL}} \int_0^{\infty} \frac{{\rm d}\sigma}{\sigma} f(\ln\sigma) = \frac{1-\mathcal{D}}{1+\delta_{\rm v}^{\rm NL}} \,, \tag{2.123}$$

were $\mathcal{D} = |\delta_{\rm v}|/(\delta_{\rm c} + |\delta_{\rm v}|)$. It follows that for various values of $\delta_{\rm v}$, this quantity exceeds unity, meaning that the volume fraction enclosed in voids exceeds the volume of the universe, which is impossible. To avoid this unphysical result, Jennings et al. [96] proposed the so called "volume conserving model", or Vdn model, to map the Lagrangian void size function into the Eulerian one. This model imposes the conservation of the volume fraction enclosed in voids in passing from Lagrangian to Eulerian space, while the total number density of voids is no longer conserved,

$$V(R)\,{\rm d}n(R) = V(R_{\rm L})\,{\rm d}n_{\rm L}|_{R_{\rm L}=R_{\rm L}(R)} \,, \tag{2.124}$$

so that

$$\begin{aligned} \frac{{\rm d}n}{{\rm d}\ln R} &= \frac{V(R_{\rm L})}{V(R)} \frac{{\rm d}n_{\rm L}}{{\rm d}\ln R_{\rm L}} \frac{{\rm d}\ln R_{\rm L}}{{\rm d}\ln R}\bigg|_{R_{\rm L}=R_{\rm L}(R)} \\ &= \frac{f(\ln\sigma)}{V(R)} \left|\frac{{\rm d}\ln\sigma}{{\rm d}\ln R_{\rm L}}\right| \left[\frac{{\rm d}\ln R_{\rm L}}{{\rm d}\ln R}\right]_{R_{\rm L}=R_{\rm L}(R)} \,. \end{aligned} \tag{2.125}$$

Assuming spherical voids, $R/R_{\rm L} = {\rm const}$ according to Eq. (2.121), so ${\rm d}\ln R_{\rm L}/{\rm d}\ln R = 1$ and the Vdn model for the void size function reads [96]

$$\frac{{\rm d}n}{{\rm d}\ln R} = \frac{f(\ln\sigma)}{V(R)} \left|\frac{{\rm d}\ln\sigma}{{\rm d}\ln R_{\rm L}}\right| \,. \tag{2.126}$$

Note, however, that this result is motivated by unphysical assumptions in considering both the multiplicity function and the map from Lagrangian to Eulerian space. Moreover, the extra assumption that the number density of voids is not conserved going from Lagrangian to Eulerian space and uniformly suppressed is strong; this would mean that only a fraction of voids in Lagrangian space will form a void in Eulerian space, independently of the size. One of the criticalities arising from the total volume problem is that voids are considered to evolve as perfect spheres, in the Lagrangian to Eulerian map. It follows that the map from Lagrangian to

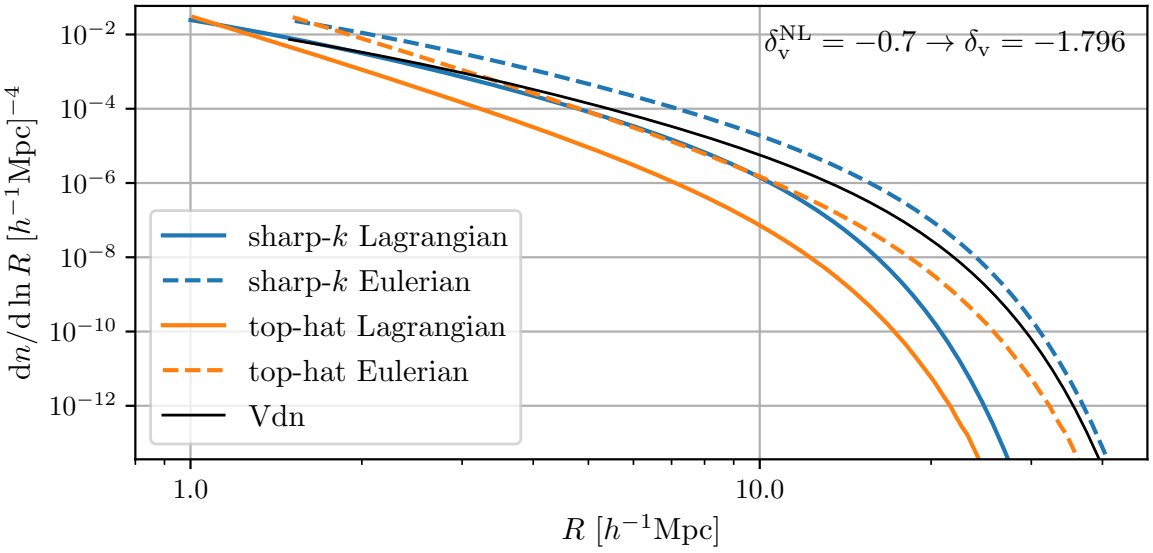

Figure 2.6: void size function for voids with nonlinear density contrast $\delta_v^{NL} = -0.7$ corresponding to a linear threshold $\delta_v = -1.796$. The blue lines represent the Lagrangian (solid) and Eulerian (dashed) void abundances, while the black line represents the Vdn model. The orange curves correspond to the top-hat results for Lagrangian (solid) and Eulerian space (dashed).

Eulerian radius is given by the relation Eq. (2.121), that does not depend on the void size. In addition to this, if a sharp-k filter is assumed, the volume fraction of voids may exceed the volume of the universe, as described by Eq. (2.123). Let us forget for a moment the problems related to the sharp-k multiplicity function and consider the Lagrangian to Eulerian map only to solve the total volume problem; then instead of imposing a uniform suppression on the void number density by a factor of $(1 + \delta_v^{NL})$, it makes more sense to consider an effective radius mapping, relaxing the non-overlapping hard sphere idealization. In this direction, a first approximation can be to consider the spherical approximation in the Lagrangian to Eulerian map valid only for voids larger than a given minimum radius, since for small voids tidal fields and nonlinear effects due to the background environment reasonably break the spherical evolution. In other words, we can assume that Eq. (2.120) together with Eq. (2.121) is valid for $R > R_{min}$, with R_{min} to be empirically determined. Note, however, that even if we assume Eq. (2.121) as a first approximation, the volume fraction of voids expressed by Eq. (2.123) is valid only for the sharp-k filter, for which the k to radius conversion is ill defined, so the volume itself is ill defined. Note that the multiplicity function for the top-hat filter assumes lower values with respect to the sharp-k one (see Fig. 2.3), and in the top-hat case we do not find that the integral Eq. (2.122) exceeds unity, at least for the threshold values considered, $\delta_v = [-2.7, -2, -1.796, -1, -0.6]$, and for the radii explored, i.e. $R_{min} = 1h^{-1}\mathrm{Mpc}$.

To conclude, we can resume the treatment of the volume fraction problem raised by Jennings et al. [96] in two points. The problem is sourced by the sharp-k multiplicity function and by the map between Lagrangian and Eulerian space. Our suggestion is that, first, there is no need of this correction when a multiplicity function with physical meaning is considered, such as for a top-hat filtering kernel; second, breaking the conservation of the number of voids (independently on their size) in passing from Lagrangian to Eulerian space is an extremely strong assumption. The spherical approximation may be not accurate enough; nevertheless, even if

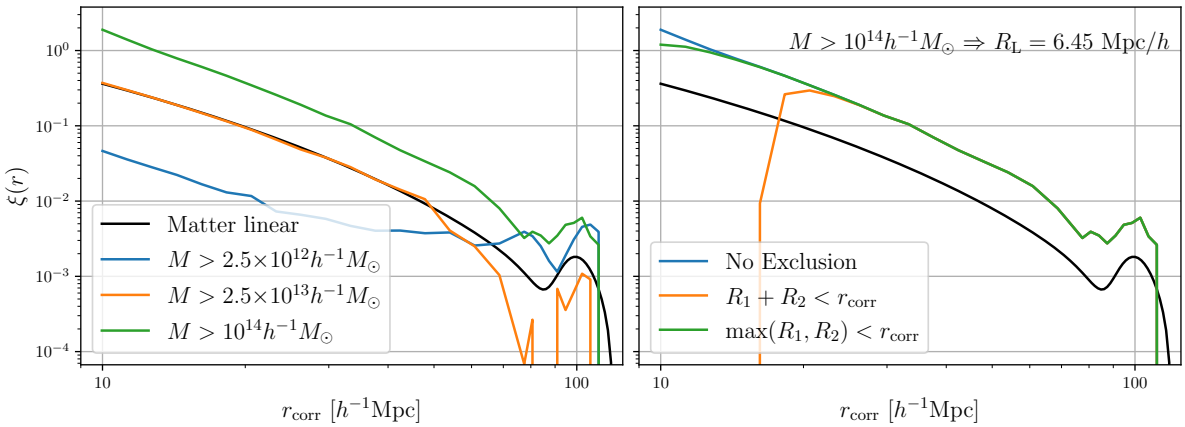

Figure 2.7: Lagrangian halo-halo correlation function obtained with the top-hat filter for various minimum mass (left panel). The right panel shows different ways of handling the exclusion effect in the correlation function with a minimum halo mass of $10^{14}h^{-1}M_\odot$, corresponding to a Lagrangian radius of $\sim 6.45h^{-1}$Mpc. The exclusion effect in the right panel is not considered for the blue curve, implemented requiring $R_1 + R_2 < r_{corr}$ (orange), and requiring $\max(R_1, R_2) < r_{corr}$, where R_1 and R_2 are the filtering smoothing radii at which the halos form.

its accuracy should be investigated both theoretically and with simulation, fixing the volume fraction is unphysical.

Fig. 2.6 shows the void size function with nonlinear density contrast of $\delta_v^{NL} = -0.7$ corresponding to a linear threshold of $\delta_v = -1.796$, obtained with the sharp-k filter (blue) and top-hat filter (orange), for both Lagrangian (solid) and Eulerian (dashed) space; the Vdn model (black) is considered only in the sharp-k filter case. The Lagrangian to Eulerian map for void radius is the spherical one of Eq. (2.121). As can be seen, the top-hat void size function is lower than the sharp-k one, and the Eulerian top-hat void size function predicts more small voids and less large voids with respect to the Vdn sharp-k model.

2.5.2 Lagrangian correlation functions

The Lagrangian correlation function is obtained via Eq. (2.57), where the joint distribution is computed as

$$P_{hh}(\langle R_i\rangle, \langle R_j\rangle, r)\Delta R_i \Delta R_j = \frac{N_\times(R_i, R_j, r)}{N(1)N(2|1)}\,, \tag{2.127}$$

where, as usual, $\langle R_i\rangle = (R_i + R_{i-1})/2$ and $\Delta R_i = (R_i - R_{i-1})/2$. The quantity $N_\times(R_i, R_j, r)$ is the number of pairs that have exceeded the threshold for the first time at R_i and R_j, respectively, $N(1)$ is the number of realizations of the first field, and $N(2|1)$ is the number of realizations of the correlated second field for each realization of the first one. The number of pairs $N_\times(R_i, R_j, r)$ that first exceed the threshold is obtained by counting the first crossing of several numerical realizations of the spatially correlated fields, via Eq. (2.103) or Eq. (2.116). In this subsection, we consider the implementation of top-hat filter only.

We start by considering the halo-halo correlation function. The left panel of Fig. 2.7 shows the results for three minimum halo masses, $2.5\times10^{12}h^{-1}M_\odot$, $2.5\times10^{13}h^{-1}M_\odot$, and $10^{14}h^{-1}M_\odot$, respectively, which correspond to Lagrangian fluctuations with radius 1.88, 4.07, and 6.45 h^{-1}Mpc. The black curve represents the unfiltered linear correlation function of matter. Note that direct solution of neither the joint distribution Eq. (2.127) nor the statistical realization of

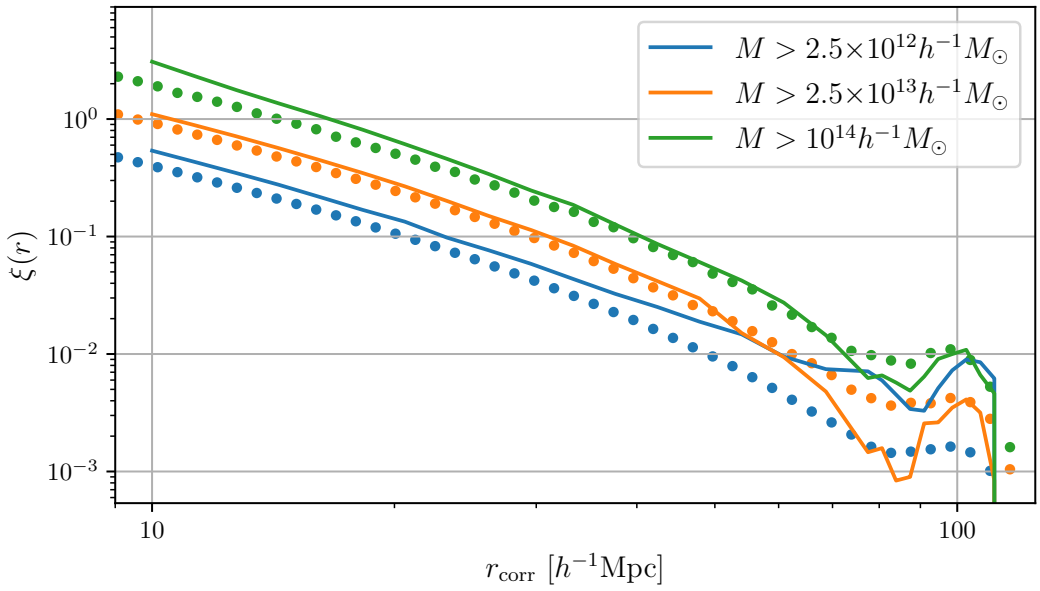

Figure 2.8: Eulerian halo-halo correlation function obtained with the top-hat for various minimum mass (solid curves) compared with measurements from DEMNUni simulations (scatters). See text for details.

the pairs (Eqs. (2.103) and (2.116)) automatically accounts for the exclusion effect, i.e. the exclusion of pairs that correspond to mass elements of the same halo in the computation of the correlation function. A simple way to account for it is not to consider pairs for which the sum of the smoothing radii is larger than the correlation distance, i.e. $R_1 + R_2 < r$. This is shown in the right panel of Fig. 2.7, where it can be seen that the impact is quite large. The basis of this argument is that halos are considered as non-overlapping spheres in Lagrangian space. Another possibility is to not consider pairs for which the Lagrangian radius of the largest halo in the pair is larger than the correlation length [133], i.e. $\max(R_1, R_2) < r_{\rm corr}$, represented by the green curve in Fig. 2.7. This is motivated by the fact that if the correlation is within the size of the halo, it will collapse within the same virialized object. Nevertheless, collapsed objects do not show a spherical radius in Lagrangian space, so other possibilities should be explored. The exclusion effect has important consequences [see e.g. 134] and the two previous definitions provide the scale at which this should happen, without entering in the details. More physical definition can be adopted considering simulations [see e.g. 135, 136].

The solution of the Langevin equations provides the Lagrangian correlation function of filtered fluctuations that first reach the collapsing threshold. This result has to be converted into a Eulerian quantity to be comparable with the observed halo distribution in the real Universe or in simulations. The map of the Lagrangian correlation function into the Eulerian one is not an easy task. A possible way to do this map is to use the perturbative approach introduced in Sect. 1.3.4. We consider the Eulerian halo density contrast truncated at the first order $\delta_{\rm h}^{\rm E} \simeq \delta_{\rm h}^{\rm L} + \delta_{\rm bg}$, since on large scale $\delta_{\rm h} \ll 1$ and $\delta_{\rm h}^{\rm bg} \ll 1$, correlating this quantity we obtain

$$\xi_{\rm hh}^{\rm E}(r) \simeq \xi_{\rm hh}^{\rm L}(r) + \sqrt{\xi_{\rm hh}^{\rm L}(r)\xi(r)} + \xi(r)\,, \tag{2.128}$$

where $\xi(r)$ is the linear correlation function of matter. Fig. 2.8 shows the results of Eq. (2.128) using the Lagrangian correlation function obtained with the top-hat filter and constant barrier,

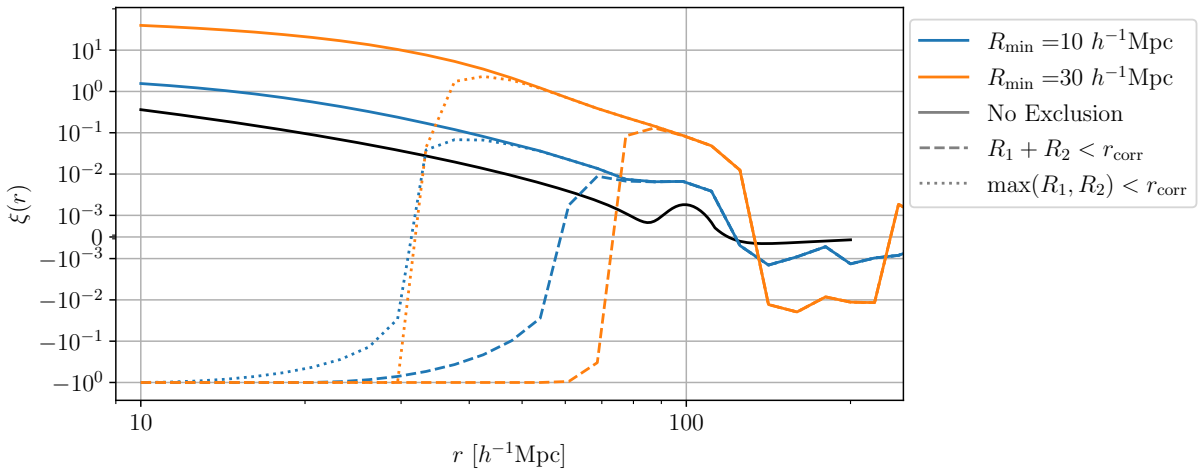

Figure 2.9: Lagrangian void-void correlation function with radius larger than 10 h^{-1}Mpc (blue) and 30 h^{-1}Mpc (orange). The linear void formation threshold considered is $\delta_{\rm v} = -0.6$. The line style labels the modelization of the exclusion effect.

together with the measured two-point correlation function of FoF halos in DEMNUni simulations. It can be seen that the numerical result fits the measurements quite well in the range of validity of the linear approximation, i.e. $r \gg \max(R_1, R_2)$ and far from the baryonic acoustic oscillation (BAO) feature, even if the simplification of considering a constant barrier. The small overestimation of the theoretical correlation function is probably due to the constant barrier implementation with the top-hat kernel. As can be seen in Fig. 2.5, the top-hat filter with a constant barrier underestimates the number of massive halos, so the bias is overestimated with respect to the true one, and the correlation function reflects this behavior. The deviation with respect to simulations around the BAO dip is mainly due to the fact that there are not enough numerical realizations. Nevertheless the BAO feature is not expected to be accurately reproduced because nonlinearities and nonlocalities that affect the BAO feature [19] are not considered in the modeling.

The void-void Lagrangian correlation function is obtained analogously to the halo-halo correlation, and the considerations concerning the exclusion effect are the same. Fig. 2.9 shows the results obtained at $z = 0$ in Lagrangian space for voids with minimum Lagrangian radius of 10 h^{-1}Mpc and 30 h^{-1}Mpc, represented by the blue and orange curves, respectively. The exclusion effect is implemented as in the halo-halo case. The solid lines represent the results for which no exclusion is considered, the dashed lines represent the results for which the exclusion is considered requiring $R_1 + R_2 < r_{\rm corr}$, while the dotted lines represent the exclusion implemented as $\max(R_1, R_2) < r_{\rm corr}$. Note that voids are larger than halos. This in Lagrangian space is true if we consider a void formation threshold lower in magnitude with respect to the one for collapse, i.e. $-\delta_{\rm v} < \delta_{\rm c}$, and in Eulerian space because of void expansion. Note that the main feature in the observed void-void correlation function is the exclusion effect that produces a peak. The exact implementation is not explored here, but we note the order of magnitude of its impact, which is relevant, since it can reach the BAO scales.

The void-halo cross-correlation function is an important quantity in analysis focused on cosmic voids. It represents the probability of finding a halo at a distance r from the void, therefore it is the expectation value of the density profile of voids in the halo distribution. This quantity is important on its own in studying redshift-space distortions and Alcock-Paczyński

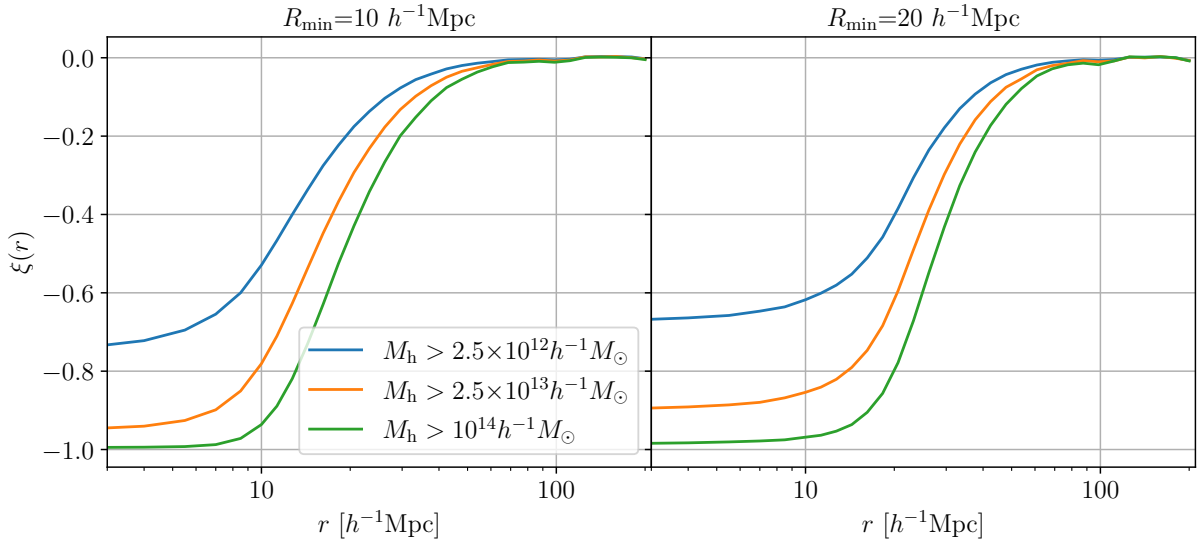

Figure 2.10: Lagrangian void-halo correlation function with radius larger than 10 h^{-1}Mpc (left) and 20 h^{-1}Mpc (right) computed for halos with various minimum mass. The linear void formation threshold considered is $\delta_{\rm v} = -0.6$.

effect around voids. Moreover, if combined with the void density profile in matter distribution (discussed later), it can provide a link between the matter distribution and the halo distribution in and around voids, useful for various cosmological analysis, such as the void size function, void-galaxy lensing, integrated Sachs-Wolfe in voids etc. [see e.g. 137–144].

The Lagrangian void-halo cross-correlation is obtained as usual, except for the exclusion effect, which has not to be considered since we are correlating different objects, and double counting is not possible. Note indeed that a fluid element can be part of a void and a halo at the same time; for instance, a halo within a void shares all its fluid elements with the void. Fig. 2.10 shows the results for voids larger than 10 h^{-1}Mpc (left panel) and 20 h^{-1}Mpc (right panel), correlated with halos with different minimum masses. The linear void formation threshold considered here is $\delta_{\rm v} = -0.6$, which corresponds to a nonlinear threshold of $\delta_{\rm v}^{\rm NL} \simeq -0.4$. This value is chosen because it can be practically considered as the lowest dark matter density contrast value that can be traced by galaxies to the end of cosmological analysis of galaxy surveys (see Chapter 3 for details). The void-halo cross-correlation function is the probability of finding a halo at distance r from a void. As expected, in the inner region of voids it is very unlikely that halos can be found, then the probability increases and reaches zero at large radii. The halo mass impacts the void-halo cross-correlation, it is less probable to find a massive halo near or within a void with respect to a smaller one. This is intuitive for the inner region of voids; nevertheless, even in the outer part, it is more probable to find less massive halos because they are more abundant than the more massive ones. The map from Lagrangian to Eulerian cross-correlation is not trivial since, differently from the halo-halo case, we are interested in the range where the correlation function, and consequently the density contrast, is of order 1, i.e. $|\xi_{\rm vh}| \sim 1$, therefore the implementation of Eq. (2.128) is not optimal. We can explore other possibilities, in particular, since we are considering the position of halos with respect voids, we can exploit the Lagrangian void density profile and their expansion in Eulerian space, which is the topic of the next subsection.

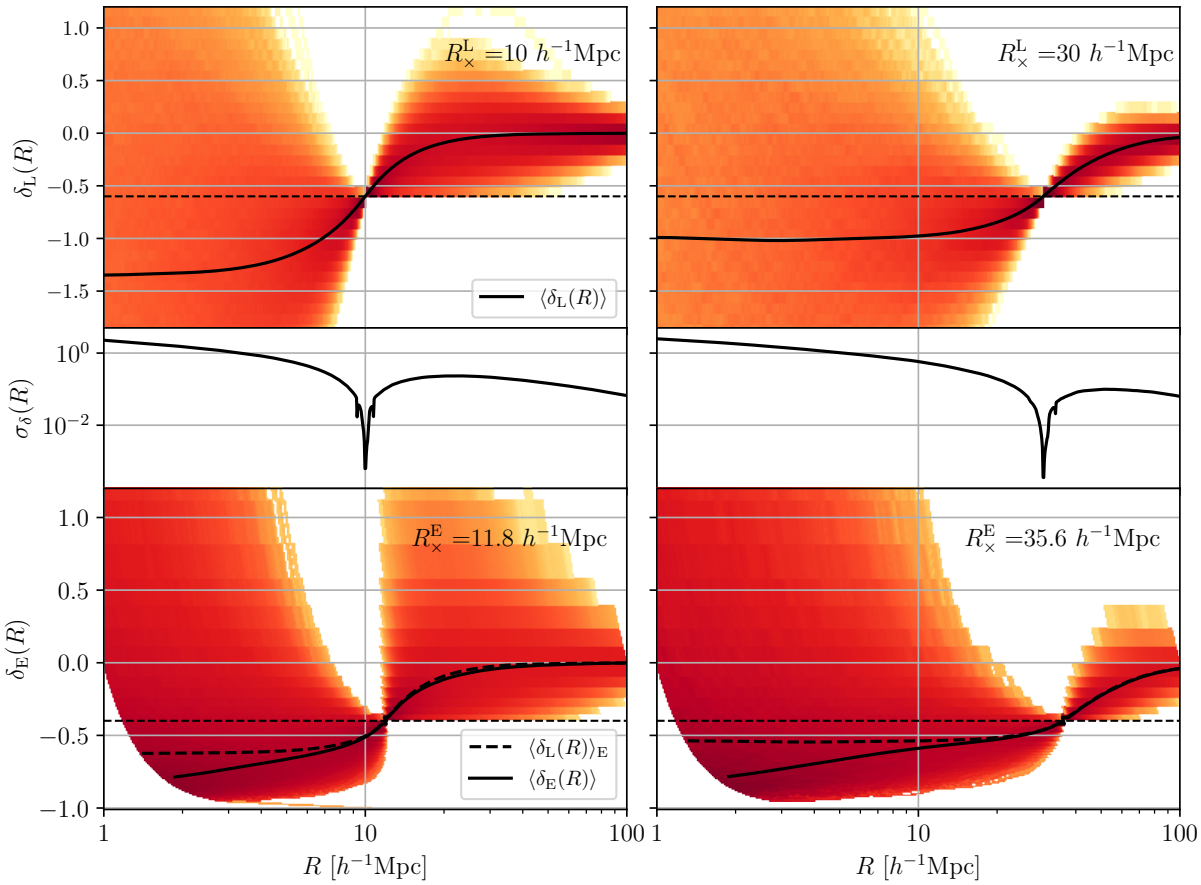

Figure 2.11: Mean value and distribution of density contrast paths that form voids at Lagrangian radius $R_\times = 10h^{-1}$Mpc (left panel) and $R_\times = 30h^{-1}$Mpc (right panel). The results are obtained with the top-hat filter. The upper plot of both panels shows the mean path (black curve) and the distribution around it (orange area) in Lagrangian space. The middle plot shows the standard deviation of the trajectory distribution (see text). The lower panels show the Lagrangian results mapped in the Eulerian space assuming spherical symmetry; the black dashed curve represents the mean Lagrangian path mapped in Eulerian space, the solid curve is the mean path inferred from the path distribution mapped in Eulerian space.

2.5.3 Lagrangian void density profile

The Langevin equations give the possibility to follow the path $\delta(R)$ along R for each stochastic realization. This allows to consider the whole trajectory $\delta(R)$ that at a given $R_\times$ forms a void in Lagrangian space. This quantity represents the extrapolated Lagrangian density contrast filtered on a sphere with radius R (in the top-hat case) around the position $\mathbf{q}$, that is the void position in the excursion-set framework. From the sample of the paths forming voids between $R_\times$ and $R_\times + {\rm d}R$, i.e. $\delta(R|R_\times)$ we can obtain the expectation path and the distribution around it. Fig. 2.11 shows the results for paths forming voids at two different Lagrangian radius bins, around 10 and 30 h^{-1}Mpc. The black curve of the upper plot of each panel shows the mean density contrast path in Lagrangian space, while the orange area shows the distribution of all the paths that satisfy the void formation condition in the given radius bin. Note that the density contrast can assume values lower than -1, this is because we are considering the linear extrapolated density contrast. The orange distribution is a histogram for each radius bin R,

with a resolution of 0.1 in density contrast. The middle plot of each panel shows the standard deviation around the mean path in Lagrangian space, $\sigma_\delta(R)$, computed fitting a Gaussian distribution, that, given the resolution of the histogram, well reproduces the distribution around the mean path for each R value. Note that the distribution of paths cannot exceed δ_v and δ_c values for $R > R_\times$ by construction. It follows that the standard deviation around the mean path reaches the minimum value in the radius bin of the crossing, since all the paths considered have reached the threshold value between $R_\times - \Delta R$ and $R_\times$. In this bin $\sigma(\delta) \to 0$ as $\Delta R \to 0$. After the threshold crossing, $\sigma(\delta)$ increases in quasi-logarithmic behavior with respect to R.

The lower panels show the Lagrangian results mapped in Eulerian space assuming spherical symmetry. The map is a nonlinear function of both density contrast and radius. Note that the first crossing radius expands from 10 to 11.6 h^{-1}Mpc and from 30 to 35.6 h^{-1}Mpc respectively, due to mass conservation expressed via Eq. (2.121). The linear threshold $\delta_v = -0.6$ is mapped in $\delta_v^{NL} \simeq -0.4$ via Eq. (2.17) and Eq. (2.22). The Lagrangian to Eulerian map greatly modifies the distribution of paths. In Lagrangian space, at each radius bin, the distribution is well fitted by a Gaussian; nevertheless, the Lagrangian negative values in the range (0,-∞) are compressed into the Eulerian range between 0 and -1, while the positive values are stretched from (0,1.686) to (0,∞), ending in a strongly skewed distribution. At the same time, the corresponding radius is stretched for underdensities and decreased for overdensities. The combination of these two effects maps the Lagrangian density contrast distribution at fixed R_L in a distribution along a curved set of points in the Eulerian (δ, R) plane. Note that the path distribution changes according to $P_L(\delta_L, R_L)d\delta_L dR_L = P_E(\delta_E, R_E)d\delta_E dR_E$. The black curves in the lower panels show the mean density contrast path. More specifically, the dashed line shows the mean Lagrangian path mapped in Eulerian space, the solid curve is the mean path computed from the path distribution obtained after the Lagrangian to Eulerian mapping. It can be seen that for radii smaller than the crossing radius, the two ways to compute the expected density contrast path start to provide different results, and the difference progressively grows as the radius decreases. This shows that the averaging procedure and the Lagrangian to Eulerian map do not commute.

This result shows how powerful the excursion-set framework is: by solving the Langevin equations it can predict the void density profile of voids. Note that this result, together with the void-halo correlation function, it can provide the relationship between the halo distribution in voids and the void density profile. However, there are at least two subtleties to be aware of. The first concerns the excursion-set void center position in Lagrangian space. According to void definition in the excursion-set framework, this position is not the minimum of the filtered density field, not even other typical void center definitions, such as the barycenter. The excursion-set void position is correlated with the minimum of the filtered field; nevertheless, this correlation has to be investigated. The second point is related to the first one and concerns the Lagrangian to Eulerian map. The spherical symmetry used in Eq. (2.121) and in Sect. 2.1 is accurate for large radii, but it becomes less and less accurate for radii smaller than $R_\times$.

2.6 Conclusion

In this chapter, we explored the void and halo statistical properties derived within the excursion-set framework; in particular, we discussed and presented the original results of our work. First of all, in Sect. 2.1, we considered the analogies and differences of the evolution and treatment of voids with respect to halos. For the first time in the literature, we showed that the shell crossing in void formation and treatment is almost irrelevant, and a perfect parallelism

between void and halo is not doable. This has consequences in detecting voids and in using theoretical predictions for observations, since the density contrast field describing a void in the Eulerian space can in principle always be mapped in the Lagrangian one, where we developed our theory. In Sect. 2.2, we introduced the fundamental quantity of the excursion-set approach, i.e. the filtered density contrast in Lagrangian space and its link to halos and voids. In that section, we discussed the meaning of the filtering function, pointing out the problems of using the sharp-k filtering function, on which the great majority of the literature is based. We then introduced the Langevin equations and their meaning in describing the statistical properties of halos and voids in Sect. 2.3, presenting some analytical results. In Sect. 2.4.1 we discussed how to numerically solve the Langevin equations for the general case, presenting our original implementation for the single and spatially correlated fields for the sharp-k filter case, considering the inter-step crossing, and for the general case, considering two different numerical implementations for which we explore the numerical stability. In particular, we extended the Cholesky decomposition to account for spatial correlation and discussed the numerical advantages and disadvantages of each presented method. At the end, in Sect. 2.5, we showed and discussed the results concerning the void size function, the halo mass function, and the two-point correlation functions. We focused on the results in the Lagrangian space obtained by solving the Langevin equations. Nevertheless, we also provided insights on the Eulerian mapping of these results. Concerning the void size function, we widely discussed the problem of the models available in the literature, showing that they rely on unphysical considerations, both in their formulation of the theory in Lagrangian space and in the Lagrangian to Eulerian mapping. For the first time in the literature, we used the top-hat filter within the excursion-set framework to model the Lagrangian halo-halo, void-void and void-halo correlation functions. We also provide a simple way to pass from the Lagrangian to Eulerian halo-halo correlation function. As a last point, for the first time we showed that the Langevin equations allow us to go beyond the 1- and 2-point statistics, predicting the Lagrangian profile for voids defined according to the excursion-set framework.

In this work, we widely explored the statistical properties of voids and halos in the Lagrangian space, whereas their map to the Eulerian space is at an initial stage. The following steps consist in exploring the Lagrangian to Eulerian map for voids, both theoretically, considering approximations beyond the spherical symmetry and using cosmological simulations to directly track dark matter particles in voids from their initial position and shape to the final ones.

Chapter 3

Cosmic voids in simulations

The primary aim of modern and future galaxy surveys is to investigate the late time acceleration of the Universe, with the goal to better understand what causes it. Aside from the dark energy problem, future cosmological surveys will reach an unprecedented precision in measuring the cosmological parameters, allowing to detect the effects of neutrino mass on the evolution and distribution of perturbations in the Universe. These effects are related to the total neutrino mass, i.e. the sum of the neutrino mass eigenvalues, and we are close to measure it with cosmological observations.

Both dark energy[1] and massive neutrinos impact the expansion of the background homogeneous Universe and the growth of perturbations. Cosmic voids are a specific kind of matter density perturbations, constituting a unique cosmological probe. They span a large range of scales and are underdense in matter density, these features make them suited to study dark energy and massive neutrinos as well. Moreover, they do not undergo extremely nonlinear gravitational evolution, as halos. It follows that the fingerprint of the physics driving their evolution is not erased by virialization. The cosmological exploitation of cosmic voids is also promising for another reason. The constraining power of voids is currently limited by the geometry and volume of surveys: voids are large, and in order to have enough statistics we need an extremely large continuous survey volume. Moreover, to study in detail the clustering properties of cosmic voids, we need spectroscopic surveys with high enough galaxy density detected, which translates in good enough spatial resolution, capable of mapping the distribution of galaxies within voids. Future galaxy surveys, such as *Euclid*, will make available a large contiguous fraction of the sky area and redshift detection up to high redshift, fulfilling the requirements to make cosmic voids an effective novel powerful cosmological probe.

The extraction of cosmological information from voids requires a detailed study of their properties and of their sensitivity to cosmological models and parameters. For this reason, in this chapter we investigate the properties of cosmic voids in large cosmological simulations, reproducing the volume and tracer distribution of a *Euclid* like survey. In this study, we consider various void observables, exploring their sensitivity to dark energy and massive neutrinos, how they are impacted by the tracer, and how to theoretically model their features. In particular, we focus on the void size function and the void density profile. We choose these two observables because they are the most simple to measure, they are widely used in the literature, they have a

[1] In this section we use the term "dark energy" in the wide sense of what causes the late time acceleration of the Universe, including under this term the cosmological constant, the vacuum energy, the proper dark energy field, and modified theories of gravity.

$\sum m_\nu$ [eV]	f_ν	$m_{\rm p}^{\rm cdm}$ [$h^{-1}M_\odot$]	$m_{\rm p}^{\nu}$ [$h^{-1}M_\odot$]
0	0	8.27×10^{10}	0
0.16	0.012	8.17×10^{10}	0.99×10^{9}
0.32	0.024	8.07×10^{10}	1.98×10^{9}

Table 3.1: Summary of cosmological parameters and derived quantities for DEMNUni simulation concerning the total neutrino masses implemented. The values $\Omega_{\rm b} = 0.05$, $\Omega_{\rm m} = 0.32$, $h = 0.67$, $n_{\rm s} = 0.96$ are shared by all the models.

clear physical interpretation, and, therefore, are easier to theoretically model.

This chapter is organized as follows. We present the set of simulations and the void finding algorithm used in Sect. 3.1; we then consider the abundance of voids in Sect. 3.2, discussing its general properties and the dependence on the tracer (Sect. 3.2.1), we then explore its sensitivity to the dark energy equation of state and the total neutrino mass, considering also geometrical effects (Sect. 3.2.2), and we present a possibility to connect the theory developed in Chapter 2 to observed voids (Sect. 3.2.3). In Sect. 3.3 we focus on the density profile of voids, extensively discussing how to estimate it in a discrete tracer distribution (Sect. 3.3.1), we then consider the relationship between matter and halo within voids, showing the properties of halos in voids and theoretically modeling halo bias in voids (Sect. 3.3.2), we conclude the section considering the sensitivity of the void density profile to dark energy and massive neutrinos Sect. 3.3.3; we then draw our conclusions in Sect. 3.4.

3.1 Voids in cosmological simulations

This work is based on the analysis of "Dark Energy and Massive Neutrino Universe" set of simulations[2] (DEMNUni) [145, 146]. The DEMNUni simulations have been conceived for the testing of different probes, including galaxy surveys, CMB lensing, and their cross-correlations, in the presence of massive neutrinos and dynamical dark energy. To this aim, this set of simulations is characterized by a volume big enough to include the very large-scale perturbation modes, and, at the same time, by a good mass resolution to investigate small-scales nonlinearities and neutrino free streaming. Moreover, for the accurate reconstruction of the light-cone back to the starting redshift of the simulations, it has been used an output time spacing small enough that possible systematic errors, due to the interpolation between neighboring redshifts along the line of sight, result to be negligible. We note that the DEMNUni simulations have a volume and resolution mimicking the data expected from large surveys such as the spectroscopic Euclid wide survey, hence allowing to explore the population of relatively large voids (which are not captured by smaller simulations), and assess their constraining power for cosmology. For this reason, in the last few years they have been used to explore the applications of cosmic voids for cosmology [97, 147, 148]. This work aims to capitalize on cosmic void properties measured in simulations and to their sensitivity to the dark energy equation of state and to the sum of neutrino masses.

The simulations were performed using the TreePM-SPH code GADGET-3, an improved ver-

[2] https://www.researchgate.net/project/DEMN-Universe-DEMNUni

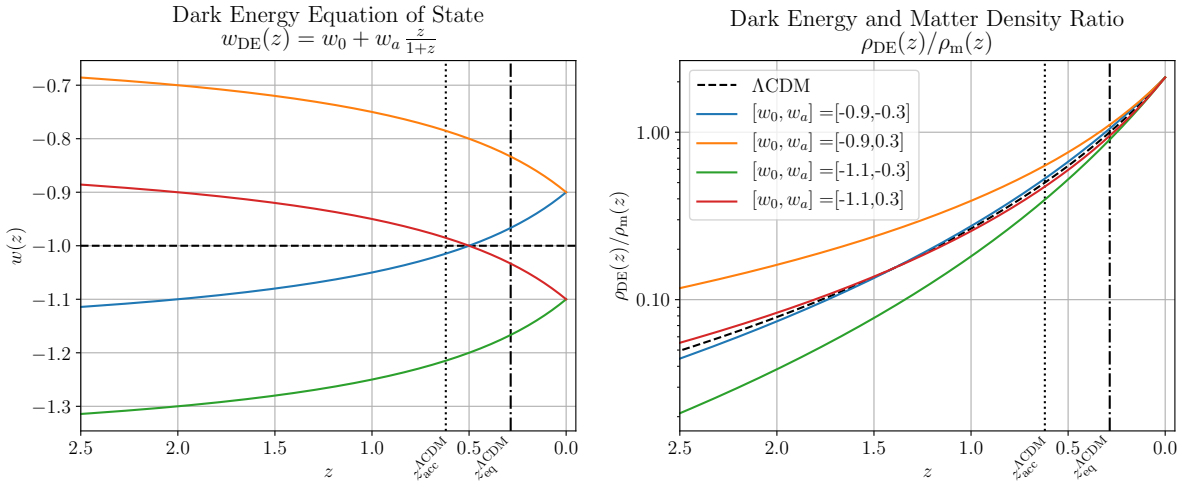

Figure 3.1: Evolution of the dark energy equation of state (left panel) and of the dark energy density $\rho_{\rm DE}$ (right panel), both as a function of redshift z. The parameter $\rho_{\rm m}$ is the matter density. In both panels the vertical lines represent respectively the redshift value at which the ΛCDM universe begins to accelerate its expansion (dotted line, $z_{\rm acc}^{\Lambda\rm CDM} = 0.620$), and the redshift of the matter-dark energy density equality (dash dotted line, $z_{\rm eq}^{\Lambda\rm CDM} = 0.286$), both for the ΛCDM model. Image from [97].

sion of GADGET-2 [149], modified to account for the presence of massive neutrinos [150]. The simulations are characterized by a softening length $\epsilon = 20\,h^{-1}$Kpc, and a comoving volume of $(2\,h^{-1}\text{Gpc})^3$ filled with 2048^3 dark matter particles and, when present, 2048^3 neutrino particles. All simulations started at redshift $z = 99$, producing 62 different time outputs, logarithmically spaced in the scale factor $a = 1/(1+z)$, down to $z = 0$. The simulations have a Planck 2013 [151] baseline ΛCDM reference cosmology and flat geometry. The set is composed of 15 simulations, considering 5 different dark energy equations of state, for the cosmological constant plus 4 dynamical equations of state for each neutrino mass considered, i.e. $\sum m_\nu = 0$, 0.16, 0.32 eV. The four dark energy variants are parameterized with the Chevallier–Polarski–Linder (CPL) parameterization [58, 59]

$$w(a) = w_0 + (1-a)w_{\rm a} \quad \Rightarrow \quad w(z) = w_0 + \frac{z}{1+z} w_{\rm a}\,, \tag{3.1}$$

with parameter values chosen on the boundaries of the 2015 Planck constraints [152]: the 4 combinations of $(w_0, w_{\rm a})$ with $w_0 = [-0.9,\ -1.1]$ and $w_{\rm a} = [-0.3,\ 0.3]$. The redshift evolution of the dark energy equations of state density, for the parameters implemented in DEMNUni simulations, are represented in Fig. 3.1. The cosmological parameters are based on ΛCDM cosmology, the entire set has $\Omega_{\rm m} = \Omega_{\rm cdm} + \Omega_{\rm b} + \Omega_\nu = 0.32$, in the presence of massive neutrinos the baryon density is kept fixed $\Omega_{\rm b} = 0.05$ and the other two density parameters are changed accordingly. Table 3.1 resumes the parameters used in the presence of massless and massive neutrinos, where $f_\nu = \Omega_\nu/\Omega_{\rm m}$.

Dark matter halos are identified using a friends-of-friends (FoF) algorithm [31] applied to dark matter particles only, with minimum number of particles fixed to 32, corresponding to a mass of $\sim 2.5 \times 10^{12}\,h^{-1}M_\odot$, and a linking length of 0.2 times the mean particle separation. FoF halos are further processed with the SUBFIND algorithm [153, 154] to get a subhalo catalog. With this procedure, some of the initial FoF parent halos are split into multiple sub-halos, with the result of an increase of the total number of identified objects, and of a lower minimum mass

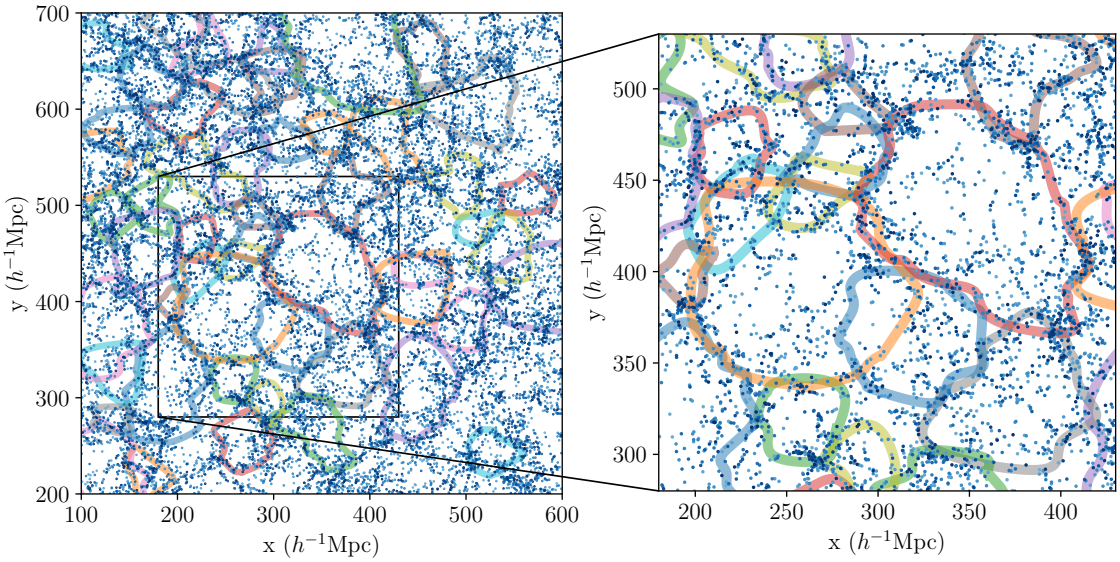

Figure 3.2: Projected void boundaries (colored lines) and halo field (blue dots represent halos, darkness is a function of halo mass) in a slice of the ΛCDM simulation at z=1.05. Image from [97].

limit. The minimum number for a subhalo is set to 20 particles in order to constitute a valid subhalo catalog, corresponding to $\sim 1.7 \times 10^{12}\, h^{-1} M_\odot$. Note that the SUBFIND algorithm is an estimator of physical halos, therefore in the following with the term "halo" we will refer to the objects identified by SUBFIND algorithm.

To identify voids and build void catalogs, we used the second version of "Void IDentification and Examination" (VIDE)[3] public toolkit [155]. VIDE is based on the tessellation plus watershed void finding formalism [156] with the Delaunay tessellation/DTFE [157] and subsequently used by ZOBOV [158] with the Voronoi tessellation of the tracer particle population, to estimate the density field based on the underlying particle positions. The algorithm first groups nearby Voronoi cells into zones, corresponding to local catchment "basins", that are identified as voids. VIDE can optionally merge basins to form larger voids with the watershed transform, we choose to do not explore this possibility in order to leave the void finding algorithm parameter free.

For each simulation, we launched VIDE on 5 different snapshots, corresponding to $z = 0$, 0.48, 1.05, 1.46, 2.05, on the particle and subhalo distribution. To find voids in dark matter particle distribution, we randomly diluted the original catalogs at 1.5%, ending up in $\sim 1.29 \times 10^8$ particles, to save computational resources. Moreover, since the void sizes we want to study are the ones detectable in upcoming galaxy surveys, such as *Euclid*, the dilution does not impact the spatial resolution required for the analysis. We then build four different catalogs for voids detected in halos, fixing four different minimum masses to the halo distribution, corresponding to $2.5 \times 10^{12} h^{-1} M_\odot$, $10^{13} h^{-1} M_\odot$, $2.5 \times 10^{13} h^{-1} M_\odot$, and $10^{14} h^{-1} M_\odot$.

Watershed voids are not spherical, they correspond to a relative minimum in the density field estimated via the Voronoi tessellation, and they extend within the watershed formed by the ridge around it (see the representation of Fig. 3.2). Therefore, this void definition is sensitive to the tracer density. It is worth noting that this is not a source of ambiguity or arbitrariness

[3] http:www.cosmicvoids.net

for voids detected in halo distributions, because the discreteness of tracers and its number density have a physical origin. This is valid for any discrete tracer with physical origin, such as galaxies, clusters, etc. On the other hand, voids detected in the dark matter particle distribution depend on the spatial resolution given by the particle density, which is numerical and not physical. Since the void finder identifies each relative minimum of the density field, the higher spatial resolution entails better identification of substructures and, therefore, a higher number of relative minima. Therefore, in the dark matter particle case, we must consider that the void scale is impacted on the numerical spatial resolution. To face this problem, it is possible to consider a threshold in `VIDE` in order to merge basins into parent voids, nevertheless, since we are more interested in voids detected in halo distribution, we do not explore this possibility.

We describe the void properties according to three parameters given by `VIDE` output. The first is the void size, that is measured via the effective radius, i.e. the radius corresponding to a sphere with the same volume of the void,

$$R_{\rm eff} = \frac{3}{4\pi}\left[\sum_i V_i\right]^{1/3}, \tag{3.2}$$

where V_i is the volume of the $i^{\rm th}$ Voronoi cell building up the void. The other quantity is the void center. In the literature there are two approaches to define the void position for watershed voids: the location of the density minimum, estimated in various ways, or the barycenter. We prefer the second option for various reasons. First, the position of the minimum of a void is numerically unstable, since it depends on few or only one particle, depending on the methodology used, so it can be more affected by sampling artifacts and the resulting position may be Poissonian dominated. On the other hand, the void is identified by all the particles belonging to it, the use of information from the whole object allows for a more robust identification of the void center. This is particularly important for the analysis we want to perform; in particular, a robust study of the density profile and statistics of voids requires a robust void center definition. Moreover, for void profiles, the barycenter is a more suitable point of reference to measure it, as we discuss in 3.3. `VIDE` estimates the barycenter by weighting the volumes of the contributing Voronoi cells:

$$\mathbf{X}_{\rm v} = \frac{1}{\sum_i V_i}\sum_i \mathbf{x}_i V_i\,, \tag{3.3}$$

where $\mathbf{x}_i$ and V_i are the position and Voronoi volume of each tracer particle i within the void.

The last quantity used in our analysis is the central density $\rho_{\rm c}$, defined as the mean tracer density within a sphere of radius $R_{\rm eff}/4$ around the volume-weighted barycenter. Since `VIDE` detects all the relative minima, we decided to keep only the ones that correspond to underdensities, therefore we pruned the void catalog requiring $\rho_{\rm c} < 1$.

3.2 The void size function

The void size function, or void abundance, is the number density of voids as a function of their size. It is sensitive to cosmological parameters [97, 147, 159–162], and it can contribute in breaking the degeneracy between different models as they are subject to different systematics with respect to the halo mass function and the two-point halo/galaxy statistics. Moreover, the void size function is potentially more sensitive since voids are dark energy dominated objects [163, 164]. In Chapter 2 we widely discussed the theoretical void size function in the

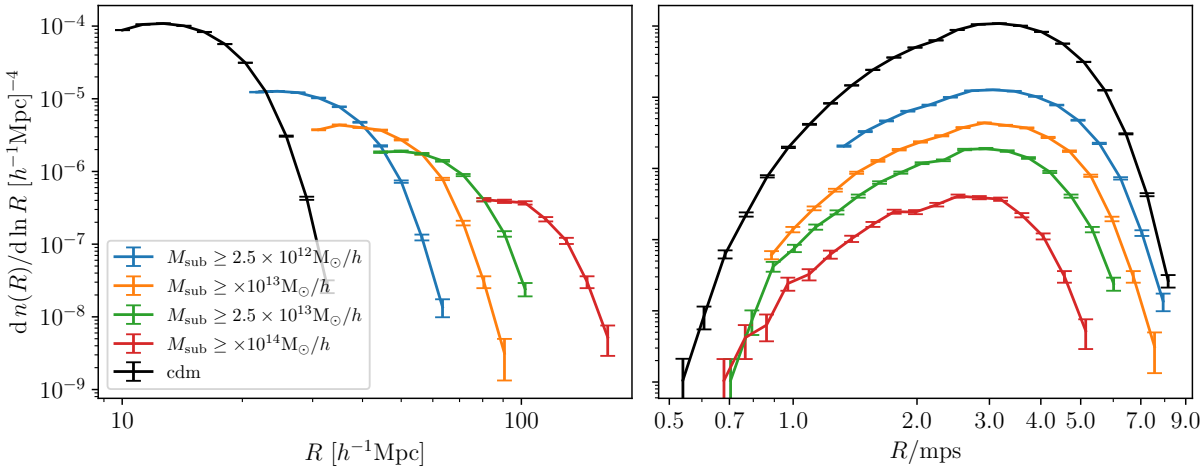

Figure 3.3: Tracer impact on the void size function. Both panels show the void size function measured in ΛCDM simulation in the $z = 0$ snapshot for voids detected in dark matter distribution (black), and halo distribution with various minimum subhalo mass, $2.5 \times 10^{12} h^{-1} M_\odot$ (blue), $10^{13} h^{-1} M_\odot$ (orange), $2.5 \times 10^{13} h^{-1} M_\odot$ (green), $10^{14} h^{-1} M_\odot$ (red). The error bars are Poissonian. The left panel shows the void abundances as a function of void effective radius considering voids with $R \geq 2.5 \times$ mps, the right panel shows the void abundances for all detected voids where the x-axis is in unit of the mean particle separation.

excursion-set framework, in this section we will mainly focus on the measured void size function in simulations (Sect. 3.2.1), to directly focus on its sensitivity to dark energy and neutrino total mass (Sect. 3.2.2), and at the end we discuss its relations with the theoretical model (Sect. 3.2.3).

3.2.1 General properties

In this subsection, we discuss how the void size function depends on the tracers and on the redshift. Fig. 3.3 shows the void size function measured in the snapshot at $z = 0$ in ΛCDM cosmology, for various tracers: dark matter particles (black) and halos with different mass cuts, $2.5 \times 10^{12} h^{-1} M_\odot$ (blue), $10^{13} h^{-1} M_\odot$ (orange), $2.5 \times 10^{13} h^{-1} M_\odot$ (green), and $10^{14} h^{-1} M_\odot$ (red). The mean particle separation (mps) is defined as $(N_{\mathrm{p}}/L^3_{\mathrm{box}})^{-1/3}$, the value is $\sim 4h^{-1}$Mpc for diluted dark matter particles and $\sim$ 8, $\sim$ 12, $\sim$ 17, and $\sim 32h^{-1}$Mpc for the halo density field, with the corresponding minimum halo mass listed above, respectively. On the left panel, the void abundances are represented as a function of the effective radius. Each population of tracers entails a different spatial resolution in mapping the underlying density fluctuations. More massive halos trace the background density field on larger scales, therefore more massive halos detect larger voids. The right panel shows the void size function as a function of the effective radius in units of the mps. This allows to better understand the impact on the spatial resolution. It can be seen that the peak of the void size function is around $3 \times$ mps for each tracer considered. We stress the fact that this is due to the sparsity of tracers, but since the number density of halos depends on the physics behind the halo formation, the suppression of small voids in the halo field has a physical origin. This is not true for voids in the dark matter particle distribution, where the number density is fixed numerically.

Fig. 3.4 shows the redshift dependence of the void size function in the halo distribution with

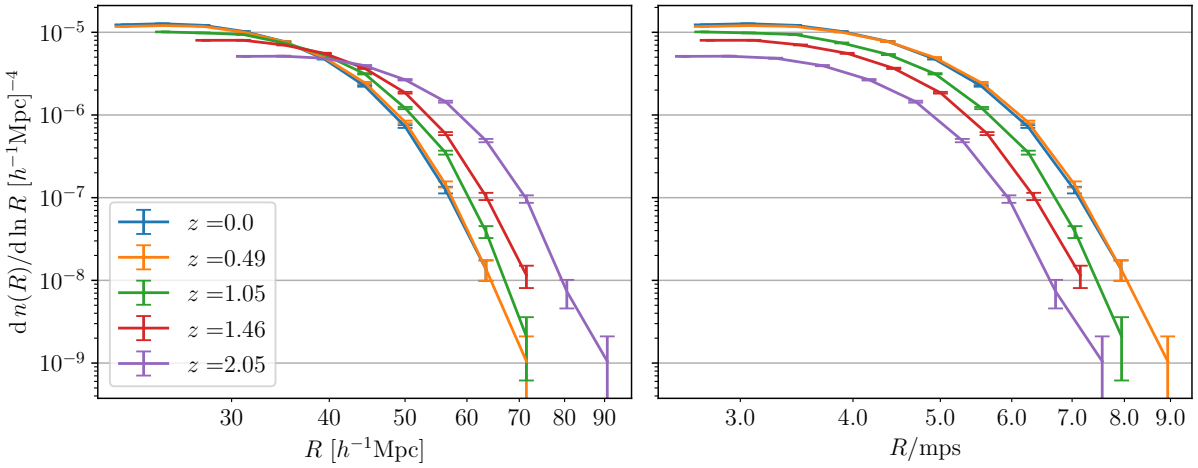

Figure 3.4: Redshift evolution of the void size function measured in the halo distribution with minimum mass of $2.5 \times 10^{12} h^{-1} M_\odot$ in ΛCDM simulation. The x-axis of the left panel is in h^{-1}Mpc units, while the left panel is in mps units.

a minimum halo mass of $2.5 \times 10^{12} h^{-1} M_\odot$ for voids with $R \geq 2.5 \times$ mps in ΛCDM cosmology. As before, the left panel shows the void size function as a function of the void effective radius in h^{-1}Mpc, while the right panel in mps units. This double representation helps to distinguish the two effects driving the evolution of the void size function with redshift. The first is that fixing the mass, halos at higher redshift are more biased with respect the ones at lower redshift, this means that they trace the underlying mass distribution at larger scales. This effect is visible in the left panel. On the other hand, perturbations, and so voids, grow as the redshift decreases, this is well visible expressing the void radius in mps units (right panel).

3.2.2 Sensitivity to dark energy and total neutrino mass

In this subsection we explore the sensitivity of the void size function of watershed voids to dark energy and to the total mass of neutrinos. The investigation of these effects on the void size function in DEMNUni simulations was explored in [97, 147], considering the impact of massive neutrinos [147] and dark energy [97] separately, both works without considering the redshift evolution. We now extend these studies by considering the combination of these effects, the impact of tracers, the redshift dependence, and geometrical distortions modifying the observed void size function.

We start considering the effect of the neutrino mass on the void abundance for voids detected in the dark matter distribution. Fig. 3.5 shows the comparison of void size function measured in the presence of massive neutrinos, $\sum m_\nu = 0.16$ and 0.32 eV, for cosmology with cosmological constant, i.e. $(w_0, w_\mathrm{a}) = (-1, 0)$, with respect to the void size function of ΛCDM. This is done for the five redshift snapshots analyzed. The quantity represented is

$$\left[\frac{\mathrm{d}n}{\mathrm{d}\ln R}\right]_{\sum m_\nu} \left[\frac{\mathrm{d}n}{\mathrm{d}\ln R}\right]^{-1}_{\Lambda\mathrm{CDM}} - 1 \tag{3.4}$$

and the error bars are Poissonian, i.e. they are estimated by counts. For a rough comparison with the upcoming data, note that a box has a volume of 8 $(h^{-1}\mathrm{Gpc})^3$, corresponding to the volume of a shell in the *Euclid* survey centered at $z \sim 1$ and spanning $\pm\Delta z$ with $\Delta z = 0.1$. Note

that the ratio of the two void size functions is equal to the ratio of the number of voids in the same radius bin. Nevertheless, this is not true when geometrical distortions are considered, as we will discuss later, therefore we prefer to express this quantity as the ratio of the void size functions from now.

It can be seen that the presence of massive neutrinos suppresses the number of large voids and enhances the number of smaller ones. To correctly interpret this result, we recall that the objects we are considering are the relative minima of the density field and the basins to which they belong. Massive neutrinos suppress perturbations on scales smaller than the free-streaming length, so the mean density contrast of fluctuations is lower in the presence of massive neutrinos with respect to the massless case. Nevertheless, their impact on the number density of relative minima in the dark matter density field and on the size of the corresponding basins is not a direct consequence, therefore it is not easy to infer. The suppression or enhancement in number does not refer to the density contrast of these objects, but to the number density of minima and the size of their basins. There is no direct relationship between the number of minima and the depth of voids. We will explore the relationship between watershed voids and the density contrast of the corresponding fluctuations in Sect. 3.2.3 and in Chapter 4. Note that the void size function in the presence of massive neutrinos is distinguishable from the massless case for each redshift. This result is particularly interesting given the fact that the minimum total neutrino mass allowed by oscillations is roughly one-third of the minimum neutrino mass implemented in DEMNUni simulations [51–53]. Moreover, the minimum mass considered in DEMNUni simulations is close to the upper limit measured by Planck [50], showing the importance of studying voids in the upcoming galaxy surveys.

Fig. 3.6 shows the ratio of the void size functions measured in the dark matter distribution for the four parameterizations of the CPL dark energy equation of state and massless neutrinos with respect to ΛCDM. The plot is organized as the previous one. Note that only two out of the four parameterizations are distinguishable from ΛCDM. Comparing these results with Fig. 3.1 it can be seen that the void size functions for cosmology with a dark energy equation of state and density that is more close to ΛCDM remain degenerate, the other two are distinguishable.

We consider now the effect of massive neutrinos and dark energy equation of state on the void size function for voids detected in the halo distribution. Fig. 3.7 and Fig. 3.8 show respectively the comparison of the void size functions in the presence of massive neutrinos and $(w_0, w_a) = (-1, 0)$ with respect to the ΛCDM case and of the void abundances in presence of dynamical dark energy and massless neutrinos with respect to ΛCDM. We report here only the results for halos with $M \geq 2.5 \times 10^{12} h^{-1} M_\odot$, since the other mass cuts show analogous features. The first thing to notice is that the behavior of the relative void size functions is inverted with respect to voids detected in the dark matter distribution. Since we are considering watershed voids, where the void depth does not enter in the void definition, an insight of the physics driving this behaviors is complex. First of all, we notice that voids detected in dark matter distribution are a different class of objects with respect to the ones in halos distribution, note that the typical scale of watershed voids is very different for instance. One of the processes driving this behavior can be addressed to the change in the number of halos formed. A different number density of halos with respect to ΛCDM corresponds to a different bias, in this way halos with the same minimum mass, but in the presence of different dark energy equations of state, track the underlying matter density at different scales. This probably is not the only reason of the void size functions behavior, more considerations will be investigated in considering the void density profiles, Sect. 3.3. Note that the impact of the neutrino mass is always detectable for

the neutrino masses considered. On the dark energy side, the impact on the void size function is reduced in the halos distribution with respect to dark matter voids, nevertheless the two dark energy equations of state that are distinguishable in the dark matter distribution remain distinguishable also in the halos distribution, even if the relative difference is reduced. Note that in the presence of dark energy, all the void size functions remain degenerate with ΛCDM at low redshift. The physical origin of this behavior is complex, since the watershed voids detected in the halo distribution depend both on void formation in dark matter distribution and halo formation, that trace the underlying dark matter voids, both impacted by dark energy. A partial explanation of this behavior is probably related to the fact that the initial scalar amplitude for simulations with dynamical dark energy is set to keep the same σ_8 at $z = 0$, together with all the cosmological density parameters. This means that at $z = 0$ the linear power spectrum is expected to be the same for each dark energy equation of state implemented, while at higher redshift the relative differences increase. As a consequence, note that the degeneracy of the void size function corresponding to the two sets of CPL parameters more distant from ΛCDM breaks as the redshift increases. This shows the importance of studying the redshift evolution.

When analyzing real data, the void size function is impacted by geometrical effects. The position of tracers on which the void finder would be launched, is measured in spherical coordinates, where the distance variable is the redshift. The conversion from redshift to comoving position depends on the cosmological model and the parameters assumed. A wrong cosmology reflects in geometrical distortions, known as Alcock-Paczyński effect [165]. The impact of the distortion can be easily derived as [141, 166–170]

$$r'_{\parallel} = \frac{H(z)}{H'(z)}\, r_{\parallel} = q_{\parallel}^{-1}\, r_{\parallel}\,, \qquad r'_{\perp} = \frac{\chi'(z)}{\chi(z)}\, r_{\perp} = q_{\perp}^{-1}\, r_{\perp}\,; \tag{3.5}$$

where $r_{\parallel}$ and $r_{\perp}$ are the comoving distances between two objects at redshift z projected along the parallel and perpendicular direction with respect to the line-of-sight, $H(z)$ is the Hubble parameter and $\chi(z)$ the comoving distance defined in Eq. (1.48). The primed quantities refer to the calculation at the fiducial cosmology, assumed to map redshifts to comoving distances, the non-primed to the true cosmology. The first step of `ZOBOV` based void finders, such as `VIDE`, is to estimate the density field via the Voronoi tessellation. Note that the redshift positions of the minima and the cells belonging to each basin do not vary under a smooth monotonic map of the redshift position. This means that the redshift position of each void does not change with the cosmological models and parameters used to map redshifts in comoving distances. Furthermore, the volume of each Voronoi cell estimated in a fiducial cosmology is modified according to $V'_i = q_{\parallel}^{-1} q_{\perp}^{-2} V_i$. It follows that the void number does not change when a wrong cosmology is used to infer distances and the void effective radius R appears modified according to $R' = q_{\parallel}^{-1/3} q_{\perp}^{-2/3} R$. The map from redshift to comoving distance impacts the inferred survey volume and consequently the estimation of the number density of detected voids. The survey volume is defined as $\Omega_{[\mathrm{rad}]}[\chi(z_{\mathrm{out}})^3 - \chi(z_{\mathrm{in}})^3]/3$, where $\Omega_{[\mathrm{rad}]}$ is the solid angle of the survey in steradian while z_{in} and z_{out} is the redshift range considered. Since we work with snapshot boxes, we take into account the change in volume effectively, considering the distortion along one Cartesian axis, so the volume correction is implemented as

$$\frac{V'_{\mathrm{box}}}{V_{\mathrm{box}}} = \frac{\chi'^2(z)}{\chi^2(z)} \frac{H(z)}{H'(z)}\,. \tag{3.6}$$

Fig. 3.9 shows the impact geometrical distortions in the void size function measured in halo distribution assuming a ΛCDM fiducial cosmology with cosmological parameters of the ΛCDM

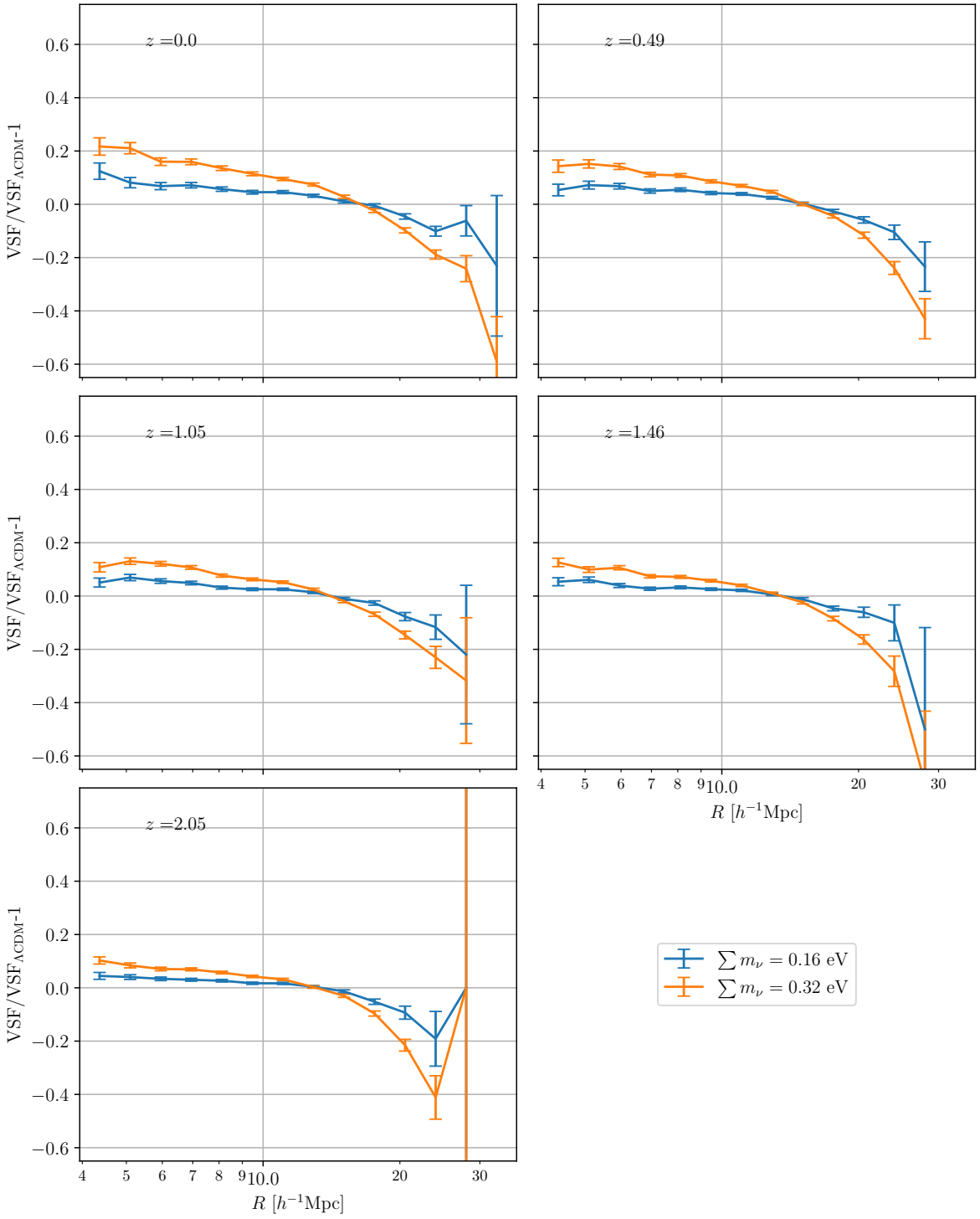

Figure 3.5: Relative void size function for voids detected in the presence of massive neutrino and cosmological constant with respect to the massless case. The voids at various redshifts are detected in dark matter distribution. "VSF" in the y-axis label stands for "void size function". The error bars are Poissonian.

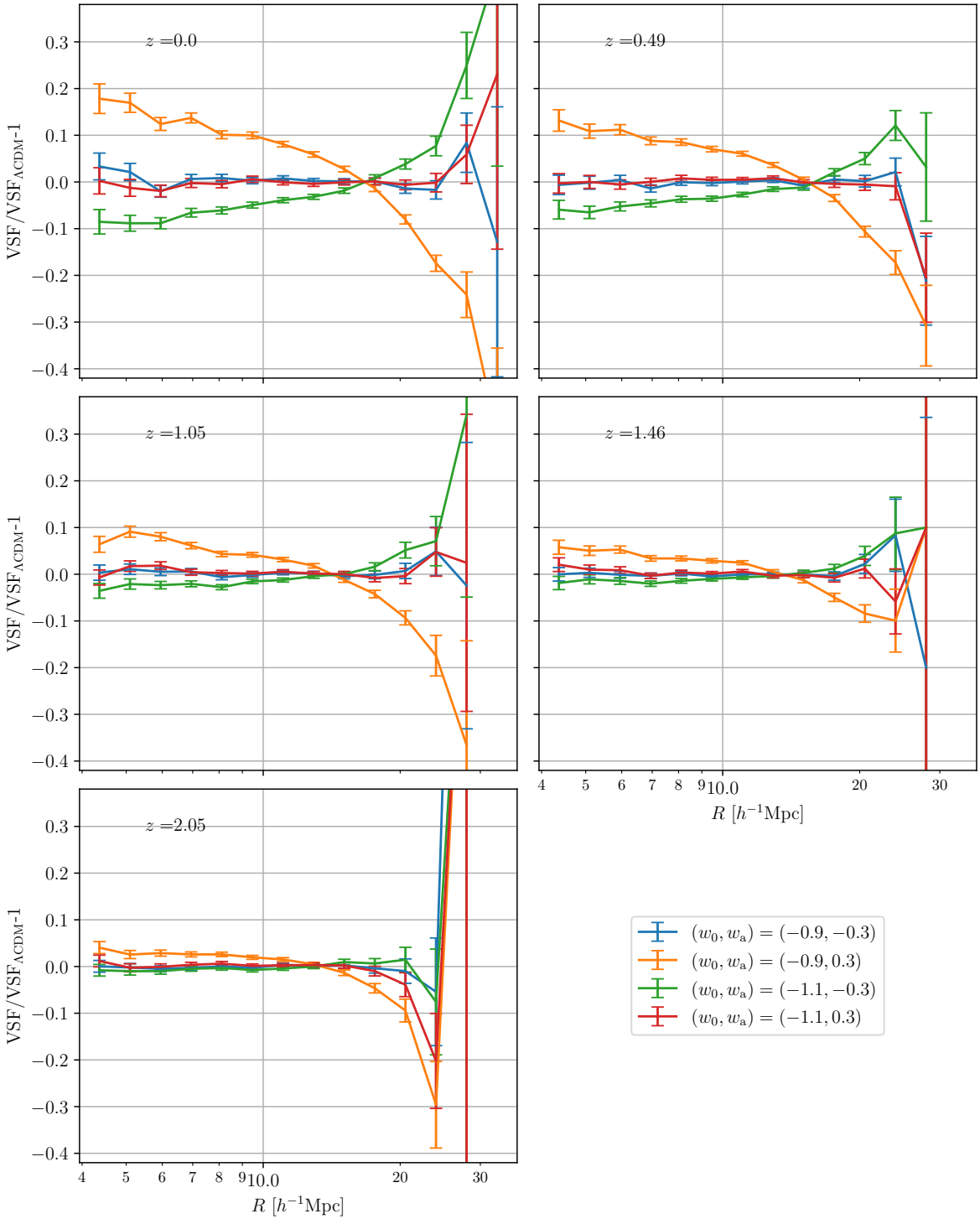

Figure 3.6: Relative void size function for voids detected in the presence of dynamical dark energy and massless neutrinos with respect to ΛCDM. The voids at various redshifts are detected in dark matter distribution. "VSF" in the y-axis label stands for "void size function". The error bars are Poissonian.

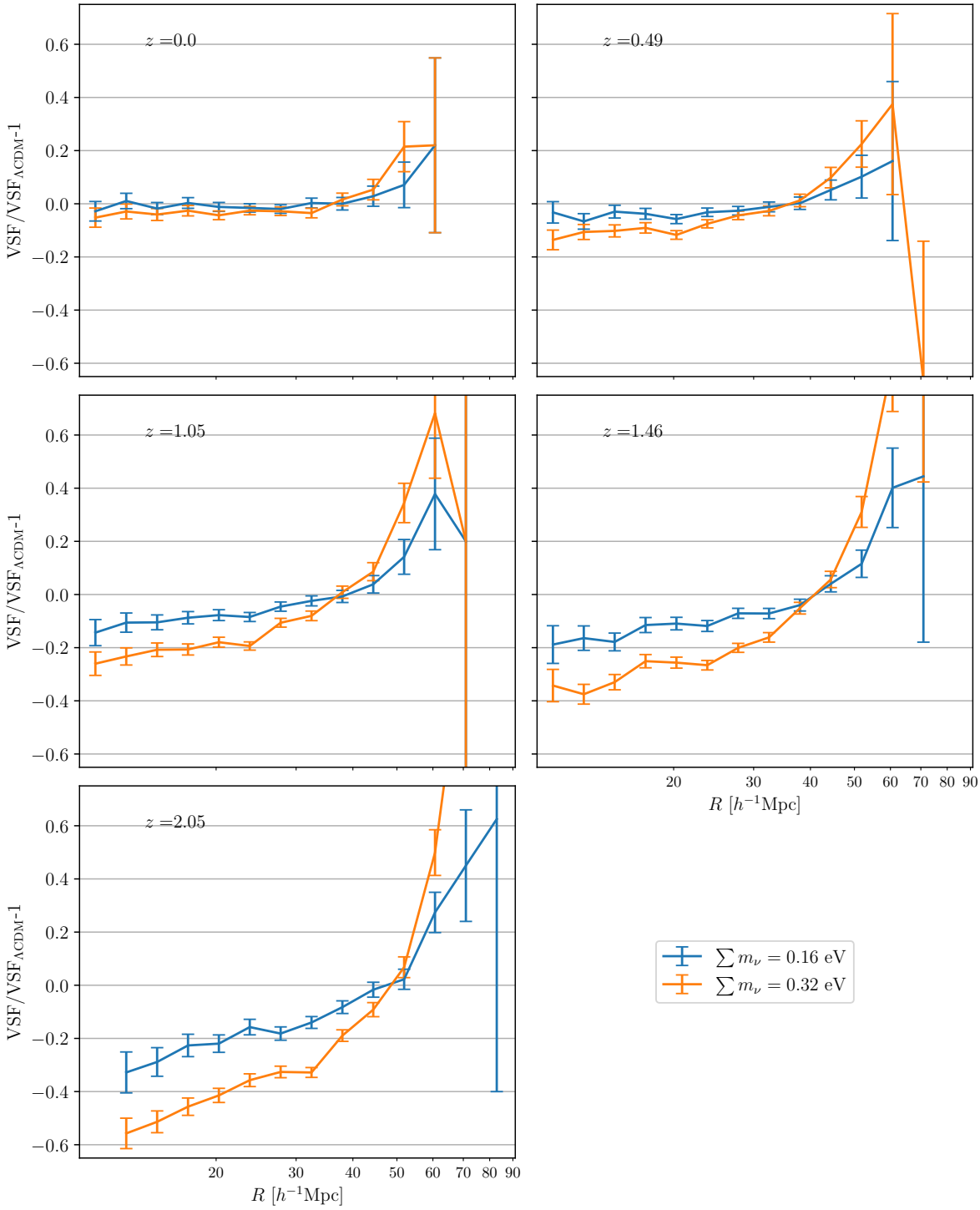

Figure 3.7: Relative void size function for voids detected in the presence of massive neutrino and cosmological constant with respect to the massless case. The voids at various redshifts are detected in the halo distribution with minimum subhalo mass of $2.5 \times 10^{12} h^{-1} M_\odot$. "VSF" in the y-axis label stands for "void size function". The error bars are Poissonian.

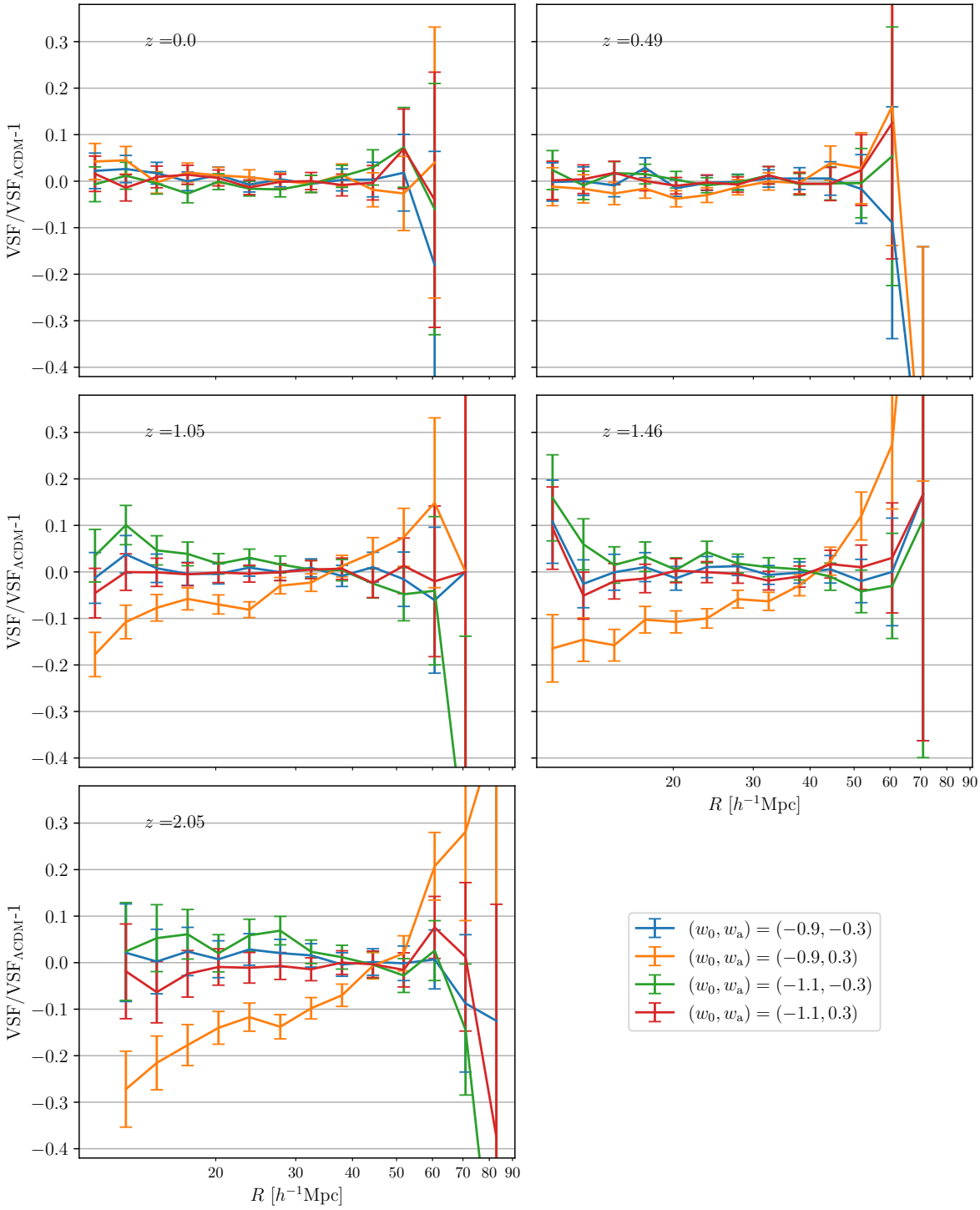

Figure 3.8: Relative void size function for voids detected in the presence of dynamical dark energy and massless neutrinos with respect to ΛCDM. The voids at various redshifts are detected in the halo distribution with minimum subhalo mass of $2.5 \times 10^{12} h^{-1} M_{\odot}$. "VSF" in the y-axis label stands for "void size function". The error bars are Poissonian.

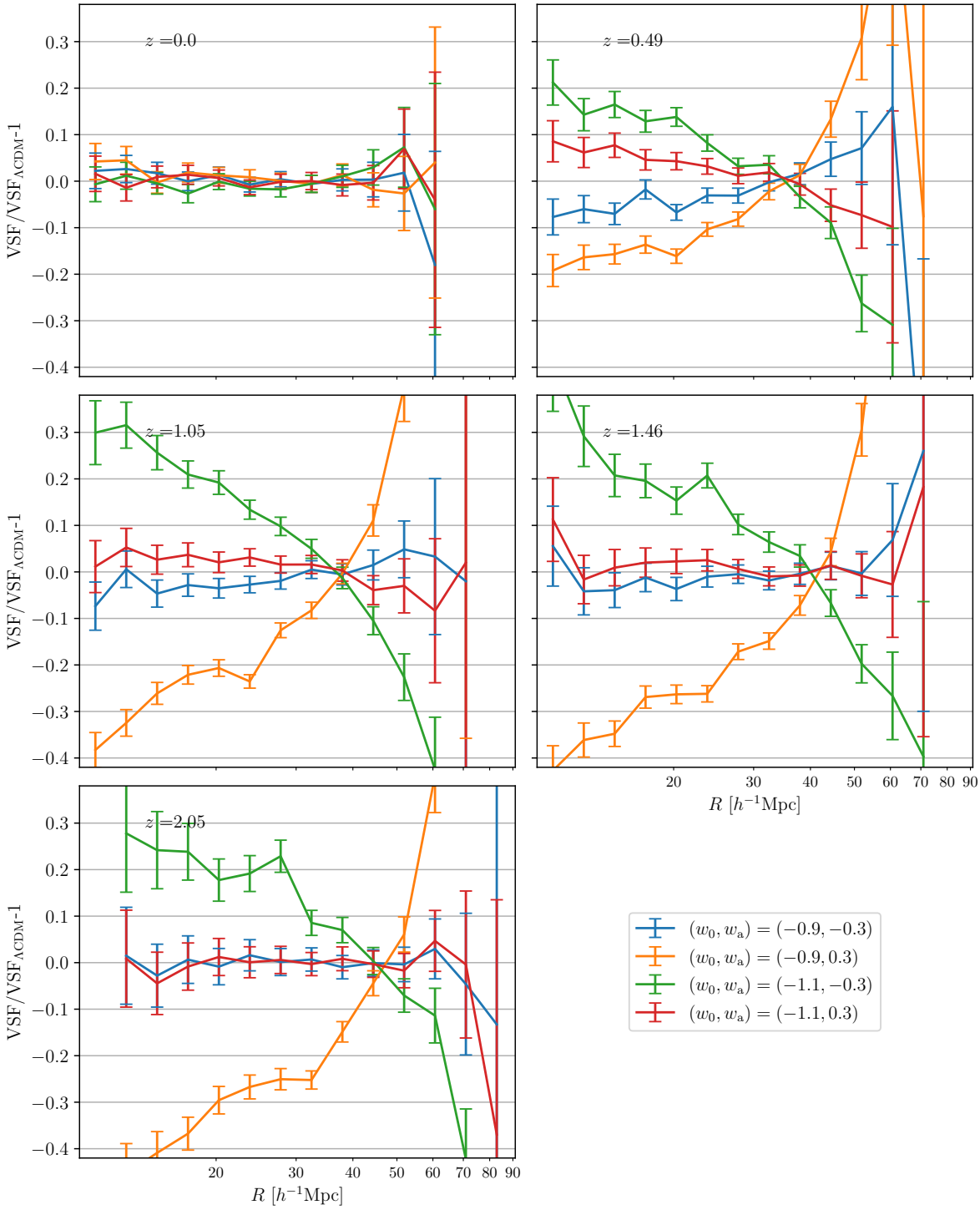

Figure 3.9: Relative void size function for voids detected in the presence of dynamical dark energy and massless neutrinos with respect to ΛCDM. The voids at various redshifts are detected in the halo distribution with minimum subhalo mass of $2.5\times10^{12}h^{-1}M_{\odot}$ and geometrical distortions are considered. "VSF" in the y-axis label stands for "void size function". The error bars are Poissonian.

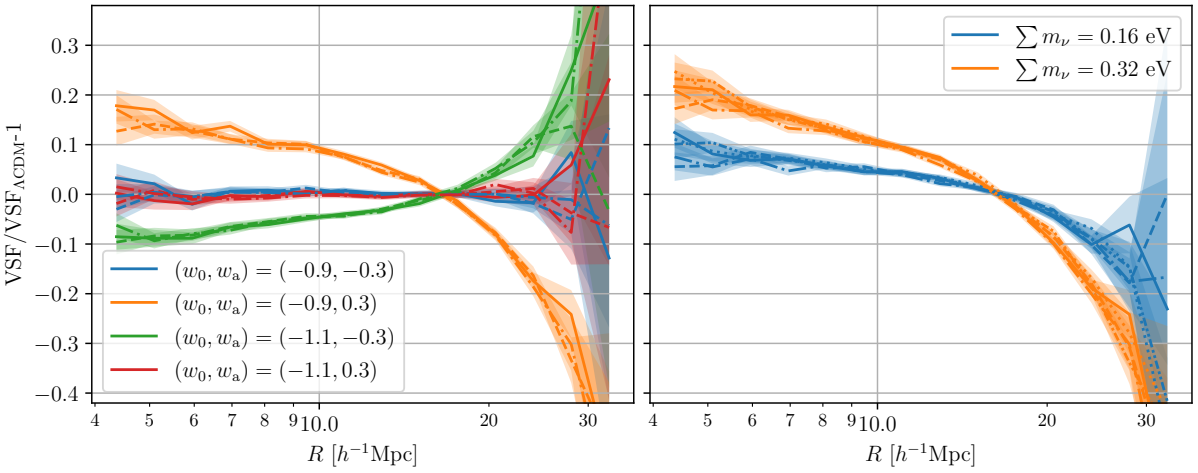

Figure 3.10: Separability of massive neutrinos effects from dark energy effects on the void size function in dark matter distribution at $z = 0$. For each fixed neutrino mass, the left panel shows the relative void size function with respect to $(w_0, w_a) = (-1, 0)$ equation of state. The color represents the dark energy equation of state, the line style the total neutrino mass to which they refer, i.e. massless (solid), 0.16 eV (dashed), 0.32 eV (dash-dotted). The right panel is analogous to the left, but fixes the dark energy equation of state and considers the relative void size function for the massive neutrinos case with respect to the massless case. The color labels the total neutrino mass, while the line style the dark energy equation of state to which they refer.

simulation. The changes in the observed void size function measured in various dark energy equations of state induced by geometrical distortions break the degeneracy. At redshift $z = 0$ there is no geometrical distortion, this is because the map between redshift and comoving distance converges to $\chi(z) \simeq cH_0 z$ and the entire DEMNUni set has $H_0 = 67 \mathrm{km\,s^{-1}Mpc^{-1}}$. The effect of geometrical distortions grows with redshift, enhancing the impact on the observed void size function where a fiducial cosmology is assumed. Note from Eq. (1.48) that massive neutrinos impact the comoving diameter distance through the Hubble factor defined by the first Friedmann equation, Eq. (1.26). At low redshift, when the radiation energy density contribution is negligible, massive neutrinos impact the Hubble factor via the matter parameter $\Omega_m(z)$, changing its value when they become non-relativistic. The transition happens at $1 + z = 1860\, m_\nu/1\mathrm{eV}$ [51, 52], so even for the minimum possible total neutrino mass, i.e. $\sum m_\nu \gtrsim 0.06\,\mathrm{eV}$, the transition occurs beyond the redshift range of galaxy surveys; therefore, the density of massive neutrinos contributes to the matter density parameter for all the redshift considered, i.e. $\Omega_m = \Omega_{cdm} + \Omega_b + \Omega_{m_\nu}$, with Ω_{m_ν} fixed after the transition. Since in DEMNUni simulations Ω_m is fixed to 0.32 in the entire set, massive neutrinos do not source geometrical distortions.

As a last point, we consider the combination of dark energy and massive neutrinos effects on the void size function. Fig. 3.10 shows that these two effects are separable. Since we are interested in the intrinsic physical modification of the void size function due to dark energy and massive neutrinos, we do not consider the geometrical observational effects that can be misleading here. Moreover, we show the results for dark matter voids to have larger statistics and more distinguishable behavior of the relative void size functions, nevertheless the halo case shows analogous results for each minimum mass considered. The left panel of this figure shows the relative void size function in the presence of dynamical dark energy with respect to

$(w_0, w_a) = (-1, 0)$ equation of state, for a fixed total neutrino mass separately. The various dark energy equations of state are represented with different colors, following the legend, while the neutrino mass for which the relative difference is considered is represented with different line styles. As can be seen, the relative void size function for each dark energy equation of state with respect to $(w_0, w_a) = (-1, 0)$ is the same for each neutrino mass, within the Poissonian uncertainty. It follows that dark energy and neutrino mass effects on the void size function are separable. The right panel shows the relative void size function in the presence of massive neutrinos with respect to the massless case, for each dark energy equation of state separately. Also in this case, the relative void size functions show the same behavior independently on the dark energy equation of state.

3.2.3 Matching theory with measurements

In this subsection, we address the problem of connecting the observed void size function with the theoretical prediction of the excursion-set framework discussed in Chapter 2. This connection was explored in the last few years in various works [96, 97, 161, 171, 172], nevertheless we will refer to [97] since is the only one that considers a fully theoretical treatment in linking voids in the tracer distribution to voids in the dark matter distribution, then only this work shows the critical points on the methodology currently adopted, its limitations, and measures the systematic effects of the algorithm adopted. The theoretical model for the void size function used in this work is the Vdn model [96], discussed in Sect. 2.5.1, Eq. (2.126). Even if further studies are showing that this model is not optimal in describing the void size function [see Sect. 2.5.1, 84], the uncertainty in the algorithms used for the analysis does not allow to appreciate what we discussed in Sect. 2.5.1.

Void catalog preparation

The theoretical definition of a void as a dark matter underdensity hardly matches the objects identified in observations, i.e. the watershed voids found in the distribution of galaxy tracers [173]. The theoretical void size function relies, in fact, on voids defined as the Lagrangian matter density field filtered at the Lagrangian position $\mathbf{q}$ that first reaches a linear density contrast value at a given smoothing length, without having reached the collapsing threshold at larger radii (see Sect. 2.2). To compare observational data with theory, we therefore have to reconcile these different void descriptions.

A first method to connect theory and observations is to leave the linear threshold of void formation δ_v in the theoretical void size function as a free parameter and calibrate it so that the theoretical void size functions fit the measured void abundance in the considered data set. This method allows to overcome the problem by parameterizing our ignorance [159, 160, 173–176]. On the other hand, as suggested in [96], it would be interesting to directly connect the top-hat definition to observed voids, matching the theoretical void size function with the measured abundances in the data set without any free parameter. This second option is more attractive, we thus implement it for the given DEMNUni void catalogs.

The most direct way to connect the top-hat void definition with voids found by `VIDE` is to reconstruct their density profile and to use the spherical shape approximation in considering the Lagrangian to Eulerian void evolution. Thus, we find the radius r inside which the mean density contrast equals the chosen threshold. This is the crucial point in the procedure, as it re-normalizes the void radii to the same density contrast. To reconstruct the spherically

averaged void density profile, we rely on the volume-weighted barycenter of VIDE defined in Eq. (3.3). Even if the void shape is not important in Lagrangian space, where a spherical filter is adopted, it impacts the void evolution from Lagrangian to Eulerian space, where voids are detected. We are interested that voids are approximately spherical at least at the radius where they reach the threshold value. The spherical approximation for void evolution can be considered accurate enough for three main reasons. First, void ellipticity is low, as discussed in [97]. We note that the low ellipticity is verified only for relatively large voids, such as the ones we consider in this work. Small voids are considerably more aspherical as they are more strongly influenced by external environmental effects such as large scale tidal forces, are embedded in collapsing overdensities, and are squeezed against neighboring voids [177]—they would need a more careful consideration that we leave for future work. Second, with our large volumes and therefore large statistics, voids can be safely considered as spherical on average [178]. Finally, the abundance is a statistical measure for which in any case individual void features are averaged out. Summarizing, to first approximation, in deriving the void abundance we can consider voids evolving as spherical objects.

We then check whether void abundances obtained with resized voids match the theoretical void size functions. Following [96], we adopt the following strategy:

- We resize all VIDE voids to the fixed threshold, which we call $\delta^{\rm h}_{\rm v,NL}$, where h stands for halos, the tracer used for the void finding procedure. We apply the resizing using the code presented in [179], available in the CosmoBolognaLib C++/Python library[4] [180]. The subscript NL here and in the following, indicates when we are considering fully nonlinear quantities.

- The theoretical void size function is referred to the statistical properties of matter, not of the tracers. We follow [181, 182] to compute the threshold in the matter $\delta^{\rm m}_{\rm v,NL}$, which corresponds to the one in the halo distribution and then invert the linear bias relation $\delta^{\rm h}_{\rm v,NL} = b_{\rm eff} \times \delta^{\rm m}_{\rm v,NL}$ (with $b_{\rm eff}$ being the tracer bias).

- In order to calculate the theoretical void size function, we need to express the matter underdensity threshold in linear theory. This is obtained in the spherical collapse model by converting the nonlinear density contrast, $\delta^{\rm m}_{\rm v,NL}$, into the corresponding linear one, $\delta_{\rm v}$ (see Sect. 2.1 and Sect. 2.1.2). We verified that this conversion is weakly dependent on the redshift and variations of the dark energy equation of state. This would allow marginalization over $\delta_{\rm v}$, when considering the different equations of state implemented in the DEMNUni simulations.

- To account for the redshift dependence of the void size function, we extrapolated all quantities using linear theory to the present time, $z = 0$, [15], as it is done in the case of the halo mass function [15, 98]:

$$\begin{aligned} \delta_{\rm v} &\to \delta_{\rm v}/D(z) \\ \delta_c &\to \delta_c/D(z) \\ \sigma &\to \sigma(z=0)\,, \end{aligned} \tag{3.7}$$

  where $D(z)$ is the growth factor in linear theory [9] normalized to unity at z=0.

[4] https://github.com/federicomarulli/CosmoBolognaLib

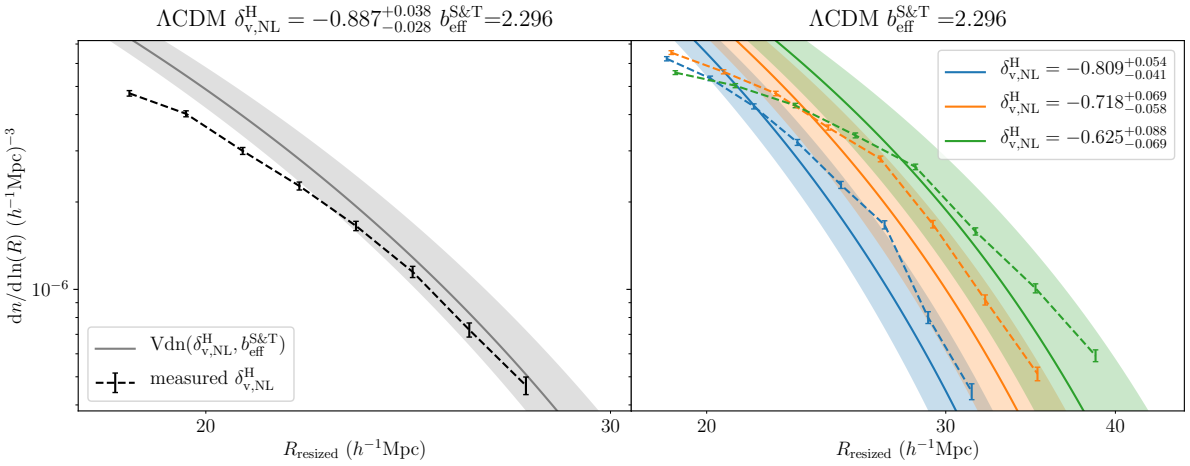

Figure 3.11: Match of the measured abundances of voids (after resizing procedure) with theoretical predictions. The left panel shows the measurement for $\delta^{\rm H}_{\rm v,NL} = -0.887$ (black curve), and the corresponding theoretical void size function obtained using $b_{\rm eff}$ as described in the text (gray curve). The right panel shows the measured void abundances for various thresholds $\delta^{\rm H}_{\rm v,NL}$ (dashed curves), and the corresponding theoretical void size functions (solid curves). Shaded areas give the uncertainty in the resizing procedure (see text). Figure from [97].

We now have all the tools to calculate the theoretical void size function and compare it with the void abundances of the resized void catalogs from the simulations.

Void abundance match: ΛCDM

We follow the strategy described: we obtain the linear threshold $\delta_{\rm v}$ corresponding to the nonlinear threshold in the matter distribution $\delta^{\rm m}_{\rm v,NL}$, obtained by inverting the linear bias relation: $\delta^{\rm h}_{\rm v,NL} = b_{\rm eff} \times \delta^{\rm m}_{\rm v,NL}$, and considering the conversion between the linear and nonlinear value. The effective bias $b_{\rm eff}$ of the FoF halos is computed using the Sheth-Tormen [14] model, Eq. (1.134) and Eq. (1.139). We study the corresponding void size function for the ΛCDM case for a fixed $\delta_{\rm v}$ value and show our results in Fig. 3.11. In the left panel, the black dashed curve represents the measured void abundance of the resized void catalog for ΛCDM, while the gray curve represents the theoretical void size function for one fixed $\delta_{\rm v}$ value in the halo distribution with $b_{\rm eff}$ the effective halo bias. We note that the radius range of this analysis is different from the one of the void size functions presented in Sect. 3.2.2, where the effective void radius was defined via the watershed approach. Here the radius $R_{\rm resized}$ refers to spherically resized voids, and therefore is smaller than the effective radius of watershed voids. The resized void range $[20-40]\,h^{-1}$Mpc of Fig. 3.11 corresponds to watershed voids in the range $R_{\rm eff} = [35-70]\,h^{-1}$Mpc.

The uncertainty in the theoretical void size functions is due to the resizing procedure and is estimated in the following way. We fix the threshold value $\delta^{\rm h}_{\rm v,NL}$ we wish to consider and we re-scale each void up to the corresponding radius using the algorithm in [179]. We then calculate independently the actual density contrast within the re-scaled radius for each void, we use the peak of the distribution of the measured density contrast as the effective threshold value, and quantify the uncertainty in the resizing procedure via the half-width-half-maximum of the measured density contrast distribution. Finally, we obtain the corresponding linear values for matter used to build the theoretical curve: $\delta_{\rm v} = -0.887^{+0.038}_{-0.028}$. While there might be ways

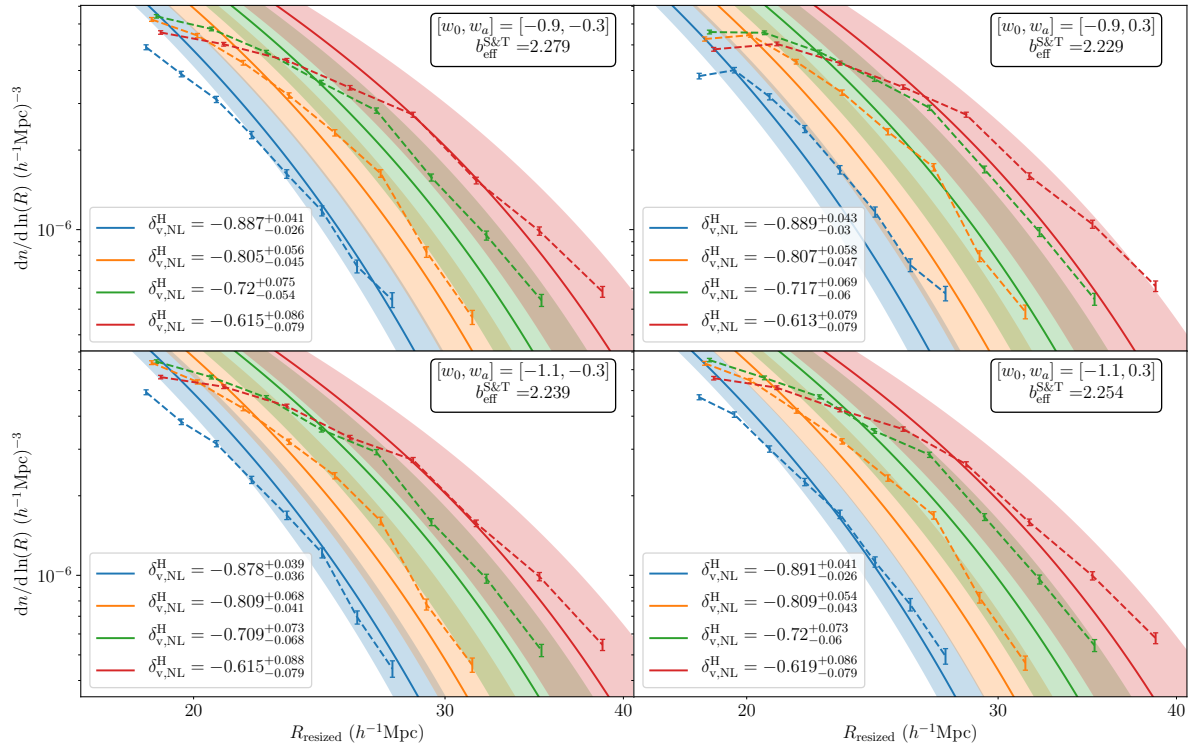

Figure 3.12: Each panel represents the measured void abundances after applying the resizing procedure (dashed curves), and the corresponding theoretical void size functions (solid curves), for various threshold values $\delta^{\rm h}_{\rm v,NL}$, and for the four sets of parameters of the CPL dark energy equation of state. The shaded areas show the uncertainty in the resizing procedure (see text). Figure from [97].

to improve the resizing procedure and reduce this uncertainty, we have focused here on testing this first version, which relies on simple principles. We will explore possible improvements in future works.

Within the estimated theoretical errors, we obtain a good agreement between data and theory without the need of free parameters, using only quantities that can be theoretically predicted: we are able to forecast the void size function with a fully theoretical approach [96, 98]. We note that, for low values of the resized radius, the measured void abundances are below theoretical predictions. This may be a consequence of the sparse statistics of small voids when approaching the mean halo separation of the considered simulations; denser surveys will have access to better void statistics in those ranges (e.g. PFS [183] and the Roman Space Telescope [184]), likely filling the gap.

Finally, we check how the theoretical void size function depends on the threshold value. We vary the threshold value $\delta^{\rm h}_{\rm v,NL}$ while keeping the effective bias fixed (right panel Fig. 3.11). All of the measured void abundances are within the uncertainty of the theoretical predictions (except at very small radii). This result confirms that the theoretical void size function prediction works for many threshold values. We notice that in observations this is a powerful tool to use, as it will allow to optimize the choice of the threshold to enhance cosmological constraints from the void size function.

Void abundance match: varying the dark energy equation of state

We repeat the above procedure for the CPL parameterization with the four sets of (w_0, w_a) parameters implemented in DEMNUni suite to check if the theoretical void size function is able to predict simulations in dynamical dark energy models. The results presented in Fig. 3.12 show that we obtain a good agreement between the measurements and the theoretical void size functions in the different cosmologies, where the dependence in the void size function is implicitly contained in the linear growth factor $D(z)$ [9]. As for the ΛCDM case, for all the analyzed threshold values the agreement is obtained by using the Sheth-Tormen [14] effective bias, computed for all the different dark energy equations of state. We note also in this case that the abundance of small voids falls below the theoretical prediction. In addition, mainly for the $(w_0, w_a) = (-0.9, 0.3)$ case (upper right panel in Fig. 3.12), the observed void abundance for two threshold values (green and orange curves) lies somewhat outside the uncertainty of the theoretical prediction. This mild mismatch could be due both to the non-optimal resizing procedure and to bias modeling.

Summarizing, we showed that the theoretical void size function agrees, within the errors, with measurements from simulations for different cosmological models, i.e. the standard ΛCDM model and the CPL parameterization with the four sets of (w_0, w_a) parameters implemented in DEMNUni. We also obtained a good agreement using various threshold values for each cosmological model. This allows on the one hand to select the best threshold δ_v for the available data, and, on the other hand, to use different threshold values to better constrain cosmological parameters with void abundances (even if, of course, measurements in this case will be correlated).

Although these results are promising, the uncertainty on the theoretical void size function is much larger than the Poissonian uncertainty in the measured voids. This can be overcome both on the theoretical side and on the data treatment side. On the theoretical side, we are making great improvements, as described in Chapter 2, and a more accurate theoretical void size function model can be tested soon in simulations [84]. On the other hand, a clearer understanding of the excursion set framework will require a proper void finder or cleaning procedure to match the voids defined in the excursion-set theory with measurements.

3.3 Void density profile

In this section, we focus on the density profile of voids measured in simulations. The void density profile can be considered in the differential or integrated version. The differential density profile is the mean density contrast in a spherical shell at distance r from the void center, the integrated density profile is the mean density contrast within a sphere of radius r. Each of these quantities has a proper application in the cosmological exploitation of cosmic voids. If the spherical approximation is accurate enough, it can be considered as the source of gravitational field, therefore of the velocity field for redshift space distortion, lensing [185–194], and integrate Sachs-Wolfe effect [140, 143, 189, 195–208]. Moreover, the integrated density profile is the quantity we used in Sect. 3.2.3 to find the threshold voids to match measurements with theory. The differential density profile in a given tracer distribution is related to the void-tracer cross-correlation that can be used extract cosmological information from the Alcock-Paczyński and redshift-space distortion effects [137–139, 141, 142, 144, 178, 209–223]. In this section,

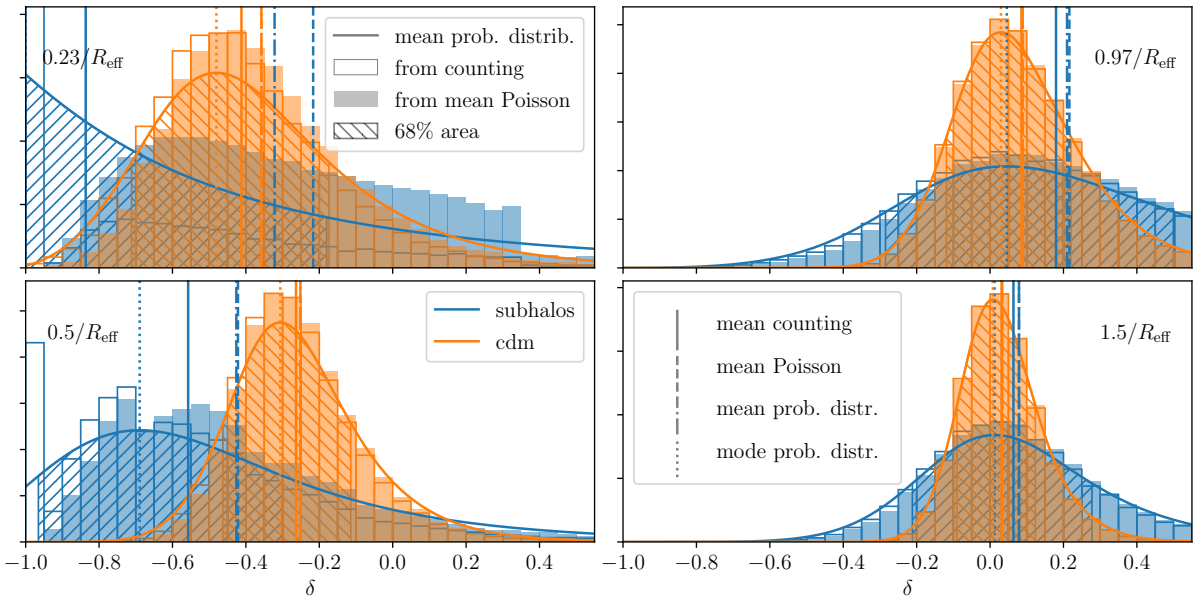

Figure 3.13: Distribution of density contrast values and stacked quantities for voids detected in ΛCDM simulation at $z = 1.05$ in the halo distribution with a minimum mass of $2.5\times10^{12}h^{-1}M_{\odot}$. The four panels show the result for the differential density profile of halos (blue) and dark matter particles (orange) at four shell radii, estimated in shells of thickness $0.0667/R_{\rm eff}$. The vertical lines show the stacked density contrast value computed by stacking the density contrast via direct counts (mean counting, solid), the expected value (mean Poisson, dashed), and considering the expected value of the stacked distribution (mean prob. distr., dash-dotted); the dotted vertical lines represent the mode of the stacked distribution (mode prob. distr.). The empty histograms represent the distribution of density contrast values in the various voids estimated via direct counting, the filled histograms show the measured distribution of the expected density contrast value of each void shell assuming a Poisson statistics, and the solid curves shows the stacked density probability of density contrast, for which the area filled with diagonal lines represents the 68% of the area subtended by the shortest interval.

we will consider the statistics that can be extracted on the measured void profile. In particular we will focus on the stacked profile, i.e. the mean density profile in which the distances are rescaled by the void effective radius. The differential stacked profile is the void-tracer cross-correlation, since it estimates the probability to find a tracer at a distance r from the center of the void. In this section we will consider various possibilities to measure the void density profile in a discrete distribution and its dependence on the tracers (Sect. 3.3.1), the halo bias and their distribution in voids (Sect. 3.3.2), and the sensitivity to dark energy and neutrino mass (Sect. 3.3.3).

3.3.1 Density estimator for a discrete particle distribution

In this subsection, we consider two possible estimators for measuring the void density profile in discrete tracer distribution and then various methods to estimate the stacked profile and other void profile related quantities. Moreover, we will consider the tracer impact in the density profiles obtained considering halos or dark matter particles. In this section we will consider only voids detected in the halos distribution, and we measure the corresponding density profile in both halos and dark matter distributions. As before, the void center from which we measure

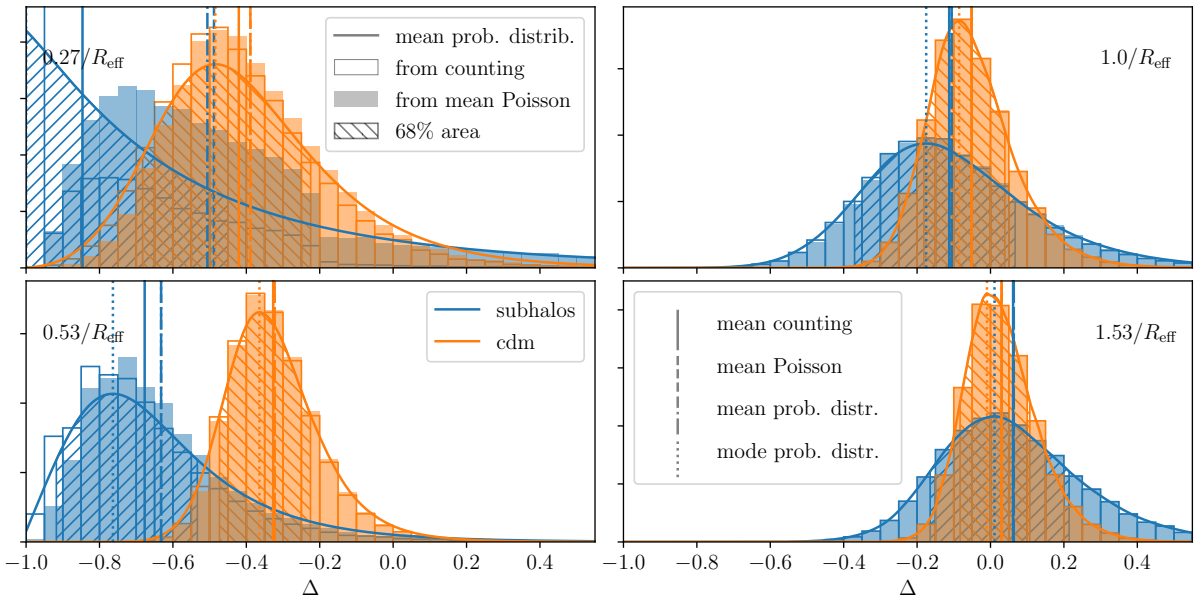

Figure 3.14: This figure is organized as Fig. 3.13 but for the density contrast within spheres with four different radii, shown at the corner of each panel.

the density profile is the volume weighted barycenter, defined in Eq. (3.3). In stacking void profiles, we renormalize the distance from the void center with respect to the void effective radius, in this way the characteristic shape of the void profile is retained in the final quantity [218].

We start by considering two possibilities to measure the density profile of voids in a discrete distribution. The most direct way to measure the density profile is via counting the number of tracers in a shell (or sphere), where the density contrast within it is

$$\frac{N_i}{V_i}\frac{V_{\rm box}}{N_{\rm tot}} - 1\,, \tag{3.8}$$

where N_i is the number of tracers contained in the $i^{\rm th}$ shell (or sphere) with volume V_i. For shells (spheres) in which there are no tracers, the corresponding density contrast obtained with direct counts is -1. Even if a zero number of halos is reasonable due to the sparsity of tracers, this is not the case in the dark matter field. Dark matter density is a continuous quantity, discretized for numerical reasons; the absence of dark matter particles in a given volume in simulations is usually due to not enough tracers, so to not enough resolution to trace the dark matter distribution and thus in measuring the density profile. In other words, we are measuring the density profile beyond the resolution provided by the simulation. This opens the problem of how to estimate the density contrast in the presence of zero or few tracers, and more generally of how to estimate the density contrast field from a discrete distribution. If the underlying continuous density contrast of the tracer field considered follows a Gaussian statistics, the corresponding randomly subsampled discrete distribution follows a Poissonian statistics [224]. Gaussian distribution is almost exact in the linear regime of the density contrast field [100], nevertheless it is a rough approximation for both the nonlinearly evolved dark matter and halo density field. In spite of this approximation, in the following, we assume a Poissonian distribution for both dark matter particles and halos. This is because the Poissonian distribution allows us a simple implementation that can be extended to more general distribution functions; moreover, it is ac-

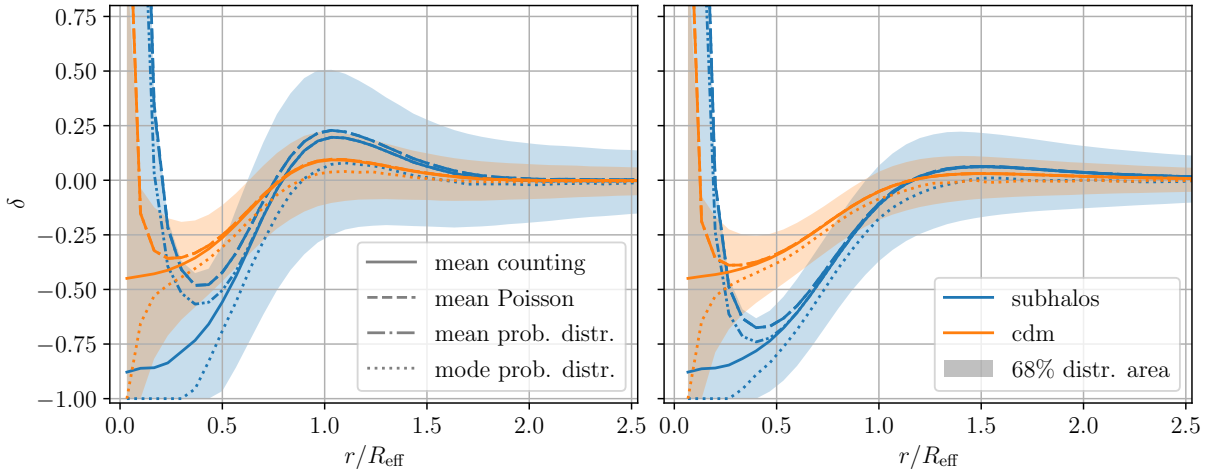

Figure 3.15: Stacked density profile measured in halo (blue) and dark matter (orange) distribution for voids detected in ΛCDM simulation at $z = 1.05$ in the halo distribution with a minimum mass of $2.5 \times 10^{12} h^{-1} M_\odot$. The left panel shows the differential density profile with shell thickness of $0.0667/R$, the right panel represents results for the integrated density profile, the two panels are organized in the same way. The stacked profile is considered via stacking the density contrast estimated via direct counts (mean counting, solid), via the expected value from the Poisson distribution (mean Poisson, dashed) and considering the expected value of the stacked distribution (mean prob. distr, dash-dotted). The two panels show also the mode of the stacked distribution (mean prob. distr., dotted) and the least interval subtending 68% of the distribution area (shaded region).

curate enough to show the impact of estimating a density probability of the density profile. The basic idea is to associate a probability distribution of density contrast values to each number of tracers N in a given volume V,

$$\mathcal{P}(N, V, \delta) = P(N, \lambda(\delta)) \frac{V N_{\rm tot}}{V_{\rm box}}, \tag{3.9}$$

where δ is the density contrast. If $P(N, \lambda)$ is a Poisson distribution,

$$P(N, \lambda) = \frac{\lambda^N e^{-\lambda}}{N!} \qquad \text{and} \qquad \lambda = (1 + \delta) \frac{V N_{\rm tot}}{V_{\rm box}}. \tag{3.10}$$

In this way, we can associate a distribution of density contrast values to each shell (sphere) of each void.

From these two estimators of the density profile for a single void, we consider various possibilities to obtain the stacked density profile and other quantities associated with the distribution of the density contrast along the void profile. Fig. 3.13 and Fig. 3.14 show the stacked density contrast values and their distribution in four radius bins, for the differential and integrated density profile, respectively. These figures refer to ΛCDM simulation at $z = 1.05$ for voids detected in the halos distribution with a minimum mass of $2.5 \times 10^{12} h^{-1} M_\odot$ and a radius larger than $2.5 \times$ mps. The density profile of each void is measured in 60 linearly spaced radial bins from 0 to $4R_{\rm eff}$, it follows that the shell thickness is $0.0667/R_{\rm eff}$. The results for the halo distribution are represented in blue, while in orange the ones for dark matter particles. The first estimator we consider for the stacked density profile is based on the direct counting method to evaluate the density profile of each void, Eq. (3.8), and it consider the stacked profile as the

mean density profile of each void, renormalized with respect to the void radius $R_{\rm eff}$. This is represented in Fig. 3.13 and Fig. 3.14 as the vertical solid line, while the distribution of the values assumed by each void is represented by the filled blue histogram. The second method to estimate the density profile of each void associates a density distribution to each radius bin, Eq. (3.9). This opens various ways to consider the stacked density profile. One possibility is to consider the expected density contrast value for each void and to stack this quantity. For a Poissonian distribution, the expectation value of the density contrast for each shell (sphere) of a single void is

$$\langle \mathcal{P}(N, V, \delta) \rangle = \frac{N+1}{V} \frac{V_{\rm box}}{N_{\rm tot}} - 1 \,. \tag{3.11}$$

Note that this result is equal to Eq. (3.8), where N is substituted with $N + 1$. The stacked density profile of the density contrast expectation value is obtained as the mean of this quantity. In the figures, this is represented as the vertical dashed line labeled as "mean Poisson", while the distribution of values assumed by each void is represented by the empty blue histogram. Note that the histograms of the density contrast distribution obtained via direct counts and via the expected Poisson value are different for low density contrast values, where the number of tracers is low and so $N+1$ is considerably larger than N, otherwise they converge, since $(N+1)/N \simeq 1$ for $N \gg 1$. Another possibility is to consider the mean density contrast distribution function given by the mean of the density contrast distributions of each void. In this way, we obtain a stacked density distribution function for each radius bin. From the stacked density distribution function, we can compute the expected density contrast. Beyond the stacked density profile, we can also consider the mode of the stacked distribution and its width. The stacked density contrast distribution is represented as solid curves, while the expected value is represented by the vertical dash-dotted line, labeled as "mean. prob. distr". The mode is represented by the vertical dotted line and the 68% of the area subtended by the least continuous interval is represented by the regions filled by diagonal lines. Note that the stacked distribution of density contrast values and the corresponding histograms overlap for enough large radius bins, where the number of tracers contained in each shell (sphere) is large enough so that the discrete distribution is well described by the continuous limit. On the other hand, in the inner region of voids, where the number of tracers is lower, the histograms and the corresponding continuous distributions are not overlapped. This is particularly evident for halos, since they are considerably less with respect to dark matter particles.

Fig. 3.15 shows the stacked density profile obtained with the various estimators described above, together with the mode and the 68% of the subtended area of the sacked distribution for both the differential (left panel) and integrated (right panel) void density profiles, for dark matter (orange) and halos (blue) density field. First, consider the behavior of the stacked density profiles at small $r/R_{\rm eff}$, where the number of tracers is low. The stacked density profile in the halo distribution obtained from direct counts (blue solid curve), reaches a negative value greater than -1. This is due to the fact that many voids have no halos in the inner regions, but since the stacked density contrast is weighted also with void containing tracers in the inner regions, the result is greater than -1. Nevertheless, the stacked profile from the expected Poissonian value (blue dashed line), diverges at low radius. This is due to the substitution $N \to N + 1$ in Eq. (3.11), therefore, if the number of tracers in a given shell (sphere) is zero, the expected density contrast value diverges as $\propto V^{-1}$. Note that this is the same behavior shown by the stacked density distribution of the halo density contrast (blue dash-dotted curve). The mode (dotted curve) of the density contrast value is -1, while the expected value, i.e. the mean, takes

large values and the distribution is wide. The different behavior of the mean and the peak of the distribution is due to the long tail of the Poisson distribution when few tracers are present, entailing a peak close to -1 and, at the same time, weighting the expected density contrast to positive values. The density contrast of dark matter particles (orange) shows a similar behavior but at lower radii, where the mode shows a steep drop and the expected value of the stacked density contrast distribution presents a fast increase. This is due to the lack of tracer caused by not enough spatial resolution of dark matter particles. At larger radii, the various estimations of the stacked density profile converge. Nevertheless, note that the mode of the distribution is always lower with respect to the mean, showing that the density contrast distribution is always asymmetric with respect to the peak. The skewness of the density contrast distribution is clearly visible in Fig. 3.13 and Fig. 3.14. One of the reasons for this asymmetry is that these distributions have a lower bound fixed by the physical condition $\delta \geq -1$, nevertheless they do not have any upper bound, allowing the possibility of long tails for $\delta > 0$.

The differential void density profile presents a peak at $r/R_{\rm eff} \sim 1$, this feature, known as the compensation wall, is useful for cosmological analysis involving redshift-space distortion and Alcock-Paczyński around cosmic voids [141, 218]. The visibility of this feature depends on the center definition of voids. In this study we are considering the volume weighet barycenter, so the void center is considered approximately at the mean distance of the overdensities forming the watershed, so at $r/R_{\rm eff} \sim 1$ we can see the overdensity peak of the watershed. This is not the case of considering the minimum of the density field, where the density profile reaches $\delta = 0$ without any particular feature [142, 221].

The difference between the density profile in the dark matter distribution with respect to the corresponding one in the halo distribution is due to halo bias. It can be seen that the halo density contrast shows the same behavior of the dark matter density contrast, but it is enhanced according to the minimum halo mass. Fig. 3.16 and Fig. 3.17 show the dependence of the density profile in halo distribution and the corresponding one in dark matter distribution, for voids detected in halo distribution with various minimum halo masses. The voids considered are the one with the effective radius larger than $2.5 \times$ mps, Fig. 3.16 shows the differential density profile, while Fig. 3.17 the integrated one. Note that, for each redshift, the density profiles in the halo fields corresponding to the various minimum masses are close to be self-similar, whereas the corresponding profile in dark matter varies greatly. This is due to the halo bias that grows with the mass and with the redshift.

3.3.2 Relationship between halos and dark matter in voids

In the previous subsection, we explored various ways to estimate the stacked void density profile, now we focus on the relationship between dark matter density contrast and halos, i.e. the bias of the tracer with voids, and on the properties of halos in voids and we focus on two possible ways to model it. In this subsection, we estimate the stacked density profile using the direct counts of tracers within voids, i.e. Eq. (3.8).

Linear bias and extrapolation

In Sect. 1.3 we introduced halo bias, showing how this phenomenon emerges in the correlation function and from the multiplicity function. Moreover, the peak-background split argument shows that halo bias depends on the background density contrast field. A comprehension of how halo (and galaxy) bias works in voids is important for many cosmological analyses with

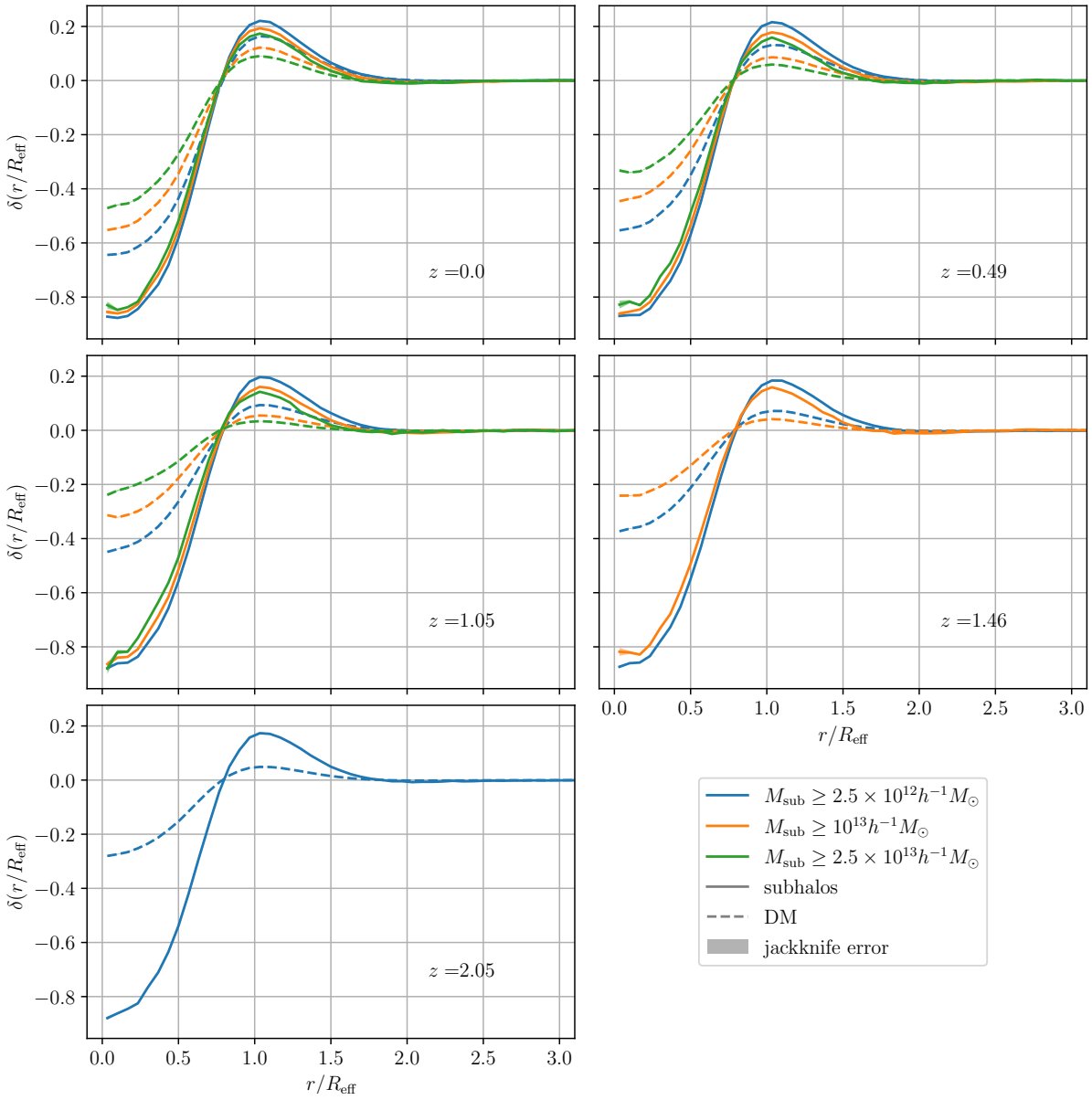

Figure 3.16: Halo (solid curves) and dark matter (dashed curves) stacked differential density profile for voids detected in the halo distribution, for minimum mass of 2.5×10^{12}, 10^{13}, and $2.5 \times 10^{13} h^{-1} M_\odot$, represented with the blue, orange, and green curves, respectively.

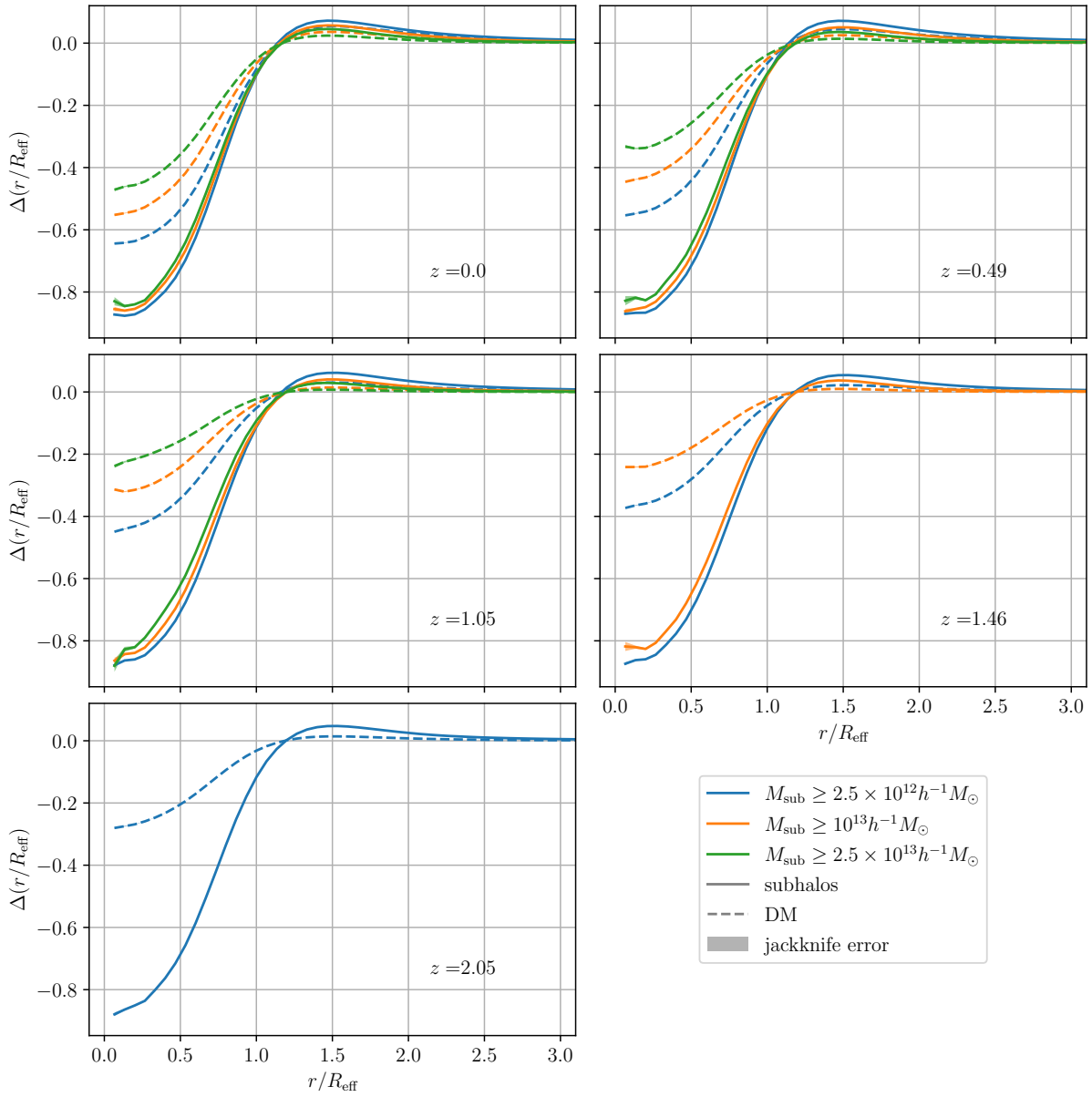

Figure 3.17: Halo (solid curves) and dark matter (dashed curves) stacked integrated density profile of voids detected in the halo distribution, for minimum mass of 2.5×10^{12}, 10^{13}, and $2.5 \times 10^{13} h^{-1} M_\odot$, represented with the blue, orange, and green curves, respectively.

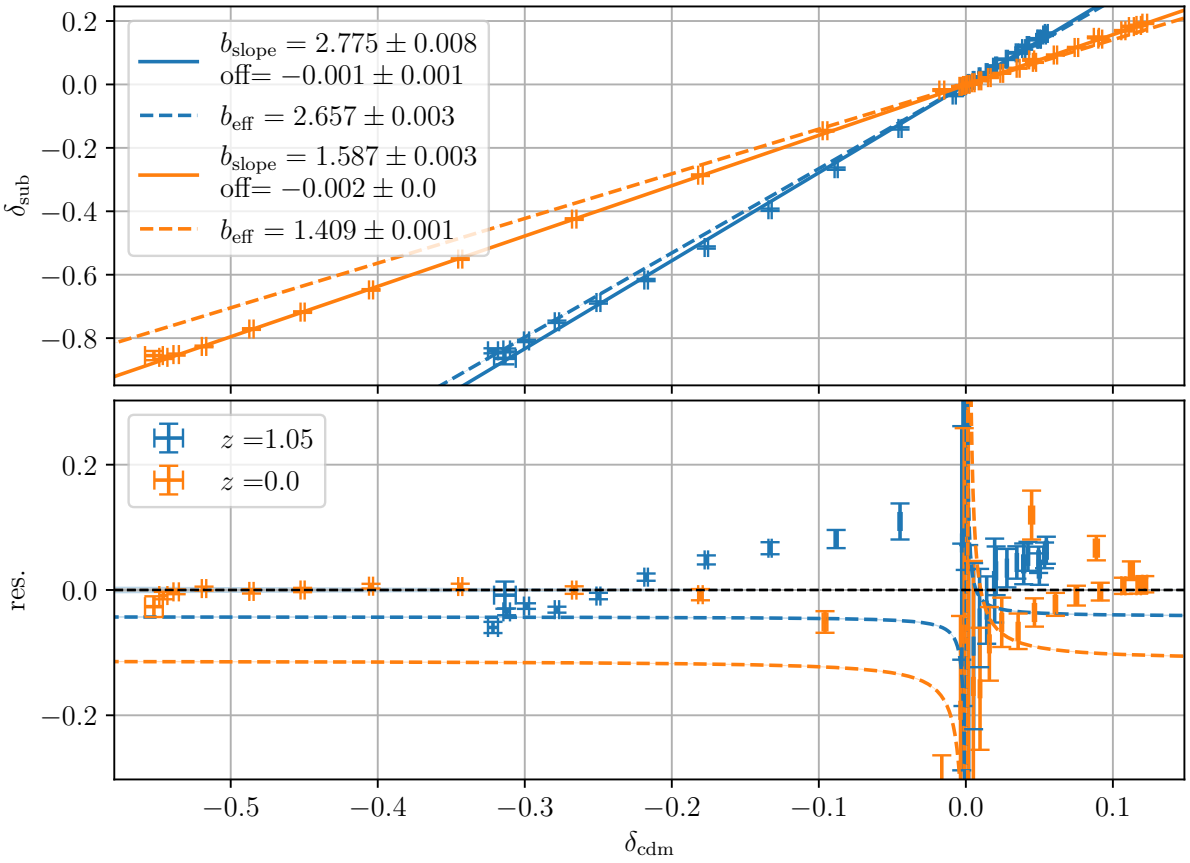

Figure 3.18: Linear modelization of the halo bias in voids in halo distribution with mass greater than $10^{13}h^{-1}M_{\odot}$ for $z = 0$ (orange) and $z = 1.05$ (blue) snapshots. The error-bars in the upper plot correspond to the measured stacked density contrast of subhalos and dark matter in the same radius bin, the solid lines represent the best fit of the linear models, and the dashed lines the theoretical linear bias. The lower panel represents the residuals with respect to the linear fit.

cosmic voids, listed at the beginning of this section. The problem of how halos, galaxies, and other kinds of objects trace the underlying matter density in cosmic voids has been widely explored in the literature [137, 172–174, 181, 182, 186, 225, 226]. For practical purposes, halo density contrast is usually modeled in linear relation with respect to the underlying matter density contrast, as described in [181]. In this subsection we will revisit the modeling proposed in [181], showing that even if a linear relation between halos and dark matter density contrast can be an effective approximation to parameterize our ignorance, the resulting halo bias is different from the linear effective bias term. Moreover, we explore the limit in which a linear modelization is valid.

To estimate a linear relation between dark matter density and halos in the stacked void density profile, we perform a linear fit, considering the relation

$$\delta_{\rm sub} = b_{\rm slope} \times \delta_{\rm cdm} + \text{offset}\,, \tag{3.12}$$

where $\delta_{\rm sub}$ is the subhalo density contrast and $\delta_{\rm cdm}$ is the density contrast of the underlying dark matter. Looking at Fig. 3.16, we note that at large distance from the void center, the density profile is very close to zeros. In order to avoid the result from being unphysically driven by the vanishing density contrast at large scale, we decided to consider the differential density profile

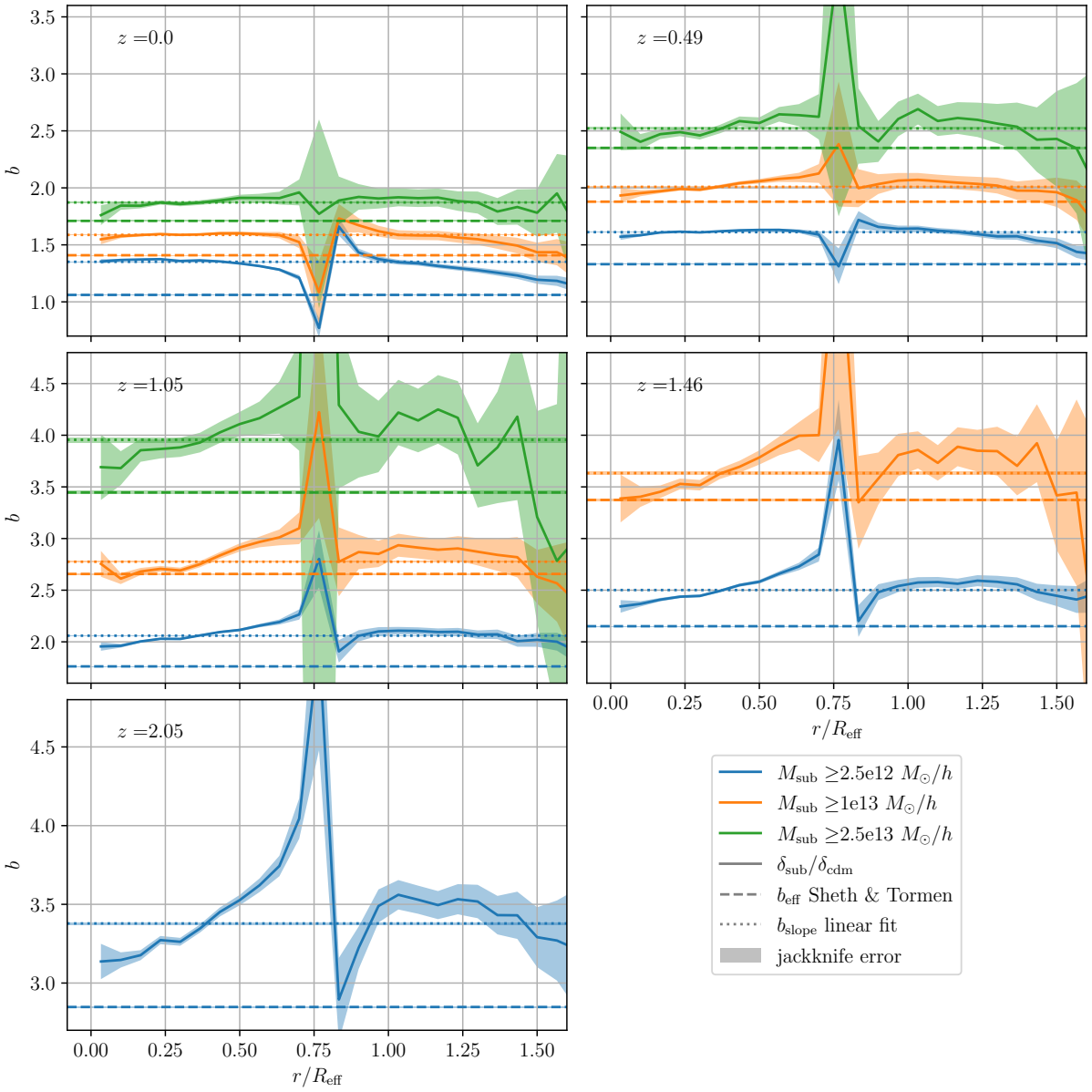

Figure 3.19: Relationship between dark matter and halo density contrast along the stacked void density profile for voids with minimum radius of 2.5 × mps detected in the halo distribution, with minimum subhalo mass of 2.5×10^{12}, 10^{13}, and $2.5\times10^{13}h^{-1}M_{\odot}$, represented with the blue, orange and green lines, respectively. The solid lines represent the measured ratio of subhalos and dark matter density contrast, where the shaded area corresponds to the jackknife error, the horizontal dashed lines correspond to the linear bias theoretically computed with the Sheth-Tormen multiplicity function, the horizontal dotted lines correspond to the slope of the linear regression fit.

in the range $0 - 2\,R_{\rm eff}$. The error of $\delta_{\rm h}$ and $\delta_{\rm cdm}$ is estimated using the jackknife method [227]. We use these data to perform a Bayesian fit, considering uniform priors for the slope and offset parameters. Fig. 3.18 shows the results for voids detected in halo distribution with minimum mass of $10^{13}h^{-1}M_{\odot}$ in the $z = 0$ and $z = 1.05$ snapshots for ΛCDM cosmology. The upper plot shows the $(\delta_{\rm cdm}, \delta_{\rm sub})$ plane, the two-dimensional error-bars correspond to measurements while the solid lines represent the best fit of the linear relation Eq. (3.12). The lower plot shows the residuals, computed as data/model − 1. It can be noted that the linear parameterization of Eq. (3.12) fits quite well the relationship between halo and dark matter density contrast in stacked voids. This is particularly true for the $z = 0$ snapshots in the underdense region. The linear model works worse for the $z = 1.05$ snapshot, nevertheless the halo density profile is recovered within 10% of precision. Note that the scatter of the residuals around $\delta_{\rm cdm} = 0$ diverge, due to the ratio with respect to values very close to 0, so they are outside the y-axis boundary of the lower panel. Interestingly, the offset is always very close to zero, and in most of the cases is consistent with zero.

The linear model for Eq. (3.12) recalls the bias expansion truncated to the first order, i.e. linear bias, described by Eq. (1.141), so we compared the linear bias relation theoretically derived with the Sheth-Tormen model, from Eq. (1.134) and Eq. (1.141). To accurately compute the Sheth-Tormen bias, we fit the p and q parameters of Sheth-Tormen multiplicity function on the halo mass function measured in each snapshot, performing a Bayesian Markov-Chain Monte Carlo (MCMC). We recall that p and q parameters model the deviation from sphericity in halo formation. We found that the theoretical linear bias does not match the measurements, nor the linear fit. It can be seen from Fig. 3.16 that the theoretical prediction and the linear fit differ for more then 10% in the $z = 0$ snapshot and around 5% in the $z = 1.05$ one.

Fig. 3.19 represents the comparison between the measured ratio $\delta_{\rm sub}/\delta_{\rm cdm}$ (solid curves), the slope of the linear fit (dotted horizontal lines), and the theoretical linear bias (dashed horizontal lines). It can be noticed that the theoretical linear bias never provides a good modelization of the bias of halos in voids. Moreover, even if the linear empirical fit provides a good modelization at low redshift, this is not true for $z \gtrsim 1$, where the linear relation does not reproduce the general behavior of the relationship between halos and dark matter in the underdensity region, corresponding to $r/R_{\rm eff} \lesssim 0.75$ (compare with Fig. 3.16). The discontinuity at $r/R_{\rm eff} \sim 0.75$ is due to the vanishing denominator.

Halo mass function within voids and theoretical bias recovered

From a theoretical point of view, it is expected that the theoretical linear bias relation cannot provide a good model for the relationship between halo and dark matter density contrast in voids. This is due to the fact that the density contrast of matter in voids is negative enough to break the linear approximation of Eq. (1.139), i.e. $|\delta_{\rm cdm}| \sim 1$, and a polynomial expansion should be more representative of the halo density contrast behavior within voids, Eq. (1.141). From the peak-background split argument, we are able to exactly model the density contrast of halo in voids from the relation

$$\delta_{\rm h} = \frac{n_{\rm h}(M_{\rm min}, r)}{\langle n_{\rm h}(M_{\rm min})\rangle} - 1\,, \tag{3.13}$$

where $n_{\rm h}(M_{\rm min}, r)$ is the number density of halos with minimum mass $M_{\rm min}$ in the shell at distance r from the void center, and $\langle n_{\rm h}(M_{\rm min})\rangle$ is the mean density of halos with minimum mass $M_{\rm min}$ in the Universe. The number density of halos is derived from the halo mass function

via

$$n_{\rm h}(M_{\rm min}) = \int_{M_{\rm min}}^{\infty} \frac{{\rm d}n}{{\rm d}M} {\rm d}M \,, \tag{3.14}$$

where ${\rm d}n/{\rm d}M$ is the halo mass function that can be theoretically modeled with a multiplicity function, Eq. (1.106). As discussed in Sect. 2.5, the halo mass function is the same in both Eulerian and Lagrangian space. This is true for the halo mass function in the entire volume of the Universe, but now we are interested in how the halo mass function behaves in patches with various background mean density, i.e. in spherical shells around voids. The number density changes according to the evolution of the background patch, as described in Sect. 1.3.3, so the halo mass function of Eq. (1.106) evolves in Eulerian space as

$$\frac{{\rm d}n}{{\rm d}M} = (1 + \delta_{\rm bg}^{\rm NL}) \frac{\rho}{M} \left[f(\nu, p, q) \frac{{\rm d}\nu}{{\rm d}M} \right]_{\delta_{\rm bg}^{\rm L}} \,. \tag{3.15}$$

In this equation, $\delta_{\rm bg}^{\rm NL}$ is the background matter density in the evolved patch, and $\delta_{\rm bg}^{\rm L}$ the corresponding linear value, obtained with Eq. (2.17) and Eq. (2.22). Note that these equations describe the evolution of a spherical fluctuation, since the patch dimension is larger than the Lagrangian dimension of halos within it, and since we are not considering extremely high density contrast values, the spherical evolution provides an accurate modeling. However, a treatment at higher order can also consider the tidal field [19]. The quantity in square brackets is evaluated at $\delta_{\rm bg}^{\rm L}$, this means that the significance of the critical density ν is treated according to the peak-background split, $\nu = (\delta_{\rm c} - \delta_{\rm bg}^{\rm L})/\sigma(M)$.

To accurately reproduce the halo mass function, we fit the parameters p and q on the measured mass function. The halo mass function is a differential quantity; therefore, to estimate it, we divide the halo catalog into mass bins ΔM, to obtain the number of halos with mass between M and $M + \Delta M$, i.e. $N(M, M + \Delta M)$. Note that the mass function is a rapidly decreasing function, so if the mass bins are not thin enough, a direct comparison between the theoretical mass function evaluated in the center of the mass bins, i.e. ${\rm d}n(M + \Delta M/2)/{\rm d}M$, and the measured $N(M, M + \Delta M)/(\Delta M\, V_{\rm box})$ may induce biases in the estimation of the parameters p and q. This leads to an error in the mean density of halos Eq. (3.14) of $\sim 10\%$ for the thicknesses of the mass bins considered in this work. To avoid biases in the results, we compared the measured differential number density of halos with the theoretical mass function integrated in the mass bin, for each bin, i.e.

$$\text{data}: \quad \frac{N(M, M + \Delta M)}{V_{\rm box}} \qquad \longleftrightarrow \qquad \text{theory}: \quad \int_{M}^{M+\Delta M} \frac{{\rm d}n(M', p, q)}{{\rm d}M'} {\rm d}M' \,. \tag{3.16}$$

Even in this approach, the theoretical modelization cannot reproduce the measurements, as can be seen from Fig. 3.22. Moreover, the disagreement between theory and data is even larger for the full nonlinear treatment (orange solid line) with respect to the linear approximation (orange dashed line). Nevertheless, this result is important for at least two reasons. First, it shows where the linear bias approximation breaks. The linear bias is a good approximation for $r/R_{\rm eff} \gtrsim 1.5$, where the density contrast of matter is very close to zero. In the ridge, $r/R_{\rm eff}$ between ~ 0.75 and ~ 1.5 the linear approximation is well distinguishable with respect to the exact peak-background split implementation, while in the underdensity region, i.e. $r/R_{\rm eff} \lesssim 0.75$, the difference between the two results increases as the matter density contrast field becomes more negative. The second reason is that, even if the linear bias approximation

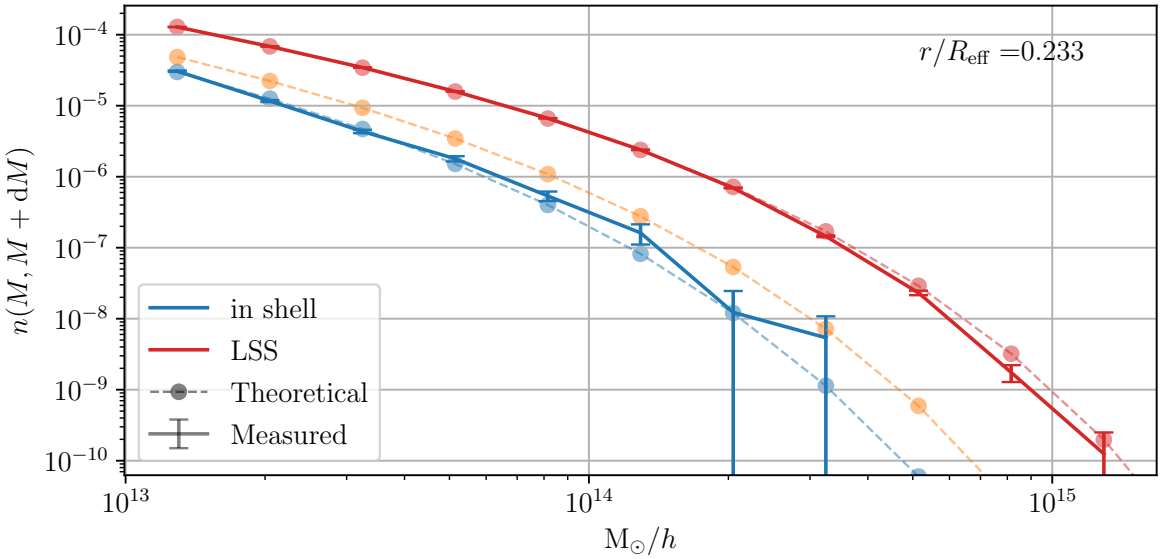

Figure 3.20: Differential number density of halos in the stacked spherical shell centered at $r/R_{\rm eff}$ = 0.233 and with $0.0667/R_{\rm eff}$ of thickness, for voids detected in subhalo distribution with a minimum mass of $10^{13}h^{-1}M_\odot$ at z = 1.05. The blue error-bars represent the measured differential halo number density function within the shell, the red ones in the entire distribution. The blue dots correspond to the theoretical differential halo number density given the background of the shell and p and p fitted within the shell, the orange dots represent the theoretical prediction given the dark matter background of the shell and p and q fitted on the large-scale halo mass function, i.e. in the entire box, the red ones the theoretical prediction with the background matter density contrast equal to zero, and large-scale values for p and q.

may better reproduce the observed density contrast, it is out of the range of its validity and thus physically unmotivated.

This result suggests that the halo mass function within voids is not universal. Exploring the halo mass function (Eq. 3.15) along the void density profile, we note that the multiplicity function modulates to reproduce measurements: the parameters p and q vary with respect to the large-scale values, i.e. the ones corresponding to the universal mass function from the entire simulation box. We measured them considering the halo mass function in each shell, following the same procedure used for the large-scale p and q values described above. Fig. 3.20 shows the differential halo number density in a representative stacked spherical shell. The error-bars represent the measurements, for the differential halo number density in shell (blue) and for the differential number density considering all the halos of the simulation (red). The dots represent the theoretical model. In particular, blue and orange dots represent the theoretical prediction of the differential halo number density given the dark matter density contrast of the shell as background, for p and q fitted on the halo distribution in the shell (blue) and using the large-scale fit values for p and q. The red dots represent the theoretical prediction with large-scale p and q and without the background, fitted on the red error-bars. Note that considering the only background in the peak-background split approach, the theoretical halo mass function does not fit the data. To reproduce the halo mass function in the shell, we must also vary p and q along the stacked void density profile. We found that the p and q values depend on the shell radius, suggesting that halo formation is environment dependent.

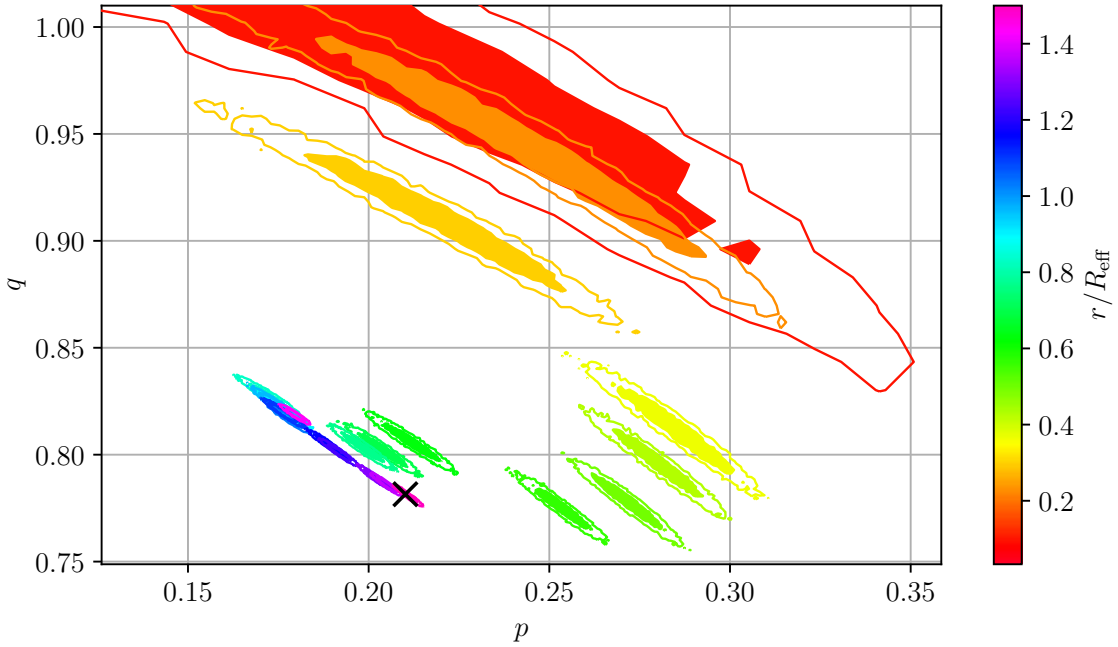

Figure 3.21: 68% and 95% confidence level for the p and q parameters of the Sheth-Tormen multiplicity function measured in stacked shells around voids, for voids detected in the subhalo distributions with minimum mass of $10^{13}h^{-1}M_{\odot}$ at $z = 1.05$. The black cross represents the universal p and q, measured in the entire box.

Fig. 3.21 shows that p and q parameters move in the (p, q) plane following a continuous path and converging to the large-scale values at large r, represented as the black cross. This behavior is reproduced in all the redshift snapshots and for each of the minimum halo masses considered. Moreover, this behavior is retained for each of the minimum void radii explored in the stacking procedure and also when we stack voids in bins of $R_{\rm eff}$. This result gives insight into the dynamics of halo collapse within voids. The parameter q refers directly to the ellipticity of the collapse [15], and $q = 1$ corresponds to spherical collapse. This allows to interpret Fig. 3.21: the q parameter approaches 1 in the inner shells, suggesting that halo collapse is more spherical in the central region of voids with respect to the outer regions. The dynamics of halo collapse modifies the multiplicity function, and in this way halo bias. It is now possible to use the corresponding halo mass function measured in each shell to model the halo distribution in the stacked void profile. To reproduce the halo density contrast with the peak-background split, we must consider that at various shells p and q varies together with the background density,

$$\delta_{\rm h}(r) = \frac{n_{\rm h}(M_{\rm min})\big|_{\delta_{\rm bg},p(r),q(r)}}{n_{\rm h}(M_{\rm min})\big|_{\delta_{\rm bg}=0,p_{\rm LSS},q_{\rm LSS}}} - 1\,. \tag{3.17}$$

where $p_{\rm LSS}$ and $q_{\rm LSS}$ are the values of the universal halo mass function. The resulting density profile in halo distribution accurately reproduces the observed one, as shown in Fig. 3.22 by the blue curve.

These results show that the peak-background split allows to accurately model the halo distribution in cosmic voids, demonstrating that the halo bias behavior in voids is sourced by the halo mass function behavior, modified both by the background underdensity and by the dynam-

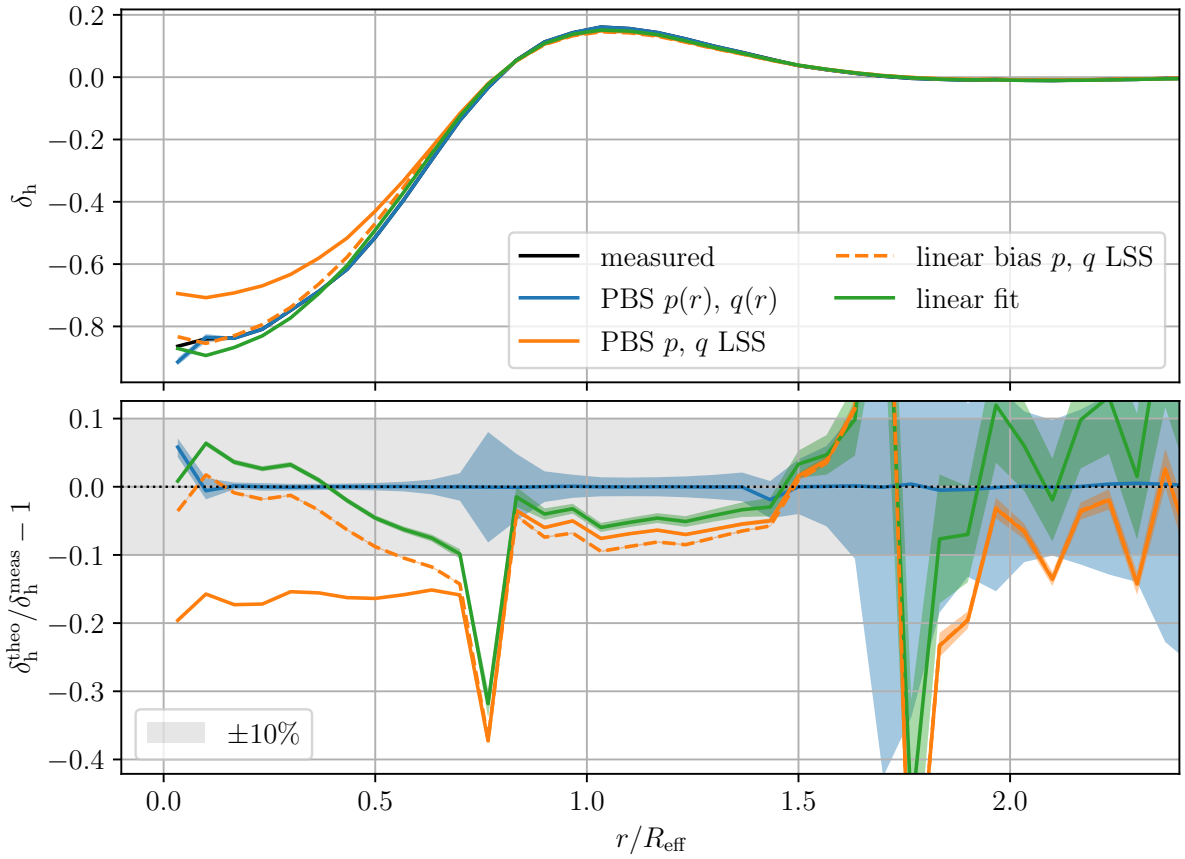

Figure 3.22: Relationship between dark matter and halo density contrast along the stacked void density profile for voids detected in the subhalo distribution with a minimum mass of $10^{13}h^{-1}M_\odot$ at $z = 1.05$. The upper panel shows the measured halo density profile (black) and the reconstructed density profile given the dark matter density obtained in the peak-background split framework with p and q depending on the distance from the void center (blue), for p and q measured at large-scale (orange, solid) and for the corresponding bias expansion truncated at the first order (orange, dashed), and reproduced by the linear fit (green). The lower panel shows the relative values.

ics of the collapse. The latter, which tends towards sphericity in the inner regions of voids, can be effectively described by the p and q parameters of the Sheth-Tormen model.

3.3.3 Sensitivity to dark energy and neutrino mass

In this subsection we explore the sensitivity of the void density profile to dark energy and massive neutrinos. Contrary to the void size function, the density profile of watershed voids is almost insensitive to both dark energy and massive neutrinos, in particular concerning voids detected in the halo distribution. Fig. 3.23 and Fig. 3.24 show the comparison of the stacked void density profile in halo distribution with respect to ΛCDM, for voids detected in the halo distribution with a minimum radius of $2.5\times$mps, and minimum halo mass of $2.5\times10^{12}h^{-1}M_\odot$ at $z = 1.05$, for the differential and integrated profiles, respectively. The uncertainty of the profile is computed using the jackknife method. In these figures, we refer to one redshift snapshot and one minimum halo mass only since in all the other cases the results are analogous. In particular, for the differential void density profile, shown in Fig. 3.23, we show the comparison

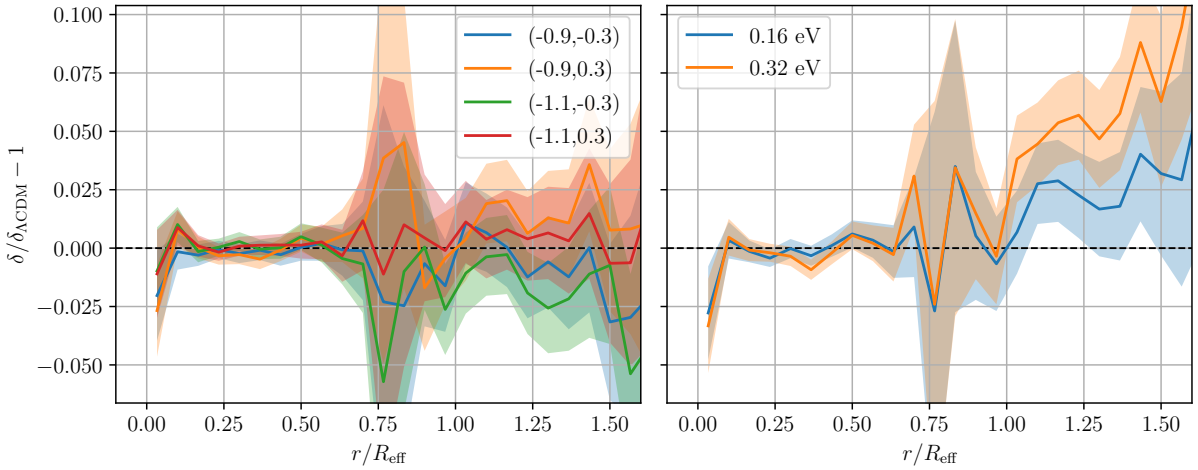

Figure 3.23: Comparison of the void stacked differential density profile in halo distribution with respect to ΛCDM, for voids with minimum radius of $2.5\times$mps detected in the halo distribution at $z = 1.05$, with minimum subhalo mass of $2.5 \times 10^{12} h^{-1} M_{\odot}$. The left panel shows the comparison between ΛCDM and the four sets of parameters of the CPL equation of state in the massless neutrinos case, the right panel shows the comparison in the presence of massive neutrinos and cosmological constant with respect to ΛCDM.

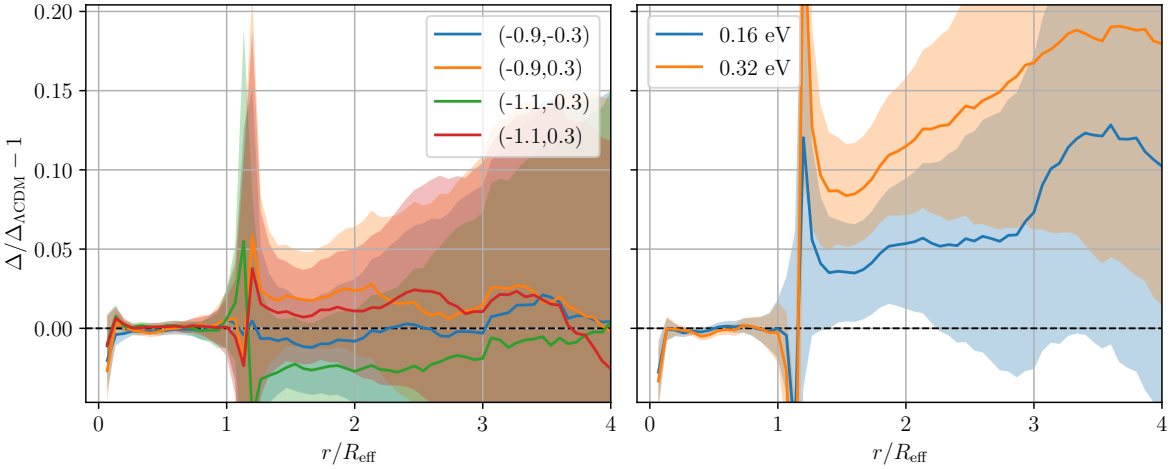

Figure 3.24: Comparison of the void stacked integrated density profile in halo distribution with respect to ΛCDM, for voids with minimum radius of $2.5 \times$ mps detected in the halo distribution at $z = 1.05$, with minimum subhalo mass of $2.5 \times 10^{12} h^{-1} M_{\odot}$. The plot is organized as Fig. 3.23

only up to $\sim 1.5\, r/R_{\mathrm{eff}}$ because beyond this radius the profiles are very close to zero and their ratio becomes numerically uncontrolled. Concerning the integrated profiles, we explore the ratio at all radii, since integrated profiles involve an increasing number of tracers as the density profile is estimated at increasing larger radii, allowing high numerical precision and stability in computing the ratio. In both figures, the discontinuity and the peak of uncertainty occur when the profiles cross zero. Note that the profile for different dark energy equations of state is always degenerate with respect to the ΛCDM one. On the other hand, in the presence of massive neutrinos, the profile seems to distinguish from the ΛCDM one at larger scales, at least in the most massive case. However, note that this ratio is evaluated at large radii, where the absolute value of the density profile is very close to zero. In particular, for the massive

neutrinos case, the profiles remain indistinguishable in the whole underdensity region up to the ridge peak, $r/R_{\rm eff} \simeq 1$. Then the density profile in the presence of massive neutrinos reaches zero values slower with respect to ΛCDM. However, comparing this plot with Fig. 3.16, note that the increase in relative difference starts where the profiles are very close to zero. The same features are present in the integrated profile case.

Note that even if the density profile in the halo distribution remains indistinguishable in the set of cosmological parameters explored, this is not the same for the corresponding dark matter one, due to the difference value of the halo bias corresponding to the various dark energy equations of state and neutrino masses considered. A cosmology independent void density profile in the halo/galaxy distribution can be usefull in analysis involving void density profiles, such as redshift-space distortion around voids or Alcock-Paczyński test. The density profile can be modeled without considering the background cosmology, assuming an empirical fit such as [228]. The dark energy, massive neutrino and other effects would not directly modify the halo/galaxy density profile, but they would impact it via observational geometrical effects, i.e. Alcock-Paczyński, and other observable related to the density profile would be impacted, such as the velocity profile and, consequently, redshift-space distortions sourced both by the dark matter density profile and the background cosmology [141].

3.4 Conclusion

In this chapter, we explored the properties of cosmic voids measured in DEMNUni simulations with the aim of investigating their exploitation as cosmological probes. We focused on the void abundance and the void density profile. For both these quantities, we have extensively discussed their estimation, the tracer impact, the role of the void finder, and their physical properties. More specifically, we investigated the impact of the number density of tracers on the detection of voids and the physical meaning of small voids depending on the tracer type considered, distinguishing between physical and numerical effects. We then also considered the impact of the discreteness of tracers in measuring the void density profile.

For both the void size function and void density profiles, we considered how to theoretically model them, focusing on the properties in the halo distribution since it is closer to cosmological observations. In particular, we consider how the excursion-set void size function can be connected to the observed void abundance for voids in the halo distribution, following [97]. Concerning the void density profile, we reviewed the empirical way to parameterize the halo void density profile with respect to the corresponding stacked dark matter density contrast profile and we presented a theoretical modelization based on the halo mass function in voids. This unprecedented modelization shows that halos collapse in a different way within cosmic voids with respect to the rest of the Universe, and it provides a physical explanation of the relationship between the void density profile in dark matter and in halo distribution, for voids detected in the halo distribution.

We considered the sensitivity of the void size function and void density profiles to the dark energy equation of state and to the total neutrinos mass. In particular, we focused on voids detected in the halo distribution, since they are close to observation. We found that the void abundance is particularly sensitive to the neutrino mass. Concerning dark energy, the sets of parameters of the CPL parameterization producing an equation of state more degenerate with respect to $w = -1$ among the four explored, do not produce detectable effects on the void size function. However, the presence of geometrical distortion in cosmological analysis will greatly

enhance the differences with respect to the ΛCDM case. We showed that since cosmic voids are extended objects, the Alcock-Paczyński effect, together with volume effects, provides a great contribution to distinguishing void abundance in different cosmologies. Moreover, we showed the importance of studying the redshift evolution of voids and that dark energy and massive neutrinos produce separable effects.

In this work, we consider the void size function and the void density profile, since they represent the principal statistical properties of voids and the most directly measurable. In future work, it will be interesting to explore other properties of voids. In particular, a detailed study of the radial void velocity profile and the distribution of velocities would be relevant for redshift-space distortion analyses around voids. The ellipticity of voids is shown to be a cosmological probe [229], so a detailed study of the intrinsic ellipticity of voids and the ellipticity profile of voids will be investigated, together to explore ways to extract this cosmological information from cosmological surveys. Another quantity that will be explored is the void-void correlation function.

In this chapter we explored the sensitivity of voids on dark energy and massive neutrinos, nevertheless, the void sizes make them an interesting probe also for primordial non-Gaussianity. The study of cosmic voids sensitivity to primordial non-Gaussianity is just at its beginning [230], so it would be a relevant new field of voids application to be considered.

Chapter 4

Cosmology with cosmic voids in the Euclid survey

Cosmic voids are vast under-dense regions filling most of the volume of the present-day Universe. With sizes up to hundreds of megaparsec [231–234] they are the largest observable structures in the *cosmic web* [235, 236] – the pattern arising in the galaxy distribution. As we discussed in the previous chapters, voids constitute a unique cosmological probe: their interiors, spanning a large range of scales and featuring low matter density, make them particularly suited to study dark energy and modified gravity [97, 159, 161, 175, 176, 237–248], as well as massive neutrinos [147, 148, 160, 164, 249, 250], primordial non-Gaussianity [230], and physics beyond the standard model [251–255]. Cosmic voids are becoming an effective and competitive new probe of cosmology thanks to the advent of current and upcoming sky surveys such as 6dFGS [256], VIPERS [257], BOSS [258] and eBOSS [259] from the SDSS [260], DES [261], DESI [262], PSF [183], the Roman Space Telescope [184], SPHEREx [263], and LSST [264]. Studying voids requires redshift surveys of very large volume, deep enough in the red band to measure a huge number of redshifts also for low-mass galaxies, and to map in detail significant contiguous fractions of the observable Universe. The *Euclid* survey, expected to sample the sky over 15 000 deg^2, will provide a unique opportunity to capitalize on cosmic voids, to leverage on measurements of the galaxy distribution at large scales and to improve our knowledge on cosmology and fundamental physics. Voids may hold the keys to shed light on some of today's open problems in cosmology [85, and references therein].

Cosmic voids from recent galaxy surveys have been used in a wide range of cosmological applications. They are sensitive to geometric effects, such as the Alcock-Paczyński effect [139, 165, 178, 209, 211, 213] and baryonic acoustic oscillations (BAO) [265–268], as well as redshift-space distortions [RSD, 137, 141, 142, 170, 210, 212, 214–223], weak lensing [169, 185–189, 191–194], and the integrated Sachs–Wolfe effect [140, 143, 189, 195–208].

In this chapter, we consider the void size function, which describes the number density of voids as a function of their size. Over the last two decades, studies of the hierarchical evolution of the void population in the excursion-set framework have allowed the construction of a theoretical void size function model built from first principles, the so-called Sheth & van de Weygaert model [98], later extended by Jennings et al. [96] (see Chapter 2). The void size function and its link to voids detected in galaxy surveys have been explored in depth with cosmological simulations [96, 97, 131, 156, 159, 161, 171–173]: it has proved to be a promising tool to constrain cosmology [97, 159, 160, 164, 172, see also Sect. 3.2.2 and Sect. 3.2.3]. The

void size function has already been measured in surveys [see e.g. 141, 223, 269–271], but it has not yet been used as a stand-alone probe to derive cosmological constraints.

This work is based on the research activities we co-led within the cosmic voids sub-group of the galaxy clustering science working group of the *Euclid* consortium, many of them presented in [162]. This study relies on the largest *Euclid*-like light-cone, the *Flagship* simulation [272], and it belongs to a series of companion projects investigating the scientific return that can be expected from voids observed by the *Euclid* mission. The purpose of this project is to measure and theoretically model the void size function from the Flagship simulation, providing a state-of-the-art forecast for void numbers to be expected from the *Euclid* survey. Our model allows us to estimate the constraining power of the void size function on the dark energy equation of state while also varying the assumed total matter density of the Universe and the total mass of neutrinos. This analysis is focused on voids found in the spectroscopic galaxy distribution, for which the identification of voids is particularly accurate and reliable. We leave for future work the measurement of the void size function in the photometric galaxy distribution from *Euclid*, for which the data treatment greatly differs from the spectroscopic one [see e.g. 182].

This chapter is organized as follows: in Sect. 4.1 we describe the *Euclid* mission; in Sect. 4.2 we introduce the Flagship simulation and describe the void finder and the cleaning algorithm used to obtain the void catalog; in Sect. 4.3 we present the theoretical model of the void size function (Sect. 4.3.1), in particular we describe how to self-consistently align the measured void catalog with the theoretical description (Sect. 4.3.2), we discuss the Bayesian statistical analysis used to perform the cosmological forecasts (Sect. 4.3.3), and finally we introduce the cosmological models considered in this work (Sect. 4.3.4). In Sect. 4.4, we fit the theoretical model to the measured void size function in the Flagship simulation (Sect. 4.4.1) to obtain constraints on the dark energy equation of state and the remaining considered cosmological parameters, with different approaches (Sect. 4.4.2); we conclude by giving a discussion and a summary of our results in Sect. 4.5.

4.1 The Euclid mission

The *Euclid* mission is designed to investigate the origin of the late time acceleration of the Universe, shedding new light on the dark energy problem. More specifically, the main scientific objectives of the *Euclid* mission are [273]:

- Search for a dark energy evolving equation of state, reaching a figure of merit (FoM) higher than 400 concerning the CPL parameterization, Eq. (1.148), see Sect. 4.4.2 for FoM details.
- Measure the growth factor index γ, Eq. (1.98), with a 1-σ precision of < 0.02, allowing to distinguish between General Relativity and a broad class of theories of modified gravity.
- Measure the sum of neutrino masses with a 1-σ better than 0.03 eV.
- Constraint the spectral index of the primordial fluctuation n_s, Eq. (1.146), with percent accuracy when combined with Planck [50] and to constrain primordial non-Gaussianity.

In order to reach these objectives, *Euclid* will measure the expansion rate of the Universe and the growth of cosmic structures using two main probes: weak gravitational lensing and

galaxy clustering, more specifically measuring the BAO pattern. These two probes require two different types of data. Weak lensing analysis relies on photometric measurements to probe the shear field sourcing the distortion of galaxies shape, dimension, and orientation and (de)magnification of their light. Galaxy clustering analyses are performed exploiting redshift measurements that provide an extremely accurate position of galaxies in redshift-space, required to measure the tridimensional galaxy correlation function and power spectrum. These two types of measurements allow to explore the *Euclid* outcome with many other additional cosmological probes.

The two primary probes require two different types of data that will be measured with two instruments: the visible imager (VIS) [274] and the Near Infrared Spectro-Photometer (NISP) [275]. The VIS focal plane is composed by a 6×6 matrix of 12-micron CCDs with a resolution of 4096×4132 pixels, covering a field of view of 0.57 square degrees with 0.1 arc-second pixels. VIS is equipped with one single broad band filter covering the wavelength range from 550 nm to 900 nm with a mean image quality of about 0.23 arc-seconds of angular resolution. The NISP instrument covers a wavelength range between 920 nm and 2000 nm. The NISP focal plane is composed of a 4×4 array of 2048×2048 pixel detectors, and it will cover a field of view of 0.53 square degrees, this results in a pixel scale of 0.3 arcsec. NISP will provide photometric measurements for all the galaxies observed by VIS, and near-infrared slitless spectroscopy. The NISP Photometric Channel (NISP-P) will image the sky with three near-infrared broad band filters: Y (600 – 1192) nm, J (1192 – 1544) nm and H (1544 – 2000) nm. The resulting dataset will have a point source detection limit (5σ) of 24 mag. The NISP Spectroscopic Channel (NISP-S) will perform slitless spectroscopy in two bands: three red grisms allow selection between the 1250 nm and 1850 nm and one blue grism for the 920−1250 nm wavelength range. The channel is a slitless spectrometer, so to help distinguish between different sources, the three red grisms will cover the same wavelength range but will provide spectra with different orientations, so that the spectrometer dispersion direction can be changed to help distinguish overlapping spectra. The channel will be sensitive enough to detect (3.5σ) a 3×10^{-16} erg cm^{-2}s^{-1} line flux [273].

The *Euclid* is a 3-mirror telescope in Korsch configuration, three mirrors allow enough degrees of freedom (three curvatures, three conic constants and two distances between mirrors) to achieve a good level of aberration correction (quasi diffraction limited images), the required image scale, and low distortion along all the *Euclid* field of view of 1.25×0.727 square degree. The primary mirror is on-axis concave with 1.25 m of diameter and will be kept below 130 K with a thermal stability better than 50 mK, the secondary mirror is an off-axis convex, the last mirror is an off-axis concave [273].

Euclid satellite will orbit around the second Lagrangian point of the Sun-Earth orbit. The survey is planned to have a nominal duration of 6 years and is subdivided into two survey types. The first one, known as the *Euclid* Wide Survey, will cover 15 000 square degrees, i.e. about one third of the sky, observing galaxies with a minimum magnitude of 24.5 for VIS and 24 for NISP. This sky area corresponds to the one out from the Galactic plane, contaminated by gas and dust of the Milky Way, and out from the Ecliptic plane, impacted by the strong Zodiacal infrared light contamination due to the dust of the Solar System plane. In the wide survey, *Euclid* will observe a strip of about 15 square degrees per day, which implies patches of about 400 square degrees per month. At the end of the mission, it is expected that *Euclid* will have measured tens of millions of galaxy spectra and billions of photometric galaxies. This is the main part of the mission, which will enable cosmological analysis of weak lensing, galaxy

clustering, and their combination. The second one will be composed of 2 *Euclid* Deep Fields which will observe objects two magnitudes fainter than the Wide survey and will cover about 40 square degrees. The Deep surveys will be used also to assess the purity and completeness of the wide surveys and calibrate their data.

4.2 Galaxy and void catalogs

We now introduce the main tools for our work: the simulation and the void catalogs. This section also includes a brief description of the void finder and of the void catalog preparation.

4.2.1 Flagship simulation

In this work we employ the version 1.8.4 of the *Euclid* Flagship mock galaxy catalog [272]. This catalog is created running a simulation of two trillion dark matter particles in a periodic box of $L = 3780\,h^{-1}$Mpc per side, with a flat ΛCDM cosmology characterized by the parameters $\Omega_{\rm m} = 0.319$, $\Omega_{\rm b} = 0.049$, $\Omega_{\rm de} = 0.681$, $\sigma_8 = 0.83$, $n_{\rm s} = 0.96$ and $h = 0.67$, as obtained by Planck in 2015 [152]. The simulation box has been converted into a light-cone and the dark matter halos have been identified using the `Rockstar` halo finder [32]. These halos have been populated with central and satellite galaxies using a halo occupation distribution (HOD) method [44], to reproduce all the observables relevant for *Euclid* 's main cosmological probes. Specifically, the HOD algorithm has been calibrated exploiting several local observational constraints, using for instance the local luminosity function for the faintest galaxies [276, 277] and the galaxy clustering as a function of luminosity and color [278]. This galaxy sample is composed of more than two billion objects and presents a cut at magnitude $H < 26$ or equivalently on the Hα flux $f_{\rm H\alpha} > 2 \times 10^{-16}$ ergs s^{-1} cm^{-2}, which mimics the observation range expected for *Euclid*. To match the completeness and the spectroscopic performance expected for the *Euclid* survey, we uniformly downsample the galaxy catalog to consider only 60% of the galaxies originally included in it. Furthermore, we associate a Gaussian error of $\sigma_z = 0.001$ to the redshift of each galaxy [279]. The same completeness percentage and the redshift error were used in all the projects of the *Euclid* Cosmic Voids group [144]. Even if more accurate values for these quantities can be explored and implemented, we note that the impact of the redshift error on the void catalog is very mild, because each void contains many galaxies, so the uncertainty on the position is averaged; moreover, the typical size of voids is much larger than the uncertainty of the galaxy position. The full catalog spans a large redshift range, up to $z = 2.3$, and covers one octant of the sky (close to 5157 deg^2).

The *Euclid* satellite will observe 15 000 deg^2 of the sky with patches that extend up to 6 000 deg^2. The total area covered by the satellite will be significantly larger than the available Flagship area. By rescaling it is possible to compute the full predicting power from *Euclid*. The larger *Euclid* survey coverage will allow us to increase statistics, reducing the size of the error bar in particular for the high radius end of the void size function, and to better account for super-sample covariance. On the other hand, the *Euclid* survey is expected to have a less regular pattern than the Flagship box, which might impact the void statistics. Conversely to galaxies, voids are strongly sensitive to survey area specifics because of their extended nature: while contiguous regions are a great advantage for void search, as they provide larger voids, void statistics can be reduced in the case of patchy survey coverage, because voids touching survey edges must be excluded from the analysis. While the interplay between these different

effects may have a role in final constraints, we do not expect this role to significantly impact the precision of constraints resulting from *Euclid*.

We focus our analysis on the expected subsample corresponding to spectroscopic data, selecting galaxies from redshift 0.9 to 1.8. We obtain a resulting mock catalog composed of about 6.5×10^6 galaxies, having the spatial distribution of a shell of sphere octant.

4.2.2 Void finding and catalog preparation

We identify cosmic voids in the Flagship light-cone with the public Void IDentification and Examination toolkit[1] [VIDE, 155], descibed in Sect. 3.1. VIDE can be launched on any catalog of tracers, both on simulation boxes with periodic boundary conditions and on galaxies from real surveys. It is also capable to handle a survey selection function and a mask. These features make VIDE a very flexible tool to study voids in data and simulations. VIDE has been extensively used for cosmological applications relying on voids in the past decade [see e.g. 141, 182, 193, 209, 211, 213, 218, 222, 280].

We build the void catalogs using VIDE from the galaxy sample both with real and redshift-space coordinates given by true and observed redshifts. Note that the redshift-space catalog is identical to the one used in the companion paper of this work [144]. In the true redshift catalog, the galaxy redshift corresponds to the cosmological one only, in the observed redshift catalog it corresponds to the cosmological plus Doppler shift due to peculiar velocity. Although VIDE is a parameter-free algorithm, the theoretical model of the void size function requires voids with the same level of embedded underdensity, so we further process the void catalog. Therefore, we apply to both obtained void catalogs a cleaning algorithm[2] [179] aimed at resizing voids around their volume-weighted barycenter in order to match a specific spherical density contrast within the sphere, $\delta_{\rm v,tr}^{\rm NL}$, in the tracer distribution, as described in Sect. 3.2.3. We modified the algorithm to be applied to catalogs with comoving coordinates, taking into account the variation of the tracer density with redshift the void in measuring the void density profile. The goal of this procedure is to align observed voids with their theoretical counterpart. We underline that any negative value of this density contrast can in principle be chosen to identify underdensities, as long as the theoretical model is consistently calculated using the same threshold (see Sect. 2.1.2 and Sect. 3.2.3, [97]). When dealing with observed voids, the threshold can be fixed to a suitable value chosen based on survey features [97]. We considered the following reasoning to select this value: on the one hand, the more negative the threshold, the stronger the impact of the cosmology on the void size function; on the other hand, an excessively negative threshold entails both a low statistic and a higher uncertainty in the rescaled void radius, caused by the sparsity of galaxies tracing such extreme underdense regions. In addition, a void catalog build with a shallow threshold is impacted by non-ideal effects, such as overlapping voids, systematics concerning the void finder and on the void center position, etc. (see [158, 281] for a discussion on spurious voids and possible treatments). Since, to date, there are no modelizations in the literature for these non-idealities (see the discussion at the end of Sect. 3.2.3), we decided to be conservative in this first application of the void size function to the *Euclid* survey, considering deep enough voids, selecting a threshold $\delta_{\rm v,tr}^{\rm NL} = -0.7$.

VIDE takes into account the presence of a survey mask, and prevents voids from including volumes outside the survey extent. We apply the mask selecting the simulated $\sim$ 5 000 deg^2

[1] https://bitbucket.org/cosmicvoids/vide_public
[2] https://gitlab.com/federicomarulli/CosmoBolognaLib

z range	shell volume $[(h^{-1}\,\mathrm{Gpc})^3]$	mgs $[(h^{-1}\,\mathrm{Mpc})^3]$	$f_{\rm cut}(z)$	VIDE voids	resized
0.950 – 1.035	1.157	10.28	2.30	4989	343
1.035 – 1.126	1.329	11.02	2.24	4935	343
1.126 – 1.208	1.269	11.74	2.18	4232	342
1.208 – 1.318	1.796	12.63	2.12	5302	341
1.318 – 1.455	2.363	13.51	2.06	5935	342
1.455 – 1.700	4.490	14.45	2.00	8435	343
0.950 – 1.700	12.40	13.69	2.15	33 828	2054

Table 4.1: Void counts measured in the redshift-space mock galaxy catalog considering the redshift bins and selections used for this analysis. The first column represents the minimum and the maximum redshift values for each bin, while the second and the third columns provide the volume in units of $(h^{-1}\,\mathrm{Gpc})^3$ corresponding to each shell of the sky octant, and the mean separation between galaxies (mgs), respectively. The fourth column reports the factor, $f_{\rm cut}(z)$, used to select voids unaffected by the incompleteness of counts. The last two columns show the number counts of voids identified by the VIDE void finder with $R > f_{\rm cut}(z) \times$ mgs and of voids obtained after the cleaning procedure with $R_{\rm eff} > f_{\rm cut}(z) \times$ mgs, respectively. In the last row we show the total volume of all redshift shells, the mean mgs, and $f_{\rm cut}(z)$ values and the total void counts corresponding to the entire range of redshifts. Table 4.2 with equi-spaced redshift bins is provided to serve as a reference for void numbers.

octant. Aiming at a very conservative void selection at the edges of the survey's footprint, we apply an additional cut to ensure the mask is not affecting the cleaning procedure: we remove all voids whose center is closer than 30 h^{-1} Mpc to the edge and correct the model accordingly for the selected volume. We then prune voids at low and high redshifts to further avoid selection effects given by redshift boundaries of the light-cone, and we divide the sample in 6 redshift bins. This number is found as the optimal compromise between maximizing the number of redshift shells and keeping void numbers in bins high enough to avoid falling in the shot-noise dominated regime. In order to have shells with roughly the same number of cleaned voids identified in redshift space and to avoid border effects at the light-cone redshift boundaries, we selected the following redshift bin edges: $z_i = [0.950, 1.035, 1.126, 1.208, 1.318, 1.455, 1.700]$, shown in Fig. 4.1. Each shell contains at least 340 voids, within the range of effective radii considered in the analysis of the measured void size function described in the following.

Tracer sparsity leads to a drop of counts for small voids in the measured void size function [97, 225], which depends on the mean galaxy separation and therefore on the redshift of the sample. Modeling the drop in counts for small voids is not trivial. To avoid falling in this regime, we conservatively exclude from the analysis voids with radii falling in the range of scales affected by incompleteness. We remove voids with radii smaller than mgs $f_{\rm cut}(z)$, where mgs is the mean galaxy separation and $f_{\rm cut}(z)$ is a factor dependent on the redshift of the sample. We computed the value of mean galaxy separation as mgs $= (V_{\rm shell}/N_{\rm gal})^{1/3}$, where $V_{\rm shell}$ is the volume of the redshift shell analyzed and $N_{\rm gal}$ is the number of galaxies present in it. The factor $f_{\rm cut}(z)$ is chosen empirically based on the drop of void counts and on the steep departure from the theoretical model. We find that values spanning from 2.3 (lowest redshift bin) to 2 (highest redshift bin) for $f_{\rm cut}(z)$ ensure the exclusion of spatially unresolved voids in redshift space.

Since we expect the resulting void size function in redshift space to be shifted towards

z range	shell volume	mgs	all voids		voids after cleaning	
	$[(h^{-1}\ \mathrm{Gpc})^3]$	$[(h^{-1}\ \mathrm{Mpc})^3]$	$R >$ mgs	> 2 mgs	$R_{\mathrm{eff}} >$ mgs	> 2 mgs
0.9 – 1.0	1.308	10.28	8928	6032	4845	726
1.0 – 1.1	1.427	11.02	8987	6637	5253	840
1.1 – 1.2	1.531	11.74	7735	5824	4690	699
1.2 – 1.3	1.622	12.63	7167	5237	4140	500
1.3 – 1.4	1.700	13.51	6575	4756	3703	321
1.4 – 1.5	1.766	14.45	5636	4078	3152	249
1.5 – 1.6	1.821	15.45	5132	3624	2719	160
1.6 – 1.7	1.867	16.48	4389	3049	2286	94
1.7 – 1.8	1.904	17.63	2248	934	851	4
0.9 – 1.8	14.95	13.69	56 797	40 171	31 639	3593

Table 4.2: Void counts in 9 equi-spaced bins in redshift, measured in the redshift-space mock galaxy catalog, provided as a reference. The first column represents the minimum and the maximum redshift values for each bin, while the second and the third columns provide the volume in units of $(h^{-1}\ \mathrm{Gpc})^3$ corresponding to each shell of the sky octant, and the mean separation between galaxies (mgs), respectively. The next 2 columns show the number of voids identified by the `VIDE` void finder, selected with an effective radius greater than 1 and 2 times the mgs, respectively. In the last 2 columns we provide the void number counts obtained after the cleaning procedure. The latter are reported with the same radius selections as described before. In the last row we show the total shells' volume, the mean mgs and the total void counts corresponding to the entire range of redshifts.

greater effective radii due to the effects of redshift-space distortions [159, 170, 269, 282], we extend the minimum radius for the real-space case, adding an extra bin at small radii while keeping the same binning of the redshift-space case for higher bins. We verified that these choices allow us to be outside of the incompleteness regime, for both the void size function in real and redshift space.

In Table 4.1 we show the number counts of voids selected from the redshift-space mock galaxy catalog. For each of the redshift bins with edges z_i we report the volume occupied by the shell and the mgs of the tracers, together with the factor $f_{\mathrm{cut}}(z)$ used to compute the minimum void radius considered in this analysis. For completeness, we show the void number counts both before and after the cleaning procedure aimed to line up observed voids to theoretical voids, according to the void size function model. The sharp decrease of the void number is an expected outcome of the cleaning procedure, that selects the largest and deepest underdensities identified by `VIDE` and rescales their sizes towards smaller values, causing a more severe rejection of voids during the removal of the spatial scales affected by the incompleteness of counts. Although this conservative approach leads to a loss of the void size function constraining power, it ensures the selection of a high-purity void sample and a robust treatment of void number counts. In future works, different approaches will be explored to improve the void selection also at small radii: among these, the application of machine learning techniques [281] is promising to carefully remove only spurious voids and consequently enhance the performance of the void size function as a cosmological tool.

For reference, in Table 4.2 we provide the number of voids identified in the redshift-space distribution of galaxies, in equi-spaced $\Delta z = 0.1$ bins. Here we report void number counts

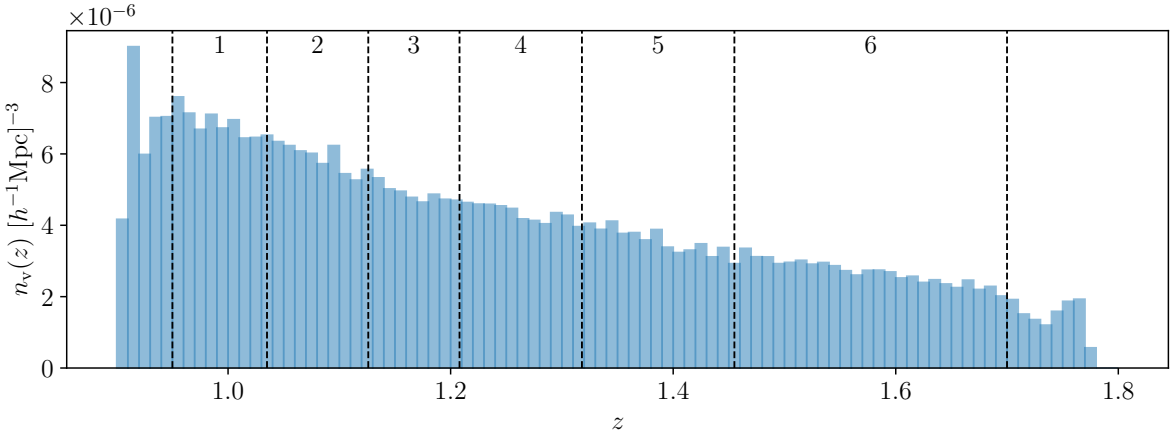

Figure 4.1: Number density of voids as a function of the redshift z for the VIDE redshift-space catalog. The vertical dashed lines show the division of the 6 redshift bins.

obtained both before and after the application of the cleaning procedure. Moreover, we show the number of cosmic voids considering both an optimistic and a pessimistic cut on smaller void radii, that correspond to select voids with radius over $1 \times$ mgs, and voids over $2 \times$ mgs, respectively. The lowering of void counts in the outermost bins is caused by survey mask effects at redshift boundaries of the simulated light-cone, as visible also from Fig. 4.1.

4.3 Theory and methods

In this section we first discuss the model of the void size function, then we present the prescriptions applied to extend this model to voids identified in the distribution of biased tracers. We describe the Bayesian statistical analysis used to provide forecasts on the dark energy equation of state and on the sum of neutrino masses. Finally, we present the cosmological scenarios considered in our analysis.

4.3.1 Theoretical void size function

To estimate the constraining power of the void size function, i.e. the distribution function of void radii, we first need a theoretical model. The void size function model most widely used in the literature relies on the excursion-set formalism, developed within the framework of the halo mass function [see Chapter 2, 12, 25, 101, 103]. As widely discussed in Chapter 2, this model was first proposed by Sheth & van de Weygaert [98] and extended by Jennings et al. [96]. The distribution of fluctuations that become voids, i.e. the multiplicity function, is obtained as the conditional first crossing distribution of the matter density contrast filtered at decreasing Lagrangian radius in a double barrier problem: a fluctuation becomes a void at a radius R_v if the filtered density contrast first crosses the void formation threshold δ_v^L at R_v, without having crossed the threshold for collapse δ_c^L at any larger scale [98]. For the sake of clarity, throughout this chapter, the density contrasts in linear and nonlinear theory are indicated with the superscripts L and NL, respectively. In the absence of any superscript, we take for granted the reference to the nonlinear counterpart. The multiplicity function of Sheth & van de Weygaert [98] is derived for spherical fluctuations in Lagrangian space, i.e. the initial density field

linearly evolved to the epoch of interest, while the observed voids live in the fully nonlinear evolved density field in comoving coordinates, i.e. the Eulerian space. The spherical approximation allows us to easily go back and forth from Lagrangian to Eulerian space in all the computations. The void size function probes the inner region of cosmic voids and in contrast to the collapsing case, i.e. halo formation [14, 15, 283], the spherical approximation is accurate enough for this purpose, at least for voids of scales detectable by *Euclid* [see Sect. 3.2.3, 97, 110].

The linear threshold for collapse is fixed at $\delta_c^L = 1.686$, according to the collapse of a spherical fluctuation. This value corresponds in a matter dominated universe to the full collapse in linear theory, when the halo virializes. The void case is different: if the initial underdensity identifying a void is deep enough, its evolution is not marked by any specific event, and it continues its outward-directed expansion forever. As we widely discussed in Sect. 2.1.2, it is common to consider the shell-crossing, but this condition strictly depends on the initial density profile of the underdensity. For an initial density profile represented as a step function, shell-crossing happens at a nonlinear matter density contrast of $\delta_{sc}^{NL} \simeq -0.8$, corresponding to a linear threshold of $\delta_{sc}^{L} \simeq -2.7$. Considering more physical density profiles [e.g. 284], shell-crossing in voids does not necessarily occur and, if it does, it may occur at even lower threshold values. Given the considered thresholds, our voids remain far from the shell-crossing regime, therefore, in principle, it is always possible to map the measured Eulerian density profile to the corresponding Lagrangian one. As we do not reach shell-crossing, we have the freedom to choose any threshold value to define void formation [97].

The theoretical void size function model that we use for this analysis is the volume conserving model (Vdn) [96].

4.3.2 Methodology

To compare the measured and the theoretical void size functions, we need to link objects found by the void finder in the tracer distribution with the ideal spherical and isolated voids described by the theoretical model void size function model [96]. Any watershed void finder defines a region spanning from its density minimum to its over-dense ridge [155, 156, 158, 285]. On the contrary, theoretical voids are matter density fluctuations for which the mean density contrast in a sphere reaches a specific threshold value at radius $R_{\rm eff}$. Previous papers attempted to mitigate this difference by modifying the threshold of the model [159, 160, 175], in particular considering marginalization over the threshold, for cosmological uses of the model.

It is useful to recall that the Vdn model describes voids evolving in the total matter density field, but that in our case (and when dealing with data) we can only identify voids in the galaxy density field. Therefore, to align these objects with those modeled by the theory, we need to relate the characteristic density threshold used in the theoretical model, δ_v^L, to the corresponding one in the galaxy density field. To accomplish this purpose, we rely on the following two steps for data preparation: first, we measure the mean density profiles of cosmic voids to find the radius of the sphere at which the mean density contrast reaches the desired value $\delta_{v,tr}$ in the galaxy distribution [96, 97, 171, 172, 179], i.e. the resized radius, $R_{\rm eff}$. Second, it is necessary to find the corresponding density contrast in the underlying matter density distribution within the resized radius. Recently, the properties of voids in the galaxy distribution, as well as of galaxies and tracer bias within cosmic voids, have been explored extensively [161, 172, 173, 181, 182, 225, 286]. To recover the matter density contrast corresponding to the threshold

value in the galaxy density field, we need to model the galaxy distribution inside cosmic voids taking into account tracer bias. To describe tracer bias, i.e the bias of the cosmological objects chosen to trace voids, various possibilities have been considered, including a full theoretical description (see discussion in [97]), or modeling the bias inside voids with simulations [161, 172, 181, 182]. In this analysis we chose to rely on the latter, considering a linear relationship between tracer and matter density contrast in cosmic voids, $\delta^{\rm NL}_{\rm v,tr}$ and $\delta^{\rm NL}_{\rm v,DM}$, with a dependence only on the large-scale effective bias $b_{\rm eff}$ [172]:

$$\delta^{\rm NL}_{\rm v,DM} = \frac{\delta^{\rm NL}_{\rm v,tr}}{\mathcal{F}(b_{\rm eff})}\,, \tag{4.1}$$

where $\delta^{\rm NL}_{\rm v,DM}$ is the value of the threshold in the dark matter field to be used in the Vdn model, after its conversion in linear theory [96]. For ΛCDM and the dark energy equations of state considered in this work, the conversion from nonlinear to linear density contrast in the matter field is cosmology and redshift independent with very high accuracy [96], allowing the use of Eq. (2.17) and Eq. (2.22). For numerical convenience, we exploit the Bernardeau formula [287], fitting the map from nonlinear to linear density contrast provided by Eq. (2.17) and Eq. (2.22):

$$\delta^{\rm L}_{\rm v} = C\left[1-(1+\delta^{\rm NL}_{\rm v})^{-1/C}\right]\,,\ \text{with}\ C = 1.594\,. \tag{4.2}$$

It was shown that the ratio between the density contrast of friends-of-friends (FoF) halos in spheres centered in the barycenter of watershed voids and with fixed density contrast and the underlying dark matter density contrast, can be modeled using a linear relation of the large-scale effective bias $b_{\rm eff}$ [172]:

$$\mathcal{F}(b_{\rm eff}) = B_{\rm slope}\, b_{\rm eff} + B_{\rm offset}\,, \tag{4.3}$$

where $B_{\rm slope}$ and $B_{\rm offset}$ are the values of the first and second coefficients of the linear function, respectively. Moreover, the cosmological dependence of the $B_{\rm slope}$ and $B_{\rm slope}$ parameters are negligible with respect to the precision in the void size function measurements [161]. In this chapter, we will denote the linear model with the function $\mathcal{F}(b_{\rm eff})$, used to parameterize the value of the tracer effective bias measured inside cosmic voids, where the measured values are denoted as $b_{\rm punct}$.

To convert the underdensity threshold of the Vdn model according to the function $\mathcal{F}(b_{\rm eff})$, we first need to compute accurately the large-scale effective linear bias of our galaxy sample. For this estimate, we exploit the galaxy two-point correlation function, performing a Bayesian statistical analysis to infer the effective bias, $b_{\rm eff}$. We compute the angle-averaged galaxy two-point correlation function $\hat{\xi}(r)$ in real space creating a random catalog 10 times larger than the original one and using the Landy & Szalay estimator [288]. We then estimate the covariance matrix, which measures the variance and correlation between the different bins of the two-point correlation function. For this purpose we apply the Bootstrap method, dividing the original catalogs in 125 sub-catalogs and constructing 100 realizations by resampling from the sub-catalogs, with replacement. In the end we perform a full Markov chain Monte Carlo (MCMC) analysis of the two-point correlation function, using a Gaussian likelihood function [289, 290]. The two-point correlation function model, $\xi_{\rm mod}(r)$, is computed as follows:

$$\xi_{\rm mod}(r) = b^2_{\rm eff}\,\xi_{\rm m}(r)\,, \tag{4.4}$$

where $\xi_{\rm m}(r)$ is the matter two-point correlation function, which is estimated by Fourier transforming the matter power spectrum, $P_{\rm m}(k)$, computed with the Code for Anisotropies in the Microwave Background [CAMB[3], 291]. The effective bias parameter $b_{\rm eff}$ is then estimated accurately by sampling its posterior distribution with the MCMC modeling in the range of scales of $[20-40]\ h^{-1}$ Mpc.

We underline that the relative error associated to $b_{\rm eff}$ is expected to be relatively small because of the strategy used to compute this quantity relying on the galaxy catalog in real space and assuming the true cosmological parameters of the simulation. A more complete and realistic treatment will be performed in the future, including in the analysis the modeling of the multipoles of the two-point correlation function, which will allow us to take into account the effects of redshift-space and geometrical distortions [see e.g. 168, 292–295].

We finally recall that another approach to compute effective bias, analogous to that applied in this work, is to measure the two-point correlation function in Fourier space and to model it via the theoretical matter power spectrum $P_{\rm m}(k)$ [see e.g. 294]. Additionally, an alternative methodology to extract Flagship galaxy bias is to follow e.g. [296], who parameterized the Flagship galaxy bias as a function of z, albeit for the photometric redshift selection.

4.3.3 Bayesian statistical analysis

In this work we use a reliable forecast method for the sensitivity of the void size function in the *Euclid* survey to constrain the cosmological model, based on a parameter extraction from Bayesian likelihood analysis with MCMC [297–305].

In order to forecast the sensitivity of void counts with an MCMC analysis in *Euclid*, we have to consider that the Flagship simulation covers about one third of the *Euclid* survey. We obtain the *Euclid* predicted void number counts relying on the theoretical void size function model validated on the Flagship simulation, see Sect. 4.4.1, that is assuming a fiducial ΛCDM cosmology with the cosmological parameters of the Flagship and the calibration in redshift space of the Vdn model described in Sect. 4.4.1. We assume the same binning of void radii employed in our Flagship analysis but consider a survey area matching the one expected for *Euclid* (roughly three times the Flagship area), rescaling the Poissonian errors of the void number counts consistently.

This allows us to use MCMC analysis to explore the likelihood distribution in the parameter space without any assumption on the Gaussianity of parameters and local approximations around the fiducial value, as in Fisher forecasts. Moreover, according to the Cramér–Rao inequality, the Fisher matrix gives a lower bound on the error on a parameter [306], while the MCMC is proven to be more realistic, in particular in the presence of degeneracies [297, 302, 305, 307]. Finally, this kind of approach allows us to compute unbiased constraints, with confidence contours centered on the cosmological parameters of the Flagship simulation and on the calibrated nuisance parameters $B_{\rm slope}$ and $B_{\rm offset}$.

According to Bayes's theorem, given a set of data $\mathcal{D}$, the distribution of a set of parameters Θ in the cosmological model considered is given by the posterior probability:

$$\mathcal{P}(\Theta|\mathcal{D}) \propto \mathcal{L}(\mathcal{D}|\Theta)\, p(\Theta)\,, \tag{4.5}$$

where $\mathcal{L}(\mathcal{D}|\Theta)$ is the likelihood and $p(\Theta)$ the prior distribution. Since in this work we consider

[3] http://camb.info

the number counts of cosmic voids, we assume a Poisson likelihood [175]:

$$\mathcal{L}(\mathcal{D}|\Theta) = \prod_{i,j} \frac{N(r_i, z_j|\Theta)^{N(r_i,z_j|\mathcal{D})} \exp\left[-N(r_i, z_j|\Theta)\right]}{N(r_i, z_j|\mathcal{D})!}, \tag{4.6}$$

where the product is over the radius and redshift bins, labeled as i and j respectively. The $N(r_i, z_j|\mathcal{D})$ quantity corresponds to the number of voids in the i^{th} radius bin and j^{th} redshift bin, while $N(r_i, z_j|\Theta)$ corresponds to the expected value in the cosmological model considered, given a set of parameters Θ. In our work, the former is obtained from the Flagship analysis (with the void size function model validated on the Flagship simulation, but considering that the *Euclid* area will be three times larger), while the latter is given by the predictions of the void size function model varying the considered cosmological parameters Θ.

In performing the MCMC analysis, the mapping between redshift and comoving distance changes with the cosmological parameters assumed at each step of the chain. This introduces geometrical distortions for all the considered sets of cosmological parameters (different from the true one), as discussed in Sect. 3.2.2. We used a fiducial cosmology to build up the void catalog, and, in computing the likelihood, we theoretically account for the distortion effects on the quantities we measured. In particular, geometrical distortions can be modeled with two effects: they vary the inferred survey comoving volume and introduce the Alcock-Paczyński distortion [165]. The effect on the survey volume impacts the number of voids expected in the survey. The theoretical void size function model predicts the number density of voids in each radius and redshift bin. Therefore, to obtain the total number of voids, the number density has to be multiplied by the volume, which is impacted by the cosmology. On the other hand, the Alcock-Paczyński distortion affects the size of voids and introduces an anisotropy between the orthogonal and the parallel direction with respect to the line-of-sight. These quantities change according to [169]:

$$r'_\parallel = \frac{H(z)}{H'(z)} r_\parallel = q_\parallel^{-1} r_\parallel, \qquad r'_\perp = \frac{\chi'(z)}{\chi(z)} r_\perp = q_\perp^{-1} r_\perp; \tag{4.7}$$

where $r_\parallel$ and $r_\perp$ are the comoving distances between two objects at redshift z projected along the parallel and perpendicular direction with respect to the line-of-sight, $H(z)$ is the Hubble parameter, and $\chi(z)$ the comoving distance. The primed quantities refer to the calculation at the fiducial cosmology, the non-primed to the true cosmology, assumed in a MCMC step. It follows that the volume of a sphere with radius R appears modified according to $R = q_\parallel^{1/3} q_\perp^{2/3} R'$ [141, 166–170], so the void size function expected in the survey is shifted accordingly. Note that, differently from watershed voids where the implementation of the Alcock-Paczyński effect is exact (see Sect. 3.2.2), for threshold voids this is approximated. A selection of tracers in a sphere estimated in the fiducial cosmology may be mapped in an ellipsoid in the true cosmology, but the multiplicity function we rely on for the theoretical model is defined for spherically filtered fluctuations. Nevertheless, for a wide range of cosmological parameters around the true cosmology, the main effect of the Alcock-Paczyński is to enlarge/contract homogeneously the inferred size of spheres, and as a secondary effect it introduces a small amount of ellipticity. It follows that for our porpoise the Alcock-Paczyński effect can be accurately modeled as modifying the void radius only. We therefore checked the validity of this relationship by varying the cosmology used to get the comoving distances from redshifts and consequently correcting the radius R_{eff} at which voids reach the underdensity threshold $\delta_{\text{v,tr}}^{\text{NL}}$.

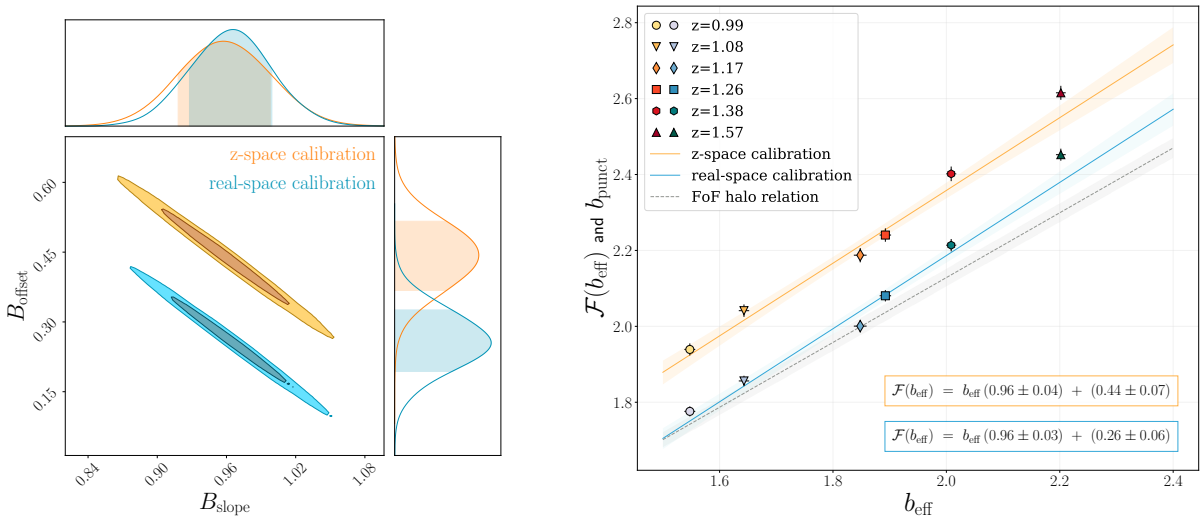

Figure 4.2: Calibration of the relation $\mathcal{F}(b_{\rm eff})$ from Eq. (4.3), required for the conversion of the threshold $\delta_{\rm v,tr}$ in Eq. (4.1). *Left*: 1σ (68%) and 2σ (95%) confidence levels in the $B_{\rm slope}$–$B_{\rm offset}$ plane for the void catalogs built both in real (blue) and in redshift space (orange). *Right*: the solid lines represent the resulting linear relations $\mathcal{F}(b_{\rm eff})$ obtained with the calibrated coefficients $B_{\rm slope}$ and $B_{\rm offset}$ for real (blue) and redshift space (orange), while the shaded regions indicate an uncertainty of 2σ on the relationships. The markers represent the calibration obtained for each bin of redshift, leaving $b_{\rm punct}$ as the only free parameter of the void size function model when fitting the measured void number counts. This alternative calibration provides a value of $b_{\rm punct}$ for each redshift of the sample and is associated with the value of the effective bias $b_{\rm eff}$ of the Flagship galaxies at that specific redshift. As a comparison we show the linear function calibrated using FoF dark matter halos in real space by [172], displayed with a dashed gray line. Figure from [162].

The redshift positions of the void centers are not affected by the Alcock-Paczyński distortion, as discussed in Sect. 3.2.2. While void shapes can suffer from symmetric geometrical distortions, this does not affect the identification of void centers. Furthermore, the variation caused by the change of the cosmological parameters on void radii is taken into account by the modeling of the Alcock-Paczyński effect, therefore the cleaning procedure (see Sect. 4.2.2) is applied only once to the void sample, considering a fiducial ΛCDM cosmology.

We note that the combination of the two effects – volume effect acting on the expected number density, and the Alcock-Paczyński effect acting on the void sizes – enhances the constraining power of the void size function.

4.3.4 Cosmological models

The aim of this work is to investigate the constraining power of the void number count statistic on cosmological parameters, focusing in particular on the dark energy equation-of-state parameters. We consider two cosmological models, extending the standard ΛCDM with different dark energy equation of states. The first model, wCDM, implements a constant dark energy equation of state w; the second one, w_0w_aCDM, parameterizes dynamical dark energy models with the popular Chevallier–Polarski–Linder (CPL) equation of state [58, 59]:

$$w_{\rm CPL}(z) = w_0 + w_a \frac{z}{z+1}\,. \tag{4.8}$$

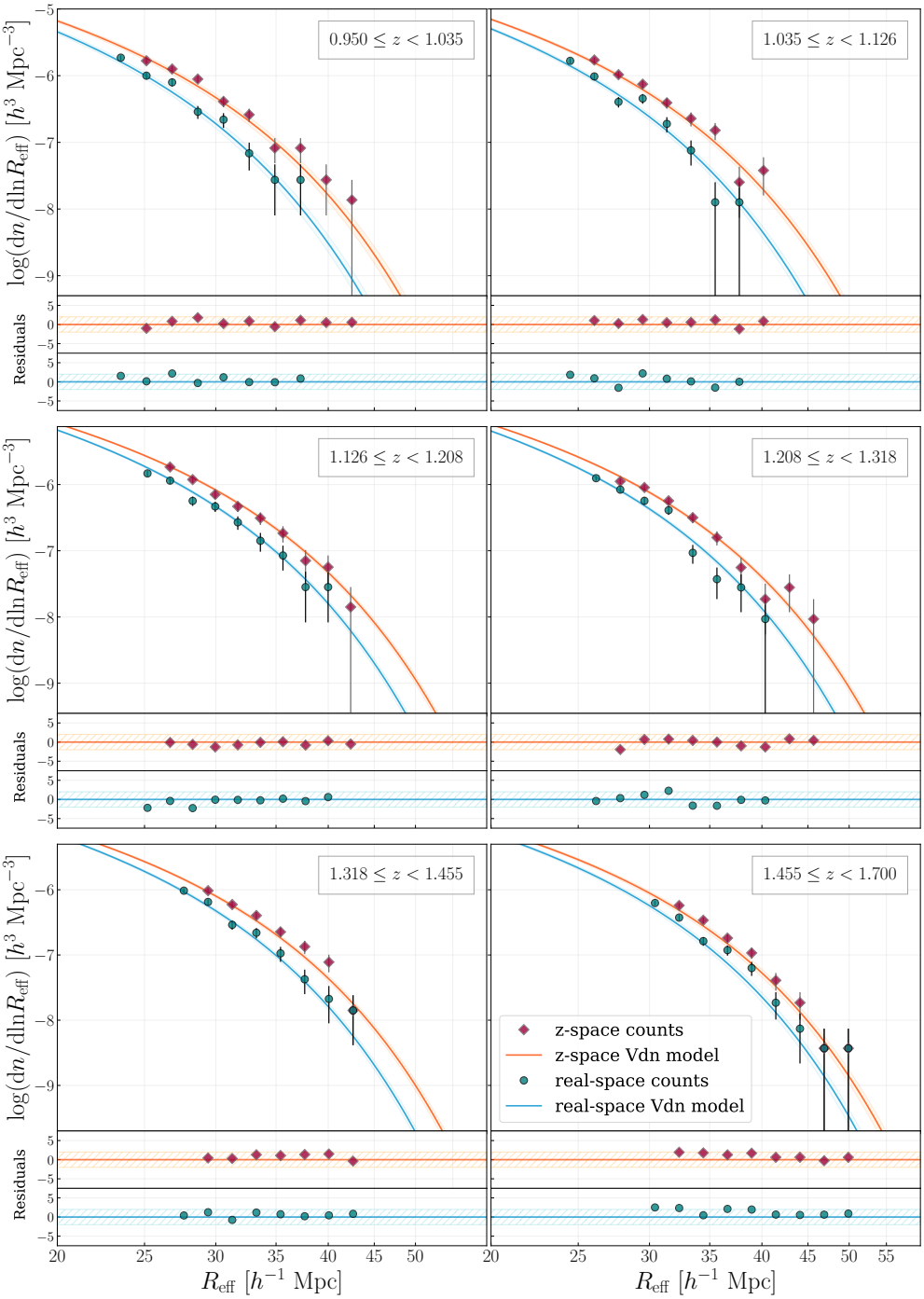

Figure 4.3: Comparison between the measured void number counts as a function of $R_{\rm eff}$ (the void radii rescaled by the cleaning algorithm), and the theoretical predictions given by the extended Vdn model, in 6 different redshift bins. The dark green circles and the dark red diamonds represent the measured void size functions in real and redshift space, respectively, while the corresponding model predictions are depicted in light blue and orange. The shaded regions indicate the uncertainty of 2σ assigned to the model through the calibration of the extended Vdn parameters. Bottom panels report the residuals computed as the difference of data points from the relative theoretical model, divided by the Poissonian error associated with each data point. The hatched regions represent a band with amplitude 2 useful to check if the data points, considered with a 2σ error, are compatible with the main theoretical curve. Figure from [162].

Both cosmological models consider a null space curvature. The MCMC analysis of each cosmological model is performed focusing on different sets of free cosmological parameters: together with the dark energy equation of state parameters (i.e. w or w_0 and w_a, depending on the cosmological model) the density parameter $\Omega_{\rm m}$ or the sum of neutrino masses M_ν are allowed to vary. Moreover, both the cases are analyzed with two different approaches:

1. fixing the parameters of the extended Vdn model, $B_{\rm slope}$ and $B_{\rm offset}$, to the median values obtained from the calibration performed with Flagship data (label: fixed calibration);

2. allowing $B_{\rm slope}$ and $B_{\rm offset}$ to vary in the parameter space described by a 2D Gaussian distribution centered on their median values and given by the calibration with the Flagship simulation (label: relaxed calibration);

The two adopted approaches are meant to demonstrate the impact of the calibration that will be performed in Sect. 4.4.1 on the cosmological forecast. In this work the constraints on the parameters $B_{\rm slope}$ and $B_{\rm offset}$ are indeed limited to the statistical relevance of the number counts of voids identified by means of the Flagship galaxies. The case in which the cosmological forecasts are computed fixing $B_{\rm slope}$ and $B_{\rm offset}$ to their exact calibrated values represents therefore an optimistic evaluation of the results that we may obtain in the future thanks to the usage of larger mock catalogs, or by means of a fully theoretical modeling of the tracer bias inside cosmic voids (see Sect. 4.3.2).

The cosmological model considered for the analysis is characterized by a primordial comoving curvature power spectrum amplitude fixed to the Flagship simulation value, $A_{\rm s} = 2.11 \times 10^{-9}$. We follow the strategy to fix this parameter in order to mimic the future application to real data, which will be supported by the impressive constraints obtained from the study of CMB anisotropies by Planck [50]. Thanks to this approach, for each MCMC step we can derive σ_8, i.e. the root mean square mass fluctuation in spheres with radius 8 h^{-1} Mpc. We rely on `CAMB` to compute this quantity as a derived parameter, which depends on all the cosmological parameters involved in the evolution of the matter power spectrum $P_{\rm m}(k)$.

The density parameter $\Omega_{\rm m}$ is computed as the sum of cold dark matter, baryon, and neutrino energy densities, $\Omega_{\rm m} = \Omega_{\rm cdm} + \Omega_{\rm b} + \Omega_\nu$, and its variation in the Bayesian statistical analysis is balanced by the changing of the dark energy density parameter, $\Omega_{\rm de}$, to keep flat the geometry of the space-time, $\Omega_{\rm de} = 1 - \Omega_{\rm m}$.

The implementation of massive neutrinos in the MCMC analysis is performed considering the sum of the mass of neutrinos as a free parameter in the cosmological model. Neutrinos are modeled with one massive eigenstate and two massless ones, assuming an effective number of neutrino species $N_{\rm eff} = 3.04$ [308, 309] and relating the neutrino mass to the neutrino density parameter as [310]:

$$\Omega_\nu = \frac{M_\nu}{93.14\ h^2\ \text{eV}}\,, \tag{4.9}$$

where we denote $M_\nu = \sum m_\nu$ as the sum of the neutrino mass eigenstates.

Since thermal free-streaming of massive neutrinos suppresses density fluctuations, the abundance of voids changes with massive neutrinos, with respect to the massless neutrinos case [see e.g. 147, 148, 161]. We include the variation of the neutrino density parameter, Ω_ν, in the MCMC analysis, by keeping the value of the total matter density $\Omega_{\rm m}$ fixed (see Sect. 4.2.1), thus rescaling consistently the cold dark matter density parameter $\Omega_{\rm cdm}$.

We rely on `CAMB` for the computation of the total matter power spectrum used to predict the theoretical model of the void size function. The region of the parameter space characterized by

a dark energy equation of state with $w_0 + w_a > 0$ is not covered by `CAMB`, since it corresponds to a non-vanishing dark energy component at the CMB epoch.

4.4 Results

The aim of this section is to compare our theoretical predictions with the void size function measured from the Flagship simulation. We then provide forecasts for the *Euclid* survey, using a Bayesian statistical analysis to predict constraints on the parameters of the dark energy equation of state, modeling the void size function according to the theoretical prescriptions reported in Sect. 4.3.

4.4.1 Void size function analysis

To compare the theoretical void size function with the number counts of voids measured in the galaxy distribution, we need to convert the threshold $\delta_{\rm v,tr}^{\rm NL}$ fixed in measurements to the corresponding one in the matter distribution, as described in Sect. 4.3.2. First of all, we verify the calibration of the relation $\mathcal{F}(b_{\rm eff})$ reported in Eq. (4.3) using the Flagship simulation. To this end, we extract the value of $B_{\rm slope}$ and $B_{\rm offset}$ by leaving them as free parameters with uniform priors of the extended Vdn model and fitting the measured void number counts in the selected redshift bins, considering a Gaussian prior for $b_{\rm eff}$ at each redshift. We notice that, since the error on the effective bias only corresponds to a few percent of its value, the variation allowed for this parameter during the fit is small. All the remaining cosmological parameters are kept fixed to the Flagship simulation values during this calibration.

With this prescription we obtain the confidence levels reported on the left panel of Fig. 4.2, for the void size function measured in both real and redshift space in light blue and orange, respectively. The resulting coefficients for the calibrated relations are:

$$\begin{aligned} \mathcal{F}(b_{\rm eff}) &= (0.96 \pm 0.04)\, b_{\rm eff} + (0.44 \pm 0.07) \quad \text{and} \\ \mathcal{F}(b_{\rm eff}) &= (0.96 \pm 0.03)\, b_{\rm eff} + (0.26 \pm 0.06) \quad , \end{aligned} \tag{4.10}$$

for the redshift-space and the real-space void abundance, respectively.

We show in the right panel of Fig. 4.2 the corresponding linear relations obtained with these calibrations, with a shaded area representing an uncertainty of 2σ. As a comparison, we present in the same plot the values computed for $b_{\rm punct}$, leaving it as the only free parameter of the model and fitting separately the measures at different redshifts. This analysis is aimed at testing the precision of the calibrated relations for each redshift: in the right plot of Fig. 4.2 the markers with the best match to the linear relations correspond in Fig. 4.3 to the redshift bins for which the calibrated model more accurately reproduces the measured void number counts, while points that depart from the linear relationship in Fig. 4.2 (right plot) will lead to a slightly worse agreement between theory and model in Fig. 4.3.

Finally, we report also the calibration obtained using the CoDECS dark matter halos [172, 311], represented in gray in the right panel of Fig. 4.2. At lower redshifts the calibration we measure in this work is in good agreement with the calibration from the CoDECS simulation, characterized by a WMAP7 cosmology [312], but it slightly deviates from the latter at higher redshift values. The reason for this deviation may be twofold. First, it is linked to the kind of cosmic tracers (i.e. dark matter halos or galaxies) and the selection criteria (i.e. minimum mass

or magnitude) used to identify voids. Second, it may be related to the fact that in [172] the calibration was performed for redshift from 0 to 1, while here we are testing this relationship beyond this range.

More importantly, since the void size function will be measured on real data from the *Euclid* survey, we have to deal with voids detected in redshift space. The overall effect of redshift-space distortions on voids, relevant for the void size function, is an apparent enlargement of the voids' volume, due to the elongation along the line of sight. This is reflected in a mean shift of the measured void size function toward larger radii. Even if this effect can in principle be theoretically modeled [159, 170], we decide to parameterize it empirically as described below. Indeed, the theoretical approach requires knowledge of the void matter density profile for the entire void population, which has to be characterized in simulations and may introduce some model dependencies. We found that the parameterization of $\mathcal{F}(b_{\rm eff})$ can be exploited to encapsulate also the modifications on the void sizes caused by the enlargement of cosmic voids in redshift space. This approach has the advantage of being both simple to model and robust, allowing us to take into account, with the same parameter, both the impact of tracer bias in voids and of the redshift-space distortions. Moreover, this approach is fully agnostic and does not require any assumption about the void density profile, nor any other modeling, making it particularly suited to survey analyses.

It is worth noting that the relation obtained for voids in redshift space shows a greater offset but almost the same slope with respect to its analog in real space. This difference reflects the increase of void sizes in redshift space. It also opens the way to test theoretical implementations in future work, indicating that a simple modeling of those effects should suffice to extract robust constraints.

Equipped with these calibrated relations, we now have all the elements necessary to compare the measured void size function with the theoretical predictions given by the Vdn model plus the $\mathcal{F}(b_{\rm eff})$ relation, in which the underdensity threshold is converted as described in Sect. 4.3.2. Figure 4.3 provides the main results of our Flagship analysis. We show the comparison between the measured void number counts and the corresponding theoretical void size functions, both in real and redshift space, for the 6 equi-populated bins in redshift. The Poissonian errors related to the data are represented by the error bars, while the uncertainty related to the theoretical model is shown as a shaded region. The latter is computed associating an error to $\mathcal{F}(b_{\rm eff})$ given by the interval delimited by the colored bands in Fig. 4.2. The residuals are reported at the bottom of each sub-plot and are calculated as the difference from the theoretical model, in units of the data errors. The latter show an excellent agreement between simulated data and theoretical models, even when considering voids identified in the Flagship galaxy catalog in redshift space. The measured void number counts are indeed within an uncertainty of 2σ, shown by the hatched colored bands in the bottom panels, represented in units of the data errors. To test the goodness of the fits shown in Fig. 4.3 we compute the reduced χ^2 using the weighted sum of squared deviations of the two data sets from their corresponding models and dividing the results by the degrees of freedom, ν, of the two systems. The results are $\chi^2_\nu = 1.60$ and $\chi^2_\nu = 1.02$ for real and redshift space, respectively.

4.4.2 Cosmological forecasts

In this subsection we provide the cosmological forecasts obtained using the void size function in redshift space in the perspective of the *Euclid* mission. We apply the statistical analysis

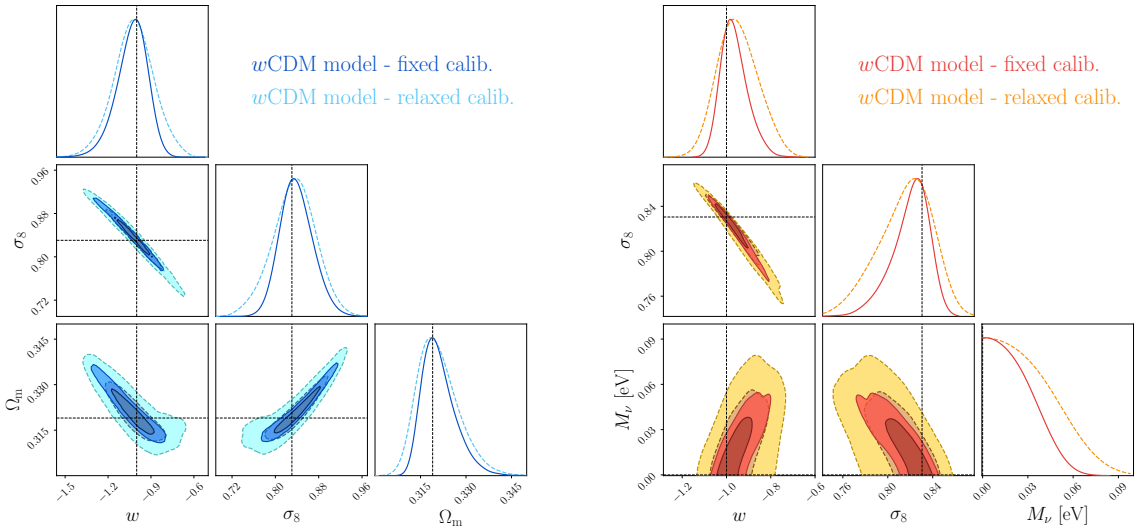

Figure 4.4: Cosmological forecasts for the *Euclid* mission from the void size function for the wCDM model, characterized by a dark energy component described by a constant w. The contours represent the 1σ (68%) and 2σ (95%) confidence levels obtained by means of the Bayesian statistical analysis described in Sect. 4.3.3. *Left*: forecasts for a cosmological model with w and $\Omega_{\rm m}$ as free cosmological parameters. We report the constraints obtained by fixing the calibration parameters with blue contours marked by a solid line and the results obtained by relaxing the calibration constraints with light-blue contours marked by a dashed line (see Sect. 4.3.4). *Right*: forecasts for a cosmological model with w and M_ν as free cosmological parameters. We represent the results of the fixed calibration case as red confidence contours having solid borders and those of the relaxed calibration case as orange contours having dashed borders. For each plot we show also the constraints on σ_8, computed as a derived parameter. The true values of the parameters are shown by a black dashed line. Figure from [162].

described in Sect. 4.3.3 to derive constraints on the parameters of the two cosmological models analyzed, labeled as wCDM and w_0w_aCDM, following the two approaches described in Sect. 4.3.4. For the model wCDM we assume a flat prior for all the remaining free cosmological parameters of the model, and for the model w_0w_aCDM we assume a Gaussian prior distribution with standard deviation $\sigma = 5$ for w_0 and $\sigma = 15$ for w_a, both centered on the true values of these parameters, given by the Flagship simulation cosmology ($w_0 = -1$, $w_a = 0$). We preferred to use very wide Gaussian priors instead of uniform ones to improve the numerical stability of the whole pipeline; nevertheless we tested that uniform priors yield consistent results. The remaining cosmological parameters analyzed in this work ($\Omega_{\rm m}$ and M_ν) are included in the void size function modeling with uniform prior distributions.

In Fig. 4.4 we present the 1σ and 2σ confidence levels of the constraints on the model wCDM. In the left plot we show the *Euclid* forecasts from a void size function model characterized by w and $\Omega_{\rm m}$ as free cosmological parameters. We represent with different colors and borders the results obtained with the two approaches described in Sect. 4.3.4: in blue with solid contours the forecasts obtained by fixing the extended Vdn parameters $B_{\rm slope}$ and $B_{\rm offset}$, in light-blue with dashed contours those obtained by relaxing the calibration constraints by means of a 2D Gaussian prior on $B_{\rm slope}$ and $B_{\rm offset}$, which distribution is represented in the left panel of Fig. 4.2. In the right plot we represent the same forecasts but considering a void size function model with the neutrino total mass M_ν as free parameter instead of the matter density $\Omega_{\rm m}$.

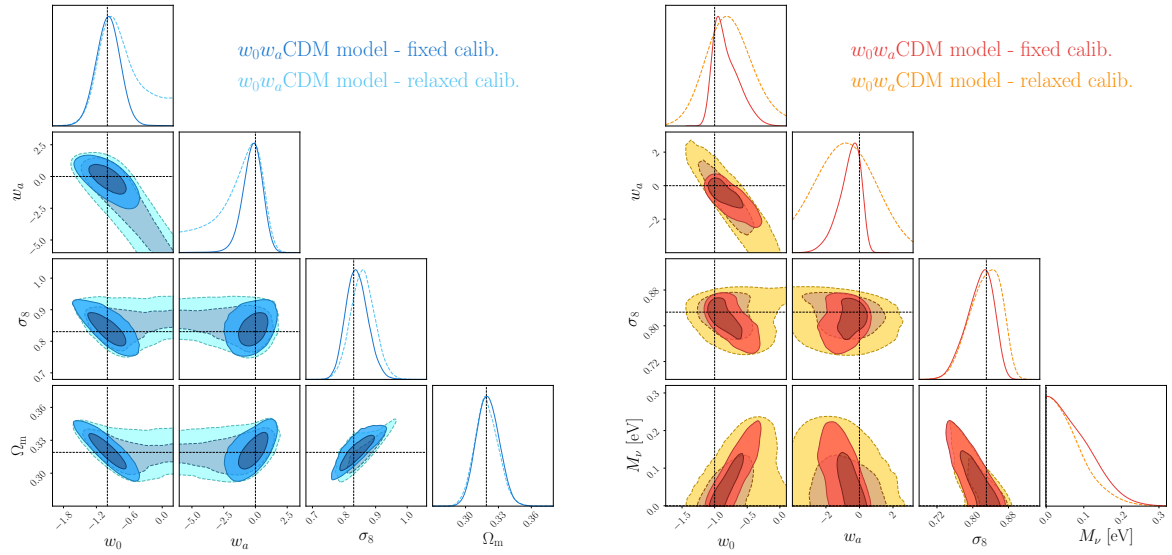

Figure 4.5: The same as Fig. 4.4 but for the cosmological model labeled as w_0w_aCDM, having a dynamical dark energy component described by the CPL parameterization (see Sect. 4.3.4). Figure from [162].

In this case we show the fixed and the relaxed calibration approach results in red and orange, respectively. In both cases presented, σ_8 is computed as derived parameter. As expected, the effect of relaxing the calibration constraints is to broaden the confidence contours.

In Fig. 4.5 we show the same contours represented in Fig. 4.4 but considering the w_0w_aCDM scenario. The free cosmological parameters of the void size function model are the coefficients of the dark energy equation of state, w_0 and w_a, together with Ω_m (left plot) or M_ν (right plot). Also in this case the relaxation of the constraining condition of the calibration parameters causes an enlargement of the confidence contours. In this scenario however, the strongest impact of the calibration constraints is on the w_0–w_a parameter plane, in particular along the diagonal where these parameters become degenerate. The effect of the calibration constraints on Ω_m and M_ν has a lower impact.

In Tables 4.3 and 4.4 we report the values, with relative 1σ errors, of the cosmological constraints derived for the wCDM and w_0w_aCDM scenario, respectively. The constraints on the sum of neutrino masses M_ν are expressed as a 1σ upper limit. For each table we show the results for the two approaches followed in this work: fixing and relaxing the calibration constraints on the void size function model. The calibration parameters are reported in the columns $B_{\rm slope}$ and $B_{\rm offset}$ for completeness. Notice that each quantity reported without any uncertainty is considered fixed in the specific scenario presented in that table's row.

For the w_0w_aCDM scenario, in order to evaluate the constraining power of the void size function on the dark energy equation of state, we derive the FoM for the coefficients of the CPL parameterization w_0 and w_a. We compute this value by following [279]:

$$\mathrm{FoM}_{w_0,w_a} = \frac{1}{\sqrt{\det \mathrm{Cov}(w_0, w_a)}}\,, \tag{4.11}$$

where $\mathrm{Cov}(w_0, w_a)$ represents the covariance matrix of the dark energy equation of state parameters. We report the FoM values in the last column of Table 4.4.

As a first exploration of the cosmic void statistics combined power, we now compare the

Model	w	σ_8	Ω_m	M_ν [eV]	B_{slope}	B_{offset}
fixed calib.	$-1.01^{+0.09}_{-0.11}$	0.83 ± 0.03	$0.319^{+0.005}_{-0.004}$	0	0.96	0.44
	$-0.99^{+0.06}_{-0.04}$	$0.83^{+0.1}_{-0.2}$	0.319	< 0.03	0.96	0.44
relaxed calib.	-1.0 ± 0.1	0.84 ± 0.04	$0.318^{+0.008}_{-0.005}$	0	0.96 ± 0.02	0.44 ± 0.04
	$-0.98^{+0.10}_{-0.07}$	$0.83^{+0.02}_{-0.03}$	0.319	< 0.06	0.95 ± 0.02	0.46 ± 0.04

Table 4.3: Cosmological forecasts computed for the *Euclid* mission from the void size function for the cosmological model wCDM. In this table we report the results of the two analysis strategies adopted in this work: considering the parameters B_{slope} and B_{offset} fixed to the respective median calibrated values (label: fixed calib.) or with a multivariate Gaussian with the same median value but a constraining power given by the calibration procedure with Flagship (label: relaxed calib.). For each of the two cases we present, in the upper and lower line, the forecasts obtained fixing M_ν or Ω_m to the Flagship simulation true values, respectively. All the constraints are reported with errors with a 1σ confidence level.

Model	w_0	w_a	σ_8	Ω_m	M_ν [eV]	B_{slope}	B_{offset}	FoM_{w_0,w_a}
fixed	-1.0 ± 0.2	$-0.1^{+0.7}_{-0.9}$	$0.84^{+0.04}_{-0.03}$	0.32 ± 0.01	0	0.96	0.44	4.9
	$-1.0^{+0.2}_{-0.6}$	$-0.1^{+0.3}_{-0.8}$	$0.83^{+0.02}_{-0.03}$	0.319	< 0.08	0.96	0.44	17
relaxed	$-0.8^{+1.6}_{-0.6}$	$-0.9^{+3.6}_{-9.6}$	0.86 ± 0.04	0.32 ± 0.01	0	$1.01^{+0.03}_{-0.04}$	$0.35^{+0.08}_{-0.05}$	0.78
	$-0.9^{+0.3}_{-0.2}$	$-0.5^{+0.9}_{-1.3}$	$0.86^{+0.02}_{-0.05}$	0.319	< 0.08	$0.99^{+0.01}_{-0.04}$	$0.38^{+0.07}_{-0.01}$	2.3

Table 4.4: The same as Table 4.3 but for the w_0w_aCDM scenario. In this case we present in the last column also the values computed with Eq. (4.11) to estimate the FoM for the dark energy equation of state.

forecasts from the void size function provided in this work with other *Euclid* forecasts. We present as a first comparison the results of the Ω_{de}–w confidence contour with the model-calibrated forecasts presented in [144]. The latter are computed by modeling the observable distortions of average shapes in redshift space via redshift-space distortions and the Alcock-Paczyński effect, for voids to be measured in the *Euclid* spectroscopic galaxy distribution. Contrary to the model-independent case, in the presented approach the nuisance parameters of the model have been calibrated by means of Flagship data. In this comparison we consider the wCDM scenario with fixed neutrino mass and we focus on the Ω_{de}–w parameter space. Given the assumption of flat spatial geometry, to compute the corresponding Ω_{de} forecasts, we converted Ω_m obtained in the MCMC analysis as $\Omega_{de} = 1 - \Omega_m$.

As a second comparison we take the results of Fisher analysis reported in the IST forecasts [279] obtained in the optimistic setting for the single weak lensing and galaxy clustering probes. In this case, we consider the w_0w_a CDM scenario with fixed neutrino mass and we focus on the Ω_m–σ_8 degeneracy. To compute the IST confidence contour we make use of the publicly available[4] Fisher matrices and we marginalize on the parameters not reported in the plot with

[4] See https://github.com/euclidist-forecasting/fisher_for_public.

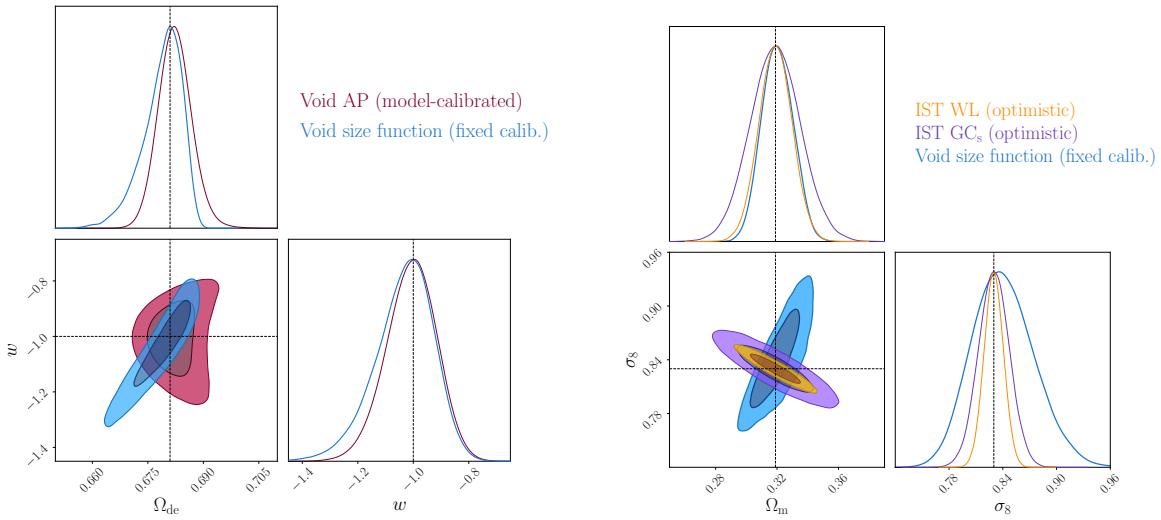

Figure 4.6: Comparison between the 1σ (68%) and 2σ (95%) confidence levels computed in this work with the void size function and different *Euclid* forecasts. *Left*: cosmological constraints on the Ω_{de}–w plane computed in this work (in blue) considering a wCDM scenario with fixed calibration parameters and in [144] (in magenta), modeling the void-galaxy cross-correlation function in redshift space, with a model-calibrated approach. *Right*: cosmological constraints on the Ω_m–σ_8 plane computed in this work considering a w_0w_aCDM scenario (in blue) with fixed calibration parameters and the marginalized IST Fisher forecasts computed in the optimistic setting with spectroscopic galaxy clustering (in purple) and weak lensing (in orange). Figure from [162].

the code `CosmicFish` [313]. We recall that the amplitude of density fluctuations at $z = 0$, σ_8, is computed as a derived parameter in our analysis and its variation is given by the modifications caused by the free cosmological parameters of the model to the total matter power spectrum. We also stress the fact that a larger set of cosmological parameters is used in IST forecasts. This includes in particular the baryon matter energy density, Ω_b, the dimensionless Hubble parameter, h, and the spectral index of the primordial density power spectrum, n_s. The impact on forecasts when including these parameters in the model will be tested in future work.

We show the presented comparisons in Fig. 4.6, representing in blue the forecasts obtained in this work considering the void size function model with fixed calibrated parameters. In the left panel we compare our results with the Ω_{de}–w confidence contour computed with the model-calibrated forecasts presented [144] (in magenta). In the right panel we show instead the comparison of Ω_m–σ_8 confidence contour provided by IST forecasts considering the optimistic setting for weak lensing (in orange) and galaxy clustering (in purple). We show in Fig. 4.7 an analogous comparison considering the cosmological forecasts presented above but with less optimistic settings for the analyses. Also in this more pessimistic scenario, the confidence contours marginalized on the analyzed parameter space are comparable and partially complementary.

In both panels we can appreciate the comparable extension of the presented contours and in the latter we can notice in particular the strong complementarity of the void size function forecasts with those of the *Euclid* standard probes. Although a more accurate analysis would require proper accounting of covariance between analyzed cosmological constraints, Fig. 4.6 shows how the presented probes explore the parameter space differently and motivates investigation on combination to be performed in future works.

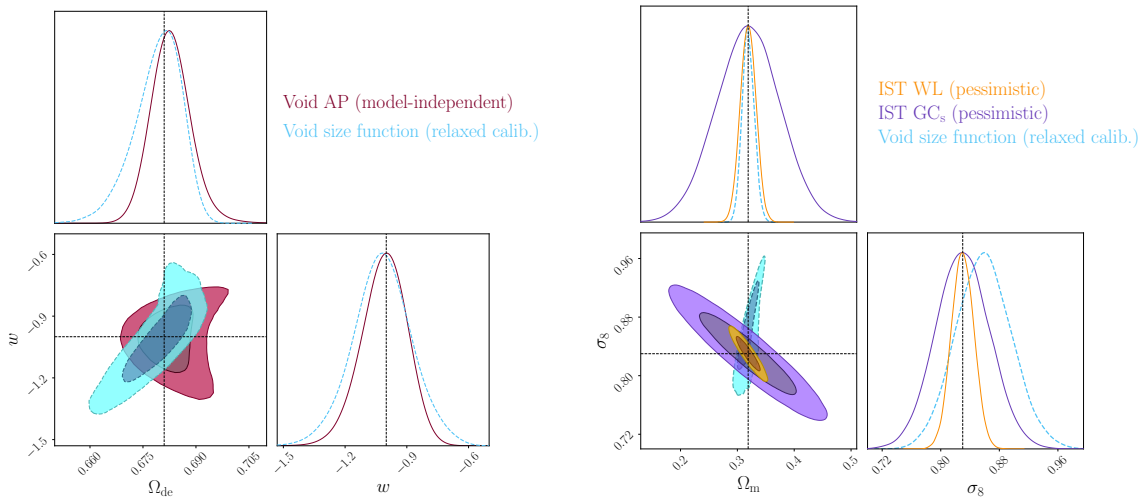

Figure 4.7: Same as Fig. 4.6 but for different forecast settings. In this case the confidence contours obtained in this work from the void size function model (light-blue contours with dashed lines) are computed relaxing the constraints given by calibration parameters. The *Euclid* forecasts derived with void cross-correlation are computed with a model-independent approach, while IST forecasts are computed with the pessimistic setting described [279]. Figure from [162].

4.5 Conclusions and discussion

In this work we presented state-of-the-art forecasts for cosmological constraints from the void size function to be expected from the *Euclid* mission. We measured the void number counts from the Flagship mock galaxy spectroscopic catalog in redshift bins and matched the measurements with the theoretical definition given by the Vdn model [96, 98]. We employed an extension of the Vdn model that conservatively accounts for the effects of the galaxy large-scale bias, $b_{\rm eff}$, on the void effective radii. With this method, we parameterized the Vdn model's characteristic threshold $\delta_{\rm v}^{\rm L}$, also verifying the calibration of the function $\mathcal{F}(b_{\rm eff})$. The parameterization method further allows us to account for the modifications on the void sizes caused by the volume change of cosmic voids in redshift space and to model and mitigate criticalities of the Vdn model.

We show that this extension of the Vdn model calibrated on Flagship data is effective in predicting the measured void number counts both in real and redshift space. Indeed, we obtain a remarkable agreement between the measured and predicted void size functions, for all the redshift bins and all the spatial scales considered in our analysis. We also perform an MCMC analysis, estimating the constraints from void number counts on two main cosmological models: assuming in one case a scenario characterized by a constant equation of state parameter (wCDM) and in the other case a scenario with a dynamical dark energy component described by the CPL parameterization (w_0w_aCDM). For each scenario we presented the *Euclid* cosmological forecasts considering both approaches: by fixing the extended Vdn model parameters and by relaxing their boundaries to those provided by the calibration with Flagship mock catalogs. The former represents the ideal situation in which the simulations used to calibrate the void size function model allow us to have no uncertainties nor systematic errors on the calibration parameters; or alternatively, the case in which the value of the tracer bias inside voids is fully determined thanks to theoretical modeling.

In the wCDM scenario we forecasted relative percentage errors on the constant dark energy

component, w, below the 10% for each analyzed case. In the w_0w_aCDM scenario, with the optimistic approach of fixing the model calibration parameters, we computed a FoM_{w_0,w_a} equal to 4.9 or 17, in the case of leaving Ω_m or M_ν, respectively, as additional free cosmological parameters of the model. The marginalized constraints on the derived parameter σ_8 are lower than 5% in every analyzed case, while the relative errors on Ω_m are of the order of 2% in the wCDM scenario and of 3% in the w_0w_aCDM scenario. The 1σ upper limit on M_ν is instead of 0.03 eV in the most optimistic case of the wCDM scenario and of 0.08 eV in the w_0w_aCDM scenario. We recall that, in the cosmological models with free neutrino mass, the total matter energy density is fixed to the Flagship simulation true value, therefore the degeneracy of Ω_ν with Ω_m is not considered in the results.

Our analysis showcases the constraining power of the void size function from the *Euclid* survey, strongly complementing the *Euclid* primary probes. This complementarity will make the combination powerful, in particular for weak lensing and galaxy clustering, additionally enhancing robustness to systematic effects in both cases.

In this work we considered extremely conservative assumptions when analyzing the void sample. Such conservative assumptions dramatically reduce the statistical power of our void catalogs, to ensure strong reliability: in the future, modeling improvements will allow a more efficient void selection, critically enhancing results while maintaining full robustness.

To enhance the constraining power of the void size function, both the theoretical model and the void identification procedure can be improved. On the theoretical side, in Sect. 2.5 we extensively discussed the criticalities of the Vdn [96] model, presenting improvements to the void size function in the excursion-set framework, concerning both the filtering function of the multiplicity function and the Lagrangian to Eulerian map. Moreover, to recover the void size function for voids in galaxy distribution, we can explore a fully theoretical approach. We can follow the same methodology described in Sect. 3.3.2, modeling the halo mass function within voids. In addition, as shown in Sect. 2.5, the excursion-set mechanism provides a framework to fully theoretically study the halo bias in voids from first principles. This can be done considering the void density profile of threshold voids in Lagrangian space, and the Lagrangian void-halo cross-correlation. In addition to this, studies of the Lagrangian to Eulerian map for the void profile and the void-halo correlation function can be exploited to theoretically describe the effect of redshift-space distortion of voids.

Beyond the improvements on the theoretical side, the void identification procedure has to be optimized to maximize the modelization provided by the theoretical model. In this work, we identified watershed voids that were then post-processed in order to obtain a catalog of threshold voids. In this procedure, systematics affecting the resulting void catalog may be present, such as in the dependence on the watershed void center. A void finder designed to directly detect threshold voids would be an improvement. It would also be interesting to reabsorb the Alcock-Paczyński effects either in the final void catalog or in the theoretical model.

Such kinds of improvements in both the theoretical and in the void identification procedure will allow to relax the conservative choices we did in this work. Among the conservative choices in modeling the void size function and in building the likelihood, we recall the treatment of both the threshold value and of the minimum void radius accepted for the analysis. We strictly restricted the range of considered radii to avoid modeling poorly sampled voids of the Flagship galaxy catalog, in order to prevent the inclusion of spatial scales affected by a loss of void counts. Different techniques will be tested in the future to better model the scales affected by numerical incompleteness [see e.g. 281] and include them in the analysis, safely

obtaining access to much larger statistics. A better modeling of these effects will lead to further improvements in the constraining power of the void size function.

Further future prospects to expand this work include exploiting void number count forecasts to predict constraints from other void applications (such as the stacked void-galaxy cross-correlation function, see [144], void lensing, the void-void correlation function, see [164]), and subsequently combine joint constraints from voids with other *Euclid* probes (primary and not, e.g. galaxy clustering, galaxy weak lensing, cluster counts and clustering, BAO, supernova distance measurements, etc.).

Other areas to explore include considering other cosmological parameters for the likelihood modeling, a more realistic treatment of observational effects (a more complex survey mask, a more realistic σ_z and further survey-related systematic effects).

This work – with a first analysis on a full mock, the *Euclid* Flagship simulation – shows the constraining capability of void number counts to tackle the properties of dark energy and neutrinos, demonstrating for the first time the feasibility of the technique with an end-to-end data-like application, and setting the ground for a robust use of the void size function for cosmology with *Euclid*.

Conclusions

In this thesis I presented the original work that I conducted during my Ph.D. on cosmic voids in the large-scale structure of the Universe as a cosmological probe. This field has become active in the last few years, thanks to the upcoming spectroscopic galaxy surveys that provide the possibility to use cosmic voids as an effective and competitive new probe for cosmology. Cosmological experiments are reaching the percent and sub-percent level accuracy in the estimation of cosmological parameters; this makes even more relevant the exploration of new cosmological probes, complementary to the classical ones, possibly producing tighter constraints when combined in data analysis. In the rather large field of cosmic void applications in cosmology, I focused on the statistics and clustering properties of voids detected in the three-dimensional matter and galaxy/halo distribution. On the theoretical side, I worked on modeling from first principles the cosmic void abundance, correlations, and cross-correlations within the excursion-set framework; I then explored and characterized their statistical properties using large cosmological simulations, evaluating their sensitivity to the dark energy equation of state and to the total neutrino mass; I then investigated the constraining power from the void abundance applied to the upcoming *Euclid* spectroscopic data.

Within the theoretical modeling of the statistical properties of voids in the excursion-set formalism, presented in Chapter 2, I reached relevant and original conclusions and results. First of all, I considered void formation in the excursion-set framework, showing that even if the halo case shows some similarities, a strict parallelism does not allow us to properly describe the formation of voids. In particular, I showed that the shell crossing condition is not relevant to define the void formation event; moreover, I showed that, contrary to the halo case, the map between the evolved density contrast field of voids in the Eulerian space and the corresponding one in the Lagrangian space (almost) always exists. I then discussed the importance of the filtering function in the Langevin equations to the end of physical describing the statistical properties of the initial density field forming proto-halos and proto-voids, and their connection to the fully evolved void and halo distributions. I then provided numerical methods for solving the Langevin equations used to describe void and halo statistics, discussing the advantages and disadvantages of each of the methodologies presented. In particular, for the first time in the literature, I presented how to use the Cholesky decomposition to numerically solve the spatially correlated Langevin equations. Then a broad class of unprecedented results were presented. I showed that the popular Vdn model for the void size function in Eulerian space is non-optimal due to two unphysical assumptions; the first is the filtering kernel adopted, the other concerns the Lagrangian to Eulerian map. I then presented the void size function model using a proper filtering function and assuming spherical symmetry in the Lagrangian to Eulerian map, discussing its range of validity. Subsequently, I explored the two-point correlation functions for the halo-halo, void-void, and void-halo. I showed that the excursion-set formalism is powerful also in the two-point statistics of halo, presenting a toy model in the Lagrangian space and

implementing a linear approximation to consider its Eulerian map. Although the simplifications, the model shows good agreement with simulations. I then considered the Lagrangian void-void and void-halo correlation functions. The latter one is particularly interesting, since it provides a way to model the halo distribution within cosmic voids, which is crucial to theoretically describe void statistics in the galaxy distribution. I concluded by presenting how the Langevin equations allow the modeling of the void density profile in Lagrangian space, the corresponding expectation values, and the distribution around it. I discussed the physical meaning of this result and the importance of its application together with the void-halo correlation function; I then considered the corresponding Eulerian map assuming spherical symmetry. In future works, I plan to explore the Lagrangian to Eulerian map for voids, both from a theoretical point of view and using cosmological simulations. In addition, a suitable void finding algorithm fitted to the excursion-set voids has to be considered and tested.

In the second work presented in this thesis, Chapter 3, I analyzed DEMNUni simulations, a set of large cosmological simulations, to explore some void properties and their sensitivity to the dark energy equation of state and to the total neutrino mass. Accurate cosmological simulations are crucial at the beginning of precision cosmology. In this context, one of the most promising ways to describe cosmological physics relies on the interplay between analytical models and simulations. In approaching this study, I followed this strategy, considering the void size function and the stacked void density profile. On the void abundance side, I discussed its dependence on the adopted void finder, the impact of various tracers, and the physical meaning of voids detected in halo and dark matter distribution. Then I considered the sensitivity to dark energy and massive neutrinos. I showed for the first time in the literature that geometric effects in the observed tracer distribution greatly contribute to enhance the void size function sensitivity to the dark energy equation of state. Contrary to halos, voids are extended objects, so even if the intrinsic effect of dark energy on the void size function is not large, geometrical distortions and volume effects modify the observed void shapes, volumes and number density when a wrong cosmological model is used to transform galaxy redshifts into comovig positions. All these geometrical effects greatly impact the observed void size function. Then I concluded by presenting a way to connect the observed void abundance to the one predicted in the excursion-set framework. I then explored the void density profile, distinguishing between the differential and the integrated one. I discussed their physical meaning and their dependence on the void center definition. Then I investigated various ways to estimate the void density profile in the tracer distribution. Subsequently, I presented a theoretical model that describes the halo bias along the stacked void density profile. This modelization is probably the most important result of this chapter: it provides a physical explanation to the empirical ways used in the literature to model halo bias in voids; it provides a full and accurate theoretical model to predict the void density profile depending on the halos properties; it explores the halo distribution within and around voids, showing that the global halo mass function is not an accurate description of the halo distribution in voids, and suggesting that halo collapse is more spherical within voids with respect to the rest of the Universe. In future work, I will explore the modeling of halo bias within voids, looking for a direct relationship between the p and q parameters of the Sheth-Tormen model and the measured properties of halos, such as halo ellipticity. Moreover, I will explore the accuracy of this model used to extract the stacked dark matter density profile from the one measured in halo distribution. Beyond the void size function and the void density profile, various other statistical void properties can be explored to evaluate their possible exploitation as a cosmological probe. In particular, the radial velocity profile, the velocity

dispersion along the void profile, and the ellipticity profile are promising targets to be studied in detail.

The last work presented in this thesis concerns the work I co-led in *Euclid*, which concerns the cosmological exploitation of voids in the upcoming *Euclid* galaxy survey, Chapter 4. This work explores how the excursion-set model of the void size function can be used to constrain cosmological parameters. The modeling relies on the Vdn model together with an empirical formula, which is used to fit the abundance of voids detected in the redshift-space galaxy distribution. For the first time in the literature, I presented the forecast of the constraining power of the void size function modeled in this way, focusing in particular on the dark energy equation of state, the matter density, and the total neutrino mass. One of the most important results concerns the comparison of the posterior distribution in the parameter space obtained in this analysis with the forecasts of galaxy clustering, weak lensing, and redshift-space distortions plus Alcock-Paczyński effect around voids. A high level of orthogonality is present for some of the explored parameters, showing that the combination of classical cosmological probes together with cosmic void analysis, and possibly with other emerging cosmological probes, may greatly tighten the constraints. This work is the first one that uses the theoretical void size function model for parameter estimation, many further developments are possible. First of all, a fully theoretical model for the void abundance in the redshift-space galaxy distribution can be tested, considering the results presented in Chapter 2 and Chapter 3. Second, this analysis was conducted in a very conservative way, greatly reducing the available statistics. Better modelization would entail better exploitation of the available statistics. As a last point, it would be interesting to consider joint analysis among other void probes, such as redshift-space distortion measurements and Alcock-Paczyński around voids, and with other galaxy clustering and weak lensing probes.

Cosmic voids are a promising probe for cosmology, and a large amount of data will be available soon. To fully exploit their power in probing cosmology, many research activities must be carried out, involving theoretical studies in interplay with cosmological simulations, to better understand their features and systematics to the end of accurate cosmological analysis, possibly extending our understanding of the physics acting in the Universe.

Bibliography

[1] L.D. Landau and E.M. Lifschits, *The Classical Theory of Fields*, vol. Volume 2 of *Course of Theoretical Physics*, Pergamon Press, Oxford (1975).

[2] C.W. Misner, K.S. Thorne and J.A. Wheeler, *Gravitation*, W. H. Freeman, San Francisco (1973).

[3] R.M. Wald, *General Relativity*, Chicago Univ. Pr., Chicago, USA (1984), 10.7208/chicago/9780226870373.001.0001.

[4] R.S. Gonçalves, G.C. Carvalho, C.A.P. Bengaly, J.C. Carvalho and J.S. Alcaniz, *Measuring the scale of cosmic homogeneity with SDSS-IV DR14 quasars*, *Mon. Not. Roy. Astron. Soc.* **481** (2018) 5270 [1809.11125].

[5] J.D. Jackson, *Classical Electrodynamics*, Wiley (1998).

[6] S. Dodelson, *Modern Cosmology*, Academic Press, Amsterdam (2003).

[7] L.D. Landau and E.M. Lifshitz, *Mechanics*, vol. Volume 1 of *Course of Theoretical Physics*, Butterworth-Heinemann, 3 ed. (1976).

[8] F. Bernardeau, S. Colombi, E. Gaztañaga and R. Scoccimarro, *Large-scale structure of the Universe and cosmological perturbation theory*, *Phys. Reports* **367** (2002) 1 [astro-ph/0112551].

[9] W.J. Percival, *Cosmological structure formation in a homogeneous dark energy background*, *Astron. Astrophys.* **443** (2005) 819 [astro-ph/0508156].

[10] W.H. Press and P. Schechter, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, *Astrophys. J.* **187** (1974) 425.

[11] R.I. Epstein, *Proto-galactic perturbations*, *Mon. Not. Roy. Astron. Soc.* **205** (1983) 207.

[12] J.A. Peacock and A.F. Heavens, *Alternatives to the Press-Schechter cosmological mass function*, *Mon. Not. Roy. Astron. Soc.* **243** (1990) 133.

[13] J.R. Bond and S.T. Myers, *The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms*, *Astrophys. J. Suppl. Ser.* **103** (1996) 1.

[14] R.K. Sheth, H.J. Mo and G. Tormen, *Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **323** (2001) 1 [astro-ph/9907024].

[15] R.K. Sheth and G. Tormen, *An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 61 [astro-ph/0105113].

[16] N. Kaiser, *On the spatial correlations of Abell clusters.*, *Astrophys. J. Lett.* **284** (1984) L9.

[17] L.G. Jensen and A.S. Szalay, *N-Point Correlations for Biased Galaxy Formation*, *Astrophys. J. Lett.* **305** (1986) L5.

[18] J.M. Bardeen, J.R. Bond, N. Kaiser and A.S. Szalay, *The Statistics of Peaks of Gaussian Random Fields*, *Astrophys. J.* **304** (1986) 15.

[19] V. Desjacques, D. Jeong and F. Schmidt, *Large-scale galaxy bias*, *Phys. Reports* **733** (2018) 1 [1611.09787].

[20] V. Desjacques, *Baryon acoustic signature in the clustering of density maxima*, *Phys. Rev. D* **78** (2008) 103503 [0806.0007].

[21] V. Desjacques, M. Crocce, R. Scoccimarro and R.K. Sheth, *Modeling scale-dependent bias on the baryonic acoustic scale with the statistics of peaks of Gaussian random fields*, *Phys. Rev. D* **82** (2010) 103529 [1009.3449].

[22] F. Bernardeau, *The Gravity-induced Quasi-Gaussian Correlation Hierarchy*, *Astrophys. J.* **392** (1992) 1.

[23] H.J. Mo, Y.P. Jing and S.D.M. White, *High-order correlations of peaks and haloes: a step towards understanding galaxy biasing*, *Mon. Not. Roy. Astron. Soc.* **284** (1997) 189 [astro-ph/9603039].

[24] P. Fosalba and E. Gaztanaga, *Cosmological Perturbation Theory and the Spherical Collapse model - I. Gaussian initial conditions*, *Mon. Not. Roy. Astron. Soc.* **301** (1998) 503 [astro-ph/9712095].

[25] H.J. Mo and S.D.M. White, *An analytic model for the spatial clustering of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **282** (1996) 347 [astro-ph/9512127].

[26] R.E. Angulo and O. Hahn, *Large-scale dark matter simulations*, *Living Reviews in Computational Astrophysics* **8** (2022) 1 [2112.05165].

[27] A. Knebe, F.R. Pearce, H. Lux, Y. Ascasibar, P. Behroozi, J. Casado et al., *Structure finding in cosmological simulations: the state of affairs*, *Mon. Not. Roy. Astron. Soc.* **435** (2013) 1618 [1304.0585].

[28] C. Lacey and S. Cole, *Merger Rates in Hierarchical Models of Galaxy Formation - Part Two - Comparison with N-Body Simulations*, *Mon. Not. Roy. Astron. Soc.* **271** (1994) 676 [astro-ph/9402069].

[29] W.H. Press and M. Davis, *How to identify and weigh virialized clusters of galaxies in a complete redshift catalog*, *Astrophys. J.* **259** (1982) 449.

[30] J. Einasto, A.A. Klypin, E. Saar and S.F. Shandarin, *Structure of superclusters and supercluster formation - III. Quantitative study of the Local Supercluster.*, *Mon. Not. Roy. Astron. Soc.* **206** (1984) 529.

[31] M. Davis, G. Efstathiou, C.S. Frenk and S.D.M. White, *The evolution of large-scale structure in a universe dominated by cold dark matter*, *Astrophys. J.* **292** (1985) 371.

[32] P.S. Behroozi, R.H. Wechsler and H.-Y. Wu, *The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores*, *Astrophys. J.* **762** (2013) 109 [1110.4372].

[33] J.L. Tinker, B.E. Robertson, A.V. Kravtsov, A. Klypin, M.S. Warren, G. Yepes et al., *The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests*, *Astrophys. J.* **724** (2010) 878 [1001.3162].

[34] Y.P. Jing, H.J. Mo and G. Börner, *Spatial Correlation Function and Pairwise Velocity Dispersion of Galaxies: Cold Dark Matter Models versus the Las Campanas Survey*, *Astrophys. J.* **494** (1998) 1 [astro-ph/9707106].

[35] R.G. Mann, J.A. Peacock and A.F. Heavens, *Eulerian bias and the galaxy density field*, *Mon. Not. Roy. Astron. Soc.* **293** (1998) 209 [astro-ph/9708031].

[36] G. Kauffmann, A. Nusser and M. Steinmetz, *Galaxy formation and large-scale bias*, *Mon. Not. Roy. Astron. Soc.* **286** (1997) 795 [astro-ph/9512009].

[37] C.-P. Ma and J.N. Fry, *Deriving the Nonlinear Cosmological Power Spectrum and Bispectrum from Analytic Dark Matter Halo Profiles and Mass Functions*, *Astrophys. J.* **543** (2000) 503 [astro-ph/0003343].

[38] U. Seljak, *Analytic model for galaxy and dark matter clustering*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 203 [astro-ph/0001493].

[39] J.A. Peacock and R.E. Smith, *Halo occupation numbers and galaxy bias*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 1144 [astro-ph/0005010].

[40] R. Scoccimarro, R.K. Sheth, L. Hui and B. Jain, *How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering*, *Astrophys. J.* **546** (2001) 20 [astro-ph/0006319].

[41] A.A. Berlind and D.H. Weinberg, *The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass*, *Astrophys. J.* **575** (2002) 587 [astro-ph/0109001].

[42] A. Cooray and R. Sheth, *Halo models of large scale structure*, *Phys. Reports* **372** (2002) 1 [astro-ph/0206508].

[43] J.F. Navarro, C.S. Frenk and S.D.M. White, *A Universal Density Profile from Hierarchical Clustering*, *Astrophys. J.* **490** (1997) 493 [astro-ph/9611107].

[44] Z. Zheng, A.L. Coil and I. Zehavi, *Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering*, *Astrophys. J.* **667** (2007) 760 [astro-ph/0703457].

[45] H. Guo, Z. Zheng, I. Zehavi, P.S. Behroozi, C.-H. Chuang, J. Comparat et al., *Redshift-space clustering of SDSS galaxies - luminosity dependence, halo occupation distribution, and velocity bias*, *Mon. Not. Roy. Astron. Soc.* **453** (2015) 4368 [1505.07861].

[46] S. Yuan, D.J. Eisenstein and L.H. Garrison, *Exploring the squeezed three-point galaxy correlation function with generalized halo occupation distribution models*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 2019 [1802.10115].

[47] A. Vale and J.P. Ostriker, *Linking halo mass to galaxy luminosity*, *Mon. Not. Roy. Astron. Soc.* **353** (2004) 189 [astro-ph/0402500].

[48] C. Conroy, R.H. Wechsler and A.V. Kravtsov, *Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time*, *Astrophys. J.* **647** (2006) 201 [astro-ph/0512234].

[49] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [1807.06211].

[50] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[51] J. Lesgourgues and S. Pastor, *Massive neutrinos and cosmology*, *Phys. Reports* **429** (2006) 307 [astro-ph/0603494].

[52] J. Lesgourgues, G. Mangano, G. Miele and S. Pastor, *Neutrino Cosmology*, Cambridge University Press (2013), 10.1017/CBO9781139012874.

[53] M. Lattanzi and M. Gerbino, *Status of neutrino properties and future prospects - Cosmological and astrophysical constraints*, *arXiv e-prints* (2017) arXiv:1712.07109 [1712.07109].

[54] J. Martin, *Everything you always wanted to know about the cosmological constant problem (but were afraid to ask)*, *Comptes Rendus Physique* **13** (2012) 566 [1205.3365].

[55] E.J. Copeland, M. Sami and S. Tsujikawa, *Dynamics of Dark Energy*, *International Journal of Modern Physics D* **15** (2006) 1753 [hep-th/0603057].

[56] L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations*, CambridgeUniversity Press (2010), 10.1017/CBO9780511750823.

[57] T. Clifton, P.G. Ferreira, A. Padilla and C. Skordis, *Modified gravity and cosmology*, *Phys. Reports* **513** (2012) 1 [1106.2476].

[58] M. Chevallier and D. Polarski, *Accelerating Universes with Scaling Dark Matter*, *International Journal of Modern Physics D* **10** (2001) 213 [gr-qc/0009008].

[59] E.V. Linder, *Exploring the Expansion History of the Universe*, *Physical Review Letters* **90** (2003) 091301 [astro-ph/0208512].

[60] A.G. Riess, L. Macri, S. Casertano, H. Lampeitl, H.C. Ferguson, A.V. Filippenko et al., *A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3*, *Astrophys. J.* **730** (2011) 119 [1103.2976].

[61] A.G. Riess, L.M. Macri, S.L. Hoffmann, D. Scolnic, S. Casertano, A.V. Filippenko et al., *A 2.4% Determination of the Local Value of the Hubble Constant*, *Astrophys. J.* **826** (2016) 56 [1604.01424].

[62] J.L. Bernal, L. Verde and A.G. Riess, *The trouble with H_0*, *JCAP* **2016** (2016) 019 [1607.05617].

[63] E. Di Valentino, A. Melchiorri and J. Silk, *Reconciling Planck with the local value of H_0 in extended parameter space*, *Physics Letters B* **761** (2016) 242 [1606.00634].

[64] G. Efstathiou, *A Lockdown Perspective on the Hubble Tension (with comments from the SH0ES team)*, *arXiv e-prints* (2020) arXiv:2007.10716 [2007.10716].

[65] A.G. Riess, W. Yuan, S. Casertano, L.M. Macri and D. Scolnic, *The Accuracy of the Hubble Constant Measurement Verified through Cepheid Amplitudes*, *Astrophys. J. Lett.* **896** (2020) L43 [2005.02445].

[66] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri et al., *In the realm of the Hubble tension-a review of solutions*, *Classical and Quantum Gravity* **38** (2021) 153001 [2103.01183].

[67] G. Efstathiou, *To H_0 or not to H_0?*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 3866 [2103.08723].

[68] A.G. Riess, S. Casertano, W. Yuan, J.B. Bowers, L. Macri, J.C. Zinn et al., *Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM*, *Astrophys. J. Lett.* **908** (2021) L6 [2012.08534].

[69] M.G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi and G. Lambiase, *On the Hubble Constant Tension in the SNe Ia Pantheon Sample*, *Astrophys. J.* **912** (2021) 150 [2103.02117].

[70] A.G. Riess, W. Yuan, L.M. Macri, D. Scolnic, D. Brout, S. Casertano et al., *A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team*, *arXiv e-prints* (2021) arXiv:2112.04510 [2112.04510].

[71] C. Heymans, E. Grocutt, A. Heavens, M. Kilbinger, T.D. Kitching, F. Simpson et al., *CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments*, *Mon. Not. Roy. Astron. Soc.* **432** (2013) 2433 [1303.1808].

[72] N. MacCrann, J. Zuntz, S. Bridle, B. Jain and M.R. Becker, *Cosmic discordance: are Planck CMB and CFHTLenS weak lensing measurements out of tune?*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 2877 [1408.4742].

[73] S. Joudaki, C. Blake, C. Heymans, A. Choi, J. Harnois-Deraps, H. Hildebrandt et al., *CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2033 [1601.05786].

[74] H. Hildebrandt, M. Viola, C. Heymans, S. Joudaki, K. Kuijken, C. Blake et al., *KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 1454 [1606.05338].

[75] M. Asgari, T. Tröster, C. Heymans, H. Hildebrandt, J.L. van den Busch, A.H. Wright et al., *KiDS+VIKING-450 and DES-Y1 combined: Mitigating baryon feedback uncertainty with COSEBIs*, *Astron. Astrophys.* **634** (2020) A127 [1910.05336].

[76] Y. Park and E. Rozo, *Concordance cosmology?*, *Mon. Not. Roy. Astron. Soc.* **499** (2020) 4638 [1907.05798].

[77] S. Joudaki, H. Hildebrandt, D. Traykova, N.E. Chisari, C. Heymans, A. Kannawadi et al., *KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear*, *Astron. Astrophys.* **638** (2020) L1 [1906.09262].

[78] T. Tröster, M. Asgari, C. Blake, M. Cataneo, C. Heymans, H. Hildebrandt et al., *KiDS-1000 Cosmology: Constraints beyond flat ΛCDM*, *Astron. Astrophys.* **649** (2021) A88 [2010.16416].

[79] M. Asgari, C.-A. Lin, B. Joachimi, B. Giblin, C. Heymans, H. Hildebrandt et al., *KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics*, *Astron. Astrophys.* **645** (2021) A104 [2007.15633].

[80] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi et al., *KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, *Astron. Astrophys.* **646** (2021) A140 [2007.15632].

[81] A. Amon, D. Gruen, M.A. Troxel, N. MacCrann, S. Dodelson, A. Choi et al., *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration*, *Phys. Rev. D* **105** (2022) 023514 [2105.13543].

[82] L.F. Secco, S. Samuroff, E. Krause, B. Jain, J. Blazek, M. Raveri et al., *Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty*, *Phys. Rev. D* **105** (2022) 023515 [2105.13544].

[83] T.M.C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon et al., *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **105** (2022) 023520 [2105.13549].

[84] G. Verza, S. Matarrese, C. Porciani, A. Renzi and C. Carbone, *The statistic of cosmic voids in Lagrangian space*, (2022, in preparation) .

[85] A. Pisani, E. Massara, D.N. Spergel, D. Alonso, T. Baker, Y.-C. Cai et al., *Cosmic voids: a novel probe to shed light on our Universe*, BAAS **51** (2019) 40 [1903.05161].

[86] M. Moresco, L. Amati, L. Amendola, S. Birrer, J.P. Blakeslee, M. Cantiello et al., *Unveiling the Universe with Emerging Cosmological Probes*, *arXiv e-prints* (2022) arXiv:2201.07241 [2201.07241].

[87] N. Straumann, *General Relativity*, Graduate Texts in Physics, Springer, Dordrecht (2013), 10.1007/978-94-007-5410-2.

[88] J.E. Gunn and I. Gott, J. Richard, *On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution*, *Astrophys. J.* **176** (1972) 1.

[89] D.M. Goldberg and M.S. Vogeley, *Simulating Voids*, *Astrophys. J.* **605** (2004) 1 [astro-ph/0307191].

[90] C. Wagner, F. Schmidt, C.T. Chiang and E. Komatsu, *Separate universe simulations.*, *Mon. Not. Roy. Astron. Soc.* **448** (2015) L11 [1409.6294].

[91] L. Dai, E. Pajer and F. Schmidt, *On separate universes*, *JCAP* **2015** (2015) 059 [1504.00351].

[92] F. Schmidt, W. Hu and M. Lima, *Spherical collapse and the halo model in braneworld gravity*, *Phys. Rev. D* **81** (2010) 063005 [0911.5178].

[93] M. Kopp, S.A. Appleby, I. Achitouv and J. Weller, *Spherical collapse and halo mass function in f(R) theories*, *Phys. Rev. D* **88** (2013) 084015 [1306.3233].

[94] L. Lombriser, *A parametrisation of modified gravity on nonlinear cosmological scales*, *JCAP* **2016** (2016) 039 [1608.00522].

[95] D. Herrera, I. Waga and S.E. Jorás, *Calculation of the critical overdensity in the spherical-collapse approximation*, *Phys. Rev. D* **95** (2017) 064029 [1703.05824].

[96] E. Jennings, Y. Li and W. Hu, *The abundance of voids and the excursion set formalism*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) 2167 [1304.6087].

[97] G. Verza, A. Pisani, C. Carbone, N. Hamaus and L. Guzzo, *The void size function in dynamical dark energy cosmologies*, *JCAP* **2019** (2019) 040 [1906.00409].

[98] R.K. Sheth and R. van de Weygaert, *A hierarchy of voids: much ado about nothing*, *Mon. Not. Roy. Astron. Soc.* **350** (2004) 517 [astro-ph/0311260].

[99] D. Lynden-Bell, *Statistical mechanics of violent relaxation in stellar systems*, *Mon. Not. Roy. Astron. Soc.* **136** (1967) 101.

[100] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [1905.05697].

[101] C. Lacey and S. Cole, *Merger rates in hierarchical models of galaxy formation*, *Mon. Not. Roy. Astron. Soc.* **262** (1993) 627.

[102] C. Porciani, S. Matarrese, F. Lucchin and P. Catelan, *Excursion set approach to the clustering of dark matter haloes in Lagrangian space*, *Mon. Not. Roy. Astron. Soc.* **298** (1998) 1097 [astro-ph/9801290].

[103] J.R. Bond, S. Cole, G. Efstathiou and N. Kaiser, *Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations*, *Astrophys. J.* **379** (1991) 440.

[104] A.N. Borodin and P. Salminen, *Handbook of Brownian motion: facts and formulae*, Springer (2002).

[105] M. Pinsky and S. Karlin, *An Introduction to Stochastic Modeling*, AN INTRODUCTION TO STOCHASTIC, Elsevier Science (2011).

[106] S. Chandrasekhar, *Stochastic problems in physics and astronomy*, *Rev. Mod. Phys.* **15** (1943) 1.

[107] D. Lynden-Bell, *On Large-Scale Instabilities during Gravitational Collapse and the Evolution of Shrinking Maclaurin Spheroids.*, *Astrophys. J.* **139** (1964) 1195.

[108] C.C. Lin, L. Mestel and F.H. Shu, *The Gravitational Collapse of a Uniform Spheroid.*, *Astrophys. J.* **142** (1965) 1431.

[109] S.D.M. White and J. Silk, *The growth of aspherical structure in the universe: Is the Local Supercluster an unusual system?*, *Astrophys. J.* **231** (1979) 1.

[110] V. Icke, *Voids and filaments*, *Mon. Not. Roy. Astron. Soc.* **206** (1984) 1P.

[111] L. Hui and E. Bertschinger, *Local Approximations to the Gravitational Collapse of Cold Matter*, *Astrophys. J.* **471** (1996) 1 [astro-ph/9508114].

[112] Y. Ohta, I. Kayo and A. Taruya, *Cosmological Density Distribution Function from the Ellipsoidal Collapse Model in Real Space*, *Astrophys. J.* **608** (2004) 647 [astro-ph/0402618].

[113] R.K. Sheth, K.C. Chan and R. Scoccimarro, *Nonlocal Lagrangian bias*, *Phys. Rev. D* **87** (2013) 083002 [1207.7117].

[114] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, *Precision measurement of the local bias of dark matter halos*, *JCAP* **2016** (2016) 018 [1511.01096].

[115] A. Paranjape, T.Y. Lam and R.K. Sheth, *Halo abundances and counts-in-cells: the excursion set approach with correlated steps*, *Mon. Not. Roy. Astron. Soc.* **420** (2012) 1429 [1105.1990].

[116] L. Appel and B.J.T. Jones, *The Mass Function in Biased Galaxy Formation Scenarios*, *Mon. Not. Roy. Astron. Soc.* **245** (1990) 522.

[117] M. Maggiore and A. Riotto, *The Halo Mass Function from Excursion Set Theory. I. Gaussian Fluctuations with Non-Markovian Dependence on the Smoothing Scale*, *Astrophys. J.* **711** (2010) 907 [0903.1249].

[118] A. de Simone, M. Maggiore and A. Riotto, *Excursion set theory for generic moving barriers and non-Gaussian initial conditions*, *Mon. Not. Roy. Astron. Soc.* **412** (2011) 2587 [1007.1903].

[119] C.-P. Ma, M. Maggiore, A. Riotto and J. Zhang, *The bias and mass function of dark matter haloes in non-Markovian extension of the excursion set theory*, *Mon. Not. Roy. Astron. Soc.* **411** (2011) 2644 [1007.4201].

[120] M. Musso and R.K. Sheth, *One step beyond: the excursion set approach with correlated steps*, *Mon. Not. Roy. Astron. Soc.* **423** (2012) L102 [1201.3876].

[121] A. Paranjape and R.K. Sheth, *Peaks theory and the excursion set approach*, *Mon. Not. Roy. Astron. Soc.* **426** (2012) 2789 [1206.3506].

[122] M. Musso, A. Paranjape and R.K. Sheth, *Scale-dependent halo bias in the excursion set approach*, *Mon. Not. Roy. Astron. Soc.* **427** (2012) 3145 [1205.3401].

[123] M. Musso and R.K. Sheth, *On the Markovian assumption in the excursion set approach: the approximation of Markovian Velocities*, *Mon. Not. Roy. Astron. Soc.* **443** (2014) 1601 [1401.8177].

[124] M. Musso and R.K. Sheth, *The importance of stepping up in the excursion set approach*, *Mon. Not. Roy. Astron. Soc.* **438** (2014) 2683 [1306.0551].

[125] T. Verechtchaguina, I.M. Sokolov and L. Schimansky-Geier, *First passage time densities in non-Markovian models with subthreshold oscillations*, *EPL (Europhysics Letters)* **73** (2006) 691 [cond-mat/0506811].

[126] M. Musso and R.K. Sheth, *The excursion set approach in non-Gaussian random fields*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 3051 [1305.0724].

[127] F. Nikakhtar, M. Ayromlou, S. Baghram, S. Rahvar, M.R. Rahimi Tabar and R.K. Sheth, *The Excursion set approach: Stratonovich approximation and Cholesky decomposition*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 5296 [1802.04207].

[128] R. Mannella, *Absorbing boundaries and optimal stopping in a stochastic differential equation*, *Physics Letters A* **254** (1999) 257 .

[129] R. Mannella, *A gentle introduction to the integration of stochastic differential equations*, in *Stochastic Processes in Physics, Chemistry, and Biology*, J.A. Freund and T. Pöschel, eds., (Berlin, Heidelberg), pp. 353–364, Springer Berlin Heidelberg, 2000.

[130] M. Viel, S. Matarrese, H.J. Mo, M.G. Haehnelt and T. Theuns, *Probing the intergalactic medium with the Lyα forest along multiple lines of sight to distant QSOs*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 848 [astro-ph/0105233].

[131] A. Paranjape, T.Y. Lam and R.K. Sheth, *A hierarchy of voids: more ado about nothing*, *Mon. Not. Roy. Astron. Soc.* **420** (2012) 1648 [1106.2041].

[132] I. Achitouv, M. Neyrinck and A. Paranjape, *Testing spherical evolution for modelling void abundances*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 3964 [1309.3799].

[133] P. Catelan, F. Lucchin, S. Matarrese and C. Porciani, *The bias field of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **297** (1998) 692 [astro-ph/9708067].

[134] T. Baldauf, U. Seljak, R.E. Smith, N. Hamaus and V. Desjacques, *Halo stochasticity from exclusion and nonlinear clustering*, *Phys. Rev. D* **88** (2013) 083507 [1305.2917].

[135] R. García and E. Rozo, *Halo exclusion criteria impacts halo statistics*, *Mon. Not. Roy. Astron. Soc.* **489** (2019) 4170 [1903.01709].

[136] R. García, E. Rozo, M.R. Becker and S. More, *A redefinition of the halo boundary leads to a simple yet accurate halo model of large-scale structure*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 1195 [2006.12751].

[137] N. Hamaus, B.D. Wandelt, P.M. Sutter, G. Lavaux and M.S. Warren, *Cosmology with Void-Galaxy Correlations*, *Phys. Rev. L* **112** (2014) 041304 [1307.2571].

[138] N. Hamaus, P.M. Sutter, G. Lavaux and B.D. Wand elt, *Testing cosmic geometry without dynamic distortions using voids*, *JCAP* **2014** (2014) 013 [1409.3580].

[139] Q. Mao, A.A. Berlind, R.J. Scherrer, M.C. Neyrinck, R. Scoccimarro, J.L. Tinker et al., *Cosmic Voids in the SDSS DR12 BOSS Galaxy Sample: The Alcock-Paczynski Test*, *Astrophys. J.* **835** (2017) 160 [1602.06306].

[140] A. Kovács, C. Sánchez, J. García-Bellido, J. Elvin-Poole, N. Hamaus, V. Miranda et al., *More out of less: an excess integrated Sachs-Wolfe signal from supervoids mapped out by the Dark Energy Survey*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 5267 [1811.07812].

[141] N. Hamaus, A. Pisani, J.-A. Choi, G. Lavaux, B.D. Wandelt and J. Weller, *Precision cosmology with voids in the final BOSS data*, *JCAP* **2020** (2020) 023 [2007.07895].

[142] S. Nadathur, A. Woodfinden, W.J. Percival, M. Aubert, J. Bautista, K. Dawson et al., *The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: geometry and growth from the anisotropic void-galaxy correlation function in the luminous red galaxy sample*, *arXiv e-prints* (2020) arXiv:2008.06060 [2008.06060].

[143] A. Kovács, R. Beck, A. Smith, G. Rácz, I. Csabai and I. Szapudi, *Evidence for a high-z ISW signal from supervoids in the distribution of eBOSS quasars*, *arXiv e-prints* (2021) arXiv:2107.13038 [2107.13038].

[144] N. Hamaus, M. Aubert, A. Pisani, S. Contarini, G. Verza, M.C. Cousinou et al., *Euclid: Forecasts from redshift-space distortions and the Alcock-Paczynski test with cosmic voids*, *Astron. Astrophys.* **658** (2022) A20 [2108.10347].

[145] C. Carbone, M. Petkova and K. Dolag, *DEMNUni: ISW, Rees-Sciama, and weak-lensing in the presence of massive neutrinos*, *JCAP* **7** (2016) 034 [1605.02024].

[146] E. Castorina, C. Carbone, J. Bel, E. Sefusatti and K. Dolag, *DEMNUni: the clustering of large-scale structures in the presence of massive neutrinos*, *JCAP* **7** (2015) 043 [1505.07148].

[147] C.D. Kreisch, A. Pisani, C. Carbone, J. Liu, A.J. Hawken, E. Massara et al., *Massive neutrinos leave fingerprints on cosmic voids*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 4413 [1808.07464].

[148] N. Schuster, N. Hamaus, A. Pisani, C. Carbone, C.D. Kreisch, G. Pollina et al., *The bias of cosmic voids in the presence of massive neutrinos*, *JCAP* **2019** (2019) 055 [1905.00436].

[149] V. Springel, *The cosmological simulation code GADGET-2*, *Mon. Not. Roy. Astron. Soc.* **364** (2005) 1105 [astro-ph/0505010].

[150] M. Viel, M.G. Haehnelt and V. Springel, *The effect of neutrinos on the matter distribution as probed by the intergalactic medium*, *JCAP* **2010** (2010) 015 [1003.2422].

[151] Planck Collaboration, P.A.R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown et al., *Planck 2013 results. XVI. Cosmological parameters*, *Astron. Astrophys.* **571** (2014) A16 [1303.5076].

[152] Planck Collaboration, P.A.R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [1502.01589].

[153] V. Springel, N. Yoshida and S.D.M. White, *GADGET: a code for collisionless and gasdynamical cosmological simulations*, *New Astronomy* **6** (2001) 79 [astro-ph/0003162].

[154] K. Dolag, S. Borgani, G. Murante and V. Springel, *Substructures in hydrodynamical cluster simulations*, *Mon. Not. Roy. Astron. Soc.* **399** (2009) 497 [0808.3401].

[155] P.M. Sutter, G. Lavaux, N. Hamaus, A. Pisani, B.D. Wandelt, M. Warren et al., *VIDE: The Void IDentification and Examination toolkit*, *Astronomy and Computing* **9** (2015) 1 [1406.1191].

[156] E. Platen, R. van de Weygaert and B.J.T. Jones, *A cosmic watershed: the WVF void detection technique*, *Mon. Not. Roy. Astron. Soc.* **380** (2007) 551 [0706.2788].

[157] W.E. Schaap and R. van de Weygaert, *Continuous fields and discrete samples: reconstruction through Delaunay tessellations*, *Astron. Astrophys.* **363** (2000) L29 [astro-ph/0011007].

[158] M.C. Neyrinck, *ZOBOV: a parameter-free void-finding algorithm*, *Mon. Not. Roy. Astron. Soc.* **386** (2008) 2101 [0712.3049].

[159] A. Pisani, P.M. Sutter, N. Hamaus, E. Alizadeh, R. Biswas, B.D. Wandelt et al., *Counting voids to probe dark energy*, *Phys. Rev. D* **92** (2015) 083531 [1503.07690].

[160] M. Sahlén, *Cluster-void degeneracy breaking: Neutrino properties and dark energy*, *Phys. Rev. D* **99** (2019) 063525 [1807.02470].

[161] S. Contarini, F. Marulli, L. Moscardini, A. Veropalumbo, C. Giocoli and M. Baldi, *Cosmic voids in modified gravity models with massive neutrinos*, *Mon. Not. Roy. Astron. Soc.* **504** (2021) 5021 [2009.03309].

[162] S. Contarini, G. Verza, A. Pisani, N. Hamaus, M. Sahlén, C. Carbone et al., *Euclid: Cosmological forecasts from the void size function*, *arXiv e-prints* (2022) arXiv:2205.11525 [2205.11525].

[163] A.E. Bayer, F. Villaescusa-Navarro, E. Massara, J. Liu, D.N. Spergel, L. Verde et al., *Detecting Neutrino Mass by Combining Matter Clustering, Halos, and Voids*, *Astrophys. J.* **919** (2021) 24 [2102.05049].

[164] C.D. Kreisch, A. Pisani, F. Villaescusa-Navarro, D.N. Spergel, B.D. Wandelt, N. Hamaus et al., *The GIGANTES dataset: precision cosmology from voids in the machine learning era*, *arXiv e-prints* (2021) arXiv:2107.02304 [2107.02304].

[165] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358.

[166] W.E. Ballinger, J.A. Peacock and A.F. Heavens, *Measuring the cosmological constant with redshift surveys*, *Mon. Not. Roy. Astron. Soc.* **282** (1996) 877 [astro-ph/9605017].

[167] D.J. Eisenstein, I. Zehavi, D.W. Hogg, R. Scoccimarro, M.R. Blanton, R.C. Nichol et al., *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, *Astrophys. J.* **633** (2005) 560 [astro-ph/0501171].

[168] X. Xu, A.J. Cuesta, N. Padmanabhan, D.J. Eisenstein and C.K. McBride, *Measuring D_A and H at z=0.35 from the SDSS DR7 LRGs using baryon acoustic oscillations*, *Mon. Not. Roy. Astron. Soc.* **431** (2013) 2834 [1206.6732].

[169] A.G. Sánchez, R. Scoccimarro, M. Crocce, J.N. Grieb, S. Salazar-Albornoz, C. Dalla Vecchia et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the configuration-space clustering wedges*, *Mon. Not. Roy. Astron. Soc.* **464** (2017) 1640 [1607.03147].

[170] C.M. Correa, D.J. Paz, A.G. Sánchez, A.N. Ruiz, N.D. Padilla and R.E. Angulo, *Redshift-space effects in voids and their impact on cosmological tests. Part I: the void size function*, *Mon. Not. Roy. Astron. Soc.* **500** (2021) 911 [2007.12064].

[171] T. Ronconi, S. Contarini, F. Marulli, M. Baldi and L. Moscardini, *Cosmic voids uncovered - first-order statistics of depressions in the biased density field*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 5075 [1902.04585].

[172] S. Contarini, T. Ronconi, F. Marulli, L. Moscardini, A. Veropalumbo and M. Baldi, *Cosmological exploitation of the size function of cosmic voids identified in the distribution of biased tracers*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 3526 [1904.01022].

[173] S.R. Furlanetto and T. Piran, *The evidence of absence: galaxy voids in the excursion set formalism*, *Mon. Not. Roy. Astron. Soc.* **366** (2006) 467 [astro-ph/0509148].

[174] K.C. Chan, N. Hamaus and V. Desjacques, *Large-scale clustering of cosmic voids*, *Phys. Rev. D* **90** (2014) 103521 [1409.3849].

[175] M. Sahlén, Í. Zubeldía and J. Silk, *Cluster-Void Degeneracy Breaking: Dark Energy, Planck, and the Largest Cluster and Void*, *Astrophys. J. Lett.* **820** (2016) L7 [1511.04075].

[176] M. Sahlén and J. Silk, *Cluster-void degeneracy breaking: Modified gravity in the balance*, *Phys. Rev. D* **97** (2018) 103504 [1612.06595].

[177] E. Platen, R. Van De Weygaert and B.J.T. Jones, *Alignment of voids in the cosmic web*, *Monthly Notices of the Royal Astronomical Society* **387** (2008) 128 [http://oup.prod.sis.lan/mnras/article-pdf/387/1/128/3187216/mnras0387-0128.pdf

[178] G. Lavaux and B.D. Wandelt, *Precision Cosmography with Stacked Voids*, *Astrophys. J.* **754** (2012) 109 [1110.0345].

[179] T. Ronconi and F. Marulli, *Cosmological exploitation of cosmic void statistics. New numerical tools in the CosmoBolognaLib to extract cosmological constraints from the void size function*, *Astron. Astrophys.* **607** (2017) A24 [1703.07848].

[180] F. Marulli, A. Veropalumbo and M. Moresco, *CosmoBolognaLib: C++ libraries for cosmological calculations*, *Astronomy and Computing* **14** (2016) 35 [1511.00012].

[181] G. Pollina, N. Hamaus, K. Dolag, J. Weller, M. Baldi and L. Moscardini, *On the linearity of tracer bias around voids*, *Mon. Not. Roy. Astron. Soc.* **469** (2017) 787 [1610.06176].

[182] G. Pollina, N. Hamaus, K. Paech, K. Dolag, J. Weller, C. Sánchez et al., *On the relative bias of void tracers in the Dark Energy Survey*, *Mon. Not. Roy. Astron. Soc.* **487** (2019) 2836 [1806.06860].

[183] N. Tamura, N. Takato, A. Shimono, Y. Moritani, K. Yabe, Y. Ishizuka et al., *Prime Focus Spectrograph (PFS) for the Subaru telescope: overview, recent progress, and future perspectives*, in *Ground-based and Airborne Instrumentation for Astronomy VI*, C.J. Evans, L. Simard and H. Takami, eds., vol. 9908 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 99081M, Aug., 2016, DOI [1608.01075].

[184] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue et al., *Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report*, *arXiv e-prints* (2015) arXiv:1503.03757 [1503.03757].

[185] P. Melchior, P.M. Sutter, E.S. Sheldon, E. Krause and B.D. Wandelt, *First measurement of gravitational lensing by cosmic voids in SDSS*, *Mon. Not. Roy. Astron. Soc.* **440** (2014) 2922 [1309.2045].

[186] J. Clampitt and B. Jain, *Lensing measurements of the mass distribution in SDSS voids*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 3357 [1404.1834].

[187] T. Chantavat, U. Sawangwit, P.M. Sutter and B.D. Wandelt, *Cosmological parameter constraints from CMB lensing with cosmic voids*, *Phys. Rev. D* **93** (2016) 043523 [1409.3364].

[188] T. Chantavat, U. Sawangwit and B.D. Wandelt, *Void Profile from Planck Lensing Potential Map*, *Astrophys. J.* **836** (2017) 156 [1702.01009].

[189] Y.-C. Cai, M. Neyrinck, Q. Mao, J.A. Peacock, I. Szapudi and A.A. Berlind, *The lensing and temperature imprints of voids on the cosmic microwave background*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 3364 [1609.00301].

[190] C. Sánchez, J. Clampitt, A. Kovacs, B. Jain, J. García-Bellido, S. Nadathur et al., *Cosmic voids and void lensing in the Dark Energy Survey Science Verification data*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 746 [1605.03982].

[191] T. Baker, J. Clampitt, B. Jain and M. Trodden, *Void lensing as a test of gravity*, *Phys. Rev. D* **98** (2018) 023511 [1803.07533].

[192] M.M. Brouwer et al., *Studying galaxy troughs and ridges using Weak Gravitational Lensing with the Kilo-Degree Survey*, *Mon. Not. Roy. Astron. Soc.* **481** (2018) 5189 [1805.00562].

[193] Y. Fang, N. Hamaus, B. Jain, S. Pandey, G. Pollina, C. Sánchez et al., *Dark Energy Survey year 1 results: the relationship between mass and light around cosmic voids*, *Mon. Not. Roy. Astron. Soc.* **490** (2019) 3573 [1909.01386].

[194] P. Vielzeuf, A. Kovács, U. Demirbozan, P. Fosalba, E. Baxter, N. Hamaus et al., *Dark Energy Survey Year 1 results: the lensing imprint of cosmic voids on the cosmic microwave background*, *Mon. Not. Roy. Astron. Soc.* **500** (2021) 464 [1911.02951].

[195] B.R. Granett, M.C. Neyrinck and I. Szapudi, *An Imprint of Superstructures on the Microwave Background due to the Integrated Sachs-Wolfe Effect*, *Astrophys. J. Lett.* **683** (2008) L99 [0805.3695].

[196] P. Pápai and I. Szapudi, *Cosmological Density Fluctuations on 100 Mpc Scales and their ISW Effect*, *Astrophys. J.* **725** (2010) 2078 [1009.0754].

[197] Y.-C. Cai, S. Cole, A. Jenkins and C.S. Frenk, *Full-sky map of the ISW and Rees-Sciama effect from Gpc simulations*, *Mon. Not. Roy. Astron. Soc.* **407** (2010) 201 [1003.0974].

[198] S. Nadathur, S. Hotchkiss and S. Sarkar, *The integrated Sachs-Wolfe imprint of cosmic superstructures: a problem for ΛCDM*, *JCAP* **2012** (2012) 042 [1109.4126].

[199] S. Flender, S. Hotchkiss and S. Nadathur, *The stacked ISW signal of rare superstructures in ΛCDM*, *JCAP* **2013** (2013) 013 [1212.0776].

[200] S. Ilić, M. Langer and M. Douspis, *Detecting the integrated Sachs-Wolfe effect with stacked voids*, *Astron. Astrophys.* **556** (2013) A51 [1301.5849].

[201] Y.-C. Cai, M.C. Neyrinck, I. Szapudi, S. Cole and C.S. Frenk, *A Possible Cold Imprint of Voids on the Microwave Background Radiation*, *Astrophys. J.* **786** (2014) 110 [1301.6136].

[202] Y.-C. Cai, B. Li, S. Cole, C.S. Frenk and M. Neyrinck, *The integrated Sachs-Wolfe effect in f(R) gravity*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 2978 [1310.6986].

[203] S. Nadathur and R. Crittenden, *A Detection of the Integrated Sachs-Wolfe Imprint of Cosmic Superstructures Using a Matched-filter Approach*, *Astrophys. J. Lett.* **830** (2016) L19 [1608.08638].

[204] A. Kovács, C. Sánchez, J. García-Bellido, S. Nadathur, R. Crittenden, D. Gruen et al., *Imprint of DES superstructures on the cosmic microwave background*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 4166 [1610.00637].

[205] A. Kovács, *The part and the whole: voids, supervoids, and their ISW imprint*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 1777 [1701.08583].

[206] F. Dong, Y. Yu, J. Zhang, X. Yang and P. Zhang, *Measuring the integrated Sachs-Wolfe effect from the low-density regions of the universe*, *arXiv e-prints* (2020) arXiv:2006.14202 [2006.14202].

[207] Q. Hang, S. Alam, Y.-C. Cai and J.A. Peacock, *Stacked CMB lensing and ISW signals around superstructures in the DESI Legacy Survey*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 510 [2105.11936].

[208] A. Kovács, N. Jeffrey, M. Gatti, C. Chang, L. Whiteway, N. Hamaus et al., *The DES view of the Eridanus supervoid and the CMB cold spot*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 216 [2112.07699].

[209] P.M. Sutter, G. Lavaux, B.D. Wandelt and D.H. Weinberg, *A First Application of the Alcock-Paczynski Test to Stacked Cosmic Voids*, *Astrophys. J.* **761** (2012) 187 [1208.1058].

[210] D. Paz, M. Lares, L. Ceccarelli, N. Padilla and D.G. Lambas, *Clues on void evolution-II. Measuring density and velocity profiles on SDSS galaxy redshift space distortions*, *Mon. Not. Roy. Astron. Soc.* **436** (2013) 3480 [1306.5799].

[211] P.M. Sutter, A. Pisani, B.D. Wandelt and D.H. Weinberg, *A measurement of the Alcock-Paczyński effect using cosmic voids in the SDSS*, *Mon. Not. Roy. Astron. Soc.* **443** (2014) 2983 [1404.5618].

[212] N. Hamaus, P.M. Sutter, G. Lavaux and B.D. Wand elt, *Probing cosmology and gravity with redshift-space distortions around voids*, *JCAP* **2015** (2015) 036 [1507.04363].

[213] N. Hamaus, A. Pisani, P.M. Sutter, G. Lavaux, S. Escoffier, B.D. Wand elt et al., *Constraints on Cosmology and Gravity from the Dynamics of Voids*, *Phys. Rev. L* **117** (2016) 091302 [1602.01784].

[214] Y.-C. Cai, A. Taylor, J.A. Peacock and N. Padilla, *Redshift-space distortions around voids*, *Mon. Not. Roy. Astron. Soc.* **462** (2016) 2465 [1603.05184].

[215] I. Achitouv, C. Blake, P. Carter, J. Koda and F. Beutler, *Consistency of the growth rate in different environments with the 6-degree Field Galaxy Survey: Measurement of the void-galaxy and galaxy-galaxy correlation functions*, *Phys. Rev. D* **95** (2017) 083502 [1606.03092].

[216] C.-H. Chuang, F.-S. Kitaura, Y. Liang, A. Font-Ribera, C. Zhao, P. McDonald et al., *Linear redshift space distortions for cosmic voids based on galaxies in redshift space*, *Phys. Rev. D* **95** (2017) 063528 [1605.05352].

[217] A.J. Hawken, B.R. Granett, A. Iovino, L. Guzzo, J.A. Peacock, S. de la Torre et al., *The VIMOS Public Extragalactic Redshift Survey. Measuring the growth rate of structure around cosmic voids*, *Astron. Astrophys.* **607** (2017) A54 [1611.07046].

[218] N. Hamaus, M.-C. Cousinou, A. Pisani, M. Aubert, S. Escoffier and J. Weller, *Multipole analysis of redshift-space distortions around cosmic voids*, *JCAP* **2017** (2017) 014 [1705.05328].

[219] I. Achitouv, *New constraints on the linear growth rate using cosmic voids in the SDSS DR12 datasets*, *Phys. Rev. D* **100** (2019) 123513 [1903.05645].

[220] C.M. Correa, D.J. Paz, N.D. Padilla, A.N. Ruiz, R.E. Angulo and A.G. Sánchez, *Non-fiducial cosmological test from geometrical and dynamical distortions around voids*, *Mon. Not. Roy. Astron. Soc.* **485** (2019) 5761 [1811.12251].

[221] S. Nadathur, P.M. Carter, W.J. Percival, H.A. Winther and J.E. Bautista, *Beyond BAO: Improving cosmological constraints from BOSS data with measurement of the void-galaxy cross-correlation*, *Phys. Rev. D* **100** (2019) 023504 [1904.01030].

[222] A.J. Hawken, M. Aubert, A. Pisani, M.-C. Cousinou, S. Escoffier, S. Nadathur et al., *Constraints on the growth of structure around cosmic voids in eBOSS DR14*, *JCAP* **2020** (2020) 012 [1909.04394].

[223] M. Aubert, M.-C. Cousinou, S. Escoffier, A.J. Hawken, S. Nadathur, S. Alam et al., *The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Growth rate of structure measurement from cosmic voids*, *arXiv e-prints* (2020) arXiv:2007.09013 [2007.09013].

[224] P. Peebles, *The Large-scale Structure of the Universe*, Princeton Series in Physics, Princeton University Press (1980).

[225] P.M. Sutter, G. Lavaux, N. Hamaus, B.D. Wand elt, D.H. Weinberg and M.S. Warren, *Sparse sampling, galaxy bias, and voids*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 462 [1309.5087].

[226] R. Voivodic, H. Rubira and M. Lima, *The Halo Void (Dust) Model of large scale structure*, *JCAP* **2020** (2020) 033 [2003.06411].

[227] R.G. Miller, *The jackknife–a review*, *Biometrika* **61** (1974) 1.

[228] N. Hamaus, P.M. Sutter and B.D. Wandelt, *Universal Density Profile for Cosmic Voids*, *Phys. Rev. L* **112** (2014) 251302 [1403.5499].

[229] D. Park and J. Lee, *Void Ellipticity Distribution as a Probe of Cosmology*, *Phys. Rev. L* **98** (2007) 081301 [astro-ph/0610520].

[230] K.C. Chan, N. Hamaus and M. Biagetti, *Constraint of void bias on primordial non-Gaussianity*, *Phys. Rev. D* **99** (2019) 121304 [1812.04024].

[231] S.A. Gregory and L.A. Thompson, *The Coma/A1367 supercluster and its environs*, *Astrophys. J.* **222** (1978) 784.

[232] A.V. Tikhonov and I.D. Karachentsev, *Minivoids in the Local Volume*, *Astrophys. J.* **653** (2006) 969 [astro-ph/0609109].

[233] L.A. Thompson and S.A. Gregory, *An Historical View: The Discovery of Voids in the Galaxy Distribution*, *arXiv e-prints* (2011) arXiv:1109.1268 [1109.1268].

[234] I. Szapudi, A. Kovács, B.R. Granett, Z. Frei, J. Silk, W. Burgett et al., *Detection of a supervoid aligned with the cold spot of the cosmic microwave background*, *Mon. Not. Roy. Astron. Soc.* **450** (2015) 288 [1405.1566].

[235] I.B. Zeldovich, J. Einasto and S.F. Shandarin, *Giant voids in the Universe*, *Nature* **300** (1982) 407.

[236] J.R. Bond, L. Kofman and D. Pogosyan, *How filaments of galaxies are woven into the cosmic web*, *Nature* **380** (1996) 603 [astro-ph/9512141].

[237] J. Lee and D. Park, *Constraining the Dark Energy Equation of State with Cosmic Voids*, *Astrophys. J. Lett.* **696** (2009) L10 [0704.0881].

[238] R. Biswas, E. Alizadeh and B.D. Wandelt, *Voids as a precision probe of dark energy*, *Phys. Rev. D* **82** (2010) 023002 [1002.0014].

[239] B. Li and G. Efstathiou, *An extended excursion set approach to structure formation in chameleon models*, *Mon. Not. Roy. Astron. Soc.* **421** (2012) 1431 [1110.6440].

[240] J. Clampitt, Y.-C. Cai and B. Li, *Voids in modified gravity: excursion set predictions*, *Mon. Not. Roy. Astron. Soc.* **431** (2013) 749 [1212.2216].

[241] D. Spolyar, M. Sahlén and J. Silk, *Topology and Dark Energy: Testing Gravity in Voids*, *Phys. Rev. L* **111** (2013) 241103 [1304.5239].

[242] Y.-C. Cai, N. Padilla and B. Li, *Testing gravity using cosmic voids*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 1036 [1410.1510].

[243] G. Pollina, M. Baldi, F. Marulli and L. Moscardini, *Cosmic voids in coupled dark energy cosmologies: the impact of halo bias*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 3075 [1506.08831].

[244] P. Zivick, P.M. Sutter, B.D. Wandelt, B. Li and T.Y. Lam, *Using cosmic voids to distinguish f(R) gravity in future galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 4215 [1411.5694].

[245] I. Achitouv, *Testing the imprint of nonstandard cosmologies on void profiles using Monte Carlo random walks*, *Phys. Rev. D* **94** (2016) 103524 [1609.01284].

[246] B. Falck, K. Koyama, G.-B. Zhao and M. Cautun, *Using voids to unscreen modified gravity*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 3262 [1704.08942].

[247] E. Paillas, M. Cautun, B. Li, Y.-C. Cai, N. Padilla, J. Armijo et al., *The Santiago-Harvard-Edinburgh-Durham void comparison II: unveiling the Vainshtein screening using weak lensing*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 1149 [1810.02864].

[248] E.L.D. Perico, R. Voivodic, M. Lima and D.F. Mota, *Cosmic voids in modified gravity scenarios*, *arXiv e-prints* (2019) arXiv:1905.12450 [1905.12450].

[249] E. Massara, F. Villaescusa-Navarro, M. Viel and P.M. Sutter, *Voids in massive neutrino cosmologies*, *JCAP* **2015** (2015) 018 [1506.03088].

[250] A. Banerjee and N. Dalal, *Simulating nonlinear cosmological structure formation with massive neutrinos*, *JCAP* **2016** (2016) 015 [1606.06167].

[251] P.J.E. Peebles, *The Void Phenomenon*, *Astrophys. J.* **557** (2001) 495 [astro-ph/0101127].

[252] D.S. Reed, A. Schneider, R.E. Smith, D. Potter, J. Stadel and B. Moore, *The same with less: the cosmic web of warm versus cold dark matter dwarf galaxies*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 4413 [1410.1541].

[253] L.F. Yang, M.C. Neyrinck, M.A. Aragón-Calvo, B. Falck and J. Silk, *Warmth elevating the depths: shallower voids with warm dark matter*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 3606 [1411.5029].

[254] M. Baldi and F. Villaescusa-Navarro, *Cosmic Degeneracies II: Structure formation in joint simulations of Warm Dark Matter and $f(R)$ gravity*, *arXiv e-prints* (2016) arXiv:1608.08057 [1608.08057].

[255] E. Lester and K. Bolejko, *Imprints of decaying dark matter on cosmic voids*, *Phys. Rev. D* **104** (2021) 123540 [2111.11593].

[256] D.H. Jones, M.A. Read, W. Saunders, M. Colless, T. Jarrett, Q.A. Parker et al., *The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures*, *Mon. Not. Roy. Astron. Soc.* **399** (2009) 683 [0903.5451].

[257] L. Guzzo, M. Scodeggio, B. Garilli, B.R. Granett, A. Fritz, U. Abbas et al., *The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large-scale structure at $0.5 < z < 1.2$*, *Astron. Astrophys.* **566** (2014) A108 [1303.2623].

[258] K.S. Dawson, D.J. Schlegel, C.P. Ahn, S.F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *Astron. J.* **145** (2013) 10 [1208.0022].

[259] K.S. Dawson, J.-P. Kneib, W.J. Percival, S. Alam, F.D. Albareti, S.F. Anderson et al., *The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *Astron. J.* **151** (2016) 44 [1508.04473].

[260] M.R. Blanton, M.A. Bershady, B. Abolfathi, F.D. Albareti, C. Allende Prieto, A. Almeida et al., *Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe*, *Astron. J.* **154** (2017) 28 [1703.00052].

[261] Dark Energy Survey Collaboration, T. Abbott, F.B. Abdalla, J. Aleksić, S. Allam, A. Amara et al., *The Dark Energy Survey: more than dark energy - an overview*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1270 [1601.00329].

[262] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L.E. Allen et al., *The DESI Experiment Part I: Science,Targeting, and Survey Design*, *arXiv e-prints* (2016) arXiv:1611.00036 [1611.00036].

[263] O. Doré, M.W. Werner, M.L.N. Ashby, L.E. Bleem, J. Bock, J. Burt et al., *Science Impacts of the SPHEREx All-Sky Optical to Near-Infrared Spectral Survey II: Report of a Community Workshop on the Scientific Synergies Between the SPHEREx Survey and Other Astronomy Observatories*, *arXiv e-prints* (2018) arXiv:1805.05489 [1805.05489].

[264] Ž. Ivezić, S.M. Kahn, J.A. Tyson, B. Abel, E. Acosta, R. Allsman et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, *Astrophys. J.* **873** (2019) 111 [0805.2366].

[265] F.-S. Kitaura, C.-H. Chuang, Y. Liang, C. Zhao, C. Tao, S. Rodríguez-Torres et al., *Signatures of the Primordial Universe from Its Emptiness: Measurement of Baryon Acoustic Oscillations from Minima of the Density Field*, *Phys. Rev. L* **116** (2016) 171301 [1511.04405].

[266] Y. Liang, C. Zhao, C.-H. Chuang, F.-S. Kitaura and C. Tao, *Measuring baryon acoustic oscillations from the clustering of voids*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 4020 [1511.04391].

[267] K.C. Chan and N. Hamaus, *Volume statistics as a probe of large-scale structure*, *Phys. Rev. D* **103** (2021) 043502 [2010.13955].

[268] D. Forero-Sánchez, C. Zhao, C. Tao, C.-H. Chuang, F.-S. Kitaura, A. Variu et al., *Cosmic Void Baryon Acoustic Oscillation Measurement: Evaluation of Sensitivity to Selection Effects*, *arXiv e-prints* (2021) arXiv:2107.02950 [2107.02950].

[269] S. Nadathur, *Testing cosmology with a catalogue of voids in the BOSS galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **461** (2016) 358 [1602.04752].

[270] Q. Mao, A.A. Berlind, R.J. Scherrer, M.C. Neyrinck, R. Scoccimarro, J.L. Tinker et al., *A Cosmic Void Catalog of SDSS DR12 BOSS Galaxies*, *Astrophys. J.* **835** (2017) 161 [1602.02771].

[271] K.A. Douglass, D. Veyrat and S. BenZvi, *Updated void catalogs of the SDSS DR7 main sample*, *arXiv e-prints* (2022) arXiv:2202.01226 [2202.01226].

[272] D. Potter, J. Stadel and R. Teyssier, *PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys*, *Computational Astrophysics and Cosmology* **4** (2017) 2 [1609.08621].

[273] R. Laureijs, J. Amiaux, S. Arduini, J.L. Augueres, J. Brinchmann, R. Cole et al., *Euclid Definition Study Report*, *arXiv e-prints* (2011) arXiv:1110.3193 [1110.3193].

[274] M. Cropper, S. Pottinger, R. Azzollini, M. Szafraniec, S. Awan, Y. Mellier et al., *VIS: the visible imager for Euclid*, in *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*, M. Lystrup, H.A. MacEwen, G.G. Fazio, N. Batalha, N. Siegler and E.C. Tong, eds., vol. 10698 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 1069828, July, 2018, DOI.

[275] T. Maciaszek, *Euclid near infrared spectrometer and photometer instrument : mid phase D status*, in *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*, M. Lystrup, H.A. MacEwen, G.G. Fazio, N. Batalha, N. Siegler and E.C. Tong, eds., vol. 10698, International Society for Optics and Photonics, SPIE, 2018, DOI.

[276] M.R. Blanton, D.W. Hogg, N.A. Bahcall, J. Brinkmann, M. Britton, A.J. Connolly et al., *The galaxy luminosity function and luminosity density at redshiftz= 0.1*, *The Astrophysical Journal* **592** (2003) 819–838.

[277] M.R. Blanton, R.H. Lupton, D.J. Schlegel, M.A. Strauss, J. Brinkmann, M. Fukugita et al., *The Properties and Luminosity Function of Extremely Low Luminosity Galaxies*, *Astrophys. J.* **631** (2005) 208 [astro-ph/0410164].

[278] I. Zehavi, Z. Zheng, D.H. Weinberg, M.R. Blanton, N.A. Bahcall, A.A. Berlind et al., *Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity*, *Astrophys. J.* **736** (2011) 59 [1005.2413].

[279] A. Euclid Collaboration: Blanchard, S. Camera, C. Carbone, V.F. Cardone, S. Casas, S. Clesse et al., *Euclid preparation. VII. Forecast validation for Euclid cosmological probes*, *Astron. Astrophys.* **642** (2020) A191 [1910.09273].

[280] F. Leclercq, J. Jasche, P.M. Sutter, N. Hamaus and B. Wandelt, *Dark matter voids in the SDSS galaxy survey*, *JCAP* **2015** (2015) 047 [1410.0355].

[281] M.C. Cousinou, A. Pisani, A. Tilquin, N. Hamaus, A.J. Hawken and S. Escoffier, *Multivariate analysis of cosmic void characteristics*, *Astronomy and Computing* **27** (2019) 53 [1805.07181].

[282] C. Zhao, C. Tao, Y. Liang, F.-S. Kitaura and C.-H. Chuang, *DIVE in the cosmic web: voids with Delaunay triangulation from discrete matter tracer distributions*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 2670 [1511.04299].

[283] P. Monaco, *The Mass Function of Cosmic Structures with Nonspherical Collapse*, *Astrophys. J.* **447** (1995) 23 [astro-ph/9406029].

[284] E. Massara and R.K. Sheth, *Density and velocity profiles around cosmic voids*, *arXiv e-prints* (2018) arXiv:1811.03132 [1811.03132].

[285] J.B.T.M. Roerdink and A. Meijster, *The watershed transform: Definitions, algorithms and parallelization strategies*, *Fundam. Inform.* **41** (2000) 187.

[286] M.C. Neyrinck, M.A. Aragón-Calvo, D. Jeong and X. Wang, *A halo bias function measured deeply into voids without stochasticity*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 646 [1309.6641].

[287] F. Bernardeau, *The Nonlinear Evolution of Rare Events*, *Astrophys. J.* **427** (1994) 51 [astro-ph/9311066].

[288] S.D. Landy and A.S. Szalay, *Bias and variance of angular correlation functions*, *Astrophys. J.* **412** (1993) 64.

[289] F. Marulli, M. Bolzonella, E. Branchini, I. Davidzon, S. de la Torre, B.R. Granett et al., *The VIMOS Public Extragalactic Redshift Survey (VIPERS) . Luminosity and stellar mass dependence of galaxy clustering at 0.5 < z < 1.1*, *Astron. Astrophys.* **557** (2013) A17 [1303.2633].

[290] F. Marulli, A. Veropalumbo, M. Sereno, L. Moscardini, F. Pacaud, M. Pierre et al., *The XXL Survey. XVI. The clustering of X-ray selected galaxy clusters at z 0.3*, *Astron. Astrophys.* **620** (2018) A1 [1807.04760].

[291] A. Lewis, A. Challinor and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *Astrophys. J.* **538** (2000) 473 [astro-ph/9911177].

[292] R. Scoccimarro, *Redshift-space distortions, pairwise velocities, and nonlinearities*, *Phys. Rev. D* **70** (2004) 083007 [astro-ph/0407214].

[293] A. Taruya, T. Nishimichi and S. Saito, *Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory*, *Phys. Rev. D* **82** (2010) 063522 [1006.0699].

[294] F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A.J. Cuesta, D.J. Eisenstein et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 2242 [1607.03150].

[295] A. Pezzotta, S. de la Torre, J. Bel, B.R. Granett, L. Guzzo, J.A. Peacock et al., *The VIMOS Public Extragalactic Redshift Survey (VIPERS). The growth of structure at 0.5 < z < 1.2 from redshift-space distortions in the clustering of the PDR-2 final sample*, *Astron. Astrophys.* **604** (2017) A33 [1612.05645].

[296] I. Tutusaus, M. Martinelli, V.F. Cardone, S. Camera, S. Yahia-Cherif, S. Casas et al., *Euclid: The importance of galaxy clustering and weak lensing cross-correlations within the photometric Euclid survey*, *arXiv e-prints* (2020) arXiv:2005.00055 [2005.00055].

[297] L. Perotto, J. Lesgourgues, S. Hannestad, H. Tu and Y. Y Y Wong, *Probing cosmological parameters with the CMB: forecasts from Monte Carlo simulations*, *JCAP* **2006** (2006) 013 [astro-ph/0606227].

[298] X. Wang, X. Chen, Z. Zheng, F. Wu, P. Zhang and Y. Zhao, *Forecasting the dark energy measurement with baryon acoustic oscillations: prospects for the LAMOST surveys*, *Mon. Not. Roy. Astron. Soc.* **394** (2009) 1775 [0809.3002].

[299] O. Lahav, A. Kiakotou, F.B. Abdalla and C. Blake, *Forecasting neutrino masses from galaxy clustering in the Dark Energy Survey combined with the Planck measurements*, *Mon. Not. Roy. Astron. Soc.* **405** (2010) 168 [0910.4714].

[300] M. Martinelli, E. Calabrese, F. de Bernardis, A. Melchiorri, L. Pagano and R. Scaramella, *Constraining modified gravitational theories by weak lensing with Euclid*, *Phys. Rev. D* **83** (2011) 023012 [1010.5755].

[301] F. de Bernardis, M. Martinelli, A. Melchiorri, O. Mena and A. Cooray, *Future weak lensing constraints in a dark coupled universe*, *Phys. Rev. D* **84** (2011) 023504 [1104.0652].

[302] L. Wolz, M. Kilbinger, J. Weller and T. Giannantonio, *On the validity of cosmological Fisher matrix forecasts*, *JCAP* **2012** (2012) 009 [1205.3984].

[303] J. Hamann, S. Hannestad and Y.Y.Y. Wong, *Measuring neutrino masses with a future galaxy survey*, *JCAP* **2012** (2012) 052 [1209.1043].

[304] S. Khedekar and S. Majumdar, *Cosmology with the largest galaxy cluster surveys: going beyond Fisher matrix forecasts*, *JCAP* **2013** (2013) 030 [1210.5586].

[305] B. Audren, J. Lesgourgues, S. Bird, M.G. Haehnelt and M. Viel, *Neutrino masses and cosmological parameters from a Euclid-like survey: Markov Chain Monte Carlo forecasts including theoretical errors*, *JCAP* **2013** (2013) 026 [1210.2194].

[306] M.G. Kendall, A. Stuart and J.K. Ord, *Kendall's Advanced Theory of Statistics*, Oxford University Press, Inc., USA (1987).

[307] E. Sellentin, M. Quartin and L. Amendola, *Breaking the spell of Gaussianity: forecasting with higher order Fisher matrices*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 1831 [1401.6892].

[308] J. Froustey, C. Pitrou and M.C. Volpe, *Neutrino decoupling including flavour oscillations and primordial nucleosynthesis*, *JCAP* **2020** (2020) 015 [2008.01074].

[309] J.J. Bennett, G. Buldgen, P.F. de Salas, M. Drewes, S. Gariazzo, S. Pastor et al., *Towards a precision calculation of $N_{\rm eff}$ in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED*, *arXiv e-prints* (2020) arXiv:2012.02726 [2012.02726].

[310] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P.D. Serpico, *Relic neutrino decoupling including flavour oscillations*, *Nuclear Physics B* **729** (2005) 221 [hep-ph/0506164].

[311] M. Baldi, *The CoDECS project: a publicly available suite of cosmological N-body simulations for interacting dark energy models*, *Mon. Not. Roy. Astron. Soc.* **422** (2012) 1028 [1109.5695].

[312] E. Komatsu, K.M. Smith, J. Dunkley, C.L. Bennett, B. Gold, G. Hinshaw et al., *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, *Astrophys. J. Suppl. Ser.* **192** (2011) 18 [1001.4538].

[313] M. Raveri, M. Martinelli, G. Zhao and Y. Wang, *CosmicFish Implementation Notes V1.0*, *arXiv e-prints* (2016) arXiv:1606.06268 [1606.06268].