The GIGANTES Data Set: Precision Cosmology from Voids in the Machine-learning Era

Christina D. Kreisch, Alice Pisani, Francisco Villaescusa-Navarro, David N. Spergel, Benjamin D. Wandelt, Nico Hamaus, Adrian E. Bayer

HAL Id: insu-03839656
https://insu.hal.science/insu-03839656

Submitted on 4 Nov 2022

The GIGANTES Data Set: Precision Cosmology from Voids in the Machine-learning Era

Christina D. Kreisch[1] , Alice Pisani[1] , Francisco Villaescusa-Navarro[1] , David N. Spergel[1,2] , Benjamin D. Wandelt[2,3,4] ,
Nico Hamaus[5] , and Adrian E. Bayer[6,7]

[1] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544 USA; ckreisch@astro.princeton.edu, apisani@astro.princeton.edu
[2] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010 USA
[3] Institut d'Astrophysique de Paris, UMR 7095, CNRS, 98 bis Boulevard Arago, F-75014 Paris, France
[4] Sorbonne Universités, Institut Lagrange de Paris, 98 bis Boulevard Arago, F-75014 Paris, France
[5] Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians Universität, Scheinerstr. 1, D-81679 München, Germany
[6] Berkeley Center for Cosmological Physics, University of California, 341 Campbell Hall, Berkeley, CA 94720, USA
[7] Department of Physics, University of California, 366 LeConte Hall, Berkeley, CA 94720, USA
Received 2021 July 14; revised 2022 June 14; accepted 2022 June 26; published 2022 August 19

Abstract

We present GIGANTES, the most extensive and realistic void catalog suite ever released—containing over 1 billion cosmic voids covering a volume larger than the observable universe, more than 20 TB of data, and created by running the void finder VIDE on QUIJOTE's halo simulations. The GIGANTES suite, spanning thousands of cosmological models, opens up the study of voids, answering compelling questions: Do voids carry unique cosmological information? How is this information correlated with galaxy information? Leveraging the large number of voids in the GIGANTES suite, our Fisher constraints demonstrate voids contain additional information, critically tightening constraints on cosmological parameters. We use traditional void summary statistics (void size function, void density profile) and the void autocorrelation function, which independently yields an error of $0.13 \, \mathrm{eV}$ on $\sum m_\nu$ for a $1 \, h^{-3} \, \mathrm{Gpc}^3$ simulation, without cosmic microwave background priors. Combining halos and voids we forecast an error of $0.09 \, \mathrm{eV}$ from the same volume, representing a gain of 60% compared to halos alone. Extrapolating to next generation multi-Gpc^3 surveys such as the Dark Energy Spectroscopic Instrument, Euclid, the Spectro-Photometer for the History of the Universe and Ices Explorer, and the Roman Space Telescope, we expect voids should yield an independent determination of neutrino mass. Crucially, GIGANTES is the first void catalog suite expressly built for intensive machine-learning exploration. We illustrate this by training a neural network to perform likelihood-free inference on the void size function, giving a $\sim 20\%$ constraint on Ω_m. Cosmology problems provide an impetus to develop novel deep-learning techniques. With GIGANTES, machine learning gains an impressive data set, offering unique problems that will stimulate new techniques.

Unified Astronomy Thesaurus concepts: N-body simulations (1083); Cosmology (343); Large-scale structure of the universe (902); Voids (1779); Cosmic web (330); Computational astronomy (293); Cosmological parameters (339); Cosmological parameters from large-scale structure (340); Cosmological neutrinos (338); Astrostatistics (1882)

1. Introduction

Cosmic voids—the large underdense regions infilling the cosmic web (Gregory & Thompson 1978; Jõeveer et al. 1978; Peebles 1980; Kirshner et al. 1981; van de Weygaert & van Kampen 1993; Bond et al. 1996)—have gained interest in the last 10 years as a robust tool for extracting cosmological information (see Pisani et al. 2019, and references therein). Until a few decades ago survey volumes were too small to provide enough statistics for voids, as voids are among the largest objects in the universe. Large-scale surveys are now ushering in the era of big data in astronomy and astrophysics, which is essential for unveiling the power of cosmic voids. With such large amounts of data and access to previously unresolved phenomena, machine learning is becoming a prudent tool in astrophysics (see, e.g., Charnock et al. 2018; Baron 2019; Chalapathy & Chawla 2019; He et al. 2019; Ntampaka et al. 2019; Villaescusa-Navarro et al. 2020a, and references therein).

Many traditional techniques, such as multilayer perceptrons (MLPs) and convolutional neural networks (CNNs), are not interpretable, making discovering new physics a challenge, and can obscure information. The inherently graphical nature of the universe has prompted substantial recent development in interpretable deep-learning methods with graph neural networks (GNNs; Cranmer et al. 2020). At the same time, symmetries are embedded in physical laws that govern the evolution of the universe. Cosmic voids offer more symmetries that can improve interpretable machine-learning techniques. The data-rich field of astrophysics, with its large surveys, offers unique problems fundamentally characterized by symmetries, interpretability, and the need for accurate error estimates, providing an impetus for the development of new machine-learning techniques to conquer these questions.

With the advent of large modern surveys, we are entering the golden age of cosmic voids. The Baryon Oscillation Spectroscopic Survey (BOSS; Alam et al. 2017) and eBOSS (Dawson et al. 2016) have already provided several thousand voids (Aubert et al. 2022; Hamaus et al. 2020). This number will dramatically increase in the coming years thanks to data from the Dark Energy Spectroscopic Instrument (DESI) experiment (DESI Collaboration et al. 2016), the Subaru survey with the Prime Focus Spectrograph (PFS; Tamura et al. 2016), the

Spectro-Photometer for the History of the Universe and Ices Explorer (SPHEREx) mission (Doré et al. 2018), the Nancy Grace Roman Space telescope (Spergel et al. 2015), the Euclid mission (Laureijs et al. 2011), and the Legacy Survey of Space and Time (LSST) survey from the Vera Rubin Observatory (The LSST Dark Energy Science Collaboration et al. 2018), each expected to detect up to an order of 10^5 or more voids. Large-scale surveys provide enough volume to measure voids over large areas of the sky while simultaneously observing low-mass galaxies. Beyond increasing the statistical power of cosmic voids, we can now intensely study their properties, such as by mapping their interiors in detail.

Devoid of matter, comic voids are inherently extremely sensitive to the properties of diffuse components; they are a novel tool for constraining neutrino masses (Villaescusa-Navarro et al. 2013; Massara et al. 2015; Banerjee & Dalal 2016; Sahlén 2019; Kreisch et al. 2019; Schuster et al. 2019; Zhang et al. 2020; Bayer et al. 2021b), modified gravity (Clampitt et al. 2013; Barreira et al. 2015; Cai et al. 2015; Zivick et al. 2015; Achitouv 2016; Falck et al. 2018; Sahlén 2019; Achitouv 2019; Perico et al. 2019; Alam et al. 2021; Contarini et al. 2021), and dark energy (Bos et al. 2012; Spolyar et al. 2013; Pisani et al. 2015; Pollina et al. 2016; Verza et al. 2019). The current understanding of our universe sets the observed accelerated expansion of the universe as one of the most puzzling mysteries for modern cosmology. The nature of dark energy, the component constituting ∼68% of the universe and postulated to explain such accelerated expansion (Planck Collaboration et al. 2020), is still poorly understood. As matter is missing in voids by definition, voids are regions in the universe that become dark energy dominated the earliest, thus providing a window toward unraveling this dark mystery (Lee & Park 2009; Lavaux & Wandelt 2012). The spherical property of void stacks in real space provides standard spheres for the Alcock–Paczyński test (Alcock & Paczynski 1979; Ryden 1995; Ryden & Melott 1996; Biswas et al. 2010; Lavaux & Wandelt 2012). Using stacked voids, it is also possible to model the redshift-space distortion (RSD) pattern around voids. The use of the void-galaxy cross-correlation function for the Alcock-Paczyński test and RSDs is an area of intense data-analysis-based activity in recent years (Lavaux & Wandelt 2010). Voids have provided results competitive to those from galaxies, and as voids do not suffer from nonlinearities to the same extent as galaxies, there are indications that modeling could be easier for voids than for galaxies (e.g., Sutter et al. 2012; Paz et al. 2013; Sutter et al. 2015; Cai et al. 2016; Hamaus et al. 2016; Achitouv et al. 2017; Cai et al. 2017; Hamaus et al. 2017; Hawken et al. 2017; Correa et al. 2019; Nadathur et al. 2019; Aubert et al. 2022; Hamaus et al. 2020; Hawken et al. 2020; Nadathur et al. 2020; Paillas et al. 2021). Aside from voids' use in cosmology, the study of galaxy properties in voids is also an active field of research (Hoyle et al. 2005; Patiri et al. 2006; Kreckel et al. 2012; Ricciardelli et al. 2014; Habouzit et al. 2020; Panchal et al. 2020).

Despite the wide use of voids in galaxy clustering analyses of survey data, simulation-based analyses of void capabilities with an extremely large set of simulations have only been performed using dark matter with neutrino particles or dark matter particles alone as tracers to find voids, and with a spherical void finder (Bayer et al. 2021b). To provide synergy with data analysis, a thorough investigation of voids extracted from the halo field and with a void finder preserving void shape and cosmic-web information is imperative.

Relying on the large set of halo catalogs from the QUIJOTE simulations (Villaescusa-Navarro et al. 2020b) and the shape-preserving void finder VIDE (Sutter et al. 2015), we build the most extensive and realistic data set of void catalogs ever released: the GIGANTES[8] void catalogs suite, containing over 1 billion cosmic voids. The suite includes 15,000 VIDE fiducial cosmology void catalogs, as well as over 9,000 catalogs in nonfiducial cosmologies, spanning various values of the following cosmological parameters $\{\Omega_m, \Omega_b, h, n_s, \sigma_8, M_\nu, w\}$ and fully leveraging the QUIJOTE simulation suite, which covers redshifts $z = 0.0, 0.5, 1.0,$ and 2 in the real and redshift space. The void finding relies on the popular public Void IDentification and Examination (VIDE; Sutter et al. 2015) toolkit, based on the ZOBOV void finder (Neyrinck 2008). VIDE is arguably the most used void finder, as testified by its use in a plethora of papers performing both simulation-based theoretical modeling and data analysis from modern surveys (e.g., Chan et al. 2014; Hamaus et al. 2014a, 2014b; Pisani et al. 2014; Sutter et al. 2014; Pisani et al. 2015; Sutter et al. 2015; Pollina et al. 2017; Cousinou et al. 2019; Fang et al. 2019; Kreisch et al. 2019; Verza et al. 2019; Vielzeuf et al. 2021; Jeffrey et al. 2021). Figure 2 provides a visualization of VIDE voids from GIGANTES. We provide further details on the QUIJOTE simulations and the VIDE void finder in Appendix A.

Voids provide a set of summary statistics uniquely sensitive to the properties of the cosmological model. However, voids have been hotly debated as to whether or not they provide different information than the halo distribution they are found within. The main void summary statistics are the void–halo cross-correlation function ξ_{vh}, equivalent to the void density profile, the void size function n_v, a histogram of void sizes, and the void autocorrelation function ξ_{vv}. With our suite of over 1 billion voids, we show in Figure 1 that void summary statistics carry novel and complementary information to halos, evidenced by the different degeneracies in the cosmological parameters for a volume of 1 h^{-3} Gpc^3 (see Section 3 for details).

Due to its unprecedented size and exploration of the power of different void summary statistics, including well-known summary statistics such as ξ_{vh} and n_v, but also recent summary statistics such as ξ_{vv}, and its power in accurately estimating covariance matrices, the GIGANTES voids data set constitutes a powerful benchmark for machine-learning applications. Its massive number of voids coupled with a large number of realizations permits the extraction of a strong void signal, information that is usually obscured by noise due to a low number of voids. As such, the GIGANTES suite opens the realm of machine-learning exploration for voids, a new avenue for these objects. With such a wealth of data for voids, the GIGANTES suite provides a test bed not only for developing new cosmological theories, but also for developing novel machine-learning techniques. Given the nature of voids, they can be analyzed with traditional techniques, such as CNNs (when projected into 2D or 3D grids) albeit with the caveats mentioned earlier, or GNNs (when exploiting the full sparsity of the data). Their symmetries, need for interpretation, and need for uncertainty estimation, however, pose questions ideally suited for the development of new machine-learning techniques.

---

[8] A reference to the giants in *Don Quijote*'s world (Miguel de Cervantes 1605, 1615).
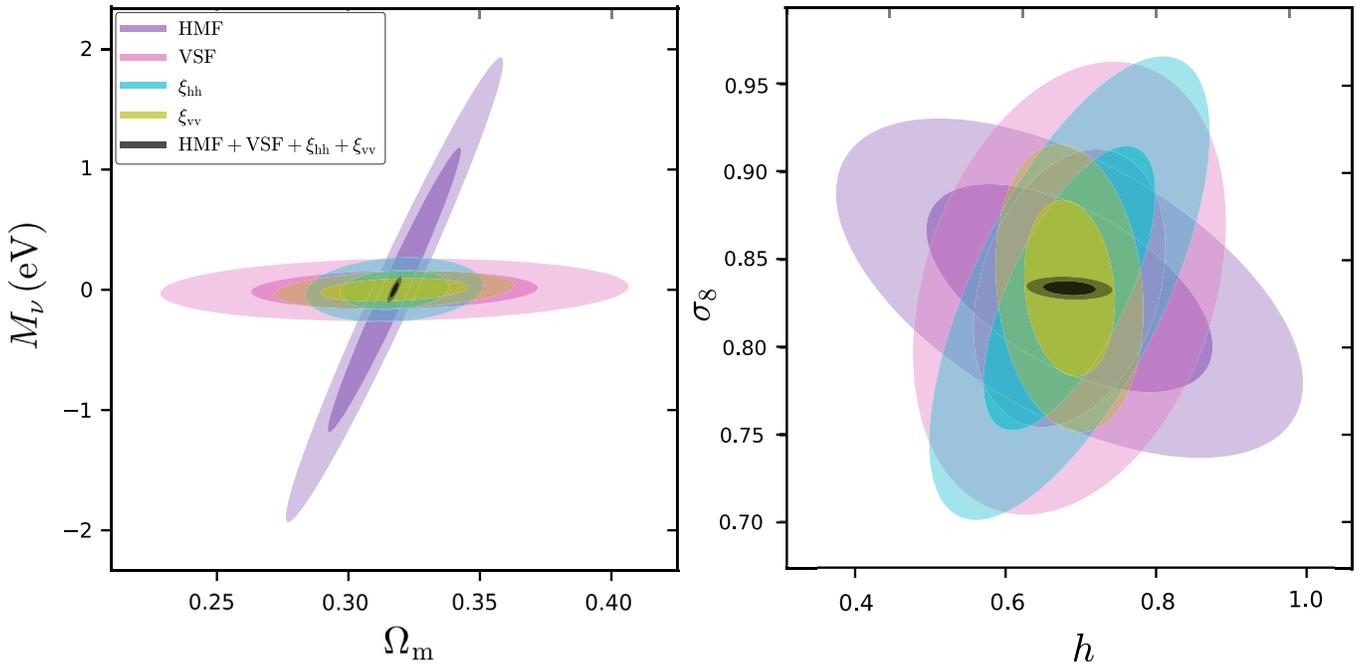
Figure 1. Constraints on the sum of neutrino masses M_ν, $\Omega_{\rm m}$, σ_8, and h from the halo mass function (HMF), void size function (VSF), halo autocorrelation function ($\xi_{\rm hh}$), void autocorrelation function ($\xi_{\rm vv}$), and combined statistics for a volume of $1\ h^{-3}\,{\rm Gpc}^3$. Void summary statistics provide strong constraints and orthogonal information to halo summary statistics.
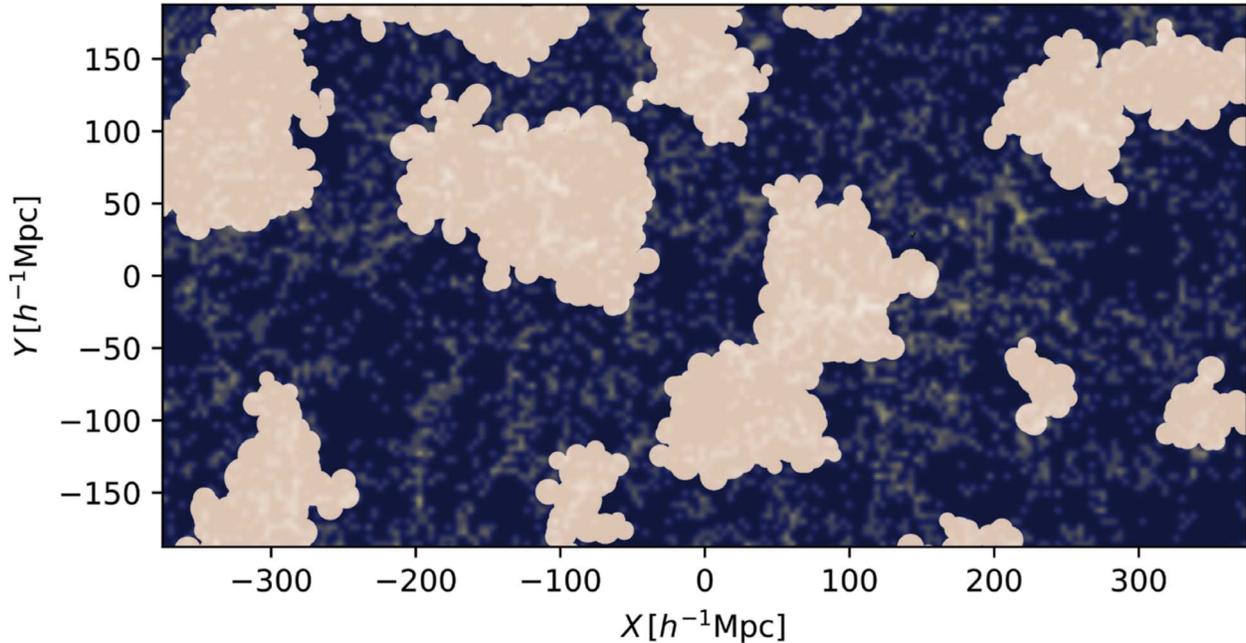
Figure 2. Several randomly selected voids are projected on the halo density field in orange, illustrating a range of sizes and shapes. The background shows the halo density field over a region of $375 \times 750 \times 250\ (h^{-1}{\rm Mpc})^3$. Void shapes are approximated by treating Voronoi cells as circles and projecting them on the halo density field. Despite the presence of projection effects, this 2D representation of voids in the slice illustrates the complex substructure of voids and the rich variety in their shapes.

This paper is organized as follows, and all results are presented for a volume of $1\ h^{-3}\,{\rm Gpc}^3$. In Section 2 we present the GIGANTES data set as well as the standard three void summary statistics. In Section 3 we address the question on whether voids carry additional information on cosmological parameters with a Fisher matrix forecast in real space at $z = 0$ exploring the constraining power of voids obtained from the halo field with

respect to traditional probes. We leave for future work an analysis in the redshift space. In Section 4 we study where the information comes from, with a computation of the correlation between and among void and halo summary statistics, a Fisher forecast in real space exploring the combined power of voids and halos, and a comparison of the power of void shape in constraints to a simplistic spherical void finder. This paper focuses on illuminating

Table 1
Summary of The GIGANTES void Catalogs

Cosmology	Redshifts	Realizations (real space)	Realizations (redshift space)	Total void #
Fiducial	[0.0, 0.5, 1.0, 2.0]	15,000 *(per z-bin)*	15,000 *(per z-bin)*	∼400,000,000
± for Fisher	[0.0, 0.5, 1.0, 2.0]	500 × 9 *(per z-bin & ± parameter)*	500 × 9 *(per z-bin & ± parameter)*	∼300,000,000
Latin Hypercube	[0.0, 0.5, 1.0, 2.0]	2000 *(per z-bin & resolution)*	2000 *(per z-bin & resolution)*	∼350,000,000

Note. The fiducial cosmology used in the simulations is: $\Omega_{\rm m} = 0.3175$, $\Omega_{\rm b} = 0.049$, $h = 0.6711$, $n_s = 0.9624$, $\sigma_8 = 0.834$, $M_\nu = 0.0$ eV, $w = -1$. Void catalogs for Fisher analyses are derived from simulations with a slight variation in each parameter.

the full information contained in these probes rather than exploring models for such probes. In Section 5 we show a machine-learning application by training a neural network to perform likelihood-free inference from the void size function. We draw the main conclusions of this work in Section 6.

2. GIGANTES

This section describes the void catalogs and the primary void summary statistics.

2.1. GIGANTES Catalogs

We build VIDE void catalogs from all the QUIJOTE simulations at different redshifts. A summary of all catalogs available is shown in Table 1. GIGANTES includes catalogs for the fiducial cosmology, catalogs varying single parameters for Fisher matrix calculations, and catalogs sampling multiple parameters at a time from a latin hypercube. We provide catalogs at different redshifts (0.0, 0.5, 1.0, and 2.0) as well as for the real space and redshift space. Overall, the GIGANTES suite provides more than 1 billion voids over thousands of different cosmological models. The GIGANTES data set will be publicly released upon acceptance of the paper.

Each void catalog provides a set of files as described in Sutter et al. (2015), giving for each void a set of properties, including:

1. the void center position (R.A., decl., z for observations; x, y, z for simulations),
2. the void radius (see Appendix A),
3. the void volume,
4. the void density contrast (as defined by ZOBOV and thus used by VIDE, given by the ratio of the minimum density along the ridge of the void versus the minimum density in the void),
5. the number of particles of each void (halos in our case),
6. the central density (density in a sphere of $0.25R_v$; see Sutter et al. 2015),
7. the void ellipticity,
8. and the void hierarchy.

As VIDE voids are found taking into account the hierarchical feature of the cosmic web, the void finder outputs all voids (parents) and their subvoids (children): for each void, the catalog includes the ID of the void parent (if applicable), the tree level, and the number of children.

2.2. Main Void Summary Statistics

The void finder provides void centers and radii, which in themselves yield two different summary statistics to be used in cosmological data analyses: the void autocorrelation function,

and the void size function. Combining void positions with halo positions defines a third summary statistic: the void–halo cross-correlation function.

#### 2.2.1. The Void–Halo Cross-correlation Function ξ_{vh}

By knowing the position of void centers and halos, it is possible to compute the probability of finding a halo at a certain distance from the void center, known as the void–halo cross-correlation function, $\xi_{\rm vh}$, or the void density profile found in the halo field. This probability will be low in the center of the void (it is not very likely to find halos, or galaxies, close to the void center), and high at the overdense rim enclosing the void, which is coincident with the average void radius. It is usually computed considering a stack of many voids. While the estimation of the void–halo cross-correlation function from data usually relies on estimators, it can be shown that $\xi_{\rm vh}$ is equivalent to the void density profile (see Equations (2.7) and (2.8) of Hamaus et al. 2015):

$$\xi_{\rm vh}(r) = \frac{\rho_{\rm vh}(r) - \bar{\rho}_{\rm h}}{\bar{\rho}_{\rm h}}, \qquad (1)$$

where r is the distance between the halos and the void center; $\rho_{\rm vh}$ is density of halos in shells around the void center; and $\bar{\rho}_{\rm h}$ is the average tracer density of the data set.

The void density profile is sensitive to cosmological parameters in both the real space and redshift space. The peak height and location of the 1D real space profile, denoting the void wall's rim and average void radius, responds to changes in cosmological parameters. In a homogeneous and isotropic universe, void density profiles do not possess a preferred direction. Thus, in real space, the 2D density profile will be spherically symmetric. In the redshift space, however, the spherical stacked 2D void profile will be distorted due to the Alcock–Paczyński effect, which introduces cosmology-dependent geometrical distortions along the line of sight (Alcock & Paczynski 1979). The 2D redshift space profile is also distorted due to RSDs. The magnitude of these effects depend on the cosmological model, so the observed 2D void density profile can be used to constrain cosmology.

The void density profile has been used to successfully measure the Alcock–Paczyński effect by using voids as standard spheres (e.g., Sutter et al. 2012) and simultaneously modeling RSDs (e.g., Hamaus et al. 2016, 2020). The profile provides the most stringent constraint on the universe's expansion history at $z = 0.5$ to date from voids: a calibration-free measurement from voids only, with a 16.8% precision level measurement of the growth rate of structure and an independent measurement of the matter content of the universe yielding $\Omega_{\rm m} = 0.312 \pm 0.020$ (Hamaus et al. 2020).
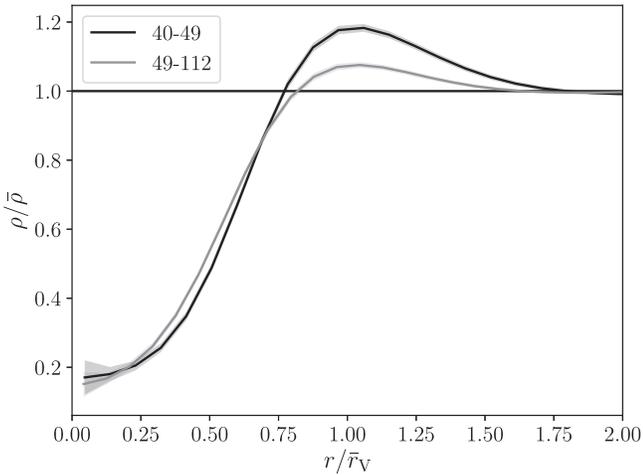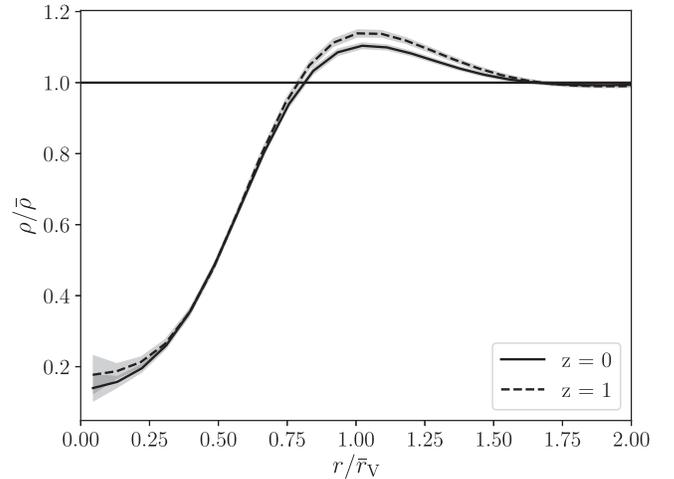
Figure 3. Void density profiles for two different radii bins in h^{-1} Mpc at $z = 0$. The void wall rim occurs at the peak of the density profile, which is also coincident with the average void radius.

Figure 4. Void density profiles for $z = 0$ and $z = 1$ with void radii spanning 50–70 h^{-1} Mpc for each.

Figure 3 shows two density profiles measured from the GIGANTES void catalogs for voids with radii 40–49 h^{-1} Mpc and 49–112 h^{-1} Mpc at $z = 0$. The distance is normalized by the average void radius for the denoted radius bin. Larger voids have a less pronounced void wall rim than smaller voids due to the fact that larger voids have evolved more than smaller voids at a fixed redshift and because we find voids in the halo field rather than in the dark matter field. Figure 4 shows profiles from two different redshift bins ($z = 0$ and $z = 1$) for voids 50–70 h^{-1} Mpc in radius.[9] For the same radius bin, voids at present are more evolved than voids at $z = 1$ as they are, on average, older. Thus, voids at $z = 0$ are more evolved and so have a lower void wall rim than voids at $z = 1$.

2.2.2. The Void Size Function

The void size function measures the number of voids as a function of their radius, also known as the void abundance, and is particularly sensitive to cosmological parameters, such as the dark energy equation of state (Pisani et al. 2015; Verza et al. 2019), modified gravity (Contarini et al. 2021), and massive neutrinos (Sahlén 2019; Kreisch et al. 2019). The theoretical modeling of the void size function is an area of intense activity (e.g., Sheth & van de Weygaert 2004; Jennings et al. 2013; Pisani et al. 2015; Sahlén & Silk 2018; Contarini et al. 2019; Verza et al. 2019) and will provide stringent constraints when applied to the next generation of surveys (e.g., Pisani et al. 2015; Doré et al. 2018). The main theoretical model to be considered for the abundance is the Sheth and van de Weygaert model (Sheth & van de Weygaert 2004), an excursion-set-based prediction of void numbers of different sizes, later extended to consider the volume conservation of voids, as small voids merge to form larger voids (Jennings et al. 2013). This model has been shown to reproduce simulation results in different cosmologies where either the dark energy equation of state (Pisani et al. 2015; Verza et al. 2019) is varied or modified gravity is considered (Contarini et al. 2021). As the model assumes dark matter voids and spherical symmetry, current work focuses on improving the match of predictions with
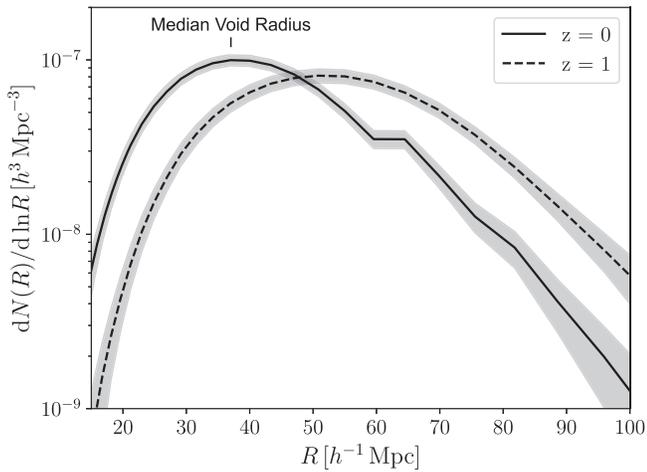
measurements of observed voids in mocks (e.g., by including tracer bias; see Pollina et al. 2019; Contarini et al. 2019, 2021).

In this work we do not attempt to model the measured void abundance with theoretical models, but we instead choose the approach of relying on the measured values of void numbers from the GIGANTES data set to assess its constraining power for cosmology. In Figure 1 we show a Fisher forecast over a volume of 1 h^{-3} Gpc^3 for constraints on a few cosmological parameters from the void size function for radii 6–60 h^{-1} Mpc. The void size function puts superior constraints on $\sum m_\nu$, for example, compared to the halo mass function (errors of 0.21 eV and 1.56 eV, respectively), and is competitive with the halo autocorrelation function (error of 0.22 eV). See Appendices B and C for forecasted constraints on additional cosmological parameters. In Appendix B we also illustrate the stability of the Fisher uncertainties by varying the number of simulations used to calculate the uncertainties and illustrating that the change in the value of the uncertainties is less than the change in the number of simulations.

Figure 5 shows the abundance of voids from the GIGANTES data set at $z = 0$ and at $z = 1$. As described below in Section 2.2.3, voids are larger in the latter case. We show the void size function for radii ranging from 15 to 100 h^{-1} Mpc. The median void radius is 35 h^{-1} Mpc for $z = 0$. Voids of lower size are less abundant due to the low simulation resolution at small scales (the mean tracer separation is ∼13 h^{-1} Mpc at $z = 0$ and ∼17 h^{-1} Mpc at $z = 1$). Nevertheless, small voids improve cosmological constraints, pointing to the constraining power of the void hierarchy, which will be further explored in future work. The large volume and, thus, strong void statistics in the GIGANTES suite promote a high signal-to-noise ratio for the void size function's response to cosmological parameters.

2.2.3. The Void Autocorrelation Function ξ_{vv}

We can compute the probability of finding a void center at a given distance from a randomly selected void in our void catalog, yielding the void autocorrelation function. As the center of a void is defined as the volume-weighted barycenter of the Voronoi cells composing the void, it carries information about the whole structure of the void. This allows us to uncover a new sensitivity of the large-scale structure of the universe to cosmological parameters (see Sections 3, 4, and Section 5). The
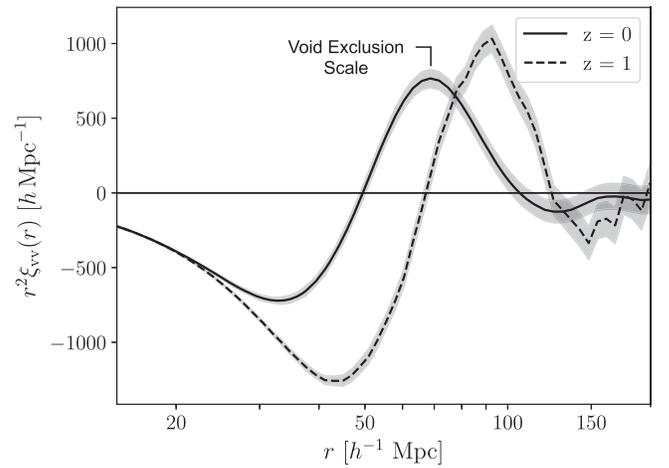
---

[9] Errors are propagated for the average correlation function based on the individual errors estimated by VIDE, which are based on scatter in the bin average.

Figure 5. Void size function, or void abundance, from the GIGANTES data set for both $z = 0$ and $z = 1$. The peak of the abundance corresponds to the median void radius.

Figure 6. Void autocorrelation function ξ_{vv} from GIGANTES voids, for both $z = 0$ and $z = 1$. The void exclusion scale is coincident with the peak of the void autocorrelation function at $\sim 2\tilde{R}_V$, where $\tilde{R}_V$ is the median void radius.

void autocorrelation function requires a great number of voids to overcome shot noise, and it has so far been measured from data only at low redshift, over a volume of $0.6 \, (h^{-1} \, \text{Gpc})^3$ (Clampitt et al. 2016), and considering relatively small effective void radii ($<30 \, h^{-1} \, \text{Mpc}$).

Given its limited application to data, the void autocorrelation function has not yet been used to derive constraints on parameters of the cosmological model. It has been investigated in simulations (Chan et al. 2014; Hamaus et al. 2014) and shows a strong sensitivity to massive neutrinos (Kreisch et al. 2019). We further illustrate this sensitivity to neutrinos for a volume of $1 \, h^{-3} \, \text{Gpc}^3$ in Figure 1, in which the void autocorrelation function puts the strongest forecasted constraints on $\sum m_\nu$ of 0.13 eV. One reason for ξ_{vv}'s strong constraints on the parameters could be that it contains a strong, clear feature at the void exclusion scale. Such strong features are useful for measuring distances, like baryon acoustic oscillations, probing the background cosmology (see also Hamaus et al. 2014). Further, void bias can be large (see, e.g., Chan et al. 2014; Schuster et al. 2019), and this boosts the large-scale correlation function amplitude, which can provide a better signal-to-noise ratio to measure the large-scale power spectrum, which depends on σ_8 and n_s. See Appendices B and C for further forecasted constraints on cosmological parameters.

With the advent of upcoming surveys, such as the DESI experiment (DESI Collaboration et al. 2016), the Roman Space Telescope (Spergel et al. 2015), the Euclid ESA satellite (Laureijs et al. 2011), the SPHEREx mission (Doré et al. 2018), the Rubin Observatory (The LSST Dark Energy Science Collaboration et al. 2018), and the PFS Subaru survey (Tamura et al. 2016), hundreds of thousands of voids will be observed (Pisani et al. 2019). The void autocorrelation function will then reach the realm of observable quantities that can be measured from data with high significance, and used to constrain cosmological parameters. Additionally, in the framework of exploiting statistics such as the void autocorrelation function and the void size function, it is important to consider that void finding may suffer from selection effects; in particular in imaging surveys these effects are expected to have a stronger impact due to photometric redshift errors distorting the line-of-sight direction (Sánchez et al. 2017). The GIGANTES set of void catalogs could serve as a mock data set to characterize and quantify such effects.

We show the void autocorrelation function measured from the GIGANTES data set in Figure 6 at $z = 0$ and $z = 1$. The void autocorrelation function has a maximum at twice the mean void radius of the sample. This corresponds to the void exclusion scale (Hamaus et al. 2014a), which is termed as such because, on average, void walls touch at twice the mean void radius as voids cannot overlap. This makes it likely to have two void centers at a distance corresponding to twice the mean void radius of the sample. Voids on average cannot be separated by smaller scales because then they will overlap. However, there is a distribution of void sizes (i.e., the void size function), and smaller voids are allowed to be separated by equivalently smaller scales. Thus, the probability of two voids being separated smoothly decreases as scales become smaller, as there are fewer small voids.

At higher redshift, the observed average radius of GIGANTES voids is larger. The universe is more dominated by dark energy at $z = 0$ than it was at $z = 1$. Large, isolated voids in the dark matter field mostly expand in their lifetime. Large voids in the dark matter field at $z = 1$ most likely continue to expand through $z = 0$. Here, however, we consider voids found in the halo field. The number density of halos of a lower fixed mass of, e.g., $10^{13} \, M_\odot$, is higher at present. This high number density of small halos results in, on average, more small voids at present than at $z = 1$. This is also why the peak of the void density profile at $z = 0$ is shallower than at $z = 1$, whereas in dark matter simulations the opposite occurs (Hamaus et al. 2014b). Thus, the average void size is larger at $z = 1$ in the GIGANTES void catalogs suite. We also note that at $z = 1$ there are less voids, so the ξ_{vv} is noisier. The theoretical modeling of the void autocorrelation function is an area that will deserve further attention in coming years, given its promising constraining power (see Section 3).

3. Do Voids Carry Additional Information? Yes

The question of whether voids, defined by relying on the halo (or galaxy) distribution, carry additional cosmological information with respect to traditional tools based on the halo (or galaxy) distribution, such as the two-point correlation function or the halo mass function, has been a debated topic.

From a theoretical perspective, there have been indications to believe voids contain higher-order information. As extended 3D objects, voids must be defined by at least four nonplanar halos. The fact that voids are defined by many halos could hint

to the possibility of accessing higher-order information, but this is not in principle guaranteed: it is possible to find extended 3D objects, and thus voids, in a Gaussian random field (or a Poisson distribution of tracers; see Cousinou et al. 2019). Nevertheless, the universe at present is not a Gaussian random field, and certain summary statistics associated to voids, like the void size function, have been expected to contain information beyond the two-point correlation function. As one-point functions, the amplitude and shape are expected to depend on all n-point functions. The probability of a random volume being a void can be written in terms of n-point correlations (Fry 1986; Fry & Colombi 2013):

$$P_0 = \exp\left[\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \bar{N}^n \xi_n\right], \tag{2}$$

where $\bar{N}$ is the mean tracer (i.e., halo or galaxy) count. Further, we are using voids found in the halo field, which already probes higher k. Another possible explanation for why voids can capture nonlinear information is that an individual void is built from halos separated by distances approximately less than or equal to its diameter, highlighting scales that are quasi-linear and approaching nonlinear.

This paper has the tools necessary to properly investigate this question, relying on a comprehensive Fisher analysis. The Fisher Matrix is given by

$$F_{ij} = \sum_{\alpha,\beta} \frac{\partial S_\alpha}{\partial \theta_i} C_{\alpha\beta}^{-1} \frac{\partial S_\beta}{\partial \theta_j}, \tag{3}$$

where S_i is the element i of the statistic $\vec{S}$, and C represents the covariance matrix

$$C_{\alpha\beta} = \langle (S_\alpha - \bar{S}_\alpha)(S_\beta - \bar{S}_\beta) \rangle; \qquad \bar{S}_i = \langle S_i \rangle. \tag{4}$$

The error on the parameter θ_i (marginalized over the other parameters), is:

$$\delta\theta_i \geqslant \sqrt{(F^{-1})_{ii}}. \tag{5}$$

The derivatives and covariance matrices are computed using the simulations and GIGANTES void catalogs (see Section 4.1 in Villaescusa-Navarro et al. 2020b for more detailed information on how the Fisher analysis is performed).

We investigate the power of the void statistics mentioned in Section 2.2 with respect to two traditional statistics defined from the halo field: the halo mass function and the halo autocorrelation function. The halo mass function measures the abundance of halos of a given mass, while the halo autocorrelation function gives the probability that, taken a random halo, there is another halo at a distance r from the first halo.

Figure 1 shows constraints from different void summary statistics, as well as for the halo mass function and the halo autocorrelation function, in the real space over a volume of $1\,h^{-3}\,\mathrm{Gpc}^3$ for selected parameters: the sum of neutrino masses $\sum m_\nu$, the matter content of the universe Ω_m, the reduced Hubble constant h, and the amplitude of linear density fluctuations inside spheres of radius $8\,h^{-1}\,\mathrm{Mpc}\,\sigma_8$. Bayer et al. (2021b) analyzed the power of voids found in the dark matter distribution, but here we use the realistic set of GIGANTES void catalogs drawn from the halo distribution to show that voids observed by galaxy surveys hold the power to

strongly reduce the available parameter space for cosmological parameters.

From this figure we clearly see that the void size function constraints have a very different orientation with respect to constraints from the halo mass function and the halo autocorrelation function. The void autocorrelation function provides even stronger constraining power as well as access to unique information, as evidenced from its orientation with respect to the other three summary statistics. The void size function, in particular, is almost fully orthogonal to the halo mass function, illustrating it provides access to information beyond that accessible from the halo statistics. Further, the orientation of the void and halo autocorrelations differ, and for some parameter planes, such as the $\sigma_8 - h$ plane, the halo mass function, void size function, halo autocorrelation function, and void autocorrelation function are almost all in different orientations. We forecast a gain in $\sum m_\nu$ for a $1\,h^{-3}\,\mathrm{Gpc}^3$ simulation of 60% combining halos and voids compared to the halo autocorrelation function and of 94% compared to the halo mass function.

The combination of all the mentioned summary statistics considerably tightens final constraints, demonstrating that voids carry complementary information to that of halos. See Appendix C for the constraining power for other parameters.

## 4. Where The Information Comes From

Aside from showing that voids provide additional information, the large GIGANTES data set allows us to estimate how correlated the different summary statistics are and where the information on cosmological parameters comes from. The GIGANTES data set contains void catalogs similar to those expected from future surveys such as PFS, as they have a similar number density to the QUIJOTE simulations, although different halo biases. Upcoming surveys such as Euclid and the Roman Space Telescope will have a higher number density than the QUIJOTE simulations, and so are expected to provide even better constraints from voids than what is illustrated here. See Appendix A for details on comparing the QUIJOTE simulations to upcoming surveys.

4.1. Summary Statistic Correlation

An important point to address is how the different void statistics are correlated, and if information from different bins is correlated. In Figure 7 we show the correlation matrix between and within the halo mass function (first 15 bins), the void size function (following 18 bins), the void autocorrelation function (following 61 bins), the halo autocorrelation function (following 61 bins), and the void–halo cross-correlation function (last 61 bins). We utilize the 15,000 fiducial simulations at $z = 0$ and compute the standard Pearson correlation coefficient for the different summary statistics:

$$\rho(x, y) = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}. \tag{6}$$

The correlations are computed for a volume of $1\,h^{-3}\,\mathrm{Gpc}^3$ for scales utilized in the Fisher forecast, namely, 15–$200\,h^{-1}\mathrm{Mpc}$ in log space for the correlation functions and radii 6–$60\,h^{-1}\mathrm{Mpc}$ for the void size function. Bins for the HMF are used from Bayer et al. (2021b). This, however, does not make a statement on whether or not the parameter constraints are independent, as 0 correlation does not imply parameter independence. In Appendix B we also
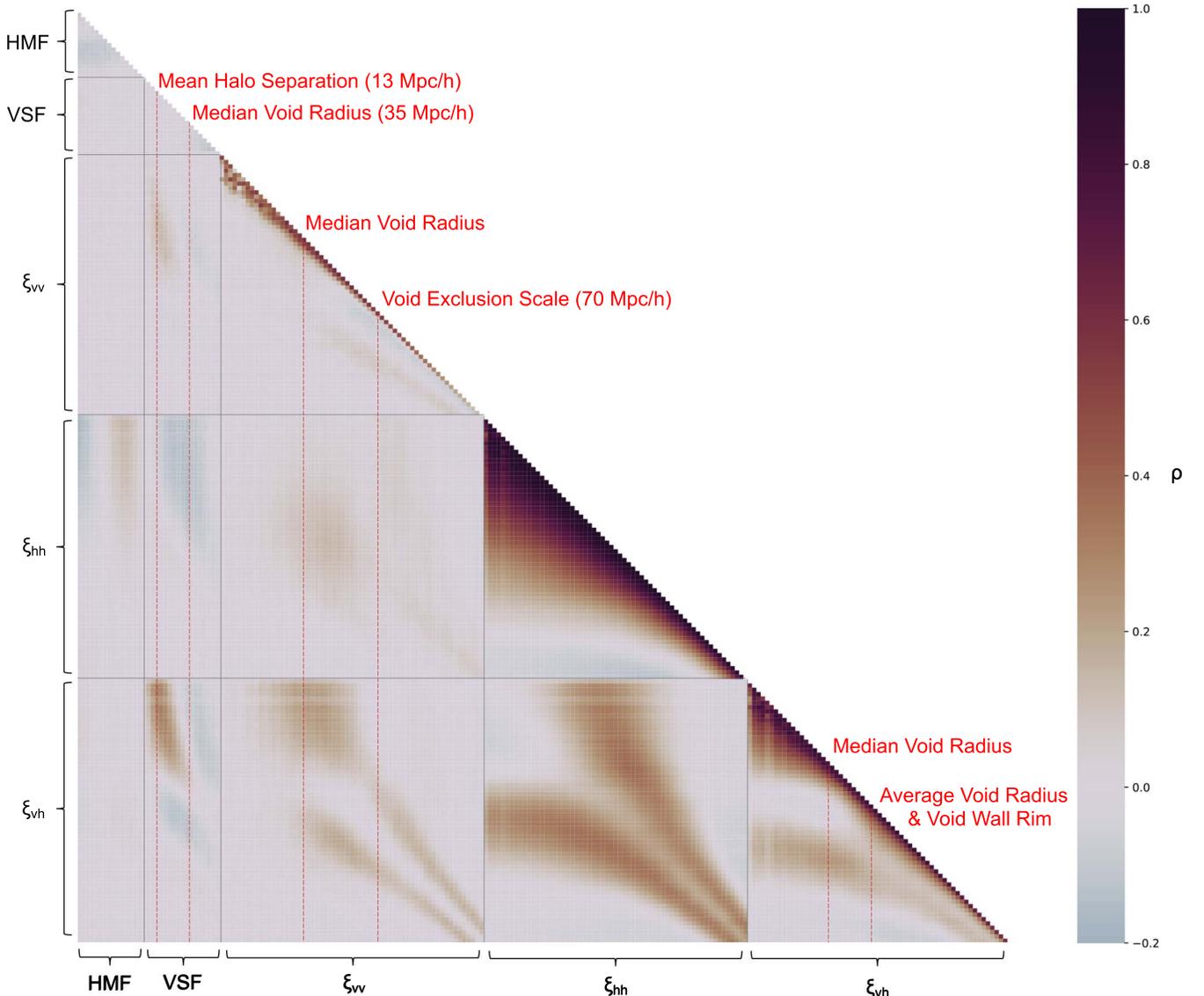
Figure 7. Correlation for the halo mass function (first 15 bins), the void size function (following 18 bins), the void autocorrelation function (following 61 bins), the halo autocorrelation function (following 61 bins), and the void–halo cross-correlation function (last 61 bins). All the correlation functions span from 15 to 200 h^{-1} Mpc in log space. The VSF goes from 6 to 60 h^{-1} Mpc in increments of 3 h^{-1} Mpc.

illustrate the stability of the covariance by varying the number of simulations used to calculate the covariance and illustrating that the change in the value of the uncertainties is less than the change in the number of simulations.

The void size function and the halo mass function each have relatively uncorrelated bins. A few bins at small distances (below the *median* radius of 35 h^{-1} Mpc) in the void autocorrelation function are correlated, indicating some coupling at those scales, while larger scales are decoupled. The halo autocorrelation function is more strongly correlated over all scales, while the void–halo correlation function shows a slightly higher correlation at low radii (that is, below the average void radius) than the void autocorrelation function. We note that scales roughly below 20 h^{-1} Mpc in the correlation matrix show some scatter. This may be physical due to `QUIJOTE` being well resolved below such scales and the covariance being stable (see Appendix B), but in the interest of transparency, we compute the Fisher uncertainties for all ξ for only scales 20–200 h^{-1} Mpc in Appendix B and Table 14.

Most changes in error are near 10%, although there are a few that exceed this. Fisher forecasts are idealized, however, and so should be trusted up to the 10% level as a rule of thumb.

The void size function and the halo mass function are uncorrelated with each other, as highlighted by the perpendicular Fisher contours already, confirming that the two probes are strongly complementary. By inspecting the correlation between the void size function and void autocorrelation function, as well as the void size function and the void–halo cross correlation, we see that large voids tend to be slightly anticorrelated at small scales, while small voids tend to be correlated at small scales. We can also see that large voids tend to be anticorrelated with halos that are close to each other, as expected. Viewing the correlation between the void autocorrelation function and the void–halo cross-correlation function, we see that at scales larger than the void exclusion scale and larger than the void wall rim, respectively, the two functions are somewhat correlated. This is because each follows the matter field at such large scales. We also see a parallel increase in
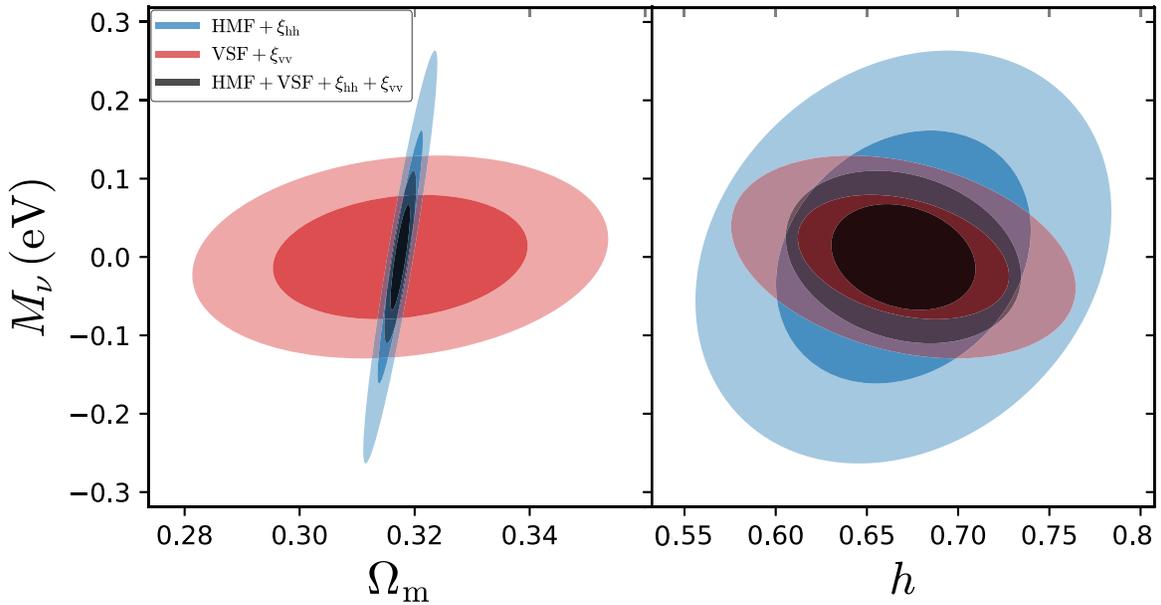
Figure 8. Comparison of halo-only with void-only constraints (therefore not including the void–halo cross correlation).

correlation for scales larger than the median void radius and larger than the void wall rim, due to the fact that voids are more likely to be clustered at these scales. Without the sheer volume of the GIGANTES suite, the correlation among and within these void summary statistics could not be understood in such detail.

4.2. Optimal Constraints Combining Voids and Halos

The previous section showed that voids provide different, additional information than other probes. Given that voids are obtained *from* the halo or galaxy distribution, we want to compare the total information from all void summary statistics (void size function and ξ_{vv}) with the total information from halo summary statistics (halo mass function and ξ_{hh})—without considering the combined statistics ξ_{vh} for this comparison. Figure 8 shows the comparison for a volume of $1 \ h^{-3} \, \mathrm{Gpc}^3$: for some parameters, voids in fact perform better than halos, and overall voids are orthogonal in their constraints to halos. Voids are particularly effective in constraining massive neutrinos and h. Constraints on neutrinos are expected to be powerful from voids (Massara et al. 2015; Kreisch et al. 2019), as diffuse components particularly impact voids. See Appendix C for further parameter comparisons.

In all cases, as highlighted above, combined constraints are powerful. We recall that for this comparison we have not considered the void–halo cross correlation, in order to compare halo-only with void-only summary statistics. If adding the void–halo cross correlation, constraining power is further increased (see Appendix C). We leave for future work the exploration of the constraining power on dark energy, expected to be large from voids (Bos et al. 2012; Pisani et al. 2015, 2019; Verza et al. 2019). Finally, we note that in this paper we do not investigate the impact of varying the density-based threshold, defining the hierarchy, because the QUIJOTE simulations have a relatively high minimum mass. To properly investigate the effect of varying this parameter on the hierarchy of voids one would need a simulation set that reaches very low masses of halos, allowing to properly detect the nested hierarchy of subvoids. In the current set of simulations, less than ~5% of

the voids are subvoids. We leave for future work the exploration of the constraining power of the void hierarchy.

4.3. Void Shape Adds Information

Void definition has been an intense area of discussion in the past (Colberg et al. 2008; van de Weygaert & Schaap 2009; Cautun et al. 2018). Recently, to exploit cosmological information, most works have chosen the approach of testing a particular definition and pushing its use in measuring cosmological signals. This approach has been particularly successful, as the point is not finding the best definition, but merely understanding the sensitivity of a particular definition to the signal to measure. In the realm of maximizing the signal and being able to model the signal, which can be more challenging for some void definitions than others, fully understanding the impact of definition choices is crucial.

There are various classes of void finders, but a striking differentiating feature is whether the algorithm assumes a spherical shape for voids. While it can be considered intuitive that a void finder with no shape assumption helps when trying to infer cosmological information from the void shape (that is for applications relying on the void–halo cross-correlation function, such as the Alcock–Paczyński test and RSD measurements from voids, or void ellipticity), it is not so trivial to understand what is the relevance of considering void shape for applications such as the void size function. Our large set of simulations allows us to address this question.

In this section we compare the information content captured when the void shape is measured in detail with the case in which a spherical assumption is made by the void finder. In other words we compare constraints obtained when selecting voids with VIDE, a void finder with no prior on void shape, and a more simplistic spherical-assumption-based void finder (see, e.g., Villaescusa-Navarro et al. 2020b; Bayer et al. 2021b). Figure 9 shows results of the comparison for a volume of $1 \ h^{-3} \, \mathrm{Gpc}^3$. For most of the cosmological parameters considered in this paper the void size function measured by VIDE provides more stringent constraints than the void size
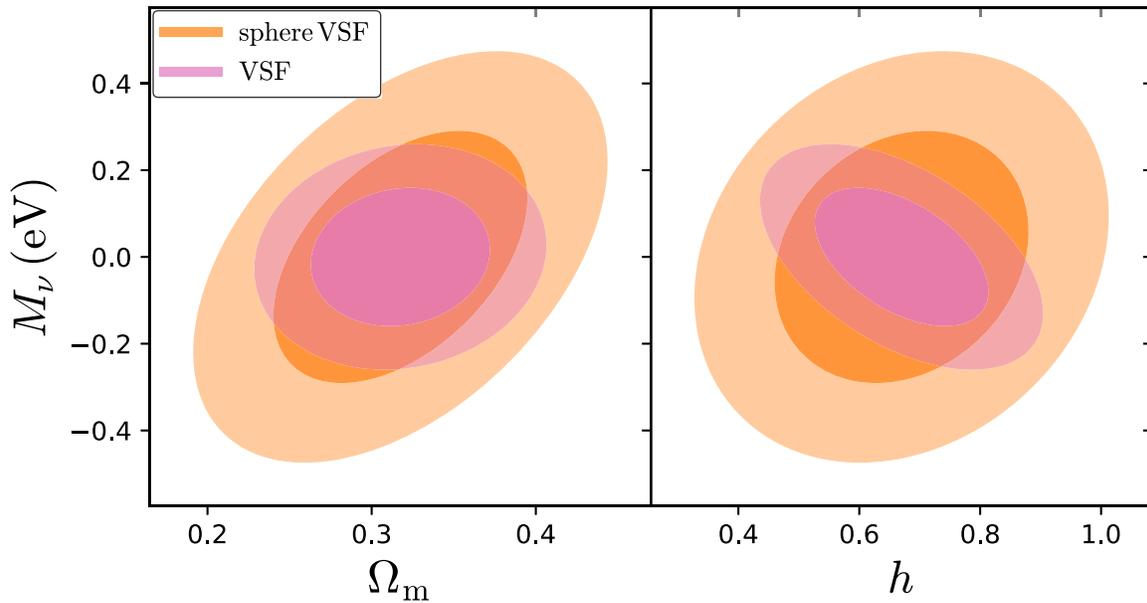
Figure 9. Comparison of constraints from a void finder with no prior on the shape of voids, and a more simplistic spherical-assumption-based void finder. Voids are found from the same halo field.

function measured by the spherical void finder (see Appendix C for the full set of contours). We notice, however, that in some cases the orientation of constraints is different. This consideration is relevant: in future work aiming to maximize the constraints extracted from the cosmic web, it could be possible to combine constraints from different void finders for the void size function. One has to keep in mind, however, that the correlations between constraints from two different void finders may be nontrivial to be accounted for, but GIGANTES contains enough catalogs to estimate this covariance accurately.

We also note that this comparison is more realistic given that it is obtained with voids found in the halo distribution, and not in the dark matter field. These results showcase for the first time that even for nonshape-based applications, such as the void size function, shape plays a strong role in determining the quality of constraints.

5. Likelihood-free Inference of Cosmological Parameters

In this section we show a machine-learning application to the GIGANTES data set. Our goal is to perform likelihood-free inference from one of the most important summary statistics associated to cosmic voids: the void size function. In order to carry out this task, we need many examples from different cosmological models in order to be able to extract unique patterns that allow us to find a connection between the void size function and the value of the cosmological parameters.

We use the void catalogs from the high-resolution QUIJOTE latin-hypercube at $z = 0$ over a volume of $1\ h^{-3}\,\mathrm{Gpc}^3$. We use 18 bins equally spaced from $R = 10.5h^{-1}$ Mpc to $R = 61.5h^{-1}$ Mpc. Our goal is to predict the mean (θ) and standard deviation of the posterior ($\delta\theta$) from the void size function

$$\theta \pm \delta\theta = f(\mathrm{VSF}), \qquad (7)$$

where we are going to use neural networks to approximate the function f. We use moment networks (Jeffrey & Wandelt 2020)

to achieve this. To carry out this task we first compute the void size function using the above bins for each of the 1000 high-resolution QUIJOTE simulations. We then split the data into training (900 VSF), validation (50 VSF), and testing (50 VSF). Next, we use fully connected layers to find the relation between the VSF and the value of the cosmological parameters, using the Jeffrey & Wandelt (2020) loss function to estimate posterior means and standard deviations. The number of layers, the number of neurons per layers, the learning rate, and the value of the weight decay are considered as hyperparameters that we optimize.

We show the results in Figure 10 for 40 different void size functions of the test set. We note that we only show results for Ω_m and n_s, as those are the only two parameters for which we have enough statistical power to find a relation between the void size function and the value of the parameters. For Ω_m we obtain a $\sim$20% constraint. For the case of Ω_b, h, and σ_8 our model just predicts the mean with a large error bar that is not meaningful due to the edges of the prior.

As it can be seen, our method is able to recover the true value within the standard deviation in most of the cases; it is expected that the true value lies outside of the posterior standard deviation for some cases.

This simple application illustrates the power of the so-called simulation-based inference, where accurate and reliable simulations can be used to perform inference over statistics whose likelihood is unknown. We emphasize that this procedure is not limited to summary statistics, but can be carried out at the field level.

6. Conclusion

We have introduced the state of the art GIGANTES void catalogs suite obtained from the QUIJOTE simulations. The catalogs, built with the popular void finder VIDE, provide a comprehensive set of more than 1 billion voids in a large range of ΛCDM cosmologies.
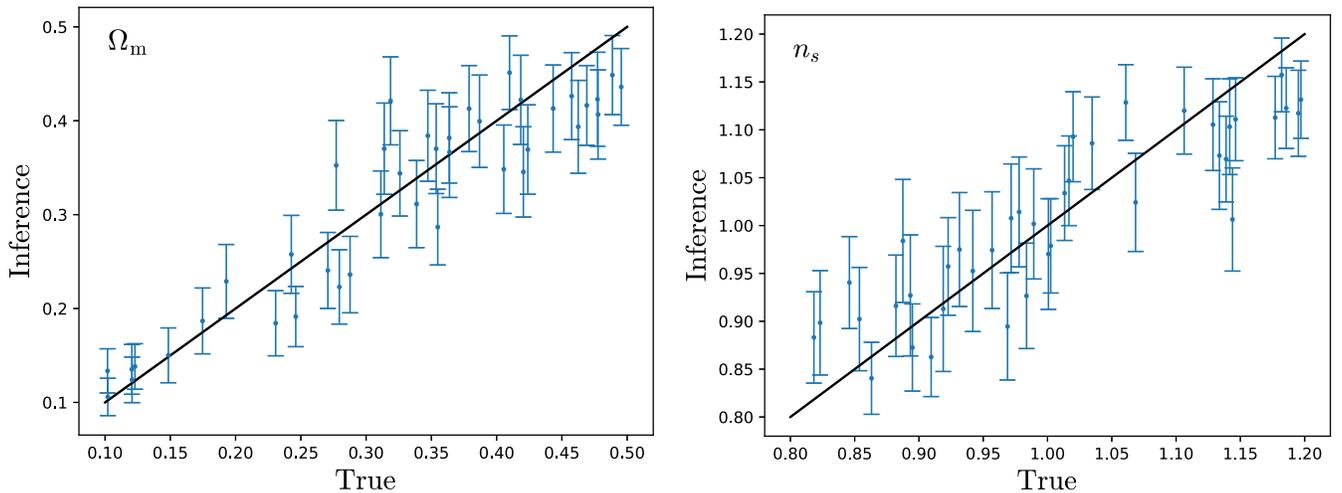
Figure 10. We use likelihood-free inference with moment networks (Jeffrey & Wandelt 2020) to estimate the mean and standard deviations of the posterior from measurements of the void size function. The above panels show the results for Ω_m (left) and n_s (right). Each point represents the posterior mean from a measurement of the void size function from one realization of the HR latin hypercube at $z = 0$. The error bars display the standard deviation of the posterior. For Ω_m (left) and n_s we are able to recover the true value in most of cases; it is expected that some points lie outside the posterior standard deviation interval. We are unable to recover the value of Ω_b, h, and σ_8 from these measurements, as we do not have enough statistical power for these parameters as we only use the VSF from 1 h^{-3} Gpc^3.

To present the GIGANTES data set we show the void size function, average profiles, and the void autocorrelation function. We also compute the correlation among and between different void summary statistics, as well as between traditional quantities such as the halo mass function and the halo autocorrelation function.

To illustrate the immense power provided by the GIGANTES catalogs, we show for the first time in a realistic setup of voids found in the halo field that voids provide complementary and independent information with respect to halos. Voids allow us to extract higher-order information from the cosmic web that would otherwise remain inaccessible. Our results show that added information is particularly effective with respect to constraining neutrinos and h but helps for most parameters of interest.

We show that, aside from the void–halo cross-correlation function and the void size function, a strong contribution to constraints comes from the void autocorrelation function. We also compare the power of halos with respect to voids when combining different summary statistics. Finally, by comparing constraints from VIDE voids with voids from a spherical void finder, we show that void shape knowledge provides an important contribution to the overall information content from voids.

While providing the largest and most realistic set of voids ever released, this paper answers many open questions relevant for void cosmology: Do voids provide additional information? How is that information correlated? Which parameters are optimally constrained by voids? Is void shape important for cosmological constraints? We have showed that voids allow us to go beyond the two-point correlation function and provide additional constraining power. GIGANTES opens up the exploration of the cosmological constraining power of void statistics beyond the ones considered in this paper. Theory developments and instrument systematics analyses for void summary statistics will allow us to fully exploit the power of the void size function, the void–halo, the void–void correlation functions, and other void statistics to constrain cosmological parameters with voids from the next generation of surveys.

Appendix A
Simulations and Void Finder

A.1. The QUIJOTE Simulations

The QUIJOTE simulations (Villaescusa-Navarro et al. 2020b) are a suite of 44,100 full *N*-body simulations exploring more than 7000 cosmological models. These simulations contain trillions of particles over a combined volume larger than the volume of the entire observable universe, and were designed for two main tasks: (1) to quantify the information content on cosmological observables, and (2) to provide enough data to train machine-learning algorithms. Given their immense combined volume, they represent a powerful tool to explore the properties of extreme objects, such as galaxy clusters and cosmic voids. The simulations have been used in a range of papers (e.g., Giusarma et al. 2019; Obuljen et al. 2019; Allys et al. 2020; Aviles & Banerjee 2020; Alves de Oliveira et al. 2020; Banerjee et al. 2020; Chen et al. 2020; Cranmer et al. 2020; de la Bella et al. 2021; Dai et al. 2020; Friedrich et al. 2020; Giri & Smith 2020; Hahn et al. 2020; Kodi Ramanah et al. 2020; Philcox et al. 2020a, 2020b; Uhlemann et al. 2020; Banerjee & Abel 2021a, 2021b; Bayer et al. 2021a, 2021b; Bernal et al. 2021; Chartier et al. 2021; Gualdi et al. 2021a, 2021b; Hahn & Villaescusa-Navarro 2021; Kuruvilla & Aghanim 2021; Lazeyras et al. 2021; Massara et al. 2021; Parimbelli et al. 2021; Philcox 2021; Philcox et al. 2021; Samushia et al. 2021). On average, each realization of the QUIJOTE simulations contains ~500,000 dark matter halos over a volume of 1 $(h^{-1}$ Gpc$)^3$ at $z = 0$, corresponding

to a halo number density of $4 \times 10^{-4} \, h^3 \, \mathrm{Mpc}^{-3}$ and mean halo separation of $\sim 13 \, h^{-1}$ Mpc. The equivalent values at $z = 1$ are $2 \times 10^{-4} \, h^3 \, \mathrm{Mpc}^{-3}$ and $\sim 17 \, h^{-1}$ Mpc, respectively. As access to the void hierarchy may depend on tracer number density, it is relevant to notice that the tracer number density of the QUIJOTE halo catalogs is of a similar order of magnitude than what we can expect from surveys such as PFS (Tamura et al. 2016), and denser than DESI (DESI Collaboration et al. 2016). With the QUIJOTE halo catalogs, however, we have a tracer number density below what will be achieved by Euclid (Laureijs et al. 2011), and the Roman Space telescope (Spergel et al. 2015), indicating that these upcoming surveys will be able to fully capture the power from the void hierarchy, and can expect even tighter constraints on cosmological parameters than what we illustrate in this work.

A.2. The Void Finder VIDE

VIDE is the most popular Voronoi-watershed-based void finder used in cosmology. Based on ZOBOV (Neyrinck 2008), VIDE provides a convenient, publicly released toolkit to find voids in simulations (dark matter/hydrodynamical) and data (spectroscopic or photometric data).

The void finder first performs a Voronoi tessellation of the tracer particle field (tracers can be dark matter particles, halos, galaxies), providing a physical splitting of the 3D particle field into cells. Each cell is assigned a density equal to the inverse of the volume of the cells. Starting from local minima (largest cells) of the tessellation, the algorithm creates basins by merging cells with a monotonic increase in density. Through the use of the watershed transform, VIDE provides the final set voids.

Void centers are defined as the volume-weighted barycenters of the Voronoi cells composing the void. Void centers are therefore sensitive to the whole structure of the object. Considering a sphere of equivalent volume to the volume V_i

of all the cells, one can define an effective radius for voids, R:

$$R = \left(\frac{3}{4\pi} \sum_i V_i \right)^{1/3}. \tag{A1}$$

Even though it provides the effective radius (that we simply refer to as radius in the rest of the paper), VIDE makes no assumption on the void shape, conversely to other void finders assuming sphericity.

The absence of a shape prior makes VIDE an excellent option for cosmological analyses for which a precise determination of void shape is a strength (e.g., Alcock-Paczyński test and redshift-space distortion analyses). Finally, VIDE is built to handle a survey mask, and can easily be applied to study voids from both data and simulations, as its wide range of applications shows. Among other results, it has been used to provide the most stringent to date constraints from cosmic voids (Hamaus et al. 2020). Recently, a Python 3 version of the code has been released.[10]

Since most applications of void finding in data for cosmology rely on Voronoi-based void finders, the work performed in this paper for this category of void finders is pertinent to current cosmological observational analyses. The large set of void catalogs presented in this work will be publicly released upon acceptance of the paper.

Appendix B
Fisher Error Stability

Tables 2, 3, 4, 5, 6, and 7 show 1 standard deviation uncertainties on cosmological parameters from the denoted summary statistic and given number of simulation realizations marginalized over the parameters listed as well as Ω_b. Percent differences relative to the error with 500 simulations are shown in parentheses. For all void summary statistics, the percent

Table 2
Error Sensitivity for HMF + VSF + ξ_{vv} + ξ_{hh}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
500	0.0022	0.052	0.030	0.0054	0.089
450 (10%)	0.0022 (0%)	0.048 (8%)	0.029 (3%)	0.0052 (4%)	0.086 (3%)
400 (20%)	0.0022 (0%)	0.048 (8%)	0.028 (7%)	0.0053 (2%)	0.083 (7%)

Table 3
Error Sensitivity for VSF

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
500	0.072	0.19	0.15	0.11	0.21
450 (10%)	0.069 (4%)	0.18 (5%)	0.14 (7%)	0.099 (10%)	0.21 (0%)
400 (20%)	0.062 (14%)	0.18 (5%)	0.13 (13%)	0.094 (15%)	0.18 (14%)

Table 4
Error Sensitivity for ξ_{vv}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
500	0.037	0.089	0.086	0.067	0.13
450 (10%)	0.036 (3%)	0.082 (5%)	0.079 (8%)	0.064 (4%)	0.12 (8%)
400 (20%)	0.032 (14%)	0.077 (13%)	0.074 (14%)	0.065 (3%)	0.12 (8%)

---

[10] https://bitbucket.org/cosmicvoids/vide_public/wiki/Home

Table 5
Error Sensitivity for ξ_{vh}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
500	0.027	0.067	0.066	0.063	0.10
450 (10%)	0.026 (4%)	0.061(9%)	0.063 (5%)	0.062 (2%)	0.10 (0%)
400 (20%)	0.024 (11%)	0.061 (9%)	0.062 (6%)	0.062 (2%)	0.090 (10%)

Table 6
Error Sensitivity for HMF

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
500	0.033	0.28	0.13	0.078	1.56
450 (10%)	0.027 (18%)	0.25 (11%)	0.12 (8%)	0.064 (18%)	1.23 (21%)
400 (20%)	0.028 (15%)	0.26 (7%)	0.12 (8%)	0.065 (17%)	1.27 (19%)

Table 7
Error Sensitivity for ξ_{hh}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
500	0.027	0.17	0.15	0.11	0.22
450 (10%)	0.028 (4%)	0.18 (6%)	0.16 (7%)	0.11 (0%)	0.22 (0%)
400 (20%)	0.029 (7%)	0.19 (12%)	0.17 (13%)	0.11 (0%)	0.22 (0%)

Table 8
Covariance Error Sensitivity for HMF + VSF + ξ_{vv} + ξ_{hh}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
15000	0.0022	0.052	0.030	0.0054	0.089
13500 (10%)	0.0022 (0%)	0.052 (0%)	0.030 (0%)	0.0054 (0%)	0.089 (0%)
12000 (20%)	0.0022 (0%)	0.052 (0%)	0.030 (0%)	0.0054 (0%)	0.089 (0%)

Table 9
Covariance Error Sensitivity for VSF

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
15000	0.072	0.19	0.15	0.10	0.21
13500 (10%)	0.072 (0%)	0.19 (0%)	0.15 (0%)	0.10 (0%)	0.21 (0%)
12000 (20%)	0.071 (1%)	0.19 (0%)	0.15 (0%)	0.10 (0%)	0.21 (0%)

Table 10
Covariance Error Sensitivity for ξ_{vv}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
15000	0.037	0.089	0.086	0.067	0.13
13500 (10%)	0.037 (0%)	0.089 (0%)	0.085 (1%)	0.066 (1%)	0.13 (0%)
12000 (20%)	0.037 (0%)	0.089 (0%)	0.085 (1%)	0.066 (1%)	0.13 (0%)

Table 11
Covariance Error Sensitivity for ξ_{vh}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
15000	0.027	0.067	0.065	0.063	0.10
13500 (10%)	0.027 (0%)	0.067 (0%)	0.065 (0%)	0.062 (2%)	0.10 (0%)
12000 (20%)	0.027 (0%)	0.067 (0%)	0.065 (0%)	0.062 (2%)	0.10 (0%)

Table 12
Covariance Error Sensitivity for HMF

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
15000	0.033	0.28	0.13	0.078	1.56
13500 (10%)	0.033 (0%)	0.28 (0%)	0.13 (0%)	0.078 (0%)	1.56 (0%)
12000 (20%)	0.033 (0%)	0.28 (0%)	0.13 (0%)	0.078 (0%)	1.56 (0%)

Table 13
Covariance Error Sensitivity for ξ_{hh}

Number of Simulations	Ω_m	h	n_s	σ_8	$\sum m_\nu$
15000	0.027	0.17	0.15	0.11	0.22
13500 (10%)	0.027 (0%)	0.17 (0%)	0.15 (0%)	0.11 (0%)	0.22 (0%)
12000 (20%)	0.027 (0%)	0.17 (0%)	0.15 (0%)	0.11 (0%)	0.22 (0%)

Table 14
Impact of Small Scales on ξ Constraints for $\{\Omega_m, h, n_s, \sigma_8, M_\nu\}$

Summary Statistic(s)	15–200 h^{-1} Mpc	20–200 h^{-1} Mpc
ξ_{vv}	0.037 (30%), 0.089 (11%), 0.086 (11%), 0.067 (17%), 0.13 (7%)	0.053, 0.10, 0.097, 0.081, 0.14
ξ_{vh}	0.027 (10%), 0.067 (7%), 0.065 (14%), 0.063 (6%), 0.10 (9%)	0.030, 0.072, 0.076, 0.067, 0.11
ξ_{hh}	0.027 (10%), 0.17 (11%), 0.15 (12%), 0.11 (27%), 0.22 (12%)	0.030, 0.19, 0.17, 0.15, 0.25
HMF + VSF + ξ_{vv} + ξ_{hh}	0.0022 (12%), 0.052 (5%), 0.030 (6%), 0.0054 (8%), 0.089 (8%)	0.0025, 0.055, 0.032, 0.0059, 0.097

change in error is less than the percent change in the number of realizations, indicating the constraints are stable.

Tables 8, 9, 10, 11, 12, and 13 also show 1 standard deviation uncertainties, but given the number of simulations used to compute the covariance. Percent differences relative to the error with 15000 simulations are shown in parentheses. Most percent changes are below 1%, indicating the covariance is stable.

Finally, in Table 14, we illustrate the impact on the Fisher uncertainties from computing the correlation functions on scales $15 - 200\,h^{-1}$ Mpc in log space compared to computing the correlation functions on scales 20–$200\,h^{-1}$ Mpc in log space. Percent differences relative to scales $20 - 200\,h^{-1}$ Mpc are shown in parentheses. Most changes are near 10%,

although a handful are larger. An increase in error is expected as less data are used.
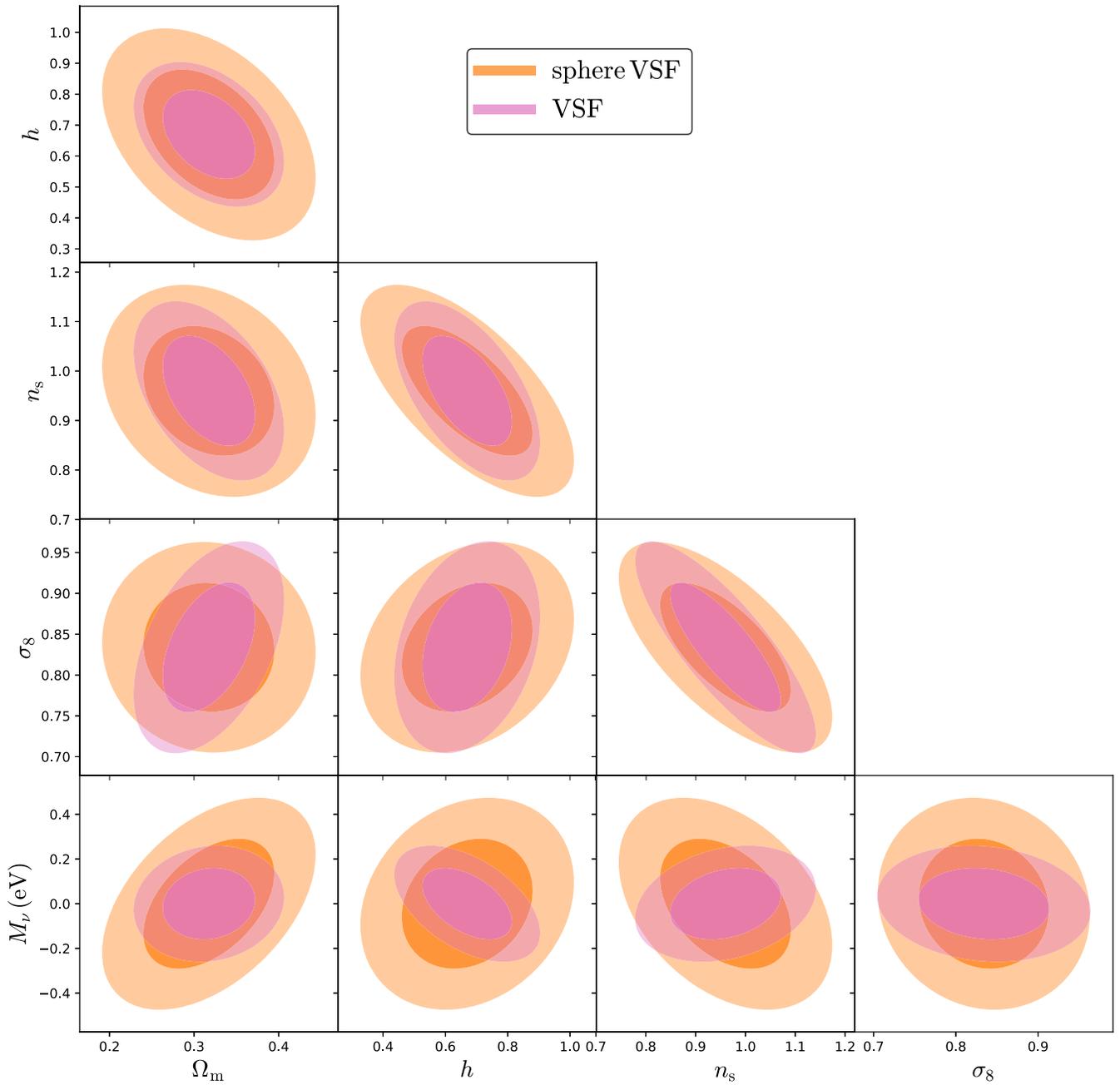
Appendix C
Full Constraints

We show in Figure 11 the constraints from different void summary statistics, as well as for the halo mass function and the halo autocorrelation function, for all the parameters considered in this paper: the sum of neutrino masses, σ_8, n_s, h, and Ω_m. In Figures 12 and 13 we show full cosmological parameter contours for combined halo and combined void summary statistics, and the spherical void finder VSF and VIDE VSF, respectively.
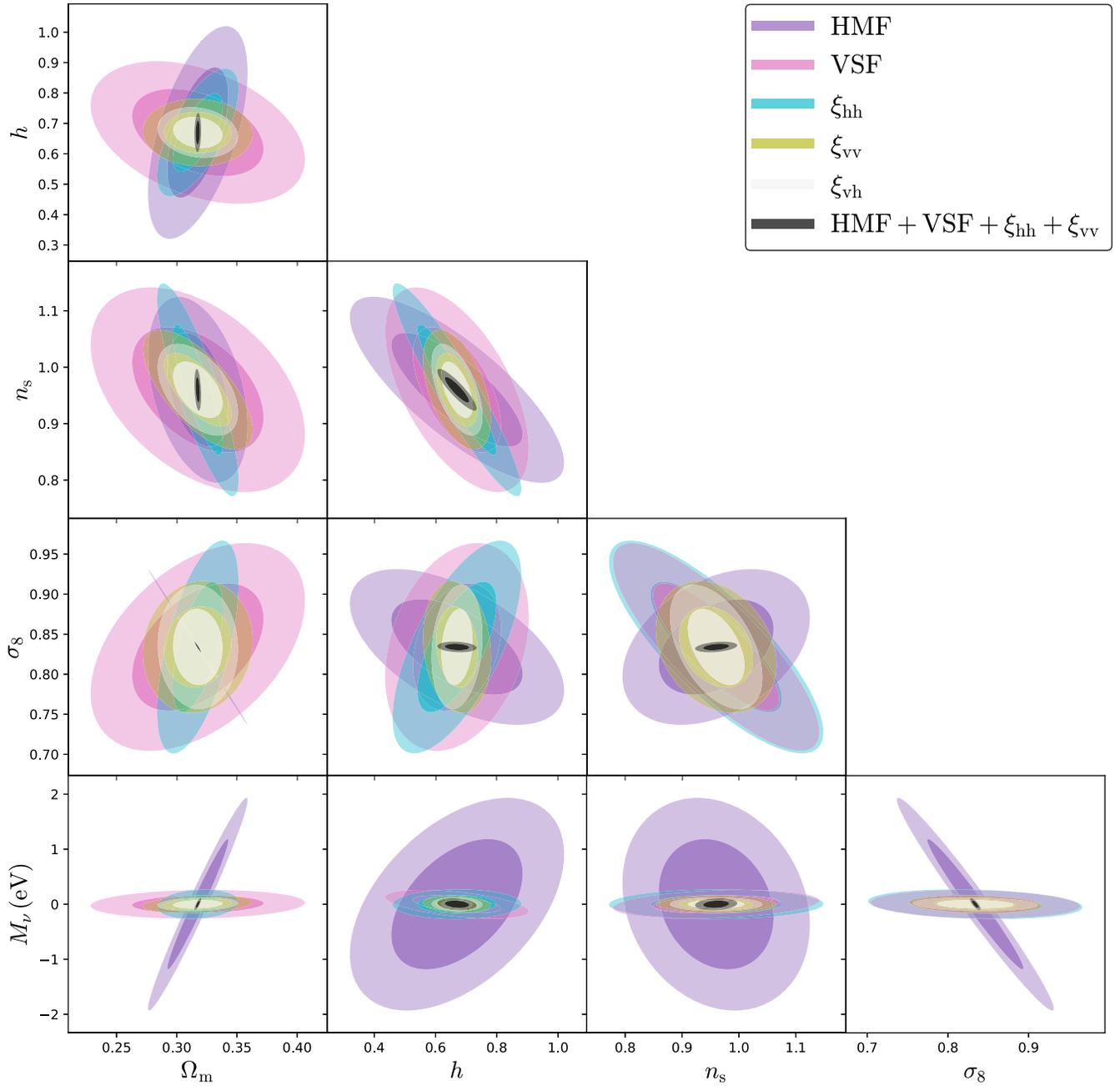
Figure 11. Constraints on cosmological parameters from the voids (void size function, void–halo, and void–void correlation functions) and halos (halo mass function, halo autocorrelation function).
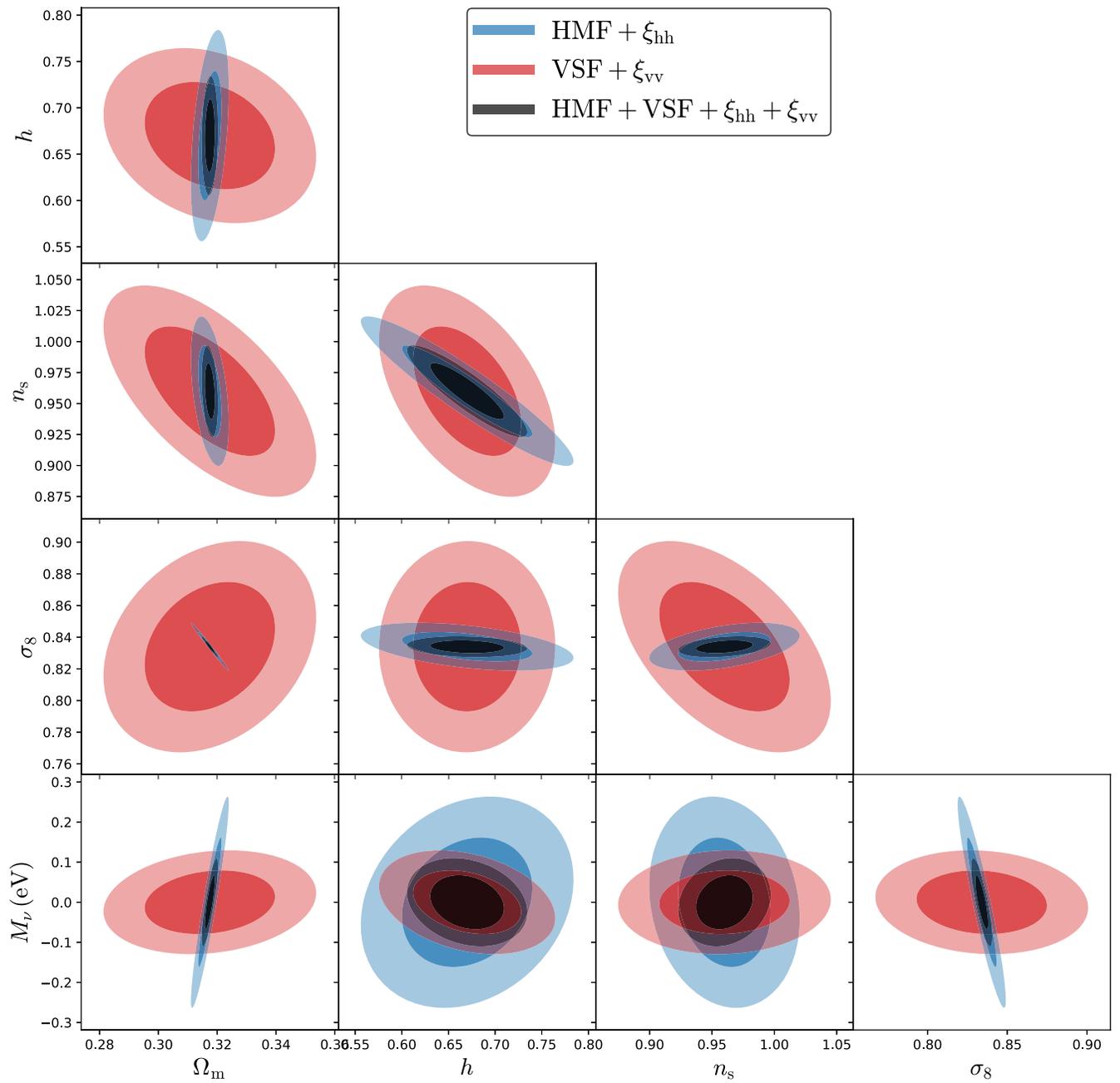
Figure 12. Comparison of halo-only with void-only constraints (therefore not including the void–halo cross correlation) for all parameters.

Figure 13. Comparison of constraints for all parameters from a void finder with no prior on the shape of voids, and a more simplistic spherical-assumption-based void finder. Voids are found from the same halo field.

ORCID iDs

Christina D. Kreisch ⓘ https://orcid.org/0000-0002-5061-7805
Alice Pisani ⓘ https://orcid.org/0000-0002-6146-4437
Francisco Villaescusa-Navarro ⓘ https://orcid.org/0000-0002-4816-0455
David N. Spergel ⓘ https://orcid.org/0000-0002-5151-0006
Benjamin D. Wandelt ⓘ https://orcid.org/0000-0002-5854-8269
Nico Hamaus ⓘ https://orcid.org/0000-0002-0876-2101
Adrian E. Bayer ⓘ https://orcid.org/0000-0002-3568-3900

References

Achitouv, I. 2016, PhRvD, 94, 103524
Achitouv, I. 2019, PhRvD, 100, 123513
Achitouv, I., Blake, C., Carter, P., Koda, J., & Beutler, F. 2017, PhRvD, 95, 083502
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alam, S., Aviles, A., Bean, R., et al. 2021, JCAP, 2021, 050
Alcock, C., & Paczynski, B. 1979, Natur, 281, 358
Allys, E., Marchand, T., Cardoso, J. F., et al. 2020, PhRvD, 102, 103506
Alves de Oliveira, R., Li, Y., Villaescusa-Navarro, F., Ho, S., & Spergel, D. N. 2020, arXiv:2012.00240
Aubert, M., Cousinou, M.-C., Escoffier, S., et al. 2022, MNRAS, 513, 186
Aviles, A., & Banerjee, A. 2020, JCAP, 2020, 034
Banerjee, A., & Abel, T. 2021a, MNRAS, 500, 5479
Banerjee, A., & Abel, T. 2021b, MNRAS, 504, 2911
Banerjee, A., Castorina, E., Villaescusa-Navarro, F., Court, T., & Viel, M. 2020, JCAP, 2020, 032
Banerjee, A., & Dalal, N. 2016, JCAP, 1611, 015
Baron, D. 2019, arXiv:1904.07248
Barreira, A., Cautun, M., Li, B., Baugh, C. M., & Pascoli, S. 2015, JCAP, 2015, 028
Bayer, A. E., Banerjee, A., & Feng, Y. 2021a, JCAP, 2021, 016
Bayer, A. E., Villaescusa-Navarro, F., Massara, E., et al. 2021b, ApJ, 919, 24
Bernal, J. L., Caputo, A., Villaescusa-Navarro, F., & Kamionkowski, M. 2021, PhRvL, 127, 131102
Biswas, R., Alizadeh, E., & Wandelt, B. D. 2010, PhRvD, 82, 023002
Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Natur, 380, 603
Bos, E. G. P., van de Weygaert, R., Dolag, K., & Pettorino, V. 2012, MNRAS, 426, 440
Cai, Y.-C., Kaiser, N., Cole, S., & Frenk, C. 2017, MNRAS, 468, 1981
Cai, Y.-C., Padilla, N., & Li, B. 2015, MNRAS, 451, 1036
Cai, Y.-C., Taylor, A., Peacock, J. A., & Padilla, N. 2016, MNRAS, 462, 2465
Cautun, M., Paillas, E., Cai, Y.-C., et al. 2018, MNRAS, 476, 3195
Chalapathy, R., & Chawla, S. 2019, arXiv:1901.03407
Chan, K. C., Hamaus, N., & Desjacques, V. 2014, PhRv, D90, 103521
Charnock, T., Lavaux, G., & Wandelt, B. D. 2018, PhRvD, 97, 083004
Chartier, N., Wandelt, B., Akrami, Y., & Villaescusa-Navarro, F. 2021, MNRAS, 503, 1897
Chen, C., Li, Y., Villaescusa-Navarro, F., Ho, S., & Pullen, A. 2020, arXiv:2012.05472
Clampitt, J., Cai, Y.-C., & Li, B. 2013, MNRAS, 431, 749
Clampitt, J., Jain, B., & Sánchez, C. 2016, MNRAS, 456, 4425
Colberg, J. M., Pearce, F., Foster, C., et al. 2008, MNRAS, 387, 933
Contarini, S., Marulli, F., Moscardini, L., et al. 2021, MNRAS, 504, 5021
Contarini, S., Ronconi, T., Marulli, F., et al. 2019, MNRAS, 488, 3526
Correa, C. M., Paz, D. J., Padilla, N. D., et al. 2019, MNRAS, 485, 5761
Cousinou, M. C., Pisani, A., Tilquin, A., et al. 2019, A&C, 27, 53
Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., et al. 2020, arXiv:2006.11287
Dai, J.-P., Verde, L., & Xia, J.-Q. 2020, JCAP, 2020, 007
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
de la Bella, L. F., Tessore, N., & Bridle, S. 2021, JCAP, 2021, 001
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, arXiv:1805.05489
Falck, B., Koyama, K., Zhao, G.-B., & Cautun, M. 2018, MNRAS, 475, 3262
Fang, Y., Hamaus, N., Jain, B., et al. 2019, MNRAS, 490, 3573
Friedrich, O., Uhlemann, C., Villaescusa-Navarro, F., et al. 2020, MNRAS, 498, 464
Fry, J. N. 1986, ApJ, 306, 358
Fry, J. N., & Colombi, S. 2013, MNRAS, 433, 581
Giri, U., & Smith, K. M. 2020, arXiv:2010.07193
Giusarma, E., Reyes Hurtado, M., Villaescusa-Navarro, F., et al. 2019, arXiv:1910.04255
Gregory, S. A., & Thompson, L. A. 1978, ApJ, 222, 784
Gualdi, D., Gil-Marin, H., & Verde, L. 2021a, JCAP, 2021, 008
Gualdi, D., Novell, S., Gil-Marín, H., & Verde, L. 2021b, JCAP, 2021, 015
Habouzit, M., Pisani, A., Goulding, A., et al. 2020, MNRAS, 493, 899
Hahn, C., & Villaescusa-Navarro, F. 2021, JCAP, 2021, 029
Hahn, C., Villaescusa-Navarro, F., Castorina, E., & Scoccimarro, R. 2020, JCAP, 2020, 040
Hamaus, N., Cousinou, M.-C., Pisani, A., et al. 2017, JCAP, 2017, 014
Hamaus, N., Pisani, A., Choi, J.-A., et al. 2020, JCAP, 2020, 023
Hamaus, N., Pisani, A., Sutter, P. M., et al. 2016, PhRvL, 117, 091302
Hamaus, N., Sutter, P. M., Lavaux, G., & Wandelt, B. D. 2014, JCAP, 1412, 013
Hamaus, N., Sutter, P. M., Lavaux, G., & Wandelt, B. D. 2015, JCAP, 11, 036
Hamaus, N., Sutter, P. M., & Wandelt, B. D. 2014b, PhRvL, 112, 251302
Hamaus, N., Wandelt, B. D., Sutter, P. M., Lavaux, G., & Warren, M. S. 2014a, PhRvL, 112, 041304
Hawken, A. J., Aubert, M., Pisani, A., et al. 2020, JCAP, 2020, 012
Hawken, A. J., Granett, B. R., Iovino, A., et al. 2017, A&A, 607, A54
He, S., Li, Y., Feng, Y., et al. 2019, PNAS, 116, 13825
Hoyle, F., Rojas, R. R., Vogeley, M. S., & Brinkmann, J. 2005, ApJ, 620, 618
Jeffrey, N., Gatti, M., Chang, C., et al. 2021, MNRAS, 505, 4626
Jeffrey, N., & Wandelt, B. D. 2020, arXiv:2011.05991
Jennings, E., Li, Y., & Hu, W. 2013, MNRAS, 434, 2167
Jõeveer, M., Einasto, J., & Tago, E. 1978, MNRAS, 185, 357
Kirshner, R. P., Oemler, A., Jr, Schechter, P. L., & Shectman, S. A. 1981, ApJL, 248, L57
Kodi Ramanah, D., Charnock, T., Villaescusa-Navarro, F., & Wandelt, B. D. 2020, MNRAS, 495, 4227
Kreckel, K., Platen, E., Aragón-Calvo, M. A., et al. 2012, AJ, 144, 16
Kreisch, C. D., Pisani, A., Carbone, C., et al. 2019, MNRAS, 488, 4413
Kuruvilla, J., & Aghanim, N. 2021, A&A, 653, A130
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Lavaux, G., & Wandelt, B. D. 2010, MNRAS, 403, 1392
Lavaux, G., & Wandelt, B. D. 2012, ApJ, 754, 109
Lazeyras, T., Villaescusa-Navarro, F., & Viel, M. 2021, JCAP, 2021, 022
Lee, J., & Park, D. 2009, ApJL, 696, L10
Massara, E., Villaescusa-Navarro, F., Ho, S., Dalal, N., & Spergel, D. N. 2021, PhRvL, 126, 011301
Massara, E., Villaescusa-Navarro, F., Viel, M., & Sutter, P. M. 2015, JCAP, 1511, 018
Miguel de Cervantes 1605, Don Quixote de la Mancha (Madrid: Francisco de Robles), https://www.cervantesvirtual.com/obra-visor/don-quixote-1605-1615–0/html/
Miguel de Cervantes 1615, Don Quixote de la Mancha (Madrid: Francisco de Robles), https://www.cervantesvirtual.com/obra-visor/don-quixote-1605-1615–0/html/
Nadathur, S., Carter, P. M., Percival, W. J., Winther, H. A., & Bautista, J. E. 2019, PhRvD, 100, 023504
Nadathur, S., Woodfinden, A., Percival, W. J., et al. 2020, MNRAS, 499, 4140
Neyrinck, M. C. 2008, MNRAS, 386, 2101
Ntampaka, M., Avestruz, C., Boada, S., et al. 2019, BAAS, 51, 14
Obuljen, A., Dalal, N., & Percival, W. J. 2019, JCAP, 2019, 020
Paillas, E., Cai, Y.-C., Padilla, N., & Sánchez, A. 2021, MNRAS, 505, 5731
Panchal, R. R., Pisani, A., & Spergel, D. N. 2020, ApJ, 901, 87
Parimbelli, G., Anselmi, S., Viel, M., et al. 2021, JCAP, 2021, 009
Patiri, S. G., Prada, F., Holtzman, J., Klypin, A., & Betancort-Rijo, J. 2006, MNRAS, 372, 1710
Paz, D., Lares, M., Ceccarelli, L., Padilla, N., & Lambas, D. G. 2013, MNRAS, 436, 3480
Peebles, P. J. E. 1980, Large-Scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)
Perico, E. L. D., Voivodic, R., Lima, M., & Mota, D. F. 2019, A&A, 632, A52
Philcox, O. H. E. 2021, MNRAS, 501, 4004
Philcox, O. H. E., Aviles, A., & Massara, E. 2021, JCAP, 2021, 038
Philcox, O. H. E., Massara, E., & Spergel, D. N. 2020a, PhRvD, 102, 043516
Philcox, O. H. E., Spergel, D. N., & Villaescusa-Navarro, F. 2020b, PhRvD, 101, 123520
Pisani, A., Lavaux, G., Sutter, P. M., & Wandelt, B. D. 2014, MNRAS, 443, 3238
Pisani, A., Massara, E., Spergel, D. N., et al. 2019, BAAS, 51, 40
Pisani, A., Sutter, P. M., Hamaus, N., et al. 2015, PhRvD, 92, 083531
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Pollina, G., Baldi, M., Marulli, F., & Moscardini, L. 2016, MNRAS, 455, 3075

18

Pollina, G., Hamaus, N., Dolag, K., et al. 2017, MNRAS, 469, 787
Pollina, G., Hamaus, N., Paech, K., et al. 2019, MNRAS, 487, 2836
Ricciardelli, E., Cava, A., Varela, J., & Quilis, V. 2014, MNRAS, 445, 4045
Ryden, B. S. 1995, ApJ, 452, 25
Ryden, B. S., & Melott, A. L. 1996, ApJ, 470, 160
Sahlén, M. 2019, PhRvD, 99, 063525
Sahlén, M., & Silk, J. 2018, PhRv, D97, 103504
Samushia, L., Slepian, Z., & Villaescusa-Navarro, F. 2021, MNRAS, 505, 628
Sánchez, C., Clampitt, J., Kovacs, A., et al. 2017, MNRAS, 465, 746
Schuster, N., Hamaus, N., Pisani, A., et al. 2019, JCAP, 2019, 055
Sheth, R. K., & van de Weygaert, R. 2004, MNRAS, 350, 517
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
Spolyar, D., Sahlén, M., & Silk, J. 2013, PhRvL, 111, 241103
Sutter, P. M., Lavaux, G., Hamaus, N., et al. 2014, MNRAS, 442, 462
Sutter, P. M., Lavaux, G., Hamaus, N., et al. 2015, A&C, 9, 1
Sutter, P. M., Lavaux, G., Wandelt, B. D., & Weinberg, D. H. 2012, ApJ, 761, 187
Tamura, N., Takato, N., Shimono, A., et al. 2016, Proc. SPIE, 9908, 99081M
The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv:1809.01669
Uhlemann, C., Friedrich, O., Villaescusa-Navarro, F., Banerjee, A., & Codis, S. 2020, MNRAS, 495, 4006
van de Weygaert, R., & Schaap, W. 2009, in Data Analysis in Cosmology, Lecture Notes in Physics, ed. V. J. Martínez, Vol. 665 (Berlin: Springer)
van de Weygaert, R., & van Kampen, E. 1993, MNRAS, 263, 481
Verza, G., Pisani, A., Carbone, C., Hamaus, N., & Guzzo, L. 2019, JCAP, 2019, 040
Vielzeuf, P., Kovács, A., Demirbozan, U., et al. 2021, MNRAS, 500, 464
Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020b, ApJS, 250, 2
Villaescusa-Navarro, F., Vogelsberger, M., Viel, M., & Loeb, A. 2013, MNRAS, 431, 3670
Villaescusa-Navarro, F., Wandelt, B. D., Anglés-Alcázar, D., et al. 2020a, arXiv:2011.05992
Zhang, G., Li, Z., Liu, J., et al. 2020, PhRvD, 102, 083537
Zivick, P., Sutter, P. M., Wandelt, B. D., Li, B., & Lam, T. Y. 2015, MNRAS, 451, 4215