The Astrophysical Journal Supplement Series, 250:2 (20pp), 2020 September https://doi.org/10.3847/1538-4365/ab9d82

The Quijote Simulations

Francisco Villaescusa-Navarro[1,2], ChangHoon Hahn[3,4], Elena Massara[1,5], Arka Banerjee[6,7,8], Ana Maria Delgado[1,9], Doogesh Kodi Ramanah[10,11], Tom Charnock[10], Elena Giusarma[1,12], Yin Li[1,3,4,13,14], Erwan Allys[15], Antoine Brochard[16,17], Cora Uhlemann[18,19], Chi-Ting Chiang[20], Siyu He[1], Alice Pisani[2], Andrej Obuljen[5], Yu Feng[3,4], Emanuele Castorina[3,4], Gabriella Contardo[1], Christina D. Kreisch[2], Andrina Nicola[2], Justin Alsing[1,21], Roman Scoccimarro[22], Licia Verde[23,24], Matteo Viel[25,26,27,28], Shirley Ho[1,2,29], Stephane Mallat[30,31], Benjamin Wandelt[1,10,11], and David N. Spergel[1,2]
[1] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA; villaescusa.francisco@gmail.com
[2] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544-0010, USA
[3] Department of Physics, University of California, Berkeley, CA 94720, USA
[4] Berkeley Center for Cosmological Physics, Berkeley, CA 94720, USA
[5] Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
[6] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA
[7] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[8] SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[9] Department of Physics, New York City College of Technology, Brooklyn, NY 11201, USA
[10] Sorbonne Universite, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, F-75014 Paris, France
[11] Sorbonne Universite, Institut Lagrange de Paris, 98 bis boulevard Arago, F-75014 Paris, France
[12] Department of Physics, Michigan Technological University, Houghton, MI 49931, USA
[13] Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS, The University of Tokyo, Chiba 277-8583, Japan
[14] Center for Computational Mathematics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA
[15] Laboratoire de Physique de l'Ecole normale superieure, ENS, Universite PSL, CNRS, Paris, France
[16] INRIA, ENS, PSL Research University Paris, France
[17] Paris Research Center, Huawei Technologies, Paris, France
[18] Centre for Theoretical Cosmology, DAMTP, University of Cambridge, CB3 0WA Cambridge, UK
[19] Fitzwilliam College, University of Cambridge, CB3 0DG Cambridge, UK
[20] Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA
[21] Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden
[22] Center for Cosmology and Particle Physics, Department of Physics, New York University, NY 10003, New York, USA
[23] Institut de Ciencies del Cosmos, University of Barcelona, ICCUB, Barcelona E-08028, Spain
[24] Institucio Catalana de Recerca i Estudis Avancats, Passeig Lluis Companys 23, Barcelona E-08010, Spain
[25] SISSA, Via Bonomea 265, I-34136 Trieste, Italy
[26] INFN, Sez. di Trieste, Via Valerio 2, I-34127 Trieste, Italy
[27] IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy
[28] INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy
[29] Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[30] Data team, Ecole Normale Supérieure, Université PSL, 45 rue d'Ulm, F-75005 Paris, France
[31] College de France, 11 place Marcelin Berthelot, F-75005, Paris, France
Received 2019 October 27; revised 2020 June 14; accepted 2020 June 15; published 2020 August 20

Abstract

The QUIJOTE simulations are a set of 44,100 full N-body simulations spanning more than 7000 cosmological models in the $\{\Omega_{\rm m}, \Omega_{\rm b}, h, n_s, \sigma_8, M_\nu, w\}$ hyperplane. At a single redshift, the simulations contain more than 8.5 trillion particles over a combined volume of 44,100 $(h^{-1}\,{\rm Gpc})^3$; each simulation follows the evolution of 256^3, 512^3, or 1024^3 particles in a box of 1 h^{-1} Gpc length. Billions of dark matter halos and cosmic voids have been identified in the simulations, whose runs required more than 35 million core hours. The QUIJOTE simulations have been designed for two main purposes: (1) to quantify the information content on cosmological observables and (2) to provide enough data to train machine-learning algorithms. In this paper, we describe the simulations and show a few of their applications. We also release the petabyte of data generated, comprising hundreds of thousands of simulation snapshots at multiple redshifts; halo and void catalogs; and millions of summary statistics, such as power spectra, bispectra, correlation functions, marked power spectra, and estimated probability density functions.

Unified Astronomy Thesaurus concepts: N-body simulations (1083); Cosmological parameters (339); Astrostatistics (1882); Large-scale structure of the universe (902); Cosmological neutrinos (338)

1. Introduction

The discovery of the accelerated expansion of the universe (Riess et al. 1998; Perlmutter et al. 1999) has revolutionized cosmology. We now believe that $\simeq$70% of the energy content of the universe is made up of a mysterious substance that is accelerating the expansion of the universe: dark energy. One of the most important tasks in modern cosmology is to unveil the properties of dark energy. This will help us to understand its nature and improve our knowledge of fundamental physics.

The spatial distribution of matter in the universe is sensitive to the nature of dark energy but also to other fundamental quantities, such as the properties of dark matter, the sum of neutrino masses, and the initial conditions (ICs) of the universe. Thus, one of the most powerful ways to learn about fundamental physics is by extracting that information from the large-scale structure of the

universe. This is the goal of many upcoming cosmological missions, such as DESI,[32] Euclid,[33] LSST,[34] PFS,[35] SKA,[36] and WFIRST.[37]

The traditional way to retrieve information from cosmological observations is to compare summary statistics from data against theory predictions. An important question is then: what statistic or statistics should be used to extract the maximum information[38] from cosmic observations?

It is well known that a Gaussian density field can be fully described by its power spectrum or correlation function (see, e.g., Verde 2007; Wandelt 2013; Leclercq et al. 2014). This is the main reason why the power spectrum/correlation function is the most prominent statistic employed when analyzing cosmological data; at high redshift, or on sufficiently large scales at low redshift, the universe resembles a Gaussian density field, and most of the information embedded on it can be extracted from the power spectrum/correlation function.

The cosmic microwave background (CMB) is an example of a Gaussian density field.[39] All of the information embedded in it can thus be retrieved through the power spectrum. Notice that for simplicity, we are ignoring the non-Gaussian information that can be extracted from the CMB, e.g., through CMB lensing. Currently, some of the tightest and more robust constraints on the value of the cosmological parameters arise from CMB data (Planck Collaboration et al. 2018). Unfortunately, the primary CMB is limited to a plane on the sky at high redshift and is insensitive to low-redshift phenomena, such as the transition from the matter-dominated epoch to the dark energy–dominated epoch.

Since the number of modes in 3D surveys is potentially much larger than in CMB observations, it is expected that the constraining power of those surveys will surpass that of CMB observations. Unfortunately, in 3D surveys, most of the modes are on mildly to nonlinear scales. In the regime where the density field is non-Gaussian, it is currently unknown what statistic or set of statistics will place the tightest constraints on the parameters. From a formal perspective, that question is also mathematically intractable. Being able to extract the cosmological information embedded in nonlinear modes will enable us to tighten the value of the cosmological parameters and therefore improve our understanding of fundamental physics.

One way to tackle this problem is to consider a given statistic/statistics and quantify the information content of it from linear to nonlinear scales. Numerical simulations are needed in this case, as they are one of the most powerful ways to obtain theory predictions in the fully nonlinear regime in real and redshift space for any considered statistic. This is the motivation that brought us to develop the Quijote simulations; we designed them to allow the community to easily quantify the information content of different statistics into the fully nonlinear regime.

[32] https://www.desi.lbl.gov
[33] https://www.euclid-ec.org
[34] https://www.lsst.org
[35] https://pfs.ipmu.jp/index.html
[36] https://www.skatelescope.org
[37] https://wfirst.gsfc.nasa.gov/index.html
[38] By information, we mean the constraints on the value of the cosmological parameters.
[39] To date, there is no significant evidence that points toward the CMB being non-Gaussian (Planck Collaboration et al. 2019).

Another way to approach the problem is to use advanced statistical techniques, such as machine/deep learning, to identify new and optimal statistics to extract cosmological information (Ravanbakhsh et al. 2017; Charnock et al. 2018; Alsing et al. 2019). One of the requirements of these methods is to have a sufficiently large data set to train the algorithms. The Quijote simulations have been designed to provide the community with a very big data set of cosmological simulations.

In this paper, we present the Quijote suite, the largest set of full N-body simulations[40] run at this mass and spatial resolution to date. The Quijote simulations contain 44,100 full N-body simulations expanding more than 7000 cosmological models, and at a single redshift, they contain more than 8.5 trillion particles. The computational cost of the simulations exceeds 35 million CPU hr, and over 1 PB of data were generated.

We note that our simulations have a relatively low resolution; they resolve halos with masses above $\simeq 3 \times 10^{12}\, h^{-1}\, M_\odot$ (high resolution) or $\simeq 2 \times 10^{13}\, h^{-1}\, M_\odot$ (fiducial resolution). Running the Quijote simulation suite at higher resolution would have been impossible due to computational and storage constraints.

The reason we chose to run simulations at this resolution is twofold: (1) we need to run a large set of simulations to evaluate the Fisher matrix with all of its ingredients converged, and (2) we wanted to sample a very large cosmological volume first and then use machine learning to increase the resolution of the simulations (see below).

The Quijote simulations, therefore, represent a very useful tool to quantify information content on the matter field and for the most massive halos/luminous galaxies. While the matter field is not directly observable in 3D (its projection is observed through weak lensing), quantifying the information content on different observables of it is still a useful exercise. If an observable provides competitive constraints on a given parameter for the matter field, it is worth exploring how much information will remain when considering galaxies. On the other hand, if a statistic does not accurately constrain the value of the parameters for the matter field, it is unlikely that it will do so for galaxies.

This paper is organized as follows. In Section 2 we describe in detail the Quijote simulations. We outline the data products generated by the simulations in Section 3. We present a few applications of the Quijote simulations in Section 4. In Section 5 we show several convergence tests in order to quantify the limitations of the simulations. Finally, we draw our conclusions in Section 6.

2. Simulations

All of the simulations in the Quijote suite are N-body simulations. They have been run using the TreePM code GADGET-III, an improved version of GADGET-II (Springel 2005).

The ICs of all simulations are generated at $z = 127$. We obtain the input matter power spectrum and transfer functions by rescaling the $z = 0$ matter power spectrum and transfer functions from CAMB (Lewis et al. 2000). For models with massive neutrinos, we use the rescaling method developed in Zennaro et al. (2017), while for models with massless neutrinos, we employ the traditional scale-independent rescaling,

$$P_{\rm m}(k, z_i) = \left(\frac{D(z_i)}{D(z)}\right)^2 P_{\rm m}(k, z = 0), \qquad f(z_i) \simeq \Omega_{\rm m}^{\gamma}(z_i)\,, \quad (1)$$

[40] To the best of our knowledge.

Table 1
Characteristics of the QUIJOTE Simulations

Name	Ω_m	Ω_b	h	n_s	σ_8	M_ν(eV)	w	δ_b	Realizations	Simulations	ICs	$N_c^{1/3}$	$N_\nu^{1/3}$
Fid	0.3175	0.049	0.6711	0.9624	0.834	0	−1	0	15,000	Standard	2LPT	512	0
									500	Standard	Zel'dovich	512	0
									500	Paired fixed	2LPT	512	0
									1000	Standard	2LPT	256	0
									100	Standard	2LPT	1024	0
Ω_m^+	0.3275	0.049	0.6711	0.9624	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
Ω_m^-	0.3075	0.049	0.6711	0.9624	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
Ω_b^{++}	0.3175	0.051	0.6711	0.9624	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
Ω_b^+	0.3175	0.050	0.6711	0.9624	0.834	0	−1	0	500	Paired fixed	2LPT	512	0
Ω_b^-	0.3175	0.048	0.6711	0.9624	0.834	0	−1	0	500	Paired fixed	2LPT	512	0
Ω_b^{--}	0.3175	0.047	0.6711	0.9624	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
h^+	0.3175	0.049	0.6911	0.9624	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
h^-	0.3175	0.049	0.6511	0.9624	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
n_s^+	0.3175	0.049	0.6711	0.9824	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
n_s^+	0.3175	0.049	0.6711	0.9424	0.834	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
σ_8^+	0.3175	0.049	0.6711	0.9624	0.849	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
σ_8^-	0.3175	0.049	0.6711	0.9624	0.819	0	−1	0	500	Standard	2LPT	512	0
									500	Paired fixed			
M_ν^{+++}	0.3175	0.049	0.6711	0.9624	0.834	0.4	−1	0	500	Standard	Zel'dovich	512	512
									500	Paired fixed			
M_ν^{++}	0.3175	0.049	0.6711	0.9624	0.834	0.2	−1	0	500	Standard	Zel'dovich	512	512
									500	Paired fixed			
M_ν^+	0.3175	0.049	0.6711	0.9624	0.834	0.1	−1	0	500	Standard	Zel'dovich	512	512
									500	Paired fixed			
w^+	0.3175	0.049	0.6711	0.9624	0.834	0	−1.05	0	500	Standard	Zel'dovich	512	0
w^-	0.3175	0.049	0.6711	0.9624	0.834	0	−0.95	0	500	Standard	Zel'dovich	512	0

Table 1
(Continued)

Name	$\Omega_{\rm m}$	$\Omega_{\rm b}$	h	$n_{\rm s}$	σ_8	M_ν(eV)	w	$\delta_{\rm b}$	Realizations	Simulations	ICs	$N_c^{1/3}$	$N_\nu^{1/3}$
DC$^+$	0.3175	0.049	0.6711	0.9624	0.834	0	−0.95	0.035	500	Standard	2LPT	512	0
DC$^-$	0.3175	0.049	0.6711	0.9624	0.834	0	−0.95	−0.035	500	Standard	2LPT	512	0
LH	[0.1, 0.5]	[0.03, 0.07]	[0.5, 0.9]	[0.8, 1.2]	[0.6, 1.0]	0	−1	0	2000	Standard	2LPT	512	0
									2000	Fixed		512	
									2000	Standard		1024	
LHνw	[0.1, 0.5]	[0.03, 0.07]	[0.5, 0.9]	[0.8, 1.2]	[0.6, 1.0]	[0, 1]	[−1.3, −0.7]	0	5000	Standard	Zel'dovich	512	512
Total	…	…	…	…	…	…	…	…	44,100	…	…	…	…
	…	…	…	…	…	…	…	…	…	…	…	19,811	10,240

Note. The simulations in the first block have been designed to quantify the information content of cosmological observables. They have a large number of realizations for a fiducial cosmology (needed to estimate the covariance matrix) and simulations varying just one cosmological parameter at a time (needed to compute numerical derivatives). The simulations in the second block arise from Latin hypercubes expanding a large volume in parameter space. The ICs of all simulations were generated at $z = 127$ using 2LPT, except for the simulations with massive neutrinos, where we used the Zel'dovich approximation. All simulations have a volume of $1(h^{-1}\,{\rm Gpc})^3$, and we have three different resolutions: low (256^3 particles), fiducial (512^3 particles), and high (1024^3 particles). In the simulations with massive neutrinos, we assume three degenerate neutrino masses. Simulations have been run with the TreePM+SPH GADGET-III code. We save snapshots at redshifts 3, 2, 1, 0.5, and 0. The parameter δ_b stands for background overdensity, and it only applies to the separate universe simulations.

where $D(z)$ is the growth factor at redshift z, f is the growth rate, and $\gamma \simeq 0.6$ for ΛCDM. From the input matter power spectrum and transfer functions, we compute displacements and peculiar velocities employing the Zel'dovich approximation (Zel'dovich 1970; for cosmologies with massive neutrinos) or second-order perturbation theory (for cosmologies with massless neutrinos). The displacements and peculiar velocities are then assigned to particles that are initially laid on a regular grid. In models with massive neutrinos, we use two different grids that are offset by half a grid size: one grid for CDM and one grid for neutrinos. For 2LPT, we use the code in https://cosmo.nyu.edu/roman/2LPT/, while for neutrinos, we use a modified version of N-GenIC, publicly available at https://github.com/franciscovillaescusa/N-GenIC_growth. The rescaling code used for massive neutrino cosmologies is publicly available in https://github.com/matteozennaro/reps.

All simulations have a cosmological volume of $1(h^{-1}\,\mathrm{Gpc})^3$. The majority of the simulations follow the evolution of 512^3 CDM particles (plus 512^3 for simulations with massive neutrinos): our fiducial resolution. However, we also have simulations with 256^3 (low resolution) and 1024^3 (high resolution) CDM particles. The gravitational softening length is set to 1/40 of the mean interparticle distance, i.e., 100, 50, and 25 h^{-1} kpc for the low-, fiducial-, and high-resolution simulations, respectively. The gravitational softening is the same for CDM and neutrino particles. We save snapshots at redshifts 0, 0.5, 1, 2, and 3. We also save the ICs and the scripts to generate them.

Table 1 summarizes the main features of all of the QUIJOTE simulations.

2.1. Simulations with Massive Neutrinos

In simulations with massive neutrinos, we use the traditional particle-based method (Brandbyge et al. 2008; Viel et al. 2010) to model the cosmic neutrino background. In that method, neutrinos are described as a collisionless and pressureless fluid that is discretized into a set of neutrino particles. Those particles are assigned thermal velocities (on top of peculiar velocities) that are randomly drawn from their Fermi–Dirac distribution at the simulation starting redshift.

One of the well-known problems of this method is that a significant fraction of the neutrino particles will cross the simulation box several times (due to their large thermal velocities). This will erase the clustering of neutrinos on small scales, producing a white power spectrum (or shot noise). This effect is, however, negligible on most of the observational quantities, e.g., the total matter power spectrum and the halo/galaxy power spectrum.

New methods have been developed to address this problem (see, e.g., Banerjee et al. 2018). The 5000 simulations of the LHνw Latin hypercube have been run using this method, which provides a neutrino density field with a negligible level of shot noise.

2.2. Paired Fixed Simulations

The QUIJOTE simulations contain (a) standard, (b) fixed, and (c) paired fixed simulations. The difference between those is the way the ICs are generated. Consider a Fourier-space mode, $\delta(\boldsymbol{k})$. Since it is in general a complex number, we can write it as $\delta(\boldsymbol{k}) = Ae^{i\theta}$, where both the amplitude A and the phase θ depend on the considered wavenumber $\boldsymbol{k}$. For Gaussian density fields, A follows a Rayleigh distribution and θ is drawn from a uniform distribution between zero and 2π. This is the standard way to generate ICs for cosmological simulations. In fixed simulations, while θ is still drawn from a uniform distribution between zero and 2π, the value of A is fixed to the square root of the variance of the previous Rayleigh distribution. Finally, paired fixed simulations are two fixed simulations where the phases of the two pairs differ by π. We refer the reader to, Angulo & Pontzen (2016), Pontzen et al. (2016), and Villaescusa-Navarro et al. (2018) for further details.

Fixed and paired fixed simulations have received a lot of attention recently, since it has been shown that they can significantly reduce the amplitude of cosmic variance on different statistics (e.g., the power spectrum) without inducing a bias on the results (Angulo & Pontzen 2016; Pontzen et al. 2016; Anderson et al. 2018; Villaescusa-Navarro et al. 2018; Chuang et al. 2019; Klypin et al. 2020). While these simulations cannot be used to estimate covariance matrices, they may be useful to compute numerical derivatives or provide an effective larger cosmological volume. For this reason, some of the simulations we have run are fixed and paired fixed.

2.3. Separate Universe Simulations

In a conventional simulation, it is assumed that the mean density in the box is equal to that of the whole observable universe. So all the above simulations have a mean background overdensity, δ_b, equal to 0. This effectively truncates the sampling of matter power spectrum at the box scale. In reality, it is expected that regions of $1(h^{-1}\mathrm{Gpc})^3$, as the ones simulated by the Quijote suite, will have background densities different to 0, due to the long-wavelength modes beyond the box size.

The non-vanishing δ_b modulates the local growth factor and the expansion rate, known as the growth and dilation effects (Li et al. 2014, 2018), respectively. As an example, a positive δ_b enhances the growth of structures and, at the same time, slows down the expansion of the local region ascompared to the global universe. Together, the growth and dilation effects are known as the super-sample effect, given their origin from the long-wavelength modes beyond the sampled (simulation or survey) region. To be able to quantify the impact of the super-sample effect, we have run a set of 1000 simulations with non-vanishing δ_b, using atechnique called the separate universe simulation.

In a ΛCDM universe, there is a well-known symmetry that allows one to absorb the mean overdensity into a redefinition of cosmology with curvature (Sirko 2005). This allows us to use existing code to perform the simulations while only modifying their cosmological parameters. In particular, we care about the linear response to δ_b, which we can measure with central difference of a pair of separate universe simulations. We have run two sets of 500 simulations each, with the fiducial cosmology and $\delta_b = \pm 0.035$ and refer the readers to Li et al (2014) for more details on the setup of the paired simulations.

An import consequence of the super-sample effect is that it introduces additional covariance of generic observables, called the super-sample covariance (Takada & Hu 2013; Li et al 2018; Philcox 2020). This extra covariance arises from the coherent modulation of the long modes on all observables within the local region, and the unknown nature of δ_b. Intuitively, the super-sample covariance between two observables O_1 and O_2 is $C^{\mathrm{SSC}}_{o_1 o_2} = \sigma_b^2 \frac{\mathrm{d}O_1}{\mathrm{d}\delta_b}\frac{\mathrm{d}O_2}{\mathrm{d}\delta_b}$, where σ_b^2 is the variance of δb, and the other two terms are the linear response of the observables. This allows us to evaluate the impact of the super-sample covariance using the Fisher matrix formalism.

2.4. Fiducial Cosmology

The values of the cosmological parameters for our fiducial model are $\Omega_{\rm m} = 0.3175$, $\Omega_{\rm b} = 0.049$, $h = 0.6711$, $n_s = 0.9624$, $\sigma_8 = 0.834$, $M_\nu = 0.0\,{\rm eV}$, and $w = -1$. The values of those parameters are in good agreement with the latest constraints by Planck (Planck Collaboration et al. 2018).

For this model, we have run a total of 17,100 simulations. Of those, 15,000 are standard simulations run at the fiducial resolution with 2LPT ICs. The main purpose of these simulations is to compute covariance matrices. We also have a set of 500 paired fixed simulations with 2LPT ICs at the fiducial resolution that can be used to study the properties of paired fixed simulations and compute numerical derivatives.

Furthermore, we have a set of 500 standard simulations with Zel'dovich ICs at the fiducial resolution needed to compute the derivatives with respect to neutrino masses (see Section 5.1). Finally, a set of 1000 standard simulations at low resolution and 100 standard simulations at high resolution are available to carry out resolution tests and apply superresolution techniques (see Section 4.6).

2.5. Simulations for Numerical Derivatives

One of the ingredients needed to quantify the information content of a statistic is the partial derivatives of it with respect to the cosmological parameters (see Section 4.1). For $\Omega_{\rm m}$, $\Omega_{\rm b}$, h, n_s, σ_8, and w, we compute the partial derivatives as

$$\frac{\partial \boldsymbol{S}}{\partial \theta} \simeq \frac{\boldsymbol{S}(\theta + d\theta) - \boldsymbol{S}(\theta - d\theta)}{2d\theta}, \tag{2}$$

where $\boldsymbol{S}$ is the considered statistic (e.g., the matter power spectrum at different wavenumbers), and θ is the cosmological parameter. We thus need to evaluate the statistic on simulations where only the considered parameter is varied above and below its fiducial value. In order to fulfill this requirement, we have run simulations varying only one cosmological parameter at a time. For instance, the simulations coined $\Omega_{\rm m}^+/\Omega_{\rm m}^-$ have the same value of $\Omega_{\rm b}$, h, n_s, σ_8, M_ν, and w as the fiducial model, but the value of $\Omega_{\rm m}$ is slightly larger/smaller. In this case, $d\Omega_{\rm m}/\Omega_{\rm m} \simeq 1.8\%$: $\Omega_{\rm m} = 0.3275$ for $\Omega_{\rm m}^+$ and $\Omega_{\rm m} = 0.3075$ for $\Omega_{\rm m}^-$.

In the simulations $\Omega_{\rm b}^{++}$ and $\Omega_{\rm b}^{--}$, we vary $\Omega_{\rm b}$ by $d\Omega_{\rm b}/\Omega_{\rm b} \simeq 4\%$. When varying the other parameters, h, n_s, σ_8, and w, we have $dh/h \simeq 3\%$, $dn_s/n_s \simeq 2\%$, $d\sigma_8/\sigma_8 \simeq 1.8\%$, and $dw/w = 5\%$, respectively. These numbers were chosen so that the difference is small enough to approximate the derivatives but not too small to be dominated by numerical noise. In the $\Omega_{\rm b}^+$ and $\Omega_{\rm b}^-$ simulations, we have $d\Omega_{\rm b}/\Omega_{\rm b} \simeq 2\%$. For most of the statistics we have considered, this difference is too small, and the derivatives are slightly affected by numerical noise.

For all of these models, we have run 500 standard simulations and 500 paired fixed simulations using 2LPT at the fiducial resolution. The exception is the models with $w \neq -1$, where we only run 500 standard simulations and $\Omega_{\rm b}^+/\Omega_{\rm b}^-$ that only have 500 paired fixed simulations.

To compute numerical derivatives with respect to massive neutrinos, we cannot use Equation (2), since the second term in the numerator will correspond to a universe with negative neutrino masses.[41] For this reason, we have run simulations at several values of the neutrino masses: $M_\nu^+ = 0.1$ eV, $M_\nu^{++} = 0.2$ eV, and $M_\nu^{+++} = 0.4$ eV. From these simulations, several derivatives can be computed:

$$\frac{\partial \boldsymbol{S}}{\partial M_\nu} \simeq \frac{\boldsymbol{S}(M_\nu) - \boldsymbol{S}(M_\nu = 0)}{M_\nu},$$
$$\frac{\partial \boldsymbol{S}}{\partial M_\nu} \simeq \frac{-\boldsymbol{S}(2M_\nu) + 4\boldsymbol{S}(M_\nu) - 3\boldsymbol{S}(M_\nu = 0)}{2M_\nu},$$
$$\frac{\partial \boldsymbol{S}}{\partial M_\nu} \simeq \frac{\boldsymbol{S}(4M_\nu) - 12\boldsymbol{S}(2M_\nu) + 32\boldsymbol{S}(M_\nu) - 21\boldsymbol{S}(M_\nu = 0)}{12M_\nu},$$

where the first equation can be used for $M_\nu = 0.1$, 0.2, or 0.4 eV. The second equation can instead be evaluated with $M_\nu = 0.1$ or 0.2 eV, while the last equation requires $M_\nu = 0.1$ eV. Notice that if the differences between the fiducial model and the cosmology with 0.1 eV neutrinos are not dominated by noise, the last equation will provide the most precise estimation of the derivative. In some cases, e.g., with the halo mass function, differences between the fiducial model with massless neutrinos and cosmology with 0.1 eV neutrinos is too small, and therefore dominated by noise. In these cases, it is recommended to use the above second equation with $M_\nu = 0.2$ eV.

For the models with 0.1, 0.2, and 0.4 eV, we have run 500 standard and 500 paired fixed simulations at the fiducial resolution. As stated above, for models with massive neutrinos, the ICs have been generated using the Zel'dovich approximation.

The top panel of Figure 1 shows the spatial distribution of matter in a realization of the fiducial cosmology. The other panels show the derivative of the density field of that particular realization with respect to the parameters $\Omega_{\rm m}$, $\Omega_{\rm b}$, h, n_s, σ_8, and M_ν. It can be seen how the morphology of the density field responds differently to each cosmological parameter. Neural networks are trained to identify these changes and constrain parameter values using them.

2.6. Latin Hypercubes

Besides the simulations described above, we have also run a set of 11,000 simulations on different Latin hypercubes. The main purpose of these simulations is to provide enough data to train machine-learning algorithms. In Section 4 we outline some applications of these simulations.

The simulations can be split into two main sets. In the first one, called LH, we use a Latin hypercube where we vary the value of $\Omega_{\rm m}$ between 0.1 and 0.5, $\Omega_{\rm b}$ between 0.03 and 0.07, h between 0.5 and 0.9, n_s between 0.8 and 1.2, and σ_8 between 0.6 and 1.0, and we keep M_ν fixed to 0.0 eV and w to -1. The LH set is made up of three different sets, but the value of the cosmological parameters is the same among the three sets. The first one is made up of standard simulations with different random seeds. The second one is made up of fixed simulations, all of them having the same random seed. Those two sets have been run at the fiducial resolution. The last set is made up of standard simulations with different random seeds but at high resolution: 1024^3 CDM particles. The ICs of all of these simulations were generated with 2LPT. The three different sets contain 2000 simulations each. The set with fixed simulations can be used to create an accurate emulator, while the other two sets can be used to train machine-learning algorithms accounting for the presence of cosmic variance.

The second Latin hypercube, called LHνw, is a set of 5000 standard simulations with different random seeds where the value of the cosmological parameters is changed within $\Omega_{\rm m} \in [0.1, 0.5]$, $\Omega_{\rm b} \in [0.03, 0.07]$, $h \in [0.5, 0.9]$, $n_s \in [0.8, 1.2]$, $\sigma_8 \in [0.6, 1.0]$, $M_\nu \in [0, 1]$, and $w \in [-1.3, -0.7]$. Since these

[41] Notice that our fiducial cosmology is for a universe with massless neutrinos.

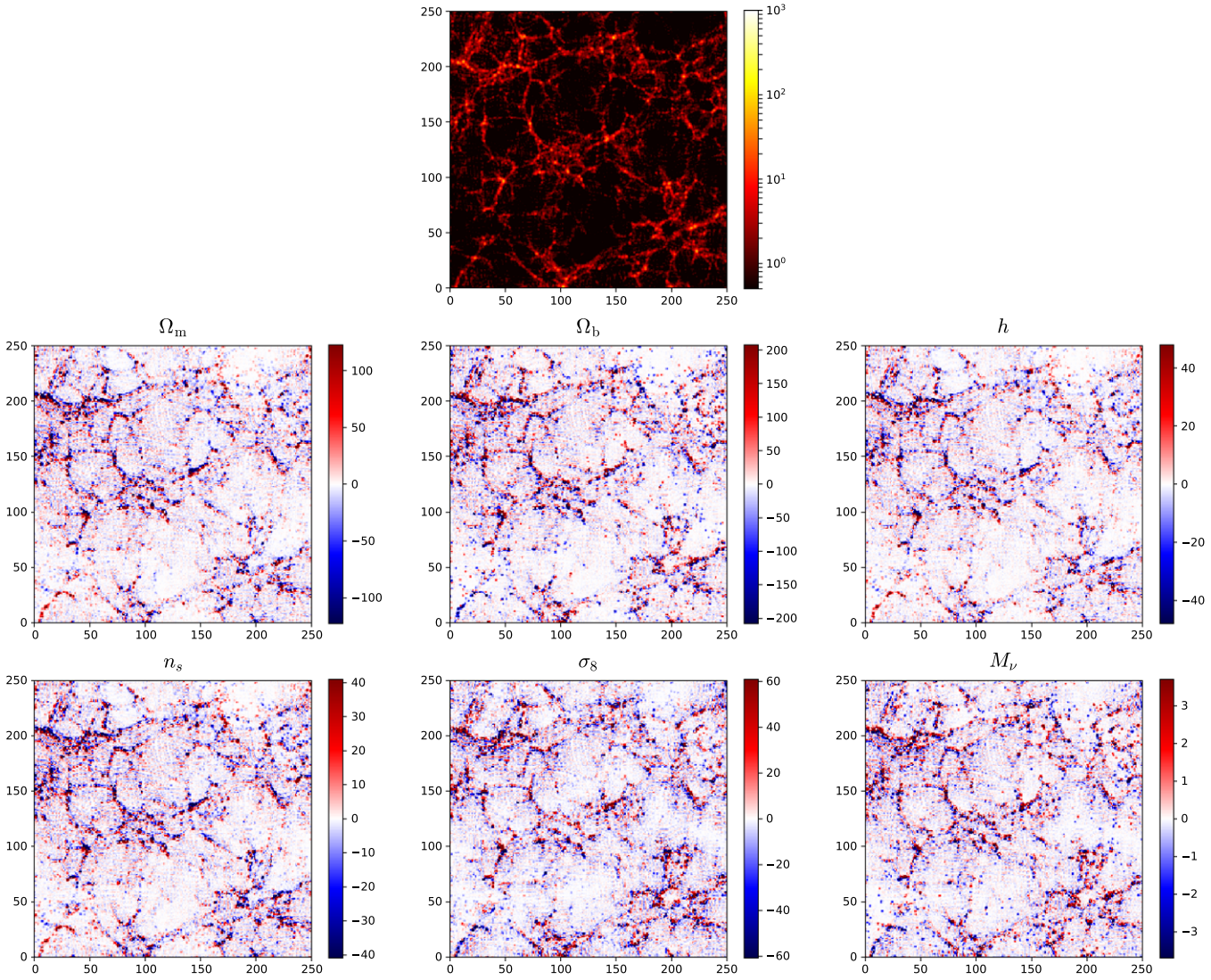

Figure 1. The image on the top shows the large-scale structure in a region of $250 \times 250 \times 15(h^{-1}\,\mathrm{Mpc})^3$ at $z = 0$ for the fiducial cosmology. We have taken simulations with the same random seed but different values of just one parameter and used them to compute the derivative of the density field with respect to the parameters. The panels in the middle and bottom rows show those derivatives with respect to $\Omega_{\rm m}$ (middle left), $\Omega_{\rm b}$ (middle center), h (middle right), n_s (bottom left), σ_8 (bottom center), and M_ν (bottom right). It can be seen how different parameters affect the large-scale structure in different manners. For instance, the filament on the bottom left part of the plot responses differently to each parameter. Neural networks can use these features to extract information from nonstandard summary statistics.

simulations contain massive neutrinos, the ICs were generated using the Zel'dovich approximation. All of the simulations follow the evolution of 512^3 CDM particles plus 512^3 neutrino particles. Each simulation is run with a different random seed.

Figure 2 shows the spatial distribution of matter in six different cosmological models of the high-resolution LH simulations at $z = 0$. Different features show up in the different images, from very long and thick filaments to highly clustered structures. This reflects the broad range covered by the QUIJOTE simulations in the parameter space, from realistic models to extreme scenarios.

3. Data Products

In this section, we describe the data products of the QUIJOTE simulations.

3.1. Snapshots

We provide access to the full snapshots of the simulations at redshifts 0, 0.5, 1, 2, and 3 and the ICs at $z = 127$. The snapshots only have four different fields: (1) header, (2) positions, (3) velocities, and (4) IDs.

The header contains information about the snapshot, such as redshift, value of $\Omega_{\rm m}$, Ω_Λ, number of particles, number of files, etc. The position block stores the positions of the particles in comoving h^{-1} kpc. The velocities of the particles are in the velocities block, while the IDs block hosts the unique IDs of the particles. The positions and velocities are saved as 32-bit floats, while the IDs are 32-bit integers. The snapshots are stored in either GADGET-II or hdf5 format. PYLIANS[42] can be used to read the snapshots independently of the format.

3.2. Halo Catalogs

We save halo catalogs at each redshift for all of the simulations, a total of 215,500 halo catalogs. Halos are identified using the friends-of-friends (FoF) algorithm (Davis et al. 1985). We set the value of the linking length parameter to

[42] https://github.com/franciscovillaescusa/Pylians

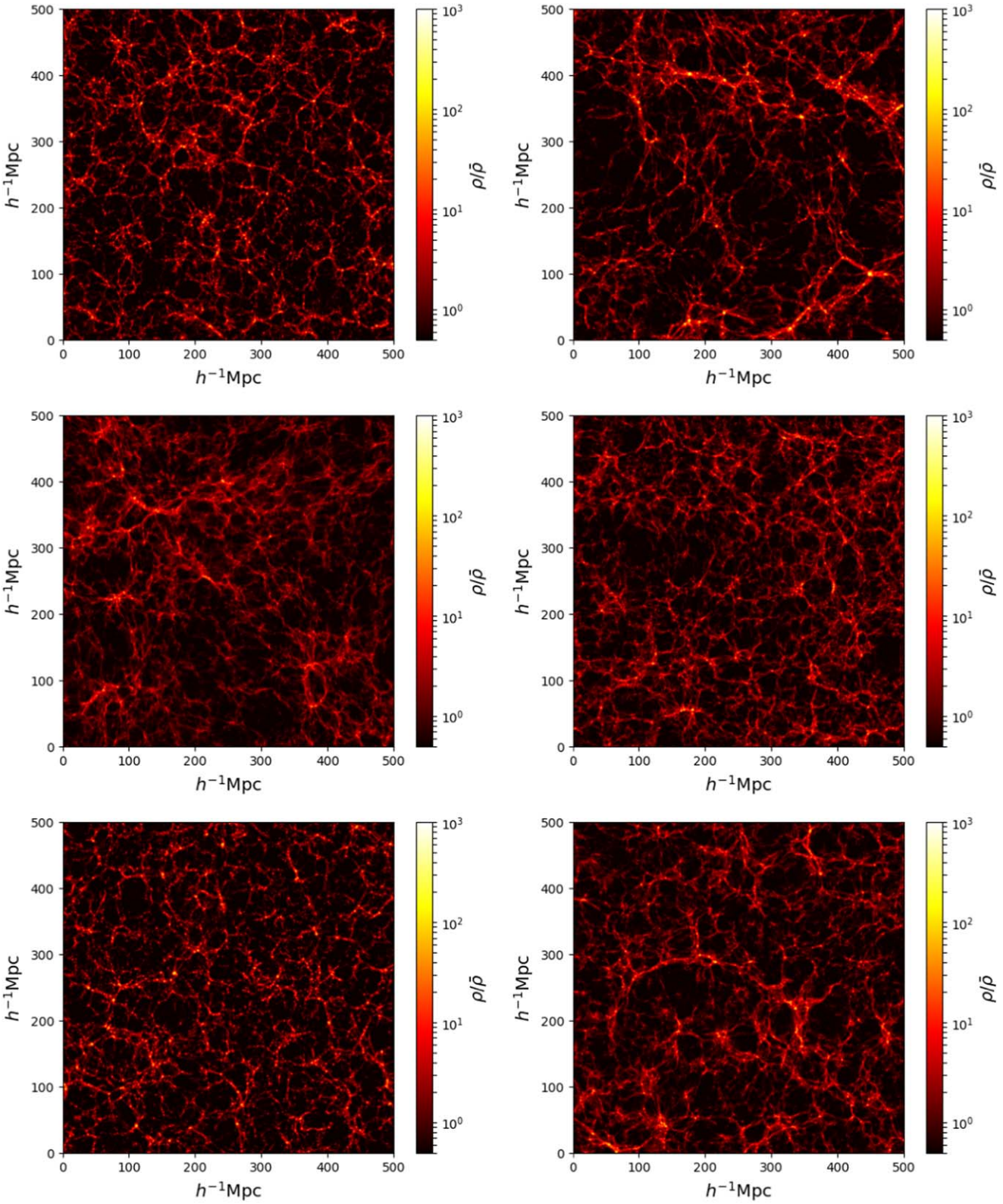

Figure 2. Projected density field of a region of $500 \times 500 \times 10(h^{-1}\,\mathrm{Mpc})^3$ from six different cosmologies of the high-resolution LH simulations at $z = 0$. The top left panel corresponds to a model close to Planck: $\{\Omega_\mathrm{m}, \Omega_\mathrm{b}, h, n_\mathrm{s}, \sigma_8\} = \{0.3223, 0.04625, 0.7015, 0.9607, 0.8311\}$. The other panels represent cosmologies with $\{0.1005, 0.04189, 0.5133, 1.0107, 0.9421\}$ (top right), $\{0.1029, 0.06613, 0.7767, 1.0115, 0.6411\}$ (middle left), $\{0.1645, 0.05257, 0.7743, 1.0311, 0.6149\}$ (middle right), $\{0.4487, 0.03545, 0.5167, 1.0387, 0.9291\}$ (bottom left), and $\{0.1867, 0.04503, 0.6189, 0.8307, 0.7187\}$ (bottom right). These images show the wide range in parameter variations the Quijote simulations cover; some cosmologies exhibit very thick and/or long filaments, while others exhibit unusual clustering patterns.

$b = 0.2$. Each halo catalog contains the positions, velocities, masses, and total number of particles of each halo. Only CDM particles are linked in the FoF halos, as the contribution of neutrinos to the total halo mass is expected to be negligible (Villaescusa-Navarro et al. 2011, 2013; Ichiki & Takada 2012; LoVerde & Zaldarriaga 2014). For simulations with large

neutrino masses (e.g., the simulations in the LHνw), we also provide halo catalogs with the mass of halos being CDM plus neutrinos. The halo catalogs are saved in a binary format. PYLIANS can be used to read the catalogs. We only save halos that contain at least 20 CDM particles. We have not carried out an unbinding step to remove spurious halos. We thus caution the user to take this into account when using FoF halos with a low number of particles.

For a subset of the simulations, we have also identified halos using the AMIGA halo finder (Knollmann & Knebe 2009).

3.3. Void Catalogs

We provide void catalogs from every simulation at each redshift. For simulations with massive neutrinos, we provide two void catalogs: one in which the voids were identified using the total matter field, and one in which the voids were identified in only the CDM + baryon field. For cosmologies with massless neutrinos, we only provide the latter. More than 250,000 void catalogs are thus provided by the QUIJOTE simulations.

Voids are identified in the simulations using the void finder used in Banerjee & Dalal (2016). The algorithm is as follows. First, the relevant overdensity field (CDM or CDM+neutrinos) is computed on a regular grid. This overdensity field is then smoothed on some scale $R_{\rm smooth}$ using a top-hat filter. All voxels at which the value of the smoothed overdensity field is below some threshold $\delta_{\rm threshold}$ are stored. Note that the initial $R_{\rm smooth}$ is chosen to be quite large ($\sim$100 h^{-1} Mpc). The grid voxels are then sorted in order of increasing overdensity, and the voxel with the lowest overdensity (or most underdense) is labeled as a void center with void radius $R_{\rm smooth}$. Since we use spherical top-hat smoothing, we can also associate a mass with the void: $M_v = 4/3\pi R^3_{\rm smooth}\bar{\rho}(1+\delta_{\rm threshold})$. We also tag all voxels within radius $R_{\rm smooth}$ so that they cannot later be labeled as void centers. We then work down the list of points that crossed the threshold, i.e., to higher overdensities (less underdense), identifying them as new void centers with radius $R_{\rm smooth}$ if they do not overlap with previously identified voids.

Once all voxels below threshold for a given $R_{\rm smooth}$ have been checked, we move to a smaller value of $R_{\rm smooth}$ and repeat the procedure outlined above. In this way, the largest voids are the first identified, and then progressively smaller voids are found and stored in the void catalog. Note that by this definition, we do not have nested void regions in the provided void catalogs.

By default, our void find was run using $\delta_{\rm threshold} = -0.7$, but we also provide void catalogs with different values of $\delta_{\rm threshold}$, such as -0.5 or -0.3. Our void catalogs contain the positions, radii, and void size functions (number density of voids per unit of radius). The void catalogs are stored in HDF5 files.

3.4. Power Spectra

We compute power spectra for (1) the total matter field, (2) CDM + baryons (only for simulations with massive neutrinos), and (3) halos with different masses. The power spectra are computed for all simulations at the redshifts 0, 0.5, 1, 2, and 3 and also at $z = 127$ for (1) and (2). We compute the power spectra in both real and redshift space. In redshift space, we place the redshift-space distortions along one Cartesian axis and compute the monopole, quadrupole, and hexadecapole. We repeat the procedure for the three Cartesian axes; i.e., in redshift space, we compute three power spectra instead of one. In total, the QUIJOTE simulations contain over 1 million power spectra.

3.5. Marked Power Spectra

Marked correlation functions are special two-point statistics where correlations are weighted according to a mark, e.g., some environmental property. They have been shown to be interesting tools to study the galaxy clustering dependence on galaxy properties such as morphology, luminosity, etc. (Beisbart & Kerscher 2000; Sheth et al. 2005); halo clustering dependence on merger history (Gottloeber et al. 2002); modified theories of gravity (White 2016; Armijo et al. 2018; Hernández-Aguayo et al. 2018; Valogiannis & Bean 2018); and neutrinos' masses (Massara et al. 2020).

We compute the marked power spectra of the matter density field, which are the Fourier counterparts of marked correlations. Inspired by White (2016), we consider the mark $M(\boldsymbol{x})$ of the form

$$M(\boldsymbol{x}) = \left[\frac{1+\delta_s}{1+\delta_s+\delta_R(\boldsymbol{x})}\right]^p, \tag{3}$$

which depends on the local matter density $\delta_R(\boldsymbol{x})$, a parameter δ_s, and an exponent p. The density $\delta_R(\boldsymbol{x})$ is obtained by smoothing the matter density field with a top-hat filter at scale R and can be evaluated at each point in the space $\boldsymbol{x}$. Thus, the mark depends on three parameters: R, p, and δ_s. When $\delta_s \to 0$, $M(\boldsymbol{x}) \to [\bar{\rho}/\rho_R(\boldsymbol{x})]^p$, with $\bar{\rho}$ being the mean matter density of the universe and $\rho_R(\boldsymbol{x})$ the density inside a sphere of radius R around $\boldsymbol{x}$. If $p > 0$, the mark gives more weight (and therefore more importance) to points that are in underdense regions, while if $p < 0$, points in overdensities are weighted more. One can adjust these parameters to obtain different types of marks that can weight the various components of the large-scale structure in different ways.

The marked power spectra are computed as follows. First, the smoothed density field $\delta_R(\boldsymbol{x})$ is calculated on the vertex of a grid; the values of δ_R at the position of each matter particle are then computed via interpolation, and a mark is assigned to each particle. Second, the marked power spectrum is computed as a power spectrum with each particle weighted by its mark.

We consider five different values for each of the three mark parameters: $R = 5$, 10, 15, 20, 30 h^{-1} Mpc, $p = -1$, 0.5, 1, 2, 3, and $\delta_s = 0$, 0.25, 0.5, 0.75, 1, giving a total of 125 different mark models. In total, millions of marked power spectra are available in the QUIJOTE simulations.

3.6. Correlation Functions

We compute correlation functions for (1) the total matter field and (2) the CDM + baryon field (only for simulations with massive neutrinos). The correlation functions are computed at redshifts 0, 0.5, 1, 2, and 3 in both real and redshift space. In the same way as for the power spectrum, redshift-space distortions are placed along one Cartesian axis and three correlation functions are computed, one for each axis.

The procedure we use to compute the correlation functions is as follows. First, we assign particle masses to a regular grid with N^3 cells using the cloud-in-cell (CIC) mass assignment scheme and compute the density contrast: $\delta(\boldsymbol{x}) = \rho(\boldsymbol{x})/\bar{\rho} - 1$. We then Fourier transform the density contrast field to get $\delta(\boldsymbol{k})$.

Next, we compute the modulus of each Fourier mode, $|\delta(\boldsymbol{k})|^2$, and Fourier transform back that field. Finally, we compute the correlation function by averaging modes that fall within a given radius interval. In redshift space, the quadrupole and hexadecapole are computed in the same way as the monopole by weighing each mode by the contribution of the corresponding Bessel function.

By default, we set N to be equal to the cubic root of the number of particles in the simulation, but we also compute correlation functions in finer grids. In total, we provide over 1 million correlation functions.

3.7. Bispectra

We compute bispectra for the total matter field, as well as for halo catalogs, in both real and redshift space at redshifts 0, 0.5, 1, 2, and 3. We use a fast Fourier transform (FFT)–based estimator similar to the estimators described in Sefusatti & Scoccimarro (2005), Scoccimarro (2015), and Sefusatti et al. (2016). We first interpolate matter particles/halos to a grid to compute the density contrast field, $\delta(\boldsymbol{x})$, using a fourth-order interpolation to get interlaced grids and then Fourier transform the grid to get $\delta(\boldsymbol{k})$. We then measure the bispectrum monopole using

$$B_0(k_1, k_2, k_3) = \frac{1}{V_B}\int_{k_1} d^3q_1 \int_{k_2} d^3q_2 \int_{k_3} d^3q_3\, \delta_{\rm D}(\boldsymbol{q}_{123}) \times \delta(\boldsymbol{q}_1)\,\delta(\boldsymbol{q}_2)\,\delta(\boldsymbol{q}_3) - B^{\rm SN}, \quad (4)$$

where $\delta_{\rm D}$ is the Dirac delta function; V_B is a normalization factor proportional to the number of triplets in the triangle bin defined by k_1, k_2, and k_3; and $B^{\rm SN}$ is the correction for Poisson shot noise. To evaluate the integral, we take advantage of the plane-wave representation of $\delta_{\rm D}$. For more details, we refer readers to Hahn et al. (2019).[43] We use $\delta(\boldsymbol{x})$ grids with $N_{\rm grid} = 360$ and triangle configurations defined by k_1, k_2, and k_3 bins of width $\Delta k = 3k_f = 0.01885\ h\ {\rm Mpc}^{-1}$. For $k_{\rm max} = 0.5\ h\ {\rm Mpc}^{-1}$, there are 1898 triangle configurations. Redshift-space distortions are imposed along one Cartesian axis, same as the power spectrum, so we measure three bispectra, one for each axis.

In total, the Quijote simulations provide over 1 million bispectra.

3.8. PDFs

We estimate the probability density functions (PDFs) of the matter, CDM + baryon, and halo fields in all of the simulations at all redshifts. The PDFs are computed as follows. First, we deposit particle masses (or halo positions) onto a regular grid with N^3 cells using the CIC mass assignment scheme. We then smooth that field with a Gaussian filter of radius R. Finally, the PDF is calculated by computing the fraction of cells whose overdensities lie within a given interval. We compute the PDFs for many different values of R. By default, we take N to be the cubic root of the number of CDM particles in the simulation. In total, the Quijote simulations provide more than 1 million PDFs.

[43] The code that we use to evaluate B_0 is publicly available at https://github.com/changhoonhahn/pySpectrum.

4. Applications

The Quijote simulations have been designed to address two main goals: (1) to quantify the information content on cosmological observables and (2) to provide enough statistics to train machine-learning algorithms. In this section, we describe a few examples of applications of the simulations.

4.1. Information Content from Observables

As discussed in the introduction, it is currently unknown what statistic or statistics should be used to retrieve the maximum cosmological information from non-Gaussian density fields. One way to quantify the information content on a set of cosmological parameters, $\boldsymbol{\theta}$, given a statistic $\boldsymbol{S}$, is through the Fisher matrix formalism. The Fisher matrix is defined as

$$F_{ij} = \sum_{\alpha,\beta} \frac{\partial S_\alpha}{\partial \theta_i} C^{-1}_{\alpha\beta} \frac{\partial S_\beta}{\partial \theta_j}, \quad (5)$$

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

$$C_{\alpha\beta} = \langle (S_\alpha - \bar{S}_\alpha)(S_\beta - \bar{S}_\beta)\rangle; \qquad \bar{S}_i = \langle S_i \rangle\ . \quad (6)$$

Notice that in Equation (5) ,we have set to zero the term (see, e.g., Tegmark et al. 1997)

$$\frac{1}{2}{\rm Tr}\left[C^{-1}\frac{\partial C}{\partial \theta_\alpha} C^{-1}\frac{\partial C}{\partial \theta_\beta}\right]. \quad (7)$$

This is because this term is expected to be small (Kodwani et al. 2019), but including it will also lead to an underestimation of the parameter errors (Carron 2013; Alsing & Wandelt 2018).

The error on the parameter θ_i, marginalized over the other parameters, is given by

$$\delta\theta_i \geqslant \sqrt{(F^{-1})_{ii}}\ . \quad (8)$$

Thus, in order to quantify the constraints that a given statistic can place on the value of the cosmological parameters, we only need two ingredients: (1) the covariance matrix of the statistic (s) and (2) the derivatives of the statistic(s) with respect to the cosmological observables. As discussed in detail in Section 2, the Quijote simulations have been designed to numerically evaluate those two pieces.

In this paper, we consider one of the simplest applications of our simulations: the information content on the matter power spectrum. In Figure 3, we plot the correlation matrix of the matter power spectrum at $z = 0$, defined as

$$\frac{C_{k_ik_j}}{\sqrt{C_{k_ik_i}C_{k_jk_j}}} \quad (9)$$

when computed using 100 (left), 1000 (middle), and 15,000 (right) realizations of the fiducial cosmology. As can be seen, the results are noisy when computing the covariance with few realizations; this, in turn, affects the results of the Fisher matrix analysis.

As is well known, on large scales, the different Fourier modes are decoupled, and the covariance matrix is almost diagonal. On small scales, modes with different wavenumbers are coupled, giving rise to nondiagonal elements whose amplitude increases on smaller scales. Notice that previous works have investigated in detail the properties of the

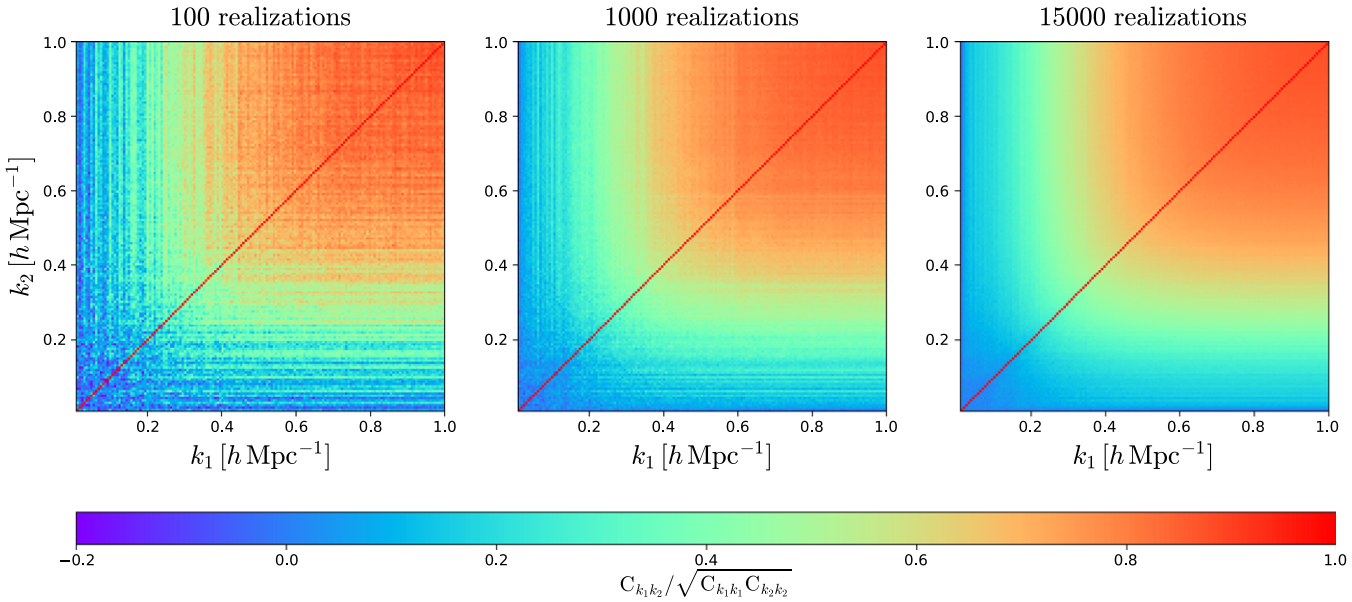

Figure 3. Correlation matrix of the matter power spectrum at $z = 0$ computed using 100 (left), 1000 (middle), and 15,000 (right) realizations. On large scales, modes are decoupled, so correlations are small. On small scales, modes are tightly coupled, and the amplitude of the correlation is high. As expected, the noise in the covariance matrix shrinks with the number of realizations.

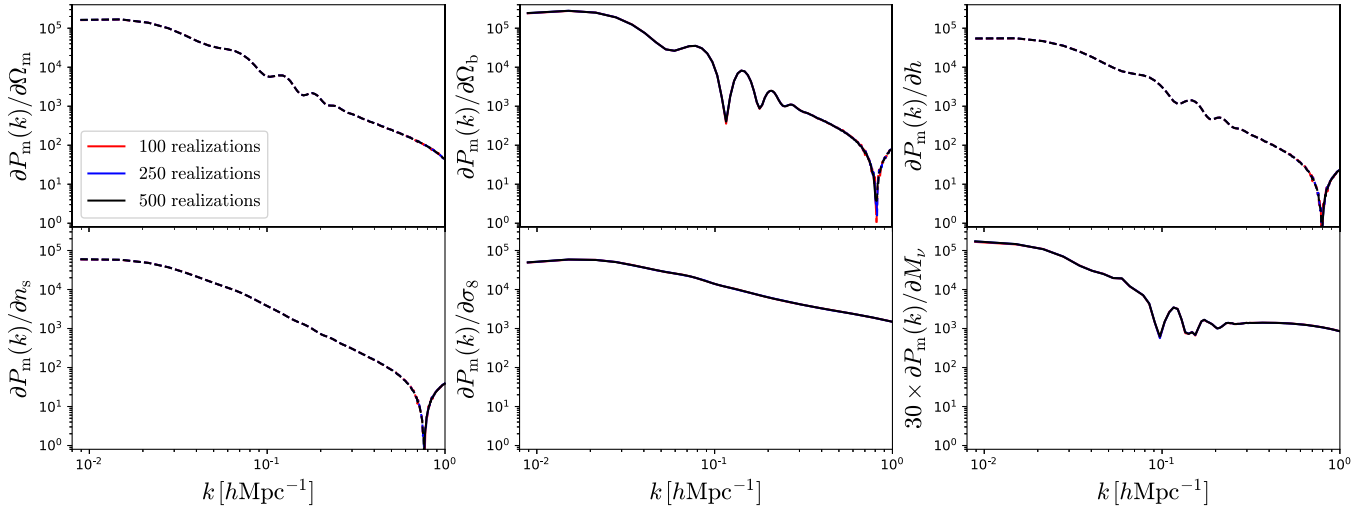

Figure 4. Derivatives of the matter power spectrum in real space with respect to Ω_m (top left), Ω_b (top middle), h (top right), n_s (bottom left), σ_8 (bottom middle), and M_ν (bottom right) at $z = 0$. Solid and dashed lines represent positive and negative values of the derivatives, respectively. We show the derivatives when computed using 100 (red), 250 (blue), and 500 (black) realizations. It can be seen how the results are very converged against the number of realizations.

covariance matrix using a very large set of simulations (Blot et al. 2015, 2016).

In Figure 4, we show the second ingredient we need to evaluate the Fisher matrix: the partial derivatives of the matter power spectrum with respect to the cosmological parameters. In our case, we only consider Ω_m, Ω_b, h, n_s, σ_8, and M_ν and show results at $z = 0$. In that figure, we show the derivatives when computed using a different number of realizations. It can be seen how the results are well converged, all the way to $k = 1\,h\,\mathrm{Mpc}^{-1}$. We can also see how the derivatives are different among the parameters, pointing out that the matter power spectrum alone can provide information on each parameter separately.

With the covariance matrix and the derivatives, we can evaluate the Fisher matrix and determine the constraints on the cosmological parameters. We have verified that our results are converged; i.e., the constraints do not change if the covariance and derivatives are evaluated with fewer realizations. We have also checked that our results are robust against different evaluations of the neutrino derivatives. We show the results in Figure 5 when we consider the matter power spectrum down to $k_\mathrm{max} = 0.1$ (red), 0.2 (blue), and 0.5 (green) $h\,\mathrm{Mpc}^{-1}$. As expected, the smaller the scales, the more cosmological information we can extract and the tighter the constraints on the parameters. However, the gain with scale does not scale proportional to k^3_max, as naively expected just by counting the number of modes. There are two main reasons for this behavior. (1) The covariance becomes nondiagonal on small scales; modes become correlated, and therefore the number of independent modes does not scale as k^3_max. (2) Degeneracies among parameters limit the amount of information that can be extracted.

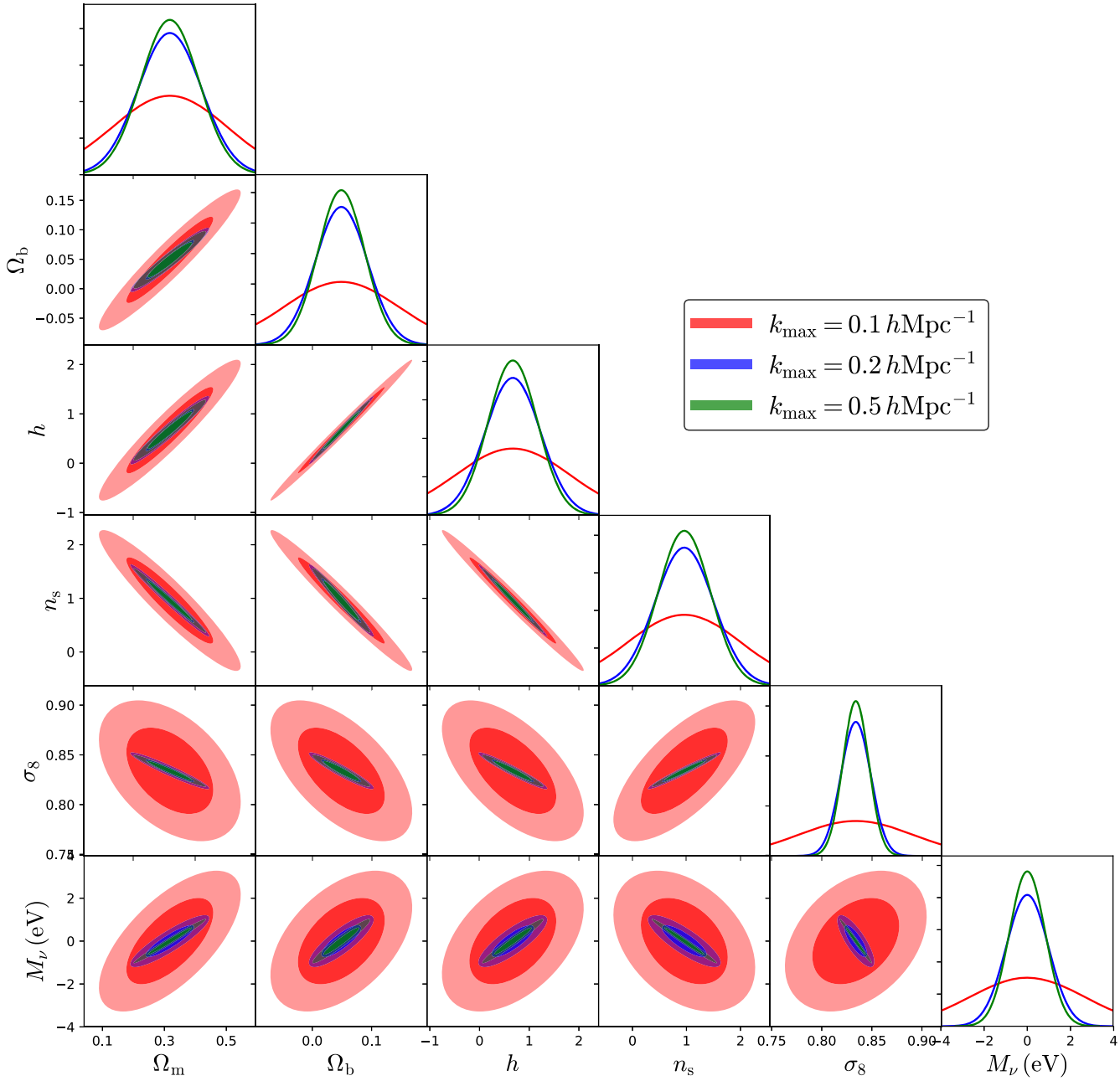

Figure 5. Constraints on the value of the cosmological parameters from the matter power spectrum in real space at $z = 0$ for $k_{\rm max} = 0.2$ (red), 0.5 (blue), and 1.0 (green) h Mpc^{-1}. The small and big ellipses represent the 1σ and 2σ constraints, respectively. The panels with the solid lines represent the probability distribution function of each parameter. As we move to smaller scales, the constraints on the parameters improve. On the other hand, the fact that modes on small scales are highly coupled limits the amount of information that can be extracted from the matter power spectrum by going to smaller scales.

In Figure 6, we show the marginalized 1σ constraints on the value of the cosmological parameters as a function of $k_{\rm max}$. As can be seen, the constraints on the parameters tend to saturate on small scales. We note that this result is mainly driven by degeneracies among parameters rather than the covariance becoming nondiagonal.

The cosmological information that was present on the matter power spectrum at high redshift on small scales has now leaked into other statistics due to nonlinear gravitational evolution. The QUIJOTE simulations can be used to quantify it. The information content on the full halo bispectrum in redshift space is estimated in Hahn et al. (2019). The constraints on the parameters by combining the power spectrum, halo mass function, and void size function are presented in F. Villaescusa-Navarro et al. (2019, in preparation), while the sensitivity of the cosmological parameters to the marked power spectrum is shown in Massara et al. (2020). Uhlemann et al. (2020) quantified the information content on the PDF of the 3D matter field.

4.2. Information Content from Neural Nets

A way of searching for new statistics is using information-maximizing neural networks (IMNNs; Charnock et al. 2018). An IMNN is designed to automatically find informative, nonlinear summaries of the data. The method uses neural networks to transform non-Gaussian data into a set of optimally compressed, Gaussianly distributed summaries via maximization of the Fisher information. These summaries can then be used in a likelihood-free

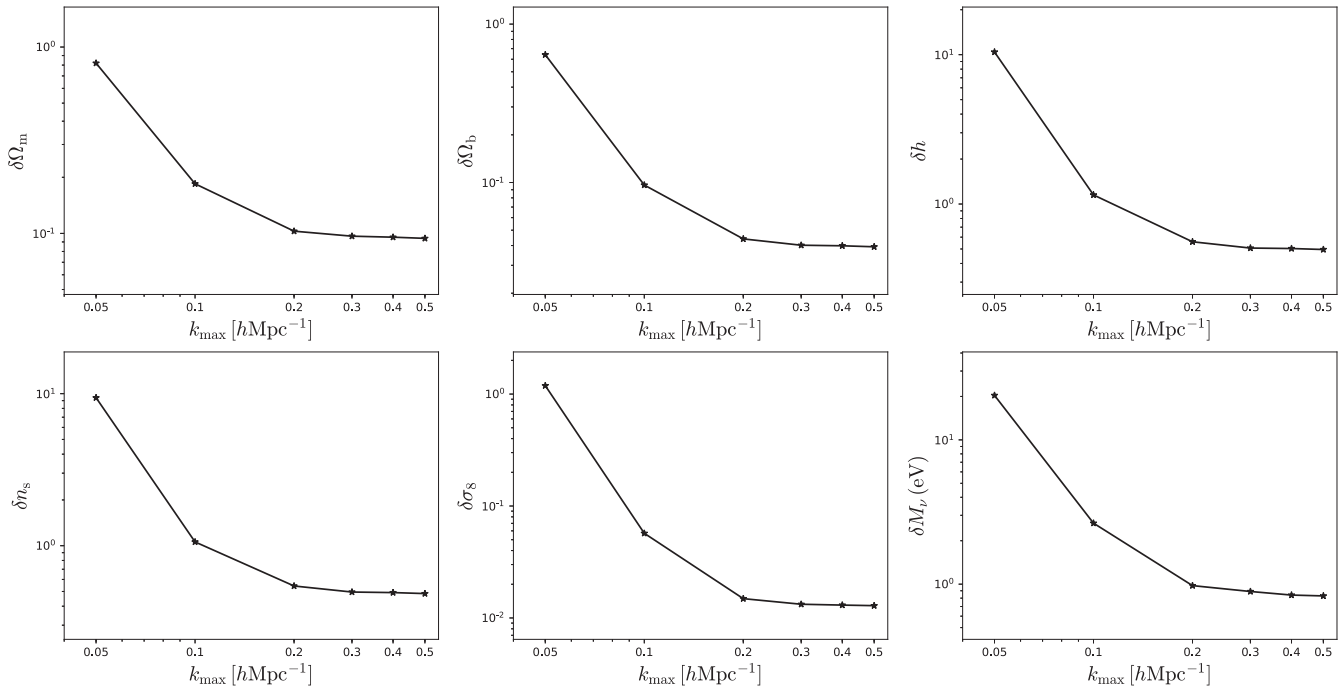

Figure 6. Marginalized 1σ constraints on the value of the cosmological parameters from the matter power spectrum in real space at $z = 0$ as a function of $k_{\rm max}$. As we go to smaller scales, the information content on the different parameters tends to saturate. This effect is mainly driven by degeneracies among the parameters.

inference setting or even directly as pseudo-maximum-likelihood estimators of the parameters. By building neural networks using physically motivated principles, not only will we obtain informative summaries of the data, we will also be able to attribute these summaries to real-space effects, hence learning even more about the connection between data and the underlying cosmological model. As an input, the IMNN requires simulated data to compute the covariance of the summaries and the derivative of the summaries with respect to model parameters. The design of the QUIJOTE simulations enables this novel approach to identify and quantify information content from new observables.

4.3. Likelihood-free Inference

Besides quantifying the information content on cosmological observables, the QUIJOTE simulations have been designed to provide enough data to train machine-learning algorithms. In this subsection, we present a very simple application using a well-known machine-learning algorithm: the random forest.

We use the 2000 standard simulations of the LH Latin hypercube run at fiducial resolution. For each simulation, we compute the 1D PDF when the density field is smoothed on a scale of $5\,h^{-1}$ Mpc using a top-hat filter (see Section 3.8 for further details) at $z = 0$. For each simulation, we thus have an input, the value of the PDF on a set of overdensity bins, and a label, the value of the five cosmological parameters that we vary in those simulations: $\Omega_{\rm m}$, $\Omega_{\rm b}$, h, n_s, and σ_8. Our purpose is to find the function that maps these two vectors, i.e.,

$$\boldsymbol{\theta} = f(\boldsymbol{PDF}(1 + \delta)). \tag{10}$$

The standard way to find the function f is to develop a theoretical model that outputs the PDF for a given value of the cosmological parameters (Uhlemann et al. 2016, 2017, 2018; Gruen et al. 2018). A different approach is to identify features in the data that can be used as a link to the value of the labels. In our case, we search features on the input data using a simple random forest regressor.

We split our data into two sets: (1) a training set with the results of 1600 simulations and (2) a test set with the remaining 400 simulations. We train the random forest algorithm using the input and output of the training set. We then use the trained random forest to predict the value of the cosmological parameters from the PDF of the simulations of the test set. We emphasize that the random forest has never seen the data from the test set; therefore, the output from the test set is a true prediction.

We show the results of this exercise in Figure 7. In each panel, the y-axis represents the prediction of the random forest, while the x-axis is the true value. As can be seen, the random forest learns how to accurately predict the value of σ_8 from the fully nonlinear PDF without need of developing a theory model. Another parameter that the random forest is able to predict is $\Omega_{\rm m}$, although less accurately. However, $\Omega_{\rm b}$, h, and n_s are unconstrained by the random forest; failing to capture the parameter dependence, the random forest regressor minimizes the training loss by outputting values close to the mean of the training set. Notice that it is physically expected that for a volume of 1 $(h^{-1}\,{\rm Gpc})^3$, and only using the 1D PDF at a single smoothing scale, the constraints on those parameters will not be very tight.

It is, however, possible to improve these results by identifying features in the 3D density field, instead of on summary statistics. For instance, A. M. Delgado et al. (2019, in preparation) used convolutional neural networks (CNNs) to identify features that allow constraining the value of the cosmological parameter directly from the 3D density field of the QUIJOTE simulations.

4.4. New Non-Gaussian Statistics

In previous years, it was shown that particular low-variance representations inspired from deep neural networks can efficiently characterize non-Gaussian fields. Based on the multiscale decomposition achieved by the wavelet transform,

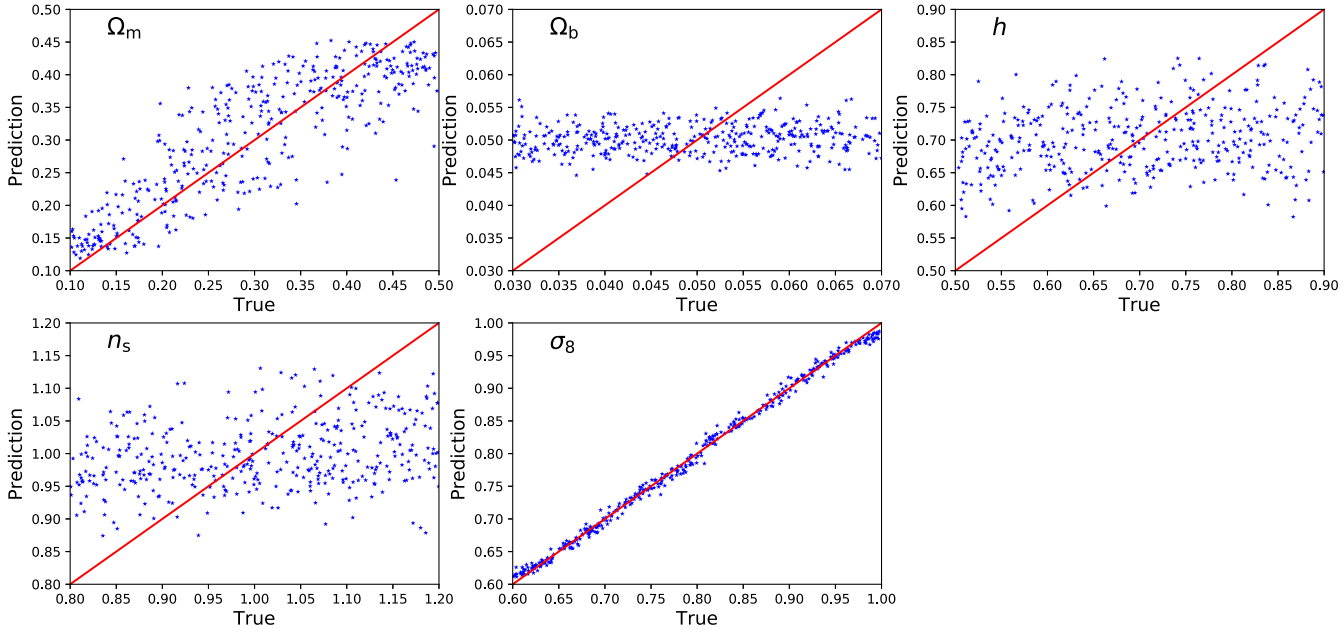

Figure 7. For each of the 2000 standard simulations at fiducial resolution in the LH hypercube, we have measured the one-point PDF of the matter density field smoothed on a scale of 5 h^{-1} Mpc. We have then split the data into two different sets: (1) a training set (1600 simulations) and (2) a test set (400 simulations). We have trained a random forest algorithm to find the mapping between the measured values of the one-point PDF and the value of the cosmological parameters using the training set. Once trained, we have used the test set (that the algorithm has never seen) to see how well we can predict the cosmological parameters from unlabeled PDF measurements. Each panel shows the predicted value vs. the true one for Ω_m (top left), Ω_b (top middle), h (top right), n_s (bottom left), and σ_8 (bottom middle). We find that the random forest can only predict the value of σ_8 and Ω_m from the PDF. We emphasize that no theory model/template has been used to relate the PDF with the value of the parameters.

these representations are built from successive applications of the so-called scattering operator on the field under study (convolution by a wavelet followed by a modulus operator; Mallat 2012) and/or from the phase harmonics of its wavelet coefficients (multiplication of their phase by an integer; Mallat et al. 2018). They can then be analyzed directly, as well as from their covariance matrix, and have obtained state-of-the-art classification results when applied to handwritten and texture discrimination (Bruna & Mallat 2013; Sifre & Mallat 2013).

The use of tailored representations to comprehensively characterize non-Gaussian fields has several advantages with respect to what can be achieved with deep neural networks. Indeed, as the structures of these representations are given and do not necessitate any training stage, they open a path to the interpretability of the results obtained (Allys et al. 2019). For the same reasons, these statistical descriptions can be used even when a large amount of data is not available, since they do not need any training to be constructed. This is illustrated by the ability to synthesize very good-looking synthetic fields from only one given sample; see below.

An important application of these non-Gaussian representations is to model the statistics of cosmological observations, e.g., of a projected density field. This unsupervised learning problem amounts to estimating the probability distribution of such observations, which are stationary, given one or more samples. One can then generate new maps by sampling this distribution. Following standard statistical physics approaches, the probability distribution of Quijote simulations is modeled as a maximum-entropy distribution conditioned by moments (Bruna & Mallat 2018). The main difficulty is to define appropriate moments that are sufficient to capture the statistics of the field. The right panel of Figure 8 was sampled from a Gaussian process, which is a maximum-entropy process conditioned by second-order moments. It thus has the same power spectrum as the original. The middle panel of Figure 8 was sampled from a nearly maximum-entropy distribution conditioned by wavelet harmonic covariance coefficients (Mallat et al. 2018). One can observe from this figure that the image obtained from wavelet harmonic covariances better captures the statistics of the original, including the geometry of high-amplitude outliers and filaments, although it uses fewer moments than the Gaussian model. Indeed, wavelet harmonic moments also depend upon the correlation of phases across scales, which are responsible for the creation of these outliers, whereas Gaussian fields have independent random phases.

A second application of these new statistical descriptors is to infer relevant physical parameters from cosmological observations, which is the goal of ongoing work. Such results would then be compared as a benchmark to the results obtained using standard statistics as, e.g., the power spectrum or the bispectrum. The Quijote simulations, being designed to quantify the information content on cosmological observables, form an ideal set of data for this purpose.

4.5. Forward Modeling and Simulation Emulators

The relatively high resolution[44] and large parameter space of the Quijote simulation suites enable us to build more accurate machine-learning models of the structure formation. He et al. (2019) showed that the highly nonlinear structure formation process, simulated with a particle-mesh (PM) gravity solver (Feng et al. 2016) with fixed cosmological parameters, can be emulated with CNNs. The CNN model is trained to predict the simulation outputs given their ICs (linearly extrapolated to the target redshift). Its accuracy is comparable to that of the

[44] This statement applies in comparison with the simulations used in He et al. (2019).

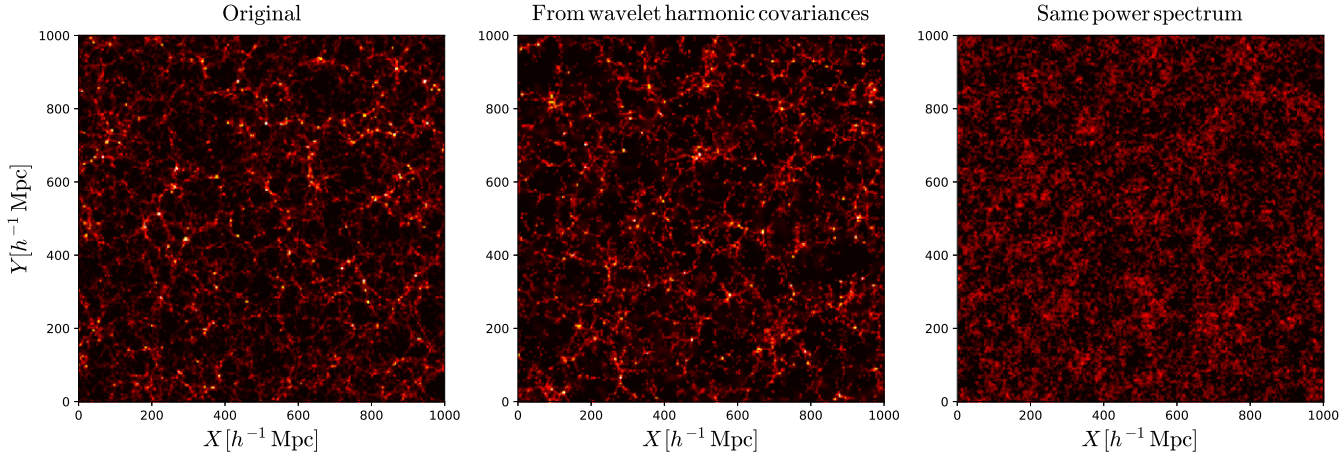

Figure 8. Left: projected density field of a $1000 \times 1000 \times 60(h^{-1}\,\mathrm{Mpc})^3$ region at $z = 0$ for a fiducial cosmology. Middle: simulated field with the same covariance coefficients of wavelet harmonics as the original one (around 1000 coefficients). Right: Gaussian field of the same power spectrum as the original one. All of these images have a resolution of 256×256 pixels. One sees that in contrast to the power spectrum, the phase harmonic coefficients are efficient for extracting statistical features from an image.

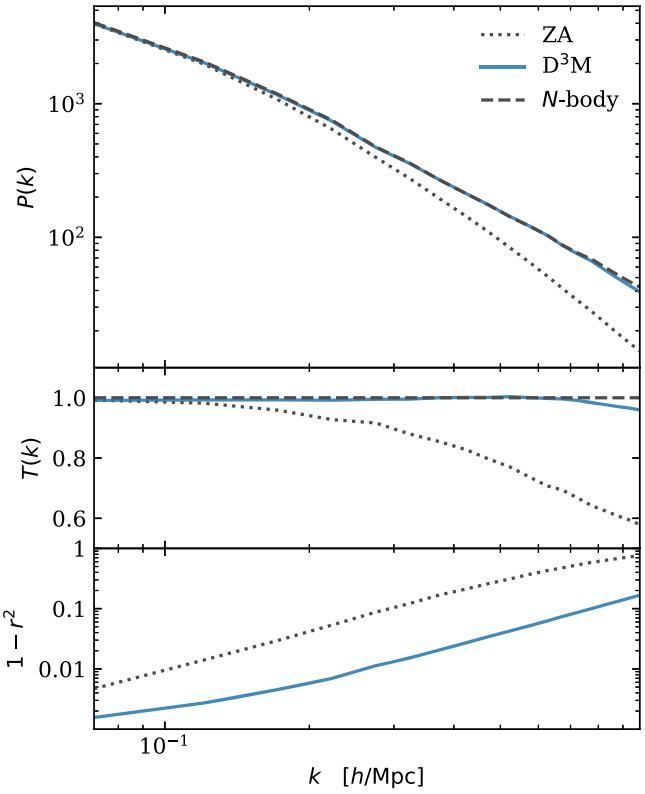

Figure 9. The top panel shows the power spectra $P(k)$ predicted by the Zel'dovich approximation (gray dotted), Quijote simulation (gray dashed), and CNN model dubbed D^{3}M (blue solid). The middle panel shows the transfer function T, defined as the square root of the ratio between the predicted power spectrum and the true one (from the simulations). In the bottom panel, we show the fraction of variance that cannot be explained by each model by the quantity $1 - r^2$, where r is the correlation coefficient between the predicted and true fields. Here T and r capture the quality of the model predictions. As T and r approach 1, the model prediction asymptotes to the ground truth (He et al. 2019). On both benchmarks, the D^{3}M predictions are nearly perfect from linear to nonlinear scales.

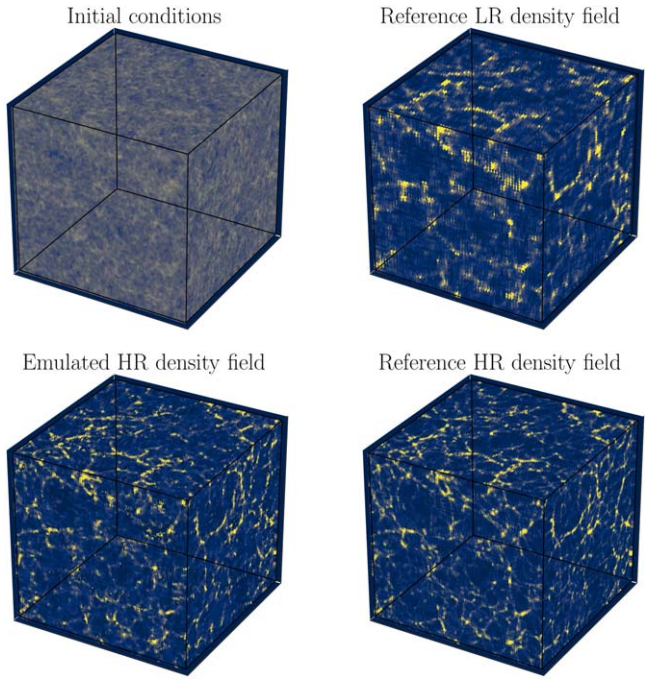

Figure 10. Example of how to increase the resolution of a simulation using deep learning via the superresolution emulator (Kodi Ramanah et al. 2020). We combine the high-resolution ICs (top left) with the $z = 0$ low-resolution snapshot (top right) to emulate a high-resolution snapshot at $z = 0$ (bottom left). The reference high-resolution simulation at $z = 0$ is shown in the bottom right panel for comparison.

training simulations and much more than that of the 2LPT commonly used to generate galaxy mocks (e.g., Scoccimarro & Sheth 2002), while at a much lower computation cost. The gain in both accuracy and efficiency proves that machine learning is a promising forward model of the universe.

The Quijote suite's full N-body simulations resolve gravity to smaller scales than a PM solver. This enables training more accurate CNN models deeper into the nonlinear regime. Figure 9 presents such an example, showing that the machine-learning model from He et al. (2019) can make accurate predictions once trained with the Quijote data.

Furthermore, with the set of Latin hypercube simulations (see Section 2.6), we are able to train CNN models that depend on chosen parameters in addition to the ICs. This allows us to

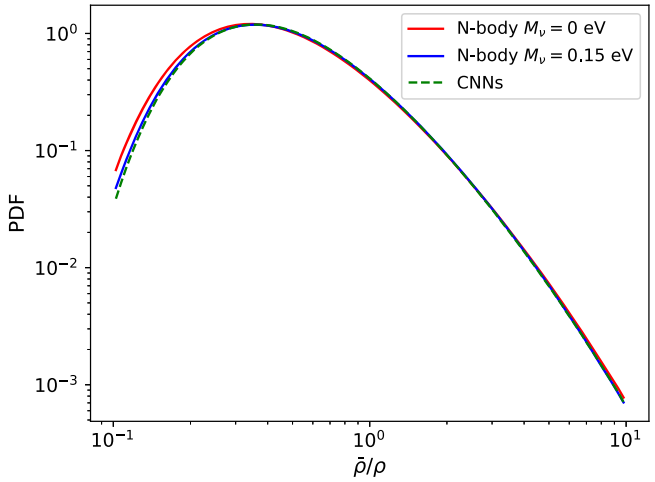

Figure 11. The red and blue lines show the probability distribution function of the CDM + baryon field for a cosmology with massless and massive neutrinos, respectively. We train neural networks to find the mapping between the massless and massive neutrino cosmologies. The dashed green line displays the probability distribution function of the generated CDM + baryon field from the massless neutrino density field, showing a very good agreement with the expected blue line.

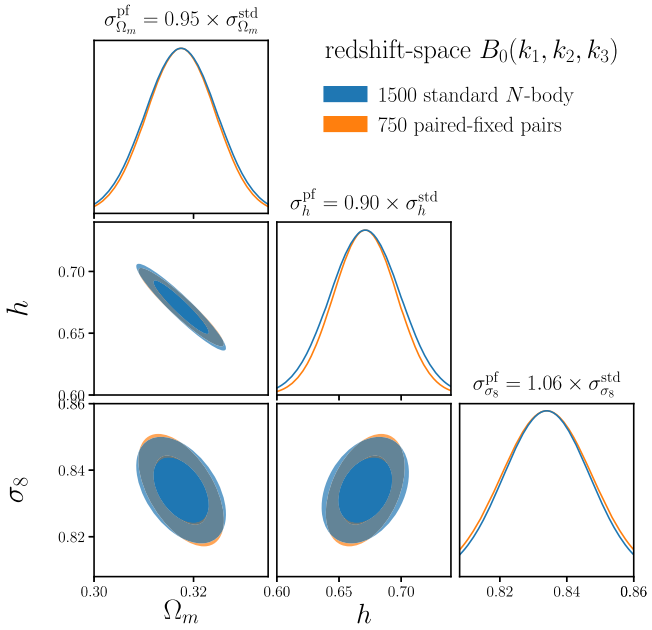

Figure 12. We use the Fisher matrix formalism to quantify how accurately the redshift-space halo bispectrum (down to $k_{\rm max} = 0.5\,h\,{\rm Mpc}^{-1}$) can constrain $\Omega_{\rm m}$, σ_8, and h. In blue and orange, we show the results when the partial derivatives are computed using standard and paired fixed simulations. We find that the results are consistent; therefore, paired fixed simulations do not introduce a significant bias for the halo bispectrum.

build an emulator at the field level. Most of the existing emulators (e.g., Heitmann et al. 2014; Knabenhans et al. 2019; McClintock et al. 2019a, 2019b; Zhai et al. 2019; Nishimichi et al. 2019; Wibking et al. 2019) are aimed mainly at predicting the ensemble-averaged two-point statistics and halo abundance. A CNN model conditional on cosmological parameters will open up the opportunity to fully exploit the information encoded in the higher-order statistics of the field.

4.6. *Superresolution Simulations*

Using the large quantity of high-quality data available with the QUIJOTE simulations, we are able to find methods with which we can accurately paint high-resolution features from computationally cheaper low-resolution simulations (Kodi Ramanah et al. 2020). This superresolution emulator relies on using physically motivated networks (Kodi Ramanah et al. 2019) to perform a mapping of the distribution of the low-resolution cosmological distribution to the space of the high-resolution small-scale structure. Since the information content of the high-resolution simulations is far greater than in the low-resolution simulations, we can use the information contained in the high-resolution ICs as a well-constructed prior from which to draw the data to in-paint the small-scale structure with statistical properties that mimic those of the high-resolution training data. In Figure 10, we show an example of the output of our superresolution emulator and its comparison with the reference high-resolution simulation.

By using this approach, not only do we obtain high-resolution simulations at a low cost, we also are able to inspect the physical network to learn about how the large-scale modes affect the small-scale structure in real space.

4.7. *Mapping between Simulations*

It is possible to use machine-learning algorithms to find the mapping between the positions of particles in simulations with different cosmologies. In this way, from one simulation with a given cosmology, it is possible to get new simulations with different cosmologies. This can be very useful in order to densely sample the parameter space or compute covariance matrices in different regions of the parameter space.

Giusarma et al. (2019) used deep CNNs to establish the link between the displacement field,

$$\boldsymbol{d}_k = \boldsymbol{x}_{f,k} - \boldsymbol{x}_{i,k}, \tag{11}$$

where $\boldsymbol{x}_{f,k}$ and $\boldsymbol{x}_{i,k}$ are the final and initial positions of particle k, in simulations with massless neutrinos and simulations with massive neutrinos (see Zennaro et al. 2019, for other methods to carry out this task). In Figure 11, we show an example of the results for a simple summary statistics: the 1D PDF.

4.8. *Statistical Properties of Paired Fixed Simulations*

The large number of paired fixed simulations available in the QUIJOTE simulations allows one to investigate their statistical properties in detail. These simulations can save a lot of computational resources, since they have been shown to largely reduce the amplitude of cosmic variance on certain statistics. Thus, they can be used to build emulators, evaluate likelihoods, etc.

Hahn et al. (2019) studied the impact of paired fixed simulations on the halo bispectrum and performed a Fisher matrix analysis using both standard and paired fixed simulations to evaluate the derivatives. They quantify how the constraints on the cosmological parameters are affected by using standard versus paired fixed simulations to evaluate the numerical derivatives. We show some results for a subset of the parameters in Figure 12.

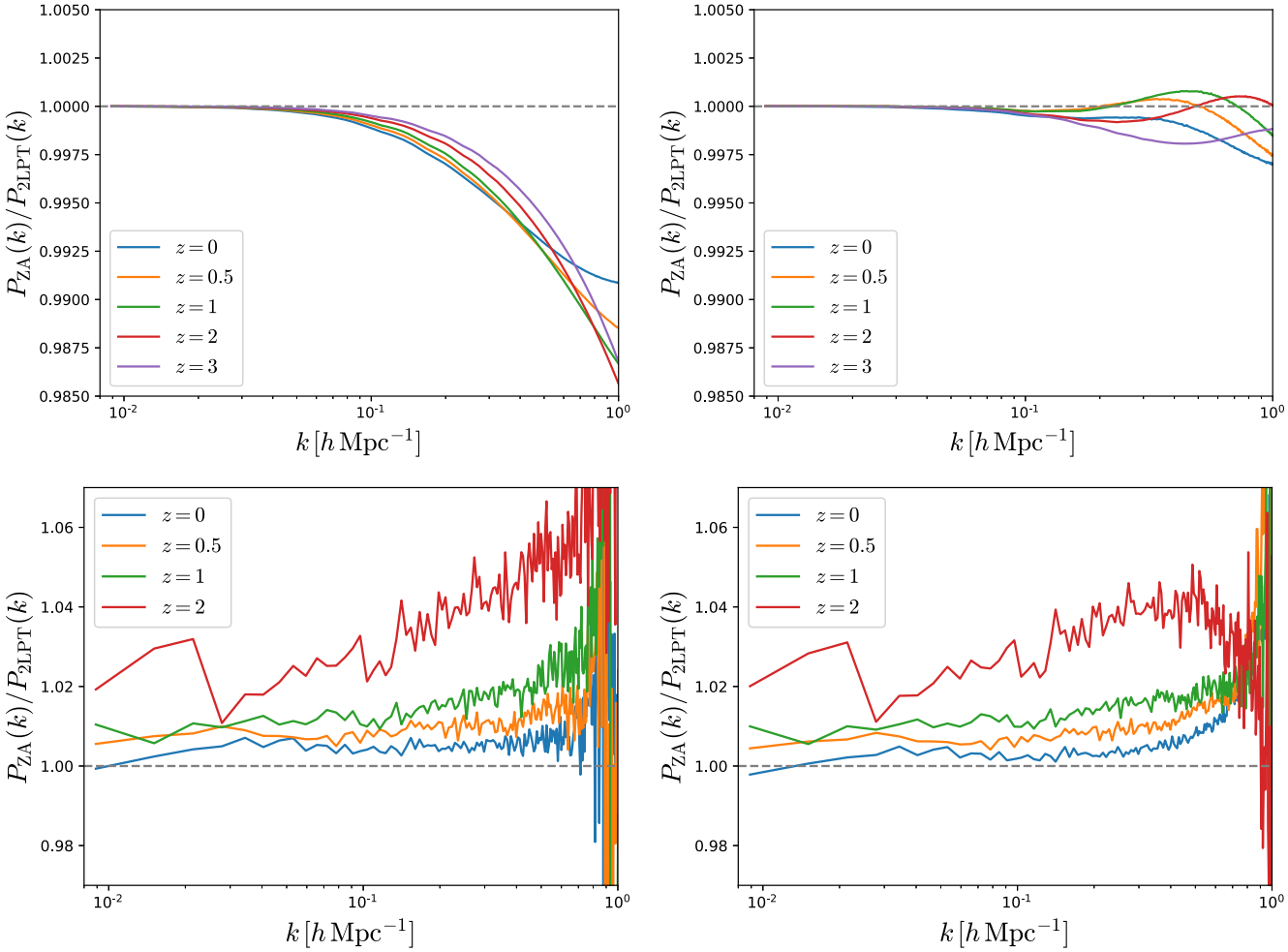

Figure 13. Effect on the matter power spectrum in real (top left) and redshift (top right) space of generating the ICs using the Zel'dovich approximation vs. 2LPT. The bottom panels show the same for the power spectrum of halos with masses above $3.2 \times 10^{13}\,h^{-1}\,M_\odot$ in real (bottom left) and redshift (bottom right) space. The plots show the ratio between the two power spectra as a function of wavenumber for different redshifts. For matter in real space, the effect is below 1.5%, while in redshift space, the effect is below 0.5% on all scales. For halos at low redshift, the effect is $\lesssim$1%. Near $k = 1\,h\,{\rm Mpc}^{-1}$, the halo power spectrum becomes negative (after subtracting shot noise) and is severely affected by numerical noise. Since the ICs of the massive neutrino simulations have been generated using the Zel'dovich approximation, in comparison with 2LPT for the other models, it is important to keep this effect in mind when computing numerical derivatives.

5. Resolution Tests

In this section, we present some tests performed on the QUIJOTE simulations to quantify the convergence of the simulations on several properties.

5.1. Zel'dovich versus 2LPT

The reason we use the Zel'dovich approximation, not 2LPT, to generate ICs for cosmologies with massive neutrinos is because, to our knowledge, it is unknown how to estimate the second-order scale-dependent growth factor and rate needed to use 2LPT in massive neutrino models.

Generating the ICs via Zel'dovich instead of 2LPT can induce small changes in the dynamics of the simulation particles that can lead to small statistical differences (Crocce et al. 2006). In order to quantify this effect, we have computed the matter and halo power spectra (for halos with masses above $3.2 \times 10^{13}\,h^{-1}\,\,M_\odot$) in 200 simulations of the fiducial cosmology: 100 simulations with Zel'dovich ICs and 100 simulations with 2LPT ICs. The random seeds are matched among the two sets. We show the results in Figure 13.

In the top panels, we show the results for the matter power spectrum in real (top left) and redshift (top right) space. We find that the differences in real space are below 1.5% at all redshifts, while in redshift space, the effects are much smaller, below 0.25%. In the bottom panels, we show the results for the power spectrum of halos in real (bottom left) and redshift (bottom right) space. We have corrected for Poissonian shot noise by subtracting $1/\bar{n}$ from the measurements, where $\bar{n}$ is the number density of halos. The results at $z = 3$ are very noisy due to the very low number density of halos; thus, for clarity, we do not show them. We find that differences in real and redshift space at low redshift are below $\simeq$1%. The higher the redshift, the larger the differences. The large variations we observe around $k = 1\,h\,{\rm Mpc}^{-1}$ are due to the halos' power spectra becoming very small and therefore highly affected by numerical noise. Notice that at low redshift, most of the differences we observe between the halos' power spectra have a very mild scale dependence. Thus, marginalizing over an overall amplitude can get rid of most of this effect.

We also carry out the above analysis for the bispectrum of halos in real and redshift space at $z = 0$ down to

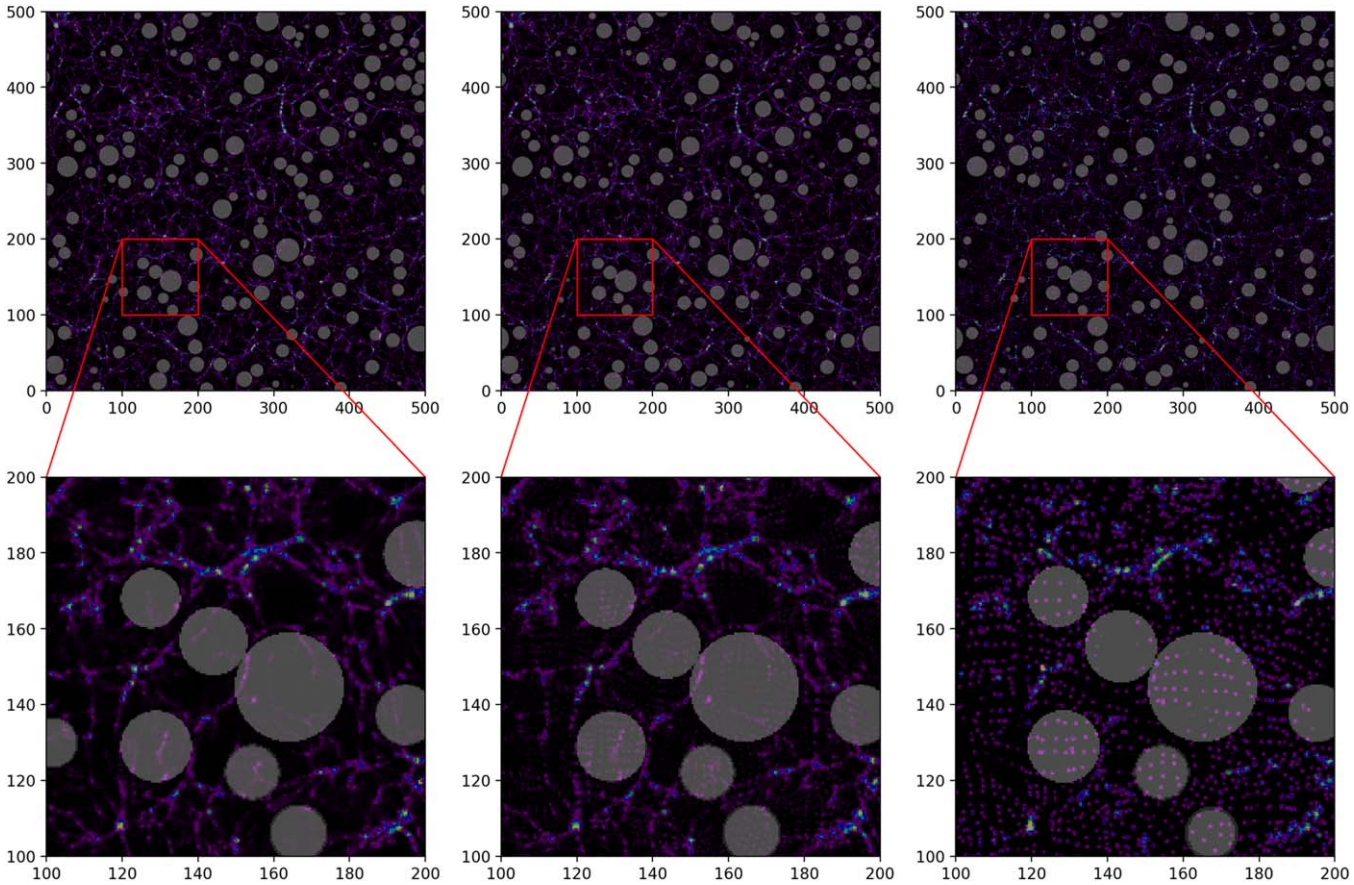

Figure 14. We have identified voids (gray spheres) in three simulations with the same random seed but different masses and spatial resolutions at $z = 0$. As can be seen, our void finder is relatively robust against these changes, at least for the largest voids.

$k_{\rm max} = 0.5\,h\,{\rm Mpc}^{-1}$. We find that differences in redshift space can be around 10% and slightly larger in real space.

When making a Fisher forecast analysis, it is important to keep this effect in mind, as the additional scale dependence present in the models with massive neutrinos may slightly affect the results. For this reason, when computing derivatives with respect to neutrino masses, we recommend using the simulations with Zel'dovich ICs from the fiducial model, instead of the 2LPT ones.

5.2. Clustering

One important aspect to consider when analyzing numerical simulations is the range of scales where results are converged. In order to quantify this, we have used three simulations, all within the fiducial cosmology but run at different resolutions: high (1024^3 particles), fiducial (512^3 particles), and low (256^3 particles).

In Figure 14, we show the projected matter overdensity field in a slice of $500 \times 500 \times 10\ (h^{-1}\,{\rm Mpc})^3$ for the three different simulations. As the amplitudes and phases of the modes that are common across the simulations are the same, the large-scale density field in the three images is basically the same. Differences show up on small scales, where different modes across simulations are present/absent. Resolution effects are clearly visible in the image; while in the low-resolution simulation, we can see individual particles in cosmic voids, in the high-resolution simulation, the density field is much smoother.

We have computed the matter power spectrum for those three simulations at redshifts 0, 0.5, 1, 2, and 3. We show the results in Figure 15. We find that at $z = 0$, the results of the fiducial-resolution run are converged all the way to $k = 1\,h\,{\rm Mpc}^{-1}$ at 2.5%. At higher redshifts, the results are only converged on larger scales; e.g., at $z = 3$, only scales $k \simeq 0.4\,h\,{\rm Mpc}^{-1}$ are converged at the fiducial resolution. We note that although the relative small-scale error increases with redshift, the absolute error decreases, since the amplitude of the power spectrum shrinks with redshift.

We emphasize that these tests indicate the range of scales where the absolute amplitude of the clustering should be trusted within a given accuracy. Numerical derivatives of statistics with respect to cosmological parameters may be converged to smaller scales, since it is expected that relative differences propagate among models in a systematic manner, such as taking differences cancels the systematic bias.

5.3. Void Finder

The void finder (see Section 3.3) we have run on the QUIJOTE simulations has some nice properties. One of them is that the positions and sizes of cosmic voids are not largely affected by the mass and spatial resolution of the simulation.[45]

In Figure 14, we show the locations and sizes of voids identified in three different simulations with the same random seed but different masses and spatial resolutions. As can be

[45] We are, of course, assuming that the sizes of the voids are larger than the spatial resolution of the density field.

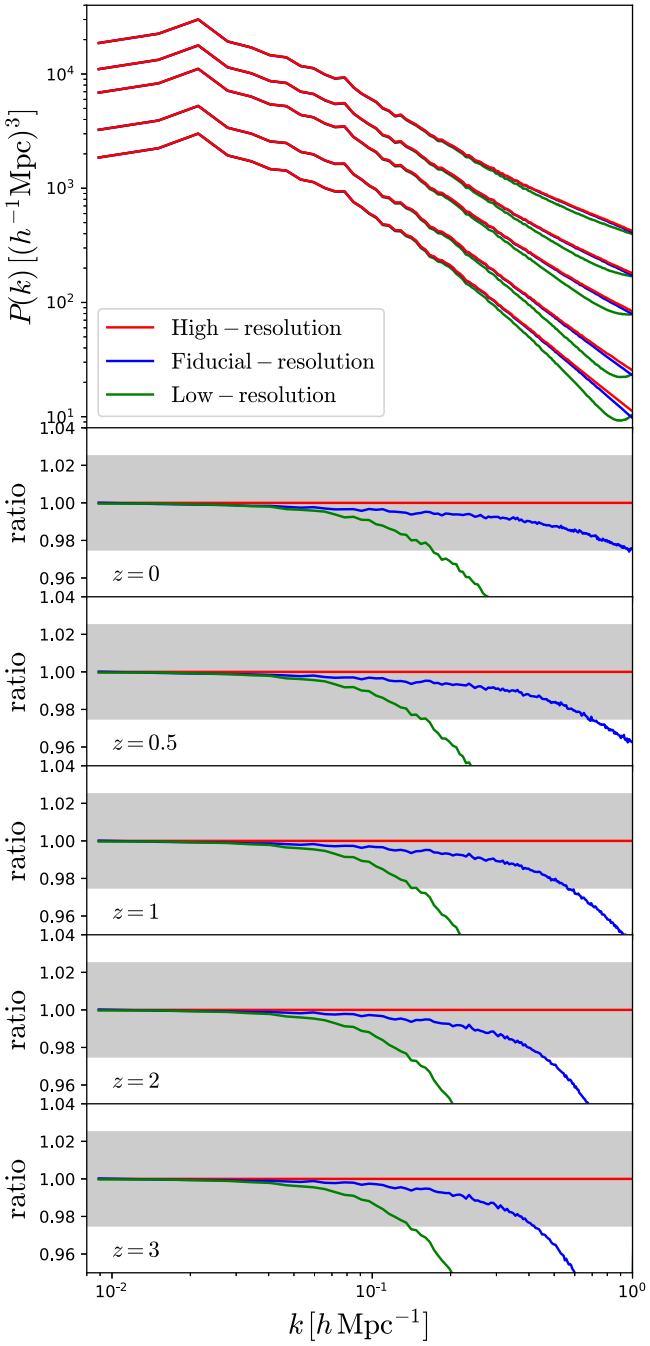

Figure 15. Matter power spectrum for a realization of the fiducial cosmology run at three different resolutions: (1) high (red lines), (2) fiducial (blue lines), and (3) low (green lines). The top panel shows the different power spectra from $z = 0$ (top lines) to 3 (bottom lines). The bottom panels display the ratio between the different power spectra at different redshifts. At $z = 0$, the matter power spectrum is converged all the way to $k = 1\ h\,\mathrm{Mpc}^{-1}$ at 2.5%, while at $z = 3$, this scale shrinks to $k \simeq 0.4\ h\,\mathrm{Mpc}^{-1}$.

seen, the locations and sizes of the voids among the simulations are very similar, pointing out the robustness of our void finder against mass and spatial resolution.

6. Summary

In this paper, we have introduced the QUIJOTE simulations, a large set of 44,100 full N-body simulations spanning thousands of different cosmologies and containing more than 8.5 trillion (8.5×10^{12}) particles at a single redshift. Each simulation follows the evolution of 256^3 (low resolution), 512^3 (fiducial resolution), or 1024^3 (high resolution) CDM particles in a periodic volume of $1(h^{-1}\,\mathrm{Gpc})^3$. Billions of dark matter halos and cosmic voids have been identified in the simulations, which required more than 35 million CPU hours to run.

The QUIJOTE simulations have been designed to accomplish two main goals:

1. quantify the information content on cosmological observables and
2. provide enough statistics to train machine-learning algorithms.

It is clear that there are many possible uses for these simulations beyond the ones we have mentioned here (see, e.g., Obuljen et al. 2019). We make the data from the QUIJOTE simulations freely available to the community with the goal of allowing the broadest possible exploration of their applications.

We believe the QUIJOTE simulations will complement very well the large efforts carried out by the community (see, e.g., Heitmann et al. 2014; Garrison et al. 2018; Nishimichi et al. 2019; DeRose et al. 2019; Knabenhans et al. 2019).

Instructions on how to download the data can be found at https://github.com/franciscovillaescusa/Quijote-simulations. As far as our storage resources allow, we will distribute all data products, e.g., halo and void catalogs, power spectra, marked power spectra, correlation functions, bispectra, PDFs, and full snapshots. The total amount of data generated by the QUIJOTE simulations exceeds 1 PB.

We also provide a set of Python libraries, PYLIANS, developed for many years, to help with the analysis of the data. PYLIANS can be found at https://github.com/franciscovillaescusa/Pylians.

We are especially thankful to Nick Carriero and Dylan Simon from the Flatiron Institute and Mahidhar Tatineni from the San Diego Supercomputer Center for their immense help with the multiple technical problems we faced while running the simulations. We thank Volker Springel for giving us access to Gadget-III. The work of S.H., E.G., E.M., D.S., F.V.N., and B.D.W. is supported by the Simons Foundation. C.D.K. acknowledges the support of National Science Foundation award No. DGE1656466 at Princeton University. A.P. is supported by NASA grant 15-WFIRST15-0008 to the WFIRST Science Investigation Team "Cosmology with the High Latitude Survey." L.V. acknowledges support from the European Union Horizon 2020 research and innovation program ERC (BePreSySe, grant agreement 725327) and MDM-2014-0369 of ICCUB (Unidad de Excelencia Maria de Maeztu). T.C. and B.D.W. also acknowledge financial support from the ANR BIG4 project under reference ANR-16-CE23-0002. A.M.D. acknowledges support from AstroCom NYC, NSF award AST-1831412, and Simons Foundation award No. 533845. S.H. thanks NASA for their support with NASA grant 15-WFIRST15-0008 and NASA Research Opportunities in Space and Earth Sciences grant 12-EUCLID12-0004.

ORCID iDs

ChangHoon Hahn https://orcid.org/0000-0003-1197-0902
Tom Charnock https://orcid.org/0000-0002-3479-3542
Elena Giusarma https://orcid.org/0000-0003-3052-3059
Alice Pisani https://orcid.org/0000-0002-6146-4437

Christina D. Kreisch https://orcid.org/0000-0002-5061-7805
Justin Alsing https://orcid.org/0000-0003-4618-3546
Licia Verde https://orcid.org/0000-0003-2601-8770

References

Allys, E., Levrier, F., Zhang, S., et al. 2019, A&A, 629, A115
Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440
Alsing, J., & Wandelt, B. 2018, MNRAS, 476, L60
Anderson, L., Pontzen, A., Font-Ribera, A., et al. 2018, arXiv:1811.00043
Angulo, R. E., & Pontzen, A. 2016, MNRAS, 462, L1
Armijo, J., Cai, Y.-C., Padilla, N., Li, B., & Peacock, J. A. 2018, MNRAS, 478, 3627
Banerjee, A., & Dalal, N. 2016, JCAP, 2016, 015
Banerjee, A., Powell, D., Abel, T., & Villaescusa-Navarro, F. 2018, JCAP, 09, 028
Beisbart, C., & Kerscher, M. 2000, ApJ, 545, 6
Blot, L., Corasaniti, P. S., Alimi, J.-M., Reverdy, V., & Rasera, Y. 2015, MNRAS, 446, 1756
Blot, L., Corasaniti, P. S., Amendola, L., & Kitching, T. D. 2016, MNRAS, 458, 4462
Brandbyge, J., Hannestad, S., Haugbølle, T., & Thomsen, B. 2008, JCAP, 8, 20
Bruna, J., & Mallat, S. 2013, ITPAM, 35, 1872
Bruna, J., & Mallat, S. 2018, arXiv:1801.02013
Carron, J. 2013, A&A, 551, A88
Charnock, T., Lavaux, G., & Wandelt, B. D. 2018, PhRvD, 97, 083004
Chuang, C.-H., Yepes, G., Kitaura, F.-S., et al. 2019, MNRAS, 487, 48
Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, MNRAS, 373, 369
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
DeRose, J., Wechsler, R. H., Tinker, J. L., et al. 2019, ApJ, 875, 69
Feng, Y., Chu, M.-Y., & Seljak, U. 2016, MNRAS, 463, 2273
Garrison, L. H., Eisenstein, D. J., Ferrer, D., et al. 2018, ApJS, 236, 43
Giusarma, E., Reyes Hurtado, M., Villaescusa-Navarro, F., et al. 2019, arXiv:1910.04255
Gottloeber, S., Kerscher, M., Kravtsov, A. V., et al. 2002, A&A, 387, 778
Gruen, D., Friedrich, O., Krause, E., et al. 2018, PhRvD, 98, 023507
Hahn, C., Villaescusa-Navarro, F., Castorina, E., & Scoccimarro, R. 2019, JCAP, 03, 40
He, S., Li, Y., Feng, Y., et al. 2019, PNAS, 116, 13825
Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, ApJ, 780, 111
Hernández-Aguayo, C., Baugh, C. M., & Li, B. 2018, MNRAS, 479, 4824
Ichiki, K., & Takada, M. 2012, PhRvD, 85, 063521
Klypin, A., Prada, F., & Byun, J. 2020, MNRAS, 496, 3862
Knabenhans, M., Stadel, M., Marelli, S., et al. 2019, MNRAS, 484, 5509
Knollmann, S. R., & Knebe, A. 2009, ApJS, 182, 608
Kodi Ramanah, D., Charnock, T., & Lavaux, G. 2019, PhRvD, 100, 043515
Kodi Ramanah, D., Charnock, T., Villaescusa-Navarro, F., & Wandelt, B. D. 2020, MNRAS, 495, 4227
Kodwani, D., Alonso, D., & Ferreira, P. 2019, OJAp, 2, 3
Leclercq, F., Pisani, A., & Wandelt, B. D. 2014, arXiv:1403.1260
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Li, Y., Hu, W., & Takada, M. 2014, PhRvD, 89, 083519
Li, Y., Schmittfull, M., & Seljak, U. 2018, JCAP, 02, 022
LoVerde, M., & Zaldarriaga, M. 2014, PhRvD, 89, 063502
Mallat, S. 2012, Commun. Pure Appl. Math., 65, 1331
Mallat, S., Zhang, S., & Rochette, G. 2018, arXiv:1810.12136
Massara, E., Villaescusa-Navarro, F., et al. 2020, arXiv:2001.11024
McClintock, T., Rozo, E., Banerjee, A., et al. 2019a, arXiv:1907.13167
McClintock, T., Rozo, E., Becker, M. R., et al. 2019b, ApJ, 872, 53
Nishimichi, T., Takada, M., Takahashi, R., et al. 2019, ApJ, 884, 29
Obuljen, A., Dalal, N., & Percival, W. J. 2019, JCAP, 10, 20
Perlmutter, S., Aldering, G., Galdhaber, G., et al. 1999, ApJ, 517, 565
Philcox, O. H. E., Spergel, D. N., & Villaescusa-Navarro, F. 2020, PhRvD, 12, 123520
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv:1807.06209
Planck Collaboration, Akrami, Y., Arroja, F., et al. 2019, arXiv:1905.05697
Pontzen, A., Slosar, A., Roth, N., & Peiris, H. V. 2016, PhRvD, 93, 103519
Ravanbakhsh, S., Oliva, J., Fromenteau, S., et al. 2017, arXiv:1711.02033
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Scoccimarro, R. 2015, PhRvD, 92, 083532
Scoccimarro, R., & Sheth, R. K. 2002, MNRAS, 329, 629
Sefusatti, E., Crocce, M., Scoccimarro, R., & Couchman, H. M. P. 2016, MNRAS, 460, 3624
Sefusatti, E., & Scoccimarro, R. 2005, PhRvD, 71, 063001
Sheth, R. K., Connolly, A. J., & Skibba, R. 2005, arXiv:astro-ph/0511773
Sifre, L., & Mallat, S. 2013, in Proc. IEEE Conf. Comput. Vision Pattern Recogn (Piscataway, NJ: IEEE), 1233
Sirko, E. 2005, ApJ, 634, 728
Springel, V. 2005, MNRAS, 364, 1105
Takada, M., & Hu, W. 2013, PhRvD, 87, 123504
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Uhlemann, C., Codis, S., Kim, J., et al. 2017, MNRAS, 466, 2067
Uhlemann, C., Codis, S., Pichon, C., Bernardeau, F., & Reimberg, P. 2016, MNRAS, 460, 1529
Uhlemann, C., Feix, M., Codis, S., et al. 2018, MNRAS, 473, 5098
Uhlemann, C., Friedrich, O., Villaescusa-Navarro, F., et al. 2020, MNRAS, 495, 4006
Valogiannis, G., & Bean, R. 2018, PhRvD, 97, 023535
Verde, L. 2007, arXiv:0712.3028
Viel, M., Haehnelt, M. G., & Springel, V. 2010, JCAP, 6, 015
Villaescusa-Navarro, F., Bird, S., Peña-Garay, C., & Viel, M. 2013, JCAP, 3, 019
Villaescusa-Navarro, F., Miralda-Escudé, J., Peña-Garay, C., & Quilis, V. 2011, JCAP, 6, 027
Villaescusa-Navarro, F., Naess, S., Genel, S., et al. 2018, ApJ, 867, 137
Wandelt, B. D. 2013, in Astrostatistical Challenges for the New Astronomy, ed. J. M. Hilbe (New York: Springer), 1013
White, M. 2016, JCAP, 1611, 057
Wibking, B. D., Salcedo, A. N., Weinberg, D. H., et al. 2019, MNRAS, 484, 989
Zel'dovich, Y. B. 1970, A&A, 5, 84
Zennaro, M., Angulo, R. E., Aricò, G., Contreras, S., & Pellejero-Ibáñez, M. 2019, MNRAS, 489, 5938
Zennaro, M., Bel, J., Villaescusa-Navarro, F., et al. 2017, MNRAS, 466, 3244
Zhai, Z., Tinker, J. L., Becker, M. R., et al. 2019, ApJ, 874, 95