

Cosmic variance or galaxy bias? Disentangling finite-volume and galaxy formation effects in cosmological analysis

Francesco Sinigaglia^{1,2,3,4,*} and Francisco-Shu Kitaura^{5,6}

¹*Institute for Fundamental Physics of the Universe (IFPU), Via Beirut 2, I-34151 Trieste, Italy*

²*SISSA—International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy*

³*INAF—Osservatorio Astronomico di Trieste, Via Gian Battista Tiepolo 11, I-34131 Trieste, Italy*

⁴*INFN—National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy*

⁵*Instituto de Astrofísica de Canarias, Calle via Láctea s/n, E-38205, La Laguna, Tenerife, Spain*

⁶*Departamento de Astrofísica, Universidad de La Laguna, E-38206, La Laguna, Tenerife, Spain*



(Received 12 January 2026; accepted 9 July 2026; published 28 July 2026)

Current and forthcoming galaxy redshift surveys, such as DESI and Euclid, are going to bring cosmological analysis to an unprecedentedly exquisite level of precision in the determination of the cosmological parameters. However, these efforts require a high degree of control over theory and systematics, to obtain unbiased results. In this sense, the cosmic variance associated to finite-volume effects represents a major challenge and should adequately accounted for. In this work, we revisit the definition of cosmic variance and develop a novel framework to describe it using a “galaxy biasing” formalism. In particular, we use halo/galaxy Eulerian perturbation theory to relate the density field from an arbitrary cosmic realization to its counterpart having statistical properties reproducing the ensemble average, introducing a new set of bias parameters. We then apply this idea to the description of the nonlinear shift of baryon acoustic oscillations (BAO), disentangling the source of uncertainty from cosmic variance and galaxy biasing associated with the measurement of the BAO scale. We finally check our analytical argument against cosmological variance-suppressed N -body simulations, finding an expected reduction in the uncertainty on the BAO peak position. We conclude that extra care should be used when inferring cosmological information from perturbative approaches involving the estimation of bias parameters and propose new practical strategies to optimally leverage the novel formulation of cosmic variance presented herein in cosmological analysis.

DOI: [10.1103/1dqk-pq4q](https://doi.org/10.1103/1dqk-pq4q)

I. INTRODUCTION

Current and forthcoming large-scale galaxy redshift surveys, such as DESI [1] and Euclid [2], are going to deliver unprecedented three-dimensional maps of the Universe, whose analyses promise to shed light onto unsolved key questions about cosmology and fundamental physics, such as unveiling the nature of dark matter (e.g., [3,4]) and dark energy (e.g., [5–9]), testing the validity of general relativity as theory of gravity (e.g., [10–12]), and probing primordial non-Gaussianities and the inflationary scenario (e.g., [13,14]), among others. In particular, the DESI collaboration has recently presented the strongest detection ever of baryon acoustic oscillations (hereafter BAO (baryon acoustic oscillations)) achieving subpercent

precision in the determination of the position of the acoustic peak [15,16].

In this golden era of precision cosmology, it is very timely to develop a complete understanding of all the ingredients associated to the measurements of cosmological parameters and, in particular, their related uncertainties. In this work, we analyze two important uncertainty contributions, which are strongly coupled: (i) cosmic variance, arising from the intrinsic variance imprinted in the initial density field, and (ii) galaxy biasing, describing galaxy formation effects. The former stems from the particular random realization of the Gaussian random field. The latter concerns the complex nonlinear processes driving galaxy formation under different physical conditions, and depends on the characteristics of the host halo and of the cosmic web environment where it resides, among others. Theoretical analytical models typically include a set of free parameters aimed at describing galaxy biasing and fit them to observations. However, in this practical framework, it is impossible to know the relative contribution of galaxy bias and cosmic variance and the latter is not taken into account at the level of the estimation of the bias parameters, but only

*Contact author: fsinigag@sissa.it

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

as an additional uncertainty contribution. In other words, the estimation of the bias parameters is performed by assuming the average matter power spectrum, and neglecting the cosmic variance effect coming from surveying a limited cosmological volume.

In this paper we aim at disentangling cosmic variance and galaxy biasing and estimate separately the contributions coming from these two effects. To do so, we present a novel formalism based on halo/galaxy Eulerian perturbation theory, to treat cosmic variance in a similar way as done for galaxy biasing, and to mathematically disentangle the two effects. This is particularly relevant for several applications, such as e.g. (i) BAO reconstruction, in which one needs to make assumptions on the numerical value for the galaxy bias, (ii) cosmological information extraction, such as inferring cosmological parameters from BAO, and (iii) galaxy evolution studies, which use the galaxy bias as observational constraints to shed light onto the physics driving galaxy formation.

The paper is structured as follows. In Sec. II, we briefly review the Eulerian perturbation theory approach. In Sec. III, we present the formalism to model cosmic variance underlying this work. In Sec. IV, we discuss the integral constraint and incorporate it in the cosmic variance treatment presented herein. In Sec. V, we apply the novel cosmic variance formulation to the nonlinear BAO shift. Section VI presents the numerical results validating the analytical framework. In Sec. VII, we present a few strategies to practically apply the ideas presented in this work to galaxy surveys and optimally extract cosmological information. We conclude in Sec. VIII.

II. EULERIAN PERTURBATION THEORY

According to [17], the expansion of the matter density contrast in Eulerian perturbation theory can be written as a sum of perturbative contributions,

$$\delta(\mathbf{k}, \eta) = \sum_{n=1}^{\infty} \delta^{(n)}(\mathbf{k}, \eta), \quad (1)$$

where $\delta^{(n)}$ denotes the n th order contribution and η is the time variable (e.g., conformal time or $\ln a$).

If we discretize Fourier modes (or grid cells) and collect the density field $\delta(\mathbf{k})$ into a vector with indices $i, j, k, \dots$, the convolution integrals appearing in perturbation theory can be written as discrete sums with coupling matrices (or tensors). The nonlinear density field then takes the form

$$\delta_i = \delta_i^L + \sum_{jk} F_{ijk}^{(2)} \delta_j^L \delta_k^L + \sum_{jkl} F_{ijkl}^{(3)} \delta_j^L \delta_k^L \delta_l^L + \dots, \quad (2)$$

where δ^L denotes the linear density field and $F_{ij_1 \dots j_n}^{(n)}$ are the discretized Eulerian perturbation theory kernels encoding mode coupling (for details see, e.g., [17]).

In the continuum limit, the n th order contribution in Fourier space can be written explicitly as

$$\begin{aligned} \delta^{(n)}(\mathbf{k}, \eta) = & D^n(\eta) \int \frac{d^3 q_1}{(2\pi)^3} \cdots \frac{d^3 q_n}{(2\pi)^3} (2\pi)^3 \\ & \times \delta_D\left(\mathbf{k} - \sum_{a=1}^n \mathbf{q}_a\right) F_n(\mathbf{q}_1, \dots, \mathbf{q}_n) \prod_{a=1}^n \delta_L(\mathbf{q}_a), \end{aligned} \quad (3)$$

where $D(\eta)$ is the linear growth factor, δ_D denotes the Dirac delta distribution enforcing momentum conservation, and F_n are the (symmetrized) Eulerian perturbation theory kernels.

Collecting the perturbative contributions and transforming back to real space, the nonlinear matter density contrast can be written schematically as

$$\delta(\mathbf{x}) = \delta_L(\mathbf{x}) + \delta^{(2)}(\mathbf{x}) + \delta^{(3)}(\mathbf{x}) + \dots \quad (4)$$

At second order, the density field admits the operator expansion

$$\delta^{(2)}(\mathbf{x}) = c_1 \delta_L^2(\mathbf{x}) + c_2 s_{ij}(\mathbf{x}) s^{ij}(\mathbf{x}) + c_3 \delta_L(\mathbf{x}) \nabla^2 \Phi(\mathbf{x}), \quad (5)$$

where the coefficients c_1 , c_2 , and c_3 are fixed by the underlying gravitational dynamics (e.g., $c_1 = 17/21$ and $c_2 = 2/7$ in an Einstein–de Sitter universe), s_{ij} is the tidal tensor,

$$s_{ij}(\mathbf{x}) = \left(\frac{\partial_i \partial_j}{\nabla^2} - \frac{1}{3} \delta_{ij} \right) \delta_L(\mathbf{x}). \quad (6)$$

Neglecting total-derivative and convective terms, or equivalently working with locally smoothed fields, the nonlinear density contrast may be written in the schematic form

$$\delta(\mathbf{x}) = \delta_L(\mathbf{x}) + c_1 \delta_L^2(\mathbf{x}) + c_2 s_{ij}(\mathbf{x}) s^{ij}(\mathbf{x}) + \mathcal{O}(\delta_L^3). \quad (7)$$

This expression makes it explicit that, although gravitational evolution is intrinsically nonlocal, Eulerian perturbation theory necessarily generates local composite operators such as $\delta_L^2(\mathbf{x})$, accompanied by nonlocal tidal contributions. In the remainder of the paper, we neglect for simplicity nonlocal terms, and use only the following local relation

$$\delta(\mathbf{x}) = \delta_L(\mathbf{x}) + c_1 \delta_L^2(\mathbf{x}) + c_4 \delta_L^3(\mathbf{x}) + \dots \quad (8)$$

Afterwards, on sufficiently large scales where the density perturbations are small, the galaxy density can be modeled in a similar perturbative way as powers of the nonlinear density field through the series,

$$\delta_g(\mathbf{x}) = b_0 + b_1\delta(\mathbf{x}) + \frac{1}{2}b_2\delta^2(\mathbf{x}) + \dots \quad (9)$$

We notice that we will refer to “galaxy bias” throughout the paper, but the same argument holds true for any biased tracer of the dark matter field. This expression provides a practical framework to describe the clustering of dark matter tracers by means of perturbation theory techniques. In this framework, the two-point correlation function of the dark matter field can be computed as $\xi = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle$, and similarly for galaxies $\xi_g = \langle \delta_g(\mathbf{x}_1)\delta_g(\mathbf{x}_2) \rangle$. While for this paper we have limited the Eulerian expression to just a local relation for simplicity, it is a well-established result that the galaxy bias has crucial nonlocal contributions up to third order (e.g., [18–20]). This simple formalism sets the basis for our treatment and will be used in the remainder of the paper.

From Eq. (9), it is evident that the clustering and phenomenology of the biased field is determined by the parameters b_1 , b_2 , and others used to describe the model. These values for these parameters depend on redshift, on the analyzed tracer and its properties (e.g., the bias parameters will be different for different galaxy populations), and are meant to model all the physical processes related to the considered tracers: halo/galaxy formation and its dependence on the large-scale structure environment and on the anisotropic clustering, galaxy evolution, and baryon physics, among others.

III. A NOVEL FRAMEWORK FOR COSMIC VARIANCE

The Eulerian framework developed in Sec. II implicitly assumes an infinite volume and provides analytical expressions corresponding to the ensemble average of the dark matter and halo/galaxy fields. How does this picture change when we consider an arbitrary realization of the Universe in a finite volume? The statistical properties of such a realization will tend to converge to the ensemble average on small scales, but they may diverge significantly at the largest scales due to their poor sampling. This is a well-known situation for the large-scale modes of the power spectrum and analogously for the two-point correlations functions measurements on large scales. While this issue is normally treated as just a source of uncertainty, it has a deeper physical meaning: in fact, whenever a perturbative framework is applied to perform cosmological measurements from a clustering observables, the found bias parameters b_1 and b_2 will include the degenerate effect of cosmic variance and of galaxy biasing. This aspect may become a problem in several aspects of cosmological analysis, in which a value for the bias needs to be assumed, as is the case for BAO reconstruction.

To disentangle the two aforementioned effects, we introduce in this work a novel way of treating the density field from arbitrary realization as a “biased” tracer of its

counterpart version featuring statistical properties identical to the ensemble average, which is the quantity described in Eq. (8). In particular, by calling δ' the density field from an arbitrary realization in a finite volume, we can write,

$$\delta'(\mathbf{x}) = b_{0,\text{dm}} + b_{1,\text{dm}}\delta(\mathbf{x}) + \frac{1}{2}b_{2,\text{dm}}\delta^2(\mathbf{x}) + \dots \quad (10)$$

This expression assumes again that the “biased” dark matter field has just tiny fluctuations with respect to its counterpart with average properties. In the limit for the volume that tends to infinity, $b_{1,\text{dm}} \rightarrow 1$, $b_{2,\text{dm}} \rightarrow 0$, and $\delta' \rightarrow \delta$.

In this picture, we have disentangled the effects of cosmic variance and of galaxy biasing, at the expense of introducing a new set of parameters modeling the former. In this way, the bias parameters will now be variance-free and will truly correspond to the biasing effect.

To develop an intuitive understanding of the meaning of $b_{1,\text{dm}}$ and $b_{2,\text{dm}}$, let us use the formalism of Lagrangian perturbation theory (LPT). In LPT, the Lagrangian and Eulerian coordinates $\mathbf{q}$ and $\mathbf{x}$ are related through the displacement field Ψ as: $\mathbf{x} = \mathbf{q} + \Psi(\mathbf{q})$. The displacement field in first-order LPT (Zel’dovich approximation [21]) is related to the initial linear density field through the relation

$$\delta_l \sim -\nabla \cdot \Psi(\mathbf{q}). \quad (11)$$

Given that the cosmic variance stems from the initial conditions, one can write an expression similar to Eq. (10) also for the initial density field from an arbitrary realization. In this sense, the two bias parameters $b_{1,l}$ and $b_{2,l}$ will control the displacement field through Eq. (11), and hence, determine the nonlinear dynamics of the system. The values of the initial density field are fully characterized mode by mode in Fourier space by the amplitude—regulated by the linear power spectrum—and the phase. Both the amplitude and the phase contribute to cosmic variance. It is straightforward to understand how the amplitude, by definition, sources cosmic variance in the two-point statistics. The phases, instead, do not necessarily impact two-point statistics, but induce cosmic variance in the one-point statistics. This is clear if one considers paired simulations, whose initial density values are related through a rotation of angle $\theta \rightarrow \theta + \pi$, i.e., a change of sign in configuration space. In this case, such simulations will have exactly the same amplitude mode by mode, but the one-point PDF will be symmetric with respect to zero density. These effects in the initial conditions are then propagated in a nontrivial way through the nonlinear evolution of cosmic structures.

IV. INTEGRAL CONSTRAINT

In large scale structure cosmology, another source of uncertainty stemming from surveying a finite volume is the

so called *integral constraint*, which consists in a potential bias arising from the estimation of the mean density of the Universe using the surveyed volume itself. In particular, inside a survey volume the estimator forces the average density inside the observed volume to vanish,

$$\int_V \delta(\mathbf{x}) d^3\mathbf{x} = 0. \quad (12)$$

However, the mean overdensity δ_V within the survey volume does not necessarily vanish,

$$\delta_V \equiv \int_V \delta(\mathbf{x}) d^3\mathbf{x} \neq 0. \quad (13)$$

Therefore, the measured density fluctuations are actually $\hat{\delta}(\mathbf{x}) = \delta(\mathbf{x}) - \delta_V$. In this picture, the largest-scale modes are suppressed, the clustering amplitude is biased low on large scales, and the measured correlation function differs from the true one.

Therefore, cosmic variance quantifies the statistical fluctuations of the survey mean due to long modes, while the integral constraint represents the bias that is introduced by forcing those long modes to zero by the estimator. In this sense, the integral constraint is, strictly speaking, not a cosmic variance term, but a distortion that arises from the estimator due to the cosmic variance. While in a standard analysis the integral constraint is treated by modifying the theory at the clustering level, in this formalism we simply express the final observed galaxy density field by using Eq. (9),

$$\hat{\delta}_g(\mathbf{x}) = \delta_g(\mathbf{x}) - \delta_{g,V}, \quad (14)$$

where $\delta_{g,V}$ is the measured mean galaxy number density within the survey volume. By expanding Eq. (9) and defining $b'_0 = b_0 - \delta_V$, the contribution from the integral constraint gets reabsorbed in the definition of the constant term b'_0 , therefore it implies no extra bias parameters. In this way, the cosmic variance contribution can be disentangled from galaxy bias as explained in the previous sections with no further complications.

V. THE NONLINEAR BAO SHIFT

In this section, we apply the model for cosmic variance introduced in Sec. III to study its impact on the determination of the nonlinear BAO scale from clustering measurements. In particular, we rely on the formalism by [22], wherein the authors derive analytical expressions for the nonlinear BAO shift of the density field and of biased tracers by means of Eulerian perturbation theory. Let us express the density field as $\delta(\mathbf{x}) = \delta_l(\mathbf{x}) + \delta_s(\mathbf{x})$, i.e., as the sum of a long-wavelength component δ_l —sourcing the BAO shift—and a short-wavelength component δ_s on scales comparable to the BAO scale. In this framework,

the nonlinear BAO shift of the real space correlation function for the dark matter field reads [22],

$$\Delta\alpha = \alpha - 1 = \frac{131}{105} \langle \delta_l^2 \rangle \quad (15)$$

and for biased tracers,

$$\Delta\alpha = \alpha - 1 = \frac{131}{105} \left(1 + \frac{70}{131} \frac{b_2}{b_1} \right) \langle \delta_l^2 \rangle, \quad (16)$$

where b_1 and b_2 are the Eulerian bias parameters defined in Eq. (9). At this stage, b_1 and b_2 include the degenerate effect from true galaxy biasing and from cosmic variance. To disentangle the two, let us rewrite Eq. (16) in two parts: one part related to the intrinsic cosmic variance of the dark matter field with respect to the ensemble average, controlled by the bias parameters $b_{1,\text{dm}}$ and $b_{2,\text{dm}}$

$$\Delta\alpha_{\text{dm}} = \frac{131}{105} \left(1 + \frac{70}{131} \frac{b_{2,\text{dm}}}{b_{1,\text{dm}}} \right) \langle \delta_l^2 \rangle \quad (17)$$

and the other related to galaxy biasing and controlled by the parameters $b_{1,g}$ and $b_{2,g}$

$$\Delta\alpha_g = \frac{131}{105} \left(1 + \frac{70}{131} \frac{b_{2,g}}{b_{1,g}} \right) \langle \delta_l^2 \rangle, \quad (18)$$

where we have expressed also the “biased” dark matter field as $\delta'(\mathbf{x}) = \delta'_l(\mathbf{x}) + \delta'_s(\mathbf{x})$. We notice that in Eq. (17) we are considering only the long-range component δ_l and neglecting the short-wavelength part, which has negligible contribution to the BAO shift [22]. In this formulation, it is straightforward to understand that cosmic variance alone can induce an additional non-negligible BAO shift, especially in small volumes, which can be comparable to the one induced by galaxy biasing. Therefore, one should pay extra care when estimating the BAO shift using perturbative approaches, such as, e.g., effective field theory, applied to small volumes. These results can be intuitively understood by considering that the BAO shift is driven at first order by the Zel’dovich displacement [21], which will be sensitive to an excess overdensity in a simulation volume with respect to the ensemble average, and will typically result in an extra shrinking of the BAO scale compared to the mean BAO shift. It is important to stress that in a realistic cosmological analysis, the estimation of the bias parameters is performed by relating the “ensemble average” dark matter field to the biased tracers of a specific realizations. Therefore, by construction the estimated bias parameters will include the degenerate contribution from the bias and from cosmic variance. While this is in principle accounted for in the estimation of the uncertainty of the bias parameters, in Sec. VII we propose a few strategies to disentangle these effects and suppress the contribution to the uncertainty on

the estimated cosmological parameters due to cosmic variance.

VI. EMPIRICAL VALIDATION WITH SIMULATIONS

In this section, we investigate the variance associated to the measurements of the BAO scale by using N-body simulations, with the aim of showing that controlling the variance in the initial conditions results in a suppression of the uncertainty on the BAO scale. To this end, we consider the UNIT simulations suite [23], designed to provide precise predictions for nonlinear statistics of the galaxy distribution. The simulations of the UNIT project were run either with the standard GADGET TreePM code [24,25] or with the particle mesh FastPM code [26], and include a variety of different realizations in different volumes and resolutions. Specifically, the flagship runs, as well as other simulation within the suite, were run by adopting the fixing and pairing techniques for initial conditions, to suppress cosmic variance and provide a robust framework for galaxy clustering analysis.

In this work, we consider the suite of 2×100 fixed-and-paired simulations run with FastPM in $V = (1 \text{ } h^{-1} \text{ Gpc})^3$ volumes and with $N = 1024^3$ dark matter particles. In particular, we consider the output at $z = 0$ and make use of the measurements of the isotropic two-point correlation function in real space for the dark matter particles and both in real and in redshift space for the haloes made available by the UNIT consortium.

We estimate the BAO peak position in a model-independent fashion as follows. We first fit a broadband shape polynomial of the type $A(r) = ar^{-2} + br^{-1} + c$, with a , b , c free parameters (see, e.g., [27,28]). Afterward, we subtract it from the full correlation function, to keep only the BAO peak “wiggly” part. Then, we fit a Gaussian function to the peak component of the correlation function, in the scales range $r \in [90, 120] h^{-1} \text{ Mpc}$, and consider the centroid of the Gaussian as a model-independent position of the acoustic peak. We notice that more sophisticated ways of fitting the BAO peak position based on templates of the input linear power spectrum are available from the literature (e.g., [29] and references therein). However, since here we are only interested in a model-independent estimate of the position of the acoustic peak and not in determining the α parameter, we limit our methodology to what is described above. Finally, we build the distribution of BAO peak positions for each of the studied cases mentioned above and estimate the ensemble average value as the median of the distributions and use the (84th–50th, 50th–16th) percentiles as associated upper and lower uncertainties.

We apply this BAO fitting procedure on the following sets of simulations:

- (i) 100 standard simulations;
- (ii) 100 fixed-amplitude simulations;

- (iii) 2×50 paired simulations;
- (iv) 2×50 fixed-and-paired simulations.

We show the resulting BAO peak position distributions in Fig. 1 and report the numerical results in Table I. We notice that in all the studied cases, the BAO peak position is well converged at subpercent level in the different studied cases and there is no apparent systematics introduced by using fixing and/or pairing techniques with respect to the ensemble of the 100 standard simulations. The sample of 100 fixed-and-paired simulations has always the smallest variance among the studied cases, by a factor ~ 5.4 with respect to the traditional ensemble average of the dark matter in real space, by a factor ~ 3.4 for the haloes in real space, and by a factor ~ 3.9 for the haloes in redshift space. The fixing and pairing techniques produce a suppression in the uncertainty on the BAO peak: a factor ~ 2.1 (fixing) and ~ 1.5 (pairing) for the real-space dark matter field, a factor ~ 1.6 (fixing) and ~ 1.7 (pairing) for the real-space halo field, and a factor ~ 1.8 (fixing) and ~ 1.7 (pairing) for the redshift-space halo field. The same results can be also visually appreciated by looking at the distributions shown in Fig. 1. From these results, we infer that fixing and pairing have different, complementary effects on the overall error budget on the BAO peak position: the fixing technique reduces the large-scale bias of each individual realization with respect to the ensemble average, while stacking couples of paired simulations reduced the uncertainty due to the deviation from the average nonlinear shift.

VII. APPLICATION TO GALAXY SURVEYS

In this section, we briefly illustrate two potential different ways stemming from the idea presented in this work, to separate the contributions from cosmic variance and galaxy biasing in BAO analysis. In particular, this could be achieved through:

- (i) An optimal combination of the clustering from different complementary tracers. It has been shown that adequately fitting the postreconstruction combined BAO signal from different tracers, such as, e.g., galaxies and cosmic voids, can lead up to a $\sim 10\%$ – 20% precision improvement in the determination of the position of the acoustic peak (e.g., [29,30]). Here, we propose to shift the focus and tailor the combination in such a way not to necessarily maximize the achieved statistical precision, but rather to cancel out cosmic variance. This can be achieved by (i) revisiting the definition of the two prereconstruction samples in such a way that they trace exactly complementary density regimes, or (ii) finding the optimal set of weights in the computation of two-point correlation function from the combined postreconstruction samples. This strategy aims at mimicking the usage of two paired realizations, fully exploiting the information from

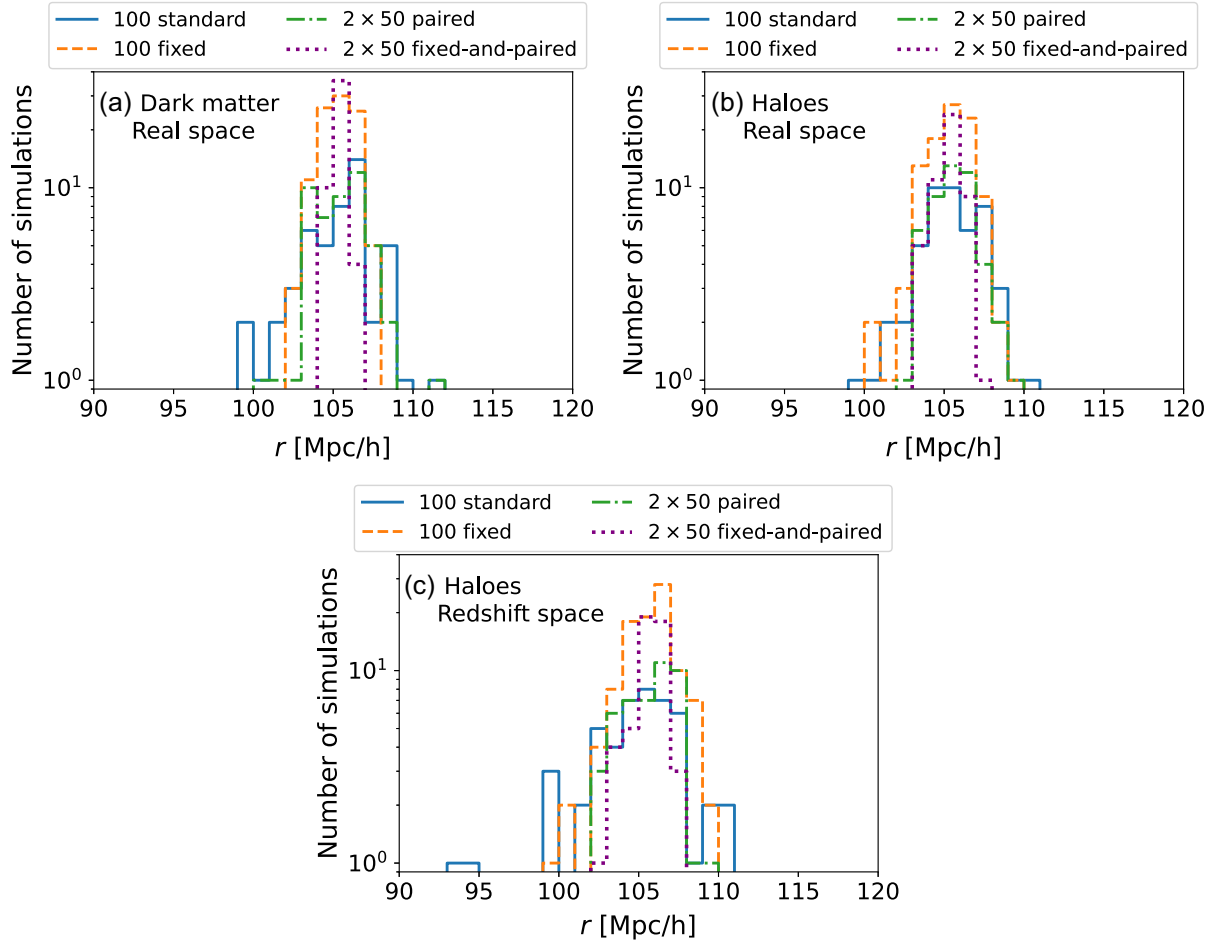


FIG. 1. Distributions of the BAO peak position measured from the two-point correlation function from FastPM simulation from the UNIT project, as measured from the dark matter field in real space (a), the halo field in real space (b) and in redshift space (c) at $z = 0$. We display the distribution from 100 standard simulations as a blue solid line, from 100 fixed-amplitude simulations as an orange dashed line, from 2×50 paired realizations as a green dotted-dashed line, and from 2×50 fixed-and-paired simulations as a purple dotted line.

the full density range and should be accurately tested on detailed mock catalogs;

- (ii) The application of cosmic variance suppression techniques (e.g., fixing and pairing techniques) both to the true observed data and to the mock catalogs adopted to estimate the covariance matrices. In particular, the application of backward inversion methods relying on the negative displacement (see, e.g., [31,32] and references therein) field or forward Bayesian reconstruction techniques (e.g., [33–38] and references therein) to obtain the initial conditions underlying the observed data, coupled to fixing and pairing techniques, would yield a variance-suppressed version of the observed Universe and its (paired) anti-Universe. While fixed-and-paired mock catalogs feature a significant suppression of cosmic variance and are therefore not suited to be employed in traditional cosmological analysis [39], we argue that their usage would be consistent and legitimate if applied to the analysis of

“fixed-and-paired” observed clustering measurements. Extra care in this sense should be paid when performing the reconstruction, both in inversion and Bayesian forward models. In the former, the accuracy of the reconstruction method will significantly impact the final result when evolving the initial conditions forward in time after having applied the cosmic variance suppression techniques. As for the latter, one does not want to impose a given BAO shift through the cosmology adopted to compute the linear power spectrum, otherwise the final result will be naturally biased toward it. Potential ways of coping with this issue are sampling either the power spectrum, or the BAO shift as part of the reconstruction procedure. The former allows the practitioner to have a cosmology-agnostic reconstruction of the initial conditions, while the latter to obtain a posterior distribution for α to draw values from when generating samples for the initial conditions.

TABLE I. Numerical results from the fitting of the BAO scale in different samples of N -body simulations from the UNIT project. The first column summarizes the investigated sample of simulations, the second column reports the median BAO scale value from the full distribution of the sample, and the third and fourth column report the upper and lower uncertainty values associated to the BAO peak position, estimated as (84th–50th, 50th–16th) percentiles. The three sections of the table refer to the results for the real-space dark matter field (top), the real-space halo field (mid), and the redshift-space halo field (bottom) at $z = 0$.

Sample	r_B	$\Delta r_{B,\text{up}}$	$\Delta r_{B,\text{low}}$
<i>DM real space</i>			
100 standard	105.6	2.4	2.6
100 fixed	105.3	1.3	1.3
100 paired	105.5	1.6	1.9
100 fixed-and-paired	105.4	0.5	0.5
<i>Halo real space</i>			
100 standard	105.3	2.6	2.0
100 fixed	105.4	1.3	1.6
100 paired	105.8	1.4	1.4
100 fixed-and-paired	105.4	0.6	0.7
<i>Halo redshift space</i>			
100 standard	105.8	3.8	2.8
100 fixed	105.9	1.7	1.8
100 paired	105.9	1.8	2.1
100 fixed-and-paired	105.8	0.7	0.9

We leave the exploration of these ideas for future publications.

VIII. CONCLUSIONS

In this work, we have presented a novel formulation of cosmic variance in an Eulerian perturbation theory fashion. In particular, we have proposed a framework to study cosmic variance leveraging the formalism usually adopted to treat biased tracers of the dark matter field. Such formulation effectively disentangles the contribution of cosmic variance from the one of true galaxy bias by means of a new set of free parameters modeling the “cosmic variance bias” effect. We have applied this formalism to the perturbative approach to compute the nonlinear BAO shift originally proposed by [22], showing that cosmic variance alone can account for a significant bias on the nonlinear BAO shift, which is degenerate with the nonlinear shift stemming from the nonlinear gravitational growth of cosmic structures.

Afterward, we have tested the impact of cosmic variance on the determination of the BAO peak position from cosmological simulations. In particular, we considered a suite of simulations comprising 100 standard, 100 fixed, 2×50 paired, and 2×50 fixed-and-paired simulations. We have determined the BAO peak position for each realization in a model-independent fashion and built the resulting distributions for each aforementioned class of simulations. In this way, we have shown that both the fixing and pairing

techniques reduce the variance of the BAO scale distribution by a factor ~ 1.5 – 2 both for the dark matter field in real space and for the haloes in real and redshift space at $z = 0$. This indicates that each of them removes one of the source of uncertainties, which are found to have a similar contribution to the total error budget. Specifically, the pairing technique averages out the shift from couples of simulations, while the fixing technique alleviates the contribution to the variance due to the large-scale bias effect. The combination of the two techniques yields a suppression of the variance by a factor ~ 4 – 5 .

Finally, we have proposed two practical strategies, to be tested in future work, to disentangle cosmic variance and galaxy biasing effects in a realistic cosmological analysis. Specifically, one strategy consists in optimally leveraging the complementary properties of different samples in a multitracers analysis, either through a proper definition of the samples or via a weighting scheme. Cosmic voids have been proposed as a complementary tracer for BAO measurements [40–42] and shown to improve galaxy-only BAO constraints when combined with galaxy clustering [29,30]. These approaches could be naturally extended to a paired BAO analysis by optimizing the BAO scale shift between tracers, rather than the BAO signal to noise, which was the primary objective in those works. The more challenging possibility consists in applying variance and pairing suppression techniques to the reconstructed initial conditions used to generate paired constrained simulations of the Universe conditions underlying real observations, as well as to the initial conditions used to generate the mock catalogs used to estimate the covariance matrices. To this end, we need to have a complex nonlinear and nonlocal bias description (see, e.g., [20,43–45]) implemented in a full Bayesian framework [38]. In future work, we will test these techniques and assess the realistic gain in using them by applying this framework to DESI-like mock galaxy catalogs and combine it with state-of-the-art analysis techniques such as BAO reconstruction and effective field theory.

ACKNOWLEDGMENTS

F.S. acknowledges the support from the IFPU Postdoctoral Fellowship scheme. F.S.K. acknowledges the Spanish Ministry of Economy and Competitiveness (MINECO) for financing the Big Data of the Cosmic Web project: Project No. PID2020–120612GB-I00/AEI/10.13039/501100011033. F.S. is also grateful to the *Instituto de Astrofísica de Canarias* for hospitality during the completion of this project.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they are owned by a third party and the terms of use prevent public distribution. The data are available from the authors upon reasonable request.

- [1] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein, B. Flaugher, K. Honscheid, R. Kron, O. Lahav, P. McDonald, N. Roe, D. Schlegel, and representing the DESI Collaboration, The DESI experiment, a whitepaper for snowmass 2013, [arXiv:1308.0847](https://arxiv.org/abs/1308.0847).
- [2] L. Amendola *et al.* (The Euclid Theory Working Group), Cosmology and fundamental physics with the Euclid satellite, [Living Rev. Relativity](https://arxiv.org/abs/1708.07248) **21**, 2 (2018).
- [3] M. Braglia, X. Chen, and A. Loeb, Exotic dark matter and the DESI anomaly, *J. Cosmol. Astropart. Phys.* **11** (2025) 064.
- [4] U. Kumar, A. Ajith, and A. Verma, Evidence for non-cold dark matter from DESI DR2 measurements, [arXiv:2504.14419](https://arxiv.org/abs/2504.14419).
- [5] M. Cortês and A. R. Liddle, Interpreting DESI's evidence for evolving dark energy, *J. Cosmol. Astropart. Phys.* **12** (2024) 007.
- [6] T.-N. Li, P.-J. Wu, G.-H. Du, S.-J. Jin, H.-L. Li, J.-F. Zhang, and X. Zhang, Constraints on interacting dark energy models from the DESI baryon acoustic oscillation and DES supernovae data, *Astrophys. J.* **976**, 1 (2024).
- [7] J. Zheng, D.-C. Qiang, and Z.-Q. You, Cosmological constraints on dark energy models using DESI BAO 2024, *J. Cosmol. Astropart. Phys.* **08** (2025) 056.
- [8] L. A. Ureña-López *et al.*, Updated cosmological constraints on axion dark energy with DESI, *Phys. Rev. D* **112**, 103505 (2025).
- [9] K. Lodha *et al.* (DESI Collaboration), DESI 2024: Constraints on physics-focused aspects of dark energy using DESI DR1 BAO data, *Phys. Rev. D* **111**, 023532 (2025).
- [10] F. Beutler and E. Di Dio, Modeling relativistic contributions to the halo power spectrum dipole, *J. Cosmol. Astropart. Phys.* **07** (2020) 048.
- [11] C. Bonvin, F. Lepori, S. Schulz, I. Tutusaus, J. Adamek, and P. Fosalba, A case study for measuring the relativistic dipole of a galaxy cross-correlation with the dark energy spectroscopic instrument, *Mon. Not. R. Astron. Soc.* **525**, 4611 (2023).
- [12] F. Lepori *et al.*, Euclid: Relativistic effects in the dipole of the two-point correlation function, *Astron. Astrophys.* **694**, A321 (2025).
- [13] J. R. Bermejo-Climent *et al.*, Constraints on primordial non-Gaussianity from the cross-correlation of DESI luminous red galaxies and Planck CMB lensing, *Astron. Astrophys.* **698**, A177 (2025).
- [14] E. Chaussidon *et al.*, Constraining primordial non-Gaussianity with DESI 2024 LRG and QSO samples, *J. Cosmol. Astropart. Phys.* **06** (2025) 029.
- [15] M. Abdul-Karim *et al.* (DESI Collaboration), DESI DR2 results II: Measurements of baryon acoustic oscillations and cosmological constraints, *Phys. Rev. D* **112**, 083515 (2025).
- [16] M. Abdul-Karim *et al.* (DESI Collaboration), DESI DR2 results I: Baryon acoustic oscillations from the Lyman alpha forest, *Phys. Rev. D* **112**, 083514 (2025).
- [17] F. Bernardreau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Large-scale structure of the Universe and cosmological perturbation theory, *Phys. Rep.* **367**, 1 (2002).
- [18] P. McDonald and A. Roy, Clustering of dark matter tracers: Generalizing bias for the coming era of precision LSS, *J. Cosmol. Astropart. Phys.* **08** (2009) 020.
- [19] K. F. Werner and C. Porciani, Renormalization of linear halo bias in N-body simulations, *Mon. Not. R. Astron. Soc.* **492**, 1614 (2020).
- [20] F. S. Kitaura, A. Balaguera-Antolínez, F. Sinigaglia, and M. Pellejero-Ibáñez, The cosmic web connection to the dark matter halo distribution through gravity, *Mon. Not. R. Astron. Soc.* **512**, 2245 (2022).
- [21] Y. B. Zel'dovich, Gravitational instability: An approximate theory for large density perturbations, *Astron. Astrophys.* **5**, 84 (1970), <https://ui.adsabs.harvard.edu/abs/1970A%26A...5...84Z/abstract>.
- [22] B. D. Sherwin and M. Zaldarriaga, Shift of the baryon acoustic oscillation scale: A simple physical picture, *Phys. Rev. D* **85**, 103523 (2012).
- [23] C.-H. Chuang, G. Yepes, F.-S. Kitaura, M. Pellejero-Ibanez, S. Rodríguez-Torres, Y. Feng, R. B. Metcalf, R. H. Wechsler, C. Zhao, C.-H. To, S. Alam, A. Banerjee, J. DeRose, C. Giocoli, A. Knebe, and G. Reyes, UNIT project: Universe N-body simulations for the investigation of theoretical models from galaxy surveys, *Mon. Not. R. Astron. Soc.* **487**, 48 (2019).
- [24] V. Springel, N. Yoshida, and S. D. M. White, GADGET: A code for collisionless and gasdynamical cosmological simulations, *New Astron.* **6**, 79 (2001).
- [25] V. Springel, The cosmological simulation code GADGET-2, *Mon. Not. R. Astron. Soc.* **364**, 1105 (2005).
- [26] Y. Feng, M.-Y. Chu, U. Seljak, and P. McDonald, FastPM: A new scheme for fast simulations of dark matter and haloes, *Mon. Not. R. Astron. Soc.* **463**, 2273 (2016).
- [27] X. Xu, N. Padmanabhan, D. J. Eisenstein, K. T. Mehta, and A. J. Cuesta, A 2 per cent distance to $z = 0.35$ by reconstructing baryon acoustic oscillations—II. Fitting techniques, *Mon. Not. R. Astron. Soc.* **427**, 2146 (2012).
- [28] M. Vargas-Magaña, S. Ho, X. Xu, A. G. Sánchez, R. O'Connell, D. J. Eisenstein, A. J. Cuesta, W. J. Percival, A. J. Ross, E. Aubourg, J. R. Brownstein, S. Escoffier, D. Kirkby, M. Manera, D. P. Schneider, J. L. Tinker, and B. A. Weaver, The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Potential systematics in fitting of baryon acoustic feature, *Mon. Not. R. Astron. Soc.* **445**, 2 (2014).
- [29] C. Zhao, C.-H. Chuang, F.-S. Kitaura, Y. Liang, M. Pellejero-Ibanez, C. Tao, M. Vargas-Magaña, A. Variu, and G. Yepes, Improving baryon acoustic oscillation measurement with the combination of cosmic voids and galaxies, *Mon. Not. R. Astron. Soc.* **491**, 4554 (2020).
- [30] C. Zhao, A. Variu, M. He, D. Forero-Sánchez, A. Tamone, C.-H. Chuang, F.-S. Kitaura, C. Tao, J. Yu, J.-P. Kneib, W. J. Percival, H. Shan, G.-B. Zhao, E. Burtin, K. S. Dawson, G. Rossi, D. P. Schneider, and A. de la Macorra, The completed SDSS-IV extended baryon oscillation spectroscopic survey: Cosmological implications from multitracers BAO analysis with galaxies and voids, *Mon. Not. R. Astron. Soc.* **511**, 5492 (2022).
- [31] M. Schmittfull, T. Baldauf, and M. Zaldarriaga, Iterative initial condition reconstruction, *Phys. Rev. D* **96**, 023505 (2017).
- [32] R. Hada and D. J. Eisenstein, An iterative reconstruction of cosmological initial density fields, *Mon. Not. R. Astron. Soc.* **478**, 1866 (2018).

- [33] F. S. Kitaura, The initial conditions of the universe from constrained simulations, *Mon. Not. R. Astron. Soc.* **429**, L84 (2013).
- [34] J. Jasche and B. D. Wandelt, Bayesian physical reconstruction of initial conditions from large-scale structure surveys, *Mon. Not. R. Astron. Soc.* **432**, 894 (2013).
- [35] H. Wang, H. J. Mo, X. Yang, and F. C. van den Bosch, Reconstructing the initial density field of the local universe: Methods and tests with mock catalogs, *Astrophys. J.* **772**, 63 (2013).
- [36] J. Jasche and G. Lavaux, Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe, *Astron. Astrophys.* **625**, A64 (2019).
- [37] F.-S. Kitaura, M. Ata, S. A. Rodríguez-Torres, M. Hernández-Sánchez, A. Balaguera-Antolínez, and G. Yepes, COSMIC BIRTH: Efficient Bayesian inference of the evolving cosmic web from galaxy surveys, *Mon. Not. R. Astron. Soc.* **502**, 3456 (2021).
- [38] P. Rosselló, F. S. Kitaura, D. Forero-Sánchez, F. Sinigaglia, and G. Favole, Differentiable fuzzy cosmic-web for field level inference, *Astron. Astrophys.* **709**, A110 (2026).
- [39] A. Klypin, F. Prada, and J. Byun, Suppressing cosmic variance with paired-and-fixed cosmological simulations: Average properties and covariances of dark matter clustering statistics, *Mon. Not. R. Astron. Soc.* **496**, 3862 (2020).
- [40] F.-S. Kitaura, C.-H. Chuang, Y. Liang, C. Zhao, C. Tao, S. Rodríguez-Torres, D. J. Eisenstein, H. Gil-Marín, J.-P. Kneib, C. McBride, W. J. Percival, A. J. Ross, A. G. Sánchez, J. Tinker, R. Tojeiro, M. Vargas-Magana, and G.-B. Zhao, Signatures of the primordial universe from its emptiness: Measurement of baryon acoustic oscillations from minima of the density field, *Phys. Rev. Lett.* **116**, 171301 (2016).
- [41] Y. Liang, C. Zhao, C.-H. Chuang, F.-S. Kitaura, and C. Tao, Measuring baryon acoustic oscillations from the clustering of voids, *Mon. Not. R. Astron. Soc.* **459**, 4020 (2016).
- [42] C. Zhao, C. Tao, Y. Liang, F.-S. Kitaura, and C.-H. Chuang, DIVE in the cosmic web: Voids with Delaunay triangulation from discrete matter tracer distributions, *Mon. Not. R. Astron. Soc.* **459**, 2670 (2016).
- [43] F. Sinigaglia, F.-S. Kitaura, K. Nagamine, Y. Oku, and A. Balaguera-Antolínez, Field-level Lyman- α forest modeling in redshift space via augmented nonlocal fluctuating Gunn-Peterson approximation, *Astron. Astrophys.* **682**, A21 (2024).
- [44] J. M. Coloma-Nadal, F.-S. Kitaura, J. E. García-Farieta, F. Sinigaglia, G. Favole, and D. Forero Sánchez, The hierarchical cosmic web and assembly bias, *J. Cosmol. Astropart. Phys.* **07** (2024) 083.
- [45] F. Sinigaglia, F.-S. Kitaura, K. Nagamine, and Y. Oku, The negative baryon acoustic oscillation shift in the Ly α forest from cosmological simulations, *Astrophys. J. Lett.* **971**, L22 (2024).