

Volume statistics as a probe of large-scale structure

Kwan Chuen Chan^{1,*} and Nico Hamaus²

¹*School of Physics and Astronomy, Sun Yat-sen University, 2 Daxue Road, Tangjia, Zhuhai 519082, China*

²*Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians Universität, Scheinerstraße 1, 81679 München, Germany*

 (Received 30 October 2020; accepted 14 January 2021; published 1 February 2021)

We investigate the application of volume statistics to probe the distribution of underdense regions in the large-scale structure of the Universe. This statistic measures the distortion of Eulerian volume elements relative to Lagrangian ones and can be built from tracer particles using tessellation methods. We apply Voronoi and Delaunay tessellation to study the clustering properties of density and volume statistics. Their level of shot-noise contamination is similar, as both methods take into account all available tracer particles in the field estimator. The tessellation causes a smoothing effect in the power spectrum, which can be approximated by a constant window function on large scales. The clustering bias of the volume statistic with respect to the dark matter density field is determined and found to be negative. We further identify the baryon acoustic oscillation (BAO) feature in the volume statistic. Apart from being smoothed out on small scales, the BAO is present in the volume power spectrum as well, without any systematic bias. These observations suggest that the exploitation of volume statistics as a complementary probe of cosmology is very promising.

DOI: [10.1103/PhysRevD.103.043502](https://doi.org/10.1103/PhysRevD.103.043502)

I. INTRODUCTION

Modern large-scale structure surveys probe both the background expansion and the growth of structure, which can be used to constrain the energy content of the Universe and help resolve the mystery behind the cosmic expansion [1]. These surveys have been remarkably successful in constraining the parameters in Λ CDM and its extensions [2–9]. Because most observed galaxies trace the high-density regions of the Universe [10], previous analyses typically focus on the clustering statistics of overdensities. However, the dominant part of the Universe exhibits a low density compared to its mean value. The clustering of underdensities may therefore contain additional information that is complementary to that obtained via conventional methods.

Cosmic voids are extended regions of relatively low matter content in the large-scale structure; they are potentially ideal proxies for studying the underdense parts of the Universe [11–13]. Because of their unique environment, interesting phenomena such as the dark energy [14–17], modified gravity [18–22], and the influence of massive neutrinos [23–26] become more visible in voids. Recently, there have been numerous studies on the clustering of voids using samples derived from cosmological surveys, exploiting effects such as redshift-space distortions (RSDs) in the void-galaxy correlation function [27–35], the

Alcock-Paczynski test [36–41], and the baryon acoustic oscillations (BAOs) [42,43]. However, the spatial number density of voids is generally low due to their large size; hence, their auto-clustering statistic exhibits a considerable shot-noise contribution [44–46]. While small voids are more abundant, they actually reside in regions of higher density (so-called void-in-clouds [13]), only larger voids truly trace the underdense regime of large-scale structure. To mitigate the shot noise, cross-correlations with other more abundant tracers, such as galaxies [44] or clusters [47,48], are considered. Another option is the use of overlapping spherical voids, as demonstrated by Ref. [43] in the measurement of the BAO in the void auto-correlation function. Allowing voids to overlap increases their number density by up to 2 orders of magnitude, but the signal-to-noise ratio of the resulting clustering statistic is more difficult to interpret.

In practice voids are identified via sparse tracers, as the full mass distribution cannot be observed directly in all three dimensions. Since many tracers are needed to define a single void, the resulting void distribution is even more sparse. Therefore, the question arises whether one can construct a field estimator for the low-density regions with low level of shot noise and without overlap. The *volume statistic* proposed in this paper satisfies these conditions. In Sec. II, we define the volume statistic in detail. In order to implement it in data, we need to estimate the density field using tessellation methods. We briefly review two tessellation interpolation methods and contrast them with the

*chankc@mail.sysu.edu.cn

conventional mass assignment scheme in Sec. III. The possibility of approximating its impact on the power spectrum by an effective window is discussed in Sec. IV. Section V is devoted to studying the overall clustering properties of the volume statistic, such as the bias parameter and the BAO feature. We conclude in Sec. VI. In the Appendices, we review the kernel density estimation (KDE) method (Appendix A) and the effects of a window function on the power spectrum (Appendix B).

II. VOLUME STATISTICS

Voids are extended underdense regions in the Universe and it has been shown that they furnish a biased tracer of the large-scale structure [13,25,44,49–53]. Small voids tend to reside in overdense environments and therefore exhibit positive bias parameters; they are in the so-called void-in-cloud regime [13]. Large voids, on the contrary, are anticorrelated with overdense structures and thus exhibit a negative clustering bias [44], i.e., they trace large-scale underdensities (it is interesting to note that the quadratic bias of voids follows a similar trend to that of halos [52]). However, their number density is significantly lower than that of small voids (owing to their size), which results in a high shot-noise contribution and hence a low signal-to-noise ratio of their large-scale auto-power spectrum. In this work, instead of defining single objects from a large number of tracers, we assign a volume to each individual tracer particle, which helps to mitigate the shot noise.

In the Lagrangian picture of large-scale structure, we can think of each tracer particle carrying a volume element that is being distorted over time. This evolution is governed by mass conservation, and for the case of dark matter we have [54]

$$d^3q = (1 + \delta)d^3x, \quad (1)$$

where d^3q and d^3x are the volume elements in the Lagrangian (initial) and Eulerian (evolved) space, respectively, and δ is the matter density contrast in Eulerian space. Then we define

$$\mathcal{J} \equiv \frac{d^3x}{d^3q} = \frac{1}{1 + \delta}. \quad (2)$$

\mathcal{J} is the Jacobian of the transformation from the Lagrangian coordinate \mathbf{q} to the Eulerian coordinate \mathbf{x} . It quantifies the change in volume compared to that in the Lagrangian space. When the displacement is small, \mathcal{J} reduces to $1 - \delta$ with a symmetric distribution around a mean equal to 1. For large displacements and hence nonlinear δ , \mathcal{J} has a mean larger than 1 and the distribution becomes skew symmetric with a singular tail at $\delta = -1$. This indicates that inside empty regions in Eulerian space, a tremendous amount of expansion in volume has occurred. After shell crossing, the volume element d^3q can split into

multiple parts and the fluid description by Eq. (1) is no longer valid. Nonetheless, shell crossing occurs mainly in the high-density regions, which are down-weighted in \mathcal{J} , and hence its impact on our statistic is minimal.

In analogy to the density contrast, we can define the volume statistic as

$$\mathcal{V} \equiv \frac{\mathcal{J}}{\bar{\mathcal{J}}} - 1 = \frac{1}{\bar{\mathcal{J}}(1 + \delta)} - 1, \quad (3)$$

where $\bar{\mathcal{J}}$ is the mean of \mathcal{J} . For small δ , it reduces to $-\delta$.

For halos, the situation is more complicated, because the density is no longer uniform in Lagrangian space, and we have instead (e.g., [55,56])

$$(1 + \delta_L)d^3q = (1 + \delta_h)d^3x, \quad (4)$$

where δ_L and δ_h are the halo-density contrast in Lagrangian and Eulerian space, respectively. However, in observations, we have no direct access to δ_L . One practical way to bypass this difficulty is to define \mathcal{J} as

$$\mathcal{J} \equiv \frac{d^3x}{(1 + \delta_L)d^3q} = \frac{1}{1 + \delta_h}, \quad (5)$$

even for the case of halos. We can think of it as the volume change inversely weighted by the halo density in Lagrangian space.

The advantage of the volume statistic is as follows. First, it is a fundamental measure of the volume change induced by the large-scale structure evolution. From a theoretical perspective, it is likely more amenable to models than the statistics of voids, which are defined by the complicated nonlinear topology of the cosmic web. In the linear regime, the volume statistic reduces to the one of linear density fluctuations. In the weakly nonlinear regime, it can be treated with perturbation theory or other analytic methods. Second, the volume statistic makes use of all the available tracer particles, resulting in a shot-noise level comparable to its original tracer field. Nevertheless, because the majority of particles are located inside collapsed regions, the sampling is still poor compared to the conventional density statistic. On the other hand, because of the denominator $1 + \delta$, \mathcal{V} is undefined when $\delta = -1$. This can pose problems in the density estimation for sparse samples, such as massive halos. We will therefore consider tessellation methods to avoid such singularities.

The volume statistic up-weights underdense regions and down-weights overdense ones; hence, it is similar to certain types of marked correlation functions in spirit.¹ If each particle is assigned a mark m , the marked correlation M for the mark can be generally defined as [59]

¹This can be generalized to the marked power spectrum [57,58].

$$M(r) = \frac{1}{N(r)\bar{m}^2} \sum_{ij} m_i m_j, \quad (6)$$

where $N(r)$ is the total number of pairs with separation r and \bar{m} is the mean of the mark. The mark can be chosen to highlight any particular physics of interest. For example, [60] proposed to consider a mark of the form

$$m = \left(\frac{\rho_* + 1}{\rho_* + 1 + \delta} \right)^p, \quad (7)$$

where ρ_* and p are free parameters. This mark has been applied to study modified gravity [61]. If we set $\rho_* = 0$ and $p = 1$, it reduces to the volume statistic. However, the value of ρ_* is often taken to be order unity, e.g., in [61], so as to avoid the singularity problem mentioned above.

Moreover, the definition of the volume statistic is similar to the log-transform of the density field [62]

$$A = \log(1 + \delta). \quad (8)$$

It was found that the field A is more Gaussian than δ at late times, and some higher-order information in δ is pulled back into the two-point statistics of A [62,63]. Both transformations up-weight the low-density regions and down-weight the high-density ones. In fact, both the log-field and the volume statistic diverge at $\delta = -1$. The log-transform aims to recover the information in the initial density field, as it accommodates both the low-density and the high-density regions. Our goal is less ambitious and we merely want to recover the information from the underdense regions. By restricting our target, it is easier to define a density reconstruction method that suits precisely this purpose.

III. INTERPOLATION METHODS

In order to compute the volume statistic, one needs a density estimate that is finite everywhere. To this end, we use interpolation by tessellation to reconstruct the density field from a discrete particle distribution. As this type of method is rarely used directly for clustering analyses, we shall first compare it against the commonly used mass-interpolation method.

A. Mass interpolation

For the large-scale clustering analysis, the density field is often interpolated to a grid.² The interpolation methods usually employed include nearest grid point (NGP), cloud-in-cell (CIC), and triangular-shaped cloud (TSC). We collectively denote them as mass-interpolation methods; they are discussed extensively in the monograph [64].

²The configuration-space estimation of correlation functions uses counting of pairs directly without the need for a grid.

In these methods, particles are smoothed by a window function (or cloud function) W of fixed size and these windows can overlap with each other. The density at the grid point x_p is given by

$$\rho(x_p) = \sum_{\alpha} \int_{[x_p]} d^3x W(x - x_{\alpha}), \quad (9)$$

where $[x_p]$ denotes the grid cell enclosing x_p and the sum is over all the particles.

The three-dimensional (3D) window can be constructed from its 1D form as $W(\mathbf{x}) = W(x)W(y)W(z)$. In 1D, the NGP window function is the Dirac delta, for CIC it is a top hat, and for TSC it is a triangular-shaped cloud centered on the particle position. In fact, these are the first three members of an infinite hierarchy of interpolating functions that can be generated via iterative convolution with the top-hat window function [64]. In NGP, the particle mass is interpolated to the nearest grid point, in CIC the mass is assigned to the grid points of the cell enclosing the particle (8 points in 3D), and the particle mass is interpolated to the neighboring 27 points in TSC. Thus, the window becomes more extended as the order of the interpolation scheme increases. Because the window function is a constant linear operator, its effect can be removed by division of W in Fourier space. To represent the density field for clustering analyses, the grid size is usually chosen sufficiently large to mitigate resolution effects.³ Among the mass-interpolation methods, CIC is the most common one, so we use it as our benchmark.

B. Tessellation interpolation

The tessellation method can generate a space-filling field with finite values (nonvanishing) everywhere in configuration space; it is often used as an intermediate step to obtain a smooth density field and to define extended structures in the cosmic web, e.g., [67–73]. However, it has rarely been used directly for clustering analyses of the density field. Given a set of points (generators) x_{α} , a tessellation generates cells that are space-filling and mutually disjoint [74,75]. Applications of the tessellation methods in cosmology were introduced in [75–80].

In the Voronoi tessellation, each Voronoi cell encloses one generator and every point x belongs to a Voronoi cell V_{α} if the distance $|x - x_{\alpha}|$ is smaller than $|x - x_{\beta}|$ for any other β . The simplest possibility is to assume that the particle mass spreads out uniformly within a Voronoi cell and the density within each cell is estimated by the mass of the particle and the volume of the Voronoi cell. This method

³Discrete sampling of the continuous field introduces aliasing effect [64–66], which is especially serious for modes close to the Nyquist frequency. We are limited to the large-scale modes because of the tessellation interpolation, we shall not consider it here.

yields a piecewise constant field and the field value is discontinuous across the boundaries. We shall refer to the technique to estimate densities in this way as the Voronoi tessellation field estimator (VTFE).

To obtain a continuous field, we can tessellate the space into Delaunay cells. In 3D, a triangulation divides space into tetrahedrons. A Delaunay tessellation is such a triangulation with the additional property that the circumsphere of each tetrahedron does not contain any generators. The field value inside a tetrahedron can then be obtained by linear interpolation from the vertices. Explicitly, the number density at a point \mathbf{x} inside a Delaunay cell with vertices \mathbf{x}_0 , \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 , is given by [75]

$$n(\mathbf{x}) = n(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)_i J_{i\alpha}^{-1} \Delta n_\alpha, \quad (10)$$

where Δn_α indicates the difference in density between vertex \mathbf{x}_α and the reference vertex \mathbf{x}_0 , $n(\mathbf{x}_\alpha) - n(\mathbf{x}_0)$. A Latin index denotes the spatial component of the vector and a Greek index represents the vertex (also from 1 to 3). The Jacobian J is the transformation matrix between the coordinate \mathbf{x} and the barycentric coordinate ζ_α ,

$$J_{i\alpha} = \frac{\partial x_i}{\partial \zeta_\alpha} \approx (\mathbf{x}_\alpha - \mathbf{x}_0)_i. \quad (11)$$

The number density at the generator can be estimated by [75,80]

$$n(\mathbf{x}_\alpha) = \frac{1 + D}{\sum_{\mathcal{D}_{x_\alpha}^{\text{adj}}} \mathcal{V}_{\text{adj}}(\mathbf{x}_\alpha)}, \quad (12)$$

where D is the dimension of space, $\mathcal{D}_{x_\alpha}^{\text{adj}}$ denotes all the adjacent Delaunay cells with \mathbf{x}_α as one of their generators, and \mathcal{V}_{adj} is the volume of the adjacent Delaunay cells. This construction ensures that the integral of the density field obtained from Eq. (10) is equal to the total particle mass. We refer to this method of estimating the density as the Delaunay tessellation field estimator (DTFE) in this paper.

Effectively, tessellation methods smooth out the particle mass adaptively based on the local density. For high-density regions, the effective window size is small, while it is large in the low-density regime. For the mass-interpolation methods, such as the CIC, as the size of the grid increases the resulting density field approaches a sum of Dirac delta distributions. Thus, this method can produce the same type of distribution as the point set. On the other hand, the tessellation method aims at reconstructing a smooth (piecewise constant for VTFE and continuous for DTFE) density field from the discrete particle distribution. This additional requirement introduces a smoothing to the small-scale power spectrum. The smoothing scale is dictated by the mean particle separation of the distribution. In summary, in the mass-interpolation scheme, the effective window-function size is controlled by the grid size, while

that of the tessellation method is determined by the mean particle separation.

C. Effects of the tessellation: An analytic estimate

In the mass-interpolation scheme, a fixed window is used to smooth the particle distribution. An analytic estimate can be made if the particles are assumed to Poisson sample the underlying density field.⁴ The density estimated by smoothing the particle distribution is akin to the KDE method in statistics, which is used to estimate the probability density distribution (see, e.g., [83]). In Appendix A, we review the KDE and show that given a sample of particles it can be used to estimate the number density of the sample, n [Eq. (A11)]. In particular, it is shown that smoothing introduces a bias to the density estimate [Eq. (A12)]. At a peak (trough) of the density field, the window-estimated density field is biased low (high). However, the advantage of smoothing the density field is that it can suppress the variance of the estimator [Eq. (A13)]: it is inversely proportional to the average volume of the window size. In statistics, the mean squared error, which is a sum of the variance and the bias squared, matters if the signal-to-noise ratio of the measurement is not high. Thus, the KDE is useful if its suppression to the variance is larger than its bias.

For the case of density estimation by tessellation, the density at the position \mathbf{x} depends on the particles spanning the cell. It is hard to make an analytic estimate of its bias and variance, because of the irregular shape of the cell and the interpolation scheme depending on all the points in the cell. For the case of VTFE, because there is only one generator inside a Voronoi cell, we can write down the density estimate as ($N = 1$)

$$\hat{n}_{\text{VTFE}}(\mathbf{x}) = \frac{1}{h_\alpha^3} W_{\text{VTFE}} \left(\frac{\mathbf{x} - \mathbf{x}_\alpha}{h_\alpha} \right), \quad (13)$$

where W_{VTFE} describes the Voronoi cell around a point \mathbf{x} , \mathbf{x}_α is the generator of the Voronoi cell, and h_α schematically represents the characteristic size of the cell.

Because the Voronoi cell is generally not symmetric about its generator, the bias reads

$$\langle \hat{n}(\mathbf{x}) \rangle - n(\mathbf{x}) = -h_\alpha \sum_i I_i^{(1)} \frac{\partial n}{\partial x_i} + \frac{h_\alpha^2}{2} \sum_{ij} \frac{\partial^2 n}{\partial x_i \partial x_j} I_{ij}^{(1)}, \quad (14)$$

where the function I is defined in general in Eq. (A6). In particular, because of the irregular shape of the Voronoi cells, $I^{(1)}$ is in general nonvanishing. However, if we now further average over the shape of the Voronoi cell W_{VTFE} , $\langle I_i^{(1)} \rangle$ vanishes and we obtain

⁴In the case of halos and voids, there are exclusion effects that violate this assumption [44,49,81,82].

$$\langle \hat{n}(\mathbf{x}) \rangle - n(\mathbf{x}) = \frac{\langle h_\alpha^2 \rangle}{2} \sum_{ij} \frac{\partial^2 n}{\partial x_i \partial x_j} \langle I_{ij}^{(1)} \rangle. \quad (15)$$

Similar to the standard KDE case [Eq. (A12)], we recover the result that the bias is sourced by the second derivatives of the density. The variance of the estimator reads

$$\begin{aligned} \text{Var}(\hat{n}(\mathbf{x})) &= \frac{1}{h_\alpha^6} \text{Var} \left(W_{\text{VTFE}} \left(\frac{\mathbf{x} - \mathbf{x}_\alpha}{h_\alpha} \right) \right) \\ &\approx \frac{I^{(2)}}{h_\alpha^3} n(\mathbf{x}). \end{aligned} \quad (16)$$

Further averaging over the shape of the Voronoi cell, we have

$$\text{Var}(\hat{n}(\mathbf{x})) = \left\langle \frac{1}{h_\alpha^3} \right\rangle \langle I^{(2)} \rangle n(\mathbf{x}). \quad (17)$$

For DTFE, the density at \mathbf{x} [Eq. (10)] depends on the position of the vertices nonlinearly and the previous analytic arguments do not apply. Nonetheless, we shall see that the features of the DTFE are qualitatively similar to those of the VTFE.

IV. EFFECTIVE WINDOW FUNCTION

In this section, we investigate whether the effects of the tessellation can be modeled by an effective window function. The width of the window involved in the mass interpolation is controlled by the grid size, which can be chosen such that the grid effects are negligible, or the window function effect can be divided out in Fourier space. Hence, we may treat the density field obtained from the CIC, δ_{CIC} , as the unsmoothed field for comparison.

In Appendix B, we review the effects of the window function on the power spectrum. We show that if the particle distribution is smoothed by a constant window function W , the effect on the power spectrum is captured by an overall factor of $|W|^2$ (or W for the case of cross-power spectrum). Note that the same window function factor applies to both the continuous clustering and the shot-noise term. Thus, if the effective window function can be taken to be a nonstochastic function, the power spectrum of the density field obtained with the tessellation method, δ_{tess} , can be written as⁵

$$\langle \delta_{\text{tess}}(\mathbf{k}) \delta_{\text{tess}}(\mathbf{k}') \rangle = (2\pi)^6 |W_{\text{tess}}(\mathbf{k})|^2 \langle \delta_{\text{CIC}}(\mathbf{k}) \delta_{\text{CIC}}(\mathbf{k}') \rangle. \quad (18)$$

Similarly, for consistency, the cross-correlation is given by

$$\langle \delta_{\text{tess}}(\mathbf{k}) \delta_{\text{CIC}}(\mathbf{k}') \rangle = (2\pi)^3 W_{\text{tess}}(\mathbf{k}) \langle \delta_{\text{CIC}}(\mathbf{k}) \delta_{\text{CIC}}(\mathbf{k}') \rangle. \quad (19)$$

We test these relations using simulation data, the details of which are described in the following. The assumed cosmology is a flat Λ CDM model with parameters $\Omega_m=0.3$, $\Omega_\Lambda=0.7$, $h_0=0.7$, $n_s=0.967$, and $\sigma_8=0.85$. Each simulation consists of a cubic box of $1000 \text{ Mpc } h^{-1}$ side length and contains 512^3 particles. The Gaussian initial conditions are generated by CLASS [84] at redshift $z=49$. In order to study the BAO signature, we ran another set of simulations with Einsteinian-Hu initial conditions without the BAO wiggles [85] (NoWiggle). The particle displacements are implemented using 2LPTIC [86] and are evolved with the N -body code GADGET2 [87]. Halos are identified using the halo finder AHF [88]. They are defined via a spherical overdensity threshold of 200 times the background density and contain at least 20 particles. For each simulation set, 20 realizations are used. The results derived from the halo samples will be our primary interest in this paper. We shall illustrate our results using four halo samples at $z=0$ and 1, with two sets of initial conditions, respectively. Unless otherwise stated, the error bars are the standard error of the mean of the measurements among all realizations.

The procedures for density estimation using the tessellation method are similar to those in the public code DTFE [89], which implements the density estimation by the Delaunay tessellation efficiently. We implement both Voronoi and Delaunay tessellation methods using the python libraries based on the Qhull code [90]. After generating the tessellation cells, we use a cubic grid to estimate the density at the grid points, which can then be used for the estimation of power spectra. We use a grid size of 256^3 and average over 2^3 uniformly spaced sampling points within each grid cell to calculate the mean density at the grid point. We have checked that these choices are sufficiently accurate for our results.

A. Measurement of the effective window

We measure the effective window function of the tessellation using random catalogs and halo samples. Without intrinsic clustering, it is easier to identify the impact of the effective window on the power spectrum, since the remaining clustering amplitude is simply white noise. To generate the random catalogs, particle positions are randomly placed in a cubic box of size $1000 \text{ Mpc } h^{-1}$ until the desired number density is reached (we match it with the halo sample density). Assuming that halos Poisson sample the underlying continuous density field, the shot-noise contribution to their power spectrum is given by [91]

$$P_{\text{shot}} = \frac{1}{(2\pi)^3 \bar{n}}, \quad (20)$$

⁵The Fourier convention used in this paper is $f(\mathbf{k}) = \int \frac{d^3 \mathbf{x}}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x})$ and $f(\mathbf{x}) = \int d^3 k e^{i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{k})$.

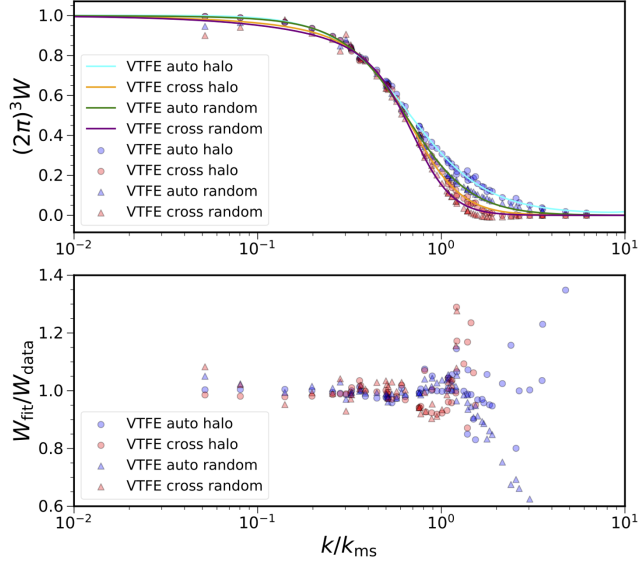


FIG. 1. Top: the VTFE effective window function measured from the random catalog (triangles) and the halos (circles). Measurements from the auto-power spectrum (blue) and the cross-power spectrum (red) are shown. The fits using Eq. (21) are overplotted (cyan for W_a from halos, green for W_a from randoms, orange for W_c from halos, and purple for W_c from randoms). Bottom: the ratio between the fits and the measurements (circles for halos and triangles for randoms, blue for auto- and red for cross-power spectra).

where \bar{n} is the mean halo density. We show in Appendix B that there is an additional factor of $|W|^2$, if the particle distribution is smoothed by a window W .

The effective VTFE window obtained using Eqs. (18) and (19) from the random catalog is shown in Fig. 1. In this case, the CIC power spectrum is given by Eq. (20). The size of the effective window can be characterized by the mean separation wave number $k_{\text{ms}} \equiv \pi\bar{n}^{1/3}$. We find that using the variable $q \equiv k/k_{\text{ms}}$, the effective window obtained from different tracers falls on a universal curve. Thus, k_{ms} represents the characteristic size of the tessellation cell in Fourier space.⁶ Although the window from the auto- and cross-power spectra (denoted as W_a and W_c , respectively) agree with each other at low k , W_a is more extended than W_c . In fact, W_c turns negative at $q \sim 1.8$, while W_a remains positive everywhere. Equations (18) and (19) imply that we can extract the same window function from the auto- and cross-power spectrum measurements. However, in practice, this only works on large scales, so the assumption that $W(\mathbf{k})$ factorizes and does not explicitly depend on the density field $\delta(\mathbf{k})$ must fail on smaller scales.

⁶We note that in Lagrangian space, the typical extension of halo-density profiles is controlled by halo mass. When the density profile is rescaled by this characteristic extension, most of the mass dependence is removed and the halo-density profiles fall on a rather universal curve [92]. Thus, k_{ms} is analogous to the role of mass for Lagrangian halos.

TABLE I. The best-fit parameters for Eq. (21) for the VTFE and DTFE methods. The results derived from W_a and W_c for the random and halo samples are shown.

	VTFE random		VTFE halo		DTFE random		DTFE halo	
	W_a	W_c	W_a	W_c	W_a	W_c	W_a	W_c
a	0.09	0.50	-0.08	0.37	0.07	0.37	-0.09	0.20
b	1.49	1.36	1.55	1.35	1.36	1.29	1.37	1.28
n	2.47	3.94	2.08	3.14	2.75	4.18	2.20	3.25

Even if the particle distribution is intrinsically clustered, the window function still factors out. In Fig. 1, we also show the effective window function extracted from $P_{\text{tess}}/P_{\text{CIC}}$ and $P_{\text{tess,CIC}}/P_{\text{CIC}}$, with P_{tess} and P_{CIC} being the halo auto-power spectra obtained with the tessellation and the CIC method, and $P_{\text{tess,CIC}}$ the cross-power spectra between them. Note that we have not subtracted shot noise in any of those cases. The curves from halo samples are somewhat less universal than for the randoms, especially at high q . This is expected, as the halos exhibit intrinsic clustering that is scale dependent, making their effective window more extended due to stronger clustering on smaller scales.

We find that the overall window function shape can be described well by the functional form

$$(2\pi)^3 W(q) = \frac{1}{(1 + aq)[1 + (bq)^n]}, \quad (21)$$

with $q = k/k_{\text{ms}}$.

We have shown the best-fit values obtained from the VTFE and the DTFE in Table I. The results for the halo and random samples obtained with W_a and W_c are compared. These best fits are also plotted in Figs. 1 and 2. Up to $q \lesssim 1$, the accuracy of the fit is about 10% within the scatter of the data. As there are significant differences between W_a and W_c , obtaining a universal fit that is accurate at $q \gtrsim 1$ is not possible. Overall, the DTFE results are in qualitative agreement with the VTFE ones. However, we find that the measurements of W_{DTFE} are slightly more extended than W_{VTFE} . This implies that the DTFE window is more compact in configuration space.

B. Power spectrum ratio

After having measured the effective window functions of the tessellation methods, we are now in the position to test whether we can use this window to recover the unsmoothed power spectrum. We utilize the ratio

$$R = \frac{\frac{P_{\text{tess}}^X}{[(2\pi)^3 W(k)]^n} - P_{\text{shot}}}{P_{\text{CIC}} - P_{\text{shot}}} \quad (22)$$

to quantify the accuracy of this reconstruction, where X denotes auto- ($n = 2$) or cross-power spectra ($n = 1$). In Fig. 3, we show the ratio R for the auto-power spectrum

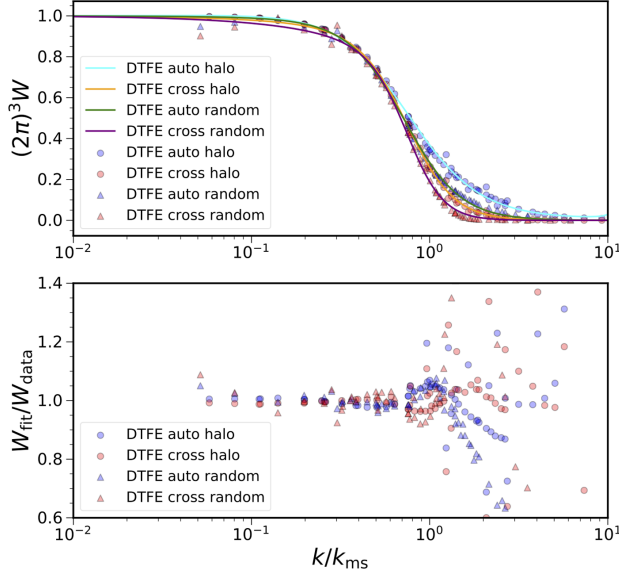


FIG. 2. As Fig. 1, but for the DTFE results.

using the VTFE windows W_a and W_c from halo and random samples. For the rest of this paper, we will exclusively show the results based on these four samples. At low k , W_a appears to yield a better agreement than W_c . However, inspection of Fig. 1 reveals that this can be explained by the fact that W_c is a bit lower than W_a at small q . This is an artifact of the fitting formula. On the other hand, when k is close to k_{ms} , the ratio deviates substantially from 1. A division by the window function is prone to noise and systematic error, which become inflated at high k when W is small. Even if we use W_c from a direct measurement,

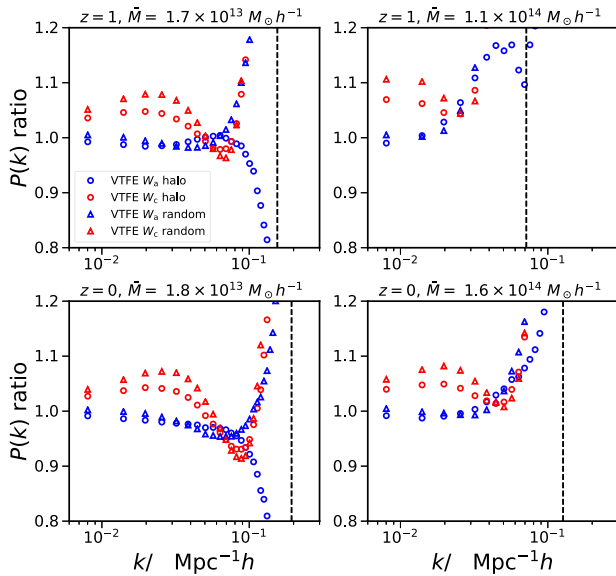


FIG. 3. Auto-power spectrum ratio Eq. (22) from the VTFE window for W_a (blue) and W_c (red) from the halo (circles) and random (triangles) samples. The mean mass and redshift of the halo samples are shown on top of each panel.

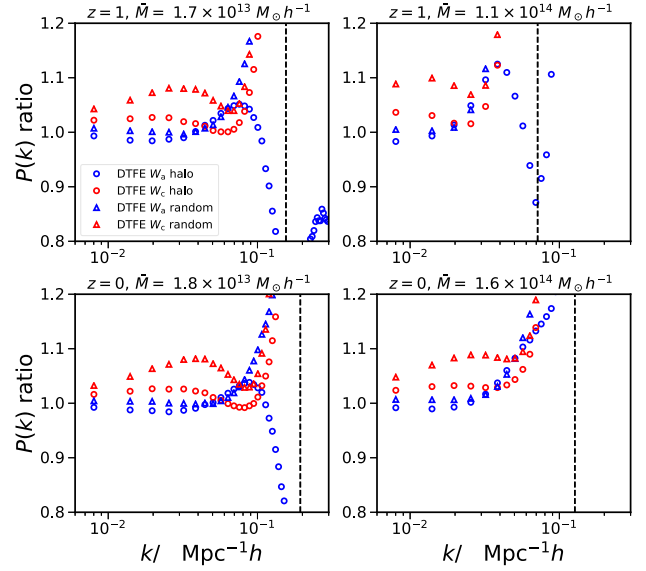


FIG. 4. As Fig. 3, but for the DTFE results.

the pattern at high k in R remains unchanged. Also, the trends are very similar for both the halo and the random samples. In Fig. 4, we show the corresponding results for the DTFE, which are qualitatively similar to the VTFE case. The ratios using the cross-power spectra in Eq. (22) are very similar, so we do not show them here.

Our results suggest that approximating the tessellation with a constant effective window function is valid on large scales, but above k_{ms} the density-dependent nature of the tessellation method is important and smoothing effects cannot be removed by division of a window function. In this case, it is necessary to determine the tessellation-smoothed density field numerically.

V. CLUSTERING OF THE VOLUME FIELD

A. The shot noise of the volume field

We estimate the shot noise of the volume field P_{Vn} using the random catalog again and showcase the auto-power spectra of the volume statistic before and after shot-noise subtraction in Fig. 5. Although the volume statistic is obtained via nonlinear transformation of the halo tracer field, on large scales the shot noise still approaches P_{shot} as given by Eq. (20). However, it is suppressed by the effects of the tessellation on smaller scales. After shot noise is subtracted from the power spectra, the results from the VTFE and DTFE agree with the shot-noise subtracted CIC case. Thus, since all available tracer particles are used to construct the volume field, its shot-noise contamination is indeed similar to that of the original tracer field.

B. The bias of the volume field

The volume field can be regarded as a biased tracer of the underlying dark matter density field. As it traces the

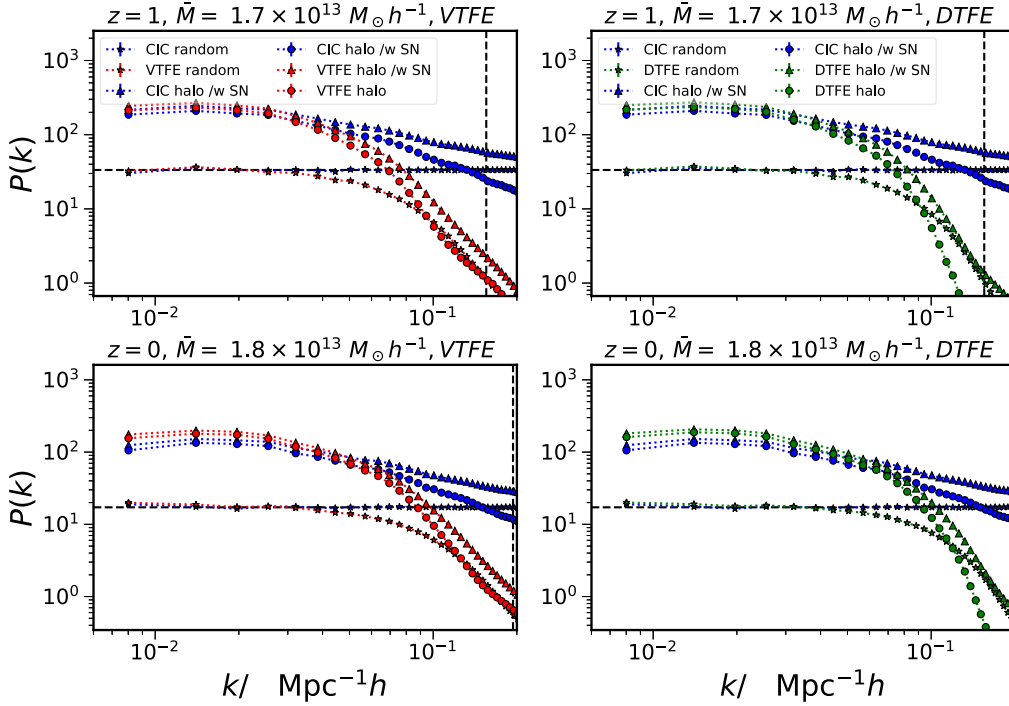


FIG. 5. Auto-power spectra of the volume statistic before (triangles) and after shot-noise subtraction (circles) for the VTFE (left panels, red) and the DTFE (right panels, green). The conventional halo-density power spectra are shown for comparison (blue), along with the shot-noise contamination estimated from the randoms (stars). The vertical dashed line indicates the mean tracer separation scale k_{ms} and the horizontal one shows the Poisson power spectrum P_{shot} . The mean mass and redshift of the halo samples are shown on top of each panel. On large scales, the shot-noise level of the volume field is comparable to that of the halo sample used to construct it.

underdense regions of large-scale structure, we expect it to exhibit a negative clustering bias, similar to large voids [13,44,49]. To guide our interpretation of the volume statistic, we first investigate the bias of the more common density field obtained via the tessellation method in Fig. 6. We compare the bias parameters b_{hh} and b_{hm} derived from the halo auto- and halo-matter cross-power spectra, P_{hh} and P_{hm} ,

$$b_{\text{hh}} = \sqrt{\frac{P_{\text{hh}} - P_{\text{hn}}}{P_{\text{mm}}}}, \quad (23)$$

$$b_{\text{hm}} = \frac{P_{\text{hm}}}{P_{\text{mm}}}, \quad (24)$$

where P_{mm} denotes the matter auto-power spectrum and P_{hn} is the halo shot-noise contribution obtained from the random catalog. As a comparison, we also show b_{hh} and b_{hm} obtained from the CIC method.

On large scales, the density-bias parameter obtained via the tessellation methods approaches the one obtained via the CIC mass assignment, which already reaches a constant value at $k \lesssim 0.08 \text{ Mpc}^{-1} h$. On smaller scales, the bias parameter from the tessellation is damped by the effective window function, as discussed in the previous section. The differences between b_{hh} and b_{hm} are apparent for $k \gtrsim 0.1 \text{ Mpc}^{-1} h$. Consistent with the findings in

Sec. IV A, Fig. 6 shows that the DTFE window is more compact than the VTFE window in configuration space. To fit the large-scale bias function, the scale dependence of the window must be taken into account and a possible functional form is a quartic polynomial

$$b(k) = c_0 + c_2 k^2 + c_4 k^4, \quad (25)$$

where c_0 , c_2 , and c_4 are the fit parameters. In Fig. 6, we have also plotted the best-fit curves, using modes up to $k_{\text{max}} = 0.06 \text{ Mpc}^{-1} h$, yielding a good agreement with the simulation data.

We now turn to the bias of the volume field, presented in Fig. 7. The auto- and cross-bias for the volume field is defined in analogy to Eqs. (23) and (24),

$$b_{\text{vv}} = -\sqrt{\frac{P_{\text{vv}} - P_{\text{vn}}}{P_{\text{mm}}}}, \quad (26)$$

$$b_{\text{vm}} = \frac{P_{\text{vm}}}{P_{\text{mm}}}, \quad (27)$$

with P_{vv} being the auto-power spectrum of the volume field and P_{vm} the cross-power spectrum between the volume field and the matter density field. The shot noise of the volume field P_{vn} is estimated using the random catalog again. Note that because the volume field

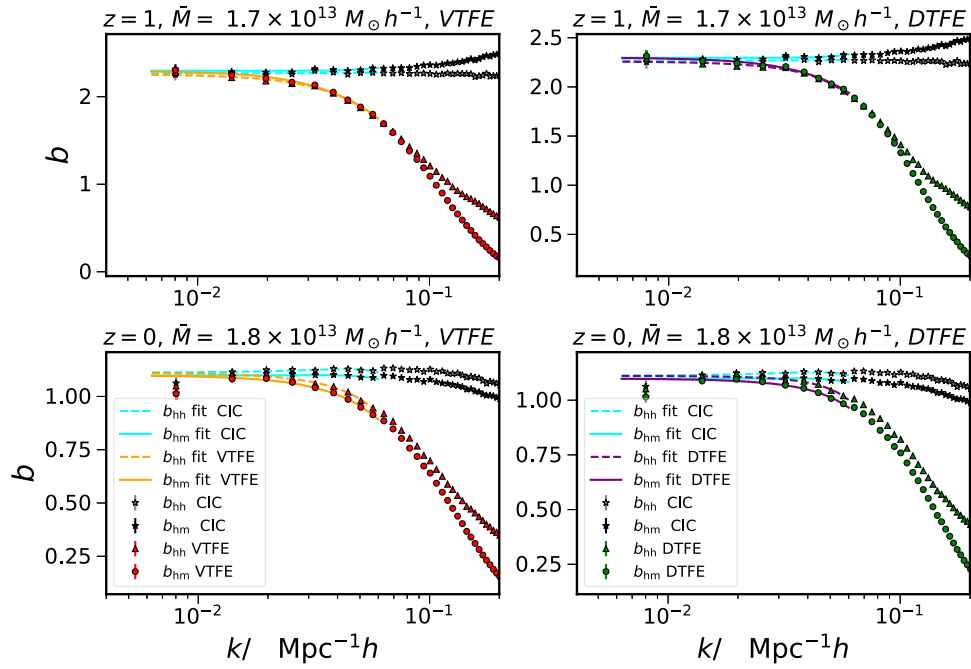


FIG. 6. The clustering bias of the halo-density field obtained with the CIC, VTFE (left panels), and DTFE (right panels) methods. The measurements originate from the halo auto-power spectrum, b_{hh} (CIC: gray stars, VTFE: red triangles, DTFE: green triangles) and the halo-matter cross-power spectrum, b_{hm} (CIC: black stars, VTFE: red circles, DTFE: green circles). The fit using Eq. (25) is overlotted, dashed for b_{hh} and solid for b_{hm} : cyan, orange, and purple for CIC, VTFE, and DTFE measurements, respectively. Redshifts and mean halo masses are indicated on top of each panel. On large scales, the bias measurements obtained with the tessellation methods are in good agreement with the CIC ones.

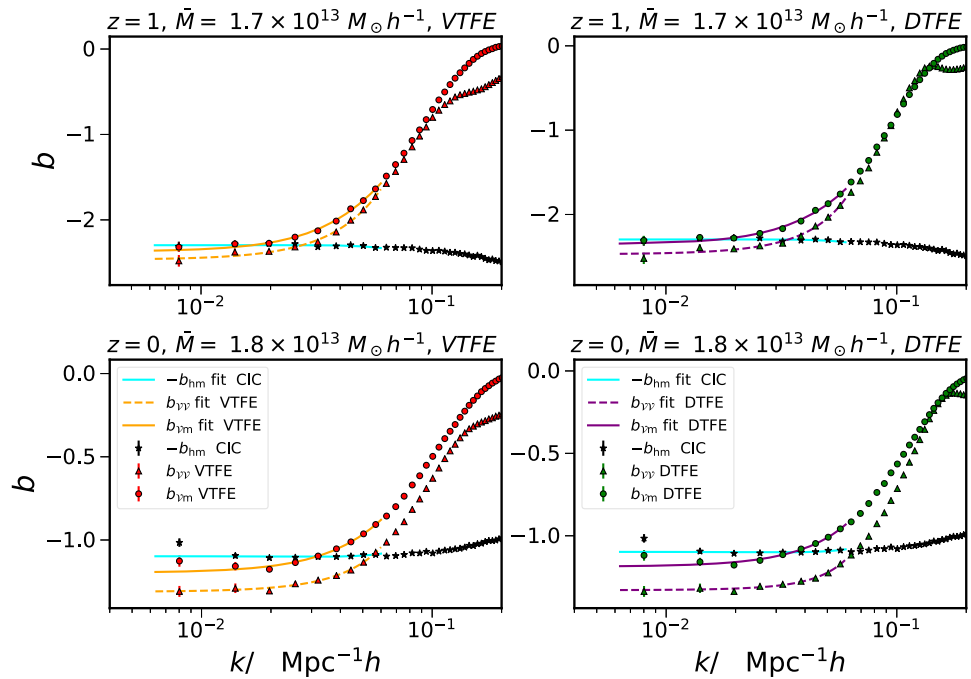


FIG. 7. As Fig. 6, but for the clustering bias of the volume field \mathcal{V} . To compare with the magnitude of the conventional halo-density bias, the negative of b_{hm} is shown. The properties of the halo samples used to construct the volume statistics are indicated on top of each panel. The bias of the volume field is negative and its magnitude is similar to that of the halo sample.

anticorrelates with the dark matter density field, its bias is *negative*. To ease comparison with the magnitude of the halo bias, we have also shown $-b_{\text{hm}}$. The overall shape of the bias functions of the volume field is similar to the density ones. On large scales, the bias of the volume field approaches the halo bias in magnitude. This is expected for large-scale fluctuations that are small, since in that limit \mathcal{V} reduces to $-\delta_{\text{h}}$. However, although $b_{\nu\text{m}}$ agrees with the halo bias well, there is a marked deviation of $b_{\nu\nu}$ from the former. We presume there to be loop corrections to the power spectrum of quadratic order in density, analogous to the shot-noise renormalization effect in the local bias case ([93], see also [63]). These additional shot-noise-like contributions cause deviations from the linear bias on large scales. On small scales, the density field is suppressed by the tessellation and the definition of \mathcal{V} is designed such that it approaches zero in the limit of vanishing δ_{h} . Hence, the behavior of the bias of \mathcal{V} is qualitatively similar to that of the tessellation density field.

Like cosmic voids, the volume field furnishes a negatively biased tracer of large-scale structure. Although the small-scale power is suppressed by the tessellation, the large-scale field is proportional to the underlying density field. In scenarios involving local primordial non-Gaussianity (PNG), the void bias exhibits a scale dependence on large linear scales [51]. Because the amplitude of void bias can be negative, this may be used to complement the traditional halo bias in constraining PNG [94]. Reference [51] demonstrated that a combination of halos and voids, taking advantage of the so-called multitracer approach [95,96], allows to substantially tighten constraints on the non-Gaussianity parameter f_{NL} . However, the gain in the constraining power of that analysis is limited by the shot noise in void autoclustering statistics. As the volume field also exhibits a negative bias, similar gains can be expected, but with a lower level of shot-noise contamination. One can think of the volume field as a “dual” of the density field with negative bias within the same survey volume. Thanks to their very different bias amplitudes, but comparable shot-noise levels, the density and volume fields together may provide optimal conditions for conducting a multitracer analysis [97–99].

Although $b_{\nu\text{m}}$ is not directly observable in galaxy surveys, this can be circumvented by considering

$$b_{\text{h}\nu} = \frac{P_{\text{h}\nu}}{P_{\text{mm}}}, \quad (28)$$

where $P_{\text{h}\nu}$ is the observable cross-power spectrum between δ_{h} and \mathcal{V} . The dark matter power spectrum P_{mm} can be modeled numerically or using perturbation theory. On large scales, where the fluctuations are small and the bias is linear, we have $b_{\text{h}\nu} \approx b_{\text{hm}}b_{\nu\text{m}}$. In Fig. 8, we plot the $b_{\text{h}\nu}$ measurement from our simulation. We note that since the volume field is constructed from the halos, there is a residual correlation analogous to the standard shot noise in the halo

auto-power spectrum. We measure the cross-power spectrum between the density field and the volume field obtained from the random catalog with the same number density as the halo field and subtract it from $P_{\text{h}\nu}$. Both the results before and after this shot-noise subtraction are shown. We indeed find that the shot-noise subtracted results are in good agreement with the prediction $b_{\text{hm}}b_{\nu\text{m}}$, for which we have used the fit results from b_{hm} and $b_{\nu\text{m}}$ obtained with Eq. (25).

C. The BAO in the volume field

In the early Universe, photons and baryons couple to form a hot plasma in which acoustic oscillations are excited. These oscillations leave important imprints in large-scale structure of the late Universe [100,101]. The BAO features are regarded as one of the most important probes of the large-scale structure and have been detected in numerous galaxy surveys [1].

Physically, the BAO manifests itself as an excess probability of finding galaxies at a distance r_{d} , the sound horizon at the drag epoch. This appears as a peak at the scale r_{d} in the galaxy density correlation function and as oscillations (wiggles) in Fourier space. Analogously, given a depression of \mathcal{V} at some location, it is more likely to find another depression of \mathcal{V} at a distance of r_{d} . This also gives rise to a positive enhancement in the correlation function of \mathcal{V} at the scale of r_{d} . We now go on to investigate the anticipated BAO signals in the volume statistic in more detail. To do so, we use two sets of simulations, one with the fiducial setup and another one with the Eisenstein-Hu initial conditions without BAO wiggles [85].

1. Real space

We begin with the BAO features measured in the real-space density field using the tessellation methods. To highlight the BAO features, we show the ratio between power spectra from the fiducial and the no-wiggle initial conditions in Fig. 9. The results for different halo groups obtained with three interpolation methods are compared. The BAO feature in the dark matter field is also shown for reference, which is determined with the CIC method. The number density of dark matter particles is so high that its density field can be regarded as continuous here.

On large scales (small k), all methods produce similar results, as the large-scale modes are unaffected by the interpolation methods. At higher k (compared to k_{ms}), the CIC halo field is still able to reproduce the BAO features imprinted in the dark matter density field, albeit with more noise. On the other hand, for the VTFE and DTFE fields, the BAO wiggles start to be smoothed out close to k_{ms} and are suppressed significantly at $k \gtrsim k_{\text{ms}}$. Following Appendix B, when the particle distribution is smoothed by a constant window function, the window factors out and cancels in the power spectrum ratio. Thus, the fact that there is smoothing of the ratio between the wiggle and no-wiggle power spectrum implies that the smoothing effect

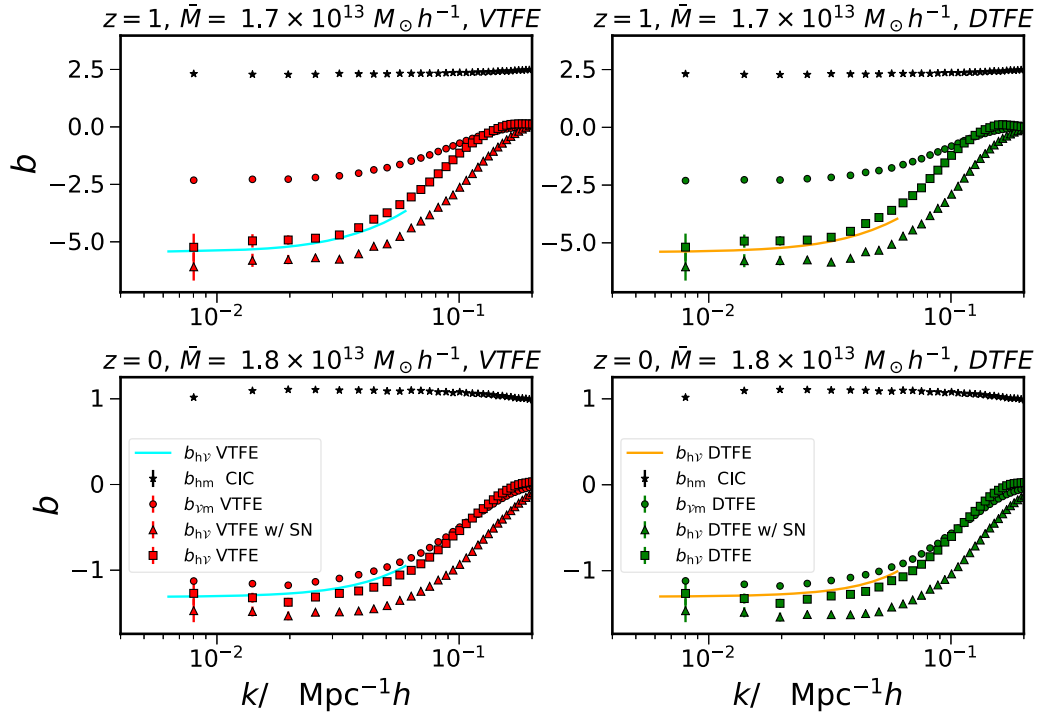


FIG. 8. The clustering bias $b_{h\nu}$ obtained from the cross-power spectrum between the halo-density field and the volume field using the VTFE (left panels, red) and the DTFE (right panels, green) methods. Because the volume field is derived from the halo distribution, there is a shot-noise contribution. Both the results before (triangles) and after shot-noise subtraction (squares) are compared. The halo bias b_{hm} (black stars) and volume bias $b_{\nu\text{m}}$ (VTFE: red circles, DTFE: green circles) are shown for reference. The solid curves (cyan for VTFE and orange for DTFE) are the predictions obtained using the fit results from Eq. (25), which are in good agreement with the direct measurements.

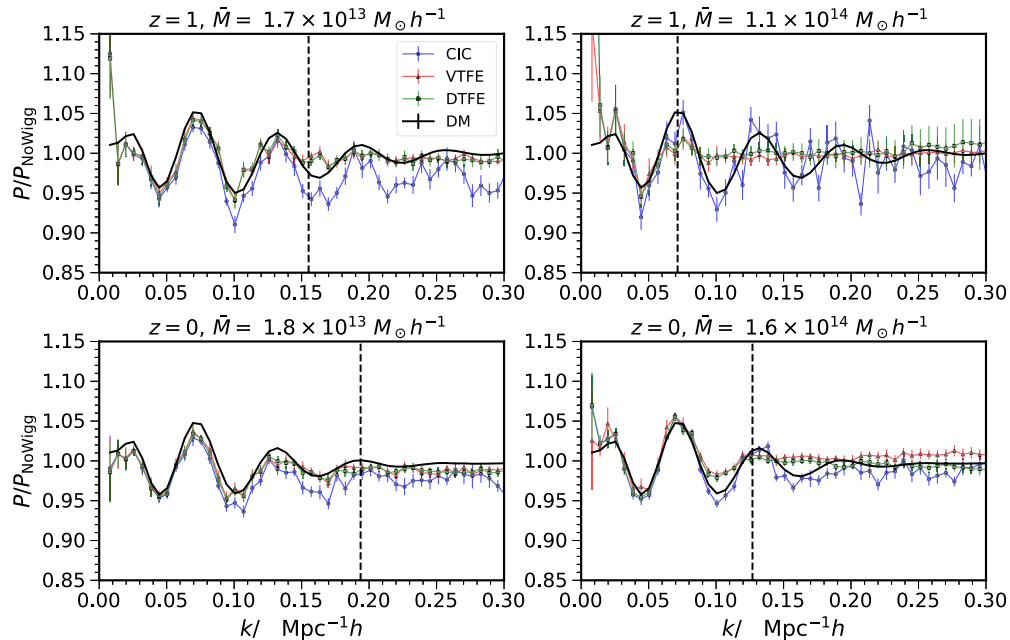


FIG. 9. Ratio between real-space halo-density power spectra from simulations with fiducial and no-wiggle initial conditions. The results for different halo-density fields obtained with the CIC (blue), VTFE (red), and DTFE (green) method are shown. The corresponding measurements from the dark matter (black solid curve) are overplotted as reference. The vertical dashed line indicates the mean separation scale k_{ms} of the halos. The halo-density power spectra from the tessellation methods can reproduce the BAO features on large scales, while the power beyond k_{ms} is suppressed.

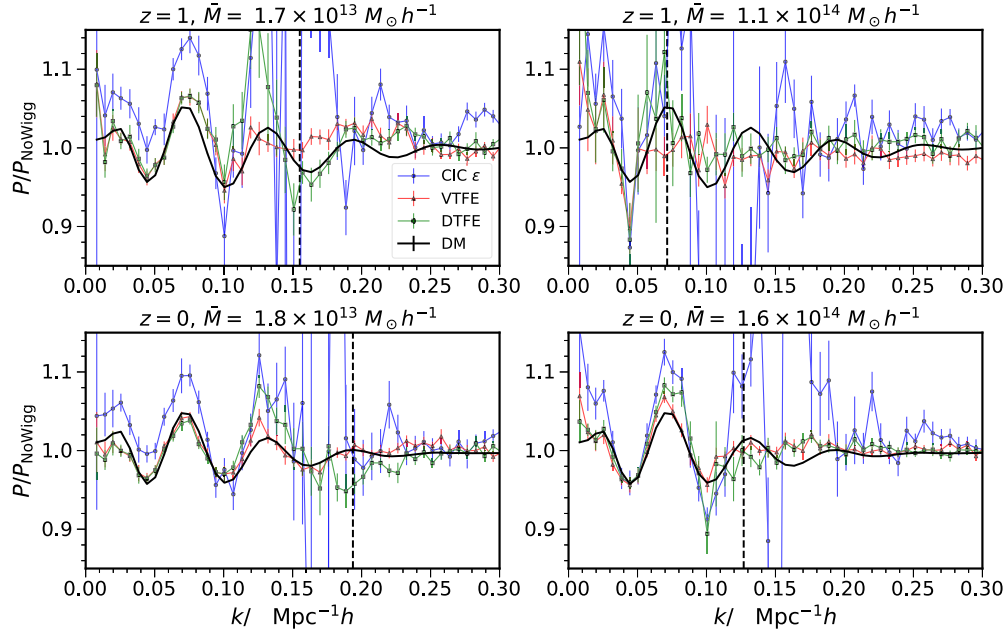


FIG. 10. As Fig. 9, but for the volume field constructed from the halo distribution. The black curve shows the BAO from the dark matter density field. The blue data points are the CIC measurement of the modified \mathcal{V}_ϵ statistic with $\epsilon = 0.1$. On large scales, the volume statistics computed with the tessellation methods exhibit BAO features without systematic bias.

cannot be attributed to a constant window and it must arise from the density-dependent nature of the tessellation methods. Numerous works have shown that the nonlinearity of density fields causes a smoothing of the BAO and its effect on the power spectrum can be approximated by a Gaussian window [55, 102–104]. This damping of the BAO is primarily driven by the large-scale bulk flow motion. Because the tessellation methods adaptively track the evolution of the particle distribution, the associated window has a similar effect, resulting in a smoothing of the BAO wiggles. This interpretation on the tessellation window is consistent with results in Sec. IV.

We now turn to the BAO imprints in the volume field, as presented in Fig. 10. The BAO measurement in the volume field is noisier than that in the density. In particular, although the VTFF and DTFE behave similarly for the density case, the DTFE yields more noisy results than the VTFF for the volume field. Overall, the BAO wiggles imprinted in the volume field follow those in the matter density field without any systematic bias. As in the density case above, beyond k_{ms} the BAO signature is washed out.

In the standard CIC interpolation, \mathcal{V} is ill-defined for regions with $\delta_h = -1$, which happens for empty grid cells. To overcome this problem, we can instead define

$$\begin{aligned} \mathcal{J}_\epsilon &= \frac{1}{1 + \delta_h + \epsilon}, \\ \mathcal{V}_\epsilon &= \frac{\mathcal{J}_\epsilon}{\bar{\mathcal{J}}_\epsilon} - 1, \end{aligned} \quad (29)$$

with some constant $\epsilon > 0$, which ensures that \mathcal{V}_ϵ is always well defined. We show the results for $\epsilon = 0.1$ in Fig. 9. This

method yields noisier results than the tessellation methods. We have checked that other values of ϵ , such as 0.01 or 0.2, do not improve this.

2. Redshift space

Observationally, we can only perform BAO measurements in redshift space. As the coordinates of the galaxies along the line-of-sight direction are deduced from redshifts in galaxy surveys, the density field is subject to additional perturbations due to their peculiar motion. These perturbations cause RSD, and we shall consider them in the plane-parallel limit, with the comoving coordinate in the z -direction x_z modified to

$$s_z = x_z + \frac{v_z}{aH}, \quad (30)$$

where v_z is the peculiar velocity in z -direction and H is the Hubble parameter. We can adapt Eq. (4) to redshift space as

$$(1 + \delta_L) d^3 q = (1 + \delta_h(x)) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{s}} \right| d^3 s = (1 + \delta_h(s)) d^3 s. \quad (31)$$

Accordingly, we have

$$\mathcal{J} = \frac{d^3 s}{(1 + \delta_L) d^3 q} = \frac{1}{1 + \delta_h(s)}. \quad (32)$$

Hence, the extension of the volume statistic to redshift space is straightforward.

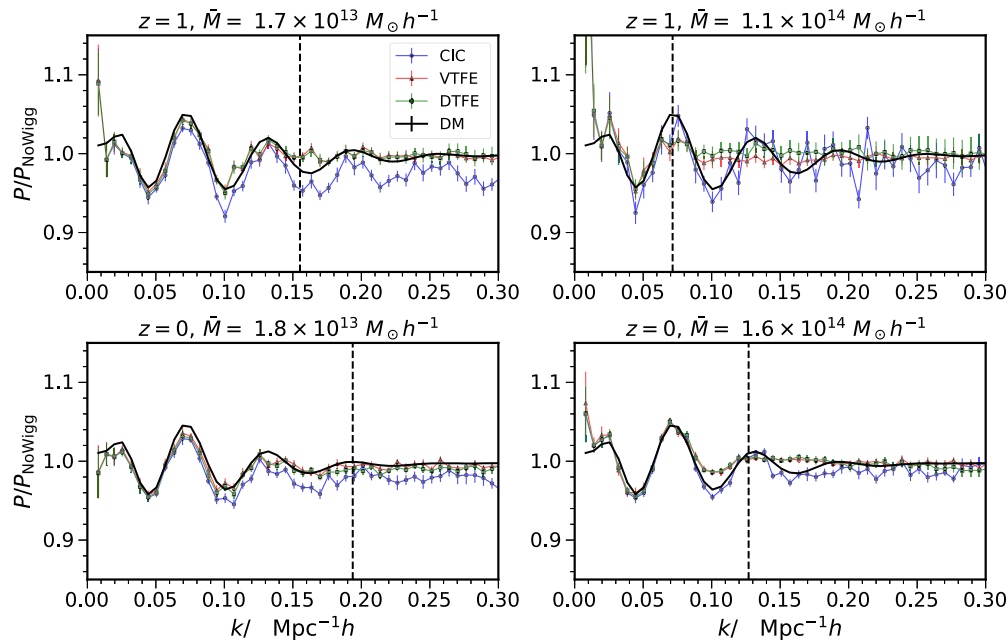


FIG. 11. As Fig. 9, but for the monopole power spectrum of the density statistics in redshift space. Similar to the results in real space, apart from the smoothing for $k \gtrsim k_{\text{ms}}$, the tessellation methods accurately reproduce the BAO wiggles in redshift space.

The redshift-space power spectrum can be expressed in terms of multipoles,

$$P_\ell = \frac{2\ell + 1}{2} \int_{-1}^1 d\mu P_s(k, \mu) \mathcal{L}_\ell(\mu), \quad (33)$$

where P_s is the power spectrum in redshift space, μ the cosine of the angle between $\hat{\mathbf{k}}$ and the line of sight, and \mathcal{L}_ℓ is the Legendre polynomial of order ℓ . Here we only show results for the monopole of the halo power spectrum, as the quadrupole measurements are noisy even for 20 realizations.

The monopole power spectrum for the halo-density field in redshift space is shown in Fig. 11. The results are similar to the real-space case. The halo monopole from the CIC interpolation appears to trace the wiggles in the dark matter monopole power spectrum well. While the tessellation results are slightly less noisy than the CIC ones for $k < k_{\text{ms}}$, they are suppressed for $k \gtrsim k_{\text{ms}}$. Figure 12 displays the monopole power spectrum for the volume statistics. As in real space, \mathcal{V}_e yields the most noisy estimator and the VTFF results are the most robust in reproducing the dark matter BAO. We note that the BAO amplitude in the volume field appears slightly enhanced at some scales. This is particularly apparent for the \mathcal{V}_e estimator, which is most sensitive to noise. We therefore attribute this effect to the discreteness of the tracer distribution.

3. Discussion

Besides the auto-power spectrum, we can measure the BAO using the cross-power spectrum between δ_{h} and \mathcal{V} .

The results are similar to those obtained from the auto-power spectrum on large scales, but they are more noisy for $k \gtrsim k_{\text{ms}}$, since the volume field lacks the BAO feature on small scales. So far, we have exclusively investigated Fourier-space statistics. In configuration space, the effect of the tessellation is a smoothing of the BAO peak as well. Thus, the BAO feature measured from the correlation function of the volume field is broadened and becomes less sharp, making it harder to differentiate from the broadband correlation function.

In order to fully capture the BAO features in the power spectrum of the volume statistics, the number density of the tracer sample must be sufficiently high. For example, at $z = 0$, the halo sample with mean mass $1.8 \times 10^{13} M_\odot h^{-1}$ and number density $2.3 \times 10^{-4} (\text{Mpc } h^{-1})^{-3}$ is sufficient. At higher redshift, the BAO wiggles are less damped by nonlinearities, so a higher number density is necessary to push k_{ms} to a larger value. For instance, at $z = 1$, the halo number density needs to be at least $7 \times 10^{-4} (\text{Mpc } h^{-1})^{-3}$ to fully capture the BAO wiggles.

To summarize the virtues of each method: for the density statistics, the CIC interpolation method is recommended, as it offers an unbiased estimate of the small-scale density field, while the VTFF and DTFE smooth out the field for scales above k_{ms} . The tessellation imposes additional conditions such that the resultant field is smooth and space-filling. These requirements modify the small-scale behavior of the field. On the other hand, in order to exploit volume statistics, the tessellation method is preferred, as it is able to construct a space-filling field with nonvanishing density everywhere. The results from the VTFF are similar

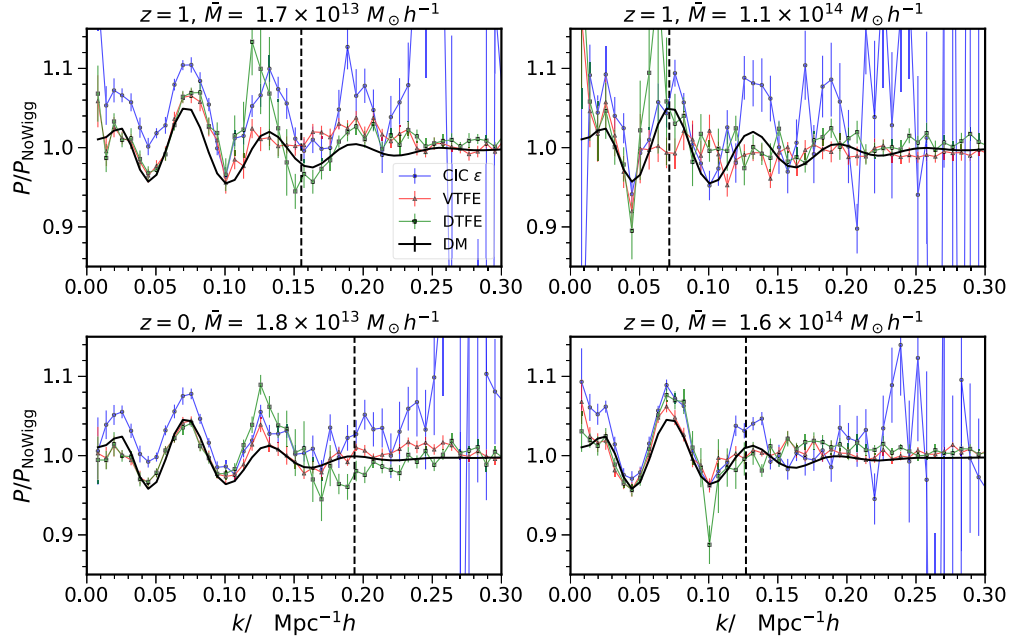


FIG. 12. As Fig. 10, but for monopole power spectrum of the volume statistics in redshift space. Similar to the results in real space, the large-scale BAO features are imprinted in the redshift-space volume power spectra without systematic bias.

to the DTFE for the density statistics, but VTFE yields more robust results for the volume statistics. Another advantage of the VTFE is its lower computational overhead compared to DTFE.⁷ Hence, for the clustering analysis of volume statistics, we recommend usage of the VTFE.

A measurement of the BAO feature from the distribution of underdense regions is interesting on its own, but it may also provide valuable information on cosmology in addition to what is available from halo clustering alone. We have demonstrated that the volume statistic traces the large-scale structure with a negative bias parameter. What remains to be shown is how correlated the volume statistic and the traditional halo-density statistic are. Since the volume statistic is constructed via the halo tracer distribution, the answer is not obvious. For example, one can construct a trivial field $-\delta_h$, which exhibits negative linear bias,⁸ but perfectly correlates with δ_h . In order to investigate the correlation between δ_h and \mathcal{V} , one could determine the covariance of the power spectrum from both the density and the volume field using many different realizations, which is beyond the scope of this paper.

A simple (albeit less conclusive) test is to consider the cross-correlation coefficient between δ_h and \mathcal{V} ,

$$r = \frac{P_{h\mathcal{V}}}{\sqrt{P_{hh}P_{\mathcal{V}\mathcal{V}}}}, \quad (34)$$

where P_{hh} , $P_{\mathcal{V}\mathcal{V}}$, and $P_{h\mathcal{V}}$ are the halo auto-, volume auto-, and halo-volume cross-power spectra.

The results for fields constructed from four different halo groups are shown in Fig. 13. We have used the full redshift-space monopole power spectra including shot noise. On large scales, both δ_h (estimated via CIC) and \mathcal{V} (estimated via VTFE or DTFE) are (anti-)correlated with the dark matter δ and with each other, but due to the presence of shot noise and other sources of stochastic noise in the volume field, r is slightly above -1 . The characteristic shape of the curve is due to the exclusion between δ_h and \mathcal{V} . At smaller scales, exclusion can lead to oscillations before r approaches to zero, but they are quickly suppressed when the shot noise kicks in at high k . If the volume statistics were as trivial as $-\delta_h$, r would be equal to -1 on all scales.

We note that the BAO features are also detectable using overlapping voids [42,43], which are defined to be the circumspheres of the tetrahedrons resulting from the Delaunay tessellation of the galaxies [105]. Although this approach is similar to ours in using the tessellation to partition the point set, we do not define objects on it. There can be substantial overlap between the circumspheres of neighboring tetrahedrons. In the void construction algorithm, the void center and size are determined entirely by the four particles spanning the tetrahedron. The pointlike nature of the void-center position prevents any additional smoothing of the BAO; however, the resultant BAO measurement may be more correlated with the density statistic.

⁷In a typical run with a single core on an Intel Xeon E5-2686 cluster, the VTFE takes about 40 minutes, while the DTFE about 24 hours. Perhaps the DTFE algorithm can be more efficient after further optimization.

⁸However, for voids, the behavior of quadratic bias is similar to that of halos [49,52], so at second order the bias of this artificial field is opposite to that of the genuine underdense tracer.

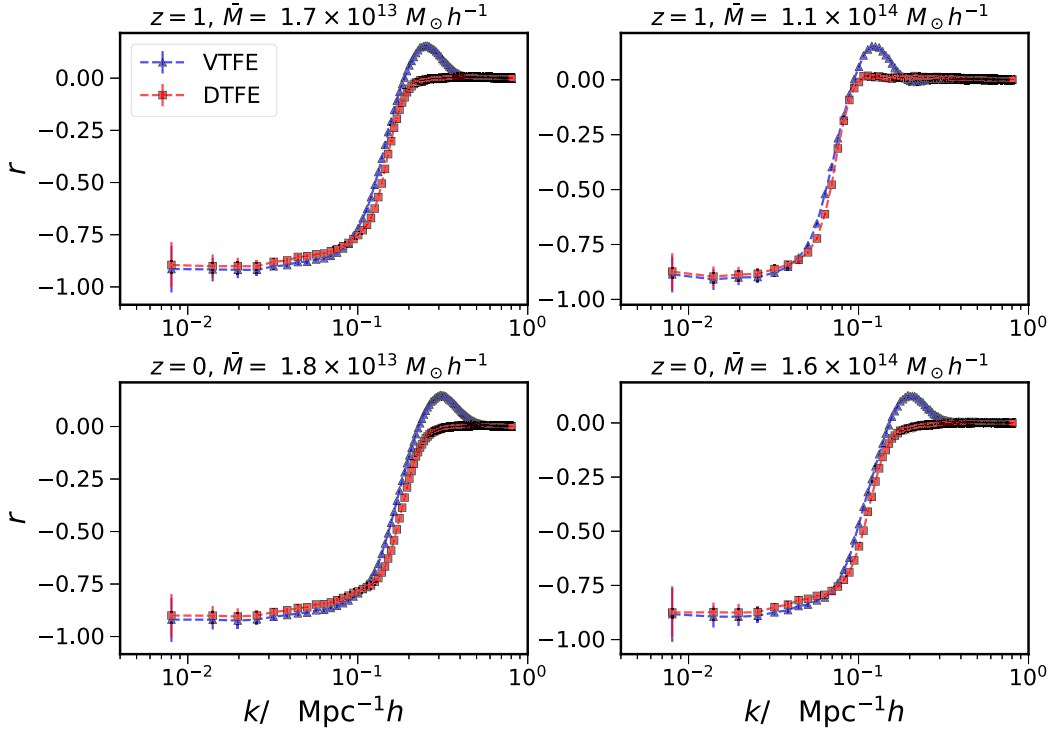


FIG. 13. Cross-correlation coefficient between the halo-density field δ_h (estimated via CIC) and the volume field \mathcal{V} . The latter is estimated using the distribution halos as tracers, either via VTFE (blue triangles), or via DTFE (red squares).

Finally, we comment on the possible effects of the BAO reconstruction [103]. BAO reconstruction is often applied in galaxy surveys (e.g., [106]) to undo part of the large-scale gravitational evolution and RSD effect so as to boost the BAO signal. This is because the reconstructed field becomes more correlated with the initial conditions, so the BAO signal in the volume field is expected to increase as well. For example, \mathcal{V} reduces to $-\delta_h$ for weak fluctuations and we expect the BAO signal to be well correlated between these two fields. However, a detailed study is required to access the overall gain in BAO information from the reconstructed volume field.

VI. CONCLUSIONS

In large-scale structure analyses, it is common to exploit statistics based on the galaxy distribution, which predominantly traces high-density regions in the Universe. However, a large fraction of its volume exhibits relatively low density. Hence, the clustering of underdense regions may bear cosmologically relevant information that is complementary to the conventional clustering of overdensities. In principle, cosmic voids are ideal proxies for probing the underdense regions, but their auto-correlation statistics suffer a large shot-noise contamination due to their low number density. In this work, we propose to apply the volume statistic \mathcal{V} to probe the volume distribution of large-scale structure. This statistic provides a measure of the volume change between Eulerian and Lagrangian

space. As it makes use of all the tracer particles available, its shot-noise level is similar to that of the conventional tracer density field. Furthermore, the definition of the volume statistic is closely related to the density contrast, so it may be more amenable to theoretical models based on perturbation theory than objects that are defined via non-linear topological characteristics of the cosmic web, such as voids. Also, an extension of its definition to redshift space is straightforward.

Traditional mass-assignment methods, such as CIC interpolation, yield empty regions with $\delta_h = -1$ for sparse halo samples, which make the volume statistic estimator ill-defined there. To overcome this difficulty, we apply tessellation interpolation methods to estimate the volume field. We study the clustering statistics obtained via the tessellation methods in detail. The power spectrum from the tessellated field is smoothed by an effective window function, which on large scales can be approximated by the same simple function for different tessellation methods. On small scales, this approximation fails due to the density dependence of the tessellation.

For the density clustering, the conventional mass-interpolation methods are recommended because they can reproduce the original discrete point distribution for sufficiently large grid size. Because of the additional smoothness condition imposed, the tessellation methods cause smoothing of the field at scales smaller than the mean particle separation. When the tessellation is used for the volume statistics, the VTFE provides the most

satisfactory results in terms of signal-to-noise ratio. At the same time, the VTFE is computationally less expensive than the DTFE.

The clustering amplitude of the volume statistic is negatively biased with respect to the dark matter density field δ on large scales, reflecting the fact that it traces underdense regions in the Universe. This bias can also be measured without knowledge of δ via the observable cross-power spectrum between the fields δ_h and \mathcal{V} , which are strongly anticorrelated on large scales. Furthermore, the BAO features are imprinted in the volume statistic as well. Apart from a smoothing beyond the mean tracer separation scale $k \gtrsim k_{\text{ms}}$, they are reproduced without systematic bias. Hence, to avoid loss of any clustering information, the number density of the tracer used to generate the volume statistic must be sufficiently high. An investigation of the covariance between the BAO information in the density and volume statistic is left for future work.

As the signal-to-noise ratio of void clustering is significantly limited by shot noise, we expect volume statistics to complement the available information from the clustering of underdense regions. For example, a cross-correlation with galaxy positions and shapes, quasars, the CMB, and even three-point statistics of the volume statistic should be detectable at high significance. Voids have been suggested to be a sensitive probe for rich phenomena, such as the imprints of dark energy, modified gravity, and massive neutrinos [107]. The volume statistic opens up a new avenue toward exploring these ideas. For example, in analogy to Ref. [51], the bias of the volume statistic can be used to constrain the inflationary paradigm via PNG. This can be combined with the multitracer approach [95–99] to optimize the statistical gain of this method.

ACKNOWLEDGMENTS

We thank Xin Wang and Yi Zheng for useful discussions. The simulations in this work were run at the Kunlun cluster at the School of Physics and Astronomy of the Sun-Yat Sen University. K. C. C. acknowledges the support from the National Science Foundation of China under the Grant No. 11873102 and the Science and Technology Program of Guangzhou, China (Grant No. 202002030360). The work of N.H. was supported by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy—EXC-2094–390783311.

APPENDIX A: KERNEL DENSITY ESTIMATION

In this Appendix, we show that the density can be estimated using the method of KDE in statistics. See, e.g., [83] for more details on KDE. Given a number of data points \mathbf{x}_α sampling the underlying probability distribution, instead of estimating the probability density by histogram,

the sample points are smoothed by a window (or kernel) W of characteristic size h_α . Note that the stochastic variables are the sampling points and the underlying probability density is fixed. The probability density p can be estimated by the estimator

$$\hat{p}(\mathbf{x})\Delta V = \frac{1}{N} \sum_{\alpha=1}^N W\left(\frac{\mathbf{x}-\mathbf{x}_\alpha}{h_\alpha}\right), \quad (\text{A1})$$

where N is the total number of particles contributing to density estimation. For overlapping windows, it includes all the particles available. The volume element ΔV is to be fixed later on. The expectation of $\hat{p}(\mathbf{x})$ is

$$\begin{aligned} \langle \hat{p}(\mathbf{x}) \rangle &= \frac{1}{N\Delta V} \sum_{\alpha=1}^N \left\langle W\left(\frac{\mathbf{x}-\mathbf{x}_\alpha}{h_\alpha}\right) \right\rangle \\ &= \frac{1}{N\Delta V} \sum_{\alpha=1}^N \int d^3x_\alpha W\left(\frac{\mathbf{x}-\mathbf{x}_\alpha}{h_\alpha}\right) p(\mathbf{x}_\alpha) \\ &= \frac{1}{N\Delta V} \sum_{\alpha=1}^N \int d^3y_\alpha h_\alpha^3 W(\mathbf{y}_\alpha) p(\mathbf{x}-h_\alpha\mathbf{y}_\alpha). \end{aligned} \quad (\text{A2})$$

By expanding the probability density to second order, we have

$$\begin{aligned} \langle \hat{p}(\mathbf{x}) \rangle &\approx \frac{1}{N\Delta V} \sum_{\alpha=1}^N \int d^3y_\alpha h_\alpha^3 W(\mathbf{y}_\alpha) \\ &\quad \times \left[p(\mathbf{x}) - h_\alpha \sum_i y_{\alpha i} \frac{\partial p}{\partial x_i} + \frac{1}{2} \sum_{ij} h_\alpha^2 y_{\alpha i} y_{\alpha j} \frac{\partial^2 p}{\partial x_i \partial x_j} \right]. \end{aligned} \quad (\text{A3})$$

The window satisfies the property that

$$\int d^3y W(\mathbf{y}) = 1, \quad (\text{A4})$$

and it is chosen to be an even function. Therefore, it simplifies to

$$\langle \hat{p}(\mathbf{x}) \rangle = \frac{\sum_{\alpha=1}^N h_\alpha^3}{N\Delta V} p(\mathbf{x}) + \frac{\sum_{\alpha=1}^N h_\alpha^5}{2N\Delta V} \sum_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j} I_{ij}^{(1)}, \quad (\text{A5})$$

where for convenience we define the notation

$$I_{ij\dots k}^{(n)} \equiv \int d^3y W^n(\mathbf{y}) y_i y_j \dots y_k. \quad (\text{A6})$$

For \hat{p} to be unbiased to the lowest order, we require

$$\Delta V = \frac{1}{N} \sum_{\alpha=1}^N h_\alpha^3. \quad (\text{A7})$$

Then the bias arises from the second derivative term

$$\langle \hat{p}(\mathbf{x}) \rangle - p(\mathbf{x}) = \frac{\sum_{\alpha=1}^N h_{\alpha}^5}{2 \sum_{\alpha=1}^N h_{\alpha}^3} \sum_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j} I_{ij}^{(1)}. \quad (\text{A8})$$

In particular, at the peak (trough), the estimated probability density is lower (higher) than the true one.

If we assume that the sampling points are independent of each other, the variance of \hat{p} is given by

$$\begin{aligned} \text{Var}(\hat{p}(\mathbf{x})) &= \frac{1}{(N\Delta V)^2} \sum_{\alpha=1}^N \text{Var}\left(W\left(\frac{\mathbf{x}-\mathbf{x}_{\alpha}}{h_{\alpha}}\right)\right) \\ &= \frac{1}{(N\Delta V)^2} \sum_{\alpha=1}^N \left\{ \int dx_{\alpha} W^2\left(\frac{\mathbf{x}-\mathbf{x}_{\alpha}}{h_{\alpha}}\right) p(x_{\alpha}) \right. \\ &\quad \left. - \left[\int dx_{\alpha} W\left(\frac{\mathbf{x}-\mathbf{x}_{\alpha}}{h_{\alpha}}\right) p(x_{\alpha}) \right]^2 \right\}. \end{aligned} \quad (\text{A9})$$

To the lowest order in h , we have

$$\text{Var}(\hat{p}(\mathbf{x})) \approx \frac{I^{(2)} p(\mathbf{x})}{\sum_{\alpha=1}^N h_{\alpha}^3}. \quad (\text{A10})$$

Note that the variance is inversely proportional to the total volume of the windows.

Now suppose that we have N particles, and we would like to estimate the number density of the sample, $n(\mathbf{x})$, by applying a smoothing window to the particles. This can be done by simply replacing p by $n(\mathbf{x})/N$, which is non-negative and normalized to 1. The estimator for n reads

$$\hat{n}(\mathbf{x}) = \frac{N}{\sum_{\alpha=1}^N h_{\alpha}^3} \sum_{\alpha=1}^N W\left(\frac{\mathbf{x}-\mathbf{x}_{\alpha}}{h_{\alpha}}\right). \quad (\text{A11})$$

The bias and the variance of the estimator are given by

$$\langle \hat{n}(\mathbf{x}) \rangle - n(\mathbf{x}) = \frac{\sum_{\alpha=1}^N h_{\alpha}^5}{2 \sum_{\alpha=1}^N h_{\alpha}^3} \sum_{ij} \frac{\partial^2 n}{\partial x_i \partial x_j} I_{ij}^{(1)}, \quad (\text{A12})$$

$$\text{Var}(\hat{n}(\mathbf{x})) = \frac{n(\mathbf{x}) I^{(2)}}{\frac{1}{N} \sum_{\alpha=1}^N h_{\alpha}^3}. \quad (\text{A13})$$

These results apply to the density estimation using the usual mass-interpolation method. It is important to remember that these results are obtained assuming that the sampling points Poisson sample the underlying density field.

APPENDIX B: EFFECTS OF THE WINDOW FUNCTION ON THE POWER SPECTRUM

The standard shot-noise power spectrum result [Eq. (20)] can be derived by considering the particle distribution given by a sum of Dirac deltas; see, e.g., the Appendix of [108]. Using the number density given by

$$n(\mathbf{x}) = \sum_{\alpha} \delta_{\text{D}}(\mathbf{x}-\mathbf{x}_{\alpha}), \quad (\text{B1})$$

the discrete two-point correlation can be written as

$$\begin{aligned} \langle n(\mathbf{x}_1) n(\mathbf{x}_2) \rangle &= \delta_{\text{D}}(\mathbf{x}_1 - \mathbf{x}_2) \left\langle \sum_{\alpha} \delta_{\text{D}}(\mathbf{x}_1 - \mathbf{x}_{\alpha}) \right\rangle + \bar{n}^2 [1 + \xi(x_{12})] \\ &= \bar{n} \delta_{\text{D}}(\mathbf{x}_1 - \mathbf{x}_2) + \bar{n}^2 [1 + \xi(x_{12})], \end{aligned} \quad (\text{B2})$$

where \bar{n} is the mean number density given by

$$\bar{n} = \left\langle \sum_{\alpha} \delta_{\text{D}}(\mathbf{x}-\mathbf{x}_{\alpha}) \right\rangle, \quad (\text{B3})$$

and ξ is the correlation function due to continuous clustering. The discrete correlation function then reads

$$\xi_{\text{d}}(\mathbf{x}_1 - \mathbf{x}_2) = \frac{1}{\bar{n}} \delta_{\text{D}}(\mathbf{x}_1 - \mathbf{x}_2) + \xi(x_{12}). \quad (\text{B4})$$

Upon Fourier transform, we obtain

$$P_{\text{d}}(k) = \frac{1}{(2\pi)^3 \bar{n}} + P(k). \quad (\text{B5})$$

with the first term being Eq. (20).

The Dirac delta function can be thought of as a special kind of window function. We now consider the case when the particles are smoothed by an extended window function W . The number density then reads

$$n_W(\mathbf{x}) = \sum_{\alpha} W(\mathbf{x}-\mathbf{x}_{\alpha}). \quad (\text{B6})$$

Note that because the window smoothing respects mass conservation, $\bar{n} = \langle n_W \rangle$. Then the two-point correlation reads

$$\begin{aligned} \langle n_W(\mathbf{x}_1) n_W(\mathbf{x}_2) \rangle &= \bar{n}^2 \int d^3 x' \int d^3 x'' W(\mathbf{x}_1 - \mathbf{x}') W(\mathbf{x}_2 - \mathbf{x}'') \xi(|\mathbf{x}' - \mathbf{x}''|) \\ &\quad + \bar{n}^2 + \bar{n} \int d^3 x' W(\mathbf{x}_1 - \mathbf{x}') W(\mathbf{x}_2 - \mathbf{x}'). \end{aligned} \quad (\text{B7})$$

The discrete correlation can then be written as

$$\begin{aligned} \xi_{\text{WW}}(|\mathbf{x}_1 - \mathbf{x}_2|) &= \frac{1}{\bar{n}} \int d^3 x' W(\mathbf{x}_1 - \mathbf{x}') W(\mathbf{x}_2 - \mathbf{x}') \\ &\quad + \int d^3 x' \int d^3 x'' W(\mathbf{x}_1 - \mathbf{x}') \\ &\quad \times W(\mathbf{x}_2 - \mathbf{x}'') \xi(|\mathbf{x}' - \mathbf{x}''|). \end{aligned} \quad (\text{B8})$$

In Fourier space, we have

$$P_{wW}(k) = (2\pi)^6 |W(k)|^2 \left[P(k) + \frac{1}{(2\pi)^3 \bar{n}} \right]. \quad (\text{B9})$$

The first term in the square bracket is analogous to the two-halo term and the second is the analog of the one-halo term in the halo model [109]. For both the continuous and the shot-noise term, the effect of the window function is the same and can be accounted for by a $|W(k)|^2$ factor.

Similarly, the cross-correlation function between n and n_w is given by

$$\xi_w(|\mathbf{x}_1 - \mathbf{x}_2|) = \frac{1}{\bar{n}} W(\mathbf{x}_1 - \mathbf{x}_2) + \int d^3x' \int d^3x'' \times W(\mathbf{x}_1 - \mathbf{x}') \delta_D(\mathbf{x}_2 - \mathbf{x}'') \xi(|\mathbf{x}' - \mathbf{x}''|), \quad (\text{B10})$$

and the cross-power spectrum reads

$$P_w(k) = (2\pi)^3 W(k) \left[P(k) + \frac{1}{(2\pi)^3 \bar{n}} \right]. \quad (\text{B11})$$

Again, we find that the effect of the window function is captured by a factor of $W(k)$.

-
- [1] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo, *Phys. Rep.* **530**, 87 (2013).
 - [2] S. Cole *et al.* (2dFGRS Collaboration), *Mon. Not. R. Astron. Soc.* **362**, 505 (2005).
 - [3] D. J. Eisenstein *et al.* (SDSS Collaboration), *Astrophys. J.* **633**, 560 (2005).
 - [4] F. Beutler, C. Blake, M. Colless, D. Heath Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, *Mon. Not. R. Astron. Soc.* **416**, 3017 (2011).
 - [5] C. Blake *et al.*, *Mon. Not. R. Astron. Soc.* **425**, 405 (2012).
 - [6] L. Anderson *et al.*, *Mon. Not. R. Astron. Soc.* **427**, 3435 (2012).
 - [7] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, *Mon. Not. R. Astron. Soc.* **449**, 835 (2015).
 - [8] S. Alam *et al.* (BOSS Collaboration), *Mon. Not. R. Astron. Soc.* **470**, 2617 (2017).
 - [9] S. Alam *et al.* (eBOSS Collaboration), arXiv:2007.08991.
 - [10] N. Kaiser, *Astrophys. J. Lett.* **284**, L9 (1984).
 - [11] R. van de Weygaert and E. van Kampen, *Mon. Not. R. Astron. Soc.* **263**, 481 (1993).
 - [12] S. Gottloeber, E. L. Lokas, A. Klypin, and Y. Hoffman, *Mon. Not. R. Astron. Soc.* **344**, 715 (2003).
 - [13] R. K. Sheth and R. van de Weygaert, *Mon. Not. R. Astron. Soc.* **350**, 517 (2004).
 - [14] A. Pisani, P. M. Sutter, N. Hamaus, E. Alizadeh, R. Biswas, B. D. Wandelt, and C. M. Hirata, *Phys. Rev. D* **92**, 083531 (2015).
 - [15] G. Pollina, M. Baldi, F. Marulli, and L. Moscardini, *Mon. Not. R. Astron. Soc.* **455**, 3075 (2016).
 - [16] M. Sahlén, *Phys. Rev. D* **99**, 063525 (2019).
 - [17] G. Verza, A. Pisani, C. Carbone, N. Hamaus, and L. Guzzo, *J. Cosmol. Astropart. Phys.* **12** (2019) 040.
 - [18] J. Clampitt, Y.-C. Cai, and B. Li, *Mon. Not. R. Astron. Soc.* **431**, 749 (2013).
 - [19] Y.-C. Cai, N. Padilla, and B. Li, *Mon. Not. R. Astron. Soc.* **451**, 1036 (2015).
 - [20] P. Zivick, P. M. Sutter, B. D. Wandelt, B. Li, and T. Y. Lam, *Mon. Not. R. Astron. Soc.* **451**, 4215 (2015).
 - [21] I. Achitouv, *Phys. Rev. D* **94**, 103524 (2016).
 - [22] E. Perico, R. Voivodic, M. Lima, and D. Mota, *Astron. Astrophys.* **632**, A52 (2019).
 - [23] E. Massara, F. Villaescusa-Navarro, M. Viel, and P. Sutter, *J. Cosmol. Astropart. Phys.* **11** (2015) 018.
 - [24] C. D. Kreisch, A. Pisani, C. Carbone, J. Liu, A. J. Hawken, E. Massara, D. N. Spergel, and B. D. Wandelt, *Mon. Not. R. Astron. Soc.* **488**, 4413 (2019).
 - [25] N. Schuster, N. Hamaus, A. Pisani, C. Carbone, C. D. Kreisch, G. Pollina, and J. Weller, *J. Cosmol. Astropart. Phys.* **12** (2019) 055.
 - [26] S. Contarini, F. Marulli, L. Moscardini, A. Veropalumbo, C. Giocoli, and M. Baldi, arXiv:2009.03309.
 - [27] D. Paz, M. Lares, L. Ceccarelli, N. Padilla, and D. G. Lambas, *Mon. Not. R. Astron. Soc.* **436**, 3480 (2013).
 - [28] N. Hamaus, P. Sutter, G. Lavaux, and B. D. Wandelt, *J. Cosmol. Astropart. Phys.* **11** (2015) 036.
 - [29] Y.-C. Cai, A. Taylor, J. A. Peacock, and N. Padilla, *Mon. Not. R. Astron. Soc.* **462**, 2465 (2016).
 - [30] I. Achitouv, C. Blake, P. Carter, J. Koda, and F. Beutler, *Phys. Rev. D* **95**, 083502 (2017).
 - [31] A. Hawken *et al.*, *Astron. Astrophys.* **607**, A54 (2017).
 - [32] C.-H. Chuang, F.-S. Kitaura, Y. Liang, A. Font-Ribera, C. Zhao, P. McDonald, and C. Tao, *Phys. Rev. D* **95**, 063528 (2017).
 - [33] N. Hamaus, M.-C. Cousinou, A. Pisani, M. Aubert, S. Escoffier, and J. Weller, *J. Cosmol. Astropart. Phys.* **07** (2017) 014.
 - [34] M. Aubert *et al.*, arXiv:2007.09013.
 - [35] S. Nadathur *et al.*, *Mon. Not. R. Astron. Soc.* **499**, 4140 (2020).
 - [36] G. Lavaux and B. D. Wandelt, *Mon. Not. R. Astron. Soc.* **403**, 1392 (2010).
 - [37] P. M. Sutter, G. Lavaux, B. D. Wandelt, and D. H. Weinberg, *Astrophys. J.* **761**, 187 (2012).
 - [38] N. Hamaus, A. Pisani, P. M. Sutter, G. Lavaux, S. Escoffier, B. D. Wandelt, and J. Weller, *Phys. Rev. Lett.* **117**, 091302 (2016).
 - [39] Q. Mao, A. A. Berlind, R. J. Scherrer, M. C. Neyrinck, R. Scoccimarro, J. L. Tinker, C. K. McBride, and D. P. Schneider, *Astrophys. J.* **835**, 160 (2017).

- [40] C. M. Correa, D. J. Paz, N. D. Padilla, A. N. Ruiz, R. E. Angulo, and A. G. Sánchez, *Mon. Not. R. Astron. Soc.* **485**, 5761 (2019).
- [41] N. Hamaus, A. Pisani, J.-A. Choi, G. Lavaux, B. D. Wandelt, and J. Weller, *J. Cosmol. Astropart. Phys.* **12** (2020) 023.
- [42] Y. Liang, C. Zhao, C.-H. Chuang, F.-S. Kitaura, and C. Tao, *Mon. Not. R. Astron. Soc.* **459**, 4020 (2016).
- [43] F.-S. Kitaura *et al.*, *Phys. Rev. Lett.* **116**, 171301 (2016).
- [44] N. Hamaus, B. D. Wandelt, P. M. Sutter, G. Lavaux, and M. S. Warren, *Phys. Rev. Lett.* **112**, 041304 (2014).
- [45] P. Sutter, G. Lavaux, N. Hamaus, B. D. Wandelt, D. H. Weinberg, and M. S. Warren, *Mon. Not. R. Astron. Soc.* **442**, 462 (2014).
- [46] N. Hamaus, P. Sutter, G. Lavaux, and B. D. Wandelt, *J. Cosmol. Astropart. Phys.* **12** (2014) 013.
- [47] G. Pollina, N. Hamaus, K. Dolag, J. Weller, M. Baldi, and L. Moscardini, *Mon. Not. R. Astron. Soc.* **469**, 787 (2017).
- [48] G. Pollina *et al.* (DES Collaboration), *Mon. Not. R. Astron. Soc.* **487**, 2836 (2019).
- [49] K. C. Chan, N. Hamaus, and V. Desjacques, *Phys. Rev. D* **90**, 103521 (2014).
- [50] J. Clampitt, B. Jain, and C. Sánchez, *Mon. Not. R. Astron. Soc.* **456**, 4425 (2016).
- [51] K. C. Chan, N. Hamaus, and M. Biagetti, *Phys. Rev. D* **99**, 121304 (2019).
- [52] K. C. Chan, Y. Li, M. Biagetti, and N. Hamaus, arXiv:1909.03736.
- [53] D. Jamieson and M. Loverde, *Phys. Rev. D* **100**, 123528 (2019).
- [54] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, *Phys. Rep.* **367**, 1 (2002).
- [55] T. Matsubara, *Phys. Rev. D* **78**, 083519 (2008); **78**, 109901(E) (2008).
- [56] T. Matsubara, *Phys. Rev. D* **83**, 083518 (2011).
- [57] E. Massara, F. Villaescusa-Navarro, S. Ho, N. Dalal, and D. N. Spergel, *Phys. Rev. Lett.* **126**, 011301 (2021).
- [58] O. H. Philcox, A. Aviles, and E. Massara, arXiv:2010.05914.
- [59] R. K. Sheth, A. J. Connolly, and R. Skibba, arXiv:astro-ph/0511773.
- [60] M. White, *J. Cosmol. Astropart. Phys.* **11** (2016) 057.
- [61] G. Valogiannis and R. Bean, *Phys. Rev. D* **97**, 023535 (2018).
- [62] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, *Astrophys. J. Lett.* **698**, L90 (2009).
- [63] X. Wang, M. Neyrinck, I. Szapudi, A. Szalay, X. Chen, J. Lesgourgues, A. Riotta, and M. Sloth, *Astrophys. J.* **735**, 32 (2011).
- [64] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (A. Hilger, 1981).
- [65] Y. Jing, *Astrophys. J.* **620**, 559 (2005).
- [66] E. Sefusatti, M. Crocce, R. Scoccimarro, and H. Couchman, *Mon. Not. R. Astron. Soc.* **460**, 3624 (2016).
- [67] M. C. Neyrinck, N. Y. Gnedin, and A. J. S. Hamilton, *Mon. Not. R. Astron. Soc.* **356**, 1222 (2005).
- [68] M. C. Neyrinck, *Mon. Not. R. Astron. Soc.* **386**, 2101 (2008).
- [69] T. Sousbie, *Mon. Not. R. Astron. Soc.* **414**, 350 (2011).
- [70] M. Cautun, R. van de Weygaert, and B. J. T. Jones, *Mon. Not. R. Astron. Soc.* **429**, 1286 (2013).
- [71] M. Cautun, R. van de Weygaert, B. J. T. Jones, and C. S. Frenk, *Mon. Not. R. Astron. Soc.* **441**, 2923 (2014).
- [72] P. Busch and S. D. White, *Mon. Not. R. Astron. Soc.* **493**, 5693 (2020).
- [73] A. Paranjape and S. Alam, *Mon. Not. R. Astron. Soc.* **495**, 3233 (2020).
- [74] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, New York, NY, 2007).
- [75] R. van de Weygaert and W. Schaap, The cosmic web: Geometric analysis, in *Data Analysis in Cosmology, Lecture Notes in Physics*, edited by V. J. Martínez, E. Saar, E. Martínez-González, and M.-J. Pons-Bordería (Springer, Berlin, 2009), Vol. 665, pp. 291–413.
- [76] V. Icke and R. van de Weygaert, *Astron. Astrophys.* **184**, 16 (1987).
- [77] S. Yoshioka and S. Ikeuchi, *Astrophys. J.* **341**, 16 (1989).
- [78] R. van de Weygaert, *Astron. Astrophys.* **283**, 361 (1994).
- [79] F. Bernardeau and R. van de Weygaert, *Mon. Not. R. Astron. Soc.* **279**, 693 (1996).
- [80] W. E. Schaap and R. van de Weygaert, *Astron. Astrophys.* **363**, L29 (2000).
- [81] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith, and T. Baldauf, *Phys. Rev. D* **82**, 043515 (2010).
- [82] T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus, and V. Desjacques, *Phys. Rev. D* **88**, 083507 (2013).
- [83] Q. Li and J. S. Racine, *Nonparametric Econometrics: Theory and Practice* (Princeton University Press, New Jersey, 2007).
- [84] D. Blas, J. Lesgourgues, and T. Tram, *J. Cosmol. Astropart. Phys.* **11** (2011) 034.
- [85] D. J. Eisenstein and W. Hu, *Astrophys. J.* **496**, 605 (1998).
- [86] M. Crocce, S. Pueblas, and R. Scoccimarro, *Mon. Not. R. Astron. Soc.* **373**, 369 (2006).
- [87] V. Springel, *Mon. Not. R. Astron. Soc.* **364**, 1105 (2005).
- [88] S. R. Knollmann and A. Knebe, *Astrophys. J. Suppl.* **182**, 608 (2009).
- [89] M. C. Cautun and R. van de Weygaert, arXiv:1105.0370.
- [90] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, *ACM Trans. Math. Softw.* **22**, 469 (1996).
- [91] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, New Jersey, 1980).
- [92] K. C. Chan, R. K. Sheth, and R. Scoccimarro, *Phys. Rev. D* **96**, 103543 (2017).
- [93] P. McDonald, *Phys. Rev. D* **74**, 103512 (2006).
- [94] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, *Phys. Rev. D* **77**, 123514 (2008).
- [95] U. Seljak, *Phys. Rev. Lett.* **102**, 021302 (2009).
- [96] P. McDonald and U. Seljak, *J. Cosmol. Astropart. Phys.* **10** (2009) 007.
- [97] N. Hamaus, U. Seljak, and V. Desjacques, *Phys. Rev. D* **84**, 083509 (2011).
- [98] N. Hamaus, U. Seljak, and V. Desjacques, *Phys. Rev. D* **86**, 103513 (2012).
- [99] L. R. Abramo and K. E. Leonard, *Mon. Not. R. Astron. Soc.* **432**, 318 (2013).
- [100] P. J. E. Peebles and J. T. Yu, *Astrophys. J.* **162**, 815 (1970).
- [101] R. A. Sunyaev and Y. B. Zeldovich, *Astrophys. Space Sci.* **7**, 3 (1970).

- [102] S. Bharadwaj, *Astrophys. J.* **472**, 1 (1996).
- [103] D. J. Eisenstein, H.-J. Seo, and M. White, *Astrophys. J.* **664**, 660 (2007).
- [104] M. Crocce and R. Scoccimarro, *Phys. Rev. D* **77**, 023533 (2008).
- [105] C. Zhao, C. Tao, Y. Liang, F.-S. Kitaura, and C.-H. Chuang, *Mon. Not. R. Astron. Soc.* **459**, 2670 (2016).
- [106] N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, K. T. Mehta, and E. Kazin, *Mon. Not. R. Astron. Soc.* **427**, 2132 (2012).
- [107] A. Pisani *et al.*, arXiv:1903.05161.
- [108] K. C. Chan and L. Blot, *Phys. Rev. D* **96**, 023528 (2017).
- [109] A. Cooray and R. K. Sheth, *Phys. Rep.* **372**, 1 (2002).