



Biased Tracer Reconstruction with Halo Mass Information

Yu Liu¹, Yu Yu¹, and Baojiu Li²

¹ Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China; yuyu22@sjtu.edu.cn

² Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

Received 2020 September 28; revised 2021 February 12; accepted 2021 February 20; published 2021 April 23

Abstract

Plenty of crucial information about our universe is encoded in the cosmic large-scale structure (LSS). However, extractions of this information are usually hindered by the nonlinearities of the LSS, which can be largely alleviated by various techniques known as reconstruction. In realistic applications, the efficiencies of these methods are always degraded by many limiting factors, a quite important one being the shot noise induced by the finite number density of biased matter tracers (i.e., luminous galaxies or dark matter halos) in observations. In this work, we explore the gains of biased tracer reconstruction achieved from halo mass information, which can suppress the shot-noise component and dramatically improves the cross-correlation between tracer field and dark matter. To this end, we first closely study the clustering biases and the stochasticity properties of halo fields with various number densities under different weighting schemes, i.e., the uniform, mass, and optimal weightings. Then, we apply the biased tracer reconstruction method to these different weighted halo fields and investigate how linear bias and observational mass scatter affect the reconstruction performance. Our results demonstrate that halo masses are critical information for significantly improving the performance of biased tracer reconstruction, indicating great application potential for substantially promoting the precision of cosmological measurements (especially for baryon acoustic oscillations) in ambitious ongoing and future galaxy surveys.

Unified Astronomy Thesaurus concepts: [Cosmology \(343\)](#); [Large-scale structure of the universe \(902\)](#); [Baryon acoustic oscillations \(138\)](#)

1. Introduction

By measuring galaxy distribution in the universe, ambitious ongoing and future galaxy surveys (e.g., 4MOST, de Jong et al. 2016; PFS, Takada et al. 2014; DESI, Levi et al. 2013; LSST, LSST Science Collaboration et al. 2009; WFIRST, Doré et al. 2018; and Euclid, Laureijs et al. 2011, etc.) will map the cosmic large-scale structure (LSS) with high precision, which can further improve the accuracy of cosmological parameter inferences and greatly deepen the understanding of our universe. However, the LSS has nonlinearly evolved to become highly non-Gaussian in the late universe, which leads to the observed signatures deviating from the theoretical predictions by linear theory (i.e., nonlinear effects), makes statistical information leak into higher-order statistics, and especially blurs some critical information encoded in the LSS of the early universe (e.g., baryon acoustic oscillations (BAOs), primordial non-Gaussianities, etc.). These defects can induce undesired systematics in the analysis of observables and thus limit the constraining power of cosmic probes on cosmological parameters and various promising candidates of new physics (e.g., dark energy models, modified gravity theories, inflation models, etc.).

To reduce nonlinearities in two-point statistics, Gaussianization methods (e.g., Weinberg 1992; Neyrinck et al. 2009) to some degree can increase information content and alleviate mode coupling (Neyrinck et al. 2011) by Gaussianizing the one-point probability function of the density field. Actually, a large part of the nonlinearities are caused by large-scale bulk flows (Eisenstein et al. 2007b), which can move galaxies on average by approximately $10 h^{-1}$ Mpc from their initial locations (Park & Kim 2010; Burden et al. 2014). Thus, these local transformation methods cannot genuinely convert the matter distribution back to its earlier stage with basically no improvement in the correlation

between the final nonlinear field and its initial condition (Harnois-Déraps et al. 2013b). Several different strategies have been proposed to tackle this problem, including forward modeling (e.g., Jasche & Wandelt 2013; Kitaura 2013; Wang et al. 2013; Seljak et al. 2017; Feng et al. 2018; Modi et al. 2018, 2019; Jasche & Lavaux 2019, etc.), backward reconstruction (hereafter reconstruction, e.g., Frisch et al. 2002; Brenier et al. 2003; Eisenstein et al. 2007a; Tassev & Zaldarriaga 2012; Burden et al. 2015; Schmittfull et al. 2015, 2017; Obuljen et al. 2017; Zhu et al. 2017; Hada & Eisenstein 2018; Shi et al. 2018; Sarpa et al. 2019, etc.), and machine learning (e.g., Mao et al. 2020), which can reproduce the initial density field at different levels.

In particular, reconstruction techniques can directly reverse bulk motions by estimating the displacement field based on the observed data and have been commonly applied in BAO measurements and also have inspired many other cosmological applications (e.g., redshift space distortions (RSDs), Zhu et al. 2018; Wang et al. 2020a; velocity reconstruction, Yu & Zhu 2019, etc.). Based on the Zel'dovich approximation, Eisenstein et al. (2007a) proposed a standard reconstruction algorithm to improve BAO measurement accuracy. This technique has been tested with simulated data (e.g., Seo et al. 2008, 2010; Mehta et al. 2011; Burden et al. 2014; Achitouv & Blake 2015) and been theoretically studied (e.g., Noh et al. 2009; Padmanabhan et al. 2009; White 2015; Seo et al. 2016; Hikage et al. 2017, 2020; Chen et al. 2019), and has also been extensively applied to observation data analysis (e.g., Padmanabhan et al. 2012; Xu et al. 2012, 2013; Anderson et al. 2014; Kazin et al. 2014; Ross et al. 2015; Beutler et al. 2016, 2017; Hinton et al. 2017). In recent years, several improved new reconstruction algorithms (e.g., the isobaric reconstruction technique, Zhu et al. 2017; the iterative reconstruction technique, Schmittfull et al. 2017; Hada & Eisenstein 2018; the multigrid relaxation method, Shi et al. 2018; the extended fast action minimization method, Sarpa et al. 2019; and the fast

semidiscrete optimal transport algorithm, Lévy et al. (2020) were proposed and tested not only on matter field (e.g., Pan et al. 2017; Wang et al. 2017) but also on more realistic halo/galaxy fields (e.g., Yu et al. 2017b; Birkin et al. 2019; Hada & Eisenstein 2019; Wang & Pen 2019; Sarpa et al. 2021), showing that they can substantially bring back the initial information and recover the linear BAO signal.

The observed halos/galaxies are discrete and biased tracers of the underlying dark matter field. When applied to a biased tracer field, these reconstruction techniques will be largely limited by additional complications (e.g., halo/galaxy bias and shot noise, etc.), which can physically or numerically affect the reconstruction process, leading to a worse estimation of the displacement field compared to the case of the matter field, and then consequently induce larger errors in the recovery of the initial matter field and degrade their powers in cosmological applications especially for the reconstruction of BAO wiggles (Yu et al. 2017b; Ding et al. 2018; Birkin et al. 2019; Wang & Pen 2019). Thus, in observations, these practical issues should be addressed or taken into account to further improve the reconstruction performance.

Recently, the effects of linear bias on BAO isobaric reconstruction were theoretically investigated and modeled (Wang & Pen 2019). Also, Birkin et al. (2019) developed a biased tracer reconstruction technique by extending the method proposed in Shi et al. (2018) to include a bias scheme up to the quadratic order. Based on mass conservation, this method transforms the reconstruction problem into solving a Monge–Ampère-type equation, which can be numerically solved by the multigrid relaxation method. It was demonstrated that this biased tracer reconstruction method can help further substantially improve the recovery of initial density and linear BAO wiggles from the biased tracer field, by performing debiasing in the process of reconstruction.

With major progress in biased tracer reconstruction, the difficulties caused by halo/galaxy bias have been alleviated now, while the shot noise (due to the limited number density of tracers) is still a significant limiting factor in realistic scenarios. To mitigate this critical problem, additional information should in principle be taken into account in constructing the tracer field. The most natural idea could be the considerations of halo masses or galaxy luminosities (or stellar masses), which can be directly measured or indirectly inferred in observations (e.g., Vale & Ostriker 2004; Yang et al. 2005; Zheng et al. 2005; Xu et al. 2018; Wechsler & Tinker 2018 and references therein). Interestingly, Seljak et al. (2009) found that weighting halos by their masses can indeed suppress the halo field’s shot-noise component and can tighten the correlation between the halo field and underlying dark matter (i.e., reducing the stochasticity), dramatically improving upon the commonly used uniform weighting scheme (i.e., weighting the halos/galaxies uniformly). Afterward, Hamaus et al. (2010) developed an optimal mass-dependent halo-weighting technique. Compared to the previous mass weighting, this optimal weighting scheme can further suppress shot noise and minimize the stochasticity, which was also confirmed by Cai et al. (2011) by using a different methodology. Then, these findings triggered several specific applications (e.g., constraints on primordial non-Gaussianity, Hamaus et al. 2011, and growth rate of structure formation, Hamaus et al. 2012, etc.) and many other cosmological investigations (e.g., Baldauf et al. 2013; Jennings et al. 2015; Smith & Marian 2016; Ginzburg et al. 2017; Schmittfull et al. 2019, etc.).

Motivated by such progress, in this work we aim to extend the study of Birkin et al. (2019) with the goal of further improving the bias tracer reconstruction performance by including information on halo masses. The biased tracer reconstruction method adopted in this work uses mass conservation to relate the density fields in Lagrangian and Eulerian spaces. Naturally, we expect great improvement in the recovery of the initial matter field when the method is performed on a halo field with more information on the underlying matter distribution, which can be obtained from the information on halo masses. To corroborate this idea, in our work, we systematically perform careful analyses of various halo fields under different halo-weighting schemes, i.e., the uniform, mass, and optimal weightings (the latter two weightings are both related to halo masses). Our results suggest that halo masses undoubtedly are important information for highly improving the performance of biased tracer reconstruction, thus demonstrating enormous application potential for significantly improving the power of extracting cosmological information from ambitious current and future galaxy surveys.

This paper is organized as follows. We review briefly the biased tracer reconstruction method in Section 2. In Section 3, we introduce in detail the data used in this work. Section 4 formulates the definitions of halo-clustering bias and halo stochasticity. The results of our investigations are presented in Section 5. Finally, in Section 6, we summarize our work, give discussions, and draw our conclusion.

2. The Biased Tracer Reconstruction Method

With the assumptions of mass conservation and uniform initial matter distribution,³

$$\rho(\mathbf{x})d^3\mathbf{x} = \rho(\mathbf{q})d^3\mathbf{q} \approx \bar{\rho}d^3\mathbf{q}, \quad (1)$$

there is a unique mapping from the initial Lagrangian coordinates $\mathbf{q}$ to the final Eulerian coordinates $\mathbf{x}$ before shell crossing of structure formation (see Frisch et al. 2002; Brenier et al. 2003 and references therein),

$$\det\left(\frac{\partial q^i}{\partial x^j}\right) = \det[\nabla^i \nabla_j \Theta(\mathbf{x})] = \frac{\rho(\mathbf{x})}{\bar{\rho}} \equiv 1 + \delta_m(\mathbf{x}), \quad (2)$$

where $\det\left(\frac{\partial q^i}{\partial x^j}\right)$ is the Jacobian determinant of mapping from $\mathbf{q}$ to $\mathbf{x}$, $\delta_m(\mathbf{x})$ is the dark matter overdensity (hereafter, we also refer to it as the dark matter field), and the displacement potential Θ is defined by $\mathbf{q} = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$. Actually, with the nonlinear evolution of the LSS, shell crossing is inevitable on small scales in the late universe, which makes the mapping no longer unique and thus leads to meaningless reconstruction results below the shell-crossing scale. Nevertheless, the assumption of no shell crossing in the reconstruction process is indeed valid on relatively larger scales, which guarantees that

³ Based on the same basic assumptions, the reconstruction problem can also be treated as an optimized mass transportation problem which can be solved by using the Monge–Ampère–Kantorovich (MAK) method (see Frisch et al. 2002; Mohayaee et al. 2003; Brenier et al. 2003; Mohayaee et al. 2006; Mohayaee & Sobolevskii 2008). For a sample with N points, the fully discrete combinatorial algorithms of MAK reconstruction have a complexity of N^3 , which thus hampers their applications to big data samples in LSS study. Very recently, Lévy et al. (2020) reported a new semidiscrete algorithm with an empirical complexity of $N \log N$, which makes it significantly more efficient than previous combinatorial ones.

the algorithm can work well to recover the large-scale initial information (see Shi et al. 2018; Birkin et al. 2019). This is particularly useful for its cosmological applications (e.g., BAO and primordial non-Gaussianity measurements, etc.), because the large-scale information is more cosmologically relevant.

By expanding $\det[\nabla^i \nabla_j \Theta(\mathbf{x})]$, Equation (2) can be rewritten as

$$\frac{1}{6}(\nabla^2 \Theta)^3 - \frac{1}{2} \nabla^i \nabla_j \Theta \nabla^j \nabla_i \Theta \nabla^2 \Theta + \frac{1}{3} \nabla^i \nabla_j \Theta \nabla^j \nabla_k \Theta \nabla^k \nabla_i \Theta = 1 + \delta_m(\mathbf{x}), \quad (3)$$

where the Einstein summation convention is used. Equation (3) is a nonlinear elliptical partial differential equation (PDE), which can be numerically solved. For the convenience of numerical implementation, $\nabla^i \nabla_j \Theta$ should be split into diagonal and traceless parts by defining barred derivatives (see Shi et al. 2018 and Birkin et al. 2019 for more discussions),

$$\nabla^i \nabla_j \Theta \equiv \frac{1}{3} \delta_j^i \nabla^2 \Theta + \bar{\nabla}^i \bar{\nabla}_j \Theta, \quad (4)$$

where δ_j^i is the Kronecker delta. Thus, Equation (3) is further rewritten as

$$(\nabla^2 \Theta)^3 - \frac{9}{2} \bar{\nabla}^i \bar{\nabla}_j \Theta \bar{\nabla}^j \bar{\nabla}_i \Theta \nabla^2 \Theta + 9 \bar{\nabla}^i \bar{\nabla}_j \Theta \bar{\nabla}^j \bar{\nabla}_k \Theta \bar{\nabla}^k \bar{\nabla}_i \Theta - 27[1 + \delta_m(\mathbf{x})] = 0, \quad (5)$$

which is called the reconstruction equation.

However, the dark matter field $\delta_m(\mathbf{x})$ is not actually observable. In galaxy surveys, we measure the biased tracer field, i.e., $\delta_{\text{tracer}}(\mathbf{x}) \equiv n_{\text{tracer}}(\mathbf{x})/\bar{n}_{\text{tracer}} - 1$, where $n_{\text{tracer}}(\mathbf{x})$ is the tracer number density and $\bar{n}_{\text{tracer}}$ is its mean value. The connection between them can be described by a series of bias parameters. Here, we only consider the bias parameters up to second order, and their relation is expressed as

$$\delta(\mathbf{x})_{\text{tracer}} = b_1 \delta_m(\mathbf{x}) + \frac{b_2}{2} \delta_m^2(\mathbf{x}) + \gamma_2 \mathcal{G}_2, \quad (6)$$

where b_1 , b_2 , γ_2 are the linear, quadratic, and nonlocal bias parameters, respectively, and the nonlocal bias term $\mathcal{G}_2$ is related to the velocity potential Φ_v (see Chan et al. 2012; Birkin et al. 2019 for more details). Replacing $\delta_m(\mathbf{x})$ with $\delta_{\text{tracer}}(\mathbf{x})$ in Equation (5), we then obtain a more general version of the reconstruction equation,

$$\begin{aligned} & (\nabla^2 \Theta)^3 - \frac{9}{2} \bar{\nabla}^i \bar{\nabla}_j \Theta \bar{\nabla}^j \bar{\nabla}_i \Theta \nabla^2 \Theta + 9 \bar{\nabla}^i \bar{\nabla}_j \Theta \bar{\nabla}^j \bar{\nabla}_k \Theta \bar{\nabla}^k \bar{\nabla}_i \Theta \\ & - 27 \left[1 + \frac{\delta_{\text{tracer}}(\mathbf{x})}{b_1} - \frac{b_2}{2b_1^3} \delta_{\text{tracer}}^2(\mathbf{x}) - \frac{\gamma_2}{b_1} (\nabla^i \nabla_j \Theta \nabla^j \nabla_i \Theta \right. \\ & \left. + 4 \nabla^2 \Theta - (\nabla^2 \Theta)^2 - 6) \right] = 0. \end{aligned} \quad (7)$$

If $b_1 = 1$ and $\gamma_2 = b_2 = 0$, Equation (7) will be reduced to Equation (5). From now on, we call Equation (7) the biased tracer reconstruction equation.

Following the methodology formulated in Shi et al. (2018) and Birkin et al. (2019), Equation (7) can be iteratively solved for $\Theta(\mathbf{x})$ and $\nabla_x \Theta(\mathbf{x})$ on discrete mesh cells by using the multigrid Gauss–Seidel technique. The algorithm is

implemented by modifying the code ECOSMOG (Li et al. 2012, 2013), which is based on the publicly available N -body code RAMSES (Teyssier 2002). Then, the displacement field is given by $\Psi(\mathbf{q}) = \mathbf{x} - \mathbf{q}$, where $\mathbf{q}(\mathbf{x}) = \nabla_x \Theta(\mathbf{x})$. Based on the linear Lagrangian perturbation theory (LPT), the reconstructed density field is finally obtained by the negative divergence of the displacement field with respect to $\mathbf{q}$:

$$\delta_r = -\nabla_q \cdot \Psi(\mathbf{q}), \quad (8)$$

which indicates that only the curl-less ‘‘E-mode’’ component of $\Psi(\mathbf{q})$ is used for initial information recovery (Yu et al. 2017a). Here, the divergence is implemented using the publicly available DTFE⁴ (Delaunay Tessellation Field Estimator) code (Schaap & van de Weygaert 2000; Cautun & van de Weygaert 2011a), which is based on Delaunay tessellation.

3. Data

3.1. N-body Simulation and Halo Samples

In this work, we adopt a cosmological N -body simulation realized using the publicly available code CUBEP³M (Harnois-Déraps et al. 2013a). This simulation is initialized at redshift $z = 127$ and evolves 2048³ cold dark matter (CDM) particles with a mass resolution of $2.15 \times 10^9 h^{-1} M_\odot$ in a periodic cubic box of width $600 h^{-1}$ Mpc. Here, a Particle–Mesh process with 4096^3 grids is used for gravitational force calculation. To increase the force resolution below the mesh scale, a Particle–Particle algorithm is also involved.

CUBEP³M’s own on-the-fly spherical overdensity (SO) halo finder is employed to identify CDM halos, the masses of which are resolved down to $2.15 \times 10^{10} h^{-1} M_\odot$ with a minimum of 10 CDM particles per halo. We intend to investigate the dependence of reconstruction performances on different halo number densities (i.e., different shot-noise levels). To this end, three halo samples with number densities of $2.77 \times 10^{-2} (h^{-1} \text{Mpc})^{-3}$, $2.77 \times 10^{-3} (h^{-1} \text{Mpc})^{-3}$, and $2.77 \times 10^{-4} (h^{-1} \text{Mpc})^{-3}$ are constructed by discarding halos with masses below the mass cutoffs of $M_{\text{min}} \simeq 2.15 \times 10^{10} h^{-1} M_\odot$, $M_{\text{min}} \simeq 1.84 \times 10^{12} h^{-1} M_\odot$, and $M_{\text{min}} \simeq 2.10 \times 10^{13} h^{-1} M_\odot$, respectively. The number densities of our halo samples are roughly comparable to those of targets of various galaxy surveys listed in Table 1. These halo samples were also used in Yu et al. (2017b), where the isobaric reconstruction method was adopted for nonlinear halo reconstruction without considering halo mass and bias, etc. Note that throughout this paper, subhalos are excluded from the analyses.

3.2. Different Weighted Halo Fields

If one halo sample is divided into N mass bins and halos in the i th bin are assigned the same weight $w(M_i)$, the weighted halo field $\delta_w(\mathbf{x})$ in configuration space can be written as

$$\delta_w(\mathbf{x}) = \frac{\sum_i N_i w(M_i) \delta_i(\mathbf{x})}{\sum_i N_i w(M_i)}, \quad (9)$$

where N_i is the number of halos in the i th bin and the uniformly weighted halo overdensity $\delta_i(\mathbf{x})$ of the i th bin is defined as

$$\delta_i(\mathbf{x}) \equiv \frac{n_i(\mathbf{x})}{\bar{n}_i} - 1, \quad (10)$$

⁴ <https://www.astro.rug.nl/~voronoi/DTFE/dtfe.html>

Table 1
Various Halo Samples Used in the Reconstruction Performance Tests

Redshift	Mass Range [$h^{-1}M_{\odot}$]	Number Density [$(h^{-1}\text{Mpc})^{-3}$]	Linear Bias Parameter	The Galaxy Survey
$z = 0$	$M_h \in [2.15 \times 10^{10}, 2.11 \times 10^{15}]$	$\bar{n}_h = 2.77 \times 10^{-2}$	$b_U = 0.68, b_M = 1.52, b_O = 1.10$	LSST (LSST Science Collaboration et al. 2009); DESI BGS (DESI Collaboration et al. 2016)
$z = 0$	$M_h \in [1.84 \times 10^{12}, 2.11 \times 10^{15}]$	$\bar{n}_h = 2.77 \times 10^{-3}$	$b_U = 0.94, b_M = 1.75, b_O = 1.35$	DESI ELG; SPHEREx (Doré et al. 2014)
$z = 0$	$M_h \in [2.10 \times 10^{13}, 2.11 \times 10^{15}]$	$\bar{n}_h = 2.77 \times 10^{-4}$	$b_U = 1.55, b_M = 2.30, b_O = 1.95$	PFS (Takada et al. 2014); BOSS CMASS (Dawson et al. 2013); Euclid (Laureijs et al. 2011)

Note. Note that our analyses are based on the halo samples at $z = 0$, which is inconsistent with the redshifts of targets of these surveys. Because the primary goal of this work is to investigate the gains of bias tracer reconstruction achieved from halo mass information, this inconsistency will not affect our conclusions.

where $n_i(\mathbf{x})$ and $\bar{n}_i$ are the local and mean halo number density of the i th bin, respectively. We denote by $w(M_i) = 1$, M_i , $b(M_i)$, and $W_{\text{opt}}(M_i)$ the uniform, mass, bias,⁵ and optimal weighting schemes, respectively, for constructing the weighted halo field $\delta_w(\mathbf{x})$, where M_i is the average mass of halos in the i th bin and “optimal” can refer to the weighting scheme that minimizes the stochasticity of $\delta_w(\mathbf{x})$ with respect to the underlying dark matter (Hamaus et al. 2010; Cai et al. 2011). Hamaus et al. (2010) states that the “optimal” weighting corresponds to a nontrivial eigenvector of the stochasticity matrix,⁶ which yields the lowest eigenvalue. The eigenvector can be well fitted by a simple function with the form of $w(M) = M + M_0$, where M_0 is a free parameter.

Actually, except for the uniform weighting, increasing the number of halo bins can further suppress the halo stochasticity (Hamaus et al. 2010). Thus, in practical applications, each halo of mass M should be assigned a weight $w(M)$ according to a smooth weighting function, corresponding to the number of bins approaching infinity (i.e., $N \rightarrow +\infty$). In our following performance tests, the smooth uniform, mass, and optimal weighting functions are adopted with the forms of $w(M) = 1$, M , $M + M_0$, respectively. Note that the last two weightings are both dependent on the halo mass information and thus are also collectively called mass-dependent weightings in this paper. In Hamaus et al. (2010), the M_0 was optimized iteratively by minimizing the $\delta_w(\mathbf{x})$ ’s shot-noise level, and they found the best value of M_0 depends on the cutoff mass M_{min} and roughly satisfies the relation of $M \simeq 3M_{\text{min}}$ within their tested domain.

However, the simulation used in this work has a higher mass resolution and resolves halos with lower masses. The halo-finder algorithm applied here (the SO method) is also different from theirs (the friends-of-friends method). Considering these differences, the empirical relation $M_0 \simeq 3M_{\text{min}}$ may not hold in our cases, in particular for our halo sample with the highest number density (i.e., $2.77 \times 10^{-2} (h^{-1}\text{Mpc})^{-3}$). Moreover, in our study, we prefer to seek out the optimal halo weighting for maximizing the performance of bias tracer reconstruction, which not only depends on the $\delta_w(\mathbf{x})$ shot-noise level but also on the bias of $\delta_w(\mathbf{x})$ with respect to dark matter. Given all that, in our work, the optimal M_0 is determined by maximizing the reconstruction performance (rather than minimizing the shot-noise level of $\delta_w(\mathbf{x})$; e.g., Hamaus et al. 2010). For the three halo samples with number densities in descending order, we find the optimal M_0 are approximately $70M_{\text{min}} \simeq 1.5 \times 10^{12} h^{-1} M_{\odot}$, $6M_{\text{min}} \simeq 1.1 \times 10^{13} h^{-1} M_{\odot}$, and $3M_{\text{min}} \simeq 6.3 \times 10^{13} h^{-1} M_{\odot}$, respectively.

3.3. The Mass Assignment Method

Yu et al. (2017b) opted for the DTFE mass assignment (Cautun & van de Weygaert 2011b) to generate halo fields to

⁵ Various previous works (e.g., Seljak et al. 2009; Hamaus et al. 2010; Cai et al. 2011) found that bias weighting only can lead to marginal or even basically no improvements relative to the uniform weighting for suppressing halo stochasticity, which is also confirmed in our check. This is because bias weighting is close to uniform weighting, which makes the implementation of local mass and momentum conservation inefficient for LSS evolution (Seljak et al. 2009). Therefore, investigations on the gains of bias tracer reconstruction achieved from the bias weighting scheme are omitted in this paper.

⁶ In Hamaus et al. (2010), multiple mass bins with an equal number of halos are used for the derivation. The stochasticity matrix is defined as $C_{ij} \equiv \langle (\delta_i(\mathbf{k}) - b_i(k)\delta_m(\mathbf{k}))(\delta_j(\mathbf{k}) - b_j(k)\delta_m(\mathbf{k}))^* \rangle$ in Fourier space, where the bias of the i th halo bin is determined by $b_i(k) = \frac{\langle \delta_i(\mathbf{k})\delta_m(\mathbf{k})^* \rangle}{\langle |\delta_m(\mathbf{k})|^2 \rangle}$.

avoid the instability of the isobaric reconstruction algorithm due to the sparseness of the halo samples. However, the DTFE scheme can induce excessive smoothing of the low-density regions of the halo field through a special window function of Delaunay tessellation, which inevitably erases some spatial distribution information of halos, thus weakening the reconstruction performance (Yu et al. 2017b; Birkin et al. 2019). This halo sparseness issue can be tackled well by the biased tracer reconstruction technique as demonstrated in Birkin et al. (2019). They found that the traditional cloud-in-cell (CIC) and triangular-shaped-cloud mass assignments can achieve similar reconstruction performance and present great improvements over the DTFE method. Given that the CIC method is more commonly used, in our work, the matter/halo fields are constructed via CIC interpolation of particles/halos onto cubical meshes with 512^3 grids (see Figure 1), where the grid cell size ($\sim 1.2h^{-1}\text{Mpc}$) is sufficient for the convergence of the biased tracer reconstruction performance (Birkin et al. 2019).

4. Bias and Stochasticity

Forming in density peak regions, halos are merely biased and stochastic tracers of the underlying dark matter field. In Fourier space, the relation between halo overdensity $\delta_h(\mathbf{k})$ and dark matter overdensity $\delta_m(\mathbf{k})$ can be expressed as

$$\delta_h(\mathbf{k}) = b(k)\delta_m(\mathbf{k}) + \varepsilon(\mathbf{k}), \quad (11)$$

where the bias is defined as

$$b(k) = \frac{\langle \delta_h(\mathbf{k})\delta_m(\mathbf{k})^* \rangle}{\langle |\delta_m(\mathbf{k})|^2 \rangle}, \quad (12)$$

and the stochastic term $\varepsilon(\mathbf{k})$ is assumed to be uncorrelated with the dark matter field (i.e., $\langle \varepsilon(\mathbf{k})\delta_m(\mathbf{k})^* \rangle = 0$). Note that $b(k)\delta_m(\mathbf{k})$ is completely correlated with the dark matter field. Hence, the measured halo spectrum $\hat{P}_{hh}(k)$ can be decomposed into two terms:

$$\hat{P}_{hh}(k) = P_{hh}(k) + \sigma^2(k) \quad (13)$$

with

$$P_{hh}(k) = b^2(k)P_{mm}(k), \quad (14)$$

where $\sigma^2(k) = \langle |\varepsilon(\mathbf{k})|^2 \rangle$ is the stochastic noise and $P_{mm}(k) = \langle |\delta_m(\mathbf{k})|^2 \rangle$ is the dark matter spectrum. This decomposition makes the stochastic noise $\sigma^2(k)$ contain all sources of stochasticity between halos and dark matter (Seljak et al. 2009). For a uniformly weighted halo field, the stochastic noise $\sigma^2(k)$ is usually modeled as the Poisson shot noise, which is given by the inverse of the halo number density $1/\bar{n}_h$. If each halo is weighted by w_i , the expected shot noise is generalized to be

$$\sigma_{\text{exp}}^2(k) = \frac{V \sum_{i=1}^N w_i^2}{(\sum_{i=1}^N w_i)^2}, \quad (15)$$

where V is the volume of the simulation box and N is the number of halos.

Halo stochasticity can lead to a lack of coherence between the halo field and dark matter field, and thus is usually

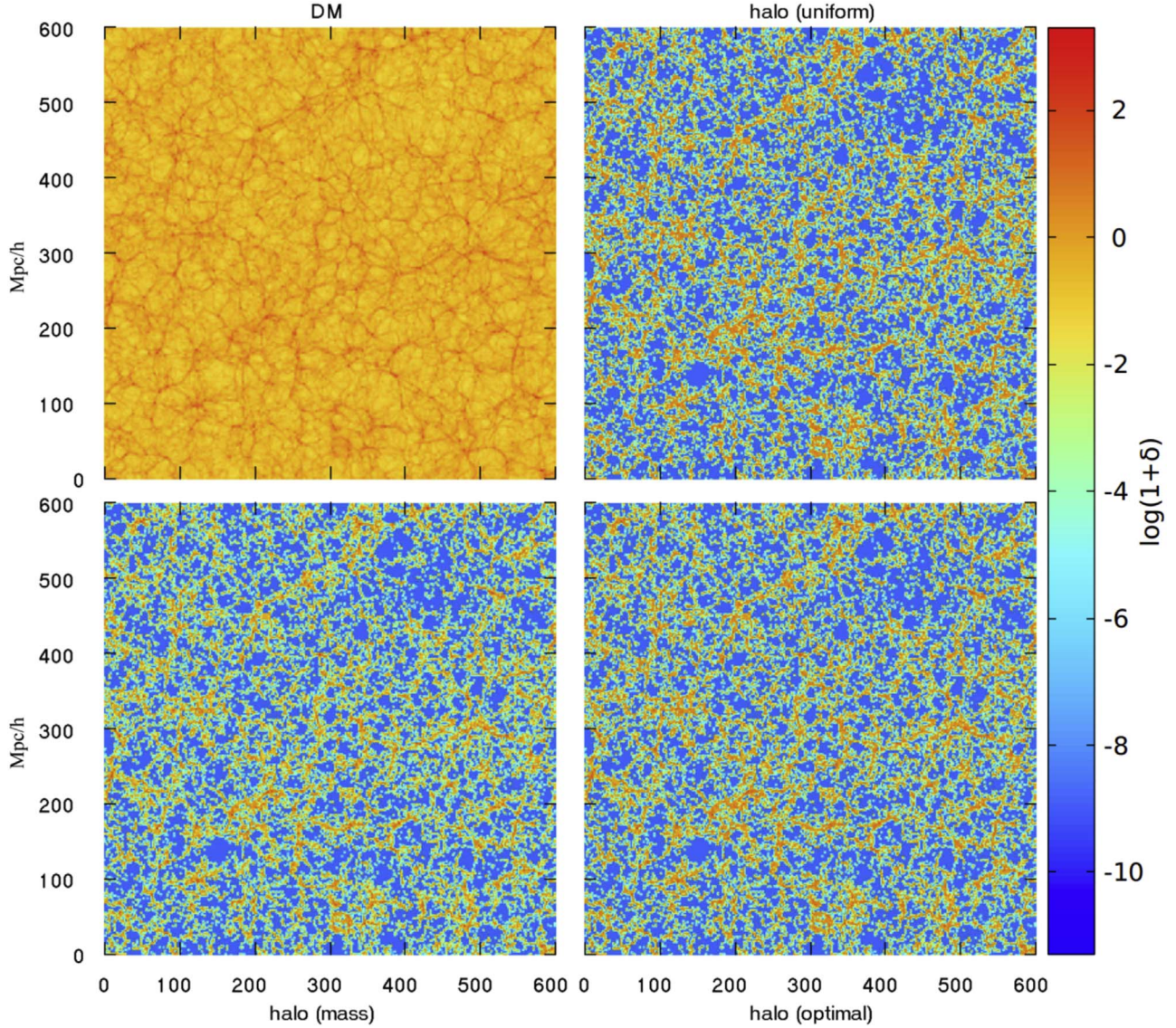


Figure 1. The two-dimensional slices of the dark matter field (top-left) and different weighted halo fields with the halo number density of $2.77 \times 10^{-2} (h^{-1} \text{Mpc})^{-3}$ before reconstruction, at $z = 0$. These fields are constructed by the CIC method. Each slice shows the same $600 \times 600 (h^{-1} \text{Mpc})^2$ region with a $5.86 h^{-1} \text{Mpc}$ thickness of the simulation box.

described by the cross-correlation coefficient:

$$r_{hm}(k) \equiv \frac{P_{hm}(k)}{\sqrt{\hat{P}_{hh}(k)P_{mm}(k)}} = \left(1 + \frac{\sigma^2(k)}{b^2(k)P_{mm}(k)}\right)^{-1/2}, \quad (16)$$

where $p_{hm}(k)$ is the cross-power spectrum of the two fields. Here, we utilized Equations (13) and (14) to derive the right-hand side of Equation (16), which explicitly indicates that the existence of stochastic noise $\sigma^2(k)$ will make $r_{hm}(k)$ deviate from unity, and the extent of the deviation also depends on the bias $b(k)$. If there is no stochasticity (i.e., $r_{hm}(k) = 1$), it means that the spatial distribution of dark matter can be derived from that of halos once the bias $b(k)$ is known (Bonoli & Pen 2009). In particular, Equation (16) can be further rewritten as

$$\frac{\sigma^2(k)}{P_{hh}(k)} = \frac{1 - r_{hm}^2(k)}{r_{hm}^2(k)}, \quad (17)$$

which is called the inverse signal-to-noise ratio (hereafter N/S ratio) of the halo spectrum and is also used to describe the halo stochasticity⁷ (see Hamaus et al. 2010).

5. Results

5.1. The Halo-clustering Bias

The halo-clustering bias is scale dependent, especially on small scales. A significant scale dependence of the bias can be treated as a sign of non-negligible deviations from zero for the higher-order bias parameters in Equation (6). Nevertheless, on large scales ($k \lesssim 0.1 h \text{Mpc}^{-1}$), it is expected to be a constant, which corresponds to the linear bias parameter in Equation (6).

⁷ Other statistical descriptions of halo stochasticity also can be found in the literature, e.g., $S(k) \equiv \sqrt{2(1 - r_{hm}(k))}$ (Seljak & Warren 2004; Bonoli & Pen 2009) and $E(k) \equiv \sqrt{1 - r_{hm}^2(k)}$ (Cai et al. 2011). Nevertheless, they all depend only on the cross-correlation coefficient between halos and dark matter, including Equation (17).

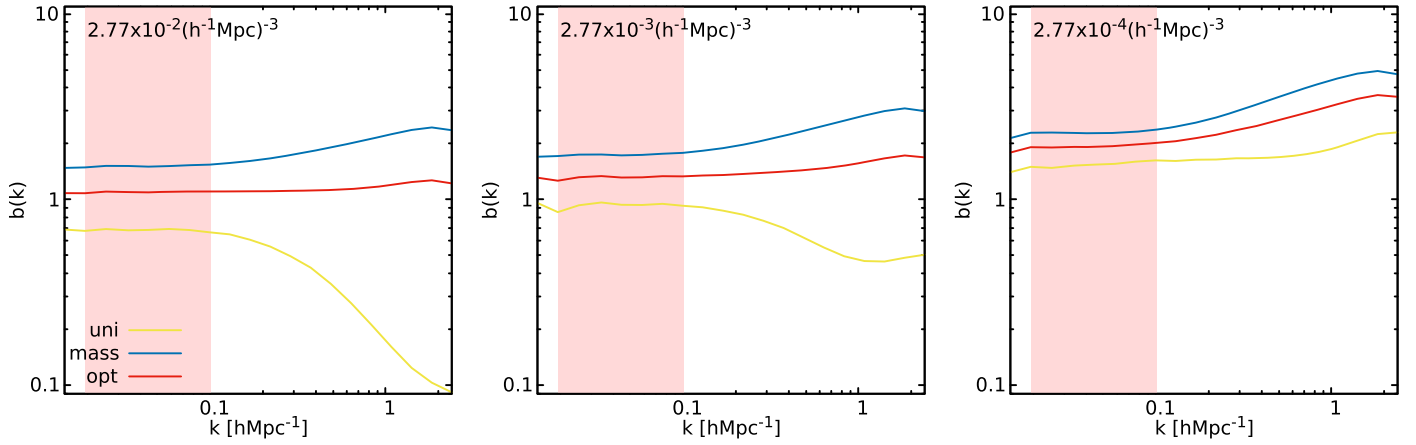


Figure 2. The clustering biases of different weighted halo fields. From left to right, the three panels correspond to different halo samples with number densities of 2.77×10^{-2} , 2.77×10^{-3} , and $2.77 \times 10^{-4} (h^{-1} \text{Mpc})^{-3}$, respectively (the same below). The yellow, blue, and red curves correspond to the uniform, mass, and optimal weighting schemes (the same below). The shaded regions indicate the scales of $0.025 h \text{ Mpc}^{-1} \lesssim k \lesssim 0.099 h \text{ Mpc}^{-1}$ for calculating the linear bias parameters.

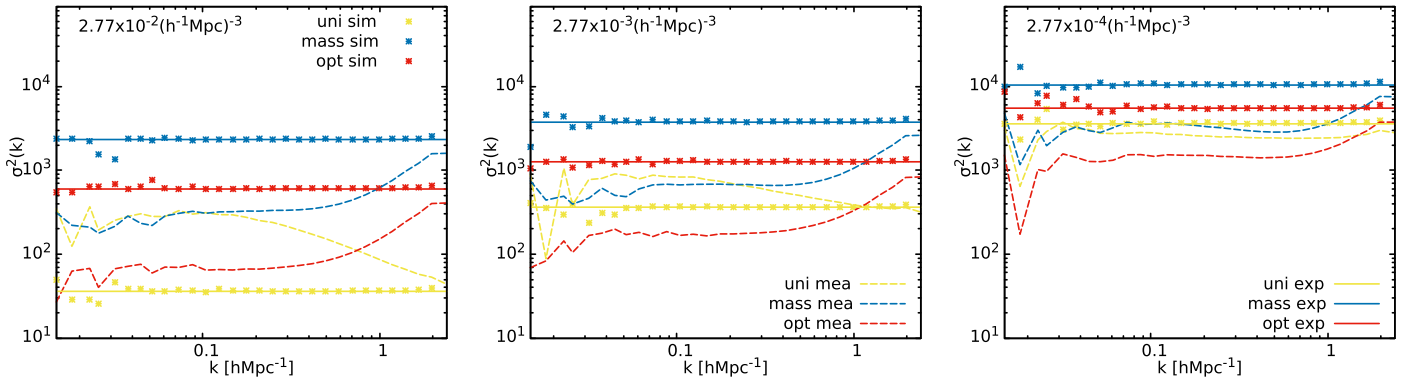


Figure 3. Various shot noises under different weighting schemes. The asterisks, dashed lines, and solid lines correspond to the simulated, measured, and expected shot noises, respectively.

This constant offset in the large-scale clustering amplitude relative to dark matter can be corrected to reconstruct the dark matter power spectrum (Sheth & Tormen 1999), while due to the scale dependence of the bias and other uncertainties, the small-scale galaxy/halo-clustering information is usually discarded. Moreover, the more massive halos should be more biased, because they have relevant higher clustering.

In Figure 2, we plot the biases of the different weighted halo fields by using Equation (12). For the uniform weighting case, we see that with the increase of halo number density, the bias on small scales (i.e., $k \gtrsim 0.1 h \text{ Mpc}^{-1}$) becomes more scale dependent. We checked that this is a combined result of various bias behaviors of halo populations with different masses⁸, which is also applicable to other weighting cases.

In our study, we take the average of the bias on scales $0.025 h \text{ Mpc}^{-1} \lesssim k \lesssim 0.099 h \text{ Mpc}^{-1}$ (corresponding to the shaded regions in Figure 2) to serve as the linear bias parameter (see Table 1). It is worth noting that for the two halo samples with higher halo number densities, the optimal weighting makes their biases more scale independent relative to the other two weighting schemes. This may provide a more

reliable way to determine dark matter spectrum directly from the optimally weighted halo spectrum by only using the linear bias model, which thus helps to extract precise information from LSS surveys (e.g., the primordial non-Gaussianity, signatures of massive neutrinos, etc.) and to minimize the systematic shifts in the BAO position relative to the dark matter (see Angulo et al. 2008; Smith et al. 2008; Zhang 2008). In addition, for this study, bias being more scale independent also helps to adequately maximize the performance of bias tracer reconstruction by only considering the linear bias parameter of Equation (6) (i.e., setting higher-order bias parameters to zero).

5.2. The Stochastic Noises of Different Weighted Halo Fields

In Figure 3, we show the expected shot noises of different weighted halo fields (i.e., the solid lines), obtained by using Equation (15). And, if halos are randomly distributed in space, for any weighting scheme, the corresponding bias should be zero (as the corresponding random halo field has no correlation with the dark matter field, i.e., $\langle \delta_h(\mathbf{k}) \delta_m(\mathbf{k})^* \rangle = 0$; see Equation (12)), thus, the corresponding halo power spectrum is completely contributed by the stochastic noise (see Equations (13) and (14)). Here, we call this noise the simulated shot noise, which should in principle be equal to its Poisson expectation. Indeed, as shown in Figure 3, the simulated shot noises (i.e., the asterisk points) are well matched with their expectations, which could serve as a nicety test of our numerical implementations.

⁸ To check this, we also measured the biases of different halo mass bins. Specifically, we split our largest sample into 10 mass bins with equal logarithmic mass intervals, given that halo abundances drop sharply with the increase in mass. Indeed, we find that the biases of different halo populations have different trends on small scales, i.e., it tends to rise for high-mass halos and tends to fall for low-mass halos.

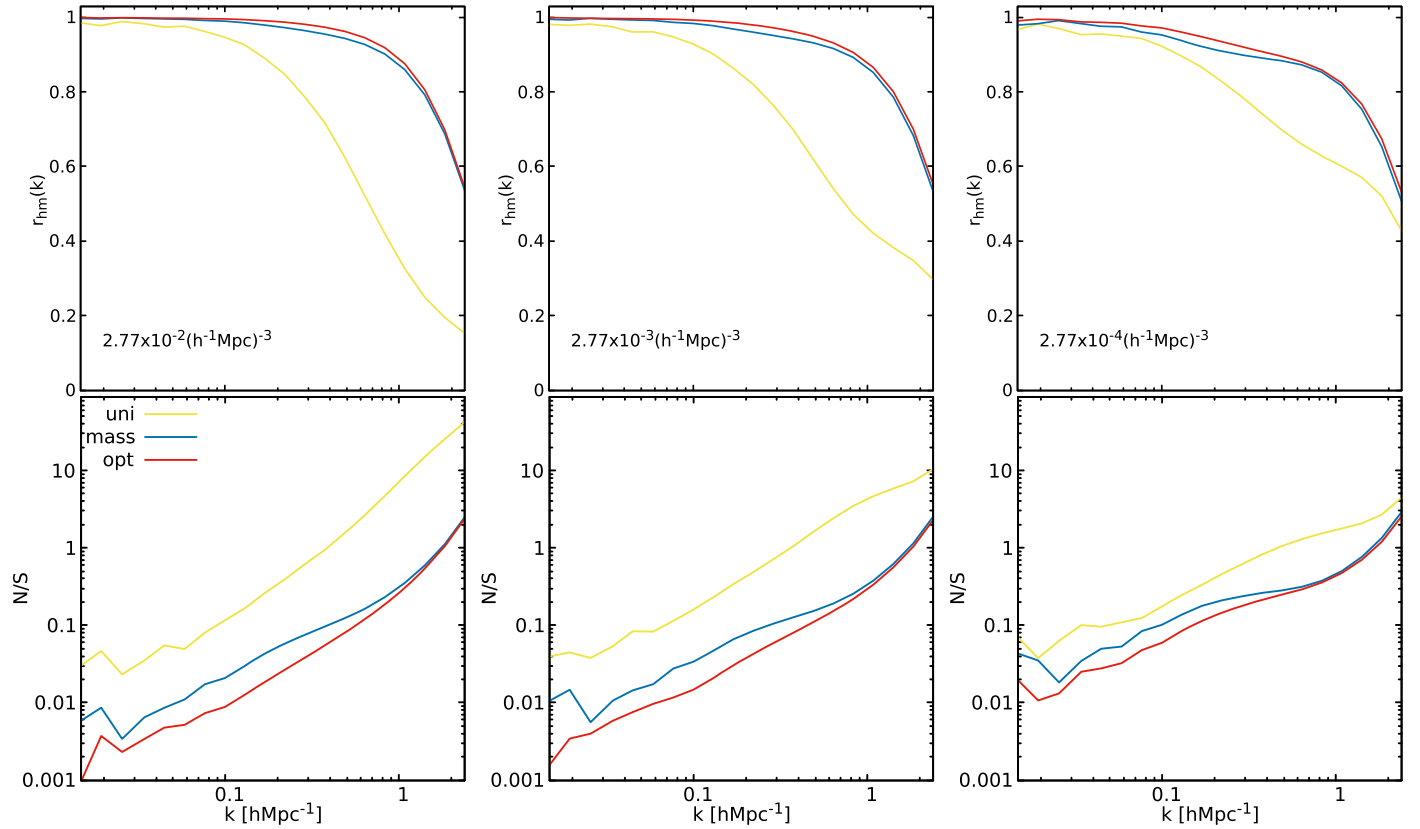


Figure 4. Top: the cross-correlation coefficients between different weighted halo fields and dark matter field, $r_{hm}(k)$. Bottom: the N/S ratios of different weighted halo power spectra, defined as $\frac{1 - r_{hm}^2(k)}{r_{hm}^2(k)}$ (see Equation (17)). The N/S ratios can be used to amplify the differences at large scales where the r_{hm} are very close to unity.

Actually, due to halo exclusion⁹ and clustering, which violate the Poisson assumption of placing point particles randomly in space, the halo stochastic noise turns out to be non-Poissonian (see Baldauf et al. 2013; Ginzburg et al. 2017; Paech et al. 2017; and references therein). Here, we also call the halo stochastic noise as the measured shot noise, which is obtained by using Equation (13). The halo exclusion and nonlinear clustering can lead to the measured shot noise being sub- and super-Poissonian, respectively. And, the amplitudes of these deviations from the Poisson expectation also depend on the bias parameters and mass ranges (Baldauf et al. 2013). Specifically, with larger exclusion scales and higher linear bias parameter, the high-mass halos are most affected by the exclusion effect, while the nonlinear clustering effect is most important for low-mass halos, where the value of the second-order bias parameter is nonzero and the exclusion effect is small (see Baldauf et al. 2013; Paech et al. 2017 for more detailed discussions). As shown in Figure 3, for the uniform weighting case, the non-Poissonian behaviors of the measured shot noises are actually caused by the two competing effects.

For the mass and optimal weighting cases, the amplitudes of the measured shot noises are significantly below the Poissonian predictions (see Figure 3), which have been demonstrated in Seljak et al. (2009) and Hamaus et al. (2010). This phenomenology indicates that mass-dependent weightings can bring some extra information to considerably suppress the shot-noise component. We argue that this information should originate from

the environment dependence of halos to meet the requirements of local mass and momentum conservation (Seljak et al. 2009), considering that the masses of halos have tight relations with the halo local environments (e.g., the local halo number density within some distance; Haas et al. 2012; Zhao et al. 2015). We can erase this information by shuffling the masses and positions of halos separately and randomly recombining them. After this procedure, the measured shot noises of the shuffled halos therefore have similar non-Poisson behaviors to the uniform weighting case, where these noises are dubbed the shuffled shot noise in this paper (see Appendix A).

5.3. The Correlation between Halos and Dark Matter

In Figure 4, we present the cross-correlation coefficients between different weighted halo fields and dark matter field, which describe the similarities with the dark matter field. In addition, we also show the N/S ratios of different weighted halo power spectra (see Equation (17)) in Figure 4, which helps amplify the differences at large scales where the cross-correlation coefficients are very close to unity. Benefiting from suppression of shot noise and boosting of clustering bias (see Figure 2 and Equation (16)), the mass and optimally weighted halo fields have markedly stronger correlations with the dark matter field (see Seljak et al. 2009; Hamaus et al. 2010), which means that these weighted halo fields retain more information of the dark matter distribution.

Generally, all above presented results suggest that mass-dependent weightings could be useful to improve the precision of cosmological parameter estimation and to reduce undesirable systematics of BAO measurements by using biased tracers. In

⁹ Halo exclusion refers to the fact that two halos cannot be too close to each other arbitrarily, corresponding to the halo two-point correlation function $\xi_{hh}(r)$ approaching -1 at the halo separation r less than the average diameter of halos.

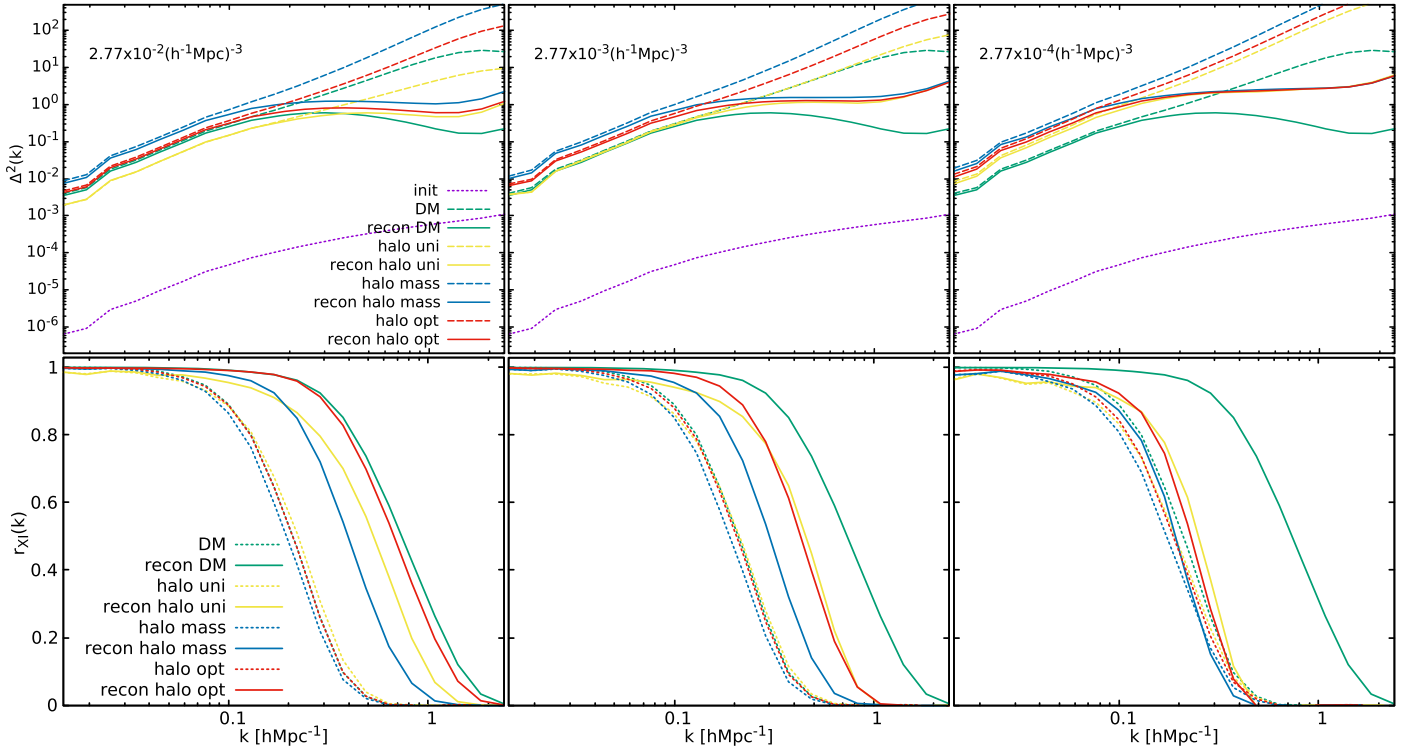


Figure 5. Top: the power spectra of different fields. The dashed lines and solid lines correspond to the fields before and after reconstruction, respectively. The green curves are for the dark matter field (the same below). For reference, we also plot the power spectrum of the initial condition, which is presented by the dotted blue line. Bottom: the cross-correlation coefficients with the initial condition for different fields pre- and post-reconstruction. The cross-correlation coefficient between the reconstructed dark matter field and initial condition (solid green line) is used to serve as an upper bound of the performance of biased tracer reconstruction. Note that here we perform the biased tracer reconstruction directly without preprocessing halo-clustering bias.

particular, mass-dependent weightings could potentially lead to dramatic improvements in the efficiency of the biased tracer reconstruction, which we will investigate in the following sections.

5.4. Direct Reconstruction without Preprocessing Bias

To quantify the initial information successfully recovered by the reconstruction algorithm (or called as reconstruction performance), we calculate the cross-correlation coefficients between the simulation’s initial condition and the different weighted halo fields before and after reconstruction, i.e., $r_{\chi}(k) \equiv \frac{P_{\chi I}(k)}{\sqrt{P_{\chi\chi}(k)P_{II}(k)}}$, where “X” and “I” stand for the field pre- or post-reconstruction and the initial condition respectively (see Figure 5). As reference, the results for the dark matter field case are also shown. Compared to the underlying dark matter, the discrete and biased tracers suffer from various additional complications (e.g., the sparseness and bias etc.), which will undoubtedly degrade the reconstruction performance in practical applications. Therefore, in the following tests, the cross-correlation coefficient between the reconstructed dark matter field and the initial condition could serve as an upper bound on the initial information, which can be recovered from a halo field.

We first investigate the consequences of direct reconstruction without any preprocessing of clustering bias (i.e., setting $b_1 = 1$ and $\gamma_2 = b_2 = 0$ respectively), which can help us understand the bias effects on the reconstruction performance. The results are presented in Figure 5, where the upper and bottom panels respectively show the power spectra and the cross-correlation coefficients with the initial condition for different fields pre- and post-reconstruction. Additionally, in the top panels, the power spectrum of the initial condition is also plotted as a reference.

We see that the process of reconstruction mainly modifies the power spectra on relatively small scales (i.e., $k > 0.1 h\text{Mpc}^{-1}$) and the power spectra pre- and post-reconstruction have similar shapes on sufficiently large scales (also see Yu et al. 2017b). In general, the reconstruction improves the correlation coefficients depending on the considered halo number density (see Yu et al. 2017b; Birkin et al. 2019). We find that the reconstruction performances for different weighting schemes obey $r_{\text{UI}}(k) < r_{\text{MI}}(k) < r_{\text{OI}}(k)$ (“U,” “M,” and “O” denote uniform, mass, and optimal weighting respectively) on very large scales, while the relationship among them becomes uncertain on relatively small scales, which should be attributed to the wrong assumption of bias parameters (see Birkin et al. 2019; Wang & Pen 2019). In particular, for the optimal weighting, we show an excellent result for the halo sample with the highest number density, as the corresponding bias is quite close to unity and also very scale independent (see Figure 2 and Table 1).

5.5. Reconstruction Considering Linear Bias

The bias between tracers and underlying dark matter is an important limiting factor for the efficiency of reconstruction. Inappropriate treatment of bias in reconstruction will lead to errors in the estimation of the displacement field and thus significantly degrade the recovery of the initial matter distribution (Wang & Pen 2019). Birkin et al. (2019) found that the performance of biased tracer reconstruction is most sensitive to the linear bias parameter and is only marginally affected by higher-order bias parameters in their tests. The main purpose of this work is to validate the improvements in the performance of biased tracer reconstruction by considering halo mass information. In view of the quasi-scale-independent

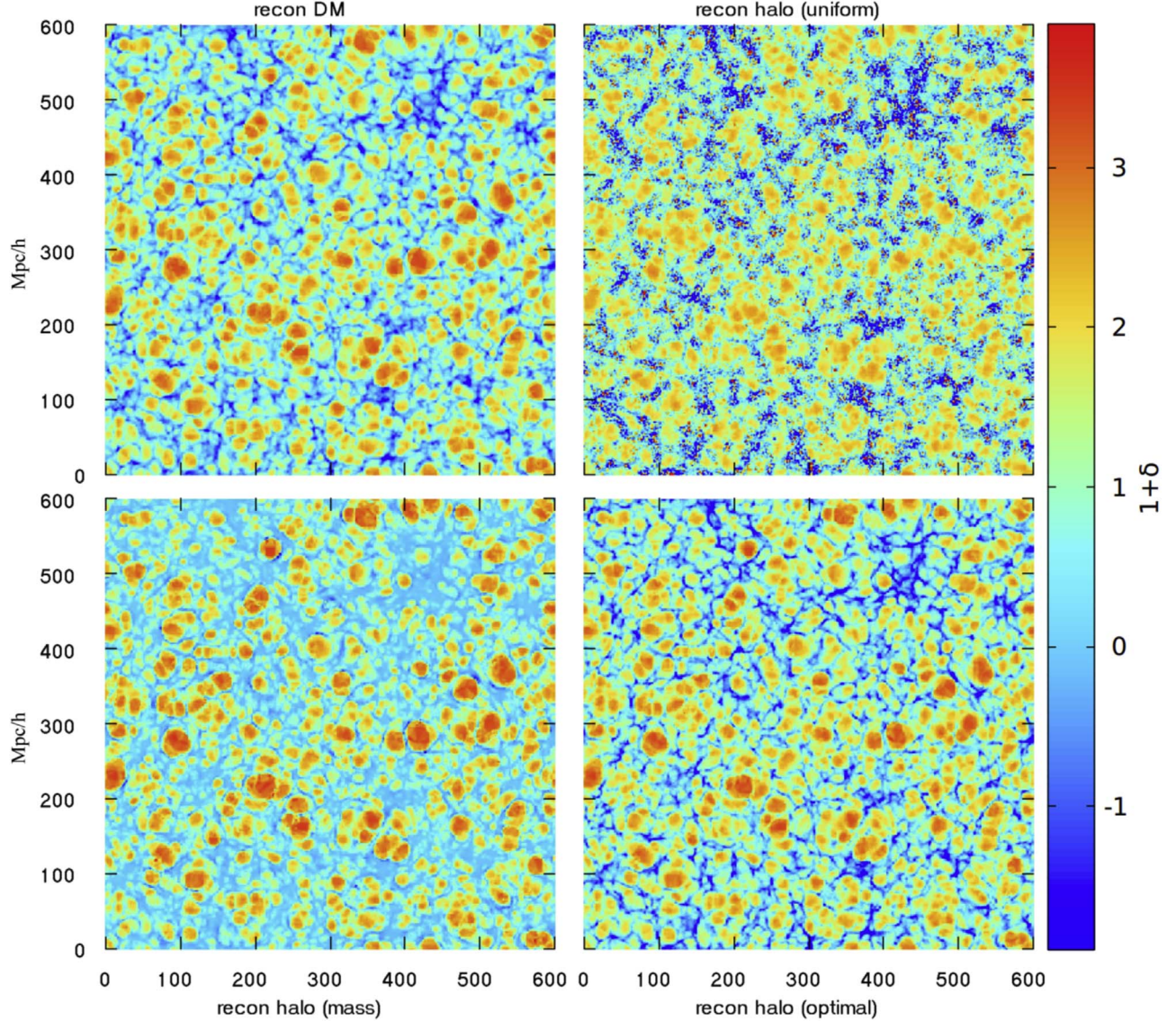


Figure 6. The two-dimensional slices of the dark matter field (top-left) and different weighted halo fields with the halo number density of $2.77 \times 10^{-2} (h^{-1} \text{Mpc})^{-3}$ after reconstruction. Each slice shows the same region (with the same projection depth) as the corresponding slice in Figure 1. Note that the linear biases are corrected in the halo cases.

features of the biases under mass-dependent weightings (mostly for the scenarios of higher halo number densities, see Figure 2), we will only consider b_1 (i.e., setting $\gamma_2 = b_2 = 0$) in the reconstruction procedure for our performance tests.

Before proceeding to test the performance quantitatively, we first show a visual comparison of different fields after reconstruction. The two-dimensional slices of projected overdensities are presented in Figure 6, where for the reconstructed halo fields we only show the case of $\bar{n}_h = 2.77 \times 10^{-2} (h^{-1} \text{Mpc})^{-3}$ to quote our results.¹⁰ We see that the reconstructed halo field under uniform weighting seems to be somewhat noisy, while for the other two mass-dependent weighting cases, the reconstructed halo fields are relatively smoother, which could be treated as a benefit brought by the information of halo masses. Visually, we find that the reconstruction results for the dark matter and the optimally weighted halos are quite similar (see Figure 6), showing the great potential of optimal weighting to highly improve the efficiency of biased tracer reconstruction.

The quantitative results are presented in Figure 7. We also plot $S_{XI}(k) \equiv \sqrt{2(1 - r_{XI}(k))}$ to amplify the differences at large scales, where $r_{XI}(k)$ is very close to unity. We find that linear debiasing can largely improve the performance for the mass and optimal weighting cases, while it is a little complicated in the uniform weighting case. Specifically, for the uniform weighting case, the bias tends to be more scale dependent (i.e., the bias changes more significantly with increase in k) on small scales with the increase in halo number density (see the yellow curves in Figure 2). For the scenarios with higher halo number densities, this implies that a simple linear debiasing is inadequate, and more sophisticated nonlinear debiasing schemes are needed to fully maximize the performance. Therefore, for the $\bar{n}_h = 2.77 \times 10^{-2}$ and $2.77 \times 10^{-3} (h^{-1} \text{Mpc})^{-3}$ cases, the $r_{UI}(k)$ drop more steeply with the increase of k on small scales, compared to the direct reconstruction (see the solid yellow lines in the bottom panels of Figure 5 and middle panels of Figure 7).

Nevertheless, under uniform weighting, linear debiasing indeed boosts $r_{UI}(k)$ on relatively large scales ($k < 0.3 h \text{Mpc}^{-1}$)

¹⁰ Other halo number density cases can be found in Appendix B.

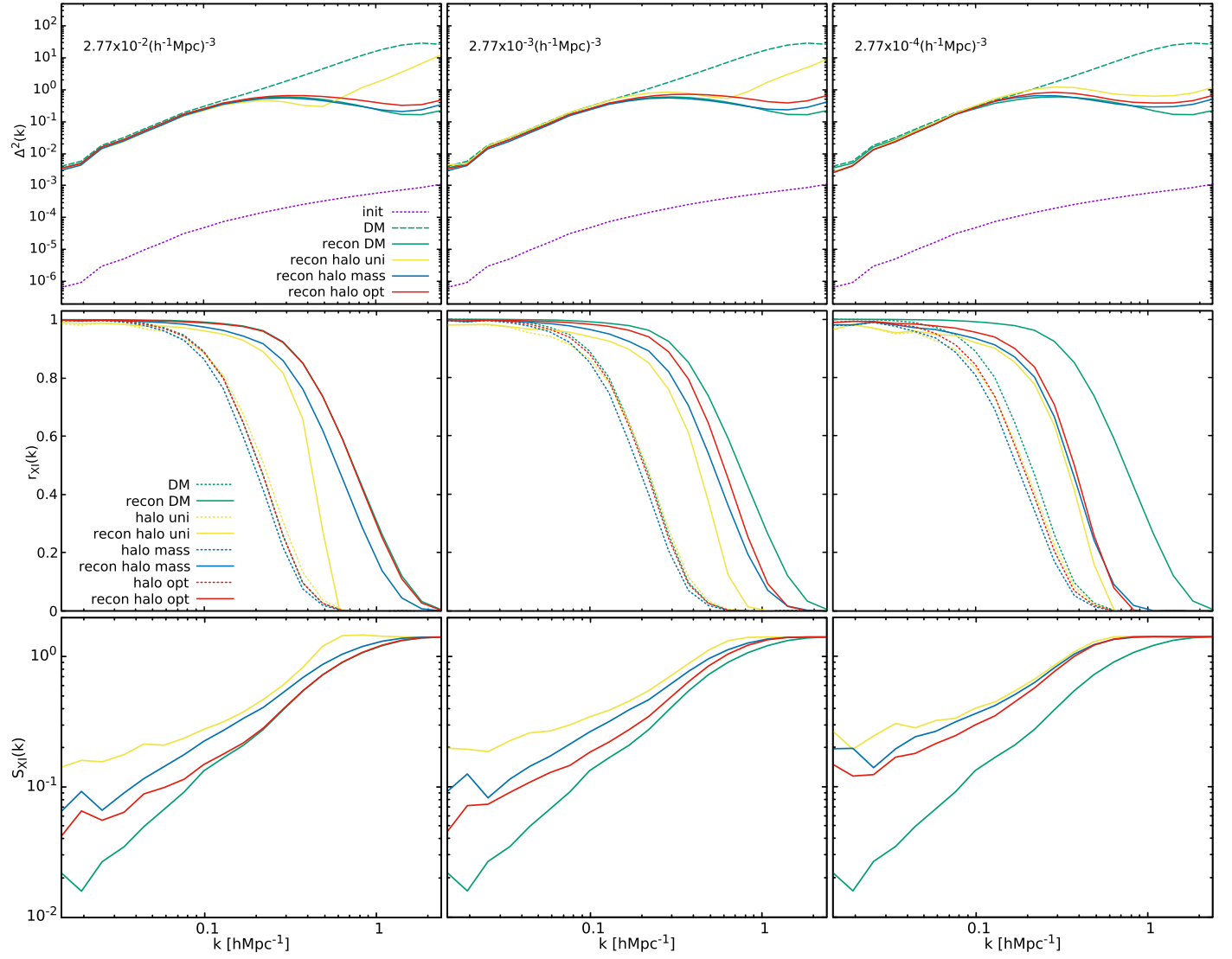


Figure 7. Top: the power spectra of different reconstructed fields (solid lines). For comparison, we also plot the power spectra of the pre-reconstructed dark matter field (dashed green line) and initial condition (dotted blue line). Middle: the cross-correlation coefficients with initial condition for different fields pre- and post-reconstruction, $r_{XI}(k)$. Bottom: the $S_{XI}(k)$. The $S_{XI}(k)$ is defined as $S_{XI}(k) \equiv \sqrt{2(1 - r_{XI}(k))}$, which can help amplify the differences at large scales where the r_{XI} are too close to unity. Here, we consider the linear clustering bias to perform the biased tracer reconstruction.

for $\bar{n}_h = 2.77 \times 10^{-2} (h^{-1}\text{Mpc})^{-3}$, while for $2.77 \times 10^{-3} (h^{-1}\text{Mpc})^{-3}$, $r_{UI}(k)$ basically has no improvements on scales of $k < 0.2h \text{ Mpc}^{-1}$ after debiasing b_1 , which should be due to the fact that b_1 is close to unity (here $b_1 = 0.94$; see Figure 2 and Table 1). And, for the massive halo sample with the lowest number density, we find that linear debiasing is quite important to improve the performance because the bias deviates from unity and is also very scale independent (see Figure 2). These results suggest the importance of linear debiasing in reconstruction-related cosmological studies, where large-scale information is more concerned. For example, the power spectrum from a typical galaxy survey basically contains BAO signals extending to the scale of $k = 0.3 \sim 0.4h \text{ Mpc}^{-1}$, signals on scales smaller than which are not detectable due to the poor signal-to-noise ratio, thus the recovery of information larger than this scale is enough to restore the BAO signals.

On the whole, as expected, the performance indeed obeys $r_{UI}(k) < r_{MI}(k) < r_{OI}(k)$ after linear debiasing in the procedure of reconstruction, consistent with the conjecture in the former sections. This due to fact that halo fields under the mass-dependent

weighting schemes reserve more information of dark matter distribution and have lower shot noises. On the other hand, the validity of this reconstruction algorithm is based on the assumption of mass conservation in LSS evolution (see Equation (1)), and mass-dependent weightings can be treated as suitable implementations of the idea to enforce the local mass and momentum conservation, which reasonably improve the efficiency of the algorithm. This conclusion should also apply to the isobaric reconstruction (Zhu et al. 2017) and other improved algorithms (Schmittfull et al. 2017; Hada & Eisenstein 2018; Lévy et al. 2020), which also use mass conservation to relate the density fields in Lagrangian and Eulerian spaces. In particular, the optimal weighting has much greater advantages to improve the reconstruction performance relative to the usually used uniform weighting scheme, and can even work perfectly for the halo sample with the highest number density. This should make optimal weighting quite interesting for ongoing and future galaxy surveys (e.g., DESI and LSST) to measure the BAO characteristic scale and other cosmological observables, which suffer from the nonlinearities of LSS.

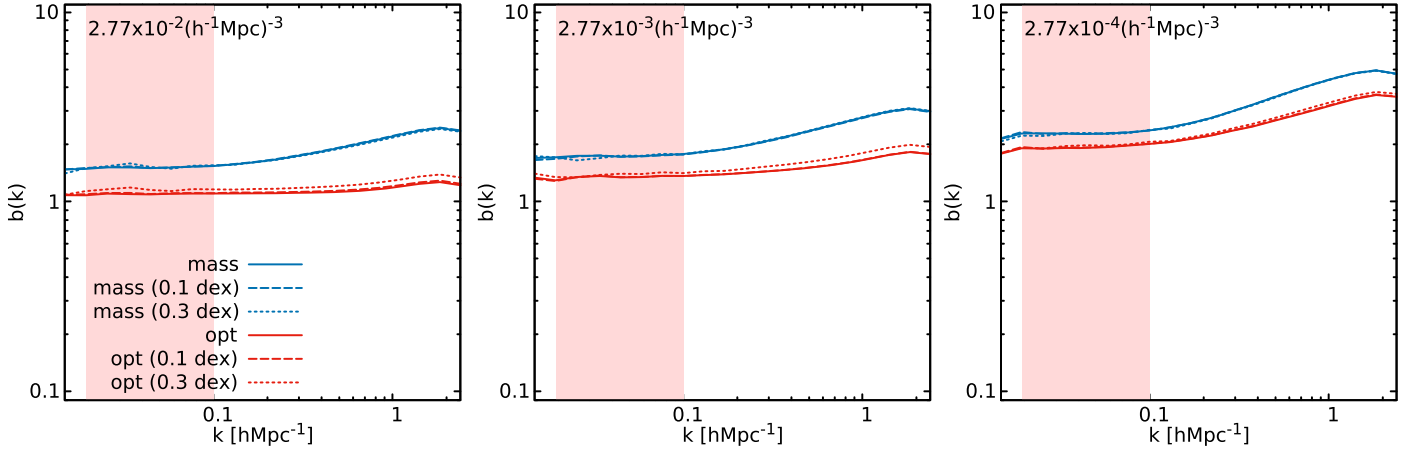


Figure 8. The clustering biases of the mass and optimally weighted halo fields with various halo mass scatters. The dashed lines and dotted lines correspond to the mass scatters of 0.1 dex and 0.3 dex, respectively (the same below). For comparison, the clustering biases of mass and optimally weighted halo fields without mass scatters (solid lines) are also presented.

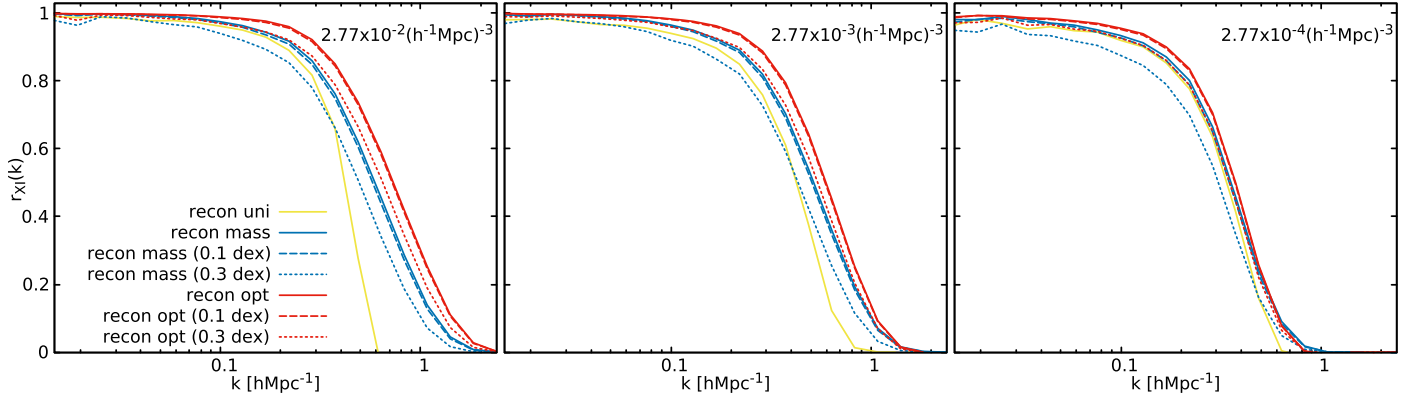


Figure 9. The cross-correlation coefficients with the initial condition for the reconstructed mass and optimally weighted halo fields with various halo mass scatters. For comparison, we also plot the cross-correlation coefficients with the initial condition for various reconstructed halo fields without mass scatters (solid lines).

5.6. The Effects of Mass Scatters on Reconstruction Performances

The key information required to achieve these gains are the halo masses, which, however, are not directly observable in reality. Nevertheless, in principle, they are achievable via relations between observables (e.g., galaxy luminosity, Vale & Ostriker 2004; X-ray luminosity, galaxy richness, weak lensing shear, velocity dispersion, or the thermal Sunyaev–Zeldovich (SZ) effect, Allen et al. 2011, etc.) and halo masses. A better understanding of these relations should be essential and critical to minimize the inaccuracies of inferred masses and maximize the gains of biased tracer reconstruction.

The SZ effect and X-ray properties yield very tight correlations with the halo masses with a scatter of about 0.1 dex, whereas these observables can only be used to probe the masses of massive systems (e.g., galaxy clusters), while the optical estimators can be used to infer relatively lower halo masses with a scatter of about 0.18 dex (see Henden et al. 2019; Bradshaw et al. 2020 and references therein). Alternatively, galaxy grouping methods (Yang et al. 2005), which group galaxies residing in the same halo, provide a direct way of studying the galaxy–halo connection to estimate halo masses. These techniques can reliably estimate halo mass over a wide range of masses, even for a poor system including isolated galaxy in a small halo. By using a halo-based group finder, Yang et al. (2007) demonstrated that more than 90% of true halos can be successfully identified in mock catalogs, and

the scatter between estimated and true halo masses is about 0.3 dex. This group finder was also improved for applications in low-redshift galaxy surveys, showing a lower mass scatter of about 0.2 dex (Lim et al. 2017). Recently, Wang et al. (2020b) developed a machine-learning-based group finder, which was applied to high-redshift incomplete spectroscopic data, estimating halo masses with scatters smaller than 0.25 dex.

In order to mimic the observational uncertainties on the estimated halo masses, we artificially add two constant mass scatters (i.e., 0.1 dex and 0.3 dex for optimistic and pessimistic scenarios respectively) to our halo samples, and then repeat the analyses for all halo number density cases shown in Figures 8 and 9. For the two mass-dependent weightings, we find that the effects of a scatter of 0.1 dex on the reconstruction performances are basically negligible and a scatter of 0.3 dex can result in noticeable degradation of the performances. For the scatter of 0.3 dex, the benefits from mass weighting are completely lost, while the benefits from optimal weighting still remain substantial in the higher number density cases, which again demonstrates the interests of optimal weighting in the applications of ongoing and future galaxy surveys. Here, note that the optimal weighting adopts the same M_0 (see Section 3.2) as before. We also see that the biases are hardly affected by the mass uncertainties (see Figure 8), which indicates that the performance degradations are purely caused by the boosts of shot noises induced by the mass scatters (see Seljak et al. 2009).

6. Summary and Conclusion

In realistic scenarios, the efficiency of reconstruction is prone to be hampered by various complications (e.g., galaxy bias, shot noise, RSDs, survey boundary, etc.). In principle, the problem could be alleviated by considering extra information (e.g., the radial velocities for reducing boundary effects, Zhu et al. 2020; and the local environments for sharpening the BAO peak, Achitouv & Blake 2015; Blake et al. 2019 etc.).

The biased tracer reconstruction method (Birkin et al. 2019) adopted in this work uses mass conservation to relate the density fields in Lagrangian and Eulerian spaces. When the method is applied to a realistic tracer field, the reconstruction performance will be degraded by the loss of information on the underlying total matter. The basic idea for improving the reconstruction performance should be to apply reconstruction to a tracer field with more matter distribution information, which can be achieved from some weighting schemes. Our work corroborated this idea by using halo-mass-dependent weighting schemes because halo masses are natural information to enforce the local mass and momentum conservation.

For our study, we performed careful analyses on three halo samples with various number densities corresponding to different targets in various galaxy surveys. The halo fields are created under different halo-weighting schemes, i.e., uniform, mass, and optimal weightings (because the mass and optimal weightings are both dependent on the information of halo masses, we collectively call them mass-dependent weightings). We summarize our work and discuss it as follows:

1. Before our performance tests, we first investigated the clustering biases and stochasticity properties of different weighted halo fields, demonstrating how mass information suppresses the shot-noise component and tightens the cross-correlation between the halo field and the dark matter. We argue that the information for suppressing the halo fields' shot noise should originate from the environmental dependence of halo mass (see Section 5.2 and Appendix A). For our halo samples with $\bar{n}_h = 2.77 \times 10^{-2}(h^{-1}\text{Mpc})^{-3}$ and $\bar{n}_h = 2.77 \times 10^{-3}(h^{-1}\text{Mpc})^{-3}$, we find that halo-clustering biases are more scale independent under mass-dependent weightings, compared with the uniform weighting case. This finding is interesting because bias being more scale independent will help to adequately maximize the performance of the bias tracer reconstruction by only performing linear debiasing in the process of reconstruction and to minimize the systematic shifts in the BAO position relative to the dark matter. For this reason, we only considered the clustering bias up to linear order in our performance tests.
2. For performance tests, we then performed the biased tracer reconstruction method on the different weighted halo fields. We investigated and discussed in detail how linear debiasing improves the reconstruction performance. After linear debiasing, we both visually and quantitatively compared the qualities of the reconstructed initial matter fields, which are obtained under different weighting schemes. We showed that the reconstruction performance can be substantially enhanced when the reconstruction method is applied to the mass and optimally weighted halo fields, compared to the uniform weighting scenario. As the halo number density increases, the gains achieved from mass information can be more significant. In particular, we showed a

compelling result that the reconstruction performance from the optimally weighted halo field with $\bar{n}_h = 2.77 \times 10^{-2}(h^{-1}\text{Mpc})^{-3}$ can almost be comparable to that from the dark matter field, which implies that the initial information can be greatly recovered from the dense data (e.g., the DESI Bright Galaxy Survey (BGS) sample; DESI Collaboration et al. 2016). Nevertheless, the reconstruction performance also depends on the accuracy of the inferred halo masses in the observations, thus the impacts of realistic halo mass scatters on our results were also investigated in this work. A better understanding of the relation between observables and underlying halos in the galaxy-halo connection and observable-mass relation studies (see Wechsler & Tinker 2018; Bradshaw et al. 2020 and references therein) should be essential and critical to decreasing the mass scatters and consequently to maximizing the gains of biased tracer reconstruction.

Based on our findings, we can safely conclude that the halo masses are critical information for highly improving the performance of biased tracer reconstruction. We note that this conclusion should also hold true for other reconstruction methods, as the effectiveness of almost all reconstruction methods relies on the information of the total matter distribution.

Our work is particularly relevant to the context of recovering the BAO signal, the characteristic scale of which is renowned as a standard ruler to measure the expansion history of the universe and are therefore of critical importance for future BAO measurements. We leave the quantification of the gains of BAO reconstruction benefiting from the halo mass information to future investigations. Moreover, because various other cosmological measurements (e.g., RSDs, Hamaus et al. 2012; primordial non-Gaussianities, Darwish et al. 2020; and even neutrino signatures, Liu et al. 2020) are also usually limited by strong data nonlinearities, we expect our work to be quite useful for significantly improving the scientific returns of current and future galaxy surveys.

We would like to thank Pengjie Zhang, Haojie Xu, Zhejie Ding, Ji Yao, and Hong-Ming Zhu for helpful discussions and several useful suggestions, and the anonymous referee for the helpful comments. Y.L. would also like to thank Xiaohu Yang and Qingyang Li for useful communications about galaxy group finder. This work was supported by the National Key Basic Research and Development Program of China (No. 2018YFA0404504), the National Science Foundation of China (grant Nos. 11773048, 11621303, 11890691), and the "111" Project of the Ministry of Education under grant No. B20019. B.L. is supported by the European Research Council (ERC) through a Starting Grant, ERC-StG-716532-PUNCA, and by the UK Science and Technology Facility Council through grants ST/T000244/1 and ST/P000541/1.

Appendix A Shuffled Shot Noises

In this appendix, we present shuffled shot noises under mass-dependent weightings, which are mentioned in Section 5.2. The results are shown in Figure A1.

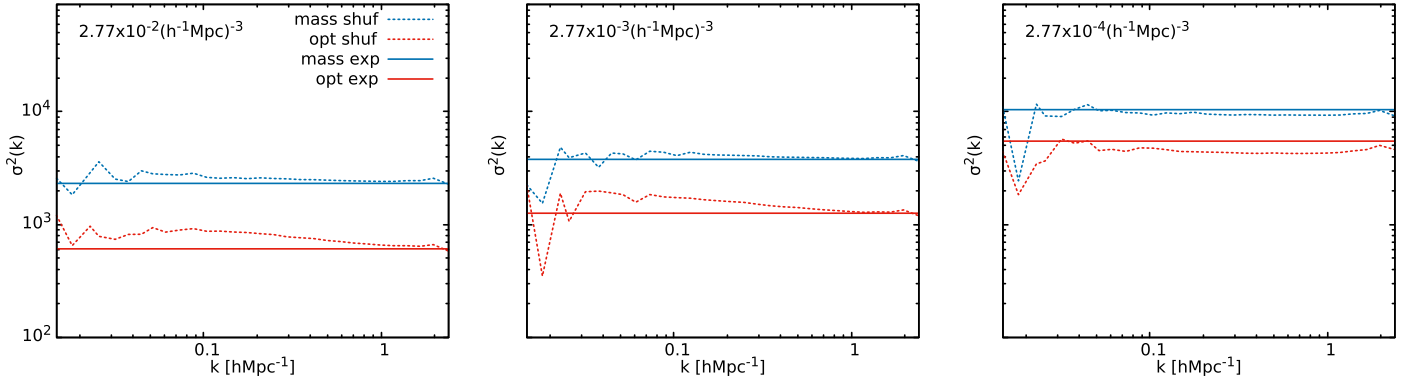


Figure A1. Shuffled shot noises under mass and optimal weighting schemes (dotted lines). For comparison, the expected shot noises under mass and optimal weighting schemes (solid lines) are also presented.

Appendix B

Different Weighted Halo Fields before and after Reconstruction

In this appendix, we present the two-dimensional slices of different weighted halo fields with number densities of 2.77×10^{-3} and $2.77 \times 10^{-4} (h^{-1} \text{Mpc})^{-3}$ before and after reconstruction. The results are shown in Figure B2 and Figure B3.

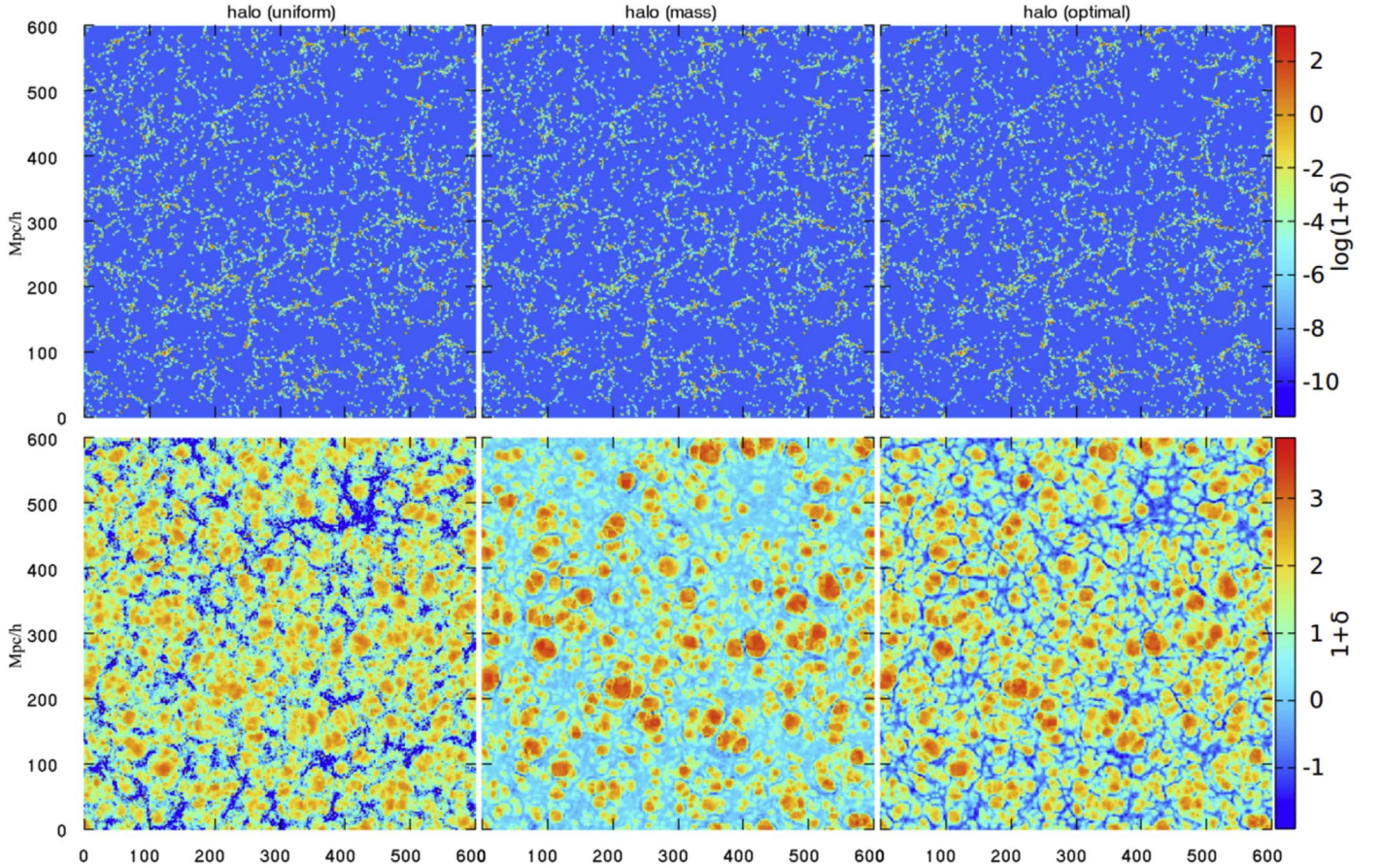


Figure B2. The two-dimensional slices of different weighted halo fields before (upper panels) and after (bottom panels) reconstruction. The halo fields are produced by the halo sample with number density of $2.77 \times 10^{-3} (h^{-1} \text{Mpc})^{-3}$. Each slice shows the same region (with the same projection depth) as the corresponding slice in Figures 1 and 6

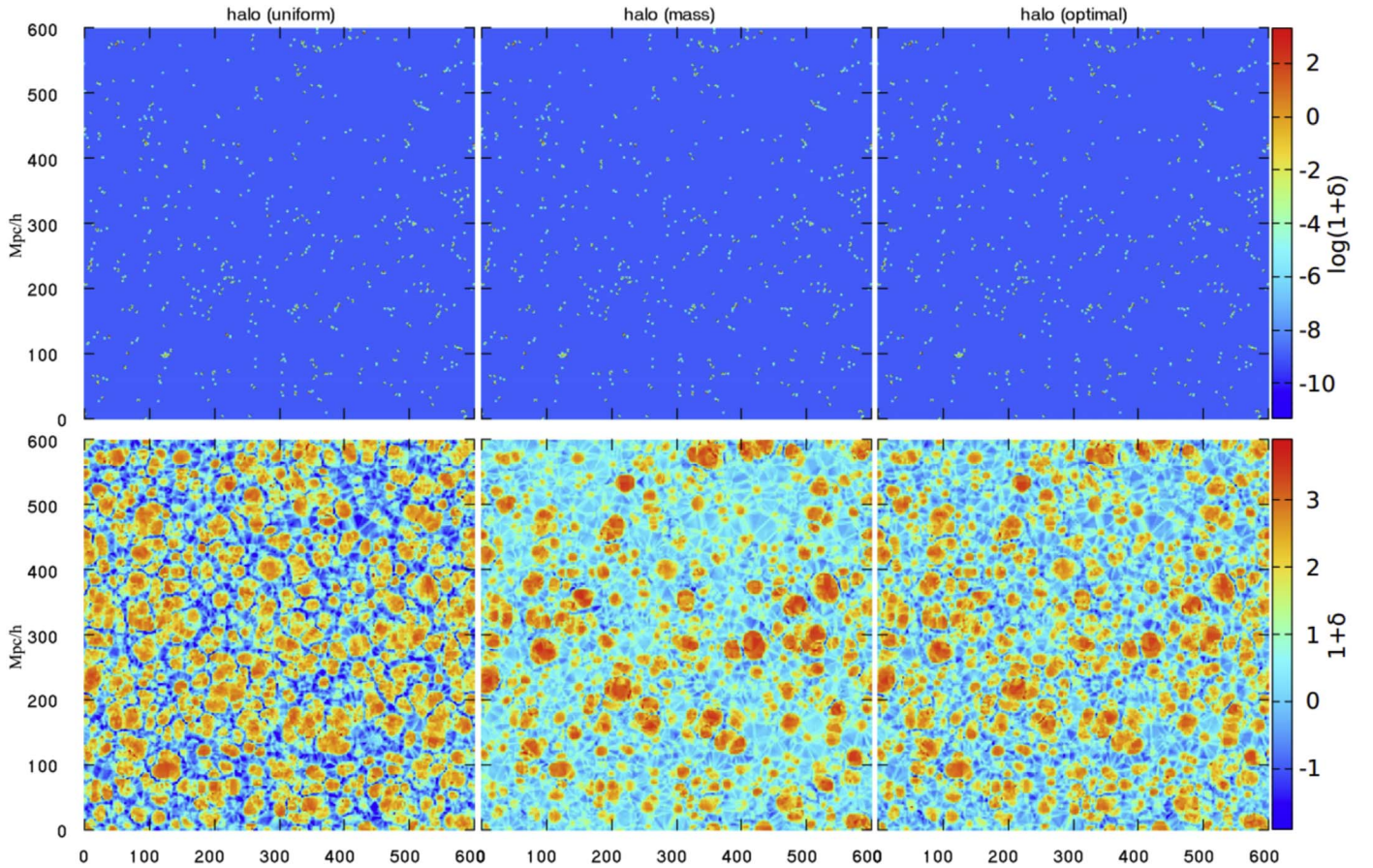


Figure B3. Same as Figure B2, but for the halo sample with number density of $2.77 \times 10^{-4}(h^{-1}\text{Mpc})^{-3}$.

ORCID iDs

Yu Liu <https://orcid.org/0000-0002-9734-906X>
 Yu Yu <https://orcid.org/0000-0002-9359-7170>
 Baojiu Li <https://orcid.org/0000-0002-1098-9188>

References

- Achitouv, I., & Blake, C. 2015, *PhRvD*, **92**, 083523
 Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, **49**, 409
 Anderson, L., Aubourg, É., Bailey, S., et al. 2014, *MNRAS*, **441**, 24
 Angulo, R. E., Baugh, C. M., Frenk, C. S., & Lacey, C. G. 2008, *MNRAS*, **383**, 755
 Baldauf, T., Seljak, U., Smith, R. E., Hamaus, N., & Desjacques, V. 2013, *PhRvD*, **88**, 083507
 Beutler, F., Blake, C., Koda, J., et al. 2016, *MNRAS*, **455**, 3230
 Beutler, F., Seo, H.-J., Ross, A. J., et al. 2017, *MNRAS*, **464**, 3409
 Birkin, J., Li, B., Cautun, M., & Shi, Y. 2019, *MNRAS*, **483**, 5267
 Blake, C., Achitouv, I., Burden, A., & Rasera, Y. 2019, *MNRAS*, **482**, 578
 Bonoli, S., & Pen, U. L. 2009, *MNRAS*, **396**, 1610
 Bradshaw, C., Leauthaud, A., Hearin, A., Huang, S., & Behroozi, P. 2020, *MNRAS*, **493**, 337
 Brenier, Y., Frisch, U., Hénon, M., et al. 2003, *MNRAS*, **346**, 501
 Burden, A., Percival, W. J., & Howlett, C. 2015, *MNRAS*, **453**, 456
 Burden, A., Percival, W. J., Manera, M., et al. 2014, *MNRAS*, **445**, 3152
 Cai, Y.-C., Bernstein, G., & Sheth, R. K. 2011, *MNRAS*, **412**, 995
 Cautun, M. C., & van de Weygaert, R. 2011a, The DTFE Public Software: the Delaunay Tessellation Field Estimator Code v1.1.1, Astrophysics Source Code Library, ascl:**1105.003**
 Cautun, M. C., & van de Weygaert, R. 2011b, arXiv:**1105.0370**
 Chan, K. C., Scoccimarro, R., & Sheth, R. K. 2012, *PhRvD*, **85**, 083509
 Chen, S.-F., Vlah, Z., & White, M. 2019, *JCAP*, **2019**, 017
 Darwish, O., Foreman, S., Abidi, M. M., et al. 2020, arXiv:**2007.08472**
 Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
 de Jong, R. S., Barden, S. C., Bellido-Tirado, O., et al. 2016, *Proc. SPIE*, **9908**, 990810
 DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:**1611.00036**
 Ding, Z., Seo, H.-J., Vlah, Z., et al. 2018, *MNRAS*, **479**, 1021
 Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:**1412.4872**
 Doré, O., Hirata, C., Wang, Y., et al. 2018, arXiv:**1804.03628**
 Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. N. 2007a, *ApJ*, **664**, 675
 Eisenstein, D. J., Seo, H.-J., & White, M. 2007b, *ApJ*, **664**, 660
 Feng, Y., Seljak, U., & Zaldarriaga, M. 2018, *JCAP*, **2018**, 043
 Frisch, U., Matarrese, S., Mohayaee, R., & Sobolevski, A. 2002, *Natur*, **417**, 260
 Ginzburg, D., Desjacques, V., & Chan, K. C. 2017, *PhRvD*, **96**, 083528
 Haas, M. R., Schaye, J., & Jeesson-Daniel, A. 2012, *MNRAS*, **419**, 2133
 Hada, R., & Eisenstein, D. J. 2018, *MNRAS*, **478**, 1866
 Hada, R., & Eisenstein, D. J. 2019, *MNRAS*, **482**, 5685
 Hamaus, N., Seljak, U., & Desjacques, V. 2011, *PhRvD*, **84**, 083509
 Hamaus, N., Seljak, U., & Desjacques, V. 2012, *PhRvD*, **86**, 103513
 Hamaus, N., Seljak, U., Desjacques, V., Smith, R. E., & Baldauf, T. 2010, *PhRvD*, **82**, 043515
 Harnois-Déraps, J., Pen, U.-L., Iliev, I. T., et al. 2013a, *MNRAS*, **436**, 540
 Harnois-Déraps, J., Yu, H.-R., Zhang, T.-J., & Pen, U.-L. 2013b, *MNRAS*, **436**, 759
 Henden, N. A., Puchwein, E., & Sijacki, D. 2019, *MNRAS*, **489**, 2439
 Hikage, C., Koyama, K., & Heavens, A. 2017, *PhRvD*, **96**, 043513
 Hikage, C., Koyama, K., & Takahashi, R. 2020, *PhRvD*, **101**, 043510
 Hinton, S. R., Kazin, E., Davis, T. M., et al. 2017, *MNRAS*, **464**, 4807
 Jasche, J., & Lavaux, G. 2019, *A&A*, **625**, A64
 Jasche, J., & Wandelt, B. D. 2013, *MNRAS*, **432**, 894
 Jennings, E., Baugh, C. M., & Hatt, D. 2015, *MNRAS*, **446**, 793
 Kazin, E. A., Koda, J., Blake, C., et al. 2014, *MNRAS*, **441**, 3524
 Kitaura, F. S. 2013, *MNRAS*, **429**, L84
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:**1110.3193**
 Levi, M., Bebek, C., Beers, T., et al. 2013, arXiv:**1308.0847**
 Lévy, B., Mohayaee, R., & von Hausegger, S. 2020, arXiv:**2012.09074**
 Li, B., Barreira, A., Baugh, C. M., et al. 2013, *JCAP*, **2013**, 012

- Li, B., Zhao, G.-B., Teyssier, R., & Koyama, K. 2012, *JCAP*, 2012, 051
- Lim, S. H., Mo, H. J., Lu, Y., Wang, H., & Yang, X. 2017, *MNRAS*, 470, 2982
- Liu, Y., Yu, Y., Yu, H.-R., & Zhang, P. 2020, *PhRvD*, 101, 063515
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
- Mao, T.-X., Wang, J., Li, B., et al. 2020, arXiv:2002.10218
- Mehta, K. T., Seo, H.-J., Eckel, J., et al. 2011, *ApJ*, 734, 94
- Modi, C., Feng, Y., & Seljak, U. 2018, *JCAP*, 2018, 028
- Modi, C., White, M., Slosar, A., & Castorina, E. 2019, *JCAP*, 2019, 023
- Mohayaee, R., Frisch, U., Matarrese, S., & Sobolevskii, A. 2003, *A&A*, 406, 393
- Mohayaee, R., Mathis, H., Colombi, S., & Silk, J. 2006, *MNRAS*, 365, 939
- Mohayaee, R., & Sobolevskii, A. 2008, *PhyD*, 237, 2145
- Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2009, *ApJL*, 698, L90
- Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2011, *ApJ*, 731, 116
- Noh, Y., White, M., & Padmanabhan, N. 2009, *PhRvD*, 80, 123501
- Obuljen, A., Villaescusa-Navarro, F., Castorina, E., & Viel, M. 2017, *JCAP*, 2017, 012
- Padmanabhan, N., White, M., & Cohn, J. D. 2009, *PhRvD*, 79, 063523
- Padmanabhan, N., Xu, X., Eisenstein, D. J., et al. 2012, *MNRAS*, 427, 2132
- Paech, K., Hamaus, N., Hoyle, B., et al. 2017, *MNRAS*, 470, 2566
- Pan, Q., Pen, U.-L., Inman, D., & Yu, H.-R. 2017, *MNRAS*, 469, 1968
- Park, C., & Kim, Y.-R. 2010, *ApJL*, 715, L185
- Ross, A. J., Samushia, L., Howlett, C., et al. 2015, *MNRAS*, 449, 835
- Sarpa, E., Schimd, C., Branchini, E., & Matarrese, S. 2019, *MNRAS*, 484, 3818
- Sarpa, E., Veropalumbo, A., Schimd, C., Branchini, E., & Matarrese, S. 2021, *MNRAS*, 503, 540
- Schaap, W. E., & van de Weygaert, R. 2000, *A&A*, 363, L29
- Schmittfull, M., Baldauf, T., & Zaldarriaga, M. 2017, *PhRvD*, 96, 023505
- Schmittfull, M., Feng, Y., Beutler, F., Sherwin, B., & Chu, M. Y. 2015, *PhRvD*, 92, 123522
- Schmittfull, M., Simonović, M., Assassi, V., & Zaldarriaga, M. 2019, *PhRvD*, 100, 043514
- Seljak, U., Aslanyan, G., Feng, Y., & Modi, C. 2017, *JCAP*, 2017, 009
- Seljak, U., Hamaus, N., & Desjacques, V. 2009, *PhRvL*, 103, 091303
- Seljak, U., & Warren, M. S. 2004, *MNRAS*, 355, 129
- Seo, H.-J., Beutler, F., Ross, A. J., & Saito, S. 2016, *MNRAS*, 460, 2453
- Seo, H.-J., Eckel, J., Eisenstein, D. J., et al. 2010, *ApJ*, 720, 1650
- Seo, H.-J., Siegel, E. R., Eisenstein, D. J., & White, M. 2008, *ApJ*, 686, 13
- Sheth, R. K., & Tormen, G. 1999, *MNRAS*, 308, 119
- Shi, Y., Cautun, M., & Li, B. 2018, *PhRvD*, 97, 023505
- Smith, R. E., & Marian, L. 2016, *MNRAS*, 457, 2968
- Smith, R. E., Scoccimarro, R., & Sheth, R. K. 2008, *PhRvD*, 77, 043525
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, *PASJ*, 66, R1
- Tassev, S., & Zaldarriaga, M. 2012, *JCAP*, 2012, 006
- Teyssier, R. 2002, *A&A*, 385, 337
- Vale, A., & Ostriker, J. P. 2004, *MNRAS*, 353, 189
- Wang, H., Mo, H. J., Yang, X., & van den Bosch, F. C. 2013, *ApJ*, 772, 63
- Wang, K., Mo, H. J., Li, C., Meng, J., & Chen, Y. 2020b, *MNRAS*, 499, 89
- Wang, X., & Pen, U.-L. 2019, *ApJ*, 870, 116
- Wang, X., Yu, H.-R., Zhu, H.-M., et al. 2017, *ApJL*, 841, L29
- Wang, Y., Li, B., & Cautun, M. 2020a, *MNRAS*, 497, 3451
- Wechsler, R. H., & Tinker, J. L. 2018, *ARA&A*, 56, 435
- Weinberg, D. H. 1992, *MNRAS*, 254, 315
- White, M. 2015, *MNRAS*, 450, 3822
- Xu, H., Zheng, Z., Guo, H., et al. 2018, *MNRAS*, 481, 5470
- Xu, X., Cuesta, A. J., Padmanabhan, N., Eisenstein, D. J., & McBride, C. K. 2013, *MNRAS*, 431, 2834
- Xu, X., Padmanabhan, N., Eisenstein, D. J., Mehta, K. T., & Cuesta, A. J. 2012, *MNRAS*, 427, 2146
- Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2007, *ApJ*, 671, 153
- Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005, *MNRAS*, 356, 1293
- Yu, H.-R., Pen, U.-L., & Zhu, H.-M. 2017a, *PhRvD*, 95, 043501
- Yu, Y., & Zhu, H.-M. 2019, *ApJ*, 887, 265
- Yu, Y., Zhu, H.-M., & Pen, U.-L. 2017b, *ApJ*, 847, 110
- Zhang, P. 2008, arXiv:0802.2416
- Zhao, C., Kitaura, F.-S., Chuang, C.-H., et al. 2015, *MNRAS*, 451, 4266
- Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, *ApJ*, 633, 791
- Zhu, H.-M., White, M., Ferraro, S., & Schaan, E. 2020, *MNRAS*, 494, 4244
- Zhu, H.-M., Yu, Y., & Pen, U.-L. 2018, *PhRvD*, 97, 043502
- Zhu, H.-M., Yu, Y., Pen, U.-L., Chen, X., & Yu, H.-R. 2017, *PhRvD*, 96, 123502