MAPPING THE REAL-SPACE DISTRIBUTIONS OF GALAXIES IN SDSS DR7. I. TWO-POINT CORRELATION FUNCTIONS

Feng Shi1,2, Xiaohu Yang3,4, Huiyuan Wang5, Youcai Zhang1, H. J. Mo6,7, Frank C. van den Bosch8, Shuie Li1, Chengze Liu3, Yi Li1, Dylan Tweed3, and Lei Yang3

1 Shanghai Astronomical Observatory, Nandan Road 80, Shanghai 200030, China; sfeng@shao.ac.cn
2 University of Chinese Academy of Sciences, 19A, Yuquan Road, Beijing, China
3 Center for Astronomy and Astrophysics, Shanghai Jiao Tong University, Shanghai 200240, China; xyang@sjtu.edu.cn
4 IFSA Collaborative Innovation Center, Shanghai Jiao Tong University, Shanghai 200240, China
5 Key Laboratory for Research in Galaxies and Cosmology, University of Science and Technology of China, Hefei, Anhui 230026, China
6 Department of Astronomy, University of Massachusetts, Amherst, MA 01003-9305, USA
7 Physics Department and Center for Astrophysics, Tsinghua University, Beijing 10084, China
8 Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06520-8101, USA

Received 2016 July 28; revised 2016 October 12; accepted 2016 October 30; published 2016 December 20

ABSTRACT

Using a method to correct redshift-space distortion (RSD) for individual galaxies, we mapped the real-space distributions of galaxies in the Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7). We use an ensemble of mock catalogs to demonstrate the reliability of our method. Here, in the first paper in a series, we focus mainly on the two-point correlation function (2PCF) of galaxies. Overall the 2PCF measured in the reconstructed real space for galaxies brighter than \(0.1M_\odot - 5\log h = -19.0\) agrees with the direct measurement to an accuracy better than the measurement error due to cosmic variance, if the reconstruction uses the correct cosmology. Applying the method to the SDSS DR7, we construct a real-space version of the main galaxy catalog, which contains 396,068 galaxies in the North Galactic Cap with redshifts in the range \(0.01 \leq z \leq 0.12\). The Sloan Great Wall, the largest known structure in the nearby universe, is not as dominant an overdense structure as it appears to be in redshift space. We measure the 2PCFs in reconstructed real space for galaxies of different luminosities and colors. All of them show clear deviations from single power-law forms, and reveal clear transitions from one-halo to two-halo terms. A comparison with the corresponding 2PCFs in redshift space nicely demonstrates how RSDs boost the clustering power on large scales (by about 40%–50% at scales \(\sim 10\ h^{-1}\ \text{Mpc}\)) and suppress it on small scales (by about 70%–80% on a scale of \(0.3\ h^{-1}\ \text{Mpc}\)).

Key words: dark matter – galaxies: halos – large-scale structure of universe – methods: statistical

1. INTRODUCTION

One of the important properties of the population of galaxies is their distribution in space (e.g., Peebles 1980; Mo et al. 2010). This distribution can be used to infer the large-scale mass distribution in the universe, thereby constraining cosmological models (e.g., Fisher et al. 1994; Peacock et al. 2001; Hawkins et al. 2003; Yang et al. 2004; Tinker et al. 2005). Furthermore, the spatial clustering of galaxies is also one of the key pieces of observational data needed for us to establish the relation between galaxies and dark matter (halos) statistically (e.g., Jing et al. 1998; Peacock & Smith 2000; Yang et al. 2003, 2012) and to understand how galaxies form and evolve in the cosmic density field.

One of the main goals of large redshift surveys of galaxies, such as the 2 degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000) is, therefore, to provide a database for studying the three-dimensional distribution of galaxies as accurately as possible. However, a key problem of this endeavor is that redshifts of galaxies are not exact measures of distances because of the peculiar motions of galaxies. The spatial distribution and clustering of galaxies observed in redshift space are thus distorted with respect to the real-space distribution and clustering (e.g., Sargent & Turner 1977; Davis & Peebles 1983; Kaiser 1987; Regos & Geller 1991; Hamilton 1992; van de Weygaert & van Kampen 1993). Take the two-point correlation function (2PCF) of galaxies as an example. The 2PCF in two-dimensional space, with one dimension along the line of sight and the other in the perpendicular direction, appears elongated on small scales and squashed on large scales along the line-of-sight direction, in contrast to the isotropic pattern expected from a statistically homogeneous and isotropic distribution in real space. Such anisotropies are clearly produced by redshift distortions and need to be corrected in order to get the true distribution of galaxies in space. Theoretically, models of the pairwise peculiar velocities of galaxies have been used to model the effects of redshift distortions on the measured 2PCF in redshift space (e.g., Davis & Peebles 1983; Fisher et al. 1994; Jing et al. 1998). Alternatively, one simply measures the projected 2PCF and uses it to infer the three-dimensional 2PCF (e.g., Jing et al. 1998; Li et al. 2006; Zehavi et al. 2011).

In the gravitational instability scenario of structure formation, the redshift distortion is not just a contamination one has to correct in order to get the real clustering of galaxies; in fact it contains useful information about cosmology as well as the mass distribution in the universe. The in-fall motions of galaxies, which produce the squashing in the 2D redshift-space 2PCF (the Kaiser effect, Kaiser 1987), are linearly proportional to the amplitudes of the mass density fluctuations on large scales. In this case, one can compute the quadrupole-to-monopole ratio of the 2D 2PCF to get \(\beta \equiv f \left(\Omega_m\right)/b\), where \(\Omega_m\) is the density parameter of mass, and \(b\) is the effective linear bias of the galaxies in question (e.g., Guzzo et al. 2008; Samushia et al. 2012; Dawson et al. 2016, and references therein). When the measurement is combined with results from
weak gravitational lensing, it can also be used as a sensitive probe of (modified) theories of gravitational cosmological scales (Zhang et al. 2007; Reyes et al. 2010; Blake et al. 2016). On smaller scales, the modeling of redshift-space distortion (RSD) (the Finger of God (FOG) effect, Jackson 1972; Tully & Fisher 1978) is complicated by the nonlinear mapping between real space and redshift space. Great efforts have been made not only to understand its impacts on galaxy clustering (e.g., Zhang et al. 2013, 2015; Zheng et al. 2013, 2015a, 2015b), but also to extract useful cosmological information (Mo et al. 1993; Jing et al. 1998; Yang et al. 2004; Li et al. 2012).

The approaches adopted earlier to deal with redshift distortions in galaxy clustering have been hampered by the fact that the large-scale Kaiser effect and the small-scale FOG effect are intertwined, and models based on a simple pairwise peculiar velocity distribution can only serve as an approximation. The situation is complicated even more by the fact that the effect of bias in galaxy distribution may be nonlinear and its form is not known a priori. Models based on the projected correlation function have their own problem, because the projection mixes clustering on different scales so that the conversion from the projected function to the three-dimensional function can be uncertain. Thus, in order to make full use of galaxy redshift surveys to study the large-scale structure of the universe, a change of tactics is needed.

One possible way is first to make corrections of redshift distortions for individual galaxies, and then to use the “pseudo” real-space distribution of galaxies to derive statistical measures of galaxy clustering in real space. As mentioned above, redshift distortions are of two different kinds. One is the Kaiser effect produced by the coherent flow due to the gravitational action of a large-scale structure (Kaiser 1987); the other is the FOG effect generated by the random motions of galaxies within virialized halos on small scales. To deal with the FOG effect, Tegmark et al. (2002) used a friends-of-friends method to link galaxies and suppressed the overdensity of the pairs along the line of sight by a factor of 10. They applied this FOG suppression to the 2dFGRS (Tegmark et al. 2002) and SDSS (Tegmark et al. 2004) in their estimates of the power spectra of galaxy distribution. In a paper aimed at reconstructing the cosmic web from 2dFGRS, Erdogdu et al. (2004) attempted to deal with the FOG effect by compressing 25 fingers seen in redshift space using groups identified by Eke et al. (2004). For the Kaiser effect, Yahil et al. (1991) used a bias model to get the density field from the galaxy distribution and iteratively corrected the infall motions of galaxies. A number of approaches along the same lines have been made to recover/correct the infall motions on the basis of galaxy distribution (e.g., Monaco & Efstathiou 1999; Lavaux et al. 2008; Wang et al. 2009, 2012; Branchini et al. 2012; Kitaura et al. 2012, 2016; Ata et al. 2016; Granett et al. 2015; Jasche et al. 2015). In particular, Wang et al. (2009, 2012) used galaxy groups as proxies for dark matter halos to reconstruct the density field, which in turn was used to obtain the velocity field.

So far there has been no real attempt to correct for both the large-scale velocities and small-scale random motions of galaxies in a systematic way. The main purpose of the present paper is to carry out such an investigation, using galaxies observed in the SDSS DR7, which is still among the best redshift surveys available. Based on this galaxy catalog, Yang et al. (2007, hereafter Y07) have constructed a catalog of galaxy groups using an adaptive halo-based group finder (see also Yang et al. 2005). Detailed tests with mock galaxy catalogs have shown that the group finder is very successful in associating galaxies according to their common dark matter halos. In particular, the group finder performs reliably not only for rich systems, but also for poor systems, including isolated central galaxies in low-mass halos. The reliable memberships of galaxies in groups provide a unique opportunity to correct for the FOG effects for individual galaxy systems. In addition, as shown in Wang et al. (2012, hereafter W12), the group catalog can also be used to reconstruct the mass density, tidal, and velocity (MTV) fields in the SDSS DR7 volume, using the halo-domain method developed in Wang et al. (2009). Since the relation between halo and mass distributions is better understood than that between galaxies and mass, the mass and velocity fields constructed are much more accurate than those constructed directly from the galaxy distribution. The redshift distortions on large scales can, therefore, also be modeled accurately for individual galaxies. With all these, we can obtain a catalog of galaxies in quasi-real space. We can then not only examine various types of redshift distortions in detail, but also measure the real-space clustering of galaxies.

This paper is organized as follows. In Section 2 we present the galaxy and group catalogs used in this paper. Section 3 introduces the methods to correct for the redshift distortions and to characterize the galaxy clustering. In Section 4 we use mock galaxy catalogs to test the reliability of our correction model. The application to the SDSS DR7 and the results are presented in Section 5. Finally, we summarize our main findings in Section 6. Throughout this paper, unless stated otherwise, physical quantities are quoted using the WMAP9 cosmological parameters (Hinshaw et al. 2013): \( \Omega_m = 0.282, \Omega_{\Lambda} = 0.718, \Omega_b = 0.046, n_s = 0.965, h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.697, \) and \( \sigma_8 = 0.817. \)

2. THE SDSS GALAXY AND GROUP CATALOGS

The galaxy sample used in this paper is constructed from the New York University Value-Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005), which is based on SDSS DR7 (Abazajian et al. 2009), but with an independent set of significantly improved reductions over the original pipeline. In addition, as galaxy groups play a key role in our approach to correct for the redshift distortions, we make use of the group catalog constructed in (Yang et al. 2012) for SDSS DR7. This group catalog is based on all galaxies in the Main Galaxy Sample with extinction-corrected apparent magnitude brighter than \( r = 17.72, \) with redshifts in the range \( 0.01 \leq z \leq 0.20, \) and with a redshift completeness \( C_z \geq 0.7. \) The catalog contains a total of 639,359 galaxies with a sky coverage of 7748 deg\(^2\). Moreover, the galaxy catalog mainly covers two sky regions: a larger contiguous region in the Northern Galactic Cap (NGC) and a smaller three-stripe region in the Southern Galactic Cap. The former contains 584,473 galaxies with a sky coverage of 7748 deg\(^2\). The catalog contains a total of 639,359 galaxies with a sky coverage of 7748 deg\(^2\). Moreover, the catalog mainly covers two sky regions: a larger contiguous region in the Northern Galactic Cap (NGC) and a smaller three-stripe region in the Southern Galactic Cap. The former contains 584,473 galaxies with a sky coverage of 7047 deg\(^2\).

Based on this SDSS DR7 galaxy catalog, Yang et al. (2012) used the adaptive halo-based group finder developed by Yang et al. (2005) to select galaxy groups. This group finder has been applied to the SDSS DR4 in Y07. Following Y07, the masses of the associated dark matter halos are estimated based on the ranking of the total characteristic luminosities of groups or the total characteristic stellar masses using group member galaxies more luminous than \( 0.1M_\odot - 5 \log h = -19.5. \) The two estimates of halo masses agree very well with each other, and we adopt the halo masses based on the characteristic luminosity.
ranking in this paper. In addition, we have updated group membership as well as halo masses according to WMAP9 cosmology.

Using this group catalog, W12 reconstructed the velocity field, which we use in this paper to correct for the RSDs. The method of W12 depends explicitly on the density field as represented by dark matter halos above a given mass threshold, \( M_{\text{th}} \). We adopt \( M_{\text{th}} = 10^{12.5} h^{-1} M_{\odot} \) and so, to be complete, restrict our sample to the nearby volume covering the redshift range \( 0.01 \leq z \leq 0.12 \). In addition, since the W12 reconstruction method can be significantly impacted by survey boundaries, we focus only on the more contiguous NGC region.

Applying all these selection criteria to the galaxy and group catalogs leaves us with a set of 286,043 groups, hosting a total of 396,068 galaxies in the NGC region with redshifts in the range \( 0.01 \leq z \leq 0.12 \). Finally, using this sample we construct both flux-limited and volume-limited subsamples for galaxies in the following six bins of absolute \( r \)-band magnitude: \( 0.1 M_{\odot} - 5 \log h = [-23.0, -22.0], \ [-22.0, -21.0], \ [-21.0, -20.0], \ [-20.0, -19.0], \ [-19.0, -18.0], \ [-18.0, -17.0] \). The corresponding redshift ranges, numbers of galaxies, and averaged magnitude are indicated in Table 1. These luminosity samples are further divided into blue and red subsamples, as detailed in Section 5.2. Note that there is no difference in the redshift limit between the flux-limited and volume-limited samples for the two brightest ones, because all the galaxies with such luminosities can be observed to \( z = 0.12 \). For a fainter sample, even the brightest galaxies in the luminosity bin can be observed only to redshift \( z < 0.12 \). In most cases we show only results obtained from the flux-limited samples, because the results obtained from the volume-limited samples are very similar. Note also that in the reconstructed real space, which we will introduce later, the number of galaxies in a sample will change very slightly.

### 3. METHODOLOGY AND BASIC ANALYSIS

We now turn to our main goal: correcting the SDSS redshifts for RSDs induced by peculiar velocities, thus allowing for a direct measurement of the two-point correlation functions of galaxies in real space. Before delving into details, we first introduce some concepts regarding RSDs and our approach to correcting for them.

#### 3.1. Redshift-space Distortions

In the absence of peculiar velocities, the redshift of a galaxy, \( z \), is directly related to its comoving distance, \( r \). For a flat universe, this relation is given by

\[
\frac{dz}{dz} = \frac{1}{H_0} \int_0^z \sqrt{\Omega_\Lambda + \Omega_m (1 + z)^3} dz, \tag{1}
\]

with \( H_0 \) the Hubble constant. In reality, though, the observed redshift of a galaxy, \( z_{\text{obs}} \), consists of a cosmological contribution, \( z_{\text{cos}} \), arising from the Hubble expansion plus a Doppler contribution, \( z_{\text{pec}} \), due to the galaxy’s peculiar velocity along the line of sight, \( \nu_{\text{pec}} \). In the non-relativistic case we have that

\[
z_{\text{obs}} = z_{\text{cos}} + z_{\text{pec}} = z_{\text{cos}} + \frac{v_{\text{pec}}}{c} (1 + z_{\text{cos}}), \tag{2}
\]

with \( c \) the speed of light.

The redshift distance, \( r(z_{\text{obs}}) \), of a galaxy inferred from its observed redshift differs from its true comoving distance, which is given by \( r(z_{\text{cos}}) \). Hence, peculiar velocities give rise to RSDs, which complicate the interpretation of galaxy clustering but also contain important additional information about the cosmic mass distribution. After all, the peculiar velocities are induced by this matter distribution, which is itself correlated with the distribution of galaxies. On small scales the virialized motion of galaxies within dark matter halos causes a reduction in the correlation power, known as the FOG effect, while on larger scales the correlations are boosted due to the infall motion of galaxies toward overdensity regions, known as the Kaiser effect (Kaiser 1987).

Since each galaxy is believed to reside in a dark matter halo, it is useful to split the peculiar velocity of a galaxy into two components:

\[
v_{\text{pec}} = v_{\text{cen}} + v_{\gamma}. \tag{3}
\]

Here \( v_{\text{cen}} \) is the line-of-sight velocity of the center of the halo, and \( v_{\gamma} \) is the line-of-sight component of the velocity vector of the galaxy with respect to that center. Roughly speaking, \( v_{\text{cen}} \) is a manifestation of the Kaiser effect (at least on large scales), while \( v_{\gamma} \) mainly contributes to the FOG effect. Hence, for convenience in what follows, we define the Kaiser and FOG redshifts as

\[

z_{\text{Kaiser}} = z_{\text{cos}} + \frac{v_{\text{cen}}}{c} (1 + z_{\text{cos}}), \tag{4}
\]

\[
z_{\text{FOG}} = z_{\text{cos}} + \frac{v_{\gamma}}{c} (1 + z_{\text{cos}}). \tag{5}
\]

The various redshifts thus defined allow us to define a number of different spaces, in addition to the standard real and redshift spaces. Table 2 gives a brief description of the various spaces used in this study. In each space, galaxy distances are computed using their corresponding redshifts inserted into

| Absolute Magnitude | Flux-limited | Volume-limited |
|--------------------|--------------|----------------|
| \( [0.01, 0.075] \) | [2200(379/1821)] | [2200(379/1821)] |
| [0.075, 0.113] | [42207(11997/30210)] | [42207(11997/30210)] |
| [0.113, 0.12] | [134801(55572/79229)] | [134801(55572/79229)] |

### Table 1

| Redshift | \( N_{\text{gal}} N_{\text{blue}} / N_{\text{real}} \) | Averaged Magnitude |
|------------|-----------------|------------------|
| [0.01, 0.075] | [2200(379/1821)] | −22.22 |
| [0.075, 0.113] | [42207(11997/30210)] | −21.34 |
| [0.113, 0.12] | [134801(55572/79229)] | −20.43 |

\( \log z \) in practice, to keep the large-scale mode at \( z = 0.12 \), we use groups in the redshift range \( 0.01 \leq z \leq 0.13 \) for our velocity reconstruction.

3
Equation (1). All spaces have the geometry of the SDSS DR7. The top four spaces listed are based on true velocities and true groups (dark matter halos), without observational errors, or errors in group identifications and/or membership. The bottom three spaces (those starting with “Re”), on the other hand, are reconstructed spaces, obtained by correcting for the corresponding redshift distortions. These are based on the reconstructed velocity field, and on groups identified by applying the group finder in redshift space (see Section 3.2 below). In what follows, we refer to the top four spaces as “true” spaces, and the lower three spaces as “reconstructed” spaces.

### 3.2. Correcting for RSDs

We now describe our method to correct the redshifts in the SDSS DR7 survey volume for RSDs. The method treats the Kaiser effect and the FOG effect separately, as detailed below.

#### 3.2.1. Correcting for the Kaiser Effect

In order to correct for the Kaiser effect, we reconstruct the velocity field in the linear regime using the method of W12. Here we briefly summarize the main ingredients of this reconstruction method, and refer the reader to W12 for more details. In the linear regime, the peculiar velocities are induced by, and proportional to, the perturbations in the matter distribution. If we write the velocity field, \( v(x) \), as a sum of Fourier modes,

\[
v(x) = \sum_k v(k) e^{i k \cdot x},
\]

then, in the linear regime, each mode can be written as

\[
v(k) = H a f(\Omega) \frac{ik}{k^2} \delta(k).
\]

Here \( H = \dot{a}/a \) is the Hubble parameter, \( a \) is the scale factor, \( \delta(k) \) is the Fourier transform of the density perturbation field \( \delta(x) \), and \( f(\Omega) = d \ln D / d \ln a = \Omega_m^{\gamma} + \frac{1}{2} \Omega_b (1 + \Omega_m/2) \) (e.g., Lahav et al. 1991).

Hence, for a given cosmology one can directly infer the linear velocity field from the density perturbation field, \( \delta(x) \). The challenge, however, is to reconstruct the matter field from observations in redshift space. The unique aspect of the W12 method is that it does not try to reconstruct \( \delta \), but instead focuses on the matter density field, \( \delta_b \), which is the (large-scale) matter distribution due to dark matter halos with a mass \( M_b \gtrsim M_{th} \). As is well known, dark matter halos are biased tracers of the mass distribution (e.g., Mo & White 1996). On large, linear scales we have that \( \delta_b(x) = b_{hm} \delta(x) \), where \( b_{hm} \) is the linear bias parameter for dark matter halos with mass \( M_b \gtrsim M_{th} \), which is given by

\[
b_{hm} = \frac{\int_{M_{th}} M b_h(M) n(M) dM}{\int_{M_{th}} M n(M) dM}\]

where \( n(M) \) and \( b_h(M) \) are the halo mass function and the halo bias function, respectively. Hence, one can reconstruct the peculiar velocity field (on linear scales) from \( \delta_b(x) \) using

\[
v(k) = H a f(\Omega) \frac{ik}{k^2} \frac{\delta_b(k)}{b_{hm}}.
\]

In other words, the velocity field can be reconstructed even if we have only the distribution of dark matter halos above some mass threshold. This is fortunate, since it means that we can use our galaxy group catalog, in which galaxy groups are linked with dark matter halos above some mass threshold.

In order to reconstruct the velocity field in the SDSS survey volume, we proceed as follows. We first embed the survey volume in a periodic cubic box of \( 726 \ h^{-1} \ Mpc \) on a side. The size of this “survey box” is chosen to be about 100 \( h^{-1} \ Mpc \) larger than the maximum scale of the survey volume among the three axes. Next, we divide the box into \( 1024^3 \) grid cells, and use groups with an assigned mass \( M_{th} \gtrsim M_{th} = 10^{12.5} \ h^{-1} M_{\odot} \), to compute \( \delta_b(x) \) on that grid using the method described in detail in W12. In order to suppress nonlinear velocities that are not captured by the linear model, we smooth \( \delta_b(x) \) using a Gaussian smoothing kernel with a mass scale of \( 10^{14.75} \ h^{-1} M_{\odot} \) (see Wang et al. 2009). Next, we fast Fourier transform this smoothed overdensity field, and compute \( v(k) \) using Equation (9), where \( b_{hm} \) is computed using Equation (8) adopting the halo mass and bias functions of Tinker et al. (2008). Fourier transforming \( v(k) \) then yields the velocity field, which we interpret as \( v_{cen}(x) \), the velocity field of group centers. Finally, the comoving distance of each galaxy, corrected for the Kaiser effect, is computed as \( r(z_{corr}) \) (compare with Equation (1)). Here

\[
z_{corr} = \frac{z_{obs} - v_{cen}/c}{1 + v_{cen}/c}
\]

with \( v_{cen} \) the inferred line-of-sight velocity at the location of the group to which the galaxy belongs. The location of the group is defined as the luminosity-weighted center of all group members.

---

**Table 2**  
**Description of Different Spaces**

| Space                      | Description                                                                 |
|----------------------------|-----------------------------------------------------------------------------|
| Real space                 | survey geometry without redshift distortions                                 |
| FOG space                  | distorted only by FOG effect:                                               |
| Redshift space             | \( z_{obs} = z_{cos} + \frac{\nu}{c} (1 + z_{cos}) \)                       |
| Re-FOG space               | corrected real space; based on correcting redshift-space distortions         |
| Re-Kaiser space            | reconstructed Kaiser space; based on correcting FOG effect only              |
| Redshift space             | reconstructed real space; based on correcting redshift-space distortions     |

*Note.* The first four spaces are “true” spaces, based on true groups (all galaxies belonging to the same dark matter halo). The final three spaces are “reconstructed” spaces based on groups identified by applying the group finder in redshift space.
Since the velocity field is computed using the redshift-space distribution of the groups, this method needs to be iterated until convergence is achieved. Using the inferred \(v_{\text{cen}}(x)\), we correct the redshifts of all groups with an inferred mass \(M_b \geq M_{\text{th}}\) for their (inferred) peculiar velocity, and reimpute \(\delta_b(x)\) and \(v_{\text{cen}}(x)\) using the same method. As shown in Wang et al. (2009, 2012), typically two iterations suffice to reach convergence, yielding an unbiased estimate of the linear velocity field.

### 3.2.2. Correcting for the FOG Effect

The FOG effect arises due to the motion of galaxies inside their dark matter halos. To first order, one can simply correct for the FOG effect by assigning all group galaxies the redshift of the group, and then computing the comoving distance using Equation (1). However, this ignores the spatial extent of dark matter halos, which can be quite substantial.

Unfortunately, it is impossible to infer a galaxy’s line-of-sight location from its peculiar velocity along that line of sight. Hence, one can only correct for the FOG effect in a statistical sense, which we do as follows. We assume that group galaxies are unbiased tracers of the halo’s mass distribution, and therefore follow an NFW (Navarro–Frenk–White: Navarro et al. 1997), radial number density profile

\[
n_{\text{gal}}(r) = \frac{n_0}{(r/r_s)(1 + r/r_s)^2},
\]

where \(r_s\) is the characteristic radius, and the normalization parameter \(n_0\) can be expressed in terms of the halo concentration parameter \(c = r_{180}/r_s\) as

\[
n_0 = \frac{N_{\text{gal}}}{4\pi r_s^3} \left[ \ln(1 + c) - c/(1 + c) \right]^{-1}.
\]

Here \(N_{\text{gal}}\) is the number of group member galaxies, and \(r_{180}\) is the radius inside which the halo has an average overdensity of 180. Numerical simulations show that halo concentration depends on halo mass, and we use the relation given by Zhao et al. (2009), converted to the \(c\) appropriate for our definition of halo mass.

In practice, we proceed as follows. We do not displace central galaxies, which are defined to be the brightest group members. For satellite galaxies (all members other than centrals), we first calculate the projected distance \(r_p\) between the galaxy and the luminosity-weighted center of its group. Then we randomly draw a line-of-sight distance, \(r_s\), for the galaxy whose probability follows Equation (11) with

\[
r = \sqrt{r_p^2 + r_s^2}.
\]

The galaxy is then assigned a comoving distance given by \(r(z_{\text{cen}}) + r_s\), with the \(z_{\text{cen}}\) of Equation (10). We have verified that using the location of the central galaxy yields results that are virtually indistinguishable to using the luminosity-weighted center of the group.

### 3.3. Two-point Correlation Functions

In this paper, we use 2PCFs to characterize the clustering of galaxies. We estimate the two-dimensional 2PCF, \(\xi(r_p, r_s)\), for galaxies in each sample using the following estimator:

\[
\xi(r_p, r_s) = \frac{\langle RR \rangle \langle DD \rangle}{\langle DR \rangle^2} - 1,
\]

where \(\langle DD \rangle\), \(\langle RR \rangle\), and \(\langle DR \rangle\) are, respectively, the numbers of galaxy–galaxy, random–random, and galaxy–random pairs with separation \((r_p, r_s)\) (Hamilton 1993). The variables \(r_p\) and \(r_s\) are the pair separations perpendicular and parallel to the line of sight, respectively. Explicitly, for a pair of galaxies, one located at \(s_1\) and the other at \(s_2\), where \(s_i\) is computed using Equation (1), we define

\[
r_p = \frac{s \cdot l}{|l|}, \quad r_s = \sqrt{s \cdot s - r_p^2}.
\]

Here \(l = (s_1 + s_2)/2\) is the line of sight intersecting the pair and \(s = s_1 - s_2\).

The projected 2PCF, \(w_p(r_p)\), is estimated using

\[
w_p(r_p) = \int_{-\infty}^{\infty} \xi(r_p, r_s) \, dr_s = 2 \sum \xi(r_p, r_s) \Delta r_p
\]

(Davis & Peebles 1983). In our analysis, the summation is over 100 bins of \(\Delta r_p = 1 \, h^{-1} \text{Mpc}\), corresponding to an integration from \(r_s = -100 \, h^{-1} \text{Mpc}\) to \(+100 \, h^{-1} \text{Mpc}\). The one-dimensional, redshift-space 2PCF, \(\xi(s)\), is estimated by averaging \(\xi(r_p, r_s)\) along constant \(s = \sqrt{r_p^2 + r_s^2}\) using

\[
\xi(s) = \frac{1}{2} \int_{-1}^{1} \xi(r_p, r_s) \, d\mu,
\]

where \(\mu\) is the cosine of the angle between the line of sight and the redshift-space separation vector \(s\). Alternatively, one can also measure \(\xi(s)\) by directly counting \(\langle DD \rangle\), \(\langle RR \rangle\), and \(\langle DR \rangle\) pairs as a function of redshift-space separation \(s\).

Whereas \(\xi(r_p, r_s)\) and \(\xi(s)\) are affected by RSDs, and can therefore differ dramatically in different spaces (real space, redshift space, Kaiser space, or FOG space), the projected correlation function, which is integrated along the line of sight, is insensitive to RSDs. In practice, though, since we only integrate over a finite extent, the projected correlation function is hampered by residual redshift-space distortions (RSDs). However, as we explicitly demonstrate below, for an integration limit of \(100 \, h^{-1} \text{Mpc}\) these RSDs are sufficiently small and do not significantly impact our results (see also van den Bosch et al. 2013, and references therein).

### 4. Tests Based on Mock Data

Before applying our reconstruction method to SDSS data, we test its accuracy and reliability using a variety of mock SDSS DR7 surveys. In particular, we construct mock galaxy surveys in real space, Kaiser space, FOG space, and redshift space, which allows us to separately test the corrections for the Kaiser and the FOG effects. In order to gauge the accuracy of our reconstruction, we compare clustering statistics from the reconstructed spaces with those obtained from their respective true spaces.

Briefly, our tests therefore consist of the following four steps:

1. Construct mock galaxy samples in real, Kaiser, FOG, and redshift space.
2. Run the galaxy group finder over each of these spaces.
3. Using these galaxy group catalogs, and the reconstruction methods described in Section 3.2, reconstruct the mock galaxy samples in re-Kaiser, re-FOG, and re-real space by correcting for the Kaiser effect, the FOG compression, and both, respectively.
4. Measure the two-dimensional 2PCF $\xi(r_p, r_c)$, the projected 2PCF $w_p(r_p)$, and the redshift-space 2PCF $\xi(s)$, and compare the results from the reconstructed spaces with those from their corresponding true spaces.

### 4.1. The Mock Catalogs

For our study, we use a high-resolution N-body simulation that evolves the distribution of 3072$^3$ dark matter particles in a periodic box of 500 $h^{-1}$ Mpc on a side (Li et al. 2016). This simulation was carried out at the Center for High Performance Computing at Shanghai Jiao Tong University and was run with L-GADGET, a memory-optimized version of GADGET2 (Springel 2005). The cosmological parameters adopted by this simulation are consistent with the WMAP9 results (Hinshaw et al. 2013), and each particle has a mass of $3.4 \times 10^8 h^{-1} M_\odot$. Dark matter halos are identified using the standard friends-of-friends algorithm (e.g., Davis et al. 1985) with a linking length that is 0.2 times the mean interparticle separation. The mass of halos, $M_h$, is simply defined as the sum of the masses of all the particles in the halos, and we remove halos with fewer than 20 particles. We refer to these halos as “real halos” in what follows in order to distinguish them from the groups identified by the group finder that is applied to the mock galaxy catalogs described below.

Based on the halo catalog, we populate galaxies using the conditional luminosity function (CLF) model of Yang et al. (2003). The algorithm for populating galaxies is similar to that outlined in Yang et al. (2004), but here updated to the CLF in the SDSS r-band (see Lu et al. 2015, for a recent application). For completeness, we briefly describe our method used to assign mock galaxies to our dark matter halos.

We write the total CLF as the sum of a central galaxy component and a satellite galaxy component:

$$\Phi(L|M_h) = \Phi_{\text{cen}}(L|M_h) + \Phi_{\text{sat}}(L|M_h).$$

The central component is assumed to follow a log-normal distribution:

$$\Phi_{\text{cen}}(L|M_h) \ d \log L = \frac{1}{\sqrt{2\pi} \sigma_c} \exp \left[-\frac{(\log L - \log L_c)^2}{2\sigma_c^2}\right] \ d \log L. \quad (18)$$

Here $\sigma_c$ is a free parameter that expresses the scatter in $\log L$ of central galaxies at fixed halo mass, and $\log L_c$ is the expectation value for the (base 10) logarithm of the luminosity of the central galaxy. For the contribution from the satellite galaxies we adopt a modified Schechter function:

$$\Phi_{\text{sat}}(L|M) \ d \log L = \phi^* \left(\frac{L}{L^*}\right)^{(\alpha_\delta + 1)} \exp \left[-\left(\frac{L}{L^*}\right)^\delta\right] \ln(10) \ d \log L. \quad (19)$$

Note that the parameters $L_c$, $\sigma_c$, $\phi^*$, $\alpha_\delta$, and $L^*$ are all functions of the halo mass $M_h$.

Following Cacciato et al. (2009), and motivated by the results of Yang et al. (2008) and More et al. (2009), we assume that $\sigma_c$ is a constant (i.e., independent of halo mass) and that the $L_c-M_h$ relation has the following functional form:

$$L_c(M_h) = L_0 \frac{(M_h/M_\odot)^{\gamma_1}}{(1 + M_h/M_\odot)^{\gamma_2}}. \quad (20)$$

This model contains four free parameters: a normalized luminosity, $L_0$, a characteristic halo mass, $M_\odot$, and two slopes, $\gamma_1$ and $\gamma_2$. For satellite galaxies we use

$$\log \phi^*(M_h) = \log L_c(M_h) - 0.25\alpha_\delta(M_h) = \alpha_s$$

(i.e., the faint-end slope of $\phi_{\text{sat}}(L|M_h)$ is independent of halo mass), and

$$\log [\phi^*(M_h)] = b_0 + b_1 (\log M_{12}) + b_2 (\log M_{12})^2,$$  

with $M_{12} = M_h/(10^{12} h^{-1} M_\odot)$. Thus defined, the CLF model has a total of nine free parameters, characterized by the vector

$$\chi^{\text{CLF}} = (\log M_1, \log L_0, \gamma_1, \gamma_2, \sigma_c, \alpha_\delta, b_0, b_1, b_2).$$

We emphasize that this functional form for the CLF accurately describes the observational results obtained by Yang et al. (2008) from the SDSS galaxy group catalog. The same functional form was adopted in Cacciato et al. (2009) to model galaxy–galaxy lensing, and, more recently, in van den Bosch et al. (2013), More et al. (2013), and Cacciato et al. (2013) to simultaneously constrain cosmological parameters and the galaxy–dark matter connection using a combination of SDSS clustering and weak lensing measurements. Here we adopt the set of best-fit CLF parameters listed in Cacciato et al. (2013) for cosmological parameters that are consistent with those used for our numerical simulation: $\log M_1 = 11.24$, $\log L_0 = 9.95$, $\gamma_1 = 3.18$, $\gamma_2 = 0.245$, $\sigma_c = 0.157$, $\alpha_\delta = -1.18$, $b_0 = -1.17$, $b_1 = 1.53$, and $b_2 = -0.217$.

We populate the dark matter halos in our simulation with mock galaxies with luminosities $\log(L/h^{-2} L_\odot) \gtrsim 7.0$ using the following approach. First, each halo is assigned a central galaxy whose luminosity is drawn from the log-normal distribution of Equation (18). The central galaxy is assumed to be located at rest at the center of the corresponding halo. Next, we populate the halo with satellite galaxies via the following steps: (1) obtain the mean number of satellite galaxies according to the integration of Equation (19) with luminosities $\log L \gtrsim 7.0$; (2) draw the actual number of satellite galaxies for the halo in question from a Poisson distribution with the mean obtained in step (1); (3) assign a luminosity to each satellite galaxy according to Equation (19). Note that satellite galaxies are allowed to be brighter than their central galaxy. Finally the phase-space coordinates (positions and velocities) of the satellite galaxies are drawn from the randomly selected dark matter particles in the halos. As we have tested, populating satellite galaxies in phase space according to an NFW profile yields quite similar results.

Next, we proceed to construct mock galaxy samples that have the same survey selection effects as the SDSS DR7. We stack $3 \times 3 \times 3$ replicas of the populated simulation box and place a virtual observer at the center of central box. We define an $(\alpha, \delta)$ coordinate system, and remove all mock galaxies that are located outside the SDSS DR7 survey region. We then assign each galaxy a redshift and $r$-band apparent magnitude according to its distance, line-of-sight velocity, and luminosity, and select galaxies according to the position-dependent
The Astrophysical Journal, 833:241 (19pp), 2016 December 20

Shi et al.

magnitude limit. Finally, we mimic the position-dependent completeness by randomly sampling each galaxy using the completeness masks provided by the SDSS DR7. In order to have a rough estimation of the cosmic variance, we construct a total of 10 such mock samples by randomly rotating and shifting the boxes in the stack. Note that in order to get a more accurate estimation of the cosmic variance, many more mocks are needed. From each mock sample, six flux-limited (and volume-limited) subsamples are constructed using the ranges of redshift and absolute magnitude listed in Table 1.

Finally, in order to disentangle the various redshift distortions, for each mock galaxy redshift catalog we construct four different versions that differ only in the redshift \( z_{\text{obs}} \) assigned to each mock galaxy: a real-space version in which \( z_{\text{obs}} = z_{\text{obs}} \), a Kaiser-space version in which \( z_{\text{obs}} = z_{\text{Kaiser}} \) (Equation (4)), a FOG-space version in which \( z_{\text{obs}} = z_{\text{FOG}} \) (Equation (5)), and a redshift-space version in which \( z_{\text{obs}} \) is given by Equation (2).  

4.2. Results for Mock Catalogs

In order to gauge the impact of the various redshift distortions, we now carry out clustering analyses of the various mock galaxy catalogs described above. We start our investigation by computing the two-dimensional 2PCF, \( \xi(r_p, r_p) \). Figure 1 shows the average results (black solid lines) from the 10 mock samples for the four true spaces. Here we show only the results for the \([-21.0, -20.0]\) subsample, but note that the results for the other subsamples are qualitatively very similar. The red dashed lines in the upper left panel show the \( \pm 1\sigma \) cosmic variance as inferred from our 10 mock samples. For enhanced clarity, we show these only in real space. Note that the variance causes small fluctuations at small transverse separations, \( r_p \), especially at larger line-of-sight separations, \( r_p \).

Clearly, the shape of the two-dimensional correlation function is very different in different spaces: whereas \( \xi(r_p, r_p) \) is isotropic in real space, it is squashed along the line of sight on large scales in Kaiser space, and elongated along the line of sight on small scales in FOG space. Finally, in redshift space \( \xi(r_p, r_p) \) reveals the characteristics of both Kaiser and FOG space. All of this has been well known since the seminal work by Davis & Peebles (1983).

Since redshift distortions only displace galaxies along the line of sight, they should not affect the projected correlation function, \( w_p(r_p) \), notwithstanding RRSDs that arise from the use of a finite integration range (see discussion in Section 3.3). The lines in Figure 2 show the projected 2PCFs in all four true spaces, and for all six absolute magnitudes bins: \([-23.0, -22.0]\), \([-22.0, -21.0]\), \([-21.0, -20.0]\), \([-20.0, -19.0]\), \([-19.0, -18.0]\), \([-18.0, -17.0]\). Error bars reflect the \( \pm 1\sigma \) variance among the 10 mock samples, and, for clarity, are plotted only for the real-space results (they are very similar in all other spaces). As expected, the various \( w_p(r_p) \) are in good agreement with each other, indicating that the impact of RRSDs is small compared to cosmic variance errors.

Finally, Figure 3 shows the 2PCF, \( \xi(s) \), for the same magnitude bins and the same four spaces. As before, error bars are obtained from the 10 mock samples, and plotted only for the real and redshift spaces for clarity. Unlike the projected correlation function, \( \xi(s) \) clearly reveals the impact of redshift distortions. Compared to the real-space correlation function, \( \xi(s) \) in Kaiser space is significantly boosted on large scales due to the large-scale flows toward overdense regions (Kaiser effect). On small scales, however, the Kaiser-space correlation function is virtually indistinguishable from the real-space correlation function. \( \xi(s) \) in FOG space, on the other hand, is identical to the real-space \( \xi(s) \) on large scales, but dramatically suppressed on small scales. Finally, \( \xi(s) \) in redshift space clearly reveals redshift distortions from both the Kaiser effect and the FOG effect.

Figure 1. The two-dimensional 2PCFs for mock galaxies with absolute r-band magnitudes in the range \(-21 \leq M < -20\) for four different spaces (see Table 2): real space (upper left), Kaiser space (upper right), FOG space (lower left), and redshift space (lower right). Black contours indicate the average values inferred from 10 mock samples. The contour levels correspond to \( \pm 1\sigma \) cosmic variance.

Figure 2. Comparison of the projected two-point correlation functions in all seven mock spaces. Different panels correspond to different bins in absolute r-band magnitude, as indicated. For clarity, the error bars, which are obtained from the 10 mock samples, are plotted only for the real-space results. Note that, as expected, all projected correlation functions are virtually indistinguishable.
Figure 3. The two-point correlation functions of mock galaxies in different true spaces. Results are shown for six different intervals in absolute $r$-band magnitude, as indicated. For clarity, we only plot error bars (expressing the variance among our 10 mock samples) for the real-space and redshift-space results.

4.3. Results for Reconstructed Catalogs

Thus far we have constructed mock SDSS DR7 galaxy catalogs in four true spaces that allow us to disentangle the impact of the FOG effect on small scales from the Kaiser effect on large scales. We have shown that the results from statistical analyses of galaxy clustering in these different spaces agree with expectations. We now proceed with using these mock catalogs to test the reliability and accuracy of the reconstruction method described in Section 3.2. We start by running the halo-based group finder of Yang et al. (2005, 2007) over each of the separate true-space mock galaxy catalogs. This yields corresponding mock group catalogs, in which each group is assigned a halo mass based on its characteristic luminosity, as described in Y07. Similar to the SDSS group catalog, the mock group catalogs are also complete to $z \sim 0.12$ for groups with an assigned halo mass $M_{h} \geq 10^{12.5} h^{-1} M_{\odot}$. We thus adopt a threshold mass of $M_{h} = 10^{12.5} h^{-1} M_{\odot}$ and restrict our reconstruction to the volume covering the redshift range $0.01 \leq z \leq 0.12$.

Next we use the method for correcting redshift distortion described in Section 3.2 to obtain mock galaxy catalogs in re-FOG, re-Kaiser, and re-real space. In this subsection we focus on comparing the clustering of galaxies in the reconstructed spaces with that in the corresponding true spaces. The goal is to investigate the accuracy with which the reconstruction method can recover the distribution of galaxies in real space. Throughout we characterize the clustering using the various two-point correlation functions introduced above and we use the 10 independent mock samples to gauge the impact of sample variance.

4.3.1. The Two-dimensional Correlation Function $\xi(r_p, r_x)$

We start with a qualitative, visual comparison based on the two-dimensional 2PCF $\xi(r_p, r_x)$. Different rows in Figure 4 correspond to different magnitude bins, as indicated to the right of each row. From left to right, the different columns show the results in redshift space, a comparison of FOG versus re-FOG, a comparison of Kaiser versus re-Kaiser, and a comparison of real versus re-real. In each case black and red contours correspond to the true and reconstructed spaces, respectively.

The $\xi(r_p, r_x)$ in redshift space is clearly anisotropic, revealing fingers-of-God on small scales and the impact of the Kaiser effect on large scales. After correcting for the Kaiser effect, the resulting $\xi(r_p, r_x)$ in re-FOG space is clearly more isotropic on large scales. As expected, it still reveals the impact of the FOG effect, which distorts the contours from being perfectly round. A comparison with the $\xi(r_p, r_x)$ in FOG space shows that the correction for the Kaiser effect is overall very successful, except for small differences in the outer contour (corresponding to $\xi = 0.1$).

Comparing the $\xi(r_p, r_x)$ in re-Kaiser space (third column from the left) with that in redshift space (left-hand column) shows that our method of FOG compression is fairly accurate. However, a comparison with the true Kaiser-space results (black contours in third column) reveals that the method is not perfect. On small scales, the $\xi(r_p, r_x)$ in re-Kaiser space shows very nice agreement with that in real space. On large scales, however, the correlation function in re-Kaiser space reveals residual FOG effects. These shortcomings of the FOG compression may arise from problems with the group finder, including errors in determining group membership ("fracturing" and "fusing" of groups), errors in the designation of centrals and satellites, and errors in the assignment of halo mass. These errors are characteristic of all group finders, and are virtually impossible to avoid (see Campbell et al. 2015, for details).

Finally, the results in the rightmost column show that the reconstruction of $\xi(r_p, r_x)$ in real space manifests both the problems with the Kaiser correction and the FOG compression. Overall, though, by comparing the correlation function in real space with that in redshift space, it is clear that the reconstruction method has successfully corrected for the majority of RSDs. In order to make this more quantitative, we now focus on $\xi(s)$.

4.3.2. The One-dimensional Correlation Function $\xi(s)$

Figure 5 compares the 2PCF, obtained by averaging results from all 10 mocks, in a true space ($\xi_{true}$, solid lines) to that in the corresponding reconstructed space ($\xi_{recon}$, blue filled circles). From top to bottom, the three parts of this figure show a comparison of (a) FOG space versus re-FOG space, (b) Kaiser space versus re-Kaiser space, and (c) real space versus re-real space. Different columns correspond to different magnitude bins, as indicated, and error bars indicate the variance among the 10 mock samples. In each part, the upper panels show the actual 2PCFs, while the lower panels plot $\xi_{recon}/\xi_{true}$. Overall, the correlation functions in the reconstructed spaces are in excellent agreement with those in their corresponding true spaces, with the vast majority of data points being consistent with $\xi_{recon}/\xi_{true} = 1$ within $1\sigma$. Recall that $\sigma$ reflects the measurement error due to cosmic variance in an SDSS-like survey.

As is evident from the middle part (b), the FOG compression seems to systematically underpredict the Kaiser-space 2PCF for faint galaxies. The effect, which results from inaccuracies in the group finder, is somewhat significant in the two low-mass bins. Thus in an accurate modeling of the distribution of halo occupation of galaxies for these faint galaxies, one needs to take this effect into account. For brighter galaxies, over the
range of scales $0.2 \leq s/ h^{-1}\text{Mpc} \leq 20$, the average value of $\xi_{\text{re-real}}/\xi_{\text{real}}$ is $1.00 \pm 0.050$. Hence, we conclude that over those scales the reconstruction of the real-space correlation function is accurate at the level of 5%. For comparison, the dashed lines in the bottom part (c) of Figure 5 correspond to the 2PCF in redshift space. On small scales ($r < 1 h^{-1}\text{Mpc}$), the clustering strength in redshift space is suppressed by $\sim 70\%$ on average, compared to that in real space. On large scales, ($r > 2 h^{-1}\text{Mpc}$) it is boosted by $\sim 30\%$ on average.

4.3.3. The Projected Correlation Function $w_p(r_p)$

Moreover, since our reconstruction only “displaces” galaxies along the line of sight, the reconstruction method has no impact on the projected correlation function, $w_p(r_p)$, other than scattering a few galaxy pairs in and out of the sample due to the finite integration range used ($|r_p| \leq 100 h^{-1}\text{Mpc}$). This effect is entirely negligible, however, as is evident from Figure 2, which shows the results for all our seven spaces (four true spaces and three reconstructed spaces). There are no
significant differences among these different projected correlation functions.

### 4.3.4. The Bias Factor

The correlation function of galaxies relative to that of dark matter is usually described by a bias factor, which is defined as

\[ \xi_{gg}(s) = b^2 \xi_{mm}(s), \]  

where \( \xi_{gg} \) and \( \xi_{mm} \) are the correlation functions of galaxies and mass, respectively. In general, the bias factor \( b \) may depend on \( s \).

Figure 6 shows the best-fitting bias factor, as a function of galaxy luminosity, obtained from the measured \( \xi(s) \) for mock galaxies relative to the correlation function of dark matter at \( z = 0.1 \). The real-space and reconstructed real-space \( b \) shown in the left panel are obtained from using the values of \( \xi(s) \) on large scales, 4 \( h^{-1} \) Mpc < \( s \) < 20 \( h^{-1} \) Mpc, while in the right panel they are obtained using the correlation functions on small scales, 0.5 \( h^{-1} \) Mpc < \( s \) < 2 \( h^{-1} \) Mpc. For comparison, we also show in Figure 6 the bias factor based on the projected 2PCFs (red lines), defined as the ratios of \( w_p(r_p) \) between galaxies and dark matter over the ranges 4 \( h^{-1} \) Mpc < \( r_p \) < 20 \( h^{-1} \) Mpc (left panel) and 0.5 \( h^{-1} \) Mpc < \( r_p \) < 2 \( h^{-1} \) Mpc (right panel). As one can see, the reconstructed real-space \( b \) closely matches that in real space, while the traditional method based on \( w(r) \) leads to larger errors and biased results relative to the true real-space values.

### 4.3.5. The Quadrupole-to-monopole Ratio \( q(s) \)

As a final diagnostic of our reconstruction performance, we consider the quadrupole-to-monopole ratio, which is defined as

\[ q(s) = \frac{\xi_2(s)}{\frac{3}{2} \int_0^s \xi_0(s') s'^2 ds' - \xi_0(s)} \]
become negative on large scales, tending asymptotically toward
\[ q(s) = \frac{-4}{3} \beta - \frac{4}{7} \beta^2. \]  
(28)
with \( \beta = f(\Omega)/b \) with \( b \) the bias parameter of the galaxy population under consideration (e.g., Hamilton 1992; Cole et al. 1994). On small scales the FOG effect causes \( q(s) \) to become positive. In real space, however, we expect isotropy to result in a quadrupole \( \xi_2(s) = 0 \). Hence, if the correction for redshift distortions is successful, the resulting clustering should have a vanishing quadrupole, and thus \( q(s) = 0 \).

Figure 7 shows the quadrupole-to-monopole ratio for our mock galaxies in real space, re-real space, and redshift space. Different panels correspond to different magnitude bins, as indicated. As expected, in redshift space \( q(s) \) has large deviations from zero on both small and large scales, while in real space \( q(s) \) is close to zero (except for a small positive signal for \( r < 3 \, h^{-1} \) Mpc, which is due to noise). In re-real space, the quadrupole-to-monopole ratio in the re-real space is consistent with zero within the error bars on large scales (\( \geq 12 \, h^{-1} \) Mpc). On smaller scales, all magnitude bins reveal a slightly negative \( q(s) \). This is a consequence of the over-correction for the FOG effect on small scales discussed in Section 4.3.1 (see Figure 4), which has its origin in inaccuracies associated with the galaxy group finder.

5. APPLICATION TO THE SDSS

Based on the analyses of the mock galaxy samples discussed in Section 4, we conclude that our reconstruction method can accurately correct for RSDs in a statistical sense. In this section we apply exactly the same method to the SDSS DR7. As described in Section 2 we follow W12 and reconstruct the velocity field on quasi-linear scales using the mass distribution reconstructed from galaxy groups of Y07 in the redshift range \( 0.01 \leq z \leq 0.12 \) and with assigned halo masses \( \log(M_\text{h}/h^{-1} \text{M}_\odot) \geq 12.5 \). We use the velocities derived to correct for the Kaiser effect using the method described in Section 3.2.1. Finally, we correct for the FOG effect by assigning all galaxies new positions within their groups based on the method described in Section 3.2.2. We apply this method to all the 396,068 galaxies in the NGC region. The reconstructed real-space galaxy catalog is publicly available through [http://gax.shao.ac.cn/data/data1/SDSS_REAL/SDS7_REAL.tar.gz](http://gax.shao.ac.cn/data/data1/SDSS_REAL/SDS7_REAL.tar.gz).

5.1. The Galaxy Distribution

To visualize the effects of our reconstruction method on galaxy distribution, we show in Figure 8 the distributions of galaxies with decl. \( |\delta| < 4^\circ \), R.A. \( 10^5 \leq \alpha \leq 14^b \), and redshift \( 0.01 \leq z \leq 0.1 \). The four different panels show the galaxy distributions in redshift space (upper left panel), re-FOG space (upper right panel), re-Kaiser space (lower left panel), and re-real space (lower right panel). Note that the volume chosen includes the Sloan Great Wall, which is readily visible in the upper left corner (\( z \sim 0.085 \) and \( 12^b \leq \alpha \leq 14^b \)).

There are a few noteworthy trends. First of all, the prominent “finger” structures clearly visible in redshift space are no longer visible in the re-Kaiser space, indicating that our FOG compression is successful. Comparing the distribution in
redshift space with that in re-real space, one sees that the latter appears more diffused on large scales, more compressed on small scales. In particular, the Sloan Great Wall is clearly much broader, and thus less pronounced, in the re-FOG and re-real spaces. This suggests that the Sloan Great Wall is not as dominant an overdense structure as it appears to be in redshift space, but that its apparent overdensity is strongly enhanced by the Kaiser effect.

It is also clear from Figure 8 that some geometrical properties of the large-scale structure may also be affected as one goes from real space to RSD. For example, the voids appear to be smaller and the filamentary structures less prominent in real space. Clearly, detailed analyses are needed in order to quantify the effects, and our reconstructed real-space catalog of SDSS DR7 provides a unique resource for such studies.

5.2. The Clustering of Galaxies

Next we investigate the galaxy clustering properties. It is important to note that the reconstruction to obtain the re-real space is cosmology-dependent. The bias parameter $b$, the halo mass assignments to galaxy groups, and the distance–redshift relation are all cosmology-dependent. In the reconstruction of the SDSS DR7, we have adopted the cosmological parameters as inferred from WMAP9. To check the impact of cosmology on our results, we also adopt a Planck cosmology ($\Omega_m = 0.308$, $\Omega_L = 0.692$, $n_s = 0.968$, $h = H_0/(100 \text{ km s}^{-1} \text{Mpc}^{-1}) = 0.678$, and $\sigma_8 =$...
In what follows, we mainly focus on the results for the WMAP9 cosmology; results for Planck cosmology are also presented where necessary.

The black contours in Figure 9 show the two-dimensional 2PCFs $\xi(r_p, r_z)$, for galaxies in four luminosity bins, in redshift space (upper panels) and re-real space (lower panels) for WMAP9 cosmology. The green contours are results for Planck cosmology, and show quite good agreement with those for WMAP9 cosmology. After the correction of the redshift distortion, $\xi(r_p, r_z)$ is clearly much more isotropic than in redshift space. However, it is also clear that the correction is not perfect, especially on small transverse scales where residual deviations from isotropy are apparent. To assess the significance of these deviations, we use the 10 mock galaxy samples in re-real space.

Figure 10 shows the one-dimensional 2PCFs in redshift space (red lines) and in re-real space (black lines) for WMAP9 cosmology, for all the six magnitude samples, as indicated. The green lines are results for Planck cosmology, and again show very good agreement with those for WMAP9 cosmology. For comparison, the results of both the flux-limited and volume-limited samples are shown. Note that for the two brightest samples, flux-limited and volume-limited samples are identical. For the other samples, the correlation functions obtained from the two types of samples are very similar, even though the samples themselves are quite different, especially for the faint magnitude bins (see Table 1). Error bars for the real-space correlation function indicate the $\pm 1\sigma$ variance among the 10 re-real mock samples described in Section 4. All the results shown in the figure are also listed in Table 3. To our knowledge, this is the first attempt to infer the real-space correlation function of galaxies in the SDSS directly from a reconstructed real-space galaxy catalog. Note that the real-space 2PCFs clearly deviate from a simple, single power law, revealing a clear one-halo to two-halo transition on scales of $1-3$ h$^{-1}$ Mpc. As demonstrated in Section 4, this transition is more pronounced in real space than in the projected space. It is therefore expected that fitting halo occupation models directly to the real-space correlation functions presented here will provide more stringent constraints on the connection between galaxies and dark matter halos—something we will pursue in a forthcoming paper. Finally, the lower panels of Figure 10 show the ratio $\xi_l/\xi_r$, where $\xi_l(s)$ and $\xi_r(s)$ are the 2PCFs in redshift space and re-real space, respectively. This nicely shows how RSDs boost the correlation power on large scales (by about 40%–50% on a scale of $10$ h$^{-1}$ Mpc), while suppressing it on small scales (by about 70%–80% on a scale of $0.3$ h$^{-1}$ Mpc).

To study how galaxy clustering depends on galaxy color, we use the bimodal distribution in the color–magnitude plane (e.g., Strateva et al. 2001; Baldry et al. 2004) to divide each of the luminosity samples into “blue” and “red” subsamples. Specifically, the demarcation line we use is $(g - r) = 0.21-0.03M_r$, as is in Zehavi et al. (2011). Information about these subsamples is given in Table 1.

Figure 11 shows the 2PCFs of red (red lines) and blue (blue lines) galaxies in re-real space for different magnitude bins, as indicated. The result of the full sample in each magnitude bin is...
also shown in each panel by the black line. Green lines show the cross-correlation functions between blue and red galaxies. The cross-correlation is obtained by replacing $\xi_{DD}$, $\xi_{RR}$, and $\xi_{DR}$ with $\xi_{DD}^{12}$, $\xi_{RR}^{12}$, and $(\xi_{DR}^{12} + \xi_{DR}^{21})/2$, respectively, in Equation (13). Here subscripts “1” and “2” denote red and blue galaxies, respectively, so that $D_1 D_2$ is the number of cross pairs between red and blue galaxies, and so on. Error bars are obtained from the 10 mock samples. All the data shown in this plot are also listed in Table 4 for reference. As one can see, red galaxies exhibit a higher clustering amplitude than blue ones in the same luminosity bin, and the cross-correlation lies in between. The difference between red and blue galaxies appears to be larger for fainter galaxies.

Figure 12 shows the bias factors defined in the same way as those in Figure 6. Solid lines in the left panel show the bias factors obtained from using the values of $\xi(s)$ on large scales, $4 \, h^{-1} \text{Mpc} < s < 20 \, h^{-1} \text{Mpc}$, while solid lines in the right panel show the bias factors obtained by using data on small scales, $0.5 \, h^{-1} \text{Mpc} < s < 2 \, h^{-1} \text{Mpc}$. Black, red, and blue lines show the results for all, red, and blue galaxies in each $M_r$ bin, respectively. Clearly, the bias factor depends on galaxy luminosity, but the dependence is not the same for red and blue galaxies. Overall, red galaxies have a higher bias factor than their blue counterparts in the same luminosity bin. The difference is largest for faint galaxies on small scales. For the total and blue populations, the bias factor on large scales increases with luminosity. In contrast, for red galaxies, the bias factor on large scales remains more or less constant all the way to $M_r = 5 \log h \sim -21.5$, and only increases with luminosity for the brightest galaxies. On small scales, the bias factor is quite independent of luminosity for both the total and blue populations at $M_r = 5 \log h > -20.5$, and increases with luminosity for higher luminosities. In contrast, the bias factor for red galaxies decreases with increasing luminosity, especially for faint galaxies. This indicates that faint red galaxies

Figure 10. The 2PCFs and 2PCF ratios for SDSS galaxies in redshift space (red lines) and the reconstructed real space (WMAP9 with black lines, Planck with green lines). For comparison, the 2PCFs of the flux-limited samples (solid lines) and volume-limited samples (dashed lines) are both shown. Error bars, shown only for the re-real space results, are the $\pm 1\sigma$ variance among the 10 mock samples discussed in Section 4.1. The red curves in the lower panels are the ratios of the redshift-space to re-real-space 2PCFs. Different columns correspond to different bins in absolute $r$-band magnitude, as indicated.
Table 3
The 2PCFs Obtained from SDSS DR7 in Redshift Space and Reconstructed Real Space

| $r$  | [-23, -22] | [-22, -21] | [-21, -20] | [-20, -19] | [-19, -18] | [-18, -17] |
|------|------------|------------|------------|------------|------------|------------|
|      | $\xi(\xi')$ | $\Delta \xi$ | $\xi(\xi')$ | $\Delta \xi$ | $\xi(\xi')$ | $\Delta \xi$ | $\xi(\xi')$ | $\Delta \xi$ | $\xi(\xi')$ | $\Delta \xi$ | $\xi(\xi')$ | $\Delta \xi$ |
| Redshift space | 0.14 | 113.849 | 34.734 | 59.898 | 14.312 | 47.269 | 6.318 | 38.225 | 12.109 | 22.684 | 19.146 |
|      | 0.28 | 49.288 | 8.310 | 30.435 | 2.486 | 25.203 | 1.322 | 20.512 | 0.770 | 17.377 | 11.908 |
|      | 0.56 | 42.401 | 22.572 | 23.400 | 1.410 | 15.645 | 0.730 | 14.101 | 0.519 | 13.083 | 0.954 |
|      | 1.12 | 19.626 | 10.093 | 11.659 | 0.382 | 8.686 | 0.231 | 7.078 | 0.226 | 6.540 | 0.356 |
|      | 2.24 | 9.474 | 4.009 | 5.424 | 0.136 | 4.070 | 0.086 | 3.313 | 0.097 | 3.070 | 0.225 |
|      | 4.47 | 5.308 | 0.645 | 2.191 | 0.055 | 1.750 | 0.034 | 1.479 | 0.047 | 1.346 | 0.117 |
|      | 8.91 | 1.516 | 0.090 | 0.773 | 0.032 | 0.652 | 0.022 | 0.539 | 0.030 | 0.470 | 0.062 |
|      | 17.78 | 0.458 | 0.058 | 0.221 | 0.022 | 0.190 | 0.014 | 0.137 | 0.020 | 0.116 | 0.024 |
|      | 35.48 | 0.109 | 0.023 | 0.055 | 0.007 | 0.048 | 0.006 | 0.028 | 0.006 | 0.024 | 0.008 |
| Re-real space | 0.14 | 1386.112 | 517.165 | 401.879 | 63.661 | 314.918 | 75.121 | 315.778 | 52.512 | 203.865 | 113.507 |
|      | 0.28 | 403.161 | 91.274 | 145.658 | 17.368 | 137.039 | 18.983 | 112.085 | 16.460 | 85.958 | 74.667 |
|      | 0.56 | 1308.251 | 425.252 | 82.973 | 178.764 | 31.54 | 314 | 44.122 | 11.189 | 40.043 | 39.228 |
|      | 1.12 | 90.370 | 27.875 | 61.745 | 0.705 | 13.028 | 0.905 | 13.116 | 0.878 | 13.187 | 11.799 |
|      | 2.24 | 7.865 | 3.316 | 4.576 | 0.161 | 3.757 | 0.132 | 3.248 | 0.229 | 3.321 | 0.687 |
|      | 4.47 | 3.645 | 0.730 | 1.737 | 0.051 | 1.386 | 0.028 | 1.164 | 0.037 | 1.092 | 0.071 |
|      | 8.91 | 1.175 | 0.125 | 0.567 | 0.027 | 0.478 | 0.015 | 0.385 | 0.019 | 0.342 | 0.048 |
|      | 17.78 | 0.413 | 0.034 | 0.180 | 0.017 | 0.152 | 0.012 | 0.110 | 0.016 | 0.097 | 0.021 |
|      | 35.48 | 0.105 | 0.019 | 0.044 | 0.006 | 0.037 | 0.005 | 0.021 | 0.005 | 0.023 | 0.005 |

Note. $r$: the comoving distances in units of $h^{-1}$ Mpc. $\xi$: the two-point correlation function for flux-limited samples. $\xi'$: the two-point correlation function for volume-limited samples (the flux- and volume-limited samples are the same for the first two samples). $\Delta \xi$: the $1\sigma$ error of $\xi(s)$ estimated using 10 mock samples.
are preferentially satellites located in relatively big halos, consistent with the results of Lan et al. (2016) based on the luminosity functions of galaxies in groups.

For comparison, the dashed lines in Figure 12 show the bias parameters obtained from the projected 2PCFs, $w_p(r_p)$, again estimated in the same way as those for mock galaxies (see Figure 6). The results show again that the bias parameter, $b$, estimated from the projected 2PCF has larger errors and is biased relative to that obtained from the reconstructed real-space $\xi(s)$, as is demonstrated using mock samples shown in Figure 6. This suggests that the bias parameters obtained earlier in the literature on the basis of $w_p(r_p)$ may be significantly biased. We will come back to a detailed analysis of this in a forthcoming paper.

Finally, we compute the quadrupole-to-monopole ratio $q(s)$ for the SDSS DR7 galaxies. Figure 13 shows $q(s)$ for two luminosity samples, $M_r = [-22.0, -21.0]$ and $[-21.0, -20.0]$. In each panel, results are shown for galaxies in both redshift and re-real spaces using lines with different colors, as indicated. The error bars on the zero line correspond to $1\sigma$ variances obtained from 10 mock samples in re-real space. We see that $q(s)$ in re-real space in SDSS DR7 has a systematic deviation from the zero line at the $2\sigma$ level, especially for the high-luminosity bin. This deviation may indicate that at $z \lesssim 0.12$ the SDSS DR7 volume still suffers from cosmic variance, likely produced by the existence of rare large-scale structures, such as the Sloan Great Wall. To check this we estimate $q(s)$ excluding galaxies with redshifts $0.065 \leq z \leq 0.09$, which effectively excludes the Sloan Great Wall. The results are shown in Figure 13 as the blue lines. The deviations from the zero line are significantly reduced at large $s$. This test result suggests that the quadrupole-to-monopole ratio is sensitive to the presence of large-scale structures, and a much larger volume is required to get a reliable estimate of this quantity.

On the other hand, as discussed at the beginning of this subsection, the reconstruction to obtain the re-real-space distribution of galaxies is cosmology-dependent. If the real universe deviates from the assumed cosmology, systematic errors can also be introduced in our reconstruction. The ratios $q(s)$ for Planck cosmology, which are shown in Figure 13 as the green dashed lines, do show some differences from those for the WMAP9 cosmology. After the removal of the Sloan Great Wall, the deviation of $q(s)$ from zero is about 0.1 at $s \sim 20 h^{-1}$ Mpc. This corresponds to an underestimate of $\beta$ by about 0.07 in the linear regime by WMAP9. We will perform a detailed cosmological probe in a subsequent paper.

6. SUMMARY

We have presented a method to correct RSDs in redshift surveys of galaxies. Adopting the method introduced in W12, we use galaxy groups identified with the halo-based group finder of Yang et al. (2005) to reconstruct the large-scale velocity field, which in turn is used to correct the observed redshifts for the Kaiser effect. The same galaxy groups are also used to correct the FOG effect produced by the virial motions of galaxies within their host dark matter halos. Our FOG correction is based on the assumption that satellite galaxies are an unbiased tracer of the mass profile and velocity structure of the host halo.

To test the method, we have constructed 10 mock SDSS DR7 galaxy catalogs in four different spaces: redshift space (equivalent to the observational space), Kaiser space (in which the FOG effect is absent), FOG space (in which the Kaiser effect is absent), and real space (in which redshift distortions are absent). We test the various components of our reconstruction method by comparing the two-point clustering statistics in these different spaces.

The contours of the two-dimensional 2PCFs $\xi(r_p, r_v)$ calculated in different spaces show that the clustering in our reconstructed space is in good agreement with that in the corresponding true space given directly by numerical simulations. On small transverse scales $r_p$, residual FOG effects are apparent, which arise mainly from the uncertainties in the
Table 4
The Color Dependence of the 2PCF Measured from SDSS DR7 in Reconstructed Real Space

|   | [-23, -22] | [-22, -21] | [-21, -20] | [-20, -19] | [-19, -18] | [-18, -17] |
|---|-----------|-----------|-----------|-----------|-----------|-----------|
| r | \(\xi\)   | \(\Delta \xi\) | \(\xi\)   | \(\Delta \xi\) | \(\xi\)   | \(\Delta \xi\) | \(\xi\)   | \(\Delta \xi\) | \(\xi\)   | \(\Delta \xi\) | \(\xi\)   | \(\Delta \xi\) |
| Blue Galaxies | 0.14 | 1771.494 | 367.491 | 429.502 | 348.935 | 218.154 | 154.158 | 213.160 | 134.140 | 108.841 | 106.506 |
|   | 0.28 | 240.858 | 171.663 | 70.223 | 31.927 | 42.708 | 20.893 | 50.913 | 39.207 | 51.447 | 43.761 |
|   | 0.56 | 40.295 | 33.505 | 20.414 | 2.658 | 17.686 | 4.379 | 16.873 | 12.391 | 19.772 | 16.476 |
|   | 1.12 | 8.188 | 3.437 | 6.342 | 1.239 | 5.974 | 2.135 | 6.896 | 5.297 | 8.139 | 3.767 |
|   | 2.24 | 3.336 | 0.445 | 2.458 | 0.219 | 2.162 | 0.511 | 2.357 | 1.391 | 2.791 | 0.206 |
|   | 4.47 | 3.294 | 2.496 | 1.185 | 0.062 | 0.963 | 0.036 | 0.846 | 0.054 | 0.903 | 0.343 |
|   | 8.91 | 0.461 | 0.407 | 0.376 | 0.050 | 0.323 | 0.022 | 0.257 | 0.034 | 0.244 | 0.108 |
|   | 17.78 | 0.369 | 0.217 | 0.129 | 0.027 | 0.104 | 0.018 | 0.064 | 0.021 | 0.061 | 0.045 |
|   | 35.48 | 0.122 | 0.076 | 0.035 | 0.007 | 0.029 | 0.007 | 0.010 | 0.013 | 0.011 | 0.015 |
| Red Galaxies | 0.14 | 1203.324 | 828.281 | 762.580 | 387.987 | 1000.507 | 237.858 | 267.352 | 552.352 |
|   | 0.28 | 741.101 | 152.922 | 288.323 | 28.802 | 400.555 | 33.271 | 717.431 | 262.018 |
|   | 0.56 | 122.181 | 87.634 | 85.996 | 4.259 | 125.857 | 7.515 | 286.505 | 47.521 |
|   | 1.12 | 83.521 | 61.860 | 20.250 | 1.187 | 20.914 | 2.145 | 61.773 | 14.289 |
|   | 2.24 | 6.535 | 3.474 | 5.244 | 0.301 | 5.073 | 0.261 | 11.068 | 2.652 |
|   | 4.47 | 5.235 | 0.688 | 2.000 | 0.077 | 1.771 | 0.039 | 2.063 | 0.215 |
|   | 8.91 | 1.344 | 0.156 | 0.666 | 0.032 | 0.613 | 0.023 | 0.660 | 0.111 |
|   | 17.78 | 0.463 | 0.060 | 0.209 | 0.020 | 0.196 | 0.017 | 0.218 | 0.048 |
|   | 35.48 | 0.119 | 0.033 | 0.051 | 0.006 | 0.050 | 0.007 | 0.035 | 0.015 |
| Blue–Red | 0.14 | 7239.943 | 1692.040 | 378.739 | 171.435 | 395.622 | 151.954 | 298.611 | 245.189 |
|   | 0.28 | 322.883 | 113.649 | 102.018 | 22.364 | 108.949 | 25.905 | 141.057 | 71.610 |
|   | 0.56 | 29.292 | 11.156 | 70.285 | 6.320 | 31.018 | 2.973 | 46.072 | 27.418 |
|   | 1.12 | 24.140 | 12.885 | 11.160 | 2.338 | 9.037 | 1.147 | 16.782 | 14.033 |
|   | 2.24 | 3.424 | 5.895 | 3.949 | 0.280 | 3.086 | 0.235 | 4.082 | 2.574 |
|   | 4.47 | 3.477 | 0.747 | 1.525 | 0.060 | 1.200 | 0.034 | 1.291 | 0.189 |
|   | 8.91 | 0.701 | 0.244 | 0.485 | 0.032 | 0.388 | 0.021 | 0.395 | 0.106 |
|   | 17.78 | 0.371 | 0.062 | 0.155 | 0.022 | 0.102 | 0.016 | 0.120 | 0.046 |
|   | 35.48 | 0.100 | 0.033 | 0.038 | 0.007 | 0.004 | 0.007 | 0.021 | 0.014 |

Note. Here \(r\) is the comoving distances in units of \(h^{-1}\) Mpc; \(\xi\) is the two-point correlation function for a volume-limited sample; \(\Delta \xi\) is the 1\(\sigma\) error of \(\xi(s)\) estimated from 10 mock samples. The autocorrelations of blue and red galaxies, and the cross-correlations between blue and red galaxies, are shown in the upper, middle, and lower parts, respectively.
We have shown, though, that the one-dimensional 2PCF, $\xi(s)$, inferred directly from the reconstructed real space is not significantly affected, with deviations typically being smaller than the uncertainties arising from cosmic variance (at least for an SDSS-like survey) for galaxies brighter than $M_r = 5 \log h - 19.0$. In fact, over the range of scales $0.2 \ h^{-1} \text{Mpc} \lesssim r \lesssim 20 \ h^{-1} \text{Mpc}$, the average error on the reconstructed real-space 2PCF is less than 5%. Hence, our method is capable of correcting redshift distortions in redshift surveys to a level that allows for an accurate, unbiased measurement of the real-space correlation function.

We have applied our reconstruction method to the SDSS DR7, giving a real-space version of the main galaxy catalog, which contains 396,068 galaxies in the NGC with redshifts in the range $0.01 \lesssim z \lesssim 0.12$. This real-space galaxy catalog is publicly available at http://gax.shao.ac.cn/data/data1/SDSS_REAL/SDSS7_REAL.tar.gz. We emphasize that the FOG correction is only statistical in nature, and that the line-of-sight positions of satellite galaxies in the catalog have been assigned at random, in accordance with our assumption that satellite galaxies are an unbiased tracer of the mass distribution of their host halo.

Using the reconstructed real-space data we have shown that the Sloan Great Wall, the largest known structure in the universe, is not as dominant an overdense structure as it appears in redshift space, but that its apparent overdensity is strongly enhanced by the Kaiser effect. We have measured the 2PCFs in reconstructed real space in different absolute magnitude bins. They all deviate clearly from a simple power law, revealing a clear one-halo to two-halo transition. A comparison with the corresponding 2PCFs in redshift space nicely demonstrates how RSDs boost the correlation on large scales (by about 40%–50% on a scale of $10 \ h^{-1} \text{Mpc}$), while suppressing it on small scales (by about 70%–80% on a scale of $0.3 \ h^{-1} \text{Mpc}$). We have also measured the real-space autocorrelation functions of blue and red galaxies, and their cross-
correlations. Using the real-space (color-dependent) $\xi(s)_r$, we have investigated how the bias factor depends on galaxy luminosity and color, and how our method provides more reliable measurements of galaxy bias factors than the traditional method that uses the projected 2PCF, $w_p(r_p)$.

The present paper, the first in a series, is focused on the methodology. In a forthcoming paper we will use our reconstructed, real-space SDSS galaxy catalog to study in more detail how the real-space clustering of galaxies depends on their intrinsic properties, such as luminosity, stellar mass, color, and star formation rate. We will also use our reconstruction method to put constraints on cosmological parameters as well as halo occupation models. As briefly mentioned in Section 5.2, the actual reconstruction is cosmology-dependent, because the bias parameter $b_{hm}$, the halo masses assigned to galaxy groups, and the distance-redshift relation are all cosmology-dependent. Consequently, assuming an incorrect cosmology can result in systematic errors in our reconstruction and distortions in the correlation functions. We can then model such distortions and constrain cosmological parameters by searching for the model that gives the best reconstructed real space, so that $\xi(r_p, r_s)$ is isotropic (i.e., the quadrupole-to-monopole ratio is close to zero).

We thank the anonymous referee for helpful comments that greatly improved the presentation of this paper. This work is supported by the 973 Program (No. 2015CB857002), national science foundation of China (grant Nos. 11123806, 11121062, 11233005, 11073017), NCET-11-0879, the Strategic Priority Research Program “The Emergence of Cosmological Structures” of the Chinese Academy of Sciences, Grant No. XDB09000000 and the Shanghai Committee of Science and Technology, China (grant No. 12ZR1452800). We also thank the support of a key laboratory grant from the Office of Science and Technology, Shanghai Municipal Government (No. 11DZ2260700). H.J.M. would like to acknowledge the support of NSF AST-1516962 and the support of a key laboratory grant from the Office of NSF AST-1517528, and F.v.d.B. is supported by the US National Science Foundation through grant AST 11233005, 11073017.

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Ata, M., Kitsura, F.-S., Chuang, C.-H., et al. 2016, arXiv:1605.09745
Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681
BICEP2/Keck Collaboration and Planck Collaborations 2015, PrRvL, 114, 101301
Blake, C., Joudaki, S., Heymans, C., et al. 2016, MNRAS, 456, 2806
Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, AJ, 129, 2562
Branchini, E., Davis, M., & Norberg, A. 2012, MNRAS, 424, 472
Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, MNRAS, 394, 929
Cacciato, M., van den Bosch, F. C., More, S., Ho, M. J., & Yang, X. 2013, MNRAS, 430, 767
Campbell, D., van den Bosch, F. C., Hearin, A., et al. 2015, MNRAS, 452, 444
Cole, S., Fisher, K. B., & Weinberg, D. H. 1994, MNRAS, 267, 785
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
Davis, M., & Peebles, P. J. E. 1983, ApJ, 267, 465
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44