ELUCID. V. Lighting Dark Matter Halos with Galaxies

Xiaohu Yang1,2, Youcai Zhang3, Huiyuan Wang4,5, Chengze Liu1, Tianhuan Lu6, Shijie Li1, Feng Shi3,4, Y. P. Jing1,2, H. J. Mo7,8, Frank C. van den Bosch9, Xi Kang10, Weiguang Cui11, Frank C. van den Bosch9, Xi Kang10, Weiguang Cui11

1 Department of Astronomy, Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China
2 IFSA Collaborative Innovation Center, and Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China
3 Shanghai Astronomical Observatory, Nandan Road 80, Shanghai 200030, People’s Republic of China
4 Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei, Anhui 230026, People’s Republic of China
5 School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, People’s Republic of China
6 Zhiyuan College, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China
7 Department of Astronomy, University of Massachusetts, Amherst MA 01003-9305, USA
8 Astronomy Department and Center for Astrophysics, Tsinghua University, Beijing 10084, People’s Republic of China
9 Astronomy Department, Yale University, P.O. Box 208101, New Haven, CT 06520-8101, USA
10 Purple Mountain Observatory, the Partner Group of MPI für Astronomie, 2 West Beijing Road, Nanjing 210008, People’s Republic of China
11 Departamento de Física Teórica, Módulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

Received 2017 December 3; revised 2018 April 28; accepted 2018 May 3; published 2018 June 8

Abstract

In a recent study, using the distribution of galaxies in the north galactic pole of the SDSS DR7 region enclosed in a 500 h−1 Mpc box, we carried out our ELUCID simulation (ELUCID III). Here, we light the dark matter halos and subhalos in the reconstructed region in the simulation with galaxies in the SDSS observations using a novel neighborhood abundance matching method. Before we make use of the galaxy–subhalo connections established in the ELUCID simulation to evaluate galaxy formation models, we set out to explore the reliability of such a link. For this purpose, we focus on the following few aspects of galaxies: (1) the central–subhalo luminosity and mass relations, (2) the satellite fraction of galaxies, (3) the conditional luminosity function (CLF) and conditional stellar mass function (CSMF) of galaxies, and (4) the cross-correlation functions between galaxies and dark matter particles, most of which are measured separately for all, red, and blue galaxy populations. We find that our neighborhood abundance matching method accurately reproduces the central–subhalo relations, satellite fraction, and the CLFs, CSMFs, and biases of galaxies. These features ensure that galaxy–subhalo connections thus established will be very useful in constraining galaxy formation processes. We provide some suggestions for the three levels of using the galaxy–subhalo pairs for galaxy formation constraints. The galaxy–subhalo links and the subhalo merger trees in the SDSS DR7 region extracted from our ELUCID simulation are available upon request.

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

1. Introduction

To fully model structure formation in the universe and to probe detailed galaxy formation processes, one needs to have a fair sampling of the universe with sufficiently large volume and high resolution. Thanks to large redshift surveys, such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), we are now able to carry out such investigations to unprecedented accuracy. However, in order to make full use of the potential of the observational data, one still has to develop or make use of optimal strategies. One of the most efficient ways of using observational data is to carry out constrained simulations where the initial density field is indeed extracted from observations, which is the basic idea of our ELUCID project.

Along this line, numerous attempts have been made to develop methods to reconstruct the initial conditions of structure formation in the local universe using galaxy distributions and/or peculiar velocities (Sousa et al. 2007; Jasche & Wandelt 2013; Jasche et al. 2015; Sousa et al. 2015; Seljak et al. 2017). Hoffman & Ribak (1991) developed a method to construct Gaussian random fields that are subject to various constraints (see also Bertschinger 1987; van de Weygaert & Bertschinger 1996; Klypin et al. 2003; Kitaura & Enßlin 2008). Klypin et al. (2003) improved this method by using the Wiener filter (see, e.g., Zaroubi et al. 1995) to deal with sparse and noisy data. Gaussian density fields constrained by the peculiar velocities of galaxies in the local universe have also been used to set up the initial conditions for constrained simulations (e.g., Kravtsov et al. 2002; Klypin et al. 2003; Gottloeber et al. 2010). Note, however, that the basic underlying assumption in this method is that linear theory is valid on all scales (Klypin et al. 2003; Doumler et al. 2013).

In a recent paper, Wang et al. (2014, hereafter ELUCID I) developed a method combining the Bayesian reconstruction approach with a much more accurate dynamic model of structure evolution, the Particle Mesh (PM) model. The PM technique has been commonly adopted in N-body codes to evaluate gravitational forces on relatively large scales (see, e.g., Klypin & Shandarin 1993; White et al. 1983; Jing & Suto 2002; Springel 2005) and can follow the structure evolution accurately as long as the chosen grid cells and time steps are sufficiently small. Tests show that this method can achieve a much higher reconstruction accuracy than any other method in the literature. To apply this method to observation, one needs to reconstruct the cosmic density field of the local universe. As illustrated in Wang et al. (2009), this density field can be fairly well reconstructed using the distribution of (relatively massive) galaxy groups (i.e., dark matter halos). Using the galaxy groups
(Yang et al. 2007, 2012) extracted from SDSS Data Release 7 (Abazajian et al. 2009, DR7), Wang et al. (2012) obtained the mass, tensor, and velocity fields of the local universe in the SDSS DR7 region.

With all of these preparations, in a recent study, Wang et al. (2016, hereafter ELUCID III) made use of the density field reconstructed from the north galactic pole of the SDSS DR7 region with an improved domain mass assignment method (Wang et al. 2013) to predict the evolution of the structure of the local universe enclosed in a 500 $h^{-1}$ Mpc length cubic box. As shown in ELUCID I, the reconstruction can recover more than half of the phase information down to a scale $k \sim 3.4$ $h$Mpc$^{-1}$ at $z = 0$. Tests using original and reconstructed simulations show that more than half of the halos with mass $\geq 10^{13.5} M_{\odot}$, wherein more than 50% of the particles are in common with the counterpart halos in the original simulation (see Tweed et al. 2017, hereafter ELUCID II), can be reliably reproduced. These features indicate that the halos formed in the SDSS DR7 region in our ELUCID simulation have a large-scale environment roughly consistent with that of the true universe, and the evolution of massive halos can be modeled quite well.

In this paper, we propose a novel neighborhood abundance matching method to link galaxies observed in the SDSS DR7 region to halos/subhalos in our ELUCID simulation in the same local small volumes. Once the galaxy–subhalo connections are generated, we can use them to constrain semi-analytical galaxy formation models (SAMs) in halo-based and local-environment-based apples-to-apples comparisons. Note that, technically, we can perform the neighborhood abundance matching between galaxy groups and dark matter halos as well. However, because of the following reasons, we decide to use galaxies rather than groups. The main reason is that since some massive groups in the observation may overpredict the galaxy population in one halo and underpredict it in another. The second reason is that since only massive groups can be well reproduced in the ELUCID simulation, we are not able to use the group–halo matching in individual low-mass halos. In addition, it would be interesting to see the impact of interlopers in galaxy groups, which is not available in group–halo matching (see Campbell et al. 2015 for related discussions).

The structure of this paper is organized as follows. Section 2 gives a detailed description of the data we used in this study, including the halos/subhalos extracted from the ELUCID simulations, the SDSS DR7 galaxy catalog, as well as the neighborhood abundance matching between galaxies and dark matter subhalos. In Section 3, we probe the central–subhalo relations, satellite fraction, and conditional luminosity functions (CLFs) and conditional stellar mass functions (CSMFs) of galaxies. In Section 4, we measure the cross-correlation functions between galaxies and dark matter particles. In Section 5, we provide some suggestions for the use of the galaxy–subhalo connections established in this work. Finally, we summarize our results in Section 6. Throughout the paper, we adopt a $\Lambda$CDM cosmology with parameters that are consistent with the fifth-year data release of the WMAP mission (hereafter, the WMAP5 cosmology): $\Omega_b = 0.258$, $\Omega_\Lambda = 0.742$, $\Omega_m = 0.044$, $h = H_0/(100 \text{ km s}^{-1}\text{Mpc}^{-1}) = 0.72$, and $\sigma_8 = 0.80$ (Dunkley et al. 2009).

2. Data

2.1. The Halos/Subhalos in the ELUCID Simulation

In this study, we use dark matter halos/subhalos extracted from the ELUCID simulation. This simulation, which evolves the distribution of $3072^3$ dark matter particles in a periodic box $500 ~ h^{-1}$ Mpc on a side, was carried out in the Center for High Performance Computing, Shanghai Jiao Tong University. The simulation was run with L-GADGET, a memory-optimized version of GADGET2 (Springel 2005). The cosmological parameters adopted in this simulation are consistent with the WMAP5 results, with each particle having a mass of $3.0875 \times 10^8 h^{-1} M_{\odot}$. In our ELUCID simulation, we make use of the mass density field extracted from the galaxy/group distribution in the north galactic pole region of the SDSS DR7 to constrain the initial conditions using a Hamiltonian Markov Chain Monte Carlo method (HMCMC) with PM dynamics (see ELUCID I and ELUCID III for details). As an illustration, we show in Figure 1 the distributions of the related galaxies and the reconstructed and simulated mass density fields. Shown in the left-hand panel is a slice of the galaxy distributions in the north galactic pole of SDSS DR7. In the middle panel is a slice of the density field constructed from these galaxy distributions. In the right panel is the mass density field revealed (evolved to redshift $z = 0$) in our ELUCID simulation. The region in the left panel enclosed by solid lines is the density field reproduced in our ELUCID simulation. The basic properties of our ELUCID simulation, including the algorithm to perform the simulation, as well as the output power spectrum, halo mass functions, etc., can be found in ELUCID III.

From the ELUCID simulation, dark matter halos were first identified using a friends-of-friends (FOF) algorithm with a linking length 0.2 times the mean particle separation and containing at least 20 particles. The dark matter halo mass function of this simulation at redshift $z = 0$ is checked and agrees very well with the model predictions of Sheth et al. (2001) and Tinker et al. (2008). Based on halos at different outputs, we first use the SUNFIND algorithm (Springel et al. 2001) to identify bound substructures in the FOF halos. The most massive substructure in an FOF halo is considered to be the main halo of this FOF, and all the other subhalos in this FOF are called subhalos. For a given subhalo or main halo, each particle is assigned a weight, which decreases with the binding energy. We then find all main halos and subhalos in the subsequent snapshot that contain some of these particles. We count these particles, weighting them for potential descendants. The candidate with the highest weighted count is selected as the descendant. Please see Springel et al. (2005) for the details.

In order to properly link galaxies, especially satellite galaxies, to dark matter halos and subhalos, one needs to properly treat the subhalos in the simulations (see Jiang & van den Bosch 2016 for related discussions). For a widely adopted subhalo population in SHMs, the mass or circular velocity of the surviving subhalos is extracted by subfinders in the simulation, but their masses/velocities are updated to the maximum values along their accretion histories (e.g., Conroy et al. 2006; Hearin et al. 2013). This method implies that (1) each subhalo can form only one galaxy, (2) the central–host halo relation does not evolve significantly with redshift, and (3) satellite galaxies are disrupted whenever subhalos can no longer be identified in the simulation, either because of limiting mass resolution or because the subhalo is tidally disrupted. An
alternative method in the SHAM is to separate the central and satellite galaxies, i.e., one can perform abundance matching separately for central galaxies versus main halos and for satellite galaxies versus subhalos (see Rodríguez-Puebla et al. 2015 for a similar attempt). In this regard, the evolution of satellite galaxies will be automatically taken into account. In this study, we will perform these two types of abundance matching methods and compare the differences between them. We refer to the former as “Match1” and to the latter as “Match2”.

In order to match the galaxies in the SDSS observations, we first rotate the simulation box so that the resimulated density field is superposed on the SDSS observation region. We then discard the dark matter subhalos that are outside the survey sky coverage region used for our density field reconstruction. The R.A., decl., and $z_{\text{com}}$ of the subhalos are calculated from their real space positions. Then, their final redshifts are obtained by adding the peculiar velocities, with $z_{\text{obs}} = z_{\text{com}} + v_{\text{pec}}/c$. We trim subhalos within the redshift range $0 < z < 0.12$ for our subsequent matching with galaxies in observations.

2.2. SDSS DR7

The galaxy catalog we used for finding galaxy groups, reconstructing density fields, and performing the ELUCID simulation is the New York University Value-Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005). The catalog is compiled based on SDSS DR7 (Abazajian et al. 2009) but with an independent set of significantly improved reductions. From the NYU-VAGC, we select all galaxies in the Main Galaxy Sample with an 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 $C_r > 0.7$. This gives a sample of 639,359 galaxies with a sky coverage of 7748 square degrees. In this study, we make use of all the galaxies in this sample for our investigation. Within these 639,359 galaxies, 35,678 do not have spectroscopic redshifts and are assigned the redshifts of their nearest neighborhoods.

According to Yang et al. (2007, hereafter Y07), the absolute magnitudes of galaxies in bandpass $Q$ are computed using

\[
0.1M_Q - 5 \log h = m_Q + \Delta m_Q - DM(z) - K_Q - E_Q.
\]  

Here, $DM(z) = 5 \log [D_L/(h^{-1}\text{Mpc})] + 25$ is the bolometric distance modulus calculated from the luminosity distance $D_L$ using a WMAP5 cosmology. $\Delta m_Q$ is the latest zero-point correction for the apparent magnitudes, which converts the SDSS magnitudes to the AB system, and for which we adopt $\Delta m_Q = (-0.036, +0.012, +0.010, +0.028, +0.040)$ for $Q = (u, g, r, i, z)$. All absolute magnitudes are $K+E$-corrected to $z = 0.1$. For the $K$ corrections, we use the latest version of $K$correct (v4) described in Blanton et al. (2003; see also Blanton & Roweis 2007), which we apply to all galaxies that have meaningful magnitudes and redshifts, including those that have redshifts from alternative sources and those that have been assigned the redshift of their nearest neighbor. Finally, the evolution corrections to $z = 0.1$ are computed using $E_Q = A_Q(z - 0.1)$, with $A_Q = (-4.22, -2.04, -1.62, -1.61, -0.76)$ for $Q = (u, g, r, i, z)$ (see Blanton et al. 2003). Note that these evolution corrections imply that galaxies were brighter in the past (at higher redshifts).

In addition to the absolute magnitudes, we also compute for each galaxy its stellar mass, $M_\star$. Using the relation between the stellar mass-to-light ratio and the color of Bell et al. (2003), we obtain

\[
\log \left( \frac{M_\star}{h^{-2} M_\odot} \right) = -0.306 + 1.097 \left[ 0.0(g - r) \right] - 0.1
\]

\[
- 0.4( 0.0M_r - 5 \log h - 4.64).
\]

Here, $0.0(g - r)$ and $0.0M_r - 5 \log h$ are the $(g - r)$ color and $r$-band magnitude $K+E$-corrected to $z = 0.0$. 4.64 is the $r$-band magnitude of the Sun in the AB system (Blanton & Roweis 2007), and the $-0.10$ term effectively implies that...
we adopt a Kroupa (2001) IMF (Borch et al. 2006). For a small fraction (about 2%) of all galaxies, the $g - r$ color that results from the photometric SDSS pipeline is unreliable. These galaxies typically have $g - r$ colors that are clearly unrealistic (they are catastrophic outliers in the color–magnitude distribution). If this is not accounted for, Equation (2) assigns these galaxies stellar masses that are unrealistically high or low. To account for these outliers, we follow Y07 using the color bi-Gaussian distributions of galaxies obtained in Yang et al. (2008; see also Li et al. 2006). For any galaxy that falls outside the 3$\sigma$ ranges of the mean color–magnitude relations of both the red sequence and the blue cloud (about 2% of all galaxies), we compute its stellar mass using the mean color of either the red sequence (when the galaxy is too red) or the blue cloud (when the galaxy is too blue).

In order to probe the color dependence of galaxies, following Yang et al. (2008), we separate our galaxies into red and blue subsamples using the criterion

$$0.1 (g - r) = 1.022 - 0.0651x - 0.00311x^2,$$

where $x = 0.1M_r - 5 \log h + 23.0$.

From the above galaxy catalog, we only select galaxies within groups that were used to map the density field and thus the initial density field in our ELUCID simulation. That is, we only select galaxies that are located within the range $99^\circ < R.A. < 283^\circ$, $-7^\circ < \text{decl.} < 75^\circ$ (i.e., in the north galactic pole) and redshift $0.01 < z < 0.12$. After this selection, a total of 396,069 galaxies remain for our subsequent probes.

### 2.3. The Neighborhood Abundance Matching Method

In order to link galaxies to dark matter (sub)halos, one can use either the halo occupation distribution (HOD) or CLF models (e.g., Jing et al. 1998; Peacock & Smith 2000; van den Bosch et al. 2003; Yang et al. 2003, 2012; Tinker et al. 2005; Zheng et al. 2005; Cooray 2006; van den Bosch et al. 2007; Brown et al. 2008; Cacciato et al. 2009; More et al. 2009; Avila-Reese & Firmani 2011; Leauthaud et al. 2011; Neistein et al. 2011; Rodríguez-Puebla et al. 2015, 2017; Li et al. 2016; Zu & Mandelbaum 2016; Bull 2017; Cohn 2017; Contreras et al. 2017), or subhalo abundance matching processes (e.g., Vale & Ostriker 2004; Conroy et al. 2006, 2009; Shankar et al. 2006; Vale & Ostriker 2006; Behroozi et al. 2010; Guo et al. 2010; Moster et al. 2010; Hearin et al. 2013). These probes have revealed many observational features of galaxies and were widely used to constrain the galaxy formation models. In these studies, however, only the global properties of galaxies, such as stellar mass functions/luminosity functions, and clustering are used to establish the galaxy–(sub)halo connections. In our probe, as the structures in our ELUCID simulation are supposed to trace the evolution of real structures in the SDSS DR7 region we are considering, we set out to match galaxies with the dark matter subhalos in their neighborhood, i.e., using a neighborhood abundance matching method.

We first sort the stellar masses of the galaxies. Starting from the most massive galaxy, in a small volume of its neighborhood, we search for the most likely subhalo in the redshift space of each subhalo sample; this is then marked as the galaxy’s counterpart. The likelihood of the subhalo to be linked to the candidate galaxy is modeled as follows,

$$P(r_p, \pi, M_{\text{sh}}) = M_{\text{sh}} \exp\left(-\frac{r_p^2}{2\sigma_{\text{off}}^2}\right) \exp\left(-\frac{\pi^2}{2\sigma_{\text{off}}^2}\right).$$

Here, $r_p$ and $\pi$ are the separations between the galaxy and subhalo in the perpendicular and along the line-of-sight directions, respectively. $M_{\text{sh}}$ is the mass of the subhalo under consideration, while $\sigma_{\text{off}}$ and $\sigma_{\text{off}}$ are the two free parameters we choose for our neighborhood abundance matching. In the extreme case where $\sigma_{\text{off}} = \infty$ and $\sigma_{\text{off}} = \infty$, the neighborhood abundance matching method degrades to the traditional abundance matching method. We use two sets of parameters to perform our neighborhood abundance matching: (1) $r_{\text{off}} = 2.5$ $h^{-1}$ Mpc and $\pi_{\text{off}} = 500$ km s$^{-1}$ and (2) $r_{\text{off}} = 5.0$ $h^{-1}$ Mpc and $\pi_{\text{off}} = 1000$ km s$^{-1}$, and compare their performances. Note that in our ELUCID simulation the reconstructed density field is smoothed using a Gaussian kernel of radius $2.0$ $h^{-1}$ Mpc. Here, these two sets of choices are made according to a compromise of the scatters and separations between the galaxies and subhalos, which can be illustrated as follows. Using these criteria, we sequentially search for the counterparts of all galaxies in the redshift space within a maximum distance $\leq 30$ $h^{-1}$ Mpc. For the total of 396,069 galaxies, according to criterion (2), there are 296,488 central galaxies that are linked to the main halos and 99,581 satellite galaxies that are linked to the subhalos for the Match1 method. Criterion (1) gives very similar numbers, with typical differences of a few hundred. Compared to the numbers specified in the group catalog, 277,139 centrals and 118,930 satellites, the Match1 method roughly underestimated the satellite galaxy population by about $\sim 20\%$. On the other hand, by definition, the Match2 method will give the same central/satellite separation as that in the group catalog.

We show in Figure 2 the distributions of the separation of the galaxy–subhalo pairs in different subhalo mass bins. Shown in the upper-left panel are the $r_p$ distributions for matching criterion (1): $r_{\text{off}} = 2.5$ $h^{-1}$ Mpc and $\pi_{\text{off}} = 500$ km s$^{-1}$. Note that as the results for the Match1 and Match2 methods are very similar, here we only show those obtained from the Match1 method. As we can see, the offset between galaxies and subhalos in the most massive mass bin with log $M_h \geq 14.0$, which peaks at $\sim 0.5$ $h^{-1}$ Mpc, is the smallest. About 50% of the matched pairs have projected separations less than 2 $h^{-1}$ Mpc. The lower-mass subhalos have slightly larger separations, and the distributions peak at about 1.5–2.0 $h^{-1}$ Mpc. About 50% of the matched pairs have projected separations less than 2.5 $h^{-1}$ Mpc. Shown in the upper-right panel are the $\pi$ distributions of the galaxy–subhalo pairs for matching criterion (1). The offsets for subhalos in different mass bins are quite similar. All of the distributions peak at $\sim 50$ km s$^{-1}$, and about 50% of the matched pairs have line-of-sight separations less than 200 km s$^{-1}$.

The results shown in the middle panels of Figure 2 are similar to those shown in the upper panels, but for matching criterion (2): $r_{\text{off}} = 5.0$ $h^{-1}$ Mpc and $\pi_{\text{off}} = 1000$ km s$^{-1}$. The overall distribution properties are quite similar to those of matching criterion (1), but with slightly larger offsets. That is, about 50% of the matched pairs have projected separations less
than 3.0 $h^{-1}$ Mpc and line-of-sight separations less than 250 km s$^{-1}$.

As we have matched galaxies with subhalos, it is quite straightforward to check their luminosity (stellar mass)–subhalo mass relations. Shown in the upper panels of Figure 3 are the luminosity–subhalo mass (left panel) and stellar mass–subhalo mass (right panel) relations for our matching criterion (1). Shown in the middle panels are the results for our matching criterion (2). Here again, we only show results obtained from the Match1 method; the results obtained from Match2 are very similar. In each panel, the unfilled squares with error bars indicate the median and 68% confidence levels of these relations for all galaxies. Comparing the results for the two matching criteria, the latter shows somewhat tighter luminosity (stellar mass)–subhalo mass relations, especially in the low-mass subhalo. We thus think the latter matching criterion works better. In what follows, we only present the results obtained using matching criterion (2), i.e., with $r_{\text{off}} = 5.0$ $h^{-1}$ Mpc and $v_{\text{off}} = 1000$ km s$^{-1}$.

Before we proceed to provide more detailed tests of the performance of our neighborhood abundance matching method in the ELUCID simulation, it would be interesting to check the above separation distribution and luminosity (stellar mass)–subhalo mass relations when the neighborhood abundance matching method is applied to a simulation that does not have good correspondence with the observation. For this purpose, we rotate the ELUCID simulation box by 90 degrees and shift it by 250 $h^{-1}$ Mpc, and then perform the same procedures using matching criterion (2). Note that after such a treatment, the simulation density field is no longer matched with the SDSS density field. Shown in the lower panels of Figures 2 and 3 are the resulting separation distributions and luminosity (stellar mass)–subhalo mass relations. We see that the separation distributions of galaxy–subhalo pairs in this situation are very different from our fiducial case, especially for massive clusters.
The very large separation between galaxies and massive (sub)halos indicates that the galaxy–subhalo pairs might have different origins. For small (sub)halos with mass \( \lesssim 10^{12.0} h^{-1} M_{\odot} \), the difference is quite small, indicating that the low-mass galaxy–subhalo pairs, even in the ELUCID simulation, might be dominated by Poisson errors (see Tweed et al. 2017).

In addition, the luminosity (stellar mass)–subhalo mass relations in this case are much worse, i.e., with much larger scatters, than those of our fiducial cases. In general, as we mentioned, if we set \( r_{\text{eff}} = \infty \) and \( v_{\text{eff}} = \infty \), the neighborhood abundance matching method will degrade to the traditional abundance matching method, which will provide monotonic luminosity (stellar mass)–subhalo mass relations.

### 3. The Halo Occupation Distributions of Galaxies

After we matched galaxies with subhalos in the ELUCID simulation, we proceed to measure a few galaxy statistics within host halos of different masses. These statistics are compared to those obtained from galaxy groups (e.g., Yang et al. 2008, 2009) to demonstrate the feasibility of populating dark matter subhalos with observed galaxies via the neighborhood abundance matching method outlined in Section 2.3.

#### 3.1. The Galaxy–Subhalo Luminosity–Mass Relations

Here we start our probing by using the data obtained from the Match2 method. Shown in the upper and lower panels of Figure 4 are the luminosity (stellar mass)–subhalo mass relations for satellite and central galaxies separately. We follow Yang et al. (2008) in using the following \( L–M_{\text{sh}} \) functional form to describe the median luminosity–subhalo mass relation,

\[
L = L_0 \left( \frac{M_{\text{sh}}/M_\bullet}{1 + M_{\text{sh}}/M_\bullet} \right)^{\alpha + \beta}.
\]
This model contains four free parameters: a normalized luminosity, $L_0$, a characteristic halo mass, $M_1$, and two slopes, $\alpha$ and $\beta$. The blue and red solid lines shown in the left panels are the best fits to the average $L$–$M_{\text{sh}}$ relations for satellite and central galaxies, respectively. The best-fitting parameters are listed in Table 1 in the first and second rows. Although not very significant, we do see some differences between the luminosity–subhalo mass relations of central and satellite galaxies, indicating that satellite galaxies may experience different stripping or disruption effects from subhalos.

For comparison, we also show in the lower-left panel the best-fit results obtained by Yang et al. (2008) using a dashed line, where the set of best-fit parameters are listed in the third row of Table 1. This set of results is obtained from the SDSS galaxy group catalog directly, in which (i) it is assumed that central galaxies are the brightest group members, and (ii) halo masses are inferred by abundance-matching the host halos to the total stellar mass of the groups. The dotted line in that panel shows the CLF model constraints obtained by Cacciato et al. (2013) using the clustering and weak-lensing measurements of galaxies. Overall, our neighborhood abundance matching method gives an $L$–$M_{\text{sh}}$ relation quite consistent with these previous measurements, except that of Cacciato et al. (2013). The slight systematic deviation from Cacciato et al. (2013) is mainly caused by the different definition of halo mass and the cosmology they used. As shown in Figure 7 of Cacciato et al. (2013), if the halo mass definition and cosmology are properly converted, their results agree very well with those obtained by Yang et al. (2008).
For the $M_h$–$M_{sh}$ relations shown in the right panels, we use a similar function to fit the data:

$$M_h = M_0 \left( \frac{M_{sh}}{M_1} \right)^{\alpha + \beta} \left( 1 + \frac{M_{sh}}{M_1} \right)^{\gamma}. \quad (6)$$

The blue and red solid lines shown in the right panels are the best fits of this model to the data for the satellite and central galaxies, respectively, where the best-fit parameters are listed in the fourth and fifth rows of Table 1.

For comparison, we also show in the lower-right panel, using a dashed line, the model constraints obtained by Yang et al. (2012), where the related set of parameters is listed in the sixth row of Table 1. The model constraints obtained by Moster et al. (2013), Behroozi et al. (2013), and Rodríguez-Puebla et al. (2015) are shown as the dotted, long dashed, and dotted–dashed lines, respectively. Here again, we see that our neighborhood abundance matching method gives an $M_h$–$M_{sh}$ relation quite consistent with these previous probes, except that of Behroozi et al. (2013), which is somewhat lower, especially at massive end. The difference is mainly caused by a different stellar mass estimation method adopted (see Behroozi et al. 2013 for a similar trend and related discussions).

3.2. The Satellite Fraction

The second quantity we probe is the satellite fraction of galaxies. Since a satellite galaxy resides in a halo more massive than a central galaxy of the same luminosity or stellar mass (e.g., Yang et al. 2003), the fraction of satellite galaxies as a function of luminosity, $f_{sat}(L)$, or stellar mass, $f_{sat}(M_h)$, thus plays an important role in modeling both the small- and large-scale clustering of galaxies of a given luminosity/stellar mass (e.g., Jing et al. 1998; Berlind & Weinberg 2002; Yang et al. 2003; van den Bosch et al. 2007; Yang et al. 2012). The satellite fraction as function of luminosity, $f_{sat}(L)$, is also important for a proper interpretation of the measurements of galaxy–galaxy lensing signals (e.g., Guzik & Seljak 2002; Mandelbaum et al. 2006; Yang et al. 2006) and pairwise velocity dispersion of galaxies (e.g., Jing & Börner 2004; Yang et al. 2004), and to understand the quenching of galaxies (e.g., Bluck et al. 2016; Wang et al. 2017).

Here, we estimate $f_{sat}(L)$ and $f_{sat}(M_h)$ directly from our matched galaxy–subhalo pairs. In the left-hand panels of Figure 5, we show $f_{sat}(L)$ as a function of galaxy luminosity. The results are plotted separately for all (upper panels), red, and blue (lower panels) galaxies. Since in our Match2 method the central and satellite galaxies are matched with the main halos and subhalos, respectively, by definition, the satellite fractions

![Figure 5](image-url)
in our Match2 method follow those of SDSS galaxy groups. We show the resulting satellite fraction in Figure 5 using solid dots with error bars. Note that, in our probe, we have made use of the modelIC sample in Yang et al. (2007), where about 5% of galaxies that lack spectroscopic redshifts due to the fiber collision effect are assigned the redshifts from their nearest neighbors. As pointed out in Yang et al. (2008), fiber collisions are expected to significantly impact the number of close pairs and hence the satellite fractions \( f_{\text{sat}}(L) \). The typical uncertainties induced by adding or removing the fiber collision galaxies are about 5%. Here we adopt this uncertainty value as the error bars shown in Figure 5.

First, for all galaxies, by comparing the model predictions of the Match1 method with those of the Match2 method, we can see that the Match1 method predicts roughly consistent satellite fractions for relatively bright galaxies. However, if one goes to fainter galaxies with \( 0.2 M_\odot c^{-1} h^{-1} \), the satellite fractions are significantly underestimated. This discrepancy indicates that the widely used subhalo abundance matching method in the literature may not predict the low-mass satellite galaxies correctly. One can either add more subhalos (e.g., the disrupted subhalos) in their abundance matching with galaxies, or match central and satellite galaxies separately, as we did here with the Match2 method.

Next, for galaxies that are separated into red and blue populations, the model predictions for the Match1 method show much smaller segregations between red and blue galaxies compared to those measured from the galaxy groups or using the Match2 method. Note that since in our neighborhood abundance matching procedures we did not perform any special treatments for the red and blue galaxies, the lack of segregation in the Match1 method is somewhat expected. In general, one may treat red and blue galaxies differently to have a better model prediction of red/blue satellite fractions (e.g., Rodríguez-Puebla et al. 2015), or more straightforwardly, by matching central and satellite galaxies separately.

The satellite fraction as a function of galaxy stellar mass is shown in the right panels of Figure 5. The overall behaviors are quite similar to those shown in the left panels.

Finally, as a comparison, we also show in the lower panels of Figure 5 the satellite fraction obtained by Mandelbaum et al. (2006) for early- and late-type galaxies (open squares with 95% confidence level error bars) from the galaxy—galaxy weak-lensing measurements. Although their samples are defined differently from ours (early and late types according to galaxy morphologies, versus red and blue galaxies according to color), the two measurements agree very well.

### 3.3. The Conditional Luminosity Functions

The CLF of galaxies in dark halos, \( \Phi(L|M) \), which describes the average number of galaxies as a function of galaxy luminosity in a dark matter halo of a given mass, plays an important role in our understanding of how galaxies form in dark matter halos (e.g., Yang et al. 2003, 2012 and references therein). Here we directly measure \( \Phi(L|M) \) from our matched galaxy—subhalo pairs, and compare them to those obtained from the galaxy group catalogs. In order to make proper comparisons, we updated the halo masses of galaxy groups according to the WMAP5 cosmology adopted in this study.

The CLF can be estimated by directly counting galaxies in halos and groups. However, since the galaxies used for our study are flux limited to \( r = 17.72 \), for a given galaxy luminosity \( L \), there is a limiting redshift, \( z_L \), beyond which galaxies with such a luminosity are not included in the sample. In order to estimate \( \Phi(L|M) \) at a given \( L \), we only use halos and groups that are complete to the redshift limit \( z_L \). The CLF is obtained by simply counting the average number of galaxies (in luminosity bins) in halos or groups of a given \( M \). We show in Figure 6 the resulting CLFs obtained from galaxy groups of different masses using symbols with error bars, where the error bars are obtained using 200 bootstrap resamplings of the groups. The contributions of the central and satellite galaxies are plotted separately using filled and unfilled symbols, respectively. The solid lines shown in Figure 6 are the results obtained for our fiducial Match2 method. For comparison, we also show using dotted lines the results obtained for the Match1 method. First, for the central galaxy component, we see that both Match2 and Match1 methods give very similar predictions. According to the comparisons with the data points alone, we see that both Match2 and Match1 methods only agree with data in the most massive bin. In the other three halo mass bins, the CLFs of central galaxies show significant deviations. On the other hand, however, if we model the CLFs for central galaxies with a lognormal distribution (e.g., Yang et al. 2008),

\[
\Phi_{\text{cent}}(L_c|M) = \frac{1}{\sqrt{2\pi} \sigma_c} \exp \left[ -\frac{(\log L_c - \log L)^2}{2\sigma_c^2} \right], \tag{7}
\]

where \( L \) is the peak luminosity and \( \sigma_c \), the lognormal scatter, the discrepancies are indeed not that significant. The two methods both predicted the correct peak luminosities, \( L \), of the central galaxies, although the lognormal scatters, \( \sigma_c \), are slightly smaller in the intermediate halo mass range and slightly larger in the lowest halo mass bin at \( \sim 0.01 \) levels. Note that since the halo mass estimations in the group catalogs are based on the ranking of characteristic group luminosity/stellar masses, where the central galaxy luminosity/stellar mass and halo mass are somewhat correlated (Yang et al. 2008), the typical uncertainty in the \( \sigma_c \) constraints is at \( \sim 0.02 \) (see their Figure 4). In addition, in the CLF/SHAM modelings, the typical \( \sigma_c \) assumed in the literature also spans quite a large range, 0.15 to \( \sim 0.20 \). Thus, for the general behaviors of our CLF model predictions for central galaxies, such amounts of differences are expected.

Next, for the satellite galaxies, compared to the CLF obtained from galaxy groups, our fiducial Match2 method gives very nice CLF model predictions in halos with mass \( \geq 10^{13.5} h^{-1} M_\odot \), while the satellite galaxies for our Match1 method at the relatively low-mass end are significantly underpredicted. On the other hand, however, in relatively lower-mass halos, the situation is quite different. Our fiducial Match2 method overpredicted the CLF at about the \( \sim 40\% \) level at \( L \sim 10^{9.5} h^{-2} L_\odot \) while the Match1 method prediction is much better. Shown in Figure 7 are the CLFs measured separately for the red and blue galaxies, respectively. Symbols with error bars are the results obtained from SDSS galaxy groups. The lines are the results obtained for our fiducial Match2 method. Even if we did not perform special treatments on the color of galaxies for our neighborhood abundance matching, we still find color dependence very similar to the galaxy groups, where massive halos clearly contain more red galaxies than blue galaxies (both
centrals and satellites), while the opposite applies to low-mass halos. Note that such a halo-mass dependence is indeed quite consistent with the halo-quenching mechanism (see, e.g., Wang et al. 2017 and references therein).

3.4. The Conditional Stellar Mass Functions

Apart from the CLF, which is more observationally related, we can also measure the CSMF of galaxies. The CSMF, \( \Phi(M_\ast | M_h) \), which describes the average number of galaxies as a function of galaxy stellar mass in a dark matter halo of a given mass, is more straightforwardly related to theoretical predictions of galaxy formation models than the CLF, because the conversion from stellar mass to luminosity in theoretical models requires detailed modeling of the stellar population and dust extinction. The CSMF can be estimated by directly counting the number of galaxies in groups or halos. However, as the galaxies used here are flux limited, here we need to take into account the completeness limits of galaxies as a function of stellar mass as well.

According to van den Bosch et al. (2008), for the stellar masses of the SDSS galaxies, at a given redshift \( z \), the stellar mass is complete above

\[
\log[M_{\ast,\text{lim}}/(h^{-2}M_\odot)] = 4.852 + 2.246\log D_L(z) + 1.123\log(1+z) - 1.186z
\]

Using this relation, we can obtain the redshift completeness limit \( z_{M_\ast} \) for a given stellar mass \( M_\ast \). Similar to the redshift limit for luminosities, here we only use galaxies and groups (halos) that are below the redshift limit \( z_{M_\ast} \) to estimate the CSMF, \( \Phi(M_\ast | M_h) \). In Figure 8, we show the resulting CSMFs for groups of different masses using symbols with error bars. The contributions of the central and satellite galaxies again are plotted separately using filled and unfilled symbols.

The solid and dashed lines shown in each panel of Figure 8 are the results measured for the Match2 and Match1 methods, respectively. Using the SDSS galaxy group results as references, we find that the general behaviors of the model predictions of the Match2 and Match1 methods are quite similar to those of the CLFs. The model prediction of the Match2 method agrees better with that in the SDSS galaxy groups for massive halos with mass \( \gtrsim 10^{13.5} h^{-1} M_\odot \) while the
model prediction of the Match1 method is better in lower-mass halos.

In Figure 9, we show the CSMFs separately for red and blue galaxies. Symbols with error bars are the results obtained from SDSS galaxy groups, while the lines are the results obtained for our fiducial Match2 method. Here again, we see that CSMFs for both red and blue galaxies can be well recovered.

4. The Biases of Galaxies

Having checked the performance of our galaxy—subhalo connections established using a neighborhood abundance matching approach in the HOD framework, we proceed to check their performance on larger scales. Note that since galaxies are only slightly moved to match nearby main halos or subhalos, the autocorrelation functions (ACFs) of galaxies on large scales will not change significantly. Here we use the cross-correlation functions (CCFs) between galaxies and dark matter particles to check if the large-scale environment is properly reproduced in the ELUCID simulation.

4.1. Cross-correlation between Galaxies and Dark Matter

With all of the galaxies linked to subhalos, we proceed to measure the CCF between subhalos (galaxies) and dark matter particles,

$$\xi_{\text{CCF}}(r) = \frac{P_{\text{HD}}(r)}{P_{\text{HR}}(r)} - 1,$$

where \(P_{\text{HD}}(r)\) and \(P_{\text{HR}}(r)\) are the number of subhalo—dark matter particle and subhalo—random pairs, respectively. For our investigations, the number of random points has been set to be the same as the number of dark matter particles within the simulation box. Those points follow a uniform distribution within the simulation volume.

We first measure the CCFs between galaxies (subhalos) and dark matter particles in the ELUCID simulation for the overall galaxy population. We divide the galaxies (subhalos) into six subsamples within different absolute magnitude bins: 

\[-17.0 \geq M_{h} - 5 \log h > -18.0, \quad -18.0 \geq M_{h} - 5 \log h > -19.0 \ldots -22.0 \geq M_{h} - 5 \log h.\]

The unfilled squares shown in Figure 10 are the CCFs measured for galaxies within these absolute magnitude bins for our fiducial Match2 method. The error bars are obtained from 100 jackknife resamplings of the galaxies. As a reference, we also show the ACF of dark matter particles in the ELUCID simulation in each panel of Figure 10 using a dotted line. Comparing to the ACF of dark matter, the CCFs of galaxies show somewhat weaker and stronger
clustering strengths for fainter and brighter subsamples, respectively, which are qualitatively consistent with the observational measurements of galaxy biases using ACFs (e.g., Zehavi et al. 2005; Wang et al. 2007).

In addition to the ACF of dark matter particles, for comparison, we also show using a solid line the results obtained for the Match1 method. The main difference induced by the Match2 and Match1 methods is in the satellite fraction of galaxies, especially at the faint end. According to the CCF comparison, we see that:

1. The galaxies in the two samples give very similar results on large scales at $r \gtrsim 5 \, h^{-1}$ Mpc.
2. On small scales, we see that the Match2 method has a stronger clustering strength, especially in the faint galaxy subsamples.

These features indicate that the clustering measurements of galaxies on small scales are also very important for HOD/CLF modeling, especially in constraining the satellite components.

Aside from the CCFs of the overall galaxy population, we also measure the CCFs of galaxies that are separated into red and blue subsamples. We show in Figure 11 the CCFs measured separately for red and blue galaxies for the Match2 method using solid dots and unfilled squares, respectively. The CCFs of the red and blue subsamples show quite different behaviors where the red galaxies show an overall stronger clustering strength than blue galaxies except in the brightest magnitude bin. Note that in our neighbor abundance matching approach, we did not distinguish between red and blue galaxies. Thus, the different clustering behaviors of red and blue galaxies are caused by their large-scale environments, e.g., satellite fraction, host halo masses, etc.

Other than the luminosities of galaxies, we also considered galaxies with different stellar masses. Similar to the treatment of luminosities, we separate galaxies into six subsamples within different stellar mass bins: $9.0 \lesssim \log M_*/(h^{-1}M_\odot) \lesssim 9.4$, $9.4 \lesssim \log M_*/(h^{-1}M_\odot) \lesssim 9.8$ ... $11.0 \lesssim \log M_*/(h^{-1}M_\odot)$. The clustering properties of galaxies in different stellar mass bins are very similar to those in different absolute magnitude bins, which, for simplicity, are not explicitly shown here.

To quantify the clustering strengths of galaxies, we show in Figure 12 the ratios of the galaxy–dark matter CCFs and the dark matter–dark matter ACFs, which are indeed the biases of galaxies as a function of radius. Here, the results are shown separately for galaxies of different colors and in different absolute magnitude bins as indicated in the plot. The

Figure 8. The conditional stellar mass functions (CSMFs) of galaxies in groups and host halos of different masses. Symbols correspond to the CSMFs obtained from SDSS galaxy groups, with solid and unfilled circles indicating the contributions from central and satellite galaxies, respectively. The error bars reflect the 1σ scatter obtained from 200 bootstrap samples. The solid and dotted lines are the results obtained for the Match2 and Match1 methods, respectively.

12
solid line in each panel are the results obtained from our Match2 method. For comparison, we also show the resulting biases extracted from the reconstructed real space ACFs of the galaxies obtained by Shi et al. (2016) using dots with error bars. In their study, the redshift space distributions of galaxies are mapped to real space by correcting redshift distortions on both small and large scales. Based on the real space distributions of galaxies thus reconstructed, Shi et al. (2016) measured the real space ACFs for galaxies in different absolute magnitude bins, which are the same as the ones used in this study. The biases of galaxies are then obtained using the square root of the ratios between the ACFs of galaxies and dark matter particles.

By comparing our model predictions with those obtained by Shi et al. (2016), we find that these two measurements agree quite well, especially for all and blue galaxies. In most cases, the data points agree with each other within the 1σ level, except for a few slightly larger than the 1σ level, while the discrepancies are somewhat larger for red galaxies. There are quite a number of data points that deviate from each other at about the 2σ level. Apart from these agreement checks, we also find that both of these bias measurements reveal some curvatures in the $-21.0 \geq 0.3M_h - 5 \log h > -22.0$ magnitude bin. According to the error bars, we believe that the curvature around $1 \ h^{-1} \text{Mpc}$, which roughly corresponds to a transition scale from a one-halo to a two-halo term is robust. The curvature at this scale, which is quite different for red and blue galaxies, might be useful for galaxy formation constraints.

The overall agreement of the bias for our galaxy–subhalo matched pairs indicate again that our neighborhood abundance matching method works very well and that the large-scale environments in our ELUCID simulation are quite reliably reproduced.

### 5. How to Use the Matched Data

Theoretically, if one can provide a perfect link between the observed galaxies and the subhalos in the simulation, one can then use the properties of individual galaxies to constrain galaxy formation models, e.g., via SAMs, etc., to unprecedented precision. The galaxy–subhalo connections obtained in this study from the ELUCID simulation, although not perfect, are already much better than the traditional subhalo abundance matching approach.

With all of the above tests on both small (halo-based) and large scales for the feasibility and reliability of our neighborhood abundance matching method, we proceed to provide some suggestions for the use of the matched galaxy–subhalo connections. Here, we suggest dividing all of the
galaxy−subhalo pairs into three categories: (1) halo-based pairs, (2) mass and local volume pairs, and (3) local volume pairs.

To make this separation, we first extract all of the galaxy−subhalo pairs that are either central−main pairs or satellite−subhalo pairs. We show in Figure 13 the group mass versus...
halo mass for these two kinds of galaxy—subhalo pairs. Shown in the left and right panels are the results for the central—main and satellite—subhalo pairs, respectively. For the central—main pairs, although we see that there are some pairs that are quite off the consistency line, caused by various reasons, e.g., the survey edge effect, mismatch, etc., the vast majority are consistent with each other. For the satellite—subhalo pairs, the situation is somewhat worse. We can see that quite a large fraction of them are quite offset from the consistency line, this is mainly caused by the mismatch of satellite galaxies with different host halos. As an illustration, we use two dotted lines, log $M_G - log M_h = \pm 0.3$, to separate the galaxy—subhalo pairs. In total, there are 212,798 central—main and 43,178 satellite—subhalo pairs with $|log M_G - log M_h| \leq 0.3$. Comparing to the total number of 277,139 central—main and 118,930 satellite—subhalo pairs, roughly 77% and 36% are central and satellite populations. If we only consider galaxies in halos with mass $log M_h \geq 13.5$, 51% and 54% are central and satellite populations that have $|log M_G - log M_h| \leq 0.3$.

Using the above behaviors of galaxy—subhalo pairs, we can separate them into three categories:

1. Cat 1 (halo-based pairs): as pointed out in Tweed et al. (2017), the reconstructed simulation can roughly reproduce more than half of the halos with mass $\gtrsim 10^{13.5}$ $h^{-1} M_{\odot}$ (e.g., with more than half of the particles in common). Here we select galaxy—subhalo pairs that have $log M_h \geq 13.5$ and $s \leq 3$ $h^{-1}$ Mpc (where $s$ is the galaxy—subhalo pair separation in redshift space), and $log M_{sh} \geq 11.5$. In total, there are 830 central—main and 7557 satellite—subhalo pairs that fall into this category. In contrast, for the same criteria, there are 13 central—main and 108 satellite—subhalo pairs that fall into this category for a rotated version of the ELUCID simulation. For these pairs, we suggest that one can use the related galaxy properties for those individual main or subhalos to evaluate the galaxy properties predicted by SAMs individually.

Figure 12. Biases for all, red, and blue galaxies in different luminosity bins as indicated. In each panel, the biases of galaxies obtained from the CCFs in this study is shown by the solid line with 1σ error bars. For comparison, the dots with error bars are the results obtained by Shi et al. (2016). Here again, we only show the results obtained from the Match2 method.

The Astrophysical Journal, 860:30 (17pp), 2018 June 10 Yang et al.
Figure 13. The halo mass obtained from galaxy groups, \( \log M_{\text{G}} \) vs. the host halo mass of the subhalos, \( \log M_{h} \), in the ELUCID simulation of the matched galaxy–subhalo pairs. Shown in the left and right panels are the results for the central–main pairs and satellite–subhalo pairs, respectively.

2. Cat 2 (mass and local volume pairs): all other galaxy–subhalo pairs that have \( | \log M_{G} - \log M_{h} | \leq 0.3 \). There are 211,968 central–main and 38,195 satellite–subhalo pairs that fall into this category. If we only consider galaxies in halos with mass \( \log M_{h} \geq 13.5 \), the related numbers are 809 for central–main and 18,250 for satellite–subhalo pairs, respectively. For these pairs, we suggest comparing the overall galaxy properties in similar-mass halos within the same local volume, i.e., within a radius of \( \sim 20 \ h^{-1} \) Mpc.

3. Cat 3 (local volume pairs): all other galaxy–subhalo pairs. For these pairs, one may compare the overall galaxy properties predicted by SAMs in given spherical regions with radius \( \sim 20 \ h^{-1} \) Mpc with those SDSS galaxies linked to subhalos in the same regions.

Based on these criteria, we will evaluate a few SAMs in a subsequent paper.

6. Summary

In this paper, we have proposed a novel neighborhood abundance matching method to link galaxies in the SDSS DR7 observation to dark matter main and subhalos in the ELUCID simulation. Here we used two matching methods for the abundance matching: Match1, which is quite popular in SAMs where galaxies are linked to all the surviving main halos and subhalos, and Match2, where the central galaxies are linked to the main halos and satellite galaxies to the surviving subhalos separately, and all of which use the maximum masses of the subhalos along their accretion histories. We made a list of tests on galaxy–subhalo connections established in this manner, and the main features are listed as follows:

1. Based on the Match2 method, we measured and modeled the luminosity (stellar mass)–subhalo mass relations for central and satellite galaxies separately and found that they have quite different behaviors.

2. We checked the satellite fractions of galaxies as a function of luminosity and stellar mass, and found that the Match1 method somewhat underestimates the related values, especially for low-mass galaxies. In addition, unlike the Match2 method, the color segregation of the satellite fraction is not well reproduced in the Match1 method.

3. We have measured the CLFs and CSMFs of galaxies in halos of different masses. Compared to the observational results, the model prediction of the Match2 method agrees better with that for massive halos with mass \( \geq 10^{13.5} \ h^{-1} M_{\odot} \) in SDSS galaxy groups, while the model prediction for the Match1 method is better for lower-mass halos.

4. We have measured the biases of galaxies as a function of radius, which overall show quite a nice agreement with the observational results obtained by Shi et al. (2016).

5. We have also checked the above quantities separately for red and blue galaxies. All of the results for our Match2 method agree fairly well with the direct measurements from observation.

The above tests show that the Match2 method performs somewhat better than the Match1 method. We thus suggest making use of the galaxy–subhalo connections established in this sample for galaxy formation studies, e.g., SAMs performed on the ELUCID simulation. In addition, we suggest that those galaxy–subhalo pairs can be divided into three categories: (1) halo-based pairs, which can be used to evaluate galaxy properties in individual subhalos, (2) mass and local volume pairs, which can be used to evaluate the overall galaxy properties in similar-mass halos in the same local volumes, and (3) local volume pairs, which can be used to evaluate the overall galaxy properties in the same small volumes. Finally, the galaxy–subhalo links and the subhalo merger trees in the SDSS DR7 region for our ELUCID simulation are available upon request.

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. 11233005, 11421303, 11522324, 11503064, 11621303, and 11733004) and Shanghai Natural Science Foundation, grant No. 15ZR1446700. We also thank the support of the Key Laboratory for Particle Physics, Astrophysics and Cosmology, Ministry of Education. H.J.M. would like to acknowledge the support of NSFC-11673065 and
