Photometric Objects around Cosmic Webs (PAC) Delineated in a Spectroscopic Survey.

I. Methods

Kun Xu1, Yun Zheng1, and Yipeng Jing1,2

1 Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, People’s Republic of China; yijing@sjtu.edu.cn
2 Tsung-Dao Lee Institute, and Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai, 200240, People’s Republic of China

Received 2021 September 22; revised 2021 November 4; accepted 2021 November 9; published 2022 January 21

Abstract

We provide a method for estimating the projected density distribution $\tilde{n}_{2wp}(r_p)$ of photometric objects around spectroscopic objects in a spectroscopic survey. This quantity describes the distribution of photometric sources with certain physical properties (e.g., luminosity, mass, and color) around cosmic webs (PAC) traced by the spectroscopic objects. The method can make full use of current and future deep and wide photometric surveys to explore the formation of galaxies up to medium redshift ($z < 2$) with the aid of cosmological spectroscopic surveys that sample only a fairly limited species of objects (e.g., emission line galaxies). As an example, we apply the PAC method to the CMASS spectroscopic and HSC-SSP PDR2 photometric samples to explore the distribution of galaxies for a wide range of stellar masses from $10^{9.0}$ to $10^{12.0} M_\odot$ around massive galaxies at $z_s \approx 0.6$. Using the abundance-matching method, we model $\tilde{n}_{2wp}(r_p)$ in N-body simulation using Markov chain Monte Carlo sampling, and we accurately measure the stellar–halo mass relation and stellar mass function for the whole mass range. We can also measure the conditional stellar mass function of satellites for central galaxies of different mass. The PAC method has many potential applications for studying the evolution of galaxies.

Unified Astronomy Thesaurus concepts: Galaxy formation (595); Observational cosmology (1146); Galaxy dark matter halos (1880)

1. Introduction

Modeling galaxy formation in the cosmological context is one of the greatest challenges in astrophysics and cosmology today. In the past few decades, the broad contours of galaxy formation physics were investigated and built (Mo et al. 2010; Silk & Mamon 2012; Somerville & Davé 2015). In theory, cosmic structures grow by gravitational instability from the initial tiny quantum fluctuations generated during the inflationary epoch. Dark matter (DM) halos, which are defined as DM objects with the density hundreds times the density of the Universe, are formed around the density peaks of the Universe. Halos grow by accreting surrounding smaller DM halos, and they diffuse matter, including DM and gas (Zhao et al. 2009). The gas is heated by shocks when accreted into a halo (Dekel & Birnboim 2006). Through radiation, the hot gas loses its energy and cools, and the cold gas spirals into the halo center to form a galaxy, which typically is a spiral. Under the hierarchical growth of the DM halo, surrounding galaxies can also be accreted into the halo and become its satellites. Some satellite galaxies, especially relatively massive ones, spiral into the center and coalesce with the central galaxy due to dynamical friction (Lacey & Cole 1994; Jiang et al. 2008). A merger of a central disk galaxy with a satellite of comparable mass (e.g., a mass ratio $\gtrsim 0.2$) may significantly change its internal structure, either producing an elliptical galaxy or a disk galaxy with a significant bulge (Naab et al. 2006; Cox et al. 2006; Bekki & Couch 2011).

Studies find that the black hole (BH) mass is highly correlated with the bulge mass (Kormendy & Ho 2013), so we expect that there are supermassive black holes (SMBHs) within elliptical galaxies or bulges. The strong energy output and/or material outflow produced by an SMBH can heat and/or blow out the surrounding cold gas, thus suppressing the gas reservoir from which stars form and causing the host galaxy to look red (Fabian 2012). Therefore, the bimodal distribution of galaxy color is highly correlated with the galaxy morphology distribution, with ellipticals being red and spirals being blue.

To fully understand the galaxy properties and distributions, one needs to consider all the complicated physical process involved in galaxy formation and evolution mentioned above. Although there exists a broad contour of the galaxy formation process, a fully predictive framework from first principles has yet to be established (Naab & Ostriker 2017). Physical models such as semianalytic models (White & Frenk 1991; Kauffmann et al. 1993; Kang et al. 2005; Guo et al. 2013) and hydrodynamic simulations (Schaye et al. 2015; Pillepich et al. 2018; Davé et al. 2019) approximate physics below their respective resolution scales to simulate the effects of supernovae, radiation pressure, multiphase gas, BH accretion, active galactic nucleus (AGN) feedback, and metallicity evolution in galaxy formation. However, different approximations lead to different galaxy properties (Somerville et al. 2008; Lu et al. 2014), and the uncertainty still remains.

Empirical modeling such as the halo occupation distribution (HOD, Jing et al. 1998; Peacock & Smith 2000; Ma & Fry 2000; Seljak 2000; Berlind & Weinberg 2002; Yang et al. 2003; Zheng et al. 2005; Zu & Mandelbaum 2015) and abundance matching (AM, Wechsler et al. 1998; Wang & Jing 2010; Moster et al. 2013; Behroozi et al. 2019) uses significantly weaker priors, and the physical constraints come almost entirely from observations. These models connect...
average galaxy properties such as occupation numbers and stellar mass to halos as a function of halo mass or halo circular velocity, and determine the parameters by fitting the observed properties of galaxies. Although it has been attempted recently in more complex models to incorporate properties of gas (Popping et al. 2015), metallicity (Rodríguez-Puebla et al. 2016), and dust (Imara et al. 2018) for a comparison with observations, the stellar–halo mass relation (SHMR) is still one of the most commonly used relations to model galaxy-halo connection (Guo et al. 2010; Wang & Jing 2010; Moster et al. 2013), in which larger halos host larger galaxies with relatively tight scatter. Recent studies found that galaxies with different properties, such as color, may have different SHMRs, which may indicate the so-called galaxy assembly bias, which causes the scatter in the average SHMR (Cooper et al. 2010; Wang et al. 2013; Zentner et al. 2014; Hearin et al. 2015; Mandelbaum et al. 2016; Cui et al. 2021). Similarly, for the HOD, the galaxy occupation number may also depend on halo properties such as halo concentration and environment other than the halo mass (Hadzhiyska et al. 2020). The best secondary parameter for the halo characterization is still being sought in the development of HOD and AM.

The stellar mass function (SMF) and galaxy clustering (GC) are the two most commonly used properties to constrain the parameters in HOD and AM. The measurement of these quantities usually relies on spectroscopic surveys with redshift information. In the past two decades, there has been significant progress in large spectroscopic surveys (York et al. 2000; Colless et al. 2001; Steidel et al. 2003; Le Fèvre et al. 2005; Ahn et al. 2012; Bolton et al. 2012; Garilli et al. 2014; Takada et al. 2014; DESI Collaboration et al. 2016; Ahumada et al. 2020). In the local universe \(z \approx 0\), large spectroscopic surveys, in particular, the Two Degree Field Galaxy spectroscopic survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000), have been used to measure the SMF and GC down to \(10^{9.0} \, M_\odot\) (Cole et al. 2001; Norberg et al. 2002; Li et al. 2006; Li & White 2009; Peng et al. 2010), although the accuracy of the measurements, especially the GC, is still very limited by the survey volume for faint galaxies. At a higher redshift \(z \approx 0.5 \sim 1.0\), the DEEP2 Galaxy spectroscopic survey (Davis et al. 2003), the VIMOS- VLT Deep Survey (VVDS; Le Fèvre et al. 2005), and the VIMOS Public Extragalactic spectroscopic survey (VIPERS; Garilli et al. 2014) have been used to successfully measure the SMF and GC for galaxies with \(M_\star > 10^{9.0} \, M_\odot\) (Poizetti et al. 2007; Meneux et al. 2008; Davidzon et al. 2013; Marulli et al. 2013; Mostek et al. 2013). However, despite the huge efforts, measuring fainter objects is still very difficult, and stellar mass-limited samples are usually very small at even higher redshift. Fortunately, huge next-generation spectroscopic surveys are being constructed for cosmological studies at intermediate to high redshift (Laureijs et al. 2011; Takada et al. 2014; DESI Collaboration et al. 2016). Due to the limited wavelength coverage and sensitivity of the spectrographs, different populations of galaxies, such as emission line galaxies (ELGs), QSOs, and Lyman break galaxies (LBGs), are targeted at different redshifts. These populations are all expected to trace large-scale structures or the cosmic webs, so they can be used to extract information for cosmological studies. However, it is difficult to use these surveys to construct stellar mass-limited samples for the target selections used by the surveys.

Compared to spectroscopic surveys, photometric surveys, which take the images and obtain the photometric information, are usually deeper and more complete in terms of the stellar mass. However, without precise redshift measurement, the usefulness of photometric surveys is limited. People attempt to infer the photometric redshift from their broadband magnitudes in the photometric surveys (Ilbert et al. 2006, 2009; Sánchez et al. 2014; Nishizawa et al. 2020). Photo-z has been used to measure the SMF (Fontana et al. 2006; Pérez-González et al. 2008; Ilbert et al. 2010; Bielby et al. 2012; McLeod et al. 2021) and GC (Crocco et al. 2016; Ishikawa et al. 2020; Wang et al. 2021b), although one has to be very careful about the error and systematics of photo-z, especially for faint galaxies. The next-generation large and deep multiband photometric surveys, such as Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) and Euclid (Laureijs et al. 2011), are expected to fairly sample galaxies to an unprecedented faint limit.

As mentioned above, photometric and spectroscopic surveys both have their advantages and disadvantages. Future large spectroscopic surveys are expected to have precise redshift information, but only for bright objects and/or selected (or biased) populations, while photometric surveys are deeper and more complete, but lack accurate redshift measurements. So far, the measurement of the SMF and GC mainly stems from the spectroscopic surveys. Thus, studies based on the SMF and GC, such as HOD and AM, are focused on the local universe or the massive end at higher redshift, resulting in a poor understanding of the faint end and the redshift evolution. To study small and faint objects, quite a few studies attempted to combine a spectroscopic survey with a photometric survey, which can reach several magnitudes fainter than pure spectroscopic surveys. With the cosmic webs traced by objects in spectroscopic surveys, the properties and distribution of photometric objects around the cosmic web can be studied using photometric surveys. Results can be achieved by stacking satellite and neighbor counts around a large sample of central galaxies, with foreground and background sources are subtracted statistically (Phillips & Shanks 1987; Lrrimore et al. 1994; Wang et al. 2011; Guo et al. 2012; Ménard et al. 2013; Newman et al. 2015; Lan et al. 2016). However, most previous studies only use luminosity and color properties of a photometric sample, which is not suitable for quantitative HOD and AM studies, which require physical properties of galaxies such as stellar mass, rest-frame color, and star formation rate (SFR).

In this paper, we provide a method for measuring the distributions and properties (luminosity, mass, color, SFR, morphology, etc.) of photometric objects around cosmic webs (PAC) represented by spectroscopic objects in a spectroscopic survey. The basic idea is that for a spectroscopic source \(i\) at redshift \(z_{e,i}\), only those objects in the photometric sample around \(z_{e,i}\) are correlated to source \(i\) and share a similar redshift. Thus for source \(i\), we calculate the physical properties for all sources in the whole photometric sample by assuming they were all at redshift \(z_{e,i}\). Through the cross-correlation of the photometric and spectroscopic samples, foreground or background galaxies with wrong redshift information can be canceled out with the help of random samples, and the true distribution of photometric sources with specified properties around the spectroscopic sources can be obtained. Because both the spectroscopic and photometric samples are huge in
next-generation surveys, we will develop a method to speed up the computation for the physical properties.

We introduce the details of PAC in Section 2. In Section 3 we apply PAC to observations. The measurement is modeled in N-body simulation using AM in Section 4. A brief conclusion is given in Section 5. We adopt a cosmology with $\Omega_m = 0.268$, $\Omega_\Lambda = 0.732$, and $H_0 = 71$ km/s/Mpc throughout the paper.

2. Methodology

In this section, we introduce a method for estimating $\bar{n}_2 w_p(r_p)$ from $w_{12}(\theta)$, where $w_p(r_p)$ and $w_{12}(\theta)$ are the projected cross-correlation function (PCCF) and the angular cross-correlation function (ACCF) between a given set of spectroscopically identified galaxies and a large sample of photometric galaxies, and $\bar{n}_2$ is the mean number density of the photometric galaxies. The quantity $\bar{n}_2 w_p(r_p)$ has a clear physical meaning: it measures the true excess of the photometric objects around the spectroscopic objects on the sky projection. If we choose photometric galaxies at a given stellar mass, the measurement over a range of the stellar mass gives the information about the stellar mass function and the clustering as a function of the stellar mass, which are the key ingredients for understanding the connection of galaxies to DM halos. Therefore, we extend this method to statistically measuring the distribution of the photometric galaxies with specified physical properties (i.e., stellar mass, SFR, and color) around spectroscopically identified galaxies.

2.1. Estimating $\bar{n}_2 w_p(r_p)$ from $w_{12}(\theta)$

Throughout this section, we call a spectroscopic sample population 1 and a photometric sample population 2.

Assuming an object in population 1 is at distance $r_1$, the number of objects $dN_2$ in population 2 within a solid angle element $d\Omega_2$ in the direction $r_2$ is

$$dN_2 = \int n_2(r_2) r_2^2 dr_2 d\Omega_2.$$  

Here $n_2$ is the mean number density of population 2 at distance $r_2$. The expected number of population 2 objects around a population 1 object is

$$\langle dN_2 \rangle = \int \langle n_2(r_2) \rangle r_2^2 dr_2 d\Omega_2$$

$$= d\Omega_2 \int \bar{n}_2[1 + \xi_{12}(r_1)r_2^2] dr_2$$

$$\approx d\Omega_2[\bar{S}_2 + \bar{n}_2 w_p(r_1 r_1')]$$.

$\bar{S}_2$ is the mean angular surface density of population 2, $\xi_{12}$ is the CCF between the two populations, and $w_p(r_p) = \bar{r}_1$ implies the projected CCF. The approximation holds if $\theta$ is small. Then we have

$$\bar{n}_2 w_p(r_1 r_1') = \frac{\langle dN_2 \rangle}{d\Omega_2} - \bar{S}_2$$

$$= \bar{S}_2 \frac{\langle dN_2 \rangle - \langle dR_2 \rangle}{\langle dR_2 \rangle}$$

$$= \bar{S}_2 \frac{\langle D_1 D_2 \rangle - \langle D_1 R \rangle}{\langle D_1 R \rangle}$$

$$= \bar{S}_2 w_{12}(\theta).$$

Here $\langle D_1 D_2 \rangle$ and $\langle D_1 R \rangle$ are the cross pair counts between population 1 and population 2, and between population 1 and the random sample of population 2.

Hence, with our method, $\bar{n}_2 w_p(r_1 r_1')$ can be estimated from $\bar{S}_2 w_{12}(\theta)$ with only the redshift of population 1. The physical meaning of the quantity $\bar{n}_2 w_p(r_1 r_1')$ is the excess number of population 2 around population 1.

2.2. Physical Properties of Photometric Sources around Spectroscopically Identified Sources

To obtain physical properties for galaxies, the distance or redshift information is needed. Unfortunately, wide deep photometric surveys do not have a measured redshift for most of the galaxies. The photometric redshift $z_p$ is often used to approximate the distance of galaxies, but the errors of $z_p$ are very difficult to estimate, especially for faint galaxies. In our method we do not use any information of $z_p$. Instead, because the photometric sources that are correlated with the spectroscopically identified galaxies must share the same redshift, we can use the spectroscopic redshift $z_s$ for the physical properties for photometric sources around the spectroscopic object. As $\bar{n}_2 w_p(r_1 r_1')$ is just the number excess of neighbors around population 1, we can extend our method to estimating the distributions and properties of PAC delineated by a spectroscopic survey.

Assuming we only have a spectroscopic sample (population 1) and a large photometric sample, and we wish to calculate $\bar{n}_2 w_p(r_1 r_1')$ between these sources in population 1 at a single redshift $z_s$ and sources with physical property X included in the photometric sample. To select population 2, we assume that the whole photometric sample is at redshift $z_s$ and calculate their physical properties using a spectral energy distribution (SED) fitting. As one may note, the calculation is correct only for sources around population 1. The physical properties that are calculated are incorrect for foreground and background sources. However, because foreground and background sources along the line of sight (LOS) to population 1 are distributed statistically in the same way as those along a random LOS, the foreground and background will cancel out when we calculate $w_{12}(\theta)$ (see Section 3.3.1, Wang et al. 2011, for a more rigorous derivation). Therefore, population 2 is just selected as the sources with the physical property X in the photometric sample even if we assume that the whole sample is at redshift $z_s$ in the calculation. With this method, we can study the distribution of satellites and neighbors with specified physical properties around spectroscopically identified galaxies.

2.3. Spectroscopic Sample with a Redshift Distribution

So far, we have assumed that all the sources in population 1 have the same redshift. However, in reality, spectroscopic samples all have a redshift distribution. If the redshift range is relatively narrow and the evolution of the universe can be neglected, $\bar{n}_2 w_p(r_p)$ will not vary much, while the change of $r_1$ and $\theta = r_p/r_1$ may not be negligible from one spectroscopic source to another. Thus, $\bar{n}_2 w_p(r_p)$ is a better statistic quantity than $\bar{n}_2 w_p(r_p) r_1^2$. We
We can also change the angular separation used to measure the number counts and make $r_p = r_{1,i}$, the same for each galaxy in population 1. Then, we sum the counts around each galaxy without weighting. These two methods are equivalent.

To minimize the effects of complex survey geometries, we adopt an estimator in analogy to the Landy–Szalay estimator (Landy & Szalay 1993):

$$w_{12,\text{weight}}(\theta) = \frac{\langle D_1 D_2 \rangle_w - \langle D_1 R_2 \rangle_w - \langle D_2 R_1 \rangle_w + \langle R_1 R_2 \rangle_w}{\langle R_1 R_2 \rangle_w}$$

where $R_1$ and $R_2$ are the random points for spectroscopic and photometric samples, respectively.

The above method has accounted for the variance of $\theta$ with redshift for sources in population 1, while $r_{1,i}$ in $S_{w_{12,\text{weight}}}(\theta)$ still varies with redshift. Therefore, we divide population 1 into narrower redshift bins and reduced the error from the change of $r_{1,i}$. For population 1 with a redshift distribution, we can divided them into $m$ redshift bins. $m$ can be as large as possible only if there are sufficient galaxies in each bin. If the mean redshifts for these bins are $\{z_{s,j}\}_{j=1,m}$, we calculate the physical properties for the whole photometric catalog and select a population 2 for each $z_{s,j}$. Then, we calculate $S_{w_{12,\text{weight}}(\theta)}(\theta)$ for each redshift bin, and the mean $\bar{n}_2 w_p(r_1 \theta)$ of population 1 can be obtained by averaging over these redshift bins:

$$\bar{n}_2 w_p(r_1 \theta) = \frac{1}{m} \sum_{j=1}^{m} \frac{S_{w_{12,\text{weight}}}(\theta)}{r_{1,i}}.$$ 

### 3. Applications to CMASS and HSC samples

In this section, we apply PAC to the CMASS spectroscopic sample in the Baryon Oscillation Spectroscopic Survey (BOSS; Ahn et al. 2012; Bolton et al. 2012) and to the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2019) PDR2 wide-field photometric sample.

#### 3.1. Observational Data

We use the HSC-SSP PDR2 wide-field photometric catalog (Aihara et al. 2019) as the photometric sample. To obtain more accurate physical properties, we choose sources in the footprints observed with all five bands (grizy) to ensure that there are enough bands for the SED. Sources around bright objects are masked using the $\{\text{grizy}_\text{mask}\}_{pdr2}$ bright object center flag provided by the HSC collaboration (Coupon et al. 2018). We also use the $\{\text{grizy}_\text{extendedness\_value}\}$ flag to exclude stars in the sample. Finally, there are around $2 \times 10^8$ galaxies in our photometric sample. We can also construct a random point catalog (100 arcmin$^2$) with the same selection criteria from the HSC database for the ACCF analysis. The effective area calculated from the random point number is 501 deg$^2$.

To ensure that the distribution of foreground and background galaxies is the same for all LOS directions, the easiest way is to ensure that the sample is complete for galaxies with specified physical properties at the required redshift. Because the survey depth is not uniform across the HSC survey, for a low stellar mass limit, some patches remain complete while others may not be complete. Therefore, we use the HSC-SSP PDR2 deep-field catalog and DECam photo-z (Nishizawa et al. 2020) to study the completeness of galaxies for different stellar mass. We select galaxies with photo-z between 0.5 and 0.7 and calculate the physical properties for these galaxies with five bands grizy using the SED code CIGALE (Boquien et al. 2019).

The stellar population synthesis models of Bruzual & Charlot (2003) are used to rederive Equation (3):

$$\bar{n}_2 w_p(r_1 \theta) = \frac{\langle dN_2 \rangle / r_1^2 - \bar{S}_2}{r_1^2} = \frac{\langle dN_2 \rangle / r_1^2 - \langle dR_2 \rangle / r_1^2 \bar{S}_2}{r_1^2} = \frac{\langle D_1 D_2 \rangle_w - \langle D_1 R \rangle_w \bar{S}_2}{\langle D_1 R \rangle_w r_1^2} = \frac{\bar{S}_2}{r_1^2} w_{12,\text{weight}}(\theta).$$

(4)
to compute the physical properties of galaxies. In these calculations, the Chabrier (2003) initial mass function is adopted. We assume a delayed star formation history \( \phi(t) \approx 1 \exp(-t/\tau) \), where \( \tau \) is taken from 10^7 to \( 1.258 \times 10^{10} \) yr with an equal logarithmic interval \( \Delta \lg \tau = 0.1 \). Three metallicities, \( Z/Z_\odot = 0.4, 1, \) and 2.5, are considered, where \( Z_\odot \) is the metallicity of the Sun. We use the extinction law of Calzetti et al. (2000) with dust reddening in the range \( 0 < E(B-V) < 0.5 \). As shown in Figure 1, the stellar mass shows a clear correlation with \( z \)-band magnitude at \( 0.5 < z_p < 0.7 \), so that the magnitude limit can be used to derive a complete sample for a specified stellar mass. Particularly, we define the \( z \)-band completeness limit \( C_{95}(M) \) that 95% of the galaxies are brighter than \( C_{95}(M) \) in the \( z \)-band for a given stellar mass \( M \). Therefore, we adopt a stellar mass cut at \( 10^{11.3} M_\odot \) in this study. Moreover, we only consider central galaxies in the spectroscopic sample, so we select galaxies that do not have more massive neighbors within the projected distance of 1 Mpc h^{-1}. Finally, there are 8028 massive \( (>10^{11.3} M_\odot) \) central galaxies left in the spectroscopic sample.

3.2. PAC for Different Mass Bins

We first divide the spectroscopic sample into three mass bins, from \( 10^{11.3} M_\odot \) to \( 10^{11.9} M_\odot \). In each mass bin, galaxies are divided into four redshift bins with an equal interval \( \Delta z = 0.05 \). Then, we perform an SED for the whole photometric sample at redshifts 0.525, 0.575, 0.625, and 0.675, respectively. After this, we obtain the magnitudes in seven bands \( grizyW1W2 \) for each CMASS galaxy in the footprint of HSC. We calculate the physical properties for these galaxies using the same SED templates, but with seven bands \( grizyW1W2 \). As noted by previous studies (Maraston et al. 2013; Leauthaud et al. 2016; Guo et al. 2018), the CMASS sample is complete to stellar mass \( M_* \approx 10^{11.3} M_\odot \). Therefore, we adopt a stellar mass cut at \( 10^{11.3} M_\odot \) in this study. Moreover, we only consider central galaxies in the spectroscopic sample, so we select galaxies that do not have more massive neighbors within the projected distance of 1 Mpc h^{-1}. Finally, there are 8028 massive \( (>10^{11.3} M_\odot) \) central galaxies left in the spectroscopic sample.

The Astrophysical Journal, 925:31 (11pp), 2022 January 20 Xu, Zheng, & Jing

![Figure 2](image-url)
redshift range \(0.5 < z < 0.7\) can be obtained. During the calculation, only footprints deeper than \(C_{05}(m)\) are used for each mass bin.

To evaluate the statistical error, we use the jackknife resampling technique (Efron 1982). The mean value and error of the mean value of \(\bar{n}_2 w_p(r_p)\) for each mass bin can be calculated as

\[
\bar{n}_2 w_p(r_p) = \frac{1}{N_{\text{sub}}} \sum_{k=1}^{N_{\text{sub}}} n_2^{k, w_p,k}(r_p)
\]

\[
\sigma^2 = \frac{N_{\text{sub}} - 1}{N_{\text{sub}}} \sum_{k=1}^{N_{\text{sub}}} (n_2^{k, w_p,k}(r_p) - \bar{n}_2 w_p(r_p))^2.
\]

where \(N_{\text{sub}}\) is the number of jackknifed realizations, and \(n_2^{k, w_p,k}(r_p)\) is the excess of the projected density of the \(k\)th realization. In this work, we adopt \(N_{\text{sub}} = 50\).

The results are shown in Figure 2 with colored dots. Each panel shows the results for the same mass bin of the photometric sample. Because \(\bar{n}_2\) is the same in the same panel, the difference of \(\bar{n}_2 w_p(r_p)\) reflects the difference of \(w_p(r_p)\) between different spectroscopic mass bins. Although the footprints become smaller for the lowest mass bin \(10^{9.0}\) \(\text{M}_\odot\), the clustering signal is still quite good for all the three spectroscopic subsamples, so we can study the properties such as SMF and SHMR down to the low-mass end.

To test the robustness of the PAC to the redshift bin used, we compare the results when different numbers of redshift bins are used to divide the spectroscopic subsample. As we can see from Figure 3, for a narrow redshift range as in our study \((0.5 < z < 0.7)\), the measurement is nearly independent of the number of redshift bins used, indicating that our algorithm is robust.

Figure 3. Comparison of the \(\bar{n}_2 w_p(r_p)\) measured with different divisions of redshift bins (as indicated in the figure) for the spectroscopic sample. The result is insensitive to the number of redshift bins used.

Figure 4. Subhalo mass \(m_{\text{peak}}\) function in halos with \(M_{200} = 10^{14.0} \text{M}_\odot\). Dots show the results from the CosmicGrowth simulation after correction for subhalos with fewer than 20 particles. The solid line shows the result from Han et al. (2016) based on high-resolution zoom-in simulations.

4. Abundance Matching with an N-body Simulation

With the density and clustering information of galaxies, we can study the galaxy-halo connection using an N-body simulation. HOD and AM are the two most commonly used methods for populating galaxies to DM halos. We follow Wang & Jing (2010) and use AM to study the galaxy–halo relation based on \(\bar{n}_2 w_p(r_p)\) for different mass bins obtained in the observation. As we show below, the SHMR, the stellar mass function (SMF) and the conditional stellar mass function (CSMF) for satellites can be inferred from the AM results for galaxies in a wide range of stellar masses \((10^{9.0} - 10^{12.0} \text{M}_\odot)\).

4.1. CosmicGrowth Simulation

We use the CosmicGrowth simulation (Jing 2019) for our studies. The CosmicGrowth simulation is a grid of high-resolution N-body simulations that are run in different cosmologies using an adaptive parallel P3M code (Jing & Suto 2002). We use the \(\Lambda\)CDM simulation with cosmological parameters \(\Omega_m = 0.268, \Omega_k = 0.732,\) and \(\sigma_8 = 0.831\). The box size is 600 Mpc \(h^{-1}\) with 3072\(^3\) DM particles and a softening length \(\eta = 0.01 \text{Mpc} h^{-1}\).

Groups are identified with the friends-of-friends algorithm with a linking length 0.2 times the mean particle separation. The halos are then processed with HBT+ (Han et al. 2012, 2018) to obtain subhalos and their evolution histories. We use the catalog of Snapshot 83 at a redshift of about 0.57 to compare with the observation.

We also use the fitting formula in Jiang et al. (2008) to evaluate the merger timescale for subhalos with fewer than 20 particles (including orphans), which may be unresolved, and abandon those that have already merged into central subhalos. In Figure 4 we compare our subhalo mass function in halos with \(M_{200} = 10^{14.0} \text{M}_\odot \text{h}^{-1}\) to the universal subhalo mass function of Han et al. (2016), who used high-resolution zoom-in simulations.
mass resolution $\sim 10^3 M_\odot h^{-1}$ for the highest one) and carefully corrected for the resolution effect. In this comparison, the halos are defined as objects of radius $R_{200}$ within which the average density equals 200 times the critical density of the universe, and the subhalo mass is defined as its peak halo mass $m_{\text{peak}}$ in its history, to be consistent with Han et al. (2016). Our subhalo mass function is in good agreement with

\begin{align*}
\log(M_0) &= 11.65^{+0.24}_{-0.19} \\
\alpha &= 0.33^{+0.11}_{-0.11} \\
\beta &= 2.39^{+0.49}_{-0.34} \\
\log(k) &= 10.20^{+0.17}_{-0.15} \\
\sigma &= 0.22^{+0.04}_{-0.05}
\end{align*}

\textbf{Figure 5.} Constraints on the parameters of the SHMR model using Markov Chain Monte Carlo sampling. The central value is a median, and the error means 16 $\sim$ 84 percentiles after other parameters are marginalized over.

\textbf{Table 1}

| model   | $M_0$ ($M_\odot h^{-1}$) | $\alpha$  | $\beta$    | $k$ ($M_\odot$) | $\sigma$    |
|---------|--------------------------|-----------|------------|-----------------|------------|
| Unified | $10^{11.65^{+0.11}_{-0.11}}$ | $0.33^{+0.11}_{-0.11}$ | $2.39^{+0.49}_{-0.34}$ | $10^{10.20^{+0.17}_{-0.15}}$ | $0.22^{+0.04}_{-0.05}$ |
Han et al. (2016) down to $10^{10.5} M_\odot h^{-1}$, which is good enough for this study.

4.2. Abundance Matching

The SHMR can be described by a formula of a double power-law form:

$$M_* = \left( \frac{M_{\text{acc}}}{M_\odot} \right)^{-\alpha} + \left( \frac{M_{\text{acc}}}{M_\odot} \right)^{-\beta} k,$$

(11)

where $M_{\text{acc}}$ is defined as the virial mass $M_{\text{vir}}$ of the halo at the time when the galaxy last was the central dominant object. We use the fitting formula in Bryan & Norman (1998) to find $M_{\text{vir}}$. The scatter in $\log(M_*)$ at a given $M_{\text{acc}}$ is described with a Gaussian function of width $\sigma$. We use the same set of parameters for centrals and satellites (unified model) as in many studies (Wang & Jing 2010; Behroozi et al. 2019).

Once the parameters $\{M_\alpha, \alpha, \beta, k, \sigma\}$ are fixed, galaxies can be assigned to each DM halo. To compare $\tilde{n}_2 w_p(r_p)$ with observation, we define $\chi^2$ as

$$\chi^2 = \frac{1}{N_p} \sum_{N_p} \left[ \frac{\log(\tilde{n}_2 w_p(r_p))_{\text{sim}} - \log(\tilde{n}_2 w_p(r_p))_{\text{obs}}}{\sigma(\log(\tilde{n}_2 w_p(r_p))_{\text{ob}})} \right]^2,$$

(12)

where $N_p$ is the total number of points over which $\tilde{n}_2 w_p(r_p)$ is compared. We only consider the radius range of $0.1 < r_p < 10$ Mpc $h^{-1}$ in order to avoid the deblending problem of the HSC catalog at smaller $r_p$ (Wang et al. 2021a), and large errors at $r_p > 10$ Mpc $h^{-1}$. In order to perform a maximum likelihood analysis, we use the Markov Chain Monte Carlo (MCMC) sampler emcee (Foreman-Mackey et al. 2013).

Posteriors probability density functions (PDFs) of the parameters of the SHMR model from the MCMC are shown in Figure 5 and Table 1. The corresponding $\tilde{n}_2 w_p(r_p)$ and errors for each mass bin are shown by solid lines with shadows in
discussed in literature (Li et al. 2006; Pozzetti et al. 2007; Davidzon et al. 2013). Because it is hard to detect the faint objects and the survey areas of the high-redshift deep surveys are usually small, completeness, selection effects, and cosmic variance should be carefully considered. Different weighting methods for compensating for the incompleteness and regions with different densities that the surveys observed can produce a huge difference in the SMF. Therefore, the SMF of the faint end at higher redshift is still very hard to measure, for which PAC is a very promising method that combines the advantages of both the photometric and spectroscopic surveys. Interestingly, the SMF from our work using PAC and AM is well consistent with the very recent measurement of McLeod et al. (2021) down to $10^{9.0} M_{\odot}$.

### 4.4. CSMF of Satellites

The CSMF of satellites around a central galaxy of given stellar mass can be derived from the PAC measurement and the AM modeling. We define a satellite to be a galaxy within $R_{\text{vir}}$ of the DM halos of the central galaxies. To obtain the number of satellites, the most straightforward way is to sum the excess surface density $n_{2} w_{p}(r_{p})$ weighted by area within $R_{\text{vir}}$. However, two effects should be corrected for in our study. The surface density includes all excess of the galaxies along the LOS direction rather than within $R_{\text{vir}}$ (see Jiang et al. 2012), and the measurement within $0.1 R_{\text{vir}}$ is unreliable for the HSC PDR2 photometric catalog due to the deblending problem (Wang et al. 2021a).

Therefore, we use the results from AM to compensate for these two effects. After populating the galaxies to halos, we measure the average number of satellites within the projected radius range $0.1 R_{\text{vir}} < r_{p} < R_{\text{vir}}$ and within the virial radius $R_{\text{vir}}$ for each central and satellite mass bin. Then, we calculate the average number of satellites within $0.1 R_{\text{vir}} < r_{p} < R_{\text{vir}}$ in the observation and infer the number within $R_{\text{vir}}$ using the ratio calculated from the simulation. We show the CSMF from $10^{9.0}$ to $10^{11.6} M_{\odot}$ for the central galaxies with different mass bins in Figure 7. Colored dots show the results from the observation, and the solid lines are the results from AM. The results are consistent with each other at all mass bins for observation and simulation.

![Figure 7. Conditional stellar mass function of satellites for centrals with different mass.](image-url)

5. Conclusion

In this paper, we provide a method for estimating the projected density distribution $\tilde{n}_{2} w_{p}(r_{p})$ from $w_{12}(\theta)$ and extended this method to measure the distributions and properties of PAC traced by spectroscopic surveys. Basically, assuming that the whole photometric sample at the same redshift consists of spectroscopic sources, we can calculate the physical properties of the photometric sample. Through cross-correlation, foreground and background galaxies with wrong properties are canceled out, and the true distribution of photometric sources with specified physical properties around the spectroscopic sources can be obtained.

We apply PAC to massive ($>10^{11.2} M_{\odot}$) central galaxies in the BOSS CMASS sample ($0.5 < z < 0.7$) and in the HSC-SSP PDR2 wide-field photometric sample. We calculate $n_{2} w_{p}(r_{p})$ for several stellar mass bins (three for CMASS times four for HSC) from $10^{9.0}$ to $10^{12.0} M_{\odot}$, and the measurement is good overall at $0.1 \text{ Mpc} h^{-1} < r_{p} < 10 \text{ Mpc} h^{-1}$ for all the mass bins. Then, we use AM to model $n_{2} w_{p}(r_{p})$ in an $N$-body simulation with an MCMC sampling. We use the same set of parameters for central and satellite galaxies to model the observation. All the parameters are constrained well, and the fitting to observation is good overall for all mass bins. Our AM model can accurately reproduce the SMF compared to observations.

The SHMR from our results also agrees well with previous works. Using PAC and AM, we also calculate the CSMF of satellites for centrals with different masses.

We expect that PAC will have many applications with ongoing and upcoming photometric and spectroscopic surveys. However, because the wide spectroscopic surveys at higher redshifts only target specific populations of galaxies such as ELGs and QSOs, the galaxy-halo connection of these populations, which is also one of the key challenges for galaxy formation and cosmological studies, should be well established to make full use of PAC. Recently, there has been some progress made in ELG HOD and AM modeling with spectroscopic or narrow-band data (H. Gao et al. 2021, in preparation; Guo et al. 2019; Okumura et al. 2021). PAC can provide more information for studying the galaxy-halo connection of ELGs. As in this work, the SHMR of normal galaxies can be obtained using PAC with a stellar mass-limited spectroscopic sample such as large red galaxies (LRGs). If the redshift range of ELGs overlaps with that of LRGs, by applying PAC to ELGs and the same photometric sample, we can obtain the galaxy bias of ELGs with respect to the underlying matter distribution. With the better understanding of the galaxy-halo connection for ELGs, we can extend PAC to a higher redshift where LRGs cannot be reached. It is even worth trying to simultaneously study the connections of the ELGs and the normal galaxy population to DM halos because the projected density distribution $\tilde{n}_{2} w_{p}(r_{p})$ is expected to be precisely measured with next-generation galaxy surveys.

With PAC and future surveys, we can study the galaxy-halo connection for galaxies with different physical properties other than mass, such as SFR, color, and morphology (Xu & Jing 2021). We can also push the understanding of SHMR, SMF, and other properties to higher redshift and to the fainter luminosity end. With the properties and distribution of satellites, we can also study galaxy evolution, such as the galaxy merger rate, the merger timescale, and environment quenching. Moreover, because PAC has a very strong signal at the small scale, we can also quantify the fiber collision effect in
spectroscopic samples. We may also apply the method to photo-$z$ samples to quantify photo-$z$ errors. We will explore these applications in our future studies.

The work is supported by NSFC (grant Nos. 12133006, 11890691, and 11621303) and by 111 project No. B20019. This work made use of the Gravity Supercomputer at the Department of Astronomy, Shanghai Jiao Tong University.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the Tokyo High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This publication has made use of data products from the Sloan Digital Sky Survey (SDSS). Funding for SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England.

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall $z$-band Legacy Survey (MzLS; Prop. ID #2016A-0453; Pf: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF’s NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Duag (Kitt Peak), a mountain with particular significance to the Tohono Oodham Nation.

NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministerio da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eötvös Loránd University, the Institut de Ciencies de l’Espai (IEEC/CSIC), the Institut de Fisica d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF’s NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program The Emergence of Cosmological Structures Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant # 114A11KYSB20160057), and Chinese National Natural Science Foundation (Grant # 11433005).

The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy underContract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO.

ORCID iDs
Kun Xu @ https://orcid.org/0000-0002-7697-3306
Yipeng Jing @ https://orcid.org/0000-0002-4534-3125

References
Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Alhara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Bekki, K., & Couch, W. J. 2011, MNRAS, 415, 1783
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Bielby, R., Hudelot, P., McCracken, H. J., et al. 2012, A&A, 545, A23
Bolton, A. S., Schlegel, D. J., Aubourg, E., et al. 2012, AJ, 144, 144
Boquien, M., Burgarella, D., Roshly, Y., et al. 2019, A&A, 622, A103
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Xu, Zheng, & Jing

10
