The clustering of galaxies in the DESI imaging legacy surveys DR8: I. The luminosity and color dependent intrinsic clustering

ZhaoYu Wang¹, HaoJie Xu¹, XiaoHu Yang¹,²⁺, YiPeng Jing¹,²⁺, Kai Wang³,⁴, Hong Guo⁵, FuYu Dong⁶, and Min He¹

¹Department of Astronomy, School of Physics and Astronomy, and Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai 200240, China;
²Tsung-Dao Lee Institute and Key Laboratory for Particle Physics, Astrophysics and Cosmology, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China;
³Department of Astronomy, Tsinghua University, Beijing 100084, China;
⁴Department of Astronomy, University of Massachusetts Amherst, Amherst 01003, USA;
⁵Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Shanghai 200030, China;
⁶School of Physics, Korea Institute for Advanced Study, Seoul 02455, Korea

Received February 8, 2021; accepted April 23, 2021; published online July 1, 2021

In a recent study, we developed a method that models the impact of photometric redshift uncertainty on the two-point correlation function (2PCF). Using this method, we simultaneously obtained both the intrinsic clustering strength and the photometric redshift errors by fitting the projected 2PCF with two integration depths along the line-of-sight. Herein, we apply this method to the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys Data Release 8 (LS DR8), which is the largest galaxy sample currently available. We separate the galaxies into 20 samples in eight redshift bins, from $z = 0.1$ to $z = 1.0$, and several $z$-band absolute magnitude bins, with $M_z \leq -20$. These galaxies are further divided into red and blue subsamples according to their $M^{0.5}_r - M^{0.5}_z$ colors. We measure the projected 2PCFs for all these galaxy subsamples and fit them using our photometric redshift 2PCF model. We find that the photometric redshift errors are smaller in the red subsamples than in the overall population. In contrast, some systematic photometric redshift errors exist in the blue subsamples, such that some of the subsamples show a significantly enhanced 2PCF on large scales. Therefore, separately focusing only on the red subsamples and on all the subsamples, we find that the biases of the galaxies in these subsamples exhibit clear color, redshift, and luminosity dependencies; the brighter red galaxies at higher redshift are more biased than their bluer and low-redshift counterparts. Apart from the best-fit set of parameters, $\sigma_z$ and $b$, from this state-of-the-art photometric redshift survey, we obtain high-precision intrinsic clustering measurements for these 40 red- and all-galaxy subsamples. These measurements, on large and small scales, hold important information regarding cosmology and galaxy formation that will be used in our subsequent probes in this series.

dark matter, large-scale structure, cosmology

PACS number(s): 95.35.+d, 98.65.-r, 98.80.-k

Citation: Z. Y. Wang, H. J. Xu, X. H. Yang, Y. P. Jing, K. Wang, H. Guo, F. Y. Dong, and M. He, The clustering of galaxies in the DESI imaging legacy surveys DR8: I. The luminosity and color dependent intrinsic clustering, Sci. China-Phys. Mech. Astron. 64, 289811 (2021), https://doi.org/10.1007/s11433-021-1707-6

*Corresponding authors (XiaoHu Yang, email: xyang@sjtu.edu.cn; YiPeng Jing, email: ypjing@sjtu.edu.cn)
1 Introduction

Previous decades have witnessed the flourishing of large galaxy redshift surveys. Many of these surveys, such as the Las Campanas Redshift Survey (LCRS) [1], the 2-degree Field Galaxy Redshift Survey (2dFGRS) [2], the Sloan Digital Sky Survey (SDSS) [3], the DEEP2 Galaxy Redshift Survey (DEEP2) [4, 5], and the VIMOS Extragalactic Redshift Survey (VIPERS) [6] have provided incredible measurements of the galaxy spatial distributions, enabling the exploration of precise cosmological parameters. The baryon acoustic oscillations (BAO) [7-11], gravitational lensing [12-20], power spectrum [21-29], and correlation functions (such as the two-point and three-point correlation functions (2PCF and 3PCF, respectively)) [30-32] are often used to depict the universe.

The 2PCF, the second-order moment statistics, or its integrated version, the projected 2PCF, is typically used to characterize the clustering strength of galaxies. Numerous studies with galaxy samples at low-z and intermediate-z have revealed strong correlations between the clustering strength and galaxy properties, such as color, luminosity, and spectral type [33-43].

To explain the correlation between the clustering strength and the galaxy properties, some statistical models based on the halo-galaxy connection are proposed, such as the halo occupation distribution (HOD) [33, 38, 42, 44-50], the conditional luminosity function model (CLF) [51-55], and halo abundance matching [46, 56]. These models transform galaxy clustering measurements to the informative, physically meaningful parameters encoding the complex physics of galaxy formation and cosmological relationships between galaxies and the dark matter halos, allowing for the physical interpretation of the 2PCF results.

The majority of the previous research on galaxy clustering has focused on galaxies with spectroscopic redshift (called speck hereafter); this is relatively easy to measure in the local universe. For high-z galaxies, owing to the observational limitation and the measurement difficulty, only a relatively small galaxy sample in a small volume can be obtained, such as DEEP2 [4, 5], zCOSMOS [60]. Hence, the 2PCF measurement will become noisy owing to the Poisson noise and the cosmic variance, particularly on large scales.

In general, two types of approaches are available for inferring the galaxy-halo connection at these high redshifts to increase the signal-to-noise ratios of the clustering measurements. One approach is to measure the cross-correlation between the photometric and spectroscopic samples [61-64], and the other is to directly model the angular galaxy clustering measurements [65-69]. However, both approaches are subject to their own limitations. The cross-correlation method is limited by the size of the spectroscopic sample and requires careful treatment of the interlopers, whereas the galaxy-halo connection in the angular clustering method is less well-constrained owing to the lack of redshift information; thus, probing the redshift evolution of galaxies using this approach is difficult.

To overcome the abovementioned problems, we propose a new method in ref. [70] (hereafter called W19) to directly measure the projected 2PCFs from galaxies with only photometric redshifts. In W19, we constructed a realistic mock light-cone through N-body simulation in which the galaxies were populated using the HOD model. Tests showed that our method could not only recover the 2PCF but could also constrain the photometric redshift uncertainties, which could not be achieved using previously existing methods [64, 66, 71].

Herein, we will apply the method developed in W19 to the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys. We aim to obtain the intrinsic (i.e., corrected for the impact of photometric redshift errors) projected 2PCFs for sets of galaxy subsamples of different luminosities and in different redshift bins. In this regard, before the spectroscopic redshifts from the subsequent DESI observations become available, we can still obtain much more accurate clustering measurements than those obtained before, for the current galaxy formation and cosmological studies. This paper is structured as follows. First, we introduce the DESI Legacy Imaging Surveys data and our clustering measurement method in sect. 2. Then, the photometric redshift modeling of the 2PCF is presented in sect. 3. We present the main results in sect. 4 and the summary in sect. 5. Throughout this paper, we assume a ΛCDM cosmology that is consistent with the Planck 2018 results [72]: $\Omega_m = 0.315$, $\Omega_\Lambda = 0.685$, $n_s = 0.965$, $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.674$ and $\sigma_8 = 0.811$.

2 Galaxy samples and clustering measurements

We investigate the luminosity and color dependence of intrinsic galaxy clustering with the DESI Legacy Imaging Surveys Data Release 8 (LS DR8 [73]). LS DR8 provides target catalogs for DESI [74]; the next generation spectroscopic survey for measuring the dark energy effect on the expansion of the universe. All the galaxy positions and properties are extracted from https://www.legacysurvey.org/dr8/files/.

Following the target selection criteria presented in Yang et
al. [75], we select galaxies based on morphological classification as type REX, EXP, DEV, and COMP from TRACTOR [76] fitting results. We also remove those objects at low galactic latitudes and in the vicinity of masked pixels and bright stars (see the selection details in Section 2.1 [75]).

The corresponding photometric redshift (called photoz hereafter) catalog is adopted from the Photometric Redshifts for the Legacy Surveys (PRLS) [77], which utilizes a random forest algorithm [78] to estimate the photoz. Similar to the procedures performed in ref. [75], we select galaxies with \( z \leq 21 \) and \( 0 < z_{\text{phot},\text{median}} \leq 1 \), wherein the \( z_{\text{phot},\text{median}} \) is the median value of the photoz probability distribution function. For those provided with spectroscopic redshifts [79], we replace the photometric redshift with the spectroscopic redshift. We also include additional spectroscopic redshifts that match the 2MASS Redshift Survey (2MRS [79]), the 6dF Galaxy Survey Data Release 3 (6dFGRS [80]), and the 2dFGRS [2].

All magnitudes and colors used in this work are in the AB system and are corrected for galactic extinction. We convert the apparent magnitude to the absolute magnitude using the following equation:

\[
M_x - 5 \log h = m_x - \text{DM}(z) - K^{0.5}_x,
\]

where \( x \equiv r, z \) and DM(z) represent the distance module corresponding to redshift \( z \),

\[
\text{DM}(z) = 5 \log D_L(z) + 25,
\]

with \( D_L(z) \) being the luminosity distance in \( h^{-1} \) Mpc. \( K^{0.5}_x \) represents the K-correction in the \( x \)-band to the sample median redshift \( z \approx 0.5 \). Here, K-correction is obtained for each galaxy according to its three optical (grz) bands, two mid-infrared (W1, 3.4 µm; W2, 4.6 µm) band photometries, and the photoz information obtained using the “K-correct” model (e.g., v4.3) described in ref. [81]. An illustration of the z-band K-correction distributions as a function of the redshift can be found in Figure 2 [75]. In general, the photoz error can impact both the K-correction and the luminosity distance measurements, thus impacting the absolute magnitude estimation. However, as the overall quality of the photoz estimation in ref. [77] is remarkably good, typically, the distance error is smaller than 5%. Consequently, we neglect this error impact on the absolute magnitude estimation for each galaxy.

2.1 Galaxy sample construction with luminosity, redshift, and color

To first investigate the luminosity dependence of the intrinsic galaxy clustering and the photoz errors, we construct galaxy samples in the redshift bins and the galaxy luminosities considering the fact that both the galaxy bias and the photoz quality are primarily dependent on the galaxy luminosity and the redshift. Zhou et al. [77] suggested that overall galaxy photoz values are most accurate with \( z < 21 \) (see Appendix for details). Therefore, we adopt the z-band absolute magnitude \( (M^0_z - 5 \log h) \) as our galaxy luminosity indicator. Because the photoz error is considerably larger than the specz error, a relatively large redshift bin (\( \Delta z = 0.2 \) in this work) for the galaxy samples is preferred to mitigate the effect due to galaxy interlopers from neighboring redshift bins. In principle, we divide the galaxies with absolute magnitude bin width \( \Delta(M^0_z - 5 \log h) = 1 \) mag, but the enormous number of galaxies in some galaxy samples (over 20 million galaxies) allows us to further probe for subtle luminosity dependence. Consequently, in such galaxy samples, we further divide the galaxies into the bin with \( \Delta(M^0_z - 5 \log h) = 0.5 \) mag, regardless of the number of galaxy counts in these samples afterward.

Considering the above, the redshift of the galaxy samples ranges from centering at \( z = 0.2 \) to \( z = 0.9 \); the galaxy samples are constructed in a volume-limited manner. We finish with 20 galaxy samples defined by bins in \( M^0_z - 5 \log h \) and photoz, covering the galaxy Luminosity-redshift diagram, which is shown in Figure 1. More detailed galaxy sample information is summarized in Table 1.

In addition to the luminosity and the redshift, galaxy clustering also depends on the color, the spectra type, the morphology, and the surface brightness [36, 38, 43, 82]. These properties are strongly correlated with each other and exhibit a similar change of clustering when the galaxies are divided based on them [82]. In this work, we choose color because it is the immediately available quantity in addition to being the one that is least vulnerable to measurement uncertainties. Moreover, Blanton et al. [83] found that luminosity and color are the two most predictive quantities for the local galaxy density, with weak residual dependence on the morphology or surface brightness once the luminosity and color are fixed.

Following the color division practice at the low redshift universe, we plot the color-magnitude diagram constructed from a random down-sampling (50000) of previous volume-limited luminosity samples as a function of the redshift, as

---

1) In the TRACTOR fitting procedure, REX represents round exponential galaxies with a variable radius, EXP for exponential profiles (spiral galaxies), DEV for de Vaucouleurs profiles (elliptical galaxies), and COMP for composite profiles that are de Vaucouleurs plus exponential (with the same source center).

2) [https://www.legacysurvey.org/dr8/files/#photometric-redshift-files-8-0-photo-z-sweep-brickmin-brickmax-pz-fits](https://www.legacysurvey.org/dr8/files/#photometric-redshift-files-8-0-photo-z-sweep-brickmin-brickmax-pz-fits)

3) The spectroscopic redshifts are adopted from BOSS, SDSS, WiggleZ, GAMA, COSMOS2015, VIPERS, eBOSS, DEEP2, AGES, 2dFLenS, VVDS, and OzDES, see details in Section 3.2 [77]
shown in Figure 2. The contours demonstrate two galaxy populations with a tight red sequence and a loose blue cloud in the entire redshift, allowing division of the galaxy into red/blue (or passive/star-forming) subsamples using the following equation:

\[ M_r^{0.5} - M_z^{0.5} = -0.8 -0.08\times(M_z^{0.5} - 5\log h) -0.22\times(z - 0.5), \]  

where the \( M_r^{0.5} - 5\log h \) represents the r-band absolute magnitude. It appears that the \( M_r^{0.5} - M_z^{0.5} \) color serves well for the passive/star-forming subsamples division. This may be due to the fact that the 4000 Å break moves to the r-band at \( z \sim 0.5 \), where the K-correction is performed. Notably, at a relatively high redshift, the dust could play a major role in linking the color and star-forming activity such that the star-forming galaxies may exhibit a red color. To what extent this effect could contaminate our color subsamples requires further study and exceeds the scope of this work.

### 2.2 Galaxy clustering measurements

In W19, we demonstrated that the intrinsic clustering and photoz uncertainty could be simultaneously constrained using the projected 2PCF. Herein, we measure the projected 2PCF using the following equation:

\[ v_{p}^{\text{obs}}(r_p|\pi_{\text{max}}) = 2 \int_0^{\pi_{\text{max}}} \xi_{\text{obs}}^{\text{obs}}(r_p, \pi) d\pi, \]  

where \( r_p \) and \( \pi \) represent the transverse and line-of-sight separation between the galaxy pair, and \( \pi_{\text{max}} \) represents the redshift

### Table 1: The subsample selection criteria, number, number density, the related \( \sigma_g \), and the galaxy bias best-fitting results for all galaxies and the red galaxies

| Sample | Redshift | Absolute magnitude \( M_r^{0.5} - 5\log h \) | Numbers (all/red) | Number density (all/red) \( 10^{-4} h^{-3} \text{Mpc}^{-3} \) | \( \sigma_g \) (all/red) | Bias (all/red) |
|--------|----------|-----------------------------------------|------------------|-------------------------|----------------|----------------|
| 1      | [0.1,0.3] | [−21.0,−20.0]                          | 6636956/4162406  | 64.6715/40.5591         | 0.0149/0.0144 | 1.0591/1.2402 |
| 2      | [0.1,0.3] | [−22.0,−21.0]                          | 2502266/1908969  | 24.3824/18.6013         | 0.0110/0.0114 | 1.2065/1.3209 |
| 3      | [0.1,0.3] | [−23.0,−22.0]                          | 264990/226065    | 2.5821/2.2028           | 0.0076/0.0079 | 1.6373/1.7131 |
| 4      | [0.2,0.4] | [−20.5,−20.0]                          | 768418/4408031   | 38.2338/22.036          | 0.0242/0.0201 | 1.1045/1.2765 |
| 5      | [0.2,0.4] | [−21.0,−20.5]                          | 532286/3526789   | 26.6903/17.6305         | 0.0197/0.0178 | 1.7711/1.2579 |
| 6      | [0.2,0.4] | [−22.0,−21.0]                          | 507090/385127    | 25.3457/19.247         | 0.0151/0.0148 | 1.2914/1.4010 |
| 7      | [0.2,0.4] | [−23.0,−22.0]                          | 610255/531282    | 3.0507/2.6559           | 0.0087/0.0078 | 1.6975/0.8089 |
| 8      | [0.3,0.5] | [−21.0,−20.5]                          | 9482610/5793204  | 30.1603/18.4258         | 0.0265/0.0232 | 1.2572/1.3470 |
| 9      | [0.3,0.5] | [−22.0,−21.0]                          | 8633198/6279861  | 27.4587/19.9737         | 0.0190/0.0167 | 1.3494/1.4553 |
| 10     | [0.3,0.5] | [−23.0,−22.0]                          | 1082075/899869   | 3.4416/2.8621           | 0.0099/0.0089 | 1.7283/1.8351 |
| 11     | [0.4,0.6] | [−21.5,−21.0]                          | 8313880/5541092  | 19.0616/12.7043         | 0.0213/0.0194 | 1.3579/1.4873 |
| 12     | [0.4,0.6] | [−22.0,−21.5]                          | 3796681/2715952  | 8.7048/6.2270          | 0.0158/0.0147 | 1.4767/1.6528 |
| 13     | [0.4,0.6] | [−23.0,−22.0]                          | 1813560/1355864  | 4.1580/3.1087          | 0.0099/0.0089 | 1.7552/1.9276 |
| 14     | [0.5,0.7] | [−21.5,−21.0]                          | 10118646/6161085 | 18.1256/11.0364        | 0.0221/0.0212 | 1.3865/1.6075 |
| 15     | [0.5,0.7] | [−22.0,−21.5]                          | 5509779/3693934  | 8.8697/6.5192          | 0.0192/0.0187 | 1.5620/1.7465 |
| 16     | [0.5,0.7] | [−23.0,−22.0]                          | 2915483/202685   | 5.2225/3.5874          | 0.0129/0.0118 | 1.7861/1.9824 |
| 17     | [0.6,0.8] | [−22.0,−21.5]                          | 7093064/4279621  | 10.4952/6.3323         | 0.0230/0.0223 | 1.6575/1.8474 |
| 18     | [0.6,0.8] | [−23.0,−22.0]                          | 3846490/2478615  | 5.6914/3.6675         | 0.0178/0.0165 | 1.9046/2.1385 |
| 19     | [0.7,0.9] | [−23.0,−22.0]                          | 5096945/3130438  | 6.4861/3.9836         | 0.0240/0.0244 | 2.0244/2.3440 |
| 20     | [0.8,1.0] | [−23.0,−22.0]                          | 5854132/3323035  | 6.6040/3.7487         | 0.0261/0.0243 | 1.9435/1.6192 |

**Figure 1** (Color online) Construction of galaxy samples in bins of \( M_r^{0.5} - 5\log h \) and photoz. Each patch represents a luminosity volume-limit sample, with different colors corresponding to different redshift bins. All luminosity samples span the redshift with \( \Delta z = 0.2 \), and the artificial \( z \geq 21 \) flux limit we choose determines how faint we construct the samples. As discussed in sect. 2.1, some luminosity bins have an absolute magnitude bin width of \( \Delta(M_r^{0.5} - 5\log h) = 0.5 \) mag, depending on the number of galaxies. The maximum luminosity becomes \( M_r^{0.5} - 5\log h = -23 \) for all redshift bins. To avoid saturation, here, we only plot the distribution of the randomly selected 0.2% galaxies.
the upper bound of the integration along the line-of-sight direction. Because the methodology requires measurement of the projected 2PCF with two different $r_{\pi,\text{max}}$, we choose $r_{\pi,\text{max}} = 50$ and $100 \, h^{-1} \, \text{Mpc}$ herein for the $w_p$ measurements of all galaxy subsamples. $\xi^{\text{obs}}$ represents the 2D correlation function in the redshift space. We estimate $\xi^{\text{obs}}(r_p, r_{\parallel})$ using the Landy-Szalay estimator [84-86],

$$\xi^{\text{obs}}(r_p, r_{\parallel}) = \frac{DD - 2DR + RR}{RR},$$  

(3)

where $DD$, $RR$, and $DR$ represent the normalized number of galaxy-galaxy, random-random, and galaxy-random pairs within the projected separation $r_p$ and the line-of-sight separation $r_{\parallel}$, respectively.

We use the public random catalogs (randoms-inside4) released along with LS DR8 to consider the survey geometry and the angular selection function. When creating the random catalogs, we adopt the same cut (i.e., we require objects to have at least one exposure in each optical band, remove the low galactic altitude region, and apply the same masks) to ensure that the random catalog has the same angular selection with the corresponding galaxy sample. Here, we use the “shuffled” redshift method to generate the random samples to compensate for the possible systematic selection effects as a function of the redshift in the observational subsamples. The random catalog for each galaxy subsample contains ~5 times as many galaxies. We employ the Python package $\text{CorrFunc}$ to compute the pair-counts [87, 88].

We adopt the sample mean obtained using the jackknife resampling method as our estimation of the projected 2PCF. The covariance matrix is also estimated using the jackknife method. The footprint of the galaxy sample is divided into $N = 120$ spatially contiguous and equal area subregions using the Python package $\text{kmeans\_radec}$5. We measure $w_p$ 120 times, leaving out one different subregion each time, and the covariance is calculated as 119 times the variance of the 120 measurements [36, 38, 40, 49, 89, 90].

### 3 Methodology

The method and performance tests of our photometric redshift modeling of the 2PCFs were presented in W19 using $N$-body simulations; here, we briefly summarize the key points.

We model the photometric redshift uncertainty as a Gaussian distribution, which is

$$P(z_{\text{phot}} - z_{\text{spec}}) = N(0, \sigma_z(1 + z_{\text{spec}})),$$  

(4)

---

4) https://www.legacysurvey.org/dr8/files/#random-catalogs
5) https://github.com/esheldon/kmeans\_radec
where $z_{\text{phot}}$ is the photometric redshift, $z_{\text{spec}}$ is the spectroscopic redshift, $N(\mu, \sigma)$ is a Gaussian distribution with mean $\mu$ and scale $\sigma$, and $\sigma_z$ is the uncertainty of the photometric redshift that we wish to infer.

To calculate the 2PCF, we need to convert the redshift to the comoving distance as follows:

$$D_c(z_{\text{phot}}) = \frac{c}{H_0} \int_0^{z_{\text{phot}}} \frac{dz'}{E(z')}$$

$$\approx D_c(z_{\text{spec}}) + \frac{c(z_{\text{phot}} - z_{\text{spec}})}{H_0 E(z_{\text{spec}})},$$

where $E(z) = \sqrt{\Omega_m (1 + z)^3 + \Omega_L}$. In the 2PCF measurements, we count galaxy pairs as a function of their separations $DD(r_p, r_{\pi})$, where $r_{\pi} = D_c(z_{\pi1}) - D_c(z_{\pi2})$. Because of the photoz error, the probability distribution of the difference between the photoz-derived separation and the specz-derived separation follows:

$$P(R) = \frac{1}{\sqrt{2\pi}\sigma_R} \exp \left( -\frac{R^2}{2\sigma_R^2} \right),$$

$$\sigma_R = \frac{\sqrt{2}c\sigma_c(1 + z_{\text{spec}})}{H_0 E(z_{\text{spec}})},$$

where $R = r_{\pi}(z_{\text{phot}}) - r_{\pi}(z_{\text{spec}})$.

Theoretically, the 2PCF of the galaxies on large scales in the redshift space can be described by

$$\xi^{\text{model}}(r_{\pi}, z_{\pi}) = \frac{1}{2\pi} b^2 \int_{-\infty}^{\infty} \xi^{\text{mm}}(r_p, r_{\pi} - R)P(R) dR,$$

where $\xi^{\text{mm}}$ represents the correlation function of the matter calculated from CAMB [91], and $b$ represents the galaxy linear bias$^6$. The projected 2PCF in the model can also be calculated as:

$$w^{\text{model}}_p(r_{\pi} | r_{\pi, \text{max}}) = 2 \int_0^{r_{\pi, \text{max}}} \xi^{\text{model}}(r_p, r_{\pi}) dr_{\pi}.$$

Here, we constrain the two free parameters, $\sigma_z$ and $b$, by maximizing the posterior distribution:

$$P_{\text{posterior}}(\sigma_z, b) = P_{\text{prior}}(\sigma_z, b) \times L(\sigma_z, b),$$

where the likelihood is as follows:

$$\log L \propto (w^{\text{obs}}_p - w^{\text{model}}_p)^T C^{-1} (w^{\text{obs}}_p - w^{\text{model}}_p),$$

where $w^{\text{obs}}_p$ represents the concatenated vector of the projected 2PCF integrated to a different upper bound, i.e., $w^{\text{obs}}_p = [w^{\text{obs}}_p(50 \ h^{-1} \ \text{Mpc}), w^{\text{obs}}_p(100 \ h^{-1} \ \text{Mpc})]$, and $C$ represents the error covariance matrix of the data vector inferred from the jackknife method [39, 92, 93]. The prior distributions for $\sigma_z$ and $b$ are uniformly distributed and are in the range of $[0.0001, 0.5 \times z_{\text{max}}]$, and $[0.01, 10.0]$, respectively, where $z_{\text{max}}$ represents the maximum redshift in each sample because $\sigma_z$ increases as the redshift becomes higher. Herein, we utilized the emcee [94] package to sample the posterior distribution of $\sigma_z$ and $b$.

### 4 Results

#### 4.1 Model fitting

We begin our investigation with all 20 galaxy samples. In Figure 3, we show the $w^{\text{obs}}_p$ measurements (open squares) in different redshift and magnitude bins, where $r_p$ is from $10^{-1}$ to $10^{-3} \ h^{-1} \ \text{Mpc}$ in 15 logarithmic bins. Here, the results are shown for the projected 2PCFs calculated with two different $r_{\pi, \text{max}} \approx 50/100 \ h^{-1} \ \text{Mpc}$ values. As shown in the figure, galaxies in the brighter magnitude bins typically have strong clustering strengths, whereas the clustering amplitude differences between the two different $r_{\pi, \text{max}}$ become closer as the galaxies become brighter, or in the lower redshift bins, indicating a relatively smaller $\sigma_z$.

Quantitatively, we use $w^{\text{obs}}_p(r_{\pi} | r_{\pi, \text{max}})$ on scales from $r_p = 1 \ h^{-1} \ \text{Mpc}$ to constrain our model parameters, $\sigma_z$ and $b$. The reasons for using these scales are two-fold: (1) on smaller scales, the galaxy bias becomes nonlinear and (2) on larger scales, the 2PCFs may somehow suffer from systematic photometric redshift errors.

For each galaxy sample, we run 10 Markov Chain Monte Carlo (MCMC) chains, each with 50000 steps. The chains mostly converge within the first 10000 steps; these steps are discarded because they are regarded as the burn-in stage. Finally, we obtain approximately 40000 MCMC models. The posterior distributions of the model parameters are shown in Figure 4. The best-fit $\sigma_z$ and galaxy bias are listed in Table 1. We show the best-fitting model predictions in Figure 3 using solid lines. Overall, our model can accurately recover the clustering signals in the different redshift and luminosity bins.

Looking into the best-fitting parameters, as shown in Figure 5(a) using open circles, we found that the photoz errors became larger when the galaxies became fainter and were located at larger redshifts. Interestingly, our model constraints on this value agreed well with the direct measurements of the photoz errors from a small sample of galaxies that have spectroscopic redshifts [77]. The open circles shown in Figure 5(b) are the best-fit galaxy biases in different absolute magnitudes and redshift bins. Obviously, the

---

$^6$ More accurate modeling of the galaxy 2PCFs, particularly on small one-halo term scales, can be achieved using the HOD/CLF models, which will be presented in a subsequent probe.
galaxy biases become larger when the galaxies are brighter and are located at higher redshifts. For the brightest galaxy samples, the bias increases from $\sim 1.6$ at $z \sim 0.2$ to $\sim 2.0$ at $z \sim 0.8$.

Next, we consider the 20 red-galaxy subsamples. Similar to the 20 all-galaxy samples, we also measure their $w_{p}^{\text{obs}}$ with two different $r_{\text{NLmax}} = 50/100$ $h^{-1}$ Mpc values. The results are shown in Figure 6 using open squares. The overall behaviors are relatively similar to those of the all-galaxy samples. Here, we can observe that the gaps between $w_{p}^{\text{obs}}$ for different $r_{\text{NLmax}}$ values become even smaller in these red subsamples.

Following the same procedures for each of the 20 all-galaxy samples, we run 10 MCMC chains, each with 50000 steps for each of our red-galaxy subsamples. After discarding the burn-in stage models, we use the remaining 400000 MCMC models to describe the posterior distributions of the model parameters. The results are shown in Figure 7 with contours. We also list the best-fit $\sigma_z$ and the galaxy bias for our 20 red-galaxy subsamples in Table 1. We illustrate the best-fitting model predictions using solid lines in Figure 6. Here, again, our models accurately recover the clustering signals in the different redshift and luminosity bins.
In addition to the red-galaxy subsamples, we examine the 20 blue-galaxy subsamples. However, as will be demonstrated in Appendix, the blue subsamples may suffer from systematic photometric redshift uncertainties, because of which their clustering measurements cannot be well modeled.

Because of these, we only focus on the red- and the all-galaxy subsamples to model their color dependence. As shown in Figure 5, the red-galaxy subsamples exhibit a smaller $\sigma_z$ and a higher galaxy bias in the same redshift and absolute magnitude bins. The difference is larger in the less bright galaxy subsamples.

Finally, we compare our bias measurements with the ones obtained in a recent similar study [77] in Figure 8. The black squares with error bars shown in this figure are the bias measurements obtained by Zhou et al. [77] from the DECaLS DR7 using an HOD fitting algorithm, whereas our results are denoted using magenta squares. In Zhou et al.’s study, they also used the projected 2PCFs measured from the photometric redshift data and formulated their model constraints.
Figure 5  (Color online) The best-fit $c_r$ (a) and bias (b) measurements for the galaxy subsamples in the DESI Imaging Legacy Surveys within different absolute magnitudes and redshift bins. The different colors represent the galaxy subsamples in different redshift bins.

Figure 6  (Color online) Similar to Figure 3 but considering the 20 red-galaxy subsamples.
However, unlike our two integration depths method, they treated the photoz error as a priori to assign a redshift error to each galaxy in their mock samples. By comparing the observed 2PCFs with the mock 2PCFs, they fitted the related HOD model parameters, and hence, obtained the biases of the galaxies. To obtain fair comparisons with their LRG samples, we only show our best-fit bias measurements for the brightest red-galaxy subsamples. Considering the different galaxy number densities between the two sets of galaxy samples, overall, our results agree quite well with their measurements at redshift $z > 0.6$. In practice, the discrepancy in the highest redshift bin is mainly induced by the galaxy selection in which their galaxy number density in that bin is much lower than ours. In contrast, their biases measured from the two lower redshift bins are somewhat larger than ours. We argue that this might have been caused by the assigned photoz errors being somewhat large in their modeling. Notably, our measurements and the model constraints show that the photoz errors significantly decrease at the lower redshifts.

4.2 The intrinsic projected 2PCFs

After we modeled the projected 2PCFs to different $r_{\text{max}}$ using the $\sigma_2$ and bias parameters, we could, generally, obtain an
estimate of their intrinsic values, i.e., the projected 2PCFs in real space, which correspond to \( w_{\text{p}}^{\text{obs}}(r_p) = w_{\text{p}}^{\text{obs}}(r_p|\pi_{\text{max}} = \infty) \) here. This set of measurements is critical for galaxy formation and the cosmological constraints in the HOD models. We use the following equation to estimate the observational intrinsic projected 2PCFs:

\[
w_{\text{p}}^{\text{obs}}(r_p) = w_{\text{p}}^{\text{obs}}(r_p|\pi_{\text{max}}) \times w_{\text{p}}^{\text{model}}(r_p)/w_{\text{p}}^{\text{model}}(r_p|\pi_{\text{max}}),
\]

where we use the \( r_{\pi_{\text{max}}} = 100 \, h^{-1} \text{ Mpc} \) cut to make the estimation. Note that this equation is accurate if the observed 2PCFs are related with the dark matter 2PCF by a constant bias factor. Unless the bias factor has a very strong scale dependence, particularly on small scales, the correction factor \( w_{\text{p}}^{\text{model}}(r_p)/w_{\text{p}}^{\text{model}}(r_p|\pi_{\text{max}}) \) may be slightly different for the observed galaxies.

To test the reliability and assess the accuracy of using the model \( w_{\text{p}}^{\text{model}}(r_p)/w_{\text{p}}^{\text{model}}(r_p|\pi_{\text{max}}) \) ratio based on the autocorrelation function of the dark matter particles to account for those of the galaxies, we show, in Figure 9, for the 20 all-galaxy samples, a comparison of the model \( w_{\text{p}}^{\text{model}}(r_p|\pi_{\text{max}}) \) ratios and the observed \( w_{\text{p}}^{\text{obs}}(r_p|\pi_{\text{max}}) \) ratios for the two \( r_{\pi_{\text{max}}} = 50 \) and 100 \( h^{-1} \text{ Mpc} \) cuts, respectively. The open circles with error-bars shown in Figure 9 are the results for the observed \( w_{\text{p}}^{\text{obs}}(r_p|\pi_{\text{max}}) \) ratios, which show very weak scale dependence. The purple solid lines represent the results for the model \( w_{\text{p}}^{\text{model}}(r_p|\pi_{\text{max}}) \) ratios. Note that although our model parameters are only constrained using the \( w_{\text{p}}^{\text{obs}}(r_p|\pi_{\text{max}}) \) values on scales between \( r_p = 1 \) and 10 \( h^{-1} \text{Mpc} \), overall, the predicted model ratios agree very well with the observed ratios, i.e., they are within \( 1 - \sigma \) error-bars on all scales. Next, the model predictions of \( w_{\text{p}}^{\text{model}}(r_p|\pi_{\text{max}}) \) for the \( r_{\pi_{\text{max}}} = 100 \, h^{-1} \text{ Mpc} \) cut are shown in Figure 9 using yellow solid lines. In all cases, the correction factors in this set are all found to be larger (i.e., less significant) than those between the \( r_{\pi_{\text{max}}} = 50 \) and 100 \( h^{-1} \text{ Mpc} \) cuts. Additionally, the scale dependence is weaker. Based on these features, the overall correction in eq. (13) is inferred to be reliable, and the errors induced in this step should not exceed \( 1 - \sigma \) errors of the observational data. To be conservative, the error-bars will be enlarged in our extracted intrinsic projected 2PCFs by a factor of two for future use. We also checked the situation for the 20 red-galaxy subsamples; for simplicity, not explicitly shown here, the overall behaviors are relatively similar to the 20 all-galaxy samples.

The intrinsic projected 2PCFs obtained for the red- and all-galaxy subsamples are shown in Figure 10. As previously mentioned, herein, the error-bars in each subsample are enlarged by a factor of 2 with respect to the direct measurements in the \( w_{\text{p}}^{\text{obs}}(r_p|\pi_{\text{max}}) \) with \( r_{\pi_{\text{max}}} = 100 \, h^{-1} \text{ Mpc} \) cut. Overall, in the same redshift and luminosity bin, the red-galaxy subsamples exhibit somewhat stronger clustering strength than the all-galaxy subsamples. The projected 2PCFs on small scales are more enhanced in the red subsamples than that in the all-galaxy subsamples. These features, in general, hold important information with regard to how galaxies with different colors populate the dark matter halos and the related galaxy formation and evolution processes. This topic will be further explored in subsequent probes.

5 Summary

In this study, we constructed 60 volume-limited galaxy sub-
samples, sampling eight redshift bins from $z = 0.1$ to $z = 1.0$, and several $z$-band absolute magnitude bins with $M_z \leq -20$, from the photometric redshift galaxy catalogs of the DESI LS DR8. We measure the projected 2PCFs for all these 60 subsamples with two $r_{\text{max}} = 50$ and 100 $h^{-1}$ Mpc cuts along the line-of-sight directions, respectively. Using the photometric redshift modeling of the 2PCFs developed in W19, we constrain the photoz errors and the galaxy biases for all 60 volume-limited galaxy subsamples. Our main results are summarized as follows.

- Our model can well-describe the clustering properties of the red- and all-galaxy subsamples, whereas the blue-galaxy subsamples might suffer from systematic redshift errors, particularly for low redshift bins, where the clustering on large scales is significantly enhanced and cannot be described by the theoretical models.
- Focusing only on the 40 red- and all-galaxy subsamples, we find that galaxies exhibit better photoz performance and have higher biases when they become redder, brighter, or are in a lower redshift bin.

**Figure 9** (Color online) The ratio of the projected 2PCFs. The purple lines and the open circles with error-bars are representative of the results for the model $w_p^{\text{model}}(r_p|r_{\text{max}})$ ratios and the observed $w_p^{\text{obs}}(r_p|r_{\text{max}})$ ratios for the two $r_{\text{max}} = 50$ and 100 $h^{-1}$ Mpc cuts, respectively. The yellow solid lines represent model predictions of the $w_p^{\text{model}}(r_p|r_{\text{max}})/w_p^{\text{model}}(r_p)$ for $r_{\text{max}} = 100 h^{-1}$ Mpc cut.
Based on the projected 2PCFs for the \( r_{\text{max}} = 100\ h^{-1}\ \text{Mpc} \) cut and the theoretical model prediction of the correction factor, we obtain the intrinsic projected 2PCFs for the 40 red- and all-galaxy subsamples. Here, notably, the photoz redshift outliers were not considered, which, according to ref. [77], are at less than the 1% level. Additionally, we assumed that the photometric redshift errors followed a Gaussian distribution; this is also quite well-demonstrated in ref. [77]. With the intrinsic clustering measurements for the 40 red- and all-galaxy subsamples, we will delve deeper into the HOD/CLF framework to explore the halo-galaxy connection in a subsequent probe.

Figure 10 (Color online) The extracted intrinsic projected 2PCFs for the red- and all-galaxy subsamples. Here, the error-bars in each subsample are enlarged by a factor of two with respect to the direct measurements in the \( w_{\text{obs}}(r_p|r_{\pi,\text{max}}) \) with \( r_{\pi,\text{max}} = 100\ h^{-1}\ \text{Mpc} \) cut.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 11833005, 11890691, 11890692, 11533006, and 11621303), Shanghai Natural Science Foundation (Grant No. 15ZR1446700), the 111 Project (Grant No. B20019), and the China Manned Space Project (Grant No. CMS-CSST-2021-A02). The Photometric Redshifts for the Legacy Surveys (PRLS) catalog used in this paper was produced thanks to funding from the U.S. Department of Energy Office of Science, Office of High Energy Physics (Grant No. DE-SC0007914).

1 S. A. Shectman, S. D. Landy, A. Oemler, D. L. Tucker, H. Lin, R. P. Kirshner, and P. L. Schechter, *Astrophys. J.* **470**, 172 (1996), arXiv:
and T. Yamashita, *Publ. Astron. Soc. Jpn.* 70, S33 (2018), arXiv: 1704.08461.

70. Z. Wang, X. Yang, Y. P. Jing, H. Guo, Z. Zheng, Y. Zu, Z. Li, and C. Liu, *Astrophys. J.* 879, 71 (2019), arXiv: 1906.00516.

71. D. J. Eisenstein, *Astrophys. J.* 586, 718 (2003), arXiv: astro-ph/0212084.

72. N. Aghanim, et al. (Planck Collaboration), *Astron. Astrophys.* 641, A6 (2020), arXiv: 1807.06209.

73. A. Dey, D. J. Schlegel, D. Lang, R. Blum, K. Burleigh, X. Fan, J. R. Findlay, D. Finkbeiner, D. Herrera, S. Janeu, M. Landriau, M. Levi, I. McGreer, A. Meisner, A. D. Myers, J. Moustakas, P. Nugent, A. Patej, E. F. Schlafly, A. R. Walker, F. Valdes, B. A. Weaver, C. Yéche, H. Zou, X. Zhou, B. Abarth, A. M. L. Breiman, Mach. Learn.

76. D. Lang, D. W. Hogg, and D. Mykytyn, The Tractor: Probabilistic

77. R. Zhou, J. A. Newman, Y. Y. Mao, A. Meisner, J. Moustakas, A. D. Myers, A. Prakash, A. R. Zentner, D. Brooks, Y. Duan, M. Landriau, M. E. Levi, F. Prada, and G. Tarle, Mon. Not. R. Astron. Soc. 501, 3309 (2020), arXiv: 2001.06018.

78. L. Breiman, *Mach. Learn.* 45, 5 (2001).

79. J. P. Huchra, L. M. Macri, K. L. Masters, T. H. Jarrett, P. Berlind, M. Calkins, A. C. Crook, R. Cutri, P. Erdogdu, E. Falco, T. George, C. M. Hutcheson, O. Lahav, J. Mader, J. D. Mink, N. Martimbeau, S. Schneider, M. Skrutskie, S. Tokarz, and M. Westover, Astrophys. J. Suppl. Ser. 199, 26 (2012), arXiv: 1108.0669.

80. D. H. Jones, M. A. Read, W. Saunders, M. Colless, T. Jarrett, Q. A. Parker, A. P. Fairall, T. Mauch, E. M. Sadler, F. G. Watson, D. Burton, L. A. Campbell, P. Cass, S. M. Croom, J. Dawe, K. Fiegert, L. Frankcombe, M. Hartley, J. Huchra, D. James, E. Kirby, O. Lahav, J. Lucey, G. A. Mann, L. Moore, B. A. Peterson, S. Prior, D. Proust, K. Russell, S. Safouris, K. Wakamatsu, E. Westra, and M. Williams, Mon. Not. R. Astron. Soc. 399, 683 (2009), arXiv: 0903.5451.

81. M. R. Blanton, and S. Roweis, Astronom. J. 133, 734 (2007), arXiv: astro-ph/0606170.

82. I. Zehavi, M. R. Blanton, J. A. Frieman, D. H. Weinberg, H. J. Mo, M. A. Strauss, S. F. Anderson, J. Annis, N. A. Bahcall, M. Bernardi, J. W. Briggs, J. Brinkmann, S. Burles, L. Carey, F. J. Castander, A. J. Connolly, I. Csabai, J. J. Dalcanton, S. Dodelson, M. Doi, D. Eisenstein, M. L. Evans, D. P. Finkbeiner, S. Friedman, M. Fukugita, J. E. Gunn, G. S. Hennessy, R. B. Hindsley, Ivezic, S. Kent, G. R. Knapp, R. Kron, P. Kunz, D. Q. Lamb, R. F. Leger, D. C. Long, J. Lovey, D. R. Lupton, T. McKay, A. Meiksin, A. Merrelli, J. A. Munn, V. Narayanan, M. Newcomb, R. C. Nichol, R. Owen, J. Peoples, A. Pope, C. M. Rockosi, D. Schlegel, D. P. Schneider, R. Scoccolimaro, R. K. Sheth, W. Siegmund, S. Smee, Y. Snir, A. Streblyanska, K. SubbaRao, A. S. Szalay, I. Szapudi, M. Tegmark, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, Y. Yanny, and D. G. York, *Astrophys. J.* 571, 172 (2002), arXiv: astro-ph/0106476.

83. M. R. Blanton, D. Eisenstein, D. W. Hogg, D. J. Schlegel, and J. Brinkmann, *Astrophys. J.* 629, 143 (2005), arXiv: astro-ph/0310453.

84. S. D. Landy, and A. S. Szalay, *Astrophys. J.* 412, 64 (1993).

85. I. Szapudi, and A. S. Szalay, *Astrophys. J.* 494, L41 (1998).

86. M. Kerscher, I. Szapudi, and A. S. Szalay, *Astrophys. J.* 535, L13 (2000), arXiv: astro-ph/9912088.

87. M. Sinha, and L. Garrison, in *Software Challenges to Exascale Computing*, edited by A. Majumdar, and R. Arora (Springer, Singapore, 2019).

88. M. Sinha, and L. H. Garrison, Mon. Not. R. Astron. Soc. 491, 3022 (2020), arXiv: 1911.03545.

89. H. Xu, Z. Zheng, H. Guo, J. Zhu, and I. Zehavi, Mon. Not. R. Astron. Soc. 460, 3647 (2016), arXiv: 1510.00393.

90. H. Guo, C. Li, Z. Zheng, H. J. Mo, Y. P. Jing, Y. Zu, S. H. Lim, and H. Xu, *Astrophys. J.* 846, 61 (2017), arXiv: 1707.01999.

91. A. Lewis, A. Challinor, and A. Lasenby, *Astrophys. J.* 537, 473 (2000), arXiv: astro-ph/9911177.

92. P. Norberg, C. M. Baugh, E. Gaztanaga, and D. J. Croton, Mon. Not. R. Astron. Soc. 396, 19 (2009), arXiv: 0810.1885.

93. H. Guo, I. Zehavi, Z. Zheng, D. H. Weinberg, A. A. Belkina, M. Blanton, Y. Chen, D. J. Eisenstein, S. Ho, E. Kazin, M. Manera, C. Maraston, C. K. McBride, S. E. Nuza, N. Padmanabhan, J. K. Parejko, W. J. Percival, A. J. Ross, N. P. Ross, L. Samushia, A. G. Sánchez, D. J. Schlegel, D. P. Schneider, R. A. Sibthorpe, M. E. C. Swanson, J. L. Tinker, R. Tojeiro, D. A. Wake, M. White, N. A. Bahcall, D. Bizyaev, H. Brewington, K. Bundy, L. N. A. Da Costa, G. Ebelke, V. Malanushenko, V. Malanushenko, D. Oravetz, G. Rossi, A. Simmons, S. Snedden, A. Streblyanska, and D. Thomas, Astrophys. J. 767, 122 (2013), arXiv: 1212.1211.

94. D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. 125, 306 (2013), arXiv: 1202.3665.

Appendix Results for the blue subsamples

Similar to those clustering measurements and model fittings performed for the 40 red- and all-galaxy subsamples in sect. 4.1, we also examined the 20 blue-galaxy subsamples. Similar to the 20 all-galaxy samples, we also measured their

\[ \sigma_{\text{cl}} \] with two different \( r_{\text{lim}} \) Mpc values; the corresponding results are shown in Figure a1 using open squares. The overall behaviors are approximately similar to those of the all-galaxy subsamples. Here, we can observe that
the gap between $w_{\text{obs}}$ for different $r_{\text{max}}$ values is larger in these blue subsamples. Additionally, we find in a few subsamples that the projected 2PCFs on large scales are significantly boosted.

Following the same procedures for each of the 20 all-galaxy samples, we run 10 MCMC chains, each with 50000 steps for each of our blue-galaxy subsamples. After discarding the burn-in stage models, we use the remaining 400000 MCMC models to describe the posterior distributions of the model parameters. The results are shown in Figure a2 with contours. Here, again, we show the best-fitting model predictions using solid lines in Figure a1. Unlike the 40 all- and red-galaxy subsamples, here, there are some blue subsamples for which our models cannot describe their behaviors well, particularly those with boosted clustering strengths on large scales. For these subsamples, our constraints on the photoz error and bias parameters are also markedly poor.

Note that even in the framework of a more sophisticated HOD model, the parameters, in general, can change the shape of $w_p$ on small scales and the amplitude on large scales. Because the clustering of galaxies on large scales is modeled through the combination of a constant galaxy bias and the
autocorrelation of the dark matter particles, the very different shape of \( w_p \) on scales \( r_p > 3 \ h^{-1} \text{Mpc} \) shown in Figure a1 will not be well modeled in the HOD models. Thus, we believe that there should be some systematic photoz errors or photometry errors in these blue subsamples such that their projected 2PCFs on large scales are boosted and cannot be well modeled by theory. Because of this, we omit further analysis for these 20 blue-galaxy subsamples in this study.