The Astrophysical Journal, 961:27 (20pp), 2024 January 20 https://doi.org/10.3847/1538-4357/ad0896
© 2024. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

LATIS: Constraints on the Galaxy–Halo Connection at $z \sim 2.5$ from Galaxy–Galaxy and Galaxy–Lyα Clustering

Andrew B. Newman[1], Mahdi Qezlou[1,2], Nima Chartab[1], Gwen C. Rudie[1], Guillermo A. Blanc[1,3], Simeon Bird[2], Andrew J. Benson[1], Daniel D. Kelson[1], and Brian C. Lemaux[4,5]
[1] Observatories of the Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101, USA; anewman@carnegiescience.edu
[2] Department of Physics and Astronomy, University of California Riverside, 900 University Ave., Riverside, CA 92521, USA
[3] Departamento de Astronomía, Universidad de Chile, Camino del Observatorio 1515, Las Condes, Santiago, Chile
[4] Gemini Observatory, NSF's NOIRLab, 670 N. A'ohoku Place, Hilo, HI 96720, USA
[5] Department of Physics and Astronomy, University of California, Davis, One Shields Ave., Davis, CA 95616, USA
Received 2023 September 6; revised 2023 October 27; accepted 2023 October 30; published 2024 January 11

Abstract

The connection between galaxies and dark matter halos is often quantified using the stellar mass–halo mass (SMHM) relation. Optical and near-infrared imaging surveys have led to a broadly consistent picture of the evolving SMHM relation based on measurements of galaxy abundances and angular correlation functions. Spectroscopic surveys at $z \gtrsim 2$ can also constrain the SMHM relation via the galaxy autocorrelation function and through the cross-correlation between galaxies and Lyα absorption measured in transverse sight lines; however, such studies are very few and have produced some unexpected or inconclusive results. We use ∼3000 spectra of $z \sim 2.5$ galaxies from the Lyα Tomography IMACS Survey (LATIS) to measure the galaxy–galaxy and galaxy–Lyα correlation functions in four bins of stellar mass spanning $10^{9.2} \lesssim M_*/M_\odot \lesssim 10^{10.5}$. Parallel analyses of the MultiDark *N*-body and ASTRID hydrodynamic cosmological simulations allow us to model the correlation functions, estimate covariance matrices, and infer halo masses. We find that results of the two methods are mutually consistent and broadly accord with standard SMHM relations. This consistency demonstrates that we are able to measure and model Lyα transmission fluctuations δ_F in LATIS accurately. We also show that the galaxy–Lyα cross-correlation, a free by-product of optical spectroscopic galaxy surveys at these redshifts, can constrain halo masses with similar precision to galaxy–galaxy clustering.

Unified Astronomy Thesaurus concepts: Galaxy dark matter halos (1880); Lyα forest (980); Lyman-break galaxies (979)

1. Introduction

Although the luminosity and stellar mass of a galaxy are more readily observed, the mass of its dark matter halo is considered to be the most fundamental parameter for theoretical galaxy evolution models. Measurements of the evolving statistical connection between galaxies' luminosities or stellar masses and their halo masses are therefore crucial for testing and constraining such models. The simplest measure of the galaxy–halo connection is the stellar mass–halo mass (SMHM) relation, which relates the mass of a dark matter halo to its average stellar content.

The main empirical tools for constraining the SMHM relation—and the only ones generally applicable beyond $z \sim 1$—have been measures of galaxy abundances and clustering. The subhalo abundance matching hypothesis (Vale & Ostriker 2004; Conroy & Wechsler 2009; Behroozi et al. 2010; Guo et al. 2010; Moster et al. 2010) enables the SMHM relation to be derived from an observed luminosity or stellar mass function by assuming a monotonic relation between these quantities and the halo[6] mass or related properties. Galaxy clustering is also used to constrain the SMHM relation by relating the galaxy–galaxy two-point correlation function to that of dark matter halos; more massive halos exhibit stronger clustering (e.g., Mo & White 1996) because they occupy denser environments (Pujol et al. 2017; Shi & Sheth 2018). The galaxy correlation function can be modeled analytically using a halo occupation distribution (HOD) model (Benson et al. 2000; Berlind & Weinberg 2002; Cooray & Sheth 2002; Zheng et al. 2007) or the related conditional luminosity (or stellar mass) function (Yang et al. 2003; Cooray 2006; van den Bosch et al. 2007). Galaxy clustering may also be modeled by matching directly to subhalos in cosmological simulations (Reddick et al. 2013; Guo et al. 2016; Zheng & Guo 2016). These techniques have been widely applied to galaxy survey data to constrain the galaxy–halo connection (e.g., Jing et al. 1998; Zehavi et al. 2005, 2011; Zheng et al. 2007; Conroy & Wechsler 2009; Wake et al. 2011; Leauthaud et al. 2012; Behroozi et al. 2013a; Moster et al. 2013; Behroozi et al. 2019; Shuntov et al. 2022, and see references below).

Beyond quasars, studies of galaxy clustering at high redshifts initially focused on Lyman-break galaxies (LBGs; Adelberger et al. 1998, 2005; Giavalisco et al. 1998; Ouchi et al. 2001; Foucaud et al. 2003), extremely red objects (EROs) seen in infrared surveys (Daddi et al. 2000; Moustakas & Somerville 2002; Roche et al. 2002; Zheng 2004), and submillimeter-selected sources (SMGs; Webb et al. 2003; Blain et al. 2004). These initial studies showed that LBGs occupy a substantial fraction of halos with masses $M_h \approx 10^{11-12}\,M_\odot$, whereas EROs and SMGs occupy somewhat more massive halos. UV-brighter LBGs were found to occupy more massive halos (Giavalisco & Dickinson 2001; Foucaud et al. 2003; Ouchi et al. 2004;

[6] We use "halo" to refer to distinct halos and subhalos collectively.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Adelberger et al. 2005; Allen et al. 2005; Lee et al. 2006; Hildebrandt et al. 2009). With the advent of deep and wide near-infrared imaging surveys, the focus shifted to quantifying stellar mass–dependent clustering at redshifts $z < 2$, with samples usually selected based on photometric redshifts rather than color criteria tailored to galaxy subpopulations (Foucaud et al. 2010; Wake et al. 2011; Bielby et al. 2014; McCracken et al. 2015). More recently such efforts have been extended to $z = 5$ (Ishikawa et al. 2017; Cowley et al. 2018; Shuntov et al. 2022). These studies revealed that the SMHM relation retains a similar shape over time, with a peak $M_*/M_h \approx 0.02$ occurring around a characteristic mass $M_h \approx 10^{12}\ M_\odot$ that gradually increases with redshift, and they also provided constraints on quenching and the evolution of satellite galaxies.

All of these studies measured the angular correlation function from imaging data in which the distances to most individual galaxies are poorly known. Using a sufficiently large spectroscopic survey, it possible to measure 3D galaxy–galaxy correlation functions. Usually such functions are projected along the line of sight, using well-measured redshifts to eliminate most unassociated galaxy pairs. In addition to measuring the halo mass of the targeted population via its autocorrelation (Trainor & Steidel 2012; Bielby et al. 2013; Durkalec et al. 2018), cross-correlations enable the host halo mass of rare objects to be measured (e.g., hyperluminous quasars; Trainor & Steidel 2012). Precise measurements of high-redshift galaxy clustering and redshift-space distortions are a premier probe of inflation and early dark energy and motivate massive spectroscopic surveys (e.g., Schlegel et al. 2022).

To date only Durkalec et al. (2015, 2018) have measured stellar mass–dependent clustering within a spectroscopic galaxy sample at $z > 2$. Although their initial study (Durkalec et al. 2015) suggested reasonable agreement with the canonical models discussed above, their final analysis based on the full VIMOS Ultra Deep Survey (VUDS; Durkalec et al. 2018) indicated extremely low halo masses for the lowest-mass galaxies with $M_* \lesssim 10^{9.2}\ M_\odot$, suggesting a remarkably high stellar mass fraction $M_*/M_h \approx 0.1$. The authors interpreted this as evidence of very efficient star formation in low-mass $z \approx 3$ galaxies, which might have been missed by earlier surveys that did reach the lowest masses. If confirmed, this finding would have major implications for galaxy formation models, but it clearly requires verification.

Another tool to constrain the halo masses of high-z galaxies is to measure the cross-correlation between galaxies and the surrounding H I absorption, as observed in transverse sight lines toward quasars or background galaxies. Excess H I is found around $z > 2$ galaxies to very large distances. The simplest tracer of the galaxy–H I cross-correlation is the Lyα flux (Adelberger et al. 2003, 2005; Crighton et al. 2011; Bielby et al. 2017; Chen et al. 2020), but Voigt profile fitting (Rudie et al. 2012) or pixel optical depth methods (Rakic et al. 2012) can be applied to high-resolution quasar spectra. These studies have examined the galaxy–H I correlation at small separations to constrain the spatial extent, kinematics, and temperature of inflows and outflows in the circumgalactic region. At separations well beyond $\sim$1 cMpc (comoving Mpc), on the other hand, the correlation is instead dominated by large-scale structure and is expected to depend mainly on the halo mass (Momose et al. 2021b).

Kim & Croft (2008) made an initial estimate of the halo mass of LBGs based on the surrounding Lyα absorption, although the data of Adelberger et al. (2003) permitted only loose constraints with an uncertainty of 0.6 dex. Rakic et al. (2013) refined the technique and applied it to an improved data set, deriving a halo mass estimate with a precision of 0.2 dex that was consistent with estimates based on galaxy–galaxy clustering. Momose et al. (2021a) were the first to measure the galaxy–Lyα cross-correlation in bins of stellar mass using spectra from the COSMOS Lyα Mapping and Tomography Observations (CLAMATO) survey (Lee et al. 2018; Horowitz et al. 2022). Their models based on a cosmological hydrodynamic simulation showed differences from the measured cross-correlations at distances up to several comoving megaparsecs, but these were not used to infer halo masses.

Altogether, optical to near-infrared imaging surveys have assembled a relatively consistent picture of the SMHM around $z \approx 2$–3, whereas there are very few measurements of stellar mass–dependent galaxy–galaxy or galaxy–Lyα clustering based entirely on spectroscopic surveys at these redshifts, and these have yielded unexpected or inconclusive results. In this paper we use LBG spectra from the Lyα Tomography IMACS Survey (LATIS; Newman et al. 2020) to examine both the galaxy–galaxy and galaxy–Lyα correlations consistently. Key advantages of LATIS include a large sample of $\sim$3000 galaxies at $z \approx 2$–3, precise and calibrated redshifts, and coverage of the Lyα forest at a spectral resolving power of $R \sim 900$.

Our goals are (1) to constrain the SMHM relation at $z = 2.5$ using galaxy–galaxy and galaxy–Lyα clustering, both to inform the galaxy–halo connection generally and as a key input to future papers examining the galaxy–intergalactic medium (IGM) connection using LATIS; (2) to assess the consistency of the methods as a precise test of the methodology and of systematic errors in the LATIS Lyα forest data; and (3) to evaluate the relative precision of galaxy–galaxy and galaxy–Lyα clustering as constraints on halo occupation to help inform future surveys.

We introduce our galaxy–galaxy clustering analysis in Section 2. We then turn to measurements of the galaxy–Lyα cross-correlation (Section 3), calculation of the halo–Lyα cross-correlation in a cosmological simulation (Section 4), and the resulting constraints on halo masses (Section 5). We compare the SMHM relations from galaxy–galaxy and galaxy–Lyα clustering to one another and to the literature (Section 6) and summarize their implications (Section 7).

2. Galaxy–Galaxy Clustering

We begin with our methods for estimating the observed galaxy and simulated (sub)halo correlation functions. A key challenge is the construction of an accurate covariance matrix, which we address using mock surveys within a large cosmological simulation. After developing the method, we first estimate the typical halo mass of the LATIS galaxies, undifferentiated by stellar mass, based on their autocorrelation function. This measurement is required for some analyses in which the galaxy sample is not divided by mass, e.g., Newman et al. (2022). Second, we consider subsets of galaxies defined by stellar mass and derive halo masses using a new method that incorporates all $N(N+1)/2$ auto- and cross-correlations among the N subsamples.

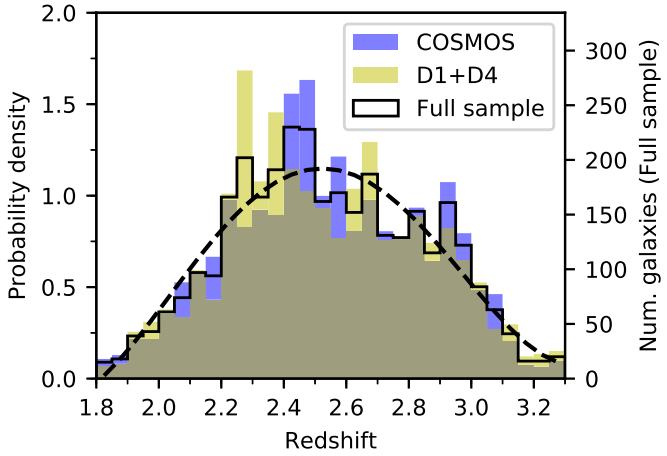

Figure 1. Redshift distribution of galaxies used to measure galaxy–galaxy clustering. The full sample is shown along with the COSMOS and D1+D4 subsamples plotted separately to illustrate field-to-field variations. Each histogram is a normalized probability density function (pdf); the right axis shows the corresponding number of galaxies per bin in the full sample. The dashed line shows the polynomial fit used to generate random points.

2.1. Observations

We measure the clustering of LATIS LBGs with high-confidence (`zqual` = 3 or 4) redshifts $z = 2$–3, approximately the central 90% of the redshift distribution shown in Figure 1, combining all three survey fields (COSMOS and Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) D1 and D4). Redshifts are derived from full spectral modeling, and for the high-confidence subset used in this paper, we visually identified multiple spectral features. Newman et al. (2020) estimated that the catastrophic error rate for the high-confidence redshifts was <2%, which we will consider as negligible. We exclude galaxies that were observed only as part of "bright target" masks intended for poor weather conditions (Newman et al. 2020), since they do not cover the entire survey area; galaxies whose spectra are heavily blended; quasars; and targets not covered by the near-infrared imaging required to estimate stellar masses robustly, which affects only the CFHTLS-D4 field.

The photometric catalogs and derivation of stellar masses from spectral energy distribution fitting are described by N. Chartab et al. (2023, in press). Figure 2 shows the stellar mass distribution. Although its full range spans $\log M_*/M_\odot = 8.4$–11.4, galaxies are concentrated around the median $M_* = 10^{9.7}\ M_\odot$, and the central 90% range is $\log M_*/M_\odot = 9.1$–10.4, a factor of 20. The total number of galaxies used to measure galaxy–galaxy clustering is 2991.

2.2. Clustering Estimator

The two-point correlation function $\xi(r_p, \pi)$ is the excess probability of finding a galaxy at a transverse separation r_p and line-of-sight separation π (in redshift space) from another galaxy, relative to a uniformly distributed population. We will primarily use the two-point correlation function projected along the line of sight:

$$w_p(r_p) = \int_{-\pi_{\rm max}}^{\pi_{\rm max}} \xi(r_p, \pi) d\pi. \quad (1)$$

Note that in this convention w_p has units of distance. The integral limits $\pm\pi_{\rm max}$ are generally selected to be large enough to desensitize the measurement to redshift-space distortions and galaxy redshift errors, while being small enough to avoid adding excessive noise from uncorrelated pairs. Typical values of $\pi_{\rm max}$ are 15–30 h^{-1} cMpc (e.g., Trainor & Steidel 2012; Marulli et al. 2013; Kashino et al. 2017; Durkalec et al. 2018), and we set $\pi_{\rm max} = 20\ h^{-1}$ cMpc. Our mock surveys, described below, showed that the precision of the inferred halo masses is not very sensitive to the choice of $\pi_{\rm max}$ within this range.

To estimate $\xi(r_p, \pi)$, we generate a large number of unclustered random points (see below), convert the celestial coordinates and redshifts of the observed galaxies and random points into comoving Cartesian coordinates, and use the `Corrfunc` code (Sinha & Garrison 2020) to compute the Landy & Szalay (1993) estimator rapidly. In the case of an autocorrelation, the simple version of this estimator is:

$$\xi(r_p, \pi) = \frac{DD - 2DR + RR}{RR}, \quad (2)$$

where DD, DR, and RR are the normalized number of galaxy–galaxy, galaxy–random, and random–random pairs, respectively, within a given bin of r_p and π. These counts are normalized to account for the oversampling of random points: $DD = N_{DD}/[N_D(N_D - 1)]$, $DR = N_{DR}/(N_D N_R)$, and $RR = N_{RR}/[N_R(N_R - 1)]$, where N_{DD}, N_{DR}, and N_{RR} represent the raw pair counts (e.g., Kerscher et al. 2000). In the case of a cross-correlation between two galaxy or halo samples, we have:

$$\xi(r_p, \pi) = \frac{D_1D_2 - D_1R_2 - D_2R_1 + R_1R_2}{R_1R_2}, \quad (3)$$

where the normalized counts are $D_1D_2 = N_{D_1D_2}/(N_{D_1}N_{D_2})$, $D_1R_2 = N_{D_1R_2}/(N_{D_1}N_{R_2})$ (analogous for D_2R_1), and $R_1R_2 = N_{R_1R_2}/(N_{R_1}N_{R_2})$, and $N_{D_1D_2}$, $N_{D_1R_2}$, $N_{D_2R_1}$, and $N_{R_1R_2}$ are the raw galaxy–galaxy, galaxy–random, and random–random pair counts. Once the 2D correlation function $\xi(r_p, \pi)$ has been estimated, we integrate along the line of sight to obtain $w_p(r_p)$. Numerically the integration consists of summing estimates of ξ over π bins with a width of 1 h^{-1} cMpc.

An essential aspect of estimating ξ is the generation of random coordinates of unclustered objects that must respect the survey geometry and redshift distribution. For each of the three survey fields, we randomly draw $N_R = 100 \times N_D$ celestial coordinates within the survey area, excluding masked regions around bright stars (see Appendix B), where N_D is the number of observed galaxies. The redshift distribution of an unclustered population would not be uniform due to the LATIS target selection function, which depends on a galaxy's r-band magnitude and colors. The best estimate of the selection function is empirical: the redshift pdf of the observed sample. Although this estimate inevitably and undesirably includes large-scale structure, its influence can be mitigated by averaging over independent fields and smoothing the distribution (e.g., Trainor & Steidel 2012; Marulli et al. 2013). Figure 1 shows the redshift pdf constructed in bins of $\Delta z = 0.05$ and modeled with a polynomial of degree $n = 4$, which we use to draw random samples. We find that our results are extremely insensitive to n for any reasonable value $n \geqslant 2$. Although only $z = 2$–3 galaxies are used to estimate ξ, we include galaxies over a slightly wider redshift range when fitting this polynomial to constrain the endpoints better.

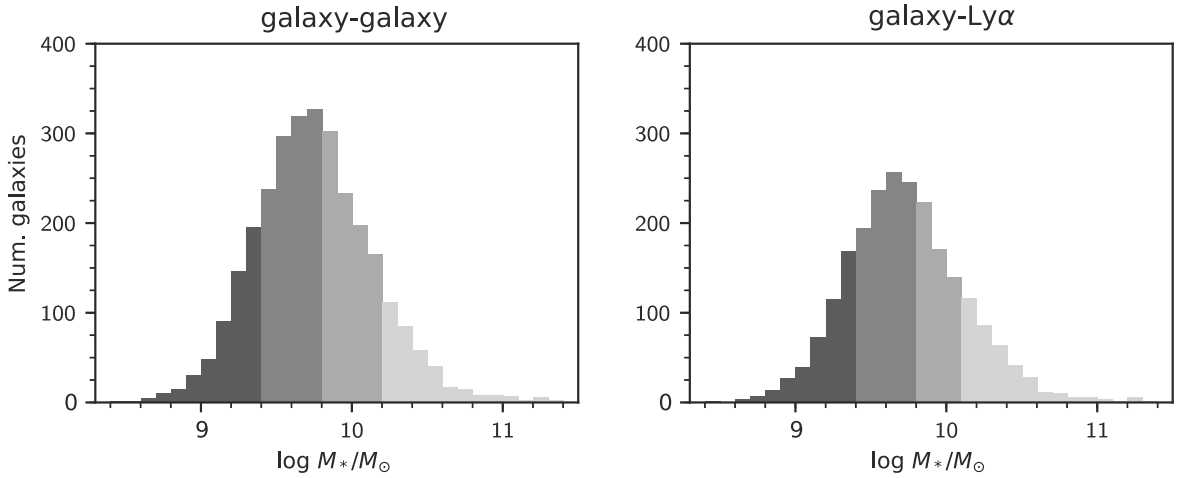

Figure 2. Stellar mass distributions of galaxies used in the galaxy–galaxy (left panel) and galaxy–Lyα (right) clustering analyses. Different shadings indicate the four stellar mass bins.

We use a weighting scheme to account for nonuniformity in the survey sensitivity and sampling. Appendix B describes our estimates of the target sampling rate (TSR) and spectroscopic success rate (SSR) as a function of celestial coordinates. The TSR gives the fraction of the LATIS parent sample that was observed, while the SSR gives the fraction of the observed sample for which a high-confidence spectroscopic redshift was determined. The product of these rates is the effective sampling rate (ESR). The ESR exhibits a variety of angular dependencies arising from slit mask design constraints, instrumental effects, and time-dependent observing conditions. Similar to other studies (e.g., Marulli et al. 2013) we account for angular variation of the ESR by weighting galaxy i by $w_i = \mathrm{ESR}^{-1}$. Specifically, in the weighted estimator of $\xi(r_p, \pi)$, the pair counts above are replaced by the sum of the product of weights $\sum \mathrm{ESR}_i^{-1} \times \mathrm{ESR}_j^{-1}$, and all random points receive the mean weight. (Recall that random points are already constrained to lie within the observed survey area.)

Beyond contributing to these large-angle variations in the ESR, slit mask design constraints lead to a deficit of close galaxy pairs in LATIS relative to the parent sample. However, this deficit requires no correction because it exceeds 5% only for separations less than 0.′3 (see also Newman et al. 2022), which is smaller than any separation bin that we will consider.

These procedures are initially used to estimate $w_p(r_p)$ in each of the three LATIS fields separately, and we then compute a final $w_p(r_p)$ from a weighted mean over the fields, where field i receives weight $\mathcal{W}_i$. In the limit of Poisson errors, inverse-variance weighting would assign a weight proportional to the number of pairs in a given radial bin. In the limit that r_p is much smaller than the survey volume, this expected pair number is proportional to $n_i^2 V_i$, where n_i is the galaxy space density and V_i is the volume of field i. We therefore use $\mathcal{W}_i = n_i^2 V_i / \sum_{i=1}^{3} n_i^2 V_i$; the weights for the COSMOS, D1, and D4 fields are 0.59, 0.36, and 0.05, respectively. Experiments with the mock surveys introduced below shows that these simple weights are close to optimal in terms of minimizing variance.

2.3. Halo Correlation Functions

We compute halo correlation functions using the MultiDark Planck 2 simulation (MDPL2; Klypin et al. 2016), which was selected to enable a sufficient number of mock surveys (see below) distributed within its 1 h^{-3} Gpc^3 volume. We use the snapshot at $z_{\rm sim} = 2.535$, which is very close to the mean redshift of the galaxies $\langle z \rangle = 2.52$. We compute the correlation functions of all halos (i.e., distinct halos and subhalos) in the `RockStar` (Behroozi et al. 2013b) catalog, as appropriate for our comparison to galaxy clustering; for reference, about 8% of halos with masses $M_h = 10^{11.7}\ M_\odot$ are subhalos. We define M_h using the virial mass `Mvir`. Due to the small expected fraction of subhalos in the relevant mass and redshift ranges along with our limited constraining power on submegaparsec scales, we prefer our single-parameter approach over fitting a more flexible HOD model, which could suffer from significant degeneracies.

First, we compute $w_p(r_p)$ for thresholded samples of halos, i.e., those with a virial mass $M_{\rm vir} > M_h^{\rm thresh}$ for a selected $M_h^{\rm thresh}$. We find the positions of these halos in redshift space by perturbing their comoving z-coordinates by $v_z(1+z)/H(z)$, where v_z is the physical halo velocity in the z-direction. For speed, we randomly select a subsample of at most 5×10^5 halos, which we find is sufficient to estimate w_p to $\sim$1% precision. We use `Corrfunc` to compute w_p on a grid of $\log M_h^{\rm thresh}/M_\odot$ spanning 10.7 to 13.5 in steps of 0.1, and we use linear interpolation to evaluate w_p between grid points.

Second, we consider the cross-correlations between two halo samples, each consisting of an essentially "mono-mass" halo population defined by a narrow range of $\pm$0.1 dex around a specified mass M_h. We compute $w_p(r_p)$ on a 2D grid and use bilinear interpolation where necessary.

Our galaxy sample spans a range of redshifts $z = 2$–3 over which clustering evolves, whereas the halo clustering is calculated at a single redshift. In the limit of linear growth and biasing, the halo autocorrelation function is:

$$\xi_{\rm h}(z) = b(z)^2 \xi_{\rm dm}(z) = \left[\frac{b^2(z) D^2(z)}{b^2(z_{\rm sim}) D^2(z_{\rm sim})} \right] \xi_{\rm h}(z_{\rm sim}), \qquad (4)$$

where $b(z)$ is the bias, $D(z)$ is the linear growth factor, and $\xi_{\rm dm}$ is the dark matter correlation function. We find that $\langle b^2(z_i) D^2(z_i) \rangle / [b^2(z_{\rm sim}) D^2(z_{\rm sim})] = 1.0002$, where the average is taken over the redshifts z_i of galaxies in our sample, which means that we can directly compare the galaxy and halo autocorrelation functions when the full galaxy sample is considered. However, when we consider auto- and cross-

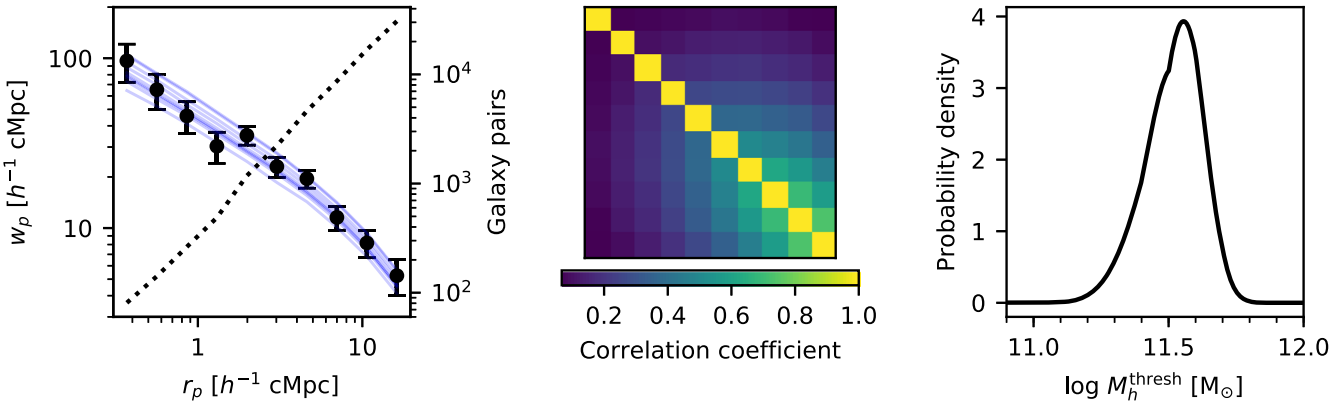

Figure 3. Left: the projected galaxy–galaxy correlation function $w_p(r_p)$ for the entire LATIS sample, undivided by stellar mass. Black points show measured values with error bars $C_{ii}^{0.5}$ derived from the mock surveys. These data are modeled by the autocorrelation function of mass-thresholded halos; blue curves show 10 posterior samples. The dotted line and right axis show the number of galaxy pairs (within $|\Delta z| < \pi_{\rm max}$) in each r_p bin. Middle: the correlation matrix $C_{ij}/(C_{ii}C_{jj})^{0.5}$ derived from the mock surveys. Each row or column corresponds to a bin of r_p, increasing down and to the right. Right: posterior probability density for the threshold halo mass $M_h^{\rm thresh}$.

correlations of galaxies in bins of stellar mass, we make small corrections for clustering evolution, because the mean redshift of the galaxies varies slightly with stellar mass. We do this by rescaling the halo w_p by the factor $f = b(z_1)D(z_1)b(z_2)D(z_2)/[b^2(z_{\rm sim})D^2(z_{\rm sim})]$, where z_1 and z_2 are the mean redshifts of the galaxy subsamples. We use the Tinker et al. (2010) bias model and the `Colossus` code (Diemer 2018) to evaluate f, which is quite weakly sensitive to halo mass. We find that this amounts to a very minor correction with $0.976 < f < 1.019$.

2.4. The Integral Constraint

Consider the average value of an estimator $\xi^{\rm est}$ of the correlation function evaluated over random pairs of points within a survey volume. Because $\xi^{\rm est}$ is an excess probability relative to random pairs, this average must be zero regardless of the true value $\langle\xi^{\rm true}\rangle$ that would be obtained over an infinite volume. The difference is known as the integral constraint (IC; Infante 1994; Roche et al. 2002). The IC is approximately a negative additive bias that can be estimated by averaging $\xi^{\rm true}$ over random pairs in the survey volume; the projected correlation function then satisfies $w_p^{\rm est}(R) \approx w_p^{\rm true}(R) - {\rm IC} \times 2\pi_{\rm max}$. We estimate the IC as part of our fitting process by evaluating it for each proposed halo population (Wake et al. 2011). Although the IC is often a significant correction to angular correlation functions, it has far less impact on a spectroscopic survey, because the distance between random points is dominated by the survey's line-of-sight depth, which is much larger than $\pi_{\rm max}$. We correct for the IC but find that it changes w_p by at most a few percent in the outermost radial bins.

2.5. Clustering of the Full Galaxy Sample

Figure 3 (left panel) shows the autocorrelation function of the LATIS galaxies, computed in 10 logarithmic bins of r_p spanning 0.3–20 h^{-1} cMpc. Our first aim is to model the galaxy clustering with a thresholded halo sample defined by $M_h > M_h^{\rm thresh}$. The main difficulty in such modeling is that w_p errors in different radial bins are correlated, and on sufficiently large scales, this correlation is dominated by large-scale structure. As in many prior clustering studies (e.g., de la Torre et al. 2013; Durkalec et al. 2015), we use mock surveys to estimate the data covariance matrix. These mock surveys require an initial estimate of $M_h^{\rm thresh}$, which we derive by assigning independent Poisson errors to the w_p measurement in each radial bin, and then fitting these data with our set of halo $w_p(r_p)$ profiles using a maximum likelihood procedure. The initial estimate is $\log M_{h,0}^{\rm thresh}/M_\odot = 11.56$.

For each LATIS survey field, we then select nonoverlapping subvolumes that fill the MDPL2 simulation box. For the largest field, there are 135 such subvolumes. With each subvolume, we assign a mock redshift z_0 to each halo such that the comoving line-of-sight distance from $z_{\rm min} = 2$ to z_0 matches the distance of the halo to the front of the simulation box. We also assign mock celestial coordinates (R.A.$_0$, decl.$_0$) to each halo by converting its x- and y-coordinates to angular transverse distances, appropriate to the halo's mock redshift, from a sight line passing through the center of the subvolume; the celestial coordinates of this sight line are set to the center of the observed field. These calculations assume the distant observer approximation. We then randomly select halos with masses $M_h > M_{h,0}^{\rm thresh}$ with a probability proportional to ESR(R.A.$_0$, decl.$_0$) $\times\, z_{\rm pdf}(z_0)$. The total number of selected halos matches the total number of LATIS galaxies in the field. We then use the same code to compute w_p that is applied to the observed data. This process is repeated to realize 100 random halo selections per subvolume. Finally, we select many random triplets (i, j, k), where i, j, and k identify a realization of a subvolume associated with each of the three survey fields, and compute the mean w_p using the same weights applied to the observations. From these realizations we estimate the covariance matrix C_{ij} of w_p. The correlation matrix $C_{ij}/(C_{ii}C_{jj})^{1/2}$ is shown in Figure 3. Note that the bins are increasingly correlated at larger r_p.

We can now perform a Bayesian inference on $M_h^{\rm thresh}$. We take a broad, uniform prior on $\log M_h$ and assume a Gaussian likelihood with the covariance matrix C_{ij} estimated from the mock surveys. The resulting posterior is shown in Figure 3 (right panel). We summarize the pdf using the mode and the 16th and 84th percentiles: $\log M_h^{\rm thresh}/M_\odot = 11.56^{+0.06}_{-0.15}$. The mock surveys showed the pdf mode to be an unbiased estimator, whereas the median is biased low because the pdfs are left skewed. The minimum $\chi^2 = 8.3$ for nine degrees of freedom, so deviations from the model are consistent with the

uncertainties. Since our inferred $M_h^{\rm thresh}$ matches the initial estimate $M_{h,0}^{\rm thresh}$ (but with much more realistic errors), we did not find it necessary to iterate this procedure.

To facilitate comparisons with future work, we also fit a power-law model, though we stress that this plays no role in our halo mass inference. If the 3D correlation function follows a power law $\xi(r) = (r/r_0)^{-\gamma}$, then the projected correlation function w_p can be expressed as a power law multiplied by a hypergeometric function, which appears due to the finite integration window over π (see Trainor & Steidel 2012, Equation (10), which must be multiplied by $2\pi_{\rm max}$ to match our w_p convention). We fix $\gamma = 1.5$ since, like Trainor & Steidel (2012), we found this value to match the shape of the halo autocorrelation function within the relevant mass range, and measure $r_0 = 4.9 \pm 0.3\ h^{-1}$ cMpc.

2.6. Stellar Mass–dependent Clustering

We now consider the clustering of LATIS galaxies as a function of their stellar mass. A common method to analyze mass-dependent galaxy clustering is to compute w_p for a series of stellar mass–thresholded subsamples, i.e., $M_* > M_*^{\rm thresh}$, and to fit such data with an HOD model (e.g., Wake et al. 2011; Durkalec et al. 2018). This approach has drawbacks for analyzing high-redshift spectroscopic samples. First, no such sample is stellar mass limited; in the case of LATIS, the stellar mass distribution has a shape closer to Gaussian than a truncated mass function. Therefore observed galaxies at relatively undersampled masses need to be up-weighted, and this introduces a dependence on the assumed stellar mass function. Second, in the high stellar mass bins, the threshold approach incorporates only the pairs of massive galaxies with themselves, omitting their many pairs with lower-mass galaxies. This is not optimal when sample sizes are limited.

We introduce a different method in which galaxies are divided into N_b bins of stellar mass, and we compute and simultaneously model all of the auto- and cross-correlations (hereafter simply cross-correlations) between the $N_b \times (N_b + 1)/2$ pairs of bins. This approach incorporates all galaxy pairs into the analysis. Figure 4 shows the 10 cross-correlations arising from the four stellar mass bins listed in Table 1. Given the smaller number of galaxies in the individual mass bins, we calculate $w_p(r_p)$ using a reduced set of six r_p bins.

We compare these galaxy cross-correlations to halo cross-correlations, modeling each galaxy stellar mass bin as a mono-mass halo population. The procedure is analogous to that described in Section 2.5. We first estimate an initial $M_{h,0}^i$ for each of the four bins, indexed by i, by using Poisson error estimates and simultaneously fitting the 10 cross-correlations. We then build mock surveys; in each realization, we now select four random samples of halos around $M_{h,0}^i$, with the number of halos matching the number of observed galaxies in stellar mass bin i. Halos are selected according to the $z_{\rm pdf}$ described by galaxies in the associated stellar mass bin. The covariance matrix C_{ij} is calculated as before, but it is now 60×60. We then proceed to a Bayesian inference, as before. To sample the posterior, we use `emcee` (Foreman-Mackey et al. 2013). The main difference compare to Section 2.5 is that we find that three iterations are necessary: after each iteration, we use the resulting estimates of the halo masses to create a new suite of mock surveys. After three iterations, the inferred halo masses are sufficiently converged ($\Delta \log M_h \leqslant \sigma/4$, where σ is the random error). The change in $\log M_h$ between the first and last iterations is $\leqslant 0.08$ in all stellar mass bins; therefore, by iteratively bringing the mocks and observations into agreement, we improve internal consistency but ultimately make a small refinement to the halo masses.

Figure 4 shows that all cross-correlations can be simultaneously modeled within their uncertainties, with a minimum $\chi^2 = 50.1$ for 56 degrees of freedom. The corner plot in Appendix A shows that the posteriors are unimodal and covariance among the four halo mass parameters is low. The halo masses that we infer (Table 1) increase monotonically with stellar mass. These results are consistent with our analysis of the LATIS sample as a whole (Section 2.5): above a threshold of $M_h^{\rm thresh} = 11.56$, the median halo mass is $\log M_h = 11.80$, which is consistent with the interpolation of the halo mass estimates in Table 1 to the median $\log M_* = 9.72$. In Section 6, we will return to these measurements and compare them to the literature and our Lyα analysis.

To facilitate comparisons with future work and to help clarify the significance of differences among the stellar mass bins in Figure 4, we also fit a power-law model (see Section 2.5) to each of the the auto- and cross-correlation functions independently. Values of r_0 assuming a fixed $\gamma = 1.5$ are shown in each subpanel.

3. Galaxy–Lyα Cross-correlation: Observations

We now turn to the second method for measuring halo masses: the galaxy–Lyα cross-correlation. In this section we describe the observations and measurements, while Section 4 describes our calculation of the halo–Lyα cross-correlation from a cosmological hydrodynamic simulation, and Section 5 presents the inference procedure and the resulting halo mass estimates.

3.1. Lyα Forest Spectra

We measure Lyα absorption in a set of 3007 sight lines toward galaxies and quasars. Among the full LATIS sample, we select objects with $z > 2.277$ so that the Lyα forest is at least partly observed, and we exclude sources identified as having major data reduction flaws, poor fits of the spectral models that describe the unabsorbed continuum, or severe blending. The processing of spectra was described by Newman et al. (2020). The key measurement is the Lyα transmission fluctuation:

$$\delta_F = \frac{F}{\langle F \rangle (z)} - 1, \tag{5}$$

where F is the flux at a given wavelength normalized by the unabsorbed continuum model (see Newman et al. 2020) and $\langle F \rangle (z)$ is the redshift-dependent mean flux transmitted through the IGM (Faucher-Giguère et al. 2008). We consider only spectral pixels for which the uncertainty $\sigma_{\delta_F} < 2$. This leaves 4.7×10^5 Lyα forest pixels in our data set.

There are two further processing steps. First, we attempt to exclude high-column-density absorption lines from our cross-correlation. Although not an essential step, we find that their inclusion does not substantially improve constraining power, but it increases model dependence because subgrid models are required to simulate these lines accurately. We use a simple technique that masks lines with a high equivalent width (EW; see Newman et al. 2020). Briefly, we convolve the continuum-normalized spectrum with a 5 pixel boxcar, identify local minima, and measure the EW. The EW is measured within the smallest wavelength interval over which $F < \langle F \rangle (z)$, truncated to a maximum of ± 1000 km s^{-1} from the minimum in order to

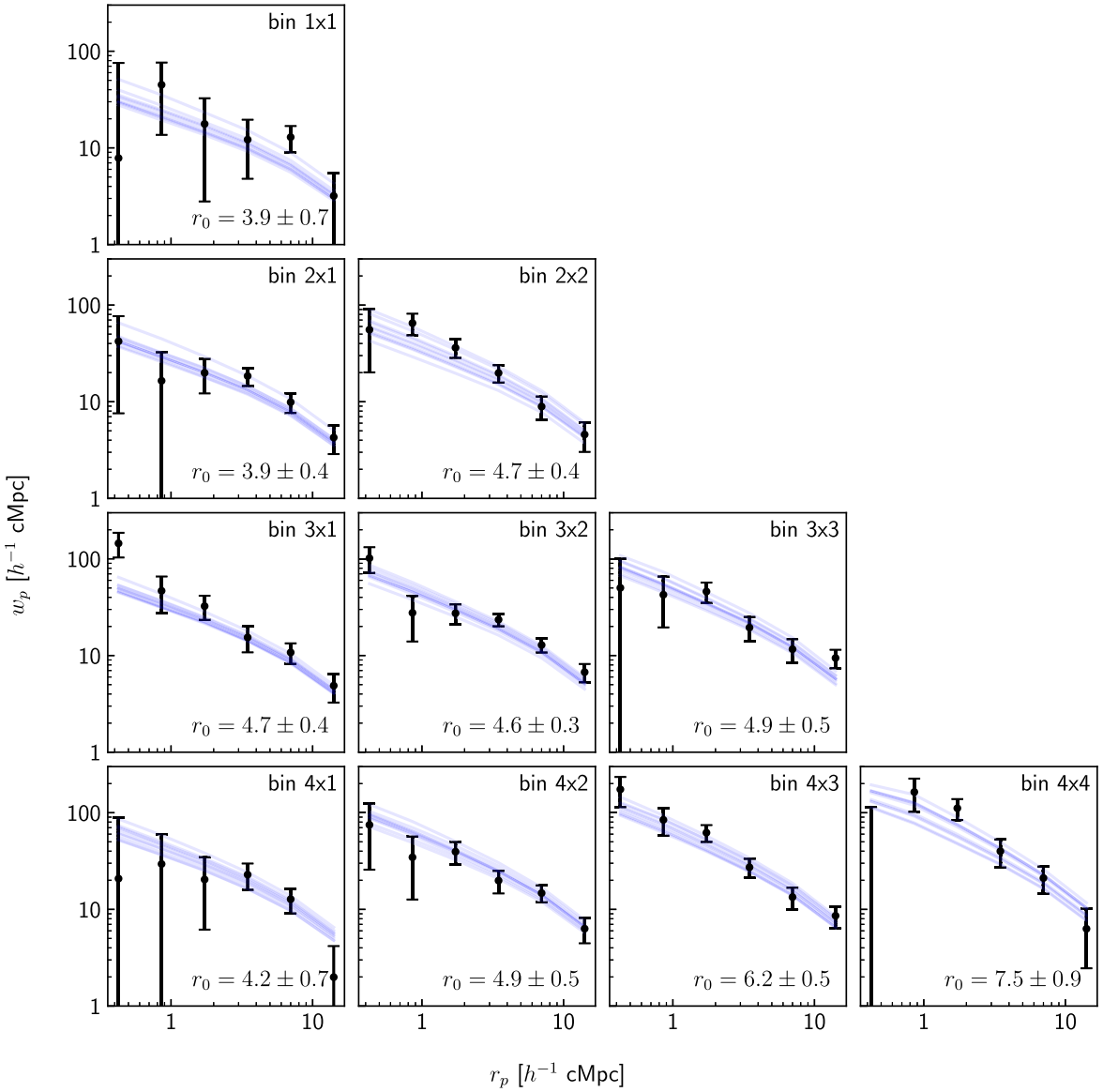

Figure 4. The projected auto- and cross-correlations among LATIS galaxies in four bins of stellar mass listed in Table 1. Each panel shows the measured $w_p(r_p)$ (black points) with $C_{ii}^{0.5}$ error bars based on the covariance matrix C_{ij} derived from the mock surveys. Blue curves show 10 posterior samples of the halo model. The correlation length r_0, expressed in units of h^{-1} cMpc, is shown in each subpanel.

Table 1
Halo Mass Constraints from Galaxy–Galaxy Clustering

Bin num.	$\log M_*$	$\langle \log M_* \rangle$	$N_{\rm gal}$	$\langle z \rangle$	$\log M_h$
1	<9.4	9.21	547	2.58	$11.39^{+0.19}_{-0.36}$
2	9.4–9.8	9.61	1181	2.53	$11.78^{+0.11}_{-0.18}$
3	9.8–10.2	9.97	899	2.49	$12.03^{+0.12}_{-0.12}$
4	>10.2	10.46	364	2.47	$12.44^{+0.13}_{-0.22}$

Note. Masses are expressed in units of $M_\odot$.

avoid identifying broad, shallow features. If the rest-frame EW > 5 Å and the feature is significantly detected (EW/$\sigma_{\rm EW}$ > 5), the line is masked over the wavelength interval where $F < \langle F \rangle(z)$. Based on the simulations described in Section 4, we find that this simple method identifies 95% of lines with column densities $N_{\rm HI} > 10^{20.7}$ cm^{-2}; it inevitably also masks some lower-$N_{\rm HI}$ lines that are blended at our spectral resolution.

Second, we refine the unabsorbed continuum model for each sight line using a multiplicative polynomial that forces $\langle \delta_F \rangle \approx 0$ on large scales. The polynomial order depends on the length of the Lyα forest that is covered by the spectrum (see Newman et al. 2020), but the filtering scale is typically $\Delta z \approx 0.2$. This process is called mean flux regularization (MFR; Lee et al. 2012). MFR mitigates errors in the spectrophotometric calibration, but it also suppresses large-scale correlations.

It is important to bear in mind that we measure a processed version of δ_F, and that the same procedures that exclude high-EW lines and perform MFR will be applied to the simulated Lyα forest spectra to enable an accurate comparison.

3.2. Foreground Galaxies and Redshift Estimates

The foreground sample consists of the galaxies whose positions will be cross-correlated with Lyα absorption. We use a subset of the galaxy sample described in Section 2.1, restricting to the redshift range $z_{\rm fg} = 2.22$–2.90. The lower limit

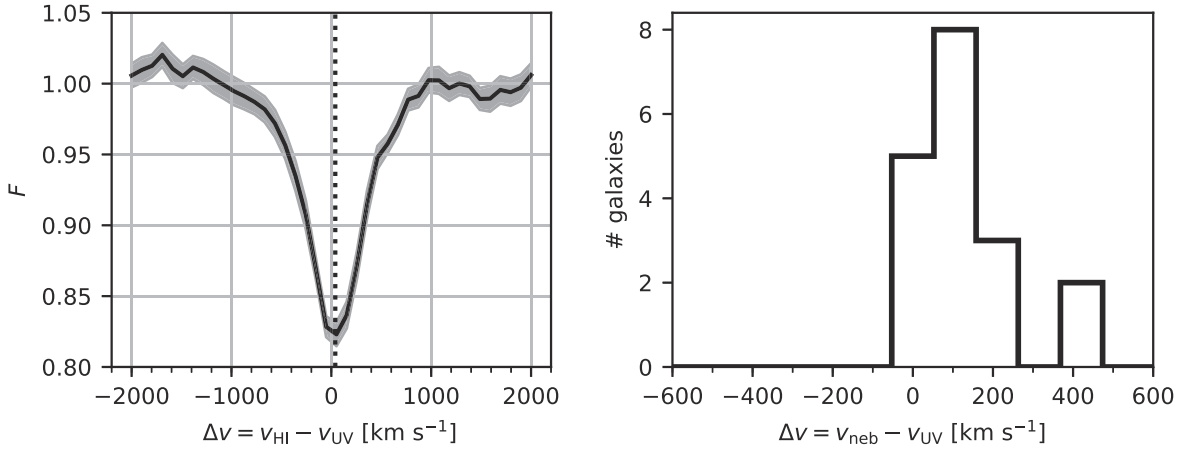

Figure 5. Calibration of LATIS redshifts using transverse absorption and nebular emission. Left: average spectrum (and 1σ uncertainty band) of sight lines within $r_p < 4\ h^{-1}$ cMpc of LATIS galaxies, shifted into each galaxy's rest frame using the LATIS UV redshifts. We detect a small asymmetry around $\Delta v = 0$ attributed to a systematic offset in the UV redshifts. Right: comparison of LATIS UV redshifts to MOSDEF nebular redshifts for 18 galaxies in common.

ensures that the region within ± 2000 km s^{-1} of Lyα lies within our observed spectral range, while the upper limit is motivated by the rapidly declining number of background sources.

Rest-frame UV redshifts rely mainly on the resonant Lyα line (in emission and absorption) and metallic absorption lines produced in part by outflowing gas; neither traces the galaxy systemic redshift directly. Rest-frame UV surveys often have systematic velocity offsets compared to indicators of galaxies' systemic redshifts. Although these are irrelevant to galaxy–galaxy clustering, such offsets would affect the galaxy–Lyα cross-correlation. One way to estimate a systematic velocity offset is to examine the average transverse Lyα absorption profile, which should be symmetric about galaxies' systemic redshifts (Rakic et al. 2012). For each sight line–foreground galaxy pair with a transverse separation $r_p < 4\ h^{-1}$ cMpc, we shift the continuum-normalized spectrum into the galaxy's rest frame using the LATIS redshift (see Newman et al. 2020), and we then compute the inverse-variance weighted mean over pairs (Figure 5). We find that the absorption is nearly symmetric about $\Delta v = v_{\rm HI} - v_{\rm UV} = 0$, but we do detect a small offset $\Delta v = 39 \pm 16$ km s^{-1}. The offset shows no clear trend with spectral morphology as indicated by the Lyα EW, so we simply add the same Δv to all LATIS redshifts.

We can also evaluate systematic offsets, as well as random errors, in the LATIS redshifts by comparing them to nebular redshifts for 18 galaxies in common with the MOSDEF survey (Kriek et al. 2015). We find that the median offset of $\Delta v = v_{\rm neb} - v_{\rm UV} = 96 \pm 36$ km s^{-1} is consistent with, but less precise than, the estimate from transverse Lyα absorption. (The uncertainties are determined from bootstrap resampling.) The standard deviation of velocity difference is $\sigma_v = 120 \pm 24$ km s^{-1}, which may be taken as an error in the UV redshifts given that the errors in MOSDEF redshifts are expected to be much smaller; however, inspecting the histogram in Figure 5 shows that σ_v is heavily influenced by the two largest outliers. The LATIS redshifts for these two galaxies are very sensitive to whether the Lyα region is included in the spectral model fit; if it is not, the redshifts obtained from the interstellar absorption lines agrees well with the MOSDEF redshifts. This situation is rare: a comparably large (>300 km s^{-1}) dependence on the inclusion of Lyα occurs in only 0.7% of the LATIS sample, which indicates that the subset in common with MOSDEF is likely unrepresentative. Excluding those two galaxies, the estimated redshift error becomes $\sigma_v = 66 \pm 8$ km s^{-1}. Such an exclusion is reasonable but still uncomfortably ad hoc. We use these results only to estimate a reasonable prior on σ_v for our galaxy–Lyα analysis: we will use a normal distribution centered on the average of our two estimates (120 and 66 km s^{-1}) with a dispersion set to their half-difference, i.e., 93 ± 27 km s^{-1}.

3.3. Estimating the Galaxy–Lyα Cross-correlation

To estimate the cross-correlation between LATIS galaxies and Lyα absorption, we follow a similar procedure as Font-Ribera et al. (2012), who measured the cross-correlation between damped Lyα absorbers and Lyα fluctuations measured in quasar spectra. The cross-correlation is defined as:

$$\xi^{{\rm Ly}\alpha}(r_p, \pi) = \frac{\sum_{i \in A} w_i \delta_{F,i}}{\sum_{i \in A} w_i}, \tag{6}$$

where the summations are over a list A of those pixels located within a separation bin $(r_p \pm \Delta r_p, \pi \pm \Delta\pi)$ from a galaxy in the foreground sample; note the same pixel may appear multiple times. This is simply a weighted mean of δ_F, and $\xi^{{\rm Ly}\,\alpha}$ can be equivalently thought of as an Lyα absorption profile. We use inverse-variance weights that incorporate both observational noise and the intrinsic variance in the Lyα forest:

$$w_i = [\sigma_F^2(z_i) + \sigma_{{\rm rand},i}^2 + \sigma_{{\rm cont},i}^2]^{-1}, \tag{7}$$

where $\sigma_F^2(z_i)$ is the intrinsic variance at the redshift z_i of pixel i, $\sigma_{{\rm rand},i}$ is the random noise in $\delta_{F,i}$, and $\sigma_{{\rm cont},i}$ is the estimated noise in $\delta_{F,i}$ attributed to errors in the continuum model. We approximate $\sigma_F(z) = 0.20[(1+z)/3.5]^{2.4}$ based on the simulated spectra that we introduce below, and therefore σ_F^2 represents the total variance in noiseless spectra with the same spectral resolution and processing as the LATIS data. We evaluate $\sigma_{{\rm cont},i}$ as a function of the continuum-to-noise ratio following (Newman et al. 2020, Section 7.4). Note that these weights naturally prevent the dominance of a few spectra with high signal-to-noise ratios.

We use linear bins of r_p spanning 0–20 h^{-1} cMpc with widths of 1 h^{-1} cMpc and bins of π spanning -15 to 15 h^{-1} cMpc with widths of 1.2 h^{-1} cMpc, which approximately match the LATIS pixel scale (1.8 Å) at the mean redshift. We then fold

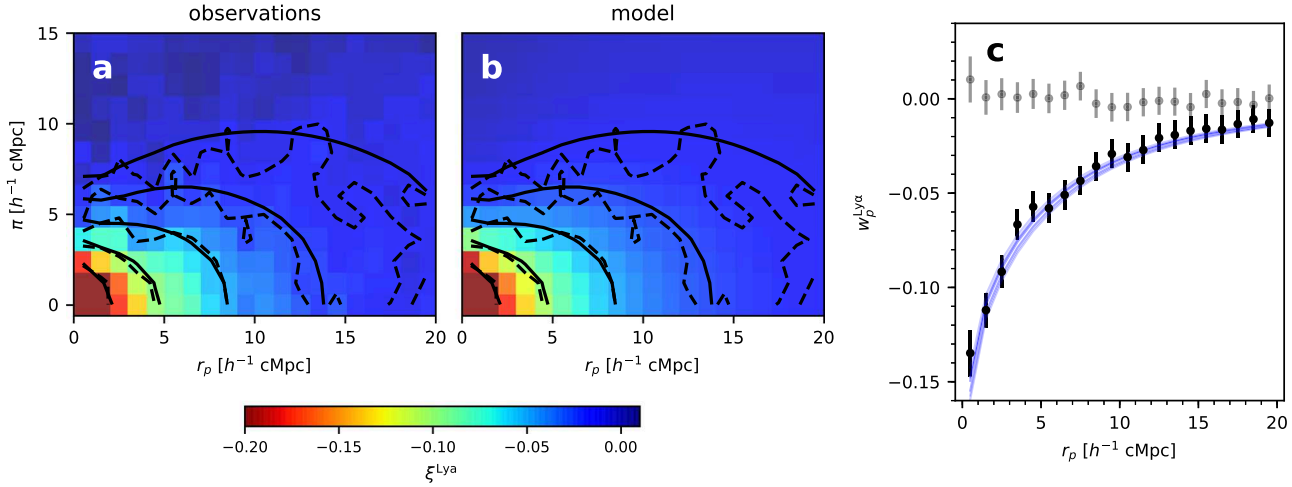

Figure 6. The galaxy–Lyα cross-correlation $\xi^{\mathrm{Ly}\,\alpha}$ computed from the full LATIS foreground galaxy sample, independent of stellar mass. Panel (a): the observed 2D correlation function. Contours begin at $\xi^{\mathrm{Ly}\,\alpha} = -0.2$ and decrease to -0.0125 in factors of two. Dashed and solid contours indicate the observed and model correlation functions, respectively. Panel (b): the model 2D correlation function; contours are the same as in panel (a). Panel (c): the projected cross-correlation $w_p^{\mathrm{Ly}\alpha}$. Black points show observed values with 1σ errors derived from the covariance matrix, and blue curves show 10 posterior samples of the model. Gray points show a null test constructed by scrambling the Lyα forest spectra among the sight lines.

he cross-correlation about $\pi = 0$, averaging estimates in the fore- and background; therefore the first π bin covers 0 to 0.6 h^{-1} cMpc.

Figure 6(a) shows the 2D galaxy–Lyα cross-correlation derived from the entire LATIS sample, independent of stellar mass. A strong detection of excess Lyα absorption to large distances is clearly apparent, along with an anisotropic shape. We will discuss the 2D structure of the cross-correlation further in Section 4.

It is useful to consider a projected cross-correlation function akin to the w_p statistic that we employed to measure galaxy–galaxy clustering. We define $w_p^{\mathrm{Ly}\alpha}$ by averaging $\xi^{\mathrm{Ly}\,\alpha}$ within $|\pi| < \pi_{\max} = 6.6\ h^{-1}$ cMpc:

$$w_p^{\mathrm{Ly}\alpha}(r_p) = \frac{1}{2\pi_{\max}} \int_{-\pi_{\max}}^{\pi_{\max}} \xi^{\mathrm{Ly}\alpha}(r_p, \pi) d\pi. \tag{8}$$

Numerically this corresponds to averaging the first six π bins (see Figure 6) with half the weight given to the first bin as to the others, since it is centered on 0. Note that this convention differs from that for the galaxy–galaxy w_p, which is an integral over π rather than an average. The disadvantage of our definition for $w_p^{\mathrm{Ly}\alpha}$ is that it does not converge as $\pi_{\max} \to \infty$, but we find it convenient for $w_p^{\mathrm{Ly}\alpha}$ to maintain a dimensionless form to facilitate comparisons with systematic uncertainties in δ_F. No single choice of $\pi_{\max}$ is optimal at all r_p, but based on the simulations introduced below, we find that the signal-to-noise ratio of $w_p^{\mathrm{Ly}\alpha}$ is close to maximal over a wide range of r_p for our choice of $\pi_{\max}$.

When analyzing galaxy–galaxy clustering we had to account for angular variations in the ESR arising from any source, since they directly affect the number of pairs that are counted. In LATIS such variations mainly depend on a galaxy's position within the IMACS field of view. Such sampling variations are not relevant to the galaxy–Lyα cross-correlation, since they do not bias the measured δ_F. Potentially concerning are variations in ESR that correlate with local density, and thus δ_F. Newman et al. (2022) simulated the recovery of galaxy density in a two-pass LATIS-like survey, respecting geometrical constraints imposed by slit mask design. On 2–8 h^{-1} cMpc scales, where $\xi^{\mathrm{Ly}\,\alpha}$ is best constrained, the galaxy density was biased $\lesssim$2% within density fluctuations $<3\sigma$. Only a small fraction of LBGs reside in higher-density regions, so we can safely assume that any such sampling bias has a negligible effect on $\xi^{\mathrm{Ly}\,\alpha}$.

3.4. Data Covariance Matrix

Measurements of $\xi^{\mathrm{Ly}\,\alpha}$ in different bins are highly covariant for several reasons. First, a given Lyα forest pixel forms pairs with many galaxies and so appears in many bins. Second, continuum errors are coherent along sight lines. Third, Lyα forest fluctuations are correlated by large-scale structure. We follow Font-Ribera et al. (2012) to estimate the covariance matrix analytically.

Consider two separation bins A and B of $\xi^{\mathrm{Ly}\,\alpha}$. The covariance of the cross-correlation $\tilde{C}_{AB}$ is:

$$\tilde{C}_{AB} = \frac{\sum_{i\in A}\sum_{j\in B} w_i w_j C_{ij}}{\sum_{i\in A} w_i \sum_{j\in B} w_j}, \tag{9}$$

which is based on Equation (3.11) of Font-Ribera et al. (2012), where $C_{ij} = \langle \delta_{F,i}\delta_{F,j} \rangle$ is the correlation of δ_F measured in pixels i and j. Taking the transverse and perpendicular separations of these pixels to be $r_{p,ij}$ and π_{ij}, we estimate the correlation as:

$$C_{ij} = \xi_F(r_{p,ij}, \pi_{ij}) + \sigma^2_{\mathrm{rand,i}}\delta^K_{ij} + \sigma^2_{\mathrm{cont,i}}\delta^D(r_{p,ij}), \tag{10}$$

which is analogous to Equation (3.12) of Font-Ribera et al. (2012). The first term is the intrinsic autocorrelation of δ_F, which we compute directly from the simulations introduced below. This term encodes cosmic variance. The second term represents the random noise in the spectra, which appears only with $i = j$ such that the Kronecker delta function $\delta^K_{ij} = 1$ ($\delta^K_{ij} = 0$ when $i \neq j$). The last term represents the continuum error, which appears only when the pixels are members of the same sight line so that the Dirac function $\delta^D(r_{p,ij}) = 1$ and $(\delta^D(r_{p,ij}) = 0$ when $r_{p,ij} \neq 0$. We approximate the continuum error as being perfectly correlated along a sight line with a variance σ^2_{cont} that is a function of the median continuum-to-noise ratio along the sight line (see Section 7.4 in Newman et al. 2020). Thus

$\sigma^2_{\mathrm{cont},i} = \sigma^2_{\mathrm{cont},j}$ whenever the third term is active. Given the covariance matrix $\tilde{C}$ of $\xi^{\mathrm{Ly}\,\alpha}$, it is straightforward to calculate the covariance matrix of the projected $w_p^{\mathrm{Ly}\alpha}$.

The matrix $\tilde{C}$ is slow to compute. Each bin of $\xi^{\mathrm{Ly}\,\alpha}$ is formed by averaging pixels from many galaxy–Lyα pairs. Computing $\tilde{C}_{AB}$ then requires calculating a weighted mean of C_{ij} over every pair of pixels in bins A and B. To speed the calculation, we parallelize it and approximate $\xi_F = 0$ when $r_{p,ij}$ or $|\pi_{ij}|$ exceeds 30 h^{-1} cMpc.

4. Galaxy–Lyα Cross-correlation: Simulations

In order to estimate a halo mass from the observed galaxy–Lyα cross-correlation, we compute simulated halo–Lyα cross-correlations using the large and high-resolution (2×5500^3 particles in a volume of 250^3 h^{-3} cMpc^3) cosmological hydrodynamic simulation ASTRID (Bird et al. 2022; Ni et al. 2022). In this section we describe our construction of Lyα skewers from ASTRID, the processing of these skewers into mock LATIS spectra, and the calculation of redshift- and mass-dependent halo–Lyα cross-correlations.

4.1. Lyα Optical Depths

Our construction of Lyα absorption spectra from the ASTRID simulations follows the procedures described in Section 2.2 of Qezlou et al. (2022) and Section 3.1 of Qezlou et al. (2023). Briefly, each gas particle in the simulation is treated as an individual absorber with internal physical properties smoothed by a quintic spline kernel. The absorption spectrum is then composed of all the Voigt profiles of the absorbers along a given line of sight. We estimate the neutral hydrogen fraction by assuming a uniform ultraviolet background (UVB) and solving the collisional/photoionization and recombination rate networks from Katz et al. (1996). The self-shielding of dense neutral gas is modeled using fitting formulae derived from radiative transfer models (Rahmati et al. 2013). We effectively set the UVB intensity by scaling the Lyα optical depths to match the mean flux evolution observed by Faucher-Giguère et al. (2008), as corrected for metal absorption. We compute a grid of 1000^2 sight lines parallel to the z-axis with a transverse spacing of 0.25 h^{-1} cMpc. The spectral pixel size is also 0.25 h^{-1} cMpc, or about 27 km s^{-1}. This procedure is repeated for snapshots at $z = 2.3$, 2.5, 2.7, and 2.9, approximately matching the range of our foreground galaxy sample.

4.2. Halo Masses in ASTRID versus MultiDark

Among our goals is to compare the consistency of halo masses estimated from galaxy–galaxy and galaxy–Lyα clustering. We use different simulations that are most appropriate for each analysis: the large, N-body simulation MDPL2 to enable a large suite of mocks for galaxy–galaxy clustering, and the smaller but hydrodynamic simulation ASTRID to compute Lyα spectra. These simulations are based on slightly different cosmological parameters, particularly σ_8, and different methods were used to construct the subhalo catalogs: `RockStar` (Behroozi et al. 2013b) for MDPL2 and `SUBFIND` (Springel et al. 2001; see Bird et al. 2022) for ASTRID.

We present halo masses derived from each method (galaxy–galaxy and galaxy–Lyα clustering) based on the associated simulation. However, when we compare the results from these methods in Section 6, we will account for the expected mass differences arising from differences in the simulations. We estimate these mass differences by comparing halo–halo clustering in MDPL2 and ASTRID. Figure 7 shows that halo populations with matching autocorrelation functions w_p have slightly different masses in MDPL2 and ASTRID, with $|\Delta \log M_h| < 0.1$ at masses relevant for our analysis. In other words, if we were able to conduct both analyses self-consistently within ASTRID, we expect the halo masses from galaxy–galaxy clustering would be slightly higher, by $|\Delta \log M_h|$, than those we inferred using MDPL2.

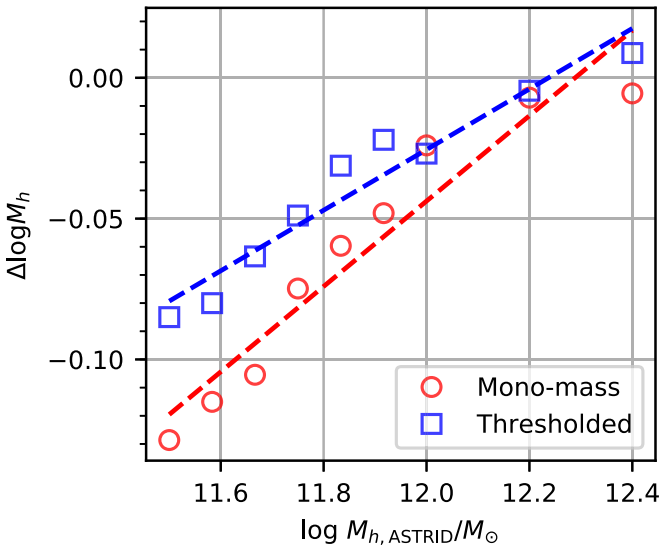

Figure 7. Differences in the masses of halo populations with matched autocorrelation functions in MDPL2 and ASTRID. For each halo population in ASTRID (mono-mass: $|\log M_h/M_{h,\mathrm{ASTRID}}| < 0.1$; thresholded: $M_h > M_{h,\mathrm{ASTRID}}$), we adjust the mass parameter $M_{h,\mathrm{MDPL2}}$ in MDPL2 to achieve the best match to w_p, defined as the rms fractional difference over $1 < r_p/(h^{-1}\ \mathrm{cMpc}) < 10\ h^{-1}$. Linear fits are shown: $\Delta \log M_h = \log M_{h,\mathrm{MDPL2}}/M_{h,\mathrm{ASTRID}} = a(\log M_{h,\mathrm{ASTRID}}/10^{12} M_\odot) + b$, with a, $b = 0.15$, -0.04 (mono-mass) or 0.11, -0.03 (thresholded).

4.3. Synthesizing Mock LATIS Spectra

The high-resolution synthetic spectra are then processed to mimic the LATIS data. First, we smooth the spectra by the LATIS line-spread function. Based on the average extent of galaxies measured in the 2D spectra, the line-spread function can be approximated as a Gaussian with $\sigma/(\mathrm{km\ s^{-1}}) = 227$–$98(\lambda/500$ nm). We evaluate this expression for each snapshot at the observed wavelength of Lyα. The convolved spectra are then resampled to a velocity grid with a cell size that matches the LATIS pixel size of 1.8 Å in the observed frame.

The spectra must then be further processed to account for the masking of high-EW lines and MFR (Section 3.1). Our ultimate aim of extracting the mean Lyα absorption signal around halos does not directly require us to inject noise. However, correctly capturing the average effect of the high-EW line masking procedure does require the spectra to have noise properties similar to LATIS. Therefore we temporarily inject random noise and continuum errors into the synthetic spectra. The random noise is Gaussian white noise with a dispersion randomly chosen for each sight line by drawing a pixel from LATIS at a redshift close to that of the snapshot. We also inject continuum error by multiplying F by a sinusoid scaled so that its variance is σ^2_{cont}, which in turn is a function of the sight

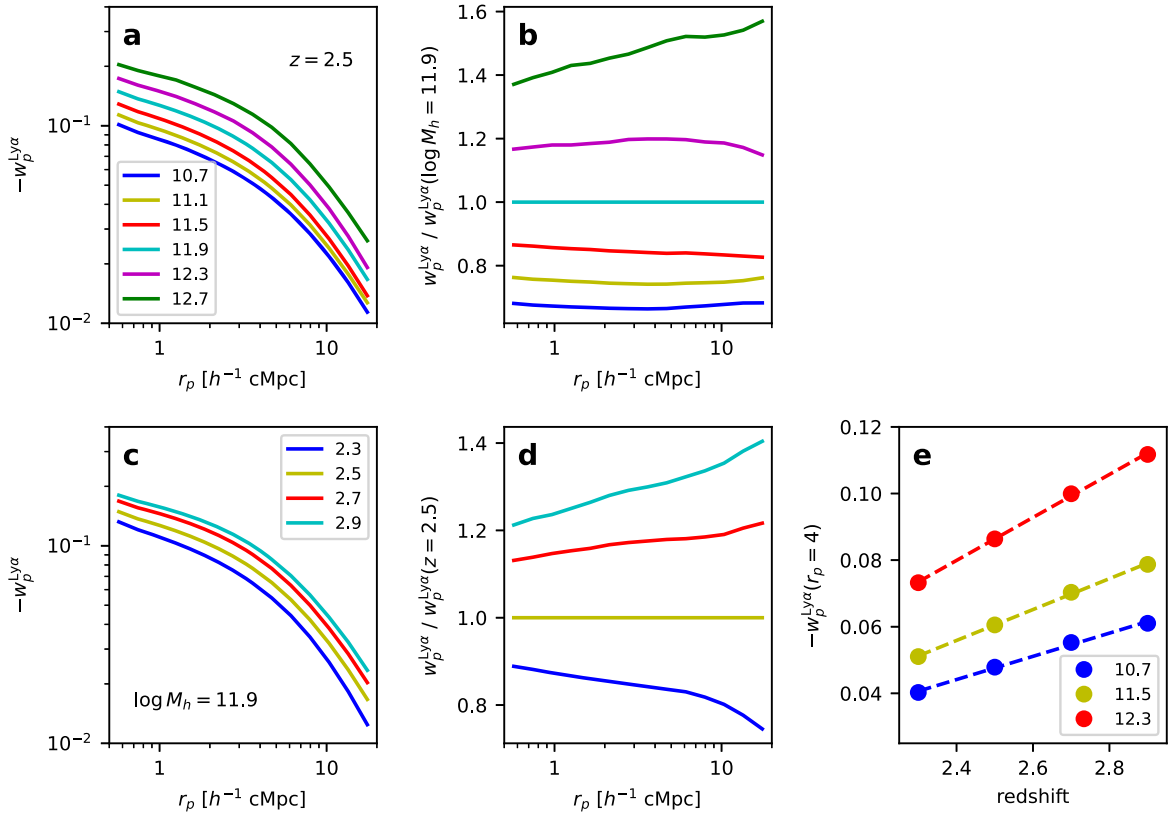

Figure 8. Trends in the projected halo–Lyα cross-correlation with halo mass and redshift. Panel (a): the variation in $w_p^{\mathrm{Ly}\alpha}(r_p)$ with halo mass ($\log M_h/M_\odot$ indicated in legend) in the $z = 2.5$ simulation snapshot. Panel (b): the ratio of each curve in panel (a) to the $M_h = 10^{11.9}\,M_\odot$ curve, highlighting the weak dependence of the shape of $w_p^{\mathrm{Ly}\alpha}(r_p)$ on M_h. Panel (c): the variation of $w_p^{\mathrm{Ly}\alpha}(r_p)$ with redshift at the fixed halo mass indicated. Panel (d): the ratio of each curve in panel (c) to the one at $z = 2.5$. Panel (e): the variation of $w_p^{\mathrm{Ly}\alpha}$ with redshift at fixed M_h (indicated in legend) and fixed $r_p = 4\ h^{-1}$ cMpc, highlighting the linearity of the redshift dependence.

line's continuum-to-noise ratio (Newman et al. 2020). We find that our results are not sensitive to the details of the continuum error modeling; they have only an indirect effect of modulating which high-EW lines are masked, and these are relatively rare and do not dominate the mean absorption signal.

We then identify the high-EW lines and perform MFR on the noisy synthetic spectra, following methods applied to the observations (Section 3.1). One complication is that the MFR polynomial order depends on the observed forest length, which is a distribution that varies with redshift, whereas the synthetic sight lines all span the full box. We mimic MFR as follows. For a given snapshot and sight line, we first draw a random LATIS background galaxy whose Lyα forest probes the snapshot redshift, and we compute the observed length L of the forest as limited by our blue cutoff $\lambda > 3890$ Å. Starting from a random position along the sight line, we partition it into (one or two) chunks of length L, wrapping around the periodic boundary when necessary, and then perform MFR separately on each. In many cases L exceeds the ASTRID box size, and the simulation is therefore missing power on scales that are observed. However, MFR removes power on scales $\Delta z \lesssim 0.3$, depending on L, while the depth of the ASTRID box is $\Delta z = 0.31$ at $z = 2.5$. Therefore the missing power would be removed by MFR anyway, and we conclude that the simulation box is large enough to mimic MFR adequately. In a small fraction 7×10^{-5} of sight lines, MFR makes a large $>4\times$ correction to the continuum level that can result in extreme values of δ_F and even unphysically negative continua. We mask these rare sight lines; they have no counterparts in the observed data set.

Finally we apply these high-EW line masks and MFR polynomials to the noiseless synthetic spectra. We are left with 10^6 sight lines per snapshot that mimic the LATIS resolution, sampling, and processing.

4.4. Halo–Lyα Cross-correlations

We next compute the average Lyα absorption around halos as a function of their mass and redshift. As in our galaxy–galaxy clustering analysis, we consider both thresholded samples with $M_h > M_h^{\mathrm{thresh}}$ and mono-mass samples with masses in a narrow range $|\log(M_h/M_{h,0})| < 0.1$. For a given snapshot and halo sample, we first randomly select a subsample of up to 5×10^5 halos. We compute the redshift-space position of each halo and round it to the nearest point defined by the grid of synthetic sight lines. (In the spectral direction, this nearest-neighbor assignment mimics our procedure of assigning LATIS pixels to the nearest π bin when constructing $\xi^{\mathrm{Ly}\,\alpha}$.) We extract a 3D cube of δ_F measurements around this point, and we average these cubes over the halo subsample. We verified that the absorption is centered in the resulting flux cube. We then project each flux cube into a 2D (r_p, π) grid with the same cell sizes as the observed $\xi^{\mathrm{Ly}\,\alpha}$, averaging over r_p and using spline interpolation over π.

Figure 8 illustrates the main trends. We first consider $w_p^{\mathrm{Ly}\alpha}$ for ease of visualization. Panels (a) and (b) show that, at a fixed redshift, Lyα absorption around halos is approximately scale independent, i.e., it differs by a factor that depends on M_h but

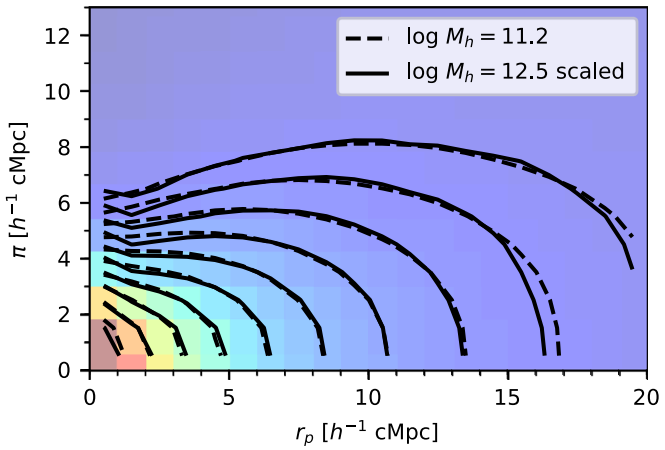

Figure 9. The insensitivity of the shape of the halo–Lyα correlation function $\xi^{\mathrm{Ly}\,\alpha}$ to halo mass is demonstrated by comparing $\xi^{\mathrm{Ly}\,\alpha}(M_h = 10^{11.2}\,M_\odot)$ (colors following Figure 6, and dashed contours) to $0.54 \times \xi^{\mathrm{Ly}\,\alpha}(M_h = 10^{12.5}\,M_\odot)$ (solid contours). Both are computed at $z = 2.5$.

little on r_p. Only at the highest masses $M_h \gtrsim 10^{12.7}\,M_\odot$ does the shape of $w_p(r_p)$ appreciably change. Panels (c) and (d) show that at fixed M_h, the shape of $w_p(r_p)$ varies with redshift. When we fix both M_h and r_p and examine the dependence on redshift (panel (e)), we find that it is quite linear over the range $z = 2.3$–2.9. This permits a simplification: we need not consider the distribution of redshifts of our foreground galaxy sample when fitting a model; only the mean redshift is relevant.

The spatial distribution of Lyα absorption is insensitive to the halo mass not only in projection, but also in its 2D structure, as noted by Kim & Croft (2008). Figure 9 shows that contours of $\xi^{\mathrm{Ly}\,\alpha}(M_h)$, at fixed redshift, have an almost identical shape even for halo masses that differ by a factor of 20. It is only the amount of Lyα absorption that constrains the halo mass, not its radial or line-of-sight distribution.

5. Galaxy–Lyα Cross-correlation: Fitting and Results

Finally we compare our observed 2D and 1D galaxy–Lyα cross-correlations, $\xi^{\mathrm{Ly}\,\alpha}$ and $w_p^{\mathrm{Ly}\alpha}$, to the computed halo–Lyα cross-correlations and infer halo masses. Fitting $\xi^{\mathrm{Ly}\,\alpha}$ in 2D should provide the most constraining power, not because the redshift-space anisotropies it encodes are sensitive to halo mass (see Section 4.4), but simply because it is analogous to a matched filter. Our simple definition of $w_p^{\mathrm{Ly}\alpha}$, which uses fixed integration limits $\pi_{\max}$, cannot provide an optimal estimator of the total absorption at every r_p and so must lose some statistical power. On the other hand, the projected cross-correlation function $w_p^{\mathrm{Ly}\alpha}$ has the merits of being insensitive to the modeling of redshift errors and redshift-space distortions; it should be less precise but possibly more robust. We will fit both and compare results. As in our galaxy–galaxy clustering analysis, we model the full LATIS galaxy sample with a thresholded halo population, and galaxy subsamples binned by stellar mass with mono-mass halo populations.

5.1. Inference

Consider a division of the LATIS foreground galaxies into N stellar mass bins: we will use $N = 1$ for the full sample and $N = 4$ for the stellar mass–defined subsamples. We begin by describing the inference of halo masses from the 2D cross-correlation $\xi^{\mathrm{Ly}\,\alpha}$. Since $\xi^{\mathrm{Ly}\,\alpha}$ has 20 bins of r_p and 13 bins of π, the length of the data vector is $260N$, and the covariance matrix $\tilde{C}$ described in Section 3.4 is $260N \times 260N$. There are $N + 1$ parameters: N describing the halo mass M_h associated with each galaxy bin, and one additional parameter σ_z that encodes random errors in the galaxy redshifts, which is assumed to be the same for all mass bins. For each mass bin i and estimated $\xi_i^{\mathrm{Ly}\alpha}$, we compute the weighted mean redshift $\langle z \rangle_i$ of all the δ_F measurements that entered $\xi_i^{\mathrm{Ly}\alpha}$ using the same weights (Equation (7)). As justified by Figure 8(e), we linearly interpolate between snapshots to produce a grid of synthetic $\xi^{\mathrm{Ly}\,\alpha}$ that is matched to $\langle z \rangle_i$ and varies over M_h. We then linearly interpolate over this grid to the target $\log M_{h,i}$. Finally we convolve the resulting model along the π direction using a Gaussian kernel with dispersion σ_z. We compute a standard Gaussian likelihood $\mathcal{L}$ from the difference between the data and model vectors and the covariance matrix $\tilde{C}$. We take broad uniform priors on $\log M_{h,i}/M_\odot$ spanning 10.7–12.9 and the Gaussian prior on σ_z described in Section 3.2. The posteriors are then sampled using `emcee`.

We also infer halo masses from the projected cross-correlation function $w_p^{\mathrm{Ly}\alpha}$. Motivated by maximizing the robustness of this analysis, we further eliminate the innermost bin with $r_p < 1\ h^{-1}$ cMpc where feedback is expected to have the strongest impact. This reduces the data vector size to $19N$; it is straightforward to compute its associated covariance matrix from $\tilde{C}$.

5.2. Results: Full Sample

Figure 6 shows the results of fitting $\xi^{\mathrm{Ly}\,\alpha}$ as measured on the full foreground galaxy sample, undifferentiated by mass, to a thresholded halo sample. Visually the model provides an excellent fit both to the observed 2D structure (compare panels (a) and (b)) and the projected absorption profile (panel (c)). Some systematic differences are apparent beyond $r_p \sim 10\ h^{-1}$ cMpc, but these measurements are highly correlated and, as we will show, the differences are not significant. We find $\log M_h^{\mathrm{thresh}}/M_\odot = 11.54 \pm 0.09$ and $\sigma_z = 65 \pm 17$ km s^{-1} (see corner plot in Appendix A). Note that the σ_z posterior is narrower than and shifted from the center of the prior, indicating that the galaxy–Lyα clustering favors redshift errors at the lower end of our range of estimates based on comparing to nebular redshifts (Section 3.2). As a metric of fit quality, we note that at the MAP parameters, $\chi^2 = 255$ for a data vector of size 260, indicating that the errors are well described by our covariance matrix.

We also performed a fit to the projected $w_p^{\mathrm{Ly}\alpha}$ alone. The resulting halo mass estimate $\log M_h^{\mathrm{thresh}}/M_\odot = 11.49 \pm 0.13$ is slightly lower but still consistent with that obtained from the 2D $\xi^{\mathrm{Ly}\,\alpha}$. The uncertainty in $\log M_h^{\mathrm{thresh}}$ increases in this case, as expected, by about 50%, and the posterior of σ_z matches the prior, since all sensitivity to redshift errors was destroyed by the projection.

5.3. Results: Mass-dependent Galaxy–Lyα Clustering

We now split the foreground galaxies into four bins of stellar mass, listed in Table 2. The bins have the same stellar mass ranges as those used in Section 2, except that the last dividing point is lower by 0.1 dex. We made this slight adjustment because we found that a usefully precise measurement required the highest-mass bin to be slightly wider, owing to the smaller galaxy sample available for the Lyα analysis.

Measurements of $\xi^{\mathrm{Ly}\,\alpha}$ and $w_p^{\mathrm{Ly}\alpha}$ are shown in Figure 10 for each stellar mass bin. Visually, we see more extended

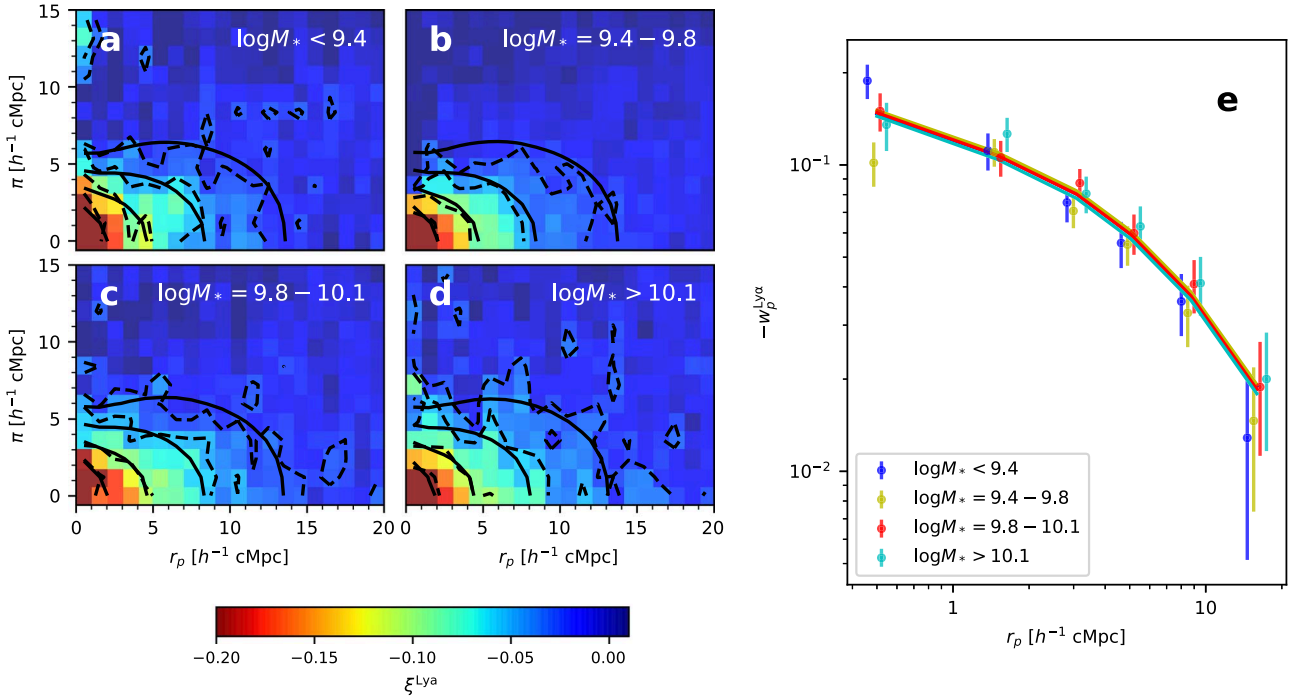

Figure 10. Panels (a–d): the galaxy–Lyα correlation function measured in four bins of M_*. Color and dashed contours show the measured $\xi^{\rm Ly\,\alpha}$, while solid contours show the fitted models. Panel (e): the projected galaxy–Lyα correlation function $w_p^{\rm Ly\alpha}$. Colored data points correspond to the stellar mass bins indicated in the legend and are horizontally offset for clarity. Colored lines (heavily overlapping) show the corresponding model for each bin.

Table 2
Halo Mass Constraints from Galaxy–Lyα Clustering

Bin number	$\log M_*$ range	$\langle \log M_* \rangle$	$\langle z \rangle$	$\log M_h$ (2D fit)	$\log M_h$ (1D fit)
1	<9.4	9.21	2.51	$11.75^{+0.17}_{-0.14}$	$11.82^{+0.16}_{-0.26}$
2	9.4–9.8	9.61	2.48	$11.86^{+0.08}_{-0.11}$	$11.74^{+0.12}_{-0.20}$
3	9.8–10.1	9.93	2.45	$11.86^{+0.09}_{-0.17}$	$12.02^{+0.11}_{-0.22}$
4	>10.1	10.36	2.45	$11.80^{+0.12}_{-0.17}$	$12.07^{+0.09}_{-0.30}$

Note. Masses are expressed in units of $M_\odot$. 2D and 1D fits are to $\xi^{\rm Ly\,\alpha}$ and $w_p^{\rm Ly\alpha}$, respectively.

absorption around galaxies in the two more massive bins (panels (c) and (d)) than in the lower-mass bins (panels (a) and (b)), which is reflected in $-w_p$ being generally higher (panel (e)) in the higher-mass bins at large r_p. However, the covariance in $\xi^{\rm Ly\,\alpha}$ at large r_p makes it impossible to gauge the significance of such differences "by eye."

In fact, we infer statistically compatible halo masses in all stellar mass bins, as listed in Table 2. We will discuss the significance of this result in Section 6. The models again produce an acceptable fit quality, as judged by χ^2, in all cases. We found that these mass estimates were not significantly changed by fitting each stellar mass bin independently (thus removing interbin covariances in $\xi^{\rm Ly\,\alpha}$ as well as decoupling σ_z) or by fitting $\xi^{\rm Ly\,\alpha}$ over a reduced range of r_p and π. Turning to the projected $w_p^{\rm Ly\alpha}$, we found again that 1D fits produced noisier but consistent M_h estimates, which are listed in the last column of Table 2. Corner plots for both fits are shown in Appendix A.

6. Discussion

Our constraints on the SMHM relation at $z = 2.5$ from both galaxy–galaxy and galaxy–Lyα clustering are shown in Figure 11 and compared to earlier work. We begin by considering our galaxy–galaxy clustering results. We find moderate evidence of M_*-dependent galaxy clustering at 3.5σ significance: the slope of the SMHM relation, based on a linear fit to the halo masses derived in Section 2, is $d \log M_h / d \log M_* = 0.85 \pm 0.24$. Both the slope and normalization agree well with the variety of literature results shown in the left panel of Figure 11 that are ultimately derived (with the exception of the ASTRID curve) from analyses of imaging surveys, primarily M_*- and photometric redshift-dependent measures of galaxy abundances and angular correlation functions. Specifically, the average slope of the UniverseMachine (Behroozi et al. 2019) relation over the M_* range plotted in Figure 11 is $d \log M_h / d \log M_* = 0.58$, which is representative of the empirical curves plotted. This slope is consistent with our galaxy–galaxy clustering–based slope at the 1σ level. The SMHM relation slope in the ASTRID simulation is steeper ($d \log M_h / d \log M_* = 0.73$), closer to our galaxy–galaxy measure but, as we will see, farther from our galaxy–Lyα result.

Our measurements support an evolution in the SMHM relation, since they agree much more closely with the UniverseMachine relation at $z = 2.5$ than at $z = 0$. Furthermore, the consistency (Figure 11) of the halo masses inferred for LATIS galaxies, which consists of fairly UV-bright ($r < 24.8$) LBGs, and for photometric samples suggests that LBG clustering is representative of the full galaxy population at matched M_*.

Compared to earlier measures of galaxy two-point correlation functions (ξ and w_p) at similar redshifts based on spectroscopic samples, we find good agreement with bulk estimates of the halos hosting similarly selected LBGs, undifferentiated by stellar mass. Our threshold mass for the LATIS sample as a whole, $\log M_h^{\rm thresh}/M_\odot = 11.56^{+0.06}_{-0.15}$, is consistent at the $<2\sigma$ level with Trainor & Steidel (2012; $\log M_h^{\rm thresh}/M_\odot = 11.7 \pm 0.1$) and Bielby et al. (2013; $\log M_h^{\rm thresh}/M_\odot = 11.73 \pm 0.07$ and 11.57 ± 0.15 for their

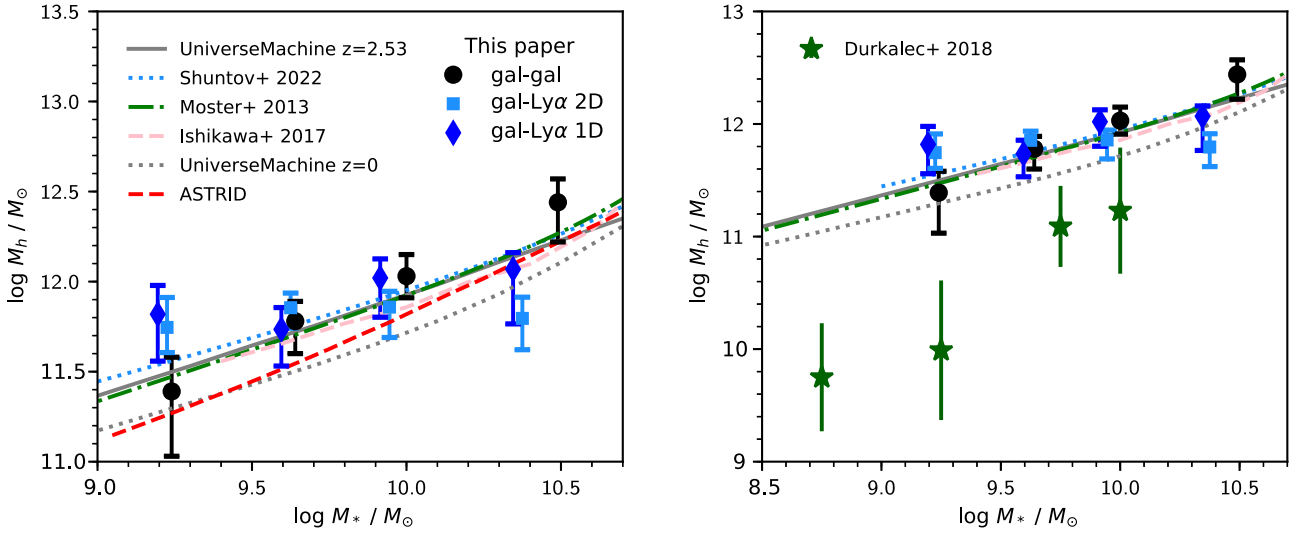

Figure 11. Left panel: the SMHM relation at $z = 2.5$, derived in this paper using galaxy–galaxy and galaxy–Lyα clustering (symbols indicated in legend), is compared to several literature results. The galaxy–Lyα points are slightly offset in $\log M_*$ for clarity. The Shuntov et al. (2022) results are based on M_*- and redshift-dependent abundance and clustering measurements using the COSMOS2020 photometric catalog; we interpolate their redshift bins to $z = 2.5$. The Ishikawa et al. (2017) relation is derived from the M_*-dependent angular correlation functions of u-band dropouts at $z \sim 3$ in CFHTLS imaging. The UniverseMachine (Behroozi et al. 2019) and Moster et al. (2013) relations are based on self-consistent multiepoch modeling of the observed stellar mass function (Moster) along with galaxy clustering and other inputs (Behroozi). We evaluate the analytic Moster et al. (2013) SMHM relation at $z = 2.5$, and we use the UniverseMachine DR1 tabulation of $\langle \log M_h | M_* \rangle$ at $z = 2.53$ and $z = 0$. The SMHM relation in the ASTRID simulation is shown, using the median M_h given M_* in the $z = 2.5$ snapshot. Right panel: the same results are shown on an expanded scale with the Durkalec et al. (2018) SMHM relation at $z \sim 3$ derived from galaxy–galaxy clustering in VUDS. We plot the HOD parameter $M_{\rm min}$ to represent the halo mass of their M_*-thresholded subsamples.

Very Large Telescope (VLT)+Keck and VLT-only samples, respectively). Compared to our initial estimate in Newman et al. (2022), $\log M_h^{\rm thresh}/M_\odot = 11.8 \pm 0.1$, the improved estimate in this paper is about 2σ lower.[7]

Prior to this work, the only derivation of the $z \sim 3$ SMHM relation from a spectroscopic survey was performed by Durkalec et al. (2015, 2018) based on VUDS. Durkalec et al. (2018) found a steep SMHM that would imply an extremely high star formation efficiency in low-mass, high-redshift galaxies. Although our sample does not reach their lowest-mass bin, it is clear that our galaxy–galaxy and galaxy–Lyα results are not consistent with Durkalec et al. (2018) in the three overlapping mass bins.[8] Small differences in the sample selection (e.g., limiting magnitude and mean redshift) are not a plausible explanation for this difference. Although we cannot determine the origin of the difference, we consider that the consistency of our SMHM measures with each other and with the other results in Figure 11 (left panel) strongly support a more conventional SMHM relation.

Now we turn to the galaxy–Lyα cross-correlation. We modeled and fit both the full 2D $\xi^{{\rm Ly}\,\alpha}$ and a projected 1D statistic $w_p^{{\rm Ly}\alpha}$. Although the velocity structure of the Lyα absorption provides no constraining power on the halo mass, the 2D fits nonetheless give more precise constraints on M_h, because they naturally weight the observed absorption as a function of π. The 1D fits, on the other hand, are insensitive to our modeling of both redshift measurement errors and the velocity structure of the gas, and we further removed sight lines within $r_p < 1\ h^{-1}$ cMpc that are expected to be most sensitive to feedback. Figure 11 shows that the masses derived from the 2D and 1D analyses are consistent. We do not detect a significant M_* dependence of the surrounding Lyα absorption, but we find that this is only moderately surprising given the uncertainties. Specifically we find slopes of $d \log M_h / d \log M_* = 0.00 \pm 0.16$ and 0.22 ± 0.22 for the 2D and 1D fits, respectively. Compared to the UniverseMachine SMHM relation, which is representative of the other models plotted in Figure 11, the measured slopes are shallower with significances of 3.5σ and 1.6σ. Thus the SMHM slope from the 1D fits is consistent with standard models, whereas the 2D fits give a notably shallow slope. The largest source of this flat slope is moderate tension in the highest-M_* bin between the halo masses inferred from 2D galaxy–Lyα clustering and galaxy–galaxy clustering (as well as the plotted models). This tension is much reduced in the 1D analysis, which could indicate that the velocity structure of the H I around massive galaxies in not correct in the simulations. However, the size of the uncertainties prevents any firm conclusion.

Compared to earlier studies of the galaxy–Lyα cross-correlation, our results are considerably more precise and are the first to estimate halo masses in multiple M_* bins quantitatively. For instance, our 1σ uncertainty in $\log M_h^{\rm thresh}$ is 0.09 dex compared to 0.2 and 0.6 dex in the Rakic et al. (2013) and Kim & Croft (2008) studies, respectively. This is a statistical error, which does not incorporate uncertainties in the estimation of the halo–Lyα cross-correlation from simulations. Rakic et al. (2013) found that the inclusion of active galactic nucleus (AGN) feedback in the OWLS simulations shifted their inferred halo mass by 0.3 dex. They regarded this as a mild uncertainty since it was comparable to their statistical errors, but it would be substantially larger than ours. However, we expect the effect of AGN feedback on our halo–Lyα cross-correlation to be much

[7] The main result of Newman et al. (2022), the low galaxy overdensity $\delta_{\rm gal}$ within the regions of strongest large-scale Lyα absorption, is not very sensitive to changes to $M_h^{\rm thresh}$. A forthcoming paper (A. B. Newman et al., 2023, in preparation.) will revisit and expand upon the Newman et al. (2022) analysis using the full LATIS maps.

[8] The significance of the difference in $\log M_h$ ranges from 1.4–2.0σ per M_* bin, with the Durkalec M_h always being lower. The average difference in $\log M_h$ is nonzero at a significance of $\sim 3\sigma$.

smaller. First, our measurements focus on larger scales; second, the AGN feedback implementation in OWLS was very strong. This can be seen, for instance, by comparing its dramatic effect on the star formation rate density (Schaye et al. 2010, Figure 18) to the subtle effect in the more recent simulations ASTRID, TNG, and SIMBA (Ni et al. 2023, Figure 4).

We clearly detect redshift-space distortions in the 2D $\xi^{\mathrm{Ly}\,\alpha}$ maps, with substantial compression in the π direction that we attribute to large-scale gas infall onto overdensities. Such infall has been previously observed around similar galaxy samples (Rudie et al. 2012; Rakic et al. 2013; Tummuangpak et al. 2014; Bielby et al. 2017; Chen et al. 2020). We found that the detailed shape of $\xi^{\mathrm{Ly}\,\alpha}$, especially the pinching at small r_p, is influenced by our spectral processing steps, specifically MFR. By carefully matching these steps in our synthesis of halo–Lyα cross-correlation models, we match the observed shape closely. Although the observed redshift-space distortions are not diagnostic of the halo mass and so are not a focus of this paper, they support the precision of the LATIS redshifts ($\sigma_z = 66 \pm 8$ km s^{-1} in our $\xi^{\mathrm{Ly}\,\alpha}$ fits).

Altogether, we find a consistent picture in which our M_*-dependent galaxy–galaxy and galaxy–Lyα correlation function measurements are accurately modeled by careful treament of cosmological simulations, leading to estimates of the SMHM that are generally consistent with each other and with canonical models largely based on photometric surveys. This is significant for several reasons. First, future analyses of the LATIS maps will rely on understanding the halo population occupied by our galaxy sample. A prime example is Newman et al. (2022), in which observed galaxy overdensities in different large-scale environments, as traced using 3D Lyα absorption maps, were compared to halo overdensities at matched M_h and environment in simulations.

Second, comparing halo masses estimated from galaxy–galaxy and galaxy–Lyα clustering provides a strong test of our ability to measure and model subtle differences in Lyα absorption. The most sensitive comparison is between the two analyses of the full LATIS galaxy sample. Galaxy–galaxy clustering led to a threshold mass of $\log M_h/M_\odot = 11.56^{+0.06}_{-0.15}$, and galaxy–Ly$\alpha$ clustering (2D) gave 11.54 ± 0.09. However, as discussed in Section 4.2, we expect small differences due to the different cosmological parameters and subhalo finders employed by the two simulations underlying these analyses. Based on Figure 7, the mass-thresholded halo population with $\log M_h^{\mathrm{thresh}}/M_\odot = 11.54$ in ASTRID has the same autocorrelation function the halo population with $\log M_h^{\mathrm{thresh}}/M_\odot > 11.47$ in MDPL2. Correcting for this known difference, we find that the threshold masses differ by $\Delta \log M_h = 0.09^{+0.11}_{-0.17}$. Here we have estimated the uncertainty assuming that errors in the galaxy–galaxy and galaxy–Lyα results are uncorrelated; the true error in $\Delta \log M_h$ must be smaller. Thus we find consistency at about the 0.1 dex level. This corresponds to 4% multiplicative changes in $\xi^{\mathrm{Ly}\,\alpha}$ and the δ_F measurements that underlie it, suggesting that we are able to measure and model Lyα absorption to a fairly high precision. Note for typical $w_p^{\mathrm{Ly}\alpha} \approx -0.1$ at $r_p \approx 4\ h^{-1}$ cMpc, 4% multiplicative errors correspond to $\Delta\delta_F = 0.004$.

Third, we find that galaxy–Lyα clustering gives similarly precise constraints on halo occupation as galaxy–galaxy clustering. We initially suspected that this conclusion would be somewhat particular to LATIS, since it is designed around deep exposures required to measure the Lyα forest in galaxy spectra. However, we found that increasing the random errors in δ_F by a factor of $\sqrt{2}$ when constructing our covariance matrix, effectively halving the exposure time, would increase the uncertainty in $\log M_h$ by only 10%. This modest increase indicates that intrinsic Lyα correlations are the major part of the error budget, and a wider, shallower survey would be more powerful for global correlation analyses. This suggests that galaxy–Lyα correlations come "for free" and will be a powerful application of any future large optical spectroscopic surveys of high-redshift galaxies.

7. Summary

We constrained the galaxy–halo connection at $z = 2.5$ through measurements of both galaxy two-point correlation functions and the cross-correlation between galaxies and transverse Lyα absorption, as detected in background galaxy spectra. The measurements are based on ∼3000 spectra collected by the LATIS survey. We presented a new method of measuring all auto- and cross-correlations among galaxies binned according to stellar mass, and by analyzing mock surveys within the MultiDark MDPL2 simulation, we derived a covariance matrix and estimated the typical halo mass associated with each bin. We then measured the galaxy–Lyα cross-correlation, both for the galaxy sample as a whole and in bins of stellar mass. We generated a large sample of synthetic Lyα spectra in the ASTRID simulation, carefully matched to the characteristics of LATIS spectra, and inferred halo masses by comparing the simulated halo–Lyα and observed galaxy–Lyα cross-correlations, both in 2D and 1D.

We found that the constraints from the two methods are consistent with one another and with standard SMHM relations derived from photometric surveys. We did not find evidence of an unusually steep SMHM relation (Durkalec et al. 2018) or of differences between the simulated and observed galaxy–Lyα cross-correlation functions (Momose et al. 2021b). These results will inform future work on the galaxy–IGM connection using the LATIS survey data, and they tightly constrain systematic errors in our measurements and modeling of the Lyα transmission fluctuations. Our results also highlight galaxy–Lyα clustering as a tool whose power will increase with future large optical spectroscopic surveys of the distant Universe.

Acknowledgments

This paper includes data gathered with the 6.5 m Magellan Telescopes located at Las Campanas Observatory, Chile. We gratefully acknowledge the support of the observatory staff. We thank the anonymous referee for thoughtful comments that improved the manuscript. This material is based upon work supported by the National Science Foundation under grant No. AST-2108014. B.C.L. acknowledges the support from the National Science Foundation under grant No. 1908422. S.B. was supported by NASA-80NSSC21K1840. M.Q. is supported by NSF grant AST-2107821. The authors acknowledge the Frontera computing project at the Texas Advanced Computing Center (TACC) for providing HPC and storage resources that have contributed to the research results reported within this paper. Frontera is made possible by NSF OAC-1818253.

Appendix A
Corner Plots

Figures 12–14 demonstrate 1D and 2D marginalized posteriors for our models fit to galaxy–galaxy and galaxy–Lyα clustering data.

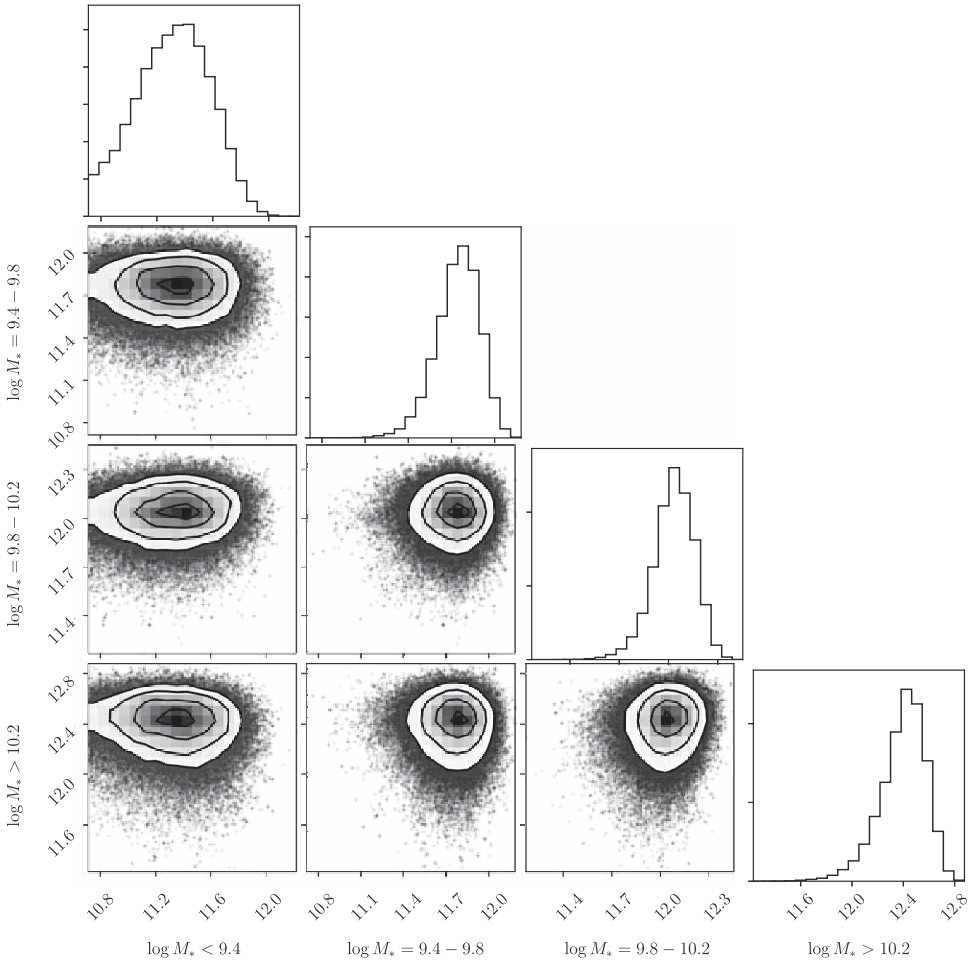

Figure 12. Corner plot illustrating posteriors for our model fit to the stellar mass–dependent galaxy–galaxy clustering data in Figure 4. The four parameters correspond to the halo mass $\log M_h/M_\odot$ for LATIS galaxies in the stellar mass intervals indicated. Contours show the 1, 2, and 3σ levels while fully marginalized posteriors are shown at the top of each row.

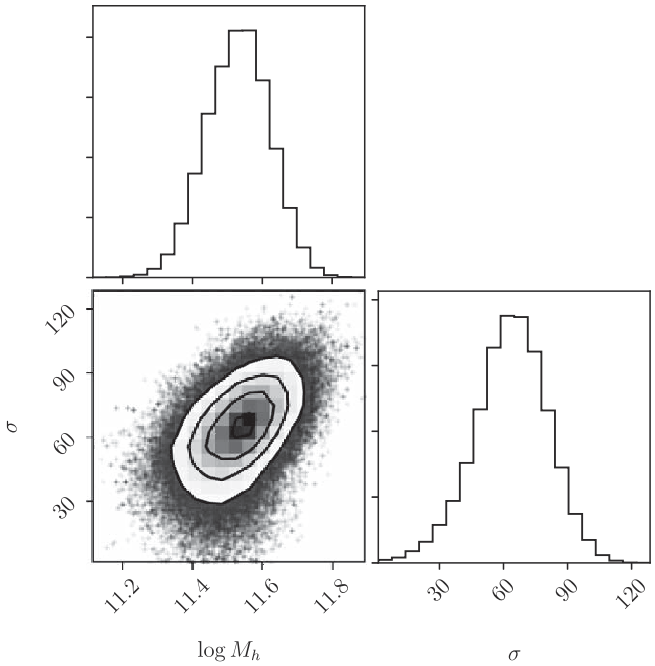

Figure 13. Corner plot illustrating posteriors for our fit to the 2D galaxy–Lyα cross-correlation $\xi^{\mathrm{Ly}\,\alpha}$ for the entire LATIS sample; data are shown in Figure 6.

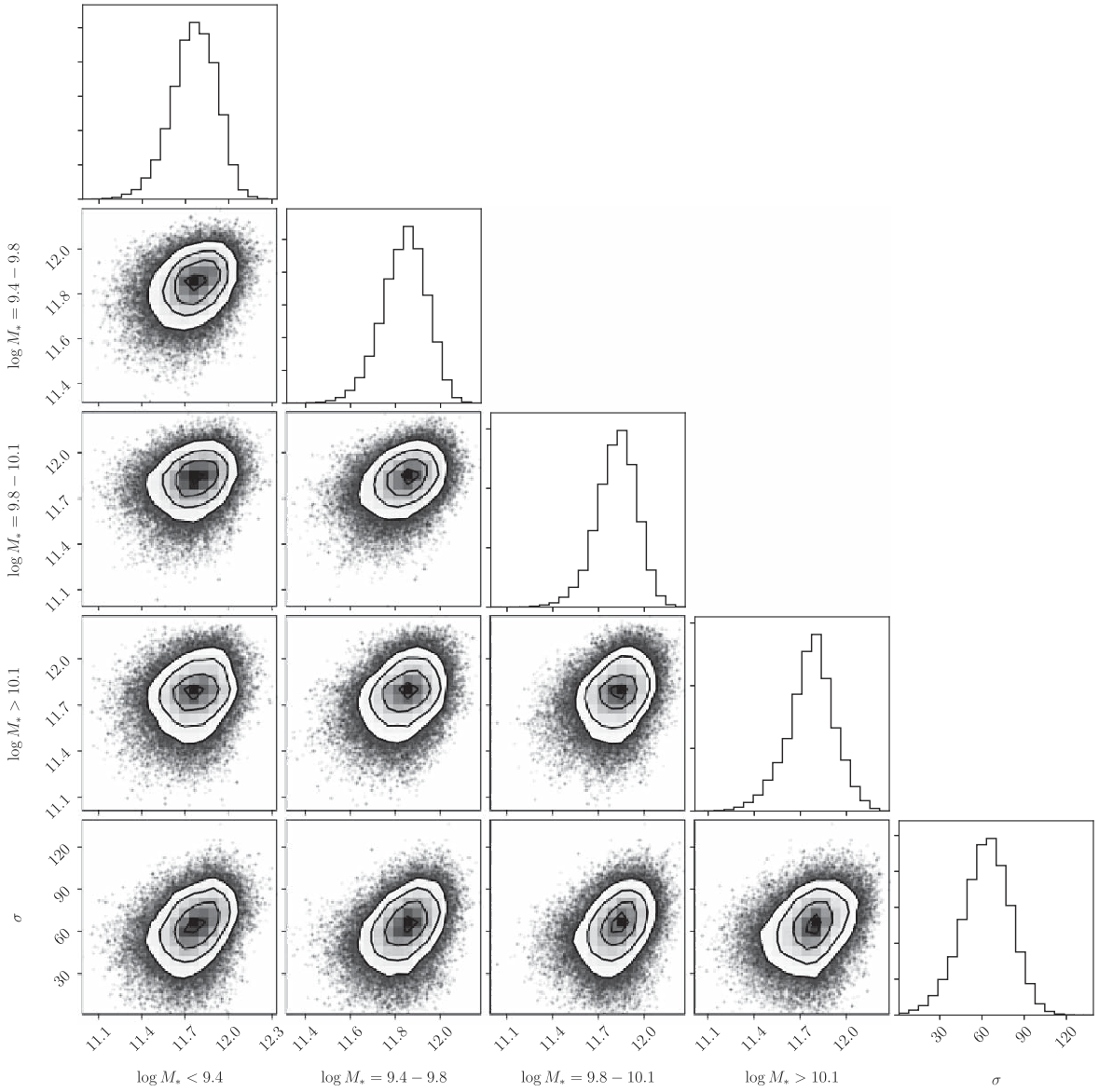

Figure 14. Corner plot illustrating posteriors for our fit to the 2D galaxy–Lyα cross-correlation $\xi^{\mathrm{Ly}\,\alpha}$ in four bins of stellar mass. Similar to Figure 12, each parameter describes the halo mass $\log M_h/M_\odot$ for LATIS galaxies in the stellar mass intervals indicated.

Appendix B
Angular Variation of the TSR and SSR

The target sampling rate (TSR) and spectroscopic success rate (SSR) vary within the LATIS survey footprint. Here we describe a model of the angular dependence of these quantities.

B.1. TSR

We first consider the TSR. We built an initial model of the area surveyed by each of the 12 LATIS footprints, based on the IMACS field of view excluding masked regions surrounding bright stars. Note that the field of view has an asymmetric shape in the east–west direction due to a lateral shift of the dispered spectra as described by Newman et al. (2020). We note that in the D1 field, the CFHTLS imaging boundary sets the northern edge. We created a model $\mathrm{TSR}_{\mathrm{foot}}$ that accounts for variation of the TSR from footprint to footprint and in regions of overlapping footprints. First, within the unmasked and nonoverlapping portion of each footprint, we compute the TSR and assign this to $\mathrm{TSR}_{\mathrm{foot}}$. For this calculation, and throughout this paper, we exclude galaxies that were observed only as part of a "bright target" mask (see Newman et al. 2020) intended for poor weather conditions, since these masks do not cover the full survey area and were not uniformly observed. Second, within each field, we calculated the average TSR within regions where two footprints overlap. We considered separately overlaps in the east–west and north–south directions, since we attempted to reobserve objects in the east–west overlap regions more frequently, partly to compensate for the poorer image quality at the edge of the IMACS field (see below). We found that the individual east–west or north–south overlap regions within each field were consistent with having the same TSR, and we therefore set $\mathrm{TSR}_{\mathrm{foot}}$ to the mean to reduce noise.

We then considered the probability of target selection as a function of position relative to the footprint center, $\mathrm{TSR}_{\mathrm{mask}}$. In computing this refined TSR, we weighted targets by the inverse of $\mathrm{TSR}_{\mathrm{foot}}$ in order to quantify the residual TSR variation. We

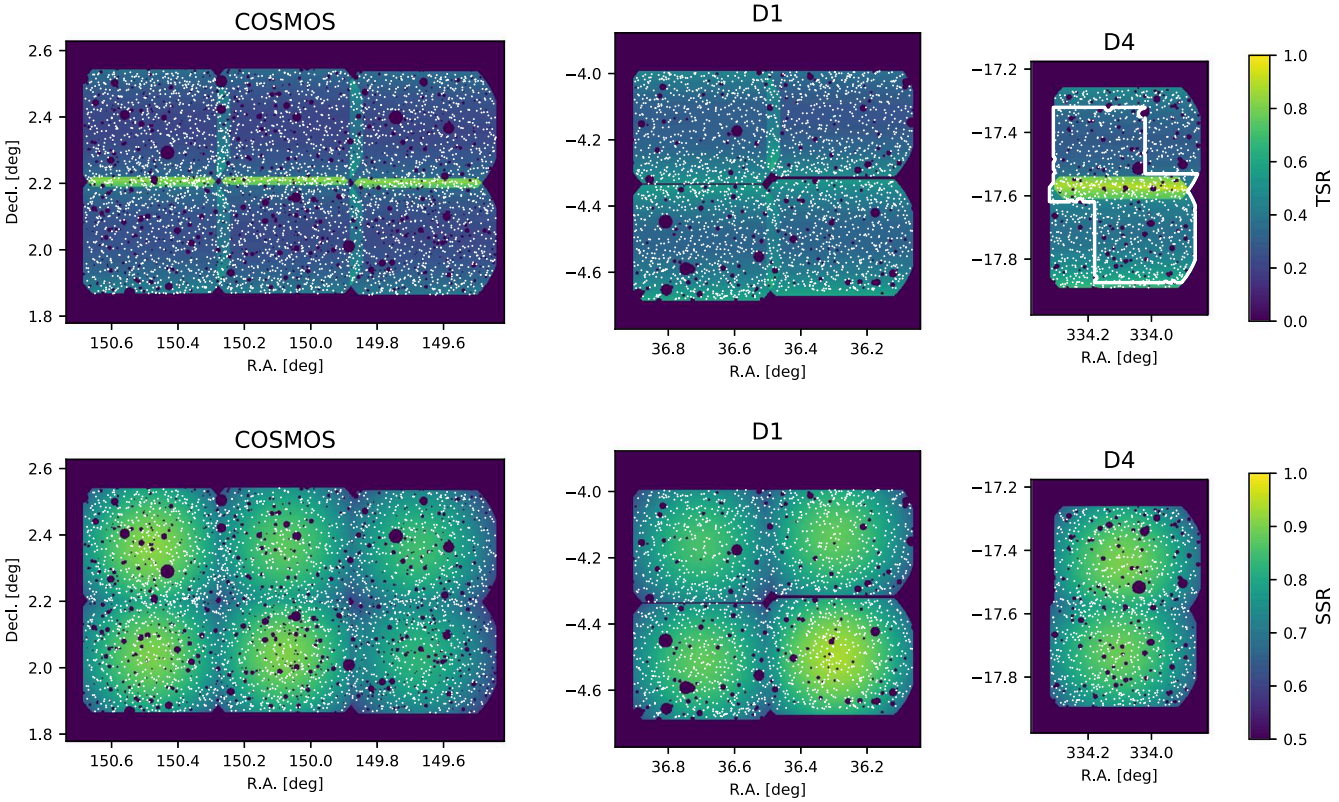

Figure 15. Top row: the variation of the TSR within each of the LATIS survey fields is indicated by the color bar. White points show the positions of targeted galaxies. The white contour in the D4 field shows the subregion covered by near-infrared imaging that is used in this paper. Bottom row: the variation of the SSR, with white points showing the position of galaxies with secure redshifts. Note the compressed color scale.

found that $\mathrm{TSR}_{\mathrm{mask}}$ is primarily a function of vertical (i.e., north–south) position relative to the mask center, as shown in Figure 16, such that the TSR is higher near the top and bottom of a footprint than at the midline. This dependence arises from the direction of dispersion (north–south) and the requirement that spectra not overlap, which tends to sort targets into ranks. We modeled $\mathrm{TSR}_{\mathrm{mask}}$ as a quadratic function of vertical mask position. The final TSR model is then the product of $\mathrm{TSR}_{\mathrm{foot}}$ and $\mathrm{TSR}_{\mathrm{mask}}$.

Figure 15 shows our model of the TSR within each field. The median TSR is lower in COSMOS (0.28) than in D1 (0.37) and D4 (0.39) because the parent sample is larger: we include both z_{phot}- and *ugr*-selected galaxies in COSMOS and only the latter in D1 and D4. For the purpose of this paper, only the relative variation of the TSR within each field is relevant, but for other applications, it may be important to recognize that the TSR depends on magnitude and is significantly higher for galaxies with $r < 24.4$ (see Newman et al. 2020). The NMAD (normalized median absolute deviation) of the TSR within each field is 6%–10% of the median, so despite the structure seen in the fgure, the overall target sampling is fairly uniform within the LATIS fields.

In this paper, an additional criterion for a galaxy to be included in our sample is the availability of adequately deep near-infrared imaging. Such imaging in the D4 field is shallower than in the other fields and does not cover the entire LATIS area. We found that galaxies outside the white contour in Figure 15 lack a near-infrared detection much more frequently than in the rest of the D4 field, and we therefore do not consider the region outside the contour in this paper.

B.2. SSR

We now consider the SSR, defined as the fraction of targeted sources for which a high-confidence redshift was determined (`zqual` = 3 or 4). Low-confidence redshifts are not used in this paper. The most significant spatial dependence of the SSR was found to be the distance of a target from the footprint center. This dependence arises because the size of images produced by the IMACS f/2 camera increases significantly with field radius, which reduces the signal-to-noise ratio of the spectra. The right panel of Figure 16 shows the relative variation of the SSR with this radial distance, defined to average to unity. We use a quadratic fit to estimate $\mathrm{SSR}_{\mathrm{mask}}$. Galaxies with increased exposure due to repeat observations were not included in this calculation in order to isolate the radial dependence of SSR from any spatial dependence of exposure time, which will be considered in the next step.

We then consider large-scale variations $\mathrm{SSR}_{\mathrm{foot}}$, which may arise from different observing conditions when different footprints are observed, or from the fact that some sources that fall within two footprints receive a longer integration time. When calculating $\mathrm{SSR}_{\mathrm{foot}}$, we weight galaxies by the inverse of $\mathrm{SSR}_{\mathrm{mask}}$ in order to determine the residual spatial variations. We calculate the weighted SSR within each footprint, first considering only the nonoverlapping part, and find that footprint-to-footprint variations in LATIS are small (mean of 0.79; standard deviation of 0.026). We next compared the weighted SSR within overlapping and nonoverlapping regions, and found that SSR differences were small (a few percent) and within the expected Poisson fluctuations. Since the overlap regions are small and the overall variation in SSR is weak, we

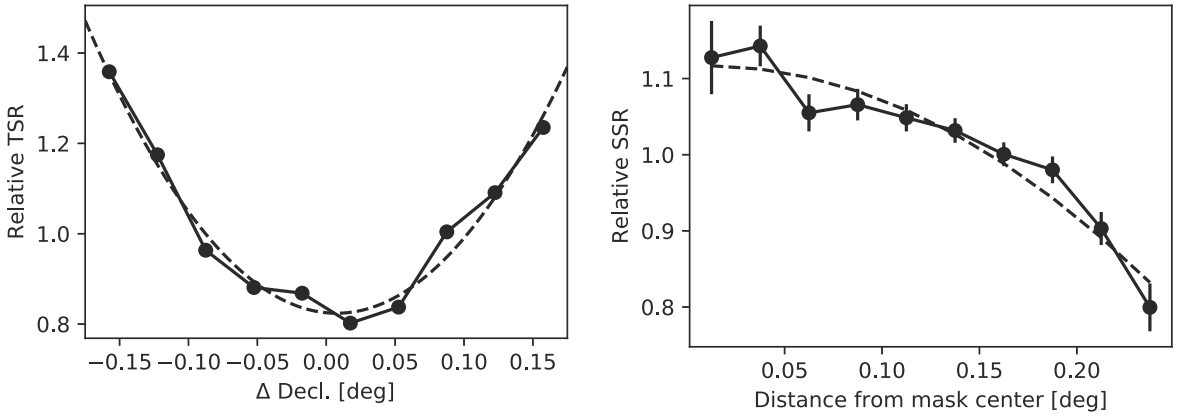

Figure 16. Left: the variation of TSR_{mask} with the decl. of a target relative to the mask center. Black points and solid line show the measurements, averaged over all footprints, while the dashed line shows the quadratic fit used in our modeling. Right: the variation of SSR_{mask} with radial distance of a target relative to the mask center, along with a quadratic fit as in the left panel.

simply set SSR_{foot} within overlap regions to the average value of the two footprints.

The overall SSR is given by the product of SSR_{mask} and SSR_{foot} and is shown in Figure 15. The ESR is the product of the TSR and SSR.

ORCID iDs

Andrew B. Newman https://orcid.org/0000-0001-7769-8660
Mahdi Qezlou https://orcid.org/0000-0001-7066-1240
Nima Chartab https://orcid.org/0000-0003-3691-937X
Gwen C. Rudie https://orcid.org/0000-0002-8459-5413
Guillermo A. Blanc https://orcid.org/0000-0003-4218-3944
Simeon Bird https://orcid.org/0000-0001-5803-5490
Andrew J. Benson https://orcid.org/0000-0001-5501-6008
Daniel D. Kelson https://orcid.org/0000-0003-4727-4327
Brian C. Lemaux https://orcid.org/0000-0002-1428-7036

References

Adelberger, K. L., Erb, D. K., Steidel, C. C., et al. 2005, ApJL, 620, L75
Adelberger, K. L., Steidel, C. C., Giavalisco, M., et al. 1998, ApJ, 505, 18
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Allen, P. D., Moustakas, L. A., Dalton, G., et al. 2005, MNRAS, 360, 1244
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJ, 770, 57
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013b, ApJ, 763, 18
Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. 2000, MNRAS, 311, 793
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Bielby, R., Hill, M. D., Shanks, T., et al. 2013, MNRAS, 430, 425
Bielby, R. M., Gonzalez-Perez, V., McCracken, H. J., et al. 2014, A&A, 568, A24
Bielby, R. M., Shanks, T., Crighton, N. H. M., et al. 2017, MNRAS, 471, 2174
Bird, S., Ni, Y., Di Matteo, T., et al. 2022, MNRAS, 512, 3703
Blain, A. W., Chapman, S. C., Smail, I., & Ivison, R. 2004, ApJ, 611, 725
Chen, Y., Steidel, C. C., Hummels, C. B., et al. 2020, MNRAS, 499, 1721
Conroy, C., & Wechsler, R. H. 2009, ApJ, 696, 620
Cooray, A. 2006, MNRAS, 365, 842
Cooray, A., & Sheth, R. 2002, PhR, 372, 1
Cowley, W. I., Caputi, K. I., Deshmukh, S., et al. 2018, ApJ, 853, 69
Crighton, N. H. M., Bielby, R., Shanks, T., et al. 2011, MNRAS, 414, 28
Daddi, E., Cimatti, A., Pozzetti, L., et al. 2000, A&A, 361, 535
de la Torre, S., Guzzo, L., Peacock, J. A., et al. 2013, A&A, 557, A54
Diemer, B. 2018, ApJS, 239, 35
Durkalec, A., Le Fèvre, O., de la Torre, S., et al. 2015, A&A, 576, L7
Durkalec, A., Le Fèvre, O., Pollo, A., et al. 2018, A&A, 612, A42
Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, ApJ, 688, 85
Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, JCAP, 2012, 059
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Foucaud, S., Conselice, C. J., Hartley, W. G., et al. 2010, MNRAS, 406, 147
Foucaud, S., McCracken, H. J., Le Fèvre, O., et al. 2003, A&A, 409, 835
Giavalisco, M., & Dickinson, M. 2001, ApJ, 550, 177
Giavalisco, M., Steidel, C. C., Adelberger, K. L., et al. 1998, ApJ, 503, 543
Guo, H., Zheng, Z., Behroozi, P. S., et al. 2016, MNRAS, 459, 3040
Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, MNRAS, 404, 1111
Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, A&A, 498, 725
Horowitz, B., Lee, K.-G., Ata, M., et al. 2022, ApJS, 263, 27
Infante, L. 1994, A&A, 282, 353
Ishikawa, S., Kashikawa, N., Toshikawa, J., et al. 2017, ApJ, 841, 8
Jing, Y. P., Mo, H. J., & Börner, G. 1998, ApJ, 494, 1
Kashino, D., More, S., Silverman, J. D., et al. 2017, ApJ, 843, 138
Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Kerscher, M., Szapudi, I., & Szalay, A. S. 2000, ApJL, 535, L13
Kim, Y.-R., & Croft, R. A. C. 2008, MNRAS, 387, 377
Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159
Lee, K.-G., Krolewski, A., White, M., et al. 2018, ApJS, 237, 31
Lee, K.-G., Suzuki, N., & Spergel, D. N. 2012, AJ, 143, 51
Lee, K.-S., Giavalisco, M., Gnedin, O. Y., et al. 2006, ApJ, 642, 63
Marulli, F., Bolzonella, M., Branchini, E., et al. 2013, A&A, 557, A17
McCracken, H. J., Wolk, M., Colombi, S., et al. 2015, MNRAS, 449, 901
Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
Momose, R., Shimasaku, K., Kashikawa, N., et al. 2021a, ApJ, 909, 117
Momose, R., Shimizu, I., Nagamine, K., et al. 2021b, ApJ, 911, 98
Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121
Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903
Moustakas, L. A., & Somerville, R. S. 2002, ApJ, 577, 1
Newman, A. B., Rudie, G. C., Blanc, G. A., et al. 2020, ApJ, 891, 147
Newman, A. B., Rudie, G. C., Blanc, G. A., et al. 2022, Natur, 606, 475
Ni, Y., Di Matteo, T., Bird, S., et al. 2022, MNRAS, 513, 670
Ni, Y., Genel, S., Anglés-Alcázar, D., et al. 2023, ApJ, 959, 136
Ouchi, M., Shimasaku, K., Okamura, S., et al. 2001, ApJL, 558, L83
Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, ApJ, 611, 685
Pujol, A., Hoffmann, K., Jiménez, N., & Gaztañaga, E. 2017, A&A, 598, A103
Qezlou, M., Bird, S., Lidz, A., et al. 2023, MNRAS, 524, 1933
Qezlou, M., Newman, A. B., Rudie, G. C., & Bird, S. 2022, ApJ, 930, 109
Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013, MNRAS, 430, 2427
Rakic, O., Schaye, J., Steidel, C. C., et al. 2013, MNRAS, 433, 3103
Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2012, ApJ, 751, 94
Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30
Roche, N. D., Almaini, O., Dunlop, J., Ivison, R. J., & Willott, C. J. 2002, MNRAS, 337, 1282

Rudie, G. C., Steidel, C. C., Trainor, R. F., et al. 2012, ApJ, 750, 67
Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, MNRAS, 402, 1536
Schlegel, D. J., Kollmeier, J. A., Aldering, G., et al. 2022, arXiv:2209.04322
Shi, J., & Sheth, R. K. 2018, MNRAS, 473, 2486
Shuntov, M., McCracken, H. J., Gavazzi, R., et al. 2022, A&A, 664, A61
Sinha, M., & Garrison, L. H. 2020, MNRAS, 491, 3022
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Trainor, R. F., & Steidel, C. C. 2012, ApJ, 752, 39
Tummuangpak, P., Bielby, R. M., Shanks, T., et al. 2014, MNRAS, 442, 2094
Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189
van den Bosch, F. C., Yang, X., Mo, H. J., et al. 2007, MNRAS, 376, 841
Wake, D. A., Whitaker, K. E., Labbé, I., et al. 2011, ApJ, 728, 46
Webb, T. M., Eales, S. A., Lilly, S. J., et al. 2003, ApJ, 587, 41
Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zheng, Z. 2004, ApJ, 610, 61
Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760
Zheng, Z., & Guo, H. 2016, MNRAS, 458, 4015