



Measuring the Density Fields around Bright Quasars at $z \sim 6$ with XQR-30 Spectra

Huanqing Chen¹, Anna-Christina Eilers^{2,12}, Sarah E. I. Bosman³, Nickolay Y. Gnedin^{1,4,5}, Xiaohui Fan⁶, Feige Wang^{6,12}, Jinyi Yang^{6,13}, Valentina D’Odorico^{7,8,9}, George D. Becker¹⁰, Manuela Bischetti⁷, Chiara Mazzucchelli^{11,14}, Andrei Mesinger⁸, and Andrea Pallottini⁸

¹ Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA; hqchen@uchicago.edu

² MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

³ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

⁴ Theoretical Physics Department, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA

⁵ Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, USA

⁶ Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA

⁷ INAF-Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy

⁸ Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

⁹ IFPU-Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy

¹⁰ Department of Physics & Astronomy, University of California, Riverside, CA 92521, USA

¹¹ European Southern Observatory, Alonso de Cordova 3107, Vitacura, Region Metropolitana, Chile

Received 2021 October 25; revised 2022 April 5; accepted 2022 April 6; published 2022 May 20

Abstract

Measuring the density of the intergalactic medium using quasar sight lines in the epoch of reionization is challenging due to the saturation of Ly α absorption. Near a luminous quasar, however, the enhanced radiation creates a proximity zone observable in the quasar spectra where the Ly α absorption is not saturated. In this study, we use 10 high-resolution ($R \gtrsim 10,000$) $z \sim 6$ quasar spectra from the extended XQR-30 sample to measure the density field in the quasar proximity zones. We find a variety of environments within 3 pMpc distance from the quasars. We compare the observed density cumulative distribution function (CDF) with models from the Cosmic Reionization on Computers simulation and find a good agreement between 1.5 and 3 pMpc from the quasar. This region is far away from the quasar hosts and hence approaching the mean density of the universe, which allows us to use the CDF to set constraints on the cosmological parameter $\sigma_8 = 0.6 \pm 0.3$. The uncertainty is mainly due to the limited number of high-quality quasar sight lines currently available. Utilizing the more than 200 known quasars at $z \gtrsim 6$, this method will allow us to tighten the constraint on σ_8 to the percent level in the future. In the region closer to the quasar within 1.5 pMpc, we find that the density is higher than predicted in the simulation by 1.23 ± 0.17 , suggesting that the typical host dark matter halo mass of a bright quasar ($M_{1450} < -26.5$) at $z \sim 6$ is $\log_{10}(M_h/M_\odot) = 12.5^{+0.4}_{-0.7}$.

Unified Astronomy Thesaurus concepts: Reionization (1383); Quasars (1319); Interstellar medium (847); Cosmology (343); Large-scale structure of the universe (902); AGN host galaxies (2017)

1. Introduction

Modern cosmology shows that the structures we see today, including galaxies and quasars, emerge from the small fluctuation in the initial density field. Measuring the cosmic density field is, therefore, an important step in studying both galaxy and quasar formation. On the astrophysics side, the large-scale density environment impacts the accretion history of galaxies, thus affecting small-scale processes like star formation and quasar activities (e.g., Springel et al. 2005; Narayanan et al. 2008; Overzier et al. 2009; Di Matteo et al. 2017; Habouzit et al. 2019). On the cosmology side, the large-scale density field contains rich information such as the initial density fluctuation and matter contents of the universe, and measuring it allows us to constrain cosmology (e.g., Croft et al.

2002; Abbott et al. 2018; Chabanier et al. 2019; Hawken et al. 2020).

There are two primary methods for measuring the large-scale structure: (1) mapping the position of discrete galaxies and (2) measuring the continuous density field using Ly α absorption from the intergalactic medium (IGM). At intermediate redshift ($z < 5$), multiple surveys utilized these methods to return fruitful results on cosmology and galaxy formation. For example, Eisenstein et al. (2005) detected the baryon acoustic oscillations (BAOs) imprinted on the cosmic web and put tight constraints on the matter content of the universe. Croft et al. (2002) and McDonald et al. (2006) measured the Ly α opacity in the Ly α forest and inferred the linear matter power spectrum. Apart from constraining cosmology, such surveys also shed light on the environments of different types of galaxies, notably quasar host galaxies. For example, Shen et al. (2007, 2009) measured their clustering signals in the Sloan Digital Sky Survey (SDSS) quasar sample and constrained not only the quasar host halo masses at $z \lesssim 5$ but also their duty cycles. Such results provide insights for modeling quasars in cosmological simulations (e.g., Schaye et al. 2015; Pillepich et al. 2018).

However, at higher redshift ($z \gtrsim 6$), measuring the density field is extremely hard. The number of observed galaxies is low

¹² NASA Hubble Fellow.

¹³ Strittmatter Fellow.

¹⁴ ESO Fellow.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

(Bouwens et al. 2017), several orders of magnitude smaller than the number of low- z galaxies mapped by surveys like SDSS and the Dark Energy Survey (Sánchez et al. 2014). Therefore, it is not practical to constrain cosmology with methods commonly used at low redshift, such as detecting the BAO signal. Although surveying a large region of the sky to detect enough $z \gtrsim 6$ galaxies for constraining cosmological parameters is not practical in the near future, detecting $z \gtrsim 6$ galaxies in a few small deep fields is still manageable. In fact, many recent studies have been focusing on mapping galaxies around $z \gtrsim 6$ quasars to uncover the underlying density environments of the first quasars (see, e.g., Kim et al. 2009; Bañados et al. 2013; McGreer et al. 2014; Simpson et al. 2014; Champagne et al. 2018; Ota et al. 2018; Habouzit et al. 2019; García-Vergara et al. 2022, and references therein). However, because galaxies are biased tracers of the density field and may be impacted by the radiation feedback from the quasar (Kitayama et al. 2000; Kashikawa et al. 2007; Bosman et al. 2020; Chen 2020), it is hard to estimate the real underlying density field using a few detected galaxies. Furthermore, galaxy detecting techniques suffer severely from observational issues. In optical wavelengths, because the selection is done based on only a few wave bands, the completeness of the observational samples is hard to quantify (Kashikawa et al. 2007). Using interferometers in the radio, the foreground source contamination is less of a problem; however, the fields of view of the interferometers are usually very small, limiting our ability to map the large-scale environments of the first quasars (Champagne et al. 2018).

One complementary method for measuring the density field that is independent from galaxy mapping is using the Ly α absorption in quasar spectra. Although this method has been widely used at redshifts $z < 5$ for decades (Rauch 1998; Weinberg et al. 2003), so far, there has not been any attempt to measure the large-scale density field using the Ly α absorption spectra at $z \gtrsim 6$. This is because at $z \gtrsim 6$, the relatively large amount of residual neutral hydrogen in the IGM causes the Ly α absorption to saturate in most places. However, there are still some large regions where the Ly α absorption is not saturated: quasar proximity zones (Mesinger et al. 2004; Bolton & Haehnelt 2007). A bright quasar at $z \gtrsim 6$ can reduce the neutral fraction of hydrogen near it, creating a large zone where the Ly α absorption is not saturated. Such regions can easily extend to dozens of comoving Mpc (cMpc).¹⁵ Recently, Chen & Gnedin (2021a) showed that we can recover the density field along the quasar lines of sight in the proximity zones. This provides a new, independent way to measure the large-scale density field at $z \gtrsim 6$. With a large sample of quasar spectra, we can use the measured density field not only to constrain cosmology but also to understand the typical density environment of the first quasars. Although measuring lines of sight alone only gives us 1D information, we expect that quasars have no preferential pointing directions, at least on large spatial scales. Therefore, using a sample of them will give us an “averaged” measurement of the density environments of the first quasars. In recent decades, we have gained high-resolution spectra, which show detailed features in the proximity zones (see, e.g., Eilers et al. 2017). In this paper, we use a sample of high-resolution, high signal-to-noise ratio (S/N) spectra from the XQR-30 (main + extended) sample to recover the density

field surrounding $z \sim 6$ quasars. By comparing the observed density field with the Cosmic Reionization on Computers (CROC) simulation (Gnedin 2014; Chen & Gnedin 2021b), we also put constraints on the cosmological parameters and quasar properties at $z \sim 6$.

This paper is organized as follows. In Section 2, we describe the basics of the density recovery method, quasar sample, and measurement process. In Section 3, we show the recovered density field of the 10 quasar sight lines in our sample and their cumulative distribution functions (CDFs). In Section 4, we compare the observed CDF with our simulation, investigate various factors that impact the CDF, and study how the CDF at different distance ranges helps us constrain the cosmological parameter σ_8 , the typical halo mass of quasar hosts, and the quasar lifetime. A summary is provided in Section 5.

2. Method

In this study, we use a sample of high-S/N quasars at $z > 6$ to measure the density field, following the method described in Chen & Gnedin (2021a). We briefly describe the method here and refer the reader to that paper for details. The method is based on the robust assumption that in the proximity zone, the IGM is in ionization equilibrium,

$$\Gamma(d)n_{\text{HI}} = \alpha n_{\text{HII}}n_e = \alpha \left(\frac{n_e}{n_{\text{H}}} \right) n_{\text{H}}^2, \quad (1)$$

where Γ is the ionization rate of H I of the gas at the location with distance d away from the quasar, α is the recombination coefficient of H II, and n_{HI} , n_{HII} , n_{H} , and n_e are the number density of neutral hydrogen, ionized hydrogen, total hydrogen, and free electrons, respectively. Inside the proximity zones, the quasar dominates the radiation field (Calverley et al. 2011; Chen & Gnedin 2021a), and the IGM is completely ionized and transparent ($\Gamma(d) \propto d^{-2}$) for the majority of the sight lines. The recombination rate α can be treated as independent of density to first order in the very aftermath of reionization. The number density of neutral hydrogen is proportional to the optical depth at the corresponding pixel. Thus, we have

$$\tau = \text{const} \times \frac{\alpha}{\Gamma(d)} \left(\frac{n_e}{n_{\text{H}}} \right) n_{\text{H}}^2. \quad (2)$$

To obtain the density field, we only need to factor out the “constants” $\Gamma(d)$, α , and n_e/n_{H} by dividing this equation by the same equation for a baseline model.

Such a baseline model is obtained by running a 1D radiative transfer (RT) code on a uniform sight line. The 1D RT code is described in Chen & Gnedin (2021b), who used an adaptive time step to evolve the ionization state of hydrogen and helium, as well as gas temperature. For each quasar, we create a sight line of the uniform IGM density at the cosmic mean at the corresponding redshift. The initial values of the IGM we adopt before the quasar turns on are, based on the mean value of the CROC simulation, $x_{\text{HI}} = 10^{-4}$, $x_{\text{HeI}} = 10^{-4}$, $x_{\text{HeII}} = 0.9$, and $T = 10^4$ K. We then postprocess the uniform sight line with the same quasar ionizing spectra of the observed quasar (Section 2.3). We set the quasar lifetime to be 30 Myr for the fiducial baseline model. This way, we obtain the optical depth $\bar{\tau}$ for this uniform sight line, and the density field of the

¹⁵ We use cMpc to denote comoving Mpc and pMpc to denote physical Mpc.

observed quasar is

$$\Delta \equiv 1 + \delta = \sqrt{\frac{\tau}{\bar{\tau}}}, \quad (3)$$

where τ is the observed optical depth for the quasar, and $\delta = \rho/\bar{\rho} - 1$ is the cosmic overdensity. Note that the recovered density Δ is a geometric mean of real space and redshift space density $\sqrt{\Delta_r \Delta_z}$ (Chen & Gnedin 2021a),

$$\Delta = \sqrt{\Delta_r \Delta_z} \equiv \Delta_r \sqrt{\left| \frac{H}{H + dv_{\text{pec}}/dr} \right|}, \quad (4)$$

where Δ_r is the real space gas density (in units of the cosmic mean), H is the Hubble constant at the quasar redshift, v_{pec} is the peculiar velocity along the line of sight, and r is the proper distance from the quasar. As shown in Chen & Gnedin (2021a), Equation (3) accurately recovers the true density, with a moderate scatter of $\lesssim 10\%$ due to some temperature fluctuation and slight deviations in the radiation profiles from r^{-2} .

2.1. Quasar Sample

We use a sample of high spectral resolution and S/N quasars to measure the optical depth in the proximity zone and recover the density field. Currently, the most suitable sample for this task is the XQR-30 sample,¹⁶ an ongoing survey of bright $z > 5.8$ quasars using the high-resolution X-Shooter spectrograph (Vernet et al. 2011) on the Very Large Telescope. The main sample of XQR-30 consists of 30 quasars with new high-S/N observations. The extended XQR-30 sample also includes archival spectra taken with X-Shooter, which are reduced in a similar manner to the XQR-30 main sample and have comparable or higher S/Ns (Bosman et al. 2021a; D’Odorico et al. in preparation). The S/Ns of these spectra in the proximity zones, the regions we are interested in, are ~ 100 .

In order to avoid contamination of the density measurement by other absorption, we exclude quasars with broad absorption lines (BALs) or proximate damped Ly α systems. Moreover, using Ly α spectra in the proximity zone to recover the density field requires an accurate measurement of the quasar redshift. Among all of the redshift measurement techniques, submillimeter lines ([C II] or CO) provide the most reliable and precise measurement. Because they are thought to trace the cold interstellar medium, the systematic uncertainty is estimated to be smaller than 100 km s^{-1} (equivalent to $\Delta z \approx 0.002$). With spectra measured from the Atacama Large Millimeter/submillimeter Array or NOEMA, the error (usually a few 10 km s^{-1}) in fitting the submillimeter lines is far smaller than the systematic uncertainty. Therefore, we further constrain the sample by only including quasars with redshift measurements from submillimeter [C II] or CO lines. With these constraints, our final sample is listed in Table 1.

In column 6 of Table 1, we list the absolute magnitude M_{1450} of each quasar at rest-frame wavelength $1450 \text{ \AA}$, collected from the literature listed in column 7. In column 8, we list the average FWHM spectral resolution determined from the measurement of the FWHM of the telluric lines in the X-Shooter single frames. The two reported values, 22 and 26 km s^{-1} , correspond to the different slit widths of 0.7 and $0''.9$, respectively, adopted for the VIS arm (see V. D’Odorico

et al. 2022, in preparation, for more details). At $z \sim 6$, the Hubble parameter is $H \approx 700 \text{ km s}^{-1} \text{ Mpc}^{-1}$. This means that with the high-quality XQR-30 spectra, we can recover the density field with the equivalent spatial resolution of $\sim 20 \text{ pkpc}$. The quasar spectra are shown in Figure 1. Note that in the quasar proximity zones, all 10 quasars have $S/N \approx 100$. At $\lambda_{\text{rest}} \gtrsim 1230 \text{ \AA}$, some quasars have a slightly lower S/N due to the presence of residuals of telluric line subtraction. However, the small change in the S/N on the red side of the spectrum does not influence our results, since the analysis is done on the spectral pixels on the blue side of the Ly α . For details of the reduction procedure, see Bosman et al. (2021a), Zhu et al. (2021), and D’Odorico et al. in preparation.

2.2. Continuum Fitting

We reconstruct the unabsorbed quasar emission in the Ly α line by performing a principal component analysis (PCA) on a large set of quasars at lower z . Our procedure follows the log-PCA approach of Davies et al. (2018) with the improvements introduced in Bosman et al. (2021b). The specific PCA used in this paper differs from previous work mainly in the wavelength range we predict and the training/testing sample. Here we broadly summarize the procedure and highlight the ways in which it differs from previous work.

The PCA reconstructions function by obtaining optimal linear decompositions of quasar spectra over two wavelength ranges: the unabsorbed “red side” of the spectrum (in this case, $1220 \text{ \AA} < \lambda_{\text{rest}} < 2000 \text{ \AA}$) and the “blue side,” which we wish to reconstruct (in this case, $1170 \text{ \AA} < \lambda_{\text{rest}} < 1260 \text{ \AA}$). We use samples of low- z quasars as training sets to learn the shape of the intrinsic quasar Ly α emission. In practice, we use quasars selected from the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS) and the SDSS-IV Extended BOSS (eBOSS) DR14 (Dawson et al. 2013, 2016; Pâris et al. 2018). We select quasars with $S/N > 10$ at $\lambda = 1290 \pm 2.5 \text{ \AA}$ and redshifts $2.25 < z < 3.5$ set by the visibility of the red and blue sides in the eBOSS spectral range that were not flagged as being BAL quasars. We added further quality checks to exclude quasars with missing data and identified further BALs by requiring that a smoothly fit continuum does not drop below 0.7 in the range $1290 \text{ \AA} < \lambda < 1570 \text{ \AA}$. Finally, we performed a visual inspection of the remaining quasars to exclude objects with strong proximate hydrogen absorption, which precludes the recovery of the intrinsic Ly α emission line. The final sample consists of 14,029 quasars.

We fit both the red and blue sides of each quasar spectrum with a slow-varying spline function to which the PCA will be applied. This step enables the PCA to be less biased by random noise in the spectra. The spline is fit automatically following the procedure of Young et al. (1979) and Carswell et al. (1982) as implemented by Dall’Aglio et al. (2008) and refined by Bosman et al. (2021b). We then randomly divide the sample into training and testing samples of equal size. We retain 15 and 10 components for the red- and blue-side spectra, respectively. The projection between the two sides is obtained by dividing the weight matrices of the PCA components on both sides (e.g., Pâris et al. 2011).

We evaluate the accuracy of the PCA prediction using the testing sample. The differences between the PCA predictions and the automatically fit splines at each rest-frame wavelength are encoded into a covariance matrix. Over the entire range of the blue side, the PCA prediction recovers the true continua

¹⁶ <http://xqr30.inaf.it/>

Table 1
Quasar Sample of This Study

Subsample (1)	Name (2)	Redshift (3)	Method (4)	Redshift Ref. (5)	M_{1450} (6)	M_{1450} (NIR) Ref. (7)	Res. [km s ⁻¹] (8)
Main	VDES J0224–4711	6.5222	[C II]	Wang et al. (2021)	−26.67	Wang et al. (2021)	26
Main	PSO J217–16	6.1498	[C II]	Decarli et al. (2018)	−26.93	Bañados et al. (2016)	26
Main	PSO J323+12	6.5872	[C II]	Venemans et al. (2020)	−27.08	Wang et al. (2021)	26
Main	PSO J359–06	6.1719	[C II]	Venemans et al. (2020)	−26.62	Schindler et al. (2020)	26
Extended	CFHQS J1509–1749	6.1225	[C II]	Decarli et al. (2018)	−26.56	Decarli et al. (2018)	26
Extended	PSO J036+03	6.5405	[C II]	Venemans et al. (2020)	−27.26	Wang et al. (2021)	26
Extended	SDSS J0927+2001	5.7722	CO	Wang et al. (2010)	−26.76	Bañados et al. (2016)	22
Extended	SDSS J1306+0356	6.0330	[C II]	Venemans et al. (2020)	−26.70	Schindler et al. (2020)	22
Extended	ULAS J1319+0950	6.1347	[C II]	Venemans et al. (2020)	−26.80	Schindler et al. (2020)	22
Extended	SDSS J0100+2802	6.3269	[C II]	Venemans et al. (2020)	−29.02	Schindler et al. (2020)	26

within $+8.0\%/ -8.2\%$ ($+19.8\%/ -16.4\%$) at 1σ (2σ). The reconstructions have a slight wavelength-dependent bias of $+0.7\%$, on average, which we correct for. Figure 2 shows the mean bias and $\pm 1\sigma$ and 2σ bounds of the PCA reconstructions performed on the testing sample. Note that for each individual quasar, there may be an error of $\sim 10\%$ in the continuum fit (Figure 1); however, the bias of the whole sample should be $\approx 10\%/\sqrt{10} \approx 3\%$. We will study such systematics in Section 4.2.

2.3. Ionizing Spectra

To calculate the baseline $\tilde{\tau}$ necessary to recover the density field, we need to know the ionizing spectrum for each quasar. However, the ionizing part of the quasar spectra is not directly observable due to the large Gunn–Peterson optical depth at $z > 4$. Therefore, we have to use the quasar templates developed at lower redshifts to convert the UV magnitude to ionizing flux. We calculate the ionizing photon rate using Equation (9) in Runnoe et al. (2012),

$$\log_{10} L_{\text{iso}} = 4.74 + 0.91 \log_{10}(1450 L_{1450}), \quad (5)$$

to convert M_{1450} to the isotropic bolometric luminosity. We then use the average breaking power-law ($L_\nu \propto \nu^{\alpha_\nu}$) spectral shape measured by Lusso et al. (2015), where

$$\alpha_\nu = \begin{cases} -1.70 & \lambda < 912 \text{Å} \\ -0.61 & \lambda > 912 \text{Å} \end{cases}, \quad (6)$$

to obtain the ionizing photon rate. For each quasar, we use this spectral shape with the corresponding ionizing photon rate to calculate the baseline model and obtain $\tilde{\tau}$. With this baseline $\tilde{\tau}$ and the observed τ described in the last subsection, the density field is obtained.

3. Results

3.1. Recovered Density

In the density recovery process, we consider the error in τ from continuum fitting and the error in $\tilde{\tau}$ from quasar redshift measurement. We use a correlation matrix to calculate the uncertainty in the continuum fitting (see Section 2.2) and

assume that the error in redshift obeys a Gaussian distribution with $\sigma = \Delta z = 0.002$, which corresponds to the typical systematic uncertainty for redshifts determined from [C II] or CO lines (see Section 2.1). We use Monte Carlo sampling to propagate these errors. For each quasar, we repeat the process 1000 times, every time drawing a random continuum and redshift to obtain $\tilde{\tau}$ and τ to compute density.

In Figures 3 and 4, we show 100 out of 1000 realizations of the recovered density field of each quasar with black lines. The spikes are regions where Ly α absorption saturates. In Figure 3, we show the observed flux in orange. The blue lines are 100 realizations of the continuum fittings. In Figure 4, we show the 100 realizations of the observed optical depth (τ) in orange and the 100 realizations of the modeled optical depth ($\tilde{\tau}$) in red. The uncertainty in the observed optical depth (orange lines) is due to continuum fitting, while the uncertainty in the modeled optical depth (red lines) is due to quasar redshift.

Certain regions have large uncertainties, especially underdense regions and regions very close to the quasar ($\lesssim 1$ pMpc). For the former, the uncertainty is mostly due to continuum fitting. The relative error in the observed optical depth is

$$\frac{\Delta\tau}{\tau} = \frac{\ln(C + \Delta C) - \ln C}{\ln C - \ln F}, \quad (7)$$

where C is the continuum, and F is the observed flux. Since an underdense region corresponds to larger flux, a given error in the quasar continuum results in a larger error in the recovered density in an underdense region as compared to an overdense region. Regions closer to the quasar also have relatively large uncertainties. This is because the baseline $\tilde{\tau} \propto d^2$, and the relative uncertainty due to the error in the quasar redshift Δz is

$$\frac{\Delta\tilde{\tau}}{\tilde{\tau}} \approx \frac{2\Delta d_z}{d},$$

where $\Delta d_z \approx c\Delta z/(1+z)/H \approx 0.1$ pMpc. The relative uncertainty is thus larger at a smaller distance.

One other thing to note concerns smoothing. Because of the instrumental broadening (FWHM ≈ 26 km s⁻¹, equivalent to 40 pkpc) of the sight lines, the recovered density field has extra smoothing in addition to the intrinsic thermal broadening. Although we may not say too much about structures smaller

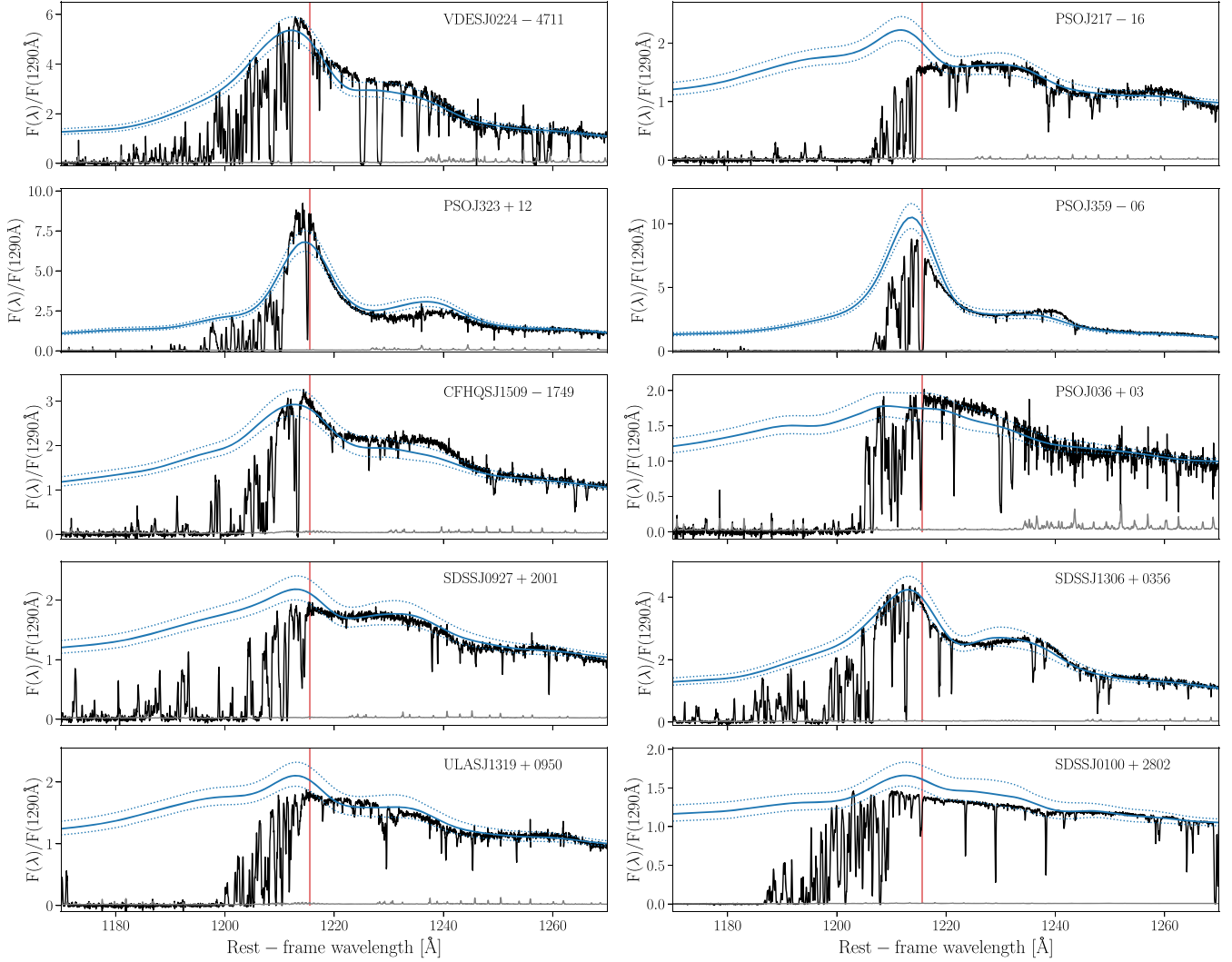


Figure 1. Rest-frame spectra for all quasars in this study. In each panel, the black line is the observed flux normalized to unity at $\lambda_{\text{rest}} = 1290\text{\AA}$, and the gray line is the noise. The solid blue line is the continuum fit, and the dotted blue lines show the 1σ uncertainty in continuum fitting. The vertical red line shows the position of $1215.6\text{\AA}$, the wavelength of Ly α .

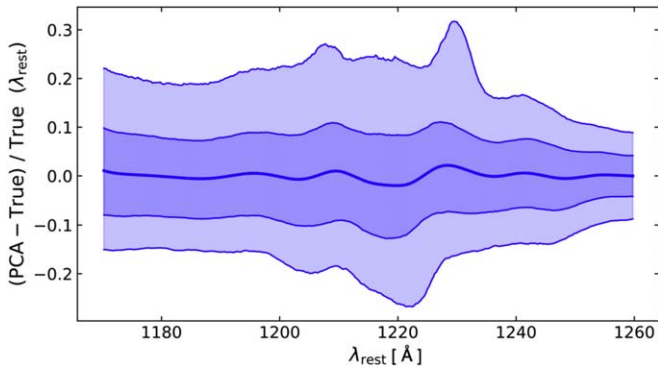


Figure 2. Performance of the PCA reconstruction on a testing sample of 7015 eBOSS quasars at $2.25 < z < 3.5$. The shaded regions show the envelope containing 68% and 95% of the deviations in the testing sample at each wavelength.

than 40 pkpc, larger-scale ($\gtrsim 200$ pkpc) structures should be robust. From Figures 3 or 4, we can see a large variation in the large-scale fields around quasars. It is typical to see an overdensity region of size $\lesssim 0.5$ pMpc, like the one at

2.0–2.5 pMpc in VDES J0224–4711. According to the CROC simulation (Chen & Gnedin 2021a), structures like this usually have $N_{\text{HI}} \sim 10^{15}\text{ cm}^{-2}$ and a neutral hydrogen fraction of 10^{-5} and most likely correspond to filaments of overdensity ~ 10 –100. There are also large voids of size ~ 1 pMpc, like the regions at 1–2 pMpc in CFHQS J1509–1749 and 2–3 pMpc in SDSS J1306+0356.

3.2. CDF of Density

The probability distribution function of the density is a simple and powerful tool to study the large-scale structure. In this section, we measure the density CDF for our XQR-30 sample. The CDF can help us constrain the cosmological parameters and quasar properties, which we will discuss in the next section.

In Figure 5, we show the reconstructed density CDF in the distance range 1.5–3.0 pMpc (corresponding to $1211\text{\AA} \gtrsim \lambda_{\text{rest}} \gtrsim 1207\text{\AA}$) for each quasar. We only choose such a narrow distance range because closer to the quasar, the density error is large due to the uncertainty on the quasar redshift, while farther from the quasar, the ionization timescale is larger ($t_{\text{ion}} \sim 0.1$

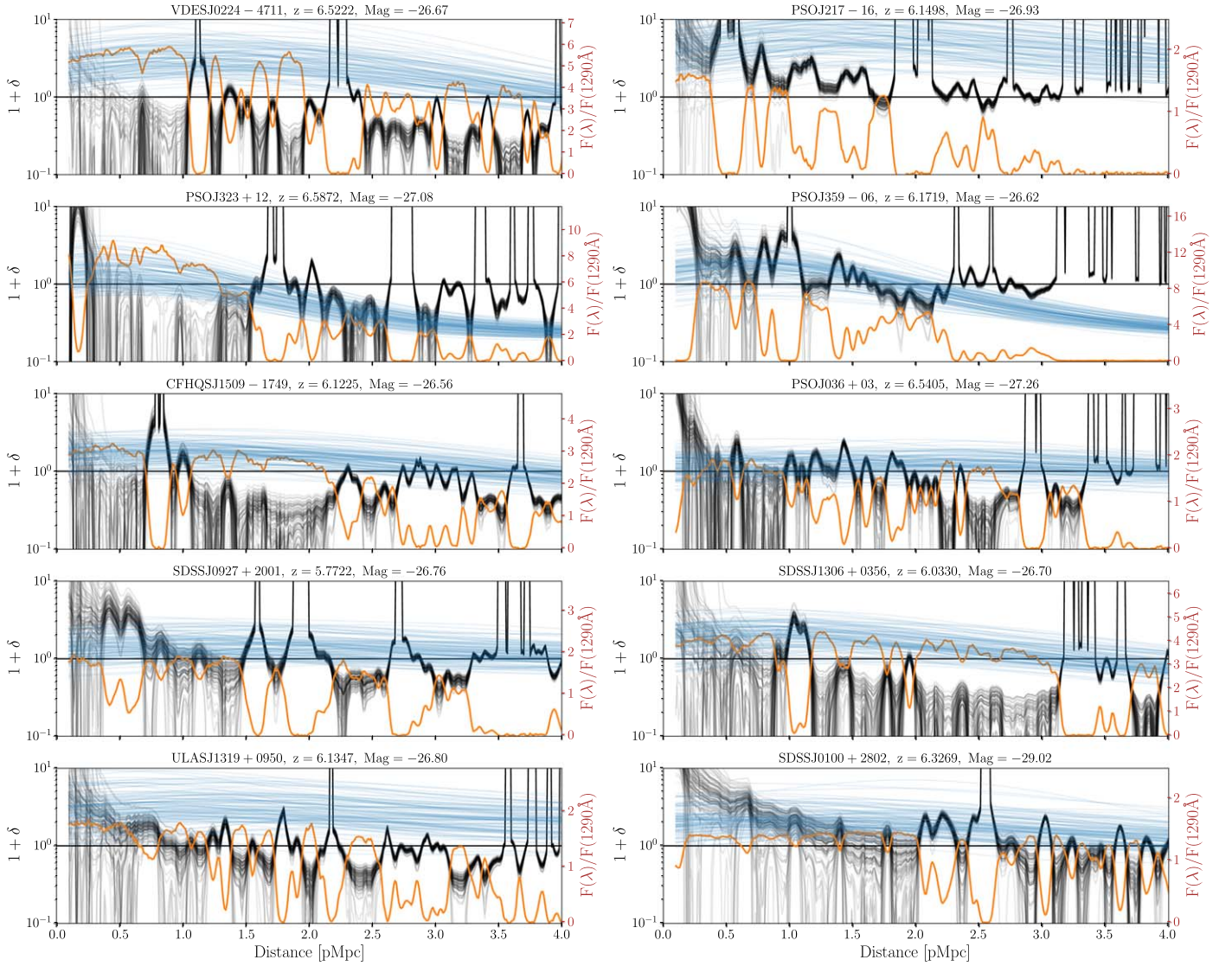


Figure 3. Each panel shows the density recovery result for each quasar. The orange line is the flux, while the blue lines are 100 draws of continuum fitting. The black lines are the corresponding recovered density field.

Myr), and it is possible that the gas has not yet fully reached ionization equilibrium. Therefore, this is a very conservative choice to ensure the validity and accuracy of the density recovery procedure from Equation (3).

In Figure 5, we show the CDF for each quasar using different colors. Note that because instrumental broadening affects the shape of the CDF (see Section 4.2), we apply additional Gaussian smoothings to bring the effective spectral resolution to $\text{FWHM} = 30 \text{ km s}^{-1}$ for all quasars. Also note that when recovering the density with Equation (3), some random draws of continuum may have flux lower than the observed flux at some wavelength, resulting in negative optical depth (Sections 2.2 and 3.1). In such cases, we assign zero density to those pixels. Therefore, the CDFs in Figure 5 do not always have $\text{CDF} \approx 0$ at the very low density end ($1 + \delta = 0.1$). On the other hand, some pixels are saturated because of the spectral noise, and we assign an infinitely large value to such pixels. This is why there are plateaus at the high-density end in Figure 5. From Figure 5, we can see that at the cosmic mean density, the majority of the CDFs vary between 0.4 and 0.8. The only exception is SDSS J1306+0356, which

is significantly more underdense than the others. This is mainly due to the voids at 2–3 pMpc from this quasar (Figure 3). The faint bands show the 68% uncertainty due to the uncertainty in the quasar continuum and redshift. Again, at the low-density end, the uncertainty is large for all quasars because of the relatively large uncertainty in the continuum fitting. The variation from sight line to sight line is also large, as expected, because we have chosen a small region, resulting in large sample variance.

4. Discussion

The recovered density fields at $z \sim 6$ can help us understand both cosmology and quasar physics. In this section, we explore how to constrain cosmological parameters and quasar environments and their lifetime.

4.1. Comparing the CDF with Simulations

Simulations help us understand the complex information encoded in the spectra. In this section, we compare the

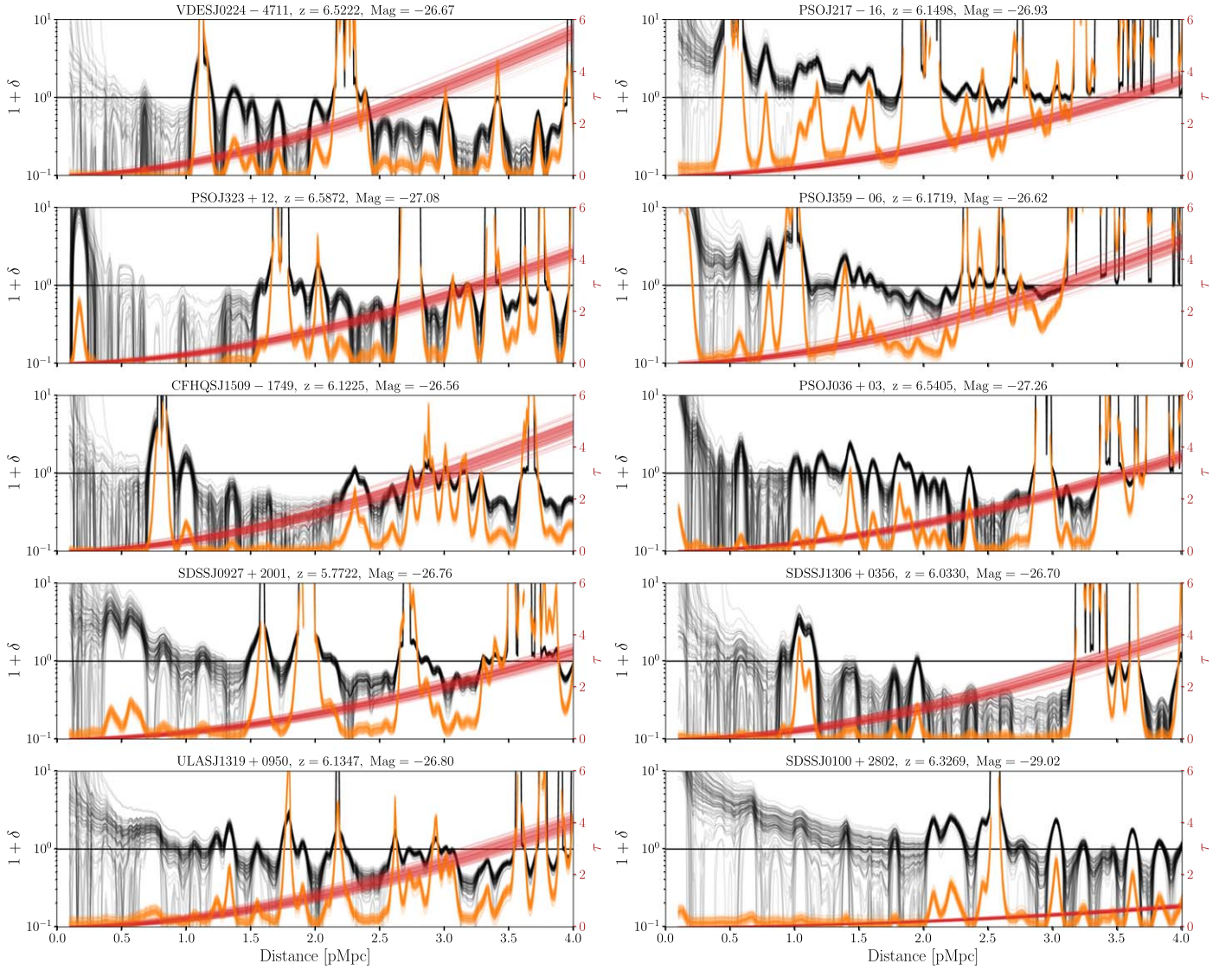


Figure 4. Similar to Figure 3, but here the orange lines are optical depths calculated from the 100 random draws of continuum fitting. The error of the observed optical depth is thus due to the uncertainty in continuum fitting. The red lines show the modeled optical depth if the universe is uniform. The error of the modeled optical depth is from the uncertainty of the exact quasar redshift. The black lines are the corresponding recovered density field, same as in Figure 3.

recovered density CDF from the observational sample to the simulated one from the CROC simulation.

It is reasonable to assume that quasar sight lines sample similar environments at 1.5–3 pMpc (~ 10 –20 cMpc), a region far away from the halo where the halo-mass bias is small. In this case, using the mean of the CDFs from these 10 quasars can greatly reduce the statistical uncertainty. In Figure 6, we show the mean CDF of the 10 quasar sight lines from our observational sample in blue. To compute the uncertainty, we consider both the uncertainty from the measurement σ_{obs} and that from the sample variance σ_{jk} . To compute σ_{obs} , we calculate the mean CDF 1000 times, each time using one realization for each of the 10 quasars. We quote the 68% range as σ_{obs} . We use the jackknife estimator to calculate the sample variance. Specifically, we exclude one quasar at a time to calculate the CDF of nine quasars. Then we quote the $\sqrt{9} \times$ the standard deviation of these 10 CDFs as the error from sample variance σ_{jk} . This error also captures the error from the continuum bias of one quasar. In Figure 6, we present the total error $\sqrt{\sigma_{\text{obs}}^2 + \sigma_{\text{jk}}^2}$ as the blue band.

As a model for the mean CDF, we use the 6001 sight lines drawn from one of the CROC simulations. Here we briefly describe the simulation and refer interested readers to Chen & Gnedin (2021b) for more details. The CROC simulation we used in this study is run in a $40 h^{-1}$ Mpc box using the adaptive mesh refinement code ART (Kravtsov 1999; Kravtsov et al. 2002; Rudd et al. 2008). The initial conditions are sampled on a uniform grid size with a cell size of $40 h^{-1}$ ckpc, and the peak spatial resolution during the simulation is maintained at 100 pc in proper units. The simulation includes physical processes like radiation field-dependent gas heating and cooling, star formation, and stellar feedback. The stars are the main driving source of reionization, and the radiation transfer is fully coupled with gas dynamics using the OTVET algorithm (Gnedin & Abel 2001). There are no individual quasars as ionizing sources in the simulation box, although the background radiation from the population of quasars is included. To model the quasar proximity-zone spectra, we randomly draw sight lines centered on the 63 most massive halos with total mass (dark matter + baryon) $M_h > 1.8 \times 10^{11} M_\odot$ at $z = 6.1$.

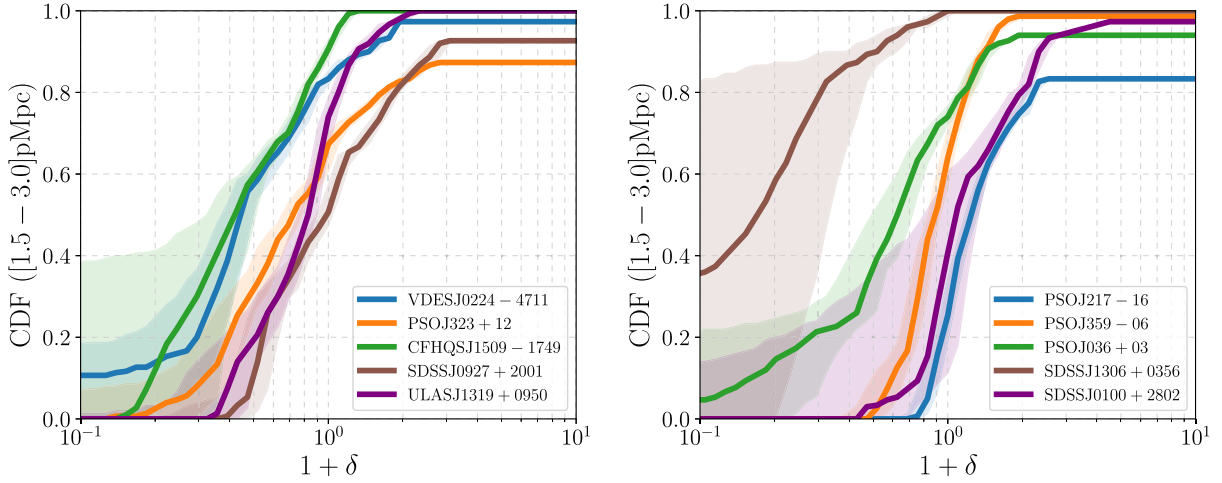


Figure 5. The CDFs of density within 1.5–3.0 pMpc for each quasar in the sample. Bands show the 68% uncertainty due to the uncertainties in the continuum fitting and quasar redshift.

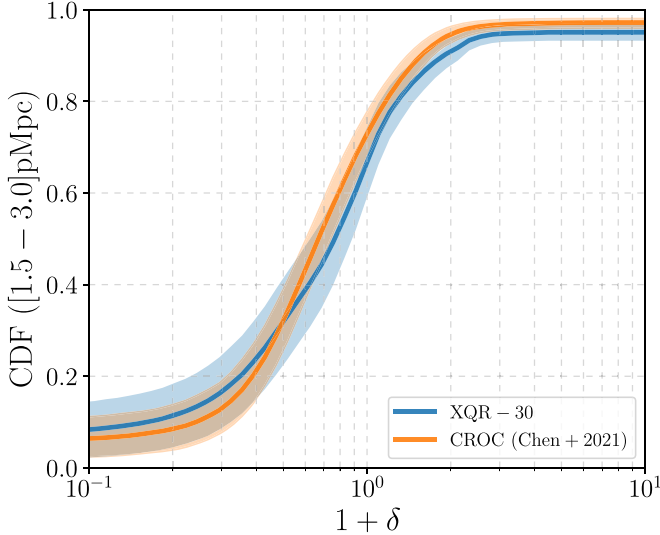


Figure 6. Mean density CDF from the 10 quasar sight lines in the XQR-30 sample (blue line). The orange line is calculated from the synthetic spectra from the $40 h^{-1}$ Mpc box CROC simulation. The blue band shows the 68% uncertainty due to the errors in the continuum fitting and quasar redshift, as well as the sample variance calculated by the jackknife estimator. The orange band shows the 68% uncertainty due to the uncertainty in the continuum fitting, quasar redshift, and spectral noise, as well as the sample variance of 10 sight lines calculated from the total set of 6001 sight lines from the $40 h^{-1}$ Mpc CROC simulation.

(Chen & Gnedin 2021b) and then run a time-dependent 1D RT code with quasar spectra.

To conduct an apples-to-apples comparison, we rerun all sight lines with the properties of each observed quasar in our sample, folding in instrumental factors like spectral resolution and noise. Specifically, we postprocess the sight lines with a spectral index of $\alpha_\nu = -1.7$ and a grid of quasar ionizing luminosities $\dot{N}_{\text{tot}} = 1 \times 10^{57}, 1.5 \times 10^{57}, 2 \times 10^{57}, 2.5 \times 10^{57}, 3 \times 10^{57}, 3.5 \times 10^{57}, 4 \times 10^{57},$ and $1 \times 10^{58} \text{ s}^{-1}$. We postprocess all 6001×8 sight lines for a fiducial quasar lifetime of 30 Myr and calculate the transmitted spectra. For each observed quasar, we pick the luminosity in the grids that is closest to the observed luminosity. We then multiply it with a random continuum drawn from the covariance matrix (Section 2.2), obtaining the modeled flux. We also add Gaussian noise to

each pixel according to the average noise in the proximity zone of each quasar. We further use a Gaussian kernel of $\text{FWHM} = 30 \text{ km s}^{-1}$ to smooth the synthetic spectra so that they have the same spectral resolution as the observed ones. We then recover the density using the exact same method as the observed one. Finally, we calculate the mean CDF of the 10 sight lines and repeat the process 1000 times to calculate the uncertainty. We show this simulated CDF as the orange line in Figure 6 and the 68% uncertainty as the orange band. This uncertainty includes quasar continuum, quasar redshift, spectral noise, and sample variance from a $40 h^{-1}$ Mpc box.

Comparing the blue and orange lines in Figure 6, we find that they agree very well. This agreement indicates that our assumptions (optically thin proximity zones and ionization equilibrium of the IGM) that go into the density reconstruction are correct.

4.2. Factors that Impact the Observed CDF

There are several factors that impact the accuracy of the recovered density, including uncertainties in the continuum fitting, observational noise, limited spectral resolution, and uncertainty in the quasar ionizing flux. These uncertainties affect different parts of the CDF.

Among these, we have a relatively good handle on the continuum fitting and noise level. In the left panel of Figure 7, we show how the scatter in the continuum fitting and the S/N affect the CDF. All of the CDFs shown in this section are the mean using the 6001 sight lines drawn from the same simulation and run with a typical ionizing photon rate $\dot{N}_{\text{tot}} = 1.5 \times 10^{57} \text{ s}^{-1}$. The blue dotted line is for a perfect instrument, i.e., with infinite spectral resolution and S/N, and exactly known quasar continua. The red dotted line adds the realistic scatter (but no bias) in the continuum fitting (Section 2.2). Due to the scatter in the continuum fitting, in some pixels, the continuum estimate will fall below the actual observed flux. Therefore, the optical depth becomes negative, and the density cannot be recovered; we set the density to zero in such pixels with the result that the CDF does not approach zero in the limit of low δ . If the uncertainty in the continuum fitting is twice as large, the value of the CDF at zero density increases further, as shown by the orange dotted line. However, as long as there is no bias in the continuum fitting, with the

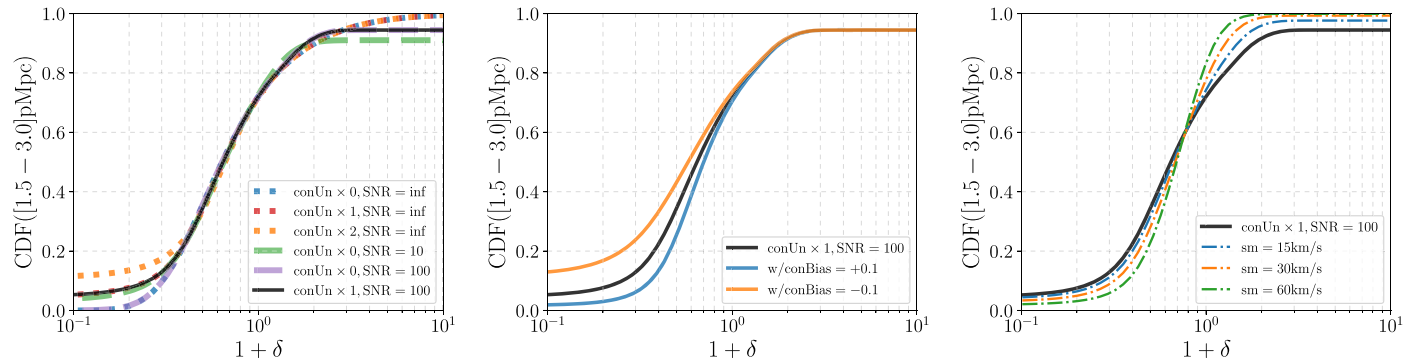


Figure 7. Left: effects of the continuum uncertainty and S/N on the CDF. The blue dotted line is assuming that we know the quasar continuum exactly and the observation is perfect. The red and orange dotted lines show how the CDF changes when the uncertainty in the quasar continuum is included with $1\times$ and $2\times$ our current estimate, respectively. The green and purple dashed lines are for S/Ns of 10 and 100. The black solid line shows the CDF when we have both the continuum uncertainty and the finite S/N to match the actual XQR-30 data. The segment between 0.3 and 0.7 is robust to these uncertainties. Middle: the blue (orange) line shows the CDF when the continuum is biased high (low) by 10%. Right: spectrum smoothing steepens the CDF. The numbers listed for the dashed-dotted lines show the equivalent Gaussian σ .

continuum underestimate always balanced by an overestimate in some other pixels, the CDF values of $\gtrsim 0.3$ are not impacted. In the same panel, we also show how changing the S/N changes the CDF. We model the spectra with noise the same way as above; i.e., in each pixel, we randomly add and subtract Gaussian noise at 10% of the continuum level to model the S/N = 10 (green dashed line). Limited S/N means there are saturated pixels where we cannot measure the true optical depth. As described earlier, for all saturated pixels, we assign an infinitely large density, and for pixels with a negative optical depth, we assign zero density. Adding a 10% level of noise results in a plateau in the CDF at high densities due to the saturation, as well as an increase at the low-density end. Increasing the S/N to 100 (purple dashed line), which is the S/N close to the actual XQR-30 sample, results in a smaller plateau at the high-density end and no significant increase at the low-density end. If both the continuum uncertainty and the S/N are at the XQR-30 level (but with infinite spectral resolution), the CDF is that traced by the black solid line, which has a nonzero value at the low-density end and a plateau at the high-density end. From this panel, we find that although the scatter in the continuum fitting and the spectral noise impact the low- and high-density ends of the CDF, the segment $0.3 < CDF < 0.7$ is robust.

While the uncertainty in the continuum fitting does not affect the CDF in the range 0.3–0.7, the bias does. In the middle panel of Figure 7, we show how an average bias of 10% in the continuum fitting affects the recovered CDF. Such a bias in the continuum impacts the lower-density end more. If, for some reason, the continuum fitting method has an average bias of 10% over the entire sample, there will be an $\sim 10\%$ error in $1 + \delta$ at $CDF = 0.5$. We have tested our continuum reconstruction technique rigorously using lower-redshift quasars and found that the mean bias is $\approx 0.7\%$ (Figure 2), well below 10%. However, the bias level of this technique for very high redshift quasars, especially a relative small sample of them, is still to be investigated. This potential bias is something to keep in mind when pushing the accuracy of the recovered CDF to the percent level.

In the right panel of Figure 7, we show how the spectral resolution affects the CDF. Thermal broadening has an effective Gaussian smoothing of $\sigma \approx 17 \text{ km s}^{-1}$ (Chen & Gnedin 2021a). If the spectral resolution is significantly smaller than this value, it becomes irrelevant. Unfortunately, all real-

life observations have limited spectral resolution, introducing some level of smoothing. Increasing the smoothing makes the CDF steeper and more “Heaviside-like.” Also note that the smoothing from the finite spectral resolution is applied directly to the flux, which is an exponential function of optical depth. This nonlinearity is the reason why the CDF recovered from the spectra with finite resolution does not have a mean density value at $\delta = 0$. Many saturated pixels can be smoothed out by a large smoothing kernel, so the saturation plateau at high densities also disappears. Extra smoothing by the instrument downgrades the constraining power of the CDF. Therefore, to make improvements in the constraining quasar properties with proximity-zone spectra, it is crucial to have at least as high a spectral resolution as in the XQR-30 sample and accurately model the broadening effect from slit width, seeing, spectrograph, etc. For X-Shooter, the uncertainty in the spectral resolution is $\approx 10\%$. According to the right panel of Figure 7, this level of uncertainty will not impact the conclusions in any significant way, as long as we use the part where $CDF = 0.3\text{--}0.7$.

The most uncertain quantity that we have to assume in the density recovery process is the ionizing radiation for each quasar. Because the ionizing part of the spectrum is heavily absorbed at $z \gtrsim 6$, we cannot directly measure it. In this study, we use the average quasar spectrum measured at $z \sim 2.4$ (Lusso et al. 2015). We note that they quoted an uncertainty of 20% in the mean ionization rate of H I, while the scatter in the ionization rate of H I between individual quasars may be a few times larger. Currently, we have not found a good way to rigorously measure the uncertainty in this quantity. Perhaps in the future, machine-learning techniques could better predict the ionizing spectra using the red part of the quasar spectra, similar to how the continua are estimated. Note that there is also an uncertainty in the observed quasar magnitude, which is usually $\Delta \text{mag} \approx 0.1$ in apparent magnitude. This translates to an uncertainty of 10% in ionizing flux for a typical $\text{mag} = 20$ quasar. Therefore, for a single quasar, the total uncertainty in the ionizing radiation may be up to 50%. Because the optical depth depends on the square root of the ionizing flux, this uncertainty translates up to 30% on recovered density on a single quasar. Increasing the sample size can decrease such uncertainty, much the same way as in averaging out uncertainty in the continuum fitting. There are at least several hundred

bright quasars at $z \sim 6$, so the total uncertainty can go down to the percent level.

4.3. Application to Cosmology

The measurement of the density field at $z \sim 6$ provides a unique way to study cosmology. Currently, apart from the cosmic microwave background (CMB), clustering measurements for matter, gas, or galaxies come from the low- or middle-redshift universe ($z \lesssim 3$), where the effect of dark energy is not completely negligible. Different existing techniques have their own unique strengths and systematics. For example, measurements based on CMB lensing put strong constraints on a combination of Ω_m and σ_8 , and, if combined with BAO measurements, they become powerful for constraining the dark energy (e.g., Wu et al. 2020). At $z \sim 6$, cosmological measurements are extremely scarce. On the other hand, at this relatively high redshift, the universe is substantially different from the $z \lesssim 3$ universe because it is completely matter-dominated (at least in the absence of early dark energy). Therefore, studying the recovered density field at $z \sim 6$ can potentially help us understand cosmology better.

Cosmological parameters like σ_8 (the amplitude of density fluctuations at the $8 h^{-1}$ cMpc scale) and n_s (the spectral index of the matter power spectrum) control the shape of the density CDF (e.g., Chen et al. 2022). Therefore, our measurement can potentially be used to constrain these cosmological parameters. In reality, however, the density CDF only mildly depends on n_s , and its sensitivity is mostly at the very low density end (Chen et al. 2022). Because this is also the place where measurement uncertainties are large, it is impossible to achieve meaningful constraints on n_s with the current data. On the other hand, the density CDF as a whole is very sensitive to σ_8 . The larger the σ_8 , the flatter the CDF.

To constrain σ_8 , we need to model the density CDF in cosmologies with different σ_8 . We do so using the method described in Chen et al. (2022). In that paper, it is shown that one can approximate the CDF of the geometric mean of the real space and redshift space density ($\sqrt{\Delta_r \Delta_z}$, the quantity we recover from the proximity zones; see Section 2) in Λ CDM cosmology with a set of self-similar simulations. The three parameters to describe the density CDF are the rms linear density fluctuation at a given smoothing scale, σ ; the spectral index of the matter power spectrum at that scale, n_s ; and a parameter describing the redshift space distortions, f . We consider the Λ CDM cosmology with different σ_8 at $z \sim 6$. At this early time, the universe is matter-dominated, so $f = 1$. The smoothing scale should be the equivalent thermal smoothing scale ($\sim 17 \text{ km s}^{-1}$), which is $\approx 0.25 \text{ cMpc}$ at $z \sim 6$. The spectral index at this scale is $n_s \approx -2$, and σ is the parameter we vary to test the effect of σ_8 . We thus calculate the density CDF from the self-similar simulations with $f = 1$, $n_s = -2$, and a range of $\sigma = 0.4\text{--}1.2$.

Because the recovered density field from observed quasars has extra smoothing due to limited spectral resolution, we mimic this effect by doing the following. For each sight line drawn from the self-similar simulation, we calculate the transmitted flux, $\tilde{F} = e^{-\tau \Delta_r \Delta_z}$. Since we do not have any way to compute the baseline model $\tilde{\tau}$ in self-similar simulations, we adopt $\tilde{\tau} = 1$, which is a typical value for XQR-30 quasars in the distance range 1.5–3 pMpc we focus on here. The result is a mock spectrum with infinite spectral resolution. We then need to smooth it with a specific scale. Because the instrumental

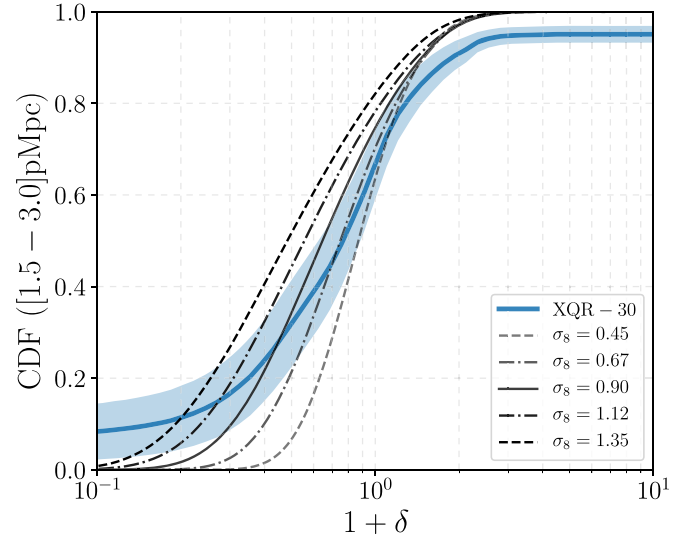


Figure 8. Density CDF (blue line, with the uncertainty shown as the blue band, same as Figure 6) measured from the XQR-30 sample, overlapped by CDFs in cosmologies with different σ_8 . The progressively darker shades represent larger σ_8 .

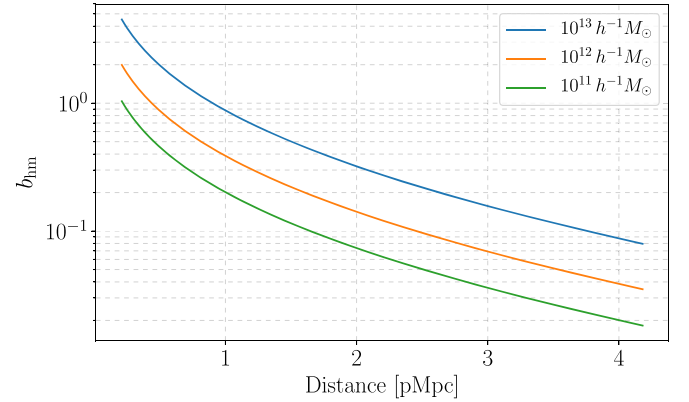


Figure 9. Halo-mass bias around halos of different masses at $z = 6$ calculated using the python toolkit Colossus (Diemer 2018).

smoothing scale ($\sigma \approx 12 \text{ km s}^{-1}$) is almost $0.75 \times$ the intrinsic thermal broadening scale ($\sigma \approx 17 \text{ km s}^{-1}$), we use a 1D Gaussian kernel of $0.75 \times$ the size as the smoothing scale corresponding to σ . Then, from this smoothed transmitted flux $\tilde{F}$, we get back the synthetic recovered density by taking $\sqrt{-\ln \tilde{F}}$. This way, we calculate the synthetic density CDFs for a grid of $\sigma = 0.4, 0.6, 0.8, 1.0$, and 1.2 . To translate these values of σ to the cosmological parameter σ_8 , there are two options. One is to analytically translate it between two spatial scales. However, it is complicated due to the involvement of both 1D and 3D smoothing. The other one, which we adopt, is to take a shortcut by calibrating them with the CROC simulation. The CROC simulation has $\sigma_8 = 0.8285$, and its CDF lies between $\sigma = 0.6$ and 0.8 , agreeing the best with $\sigma = 0.74$. Therefore, we multiply each σ by a factor of $0.8285/0.74 = 1.12$ to obtain the effective σ_8 . The five modeled CDFs with different σ_8 are shown as black lines in Figure 8.

In Figure 8, we overplot the density CDFs of different σ_8 using black lines, with larger σ represented by progressively darker shades. Note that when calculating the CDFs of different σ_8 , we do not fully model the continuum fitting uncertainty and

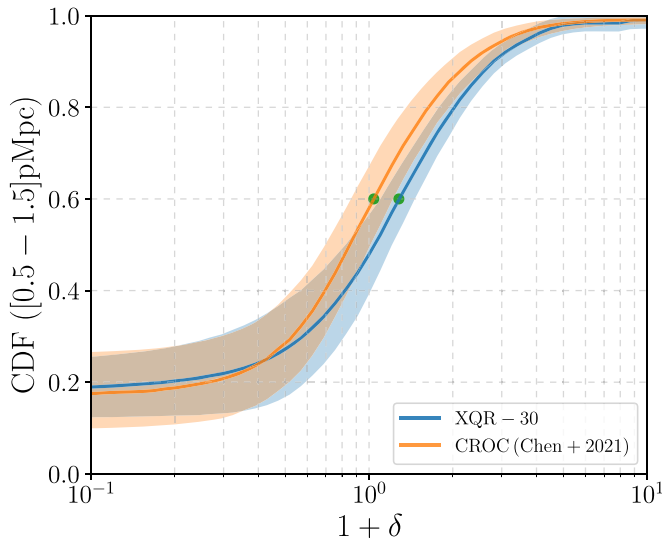


Figure 10. Same as Figure 6 but in an inner region of 0.5–1.5 pMpc, where the halo-mass bias is significant. The green dots mark locations where CDF = 0.6, which corresponds to $(1 + \delta) = 1.28$ for XQR-30 and 1.04 for CROC.

spectral noise. Therefore, we should only compare the section with $0.3 < \text{CDF} < 0.7$, which is robust against these factors (see the left panel of Figure 7). Using the value at CDF = 0.6, which is also robust against slight differences in smoothing scale, the constraint we obtain for σ_8 is 0.6 ± 0.3 , which is consistent with the concurrent cosmology measurement (Planck Collaboration et al. 2020; Porredon et al. 2021). Compared to the Planck measurement of $\sigma_8 = 0.811 \pm 0.006$, our inferred value seems to lie in a lower end, which may be because the halo bias is not completely negligible (see the next section and Figure 9). Also, as mentioned in the previous sections, there are other observational and physical factors that impact the observational CDF, like the quasar continuum fitting and the ionizing spectrum. The large uncertainty mainly comes from the current limited sample size. There are hundreds of bright quasars at $z \sim 6$, which means we can potentially reduce such uncertainty to $\approx 10\%$, rivaling current constraints from low- z BAO measurements. Developing an efficient procedure for jointly constraining them together with cosmological parameters requires substantially more research and is beyond the scope of this paper. That effort is, however, timely, as in the near future, 30 m class telescopes are expected to boost the number of available high-quality spectra by a factor of 30, paving the way to a likely breakthrough in our understanding of the quasar physics and cosmology from the proximity-zone spectra.

4.4. Constraining Properties of First Quasars

The halo-mass bias (b_{hm}) is very small at $\gtrsim 2$ pMpc (at $z \sim 6$) from a massive halo; however, it increases dramatically closer to the halo. In Figure 9, we show b_{hm} as a function of distance around halos of different masses, calculated using the python toolkit Colossus¹⁷ (Diemer 2018). Around 1 pMpc, for a halo of $10^{13} h^{-1} M_\odot$, the bias is still around unity, but for halos of $10^{12} h^{-1} M_\odot$ or lower, the bias is less than 0.4. Therefore, measuring the CDF at ~ 1 pMpc offers us a way to constrain the halo mass of the first quasars statistically.

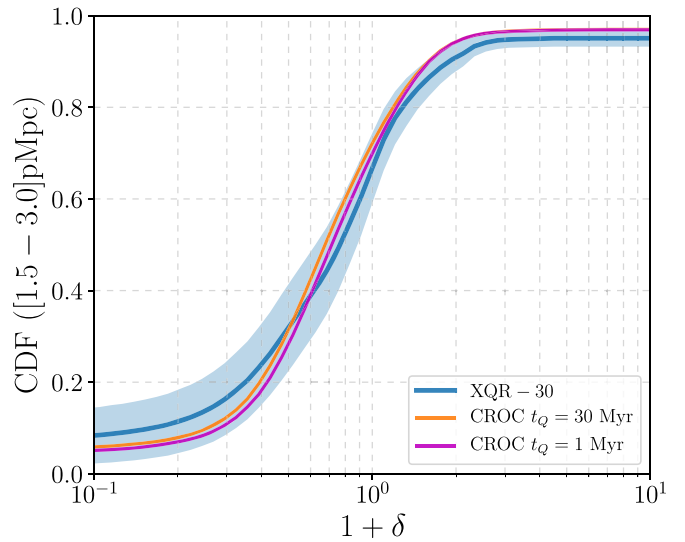


Figure 11. Density CDF in the distance range of 1.5–3.0 pMpc. The blue and orange lines are the same as in Figure 6. The magenta line is the recovered CDF from synthetic CROC spectra if the quasar lifetime is only $t_Q = 1$ Myr.

In Figure 10, we compare the CDF of the XQR-30 sample and the one from CROC in the inner region of 0.5–1.5 pMpc. Compared with Figure 6, the uncertainty in the CDF of the XQR-30 sample in this distance range is slightly larger than that in the 1.5–3 pMpc region because the distance range is smaller and the error in the quasar redshift has a larger impact at shorter distances. Also, compared with Figure 6, both curves shift to the right significantly. At CDF = 0.6, in the outer region of 1.5–3.0 pMpc, the XQR-30 sample (CROC sample) has a density $1 + \delta \approx 0.9$ (0.8), while in the inner region of 0.5–1.5 pMpc, the corresponding densities are $1 + \delta = 1.28$ for XQR-30 and $1 + \delta = 1.04$ for CROC. This bias agrees with the common assumption that bright quasars live in massive halos at $z \sim 6$. Furthermore, although this is not very statistically significant, there is a hint that the observed sample sits in a denser region, on average, than the massive CROC halos; at CDF = 0.6, the density ratio is 1.23 ± 0.17 . If we use this as the ratio of $(1 + b_{\text{hm}})_{\text{XQR-30}} / (1 + b_{\text{hm}})_{\text{CROC}}$, then, considering that the halo masses for the halos we used in the CROC simulation are between 1.8×10^{11} and $1.8 \times 10^{12} M_\odot$ with a median of $2.3 \times 10^{11} M_\odot$, this indicates a typical halo mass of an XQR-30 quasar host halo of $\log_{10}(M_h/M_\odot) = 12.5^{+0.4}_{-0.7}$. Note that this value is consistent with the dynamical mass estimated from the [C II] 158 μm velocity-width measurements (Shimasaku & Izumi 2019).

Apart from constraining the halo mass, the recovered density can also potentially be used for constraining the quasar lifetime in the range of 1–100 Myr. In Chen & Gnedin (2021a), it is shown that the quasar lifetime can bias the reconstructed density due to photoheating of He II. For a sufficiently short quasar lifetime t_Q , the ionized He III bubble has not yet propagated beyond 3 pMpc, resulting in cooler and thus slightly more neutral hydrogen gas. We model the recovered density CDF with the assumption that all of the 6001 simulated CROC spectra have $t_Q = 1$ Myr (the average quasar lifetime at $z \sim 6$ measured by Morey et al. 2021) and show it with the magenta line in Figure 11. There is a slight shift toward higher densities, and it is not statistically significant with just 10 quasar sight lines from the XQR-30 survey. However, in the future, after the 30 m class telescopes come online, we could

¹⁷ <http://www.benediktdiemer.com/code/colossus/>

potentially have more than 100 spectra with even higher spectral resolution. This improvement in data quantity and quality could help us to put constraints on the quasar lifetime in the megayear range. Also, the CDF may not be a particularly good statistic to constrain the quasar lifetime. Other statistical tools that we have not yet explored, such as the power spectrum, may be more useful in diagnosing a more subtle effect of spatial variation in thermal broadening, and this is a good topic for future work.

5. Summary

Using a sample of high-S/N, high-resolution $z \sim 6$ quasar spectra from the XQR-30 (main + extended) survey, we measure the density fields in their proximity zones out to ~ 20 cMpc for the first time. We compare the recovered density CDF with the modeled one from the CROC simulation in the outer region, where the halo-mass bias is low, and find excellent agreement. We also test different factors that may impact the CDF and find that the range between $\text{CDF} = 0.3$ and 0.7 is robust against uncertainties in the quasar continuum fitting and the spectral noise of the data.

We also explore how the recovered density CDF can constrain cosmology and quasar properties. We find that in the region $1.5\text{--}3.0$ pMpc away from the quasars, using the sample of 10 quasars, one can reach a precision of 30% in $1 + \delta$ at $\text{CDF} = 0.6$, although the uncertainty from quasar ionizing flux is an important systematic that requires further investigation. Using the CDF in that region, we can put a constraint on $\sigma_8 = 0.6 \pm 0.3$. We also investigate the CDF in different distance ranges and find that the density field close to a quasar is systematically denser than in the outer region. The level of overdensity is comparable to the halo-mass bias of massive halos at $z \sim 6$. A comparison of the CDFs in the inner region between our observational sample and the CROC simulation suggests that the typical host halo mass of $M_{1450} < -26.5$ quasars at $z \sim 6$ is $\log_{10}(M_h/M_\odot) = 12.5^{+0.4}_{-0.7}$.

We also investigate the change in the recovered density CDF due to a finite quasar lifetime. We find that the change is very small compared with the uncertainty of the existing XQR-30 data. However, we expect that the large increase in the quantity and quality of observational data from the future 30 m class telescopes will lead to meaningful constraints on the quasar lifetime. In the future, we plan to investigate other statistical measures of the recovered density field, like the power spectrum, and extract more information about reionization and the first quasars. We also expect that with the synergy of line-of-sight density measurements from quasar spectra and plan-of-sky galaxy maps from the James Webb Space Telescope, which will offer much deeper observations than currently available, we can provide significantly better large-scale density measurements to characterize the environments of the first quasars.

H.C. is thankful for the support by NASA through NASA FINESST grant NNH19ZDA005K. A.C.E. acknowledges support by NASA through NASA Hubble Fellowship grant No. HF2-51434 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. X.F. and J.Y. acknowledge support by the NSF through grant AST-1908284. F.W. acknowledges support by NASA through NASA Hubble Fellowship grant No. HF2-

51448 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. G.D.B. is supported by the NSF through grant AST-1751404. M.B. acknowledges support from PRIN MIUR project “Black hole winds and the Baryon Life Cycle of Galaxies: the stone-guest at the galaxy evolution supper,” contract No. 2017PH3WAT. A.P. acknowledges support from the ERC Advanced Grant INTERSTELLAR H2020/740120. This manuscript has been coauthored by the Fermi Research Alliance, LLC, under contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The simulations used in this work were completed at the Argonne Leadership Computing Facility and the Blue Waters supercomputer. The Argonne Leadership Computing Facility is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research is also part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

ORCID iDs

Huanqing Chen  <https://orcid.org/0000-0002-3211-9642>
 Anna-Christina Eilers  <https://orcid.org/0000-0003-2895-6218>
 Sarah E. I. Bosman  <https://orcid.org/0000-0001-8582-7012>
 Nickolay Y. Gnedin  <https://orcid.org/0000-0001-5925-4580>
 Xiaohui Fan  <https://orcid.org/0000-0003-3310-0131>
 Feige Wang  <https://orcid.org/0000-0002-7633-431X>
 Jinyi Yang  <https://orcid.org/0000-0001-5287-4242>
 Valentina D’Odorico  <https://orcid.org/0000-0003-3693-3091>
 George D. Becker  <https://orcid.org/0000-0003-2344-263X>
 Chiara Mazzucchelli  <https://orcid.org/0000-0002-5941-5214>
 Andrei Mesinger  <https://orcid.org/0000-0003-3374-1772>

References

- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, *PhRvD*, **98**, 043526
- Bañados, E., Venemans, B., Walter, F., et al. 2013, *ApJ*, **773**, 178
- Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, *ApJS*, **227**, 11
- Bolton, J. S., & Haehnelt, M. G. 2007, *MNRAS*, **374**, 493
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2021a, *arXiv:2108.03699*
- Bosman, S. E. I., Ďurovčíková, D., Davies, F. B., & Eilers, A.-C. 2021b, *MNRAS*, **503**, 2077
- Bosman, S. E. I., Kakiichi, K., Meyer, R. A., et al. 2020, *ApJ*, **896**, 49
- Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, *ApJ*, **843**, 129
- Calverley, A. P., Becker, G. D., Haehnelt, M. G., & Bolton, J. S. 2011, *MNRAS*, **412**, 2543
- Carswell, R. F., Whelan, J. A. J., Smith, M. G., Boksenberg, A., & Tytler, D. 1982, *MNRAS*, **198**, 91
- Chabanier, S., Palanque-Delabrouille, N., Yèche, C., et al. 2019, *JCAP*, **2019**, 017
- Champagne, J. B., Decarli, R., Casey, C. M., et al. 2018, *ApJ*, **867**, 153
- Chen, H. 2020, *ApJ*, **893**, 165
- Chen, H., & Gnedin, N. Y. 2021a, *ApJ*, **916**, 118
- Chen, H., & Gnedin, N. Y. 2021b, *ApJ*, **911**, 60
- Chen, H., Gnedin, N. Y., & Mansfield, P. 2022, *ApJ*, **929**, 135

- Croft, R. A. C., Weinberg, D. H., Bolte, M., et al. 2002, *ApJ*, **581**, 20
- Dall’Aglio, A., Wisotzki, L., & Worseck, G. 2008, *A&A*, **491**, 465
- Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018, *ApJ*, **864**, 143
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, **151**, 44
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- Decarli, R., Walter, F., Venemans, B. P., et al. 2018, *ApJ*, **854**, 97
- Di Matteo, T., Croft, R. A. C., Feng, Y., Waters, D., & Wilkins, S. 2017, *MNRAS*, **467**, 4243
- Diemer, B. 2018, *ApJS*, **239**, 35
- Eilers, A.-C., Davies, F. B., Hennawi, J. F., et al. 2017, *ApJ*, **840**, 24
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *ApJ*, **633**, 560
- García-Vergara, C., Rybak, M., Hodge, J., et al. 2022, *ApJ*, **927**, 65
- Gnedin, N. Y. 2014, *ApJ*, **793**, 29
- Gnedin, N. Y., & Abel, T. 2001, *NewA*, **6**, 437
- Habouzit, M., Volonteri, M., Somerville, R. S., et al. 2019, *MNRAS*, **489**, 1206
- Hawken, A. J., Aubert, M., Pisani, A., et al. 2020, *JCAP*, **2020**, 012
- Kashikawa, N., Kitayama, T., Doi, M., et al. 2007, *ApJ*, **663**, 765
- Kim, S., Stiavelli, M., Trenti, M., et al. 2009, *ApJ*, **695**, 809
- Kitayama, T., Tajiri, Y., Umemura, M., Susa, H., & Ikeuchi, S. 2000, *MNRAS*, **315**, L1
- Kravtsov, A. V. 1999, PhD thesis, New Mexico State University
- Kravtsov, A. V., Klypin, A., & Hoffman, Y. 2002, *ApJ*, **571**, 563
- Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, *MNRAS*, **449**, 4204
- McDonald, P., Seljak, U., Burles, S., et al. 2006, *ApJS*, **163**, 80
- McGreer, I. D., Fan, X., Strauss, M. A., et al. 2014, *AJ*, **148**, 73
- Mesinger, A., Haiman, Z., & Cen, R. 2004, *ApJ*, **613**, 23
- Morey, K. A., Eilers, A.-C., Davies, F. B., Hennawi, J. F., & Simcoe, R. A. 2021, *ApJ*, **921**, 88
- Narayanan, D., Li, Y., Cox, T. J., et al. 2008, *ApJS*, **174**, 13
- Ota, K., Venemans, B. P., Taniguchi, Y., et al. 2018, *ApJ*, **856**, 109
- Overzier, R. A., Guo, Q., Kauffmann, G., et al. 2009, *MNRAS*, **394**, 577
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, *A&A*, **613**, A51
- Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, *A&A*, **530**, A50
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, **473**, 4077
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Porredon, A., Crocce, M., Fosalba, P., et al. 2021, *PhRvD*, **103**, 043503
- Rauch, M. 1998, *ARA&A*, **36**, 267
- Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, *ApJ*, **672**, 19
- Runnoe, J. C., Brotherton, M. S., & Shang, Z. 2012, *MNRAS*, **422**, 478
- Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, *MNRAS*, **445**, 1482
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Schindler, J.-T., Farina, E. P., Bañados, E., et al. 2020, *ApJ*, **905**, 51
- Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, *AJ*, **133**, 2222
- Shen, Y., Strauss, M. A., Ross, N. P., et al. 2009, *ApJ*, **697**, 1656
- Shimasaku, K., & Izumi, T. 2019, *ApJL*, **872**, L29
- Simpson, C., Mortlock, D., Warren, S., et al. 2014, *MNRAS*, **442**, 3454
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Natur*, **435**, 629
- Venemans, B. P., Walter, F., Neeleman, M., et al. 2020, *ApJ*, **904**, 130
- Vernet, J., Dekker, H., D’Odorico, S., et al. 2011, *A&A*, **536**, A105
- Wang, F., Fan, X., Yang, J., et al. 2021, *ApJ*, **908**, 53
- Wang, R., Carilli, C. L., Neri, R., et al. 2010, *ApJ*, **714**, 699
- Weinberg, D. H., Davé, R., Katz, N., & Kollmeier, J. A. 2003, in AIP Conf. Ser., 666, The Emergence of Cosmic Structure, ed. S. H. Holt & C. S. Reynolds (Melville, NY: AIP), 157
- Wu, W. L. K., Motloch, P., Hu, W., & Raveri, M. 2020, *PhRvD*, **102**, 023510
- Young, P. J., Sargent, W. L. W., Boksenberg, A., Carswell, R. F., & Whelan, J. A. J. 1979, *ApJ*, **229**, 891
- Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2021, *ApJ*, **923**, 223