Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **531,** 3069–3087 (2024) https://doi.org/10.1093/mnras/stae1293
Advance Access publication 2024 May 30

Measurements of the $z > 5$ Lyman-α forest flux autocorrelation functions from the extended XQR-30 data set

Molly Wolfson [1★] Joseph F. Hennawi [1,2] Sarah E. I. Bosman [3,4] Frederick B. Davies [3] Zarija Lukić [5] George D. Becker [6] Huanqing Chen [7] Guido Cupani [8,9] Valentina D'Odorico [8,9,10] Anna-Christina Eilers [11] Martin G. Haehnelt [12] Laura C. Keating [13] Girish Kulkarni [14] Samuel Lai [15] Andrei Mesinger [10] Fabian Walter [3] and Yongda Zhu [6]

[1] *Department of Physics, University of California, Santa Barbara, CA 93106, USA*
[2] *Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands*
[3] *Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[4] *Institute for Theoretical Physics, Heidelberg University, Philosophenweg 12, D-69120 Heidelberg, Germany*
[5] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA*
[6] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[7] *Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St George St, Toronto, ON M5S 3H8, Canada*
[8] *INAF – Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34143 Trieste, Italy*
[9] *IFPU-Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy*
[10] *Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[11] *MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[12] *Kavli Institute for Cosmology and Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK*
[13] *Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK*
[14] *Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[15] *Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia*

Accepted 2024 May 3. Received 2024 April 30; in original form 2023 September 6

ABSTRACT

We present the first observational measurements of the Lyman-α (Ly α) forest flux autocorrelation functions in ten redshift bins from $5.1 \leq z \leq 6.0$. We use a sample of 35 quasar sightlines at $z > 5.7$ from the extended XQR-30 data set; these data have signal-to-noise ratios of >20 per spectral pixel. We carefully account for systematic errors in continuum reconstruction, instrumentation, and contamination by damped Ly α systems. With these measurements, we introduce software tools to generate autocorrelation function measurements from any simulation. Our measurements of the smallest bin of the autocorrelation function increase with redshift when normalizing by the mean flux, $\langle F \rangle$. This increase may come from decreasing $\langle F \rangle$ or increasing mean free path of hydrogen-ionizing photons, $\lambda_{\rm mfp}$. Recent work has shown that the autocorrelation function from simulations at $z > 5$ is sensitive to $\lambda_{\rm mfp}$, a quantity that contains vital information on the ending of reionization. For an initial comparison, we show our autocorrelation measurements with simulation models for recently measured $\lambda_{\rm mfp}$ values and find good agreements. Further work in modelling and understanding the covariance matrices of the data is necessary to get robust measurements of $\lambda_{\rm mfp}$ from this data.

Key words: methods: data analysis – intergalactic medium – quasars: absorption lines – dark ages, reionization, first stars.

1 INTRODUCTION

The reionization of the neutral hydrogen in the intergalactic medium (IGM) is one of the major phase changes in our Universe's history. Understanding the timing of this process has been the focus of many recent studies. Current Planck constraints put the midpoint of reionization at $z_{\rm re} = 7.7 \pm 0.7$ (Planck Collaboration VI 2020) with mounting evidence that it was not completed until after $z \leq 6$ (Fan et al. 2006; Becker et al. 2015, 2018; Bosman et al. 2018; Eilers, Davies & Hennawi 2018; Boera et al. 2019; Jung et al. 2020; Kashino et al. 2020; Yang et al. 2020; Morales et al. 2021; Bosman et al. 2022).

Before the end of reionization, the mean free path of hydrogen-ionizing photons ($\lambda_{\rm mfp}$) is expected to be short due to the significant neutral hydrogen remaining in the IGM which will absorb these photons close to their sources. In some models, as reionization ends $\lambda_{\rm mfp}$ will rapidly increase due to the overlap of initially isolated ionized bubbles and the photoevaporation of dense photon sinks (Gnedin 2000; Shapiro, Iliev & Raga 2004; Furlanetto & Oh 2005; Gnedin & Fan 2006; Wyithe, Bolton & Haehnelt 2008; Sobacchi & Mesinger 2014; Park et al. 2016; Kulkarni et al. 2019; Keating et al.

★ E-mail: mawolfson@ucsb.edu

2020a, b; Nasir & D'Aloisio 2020; Cain et al. 2021; Gnedin & Madau 2022). Thus detecting an increase in $\lambda_{\rm mfp}$ will provide insights into the end of reionization.

Becker et al. (2021) reported the first direct measurement of $\lambda_{\rm mfp}$ at $z \sim 6$ from stacked quasar spectra. Zhu et al. (2023) updated this measurement and added two additional redshift bins at $z = 5.31$ and $z = 5.65$. They found that $\lambda_{\rm mfp} = 9.33^{+2.06}_{-1.80}$, $5.40^{+1.47}_{-1.40}$, $3.31^{+2.74}_{-1.34}$, and $0.81^{+0.73}_{-0.48}$ pMpc at $z = 5.08$, 5.31, 5.65, and 5.93, respectively. Becker et al. (2021) and Zhu et al. (2023) expanded on previous measurements of $\lambda_{\rm mfp}$ at $z \leq 5.1$ (Prochaska, Worseck & O'Meara 2009; Fumagalli et al. 2013; O'Meara et al. 2013; Worseck et al. 2014). The Zhu et al. (2023) measurement has $\lambda_{\rm mfp}$ rapidly increasing between $z = 6$ and $z = 5.1$, potentially signalling the end of reionization. The values at $z \geq 5.3$ are significantly smaller than extrapolations from previous lower z measurements (Worseck et al. 2014) based on a fully ionized IGM. In addition, the value at $z \sim 6$ may cause tension with measurements of the ionizing output from galaxies (Cain et al. 2021; Davies et al. 2021).

Alternative methods to constrain $\lambda_{\rm mfp}$ are needed to check the measurements discussed above and to constrain the timing of reionization in finer redshift bins. One such method from Bosman (2021) used lower limits on individual free paths (the distance ionizing radiation travels from an individual source) towards high-z sources to place a 2σ limit of $\lambda_{\rm mfp} > 0.31$ proper Mpc at $z = 6.0$. This Bosman (2021) method is similar to other measurements using individual free paths (Songaila & Cowie 2010; Rudie et al. 2013; Romano et al. 2019). Additionally, Gaikwad et al. (2023) constrained $\lambda_{\rm mfp}$ for $4.9 < z < 6.0$ with $\Delta z = 0.1$ by comparing the observed probability distribution function of the Lyα optical depth to predictions from simulations with a fluctuating ultraviolet background (UVB) driven by a short $\lambda_{\rm mfp}$. The measurement of $\lambda_{\rm mfp}$ at $z < 5.1$ in Gaikwad et al. (2023) shows a good agreement with the measurements from Worseck et al. (2014) and Becker et al. (2021). At $z = 6.0$ Gaikwad et al. (2023) measured $\lambda_{\rm mfp} = 8.318^{+7.531}_{-4.052}$ comoving Mpc (cMpc) h^{-1}, which agrees with the Zhu et al. (2023) measurement at the 1.2σ level and also falls above the lower limit found by Bosman (2021).

The level of fluctuations in the UVB are set by the distribution of ionizing photon sources and $\lambda_{\rm mfp}$. For large values of $\lambda_{\rm mfp}$, photons travel further from their sources and effectively create a more uniform UVB (Mesinger & Furlanetto 2009). Alternatively, small values of $\lambda_{\rm mfp}$ lead to greater fluctuations in the UVB, causing some regions to have very large $\Gamma_{\rm H\,I}$ values. These fluctuations then imprint themselves on the Lyα forest flux transmission in high-z quasar spectra via the Lyα opacity, $\tau_{{\rm Ly}\alpha}$ where $\tau_{{\rm Ly}\alpha} = n_{\rm HI}\sigma_{{\rm Ly}\alpha} \propto 1/\Gamma_{\rm HI} \propto 1/\lambda^{\alpha}_{\rm mfp}$ where $3/2 < \alpha < 2$ (see e.g. Rauch 1998; Haardt & Madau 2012). Many previous studies have investigated the effect of large-scale variations in the UVB on the structure of the Lyα forest (Zuo 1992a, b; Croft 2004; Meiksin & White 2004; McDonald et al. 2005; Gontcho A Gontcho, Miralda-Escudé & Busca 2014; Pontzen 2014; Pontzen et al. 2014; D'Aloisio et al. 2018; Meiksin & McQuinn 2019; Oñorbe et al. 2019). This is similar to the argument explored by Gaikwad et al. (2023) in using the probability distribution function of the Lyα optical depth to constrain $\lambda_{\rm mfp}$. The probability distribution function of the Lyα optical depth does not consider the 2-point clustering, which can be quantified through the autocorrelation function and the power spectrum, which is the Fourier transform of the autocorrelation function, of the Lyα forest flux. Beyond the effect of UVB fluctuations, the power spectrum of the Lyα forest flux contrast has been measured at high z and used to constrain the thermal state of the IGM (Boera et al. 2019; Walther et al. 2019; Gaikwad et al. 2021) as well as warm dark matter particle mass (Viel et al. 2013; Garzilli, Boyarsky & Ruchayskiy 2017; Iršič et al. 2017).

This work is specifically building on Wolfson et al. (2023b) which investigated the effect of a fluctuating UVB on small scales in Lyα forest transmission at $z \geq 5.4$. They found that the Lyα forest transmission on small scales will be boosted for small values of $\lambda_{\rm mfp}$ and that this can be quantified with the Lyα forest flux autocorrelation function. They used the autocorrelation function to recover $\lambda_{\rm mfp}$ from simulated mock data. The Lyα forest flux autocorrelation function has yet to be measured at $z \gtrsim 5.5$ for observational data. Many previous studies have measured the Lyα forest flux autocorrelation function at lower redshifts for a wide range of applications (McDonald et al. 2000; Rollinde et al. 2003; Becker, Sargent & Rauch 2004; D'Odorico et al. 2006).

In this paper, we use the XQR-30 extended data set to measure the Lyα forest flux autocorrelation function. We discuss this observational data in Section 2. The details on the data selection and measurement process with a full account of relevant errors are described in Section 3. We then discuss our resulting measurements in Section 4 and some preliminary comparisons to simulations in Section 5. We summarize our results in Section 6.

2 DATA

The quasar spectra used in this work are a subset of those presented in Bosman et al. (2022). The data reduction was performed and discussed in detail there but will be summarized again in this work for the sake of completeness. Additionally, more information on the continuum reconstructions can be found in Bosman et al. (2021).

All of the observations used in this work comes from the XQR-30 programme[1] (1103.A0817(A), D'Odorico et al. 2023), which consists of a sample of 30 very luminous quasars at $z \gtrsim 5.8$ observed with the X-Shooter instrument (Vernet et al. 2011) on ESO's Very Large Telescope. We use 24 quasars from the XQR-30 sample which do not show strong broad absorption lines (BALs) that would create issues in the modelling of the intrinsic continuum (Bischetti et al. 2022) and could also possibly contaminate the Lyα forest region. Three additional spectra (PSO J231-20, ATLAS J2211-3206, and SDSS J2310+1855) were identified as hi-BALs so we exclude regions of the spectra where there is possible strong OVI contamination (7770 Å $< \lambda_{\rm obs} <$ 7870 Å, $\lambda_{\rm obs} <$ 7280 Å, and $\lambda_{\rm obs} <$ 6700 Å, respectively). All XQR-30 spectra have signal-to-noise ratios (SNRs) larger than 20 per 10 km s^{-1} pixel measured over 1165 Å $< \lambda_{\rm rest} <$ 1170 Å (Table 1). In addition to the 24 XQR-30 quasars, we use 11 archival X-Shooter spectra that are from the extended XQR-30 sample (D'Odorico et al. 2023). These spectra have SNR $>$ 40 per 10 km s^{-1} pixel from the literature (Table 1, marked with $*$). The extended XQR-30 sample has a median effective resolving power over all 42 quasars of $R \simeq 11\,400$ and 9800 in the visible (5500 Å $< \lambda_{\rm obs} <$ 10200 Å) and infrared arm (10 200 Å $< \lambda_{\rm obs} <$ 24 800 Å) of X-Shooter, respectively (D'Odorico et al. 2023).

All quasars are reduced with the same procedure. Observations are first flat-fielded and sky-subtracted following the method of Kelson (2003). The spectra are extracted (Horne 1986) separately for the visible and infrared arms of the instrument which are then stitched together over the 10 110 Å $< \lambda_{\rm obs} <$ 10 130 Å spectral window. The infrared spectrum is re-scaled to match the observed mean flux in the optical arm. The spectrum is then interpolated over the overlap

[1] https://xqr30.inaf.it/

Table 1. The extended XQR-30 quasars included in this work. Those with a ∗ represent the extended data set quasars which did not get new spectra in the XQR-30 programme. References correspond to: discovery, redshift determination.

Quasar ID	z_{qso}	SNR pix^{-1}	Refs.
PSO J323+12	6.5872	35.9	Mazzucchelli et al. (2017), Venemans et al. (2020)
PSO J231-20	6.5869	42.3	Mazzucchelli et al. (2017), Venemans et al. (2020)
VDES J0224-4711	6.5223	24.4	Reed et al. (2017), Wang et al. (2021)
PSO J036+03∗	6.5405	61.4	Venemans et al. (2015), Venemans et al. (2020)
PSO J1212+0505	6.4386	55.8	Mazzucchelli et al. (2017), Decarli et al. (2018)
DELS J1535+1943	6.3932	22.6	Wang et al. (2019), Bosman et al. (2022)
ATLAS J2211-3206	6.3394	37.5	Chehade et al. (2018)/Farina et al. (2019), Decarli et al. (2018)
SDSS J0100+2802∗	6.3269	560.5	Wu et al. (2015), Venemans et al. (2020)
ATLAS J025-33∗	6.318	127.3	Carnall et al. (2015), Becker et al. (2019)
SDSS J1030+0524∗	6.309	69.6	Fan et al. (2001), Jiang et al. (2007)
PSO J060+24	6.192	49.7	Bañados et al. (2016), Bosman et al. (2022)
PSO J065-26	6.1871	77.9	Bañados et al. (2016), Venemans et al. (2020)
PSO J359-06	6.1722	68.8	Wang et al. (2016), Eilers et al. (2021)
PSO J217-16	6.1498	73.0	Bañados et al. (2016), Decarli et al. (2018)
ULAS J1319+0950∗	6.1347	81.7	Mortlock et al. (2009), Venemans et al. (2020)
CFHQS J1509-1749∗	6.1225	43.0	Willott et al. (2007), Decarli et al. (2018)
PSO J239-07	6.1102	56.3	Bañados et al. (2016), Eilers et al. (2021)
SDSS J0842+1218	6.0754	83.2	De Rosa et al. (2011)/Jiang et al. (2015), Venemans et al. (2020)
ATLAS J158-14	6.0685	60.3	Chehade et al. (2018), Eilers et al. (2021)
VDES J0408-5632	6.0345	86.6	Reed et al. (2017), Reed et al. (2017)
SDSS J1306+0356∗	6.033	65.3	Fan et al. (2001), Venemans et al. (2020)
ATLAS J029-36	6.021	57.1	Carnall et al. (2015), Becker et al. (2019)
SDSS J2310+1855	6.0031	113.4	Jiang et al. (2016), Wang et al. (2013)
PSO J007+04	6.0015	54.4	Jiang et al. (2015)/Bañados et al. (2014), Venemans et al. (2020)
ULAS J0148+0600∗	5.998	152.0	Jiang et al. (2015), Becker et al. (2019)
SDSS J0818+1722∗	5.997	132.1	Fan et al. (2006), Becker et al. (2019)
PSO J029-29	5.984	65.6	Bañados et al. (2016), Bañados et al. (2016)
PSO J108+08	5.9485	104.8	Bañados et al. (2016), Bañados et al. (2016)
PSO J183-12	5.917	61.8	Bañados et al. (2014), Bosman et al. (2022)
PSO J025-11	5.844	50.6	Bañados et al. (2016), Bosman et al. (2022)
PSO J242-12	5.837	22.9	Bañados et al. (2016), Bosman et al. (2022)
PSO J065+01	5.833	25.1	D'Odorico et al. (2023), Bosman et al. (2022)
SDSS J0836+0054∗	5.804	73.8	Fan et al. (2001), Bosman et al. (2022)
PSO J308-27	5.7985	53.2	Bañados et al. (2016), D'Odorico et al. (2023)
SDSS J0927+2001∗	5.7722	53.8	Fan et al. (2006), Wang et al. (2010)

window in order to minimize the risk of creating an artificial step in the spectrum between the arms to which the continuum-fitting method may be non-linearly sensitive (see discussion in Bosman et al. 2022). The reduction routines are described in more detail in Becker, Rauch & Sargent (2009). Further details are presented in D'Odorico et al. (2023).

An example spectrum from the programme is shown in Fig. 1 for PSO J029-29. The black line shows the reduced XQR-30 spectrum and the red line shows the noise vector. The intrinsic continuum reconstructed with the method described in Section 3.1 is shown by the solid blue line, while the continuum fit to the red side of the quasars emission is shown in green. The light blue lines show draws of the continuum reconstruction with the appropriate scatter from the covariance matrix of the PCA reconstruction. The sampling procedure for these draws are also discussed in Section 3.1.

3 METHODS

3.1 Continuum reconstruction

For each quasar, the continuum, $F_{\rm cont}(\lambda_{\rm rest})$, was reconstructed using Principal Component Analysis (PCA). To do this, we consider both the red side ($\lambda_{\rm rest} > 1280$ Å) and the blue side ($\lambda_{\rm rest} < 1220$ Å) of the quasar continuum with respect to the Ly α emission. At low-z, both sides of the quasar continuum are transmitted through the IGM, as the IGM is mainly ionized. Thus we can use PCA to find the optimal linear decomposition of both the red side and the blue side of the low-z quasar continuum, then construct an optimal mapping between the the linear coefficients from the two decompositions. At high-z, the red side of quasar continua will be transmitted while the blue side is absorbed by remaining neutral hydrogen in the IGM, see e.g. Fig. 1. We can thus get the PCA decomposition for the red side of the continuum then use the optimal mapping, determined from low-z quasars, to predict the blue side coefficients and thus the continuum (Francis et al. 1992; Yip et al. 2004). This method has been historically used to get the continuum for the Ly α forest in Suzuki et al. (2005) then it was further expanded, for example by: McDonald et al. (2005); Pâris et al. (2011); Davies et al. (2018b, c); Ďurovčíková et al. (2020). Previously, Bosman et al. (2021) determined the most accurate PCA method and Bosman et al. (2022) further improved this method with the log-PCA approach of Davies et al. (2018b, c).

This work uses the same reconstructions that were generated for Bosman et al. (2022) using the log-PCA approach. The PCA consists of 15 red-side components and 10 blue-side components. The training set amounted to 4597 quasars from the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and the SDSS-IV Extended BOSS (eBOSS; Dawson et al. 2016) at $2.7 < z <$

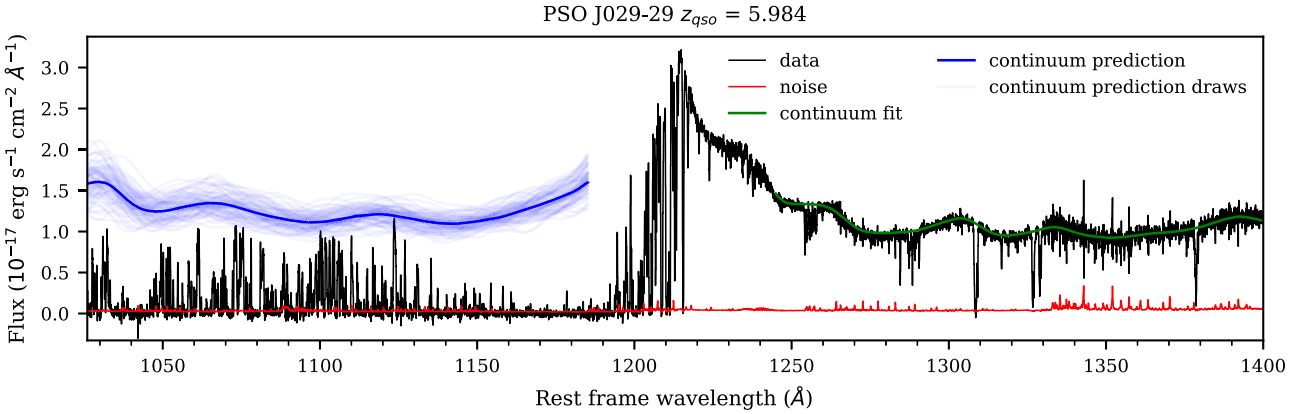

Figure 1. The X-Shooter spectrum of the Ly α transmission region for the quasar PSO J029-29 from the XQR-30 sample. The noise vector is shown in red and the PCA-reconstructed continuum is shown in blue. The light blue lines show draws of the continuum reconstruction with the appropriate scatter from the covariance matrix of the PCA reconstruction. The pixel scale is 10 km s^{-1} and the SNR of the Ly α region (reconstruction divided by uncertainty) is SNR = 50.6.

3.5 with SNR > 7. Intrinsic continua were obtained automatically using a modified version of the method of Dall'Aglio, Wisotzki & Worseck (2008), originally based on the procedures outlined in Young et al. (1979) and Carswell et al. (1982). These continua are re-normalized so that they match the observed mean Ly α transmission at $z \sim 3$ that was measured from high-resolution spectra (Faucher-Giguère et al. 2008; Becker et al. 2013) to prevent bias from the low spectral resolution of the SDSS spectrograph (as described in Dall'Aglio, Wisotzki & Worseck 2009). The reconstructions were tested with an independent set of 4597 quasars from eBOSS. As described in Bosman et al. (2022), this testing revealed that there is no bias in reconstructing the blue-side emission lines and that the method predicts the underlying continuum within 8 per cent. The reconstruction error on this testing set gives us the mean, $\boldsymbol{\mu}_{\rm cont}$, and covariance, $\boldsymbol{\Sigma}_{\rm cont}$, of the PCA reconstruction as shown in fig. 2 of Bosman et al. (2022).

In the following steps, we always forward-model the full wavelength-dependent uncertainties from the reconstruction of $F_{\rm cont}(\lambda_{\rm rest})$ into all measurements and model comparisons. We do this by randomly drawing realizations of the continuum error, $\boldsymbol{E}_{\rm cont} \sim N(\boldsymbol{\mu}_{\rm cont}, \boldsymbol{\Sigma}_{\rm cont})$, where N is the normal distribution. We create a realization of the predicted continuum with this error, $\boldsymbol{C}_{\rm pred}$, from the fit quasar continuum, $\boldsymbol{C}_{\rm fit}$, via:

$$\boldsymbol{C}_{\rm pred} = \boldsymbol{C}_{\rm fit} \times \boldsymbol{E}_{\rm cont}. \quad (1)$$

We use 500 of these continuum draws to analyse each quasar's spectrum. When we performed bootstrap re-sampling as described in Section 4.3, each draw uses a random selection of these 500 continua. Figures showing all PCA fits and blue-side predictions for all extended XQR-30 quasars are shown in Zhu et al. (2021).

3.2 Pixel masking

We want to use flux from the quasar continuum that exclusively corresponds to Ly α forest absorption. To do this, we only use wavelengths larger than the Ly β emission at the redshift of the quasar, or $\lambda_{\rm rest} > 1026$ Å. Additionally, we want to exclude the quasars proximity zone, which is the region close to the quasar where the IGM has been ionized by the quasar's own emission and the transmission is enhanced. For this reason, we consider $\lambda_{\rm rest} < 1185$ Å following Bosman et al. (2022) which corresponds to ~7650 km s^{-1} from emission at $z \sim 6$. This is a conservative estimate based on Bosman et al. (2018), which found no effect on the Ly α transmission in spectral stacks over this wavelength.

The data reduction procedure should automatically reject outlier pixels. However, we check for and exclude anomalous pixels that meet either of the following conditions: the SNR at the unabsorbed continuum level is < 2 per pixel or if pixels have negative flux at $>3\sigma$ significance. This excludes 0 per cent of pixels for the SNR cut at all redshifts and 0.07 − 0.47 per cent of pixels for the negative flux cut depending on redshift.

3.3 DLA exclusion

Damped Ly α absorption systems (DLAs) are intervening systems along quasar sightlines with hydrogen column densities $N_{\rm H\,I} \geq 10^{20.3}$ cm^{-2}. These systems result in significant damping wings in the Ly α absorption profile (Wolfe, Gawiser & Prochaska 2005; Rafelski et al. 2012). DLAs in quasar spectra at $z \gtrsim 6$ can cause complete absorption of Ly α transmission over $\Delta v = 2000$ km s^{-1} and additional suppression over $\Delta v \gtrsim 5000$ km s^{-1} intervals (D'Odorico et al. 2018; Bañados et al. 2019; Davies 2020). DLAs can arise in the circumgalactic medium (CGM) of galaxies which are not typically included in reionization simulations, including those discussed in Section 5. For this reason, we attempt to remove DLAs from our observations based on the presence of metals in the spectra. This does leave open the possibility that DLAs from pristine neutral patches of the IGM remain in our observations.

We remove DLAs by identifying and masking out their locations in our spectra. The detection of $z \gtrsim 5$ DLAs relies on the identification of associated low-ionization metal absorption lines, since the Ly α absorption from the DLA may not be distinguishable from the highly opaque IGM. The typical transitions are C II, O I, Si II, and Mg II. DLA metallicities at $z \gtrsim 5$ vary so even relatively weak metal absorption could indicate a DLA. The identification of intervening metal absorbers in the extended XQR-30 sample has been described in detail in Davies et al. (2023) and Sodini et al. (2024). Due to the high SNR of the X-Shooter spectra, we expect to be > 90 per cent complete to absorption corresponding to $\log N_{\rm Mg\,II}/{\rm cm}^{-2} \gtrsim 13$.

We adopt the following criteria for our masks, following Bosman et al. (2022). We mask the central $\Delta v = 3000$ km s^{-1} for systems with metal column densities $\log N_{\rm C\,II}/{\rm cm}^{-2} > 13$, $\log N_{\rm O\,I}/{\rm cm}^{-2} > 13$, or $\log N_{\rm Si\,II}/{\rm cm}^{-2} > 12.5$, measured through the $\lambda_{\rm rest} = 1334.53$ Å, 1302.16 Å, and 1526 Å transitions, respectively. When none of these

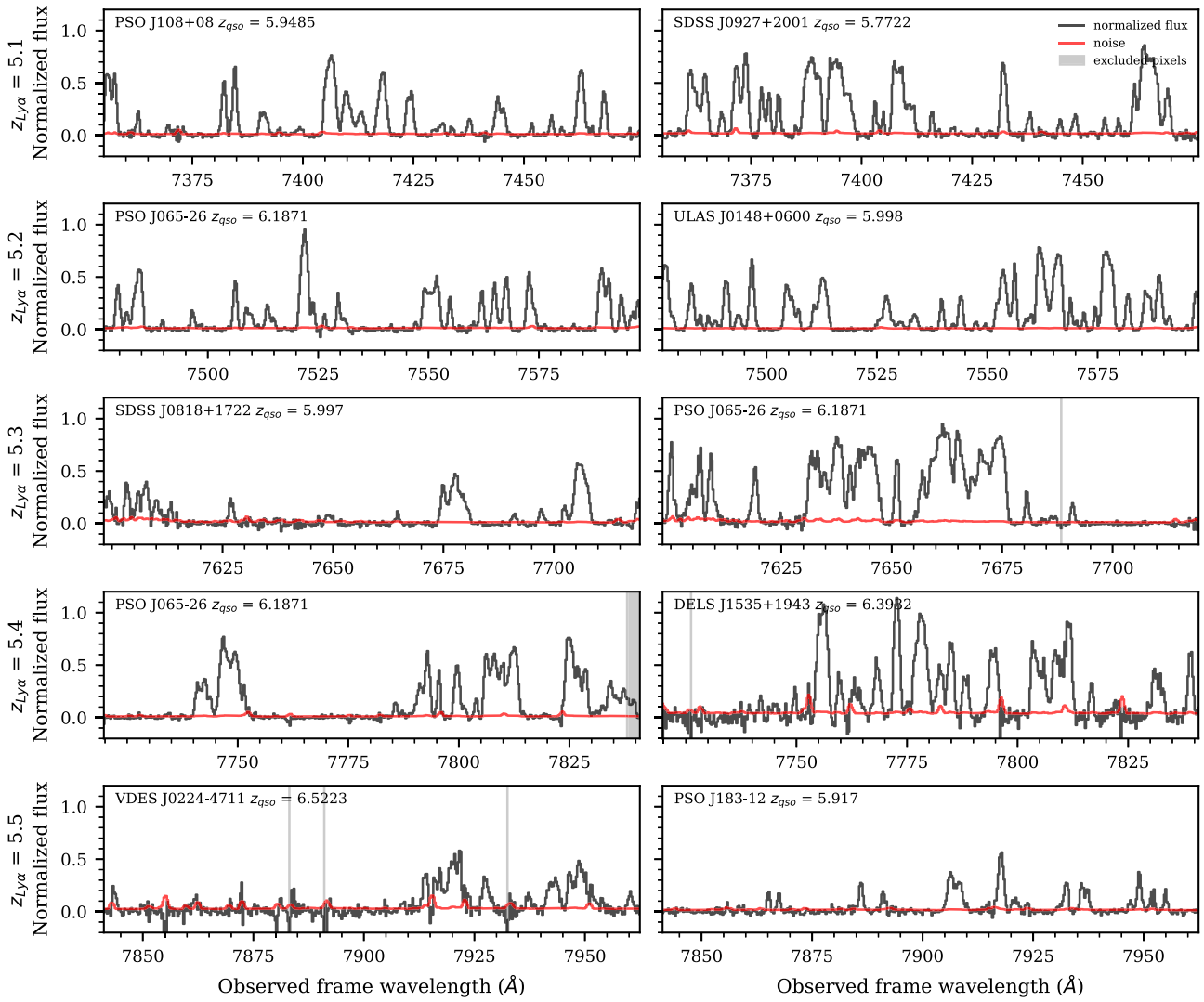

Figure 2. This figure shows the continuum normalized flux for two randomly selected quasars in each row at five values of z of the Ly α forest from $5.1 \leq z \leq 5.5$. These sections are centred on the given z and span $\Delta z = 0.05$. The continuum normalized flux is shown in black with the continuum normalized uncertainty in red. The shaded regions indicate excluded pixels based on the masking procedure described in Sections 3.2 and 3.3. Each row shares the same y-axis to demonstrate the decrease in $\langle F \rangle$ with increasing z (down the rows). Note that the normalized flux for all the quasars considered in each measurement can be found in the online supplementary material.

ions are accessible, we also exclude the central $\Delta v = 3000\,\mathrm{km\,s^{-1}}$ for systems with $\log N_{\mathrm{Mg\,II}}/\mathrm{cm}^{-2} > 13$ based on the high rates of co-occurrence of the Mg II 2796.35, 2803.53 Å doublet (Cooper et al. 2019). We exclude a larger window of $\Delta v = 5000\,\mathrm{km\,s^{-1}}$ around intervening systems with $\log N_{\mathrm{O\,I,\,C\,II,\,Si\,II,\,Mg\,II}}/\mathrm{cm}^{-2} > 14$ due to the likely presence of extended damping wings. We do not exclude systems based on the presence of highly ionized ions alone (e.g. C IV, Si IV) since the corresponding gas is likely highly ionized (Cooper et al. 2019).

We investigate the effect of this mask on the measurement of the autocorrelation function in Appendix B.

3.4 Resulting normalized flux

After combining the masks of the bad pixels discussed in Section 3.2 and the DLAs discussed in Section 3.3, we only considered sightlines that maintain at least 10 per cent of the pixels in a given redshift bin. Only using spectra that maintain at least 10 per cent of pixels limits noisy contributions to the measurement from short spectra that may only consist of one transmission spike. Two random examples of the normalized flux from quasars in our sample at each redshift are shown in Figs 2 and 3. The normalized flux for all the sightlines used in each redshift bin can be found in the online supplementary material, which demonstrate the variance between the sightlines at a given redshift.

Fig. 2 shows the normalized flux for two quasar sightlines for $5.1 \leq z \leq 5.5$ while Fig. 3 has the same for $5.6 \leq z \leq 6.0$. Each row has the same z and each column shows a random quasar sightline. The value of z increases down the rows. The y-axis is fixed within Figs 2 and 3 though it varies between the two figures. The fixed y-axis illustrates the rough trend of decreasing $\langle F \rangle$ with increasing z. Both of the random sightlines shown at $z = 6$ have very limited transmission, which highlights the difficulty in making statistical measurements of the Ly α forest at high redshifts.

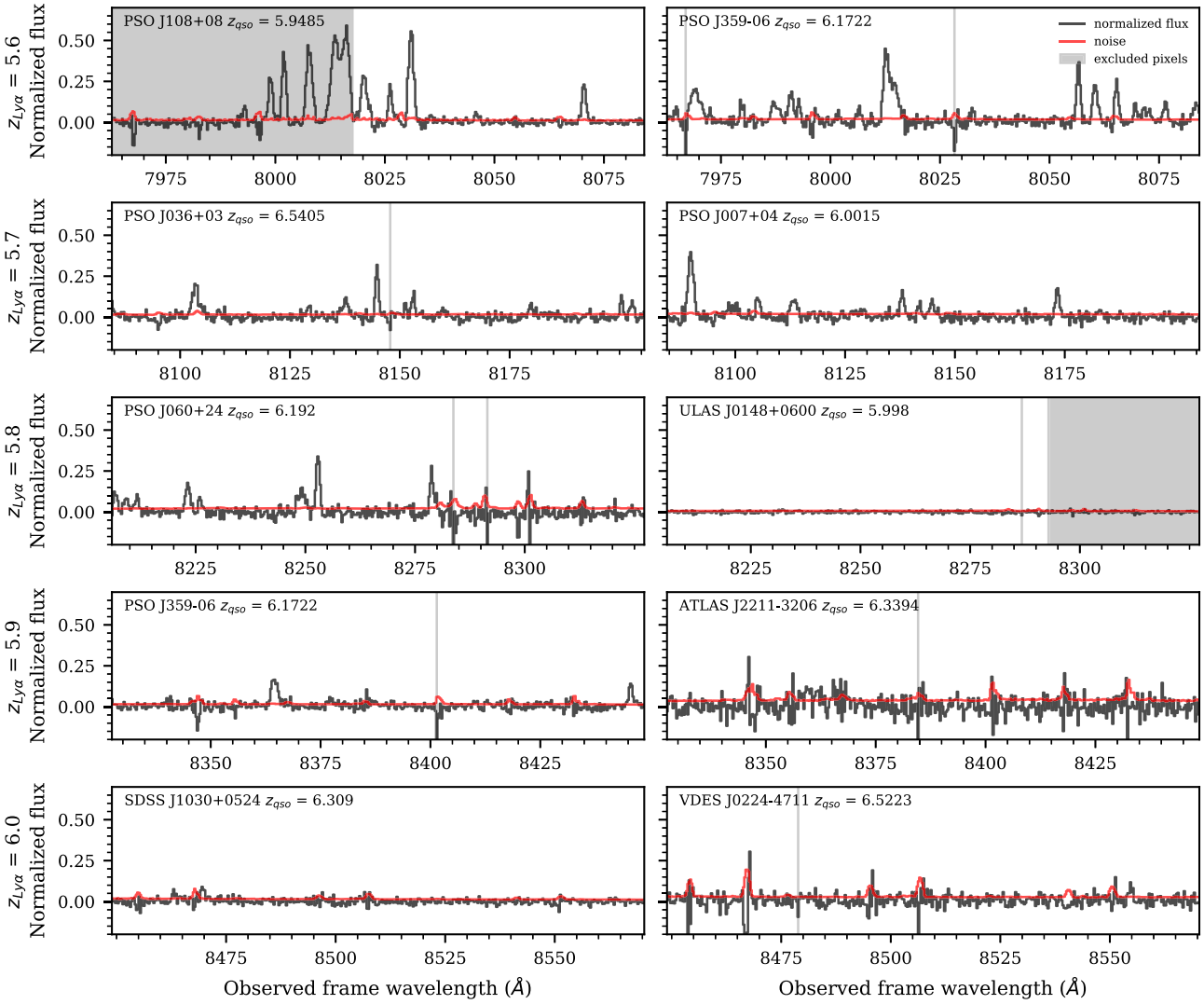

Figure 3. The same as Fig. 2 except for $5.6 \leq z \leq 6.0$. The y-axis spans a smaller range than that in Fig. 2.

4 RESULTS

4.1 Mean flux

The mean flux in this paper was calculated as the average of the normalized flux values for the non-excluded pixels as shown in Figs 2 and 3. The resulting values are reported in Table 2 and plotted as a function of redshift in Fig. 4. The error on the $\langle F \rangle$ values were computed by bootstrap re-sampling the quasar sightlines considered at each z for 500 000 data set realizations and computing the variance on these values. See Section 4.3 for more information on how the bootstrap realizations were generated.

Fig. 4 shows the $\langle F \rangle$ values computed in this work in red, the previous measurement of Bosman et al. (2022) in black, and the measurements of Becker et al. (2013), Bosman et al. (2018), and Eilers et al. (2018) in blue, orange, and green, respectively. Our measurement is in agreement with that from Bosman et al. (2022), as is expected since the data used here is a subset of that used in that work and our method is the same. In addition, we use the same continuum reconstruction and masking procedure as in Bosman et al. (2022). At $z = 5.1$ and $z = 5.2$ our measurement appears greater than that from Bosman et al. (2022), but the data set we considered is much smaller and the measurements are consistent within the error bars. A discussion of the agreement of $\langle F \rangle$ with previous work can be found in Bosman et al. (2022).

Table 2. The second column lists the numbers of lines of sight at each z in our sample. The third column reports the mean flux, $\langle F \rangle$, value that was directly computed from this sample. The error on $\langle F \rangle$ comes from bootstrap re-sampling of the sightlines.

z	$N_{\rm los}$	$\langle F \rangle$
5.1	24	0.1456 ± 0.0075
5.2	29	0.1314 ± 0.0072
5.3	29	0.1097 ± 0.0087
5.4	33	0.0830 ± 0.0086
5.5	34	0.0567 ± 0.0055
5.6	34	0.0474 ± 0.0053
5.7	29	0.0269 ± 0.0044
5.8	26	0.0181 ± 0.0035
5.9	15	0.0089 ± 0.0018
6.0	14	0.0090 ± 0.0023

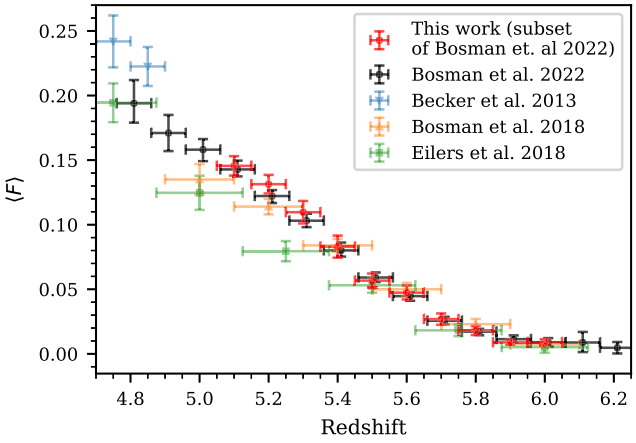

Figure 4. Recent measurements of the average Ly α transmission, $\langle F \rangle$, at high-z. The measured $\langle F \rangle$ from this work are shown in red. This is computed directly by taking the average of the non-excluded normalized flux values from the masks created as discussed in Sections 3.2 and 3.3. The errors come from bootstrap re-sampling the quasar sightlines. Note that the measurement shown in red comes from a subset of the quasar sightlines used in Bosman et al. (2022) which are plotted in black. Additional data points from previous works are shown in blue, orange, and green over the same z range (Becker et al. 2013; Bosman et al. 2018; Eilers et al. 2018).

4.2 Autocorrelation function

The autocorrelation function of the flux ($\xi_F(\Delta v)$) is defined as

$$\xi_F(\Delta v) = \langle \boldsymbol{F}(v)\boldsymbol{F}(v + \Delta v) \rangle, \quad (2)$$

where $\boldsymbol{F}(v)$ is the normalized flux of the Ly α forest and the average is performed over all pairs of pixels at the same velocity lag (Δv). The pixels that have been masked as discussed in Sections 3.2 and 3.3 are not used when computing the autocorrelation function for each quasar. See Appendix B for a discussion of the effect of the DLA exclusion on the measurement of the autocorrelation function. Note that different quasar sightlines will have a different number of pixel pairs contributing to the same velocity bin. Thus, when combining the different quasar sightlines, we weight each quasar's contribution by the numbers of pixel pairs in each bin. The number count of pixel pairs contributing to each autocorrelation function bin is output during the autocorrelation function computation.

We compute the autocorrelation function with the following consideration for the velocity bins. We start with the left edge of the smallest bin to be 40 km s^{-1} and use linear bins with a width of 40 km s^{-1} up to 280 km s^{-1}. The choice of 40 km s^{-1} was done as it is roughly the size of a resolution element for these observations. Then we switch to logarithmic bin widths where $\log_{10}(\Delta v) = 0.058$ out to a maximal distance of 2700 km s^{-1}. This results in 22 velocity bins considered where the first 6 have linear spacing. The centre of our smallest bin was 60 km s^{-1} and our largest bin was 2223 km s^{-1}, which corresponds to ∼16 cMpc h^{-1} at $z = 5.5$. We chose to use linear bins on the smallest scales because the effect of $\lambda_{\rm mfp}$ is greatest on small scales and these scales already have access to the most pixel pairs which reduces noise. Larger scales are more sensitive to $\langle F \rangle$ than $\lambda_{\rm mfp}$ so having fewer bins here is not as important. In addition, there are fewer pixel pairs at large scales to begin with so using larger bins will increase the pixel pairs per bin and reduce noise.

Previously, Wolfson et al. (2023b) demonstrated the sensitivity of the autocorrelation function to $\lambda_{\rm mfp}$ for mock data at $z \geq 5.4$. Generally, they found that shorter $\lambda_{\rm mfp}$ values cause a greater boost in the autocorrelation function on the smallest scales. We compute the autocorrelation functions of the XQR-30 data set discussed in Section 2. The measured autocorrelation function from the extended XQR-30 data set can be seen in Figs 5 and 6. The errors on these plots come from bootstrap sampling of the quasar sightlines when computing the mean autocorrelation function and will be discussed in more detail in Section 4.3. The first few velocity bins of the final measurement with error from the diagonal of the covariance matrix estimated via bootstrap re-sampling are in Table 3. The full measurement, error bars, as well as the full bootstrap covariance matrices for each redshift are available to download online.[2]

Fig. 5 has two panels that show the autocorrelation function of this data set at different z. The top panel shows $5.1 \leq z \leq 5.5$ while the bottom panel shows $5.6 \leq z \leq 6.0$. They are shown in two different panels in order to better accommodate the dynamic range of the autocorrelation function over our range of z. The overall amplitude of the autocorrelation function of the flux is set by $\langle F \rangle^2$, which decreases with increasing z.

In order to better visually demonstrate the differences in the shape of the autocorrelation function on small scales, we also plot the measured autocorrelation function normalized and shifted by $\langle F \rangle^2$ at each z in Fig. 6. Note that the $\langle F \rangle$ value used is redshift dependent and is reported in Table 2. This is equivalent to the autocorrelation function of the flux density field. The colour of the normalized autocorrelation function at each z matches those from Fig. 5. This has been split into two panels for visual clarity to more easily see the behaviour in each redshift bin. The top panel has $z = 5.1, 5.3, 5.5, 5.7, 5.9$ while the bottom panel has 5.2, 5.4, 5.6, 5.8, 6.0. By looking at the smallest scales, $v < 500$ km s^{-1} or $x < 4$ cMpc h^{-1} at $z = 5.5$, there is a trend of increasing small-scale values of the autocorrelation function with increasing redshift. For example, the lines for $5.8 \leq z \leq 6.0$ have the greatest autocorrelation value (in shades of purple). Note that these points have the largest error bars, likely caused by both the limited number of sightlines and the low transmission at these redshifts. Both $\langle F \rangle$ and $\lambda_{\rm mfp}$ affect the small scale boost in the autocorrelation function. Smaller $\langle F \rangle$ will lead to larger fluctuations in the flux contrast field and thus a boost on the small scales. Wolfson et al. (2023b) found that shorter $\lambda_{\rm mfp}$ values also cause a boost in the autocorrelation function on the smallest scales. These effects are not completely degenerate since the overall autocorrelation function shape differs as shown in the forecast measurements of Wolfson et al. (2023b).

We isolate the redshift evolution of the smallest velocity bin (60 km s^{-1}) of the normalized autocorrelation function in Fig. 7. Again, the $\langle F \rangle$ value used is redshift dependent and is reported in Table 2. The errors are computed by propagating the statistical uncertainty from bootstrap re-sampling both the autocorrelation function and $\langle F \rangle$. In general, these values increase with redshift, which is expected from decreasing $\langle F \rangle$ as well as $\lambda_{\rm mfp}$. However, the errors also increase with redshift and the values at the highest redshift are consistent with each other within errors.

4.3 Bootstrap covariance matrices

In order to calculate the error on $\langle F \rangle$ and the autocorrelation functions we used bootstrap re-sampling. To compute the values we performed averages over $N_{\rm boot} = 500\,000$ realizations of the data set. Each realization is a random selection of $N_{\rm los}$ quasars with replacement. In addition, each choice of quasar goes along with a choice of the 500 continuum realizations that were generated as described at the end

[2] https://github.com/mollywolfson/lya_autocorr/

Table 3. The table lists the autocorrelation function measurement for the first six bins of the autocorrelation function at all z with errors from the diagonal of the covariance matrix estimated from bootstrap re-sampling. The full measurement values of the autocorrelation function at all z can be found online.

z	Central velocity (km s^{-1})						
	60	100	140	180	220	260	...
5.1	0.0413 ± 0.0040	0.0331 ± 0.0035	0.0291 ± 0.0033	0.0271 ± 0.0032	0.0270 ± 0.0032	0.0262 ± 0.0031	...
5.2	0.0365 ± 0.0043	0.0290 ± 0.0037	0.0256 ± 0.0032	0.0235 ± 0.0028	0.0222 ± 0.0025	0.0210 ± 0.0024	...
5.3	0.0291 ± 0.0047	0.0227 ± 0.0042	0.0201 ± 0.0038	0.0191 ± 0.0037	0.0178 ± 0.0037	0.0168 ± 0.0036	...
5.4	0.0185 ± 0.0032	0.0143 ± 0.0026	0.0130 ± 0.0024	0.0122 ± 0.0023	0.0117 ± 0.0023	0.0113 ± 0.0022	...
5.5	0.0105 ± 0.0020	0.0076 ± 0.0016	0.0064 ± 0.0013	0.0054 ± 0.0012	0.0048 ± 0.0010	0.00436 ± 0.00081	...
5.6	0.0078 ± 0.0016	0.0057 ± 0.0012	0.0047 ± 0.0011	0.00406 ± 0.00093	0.00365 ± 0.00083	0.00361 ± 0.00084	...
5.7	0.00298 ± 0.00082	0.00206 ± 0.00062	0.00193 ± 0.00065	0.00182 ± 0.00060	0.00158 ± 0.00048	0.00141 ± 0.00044	...
5.8	0.00197 ± 0.00065	0.00120 ± 0.00040	0.00082 ± 0.00026	0.00072 ± 0.00022	0.00070 ± 0.00027	0.00083 ± 0.00032	...
5.9	0.00055 ± 0.00023	0.00030 ± 0.00013	0.000150 ± 0.000045	0.000124 ± 0.000085	0.00020 ± 0.00012	0.000189 ± 0.000095	...
6.0	0.00053 ± 0.00023	0.00027 ± 0.00014	0.000180 ± 0.000096	0.00023 ± 0.00014	0.00028 ± 0.00017	0.00019 ± 0.00012	...

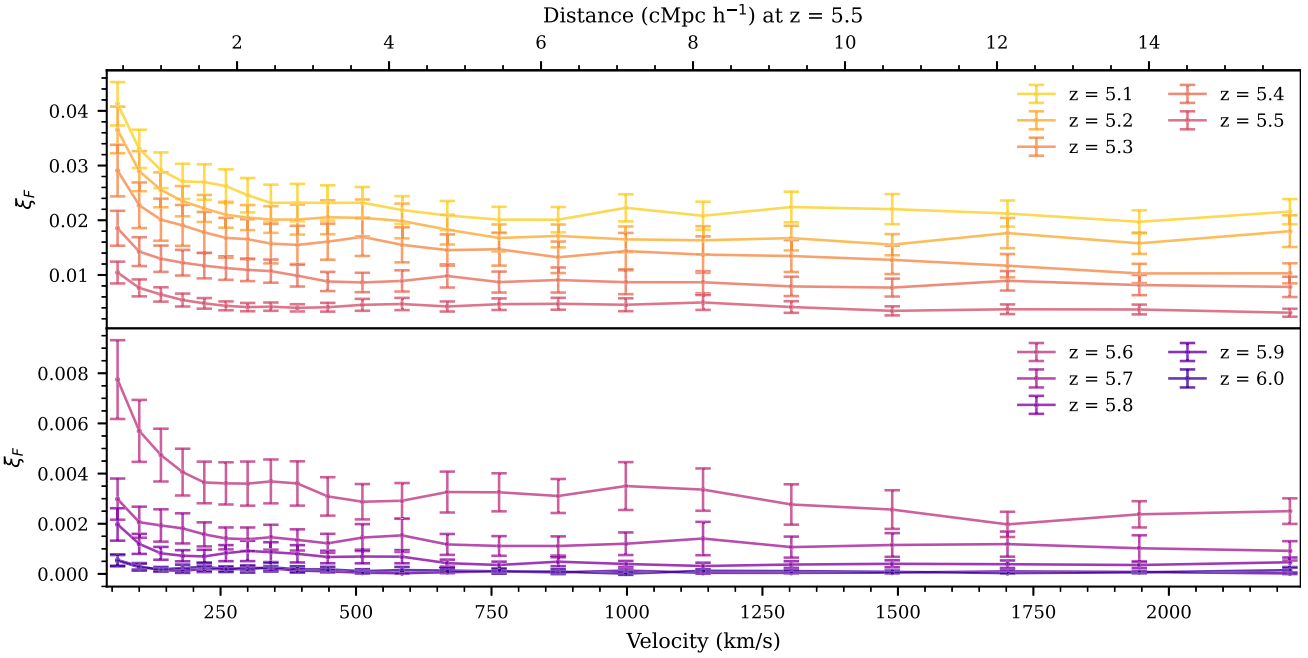

Figure 5. The autocorrelation function of Ly α transmission in ten redshift bins for XQR-30 data. The top panel shows the lower z bins, $5.1 \leq z \leq 5.5$ while the lower panel shows the higher z bins, $5.6 \leq z \leq 6.0$. The main trend seen in these plots is the evolution of $\langle F \rangle$ which is very small at high-z.

of Section 3.1. The computed mean flux for the ith sample is thus $\langle F \rangle_i$ and the error on $\langle F \rangle$, σ_F is:

$$\sigma_F = \left(\frac{1}{N_{\rm boot} - 1} \sum_{i=1}^{N_{\rm boot}} (\langle F \rangle_i - \langle F \rangle)^2 \right)^{1/2}. \quad (3)$$

These errors are reported in Table 2 and shown in Fig. 4.

For the autocorrelation function, $\boldsymbol{\xi}$, we compute the entire bootstrap covariance matrix, not only the diagonal error. Again we chose $N_{\rm boot}$ realizations of the observed data set by randomly selecting $N_{\rm los}$ quasars with replacement each with their own random selection of the continuum realization. For any given bootstrap realization we computed the average of the autocorrelation function over the chosen sightlines to construct a realization of the average autocorrelation function, $\boldsymbol{\xi}_i$. The covariance matrix was then computed by averaging over the ensemble of bootstrap realizations in the following way:

$$\boldsymbol{\Sigma}_{\rm boot} = \frac{1}{N_{\rm boot} - 1} \sum_{i=1}^{N_{\rm boot}} (\boldsymbol{\xi}_i - \boldsymbol{\xi}_{\rm data})(\boldsymbol{\xi}_i - \boldsymbol{\xi}_{\rm data})^{\rm T}. \quad (4)$$

For visualization purposes, we use the diagonal of the bootstrap covariance matrices to estimate the error bars on the autocorrelation function shown in Fig. 5. Specifically we define $\sigma_{\rm boot} = \sqrt{(\rm diag(\boldsymbol{\Sigma}_{boot}))}$. The diagonal of the covariance matrix is not a full description of the error since the bins of the autocorrelation function are highly correlated and should thus fluctuate in a correlated way, thus making the full covariance matrix necessary in any computations. The error bars in Fig. 6, $\boldsymbol{\sigma}_\Delta$, come from combining the bootstrap estimate of the errors for $\boldsymbol{\xi}_F$ with bootstrap estimate of the errors on $\langle F \rangle$ via:

$$\boldsymbol{\sigma}_\Delta = \frac{\boldsymbol{\xi}_F - \langle F \rangle^2}{\langle F \rangle^2} \sqrt{\left(\frac{\boldsymbol{\sigma}_{\rm boot}}{\boldsymbol{\xi}_F} \right)^2 + 2 \left(\frac{\sigma_F}{\langle F \rangle} \right)^2} \quad (5)$$

Additionally we define the correlation matrix, $\mathbf{C}$, which expresses the covariances between jth and kth bins in units of the diagonal elements of the covariance matrix. This is done for the jth, kth element by

$$C_{jk} = \frac{\Sigma_{jk}}{\sqrt{\Sigma_{jj} \Sigma_{kk}}}. \quad (6)$$

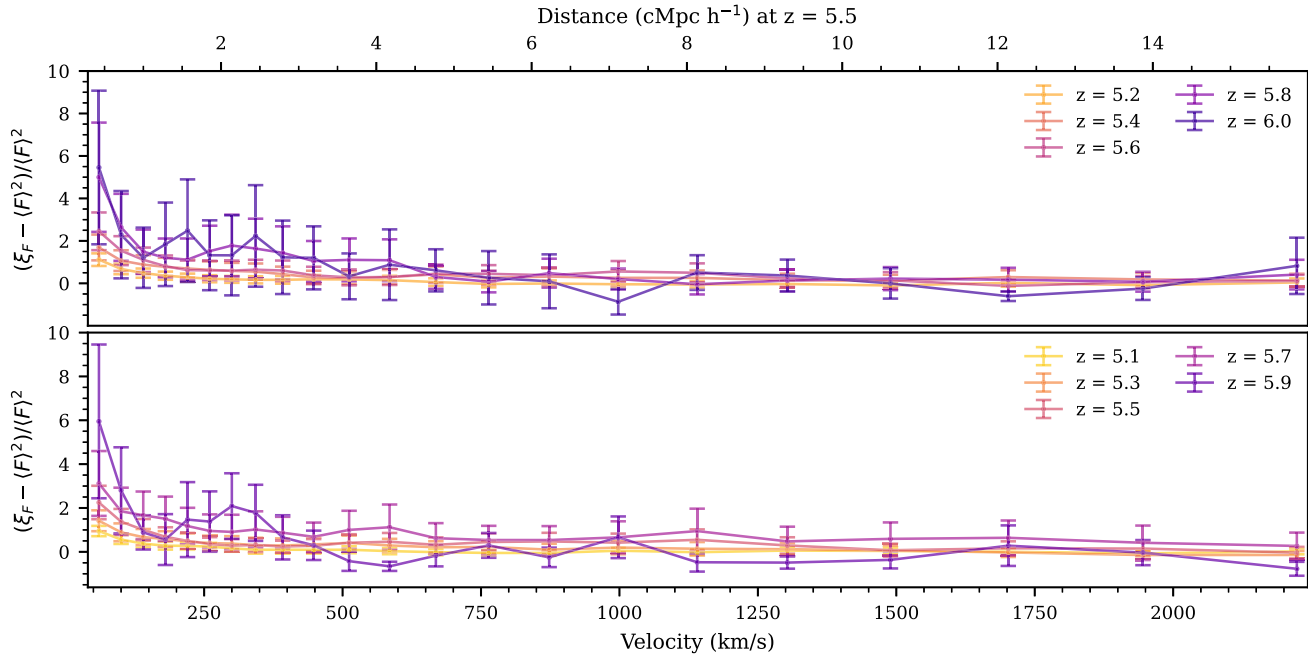

Figure 6. The autocorrelation function of Ly α transmission normalized and shifted by the mean transmission, $\langle F\rangle$, in ten redshift bins for XQR-30 data. This is equivalent to the autocorrelation function of the flux density field. The errors are computed by propagating the statistical uncertainty from bootstrap re-sampling both the autocorrelation function and $\langle F\rangle$. This is split into two panels for visual clarity, so as to not overcrowd the panels. The top panel has z = 5.1, 5.3, 5.5, 5.7, 5.9 while the bottom panel has 5.2, 5.4, 5.6, 5.8, 6.0. This figure makes the trend of higher redshift bins having larger boosts of the autocorrelation function on small scales when dividing out the flux evolution more visible.

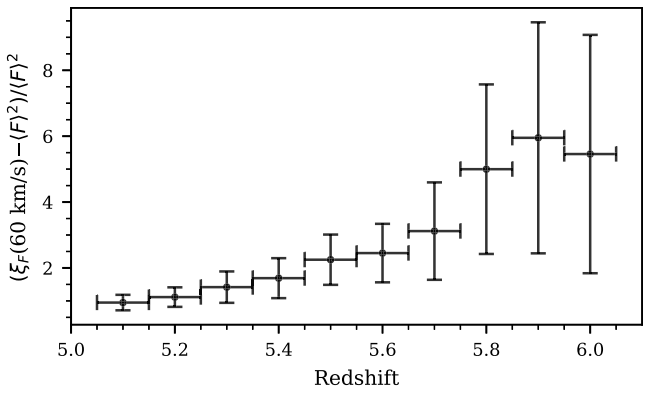

Figure 7. The value of the first bin of the autocorrelation function of Ly α transmission normalized and shifted by the mean transmission, $\langle F\rangle$, as a function of redshift. The errors are computed by propagating the statistical uncertainty from bootstrap re-sampling both the autocorrelation function and $\langle F\rangle$. These values are also shown in Fig. 6. There is a general trend of increasing value with redshift, though the errors also increase. The highest redshift values are consistent with no evolution.

The bootstrap correlation matrices for the measured autocorrelation functions at each z are shown in Fig. 8. Based on the simulated correlation matrices from Wolfson et al. (2023b), we expect there to be significant off-diagonal values of these bootstrap correlation matrices. This is because, generally, each pixel in the Ly α forest contributes to every bin of the autocorrelation function so the different velocity bins in the autocorrelation function are highly covariant. Large off-diagonal values are seen in the bootstrap correlation matrices in Fig. 8 for $z < 5.8$. At the highest three redshifts, especially $z = 5.9$ and $z = 6.0$, the number of quasar lines of sight are quite small and the transmission is quite low, leading to large noise fluctuations and non-converged off-diagonal values. In particular, there are negative values off the diagonal for $z = 5.9$ and $z = 6.0$ which we do not see in our simulated covariance matrices. We expect noisy fluctuations in the off-diagonal covariance matrix values to go away with the addition of more quasar sightlines, though low transmission at the highest redshifts will still make this computation difficult.

5 MODELLING THE MEASUREMENT

In order to interpret the physical implications of the measured autocorrelation function, we construct forward models with the properties of the observed quasars. Functions to convert any set of simulation skewers into autocorrelation function measurements are available online at `https://github.com/mollywolfson/lya_autocorr/`. In addition, there is a `JUPYTER NOTEBOOK` that goes through an example of forward-modelling simulation skewers and then computing the autocorrelation function. The simulation method used here was introduced in Wolfson et al. (2023b) for a simplified mock data set. We have updated this method to include continuum uncertainty, noise vectors from observational data, and a $\Gamma_{\mathrm{H\,I}}$ box that matches the density field of the main simulation suite. We will briefly describe this updated method here, for more information see Wolfson et al. (2023b).

Note that this paper is using a simple model that only varies λ_{mfp} and $\langle F\rangle$. More sophisticated modelling that includes variation in the IGM thermal state, patchy reionization, and more robust UVB modelling is left for future work. For an initial investigation into the effect of the IGM thermal state and inhomogeneous reionization on the Ly α forest flux autocorrelation function see Wolfson et al. (2023a). They found that these mainly affect scales $v < 100\,\mathrm{km\,s^{-1}}$, which corresponds to only the smallest bin considered here. Thus,

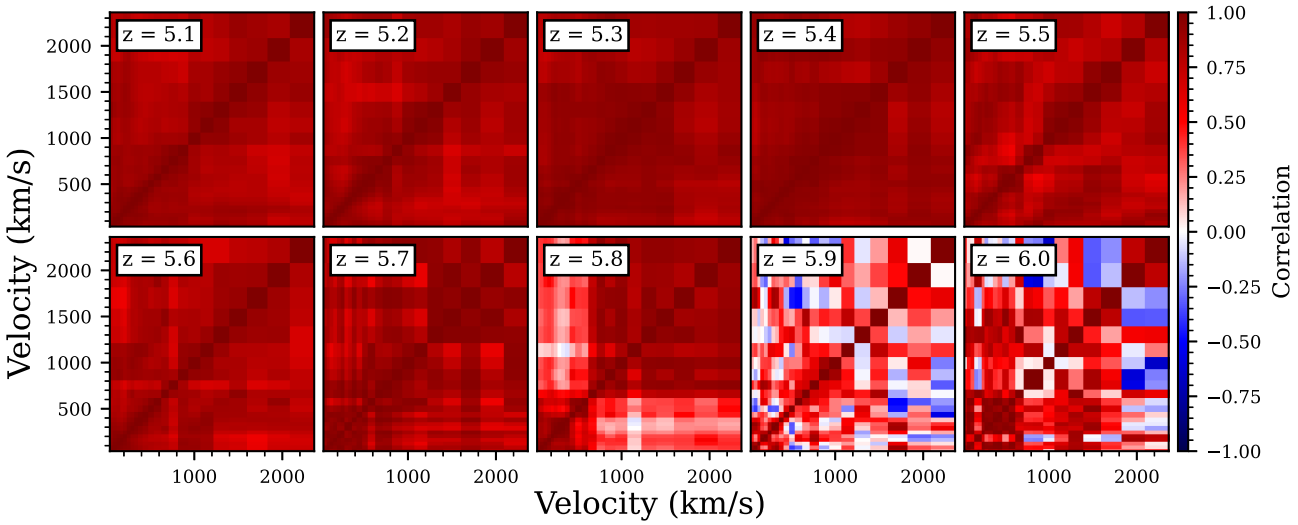

Figure 8. The correlation matrices from bootstrap re-sampling the autocorrelation function in the ten redshift bins considered in this work. For $z < 5.8$ we see very strong positive off-diagonal values of the correlation matrices. This behaviour is expected since each pixel in the Ly α forest contribute to every bin of the autocorrelation function, making these bins highly correlated. The fluctuations in the correlation matrix values are caused by noise due to the limited sightlines available to bootstrap. At $z \geq 5.8$ the number of sightlines is small and the transmission is low, causing large noise fluctuations. For $z = 5.9$ and $z = 6.0$, the sightlines are so few and so non-transmissive that noise fluctuations lead to negative values in the correlation matrices. There is no physical explanation for these negative values. The numbers of sightlines used at each z are listed in Table 2.

while additional simulation work is necessary to include all relevant parameters, the models presented here are sufficient for an initial comparison.

5.1 Simulation box

To begin, we use a NYX simulation box (Almgren et al. 2013). NYX is a hydrodynamical simulation code designed to simulate the Ly α forest with updated physical rates from Lukić et al. (2015). The NYX box has a size of $L_{\rm box} = 100$ cMpc h^{-1} with 4096^3 dark matter particles and 4096^3 baryon grid cells. This box is reionized by a Haardt & Madau (2012) uniform UVB that is switched on at $z \sim 15$, which means these simulation boxes do not include the effects of a patchy, inhomogeneous reionization.

We have three snapshots of this simulation at $z = 5.0$, $z = 5.5$, and $z = 6$ and we want to model all ten redshifts $5.1 \leq z \leq 6.0$ with $\Delta z = 0.1$. In order to consider the redshifts for which we do not have a simulation output, we select the nearest snapshot and use the desired redshift when calculating the proper size of the box and the mean density. This means we use the density fluctuations, temperature, and velocities directly from the nearest NYX simulation output. Previously, in Wolfson et al. (2023b) we tested this choice of simulation interpolation by using the $z = 6.0$ simulation snapshot to generate skewers at $z = 5.7$ and found no change in the results.

In addition, we have a grid of boxes of $\Gamma_{\rm H\,I}/\langle\Gamma_{\rm H\,I}\rangle$ values generated with the seminumerical method of Davies & Furlanetto (2016) corresponding to a fluctuating UVB for different $\lambda_{\rm mfp}$ values, all at $z = 5.5$. These boxes have a size of $L_{\rm box} = 100\,h^{-1}$ cMpc, 64^3 pixels, and are generated from the density field of the NYX simulation box. The method of Davies & Furlanetto (2016) uses Mesinger & Furlanetto (2007) and Bouwens et al. (2015) to create haloes and assign UV luminosities from the density field. They then get the ionizing luminosity of each galaxy by assuming it to be proportional to its UV luminosity where the constant of proportionality is left as a free parameter. Finally the ionizing background radiation intensity, J_ν, is computed by a radiative transfer algorithm and $\Gamma_{\rm H\,I}$ is finally calculated by integrating over J_ν. For more information on this method of generating $\Gamma_{\rm H\,I}$ boxes see Davies & Furlanetto (2016), Davies et al. (2018a), or Wolfson et al. (2023b). Note that this modelling assumes a number of relations, such as local $\lambda \propto \Gamma_{\rm H\,I}^{2/3}\Delta^{-1}$. Additional work looking into the effect on the UVB from varying these assumptions is necessary to get robust constraints on $\lambda_{\rm mfp}$ from these models. We leave this for future work and use this simple one parameter model for an initial, qualitative comparison with the data.

To combine the NYX box with the $\Gamma_{\rm H\,I}$ values generated via the Davies & Furlanetto (2016) method, we linearly interpolated $\log(\Gamma_{\rm H\,I}/\langle\Gamma_{\rm H\,I}\rangle)$ onto the higher resolution grid of the NYX simulation box. We then re-scale the optical depths from the NYX box with a constant UVB, $\tau_{\rm const.}$, by these fluctuating $\Gamma_{\rm H\,I}$ values to get the optical depths for a fluctuating UVB, $\tau_{\rm mfp} = \tau_{\rm const.}/(\Gamma_{\rm H\,I}/\langle\Gamma_{\rm H\,I}\rangle)$. This implies that we need to know $\langle\Gamma_{\rm H\,I}\rangle$ to compute our final optical depths, which is not known a priori. We therefore determine this value by matching an overall mean flux $\langle F\rangle$, where we vary $\langle F\rangle$ over a range of models based off the measurement of Bosman et al. (2022). We look at the relationship between $\langle F\rangle$, $\lambda_{\rm mfp}$, and $\langle\Gamma_{\rm H\,I}\rangle$ in Appendix C. We generate 1000 skewers from this simulation method for each $\lambda_{\rm mfp}$ and $\langle F\rangle$ at each z for $5.1 \leq z \leq 6.0$. These skewers come from the same location in the simulation box for all parameter values and z.

5.2 Forward modelling

Our simulations provide skewers of the optical depth of the Ly α forest for given $\lambda_{\rm mfp}$ and $\langle F\rangle$ values. In order to compare these (or any) simulated skewers with the results of our observational measurement, we forward model the telescope resolution, the noise properties of our observed sightlines, and the continuum uncertainty from the PCA continuum fit. This section will describe how each property is modelled for our simulation skewers, see the `lya-autocorr` git repository to follow along with an example simulation skewers being forward modelled.

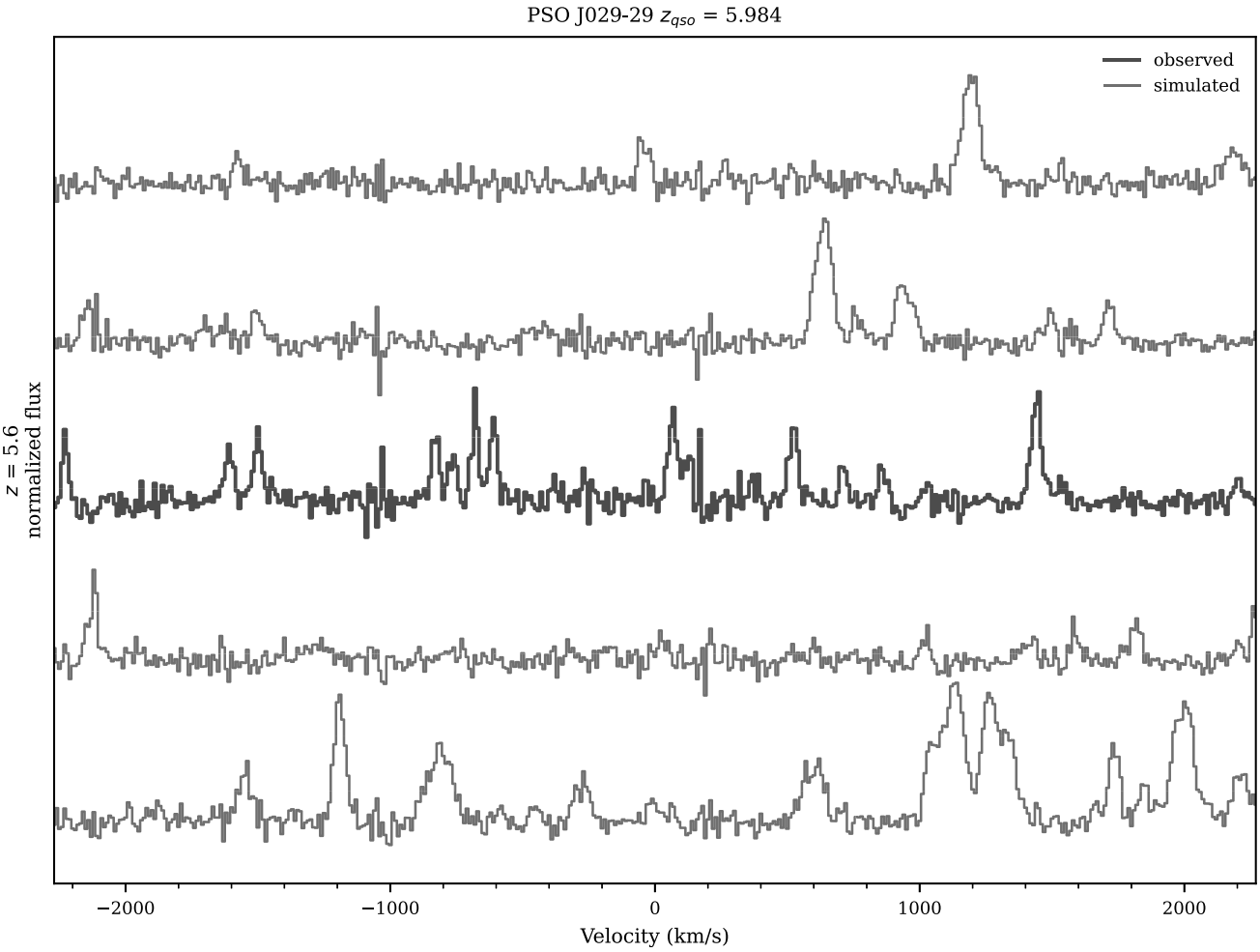

Figure 9. This figure compares the observed Ly α forest flux at $z = 5.6$ from PSO J029-29 with forward modelled simulation skewers modelled to have the same noise properties as this quasar. The thick line in the middle is the observed flux while the other four thinner lines are from the forward modelled simulations. The visual similarities between the observed and simulated Ly α forest flux shown here demonstrates the success of our forward-modelling procedure.

To model the resolution of X-shooter for visible light with a 0.9 arcsec slit, we convolved the flux by a Gaussian line-spread function with FWHM $\approx 34\,\mathrm{km\,s^{-1}}$. This corresponds the nominal resolving power ($R \sim 8800$) of the X-Shooter setup used for the XQR-30 data. However, as noted in Section 2 the actual data has a higher median resolving power in the visible of $R = 11\,400$ (D'Odorico et al. 2023). Future work will use the measured resolving power for each quasar in the modelling but using the nominal value for all is sufficient for this initial comparison. After using this Gaussian filter we interpolated the line-spread-function convolved flux onto the exact velocity grid from the observation. This step also reduced the simulation skewers from the box size to the same length as our observations, as 100 cMpc h^{-1} corresponds to $\Delta z \sim 0.3$ at the relevant redshifts and our observations have $\Delta z = 0.1$.

We add noise to the interpolated, line-spread-function convolved flux, $\boldsymbol{F}_{\rm res}$, according to the noise vectors for each quasar sightline, $\boldsymbol{\sigma}_{\rm qso}$, with random normal distribution realization, $\boldsymbol{N}_{\rm qso} \sim N(0, 1)$, via

$$\boldsymbol{F}_{\rm noise} = \boldsymbol{F}_{\rm res} + \left(\boldsymbol{N}_{\rm qso} \times \boldsymbol{\sigma}_{\rm qso}\right). \tag{7}$$

$\boldsymbol{F}_{\rm noise}$ is thus the flux modelled with both the resolution of the telescope and the noise properties of our observed sightlines. This modelling choice is valid because of the low flux in the Ly α forest at these redshifts such that we are background limited in the observations.

To model continuum error, we used the mean, $\boldsymbol{\mu}_{\rm cont}$, and covariance, $\boldsymbol{\Sigma}_{\rm cont}$, of the PCA reconstruction just as we do for the data as described in Section 3.1. We randomly draw realizations of the continuum error, $\boldsymbol{E}_{\rm cont} \sim N(\boldsymbol{\mu}_{\rm cont}, \boldsymbol{\Sigma}_{\rm cont})$, where N is the normal distribution. In our simulations, we do not fit and normalize by the quasar continuum so we model continuum error by:

$$\boldsymbol{F}_{\rm cont} = \boldsymbol{F}_{\rm noise} / \boldsymbol{E}_{\rm cont}, \tag{8}$$

where $\boldsymbol{F}_{\rm cont}$ is the final fully forward-modelled Ly α forest spectra. We investigate the effect of the continuum modelling on the resulting models of the autocorrelation function in Appendix A.

Ultimately, we generate $N_{\rm skewer}$ forward-modelled copies of each of the $N_{\rm los}$ quasars in the sample, where $N_{\rm skewer} = 1000$ from the simulation and $N_{\rm los}$ is the number of quasar sightlines at each redshift as listed in Table 2. For example at $z = 5.1$ we have $N_{\rm skewers} \times N_{\rm los} = 1000 \times 24 = 24\,000$ total forward-modelled Ly α forest spectra.

Fig. 9 shows the normalized flux of the $z = 5.6$ Ly α forest from PSO J029+29 with four examples of the normalized flux from our simulations that were forward modelled with this quasar's properties.

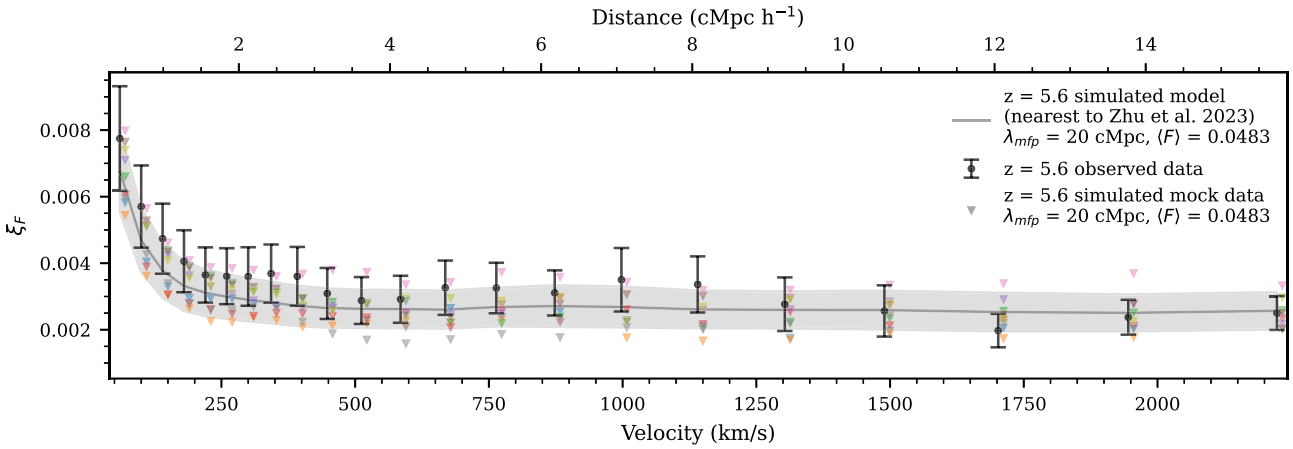

Figure 10. The black points show the observed autocorrelation function from the extended XQR-30 data discussed in this work at $z = 5.6$. The coloured triangles show the autocorrelation value for nine different simulated mock data sets. The mock data sets shown here were all modelled with $\lambda_{\rm mfp} = 20$ cMpc and $\langle F \rangle = 0.0483$, the closest $\lambda_{\rm mfp}$ value to the Zhu et al. (2023) measurement and the closest $\langle F \rangle$ value to our measurement listed in Table 2. The model value of the autocorrelation is shown as the grey line with the shaded region representing the diagonal elements from the corresponding simulated covariance matrix.

The thick line in the middle is the flux from the quasar while the other four thinner lines are from the simulation. The visual similarities between the observed data and the forward modelled data highlights the ability of our forward modelling methods to mimic realistic data. The remaining figures all show data and simulations at $z = 5.6$ because this redshift has the maximal observed sightlines with $N_{\rm los} = 34$ and there is a nearby measurement of $\lambda_{\rm mfp}$ at $z = 5.6$ by Zhu et al. (2023). Note that $N_{\rm los}$ does not affect the convergence of our simulated models but it determines the convergence of the bootstrap covariance matrix estimate which we will compare to later in the section.

5.3 Modelled autocorrelation function

We then computed the autocorrelation function of these forward modelled skewers in the same way as the actual data, with equation (2), for each copy of the skewer. We used the same mask from the observed quasar when computing the autocorrelation function. This includes the DLA mask as described in Section 3.3. In the observations, the DLA mask corresponds to regions in the spectra where the transmission is low. However, for the simulations the DLA mask corresponds to random parts of the spectra. We choose to include this part of the mask for the simulation data in order to keep the number of pixel pairs used per quasar sightline the same between simulations and observations. A discussion on the effect of the DLA mask on the measured autocorrelation function can be found in Appendix B.

To create a mock data set, we randomly selected $N_{\rm los}$ quasars from the 1000 forward modelled skewers without replacement. We then assigned each of the randomly selected skewers one of each of the $N_{\rm los}$ quasars, so each mock data set had exactly one skewer forward modelled with the properties of each quasar. The value of the autocorrelation function from the mock data set, ξ_i, is then the weighted average of the autocorrelation function from these $N_{\rm los}$ forward modelled skewers, where the weights are the number of pixels pairs in each bin of the autocorrelation function. We defined the model value of the autocorrelation function, $\boldsymbol{\xi}_{\rm model} = \boldsymbol{\xi}_{\rm model}(\lambda_{\rm mfp}, \langle F \rangle)$, to be the weighted average of the autocorrelation functions from all $N_{\rm los} \times N_{\rm skewers}$ skewers generated. The simulated covariance matrices, $\boldsymbol{\Sigma}_{\rm sim}$, are computed for each $\lambda_{\rm mfp}$ and $\langle F \rangle$ values from $N_{\rm mocks}$ mock data sets in the following way:

$$\boldsymbol{\Sigma}_{\rm sim}(\boldsymbol{\xi}_{\rm model}) = \frac{1}{N_{\rm mocks}} \sum_{i=1}^{N_{\rm mocks}} (\boldsymbol{\xi}_i - \boldsymbol{\xi}_{\rm model})(\boldsymbol{\xi}_i - \boldsymbol{\xi}_{\rm model})^{\rm T}. \quad (9)$$

Fig. 10 shows nine measurements of the autocorrelation function from nine different mock data sets generated from the simulations at $z = 5.6$ (coloured triangles). These mock measurements were generated from the $\lambda_{\rm mfp} = 20$ cMpc and $\langle F \rangle = 0.0483$ simulation, the closest $\lambda_{\rm mfp}$ value to the Zhu et al. (2023) measurement and the closest $\langle F \rangle$ value to our measurement listed in Table 2. This model value of the autocorrelation function is shown as the grey line where the grey shaded region shows the error from the diagonal of the simulated covariance matrix. The black points show the measured autocorrelation function at $z = 5.6$ with error bars from the bootstrap covariance matrix. This plot demonstrates that our forward modelling procedure leads to mock correlation function measurements that are visually similar to our actual measurement. This plot also shows that our measured autocorrelation function and the model with the value from Zhu et al. (2023) agree within $1\sigma_{\rm boot}$ for nearly all the points, though again these errors come from the diagonal of the covariance matrix only and therefore do not include information on the strong off-diagonal covariance between autocorrelation function bins. We discuss the comparison of our measured autocorrelation function and the measurements of Zhu et al. (2023) and Gaikwad et al. (2023) in Section 5.5.

5.4 Model based covariance matrices

Fig. 11 shows correlation matrices from the forward modelled data for six different parameter values at $z = 5.6$. The parameter values shown are $\lambda_{\rm mfp} = 5$, 15, 150 cMpc going down the rows and then $\langle F \rangle = 0.0303$, 0.0591 across the columns, both of which span the full range of parameter values available to us. Going from the left to the right column, we see that increasing the $\langle F \rangle$ weakly increases the off-diagonal values of the correlation matrices, however the effect going down the rows is much stronger. Going down the rows shows that an increase in $\lambda_{\rm mfp}$ decreases the off-diagonal values for the correlation matrix. This means that shorter $\lambda_{\rm mfp}$ models have more highly covariant bins in the autocorrelation function.

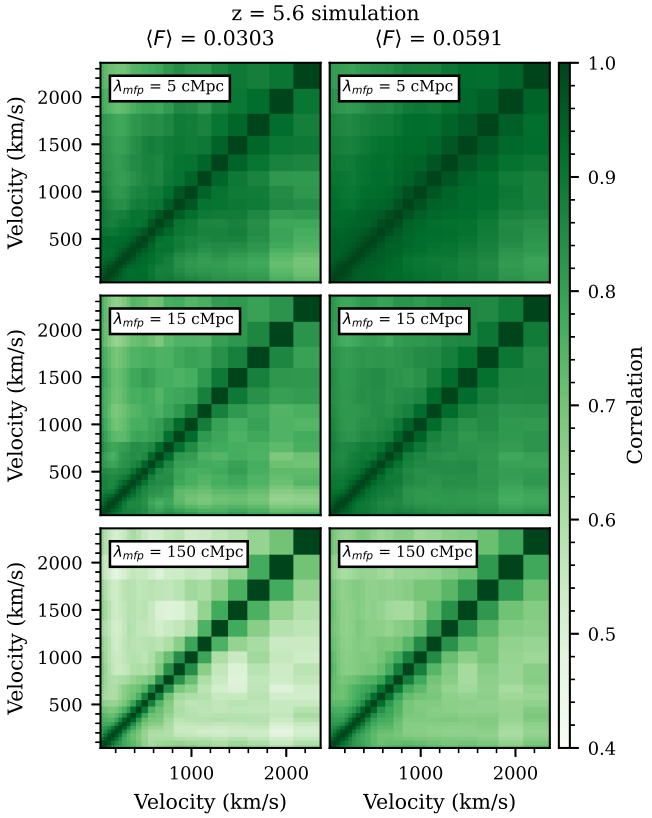

Figure 11. Correlation matrices for six different simulation model values at $z = 5.6$. These six covariance matrices come from the combination of $\lambda_{\rm mfp} = 5$, 15, 150 cMpc and $\langle F \rangle = 0.0303$, 0.0591 as labelled in the title of each subplot. These include the maximal and minimal $\lambda_{\rm mfp}$ and $\langle F \rangle$ values simulated at $z = 5.6$. This shows the model-dependence of the correlation (and thus covariance) matrices. Larger values of $\lambda_{\rm mfp}$ result in weaker off-diagonal correlation matrix values, as is seen going down the rows. Smaller $\langle F \rangle$ values also appear to cause weaker off-diagonal correlation matrix values (as seen when comparing the left and right columns) but this effect is weaker than the effect of $\lambda_{\rm mfp}$.

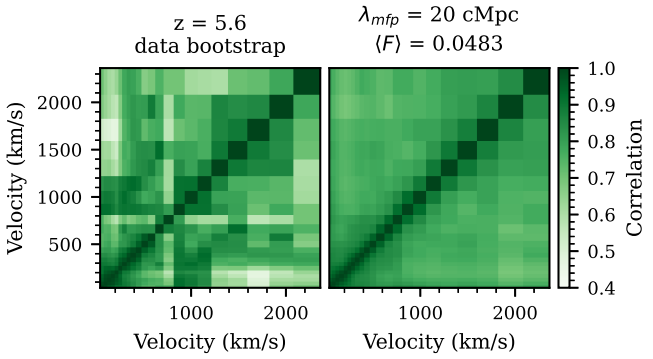

Figure 12. The correlation matrix computed via bootstrap re-sampling the data at $z = 5.6$ (left) and the simulated correlation matrix from the model with $\lambda_{\rm mfp} = 20$ cMpc and $\langle F \rangle = 0.0483$ (right). This model was chosen as the model with the closest $\lambda_{\rm mfp}$ value to the Zhu et al. (2023) measurement and the closest $\langle F \rangle$ value to our measurement. This bootstrap correlation matrix is also shown in Fig. 8 with a different colour bar and has been reproduced here with the colour bar used in Fig. 11, to more easily compare the values of the correlation matrix from data to the simulated examples. The bootstrap covariance matrix is noisy due to the limited data available, though this redshift was selected as the bin with the maximal value of $N_{\rm los} = 34$.

To compare a bootstrap covariance matrix from the data with the forward modelled covariance matrices, Fig. 12 shows the bootstrap correlation matrix at $z = 5.6$ with the same colour bar as Fig. 8. Additionally, Fig. 12 shows the simulated correlation matrix for the $\lambda_{\rm mfp} = 20$ cMpc and $\langle F \rangle = 0.0483$ model to directly compare to the bootstrapped matrix. Again, this is the model with the closest $\lambda_{\rm mfp}$ value to the Zhu et al. (2023) measurement and the closest $\langle F \rangle$ value to our measurement. The bootstrap covariance matrix is still quite noisy due to the limited data available so it is difficult to determine the best matching simulated covariance matrix. The bootstrap correlation matrix has regions of high off-diagonal values, such as $1200\,{\rm km\,s^{-1}} < v < 2000\,{\rm km\,s^{-1}}$ as well as individual pixels with relatively small off-diagonal values, such as the combination of $v = 60\,{\rm km\,s^{-1}}$ and $v = 1702\,{\rm km\,s^{-1}}$. This potentially suggests additional structure in the bootstrap covariance matrix compared to the simulated covariance data, but these fluctuations appear consistent with the noise.

As can be seen in Fig. 11, the correlation matrices, and therefore the covariance matrices, strongly depend on the model value of $\lambda_{\rm mfp}$. For this reason, when attempting to fit this data to a model, we would be fitting both the measured autocorrelation function as well as the covariance structure between the bins. While the amplitude of the correlation function might favour one combination of model parameters, it is conceivable that the level of fluctuations between two correlated correlation function bins, which is quantified by the covariance matrix, could favour a different combination of parameters. For this reason, fitting these models to our measurements is quite challenging and we leave this discussion for future work.

5.5 Comparison to previous work

We model the autocorrelation function at any value of $\lambda_{\rm mfp}$ and $\langle F \rangle$ via nearest grid-point emulation from our initial grid of values. Therefore, we can compare our autocorrelation function measurement to the models with the $\lambda_{\rm mfp}$ values measured in Gaikwad et al. (2023) and Zhu et al. (2023) which updated the measurements of Becker et al. (2021). Since we need to specify both $\lambda_{\rm mfp}$ and $\langle F \rangle$ to get our models, we use the measured $\langle F \rangle$ from this work to get the models representing the $\lambda_{\rm mfp}$ values from the corresponding alternative measurements. Fig. 13 has ten panels, each of which has one of our measured $\frac{\xi_{F-\langle F \rangle^2}}{\langle F \rangle^2}$ values (shown as the black points) at a given z. We have chosen to show $\frac{\xi_{F-\langle F \rangle^2}}{\langle F \rangle^2}$ instead of the regular autocorrelation function because we have to use the nearest grid point on a coarse $\langle F \rangle$ grid which could be quite far from the measured $\langle F \rangle$ value. This would have a large effect on the autocorrelation function value and a smaller effect on $\frac{\xi_{F-\langle F \rangle^2}}{\langle F \rangle^2}$.

Gaikwad et al. (2023) measured $\lambda_{\rm mfp}$ at each of these redshifts and so each panel has our model with their $\lambda_{\rm mfp}$ values (green lines). Zhu et al. (2023) has measured $\lambda_{\rm mfp}$ for $z = 5.08$, 5.31, 5.65, and 5.93. We show the models for the measured $\lambda_{\rm mfp}$ values from Zhu et al. (2023) in the $z = 5.1$, 5.3, (5.6 and 5.7), and 6.0 panels respectively (red lines). Finally, we also show the model for $\lambda_{\rm mfp} = 150$ cMpc, our most uniform UVB (blue line).

Making a quantitative comparison of these models with the measured autocorrelation function is difficult due to the expected large off-diagonal values of the covariance matrix as well as the noise in the bootstrap covariance matrices as shown in Fig. 8. For this reason, we leave detailed quantitative comparisons and fitting

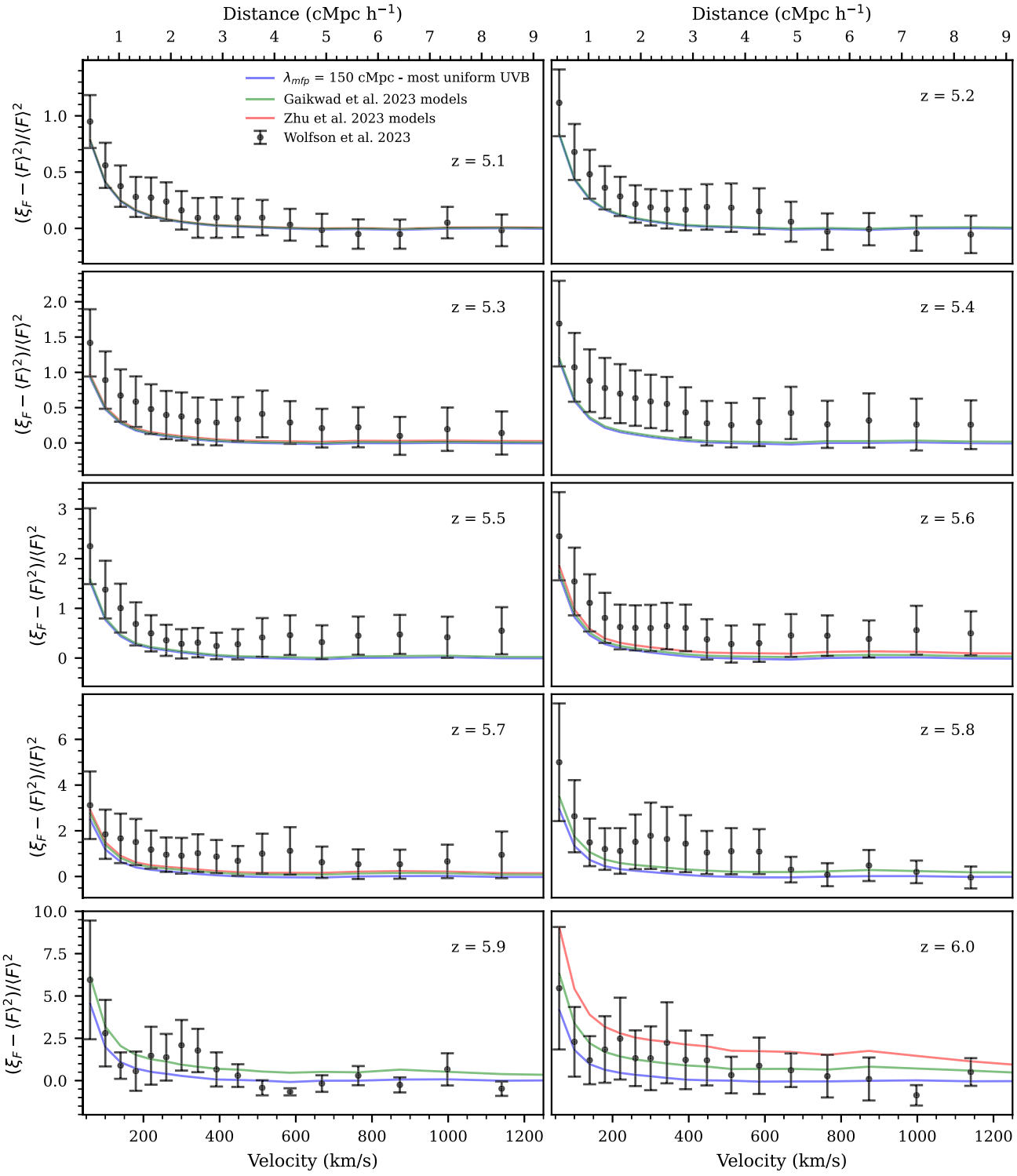

Figure 13. The autocorrelation function of Ly α transmission normalized and shifted by the mean transmission, $\langle F \rangle$, in ten redshift bins measured in this work. Gaikwad et al. (2023) measured $\lambda_{\rm mfp}$ at each of these redshifts and so each panel has our model with their $\lambda_{\rm mfp}$ values (green lines). Zhu et al. (2023) has measured $\lambda_{\rm mfp}$ for $z = 5.08$, 5.31, 5.65, and 5.93. We show the model models for the measured $\lambda_{\rm mfp}$ values from Zhu et al. (2023) in the $z = 5.1$, 5.3, (5.6 and 5.7), and 6.0 panels, respectively (red lines). The model for a uniform UVB value (blue line) is also shown as a comparison.

for future work. It is interesting to note that our measurements fall above the models from Zhu et al. (2023), Gaikwad et al. (2023), and $\lambda_{\rm mfp} = 150$ cMpc for $z < 5.8$. Also note that models from Zhu et al. (2023) and Gaikwad et al. (2023) show a small boost over the most uniform UVB model for $z < 5.8$.

6 CONCLUSIONS

In this work, we have measured the autocorrelation function of the Ly α forest flux from the extended XQR-30 data set in 10 redshift bins, $5.1 \leq z \leq 6.0$. This is the first measurement of

the autocorrelation function of the Ly α forest at these redshifts. Our final assembled data set includes 36 $z > 5.7$ quasars with SNR >20 per spectral pixel. This data set was analysed while fully accounting for the error from continuum reconstruction, instrumentation, and contamination from DLAs. We measured the average transmission, $\langle F \rangle$, from this data and found good agreement with previous work. We found that the boost in the autocorrelation function on the smallest scales increases when increasing z, which may suggest a decrease in $\lambda_{\rm mfp}$. We additionally measured covariance matrices of the autocorrelation function by bootstrap re-sampling the available data. The convergence of these matrices was hindered by noise from the limited number of sightlines and low transmission, especially for the highest redshift bins, $z \geq 5.8$. The autocorrelation function measurements as well as the bootstrap covariance matrices are available to download online at `https://github.com/mollywolfson/lya_autocorr/`. Note that this is the best available sample of quasars at these redshifts in terms of size, resolution, and SNR. Increasing the number of observations, especially at $z \gtrsim 6.5$, with the same quality would greatly improve these measurements.

In addition, we introduced Ly α forest simulations with a fluctuating UVB model described by $\lambda_{\rm mfp}$. This comparison indicates preliminary agreement between these models and our measurements. We found that the covariance matrices produced from the simulations had a strong dependence on $\lambda_{\rm mfp}$. In order to fit these models to our data, we would need to use an estimate of the covariance matrix for the bins of the autocorrelation function. In this work, we have presented two options for this covariance matrix: the bootstrap estimate, $\mathbf{\Sigma}_{\rm boot}$, and the simulation covariance matrices, $\mathbf{\Sigma}_{\rm sim}$. Ideally we would like to use $\mathbf{\Sigma}_{\rm boot}$ when fitting, however as seen in Fig. 8, these covariance matrices are quite noisy and non-converged. Therefore, we could hope to use $\mathbf{\Sigma}_{\rm sim}$, where the off-diagonal structure depends strongly on the value of $\lambda_{\rm mfp}$. This dependence of $\mathbf{\Sigma}_{\rm sim}$ on $\lambda_{\rm mfp}$ means that fitting the models to the data would require fitting both the mean line as well as this covariance structure, which is subtle. Thus, additional work is necessary to get robust measurements of $\lambda_{\rm mfp}$, which we leave to the future. We did show a preliminary comparison of our measured auto-correlation function to models with the $\lambda_{\rm mfp}$ values measured by Gaikwad et al. (2023) and Zhu et al. (2023), leaving a quantitative comparison of these results to future work.

With this work we have included a link to a Git repository with the code necessary to measure the autocorrelation function from any set of simulation skewers. This will allow other simulation groups to compare the autocorrelation function from their simulations to our measured autocorrelation function and thus foster more work on this statistic.

Future work to get a robust measurement of $\lambda_{\rm mfp}$ from the Ly α forest autocorrelation function include further considerations in the modelling methods. The Davies & Furlanetto (2016) method to generate $\Gamma_{\rm H\,I}$ for various $\lambda_{\rm mfp}$ assumes a fixed source model. Other source model choices could impact the fluctuations in $\Gamma_{\rm H\,I}$ seen at a fixed $\lambda_{\rm mfp}$ value, and thus bias measurements from observation data when compared with these models. Additionally, rare bright sources could cause boosts in the autocorrelation function for individual sightlines that are not modelled in our simulations. We leave a detailed investigation into these effects on the autocorrelation function models and covariance matrices to future work.

Note that in order to generate UVB fluctuations due to different $\lambda_{\rm mfp}$ values that matched the density field of our NYX simulation, we also generated UVB fluctuations in a 100 cMpc h^{-1} box. Wolfson et al. (2023b) found that using a 40 cMpc h^{-1} box to generate UVB fluctuations significantly reduced the autocorrelation function on all scales when compared to a 512 cMpc box. Future work would be needed to understand the effect of the box size on any measured $\lambda_{\rm mfp}$ from the autocorrelation function with a 100 cMpc h^{-1} UVB box.

Additionally, this work ignored the effect of inhomogeneous reionization beyond a fluctuating UVB. It is expected that a patchy, inhomogeneous reionization process would have other physical effects, such as additional fluctuations in the thermal state of the IGM. We leave an exploration of the effect of the temperature of the IGM on the Ly α forest flux autocorrelation function, including the effect of temperature fluctuations, to a future work.

Overall, this first measurement of the $z > 5$ Ly α forest flux autocorrelation functions opens up an exciting new way to measure $\lambda_{\rm mfp}$ at the tail-end of reionization.

ACKNOWLEDGEMENTS

We acknowledge helpful conversations with the entire XQR-30 collaboration as well as the ENIGMA group at UC Santa Barbara and Leiden University. JFH acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 885301) and from the National Science Foundation under Grant No. 1816006. MH acknowledges support by STFC (grant number ST/N000927/1). GK is partly supported by the Department of Atomic Energy (Government of India) research project with Project Identification Number RTI 4002, and by the Max Planck Society through a Max Planck Partner Group.

Based on observations collected at the European 761 Southern Observatory under ESO programme 1103.A0817. This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

DATA AVAILABILITY

The XQR-30 spectra and associated meta-data have been made public in D'Odorico et al. (2023). The measurements of the autocorrelation function are available to download online at `https://github.com/mollywolfson/lya_autocorr/` along with functions to forward model any simulation skewers with the properties of our data set.

The simulation data analysed in this article will be shared on reasonable request to the corresponding author.

REFERENCES

Almgren A. S., Bell J. B., Lijewski M. J., Lukić Z., Van Andel E., 2013, ApJ, 765, 39
Bañados E. et al., 2014, AJ, 148, 14
Bañados E. et al., 2016, ApJS, 227, 11
Bañados E. et al., 2019, ApJ, 885, 59
Becker G. D., Sargent W. L. W., Rauch M., 2004, ApJ, 613, 61
Becker G. D., Rauch M., Sargent W. L. W., 2009, ApJ, 698, 1010
Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92
Becker G. D. et al., 2019, ApJ, 883, 163
Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853
Bischetti M. et al., 2022, Nature, 605, 244
Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, ApJ, 872, 101
Bosman S. E. I., 2021, preprint (arXiv:2108.12446)
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bosman S. E. I., Ďurovčíková D., Davies F. B., Eilers A.-C., 2021, MNRAS, 503, 2077
Bosman S. E. I. et al., 2022, MNRAS, 514, 55
Bouwens R. J. et al., 2015, ApJ, 803, 34
Cain C., D'Aloisio A., Gangolli N., Becker G. D., 2021, ApJ, 917, L37
Carnall A. C. et al., 2015, MNRAS, 451, L16
Carswell R. F., Whelan J. A. J., Smith M. G., Boksenberg A., Tytler D., 1982, MNRAS, 198, 91
Chehade B. et al., 2018, MNRAS, 478, 1649
Cooper T. J., Simcoe R. A., Cooksey K. L., Bordoloi R., Miller D. R., Furesz G., Turner M. L., Bañados E., 2019, ApJ, 882, 77
Croft R. A. C., 2004, ApJ, 610, 642
D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018, MNRAS, 473, 560
D'Odorico V. et al., 2006, MNRAS, 372, 1333
D'Odorico V. et al., 2018, ApJ, 863, L29
D'Odorico V. et al., 2023, MNRAS, 523, 1399
Dall'Aglio A., Wisotzki L., Worseck G., 2008, A&A, 491, 465
Dall'Aglio A., Wisotzki L., Worseck G., 2009, preprint (arXiv:0906.1484)
Davies F. B., 2020, MNRAS, 494, 2937
Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328
Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018a, ApJ, 855, 106
Davies F. B. et al., 2018b, ApJ, 864, 142
Davies F. B. et al., 2018c, ApJ, 864, 143
Davies F. B., Bosman S. E. I., Furlanetto S. R., Becker G. D., D'Aloisio A., 2021, ApJ, 918, L35
Davies R. L. et al., 2023, MNRAS, 521, 289
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
De Rosa G., Decarli R., Walter F., Fan X., Jiang L., Kurk J., Pasquali A., Rix H. W., 2011, ApJ, 739, 56
Decarli R. et al., 2018, ApJ, 854, 97
Ďurovčíková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A., 2020, MNRAS, 493, 4256
Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53
Eilers A.-C. et al., 2021, ApJ, 914, 74
Fan X. et al., 2001, AJ, 122, 2833
Fan X. et al., 2006, AJ, 132, 117
Farina E. P. et al., 2019, ApJ, 887, 196
Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 681, 831
Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., 1992, ApJ, 398, 476
Fumagalli M., O'Meara J. M., Prochaska J. X., Worseck G., 2013, ApJ, 775, 78
Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031
Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2021, MNRAS, 506, 4389
Gaikwad P. et al., 2023, MNRAS 525 4093
Garzilli A., Boyarsky A., Ruchayskiy O., 2017, Phys. Lett. B, 773, 258
Gnedin N. Y., 2000, ApJ, 535, 530
Gnedin N. Y., Fan X., 2006, ApJ, 648, 1
Gnedin N. Y., Madau P., 2022, Living Rev. Comput. Astrophys., 8, 3
Gontcho A Gontcho S., Miralda-Escudé J., Busca N. G., 2014, MNRAS, 442, 187
Haardt F., Madau P., 2012, ApJ, 746, 125
Horne K., 1986, PASP, 98, 609
Iršič V. et al., 2017, Phys. Rev. D, 96, 023522
Jiang L., Fan X., Vestergaard M., Kurk J. D., Walter F., Kelly B. C., Strauss M. A., 2007, AJ, 134, 1150
Jiang L., McGreer I. D., Fan X., Bian F., Cai Z., Clément B., Wang R., Fan Z., 2015, AJ, 149, 188
Jiang L. et al., 2016, ApJ, 833, 222
Jung I. et al., 2020, ApJ, 904, 144
Kashino D., Lilly S. J., Shibuya T., Ouchi M., Kashikawa N., 2020, ApJ, 888, 6
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020a, MNRAS, 491, 1736
Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020b, MNRAS, 497, 906
Kelson D. D., 2003, PASP, 115, 688
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lukić Z., Stark C. W., Nugent P., White M., Meiksin A. A., Almgren A., 2015, MNRAS, 446, 3697
Mazzucchelli C. et al., 2017, ApJ, 849, 91
McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, ApJ, 543, 1
McDonald P., Seljak U., Cen R., Bode P., Ostriker J. P., 2005, MNRAS, 360, 1471
Meiksin A., McQuinn M., 2019, MNRAS, 482, 4777
Meiksin A., White M., 2004, MNRAS, 350, 1107
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., 2009, MNRAS, 400, 1461
Morales A. M., Mason C. A., Bruton S., Gronke M., Haardt F., Scarlata C., 2021, ApJ, 919, 120
Mortlock D. J. et al., 2009, A&A, 505, 97
Nasir F., D'Aloisio A., 2020, MNRAS, 494, 3080
Oñorbe J., Davies F. B., Lukić Z., Z., Hennawi J. F., Sorini D., 2019, MNRAS, 486, 4075
O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137
Pâris I. et al., 2011, A&A, 530, A50
Park H., Shapiro P. R., Choi J.-h., Yoshida N., Hirano S., Ahn K., 2016, ApJ, 831, 86
Planck Collaboration VI 2020, A&A, 641, A6
Pontzen A., 2014, Phys. Rev. D, 89, 083010
Pontzen A., Bird S., Peiris H., Verde L., 2014, ApJ, 792, L34
Prochaska J. X., Worseck G., O'Meara J. M., 2009, ApJ, 705, L113
Rafelski M., Wolfe A. M., Prochaska J. X., Neeleman M., Mendez A. J., 2012, ApJ, 755, 89
Rauch M., 1998, ARA&A, 36, 267
Reed S. L. et al., 2017, MNRAS, 468, 4702
Rollinde E., Petitjean P., Pichon C., Colombi S., Aracil B., D'Odorico V., Haehnelt M. G., 2003, MNRAS, 341, 1279
Romano M., Grazian A., Giallongo E., Cristiani S., Fontanot F., Boutsia K., Fiore F., Menci N., 2019, A&A, 632, A45
Rudie G. C., Steidel C. C., Shapley A. E., Pettini M., 2013, ApJ, 769, 146
Shapiro P. R., Iliev I. T., Raga A. C., 2004, MNRAS, 348, 753
Sobacchi E., Mesinger A., 2014, MNRAS, 440, 1662
Sodini A. et al., 2024, preprint (arXiv:2404.10722)
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448
Suzuki N., Tytler D., Kirkman D., O'Meara J. M., Lubin D., 2005, ApJ, 618, 592
Venemans B. P. et al., 2015, MNRAS, 453, 2259
Venemans B. P. et al., 2020, ApJ, 904, 130
Vernet J. et al., 2011, A&A, 536, A105
Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502
Walther M., Oñorbe J., Hennawi J. F., Lukić Z., 2019, ApJ, 872, 13
Wang R. et al., 2010, ApJ, 714, 699
Wang R. et al., 2013, ApJ, 773, 44

Wang F. et al., 2016, ApJ, 819, 24
Wang F. et al., 2019, ApJ, 884, 30
Wang F. et al., 2021, ApJ, 908, 53
Willott C. J. et al., 2007, AJ, 134, 2435
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Wolfson M., Hennawi J. F., Davies F. B., Lukić Z., Oñorbe J., 2023a, preprint (arXiv:2309.05647)
Wolfson M., Hennawi J. F., Davies F. B., Oñorbe J., 2023b, MNRAS, 521, 4056
Worseck G. et al., 2014, MNRAS, 445, 1745
Wu X.-B. et al., 2015, Nature, 518, 512
Wyithe J. S. B., Bolton J. S., Haehnelt M. G., 2008, MNRAS, 383, 691
Yang J. et al., 2020, ApJ, 904, 26
Yip C. W. et al., 2004, AJ, 128, 2603
Young P. J., Sargent W. L. W., Boksenberg A., Carswell R. F., Whelan J. A. J., 1979, ApJ, 229, 891
Zhu Y. et al., 2021, ApJ, 923, 223
Zhu Y. et al., 2023, ApJ 955 115
Zuo L., 1992a, MNRAS, 258, 45
Zuo L., 1992b, MNRAS, 258, 36

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: CONTINUUM UNCERTAINTY MODELLING EFFECT

Fig. A1 quantifies the difference in the autocorrelation models calculated from forward-modelled skewers with or without continuum uncertainty multiplied in, as described in Section 5.2. The first and third panels show the autocorrelation function from the simulations with (solid line) and without (dashed line) modelling continuum uncertainty at $z = 5.1$ and 6. The different colours represent different parameter values of $\lambda_{\rm mfp}$ and $\langle F \rangle$ used. The second and fourth panels show the relative difference in percent, defined as:

$$\frac{\xi_{\rm cont} - \xi_{\rm no\ cont}}{\xi_{\rm no\ cont}}. \tag{A1}$$

At $z = 5.1$ there is a < 1 per cent of a difference between the autocorrelation models with and without the continuum error. At $z = 6.0$ there is a larger difference between the models where the difference is < 8 per cent for all the parameter values. However, the effect is most noticeable when $\langle F \rangle$, and hence the autocorrelation function which goes as $\langle F \rangle^2$, is quite small. For the other $\langle F \rangle$ value at this redshift the error is < 2 per cent. These values are typically positive because of the bias in the continuum reconstruction as seen in fig. 2 of Bosman et al. (2022).

We computed the difference in our measured autocorrelation function at all z with and without continuum error. The difference in the measured data ranges from at most 0.4 per cent to 1.8 per cent with a stronger effect at the highest redshifts.

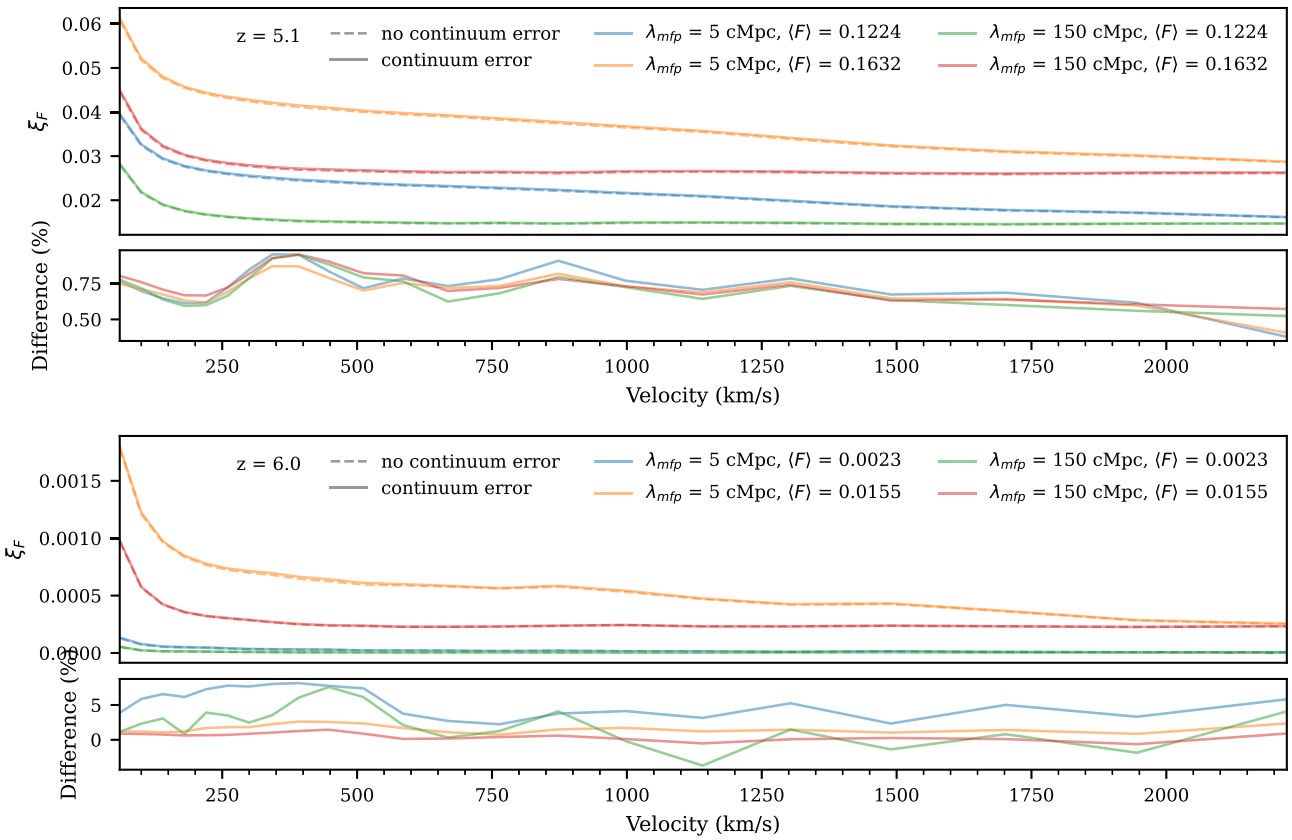

Figure A1. The first and third panels show the autocorrelation function from the simulations with (solid line) and without (dashed line) modelling continuum uncertainty at redshifts of 5.1 and 6. The different colours represent different parameter values of $\lambda_{\rm mfp}$ and $\langle F \rangle$ used. The second and fourth panels show the relative difference between these lines defined by equation (A1), in per cent.

APPENDIX B: DLA MODELLING EFFECT

In order to investigate how the DLA mask that was described in Section 3.3 we compute the measured autocorrelation functions without using this mask. This is shown for all redshifts in Fig. B1 in red. The original measurement including this mask is shown in black.

The measurement at $z = 5.2$ is not impacted at all by the DLA mask as no sightline has a detected DLA in this redshift range. Otherwise, for most scales at most redshifts ignoring the DLA mask reduces the autocorrelation function values. This follows as generally the regions masked in our procedure are regions with high absorption.

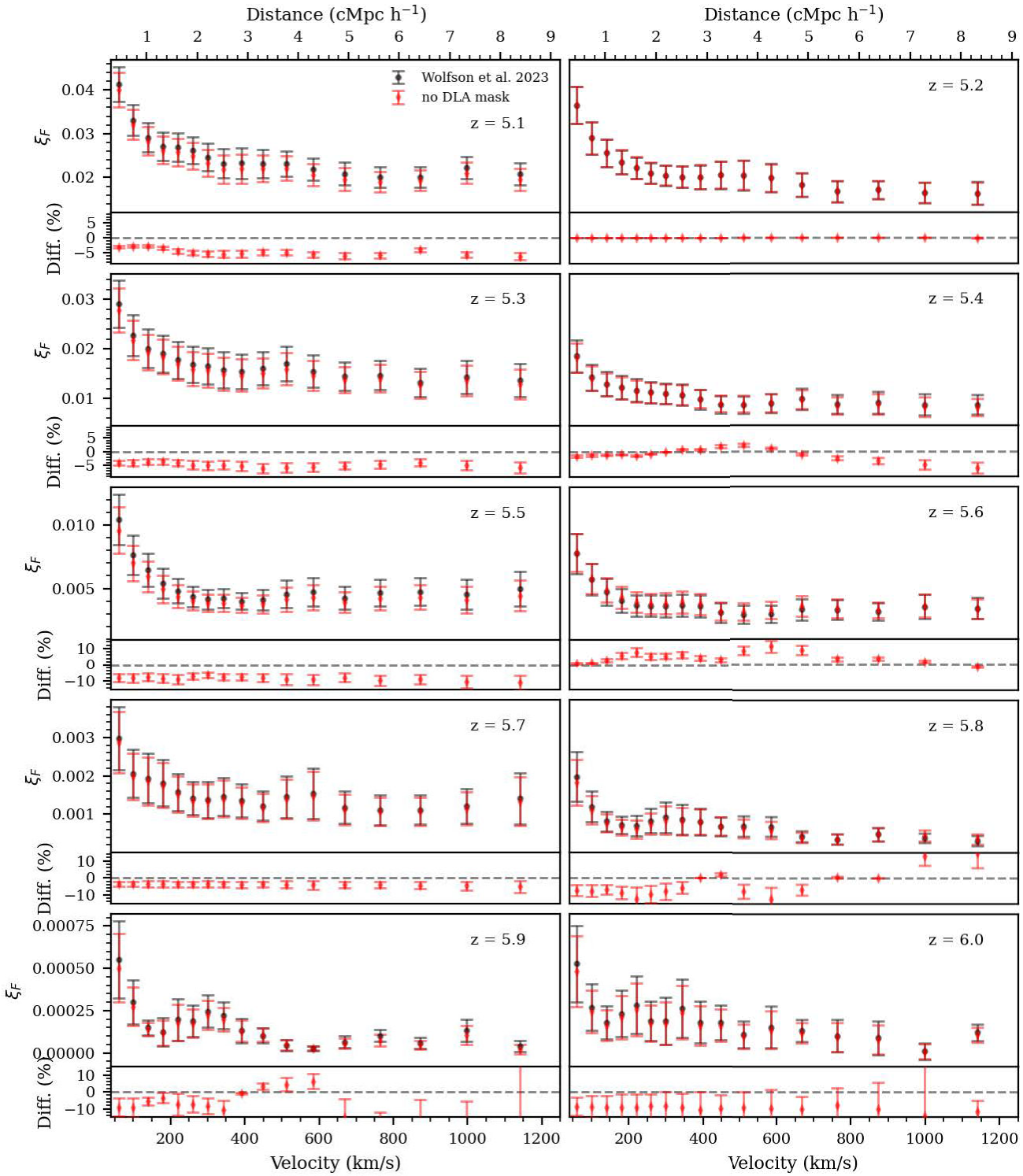

Figure B1. The black points show autocorrelation function of Ly α transmission in ten redshift bins measured in this work. The red points show the measured autocorrelation function of the Ly α transmission when ignoring the masks for the DLAs as described in Section 3.3. In general, ignoring the DLA mask decreases the autocorrelation function values.

APPENDIX C: MEAN UVB FOR FIXED MEAN FLUX

We computed the $\langle\Gamma_{\rm H\,I}\rangle$ values that arose in our simulations for given values of $\langle F\rangle$ and $\lambda_{\rm mfp}$. These are shown in Fig. C1 for three fixed values of $\langle F\rangle$. Each fixed $\langle F\rangle$ value is shown in a different colour. These lines demonstrate that increasing $\langle\Gamma_{\rm H\,I}\rangle$ is required in order to maintain a given $\langle F\rangle$ when decreasing $\lambda_{\rm mfp}$. This follows from the effect of small $\lambda_{\rm mfp}$ on $\langle F\rangle$. Consider fig. 3 from Wolfson et al. (2023b) which shows the flux along the line of sight of a skewer for different $\lambda_{\rm mfp}$ values. Small $\lambda_{\rm mfp}$ causes there to be large regions of the Ly α forest with no transmitted flux. Therefore, higher flux values in regions where there is some transmitted flux are required to match the average to models with more areas of transmitted flux (in this case larger $\lambda_{\rm mfp}$ models), meaning a larger $\langle\Gamma_{\rm H\,I}\rangle$. This may seem in conflict to the assumption that $\lambda \propto \Gamma_{\rm H\,I}^{2/3}$, however this is a local relation and the overall average has additional influences.

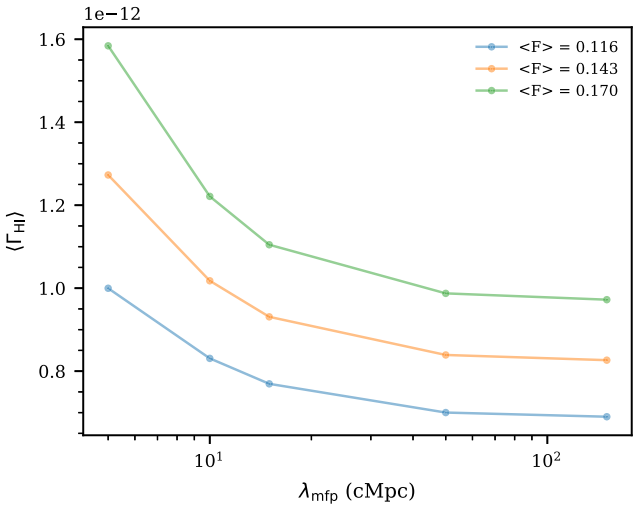

Figure C1. The $\langle\Gamma_{\rm H\,I}\rangle$ values as a function of $\lambda_{\rm mfp}$ for fixed $\langle F\rangle$ at $z = 5.1$. There are three values of $\langle F\rangle$ shown in three colours. This demonstrates that increasing $\langle\Gamma_{\rm H\,I}\rangle$ if required when decreasing $\lambda_{\rm mfp}$ in order to maintain a given $\langle F\rangle$.

This paper has been typeset from a TEX/LATEX file prepared by the author.