Universiteit Leiden
The Netherlands

Forecasting constraints on the mean free path of ionizing photons at z ≥ 5.4 from the Lyman-α forest flux autocorrelation function

Wolfson, M.; Hennawi, J.F.; Davies, F.B.; Oñorbe, J.

Forecasting constraints on the mean free path of ionizing photons at $z \geq 5.4$ from the Lyman-α forest flux autocorrelation function

Molly Wolfson [ORCID],[1]★ Joseph F. Hennawi [ORCID],[1,2] Frederick B. Davies [ORCID][3] and Jose Oñorbe [ORCID][4]

[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]*Facultad de Físicas, Universidad de Sevilla, Avda. Reina Mercedes s/n, Campus de Reina Mercedes, E-41012 Sevilla, Spain*

ABSTRACT

Fluctuations in Lyman-α (Lyα) forest transmission towards high-z quasars are partially sourced from spatial fluctuations in the ultraviolet background, the level of which are set by the mean free path of ionizing photons (λ_{mfp}). The autocorrelation function of Lyα forest flux characterizes the strength and scale of transmission fluctuations and, as we show, is thus sensitive to λ_{mfp}. Recent measurements at $z \sim 6$ suggest a rapid evolution of λ_{mfp} at $z > 5.0$ which would leave a signature in the evolution of the autocorrelation function. For this forecast, we model mock Lyα forest data with properties similar to the XQR-30 extended data set at $5.4 \leq z \leq 6.0$. At each z, we investigate 100 mock data sets and an ideal case where mock data matches model values of the autocorrelation function. For ideal data with $\lambda_{\mathrm{mfp}} = 9.0$ cMpc at $z = 6.0$, we recover $\lambda_{\mathrm{mfp}} = 12^{+6}_{-3}$ cMpc. This precision is comparable to direct measurements of λ_{mfp} from the stacking of quasar spectra beyond the Lyman limit. Hypothetical high-resolution data leads to a ~ 40 per cent reduction in the error bars over all z. The distribution of mock values of the autocorrelation function in this work is highly non-Gaussian for high-z, which should caution work with other statistics of the high-z Lyα forest against making this assumption. We use a rigorous statistical method to pass an inference test, however future work on non-Gaussian methods will enable higher precision measurements.

Key words: *(cosmology:)* dark ages, reionization, first stars – *(galaxies:)* intergalactic medium – *(galaxies:)* quasars: absorption lines – methods: statistical.

1. INTRODUCTION

The neutral hydrogen in the intergalactic medium (IGM) was reionized by the first luminous sources during the epoch of reionization. This period was one of the most dramatic changes in the history of the Universe. Current Planck constraints from the cosmic microwave background put the midpoint of reionization at $z_{\mathrm{re}} = 7.7 \pm 0.7$ (Planck Collaboration VI 2020). There have also been multiple measurements that suggest reionization was not completed until after $z \leq 6$ (Fan et al. 2006; Becker et al. 2015, 2018; Bosman et al. 2018, 2022; Eilers, Davies & Hennawi 2018; Boera et al. 2019; Jung et al. 2020; Kashino et al. 2020; Yang et al. 2020; Morales et al. 2021). However, much is still unknown about this process such as the exact timing, the impact on the thermal state of the IGM, the driving sources, and the number of photons that must be produced to complete reionization.

Characterizing the IGM both during and immediately after reionization will give vital information to answer these remaining questions. Of particular interest is the average distance that the ionizing photons travel through the IGM before interacting with its neutral hydrogen – also known as the mean free path of ionizing

photons, λ_{mfp}. The end of reionization results in a rapid increase in λ_{mfp} as the initially isolated regions of ionized hydrogen overlap to form a mostly ionized universe (Gnedin 2000; Gnedin & Fan 2006; Wyithe, Bolton & Haehnelt 2008; D'Aloisio et al. 2018; Kulkarni et al. 2019; Keating et al. 2020a,b; Nasir & D'Aloisio 2020; Cain et al. 2021; Gnedin & Madau 2022). Detecting this rapid increase is therefore a clear signal of the end of reionization.

Direct measurements of λ_{mfp} at $z \leq 5.2$ have been achieved from stacked quasar spectra (Prochaska, Worseck & O'Meara 2009; Fumagalli et al. 2013; O'Meara et al. 2013; Worseck et al. 2014). Using a similar method, Becker et al. (2021) recently reported measurements of $\lambda_{\mathrm{mfp}} = 9.09^{+1.62}_{-1.28}$ proper Mpc at $z = 5.1$ and $\lambda_{\mathrm{mfp}} = 0.75^{+0.65}_{-0.45}$ proper Mpc at $z = 6$. This value at $z = 6$ is significantly smaller than extrapolations from previous lower z measurements (Worseck et al. 2014), causes tension with measurements of the ionizing output from galaxies (Cain et al. 2021; Davies et al. 2021), and also suggests a roughly 12-fold increase in λ_{mfp} between $z = 6$ and $z = 5.1$, potentially signalling the end of reionization. An alternative method presented in Bosman (2021) used lower limits on individual free paths towards high-z sources to place a 2σ limit of $\lambda_{\mathrm{mfp}} > 0.31$ 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). Additional independent methods of measuring λ_{mfp} are necessary to verify these measurements.

★ E-mail: mawolfson@ucsb.edu

Of particular interest are methods that can be used at several redshift bins at $z > 5$ in order to study the evolution of λ_{mfp} in finer detail.

In this paper, we investigate using the autocorrelation function of Ly α forest flux in high-z quasar sightlines to constrain λ_{mfp}. The Ly α opacity, $\tau_{\mathrm{Ly}\alpha}$, is related to λ_{mfp} via $\tau_{\mathrm{Ly}\alpha} = n_{\mathrm{HI}}\sigma_{\mathrm{Ly}\alpha} \propto 1/\Gamma_{\mathrm{HI}} \propto 1/\lambda_{\mathrm{mfp}}^{\alpha}$ where α is typically between 3/2 and 2 (see e.g. Rauch 1998; Haardt & Madau 2012). Additionally, during reionization the existence of significant neutral hydrogen in the IGM will cause a short mean free path value to also result in large spatial fluctuations in the ultraviolet background (UVB). This is because, during reionization ionizing photons are produced from the first sources and then quickly absorbed by the remaining neutral hydrogen. Thus there are large values of the UVB where the photons are produced and very small values where neutral hydrogen remains. If the mean free path is large, photons will travel further and effectively smooth the UVB (Mesinger & Furlanetto 2009). The positive fluctuations in the UVB on small scales that accompany a short mean free path would then boost the flux of the Ly α forest on small scales, which could then be detected in the autocorrelation function. Various previous studies have investigated the effect of large-scale variations in the UVB on the autocorrelation function and power spectrum of the Ly α 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). Our work is focused on determining if the effect of the fluctuating UVB on the autocorrelation function can lead to a constraint on λ_{mfp}.

While the power spectrum has been a more popular statistic used on the high-z Ly α forest to date (Boera et al. 2019; Walther et al. 2019; Gaikwad et al. 2021), the autocorrelation function has a few characteristics that make it easier to work with than the power spectrum. The two most obvious are the effect of noise and masking on the autocorrelation function when compared to the power spectrum. Astronomical spectrograph noise is expected to be white or uncorrelated. Uncorrelated noise only impacts the autocorrelation function at zero lag, since at all other lags the uncorrelated noise will average to zero. Therefore, by not measuring the autocorrelation at zero lag we have fully removed the effect of white noise. On the other hand, white noise is a constant positive value at all scales for the power spectrum. Thus the unknown noise level must be calculated and subtracted from power spectrum measurements which will add additional uncertainty to the final measurement. Additionally, real data often has regions of spectra that need to be removed from the quasar spectrum (e.g. for metal lines). Masking out these and other regions introduces a complicated window function to the power spectrum that must be corrected for (see e.g. Walther et al. 2019) and will again increasing the uncertainty in the measurement. The autocorrelation function does not require a similar correction since masking only result in fewer points in bins for certain lags.

The structure of this paper is as follows. We discuss our simulation data in Section 2. The autocorrelation function and our other statistical methods are described in Section 3. We then discuss our results in Section 4 and summarize in Section 5. Here, we also touch on how additional work on modelling λ_{mfp} in simulations as well as better statistical methods will improve these constraints.

2. SIMULATION DATA

2.1 Models

In this work, we use a simulation box run with Nyx code (Almgren et al. 2013). Nyx is a hydrodynamical simulation code that was designed for simulating the Ly α forest with updated physical rates from Lukić et al. (2015). The Nyx box has a size of $L_{\mathrm{box}} = 100$ 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 $z \sim 15$. We have two snapshots of this simulation at $z = 5.5$ and $z = 6$. In this work, we want to consider these models at seven redshifts: $5.4 \leq z \leq 6$ 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. We additionally used the $z = 6.0$ simulation snapshot to generate low-resolution skewers at $z = 5.7$ and found no significant change in our finally results, confirming that using the nearest simulation snapshot in this way is sufficient.

We also have separate boxes of fluctuating Γ_{HI} values generated with the seminumerical method of Davies & Furlanetto (2016). These boxes have a size $L_{\mathrm{box}} = 512$ cMpc and 128^3 pixels. We have one snapshot of these Γ_{HI} boxes at $z = 5.5$. To get the flux skewers used in this work, we combine random skewers of Γ_{HI} from these UVB boxes with the skewers from the Nyx box. The UVB boxes have a different resolution than the Nyx box, to generate a skewer of Γ_{HI} values we randomly selected a starting location and direction in the UVB box then linearly interpolated the $\log(\Gamma_{\mathrm{HI}})$ values onto the same length and resolution as the Nyx skewers.

The method of Davies & Furlanetto (2016) allows for a spatially varying mean free path generated from fluctuations in the density of the sources of ionizing radiation with $\lambda \propto \Gamma_{\mathrm{HI}}^{2/3} \Delta^{-1}$, for λ, the local mean free path, and Δ, the local matter density. These simulations are scaled such that the mean value $\langle \lambda \rangle = \lambda_{\mathrm{mfp}}$ as desired. A brief summary of the Davies & Furlanetto (2016) method is as follows. Cosmological initial conditions, independent of those from the 100-cMpc h^{-1} Nyx boxes, were generated for the 512 cMpc box and evolved to $z = 5.5$ via the Zel'dovich approximation (Zel'dovich 1970). Haloes were created via the approach of Mesinger & Furlanetto (2007) down to a minimum halo mass of $M_{\mathrm{min}} = 2 \times 10^9 \, \mathrm{M}_{\odot}$. The ionizing luminosity of galaxies corresponding to each halo were determined following two steps: First the UV luminosities of galaxies were assigned by abundance matching to the Bouwens et al. (2015) UV luminosity function and then the ionizing luminosity of each galaxy was assumed to be proportional to its UV luminosity where the constant of proportionality is left as a free parameter. The ionizing background radiation intensity, J_{ν}, is then computed by a radiative transfer algorithm. The photoionization rate, Γ_{HI}, is finally calculated by integrating over J_{ν}. For more details on the method see Davies & Furlanetto (2016), Davies et al. (2018), or Davies et al. (in preparation) where they also use this stitching procedure. Note that this method of generating UVB fluctuations ignores the effect of correlations between the baryon density in the Nyx boxes and the UVB. This is sufficient for the aims of this work but see Section 3.2 for a discussion on the effects of ignoring these correlations on the resulting autocorrelation function and therefore future measurements of λ_{mfp} from real data.

Example slices through the UVB boxes for four values of λ_{mfp} are shown in Fig. 1 with a lower cutoff of $\log(\Gamma_{\mathrm{HI}}/\langle \Gamma_{\mathrm{HI}} \rangle) = -1$ for visual purposes. The top-left box shows a slice of Γ_{HI} for the UVB simulation with the shortest $\lambda_{\mathrm{mfp}} = 5$ cMpc and has the greatest fluctuations. The bottom-right box shows a slice of Γ_{HI} for the UVB simulation with the longest $\lambda_{\mathrm{mfp}} = 150$ cMpc and has the weakest fluctuations. This follows since overall longer λ values means that photons travel further and effectively smooth the UVB over these large scales.
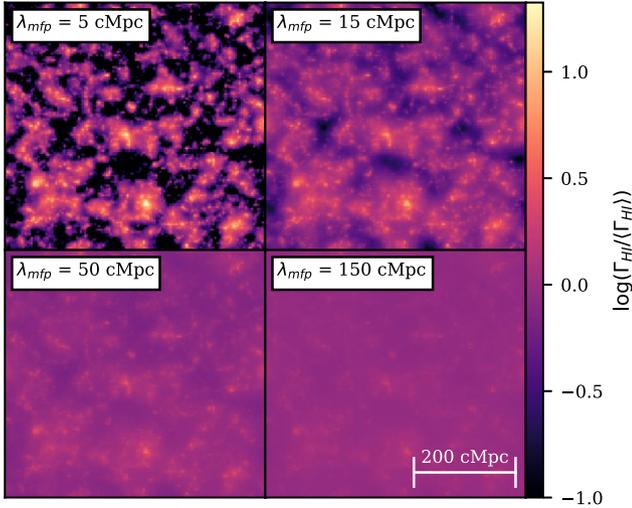
Figure 1. Each quadrant of this figure shows a slice through the box of the $z = 5.5$ UVB model used for four example values of λ_{mfp} (5, 15, 50, and 150 cMpc). The colour bar is cut off at $\log(\Gamma_{HI}/10^{-12}\,\mathrm{s}^{-1}) = -1$ in order to better visualize the differences between the models. The models with smaller λ_{mfp} values show greater variation in the UVB than those with larger λ_{mfp}, as expected.
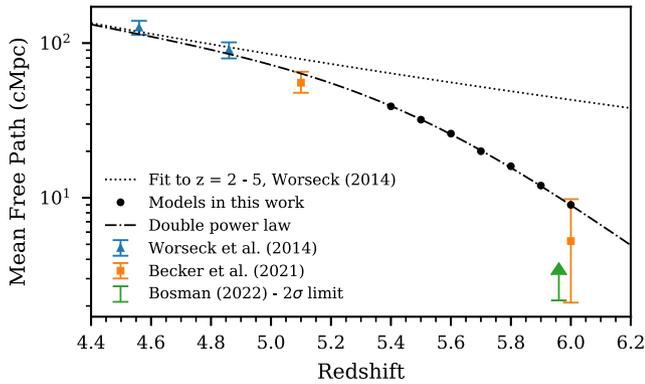
Figure 2. The blue triangles and orange squares show previous measurements of λ_{mfp} at high-z from Becker et al. (2021) and Worseck et al. (2014), respectively. The green limit is from Bosman (2021). Additionally, the dotted line shows the results of the power-law fit to data from $z = 2-5$ from Worseck et al. (2014). For this work, we modified this power-law fit into a double power law using the same low-z scaling by eye in order to agree with the Becker et al. (2021) points. This new scaling is shown by the dot–dashed line. We used this double power law as an example redshift evolution of λ_{mfp}, where the values we modelled are shown as black circles.

We ran UVB boxes for 14 values of λ_{mfp} (in cMpc): 5, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, and 150. To generate UVB boxes for additional values of λ_{mfp} we linearly interpolated the $\log(\Gamma_{HI})$ values at each location in the box between the two UVB boxes with the nearest λ_{mfp} values. This was done for three linearly spaced values between each existing λ_{mfp} values, resulting in a total of 53 UVB boxes.

To model a hypothetical evolution of λ_{mfp} as a function of redshift we used the double power law shown as the dot–dashed line shown in Fig. 2. This double power law was fit by eye with the following two considerations. We fixed the low-z behaviour to the power-law fit from Worseck et al. (2014) for $z < 5$: $\lambda_{mfp}(z) = (37 \pm 2)h_{70}^{-1}[(1 + z)/5]^{-5.4 \pm 0.5}$ Mpc (proper). We also required consistency with the new measurements at higher z from Becker et al. (2021).

Table 1. This table lists several relevant parameters for our simulations and mock data set. The second column lists the 'true' values of the redshift-dependent λ_{mfp} calculated from equation (1). The third column gives the true values of $\langle F \rangle$ at each z from Bosman et al. (2022). These $\langle F \rangle$ values are the central value for the grid of values considered. The final column contains the number of quasar sightlines we modelled for one mock data set, which is the data set size in Bosman et al. (2022). These sightlines each have a length of $\Delta z = 0.1$.

z	λ_{mfp} (cMpc)	$\langle F \rangle$	# QSOs
5.4	39	0.0801	64
5.5	32	0.0591	64
5.6	26	0.0447	59
5.7	20	0.0256	51
5.8	16	0.0172	45
5.9	12	0.0114	28
6.0	9	0.0089	19

The resulting double power law is

$$\lambda_{\mathrm{mfp}}(z) = \frac{37 h_{70}^{-1} \left(\frac{5}{6.55}\right)^{5.4}}{\left(\frac{1+z}{6.55}\right)^{5.4} + \left(\frac{1+z}{6.55}\right)^{25.5}} \text{ Mpc (proper)}. \tag{1}$$

We then evaluated equation (1) at centre of the seven redshift bins we considered and rounded to the nearest integer. The resulting true model λ_{mfp} values are listed in Table 1 and are plot as the black circles in Fig. 2. If these values were already in our set of 53 models then nothing else was done. If not, we linearly interpolated the value of $\log(\Gamma_{HI})$ at each point in the UVB simulation box between the two UVB boxes with the closest values of λ_{mfp} to get the final desired UVB box. This ultimately caused some redshifts to have 53 models of λ_{mfp} while others have 54. To generate the final flux skewers, we calculated the optical depths assuming a constant UVB then rescaled $\tau_{mfp} = \tau_{const.}/(\Gamma_{HI}/\langle\Gamma_{HI}\rangle)$. The $z = 5.5$ values of Γ_{HI} are used when generating flux skewers at all redshifts. This is justified because the value of λ_{mfp} is more important than the redshift evolution of the bias of the source population between $5 \leq z \leq 6$ (Furlanetto et al. 2017).

The overall average of Γ_{HI} calculated in the UVB fluctuation simulations is not uniquely determined since this originates from complicated galaxy physics. Thus, we force the average mean flux, $\langle F \rangle$, to be the same for each model where the average is taken over all flux skewers considered. This is achieved by calculating a constant, a, such that $\langle e^{-a\tau} \rangle = \langle F \rangle$. Additionally, we want to consider how changes in $\langle F \rangle$ would affect the autocorrelation function and determine if there is a degeneracy with λ_{mfp}. Therefore, we create a grid of nine values of $\langle F \rangle$ at each redshift. We chose the central value of $\langle F \rangle$ for a grid from Bosman et al. (2022) and chose the range of values to keep $\langle F \rangle > 0$ while not running into boundary issues during our inference.

2.2 Comparison of flux skewers

We drew 1000 skewers from the Nyx simulation and 1000 independent skewers of Γ_{HI} from the UVB boxes to use in this work. One example flux skewer, which combines the Nyx simulation skewer and the Γ_{HI} values from the UVB boxes, at $z = 5.4$ is shown in Fig. 3 for three different values of λ_{mfp} all normalized to $\langle F \rangle = 0.0801$. The bottom panel of this figure shows the corresponding UVB skewers that were used to calculate the flux. Two-dimensional (2D) slices of the UVB boxes these skewers came from are shown in Fig. 1. The model shown with the shortest λ_{mfp}, 5 cMpc (blue),
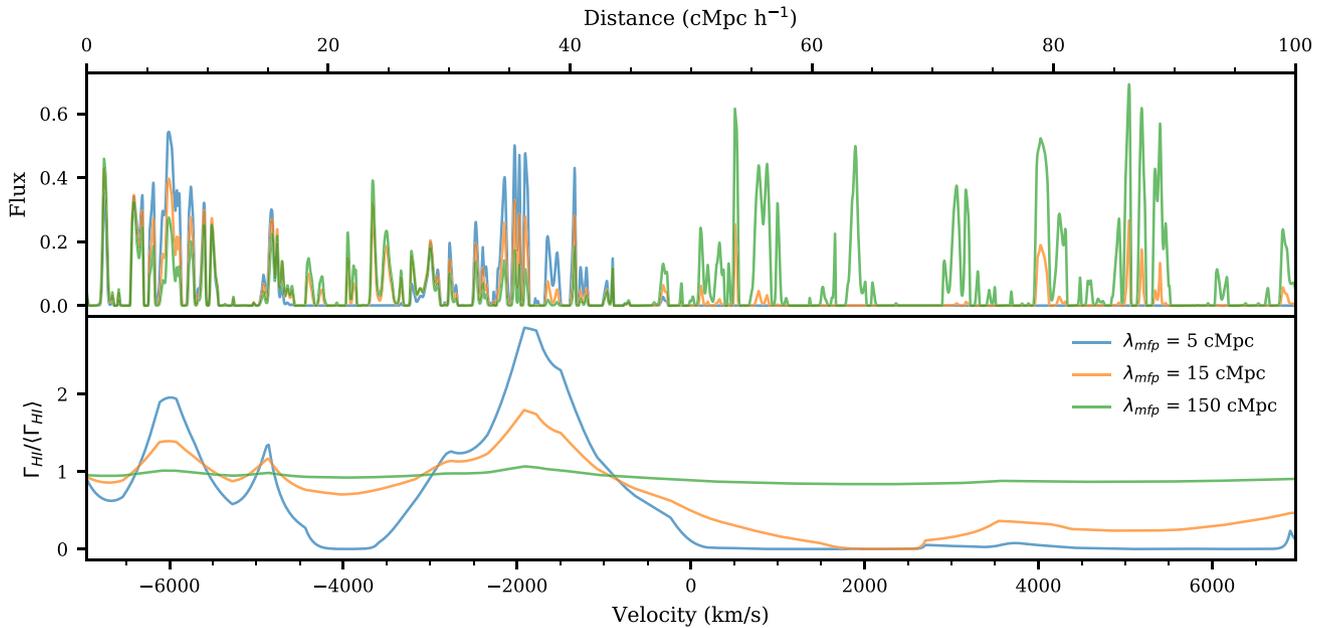
Figure 3. This figure shows the flux for one skewer of our simulation at $z = 5.4$ with different values of λ_{mfp} all normalized to $\langle F \rangle = 0.0801$ in the top panel. The bottom panel shows the corresponding UVB skewer used to calculate the flux. Smaller λ_{mfp} values (such as $\lambda_{\text{mfp}} = 5$ cMpc in blue) has greater variations in Γ_{HI} while the larger λ_{mfp} values (such as $\lambda_{\text{mfp}} = 150$ cMpc in green) are more uniform. Larger values of Γ_{HI} leads to increased flux in that region which can be seen when comparing the two panels. Consider $\Delta v = -2000$ km s^{-1}, here $\lambda_{\text{mfp}} = 5$ cMpc model (blue) has a peak in Γ_{HI}. The corresponding flux is boosted when compared to the other models. Additionally, for $\lambda_{\text{mfp}} = 5$ cMpc (blue) the Γ_{HI} values are very small for $\Delta v \geq 0$ km s^{-1} resulting in $F \sim 0$.

results in the greatest variation of $\Gamma_{\text{HI}}/\langle \Gamma_{\text{HI}} \rangle$. In particular, note that at $\Delta v = -2000$ km s^{-1}, the $\lambda_{\text{mfp}} = 5$ cMpc model (blue) has a peak in Γ_{HI} and the corresponding flux is boosted when compared to the other models. Additionally, for $\lambda_{\text{mfp}} = 5$ cMpc (blue) the Γ_{HI} values are very small for $\Delta v \geq 0$ km s^{-1} resulting in $F \sim 0$. The model with the largest λ_{mfp}, 150 cMpc (green), shows a mostly uniform Γ_{HI} skewer throughout the whole velocity range leading to more consistent flux levels.

2.3 Forward modelling

For this work, we aim to model the resolution, noise, and size properties of a realistic data set. We first chose to model a simplified version of the XQR-30 (main and extended) data set.[1] The main XQR-30 data set consists of 30 spectra of the brightest $z > 5.8$ quasars observed with VLT/X-shooter (Vernet et al. 2011). These spectra are supplemented with an extended data set consisting of 12 archival X-shooter spectra with comparable signal-to-noise ratio (SNR). See D'Odorico (in preparation) for additional information on these data. For this work, we specifically model properties similar to the data set of Bosman et al. (2022) which consists of the extended XQR-30 data supplemented with additional archival X-Shooter data and archival Keck/ESI spectra which have a lower resolution than the X-shooter spectra.

For our simplified modelling, we use the resolving power of X-shooter for visible light with a 0.9-arcsec slit, so $R = 8800$. We also use a typical SNR per 10 km s^{-1} pixel (SNR$_{10}$) of SNR$_{10} = 35.9$, which is the median of all the data presented in Bosman et al. (2022). Additionally, we investigate how higher resolution data with access to smaller scales in the Ly α forest would impact measurements of

λ_{mfp} from the autocorrelation function. To achieve this, we consider a 'high-resolution' data set with the same SNR$_{10}$ and size properties as the 'low-resolution' ($R = 8800$) data set but with $R = 30\,000$. This resolution is achievable with instruments such as Keck/HIRES, VLT/UVES, and Magellan/MIKE though the number of sightlines and noise properties used here do not represent a high-resolution data set currently in existence.

We model the resolution by smoothing the flux by a Gaussian filter then after smoothing we re-sampled such that there are four pixels per resolution element, where the resolution element is the full width at half-maximum (FWHM). This means, for the low-resolution data set, we smoothed by a Gaussian filter with FWHM ≈ 34 km s^{-1} then re-sampled so the pixel size was $\Delta v = 8.53$ km s^{-1}. For the high-resolution data set, we smoothed by a Gaussian filter with FWHM $= 10$ km s^{-1} then re-sampled so the pixel size was $\Delta v = 2.5$ km s^{-1}.

As stated above, we modelled an SNR$_{10} = 35.9$. Using SNR$_{\Delta v} =$ SNR$_{10} \sqrt{\Delta v / 10 \text{ km s}^{-1}}$, this corresponds to an SNR of 33.2 per 8.53 km s^{-1} low-resolution pixel and an SNR of 18.0 per 2.5 km s^{-1} high-resolution pixel. For simplicity, we add flux-independent noise in the following way. We generate one realization of random noise drawn from a Gaussian with $\sigma_N = 1/\text{SNR}_{\Delta v}$ for each SNR value and add this noise realization to every model at every redshift. The size of each noise realization is the number of skewers created (1000) by the number of pixels in the re-sampled flux skewers (1705 pixels for low resolution and 5814 pixels for high resolution). Using the same noise realization over the different models prevents stochasticity from different realizations of the noise from causing a noisy likelihood, which means the likelihood will be smooth as a function of model parameter. Thus, the noise modelling will not unduly, adversely affect the parameter inference.

A section of one skewer for both the initial and forward-modelled flux is shown in Fig. 4. Both panels show a skewer at $z = 5.4$ with $\lambda_{\text{mfp}} = 39$ cMpc and $\langle F \rangle = 0.0801$, our assumed true parameter values at this redshift. The initial flux in both panels is the same
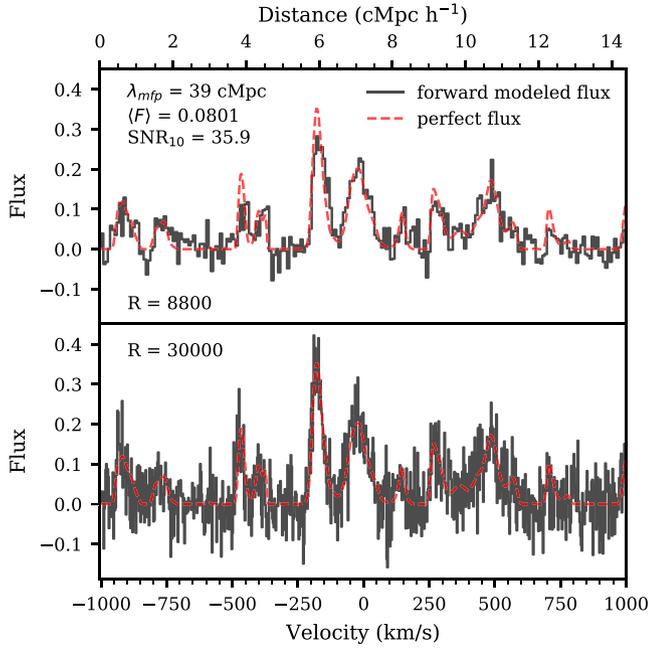
Figure 4. Both panels show initial and forward-modelled flux from a skewer with $\lambda_{mfp} = 39$ cMpc and $\langle F \rangle = 0.0801$ at $z = 5.4$. The initial flux is the same in both panels (red dashed line) while the forward modelled flux (black histogram) varies. The top panel shows the low-resolution flux with $R = 8800$, which represents XQR-30 data. The bottom panel shows the high-resolution flux with $R = 30\,000$. Both of these resolutions have $SNR_{10} = 35.9$ which leads to differing $SNR_{\Delta v}$ as can be seen when comparing the two panels.

and is shown as a red dashed line. The top panel shows the low-resolution forward-modelled flux (black histogram) with $R = 8800$. The bottom panel shows the high-resolution forward-modelled flux (black histogram) with $R = 30\,000$. Again both of these panels have the same $SNR_{10} = 35.9$ which results in different noise levels per pixel, as can be seen when comparing the two panels.

We assume a fiducial data set size that matches the number of sightlines reported in table 4 of Bosman et al. (2022) each with a length of $\Delta z = 0.1$. The number of sightlines are reported in the last column of Table 1 where to total path-length considered is equal to these values multiplied by $\Delta z = 0.1$. Redshift bins of $\Delta z = 0.1$ correspond to distances of 33–29 cMpc h^{-1} when centred at $z = 5.4$ to $z = 6.0$. However, the Nyx simulation box is 100 cMpc h^{-1} long, much longer than these redshift bins. If we were to use the full 100 cMpc h^{-1} skewers in our calculation we would be averaging over fewer skewers to get the same total Δz path. We wanted to use greater number independent skewers with more accurate lengths when compared with observed Ly α forest regions. For simplicity, we split all our skewers into two 40 cMpc h^{-1} regions which we treated as independent, giving us an effective number of 2000 independent skewers.

Note that unless otherwise specified the plots in this work mainly show results from the low resolution, $R = 8800$ data, since it represents existing XQR-30 data.

3. METHODS

3.1 Autocorrelation function

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

$$\xi_F(\Delta v) = \langle F(v)F(v + \Delta v) \rangle \tag{2}$$

where $F(v)$ is the flux of the Ly α forest and the average is performed over all pairs of pixels at the same velocity lag (Δv). The autocorrelation function is related to the power spectral density ($P_F(k)$) as

$$P_F(k) = \langle F \rangle^{-2} \int_{-\infty}^{\infty} \xi_F(\Delta v)e^{-ik\Delta v} d(\Delta v). \tag{3}$$

Note that this implies that the autocorrelation function should be sensitive to the same physical parameters as the power spectrum. Additionally, the autocorrelation function has nice properties with respect to white noise and spectral masks that make it a promising statistic to measure. Conventionally, the flux contrast field, $(F - \langle F \rangle)/\langle F \rangle$, is used when measuring statistics of the Ly α forest. Here, we chose to use the flux since $\langle F \rangle$ is small and has large uncertainties at high-z where we are most interested in this measurement. Using the flux thus prevents us from dividing by a small number which would come from an independent measurement and could potentially blow up the value of the flux contrast. This leads to the factor of $\langle F \rangle^{-2}$ in equation (3).

For each resolution and model we compute the autocorrelation function with a bin size of one FWHM of the resolution (either 34 km s^{-1} or 10 km s^{-1}) starting from this resolution size out to 20 cMpc h^{-1} (half the length of the skewer) which corresponds to $\sim 2900 \text{ km s}^{-1}$ at $z = 5.4$. The model value of the autocorrelation function was determined by taking the average of the autocorrelation function over all 2000 forward-modelled skewers. Each mock data set of the autocorrelation were calculated by taking an average over the appropriate number of random skewers for the number of quasars at that redshift from the initial 2000 forward-modelled skewers. The value of the autocorrelation function for small-scale bins is affected by the finite resolution. This effect is left in both the models and the mock data. We determine the errors on the models via the following estimate of the covariance matrix from mock draws of the data:

$$\Sigma(\boldsymbol{\xi}_{\text{model}}) = \frac{1}{N_{\text{mocks}}} \sum_{i=1}^{N_{\text{mocks}}} (\boldsymbol{\xi}_i - \boldsymbol{\xi}_{\text{model}})(\boldsymbol{\xi}_i - \boldsymbol{\xi}_{\text{model}})^{\text{T}} \tag{4}$$

where $\boldsymbol{\xi}_i$ is the autocorrelation function calculated for the ith mock data set, $\boldsymbol{\xi}_{\text{model}}$ is the average value of the autocorrelation function over all 2000 skewers, and N_{mocks} is the number of forward-modelled mock data sets used. Both the mock data sets and the overall average have the same values of λ_{mfp} and $\langle F \rangle$ in this calculation, so we end up with a covariance matrix at each parameter grid point. We use $N_{\text{mocks}} = 500\,000$ for all models and redshifts in this work, see Appendix A for a discussion on the convergence of the covariance matrix.

Fig. 5 shows the model value of the autocorrelation function with different parameter values at $z = 5.4$. The top panel shows models with a changing λ_{mfp} and constant $\langle F \rangle = 0.0801$. The solid lines show the model values calculated by averaging the autocorrelation function from all forward modelled skewers while the shaded regions show the errors from the diagonal elements of the covariance matrix as estimated in equation (4). Smaller λ_{mfp} values (such as $\lambda_{\text{mfp}} = 5$ cMpc – blue) result in a greater correlation function at all scales, though mainly at small scales, and larger error bars than large λ_{mfp} values (such as $\lambda_{\text{mfp}} = 150$ cMpc – green). These models are non-linearly spaced with greater differences between the models at small λ_{mfp} (blue versus orange) than large λ_{mfp} (orange versus green) which will result in variable sensitivity to λ_{mfp} from the autocorrelation function at different λ_{mfp} values. The bottom panel shows models with varying $\langle F \rangle$ and constant $\lambda_{\text{mfp}} = 39$ cMpc. $\langle F \rangle$ sets the overall amplitude of the autocorrelation function. Here the differences between models are linear where larger $\langle F \rangle$ leads to larger autocorrelation values.
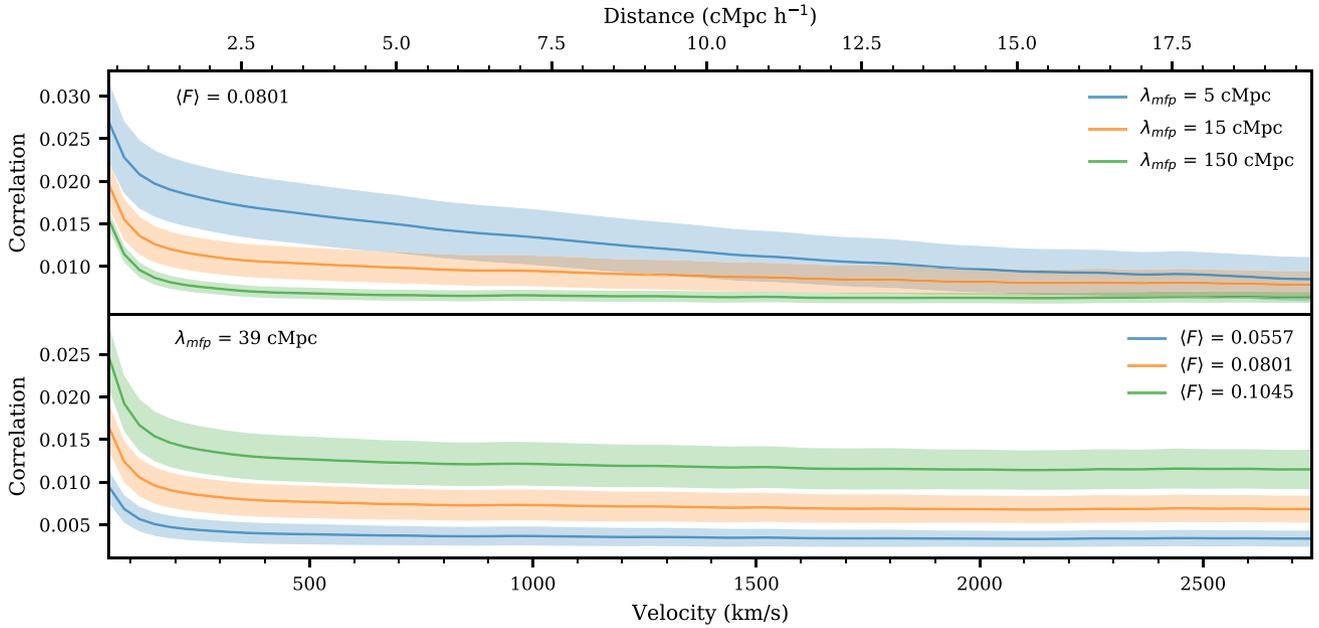
Figure 5. This figure demonstrates the effects of varying λ_{mfp} and $\langle F \rangle$ on the model values of the autocorrelation function at $z = 5.4$ and $R = 8800$. The solid lines show the model values calculated by averaging the autocorrelation function from all forward modelled skewers available while the shaded regions show the errors from the covariance matrix as estimated in equation (4). The top panel varies λ_{mfp} with a constant $\langle F \rangle$ labelled in the top-left corner while the bottom panel does the opposite. Both λ_{mfp} and $\langle F \rangle$ change the autocorrelation function on all scales shown, though λ_{mfp} appears to affect small scales more than large scales. In the top panel, the model value of the autocorrelation function are further apart for $\lambda_{\mathrm{mfp}} = 5$ cMpc (blue) and $\lambda_{\mathrm{mfp}} = 15$ cMpc (orange) than for $\lambda_{\mathrm{mfp}} = 15$ cMpc (orange) and $\lambda_{\mathrm{mfp}} = 150$ cMpc (green), which is a greater difference in λ_{mfp} value. This means the autocorrelation function is more sensitive to small λ_{mfp} values than large λ_{mfp} values. Comparatively, in the bottom panel, the differences in the mean autocorrelation function appear roughly linear with varying $\langle F \rangle$ which should result in similar sensitivity for all $\langle F \rangle$ values.

This scaling is roughly $\propto \langle F \rangle^2$, which follows from the definition of the autocorrelation function.

To visualize the covariance matrix, we define the correlation matrix, C. The correlation matrix is the covariance matrix with the diagonal normalized to 1. This is done to the jth, kth element by

$$C_{jk} = \frac{\Sigma_{jk}}{\sqrt{\Sigma_{jj}\Sigma_{kk}}}. \tag{5}$$

One example correlation matrix is shown in Fig. 6 for $z = 5.4$, $\lambda_{\mathrm{mfp}} = 39$ cMpc, and $\langle F \rangle = 0.0801$. All bins of the autocorrelation function are very-highly correlated which is due to the fact that each pixel in the Ly α forest contribute to multiple (in fact almost all) bins in the autocorrelation function.

3.2 Effect of model limits on the autocorrelation function

As stated in Section 2, the seminumerical method to generate the fluctuating UVB with various λ_{mfp} ignores the correlation between the density and Γ_{HI}. This is a result of the current limitations on available simulation boxes. We require that the UVB boxes are large enough to avoid suppressing UVB fluctuations and we require that the underlying hydrodynamical simulation boxes of the IGM have a grid that is fine enough to resolve the small structures in the Ly α forest. Lukić et al. (2015) found that this grid needs to have a grid resolution of 20 h^{-1} kpc to produce 1 per cent convergence of Ly α forest flux statistics. Davies & Furlanetto (2016) found that, with their 400-Mpc box of Γ_{HI} values, the tail of their optical depth distribution was impacted by cosmic variance, highlighting the need to go to even larger boxes. Having both a large box with a fine grid, which would be required to correlate the UVB and simulation box density, is currently too computationally expensive to be feasible.
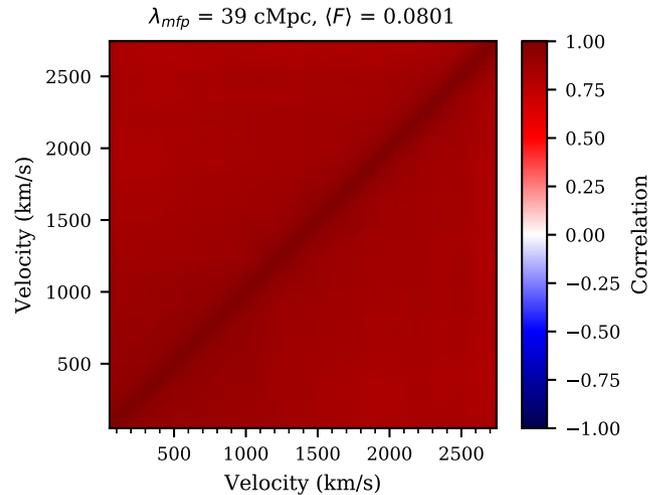
Figure 6. This figure shows the correlation matrix calculated with equation (5) with $N_{\mathrm{mocks}} = 500\,000$ for the model at $z = 5.4$ with $\lambda_{\mathrm{mfp}} = 39$ cMpc, $\langle F \rangle = 0.0801$, and $R = 8800$. The colour bar is fixed to span from -1 to 1, which is all possible values of the correlation matrix. Here, it is clear that all bins in the autocorrelation function are highly correlated with each other.

In general, there is a positive correlation between density and Γ_{HI} and a negative correlation between density and transmitted flux. This means that in areas with high Γ_{HI} there should also be higher density which would in turn decrease the transmitted flux, therefore reducing the extra signal from the short λ_{mfp}. To quantitatively explore this, we used a Nyx simulation box with a size of $L_{\mathrm{box}} = 40$ cMpc h^{-1} at $z = 5.8$. This box size has associated UVB values for $\lambda_{\mathrm{mfp}} = 15$
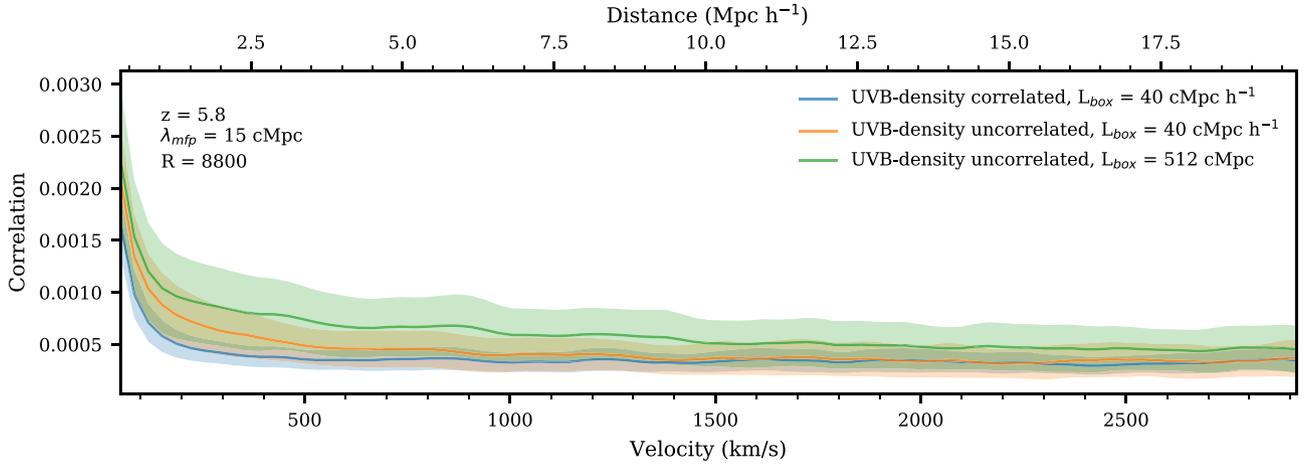
Figure 7. This figure demonstrates the effect of ignoring density correlations as well as using a small box size when generating Γ_{HI}. The blue line shows the autocorrelation function when using a Γ_{HI} calculated with the appropriate density field and a box size of $L_{\text{box}} = 40$ cMpc h^{-1}. The orange line shows the same for a Γ_{HI} calculated with a random density field and a box size of $L_{\text{box}} = 40$ cMpc h^{-1}, isolating the effect of density correlations when compared to blue. The green line shows the same for a Γ_{HI} calculated with a random density field and a box size of $L_{\text{box}} = 512$ cMpc, isolating the effect of the box size when compared with orange. Here, we see that the correct density field will cause the signal on small scales to be reduced and that using a larger box size will increase the signal for all scales.

cMpc generated with the same method of Davies & Furlanetto (2016) as described in Section 2.1. For these UVB boxes the local matter density matches that of the Nyx simulations of the IGM. We selected skewers from the UVB boxes in two ways: from the same location as the Nyx skewers or from a random location in the box. When the UVB skewers come from the same location as the Nyx skewers the density field and UVB field are correlated. When the UVB skewers come from a random location these two fields will not be correlated, which is analogous to the uncorrelated modelling adopted in the main text. The resulting autocorrelation function models are shown in Fig. 7 as the blue and orange lines. The blue line is the model with UVB skewers from the 40-h^{-1} cMpc box that were derived from the same density field as the Nyx simulations. The orange line is the model with UVB skewers from the 40-h^{-1} cMpc box that were derived from a random density field. Comparing these two lines isolates the effects of ignoring the UVB-density correlations. Here, we see that the density correlations reduce the autocorrelation signal at small scales while leaving the large-scale signal unchanged. When correlated, Γ_{HI} is proportional to the local density field so the regions of high Γ_{HI} values will also be regions of higher density. Since the optical depth scales as a power of the local density field, the boosted signal on small scales from regions of high Γ_{HI} in the orange model will be reduced by the corresponding increased local density leading to the reduction in small scales in the blue model. Since the reduction is happening on small scales, this mimics the effect of instead having a model with a larger λ_{mfp}.

Additionally, we investigated the effect of the box size used to generate the UVB on the amount of fluctuations in Γ_{HI} seen at a fixed λ_{mfp}. Using a smaller box size, such as the 40-h^{-1} cMpc box considered in Oñorbe et al. (2019), can suppress fluctuations in the local λ value since there is a smaller volume that must average to λ_{mfp}. For this comparison, we use randomly selected UVB skewers from the 40-h^{-1} cMpc box as well as randomly selected skewers from our 512 cMpc UVB box with $\lambda_{\text{mfp}} = 15$ cMpc from the main text of this work as described in Section 2.1. The UVB skewers chosen with both of these methods are uncorrelated with the density field, so we isolate the effect of only the box size. The two resulting autocorrelation function models are also shown in Fig. 7. Again, the orange line

is the model with UVB skewers from the 40-h^{-1} cMpc box that were derived from a random density field. The green line shows the model with UVB skewers from the 512-cMpc box that has a random density field compared to the Nyx simulation. Comparing this green line to the orange line thus isolates the effect of the small box size where again the large box size is required for UVB fluctuations to converge for a given λ_{mfp}. Here, we see that the green model has a greater signal than the orange at all scales. Therefore, both the blue and orange models with UVB skewers generated in a 40-h^{-1} cMpc box are likely underestimating the autocorrelation function on all scales. This makes it difficult to quantify the level of excess signal in the autocorrelation function that we get from ignoring correlations between the UVB and the local density field since the signal is underestimated on all scales when using the smaller UVB box. For this reason, we choose not to correct the mock data to account for the effect of using an uncorrelated UVB.

The mock data and models of the autocorrelation function from this study are self-consistently generated since both ignore the correlations between the UVB and the local density field. Therefore, the excess signal on small scales from modelling with an uncorrelated UVB will not bias the constraints obtained in this work. However, this excess on small scales needs to be accounted for when using the models of the autocorrelation function from the main text to constrain λ_{mfp} with observational data. We would expect measurements made by comparing data to models generated without UVB-density correlations to be biased towards larger values of λ_{mfp}, since the reduced signal on small scales from real density correlations would look like a larger λ_{mfp} in our models. We cannot quantify this potential bias with these simulations because, again, the small box size of 40 h^{-1} cMpc reduces the autocorrelation function signal on all scales. Modelling the UVB consistently with Ly α forest simulations in larger boxes is necessary to conclusively study the limitations of the model used in this work. We therefore leave this to future work.

3.3 Parameter estimation

To quantify the precision with which λ_{mfp} can be measured we use Bayesian inference with a multivariate Gaussian likelihood
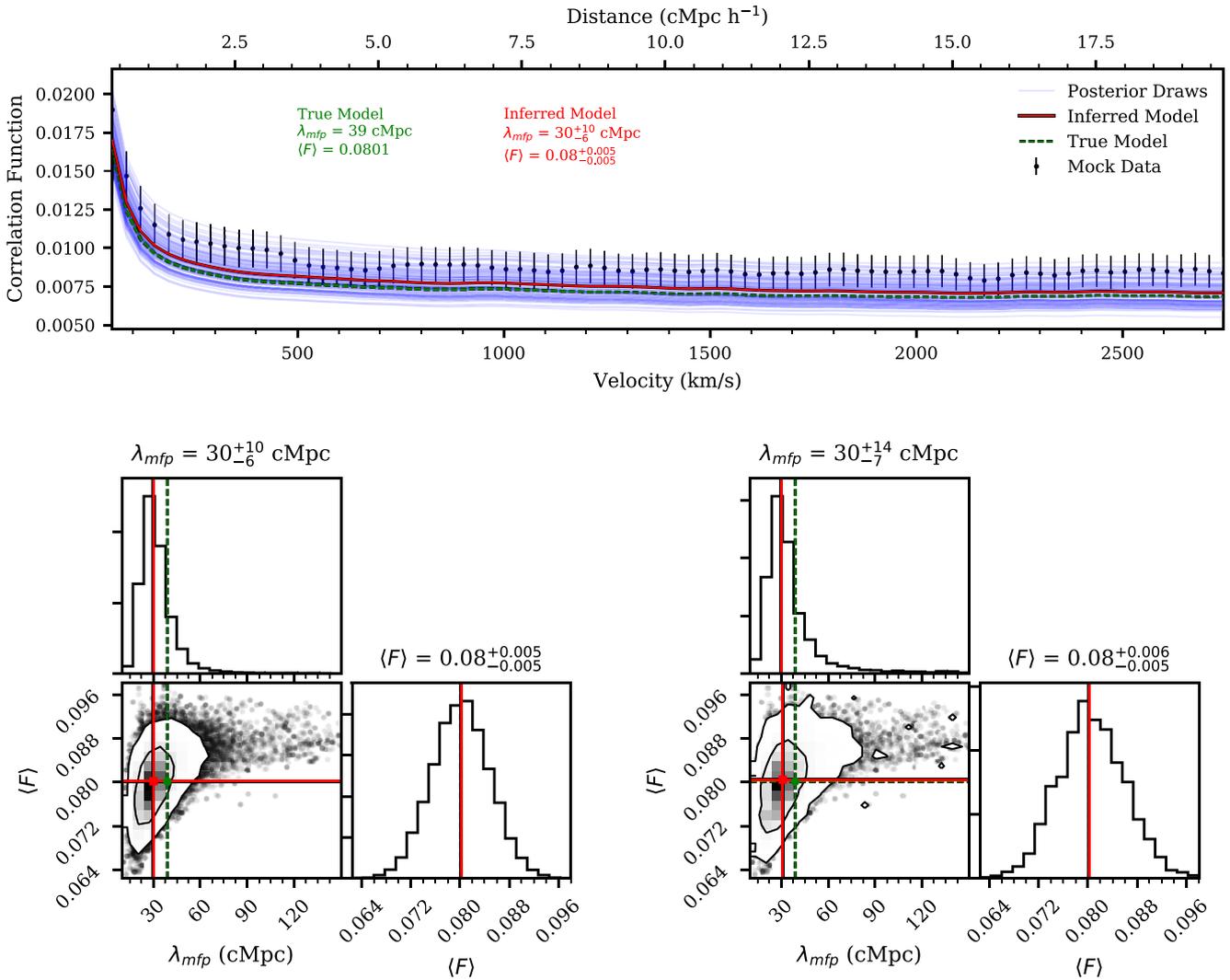
Figure 8. This figure illustrated the results of our inference procedure applied to one mock data set at $z = 5.4$. The top panel shows the resulting models from our inference procedure without re-weighting while the bottom panel has two corner plots that show the resulting parameters, the left without re-weighting and the right with re-weighting. In the top panel, the black points with error bars are the mock data with error bars from the inferred model. The inferred model was calculated by the median (50th percentile) of the MCMC chains of each parameter independently. The inferred model is shown as a red line while the accompanying red text reports errors calculated from the 16th and 84th percentiles of each parameter. In comparison, the true model the data was drawn from is the green dotted line and accompanying text. To demonstrate the width of the posterior, multiple faint blue lines are shown, which are the models corresponding to the parameters from 100 random draws of the MCMC chain. The bottom-left panel shows a corner plot of the values of λ_{mfp} and $\langle F \rangle$ that immediately result from our inference procedure. The bottom-right panel shows the corner plot of the values of λ_{mfp} and $\langle F \rangle$ from our inference procedure using the re-weighting approach. This means the corner plot has been made with the weights calculated from our inference test as described in Section 3.4.

and a flat prior over the parameters of interest. This likelihood ($\mathcal{L} = p(\xi | \lambda_{mfp}, \langle F \rangle)$) has the form

$$\mathcal{L} = \frac{1}{\sqrt{\det(\Sigma)(2\pi)^n}} \exp\left(-\frac{1}{2}(\xi - \xi_{model})^T \Sigma^{-1}(\xi - \xi_{model})\right) \quad (6)$$

where ξ is the autocorrelation function from our mock data, $\Sigma = \Sigma(\xi_{model})$ is the model-dependent covariance matrix estimated by equation (4), and n is the number of points in the autocorrelation function. We discuss the assumption of using a multivariate Gaussian likelihood in Appendix B.

Our models are defined by two parameters: λ_{mfp} and $\langle F \rangle$. We compute the posteriors for these parameters using Markov Chain Monte Carlo (MCMC) with the EMCEE package (Foreman-Mackey et al. 2013). We linearly interpolate the model values and covariance matrix elements onto a finer 2D grid of λ_{mfp} and $\langle F \rangle$ then use the

nearest model during the MCMC. This fine grid has 137 values of λ_{mfp} and 37 values of $\langle F \rangle$. Our MCMC was run with 16 walkers taking 5000 steps each and skipping the first 500 steps of each walker as a burn-in.

Fig. 8 shows the result of our inference procedure for one mock data set at $z = 5.4$. The top panel shows the mock data set with various lines relating to the inference procedure as follows. The green dotted line and accompanying text presents the true model that the mock data was drawn from. The mock data set is plot as the black point with error bars that come from the diagonal elements of the covariance matrix of the model that is nearest to the inferred model. The inferred model is the model that comes from the median of each parameter determined independently via the 50th percentile of the MCMC chains. The red lines and accompanying text shows the inferred model from MCMC. The errors on the inferred model written in the text are
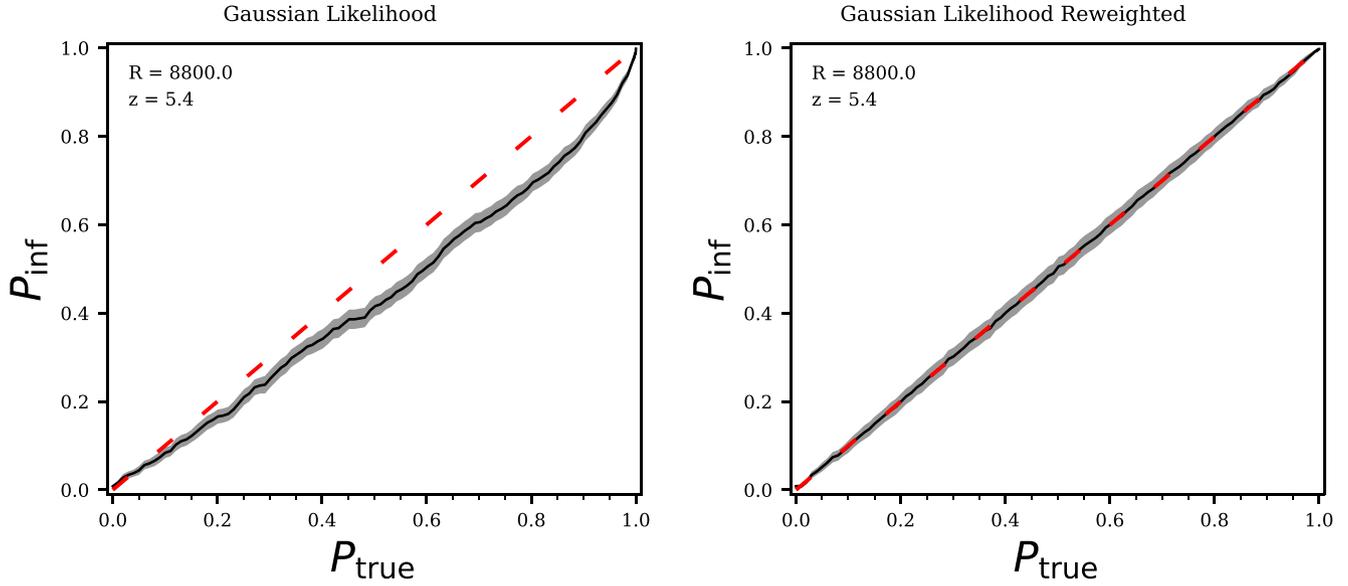
Figure 9. The left-hand panel of this figure shows the coverage resulting from the inference test from 500 models at $z = 5.4$ and $R = 8800$ drawn from our priors on λ_{mfp} and $\langle F \rangle$. Here, we see that the true parameters for the models fall above the 60th percentile in the MCMC chain ~ 50 per cent of the time, for example. The right-hand panel of this figure shows the coverage resulting from the inference test with the use of one set of weights to re-weight the posteriors. With these weights we are able to pass the inference test.

the distance between the 16th, 50th, and 84th percentiles of the MCMC chains. The blue lines show the models corresponding to 100 random draws from the MCMC chain to visually demonstrate variety of models that come from the resulting posterior. The bottom-left panel shows a corner plot of the posteriors for both λ_{mfp} and $\langle F \rangle$. Here, we see evidence of an extended tail out towards larger λ_{mfp} which is quantified in the asymmetric errors reported on the inferred value of λ_{mfp}. This asymmetry comes from the non-linear spacing of the autocorrelation function models as discussed in Section 3.1.

3.4 Inference test and re-weighting

We perform a test to check the fidelity of our inference procedure in order to ensure that our resulting posteriors act in a statistically valid way. This will ensure any assumptions we make during our inference are justified. For example, in this work, we used an approximate likelihood in the form of a multivariate Gaussian likelihood. The Ly α forest is known to be a non-Gaussian random field. By adopting a multivariate Gaussian likelihood here, we are tacitly assuming that averaging over all pixel pairs when calculating the autocorrelation function will Gaussianize the resulting distribution of the values of the autocorrelation function, as is expected from the central limit theorem. We discuss the distribution of these values for our mock data in detail in Appendix B. If this assumption is not valid our reported errors may be either underestimated or overestimated.

The general idea of our inference test is to compare the true probability contour levels with the 'coverage' probability. The coverage probability is the per cent of time the probability of the true parameters of a mock data set fall above a given probability level over many mock data sets. In our case, we compute this over 500 mock data sets where the true parameters considered are samples from our priors. Ideally, this coverage probability should be equal to the chosen probability contour level. This calculation can be done at many chosen probabilities resulting in multiple corresponding coverage probabilities. Existing work that explore this coverage probability

include Prangle et al. (2014), Ziegel & Gneiting (2014), Morrison & Simon (2018), and Sellentin & Starck (2019).

When considering multiple chosen probabilities, P_{true}, and resulting coverage probabilities, P_{inf}, the results can be plot against each other. The results of our inference test at $z = 5.4$ from 500 posteriors with true parameters randomly drawn from our priors are shown in the left-hand panel of Fig. 9. The grey-shaded regions around our resulting line show the Poisson errors for our results. Again we expect $P_{\mathrm{true}} = P_{\mathrm{inf}}$ which would give the red dashed line in this figure. To interpret this plot, first consider one point, e.g. $P_{\mathrm{true}} \approx 0.6$. This represents the 60th percentile contour, which was calculated by the 60th percentile of the probabilities from the draws of the MCMC chain for each mock data set. Here, the true parameters fall within the 60th percentile contour only ≈ 50 per cent of the time. This implies that our posteriors are too narrow and should be wider such that the true model parameters will fall in the 60th percentile contour more often, so we are in fact underestimating our errors. We run this inference test at all z considered in this work and found the deviation from the 1–1 line is worse at higher redshifts. See Appendix C for a discussion of the inference test at $z = 6$. We additionally run the inference test for mock data generated from a multivariate Gaussian distribution in Appendix D.

There has also been past work trying to correct posteriors that do not pass this coverage probability test (Prangle et al. 2014; Grünwald & van Ommen 2017; Sellentin & Starck 2019). In this work, we are using the method of Hennawi et al. (in preparation), where we can calculate one set of weights for the MCMC draws that broaden the posteriors in a mathematically rigorous way.

A brief description of the re-weighting method from Hennawi et al. (in preparation) is as follows. Consider one data set which gives a corresponding posterior probability density function (PDF). Initially, we have

$$\int \mathrm{d}\hat{x}\, p(\hat{x}) H(p(\hat{x}) > p_0) = \alpha_0 \qquad (7)$$

where $p(\hat{x})$ is the PDF of the posterior of some parameters $\hat{x}$, p_0 is a chosen posterior probability, H is the Heaviside function – causing the integrand to be 0 when the probability is less than the given contour p_0, and α_0 is the corresponding volume of the PDF where the probability of $\hat{x}$ is greater than p_0. This means that $\alpha_0 = 1 - C(p_0)$ where C is the cumulative distribution function.

If we instead consider our MCMC chain used to estimate the posterior with N_{samples} points each with probability $1/N_{\text{samples}}$ in the chain we get

$$\frac{1}{N_{\text{samples}}} \sum_i^{N_{\text{samples}}} H(p_i > p_0) = \frac{\# \text{ of samples with } p > p_0}{N_{\text{samples}}} = \alpha_0 \quad (8)$$

where the last equality comes from the fact that this sum is a Monte Carlo integral.

Consider the corresponding percentile, P_{true}, of this probability contour. By definition, $C(P_{\text{true}}) = 1 - P_{\text{true}}$ (because the greatest values of the probability correspond to the smallest percentile contours). Thus, we have

$$\frac{1}{N_{\text{samples}}} \sum_i^{N_{\text{samples}}} H(p_i > p_0) = P_{\text{true}}. \quad (9)$$

However, as discussed above, after running an inference test what was thought of as the P_{true} percentile contour is in reality the inferred percentile, P_{inf}, contour. Previous work Sellentin & Starck (2019) suggested relabeling the P_{true} contour as the P_{inf} contour. However, another method to broaden (or condense) the probability contour is by using a set of weights. Consider re-writing equation (9) using weights, w:

$$\frac{1}{N_{\text{samples}}} \sum_i^{N_{\text{samples}}} w(x_i) H(p_i > p_0) \approx f(P_{\text{true}}). \quad (10)$$

You can then consider multiple values of P_{true} and absorb the factor of $\frac{1}{N_{\text{samples}}}$ into the weights

$$A\boldsymbol{w} = f(\boldsymbol{P}_{\text{true}}) \quad (11)$$

where A is matrix of only 1s and 0s from the Heaviside function, $\boldsymbol{w}$ is the vector of weights, and $\boldsymbol{P}_{\text{true}}$ is the vector of probability contours considered. In fact, we can order the samples from the smallest probability value to the largest probability value such that A is an upper triangular matrix. To guarantee the new weighted probability contours behave as they should statistically (i.e. the true value falls in the Pth percentile contour P per cent of the time), we set $f(P_{\text{true}}) = P_{\text{inf}}$. This works because P_{inf} is the measured statistically correct percentile contour for this P_{true} value from the previous inference test. Therefore, we can compute weights by

$$\boldsymbol{w} = A^{-1}\boldsymbol{P}_{\text{inf}}. \quad (12)$$

Note that this equation implies that we must run the inference test for the number of probability contours equal to the number of MCMC probability samples we have for each posterior. However, in practice we compute much fewer P_{inf} values during the inference test and then interpolate this vector onto one with the number of MCMC samples we have.

Thus, we are able to calculate one set of $(5000 - 500) \times 16 = 72\,000$ weights that would be used for all 500 mock data sets to broaden the posteriors and pass this inference test. The weights calculated by this method, for a given set of MCMC chains, are unique. The line resulting from the inference test after calculating and using a set of weights is shown in the right-hand panel of Fig. 9. This line clearly falls along the 1–1 line as expected so our calculated

weights allow us to re-weight our posteriors into a statistically correct form. See Appendix C for a discussion of the re-weighting at $z = 6$.

We show the re-weighted posteriors on λ_{mfp} and $\langle F \rangle$ in the bottom-right part of Fig. 8. The weights give greater importance to larger values of λ_{mfp} in the tail of values to the right, effectively broadening the posteriors and increasing the errors on the fit. For the mock data set in Fig. 8 the re-weighted marginalized posterior for λ_{mfp} gives $\lambda_{\text{mfp}} = 30^{+14}_{-7}$ cMpc whereas before the inferred value was 30^{+10}_{-6} cMpc, so the new errors are ~ 30 per cent larger. When looking at the 2D distribution in this panel we see that the weights do introduce an additional source of noise to the posterior distribution.

This whole inference procedure is not the optimal and will not give the best constraints on λ_{mfp} possible from this statistic. The need to use re-weighting, or some method to correct our posteriors to pass an inference test, comes from our incorrect (though frequently used) assumption of a multivariate Gaussian likelihood. The values of the autocorrelation function at these high z do not sufficiently follow a multivariate Gaussian distribution to justify this assumption, which should be a warning for other studies of the Ly α forest at these z. Using a more correct form of the likelihood (such as a skewed distribution) or likelihood-free inference (such as approximate Bayesian computation as used in Davies et al. (2018) or other machine learning methods) would lead to more optimal posteriors that better reflect the information in the autocorrelation function. Therefore, future work on this inference procedure will improve the constraints on λ_{mfp}.

4. RESULTS

In order to consider the range of observational constraints possible from one set of λ_{mfp} and $\langle F \rangle$ values because of cosmic variance, we study the distribution of measurements for 100 mock data sets. For each z, we use the λ_{mfp} and $\langle F \rangle$ values reported in Table 1. Each mock data set is chosen by randomly selecting the appropriate number of skewers given the data set size at each redshift, and averaging the autocorrelation function for each individual skewer. For each mock data set, we perform MCMC as described in Section 3.3 and then re-weight the resulting posteriors following Section 3.4. Once we have the weights and the chains resulting from our inference procedure we can calculate the marginalized posterior for λ_{mfp}.

We calculate the marginalized re-weighted posteriors for 100 mock data sets at each z and R. All 100 marginalized re-weighted posteriors are shown as the faint blue lines in Fig. 10 at $z = 5.4$ for $R = 8800$ (top panel) and $R = 30\,000$ (bottom panel). In addition to the randomly selected mock data sets, we computed the re-weighted posterior using the model value of the autocorrelation. This is shown as the blue histogram in Fig. 10. Using the model value as mock data is the ideal case and removes the luck of the draw from affecting the precision of this posterior. The measurement resulting from the model data is written in the blue text of this figure and the values at each z and R are reported in Table 2.

Fig. 10 shows the results from all 100 mock data sets (blue lines) at $z = 5.4$ in order to visualize the various shapes of the resulting re-weighted posteriors. These data sets all have the true values of $\lambda_{\text{mfp}} = 39$ cMpc, $\langle F \rangle = 0.0801$, and a 64 quasar data set. The top panel shows the low-resolution $R = 8800$ results and the bottom panel shows the high-resolution $R = 30\,000$ results. The re-weighted histograms in Fig. 10 are noisy; much like is seen in the bottom-right panel of Fig. 8. This is a direct consequence of our re-weighting procedure and will be improved with further work on likelihood-free inference. There are roughly two behaviours of posteriors shown here: those
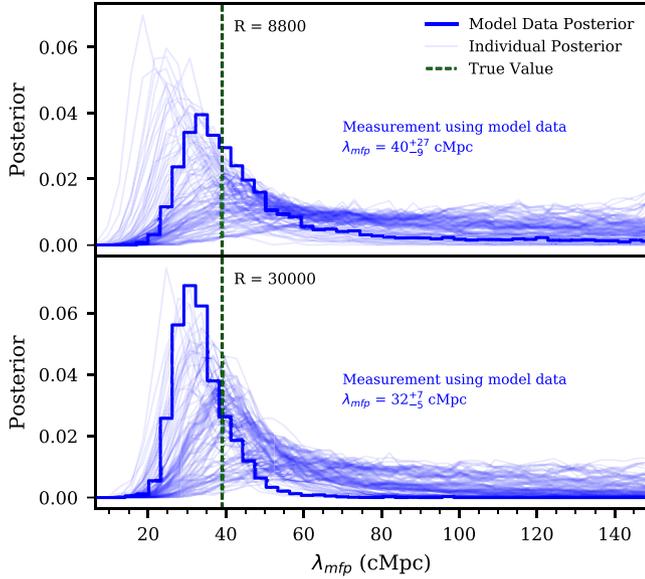
Figure 10. This figure shows 100 re-weighted posteriors of λ_{mfp} at $z = 5.4$ with true $\lambda_{mfp} = 39$ cMpc and $\langle F \rangle = 0.0801$ (blue faint lines). The top panel shows the low-resolution $R = 8800$ results and the bottom panel shows the high-resolution $R = 30\,000$ results. It also displays the re-weighted posterior (thick blue histogram) from the model's value of the autocorrelation function with the measurement of these average posteriors written in blue text. This demonstrates the different possible behaviours the posterior can have from our method. Overall, the higher resolution data do produce more precise posteriors, including the average posterior which is seen in the higher peak and smaller reported errors.

Table 2. This table contains the results of analysing the λ_{mfp} posteriors for the model value of the autocorrelation function. The mock data at each z have the same value of λ_{mfp} as recorded in Table 1. The first column contains the modelled value of λ_{mfp} at each z that was used in this measurement. The next column contains the resulting measurements at each z for $R = 8800$ data while the last column has the resulting measurements for $R = 30\,000$ data. In general the trend of the errors is to initially decrease with increasing redshift and then stay about flat beyond $z = 5.7$. This trend follows from the evolution in the true value of λ_{mfp} and the data set size at each z.

z	Model λ_{mfp} (cMpc)	Measured λ_{mfp} (cMpc)	
		$R = 8800$	$R = 30\,000$
5.4	39	40^{+27}_{-9}	32^{+7}_{-5}
5.5	32	35^{+12}_{-6}	33^{+6}_{-4}
5.6	26	28^{+7}_{-4}	27^{+5}_{-3}
5.7	20	22^{+7}_{-4}	20^{+4}_{-3}
5.8	16	18^{+6}_{-4}	16^{+3}_{-3}
5.9	12	14^{+5}_{-4}	13^{+3}_{-3}
6.0	9	12^{+6}_{-3}	11^{+4}_{-2}

that have a large peak at low values and those that are lower limits starting at low value and staying non-zero at our upper boundary of 150 cMpc. For both the lower resolution and higher resolution data, the model values of the autocorrelation function give posteriors with typical widths when compared to the mock data. Both model posteriors also contain the true value of λ_{mfp} within their 1σ error bars. Overall, the higher resolution data do produce tighter, more precise posteriors for both the model value and the mock data.

Table 2 reports the measurements that result from using the model values of the autocorrelation function as our data. This is

an ideal scenario that removes luck of the draw from the resulting measurement. The first column contains the modelled value of λ_{mfp} at each z that was used in this measurement, which also appears in Table 1. The next column has the resulting measurements for $R = 8800$ data while the last column has the resulting measurement for the $R = 30\,000$ data. The errors initially decrease with increasing redshift and then stay about the same beyond $z = 5.7$. There are two important factors to consider when looking at this trend. First is the trend of the true value of λ_{mfp} with z where λ_{mfp} decreases with increasing z. The autocorrelation function is more sensitive to smaller values of λ_{mfp} as discussed in Section 3.1. Briefly this is due to the fact that smaller λ_{mfp} produces greater fluctuations resulting in a larger signal. This means we would expect the results to get more precise and thus have smaller errors at higher z. The other factor is the size of the data set, which is greatest at the lowest z. We would expect the measurements to be less precise and thus have larger errors for the smaller data sets at high z. These effects combine resulting in the trend we see. When comparing the $R = 8800$ and $R = 30\,000$ measurements, we find that the $R = 30\,000$ values are on average 40 per cent more precise. Note that it also appears that the measured values of λ_{mfp} are always biased high. However, these posterior distributions have tails to greater values of λ_{mfp} which causes the reported median measured value of λ_{mfp} to be greater than the most likely value of λ_{mfp}.

In order to visualize the differences between measurements at different redshifts, we plot the results for five random mock data sets with $R = 8800$ in Fig. 11. Each violin is the re-weighted marginalized posterior for one randomly selected mock data set at the corresponding redshift. The light blue-shaded region demarcates the 2.5th and 97.5th percentiles (2σ) of the MCMC draws while the darker blue-shaded region demarcates the 16th and 84th percentiles (1σ) of the MCMC draws. The dot–dashed line is the double power law, equation (1), which we used to determine the true λ_{mfp} evolution as shown in Fig. 2.

Looking at the posteriors for a given redshift (one column in the figure), the only difference between the posteriors is the random mock data set drawn. This still produces different precision results as seen in Fig. 10 for $z = 5.4$. There are then three differences between mock data sets shown for a given panel. First is again the mock data is chosen at random so there will be fluctuations in the precision with the luck of the draw. The mock data at different redshift also have different true λ_{mfp} values, shown in the dot–dashed black line, where the smallest λ_{mfp} value is at the highest z. The autocorrelation function is most precise at small inferred λ_{mfp} values which are more likely at the highest z. Additionally, the redshifts each have different data set sizes, as reported in Table 1. The highest redshifts have the smallest data set sizes, leading to greater scatter in the precision of the posteriors. Again, the individual posteriors are noisy, resulting from the re-weighting procedure as described in Section 3.4.

Here all the mock data sets are at our lower resolution, $R = 8800$, which was chosen to mimic the existing XQR-30 data. In Appendix E, we discuss the same plot (Fig. E1) but with the higher resolution, $R = 30\,000$, data. The only difference between the data used in Figs 11 and E1 is the resolution of the mock data. The randomly chosen mock data sets, the data set sizes, and the true values are the same.

5. CONCLUSIONS

In this work, we have investigated to what precision λ_{mfp} can be constrained using the autocorrelation function of Lyα forest flux in quasar sightlines. Overall, we found that the autocorrelation function
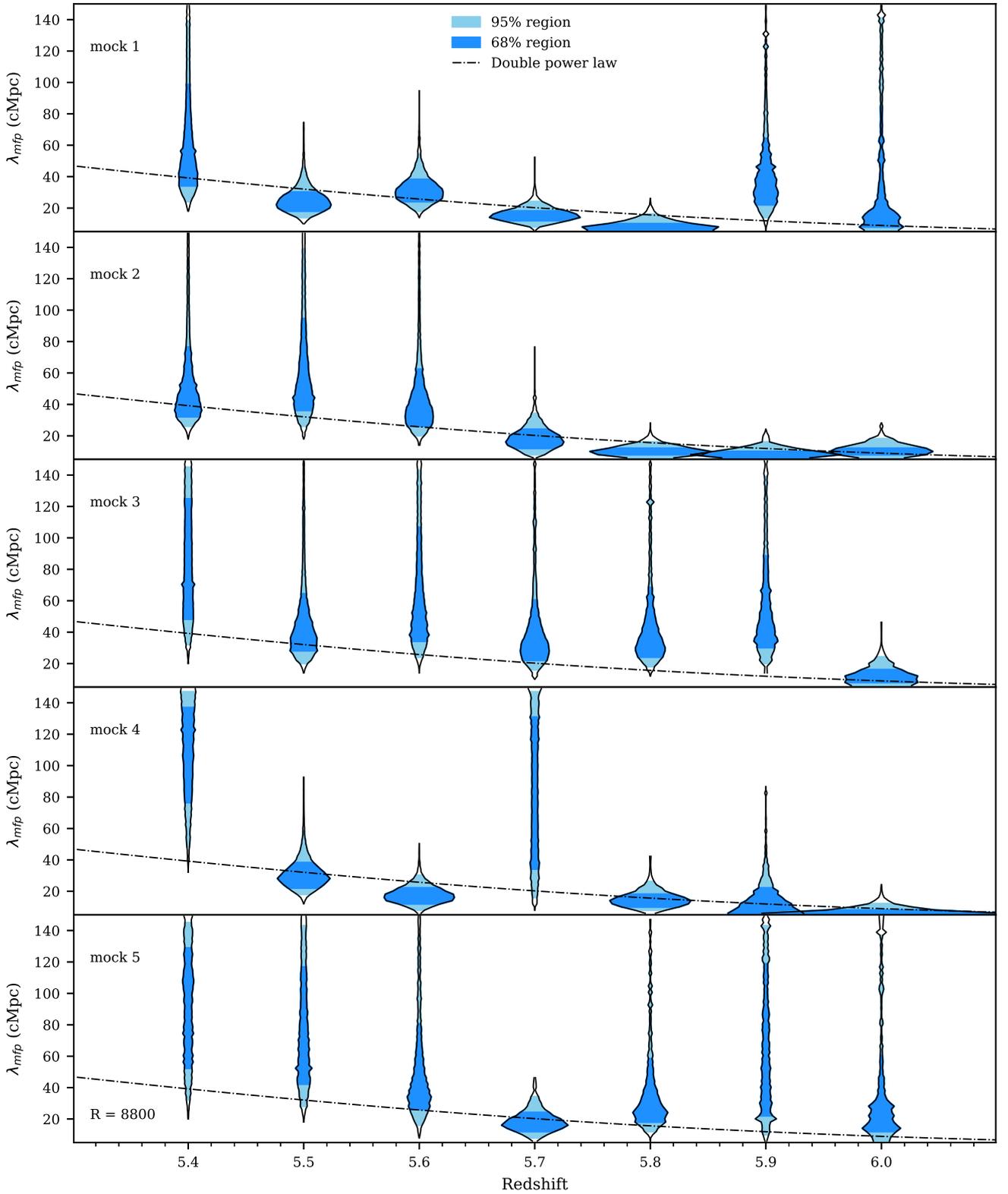
Figure 11. Each panel of this figure shows one posterior for a different randomly selected low-resolution ($R = 8800$) mock data set at each z. For each posterior, the light blue-shaded region demarcates the 2.5th and 97.5th percentile of the MCMC draws while the dark blue-shaded region demarcates the 17th and 83rd percentile of the MCMC draws. The black dot–dashed line shows the double power law from equation (1) and Fig. 2. The behaviour of each posterior at the different z is determined by the luck of the draw when selecting the mock data, the true λ$_{mfp}$ value at each z, and the data set size at each z. The true λ$_{mfp}$ values and data set sizes are reported in Table 1.

is sensitive to the value of λ_{mfp} across multiple redshift bins and for realistic mock data with both high and low resolution. We computed the marginalized re-weighted posterior for λ_{mfp} for 100 mock data sets with properties similar to the XQR-30 extended data set at $5.4 \leq z \leq 6.0$. We additionally considered 100 mock data sets with $R = 30\,000$, over three times greater than XQR-30 data resolution. The re-weighted posterior showed a variety of behaviours based on the luck of the draw of the mock data chosen, the true value of λ_{mfp} for the mock data, and the data set size at each z.

We considered an ideal data set which had the model value of the autocorrelation function, effectively removing the luck of the draw from our measurement. The error on these measurements for both the high- and low-resolution data initially got smaller (more precise) with increasing redshift then stayed about the same beyond $z = 5.7$. This followed from the changing true value of λ_{mfp} and the size of the data set at each z. Small values of λ_{mfp} lead to greater fluctuations in the UVB and thus produce an increased signal in the autocorrelation function. This makes the autocorrelation function more sensitive to smaller values of λ_{mfp} than larger values of λ_{mfp} where the fluctuations are smaller. This work has opened up the possibility for future measurements of λ_{mfp} with the autocorrelation function by quantifying the sensitivity of this method.

Of particular interest is the measurement at $z = 6.0$, where recent measurements imply a rapid evolution of λ_{mfp}. For our ideal model data at $z = 6.0$ with $R = 8800$, we get $\lambda_{mfp} = 12^{+6}_{-3}$ cMpc where the true value we modelled was $\lambda_{mfp} = 9$ cMpc. In comparison, the measurement from Becker et al. (2021) at $z = 6.0$ is $0.75^{+0.65}_{-0.45}$ proper Mpc (or $5.25^{+4.25}_{-3.15}$ cMpc). Thus, our ideal measurement with this new statistical method has comparable error bars as those from Becker et al. (2021). We therefore expect that a measurement using this technique on real data will provide a competitive, secondary check on the value of λ_{mfp} at $z = 6.0$. Additionally, we have shown that our method can be applied to multiple fine redshift bins from $5.4 \leq z \leq 6.0$ to precisely constrain the evolution of λ_{mfp}.

Note that our procedure uses a multivariate Gaussian likelihood, MCMC, and a set of weights for the MCMC chains that ensures our posteriors pass an inference test. The original failure of our procedure to pass an inference test is likely due to the incorrect assumption that the autocorrelation function follows a multivariate Gaussian distribution, as discussed in Appendix B. This result should caution against using a multivariate Gaussian likelihood with other statistics, such as the power spectrum, when making measurements at $z > 5$ as the same issue of non-Gaussian data likely applies. This is especially concerning if the low value of λ_{mfp} with high corresponding fluctuations in the UVB at high-z holds true. In the future, better likelihoods or likelihood-free inference will allow for a more optimal inference procedure (see e.g. Davies et al. 2018 or Alsing et al. 2019). This will lead to tighter constraints on λ_{mfp} from the autocorrelation function.

For this work, we used the method of Davies & Furlanetto (2016) to generate the UVB boxes as described in Section 2.1. This assumes a fixed source model which could potentially prove to be incorrect. For example, if fainter galaxies had higher escape fractions it would reduce the strength of UVB fluctuations at fixed λ_{mfp}, also reducing the autocorrelation signal. This would bias λ_{mfp} measurements through this method from real data compared to these models (though it is consistent for our mock data generated from our models). We leave a detailed consideration of the effect of other source models to future work.

Our work also discussed the effect of the current limitations in modelling the UVB and Ly α forest on the autocorrelation function. Namely, our UVB boxes are not correlated with the density of

our Nyx simulation box, where in reality these quantities are physically correlated. We considered the effect of a correlated UVB in Section 3.2. We found that the correlation between high-density areas with increased UVB values would reduce the autocorrelation signal for a fixed λ_{mfp} on small scales, since higher density leads to reduced transmission. This would again bias a measurement from real data, where these correlations would exist, because the true signal for a given λ_{mfp} should be lower than it is in our models, which mimics a model with a larger λ_{mfp} value. However, this comparison was done in a small box (40 cMpc h^{-1}) which suppresses UVB fluctuations on all scales as is also discussed in Section 3.2. Suppressing fluctuations in the UVB causes the autocorrelation signal to be lower in these boxes. Thus in this comparison the signal is smaller from the density correlations but the UVB fluctuations are also underestimated due to the box size. The existence of both of these effects means that we were not able to quantify any potential bias from the uncorrelated UVB boxes. The mock data used in this work are generated in the same ways as the models they are compared to, so the measurements here are self-consistent. However, any attempts to compare these models with actual data will need to take into account the effect of using an uncorrelated UVB in the modelling. Thus, future work on UVB models will be necessary before observational λ_{mfp} constraints can be produced.

Another potential physical impact on the autocorrelation signal is fluctuation in the temperature of the IGM. Oñorbe et al. (2019) showed that fluctuations in the temperature of the IGM impacted the largest scales of the power spectrum at $z > 5$. We therefore would conclude these fluctuations would also impact the autocorrelation function, which is the Fourier transform of the power spectrum. However, Oñorbe et al. (2019) also considered a fluctuating UVB and found that this effectively cancelled out the impact of the thermal fluctuations on the largest scales of the power spectrum. We leave further work on the impact of temperature fluctuations along with UVB fluctuations to future work.

Continuum errors will affect the measurement of the autocorrelation on larger scales which are less important than the small scales when considering λ_{mfp}. The reconstruction done in Bosman et al. (2022) is shown to reconstruct the continuum within 8 per cent. Additionally, Eilers, Hennawi & Lee (2017) showed that continuum errors had minimal effect on the shape of the normalized flux PDF at $z = 5$ where transmission is low. We have left a detailed exploration of the effect of continuum errors on the autocorrelation function for future work.

We also note that there is additional $z > 5$ Ly α forest data in telescope archives with lower SNR that could be used in our analysis. Here, we limited the consideration to mock XQR-30 data (and a high-resolution analogue) but will consider the impact of adding noisier data in future work.

The value of λ_{mfp} and its evolution at high z is important for understanding reionization. Measuring λ_{mfp} at high z is a difficult task that so far has been restricted to two redshift bins at $z > 5$. This work has shown that the autocorrelation function of the Ly α forest flux provides a new, competitive way to constrain λ_{mfp} in multiple redshift bins at $z \geq 5.4$.

DATA AVAILABILITY

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

Alsing J., Charnock T., Feeney S., Wandelt B., 2019, MNRAS, 488, 4440

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., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853

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. 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

Croft R. A. C., 2004, ApJ, 610, 642

D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018, MNRAS, 473, 560

Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328

Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018, ApJ, 855, 106

Davies F. B., Bosman S. E. I., Furlanetto S. R., Becker G. D., D'Aloisio A., 2021, ApJ, 918, L35

Eilers A.-C., Hennawi J. F., Lee K.-G., 2017, ApJ, 844, 136

Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53

Fan X. et al., 2006, AJ, 132, 117

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Fumagalli M., O'Meara J. M., Prochaska J. X., Worseck G., 2013, ApJ, 775, 78

Furlanetto S. R., Mirocha J., Mebane R. H., Sun G., 2017, MNRAS, 472, 1576

Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2021, MNRAS, 506, 4389

Gnedin N. Y., 2000, ApJ, 535, 530

Gnedin N. Y., Fan X., 2006, ApJ, 648, 1

Gnedin N. Y., Madau P., 2022, Liv. Rev. Comput. Astrophys., 8, 3

Gontcho A Gontcho S., Miralda-Escudé J., Busca N. G., 2014, MNRAS, 442, 187

Grünwald P., van Ommen T., 2017, Bayesian Anal., 12, 1069

Haardt F., Madau P., 2012, ApJ, 746, 125

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

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

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

Morrison J., Simon N., 2018, J. Comput. Graph. Stat., 27, 648

Nasir F., D'Aloisio A., 2020, MNRAS, 494, 3080

O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137

Oñorbe J., Davies F. B., Lukić Z., Hennawi J. F., Sorini D., 2019, MNRAS, 486, 4075

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

Prangle D., Blum M. G. B., Popovic G., Sisson S. A., 2014, Aust. N.Z. J. Stat., 56, 309

Prochaska J. X., Worseck G., O'Meara J. M., 2009, ApJ, 705, L113

Rauch M., 1998, ARA&A, 36, 267

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

Sellentin E., Starck J.-L., 2019, J. Cosmol. Astropart. Phys., 2019, 21

Songaila A., Cowie L. L., 2010, ApJ, 721, 1448

Vernet J. et al., 2011, A&A, 536, A105

Walther M., Oñorbe J., Hennawi J. F., Lukić Z., 2019, ApJ, 872, 13

Worseck G. et al., 2014, MNRAS, 445, 1745

Wyithe J. S. B., Bolton J. S., Haehnelt M. G., 2008, MNRAS, 383, 691

Yang J. et al., 2020, ApJ, 904, 26

Zel'dovich Y. B., 1970, A&A, 5, 84

Ziegel J. F., Gneiting T., 2014, Electron. J. Stat., 8, 2619

Zuo L., 1992a, MNRAS, 258, 45

Zuo L., 1992b, MNRAS, 258, 36

APPENDIX A: CONVERGENCE OF THE COVARIANCE MATRICES

We calculate the covariance matrices for our models with mock draws, as defined in equation (4). Using mock draws is inherently noisy and it should converge as $1/\sqrt{N}$ where N is the number of draws used. As stated in the text, we used 500 000 mock draws. To check that this number is sufficient to minimize the error in our
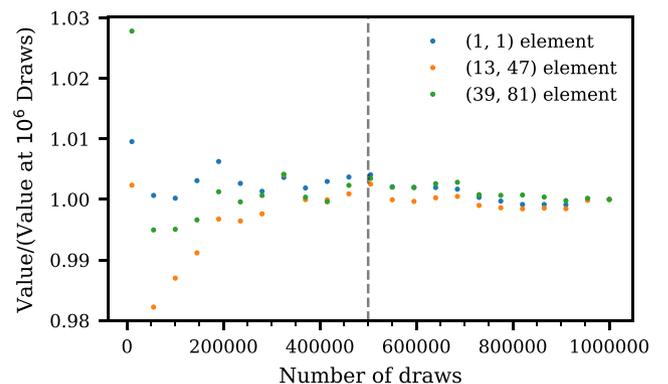
Figure A1. This figure shows the behaviour of three elements of the model covariance matrix ($z = 6$, $R = 8800$, $\lambda_{mfp} = 9$ cMpc, and $\langle F \rangle = 0.0089$) for different numbers of mock draws. At all values of the number of mocks considered, the covariance elements fall within 3 per cent of their final value. By around $\sim 200\,000$ draws, all of the values fall within 1 per cent of the final value. For this reason, we believe using 500 000 mock draws is sufficient to generate the covariance matrices used in this study. 500 000 mock draws is represented by the vertical dashed black line.

calculation, we looked at the behaviour of elements of one covariance matrix in Fig. A1. This covariance matrix is for the model with $z = 6$, $R = 8800$, $\lambda_{\mathrm{mfp}} = 9$ cMpc, and $\langle F \rangle = 0.0089$, chosen because $z = 6$ has the lowest 'true' λ_{mfp} value which would lead to the largest fluctuations in the UVB. The values in the plot have been normalized to 1 at 10^6 draws. The three elements have been chosen such that there is one diagonal value and two off-diagonal values in different regions of the matrix. Looking at the correlation matrix in Fig. 6 (which is for a different model but the qualitative behaviour is the same for this model) we see that at all values both on and off the diagonal of the correlation matrix are high and positive, so we expect the convergence for all elements to be roughly the same. At all values of the number of mock draws considered, the covariance elements fall within 3 per cent of their final value. By around $\sim 200\,000$ draws, the values fall within 1 per cent of the final value. For this reason, we believe using $500\,000$ mock draws is sufficient to generate the covariance matrices used in this study. In Fig. A1, $500\,000$ mock draws are represented by the vertical dashed black line.

APPENDIX B: NON-GAUSSIAN DISTRIBUTION OF THE VALUES OF THE AUTOCORRELATION FUNCTION

For our inference, we used the multivariate Gaussian likelihood defined in equation (6). This functional form assumes that the distribution of mock draws of the autocorrelation function is Gaussian distributed about the mean for each bin. In order to visually check this, we will look at the distribution of mock draws from two bins of the autocorrelation function for two different models.

Both Figs B1 and B2 show the distribution of 1000 mock data sets from the velocity bins of the autocorrelation function with $\Delta v = 85.0\,\mathrm{km\,s^{-1}}$ and $\Delta v = 289.0\,\mathrm{km\,s^{-1}}$. The bottom-left panels show the 2D distribution of the autocorrelation values from these bins. The blue (green) ellipses represent the theoretical 68 per cent (95 per cent) percentile contour calculated from the covariance matrix calculated for each model from equation (4). The red crosses show the calculated mean. The top panels show the distribution of only the $v = 289.0\,\mathrm{km\,s^{-1}}$ bins while the right-hand panels show the distribution of only the $v = 85.0\,\mathrm{km\,s^{-1}}$ bins.

Fig. B1 shows mock values of two bins of the autocorrelation function for the model at $z = 5.4$ with $R = 8800$, $\lambda_{\mathrm{mfp}} = 39$ cMpc, and $\langle F \rangle = 0.0801$. These mock data sets consist of 64 quasar sightlines of length $\Delta z = 0.1$. Both the 1D and 2D distributions seem relatively well described by Gaussian distributions by eye though they do show some evidence of non-Gaussian tails to larger values. The numbers of points falling in each contour both fall within 2 per cent of the expected values. In the bottom-left panel with the 2D distribution, there are more mock values falling outside the 95 per cent contour to the top-right side (higher values) than in any other direction. For this reason the distribution is not exactly Gaussian but a Gaussian visually appears as an acceptable approximation.

Fig. B2 shows mock values of two bins of the autocorrelation function for the model at $z = 6$ with $R = 8800$, $\lambda_{\mathrm{mfp}} = 9$ cMpc, and $\langle F \rangle = 0.0089$. These mock data sets consist of 19 quasar sightlines of length $\Delta z = 0.1$. In both the top panel and right-hand panels, which show the distribution of values for one bin of the autocorrelation function, it is clear that the distribution of mock draws is skewed and a Gaussian is not a good approximation for the distributions. This is quantified by the per cent of points in the two ellipses from the bottom-left panel labelled in the top-right panel with 79.0 per cent of the mock draws falling within the 68 per cent contour
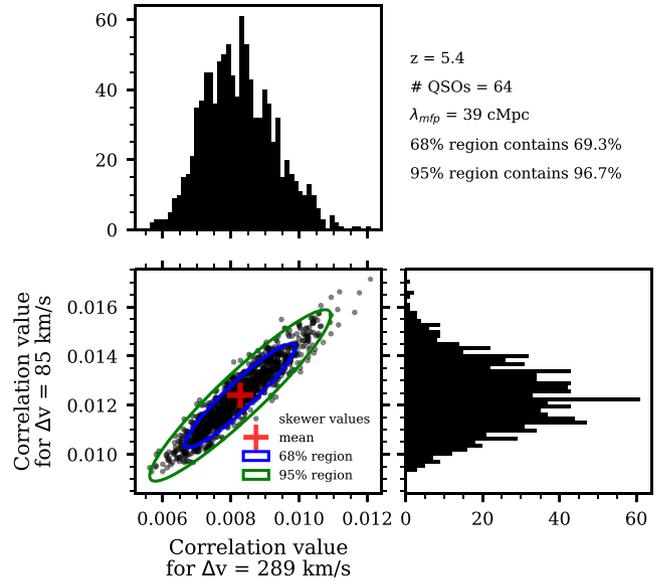
Figure B1. This figure shows the distribution 1000 mock draws from two bins of the autocorrelation function ($\Delta v = 85.0\,\mathrm{km\,s^{-1}}$ and $\Delta v = 289.0\,\mathrm{km\,s^{-1}}$) for one model ($z = 5.4$, $R = 8800$, $\lambda_{\mathrm{mfp}} = 39$ cMpc, and $\langle F \rangle = 0.0801$). The top panel shows the distribution of only the $\Delta v = 289.0\,\mathrm{km\,s^{-1}}$ bin while the right-hand panel shows the distribution of only the $\Delta v = 85.0\,\mathrm{km\,s^{-1}}$ bin. The blue (green) circle represents the 68 per cent (95 per cent) ellipse calculated from the covariance matrix calculated for this model from equation (4). The red plus shows the calculated mean. Additionally the per cent of mock draws that fall within each of these contours is written in the top-right side. Both the 1D and 2D distributions seem relatively well described by a Gaussian distribution. In the 2D plot, there are more points outside the 95 per cent contour to the top-right side than on any other side but it is not extreme.

and 92.2 per cent of the mock draws falling within the 95 per cent contour. The points outside of the contours are highly skewered towards the top-right side (higher values). It is only possible for the autocorrelation function to be negative due to noise, which generally averages to very small values approaching zero at the non-zero lags of the autocorrelation function. However, real fluctuations in the UVB cause the positive fluctuations, making them much more likely and cause the resulting skewed distribution at high z where the overall signal is closer to zero.

Figs B1 and B2 show the changing distribution of the autocorrelation value with λ_{mfp}, $\langle F \rangle$, and mock data set size. There is a greater deviation from a multivariate Gaussian distribution at higher z. It is possible that adding additional sightlines will cause the autocorrelation function to better follow a multivariate Gaussian distribution due to the central limit theorem, though investigating this in detail is beyond the scope of the paper. The incorrect assumption of the multivariate Gaussian likelihood thus contributes to the failure of our method to pass an inference test as discussed in Section 3.4 for $z = 5.4$ and Appendix C for $z = 6$. For our final constraints, we calculated weights for our MCMC chains such that the resulting posteriors do pass our inference test, as discussed in Section 3.4. The whole method of assuming a multivariate Gaussian then re-weighting the posteriors in non-optimal and future work using a more correct likelihood or likelihood-free inference will improve our results.
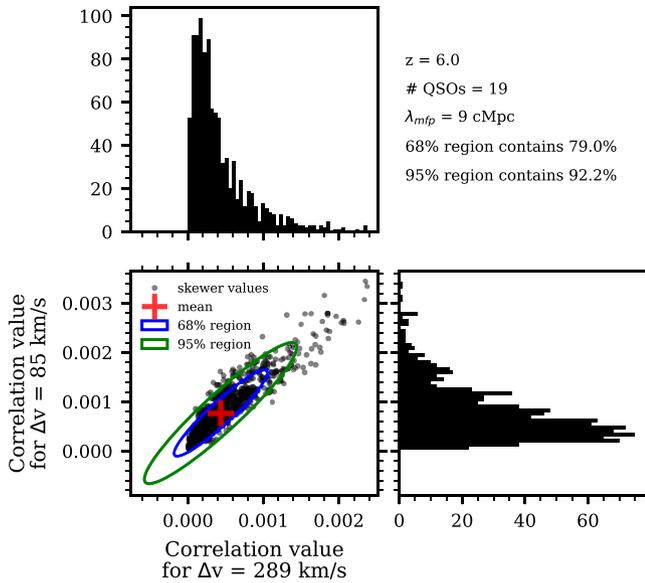
Figure B2. This figure shows the distribution 1000 mock draws from two bins of the autocorrelation function ($\Delta v = 85.0\,\mathrm{km\,s^{-1}}$ and $\Delta v = 289.0\,\mathrm{km\,s^{-1}}$) for one model ($z = 6$, $R = 8800$, $\lambda_{mfp} = 9$ cMpc, and $\langle F \rangle = 0.0089$). The top panel shows the distribution of only the $\Delta v = 289.0\,\mathrm{km\,s^{-1}}$ bin while the right-hand panel shows the distribution of only the $\Delta v = 85.0\,\mathrm{km\,s^{-1}}$ bin. The blue (green) circle represents the 68 per cent (95 per cent) ellipse calculated from the covariance matrix calculated for this model from equation (4). The red plus shows the calculated mean. Additionally the per cent of mock draws that fall within each of these contours is written in the top-right side. Both the 1D and 2D distributions do not seem to be well described by a Gaussian with 79.0 per cent of the mock draws falling within the 68 per cent contour and 92.2 per cent of the mock draws falling within the 95 per cent contour.

APPENDIX C: INFERENCE TEST AT HIGH REDSHIFT

Here, we present the results of the inference test at $z = 6$. This calculation was done following the procedure described in Section 3.4. Fig. C1 shows the results for $z = 6$ and can be compared to the $z = 5.4$ results in Fig. 9. The left-hand panel here shows the initial coverage plot which deviates greatly from the expected $P_{inf} = P_{true}$ line, much more so than the $z = 5.4$. This likely comes from a greater deviation from the assumption of a multivariate Gaussian likelihood as described in Appendix B. The $z = 6$ mock data show highly skewed distributions that are not well described by a Gaussian likelihood.

This initial coverage plot only ever reaches a value of $P_{inf} \approx 0.8$, which becomes an issue for the re-weighting. In the right-hand panel of Fig. C1, the re-weighted inference line thus still only able to reach $P_{inf} \approx 0.8$ creating a plateau in the line once it reaches this value. One way to reach higher values is to increase the number of steps in the MCMC chain. We tried to triple the number of steps but did not see much improvement in the inference test. For computational reasons we stick with the numbers used at other redshifts resulting in 72 000 total steps as described in Section 3.4. This plateau at $P_{true} = 0.8$ means that our 1σ (68th percentile) contours are robust but our 2σ (95th percentile) contours are underestimated since we can only correct up to ~80th percentile.

The inference lines at other redshifts are available upon request. For $5.4 \leq z \leq 5.8$ the coverage plots after re-weighting do not plateau, like the re-weighted coverage plot shown in Fig. 9. Both the $z = 6.0$ and the $z = 5.9$ coverage plots plateau after re-weighting, like that in Fig. C1. This means our re-weighted posteriors at $z = 5.9$ and $z = 6$ may still need additional work to further enlarge probability contours above the value of the plateau.

Figure C1. The left-hand panel of this figure shows the coverage plot resulting from the inference test from 500 models at $z = 6$ and $R = 8800$ drawn from our priors on λ_{mfp} and $\langle F \rangle$. Here, we see that the true parameters for the models fall above the 60th percentile in the MCMC chain ~ 20 per cent of the time, for example. The right-hand panel of this figure shows the coverage plot resulting from the inference test with the use of one set of weights to re-weight the posteriors. With these weights the true parameters for the models fall on the $P_{inf} = P_{true}$ line up to $P_{true} \sim 0.8$. This is because the original coverage plot was only able to reach $P_{inf} \sim 0.8$ so our re-weighting could only match up to this value.

Figure D1. Both panels of this figure show the coverage plot resulting from the inference test from 500 data sets generated by randomly drawing points from the mean model and covariance matrix. The means and covariance matrices used come from $z = 5.4$ and $R = 8800$ in the left-hand panel and $z = 6.0$ and $R = 8800$ in the right-hand panel. The true parameter values for both panels were drawn from our priors on λ_{mfp} and $\langle F \rangle$. In both panels, the Gaussian mock data produced inference lines that fall on top of the 1-1 line within errors, as expected for the statistically correct posteriors.

APPENDIX D: GAUSSIAN DATA INFERENCE TEST

As shown in Appendix B, the distribution of mock values of the autocorrelation function is not exactly Gaussian distributed. In order to confirm the failure of our mock data to pass an inference test (as discussed in Section 3.4 and Appendix C) comes from the use of a multivariate Gaussian likelihood, we generate Gaussian distributed data and run inference tests. For one value of λ_{mfp} and $\langle F \rangle$, we randomly generate a mock data set from a multivariate Gaussian with the given mean model and covariance matrix that we calculated for our mock data in Section 3.1. We can then continue with the inference test as described in Section 3.4. The results for this inference test for $z = 5.4$ and $z = 6.0$ (both with $R = 8800$) are shown in Fig. D1. Here, both redshifts inference lines fall along the 1-1 line that is expected for all probability contour, P_{true}, values. This behaviour is also seen at the other redshifts and $R = 30\,000$. The fact that perfectly Gaussian data pass an inference test with the same likelihood, priors, and method as was used on mock data confirms that the failure of our mock data to pass an inference test is due to the non-Gaussian distribution of the mock data.

APPENDIX E: HIGH-RESOLUTION RESULTS

In Section 4, we only show the posteriors for multiple mock data sets at different redshifts for $R = 8800$. Here, we present the same results but for mock data with $R = 30\,000$. Each violin plot in Fig. E1 is the re-weighted marginalized posterior for one randomly selected mock data set at the corresponding redshift. The light red-shaded region

demarcates the 2.5th and 97.5th percentiles (2σ) of the MCMC draws while the darker red-shaded region demarcates the 16th and 84th percentiles (1σ) of the MCMC draws. Beneath the red violins are blue violins for the posteriors for the same data with $R = 8800$ as shown in Fig. 11. The dot–dashed line is the double power law, equation (1), which we used to determine the true λ_{mfp} evolution as shown in Fig. 2. The random mock data selected for this figure matches exactly with the random mock data used to make Fig. 11. The only difference between the data used in these two figures is the resolution. Generally, the posteriors from the $R = 30\,000$ data shown in Fig. E1 are more precise than those from the $R = 8800$ data.

Again, looking at the posteriors for a given redshift (one column in the figure), the only difference between the posteriors is the random mock data set drawn. These results still have varying precision as is expected from luck of the draw with the mock data sets. There are then three differences between mock data sets shown for a given panel. First is the same as the difference between mocks at one redshift: The mock data is chosen at random so there is just the luck of the draw. The mock data at each redshift also vary with the true λ_{mfp} value, shown in the dot–dashed black line, where the smallest λ_{mfp} value is at the highest z. The autocorrelation function is most precise at small inferred λ_{mfp} values which are more likely at the highest z. Additionally, the redshifts each have different data set sizes, as reported in Table 1. The highest redshifts have the smallest data set sizes, leading to greater scatter in the precision of the posteriors. Again, the individual posteriors are noisy, resulting from the re-weighting procedure as described in Section 3.4.
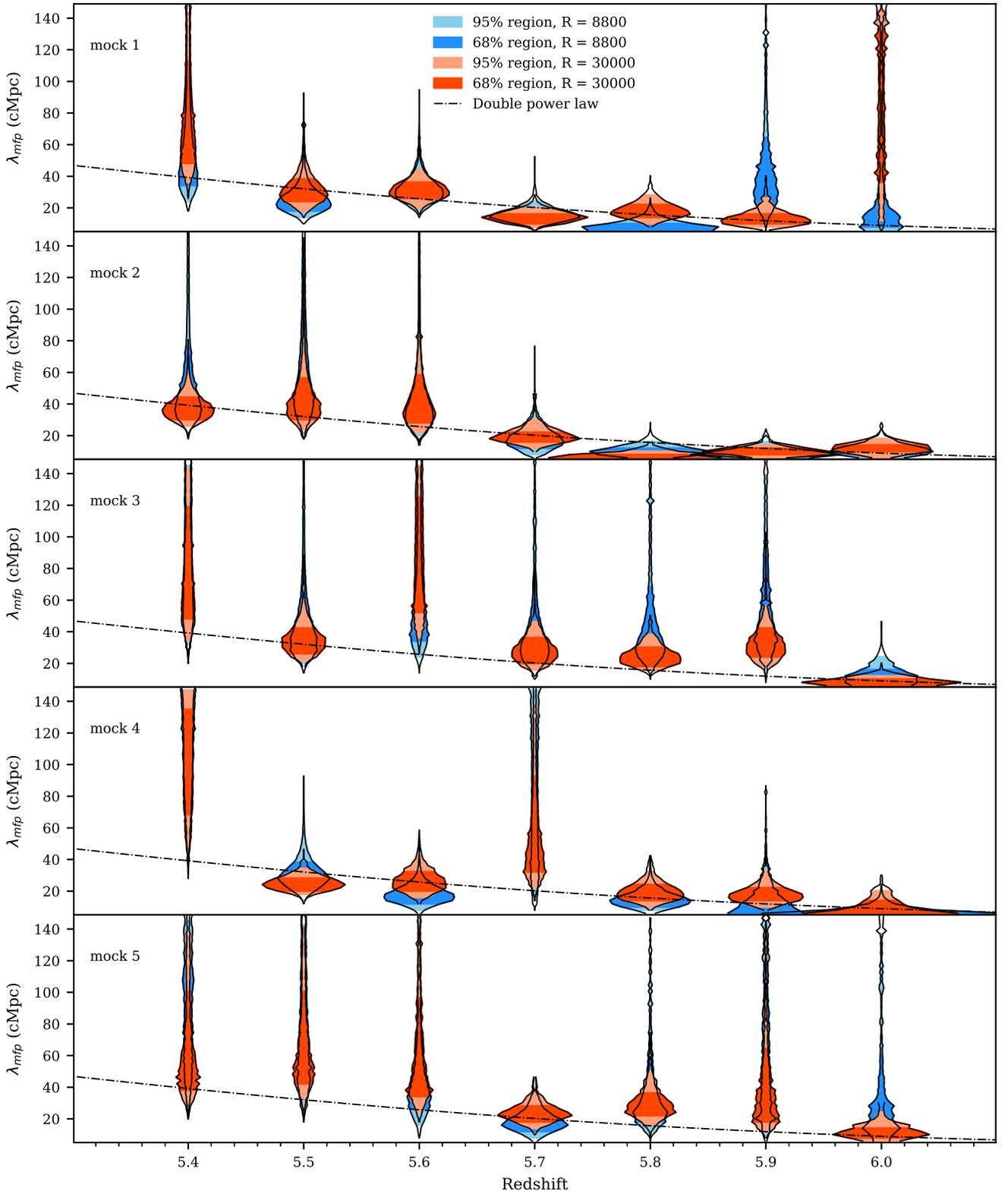
Figure E1. Each panel of this figure shows one posterior for a different randomly selected high-resolution ($R = 30\,000$) mock data set at each z in shades of red. Note that the low-resolution ($R = 8800$) mock data posteriors are plot below the high-resolution posteriors in blue. For each posterior, the light red-shaded region demarcates the 2.5th and 97.5th percentile of the MCMC draws while the dark red-shaded region demarcates the 16th and 84th percentile of the MCMC draws. The black dot–dashed line shows the double power law from equation (1) and Fig. 2. The behaviour of each posterior at the different z is determined by the luck of the draw when selecting the mock data, the true λ$_{mfp}$ value at each z, and the data set size at each z. The true λ$_{mfp}$ values and data set sizes are reported in Table 1.

This paper has been typeset from a T$_E$X/L^{A}T$_E$X file prepared by the author.