



# A measurement of the Ly $\beta$ forest power spectrum and its cross with the Ly $\alpha$ forest in X-Shooter XQ-100

Bayu Wilson<sup>1</sup>,<sup>\*</sup> Vid Iršič<sup>1,2,3</sup> and Matthew McQuinn<sup>1</sup>

<sup>1</sup>*Department of Astronomy, University of Washington, 3910 15th Avenue NE, Seattle, WA 98195-1580, USA*

<sup>2</sup>*Kavli Institute for Cosmology, Department of Physics, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

<sup>3</sup>*Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Avenue, Cambridge CB3 0HE, UK*

Accepted 2021 October 11. Received 2021 September 28; in original form 2021 July 4

## ABSTRACT

The Ly  $\alpha$  forest is the large-scale structure probe for which we appear to have modelling control to the highest wavenumbers. This makes the Ly  $\alpha$  forest of great interest for constraining the warmth/fuzziness of dark matter and the timing of reionization processes. However, the standard statistic, the Ly  $\alpha$  forest power spectrum, is unable to strongly constrain the intergalactic medium (IGM) temperature–density relation, and this inability further limits how well other high-wavenumber-sensitive parameters can be constrained. With the aim of breaking these degeneracies, we measure the power spectrum of the Ly  $\beta$  forest and its cross-correlation with the coeval Ly  $\alpha$  forest using the 100 spectra of  $z = 3.5$ – $4.5$  quasars in the VLT/X-Shooter XQ-100 Legacy Survey, motivated by the Ly  $\beta$  transition’s smaller absorption cross-section that makes it sensitive to somewhat higher densities relative to the Ly  $\alpha$  transition. Our inferences from this measurement for the IGM temperature–density relation appear to latch consistently on to the recent tight lower redshift Ly  $\alpha$  forest constraints. The  $z = 3.4$ – $4.7$  trends we find using the Ly  $\alpha$ –Ly  $\beta$  cross-correlation show a flattening of the slope of the temperature–density relation with decreasing redshift. This is the trend anticipated from ongoing He II reionization and there being sufficient time to reach the asymptotic temperature–density slope after hydrogen reionization completes. Furthermore, our measurements provide a consistency check on IGM models that explain the Ly  $\alpha$  forest, with the cross-correlation being immune to systematics that are uncorrelated between the two forests, such as metal line contamination.

**Key words:** intergalactic medium.

## 1 INTRODUCTION

The Ly  $\alpha$  forest has been used to constrain the Universe’s initial conditions (McDonald et al. 2000; Zaldarriaga, Seljak & Hui 2001; Croft et al. 2002; McDonald 2003; Seljak, McDonald & Makarov 2003; Zaldarriaga, Scoccimarro & Hui 2003; Viel, Haehnelt & Springel 2004b; Viel et al. 2004a; Viel, Weller & Haehnelt 2004c; McDonald et al. 2005, 2006; Seljak, Slosar & McDonald 2006a; Slosar et al. 2011; Busca et al. 2013; Palanque-Delabrouille et al. 2013, 2015; Slosar et al. 2013; Bautista et al. 2015; Baur et al. 2017; Bautista et al. 2017; du Mas des Bourboux et al. 2017), the timing of reionization processes (McDonald et al. 2000; Ricotti, Gnedin & Shull 2000; Schaye et al. 2000; Theuns & Zaroubi 2000; Theuns et al. 2002; Viel & Haehnelt 2006; Bolton et al. 2008, 2010, 2014; Lidz et al. 2010; Becker et al. 2011; Garzilli et al. 2012; Rudie, Steidel & Pettini 2012; Boera et al. 2014; Lee et al. 2015; Upton Sanderbeck, D’Aloisio & McQuinn 2016; Rorai et al. 2017; Hiss et al. 2018; Walther et al. 2019; Wu et al. 2019), and the warmth or fuzziness of the dark matter (Narayanan et al. 2000; Viel et al. 2005, 2008, 2013a; Seljak et al. 2006b; Bird et al. 2011; Baur et al. 2016; Armengaud et al. 2017; Iršič et al. 2017a, b; Yèche et al. 2017; Garzilli et al. 2019a, b; Iršič, Xiao & McQuinn 2020; Rogers & Peiris 2021). Often the Ly  $\alpha$  forest is the standard bearer for the said

constraints. This superiority owes to having modelling control over its spectrum of fluctuations to higher wavenumbers than all other established large-scale structure probes (e.g. McQuinn 2016).

However, it is likely that better constraints can be extracted from intergalactic Lyman-series absorption. Many studies have found that large degeneracies in the Ly  $\alpha$  forest’s constraints on parameters (particularly those constrained best by the highest wavenumbers probed) when using the standard statistic, the power spectrum. Namely, significant degeneracies exist between any two of the following: the particle mass in warm/fuzzy dark matter models, the gas temperature at mean density, and the trend in temperature with density (e.g. Lidz et al. 2010; Becker et al. 2011; Iršič et al. 2017b). The latter two thermal parameters constrain reionization (as well as any other heating processes), and the degeneracies are so severe that the trend of temperature with density is essentially unconstrained by previous Ly  $\alpha$  power spectrum analyses. While some degeneracies can be broken by measurements at multiple redshifts (McDonald et al. 2005), substantially improving constraints over existing ones likely either requires (1) combining Ly  $\alpha$  power spectrum measurements with that of other Lyman-series transitions such as Ly  $\beta$  (Dijkstra, Lidz & Hui 2004) or (2) using Ly  $\alpha$  absorption statistics beyond just the power spectrum (Zaldarriaga, Seljak & Hui 2001; Fang & White 2004; Gaikwad et al. 2020).

Here, we present the first measurement of the Ly  $\beta$  forest auto-power spectrum at the wavenumbers most sensitive to the gas temperature, as well as its cross-power spectrum with the Ly  $\alpha$

\* E-mail: [bwils033@ucr.edu](mailto:bwils033@ucr.edu)

forest. The cross-power spectrum is estimated using a Fourier transformation of Ly  $\alpha$  and Ly  $\beta$  absorption features. An intervening gas overdensity creates both a Ly  $\alpha$  and Ly  $\beta$  forest absorption feature but the Ly  $\beta$  forest would also contain lower redshift Ly  $\alpha$  absorption that biases the Ly  $\beta$  autopower spectrum. The cross power turns out to be our most constraining new diagnostic because the effective noise for the Ly  $\beta$  forest, set by the lower redshift Ly  $\alpha$  absorption, is higher than Ly  $\alpha$  forest. The Ly  $\beta$  transition has a smaller cross-section for absorption than Ly  $\alpha$ , which makes it more sensitive to higher density gas for which the Ly  $\alpha$  absorption is more saturated (Dijkstra et al. 2004; Iršič & Viel 2014). This sensitivity to higher densities breaks the aforementioned degeneracies, most obviously between the temperature at mean density ( $T_0$ ) and its power-law trend with density (with index  $\gamma - 1$ ).

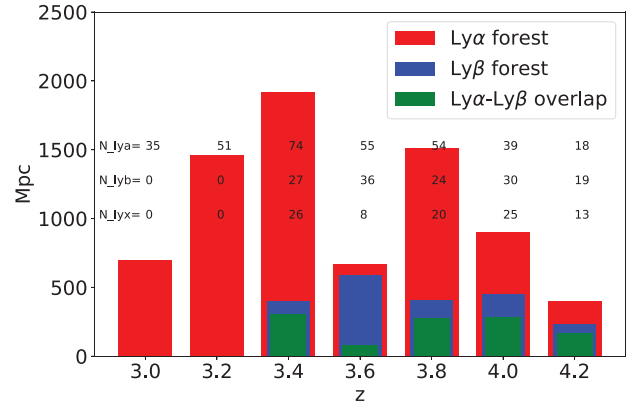
A precision measurement requires a large sample of Ly  $\beta$  forest spectra. However, this goal is hindered by the shorter path-length probed by each sightline relative to the Ly  $\alpha$  forest (with about a third as much of useful absorption), by foreground Ly  $\alpha$  forest absorption that contaminates the Ly  $\beta$  forest, and by the blueness of the Ly  $\beta$  transition that results in this forest being less likely to be captured than the Ly  $\alpha$  forest in existing quasi-stellar object (QSO) spectra. Perhaps as a result of these difficulties, there has been only one attempt to measure the Ly  $\beta$  power spectrum: Iršič et al. (2013) presented a measurement using 60 000 SDSS/BOSS quasar spectra. The low resolution of these spectra, with  $\mathcal{R} \equiv \lambda/\Delta\lambda \approx 2000$ , inhibits constraining thermal scales and breaking the associated degeneracies. Here, we measure the Ly  $\beta$  forest power spectrum (and its cross with the Ly  $\alpha$  forest) using the 100,  $\mathcal{R} \sim 10^4$  spectra observed as part of the VLT/XSHOOTER XQ-100 survey (López et al. 2016). We also present a preliminary measurement using the  $\mathcal{R} \sim 10^5$  VLT/UVES SQUAD DR1 sample (Murphy et al. 2019).

Additional motivations for measuring the Ly  $\beta$  forest power spectrum, and especially its cross with Ly  $\alpha$ , are (1) to test consistency with parameters derived with Ly  $\alpha$  forest and (2) to further test the standard paradigm that cosmological simulations reproduce the low-density intergalactic medium (built principally from comparing these simulations to Ly  $\alpha$  forest power spectrum measurements). The Ly  $\alpha$  forest power spectrum at high wavenumbers can be contaminated by metal lines owing to their smaller thermal widths (e.g. Lidz et al. 2010) and there is some controversy over the severity of this effect (Day, Tytler & Kambalur 2019). The cross-power spectrum is unbiased by contaminants that are not correlated between the two forests, including the majority of metals (with a small bias for select transitions that fall near Ly  $\alpha/\beta$ ) as well as instrumental noise. Thus, the cross power could serve as a check on parameters derived from the Ly  $\alpha$  forest alone.

This paper is organized as follows: Section 2 describes the data analysis pipeline, as well as how mock spectra are created and used to test the pipeline. Section 3 describes how the power spectrum is estimated and how it is corrected for noise and resolution, concluding by presenting our power spectrum estimates. We also compare our XQ-100 measurement with a preliminary Ly  $\beta$  measurement using archival VLT/UVES spectra. Section 4 interprets our measurements in terms of the IGM thermal history. A series of appendices present details regarding our resolution, noise, DLAs, and metal corrections.

## 2 DATA SELECTION AND MOCKS

Our primary measurement uses the XQ-100 Legacy Survey (López et al. 2016), consisting of one hundred  $3.4 < z < 4.7$  QSO spectra observed with the X-Shooter spectrograph on the Very Large Telescope (Vernet et al. 2011). The number of quasars and path-



**Figure 1.** The total comoving path-length of our XQ-100 sample for the Ly  $\alpha$  forest, Ly  $\beta$  forest, and the intersection of these two paths (used to calculate the cross power), using the selection criteria outlined in Section 2. The annotations give the total number of quasars that contribute in each redshift bin to the three path-lengths. For the highest wavenumbers used in our power spectrum estimates, these distances are marginally smaller as we use spectra for which the resolution correction is the most certain.

lengths of this sample are presented in Fig. 1, quoting these quantities in the  $\delta z = 0.2$  redshift bins we use in our analysis.

Each quasar was observed for 2–12 exposures with a 1.0 arcsec slit for the UV arm or a 0.9 arcsec slit for the VIS arm. The wavelengths for each exposure were calibrated off of known skylines (Appendix C). The individual exposures then used nearest grid point interpolation to average the exposures on to a  $20 \text{ km s}^{-1}$  grid for the UV arm and  $11 \text{ km s}^{-1}$  for the visible arm. In our mocks, discussed shortly, we replicate this binning to show that our measurements are not affected. This binning is chosen to be somewhat smaller than the full width at half-maximum (FWHM) spectral resolution of the X-shooter spectrograph for a fully illuminated slit of  $56 \text{ km s}^{-1}$  for the UV arm and  $34 \text{ km s}^{-1}$  for the VIS arm. There is an overlap region at  $5500\text{--}5600 \text{ \AA}$  where both arms of the spectrograph process a significant fraction of the light, corresponding to Ly  $\alpha$  at  $z \approx 3.6$ . We exclude this region from our analysis, which results in significantly less Ly  $\alpha$  (and hence cross) data in the  $z = 3.6$  redshift bin (see Fig. 1). (This cut also results in this redshift bin always using data from the spectrograph’s VIS arm.)

Our analysis uses quasar continuum estimates developed by the XQ-100 Legacy Survey team (Berg et al. 2016), which fitted a spline over several wavelength ranges within each spectrum. We use these estimates to calculate the continuum-normalized flux in the Lyman forests. The same continuum estimate was used in the previous XQ-100 Ly  $\alpha$  forest analysis (Iršič et al. 2017c).

Our Ly  $\alpha$  forest measurement uses the pixels within the  $1045\text{--}1185 \text{ \AA}$  QSO-frame wavelength range, a range chosen to omit absorption in the broad Ly  $\alpha$  and Ly  $\beta$  emission lines of the quasar where continuum fitting can be more challenging (e.g. McDonald et al. 2005). Furthermore, this cut omits regions where the ionizing background is enhanced even modestly by the QSO owing to the proximity effect. For similar reasons, our measurement of the Ly  $\beta$  forest uses the QSO-frame wavelength range  $978\text{--}1014 \text{ \AA}$ , following Iršič et al. (2013). This wavelength range for Ly  $\beta$  corresponds to using pixels somewhat closer to the quasar than in Ly  $\alpha$  (to an equivalent wavelength of  $1202 \text{ \AA}$  in the Ly  $\alpha$  forest) as the Ly  $\beta$  line of the QSO is less broad than its Ly  $\alpha$  counterpart.

We mask regions around Damped Ly  $\alpha$  (DLA) systems using the DLA sample of the XQ-100 survey team (Sánchez-Ramírez et al.

2016). 30 per cent of sightlines show a DLA, with the probability that a DLA falls in both Ly  $\alpha$  and Ly  $\beta$  reduced by the path-length ratio (Fig. 1). We do not use data within 1/2 the equivalent width of Ly  $\alpha$  and Ly  $\beta$  line centres of each DLA and additionally correct the mean flux outside of this range for the wings of the line. Appendix B quantifies the effect of this masking on our measurements.

Our data analysis pipeline to estimate the mean flux and power spectra of the Ly  $\alpha$  and Ly  $\beta$  forests was tested with synthetic Lyman-forest data that were generated following the method for creating synthetic Ly  $\alpha$  forest mocks presented in more detail in Iršič et al. (2017b), with the most significant difference here being the inclusion of Ly  $\beta$  forest absorption. In summary, we generate a realistic flux field with a QSO redshift distribution matching that of the XQ-100 sample as well as approximating the XQ-100 pixel, resolution, and noise specifications. 5000 light-cone spectra are created using the simulation outputs spaced at  $\Delta z = 0.1$ . We use the Sherwood simulation suite of high-resolution hydrodynamical simulations, with  $2 \times 2048^3$  particles in a  $40 \text{ Mpc } h^{-1}$  box (Bolton et al. 2017). This simulation appears to be converged in its estimate for the Ly  $\alpha$  forest power spectrum at the redshifts of interest to better than  $\sim 5$  per cent (Bolton et al. 2017; Iršič et al. 2017a). The mean flux of our mocks is rescaled to match measurements. In addition to testing our pipeline, bootstraps of our mocks are used for calculating the covariance matrix.

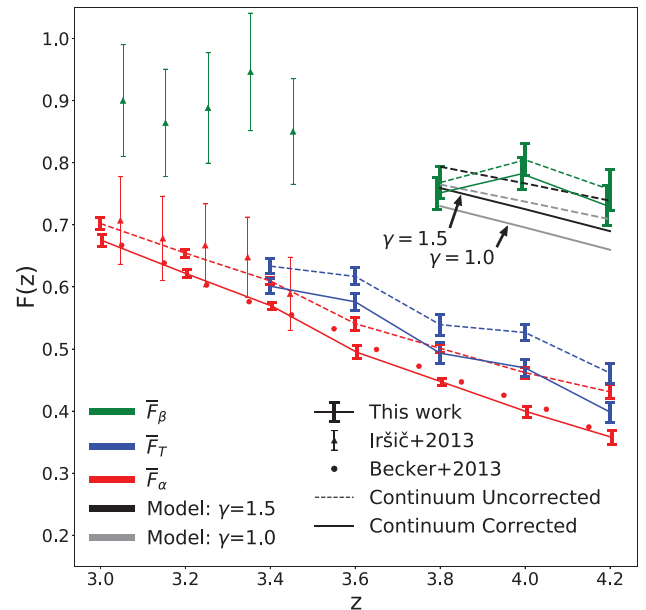
### 3 DATA ANALYSIS AND MEASUREMENT

This section describes both our mean flux and power spectrum measurements. The same analysis strategy was adopted for both real and synthetic data.

#### 3.1 Mean flux

We average the continuum-normalized flux (i.e. the estimated transmission) in all pixels that fall into a redshift bin to obtain the mean transmission in Ly  $\alpha$  ( $\bar{F}_\alpha$ ) and Ly  $\alpha$  + Ly  $\beta$  ( $\bar{F}_T$ ). Error bars are estimated by bootstrap resampling. Each sample constitutes pixels from a sightline's Ly  $\alpha$  forest spectrum within a redshift bin. The mean transmission in Ly  $\beta$  ( $\bar{F}_\beta$ ) in a redshift bin is estimated by dividing the value of  $\bar{F}_\alpha$  for the foreground redshift that contributes Ly  $\alpha$  absorption from  $\bar{F}_T$ . We only measure the foreground  $\bar{F}_\alpha$  and, hence, can perform this subtraction for our three highest Ly  $\beta$  redshifts. The error bars that are connected with dashed lines in Fig. 2 show our measurements of  $\bar{F}_\alpha$ ,  $\bar{F}_T$ , and  $\bar{F}_\beta$  (red, blue, and green curves, respectively).

The most significant systematic in estimating the mean flux is the placement of the quasar continuum. The error bars connected with the solid lines are our mean flux measurement corrected for the human bias from by-eye continuum fitting, using the correction factor in Faucher-Giguère et al. (2008). Again red, blue, and green solid curves, respectively, show our continuum-corrected estimates for  $\bar{F}_\alpha$ ,  $\bar{F}_T$ , and  $\bar{F}_\beta$ . The Faucher-Giguère et al. (2008) correction,  $\Delta C/C_{\text{true}} = 1.58 \times 10^{-5}(1+z)^{5.63}$ , was estimated by fitting the continua of mock spectrum corrections to the mean flux in Ly  $\alpha$ . While this correction was estimated on mocks that simulated the Keck/ESI spectra assuming FWHM =  $40 \text{ km s}^{-1}$  and S/N = 20, these specifications are not dissimilar to our X-Shooter spectra. Faucher-Giguère et al. (2008) further found that the corrections were similar if they considered higher S/N and higher resolution mock spectra reminiscent of Keck/HIRES. However, the Faucher-Giguère et al. (2008) estimates could overestimate the true continuum correction



**Figure 2.** Mean flux estimates for Ly  $\beta$  ( $\bar{F}_\beta$ ; green), Ly  $\alpha$  + Ly  $\beta$  at redshift that corresponds to Ly  $\beta$  ( $\bar{F}_T$ ; blue), and Ly  $\alpha$  ( $\bar{F}_\alpha$ ; red). The error bars that are linked by dashed lines are those estimated here using the XQ-100 data, and those linked with solid lines show our measurement corrected for the human bias in continuum placement using the correction of Faucher-Giguère et al. (2008). For comparison, the triangles with error bars are the  $\bar{F}_\alpha$  and  $\bar{F}_\beta$  estimates of Iršič et al. (2013) using SDSS/BOSS, and the circles are the  $\bar{F}_\alpha$  estimates of Becker et al. (2013) again on the SDSS/BOSS data using a novel method for removing continuum (which results in minute statistical error bars). The thin grey and black ‘lines’ spanning  $z = 3.8$ – $4.2$  show the predicted mean flux in Ly  $\beta$  ( $\bar{F}_\beta$ ) using simulations that have, respectively,  $\gamma = 1.5$  and  $1.0$  and our measured  $\bar{F}_\alpha$ . The solid and dashed versions of these lines calibrate the simulations to our continuum-corrected and continuum-uncorrected values of  $\bar{F}_\alpha$ . The fact that Becker et al. (2013) measurement falls near our continuum-corrected estimates makes us favour our solid lines.

as they used low-resolution  $N$ -body simulations to model the forest and a steep  $T$ – $\Delta$  relation with  $\gamma = 1.6$ .

Fig. 2 compares our measurements with those of Iršič et al. (2013) and Becker et al. (2013). Both of these measurements use different methodologies than our more traditional mean transmission measurement. The Iršič et al. (2013) measurements are done by fitting a parametric model to the measured power spectrum in a sample of thousands of SDSS/BOSS quasars rather than directly measuring the mean transmission. Also applied to the SDSS/BOSS sample, the Becker et al. (2013) measurement uses a different method still that estimates the mean transmission using stacked quasar spectra. This measurement assumes that in aggregate the stack’s mean continuum shows little redshift evolution. Both methods likely avoid continuum overfitting issues, which is a major systematic in our measurement. The measurements of Iršič et al. (2013) have large error bars and are consistent at the  $1\sigma$  level with our  $\bar{F}_\alpha$ . Their  $\bar{F}_\beta$  measurements do not overlap in redshift with ours. The results by Becker et al. (2013) fall much closer to our continuum-corrected estimate.<sup>1</sup> However, we

<sup>1</sup> Another common correction made in mean flux measurements is for metal absorption contamination. We do not apply such a correction in our analysis. Metals result in a 9 per cent correction at  $z = 3$  and 5 per cent at  $z = 4$  using the metal correction estimates based on direct identification of Schaye et al. (2003), and a 6 per cent at  $z = 3$  and 2 per cent at  $z = 4$  using the statistical



note that the uncertainty in the continuum correction divides out in our power spectrum measurements (presented in the next section) and so is not a concern there.

These mean flux measurements alone have the potential to constrain the intergalactic temperature–density relation. Fig. 2 investigates this possibility. The thin grey and black lines spanning  $z = 3.8$ – $4.2$  show the predicted mean flux in Ly  $\beta$  ( $\bar{F}_\beta$ ) using simulations that have, respectively, temperature–density relations with power-law slope  $\gamma = 1.5$  and  $1.0$  and our measured Ly  $\alpha$  mean flux (which falls approximately on a single line at these redshifts). These values for  $\gamma$  span most of the theoretically motivated range of  $1$ – $1.6$  (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016). The solid and dashed versions of these lines calibrate the simulations to the continuum-corrected and continuum-uncorrected values of the Ly  $\alpha$  mean flux.<sup>2</sup> As the grey and solid models span the  $1\sigma$  error bar of our measured  $\bar{F}_\beta$  (the connected green points with error bars), we conclude that our mean flux measurements alone are not strongly constraining of  $\gamma$ . There is a slight preference to  $\gamma = 1.5$ , especially if one takes the Becker et al. (2013) Ly  $\alpha$  mean flux measurement to indicate that our continuum-corrected measurement is closer to the truth, as one should expect. The value  $\gamma = 1.5$  is also closer to what we find in our power spectrum analysis (Section 4).<sup>3</sup>

### 3.2 Power spectra

To estimate the power spectrum from the data, we follow the standard approach of Fourier transforming segments of our data that pass our cuts (Croft et al. 1999, 2002; Kim et al. 2004; Viel et al. 2004b, 2013a). A disadvantage of this approach is that the power spectrum estimate is not weighted optimally to minimize errors. Additionally, if the analysis uses segments that are too short, the cut-offs at the edges result in spurious high- $k$  power. The other approach that has been adopted is to use a quadratic estimator, which mitigates these effects (McDonald et al. 2005). Quadratic estimators have predominantly been used only for SDSS data sets as they have difficulty converging for smaller data sets for which the power spectrum is less constrained. Because our data set is small relative to SDSS, we adopt the direct Fourier transform approach.

The flux power spectrum used in the analysis has been calculated in five redshift bins, each with  $\delta z = 0.2$ , spanning  $3.4 < z < 4.2$ . For each of the quasar spectra, we sort the data into redshift bins that correspond to Ly  $\alpha$  and Ly  $\beta$  absorption in that redshift window. Spectral segments that fall within a given redshift bin are selected for the power estimate in the bin. The flux in each segment is divided by our mean flux estimate,  $\bar{F}_X$ , to convert to an overdensity

$$\hat{\delta}_{F_X}^s(\lambda, z_i) = \hat{F}_X^s(\lambda) / \bar{F}_X(z_i) - 1, \quad (1)$$

where  $X = \{\alpha, \beta\}$  denotes which Lyman-series forest is measured and  $s$  indexes the quasar spectrum. Remember our convention that

results of Tytler et al. (2004). For the principal aim of our analysis, a Ly  $\alpha$  and Ly  $\beta$  power spectrum measurement, the mean flux that is used to calculate the flux overdensity does not need to be corrected for metal absorption. Metals would not affect our  $\bar{F}_\beta$  estimates to the extent that the mean metal absorption does not differ between the coeval Ly  $\alpha$  and Ly  $\beta$  forests.

<sup>2</sup>We note that simulations with different values for the temperature at the mean density,  $T_0$ , predict essentially the same  $\bar{F}_\beta$  when calibrated to the same  $\bar{F}_\alpha$ .

<sup>3</sup>The inferred Ly  $\alpha$  mean flux from the much different power spectrum analysis is the same to a percent fractional level to the continuum-corrected curve. This analysis did use a prior centred on Becker et al. (2013) with a 5 per cent error.

$T$  indicates the Ly  $\beta$  forest plus the foreground Ly  $\alpha$  that falls in the same spectral region. The  $\bar{F}_X(z_i)$  are our estimates for the mean flux presented in 3.1 (which are uncorrected for continuum bias and metal absorption as these then essentially cancel out in the computation of 1). Next,  $\hat{\delta}_{F_X}^s(\lambda, z_i)$  is Fourier transformed yielding  $\tilde{\delta}_{F_X}^s(k, z_i)$  and the auto- and cross-power spectrum is estimated as

$$\begin{aligned} \hat{P}_{XY}(k, z_j) = N_k^{-1} \sum_{\substack{|k_i - k| < \frac{\Delta k}{2} \\ s \in \mathcal{S}(k)}} \left\{ \left[ \tilde{\delta}_{F_X}^s(k_i, z_j) \tilde{\delta}_{F_Y}^s(k_i, z_j)^* L_s^{z_i} \right. \right. \\ \left. \left. - P_{N,X}^s(k_i, z_j) \delta_{X,Y}^K \right] W_{s,X}(k_i, z_j)^{-1} W_{s,Y}(k_i, z_j)^{-1} \right. \\ \left. - \hat{P}_{M,X}(k, z_j) \delta_{X,Y}^K \right\}, \quad (2) \end{aligned}$$

where the sum runs over all  $N_k$  modes that fall in the band power in all segments that correspond to the desired redshift bin,  $\mathcal{S}(k)$  is the sample of all quasar spectra where the effective resolution is reliably known to estimate the power at  $k$ ,  $L_s^{z_i}$  is the length of each spectral segment,  $W_{s,X}$  are the kernels correcting for the effects of spectral resolution and pixel size, and  $P_{N,X}^s$  ( $\hat{P}_{M,X}$ ) are the noise (estimated metal power). The details of the resolution kernel and noise/metal power are described below. Since the noise between different Lyman-series forests is uncorrelated, it only contributes to the autopower and, hence, the Kronecker delta function yields  $\delta_{X,Y}^K = 1$  if  $(X = Y)$  and 0 otherwise. Band power measurements are made in 13 logarithmic wavenumber bins, with bin centres spanning the range of  $-2.5 \leq (k/\text{km}^{-1} \text{ s}) \leq -1.2$ . For cross-power spectra ( $X \neq Y$ ), the power spectra can be imaginary if translational invariance is broken, which can occur because of resonant metal contamination or because of imperfect wavelength calibration. Appendix C uses the value of the imaginary to test the wavelength calibration. We finally note that the minimum variance estimator would weight by the signal-to-noise squared, but since the statistics at all wavenumbers we report are limited by sample variance rather than detector noise, the above estimator should essentially be minimum variance. In what follows, we provide details regarding the treatment of noise, metal absorption, instrumental resolution, and wavelength calibration.

#### 3.2.1 Noise

For the noise power spectrum  $P_{N,X}^s$  in equation (2), we assume it is white such that  $P_{N,X}^s(k, z) = \sigma_s(z)^2 w_X(k, z)^2$ , where  $\sigma_s(z)^2$  is the noise variance in each spectral segment for quasar  $s$  in redshift bin  $z$  and  $w_X(k, z) \equiv \text{sinc}(k \Delta v_X^z / 2)$  owes to the boxcar spectral bins with velocity width  $\Delta v_X^z$ . The noise power  $P_{N,X}^s$  is at least two orders of magnitude below the Ly  $\alpha$  power spectrum signal (Iršič et al. 2017c), and there is no noise correction for our cross-power measurement. We have also tested a correction that accounts for spatial inhomogeneities in the noise and concluded that the associated correction would be negligible.

#### 3.2.2 Metals

Our measurements are at sufficiently high redshifts that the contamination from metal absorbers is a percent-level correction to the total power, a correction well below our quoted error bars. Furthermore, the Ly  $\alpha$ –Ly  $\beta$  cross power is unbiased by non-resonant metal absorption. Nevertheless, for the autopower, we do correct for non-resonant metals using the standard procedure of using the absorption redward of the forest in our spectra to subtract their contribution as represented in equation (2). The Kronecker delta

function that multiplies the metal power  $\hat{P}_M(k_i, z_j)$  in that equation is only non-zero for the autopower spectra. For the power spectrum of the metals  $\hat{P}_M(k_i, z_j)$ , we use the measurement of Iršič et al. (2017c) with the same XQ-100 data set from the red-side power spectrum, which decreases our autopower by a few per cent (Appendix B). We further find that the contamination from the resonantly enhanced metals, defined by that they fall near our Ly  $\alpha$  and Ly  $\beta$  (namely O VI  $\lambda\lambda$  1032, 1038 Å and Si III 1207 Å), is at a similar level, and we do not correct for resonant metals. See Appendix B for additional details.

### 3.2.3 Resolution

The correction for spectral resolution is complex as slit spectrographs have resolutions that depend on the illumination of the slit and, hence, the seeing of individual observations. Yet, for IGM thermal constraints, it is advantageous to use as high a wavenumber as possible, but high wavenumbers are also the most sensitive to uncertainties in the spectral resolution. This issue has led to there being debate over the reliability of XQ-100 inferences from the Ly  $\alpha$  forest power spectrum, especially inferences from the lower resolution UV arm (Iršič et al. 2017c; Yèche et al. 2017; Walther et al. 2019). Here, we develop a better model for the seeing-dependent resolution of XQ-100. Our approach suggests that some of the previous discrepancy in the quoted X-Shooter resolution may owe to the functional form that was assumed for the line spread function (Appendix A).

We model the line spread function  $W_s(k_i, z_i)$  as a Gaussian multiplied by a sinc function:

$$W_{s,X} = e^{-\frac{1}{2}k^2\sigma_R(X,s,z)^2} w_X(k, z), \quad (3)$$

where  $w_X(k, z) = \text{sinc}(k\Delta v_X^z/2)$  with  $v_X^z = 11$  (20) km s $^{-1}$  for wavelengths that fall on the visible (ultraviolet) arm. The Gaussian kernel that multiplies  $w_X$  approximates the Fourier transform of the X-Shooter line spread function. In Appendix A, we show that really the line spread function is better modelled as a Gaussian convolved with a top-hat function. We calibrate this model off of arc-lamp spectra. However, we find that the Fourier transform of our model to yield  $W_{s,X}$  is well approximated over the measured wavenumber range by a Gaussian with standard deviation  $\sigma_R(X, s, z)$ .

Our model for  $\sigma_R$  depends on the seeing conditions for each spectrum. Among the exposures combined for a single quasar spectrum, the seeing can vary significantly (although for 80 per cent of the quasars the FWHM of seeing varies by  $<0.2$  arcsec over the exposures). We estimate the minimum and maximum of  $\sigma_R$  using the minimum and maximum seeing of the exposures on an individual quasar. We discard modes measured from spectra where using  $\sigma_R^{\min}$  rather than  $\sigma_R^{\max}$  to estimate the  $W_{s,X}W_{s,Y}$  would lead to a 10 per cent difference in the estimated power; i.e. we discard modes for which

$$[\sigma_R^{\max}(X, s, z)^2 + \sigma_R^{\max}(Y, s, z)^2] - [\sigma_R^{\min}(X, s, z)^2 + \sigma_R^{\min}(Y, s, z)^2] > 0.1 k^{-2}.$$

This selection criterion is combined with a 20 per cent allowance for uncertainty in the mean  $\sigma_R$  when constraining thermal parameters in Section 4.

### 3.2.4 Wavelength calibration

Our Ly  $\beta$  power spectrum requires wavelengths to be calibrated to the accuracy of 10(7) km s $^{-1}$  in order to make a 20 (10) per cent error at the bin centre of the maximum wavenumber we report,  $k =$

10 $^{-1.2}$  s km $^{-1}$ . These errors would be approximately halved in our next-to-largest wavenumber bin. These numbers hold both if the wavelength calibration is systematically offset or Gaussian random between sightlines with the accuracy quoted above being the standard deviation (Appendix C).

The wavelength calibration of our XQ-100 data set is done using skylines, first calibrated on a master integration and then adjusted for each exposure, and interpolating using the pipeline used by the XQ-100 team (López et al. 2016). Tests show that the precision is likely better than 5 km s $^{-1}$ , with the dominant error being interpolation and being more significant for the UV arm where there are fewer skylines (George Becker, private communication). There can also be offsets owing to the positioning of the source within the slit for each arm; any offsets from centre would result in a shift for that arm. The quoted VLT/X-Shooter precision of the alignment of the arms indicates that such offsets are likely controlled to a few km s $^{-1.4}$  and, if correct, such offsets would not be a significant systematic.<sup>5</sup>

We can test that the wavelength calibrations likely meet the required calibration level by using the imaginary component of the Ly  $\alpha$ –Ly  $\beta$  cross-power spectrum (Appendix C). Our measurement of this statistic limits any systematic offset to  $\lesssim 10$  km s $^{-1}$ , with some evidence for an offset at this level at  $z = 3.8$  and 4.0. We have rerun our analysis presented in Section 4.2 with a 5 km s $^{-1}$  correction to offset this apparent shift, as the error is quadratic in this shift and so this splits the difference, and find our results for temperature only change at the  $\sim 0.1\sigma$  level. Part of this insensitivity is because of our large allowance for resolution error also effectively increases the variance at the highest wavenumbers that are affected by such a shift. The imaginary component of the cross power is less sensitive to positive and negative offsets that are random between each of the spectra, with our measurements of the imaginary power suggesting that the random offsets have an RMS of  $<15$  km s $^{-1}$ .

Seven of the XQ-100 quasars were observed without their atmospheric dispersion corrector, which corrects for differential atmospheric refraction. A lack of correction could result in larger offsets, particular for objects observed with large zenith angles. We have done the analysis with and without these quasars included and find negligible differences.

### 3.2.5 Masks

After masking to account for our rest-frame wavelength cuts and DLA contamination, the minimum contiguous number of pixels in a segment that we still use to make a power spectrum measurement is 100 pixels. A minimum allowed segment reduces the extra power at high wavenumbers that owes to the discontinuities at the end of each segment. We wrote a lognormal mock code to understand what biases result from this cut (and from the non-periodicity of the segment at the edge), finding that even if all of our segments are at the 100 pixel threshold, the biases are negligible:  $\lesssim 3$  per cent for the UVB arm and double this bias for the VIS where 100 pixels corresponds to a shorter path-length. In our data, only  $\sim 10$  per cent of segments tend

<sup>4</sup>[https://www.eso.org/sci/facilities/paranal/instruments/xshooter/doc/XS\\_wlc\\_shift\\_150615.pdf](https://www.eso.org/sci/facilities/paranal/instruments/xshooter/doc/XS_wlc_shift_150615.pdf)

<sup>5</sup>Another potential wavelength offset occurs owing to fine structure of the transitions shifts the wavelengths by  $\Delta\lambda_\alpha = 0.006$  Å (1.5 km s $^{-1}$ ) and  $\Delta\lambda_\beta = 0.002$  Å (0.6 km s $^{-1}$ ), which are insignificant. We use  $\lambda_\alpha = 1215.67$  Å and  $\lambda_\beta = 1025.72$  Å as our mean vacuum wavelengths.

to fall within a factor of 2 of this minimum pixel threshold and so these percentiles significantly overestimate the effect.<sup>6</sup>

### 3.2.6 Covariance matrix

We estimate the measurement covariance matrix by first bootstrap sampling our mocks to measure the covariance cross-correlation coefficient matrix

$$\hat{C}_{ij} = \langle (\hat{P}_i - \langle \hat{P}_i \rangle) (\hat{P}_j - \langle \hat{P}_j \rangle) \rangle, \quad (5)$$

where  $i, j$  enumerate both the types of power spectra,  $P_{\alpha\alpha}, P_{TT}, P_{\alpha T}$ , the redshift bins, and the  $k$ -bin. We use the 3000 mock spectra to estimate the full covariance, denoting this as  $\hat{C}_{ij}^M$ , although we also will use the diagonals of the covariance matrix measured with XQ-100, which we denote as  $\hat{C}_{ii}^{XS}$ . We further set correlations between different redshift bins to zero as these should be nearly zero. Because our mocks use the same skewers through different times in the simulations, there are spurious correlations that this step suppresses. We evaluate the correlation coefficient from the mock sample:

$$\hat{r}_{ij}^M \equiv \frac{\hat{C}_{ij}^M}{\sqrt{\hat{C}_{ii}^M \hat{C}_{jj}^M}}, \quad (6)$$

which carries the information about the off-diagonal structure in the covariance matrix. As the last step, we rescale the correlation coefficient from the mocks by the diagonal elements of  $\hat{C}_{ii}^{XS}$ , as measured on the XQ-100 data. This rescaling allows us to potentially capture additional variance that is not in our mocks (such as from e.g. continuum fitting errors or large-scale modes not in our simulations) and also to likely take out some of the model dependence of the mock covariance matrix. Since  $\hat{C}_{ii}^{XS}$  is measured on far fewer sightlines than the mock covariance matrix, there is roughly 10 per cent scatter around the mean relation from the mocks (re-scaled to the same path-length). In the subsequent Monte Carlo Markov chain (MCMC) analysis, we have tested that replacing  $\hat{C}_{ii}^{XS}$  with the mean values from the re-scaled mock sample does not impact the conclusions of this paper. Due to re-sampling from the fixed pool of sightlines, the bootstrap method can underestimate the variance (Rollinde et al. 2013; Viel, Schaye & Booth 2013b; Iršič et al. 2017c). This effect is more severe for the redshift bins with shorter path-length. To correct for that effect, we multiply the full covariance matrix by a factor of

$$\hat{C}_{ij}^2 = [\hat{C}_{ij}^{XS}]^2 \left( 1 + 0.3 \sqrt{\frac{\Delta x_{\max}}{\Delta x(z_i)}} \right) \left( 1 + 0.3 \sqrt{\frac{\Delta x_{\max}}{\Delta x(z_j)}} \right), \quad (7)$$

where  $\Delta x(z_{i,j})$  denotes the path-length for the type of power spectrum ( $P_{\alpha\alpha}, P_{\alpha T}, P_{TT}$ ), index  $i, j$  corresponds to the redshift bin, and  $\Delta x_{\max}$  is the maximum path-length in the sample of our redshift bins and spectra. The path-lengths used in this rescaling correspond to the values in Fig. 1, with  $\Delta x_{\max}$  corresponding to the path-length of Ly  $\alpha$  forest at  $z = 3.4$ . This procedure effectively boosts the covariance matrix by an average factor of 1.44 (1.9) for the covariance corresponding to  $P_{\alpha\alpha}$  ( $P_{\alpha T}$ ), with larger boosts for smaller path-lengths. This corresponds to the boost in the power spectrum errors by roughly 20 per cent (40 per cent) on average for the  $P_{\alpha\alpha}$  ( $P_{\alpha T}$ ). The amplitude of 0.3 is chosen to reproduce the typical value of 0.44 in the XQ-100 Ly  $\alpha$  analysis of Iršič et al. (2013) and the functional

form for this correction is motivated by a sampling argument.<sup>7</sup> We note that our results are not significantly changed if we instead use a constant  $(1 + 0.44)^2$  enhancement over all redshifts as in Iršič et al. (2013) rather than the terms in parentheses in equation (7).

In the MCMC analysis, we also include a systematic error budget owing primarily to uncertainty in the resolution measurement. The systematic error is modelled as uncorrelated – contributing in quadrature only to diagonal elements of the covariance matrix.<sup>8</sup> In this simple model, the resolution uncertainty in UVB ( $\Delta\sigma_R^{\text{UVB}}/\sigma_R^{\text{UVB}}$ ) and VIS ( $\Delta\sigma_R^{\text{VIS}}/\sigma_R^{\text{VIS}}$ ) arms is propagated to the power spectrum measurement between  $X$  and  $Y$  fields as

$$\left( \frac{\sigma_{P_{XY}}(k, z_j)}{\hat{P}_{XY}(k, z_j)} \right)^2 = \sum_{W \in \{X, Y\}} \frac{\Delta\sigma_R}{\sigma_R} \bigg|_{z_j^W} (k \sigma_R(z_j^W))^2, \quad (8)$$

with the sum accounting for the contributions of both fields  $X$  and  $Y$ . In the case of autopower ( $X = Y$ ), the two terms in the sum are identical. Each of the parts in the sum depends on the *observed wavelength* range for that transition, as the distinction between UVB and VIS arms is in the frame of the spectrograph. Thus, each of the arms covers a range of redshifts of absorptions that depends on the transition. For  $W = \alpha$ , the redshifts are already expressed for the Ly  $\alpha$  transition, and so  $z_j^W = z_j$ . For the case of  $W = T$ , the redshifts have to be shifted by the ratio of the wavelengths of the Ly  $\beta$  and Ly  $\alpha$  transitions, and thus  $1 + z_j^W = (1 + z_j)\lambda_\beta/\lambda_\alpha$ . We then imposed the condition that if  $z_j^W < 3.6$  the contribution is coming from the UVB arm. This is exact, as in our measurement of the  $z = 3.6$  bin we have only included data from the VIS arm (Section 2). For our analysis, we take  $\Delta\sigma_R^{\text{UVB}}/\sigma_R^{\text{UVB}} = \Delta\sigma_R^{\text{VIS}}/\sigma_R^{\text{VIS}} = 0.2$ , as motivated earlier.

### 3.3 Measurement

Fig. 3 shows the resulting power spectrum estimates, with the diagonal errors computed from bootstrapping the data in the five redshift bins in which the Ly  $\beta$  forest can be measured. For our three redshift bins with  $z \geq 3.8$ , we are able to isolate the Ly  $\beta$  autopower spectrum as we are able to estimate and then subtract the foreground Ly  $\alpha$  power from  $\hat{P}_{TT}$ . These estimates, which are the purple errors in Fig. 4, are not exactly the Ly  $\beta$  autospectrum  $\hat{P}_{\beta\beta}$  as  $P_{TT}$  is not just a sum of  $P_{\alpha\alpha}$  and  $P_{\beta\beta}$  but rather given by

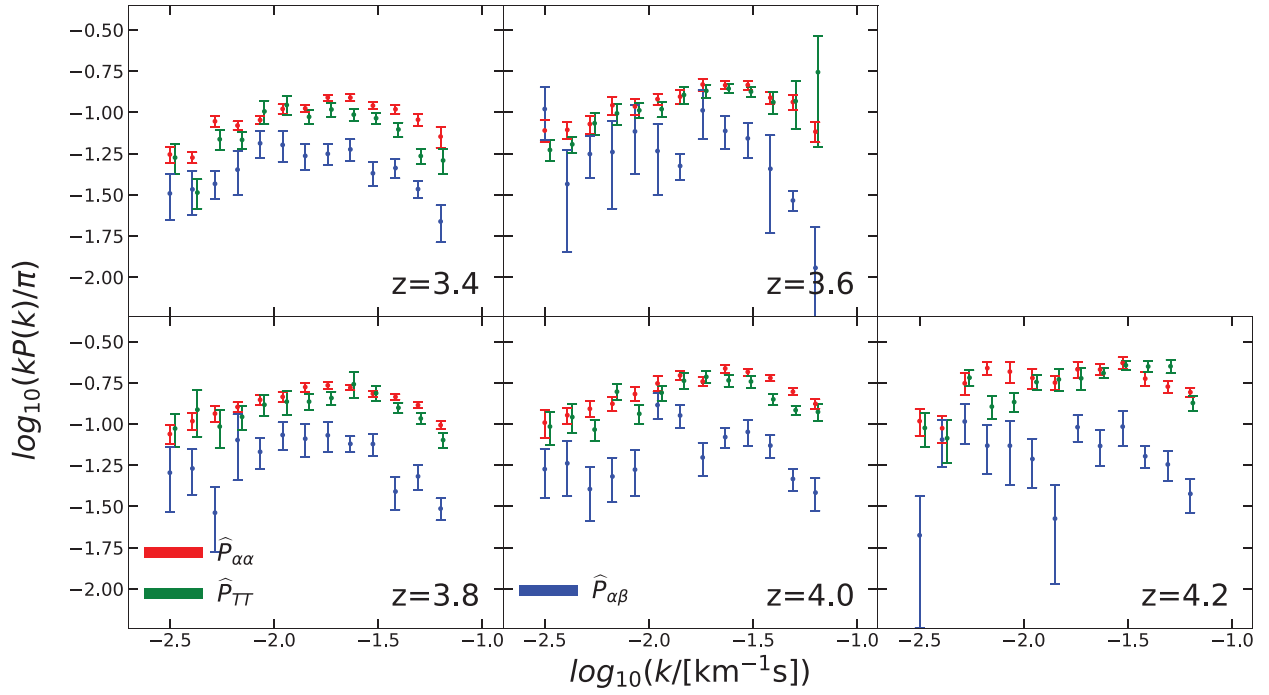
$$P_{TT} = P_{\alpha\alpha}(k, z_f) + \overbrace{P_{\beta\beta}(k, z) + \int \frac{dk'}{2\pi} P_{\beta\beta}(k', z) P_{\alpha\alpha}(k - k', z_f)}^{\mathcal{P}_{\beta\beta}}, \quad (9)$$

where  $z_f \equiv \lambda_\beta/\lambda_\alpha(1 + z) - 1$ . Thus, we denote what we measure when we subtract  $P_{\alpha\alpha}$  as  $\mathcal{P}_{\beta\beta}$ , which is the quantity indicated by the overbrace in the previous formula. A rough estimate for the size of  $P_{\beta\beta}$ -contaminating convolution term is  $\sigma_\alpha^2 P_{\beta\beta}(k, z)$ , where  $\sigma_\alpha^2 \sim$

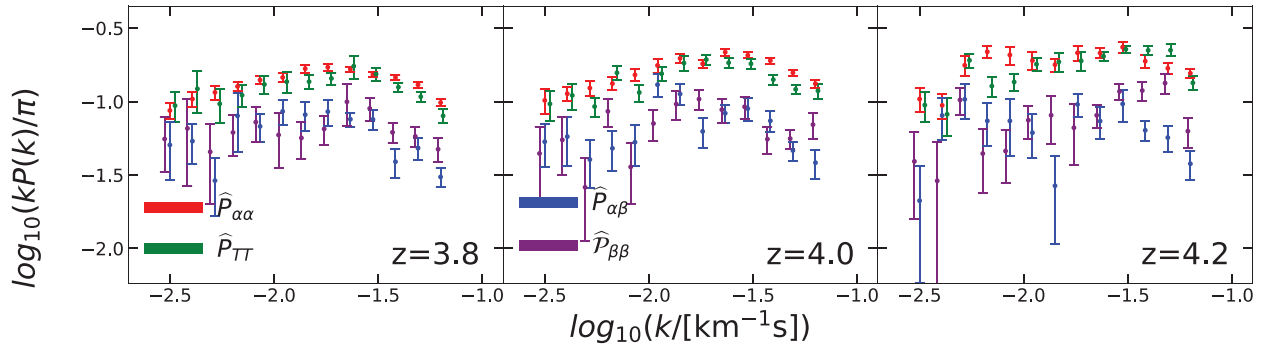
<sup>7</sup>The boost aims to account for that our estimate for the covariance matrix diagonals should have an error that scales as the inverse of the number of samples, which we assume is proportional to the path-length  $\Delta x(z)$ . Equation (7) is adding back this typical error so that we are unlikely to substantially underestimate the covariance in any redshift bin.

<sup>8</sup>Generally, one would expect that the resolution uncertainty induces correlated systematic uncertainty across the sightlines, as the uncertainty is fixed per wavelength calibration and thus the same for all sightlines. This is further complicated in the presence of seeing corrections that could add sightline-to-sightline variations. While these effects are accounted for in the mean measurement, a simplistic model is sufficient for the covariance matrix that we use in the MCMC analysis of this paper.

<sup>6</sup>This bias can be substantial at higher wavenumbers that are used in high-resolution data sets.



**Figure 3.** The panels show our estimates for  $\hat{P}_{\alpha\alpha}$ ,  $\hat{P}_{TT}$ , and  $\hat{P}_{\alpha T} (= \hat{P}_{\alpha\beta})$  in the five redshift bins. Note that  $\hat{P}_{\alpha T}$  is equivalent to  $\hat{P}_{\alpha\beta}$  as the two Ly  $\alpha$  forest segments are at such large distances that they essentially do not correlate. The error bars do not include the allowance for 20 per cent uncertainty in the resolution parameter  $\sigma_R$  and also the additional error for undersampling used in our final analysis (cf. Section 3.2).



**Figure 4.** The same as Fig. 3 but where we also show  $\hat{P}_{\beta\beta} \equiv P_{TT}(z) - P_{\alpha\alpha}(z_f)$ , where  $z_f \equiv \lambda_\beta/\lambda_\alpha(1+z) - 1$ . The three panels correspond to the  $z$  where  $P_{\alpha\alpha}(z_f)$  is estimated so that this subtraction can be performed within our measurements. Note that  $\hat{P}_{\beta\beta}$  is not exactly the Ly  $\beta$  autopower spectrum as there is a convolution term that contributes at the 10 per cent level (cf. equation 9).

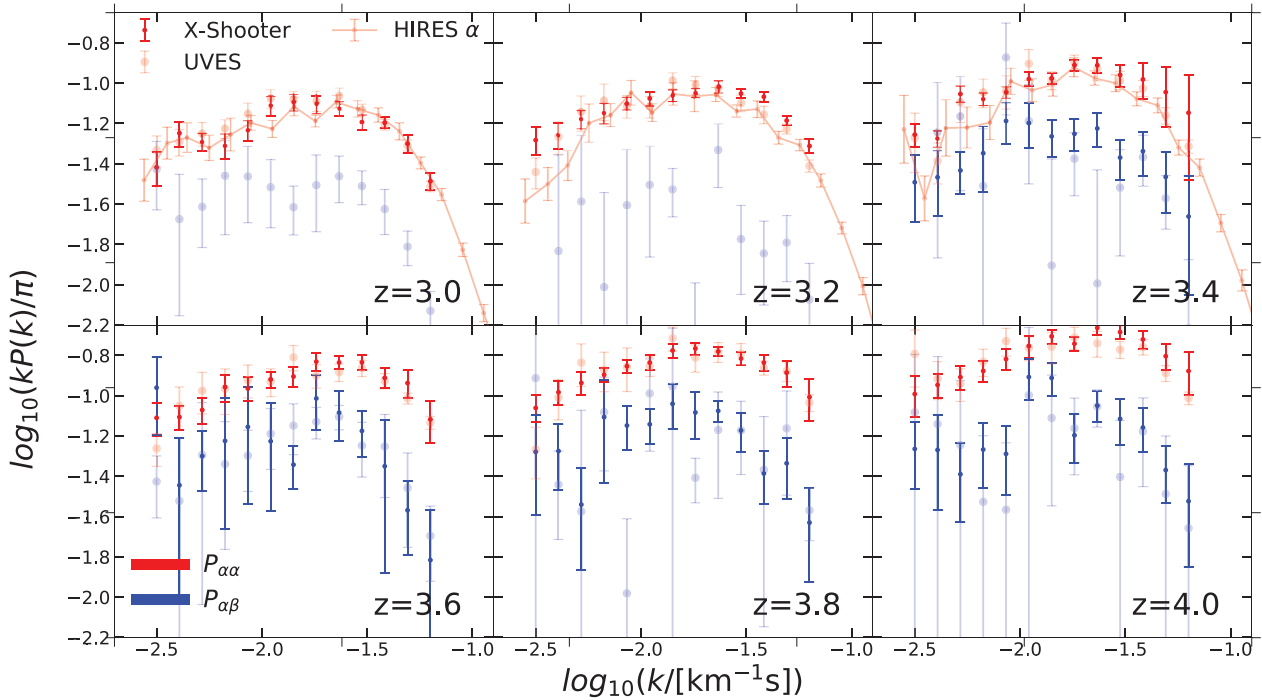
$kP_{\alpha\alpha}(k)/\pi \sim 0.1$  approximates the variance in the Ly  $\alpha$  forest over our surveyed redshifts. (We have done full calculations that verify this estimate.) This additional contribution needs to be modelled for accurate inference from  $\mathcal{P}_{\beta\beta}$ , although we suspect generally it will be more useful to forward model  $P_{TT}$ . As our measurement of  $P_{\alpha\beta}$  is more constraining, we do not use  $\mathcal{P}_{\beta\beta}$  nor  $P_{TT}$  in our analysis to estimate the thermal properties of the IGM (Section 4). Note that  $P_{\alpha T}$  is equivalent to  $P_{\alpha\beta}$  as the two Ly  $\alpha$  forest segments are at such large distances that they essentially do not correlate, and so we subsequently refer to this cross power as  $P_{\alpha\beta}$ .

Fig. 5 compares our measurement to the Keck/HIRES  $P_{\alpha\alpha}$  measurement of Walther et al. (2018) using the KODIAQ reductions (O’Meara et al. 2015) and a preliminary measurement of  $P_{\alpha\alpha}$ ,  $P_{\alpha\beta}$ ,  $P_{TT}$  using the SQUAD DR1 reductions of the VLT/UVES archival data (Murphy et al. 2019). Both the HIRES and UVES data sets are at significantly higher resolution (with  $\langle R \rangle \sim 50\,000$ ) than

our XQ-100 data, which makes the resolution correction essentially negligible over the wavenumbers we report for the XQ-100 data.<sup>9</sup> We first concentrate on the comparison with the Walther et al. (2018) Keck/HIRES measurement of  $P_{\alpha\alpha}$ . The top panels show the power at  $z = 3.0, 3.2$ , and  $3.4$ . The former two redshifts are lower than the redshifts where we measure Ly  $\beta$  power, but they overlap with this Keck analysis. Our Ly  $\alpha$  power spectrum measurement agrees better with this Keck/HIRES measurement in the first two redshift bins, but overshoots in the  $z = 3.4$  bin at the highest wavenumbers. A similar overshoot has been seen before when comparing with the

<sup>9</sup>The purpose of including the preliminary VLT/UVES measurement in this paper is to evaluate the spectral resolution correction for the XQ-100 data. The joint analysis between the VLT/UVES and XQ-100 data is left for a follow-up paper.





**Figure 5.** A comparison of our XQ-100 power spectrum measurements with the Keck/HIRES  $P_{\alpha\alpha}$  measurement of Walther et al. (2018) at  $z = 3.0, 3.2$ , and  $3.4$  and our measurement of  $P_{\alpha\alpha}$  and  $P_{\alpha\beta}$  using the VLT/UVES archival data in all the redshift bins. The systematic error for resolution has been added to the X-Shooter measurements, except at  $z = 3.0$  and  $3.2$   $P_{\alpha\alpha}$ , which are not used in our thermal parameter analysis.

XQ-100 measurement of Iršič et al. (2017c), with Walther et al. (2018) contending that it owed to Iršič et al. (2017c) using a value of  $\sigma_R$  that is too small. While our resolution correction is different from Iršič et al. (2017c), the results are not dissimilar. In our analysis (and the error bars shown here), we allow for a 20 per cent uncertainty in the resolution parameter  $\sigma_R$  that mitigates this discrepancy such that the overshoot is now within the error bar. Note that this error is particularly significant for the lower resolution UV arm that is used for the  $z = 3.4$   $P_{\alpha\alpha}$  measurement.

Next, let us compare with our preliminary measurement of the VLT/UVES SQUAD data set. This was analysed using a slightly adapted pipeline as that used here for XQ-100. We further chose a minimum continuum-to-noise of 20 in a  $2.5 \text{ km s}^{-1}$  pixel at  $5500 \text{ \AA}$ . This resulted in our UVES analysis considering 62  $z > 3$  quasars, about half of the full SQUAD sample at these redshifts. In future work, we intend to use lower quality spectra to improve the S/N, and more thoroughly investigate systematics that crop up especially in the lower quality spectra. However, one can see that our UVES measurement is less precise at higher redshifts for estimating the cross power; we do not expect this property to change. Rather, the VLT/UVES data set excels at lower redshifts. This figure also presents  $z = 3$  and  $3.2$  cross-power estimates with VLT/UVES, lower redshifts than where this can be measured with X-Shooter. By and large, the VLT/UVES measurements appear consistent with X-Shooter measurements. We take this agreement as further support for our new X-Shooter resolution correction.

#### 4 THERMAL HISTORY ESTIMATION

While the temperature at the mean gas density has been routinely measured, more previous measurements of  $P_{\alpha\alpha}$  were not sensitive

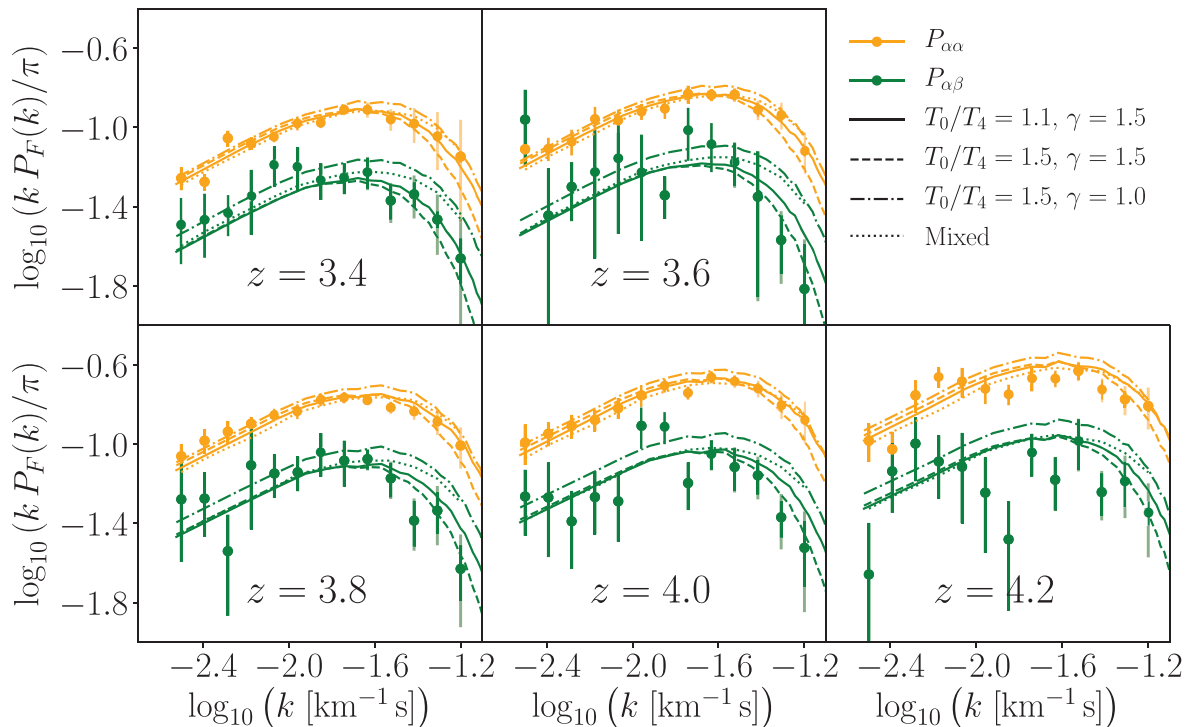
enough to draw conclusions on the slope of the temperature–density relation  $\gamma - 1 \equiv \text{dln } T / \text{dln } \rho$ . This stems from the fact that  $P_{\alpha\alpha}$  is sensitive to a fairly narrow range of gas densities, and effectively traces the temperature at one characteristic gas overdensity (Becker et al. 2011). To the extent that it does not probe a range of overdensities, the measurement of the temperature is not sensitive to the trend with density (Lidz et al. 2010; Becker et al. 2011; McQuinn et al. 2011; Iršič & McQuinn 2018). To alleviate this problem, several suggestions have been put forward including (a) other statistics than the two-point function of the Ly  $\alpha$  forest (Dijkstra et al. 2004; Bolton et al. 2014; Hiss et al. 2018; Telikova, Balashev & Shternin 2018; Gaikwad et al. 2020) or (b) bluer Lyman-series transitions (Dijkstra et al. 2004; Iršič et al. 2013; Iršič & Viel 2014; Boera et al. 2016). We follow the latter, presenting the first measurements of the IGM thermal history from the Ly  $\beta$  forest power spectra and its cross with Ly  $\alpha$ .

#### 4.1 Simulations

In our Bayesian analysis, we sample the parameter space  $\{\bar{F}(z_i), T_0(z_i), \gamma(z_i), z_{\text{rei}}, \sigma_8, n_{\text{eff}}\}$ , where  $z_i$  goes over the observed redshift bins (3.4, 3.6, 3.8, 4.0, and 4.2), using MCMC sampler based on the Metropolis–Hastings algorithm (e.g. Iršič et al. 2017a). All parameters are computed self-consistently in the simulation (and not in post-processing), so effects like heating from structure formation shocks are captured. To compare the models to the measurement, we compute mock Ly  $\alpha$  forest absorption spectra from a suite of hydrodynamical simulations and, then, compute from each simulation their 1D flux power spectra, namely  $P_{\alpha\alpha}$ ,  $P_{\alpha\beta}$ , and  $P_{\beta\beta}$ .

The grid of simulations used to cover the parameter space is described in detail in Iršič et al. (2017a). The simulations were





**Figure 6.** Model predictions (and not best-fitting models!) for several temperature–density relations (curves) alongside our XQ-100 measurement (points with error bars). Three of the models show power-law temperature–density relations specified by  $T_0/T_4$ ,  $\gamma$ , where  $T_4 \equiv 10^4$  K. In addition, a ‘Mixed’ model, where the IGM is filled half with gas at  $(T_0/T_4, \gamma) = (0.7, 1.5)$  and half with  $(T_0/T_4, \gamma) = (1.5, 1.0)$ , exaggerates the features anticipated for midway through He II reionization. That the differences between the various models are larger in  $P_{\alpha\beta}$  compared to  $P_{\alpha\alpha}$  illustrates how our measurement of  $P_{\alpha\beta}$  has the potential to break degeneracies in the thermal model.

run using the GADGET-2 cosmology  $N$ -body + Smooth particle hydrodynamics code with  $768^3$  dark matter particles and  $768^3$  gas particles in a  $20 \text{ Mpc } h^{-1}$  comoving box. Additional simulations were run using  $2 \times 1024^3$  particles in the same box size to calibrate a correction for the mass resolution. The resulting corrected power spectra are converged in both resolution and box size to within 5 per cent over the redshift and wavenumber range ( $k < 0.1 \text{ s km}^{-1}$ ) considered in our measurements (Iršič et al. 2017a).

Fig. 6 showcases the power spectra computed from a sample of our simulations (curves), alongside our flux power spectrum measurements (points with error bars). We caution, however, of the direct comparison here of models and data, as the models are not best fit and such parameters like mean flux have not been chosen carefully. The simulations span a large range of the IGM mean temperatures and temperature–density relations (see the  $T_0$  and  $\gamma$  labels), and the flux power spectra are sensitive to the differences between the model at highest wavenumbers ( $k > 0.02 \text{ s km}^{-1}$ ). Moreover, the differences between the models are larger for the cross-power spectra,  $P_{\alpha\beta}$  compared to the Ly  $\alpha$  autopower spectrum  $P_{\alpha\alpha}$ . Therefore, the addition of Ly  $\beta$  power spectrum measurements is likely to increase the sensitivity to the IGM parameters. We explicitly show this to be the case in the next section.

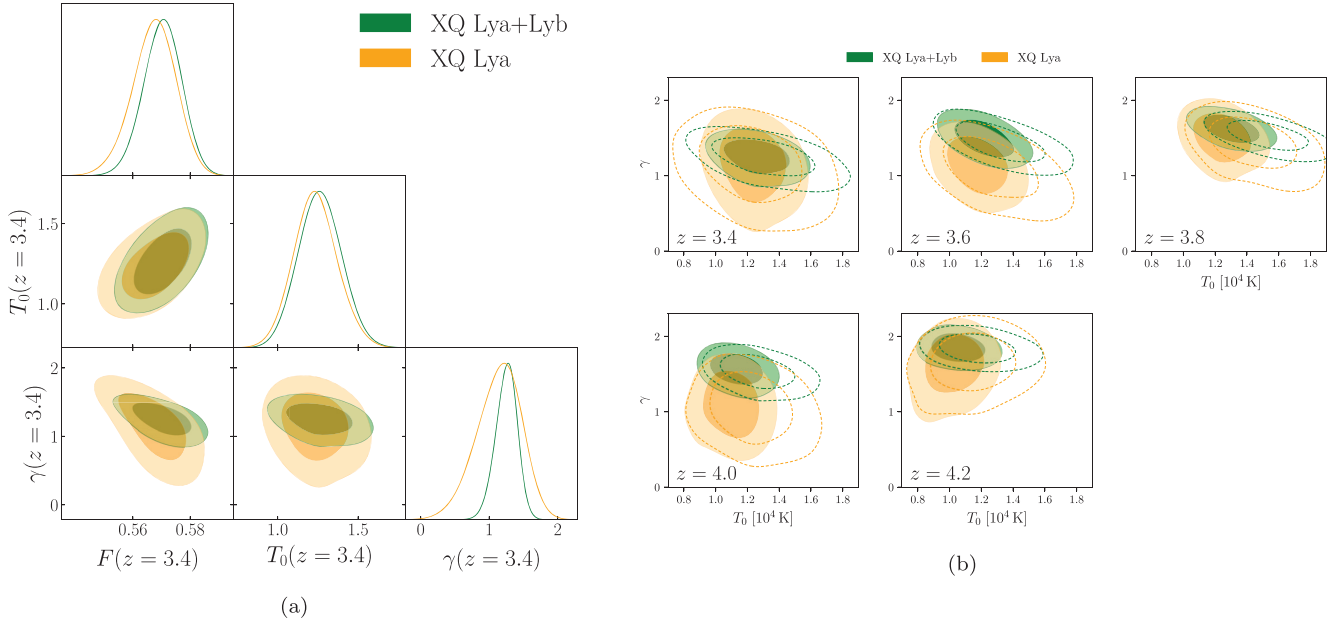
Also shown in Fig. 6 is a ‘Mixed’ case, which combines two simulations so the power spectrum is computed so that half the skewers are through a simulation with  $(T_0, \gamma) = (7000 \text{ K}, 1.5)$  and the other half from a simulation with  $(T_0, \gamma) = (15000 \text{ K}, 1.0)$ . This mixed case idealizes the situation that may be expected halfway through (inhomogeneous) He II reionization (McQuinn et al. 2011). Qualitatively, the flattening does not appear consistent with our measurements of the cross power, but the mixed model likely

exaggerates the actual picture and, again, this is not a best-fitting model. Our best-fitting models presented in Section 4.2 are shifted somewhat down relative to the models predictions here.

## 4.2 Results

Here, we describe the IGM thermal parameter analysis that results from the MCMC analysis. The analysis fits for all the redshifts together, with three IGM parameters per redshift bin ( $\bar{F}$ ,  $T_0$ ,  $\gamma$ ) and three global parameters that are not changing with redshift: the redshift of instantaneous reionization  $z_{\text{rei}}$  in the simulations (which is a standard parameter in such analyses adopted as a proxy for the amount of Jeans smoothing of the gas by sound waves) and the two cosmological parameters  $\sigma_8$  and  $n_{\text{eff}}$  (where the latter non-standard parameter is the slope of the matter power spectrum at  $k = 1 \text{ h Mpc}^{-1}$ ). For all the results in this section, we fix the cosmological parameters by imposing a tight Gaussian prior around the values as measured by Planck+2018 ( $\sigma_8 = 0.811 \pm 0.006$  and  $n_{\text{eff}} = -2.30 \pm 0.005$ ). While these two cosmological parameters are marginalized over in our analysis, including them makes little difference in the final results. For the redshift of reionization, we adopt a flat prior over  $6 < z < 14$ . We generally find that this parameter is not well constrained, which is expected as our measurements are well after the end of reionization.

Furthermore, we have used Gaussian priors on the mean transmission in the Ly  $\alpha$  forest at each redshift centred at the measurement of Becker et al. (2013) with 5 per cent standard deviation. This 5 per cent standard deviation is motivated by the differences we find when correcting and not correcting for continuum and larger than the differences between Becker et al. (2011) and our mean flux



**Figure 7.** Left: Contour plot for the thermal parameters at  $z = 3.4$  using our  $\hat{P}_{\alpha\alpha}$  estimates alone (orange) and these plus our  $\hat{P}_{\alpha\beta}$  estimates (green). The constraint on  $\gamma$  is especially improved by the inclusion of the cross-power spectrum. Right: Contour plot in the  $T_0$ - $\gamma$  plane for all the measured redshifts. The colour scheme is the same as in the left-hand panel, and  $T_0$  is again in units of  $10^4$  K. The full contours are for the same analysis as in the left-hand panel, while the dashed contours are showing the results when weaker  $T_0$  priors were used in the analysis (T3; using  $\Delta T_0 = 3000$  K as the standard deviation in our prior). The constraints on  $T_0$  are largely set by the  $T_0$  prior, but constraints on  $\gamma$  are not driven by any  $\gamma$  prior nor are they shaped by which  $T_0$  prior is used.

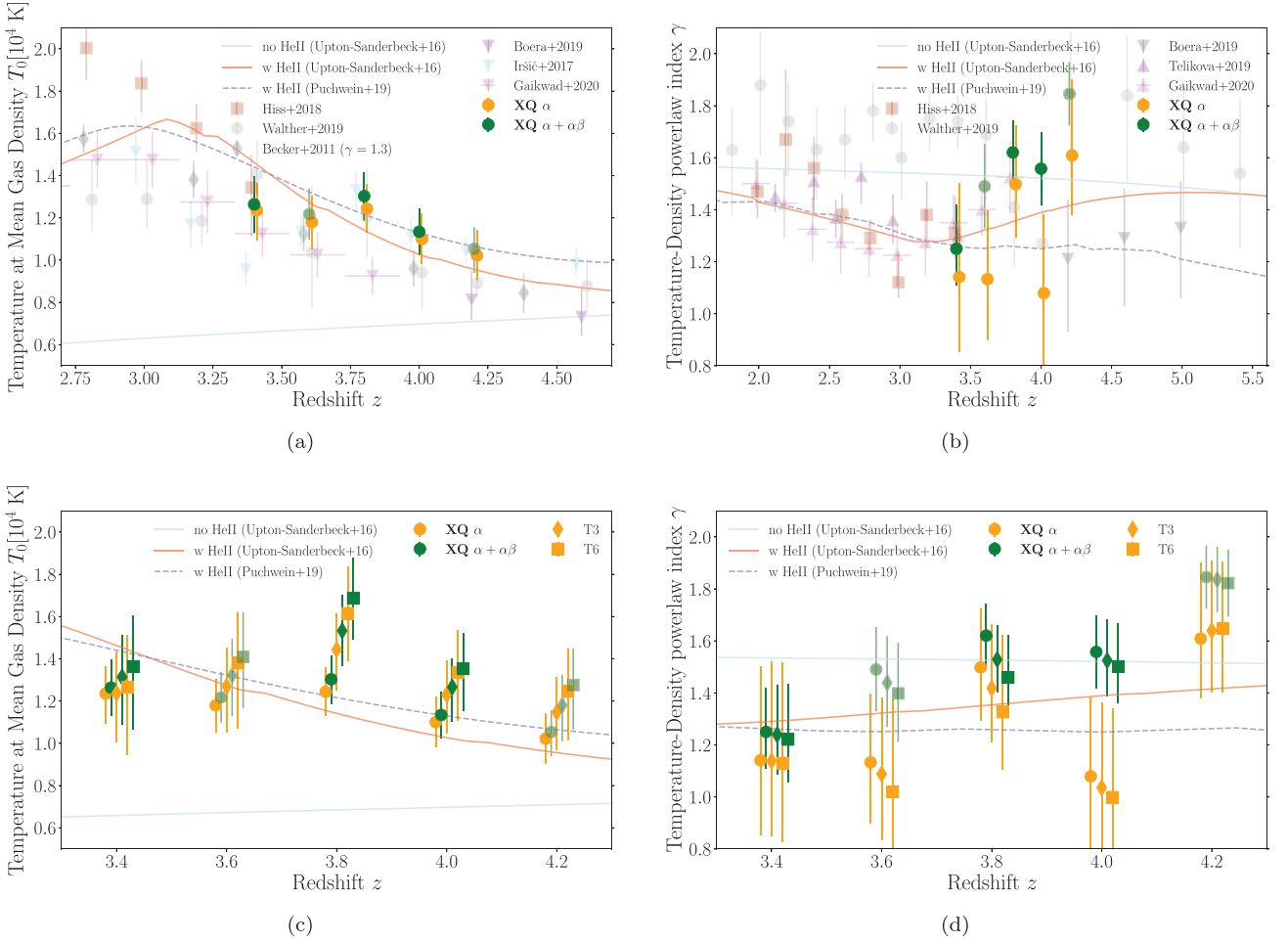
measurement (Section 3.1). We note that the mean transmission in the Ly  $\beta$  forest is then a prediction of the simulations once calibrating to the Ly  $\alpha$  mean flux and so does not need to be modelled. Similarly we have used a Gaussian prior on the IGM temperature around the results of Becker et al. (2011) and taking the reported  $T_0$  values for  $\gamma = 1.3$  as the mean with a standard deviation of 1500 K as our fiducial value. As these are relatively tight priors based on prior measurements of  $T_0$ , we have also explored the effects of weakening the  $T_0$  priors by increasing the standard deviation to 3000 K (T3) and 6000 K (T6), finding that our measurement of  $\gamma$  is robust to these priors.

As described in Section 3.2, we have added a systematic error budget to the covariance matrix. The nominal values that we choose based on our estimated spectral resolution for the median seeing were  $\sigma_R^{\text{UVB}} = 18.3 \text{ km s}^{-1}$  ( $\Delta\sigma_R^{\text{UVB}} = 3.7 \text{ km s}^{-1}$ ) for the UVB arm, and  $\sigma_R^{\text{VIS}} = 12.0 \text{ km s}^{-1}$  ( $\Delta\sigma_R^{\text{VIS}} = 2.40 \text{ km s}^{-1}$ ) for the VIS arm. This corresponds to 20 per cent uncertainty in  $\sigma_R$  in each of the spectral arms. This is perhaps on the conservative side, and we have tested that reducing the resolution error bar by half does not have a large impact on our conclusions.

We now present the results of this analysis. As different redshifts are nearly independent, the degeneracy structure between the IGM parameters is similar in all the redshift bins. To illustrate this structure, we first consider one redshift. The left-hand panel of Fig. 7 shows 2D posterior distributions for the IGM parameter at  $z = 3.4$ , with the orange contours showing our Ly  $\alpha$ -only analysis and with the green contours showing our combined Ly  $\alpha$  and Ly  $\beta$  analysis (i.e. that uses also  $P_{\alpha\beta}$  in addition to  $P_{\alpha\alpha}$ ). As anticipated, adding the Ly  $\beta$  measurement helps principally to constrain  $\gamma$ , the power-law parameter of the temperature–density relation of the IGM. Adding the Ly  $\beta$  forest information does not particularly help in constraints on the IGM temperature nor the mean transmission, rather it appears to multiply our Ly  $\alpha$ -only posterior by a function  $p(\gamma)$ , thus acting to constrain only  $\gamma$ . This behaviour is present in every redshift bin.

This is illustrated in the right-hand panel of Fig. 7 that shows 2D posterior distributions in the plane of  $T_0$ - $\gamma$  for every redshift in our measurement. The solid contours in the panel are for our fiducial prior choice on  $T_0$ , whereas the empty, dashed contours are for a weaker prior choice on  $T_0$ , where the standard deviation on the parameter was 3000 K (T3). The primary effect of the weaker temperature prior is to broaden the posterior distribution in the temperature direction, without significantly affecting the width of the posterior in  $\gamma$  direction.

The recovered redshift evolution of the IGM temperature and the power-law index of the temperature–density relation are shown in Fig. 8, highlighting the difference between the Ly  $\alpha$ -only (orange) and Ly  $\alpha$  + Ly  $\alpha \times$  Ly  $\beta$  analyses (green). The top panels compare our measurements to other measurements in the literature, using the fiducial  $T_0$  prior that is centred around the Becker et al. (2011) measurement. The addition of the cross-power spectrum does not add much information to the measurement of the temperature, with the estimated errors on the best-fitting  $T_0$  improving by less than  $\sim 5$  per cent between the Ly  $\alpha$ -only and joint Ly  $\alpha$  and Ly  $\beta$  analyses. The additional information in the Ly  $\beta$  forest is primarily in the  $\gamma$  evolution. The improvement in the uncertainty on the recovered  $\gamma$  parameters is, on average, a factor of 2 in joint Ly  $\alpha$  and Ly  $\beta$  analysis compared to the Ly  $\alpha$ -only analysis. The best-fitting values for the temperature–density relation in the combined Ly  $\alpha$  analysis for the 1500 K  $T_0$  prior case are  $\gamma = (1.25^{+0.17}_{-0.14}, 1.49^{+0.16}_{-0.16}, 1.62^{+0.12}_{-0.12}, 1.56^{+0.14}_{-0.14}, 1.85^{+0.12}_{-0.12})$ , with the values corresponding to the measured redshift bins (3.4, 3.6, 3.8, 4.0, and 4.2). Note that the central values reduce by 0.1–0.2 at  $z = 3.6$  and 3.8 when we reduce this strong prior on  $T_0$ . The highest redshift bin at  $z = 4.2$  is inconsistent at  $2\sigma$  with the theoretical upper limit of 1.6 for models that assume the only source of heat is photoheating from a uniform background (Hui & Haiman 2003; McQuinn 2016). As this is also one of the two redshift bins where we have a very short path-length (and small number of quasars contributing;



**Figure 8.** Constraints on the temperature at mean density ( $T_0$ ) and the power-law index of the temperature–density relation ( $\gamma - 1$ ) of our work (green and orange) as well as other recent measurements. The solid curves are the models of Upton Sanderbeck, D’Aloisio & McQuinn (2016) with and without He II reionization, and the dashed curves are models of Puchwein et al. (2019) with He II reionization. Top: Our measurements using a fiducial  $T_0$  prior plotted against other measurements in the literature. We have indicated that our  $z = 3.6$  measurement of the cross and both measurements at  $z = 4.2$  used a much smaller path-length than our other measurements by making them transparent. Bottom: Our measurements for different choices of  $T_0$  prior, plotted only within the redshift range of our observations. The  $T_0$  priors differ only in the value of the standard deviation: The fiducial assumes 1500 K (circles); T3 assumes 3000 K (diamonds); and T6 assumes 6000 K (squares). The posterior on  $\gamma(z_i)$  does not change significantly when relaxing the priors on  $T_0$ .

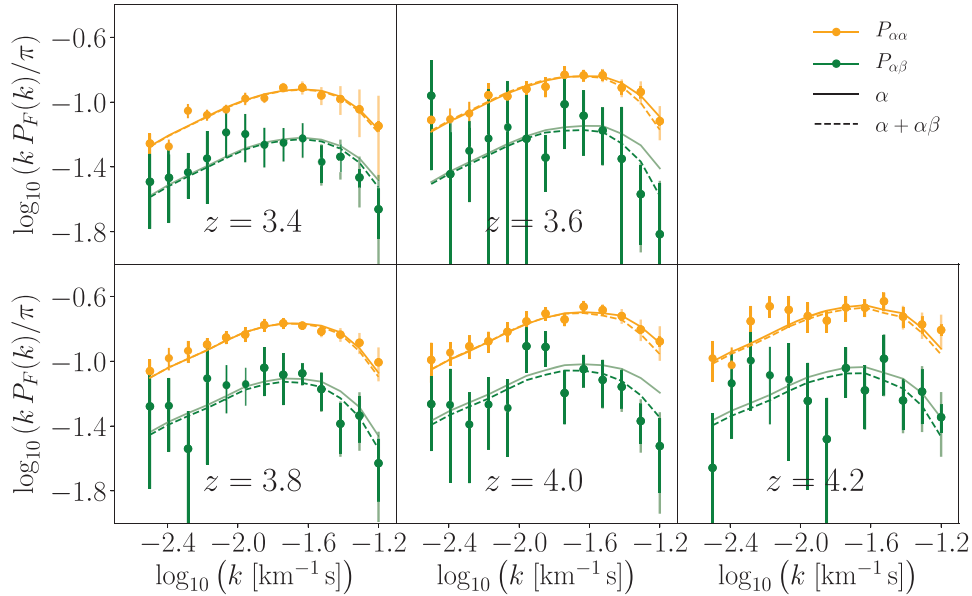
cf. Fig. 1), we are suspicious of our relatively small error bar of  $\pm 0.12$  here.

The bottom panels of Fig. 8 highlight the robustness of the recovered  $\gamma$  posteriors against the imposed priors on  $T_0$ . From the left bottom panel, it is clear that the  $T_0$  measurements from XQ-100 sample are heavily influenced by the choice of the  $T_0$  prior, where our fiducial choice is a very tight prior with 1500 K standard deviation around Becker et al. (2011). Increasing the standard deviation on the Gaussian priors by a factor of  $2 \times (4 \times)$  increases the uncertainty on  $T_0$  parameters by  $1.5 \times (2.2 \times)$ . The benefit of the joint Ly  $\alpha$  and Ly  $\beta$  analysis on the  $T_0$  measurement is slightly more pronounced for weaker priors, with the improvement on the estimated temperature uncertainty being 9 per cent (15 per cent) for the T3 and T6 prior choices, respectively. The XQ-100 power spectrum measurements do not competitively constrain  $T_0$  relative to determinations from higher resolution data, and adding the Ly  $\beta$  forest is only of minor help.

Our measurement of  $\gamma$  is very stable with respect to the choice of  $T_0$  prior and, when including the cross power, is competitive with previous constraints. The bottom right panel of Fig. 8 shows that the recovered  $\gamma$  values change very little between the different

prior choices. This is true for both the Ly  $\alpha$ -only and joint Ly  $\alpha$  and Ly  $\beta$  analyses, with the uncertainty in  $\gamma$  increasing by 4 per cent (18 per cent) and 6 per cent (19 per cent), respectively, when considering the T3 (T6) prior choice. In addition, the errors are significantly reduced when including the cross. The reason why the cross power is able to competitively constrain  $\gamma$  is illustrated in Fig. 6. The effect of different possible  $\gamma$  has more effect at high  $k$  than allowed variations in  $T_0$ . The effect on the power of changing  $\gamma$  by 0.5, a change that is in line with current scatter in  $\gamma$  values, is shown by the dot-dashed and dashed curves in this figure. This effect on the cross power is much larger than that of changing  $T_0$  by 4000 K but fixing  $\gamma$  (compare the solid and dashed curves), a change in  $T_0$  that is significantly larger than the reported errors on this parameter in the literature.

Fig. 9 shows the best-fitting model to only the Ly  $\alpha$  flux power spectrum (solid orange curve) as well as the fit that includes the cross power with Ly  $\beta$  (dashed orange and green curves, respectively). The error bars show the measurements. The best-fitting model yields  $\chi^2 = 48.7$  for 47 degrees of freedom. For the analysis that adds the cross,  $\chi^2 = 115.5$  for 112 degrees of freedom. The final  $\chi^2 = 115.5$  only



**Figure 9.** The best-fitting models of the analysis using  $P_{\alpha\alpha}$  only (solid orange) and of the analysis combining  $P_{\alpha\alpha}$  with the cross  $P_{\alpha\beta}$  (dashed). We also show the predicted  $P_{\alpha\beta}$  using the best-fitting parameters of the Ly  $\alpha$ -only analysis (solid green). The error bars shown are from statistical errors, with added systematic error budget of 20 per cent in the resolution parameter  $\sigma_R$ . The recovered goodness of fit as indicated by  $\chi^2$  (including systematic error budget) is 48.7 for 47 degrees of freedom and 115.5 for 112 degrees of freedom for Ly  $\alpha$ -only and joint  $P_{\alpha\alpha}$  and  $P_{\alpha\beta}$  analyses, respectively. The  $T_0$  prior choice was the fiducial prior with 1500 K for the value of standard deviation.

reduces by 0.5 (2.4) for the weaker T3 (T6) priors on  $T_0$ , again indicating that our measurement is not very sensitive to  $T_0$ .

While the  $\chi^2$  values indicate a good fit, one odd aspect of our result is that the combined analysis shifts  $\gamma$  in *all* of our redshift bins by  $0.5\sigma$ – $1.5\sigma$  relative to the Ly  $\alpha$  analysis. While not of high statistical significance, this trend may suggest a systematic effect. We have performed several tests to assess the robustness of our competitive constraints on  $\gamma(z)$ . As discussed in more detail in Appendixes A and C, resolution and wavelength calibration are possible sources of systematic budget at smaller scales, considerations that led to our error budget of 20 per cent in  $\sigma_R$ . Comparison with high-resolution HIRES/UVES data sets, the resolution differences are largest at  $z = 3.4$  at the level of 15 per cent in  $\sigma_R$ , which results in a factor of 2 smaller errors at our maximum wavenumber. If this lower resolution uncertainty is assumed in the analysis, the results change on average by less than  $\sim 0.2\sigma$  with the main effect in  $T_0$  values, and there are only two redshift bins where the shift is  $>0.2\sigma$  in both the combined and Ly  $\alpha$ -only analyses (and the maximum change in the estimated parameter value is always  $<0.7\sigma$ ). On average, the changes in the parameters are smaller for the combined analysis. Similarly, when including a systematic shift in the wavelength calibration between the UVB and VIS arms of the order of  $5 \text{ km s}^{-1}$  to minimize an imaginary component of the cross power that may indicate a systematic wavelength miscalibration (see Appendix A), the main change is for  $\gamma(z = 3.4)$ , where the value changes by  $1\sigma$ , a redshift where we do not expect any significant miscalibration since the cross is done using one arm of the spectrograph. We found that other redshifts have much smaller shifts of  $\sim 0.1\sigma$ .

The last test we report concerns measuring the covariance matrix on a small statistical sample. We have rerun our analysis but *only* using the mocks to estimate the covariance matrix rather than rescaling the diagonal by bootstrapping the data. To do this, we rescaled the covariance matrix to match the observed path-lengths in both the Lyman-series forests. This also yields similar results, with

the largest changes observed in  $\gamma(z = 3.8)$  and  $\gamma(z = 4.0)$  at below a  $1\sigma$  shift.

Returning to Fig. 9, the solid curves show the Ly  $\alpha$  auto and cross power from the best-fitting model that only uses the  $P_{\alpha\alpha}$  measurement, whereas the dashed curves show the best-fitting model that also includes our  $P_{\alpha\beta}$  measurement. Even though it was not used in the fit, the green solid curve shows the cross power in this model. For the redshifts where the cross has the largest effect on shifting the best-fitting model posterior, namely  $z = 3.6$  and  $4.0$ , this figure illustrates that the added constraining power from using the cross is coming from high  $k$ . (The improvement in error relative to Ly  $\alpha$  analysis alone in the other redshift owes to same cross-power sensitivity.) It is the lowness of the power at high  $k$  that is driving our high  $\gamma$  values. Thus, we should be concerned with systematics whose effect is at the higher wavenumbers: To bias towards higher  $\gamma$ , one needs to find systematics that result in lower cross power particular relative to Ly  $\alpha$ .<sup>10</sup> Since the cross power is immune to systematics that do not appear in both forests, it may be easier to put such a systematic in the Ly  $\alpha$  autopower. More autopower could be due to residual metal contamination, structure in the spectrograph’s noise, or residual continuum fitting. However, our XQ-100 Ly  $\alpha$  power normalization agrees with others’ analysis of the Keck data and our analysis of VLT/UVES (Fig. 5), and we note that all of these systematic effects would be outside the mainstream understanding of how

<sup>10</sup> Another possibility is that it adds additional variance in the cross that affects our results and that by chance this variance affects our measurements in the several redshift bins in the same direction. This variance should be picked up in the bootstrap error estimates and should be reflected in the errors. Let us take the case of variance from continuum errors since this might be singled out owing to the perception (originating we think from higher redshift studies) that the continuum may be harder to fit there. At our redshifts, it is not obviously harder to fit the continuum in Ly  $\beta$  region than Ly  $\alpha$  since the Ly  $\beta$  region is actually less absorbed at a fixed redshift (Fig. 2).



these systematics affect the Ly  $\alpha$  forest power spectrum. The cross power could be affected by resolution uncertainty and wavelength calibration. While these calibrations are discussed and tested in the appendix, we note that *if anything* our resolution correction appears to be too large when comparing with the Keck/HIRES and our VLT/UVES high-resolution measurements of the Ly  $\alpha$  autopower (see  $z = 3.2$  and  $3.4$  panels in Fig. 5 where our measurement falls slightly above these at high  $k$ , although at  $z = 3.0$  the agreement appears greater. The  $z = 3\text{--}3.4$  Ly  $\alpha$  measurements are relevant since they use the lower resolution UV arm that is used for all of our Ly  $\beta$  measurements). Also, our preliminary VLT/UVES Ly  $\beta$ –Ly  $\alpha$  cross-power spectrum appears to show similar high wavenumber behaviour: There is no evidence for a systematic underestimate at high wavenumbers.

The shift in  $\gamma$  to higher values that owes to the cross could also indicate some missing ingredient in our IGM models. For example, the standard cosmological hydrodynamical simulations, like those employed here, do not model the complexities of how reionization processes heat the gas. During He II reionization – which is the important process happening over the redshift range of  $z = 3\text{--}4$  – segments of sightlines should pierce colder IGM while others should pierce recently reionized He II regions that are hotter. We suspect that adding this inhomogeneity will not lead to large improvements in the quality of the best fit as McQuinn et al. (2011) found that the Ly  $\alpha$  power spectrum in realistic models of He II reionization could be well described by a single power-law temperature–density model (despite their simulations of He II reionization showing significant temperature dispersion). However, they did not investigate the impact on a joint measurement like ours, where maybe the same model cannot fit both. We note that our simple *mixed model* for He II reionization in Fig. 6 does show a rather large effect, although likely in the direction that would result in Ly  $\beta$  appearing colder. Further investigation is merited to see whether such missing physics could result in the systematic shift between the two analyses.

Some other source of heating/disruption that is more important in the somewhat higher density regions that are probed by the Ly  $\beta$  forest could also drive the posterior shift between our Ly  $\alpha$  analysis and the combined analysis with the cross. Perhaps the most obvious source there would be cosmic rays from galaxies as these can easily reach the low-density gas probed by our forests. However, as cosmic rays are more likely to be a source of pressure than heating, pressure alone may not be enough since much of the power spectrum cut-off at high  $k$  owes more to thermal broadening than pressure smoothing. To throw a final effect out there, feedback from galactic and quasar winds could be more disruptive to the Ly  $\beta$  forest than Ly  $\alpha$  since it probes gas somewhat closer to galaxies. A potential difficulty with all these possibilities is that the densities that Ly  $\beta$  is sensitive to are likely less than a factor of 2 larger than those for Ly  $\alpha$ .

### 4.3 Interpretation in context of standard model for IGM heating

Fig. 8 also compares our measurement to other measurements (Lidz et al. 2010; Becker et al. 2011; Rorai et al. 2017; Iršič et al. 2017b; Hiss et al. 2018; Telikova et al. 2018; Boera et al. 2019; Walther et al. 2019; Gaikwad et al. 2020). As just mentioned, when we relax the standard deviation of our prior on  $T_0$  that is centred on previous measurements, our data mildly prefer larger temperatures than the most constraining of these measurements. Also shown are two models. The solid red curve is the fiducial He II reionization model in Upton Sanderbeck et al. (2016) assuming parameters to achieve a minimum temperature at  $z \sim 5$  to match the low temperature values

at high redshifts. Thus, it is the case that the low values of many of the previous measurements are difficult to explain with thermal models, and our mildly higher temperatures appear more consistent with thermal history models. The model of Puchwein et al. (2019) favours slightly higher temperatures (dashed blue curve). Our and other temperature measurements are more consistent with the models that include He II reionization than the one that does not (solid blue curve).

More interesting is our constraints on  $\gamma$ . The right-hand panels in Fig. 8 show select prior measurements of  $\gamma$ . The error bars are generally large and there is not clear agreement among these measurements. The reason for large error bars is that Ly  $\alpha$  alone is sensitive to a narrow range of densities and so does not provide much of the lever arm needed to constrain this parameter (Lidz et al. 2010; Becker et al. 2011). Gaikwad et al. (2020) are the first to precisely measure  $\gamma$  at multiple redshifts, aided by using several Ly  $\alpha$  forest statistics rather than just the power spectrum alone.<sup>11</sup> Our measurement suggests that  $\gamma$  flattens with decreasing redshift from  $\gamma \sim 1.6$  at  $z = 4.0$  to  $\gamma \sim 1.2$  at  $z = 3.4$ , confirming the trend found at lower redshift by Gaikwad et al. (2020). This flattening is consistent with expectations from the heating from He II reionization by quasars found in simulations (McQuinn et al. 2009) and models that attempt to replicate the inhomogeneous heating seen in simulations like Upton Sanderbeck et al. (2016, red solid line).

It is interesting to ask how the  $\gamma$  curve changes for different models. The  $T_0$  measurements suggest a relatively brief He II reionization over  $z \sim 3\text{--}4.5$ , which is followed by the Upton Sanderbeck et al. (2016) model. If He II reionization is more extended,  $\gamma$  would evolve more slowly and be somewhat steeper. The Upton Sanderbeck et al. (2016) model assumes an instantaneous hydrogen reionization at  $z = 8$  that imparts a flat temperature–density relation at this redshift ( $\gamma = 1$ ). If most of the hydrogen is reionized as late as possible, at  $z \approx 6$ , the  $\gamma$  values are pushed down at high redshifts, with the  $z = 6$  reionization model in McQuinn et al. (2009) suggesting that they can have a flat value of  $\gamma \approx 1.3$  over our redshift range (similar, but somewhat higher, than the one zone Puchwein et al. 2019 model shown in Fig. 8). Thus, our measurement may tentatively constrain hydrogen reionization, favouring significant ionization at  $z > 6$  (Raskutti et al. 2012; Kulkarni et al. 2019; Keating et al. 2020). Our high values of  $\gamma$  at  $z \sim 4$  are in mild tension with the measurements of Boera et al. (2019) and more consistent with Walther et al. (2019).

## 5 CONCLUSION

We have presented a measurement of the Ly  $\beta$  forest autopower spectrum as well as the Ly  $\alpha$ –Ly  $\beta$  cross-power spectrum using 100  $\lambda/\Delta\lambda \sim 10^4$ ,  $3.4 < z < 4.2$  quasar spectra from the XQ-100 Legacy Survey. Previously, these statistics have only been investigated using low-S/N, low-resolution SDSS spectra (Iršič et al. 2013), where the low resolution inhibited much of the IGM temperature science that is our primary motivation. A secondary motivation is that the cross-power spectrum is essentially immune to the metal line contamination, which substantially complicates Ly  $\alpha$  forest temperature analyses (Lidz et al. 2010; Day et al. 2019), or any systematic that is uncorrelated between the two spectral regions.

The Ly  $\beta$  cross power helps break a strong degeneracy between the mean flux, the temperature at mean density, and the power-law index of the temperature–density relation ( $\gamma - 1$ ) that is present in Ly  $\alpha$

<sup>11</sup> Also of note is a constraining measurement at  $z = 2.4$  of  $\gamma = 1.54 \pm 0.11$  of Bolton et al. (2014), exploiting that this redshift has the best Ly  $\alpha$  forest data.

forest analyses, allowing potentially a much better constraint on  $\gamma$ . Our analysis demonstrated this advantage, reducing the error bar on  $\gamma$  considerably compared to our analysis that uses the Ly  $\alpha$  autopower alone. Our measurements suggest that  $\gamma$  flattens with decreasing redshift from  $\gamma \sim 1.6$  at  $z = 4.0$  to  $\gamma \sim 1.2$  at  $z = 3.4$ , confirming the trend found at lower redshift by Gaikwad et al. (2020). (A steep value at  $z \sim 4$  also seems consistent with our mean flux analysis in Section 3.1.) This flattening in the temperature–density relation is the expected trend from He II reionization, and relatively steep values with  $\gamma \sim 1.5$  are anticipated before He II reionization and well after hydrogen reionization. The steep values may be inconsistent with very late reionization in which the bulk of the IGM is reionized at  $z \approx 6$ .

The value of  $\gamma$  increased across *all* redshift bins at the  $0.5\sigma$ – $1.5\sigma$  when using the cross power compared to the analysis that uses the Ly  $\alpha$  autopower alone. Although this trend is not of high statistical significance, it could indicate some additional contribution to the power or some effect that makes Ly  $\beta$  appear relatively hotter. Our extensive investigation of the XQ-100 resolution and wavelength calibration (and comparison to Keck/HIRES and VLT/UVES high-resolution measurements) suggests that these are not at play. Any systematic error seems more likely to affect the Ly  $\alpha$  autopower since the cross is relatively immune to systematics. However, this would indicate contamination from e.g. metal lines that is well beyond what most Ly  $\alpha$  power spectrum studies have concluded. This trend could also indicate missing IGM physics in our models that makes more overdense regions appear hotter or more pressurized.

For our Ly  $\beta$  measurements, we had to generalize and expand methods adopted for Ly  $\alpha$  to Ly  $\beta$ . In particular, this included understanding metal contamination of our Ly  $\beta$  statistics (including especially the resonant contamination of O VI  $\lambda\lambda 1032, 1038$  Å) and how to remove foreground contamination of the Ly  $\beta$  forest (where we found that removal in the power spectrum is imperfect with just subtraction). We also had to generalize the previous XQ-100 Ly  $\alpha$  analysis to deal with a  $4 \times$  larger covariance matrix and for our MCMC code to search a larger parameter space. In addition, wavelength calibration becomes a more pressing concern when disparate lines are being used to measure correlations on  $10 \text{ km s}^{-1}$  separations. We used the imaginary part of the cross-power spectrum as a diagnostic of such systematic miscalibration.

This work shows that the Ly  $\beta$  forest can be a valuable tool for IGM thermal evolution studies at these redshifts. We put significant work into understanding the resolution of the X-Shooter spectrograph, calibrating a physical seeing-dependent model for the line spread function on arc-lamp lines. Yet, our final thermal parameter constraints were largely limited by resolution uncertainties. Our competitive constraints marginalize over a generous allowance for this uncertainty. Future Ly  $\alpha$ –Ly  $\beta$  cross-power measurements with higher resolution data sets such as SQUAD (Murphy et al. 2019) and KODIAQ (O’Meara et al. 2015) have the potential to set our tightest constraints on  $\gamma$ . We presented a preliminary analysis of the high-resolution VLT/UVES SQUAD DR1 data over the overlapping redshift interval that showed broad agreement. This data set will be even more powerful at lower redshifts than presented here owing to the redshift distribution of the quasars in this sample. We plan to perform a full Ly  $\beta$  analysis on the SQUAD spectra in future work.

## ACKNOWLEDGEMENTS

We would like to especially thank George Becker for discussion of X-Shooter and possible systematics, Guido Cupani for help

with X-Shooter arc-lamp spectra, Bob Carswell for many long discussions on the spectra resolution, Matteo Viel for discussion and help with running the simulations, and John O’Meara for a long discussion on spectral reductions. We would also like to thank Phoebe Upton-Sanderback and Ewald Puchwein for providing the theoretical models to which we compare our results. VI is supported by the Kavli Foundation. This work is supported by NSF grant AST-1514734 and NASA grant NNX17AH68G.

This work is based on observations collected at the European Organisation for Astronomical Research in the Southern hemisphere under ESO programme 189.A-0424. This work made use of the DiRAC High Performance Computing System (HPCS) and the COSMOS shared memory service at the University of Cambridge. These are operated on behalf of the STFC DiRAC HPC facility. This equipment is funded by BIS National E-infrastructure capital grant ST/J005673/1 and STFC grants ST/H008586/1 and ST/K00333X/1.

## 6 DATA AVAILABILITY

The spectroscopic data used in this article were obtained through VLT/XSHOOTER XQ-100 programme and are publicly available in the form of ESO Phase 3 material ([http://archive.eso.org/wdb/wdb/adp/phase3\\_main/form](http://archive.eso.org/wdb/wdb/adp/phase3_main/form)), as described in López et al. (2016). The mean flux and power spectrum measurements used in this paper may be accessed in the Github repository ([https://github.com/bayu-wils/lyb\\_pk/tree/main/output](https://github.com/bayu-wils/lyb_pk/tree/main/output)).

## REFERENCES

- Aguirre A., Schaye J., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2004, *ApJ*, 602, 38
- Aguirre A., Dow-Hygelund C., Schaye J., Theuns T., 2008, *ApJ*, 689, 851
- Armengaud E., Palanque-Delabrouille N., Yèche C., Marsh D. J. E., Baur J., 2017, *MNRAS*, 471, 4606
- Baur J., Palanque-Delabrouille N., Yèche C., Magneville C., Viel M., 2016, *J. Cosmol. Astropart. Phys.*, 8, 012
- Baur J., Palanque-Delabrouille N., Yèche C., Boyarsky A., Ruchayskiy O., Armengaud E., Lesgourgues J., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 013
- Bautista J. E. et al., 2015, *J. Cosmol. Astropart. Phys.*, 5, 060
- Bautista J. E. et al., 2017, *A&A*, 603, A12
- Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, *MNRAS*, 410, 1096
- Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, *MNRAS*, 430, 2067
- Berg T. A. M. et al., 2016, *MNRAS*, 463, 3021
- Bird S., Peiris H. V., Viel M., Verde L., 2011, *MNRAS*, 413, 1717
- Boera E., Murphy M. T., Becker G. D., Bolton J. S., 2014, *MNRAS*, 441, 1916
- Boera E., Murphy M. T., Becker G. D., Bolton J. S., 2016, *MNRAS*, 456, L79
- Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, *ApJ*, 872, 101
- Bolton J. S., Viel M., Kim T.-S., Haehnelt M. G., Carswell R. F., 2008, *MNRAS*, 386, 1131
- Bolton J. S., Becker G. D., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2010, *MNRAS*, 406, 612
- Bolton J. S., Becker G. D., Haehnelt M. G., Viel M., 2014, *MNRAS*, 438, 2499
- Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, *MNRAS*, 464, 897
- Busca N. G. et al., 2013, *A&A*, 552, A96
- Croft R. A. C., Weinberg D. H., Pettini M., Hernquist L., Katz N., 1999, *ApJ*, 520, 1
- Croft R. A. C., Weinberg D. H., Bolte M., Burles S., Hernquist L., Katz N., Kirkman D., Tytler D., 2002, *ApJ*, 581, 20

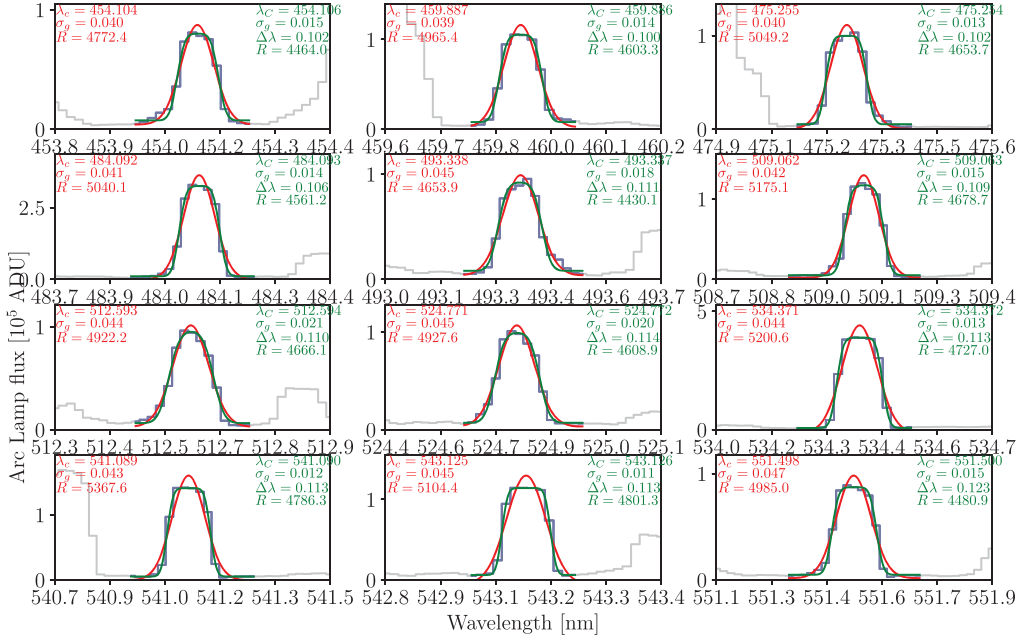
- Day A., Tytler D., Kambalur B., 2019, *MNRAS*, 489, 2536
- Dijkstra M., Lidz A., Hui L., 2004, *ApJ*, 605, 7
- du Mas des Bourboux H. et al., 2017, *A&A*, 608, A130
- Fang T., White M., 2004, *ApJ*, 606, L9
- Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, *ApJ*, 681, 831
- Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2020, *MNRAS*, 506, 4389
- Garzilli A., Bolton J. S., Kim T.-S., Leach S., Viel M., 2012, *MNRAS*, 424, 1723
- Garzilli A., Ruchayskiy O., Magalich A., Boyarsky A., 2019a, preprint ([arXiv:1912.09397](https://arxiv.org/abs/1912.09397))
- Garzilli A., Magalich A., Theuns T., Frenk C. S., Weniger C., Ruchayskiy O., Boyarsky A., 2019b, *MNRAS*, 489, 3456
- Hiss H., Walther M., Hennawi J. F., Oñorbe J., O’Meara J. M., Rorai A., Lukić Z., 2018, *ApJ*, 865, 42
- Hui L., Gnedin N. Y., 1997, *MNRAS*, 292, 27
- Hui L., Haiman Z., 2003, *ApJ*, 596, 9
- Iršič V., McQuinn M., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 026
- Iršič V., Viel M., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 024
- Iršič V. et al., 2013, *J. Cosmol. Astropart. Phys.*, 9, 16
- Iršič V. et al., 2017a, *Phys. Rev. D*, 96, 023522
- Iršič V., Viel M., Haehnelt M. G., Bolton J. S., Becker G. D., 2017b, *Phys. Rev. Lett.*, 119, 031302
- Iršič V. et al., 2017c, *MNRAS*, 466, 4332
- Iršič V., Xiao H., McQuinn M., 2020, *Phys. Rev. D*, 101, 123518
- Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, *MNRAS*, 491, 1736
- Kim T.-S., Viel M., Haehnelt M. G., Carswell R. F., Cristiani S., 2004, *MNRAS*, 347, 355
- Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, *MNRAS*, 485, L24
- Lee K.-G. et al., 2015, *ApJ*, 799, 196
- Lidz A., Faucher-Giguère C.-A., Dall’Aglio A., McQuinn M., Fechner C., Zaldarriaga M., Hernquist L., Dutta S., 2010, *ApJ*, 718, 199
- López S. et al., 2016, *A&A*, 594, A91
- McDonald P., 2003, *ApJ*, 585, 34
- McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, *ApJ*, 543, 1
- McDonald P. et al., 2005, *ApJ*, 635, 761
- McDonald P. et al., 2006, *ApJS*, 163, 80
- McQuinn M., 2016, *ARA&A*, 54, 313
- McQuinn M., Upton Sanderbeck P. R., 2016, *MNRAS*, 456, 47
- McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Hopkins P. F., Dutta S., Faucher-Giguère C.-A., 2009, *ApJ*, 694, 842
- McQuinn M., Hernquist L., Lidz A., Zaldarriaga M., 2011, *MNRAS*, 415, 977
- Murphy M. T., Kacprzak G. G., Savorgnan G. A. D., Carswell R. F., 2019, *MNRAS*, 482, 3458
- Narayanan V. K., Spergel D. N., Davé R., Ma C.-P., 2000, *ApJ*, 543, L103
- O’Meara J. M. et al., 2015, *AJ*, 150, 111
- Palanque-Delabrouille N. et al., 2013, *A&A*, 559, A85
- Palanque-Delabrouille N. et al., 2015, *J. Cosmol. Astropart. Phys.*, 11, 011
- Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2019, *MNRAS*, 485, 47
- Raskutti S., Bolton J. S., Wyithe J. S. B., Becker G. D., 2012, *MNRAS*, 421, 1969
- Ricotti M., Gnedin N. Y., Shull J. M., 2000, *ApJ*, 534, 41
- Rogers K. K., Peiris H. V., 2021, *Phys. Rev. Lett.*, 126, 071302
- Rollinde E., Theuns T., Schaye J., Pâris I., Petitjean P., 2013, *MNRAS*, 428, 540
- Rorai A. et al., 2017, *Science*, 356, 418
- Rudie G. C., Steidel C. C., Pettini M., 2012, *ApJ*, 757, L30
- Sánchez-Ramírez R. et al., 2016, *MNRAS*, 456, 4488
- Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, *MNRAS*, 318, 817
- Schaye J., Aguirre A., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2003, *ApJ*, 596, 768
- Seljak U., McDonald P., Makarov A., 2003, *MNRAS*, 342, L79
- Seljak U., Slosar A., McDonald P., 2006a, *J. Cosmol. Astropart. Phys.*, 10, 14
- Seljak U., Makarov A., McDonald P., Trac H., 2006b, *Phys. Rev. Lett.*, 97, 191303
- Slosar A. et al., 2011, *J. Cosmol. Astropart. Phys.*, 9, 1
- Slosar A. et al., 2013, *J. Cosmol. Astropart. Phys.*, 4, 26
- Telikova K. N., Balashev S. A., Shternin P. S., 2019, *ApJ*, 887, 205
- Theuns T., Zaroubi S., 2000, *MNRAS*, 317, 989
- Theuns T., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell R. F., 2002, *MNRAS*, 332, 367
- Tytler D., O’Meara J. M., Suzuki N., Kirkman D., Lubin D., Orin A., 2004, *AJ*, 128, 1058
- Upton Sanderbeck P. R., D’Aloisio A., McQuinn M. J., 2016, *MNRAS*, 460, 1885
- Vernet J. et al., 2011, *A&A*, 536, A105
- Viel M., Haehnelt M. G., 2006, *MNRAS*, 365, 231
- Viel M., Matarrese S., Heavens A., Haehnelt M. G., Kim T.-S., Springel V., Hernquist L., 2004a, *MNRAS*, 347, L26
- Viel M., Haehnelt M. G., Springel V., 2004b, *MNRAS*, 354, 684
- Viel M., Weller J., Haehnelt M. G., 2004c, *MNRAS*, 355, L23
- Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, *Phys. Rev. D*, 71, 063534
- Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., Rauch M., Sargent W. L. W., 2008, *Phys. Rev. Lett.*, 100, 041304
- Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013a, *Phys. Rev. D*, 88, 043502
- Viel M., Schaye J., Booth C. M., 2013b, *MNRAS*, 429, 1734
- Walther M., Hennawi J. F., Hiss H., Oñorbe J., Lee K.-G., Rorai A., O’Meara J., 2018, *ApJ*, 852, 22
- Walther M., Oñorbe J., Hennawi J. F., Lukić Z., 2019, *ApJ*, 872, 13
- Wu X., McQuinn M., Kannan R., D’Aloisio A., Bird S., Marinacci F., Davé R., Hernquist L., 2019, *MNRAS*, 490, 3177
- Yèche C., Palanque-Delabrouille N., Baur J., du Mas des Bourboux H., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 047
- Zaldarriaga M., Seljak U., Hui L., 2001, *ApJ*, 551, 48
- Zaldarriaga M., Scoccimarro R., Hui L., 2003, *ApJ*, 590, 1

## APPENDIX A: SPECTROGRAPH RESOLUTION CORRECTION

Studies of the small-scale Ly $\alpha$  flux power spectrum rely on correcting for spectral smoothing that owes to the spectrograph’s resolution. While this is a small correction for studies using high-resolution Keck/HIRES or VLT/UVES spectra even at our maximum wavenumber of  $k_{\max} = 10^{-1.2} \text{ s km}^{-1}$ , the correction amounts to a factor of as much several for the X-Shooter spectrograph. Previous studies have disagreed on the effective resolution of X-Shooter at the 20 per cent level, which can lead to a large factor of 2 differences in the reported power at  $k_{\max}$  (Iršič et al. 2017c; Yèche et al. 2017; Walther et al. 2018). In this section, we describe the implementation of the resolution correction that is used in this paper.

First, in order to determine the resolution when the spectrograph’s slit is fully illuminated, we use spectra taken of an arc lamp using  $1 \times 2$  binning of spectral pixels. We use the same slit widths as our data set: 1.0 arcsec for the UVB arm and 0.9 arcsec for the VIS arm. The arc-lamp spectra were reduced using the ESO pipeline. The most isolated lines as determined by visual inspection were fitted with a Gaussian. These fits roughly reproduce the results of a similar analysis in Walther et al. (2018) that found  $R \equiv \lambda/\Delta\lambda \approx 5000$ , where  $\Delta\lambda$  is the FWHM. See Fig. A1, which shows this exercise for the lines analysed in the UVB arm. The blue histograms are the observed line profiles, and the red coloured lines are the best-fitting Gaussian models. The best-fitting parameters in each of the panels are also shown in red text.



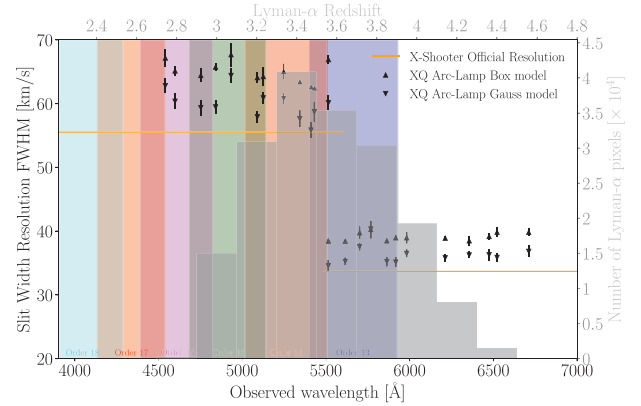


**Figure A1.** The panels show the X-Shooter UV arm ThAr calibration arc-lamp spectra of select clean lines that span the wavelength range used here. The arc-lamp spectrum in each panel is shown in grey, and highlighted in blue is the region around the line where the fits were performed. In each panel, two fitting models – Gaussian model (in red) and Box model (in green) – are overlotted for the best-fitting parameters, which are quoted in the top left (red) and right (green).

However, a Gaussian is not expected to be a good model for the line spread function of a fully illuminated slit, and several of the arc-lamp lines clearly show a more box-like profile. Thus, we explore a more physically motivated model for the line profile, where the line profile is modelled as a boxcar filter convolved with a Gaussian. This model captures that the slit is a boxcar filter, allowing light to go through spectrograph at different angles. This angular dispersion is larger than the intrinsic dispersion of the spectrograph that owes to diffraction. While this diffractive broadening depends on the details of the spectrograph, such as the number of grooves illuminated, we approximate this lesser contribution with a Gaussian kernel. This model we refer to as the ‘Box model’. Unlike our previous one-parameter Gaussian model, the Box model is described by two parameters: the full width of the boxcar  $R \equiv \lambda/\Delta\lambda$  and the standard deviation of the Gaussian  $\sigma_g$ . The broader boxcar width  $R$  essentially is the FWHM resolution in this model. The results of the Box model fits are shown in green in Fig. A1, with green text giving the best-fitting parameters.

The summary of the fits to the slit-arc spectra is presented in Fig. A2, comparing both models (Gaussian and Box), and comparing the results to the official X-Shooter resolution. The Box model recovers an  $\sim 10$  per cent larger FWHM resolution compared to the Gaussian model. (This difference should be thought of as purely parametric. These are just different functions!) The VIS arm shows very little wavelength-dependent scatter, while the UVB arm shows more structure. This could be partially due to two effects: First, the lines in the UVB are more likely to be blends as the lower resolution in this arm makes it harder to find all blends; secondly, the Echelle orders of the spectrograph (visualized by the background colours in Fig. A2) have more overlap in the UVB arm, which might cause resolution variation that owes to the variations in Echelle orders used.

We compress the results of our arc fits to a single resolution number for each of the arms. Using medians to reduce biases from blended lines, we find an FWHM of  $\Delta v = 64.9 \text{ km s}^{-1}$  for the UVB arm and



**Figure A2.** Shows the slit width resolution (FWHM in  $\text{km s}^{-1}$ ) as estimated from fitting arc-lamp lines for the UVB and VIS arms of the XQ-Shooter spectrograph. The two orange lines correspond to the official published X-Shooter resolution for the nominal slit width of the XQ-100 observations, which is 1.0 arcsec for the UVB arm ( $R = 5400$ , shown at shorter wavelengths) and 0.9 arcsec for the VIS arm ( $R = 8900$ , shown at longer wavelengths). The black data points are the measurements of the arc-lamp emission lines in both spectral arms for the nominal slit width, using two different methods: the Gaussian model (inverted triangles) and the Box model (triangles). The Box model is more physically motivated and yields larger FWHMs, possibly explaining some of the discrepant resolutions quoted in previous literature. The resolution correction applied to the measurements in this paper Fourier transforms the full profile of the Box model.

$\Delta v = 39.1 \text{ km s}^{-1}$  for the VIS arm. (These numbers set the width of the boxcar filter. For the Gaussian that is then convolved with this boxcar, we find a width of 8.76 and 6.64  $\text{km s}^{-1}$ .)

The second characteristic of the XQ-100 observations that needs to be addressed to properly account for the resolution correction is



the atmospheric seeing. The seeing of our XQ-100 observations is generally better than the slit width of the observations (Yèche et al. 2017), which would result in a lower effective FWHM resolution values when correcting the flux power spectrum measurements of the XQ-100 data (because the angular dispersion through the spectrograph is reduced). In our favoured Box car model, there is a simple way to account for this effect. In a geometric optics picture, each point illuminated the slit contributes to one point in the boxcar profile, with one side of slit contributing to one edge and the other side contributing to the other edge. Since our model is a Box car convolved with a Gaussian, in this picture the Gaussian owes to diffraction. This picture allows us to then weight each part of the slit depending on the seeing, which we model as having a Gaussian point spread function for which the width has been fitted observationally during each exposure. Thus, in contrast to the arc-lamp spectra in which the slit is uniformly illuminated and the full boxcar populated, each point in the boxcar is weighted by the Gaussian point spread function. We do this in a manner that weights each point in the slit by the Gaussian seeing profile.

Thus, this Boxcar model allows us to model the line spread function of the spectrograph for a given seeing. We are interested in the Fourier transform of the line spread function [ $W_{s,X}(k, z)$  that appears in equation 2]. We find that once we Fourier transform the Boxcar model, the resolution window function can be very well approximated by a Gaussian over the wavenumber range of interest. We then fit a Gaussian model with standard deviation  $\sigma_R$  to our calculations of the seeing-dependent window function and use the form

$$\sigma_R = a + b(x - 0.65) + c(x - 0.65)^2 + d(x - 0.65)^3$$

$$\text{if } x < 0.8;$$

$$\sigma_R = e + f(x - 1) + g(x - 1)^2 + h(x - 1)^3$$

$$\text{otherwise,}$$

where  $x \equiv \theta_{\text{atm}}(s)/\theta_{\text{slit}}$ ,  $\theta_{\text{atm}}(s)$  is the reported seeing for quasar  $s$ , and  $\theta_{\text{slit}}$  is the size of the slit. In units of  $\text{km s}^{-1}$ , we find  $(a, b, c, d, e, f, g, h) = (17.7982, -9.8992, -14.7344, -15.366, 19.9168, -3.4890, -4.6915, -4.6875)$  for the UV arm and  $(a, b, c, d, e, f, g, h) = (11.4691, -5.2146, -8.0595, -8.7484, 12.5689, -1.7810, -2.4529, -2.4894)$  for the VIS. Mathematica notebooks with the full calculation pertaining to our resolution correction can be provided upon request. These fits are then used when computing  $W_{s,X}(k, z)$  per equation (3). Our parametrization of  $\sigma_R$  is a function of the seeing, which can vary over an observation. We use the minimum and maximum reported seeing to calculate  $W_{s,X}$ , and toss out modes where this causes differences beyond our error tolerance (cf. equation 4). The average of these two  $\sigma_R$  is used as our resolution correction. Finally, we allow for 20 per cent uncertainty in the resolution parameter,  $\sigma_R$ , in our main analysis. Our seeing dependence of the resolution is considerably weaker than the model of Yèche et al. (2017) where  $\sigma_R \propto x$ . This  $\sigma_R \propto x$  scaling should only hold when  $x$  is considerably smaller than 1 and is not a good approximation over the seeings present in XQ-100.

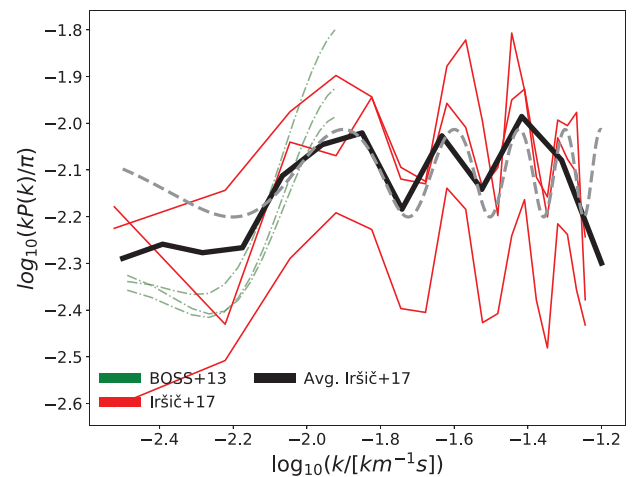
## APPENDIX B: METAL CONTAMINATION AND DLA ABSORBERS

We use estimates for the contaminating metal power spectrum,  $\hat{P}_M(k_i, z_i)$ , given in Iršič et al. (2017c) for the same XQ-100 data set. These estimates use the QSO-frame wavelength range just redward of the quasar Ly  $\alpha$  line,  $1268 \text{ \AA} < \lambda_{\text{metal}} < 1380 \text{ \AA}$ , to estimate the forest metal contamination. Isolating to these wavelengths captures

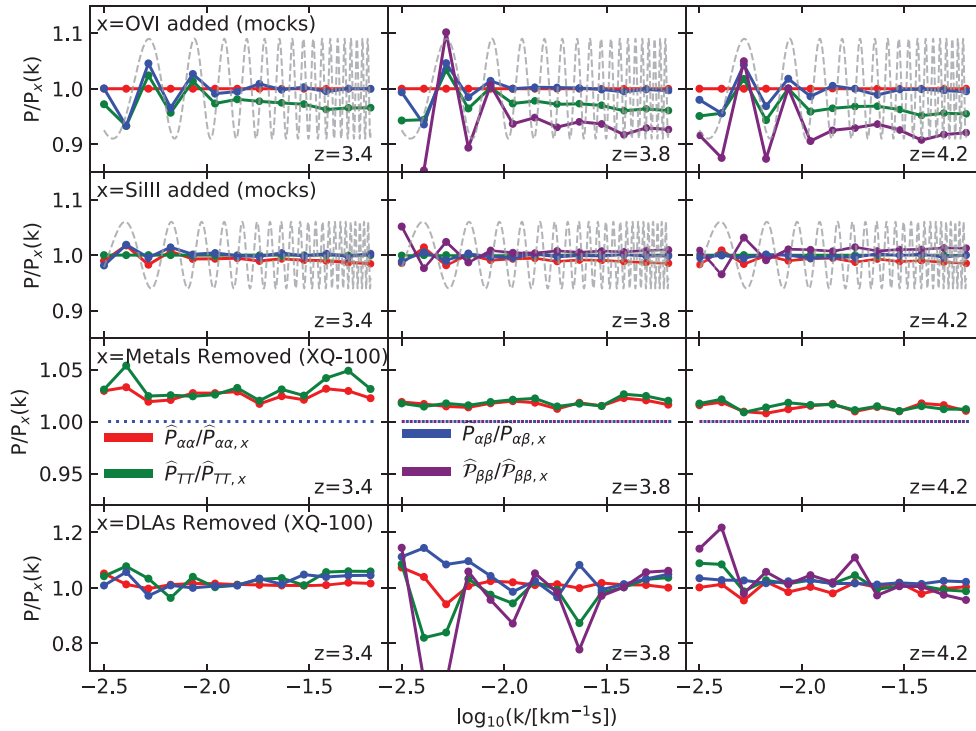
the bulk of metal absorption in the Ly  $\alpha$  forest at  $z = \lambda_{\text{metal}}/\lambda_{\alpha}(z_{\text{qso}} + 1) - 1$  from lines redward of Ly  $\alpha$ , and similarly for the Ly  $\beta$  forest. Because the redside contains the strongest metal absorption line, C IV  $\lambda\lambda 1548, 1550 \text{ \AA}$ , the redside contribution to the power should be larger than the blueside. The three solid red curves in Fig. B1 show the XQ-100 estimates of Iršič et al. (2017c) for  $z = 3.8, 4.0$ , and  $4.2$ . The oscillatory behaviour owes to the doublet structure of the C IV line. The figure also shows the measurement of BOSS in Palanque-Delabrouille et al. (2013), which used many more spectra (green curves) for the same three redshifts. The BOSS metal power estimates show less variation in redshift than those of Iršič et al. (2017c), suggesting that the variation between the three redshift bins in Iršič et al. (2017c) owes largely to cosmic variance and noise. Thus, to subtract metals we average these three Iršič et al. (2017c) measurements (which yields the solid black curve) and, then, subtract this redshift-average power off of  $P_{\alpha\alpha}$  and  $P_{TT}$  to obtain our power spectra as given in equation (2). We note that the relative size of the metal correction is less than 3 per cent, as shown in the third row in Fig. B2. The cross power,  $P_{\alpha\beta}$ , is unaffected by non-resonant metals – metals that fall far enough from our H I lines that their effect is not resonantly enhanced at the beat wavenumber. We consider the resonant effect of metals below. Furthermore, to the extent the metal power does not evolve with redshift as we have assumed, our estimates for  $\mathcal{P}_{\beta\beta}$  are unaffected by metal contamination.

This metal subtraction scheme does not capture the beating effect of metals that have wavelengths close to our transitions (McDonald et al. 2006). The metal lines that are the most problematic for our analysis are Si III  $\lambda 1207 \text{ \AA}$  ( $2000 \text{ km s}^{-1}$  from the Ly  $\alpha$  line), and O VI  $\lambda\lambda 1032, 1038 \text{ \AA}$  ( $1800, 3500 \text{ km s}^{-1}$  from Ly  $\beta$ ). The beating from Si III is famously apparent in SDSS analyses of the Ly  $\alpha$  forest power spectra at low wavenumbers (McDonald et al. 2006). O VI is by far the strongest absorption line at these redshifts for moderately overdense systems (Aguirre et al. 2008) and so a similar beating effect may be important between Ly  $\beta$  and O VI.

To estimate the potential contribution of resonant metals, we have created mock spectra with Si III and O VI absorption. To do this, we used the mean relationship between  $\tau_{\text{Si III}}(\lambda = 1207 \text{ \AA})$  or  $\tau_{\text{O VI}}(\lambda =$



**Figure B1.** Measurements of the red-side metal power spectra described in Appendix B. The solid red curves show the XQ-100 measurement of Iršič et al. (2017c) at  $z = 3.8, 4.0$ , and  $4.2$ , ordered in decreasing amplitude. The thick solid black curve averages these three estimates. The green curves are the SDSS/BOSS measurement of the red-side metal power from Palanque-Delabrouille et al. (2013) for the same redshifts.



**Figure B2.** The effect of different systematics for three of the redshift bins we consider. The first and second rows show, respectively, predictions for the fractional impact of resonant O VI and Si III absorption on our power estimates, where this metal contamination is modelled in the manner described in Appendix B. The grey dashed lines show the expected phase from resonances with O VI  $\lambda 1032$  Å and Si III  $\lambda 1207$  Å. The third row shows the fractional effect on our estimated power from our non-resonant metal power correction. Non-resonant metals have no effect on  $P_{\alpha\beta}$  and, to the extent they contaminate equally the Ly  $\beta$  and foreground Ly  $\alpha$  forests (as our procedure assumes),  $P_{\beta\beta}$ . The fourth row shows the fractional effect of our DLA correction.

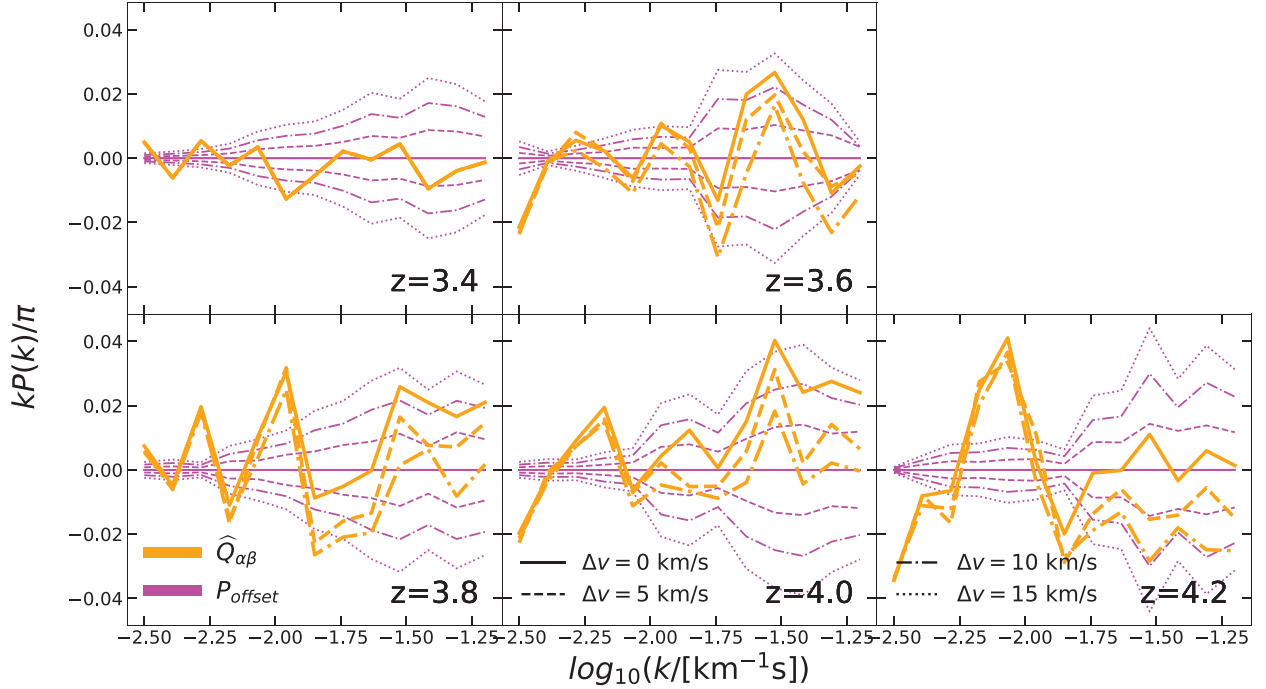
1032 Å) and the H I Ly  $\alpha$  optical depth  $\tau_{\alpha}$  measured in Aguirre et al. (2004, 2008) using the pixel optical depth method. These studies find an average mapping of  $\tau_{\text{Si III}} = [10^{-4} - 10^{-3}] \tau_{\alpha}$  and  $\tau_{\text{O VI}} = [10^{-3} - 10^{-2.5}] \tau_{\alpha}$ , where these ranges have been chosen to generously bracket the findings of these studies. We also put in the other doublet of the O VI line by using the ratio of oscillator strengths. The top two rows in Fig. B2 show the fractional effect of these resonances with O VI and Si III assuming the maximum multiplier in the ranges quoted above of  $\tau_{\text{O VI}} = 10^{-2.5} \tau_{\alpha}$  and  $\tau_{\text{Si III}} = 10^{-3.5} \tau_{\alpha}$ . The dashed curves are to guide the eye, showing the effective beat frequency of the strongest O VI line,  $\cos([1800 \text{ km s}^{-1}]k)$  for the 1800 km s $^{-1}$  shift from O VI  $\lambda 1032$  Å, and the analogous cosine for Si III. The solid lines show these estimates for fractional amplitude of this resonant metal contamination for the different power spectra we measure. Most important for our study are  $P_{\alpha\alpha}$  and  $P_{\alpha\beta}$ , where the effects are small, peaking at  $\sim 5$  per cent at low wavenumbers. However, this model predicts that for  $P_{\beta\beta}$ , resonant O VI can be a 10 per cent effect across the board. Since we have taken the maximum of our range of mapping from H I to O VI, our models may overestimate the effect. However, we are not modelling the stochasticity of the cosmic metal distribution, such that some H I systems will have much stronger lines. This results in a larger stochastic component to the power that could increase the contamination over the level predicted.

We now turn to a second form of contamination, from the damping wings of H I lines from dense hydrogen systems associated with intervening galaxies. These so-called DLA absorbers are a known source of large-scale power that can bias analyses as damping wings are not included in our mock spectra (nor do our simulations likely capture the properties of DLAs accurately). DLAs that contaminate

the XQ-100 quasar spectra are masked using the DLA sample of the XQ-100 survey team (Sánchez-Ramírez et al. 2016). For each DLA, we mask out the spectral region within 1/2 of the equivalent width from the Ly  $\alpha$  and Ly  $\beta$  line centres. In the region outside of the mask, we then divide out DLA absorption profile using the  $N_{\text{H I}}$  measured in Sánchez-Ramírez et al. (2016) to remove any remaining effect from the broad damping wings. In the bottom row in Fig. B2, the ratio between the power spectrum with and without DLAs shows that DLAs are a significant contaminant, with a 10–20 per cent effect at  $z = 3.8$  in  $P_{\text{TT}}$  and  $P_{\beta\beta}$ . The effects are smaller,  $< 5$  per cent ( $\lesssim 10$  per cent), on the  $P_{\alpha\alpha}$  ( $P_{\alpha\beta}$ ) power spectra we use for our thermal analysis. This shows that DLAs are not a negligible contaminant and so our correction is important.

## APPENDIX C: WAVELENGTH CALIBRATION

Another concern is the wavelength calibration of X-Shooter. Wavelength calibration errors become substantial once they are comparable to the thermal scale of  $\sim 10 \text{ km s}^{-1}$ , as quantified below. Within a single arm of the spectrograph, the wavelength calibration is done by centroiding either arc or skylines, and then fitting a polynomial for how the wavelength depends on pixel number. This procedure is first applied to a single well-characterized ‘master’ exposure, which is given extra scrutiny. This master fit is translated on to other exposures using offsets calculated from the positions of the skylines (López et al. 2016). The expectation is that the errors in interpolation are well below our  $10 \text{ km s}^{-1}$  tolerance (G. Becker, private communication). Another systematic shift owes to the relative positions of the objects in the slits for the UV and VIS arms (as the



**Figure C1.** The imaginary component of the Ly  $\alpha$ –Ly  $\beta$  cross-power spectrum, denoted as  $Q_{\alpha\beta}$ , as well as the real part of it multiplied by  $\tan(k\Delta v)$  for  $\Delta v = 0, 5, 10,$  and  $15 \text{ km s}^{-1}$  [pink curves denoted by  $P_{\text{offset}}$ , where  $P_{\text{offset}} \equiv P_{\alpha\beta} \tan(k\Delta v)$ ]. Also shown are the resulting  $Q_{\alpha\beta}$  if we input a shift into the analysis of the specified amounts. These pink curves represent the expected imaginary component in the presence of a systematic wavelength miscalibration of size  $\Delta v$ . The measured  $Q_{\alpha\beta}$  (solid orange) may suggest a slight systematic offset, particularly at  $z = 3.8$  and  $4.0$  ( $z = 3.4$  is only redshift where both measurements use the UV arm and so any offset would be surprising). To see this, first note that at low  $k$  behaviour of the orange curves is just noise, and focus on the offsets at the higher  $k$ .

skylines illuminate the whole slit and so calibrating off of them does not reflect these offsets). For example, observer centroiding differences will result in a correlated shift but that does not cancel as the arms have different resolutions. Additionally, the slits themselves (as spectrograph arms are observed simultaneously) may not be perfectly aligned mechanically, creating a similar effect. As the FWHM for the UVB arm is  $\sim 50 \text{ km s}^{-1}$ , a  $0.2 \text{ arcsec}$  offset in the  $1 \text{ arcsec}$  slit would result in a  $10 \text{ km s}^{-1}$  shift relative to an arm where the source is not offset in its slit. If the centroid is similarly displaced in both slits as per observer miscentring, the effect would be more than halved, reflecting the difference in resolutions between the UV and VIS arms.

For the autopower spectra, wavelength calibration is likely not a significant systematic as this measurement uses pixels that are relatively close to each other and, furthermore, systematic shifts between arms would have no effect. However, our cross-power spectra typically correlate different spectrograph arms for all redshifts besides  $z = 3.4$ : All of our Ly  $\beta$  measurements use the UVB arm, and these are correlated with Ly  $\alpha$ , which below  $z < 3.5$  also uses the UVB arm and  $z > 3.60$  uses the VIS arm.

As a shift in velocity of  $\Delta v$  results in a mode acquiring an additional phase of  $\Delta vk$ , such that the imaginary and real parts of the cross-power spectrum,  $P_{\alpha\beta}$  and  $Q_{\alpha\beta}$ , are changed from the true purely real power spectrum  $P_{\alpha\beta}^T$ , becoming

$$P_{\alpha\beta} = P_{\alpha\beta}^T(k) \cos[\Delta vk]; \quad Q_{\alpha\beta} = P_{\alpha\beta}^T(k) \sin[\Delta vk], \quad (\text{C1})$$

where  $P_{\alpha\beta}^T(k)$  is the true cross-power spectrum. For the wavenumbers we measure and the likely shifts,  $\Delta vk$  is smaller than unity and, hence, the effect is much larger in the imaginary term ( $Q_{\alpha\beta}$ ), which

we use below to diagnose any offsets. A shift of  $\Delta v = 5$  ( $10$ )  $\text{km s}^{-1}$  would bias the cross-power measurement studied in the main body of this paper by a factor of  $\cos[\Delta vk_{\text{max}}] = 0.95$  ( $0.81$ ), where  $k_{\text{max}} = 10^{-1.2} \text{ s}^{-1} \text{ km}$  is the centre of our highest wavenumber bin used in our analysis. Thus, given our errors, the tipping point where the real component (i.e. our power spectrum measurement) is appreciably biased is offsets of at least  $\Delta v \approx 10 \text{ km s}^{-1}$ . We make the previous statement for the case where there is no systematic error budget for resolution as it reflects our statistical errors: Really, the tipping point in our thermal analysis in Section 2 should be at larger  $\Delta v$  than  $10 \text{ km s}^{-1}$  as our 20 per cent resolution error budget really allows for a factor of  $\sim 2$  error at  $k_{\text{max}}$ .

Fig. C1 shows this systematic check. The yellow solid lines are the imaginary component of the Ly  $\alpha$ –Ly  $\beta$  cross-power spectrum,  $\hat{Q}_{\alpha\beta}$ . The pink curves are the real part of it multiplied by  $\tan(k\Delta v)$  for  $\Delta v = 0, 5, 10,$  and  $15 \text{ km s}^{-1}$  (denoted in the figure as  $P_{\text{offset}}$ , where  $P_{\text{offset}} \equiv \hat{P}_{\alpha\beta} \tan(k\Delta v)$ ). These pink curves represent the expected imaginary component in the presence of a systematic wavelength miscalibration of size  $\Delta v$ . Also shown are the resulting  $\hat{Q}_{\alpha\beta}$  if we input a shift into the analysis of the specified amounts. The measured  $Q_{\alpha\beta}$  (solid orange) may suggest a slight systematic offset, particularly at  $z = 3.8$  and  $4.0$  ( $z = 3.4$  is only redshift where both measurements use the UV arm and so an offset would be surprising). To see this, first note that at low  $k$  the structure is set by noise, and focus on the offsets at the higher  $k$ . For example, at  $z = 4.2$  there is no appreciable offset at high  $k$  without a shift, making it so that when we shift by  $5$  and  $10 \text{ km s}^{-1}$  the resulting orange curves follow the pink predictions for the effect of such offsets.  $\hat{Q}_{\alpha\beta}$  in the  $z = 3.8$  and  $4.0$  bins suggest a  $10 \text{ km s}^{-1}$  offset, perhaps (again look at the orange solid curve, and the dashed and dot–dashed orange curves then show how putting in

a possibly corrective offset changes  $\widehat{Q}_{\alpha\beta}$ ). A  $10 \text{ km s}^{-1}$  shift should not affect our thermal constraints in the paper: We have redone this analysis with a  $5 \text{ km s}^{-1}$  shift, splitting the difference with unoffset and possible offset of  $\Delta v = 10 \text{ km s}^{-1}$  motivated by the systematic error scaling quadratically in  $\Delta v$ , and find only shifts at the tenths of standard deviations in our  $T_0$ - $\gamma$  inferences.

Unfortunately, for the case where the calibration errors are uncorrelated between observations (and not systematic), as might occur from random observer miscentring on the slit, the imaginary component no longer constrains the miscalculation. Rather, the cross-power spectrum would be purely real with

$$P_{\alpha\beta} = P_{\alpha\beta}^T(k) \exp \left[ -\sigma_v^2 k^2 / 2 \right], \quad (\text{C2})$$

where we have assumed that  $\Delta v$  is drawn from a Gaussian with standard deviation  $\sigma_v$ . The extra Gaussian damping would bias our measurement. If  $\sigma_v = 5$  (10)  $\text{km s}^{-1}$ , our rough upper bound from observer miscentring above assuming 20 per cent offsets,  $\exp[-\sigma_v^2 k_{\text{max}}^2 / 2] = 0.95$  (0.82). Thus, similarly to with a systematic offset of  $\Delta v$ , the bias would be significant for a random offset that is larger than  $\sigma_v = 10 \text{ km s}^{-1}$ .

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.