Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 2842–2855 (2022) https://doi.org/10.1093/mnras/stab3201
Advance Access publication 2021 November 11

Optimal 1D Ly α forest power spectrum estimation – II. KODIAQ, SQUAD, and XQ-100

Naim Göksel Karaçaylı [1]★ Nikhil Padmanabhan [1,2] Andreu Font-Ribera [3,4] Vid Iršič,[5] Michael Walther [6] David Brooks,[4] Enrique Gaztañaga,[7,8] Robert Kehoe,[9] Michael Levi,[10] Pierros Ntelis [11] Nathalie Palanque-Delabrouille[6] and Gregory Tarlé[12]

[1]*Department of Physics, Yale University, New Haven, CT 06511, USA*
[2]*Department of Astronomy, Yale University, New Haven, CT 06511, USA*
[3]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, E-08193 Bellaterra (Barcelona), Spain*
[4]*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[5]*Kavli Institute for Cosmology, Department of Physics, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[6]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[7]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[8]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[9]*Department of Physics, Southern Methodist University, 3215 Daniel Ave., Dallas, TX 75205, USA*
[10]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[11]*Aix-Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France*
[12]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*

Accepted 2021 October 28. Received 2021 October 18; in original form 2021 August 25

ABSTRACT

We measure the 1D Ly α power spectrum P_{1D} from Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ), The Spectral Quasar Absorption Database (SQUAD), and XQ-100 quasars using the optimal quadratic estimator. We combine KODIAQ and SQUAD at the spectrum level, but perform a separate XQ-100 estimation to control its large resolution corrections in check. Our final analysis measures P_{1D} at scales $k < 0.1\,\mathrm{s\,km^{-1}}$ between redshifts $z = 2.0$–4.6 using 538 quasars. This sample provides the largest number of high-resolution, high-S/N observations; and combined with the power of optimal estimator it provides exceptional precision at small scales. These small-scale modes ($k \gtrsim 0.02\,\mathrm{s\,km^{-1}}$), unavailable in Sloan Digital Sky Survey and Dark Energy Spectroscopic Instrument analyses, are sensitive to the thermal state and reionization history of the intergalactic medium, as well as the nature of dark matter. As an example, a simple Fisher forecast analysis estimates that our results can improve small-scale cut-off sensitivity by more than a factor of 2.

Key words: methods: data analysis – intergalactic medium – quasars: absorption lines.

1 INTRODUCTION

The Ly α forest technique can map out the matter distribution in vast volumes and at small scales ($r \lesssim 1$ Mpc) between $2 \lesssim z \lesssim 5$. At these redshifts, the structure formation is mildly non-linear, but the physics of the Ly α forest is further enriched by the thermal state and reionization history of the intergalactic medium (IGM; Hui & Gnedin 1997; Gnedin & Hui 1998). The line-of-sight flux power spectrum P_{1D} has been at the frontier of constraining new physics including IGM thermal evolution (Boera et al. 2019; Walther et al. 2019), neutrino masses (Croft, Hu & Dave' 1999a; Seljak et al. 2006; Palanque-Delabrouille et al. 2015a, b; Yeche et al. 2017) and the nature of dark matter (Boyarsky et al. 2009; Viel et al. 2013; Baur et al. 2016; Iršič et al. 2017a; Garzilli et al. 2019).

Two categories of P_{1D} data sets have emerged over the years. The first category contains thousands of low- to medium-resolution spectra obtained by the Extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) and its predecessors; and the corresponding P_{1D} estimates (McDonald et al. 2006; Palanque-Delabrouille et al. 2013; Chabanier et al. 2019). The upcoming Dark Energy Spectroscopic Instrument (DESI; Levi et al. 2013; DESI Collaboration 2016) aims to obtain approximately one million Ly α quasar spectra. Such large sample sizes can probe large scales to constrain cosmology and neutrino masses, but these data sets are limited by noise and resolution at small scales. The second category contains tens to hundreds of high-resolution, high-S/N spectra obtained by various spectrographs including the High-Resolution Echelle Spectrograph (HIRES; Vogt et al. 1994), the Ultraviolet and Visual Echelle Spectrograph (UVES; Dekker et al. 2000), and X-Shooter spectrograph (Vernet et al. 2011). P_{1D} estimates in this category often have been limited to their respective data sets with 10–100 spectra (Croft et al. 1999b; McDonald et al. 2000; Kim et al. 2004; Viel et al. 2013; Walther et al. 2017; Iršič et al. 2017b; Yeche et al. 2017; Boera et al. 2019; Day, Tytler & Kambalur 2019). The large-scale modes (∼10 Mpc) are poorly measured due to large

★ E-mail: naimgoksel.karacayli@yale.edu

© 2021 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

sample variance from the small numbers of quasars in this category, but these spectra can probe extremely small scales (∼100 kpc) that are crucial to constrain thermal state of the IGM and non-standard dark matter models.

In this work, we measure the small-scale P_{1D} from the largest sample of high-S/N quasars using a combination of three public releases (López et al. 2016; O'Meara et al. 2017; Murphy et al. 2019). This combined sample is seven times larger than Walther et al. (2017), but requires attention when combining the different data sets. We use the optimal quadratic estimator formalism and pipeline described in Karaçayli, Font-Ribera & Padmanabhan (2020) to measure P_{1D} . This approach has a number of advantages: by working in pixel space, it is unbiased by gaps in the spectra, allows weighting by both the pixel-level pipeline noise as well as accounting for sample variance from intrinsic Ly α correlations, and naturally allows combining very different data sets (with different pixel spacings, resolutions, etc.). As we demonstrate below, the large data set and pipeline result in significant improvements to the precision with which we measure P_{1D} .

This paper is organized in seven sections. Section 2 describes the spectra and the preprocessing steps we take. Section 3 details our method with a summary of the pipeline and mean flux measurement. We also review the quadratic estimator and validate the pipeline on simulated data here. Our results and a discussion on systematics of damped Ly α absorbers, continuum errors, and metal contamination can be found in Section 4. We reflect on our results and their statistical power in Section 5. Finally, we summarize in Section 6.

2 DATA

We use three publicly available data sets in this work:

(i) Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ) Data Release 2 (DR2; Lehner et al. 2014; O'Meara et al. 2015, 2017) is observed by HIRES (Vogt et al. 1994) on the Keck I telescope.

(ii) The Spectral Quasar Absorption Database (SQUAD) DR1 (Murphy et al. 2019) is observed by UVES (Dekker et al. 2000) on the European Southern Observatory's Very Large Telescope (VLT).

(iii) XQ-100 is observed using the X-Shooter spectrograph (Vernet et al. 2011) under the European Southern Observatory Large Programme 'Quasars and their absorption lines: a legacy survey of the high-redshift Universe with VLT/XSHOOTER' (López et al. 2016).

The availability of large, high-resolution, high-S/N spectra of KODIAQ and SQUAD also kindled analysis in two- and three-point correlation functions of Ly α absorbers (Maitra, Srianand & Gaikwad 2021).

2.1 KODIAQ

KODIAQ DR2[1] has 300 reduced, continuum-fitted, high-resolution quasar spectra at $0 < z < 5.3$ with resolving power $R \gtrsim 36\,000$ (Lehner et al. 2014; O'Meara et al. 2015, 2017). The continuum is fitted by hand one echelle order at a time using Legendre polynomials. These high-resolution spectra come in 1.3 or 2.6 km s^{-1} velocity spacing. We co-add and resample different observations on to a common 3 km s^{-1} grid using exposure times as weight (Gaikwad et al. 2021). While this resampling step is not required, it significantly reduces our computational cost while not affecting any of the scales of interest.

2.2 SQUAD

SQUAD DR1[2] consists of 467 fully reduced, continuum-fitted high-resolution quasar spectra at redshifts $0 < z < 5$ with resolving power $R \gtrsim 40\,000$ (Murphy et al. 2019). The continuum fitting consists of an automatic phase and then a manual phase to eliminate the remaining artefacts. These spectra are sampled on to varying 1.3–3.0 km s^{-1} spaced grids in velocity units. As with KODIAQ, we resample these on to a common 3 km s^{-1} grid.

There are two further important corrections to the reduced spectra. First, the median seeing is smaller than the slit width for some observations. This results in underestimated nominal resolution values and consequently overcorrecting spectrograph resolution. We correct the reported nominal resolution by approximating $R_{\rm cor} = Rs/\theta$ only when $s > \theta$, where s is the slit width and θ is the median seeing, both in arc seconds. This yields 25 per cent correction on average with a maximum of 150 per cent. The net effect on P_{1D} is less than 3 per cent even at $k = 0.1$ s km^{-1} since the resolution comes into effect at significantly smaller scales.

Secondly, Murphy et al. (2019) note that their pipeline underestimates the errors in saturated absorption lines. They provide χ^2_ν of each pixel about the weighted mean when combining multiple exposures. Following King et al. (2012)'s correction, we apply a median filter of size 5 to χ^2_ν and multiply the error with $\sqrt{\text{median}[\chi^2_\nu]}$ if it is greater than 1.

2.3 XQ-100

XQ-100 contains 100 quasars at redshifts $3.5 < z < 4.5$ with resolving power ranging from $R \sim 4000$–7000 and spectra from different arms made available (López et al. 2016). These spectra are obtained from the ESO database.[3] For each arm, the continuum is manually fit by selecting absorption free points. The Ly α forest falls into VIS and UVB arms. Spectra from VIS arm have a pixel spacing of 11 km s^{-1}, whereas UVB spectra are on 20 km s^{-1} velocity spaced grids. Given the lower resolution compared with KODIAQ and SQUAD, we do not resample these spectra on to another grid. We also keep the spectra from different arms separate, which allows us to keep resolution correction more accurate. For simplicity, we ignore any correlations between overlapping regions of these two arms.

Because the seeing is smaller than the slit width for most observations, the nominal resolution is similarly underestimated for this set. We correct the resolution for this effect by interpolating the tabulated values[4] only when the seeing is smaller than the slit width (Yeche et al. 2017). We do not extrapolate below the smallest provided slit width value. Even though this yields 30 per cent resolution correction on average and doubles the resolution at maximum, the effect in P_{1D} is smaller. We see an average of 15 per cent correction for UVB arm and 5 per cent for VIS arm in P_{1D} at $k = 0.045$ s km^{-1}, our confidence limit for XQ-100.

[1] https://koa.ipac.caltech.edu/workspace/TMP_939bFW_53591/kodiaq53591.html

[2] https://doi.org/10.5281/zenodo.1345974

[3] http://telbib.eso.org/detail.php?bibcode = 2016A~per~cent26A...594A..91L

[4] https://www.eso.org/sci/facilities/paranal/instruments/xshooter/inst.html

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

3 METHOD

3.1 Summary of the pipeline

Before processing the spectra, we need to correct the nominal resolution values, identify duplicate quasar observations and mark DLAs. We do not mask metal lines, but subtract a statistical estimate of the metal power using side band regions. First, the nominal resolutions of XQ-100 and SQUAD spectra are corrected for seeings that are less than the slit width as described in their respective sections. Secondly, assuming a data set does not contain duplicates in itself, we identify quasars within 10 arcsec of each other from different sets as duplicates, and pick the spectrum that has the highest S/N per km s^{-1}. Thirdly, we visually identify and remove DLAs with the help of a simple automated DLA finder and given catalogues (see Section 4.3).

We then process the spectra as follows[5]:

(i) Our analysis region is limited to $1.7 < z < 4.7$ and $1050\,\text{Å} < \lambda_{\rm RF} < 1180$ Å. We remove pixels that fall outside of these bounds.

(ii) These spectra are still susceptible to reduction artefacts such as spikes due to continuum normalization near an echelle order, sky subtraction or cosmic rays. We remove these outlier artefacts by eliminating pixels with flux values one median absolute deviation (MAD) outside 0 and 1, and by eliminating pixels with errors 3.5 MAD above the median error. In other words, we keep pixels that satisfy the following criteria:

$$-\,\text{MAD}(F) < F(\lambda) < 1 + \text{MAD}(F) \tag{1}$$

$$0 < \sigma(\lambda) < \text{median}(\sigma) + 3.5 \times \text{MAD}(\sigma), \tag{2}$$

where MAD is computed in the Ly α region. We prefer median statistics because they are robust against outliers.

(iii) We co-add multiple KODIAQ observations, then resample these and SQUAD spectra on to a common 3 km s^{-1} spaced grid. We keep XQ-100 data in its original spacing, and do not co-add UVB and VIS arms.

(iv) We divide by the best-fitting mean flux of the corresponding data set to get $\delta_F = F/\overline{F} - 1$.

(v) We divide the forest into three equal regions in the rest frame, and split all spectra into these chunks to speed up our calculation and help continuum marginalization. We remove chunks that are shorter than 10 per cent of the entire forest.

Combining KODIAQ and SQUAD with this process results in 1276 spectral chunks from 464 quasars comprised 186 KODIAQ and 278 SQUAD quasars. We are then left with 74 unique XQ-100 quasars. UVB arm contributes 62 chunks, while VIS contributes 47.

3.2 Mean flux

We measure the mean flux of each data set independently and use the respective best fits as mean flux to calculate $\delta_F = F/\overline{F} - 1$. As we discuss in Section 4, this removes some systematic errors in continuum fitting procedure.

We add Ly α variance to the pipeline error on pixel level using a fiducial power from fitting equation (4) to previous measurements and the mean flux from Faucher-Giguère et al. (2008, FG08). We assign the square of the mean S/N per km s^{-1} of the Ly α region as weight, then simply average pixel values. We do not use inverse variance at the pixel level because the pipeline noise and flux are correlated. Specifically, low flux regions almost always have smaller pipeline error estimates. Error for a given bin is the propagated value of the modified pipeline errors that go into that bin. Note that we do not account for systematics here and therefore underestimate the errors. We then fit Becker et al. (2013) form for $\overline{F}(z) = \exp(-\tau(z))$ to our measurements, where $\tau(z)$ is given by

$$\tau(z) = C + \tau_0 \left(\frac{1+z}{1+z_0}\right)^{\beta}, \tag{3}$$

and $z_0 = 3.5$. Fig. 1 shows our mean flux measurements from each data set by using $dz = 0.1$ spaced bins. Table 1 has the corresponding best-fitted parameters.

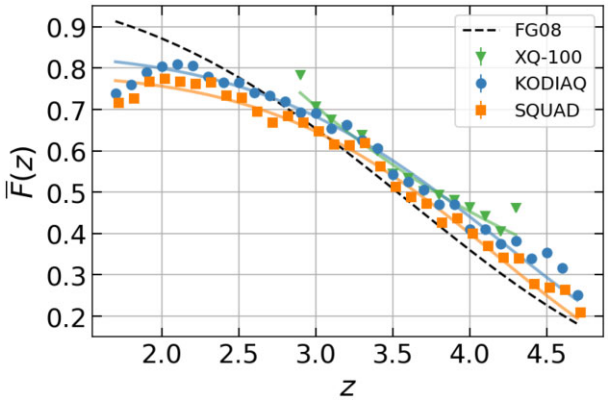

Figure 1. Mean flux measurement of each data set when DLAs are masked. Solid transparent lines show best-fitted results. KODIAQ and SQUAD have deviations at low z. FG08 is extrapolated to lower redshifts. We address the discrepancies in Section 4.

Table 1. Best-fitted parameters to our mean flux measurements of each data set when DLAs are masked.

Set	τ_0	β	C
FG08	0.675	3.92	0
KODIAQ	0.373	5.13	0.18
SQUAD	0.377	5.54	0.24
XQ-100	2.000	1.00	− 1.43

Using the errors we obtain from this analysis, we plot the S/N of each data set in Fig. 2. Since the systematics are not included, we overestimate the S/N, so this figure is for illustration purposes only. We assume signal mean flux to be from FG08. XQ-100 equally contributes to S/N distribution between $3 \lesssim z \lesssim 4$ even though it has lower resolution. Additionally, we provide the total redshift path-length covered by our final data set in Table 2.

3.3 Quadratic estimator

Our primary method is the quadratic maximum likelihood estimator (QMLE; Hamilton 1997; Tegmark, Taylor & Heavens 1997; Seljak 1998; Tegmark et al. 1998; McDonald et al. 2006). We refer the reader to Karaçayli et al. (2020) for details and extensive tests.[6] We also implemented a simple FFT estimator that does not account for noise or resolution as a rough cross-check for our results.

[5] https://bitbucket.org/naimgk/qsotools. Our pipeline code is not well documented, but we make it public for reference.

[6] https://bitbucket.org/naimgk/lyspeq

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

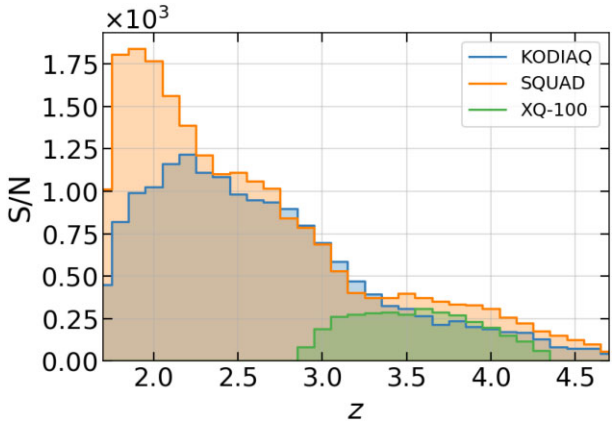

Figure 2. S/N as a function of redshift for each data set, where we define S/N to be $\overline{F}(z)/\sigma(z)$. The signal mean flux is assumed to be from FG08, $\overline{F} = \overline{F}_{\rm FG08}$. The noise in a redshift bin $\sigma(z)$ is the propagated error when mean flux is estimated, i.e. the error bars in Fig. 1. Note we do not account for systematics here, and therefore overestimate the S/N.

An important feature of our QMLE implementation is estimating deviations from a fiducial power spectrum such that $P(k,\, z) = P_{\rm fid}(k,\, z) + \sum_{m,\,n} w_{(mn)}(k,\, z)\theta_{(mn)}$, where we adopt top-hat k bands with k_n as bin edges and linear interpolation for z bins with z_m as bin centres (Font-Ribera, McDonald & Slosar 2018). Palanque-Delabrouille et al. (2013) provide a fitting function with best-fitted parameters. We further modify their fitting function with a Lorentzian decay (Karaçayli et al. 2020):

$$\frac{k\,P(k,\,z)}{\pi} = A\,\frac{(k/k_0)^{3+n+\alpha\,\ln k/k_0}}{1+(k/k_1)^2}\left(\frac{1+z}{1+z_0}\right)^{B+\beta\,\ln k/k_0}, \tag{4}$$

where $k_0 = 0.009\,{\rm s\,km^{-1}}$ and $z_0 = 3.0$. Combined with Walther et al. (2017) results, the best-fitted parameters are $A = 0.066$, $n = -2.685$, $\alpha = -0.22$, $B = 3.59$, $\beta = -0.16$, and $k_1 = 0.053\,{\rm s\,km^{-1}}$. We note that this fit has bad χ^2 and should not be used for scientific purposes, but it is sufficient for a baseline estimate.

Given a collection of pixels representing normalized flux fluctuations $\boldsymbol{\delta}_F$, the quadratic estimator is formulated as follows:

$$\hat{\theta}_\alpha^{(X+1)} = \sum_{\alpha'}\frac{1}{2}F^{-1}_{\alpha\alpha'}(d_{\alpha'} - b_{\alpha'} - t_{\alpha'}), \tag{5}$$

where X is the iteration number and

$$d_\alpha = \boldsymbol{\delta}_F^{\rm T}\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\boldsymbol{\delta}_F, \tag{6}$$

$$b_\alpha = {\rm Tr}\left(\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\mathbf{N}\right), \tag{7}$$

$$t_\alpha = {\rm Tr}\left(\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\mathbf{S}_{\rm fid}\right), \tag{8}$$

where the covariance matrix $\mathbf{C} \equiv \langle\boldsymbol{\delta}_F\boldsymbol{\delta}_F^T\rangle$ is the sum of signal and noise as usual, $\mathbf{C} = \mathbf{S}_{\rm fid} + \sum_\alpha \mathbf{Q}_\alpha\theta_\alpha + \mathbf{N}$, $\mathbf{Q}_\alpha = \partial\mathbf{C}/\partial\theta_\alpha$ and the estimated Fisher matrix is

$$F_{\alpha\alpha'} = \frac{1}{2}{\rm Tr}\left(\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\mathbf{Q}_{\alpha'}\right). \tag{9}$$

The covariance matrices in the right-hand side of equation (5) are computed using parameters from the previous iteration $\theta_\alpha^{(X)}$. Assuming different quasar spectra are uncorrelated, the Fisher matrix $F_{\alpha\alpha'}$ and the expression in parentheses in equation (5) can be computed for each quasar, then accumulated, i.e. $\mathbf{F} = \sum_q \mathbf{F}_q$, etc.

In order to adapt this to Ly α forest analysis, we first convert a pixel's wavelength to velocity using logarithmic spacing as has been the cosmology independent convention,

$$v_i = c\,\ln(\lambda_i/\lambda_{{\rm Ly}\,\alpha}) \tag{10}$$

$$z_i = {\rm e}^{v_i/c} - 1, \tag{11}$$

where $\lambda_{{\rm Ly}\,\alpha} = 1216$ Å.

Secondly, since the resolution matrix is not provided, we make the approximation that the resolution does not change with wavelength and is Gaussian for the rest of the paper. Then, the signal is the power spectrum multiplied with the spectrograph window function $W(k)$ in Fourier space.

$$S_{ij}^{\rm fid} = \int_0^\infty \frac{{\rm d}k}{\pi}\cos(k\,v_{ij})W^2(k)P_{\rm fid}(k,\,z_{ij}), \tag{12}$$

where $v_{ij} \equiv v_i - v_j$ and $1 + z_{ij} \equiv \sqrt{(1+z_i)(1+z_j)}$. The spectrograph window function is given by

$$W(k) = e^{-k^2R^2/2}{\rm sinc}(k\triangle v/2), \tag{13}$$

where R is the 1σ resolution and $\triangle v$ is the pixel width, both in velocity units. The derivative matrix for redshift bin m and wavenumber bin n is

$$Q_{ij}^{(mn)} = I_m(z_{ij})\int_{k_n}^{k_{n+1}}\frac{{\rm d}k}{\pi}\cos(k\,v_{ij})W^2(k), \tag{14}$$

where $I_m(z)$ is the interpolation kernel which is 1 when $z = z_m$ and 0 when $z = z_{m\pm1}$. We compute these matrices for as many redshift bins as necessary for a given spectrum.

Finally, we assume that the noise of every pixel is independent. This results in a diagonal noise matrix with $N_{ii} = \sigma_i^2$, where σ_i is the continuum normalized pipeline noise divided by the mean normalized flux $\overline{F}(z)$. We use the best-fitting mean flux of the corresponding data set for $\overline{F}(z)$.

3.4 Validation

Given the resolution and S/N diversity of KODIAQ, SQUAD, and XQ-100 data, it is crucial to verify the power spectrum estimates are unbiased and the errors are correctly estimated in an ideal statistical limit. To validate our method in these both aspects, we generate 100 independent lognormal mock data sets with exact redshift distribution, resolution, and noise properties of the data using the procedure described in Karaçayli et al. (2020). These synthetic spectra approximately produce expected mean flux redshift evolution (Faucher-Giguère et al. 2008) and power spectrum similar to Palanque-Delabrouille et al. (2013) and Walther et al. (2017). Even though these mocks cannot capture all the richness of data, they form the baseline with which we validate QMLE. We use bootstrap error estimates to partially capture effects not present in the mocks. We also discuss some astrophysical and instrumental effects in Section 4.6.

We use the true values for fiducial power spectrum and mean flux to estimate $P_{\rm 1D}$ with one iteration.[7] Fig. 3 shows a sample result from one mock set. From the average of these 100 estimates, we find that our results are unbiased in the range of interest $k < 0.1\,{\rm s\,km^{-1}}$, but they diverge from the truth for $k \gtrsim 0.2\,{\rm s\,km^{-1}}$. We note that these scales are noise-dominated, and further complicated by resolution effects and narrow metal lines in real data. Therefore, they are

[7]We note that using a fiducial power that is close to the truth yields better $P_{\rm 1D}$ estimates than not using any fiducial (Karaçayli et al. 2020).

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

Table 2. Total redshift path-lengths covered by combined KODIAQ and SQUAD sample (KS) and remaining unique XQ-100 quasars in our analysis z bins.

	z bin	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0	4.2	4.4	4.6
Redshift	KS	14.7	18.5	18.8	16.5	18.1	17.3	14.3	7.3	5.5	5.2	5.3	5.1	3.8	1.7	1.1
path	XQ-100	–	–	–	–	–	–	3.0	7.0	9.0	7.1	6.3	4.0	1.5	–	–
length	Total	14.7	18.5	18.8	16.5	18.1	17.3	17.3	14.3	14.5	12.3	11.6	9.1	5.3	1.7	1.1

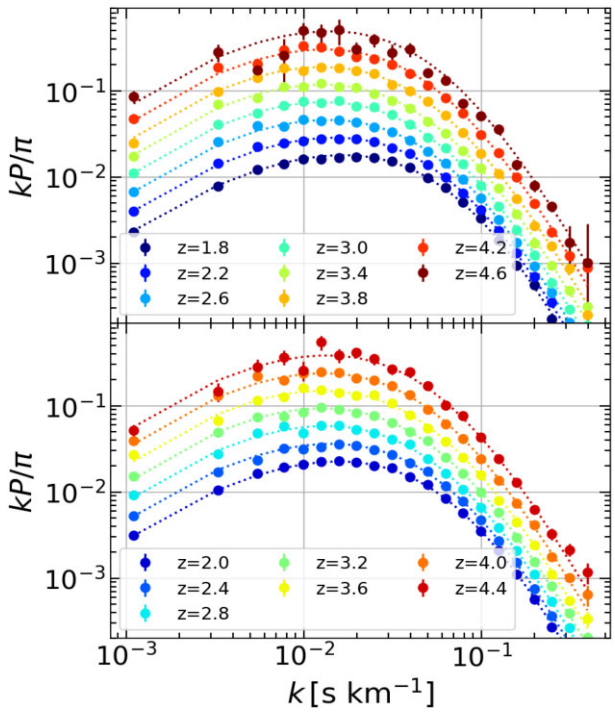

Figure 3. Power spectrum measurement from one mock set (circles) compared to the truth (dotted lines). Error bars are replaced with bootstrap estimates.

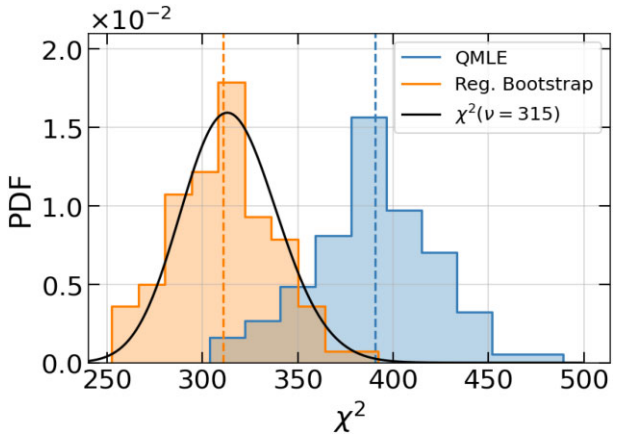

Figure 4. To validate our method, we performed χ^2 analysis on 100 mock sets using the Fisher matrices from QMLE and 25 000 bootstrap realizations from one set. We regularized the bootstrap Fisher matrix by exploiting its sparsity pattern and by flooring the eigenvalues to their Gaussian limits, which is noted as 'Reg. Bootstrap'. This produces the expected χ^2 distribution.

significantly hard to measure and have been absent in previous studies.

We then investigate the accuracy of estimated errors by a χ^2 analysis, and set the average of 100 results as the truth. First, we calculate the χ^2 of each result using the Fisher matrix from QMLE,[8] which yields larger values than expected. We identify the cause to be the violation of QMLE Gaussianity assumption at small scales by generating Gaussian mocks, which yields the correct χ^2 distribution. Therefore, we get another estimate of the covariance matrix by generating 25 000 bootstrap realizations from one set. We find that the diagonals of the bootstrap covariance matrix converge rapidly, but the off-diagonal terms remain noisy (and probably correlated) even with 250 000 realizations. We believe our sample size of approximately 1500 chunks is not enough for the degrees of freedom in the covariance matrix, which is a 315 × 315 matrix, or for its condition number $\mathcal{O}(10^{11})$. To achieve a stable covariance matrix, we apply a two-step regularization scheme.

After generating bootstrap realizations with spectral chunks, our final algorithm is based only on data and as follows:

(i) To prevent off-diagonal noise from leaking, we directly estimate the Fisher matrix element-wise, where $\left|r_{ij}^{\rm QMLE}\right| > 0.01$ and $r_{ij} \equiv F_{ij}/\sqrt{F_{ii}F_{jj}}$ using the algorithm in Padmanabhan et al. (2016). This algorithm further refines the element-wise estimate to the 'closest' positive-definite matrix.

(ii) We then find the eigenvalues λ_i and eigenvectors $\boldsymbol{e}_i$ of this Fisher matrix. We calculate the precision of these eigenvectors under Gaussianity: $\lambda_i^{\rm QMLE} = \boldsymbol{e}_i^T \mathbf{F}^{\rm QMLE} \boldsymbol{e}_i$. This is the theoretical maximum (minimum for the covariance), so we replace $\lambda_i \rightarrow \min(\lambda_i, \lambda_i^{\rm QMLE})$ and rebuild the Fisher matrix (McDonald et al. 2006).

We note that using the bootstrap eigenvectors as basis in step (ii) captures the non-Gaussian mode couplings. However, bootstraps are also missing fluctuations in certain modes due to low statistics; for example, there are only 5 chunks in the last redshift bin. This eigenvalue regularization scheme takes these modes to their Gaussian limit, and is a slight modification of McDonald et al. (2006).

Fig. 4 shows that the regularized bootstrap Fisher matrix produces the expected χ^2 distribution. These tests give us the confidence that our power spectrum estimates are unbiased, and the covariance can be estimated using bootstrap method at the scales of interest.

We tested our regularization scheme on the bootstrap covariance matrix and observed a marginally worse performance. The element-wise estimate is not positive-definite, but eigenvalue regularization fixes that problem. However, the Fisher matrix has a simpler structure than the covariance matrix. Furthermore, the χ^2 calculation is a linear operation with the Fisher matrix elements that further justifies the element-wise approximation, whereas inverting the covariance matrix is highly non-linear. Therefore, we opted for a direct estimation of the Fisher matrix in this paper.

[8]We note that for all sets QMLE gives the same Fisher matrix, and so the same covariance matrix.

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

We also test our mean flux measurement process with one mock set. We find that we obtain the correct mean flux, but the propagated errors are notably underestimated even with added Ly α variance, which can be suspected from Fig. 1 as well. We remind the reader that we are not interested in a precise mean flux measurement, but only in using it to remove some continuum fitting systematics from spectra. However, we perform another test to quantify the underestimation of these errors. We resample the spectra on to a coarse grid (300 km s^{-1}) using inverse variance (note signal and errors are not correlated for the mocks). We then estimate the standard deviation using these coarse pixels in a given redshift bin, and find that they are four times the propagated errors. We also note that the correlations between pixels actually matter for the resolution and S/N of our data when resampling (Slosar et al. 2013), but it is still unclear if they could yield correct error bars when propagated. We speculate that the systematics are most likely a significant source of error for the measurement from data however.

4 RESULTS

We measure the power spectrum in 15 redshift bins from $z = [1.8, 4.6]$ with $dz = 0.2$ spacing for 21 k bins of which the first 4 are linearly spaced with $\Delta k_{\rm lin} = 0.0022$ s km^{-1} and the rest are logarithmically spaced with $\Delta k_{\rm log} = 0.1$.

To remind the reader our pipeline, we mask DLAs by using the accompanying catalogues and visually identifying remaining absorbers with damping wings. We do not mask metal lines, but provide a statistical estimate of the metal contamination using side band regions. Furthermore, we marginalize out the constant and the slope ($\ln \lambda$) terms of the continuum errors. Because the continuum is fitted piece-wise and corrected manually, these are not the exact expressions for the continuum errors, but the marginalization should still remove some or most of the contamination in the largest scales. These points and their contribution to systematic error budget are further discussed below.

We perform an initial run on the combined set with the fiducial power spectrum as described in Section 3. We fit the results for $k < 0.1$ s km^{-1} to get a new estimate for the fiducial power, and then perform one iteration. The new parameters are $A = 0.084$, $n = -2.655$, $\alpha = -0.155$, $B = 3.64$, $\beta = 0.32$, and $k_1 = 0.048$ s km^{-1}. We showed this process yields better power spectrum estimates in Karaçayli et al. (2020).

The estimated errors from QMLE are under Gaussian assumption and significantly underestimated at small scales and high redshifts. We estimate the error bars by generating bootstrap realizations and using our regularization scheme as discussed in Section 3.4. Since our method divides pixel pairs into two redshift bins, the power spectrum estimates are correlated between these redshift bins. Fig. 5 shows an example normalized covariance matrix from bootstrap estimation for $k = 1.3 \times 10^{-2}$ s km^{-1} bin.

4.1 Consistency between data sets

Our aim in this work is to utilize the combined data for the best S/N, but first we compare individual power spectrum estimates to check our treatment of the data and consistency between data sets. Fig. 6 lays out the ratios of individual power spectra to our raw measurement from KODIAQ + SQUAD (KS).

The first notable difference is between KODIAQ and SQUAD at $z = 1.8$, but they both still fall within the large errors of Walther et al. (2017). Otherwise, these two sets agree with each other. To quantify the agreement between them, we calculate χ^2 from the power difference:

$$\chi^2 = \boldsymbol{d}^T (\mathbf{C}_1 + \mathbf{C}_2)^{-1} \boldsymbol{d}, \tag{15}$$

where $\boldsymbol{d} = \boldsymbol{P}_1 - \boldsymbol{P}_2$. This yields $\chi^2 = 218$ for 180 degrees of freedom (dof) while removing $z = 1.8$ and $k > 0.1$ s km^{-1} bins. This value is 2σ away from the mean; or in other words, probability of getting $\chi^2 > 218$ is 3 per cent. This is a reasonable probability when comparing two data sets, so we conclude they are consistent and remove $z = 1.8$ from our conservative range.[9]

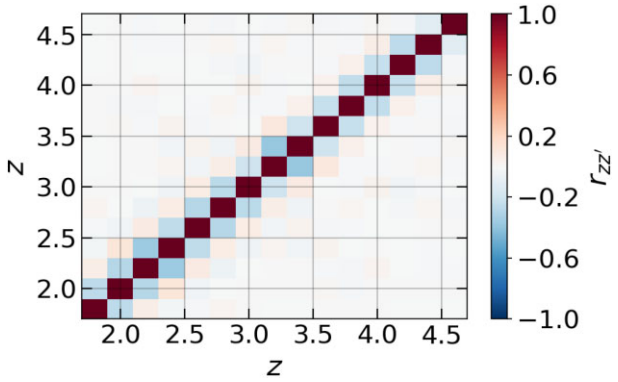

Figure 5. The normalized bootstrap covariance matrix of data for $k = 0.013$ s km^{-1} bin. Neighbouring redshift bins are anticorrelated because of assigning pixel pairs into two redshift bins.

We find that our XQ-100 results agree with Iršič et al. (2017b) in their provided range. The largest inconsistency here most visibly comes from $k > 0.06$ s km^{-1} values. We remind the reader that we corrected the resolution for seeing to the best of our knowledge, but it can still carry some errors especially at this range given the Nyquist frequency for XQ-100 is $k \sim 0.1$ s km^{-1}. So, accurately measuring these modes is an ambitious goal for XQ-100 alone. Given our limited understanding of the resolution, we assign a conservative $k < 0.045$ s km^{-1} range for our XQ-100 measurement, where $\chi^2 = 83$ for 77 dof when systematic errors are included. Unfortunately, this makes combining XQ-100 at the chunk level complicated as QMLE in its current form cannot account for systematic errors. We defer how these systematic errors can be incorporated into QMLE to a future work. Instead, we provide a separate measurement and a covariance matrix for XQ-100 in this paper.

4.2 Final combined results

Let us start this section with a review of the weighted average and its error propagation. Assume for a data set m, we have measurements $\boldsymbol{x}_m$ with respective errors $\mathbf{C}_m$. Using normalized weights $\mathbf{W}_m$, the weighted average $\boldsymbol{x}_W$ and resulting error $\mathbf{C}_W$ are given by

$$\boldsymbol{x}_W = \textstyle\sum_m \mathbf{W}_m \boldsymbol{x}_m, \tag{16}$$

$$\mathbf{C}_W = \textstyle\sum_m \mathbf{W}_m \mathbf{C}_m \mathbf{W}_m^T + \sum_{m \neq n} \mathbf{W}_m \mathbf{C}_{mn} \mathbf{W}_n^T, \tag{17}$$

where $\mathbf{C}_{mn}$ is the cross-covariance between measurements m and n. These become the usual inverse variance weighted average ex-

[9]We have a number of hypotheses for this discrepancy. First, the Ly α forest at low redshifts will be in the UV where the response of most spectrographs rapidly falls. Also, metal contamination would have the largest effect at these redshifts and is hard to measure for these samples. However, we were not able to isolate a particular cause, which is why we exclude this redshift bin from our final results.

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

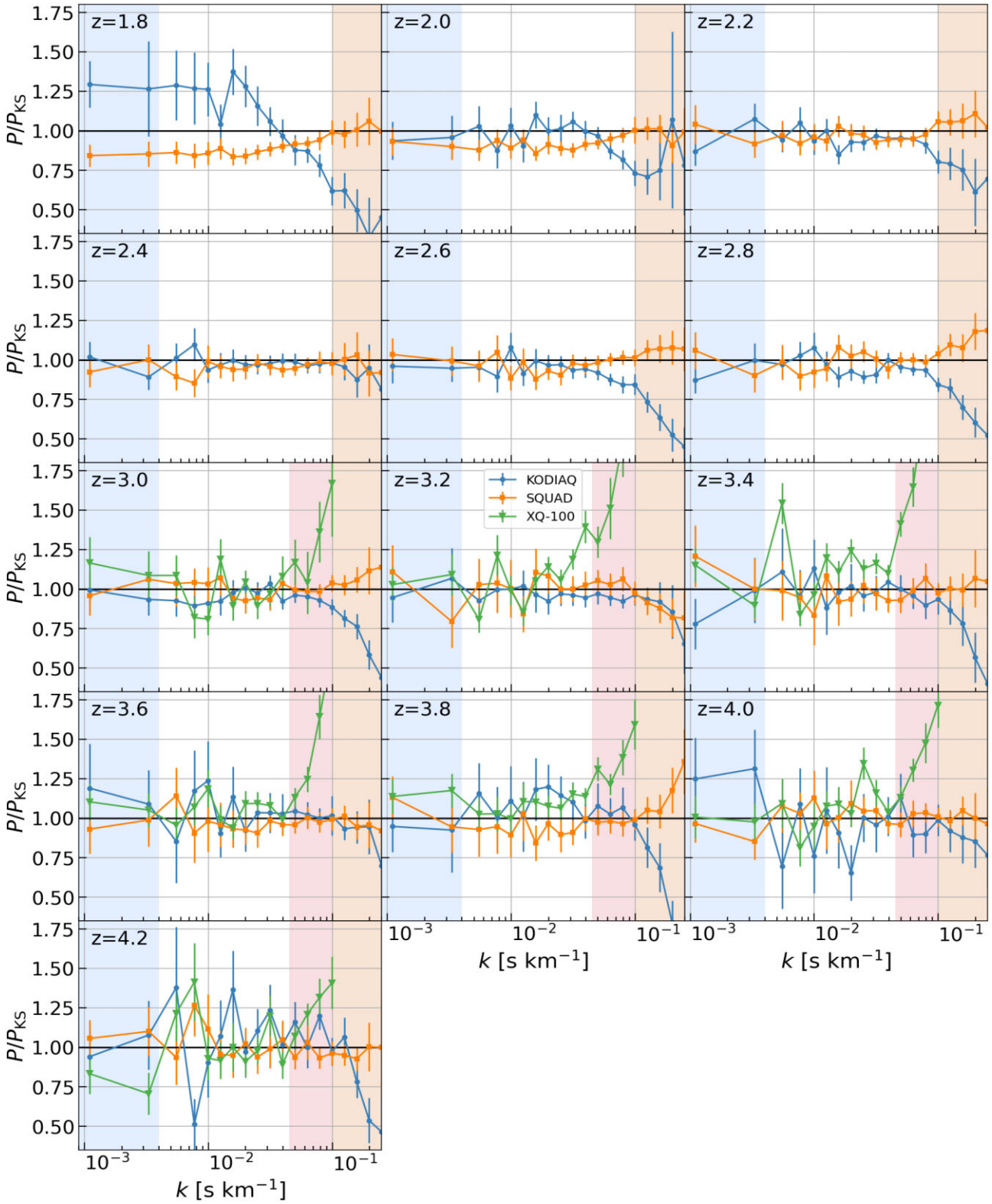

Figure 6. The raw power spectrum measurements of KODIAQ, SQUAD, and XQ-100 divided by the raw power from KODIAQ + SQUAD ($P_{\rm KS}$) when DLAs are masked. Error bars are from bootstrap. We highlight the large-scale modes $k \lesssim 4 \times 10^{-3}\,{\rm s\,km^{-1}}$ that are susceptible to continuum errors in light blue, and small-scale modes $k \gtrsim 0.1\,{\rm s\,km^{-1}}$ that are affected by noise, metal contamination and resolution effects in light orange. Light pink region is excluded from our XQ-100 analysis.

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

pressions with $\mathbf{W}_m \propto (\mathbf{C}_m)^{-1}$ if the measurements are uncorrelated. In our case, however, some systematic errors (DLAs, continuum and metals) between KS and XQ-100 are data-independent and correlated, whereas resolution systematics for example are not. Let us define $\Lambda^{\rm cor}$ to be these data-independent, correlated systematics between KS and XQ-100, and define $\Lambda_m^{\rm uncor}$ to be the data-dependent, uncorrelated systematics for measurement m. Then, we substitute $\mathbf{C}_{mn} = \Lambda^{\rm cor}$ and $\mathbf{C}_m = \mathbf{C}_m^{\rm stat} + \Lambda_m^{\rm uncor} + \Lambda^{\rm cor}$, and correlated systematic comes out since $\sum_m \mathbf{W}_m \equiv \mathbf{1}$.

$$\mathbf{C}_W = \sum_m \mathbf{W}_m \left(\mathbf{C}_m^{\rm stat} + \Lambda_m^{\rm uncor}\right) \mathbf{W}_m^T + \Lambda^{\rm cor}. \tag{18}$$

In our case, $\Lambda^{\rm cor} = \Lambda^{\rm DLA} + \Lambda^{\rm Cont} + \Lambda^{\rm Metal}$ and $\Lambda_m^{\rm uncor}$ is the resolution systematics. In short, uncorrelated systematics will become smaller when different measurements are combined, while correlated systematics stay the same.

We now use the inverse variance weights with systematic errors, which also prevents underestimated statistical errors from dominating the average:

$$\mathbf{W}_m \propto \left(\mathbf{C}_m^{\rm stat} + \Lambda_m^{\rm uncor} + \Lambda^{\rm cor}\right)^{-1}. \tag{19}$$

We again generate 25 000 bootstrap realizations to calculate the statistical Fisher matrix. Our systematic error budget is the same for KS and XQ-100 measurements except for the resolution (see Section 4.6) and is added in quadrature to the diagonal. We also add large numbers to the diagonal of XQ-100 covariance at $k > 0.045\,{\rm s\,km^{-1}}$ to remove these modes from the average. For both measurements, we also subtract the metal power and add its covariance first, so that the weights have all the statistical and systematic errors in place.

Our results from this weighted average are shown in Fig. 7, which is in good agreement with previous measurements even at the largest scales (Iršič et al. 2017b; Walther et al. 2017; Chabanier et al. 2019). Note that we do not directly compare with Gaikwad et al. (2021) since their method involves forward modelling the spectra and does not subtract noise nor deconvolve the spectrograph window function. We caution the reader that the number of quasars and different continuum treatments limit the accuracy of our measurement for $k \lesssim 4 \times 10^{-3}\,{\rm s\,km^{-1}}$; and refer the reader to eBOSS (Chabanier et al. 2019) for a better measurement at these scales. Nevertheless, we are encouraged by this agreement at the large scales.

4.3 Damped Ly α absorbers

The net result of damped Ly α absorbers (DLA) on the power spectrum comprises three effects. The primary effect is the large amount of power added to the largest scales due to their damping wings. The central region of complete absorption has two competing effects: suppression of power and amplification because of lower mean flux measurement (McDonald et al. 2005; Rogers et al. 2018). These high-column density absorbers are difficult to model and simulate, so they are often removed from spectra. We describe our DLA identification and removal in this section.

The SQUAD data set comes with a DLA catalogue; and the ones in XQ-100 are identified in Sánchez-Ramírez et al. (2016). We mask a reported DLA at $z_{\rm abs}$ between $[\lambda_C - W/2, \lambda_C + W/2]$, where $\lambda_C = (1 + z_{\rm abs})\lambda_{{\rm Ly}\,\alpha}$ and W is the equivalent width given by the following equation (section 16.4.4 of Mo, van den Bosch & White 2010):

$$W = 7.3(1 + z_{\rm abs})\sqrt{\frac{N_{\rm HI}}{10^{20}\ {\rm cm^{-2}}}}\ {\rm Å}. \tag{20}$$

We also apply a simple automated DLA finder which first finds regions that are consecutively $F(\lambda) < \overline{F}(\lambda) + \sigma(\lambda)$. If a region is longer than the equivalent width for an absorber with $N_{\rm HI} = 10^{19}$ cm^{-2} in equation (20) at the central redshift, we mark it as a DLA candidate. We then visually inspect all these candidates for damping wings to remove from data.

Our results in Fig. 7 are already without DLAs. We performed a run where we kept in the DLAs as well. Fig. 8 shows the effect described in the beginning of this section. The presence of DLAs significantly affects the largest scales $k \lesssim 3 \times 10^{-1}\,{\rm s\,km^{-1}}$, while boosting the intermediate scales by few per cent because of lower mean flux estimates.

4.4 Continuum

The observed flux $f(\lambda)$ is divided by the quasar continuum $C(\lambda)$ and the mean normalized flux $\overline{F}(\lambda)$ to obtain the flux fluctuations $\delta_F(\lambda)$. Since $C(\lambda)$ is smooth, errors in this process propagate to mostly large scales and are called the continuum errors.

We start our discussion by considering a systematic scaling of the continuum in a given data set. In other words, we assume the fitted continuum of a quasar q is the true continuum times a constant, $C_q^*(\lambda) = aC_q(\lambda)$, for all quasars in the set. This inevitably scales the normalized flux by the same factor, $F_q^* = f_q/C_q^* = F_q/a$, and hence the measured mean flux $\overline{F}^* = \overline{F}/a$ since a does not depend on q. Therefore, using the measured mean flux $\overline{F}^*$ of a given data set removes this systematic bias from the fluctuations $\delta_{F,q}^* = F_q^*/\overline{F}^* - 1 = \delta_{F,q}$. In fact, we see this type of measured mean flux scaling across data sets, where KODIAQ has 6 per cent and XQ-100 has 15 per cent excess mean flux on average compared to SQUAD. We are further encouraged by the agreement of $P_{\rm 1D}$ without any scaling between data sets. Finally, note that this multiplicative scaling would cancel even if it depended on observed wavelength $a = a(\lambda)$ as long as a does not change from quasar to quasar.

Now, let us introduce a quasar dependent error $\eta_q(\lambda)$ that remains after this systematic scaling, i.e. $C_q^*(\lambda) = aC_q(\lambda)[1 + \eta_q(\lambda)]$. Since the continuum itself is smooth, $\eta_q(\lambda)$ will also be smooth and slowly changing given a well-behaving continuum fitting procedure. The effect of this quasar dependent error will propagate to the mean flux measurement by some average.

$$\overline{F}^* = \frac{\overline{F}}{a}\left\langle \frac{1}{1+\eta_q} \right\rangle_q \approx \frac{\overline{F}}{a}(1 - \bar{\eta}), \tag{21}$$

where we have assumed η_q is small and not correlated with F_q. Then,

$$\delta_{F,q}^* = \frac{F_q^*}{\overline{F}^*} - 1 = \frac{F_q/\overline{F}}{(1+\eta_q)(1-\bar{\eta})} - 1 \tag{22}$$

$$= (1 + \delta_{F,q})(1 - \Delta\eta_q) - 1. \tag{23}$$

Finally,

$$\delta_{F,q}^*(\lambda) = \delta_{F,q}(\lambda)[1 - \Delta\eta_q(\lambda)] - \Delta\eta_q(\lambda), \tag{24}$$

where we have defined $\Delta\eta_q(\lambda) \equiv \eta_q(\lambda) - \bar{\eta}(\lambda)$. Unfortunately, this error does not fully cancel, but it is possible to partially marginalize out the additive term $\Delta\eta_q(\lambda)$ by approximating its form. This error is described by a constant shift and a slope in eBOSS pipeline, $\Delta\eta_q(\lambda) = a_q + b_q \ln\lambda$. We marginalize out these modes in this work as well, even though they are not the exact expressions for the continuum errors because the continuum is fitted piece-wise and then corrected manually. Dividing the forest into three chunks also helps by limiting the wavelength range of $\Delta\eta_q(\lambda)$ that needs to be described. Note that since $\Delta\eta_q(\lambda)$ is smooth, this error affects mostly large scales even when uncorrected. We ignore the multiplicative term within our approximation. Its major complication would stem

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

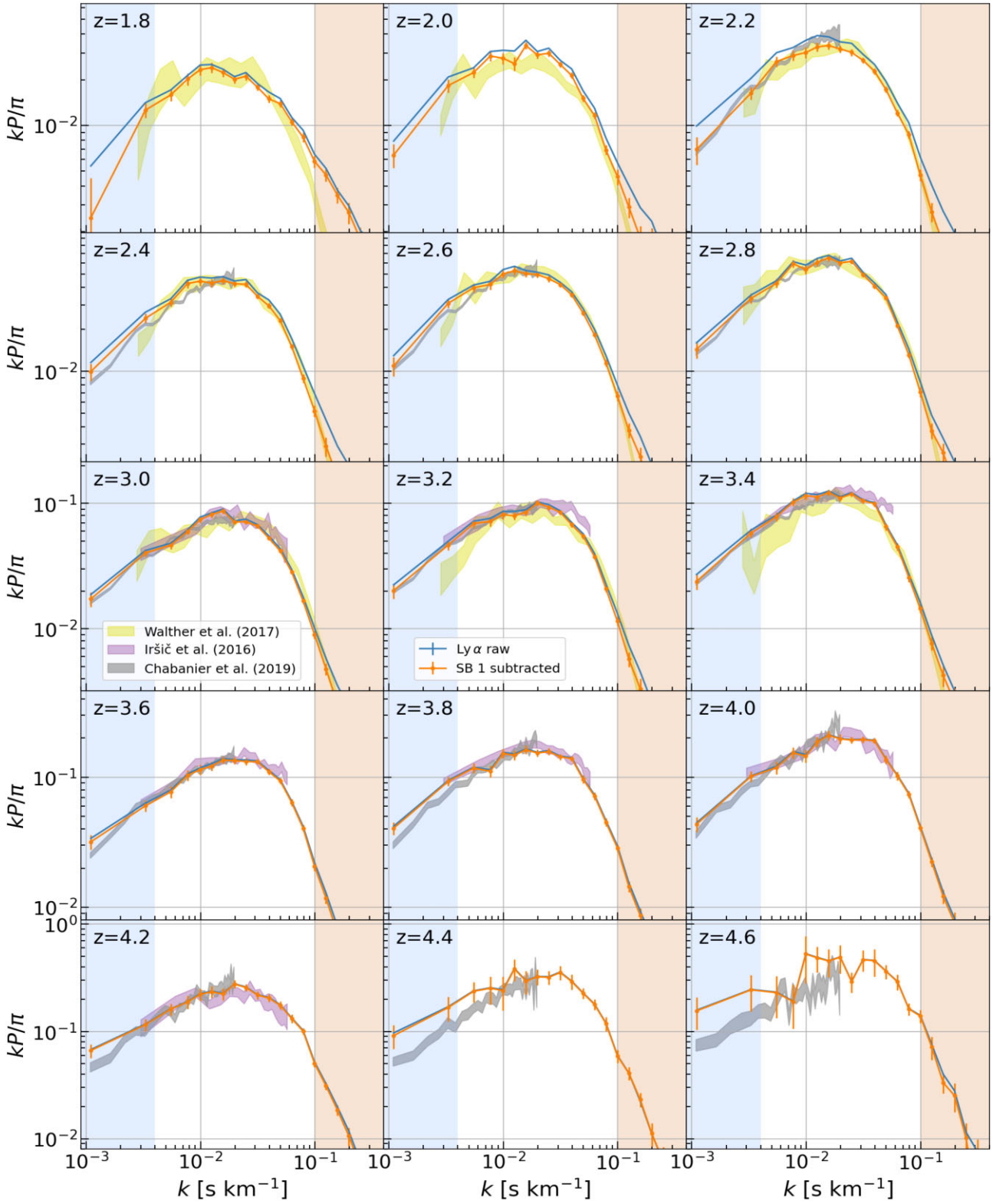

Figure 7. Our final raw and side band subtracted power spectrum measurements when DLAs are masked, where all three sets are combined as described in Section 4.2. The error bars are from bootstraps; and systematic error budget is added in quadrature. We highlight the large-scale modes $k \lesssim 4 \times 10^{-3}\,\mathrm{s\,km^{-1}}$ that are susceptible to continuum errors in light blue, and small-scale modes $k \gtrsim 0.1\,\mathrm{s\,km^{-1}}$ that are affected by noise, metal contamination and resolution effects in light orange. We remind the reader that the last two z bins have low statistics as well as further continuum fitting complications.

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

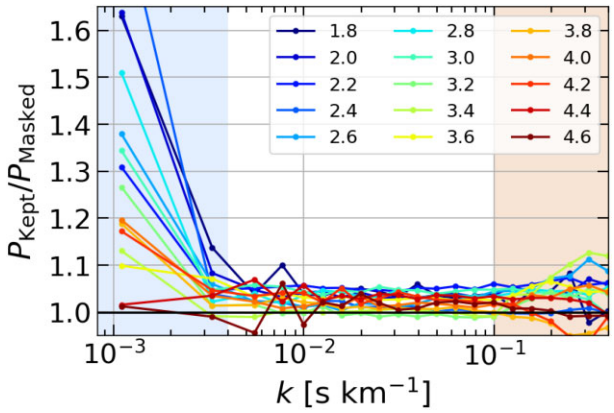

Figure 8. The ratio of the measured power spectrum from the combined data set between DLAs kept $P_{\rm Kept}$ and DLAs masked $P_{\rm Masked}$. The damping wings add power to large scales, and the lower mean flux overcomes the full absorption suppression from medium to small scales by changing it ≈ 5 per cent.

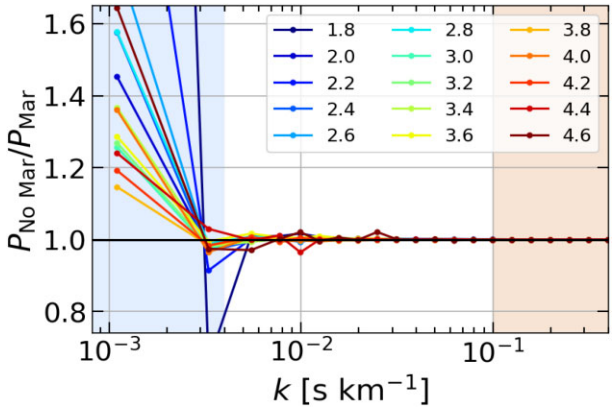

Figure 9. The ratio of the measured power spectrum without continuum marginalization $P_{\rm No\ Mar}$ with marginalization $P_{\rm Mar}$. Continuum errors significantly contaminate the largest scales; the rest fluctuates around less than one per cent.

from the correlations between $\eta_q(\lambda) - F_q(\lambda)$ and $\eta_q(\lambda) - \eta_q(\lambda')$, which can ultimately be treated by smaller errors or by more uniform continuum fitting procedures.

We performed another run where we turned off the continuum marginalization to show its effect. Fig. 9 shows the largest scales $k \lesssim 3 \times 10^{-3}$ s km^{-1} are again significantly affected by the continuum. However, the intermediate scales deviate by less than one per cent.

4.5 Side band power

The red side of the Ly α line in the spectrum is free from H I absorption, so this region is used to estimate the power from metals and other systematics (McDonald et al. 2006; Palanque-Delabrouille et al. 2013; Chabanier et al. 2019). We define the first side band (SB 1) region to be between 1268–1380 Å, below the Si IV line, in quasar's rest frame, and the second side band region to be between 1409–1524 Å, below the C IV line. We use $\overline{F}(z) = 1$, and estimate the power on the same k bins. We note that this only removes power due to metals with $\lambda_{\rm RF} \gtrsim 1400$ Å, and hence some metal contamination still remains and produces oscillatory features such as Si III-Ly α cross-correlation (McDonald et al. 2006; Palanque-Delabrouille et al. 2013).

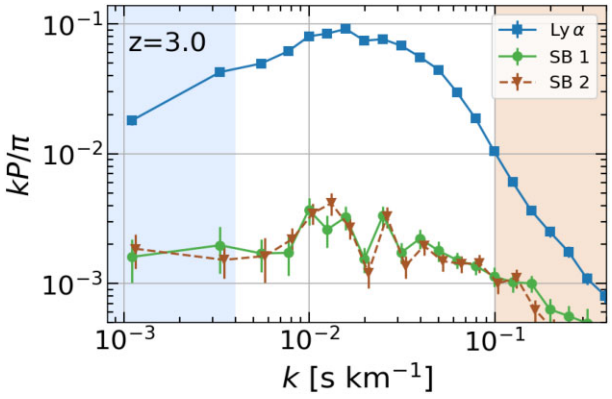

Figure 10. The power spectra measured from the two side bands at $z = 3.0$. The C IV doublet causes the oscillation peaks at $k = 1.3 \times 10^{-2}$ s km^{-1} and 2.5×10^{-2} s km^{-1}. This doublet and others are apparent in the correlation function (see Fig. 11).

We limit the side band power estimate to KODIAQ and SQUAD data sets to not further complicate analysis. The final metal power is estimated from the Si IV region (SB 1), and uses 1391 chunks with 221 KODIAQ and 271 SQUAD quasars.

The initial runs for both side band regions use 10 per cent of the fiducial power in Section 3. We then switch to the best-fitted parameters for $k < 0.1$ s km^{-1} as the new fiducial and estimate side band powers with one iteration. These new parameters are $A = 0.0027$, $n = -2.92$, $\alpha = -0.174$, $B = 0.236$, $\beta = -0.01$ where the Lorentzian term can be ignored as $k_1 \sim 10^{10}$ s km^{-1}. We again estimate the covariance matrix using 25 000 bootstrap realizations and use these instead of QMLE errors. We find that the two side band powers mostly agree with each other with some offset at $z = 2.0$, 3.2, and 3.6 bins, while $z = 2.2$ has a larger offset. Fig. 10 shows and compares the side band powers to the Ly α power at $z = 3.0$.

Noteworthy features in both side band powers are the two visible peaks due to oscillations in all redshift bins at $k = 1.3 \times 10^{-2}$ s km^{-1} and 2.5×10^{-2} s km^{-1}. While bootstrap method shows that the QMLE errors are underestimated for $z < 3$ and overestimated for $z > 3$, these peaks remain visible in QMLE error to bootstrap error ratio. We find these oscillations are due to C IV doublet, which is manifested as a peak in the 1D correlation function at $v \approx 500$ km s^{-1} separation. Moreover, we identified more doublet features in the 1D correlation function with a simple model. First, we estimate the correlation function by inverse FFTing the power spectrum before binning into k bins. This yields the correlation function on the full-resolution grid, which we then bin into v bins. Doublets of the same absorber manifest as peaks in this correlation function. To increase S/N, we average all redshift bins.

We found that the correlation function has a smooth component, which we model as follows:

$$\xi_{\rm smooth}(v) = C + \frac{x_0}{1 + (v/v_0)^\gamma}, \tag{25}$$

where C, x_0, v_0, and γ are fitting parameters. We then model the peaks as Gaussian functions:

$$\xi_p(v) = A_p e^{-(v-\mu_p)^2/2\sigma_p^2}, \tag{26}$$

where A_p and σ_p are free parameters, and the doublet separation is fixed to tabulated values: $\mu_p = c \ln(\lambda_2^p/\lambda_1^p)$. We fit for known metal doublets in Table 3 by summing these functions: $\xi = \xi_{\rm smooth}$

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

Table 3. Doublets used in SB correlation function analysis.

Doublet	$\lambda_1 - \lambda_2$ [Å]	v [km s^{-1}]
C IV	1548.20 – 1550.77	497.2
Mg II	2796.35 – 2803.53	768.7
Fe II	2374.46 – 2382.76	1046.6
Fe II'	2586.65 – 2600.17	1563.2
Al III	1854.72 – 1862.79	1302.2
Si IV	1393.76 – 1402.77	1932.8

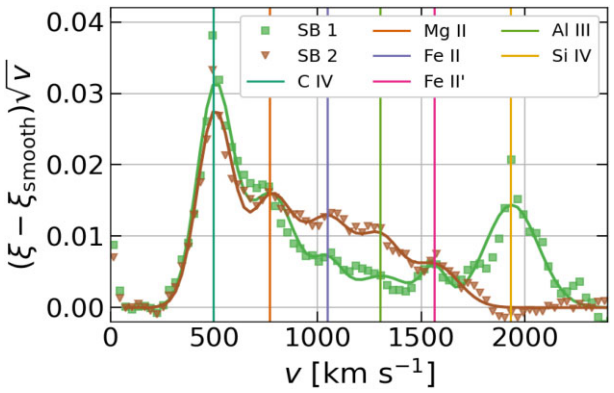

Figure 11. The residual correlation functions for SB 1 (green squares) and SB 2 (brown triangles). Curves with solid lines are our fitting model for known metal doublets. Our simple fitting function captures the important features. Note again that Si IV line is outside of SB 2, so there is no peak at $v \approx 2000$ km s^{-1}.

$+ \sum_p \xi_p$. The residuals and their respective fits are in Fig. 11. This simple model accurately describes the important features in both correlation functions.

4.6 Systematic error budget

We identified four possible sources of systematic errors in our analysis, which are difficult to rectify by models. The continuum is fitted by hand and hard to reproduce. DLA removal is also partly manual in our measurement. For metals, no model at sufficient resolution exists. Our systematic error budget is similar to Chabanier et al. (2019) and based on the fiducial power parameters given in Section 4. We provide each systematic error separately in our files, so that they can be scaled to other values readers see fit. However, we recommend that these numbers are not modified for immediate cosmological fits. Our intention is to leave room for focused studies of each contaminant such as sub-DLAs, Lyman-limit systems (LLS), resolution modelling, and metal line blending. These require dedicated studies; and the reader can adjust the error budget according to their findings.

(i) **Resolution:** As we have discussed in Section 2, seeing conditions alter the resolution. Due to this correction, we expect resolution accuracy to be smaller than normal. XQ-100 reports seeings with an average of 10 per cent precision. SQUAD, however, is much worse at 40 per cent. We then assign corresponding inaccuracies to R values.

$$\sigma_{\rm Res} = r_m \times P_{\rm fid}(k) \times 2k^2\overline{R}^2, \tag{27}$$

where $\overline{R}$ is the average resolution in a given redshift bin and r_m is 0.1 for XQ-100 and 0.4 for KS. Note that even an undesirable 40 per cent change in resolution for SQUAD-like spectrum results in only 1–3 per cent change in $P_{\rm 1D}$ at $k = 0.1$ s km^{-1}. Our reported precision at this value is at least twice as large. Here, we assumed the resolving power is provided with perfect precision by the data sets, and ignored its contribution to the budget.

(ii) **Metal:** There are two reasons to include a systematic error budget for metals: (1) Our statistical estimate of the metal power might be off. (2) Bootstrap error estimates might not reflect the truth. Assuming metal power is nearly constant with k and z, we compared the fluctuations between redshifts and bootstrap error estimates. We found on average fluctuations between redshift bins are 6 per cent larger than the bootstrap error estimates. Leaving room for the blending of metal lines we pick 10 per cent for our budget (Day et al. 2019). Using the fiducial power of the side band estimates, the systematic error is given by

$$\sigma^{\rm (Syst)}_{\rm Metal} = 0.10 \times P^{\rm (SB1)}_{\rm fid}(k). \tag{28}$$

(iii) **Incomplete DLA removal:** Even though we used visual identification and catalogues, we leave room for missed DLAs or other high-density absorbers by assigning 1 per cent incompleteness to this possibility. We multiply the fiducial power estimate with redshift average of the absolute ratio of $P_{\rm Kept}/P_{\rm Masked} - 1$ in Fig. 8 and DLA incompleteness ratio.

$$\sigma_{\rm DLA} = 0.01 \times P_{\rm fid}(k) \times \left\langle |P_{\rm Kept}/P_{\rm Masked} - 1| \right\rangle_z \tag{29}$$

We note that the effects of sub-DLAs and LLS that are absent in catalogues can be larger than 1 per cent (Rogers et al. 2018).

(iv) **Continuum:** We assign 10 per cent inefficiency to continuum marginalization and use the redshift average of the absolute ratio of $P_{\rm NoMar}/P_{\rm Mar} - 1$ in Fig. 9. This error is also scaled with the fiducial power.

$$\sigma_{\rm Cont} = 0.10 \times P_{\rm fid}(k) \times \langle |P_{\rm NoMar}/P_{\rm Mar} - 1| \rangle_z \tag{30}$$

Note that continuum errors themselves could be larger than 10 per cent. Our method, however, prevents these errors from contaminating $P_{\rm 1D}$ measurements by marginalizing out bulk of them, namely a constant and a slope per chunk (not spectrum). Here, our systematic budget really comes from the efficiency of this removal, which we assume to be 90 per cent effective. Furthermore, the modes that are most affected are not in our conservative range.

The results are summarized in Fig. 12. DLA and continuum systematics affect the largest scales as expected, but DLA systematics are an order of magnitude smaller. Errors due to resolution inaccuracy become relevant near $k = 0.1$ s km^{-1}, but they do not overcome statistical errors. Metal systematics are relevant at low redshifts. In our conservative 0.004 s km^{-1} $< k < 0.1$ s km^{-1} and $z > 1.8$ range, systematic errors are 19 per cent of the statistical errors on average.

5 DISCUSSION

5.1 Statistical power of our results

We would like to compare the statistical power of our results to the existing measurements. In order to do so, we come up with a crude Fisher forecasting analysis, which replaces the actual $P_{\rm 1D}$ measurement with a fiducial power spectrum and only takes the covariance matrix into account. Our model is unquestionably primitive and ignores many complications including thermal state of the IGM. But it will serve as an adequate frame of reference.

We limit our simple forecast to a single parameter: the cut-off scale $k_{\rm cut}$. We assume a fiducial model as the 'truth' for all data sets.

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

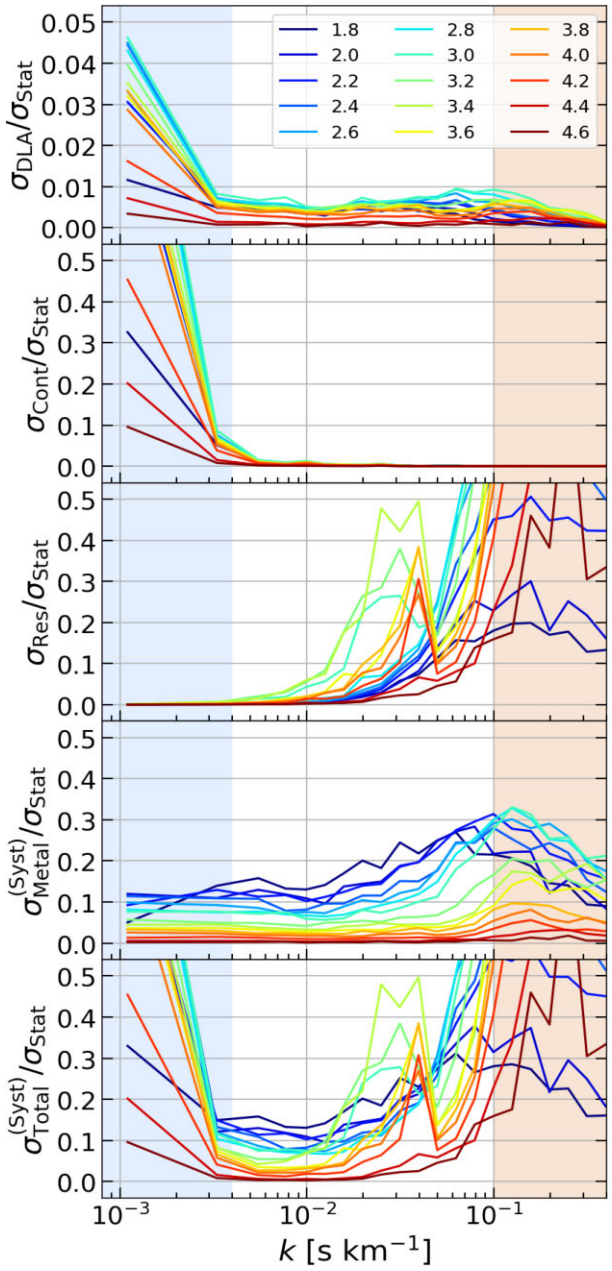

Figure 12. Systematic error estimates divided by the SB 1 subtracted bootstrap estimates. Continuum systematics affect the largest scales that are outside our conservative range. DLA systematics behave similarly, but they are an order of magnitude smaller. Metal systematics are relevant at low redshifts or high k.

Then, the error on $k_{\rm cut}$ is given by

$$\frac{1}{\sigma^2_{\rm cut}} = \frac{\partial^2 \mathcal{L}}{\partial k^2_{\rm cut}} = \boldsymbol{P}^T_{,k_{\rm cut}} \mathbf{F} \boldsymbol{P}_{,k_{\rm cut}}, \tag{31}$$

where $\boldsymbol{P}_{,k_{\rm cut}}$ is analytically calculated and the Fisher matrix $\mathbf{F}$ is taken from actual measurements. We assume the following relative suppression transfer function for the 3D power spectrum.

$$T^2(k) = [1 + (k/k_{\rm cut})^2]^{-10}. \tag{32}$$

This expression comes from warm dark matter models as a fitting function (Bode, Ostriker & Turok 2001; Viel et al. 2005). We pick $k_{\rm cut} = 100\,h\,{\rm Mpc}^{-1}$ as our fiducial cut-off scale. We keep the modelling of Ly α forest $P_{\rm 3D}$ out of scope of this estimation, but note that $P_{\rm 1D}$ is highly sensitive to small-scale physics including Jeans smoothing, reionization history (Hui & Gnedin 1997; Gnedin & Hui 1998) and redshift space distortions. Instead, we use our fiducial fit in equation (4), and integrate by parts to find the suppression on $P_{\rm 1D}$.

$$P_{\rm 1D}(k; k_{\rm cut}) = \int_k^\infty \tfrac{q\,{\rm d}q}{2\pi} P_{\rm 3D}(q) T^2(k; k_{\rm cut}) \tag{33}$$

$$P_{\rm 3D}(k) = -\tfrac{2\pi}{k} \tfrac{{\rm d}P_{\rm fit}}{{\rm d}k} \tag{34}$$

Finally, we use Planck 2018 as our fiducial cosmology (Planck Collaboration I 2020), and convert from distance units in Mpc h^{-1} to velocity units in Ly α forest using $k_{{\rm Ly}\,\alpha} = k_{\rm Mpc}(1+z)/H(z)$, which roughly corresponds to a factor of 0.01.

For Iršič et al. (2017b), we add systematic errors in quadrature to the diagonal of the covariance matrix and use all $P_{\rm 1D}$ points. This yields $\sigma_{\rm cut} = 32.9\,h\,{\rm Mpc}^{-1}$. The same process yields $\sigma_{\rm cut} = 20.1\,h\,{\rm Mpc}^{-1}$ for Chabanier et al. (2019). We use metal subtracted result from Walther et al. (2017) and limit to $k < 0.1\,{\rm s\,km}^{-1}$ range. They do not provide a separate systematic error budget, so the covariance matrix is not modified. This yields $\sigma_{\rm cut} = 19.0\,h\,{\rm Mpc}^{-1}$. These numbers are also listed in Table 4.

For the statistical power of our measurement, we limit ourselves to a conservative $0.004\,{\rm s\,km}^{-1} < k < 0.1\,{\rm s\,km}^{-1}$ and $z > 1.8$ range. This analysis with only statistical errors from bootstrap yields $\sigma_{\rm cut} = 6.4\,h\,{\rm Mpc}^{-1}$ for our results. We then add our systematic error budget in quadrature to the diagonal of the covariance matrix, which increases the final error on $k_{\rm cut}$ to $\sigma_{\rm cut} = 7.8\,h\,{\rm Mpc}^{-1}$. Adding the largest scales $k < 0.004\,{\rm s\,km}^{-1}$ does not improve this number as expected. However, $z = 1.8$ bin brings this value down to $5.8\,h\,{\rm Mpc}^{-1}$. To summarize, we expect our results in conservative range to improve the cut-off scale sensitivity by a factor of 2.4.

Interestingly, including XQ-100 at the chunk level actually yields higher $\sigma_{\rm cut} = 7.7\,h\,{\rm Mpc}^{-1}$ with only statistical errors from bootstrap. This increased error from the bootstrap realizations reflects the resolution disparity between data sets.

Table 4. Comparison of different data sets' statistical cut-off scale constraining power in terms of estimated errors, $\sigma_{\rm cut}$, for fiducial $k_{\rm cut} = 100\,h\,{\rm Mpc}^{-1}$. Without systematic error budget, our results can improve cut-off scale sensitivity by a factor of 3. Additional error budget decreases this factor to 2.4.

	$\sigma_{\rm cut}$ [$h\,{\rm Mpc}^{-1}$]
This work (stat. only)	6.4
This work (stat. + syst.)	7.8
Walther et al. (2017)	19.0
Chabanier et al. (2019)	20.1
Iršič et al. (2017b)	32.9

5.2 Remaining challenges

We assumed the noise is uncorrelated throughout, even though QMLE is capable of including pixel level correlations. Correlated noise has two effects. First, it changes the weighting in QMLE. This is not a big effect, since these weights roughly stay constant across k bins. A second effect is the uncorrected spurious power. This is removed through the side band power subtraction, though not exactly because we combine data from different instruments. Furthermore, if the noise is correlated, it should be correlated at the pixel level, which corresponds to very small scales that are outside our conservative range. For example, $k = 0.1\,{\rm s\,km}^{-1}$ corresponds to approximately 20 pixels for KODIAQ and SQUAD. Finally, our error

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

estimates are from bootstraps, which will account for the correlated noise in the error bars. Therefore, we expect our P_{1D} estimates not to be biased due to correlated noise. However, the impact of correlated noise should be revisited as the precision of P_{1D} increases in the future.

One shortcoming of high-resolution quasar observations is that they are targeted for specific studies of e.g. strong absorbers or metallicity distribution. Murphy et al. (2019) note that some UVES quasars were specifically targeted due to known DLAs. The original goal of the KODIAQ survey was to study O VI absorption at $z > 2.2$ (O'Meara et al. 2017). Even though we mask or subtract contaminants, their cross-correlations with the IGM remain. Therefore, unbiased high-resolution quasar observations are crucial to remedy the effect of this sampling bias.

The continuum errors remain a big challenge for P_{1D} measurements at large scales in general, but especially for these types of quasar spectra. Our marginalization limits the shape of the errors to a constant and a slope, and is not fully descriptive given hand-fitted continua. One could imagine marginalizing out higher order terms and trying to find convergence. However, it is possible that this recipe is not convergent and extra degrees of freedom will eventually wipe out all information. We defer a dedicated analysis to future work. Another option would be finding and marginalizing out dominant terms in continuum by a PCA analysis. Although such templates exist, they come with non-negligible errors with a range of 3–30 per cent (Suzuki et al. 2005). This would further require an analytically well-defined, uniform continuum fitting procedure. In short, accurately describing the quasar continuum remains a much-needed but difficult task.

A challenge for P_{1D} analyses is estimating systematic error budget for DLAs and continuum errors. Our data is inhomogeneous, so it is additionally difficult to gauge. Our best understanding showed these errors contaminate low k values. We decided to remove these most contaminated bins from our measurement. As precision increases with future surveys, rigorous description of DLA and continuum systematics (as well as their careful treatments) will be needed. It is best if future surveys build these into their pipeline.

6 SUMMARY

High-resolution, high-S/N quasar spectra can measure 1D Ly α forest power spectrum at smaller scales compared to large-scale structure surveys. At these scales, P_{1D} is able to constrain the thermal state of the IGM and different dark matter models.

We applied the optimal quadratic estimator to the largest such data set by combining two publicly available data releases (KODIAQ + SQUAD) at the spectrum level, and performed a separate analysis for another publicly available data set XQ-100. We then presented the weighted average of these two as our final P_{1D} measurement, and found that it agreed with previous studies with reduced error bars. We identified four systematic error sources for our analysis: incomplete DLA removal, inefficient continuum marginalization, resolution errors, and metals. As an advantage of the quadratic estimator, our method is not biased due to gaps and hence free from the respective systematic error. These four systematics are scale and redshift dependent, but smaller than the statistical errors.

Finally, to demonstrate the constraining power of this work, we performed a crude, single-parameter Fisher forecast analysis for the cut-off scale. We estimate that our results are more sensitive to this cut-off scale by a factor of 2.4 than the previous measurements.

ACKNOWLEDGEMENTS

NGK and NP were supported by DOE, reference number DE-SC0017660. AFR acknowledges support by FSE funds through the program Ramon y Cajal (RYC-2018-025210) of the Spanish Ministry of Science and Innovation. VI was supported by the Kavli foundation. PN acknowledges financial support from Centre National d'Études Spatiales (CNES).

Some of the data presented in this work were obtained from the Keck Observatory Database of Ionized Absorbers toward QSOs (KODIAQ), which was funded through NASA ADAP grants NNX10AE84G and NNX16AF52G along with NSF award number 1516777".

This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico; the Ministry of Economy of Spain, and by the DESI Member Institutions. The authors are honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material. Our quadratic estimator code is publicly available and can be downloaded from https://bitbucket.org/naimgk/lyspeq. Our implementation of further reductions is a collection of personal scripts, and hence less documented. We also make these publicly available for reference. These can be found in https://bitbucket.org/naimgk/qsotools.

REFERENCES

Baur J., Palanque-Delabrouille N., Yèche C., Magneville C., Viel M., 2016, J. Cosmol. Astropart. Phys., 2016, 012
Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, ApJ, 872, 101
Boyarsky A., Lesgourgues J., Ruchayskiy O., Viel M., 2009, J. Cosmol. Astropart. Phys., 2009, 012
Chabanier S. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 017
Croft R. A. C., Hu W., Dave' R., 1999a, Phys. Rev. Lett., 83, 1092
Croft R. A. C., Weinberg D. H., Pettini M., Hernquist L., Katz N., 1999b, ApJ, 520, 1
Dawson K. S. et al., 2016, AJ, 151, 44
Day A., Tytler D., Kambalur B., 2019, MNRAS, 489, 2536
Dekker H., D'Odorico S., Kaufer A., Delabre B., Kotzlowski H., 2000, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Ser. Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. SPIE, Bellingham, p. 534
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 681, 831 (FG08)

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021

Font-Ribera A., McDonald P., Slosar A., 2018, J. Cosmol. Astropart. Phys., 2018, 003
Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2021, MNRAS, 506, 4389
Garzilli A., Magalich A., Theuns T., Frenk C. S., Weniger C., Ruchayskiy O., Boyarsky A., 2019, MNRAS, 489, 3456
Gnedin N. Y., Hui L., 1998, MNRAS, 296, 44
Hamilton A. J. S., 1997, MNRAS, 289, 285
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Iršič V. et al., 2017b, MNRAS, 466, 4332
Iršič V., Viel M., Haehnelt M. G., Bolton J. S., Becker G. D., 2017a, Phys. Rev. Lett., 119, 031302
Karaçayli N. G., Font-Ribera A., Padmanabhan N., 2020, MNRAS, 497, 4742

Kim T.-S., Viel M., Haehnelt M. G., Carswell R. F., Cristiani S., 2004, MNRAS, 347, 355
King J. A., Webb J. K., Murphy M. T., Flambaum V. V., Carswell R. F., Bainbridge M. B., Wilczynska M. R., Koch F. E., 2012, MNRAS, 422, 3370
Lehner N., O'Meara J. M., Fox A. J., Howk J. C., Prochaska J. X., Burns V., Armstrong A. A., 2014, ApJ, 788, 119
Levi M. et al., 2013, preprint (arXiv:1308.0847)
López S. et al., 2016, A&A, 594, A91
Maitra S., Srianand R., Gaikwad P., 2021, MNRAS, 509, 1536
McDonald P. et al., 2006, ApJS, 163, 80
McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, ApJ, 543, 1
McDonald P., Seljak U., Cen R., Bode P., Ostriker J. P., 2005, MNRAS, 360, 1471
Mo H., van den Bosch F., White S., 2010, Galaxy Formation and Evolution. Cambridge Univ. Press, Cambridge
Murphy M. T., Kacprzak G. G., Savorgnan G. A. D., Carswell R. F., 2019, MNRAS, 482, 3458
O'Meara J. M. et al., 2015, AJ, 150, 111
O'Meara J. M., Lehner N., Howk J. C., Prochaska J. X., Fox A. J., Peeples M. S., Tumlinson J., O'Shea B. W., 2017, AJ, 154, 114
Padmanabhan N., White M., Zhou H. H., O'Connell R., 2016, MNRAS, 460, 1567
Palanque-Delabrouille N. et al., 2013, A&A, 559, A85
Palanque-Delabrouille N. et al., 2015a, J. Cosmol. Astropart. Phys., 2015, 045
Palanque-Delabrouille N. et al., 2015b, J. Cosmol. Astropart. Phys., 2015, 011
Planck Collaboration I, 2020, A&A, 641, A1
Rogers K. K., Bird S., Peiris H. V., Pontzen A., Font-Ribera A., Leistedt B., 2018, MNRAS, 474, 3032
Sánchez-Ramírez R. et al., 2016, MNRAS, 456, 4488
Seljak U., 1998, ApJ, 506, 64
Seljak U., Makarov A., McDonald P., Trac H., 2006, Phys. Rev. Lett., 97, 191303
Slosar A. et al., 2013, J. Cosmol. Astropart. Phys., 2013, 026
Suzuki N., Tytler D., Kirkman D., O'Meara J. M., Lubin D., 2005, ApJ, 618, 592
Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, ApJ, 499, 555
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Vernet J. et al., 2011, A&A, 536, A105
Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502
Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Phys. Rev. D, 71, 063534
Vogt S. S. et al., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE Conf. Ser. Vol. 2198, Instrumentation in Astronomy VIII. SPIE, Bellingham, p. 362
Walther M., Hennawi J. F., Hiss H., Oñorbe J., Lee K.-G., Rorai A., O'Meara J., 2017, ApJ, 852, 22
Walther M., Oñorbe J., Hennawi J. F., Lukic Z., 2019, ApJ, 872, 13
Yeche C., Palanque-Delabrouille N., Baur J., BourBoux H. d. M. d., 2017, J. Cosmol. Astropart. Phys., 2017, 047

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

karacayli_etal2021_data.zip

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

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

Downloaded from https://academic.oup.com/mnras/article/509/2/2842/6425772 by UCL (University College London) user on 29 November 2021